

[TYPE THE COMPANY NAME]

# STATISTIK

## TEORI DAN APLIKASI

**Dr. Indasah, Ir., M.Kes**

[Pick the date]



[Type the abstract of the document here. The abstract is typically a short summary of the contents of the document. Type the abstract of the document here. The abstract is typically a short summary of the contents of the document.]

# KATA PENGANTAR

---

# DAFTAR ISI

---

<b>KATA PENGANTAR .....</b>	<b>2</b>
<b>DAFTAR ISI .....</b>	<b>3</b>
<b>BAB 1. PENGANTAR STATISTIK .....</b>	<b>6</b>
A. PENGERTIAN STATISTIK DAN STATISTIKA .....	6
B. PERKEMBANGAN STATISTIKA.....	11
C. PERANAN DAN FUNGSI STATISTIKA .....	19
D. PEMBAGIAN STATISTIKA .....	22
<b>BAB 2. DATA STATISTIK .....</b>	<b>28</b>
A. PENGERTIAN DATA .....	28
B. KARAKTERISTIK DATA .....	29
C. JENIS DATA.....	30
1. Berdasarkan Nilainya.....	30
2. Menurut Sumbernya .....	30
3. Menurut Sumber dan Pengguna.....	30
4. Menurut Cara Memperoleh.....	31
5. Menurut Sifat .....	31
6. Menurut Waktu Pengumpulannya .....	32
D. PENGUMPULAN DATA.....	33
E. PENGUKURAN DAN SKALA DATA .....	34
1. Skala Nominal .....	35
2. Skala Ordinal .....	37
3. Skala Interval .....	39
4. Skala Ratio.....	40
F. PENGOLAHAN DATA.....	42
G. PENYAJIAN DATA .....	42
1. Tabel Data.....	42
2. Grafik Data. ....	42
a) Piktogram.....	43
b) Grafik Batang atau Balok.....	43
c) Grafik Garis.....	43
d) Grafik Lingkaran.....	43
e) Kartogram. ....	43

<b>BAB 3. DISTRIBUSI FREKUENSI</b> .....	44
A. PENGERTIAN DISTRIBUSI FREKUENSI.....	44
B. BAGIAN-BAGIAN DISTRIBUSI FREKUENSI.....	45
C. PENYUSUNAN DISTRIBUSI FREKUENSI.....	45
D. PENYAJIAN GRAFIK DISTRIBUSI FREKUENSI .....	46
1. Tabel Distribusi Frekuensi Data Tunggal.....	47
2. Tabel Distribusi Frekuensi Data Kelompok .....	47
3. Teknik Pembuatan Tabel Distribusi Frekuensi (TDF) .....	54
a) Langkah-langkah dalam menyusun tabel distribusi frekuensi :.....	54
4. Distribusi Frekuensi Relatif dan Kumulatif.....	57
a) Distribusi Frekuensi Kumulatif.....	58
b) Histogram.....	58
c) Poligon Frekuensi .....	59
d) Ogive.....	59
E. JENIS-JENIS DISTRIBUSI FREKUENSI.....	60
 <b>BAB 4. DATA EDITOR (SPSS).</b> .....	62
A. DUA BAGIAN UTAMA DATA EDITOR SPSS.....	62
1. Jendela Data Editor.....	63
2. Jendela Output .....	64
B. PEMBUATAN, PENGISIAN, DAN PENYIMPANAN DATA SPSS .....	65
1. Buka software SPSS .....	66
2. Memasukkan metadata pada variable view. ....	70
3. Menyimpan Data.....	76
C. MENU EDIT PADA SPSS .....	77
D. MENCETAK DATA DAN KELUAR DARI SPSS .....	85
E. INSTALASI SPSS VERSI 18.....	86
 <b>BAB 5. STATISTIK DISKRIPITIF</b> .....	97
A. PENGERTIAN STATISTIK DESKRIPITIF .....	97
B. UKURAN PEMUSATAN DATA .....	97
1. Mean (arithmetic mean).....	97
a) Rata-rata hitung (Mean) untuk data tunggal .....	98
b) Rata-rata hitung (Mean) dari data distribusi Frekuensi atau dari gabungan .....	99
c) Rata-rata Gabungan atau rata-rata terboboti (Weighted Mean).....	101
2. Median .....	101
a) Median data tunggal.....	102

b) Median dalam distribusi frekuensi:.....	103
3. Mode.....	104
a) Modus Data Tunggal:.....	105
b) Mode dalam Distribusi Frekuensi:.....	106
4. Rata-rata Ukur (Geometric Mean).....	107
a) Rata-rata ukur untuk data tunggal.....	107
b) Distribusi Frekuensi:.....	108
c) Rata-rata Harmonik (H).....	108
d) Rata-rata harmonic untuk data tunggal.....	109
e) Rata-rata Harmonik untuk Distribusi Frekuensi:.....	109
6. Karakteristik penting untuk ukuran tendensi sentral yang baik.....	110
C. UKURAN PENYEBARAN DATA.....	111
1. Range.....	112
a) Pengertian Range.....	112
b) Cara Mencari Range.....	113
c) Range Untuk Data Tidak Berkelompok.....	114
d) Range Untuk Data Berkelompok.....	114
e) Penggunaan Range.....	115
f) Kebaikan dan Kelemahan Range.....	115
2. Deviasi.....	116
a) Pengertian Deviasi.....	116
b) Deviasi Rata-rata.....	117
c) Cara Mencari Deviasi Rata-rata.....	117
D. PROSEDUR KERJA ANALISIS DISKRIPSTIF DAN INTERPRETASI OUTPUT ANALISIS DISKRIPSTIF MELALUI SPSS.....	118
1. Entry Data.....	119
2. Frequencies.....	122
3. Descriptives.....	125
5. Crosstabs.....	130
<b>BAB 6. KONSEP DASAR PROBABILITAS.....</b>	<b>133</b>
A. PENGERTIAN PROBABILITAS.....	133
1. Percobaan (Experiment).....	133
2. Hasil (Outcome).....	133
3. Peristiwa (Event).....	133
B. MANFAAT MEMPELAJARI PROBABILITAS.....	134
C. PENDEKATAN PROBABILITAS.....	134

1.	Pendekatan klasik .....	135
2.	Pendekatan relatif .....	135
3.	Pendekatan subjektif .....	136
D.	HUKUM PROBABILITAS .....	136
1.	Hukum penjumlahan.....	136
a)	Peristiwa atau kejadian bersama .....	137
b)	Kejadian saling lepas (mutually exclusive) .....	138
2.	Hukum Perkalian .....	139
a)	Probabilitas bersyarat (conditional probability).....	139
b)	Persitiwa pelengkapp (complementary event) .....	140
c)	Diagram Pohon Probabilitas .....	140
E.	TEOREMA BAYES.....	141
1.	Frequentist versus Bayesian .....	142
2.	Teorema Bayes .....	143
<b>BAB 7. DISTRIBUSI PROBABILITAS.....</b>		<b>147</b>
A.	PENGERTIAN DISTRIBUSI PROBABILITAS .....	147
B.	JENIS-JENIS DISTRIBUSI PROBABILITAS NORMAL BAKU .....	147
1.	Distribusi Probabilitas Berdasarkan Teori.....	147
2.	Distribusi Probabilitas Berdasarkan Subjektif.....	148
3.	Distribusi Probabilitas Berdasarkan Pengalaman.....	149
4.	Distribusi Probabilitas dan Kurva Normal dengan $\mu$ dan $\sigma$ Berbeda .....	149
5.	Distribusi Probabilitas dan Kurva Normal dengan $\mu$ Berbeda dan $\sigma$ Sama.....	150
6.	Distribusi Probabilitas dan Kurva Normal dengan $\mu$ dan $\sigma$ Berbeda .....	150
7.	Distribusi Probabilitas Normal Baku.....	150
C.	KARAKTERISTIK KURVA DISTRIBUSI NORMAL.....	151
D.	PENTINGNYA DISTRIBUSI NORMAL DALAM STATISTIKA .....	151
E.	CIRI-CIRI DISTRIBUSI NORMAL .....	152
1.	Distibusi normal standar .....	152
a)	Cara Ordinat:.....	152
b)	Cara luas.....	153
F.	PENGGUNAAN TABEL DISTRIBUSI NORMAL DAN APLIKASINYA.....	154
1.	Tabel distribusi normal standar terdiri dari kolom dan baris.....	154
2.	Aplikasi Distribusi Normal.....	154
G.	LUAS DAERAH DI BAWAH KURVA NORMAL .....	155
H.	PENERAPAN KURVA NORMAL .....	155
I.	PENDEKATAN NORMAL TERHADAP BINOMINAL.....	156

J.	FAKTOR KOREKSI KONTINUITAS.....	156
K.	VARIABEL ACAK / RANDOM.....	157
	1. Variabel Acak Diskrit.....	157
	2. Variabel Acak Kontinu.....	157
	3. Distribusi Probabilitas Variabel Acak Diskrit.....	158
	4. Fungsi Probabilitas Kumulatif Variabel Acak diskrit.....	159
	5. Distribusi Probabilitas Variabel Acak Kontinu.....	159
	6. Fungsi Probabilitas Kumulatif Variabel Acak Kontinu.....	160
	7. Fungsi Probabilitas Bersama.....	160
	8. Nilai Harapan Dan Varians Dari Variabel Acak Diskrit.....	160
	9. Nilai Harapan dari Fungsi Probabilitas Bersama.....	161
	10. Aturan-aturan dalam Menghitung Nilai Harapan.....	162
	11. Kovarians Dan Aplikasinya Dalam Keuangan.....	162
	a) Kovarians.....	162
	12. Nilai Harapan dari Penjumlahan Dua Variabel.....	162
	13. Varians dari Penjumlahan Dua Variabel.....	162
	a) Portfolio Expected Return dan Fortfolio Risk.....	162
	<b>BAB 8. HIPOTESIS.....</b>	<b>164</b>
A.	PENGERTIAN HIPOTESIS.....	164
B.	KONSEP HIPOTESIS.....	165
C.	MACAM-MACAM HIPOTESIS.....	165
	1. Hipotesis Deskriptif.....	166
	2. Hipotesis Komparatif.....	166
	3. Hipotesis Asosisatif.....	167
	4. Berdasarkan Jenis Parameternya.....	167
	5. Berdasarkan Jumlah Sampelnya.....	168
	6. Berdasarkan Jenis Distribusinya.....	169
	7. Berdasarkan Arah atau Bentuk Formulasi Hipotesisnya.....	170
D.	PENGUJIAN HIPOTESIS.....	171
	1. Menentukan Formulasi Hipotesis.....	171
	2. Menentukan Taraf Nyata ( $\alpha$ ).....	172
	3. Menentukan Kriteria Pengujian.....	172
	4. Menentukan Nilai Uji Statistik.....	173
	5. Membuat Kesimpulan.....	173
E.	JENIS KESALAHAN DALAM HIPOTESIS.....	174
F.	PERUMUSAN HIPOTESIS.....	177

1.	Teori Sebagai Acuan Perumusan Hipotesis.....	177
2.	Fakta Ilmiah Sebagai Acuan Perumusan Hipotesis .....	177
G.	PENGUJIAN HIPOTESA RATA-RATA.....	178
1.	Pengujian Hipotesis Satu Rata-Rata .....	178
a)	Sampel besar ( $n > 30$ ) .....	178
b)	Sampel Kecil ( $n \leq 30$ ) .....	180
2.	Pengujian Hipotesis Beda Dua Rata-Rata .....	183
a)	Sampel besar ( $n > 30$ ) .....	183
b)	Sampel kecil ( $n \leq 30$ ) .....	186
 <b>BAB 9. KORELASI BIVARIATE.....</b>		<b>191</b>
A.	PENGERTIAN KORELASI .....	191
B.	KOEFISIEN KORELASI.....	194
1.	Signifikansi.....	196
2.	Interpretasi Korelasi.....	197
3.	Koefisien Determinasi.....	198
C.	MACAM KORELASI DAN TEKNIK PENGHITUNGAN.....	203
1.	Korelasi Untuk Uji Parametrik .....	203
a)	Uji Pearson Product Moment dan Asumsi Klasik .....	204
b)	Signifikansi Atau P Value Uji Pearson Product Moment .....	207
c)	Asumsi Klasik Uji Pearson Product Moment .....	208
2.	Korelasi Untuk Uji Non Parametrik .....	212
a)	Koefisien Phi.....	212
b)	Korelasi Koefisien Kontingency C .....	239
c)	Korelasi Rank Spearman.....	221
D.	PROSEDUR KERJA DAN INTERPRETASI OUTPUT DI SPSS .....	226
1.	Korelasi Untuk Uji Parametrik .....	226
a)	Uji Pearson Product Moment dan Asumsi Klasik .....	235
2.	Korelasi Untuk Uji Non Parametrik .....	226
a)	Korelasi Phi.....	236
b)	Korelasi Koefisien Kontingency C .....	239
c)	Korelasi Rank Spearman.....	239
 <b>BAB 10. UJI CHI-KUADRAT.....</b>		<b>243</b>
A.	STATISTIK NONPARAMETRIK .....	243
1.	Penggunaan statistik non parametrik .....	243
2.	Kesesuaian Skala Pengukuran Dengan Tes Statistik.....	244



B.	CHI KUADRAT UNTUK UJI GOODNESS OF FIT .....	245
1.	Penetapan Hipotesis Awal dan Hipotesis Alternatif.....	246
2.	Penetapan Derajat Kebebasan (df).....	246
3.	Contoh Kasus uji kecocokan/ goodness of fit test .....	247
C.	Uji Keselarasan dengan Frekuensi Harapan Sama .....	248
1.	“Fisher Exact atau Koreksi Yates” .....	250
2.	Menguji Independensi antara 2 faktor (independensi) .....	251
D.	UJI CHI-KUADRAT UNTUK UJI KENORMALAN .....	255
E.	UJI CHI-KUADRAT UNTUK UJI INDEPENDENSI.....	265
F.	PROSEDUR KERJA DAN CARA INTERPRETASI OUTPUT DARI SPSS.....	267
 <b>BAB 11. REGRESI .....</b>		<b>273</b>
A.	PENGERTIAN REGRESI .....	273
B.	ASUMSI PENGGUNAAN REGRESI.....	276
C.	MACAM REGRESI DAN TEKNIK PENGHITUNGAN.....	276
1.	Regresi untuk Uji Parametrik .....	276
a)	Regresi Linier.....	277
b)	Regresi Linier Berganda .....	279
c)	Regresi Dummy .....	283
2.	Regresi untuk Uji Non Parametrik .....	291
a)	Regresi Logistik .....	291
b)	Regresi Ordinal .....	294
D.	PROSEDUR KERJA DAN INTERPRETASI OUTPUT DARI SPSS.....	300
1.	Regresi untuk Uji Parametrik .....	300
a)	Regresi Linier.....	300
b)	Regresi Linier Berganda .....	304
c)	Regresi Dummy .....	314
2.	Regresi untuk Uji Non Parametrik .....	325
a)	Regresi Logistik .....	325
b)	Regresi Ordinal .....	337
 <b>BAB 12. UJI BEDA BIVARIATE.....</b>		<b>342</b>
A.	UJI BEDA UNTUK UJI PARAMETRIK.....	342
1.	Paired sample t-test (Uji t berpasangan).....	342
a)	Syarat jenis uji ini adalah: .....	343
b)	Rumus t-test yang digunakan untuk sampel berpasangan (paired).....	343
c)	Cara menganalisis dan interpretasi SPSS.....	344

2.	Independent sample t-test (Uji t independent).....	348
	a) Asumsi Independen T Test .....	348
	b) Cara menganalisis dan interpretasi SPSS.....	352
<b>B.</b>	<b>UJI BEDA UNTUK UJI NON PARAMETRIK .....</b>	<b>356</b>
1.	Uji Beda Berpasangan .....	356
	a) Uji Mc Nemar .....	356
	b) Uji Chocran .....	363
	c) Uji Wilcoxon.....	368
	d) Uji Friedman .....	375
2.	Uji Beda Independen .....	382
	a) Uji Mann Whitney .....	382
	b) Uji Kruskal Walis.....	395
<b>BAB 13.</b>	<b>ANALYSIS OF VARIANCE (ANOVA).....</b>	<b>403</b>
<b>A.</b>	<b>PENGERTIAN ANOVA.....</b>	<b>403</b>
<b>B.</b>	<b>KEGUNAAN ANOVA .....</b>	<b>403</b>
<b>C.</b>	<b>CIRI-CIRI ANOVA .....</b>	<b>405</b>
<b>D.</b>	<b>JENIS ANOVA .....</b>	<b>406</b>
	1. Univariat: .....	406
	2. Multivariat: .....	406
<b>E.</b>	<b>ONE WAY ANOVA.....</b>	<b>406</b>
	1. Langkah-langkah dalam perhitungan ANOVA satu jalur: .....	407
	2. Contoh Kasus ANOVA satu arah: .....	408
	3. Cara Analisis dan Interpretasi SPSS.....	410
<b>F.</b>	<b>TWO-WAY ANOVA.....</b>	<b>414</b>
	1. Langkah-langkah dalam perhitungan ANOVAdua jalur (two way ANOVA): .....	415
	2. Hipotesis dalam ANOVA(analysis of variance) dengan interaksi: .....	416
	3. Langkah-langkah melakukan uji hipotesis dengan ANOVA .....	416
	4. Contoh 1 Kasus ANOVAdua arah dengan interaksi: .....	420
	5. Contoh 2 Kasus ANOVA dua arah dengan interaksi: .....	423
	6. Cara Analisis dan Interpretasi SPSS.....	426
<b>G.</b>	<b>MULTIVARIATE WAY ANOVA .....</b>	<b>435</b>
	1. Prosedur MANOVA .....	439
	2. Tahapan dalam MANOVA:.....	442
	3. Cara Analisis dan Interpretasi SPSS.....	449
<b>DAFTAR PUSTAKA.</b>	<b>.....</b>	<b>457</b>
<b>LAMPIRAN-LAMPIRAN</b>	<b>.....</b>	<b>463</b>

# BAB 1 PENGANTAR STATISTIK

---

## A. PENGERTIAN STATISTIK DAN STATISTIKA

Statistik telah dipakai untuk menyatakan kumpulan fakta, umumnya berbentuk angka yang disusun dalam tabel atau diagram, yang melukiskan atau menggambarkan suatu persoalan. Statistik yang menjelaskan sesuatu hal biasanya diberi nama statistik mengenai hal yang bersangkutan. Misalnya kita mengenal : statistik penduduk, statistik kelahiran, statistik pendidikan, statistik produksi, statistik pertanian, statistik kesehatan dan lain sebagainya.

Statistika adalah sekumpulan konsep dan metode yang digunakan untuk mengumpulkan data, menyajikan data, analisis data dan menginterpretasikan data tentang bidang tertentu dan mengambil kesimpulan dalam situasi dimana ada ketidakpastian dan variasi.

Statistika adalah pengetahuan yang berhubungan dengan cara-cara pengumpulan fakta, pengolahan serta penganalisisannya, penarik kesimpulan serta pembuat keputusan yang cukup beralasan berdasarkan fakta dan penganalisan yang dilakukan.

Menurut sejarah kata statistika diambil dari bahasa latin, **Status** yang berarti **Negara**. Untuk beberapa decade statistic semata-mata hanya dikaitkan dengan penyajian angka-angka tentang situasi perekonomian, kependudukan dan politik yang terjadi di suatu Negara. Statistik dalam perkembangannya telah membuat lompatan yang jauh lebih maju daripada hanya sekitar kompilasi grafik-grafik dan table-tabel angka. Sebagai suatu disiplin ilmu saat ini statistika meliputi berbagai metode dan konsep yang sangat penting dalam semua penelitian yang melibatkan pengumpulan data dengan cara eksperimental dan observasi dan mengambil inferensi atau kesimpulan dengan jalan menganalisis data.

Berikut adalah beberapa penjelasan terkait definisi dan pengertian statistika menurut para ahli, baik dari dalam negeri maupun dari luar negeri.

### 1. Menurut Sigit Nugroho (2007)

Pengertian statistika menurut Sigit Nugroho (2007) adalah hasil-hasil pengolahan dan analisis data. Statistik dapat berupa mean, modus, median, dan sebagainya. Statistik dapat digunakan untuk menyatakan kesimpulan data

berbentuk bilangan yang disusun dalam bentuk tabel atau diagram yang menggambarkan karakteristik data.

**2. Menurut Marguerite F. Hall (1892)**

Pengertian statistika menurut Marguerite F. Hall (1892) merupakan suatu teknik yang digunakan untuk mengumpulkan data, menganalisa data dan menyimpulkan dan mengadakan penafsiran data yang berbentuk angka.

**3. Menurut Anderson & Bancroft**

Statistika merupakan ilmu dan seni mengembangkan dan menerapkan metoda yang paling efektif untuk mengumpulkan, mentabulasi, menginterpretasikan data kuantitatif sedemikian rupa sehingga kemungkinan salah dalam kesimpulan dan estimasi dapat diperkirakan dengan menggunakan penalaran induktif berdasarkan matematika probabilitas.

**4. Menurut Sudjana (2001)**

Arti statistika merupakan ilmu pengetahuan yang berhubungan dengan cara-cara pengumpulan fakta, pengolahan serta penganalisanya, penarikan kesimpulan, penyajian dan publikasi dari data yang berbentuk angka.

**5. Menurut Siregar (2004)**

Pengertian statistika menurut Siregar (2004) adalah ilmu yang mempelajari cara-cara mendeteksi suatu objek, mendeskripsikan objek, dan menganalisis setiap aspek-aspek yang mempengaruhi objek, untuk disimpulkan secara ilmiah tentang keberadaan objek, sebagai pedoman sains atau pengambilan keputusan.

**6. Menurut Subana & Sudrajat (2000)**

Definisi statistika menurut Sudrajat merupakan ilmu pengetahuan mengenai cara dan aturan dalam hal pengumpulan data, pengolahan, analisa, penarikan kesimpulan, penyajian dan publikasi dari kata-kata yang berbentuk angka.

**7. Menurut William Mendenhall**

Arti statistika menurut Mendenhall adalah salah satu bidang sains yang berhubungan dengan ekstraksi informasi dari sebuah data numerik dan digunakan untuk membuat keputusan dari suatu populasi dimana data itu didapatkan.

**8. Menurut Kendal & Stuart**

Arti statistik menurut Kendal & Stuart merupakan cabang dari metode ilmiah yang berkaitan dengan pengumpulan data yang dikumpulkan dengan mengukur sifat-sifat dari populasi yang ditemukan.

**9. Menurut Asher**

Pengertian statistik menurut Asher adalah hal-hal yang berkaitan dengan suatu langkah atau metode dalam menarik sebuah kesimpulan dari hasil uji coba.

**10. Menurut Mood, Graybill & Boes**

Definisi statistika adalah suatu teknologi dari salah satu metode ilmiah dan berkaitan dengan percobaan, penyelidikan dan penarikan kesimpulan.

**11. Menurut Stoel Torrie**

Stoel Torrie mengungkapkan bahwa pengertian statistik sebagai cara atau metode yang memberikan langkah-langkah guna untuk menilai ketidakpastian dari sebuah kesimpulan yang sifatnya induktif.

**12. Menurut Freund & Walpole**

Definisi statistik menurut Freund & Walpole adalah salah satu sains dan ilmu pengetahuan dalam pengambilan suatu keputusan yang belum pasti.

**13. Menurut Sugiyono**

Pengertian statistika dalam arti sempit ialah sebagai data dan alat. Sedangkan dalam arti luas, arti statistik merupakan suatu alat dalam menganalisis dan mengambil sebuah keputusan.

**14. Menurut Sutrisno Hadi**

Statistika merupakan salah satu cara untuk mengolah data dan menarik sebuah kesimpulan serta keputusan yang logis dari sebuah proses pengolahan data.

**15. Menurut Agus Irianto**

Pengertian statistik secara singkat adalah sekumpulan cara yang berhubungan dengan pengumpulan data, analisis data, penarikan kesimpulan dari data-data yang berbentuk angka dengan menggunakan asumsi tertentu.

**16. Menurut Suntoyo Yitnosumarto**

Statistika diartikan sebagai sebuah informasi yang menggunakan metodologi dan cara-cara perhitungan dalam menyelesaikan permasalahan-permasalahan praktis yang muncul.

**17. Menurut Iqbal Hasan**

Definisi statistika adalah sebagai suatu ilmu yang mempelajari tentang seluk beluk data yakni tentang tatacara pengumpulan data, pengolahan data, penganalisaan, penafsiran dan penarikan kesimpulan dari suatu data dalam bentuk angka-angka.

### **18. Menurut Singgih**

Statistika merupakan sebuah kegiatan untuk mengumpulkan data, meringkas data, menyajikan data, menganalisa data dengan metode tertentu, dan menginterpretasikan hasil analisis tersebut.

Statistika terdiri dari seni dan ilmu tentang pengumpulan penyajian, analisis dan interpretasi data maupun mengambil kesimpulan (generalisasi) yang masuk akal sehubungan dengan fenomena yang dipelajari/diselidiki. Statistika mempunyai peranan yang sangat penting dalam langkah-langkah pokok metode ilmiah pada tingkat pengumpulan informasi misalnya statistika memberi petunjuk kepada para peneliti bagaimana cara yang wajar dan baik untuk mengumpulkan data yang informative termasuk penentuan macam dan banyak data/sample sedemikian hingga kesimpulan yang ditarik dari analisis data dapat dinyatakan dengan tingkat ketepatan (presisi) yang diinginkan.

Statistik dapat memberikan teknik-teknik yang tepat dalam pengumpulan, pengklasifikasian dan penyajian data, sehingga hasil-hasil penelitian lebih mudah dimengerti. Statistik dapat memberi suatu ukuran yang dapat mensifatkan populasi, menyatakan variasi dan memberikan gambaran yang lebih akurat tentang kecenderungan-kecenderungan suatu variabel penelitian. Statistik dapat digunakan sebagai dasar untuk menjelaskan hubungan antara dua variabel atau lebih.

Statistik memegang peranan yang sangat penting dalam kegiatan penelitian, baik penelitian pendidikan maupun penelitian sosial. banyak masalah-masalah sosial dan masalah-masalah pendidikan baru dapat diterjemahkan jika diberi bobot tertentu dan dianalisis dengan rumus-rumus statistik.

Banyak persoalan, apakah itu hasil penelitian, riset ataupun pengamatan, baik yang dilakukan khusus ataupun berbentuk laporan, dinyatakan dan dicatat dalam bentuk bilangan atau angka-angka. Kumpulan angka-angka sering disusun, diatur atau disajikan dalam bentuk daftar atau tabel. Sering pula daftar atau tabel tersebut disertai dengan gambar yang biasa disebut dengan diagram atau grafik supaya lebih dapat menjelaskan lagi tentang persoalan yang sedang dipelajari. Orang menamakan

fenomena diatas statistik. Jadi, kata statistik dipakai untuk menyatakan kumpulan data, bilangan maupun non bilangan yang disusun dalam tabel atau diagram, yang melukiskan suatu persoalan.

Hasil penelitian, riset ataupun pengamatan, baik yang dilakukan khusus ataupun berbentuk laporan, sering diminta atau diinginkan suatu uraian, penjelasan atau kesimpulan tentang persoalan yang diteliti. Sebelum kesimpulan dibuat, keterangan atau data yang telah terkumpul itu terlebih dahulu dipelajari, dianalisa atau diolah dan berdasarkan pengolahan inilah baru kesimpulan dibuat. Tentulah mudah dimengerti bahwa pengumpulan data atau keterangan, pengolahan dan pembuatan kesimpulan harus dilakukan dengan baik, cermat, teliti, hati-hati, mengikuti cara-cara dan teori yang benar dan dapat dipertanggung jawabkan.

Ini semua ternyata merupakan pengetahuan tersendiri yang diberi nama statistika. Jadi statistika, adalah pengetahuan yang berhubungan dengan cara pengumpulan data, cara penyajian data, cara pengolahan data dan penarikan kesimpulan berdasarkan data dan analisis yang dilakukan.

Banyak persoalan, apakah itu hasil penelitian, riset ataupun pengamatan, baik yang dilakukan khusus ataupun berbentuk laporan, dinyatakan dan dicatat dalam bentuk bilangan atau angka-angka. Kumpulan angka-angka sering disusun, diatur atau disajikan dalam bentuk daftar atau tabel. Sering pula daftar atau tabel tersebut disertai dengan gambar yang biasa disebut dengan diagram atau grafik supaya lebih dapat menjelaskan lagi tentang persoalan yang sedang dipelajari. Orang menamakan fenomena diatas statistik. Jadi, kata statistik dipakai untuk menyatakan kumpulan data, bilangan maupun non bilangan yang disusun dalam tabel atau diagram, yang melukiskan suatu persoalan.

Hasil penelitian, riset ataupun pengamatan, baik yang dilakukan khusus ataupun berbentuk laporan, sering diminta atau diinginkan suatu uraian, penjelasan atau kesimpulan tentang persoalan yang diteliti. Sebelum kesimpulan dibuat, keterangan atau data yang telah terkumpul itu terlebih dahulu dipelajari, dianalisa atau diolah dan berdasarkan pengolahan inilah baru kesimpulan dibuat. Tentulah mudah dimengerti bahwa pengumpulan data atau keterangan, pengolahan dan pembuatan kesimpulan harus dilakukan dengan baik, cermat, teliti, hati-hati, mengikuti cara-cara dan teori yang benar dan dapat dipertanggung jawabkan.

Ini semua ternyata merupakan pengetahuan tersendiri yang diberi nama statistika. Jadi statistika, adalah pengetahuan yang berhubungan dengan cara pengumpulan data, cara penyajian data, cara pengolahan data dan penarikan kesimpulan berdasarkan data dan analisis yang dilakukan.

## B. PERKEMBANGAN STATISTIKA

Perkembangan statistika diawali sebagai suatu ilmu yang membahas cara-cara mengumpulkan angka sebagai hasil pengamatan menjadi bentuk yang lebih mudah dipahami. Menurut Spiegel (1961) statistika berasal dari kata “status” yang berarti negara. Sehingga pada awalnya statistika berkaitan dengan ilmu untuk angka-angka (keterangan) atas perintah raja suatu negara, yang ingin mengetahui kekayaan negaranya, jumlah penduduk, hewan piaraan, hasil pertanian, dan modal. Contoh tertua mengenai hal ini dapat diambil dari zaman Kaisar Agustus yang membuat pernyataan bahwa seluruh dunia harus dikenai pajak, sehingga setiap orang harus melapor kepada statistikawan terdekat (pengumpul pajak). Peristiwa lain di dalam sejarah yang dapat dikemukakan ialah sewaktu William si Penakluk memerintahkan mengadakan pencacahan jiwa dan kekayaan di seluruh wilayah Inggris untuk pengumpulan pajak dan tugas militer. Semua pengamatan dicatat di dalam sebuah buku yang dikenal dengan Domesday Book.

Dari keperluan semacam ini timbullah teknik pencatatan angka-angka pengamatan dalam bentuk daftar dan grafik. Bagian statistika yang membicarakan cara mengumpulkan dan menyederhanakan angka-angka pengamatan ini dikenal sebagai statistika deskriptif. Statistika deskriptif dapat berkembang tanpa memerlukan dasar matematika yang kuat, selain kecermatan dalam teknik berhitung.

Sejak tahun 1700-an analisis data yang dilakukan secara deskriptif berdasarkan tabel-tabel frekuensi, rata-rata, dan ragam untuk sampel (contoh) ukuran besar. Tahun 1800-an merupakan awal penggunaan grafik-grafik untuk penyajian data, seperti histogram, sejalan dengan penemuan sebaran (kurva) Normal. Florence Nightengale (1820-1920) adalah seorang perawat yang terkenal dengan inovasi di bidang ilmu perawatan merupakan pelopor dalam penyajian data secara grafik. Selama perang Crimean, Nightengale mengumpulkan data dan membuat sistem pencatatan. Dari data tersebut dapat ditentukan tingkat mortalitas yang dapat menunjukkan hasil perbaikan kondisi kesehatan yang cenderung



menurunkan tingkat kematian. Selanjutnya data tersebut disajikan dalam bentuk grafik yang merupakan suatu inovasi statistika waktu itu.

Dalam statistika deskriptif tidak ada perbedaan antara data yang diperoleh dari sampel dengan populasinya, kemudian apa yang dihitung dari sampel digunakan untuk menandai populasi. Pada taraf selanjutnya orang tidak puas hanya mengumpulkan angka-angka pengamatan saja. Mereka juga tidak puas bahwa yang diperoleh dari sampel digunakan untuk mencirikan populasi. Timbullah usaha-usaha untuk memperbaiki kesimpulan dalam melakukan ramalan-ramalan populasi berdasarkan angka-angka statistik yang dikumpulkan dari sampel tersebut. Bagian ilmu yang membahas cara-cara mengambil kesimpulan berdasarkan angka-angka pengamatan ini dinamakan statistika induktif.

Karl Pearson, Fisher, Neyman dan Wald selama setengah abad telah meletakkan dasar statistika yang berbasis matematika, sehingga penelitian-penelitian dan kuliah-kuliah statistika di Perguruan Tinggi umumnya didasarkan pada beberapapedoman atau dasar yang ditemukan oleh tokoh-tokoh tersebut. Penggunaan statistika secara luas, terkadang timbul kontroversi diantara para ahli tentang pemilihan model data, penggunaan prior probability dan interpretasi hasil. Hasil analisis terhadap data yang sama dengan lain konsultan statistika dimungkinkan terjadi perbedaan kesimpulan. Statistika induktif dapat dipakai untuk menangani masalah dimana perolehan data dirasakan perlu efisiensi atau perlu biaya mahal, sehingga umumnya dapat diatasi dengan analisis dengan sampel-sampel ukuran kecil.

Di era millenium dengan dominasi teknologi informasi, data base yang besar, interaksi dengan komputer dan informasi yang kompleks, statistika yang berdasarkan pada model-model probabilistik tidak mencukupi, sehingga metode-metode yang akan muncul diarahkan untuk menjawab tantangan zaman yang diberi nama data mining. Istilah data mining (penambangan data) ini menurut Nasoetion (2002) awalnya berasal dari para ahli ilmu komputer yang dalam sehari-harinya bekerja dalam dunia kecerdasan buatan. Untuk pekerjaan ini mereka membangkitkan dan mengumpulkan data dalam ukuran sangat besar dan mencoba menemukan pola-pola keteraturan data yang dapat diterangkan.

Pada tahun 1990-an metode data driven yang tidak terlalu ketat dengan asumsi sebaran mulai digunakan untuk analisis berbagai data, terutama untuk

eksplorasi data atau “data mining”. Berhubung data mining ini sangat computer intensive, maka diusulkan diberi nama Statistical Methods Mining oleh beberapa statistikawan USA. Menurut David M. Rpkce dari University of California metode data mining mempunyai dua prinsip dasar yaitu data cleaning dan data segmentation. Data cleaning untuk mendeteksi data pencilan sedangkan data segmentation untuk pengelompokan data, sehingga akan diketahui pola dari data.

Pada abad 21 diperkirakan metode data mining merupakan metode yang akan banyak digunakan dalam berbagai bidang terapan. Pada metode data mining spesifikasi permasalahan didasarkan pada bidang ilmunya lebih diutamakan daripada pendugaan parameter sehingga masalah tersebut dapat diformulasikan dengan benar untuk memperoleh solusi yang tepat melalui eksplorasi data. Hal ini berbeda dengan periode Fisher yang lebih mementingkan mencari metode pendugaan dan pendekatan sebaran yang tepat, sehingga spesifikasi permasalahan lebih diutamakan pada pendugaan parameter dan asumsi sebaran. Jadi Fisherian Statistics itu sebenarnya model driven yang agak beda dengan data mining yang lebih bersifat data driven. Akan tetapi pada pelaksanaannya, kedua “driven” tersebut harus dikuasai oleh statistikawan di abad millenium ini. Situasi ini akan berpengaruh terhadap model pendidikan dan pengajaran statistika dewasa ini. Emanuel Parzen (Department of statistics Texas A & M University College) baru-baru ini menulis tentang “Data Mining, Statistical Methods Mining and History of Statistics”. Dalam tulisannya tersebut dibahas juga masalah pendidikan statistika menghadapi masa depan dimana data mining akan berkembang, seperti bagaimana cara mengajar matematik statistik untuk non matematik statistik, materi yang berhubungan dengan komputer seperti teknik simulasi, analisis numerik, analisis data dan struktur data perlu ditingkatkan bagi para mahasiswa.

Belum adanya standard analisis untuk eksplorasi data dalam data ukuran besar inilah diperkirakan, metode data mining akan banyak dikembangkan dan diteliti oleh para pakar statistika.

Dilihat dari sejarah pendidikan statistika di Indonesia, Jurusan Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Institut Pertanian Bogor dapat berbangga, karena jurusan yang dirintis dan didirikan oleh Prof. Dr. Andi Hakim Nasoetion (Alm) tahun 1972 adalah Jurusan Statistika tertua di Indonesia. Awalnya dimulai dari Unit Biometrika di bawah Fakultas Pertanian IPB yang

kemudian berubah menjadi Pusat Pengolahan Data Statistika dan Komputasi dan akhirnya menjadi Departemen Statistika dan Komputasi di bawah Fakultas Pertanian. Pada waktu FMIPA disyahkan di IPB pada tahun 1982, namanya berubah menjadi Jurusan Statistika di bawah FMIPA. Jadi boleh dikatakan Departemen Statistika adalah the founding father of FMIPA IPB.

Pada dekade 60 dan 70-an statistika dikenal sebagai “tongkat pembimbing di daerah ketidaktahuan”. Pada dasarnya fungsi tersebut tidak akan hilang, karena statistika tetap berperan di dalam proses penelitian mulai dari rancangan dan analisis, sampai ke penarikan kesimpulan. Di wilayah dimana dunia penuh dengan ketidakpastian, keragaman dan proses acak itulah statistika sangat diperlukan. Tanpa bantuan statistika tidak mustahil kita terjebak oleh kesimpulan yang tidak sepatutnya (misleading conclusion). Selain itu, statistika juga perlu menerawang ke masa depan. Statistika sebagai tongkat pembantu ke masa depan itu wajar saja bila saat ini berkembang moto “Statistika adalah alat bantu untuk memecahkan masalah masa depan”, problem solver of the future.

“Statistics is not just for statistician”, memang demikian adanya. Model-model statistika sangat membantu pemahaman proses pembelajaran dalam dunia pendidikan dan psikologi, regresi dan analisis deret waktu sudah sering membuka tabir kesulitan riset dalam keteknikan, kimia, ekonomi, biologi dan ilmu-ilmu kesehatan. Dewasa ini statistika sering diminta bantuan oleh ahli-ahli hukum kriminalitas, khususnya dengan berkembangnya “statistics for forensic and DNA fingerprinting”.

### C. PERANAN DAN FUNGSI STATISTIKA

Pembahasan mengenai peranan statistik dalam penelitian, sengaja dikedepankan dalam buku ini, mengingat statistik hampir selalu ambil bagian dalam keseluruhan proses kerja peneliti. Adapun peranan-peranan itu adalah sebagai berikut:

Pada era globalisasi, hampir semua bidang tidak terlepas dengan menggunakan angka, data, dan fakta. Hal ini menunjukkan bahwa statistika sangat dibutuhkan. Statistika sebagai sarana mengembangkan cara berpikir logis, lebih dari itu statistika mengembangkan berpikir secara ilmiah untuk merencanakan (forecasting) penyelidikan, menyimpulkan dan membuat keputusan yang teliti dan meyakinkan. Baik disadari atau tidak, statistika merupakan bagian substansi dari latihan profesional

dan menjadi landasan dari kegiatan-kegiatan penelitian. Statistika berperan dalam berbagai kegiatan hidup manusia, antara lain:

### **1. Dalam aktivitas kehidupan sehari-hari**

Dalam aktivitas kehidupan sehari-hari manusia dihadapkan pada berbagai keterangan serta bahan-bahan yang berbentuk angka-angka yang perlu ditafsirkan dan alat bantu yang berperan dalam menafsirkan bahan keterangan dan bahan-bahan yang berbentuk angka tersebut adalah statistika.

### **2. Dalam ilmu pengetahuan**

Dalam ilmu pengetahuan akan didapati penyajian data-data dalam bentuk angka-angka, sehingga diperlukan statistika dalam menafsirkan dan menyimpulkan data tersebut.

### **3. Statistik Sebagai Alat Perumusan Hipotesis**

Kebanyakan peneliti kependidikan dan penelitian sosial pada umumnya dimaksudkan untuk menguji hipotesis. Hipotesis merupakan pernyataan mengenai hubungan atau perbedaan beberapa gejala. Rumusan hipotesis pada hakikatnya mengacu kepada perumusan model matematik. Untuk membuktikan apakah hipotesis yang dirumuskan oleh peneliti diterima atau di tolak, perlu diuji dengan rumus statistik.

### **4. Statistik Sebagai Alat Pengembangan Instrumen**

Instrumen peneliti harus diuji validitas dan reliabilitasnya. Untuk uji validitas dan reabilitas tersebut diperlukan rumus-rumus statistik. Instrumen yang tidak diuji validitas dan reliabilitasnya, diragukan kedudukannya sebagai alat menjaring data yang dapat dipertanggungjawabkan secara ilmiah. Meskipun demikian, memang betul ada sejumlah instrumen penelitian yang tidak memerlukan uji validitas dan reliabilitas, seperti: blangko isian, format-format, dan sejenisnya.

### **5. Statistik Sebagai Alat Penyusun Desain Penelitian**

Penelitian eksperimental, misalnya, perlu disajikan dengan ketentuan ilmiah yang benar. Itu sebabnya, maka desain penelitian eksperimental tersebut perlu diletakkan pada landasan statistik yang benar. Pertimbangan-pertimbangan statistiklah yang menentukan apakah desain itu memenuhi ketentuan metodologis atau tidak.

## **6. Statistik Sebagai Alat Penentuan Sampel Penelitian**

Populasi penelitian terkadang sangat besar jumlahnya, mulai dari satu sampai tak jelas batasnya. Dalam hal dimana populasi penelitian sangat banyak, teknik sampling mutlak diperlukan, mulai dari ukuran sampel (sample size) sampai teknik pengambilan dari jumlah sampel itu. Rumus statistik yang dituangkan merupakan salah satu contoh. Bagi populasi penelitian yang relatif kecil, penentuan sampel sering tak mengalami kesulitan. Sampel penelitian diambil sebesar jumlah populasi. Namun di sini diperlukan asumsi-asumsi statistik untuk menentukan keputusan itu.

## **7. Statistik Sebagai Alat Analisis Data**

Mengolah data merupakan kegiatan penting dalam penelitian. Data yang perlu diolah dengan benar dan dengan penerapan rumus-rumus yang benar pula. Tidak hanya menggunakan formula itu yang perlu benar, tapi juga asumsi-asumsi yang mendasari digunakannya formula tersebut. Dengan statistik pula, data penelitian diklasifikasi, dianalisis selanjutnya disimpulkan. Data yang benar, tapi diolah secara salah tak punya arti yang memadai. Hanya merupakan kumpulan data semata.

## **8. Statistik Sebagai Alat Penarikan Kesimpulan**

Penelitian yang baik dan ilmiah kepada kesimpulan yang ilmiah pula. Statistik punya peran penting dalam penarikan kesimpulan ini. Data tertentu diolah dengan statistik tertentu dan menghasilkan kesimpulan penelitian yang tertentu pula. Misalnya, untuk penelitian yang bersifat uji hipotesis. Uji hipotesis dengan rumus statistik tertentu. Hasil uji hipotesis itu dijadikan dasar penarikan kesimpulan, dimana sebelumnya ditentukan, apakah hipotesis itu diterima atau ditolak.

## **9. Statistik Sebagai Pemberi Arah Kerja Penelitian**

Penelitian kuantitatif tak lepas dari penggunaan rumus-rumus statistik, mulai dari penentuan hipotesis, pengembangan instrumen, penyusunan desain, menganalisis data, dan sebagainya. ada beberapa catatan penting bagi kandidat sarjana yang akan melakukan tugas penelitian sebagai berikut:

- a. Statistik umumnya diperlukan dalam kerja penelitian. Untuk itu penelitian harus diletakkan pada konteks metodologis yang benar.

- b. Namun penelitian yang terlalu mengutamakan segi metodologis, seringkali mengabaikan kedalaman materi yang diteliti.
- c. Konklusinya, bahwa penelitian harus diletakkan diatas metodologi yang benar dan dengan kedalaman materi yang benar pula secara leilmuan.

### **10. Peranan Statistik Dalam Penyusunan Model Teoretis**

Dalam usaha memecahkan masalah penelitian, mula-mula orang belum mempunyai gambaran yang jelas dan detail mengenai keadaan sesungguhnya.berdasarkan penelaahan kepustakaan, apa yang dimilikinya adalah gambaran garis besar, gambaran mengenai pokok-pokoknya saja, yang merupakan abstrak dari keadaan yang sesungguhnya. Dewasa ini model yang paling banyak digunakan adalah model matematis, yaitu model yang menggunakan hukum-hukum matematis, yaitu model yang sebagai dasarnya. Model matematis ini mempunyai beberapa kelebihan-kelebihan itu antara lain:

- a. Dengan model matematis orang dapat merumuskan masalah dengan lebih singkat dan padat, sehingga struktur masalah menjadi lebih terungkap, dan hubungan antara komponen-komponen lebih jelas.
- b. Dengan model matematis orang lebih mudah melakukan kuantifikasi.
- c. Dengan penggunaan model matematis akan lebih memudahkan penggunaan teknik analisis statistik dan penggunaan jasa komputer.
- d. Dengan model matematik lebih mudah dapat dilihat apakah asumsi-asumsi yang mendasari berbagai komponen dalam penelitian itu terpenuhi atau tidak.

### **D. PEMBAGIAN STATISTIK**

Agar mendapatkan informasi lebih terperinci, kita memerlukan analisa data dengan metode statistik tertentu. Hasil analisis data akan memberikan informasi lebih rinci sehingga kita memperoleh suatu kesimpulan mengenai suatu fenomena berdasarkan sampel yang diambil. Analisa tersebut dinamakan statistika inferensial atau statistika induktif.

Ada beberapa teknik statistic yang dapat digunakan untuk menganalisis data. Tujuan dari analisis data adalah mendapatkan informasi relevan yang terkandung didalam suatu data tersebut dan menggunakan hasilnya untuk memecahkan suatu

masalah. Permasalahan yang ingin dipecahkan biasanya dinyatakan dalam bentuk satu atau lebih hipotesis nol. Sampel data yang dikumpulkan kemudian digunakan untuk menguji penolakan atau penerimaan hipotesis nol secara statistic. Jadi hipotesis nol menggambarkan permasalahan dan “informasi relevan” yang terkandung di dalam data yang digunakan untuk menguji secara statistic hipotesis nol.

### 1. Berdasarkan orientasi pembahasan

- a. **Statistika matematik:** statistika teoritis yang lebih berorientasi kepada pemahaman model dan teknik-teknik statistika secara matematis teoritis.
- b. **Statistika terapan:** statistika yang lebih berorientasi kepada pemahaman intuitif atas konsep dan teknik-teknik statistika serta penggunaannya di berbagai bidang

### 2. Berdasarkan tahapan dan tujuan analisis

#### a. Statistik Deskriptif atau Statistik Deduktif

Statistika yang berkenaan dengan metode atau cara mendeskripsikan, menggambarkan, menjabarkan, atau menguraikan data sehingga mudah dipahami. Bidang atau bagian dari pengetahuan statistik yang bertugas mempelajari tata cara pengumpulan data, pencatatan, penyusunan, dan penyajian data penelitian dalam bentuk tabel frekuensi atau grafik. Macam grafik yaitu Histogram adalah grafik yang berbentuk batang, Frekuensi Polygon adalah grafik yang berbentuk garis dari distribusi frekuensi dengan data kontinyu, sedangkan Ogive adalah grafik yang berbentuk garis dari distribusi kumulatif dengan data kontinyu. Kemudian dilakukan pengukuran/perhitungan untuk mengetahui nilai-nilai statistik sebagai berikut :

- 1) Pengukuran *Nilai-Nilai Tendensi Pusat* seperti Mean (rata-rata), Median (nilai yang letaknya sedemikian rupa sehingga memisahkan nilai-nilai menjadi dua bagian sama besar ) dan Modus (Nilai yang paling sering muncul).
- 2) Pengukuran *Nilai Fractile* yaitu menentukan nilai sedemikian rupa sehingga memisahkan nilai-nilai tersebut menjadi 4 bagian (Quartile), 10 bagian (Desil) ataupun 100 bagian (Presentil).

- 3) Pengukuran *Nilai Dispersi* adalah pengukuran penyebaran nilai-nilai pengamatan disekitar tendensi pusatnya. Dalam pengukuran dispersi faktor yang pertama harus diperhatikan adalah tingkat keseragaman (homogenitas) dari sekumpulan nilai pengamatannya. Semakin homogen akan semakin kecil dispersinya. Beberapa macam tehnik pengukuran dispersi adalah Range (interval), Deviasi Kuartil, Mean Deviation, Deviasi Standard, Variance dan Koefisien Variance.
- 4) Pengukuran *Skewness*. Kurve suatu distribusi dapat berbentuk simetris atau tidak simetris (disebut menceng, condong atau juling). Suatu distribusi frekuensi yang tidak simetris mungkin berat sebelah kanan (ekor kiri lebih panjang dari sebelah kanan) yang dinamakan *Skewness negatif*. Atau dapat pula berat sebelah kiri (ekor sebelah kanan lebih panjang dari kiri) yang disebut *Skewness positif*. Misalnya distribusi penghasilan penduduk umumnya menceng positif, artinya sebagian besar penduduk itu tingkat penghasilannya rendah. Dapat dipakai sebagai pedoman untuk mengetahui apakah kurve suatu distribusi itu simetris atau tidak simetris ialah bagaimana mean, median dan modus dari distribusi tersebut. Apabila nilai mean, median dan modus sama besar maka kurve distribusi frekuensinya simetris, dan sebaliknya apabila tidak sama maka distribusi frekuensinya tidak simetris.
- 5) Pengukuran *Kurtosis* adalah ukuran tentang keruncingan (peakedness) dari suatu distribusi frekuensi. Terdapat tiga ukuran keruncingan yaitu
- a) *Leptokurtik*, kurve distribusi ini menyempit pada bagian pucaknya atau mendekati runcing. Keadaan ini menunjukkan bahwa frekuensi cenderung tertumpuk pada daerah sekitar nilai mean atau menunjukkan hanya sedikit frekuensi yang menyebar lebih jauh dari nilai tendensi pusat.
  - b) *Platikurtik*, kurve distribusi ini agak mendatar (tumpul) pada bagian puncak, yang menunjukkan adanya frekuensi agak tersebar merata pada seluruh kelas, kecuali pada beberapa kelas dari bagian pertama dan terakhir.
  - c) *Mesokurtik*, kurve distribusi ini normal, berada di antara leptokurtik dan platikurtik.



Statistik deskriptik dapat dilakukan pada data tunggal atau data berkelompok. Sedangkan penarikan kesimpulan (kalaupun ada) pada statistik deskriptif hanya ditujukan pada kumpulan data yang ada, bukan untuk tujuan generalisasi.

**b. Statistik Induktif atau Statistik Inferensial.**

Bidang atau bagian dari pengetahuan statistik yang bertugas mempelajari tata cara penarikan kesimpulan mengenai keseluruhan populasi berdasarkan data hasil penelitian pada sampel (bagian dari populasi). Didalamnya berisi bagaimana cara membuat estimasi harga parameter, bagaimana cara menguji hipotesis, bagaimana membuat prediksi berdasarkan hubungan pengaruh antara variabel-variabel dan perhitungan derajat asosiasi antara variabel-variabel.

Pengertian tersebut diatas menunjukkan bahwa ruang lingkup statistik induktif lebih luas dibandingkan dengan statistik deskriptif. Penarikan kesimpulan pada statistik induktif merupakan generalisasi dari suatu populasi berdasarkan data sampel. Jadi statistik induktif diperlukan karena peneliti tidak menyelidiki seluruh elemen populasi melainkan hanya mendasarkan pada penelitian sebagian elemen populasi yang disebut *sampel*. Dan karena tujuan penelitian pada hakekatnya untuk membuat generalisasi tentang populasi maka penarikan sampel harus dilakukan dengan benar guna memperoleh sampel yang representatif ( benar-benar mewakili ). Selanjutnya karena penelitian hanya didasarkan pada pengamatan sampel maka peneliti tidak dapat menduga harga parameter (ukuran) populasi dengan pasti dan tidak dapat mengambil kesimpulan apakah hipotesis benar atau salah. Pendugaan harga parameter dan pengujian hipotesis tersebut dilakukan berdasarkan teori-teori probabilitas yaitu mendasarkan pada taraf kepercayaan (confidence level) atau taraf signifikansi (level of significance) tertentu.

**3. Berdasarkan ruang lingkup penggunaannya**

**a. Statistik Pendidikan**

Statistik pendidikan adalah statistik yang digunakan atau diterapkan pada bidang atau disiplin ilmu pendidikan.

**b. Statistik Sosial**

Statistik sosial adalah statistik yang digunakan atau diterapkan pada bidang atau disiplin ilmu sosial.

**c. Statistik Kesehatan**

Statistik kesehatan adalah statistik yang digunakan atau diterapkan pada bidang atau disiplin ilmu kesehatan.

**d. Statistik Ekonomi**

Statistik ekonomi adalah statistik yang digunakan atau diterapkan pada bidang atau disiplin ilmu ekonomi.

**e. Statistik Pertanian**

Statistik pertanian adalah statistik yang digunakan atau diterapkan pada bidang atau disiplin ilmu pertanian.

**4. Berdasarkan asumsi distribusi yang digunakan**

Pengambilan kesimpulan mengenai keseluruhan populasi yang didasarkan data yang ada dari sampel membutuhkan asumsi persyaratan-persyaratan atau kondisi-kondisi tertentu. Dalam statistik induktif, asumsi atau persyaratan ini adalah bahwa bentuk distribusinya diketahui, misalnya menyebar secara normal. Statistik induktif memenuhi persyaratan yang demikian termasuk dalam statistik parametrik. Apabila asumsi, persyaratan atau kondisi tersebut tidak dipenuhi, yaitu bahwa bentuk distribusi populasinya tidak diketahui maka dipakailah statistik non-parametrik. Jadi berdasarkan asumsi yang mendasarinya, statistik induktif (inferensial) dibedakan menjadi dua yaitu

**a. Statistik Parametrik.**

Pendugaan dan uji hipotesis harga parameter populasi didasarkan anggapan bahwa skor-skor yang dianalisis telah ditarik dari suatu populasi dengan distribusi tertentu, misalnya dari populasi yang berdistribusi normal. Biasanya berlaku dalam penelitian dengan data terukur (measurement data dengan skala interval atau skala rasio) dan sampelnya cukup besar.

**b. Statistik Non-Parametrik.**

Pendugaan dan uji hipotesis harga parameter populasi didasarkan anggapan bahwa skor-skor yang dianalisis telah ditarik dari suatu populasi dengan bebas sebaran (tidak mengikuti distribusi tertentu) misalnya dari

populasi yang berdistribusi normal. Biasanya berlaku dalam penelitian dengan data pencacahan (enumeration data dengan skala nominal atau skala ordinal) dan sampel-sampelnya kecil.

Pada umumnya setiap teknik pengujian data dengan teknik statistika parametrik mempunyai teknik padanannya pada statistika non parametrik. Teknik padanan pada statistika non parametrik biasa digunakan apabila data interval/rasio tidak memenuhi asumsi-asumsi tertentu, misalnya data tidak berdistribusi normal. Sebagai contoh apabila data yang akan di analisis dengan menggunakan Uji-F (Anova) tidak memenuhi asumsi-asumsi ANOVA(additive, normalitas, homoskedastisiti, independensi) meskipun sudah dilakukan transformasi, maka alternatif terakhir kita bisa mengujinya dengan menggunakan Uji Kruskal-Wallis (One Way ANOVA–RAL) atau Uji Friedman (RAK) yang merupakan teknik statistika nonparametrik.

## **5. Berdasarkan jumlah variabel**

- a. Statistika Univariat:** teknik analisis statistik yang hanya melibatkan satu variabel dependent
- b. Statistika Multivariat:** teknik analisis statistik yang melibatkan lebih dari satu variabel dependent sekaligus.

# BAB 2 DATA STATISTIK

---

## A. PENGERTIAN DATA

Data adalah bentuk jamak dari datum. Data itu sendiri merupakan keterangan-keterangan tentang suatu hal, dapat berupa sesuatu yang diketahui ataupun dianggap. Jadi dapat diartikan bahwa data itu adalah sebagai sesuatu yang diketahui atau yang dianggap (anggapan).

Data atau data statistik adalah keterangan atau fakta mengenai sesuatu persoalan bisa berbentuk kategori, misalnya rusak, baik, senang, puas, berhasil, gagal, dan sebagainya, atau juga berbentuk bilangan. Data dalam bilangan disebut data kuantitatif dan non bilangan disebut data kualitatif., harganya dapat berubah-ubah atau bersifat variabel.

Data merupakan kumpulan fakta atau angka atau segala sesuatu yang dapat dipercaya kebenarannya sehingga dapat digunakan sebagai dasar penarikan kesimpulan.

Kualitas kue ditentukan oleh bahan dasar dan cara memasaknya. Demikian halnya, kualitas kesimpulan penelitian ditentukan oleh data (bahan dasar) dan teknik analisis dan penyajiannya.

- Contoh data: Harga sekuritas, Kurs mata uang asing, Harga komoditas, Angka inflasi, Harga saham, Jumlah produk per hari.

Seperti kita ketahui bahwa statistika diterapkan untuk mengumpulkan menyajikan menganalisis dan menginterpretasikan data. Data yang kita kumpulkan dapat bersifat kualitatif maupun kuantitatif statistika khususnya bekerja dengan data kuantitatif atau data kualitatif yang sudah dikuantitatifkan dengan berbagai cara.

Data kuantitatif adalah fakta yang dipresentasikan dengan angka. Misalnya penghasilan keluarga dalam rupiah (Rp), berat sapi dalam Kg, tinggi badan dalam Cm, lama hidup suatu mikroorganisme dalam jam dan sebagainya. Data kualitatif adalah fakta yang dinyatakan dalam bentuk sifat (bukan angka). Misalnya jenis babi yaitu babi local babi saddle back, babi landrace, babi duroc dan sebagainya; jenis kandang yaitu kandang induk, kandang pejantan, kandang anak, kandang penggemukan dan sebagainya. Data kualitatif dapat kita kuantitatifkan antara lain dengan cara memberi skor, ranking, variable boneka (dummy variable) dan sebagainya

## B. KARAKTERISTIK DATA

Sebagai ilmu pengetahuan, statistika mempunyai karakteristik sebagai berikut:

### 1. Statistika selalu bekerja dengan angka atau bilangan yang disebut dengan data kuantitatif.

Hal ini dimaksudkan apabila statistika dipergunakan sebagai alat analisis bagi data kualitatif (bahan/keterangan yang tidak berwujud angka atau bilangan), maka data kualitatif tersebut harus diubah atau dikonversikan terlebih dahulu menjadi data kuantitatif, proses ini disebut kuantifikasi. Angka-angka ini dalam statistik terbagi menjadi dua bagian, yaitu :

- a) Angka statistik sebagai jumlah atau frekuensi dan angka statistik sebagai nilai atau harga. Pengertian ini mengandung arti bahwa data statistik adalah data kuantitatif.
  - Contoh : jumlah buruh di Perusahaan Sinar Dunia , jumlah tanggungan orang tua, jumlah penduduk Indonesia tiap tahunnya. , jumlah pegawai negeri di Kabupaten Klungkung.
- b) Angka statistik sebagai nilai, dimaksudkan bahwa data tersebut adalah data kualitatif yang diwujudkan dalam angka.
  - Contoh : hasil ujian nasional siswa SMA, nilai keuntungan dari perusahaan , nilai indeks prestasi mahasiswa , daftar harga Hand Phone Celuler.

### 2. Statistika bersifat obyektif

Kesimpulan dan ramalan yang dihasilkan oleh statistika didasarkan pada angka yang diolah (obyektif) dan tidak didasarkan pengaruh dari luar (subyektif). Statistik yang bersifat obyektif merupakan statistik yang bekerja sesuai dengan keadaan data yang diteliti atau sesuai berdasarkan atas kenyataan yang ada atau fakta.

- Contoh : Pengambilan suatu data mengenai jumlah penduduk Indonesia apakah setiap tahunnya jumlah penduduk Indonesia bertambah ataukah mengalami penurunan. Dengan data yang terkumpul dapat diketahui apakah setiap tahun penduduk Indonesia bertambah atau berkurang.

### 3. Statistika bersifat universal

Ruang lingkup statistika tidaklah sempit, ruang lingkungnya sangat luas dalam kehidupan manusia baik dibidang perdagangan, pertanian, kependudukan, pendidikan, dan sebagainya.

- Contoh : statistik kesehatan, statistik perusahaan, maupun statistik keuangan.

## C. JENIS DATA

### 1. Berdasarkan Nilainya:

- a) Data dari variabel diskrit atau disingkat dengan *data diskrit*, didapat dengan proses *Menghitung* atau *membilang*. Contoh: Kabupaten A sudah membangun 3010 unit rumah Tipe SS dan PT B sudah mewisuda 1200 mahasiswa dengan rincian 670 laki-laki dan 530 wanita.
- b) Data dari variabel kontinu atau disebut *data kontionu*, didapat dengan proses *mengukur*. Contoh: Luas daerah C adalah 3076 km<sup>2</sup> dan Rata-rata debit air sungai Serayu-mrica pada bulan Mei 1998 adalah 112 m<sup>3</sup>/det.

### 2. Menurut Sumbernya

- a) Data intern, Pengusaha mencatat segala aktivitas perusahaannya sendiri, misalnya : keadaan pegawai, pengeluaran, keadaan barang digudang, hasil penjualan, keadaan produksi pabriknya dan lain-lain aktivitas yang terjadi di dalam perusahaan itu.
- b) Data ekstern, untuk perbandingan misalnya, diperlukan data dari sumber lain di luar perusahaan tadi.
- c) Data mentah, Data yang baru dikumpulkan dan belum pernah mengalami pengolahan apapun.

### 3. Menurut Sumber dan Pengguna

#### a) Data Internal

Data yang menggambarkan keadaan atau kegiatan suatu badan yang dikumpulkan sendiri dan hasil datanya digunakan oleh badan itu sendiri.  
contoh:

- Data pengeluaran keuangan untuk membayar biaya produksi perusahaan tekstil
- Data hasil produksi pabrik mie “sedaap”

b) Data Eksternal

Data yang menggambarkan keadaan atau kegiatan di luar badan dan data tersebut tidak terdapat dalam aktivitas intern suatu badan. Contoh:

- Bagi perusahaan “LG”, data daya beli masyarakat terhadap barang produksinya (seperti TV “Turbo Swing”) adalah data eksternal perusahaan tersebut
- Data tingkat kepuasan masyarakat terhadap barang produksi menjadi tolok ukur dalam mengembangkan daerah pemasaran

#### 4. Menurut Cara Memperoleh

a) Data Primer

Data yang dikumpulkan dan diolah sendiri oleh suatu badan secara langsung serta diterbitkan oleh badan itu pula. Contoh:

- Sensus penduduk oleh BPS, dihasilkan data primer langsung dari penduduk
- Data pengeluaran beras di gudang penyimpanan BULOG

b) Data Sekunder

Data yang dilaporkan oleh suatu badan sedang badan ini tidak secara langsung mengumpulkan sendiri tapi diperoleh dari pihak lain yang telah mengumpulkannya.

Contoh:

- Data kenaikan atau penurunan nilai tukar rupiah terhadap mata uang asing dari BEJ
- Pemeriksaan dan pendataan ulang barang impor di pelabuhan dari pihak bea cukai

#### 5. Menurut Sifat

a) Data Kualitatif

Kemungkinan observasi yang tidak dinyatakan dengan angka-angka. Contoh:

- Nilai rupiah sangat kuat

- Pengangguran dan kemiskinan meningkat tajam

b) Data Kuantitatif

Serangkaian observasi atau pengajaran yang dapat dinyatakan dengan angka-angka. Contoh:

- Nilai rupiah Rp 9.250,00 per US\$ di akhir tahun 2006
- Jumlah pengungsi akibat banjir di Jakarta sebanyak 1423 jiwa

**6. Data kuantitatif terbagi atas:**

a) Data Diskrit

Data yang hanya mempunyai sejumlah terbatas nilai-nilai. Contoh:

- Jumlah mahasiswa di sebuah universitas
- Banyaknya masyarakat yang memakai kendaraan roda dua di daerah Purwokerto

b) Data Kontinu

Data yang secara teoritis dapat menjalani setiap nilai. Disebut juga nilai pengamatan kuantitatif kontinyu. Contoh:

- Pengukuran debit air di bendungan
- Pengukuran tingkat curah hujan di daerah Bandung

**7. Menurut Waktu Pengumpulannya**

a) Data Cross Section

Data yang dikumpulkan pada suatu waktu tertentu yang bisa menggambarkan keadaan atau kegiatan pada waktu tersebut. Contoh:

- Data jumlah TKI yang meninggal pada tahun 2006 akibat kekerasan menggambarkan kurangnya perlindungan keselamatan TKI di luar negeri
- Bencana meluasnya lumpur lapindo menandakan kurang seriusnya pemerintah dalam menangani korban bencana tersebut.

b) Data Time Series

Data yang dikumpulkan dari waktu ke waktu sehingga ada perkembangannya (trend) yang menunjukkan arah secara umum. Garis trend sangat berguna untuk membuat ramalan (forecasting) yang dibutuhkan bagi perencanaan. Contoh:



- Data persebaran penduduk di Indonesia dibutuhkan untuk perencanaan transmigrasi sebagai upaya pemerataan jumlah persebaran di tiap daerah
- Data tingkat curah hujan tiap tahunnya dibutuhkan untuk mengantisipasi datangnya tanah longsor atau banjir

#### D. PENGUMPULAN DATA

Data diukur secara langsung dan tidak sedikit data yang tidak dapat diukur secara langsung. Untuk data yang tidak dapat diukur secara langsung harus kita buat secara operasional dapat diukur. Operasionalisasi ini berarti harus diusahakan untuk memecah atau menguraikan pengertian itu dalam sejumlah dimensi yang dapat diukur. Misalnya operasionalisasi status social ekonomi masyarakat menjadi dimensi pendapatan dan dimensi pekerjaan. Dalam mengukur fakta validitas pengukuran harus diusahakan sebaik mungkin. Sebagai contoh apakah daging ayam yang diawetkan dengan suatu zat pengawet masih disenangi oleh konsumen, dapat diukur dengan skala pengukuran sangat disukai, disukai, sedikit suka, suka, biasa saja, sedikit tidak suka, tidak suka dan sangat tidak suka. Jika kita mengukur berat kambing maka dapat digunakan timbangan yang sudah punya skala. Demikian juga untuk mengukur suhu tubuh dapat digunakan thermometer yang sudah ada skalanya.

1. Data harus betul-betul jujur, yakni kebenaran-nya harus dapat dipercaya.
2. Proses pengumpulan data dapat dilakukan dengan jalan *sensus* atau *sampling* dengan langkah yang dapat ditempuh, antara lain :
  - a) Mengadakan penelitian langsung ke lapangan atau laboratorium
  - b) Mengambil atau menggunakan, sebagian atau seluruhnya, dan sekumpulan data yang telah dicatat atau dilaporkan oleh badan atau orang lain.
  - c) Mengadakan angket : yakni cara pengumpulan data dengan menggunakan isian atau daftar pertanyaan yang telah disiapkan dan disusun sedemikian rupa sehingga calon responden hanya tinggal mengisi atau menandainya dengan mudah dan cepat.
3. Perlu memperhatikan bahwa batas-batas populasi harus diketahui dan ditentukan dengan jelas dan tegas.

Pengumpulan data dapat dibedakan dari beberapa jenis, yaitu :

## 1. Berdasarkan Karakteristiknya

### a) Pengamatan (Observasi).

Pengamatan atau peneliti lapangan ini adalah cara pengumpulan data dengan terjun langsung atau melihat langsung ke lapangan(laboratorium) terhadap objek yang diteliti(populasi).

### b) Penelusuran Litelatur.

penelusuran litelatur atau Pengamatan tidak langsung iyalah cara pengumpulan data dengan menggunakan sebagian ataupun semua data yang ada(laporan data) dari peneliti sebelumnya.

### c) Penggunaan Kuesioner(angket).

penggunaan Kuesioner iyalah cara pengumpulan data dengan menggunakan daftar pertanyaan(angket) ataupun daftar isian terhadap objek yang diteliti(Populasi).

### d) Wawancara(interview).

wawancara iyalah cara pengumpulan data dengan langsung mengadakan banyak pertanyaan kepada objek yang diteliti atau pada perantara yang mengetahui dari objek yang diteliti.

## 2. Berdasarkan Banyaknya Data Yang Diambil.

### a) Sensus.

Sensus iyalah cara pengumpulan data dengan mengambil anggota populasi secara keseluruhan untuk diselidiki, Data yang diambil melalui pengumpulan data sensus ini adalah data sebenarnya(true value) atau parameter.

### b) Sampling.

Sampling iyalah cara pengumpulan data dengan mengambil sebagian saja dari anggota populasi untuk diselidiki. data yang diperoleh dari pengumpulan data sampling ini disebut dengan "*statistic*" tanpa(s) atau "*perkiraan (estimate value)*"

## E. PENGUKURAN DAN SKALA DATA

Skala pengukuran data sama dengan skala data pada dasarnya dimaksudkan untuk mengklasifikasikan Variabel yang akan diukur agar tidak terjadi kesalahan dalam menentukan teknik analisis data dan tahapan penelitian selanjutnya.

Skala pengukuran data merupakan seperangkat aturan yang diperlukan untuk ‘*mengkuantitatifkan*’ data dari pengukuran suatu variabel. Dalam melakukan analisis statistik, perbedaan jenis data sangat berpengaruh terhadap pemilihan model atau alat uji statistik. Tidak sembarangan jenis data dapat digunakan oleh alat uji tertentu. Untuk itu skala pengukuran data (variabel) sangat menentukan dalam uji statistik.

## 1. Skala Nominal

Skala nominal adalah pengukuran yang paling rendah tingkatannya ini terjadi apabila bilangan atau lambing lambing lain digunakan untuk mengklasifikasikan obyek, orang, hewan atau benda lain. Apabila bilangan atau lambing-lambang lain digunakan untuk mengidentifikasi kelompok dimana beberapa obyek dapat dimasukkan kedalamnya maka bilangan atau lambing itu membentuk suatu skala nominal (klasifikasi).

Adalah skala yang hanya mendasarkan pada pengelompokan atau pengkategorian peristiwa atau fakta dan apabila menggunakan notasi angka hal itu sama sekali tidak menunjukkan perbedaan kuantitatif melainkan hanya menunjukkan perbedaan kualitatif.

Suryabrata, S (2003) menyebut bahwa skala nominal adalah skala yang ditetapkan berdasarkan atas proses penggolongan yang bersifat diskrit dan saling pilah (*mutually exclusive*). Banyak variabel dalam penelitian sosial menggunakan skala nominal seperti agama, jenis kelamin, tempat lahir, asal sekolah, dsb. Untuk itu skala nominal mempunyai sifat:

- a) Kategori data bersifat *mutually exclusive* (saling memisah),
- b) Kategori data tidak mempunyai aturan yang logis (bisa sembarang).

Skala nominal merupakan skala yang paling sederhana disusun menurut jenis (katagorinya) atau fungsi bilangan hanya sebagai simbol untuk membedakan sebuah karakteristik lainnya. Skala nominal merupakan skala yang paling lemah/rendah di antara skala pengukuran yang ada. Skala nominal hanya bisa membedakan benda atau peristiwa yang satu dengan yang lainnya berdasarkan nama (predikat). Skala pengukuran nominal digunakan untuk mengklasifikasi objek, individual atau kelompok dalam bentuk kategori. Pemberian angka atau simbol pada skala nominal tidak memiliki maksud kuantitatif hanya menunjukkan ada atau tidaknya atribut atau karakteristik pada objek yang diukur. Misalnya:

- Jenis kelamin diberi kode 1 untuk laki-laki dan kode 2 untuk perempuan. Angka ini hanya berfungsi sebagai label. Kategori tanpa memiliki nilai intrinsik dan tidak memiliki arti apa pun. Kita tidak bisa mengatakan perempuan dua kali dari laki-laki. Kita bisa saja mengkode laki-laki menjadi 2 dan perempuan dengan kode 1, atau bilangan apapun asal kodenya berbeda antara laki-laki dan perempuan.
- Agama, kita bisa mengkode 1 = Islam, 2 = Kristen, 3 = Hindu, 4 = Budha, dst. Kita bisa menukar angka-angka tersebut, selama suatu karakteristik memiliki angka yang berbeda dengan karakteristik lainnya. Karena tidak memiliki nilai intrinsik, maka angka-angka (kode-kode) yang kita berikan tersebut tidak memiliki sifat sebagaimana bilangan pada umumnya.
- menggolongkan ternak dalam himpunan ternak besar, ternak kecil, ternak unggas dan aneka ternak. Dalam hal ini skala untuk mengukur variable ternak terdiri dari empat titik. Titik skala dinamakan kelas atau kategori. Demikian pula pengelompokan suatu kejadian menjadi dua kelompok yang dikenal dengan skala nominal dikotonik dan biasanya diberi lambing himpunan.
- Misalnya kejadian mati dan hidup sembuh dan sakit, tidak berhasil, berhasil, tidak ditemukan dan ditemukan.

Oleh karenanya, pada variabel dengan skala nominal tidak dapat diterapkan operasi matematika standar (aritmatik) seperti pengurangan, penjumlahan, perkalian, dan lainnya. Peralatan statistik yang sesuai dengan skala nominal adalah proposisi seperti modus, distribusi frekuensi, Chi Square dan beberapa peralatan statistik non-parametrik lainnya. Ciri-ciri Skala NOMINAL:

- a) Hasil penghitungan tidak dijumpai bilangan pecahan,
- b) Angka yang tertera hanya label saja,
- c) Tidak mempunyai urutan (ranking),
- d) Tidak mempunyai ukuran baru,
- e) Tidak mempunyai nol mutlak,
- f) Tes statistik yang digunakan adalah statistik non parametrik.

**Contoh Skala nominal sebenarnya :**

- Jenis kulit : Hitam Kuning Putih
- Suku Daerah : Jawa Madura Bugis

- Agama yang dianut : Islam Kristen Hindu
- Partai pemenang pemilu : Golkar Demokrat PKB
- Jenis kelamin : Laki Perempuan
- Jenis Pekerjaan : PNS Swasta Tani dll
- Status Perkawinan : Kawin Tidak Kawin

### **Contoh Skala nominal yang Tidak Sebenarnya**

- Kelulusan : Lulus Tidak Lulus
- Ijazah yang dipunyai : SD SMP SMA S1 S2 S3
- Tahun Produksi Kendaraan : 2004 2005 2006 2007
- Aktivitas : Bekerja Menganggur

## **2. Skala Ordinal**

Skala ordinal terjadi bila obyek yang ada dalam satu katagori suatu skala tidak hanya berbeda dengan obyek-obyek itu, tetapi juga mempunyai hubungan satu dengan yang lain. Hubungan yang ada biasa kita jumpai diantara kelas-kelas adalah : lebih tinggi, lebih disenangi, lebih sering, lebih sulit, lebih dewasa dan sebagainya.

Pengukuran di mana skala yang dipergunakan disusun berdasarkan atas jenjang dalam atribut tertentu sehingga penyusunannya disusun secara terurut dari yang rendah sampai yang tinggi menurut suatu ciri tertentu, namun antara urutan (ranking) yang satu dengan yang lainnya tidak mempunyai jarak yang sama.

Skala ordinal banyak dipergunakan dalam penelitian sosial dan pendidikan terutama berkaitan dengan pengukuran kepentingan, persepsi, motivasi serta sikap, apabila mengukur sikap responden terhadap suatu kebijakan pendidikan, responden dapat diurutkan dari mulai Sangat Setuju (1), Setuju (2), Tidak Beraardapat (3), Kurang Setuju (4), dan Tidak Setuju (5), maka angka-angka tersebut hanya sekedar menunjukkan urutan responden, bukan nilai untuk variabel tersebut. Adapun cirri dari skala ordinal adalah :

- a) Kategori data bersifat saling memisah,
- b) Kategori data mempunyai aturan yang logis,
- c) Kategori data ditentukan skalanya berdasarkan jumlah karakteristik khusus yang dimilikinya.

Dapat juga dikatakan bahwa skala ordinal merupakan skala yang didasarkan pada ranking diurutkan dari jenjang yang lebih tinggi sampai jenjang yang lebih rendah atau sebaliknya. Skala ordinal ini lebih tinggi daripada skala nominal, dan sering juga disebut dengan skala peringkat. Hal ini karena dalam skala ordinal, lambing-lambang bilangan hasil pengukuran selain menunjukkan perbedaan juga menunjukkan urutan atau tingkatan objek yang diukur menurut karakteristik tertentu.

Pengukuran yang dilakukan dalam skala ordinal adalah obyek dibedakan menurut persamaanya dan menurut urutannya. Jadi dapat dibuat urutan atau ranking yang lengkap dan teratur diantar kelas-kelas. Sebagai contoh kejadian suatu penyakit pada ternak babi dibagi menjadi sering sekali, sering, kadang-kadang, dan tidak pernah. Kejadian ini bisa dikuantitatifkan dengan memberikan lambing himpunan angka-angka **{4,3,2,1}**.

Misalnya tingkat kepuasan seseorang terhadap produk. Bisa kita beri angka dengan 5 = sangat puas, 4 = puas, 3 = kurang puas, 2 = tidak puas, dan 1 = sangat tidak puas. Atau misalnya dalam suatu lomba, pemenangnya diberi peringkat 1, 2, 3, dst. Dalam skala ordinal, tidak seperti skala nominal, ketika kita ingin mengganti angka-angkanya, harus dilakukan secara berurut dari besar ke kecil atau dari kecil ke besar. Jadi, tidak boleh kita buat 1 = sangat puas, 2 = tidak puas, 3 = puas, dst. Yang boleh adalah 1 = sangat puas, 2 = puas, 3 = kurang puas, dst.

Selain itu, yang perlu diperhatikan dari karakteristik skala ordinal adalah meskipun nilainya sudah memiliki batas yang jelas tetapi belum memiliki jarak (selisih). Kita tidak tahu berapa jarak kepuasan dari tidak puas ke kurang puas. Dengan kata lain juga, walaupun sangat puas kita beri angka 5 dan sangat tidak puas kita beri angka 1, kita tidak bisa mengatakan bahwa kepuasan yang sangat puas lima kali lebih tinggi dibandingkan yang sangat tidak puas. Sebagaimana halnya pada pada skala nominal, pada skala ordinal kita juga tidak dapat menerapkan operasi matematika standar (aritmatik) seperti pengurangan, penjumlahan, perkalian, dan lainnya. Peralatan statistik yang sesuai dengan skala ordinal juga adalah peralatan statistik yang berbasiskan (berdasarkan) jumlah dan proposisi seperti modus, distribusi frekuensi, Chi Square dan beberapa peralatan statistik non-parametrik lainnya.

### 3. Skala Interval

Pengukuran dalam skala interval lebih kuat dari skala ordinal, sebab pengukuran dicapai selain dengan persamaan dan urutannya juga mengetahui jarak (interval) antar adua kelas.

Skala pengukuran di mana jarak satu tingkat dengan tingkat lainnya sama, oleh karena itu skala interval dapat juga disebut skala unit yang sama (*equal unit scale*). Skala interval merupakan skala yang dihasilkan dari proses pengukuran, di mana dalam pengukuran tersebut diasumsikan terdapat satuan (unit) pengukuran yang sama. Contoh yang sangat dikenal adalah temperatur. Adapun ciri-ciri skala interval adalah :

- a) Kategori data bersifat saling memisah,
- b) Kategori data bersifat logis,
- c) Kategori data ditentukan skalanya berdasarkan jumlah karakteristik khusus yang dimilikinya,
- d) Perbedaan karakteristik yang sama tergambar dalam perbedaan yang sama dalam jumlah yang dikenakan pada kategori,
- e) Angka nol hanya menggambarkan suatu titik dalam skala (tidak punya nilai nol absolut).

Dengan demikian skala interval merupakan skala yang menunjukkan jarak antara satu data dengan data yang lain dan mempunyai bobot yang sama. Analisis statistik yang digunakan ialah mempunyai karakteristik uji statistik parametik. Skala interval mempunyai karakteristik seperti yang dimiliki oleh skala nominal dan ordinal dengan ditambah karakteristik lain, yaitu berupa adanya interval yang tetap. Dengan demikian, skala interval sudah memiliki nilai intrinsik, sudah memiliki jarak, tetapi jarak tersebut belum merupakan kelipatan. Pengertian “jarak belum merupakan kelipatan” ini kadang-kadang diartikan bahwa skala interval tidak memiliki nilai nol mutlak.

Missal pada pengukuran suhu. Kalau ada tiga daerah dengan suhu daerah A = 10°C, daerah B = 15°C dan daerah C = 20°C. Kita bisa mengatakan bahwa selisih suhu daerah B 5°C lebih panas dibandingkan daerah A, dan selisih suhu daerah C dengan daerah B adalah 5°C (ini menunjukkan pengukuran interval sudah memiliki jarak tetap). Tetapi, kita bisa mengatakan bahwa suhu daerah C dua kali lebih panas dibandingkan daerah A (artinya tidak bisa jadi

kelipatan). **Kenapa?** karena dalam derajat Celcius tidak memiliki NOL ABSOLUT. (Titik nolnya pada 0C Bukan berarti Tidak ada Suhu sama sekali). Contoh lainnya, misalnya dua orang murid, si A mendapat 70 sedangkan si B mendapat nilai 35. Kita tidak bisa mengatakan si A dua kali lebih pintar dibandingkan si B.

#### 4. Skala Ratio

Skala rasio disamping mempunyai sifat seperti skala interval masih juga mempunyai sifat lain yaitu titik nolnya tertentu. Dalam skala rasional perbandingan dua titik skala sembarang adalah independent dengan unit pengukuran.

Merupakan skala pengukuran yang mempunyai nilai NOL MUTLAK dan mempunyai jarak yang sama. Skala interval yang benar-benar memiliki nilai nol mutlak disebut skala rasio, dengan demikian skala rasio menunjukkan jenis pengukuran yang sangat jelas dan akurat (*precise*). Jika kita memiliki skala rasio, kita dapat menyatakan tidak hanya jarak yang sama antara satu nilai dengan nilai lainnya dalam skala, tapi juga tentang jumlah proposional karakteristik yang dimiliki dua objek atau lebih, dan contoh untuk skala ini adalah uang. Adapun ciri-ciri dari skala rasio adalah :

- a) Kategori data bersifat saling memisah,
- b) Kategori data mempunyai aturan yang logis,
- c) Kategori data ditentukan skalanya berdasarkan jumlah karakteristik khusus yang dimilikinya,
- d) Perbedaan karakteristik yang sama tergambar dalam perbedaan yang sama dalam jumlah yang dikenakan pada kategori,
- e) Angka nol menggambarkan suatu titik dalam skala yang menunjukkan ketiadaan karakteristik (punya nilai nol absolut).

Tes yang digunakan adalah tes statistik parametik. Skala rasio adalah skala data dengan kualitas paling tinggi. Pada skala rasio, terdapat semua karakteristik skala nominal, ordinal, dan skala interval ditambah dengan sifat adanya nilai nol yang bersifat adanya nilai nol bersifat mutlak. Nilai nol mutlak ini artinya adalah nilai dasar yang tidak bisa diubah meskipun menggunakan skala yang lain. Oleh karenanya, pada skala ratio, pengukuran sudah mempunyai nilai



perbandingan/rasio. Pengukuran-pengukuran dalam skala rasio yang sering digunakan adalah pengukuran tinggi dan berat.

- Misalnya berat benda A adalah 30 kg, sedangkan benda B adalah 60 kg, maka dapat dikatakan bahwa benda B lebih berat dua kali dibandingkan benda A.
- Umur manusia,
- Ukuran timbangan,
- Berat badan,
- Tinggi pohon,
- Tinggi badan manusia,
- Jarak,
- Panjang barang,
- Nilai ujian.

Dari uraian di atas jelas bahwa Skala *Ratio*, *Interval*, *Ordinal* dan *Nominal* berturut – turut memiliki nilai kuantitatif dari yang Paling Rinci ke yang Kurang Rinci. Skala Ratio mempunyai sifat – sifat yang dimiliki Skala Interval, Ordinal dan Nominal. Skala Interval memiliki ciri – ciri yang dimiliki Skala Ordinal dan Nominal, sedangkan Skala Ordinal memiliki sifat yang dimiliki Skala Nominal. Adanya perbedaan tingkat pengukuran memungkinkan terjadinya Transformasi Skala Ratio dan Interval menjadi Ordinal atau Nominal. Transformasi ini dikenal sebagai *Data Reduction* atau *Data Collapsing*. Hal ini dimaksudkan agar dapat menerapkan metode statistik tertentu, terutama yang menghendaki skala data dalam bentuk Ordinal atau Nominal.

Sebaliknya, Skala Ordinal dan Nominal **TIDAK DAPAT** diubah menjadi Interval atau Ratio. Skala Nominal yang diberi label 0, 1 atau 2 dikenal sebagai *Dummy Variable* (Variabel Rekayasa). Misalnya : Pemberian label 1 untuk laki – laki dan 2 untuk perempuan tidak mempunyai arti kuantitatif (tidak mempunyai nilai / hanya kode). Dengan demikian, perempuan tidak dapat dikatakan 1 lebih banyak dari laki – laki. Pemberian label tersebut dimaksudkan untuk mengubah kategori huruf (Alfabet) menjadi kategori Angka (Numerik), sehingga memudahkan analisis data.

## F. PENGOLAHAN DATA

Data yang telah dikumpulkan (data mentah) yang kemudian akan diolah, penguolah dimaksud ialah sebagai suatu proses untuk memperoleh data ringkasan dari data mentah tersebut dengan menggunakan rumus tertentu.

## G. PENYAJIAN DATA

Agar data yang diolah dapat dibaca serta dimengerti oleh orang lain, maka perlu disajikan kedalam bentuk-bentuk tertentu. Penyajian data memiliki fungsi antara lain :

- Menunjukkan perkembangan suatu keadaan.
- Mengadakan perbandingan pada suatu waktu. Penyajian data dapat dilakukan melalui tabel ataupun Grafik.

### 1. Tabel Data.

Tabel Data ialah penyajian data dalam bentuk kumpulan angka yang disusun menurut kategori-kategori tertentu, dalam suatu daftar. Berdasarkan pengaturan data, tabel dapat dibedakan atas beberapa jenis yaitu:

#### a) Tabel Frekuensi.

Tabel Frekuensi ialah tabel yang menunjukkan atau memuat banyaknya kejadian ataupun Frekuensi dari suatu kejadian.

#### b) Tabel Klasifikasi.

Tabel Klasifikasi ialah tabel yang menunjukkan atau memuat pengelompokan data.

#### c) Tabel Kontigen.

Tabel kontigen ialah tabel yang menunjukkan atau memuat data dengan sesuai serta rinciannya.

#### d) Tabel Kolerasi.

Tabel Kolerasi ialah tabel yang menunjukkan atau memuat adanya korelasi (hubungan) antara data yang disajikan.

### 2. Grafik Data.

Grafik data atau disebut juga dengan diagram, ialah penyajian data dalam bentuk gambar-gambar. Grafik biasanya berasal dari tabel. Grafik data

sebenarnya ialah penyajian data secara visual dari data bersangkutan. Grafik data dibedakan atas beberapa jenis yaitu :

**a) Piktogram**

Grafik data yang menggunakan gambar ataupun lambang dari data itu sendiri dengan skala tertentu.

**b) Grafik Batang atau Balok.**

Grafik data berbentuk persegi panjang yang lebarnya sama dan dilengkapi dengan skala atau ukuran dengan data yang bersangkutan.

**c) Grafik Garis.**

Grafik data berupa garis, yang diperoleh dari beberapa ruas garis yang menghubungkan titik-titik pada bidang bilangan(sistem Salib Sumbu).

**d) Grafik Lingkaran.**

Data yang berupa lingkaran yang telah dibagi menjadi juring-juring sesuai dengan data tersebut.

**e) Kartogram.**

Peta statistik ialah grafik data yang berupa peta yang menunjukkan kepadatan penduduk, curah hujan , hasil pertanian dan lain-lain.

# BAB 3 DISTRIBUSI FREKUENSI

---

## A. PENGERTIAN DISTRIBUSI FREKUENSI

Daftar nilai data (bisa nilai individual atau nilai data yang sudah dikelompokkan) yang disertai dengan nilai frekuensi yang sesuai. Pengelompokan data ini dimaksudkan agar ciri-ciri penting data tersebut dapat segera terlihat. Dan daftar frekuensi ini akan memberikan gambaran yang khas tentang bagaimana keragaman data. Karena sifat keragaman data sangat penting untuk diketahui, karena dalam pengujian statistik selanjutnya kita harus selalu memperhatikan sifat dari keragaman data nya. Tanpa memperhatikan sifat keragaman data nya, penarikan sebuah kesimpulan pada umumnya tidaklah sah.

Salah satu cara untuk mengatur, menyusun, atau meringkas data ialah dengan cara membuat distribusi frekuensi. Kata distribusi berasal dari kata *distribution* (bahasa Inggris), yang berarti penyaluran, pembagian, atau pancaran. Jadi, secara mendasar, distribusi frekuensi dapat diartikan sebagai penyaluran frekuensi, pembagian frekuensi, atau pancaran frekuensi. Sedangkan, frekuensi sendiri juga berasal dari bahasa Inggris, *frequency*, yang berarti kekerapan, keseringan, atau jarang-kerapnya. Dalam statistika, frekuensi berarti seberapa kali suatu variabel yang dilambangkan dengan angka (bilangan) berulang kali dalam deretan data angka tersebut.

Dengan demikian, distribusi frekuensi merupakan suatu keadaan yang menggambarkan bagaimana frekuensi dari gejala atau variabel yang dilambangkan dengan angka itu telah tersalur, terbagi, tersebar, dan terpancar. Penggambaran angka (bilangan) atau penyajian data angka tersebut dapat disajikan dalam bentuk tabel atau grafik/gambar, yang kemudian dikenal dengan istilah tabel distribusi frekuensi dan grafik distribusi frekuensi.

Data yang telah diperoleh dari suatu penelitian yang masih berupa data acak yang dapat dibuat menjadi data yang berkelompok, yaitu data yang telah disusun ke dalam kelas-kelas tertentu. Daftar yang memuat data berkelompok disebut distribusi frekuensi atau tabel frekuensi. Distribusi frekuensi adalah susunan data menurut kelas interval tertentu atau menurut kategori tertentu dalam sebuah daftar (Hasan, 2001).

## B. BAGIAN-BAGIAN DISTRIBUSI FREKUENSI

Bagian-bagian dari distribusi frekuensi antara lain:

1. Kelas (class) merupakan kelompok nilai data atau variabel dari sebuah data acak
2. Batas Kelas (class limit) merupakan nilai yang menjadi batas antar kelas. Batas kelas adalah batas semu dari masing-masing kelas, karena di antara kelas yang satu dengan kelas yang lain masih ada lubang tempat angka tertentu. Terdapat dua batas kelas untuk data yang telah diurutkan, yakni batas kelas bawah (lower class limit) dan batas kelas atas (upper class limit)
3. Tepi kelas atau batas nyata kelas yang tidak mempunyai lubang untuk angka tertentu antara kelas yang satu dengan yang lainnya. Terdapat dua kelas yakni tepi kelas atas dan tepi kelas bawah.
4. Titik tengah kelas atau tanda kelas merupakan angka atau nilai data yang tepat berada di tengah kelas. Titik tengah kelas merupakan nilai yang menjawi wakil kelasnya dalam data. Untuk menetapkan titik tengah kelas bisa dicari dengan rumus:  
Titik Tengah Kelas =  $\frac{1}{2}$  (batas atas kelas + batas bawah kelas)
5. Interval kelas merupakan selang yang menjadi pemisah kelas yang satu dengan kelas yang lain.
6. Panjang interval kelas atau luas kelas merupakan jarak antara tepi atas kelas dengan tepi bawah kelas
7. Frekuensi kelas merupakan banyaknya data yang termasuk ke dalam kelas tertentu dari data acak.

## C. PENYUSUNAN DISTRIBUSI FREKUENSI

Untuk melakukan penyusunan distribusi frekuensi atau tahapannya adalah sebagai berikut:

1. Urutkan data-data lebih dulu sesuai dengan besar nilai pada data.
2. Menentukan jangkauan (range) dari data. Rumus jangkauan yakni:  
Jangkauan = data terbesar-data terkecil
3. Menentukan banyaknya kelas (k). Banyaknya kelas bisa ditentukan dengan rumus sturgess:

$$K = 1 + 3.3 \log$$

**N = Keterangan**

k = banyaknya kelas

n = banyaknya data

4. Menentukan panjang interval kelas. Untuk menentukan panjang interval bisa memakai rumus:  
Panjang interval kelas (i) = Jumlah kelas (k) / Jangkauan (R)
5. Menentukan batas bawah kelas pertama. Tepi bawah kelas pertama seringkali dipilih dari data terkecil atau data yang berasal dari pelebaran jangkauan (data yang lebih kecil dari data data terkecil) dan selisihnya harus dikurangi dari panjang interval kelasnya.
6. Menuliskan frekuensi kelas didalam kolom turus atau tally (sistem turus) sesuai banyaknya data

#### D. PENYAJIAN GRAFIK DISTRIBUSI FREKUENSI

Tabel distribusi frekuensi merupakan alat penyajian data statistika yang berbentuk kolom dan lajur (baris), yang di dalamnya termuat angka-angka yang dapat melukiskan atau menggambarkan pancaran atau pembagian frekuensi dari variabel yang sedang menjadi objek penelitian. Sebelum menggunakan tabel, ada tiga tahapan yang perlu dilakukan pada data, yaitu penyusunan, penyederhanaan, dan pengelompokan.

- Contoh:

Diketahui data berikut: 60, 50, 75, 60, 80, 40, 60, 70, 100, 75.

✓ Tahap penyusunan: 40, 50, 60, 60, 60, 70, 75, 75, 80, 100.

✓ Tahap penyederhanaan.

Nilai ( X )	Banyaknya Orang (Frekuensi)
100	1
80	1
75	2
70	1
60	3
50	1
40	1
Jumlah (N)	10

- ✓ Tahap pengelompokan:

Jika data yang diperoleh banyak/besar (data kelompok), jumlah lebih besar dari 30 ( $N > 30$ ), sebaiknya data itu disusun dalam bentuk tabel distribusi frekuensi data kelompok.

## 1. Tabel Distribusi Frekuensi Data Tunggal

Tabel distribusi frekuensi data tunggal adalah salah satu jenis tabel statistik yang di dalamnya disajikan frekuensi dari data angka. Angka yang ada tersebut tidak dikelompok-kelompokkan.

- Contoh: Berikut distribusi frekuensi nilai mata kuliah statistik pendidikan semester IV dari 40 orang mahasiswa.

Nilai ( X )	Frekuensi ( f )
4,0	6
3,5	9
3,0	19
2,5	6
Jumlah ( N )	40

## 2. Tabel Distribusi Frekuensi Data Kelompok

Tabel distribusi frekuensi data kelompok adalah salah satu jenis tabel statistik yang di dalamnya disajikan pencaran atau distribusi frekuensi dari data angka. Angka-angka tersebut dikelompokkan (dalam setiap unit terdapat sekelompok angka). Adapun langkah-langkah membuat tabel distribusi frekuensi data kelompok adalah sebagai berikut:

- ✓ Mencari range (R) = penyebaran/jangkauan, dengan rumus:

$$R = X_t - X_r + 1$$

$$R = H - L + 1$$

Keterangan:

R = range

$X_t$  (H) = nilai tertinggi

$X_r$  (L) = nilai terendah

- \_\_\_\_\_
- ✓ Menentukan kelas/kelompok. Ada dua cara untuk menentukan kelas ini, yaitu dengan rumus dan dengan serampangan.

a) Dengan rumus tetap yang diusulkan Sturges.

$$K = 1 + 3,3 \log N$$

Dengan serampangan/sembarangan, yaitu dengan menambah satu kelas lagi, jika masih terdapat salah satu nilai (skor) yang belum masuk dalam distribusi. Bisa juga dengan membulatkan ke atas setiap hasil dari penghitungan penentuan kelompok/kelas.

- \_\_\_\_\_
- ✓ Menghitung lebar kelas (interval kelas) yang disimbolkan dengan  $i$  dan dirumuskan sebagai berikut:

$$i = R/K$$

- ✓ Menentukan batas kelas (batas bawah nyata dan batas atas nyata) dengan rumus berikut:

$$\text{batas bawah} = \text{ujung bawah} - 0,5$$

$$\text{batas atas} = \text{ujung atas} + 0,5$$

- \_\_\_\_\_
- ✓ Menentukan titik tengah kelas ( $X_1$ ) dengan tiga macam cara sebagai berikut:
  - a) Titik tengah kelas =  $0,5 \times (\text{batas bawah} + \text{batas atas})$
  - b) Titik tengah kelas =  $0,5 \times (\text{ujung bawah relatif} + \text{ujung atas relatif})$
  - c) Titik tengah kelas =  $(\text{ujung bawah relatif} + \text{ujung ATAS RELATIF})/2$

- \_\_\_\_\_
- ✓ Membuat tabel distribusi frekuensi berdasarkan hasil  $R$ ,  $K$ , dan  $i$ .  
Cara pengisian tabel tersebut:

a) Pada kolom interval kelas, mulailah dari angka yang paling kecil dari data mentah, kemudian diurut sampai empat belas angka sesuai hasil interval, begitu seterusnya ke atas sampai kelas tertentu sesuai hasil  $K$ .

b) Data mentah dari masing-masing angka ditabulasikan untuk menentukan frekuensi.

c) Dilakukan perhitungan  $f_k$  (frekuensi komulatif). Nilai  $f_k$  ini merupakan akumulasi penjumlahan frekuensi dari bawah ke atas atau sebaliknya, akumulasi frekuensi dari atas ke bawah. Hal ini dilakukan



untuk memastikan bahwa jumlah frekuensi sesuai dengan jumlah N, sekaligus untuk kepentingan membuat grafik ogive.

d) Menentukan titik tengah ( $X_1$ ) dari masing-masing data.

Contoh:

Terdapat data hasil ulangan bidang studi fiqh sebagai berikut:

65 54 44 32 28 82 29 15

78 64 43 17 41 76 35 50

20 96 21 43 46 29 28 52 N = 40

52 27 36 46 53 37 16 55

35 43 33 37 28 68 36 55

▪  
✓ Mencari range. Dari data tersebut, range-nya (R) adalah sebagai berikut:

$$\begin{aligned} R &= 96 - 15 + 1 \\ &= 81 + 1 \\ &= 82 \end{aligned}$$

▪  
✓ Menentukan kelas/kelompok dengan rumus Sturgess. Dari data tersebut, nilai K dapat dicari sebagai berikut:

$$\begin{aligned} K &= 1 + 3,3 \cdot 1,6 \\ &= 1 + 5,28 \\ &= 6,28 \\ &= 6 \end{aligned}$$

▪  
✓ Menghitung lebar kelas (interval kelas). Dari data tersebut, i dapat dicari sebagai berikut:

$$i = R/K = 82/6 = 13,7 = 14$$

▪  
✓ Menentukan batas kelas. Dari data tersebut diperoleh:

$$\begin{aligned} \text{Ujung bawah} &= 15 \\ \text{Ujung atas} &= 28 \\ \text{Batas bawah} &= 15 - 0,5 = 14,5 \\ \text{Batas atas} &= 28 + 0,5 = 28,5 \end{aligned}$$

✓ Menentukan titik tengah kelas ( $X_1$ ). Dari data tersebut, diperoleh:

$$(X) = 0,5 (\text{batas bawah} + \text{batas atas})$$

$$= 0,5 (14,5 + 28,5)$$

$$= 0,5 \cdot 43$$

$$= 21,5$$

■

✓ Membuat tabel distribusi frekuensi berdasarkan hasil R, K, dan i. Dari data tersebut, tabel distribusi frekuensi yang terbentuk sebagai berikut:

Interval Kelas / Kelompok	Tallys/ Tabulasi	F	fk	$X^1$
85 – 98	I	1	40 = N	91,5
71 – 84	II	3	39	77,5
57 – 70	III	3	36	63,5
43 – 56	III III III	13	33	49,5
29 – 42	III III I	11	20	35,5
15 – 28	III III	9	9	21,5
		N = 40		

**Perhatikan contoh** data pada Tabel berikut ini, daftar nilai ujian matakuliah statistik dari 80 Mahasiswa. Daftar Nilai Ujian Matakuliah Statistik

79	49	48	74	81	98	87	80
80	84	90	70	91	93	82	78
70	71	92	38	56	81	74	73
68	72	85	51	65	93	83	86
90	35	83	73	74	43	86	88
92	93	76	71	90	72	67	75
80	91	61	72	97	91	88	81
70	74	99	95	80	59	71	77
63	60	83	82	60	67	89	63
76	63	88	70	66	88	79	75

Sangatlah sulit untuk menarik sebuah kesimpulan dari daftar data tersebut. Secara sepiantas, kita belum bisa menentukan beberapa nilai ujian terkecil atau terbesar. Demikian pula, dengan kita belum bisa mengetahui dengan tepat, berapa nilai ujian yang paling banyak atau berapa banyak mahasiswa nya yang mendapatkan nilai tertentu. Dengan demikian, kita harus mengolah data tersebut terlebih dahulu agar dapat memberikan gambaran atau keterangan yang lebih baik.

Bandingkan dengan tabel yang sudah disusun dalam bentuk daftar frekuensi berikut ini. Tabel 2a ialah daftar frekuensi dari data tunggal dan Tabel 2b ialah daftar frekuensi yang disusun dari data yang sudah di kelompokkan pada kelas yang sesuai dengan selangnya. Dan kita bisa memperoleh beberapa informasi atau karakteristik dari data nilai ujian mahasiswa nya.

**Tabel 2a :**

No	Nilai Ujian	Frekuensi
	$x_i$	$f_i$
1	35	1
2	36	0
3	37	0
4	38	1
:	:	:
16	70	4
17	71	3
:	:	1
42	98	1
43	99	1
	Total	80

Pada Tabel 2a, kita bisa mengetahui bahwa ada 80 mahasiswa yang mengikuti ujian, nilai ujian terkecil ialah 35 dan tertinggi ialah 99. Nilai 70 yakni nilai yang paling banyak diperoleh oleh mahasiswa, yaitu ada 4 orang, atau kita juga bisa mengatakan ada 4 orang mahasiswa yang memperoleh nilai 70. Tidak ada satu pun mahasiswa yang mendapatkan nilai 36, atau hanya 1 orang mahasiswa yang mendapatkan nilai 35.

**Tabel 2b :**

Kelas ke-	Nilai Ujian	Frekuensi $f_i$
1	31 – 40	2
2	41 – 50	3
3	51 – 60	5
4	61 – 70	13
5	71 – 80	24
6	81 – 90	21
7	91 – 100	12
	Jumlah	80

Tabel 2b ialah daftar frekuensi dari data yang sudah dikelompokkan. Daftar ini yakni daftar frekuensi yang sering digunakan. Kita sering kali mengelompokkan data contoh ke dalam selang-selang tertentu agar memperoleh gambaran yang lebih baik mengenai karakteristik dari suatu data.

Dari daftar tersebut, kita bisa mengetahui bahwa mahasiswa yang mengikuti ujian ada 80 orang, selang kelas nilai yang paling banyak diperoleh oleh mahasiswa ialah sekitar 71 sampai 80 orang, yaitu ada 24 orang, dll. Tapi perlu diingat bahwa dengan cara ini kita bisa kehilangan identitas dari data aslinya. Sebagai contoh, kita bisa mengetahui bahwa ada 2 orang mahasiswa yang mendapatkan nilai antara 31 sampai 40.

Meskipun demikian, kita tidak akan tahu dengan persis, berapa nilai sebenarnya dari 2 orang mahasiswa itu, apakah 31 apakah 32 atau 36. Ada beberapa istilah yang harus dipahami dahulu dalam menyusun daftar frekuensi.

**Tabel 3 :**

Kelas ke-	Selang Nilai Ujian	Batas Kelas	Nilai Kelas ( $x_i$ )	Frekuensi ( $f_i$ )
1	31 – 40	30.5 – 40.5	35.5	2
2	41 – 50	40.5 – 50.5	45.5	3
3	51 – 60	50.5 – 60.5	55.5	5
4	61 – 70	60.5 – 70.5	65.5	13
5	71 – 80	70.5 – 80.5	75.5	24
6	81 – 90	80.5 – 90.5	85.5	21
7	91 – 100	90.5 – 100.5	95.5	12
	Jumlah			80

1. **Range** = Selisih antara nilai tertinggi dan terendah. Dan pada contoh ujian di atas,  $\text{Range} = 99 - 35 = 64$ .
2. **Batas bawah kelas** = Nilai terkecil yang berada pada setiap kelas. (Contohnya pada Tabel 3 di atas, batas bawah kelas nya ialah 31, 41, 51, 61, ..., 91).
3. **Batas atas kelas** = Nilai terbesar yang berada pada setiap kelas. (Contohnya pada Tabel 3 di atas, batas bawah kelas nya ialah 40, 50, 60, ..., 100).
4. **Batas kelas (Class boundary)** = Nilai yang digunakan untuk memisahkan antar kelas, tapi tanpa adanya jarak antar batas atas kelas dengan batas bawah kelas berikutnya. Contohnya pada kelas ke 1, batas kelas terkecil nya ialah 30.5 dan yang terbesar 40.5. Pada kelas ke 2, batas kelas nya ialah 40.5 dan

50.5. Nilai pada batas atas kelas yang ke 1 (40.5) sama dengan. Dan merupakan nilai batas bawah bagi kelas yang ke 2 (40.5). Batas kelas selalu dinyatakan dengan jumlah digit 1 desimal lebih banyak daripada data pengamatan asalnya. Ini dilakukan untuk menjamin tidak ada nilai pengamatan yang jatuh tepat pada batas kelasnya, sehingga menghindarkan keraguan saat data tersebut harus ditempatkan. Contohnya bila batas kelas nya di buat seperti ini:

- Kelas ke 1 = 30 – 40.
- Kelas ke 2 = 40 – 50.

5. **Panjang/lebar kelas (selang kelas)** = Selisih antar 2 nilai batas bawah kelas yang berurutan atau selisih antara 2 nilai batas atas kelas yang berurutan atau selisih antara nilai terbesar dan terkecil batas kelas bagi kelas yang bersangkutan. Biasanya lebar kelas nya memiliki lebar yang sama. Contohnya :

- lebar kelas =  $41 - 31 = 10$  (selisih antar 2 batas bawah kelas yang berurutan).
- lebar kelas =  $50 - 40 = 10$  (selisih antar 2 batas atas kelas yang berurutan).
- lebar kelas nya =  $40.5 - 30.5 = 10$ . (selisih antar nilai terbesar dengan nilai terkecil batas kelas pada kelas ke 1).

6. **Nilai tengah kelas** = Nilai kelas ialah nilai tengah dari kelas yang bersangkutan yang diperoleh dengan formula berikut =  $\frac{1}{2}$  (batas atas kelas nya + batas bawah kelas nya). Nilai ini yang dijadikan perwakilan dari selang kelas tertentu untuk perhitungan analisis statistik yang selanjutnya. Contohnya nilai kelas ke 1 ialah  $\frac{1}{2}(31+40) = 35.5$ .

7. **Banyak kelas** = Pada tabel ada 7 buah kelas.

8. **Frekuensi kelas** = Banyaknya nilai yang muncul pada selang kelas tertentu. Contohnya, pada kelas ke-1, frekuensinya adalah = 2. Nilai frekuensi nya adalah = 2 karena pada selang antara 30.5 – 40.5, hanya ada 2 angka yang muncul, yakni nilai ujian 31 dan 38.

### 3. Teknik Pembuatan Tabel Distribusi Frekuensi (TDF)

Distribusi frekuensi dibuat dengan alasan berikut ini :

- Kumpulan data nya yang besar dapat diringkas.
- Lalu kita dapat memperoleh beberapa gambaran mengenai karakteristik data.
- Dan ini merupakan dasar dalam pembuatan grafik penting (seperti histogram).

Banyak juga software yang bisa digunakan untuk membuat tabel distribusi frekuensi secara otomatis. Meski demikian, di sini tetap akan diuraikan mengenai prosedur dasar dalam membuat tabel distribusi frekuensi.

#### a) Langkah-langkah dalam menyusun tabel distribusi frekuensi :

- Urutkan data dari nilai yang paling kecil.
- Tentukan range nya (Range = nilai maksimum – nilai minimum).
- Tentukan banyak kelas yang diinginkan. Tapi jangan terlalu banyak atau sedikit, berkisar antara 5 dan 20 kelas, tergantung dari banyak datanya.
- Tentukan panjang atau lebar nya dari kelas interval (p).
- Tentukan nilai ujung bawah kelas interval yang pertama.

Pada saat menyusun TDF, pastikan bahwa kelas nya tidak tumpang tindih sehingga setiap nilai nya pengamatan nya harus masuk tepat ke dalam 1 kelas. Pastikan juga tidak akan ada data pengamatan yang tertinggal.

Lalu cobalah untuk menggunakan lebar yang sama untuk semua kelas, meskipun kadang tidak mungkin untuk menghindari interval terbuka, seperti ( $\geq 91$ ). Dan mungkin saja ada kelas tertentu dengan frekuensi 0.

**Contoh :** Kita gunakan prosedur di atas untuk menyusun sebuah tabel distribusi frekuensi nilai ujian mahasiswa.

**Berikut ialah nilai ujian yang sudah diurutkan dari terkecil ke terbesar :**

35 38 43 48 49 51 56 59 60 60  
61 63 63 63 65 66 67 67 68 70  
70 70 70 71 71 71 72 72 72 73  
73 74 74 74 74 75 75 76 76 77  
78 79 79 80 80 80 80 81 81 81  
82 82 83 83 83 84 85 86 86 87  
88 88 88 88 89 90 90 90 91 91  
91 92 92 93 93 93 95 97 98 99

1. Range ialah nilai tertinggi – nilai terendah =  $99 - 35 = 64$ .
2. Banyak nya kelas :

Tentukan banyak kelas yang diinginkan. Apabila kita lihat nilai Range = 64, mungkin banyak kelas nya sekitar 6 atau 7 kelas.

Sebagai latihan, kita gunakan aturan Sturges seperti ini:

- banyak kelas nya =  $1 + 3.3 \times \log(n)$ .
  - =  $1 + 3.3 \times \log(80)$ .
  - =  $7.28 = 7$ .
3. Panjang kelas nya = range : banyak kelas.
    - =  $64/7$
    - =  $9.14 = 10$
  4. Tentukan nilai batas bawah kelas pada kelas yang pertama.

Nilai ujian yang terkecil = 35.

Penentuan nilai batas bawah kelas nya bebas saja, asalkan nilai terkecil masih masuk ke dalam kelas itu.

Misalkan, apabila nilai batas bawah yang kita pilih yakni 26, maka interval kelas pertama nya adalah =  $26 - 35$ , nilai 35 tepat jatuh di batas atas kelas ke 1.

Namun, apabila kita memilih nilai batas bawah kelas 20 atau 25, jelas nilai terkecil nya adalah 35, dan itu tidak akan masuk ke dalam kelas tersebut. Namun, untuk kemudahan dalam penyusunan dan pembacaan TDF nya,

batas bawah 30 atau 31. Ok, saya tertarik dengan angka 31, sehingga batas bawahnya ialah 31. Dari prosedur di atas, kita dapat info sebagai berikut ini :

- Banyak kelas nya = 7.
- Panjang kelas nya = 10.

- Batas bawah kelas nya = 31.

Selanjutnya kita susun TDF nya :

Form TDF :

Kelas ke | Nilai Ujian | Batas Kelas | Turus | Frekuensi |

1 = 31 –  
 2 = 41 –  
 3 = 51 –  
 : = : –  
 6 = 81 –  
 7 = 91 –

Tabel berikut ini merupakan tabel yang sudah dilengkapi :

Kelas ke-	Nilai Ujian	Batas Kelas	Frekuensi (f <sub>i</sub> )
1	31 – 40	30.5 – 40.5	2
2	41 – 50	40.5 – 50.5	3
3	51 – 60	50.5 – 60.5	5
4	61 – 70	60.5 – 70.5	13
5	71 – 80	70.5 – 80.5	24
6	81 – 90	80.5 – 90.5	21
7	91 – 100	90.5 – 100.5	12
	Jumlah		80

Atau dalam bentuk nya yang lebih ringkas adalah :



Kelas ke-	Nilai Ujian	Frekuensi (fi)
1	31 – 40	2
2	41 – 50	3
3	51 – 60	5
4	61 – 70	13
5	71 – 80	24
6	81 – 90	21
7	91 – 100	12
	Jumlah	80

## 5. Distribusi Frekuensi Relatif dan Kumulatif

Variasi penting dari distribusi frekuensi dasar ialah dengan menggunakan nilai frekuensi relatifnya, yang disusun dengan cara membagi frekuensi setiap kelas nya dengan total dari semua frekuensi nya. Sebuah distribusi frekuensi relatif mencakup batas-batas kelas yang sama seperti TDF, namun frekuensi yang digunakan bukan frekuensi aktual melainkan frekuensi relatif. Frekuensi relatif kadang-kadang dinyatakan dengan %.

$$\text{Frekuensi relatif} = \frac{f_i}{\sum f_i} \times 100\% = \frac{f_i}{n} \times 100\%$$

Contoh nya frekuensi relatif kelas ke 1 :

**fi = 2 dan n = 80.**

Frekuensi relatif nya =  $2 : 80 \times 100\% = 2.5\%$ .

Kelas ke-	Nilai Ujian	Frekuensi relatif (%)
1	31 – 40	2.50
2	41 – 50	3.75
3	51 – 60	6.25
4	61 – 70	16.25
5	71 – 80	30.00
6	81 – 90	26.25
7	91 – 100	15.00
	Jumlah	100.00

### a) Distribusi Frekuensi Kumulatif

Variasi lain dari distribusi frekuensi standar ialah frekuensi kumulatif. Frekuensi kumulatif untuk sebuah kelas ialah nilai frekuensi untuk kelas tersebut ditambah dengan jumlah frekuensi semua kelas yang sebelumnya. Perhatikan, bahwa kolom frekuensi nya selain label header nya diganti dengan frekuensi kumulatif kurang dari, batas-batas kelas diganti dengan (kurang dari) ekspresi yang menggambarkan kisaran nilai yang baru.

Nilai Ujian	Frekuensi kumulatif kurang dari
kurang dari 30.5	0
kurang dari 40.5	2
kurang dari 50.5	5
kurang dari 60.5	10
kurang dari 70.5	23
kurang dari 80.5	47
kurang dari 90.5	68
kurang dari 100.5	80

atau kadang disusun juga dalam bentuk seperti ini :

Nilai Ujian	Frekuensi kumulatif kurang dari
kurang dari 41	2
kurang dari 51	5
kurang dari 61	10
kurang dari 71	23
kurang dari 81	47
kurang dari 91	68
kurang dari 101	80

Variasi lain nya ialah frekuensi kumulatif nya lebih dari. Prinsip nya ialah hampir sama dengan prosedur yang di atas.

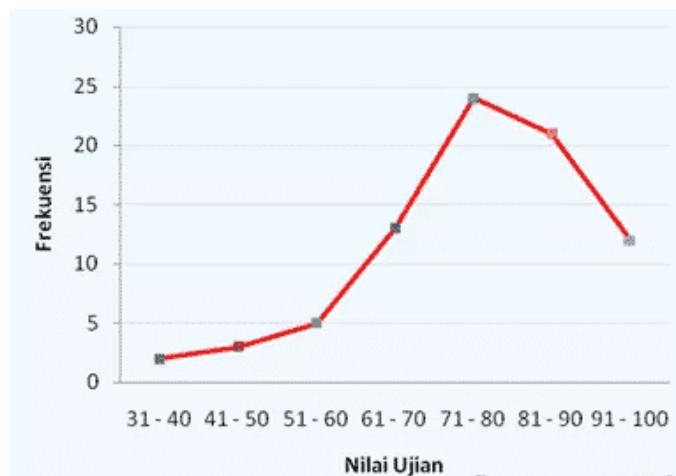
### b) Histogram

Histogram ialah bagian dari grafik batang di mana skala horisontal nya mewakili nilai data kelas dan skala vertikal yang mewakili nilai frekuensinya.

Tinggi batang nya sesuai dengan nilai frekuensinya, dan batang satu dengan batang lainnya saling berdempetan, tidak ada jarak diantara batang nya.

### c) Poligon Frekuensi

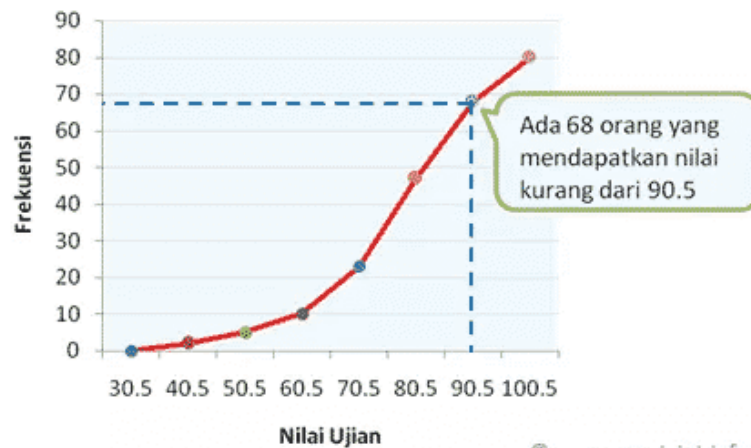
Poligon Frekuensi menggunakan segmen garis yang terhubung ke titik yang terletak di atas nilai-nilai titik tengah kelas. Ketinggian dari titik-titik tersebut sesuai dengan frekuensi kelas, dan segmen garis nya diperluas ke kanan dan kiri sehingga grafik nya dimulai dan berakhir pada sumbu horisontal.



### d) Ogive

Ogive ialah grafik garis yang menggambarkan frekuensi kumulatif, dan ini sama seperti daftar distribusi frekuensi kumulatif. Perhatikan bahwa batas kelas dihubungkan oleh segmen garis yang dimulai dari batas bawah kelas pertama dan berakhir di batas atas dari kelas terakhir.

Ogive juga berguna untuk menentukan jumlah nilai di bawah nilai tertentu. Contoh nya, pada gambar berikut ini menunjukkan bahwa ada 68 mahasiswa mendapatkan nilai yang kurang dari 90.5 :



## E. JENIS-JENIS DISTRIBUSI FREKUENSI

Adapun jenis-jenis dari distribusi frekuensi antara lain:

### 1. Distribusi Frekuensi Menurut Jenisnya

- Distribusi Frekuensi Numerik, adalah suatu deret hitung yang berdiri sendiri.
- Distribusi Frekuensi Kategorial, adalah data yang telah dikelompokkan.

### 2. Distribusi Frekuensi Menurut Nyata Tidaknya Frekuensi

- Distribusi frekuensi absolut, adalah suatu jumlah bilangan yang menjelaskan banyaknya data di suatu kelompok tertentu.
- Distribusi frekuensi relatif, adalah suatu jumlah presentase yang menjelaskan banyaknya data di suatu kelompok tertentu.

### 3. Distribusi Frekuensi Menurut Kesatuannya

- Distribusi frekuensi satuan, adalah frekuensi yang menggambarkan berapa banyak data di kelompok tertentu.
- Distribusi frekuensi kumulatif, adalah distribusi frekuensi yang menggambarkan jumlah frekuensi di sekelompok nilai tertentu mulai dari kelompok sebelumnya sampai dengan kelompok itu.

Menurut Hasan (2001) menyatakan, distribusi frekuensi bisa dibagi menjadi tiga, antara lain:

### 1. Distribusi Frekuensi Biasa

Adalah jenis distribusi frekuensi yang mengandung jumlah frekuensi dari masing-masing kelompok data. Terdapat dua jenis distribusi frekuensi yakni distribusi frekuensi numerik dan distribusi frekuensi peristiwa atau kategori.

## **2. Distribusi Frekuensi Relatif**

Adalah jenis distribusi frekuensi yang mengandung nilai hasil bagi antara frekuensi kelas dan jumlah pengamatan . Distribusi frekuensi relatif menjelaskan proporsi data yang terletak di suatu kelas interval, distribusi frekuensi relatif suatu kelas diperoleh dengan cara membagi frekuensi dengan total data yang ada dari pengamatan atau observasi.

## **3. Distribusi Frekuensi Kumulatif**

Adalah jenis distribusi frekuensi yang mengandung frekuensi kumulatif atau frekuensi yang dijumlahkan. Distribusi frekuensi kumulatif mempunyai kurva yang dinamakan ogif. Terdapat dua jenis distribusi frekuensi yakni distribusi kumulatif kurang dari dan distribusi frekuensi lebih dari.

# BAB 4 DATA EDITOR (SPSS)

---

## A. DUA BAGIAN UTAMA DATA EDITOR SPSS

SPSS adalah sebuah program aplikasi yang memiliki kemampuan analisis statistik cukup tinggi serta sistem manajemen data pada lingkungan grafis dengan menggunakan menu-menu deskriptif dan kotak-kotak dialog yang sederhana sehingga mudah untuk dipahami cara pengoperasiannya. Beberapa aktivitas dapat dilakukan dengan mudah dengan menggunakan pointing dan clicking mouse.

SPSS banyak digunakan dalam berbagai riset pemasaran, pengendalian dan perbaikan mutu (quality improvement), serta riset-riset sains. SPSS pertama kali muncul dengan versi PC (bisa dipakai untuk komputer desktop) dengan nama SPSS/PC+ (versi DOS). Tetapi, dengan mulai populernya system operasi windows. SPSS mulai mengeluarkan versi windows (mulai dari versi 6.0 sampai versi terbaru sekarang).

Pada awalnya SPSS dibuat untuk keperluan pengolahan data statistik untuk ilmu-ilmu social, sehingga kepanjangan SPSS itu sendiri adalah Statistikal Package for the Social Sciens. Sekarang kemampuan SPSS diperluas untuk melayani berbagai jenis pengguna (user), seperti untuk proses produksi di pabrik, riset ilmu sains dan lainnya. Dengan demikian, sekarang kepanjangan dari SPSS Statistical Product and Service Solutions.

SPSS dapat membaca berbagai jenis data atau memasukkan data secara langsung ke dalam SPSS Data Editor. Bagaimanapun struktur dari file data mentahnya, maka data dalam Data Editor SPSS harus dibentuk dalam bentuk baris (cases) dan kolom (variables). Case berisi informasi untuk satu unit analisis, sedangkan variable adalah informasi yang dikumpulkan dari masing-masing kasus.

Hasil-hasil analisis muncul dalam SPSS Output Navigator. Kebanyakan prosedur Base System menghasilkan pivot tables, dimana kita bisa memperbaiki tampilan dari keluaran yang diberikan oleh SPSS. Untuk memperbaiki output, maka kita dapat memperbaiki output sesuai dengan kebutuhan.

SPSS adalah sebuah program analisa statistik yang digunakan dalam berbagai bidang, dari riset pasar hingga badan pemerintahan. SPSS menyediakan banyak fungsi untuk mengolah data, tetapi Anda membutuhkan data sebelum dapat menggunakan fungsi-fungsi yang disediakan. Ada beberapa cara untuk memasukkan data ke dalam

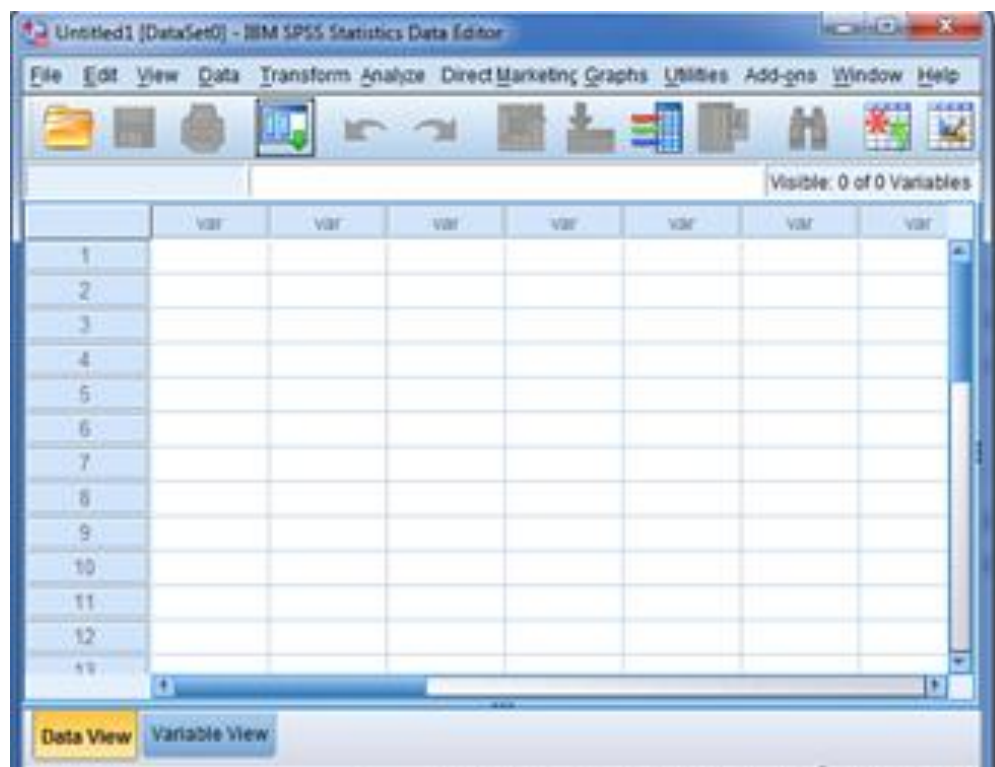
SPSS, dimulai dari memasukkannya secara manual hingga memasukkan data dari berkas lain. SPSS memiliki dua bagian utama, yaitu jendela data editor dan jendela output. Kedua bagian utam ini memiliki fungsi dan peran masing-masing.

## 1. Jendela Data Editor

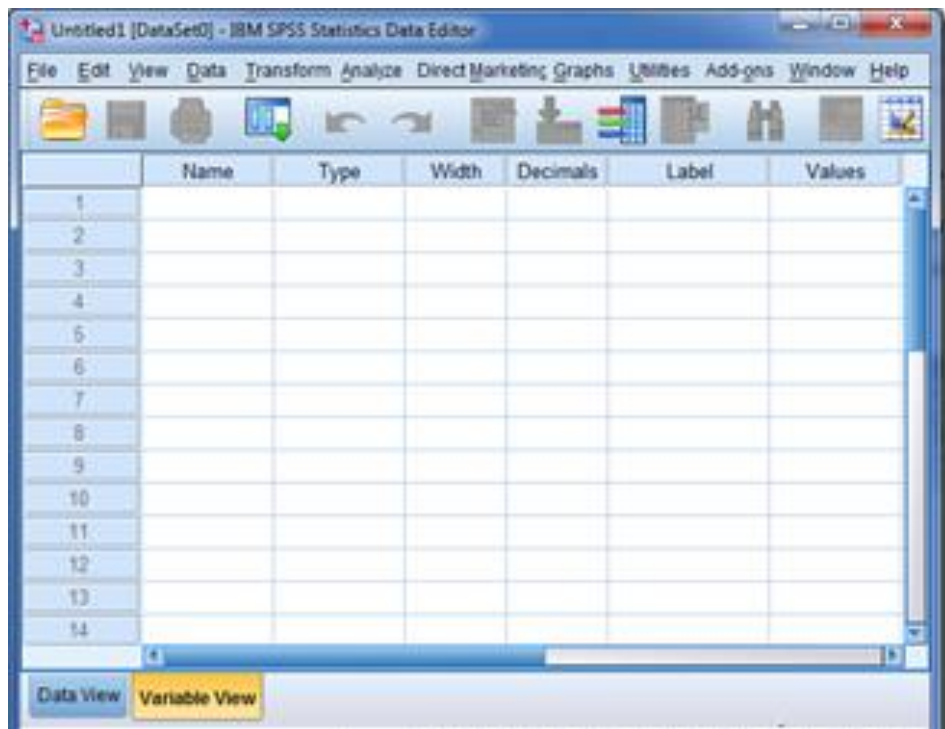
Jendel data editor berfungsi sebagai tempat kita memasukkan data. Berbagai data yang akan kita analisis kita masukkan melalui jendela data editor ini. Tampilan jendela data editor terdiri atas 2 macam, yaitu Data View dan Variabel View. Secara default, ketika SPSS kita buka maka yang akan ditampilkan adalah Data View. Menu untuk beralih dari Data View ke Variabel View dan sebaliknya berada di sebelah kiri bawah.

Data View digunakan untuk memasukkan nilai data yang akan dianalisis. Variabel View digunakan untuk memberi nama varibel serta menentukan type variabel. Pada bagian Variabel View ini kita juga bisa melakukan pengaturan variabel, misalnya jumlah desimal di belakang koma yang akan di tampilkan, jenis label, nilai, dan sebagainya.

- Tampilan Data View

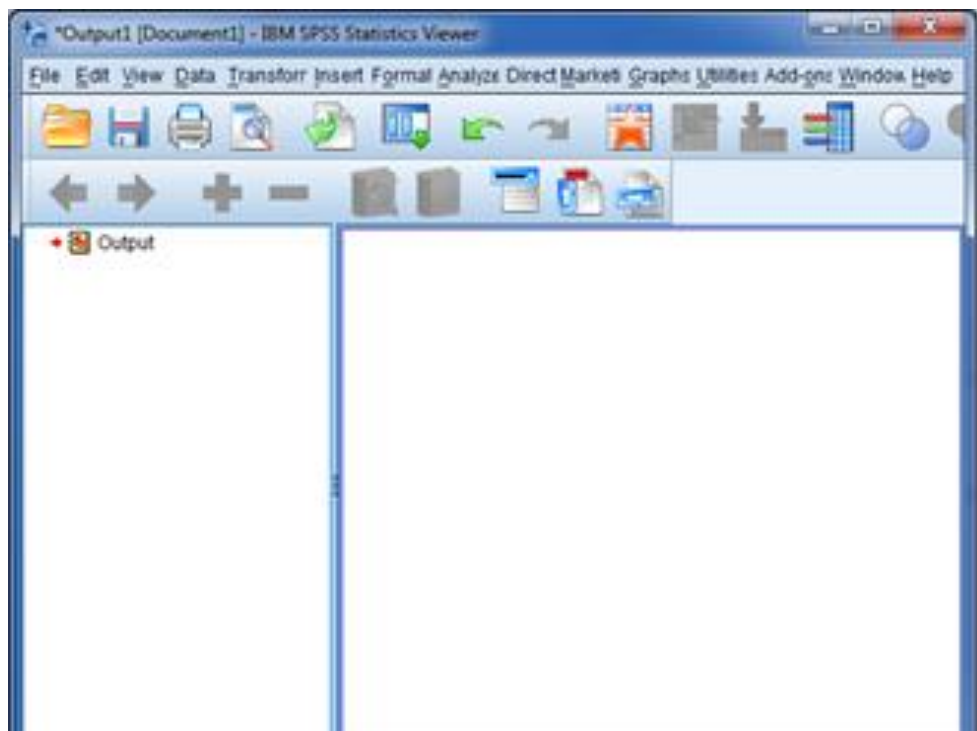


- Tampilan Variabel View



## 2. Jendela Output

Jendela output merupakan jendela untuk menampilkan analisis dari SPSS. Semua hasil perhitungan akan ditampilkan melalui jendela output ini. Tampilan jendela output adalah sebagai berikut:





## B. PEMBUATAN, PENGISIAN, DAN PENYIMPANAN DATA SPSS

Cara menggunakan SPSS dimulai dari memasukkan data dari kuesioner ke SPSS. Tahap *entry / input* data merupakan tahap awal sebelum peneliti melakukan pengujian yang lainnya. Semisalkan, data dari kuesioner saya adalah sebagai berikut

Tabel 1. Data Karyawan Perusahaan Xy Tahun 2003

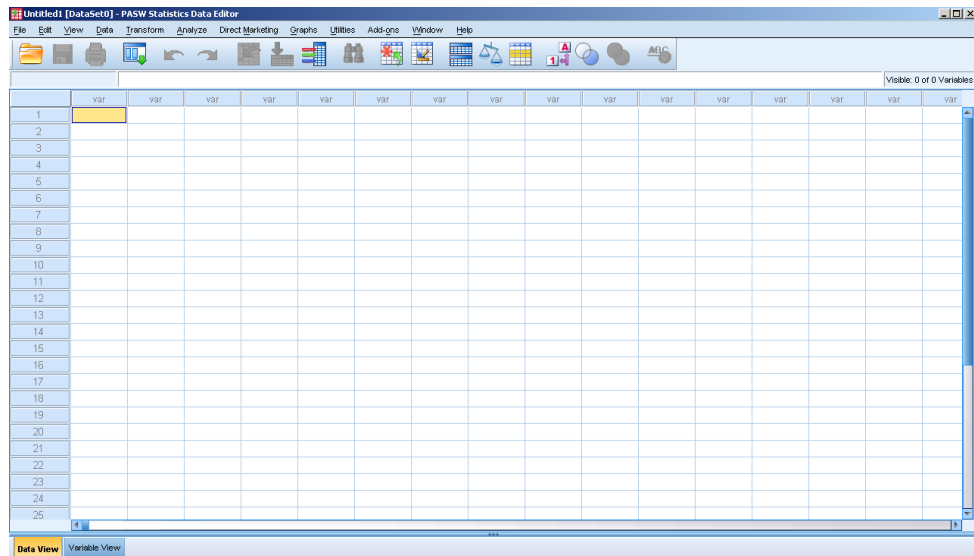
<b>Nama</b>	<b>Umur</b>	<b>Gender</b>	<b>Masa Kerja</b>	<b>Tinggi Badan</b>	<b>Berat Badan</b>
Mardiana	42	Laki - Laki	20	167	63
Sutisna	41	Laki - Laki	20	162	61
Suhardi	43	Laki - Laki	24	156	58
Suwanto	35	Laki - Laki	16	160	45
Slamet	30	Laki - Laki	13	157	59
Asep	37	Laki - Laki	18	158	57
Dana	35	Laki - Laki	13	163	63
Soleh	34	Laki - Laki	11	158	55
Saefudin	49	Laki - Laki	20	159	56
Hapipudin	44	Laki - Laki	10	155	50
Kasbi	43	Laki - Laki	21	161	56
Tujo	36	Laki - Laki	17	162	62
Supandi	35	Laki - Laki	14	162	50
Samsudin	33	Laki - Laki	15	170	70
Solihin	30	Laki - Laki	10	165	70
Agus	41	Laki - Laki	19	162	46
Rustiyah	38	Perempuan	22	162	45

Sukamtin	40	Perempuan	20	151	50
Suryadi	40	Laki - Laki	18	156	60
Musri P	38	Laki - Laki	19	159	54
Neni	37	Perempuan	24	150	44
Moh. Arif	21	Laki - Laki	1	169	63
Muksin	34	Laki - Laki	14	160	50
Dinar	36	Perempuan	11	156	53
Sarmiyati	45	Perempuan	26	146	63
Hadi	47	Laki - Laki	26	170	62
Wasiyah	39	Perempuan	20	153	57
Edy	35	Laki - Laki	12	160	68
Emi	31	Perempuan	12	154	44
Sopiah	46	Perempuan	24	155	54

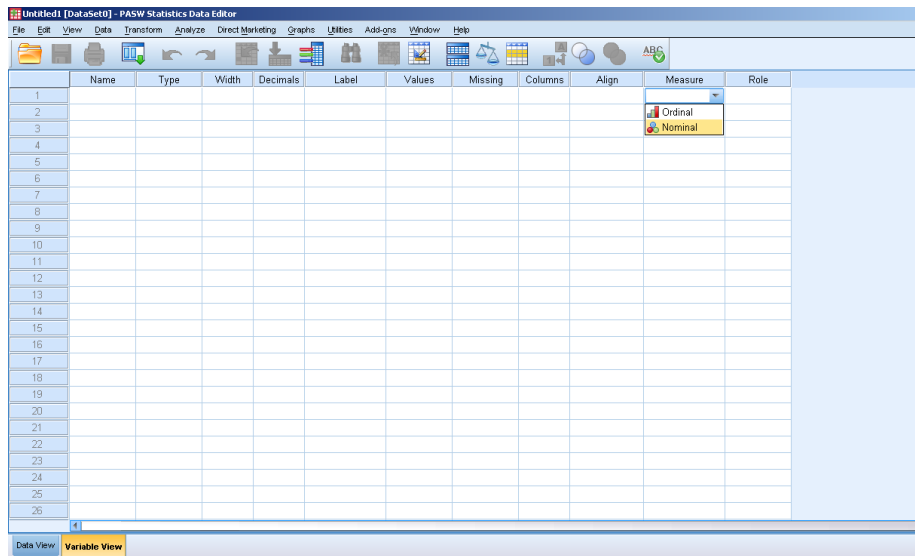
## 1. Buka software SPSS

Berikut tampilan yang akan kalian jumpai...sedikit serupa dengan microsoft excel, namun memiliki opsi pilihan yang berbeda. dan selalu terdiri dari 2 sheet, yaitu "data view" dan "variable view".

*Data View* adalah *tab sheet* yang menampilkan nilai data yang sebenarnya atau label nilai yang didefinisikan. Pada *tab sheet* inilah, anda mengentri data ke SPSS. Diatasnya terdapat menu-menu seperti File, Edit, View, Windows, dan Help seperti halnya menu-menu umum pada aplikasi under Windows lainnya.




- a) **Data.** Menu ini menampilkan submenu untuk melakukan perubahan-perubahan data, seperti mengurutkan data, memisahkan isi *file* dengan kriteria tertentu, menggabungkan data, etc.
- b) **Transform.** Menu untuk transformasi data, seperti menghitung variabel data, mengubah data, meranking data, etc.
- c) **Analyze.** Menu yang menjadi pusat pengolahan data, seperti mengolah statistik deskriptif, regresi, korelasi, etc.
- d) **Graphs.** Menu untuk menampilkan data dan hasil pengolahan data dalam bentuk grafik dan *chart*, seperti *bar charts*, *histogram*, *scatter diagram*, etc.
- e) **Utilities.** Menu pelengkap dalam pengoperasian SPSS, seperti menampilkan informasi variabel, mendefinisikan dan menampilkan variabel data, etc.



*Variable View* adalah *tab sheet* yang menampilkan kamus *metadata* di mana setiap baris mewakili sebuah variabel dan memperlihatkan nama variabel, jenis data (misal: *numeric*, *string*, *date*), lebar cetak, dan berbagai karakteristik lain. Menu yang tersedia dalam Variabel View diantaranya:

- a) **Name.** Kolom ini untuk memberikan informasi tentang nama variabel data. Nama variabel yang kita tuliskan di sini akan muncul pada Data View. Beberapa aturan penamaan variabel:
  - Nama variabel maksimal 8 karakter.
  - Nama diawali dengan huruf (tidak bisa dimulai dengan angka), sisanya dapat berisi huruf, angka, titik, atau simbol @, #, \_, atau \$. Kosong dan karakter khusus lain (misal: !, ?, ' , dan \*) tidak dapat digunakan.
  - Nama tidak bisa berakhir dengan titik dan tidak harus diakhiri dengan garis bawah “\_”.
  - nama variabel harus unik; duplikasi tidak diperbolehkan.
  - Nama-nama variabel tidak case sensitif, “Nama”, “nama”, dan “naMa” semua dianggap sama.
- b) **Type.** Kolom ini untuk memberikan jenis variabel data yang digunakan seperti angka, huruf atau desimal.
- c) **Label.** Kolom ini menunjukkan tambahan informasi dengan memberi label pada variabel data yang kita inginkan.
  - Variabel: ”nama” kita beri label: ”responden”
  - Variabel: ”umur” kita beri label ”umur”

- Variabel "gender" kita beri label "jenis kelamin", dan seterusnya
- Pemberian label dapat membantu dalam interpretasi hasil analisis (*output*) karena definisi *output* lebih jelas.
- d) **Value.** Kolom ini untuk memberikan label *string* yang diterapkan untuk nilai *numeric* tertentu, biasanya untuk data yang bersifat *ordinal* dan *interval*, misal
- Angka 1 untuk laki-laki dan angka 2 untuk perempuan.
  - Klik ikon  pada kolom Values maka akan muncul *dialog box*
- e) **Value Labels.** Misalnya untuk variabel "gender" kita akan mendefinisikan "jenis kelamin" dengan memberi label: Isi [Value] dengan angka 1 dan [Label] dengan "Laki-laki" lalu klik [Add], kemudian ulangi langkah-langkah tersebut untuk jenis kelamin "Perempuan". Nantinya pada variabel *gender*, kita tidak perlu menuliskan laki-laki dan perempuan melainkan cukup mengisi 1 untuk laki-laki dan 2 untuk perempuan
- f) **Missing.** Kolom ini menunjukkan nilai yang hilang (*missing value*) dalam data (jika ada). Responden dapat menolak untuk menjawab pertanyaan tertentu, mungkin tidak tahu jawabannya, atau mungkin menjawab dalam bentuk lain. Jika anda tidak mengidentifikasi data ini, analisis anda mungkin tidak memberikan hasil yang akurat.
- g) **Columns.** Kolom ini menunjukkan lebar kolom. baik jenis data *numeric* maupun *string*, lebar maksimal 255 digit.
- h) **Align.** Kolom ini menunjukkan posisi data pada tiap *cell*. Terdapat tiga pilihan posisi data, yaitu: left (rata kiri), right (rata kanan), dan center (rata tengah). fungsi ini sama persis seperti yang ada di microsoft office.
- i) **Measure.** Kolom ini menunjukkan jenis ukuran data yang digunakan. Terdapat tiga pilihan jenis ukuran data, yaitu:
- Scale/interval  
Kelompok data interval adalah kelompok data kuantitatif. Angka yang digunakan pada data ini menunjukkan suatu urutan dan dapat dilakukan operasi matematika. Angka nol pada data interval bukan seperti angka nol pada arti sesungguhnya. seperti usia 0-5 tahun, 6-11 tahun, 12-16 tahun
  - Nominal

Data nominal adalah kelompok data kualitatif dan merupakan level data paling sederhana. Apabila pada pengambilan data yang dihasilkan hanya berupa kategori maka data tersebut adalah data Nominal. seperti kategori (1) untuk laki-laki dan kategori (2) untuk perempuan.

- **Ordinal**

Data ordinal adalah data kelompok kualitatif di atas data Nominal. Jika pada data ordinal semua data setara maka pada data ordinal ada klasifikasi berdasarkan tingkatannya. Tingkatan ini berdasarkan kriteria tertentu pada saat pengambilan data. seperti kategori (1) untuk SD kategori (2) untuk SMP, kategori (3) untuk SMA

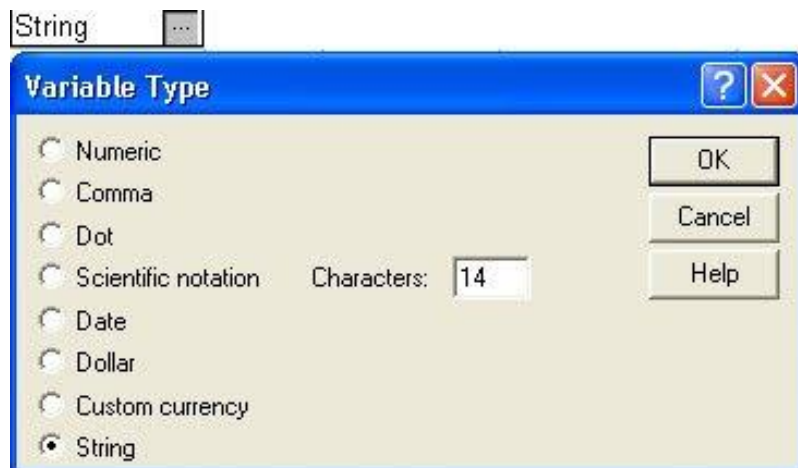
- **Rasio Data rasio** adalah tipe data level tertinggi pada pengukuran .Data ini termasuk data kuantitatif angka yang digunakan pada data ini adalah angka sesungguhnya. sehingga dapat dilakukan operasi matematika. Angka nol memiliki nilai yang sesungguhnya.

## 2. Memasukkan metadata pada *variable view*.

Langkah ini seperti membuat master sebelum data pada kuesioner kita masukkan. sebagian orang menyebut langkah ini sebagai *koding* atau pengkodean atau klasifikasi golongan data.

### a) **Variabel pertama: Nama**

Kolom Name atau nama variabel, ketik dengan nama untuk menamai responden. Pilihan Type atau tipe data, karena data berupa non-angka, maka diisi tipe string. Dengan banyak karakter 14, karena banyaknya karakter pada contoh paling banyak 14 dan klik OK, Tampak di layar:



- Karena ini Anda telah mendefinikan variable Nama mempunyai tipe variable string maka kolom decimal tidak aktif dan bernilai 0.
- Untuk menjelaskan nama dari variable “nama” maka kita masukan dalam kolom Label “ Nama Responden”.
- Kolom Values. Anda isi dengan none. Ini menyatakan bahwa tidak ada nilai numeric pada variable Nama yang perlu di beri nilai tertentu. Tentu saja hal ini akan berlaku untuk semua data bertipe string.
- Kolom missing, Anda isi dengan None. Ini menyatakan bahwa Anda tidak berharap adanya data yang hilang dan tentu saja ini sesuai dengan contoh.
- Kolom Columns, akan terlihat angka 8. ini merupakan angka default. Karena nama variable “Nama” mempunyai karakter kurang dari 8 maka tentu saja nilai ini dapat kita rubah minimal 4. dengan pertimbangan meingisikan nilai 4 pada Columns akan menyebabkan tertutupnya data nama responden yang mempunyai karakter maksimal 8 dalam hal ini maka sebaiknya dalam Columns mempunyai nilai 8 tepatnya 14.
- Kolom Align, secara default akan berisi left. Tentunya nilai left ini Anda dapat rubah sesuai keinginan Anda, yaitu dapat Right atau Center.
- Karena ini merupakan data string maka kolom Measure akan berisi nilai nominal.

Setelah semua proses pengisian kolom-kolom Data Editor (angka 1-seterusnya) pada variable view selesai, sekarang anda tinggal memasukan data kedalam variable nama, yaitu nama-nama responden. Dalam memasukan data, Anda dapat memasuknya satu per satu atau copy – paste.

**b) Variable kedua: Umur**

- Kolom Name, isi dengan Umur.
- Kolom Type atau tipe data adalah numeric /angka, dengan Width adalah 8 dan Desimal Places adalah 0. Dipilih desimal 0 karena jenis kelamin berupa kode dan bilangan bulat.
- Kolom Value, pilih None
- Kolom Decimals akan berisi 0 aktif, karena 0 telah kita definisikan dalam kolom type dan aktif dikarenakan kita mendefinisikan Jenis\_Kelamin sebagai variable bertipe numeric.
- Kolom Label atau keterangan variabel – untuk keseragaman – ketik Umur Responden dalam Tahun.
- Kolom Columns, isi dengan 10
- Kolom Align, isi dengan Right
- Kolom Measure isi dengan Scale.

Setelah semua proses pengisian kolom-kolom Data Editor pada variable view selesai, sekarang anda tinggal memasukan data kedalam variable Umur. Dalam memasukan data, Anda ketik satu per satu atau melalui copy – paste.

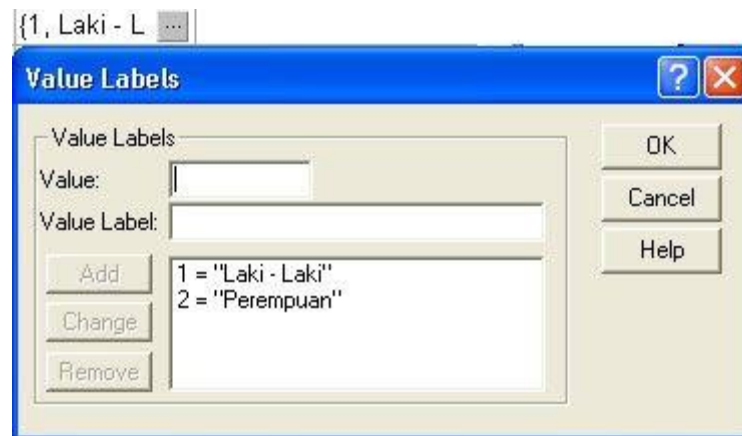
**c) Variabel ketiga: Jenis\_Kelamin**

Prinsipnya sama dengan pengisian pada variabel pertama, yaitu klik Variable View pada bagian kiri bawah dari tampilan pertama SPSS

- Kolom Name, isi dengan Jenis\_Kelamin untuk menamai jenis kelamin responden.
- Kolom Type atau tipe data adalah numeric /angka, dengan Width adalah 8 atau lebih dan Desimal Places adalah 0. Dipilih desimal 0 karena jenis kelamin berupa kode dan bilangan bulat.



- Kolom Value. Seperti diketahui, perhitungan dalam SPSS selalu untuk tipe data numeric. Untuk itu jenis kelamin harus dijadikan numerik pula, yaitu dengan tanda: 1 = tanda Laki-Laki, dan 2 = tanda Perempuan. Penulisan kode bisa bebas, misal 11 atau 12 untuk pria, dan variasi lainnya. Tapi menurut kebiasaan pria akan disebut pertama kali.



- Kolom Decimals akan berisi 0 aktif, karena 0 telah kita definisikan dalam kolom type dan aktif dikarenakan kita mendefinisikan Jenis\_Kelamin sebagai variable bertipe numeric
  - Kolom Label atau keterangan variabel – untuk keseragaman – ketik Jenis Kelamin Responden.
  - Kolom Measure isi dengan nominal. Ini menyatakan bahwa data yang Anda masukan nantinya berupa nilai yang berbeda tetapi tidak membedakan. Artinya nilai 1 dan 2 yang Anda masukan berupa nilai yang mewakili pria dan wanita dengan pria dan wanita mempunyai posisi yang sama, yaitu pria tidak lebih tinggi dari wanita begitu pula sebaliknya. (karena jenis kelamin masuk pada data nominal)
  - Kolom Columns, isi dengan 10
  - Kolom Align, isi dengan Right
- Setelah semua proses pengisian kolom-kolom Data Editor pada variable view selesai, sekarang anda tinggal memasukan data kedalam variable Jenis\_Kelamin, yaitu jenis kelamin responden. Dalam

memasukan data, Anda masukan 1 untuk responden pria dan 2 untuk wanita.

**d) Variabel ke-empat: Masa\_Kerja**

- Kolom Name, isi dengan Masa\_Kerja.
- Kolom Type atau tipe data adalah numeric /angka, dengan Width adalah 8 dan Desimal Places adalah 0. Dipilih desimal 0 karena jenis kelamin berupa kode dan bilangan bulat.
- Kolom Value, pilih None
- Kolom Decimals akan berisi 0 aktif, karena 0 telah kita definisikan dalam kolom type dan aktif dikarenakan kita mendefinisikan Jenis\_Kelamin sebagai variable bertipe numeric.
- Kolom Label atau keterangan variabel – untuk keseragaman – ketik Masa Kerja Responden dalam.
- Kolom Columns, isi dengan 10
- Kolom Align, isi dengan Right
- Kolom Measure isi dengan Scale.

Setelah semua proses pengisian kolom-kolom Data Editor pada variable view selesai, sekarang anda tinggal memasukan data kedalam variable Masa\_Kerja, yaitu masa kerja responden. Dalam memasukan data, Anda ketik satu

**e) Variabel kelima: Tinggi\_Badan**

- Kolom Name, isi dengan Tinggi\_Badan.
- Kolom Type atau tipe data adalah numeric /angka, dengan Width adalah 8 dan Desimal Places adalah 0. Dipilih desimal 0 karena jenis kelamin berupa kode dan bilangan bulat.
- Kolom Value, pilih None
- Kolom Decimals akan berisi 0 aktif, karena 0 telah kita definisikan dalam kolom type dan aktif dikarenakan kita mendefinisikan Jenis\_Kelamin sebagai variable bertipe numeric.
- Kolom Label atau keterangan variabel – untuk keseragaman – ketik Tinggi Badan Responden dalam cm.
- Kolom Columns, isi dengan 10
- Kolom Align, isi dengan right

- Kolom Measure isi dengan Scale.

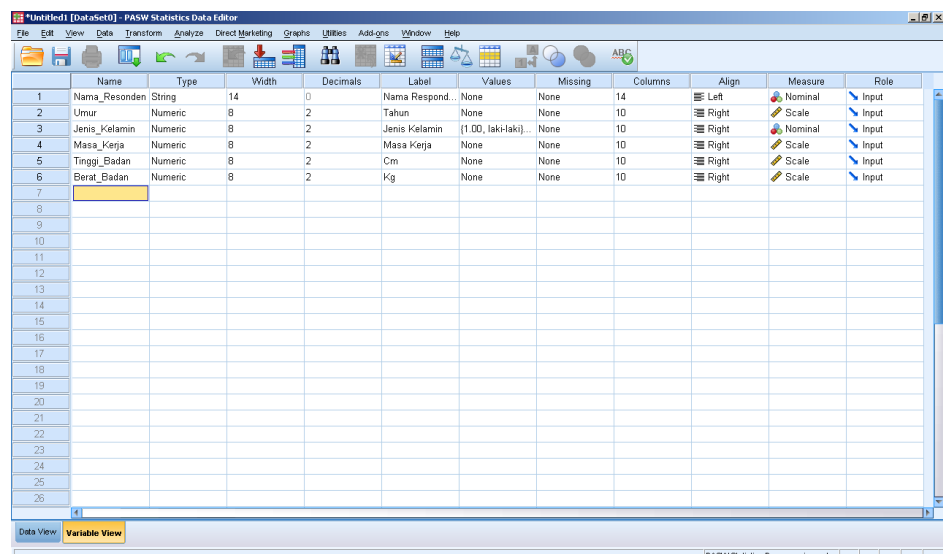
Setelah semua proses pengisian kolom-kolom Data Editor pada variable view selesai, sekarang anda tinggal memasukan data kedalam variable Tinggi\_Badan. Dalam memasukan data, Anda ketik satu per satu

e) **Variabel keenam: Berat\_Badan**

- Kolom Name, isi dengan Berat\_Badan.
- Kolom Type atau tipe data adalah numeric /angka, dengan Width adalah 8 dan Desimal Places adalah 0. Dipilih desimal 0 karena jenis kelamin berupa kode dan bilangan bulat.
- Kolom Value, pilih None
- Kolom Decimals akan berisi 0 aktif, karena 0 telah kita definisikan dalam kolom type dan aktif dikarenakan kita mendefinisikan Jenis\_Kelamin sebagai variable bertipe numeric.
- Kolom Label atau keterangan variabel – untuk keseragaman – ketik Berat Badan Responden dalam kg
- Kolom Columns, isi dengan 10
- Kolom Align, isi dengan Right
- Kolom Measure isi dengan Scale.

Setelah semua proses pengisian kolom-kolom Data Editor pada variable view selesai, sekarang anda tinggal memasukan data kedalam variable Berat\_Badan. Dalam memasukan data, Anda ketik satu per satu

- Tampilan variabel view



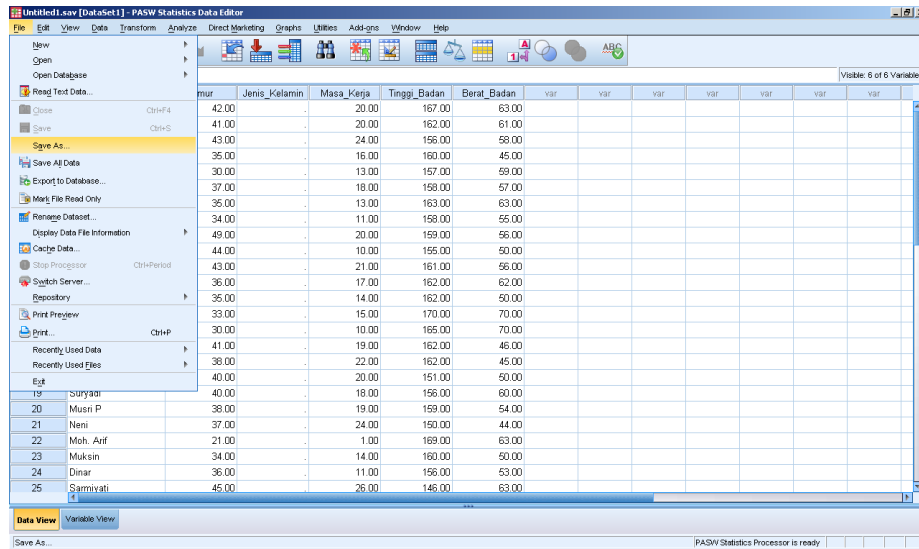
- Tampilan data view

The screenshot shows the SPSS Data Editor window with the following data:

	Nama_Responden	Umur	Jenis_Kelamin	Masa_Kerja	Tinggi_Badan	Berat_Badan	var	var	var	var	var	var	var
1	Mardiana	42.00	.	20.00	167.00	63.00							
2	Sutisna	41.00	.	20.00	162.00	61.00							
3	Suhardi	43.00	.	24.00	156.00	58.00							
4	Suwanto	35.00	.	16.00	160.00	45.00							
5	Slamet	30.00	.	13.00	157.00	59.00							
6	Asep	37.00	.	18.00	158.00	57.00							
7	Dana	35.00	.	13.00	163.00	63.00							
8	Soleh	34.00	.	11.00	158.00	55.00							
9	Saeudin	49.00	.	20.00	159.00	56.00							
10	Hapipudin	44.00	.	10.00	155.00	50.00							
11	Kasbi	43.00	.	21.00	161.00	56.00							
12	Tujo	36.00	.	17.00	162.00	62.00							
13	Supandi	35.00	.	14.00	162.00	50.00							
14	Samsudin	33.00	.	15.00	170.00	70.00							
15	Solihin	30.00	.	10.00	165.00	70.00							
16	Agus	41.00	.	19.00	162.00	46.00							
17	Rustiyah	38.00	.	22.00	162.00	45.00							
18	Sukamti	40.00	.	20.00	151.00	50.00							
19	Suryadi	40.00	.	18.00	156.00	60.00							
20	Muhsin P	38.00	.	19.00	159.00	54.00							
21	Neni	37.00	.	24.00	150.00	44.00							
22	Moh. Anif	21.00	.	1.00	169.00	63.00							
23	Muhsin	34.00	.	14.00	160.00	50.00							
24	Dinar	36.00	.	11.00	156.00	53.00							
25	Sarmiyati	45.00	.	26.00	146.00	63.00							

### 3. Menyimpan Data

Data yang telah dimasukan bisa disimpan dengan prosedur berikut :



- Dari menu utama SPSS, pilih menu File, kemudian pilih sub-menu “SAVE AS”
- Beri nama file – untuk keseragaman – tempatkan file pada directory yang dikehendaki.
- Untuk tipe data dipakai ekstensi (tipe) file SPSS adalah “sav”

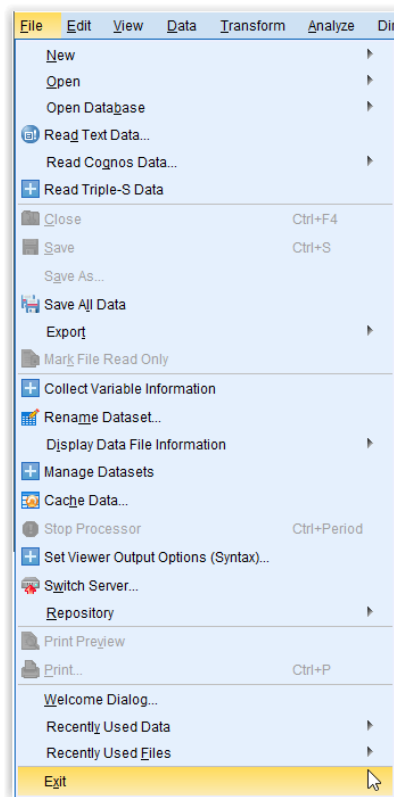
- Selain pilihan “SAVE AS”, ada juga pilihan “SAVE”. Pilihan “SAVE” dipakai jika penyimpanan data tidak memerlukan nama baru atau file sudah diberi nama

## C. MENU EDIT PADA SPSS

### 1. Menu Bar SPSS

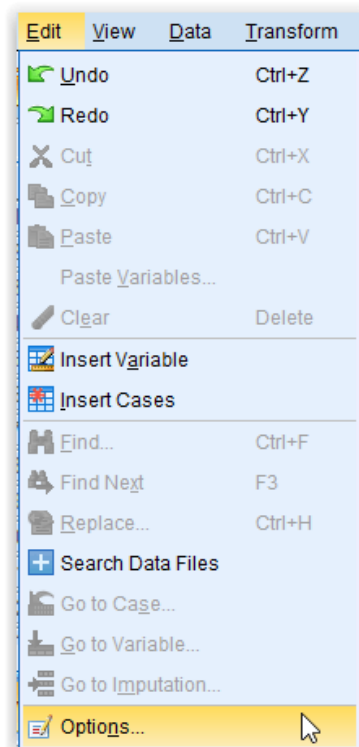
#### a) Menu File

Menu file adalah menu bar yang menyediakan perintah-perintah terkait fungsi manajemen file seperti membuka file, menyimpan file, export file, import file, dan lain-lain.



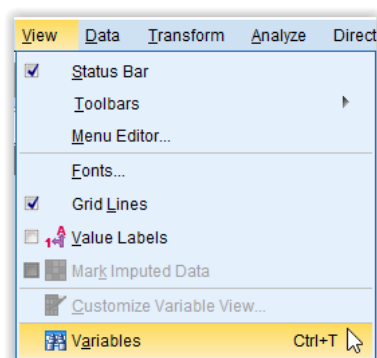
#### b) Menu Edit

Menu edit adalah menu bar dengan perintah-perintah yang berfungsi untuk melakukan sunting atau edit data. Misalnya undo, redo, cut, copy, paste, dan lain-lain.



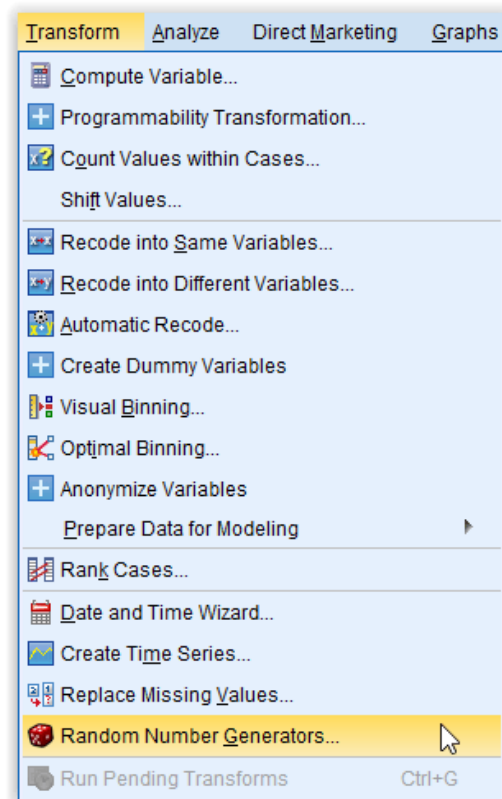
### c) Menu View

Menu View adalah menu yang berfungsi untuk melakukan konfigurasi tampilan dari jendela SPSS, terkait menampilkan dan menyembunyikan perintah. Beberapa diantaranya adalah status bar, toolbars, menu editor, font, grid lines, value labels, mark imputed data, custom variable view, dan variables.



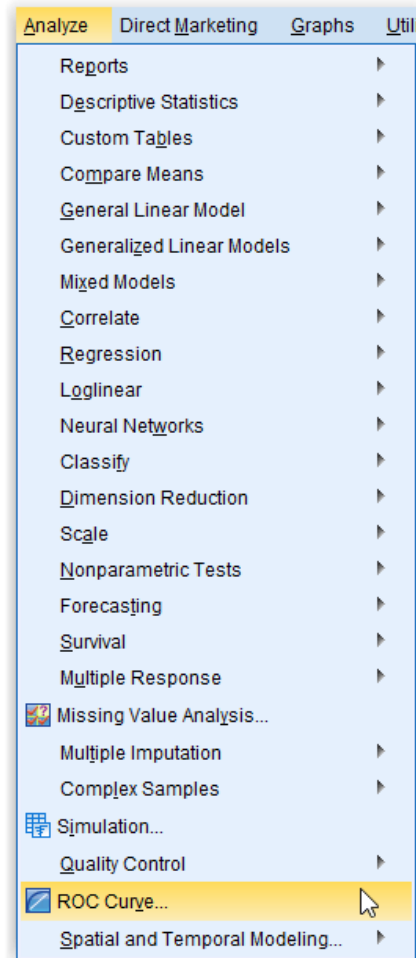
### d) Menu Transform

Menu transform adalah menu bar yang berfungsi untuk melakukan transformasi variabel terkait pembentukan variabel baru dari variabel-variabel yang telah ada.



**e) Menu Analyze**

Menu analyze adalah menu bar yang berfungsi untuk melakukan analisis statistika. Mulai dari analisis statistika deskriptif hingga analisis regresi, parametrik, serta non-parametrik.

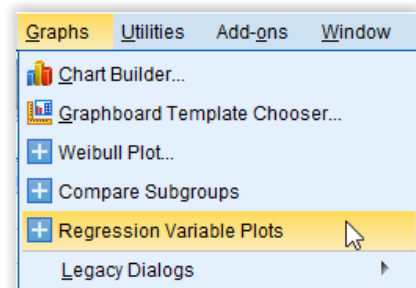


**f) Menu Direct Marketing**

Menu direct marketing adalah menu bar yang digunakan untuk analisis pemasaran untuk mengoptimalkan suatu bisnis.

**g) Menu Graph**

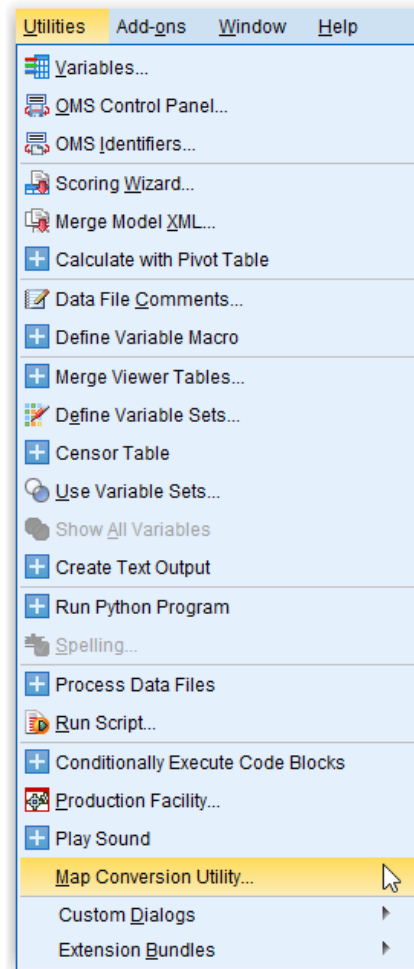
Menu graph adalah menu bar yang berfungsi untuk membuat visualisasi grafik menggunakan chart editor.



**h) Menu Utilities**

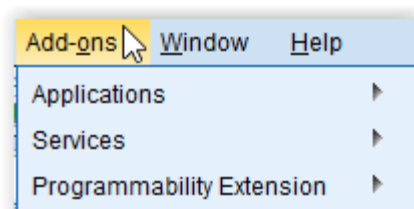
Menu utilities adalah menu bar yang berfungsi untuk menampilkan informasi data serta menjalankan script.





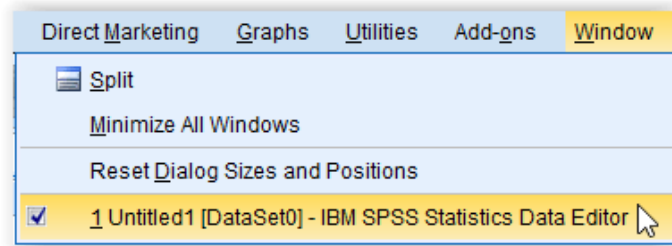
**i) Menu Add-ons**

Menu add-ons adalah menu bar yang menyediakan paket instalasi SPSS lainnya baik aplikasi, module, dan extension.



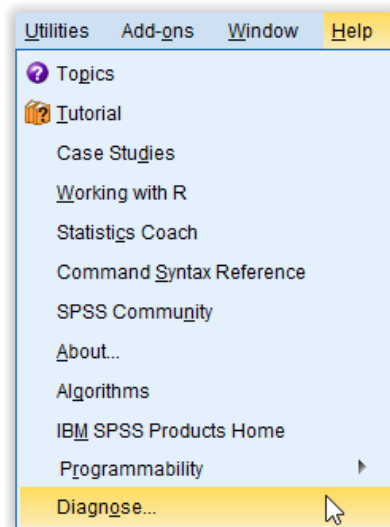
**j) Menu Window**

Menu Window adalah menu bar yang menyediakan perintah terkait jendela SPSS, misalnya split, mengganti active window, mengatur tampilan awal (reset), dan minimize.



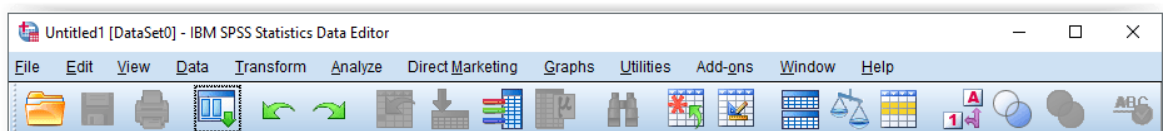
### k) Menu Help












Menu help adalah menu bar yang menyediakan informasi dan bantuan penggunaan perangkat lunak SPSS.





## 2. Toolbar SPSS

Toolbar pada SPSS memberikan shortcut perintah berupa ikon yang dapat mempercepat pekerjaan terkait beberapa fungsi menu bar pada SPSS. Berikut ilustrasi toolbar dari perangkat lunak SPSS,

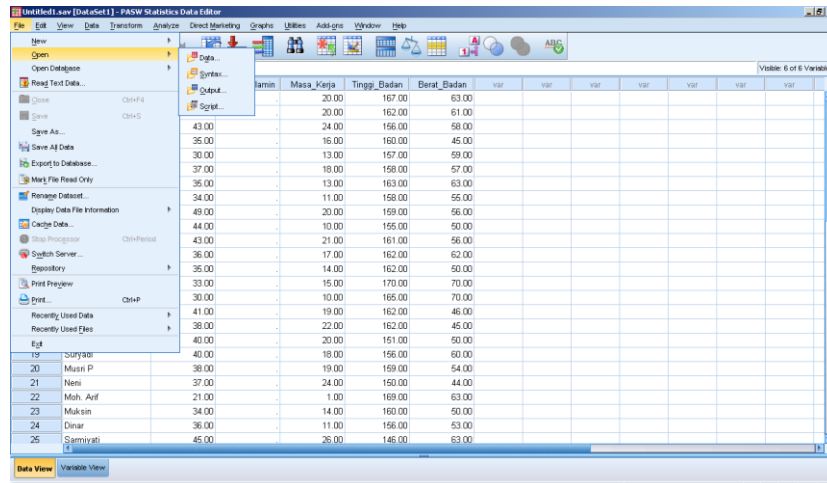


Ikon toolbar	Nama Ikon	Fungsi
	Open data document	Untuk membuka file yang tersimpan, sama dengan perintah <b>File &gt; Open &gt; Data</b>
	Save this document	Untuk menyimpan file yang sedang aktif, sama dengan <b>File &gt; Save (Ctrl+S)</b>
	Print	Untuk mencetak data yang aktif pada Data View, sama dengan <b>Ctrl+P</b>
	Recall recently used dialogs	Untuk memanggil dialog perintah yang sering digunakan
	Undo	Untuk membatalkan perintah sebelumnya, sama dengan <b>Ctrl+Z</b>
	Redo	Untuk mengembalikan perintah yang telah dibatalkan sebelumnya, sama dengan <b>Ctrl+Y</b>
	Go to case	Untuk berpindah atau menyorot suatu kasus yang spesifik, sama dengan <b>Edit &gt; Go to Case</b> .
	Go to variable	Untuk menyorot variable tertentu atau yang spesifik, sama dengan <b>Edit &gt; Go to Variable</b>
	Variable	Untuk melihat informasi setiap variabel pada lembar kerja yang aktif, sama dengan <b>Utilities &gt; Variables</b> .
	Run descriptive statistics	Menjalankan fungsi statistika deskriptif, yang secara default menampilkan frekuensi data. Variabel yang disorot pada lembar kerja aktif akan diproses untuk analisis statistika deskriptif. Sama dengan <b>Analyze &gt; Descriptive Statistics &gt; Frequencies</b>
	Find	Befungsi untuk mencari suatu nilai dari

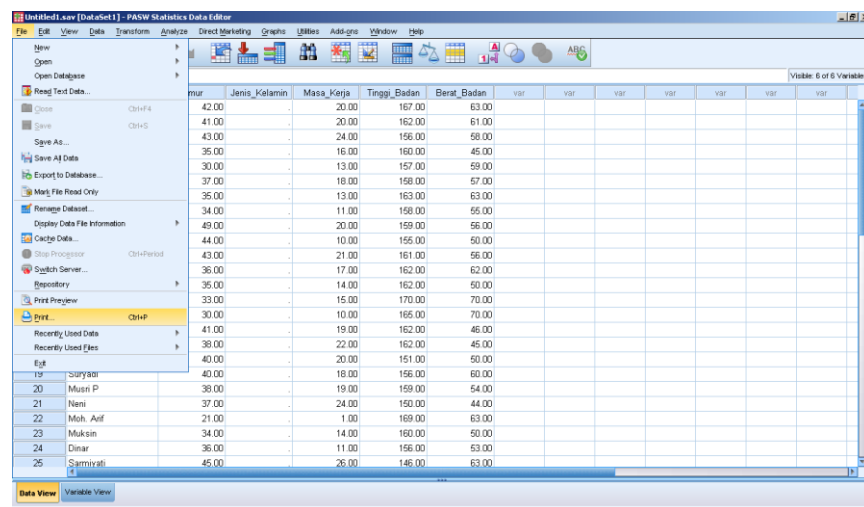
		Data View yang aktif dan juga tersedia untuk menggantikan nilai tersebut. Shortcut ini merupakan perintah dari <b>Find (Ctrl+F)</b> dan <b>Replace (Ctrl+H)</b> pada menu <b>Edit</b>
	Insert cases	Untuk menambah kasus baru di sel yang disorot, sama dengan <b>Edit &gt; Insert Variable</b>
	Insert variable	Untuk menambah atau menyisipkan variabel baru, sama dengan <b>Edit &gt; Insert Variable</b>
	Split file	Untuk memecah dataset berdasarkan variabel. Hal ini dapat mempermudah analisis khusus terhadap variabel tertentu. Sama dengan <b>Data &gt; Split File</b>
	Weight cases	Untuk menentukan bobot variabel, sama dengan <b>Data &gt; Weight Cases</b>
	Select cases	Untuk menyorot kasus yang spesifik, sama dengan <b>Data &gt; Select Cases</b>
	Value labels	Untuk mengganti tampilan value atau label pada Data View, hal ini berkaitan dengan ordinal dan nominal measurement
	Use variable sets	Untuk memilih variabel yang akan digunakan untuk analisis data secara menyeluruh, sama dengan <b>Utilities &gt; Use Variable Sets</b>
	Show all variables	Untuk menampilkan semua variabel, sama dengan <b>Utilities &gt; Show All Variables</b>
	Spell check	Untuk mengecek kesalahan eja, bergantung konfigurasi bahasa yang aktif, sama dengan <b>Utilities &gt; Spelling</b>

#### D. MENCETAK DATA DAN KELUAR DARI SPSS

Data yang telah ada di SPSS bisa dicetak oleh printer dengan menu “**PRINT**”. Menu Print bisa mencetak seluruh data yang tampil ataupun sebagian data yang dikehendaki. Langkahnya sebagai berikut :



- Menu **File – Open** (kemudian pilih file yang diinginkan hingga tampil di data editor)



- Menu **File – Print** (pilih jenis printer yang menyala, dan pilih “**all**” jika akan mencetak semua halaman jika hanya beberapa pilihan halaman berapa yang akan di cetak)

## E. INSTALASI SPSS VERSI 18

SPSS merupakan sebuah program aplikasi yang memiliki kemampuan untuk melakukan analisis statistik yang sangat tinggi serta memiliki sistem manajemen data pada lingkungan grafis dengan menggunakan menu-menu deskriptif serta kotak dialog yang sederhana sehingga mudah untuk dipahami dan mudah untuk dioperasikan oleh para pengguna.

Pada saat ini aplikasi SPSS sudah memiliki banyak versi, versi pertama yang dirilis adalah SPSS/PC+ yang berbasis teks. Aplikasi dapat menggunakan program atau kode eksternal yang berarti aplikasi ini membutuhkan bantuan dari aplikasi lainnya yaitu berupa aplikasi editor. Sejalan dengan berkembangnya sistem operasi SPSS mulai dikembangkan untuk mendukung sistem operasi yang berbasis windows yang ditandai dengan dirilisnya SPSS 6.0 for Windows serta diikuti oleh versi-versi lainnya, seperti SPSS 7, 10, 11,5, 12, hingga SPSS 18. Aplikasi ini dibuat oleh SPSS Incorporation.

Berbicara mengenai SPSS versi 18 itu sendiri memiliki beberapa fitur yang dapat menunjang kinerja seorang yang bekerja di bidang statistika yang berguna untuk melakukan proses pengolahan data statistik. Adapun fitur yang ditawarkan oleh SPSS 18 ini sebagai berikut:

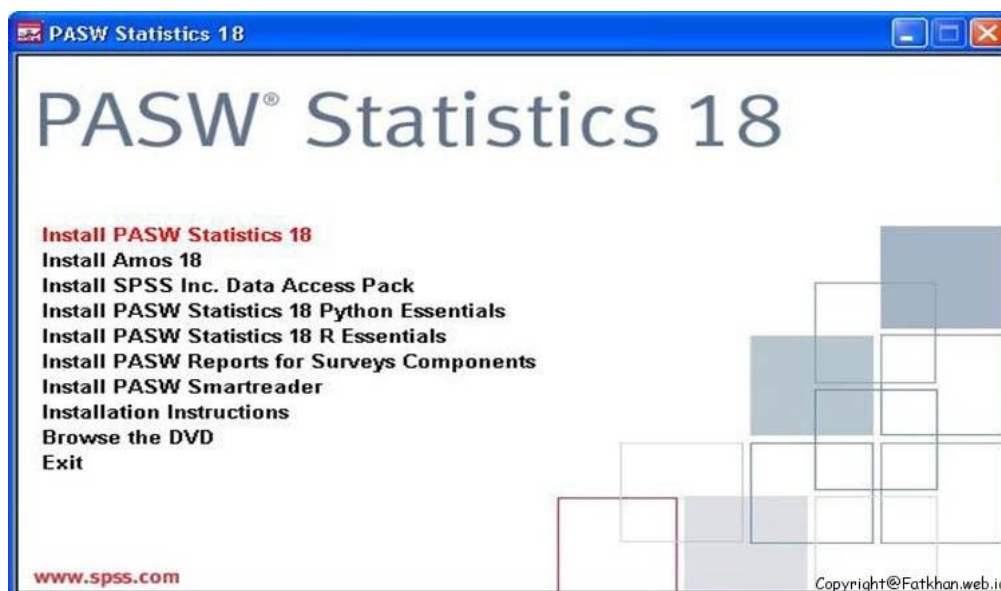
1. SPSS menyediakan tool-tool dalam pendefinisian data
2. Adanya tool copy data properties wizard yang memberikan kemudahan dalam pemakaian file data eksternal, seperti template untuk mendefinisikan file dan properti variable yang bekerja pada sebuah file data.
3. Dengan menggunakan spss 18 ini anda dapat menghasilkan output pivot table dengan berbagai bahasa yang berbeda dan dapat memilih bahasa yang ingin anda gunakan nantinya.
4. Memiliki tampilan grafik yang lebih baik.
5. Memiliki fitur manajemen data dan output yang dapat mengubah data scale menjadi data range.
6. Dapat melakukan kostumasi dan ekspor data dengan mudah.
7. Output data dapat ditampilkan menggunakan aplikasi power point.

Seperti halnya aplikasi terdahunya, proses penginstalan aplikasi ini membutuhkan hardware yang mendukung terutama dari segi resolusi monitor yang

cukup tinggi dan apabila memiliki resolusi monitor yang rendah aplikasi ini tidak dapat di install.

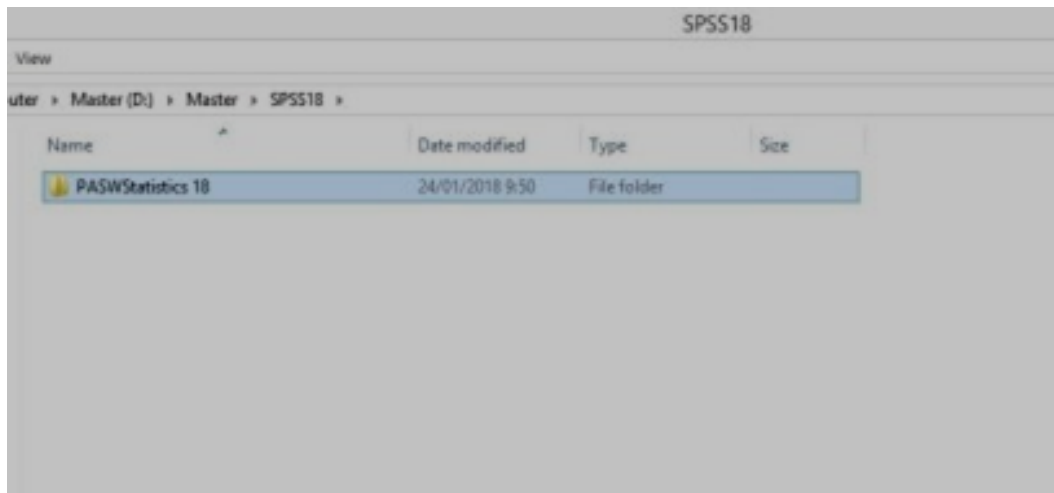
Sebelum menggunakan aplikasi SPSS ini, ada baiknya untuk memperhatikan proses instalasi sebagai berikut:

1. Run setup exe (jalankan penginstalan)
2. Single user license pilih saja (option ini)
3. Klik saja "no serial number" dan lanjutkan proses instalasinya
4. In the end uncheck "register with spss.com)
5. Close license authorization wizard (setelah selesai close program)
6. Copy content of EQX to "instal dir" (salin folder EQX yang mana crack full version ke program yang di instal)
7. Jalankan program untuk memulai belajar/menghitung statistik

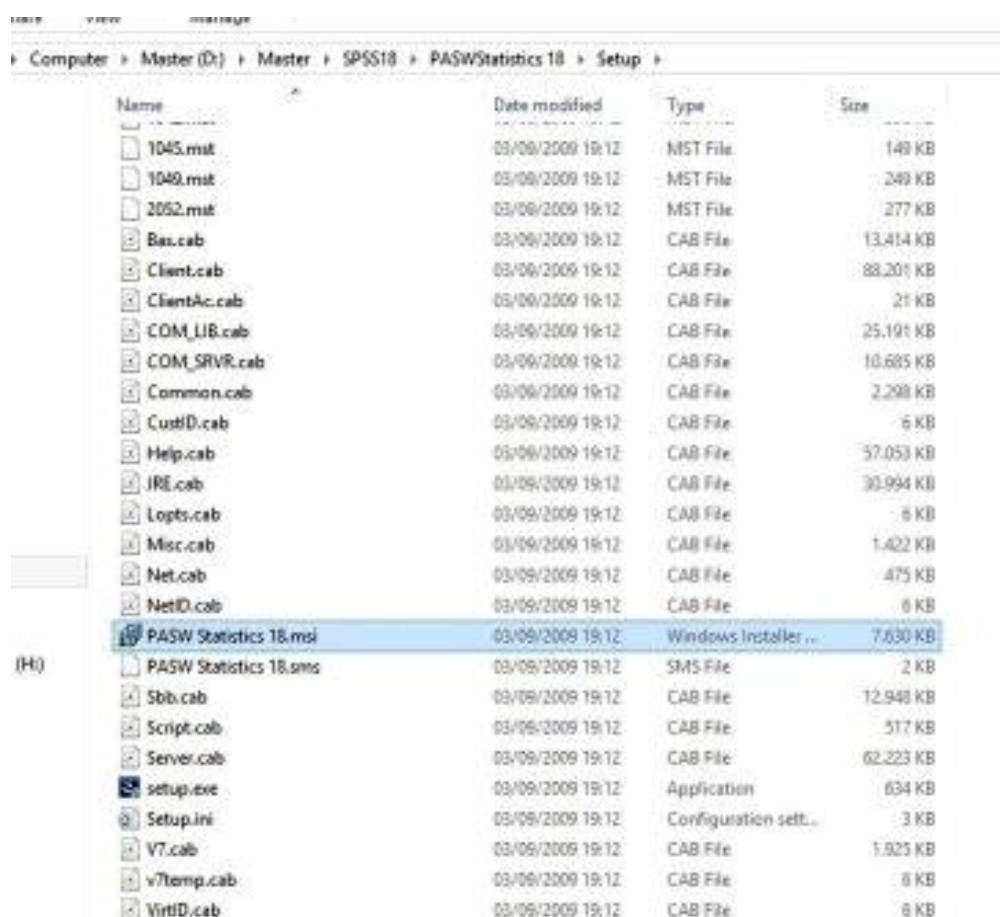


Berikut adalah cara instalansi SPSS

1. Pastikan anda sudah punya Software Installer SPSS Versi 18 Jika Belum ada silahkan download softwarena di sini (jika Link mati/broken silahkan PM saya atau berikan email anda di kolom komentar akan saya kirim link terbaru).
2. Setelah download atau jika anda sudah punya Software installernya silahkan ekstrak dan buka Folder SPSS 18

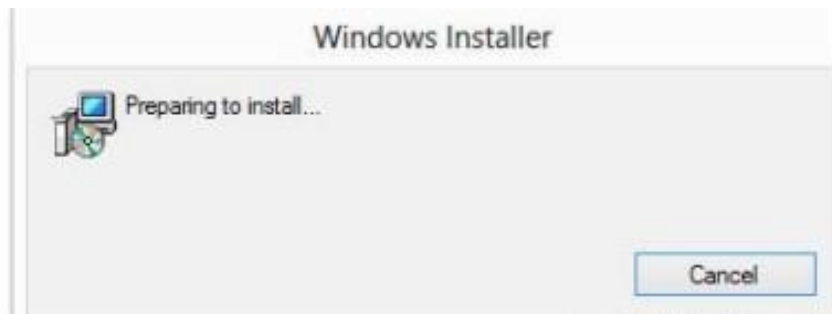


- Setelah itu cari folder Setup dan cari file "PASW Statistics 18.msi" atau "Setup.exe", setelah itu double klik salah satu file tersebut untuk memulai proses instalasi.





4. Tunggu proses instalasi berjalan.



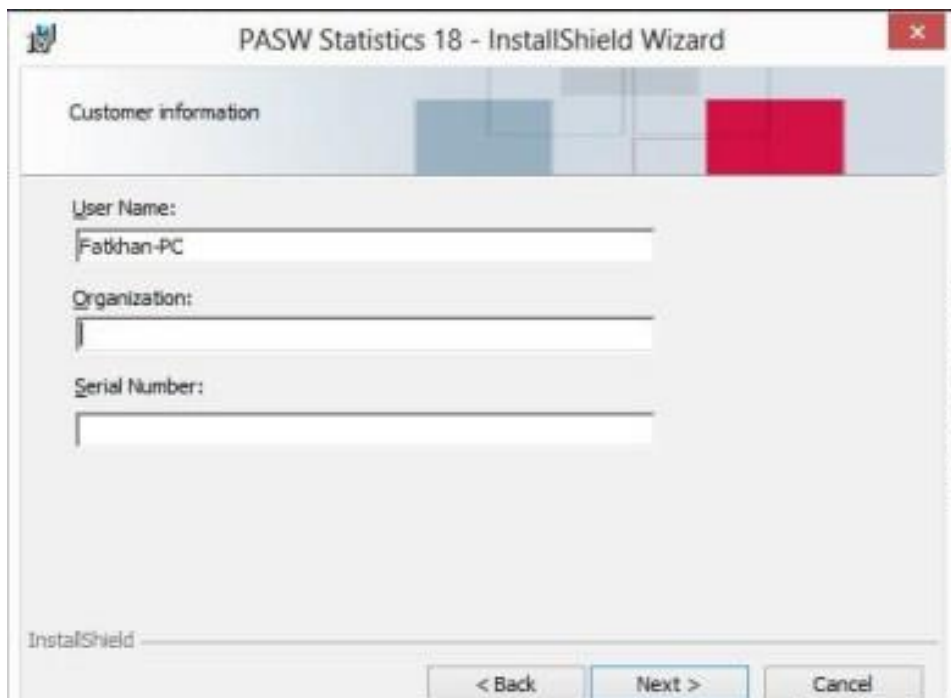
5. Sampai pada proses ini, pilih Single User License ( I purchased a single copy of the product ) kemudian klik Next



- Sampai pada tahap ini silahkan pilih " I accept the Terms in the license agreement" kemudian klik Next.



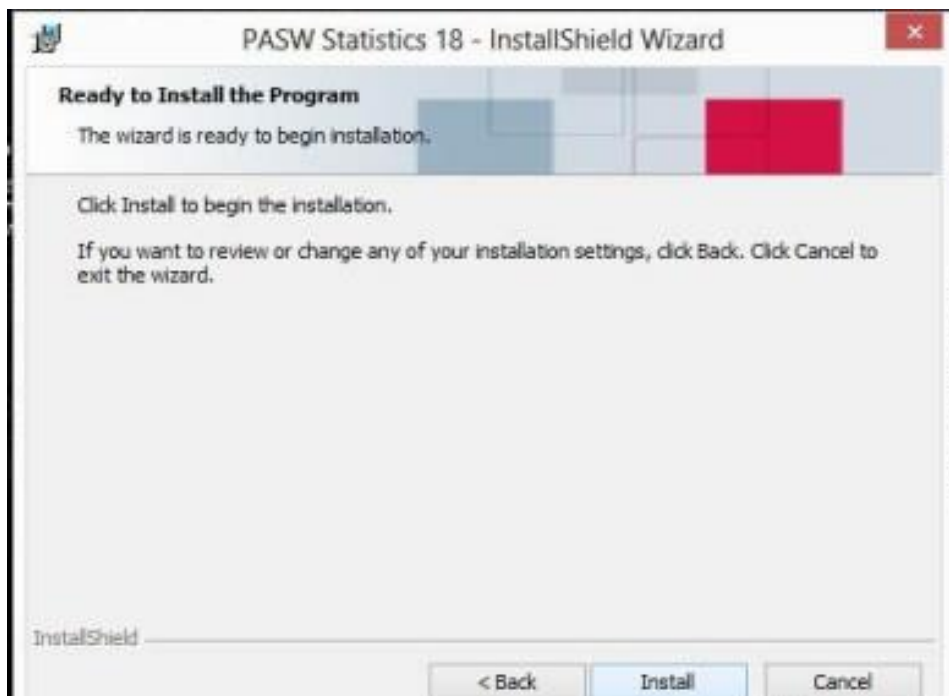
- Sampai pada tahap ini silahkan isi kotak "Organization" dan untuk kotak "Serial Number" Kosongkan saja, kemudian klik Next.



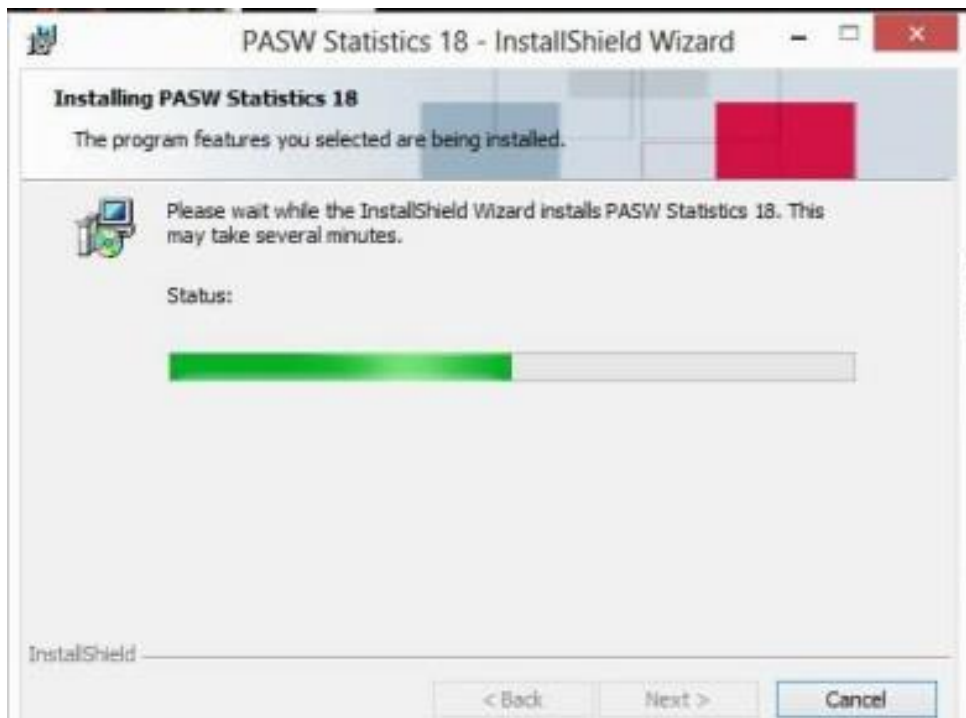
8. Sampai pada tahap ini merupakan proses penempatan file instalasi, defaultnya file instalasi ada di Drive C:\Program Files\SPSSInc\PASW Statistics18\. Jika tidak dirubah langsung saja klik Next.



9. Selanjutnya silahkan klik "Install".



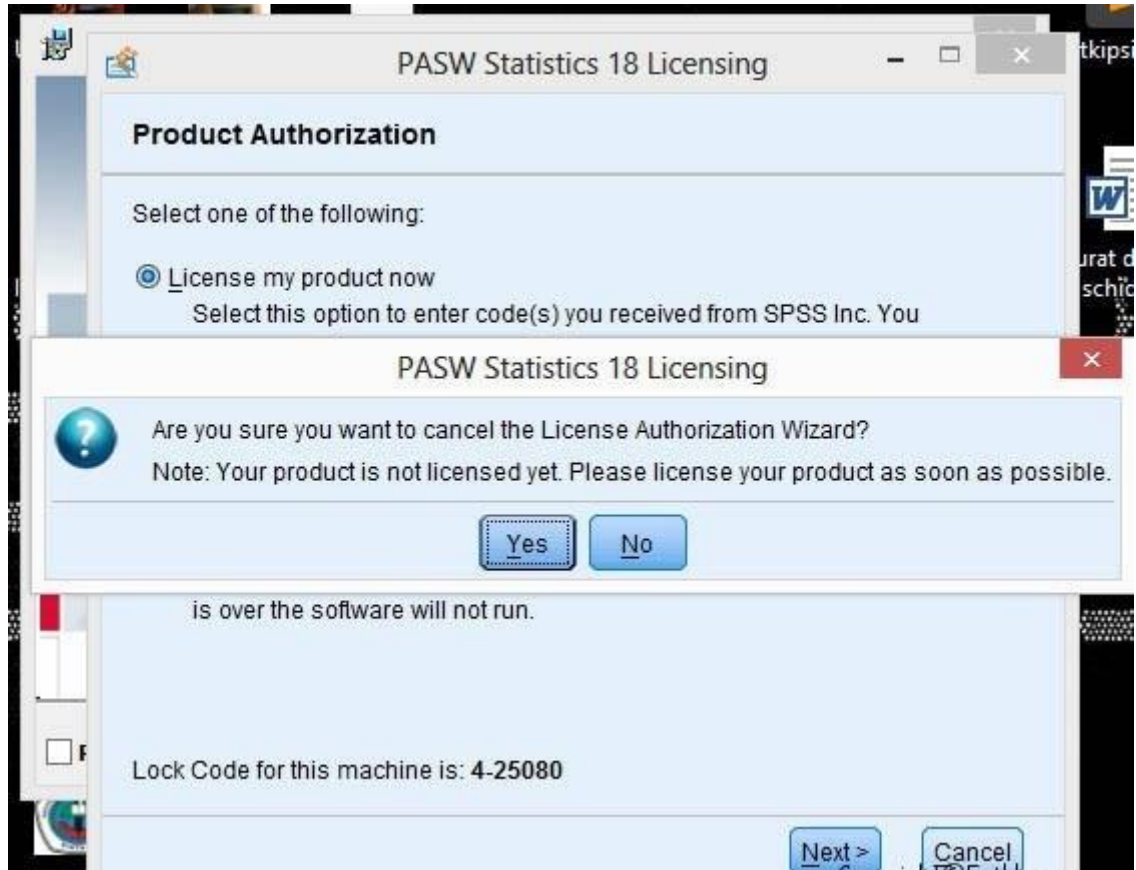
10. Tunggu sampai proses instalasi selesai seperti gambar dibawah ini.



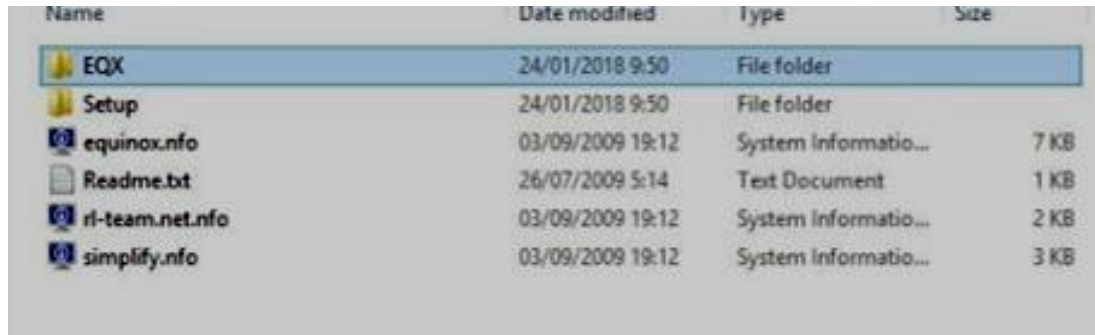
11. Setelah proses instalasi selesai seperti gambar di bawah ini. Silahkan Uncheck ( hilangkan checklist ) pada "Register with spss.com" )



12. Sampai pada tahap ini, jika diminta untuk memasukkan "Product Authorization" atau Serial Number di "Cancel" atau "No" saja. **Jangan** anda buka/jalankan dulu program SPSS anda, karena kita akan masuk ke proses aktivasi produk karena jika tidak di aktivasi maka produk/program SPSS tidak bisa digunakan krna fitur yang tersedia tidak ada.

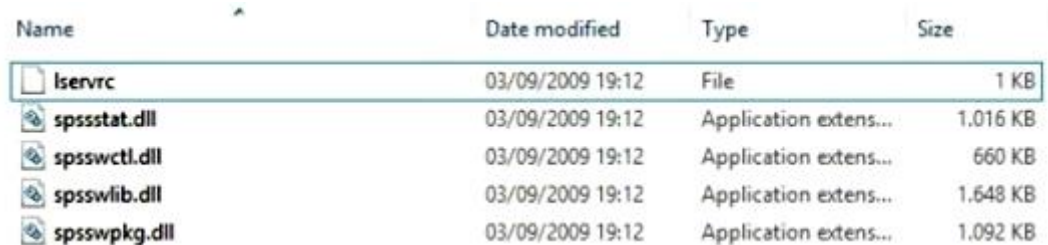


13. Langkah selanjutnya adalah masuk kembali ke folder PASWStatistics18 kemudian cari folder dengan nama "EQX" seperti pada gambar di bawah ini. Kemudian buka folder tersebut.



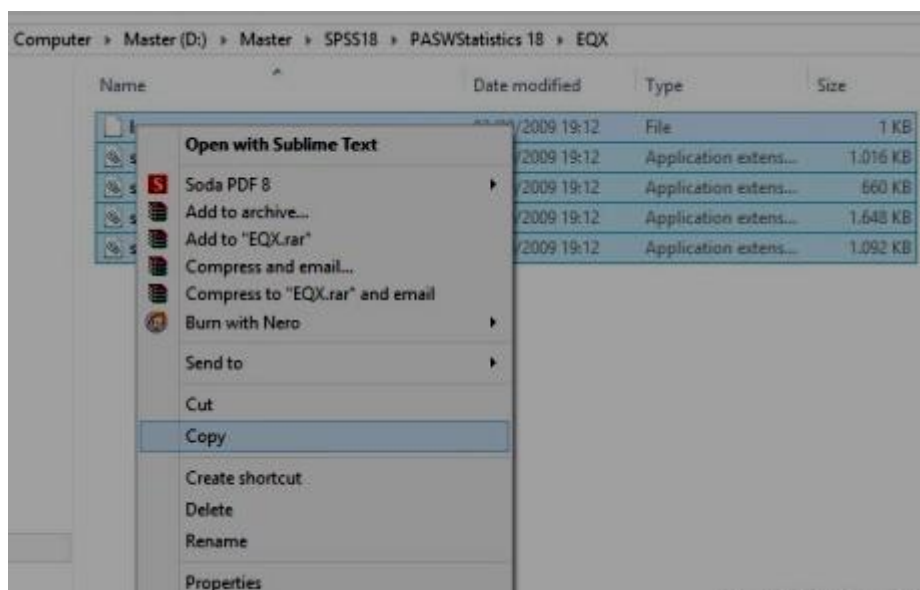
Name	Date modified	Type	Size
EQX	24/01/2018 9:50	File folder	
Setup	24/01/2018 9:50	File folder	
equinox.nfo	03/09/2009 19:12	System Informatio...	7 KB
Readme.txt	26/07/2009 5:14	Text Document	1 KB
rl-team.net.nfo	03/09/2009 19:12	System Informatio...	2 KB
simplify.nfo	03/09/2009 19:12	System Informatio...	3 KB

14. Setelah anda buka folder "EQX" di dalamnya ada lima file, copykan file tersebut ke drive instalasi SPSS anda, pada komputer saya folder instalasi terletak pada: **C:\Program Files (x86)\SPSSInc\PASWStatistics18.**

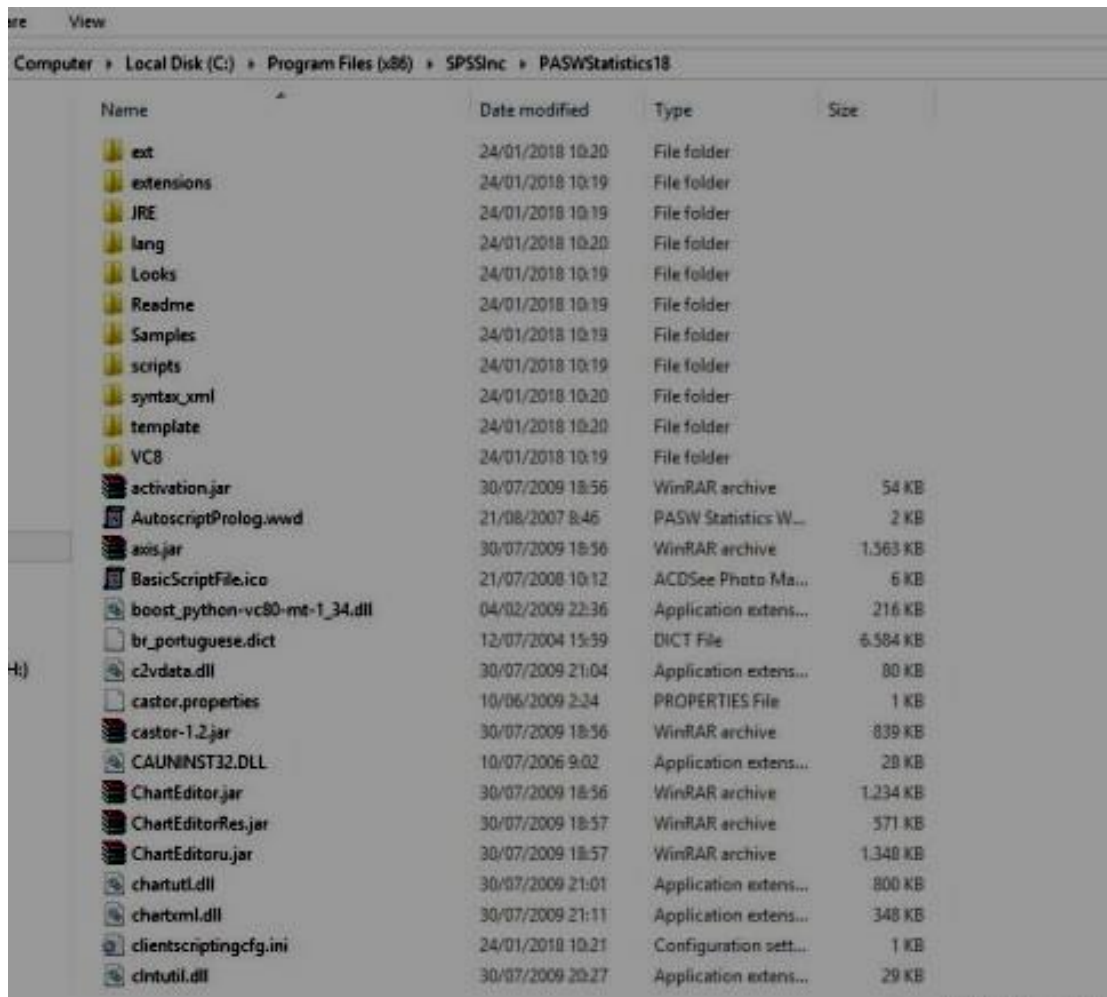


Name	Date modified	Type	Size
lservrc	03/09/2009 19:12	File	1 KB
spssstat.dll	03/09/2009 19:12	Application extens...	1.016 KB
spsswctl.dll	03/09/2009 19:12	Application extens...	660 KB
spsswlib.dll	03/09/2009 19:12	Application extens...	1.648 KB
spsswpkg.dll	03/09/2009 19:12	Application extens...	1.092 KB

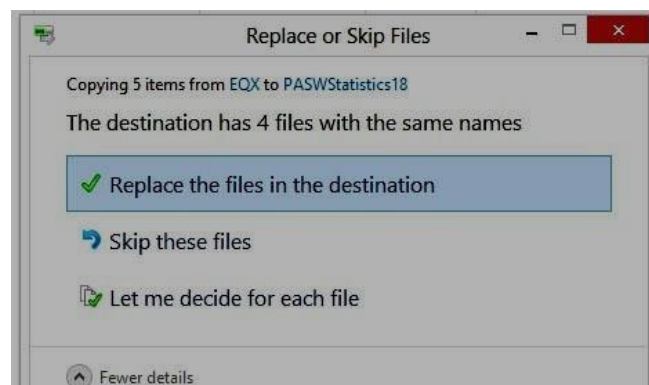
15. Copykan kelima file di dalam folder EQX tadi ke folder instalasi SPSS anda.

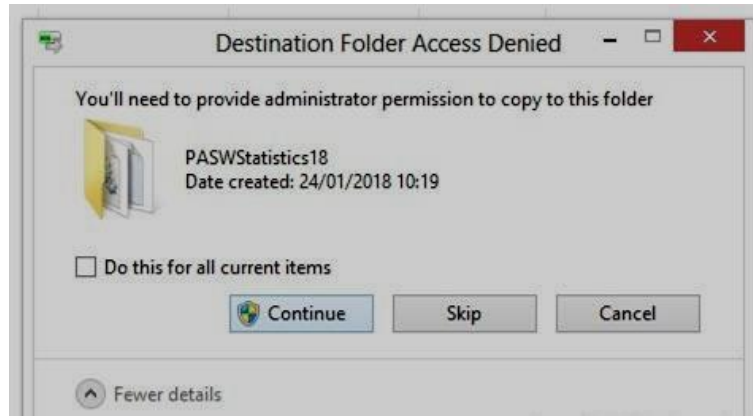


16. Folder instalasi default di komputer saya seperti gambar di bawah ini. silahkan pastekan kelima file di folder EQX tadi ke folder Instalasi program SPSS anda.

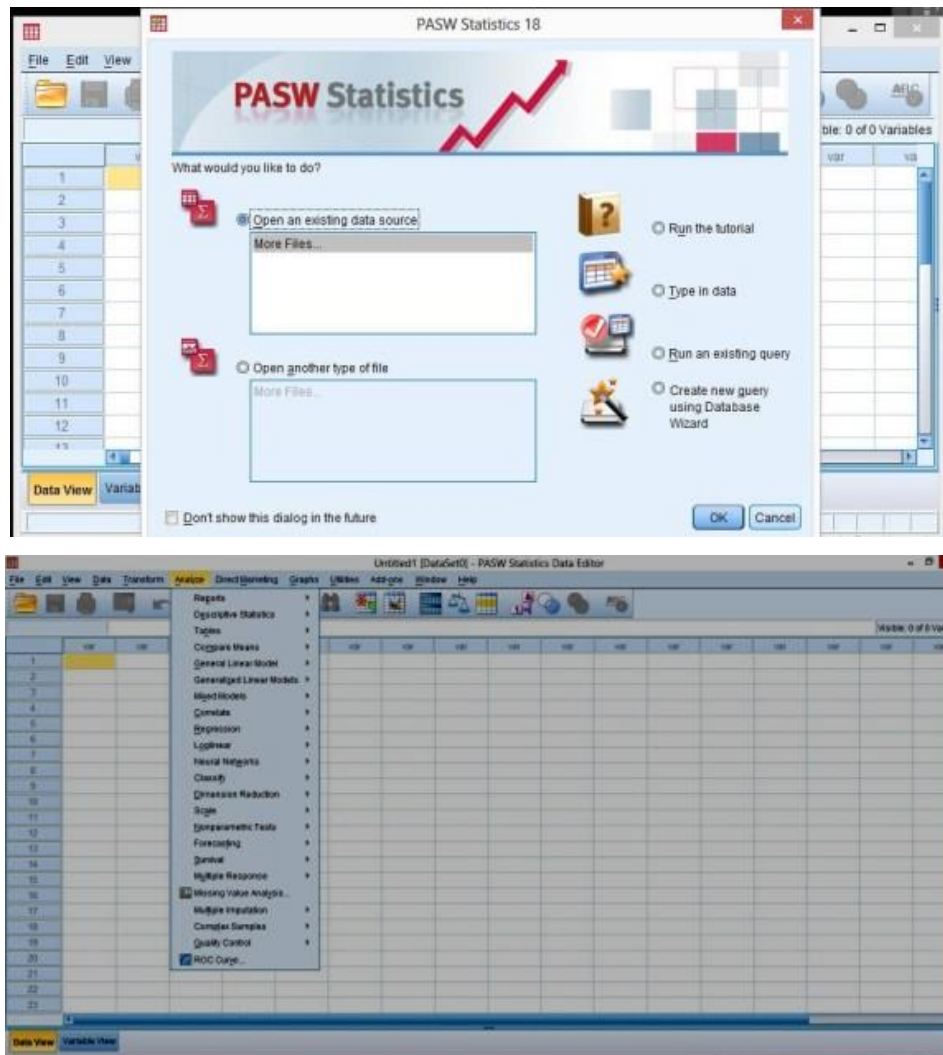


17. Jika ada pemberitahuan Replace silahkan di replace dan centang "Do this for all current items" dan pilih "Continue".





18. Jika proses mengcopy file dari folder EQX tersebut selesai maka selesai jika proses instalasi dan proses aktivasi program SPSS. Sekarang program SPSS versi 18 anda sudah siap di gunakan.





# BAB 5 STATISTIK DISKRIPITIF

---

## A. PENGERTIAN STATISTIK DESKRIPITIF

Statistik deskriptif adalah metode dari pengorganisasian, penjumlahan, dan penyajian data dalam sebuah cara yang nyaman dan informatif, termasuk teknik grafik, dan teknik penghitungan. Pada sesi tulisan ini teknik grafik tidak akan dibahas, penyajian teknik ini akan dijelaskan pada sesi tulisan yang lain. Statistik deskriptif dapat mendeskripsikan data yang sedang dianalisis, tetapi tidak boleh menarik kesimpulan apapun dari data.

Statistik deskriptif yaitu merupakan metode – metode yang juga berkaitan dengan pengumpulan dan Penyajian suatu gugus data sehingga dapat memberikan informasi yang berguna.

Statistik Deskriptif juga merupakan metode yang sangat sederhana. Metode ini hanya mendeskripsikan kondisi dari data yang sudah anda miliki Dan menyajikannya dalam bentuk tabel diagram grafik dan bentuk lainnya yang disajikan dalam uraian – uraian singkat dan juga terbatas.

## B. UKURAN PEMUSATAN DATA

Salah satu aspek yang paling penting untuk menggambarkan distribusi data adalah nilai pusat data pengamatan (*Central Tendency*). Setiap pengukuran aritmatika yang ditujukan untuk menggambarkan suatu nilai yang mewakili nilai pusat atau nilai sentral dari suatu gugus data (himpunan pengamatan) dikenal sebagai ukuran pemusatan data (*tendensi sentral*). Terdapat tiga ukuran pemusatan data yang sering digunakan, yaitu:

### 1. Mean (*arithmetic mean*)

Rata-rata hitung atau arithmetic mean atau sering disebut dengan istilah mean saja merupakan metode yang paling banyak digunakan untuk menggambarkan ukuran tendensi sentral. Mean dihitung dengan menjumlahkan semua nilai data pengamatan kemudian dibagi dengan banyaknya data. Definisi tersebut dapat di nyatakan dengan persamaan berikut:

**Sampel:**

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n} \text{ atau } \bar{x} = \frac{\sum_{i=1}^n x_i}{n} \text{ atau } \bar{x} = \frac{\sum x}{n}$$

**Populasi:**

$$\mu = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n} \text{ atau } \mu = \frac{\sum_{i=1}^n x_i}{n} \text{ atau } \mu = \frac{\sum x}{n}$$

Keterangan:

$\Sigma$  = lambang penjumlahan semua gugus data pengamatan

n = banyaknya sampel data

N = banyaknya data populasi

$\bar{x}$  = nilai rata-rata sampel

$\mu$  = nilai rata-rata populasi

Mean dilambangkan dengan  $\bar{x}$  (dibaca "x-bar") jika kumpulan data ini merupakan *contoh* (sampel) dari populasi, sedangkan jika semua data berasal dari populasi, mean dilambangkan dengan  $\mu$  (huruf kecil Yunani *mu*).

#### a) Rata-rata hitung (*Mean*) untuk data tunggal

- Contoh 1:

Hitunglah nilai rata-rata dari nilai ujian matematika kelas 3 SMU berikut ini:

2; 4; 5; 6; 6; 7; 7; 7; 8; 9

Jawab:

$$\bar{x} = \frac{\sum x}{n} = \frac{2 + 4 + 5 + 6 + 6 + 7 + 7 + 7 + 8 + 9}{10} = \frac{61}{10} = 6.10$$

- Contoh 2:

Berapa rata-rata hitung pada tabel frekuensi berikut:

$x_i$	$f_i$
70	5
69	6
45	3
80	1
56	1

Catatan: Tabel frekuensi pada tabel di atas merupakan tabel frekuensi untuk data tunggal, bukan tabel frekuensi dari data yang sudah dikelompokkan berdasarkan selang/kelas tertentu.

Jawab:

$x_i$	$f_i$	$f_i x_i$
70	5	350
69	6	414
45	3	135
80	1	80
56	1	56
Jumlah	16	1035

$$\bar{x} = \frac{\sum f_i x_i}{\sum f_i} = \frac{1035}{16} = 64.6$$

**b) Rata-rata hitung (*Mean*) dari data distribusi Frekuensi atau dari gabungan**

*Distribusi Frekuensi:* Rata-rata hitung dari data yang sudah disusun dalam bentuk tabel distribusi frekuensi dapat ditentukan dengan menggunakan formula yang sama dengan formula untuk menghitung nilai rata-rata dari data yang sudah dikelompokkan, yaitu:

$$\bar{x} = \frac{\sum f_i x_i}{\sum f_i}$$

Keterangan:

$\sum$  = lambang penjumlahan semua gugus data pengamatan  $f_i$  = frekuensi data ke-i

$\bar{x}$  = nilai rata-rata sampel

- Contoh :

Tabel berikut ini adalah nilai ujian statistik 80 mahasiswa yang sudah disusun dalam tabel frekuensi. Berbeda dengan contoh 2, pada contoh ke-3 ini, tabel distribusi frekuensi dibuat dari data yang sudah dikelompokkan berdasarkan selang/kelas tertentu (banyak kelas = 7 dan panjang kelas = 10).

Kelas ke-	Nilai Ujian	$f_i$
1	31 - 40	2
2	41 - 50	3
3	51 - 60	5
4	61 - 70	13
5	71 - 80	24
6	81 - 90	21
7	91 - 100	12
	Jumlah	80

Jawab:

Buat daftar tabel berikut, tentukan nilai wakilnya ( $x_i$ ) dan hitung  $f_i x_i$ .

Kelas ke-	Nilai Ujian	$f_i$	$x_i$	$f_i x_i$
1	31 - 40	2	35.5	71.0
2	41 - 50	3	45.5	136.5
3	51 - 60	5	55.5	277.5
4	61 - 70	13	65.5	851.5
5	71 - 80	24	75.5	1812.0
6	81 - 90	21	85.5	1795.5
7	91 - 100	12	95.5	1146.0
	Jumlah	80		6090.0

$$\bar{x} = \frac{\sum f_i x_i}{\sum f_i} = \frac{6090}{80} = 76.1$$

Catatan:

Pendekatan perhitungan nilai rata-rata hitung dengan menggunakan distribusi frekuensi kurang akurat dibandingkan dengan cara perhitungan rata-rata hitung dengan menggunakan data aktualnya. Pendekatan ini seharusnya hanya digunakan apabila tidak memungkinkan untuk menghitung nilai rata-rata hitung dari sumber data aslinya.

### c) Rata-rata Gabungan atau rata-rata terboboti (*Weighted Mean*)

Rata-rata gabungan (disebut juga *grand mean*, *pooled mean*, atau *rata-rata umum*) adalah cara yang tepat untuk menggabungkan rata-rata hitung dari beberapa sampel.

$$\bar{x} = \frac{\sum n_i x_i}{\sum n_i} = \bar{x} = \frac{\sum f_i x_i}{\sum f_i}$$

- Contoh :

Tiga sub sampel masing-masing berukuran 10, 6, 8 dan rata-ratanya 145, 118, dan 162. Berapa rata-ratanya?

Jawab:

$$\bar{x} = \frac{\sum n_i x_i}{\sum n_i} = \frac{(10)(145) + (6)(118) + (8)(162)}{10 + 6 + 8} = 143.9$$

## 2. Median

Median dari  $n$  pengukuran atau pengamatan  $x_1, x_2, \dots, x_n$  adalah nilai pengamatan yang terletak di tengah gugus data setelah data tersebut diurutkan. Apabila banyaknya pengamatan ( $n$ ) ganjil, median terletak tepat ditengah gugus data, sedangkan bila  $n$  genap, median diperoleh dengan cara interpolasi yaitu rata-rata dari dua data yang berada di tengah gugus data. Dengan demikian, median membagi himpunan pengamatan menjadi dua bagian yang sama besar, 50% dari pengamatan terletak di bawah median dan 50% lagi terletak di atas median. Median sering dilambangkan dengan

- $\bar{x}$  (dibaca "x-tilde") apabila sumber datanya berasal dari sampel
- $\mu$  (dibaca "mu-tilde") untuk median populasi.

Median tidak dipengaruhi oleh nilai-nilai aktual dari pengamatan melainkan pada posisi mereka. Prosedur untuk menentukan nilai median, pertama urutkan data terlebih dahulu, kemudian ikuti salah satu prosedur berikut ini:

- Banyak data ganjil  $\rightarrow$  mediannya adalah nilai yang berada tepat di tengah gugus data
- Banyak data genap  $\rightarrow$  mediannya adalah rata-rata dari dua nilai data yang berada di tengah gugus data

## a) Median data tunggal

Untuk menentukan median dari data tunggal, terlebih dulu kita harus mengetahui letak/posisi median tersebut. Posisi median dapat ditentukan dengan menggunakan formula berikut:

$$\text{Posisi Median} = \frac{(n + 1)}{2}$$

$n$  = banyaknya data pengamatan.

### 1) Median apabila $n$ ganjil

- Contoh :

Hitunglah median dari nilai ujian matematika kelas 3 SMU berikut ini: 8; 4; 5; 6; 7; 6; 7; 7; 2; 9; 10

Jawab:

- ✓ data: 8; 4; 5; 6; 7; 6; 7; 7; 2; 9; 10
- ✓ diurutkan: 2; 4; 5; 6; 6; 7; 7; 7; 8; 9; 10
- ✓ banyaknya data ( $n$ ) = 11
- ✓ posisi  $Me = \frac{1}{2}(11+1) = 6$
- ✓ jadi Median = 7 (data yang terletak pada urutan ke-6)

Nilai Ujian	2	4	5	6	6	<b>7</b>	7	7	8	9	10
Urutan data ke-	1	2	3	4	5	<b>6</b>	7	8	9	10	11

### 2) Median apabila $n$ genap

- Contoh :

Hitunglah median dari nilai ujian matematika kelas 3 SMU berikut ini: 8; 4; 5; 6; 7; 6; 7; 7; 2; 9

Jawab:

- ✓ data: 8; 4; 5; 6; 7; 6; 7; 7; 2; 9
- ✓ setelah diurutkan: 2; 4; 5; 6; 6; 7; 7; 7; 8; 9
- ✓ banyaknya data ( $n$ ) = 10
- ✓ posisi  $Me = \frac{1}{2}(10+1) = 5.5$
- ✓ Data tengahnya: 6 dan 7
- ✓ jadi Median =  $\frac{1}{2} (6+7) = 6.5$  (rata-rata dari 2 data yang terletak pada urutan ke-5 dan ke-6)

Nilai Ujian	2	4	5	6	<b>6</b>	<b>7</b>	7	7	8	9
Urutan data ke-	1	2	3	4	<b>5</b>	<b>6</b>	7	8	9	10

**b) Median dalam distribusi frekuensi:**

Formulasi untuk menentukan median dari tabel distribusi frekuensi adalah sebagai berikut:

$$Me = b + p \left( \frac{\frac{1}{2}n - F}{f} \right)$$

b = batas bawah kelas median dari kelas selang yang mengandung unsur atau memuat nilai median

p = panjang kelas median

n = ukuran sampel/banyak data

f = frekuensi kelas median

F = Jumlah semua frekuensi dengan tanda kelas lebih kecil dari kelas median ( $\sum f_i$ )

- Contoh :

Tentukan nilai median dari tabel distribusi frekuensi pada

Jawab:

Kelas ke-	Nilai Ujian	$f_i$	$f_{kum}$
1	31 - 40	2	2
2	41 - 50	3	5
3	51 - 60	5	10
4	61 - 70	13	23
<b>5</b>	<b>71 - 80</b>	<b>24</b>	<b>47</b>
6	81 - 90	21	68
7	91 - 100	12	80
8	Jumlah	80	

- ✓ Letak kelas median: Setengah dari seluruh data = 40, terletak pada kelas ke-5 (nilai ujian 71-80)
- ✓  $b = 70.5$ ,
- ✓  $p = 10$
- ✓  $n = 80$ ,
- ✓  $f = 24$
- ✓  $F = 2 + 3 + 5 + 13 = 23$

$$Me = b + p \left( \frac{\frac{1}{2}n - F}{f} \right) = 70.5 + 10 \left( \frac{\frac{1}{2}(80) - 23}{24} \right) = 77.58$$

### 3. Mode

Mode adalah data yang paling sering muncul/terjadi. Untuk menentukan modus, pertama susun data dalam urutan meningkat atau sebaliknya, kemudian hitung frekuensinya. Nilai yang frekuensinya paling besar (sering muncul) adalah modus. Modus digunakan baik untuk tipe data numerik atau pun data kategoris. Modus tidak dipengaruhi oleh nilai ekstrem. Beberapa kemungkinan tentang modus suatu gugus data:

- a) Apabila pada sekumpulan data terdapat dua mode, maka gugus data tersebut dikatakan bimodal.
- b) Apabila pada sekumpulan data terdapat lebih dari dua mode, maka gugus data tersebut dikatakan multimodal.
- c) Apabila pada sekumpulan data tidak terdapat mode, maka gugus data tersebut dikatakan tidak mempunyai modus.

Meskipun suatu gugus data mungkin saja tidak memiliki modus, namun pada suatu distribusi data kontinyu, modus dapat ditentukan secara analitis.

- a) Untuk gugus data yang distribusinya simetris, nilai mean, median dan modus semuanya sama.
- b) Untuk distribusi miring ke kiri (negatively skewed):  $\text{mean} < \text{median} < \text{modus}$
- c) untuk distribusi miring ke kanan (positively skewed): terjadi hal yang sebaliknya, yaitu  $\text{mean} > \text{median} > \text{modus}$ .



Hubungan antara ketiga ukuran tendensi sentral untuk data yang tidak berdistribusi normal, namun hampir simetris dapat didekati dengan menggunakan rumus empiris berikut:

$$\text{Mean} - \text{Mode} = 3 (\text{Mean} - \text{Median})$$

#### a) Modus Data Tunggal:

- Contoh :

Berapa modus dari nilai ujian matematika kelas 3 SMU berikut ini:

- ✓ 2, 4, 5, 6, 6, 7, 7, 7, 8, 9
- ✓ 2, 4, 6, 6, 6, 7, 7, 7, 8, 9
- ✓ 2, 4, 6, 6, 6, 7, 8, 8, 8, 9
- ✓ 2, 4, 5, 5, 6, 7, 7, 8, 8, 9
- ✓ 1, 2, 3, 4, 5, 6, 7, 8, 9, 10

Jawab:

- ✓ 2, 4, 5, 6, 6, 7, 7, 7, 8, 9 → Nilai yang sering muncul adalah angka 7 (frekuensi terbanyak = 3), sehingga Modus (M) = 7
- ✓ 2, 4, 6, 6, 6, 7, 7, 7, 8, 9 → Nilai yang sering muncul adalah angka 6 dan 7 (masing-masing muncul 3 kali), sehingga Modusnya ada dua, yaitu 6 dan 7. Gugus data tersebut dikatakan bimodal karena mempunyai dua modus. Karena ke-2 mode tersebut nilainya berurutan, mode sering dihitung dengan menghitung nilai rata-rata keduanya,  $\frac{1}{2}(6+7) = 6.5$ .
- ✓ 2, 4, 6, 6, 6, 7, 8, 8, 8, 9 → Nilai yang sering muncul adalah angka 6 dan 8 (masing-masing muncul 3 kali), sehingga Modusnya ada dua, yaitu 6 dan 8. Gugus data tersebut dikatakan bimodal karena mempunyai dua modus. Nilai mode tunggal tidak dapat dihitung karena ke-2 mode tersebut tidak berurutan.
- ✓ 2, 4, 5, 5, 6, 7, 7, 8, 8, 9 → Nilai yang sering muncul adalah angka 5, 6 dan 7 (masing-masing muncul 2 kali), sehingga Modusnya ada tiga, yaitu 5, 6 dan 7. Gugus data tersebut dikatakan multimodal karena modusnya lebih dari dua.

- ✓ 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 → Pada gugus data tersebut, semua frekuensi data sama, masing-masing muncul satu kali, sehingga gugus data tersebut dikatakan tidak mempunyai modusnya

**b) Mode dalam Distribusi Frekuensi:**

$$M_o = b + p \left( \frac{b_1}{b_1 + b_2} \right)$$

dimana:

$M_o$  = modal = kelas yang memuat modus

$b$  = batas bawah kelas modal

$p$  = panjang kelas modal

$b_{mo}$  = frekuensi dari kelas yang memuat modus (yang nilainya tertinggi)

$b_1 = b_{mo} - b_{mo-1}$  = frekuensi kelas modal – frekuensi kelas sebelumnya

$b_2 = b_{mo} - b_{mo+1}$  = frekuensi kelas modal – frekuensi kelas sesudahnya

- Contoh :

Tentukan nilai median dari tabel distribusi frekuensi pada Contoh 3 di atas!

Jawab:

Kelas ke-	Nilai Ujian	$f_i$	
1	31 - 40	2	
2	41 - 50	3	
3	51 - 60	5	
4	61 - 70	13	
			→ $b_1 = (24 - 13) = 11$
5	71 - 80	24	← kelas modal (frekuensinya paling besar)
			→ $b_2 = (24 - 21) = 3$
6	81 - 90	21	
7	91 - 100	12	
8	Jumlah	80	

- ✓ Kelas modul = kelas ke-5
- ✓  $b = 71 - 0.5 = 70.5$
- ✓  $b_1 = 24 - 13 = 11$
- ✓  $b_2 = 24 - 21 = 3$

$$\checkmark \quad p = 10$$

$$Mo = b + p \left( \frac{b_1}{b_1 + b_2} \right) = 70.5 + 10 \left( \frac{11}{11 + 3} \right) = 78.36$$

Selain tiga ukuran tendensi sentral di atas (mean, median, dan mode), terdapat ukuran tendensi sentral lainnya, yaitu rata-rata ukur (*Geometric Mean*) dan rata-rata harmonis (*Harmonic Mean*)

#### 4. Rata-rata Ukur (Geometric Mean)

Untuk gugus data positif  $x_1, x_2, \dots, x_n$ , rata-rata geometrik adalah akar ke- $n$  dari hasil perkalian unsur-unsur datanya. Secara matematis dapat dinyatakan dengan formula berikut:

$$U = \sqrt[n]{x_1 \cdot x_2 \cdot x_3 \dots x_n} \text{ atau } U = \sqrt[n]{\prod_{i=1}^n x_i} \text{ atau } \log(U) = \frac{\sum \log(x_i)}{n}$$

Dimana:

$U$  = rata-rata ukur (rata-rata geometrik)

$n$  = banyaknya sampel

$\Pi$  = Huruf kapital  $\pi$  (pi)

yang menyatakan jumlah dari hasil kali unsur-unsur data. Rata-rata geometrik sering digunakan dalam bisnis dan ekonomi untuk menghitung rata-rata tingkat perubahan, rata-rata tingkat pertumbuhan, atau rasio rata-rata untuk data berurutan tetap atau hampir tetap atau untuk rata-rata kenaikan dalam bentuk persentase.

##### a) Rata-rata ukur untuk data tunggal

- Contoh :

Berapakah rata-rata ukur dari data 2, 4, 8?

Jawab:

$$U = \sqrt[n]{x_1 \cdot x_2 \cdot x_3 \dots x_n} = \sqrt[3]{64} = 4$$

atau

$$\text{Log } (U) = \frac{\sum \log(x_i)}{n}$$

$$\text{Log } (U) = \frac{\log(2)+\log(4)+\log(8)}{3} = \frac{0.3010+0.6021+0.9031}{3} = 0.6021$$

$$U = 10^{0.6021} = 4$$

**b) Distribusi Frekuensi:**

$$\text{Log } (U) = \frac{\sum(f_i \cdot \log(x_i))}{\sum f_i}$$

xi = tanda kelas (nilai tengah)

fi = frekuensi yang sesuai dengan xi

- Contoh :

Tentukan rata-rata ukur dari tabel distribusi frekuensi pada Contoh 3 di atas!

Jawab

Kelas ke-	Nilai Ujian	fi	xi	log xi	fi.log xi
1	31 - 40	2	35.5	1.5502	3.1005
2	41 - 50	3	45.5	1.6580	4.9740
3	51 - 60	5	55.5	1.7443	8.7215
4	61 - 70	13	65.5	1.8162	23.6111
5	71 - 80	24	75.5	1.8779	45.0707
6	81 - 90	21	85.5	1.9320	40.5713
7	91 - 100	12	95.5	1.9800	23.7600
8	Jumlah	80			149.8091

$$\text{Log } (U) = \frac{\sum(f_i \cdot \log(x_i))}{\sum f_i} = \frac{149.8091}{80} = 1.8726$$

$$U = 10^{1.8726} = 74.5789$$

**c) Rata-rata Harmonik (H)**

Rata-rata harmonik dari suatu kumpulan data  $x_1, x_2, \dots, x_n$  adalah kebalikan dari nilai rata-rata hitung (aritmetik mean). Secara matematis dapat dinyatakan dengan formula berikut:

$$H = \frac{n}{\sum \left( \frac{1}{x_i} \right)}$$

Secara umum, rata-rata harmonic jarang digunakan. Rata-rata ini hanya digunakan untuk data yang bersifat khusus. Misalnya, rata-rata harmonic sering digunakan sebagai ukuran tendensi sentral untuk kumpulan data yang menunjukkan adanya laju perubahan, seperti kecepatan.

#### d) Rata-rata harmonic untuk data tunggal

- Contoh :

Si A bepergian pulang pergi. Waktu pergi ia mengendarai kendaraan dengan kecepatan 10 km/jam, sedangkan waktu kembalinya 20 km/jam. Berapakah rata-rata kecepatan pulang pergi?

Jawab:

Apabila kita menghitungnya dengan menggunakan rumus jarak dan kecepatan, tentu hasilnya 13.5 km/jam! Apabila kita gunakan perhitungan rata-rata hitung, hasilnya tidak tepat!

$$\bar{x} = \frac{(10 + 20)}{2} = 15 \text{ km/jam}$$

Pada kasus ini, lebih tepat menggunakan rata-rata harmonic:

$$\bar{x} = \frac{2}{\frac{1}{10} + \frac{1}{20}} = \frac{40}{3} = 13.5 \text{ km/jam}$$

#### e) Rata-rata Harmonik untuk Distribusi Frekuensi:

$$H = \frac{\sum f_i}{\sum \left( \frac{f_i}{x_i} \right)}$$

- Contoh :

Berapa rata-rata Harmonik dari tabel distribusi frekuensi pada Contoh 3 di atas!

Jawab:

Kelas ke-	Nilai Ujian	f <sub>i</sub>	x <sub>i</sub>	f <sub>i</sub> /x <sub>i</sub>
1	31 - 40	2	35.5	0.0563
2	41 - 50	3	45.5	0.0659

3	51 - 60	5	55.5	0.0901
4	61 - 70	13	65.5	0.1985
5	71 - 80	24	75.5	0.3179
6	81 - 90	21	85.5	0.2456
7	91 - 100	12	95.5	0.1257
8	Jumlah	80		1.1000

$$H = \frac{\sum f_i}{\sum \left(\frac{f_i}{x_i}\right)} = \frac{80}{1.1000} = 72.7283$$

- **Perbandingan ketiga rata-rata (Mean)**

$$\bar{x} = 76.10; U = 74.58; H = 72.73$$

$$H \leq U \leq \bar{x} = 76.10$$

## 6. Karakteristik penting untuk ukuran tendensi sentral yang baik

Ukuran nilai pusat/tendensi sentral (*average*) merupakan nilai perwakilan dari suatu distribusi data, sehingga harus memiliki sifat-sifat berikut:

- Harus mempertimbangkan semua gugus data
- Tidak boleh terpengaruh oleh nilai-nilai ekstrim.
- Harus stabil dari sampel ke sampel.
- Harus mampu digunakan untuk analisis statistik lebih lanjut.

Dari beberapa ukuran nilai pusat, Mean hampir memenuhi semua persyaratan tersebut, kecuali syarat pada point kedua, rata-rata dipengaruhi oleh nilai ekstrem. Sebagai contoh, jika item adalah 2; 4; 5; 6; 6; 6; 7; 7; 8; 9 maka mean, median dan modus semua bernilai sama, yaitu 6. Jika nilai terakhir adalah 90 bukan 9, rata-rata akan menjadi 14.10, sedangkan median dan modus tidak berubah. Meskipun dalam hal ini median dan modus lebih baik, namun tidak memenuhi persyaratan lainnya. Oleh karena itu Mean merupakan ukuran nilai pusat yang terbaik dan sering digunakan dalam analisis statistik.

Nilai ukuran pusat yang tepat untuk digunakan tergantung pada sifat data, sifat distribusi frekuensi dan tujuan. Jika data bersifat kualitatif, hanya modus yang dapat digunakan. Sebagai contoh, apabila kita tertarik untuk mengetahui jenis tanah yang khas di suatu lokasi, atau pola tanam di suatu daerah, kita hanya dapat menggunakan modus. Di sisi lain, jika data bersifat kuantitatif, kita dapat

menggunakan salah satu dari ukuran nilai pusat tersebut, mean atau median atau modus. Meskipun pada jenis data kuantitatif kita dapat menggunakan ketiga ukuran tendensi sentral, namun kita harus mempertimbangkan sifat distribusi frekuensi dari gugus data tersebut.

- a) Bila distribusi frekuensi data tidak normal (tidak simetris), median atau modus merupakan ukuran pusat yang tepat.
- b) Apabila terdapat nilai-nilai ekstrim, baik kecil atau besar, lebih tepat menggunakan median atau modus.
- c) Apabila distribusi data normal (simetris), semua ukuran nilai pusat, baik mean, median, atau modus dapat digunakan. Namun, mean lebih sering digunakan dibanding yang lainnya karena lebih memenuhi persyaratan untuk ukuran pusat yang baik.
- d) Ketika kita berhadapan dengan laju, kecepatan dan harga lebih tepat menggunakan rata-rata harmonik.
- e) Jika kita tertarik pada perubahan relatif, seperti dalam kasus pertumbuhan bakteri, pembelahan sel dan sebagainya, rata-rata geometrik adalah rata-rata yang paling tepat.

### C. UKURAN PENYEBARAN DATA

Ukuran penyebaran data merupakan ukuran yang menunjukkan seberapa jauh data suatu menyebar dari rata-ratanya. Ukuran penyebaran data yang ada dalam materi statistika meliputi jangkauan, hamparan, dan kuartil.

Ukuran Penyebaran data (Measures of Dispersion = Measures of Variability) adalah berbagai macam ukuran statistic yang dapat digunakan untuk mengetahui luas penyebaran data (Variasi data = Homogenitas Data = Stabilitas Data).

Penyebaran atau dispersi adalah pergerakan dari nilai observasi terhadap nilai rata-ratanya. Rata-rata dari serangkaian nilai observasi tidak dapat diinterpretasikan secara terpisah dari hasil dispersi nilai-nilai tersebut sekitar rata-ratanya. Makin besar variasi nilai, makin kurang representatif rata-rata distribusinya. Adapun ukuran penyebaran data yang biasa dihitung adalah range (rentang), standar deviasi (simpangan baku), kurtosis (keruncingan), skewness (kemiringan). Rentang data menunjukkan selisih antara nilai terbesar dengan nilai terkecil dalam suatu himpunan data. Simpangan baku adalah jumlah mutlak selisih setiap nilai pengamatan terhadap

nilai rata-rata dibagi dengan banyaknya pengamatan, kurtosis merupakan ukuran untuk menentukan bentuk-bentuk distribusi yang biasanya dibandingkan dengan kurva distribusi normal.

Kegunaan ukuran penyebaran antara lain sebagai berikut:

1. Ukuran penyebaran dapat digunakan untuk menentukan apakah nilai rata-ratanya benar-benar representatif atau tidak. Apabila suatu kelompok data mempunyai penyebaran yang tidak sama terhadap nilai rata-ratanya, maka dikatakan bahwa nilai rata-rata tersebut tidak representatif.
2. Ukuran penyebaran dapat digunakan untuk mengadakan perbandingan terhadap variabilitas data.
3. Ukuran penyebaran dapat membantu penggunaan ukuran statistika, misalnya dalam pengujian hipotesis, apakah dua sampel berasal dari populasi yang sama atau tidak.

Dalam dunia statistik, dikenal beberapa macam Ukuran Penyebaran Data, dari ukuran yang paling sederhana (kasar) sampai dengan ukuran yang dipandang memiliki kadar ketelitian yang tinggi, yaitu: (1) Range, (2) Deviasi ( simpangan rata-rata).

## 1. Range

Ukuran Penyebaran Data yang pertama-tama diperkenalkan disini adalah Range, yang didalam dunia statistik dikenal sebagai ukuran penyebaran data yang paling sederhana, yang karena itu juga sering disebut sebagai ukuran penyebaran data yang paling kasar.

### a) Pengertian Range

Range yang biasa diberi lambang “R” adalah salah satu ukuran statistik yang menunjukkan jarak penyebaran antara skor (nilai) yang terendah (lowest score) samapi skor (nilai) yang tertinggi (highest score). Dengan singkat dapat dirumuskan:

$$R = H - L$$

Ket :

R = Range yang kita cari

H = Skor atau nilai yang tertinggi (highest score)



$L = \text{Skor atau nilai yang terendah (lowest score)}$

### b) Cara Mencari Range

- Contoh

Perhitungan range nilai hasil tes untuk lima macam bidang study yang diikuti oleh tiga orang calon yang mengikuti tes seleksi panerimaan calon mahasiswa baru pada sebuah perguruan tinggi agama islam

NO	NAMA	Nilai yang dicapai					H	L	R=H-L	Jumlah niali	Mean
		PKN	PNJS	Bhs. indo	Bhs. arab	Bhs. ingg					
1.	Ega	85	65	76	45	65	85	45	40	325	65
2.	Nunu	58	65	72	60	70	72	58	14	325	65
3.	Yuni	65	65	65	65	65	65	65	0	325	65

Keterangan:

- Kolom 3-7 menunjukkan distribusi nilai hasil yang dicapai oleh tiga orang calon.
- Kolom 8 memuat nilai tertinggi (highest score) masing-masing calon.
- Kolom 9 memuat nilai terendah (lowest score) masing-masing calon.
- Kolom 10 menunjukkan jumlah seluruh nilai.
- Kolom 11 adalah mean (nilai rata-rata) yang dicapai oleh masing-masing calon.

Dari tabel tersebut menunjukkan bahwamakin kecil jarak penyebaran nilai dari nilai terendah sampai nilai tertinggi, akan semakin homogen (concentrated) distribusi nilai tersebut. Sebaliknya makin besar range-nya, akan makin berfariasi nilai-nilai yang dada dalam distribusi nilai tersebut.

Selain itu berdasar pada range kita juga dapat mengatakan bahwa kian kecil range dari suatu distribusi data, kian cenderung bagi diri kita untuk menganggap bahwa mean yang kita peroleh merupakan wakil yang persentatif data yang bersangkutan; sebaliknya kian besar range-nya, kita akan lebih cenderung untuk menganggap bahwa mean yang kita peroleh itu sifatnya meragukan.

### c) Range Untuk Data Tidak Berkelompok

Rumus untuk data tidak berkelompok adalah sebagai berikut :

**Jarak (range) = Nilai Terbesar – Nilai Terkecil**

- Contoh :

Data nilai UAS Statistika

Kelas A : 90 80 70 90 70 100 80 50 75 70

Kelas B : 80 80 75 95 75 70 95 60 85 60

- ✓ Urutkan dahulu kemudian dihitung berapa rentangnya.

Kelas A : 50 70 70 70 75 80 80 90 90 100

Kelas B : 60 60 70 75 75 80 80 85 95

- ✓ Rentangan kelas A :  $100 - 50 = 50$

Rentangan kelas B :  $95 - 60 = 35$

### d) Range Untuk Data Berkelompok

Jangkauan data berkelompok merupakan selisih antara nilai tengah kelas terakhir dengan nilai tengah kelas pertama. Perhatikan tabel berikut ini!

Umur	Titik Tenga	Frekuensi
30-34	32	5
35-39	37	35
40-44	42	100
45-49	47	50
50-54	52	10

Tabel menunjukkan data umur peserta yang mengikuti diklat sertifikasi guru yang berjumlah 200 orang. Bila nilai tengah kelas pertama adalah 32 dan nilai tengah kelas terakhir adalah 52, maka,

$$R = 52 - 32 = 20$$

Jadi, jangkauan data dari tabel 2 adalah 20. Dapat disimpulkan bahwa untuk menentukan jangkauan data berkelompok digunakan persamaan:

$$R = x_{maks} - x_{min}$$

Keterangan :

$$R = \text{jangkauan/range/rentang}$$

$X_{\text{maks}}$  = nilai tengah kelas terakhir

$X_{\text{min}}$  = nilai tengah kelas pertama

### e) Penggunaan Range

Range kita gunakan sebagai ukuran, apabila di dalam waktu yang sangat singkat kita ingin memperoleh gambaran tentang penyebaran data yang sedang kita selidiki dengan mengabaikan faktor ketelitian atau kecermatan.

### f) Kebaikan dan Kelemahan Range

Kebaikan Range sebagai salah satu ukuran penyebaran data ialah dengan menggunakan Range dalam waktu singkat dapat diperoleh gambaran umum mengenai luas penyebaran data yang sedang kita hadapi.

Adapun kelemahannya ialah:

1) Range akan sangat tergantung kepada nilai-nilai ekstrimnya. Dengan kata lain, besar-kecilnya Range akan sangat ditentukan oleh Nilai Terendah dan Nilai Tertinggi yang terdapat dalam distribusi data, dengan demikian Range sifatnya sangat labil dan kurang teliti.

• Contoh:

✓ Data X :  $H = 80, L = 30 - R = 80 - 30 = 50$ .

✓ Data Y :  $H = 95, L = 30 - R = 95 - 45 = 50$ .

✓ Data Z :  $H = 88, L = 38 - R = 88 - 38 = 50$ .

2) Range sebagai ukuran penyebaran data, tidak memperhatikan distribusi yang terdapat di dalam range itu sendiri. Ambillah sebagai contoh, misalnya Nilai Tertinggi dan Nilai Terendah yang berhasil dicapai oleh 8 orang mahasiswa masing-masing adalah 80 dan 40, sehingga Rangnya =  $80 - 40 = 40$ . Dengan Range sebesar 40 itu ada kemungkinan distribusi nilai itu adalah: 40, 47, 52, 59, 64, 67, 70 dan 80, mungkin juga: 40, 40, 40, 40, 40, 40, 40 dan 80; mungkin juga: 40, 40, 50, 50, 60, 60, 80, 80, atau bentuk distribusi lainnya. Yang jelas, dengan hanya mengetahui Rangnya saja, kita belum tahu secara pasti bagaimana sebenarnya bentuk distribusi data yang kita hadapi mulai dari Nilai Terendah sampai Nilai Tertinggi.

Karena kelemahan itulah maka sebagai salah-satu ukuran penyebaran data, Range sangat jarang digunakan dalam pekerjaan analisis statistik.

## 2. Deviasi

### a) Pengertian Deviasi

Dalam statistik, yang dimaksud dengan deviasi ialah selisih satu simpangan dari masing-masing skor atau interval, dari nilai rata-rata hitungannya (deviation from the mean).

Deviasi merupakan salah satu ukuran variabilitas data yang biasa dilambangkan dengan huruf kecil dari huruf yang digunakan bagi lambang skornya. Jika apabila skornya dibagi lambang X maka deviasinya berlambang x; jika skornya Y maka deviasinya y; jika skornya Z maka lambang deviasinya z.

Karena deviasi merupakan simpangan atau selisih dari masing-masing skor terhadap Mean groupnya, maka sudah tentu akan terdapat dua jenis deviasi, yaitu: (1) deviasi yang berada diatas Mean, dan (2) deviasi yang berada dibawah Mean.

Deviasi yang berada diatas Mean dapat diartikan sebagai “selisih lebih”; karenanya deviasi semacam ini akan bertanda (+), dan lazim dikenal dengan istilah deviasi positif. Adapun deviasi yang berada dibawah Mean dapat diartikan sebagai “selisih kurang” oleh karena itu, selalu bertanda minus (-), dan lazim dikenal dengan istilah deviasi negatif.

Perlu diingatkan bahwa semua deviasi – baik yang bertanda plus maupun yang bertanda minus – apabila kita jumlahkan hasilnya pasti sama dengan nol (0).

Guna memperjelas uraian yang telah dikemukakan diatas, marilah kita perhatikan contoh berikut ini.

Skor (X)	Banyaknya (f)	Deviasi ( $x=X-M_x$ )
8	1	$8-6 = +2$
7	1	$7-6 = +1$
6	1	$6-6 = 0$
5	1	$5-6 = -1$
4	1	$4-6 = -2$

$30 =$ $\sum X$	$5 = N$	$0 =$ $\sum x$
--------------------	---------	-------------------

### b) Deviasi Rata-rata

Seperti terlihat pada tabel diatas, jika seluruh deviasi kita jumlahkan, hasilnya pasti sama dengan nol ( $\sum x=0$ ). Karena jumlah deviasi akan selalu sama dengan nol, maka kalau deviasi itu kita gunakan sebagai ukuran untuk mengetahui variabilitas data tidak akan ada gunanya sama sekali. Oleh karena itulah agar deviasi dapat digunakan sebagai ukuran variabilitas, dalam menjumlahkan deviasi itu tanda-tanda aljabar (yaitu tanda + dan -) yang terdapat didepan deviasi sebaiknya diabaikan. Dengan kata lain, agar deviasi dapat dimanfaatkan sebagai ukuran variabilitas, maka penjumlahan itu dilakukan terhadap harga mutlakny. Setelah seluruh harga mutlak deviasi dijumlahkan, lalu dihitung rata-ratanya.

Dari pembicaraan diatas sebenarnya sudah cukup tergambar apa yang sebenarnya dimaksud dengan deviasi rata-rata itu, yakni: jumlah harga mutlak deviasi dari tiap-tiap skor, doibagi dengan banyaknya skor itu sendiri.

Dalam bahasa Inggris Deviasi Rata-rata dikenal dengan nama Mean Deviation (diberi lambang: MD) atau Average Deviation (diberi lambang: AD); dalam uraian selanjutnya akan digunakan lambang AD. Dengan demikian, apabila pengertian tentang Deviasi Rata-rata tadi kita formasikan dalam bentuk rumus adalah sebagai berikut:

$$\sum x = \text{jumlah harga mutlak deviasi tiap-tiap skor atau interval.}$$

$$N = \text{Number of Cases}$$

### c) Cara Mencari Deviasi Rata-rata

1) Cara Mencari Deviasi Rata-rata untuk Data Tunggal.

**Soal :**

Diberikan data sebagai berikut: 5, 6, 8, 5, 7 Tentukan nilai simpangan rata-rata data di atas!

**Pembahasan**

Menentukan simpangan rata-rata data tunggal, lebih dulu dicari rata-rata

datanya: Setelah diketahui rata-ratanya, saatnya mencari simpangan rata-rata: Sehingga nilainya

- 2) Cara mencari deviasi untuk data berkelompok

**Soal :**

Perhatikan tabel distribusi frekuensi data berikut ini

Nilai	Frekuensi
11 - 15	2
16 - 20	2
21 - 25	10
26 - 30	9
31 - 35	4

Tentukan nilai simpangan rata-rata data di atas!

**Pembahasan**

Menentukan simpangan rata-rata data berkelompok, tentukan dulu **titik tengah** setiap kelas, untuk kemudian dicari reratanya:

Nilai	Frekuensi	X
11 - 15	2	<b>13</b>
16 - 20	2	<b>18</b>
21 - 25	10	<b>23</b>
26 - 30	9	<b>28</b>
31 - 35	4	<b>33</b>

Rata-ratanya adalah: Dengan rumus yang sama soal sebelumnya saja, tapi dipake **titik tengah** kelas sebagai x diperoleh:

**D. PROSEDUR KERJA ANALISIS DISKRIPSTIF DAN INTERPRETASI OUTPUT ANALISIS DISKRIPSTIF MELALUI SPSS**

Dalam sebuah penelitian deskripsi data itu penting untuk memberikan gambaran data yang akan diteliti. Disamping itu, dengan adanya deskripsi data, peneliti akan lebih muda untuk mengetahui paparan data dalam sebuah penelitian secara lebih terperinci dan jelas.

Pada SPSS, analisis statistik deskriptif dilakukan dengan meng-klik menu

- Klik [Analyze] -> [Descriptive Statistics]

kemudian

terdapat

pilihan: **Frequencies, Descriptives, Explore, Crosstabs, dan Ratio.**

Dengan data sebagaimana ditunjukkan tabel di bawah ini, kita akan mempraktekkan operasi submenu semuanya.

- Tabel. *Data Nilai Mahasiswa (bukan data sebenarnya)*

Nama	Usia	Jenis Kelamin	Nilai APK	Nilai PPC	Nilai PLO
Suhairi	20	Laki-Laki	80	50	70
Ambon	21	Laki-Laki	70	70	90
Astri	22	Perempuan	60	80	70
Henri	21	Laki-Laki	80	90	60
Yugos	22	Laki-Laki	90	60	70
Muji	19	Perempuan	70	80	80
Tatang	20	Laki-Laki	60	70	40
Ferdi	21	Laki-Laki	60	90	60
Arsyad	21	Laki-Laki	70	70	40
Fauzan	21	Laki-Laki	90	80	60

\*) Laki-Laki (Value: 1), Perempuan (Value: 2)

## 1. Entry Data

Langkah pertama yang perlu dilakukan adalah meng-*entry* data, tentunya anda perlu paham dasar-dasar SPSS. *Entry* data dilakukan pada *tab sheet Data View* setiap baris mewakili satu responden, sedangkan setiap kolom mewakili satu variabel, dalam kasus ini variabelnya adalah: Nama, Usia, Jenis Kelamin, Nilai APK, Nilai PPC, dan Nilai PLO.

Berikut langkah-langkah *entry* datanya:

**a) Pada tab [Variable View]**

**1) Isi kolom Name**

- ✓ Nama untuk “Nama”
- ✓ Usia untuk “Usia”
- ✓ Gender untuk “Jenis Kelamin”
- ✓ APK untuk “Nilai APK”
- ✓ PPC untuk “Nilai PPC”
- ✓ PLO untuk “Nilai PLO”

**2) Isi kolom Label**

- ✓ Usia
- ✓ Jenis Kelamin
- ✓ Nilai APK
- ✓ Nilai PPC
- ✓ Nilai PLO

Hal ini berarti: variabel Gender mempunyai label “Jenis Kelamin”, variabel APK mempunyai label “Nilai APK”, dan seterusnya.

**3) Isi kolom Values untuk variabel Gender**

- ✓ Value: 1 = Laki-laki dan
- ✓ Value: 2 = Perempuan.

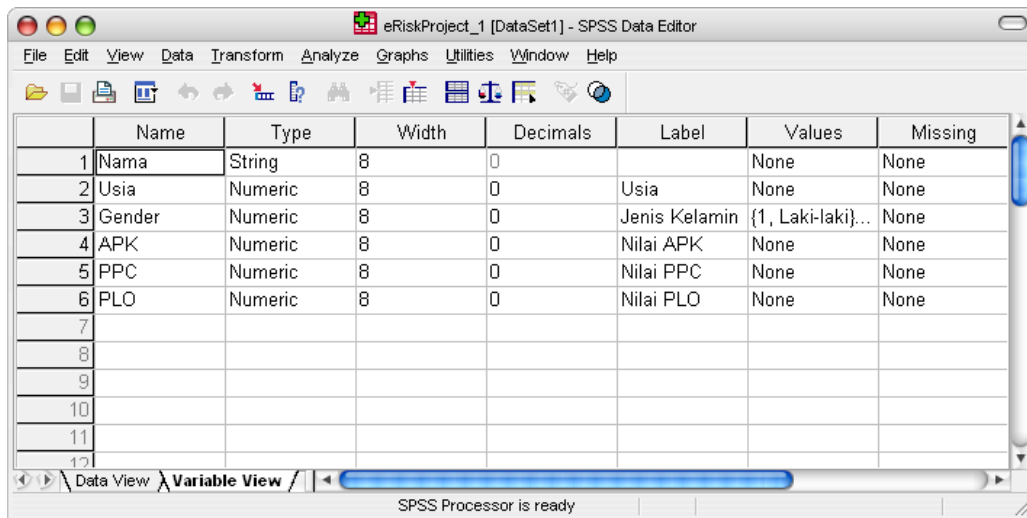
**4) Ubah kolom Type untuk variabel Nama menjadi String.**

**5) Isi kolom Decimals dengan 0 (nol) untuk semua variabel.**

**6) Biarkan *default* SPSS untuk kolom Width, Missing, dan Columns.**

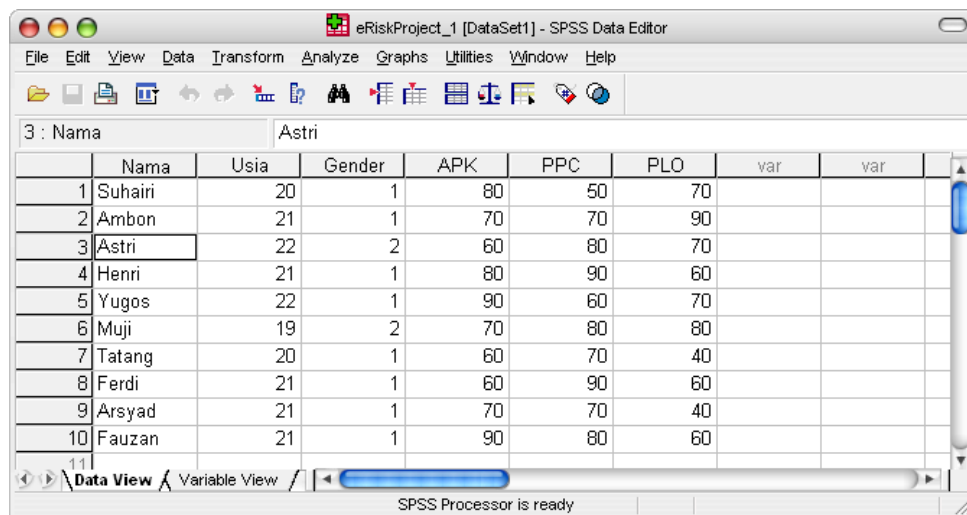
**7) Jangan lupa ”save” atau tekan Ctrl + S.**






Gambar 1. Entry Variabel pada Tab Sheet Variable View

8) Pada tab [Data View], isi data seperti Gambar 2 di bawah ini.



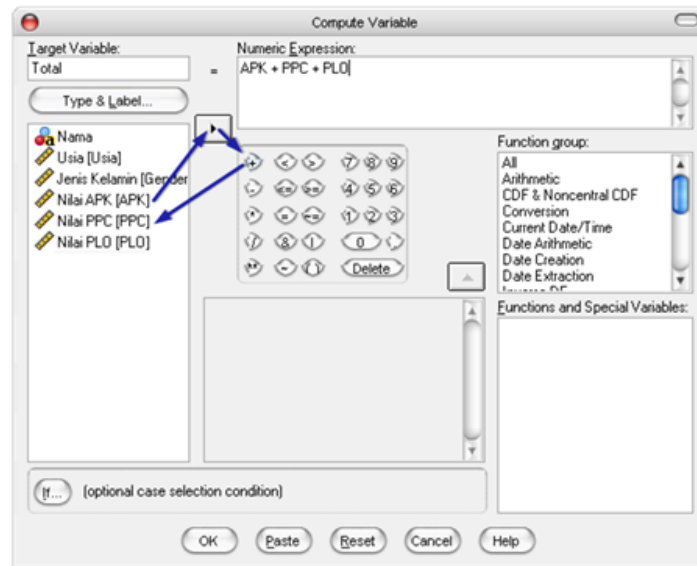
Gambar 2. Entry Data pada Tab Sheet Data View

9) Klik ikon  untuk melihat definisi Value,  
 ✓ Variabel Gender terdefinisi laki-laki dan perempuan, tidak lagi berisi angka 1 dan 2.

10) Menjumlahkan nilai APK, PPC, dan PLO:

- ✓ Klik menu [Transform] → [Compute],
- ✓ muncul dialog box **Compute Variable**.
- ✓ Tuliskan “total” pada form Target Variable.
- ✓ Klik [Type & Label], beri label “Nilai Total”.

- ✓ Ketik “APK + PPC + PLO” pada *form* Numeric Expression,



Gambar 3. Contoh Perintah Penjumlahan pada Dialog Box Compute

- ✓ KIPada *Data View* akan muncul variabel baru dengan nama “Total”

Nama	Usia	Gender	APK	PPC	PLO	Total
Suhairi	20	Laki-laki	80	50	70	200
Ambon	21	Laki-laki	70	70	90	230
Astri	22	Perempuan	60	80	70	210
Henri	21	Laki-laki	80	90	60	230
Yugos	22	Laki-laki	90	60	70	220
Muji	19	Perempuan	70	80	80	230
Tatang	20	Laki-laki	60	70	40	170
Ferdi	21	Laki-laki	60	90	60	210
Arsyad	21	Laki-laki	70	70	40	180
Fauzan	21	Laki-laki	90	80	60	230

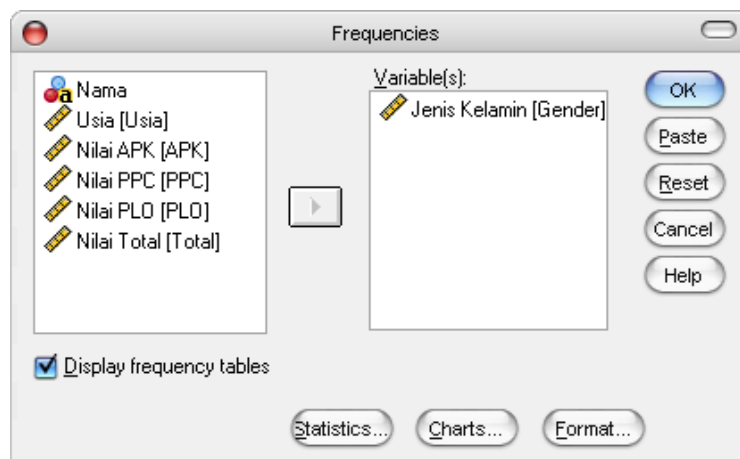
Gambar 4. Output dari Perintah Penjumlahan pada Dialog Box Compute Variable

Setelah data *di-entry*, selanjutnya memulai menggunakan perintah-perintah statistik deskriptif. Tahap pertama adalah menggunakan perintah Frequencies.

## 2. Frequencies

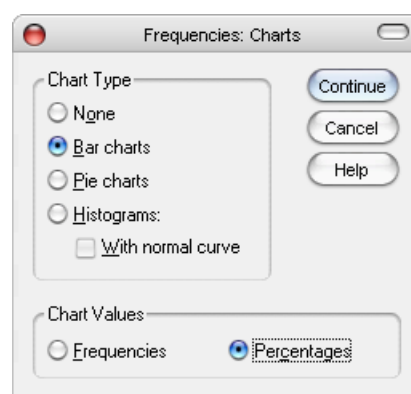
Perintah Frequencies digunakan untuk memperoleh jumlah pada nilai-nilai sebuah variabel tunggal. Langkah-langkahnya adalah sebagai berikut:

- a) Klik menu [Analyze] -> [Descriptive Statistics] -> [Frequencies], muncul *dialog box* **Frequencies**.
- b) Masukkan variabel Jenis Kelamin [Gender] ke form **Variables(s)** (untuk menganalisis variabel Jenis Kelamin).
- c) Jangan lupa centang Display frequency tables.



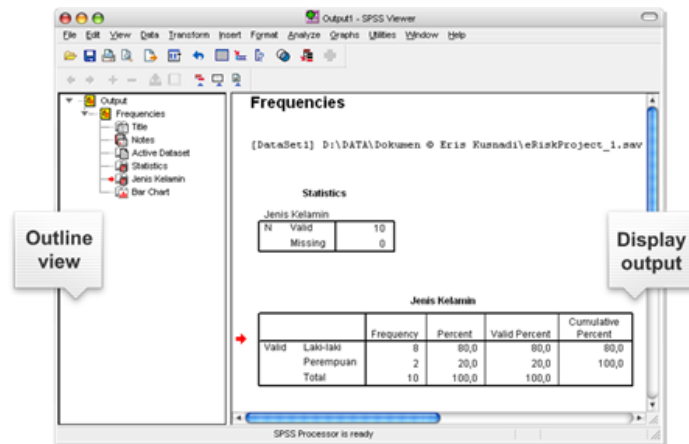
Gambar 5. Dialog Box Frequencies

- d) Agar menampilkan representasi bergambar (grafik), klik [Charts], maka akan muncul *dialog box* **Frequencies: Charts**. Saya memilih Bar charts pada *form* Chart Type. Pada form Chart Values, saya memilih Percentages.



Gambar 6. Menampilkan Bar Charts pada Dialog Box Frequencies: Charts

- e) Klik [Continues] untuk kembali ke *dialog box* **Frequencies** lalu klik [OK] maka muncul jendela *SPSS Viewer* yang menunjukkan hasil analisis frekuensi.



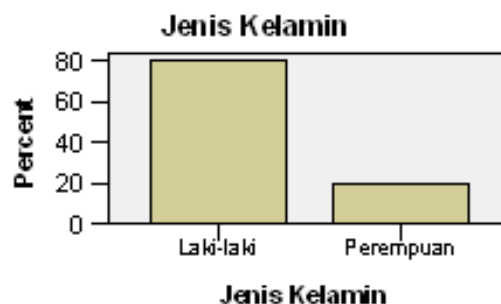
- \* ) Sebelah kiri adalah **Outline view** dalam bentuk tree files, fungsinya sebagai navigasi dalam melihat *output* analisis.
- \*\* ) Sebelah kanan adalah **Display output**, fungsinya menampilkan seluruh hasil analisis yang telah kita lakukan.

Gambar 7. Display Output pada SPSS Viewer

Pada *Gambar diatas* terlihat hasil analisis SPSS, di mana pada tabel pertama N Valid = 10 yang menunjukkan jumlah responden 10 orang dan N Missing = 0 yang berarti tidak ada data yang hilang (*missing*).

Pada tabel yang kedua terlihat hasil analisis Frequencies terhadap variabel Jenis Kelamin, di mana jumlah responden laki-laki 8 orang (80%) dan responden perempuan ada 2 orang (20%).

Jika *scroll* digeser ke bawah akan terlihat *Bar Chart* yang menunjukkan visualisasi jumlah responden laki-laki dan perempuan.



Gambar 8. Bar Chart

Perhatikan kriteria “laki-laki” dan “perempuan“, ini merupakan hasil definisi value variabel Gender pada kolom Values, di mana

- ✓ Value: 1 = Laki-laki dan
- ✓ Value: 2 = Perempuan.

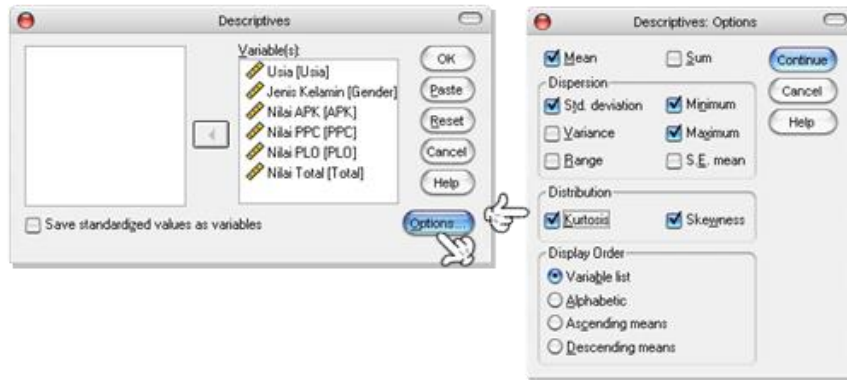
Jika definisi value diabaikan maka pada *bar chart* maupun tabel analisis yang terlihat bukan laki-laki dan perempuan melainkan 1 dan 2. Begitu juga judul tabel dan judul histogram: “jenis kelamin“, ini merupakan hasil dari proses label yang telah kita lakukan untuk variabel Gender. Fungsi label ini bermanfaat untuk para pembaca analisis. Selanjutnya adalah penggunaan perintah Descriptives.

### 3. Descriptives

Dengan menggunakan data sebelumnya langkah-langkah perintah Descriptives adalah sebagai berikut:

- a) Klik menu [Analyze] -> [Descriptives Statistics] -> [Descriptives], muncul *dialog box Descriptives*.
- b) Masukkan variabel yang akan dianalisis ke form **Variables(s)**.
- c) Klik [Options] untuk melakukan setting optional, muncul *dialog box Descriptives: Options*.
- d) Centang analisis yang diperlukan, saya memilih:
  - ✓ Mean
  - ✓ Std. deviation
  - ✓ Minimum
  - ✓ Maximum
  - ✓ Kurtosis
  - ✓ Skewness

e) Centang Variable list pada form **Display Order**.



Gambar 9. Langkah-Langkah Descriptives Statistics

f) Klik [Continue] dan [OK]. Hasil analisis akan terlihat seperti tabel yang ditunjukkan Gambar 10 di bawah ini:

Descriptive Statistics									
	N	Minimum	Maximum	Mean	Std.	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
Usia	10	19	22	20,80	,919	-,601	,687	,396	1,334
Jenis Kelamin	10	1	2	1,20	,422	1,779	,687	1,406	1,334
Nilai APK	10	60	90	73,00	11,595	,342	,687	-1,227	1,334
Nilai PPC	10	50	90	74,00	12,649	-,544	,687	-,026	1,334
Nilai PLO	10	40	90	64,00	15,776	-,195	,687	-,144	1,334
Nilai Total	10	170	230	211,00	21,833	-,959	,687	-,217	1,334
Valid N (listwise)	10								

Gambar 10. Output Descriptives

Tabel *output* di atas menunjukkan jumlah pengukuran (N), nilai minimum (Minimum), nilai maksimum (Maximum), nilai rata-rata (Mean), standar deviasi (Std.), Skewness, dan Kurtosis dari masing-masing variabel.

Nilai *skewness* merupakan ukuran kesimetrisan histogram, sedangkan *kurtosis* merupakan ukuran datar atau runcingnya histogram. Idealnya nilai skewness dan kurtosis pada distribusi normal adalah nol. Oleh karena itu:

- ✓ Jika nilai skewness positif maka distribusi data “miring ke kiri distribusi normal” (ada frekuensi nilai yang tinggi di sebelah kiri titik tengah distribusi normal), sebaliknya apabila skewness negatif maka distribusi data ”miring ke kanan distribusi normal” (kiri bagi kita yang melihatnya).

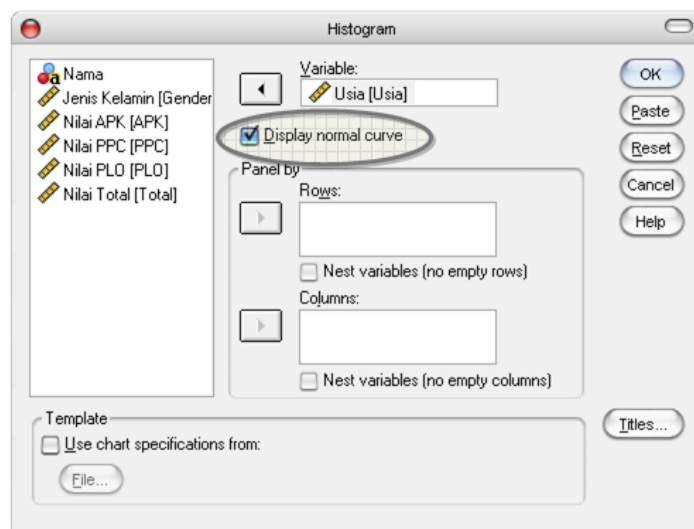
- ✓ Jika nilai kurtosis positif maka distribusi data “meruncing” (ada satu nilai yang mendominasi), sebaliknya apabila Kurtosis Negatif maka distribusi data “melandai” (varians besar).

Perhatikan *Gambar* di atas, variabel Usia memiliki skewness negatif dan kurtosis positif, artinya distribusinya “miring ke kiri distribusi normal” dan “meruncing”.

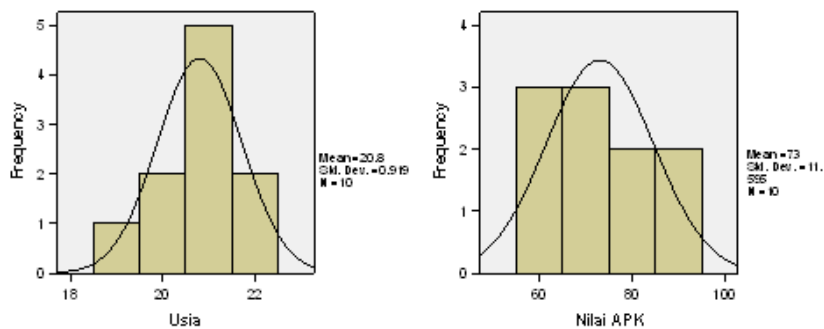
Pada variabel nilai APK, nilai skewness positif dan nilai kurtosis negatif, artinya distribusinya “miring ke kanan distribusi normal” dan “melandai”. Sebagai pembuktian, buat histogram untuk variabel Usia dan Nilai APK. Berikut caranya:

- g) Klik menu [Graphs] -> [Histogram], maka muncul *dialog box Histogram*.
- h) Pilih variabel Usia dan masukkan dalam *form Variable*.
- i) Centang Display normal curve, untuk memperlihatkan kurva normal.
- j) Selanjutnya klik [OK].

Lakukan langkah yang sama untuk variabel nilai APK.



*Gambar 11.* Menampilkan Histogram bersama Kurva Normal



Gambar 12. Analisis Skewness dan Kurtosis pada Histogram

Gambar histogram di atas menunjukkan histogram untuk variabel Usia memiliki distribusi “miring ke kiri distribusi normal” karena nilainya skewness negatif dan “meruncing” karena nilai kurtosis positif.

Sebaliknya, histogram untuk variabel Nilai APK memiliki distribusi “miring ke kanan distribusi normal” karena nilainya skewness positif dan “melandai” karena nilai kurtosis negatif.

Di sini, anda bisa menentukan apakah distribusi tersebut normal atau tidak. Anda bisa saja menyatakan normal, karena menyerupai bentuk lonceng tetapi agak serong, tapi orang lain mungkin akan menyatakan tidak normal karena jauh dari bentuk lonceng. Jika sulit mengambil keputusan, silahkan lakukan pengujian normalitas yang lebih *advance*, misal dengan Uji Kolmogorov-Smirnov. Selanjutnya, kita masuk pada penggunaan perintah Explore.

#### 4. Explore

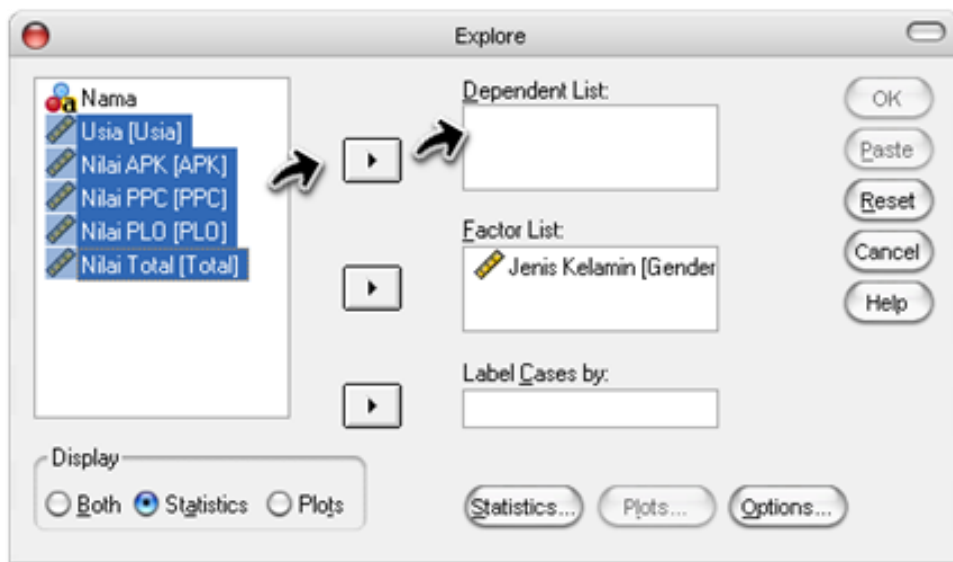
Perintah Explore digunakan untuk membandingkan antara dua atau lebih kelompok dengan satu variabel. Sebagai contoh, jika kita menggunakan Jenis Kelamin sebagai variabel independen; variabel ini mendefinisikan kelompok (Laki-Laki dan Perempuan), kemudian membandingkannya dengan variabel lain, seperti Usia. Perintah Explore; contoh dalam kasus *mean*, akan menghasilkan berapa rata-rata usia laki-laki dan berapa rata-rata usia perempuan. Ukuran-ukuran yang dihasilkan perintah Explore antara lain: ukuran-ukuran pemusatan data (*mean* dan *median*), ukuran



penyebaran (*range*, *interquartile range*, standar deviasi, varians, minimum, dan maksimum), ukuran *kurtosis*, dan *skewness*.

Berikut langkah-langkah perintah Explore:

- a) Klik menu [Analyze] -> [Descriptives Statistics] -> [Explore], muncul *dialog box Explore*.
- b) Isi variabel Jenis Kelamin pada *form Factor List*.
- c) Isi variabel Usia, Nilai APK, Nilai PPC, Nilai PLO, dan Nilai Total pada *form Dependent List*.
- d) Pada *form Display* ada tiga pilihan:
  - ✓ Klik [Plots] bila perlu grafik *boxplot*.
  - ✓ Klik [Statistics] bila tidak perlu grafik *boxplot*.
  - ✓ Klik [Both] bila perlu keduanya.
- e) Terakhir klik [OK].



Gambar 13. Dialog Box Explore

Descriptives				Statistic	Std. Error
Usia	Jenis Kelamin				
	Laki-laki	Mean		20,88	,227
		95% Confidence Interval for Mean	Lower Bound	20,34	
			Upper Bound	21,41	
		5% Trimmed Mean		20,88	
		Median		21,00	
		Variance		,411	
		Std. Deviation		,641	
		Minimum		20	
		Maximum		22	
		Range		2	
		Interquartile Range		1	
		Skewness		,068	,752
		Kurtosis		,741	1,481
	Perempuan	Mean		20,50	1,500
		95% Confidence Interval for Mean	Lower Bound	1,44	
			Upper Bound	39,56	
		5% Trimmed Mean		.	
		Median		20,50	
		Variance		4,500	
		Std. Deviation		2,121	
		Minimum		19	
		Maximum		22	
		Range		3	
		Interquartile Range		.	
		Skewness		.	.
		Kurtosis		.	.

Gambar 14. Contoh Output Explore untuk Variabel Usia

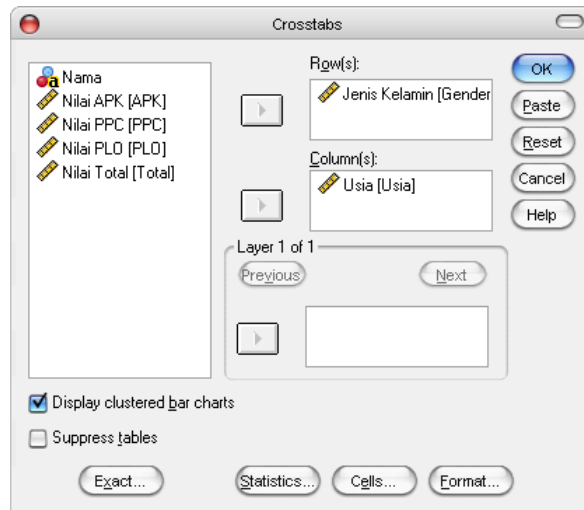
f) Selanjutnya, kita masuk pada penggunaan perintah Crosstabs.

## 5. Crosstabs

Jika perintah Frequencies digunakan untuk memperoleh jumlah pada nilai-nilai sebuah variabel tunggal, perintah Crosstabs digunakan untuk memperoleh jumlah pada nilai-nilai lebih dari satu variabel. Apabila analisis statistik deskriptif sebelumnya mengolah data secara keseluruhan dalam setiap variabel dengan menghitung perhitungan statistik seperti Mean, Standar deviasi, Kurtosis, etc. Pada Crosstabs, setiap nilai pada variabel yang dianalisis dijabarkan jumlahnya, dengan begitu kita dapat mengetahui berapa jumlah subyek laki-laki yang berusia 19 tahun, 20 tahun, dst. Deskripsi data pada Crosstabs akan disajikan dalam bentuk tabel silang (*crosstab*) yang terdiri dari baris dan kolom.

Berikut langkah-langkah perintah Crosstabs:

- a) Klik menu [Analyze] -> [Descriptives Statistics] -> [Crosstabs], muncul *dialog box Crosstabs*.
- b) Isi variabel Jenis Kelamin pada *form Row(s)*.
- c) Isi variabel yang akan dianalisis, contoh variabel Usia pada *form Column(s)*.



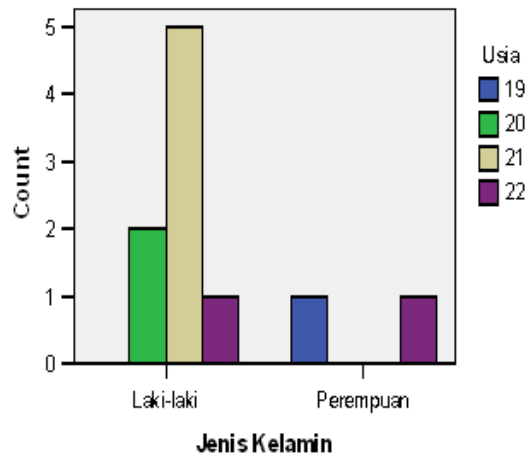
Gambar 15. Dialog Box Crosstabs

- d) Centang [Display clustered bar charts] untuk menampilkan chart bar dari output.
- e) Untuk *dialog box* [Statistics], [Cells], dan [Format] biarkan sesuai dengan *default* SPSS.
- f) Terakhir klik [OK].

**Jenis Kelamin \* Usia Crosstabulation**

Count		Usia				Total
		19	20	21	22	
Jenis Kelamin	Laki-laki	0	2	5	1	8
	Perempuan	1	0	0	1	2
Total		1	2	5	2	10

Gambar 16. Output Crosstabs



Gambar 17. Clustered Bar Charts

Statistik deskriptif memberikan informasi inti dari kumpulan data, seperti ukuran-ukuran pemusatan data (*mean* dan *median*), ukuran penyebaran (*range*, *interquartile range*, standar deviasi, varians, minimum, dan maksimum), ukuran *kurtosis*, dan *skewness* serta representasi piktorialnya. Tabel, diagram, dan grafik yang sering ditemukan di majalah dan koran-koran merupakan salah satu contoh penggunaan statistik deskriptif.

# BAB 6 KONSEP DASAR PROBABILITAS

---

## A. PENGERTIAN PROBABILITAS

Lind (2002) dalam mendefinisikan probabilitas sebagai : Suatu ukuran tentang kemungkinan suatu peristiwa (event) akan terjadi dimasa mendatang. Probabilitas dinyatakan antara ) sampai 1 atau dalam presentase”

Tiga hal penting dalam membicarakan probabilitas :

### 1. Percobaan (*Experiment*)

Pengamatan terhadap beberapa aktivitas atau proses yang memungkinkan timbulnya paling sedikit dua peristiwa tanpa memperhatikan peristiwa mana yang akan terjadi.

### 2. Hasil (*Outcome*)

Suatu hasil dari sebuah percobaan. Dalam hasil ini semua kejadian akan dicatat atau dalam artian seluruh peristiwa yang akan terjadi dalam sebuah percobaan. Misalnya dalam mengikuti ujian semester maka hasil yang akan diperoleh ada mahasiswa yang lulus dan ada yang tidak lulus. Ada yang lulus memuaskan ada yang tidak memuaskan.

### 3. Peristiwa (*Event*)

Kumpulan dari satu atau lebih hasil yang terjadi pada sebuah percobaan atau kegiatan.

Contoh :

Percobaan	Pertandingan sepak bola antara fakultas kesehatan masyarakat dan fakultas hukum Universitas Indonesia
Hasil	Fakultas kesehatan masyarakat menang, Fakultas hukum kalah, Seri, tidak ada yang menang dan tidak ada yang kalah
Peristiwa	Fakultas kesehatan masyarakat menang

Probabilitas dinyatakan dalam bentuk pecahan dari 0 sampai 1. Probabilitas ) menunjukkan sesuatu yang tidak mungkin terjadi, sedangkan

probabilitas 1 menunjukkan peristiwa yang terjadi. Contoh penulisan probabilitas dalam decimal atau persentase :

1. Pada hari Jumat adalah penutupan bursa saham, maka kebanyakan investor berusaha meraih keuntungan melalui penjualan saham atau yang biasanya diistilahkan profit taking, sehingga probabilitas menjual mencapai 0.7 sedangkan membeli 0.3.
2. Melihat kondisi kesiapan mahasiswa yang mengikuti ujian statistika inferensial, maka mahasiswa yang mempunyai probabilitas untuk lulus 70% dan kalah 30%

Probabilitas kejadian dengan nilai 0 berarti peristiwa yang tidak mungkin terjadi, seperti seorang anak balita melahirkan seorang bayi. Sedangkan probabilitas dengan nilai 1 adalah peristiwa yang terjadi, seperti manusia pasti akan meninggal.

## B. MANFAAT MEMPELAJARI PROBABILITAS

Berguna untuk pengambilan keputusan yang tepat, karena kehidupan di dunia tidak ada kepastian, sehingga diperlukan untuk mengetahui berapa besar probabilitas suatu peristiwa akan terjadi. Probabilitas dinyatakan dalam angka pecahan antara 0 sampai 1 atau dalam presentase.

Contoh :

Seluruh mahasiswa IIK STRADA Indonesia harus memiliki sertifikat bahasa inggris. Di Kota Kediri sendiri banyak terdapat tempat kursus bahasa inggris diantaranya Mozaik, Mahesa, Elfes dan lainnya. Maka akan muncul kebingungan dalam memilih tempat kursus. Untuk menentukan pilihan biasanya mahasiswa akan bertanya kepada teman-temannya, mereka mahasiswa mungkin kursus dimana?. Dari ratusan mahasiswa mungkin anda bertanya hanya pada 20 orang mahasiswa. Yang paling banyak diminati anda akan memilih tempat tersebut untuk kursus.

Dari contoh tersebut dapat dilihat bahwa keputusan diambil hanya dari beberapa contoh atau sampel dari populasi keseluruhan.

## C. PENDEKATAN PROBABILITAS

Untuk menentukan tingkat probabilitas suatu kejadian, maka ada tiga pendekatan yaitu pendekatan klasik, pendekatan relative dan pendekatan subjektif.

## 1. Pendekatan klasik

Diasumsikan bahwa semua peristiwa mempunyai kesempatan yang sama untuk terjadi (*equally likely*). Probabilitas suatu peristiwa kemudian dinyatakan sebagai rasio antara jumlah kemungkinan hasil dengan total kemungkinan hasil (rasio peristiwa terhadap hasil)

$$\text{Probabilitas} = \frac{\text{jumlah kemungkinan hasil (peristiwa)}}{\text{jumlah total kemungkinan hasil}}$$

Contoh :

Pada kegiatan mahasiswa belajar semua hasil ada yang sangat memuaskan dan terpuji. Jumlah hasil ada 3 dan hanya 1 peristiwa yang terjadi, maka probabilitas setiap peristiwa 1/3.

Pada suatu percobaan hanya 1 peristiwa yang terjadi, dan peristiwa lain tidak mungkin terjadi pada waktu yang bersamaan maka dikenal sebagai peristiwa saling lepas (*mutually exclusive*).

Pada suatu percobaan atau kegiatan semua hasil mempunyai probabilitas yang sama, dan hanya satu peristiwa yang terjadi maka peristiwa ini dikenal dengan lengkap terbatas kolektif (*collection exhaustive*).

## 2. Pendekatan relatif

Probabilitas suatu keadaan tidak dianggap sama, tergantung dari beberapa banyak suatu kejadian terjadi, yang dinyatakan sebagai berikut :

$$\text{Probabilitas kejadian relatif} = \frac{\text{jumlah peristiwa yang terjadi}}{\text{jumlah total percobaan}}$$

Contoh :

Dari kegiatan belajar mahasiswa dapat dilihat hasilnya pada wisuda sarjana Universitas Brawijaya tahun 2019 sebanyak 800 orang mahasiswa. 500 orang lulus dengan memuaskan, 200 orang dengan sangat memuaskan dan 100 orang dengan predikat terpuji.

- Probabilitas lulus memuaskan adalah :

$$\text{Probabilitas lulus memuaskan} = \frac{500}{800} = 0.625$$

- Probabilitas lulus dengan sangat memuaskan adalah :

$$\text{Probabilitas lulus sangat memuaskan} = \frac{200}{800} = 0.25$$

- Probabilitas lulus dengan terpuji adalah :

$$\text{Probabilitas lulus dengan terpuji} = \frac{100}{800} = 0.125$$

### 3. Pendekatan subjektif

Menentukan besarnya probabilitas suatu peristiwa didasarkan pada penilaian pribadi dan dinyatakan dalam derajat kepercayaan.

Contoh :

Menurut pengamat politik, Jokowi akan menang dalam Pemilu Indonesia Tahun 2019

## D. HUKUM PROBABILITAS

Dalam teori probabilitas, probabilitas kejadian dilambangkan dengan “P”, apabila kejadian jual saham dilambangkan dengan huruf “A”, maka probabilitas jual saham dilambangkan dengan  $P(A)$ . sebaliknya apabila kejadian beli saham dilambangkan dengan “B”, maka probabilitas beli saham dilambangkan dengan  $P(B)$ .

### 1. Hukum penjumlahan

Hukum penjumlahan menghendaki peristiwa yang saling lepas (*mutually exclusive*) yaitu apabila suatu peristiwa terjadi, maka peristiwa lain tidak dapat terjadi pada saat bersamaan.

Hukum ini dilambangkan sebagai :

$$P(A \text{ atau } B) = P(A) + P(B)$$

Untuk kejadian yang lebih banyak dilambangkan sampai  $n$  yaitu :

$$P(A \text{ atau } \dots n) = P(A) + P(B) + \dots + P(n)$$

- Contoh :



Berikut adalah kegiatan unit mahasiswa untuk tiga sekolah tinggi kesehatan di Kediri dengan jumlah total sebanyak 200 kegiatan

Janis transaksi	Volume transaksi
Kegiatan A	120
Kegiatan B	80
Jumlah total kegiatan	200

Penyelesaian :

Dari data diatas diketahui bahwa

$$\text{Probabilitas Kegiatan A} = P(A) = \frac{120}{200} = 0.60$$

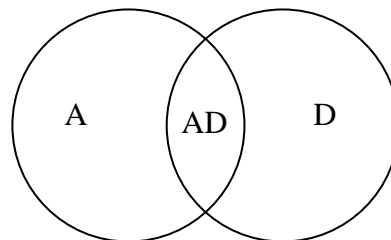
$$\text{Probabilitas Kegiatan B} = P(B) = \frac{80}{200} = 0.40$$

Sehingga probabilitas A atau B,

$$P(A \text{ atau } B) = P(A) + P(B) = 0.6 + 0.4 = 1.0$$

#### a) Peristiwa atau kejadian bersama

Pada peristiwa bersama dua atau lebih peristiwa dapat terjadi secara bersama-sama, peristiwa bersama tersebut dapat lebih mudah dilihat dengan diagram Venn seperti berikut :



Penjumlahan probabilitas dengan adanya unsure kegiatan bersama, maka rumus penjumlahan dirumuskan kembali menjadi sebagai berikut :

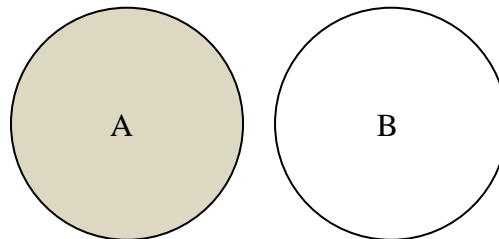
$$P(A \text{ atau } D) = P(A) + P(D) - P(AD)$$

Dimana :

- ✓ P (A atau D) : probabilitas terjadinya A atau D atau A dan D bersama sama
- ✓ P (A) : probabilitas terjadinya A
- ✓ P (D) : probabilitas terjadinya D
- ✓ P (AD) : probabilitas terjadinya A dan D bersama-sama

**b) Kejadian saling lepas (*mutually exclusive*)**

Kejadian saling lepas terjadi apabila hanya satu dari dua atau lebih peristiwa yang dapat terjadi. Dapat digambarkan dengan diagram Venn sebagai berikut :



Oleh sebab itu, untuk peristiwa yang saling lepas, probabilitas kejadian A dan B yang dinyatakan  $P(A \text{ atau } B)$

$$P(A \text{ atau } B) = P(A) + P(B) - P(AB)$$

Karena  $P(AB) = 0$  maka,

$$P(A \text{ atau } B) = P(A) + P(B) - 0$$

Sehingga

$$P(A \text{ atau } B) = P(A) + P(B)$$

• **Contoh**

Cobalah hitung beberapa probabilitas kegiatan mahasiswa yang meliputi kegiatan A dan kegiatan B disebut  $P(AB)$  dan probabilitas kejadian untuk sekolah kesehatan X, Y dan Z disebut  $P(CDE)$

Kegiatan	Sekolah tinggi kesehatan			Jumlah
	X (C)	Y (D)	Z (E)	
Kegiatan (A)	30	50	40	120
Kegiatan (B)	40	30	10	80
Jumlah	70	80	50	200

Penyelesaian :

Probabilitas kejadian A dan B adalah kejadian yang saling lepas, maka  $P(AB) = 0$ , maka hukum penjumlahan untuk peristiwa saling lepas adalah :

- $P(A \text{ atau } B) = P(A) + P(B) - P(AB)$   
 $= 0.6 + 0.4$   
 $= 1.0$

Probabilitas kejadian ketiga sekolah tinggi kesehatan juga merupakan kejadian saling lepas, maka hukum penjumlahan adalah

- $P(C \text{ atau } D \text{ atau } E) = P(C) + P(D) + P(E) - P(CDE)$   
 $= 0.35 + 0.40 + 0.25 - 0$   
 $= 1.0$

- Probabilitas  $P(C \text{ atau } D \text{ atau } E) = P(C) + P(D) + P(E) - P(CDE)$   
 $= 0.35 + 0.40 + 0.25 - 0$   
 $= 1.0$

Probabilitas  $P(C \text{ atau } D)$

- $P(C \text{ atau } D) = P(C) + P(D) - P(CD)$   
 $= 0.35 + 0.40$   
 $= 0.75$

## 2. Hukum Perkalian

Dalam hukum perkalian dikehendaki setiap peristiwa independent yaitu semua peristiwa yang terjadi tanpa harus menghalangi peristiwa lain terjadi.

Peristiwa independent adalah terjadinya peristiwa atau kejadian tidak mempengaruhi probabilitas terjadinya peristiwa lain. Dapat dinyatakan dalam bentuk :

$$P(A \text{ dan } B) = P(A) \times P(B)$$

### a) Probabilitas bersyarat (*conditional probability*)

Probabilitas bersyarat adalah suatu peristiwa akan terjadi, dengan ketentuan peristiwa lain telah terjadi. Hukum perkalian untuk probabilitas bersyarat bahwa peristiwa B terjadi dengan syarat peristiwa A telah terjadi dinyatakan sebagai berikut :

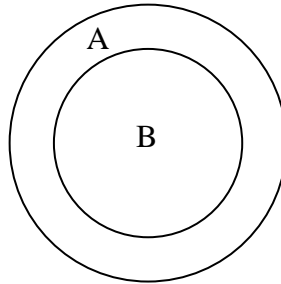
$$P(A \text{ dan } B) = P(A) \times P(B)$$

### b) Persistiwa pelengkap (*complementary event*)

Peristiwa pelengkap menunjukkan bahwa apabila ada dua peristiwa A dan B yang saling melengkapi, sehingga jika peristiwa A tidak terjadi, maka peristiwa B pasti terjadi. Maka probabilitas keduanya dapat dirumuskan sebagai berikut :

$$P(A) + P(B) = 1 \text{ atau } P(A) = 1 - P(B)$$

Dalam bentuk diagram Venn dapat digambarkan sebagai berikut :



### c) Diagram Pohon Probabilitas

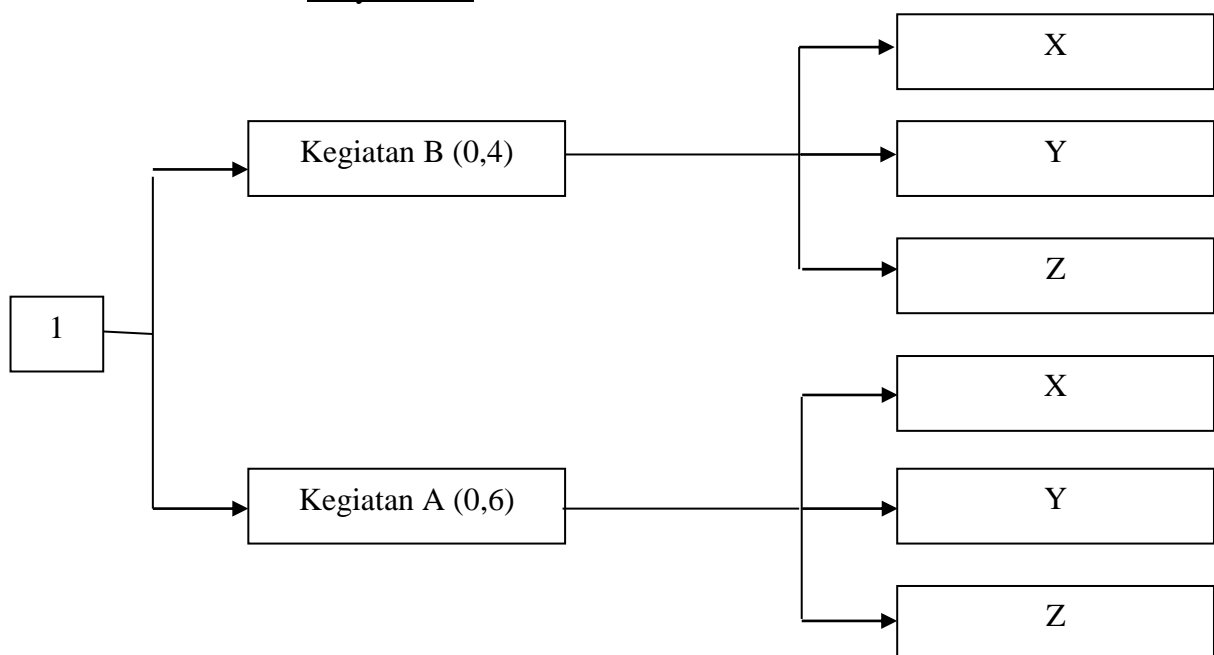
Tahapan dalam menyusun diagram pohon :

- a) Tahap 1 adalah langkah awal kegiatan, kita mulai dengan tanda titik atau bulatan dengan angka, tahap 1 diumpamakan sebagai pohonnya dengan pohon utamanya berupa kegiatan dibursa saham. Nilai probabilitas pada tahap 1 adalah 1.
- b) Tahap 2, membuat cabang. Kegiatan di sekolah kesehatan ada 2 yaitu kegiatan A dan kegiatan B dengan probabilitas kegiatan A = 0.6 dan probabilitas kegiatan B = 0.4.
  - Nilai probabilitas pada cabang =  $0.6+0.4 = 1$
- c) Tahap 3 membuat ranting. Pada setiap cabang baik kegiatan A maupun kegiatan B ada 3 ranting jenis saham. Ranting jenis sekolah kesehatan yaitu X,Y dan Z.
  - Nilai probabilitas setiap ranting =  $0.35+0.40+0.25 = 1$
- d) Tahap 4, menghitung probabilitas bersamma (joint probability) antara kejadian pertama A dan B dengan kejadian kedua C,D dan E. kita bisa menghitung probabilitas  $P(C|A)$  atau  $P(D|B)$  secara langsung.
  - Nilai probabilitas keseluruhan pada tahap empat juga harus sama dengan 1

• **Contoh :**

Hasil penelitian di Jakarta menunjukkan bahwa 60% dari usaha kecil dan menengah (UKM) tidak berbadan hukum, sedang sisanya berbadan hukum. Bank sebagai lembaga pembiayaan dengan memperhatikan aspek kehati-hatian memberikan probabilitas 80% kepada UKM berbadan hukum masih mempunyai kesempatan sebesar 20% untuk mendapatkan kredit. Hitunglah beberapa persen probabilitas UKM mendapatkan kredit dan Bank?

Penyelesaian



E. TEOREMA BAYES

Tujuan dari inferensi statistik adalah untuk menarik kesimpulan dari data sampel yang diketahui tentang populasi yang tidak ada datanya. Sebagai contoh, kita tahu dari sampel bahwa 55 persen pemilih cenderung untuk memilih pilihan A, tapi sebenarnya berapa banyak pemilih secara keseluruhan yang cenderung memilih A?

Saat ini, terdapat dua pendekatan filosofis utama dalam statistik inferens, yang pertama disebut sebagai pendekatan *frequentist* atau kadang-kadang disebut sebagai pendekatan klasik (karena berkembang lebih dulu). Dalam pendekatan ini, prosedur dikembangkan hanya dengan melihat performa seluruh kemungkinan sampel acak (*all possible random sample*) saat ini. Informasi sampel acak yang diperoleh sebelumnya

(pada percobaan/observasi lain di masa lalu) diabaikan. Kemudian pendekatan kedua, dikenal sebagai Bayesian

## 1. Frequentist versus Bayesian

Pendekatan frequentist berlandaskan pada ide-ide dibawah ini:

- a) Parameter, yaitu karakteristik dari populasi, adalah konstan namun tidak diketahui.
- b) Probabilitas selalu diinterpretasikan sebagai frekuensi relatif jangka panjang, tak peduli datanya.
- c) Prosedur statistik dinilai dengan seberapa baik prosedur itu dalam jangka panjang dengan mengulang-ulang percobaan sampai tak hingga.

Karena dalam pendekatan ini parameter adalah tetap, maka kita tidak bisa membuat pernyataan tentang peluang dari nilai parameter tersebut (bagaimana kita menyatakannya dalam peluang jika nilai parameter adalah tetap dengan kata lain pasti). Interval kepercayaan tidak memiliki arti peluang akan nilai parameter, namun hanya digunakan untuk uji hipotesis apakah nilai penduga parameter bisa kita terima atau tidak.

Misal diperoleh  $P(a < \theta \leq b) = 0.95$ , kita tidak bisa mengatakan peluang  $\theta$  diantara  $[a, b]$  adalah 95 persen karena jika kita mengatakan demikian berarti  $\theta$  adalah suatu nilai acak. Karena itu dalam frequentist interval itu selalu diartikan begini: dari 100 percobaan dengan random sampel iid maka 95 percobaan akan mendapatkan nilai penduga parameter  $\hat{\theta}$  berada pada interval  $[a, b]$ . Sedangkan Bayesian berlandaskan pada ide-ide berikut:

- a) Sejak kita tak pernah yakin akan nilai sebenarnya dari parameter, maka parameter dianggap sebagai suatu random variabel.
- b) Aturan probabilitas digunakan secara langsung untuk melakukan inferens tentang parameter.
- c) Pernyataan probabilitas tentang parameter diinterpretasikan sebagai “derajat kepercayaan”. Distribusi prior adalah subyektif. Setiap orang bisa memilih priornya sendiri, yang mengandung bobot relatif yang diberikannya pada parameter tersebut, yang mengukur bagaimana sejauh mana bisa diterima/dipercaya setiap parameter tersebut sebelum percobaan.
- d) Setelah itu kita menyesuaikan kepercayaan/penerimaan kita pada parameter tersebut setelah memperoleh data dengan menggunakan teorema

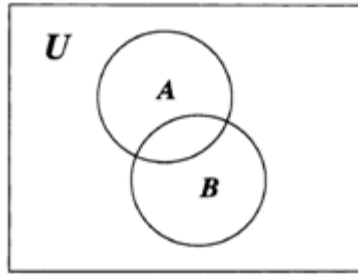
Bayes, sehingga akan menghasilkan distribusi posterior, yang memberikan bobot relatif tiap parameter setelah data dianalisis. Distribusi posterior diperoleh dari dua sumber, yaitu: **distribusi prior** dan **data pengamatan**.

Dengan pendekatan Bayesian ini kita bisa membuat pernyataan probabilitas dari parameter karena memang parameter adalah random variabel.  $P(a < \theta \leq b) = 0.95$  memang berarti peluang nilai parameter  $\theta$  berada pada interval  $[a,b]$  dengan syarat data seperti pada data observasi adalah 95 persen. Hanya dengan teorema Bayes kita bisa secara konsisten memperbaiki kepercayaan kita pada parameter berdasarkan data yang benar-benar terjadi! Selain itu pendekatan Bayesian sangat bermanfaat dalam menangani parameter pengganggu (*nuisance parameter*). Parameter pengganggu adalah suatu parameter yang kita tidak tertarik untuk melakukan inferens atasnya, tapi kita tidak ingin parameter tersebut mempengaruhi inferens tentang parameter utama (tidak kita bahas dalam artikel ini).

## 2. Teorema Bayes

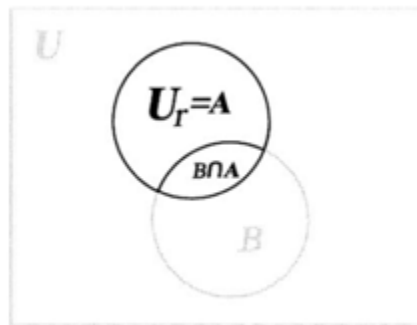
Sebelum lebih jauh mengenal Bayesian kita harus kenal lebih dalam dulu dengan teorema Bayes, yang menjadi landasan utama dalam pendekatan Bayesian. Pertama kita harus berkenalan dulu dengan Thomas Bayes, seorang pendeta dan matematikawan berkebangsaan Inggris, yang pertama kali mengemukakan teorema Bayes. Dalam tulisannya yang diterbitkan tahun 1763, 3 tahun setelah kematiannya, Bayes memperkenalkan sebuah versi dari persamaan beberapa probabilitas yang sekarang dikenal sebagai teorema Bayes. Saat paper ini pertama kali terbit, hanya ada sedikit ekspektasi bahwa persamaan sederhana ini bisa memecahkan banyak permasalahan dalam teori peluang. Namun siapa sangka jika dua ratus tahun kemudian, teorema Bayes telah menjadi sesuatu yang penting dan saat ini menjadi dasar bagi inferensi statistik Bayesian.

Untuk memahami teorema Bayes, kita harus pahami dulu peluang bersyarat. Sekarang pikirkan permasalahan ini: jika kita tahu suatu *event* (peristiwa) **telah** terjadi, apakah akan mempengaruhi peluang terjadinya event yang lain? Misal terdapat dua event A dan B yang saling berpotongan seperti digambarkan dalam diagram Venn di bawah ini.



Gambar.1 Diagram Venn dua event A dan B dalam U (semesta)

Daerah perpotongan kita sebut irisan, atau  $A \cap B$ , dimana seluruh elemennya adalah anggota A sekaligus anggota B. Misal kita tahu bahwa A telah terjadi lebih dulu, maka seluruh kemungkinan di luar peristiwa A menjadi tidak mungkin. Kini kita hanya memperhatikan seluruh hasil yang hanya ada didalam event A, digambarkan sebagai berikut:



Gambar 2. U setelah A terjadi

Kita lihat bahwa bagian peristiwa B yang masih relevan (masih mungkin terjadi) setelah peristiwa terjadi hanyalah B yang ada di dalam A, atau  $B \cap A$ .

Dengan demikian peluang terjadinya dua peristiwa berturut-turut, dimana A terjadi lebih dulu lalu B menyusul terjadi (dengan kata lain: peluang terjadinya B jika A telah terjadi lebih dulu), dinotasikan dengan  $P(B | A)$  adalah:

$$P(B | A) = \frac{n(B \cap A)}{n(A)} = \frac{\frac{n(B \cap A)}{n(A \cup B)}}{\frac{n(A)}{n(A \cup B)}}$$

$$P(B | A) = \frac{P(A \cap B)}{P(A)}$$



maka:

$$P(A \cap B) = P(B | A) P(A)$$

Dan sekarang misal peristiwa itu dibalik menjadi event B terjadi lebih dulu baru kemudian event A menyusul terjadi. Maka peluang terjadinya A dengan syarat B terjadi lebih dulu adalah:

$$P(A | B) = \frac{P(B \cap A)}{P(B)}$$

$$\Leftrightarrow P(B \cap A) = P(A | B) P(B)$$

dimana kita tahu bahwa dalam teori himpunan  $P(A \cap B) = P(B \cap A)$  (sifat komutatif), sehingga:

$$P(A | B) P(B) = P(B | A) P(A)$$

$$\Leftrightarrow P(A | B) = \frac{P(B | A) P(A)}{P(B)}$$

Dari teori himpunan juga kita tahu bahwa  $B = (A \cap B) \cup (A \cap B^c)$  dimana  $(A \cap B)$  dan  $(A \cap B^c)$  adalah disjoint (saling bebas, tidak saling berpotongan), maka:

$$\begin{aligned} P(B) &= P(A \cap B) + P(A \cap B^c) \\ &= P(A | B) P(B) + P(A | B^c) P(B^c) \end{aligned}$$

sehingga disubstitusikan ke persamaan diatas menjadi:

$$P(A | B) = \frac{P(B | A) P(A)}{P(B | A) P(A) + P(B | A^c) P(A^c)}$$

Hasil diatas adalah bentuk dasar dari teorema Bayes. Yang menjadi catatan diatas bahwa A dan  $A^c$  adalah partisi dari semesta sedemikian hingga  $A \cup A^c = S$  dan A dan  $A^c$  adalah disjoint. Sehingga seandainya pun semesta himpunan dipartisi sejumlah n partisi sedemikian hingga:

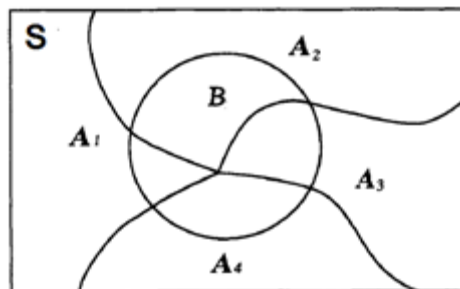
$$S = \bigcup_{i=1}^n A_i$$

maka persamaan teorema Bayes diatas disesuaikan menjadi:

$$P(A_i | B) = \frac{P(B | A_i) P(A_i)}{\sum_j P(B | A_j) P(A_j)}$$

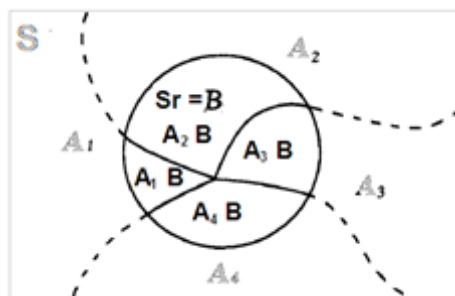
Lebih jelasnya lihat ilustrasi dibawah ini:

- Misal  $S=A_1 + A_2 + A_3 + A_4$ , yang berarti semesta S dipartisi menjadi empat partisi, kemudian didalam S juga terdapat event B. Digambarkan sebagai berikut:



Gambar 3. Empat event  $A_i$  yang mempartisi S dan satu event B

- Misal event B terjadi lebih dulu, sehingga seluruh kemungkinan event di luar B menjadi tidak mungkin terjadi. Sehingga diilustrasikan sebagai berikut:



Gambar 5. Semesta terpotong setelah event B terjadi

# BAB 7 DISTRIBUSI PROBABILITAS

---

## A. PENGERTIAN DISTRIBUSI PROBABILITAS

Probabilitas sangat dibutuhkan, karena kebenaran dari suatu kesimpulan yang dibuat dari analisis data sebetulnya tidak dapat dipastikan benar secara absolut, disebabkan data berdasarkan dari sampel

Distribusi Probabilitas adalah suatu distribusi yang menggambarkan peluang dari sekumpulan variat sebagai pengganti frekuensinya. Probabilitas kumulatif adalah probabilitas dari suatu variabel acak yang mempunyai nilai sama atau kurang dari suatu nilai tertentu.

Distribusi normal, disebut pula distribusi Gauss, adalah distribusi probabilitas yang paling banyak digunakan dalam berbagai analisis statistika. Distribusi normal baku adalah distribusi normal yang memiliki rata-rata nol dan simpangan baku satu. Distribusi ini juga dijuluki kurva lonceng (bell curve) karena grafik fungsi kepekatan probabilitasnya mirip dengan bentuk lonceng.

## B. JENIS-JENIS DISTRIBUSI PROBABILITAS NORMAL BAKU

### 1. Distribusi Probabilitas Berdasarkan Teori

Penyusunan distribusi ini berdasarkan teori peluang.

Contoh : jika kita ingin mengetahui probabilitas jenis kelamin bayi yang dilahirkan dua kali berturut-turut, maka terdapat 4 kemungkinan sebagai berikut:

- Kelahiran pertama laki-laki, kelahiran kedua laki-laki
- Kelahiran pertama laki-laki, kelahiran kedua perempuan
- Kelahiran pertama perempuan, kelahiran kedua perempuan
- Kelahiran pertama perempuan, kelahiran kedua laki-laki.

Distribusi peluang tersebut dapat disusun dalam bentuk tabel dan grafik frekuensi distribusi seperti berikut.

Tabel 1.1. probabilitas jenis kelamin dua kelahiran berturut-turut.

Kelahiran 1	Kelahiran 2	Jumlah kelahiran perempuan	probabilitas
Laki-laki	Laki-laki	0	$0,5 \times 0,5 = 0,25$
Laki-laki	Perempuan	1	$0,5 \times 0,5 = 0,25$
perempuan	Perempuan	2	$0,5 \times 0,5 = 0,25$
perempuan	Laki-laki	1	$0,5 \times 0,5 = 0,25$

Dari tabel diatas dapat disusun frekuensi distribusi probabilitas sebagai berikut:

$F_P$	$F_L$	Probabilitas
0	2	0,25
1	1	0,5
2	0	0,25

## 2. Distribusi Probabilitas Berdasarkan Subjektif

Distribusi probabilitas subjektif berarti penyusunan distribusi frekuensi atas dasar pertimbangan pribadi. Misalnya seorang dokter puskesmas menyatakan bahwa jumlah kunjungan setiap hari senin adalah 40 orang dengan peluang sebesar 10 %, hari selasa jumlah kunjungan sebesar 45 orang dengan probabilitas 30%, hari rabu jumlah kunjungan 50 orang dengan probabilitas 50% dan pada hari kamis jumlah kunjungan sebanyak 55 orang dengan probabilitas 20%.

Pernyataan dokter puskesmas tersebut dapat disusun sebagai berikut :

- Jumlah kunjungan : 40 45 50 55
- Probabilitas (dalam persen) : 10 30 50 20

### 3. Distribusi Probabilitas Berdasarkan Pengalaman.

Distribusi probabilitas ini berdasarkan pengalaman pribadi atau berdasarkan catatan dimasa lampau yang digunakan untuk meramalkan keadaan dimasa yang akan datang. Misalnya seorang dokter puskesmas mempelajari dan mengadakan analisis data penyakit berdasarkan catatan medik yang lalu, kemudian atas dasar tersebut dia membuat ramalan tentang probabilitas distribusi penyakit yang akan datang.

Misalnya pada tahun 2000 jumlah penderita muntaber sebanyak 1000 orang dengan probabilitas 0,6, penyakit kulit 700 orang dengan probabilitas 0,2 dan penderita penyakit mata sebanyak 800 orang dengan probabilitas 0,2.

Data diatas dapat disusun dalam tabel sebagai berikut

Tabel 1.3. distribusi frekuensi jenis penyakit

Jenis penyakit	Jumlah	Probabilitas
Muntaber	1000	0,6
Penyakit kulit	700	0,2
Penyakit mata	800	0,2
Jumlah	2500	1

### 4. Distribusi Probabilitas dan Kurva Normal dengan $\mu$ dan $\sigma$ Berbeda

Bentuk distribusi probabilitas dan kurva normal dengan nilai tengah sama dan standar deviasi yang berbeda, adalah bentuk leptokurtic, platykurtik dan mesokurtik. Kurva normal tersebut mempunyai  $\mu = Md = Mo$  yang sama, namun mempunyai  $\sigma$  berbeda. Semakin besar  $\sigma$ , maka kurva semakin pendek dan semakin tinggi nilai  $\sigma$ , maka semakin runcing. Oleh sebab itu,  $\sigma$  tinggi cenderung menjadi platykurtik dan  $\sigma$  rendah menjadi leptokurtik. Nilai  $\sigma$  yang tinggi menunjukkan bahwa nilai data semakin menyebar dari nilai tengahnya ( $\mu$ ). Apabila  $\sigma$  rendah, maka nilai semakin mengelompok pada nilai tengahnya.

## 5. Distribusi Probabilitas dan Kurva Normal dengan $\mu$ Berbeda dan $\sigma$ Sama

Bentuk distribusi probabilitas dan kurva normal dengan  $\mu$  berbeda dan  $\sigma$  sama mempunyai jarak antara kurva yang berbeda, namun bentuk kurva tetap sama. Hal demikian bisa terjadi karena kemampuan antar populasi berbeda, namun setiap populasi mempunyai keragaman yang hampir sama.

## 6. Distribusi Probabilitas dan Kurva Normal dengan $\mu$ dan $\sigma$ Berbeda

Distribusi kurva normal dengan  $\mu$  dan  $\sigma$  berbeda. Kurva ini mempunyai titik pusat yang berbeda pada sumbu mendatar dan bentuk kurva berbeda karena mempunyai standar deviasi yang berbeda.

## 7. Distribusi Probabilitas Normal Baku

Distribusi normal baku adalah distribusi probabilitas acak normal dengan nilai tengah nol dan simpangan baku 1. Seringkali disebut dengan distribusi z. Hal yang perlu dilakukan dalam rangka distribusi probabilitas normal baku adalah mengubah atau membakukan distribusi aktual dalam bentuk distribusi norma baku yang dikenal dengan nilai Z atau skor Z

Nilai Z adalah jarak yang berbeda antara sebuah nilai X yang dipilih dari rata-rata  $\mu$ , dibagi dengan standar deviasinya,  $\sigma$ .

Rumus nilai Z adalah :

Z = Skor Z atau nilai normal baku

X = Nilai dari suatu pengamatan atau pengukuran

$\mu$  = Nilai rata-rata hitung suatu distribusi

$\sigma$  = Standar deviasi

### Contoh Soal:

Rata-rata berat sebuah kotak adalah 283 gram dan standar deviasinya 1,6 gram. Berapakah probabilitas sebuah kotak dibawah 284,5 gram ?

Transformasi dari X ke Z

Bila nilai X berada di antara  $X = x_1$  dan  $X = x_2$ , maka variabel acak Z akan berada di antara nilai: Transformasi ini juga mempertahankan luas di bawah kurvanya,

### Contoh Soal:

Misalkan kita memilih 20 saham pada bulan Mei 2007. Harga saham ke-20 perusahaan tersebut berkisar antara Rp. 2.000 – 2.805 per lembarnya. Berapa probabilitas harga saham antara Rp. 2.500 sampai 2.805 per lembarnya. Diketahui  $\mu = 2.500$  sebagai nilai rata-rata hitung dan standar deviasinya 400.

$$Z = (X - \mu) / \sigma$$

$$Z1 = (2.500 - 2500) / 400$$

$$Z1 = 0 / 400 = 0$$

$$Z2 = (2.805 - 2.500) / 400$$

$$Z2 = 0.76$$

### C. KARAKTERISTIK KURVA DISTRIBUSI NORMAL

Distribusi probabilitas normal dan kurva normal telah dikembangkan oleh DeMoivre (1733) dan Gauss (1777 – 1855) dengan menurunkan persamaan matematis dan kurva normalnya. Oleh sebab itu, kurva normal sering juga disebut kurva Gauss.

Beberapa karakteristik dari distribusi probabilitas dan kurva normal adalah:

1. Kurva berbentuk genta atau lonceng dan memiliki satu puncak yang terletak di tengah. Nilai rata-rata hitung ( $\mu$ ) = median (Md) = modus (Mo). Nilai  $\mu = Md = Mo$  yang berada di tengah membelah kurva menjadi dua bagian yaitu setengah di bawah nilai  $\mu = Md = Mo$  dan setengah di atas nilai  $\mu = Md = Mo$ .
2. Distribusi probabilitas dan kurva normal berbentuk kurva simetris dengan rata-rata hitungnya ( $\mu$ ).
3. Distribusi probabilitas dan kurva normal bersifat asimptotis.
4. Kurva mencapai puncak pada saat  $X = \mu$ .
5. Luas daerah di bawah kurva normal adalah 1;  $\frac{1}{2}$  di sisi kanan nilai tengah dan  $\frac{1}{2}$  di sisi kiri.
6. Bila X suatu pengubah acak normal dengan nilai tengah  $\mu$ , dan standar deviasi  $\sigma$ ,

### D. PENTINGNYA DISTRIBUSI NORMAL DALAM STATISTIKA

Satu-satunya distribusi probabilitas dengan variabel random kontinu adalah distribusi normal. Ada 2 peran yang penting dari distribusi normal : Memiliki beberapa sifat yang mungkin untuk digunakan sebagai patokan dalam mengambil

suatu kesimpulan berdasarkan hasil sampel yang diperoleh. Pengukuran sampel digunakan untuk menafsirkan parameter populasi.

Distribusi normal sangat sesuai dengan distribusi empiris, sehingga dapat dikatakan bahwa semua kejadian alami akan membentuk distribusi ini. Karena alasan inilah sehingga distribusi ini dikenal sebagai distribusi normal dan grafiknya dikenal sebagai kurva normal atau kurva gauss.

1. Grafiknya selalu berada di atas sumbu x
2. Bentuknya simetris pada  $x = \mu$
3. Mempunyai satu buah modus, yaitu pada  $x = \mu$
4. Luas grafiknya sama dengan satu unit persegi, dengan rincian
5. Kira-kira 68% luasnya berada di antara daerah  $\mu - \sigma$  dan  $\mu + \sigma$
6. Kira-kira 95% luasnya berada di antara daerah  $\mu - 2\sigma$  dan  $\mu + 2\sigma$
7. Kira-kira 99% luasnya berada di antara daerah  $\mu - 3\sigma$  dan  $\mu + 3\sigma$

#### E. CIRI-CIRI DISTRIBUSI NORMAL

Distribusi normal mempunyai beberapa sifat dan ciri, yaitu:

- Disusun dari variable random kontinu
- Kurva distribusi normal mempunyai satu puncak (uni-modal)
- Kurva berbentuk simetris dan menyerupai lonceng hingga mean, median dan modus terletak pada satu titik.
- Kurva normal dibentuk dengan N yang tak terhingga.
- Peristiwa yang dimiliki tetap independen.
- Ekor kurva mendekati absis pada penyimpangan 3 SD ke kanan dan ke kiri dari rata-rata dan ekor grafik dapat dikembangkan sampai tak terhingga tanpa menyentuh sumbu absis.

#### 1. Distribusi normal standar

Suatu distribusi normal tidak hanya memiliki satu kurva, tetapi merupakan kumpulan kurva yang mempunyai ciri-ciri yang sama. sehingga harus ditentukan 1 pegangan sebagai distribusi normal yang standar.

Ada 2 cara untuk menentukan distribusi normal :

##### a) Cara Ordinasi:

Menggunakan rumus distribusi normal berikut :



$\mu$	=	rata-rata
$\sigma$	=	simpang baku
$\pi$	=	3,1416 (bilangan konstan)
$e$	=	2,7183 (bilangan konstan)
$X$	=	absis dengan batas $-\infty < X <$

Bila nilai  $\mu$  dan  $\sigma$  tetap maka setiap nilai  $x$  akan menghasilkan nilai  $y$  sehingga bila nilai  $x$  dimasukkan dalam perhitungan berkali-kali dengan jumlah tak terhingga maka akan dihasilkan suatu kurva distribusi normal. Terdapat banyak kurva normal dengan bentuk yang berlainan, tergantung besar dan kecilnya  $\sigma$ .

Bila  $\sigma$  besar, kurva yang terbentuk mempunyai puncak yang rendah, sebaliknya bila  $\sigma$  kecil akan menghasilkan puncak kurva yang tinggi.

#### b) Cara luas

Kurva normal adalah kurva yang simetris, yang berarti bahwa kurva ini akan membagi luas kurva menjadi 2 bagian yang sama. Seluruh luas kurva = 1 atau 100% dan rata-rata ( $\mu$ ) membagi luas kurva menjadi 2 bagian yang sama. Berarti luas tiap belahan adalah 50%.

Setiap penyimpangan rata-rata dapat ditentukan presentase terhadap seluruh luas kurva. penyimpangan ke kanan dan ke kiri :

- -.penyimpangan 1 SD = 68,2% dari seluruh luas kurva.
- -.penyimpangan 2 SD = 95,5% dari seluruh luas kurva.
- -.penyimpangan 3 SD, = 99,7% dari seluruh luas kurva.

Proses standarisasi dapat dilakukan dengan transformasi rumus (kurva normal standar) :

$$Z = \frac{x - \mu}{\sigma}$$

$\sigma$

$x$  = nilai variable random

$\mu$  = rata-rata distribusi

$\sigma$  = simpang baku

$Z$  = nilai standar, yaitu besarnya penyimpangan suatu nilai terhadap rata-rata yang dinyatakan dari unit SD.

Standarisasi penting dilakukan karena ada variabel random yang memiliki satuan yang berbeda-beda, seperti cm, kg, bulan. Untuk memudahkan perhitungan dapat digunakan sebuah table yang menunjukkan luas area di bawah kurva normal antara nilai rata-rata dan suatu nilai variable random yang dinyatakan dalam unit SD.

Misalnya : luas 95% adalah 1,96 SD.

Untuk transformasi distribusi normal menjadi distribusi normal standar dinyatakan  $\mu = 0$  dan  $\sigma = 1$ .

## F. PENGGUNAAN TABEL DISTRIBUSI NORMAL DAN APLIKASINYA

### 1. Tabel distribusi normal standar terdiri dari kolom dan baris.

Kolom paling kiri menunjukkan nilai Z, tertera angka 0 sampai 3 dengan satu desimal dibelakangnya. Desimal berikutnya terletak pada baris paling atas dengan angka dari 0 sampai 9.

Misalnya dari hasil perhitungan diperoleh nilai  $Z = 1,96$

- Maka di kolom kiri kita cari nilai 1,9 dan baris atas kita cari angka 6
- Dari kolom 6 bergarak ke bawah, hingga pertemuan titik yang menunjukkan angka 0,4750.
- Berarti luas daerah di dalam kurva normal antara rata-rata dengan 1,96 SD ke kanan adalah 0,475.
- Karena luas kurva ke kanan dan ke kiri sama, maka luas penyimpangan 1,96 ke kanan dan ke kiri dari rata-rata adalah 0,95 (95%).

### 2. Aplikasi Distribusi Normal

Sebagai contoh aplikasi distribusi normal, dilakukan suatu evaluasi thd pengobatan TB menggunakan Rifampicin dengan rata-rata kesimpulan 200 hari dan standar deviasinya sebesar 10. Berapakah probabilitas kesembuhan antara 190 dan 210?

Jawab :

Mula-mula dihitung nilai  $Z = 210$

$$Z = (210 - 200) / 10 = 1 = 0,3413$$

jadi probabilitas kesembuhan 190 sampai 210 =  $0,3413 + 0,3413 = 0,6826 = 68,26\%$

## G. LUAS DAERAH DI BAWAH KURVA NORMAL

Kurva setiap distribusi peluang kontinu atau fungsi padat dibuat sedemikian rupa sehingga luas di bawah kurva di antara kedua ordinat  $x = x_1$  dan  $x = x_2$  sama dengan peluang acak  $X$  mendapatkan harga antara  $x = x_1$  dan  $x = x_2$ . Luas di bawah kurva antara dua ordinat sembarang tergantung pada harga  $\mu$  dan  $\sigma$ . Luas antara nilai  $Z$  ( $-1 < Z < 1$ ) sebesar 68,26% dari jumlah data.

Berapa luas antara  $Z$  antara 0 dan sampai  $Z = 0,76$  atau biasa ditulis  $P(0 < Z < 0,76)$ ?

Dapat dicari dari tabel luas di bawah kurva normal. Nilainya dihasilkan = 0,2764

## H. PENERAPAN KURVA NORMAL

### Contoh Soal 1

PT GS mengklaim rata-rata berat buah mangga “B” adalah 350 gram dengan standar deviasi 50 gram. Bila berat mangga mengikuti distribusi normal, berapa probabilitas bahwa berat buah mangga mencapai kurang dari 250 gram, sehingga akan diprotes oleh konsumen.

Jawab:

- Transformasi ke nilai  $z$   
 $AP(x < 250); P(x=250) = (250-350)/50 = -2,00$  Jadi  $P(x < 250) = P(z < -2,00)$
- Lihat pada tabel luas di bawah kurva normal  
 $P(z < -2,00) = 0,4772$
- Luas sebelah kiri nilai tengah adalah 0,5. Oleh sebab itu, nilai daerah yang diarsir menjadi  $0,5 - 0,4772 = 0,0228$ . Jadi probabilitas di bawah 250 gram adalah 0,0228 (2,28%). Dengan kata lain probabilitas konsumen protes karena berat buah mangga kurang dari 250 gram adalah 2,28%.

### Contoh Soal 2

PT Work Electric, memproduksi Bohlam Lampu yang dapat hidup 900 jam dengan standar deviasi 50 jam. PT Work Electric ingin mengetahui berapa persen produksi pada kisaran antara 800-1.000 jam, sebagai bahan promosi bohlam lampu. Hitung berapa probabilitasnya!

Jawab:

- $P(800 < X < 1.000)$ ?
- Hitung nilai Z  
 $Z_1 = (800-900)/50 = -2,00$ ;  
 $Z_2 = (1.000-900)/50 = 2,00$
- Jadi:  $P(800 < X < 1.000) = P(-2,00 < Z < 2,00)$ ;  
 $P(-2,00 < Z) = 0,4772$  dan  $P(Z < 2,00) = 0,4772$
- Sehingga luas daerah yang diarsir adalah  $= 0,4772 + 0,4772 = 0,9544$ . Jadi  $P(800 < X < 1.000) = P(-2,00 < Z < 2,00) = 0,9544$ .
- Jadi 95,44% produksi berada pada kisaran 800-1.000 jam. Jadi jika PT Work Electric mengklaim bahwa lampu bohlamnya menyala 800-1.000 jam, mempunyai probabilitas benar 95,44%, sedang sisanya 4,56% harus dipersiapkan untuk garansi.

#### I. PENDEKATAN NORMAL TERHADAP BINOMIAL

Apabila kita perhatikan suatu distribusi probabilitas binomial, dengan semakin besarnya nilai  $n$ , maka semakin mendekati nilai distribusi normal. Gambar berikut menunjukkan distribusi probabilitas binomial dengan  $n$  yang semakin membesar.

Bila nilai  $X$  adalah distribusi acak binomial dengan nilai tengah  $\mu = np$  dan standar deviasi  $\sigma = \sqrt{npq}$ , maka nilai  $Z$  untuk distribusi normal adalah:

- di mana  $n \gg \gg \gg$  dan nilai  $p$  mendekati  $0,5$

#### J. FAKTOR KOREKSI KONTINUITAS

Nilai koreksi kontinuitas adalah sebesar 0,5 yang dikurangkan dan ditambahkan pada data yang diamati. Untuk mengubah pendekatan dari binomial ke normal, memerlukan faktor koreksi, selain syarat binomial terpenuhi: (a) hanya ada dua peristiwa, (b) peristiwa bersifat independen; (c) besar probabilitas sukses dan gagal sama setiap percobaan, (d) data merupakan hasil penghitungan. Menggunakan faktor koreksi yang besarnya 0,5.

#### Contoh Soal:

Adi merupakan pedagang buah di Tangerang. Setiap hari ia membeli 300 kg buah di Pasar Induk Kramat Jati, Jakarta Timur. Probabilitas buah tersebut laku dijual adalah 80% dan 20% kemungkinan tidak laku dan busuk. Berapa probabilitas buah sebanyak 250 kg laku dan tidak busuk ?

Jawab:

- $n = 300$ ; probabilitas laku  $p = 0.8$ , dan  $q = 1 - 0.8 = 0.2$
- $\mu = np = 300 \times 0.80 = 240$
- $\sigma = \sqrt{npq} = \sqrt{300 \times 0.80 \times 0.20} = 6.93$

Diketahui  $X = 250$ , dan dikurangi faktor koreksi 0.5 sehingga  $X = 250 - 0.5 = 249.5$  Dengan demikian nilai  $Z$  menjadi:

- $Z = (249.5 - 240) / 6.93 = 1.37$  dan  $P(Z < 1.37) = 0.4147$
- Jadi probabilitas laku adalah  $0.5 + 0.4147 = 0.9147$
- Dengan kata lain harapan buah laku 250 kg adalah 91.47%

## K. VARIABEL ACAK / RANDOM

Untuk menggambarkan hasil-hasil percobaan sebagai nilai-nilai numerik secara sederhana, kita menggunakan apa yang disebut sebagai variabel acak. Jadi variabel acak dapat didefinisikan sebagai deskripsi numerik dari hasil percobaan.

Variabel acak biasanya menghubungkan nilai-nilai numerik dengan setiap kemungkinan hasil percobaan. Karena nilai-nilai numerik tersebut dapat bersifat diskrit(hasil perhitungan) dan bersifat kontinu(hasil pengukuran) maka variabel acak dapat dikelompokkan menjadi variabel acak diskrit dan variabel acak kontinu.

### 1. Variabel Acak Diskrit

Variabel acak diskrit adalah variabel acak yang tidak mengambil seluruh nilai yang ada dalam sebuah interval atau variabel yang hanya memiliki nilai tertentu. Nilainya merupakan bilangan bulat dan asli, tidak berbentuk pecahan. Variabel acak diskrit jika digambarkan pada sebuah garis interval, akan berupa sederetan titik-titik yang terpisah.

**Contoh :**

- a) Banyaknya pemunculan sisi muka atau angka dalam pelemparan sebuah koin (uang logam).
- b) Jumlah anak dalam sebuah keluarga.

### 2. Variabel Acak Kontinu

Variabel acak kontinu adalah variabel acak yang mengambil seluruh nilai yang ada dalam sebuah interval atau variabel yang dapat memiliki nilai-nilai pada suatu interval tertentu. Nilainya dapat merupakan bilangan bulat maupun pecahan.

Variabel acak kontinu jika digambarkan pada sebuah garis interval, akan berupa sederetan titik yang bersambung membentuk suatu garis lurus.

**Contoh :**

- a) Usia penduduk suatu daerah.
- b) Panjang beberapa helai kain.

### 3. Distribusi Probabilitas Variabel Acak Diskrit

Distribusi probabilitas variabel acak menggambarkan bagaimana suatu probabilitas didistribusikan terhadap nilai-nilai dari variabel acak tersebut. Untuk variabel diskrit  $X$ , distribusi probabilitas didefinisikan dengan fungsi probabilitas dan dinotasikan sebagai  $p(x)$ .

Fungsi probabilitas  $p(x)$  menyatakan probabilitas untuk setiap nilai variabel acak  $X$ .

**Contoh :**

Jumlah mobil terjual dalam sehari menurut jumlah hari selama 300 hari

Jumlah mobil terjual dalam sehari	Jumlah hari
0	54
1	117
2	72
3	42
4	12
5	3
Total	300

Distribusi Probabilitas Jumlah Mobil Terjual dalam Sehari

$X$	$p(x)$
0	0,18
1	0,39
2	0,24
3	0,14
4	0,04
5	0,01
Total	1,00

Dalam membuat suatu fungsi probabilitas untuk variabel acak diskrit, kondisi berikut harus dipenuhi.

- $p(x) \geq 0$  atau  $0 \leq p(x) \leq 1$
- $\sum p(x) = 1$

Kita juga bisa menyajikan distribusi probabilitas dengan menggunakan grafik.

#### 4. Fungsi Probabilitas Kumulatif Variabel Acak diskrit

Fungsi probabilitas kumulatif digunakan untuk menyatakan jumlah dari seluruh nilai fungsi probabilitas yang lebih kecil atau sama dengan suatu nilai yang ditetapkan.

- Secara matematis, fungsi probabilitas kumulatif dinyatakan sebagai berikut.

$$F(x) = P(X \leq x) = \sum_{X \leq x} p(x)$$

- Dimana

$F(x) = P(X \leq x)$  menyatakan fungsi probabilitas kumulatif pada titik  $X = x$  yang merupakan jumlah dari seluruh nilai fungsi probabilitas untuk nilai  $X$  sama atau kurang dari  $x$ .

##### Contoh :

Probabilitas Kumulatif dari jumlah mobil terjual dalam sehari

X	F(x)
0	0,18
1	0,57 (= 0,18 + 0,39)
2	0,81 (= 0,18 + 0,39 + 0,24)
3	0,95 (= 0,18 + 0,39 + 0,24 + 0,14)
4	0,99 (= 0,18 + 0,39 + 0,24 + 0,14 + 0,04)
5	1,00 (= 0,18 + 0,39 + 0,24 + 0,14 + 0,04 + 0,01)

#### 5. Distribusi Probabilitas Variabel Acak Kontinu

Distribusi probabilitas variabel acak kontinu dinyatakan dengan fungsi  $f(x)$  dan sering disebut sebagai fungsi kepadatan atau fungsi kepadatan probabilitas dan bukan fungsi probabilitas. Nilai  $f(x)$  bisa lebih besar dari 1.

Fungsi kepadatan probabilitas harus memenuhi syarat sebagai berikut.

- $f(x) \geq 0$

b) (integral seluruh fungsi kepadatan probabilitas  $f(x) = 1$ )

- Catatan :  $f(x) dx = P\{x \leq X \leq (x + dx)\}$ , yaitu probabilitas bahwa nilai  $X$  terletak pada interval  $x$  dan  $x + dx$ .

## 6. Fungsi Probabilitas Kumulatif Variabel Acak Kontinu

Kalau pada variabel acak diskrit, fungsi probabilitas kumulatif dihitung dengan cara penjumlahan maka pada variabel acak kontinu, probabilitas kumulatif dicari dengan integral.

Rumusnya adalah sebagai berikut.

- $F(x) = P(X \leq x)$

Nilai-nilai dalam rumus ini harus kontinu atau dalam suatu interval.

## 7. Fungsi Probabilitas Bersama

Bila  $X$  dan  $Y$  adalah dua variabel acak diskrit, distribusi probabilitas bersamanya dapat dinyatakan sebagai sebuah fungsi  $f(x,y)$  bagi sembarang nilai  $(x,y)$  yang dapat diambil oleh peubah acak  $X$  dan  $Y$ . Sehingga dalam rumus variabel acak diskrit.

- $f(x,y) = p(X = x, Y = y)$

Dimana nilai  $f(x,y)$  menyatakan peluang bahwa  $x$  dan  $y$  terjadi secara bersamaan.

- Sedangkan distribusi probabilitas kumulatif bersama  $X$  dan  $Y$  terdiri dari nilai  $(x,y)$  dan  $f(x,y)$  untuk semua  $(X,Y)$
- Kalau dua variabel  $X, Y$  dan  $P(P = x, Y = y) = p(x,y)$  merupakan suatu fungsi yang memenuhi syarat berikut :
- $p(x,y) \geq 0$ , untuk seluruh nilai  $X$  dan  $Y$
- (penjumlahan untuk seluruh nilai  $X$  dan  $Y$ ) maka  $p(x,y)$  disebut fungsi probabilitas bersama  $X$  dan  $Y$ .
- Fungsi  $p(x)$  dan  $q(y)$  yang diperoleh langsung dari  $p(x,y)$  disebut fungsi marjinal.
- Fungsi marjinal  $p(x)$  dan  $q(y)$  dapat dilihat dalam tabel, pada beris dan kolom yang paling akhir (pada tepi tabel, marjin = tepi/pinggir).

## 8. Nilai Harapan Dan Varians Dari Variabel Acak Diskrit

Rata-rata ( $m$ ) dari distribusi probabilitas adalah nilai harapan dari variabel acaknya. Nilai harapan variabel acak diskrit adalah rata-rata tertimbang terhadap



seluruh kemungkinan hasil dimana penimbangannya adalah nilai probabilitas yang dihubungkan dengan setiap hasil.

Nilai harapan diperoleh dengan menyatakan setiap kemungkinan hasil  $x$  dengan probabilitasnya  $P(X)$  dan kemudian menjumlahkan hasil perkalian tersebut. Nilai harapan dari variabel acak diskrit  $X$  yang dinotasikan dengan  $E(X)$  dirumuskan sebagai berikut.

$$= x_1 p(x_1) + x_2 p(x_2) + \dots + x_N p(x_N)$$

dimana.

$x_i$  = nilai ke- $i$  dari variabel acak  $X$

$p(x_i)$  = probabilitas terjadinya  $x_i$

Selain rata-rata, ukuran statistic yang lain adalah varians dan standar deviasi. Varians ( $s^2$ ) dari variabel acak diskrit didefinisikan sebagai berikut. Varians dari variabel acak diskrit adalah rata-rata tertimbang dari kuadrat selisih antara kemungkinan hasil dan rata-ratanya dimana penimbangannya adalah probabilitas dari masing-masing hasil tersebut.

Varians diperoleh dengan mengalikan setiap kemungkinan kuadrat selisih  $(x_i - m)^2$  dengan probabilitasnya  $p(x_i)$  dan kemudian menjumlahkan seluruh hasil perkalian tersebut. Sehingga varians dinyatakan sebagai berikut dimana:

$x_i$  = nilai ke- $i$  dari variable acak  $X$

$p(x_i)$  = probabilitas terjadinya  $x_i$

## 9. Nilai Harapan dari Fungsi Probabilitas Bersama

Jika fungsi probabilitas bersama dinotasikan dengan  $p(x, y)$  untuk variabel acak  $X$  dan  $Y$ , maka nilai harapan dari variabel acak  $h(x, y)$  yang merupakan fungsi dari  $X$  dan  $Y$  adalah sebagai berikut.

- $E[h(x, y)] = \sum \sum h(x, y) p(x, y)$

Dimana.

$h(x, y)$  adalah sembarang fungsi dari  $X$  dan  $Y$

$p(x, y)$  adalah probabilitas terjadinya  $X$  dan  $Y$  secara bersama-sama.

- Kalau  $h(x, y) = xy$ , maka

$$E[h(x, y)] = E(XY) = \sum \sum xy p(x, y)$$

- Kalau  $h(x, y) = x + y$ , maka

$$E[h(x, y)] = e(X + Y) = \sum (x + y) p(x, y)$$

## 10. Aturan-aturan dalam Menghitung Nilai Harapan.

- a.  $E(k) = k$ ,  $k =$  bilangan konstan.
- b. Varians  $(k) = 0$  dan varians  $(X) = s^2$
- c.  $E(kX) = k E(X)$
- d. Varians  $(kX) = k^2 s^2$
- e.  $E(X \pm Y) = E(X) \pm E(Y)$ 
  - i.  $E(\sum X_i) = \sum E(X_i) \quad i = 1, 2, \dots, n$
  - ii.  $E(\sum k_i X_i) = \sum k_i E(X_i) \quad i = 1, 2, \dots, n$

## 11. Kovarians Dan Aplikasinya Dalam Keuangan

Pada sub bab ini, kita pelajari konsep kovarians antara dua variabel dan kegunaannya dalam manajemen portfolio dan keuangan.

### a) Kovarians

Kovarians adalah suatu pengukur yang menyatakan variasi bersama dari dua variable acak. Kovarians antara dua variabel acak diskrit  $X$  dan  $Y$  dinotasikan dengan  $s_{xy}$  dan didefinisikan sebagai berikut dimana

- $X_i =$  nilai variable acak  $X$  ke- $i$
  - $Y_i =$  nilai variable acak  $Y$  ke- $i$
  - $P(x_i, y_i) =$  probabilitas terjadinya  $x_i$  dan  $y_i$
- $i = 1, 2, \dots, N$

## 12. Nilai Harapan dari Penjumlahan Dua Variabel

Nilai harapan dari penjumlahan dua variable acak adalah sama dengan penjumlahan dari nilai harapan masing-masing variabel acak.

$$E(X + Y) = E(X) + E(Y)$$

## 13. Varians dari Penjumlahan Dua Variabel

### a) Portfolio Expected Return dan Portfolio Risk

Setelah kita definisikan kovarians, expected return, dan standar deviasi dari penjumlahan dua variabel acak, kita dapat menerapkan konsep-konsep tersebut pada studi mengenai sekelompok asset yang

merujuk pada apa yang disebut sebagai portfolio. Dengan menanamkan investasi yang disebarkan pada tidak hanya satu perusahaan, investor mengkombinasikan pengembalian dan meminimumkan resiko. Dalam studi portfolio, kita menggunakan penimbang untuk setiap jenis investasi dengan proporsi asset pada investasi tersebut. Hal ini memungkinkan kita untuk menghitung portfolio expected return dan portfolio risk.

Portfolio expected return untuk investasi dua asset sama dengan penimbang bagi asset X dikalikan dengan expected return dari asset X ditambah dengan penimbang bagi asset Y dikalikan dengan expected return asset Y.

$$E(P) = wE(X) + (1 - w) E(Y)$$

Dimana.

$E(P)$  = portfolio expected return

$w$  = proporsi nilai portfolio dari asset X

$(1 - w)$  = proporsi nilai portfolio dari asset Y

$E(X)$  = expected return asset X

$E(Y)$  = expected return asset Y

# BAB 8 HIPOTESIS

---

## A. PENGERTIAN HIPOTESIS

Ketika sedang melihat sebuah drama ataupun reality show di televisi, pernahkah Anda menduga-duga apa yang akan terjadi pada tokoh utama di akhir cerita? Jika pernah, apa dasar yang Anda gunakan untuk membuat dugaan tersebut?

Dalam kehidupan ini ada banyak hal yang membuat kita sering menduga-duga tentang apa yang akan terjadi selanjutnya. Seringkali dugaan-dugaan tersebut muncul karena adanya pengalaman akan hal yang sama atau setidaknya mirip dengan kejadian yang tengah kita hadapi. Dalam ranah penelitian, dugaan-dugaan juga seringkali muncul. Dugaan ini lebih sering disebut dengan istilah hipotesis.

Hipotesis (atau ada pula yang menyebutnya dengan istilah hipotesa) dapat diartikan secara sederhana sebagai dugaan sementara. Hipotesis berasal dari bahasa Yunani *hypo* yang berarti di bawah dan *thesis* yang berarti pendirian, pendapat yang ditegakkan, kepastian. Jika dimaknai secara bebas, maka hipotesis berarti pendapat yang kebenarannya masih diragukan. Untuk bisa memastikan kebenaran dari pendapat tersebut, maka suatu hipotesis harus diuji atau dibuktikan kebenarannya.

Hipotesis berasal dari bahasa Yunani, *Hupo* berarti Lemah atau kurang atau di bawah, *Thesis* berarti teori, proposisi atau pernyataan yang disajikan sebagai bukti. Sehingga dapat diartikan sebagai ***Pernyataan yang masih lemah kebenarannya dan perlu dibuktikan atau dugaan yang sifatnya masih sementara.***

Hipotesis juga dapat diartikan sebagai pernyataan keadaan populasi yang akan diuji kebenarannya menggunakan data/informasi yang dikumpulkan melalui sampel, dan dapat dirumuskan berdasarkan teori, dugaan, pengalaman pribadi/orang lain, kesan umum, kesimpulan yang masih sangat sementara. Untuk membuktikan kebenaran suatu hipotesis, seorang peneliti dapat dengan sengaja menciptakan suatu gejala, yakni melalui percobaan atau penelitian. Jika sebuah hipotesis telah teruji kebenarannya, maka hipotesis akan disebut teori.

Dalam penelitian ada dua jenis hipotesis yang seringkali harus dibuat oleh peneliti, yakni hipotesis penelitian dan hipotesis statistik. Pengujian hipotesis penelitian merujuk pada menguji apakah hipotesis tersebut betul-betul terjadi pada sampel yang diteliti atau tidak. Jika apa yang ada dalam hipotesis benar-benar terjadi, maka hipotesis penelitian terbukti, begitu pun sebaliknya. Sementara itu, pengujian

hipotesis statistik berarti menguji apakah hipotesis penelitian yang telah terbukti atau tidak terbukti berdasarkan data sampel tersebut dapat diberlakukan pada populasi atau tidak.

## B. KONSEP HIPOTESIS

Menurut Kerlinger (1973:18) dan Tuckman (1982:5) mengartikan hipotesis adalah sebagai dugaan terhadap hubungan antara dua variabel atau lebih. Selanjutnya menurut Sudjana (1992:219) mengartikan hipotesis adalah asumsi atau dugaan mengenai suatu hal yang dibuat untuk menjelaskan hal itu yang sering dituntut untuk melakukan pengecekannya. Atas dasar dua definisi di atas, maka dapat disimpulkan bahwa hipotesis adalah jawaban atau dugaan sementara yang harus diuji lagi kebenarannya.

Hipotesis penelitian adalah hipotesis kerja (Hipotesis Alternatif  $H_a$  atau  $H_1$ ) yaitu hipotesis yang dirumuskan untuk menjawab permasalahan dengan menggunakan teori-teori yang ada hubungannya (relevan) dengan masalah penelitian dan belum berdasarkan fakta serta dukungan data yang nyata di lapangan. **Hipotesis alternatif ( $H_a$ ) dirumuskan dengan kalimat positif.** Hipotesis nol adalah pernyataan tidak adanya hubungan, pengaruh, atau perbedaan antara parameter dengan statistik. **Hipotesis Nol ( $H_0$ ) dirumuskan dengan kalimat negatif).** Nilai Hipotesis Nol ( $H_0$ ) harus menyatakan dengan pasti nilai parameter.

Hipotesis Nol ( $H_0$ ) harus menyatakan dengan pasti nilai parameter.

$H_0 \rightarrow$  ditulis dalam bentuk persamaan

Sedangkan Nilai Hipotesis Alternatif ( $H_a$ ) dapat memiliki beberapa kemungkinan.

$H_1 \rightarrow$  ditulis dalam bentuk pertidaksamaan ( $<$ ;  $>$ ;  $\neq$ )

## C. MACAM-MACAM HIPOTESIS

Terdapat tiga macam hipotesis dalam penelitian, yakni hipotesis deskriptif, hipotesis komparatif, dan hipotesis asosiatif. Masing-masing dari hipotesis ini dapat digunakan sesuai dengan bentuk variabel penelitian yang digunakan. Apakah penelitian menggunakan variabel tunggal/ mandiri ataukah variabel jamak? Jika yang digunakan adalah variabel jamak, apa yang ingin diketahui oleh peneliti dalam rumusan masalah?

## 1. Hipotesis Deskriptif

Hipotesis deskriptif dapat didefinisikan sebagai dugaan atau jawaban sementara terhadap masalah deskriptif yang berhubungan dengan variabel tunggal/mandiri.

### **Contoh:**

Seorang peneliti ingin mengetahui apakah bakso di restoran Bakso Idola Malang mengandung boraks atau tidak. Maka peneliti dapat membuat rumusan masalah seperti berikut: Apakah bakso di restoran Bakso Idola Malang mengandung boraks?

Dalam penelitian ini, variabel yang digunakan adalah variabel tunggal yakni bakso di restoran Bakso Idola Malang, maka hipotesis yang digunakan adalah hipotesis deskriptif. Ada dua pilihan hipotesis yang dapat dibuat oleh peneliti sesuai dengan dasar teori yang ia gunakan, yakni:

- Ho : Bakso di restoran Bakso Idola Malang mengandung boraks
- H1 : Bakso di restoran Bakso Idola Malang tidak mengandung boraks

## 2. Hipotesis Komparatif

Hipotesis komparatif dapat didefinisikan sebagai dugaan atau jawaban sementara terhadap rumusan masalah yang mempertanyakan perbandingan (komparasi) antara dua variabel penelitian.

### **Contoh:**

Seorang peneliti hendak mengetahui bagaimana sikap loyal antara pendukung club sepakbola Manchester United jika dibandingkan dengan sikap loyal pendukung club sepakbola Chelsea. Apakah pendukung memiliki tingkat loyalitas yang sama atautkah berbeda.

Maka peneliti dapat membuat rumusan masalah seperti berikut: Apakah pendukung club sepakbola Manchester United dan Chelsea memiliki tingkat loyalitas yang sama?

Dalam penelitian ini, variabel yang digunakan adalah variabel jamak. Variabel pertama adalah loyalitas club sepakbola Manchester United, sedangkan variabel kedua adalah loyalitas club sepakbola Chelsea. Karena rumusan masalah mempertanyakan perihal perbandingan antara dua variabel, maka hipotesis yang

digunakan adalah hipotesis komparatif. Ada dua pilihan hipotesis yang dapat dibuat oleh peneliti sesuai dengan dasar teori yang ia gunakan, yakni:

- Ho: Pendukung club Manchester United memiliki tingkat loyalitas yang sama dengan pendukung club Chelsea
- H1: Pendukung club Manchester United memiliki tingkat loyalitas yang tidak sama (berbeda) dengan pendukung club Chelsea

### 3. Hipotesis Asosiatif

Hipotesis asosiatif dapat didefinisikan sebagai dugaan/jawaban sementara terhadap rumusan masalah yang mempertanyakan hubungan (asosiasi) antara dua variabel penelitian.

#### **Contoh:**

Seorang peneliti ingin mengetahui apakah sinetron berjudul “Anak Jalanan” memengaruhi gaya remaja laki-laki dalam mengendarai motor.

Maka peneliti dapat membuat rumusan masalah seperti berikut: Apakah sinetron berjudul “Anak Jalanan” memengaruhi gaya remaja laki-laki dalam mengendarai motor?

Dalam penelitian ini, variabel yang digunakan adalah variabel jamak. Variabel pertama adalah sinetron berjudul “Anak Jalanan”, sedangkan variabel kedua adalah gaya remaja laki-laki dalam mengendarai motor. Karena rumusan masalah mempertanyakan perihal hubungan antara dua variabel, maka hipotesis yang digunakan adalah hipotesis asosiatif. Ada dua pilihan hipotesis yang dapat dibuat oleh peneliti sesuai dengan dasar teori yang ia gunakan, yakni:

- Ho: Sinetron berjudul “Anak Jalanan” memengaruhi gaya remaja laki-laki dalam mengendarai motor.
- H1: Sinetron berjudul “Anak Jalanan” tidak memengaruhi gaya remaja laki-laki dalam mengendarai motor.

Pengujian hipotesis dapat di bedakan atas beberapa jenis berdasarkan criteria yang menyertainya.

### 4. Berdasarkan Jenis Parameternya

Didasarkan atas jenis parameter yang di gunakan, pengujian hipotesis dapat di bedakan atas tiga jenis, yaitu sebagai berikut .

- a) Pengujian hipotesis tentang rata-rata

Pengujian hipotesis tentang rata-rata adalah pengujian hipotesis mengenai rata-rata populasi yang di dasarkan atas informasi sampelnya.

**Contohnya:**

- Pengujian hipotesis satu rata-rata
- Pengujian hipotesis beda dua rata-rata
- Pengujian hipotesis beda tiga rata-rata

b) Pengujian hipotesis tentang proporsi

Pengujian hipotesis tentang proporsi adalah pengujian hipotesis mengenai proporsi populasi yang di dasarkan atas informasi sampelnya.

**Contohnya:**

- Pengujian hipotesis satu proporsi
- Pengujian hipotesis beda dua proporsi
- Pengujian hipotesis beda tiga proporsi

c) Pengujian hipotesis tentang varians

Pengujian hipotesis tentang varians adalah pengujian hipotesis mengenai rata-rata populasi yang di dasarkan atas informasi sampelnya.

**Contohnya:**

- Pengujian hipotesis tentang satu varians
- Pengujian hipotesis tentang kesamaan dua varians

## 5. Berdasarkan Jumlah Sampelnya

Didasarkan atas ukuran sampelnya, pengujian hipotesis dapat di bedakan atas dua jenis, yaitu sebagai berikut.

a) Pengujian hipotesis sampel besar

Pengujian hipotesis sampel besar adalah pengujian hipotesis yang menggunakan sampel lebih besar dari 30 ( $n > 30$ ).

b) Pengujian hipotesis sampel kecil

Pengujian hipotesis sampel kecil adalah pengujian hipotesis yang menggunakan sampel lebih kecil atau sama dengan 30 ( $n \leq 30$ ).



## 6. Berdasarkan Jenis Distribusinya

Didasarkan atas jenis distribusi yang digunakan, pengujian hipotesis dapat di bedakan atas empat jenis, yaitu sebagai berikut.

### a) Pengujian hipotesis dengan distribusi Z

Pengujian hipotesis dengan distribusi Z adalah pengujian hipotesis yang menggunakan distribusi Z sebagai uji statistik. Tabel pengujiannya disebut tabel normal standard. Hasil uji statistik ini kemudian di bandingkan dengan nilai dalam tabel untuk menerima atau menolak hipotesis nol ( $H_0$ ) yang di kemukakan.

#### **Contohnya :**

- Pengujian hipotesis satu dan beda dua rata-rata sampel besar
- Pengujian satu dan beda dua proporsi

### b) Pengujian hipotesis dengan distribusi t (t-student)

Pengujian hipotesis dengan distribusi t adalah pengujian hipotesis yang menggunakan distribusi t sebagai uji statistik. Tabel pengujiannya disebut tabel t-student. Hasil uji statistik ini kemudian di bandingkan dengan nilai dalam tabel untuk menerima atau menolak hipotesis nol ( $H_0$ ) yang di kemukakan.

#### **Contohnya :**

- Pengujian hipotesis satu rata-rata sampel kecil
- Pengujian hipotesis beda dua rata-rata sampel kecil

### c) Pengujian hipotesis dengan distribusi $\chi^2$ ( kai kuadrat)

Pengujian hipotesis dengan distribusi  $\chi^2$  ( kai kuadrat) adalah pengujian hipotesis yang menggunakan distribusi  $\chi^2$  sebagai uji statistik. Tabel pengujiannya disebut tabel  $\chi^2$ . Hasil uji statistik ini kemudian di bandingkan dengan nilai dalam tabel untuk menerima atau menolak hipotesis nol ( $H_0$ ) yang di kemukakan.

#### **Contohnya :**

- Pengujian hipotesis beda tiga proporsi
- Pengujian Independensi
- Pengujian hipotesis kompatibilitas

d) Pengujian hipotesis dengan distribusi F (F-ratio)

Pengujian hipotesis dengan distribusi F (F-ratio) adalah pengujian hipotesis yang menggunakan distribusi F (F-ratio) sebagai uji statistik. Tabel pengujiannya disebut tabel F. Hasil uji statistik ini kemudian di bandingkan dengan nilai dalam tabel untuk menerima atau menolak hipotesis nol ( $H_0$ ) yang di kemukakan.

**Contohnya :**

- Pengujian hipotesis beda tiga rata-rata
- Pengujian hipotesis kesamaan dua varians

## 7. Berdasarkan Arah atau Bentuk Formulasi Hipotesisnya

Didasarkan atas arah atau bentuk formulasi hipotesisnya, pengujian hipotesis di bedakan atas 3 jenis, yaitu sebagai berikut.

a) Pengujian hipotesis dua pihak (two tail test)

Pengujian hipotesis dua pihak adalah pengujian hipotesis di mana hipotesis nol ( $H_0$ ) berbunyi “sama dengan” dan hipotesis alternatifnya ( $H_1$ ) berbunyi “tidak sama dengan” ( $H_0 =$  dan  $H_1 \neq$ )

b) Pengujian hipotesis pihak kiri atau sisi kiri

Pengujian hipotesis pihak kiri adalah pengujian hipotesis di mana hipotesis nol ( $H_0$ ) berbunyi “sama dengan” atau “lebih besar atau sama dengan” dan hipotesis alternatifnya ( $H_1$ ) berbunyi “lebih kecil” atau “lebih kecil atau sama dengan” ( $H_0 =$  atau  $H_0 \geq$  dan  $H_1 <$  atau  $H_1 \leq$ ). Kalimat “lebih kecil atau sama dengan” sinonim dengan kata “paling sedikit atau paling kecil”.

c) Pengujian hipotesis pihak kanan atau sisi kanan

Pengujian hipotesis pihak kanan adalah pengujian hipotesis di mana hipotesis nol ( $H_0$ ) berbunyi “sama dengan” atau “lebih kecil atau sama dengan” dan hipotesis alternatifnya ( $H_1$ ) berbunyi “lebih besar” atau “lebih besar atau sama dengan” ( $H_0 =$  atau  $H_0 \leq$  dan  $H_1 >$  atau  $H_1 \geq$ ). Kalimat “lebih besar atau sama dengan” sinonim dengan kata “paling banyak atau paling besar”.

## D. PENGUJIAN HIPOTESIS

Pengujian Hipotesis adalah suatu prosedur yang dilakukan dengan tujuan memutuskan apakah *menerima* atau *menolak* hipotesis itu. Dalam pengujian hipotesis, keputusan yang di buat mengandung ketidakpastian, artinya keputusan bias benar atau salah, sehingga menimbulkan risiko. Besar kecilnya risiko dinyatakan dalam bentuk probabilitas. Pengujian hipotesis merupakan bagian terpenting dari statistic inferensi (statistic induktif), karena berdasarkan pengujian tersebut, pembuatan keputusan atau pemecahan persoalan sebagai dasar penelitian lebih lanjut dapat terselesaikan.

Prosedur pengujian hipotesis statistic adalah langkah-langkah yang di pergunakan dalam menyelesaikan pengujian hipotesis tersebut. Berikut ini langkah-langkah pengujian hipotesis statistic adalah sebagai berikut.

### 1. Menentukan Formulasi Hipotesis

Formulasi atau perumusan hipotesis statistic dapat di bedakan atas dua jenis, yaitu sebagai berikut;

a) Hipotesis nol / nihil ( $H_0$ )

Hipotesis nol adalah hipotesis yang dirumuskan sebagai suatu pernyataan yang akan di uji. Hipotesis nol tidak memiliki perbedaan atau perbedaannya nol dengan hipotesis sebenarnya.

b) Hipotesis alternatif/ tandingan ( $H_1 / H_a$ )

Hipotesis alternatif adalah hipotesis yang di rumuskan sebagai lawan atau tandingan dari hipotesis nol. Dalam menyusun hipotesis alternatif, timbul 3 keadaan berikut.

- 1)  $H_1$  menyatakan bahwa harga parameter lebih besar dari pada harga yang di hipotesiskan. Pengujian itu disebut pengujian satu sisi atau satu arah, yaitu pengujian sisi atau arah kanan.
- 2)  $H_1$  menyatakan bahwa harga parameter lebih kecil dari pada harga yang di hipotesiskan. Pengujian itu disebut pengujian satu sisi atau satu arah, yaitu pengujian sisi atau arah kiri.
- 3)  $H_1$  menyatakan bahwa harga parameter tidak sama dengan harga yang di hipotesiskan. Pengujian itu disebut pengujian dua sisi atau dua arah, yaitu pengujian sisi atau arah kanan dan kiri sekaligus.

Secara umum, formulasi hipotesis dapat di tuliskan :

$$\begin{aligned} H_0 : \theta &= \theta_0 \\ H_1 : \theta &> \theta_0 \\ H_1 : \theta &< \theta_0 \\ H_1 : \theta &\neq \theta_0 \end{aligned}$$

Apabila hipotesis nol ( $H_0$ ) diterima (benar) maka hipotesis alternatif ( $H_a$ ) di tolak. Demikian pula sebaliknya, jika hipotesis alternatif ( $H_a$ ) di terima (benar) maka hipotesis nol ( $H_0$ ) ditolak.

## 2. Menentukan Taraf Nyata ( $\alpha$ )

Taraf nyata adalah besarnya batas toleransi dalam menerima kesalahan hasil hipotesis terhadap nilai parameter populasinya. Semakin tinggi taraf nyata yang di gunakan, semakin tinggi pula penolakan hipotesis nol atau hipotesis yang di uji, padahal hipotesis nol benar.

Besaran yang sering di gunakan untuk menentukan taraf nyata dinyatakan dalam %, yaitu: 1% (0,01), 5% (0,05), 10% (0,1), sehingga secara umum taraf nyata di tuliskan sebagai  $\alpha_{0,01}$ ,  $\alpha_{0,05}$ ,  $\alpha_{0,1}$ . Besarnya nilai  $\alpha$  bergantung pada keberanian pembuat keputusan yang dalam hal ini berapa besarnya kesalahan (yang menyebabkan resiko) yang akan di tolerir. Besarnya kesalahan tersebut di sebut sebagai daerah kritis pengujian (critical region of a test) atau daerah penolakan ( region of rejection).

Nilai  $\alpha$  yang dipakai sebagai taraf nyata di gunakan untuk menentukan nilai distribusi yang di gunakan pada pengujian, misalnya distribusi normal (Z), distribusi t, dan distribusi  $X^2$ . Nilai itu sudah di sediakan dalam bentuk tabel di sebut nilai kritis.

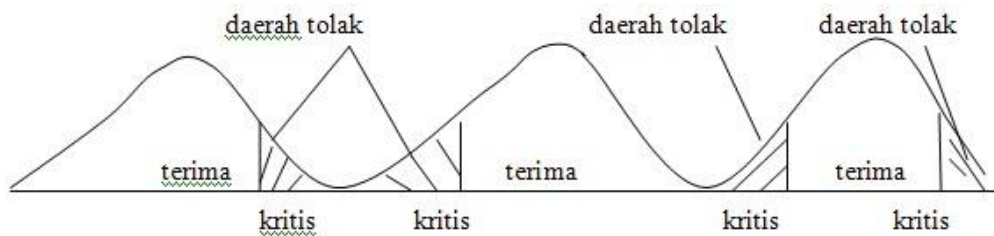
## 3. Menentukan Kriteria Pengujian

Kriteria Pengujian adalah bentuk pembuatan keputusan dalam menerima atau menolak hipotesis nol ( $H_0$ ) dengan cara membandingkan nilai  $\alpha$  tabel distribusinya (nilai kritis) dengan nilai uji statistiknya, sesuai dengan bentuk

pengujiannya. Yang di maksud dengan bentuk pengujian adalah sisi atau arah pengujian.

- a) Penerimaan  $H_0$  terjadi jika nilai uji statistiknya lebih kecil atau lebih besar daripada nilai positif atau negatif dari  $\alpha$  tabel. Atau nilai uji statistik berada di luar nilai kritis.
- b) Penolakan  $H_0$  terjadi jika nilai uji statistiknya lebih besar atau lebih kecil daripada nilai positif atau negatif dari  $\alpha$  tabel. Atau nilai uji statistik berada di luar nilai kritis.

Dalam bentuk gambar, kriteria pengujian seperti gambar di bawah ini



#### 4. Menentukan Nilai Uji Statistik

Uji statistik merupakan rumus-rumus yang berhubungan dengan distribusi tertentu dalam pengujian hipotesis. Uji statistik merupakan perhitungan untuk menduga parameter data sampel yang di ambil secara random dari sebuah populasi. Misalkan, akan di uji parameter populasi ( $P$ ), maka yang pertama-tam di hitung adalah statistik sampel ( $S$ ).

#### 5. Membuat Kesimpulan

Pembuatan kesimpulan merupakan penetapan keputusan dalam hal penerimaan atau penolakan hipotesis nol ( $H_0$ ) yang sesuai dengan kriteria pengujiaanya. Pembuatan kesimpulan dilakukan setelah membandingkan nilai uji statistik dengan nilai  $\alpha$  tabel atau nilai kritis.

- a) Penerimaan  $H_0$  terjadi jika nilai uji statistik berada di luar nilai kritisnya.
- b) Penolakan  $H_0$  terjadi jika nilai uji statistik berada di dalam nilai kritisnya.

- **Kelima langkah pengujian hipotesis tersebut di atas dapat di ringkas seperti berikut.**

1. *Langkah 1* : Menentukan formulasi hipotesis nol ( $H_0$ ) dan hipotesis alternatifnya ( $H_a$ )
2. *Langkah 2* : Memilih suatu taraf nyata ( $\alpha$ ) dan menentukan nilai table.
3. *Langkah 3* : Membuat criteria pengujian berupa penerimaan dan penolakan  $H_0$ .
4. *Langkah 4* : Melakukan uji statistic
5. *Langkah 5* : Membuat kesimpulannya dalam hal penerimaan dan penolakan  $H_0$ .

#### E. JENIS KESALAHAN DALAM HIPOTESIS

Pada dasarnya menguji hipotesis adalah menaksir parameter populasi berdasarkan data sampel. Menurut Sugiyono (2008: 224-225) menyatakan bahwa terdapat dua cara menaksir, yaitu: a point estimate dan interval estimate atau sering disebut confidence interval. A point estimate (titik taksiran) adalah suatu taksiran parameter populasi berdasarkan satu nilai data sampel. Sedangkan interval estimate (taksiran interval) adalah suatu taksiran parameter populasi berdasarkan nilai interval data sampel. Sebagai contoh, saya berhipotesis (menaksir) bahwa daya tahan belajar siswa Indonesia itu 10 jam/hari. Hipotesis ini disebut point estimate, karena daya tahan belajar siswa Indonesia ditaksir melalui satu nilai yaitu 10 jam/hari. Bila hipotesisnya berbunyi daya tahan belajar siswa Indonesia antara 8 sampai dengan 12 jam/hari, maka hal ini disebut interval estimate. Nilai intervalnya adalah 8 sampai dengan 12 jam.

Dalam menaksir parameter populasi berdasarkan data sampel, kemungkinan akan terdapat dua kesalahan, yaitu :

1. Kesalahan Tipe I adalah kesalahan dalam menolak hipotesis  $H_0$  yang benar (seharusnya diterima). Dalam hal ini tingkat kesalahan.
2. Kesalahan Tipe II adalah kesalahan dalam menerima hipotesis yang salah (seharusnya ditolak).

Berikut dapat dilihat tabel hubungan antara keputusan menolak atau menerima  $H_0$  :

Tabel 1. Hubungan Antara Keputusan Menolak dan Menerima Hipotesis

Keputusan	Keadaan Sebenarnya	
	Hipotesis Benar	Hipotesis Salah
Terima Hipotesis	Tidak membuat kesalahan	Kesalahan tipe II ( $\beta$ )
Tolak Hipotesis	Kesalahan Tipe I ( $\alpha$ )	Tidak membuat kesalahan

1. Keputusan menerima hipotesis  $H_0$  yang benar, berarti tidak terjadi kesalahan.
2. Keputusan menerima hipotesis  $H_0$  yang salah, berarti terjadi kesalahan tipe II (Beta).
3. Keputusan menolak hipotesis  $H_0$  yang benar, berarti terjadi kesalahan tipe I (Alpha).
4. Keputusan menolak hipotesis  $H_0$  yang salah, berarti tidak terjadi kesalahan.

Tingkat kesalahan ini selanjutnya dinamakan tingkat signifikansi / taraf signifikansi / *level of significant* (lihat Taraf Signifikant). Dalam prakteknya tingkat signifikansi telah ditetapkan oleh peneliti terlebih dahulu. Dalam pengujian hipotesis kebanyakan digunakan kesalahan tipe I yaitu berapa persen kesalahan untuk menolak hipotesis nol yang benar (biasa menggunakan nilai Alpha).

Dalam pengujian hipotesis kebanyakan digunakan kesalahan tipe I yaitu berapa persen kesalahan untuk menolak hipotesis nol ( $H_0$ ) yang benar (yang seharusnya diterima). Prinsip pengujian hipotesis yang baik adalah meminimalkan nilai  $\alpha$  dan  $\beta$ . Dalam perhitungan, nilai  $\alpha$  dapat dihitung sedangkan nilai  $\beta$  hanya bisa dihitung jika nilai hipotesis alternatif sangat spesifik. Pada pengujian hipotesis, kita lebih sering berhubungan dengan nilai  $\alpha$ . Dengan asumsi, nilai  $\alpha$  yang kecil juga mencerminkan nilai  $\beta$  yang juga kecil. Menurut Furqon (2004:167), kedua tipe kekeliruan tersebut berhubungan negatif (berlawanan arah). Para peneliti biasanya, secara konservatif menetapkan sekecil mungkin (0,05 atau 0,01) sehingga meminimalkan peluang kekeliruan tipe I. Dalam hal ini, mereka beranggapan bahwa menolak hipotesis nol yang seharusnya diterima merupakan kekeliruan yang serius mengingat akibat yang ditimbulkannya. Namun perlu diingat dalam menetapkan taraf signifikansi kita harus melihat situasi penelitian.

Apabila kedua jenis kesalahan tersebut dinyatakan dalam bentuk probabilitas didapatkan hal-hal berikut :

1. Kesalahan jenis I disebut kesalahan  $\alpha$  yang dalam bentuk penggunaannya disebut sebagai taraf nyata atau taraf signifikan (level of significant).  $1 - \alpha$  disebut sebagai tingkat keyakinan (level of confidence), karena dengan itu kita yakin bahwa kesimpulan yang kita buat adalah benar, sebesar  $1 - \alpha$ .
2. Kesalahan jenis II disebut kesalahan  $\beta$  yang dalam bentuk penggunaannya disebut sebagai fungsi ciri operasi (operating characteristic function).  $1 - \beta$  disebut sebagai kuasa pengujian karena memperlihatkan kuasa terhadap pengujian yang dilakukan untuk menolak hipotesis yang seharusnya ditolak.

### Hubungan $\alpha$ , $\beta$ , dan $n$

Antara kedua jenis kesalahan, yaitu kesalahan  $\alpha$  dan  $\beta$  saling berkaitan. Jika kesalahan  $\alpha$  kecil, maka kesalahan  $\beta$ , demikian pula sebaliknya. Untuk membuat suatu kesimpulan yang baik, maka kedua kesalahan tersebut harus dibuat seminimal mungkin. Hal ini biasanya dilakukan melalui caracara seperti berikut :

1. Memperbesar ukuran sampel ( $n$ ) yang akan menjadikan rata-rata ukuran sampel, mendekati ukuran populasinya. Dengan makin besarnya sampel ( $\alpha$  tetap), akan memperkecil  $\beta$  dan memperbesar  $1 - \beta$ , sehingga akan makin besar probabilitas untuk menolak hipotesis ( $H_0$ ) yang salah.
2. Menentukan terlebih dahulu taraf nyata ( $\alpha$ ).

### Contoh Soal :

Berdasarkan pengalaman masa lalu, tinggi badan calon mahasiswa sebuah akademi didistribusikan secara normal dengan rata-rata 160 cm dan simpangan baku 20 cm. Instruktur ingin menguji pada taraf nyata 5%, apakah rata-rata tinggi calon mahasiswa tahun ini di atas 160 cm. Untuk melakukan itu, dipilih sampel sebanyak 36 calon mahasiswa dan diperoleh rata-rata tinggi badan 163 cm. Berapakah nilai  $\beta$  dan  $1 - \beta$  tersebut ?

### Penyelesaian :

Dari soal, diperoleh nilai kritis  $Z_{0,05} = 1,64$  (Lihat Tabel Z)

<b>Z</b>	0,00	0,01	...	0,04
0,0	0,0000	0,0040	...	0,0160
0,1	0,0398	0,0438	...	0,0557
...	...	...	...	...
1,6	0,4452	0,4436	...	0,4495



## F. PERUMUSAN HIPOTESIS

Setelah mengetahui pengertian hipotesis, jenis-jenis hipotesis, dan ciri-ciri hipotesis yang baik, sekarang saatnya kita belajar untuk membuat hipotesis. Untuk menghasilkan sebuah hipotesis, tentunya kita harus mengikuti langkah-langkah tertentu. Dengan langkah dan cara yang benar, sebuah hipotesis yang baik akan memudahkan jalannya proses penelitian.

Awal terbentuknya hipotesis dalam sebuah penelitian biasanya diawali atas dasar terkaan atau *conjecture* peneliti. Meskipun hipotesis berasal dari terkaan, namun sebuah hipotesis tetap harus dibuat berdasarkan pada sebuah acuan, yakni teori dan fakta ilmiah.

### 1. Teori Sebagai Acuan Perumusan Hipotesis

Untuk memudahkan proses pembentukan hipotesis, seorang peneliti biasanya menurunkan sebuah teori menjadi sejumlah asumsi dan postulat. Asumsi-asumsi tersebut dapat didefinisikan sebagai anggapan atau dugaan yang mendasari hipotesis. Berbeda dengan asumsi, hipotesis yang telah diuji dengan menggunakan data melalui proses penelitian adalah dasar untuk memperoleh kesimpulan.

### 2. Fakta Ilmiah Sebagai Acuan Perumusan Hipotesis

Selain menggunakan teori sebagai acuan, dalam merumuskan hipotesis dapat pula menggunakan acuan fakta. Secara umum, fakta dapat didefinisikan sebagai kebenaran yang dapat diterima oleh nalar dan sesuai dengan kenyataan yang dapat dikenali dengan panca indera.

Fakta Ilmiah sebagai acuan perumusan hipotesis dapat diperoleh dengan berbagai cara, misalnya :

- a) Memperoleh dari sumber aslinya
- b) Fakta yang diidentifikasi dengan cara menggambarkan dan menafsirkannya dari sumber yang asli.
- c) Fakta yang diperoleh dari orang mengidentifikasi dengan jalan menyusunnya dalam bentuk *abstract reasoning* (penalaran abstrak).
- d) Selain teori dan fakta ilmiah, hipotesis dapat pula dirumuskan berdasarkan beberapa sumber lain, yakni:
- e) Kebudayaan dimana ilmu atau teori yang relevan dibentuk

- f) Ilmu yang menghasilkan teori yang relevan
- g) Analogi
- h) Reaksi individu terhadap sesuatu dan pengalaman

## G. PENGUJIAN HIPOTESA RATA-RATA

### 1. Pengujian Hipotesis Satu Rata-Rata

#### a) Sampel besar ( $n > 30$ )

Untuk pengujian hipotesis satu rata-rata dengan sample besar ( $n > 30$ ), uji statistiknya menggunakan distribusi Z. Prosedur pengujian hipotesisnya adalah sebagai berikut.

##### 1) Formulasi hipotesis

- $H_0 : \mu = \mu_0$   
 $H_1 : \mu > \mu_0$
- $H_0 : \mu = \mu_0$   
 $H_1 : \mu < \mu_0$
- $H_0 : \mu = \mu_0$   
 $H_1 : \mu \neq \mu_0$

##### 2) Penentuan nilai $\alpha$ (taraf nyata) dan nilai Z table ( $Z_\alpha$ )

Menentukan nilai  $\alpha$  sesuai soal, kemudian nilai  $Z_\alpha$  atau  $Z_{\alpha/2}$  ditentukan dari tabel.

##### 3) Kriteria Pengujian

- Untuk  $H_0 : \mu = \mu_0$  dan  $H_1 : \mu > \mu_0$   
 $H_0$  di terima jika  $Z_o \leq Z_\alpha$   
 $H_0$  di tolak jika  $Z_o > Z_\alpha$
- Untuk  $H_0 : \mu = \mu_0$  dan  $H_1 : \mu < \mu_0$   
 $H_0$  di terima jika  $Z_o \geq -Z_\alpha$   
 $H_0$  di tolak jika  $Z_o < -Z_\alpha$
- Untuk  $H_0 : \mu = \mu_0$  dan  $H_1 : \mu \neq \mu_0$   
 $H_0$  di terima jika  $-Z_{\alpha/2} \leq Z_o \leq Z_{\alpha/2}$   
 $H_0$  di tolak jika  $Z_o > Z_{\alpha/2}$  atau  $Z_o < -Z_{\alpha/2}$

##### 4) Uji Statistik

- a) Simpangan baku populasi ( $\sigma$ ) di ketahui :

$$Z_0 = \frac{\bar{X} - \mu_0}{\sigma_{\bar{x}}} = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

b) Simpangan baku populasi (  $\sigma$  ) tidak di ketahui :

$$Z_0 = \frac{\bar{X} - \mu_0}{s_{\bar{x}}} = \frac{\bar{X} - \mu_0}{\frac{s}{\sqrt{n}}}$$

#### 5) Kesimpulan

Menyimpulkan tentang penerimaan atau penolakan  $H_0$  (sesuai dengan kriteria pengujiannya).

- Jika  $H_0$  diterima maka  $H_1$  di tolak
- Jika  $H_0$  di tolak maka  $H_1$  di terima

#### Contoh Soal :

Suatu pabrik susu merek Good Milk melakukan pengecekan terhadap produk mereka, apakah rata-rata berat bersih satu kaleng susu bubuk yang di produksi dan di pasarkan masih tetap 400 gram atau sudah lebih kecil dari itu. Dari data sebelumnya di ketahui bahwa simpangan baku bersih per kaleng sama dengan 125 gram. Dari sample 50 kaleng yang di teliti, di peroleh rata-rata berat bersih 375 gram. Dapatkah di terima bahwa berat bersih rata-rata yang di pasarkan tetap 400 gram? Ujilah dengan taraf nyata 5 % !

#### Penyelesaian :

Diketahui :

$$n = 50, \bar{X} = 375, \sigma = 125, \mu_0 = 400$$

Jawab :

1) Formulasi hipotesisnya :

$$H_0 : \mu = 400$$

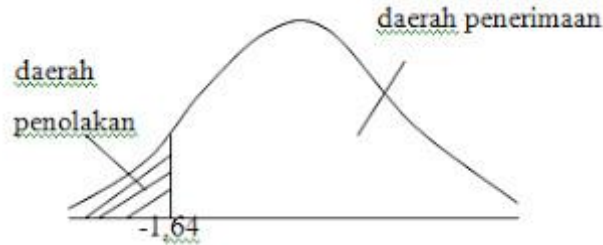
$$H_1 : \mu < 400$$

2) Taraf nyata dan nilai tabelnya :

$$\alpha = 5\% = 0,05$$

$$Z_{0,05} = -1,64 \text{ (pengujian sisi kiri)}$$

3) Kriteria pengujian :



- $H_0$  di terima jika  $Z_0 \geq -1,64$
- $H_0$  di tolak jika  $Z_0 < -1,64$

4) Uji Statistik

$$\begin{aligned} Z_0 &= \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \\ &= \frac{375 - 400}{\frac{125}{\sqrt{50}}} = -1,41 \end{aligned}$$

5) Kesimpulan

Karena  $Z_0 = -1,41 \geq -Z_{0,05} = -1,64$  maka  $H_0$  di terima. Jadi, berat bersih rata-rata susu bubuk merek GOOD MILK per kaleng yang di pasaran sama dengan 400 gram

### b) Sampel Kecil ( $n \leq 30$ )

Untuk pengujian hipotesis satu rata-rata dengan sampel kecil ( $n \leq 30$ ), uji statistiknya menggunakan distribusi t. Prosedur pengujian hipotesisnya adalah sebagai berikut.

1) Formulasi hipotesis

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu > \mu_0$$

2)  $H_0 : \mu = \mu_0$

$$H_1 : \mu < \mu_0$$

3)  $H_0 : \mu = \mu_0$

$H_1 : \mu \neq \mu_0$

4) Penentuan nilai  $\alpha$  (taraf nyata) dan nilai t- tabel

Menentukan nilai  $\alpha$  sesuai soal, kemudian menentukan derajat bebas, yaitu  $db = n - 1$ , lalu menentukan nilai  $t_{\alpha;n-1}$  atau  $t_{\alpha/2;n-1}$  ditentukan dari tabel.

5) Kriteria Pengujian

- Untuk  $H_0 : \mu = \mu_0$  dan  $H_1 : \mu > \mu_0$

$H_0$  di terima jika  $t_0 \leq t_\alpha$

$H_0$  di tolak jika  $t_0 > t_\alpha$

- Untuk  $H_0 : \mu = \mu_0$  dan  $H_1 : \mu < \mu_0$

$H_0$  di terima jika  $t_0 \geq - t_\alpha$

$H_0$  di tolak jika  $t_0 < - t_\alpha$

- Untuk  $H_0 : \mu = \mu_0$  dan  $H_1 : \mu \neq \mu_0$

$H_0$  di terima jika  $- t_{\alpha/2} \leq t_0 \leq t_{\alpha/2}$

$H_0$  di tolak jika  $t_0 > t_{\alpha/2}$  atau  $t_0 < - t_{\alpha/2}$

6) Uji Statistik

a) Simpangan baku populasi ( $\sigma$ ) di ketahui :

$$t_0 = \frac{\bar{X} - \mu_0}{\sigma_{\bar{x}}} = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

b) Simpangan baku populasi ( $\sigma$ ) tidak di ketahui :

$$t_0 = \frac{\bar{X} - \mu_0}{s_{\bar{x}}} = \frac{\bar{X} - \mu_0}{\frac{s}{\sqrt{n}}}$$

6) Kesimpulan

Menyimpulkan tentang penerimaan atau penolakan  $H_0$  (sesuai dengan criteria pengujiannya).

- Jika  $H_0$  diterima maka  $H_1$  di tolak
- Jika  $H_0$  di tolak maka  $H_1$  di terima

**Contoh soal :**

Sebuah sample terdiri atas 15 kaleng susu, memiliki isi berat kotor seperti yang di berikan berikut ini. (Isi berat kotor dalam kg/kaleng)

1,21	1,21	1,23	1,20	1,21
1,24	1,22	1,24	1,21	1,19
1,19	1,18	1,19	1,23	1,18

Jika di gunakan taraf nyata 1%, dapatkah kita menyakini bahwa populasi cat dalam kaleng rata-rata memiliki berat kotor 1,2 kg/kaleng ? (dengan alternatif tidak sama dengan). Berikan evaluasi anda !

**Penyelesaian :**

Diketahui :

$$n = 15, \alpha = 1\%, \mu_0 = 1,2$$

Jawab:

$$\sum X = 18,13$$

$$\sum X^2 = 21,9189$$

$$\begin{aligned} \bar{X} &= 18,13 / 15 \\ &= 1,208 \end{aligned}$$

$$\begin{aligned} s &= \sqrt{\frac{21,9189}{14} - \frac{(18,13)^2}{210}} \\ &= 0,02 \end{aligned}$$

1) Formulasi hipotesisnya :

$$H_0 : \mu = 1,2$$

$$H_1 : \mu \neq 1,2$$

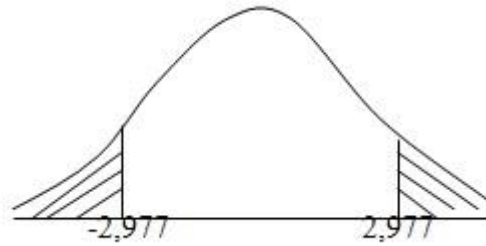
2) Taraf nyata dan nilai tabelnya :

$$\alpha = 1\% = 0,01$$

$$t_{\alpha/2} = 0,005 \text{ dengan db} = 15-1 = 14$$

$$t_{0,005;14} = 2,977$$

3) Kriteria pengujian :



$H_0$  di terima apabila :  $- 2,977 \leq t_0 \leq - 2,977$

$H_0$  di tolak :  $t_0 > 2,977$  atau  $t_0 < - 2,977$

#### 4) Uji Statistik

$$t_0 = \frac{\bar{X} - \mu_0}{\frac{s}{\sqrt{n}}}$$

$$= \frac{1,208 - 1,2}{\frac{0,02}{\sqrt{15}}} = 1,52$$

#### 5) Kesimpulan

Karena  $-t_{0,005;14} = -2,977 \leq t_0 = 1,52 \leq t_{0,005;14} = 2,977$  maka  $H_0$  di terima.  
Jadi, populasi susu dalam kaleng secara rata-rata berisi berat kotor 1,2 kg/kaleng.

## 2. Pengujian Hipotesis Beda Dua Rata-Rata

### a) Sampel besar ( $n > 30$ )

Untuk pengujian hipotesis beda dua rata-rata dengan sampel besar ( $n > 30$ ), uji statistiknya menggunakan distribusi Z. Prosedur pengujian hipotesisnya adalah sebagai berikut.

#### 1) Formulasi hipotesis

- $H_0 : \mu = \mu_0$   
 $H_1 : \mu > \mu_0$
- $H_0 : \mu = \mu_0$   
 $H_1 : \mu < \mu_0$
- $H_0 : \mu = \mu_0$

$$H_1 : \mu \neq \mu_0$$

2) Penentuan nilai  $\alpha$  (taraf nyata) dan nilai Z tabel ( $Z_\alpha$ )

Mengambil nilai  $\alpha$  sesuai soal, kemudian nilai  $Z_\alpha$  atau  $Z_{\alpha/2}$  ditentukan dari tabel.

3) Kriteria Pengujian

- Untuk  $H_0 : \mu_1 = \mu_2$  dan  $H_1 : \mu_1 > \mu_2$

$H_0$  di terima jika  $Z_0 \leq Z_\alpha$

$H_0$  di tolak jika  $Z_0 > Z_\alpha$

- Untuk  $H_0 : \mu_1 = \mu_2$  dan  $H_1 : \mu_1 < \mu_2$

$H_0$  di terima jika  $Z_0 \geq -Z_\alpha$

$H_0$  di tolak jika  $Z_0 < -Z_\alpha$

- Untuk  $H_0 : \mu_1 = \mu_2$  dan  $H_1 : \mu_1 \neq \mu_2$

$H_0$  di terima jika  $-Z_{\alpha/2} \leq Z_0 \leq Z_{\alpha/2}$

$H_0$  di tolak jika  $Z_0 > Z_{\alpha/2}$  atau  $Z_0 < -Z_{\alpha/2}$

4) Uji Statistik

- Simpangan baku populasi ( $\sigma$ ) di ketahui :

$$Z_0 = \frac{\bar{X}_1 - \bar{X}_2}{\sigma_{\bar{x}_1 - \bar{x}_2}} \quad \text{dengan} \quad \sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

- Simpangan baku populasi ( $\sigma$ ) tidak di ketahui :

$$Z_0 = \frac{\bar{X}_1 - \bar{X}_2}{s_{\bar{x}_1 - \bar{x}_2}} \quad \text{dengan} \quad \sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

5) Kesimpulan

Menyimpulkan tentang penerimaan atau penolakan  $H_0$  (sesuai dengan kriteria pengujiannya).

- Jika  $H_0$  diterima maka  $H_1$  di tolak
- Jika  $H_0$  di tolak maka  $H_1$  di terima



**Contoh Soal :**

Seseorang berpendapat bahwa rata-rata jam kerja buruh di daerah A dan B sama dengan alternatif A lebih besar dari pada B. Untuk itu, di ambil sample di kedua daerah, masing-masing 100 dan 70 dengan rata-rata dan simpangan baku 38 dan 9 jam per minggu serta 35 dan 7 jam per minggu. Ujilah pendapat tersebut dengan taraf nyata 5% ! Untuk Varians/ simpangan baku kedua populasi sama besar !

**Penyelesaian :**

Diketahui :

$n_1 = 100$	$X_1 = 38$	$s_1 = 9$
$n_2 = 70$	$X_2 = 35$	$s_2 = 7$

Jawab:

1) Formulasi hipotesisnya :

$$H_0 : \mu_1 = \mu_2$$

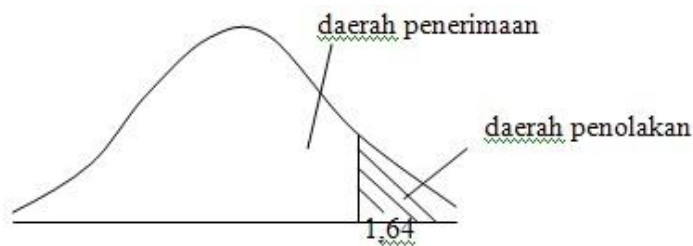
$$H_1 : \mu_1 > \mu_2$$

2) Taraf nyata dan nilai tabelnya :

$$\alpha = 5\% = 0,05$$

$$Z_{0,05} = 1,64 \text{ (pengujian sisi kanan)}$$

3) Kriteria pengujian :



- $H_0$  di terima jika  $Z_o \leq 1,64$
- $H_0$  di tolak jika  $Z_o > 1,64$

4) Uji Statistik

$$\begin{aligned} \bullet s &= \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \\ &= \sqrt{\frac{9_1^2}{100} + \frac{7_2^2}{70}} \\ &= \underline{1,23} \end{aligned}$$

$$\begin{aligned} \bullet Z_0 &= \frac{\bar{X}^1 - \bar{X}^2}{s_{\bar{X}^1 - \bar{X}^2}} \\ &= \frac{38 - 35}{1,23} \\ &= \underline{2,44} \end{aligned}$$

5) Kesimpulan

Karena  $Z_0 = 2,44 > Z_{0,05} = 1,64$  maka  $H_0$  di tolak. Jadi, rata-rata jam kerja buruh di daerah A dan daerah B adalah tidak sama.

**b) Sampel kecil (  $n \leq 30$  )**

Untuk pengujian hipotesis beda dua rata-rata dengan sampel kecil ( $n \leq 30$ ), uji statistiknya menggunakan distribusi t. Prosedur pengujian hipotesisnya adalah sebagai berikut.

1) Formulasi hipotesis

- $H_0 : \mu_1 = \mu_2$

- $H_1 : \mu_1 > \mu_2$

- $H_0 : \mu_1 = \mu_2$

- $H_1 : \mu_1 < \mu_2$

- $H_0 : \mu_1 = \mu_2$

- $H_1 : \mu_1 \neq \mu_2$

2) Penentuan nilai  $\alpha$  (taraf nyata) dan nilai t tabel ( $t_\alpha$ )

Mengambil nilai  $\alpha$  sesuai soal, kemudian nilai  $t_\alpha$  atau  $t_{\alpha/2}$  ditentukan dari tabel.

3) Kriteria Pengujian

- Untuk  $H_0 : \mu_1 = \mu_2$  dan  $H_1 : \mu_1 > \mu_2$   
 $H_0$  di terima jika  $t_0 \leq t_\alpha$   
 $H_0$  di tolak jika  $t_0 > t_\alpha$
- Untuk  $H_0 : \mu_1 = \mu_2$  dan  $H_1 : \mu_1 < \mu_2$   
 $H_0$  di terima jika  $t_0 \geq t_\alpha$   
 $H_0$  di tolak jika  $Z_0 < -t_\alpha$
- Untuk  $H_0 : \mu_1 = \mu_2$  dan  $H_1 : \mu_1 \neq \mu_2$   
 $H_0$  di terima jika  $-t_{\alpha/2} \leq t_0 \leq t_{\alpha/2}$   
 $H_0$  di tolak jika  $t_0 > t_{\alpha/2}$  atau  $t_0 < -t_{\alpha/2}$

#### 4) Uji Statistik

a. Uji pengamatan tidak berpasangan :

$$t_0 = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{(n_1-1)s^2_1 + (n_2-1)s^2_2}{n_1+n_2-2} \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$db = n_1 + n_2 - 2$$

b. Uji pengamatan berpasangan :

$$t_0 = \frac{\bar{d}}{\frac{s_d}{\sqrt{n}}}$$

Keterangan :

$d$  = rata-rata dari nilai  $d$

$s_d$  = simpangan baku dari nilai  $d$

$n$  = banyaknya pasangan

$db = n-1$

#### 5) Kesimpulan

Menyimpulkan tentang penerimaan atau penolakan  $H_0$  (sesuai dengan kriteria pengujiannya).

- Jika  $H_0$  diterima maka  $H_1$  di tolak
- Jika  $H_0$  di tolak maka  $H_1$  di terima

**Contoh Soal :**

Sebuah perusahaan mengadakan pelatihan teknik pemasaran. Sampel sebanyak 12 orang dengan metode biasa dan 10 orang dengan terprogram. Pada akhir pelatihan di berikan evaluasi dengan materi yang sama. Kelas pertama mencapai nilai rata-rata 75 dengan simpangan baku 4,5. Ujilah hipotesis kedua metode pelatihan, dengan alternative keduanya tidak sama! Gunakan taraf nyata 10%! Asumsikan kedua populasi menghampiri distribusi normal dengan varians yang sama!

**Penyelesaian :**

Diketahui :

$$\begin{array}{lll} n_1 = 12 & X_1 = 80 & s_1 = 4 \\ n_2 = 10 & X_2 = 75 & s_2 = 4,5 \end{array}$$

Jawab:

1) Formulasi hipotesisnya :

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

2) Taraf nyata dan nilai tabelnya :

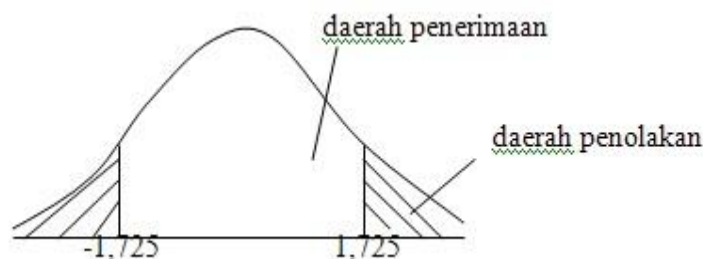
$$\alpha = 10\% = 0,10$$

$$t_{\alpha/2} = 0,05$$

$$db = 12 + 10 - 2 = 20$$

$$t_{0,05;20} = 1,725$$

3) Kriteria pengujian



- $H_0$  di terima apabila  $-1,725 \leq t_0 \leq 1,725$
- $H_0$  di tolak apabila  $t_0 > 1,725$  atau  $t_0 < -1,725$

4) Uji Statistik

$$\begin{aligned}
 t_0 &= \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \\
 &= \frac{80 - 75}{\sqrt{\frac{(12-1)4^2 + (10-1)4,5^2}{12+10-2} \left(\frac{1}{12} + \frac{1}{10}\right)}} \\
 &= 2,76
 \end{aligned}$$

5) Kesimpulan

Karena  $t_0 = 2,76 > t_{0,05;20} = 1,725$  maka  $H_0$  di tolak. Jadi, kedua metode yang digunakan dalam pelatihan tidak sama hasilnya.

**Contoh Soal :**

Untuk mengetahui apakah keanggotaan dalam organisasi mahasiswa memiliki akibat baik atau buruk terhadap prestasi akademik seseorang, diadakan penelitian mengenai mutu rata-rata prestasi akademik. Berikut ini data selama periode 5 tahun.

	Tahun				
	1	2	3	4	5
<b>Anggota</b>	7,0	7,0	7,3	7,1	7,4
<b>Bukan Anggota</b>	7,2	6,9	7,5	7,3	7,4

Ujilah pada taraf nyata 1% apakah keanggotaan dalam organisasi mahasiswa berakibat buruk pada prestasi akademiknya dengan asumsi bahwa populasinya normal !

**Penyelesaian :**

1) Formulasi hipotesisnya :

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 < \mu_2$$

2) Taraf nyata dan nilai tabelnya :

$$\alpha = 1\% = 0,01$$

$$t_{\alpha/2} = 0,05$$

$$db = 5 - 1 = 4$$

$$t_{0,01;4} = -3,747$$

3) Kriteria pengujian :

$H_0$  di terima apabila  $t_0 \geq -3,747$

$H_0$  di tolak apabila  $t_0 < -3,747$

4) Uji Statistik :

Anggota	Bukan Anggota	d	d <sup>2</sup>
7,0	7,2	-0,2	0,04
7,0	6,9	0,1	0,01
7,3	7,5	-0,2	0,04
7,1	7,3	-0,2	0,04
7,4	7,4	0,0	0,00
Jumlah		-0,5	0,13

$$\bar{d} = \frac{-0,5}{5} = -0,1$$

$$S_d^2 = \frac{0,13}{4} - \frac{(-0,5)^2}{20} = 0,02$$

$$S_d = 0,14$$

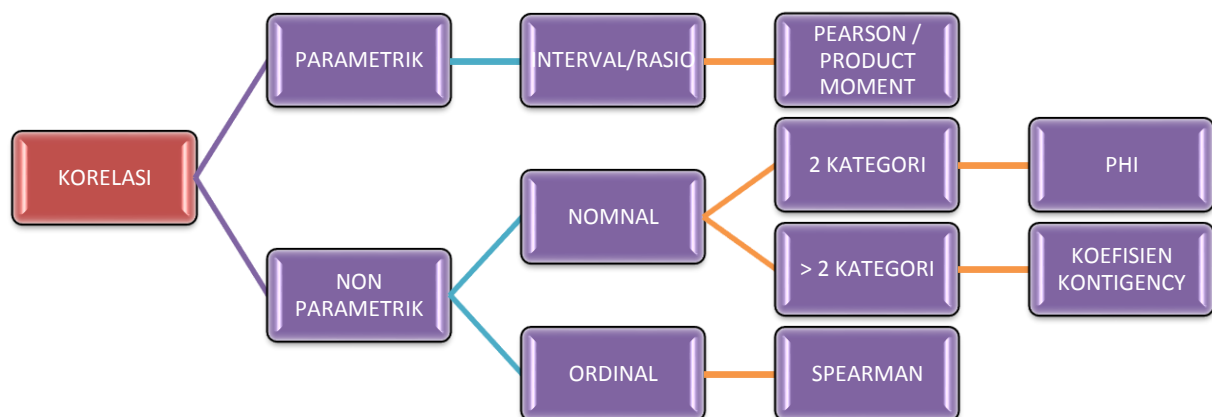
$$t_0 = \frac{-0,1}{\frac{0,14}{\sqrt{5}}} = -1,6$$

5) Kesimpulan

Karena  $t_0 = -1,6 > t_{0,01;4} = -3,747$ , maka  $H_0$  di terima. Jadi, keanggotaan organisasi bagi mahasiswa tidak membeikan pengaruh buruk terhadap prestasi akademiknya.

# BAB 9 KORELASI BIVARIATE

## KERANGKA KONSEP



### A. PENGERTIAN KORELASI

Korelasi adalah salah satu teknik analisis dalam statistik yang digunakan untuk mencari hubungan antara dua variabel yang juga bersifat kuantitatif. Sebagai contoh, kita dapat menggunakan tinggi badan dan usia siswa SD sebagai variabel dalam korelasi positif.

Statistik korelasi yaitu suatu cara atau metode untuk mengetahui ada atau tidaknya hubungan linear antar variabelnya. Dan apabila terdapat hubungan maka perubahan – perubahan yang terjadi pada salah satu variabel X akan mengakibatkan terjadinya perubahan pada variabel lainnya (Y).

Istilah tersebut juga dapat dikatakan istilah sebab akibat, dan istilah tersebut menjadi ciri khas tersendiri dari statistik korelasi. Di dalam kehidupan sehari-hari kita, hampir seluruh peristiwa yang dialami ini saling berkaitan. Misalnya seperti sebab curah hujan yang meningkat, keuntungan penjualan yang meningkat seiring penambahan dari jumlah barang ditoko, serta beberapa kasus lainnya. Lantas, apakah dengan mengetahui hubungan (korelasi) antar variabel penting? Iya, sebab bila

diketahui terjadinya hubungan di antara dua variabel tersebut, maka akan mudah pula dalam menentukan dan memprediksi dari nilai variabel lainnya.

Secara sederhana, korelasi dapat diartikan sebagai hubungan. Namun ketika dikembangkan lebih jauh, korelasi tidak hanya dapat dipahami sebatas pengertian tersebut. Korelasi merupakan salah satu teknik analisis dalam statistik yang digunakan untuk mencari hubungan antara dua variabel yang bersifat kuantitatif. Hubungan dua variabel tersebut dapat terjadi karena adanya hubungan sebab akibat atau dapat pula terjadi karena kebetulan saja. Dua variabel dikatakan berkorelasi apabila perubahan pada variabel yang satu akan diikuti perubahan pada variabel yang lain secara teratur dengan arah yang sama (korelasi positif) atau berlawanan (korelasi negatif).

Dalam Matematika, korelasi merupakan ukuran dari seberapa dekat dua variabel berubah dalam hubungan satu sama lain. Sebagai contoh, kita bisa menggunakan tinggi badan dan usia siswa SD sebagai variabel dalam korelasi positif. Semakin tua usia siswa SD, maka tinggi badannya pun menjadi semakin tinggi. Hubungan ini disebut korelasi positif karena kedua variabel mengalami perubahan ke arah yang sama, yakni dengan meningkatnya usia, maka tinggi badan pun ikut meningkat.

Sementara itu, kita bisa menggunakan nilai dan tingkat ketidakhadiran siswa sebagai contoh dalam korelasi negatif. Semakin tinggi tingkat ketidakhadiran siswa di kelas, maka nilai yang diperolehnya cenderung semakin rendah. Hubungan ini disebut korelasi negatif karena kedua variabel mengalami perubahan ke arah yang berlawanan, yakni dengan meningkatnya tingkat ketidakhadiran, maka nilai siswa justru menurun.

Kedua variabel yang dibandingkan satu sama lain dalam korelasi dapat dibedakan menjadi variabel independen dan variabel dependen. Sesuai dengan namanya, variabel independen adalah variabel yang perubahannya cenderung di luar kendali manusia. Sementara itu variabel dependen adalah variabel yang dapat berubah sebagai akibat dari perubahan variabel independen. Hubungan ini dapat dicontohkan dengan ilustrasi pertumbuhan tanaman dengan variabel sinar matahari dan tinggi tanaman. Sinar matahari merupakan variabel independen karena intensitas cahaya yang dihasilkan oleh matahari tidak dapat diatur oleh manusia. Sedangkan tinggi tanaman merupakan variabel dependen karena perubahan tinggi tanaman dipengaruhi langsung oleh intensitas cahaya matahari sebagai variabel independen.



Korelasi merupakan teknik analisis yang termasuk dalam salah satu teknik pengukuran asosiasi / hubungan (*measures of association*). Pengukuran asosiasi merupakan istilah umum yang mengacu pada sekelompok teknik dalam *statistik bivariat* yang digunakan untuk mengukur kekuatan hubungan antara dua variabel. Diantara sekian banyak teknik-teknik pengukuran asosiasi, terdapat dua teknik korelasi yang sangat populer sampai sekarang, yaitu Korelasi Pearson Product Moment dan Korelasi Rank Spearman. Selain kedua teknik tersebut, terdapat pula teknik-teknik korelasi lain, seperti Kendal, Chi-Square, Phi Coefficient, Goodman-Kruskal, Somer, dan Wilson.

Pengukuran asosiasi mengenakan nilai numerik untuk mengetahui tingkatan asosiasi atau kekuatan hubungan antara variabel. Dua variabel dikatakan berasosiasi jika perilaku variabel yang satu mempengaruhi variabel yang lain. Jika tidak terjadi pengaruh, maka kedua variabel tersebut disebut independen.

Korelasi bermanfaat untuk mengukur kekuatan hubungan antara dua variabel (kadang lebih dari dua variabel) dengan skala-skala tertentu, misalnya Pearson data harus berskala interval atau rasio; Spearman dan Kendal menggunakan skala ordinal; Chi Square menggunakan data nominal. Kuat lemah hubungan diukur diantara jarak (range) 0 sampai dengan 1. Korelasi mempunyai kemungkinan pengujian hipotesis dua arah (*two tailed*). Korelasi searah jika nilai koefisien korelasi ditemukan positif; sebaliknya jika nilai koefisien korelasi negatif, korelasi disebut tidak searah. Yang dimaksud dengan koefisien korelasi ialah suatu pengukuran statistik kovariansi atau asosiasi antara dua variabel. Jika koefisien korelasi ditemukan tidak sama dengan nol (0), maka terdapat ketergantungan antara dua variabel tersebut. Jika koefisien korelasi ditemukan +1. maka hubungan tersebut disebut sebagai korelasi sempurna atau hubungan linear sempurna dengan kemiringan (slope) positif.

Jika koefisien korelasi ditemukan -1. maka hubungan tersebut disebut sebagai korelasi sempurna atau hubungan linear sempurna dengan kemiringan (slope) negatif. Dalam korelasi sempurna tidak diperlukan lagi pengujian hipotesis, karena kedua variabel mempunyai hubungan linear yang sempurna. Artinya variabel X mempengaruhi variabel Y secara sempurna. Jika korelasi sama dengan nol (0), maka tidak terdapat hubungan antara kedua variabel tersebut.

Dalam korelasi sebenarnya tidak dikenal istilah variabel bebas dan variabel tergantung. Biasanya dalam penghitungan digunakan simbol X untuk variabel

pertama dan Y untuk variabel kedua. Dalam contoh hubungan antara variabel remunerasi dengan kepuasan kerja, maka variabel remunerasi merupakan variabel X dan kepuasan kerja merupakan variabel Y.

## B. KOEFISIEN KORELASI

Sering kali kita ingin mengetahui bagaimanakah hubungan antara satu variabel dengan variabel lainnya. Apakah variabel X mempunyai hubungan dengan variabel Y ? Apakah nilai Matematika siswa mempunyai hubungan dengan tingkat kecerdasannya (IQ) ? Apakah terdapat kesepakatan antara para juri dalam menilai para pesertanya dimana dalam penilaiannya skor tertinggi adalah 10 dan terendah adalah 1 ?

Koefisien korelasi adalah nilai yang menunjukkan kuat/tidaknya hubungan linier antar dua variabel. Koefisien korelasi biasa dilambangkan dengan huruf r dimana nilai r dapat bervariasi dari -1 sampai +1. Nilai r yang mendekati -1 atau +1 menunjukkan hubungan yang kuat antara dua variabel tersebut dan nilai r yang mendekati 0 mengindikasikan lemahnya hubungan antara dua variabel tersebut. Sedangkan tanda + (positif) dan - (negatif) memberikan informasi mengenai arah hubungan antara dua variabel tersebut. Jika bernilai + (positif) maka kedua variabel tersebut memiliki hubungan yang searah. Dalam arti lain peningkatan X akan bersamaan dengan peningkatan Y dan begitu juga sebaliknya. Jika bernilai - (negatif) artinya korelasi antara kedua variabel tersebut bersifat berlawanan. Peningkatan nilai X akan dibarengi dengan penurunan Y.

Koefisien korelasi pearson atau Product Moment Coefficient of Correlation adalah nilai yang menunjukkan keeratan hubungan linier dua variabel dengan skala data interval atau rasio. Rumus yang digunakan adalah

$$r = \frac{n \sum_{i=1}^n X_i Y_i - \sum_{i=1}^n X_i \sum_{i=1}^n Y_i}{\sqrt{n \sum_{i=1}^n X_i^2 - (\sum_{i=1}^n X_i)^2} \sqrt{n \sum_{i=1}^n Y_i^2 - (\sum_{i=1}^n Y_i)^2}}$$

Koefisien korelasi rangking Spearman atau Spearman rank correlation coefficient merupakan nilai yang menunjukkan keeratan hubungan linier antara dua variabel dengan skala data ordinal. Koefisien Spearman biasa dilambangkan dengan . Rumusnya yang digunakan adalah

$$r_s = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

Dimana  $d_i$  diselisih dari pasangan ke-i atau  $X_i - Y_i$  ;

$n$  = banyaknya pasangan rank

Jika variabel X dan Y independen maka nilai  $r = 0$ , akan tetapi jika nilai  $r=0$ , X dan Y tidak selalu independen. Variabel X dan Y hanya tidak berasosiasi. Perlu diketahui bahwa hasil dari koefisien koefisien korelasi hanya bisa digunakan sebagai indikasi awal dalam analisa. Nilai dari koefisien korelasi tidak dapat menggambarkan hubungan sebab akibat antara variabel X dan Y. Untuk sampai pada adanya hubungan sebab dan akibat diperlukan penelitian yang lebih intensif atau dapat didasarkan pada teori yang ada dimana X mempengaruhi Y atau Y yang mempengaruhi X.

Selain itu, dalam menganalisa hubungan antara X dan Y, tentunya harus didasarkan adanya hubungan yang logis antara kedua variabel tersebut. Kita tidak bisa sembarangan mengukur koefisien korelasi antara dua variabel. Misalnya, variabel Y merupakan data mengenai banyaknya angka kecelakaan yang terjadi di Jakarta pada tahun 2013 dan variabel X adalah jumlah kasus pencurian di Jakarta pada tahun 2013. Kemudian dihitung koefisien korelasi antara variabel X dan Y, diperoleh hubungannya yang kuat antara kedua variabel tersebut. Disini nilai koefisien korelasi yang didapat tentunya tidak akan memiliki makna meskipun didapat nilai korelasi yang kuat karena secara logis tingkat kecelakaan tidak memiliki hubungan dengan tingkat pencurian yang ada.

Koefisien korelasi ialah pengukuran statistik kovarian atau asosiasi antara dua variabel. Besarnya koefisien korelasi berkisar antara +1 s/d -1. Koefisien korelasi menunjukkan kekuatan (strength) hubungan linear dan arah hubungan dua variabel acak. Jika koefisien korelasi positif, maka kedua variabel mempunyai hubungan searah. Artinya jika nilai variabel X tinggi, maka nilai variabel Y akan tinggi pula. Sebaliknya, jika koefisien korelasi negatif, maka kedua variabel mempunyai hubungan terbalik. Artinya jika nilai variabel X tinggi, maka nilai variabel Y akan menjadi rendah (dan sebaliknya). Untuk memudahkan melakukan interpretasi mengenai kekuatan hubungan antara dua variabel penulis memberikan kriteria sebagai berikut

- 0 : Tidak ada korelasi antara dua variabel
- $>0 - 0,25$ : Korelasi sangat lemah
- $>0,25 - 0,5$ : Korelasi cukup
- $>0,5 - 0,75$ : Korelasi kuat
- $>0,75 - 0,99$ : Korelasi sangat kuat
- 1: Korelasi sempurna

## 1. Signifikansi

Apa sebenarnya signifikansi itu? Dalam bahasa Inggris umum, kata, “significant” mempunyai makna penting; sedang dalam pengertian statistik kata tersebut mempunyai makna “benar” tidak didasarkan secara kebetulan. Hasil riset dapat benar tapi tidak penting. Signifikansi / probabilitas /  $\alpha$  memberikan gambaran mengenai bagaimana hasil riset itu mempunyai kesempatan untuk benar. Jika kita memilih signifikansi sebesar 0,01, maka artinya kita menentukan hasil riset nanti mempunyai kesempatan untuk benar sebesar 99% dan untuk salah sebesar 1%.

Secara umum kita menggunakan angka signifikansi sebesar 0,01; 0,05 dan 0,1. Pertimbangan penggunaan angka tersebut didasarkan pada tingkat kepercayaan (confidence interval) yang diinginkan oleh peneliti. Angka signifikansi sebesar 0,01 mempunyai pengertian bahwa tingkat kepercayaan atau bahasa umumnya keinginan kita untuk memperoleh kebenaran dalam riset kita adalah sebesar 99%. Jika angka signifikansi sebesar 0,05, maka tingkat kepercayaan adalah sebesar 95%. Jika angka signifikansi sebesar 0,1, maka tingkat kepercayaan adalah sebesar 90%.

Pertimbangan lain ialah menyangkut jumlah data (sample) yang akan digunakan dalam riset. Semakin kecil angka signifikansi, maka ukuran sample akan semakin besar. Sebaliknya semakin besar angka signifikansi, maka ukuran sample akan semakin kecil. Untuk memperoleh angka signifikansi yang baik, biasanya diperlukan ukuran sample yang besar. Sebaliknya jika ukuran sample semakin kecil, maka kemungkinan munculnya kesalahan semakin ada.

Untuk pengujian dalam SPSS digunakan kriteria sebagai berikut:

- Jika angka signifikansi hasil riset  $< 0,05$ , maka hubungan kedua variabel signifikan.

- Jika angka signifikansi hasil riset  $> 0,05$ , maka hubungan kedua variabel tidak signifikan

## 2. Interpretasi Korelasi

Ada tiga penafsiran hasil analisis korelasi, meliputi: pertama, melihat kekuatan hubungan dua variabel; kedua, melihat signifikansi hubungan; dan ketiga, melihat arah hubungan.

Untuk melakukan interpretasi kekuatan hubungan antara dua variabel dilakukan dengan melihat angka koefisien korelasi hasil perhitungan dengan menggunakan kriteria sbb:

- Jika angka koefisien korelasi menunjukkan 0, maka kedua variabel tidak mempunyai hubungan
- Jika angka koefisien korelasi mendekati 1, maka kedua variabel mempunyai hubungan semakin kuat
- Jika angka koefisien korelasi mendekati 0, maka kedua variabel mempunyai hubungan semakin lemah
- Jika angka koefisien korelasi sama dengan 1, maka kedua variabel mempunyai hubungan linier sempurna positif.
- Jika angka koefisien korelasi sama dengan -1, maka kedua variabel mempunyai hubungan linier sempurna negatif.

Interpretasi berikutnya melihat signifikansi hubungan dua variabel dengan didasarkan pada angka signifikansi yang dihasilkan dari penghitungan dengan ketentuan sebagaimana sudah dibahas di bagian 2.7. di atas. Interpretasi ini akan membuktikan apakah hubungan kedua variabel tersebut signifikan atau tidak.

Interpretasi ketiga melihat arah korelasi. Dalam korelasi ada dua arah korelasi, yaitu searah dan tidak searah. Pada SPSS hal ini ditandai dengan pesan two tailed. Arah korelasi dilihat dari angka koefisien korelasi. Jika koefisien korelasi positif, maka hubungan kedua variabel searah. Searah artinya jika variabel X nilainya tinggi, maka variabel Y juga tinggi. Jika koefisien korelasi negatif, maka hubungan kedua variabel tidak searah. Tidak searah artinya jika variabel X nilainya tinggi, maka variabel Y akan rendah.

Dalam kasus, misalnya hubungan antara kepuasan kerja dan komitmen terhadap organisasi sebesar 0,86 dengan angka signifikansi sebesar 0 akan

mempunyai makna bahwa hubungan antara variabel kepuasan kerja dan komitmen terhadap organisasi sangat kuat, signifikan dan searah. Sebaliknya dalam kasus hubungan antara variabel mangkir kerja dengan produktivitas sebesar -0,86, dengan angka signifikansi sebesar 0; maka hubungan kedua variabel sangat kuat, signifikan dan tidak searah.

### 3. Koefisien Determinasi

Koefisien determinasi dengan simbol  $r^2$  merupakan proporsi variabilitas dalam suatu data yang dihitung didasarkan pada model statistik. Definisi berikutnya menyebutkan bahwa  $r^2$  merupakan rasio variabilitas nilai-nilai yang dibuat model dengan variabilitas nilai data asli. Secara umum  $r^2$  digunakan sebagai informasi mengenai kecocokan suatu model. Dalam regresi  $r^2$  ini dijadikan sebagai pengukuran seberapa baik garis regresi mendekati nilai data asli yang dibuat model. Jika  $r^2$  sama dengan 1, maka angka tersebut menunjukkan garis regresi cocok dengan data secara sempurna.

Interpretasi lain ialah bahwa  $r^2$  diartikan sebagai proporsi variasi tanggapan yang diterangkan oleh regresor (variabel bebas / X) dalam model. Dengan demikian, jika  $r^2 = 1$  akan mempunyai arti bahwa model yang sesuai menerangkan semua variabilitas dalam variabel Y. jika  $r^2 = 0$  akan mempunyai arti bahwa tidak ada hubungan antara regresor (X) dengan variabel Y. Dalam kasus misalnya jika  $r^2 = 0,8$  mempunyai arti bahwa sebesar 80% variasi dari variabel Y (variabel tergantung / response) dapat diterangkan dengan variabel X (variabel bebas / explanatory); sedang sisanya 0,2 dipengaruhi oleh variabel-variabel yang tidak diketahui atau variabilitas yang inheren. (Rumus untuk menghitung koefisien determinasi (KD) adalah  $KD = r^2 \times 100\%$ ) Variabilitas mempunyai makna penyebaran / distribusi seperangkat nilai-nilai tertentu. Dengan menggunakan bahasa umum, pengaruh variabel X terhadap Y adalah sebesar 80%; sedang sisanya 20% dipengaruhi oleh faktor lain.

Dalam hubungannya dengan korelasi, maka  $r^2$  merupakan kuadrat dari koefisien korelasi yang berkaitan dengan variabel bebas (X) dan variabel Y (tergantung). Secara umum dikatakan bahwa  $r^2$  merupakan kuadrat korelasi antara variabel yang digunakan sebagai predictor (X) dan variabel yang memberikan response (Y). Dengan menggunakan bahasa sederhana  $r^2$  merupakan koefisien korelasi yang dikuadratkan. Oleh karena itu,

penggunaan koefisien determinasi dalam korelasi tidak harus diinterpretasikan sebagai besarnya pengaruh variabel X terhadap Y mengingat bahwa korelasi tidak sama dengan kausalitas. Secara bebas dikatakan dua variabel mempunyai hubungan belum tentu variabel satu mempengaruhi variabel lainnya. Lebih lanjut dalam konteks korelasi antara dua variabel maka pengaruh variabel X terhadap Y tidak nampak. Kemungkinannya hanya korelasi merupakan penanda awal bahwa variabel X mungkin berpengaruh terhadap Y. Sedang bagaimana pengaruh itu terjadi dan ada atau tidak kita akan mengalami kesulitan untuk membuktikannya. Hanya menggunakan angka  $r^2$  kita tidak akan dapat membuktikan bahwa variabel X mempengaruhi Y.

Dengan demikian jika kita menggunakan korelasi sebaiknya jangan menggunakan koefisien determinasi untuk melihat pengaruh X terhadap Y karena korelasi hanya menunjukkan adanya hubungan antara variabel X dan Y. Jika tujuan riset hanya untuk mengukur hubungan maka sebaiknya berhenti saja di angka koefisien korelasi. Sedang jika kita ingin mengukur besarnya pengaruh variabel X terhadap Y sebaiknya menggunakan rumus lain, seperti regresi atau analisis jalur.

### Contoh Soal

Di bawah ini disajikan data tentang harga rata-rata dollar Amerika dan emas 24 karat di pasaran Jakarta tiap akhir tahun selama 1970 s/d 1978.

Tahun	Harga dollar US dalam rupiah	Harga emas 24 karat dalam rupiah
1970	382,00	485,00
1971	420,00	622,00
1972	420,00	876,00
1973	420,00	1.483,00
1974	422,00	2.303,00
1975	420,00	1.900,00

1976	421,00	1.850,00
1977	420,00	2.150,00
1978	632,00	4.300,00

Tentukan apakah terdapat Korelasi antara harga dollar US dan harga emas 24 karat?  
Jelaskan artinya!

**Penyelesaian**

Harga dollar US = X

Harga emas 24 karat = Y

Dari data yang ada, diketahui bahwa kedua variabel, yaitu harga dollat US dan harga emas 24 karat mempunyai skala data rasio. Maka koefisien korelasi yang digunakan adalah Koefisien korelasi pearson

No	X	Y	X <sup>2</sup>	Y <sup>2</sup>	XY
1	382	485	145.924	235.225	185.270
2	420	622	176.400	386.884	216.240
3	420	876	176.400	767.376	367.920
4	420	1.483	176.400	2.199.289	622.860
5	422	2.303	178.084	5.303.809	971.866
6	420	1.900	176.400	3.610.000	798.000
7	421	1.850	177.241	3.422.500	778.850
8	420	2.150	176.400	4.622.500	903.000
9	632	4.300	399.424	18.490.000	2.717.600
Jumlah	3.975	15.969	1.782.673	39.037.583	7.561.606



$$r = \frac{n \sum_{i=1}^n X_i Y_i - \sum_{i=1}^n X_i \sum_{i=1}^n Y_i}{\sqrt{n \sum_{i=1}^n X_i^2 - (\sum_{i=1}^n X_i)^2} \sqrt{n \sum_{i=1}^n Y_i^2 - (\sum_{i=1}^n Y_i)^2}}$$

$$r = \frac{9(7.561.606) - (3975)(15.969)}{\sqrt{9(1.782.673) - (3975)^2} \sqrt{9(39.037.583) - (15.969)^2}}$$

$$= \frac{68.054.454 - 63.476.775}{\sqrt{243.432} \sqrt{9.329.286}} = \frac{4.577.679}{4.842.481,879} = 0,945$$

Jadi, terdapat hubungan linier antara harga dollar US dan harga emas 24 karat dimana hubungan linier yang terjadi dapat dikatakan kuat dan positif. Dengan demikian, kenaikan harga dollar US terjadi bersama – sama dengan kenaikan harga emas 24 karat. Begitu juga sebaliknya, penurunan harga dollar US terjadi berasama – sama dengan penurunan harga emas 24 karat.

### Contoh Soal

Nur dan Hap diminta untuk menilai beberapa merek handphone. Dari 10 merek handphone yang diberikan akan dinilai manakah yang paling bagus. Penilaian yang diberikan berkisar dari nilai paling rendah, yaitu 1 dan paling tinggi 10. Hasil dari penilaian disajikan pada tabel di bawah ini

Brand	Nur	Hap
Siemens BenQ	10	9
Alcatel	6	10
Samsung	8	8
O2	9	6
Sony Ericson	5	5
Vodafone	7	7
Motorola	4	3

Sanex	2	4
Nokia	1	1
LG	3	2

Apakah terdapat kesepakatan antara Nur dan Hap dalam memberikan penilaian terhadap merek handphone?

### Penyelesaian

Dalam contoh yang ke dua ini, dapat dilihat bahwa data yang digunakan mempunyai skala ordinal karena data disusun berdasarkan rangking. Misalkan rangking yang diberikan oleh Nur dilambangkan X dan Hap diwakilkan oleh Y.

Brand	Rank X	Rank Y	$d_i$	$d_i^2$
(1)	(3)	(5)	(6)	(7)
Siemens BenQ	10	9	1	1
Alcatel	6	10	-4	16
Samsung	8	8	0	0
O2	9	6	3	9
Sony Ericson	5	5	0	0
Vodafone	7	7	0	0
Motorola	4	3	1	1
Sanex	2	4	-2	4
Nokia	1	1	0	0
LG	3	2	1	1

Total	55	55	0	32
-------	----	----	---	----

$$r_s = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

$$r_s = 1 - \frac{6(32)}{10(10^2 - 1)} = 1 - 0,19 = 0,81$$

Dengan demikian, dapat dikatakan terdapat hubungan yang linier antara Nur dan Hap dalam memberikan penilaian terhadap kesepuluh merek handphone yang diberikan. Hubungan linier yang terjadi adalah positif dan kuat.

## C. MACAM KORELASI DAN TEKNIK PENGHITUNGAN

### 1. Korelasi Untuk Uji Parametrik

Pada korelasi parametrik biasanya dilakukan dengan koefisien korelasi hasil kali momen Pearson ( $r$ ). Korelasi ini menuntut data yang digunakan sekurang-kurangnya dalam skala interval, dan uji signifikansinya tidak hanya harus memenuhi persyaratan pengukuran tersebut, tetapi harus pula menganggap data berasal dari suatu populasi berdistribusi normal.

Batas-Batas Koefisien Korelasi, menurut Umar (2002: 314) nilai koefisien korelasi berkisar antara  $-1$  sampai  $+1$ , yang kriteria pemanfaatannya dijelaskan sebagai berikut:

- Jika, nilai  $r > 0$ , artinya telah terjadi hubungan yang linier positif, yaitu makin besar nilai variabel X makin besar pula nilai variabel Y atau makin kecil nilai variabel X makin kecil pula nilai variabel Y.
- Jika, nilai  $r < 0$ , artinya telah terjadi hubungan yang linier negatif, yaitu makin besar nilai variabel X makin kecil nilai variabel Y atau makin kecil nilai variabel X maka makin besar pula nilai variabel Y
- Jika, nilai  $r = 0$ , artinya tidak ada hubungan sama sekali antara variabel X dan variabel Y.
- Jika, nilai  $r = 1$  atau  $r = -1$ , maka dapat dikatakan telah terjadi hubungan linier sempurna, berupa garis lurus, sedangkan untuk  $r$  yang makin mengarah ke angka 0 (nol) maka garis makin tidak lurus.

### a) Uji Pearson Product Moment dan Asumsi Klasik

Korelasi Pearson atau sering disebut Korelasi Product Moment (KPM) merupakan alat uji statistik yang digunakan untuk menguji hipotesis asosiatif (uji hubungan) dua variabel bila datanya berskala interval atau rasio. KPM dikembangkan oleh Karl Pearson (Hasan, 1999).

KPM merupakan salah satu bentuk statistik parametris karena menguji data pada skala interval atau rasio. Oleh karena itu, ada beberapa persyaratan untuk dapat menggunakan KPM, yaitu :

- a) Sampel diambil dengan teknik random (acak)
- b) Data yang akan diuji harus homogen
- c) Data yang akan diuji juga harus berdistribusi normal
- d) Data yang akan diuji bersifat linier

Fungsi KPM sebagai salah satu statistik inferensia adalah untuk menguji kemampuan generalisasi (signifikansi) hasil penelitian. Adapun syarat untuk bisa menggunakan KPM selain syarat menggunakan statistik parameteris, juga ada persyaratan lain, yaitu variabel independen (X) dan variabel (Y) harus berada pada skala interval atau rasio.

Nilai KPM disimbolkan dengan  $r$  (rho). Nilai KPM juga berada di antara  $-1 \leq r \leq 1$ . Bila nilai  $r = 0$ , berarti tidak ada korelasi atau tidak ada hubungan antara variabel independen dan dependen. Nilai  $r = +1$  berarti terdapat hubungan yang positif antara variabel independen dan dependen. Nilai  $r = -1$  berarti terdapat hubungan yang negatif antara variabel independen dan dependen. Dengan kata lain, tanda “+” dan “-” menunjukkan arah hubungan di antara variabel yang sedang dioperationalkan.

Uji signifikansi KPM menggunakan uji t, sehingga nilai t hitung dibandingkan dengan nilai t tabel. Kekuatan hubungan antarvariabel ditunjukkan melalui nilai korelasi. Berikut adalah tabel nilai korelasi beserta makna nilai tersebut :

Tabel Makna Nilai Korelasi Product Moment

Nilai	Makna
0,00 – 0,19	Sangat rendah / sangat lemah
0,20 – 0,39	
0,40 – 0,59	
0,60 – 0,79	
0,80 – 1,00	

Korelasi pearson digunakan untuk mengetahui ada tidaknya hubungan antara 2 variabel, yaitu variabel bebas dan variabel tergantung yang berskala interval atau rasio (parametrik) yang dalam SPSS disebut *scale*. Asumsi dalam korelasi Pearson, data harus berdistribusi normal. Korelasi dapat menghasilkan angka positif (+) dan negatif (-). Jika angka korelasi positif berarti hubungan bersifat searah. Searah artinya jika variabel bebas besar, variabel tergantung semakin besar. Jika menghasilkan angka negatif berarti hubungan bersifat tidak searah. Tidak searah artinya jika nilai variabel bebas besar, variabel tergantung semakin kecil. angka korelasi berkisar antara 0-1.

Rumus korelasi pearson :

$$r_{xy} = \frac{n \sum XY - \sum X \sum Y}{\sqrt{[n \sum X^2 - (\sum X)^2][n \sum Y^2 - (\sum Y)^2]}}$$

Dimana :

r = Pearson r correlation coefficient

N = jumlah sampel

r = nilai korelasi

x = variabel x

y = variabel y

Kekuatan hubungan korelasi, menurut Jonathan Sarwono sebagai berikut :

- 0 : Tidak ada korelasi
- 0.00 - 0.25 : korelasi sangat lemah
- 0.25 - 0.50 : korelasi cukup
- 0.50 - 0.75 : korelasi kuat
- 0.75 - 0.99 : korelasi sangat kuat
- 1 : korelasi sempurna

Korelasi Pearson merupakan salah satu ukuran korelasi yang digunakan untuk mengukur kekuatan dan arah hubungan linier dari dua variabel. Dua variabel dikatakan berkorelasi apabila perubahan salah satu variabel disertai dengan perubahan variabel lainnya, baik dalam arah yang sama ataupun arah yang sebaliknya. Harus diingat bahwa nilai koefisien korelasi yang kecil (tidak signifikan) bukan berarti kedua variabel tersebut tidak saling berhubungan. Mungkin saja dua variabel mempunyai keeratan hubungan yang kuat namun nilai koefisien korelasinya mendekati nol, misalnya pada kasus hubungan non linier. Dengan demikian, koefisien korelasi hanya mengukur kekuatan hubungan linier dan tidak pada hubungan non linier. Harus diingat pula bahwa adanya hubungan linier yang kuat di antara variabel tidak selalu berarti ada hubungan kausalitas, sebab-akibat.

Uji Pearson Product Moment adalah salah satu dari beberapa jenis uji korelasi yang digunakan untuk mengetahui derajat keeratan hubungan 2 variabel yang berskala interval atau rasio, di mana dengan uji ini akan mengembalikan nilai koefisien korelasi yang nilainya berkisar antara -1, 0 dan 1. Nilai -1 artinya terdapat korelasi negatif yang sempurna, 0 artinya tidak ada korelasi dan nilai 1 berarti ada korelasi positif yang sempurna.

Rentang dari koefisien korelasi yang berkisar antara -1, 0 dan 1 tersebut dapat disimpulkan bahwa apabila semakin mendekati nilai 1 atau -1 maka hubungan makin erat, sedangkan jika semakin mendekati 0 maka hubungan semakin lemah.

Berikut Tabel klasifikasi nilai koefisien korelasi  $r$  pearson:

Interval Koefisien	Tingkat Hubungan
0,80 – 1,000	Sangat Kuat
0,60 – 0,799	Kuat
0,40 – 0,599	Cukup Kuat
0,20 – 0,399	Rendah
0,00 – 0,199	Sangat Rendah

Tabel Klasifikasi Koefisien Pearson

Berdasarkan tabel di atas, dapat kami jelaskan tentang nilai koefisien korelasi uji pearson product moment dan makna keeratannya dalam sebuah analisis statistik atau analisis data. Berikut penjelasannya:

- 1) Nilai koefisien 0 = Tidak ada hubungan sama sekali (jarang terjadi),
- 2) Nilai koefisien 1 = Hubungan sempurna (jarang terjadi),
- 3) Nilai koefisien  $> 0$  sd  $< 0,2$  = Hubungan sangat rendah atau sangat lemah,
- 4) Nilai koefisien  $0,2$  sd  $< 0,4$  = Hubungan rendah atau lemah,
- 5) Nilai koefisien  $0,4$  sd  $< 0,6$  = Hubungan cukup besar atau cukup kuat,
- 6) Nilai koefisien  $0,6$  sd  $< 0,8$  = Hubungan besar atau kuat,
- 7) Nilai koefisien  $0,8$  sd  $< 1$  = Hubungan sangat besar atau sangat kuat.
- 8) Nilai negatif berarti menentukan arah hubungan, misal: koefisien korelasi antara penghasilan dan berat badan bernilai -0,5. Artinya semakin tinggi nilai penghasilan seseorang maka semakin rendah berat badannya dengan besarnya keeratan hubungan sebesar 0,5 atau cukup kuat (*lihat tabel di atas*).

#### **b) Signifikansi Atau P Value Uji Pearson Product Moment**

Pengujian lanjutan untuk menentukan apakah koefisien korelasi yang didapat bisa digunakan untuk generalisasi atau mewakili populasi, maka digunakan uji signifikansi dari uji t. Maka nilai r pearson yang didapat digunakan untuk menghitung nilai t hitung. Berikut rumusnya:

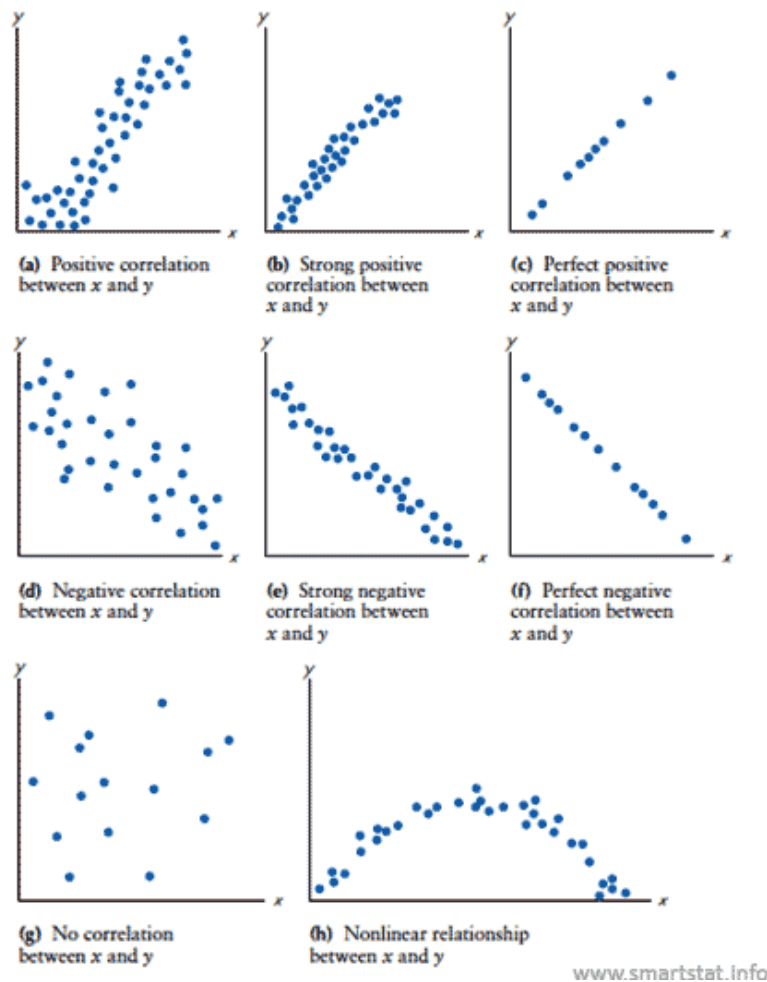
$$t_{\text{hitung}} = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

Nilai  $t$  hitung yang di dapat nantinya kita bandingkan dengan nilai  $t$  tabel. Apabila  $t$  hitung  $>$   $t$  tabel pada derajat kepercayaan tertentu, misal 95 % maka berarti signifikan atau bermakna.

### c) Asumsi Klasik Uji Pearson Product Moment

Sebelum melakukan analisis korelasi antar variabel, sebaiknya kita mengeksplorasi data tersebut terlebih dahulu secara grafis. Seringkali kita melihat pola hubungan di antara variabel dengan cara memplotkan pasangan sampel data tersebut pada diagram kartesian yang disebut dengan *scatterplot* atau diagram pencar. Setiap pasangan data  $(x, y)$  diplotkan sebagai titik tunggal.

Contoh diagram pencar dapat dilihat pada gambar berikut.



Secara sepintas kita bisa melihat pola hubungan dari grafik-grafik tersebut. Pada **Grafik a, b, c** terlihat bahwa peningkatan nilai  $y$  sejalan dengan peningkatan nilai  $x$ . Apabila nilai  $x$  meningkat, maka



nilai  $y$  pun meningkat, dan sebaliknya. Dari **Grafik a** sampai **c**, sebaran titik-titik pasangan data semakin mendekati bentuk garis lurus yang menunjukkan bahwa keeratan hubungan antara variabel  $x$  dan  $y$  semakin kuat (*sinergis*).

Hal yang sebaliknya terjadi pada **Grafik d, e, dan f**. Peningkatan nilai  $y$  tidak sejalan dengan peningkatan nilai  $x$  (*antagonis*). Peningkatan salah satu nilai menyebabkan penurunan nilai pasangannya. Sekali lagi tampak bahwa kekuatan hubungan antara kedua variabel dari **d** menuju **f** semakin kuat.

Berbeda dengan grafik sebelumnya, pada **Grafik g** tidak menunjukkan adanya pola hubungan linier antara kedua variabel. Hal ini menandakan bahwa tidak ada korelasi di antara kedua variabel tersebut. Terakhir, pada **Grafik h** kita bisa melihat adanya pola hubungan di antara kedua variabel tersebut, hanya saja polanya bukan dalam bentuk hubungan linier, melainkan dalam bentuk kuadratik.

#### **Contoh Soal :**

Berikut ini merupakan data perusahaan mengenai harga permintaan suatu komoditi ( $X$ ) dan harga rata-rata suatu komoditi ( $Y$ ) disajikan dalam tabel berikut:

X	Y
178	105
224	105
160	130
315	130
229	130
250	150
181	150

306	170
257	170
300	180

Hitunglah koefisien korelasi pada kasus tersebut dan bagaimana arti dari hasil koefisien korelasi yang didapat!

**Jawaban :**

Berikut ini hasil perhitungan tabel untuk mendapatkan nilai total untuk X, Y, X<sup>2</sup>, Y<sup>2</sup>, dan XY:

Data	X	Y	X <sup>2</sup>	Y <sup>2</sup>	XY
1	178	105	31684	11025	18690
2	224	105	50176	11025	23520
3	160	130	25600	16900	20800
4	315	130	99225	16900	40950
5	229	130	52441	16900	29770
6	250	150	62500	22500	37500
7	181	150	32761	22500	27150
8	306	170	93636	28900	52020
9	257	170	66049	28900	43690
10	300	180	90000	32400	54000
Total	2400	1420	604072	207950	348090

Dengan memasukan nilai total dari semua variabel pada tabel dan jumlah

data ke dalam rumus korelasi Pearson Product Moment maka didapat hasil sebagai berikut:

$$r = \frac{n \sum_{i=1}^n X_i Y_i - \sum_{i=1}^n X_i \sum_{i=1}^n Y_i}{\sqrt{(n \sum_{i=1}^n X_i^2 - (\sum_{i=1}^n X_i)^2)(n \sum_{i=1}^n Y_i^2 - (\sum_{i=1}^n Y_i)^2)}}$$

$$r = \frac{10(348090) - (2400)(1420)}{\sqrt{(10(604072) - (2400)^2)(10(207950) - (1420)^2)}}$$

$$r = 0.5477$$

Maka nilai koefisien korelasi Pearson Product Momentnya adalah 0.5477. Bagaimana intepretasi arti dari hasil koefisien korelasi Pearson Product Moment pada kasus tersebut?

- Koefisien Korelasi antara permintaan suatu komoditi dengan harga rata-rata komoditi sebesar 0.5477 dengan hubungan antara permintaan suatu komoditi dengan harga rata-rata komoditi memiliki tingkat hubungan yang menengah sehingga terkadang apabila ditambahkan permintaan suatu komoditi maka akan mempengaruhi peningkatan harga rata-rata suatu komoditi. Apabila diturunkan permintaan suatu komoditi maka akan mempengaruhi penurunan harga rata-rata suatu komoditi.
- Dapat juga dibalik intepretasinya, apabila ditambahkan harga rata-rata suatu komoditi maka akan mempengaruhi peningkatan permintaan suatu komoditi. Apabila diturunkan harga rata-rata suatu komoditi maka akan mempengaruhi permintaan suatu komoditi.

**Contoh Soal :**

Diketahui data rekap data hasil pendapatan perusahaan (Y) berdasarkan jumlah sales (X<sub>1</sub>) dan harga yang dijual (X<sub>2</sub>) dengan masing-masing koefisien korelasi antar variabel yang didapat sebagai berikut:

$$r_{x_1x_2} = -0.545$$

$$r_{x_1y} = 0.986$$

$$r_{x_2y} = 0.436$$

Intepretasikan hasil koefisien korelasi berdasarkan informasi yang didapat dari kasus tersebut!

### **Jawaban Soal :**

- Untuk  $r_{x_1x_2} = -0.545$  memiliki arti bahwa koefisien korelasi jumlah sales dengan harga jual sebesar  $-0.545$  dengan arah hubungan yang negatif serta tingkat hubungan yang moderate, sehingga terkadang apabila jumlah sales ditingkatkan maka akan menurunkan harga jual. Sebaliknya terkadang apabila jumlah sales diturunkan maka akan meningkatkan harga jual.
- Untuk  $r_{x_1y} = 0.986$  memiliki arti bahwa koefisien korelasi jumlah sales dengan pendapatan sebesar  $0.986$  dengan arah hubungan yang positif serta tingkat hubungan yang sangat kuat, sehingga jelas bahwa apabila jumlah sales ditingkatkan maka akan meningkatkan pendapatan. Sebaliknya jelas pula apabila jumlah sales diturunkan maka akan menurunkan pendapatan.
- Untuk  $r_{x_2y} = 0.436$  memiliki arti bahwa koefisien korelasi harga jual dengan pendapatan sebesar  $0.436$  dengan arah hubungan yang positif serta tingkat hubungan yang moderate, sehingga terkadang apabila harga jual ditingkatkan maka akan meningkatkan pendapatan. Sebaliknya terkadang apabila harga jual diturunkan maka akan menurunkan pendapatan.

## **1. Korelasi Untuk Uji Non Parametrik**

### **a) Koefisien Phi**

Teknik korelasi Phi adalah salah satu teknik analisis korelasional yang dipergunakan apabila data yang dikorelasikan adalah data yang benar-benar dikotomik (terpisah atau dipisahkan secara tajam); dengan istilah lain : variabel yang dikorelasikan itu adalah variabel diskrit murni, misalnya: Laki-laki-Perempuan, Hidup-Mati, Lulus-Tidak Lulus, dsb. Apabila variabelnya bukan merupakan variabel diskrit dan kita ingin menganalisis data tersebut dengan

menggunakan teknik ini, maka variabel tersebut harus diubah lebih dulu menjadi variabel diskrit.

Besar-kecil, kuat-lemah, atau tinggi-rendahnya korelasi antar dua variabel yang kita selidiki korelasinya pada Teknik Korelasi Phi ini, ditunjukkan oleh besar kecilnya angka indeks korelasi yang dilambangkan dengan huruf  $\phi$  (Phi). Phi besarnya berkisar antara 0,00 sampai dengan  $\pm 1,00$ .

Koefisien phi adalah uji asosiatif pada skala data nominal apabila tabel kontingensi berbentuk  $2 \times 2$ . Sehingga dapat dikatakan bahwa koefisien korelasi phi dirancang untuk peubah dikhotom. Makna dari dikhotom atau dikotomi adalah dua kategori yang bersifat nominal. Silahkan pelajari artikel kami tentang Jenis Data dan Variabel Penelitian.

Kita mempunyai dua peubah, peubah I dan peubah II yang hasil amatannya disajikan dalam bentuk tabel Kontingensi  $2 \times 2$ .

Berikut di bawah ini adalah tabel phi, yang berfungsi sebagai tabel penolong dalam proses perhitungan phi. Seperti sudah dijelaskan di atas, bahwa koefisien korelasi phi melibatkan 2 kategori pada kedua peubah atau kedua variabel.

Kategori Peubah II	Kategori Peubah I		Total
	1	2	
1	a	b	a + b
2	c	d	c + d
Total	a + c	b + d	n

Tabel Penolong Uji Phi

- 1) Rumus Koefisien Phi dijelaskan sebagai berikut:

$$\phi = \frac{ad - bc}{\sqrt{(a + b)(c + d)(a + c)(b + d)}}$$

Perlu diketahui bahwa nilai phi mempunyai rentang diantara – 1 dan 1. Nilai phi tersebut didapat dari perhitungan chi square. Hubungannya dengan chi square adalah seperti rumus di bawah ini:

$$\phi^2 = \frac{X^2}{n} \text{ atau } X^2 = n\phi^2$$

Yang mempunyai sebaran khi-kuadrat dengan 1 derajat bebas (degree of freedom).

NB: Degree of freedom = (c - 1) x (r - 1).

Dimana:

c: Jumlah kolom.

r: Jumlah baris.

Oleh karena masing-masing variabel, baik variabel bebas maupun terikat mempunyai dua kategori, maka  $c = 2-1=1$  dan  $r = 2-1=1$ . Sehingga degree of freedom =  $1 \times 1 = 1$ .

Nilai degree of freedom ini berguna dalam menentukan nilai tabel chi square dalam menentukan tingkat signifikansi chi square. Untuk lebih jelasnya silahkan baca artikel kami tentang chi square.

### **Contoh Soal**

Cara Mencari Angka Indeks Korelasi Phi dengan mendasarkan diri pada frekuensi masing-masing sel yang terdapat dalam Tabel Kerja (Tabel Perhitungan)

Misalnya dalam suatu kegiatan penelitian yang bertujuan untuk mengetahui apakah secara signifikan terdapat korelasi antara kegiatan mengikuti bimbingan tes yang dilakukan oleh para siswa lulusan SMA dan prestasi mereka dalam Tes SPMB, yang telah ditetapkan jumlah pesertanya 100 orang. Berikut adalah datanya :

Status Prestasi	Mengikuti Bimbingan Tes	Tidak Mengikuti Bimbingan Tes	Jumlah
Lulus Tes SPMB	20	20	40
Tidak Lulus Tes SPMB	25	35	60
Jumlah	45	55	100 = N

Penerimaan calon mahasiswa baru (SPMB), dalam penelitian mana telah diterapkan sampel sejumlah 100 orang lulusan SMTA berhasil diperoleh data sebagaimana tertera pada table diatas.

Kita rumuskan terlebih dahulu  $H_a$  dan  $H_o$  nya :

$H_a$  : ada korelasi yang signifikan antara keikutsertaan para lulusan SMTA dalam bimbingan tes dan keberhasilan mereka dalam tes SPMB

$H_o$  : tidak ada korelasi yang signifikan antara keikutsertaan para lulusan SMTA dalam bimbingan tes dan keberhasilan mereka dalam tes SPMB

Karena Phi disini akan dihitung berlandaskan pada frekuensi selnya, maka masing-masing sel yang terdapat pada table diatas itu kita persiapkan lebih dahulu menjadi tabel perhitungan.

Disini kita lihat: frekuensi sel  $a=20$ ;  $b=20$ ;  $c=25$ ; dan  $d=35$ .

- Rumus yang kita pergunakan adalah

$$\phi = \frac{(ad-bc)}{a+b+c+d} \sqrt{\frac{1}{n}}$$

Status Prestasi	Mengikuti Bimbingan Tes	Tidak Mengikuti Bimbingan Tes	Jumlah
Lulus Tes SPMB	20 a	20 b	40
Tidak Lulus Tes SPMB	25 c	35 d	60
Jumlah	45	55	100 = N

$$\phi = \frac{(20 \times 35 - 20 \times 25)}{20+20+25+35} \sqrt{\frac{1}{100}}$$

$$= \frac{700 - 500}{90} \sqrt{\frac{1}{100}} = \frac{200}{90} \times 0,1 = 0,222 \times 0,1 = 0,0222$$

- interpretasi;  $\phi$  disini di anggap sebagai  $r_{xy}$

$df = N - nr = 100 - 2 = 98$  ( konsultasi tabel nilai “r”) dalam tabel tidak dijumpai  $df$  sebesar 98 karena itu kita pergunakan  $df$  sebesar 100. dengan  $df$  sebesar 100, di peroleh  $r_{tabel}$  pada taraf signifikan 5% = 0,195, sedangkan pada taraf signifikan 1% = 0,254. dengan demikian  $\phi$  yang di peroleh (yaitu: 0,082) adalah lebih kecil jika di bandingkan dengan  $r_{tabel}$  (yaitu: 0,195 dan 0,254). dengan demikian hipotesis  $H_0$  diterima/disetujui. berarti tidak terdapat korelasi yang signifikan antara keikutsertaan siswa lulusan SMA dengan kegiatan bimbingan tes dan prestasi yang mereka. **Jadi** dapat di tarik sebuah kesimpulan bahwa keberhasilan para siswa lulusan SMA dalam tes SPMB itu secara signifikan tidak ada hubungannya (tidak di pengaruhi) oleh ikut tidaknya mereka dalam kegiatan Bimbingan Tes Masuk Perguruan Tinggi.



- Cara Mencari Angka Indeks Korelasi Phi dengan mendasarkan diri pada Nilai Proporsinya.

Status Prestasi	Mengikuti Bimbingan Tes	Tidak Mengikuti Bimbingan Tes	Jumlah
Lulus Tes SPMB	20 $\alpha = 20/100 = 0,200$	20 $\beta = 20/100 = 0,200$	40 $p = 0,400$
Tidak Lulus Tes SPMB	25 $\gamma = 25/100 = 0,250$	35 $\delta = 35/100 = 0,350$	60 $q = 0,600$
Jumlah	45 $p' = 0,450$	55 $q' = 0,550$	100 = 1,000

Rumus yang digunakan adalah :  $\phi = \alpha\delta - \beta\gamma / pp'(q')$

Dengan menggunakan contoh sebelumnya, maka tabel yang diperlukan adalah :

Diketahui :  $(\alpha) = 0,200$  ;  $(\beta) = 0,200$  ;  $(\gamma) = 0,250$  ;  $(\delta) = 0,350$

Kita masukkan dalam rumus:

$$\begin{aligned} \phi &= \alpha\delta - \beta\gamma / pp'(q') \\ &= 0,200 \cdot 0,350 - 0,200 \cdot 0,250 / 0,400 \cdot 0,600 / 0,450 \cdot 0,550 \\ &= 0,07 - 0,050 / 0,594 = 0,020 / 0,244 \\ &= 0,082 \end{aligned}$$

- Cara Mencari (Menghitung) Angka Indeks Korelasi Phi dengan memperhitungkan Kai Kuadrat

Kai Kuadrat di sini sekedar diperkenalkan sebagai suatu proses perhitungan atau pengolahan data. Jika perhitungan  $\phi$  didasarkan pada harga Kai Kuadrat maka menggunakan rumus sebagai berikut :  $\phi = \frac{x^2}{N}$

Dengan menggunakan contoh awal, maka untuk memperoleh harga Phi dengan menggunakan Kai Kuadrat, Tabel dan Proses perhitungannya adalah sebagai berikut :

Status Prestasi	Mengikuti Bimbingan Tes	Tidak Mengikuti Bimbingan Tes	Jumlah
Lulus Tes SPMB	20	20	40 = r <sub>N</sub>
Tidak Lulus Tes SPMB	25	35	60 = r <sub>N</sub>
Jumlah	45 = c <sub>N</sub>	55 = c <sub>N</sub>	100 = N

Seperti telah dikemukakan sebelumnya, maka rumus untuk mencari Kai Kuadrat adalah:  $X^2 = \frac{\sum (f_o - f_t)^2}{f_t}$

**Cara menghitungnya :**

Dengan demikian,  $\phi$  dapat kita peroleh dengan jalan mensubstitusikan harga Kai. Kuadrat ke dalam rumus Phi :

$$\phi = \frac{X^2}{N} = \frac{0,6733100}{100} = 0,006733 = 0,082$$

Sel	f <sub>o</sub>	f <sub>t</sub> = c <sub>N</sub> X r <sub>NN</sub>	(f <sub>o</sub> - f <sub>t</sub> )	(f <sub>o</sub> - f <sub>t</sub> ) <sup>2</sup>	(f <sub>o</sub> - f <sub>t</sub> ) / f <sub>t</sub>
1	20	45 X 40/100 = 18	+2	4	0,2222
2	20	55 X 60/100 = 33	-13	169	-0,2364
3	25	45 X 60/100 = 27	-2	4	-0,0741
4	35	55 X 60/100 = 33	+2	4	0,0606

Jumlah	100 = N	100 = N	0	-	0,6733 = (fo-ft)2ft
--------	---------	---------	---	---	------------------------

### b) Korelasi Koefisien Kontingency C

Koefisien kontingensi adalah suatu ukuran kadar asosiasi relasi antara dua himpunan atribut. Ukuran ini berguna khususnya apabila kita hanya mempunyai informasi kategori (skala nominal) mengenai satu diantara himpunan-himpunan atribut atau kedua himpunan atribut tersebut. Yaitu, pengukuran ini dapat dipergunakan kalau informasi kita tentang atribut-atribut itu terdiri dari suatu rangkaian frekuensi yang tidak berurut.

Dalam menggunakan koefisien kontingensi, tidak perlu membuat anggapan kontinuitas untuk berbagai kategori yang dipergunakan untuk mengukur salah satu atau kedua himpunan. Koefisien kontingensi, yang dihitung dari suatu tabel kontingensi, akan mempunyai harga yang sama bagaimanapun kategori kategori itu tersusun dalam baris-baris dan kolom-kolom.

Koefisien korelasi kontingensi ini digunakan untuk mengukur keeratan hubungan antara dua variabel yang datanya berbentuk data nominal (data kualitatif). Disimbolkan dengan "C" dan dirumuskan:

Uji ini sangatlah erat kaitannya dengan uji chi-square. Sebab berdasarkan rumus uji koefisien ini, bahwa tidaklah mungkin koefisien ini dapat dihitung tanpa terlebih dahulu mengetahui nilai dari chi-square. Jadi, logikanya adalah hitung terlebih dahulu chi-square, baru kemudian hitung koefisien kontingensi.

- Rumus Koefisien Kontingensi

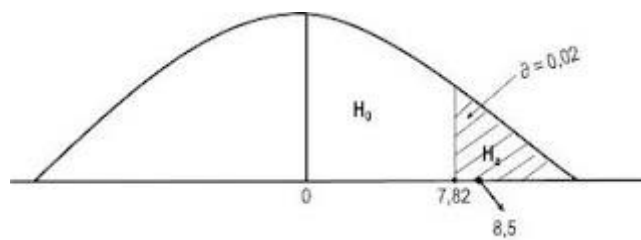
$$C = \sqrt{\frac{X^2}{N + X^2}}$$

- Rumus Kontingensi C

$$X^2 = \sum_{i=1}^b \sum_{j=1}^k \frac{(A_{ij} - H_{ij})^2}{H_{ij}}$$

- Besarnya koefisien kontingensi:

$$C = \sqrt{\frac{8,5}{96 + 8,5}}$$



- Daerah Penerimaan Koefisien Kontingensi

Dengan mempelajari tutorial di atas, diharapkan para pembaca memahami sebenarnya cara perhitungan atau rumus uji dalam tutorial ini. Selanjutnya dalam prakteknya, para peneliti atau pembaca dapat langsung menggunakan aplikasi SPSS. Dengan menggunakan SPSS, kita bisa langsung menilai tingkat signifikansi atau kemaknaan nilai koefisien. Yaitu, apabila nilai signifikansi < 0,05 atau batas kritis, maka dapat diartikan bahwa: terima H<sub>1</sub> atau bermakna secara statistik.

- Nilai Signifikansi dalam uji ini adalah berdasarkan perhitungan uji chi square. Jenis uji yang lain dan mirip dengan uji ini antara lain: Uji koefisien cramer, lambda, phi dan uncertainty coefficient.

Untuk menghitung koefisien kontingensi antara skor-skor dua himpunan kategori, misal A<sub>1</sub>, A<sub>2</sub>, A<sub>3</sub>,..., A<sub>k</sub> dan B<sub>1</sub>, B<sub>2</sub>, B<sub>3</sub>,..., B<sub>r</sub>. Dapat menyusun frekuensi frekuensinya dalam suatu tabel kontingensi, pada tabel. Dalam tabel semacam ini dapat memasukkan frekuensi yang diharapkan untuk tiap sel (E<sub>ij</sub>) dengan menentukan frekuensi manakah akan terjadi seandainya tidak terdapat asosiasi atau korelasi antara kedua variabel. Semakin besar perbedaan antara harga-harga sel yang

diobservasi, makin besar pula tingkat asosiasi antara kedua variabel dan dengan demikian semakin tinggi harga C.

Tabel : bentuk tabel kontingensi, untuk menghitung C

Baris	Kolom				Total
	A1	A2	...	Ak	
B1	(A1,B1)	(A2,B1)	...	(Ak,B1)	...
B2	(A1,B2)	(A2,B2)	...	(Ak,B2)	...
...	...	...	...	...	...
Br	(A1,Br)	(A2,Br)	...	(Ak,Br)	...
Total	...	...	....	...	N

Tingkat asosiasi antara dua himpunan, entah berurut atau tidak, dan tidak terpengaruh sifat hakekat variabelnya (dapat kontinu atau diskrit) atau tidak terpengaruh oleh distribusi yang mendasari (distribusi bisa saja normal atau senbarang bentuk distribusi lain), dapat diketahui dari suatu tabel kontingensi dengan,

#### d) Korelasi Rank Spearman

Koefisien korelasi spearman merupakan statistik nonparametrik. Statistik ini merupakan suatu ukuran asosiasi atau hubungan yang dapat digunakan pada kondisi satu atau kedua variabel yang diukur adalah skala ordinal (berbentuk ranking) atau kedua variabel adalah kuantitatif namun kondisi normal tidak terpenuhi. Simbol ukuran populasinya adalah  $\rho$  dan ukuran sampelnya  $r_s$ . Formula  $r_s$  untuk korelasi Spearman adalah sebagai berikut:

$$r_s = 1 - \frac{6 \sum_{i=1}^N d_i^2}{N^3 - N}$$

dimana:

$d_i$  adalah perbedaan antara kedua ranking

$N$  adalah banyaknya observasi.

Pembuatan ranking dapat dimulai dari nilai terkecil atau nilai terbesar tergantung permasalahannya. Bila ada data yang nilainya sama, maka pembuatan ranking didasarkan pada nilai rata-rata dari ranking-

ranking data tersebut. Apabila proporsi angka yang sama tidak besar, maka formula diatas masih bisa digunakan. Namun apabila proporsi angka yang sama cukup besar, maka dapat digunakan suatu faktor koreksi dan formula menjadi seperti berikut ini:

$$r_s = \frac{2 \left( \frac{N^3 - N}{12} \right) - \sum T_1 - \sum T_2 - \sum d_i^2}{2 \sqrt{\left( \frac{N^3 - N}{12} - \sum T_1 \right) \left( \frac{N^3 - N}{12} - \sum T_2 \right)}} \quad \text{dimana:} \quad T = \frac{t^3 - t}{12}$$

$t$  adalah banyaknya observasi yang berangka sama pada suatu ranking tertentu.

Jika dilakukan secara manual, maka tata tertib melakukan uji korelasi Spearman adalah:

- 1) Jumlahkan skor item-item di tiap variabel untuk mendapatkan skor total variabel (misalnya cari skor total variabel X dengan menotakan item-item variabel X).
- 2) Lakukan rangkin skor total x ( $r_x$ ) dan rangking skor total y ( $r_y$ ).
- 3) Cari nilai d yaitu selisih  $r_x - r_y$ .
- 4) Cari nilai  $d^2$  yaitu kuadrat d (selisih  $r_x - r_y$ ).

### Contoh Soal

Responde n	Item X #1	...	Item X #n	Total X	Item Y #1	...	Item Y #n	Total Y	$r_x$	$r_y$	d	$D^2$
1	5		4	65	5		4	57	1	1	0	0
2	4		4	62	4		4	55	4	3	1	1
3	4		3	61	3		3	50	5	6	-1	1
4	5		4	63	3		5	52	3	5	-2	4
5	4		5	64	2		4	54	2	4	-2	4
....												
30	4		5	60	5		3	56	6	2	4	16
total												2125

Setelah data dihitung dalam tabel, lalu masukkan ke dalam rumus uji korelasi Spearman:

$$r_s = 1 - \frac{6 \sum d^2}{n(n^2 - 1)} = 1 - \frac{6.2125}{30(30^2 - 1)} = 1 - \frac{12750}{30(900 - 1)} = 1 - \frac{12750}{30.799} = 0,47$$

Dengan demikian korelasi Spearman ( $r_s$ ) variabel x dengan variabel y dalam contoh adalah 0,47. Nilai korelasi Spearman hitung ini

(rs) lalu diperbandingkan dengan Spearman Tabel (rs tabel). Keputusan diambil dari perbandingan tersebut. Jika  $rs > rs \text{ tabel}$ ,  $H_0$  ditolak dan  $H_1$  diterima. Jika  $rs \text{ hitung} \leq rs \text{ tabel}$ ,  $H_0$  diterima,  $H_1$  ditolak. Pengambilan keputusan dari contoh di atas adalah karena  $rs \text{ hitung} > rs \text{ tabel}$  maka  $H_0$  ditolak dan  $H_1$  diterima. Artinya terdapat hubungan antara variabel x dengan variabel y. Lalu, bagaimana menginterpretasikannya?

### 1) Data Saya Lebih Besar dari 30 !

Rumus di atas berlaku jika jumlah sampel lebih kecil atau sama dengan 30 ( $\leq 30$ ). Lalu, bagaimana menghitung uji korelasi Spearman dengan lebih dari 30 sampel? Caranya mudah saja yaitu Cari Nilai z hitung terlebih dahulu.

Cara mencari nilai z hitung sebagai berikut:

$$z = rs\sqrt{n - 1}$$

Di mana:

*z = Nilai z hitung*

*rs = Koefisien Korelasi Spearman*

*n = Jumlah Sampel Penelitian*

Nilai rs dicari dengan cara yang sama seperti perhitungan terdahulu (di bagian atas). Dalam contoh sampel yang lebih besar dari 30 ini misalnya sampel menggunakan 50 responden. Maka perhitungannya sebagai berikut:

$$z = 0,99\sqrt{50 - 1} = 0,99\sqrt{49} = 0,99 \cdot 7 = 6,93$$

Nilai z hitung dalam sampel  $> 30$  ini adalah 6,93. Pengambilan keputusan dalam sampel  $> 30$  ini adalah membandingkan antara z hitung dengan z tabel. Z hitung sudah diperoleh sekarang tinggal z tabel.

## 2) Cara Mencari z Tabel

Nilai z tabel dicari dari tabel Z (lihat buku-buku statistik). Caranya adalah:

- a) Tentukan Taraf Keyakinan Penelitian (misalnya 95%). Taraf Keyakinan 95% berarti Interval Keyakinan-nya (alpha) 0,05. Nilai 0,05 ini merupakan bentuk desimal dari 5% yang diperoleh dari pengurangan 100% selaku kebenaran absolut dengan 95% ( $100\% - 95\% = 5\%$  atau 0,05).
- b) Tentukan Uji yang digunakan. Apakah 1 sisi (One-Tailed) atau 2 sisi (Two-Tailed). Penentuan 1 sisi atau 2 sisi ini didasarkan hipotesis penelitian. Jika hipotesis hanya menyebutkan “terdapat hubungan” maka artinya bentuk hubungan belum ditentukan apakah positif atau negatif dan dengan demikian menggunakan uji 2 sisi. Jika hipotesis menyatakan “terdapat hubungan positif” atau “terdapat hubungan negatif” maka artinya bentuk hubungan sudah ditentukan dan dengan demikian menggunakan uji 1 sisi.
- c) Jika Uji 2 Sisi (Two-Tailed) maka lihat Tabel Z. Dalam uji 2 sisi Interval Keyakinan dibagi dua yaitu  $0,05 / 2 = 0,025$ . Cari pada kolom tabel nilai yang paling mendekati 0,025. Dari nilai yang paling dekat tersebut tarik garis ke kiri sehingga bertemu dengan nilai  $1,9 + 0,060 = 1,96$ . Batas kiri pengambilan keputusan dengan kurva adalah  $-1,96$  batas kanannya  $+1,96$ . Keputusannya: Tolak  $H_0$  dan Terima  $H_1$  jika  $-z$  hitung  $<$  dari  $-1,96$  dan  $>$  dari  $+1,96$ . Sebaliknya, Terima  $H_0$  dan Tolak  $H_1$  jika  $-z$  hitung  $>$   $-1,96$  dan  $<$  dari  $+1,96$ .

### Contoh Soal

Seorang manager produksi ingin mengetahui apakah ada hubungan antara nilai tes bakat (*aptitude test*) pada waktu penerimaan kerja dengan *rating* tampilannya setelah satu semester bekerja. Nilai *aptitude test* berkisar antara 0 sampai 100. Sedangkan *rating* tampilan mempunyai skala sebagai berikut:

1 = pekerja berpenampilan sangat dibawah rata-rata



2 = pekerja berpenampilan dibawah rata-rata

3 = pekerja berpenampilan sedang (rata-rata)

4 = pekerja berpenampilan diatas rata-rata

5 = pekerja berpenampilan sangat diatas rata-rata

<u>Pekerja</u>	<u>nilai test</u>	<u>rating</u>	<u>Pekerja</u>	<u>nilai test</u>	<u>rating</u>
1	59	3	11	48	3
2	47	2	12	65	3
3	58	4	13	51	2
4	66	3	14	61	3
5	77	2	15	40	3
6	57	4	16	67	4
7	62	3	17	60	2
8	68	3	18	56	3
9	69	5	19	76	3
10	36	1	20	71	5

Solusi:

<u>Pekerja</u>	<u>nilai test</u>	<u>rank(a)</u>	<u>rating</u>	<u>rank(b)</u>	$d_i$	$d_i^2$
1	59	9	3	10,5	-1,5	2,25
2	47	3	2	3,5	-0,5	0,25
3	58	8	4	17	-9	81
4	66	14	3	10,5	3,5	12,25
5	77	20	2	3,5	16,5	272,25
6	57	7	4	17	-10	100
7	62	12	3	10,5	1,5	2,25
8	68	16	3	10,5	5,5	30,25
9	69	17	5	19,5	-2,5	6,25
10	36	1	1	1	0	0
11	48	4	3	10,5	-6,5	42,25
12	65	13	3	10,5	2,5	6,25
13	51	5	2	3,5	1,5	2,25
14	61	11	3	10,5	0,5	0,25
15	40	2	3	10,5	-8,5	72,25
16	67	15	4	17	-2	4
17	60	10	2	3,5	6,5	42,25
18	56	6	3	10,5	-4,5	20,25
19	76	19	3	10,5	8,5	72,25
20	71	18	5	19,5	-1,5	2,25
					$\Sigma$	771

$$r_s = 1 - \frac{6(771)}{20^3 - 20} = 0,4203$$

Jika menggunakan faktor koreksi  $\sum T_i = 0$  karena tidak ada nilai yang sama pada variabel nilai test, sedangkan untuk variabel rating  $\sum T_i = 90$  (nilai 2 ada 4, nilai 3 ada 10, nilai 4 ada 3, dan nilai 5 ada 2). Sehingga nilai  $r_s$  adalah sebesar:

$$r_s = \frac{2\left(\frac{20^3 - 20}{12}\right) - 0 - 90 - 771}{2\sqrt{\left(\frac{20^3 - 20}{12} - 0\right)\left(\frac{20^3 - 20}{12} - 90\right)}} = 0,3792 .$$

**Uji untuk menentukan adanya hubungan antara dua variabel.**

$$H_0 : \rho = 0 \quad \text{VS} \quad H_1 : \rho \neq 0$$

Untuk sampel kecil ( $n \leq 30$ )  $\longrightarrow$  gunakan tabel koefisien korelasi Spearman.  
Keputusan: tolak  $H_0$  jika  $r_s < -r_{\text{tabel}}$  dan  $r_s > r_{\text{tabel}}$  (pada  $n$  dan  $\alpha$  tertentu)

Untuk sampel besar ( $n > 30$ ) dapat didekati dengan distribusi normal

$$z = \frac{r_s - 0}{\frac{1}{\sqrt{n-1}}} = r_s \sqrt{n-1}$$

Daerah tolak  $H_0$ :  $Z < -Z_{\alpha/2}$  atau  $Z > Z_{\alpha/2}$

## D. PROSEDUR KERJA DAN INTERPRETASI OUTPUT DI SPSS

### 1. Korelasi Untuk Uji Parametrik

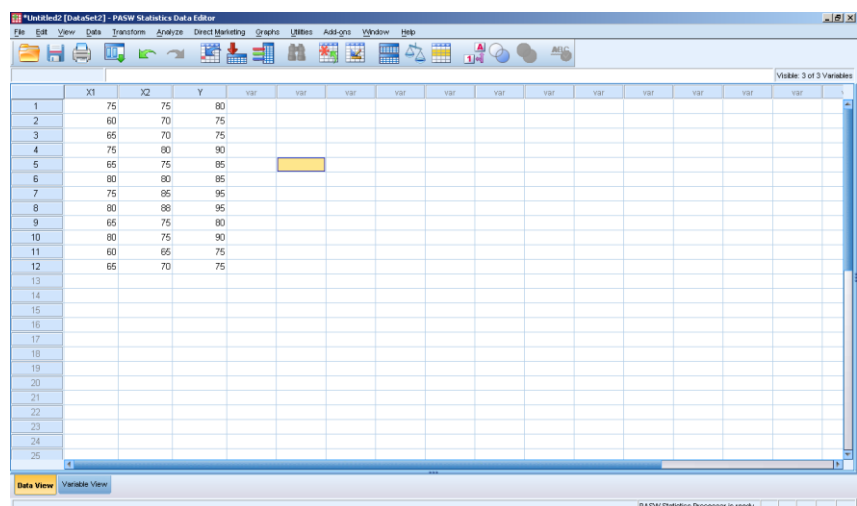
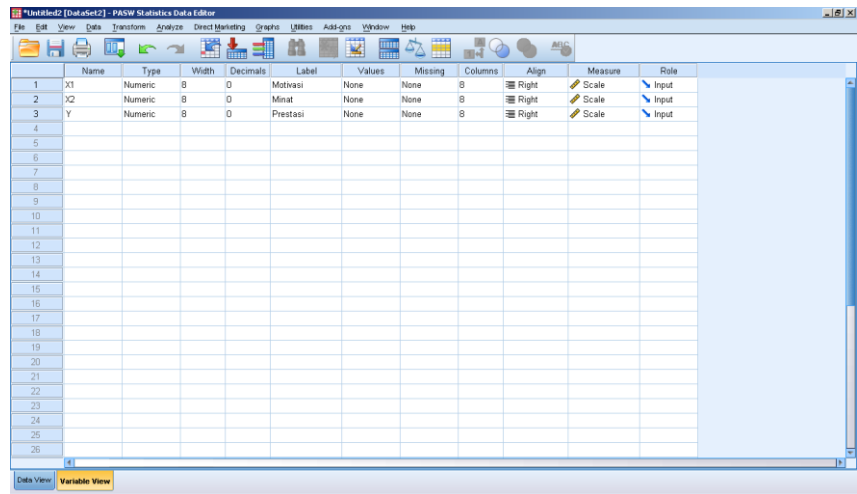
#### a) Asumsi Uji Korelasi

Sebelum diimplementasi, uji Korelasi terlebih dulu harus memenuhi serangkaian asumsi. Asumsi-asumsi uji Korelasi adalah:

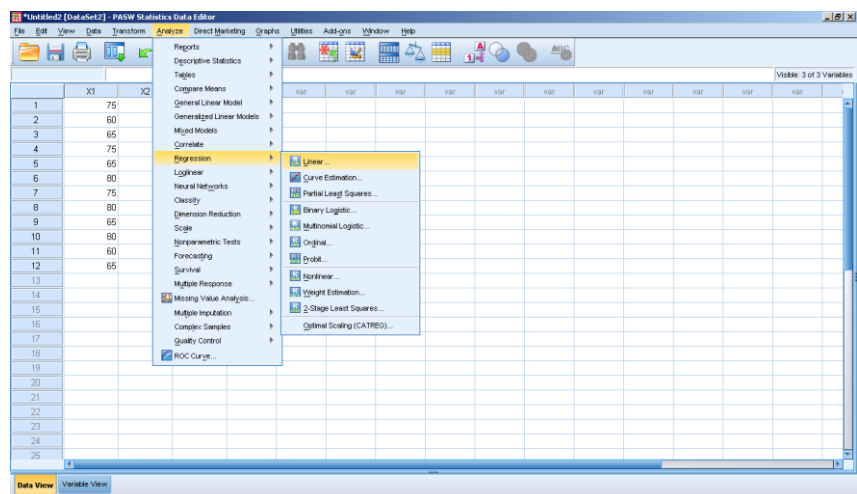
- Normalitas. Artinya, sebaran variabel-variabel yang hendak dikorelasikan harus berdistribusi normal.
- Linearitas. Artinya hubungan antara dua variabel harus linier. Misalnya ditunjukkan lewat straight-line.
- Homoskedastisitas. Artinya, variabilitas skor di variabel Y harus tetap konstan di semua nilai variabel X.

## a) Cara uji normalitas

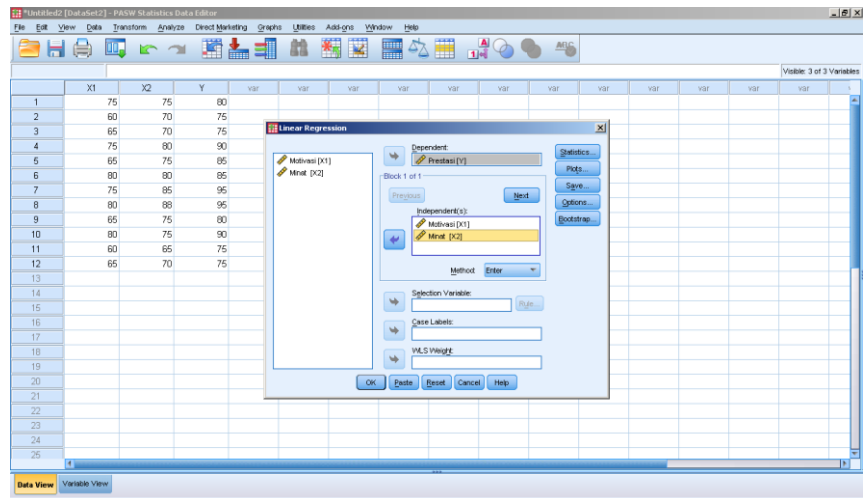
- Masukan data variabel yang akan di teliti



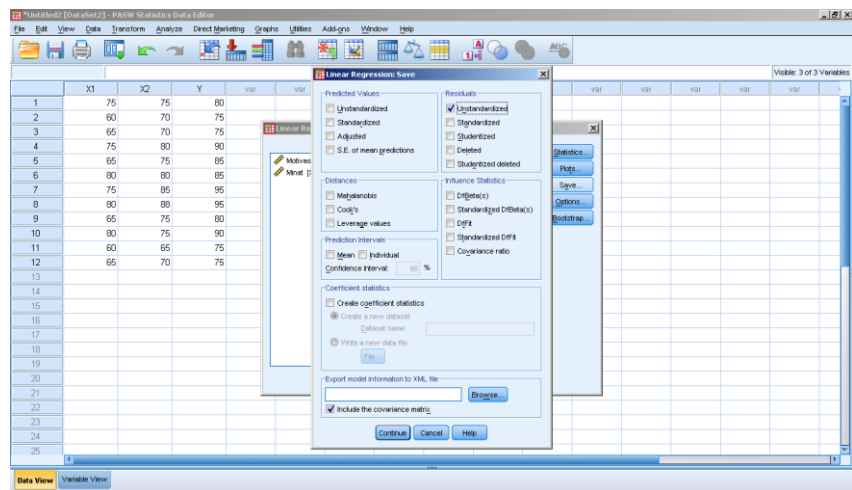
- klik Analyze – Regression – Linier untuk melakukan uji normalitas



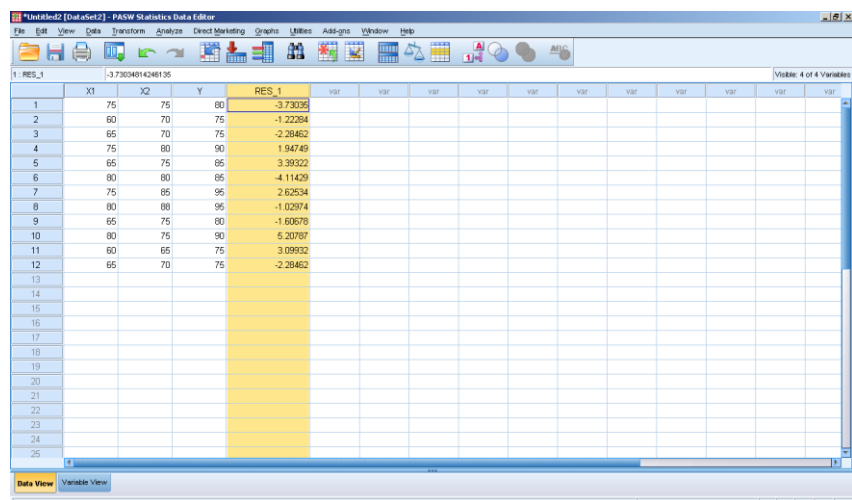
- Variabel Y dimasukan dalam Dependent, dan X masukan dalam Independen



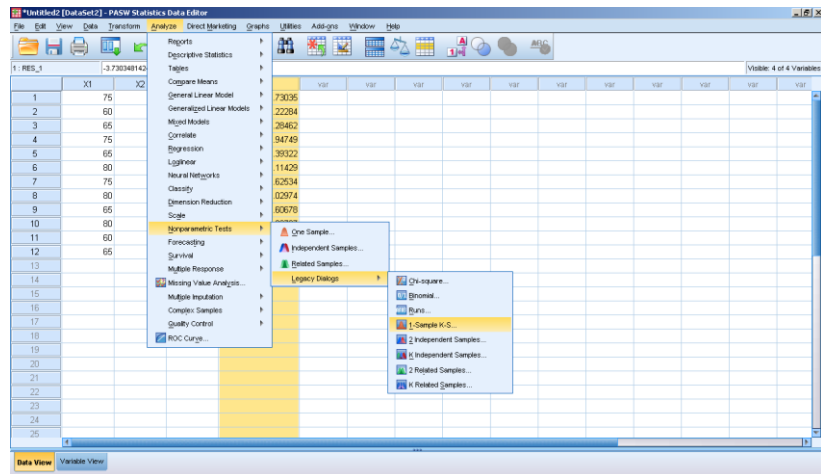
- Klik Save kemudian pilih cetang unstandadized



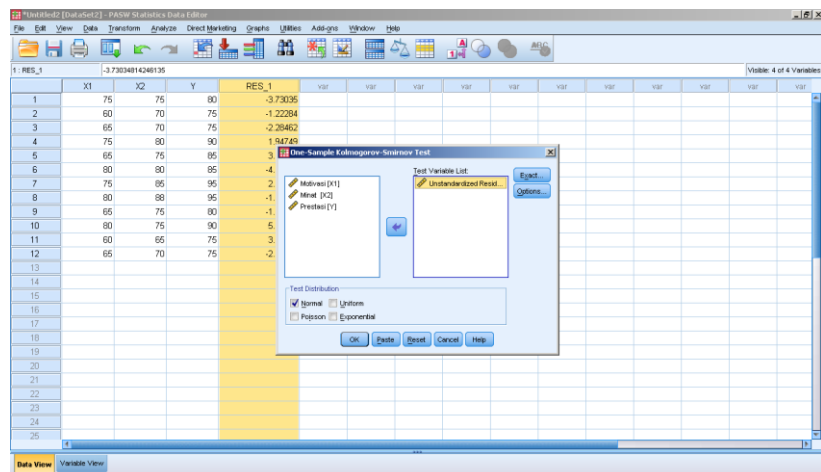
- Setelah itu akan muncul RES yang akan di uji normalitas



- Uji normalitas dengan memilih nonparametric test – legacy dialog – sampel s-k



- Masukkan RES dalam kotak sebelah kanan



▪ **Output**

One-Sample Kolmogorov-Smirnov Test		Unstandardized Residual
N		12
Normal Parameters <sup>a,b</sup>	Mean	.0000000
	Std. Deviation	3.09465898
Most Extreme Differences	Absolute	.214
	Positive	.214
	Negative	-.152
Kolmogorov-Smirnov Z		.740
Asymp. Sig. (2-tailed)		.644
a. Test distribution is Normal.		
b. Calculated from data.		

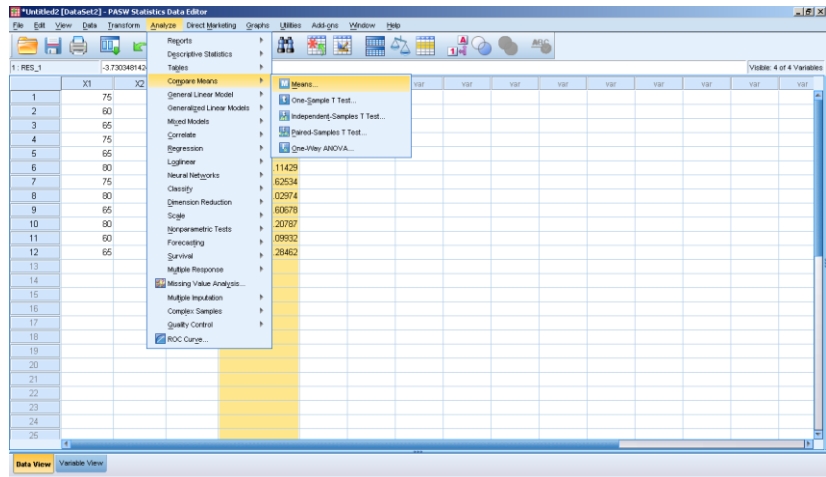
Dari sini yang dibaca adalah Asymp Sig (2-tailed) jika nilai

- sig > 0,05 maka data dikatakan normal

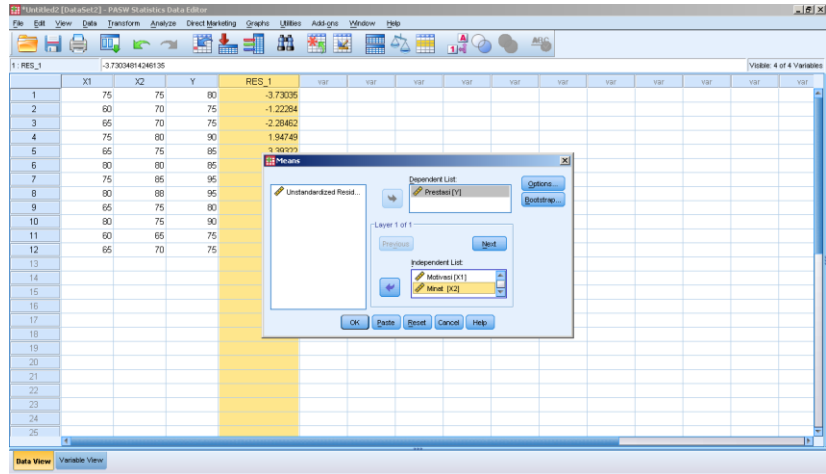
- sebaliknya jika  $\text{sig} < 0,05$  maka data dikatakan tidak normal

### b) Cara uji linieritas

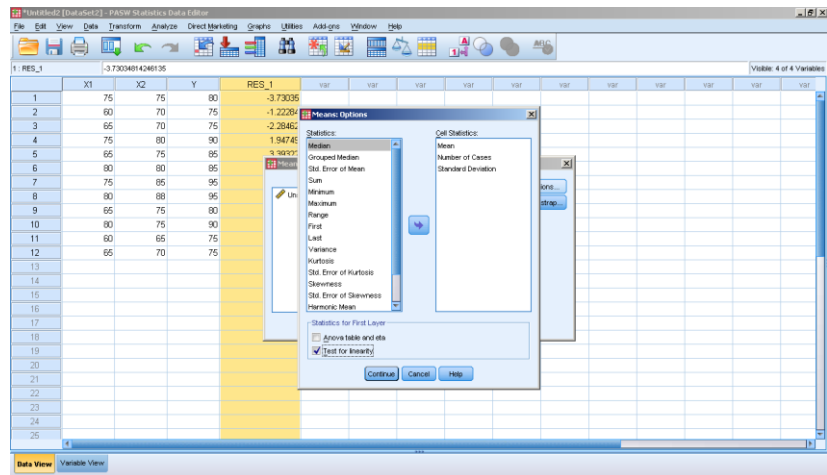
- Dari data yang sudah di masukan dalam spss, maka klik Compare Means – Means



- Masukan Y dalam variabel dependen dan X dalam variabel independen



- Klik Options kemudian perhatikan Cells Statistics dan klik *Test for linearity* pada bagian bawah



▪ **Output**

ANOVA Table							
			Sum of Squares	df	Mean Square	F	Sig.
Prestasi * Minat	Between Groups	(Combined)	585.417	5	117.083	8.646	.010
		Linearity	549.801	1	549.801	40.601	.001
		Deviation from Linearity	35.616	4	8.904	.658	.643
	Within Groups		81.250	6	13.542		
	Total		666.667	11			

Cara membaca tabel pada diatas ada 2 step yaitu :

1) Membandingkan nilai signifkansi (Sig) dengan 0,05

- Jika nilai Deviation from Linearity sig > 0,05, maka ada hubungan yang linier secara signifikan antara variabel independen dengan variabel dependen dengan variabel dependen.
- Jika nilai Deviation from Linearity sig < 0,05, maka tidak ada hubungan yang linier secara signifikan antara variabel independen dengan variabel dependen dengan variabel dependen.

2) Membandingkan nilai F hitung dengan F tabel

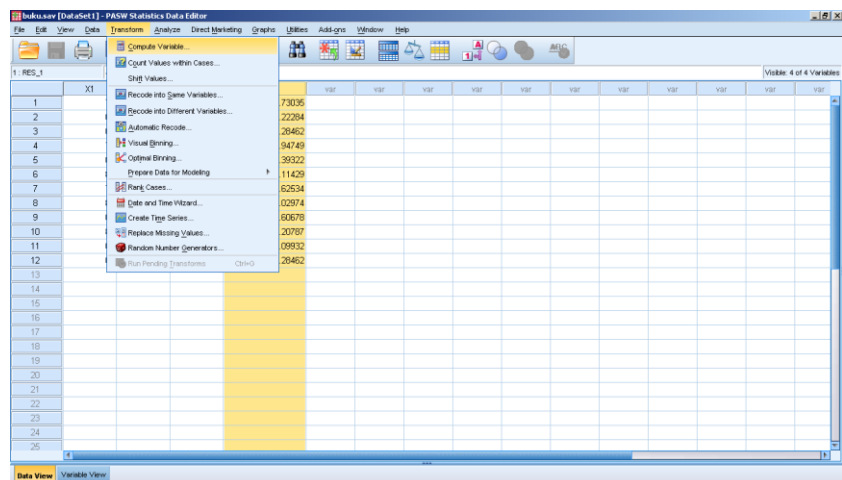
- Jika nilai F hitung < F tabel, maka ada hubungan yang linier secara signifikan antara variabel independen dengan variabel dependen.

- Jika nilai F hitung > F tabel, maka tidak ada hubungan yang linier secara signifikan antara variabel independen dengan variabel dependen.

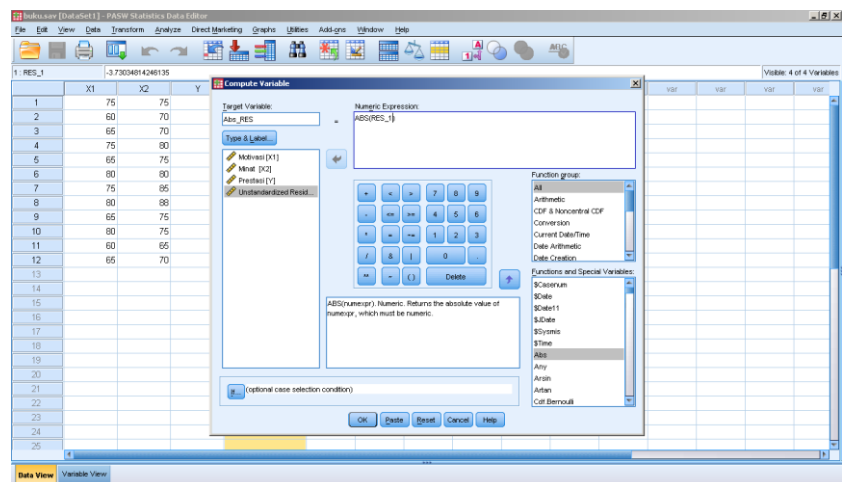
Jadi dari hasil diatas dapat disimpulkan nilai sig (0,643) > 0,05 maka ada hubungan yang linier secara signifikan. Dan dari hasil F (0,658) < F tabel (0,322) maka ada hubungan yang linier secara seignifikan.

### c) Cara Uji Homoskedastisitas

- Hasil RES yang akan di ujikan, klik Transform – Compare Means



- Di dalam Target Variabel di ketik Abs\_RES, kemudian klik bagian Function Group di pilih “All” kemudian bawahnya pilih “Abs” klik 2 kali kemudian masukan RES dibagian Type & Label ke kolom Numeric, klik OK

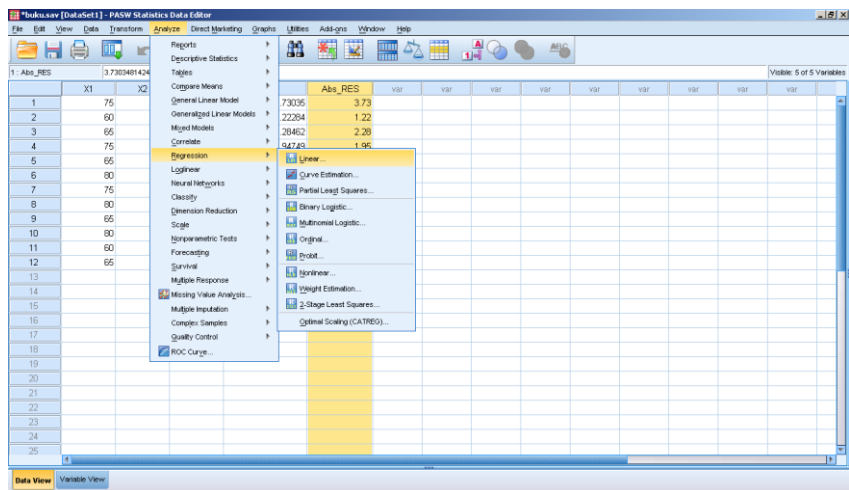




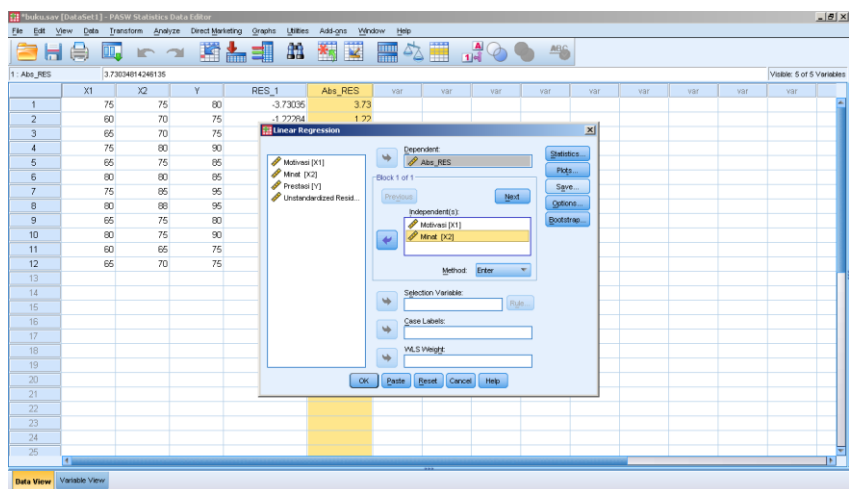
- Pada data variabel akan muncul Abs\_RES

	X1	X2	Y	RES_1	Abs_RES
1	75	75	80	-3.73035	3.73
2	60	70	75	-1.22284	1.22
3	65	70	75	-2.28462	2.28
4	75	80	90	-1.94749	1.95
5	65	75	85	3.39322	3.39
6	80	80	85	-4.11429	4.11
7	75	85	95	2.62534	2.63
8	80	88	95	-1.02974	1.03
9	65	75	80	-1.60678	1.61
10	80	75	90	5.20787	5.21
11	60	65	75	3.05932	3.10
12	65	70	75	-2.28462	2.28

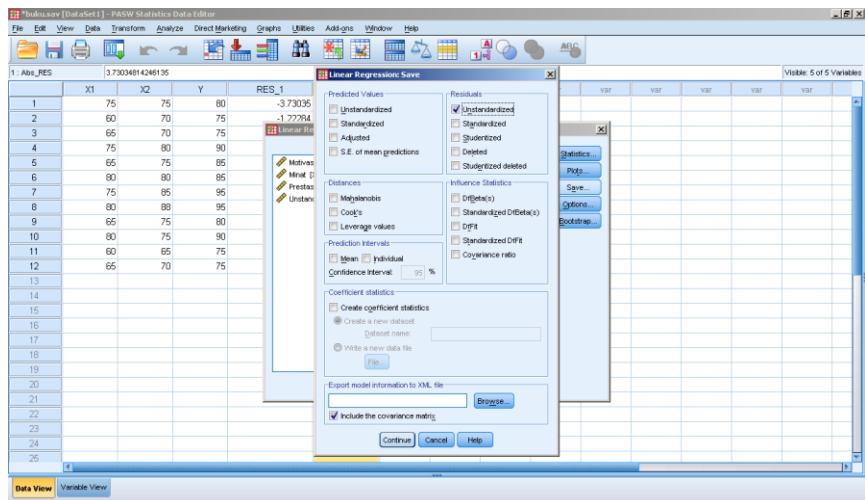
- Klik Analyze – Regression \_ Linier



- Masukkan Abs\_RES kedalam kolom Dependen dan masukan Motivasi dan Minat – kemudian klik “Save”



- Centang “Unstandardized” pada residuals – klik continue



### ■ Output

Model		Coefficients <sup>a</sup>				
		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	4.428	2.924		1.514	.164
	Motivasi	.201	.053	1.259	3.765	.004
	Minat	-.209	.063	-1.120	-3.348	.009

a. Dependent Variable: Abs\_RES

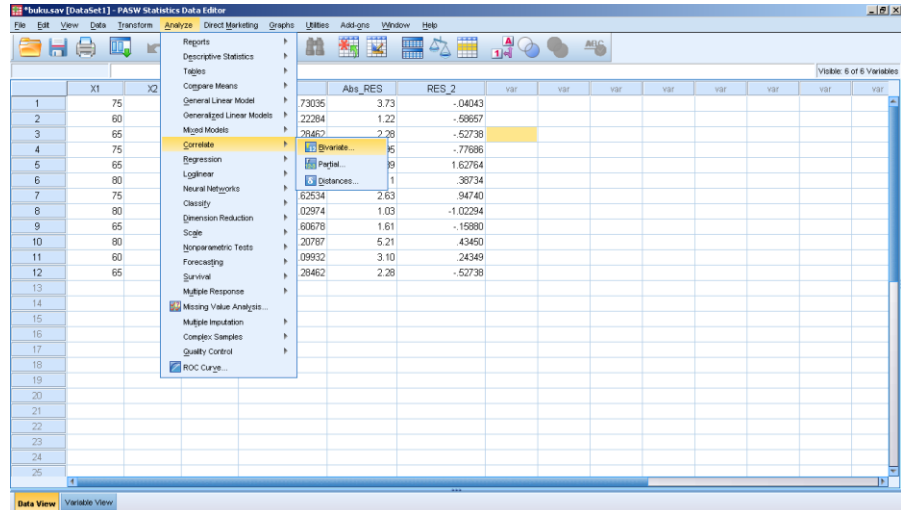
Adapun dasar pengambilan keputusan dalam uji heteroskedasitas dengan menggunakan uji glejser adalah sebagai berikut :

- 1) Jika nilai signifikansi (Sig) > 0.05, maka kesimpulannya adalah tidak terjadi gejala heteroskedasitas dalam model regresi.
- 2) Jika nilai signifikansi (Sig) < 0.05, maka kesimpulannya adalah terjadi gejala heteroskedasitas dalam model regresi.

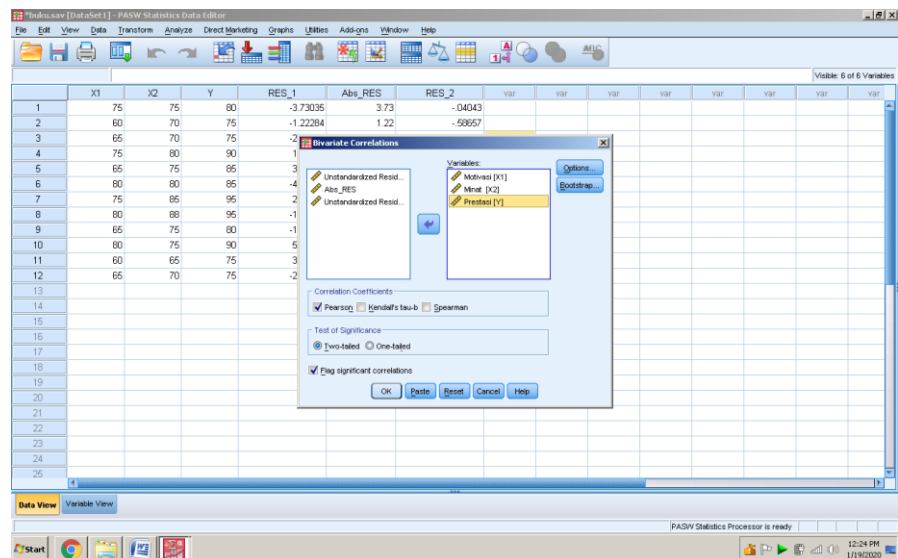
Jadi dalam penelitian dapat disimpulkan bahwa nilai Sig Motivasi (0.004) dan nilai Sig Minat (0.009) < 0.05 maka terjadi gejala heteroskedasitas dalam model regresi.

## e) Uji Pearson Product Moment dan Asumsi Klasik

- Data yang sudah di entry kemudian klik Analyze – Correlation – Bivariate



- Kemudian masukan variabel X dan Y ke dalam kolom “Variable” dan centang “Pearson”, “Two Tailed” dan bagian terbawah – OK



### ■ Output

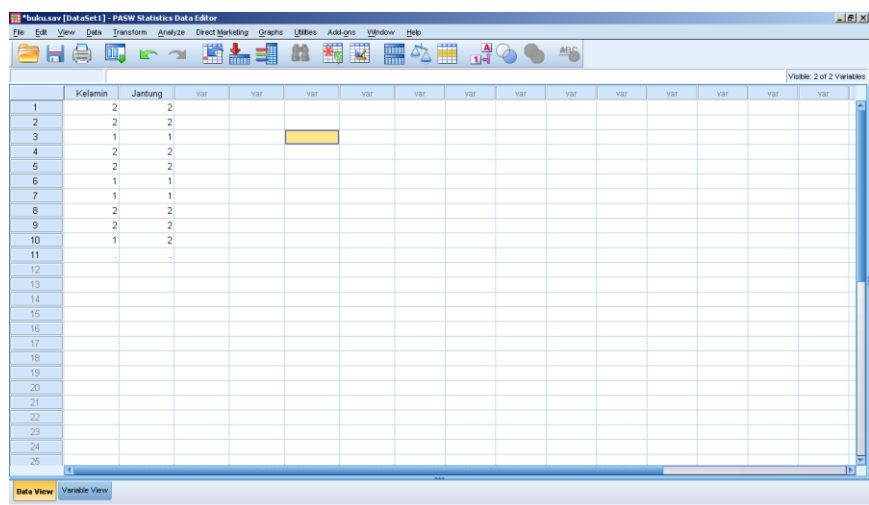
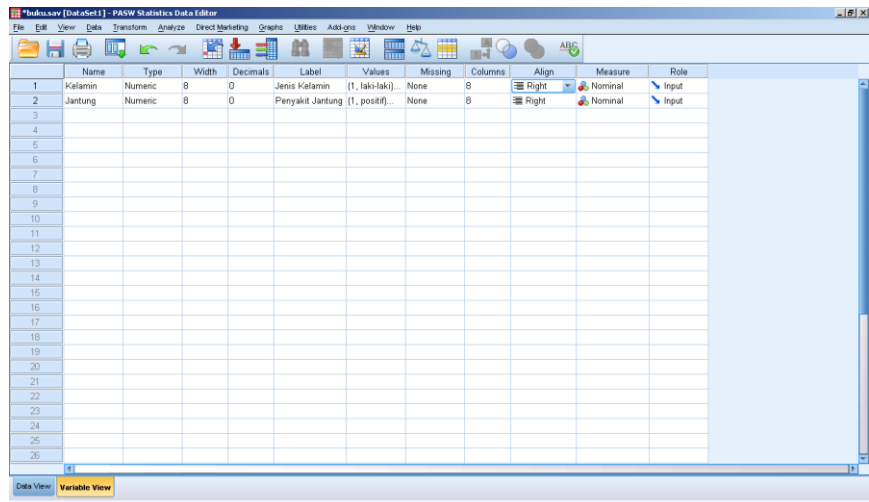
Correlations				
		Motivasi	Minat	Prestasi
Motivasi	Pearson Correlation	1	.788**	.796**
	Sig. (2-tailed)		.002	.002
	N	12	12	12

Minat	Pearson Correlation	.788**	1	.908**
	Sig. (2-tailed)	.002		.000
	N	12	12	12
Prestasi	Pearson Correlation	.796**	.908**	1
	Sig. (2-tailed)	.002	.000	
	N	12	12	12
**. Correlation is significant at the 0.01 level (2-tailed).				

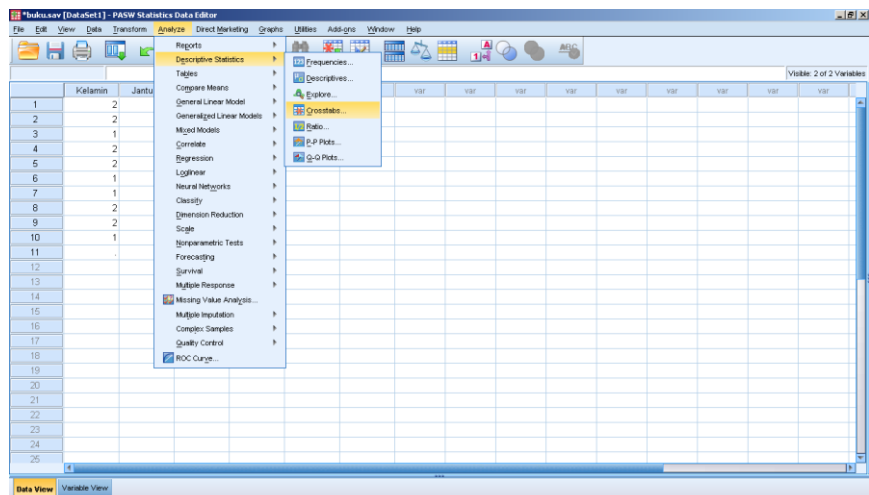
- Interpretasi analisis korelasi biavariate pearson  
Berdasarkan tabel output di atas, maka dapat disimpulkan :
  - 1) Berdasarkan nilai Sig (2-tailed) antara Motivasi (X1) dengan Prestasi (Y) adalah sebesar  $0.002 < 0.05$  yang berarti terdapat korelasi yang signifikan antara variabel Motivasi dengan variabel Prestasi.
  - 2) Hubungan antara Minat (X2) dengan Prestasi (Y) memiliki nilai Sig (2-tailed) sebesar  $0.000 < 0.05$  yang berarti terdapat korelasi yang signifikan antara variabel Minat dengan variabel Prestasi.
  - 3) Berdasarkan nilai r hitung (Pearson Correlation) diketahui nilai r hitung untuk hubungan motivasi (X1) dengan prestasi (Y) adalah sebesar  $0.796 > r \text{ tabel } 0.576$  maka dapat disimpulkan bahwa ada hubungan atau korelasi antara variabel Motivasi dengan variabel Prestasi.
  - 4) Diketahui nilai r hitung untuk hubungan Minat (X2) dengan prestasi (Y) adalah sebesar  $0.908 > r \text{ tabel } 0.576$  maka dapat disimpulkan bahwa ada hubungan atau korelasi antara variabel Minat dengan variabel Prestasi.  
(r tabel terlampir)

#### f) Korelasi Phi

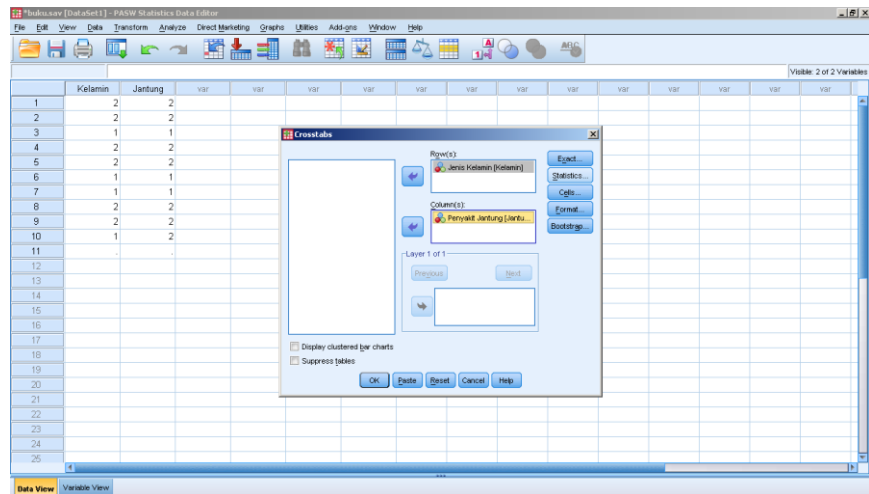
- Dalam soal ini adalah mencari hubungan antara jenis kelamin dengan kejadian penyakit jantung. Entry terlebih dahulu datanya ke SPSS seperti berikut.



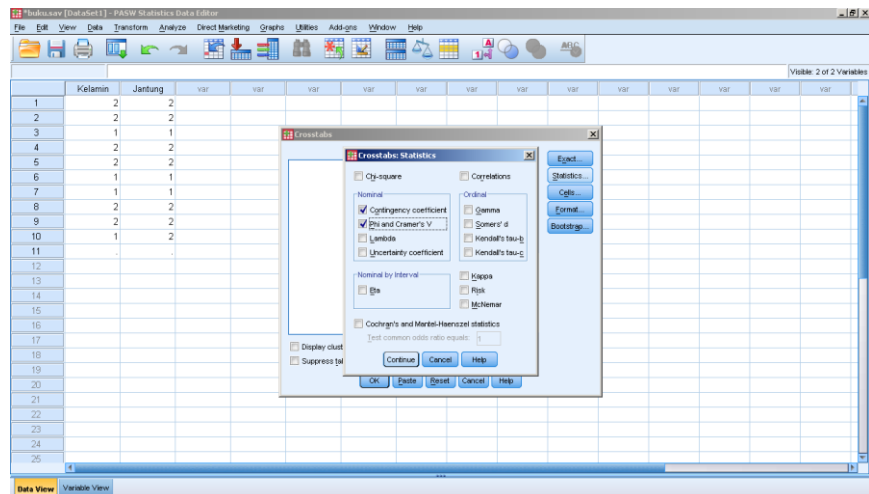
- Klik Analyze – Descriptive Statistics – Crosstabs



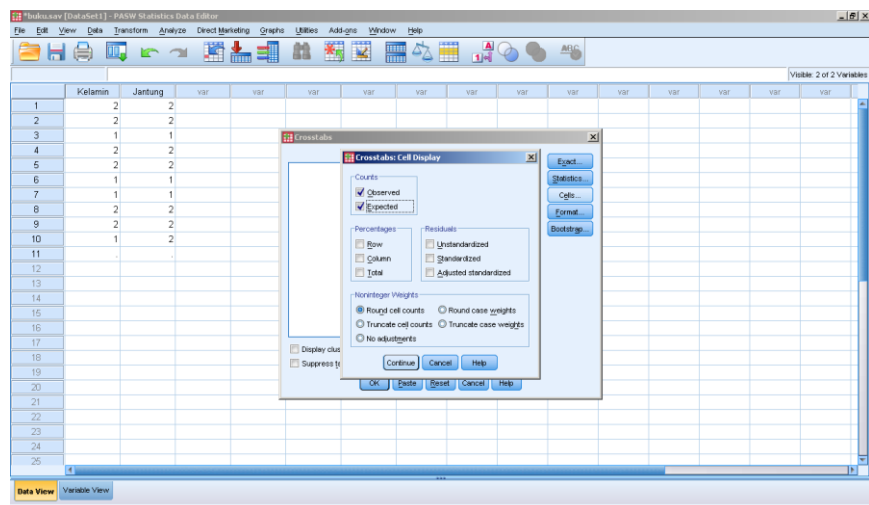
- Masukkan variabel jenis kelamin ke “Row” dan penyakit jantung ke “Column” – klik Statistics



- Centang Contingency dan Phi – Continue - Cells



- Centang bagian “Counts” dan Round Cell counts



▪ **Output**

<b>Symmetric Measures</b>			
		Value	Approx. Sig.
Nominal by Nominal	Phi	.802	.011
	Cramer's V	.802	.011
	Contingency Coefficient	.626	.011
N of Valid Cases		10	

**PENGAMBILAN KEPUTUSAN**

- Jika Sig > 0,05 maka Ho diterima
- Jika Sig < 0,05 maka Ho ditolak

Dari tabel di atas memiliki nilai Phi, Cramer's V dan Contingency coefficient masing-masing 0,802, 0,802 dan 0,626 dengan Approx. Sig. 0,011 atau lebih kecil dari 0,05 sehingga diputuskan Jenis Kelamin dan Gangguan Jantung berhubungan nyata dan signifikan.

**g) Korelasi Koefisien Kontingency C**

Caranya sama dengan Korelasi Phi di atas, hanya saja yang dibaca pada output adalah **Cramers dan Contingency Coefficient**

**h) Korelasi Rank Spearman**

Dalam contoh ini peneliti ingin mengetahui apakah terdapat hubungan antara kualitas produk dengan kepuasan konsumen. Untuk keperluan tersebut, peneliti menyebar 10 kuisisioner di toko furniture.

No	Kualitas Produk (X)	Kepuasan Konsumen (Y)
1	TB	CP
2	CB	CP
3	B	P
4	B	SP
5	CB	P
6	SB	SP
7	SB	SP
8	B	SP
9	SB	SP
10	B	P

**Cara analisis**

Jawaban responden tersebut kemudian diberi kode :

- Untuk data kualitas produk

STB = 1

TB = 2

CB = 3

B = 4

SB = 5

- Untuk data kepuasan konsumen

STP = 1

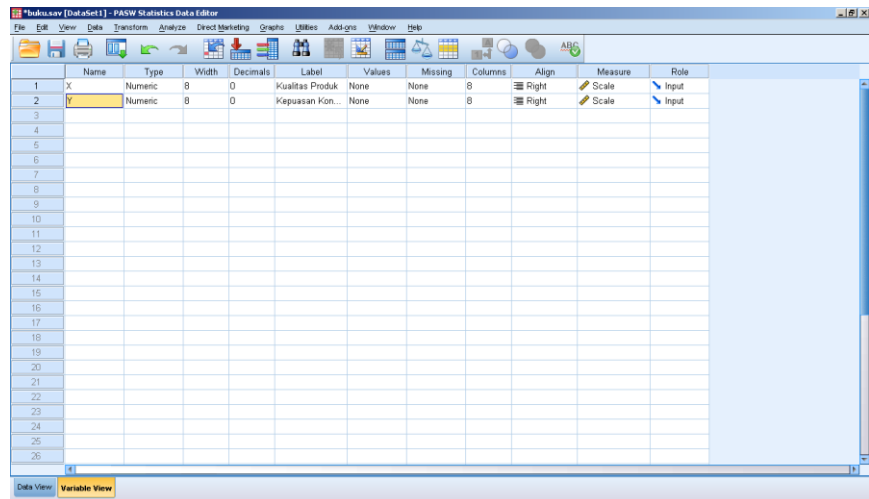
TP = 2

CP = 3

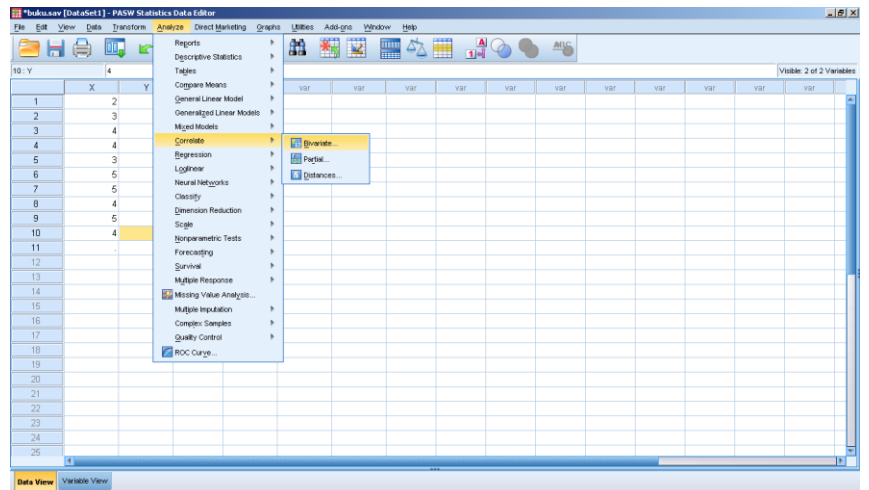
P = 4

SP = 5

- Memasukan data ke SPSS

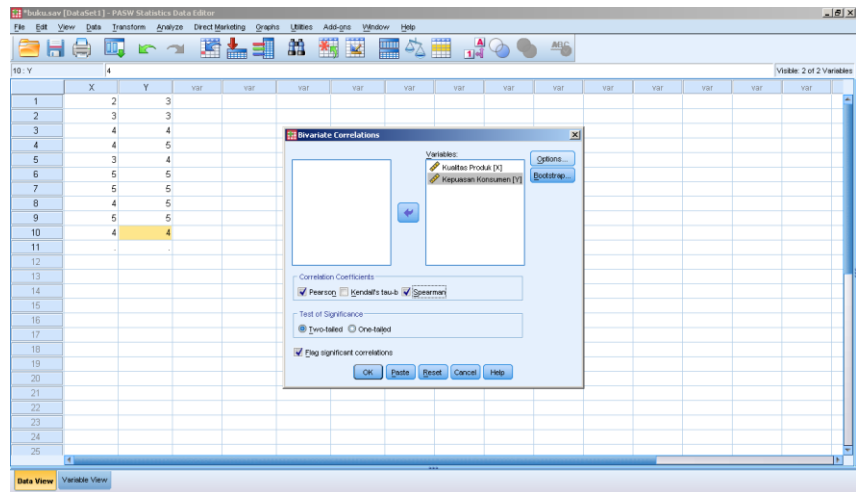


- Masukan nilai yang telah di code Correlate - Bivariat

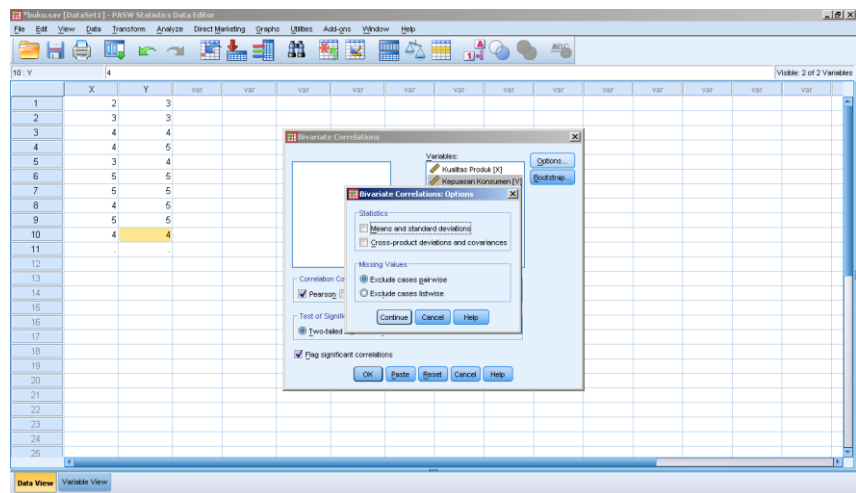




- Masukkan semua variabel X dan Y kedalam kolom variabel centang Spearman – two tailed dan bagian bawah sendiri kemudian pilih Options



- Pilih centang pada Exclude cases pairwise – Continue - OK



▪ **Output**

Correlations				
			Kualitas Produk	Kepuasan Konsumen
Spearman's rho	Kualitas Produk	Correlation Coefficient	1.000	.838**
		Sig. (2-tailed)	.	.002
		N	10	10
	Kepuasan Konsumen	Correlation Coefficient	.838**	1.000
		Sig. (2-tailed)	.002	.
		N	10	10

\*\* . Correlation is significant at the 0.01 level (2-tailed).

Dari output di atas didapatkan angka koefisien korelasi sebesar 0.838\*\*. Artinya, tingkat kekuatan hubungan (korelasi) antara variabel kualitas produk dengan kepuasan konsumen adalah sebesar 0.838 atau sangat kuat. Tanda bintang (\*\*) artinya korelasi bernilai signifikan pada angka signifikansi sebesar 0.01.

- Melihat arah (jenis) hubungan variabel kualitas produk dengan kepuasan konsumen. Angka koefisien korelasi pada hasil di atas bernilai positif, yaitu 0.838 sehingga hubungan kedua variabel bersifat searah. Dengan begitu berarti kualitas semakin ditingkatkan kualitas produk maka kepuasan konsumen juga akan meningkat.
- Melihat signifikansi hubungan kedua variabel. Diketahui nilai Sig sebesar  $0.002 < 0.05$  maka artinya ada hubungan yang signifikan (berarti) antara kualitas produk dengan kepuasan konsumen.

# BAB 10 UJI CHI-KUADRAT

---

## A. STATISTIK NONPARAMETRIK

Dalam terminologi ilmu statistika, statistik parametrik dan non parametrik merupakan dua hal yang sering digunakan. Lantas apa perbedaan keduanya? Secara sederhana sebetulnya antara statistik parametrik dan non parametrik mudah dibedakan dari istilahnya saja. Statistik non parametrik adalah statistik yang tidak mendasarkan pada parameter-parameter statistik. apa itu parameter-parameter statistik? jika anda melakukan penelitian, tentu anda melakukan pengukuran-pengukuran, nah ukuran-ukuran tersebut diistilahkan dengan parameter. dalam statistik kita mengenal mean, median, modus dan standar deviasi. itulah parameter-parameter statistik. dalam statistik non parametrik, parameter tersebut tidak dijadikan acuan. Mengapa? ketika kita menggunakan skala data nominal atau ordinal, parameter-parameter tersebut menjadi tidak relevan. itu lebih kepada membuat ranking pada data. selain itu, statistik non parametrik tidak mendasakan pada distribusi data tertentu.

### 1. Penggunaan statistik non parametrik

Statistik non parametrik banyak digunakan pada kondisi di mana peneliti dihadapkan pada data yang berupa ranking, misalnya data untuk menilai peringkat mana yang lebih penting diantara beberapa atribut produk. Begitupun ketika hendak menganalisis data berupa data nominal atau data dikotomis, misalnya kita hanya menggunakan skala 1 dan 2 untuk membedakan jenis kelamin laki-laki dan perempuan. Parameter-parameter statistik seperti rata-rata dan standar deviasi menjadi tidak relevan. Jika kita paksakan untuk menggunakannya maka tentu rata-rata data hanya menyebar di antara angka 1 dan 2.

Ketika peneliti menggunakan skala ordinal dalam mengukur suatu variabel, statistik non parametrik merupakan metode yang cocok untuk menganalisis data tersebut. Namun, kebanyakan peneliti menggunakan statistik parametrik melalui penghitungan parameter mean dan standar deviasi terlebih dahulu. Memang, dalam hal interpretasi, statistik parametrik lebih mudah dipahami dibandingkan statistik non parametrik. Kita tentu akan lebih mudah membaca rata-rata atau penyimpangan suatu data dibandingkan ranking dari data

itu sendiri. Alasan kemudahan membaca hasil inilah yang sering dijadikan justifikasi untuk menghindari statistik non parametrik.

Alasan kedua penggunaan statistik non parametrik adalah ketika data peneliti dihadapkan pada data yang tidak berdistribusi normal atau peneliti tidak memiliki cukup bukti yang kuat data berasal dari distribusi data seperti apa. Kita sering dihadapkan pada kondisi di mana data tidak berdistribusi normal, misalnya distribusi data terlalu miring ke kiri atau ke kanan. Berbagai usaha dapat dilakukan dengan mereduksi data outlier atau data ekstrim. Namun, jika hal tersebut tidak merubah distribusi data menjadi terdistribusi normal, maka metode non parametrik dapat dilakukan.

## 2. Kesesuaian Skala Pengukuran Dengan Tes Statistik.

Berdasarkan skala pengukuran data beberapa jenis tes statistik parametrik dan nonparametrik yang kiranya cocok digunakan dalam suatu analisis dapat disajikan sbb:

Tabel 1.1. Empat Skala Pengukuran dan Tes Statistik yang Cocok Digunakan

Skala	Hubungan-hubungan yang membatasi	Contoh-contoh Data Statistik	Tes Statistik yang Sesuai
Nominal	(1) Ekuivalensi	Modus Frekuensi Koefisien Kontingensi	Tes Statistik Nonparametrik.
Ordinal	(1) Ekuivalensi	Median Persentil	Tes Statistik Nonparametrik
	(2) Lebih besar dari	Sepearman Kendal	
Interval	(1) Ekuivalensi (2) lebih besar dari (3) Rasio sembarang dua interval diketahui	Mean (rata-rata) Deviiasi standar Korelasi Produk momen Tunggal dan berganda	Tes Statistik Nonparametrik dan Parametrik

Rasio	1) Ekuivalensi (2) lebih besar dari (3) Rasio sembarang dua interval diketahui (4) Rasio sembarang dua harga skala diketahui	Mean geometrik Koefesien variasi	Tes Statistik Nonparametrik dan Parametrik
-------	---	-------------------------------------	--

## B. CHI KUADRAT UNTUK UJI GOODNESS OF FIT

Dalam statistik, distribusi chi square (dilambangkan dengan  $\chi^2$ ) termasuk dalam statistik nonparametrik. Distribusi nonparametrik adalah distribusi dimana besaran-besaran populasi tidak diketahui. Distribusi ini sangat bermanfaat dalam melakukan analisis statistik jika kita tidak memiliki informasi tentang populasi atau jika asumsi-asumsi yang dipersyaratkan untuk penggunaan statistik parametrik tidak terpenuhi. Beberapa hal yang perlu diketahui berkenaan dengan distribusi chi square adalah :

- Distribusi chi square memiliki satu parameter yaitu derajat bebas (db)
- Nilai-nilai chi square di mulai dari 0 disebelah kiri, sampai nilai-nilai positif tak terhingga di sebelah kanan
- Probabilitas nilai chi square di mulai dari sisi sebelah kanan
- Luas daerah di bawah kurva normal adalah 1. Nilai dari chi square bisa dicari jika kita memiliki informasi luas daerah disebelah kanan kurva serta derajat bebas. Misalnya jika luas daerah disebelah kanan adalah 0,1 dan derajat bebas sebanyak 7, maka nilai chi square adalah 12, 017. Dalam statistic, distribusi chi square digunakan dalam banyak hal. Mulai dari pengujian proporsi data multinom, menguji kesamaan rata-rata Poisson serta pengujian hipotesis. Pengujian hipotesis yang menggunakan dasar distribusi chi square misalnya Goodness-of-fit test, pengujian independensi, pengujian homogenitas serta pengujian varians dan standar deviasi populasi tunggal.

Uji Goodness of Fit Test pada prinsipnya bertujuan untuk mengetahui apakah sebuah distribusisebuah distribusi data dari sampel mengikuti sebuah distribusi teoritis tertentu ataukah tidak. Sebagai contoh, jika sebuah dadu dilempar, maka kemungkinan mendapat angka '5' adalah 1/6, juga kemungkinan untuk angka yang lain. Inilah yang disebut distribusi teoritis sebuah dadu, karena terdiri dari 6 mata dadu yang mempunyai kemungkinan seimbang untuk muncul dalam sekali

pelemparan. Seandainya dilakukan pelemparan 120 kali, seharusnya tiap mata dadu secara teoritis akan muncul masing-masing  $1/6 \times 120 = 20$  kali (angka 1 muncul 20 kali, angka 2 muncul 20 kali dan seterusnya). Namun tentu kenyataan tidaklah persis sama, bisa saja angka 1 muncul hanya 10 kali, tapi angka 3 muncul 24 kali dan kemungkinan lain. Untuk mengetahui apakah kenyataan tersebut masih bisa dianggap selaras (fit) dengan distribusi teoritis, akan digunakan uji Goodness of Fit.

Dengan demikian, Goodness of Fit Test akan membandingkan dua distribusi data, yakni yang teoritis (frekuensi harapan) dan yang sesuai kenyataan (frekuensi observasi). Uji ini hampir sama dengan uji Binomial, hanya jika pada binomial hanya ada dua kemungkinan jawaban, pada uji Goodness of Fit ada lebih dari dua kemungkinan. Dua Kasus berikut menjelaskan penggunaan Chi Square pada Goodness of Fit Test

Uji chi squared secara yang dipernah dibahas pada postingan sebelumnya. salah satu kegunaan uji chi squared yaitu uji kecocokan atau goodness of fit test. silahkan klik link berikut untuk melihat penjelasan secara umum. Uji chi squared. Secara khusus akan dibahas untuk goodness of fit test beserta contoh kasus supaya lebih mudah dimengerti.

Uji kecocokan atau goodness of fit test, hipotesis nol merupakan suatu ketentuan tentang pola yang diharapkan dari frekuensi-frekuensi dalam kategori (- kategori) tertentu. Pola yang diharapkan harus sesuai dengan asumsi atau anggapan atas kemungkinan kejadian yang sama dan bersifat umum.

Uji Kecocokan (goodness of fit), membandingkan antara Frekuensi Observasi dengan Frekuensi Teoretis /Harapan. Apakah Frekuensi hasil Observasi menyimpang dari Frekuensi Harapan. Jika nilai (chi square) kecil, berarti kedua frekuensi tersebut sangat dekat, mengarah pada penerimaan kepada hipotesa nol ( $H_0$ ).

## 1. Penetapan Hipotesis Awal dan Hipotesis Alternatif

$H_0$  : frekuensi setiap kategori memenuhi suatu nilai/perbandingan.

$H_1$  : Ada frekuensi suatu kategori yang tidak memenuhi nilai/perbandingan tersebut.

## 2. Penetapan Derajat Kebebasan (df)

Dalam uji kecocokan model derajat kebebasan (df) sama dengan jumlah kategori dikurangi jumlah estimator yang didasarkan pada sampel dan dikurang 1.

Yang dimaksud estimator parameter adalah parameter yang diperkirakan nilainya, karena nilai parameter tidak dapat secara tepat ditentukan berdasarkan data sampel yang tersedia. Jika dirumuskan menjadi:

$$df = k - m - 1$$

**Dengan :**

k : jumlah kategori data sampel

m : jumlah nilai-nilai parameter yang diestimasi

### 3. Contoh Kasus uji kecocokan/ goodness of fit test

Sebuah distributor alat penggilingan padi membagi pasar menjadi 4 wilayah (A, B, C, dan D). Ada informasi bahwa pendistribusian alat penggilingan merata pada setiap wilayah. Untuk membuktikan pernyataan tersebut diambil 40 arsip sebagai sampel. Dari 40 arsip tersebut diperoleh informasi yang tertuang pada tabel. Gunakan tingkat signifikansi 5 persen untuk menguji hipotesis yang menyatakan bahwa distribusi alat penggilingan di keempat wilayah merata (sama)

	Wilayah				Total Baris
	A	B	C	D	
Data berdasarkan hasil sampel (O)	6	12	14	8	40
Data yang diharapkan (E)	10	10	10	10	40

Pembahasan Contoh goodness of fit

#### a) Hipotesis

Sebelum memulai pembahasan langkah pertama yang harus diketahui adalah permasalahan dan tujuan dari soal yang ingin dicapai oleh peneliti. Pertama Peneliti ingin membuktikan bahwa pendistribusian sama rata. sehingga bisa memperkirakan hipotesis. ini hipotesis yang digunakan dalam penelitian ini

- Ho : distribusi alat penggilingan di keempat wilayah merata (sama)
- Ha : distribusi alat penggilingan di keempat wilayah tidak merata (tidak sama)

### b) Nilai Kritis

Kedua yang perlu diperhatikan yaitu nilai kritis. Maksud dari nilai kritis tersebut adalah nilai batas dari penentu keputusan hipotesis mana yang di ambil. sehingga ini sangat perlu dilakukan. Berdasarkan penjelasan di atas. Dalam kasus di atas tidak perlu ada parameter yang diestimasi. oleh karena itu:

$$df = k - 1 = 4 - 0 - 1 = 3$$

k = jumlah kategori data sampel (A, B, C, dan D)

Selain itu tingkat signifikansi yang digunakan adalah 0,05(5%), sehingga nilai kritisnya adalah:

$$X_2(0,05;3) = 7,81$$

nilai 7,81 ini diperoleh dari tabel chi squared. untuk mengetahui nilai tersebut harus punya tabel chi squared. silahkan klik link berikut jika mau lihat tabel chi squared

### c) Nilai Hitung

Nilai uji statistik  $X_2$  hitung diperoleh dengan cara sebagai berikut seperti pada materi sebelumnya dalam uji chi squared:

$$\begin{aligned}\chi^2 &= \sum \frac{(O - E)^2}{E} \\ &= \frac{(6 - 10)^2}{10} + \frac{(12 - 10)^2}{10} + \frac{(14 - 10)^2}{10} + \frac{(8 - 10)^2}{10} \\ &= \frac{40}{10} = 4\end{aligned}$$

## C. Uji Keselarasan dengan Frekuensi Harapan Sama

Chi-Square disebut juga dengan Kai Kuadrat. Chi Square adalah salah satu jenis uji komparatif non parametris yang dilakukan pada dua variabel, di mana skala data kedua variabel adalah nominal. (Apabila dari 2 variabel, ada 1 variabel dengan skala nominal maka dilakukan uji chi square dengan merujuk bahwa harus digunakan uji pada derajat yang terendah).

Uji chi-square merupakan uji non parametris yang paling banyak digunakan. Namun perlu diketahui syarat-syarat uji ini adalah: frekuensi responden atau sampel yang digunakan besar, sebab ada beberapa syarat di mana chi square dapat digunakan yaitu:



1. Tidak ada cell dengan nilai frekuensi kenyataan atau disebut juga Actual Count (FO) sebesar 0 (Nol).
2. Apabila bentuk tabel kontingensi 2 X 2, maka tidak boleh ada 1 cell saja yang memiliki frekuensi harapan atau disebut juga expected count (“Fh”) kurang dari 5.
3. Apabila bentuk tabel lebih dari 2 x 2, misak 2 x 3, maka jumlah cell dengan frekuensi harapan yang kurang dari 5 tidak boleh lebih dari 20%.

Rumus chi-square sebenarnya tidak hanya ada satu. Apabila tabel kontingensi bentuk 2 x 2, maka rumus yang digunakan adalah “koreksi Yates”. Apabila tabel kontingensi 2 x 2 seperti di atas, tetapi tidak memenuhi syarat seperti di atas, yaitu ada cell dengan frekuensi harapan kurang dari 5, maka rumus harus diganti dengan rumus “Fisher Exact Test”.

Uji kai kuadrat (dilambangkan dengan “ $\chi^2$ ” dari huruf Yunani “Chi” dilafalkan “Kai”) digunakan untuk menguji dua kelompok data baik variabel independen maupun dependennya berbentuk kategorik atau dapat juga dikatakan sebagai uji proporsi untuk dua peristiwa atau lebih, sehingga datanya bersifat diskrit. Misalnya ingin mengetahui hubungan antara status gizi ibu (baik atau kurang) dengan kejadian BBLR (ya atau tidak).

Dasar uji kai kuadrat itu sendiri adalah membandingkan perbedaan frekuensi hasil observasi (O) dengan frekuensi yang diharapkan (E). Perbedaan tersebut meyakinkan jika harga dari Kai Kuadrat sama atau lebih besar dari suatu harga yang ditetapkan pada taraf signifikan tertentu (dari tabel  $\chi^2$ ).

### **Uji Kai Kuadrat dapat digunakan untuk menguji :**

1. Uji  $\chi^2$  untuk ada tidaknya hubungan antara dua variabel (Independency test).
2. Uji  $\chi^2$  untuk homogenitas antar- sub kelompok (Homogeneity test).
3. Uji  $\chi^2$  untuk Bentuk Distribusi (Goodness of Fit)

Keterangan :

O = frekuensi hasil observasi

E = frekuensi yang diharapkan.

Nilai E = (Jumlah sebaris x Jumlah Sekolom) / Jumlah data

df = (b-1) (k-1)

Dalam melakukan uji kai kuadrat, harus memenuhi syarat:

1. Sampel dipilih secara acak

2. Semua pengamatan dilakukan dengan independen
3. Setiap sel paling sedikit berisi frekuensi harapan sebesar 1 (satu). Sel-sel dengan frekuensi harapan kurang dari 5 tidak melebihi 20% dari total sel
4. Besar sampel sebaiknya  $> 40$  (Cochran, 1954)

Keterbatasan penggunaan uji Kai Kuadrat adalah tehnik uji kai kuadrat memakai data yang diskrit dengan pendekatan distribusi kontinu. Dekatnya pendekatan yang dihasilkan tergantung pada ukuran pada berbagai sel dari tabel kontingensi. Untuk menjamin pendekatan yang memadai digunakan aturan dasar “frekuensi harapan tidak boleh terlalu kecil” secara umum dengan ketentuan:

1. Tidak boleh ada sel yang mempunyai nilai harapan lebih kecil dari 1 (satu)
2. Tidak lebih dari 20% sel mempunyai nilai harapan lebih kecil dari 5 (lima)

Bila hal ini ditemukan dalam suatu tabel kontingensi, cara untuk menanggulanginya adalah dengan menggabungkan nilai dari sel yang kecil ke sel lainnya (mengcollaps), artinya kategori dari variabel dikurangi sehingga kategori yang nilai harapannya kecil dapat digabung ke kategori lain. Khusus untuk tabel  $2 \times 2$  hal ini tidak dapat dilakukan, maka solusinya adalah melakukan uji

## 1. “Fisher Exact atau Koreksi Yates”

### Contoh kasus

Perusahaan penyalur alat elektronik AC ingin mengetahui apakah ada hubungan antara gender dengan sikap mereka terhadap kualitas produk AC. Untuk itu mereka meminta 25 responden mengisi identitas mereka dan sikap atau persepsi mereka terhadap produknya.

Permasalahan :

- Apakah ada hubungan antara gender dengan sikap terhadap kualitas AC?

Hipotesis :

- $H_0$  = Tidak ada hubungan antara gender dengan sikap terhadap kualitas AC
- $H_1$  = Ada hubungan antara gender dengan sikap terhadap kualitas AC

Tolak hipotesis nol ( $H_0$ ) apabila nilai signifikansi chi-square  $< 0.05$  atau nilai chi-square hitung lebih besar ( $>$ ) dari nilai chi-square tabel.

## 2. Menguji Independensi antara 2 faktor (independensi)

Independensi (keterkaitan) antara 2 faktor dapat diuji dengan uji chi square. Masalah independensi ini banyak mendapat perhatian hampir di semua bidang, baik eksakta maupun sosial ekonomi. Kita ambil contoh di bidang ekonomi dan pendidikan. Kita bisa menduga bahwa keadaan ekonomi seseorang tidak ada kaitannya dengan tingkat pendidikannya, atau justru sebaliknya bahwa keadaan ekonomi seseorang terkait erat dengan tingkat pendidikannya. Untuk menjawab dugaan-dugaan ini, kita bisa menggunakan uji chi square. Langkah-langkahnya sebagai berikut.

- Buatlah hipotesis  
 $H_0$ : tidak ada kaitan antara keadaan ekonomi seseorang dengan pendidikannya  
 $H_1$ : ada kaitan antara keadaan ekonomi seseorang dengan pendidikannya
- Lakukan penelitian dan kumpulkan data  
Hasil penelitian adalah sebagai berikut (tentatif).

Kategori	Di bawah garis kemiskinan	Di atas garis kemiskinan	Total
Tidak tamat SD	8	4	12
SD	20	17	37
SMP	15	16	31
SMA	3	23	26
Perguruan Tinggi	2	22	24
Total	48	82	130

- Lakukan analisis

Kategori	Di bawah garis kemiskinan	Di atas garis kemiskinan	Total
Tidak tamat SD O E	8 4,43	4 7,57	12
SD O E	20 13,66	17 23,34	37
SMP O E	15 11,45	16 19,55	31
SMA O E	3 9,60	23 16,40	26
Perguruan Tinggi O E	2 8,86	22 15,14	24
Total	48	82	130

Nilai O (Observasi) adalah nilai pengamatan di lapangan

Nilai E (expected) adalah nilai yang diharapkan, dihitung sbb :

- Nilai E untuk kategori tidak tamat SD di bawah garis kemiskinan=  $(12 \times 48)/130 = 4,43$
- Nilai E untuk kategori tidak tamat SD di atas garis kemiskinan =  $(12 \times 82)/130 = 7,57$
- Nilai E untuk kategori SD di bawah garis kemiskinan =  $(37 \times 48)/130 = 13,66$

- Nilai E untuk kategori SD di atas garis kemiskinan =  $(37 \times 82)/130 = 23,34$
- Nilai E untuk kategori SMP di bawah garis kemiskinan =  $(31 \times 48)/130 = 11,45$
- Nilai E untuk kategori SMP di atas garis kemiskinan =  $(31 \times 82)/130 = 19,55$
- Nilai E untuk kategori SMA di bawah garis kemiskinan =  $(26 \times 48)/130 = 9,60$
- Nilai E untuk kategori SMA di atas garis kemiskinan =  $(26 \times 82)/130 = 16,40$
- Nilai E untuk kategori Perguruan Tinggi di bawah garis kemiskinan =  $(24 \times 48)/130 = 8,86$
- Nilai E untuk kategori Perguruan Tinggi di atas garis kemiskinan =  $(24 \times 82)/130 = 15,14$

Pengujian hipotesis kompatibilitas (goodness of fit) merupakan pengujian hipotesis untuk menentukan apakah suatu himpunan frekuensi yang diharapkan sama dengan frekuensi yang diperoleh dari suatu distribusi, seperti distribusi binomial, poisson, normal, atau dari perbandingan lain. Jadi, uji goodness of fit merupakan pengujian kecocokan atau kebaikan suai antara hasil pengamatan (frekuensi pengamatan) tertentu dengan frekuensi yang diperoleh berdasarkan nilai harapannya (frekuensi teoretis). Uji keselarasan (goodness of test) 2.

Metode ini dikembangkan oleh Pearson tahun 1900 yang merupakan perhitungan suatu kuantitas yang disebut Kai Kuadrat . Metode ini sangat bermanfaat ketika data yang tersedia hanya berupa frekuensi (disebut count), misalnya banyaknya subjek dalam kategori sakit dan tidak sakit, atau banyaknya penderita diabetes mellitus dalam kategori I, II, III, IV menurut keparahan penyakitnya.

Uji kai kuadrat untuk satu sampel dapat dipakai untuk menguji apakah data sebuah sampel yang diambil menunjang hipotesa yang menyatakan bahwa populasi asal sampel tersebut mengikuti suatu distribusi yang telah ditetapkan. Oleh karena itu uji ini disebut juga uji keselarasan (goodness of fit test), karena untuk menguji apakah sebuah sampel selaras dengan salah satu distribusi teoritis (seperti distribusi normal, uniform, binomial dan lainnya).

### Contoh Kasus :

Sebuah survei berminat menyelidiki determinasi orang dalam mencegah factor-faktor risiko penyakit jantung koroner. Setiap subjek dari sampel berukuran 200 orang diminta menyatakan sikapnya terhadap sebuah pertanyaan kuesioner sebagai berikut “ apakah anda yakin dapat menghindari makanan berkolesterol tinggi” dengan hasil 70 orang sangat yakin, 50 orang yakin, 45 orang ragu-ragu, dan 35 orang sangat ragu-ragu. Dapatkah kita menarik kesimpulan berdasarkan data tersebut, bahwa keempat sikap yang berbeda menyebar merata di dalam populasi asal sampel?

Penyelesaian :

Buat tabel seperti ini untuk memudahkan dalam perhitungan :

Jawaban	O	E
Sangat Yakin	70	50
Yakin	50	50
Ragu-ragu	45	50
Sangat Ragu-ragu	35	50
<b>Jumlah</b>	<b>200</b>	<b>200</b>

Nilai E = 50, karena kita berharap bahwa jumlah yang menjawab pada masing-masing kategorik akan berdistribusi sama. Selanjutnya masukan dalam rumus.

$$x^2 = \frac{(70 - 50)^2}{50} + \frac{(50 - 50)^2}{50} + \frac{(45 - 50)^2}{50} + \frac{(35 - 50)^2}{50}$$

$$x^2 = \frac{(20)^2}{50} + \frac{(0)^2}{50} + \frac{(-5)^2}{50} + \frac{(-15)^2}{50}$$

$$x^2 = 13$$

Dari hasil perhitungan terlihat Chi square hitung adalah 13. Selanjutnya melihat nilai tabel pada kemaknaan alfa = 0.05 pada df = 4-1 = 3.

df	α							
	0,99	0,95	0,90	0,50	0,10	0,05	0,01	0,001
1	.00157	.00393	.0158	.455	2.706	3.841	6.635	10.827
2	.0201	.103	.211	1.386	4.605	5.991	9.210	13.815
3	.115	.352	.584	2.366	6.251	7.815	11.345	16.226
.. dst								

Dari tabel chi square diperoleh chi square tabel dengan  $df= 3$  adalah 7,815, berarti Chi-square hitung  $>$  chi-square tabel, maka  $H_0$  ditolak. Artinya sikap responden terhadap pertanyaan tidak proporsional, dimana sikap responden cenderung pada sikap tertentu.

#### D. UJI CHI-KUADRAT UNTUK UJI KENORMALAN

Uji normalitas dengan menggunakan uji Chi Kuadrat disebut juga Uji Goodness of Fit. Menggunakan pendekatan penjumlahan penyimpangan data observasi tiap kelas dengan nilai yang diharapkan. Uji normalitas datanya disajikan secara berkelompok. Data berbentuk nominal atau ordinal.

##### 1. Ciri-Ciri Distribusi Chi Kuadrat

Selalu positif

$df = k - 1$ , dimana  $k$  adalah jumlah kategori (variabel).

Jadi bentuk distribusi chi-kuadrat tidak ditentukan banyaknya sampel, melainkan banyaknya derajat bebas.

Bentuk distribusi chi-kuadrat menjulur positif. Semakin besar derajat bebas, semakin mendekati distribusi normal. Rumus umum :

$$X^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}$$

Keterangan :

$O_i$  = frekuensi hasil pengamatan pada klasifikasi ke- $i$

$E_i$  = frekuensi yang diharapkan pada klasifikasi ke- $i$

$X^2$  = Nilai Chi-Kuadrat

Uji normalitas dengan menggunakan Chi-Kuadrat dapat dilakukan dengan langkah-langkah sebagai berikut :

1. Mencari nilai terbesar dan terkecil
2. Mencari nilai rentang
3. Mencari banyak kelas
4. Mencari panjang kelas interval ( $i$ )
5. Membuat tabel distribusi frekuensi
6. Mencari rata-rata (mean)

7. Mencari simpangan baku (standar deviasi)
8. Membuat daftar frekuensi yang diharapkan dengan cara sebagai berikut :
  - Menentukan batas kelas, yaitu ujung bawah kelas interval dikurangi 0.5 dan kemudian ujung atas kelas interval ditambah 0.5
  - Mencari nilai Z menggunakan batas bawah dan batas atas kelas interval dengan rumus:

$$Z = \frac{x_i - \bar{x}}{s}$$

- Mencari luas 0-Z dari Tabel Kurva Normal dari 0-Z dengan menggunakan Z hitung.
- Mencari selisih luas tiap kelas interval dengan cara mengurangkan nilai-nilai 0-Z tepi bawah dengan tepi atas.
- Mencari frekuensi yang diharapkan dengan cara mengalikan luas tiap interval dengan jumlah responden.
- Mencari Chi-Kuadrat hitung

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}$$

Uji chi-kuadrat digunakan jika ukuran sampel ( $n \geq 30$ ). Metode Chi-Square atau uji Goodness of fit Distribution Normal, menggunakan pendekatan penjumlahan penyimpangan data observasi tiap kelas dengan nilai yang diharapkan.

- Langkah-Langkah pengujian
- Rumusan Hipotesis

$H_0$  : sampel berasal dari populasi berdistribusi normal

$H_1$  : sampel tidak berasal dari populasi berdistribusi normal

$\alpha$  : taraf nyata

Data disusun dalam distribusi frekuensi sebagai berikut :

kelas interval	$f_i = O_i$	$x_i$	$z = \frac{x_i - \bar{x}}{s}$
a - b	$f_1$	$a'$	$z_1$



c - d	f2	b'=c'	z2
...	...	...	...
k - 1	fk	l'	zk

Keterangan :

- xi = batas bawah kelas
- s = simpangan baku
- zi = angka baku
- pi = luas daerah antara dua harga
- Ei = nilai ekspektasi atau harapan
- d = jumlah frekuensi (fi)

- Signifikansi

Signifikansi uji, nilai hitung dibandingkan dengan tabel (Chi-Square)

Jika nilai hitung kurang dari nilai tabel, maka diterima dan ditolak.

Jika nilai hitung lebih besar dari nilai tabel, maka ditolak dan diterima.

**Contoh Soal :**

Akan diuji normalitas untuk data Kemampuan Pemecahan Masalah Matematika Siswa Kelas VIII G SMP Negeri 407 Jepara pada Materi Bangun Ruang. Apakah data tersebut berdistribusi normal? Datanya adalah sebagai berikut :

No.	X	No.	X.	No.	X	No.	X	No.	X
1.	58	16.	51	31.	57	46.	72	61.	78
2.	57	17.	65	32.	51	47.	78	62.	71
3.	57	18.	85	33.	52	48.	66	63.	45
4.	51	19.	72	34.	51	49.	48	64.	50
5.	51	20.	78	35.	65	50.	71		
6.	52	21.	58	36.	78	51.	78		
7.	71	22.	59	37.	71	52.	70		
8.	79	23.	58	38.	71	53.	65		
9.	72	24.	64	39.	64	54.	37		
10.	75	25.	64	40.	58	55.	58		
11.	58	26.	58	41.	50	56.	50		
12.	62	27.	71	42.	44	57.	50		
13.	57	28.	64	43.	58	58.	58		
14.	57	29.	78	44.	48	59.	48		
15.	57	30.	78	45.	65	60.	67		

**Jawab :**

Hipotesis :

$H_0$  : data pada sampel berasal dari populasi yang berdistribusi normal

$H_1$  : data pada sampel berasal dari populasi yang tidak berdistribusi normal

Langkah pengujiannya sebagai berikut :

1. Mencari nilai terbesar terbesar dan terkecil

- Nilai terbesar = 85
- Nilai terkecil = 37

2. Mencari nilai rentang (R)

$R = \text{skor terbesar} - \text{skor terkecil}$

$$R = 85 - 37 \\ = 58$$

Mencari banyak kelas (BK)

$$BK = 1 + 3,3 \log n$$

$$BK = 1 + 3,3 \log 64$$

$$BK = 1 + 3,3 (1,81)$$

$$BK = 1 + 3,3 (1,81)$$

$$BK = 1 + 5,973$$

$$BK = 6,973 \text{ dibulatkan menjadi } 7$$

3. Mencari nilai panjang kelas (i)

$$i = \frac{R}{BK}$$

$$i = \frac{58}{7}$$

$$i = 8,285 \approx 8$$

4. Membuat tabel distribusi frekuensi

No	Kelas Interval	f	Nilai Tengah ( $x_i$ )
1.	37 - 44	2	40,5
2.	45 - 52	15	48,5
3.	53 - 60	16	56,5
4.	61 - 68	11	64,5

5.	69 - 76	11	72,5
6.	77 - 84	8	80,5
7.	85 - 92	1	88,5
jumlah		64	

5. Mencari rata-rata (mean)

$$\bar{x} = \frac{\sum f \cdot x_i}{n}$$

$$\bar{x} = \frac{3952}{64}$$

$$\bar{x} = 61,75$$

6. Mencari simpangan baku (standar deviasi)

No.	Kelas Interval					
1.	37-44	2	40,5	-21,25	451,5625	903,125
2.	45-52	15	48,5	-13,25	175,5625	2633,4375
3.	53-60	16	56,5	-5,25	27,5625	441
4.	61-68	11	64,5	2,75	7,5625	83,1875
5.	69-76	11	72,5	10,75	115,5625	1271,1875
6.	77-84	8	80,5	18,75	351,5625	2812,5
7.	85-92	1	88,5	26,75	715,5625	715,5625
		64				8860

$$s = \sqrt{\frac{\sum_{i=1}^n f_i (x_i - \bar{x})^2}{n - 1}} = \sqrt{\frac{8860}{63}} = 11,86$$

7. Membuat daftar frekuensi yang diharapkan dengan cara sebagai berikut :

a) Menentukan Tepi Bawah dan Tepi Atas Kelas Interval :

No	Kelas Interval	Batas bawah kelas	Batas atas kelas
1.	37 - 44	36,5	44,5
2.	45 - 52	44,5	52,5
3.	53 - 60	52,5	60,5

4.	61 - 68	60,5	68,5
5.	69 - 76	68,5	76,5
6.	77 - 84	76,5	84,5
7.	85 - 92	84,5	92,5

b) Mencari nilai Z menggunakan Tepi Bawah dan Tepi Atas Kelas Interval

$$Z = \frac{\text{Tepi Bawah} - \bar{x}}{s}$$

$$Z_1 = \frac{36,5 - 61,75}{11,86} = -2,13$$

$$Z_2 = \frac{44,5 - 61,75}{11,86} = -1,45$$

$$Z_3 = \frac{52,5 - 61,75}{11,86} = -0,78$$

$$Z_4 = \frac{60,5 - 61,75}{11,86} = -0,11$$

$$Z_5 = \frac{68,5 - 61,75}{11,86} = 0,57$$

$$Z_6 = \frac{76,5 - 61,75}{11,86} = 1,24$$

$$Z_7 = \frac{84,5 - 61,75}{11,86} = 1,92$$

$$Z = \frac{\text{Tepi Atas} - \bar{x}}{s}$$

$$Z_1 = \frac{44,5 - 61,75}{11,86} = -1,45$$

$$Z_2 = \frac{52,5 - 61,75}{11,86} = -0,78$$

$$Z_3 = \frac{60,5 - 61,75}{11,86} = -0,11$$

$$Z_4 = \frac{68,5 - 61,75}{11,86} = 0,57$$

$$Z_5 = \frac{76,5 - 61,75}{11,86} = 1,24$$

$$Z_6 = \frac{84,5 - 61,75}{11,86} = 1,92$$

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2122	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2672	0.2701	0.2730	0.2759	0.2787	0.2815	0.2842
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4250	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4720	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4831	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964

No.	Z	Luas 0-Z
1	-2,13	0,4834
2	-1,45	0,4265
3	-0,78	0,2823
4	-0,11	0,0438
5	0,57	0,2157
6	1,24	0,3925
7	1,92	0,4726

Z	Luas 0-Z Tepi Atas
-1,45	0,4265
-0,78	0,2823
-0,11	0,0438
0,57	0,2157
1,24	0,3925
1,92	0,4726
2,59	0,4952

8. Mencari selisih luas tiap kelas interval dengan cara mengurangkan nilai-nilai 0-Z tepi bawah dengan tepi atas

Selisih Luas 0-Z
0,0569
0,1442
0,2385
0,2595
0,1768
0,0801
0,0226

9. Mencari frekuensi yang diharapkan dengan cara mengalikan selisih luas tiap interval dengan jumlah responden ( $n = 64$ )

Selisih Luas 0-Z	Ei
0,0569	3,64
0,1442	9,23
0,2385	15,26
0,2595	16,61
0,1768	11,32
0,0801	5,13
0,0226	1,45

Frekuensi yang Diharapkan ( $E_i$ ) dari Hasil Pengamatan ( $O_i$ ) untuk Variabel Kemampuan Pemecahan Masalah Matematika Siswa Kelas VIII G SMP Negeri 407 Jepara pada Materi Bangun Ruang

No.	Batas Kelas	Z	Luas 0-Z	Luas Tiap Kelas Interval	( $E_i$ )	( $O_i$ )
1	36,5	-2,13	0,4834	0,0569	3,64	2
2	44,5	-1,45	0,4265	0,1442	9,23	15
3	52,5	-0,78	0,2823	0,2385	15,26	16
4	60,5	-0,11	0,0438	0,2595	16,61	11
5	68,5	0,57	0,2157	0,1768	11,32	11
6	76,5	1,24	0,3925	0,0801	5,13	8
7	84,5	1,92	0,4726	0,0226	1,45	1
	92,5	2,58	0,4952			
						$\sum O = 64$

10. Mencari Chi-Kuadrat hitung

$$x^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}$$

$$x^2 = \frac{(2 - 3,64)^2}{3,64} + \frac{(15 - 9,23)^2}{9,23} + \frac{(16 - 15,26)^2}{15,26} + \frac{(11 - 16,61)^2}{16,61}$$

$$+ \frac{(11 - 11,32)^2}{11,32} + \frac{(8 - 5,13)^2}{5,13} + \frac{(1 - 1,45)^2}{1,45}$$

$$x^2 = 0,74 + 3,61 + 0,04 + 1,89 + 0,009 + 1,61 + 0,178 = 8,077$$

11. Membandingkan  $X^2$  hitung dengan  $X^2$  tabel

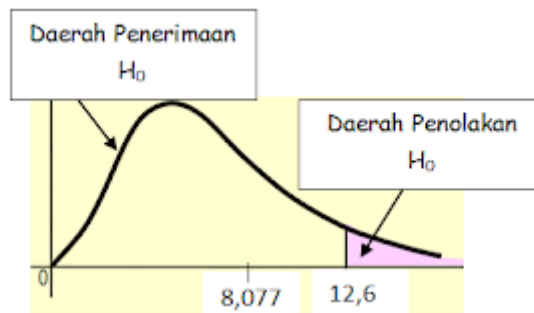
Dengan membandingkan  $X^2$  hitung dengan nilai  $X^2$  tabel untuk  $\alpha = 0,05$  dan derajat kebebasan (dk) =  $k - 1 = 7 - 1 = 6$ , maka dicari pada tabel Chi-Kuadrat didapat  $X^2_{\text{tabel}} = 12,6$  dengan kriteria pengujian sebagai berikut :

- Jika  $X^2$  hitung  $>$   $X^2$  tabel, artinya distribusi data tidak normal
- Jika  $X^2$  hitung  $<$   $X^2$  , artinya data berdistribusi normal.

Ternyata  $X^2$  hitung  $<$   $X^2$ , atau  $8,077 < 12,6$  , maka  $H_0$  diterima. Jadi, data Kemampuan Pemecahan Masalah Matematika Siswa Kelas VIII G SMP Negeri 407 Jepara pada Materi Bangun Ruang adalah berdistribusi normal. Sehingga, analisis uji selanjutnya dapat dilanjutkan.

$\nu$	$\chi^2_{0,995}$	$\chi^2_{0,99}$	$\chi^2_{0,975}$	$\chi^2_{0,95}$	$\chi^2_{0,90}$	$\chi^2_{0,75}$
1	7,88	6,63	5,02	3,84	2,71	1,32
2	10,6	9,21	7,38	5,99	4,61	2,77
3	12,8	11,3	9,35	7,81	6,25	4,11
4	14,9	13,3	11,1	9,49	7,78	5,39
5	16,7	15,1	12,8	11,1	9,24	6,63
6	18,5	16,8	14,4	12,6	10,6	7,84
7	20,3	18,5	16,0	14,1	12,0	9,04
8	22,0	20,1	17,5	15,5	13,4	10,2
9	23,6	21,7	19,0	16,9	14,7	11,4
10	25,2	23,2	20,5	18,3	16,0	12,5
11	26,8	24,7	21,9	19,7	17,3	13,7

12. Kurva daerah penerimaan  $H_0$



**Contoh Soal :**

TINGGI BADAN MASYARAKAT KALIMAS PADA TAHUN 1990

No	Tinggi Badan	Jumlah
1	140-149	6
2	150-159	22

3	160-169	39
4	170-179	25
5	180-189	7
6	190-199	1
jumlah		100

Selidikilah dengan  $\alpha = 5\%$  apakah data diatas berdistribusi normal?

Jawab :

$H_0$  = data berdistribusi normal

$H_1$  = data tidak berdistribusi normal

$\alpha = 5\% = 0,05$

No	Kelas Interval	$f_i = O_i$	$x_i$	$z = \frac{x_i - \bar{x}}{s}$	$p_i$	$E_i = p_i \cdot d$
1	140-149	6	139,5	-2,49	0,0064- 0,0643=0,0579	5,79
2	150-159	22	149,5	-1,52	0,0643- 0,2877=0,2234	22,34
3	160-169	39	159,5	-0,56	0,2877- 0,6554=0,3677	36,77
4	170-179	25	169,5	0,40	0,6554- 0,9147=0,2593	25,93
5	180-189	7	179,5	1,37	0,9147- 0,9901=0,0754	7,54
6	190-199	1	189,5	2,33	0,9901- 0,9995=0,0094	0,94
	200		199,5	3,30		
	Jumlah (d)	100				

rata-rata = 165,3

s = 10,36



- Statistik Uji :

$$\begin{aligned}\chi^2 &= \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i} \\ &= (6 - 5,79)^2/5,79 + (22 - 22,34)^2/22,34 + (39 - 36,77)^2/36,77 + \\ &\quad (25 - 25,93)^2/25,93 + (7 - 7,54)^2/7,54 + (1 - 0,94)^2/0,94 \\ &= 0,0076 + 0,0052 + 0,1352 + 0,0333 + 0,0387 + 0,0038 \\ &= 0,2238\end{aligned}$$

$$\alpha = 0,05,$$

$$db = k - 3 = 6 - 3 = 3,$$

maka berdasarkan Tabel 3 = 7,81.

- Kriteria Uji : tolak  $H_0$  jika  $\chi^2$  hitung  $\geq \chi^2_{\alpha}$  terima dalam hal lainnya , ternyata  $\chi^2$  hitung = 0,2238 >  $\chi^2_{\alpha} = 7,81$ .
- Jadi  $H_0$  diterima, artinya data diatas berdistribusi normal

#### E. UJI CHI-KUADRAT UNTUK UJI INDEPENDENSI

Uji kebebasan ini digunakan untuk memeriksa kebebasan atau independensi dari dua variabel (frekuensi observasi dan frekuensi harapan) sehingga kita dapat menyimpulkan apakah kedua peubah tersebut saling bebas (tidak berpengaruh) ataukah keduanya saling bertalian (berpengaruh).

Data untuk menguji kebebasan dua variabel tersebut disajikan dalam bentuk Tabel Kontingensi atau Tabel Berkemungkinan yang umumnya berukuran r baris x k kolom. Sebelum melakukan pengujian, terlebih dahulu kita harus mendefinisikan Hipotesis Awal ( $H_0$ ) dan Hipotesis Alternatif ( $H_1$ ), yaitu:

- $H_0$  : variabel-variabel saling bebas
- $H_1$  : variabel-variabel tidak saling bebas

Biasanya Tabel Kontingensi berisikan data berupa frekuensi observasi yang diperoleh dari suatu pengujian. Untuk itu, kita perlu mencari frekuensi ekspektasi terlebih dahulu sebelum melakukan pengujian.

$$\text{Frekuensi ekspektasi} = \frac{(\text{total kolom}) \times (\text{total baris})}{\text{total observasi}}$$

Uji kebebasan dirumuskan dalam:

$$\chi^2 = \sum_{i,j=1}^{r,k} \frac{(o_{ij} - e_{ij})^2}{e_{ij}}$$

$\chi^2$  : nilai peubah acak yang distribusi sampelnya didekati oleh distribusi **Chi-Kuadrat** dengan derajat kebebasan  $v=(r-1)(k-1)$

k : jumlah kolom

r : jumlah baris

$o_{ij}$  : frekuensi observasi baris ke-i, kolom ke-j

$e_{ij}$  : frekuensi ekspektasi baris ke-i, kolom ke-j

### Contoh Soal

Suatu percobaan dilakukan untuk mengetahui apakah pendapat penduduk pemilih di negara bagian Illinois mengenai perubahan pajak baru tidak ada hubungannya dengan tingkat penghasilannya. Suatu sampel acak 1000 pemilih yang tercatat di Illinois dikelompokkan menurut apakah penghasilan mereka rendah, sedang, atau tinggi, dan apakah mereka setuju atau tidak terhadap perubahan pajak baru dalam tabel kontingensi berikut: (gunakan taraf uji 0,05)

Perubahan Pajak	Tingkat Pendapatan			Total
	R (Rendah)	M (Menengah)	B (Berada)	
Setuju	182	213	203	598
Tidak Setuju	154	138	110	402
Total	336	351	313	1000

### Jawab :

- $H_0$  : pendapat penduduk pemilih di negara bagian Illinois mengenai perubahan pajak baru dan tingkat penghasilannya saling bebas
- $H_1$  : pendapat penduduk pemilih di negara bagian Illinois mengenai perubahan pajak baru dan tingkat penghasilannya tidak saling bebas
- $\alpha = 0,05$

Daerah kritis  $\chi^2 > 5,991$  dengan derajat kebebasan  $v = (2-1)(3-1) = 2$

Perhitungan  $\chi^2$

Frekuensi harapan untuk:

- Setuju, R =  $\frac{336 \times 598}{1000} = 200,9$
- Tidak Setuju, R =  $\frac{336 \times 402}{1000} = 135,1$
- Setuju, M =  $\frac{351 \times 598}{1000} = 209,9$
- Tidak Setuju, M =  $\frac{351 \times 402}{1000} = 141,1$
- Setuju, B =  $\frac{313 \times 598}{1000} = 187,2$
- Tidak Setuju, B =  $\frac{313 \times 402}{1000} = 125,8$

kategori	$o_{ij}$	$e_{ij}$	$(o_{ij} - e_{ij})$	$(o_{ij} - e_{ij})^2$	$(o_{ij} - e_{ij})^2 / e_{ij}$
Setuju, R	182	200,9	-18,9	357,21	1,78
Setuju, M	213	209,9	3,1	9,61	0,05
Setuju, B	203	187,2	15,8	249,64	1,33
Tidak Setuju, R	154	135,1	18,9	357,21	2,64
Tidak Setuju, M	138	141,1	-3,1	9,61	0,07
Tidak Setuju, B	110	125,8	-15,8	249,64	1,98
$\Sigma$					$\chi^2 \text{ hitung} = 7,85$

6. Keputusan :  $\chi^2 \text{ hitung} > \chi^2 \text{ tabel}$ ,  $H_0$  ditolak
7. Kesimpulan : pendapat penduduk pemilih di negara bagian Illinois mengenai perubahan pajak baru dan tingkat penghasilannya tidak saling bebas

## F. PROSEDUR KERJA DAN CARA INTERPRETASI OUTPUT DARI SPSS

### Soal

Prof D melakukan sebuah penelitian untuk mengetahui pengaruh dari les privat terhadap tingkat kelulusan ujian siswa. Prof D terlebih dahulu mengambil sampel dengan menentukan yang ikut les dan tidak ikut les. Kemudian baru dilihat berapa yang lulus dan tidak lulus. Data yang diperoleh disajikan dalam tabel berikut

		Hasil Ujian		Total
		Ya	Tidak	
Les Privat	Ya	25	10	35
	Tidak	14	51	65
Total		39	61	100

Kemudian Prof D melakukan uji Chi-Square untuk melihat apakah terdapat hubungan dari les privat dengan hasil ujian. Prof D menggunakan taraf signifikansi 5 % .

- Ho: Tidak terdapat hubungan antara les privat dengan hasil ujian atau les privat dan hasil ujian saling bebas
- H1: Terdapat hubungan antara les privat dengan hasil ujian atau les privat dan hasil ujian tidak saling bebas

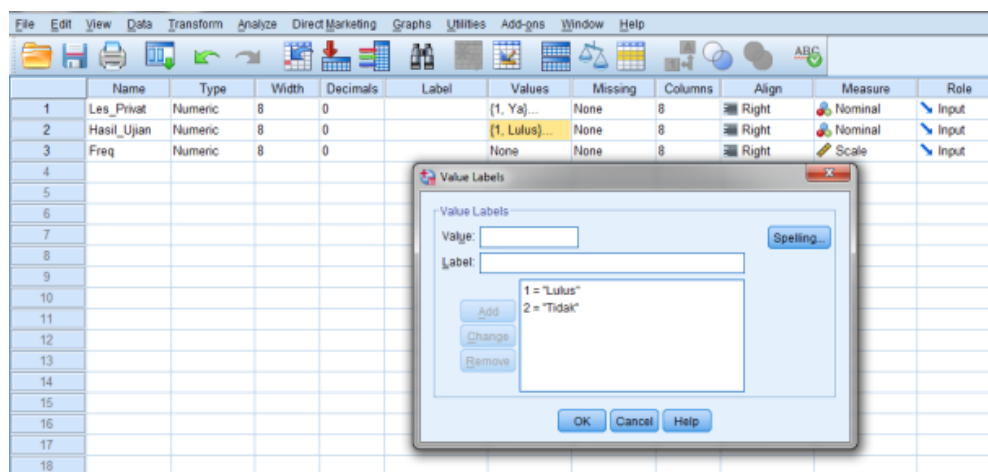
Tolak hipotesis awal (Ho) jika nilai p-value kurang dari atau sama dengan tingkat signifikansi yang dipakai

**Penyelesaian dengan SPSS adalah sebagai berikut:**

Definisikan data pada Variabel View. Disini akan dibuat tiga variabel, yaitu Hasil Ujian, Les Privat, dan Freq untuk variabel jumlahnya. Skala data untuk variabel Hasil Ujian dan les privata adalah Nominal, sedangkan Freq berskala Scale Berikan kode pada kolom Values

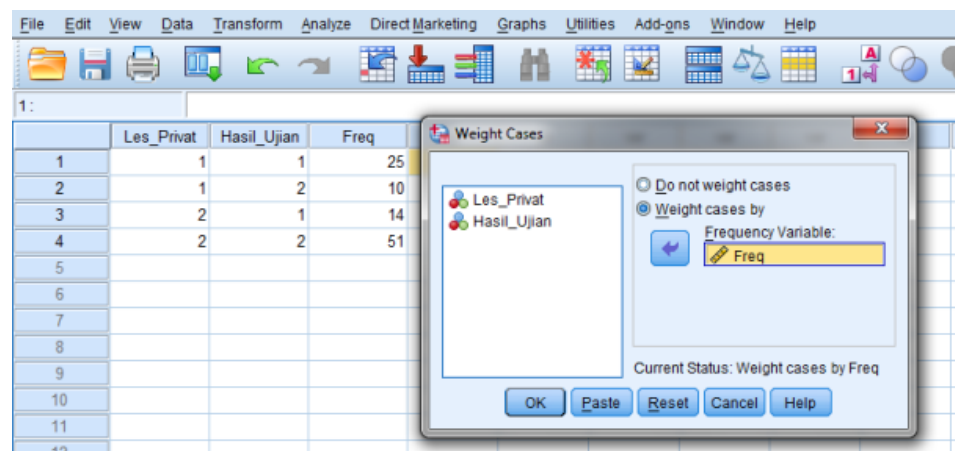
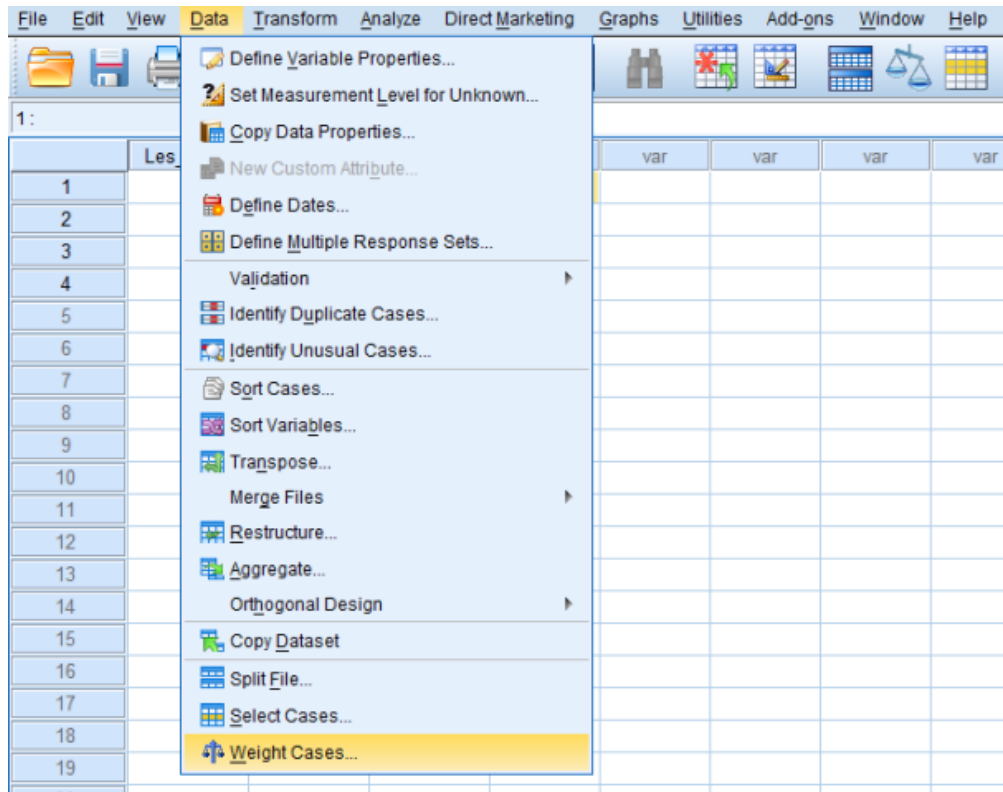
- Untuk variabel Hasil Ujian adalah  
1 =Lulus,      2 = Tidak
- Untuk variabel Les Privat  
1 =Ya,          2 = Tidak

✓ Setelah semua selesai, masukan data kedalam SPSS



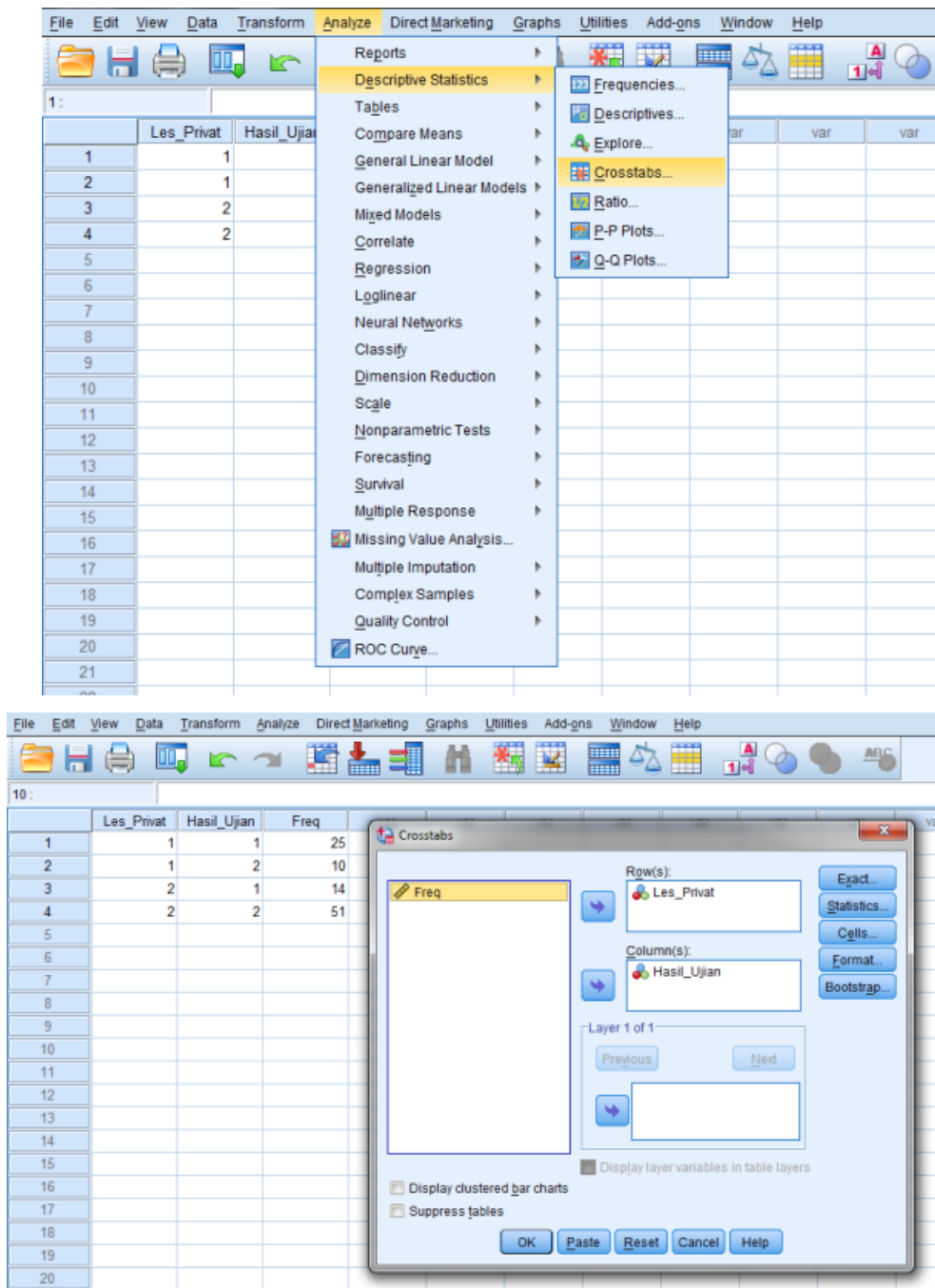
- ✓ Kemudian dilakukan pembobotan dengan Weight Case untuk menghubungkan variabel Hasil Ujian dan Les Privat dengan Freq
  - Pilih Data dan Klik Weight Case
  - Kemudian akan muncul kotak dialog Weight Cases
  - Tandai Weight Cases By, lalu pindahkan Variabel Freq ke Frequency Variable

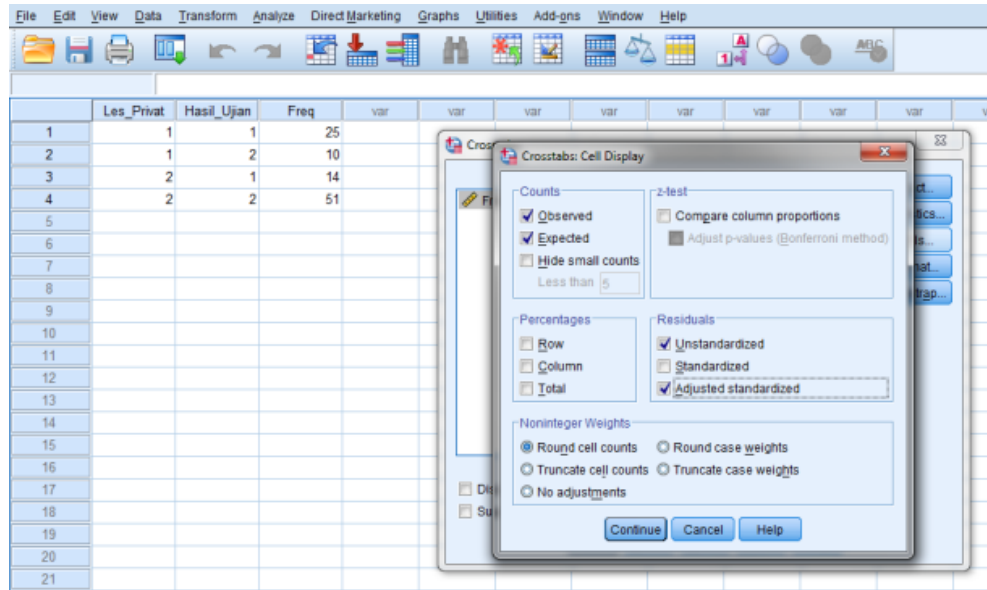
➤ OK



- Pilih Anlyze, Descriptive Statistics, lalu klik Crosstabs
- Kotak dialog Crosstabs muncul, pindahkan Les Privat ke Row(s) dan Hasil Ujian ke Colum(s)
- Klik Statistics, Pada kotak dialog Crosstabs:Statistics yang muncul centang Chi-Square
- Klik Cells, kemudian pilih opsi yang diinginkan, dalam kasus ini dipilih Observed, Expected, Unstandardized, dan Adjusted Standardized. (Step ini bisa dilompat jika tidak membutuhkan)

- Klik Continue, lalu Ok





- **Output**

**Les\_Privat \* Hasil\_Ujian Crosstabulation**

			Hasil_Ujian		Total
			Lulus	Tidak	
Les_Privat	Ya	Count	25	10	35
		Expected Count	13.7	21.4	35.0
		Residual	11.4	-11.4	
		Adjusted Residual	4.9	-4.9	
	Tidak	Count	14	51	65
		Expected Count	25.4	39.7	65.0
		Residual	-11.4	11.4	
		Adjusted Residual	-4.9	4.9	
Total	Count	39	61	100	
	Expected Count	39.0	61.0	100.0	

**Chi-Square Tests**

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
<u>Pearson Chi-Square</u>	<u>23.802<sup>a</sup></u>	1	<u>.000</u>		
Continuity Correction <sup>b</sup>	21.751	1	.000		
Likelihood Ratio	24.140	1	.000		
Fisher's Exact Test				.000	.000
Linear-by-Linear Association	23.564	1	.000		
N of Valid Cases	100				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 13.65.

b. Computed only for a 2x2 table

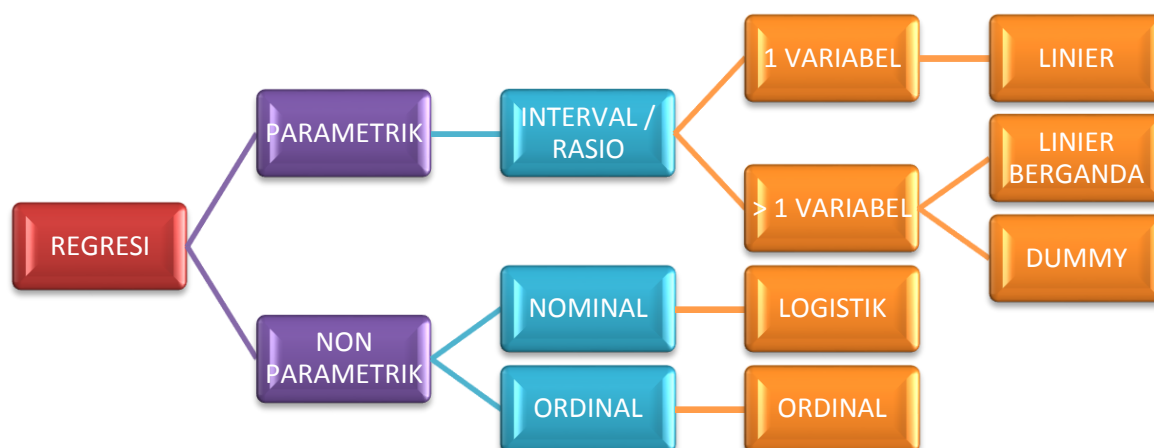
- Hasil Output untuk tabel Les\_Privat\*Hasil\_Ujian Crosstabulation dapat diinterpretasikan
- Nilai Adjusted Residual yang cukup besar (dengan nilai absolut 4,9) menunjukkan deviasi atau perbedaan yang besar pada setiap sel dari nilai yang diharapkan. Terlihat bahwa nilai Residual yang merupakan selisih antar nilai Observed dengan Expected Frequency menjauh dari nol. Jika hipotesis awal benar, maka kita mengharapkan nilai residualnya akan menjadi nol atau mendekati nol.
- Hasil Output untuk tabel Chi-Square Test dapat diinterpretasikan. Pada tabel ini tidak hanya nilai Chi-Square saja yang ditampilkan tapi ada juga nilai dari uji dengan menggunakan Likelihood Ratio dan Fisher's Exact Test. Didapa nilai dari Chi-Square adalah 23,802 dan  $p\text{-value} < 0,0001$  sehingga dapat diambil tolak hipotesis awal dan menyatakan terdapat hubungan yang signifikan antara les privat dengan hasil ujian pada taraf signifikansi 5%. Hasil ini juga sejalan dengan uji Likelihood Ratio dimana nilainya adalah 24,140 dan  $p\text{-value} < 0,001$  karena nilai  $p\text{-value} < 0,05$  tolak hipotesis awal juga.



# BAB 11 REGRESI

---

## KERANGKA KONSEP



### A. PENGERTIAN REGRESI

Istilah regresi pertama kali dikenalkan oleh Francis Galton (Gujarati & Porter, 2010: 19). Untuk mengukur besarnya pengaruh variabel bebas terhadap variabel tergantung dan memprediksi variabel tergantung dengan menggunakan variabel bebas. Gujarati (2006) mendefinisikan analisis regresi sebagai kajian terhadap hubungan satu variabel yang disebut sebagai variabel yang diterangkan (the explained variabel) dengan satu atau dua variabel yang menerangkan (the explanatory). Variabel pertama disebut juga sebagai variabel tergantung dan variabel kedua disebut juga sebagai variabel bebas. Jika variabel bebas lebih dari satu, maka analisis regresi disebut regresi linear berganda. Disebut berganda karena pengaruh beberapa variabel bebas akan dikenakan kepada variabel tergantung.

Interpretasi istilah regresi menurut Gujarati & Porter (2010: 20) Regresi adalah suatu metode analisis statistik yang digunakan untuk melihat pengaruh antara dua atau lebih variabel. Hubungan variabel tersebut bersifat fungsional yang diwujudkan dalam suatu model matematis. Pada analisis regresi, variabel dibedakan menjadi dua bagian, yaitu variabel respons (response variable) atau biasa juga disebut

variabel bergantung (*dependent variable*) dan variabel explanatory atau biasa disebut penduga (*predictor variable*) atau disebut juga variabel bebas (*independent variabel*).

Regresi dalam pengertian moderen menurut Gujarati (2009) ialah sebagai kajian terhadap ketergantungan satu variabel, yaitu variabel tergantung terhadap satu atau lebih variabel lainnya atau yang disebut sebagai variabel – variabel eksplanatori dengan tujuan untuk membuat estimasi dan / atau memprediksi rata – rata populasi atau nilai rata-rata variabel tergantung dalam kaitannya dengan nilai – nilai yang sudah diketahui dari variabel eksplanatorinya. Selanjutnya menurut Gujarati meski analisis regresi berkaitan dengan ketergantungan atau dependensi satu variabel terhadap variabel – variabel lainnya hal tersebut tidak harus menyiratkan sebab – akibat (*causation*). Dalam mendukung pendapatnya ini, Gujarati mengutip pendapat Kendal dan Stuart yang diambil dari buku mereka yang berjudul “*The Advanced Statistics*” yang terbit pada tahun 1961 yang mengatakan bahwa, ” suatu hubungan statistik betapapun kuat dan sugestifnya tidak akan pernah dapat menetapkan hubungan sebab akibat (*causal connection*); sedang gagasan mengenai sebab akibat harus datang dari luar statistik, yaitu dapat berasal dari teori atau lainnya”.

Sedang menurut Levin & Rubin (1998:648), regresi digunakan untuk menentukan sifat – sifat dan kekuatan hubungan antara dua variabel serta memprediksi nilai dari suatu variabel yang belum diketahui dengan didasarkan pada observasi masa lalu terhadap variabel tersebut dan variabel-variabel lainnya. Selanjutnya dalam regresi kita akan mengembangkan persamaan estimasi (*estimating equation*), yaitu rumus matematika yang menghubungkan variabel-variabel yang diketahui dengan variabel-variabel yang tidak diketahui. Setelah dipelajari pola hubungannya, kemudian kita dapat mengaplikasikan analisis korelasi (*correlation analysis*) untuk menentukan tingkatan dimana variabel – variabel tersebut berhubungan. Kesimpulannya, analisis korelasi mengungkapkan seberapa benar persamaan estimasi sebenarnya menggambarkan hubungan tersebut. Lebih lanjut Levin & Rubin mengatakan bahwa: “ Kita sering menemukan hubungan sebab akibat antar variabel – variabel; yaitu variabel bebas ‘menyebabkan’ variabel tergantung berubah. Sekalipun demikian mereka melanjutkan bahwa: “penting untuk kita perhatikan bahwa yang kita anggap hubungan (*relationship*) yang diketemukan melalui regresi sebagai hubungan asosiasi (*relationship of association*) tetapi tidak selalu harus sebab dan akibat (*cause and effect*). Kecuali kita mempunyai alasan –

alasan khusus untuk percaya bahwa (perubahan pada) nilai – nilai variabel tergantung disebabkan oleh nilai – nilai variabel (variabel) bebas; jangan menyimpulkan (*infer*) hubungan sebab akibat dari hubungan yang ditemukan dalam regresi.

Analisis regresi dalam statistika adalah salah satu metode untuk menentukan hubungan sebab-akibat antara satu variabel dengan variabel(-variabel) yang lain. Variabel "penyebab" disebut dengan bermacam-macam istilah: variabel penjelas, variabel eksplanatorik, variabel independen, atau secara bebas, variabel X (karena seringkali digambarkan dalam grafik sebagai absis, atau sumbu X). Variabel terkena akibat dikenal sebagai variabel yang dipengaruhi, variabel dependen, variabel terikat, atau variabel Y. Kedua variabel ini dapat merupakan variabel acak (random), namun variabel yang dipengaruhi harus selalu variabel acak.

Analisis regresi adalah salah satu analisis yang paling populer dan luas pemakaiannya. Analisis regresi dipakai secara luas untuk melakukan prediksi dan ramalan, dengan penggunaan yang saling melengkapi dengan bidang pembelajaran mesin. Analisis ini juga digunakan untuk memahami variabel bebas mana saja yang berhubungan dengan variabel terikat, dan untuk mengetahui bentuk-bentuk hubungan tersebut.

Analisis regresi digunakan hampir pada semua bidang kehidupan, baik dalam bidang pertanian, ekonomi dan keuangan, industri dan ketenagakerjaan, sejarah, pemerintahan, ilmu lingkungan, dan sebagainya. Kegunaan analisis regresi di antaranya untuk mengetahui variabel-variabel kunci yang memiliki pengaruh terhadap suatu variabel bergantung, pemodelan, serta pendugaan (estimation) atau peramalan (forecasting). Adapun tahap-tahap dalam melakukan analisis regresi, meliputi perumusan permasalahan, penyeleksian variabel potensial yang relevan, pengumpulan data, spesifikasi model, pemilihan metode yang tepat, model fitting, validasi model dan penerapan model terpilih untuk penyelesaian permasalahan.

Tujuan Penggunaan Analisis Regresi antara lain:

- a) Membuat estimasi rata-rata dan nilai variabel tergantung dengan didasarkan pada nilai variabel bebas.
- b) Untuk menguji hipotesis karakteristik dependensi.

- c) Meramalkan nilai rata-rata variabel bebas yang didasari nilai variabel bebas diluar jangkauan sample.

## B. ASUMSI PENGGUNAAN REGRESI

Dalam analisis regresi ada dua macam linearitas, yaitu linieritas dalam variabel dan linieritas dalam parameter. Linier dalam variabel merupakan nilai rata-rata kondisional variabel tergantung yang merupakan fungsi linier dari variabel (variabel) bebas. Sedangkan linieritas dalam parameter merupakan fungsi linier parameter dan dapat tidak linier dalam variabel.

Analisis regresi berbeda dengan analisis korelasi. Jika dalam analisis korelasi digunakan untuk melihat hubungan dua variable, maka analisis regresi digunakan untuk melihat pengaruh variable bebas terhadap variable tergantung serta memprediksi nilai variable tergantung dengan menggunakan variable bebas. Dalam analisis regresi variable bebas berfungsi untuk menerangkan (explanatory) sedang variable tergantung berfungsi sebagai yang diterangkan (the explained). Dalam analisis regresi data harus berskala interval atau rasio. Hubungan dua variable bersifat dependensi. Untuk menggunakan analisis regresi diperlukan beberapa persyaratan yang harus dipenuhi.

1. Model regresi harus linier dalam parameter.
2. Variabel bebas tidak berkorelasi dengan disturbance term (Error).
3. Nilai disturbance term sebesar 0 atau dengan simbol sebagai berikut:  $(E(U / X) = 0$
4. Varian untuk masing-masing error term (kesalahan) konstan.
5. Tidak terjadi otokorelasi
6. Model regresi hendaknya dispesifikasi secara benar. Tidak terdapat bias spesifikasi dalam model yang digunakan dalam analisis empiris.
7. Jika variabel bebas lebih dari satu, maka antara variabel bebas (explanatory) tidak ada hubungan linier yang nyata.

## C. MACAM REGRESI DAN TEKNIK PENGHITUNGAN

### 1. Regresi untuk Uji Parametrik

Asumsi yang digunakan sama dengan uji Korelasi untuk uji parametrik

## a) Regresi Linier

Terdapat dua syarat yang harus dipenuhi oleh data dalam menggunakan analisis regresi linier yaitu:

- **Data**

Data harus terdiri dari dua jenis variabel, yaitu dependen dan independen. Selain itu data berupa kuantitatif dan variabel berupa kategori, seperti SD, SMA, SMK, dll.

- **Asumsi**

Setiap data diasumsikan variabel dependen terdistribusi secara normal. Selain itu, antara variabel dependen dan independen harus memiliki hubungan linier dengan observasi harus saling bebas.

Regresi linear sederhana adalah hubungan secara linear antara satu variabel independen (X) dengan variabel dependen (Y). Analisis ini digunakan untuk mengetahui arah hubungan antara variabel independen dengan variabel dependen apakah positif atau negatif serta untuk memprediksi nilai dari variabel dependen apabila nilai variabel independen mengalami kenaikan atau penurunan nilai. Data yang digunakan biasanya berskala interval atau rasio.

Rumus dari analisis regresi linear sederhana adalah sebagai berikut:

$$Y = a + bX$$

**Keterangan:**

Y = subyek dalam variabel dependen yang diprediksi

a = harga Y ketika harga X = 0 (harga konstan)

b = angka arah atau koefisien regresi, yang menunjukkan angka peningkatan ataupun penurunan variabel dependen yang didasarkan pada perubahan variabel independen. Bila (+) arah garis naik, dan bila (-) maka arah garis turun.

X = subyek pada variabel independen yang mempunyai nilai tertentu.

Secara teknik harga b merupakan tangent dari perbandingan antara panjang garis variabel dependen, setelah persamaan regresi ditemukan.

$$\text{Harga } b = r \frac{s_y}{s_x} \dots\dots\dots \text{Rumus 2}$$

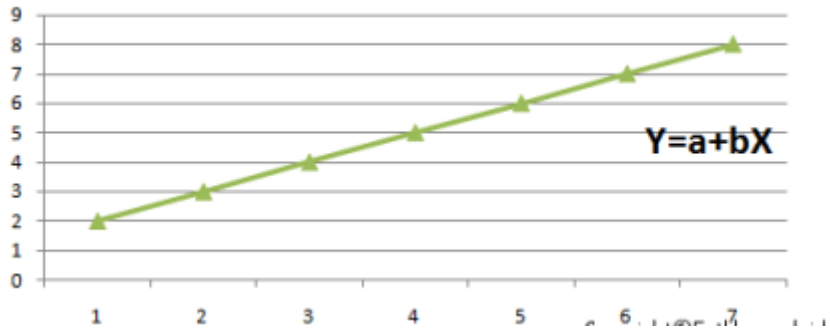
$$\text{Harga } a = Y - b X \dots\dots\dots \text{Rumus 3}$$

**Dimana :**

R = koefisien korelasi product moment antara variabel variabel X dengan variabel Y

S<sub>y</sub> = simpangan baku variabel Y

S<sub>x</sub> = simpangan baku variabel X



Jika harga b merupakan fungsi dari koefisien korelasi. Bila koefisien korelasi tinggi, maka harga b juga besar, sebaliknya bila koefisien korelasi rendah maka harga b juga rendah (kecil). Selain itu bila koefisien korelasi negatif maka harga b juga negatif, dan sebaliknya bila koefisien korelasi positif maka harga b juga positif.

Selain itu harga a dan b dapat dicari dengan rumus berikut:

$$a = \frac{(\sum Y)(\sum X^2) - (\sum X)(\sum XY)}{n \sum X^2 - (\sum X)^2} \dots\dots\dots \text{Rumus 5}$$

$$b = \frac{n \sum XY - (\sum X)(\sum Y)}{n \sum X^2 - (\sum X)^2} \dots\dots\dots \text{Rumus 6}$$

**1) Uji Koefisien Regresi Sederhana (Uji t)**

Uji ini digunakan untuk mengetahui apakah variabel independen (X) berpengaruh secara signifikan terhadap variabel dependen (Y). Signifikan berarti pengaruh yang terjadi dapat berlaku untuk populasi (dapat digeneralisasikan).

Rumus t hitung pada analisis regresi adalah sebagai berikut:

$$t \text{ hitung} = \frac{b}{s_b} \quad \text{atau} \quad t \text{ hitung} = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

Keterangan:

- b = Koefisien regresi
- S<sub>b</sub> = Standar error
- r = Koefisien korelasi sederhana
- n = jumlah data atau kasus

Langkah-langkah pengujian koefisien regresi sederhana adalah sebagai berikut:

- a. Menentukan Hipotesis
  - H<sub>0</sub> = Tidak ada pengaruh yang signifikan
  - H<sub>a</sub> = ada pengaruh yang signifikan

**2) Menentukan tingkat signifikansi**

Biasanya menggunakan α = 5% atau 0,05

**3) Menentukan t hitung**

- Menentukan t tabel
- Membandingkan t hitung dan t table dengan kriteria

H<sub>0</sub> diterima jika: t hitung ≥ t tabel

H<sub>0</sub> ditolak jika: t hitung < t tabel

H<sub>0</sub> diterima jika: -t hitung ≤ t tabel

H<sub>0</sub> di tolak jika: -t hitung > t tabel

Analisis regresi linier sederhana adalah hubungan secara linear antara satu variabel independen (X) dengan variabel dependen (Y). Analisis ini untuk mengetahui arah hubungan antara variabel independen dengan variabel dependen apakah positif atau negatif dan untuk memprediksi nilai dari variabel dependen apabila nilai variabel independen mengalami kenaikan atau penurunan.. Data yang digunakan biasanya berskala interval atau rasio.

**b) Regresi Linier Berganda**

Analisis regresi linier berganda adalah hubungan secara linear antara dua atau lebih variabel independen (X<sub>1</sub>, X<sub>2</sub>,...X<sub>n</sub>) dengan variabel dependen (Y). Analisis ini untuk mengetahui arah hubungan antara variabel independen dengan variabel dependen apakah masing-masing variabel independen

berhubungan positif atau negatif dan untuk memprediksi nilai dari variabel dependen apabila nilai variabel independen mengalami kenaikan atau penurunan. Data yang digunakan biasanya berskala interval atau rasio.

Regresi linier berganda adalah analisis regresi yang menjelaskan hubungan antara peubah respon (variabel dependen) dengan faktor-faktor yang mempengaruhi lebih dari satu prediktor (variabel independen).

Regresi linier berganda hampir sama dengan regresi linier sederhana, hanya saja pada regresi linier berganda variabel bebasnya lebih dari satu variabel penduga. Tujuan analisis regresi linier berganda adalah untuk mengukur intensitas hubungan antara dua variabel atau lebih dan membuat prediksi perkiraan nilai Y atas X.

Secara umum model regresi linier berganda untuk populasi adalah sebagai berikut:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_n X_n + \varepsilon$$

Di mana  $\beta_0, \beta_1, \beta_2, \dots, \beta_k$  adalah koefisien atau parameter model.

Model regresi linier berganda untuk populasi diatas dapat ditaksir berdasarkan sebuah sampel acak yang berukuran n dengan model regresi linier berganda untuk sampel, yaitu:

$$\hat{Y} = a_0 + a_1 X_1 + a_2 X_2 + a_3 X_3 + \dots + a_n X_n$$

Dengan:

$\hat{Y}$  = Nilai taksiran bagi variabel  $Y$

$a_0$  = Taksiran bagi parameter konstanta  $a_0$

$a_1, a_2, a_3$  = Taksiran bagi parameter koefisien regresi  $a_1, a_2, a_3$

Bentuk data yang akan diolah ditunjukkan pada table berikut:

Tabel 2.1 Bentuk Umum Data Observasi



Nomor Observasi	Responden ( $Y_i$ )	Variabel Bebas			
		$X_{1i}$	$X_{2i}$	...	$X_{ki}$
1	$Y_1$	$X_{11}$	$X_{21}$	...	$X_{k1}$
2	$Y_2$	$X_{12}$	$X_{22}$	...	$X_{k2}$
.	.	.	.	...	.
.	.	.	.	...	.

### 1) Membentuk Persamaan Regresi Linier Berganda

Dalam regresi linier berganda variabel tak bebas ( $y$ ) tergantung kepada dua atau lebih variabel bebas ( $x$ ). Bentuk persamaan regresi linier berganda yang mencakup dua atau lebih variabel dapat ditulis sebagai berikut:

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_n X_n + \varepsilon_i$$

Dengan:

$$i = 1, 2, \dots, n$$

$$n = \text{ukuran sampel}$$

$$\varepsilon_i = \text{variabel kesalahan (galat)}$$

Untuk rumus diatas, dapat diselesaikannya dengan empat persamaan oleh empat variabel yang terbentuk:

$$\sum Y_i = n\beta_0 + \sum \beta_1 X_{1i} + \sum \beta_2 X_{2i} + \sum \beta_3 X_{3i}$$

$$\sum X_{1i} Y_i = \beta_0 \sum X_{1i} + \beta_1 \sum (X_{1i})^2 + \beta_2 \sum X_{1i} X_{2i} + \beta_3 \sum X_{1i} X_{3i}$$

$$\sum X_{2i} Y_i = \beta_0 \sum X_{2i} + \beta_1 \sum X_{1i} X_{2i} + \beta_2 \sum (X_{2i})^2 + \beta_3 \sum X_{2i} X_{3i}$$

$$\sum X_{3i} Y_i = \beta_0 \sum X_{3i} + \beta_1 \sum X_{1i} X_{3i} + \beta_2 \sum X_{2i} X_{3i} + \beta_3 \sum (X_{3i})^2$$

Dengan  $\beta_0, \beta_1, \beta_2, \beta_3$  adalah koefisien yang ditentukan berdasarkan data hasil pengamatan.

Regresi Linear Berganda adalah model regresi linear dengan melibatkan lebih dari satu variabel bebas atau predictor. Dalam bahasa Inggris, istilah ini disebut dengan multiple linear regression.

Pada dasarnya regresi linear berganda adalah model prediksi atau peramalan dengan menggunakan data berskala interval atau rasio serta terdapat lebih dari satu predictor.

Skala data yang dimaksud diatas adalah pada semua variabel terutama variable terikat. Pada regresi linear, tidak menutup kemungkinan digunakannya data dummy pada variable bebas. Yaitu pada regresi linear dengan dummy.

## 2) Perbedaan Regresi Linear Berganda dan Sederhana

Seperti yang sudah saya bahas diatas, dikatakan regresi linear berganda jika jumlah variable bebas lebih dari satu. Sedangkan jika jumlah variable bebas hanya ada satu saja, maka itu yang disebut dengan regresi linear sederhana.

Model regresi linear berganda dilukiskan dengan persamaan sebagai berikut:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_n X_n + e$$

### **Keterangan:**

Y = Variabel terikat atau response.

X = Variabel bebas atau predictor.

$\alpha$  = Konstanta.

$\beta$  = Slope atau Koefisien estimate.

## 3) Contoh Regresi Linear Berganda

Salah satu contoh penelitian yang menggunakan analisis regresi linear berganda adalah penelitian dengan judul: “pengaruh ROA, NPM dan Size Terhadap Return Saham.” Untuk contoh penelitian ini sudah saya jelaskan secara rinci dalam artikel tentang contoh penelitian bisnis dengan regresi linear.

Berdasarkan contoh penelitian diatas, sudah jelas bahwasanya semua variable berskala data interval atau rasio. Dengan kata lain data yang digunakan adalah data kuantitatif atau numeric. Dalam contoh tersebut juga dijelaskan bahwasanya variable bebas lebih dari satu, yaitu ROA, NPM dan Size.

- **Asumsi Klasik regresi Linear Berganda**

Seperti halnya uji parametris lainnya, maka regresi linear juga mempunyai syarat atau asumsi klasik yang harus terpenuhi. Agar model prediksi yang dihasilkan nantinya bersifat BLUE (Best Linear Unbiased Estimation).

Asumsi klasik pada regresi linear berganda antara lain: Data interval atau rasio, linearitas, normalitas, non outlier, homoskedastisitas, non multikolinearitas dan non autokorelasi.

### c) Regresi Dummy

Kata “*dummy*” kalau diterjemahkan secara harfiah artinya “buatan”. Jadi yang dimaksud dengan variabel dummy adalah variabel yang nilainya sebenarnya adalah buatan, karena nilai variabel tersebut sebenarnya bukanlah skala. Atau secara mudahnya, variabel dummy adalah variabel independen yang wujudnya berskala non-metrik atau kategori. Jika variabel independen berukuran kategori atau dikotomi, maka dalam model regresi kita harus menyatakan sebagai variabel dummy dengan memberi kode 0 atau 1. Setiap variabel dummy menyatakan satu kategori variabel, dan setiap variabel dengan k kategori dapat dinyatakan dalam k-1 variabel dummy.

Analisis dengan variabel dummy dilakukan pada saat kita tertarik pada pengaruh variabel independen kategori, atau kita ingin memasukkan variabel kategori tersebut untuk meningkatkan kualitas penelitian kita. Analisis regresi dengan menggunakan variabel dummy memiliki kompleksitas, hasil analisis ini memiliki kemiripan dengan analisis kovarian (anakova), namun sistematis komputasinya sedikit berbeda. Ingat, prosedur yang disajikan dalam tulisan ini adalah prosedur analisis regresi dengan variabel independen berupa kategori, namun jika yang berwujud kategori adalah variabel dependen, maka teknik analisis yang dilakukan adalah dengan analisis regresi logistik.

Misalkan kita ingin variabel demografi apa saja yang mampu memprediksi agresivitas. Desain penelitian kita adalah sebagai berikut.

- Variabel dependen : agresivitas
- Variabel independen : umur, jenis kelamin, suku

Kita punya tiga variabel independen yakni umur (X1), jenis kelamin (X2), dan suku (X3). Umur adalah variabel kuantitatif, sedangkan jenis kelamin dan suku merupakan kategori. Jenis kelamin terdiri atas dua kategori, yakni laki-laki dan perempuan, sedangkan suku terdiri atas tiga kategori, yakni Jawa, Sunda, Madura. Karena jenis kelamin terdiri atas 2 kategori, maka variabel jenis kelamin hanya memiliki satu variabel dummy (2-1), sedangkan suku terdiri atas 3 kategori, sehingga variabel suku memiliki 2 variabel dummy (3-1).

Pengkodean data kategorikal memerlukan pengkategorian eksklusif. Artinya satu subjek/sampel adalah masuk dalam satu kategori, tidak boleh dua kategori. Sampel A misalnya, tidak boleh masuk ke dalam kategori laki-laki dan perempuan. Si B tidak boleh masuk ke dalam kategori PNS dan wiraswasta meskipun kedua profesi tersebut dijalannya. Aturan ini berlaku variabel dummy. Sebuah variabel dengan kategori sebanyak k akan membutuhkan seperangkat k – 1 variabel dummy untuk menjangkau semua informasi yang terkandung didalamnya. Jadi misalnya saya memiliki variabel tingkat pendidikan dari SD hingga PT (4 kategori) maka jumlah variabel kategori yang sama buat ada 3 kategori. Menggunakan pola koding biner (0,1), variabel dummy selalu variabel berbentuk dikotomi.

Semua responden yang menjadi anggota kategori yang diberi kode 1 sedangkan responden tidak dalam kategori tersebut dikode dari 0. Dengan cara seperti ini maka setiap responden akan memiliki kode 1 pada kategori yang sesuai dengannya dan kode 0 pada kategori yang tidak sesuai dengannya. Kode biner dapat dianggap sebagai mirip ke saklar listrik: kode A 1 sinyal bahwa kategori yang diberikan adalah “on” untuk responden (misalnya, dia adalah anggota dari kelompok tertentu, atau karakteristik tertentu hadir ), karena bukan anggota, variabel dummy yang menunjukkan kategori yang diaktifkan “off (yaitu, karakteristik ini tidak ada).

Variabel kualitatif tersebut harus dikuantitatifkan atributnya (cirinya). Untuk mengkuantitatifkan atribut variabel kualitatif, dibentuk variabel dummy dgn nilai 1 dan 0. Jadi, inilah yang dimaksud dengan variabel dummy tersebut. Nilai 1 menunjukkan adanya, sedangkan nilai 0 menunjukkan tidak adanya ciri kualitas tsb. Misalnya variabel jenis kelamin. Jika nilai 1

digunakan untuk laki-laki maka nilai 0 menunjukkan bukan laki-laki (perempuan), atau sebaliknya. (Kategori yg diberi nilai 0 disebut kategori dasar, dlm artian bahwa perbandingan dibuat atas kategori tsb.)

Jika variabel kualitatif tsb terdiri lebih dari dua kategori, jumlah variabel dummy yg dibentuk harus sebanyak  $n-1$ , dimana  $n$  adalah banyaknya kategori variabel tsb. Misalnya variabel jenis pekerjaan dgn tiga kategori yaitu pekerja kasar, setengah terampil dan pekerja terampil, maka dibentuk dua variabel dummy sbb:

- Alternatif 1

$Dk1 = 1$  jika kasar,  $= 0$  jika lainnya

$Dk2 = 1$  jika setengah terampil,  $= 0$  jika lainnya

Pada alternatif 1,

$Dk1 = 0, Dk2 = 0$  ® terampil

$Dk1 = 0, Dk2 = 1$  ®  $\frac{1}{2}$  terampil

$Dk1 = 1, Dk2 = 0$  ® kasar

- Alternatif 2

$Dk1 = 1$  jika setengah terampil,  $= 0$  jika lainnya

$Dk2 = 1$  jika terampil,  $= 0$  jika lainnya

Pada alternatif 2,

$Dk1 = 0, Dk2 = 0$  ® kasar

$Dk1 = 0, Dk2 = 1$  ® terampil

$Dk1 = 1, Dk2 = 0$  ®  $\frac{1}{2}$  terampil

(Catatan: untuk lebih jelasnya lihat data yang dilampirkan. Dalam data tersebut, kita menggunakan alternatif 1 untuk membentuk variabel dummy jenis pekerjaan).

- **Contoh Desain Penelitian**

Sebuah penelitian hendak menguji prediksi ketahanan kerja terhadap kepuasan kerja. Karena dari teori dinyatakan bahwa kepuasan kerja turut dipengaruhi oleh jabatan dan bidang kerja, maka peneliti turut memasukkannya dalam penelitian. Formulasi informasi variabelnya sebagai berikut :

- Variabel dependen : kepuasan kerja (Y)
- Variabel independen : ketahanan kerja (X1), jabatan (X2), bidang kerja (X3)

Variabel kepuasan kerja dan ketahanan kerja adalah variabel kuantitatif karena bersifat kontinum karena didapatkan dari pengukuran melalui skala Jabatan dan bidang kerja adalah variabel kualitatif karena bersifat kategorikal. Dalam hal ini jabatan ada 2 kategori : pelaksana & manajerial. Bidang kerja ada 3 yaitu suplai, produksi, dan marketing.

**a) Regresi Atas Satu Variabel Kualitatif (Dua Kategori)**

Dari data yang diberikan, misalnya kita ingin meregresikan variabel kualitatif jenis kelamin (sex) terhadap penghasilan, dengan kuantifikasi laki-laki =1 dan perempuan =0

Model regresi yang kita bentuk :

$$Y = b_0 + b_1 D_s$$

Dimana : Y = penghasilan

$D_s$  = variabel dummy untuk seks

Berdasarkan pengolahan data, persamaan regresi yang terbentuk sebagai berikut:

$Y =$	647884	+	552801 $D_s$	$R^2 = 0,205$
	(130773)		(176335)	df = 38
t =	(4,95)**		(3,14)**	F = 9,82**

Model ini memungkinkan untuk mengetahui perbedaan penghasilan berdasarkan jenis kelamin.

Interpretasinya :  $b_0$  = rata-rata penghasilan pekerja perempuan,  $b_1$  = besarnya perbedaan penghasilan pekerja laki-laki dengan perempuan, dan  $b_0 + b_1$  merupakan rata-rata penghasilan pekerja laki-laki. Karena variabel  $D_s$  signifikan, artinya secara statistik terdapat perbedaan nyata penghasilan laki-laki dengan perempuan

Jika  $D_s=0$ , artinya rata-rata penghasilan pekerja perempuan adalah 647884 (sebesar  $b_0$ ). Jika  $D_s=1$ , artinya rata-rata penghasilan pekerja laki-laki adalah  $647884 + 552801 = 1200685$  (sebesar  $b_0 + b_1$ )

**b) Regresi Atas Satu Variabel Kuantitatif dan Satu Variabel Kualitatif (dua kategori)**

Dari data yang sudah diberikan, misalnya kita bentuk suatu model regresi sbb:

$$Y = b_0 + b_1 D_s + b_2 P_d + e_i$$

Dimana :  $D_s$  = dummy jenis kelamin

$P_d$  = pendidikan

Berdasarkan pengolahan datanya, ditampilkan hasilnya sebagai berikut :

Y =	-1039341	+ 331928 $D_s$	+ 134978 $P_d$	$R^2 = 0,52$
	(378566)	(152620)	(27847)	df = 37
t =	(-2,90)**	(2,27)*	(4,91)**	F = 19,94**

Interpretasi : Pada setiap tingkat pendidikan, terdapat perbedaan besarnya penghasilan laki-laki dan perempuan. Perbedaan tsb besarnya adalah 331928 (sebesar  $b_1$ ) lebih tinggi untuk laki-laki

✓ Misalnya jika tahun pendidikan ( $P_d$ ) = 15, maka

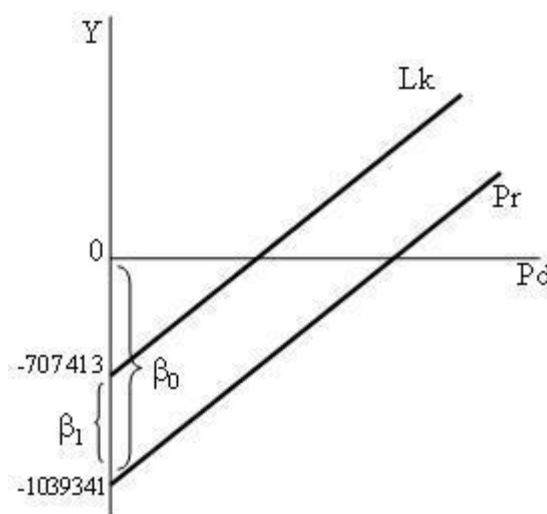
Jika  $D_s=0$ , rata-rata penghasilan perempuan adalah

$$Y = -1039341 + 331928 (0) + 134978 (15) = 985329$$

Jika  $D_s=1$ , rata-rata penghasilan laki-laki adalah

$$Y = -1039341 + 331928 (1) + 134978 (15) = 1317257$$

Secara grafik, hal ini dapat digambarkan sebagai berikut:



Ini memberikan arti fungsi penghasilan laki-laki dan perempuan sehubungan pendidikannya memiliki slope yg sama, tetapi intersep berbeda.

Dengan kata lain, rata-rata penghasilan laki-laki dan perempuan berbeda, tetapi tingkat perubahan penghasilan yg disebabkan oleh tahun pendidikan adalah sama

**c) Regresi Atas Satu Variabel Kuantitatif dan Satu Variabel Kualitatif (lebih dua kategori)**

Dari data yang diberikan, misalnya kita bentuk suatu model regresi sbb:

$$Y = b_0 + b_1 Dk1 + b_2 Dk2 + b_3 Pd + e_i$$

Dimana : Dk1 = dummy jenis pekerjaan

1 = setengah terampil

0 = lainnya

Dk2 = dummy jenis pekerjaan

1 = terampil

0 = lainnya

Pd = pendidikan

Berdasarkan pengolahan datanya, ditampilkan hasilnya sebagai berikut :

Y =	-1286520	+475244 Dk1	+590905 Dk2	+145168 Pd	R <sup>2</sup> = 0,62
	(314191)	(22870)	(145658)	(146349)	df = 36
t =	(-4,09)**	(6,35)**	(32,26)**	(4,04)**	F = 21,91**

Interpretasi :

Jika Dk1 = 0, Dk2 = 0 @ pekerja kasar

Jika Dk1 = 1, Dk2 = 0 @ pekerja setengah terampil

Jika Dk1 = 0, Dk2 = 1 @ pekerja terampil

- Pada setiap tingkat pendidikan, terdapat perbedaan penghasilan pekerja kasar, setengah terampil dan terampil.
- Pekerja setengah terampil memiliki penghasilan 475244 (sebesar b<sub>1</sub>) lebih tinggi dibandingkan pekerja kasar.
- Pekerja terampil memiliki penghasilan 590905 (sebesar b<sub>2</sub>) lebih tinggi dibandingkan pekerja kasar.

Misalnya jika tahun pendidikan (Pd) = 15, maka

- Jika Dk1=0 dan Dk2=0, rata-rata penghasilan pekerja kasar adalah:
- $Y = -1286520 + 475244 (0) + 590905 (0) + 145168 (15) = 891000$
- Jika Dk1=1 dan Dk2=0, rata-rata penghasilan pekerja ½ terampil :



- $Y = -1286520 + 475244 (1) + 590905 (0) + 145168 (15) = 1366244$
- Jika  $Dk1=0$  dan  $Dk2=1$ , rata-rata penghasilan pekerja terampil:
- $Y = -1286520 + 475244 (1) + 590905 (1) + 145168 (15) = 1481905$

#### d) Membandingkan Dua Regresi dengan Variabel Dummy

Secara implisit, teknik penggunaan variabel dummy (dari pembahasan sebelum ini), pada dasarnya mengandung asumsi bahwa variabel kualitatif mempengaruhi intersep tetapi tidak mempengaruhi koefisien kemiringan dari berbagai regresi sub-kelompok. Kita dapat menghindari asumsi ini dengan menggunakan teknik “pooling” (mengelompokkan), dengan dasar pemikiran sbb: (catatan: banyak teknik yang lain, tetapi ini salah satunya)

Misalnya, dari model regresi terdahulu:

$$Y = b_0 + b_1 D_s + b_2 P_d + e_i$$

dimana :  $D_s$  = dummy jenis kelamin,  $P_d$  = pendidikan

Kita dapat merubah model ini dengan meregresi secara terpisah antara pekerja perempuan dan laki-laki sebagai berikut:

- ✓ Regresi untuk pekerja perempuan

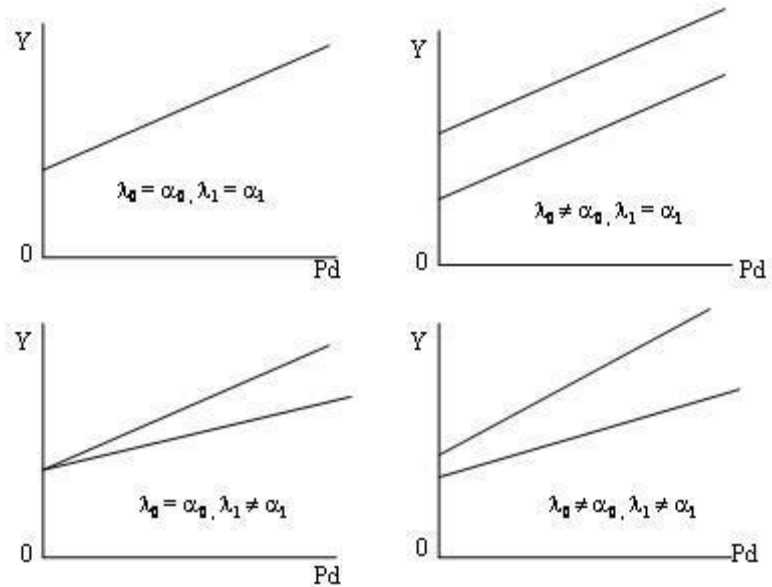
$$Y = l_0 + l_1 P_d + e_i$$

- ✓ Regresi untuk pekerja laki-laki

$$Y = a_0 + a_1 P_d + e_i$$

(Catatan: banyaknya sampel untuk laki-laki dan perempuan tidak perlu sama)

Dua persamaan regresi tersebut, memberikan empat kemungkinan hasil sebagai berikut:



Semua kemungkinan tsb dapat diuji jika kita mengelompokkan (pool) semua sampel laki-laki dan perempuan, dengan memodifikasi model menjadi :

$$Y = b_0 + b_1 D_s + b_2 Pd + b_3 (D_s.Pd) + e_i$$

Dibandingkan model sebelumnya, model ini mempunyai variabel tambahan  $D_s Pd$

Implikasi model ini sebagai berikut:

✓ Jika  $D_s=0$  (perempuan), maka model menjadi :

$$Y = b_0 + b_2 Pd + e_i, \text{ setara dgn } Y = l_0 + l_1 Pd + e_i$$

✓ Jika  $D_s=1$  (laki-laki), maka model menjadi :

$$Y = (b_0 + b_1) + (b_2 + b_3)Pd + e_i, \text{ setara dgn } Y = a_0 + a_1 Pd + e_i$$

Dari data yg telah diberikan, kita meregresikan secara terpisah untuk pekerja perempuan dan laki-laki didapatkan :

✓ Regresi pekerja perempuan :  $Y = 12971 + 50793 Pd$

✓ Regresi pekerja laki-laki :  $Y = -1242495 + 172830 Pd$

Dengan teknik “pooling” kita modifikasi model tersebut, dan melalui perhitungan, persamaan regresinya menjadi :

$$Y = 12971 - 1255466 D_s + 50793 Pd + 122036 D_s Pd$$

Jika  $D_s=0$  (perempuan), persamaan regresi tsb menjadi:

$$Y = 12971 - 1255466 (0) + 50793 Pd + 122036 (0)Pd$$

$$Y = 12971 + 50793 Pd$$

Jika  $D_s=1$  (laki-laki), persamaan regresi tsb menjadi :

$$Y = 12971 - 1255466 (1) + 50793 Pd + 122036 (1)Pd$$

$$Y = 12971 - 1255466 + 50793 Pd + 122036 (1)Pd$$

$$Y = -1242495 + 172830 Pd$$

## 2. Regresi untuk Uji Non Parametrik

Jika asumsi uji parametrik tidak terpenuhi

### a) Regresi Logistik

Regresi logistik adalah sebuah pendekatan untuk membuat model prediksi seperti halnya regresi linear atau yang biasa disebut dengan istilah *Ordinary Least Squares (OLS) regression*. Perbedaannya adalah pada regresi logistik, peneliti memprediksi variabel terikat yang berskala dikotomi. Skala dikotomi yang dimaksud adalah skala data nominal dengan dua kategori, misalnya: Ya dan Tidak, Baik dan Buruk atau Tinggi dan Rendah.

Regresi logistik (kadang disebut model logistik atau model logit), dalam statistika digunakan untuk prediksi probabilitas kejadian suatu peristiwa dengan mencocokkan data pada fungsi logit kurva logistik. Metode ini merupakan model linier umum yang digunakan untuk regresi binomial. Seperti analisis regresi pada umumnya, metode ini menggunakan beberapa variabel prediktor, baik numerik maupun kategori. Misalnya, probabilitas bahwa orang yang menderita serangan jantung pada waktu tertentu dapat diprediksi dari informasi usia, jenis kelamin, dan indeks massa tubuh. Regresi logistik juga digunakan secara luas pada bidang kedokteran dan ilmu sosial, maupun pemasaran seperti prediksi kecenderungan pelanggan untuk membeli suatu produk atau berhenti berlangganan.

Apabila pada OLS mewajibkan syarat atau asumsi bahwa error varians (residual) terdistribusi secara normal. Sebaliknya, pada regresi ini tidak dibutuhkan asumsi tersebut sebab pada regresi jenis logistik ini mengikuti distribusi logistik.

#### 1) Asumsi Regresi Logistik

Asumsi Regresi Logistik antara lain:

- Regresi logistik tidak membutuhkan hubungan linier antara variabel independen dengan variabel dependen.
- Variabel independen tidak memerlukan asumsi *multivariate normality*.
- Asumsi homokedastisitas tidak diperlukan
- Variabel bebas tidak perlu diubah ke dalam bentuk metrik (interval atau skala ratio).
- Variabel dependen harus bersifat dikotomi (2 kategori, misal: tinggi dan rendah atau baik dan buruk)
- Variabel independen tidak harus memiliki keragaman yang sama antar kelompok variabel
- Kategori dalam variabel independen harus terpisah satu sama lain atau bersifat eksklusif
- Sampel yang diperlukan dalam jumlah relatif besar, minimum dibutuhkan hingga 50 sampel data untuk sebuah variabel prediktor (independen).
- Dapat menyeleksi hubungan karena menggunakan pendekatan non linier log transformasi untuk memprediksi odds ratio. Odd dalam regresi logistik sering dinyatakan sebagai probabilitas.

## 2) Model Persamaan Regresi Logistik

Model persamaan aljabar layaknya OLS yang biasa kita gunakan adalah berikut:  $Y = B_0 + B_1X + e$ . Dimana  $e$  adalah error varians atau residual. Dengan model regresi ini, tidak menggunakan interpretasi yang sama seperti halnya persamaan regresi OLS. Model Persamaan yang terbentuk berbeda dengan persamaan OLS.

Berikut persamaannya:

$$\ln\left(\frac{\hat{p}}{1-\hat{p}}\right) = B_0 + B_1X$$

- **Persamaan Regresi Logistik**

Ln: Logaritma Natural. Di mana:

$B_0 + B_1X$ : Persamaan yang biasa dikenal dalam OLS.

Sedangkan P Aksien adalah probabilitas logistik yang didapat rumus sebagai berikut:

$$\hat{p} = \frac{\exp(B_0 + B_1 X)}{1 + \exp(B_0 + B_1 X)} = \frac{e^{B_0 + B_1 x}}{1 + e^{B_0 + B_1 x}}$$

- **Probabilitas Regresi Logistik**

Di mana:

exp atau ditulis “e” adalah fungsi exponen.

(Perlu diingat bahwa exponen merupakan kebalikan dari logaritma natural. Sedangkan logaritma natural adalah bentuk logaritma namun dengan nilai konstanta 2,71828182845904 atau biasa dibulatkan menjadi 2,72).

Dengan model persamaan di atas, tentunya akan sangat sulit untuk menginterpretasikan koefisien regresinya. Oleh karena itu maka diperkenalkanlah istilah Odds Ratio atau yang biasa disingkat Exp(B) atau OR. Exp(B) merupakan exponen dari koefisien regresi. Jadi misalkan nilai slope dari regresi adalah sebesar 0,80, maka Exp(B) dapat diperkirakan sebagai berikut:

$$2,72^{0,8} = 2,23$$

- **Nilai Odds Ratio**

Besarnya nilai Exp(B) dapat diartikan sebagai berikut:

Misalnya nilai Exp (B) pengaruh rokok terhadap terhadap kanker paru adalah sebesar 2,23, maka disimpulkan bahwa orang yang merokok lebih beresiko untuk mengalami kanker paru dibandingkan dengan orang yang tidak merokok. Interpretasi ini diartikan apabila pengkodean kategori pada tiap variabel sebagai berikut:

- 1) Variabel bebas adalah Rokok: Kode 0 untuk tidak merokok, kode 1 untuk merokok.

- 2) Variabel terikat adalah kanker Paru: Kode 0 untuk tidak mengalami kanker paru, kode 1 untuk mengalami kanker paru.

### 3) Pseudo R Square

Perbedaan lainnya yaitu pada regresi ini tidak ada nilai "R Square" untuk mengukur besarnya pengaruh simultan beberapa variabel bebas terhadap variabel terikat. Dalam regresi logistik dikenal istilah *Pseudo R Square*, yaitu nilai R Square Semu yang maksudnya sama atau identik dengan R Square pada OLS.

Jika pada OLS menggunakan uji F ANOVA untuk mengukur tingkat signifikansi dan seberapa baik model persamaan yang terbentuk, maka pada regresi ini menggunakan Nilai Chi-Square. Perhitungan nilai Chi-Square ini berdasarkan perhitungan Maximum Likelihood.

### b) Regresi Ordinal

Dalam statistik, regresi ordinal (juga disebut "klasifikasi ordinal") adalah jenis analisis regresi yang digunakan untuk memprediksi variabel ordinal, yaitu variabel yang nilainya ada pada skala arbitrer di mana hanya urutan relatif antara nilai yang berbeda yang signifikan. Ini dapat dianggap sebagai masalah antara antara regresi dan klasifikasi. Contoh regresi ordinal adalah logit terurut dan probit berurutan. Regresi ordinal sering muncul dalam ilmu-ilmu sosial, misalnya dalam pemodelan tingkat preferensi manusia (pada skala dari, katakanlah, 1-5 untuk "sangat miskin" melalui "sangat baik"), serta dalam pencarian informasi.

Dalam pembelajaran mesin, regresi ordinal juga bisa disebut pembelajaran peringkat. Regresi ordinal adalah salah satu dari berbagai jenis analisis yang khusus digunakan jika variabel dependen adalah data berskala kategorik bertingkat. Istilah kategori bertingkat juga biasa disebut dengan istilah ordinal atau ranking.

Sebelum anda memahami secara detail dan menyeluruh perihal jenis regresi ini termasuk syarat, fungsi ataupun kegunaannya, maka ada baiknya anda terlebih dahulu memahami apa itu arti kata ordinal. Ordinal atau ranking adalah salah satu jenis skala data yang mempunyai ciri kategorik namun tiap

kategorik yang ada terdapat perbedaan derajat, dimana ada yang lebih baik atau buruk dan tinggi atau rendah. Contohnya adalah tingkat pengetahuan seseorang, dimana ada kategorik tingkat pengetahuan rendah, pengetahuan sedang dan pengetahuan tinggi. Contoh lain adalah pecandu alkohol berat, sedang dan ringan.

Regresi ordinal dapat dilakukan dengan menggunakan model linear umum (GLM) yang sesuai dengan vektor koefisien dan seperangkat ambang batas untuk dataset.

Misalkan seseorang memiliki satu set pengamatan, diwakili oleh vektor  $p$ -panjang  $x_1$  hingga  $x_n$ , dengan respons terkait  $y_1$  hingga  $y_n$ , di mana setiap  $y_i$  adalah variabel ordinal pada skala  $1, \dots, K$ . Untuk kesederhanaan, dan tanpa kehilangan keumuman, kami menganggap  $y$  adalah vektor yang tidak berkurang. Untuk data ini, satu cocok dengan vektor koefisien panjang- $p$   $w$  dan satu set ambang batas  $\theta_1, \dots, \theta_{K-1}$  dengan properti yang  $\theta_1 < \theta_2 < \dots < \theta_{K-1}$ . Seperangkat ambang ini membagi garis bilangan real ke dalam segmen  $K$  disjoint, sesuai dengan tingkat respons  $K$ . Model sekarang dapat dirumuskan sebagai

$$\Pr(y \leq i | \mathbf{x}) = \sigma(\theta_i - \mathbf{w} \cdot \mathbf{x})$$

atau, probabilitas kumulatif dari respons  $y$  yang paling banyak diberikan oleh fungsi  $\sigma$  (fungsi tautan terbalik) yang diterapkan pada fungsi linier  $x$ . Ada beberapa pilihan untuk  $\sigma$ ; fungsi logistik

$$\sigma(\theta_i - \mathbf{w} \cdot \mathbf{x}) = \frac{1}{1 + e^{-(\theta_i - \mathbf{w} \cdot \mathbf{x})}}$$

memberikan model logit berurutan, sementara menggunakan fungsi probit memberikan model probit berurutan. Opsi ketiga adalah menggunakan fungsi eksponensial

$$\sigma(\theta_i - \mathbf{w} \cdot \mathbf{x}) = \exp(-\exp(\theta_i - \mathbf{w} \cdot \mathbf{x}))$$

yang memberikan model bahaya proporsional.

model variabel laten

Versi probit dari model di atas dapat dibenarkan dengan mengasumsikan keberadaan variabel laten bernilai nyata (kuantitas yang tidak teramati)  $y^*$ , ditentukan oleh

$$y^* = \mathbf{w} \cdot \mathbf{x} + \varepsilon$$

di mana  $\varepsilon$  terdistribusi normal dengan nol mean dan varians unit, dikondisikan pada  $\mathbf{x}$ . Variabel respon  $y$  dihasilkan dari "pengukuran tidak lengkap" dari  $y^*$ , di mana seseorang hanya menentukan interval ke mana  $y^*$  jatuh:

$$y = \begin{cases} 1 & \text{if } y^* \leq \theta_1, \\ 2 & \text{if } \theta_1 < y^* \leq \theta_2, \\ 3 & \text{if } \theta_2 < y^* \leq \theta_3 \\ \vdots & \\ K & \text{if } \theta_{K-1} < y^*. \end{cases}$$

Mendefinisikan  $\theta_0 = -\infty$  dan  $\theta_K = \infty$ , di atas dapat diringkas sebagai  $y = k$  jika dan hanya jika  $\theta_{k-1} < y^* \leq \theta_k$ .

- Dari asumsi-asumsi ini, seseorang dapat memperoleh distribusi bersyarat dari  $y$  sebagai

$$\begin{aligned} P(y = k | \mathbf{x}) &= P(\theta_{k-1} < y^* \leq \theta_k | \mathbf{x}) \\ &= P(\theta_{k-1} < \mathbf{w} \cdot \mathbf{x} + \varepsilon \leq \theta_k) \\ &= \Phi(\theta_k - \mathbf{w} \cdot \mathbf{x}) - \Phi(\theta_{k-1} - \mathbf{w} \cdot \mathbf{x}) \end{aligned}$$

di mana  $\Phi$  adalah fungsi distribusi kumulatif dari distribusi normal standar, dan mengambil peran fungsi tautan terbalik  $\sigma$ . Log-kemungkinan model untuk contoh pelatihan tunggal  $\mathbf{x}_i, y_i$  sekarang dapat dinyatakan sebagai

$$\log \mathcal{L}(\mathbf{w}, \theta | \mathbf{x}_i, y_i) = \sum_{k=1}^K [y_i = k] \log[\Phi(\theta_k - \mathbf{w} \cdot \mathbf{x}_i) - \Phi(\theta_{k-1} - \mathbf{w} \cdot \mathbf{x}_i)]$$

(menggunakan braket Iverson  $[y_i = k]$ .) Kemungkinan log dari model logit yang diurutkan adalah analog, menggunakan fungsi logistik alih-alih  $\Phi$ . [6]



## 1) Skala Data Regresi Ordinal

Selanjutnya kita bahas kembali perihal uji regresi ordinal. Regresi ordinal mensyaratkan skala data variabel terikat adalah ordinal dan skala data variabel bebas boleh kategorik ataupun kuantitatif (numerik). Variabel bebas kategorik dalam regresi ordinal disebut dengan istilah Factor. Sedangkan variabel bebas numerik disebut dengan istilah covariate.

## 2) Prinsip Regresi Ordinal

Jika skala data sudah sesuai dengan petunjuk di atas, selanjutnya kita bahas perihal prinsip regresi ordinal. Prinsip yang digunakan dalam regresi ordinal adalah prinsip general least square. Prinsip general least square yaitu metode pemodelan yang memprediksikan probabilitas kumulatif dari tiap kategori yang ada.

- Contoh probabilitas kumulatif yaitu:

Sebagai seseorang dengan pengetahuan rendah, kemungkinan atau probabilitasnya adalah 0,5. Sebagai seseorang dengan pengetahuan sedang, probabilitasnya adalah 0,3. Sebagai seseorang dengan pengetahuan tinggi, probabilitasnya adalah 0,2.

Sehingga dari contoh di atas, maka nilai probabilitas kumulatif jika dihitung adalah: Sebagai seseorang dengan pengetahuan rendah, probabilitas kumulatifnya adalah 0,5. Sebagai seseorang dengan pengetahuan sedang, probabilitas kumulatifnya adalah  $0,5 + 0,3 = 0,8$ . Sebagai seseorang dengan pengetahuan tinggi, probabilitas kumulatifnya adalah  $0,5 + 0,3 + 0,2 = 1$ .

Perlu anda perhatikan baik-baik, bahwa probabilitas kumulatif yang paling akhir haruslah menunjukkan angka 1 atau 100%.

Berdasarkan penjelasan di atas, dapat kita pahami bersama bahwa model regresi ordinal dapat memberikan penjelasan kepada kita dalam bentuk sebuah persamaan yang menjelaskan prediksi seseorang akan berpengetahuan rendah, sedang dan berat. Dimana berpengetahuan rendah, sedang dan berat adalah sebuah kesatuan yang utuh dimana nilai totalnya pastilah 1 (satu).

Sehingga dengan penjelasan di atas, kita pahami pula bahwa regresi ini tidak membentuk model persamaan yang dapat memprediksi seberapa besarnya seseorang dapat berpengetahuan rendah, atau sedang atau ringan.

### 3) Jenis Regresi Ordinal

Regresi ordinal mempunyai beberapa teknik perhitungan yang mana pada SPSS disebut dengan istilah Option Link. Jenis tersebut antara lain:

- Logit dengan persamaan:  $f(x) = \log(x/(1-x))$   
Jenis ini digunakan pada sebagian besar distribusi data. Maka aplikasi SPSS secara default atau bawaan aslinya menggunakan option link jenis Logit ini,
- Negative Log-log dengan persamaan  $f(x) = -\log(-\log(x))$   
Jenis ini digunakan apabila data mempunyai kecenderungan bernilai rendah,
- Complementary Log-log dengan persamaan  $f(x) = \log(-\log(1-x))$   
Jenis ini digunakan apabila data mempunyai kecenderungan bernilai tinggi,
- Cauchit (Inverse Cauchy) dengan persamaan  $f(x) = \tan(\Phi(x-0,5))$   
Jenis ini digunakan apabila variabel latent mempunyai nilai yang ekstrim,
- Probit dengan persamaan  $f(x) = O-1(x)$  dengan O-1 adalah fungsi inverse distribusi kumulatif standar normal. Jenis ini digunakan apabila variabel latent terdistribusi secara normal

Semua jenis atau option link di atas tentulah dipilih dengan menyesuaikan dengan distribusi data yang ada.

#### Contoh Model Regresi Ordinal

Setelah memahami uraian perihal model regresi jenis ordinal di atas, harapannya pembaca dapat memanfaatkannya dalam penelitian jika memang ingin membentuk model regresi dengan variabel terikat berskala data ordinal.

Contoh sebuah judul penelitian yang mana analisis statistiknya menggunakan regresi ordinal adalah: “Pengaruh Tingkat Pendidikan

terhadap Tingkat Pengetahuan.” Dimana Pengaruh Tingkat Pendidikan sebagai Factor adalah variabel independen dan Tingkat Pengetahuan sebagai variabel dependen. Tingkat pendidikan adalah data kategorik, yaitu: tidak sekolah, SD, SMP, SMA dan Perguruan Tinggi. Sedangkan tingkat pengetahuan juga kategorik bertingkat, yaitu pengetahuan rendah, sedang dan berat.

#### **4) Perbedaan Regresi Ordinal dengan Model Regresi Lainnya**

Regresi data bertingkat atau ordinal menggunakan skala data ordinal pada variabel terikat. Sedangkan regresi linear menggunakan skala data variabel terikat adalah data numerik (interval/rasio). Regresi logistik menggunakan skala data variabel terikat adalah nominal dikotomi (nominal 2 kategori). Regresi multinomial skala data variabel terikat adalah nominal lebih dari 2 kategori.

Pada regresi linear membutuhkan beberapa uji asumsi klasik yaitu uji normalitas, uji heteroskedastisitas, uji multikolinearitas, uji autokorelasi, uji linearitas, uji outlier baik univariat maupun multivariat. Sedangkan pada uji regresi ordinal, tidak dibutuhkan adanya syarat atau asumsi klasik. Sehingga uji-uji syarat di atas tidak perlu dilakukan dan juga tidak relevan untuk dilakukan. Namun perlu kita pahami bahwa uji regresi linear karena menggunakan skala data interval atau rasio dan termasuk dalam jenis uji parametris, maka informasi yang diberikan jauh lebih banyak.

#### **5) Kesimpulan Analisis Regresi Ordinal**

Setelah kita berdiskusi dalam artikel ini, mari kita coba membuat kesimpulan bersama-sama tentang analisis regresi ini. Kesimpulan regresi ini berdasarkan statistikian adalah:

- Salah satu jenis model regresi (model prediksi),
- Menggunakan prinsip general least square,
- membentuk prediksi probabilitas kumulatif,
- Jumlah probabilitas kumulatif adalah 1 (satu) atau seratus persen (100%),
- Skala data variabel terikat harus ordinal,
- Skala data variabel bebas boleh kategorik ataupun numerik.

## D. PROSEDUR KERJA DAN INTERPRETASI OUTPUT DARI SPSS

### 1. Regresi untuk Uji Parametrik

Asumsi yang digunakan sama dengan uji Korelasi untuk uji parametrik

#### a) Regresi Linier

##### Contoh kasus:

Seorang mahasiswa bernama Hermawan ingin meneliti tentang pengaruh biaya promosi terhadap volume penjualan pada perusahaan jual beli motor. Dengan ini di dapat variabel dependen (Y) adalah volume penjualan dan variabel independen (X) adalah biaya promosi. Dengan ini Hermawan menganalisis dengan bantuan program SPSS dengan alat analisis regresi linear sederhana. Data-data yang di dapat ditabulasikan sebagai berikut:

Tabel. Tabulasi Data Penelitian (Data Fiktif)

No	Biaya Promosi	Volume Penjualan
1	12,000	56,000
2	13,500	62,430
3	12,750	60,850
4	12,600	61,300
5	14,850	65,825
6	15,200	66,354
7	15,750	65,260
8	16,800	68,798
9	18,450	70,470
10	17,900	65,200
11	18,250	68,000
12	16,480	64,200
13	17,500	65,300
14	19,560	69,562
15	19,000	68,750
16	20,450	70,256

17	22,650	72,351
18	21,400	70,287
19	22,900	73,564
20	23,500	75,642

- Masuk program SPSS
- Klik variable view pada SPSS data editor
- Pada kolom Name ketik y, kolom Name pada baris kedua ketik x.
- Pada kolom Label, untuk kolom pada baris pertama ketik Volume Penjualan, untuk kolom pada baris kedua ketik Biaya Promosi.
- Untuk kolom-kolom lainnya boleh dihiraukan (isian default)
- Buka data view pada SPSS , maka didapat kolom variabel y dan x.
- Ketikkan data sesuai dengan variabelnya
- Klik Analyze - Regression - Linear
- Klik variabel Volume Penjualan dan masukkan ke kotak Dependent, kemudian klik variabel Biaya Promosi dan masukkan ke kotak Independent.
- Klik Statistics, klik Casewise diagnostics, klik All cases. Klik Continue
- Klik OK, maka hasil output yang didapat pada kolom Coefficients dan Casewise Diagnostics adalah sebagai berikut:
- **Output**

Tabel. Hasil Analisis Regresi Linear Sederhana

**Coefficients<sup>a</sup>**

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-28764.7	4229.248		-6.801	.000
	Biaya Promosi	.691	.063	.933	10.983	.000

a. Dependent Variable: Volume Penjualan

**Casewise Diagnostics<sup>a</sup>**

Case Number	Std. Residual	Volume Penjualan	Predicted Value	Residual
1	1.576	12000.00	9955.0474	2044.9526
2	-.694	13500.00	14400.901	-900.9006
3	-.430	12750.00	13308.451	-558.4514
4	-.786	12600.00	13619.592	-1019.59
5	-1.463	14850.00	16748.283	-1898.28
6	-1.475	15200.00	17114.047	-1914.05
7	-.468	15750.00	16357.629	-607.6292
8	-1.545	16800.00	18803.886	-2003.89
9	-1.164	18450.00	19959.946	-1509.95
10	1.221	17900.00	16316.144	1583.8562
11	-.002	18250.00	18252.130	-2.1296
12	.659	16480.00	15624.720	855.2798
13	.859	17500.00	16385.286	1114.7139
14	.176	19560.00	19332.133	227.8668
15	.177	19000.00	18770.697	229.3027
16	.492	20450.00	19811.981	638.0189
17	1.071	22650.00	21260.513	1389.4866
18	1.207	21400.00	19833.415	1566.5848
19	.617	22900.00	22099.210	800.7899
20	-.028	23500.00	23535.988	-35.9882

a. Dependent Variable: Volume Penjualan

- Persamaan regresinya sebagai berikut:

$$Y' = a + bX$$

$$Y' = -28764,7 + 0,691X$$

Angka-angka ini dapat diartikan sebagai berikut:

- Konstanta sebesar -28764,7; artinya jika biaya promosi (X) nilainya adalah 0, maka volume penjualan (Y') nilainya negatif yaitu sebesar -28764,7.
- Koefisien regresi variabel harga (X) sebesar 0,691; artinya jika harga mengalami kenaikan Rp.1, maka volume penjualan (Y') akan mengalami peningkatan sebesar Rp.0,691. Koefisien bernilai positif artinya terjadi hubungan positif antara harga dengan volume penjualan, semakin naik harga maka semakin meningkatkan volume penjualan.
- Nilai volume penjualan yang diprediksi (Y') dapat dilihat pada tabel Casewise Diagnostics (kolom Predicted Value). Sedangkan Residual (unstandardized residual) adalah selisih antara Volume Penjualan dengan Predicted Value, dan Std. Residual (standardized residual)

adalah nilai residual yang telah terstandarisasi (nilai semakin mendekati 0 maka model regresi semakin baik dalam melakukan prediksi, sebaliknya semakin menjauhi 0 atau lebih dari 1 atau -1 maka semakin tidak baik model regresi dalam melakukan prediksi).

- Uji Koefisien Regresi Sederhana (Uji t)

Uji ini digunakan untuk mengetahui apakah variabel independen (X) berpengaruh secara signifikan terhadap variabel dependen (Y). Signifikan berarti pengaruh yang terjadi dapat berlaku untuk populasi (dapat digeneralisasikan).

Dari hasil analisis regresi di atas dapat diketahui nilai t hitung seperti pada tabel 2. Langkah-langkah pengujian sebagai berikut:

- Menentukan Hipotesis

Ho : Ada pengaruh secara signifikan antara biaya promosi dengan volume penjualan

Ha : Tidak ada pengaruh secara signifikan antara biaya promosi dengan volume penjualan

- Menentukan tingkat signifikansi

Tingkat signifikansi menggunakan  $\alpha = 5\%$  (signifikansi 5% atau 0,05 adalah ukuran standar yang sering digunakan dalam penelitian)

- Menentukan t hitung

Berdasarkan tabel diperoleh t hitung sebesar 10,983

- Menentukan t tabel

Tabel distribusi t dicari pada  $\alpha = 5\% : 2 = 2,5\%$  (uji 2 sisi) dengan derajat kebebasan (df)  $n-k-1$  atau  $20-2-1 = 17$  (n adalah jumlah kasus dan k adalah jumlah variabel independen). Dengan pengujian 2 sisi (signifikansi = 0,025) hasil diperoleh untuk t tabel sebesar 2,110 (Lihat pada lampiran) atau dapat dicari di Ms Excel dengan cara pada cell kosong ketik =tinv(0.05,17) lalu enter.

- Kriteria Pengujian

Ho diterima jika  $-t \text{ tabel} < t \text{ hitung} < t \text{ tabel}$

Ho ditolak jika  $-t \text{ hitung} < -t \text{ tabel}$  atau  $t \text{ hitung} > t \text{ tabel}$

- Membandingkan t hitung dengan t tabel

Nilai t hitung  $> t \text{ tabel}$  ( $10,983 > 2,110$ ) maka Ho ditolak.

- **Kesimpulan**

Oleh karena nilai  $t$  hitung  $>$   $t$  tabel ( $10,983 > 2,110$ ) maka  $H_0$  ditolak, artinya bahwa ada pengaruh secara signifikan antara biaya promosi dengan volume penjualan. Jadi dalam kasus ini dapat disimpulkan bahwa biaya promosi berpengaruh terhadap volume penjualan pada perusahaan jual beli motor.

## b) Regresi Linier Berganda

### Contoh kasus:

Kita mengambil contoh kasus pada uji normalitas, yaitu sebagai berikut: Seorang mahasiswa bernama Bambang melakukan penelitian tentang faktor-faktor yang mempengaruhi harga saham pada perusahaan di BEJ. Bambang dalam penelitiannya ingin mengetahui hubungan antara rasio keuangan PER dan ROI terhadap harga saham. Dengan ini Bambang menganalisis dengan bantuan program SPSS dengan alat analisis regresi linear berganda. Dari uraian di atas maka didapat variabel dependen (Y) adalah harga saham, sedangkan variabel independen ( $X_1$  dan  $X_2$ ) adalah PER dan ROI.

Data-data yang di dapat berupa data rasio dan ditabulasikan sebagai berikut:

Tabel. Tabulasi Data (Data Fiktif)

Tahun	Harga Saham (Rp)	PER (%)	ROI (%)
1990	8300	4.90	6.47
1991	7500	3.28	3.14
1992	8950	5.05	5.00
1993	8250	4.00	4.75
1994	9000	5.97	6.23
1995	8750	4.24	6.03
1996	10000	8.00	8.75
1997	8200	7.45	7.72
1998	8300	7.47	8.00



1999	10900	12.68	10.40
2000	12800	14.45	12.42
2001	9450	10.50	8.62
2002	13000	17.24	12.07
2003	8000	15.56	5.83
2004	6500	10.85	5.20
2005	9000	16.56	8.53
2006	7600	13.24	7.37
2007	10200	16.98	9.38

- Masuk program SPSS
- Klik variable view pada SPSS data editor
- Pada kolom Name ketik y, kolom Name pada baris kedua ketik x1, kemudian untuk baris kedua ketik x2.
- Pada kolom Label, untuk kolom pada baris pertama ketik Harga Saham, untuk kolom pada baris kedua ketik PER, kemudian pada baris ketiga ketik ROI.
- Untuk kolom-kolom lainnya boleh dihiraukan (isian default)
- Buka data view pada SPSS data editor, maka didapat kolom variabel y, x1, dan x2.
- Ketikkan data sesuai dengan variabelnya
- Klik Analyze - Regression - Linear
- Klik variabel Harga Saham dan masukkan ke kotak Dependent, kemudian klik variabel PER dan ROI kemudian masukkan ke kotak Independent.
- Klik Statistics, klik Casewise diagnostics, klik All cases. Klik Continue
- Klik OK, maka hasil output yang didapat pada kolom Coefficients dan Casewise diagnostics adalah sebagai berikut:
- **Output**  
Tabel. Hasil Analisis Regresi Linear Berganda

Coefficients<sup>a</sup>

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	4662.491	668.382		6.976	.000
	PER (X1)	-74.482	59.161	-.214	-1.259	.227
	ROI (X2)	692.107	116.049	1.012	5.964	.000

a. Dependent Variable: Harga Saham (Y)

Casewise Diagnostics<sup>a</sup>

Case Number	Std. Residual	Harga Saham	Predicted Value	Residual
1	-.546	8300.00	8775.4598	-475.4598
2	1.043	7500.00	6591.4051	908.5949
3	1.382	8950.00	7746.8907	1203.1093
4	.687	8250.00	7652.0699	597.9301
5	.540	9000.00	8529.6586	470.3414
6	.264	8750.00	8520.0908	229.9092
7	-.141	10000.00	10122.569	-122.5693
8	-1.436	8200.00	9450.6644	-1250.66
9	-1.542	8300.00	9642.9647	-1342.96
10	-.018	10900.00	10915.970	-15.9703
11	.709	12800.00	12182.193	617.8070
12	-.455	9450.00	9846.3909	-396.3909
13	1.456	13000.00	11732.151	1267.8487
14	.530	8000.00	7538.5352	461.4648
15	-1.095	6500.00	7453.3174	-953.3174
16	-.382	9000.00	9332.7414	-332.7414
17	-1.352	7600.00	8777.1773	-1177.18
18	.356	10200.00	9889.7497	310.2503

a. Dependent Variable: Harga Saham

- Persamaan regresinya sebagai berikut:

$$Y' = a + b_1X_1 + b_2X_2$$

$$Y' = 4662,491 + (-74,482)X_1 + 692,107X_2$$

$$Y' = 4662,491 - 74,482X_1 + 692,107X_2$$

Keterangan:

Y' = Harga saham yang diprediksi (Rp)

a = konstanta

b<sub>1</sub>, b<sub>2</sub> = koefisien regresiX<sub>1</sub> = PER (%)X<sub>2</sub> = ROI (%)

- Persamaan regresi di atas dapat dijelaskan sebagai berikut:
  - Konstanta sebesar 4662,491; artinya jika PER ( $X_1$ ) dan ROI ( $X_2$ ) nilainya adalah 0, maka harga saham ( $Y'$ ) nilainya adalah Rp.4662,491.
  - Koefisien regresi variabel PER ( $X_1$ ) sebesar -74,482; artinya jika variabel independen lain nilainya tetap dan PER mengalami kenaikan 1%, maka harga saham ( $Y'$ ) akan mengalami penurunan sebesar Rp.74,482. Koefisien bernilai negatif artinya terjadi hubungan negatif antara PER dengan harga saham, semakin naik PER maka semakin turun harga saham.
  - Koefisien regresi variabel ROI ( $X_2$ ) sebesar 692,107; artinya jika variabel independen lain nilainya tetap dan ROI mengalami kenaikan 1%, maka harga saham ( $Y'$ ) akan mengalami peningkatan sebesar Rp.692,107. Koefisien bernilai positif artinya terjadi hubungan positif antara ROI dengan harga saham, semakin naik ROI maka semakin meningkat harga saham.

Nilai harga saham yang diprediksi ( $Y'$ ) dapat dilihat pada tabel Casewise Diagnostics (kolom Predicted Value). Sedangkan Residual (*unstandardized residual*) adalah selisih antara harga saham dengan Predicted Value, dan Std. Residual (*standardized residual*) adalah nilai residual yang telah terstandarisasi (nilai semakin mendekati 0 maka model regresi semakin baik dalam melakukan prediksi, sebaliknya semakin menjauhi 0 atau lebih dari 1 atau -1 maka semakin tidak baik model regresi dalam melakukan prediksi).

- **Analisis Korelasi Ganda (R)**

Analisis ini digunakan untuk mengetahui hubungan antara dua atau lebih variabel independen ( $X_1, X_2, \dots, X_n$ ) terhadap variabel dependen ( $Y$ ) secara serentak. Koefisien ini menunjukkan seberapa besar hubungan yang terjadi antara variabel independen ( $X_1, X_2, \dots, X_n$ ) secara serentak terhadap variabel dependen ( $Y$ ). nilai R

berkisar antara 0 sampai 1, nilai semakin mendekati 1 berarti hubungan yang terjadi semakin kuat, sebaliknya nilai semakin mendekati 0 maka hubungan yang terjadi semakin lemah.

Menurut Sugiyono (2007) pedoman untuk memberikan interpretasi koefisien korelasi sebagai berikut:

- 0,00 - 0,199 = sangat rendah
- 0,20 - 0,399 = rendah
- 0,40 - 0,599 = sedang
- 0,60 - 0,799 = kuat
- 0,80 - 1,000 = sangat kuat

Dari hasil analisis regresi, lihat pada output *model summary* dan disajikan sebagai berikut:

Tabel. Hasil analisis korelasi ganda

**Model Summary**

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.879 <sup>a</sup>	.772	.742	870.80

a. Predictors: (Constant), ROI (X2), PER (X1)

Berdasarkan tabel di atas diperoleh angka R sebesar 0,879. Hal ini menunjukkan bahwa terjadi hubungan yang sangat kuat antara PER dan ROI terhadap harga saham.

- **Analisis Determinasi (R<sup>2</sup>)**

Analisis determinasi dalam regresi linear berganda digunakan untuk mengetahui prosentase sumbangan pengaruh variabel independen (X<sub>1</sub>, X<sub>2</sub>,.....X<sub>n</sub>) secara serentak terhadap variabel dependen (Y). Koefisien ini menunjukkan seberapa besar prosentase variasi variabel independen yang digunakan dalam model mampu menjelaskan variasi variabel dependen. R<sup>2</sup> sama dengan 0, maka tidak ada sedikitpun prosentase sumbangan

pengaruh yang diberikan variabel independen terhadap variabel dependen, atau variasi variabel independen yang digunakan dalam model tidak menjelaskan sedikitpun variasi variabel dependen. Sebaliknya  $R^2$  sama dengan 1, maka prosentase sumbangan pengaruh yang diberikan variabel independen terhadap variabel dependen adalah sempurna, atau variasi variabel independen yang digunakan dalam model menjelaskan 100% variasi variabel dependen.

Dari hasil analisis regresi, lihat pada output *model summary* dan disajikan sebagai berikut:

Tabel. Hasil analisis determinasi

**Model Summary**

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.879 <sup>a</sup>	.772	.742	870.80

a. Predictors: (Constant), ROI (X2), PER (X1)

Berdasarkan tabel di atas diperoleh angka  $R^2$  (*R Square*) sebesar 0,772 atau (77,2%). Hal ini menunjukkan bahwa prosentase sumbangan pengaruh variabel independen (PER dan ROI) terhadap variabel dependen (harga saham) sebesar 77,2%. Atau variasi variabel independen yang digunakan dalam model (PER dan ROI) mampu menjelaskan sebesar 77,2% variasi variabel dependen (harga saham). Sedangkan sisanya sebesar 22,8% dipengaruhi atau dijelaskan oleh variabel lain yang tidak dimasukkan dalam model penelitian ini.

Adjusted R Square adalah nilai R Square yang telah disesuaikan, nilai ini selalu lebih kecil dari R Square dan angka ini bisa memiliki harga negatif. Menurut Santoso (2001) bahwa untuk regresi dengan lebih dari dua variabel bebas digunakan Adjusted  $R^2$  sebagai koefisien determinasi.

Standard Error of the Estimate adalah suatu ukuran banyaknya kesalahan model regresi dalam memprediksikan nilai

Y. Dari hasil regresi di dapat nilai 870,80 atau Rp.870,80 (satuan harga saham), hal ini berarti banyaknya kesalahan dalam prediksi harga saham sebesar Rp.870,80. Sebagai pedoman jika Standard error of the estimate kurang dari standar deviasi Y, maka model regresi semakin baik dalam memprediksi nilai Y.

- **Uji Koefisien Regresi Secara Bersama-sama (Uji F)**

Uji ini digunakan untuk mengetahui apakah variabel independen ( $X_1, X_2, \dots, X_n$ ) secara bersama-sama berpengaruh secara signifikan terhadap variabel dependen (Y). Atau untuk mengetahui apakah model regresi dapat digunakan untuk memprediksi variabel dependen atau tidak. Signifikan berarti hubungan yang terjadi dapat berlaku untuk populasi (dapat digeneralisasikan), misalnya dari kasus di atas populasinya adalah 50 perusahaan dan sampel yang diambil dari kasus di atas 18 perusahaan, jadi apakah pengaruh yang terjadi atau kesimpulan yang didapat berlaku untuk populasi yang berjumlah 50 perusahaan.

Dari hasil output analisis regresi dapat diketahui nilai F seperti pada tabel 2 berikut ini.

Tabel. Hasil Uji F

**ANOVA<sup>b</sup>**

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	38620594	2	19310297.00	25.465	.000 <sup>a</sup>
	Residual	11374406	15	758293.733		
	Total	49995000	17			

a. Predictors: (Constant), ROI (X2), PER (X1)

b. Dependent Variable: Harga Saham (Y)

Tahap-tahap untuk melakukan uji F adalah sebagai berikut:

- Merumuskan Hipotesis

Ho : Tidak ada pengaruh secara signifikan antara PER dan ROI secara bersama-sama terhadap harga saham.

Ha : Ada pengaruh secara signifikan antara PER dan ROI secara bersama-sama terhadap harga saham.

- Menentukan tingkat signifikansi

Tingkat signifikansi menggunakan  $\alpha = 5\%$  (signifikansi 5% atau 0,05 adalah ukuran standar yang sering digunakan dalam penelitian)

- Menentukan F hitung

Berdasarkan tabel diperoleh F hitung sebesar 25,465

- Menentukan F tabel

Dengan menggunakan tingkat keyakinan 95%,  $\alpha = 5\%$ , df 1 (jumlah variabel-1) = 2, dan df 2 ( $n-k-1$ ) atau  $18-2-1 = 15$  (n adalah jumlah kasus dan k adalah jumlah variabel independen), hasil diperoleh untuk F tabel sebesar 3,683 (Lihat pada lampiran) atau dapat dicari di Ms Excel dengan cara pada cell kosong ketik =FINV(0.05,2,15) lalu enter.

- Kriteria pengujian

Ho diterima bila F hitung < F tabel

Ho ditolak bila F hitung > F tabel

- Membandingkan F hitung dengan F tabel

Nilai F hitung > F tabel ( $25,465 > 3,683$ ), maka Ho ditolak.

- **Kesimpulan**

Karena F hitung > F tabel ( $25,465 > 3,683$ ), maka Ho ditolak, artinya ada pengaruh secara signifikan antara *price earning ratio* (PER) dan *return on investmen* (ROI) secara bersama-sama terhadap terhadap harga saham. Jadi dari kasus ini dapat disimpulkan bahwa PER dan ROI secara bersama-sama berpengaruh terhadap harga saham pada perusahaan di BEJ.

- **Uji Koefisien Regresi Secara Parsial (Uji t)**

Uji ini digunakan untuk mengetahui apakah dalam model regresi variabel independen ( $X_1, X_2, \dots, X_n$ ) secara parsial berpengaruh signifikan terhadap variabel dependen (Y).

Dari hasil analisis regresi output dapat disajikan sebagai berikut:

Tabel. Uji t

Coefficients<sup>a</sup>

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	4662.491	668.382		6.976	.000
	PER (X1)	-74.482	59.161	-.214	-1.259	.227
	ROI (X2)	692.107	116.049	1.012	5.964	.000

a. Dependent Variable: Harga Saham (Y)

Langkah-langkah pengujian sebagai berikut:

### Pengujian koefisien regresi variabel PER

- Menentukan Hipotesis

Ho : Secara parsial tidak ada pengaruh signifikan antara PER dengan harga saham.

Ha : Secara parsial ada pengaruh signifikan antara PER dengan harga saham

- Menentukan tingkat signifikansi

Tingkat signifikansi menggunakan  $\alpha = 5\%$

- Menentukan t hitung

Berdasarkan tabel diperoleh t hitung sebesar -1,259

- Menentukan t tabel

Tabel distribusi t dicari pada  $\alpha = 5\% : 2 = 2,5\%$  (uji 2 sisi) dengan derajat kebebasan (df)  $n-k-1$  atau  $18-2-1 = 15$  (n adalah jumlah kasus dan k adalah jumlah variabel independen). Dengan pengujian 2 sisi (signifikansi  $= 0,025$ ) hasil diperoleh untuk t tabel sebesar 2,131 (Lihat pada lampiran) atau dapat dicari di Ms Excel dengan cara pada cell kosong ketik  $=\text{tinv}(0.05,15)$  lalu enter.

- Kriteria Pengujian

Ho diterima jika  $-t \text{ tabel} < t \text{ hitung} < t \text{ tabel}$

Ho ditolak jika  $-t \text{ hitung} < -t \text{ tabel}$  atau  $t \text{ hitung} > t \text{ tabel}$

- Membandingkan thitung dengan t tabel

Nilai -t hitung  $> -t \text{ tabel}$  ( $-1,259 > -2,131$ ) maka Ho diterima



- **Kesimpulan**

Oleh karena nilai  $-t$  hitung  $> -t$  tabel ( $-1,259 > -2,131$ ) maka  $H_0$  diterima, artinya secara parsial tidak ada pengaruh signifikan antara PER dengan harga saham. Jadi dari kasus ini dapat disimpulkan bahwa secara parsial PER tidak berpengaruh terhadap harga saham pada perusahaan di BEJ.

• **Pengujian koefisien regresi variabel ROI**

- Menentukan Hipotesis

$H_0$  : Secara parsial tidak ada pengaruh signifikan antara ROI dengan harga saham

$H_a$  : Secara parsial ada pengaruh signifikan antara ROI dengan harga saham

- Menentukan tingkat signifikansi

Tingkat signifikansi menggunakan  $\alpha = 5\%$ .

- Menentukan  $t$  hitung

Berdasarkan tabel diperoleh  $t$  hitung sebesar 5,964

- Menentukan  $t$  tabel

Tabel distribusi  $t$  dicari pada  $\alpha = 5\% : 2 = 2,5\%$  (uji 2 sisi) dengan derajat kebebasan (df)  $n-k-1$  atau  $18-2-1 = 15$  ( $n$  adalah jumlah kasus dan  $k$  adalah jumlah variabel independen). Dengan pengujian 2 sisi (signifikansi = 0,025) hasil diperoleh untuk  $t$  tabel sebesar 2,131.

- Kriteria Pengujian

$H_0$  diterima jika  $-t$  tabel  $\leq t$  hitung  $\leq t$  tabel

$H_0$  ditolak jika  $-t$  hitung  $< -t$  tabel atau  $t$  hitung  $> t$  tabel

- Membandingkan  $t$  hitung dengan  $t$  tabel

Nilai  $t$  hitung  $> t$  tabel ( $5,964 > 2,131$ ) maka  $H_0$  ditolak

- **Kesimpulan**

Oleh karena nilai  $t$  hitung  $> t$  tabel ( $5,964 > 2,131$ ) maka  $H_0$  ditolak, artinya secara parsial ada pengaruh signifikan antara ROI dengan harga saham. Jadi dari kasus ini dapat disimpulkan

bahwa secara parsial ROI berpengaruh positif terhadap harga saham pada perusahaan di BEJ.

### c) Regresi Dummy

#### Contoh Soal

Subjek	Agresivitas	Umur	Jenis kelamin	Suku
A	45	52	Laki-laki	Jawa
B	43	41	Perempuan	Sunda
C	47	40	Laki-laki	Madura
D	53	58	Perempuan	Jawa
E	60	46	Laki-laki	Madura

Pada kategori 2 kelompok, sebenarnya tidak ada aturan baku mana yang harus diberi kode 1 mana yang 0. Tapi agar memudahkan pemaknaan terhadap hasil berilah angka 1 pada kategori yang secara teoritik akan memiliki nilai prediksi lebih tinggi. Misal secara teoritik, agresivitas ini lebih mampu diprediksi oleh laki-laki, maka laki-laki kita beri kode 1. Dari data mentah kita, kita bisa ubah input data kita seperti di bawah ini.

Subjek	Agresivitas	Umur	Laki-laki	Jawa	Sunda
A	45	52	1	1	0
B	43	41	0	0	1
C	47	40	1	0	0
D	53	58	0	1	0
E	60	46	1	0	0

Subjek A adalah laki-laki Jawa, maka pada variabel laki-laki dan Jawa kita beri skor 1, sedangkan ada variabel sunda kita beri skor 0. Sedangkan subjek E adalah laki-laki madura, maka pada variabel laki-laki kita beri skor 1, sedangkan pada variabel Jawa dan Sunda kita beri skor 0 (artinya bukan keduanya).

- **Input data di SPSS**

Karena pemberian kode 1 dan 0 terlalu melelahkan jika harus dilakukan secara manual, apalagi kalau subjek kita sampai ratusan, maka ada beberapa cara yang lebih cepat untuk membuat kode. Salah satu cara yang paling umum dilakukan adalah dengan fungsi “if” di excel. Pada tulisan ini saya akan langsung memberikan cara input kode variabel dummy di SPSS.

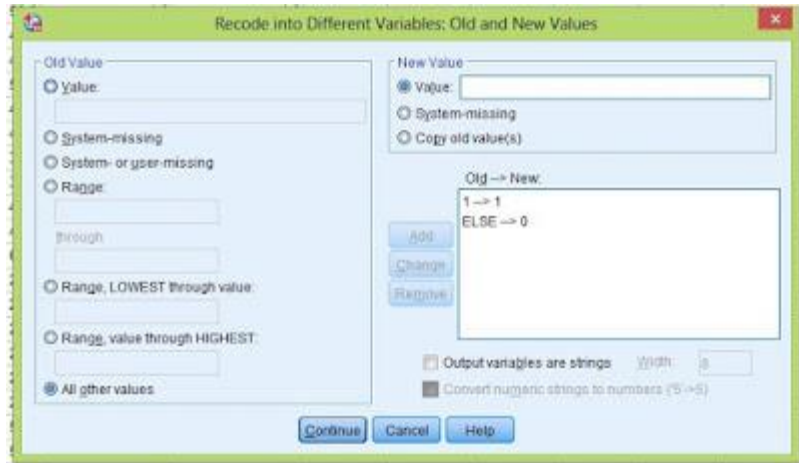
Sebelum kita membuat variabel dummy, seperti biasa, kita ubah terlebih dahulu variabel kategori kita dengan angka nominal. Misalkan dalam contoh di atas, saya ubah laki-laki=1, perempuan=2, Jawa=1, Sunda=2, Madura=3. Dengan demikian pada contoh data di atas data akan berubah seperti ini. Untuk mengubah kode ini cukup mudah, gunakan saja **find and replace** dengan menekan **ctrl+h**.

Subjek	Agresivitas	Umur	Jenis kelamin	Suku
A	45	52	1	1
B	43	41	2	2
C	47	40	1	3
D	53	58	2	1
E	60	46	1	3

Misalkan kita akan membuat variabel Dummy laki-laki. Untuk membuat kode pada variabel dummy laki-laki, ikut langkah berikut.

- Klik transform – recode into **defferent variables**
- Masukkan variabel **Jenis Kelamin**, kemudian pada **output variable name** tuliskan **laki**, lalu klik **change**. Ini artinya kita akan membuat variabel baru bernama laki dari data variabel JenisKelamin.
- Klik **old and new value**
- Pada **old value**, masukkan nilai **value** adalah 1, kemudian pada **new value** masukan nilai value adalah **1**. Kemudian tekan **Add**. Prosedur ini akan mengubah laki-laki yang sudah kita kode 1 tadi menjadi angka 1

- Kemudian pada old value klik **all other values**, dan pada new values masukkan **value** adalah **0**. Prosedur ini akan mengubah nilai yang bukan merupakan 1 (laki-laki) menjadi nilai 0.
- Klik **continue**, lalu **ok**



Jika kita kembali ke data kita, maka sudah muncul variabel baru, yakni laki. Variabel laki memiliki nilai 1 dan 0, dimana 1 menunjukkan laki-laki, dan nol menunjukkan perempuan.

- **Mulai analisis regresi dengan variabel dummy**

Untuk analisis regresi di SPSS, kita bisa menggunakan regresi berganda (*multiple regression*) atau regresi berjenjang (*hierarchical regression*). Jika kita hendak melihat peran semua variabel independen kita gunakan regresi berganda. Namun jika kita ingin melihat peran tiap variabel independen secara berurutan kita gunakan regresi berjenjang. Pada contoh kali ini kita akan menggunakan regresi berganda.

Analisis regresi berganda kali ini dilakukan untuk melihat peran umur, jenis kelamin, dan suku secara bersama-sama terhadap agresivitas. Untuk mulai analisis regresi berganda, ikuti langkah berikut.

- Klik **analyze – regression – linear**
- Masukkan variabel **agresivitas** pada kotak **dependent**
- Masukkan variabel **umur, laki, jawa, dan sunda** pada kotak **independent**
- Klik **ok**

- **Output** hasil analisis dapat dilihat pada gambar di bawah

**Model Summary**

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,511 <sup>a</sup>	,261	,232	8,5455

a. Predictors: (Constant), Sunda, Umur, laki, Jawa

**ANOVA<sup>a</sup>**

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2585,114	4	646,278	8,850	,000 <sup>b</sup>
	Residual	7302,601	100	73,026		
	Total	9887,714	104			

a. Dependent Variable: Agresifitas

b. Predictors: (Constant), Sunda, Umur, laki, Jawa

**Coefficients<sup>a</sup>**

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	53,800	3,163		17,009	,000
	Umur	-,194	,068	-,247	-2,868	,005
	laki	8,275	1,684	,426	4,912	,000
	Jawa	-1,702	2,219	-,080	-,767	,445
	Sunda	-2,147	2,035	-,109	-1,055	,294

a. Dependent Variable: Agresifitas

Terlihat pada tabel **ANOVA** bahwa secara bersama-sama semua variabel independen berperan terhadap agresivitas secara signifikan ( $F=8,850$  dan  $p<0,05$ ). Pada tabel **Model Summary**, diketahui bahwa sumbangan efektif variabel umur, jenis kelamin, dan suku terhadap agresivitas adalah 26,1% ( $R\text{ square} = 0,261$ ). Sisanya dijelaskan oleh hal lain diluar ketiga variabel tersebut.

Pada tabel **coefficients**, kita dapat melihat bahwa secara sendiri-sendiri, hanya umur dan laki saja yang signifikan. Artinya hanya kedua variabel ini saja yang berperan terhadap agresivitas.

Umur berperan terhadap agresivitas ( $B=-0,194$ , dan  $P<0,05$ ). Karena B nilainya negatif, artinya arah hubungan antara umur dan agresivitas adalah berbanding terbalik. Jika umur naik satu poin, maka agresivitas diprediksikan akan turun 0,194 poin.

Laki juga berperan terhadap agresivitas ( $B=8,275$  dan  $p<0,05$ ), atau secara sederhana kita dapat mengatakan bahwa ada perbedaan yang signifikan agresivitas antara laki-laki dan perempuan. Laki-laki

memiliki agresivitas yang lebih tinggi. Secara matematis, jika jenis kelamin berubah dari perempuan menjadi laki-laki (dari 0 menjadi 1), maka skor agresivitas diprediksikan akan naik 8,275 poin. Interpretasi matematis ini tentu harus disertai pertimbangan logis.

Pada variabel dummy suku, Jawa memiliki agresivitas yang lebih rendah dibanding Madura (*exclude group*). Hal ini ditunjukkan oleh nilai B yang negatif. Begitu pula Sunda juga memiliki agresivitas yang lebih rendah dibandingkan Madura. Namun, perlu diperhatikan bahwa semua variabel dummy pada suku tersebut tidak signifikan. Hal ini berarti sebenarnya tidak ada perbedaan agresivitas pada tiga suku Jawa, Sunda, Madura. Jadi interpretasi di atas akan berguna jika secara statistik koefisien variabel dummy signifikan.

### **Contoh Soal**

Untuk menyajikan data dengan variabel yang berbentuk dummy, kita mengkodennya dengan kode 1 dan 0. Misalnya jabatan yang terdiri dari 2 kategori, maka kita membuat 1 variabel baru yang isinya adalah kode 1 dan 0, misalnya 1 untuk manajer dan 0 untuk pelaksana. Agar memudahkan pemaknaan terhadap hasil berilah angka 1 pada kategori yang secara teoritik akan memiliki nilai prediksi lebih tinggi. Dalam hal ini saya memberikan kode 1 pada manajer. Data ini dapat diunduh [di sini](#)

Jika kategorinya lebih dari 2, maka kita juga melakukan hal yang sama. Variabel bidang kerja saya jabarkan menjadi 3 variabel baru, dengan kode 1 dan 0. Tampilannya bisa dilihat di Tabel 1.

puas	motif	jabatan	Manajer	Pelaksana	bidang	suplai	produksi	marketing
3	4	1	1	0	1	1	0	0
4	5	1	1	0	1	1	0	0
3	6	1	1	0	1	1	0	0
5	6	1	1	0	2	0	1	0
4	5	1	1	0	2	0	1	0
5	6	1	1	0	2	0	1	0
6	6	1	1	0	3	0	0	1
5	7	1	1	0	3	0	0	1
6	8	1	1	0	3	0	0	1
4	8	1	1	0	3	0	0	1
5	9	2	0	1	1	1	0	0
8	9	2	0	1	1	1	0	0
9	8	2	0	1	1	1	0	0
8	7	2	0	1	2	0	1	0
7	6	2	0	1	2	0	1	0
8	5	2	0	1	2	0	1	0
9	9	2	0	1	3	0	0	1
9	8	2	0	1	3	0	0	1
8	8	2	0	1	3	0	0	1
7	9	2	0	1	3	0	0	1

Terlihat ada variabel bidang yang terdiri dari 3 kategori (1 untuk suplai, 2 untuk produksi dan 3 untuk marketing). Lalu saya menjabarkannya menjadi 3 variabel yang memiliki nama sesuai dengan kategorinya. Misalnya untuk variabel suplai, kode 1 saya berikan pada subjek pada bagian suplai, dan kode 0 untuk variabel produksi dan marketing. Setelah siap, data bisa kita masukkan ke SPSS.

### Analisis Data

Untuk menganalisisnya, kita bisa menggunakan regresi berganda (*multiple regression*) atau regresi berjenjang (*hierarchical regression*). Jika kita hendak mencari secara langsung prediksi semua variabel independen kita bisa memiliki regresi berganda, akan tetapi jika kita ingin melihat prediksi tiap variabel independen secara berurutan kita menggunakan regresi berjenjang. Kali ini kita akan menggunakan regresi berganda.

Dalam analisis regresi berganda dengan variabel dummy, tidak semua kategori dalam variabel dummy dilibatkan analisis regresi jika jumlah kategorinya. Jika jumlah kategori lebih dari 2 maka jumlah kategori yang kita libatkan adalah k-1 (k=jumlah kategori). Misalnya jumlah kategori bidang

kerja adalah 3, maka jumlah kategori yang kita libatkan adalah  $3-1=2$  kategori. Untuk desain kita seperti ini, kita masukkan variabel kepuasan kerja (Y), ketahanan kerja (X1), jabatan (X2) dan 2 kategori pada bidang kerja (X3) yaitu produksi dan marketing. Kategori suplai tidak kita masukkan karena menjadi basis interpretasi variabel bidang kerja nantinya.

- Memprediksi kepuasan kerja (Y) berdasarkan ketahanan kerja (X1).

	puas	tahan	manajer	suplai	produksi
1	3	4	0	1	0
2	4	5	0	1	0
3	3	6	0	1	0
4	5	6	0	0	1
5	4	5	0	0	1
6	5	6	0	0	1
7	6	6	0	0	0
8	5	7	0	0	0
9	6	8	0	0	0
10	4	8	0	0	0
11	5	9	1	1	0
12	8	9	1	1	0
13	9	8	1	1	0
14	8	7	1	0	1
15	7	6	1	0	1
16	8	5	1	0	1
17	9	9	1	0	0
18	9	8	1	0	0
19	8	8	1	0	0
20	7	9	1	0	0

- Bukalah SPSS – Tekan menu ANALYZE – REGRESSION – LINIER. Pada menu yang ditampilkan masukkan variabel puas (Y) pada DEPENDENT, trus tahan (X1) pada kolom INDEPENDENT. Klik OK.

**Model Summary**

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.546 <sup>a</sup>	.298	.259	1.750

a. Predictors: (Constant), tahan

**ANOVA<sup>b</sup>**

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	23.406	1	23.406	7.640	.013 <sup>a</sup>
	Residual	55.144	18	3.064		
	Total	78.550	19			

a. Predictors: (Constant), tahan  
b. Dependent Variable: puas



Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	1.243	1.818		.684	.503
	tahan	.706	.255	.546	2.764	.013

a. Dependent Variable: puas

- Hasil analisis menunjukkan bahwa variabel ketahanan mampu memprediksi kepuasan kerja ( $F=7.640$ ;  $p<0.05$ ) yang terlihat pada kolom ANOVA dengan sumbangan efektif sebesar 29.8 persen yang terlihat dari tabel Model Summary. Artinya ketahanan dapat memprediksi sebesar 29.8 persen kepuasan individu, sisanya diprediksi oleh hal-hal selain ketahanan. Gak apa, 29.8 persen termasuk sumbangan yang sedang kok. Dari tabel koefisien kita dapat menurunkan menjadi persamaan regresi sebagai berikut :

$$Y = B_0 + B_1(\text{TAHAN}) = 1.243 + 0.706(\text{TAHAN})$$

Dari persamaan ini kita dapat memprediksi jika seseorang memiliki skor ketahanan mental sebesar 2 maka kepuasannya sebesar  $Y = 1.243 + 0.706(2) = 2.655$ . Dan jika ia memiliki ketahanan sebesar 0 maka kepuasannya sebesar 1.243. Inilah yang berkaitan dengan hal-hal lain selain ketahanan yang mempengaruhi kepuasan individu, atau  $100\% - 29.8\% = 70.20\%$

## Latihan 2

- Kita memprediksi kepuasan kerja (Y) berdasarkan jabatan (X2). Variabel jabatan telah di dummy dengan kode kategori 1 untuk manajer dan 0 untuk pelaksana.
- Bukalah SPSS – Tekan menu ANALYZE – REGRESSION – LINIER. Pada menu yang ditampilkan masukkan variabel puas (Y) pada DEPENDENT, trus jabatan (X2) pada kolom INDEPENDENT Klik OK.
- Terlihat pada tabel ANOVA bahwa jabatan mampu memprediksi kepuasan secara signifikan ( $F=40.668$ ;  $p<0.05$ ). Pada tabel Coefficient terlihat bahwa ada perbedaan kepuasan antara jabatan manajer dan pelaksana ( $b=3.300$ ;  $p<0.05$ ). Persamaan regresi untuk analisis ini adalah sebagai berikut :

$$Y = B_0 + B_1(\text{MANAJER})$$

$$Y = 4.50 + 3.30(\text{MANAJER})$$

Misalnya kita ditanya berapakah kepuasan individu PELAKSANA ? Kita masukkan ke dalam persamaan di atas, PELAKSANA adalah sama dengan MANAJER = 0 (ingat pengkodean di muka)

$$Y = 4.50 + 3.30(0) = 3.30$$

Berapa kepuasan MANAJER (MANAJER = 1)

$$Y = 4.50 + 3.30(1) = 7.80$$

Jabatan manajer memiliki kepuasan lebih tinggi dibanding dengan pelaksana (selisih 4.5 poin)

### Latihan 3

Sekarang sedikit kompleks, untuk memprediksi kepuasan kerja (Y) berdasarkan bidang kerja (X3) yang memiliki banyak kategori.

- Bukalah SPSS – Tekan menu ANALYZE – REGRESSION – LINIER. Pada menu yang ditampilkan masukkan variabel PUAS (Y) pada DEPENDENT, trus masukkan kategori SUPLAI dan PRODUKSI pada kolom INDEPENDENT Klik OK. Variabel MARKETING tidak kita masukkan karena ia akan menjadi referensi seperti halnya variabel PELAKSANA sebelumnya.
- Terlihat pada tabel ANOVA bahwa bidang kerja tidak mampu memprediksi kepuasan ( $F=.816$ ;  $p>0.05$ ). Sumbangan efektif prediksi bidang kerja hanya sekitar 0.88 persen. Sedikit sekali ya, 99,2 persen adalah prediksi oleh variabel yang lain. Kita masuk ke tabel COEFFICIENT. Sama seperti pada analisis berdasarkan jabatan di muka, prediksi kepuasan dapat kita ketahui melalui persamaan regresi. Berikut ini persamaannya.

$$Y = B_0 + B_1(\text{SUPLAI}) + B_2(\text{PRODUKSI})$$

$$Y = 6.750 - 1.417(\text{SUPLAI}) - 0.583(\text{PRODUKSI})$$

ingat MARKETING tidak kita masukkan karena menjadi kategori yang dijadikan referensi, seperti halnya PELAKSANA

Jika kita ditanya berapa KEPUASAN bidang SUPLAI ?

$$Y = 6.750 - 1.417(1) - 0.583(0) = 4.75$$

Berapa KEPUASAN bidang MARKETING ?

$$Y = 6.750 - 1.417(o) - 0.583 (0) = 6.75$$

- Nilai signifikansi di atas menunjukkan perbedaan antara kategori yang ditampilkan dengan MARKETING sebagai referensi. Terlihat bahwa tidak ada perbedaan KEPUASAN antara bidang MARKETING dan SUPLAI ( $t=-1.278$ ;  $p>0.05$ ) dan juga tidak ada perbedaan antara bidang MARKETING dengan PRODUKSI ( $t=-0.526$ ;  $p>0.05$ )

#### Latihan 4

Pada bagian ini kita akan menganalisis semua variabel yang kita miliki.

- Bukalah SPSS – Tekan menu ANALYZE – REGRESSION – LINIER. Pada menu yang ditampilkan masukkan variabel PUAS (Y) pada DEPENDENT, trus tahan (X1), MANAJER (X2), dan 2 kategori bidang kerja (X3), yaitu SUPLAI dan PRODUKSI pada INDEPENDENT(S) lalu klik OK. Cara membacanya sama seperti yang saya contohkan.

Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.884 <sup>a</sup>	.781	.722	1.071

a. Predictors: (Constant), produksi, manajer, suplai, tahan

ANOVA <sup>b</sup>						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	61.336	4	15.334	13.361	.000 <sup>a</sup>
	Residual	17.214	15	1.148		
	Total	78.550	19			

a. Predictors: (Constant), produksi, manajer, suplai, tahan  
b. Dependent Variable: puas

Coefficients <sup>a</sup>						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	5.022	1.825		2.752	.015
	tahan	.011	.252	.009	.044	.966
	manajer	3.281	.643	.828	5.107	.000
	suplai	-1.405	.635	-.325	-2.212	.043
	produksi	-.561	.774	-.130	-.724	.480

a. Dependent Variable: puas

- Terlihat pada tabel ANOVA bahwa secara simultan (bersama-sama) semua prediktor mampu memprediksi KEPUASAN secara signifikan ( $F=13.361$ )

$p < 0.05$ ). Sumbangan efektif prediksi bidang kerja hanya sekitar 78.1 persen. Kita masuk ke tabel COEFFICIENT, prediksi KEPUASAN dapat kita ketahui melalui persamaan regresi. Berikut ini persamaannya :

$$Y = B_0 + B_1(\text{TAHAN}) + B_2(\text{MANAJER}) + B_3(\text{SUPLAI}) + B_4(\text{PRODUKSI})$$

$$Y = 5.022 - 0.011(\text{TAHAN}) + 3.281(\text{MANAJER}) - 1.405(\text{SUPLAI}) + 0.561(\text{PRODUKSI})$$

Jika di tanya berapakah KEPUASAN pekerja pada jabatan PELAKSANA dalam bidang SUPLAI jika KETAHANAN mereka skornya 2 ?

$$Y = 5.022 - 0.011(2) + 3.281(0) - 1.405(1) + 0.561(0) = 3.95$$

Nilai signifikansi di atas menunjukkan bahwa ketahanan tidak mampu memprediksi secara signifikan ( $B = 0.011$ ;  $p > 0.05$ ). Ada perbedaan kepuasan antara MANAJER dan PELAKSANA ( $B = 3.281$ ;  $p < 0.05$ ) dan seterusnya, sama membacanya dengan cara sebelumnya.

- Terlihat bahwa ada perbedaan antara hasil analisis secara terpisah dengan analisis secara simultan. Meski secara umum dalam hal signifikansi tidak ada perbedaan, akan tetapi dari sisi prediksi ada perbedaan. Memang demikian adanya, yang dikarenakan ada varian-varian di data turut dilibatkan. Analisis secara simultan memiliki kelebihan dibanding secara terpisah, karena manusia adalah integral, kita tidak hanya melihat satu sisi darinya namun harus melihat sisi yang lainnya pula. Dalam hal ini kita tidak melihat KEPUASAN individu tidak hanya dari KETAHANANNYA saja akan tetapi dari JABATAN dan BIDANG KERJANYA.
- Analisis regresi dengan menggunakan variabel dummy memiliki kompleksitas, hasil analisis ini memiliki kemiripan dengan analisis kovarian, namun memiliki sistematika komputasi yang sedikit berbeda.

## 2. Regresi untuk Uji Non Parametrik

Jika asumsi uji parametrik tidak terpenuhi

### a) Regresi Logistik

Misalkan kita akan melakukan uji regresi logistik sebuah penelitian yang berjudul Pengaruh Rokok dan Riwayat Kanker Terhadap Kanker Paru. Di mana variabel bebas ada 2 yaitu rokok dan riwayat kanker pada keluarga dan variabel terikatnya adalah kejadian kanker paru. Rokok terdiri dari 2 kategori yaitu “tidak merokok (kode 0)” dan “merokok (kode 1).” Riwayat terdiri dari 2 kategori yaitu “tidak ada riwayat (kode 0)” dan “ada riwayat (kode 1).” Kanker paru terdiri dari 2 kategori yaitu “tidak mengalami kanker (kode 0)” dan “mengalami kanker (kode 1).” Sebagai catatan: kategori yang terburuk diberi kode 1 dan kategori yang terbaik diberi kode 0.

- Langsung saja anda buka aplikasi SPSS anda dan masukkan data sebagai berikut sebanyak 200 sampel. (terlampir).

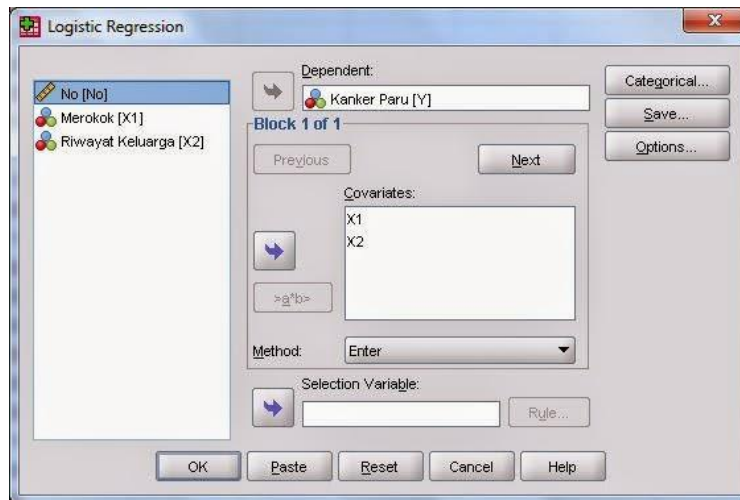
	No	X1	X2	Y
1	1	0	0	0
2	2	0	1	0
3	3	1	0	0
4	4	1	1	1
5	5	1	1	0
6	6	0	0	0
7	7	0	0	0
8	8	1	1	1
9	9	0	0	1

- Jangan lupa pilih tab variable view:

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	No	Numeric	8	0	No	None	None	8	Right	Scale
2	X1	Numeric	8	0	Merokok	{0, Tidak M...	None	8	Right	Nominal
3	X2	Numeric	8	0	Riwayat Keluarga	{0, Tidak Ad...	None	8	Right	Nominal
4	Y	Numeric	8	0	Kanker Paru	{0, Tidak M...	None	8	Right	Nominal

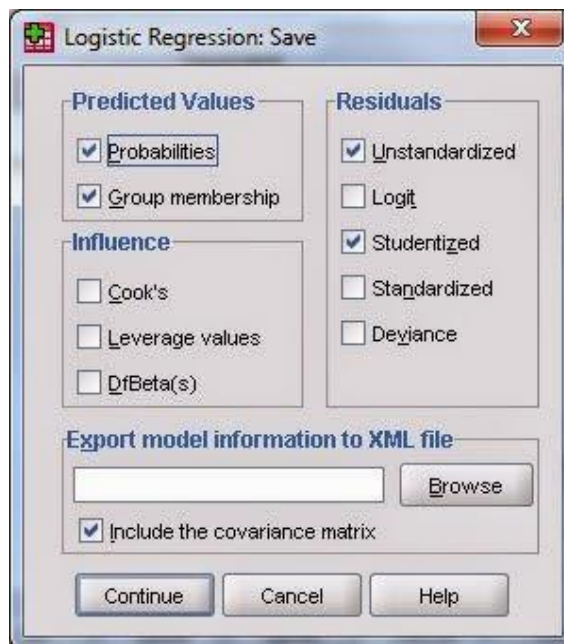
- Tahap Analisis Regresi Logistik

Kemudian pada menu, klik **Analyze -> Regression -> Binary Logistic**. Kemudian masukkan variabel terikat ke kotak **dependent** dan masukkan semua variabel bebas ke kotak **Covariates**.



- **Save Regresi Logistik**

Tekan tombol **Save** lalu centang **Probabilities**, **Group membership**, **Unstandardized** dan **Studentized** kemudian klik **Continue**.



- Tekan tombol **Options** lalu centang **Classification plots**, **Hosmer-lemeshow goodness-of-fit**, **Casewise listing residuals** dan pilih **Outliers outside** dan isi dengan angka **2**, **Correlation of estimates**, **Iteration history**, **CI for exp(B)** dan isi dengan **95**.

- **Opsi Regresi Logistik**

Sedangkan nilai *Maximum iteration* biarkan tetap sebesar 20 dan nilai *Classification Cutoff* tetap 0.5. Nilai ini disebut dengan *the*

*cut value* atau *prior probability*, yaitu peluang suatu observasi untuk masuk ke dalam salah satu kelompok sebelum karakteristik variabel penjelasnya diketahui. Apabila kita tidak mempunyai informasi tambahan tentang data kita, maka bisa langsung menggunakan nilai default yaitu 0,5. Jika tidak ada penelitian sebelumnya, dapat digunakan *classification cutoff* sebesar 0,5. Namun, jika ada penelitian lain yang telah meneliti maka bisa dinaikkan/diturunkan *classification cutoff* sesuai hasil penelitian.



- Kemudian pada jendela utama, klik **OK** dan segera lihat Output anda.
- **Output**
  - ✓ **Iteration History**

**Block 0: Beginning**

**Iteration History<sup>a, b, c</sup>**

Iteration	-2 Log likelihood	Coefficients
		Constant
Step 0	1	276.939
	2	276.939

a. Constant is included in the model.  
 b. Initial -2 Log Likelihood: 276.939  
 c. Estimation terminated at iteration number 2 because parameter estimates changed by less than .001.

- Tabel Iteration History pada block 0 atau saat variabel independen tidak dimasukkan dalam model: N=200 mendapatkan Nilai -2 Log Likelihood: 276,939.
- Degree of Freedom (DF) = N – 1 = 200-1=199. Chi-Square (X2) Tabel Pada DF 199 dan Probabilitas 0.05 = 232,912.
- Nilai -2 Log Likelihood (276,939) > X2 tabel (232,912) sehingga menolak H0, maka menunjukkan bahwa model sebelum memasukkan variabel independen adalah TIDAK FIT dengan data.

✓ **Classification Table**

**Classification Table<sup>a, b</sup>**

		Predicted		
		Kanker Paru		Percentage Correct
Observed		Tidak Mengalami Kanker	Mengalami Kanker	
Step 0	Kanker Paru	0	96	.0
	Tidak Mengalami Kanker	0	104	100.0
Overall Percentage				52.0

a. Constant is included in the model.  
b. The cut value is .500

Merupakan tabel kontingensi 2 x 2 yang seharusnya terjadi atau disebut juga frekuensi harapan berdasarkan data empiris variabel dependen, di mana jumlah sampel yang memiliki kategori variabel dependen referensi atau akibat buruk (kode 1) yaitu “Mengalami Kanker” sebanyak 104. Sedangkan yang “Tidak mengalami Kanker” sebanyak 96 orang. Jumlah sampel sebanyak 200 orang. Sehingga nilai overall percentage sebelum variabel independen dimasukkan ke dalam model sebesar:  $104/200 = 52,0\%$ .

✓ **Variabel Dalam Persamaan Tahap Beginning**

**Variables in the Equation**

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 0 Constant	.080	.142	.320	1	.572	1.083



Di atas pada tabel *Variables in The Equation*: Saat sebelum var independen di masukkan ke dalam model, maka belum ada variabel independen di dalam model. Nilai Slope atau Koefisien Beta (B) dari Konstanta adalah sebesar 0,080 dengan Odds Ratio atau Exp(B) sebesar 1,083. Nilai Signifikansi atau p value dari uji Wald sebesar 0,572.

- Perlu diingat bahwa nilai B identik dengan koefisien beta pada *Ordinary Least Square (OLS)* atau regresi linear. Sedangkan Uji Wald identik dengan t parsial pada OLS. Sedangkan Exp(B) adalah nilai eksponen dari B, maka  $Exp(0,080) = 1,083$ .

✓ **Variabel Tidak Dalam Persamaan Tahap Beginning**

	Score	df	Sig.
Step 0 Variables X1	61.191	1	.000
X2	50.261	1	.000
Overall Statistics	64.605	2	.000

Di atas Menunjukkan variabel yang belum dimasukkan ke dalam model regresi, yaitu variabel X1 dan X2. Di mana X1 adalah variabel merokok dan X2 adalah variabel riwayat keluarga.

✓ **Tahap Entry Variabel**

**Block 1: Method =**

Iteration	-2 Log likelihood	Coefficients		
		Constant	X1	X2
Step 1 1	208.765	-1.077	1.621	.790
2	207.580	-1.208	1.823	.962
3	207.575	-1.214	1.837	.973
4	207.575	-1.214	1.837	.973

a. Method: Enter  
 b. Constant is included in the model.  
 c. Initial -2 Log Likelihood: 276.939  
 d. Estimation terminated at iteration number 4 because parameter estimates changed by less than .001.

Di atas saat variabel independen dimasukkan dalam model:  $N=200$ . Degree of Freedom (DF) =  $N - \text{jumlah variabel independen} - 1 = 200-2-1=197$ . Chi-Square ( $X^2$ ) Tabel Pada DF 197 dan Prob 0.05 = 230,746.

Nilai  $-2 \text{ Log Likelihood}$  (207,575) <  $X^2$  tabel (230,746) sehingga menerima  $H_0$ , maka menunjukkan bahwa model dengan memasukkan variabel independen adalah FIT dengan data. Hal ini berbeda dengan *Block Beginning* di atas, di mana saat sebelum variabel independen dimasukkan ke dalam model, model TIDAK FIT dengan data.

✓ **Hasil Omnibus Test**

		Chi-square	df	Sig.
Step 1	Step	69.364	2	.000
	Block	69.364	2	.000
	Model	69.364	2	.000

Nilai  $X^2$  69,394 >  $X^2$  tabel pada DF 2 (jumlah variabel independen 2) yaitu 5,991 atau dengan signifikansi sebesar 0,000 (< 0,05) sehingga menolak  $H_0$ , yang menunjukkan bahwa penambahan variabel independen DAPAT memberikan pengaruh nyata terhadap model, atau dengan kata lain model dinyatakan FIT.

✓ **Jawaban Hipotesis**

Perlu diingat jika pada *OLS* untuk menguji signifikansi simultan menggunakan uji F, sedangkan pada regresi logistik menggunakan nilai *Chi-Square* dari selisih antara  $-2 \text{ Log likelihood}$  sebelum variabel independen masuk model dan  $-2 \text{ Log likelihood}$  setelah variabel independen masuk model. Pengujian ini disebut juga dengan pengujian *Maximum likelihood*.

Sehingga jawaban terhadap hipotesis pengaruh simultan variabel independen terhadap variabel dependen adalah menerima H1 dan menolak H0 atau yang berarti ada pengaruh signifikan secara simultan merokok dan riwayat keluarga terhadap kejadian kanker paru oleh karena nilai p value Chi-Square sebesar 0,000 di mana  $< \text{Alpha } 0,05$  atau nilai Chi-Square Hitung  $69,364 > \text{Chi-Square tabel } 5,991$ .

✓ **Pseudo R Square**

Model Summary			
Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	207.575 <sup>a</sup>	.293	.391

a. Estimation terminated at iteration number 4 because parameter estimates changed by less than .001.

Di atas pada tabel *Model Summary*: Untuk melihat kemampuan variabel independen dalam menjelaskan variabel dependen, digunakan nilai *Cox & Snell R Square* dan *Nagelkerke R Square*. Nilai-nilai tersebut disebut juga dengan *Pseudo R-Square* atau jika pada regresi linear (OLS) lebih dikenal dengan istilah *R-Square*.

Nilai *Nagelkerke R Square* sebesar 0,391 dan *Cox & Snell R Square* 0,293, yang menunjukkan bahwa kemampuan variabel independen dalam menjelaskan variabel dependen adalah sebesar 0,391 atau 39,1% dan terdapat  $100\% - 39,1\% = 60,9\%$  faktor lain di luar model yang menjelaskan variabel dependen.

✓ **Hosmer and Lemeshow Test**

Hosmer and Lemeshow Test adalah uji *Goodness of fit test (GoF)*, yaitu uji untuk menentukan apakah model yang dibentuk sudah tepat atau tidak. Dikatakan tepat apabila tidak ada perbedaan signifikan antara model dengan nilai observasinya.

### Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	13.671	1	.000

Nilai *Chi Square* tabel untuk DF 1 (Jumlah variabel independen – 1) pada taraf signifikansi 0,05 adalah sebesar 3,841. Karena nilai *Chi Square Hosmer and Lemeshow* hitung 13,671 > *Chi Square table* 3,841 atau nilai signifikansi sebesar 0,000 (< 0,05) sehingga menolak H<sub>0</sub>, yang menunjukkan bahwa model TIDAK dapat diterima dan pengujian hipotesis TIDAK dapat dilakukan sebab ada perbedaan signifikan antara model dengan nilai observasinya.

Oleh karena dalam tutorial Interpretasi regresi logistik dengan SPSS ini, nilai Hosmer and Lemeshow Test menolak H<sub>0</sub>, sebaiknya anda mencoba untuk membuat agar nilai Hosmer and Lemeshow Test menerima H<sub>0</sub>.

#### ✓ Cara Mengatasi Hosmer Lemeshow

Caranya adalah dengan mengubah model persamaan regresi logistik dengan menambahkan variabel interaksi antar variabel independen. Misal pada kasus di sini, dengan menambahkan variabel interaksi antara X<sub>1</sub> dan X<sub>2</sub>.

#### - Classification Result

		Predicted			
		Kanker Paru		Percentage Correct	
Observed		Tidak Mengalami Kanker	Mengalami Kanker		
Step 1	Kanker Paru	Tidak Mengalami Kanker	78	18	81.2
		Mengalami Kanker	27	77	74.0
Overall Percentage					77.5

a. The cut value is .500

Berdasarkan tabel *Classification Table* di atas, jumlah sampel yang tidak mengalami kanker 78 + 18 = 96 orang. Yang benar-benar tidak mengalami kanker sebanyak 78 orang dan yang seharusnya tidak mengalami kanker namun mengalami, sebanyak 18

orang. Jumlah sampel yang mengalami kanker  $27 + 77 = 104$  orang. Yang benar-benar mengalami kanker sebanyak 27 orang dan yang seharusnya mengalami kanker namun tidak mengalami, sebanyak 77 orang.

Dalam Interpretasi regresi logistik dengan SPSS: Tabel di atas memberikan nilai *overall percentage* sebesar  $(78+77)/200 = 77,5\%$  yang berarti ketepatan model penelitian ini adalah sebesar 77,5%.

- **Pendugaan Parameter**

Variables in the Equation									
		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I. for EXP(B)	
								Lower	Upper
Step 1 <sup>a</sup>	X1	1.837	.455	16.301	1	.000	6.277	2.573	15.313
	X2	.973	.452	4.638	1	.031	2.645	1.091	6.411
	Constant	-1.214	.240	25.578	1	.000	.297		

a. Variable(s) entered on step 1: X1, X2.

Lihat tabel *Variabel in the equation* di atas: semua variabel independen nilai P value uji wald (Sig) < 0,05, artinya masing-masing variabel mempunyai pengaruh parsial yang signifikan terhadap Y di dalam model. X1 atau merokok mempunyai nilai Sig Wald  $0,000 < 0,05$  sehingga menolak H0 atau yang berarti merokok memberikan pengaruh parsial yang signifikan terhadap kejadian kanker paru. X2 atau riwayat keluarga mempunyai nilai Sig Wald  $0,031 < 0,05$  sehingga menolak H0 atau yang berarti riwayat keluarga memberikan pengaruh parsial yang signifikan terhadap kejadian kanker paru.

- **Odds Ratio**

Besarnya pengaruh ditunjukkan dengan nilai EXP (B) atau disebut juga ODDS RATIO (OR). Variabel Merokok dengan OR 6,277 maka orang yang merokok (kode 1 variabel independen), lebih beresiko mengalami kanker paru (kode 1 variabel dependen)

sebanyak 6,277 kali lipat di bandingkan orang yang tidak merokok (kode 0 variabel independen). Nilai B = Logaritma Natural dari 6,277 = 1,837. Oleh karena nilai B bernilai positif, maka merokok mempunyai hubungan positif dengan kejadian kanker.

Variabel Riwayat Keluarga dengan OR 2,645 maka orang yang ada riwayat keluarga (kode 1 variabel independen), lebih beresiko mengalami kanker paru (kode 1 variabel dependen) sebanyak 2,645 kali lipat di bandingkan orang yang tidak ada riwayat keluarga (kode 0 variabel independen). Nilai B = Logaritma Natural dari 2,645 = 0,973. Oleh karena nilai B bernilai positif, maka riwayat keluarga mempunyai hubungan positif dengan kejadian kanker.

- **Persamaan Regresi Logistik**

Berdasarkan nilai-nilai B pada perhitungan di atas, maka model persamaan yang dibentuk adalah sebagai berikut:

$$\mathbf{\ln P/1-P = -1,214 + 1,837 \text{ Rokok} + 0,973 \text{ Riwayat Keluarga.}}$$

Atau bisa menggunakan rumus turunan dari persamaan di atas, yaitu:

$$\mathbf{\text{Probabilitas} = \exp(-1,214 + 1,837 \text{ Rokok} + 0,973 \text{ Riwayat Keluarga}) / 1 + \exp(-1,214 + 1,837 \text{ Rokok} + 0,973 \text{ Riwayat Keluarga}).}$$

- **Contoh Penggunaan Persamaan Regresi Logistik**

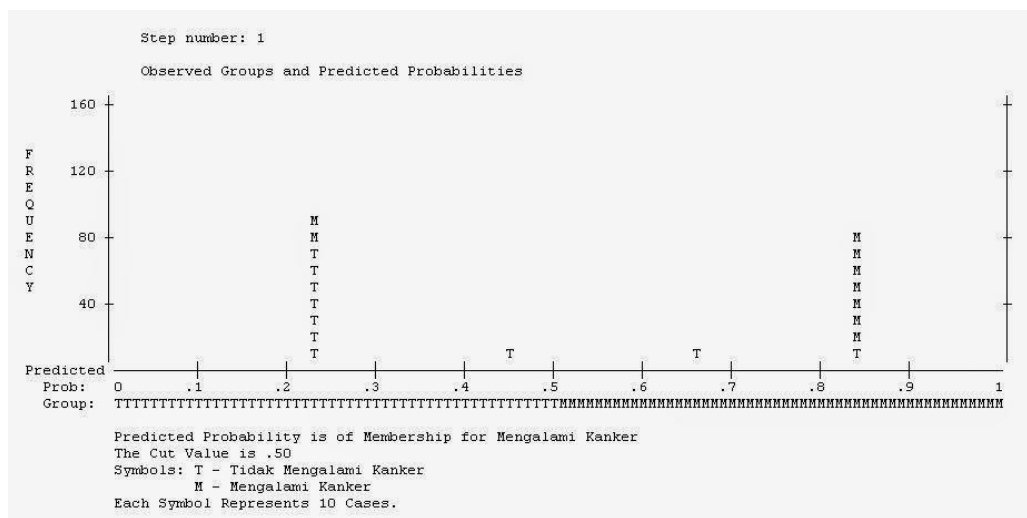
Misalkan sampel yang merokok dan ada riwayat keluarga, maka merokok=1 dan riwayat keluarga=1. Jika dimasukkan ke dalam model persamaan di atas, maka sebagai berikut:

- Probabilitas atau *Predicted* =  $(\exp(-1,214 + (1,837 \times 1) + (0,973 \times 1))) / (1 + \exp(-1,214 + (1,837 \times 1) + (0,973 \times 1)))$ .
- Probabilitas atau *Predicted* =  $(\exp(-1,214 + (1,837) + (0,973))) / (1 + \exp(-1,214 + (1,837) + (0,973)))$ .
- Probabilitas atau *Predicted* = 0.8315.

✓ **Nilai Predicted Regresi Logistik**

Oleh karena *Predicted*: 0,8315 > 0,5 maka nilai *Predicted Group Membership* dari sampel di atas adalah 1. Di mana 1 adalah kode mengalami kanker. Jadi jika sampel merokok (kode 1) dan ada riwayat keluarga (kode 1) maka prediksinya adalah mengalami kanker (kode 1). Jika seandainya sampel yang bersangkutan ternyata faktanya tidak mengalami kanker (kode 0) maka sampel tersebut keluar dari nilai prediksi.

Besarnya perbedaan atau yang disebut dengan *Residual = Predicted Group Membership - Predicted*. Pada kasus di atas di mana orang yang merokok dan ada riwayat keluarga namun faktanya tidak mengalami kanker, maka *Residual* = 0 – 0,8315 = -0,8315.



Di atas adalah grafik klasifikasi dari *observed group* dan *predicted group membership*.

## ✓ **Outlier**

### **Casewise List<sup>a</sup>**

a. The casewise plot is not produced because no outliers were found.

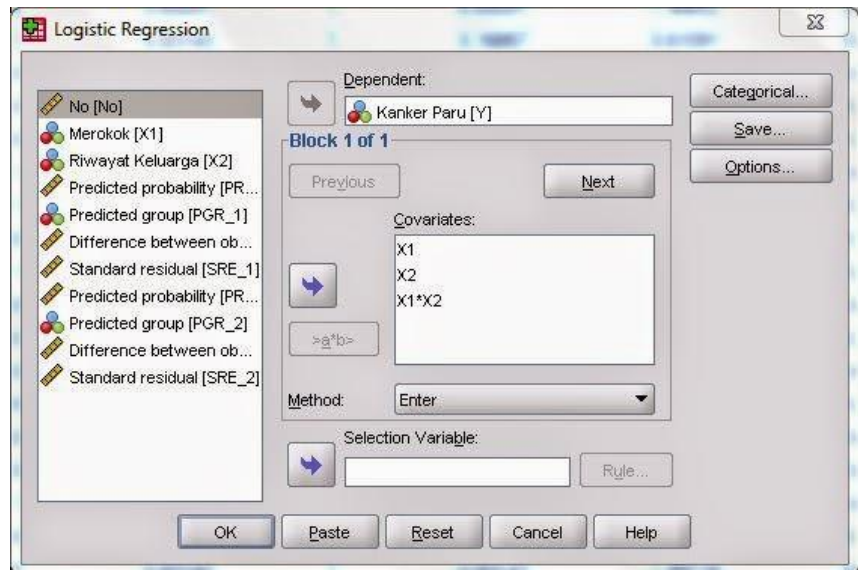
Agar anda paham Interpretasi Regresi Logistik dengan SPSS, maka perhatikan output di atas, yaitu output hasil deteksi outlier atau data pencilan. Pada kasus dalam tutorial ini tidak ada outlier dengan notifikasi seperti di atas, yaitu: *The casewise plot is not produced because no outliers were found.* Seandainya ada outlier, maka tampilan akan berubah dalam bentuk tabel yang berisi daftar sampel yang menjadi outlier beserta nilai *Studentized Residual*. Dinyatakan outlier apabila nilai Absolut dari *Studentized Residual* dari sampel tersebut lebih dari 2 ( $> 2$ ). Di mana *Studentized Residual* adalah standarisasi berdasarkan *Mean* dan *standart deviasi* dari nilai *Residual* yang telah dibahas di atas.

## ✓ **Variabel Interaksi**

Variabel interaksi adalah variabel yang merupakan hasil interaksi antar variabel independen. Variabel ini ada kalanya diperlukan untuk dimasukkan ke dalam model regresi logistik dengan alasan karena secara substantif memang ada hubungan antar variabel independen atau untuk mengatasi masalah *Goodness of Fit Test* di mana menolak  $H_0$ .

dengan menambahkan variabel interaksi ke dalam variabel independen, yaitu pada saat memasukkan variabel independen, seleksi variabel-variabel independen yang akan diinteraksikan, kemudian klik tombol “>a\*b>”.





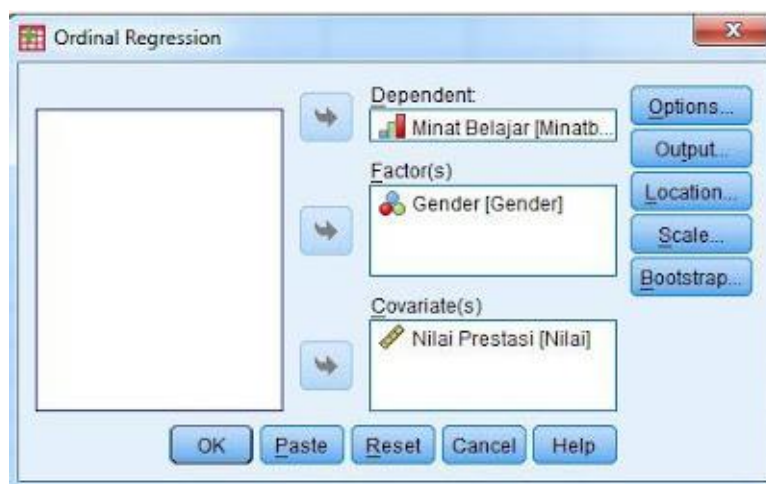
## b) Regresi Ordinal

### Contoh kasus :

Seorang dosen ingin mengetahui hubungan antara gender dan minat belajar di sebuah universitasnya. Variabel independen terdiri dari gender dan nilai prestasi sedangkan variabel dependen (Y) adalah minat belajar yang diukur dalam 3 tingkatan yaitu rendah, sedang dan tinggi.

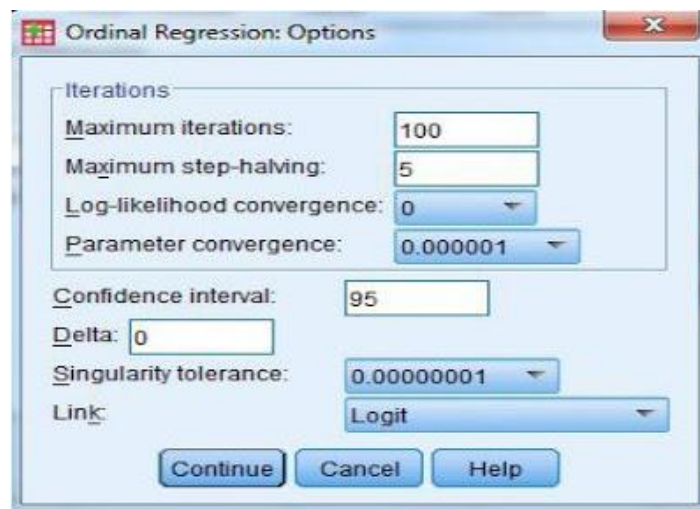
Langkah-langkah analisis regresi ordinal :

- Analyze >> Regression >> Ordinal
- Masukkan variabel Minat belajar ke kotak Dependent, Gender ke kotak Factor(s) dan Nilai prestasi ke Covariate.



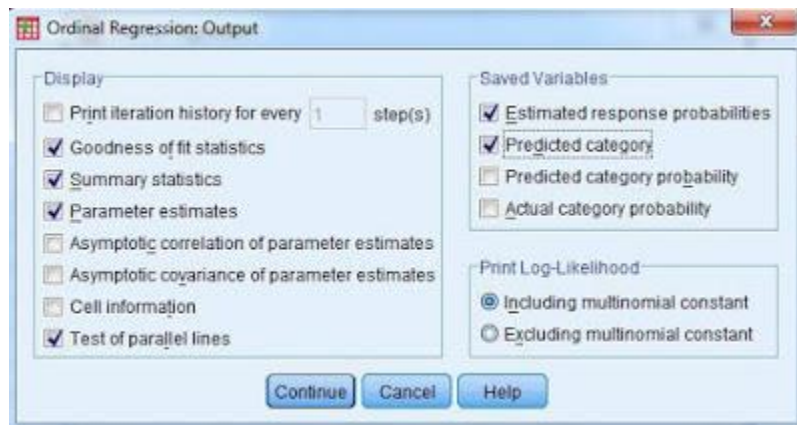
Menu Regresi Ordinal

- Pilih Option. Kita pilih *Link logit*. Klik Continue.



Ordinal Regression Option

- Pilih Output dan tik kotak *Predicted category, Estimated response probabilities* dan *Test of parallel lines*. Klik Continue.



Ordinal Regression Output

- Klik OK
- **Hasil Output** seperti di bawah ini.

Model	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	522.977			
Final	505.169	17.808	2	.000

Model Fitting Information

Pada **Model Fitting Information -2log Likelihood** menerangkan bahwa tanpa memasukkan variabel independen (*intercept only*) nilainya 522.977. Namun dengan memasukkan variabel independen ke model (*final*) terjadi penurunan nilai menjadi 505.167. Perubahan nilai ini merupakan nilai chi-square yaitu 17,808 dan signifikan pada taraf nyata 5% (sig.0.00).

Goodness-of-Fit			
	Chi-Square	df	Sig.
Pearson	317.892	380	.991
Deviance	350.797	380	.856

Goodness of fit

Tabel **Goodness of Fit** menunjukkan uji kesesuaian model dengan data. Nilai Pearson sebesar 317,892 dengan signifikansi 0,991 ( $> 0,05$ ) dan Deviance sebesar 350,797 dengan signifikansi 0,856 ( $> 0,05$ ). Hal ini berarti model sesuai dengan data empiris atau model layak digunakan.

Pseudo R-Square	
Cox and Snell	.044
Nagelkerke	.052
McFadden	.024

Pseudo R-Square

Tabel **Pseudo R-Square** menunjukkan bahwa seberapa besar variabel bebas (gender dan nilai prestasi) mampu menjelaskan variabel independen (minat belajar). Nilai ini seperti halnya koefisien determinasi pada regresi. Nilai Cox and Snell sebesar 0,044 (4,4%) dan Nagelkerke sebesar 0,052 (5,2%).

Parameter Estimates						
		Estimate	Std. Error	Wald	df	Sig.
Threshold	[Minatbelajar = .00]	1.479	.832	3.164	1	.075
	[Minatbelajar = 1.00]	3.543	.849	17.424	1	.000
Location	Nilai	.634	.255	6.177	1	.013
	[Gender = .00]	-.761	.252	9.163	1	.002
	[Gender = 1.00]	0 <sup>a</sup>			0	

#### Parameter Estimates

Tabel **Parameter Estimate** di atas, perhatikan nilai Wald dan nilai signifikansinya. Variabel nilai prestasi sebesar 6.177 dengan sig. 0,013 ( $< 0,05$ ) dan variabel gender sebesar 9,163 dengan sig.0,02 ( $< 0,05$ ). Hal ini menunjukkan faktor nilai prestasi dan gender berpengaruh terhadap minat belajar.

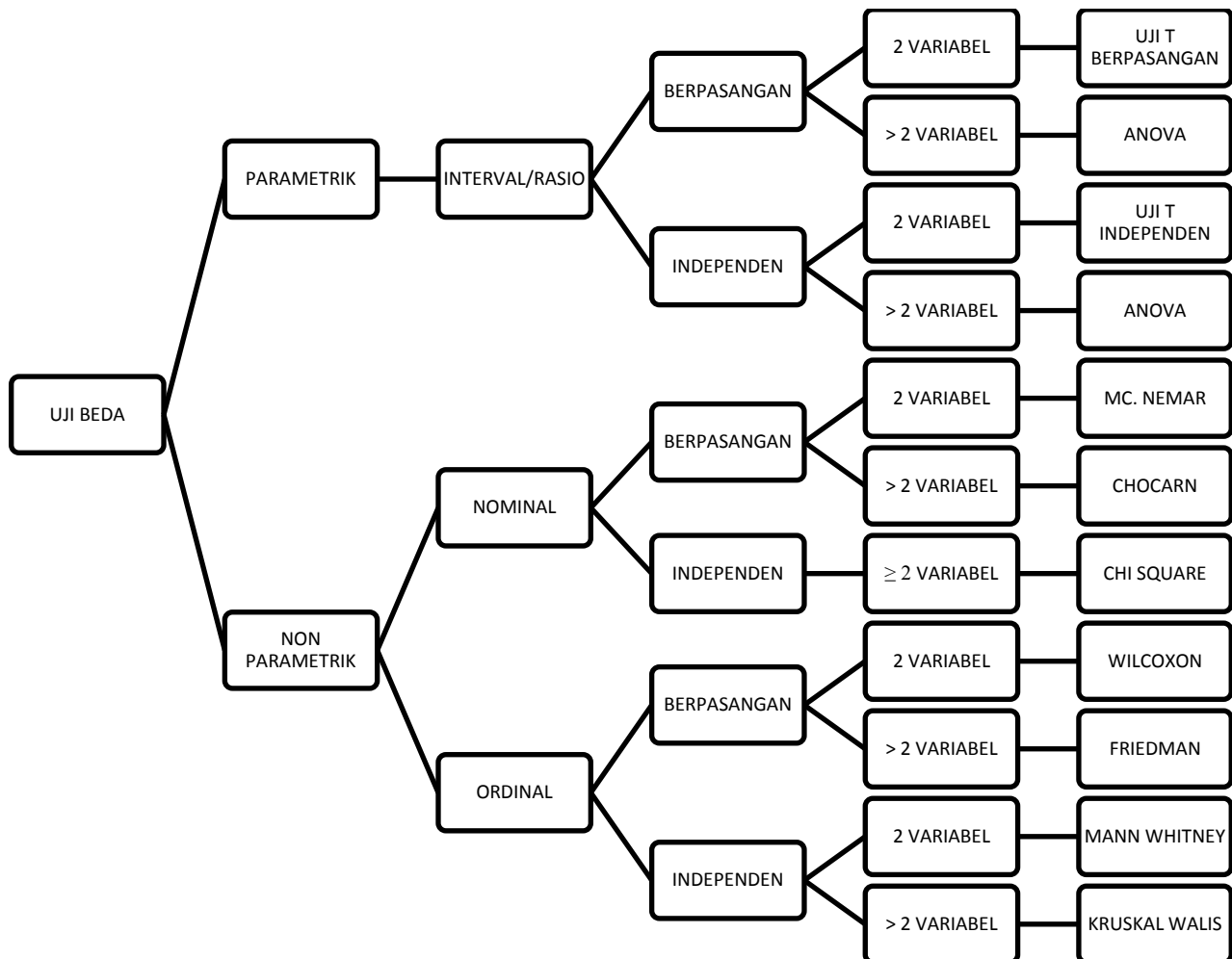
Test of Parallel Lines <sup>a</sup>				
Model	-2 Log Likelihood	Chi-Square	df	Sig.
Null Hypothesis	505.169			
General	504.301	.867	2	.648

#### Test of Parallel Lines

Tabel **Test of Parallel Lines** digunakan untuk menguji asumsi bahwa setiap kategori memiliki parameter yang sama atau hubungan antara variabel independen dengan logit adalah sama untuk semua persamaan logit. Oleh karena nilai signifikansi 0,648 ( $> 0,05$ ), maka terima  $H_0$  bahwa model yang dihasilkan memiliki parameter yang sama sehingga pemilihan *link function* adalah sesuai. Namun sebaliknya bila asumsi ini tidak terpenuhi, maka pemilihan *link function* logit tidak tepat.

# BAB 12 UJI BEDA BIVARIATE

## KERANGKA KONSEP



Independent Samples T-Test dan Paired Samples T-Test (berpasangan) termasuk dalam kelompok uji beda rerata (mean). Ada perbedaan mendasar dari dua jenis analisis data tersebut.

Independent Samples T-Test dan Paired Samples T-Test termasuk jenis uji beda rerata yang sama-sama menggunakan dua kelompok sampel, yang membedakan adalah: Independent Samples T-Test menggunakan dua kelompok yang anggotanya berbeda satu dengan yang lain, sedangkan Paired Samples T-Test menggunakan dua kelompok tetapi anggota dari dua kelompok tersebut sama.

- **Contoh**

Misalkan ada seorang mahasiswa yang akan melakukan penelitian eksperimen di kelas A dan B. Dia menerapkan model pembelajaran jigsaw di kelas A dan jigsaw 1 di kelas B, kemudian dia ingin membandingkan prestasi belajar kelas A dengan kelas B setelah diberi perlakuan tersebut.

Jelas ada dua kelompok di situ, yaitu kelas A dan kelas B, karena berbeda kelas maka tentunya berbeda pula anggotanya, sehingga analisis yang bisa digunakan adalah Independent Samples T-Test.

- **Contoh**

Misalkan ada seorang mahasiswa yang akan melakukan penelitian eksperimen di kelas A. Dia ingin membandingkan prestasi belajar siswa kelas A sebelum dan sesudah diterapkan model pembelajaran jigsaw.

Dari contoh tersebut sudah jelas bahwa ada satu kelas, tetapi ada dua kelompok. Dua kelompok tersebut adalah kelompok sebelum diberi perlakuan (model pembelajaran jigsaw) dan setelah diberi perlakuan (model pembelajaran jigsaw).

Karena hanya ada satu kelas, yaitu kelas A maka sudah tentu anggotanya adalah sama, sehingga analisis yang tepat untuk digunakan adalah Paired Samples T-Test.

## A. UJI BEDA UNTUK UJI PARAMETRIK

Bila seorang peneliti ingin mengetahui apakah parameter dua populasi berbeda atau tidak, maka uji statistik yang digunakan disebut uji beda dua mean. Umumnya, pendekatan yang dilakukan bisa dengan distribusi Z (uji Z), ataupun distribusi t (uji t).

Uji Z dapat digunakan bila (1) standar deviasi populasi ( $\sigma$ ) diketahui, dan (2) jumlah sampelnya besar ( $> 30$ ). Bila kedua syarat tersebut tidak terpenuhi, maka jenis uji yang digunakan adalah uji t dua sampel (two sample t-test).

Berdasarkan hubungan antar populasinya, uji t dapat digolongkan kedalam dua jenis uji, yaitu dependent sample t-test, dan independent sample t-test:

### 1. Paired sample t-test (Uji t berpasangan)

Dependent sample t-test atau sering diistilakan dengan Paired Sampel t-Test, adalah jenis uji statistika yang bertujuan untuk membandingkan rata-rata dua grup yang saling berpasangan. Sampel berpasangan dapat diartikan

sebagai sebuah sampel dengan subjek yang sama namun mengalami 2 perlakuan atau pengukuran yang berbeda, yaitu pengukuran sebelum dan sesudah dilakukan sebuah treatment.

**a) Syarat jenis uji ini adalah:**

- Data berdistribusi normal;
- Kedua kelompok data adalah dependen (saling berhubungan / berpasangan);
- Jenis data yang digunakan adalah numeric dan kategorik (dua kelompok).

**b) Rumus t-test yang digunakan untuk sampel berpasangan (paired)**

$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} - 2r\left(\frac{s_1}{\sqrt{n_1}}\right)\left(\frac{s_2}{\sqrt{n_2}}\right)}}$	<p><b>KETERANGAN :</b></p> <p><math>\bar{x}_1</math> = Rata-rata sampel 1</p> <p><math>\bar{x}_2</math> = Rata-rata sampel 2</p> <p><math>s_1</math> = Simpangan baku sampel 1</p> <p><math>s_2</math> = Simpangan baku sampel 2</p>
<p><math>s_1^2</math> = Varians sampel 1</p> <p><math>s_2^2</math> = Varians sampel 2</p> <p><math>r</math> = Korelasi antara dua sampel</p>	

**Contoh**

adalah bila seorang Manajer perusahaan ingin mengetahui apakah ada perbedaan prestasi penjualan seles setelah mengikuti pelatihan marketing. Setelah dilakukan rekapitulasi jumlah penjualan terhadap 15 orang sales, diperoleh data sebagai berikut:

Kode Sales	Rata-rata Penjualan (Unit/hari)	
	Sebelum Mengikuti Program Pelatihan	Setelah Mengikuti Program Pelatihan
Sales 1	2	4
Sales 2	3	3
Sales 3	3	4
Sales 4	2	3
Sales 5	3	4
Sales 6	3	5
Sales 7	4	3
Sales 8	2	3
Sales 9	3	3
Sales 10	4	4
Sales 11	4	4
Sales 12	4	4
Sales 13	2	4
Sales 14	3	3
Sales 15	3	3

- Merumuskan hipotesis, yaitu:  
 $H_0 = \text{Rata-rata penjualan sebelum mengikuti program pelatihan} = \text{Rata-rata penjualan setelah mengikuti program pelatihan},$   
 $H_1 = \text{Rata-rata penjualan sebelum mengikuti program pelatihan} \neq \text{Rata-rata penjualan setelah mengikuti program pelatihan}.$
- Kesimpulan yang dapat ditarik adalah :  $t \text{ hitung } (2.553) > t \text{ tabel } (2.145),$  yang berarti  $H_0$  ditolak dan  $H_1$  diterima, sehingga disimpulkan bahwa : rata-rata penjualan sebelum mengikuti program pelatihan  $\neq$  Rata-rata penjualan setelah mengikuti program pelatihan. Atau dengan kata lain, Program pelatihan marketing memberikan pengaruh yang signifikan terhadap prestasi penjualan sales.

### c) Cara menganalisis dan interpretasi SPSS

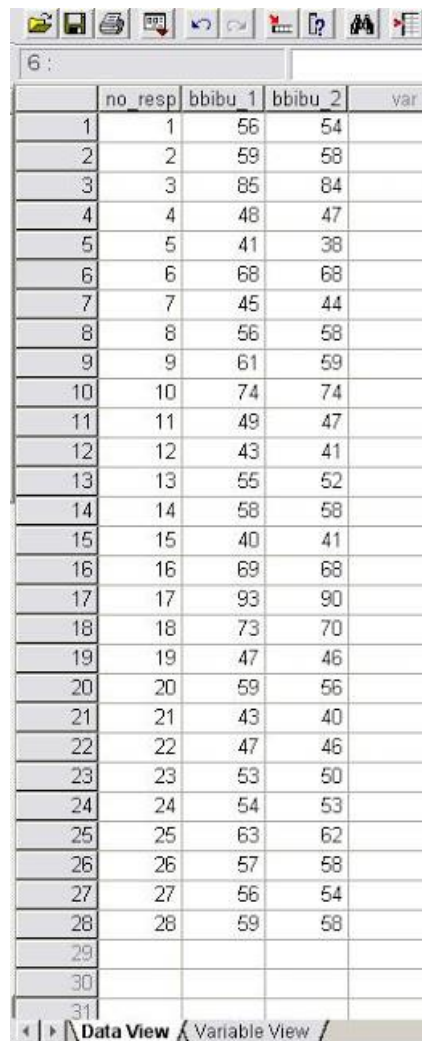
Uji-t untuk data berpasangan berarti setiap subjek diukur dua kali. Misalnya sebelum dan sesudah dilakukannya suatu intervensi atau pengukuran yang dilakukan terhadap pasangan orang kembar. Dalam contoh ini akan membandingkan data sebelum dengan sesudah intervensi.



### **Contoh Kasus :**

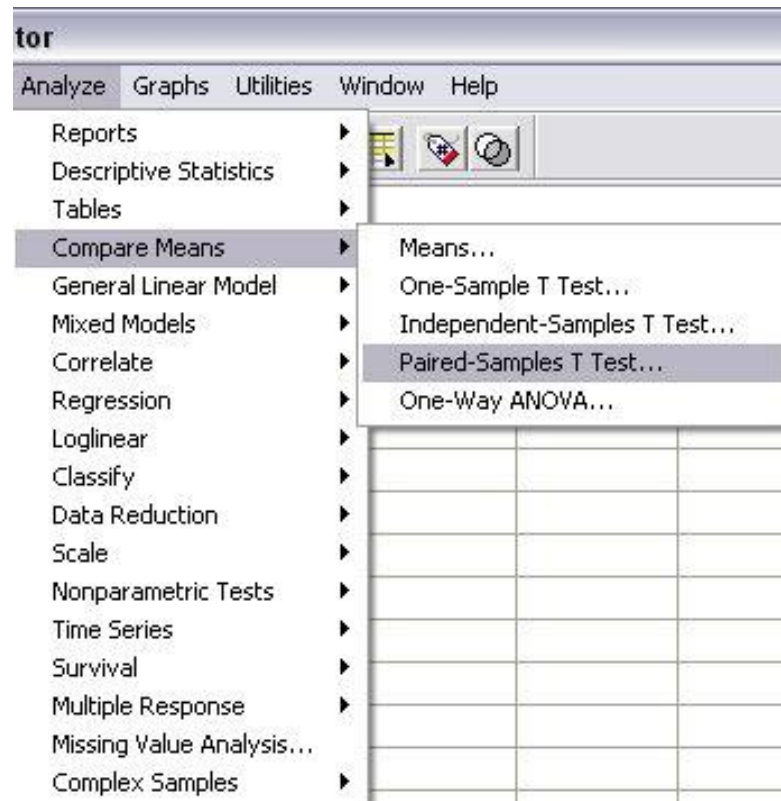
Suatu studi ingin mengetahui pengaruh suatu metode diet, lalu diambil 28 ibu sebagai sampel untuk menjalani program diet tersebut. Pengukuran berat badan yang pertama (BBIBU\_1) dilakukan sebelum kegiatan penyesuaian diet dilakukan, dan pengukuran berat badan yang kedua (BBIBU\_2) dilakukan setelah dua bulan menjalani penyesuaian diet.

- Buka SPSS, dan masukan datanya seperti ini :

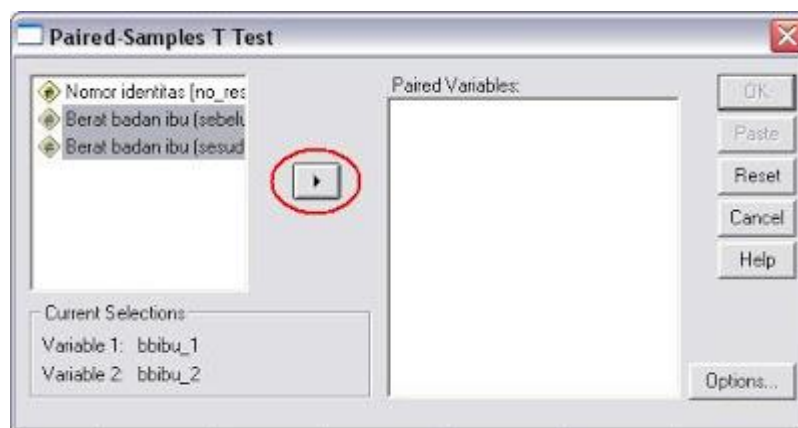


	no_resp	bbibu_1	bbibu_2	var
1	1	56	54	
2	2	59	58	
3	3	85	84	
4	4	48	47	
5	5	41	38	
6	6	68	68	
7	7	45	44	
8	8	56	58	
9	9	61	59	
10	10	74	74	
11	11	49	47	
12	12	43	41	
13	13	55	52	
14	14	58	58	
15	15	40	41	
16	16	69	68	
17	17	93	90	
18	18	73	70	
19	19	47	46	
20	20	59	56	
21	21	43	40	
22	22	47	46	
23	23	53	50	
24	24	54	53	
25	25	63	62	
26	26	57	58	
27	27	56	54	
28	28	59	58	
29				
30				
31				

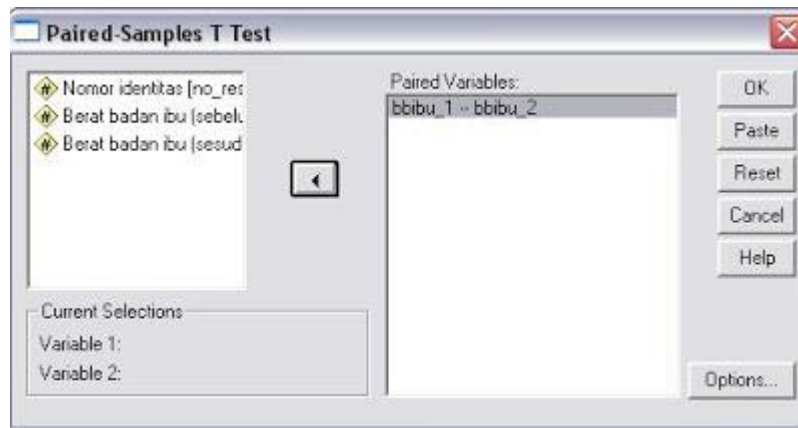
- Kita akan melakukan uji hipotesis untuk menilai apakah ada perbedaan berat badan ibu antara sebelum dengan sesudah mengikuti program diet, langkah-langkahnya sebagai berikut.



- Dari menu utama, pilihlah: Analyze-->Compare Mean-->Paired-Sample T-test....



- Pilih variabel BBIBU\_1 dan BBIBU\_2 dengan cara mengklik masing-masing variable tersebut. Kemudian klik tanda 'segitiga' untuk memasukkannya ke dalam kotak Paired-Variables. Seperti nampak di bawah ini :



- Selanjutnya klik OK untuk menjalankan prosedur. Pada layar Output tampak hasil seperti berikut:

**Paired Samples Statistics**

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Berat badan ibu (sebelum diet)	57,54	28	12,773	2,414
	Berat badan ibu (sesudah diet)	56,21	28	12,790	2,417

- Dari 28 subjek yang diamati terlihat bahwa rata-rata (mean) berat badan dari ibu sebelum intervensi (BBIBU\_1) adalah 57.54, dan rata-rata berat badan sesudah intervensi (BBIBU\_2) adalah 56,21. Uji 't' yang dilakukan terlihat pada tabel berikut:

**Paired Samples Test**

		Paired Differences				t	df	Sig. (2-tailed)	
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower				Upper
Pair 1	Berat badan ibu (sebelum diet) - Berat badan ibu (sesudah diet)	1,321	1,362	,257	,793	1,950	5,133	,000	

- Dari hasil uji-t berpasangan tersebut terlihat bahwa rata-rata perbedaan antara BBIBU\_1 dengan BBIBU\_2 adalah sebesar 1.321. Artinya ada penurunan berat badan sesudah intervensi dengan rata-rata penurunan sebesar 1.32 kg.
- Hasil perhitungan nilai "t" adalah sebesar 5,133 dengan p-value 0.000 dapat ditulis 0,001 (uji 2-arah). Hal ini berarti kita menolak Ho dan menyimpulkan bahwa secara statistik ada

perbedaan yang bermakna antara rata-rata berat badan sebelum dengan sudah intervensi.

- Dari hasil di atas kita bisa menilai bahwa program diet tersebut berhasil.

## 2. Independent sample t-test (Uji t independent)

Independent sample t-test adalah jenis uji statistika yang bertujuan untuk membandingkan rata-rata dua grup yang tidak saling berpasangan atau tidak saling berkaitan. Tidak saling berpasangan dapat diartikan bahwa penelitian dilakukan untuk dua subjek sampel yang berbeda.

Prinsip pengujian uji ini adalah melihat perbedaan variasi kedua kelompok data, sehingga sebelum dilakukan pengujian, terlebih dahulu harus diketahui apakah variannya sama (equal variance) atau variannya berbeda (unequal variance).

### a) Asumsi Independen T Test

Asumsi yang harus dipenuhi pada independen t test antara lain:

- 1) Skala data interval/rasio.
  - 2) Kelompok data saling bebas atau tidak berpasangan.
  - 3) Data per kelompok berdistribusi normal.
  - 4) Data per kelompok tidak terdapat outlier.
  - 5) Varians antar kelompok sama atau homogen.
- Homogenitas varian diuji berdasarkan rumus:

$F = \frac{S_1^2}{S_2^2}$	<b>KETERANGAN :</b> F = Nilai F hitung $S_1^2$ = Nilai varian terbesar $S_2^2$ = Nilai varian terkecil
---------------------------	---

Data dinyatakan memiliki varian yang sama (equal variance) bila F-Hitung < F-Tabel, dan sebaliknya, varian data dinyatakan tidak sama (unequal variance) bila F-Hitung > F-Tabel.

Bentuk varian kedua kelompok data akan berpengaruh pada nilai standar error yang akhirnya akan membedakan rumus pengujiannya.

- Uji t untuk varian yang sama (equal variance) menggunakan rumus Polled Varians:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

- Uji t untuk varian yang berbeda (unequal variance) menggunakan rumus Separated Varians:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

Contohnya adalah seorang mahasiswa pertanian melakukan penelitian, ingin membandingkan efektivitas sistem terasering dalam mereduksi laju erosi. Mahasiswa tersebut kemudian melakukan pengukuran besaran erosi pada 15 unit lahan pertanian tanpa sistem konservasi, dan lahan pertanian yang menggunakan sistem konservasi terasering. Data yang dihasilkan sebagai berikut:

Kode Unit Lahan	Lahan Tanpa Sistem Konservasi	Lahan Dengan Sistem Terasering
1	186.00	150.00
2	176.00	152.00
3	168.00	155.00
4	189.00	140.00
5	190.00	132.00
6	193.00	133.00
7	185.00	135.00
8	187.00	138.00
9	186.00	144.00
10	180.00	148.00
11	191.00	147.00
12	192.00	145.00
13	191.00	148.00
14	184.00	149.00
15	188.00	149.00

- Hipotesis yang disusun adalah hipotesis dua arah, yaitu:  
 $H_0 = \text{Besaran erosi lahan pertanian tanpa sistem konservasi} = \text{Besaran erosi lahan pertanian dengan sistem konservasi terasering},$   
 $H_1 = \text{Besaran erosi lahan pertanian tanpa sistem konservasi} \neq \text{Besaran erosi lahan pertanian dengan sistem konservasi terasering}.$
- Hasil perhitungannya dapat dilihat pada tabel berikut:

t-Test: Two-Sample Assuming Equal Variances

	<i>Lahan Tanpa Sistem Konservasi</i>	<i>Lahan Dengan Sistem Terasering</i>
Mean	185.73	144.33
Variance	44.92	50.67
Observations	15.00	15.00
Pooled Variance	47.80	
Hypothesized Mean Differer	0.00	
df	28.00	
t Stat	16.40	
P(T<=t) one-tail	0.00	
t Critical one-tail	1.70	
P(T<=t) two-tail	0.00	
t Critical two-tail	2.048	

- Kesimpulan yang dapat ditarik adalah : t stat (16.4) > t tabel (2.048), yang berarti  $H_0$  ditolak dan  $H_1$  diterima, sehingga disimpulkan bahwa : Besaran erosi lahan tanpa sistem konservasi  $\neq$  Besaran erosi lahan pertanian dengan sistem konsrvasi terasering. Atau dengan kata lain, Penggunaan sistem terasering mampu mereduksi laju erosi di lahan pertanian secara signifikan.

Pada Paired Sample T Test digunakan uji beda mean untuk satu sampel yang diberikan perlakuan yang berbeda. Jumlah sampel harus sama, dan pengujiannya kuga sama dengan sebelumnya untuk melihat perbedaan mean dari sampel tersebut sebelum dan sesudah diberi perlakuan dan manakah yang lebih tinggi/rendah apakah sampel yang sebelum /sesudah diberi perlakuan.

### Contoh

Seorang mahasiswa dalam penelitiannya ingin mengetahui apakah ada peningkatan rata-rata antara nilai matematika sebelum diadakan les dengan

sesudah diadakan les matematika. Sampel yang diambil sebanyak 20 siswa, yaitu :

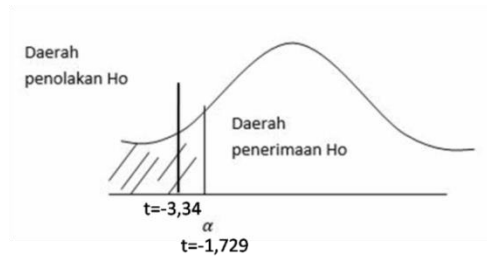
No.	Sebelum Les	Sesudah Les
1	6,34	6,92
2	6,58	7,25
3	5,38	6,45
4	5,6	7,5
5	6,68	7,25
6	7,42	7,27
7	7,2	7,86
8	6,24	6,9
9	5,78	5,47
10	5,47	6,98
11	6,53	7,28
12	6,83	6,28
13	7,75	7,8
14	5,28	6,68
15	7,83	6,93
16	6,77	6,97
17	5,37	6,28
18	6,57	6,78
19	4,82	6,73
20	6,85	7,14

- Perhitungan nilai t :  
dengan nilai korelasi  $r=0,479$

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} - 2r\left(\frac{s_1}{\sqrt{n_1}}\right)\left(\frac{s_2}{\sqrt{n_2}}\right)}} = \frac{6,34 - 6,94}{\sqrt{\frac{0,86^2}{20} + \frac{0,55^2}{20} - 2(0,479)\left(\frac{0,86}{\sqrt{20}}\right)\left(\frac{0,55}{\sqrt{20}}\right)}} = -3,34$$

- Hipotesis Uji :  
Ho : Tidak terjadi peningkatan rata-rata nilai matematika siswa sebelum dengan sesudah diadakan les  
Ha : Terjadi peningkatan rata-rata nilai matematika siswa sebelum dengan sesudah diadakan les
- Hipotesis Statistik :  
Ho :  $M1 \leq M2$   
Ha :  $M1 > M2$
- Uji Statistik :  
Pilih nilai sigfinikansi alpha 5% dan uji satu arah dengan nilai derajat bebas  $df=20-1=19$  diperoleh  $t\text{-tabel}=1,729$ .

- Keputusan :



Nilai t-hitung = -3,34 berarti uji pihak kiri sehingga  
 $t\text{-hitung} (-3,34) < t\text{-tabel} (-1,729)$   
 $H_0$  ditolak dan  $H_a$  diterima.

- Kesimpulan :  
 Terjadi peningkatan rata-rata nilai matematika siswa sebelum dengan sesudah diadakan les.

## b) Cara menganalisis dan interpretasi SPSS

### Contoh Kasus

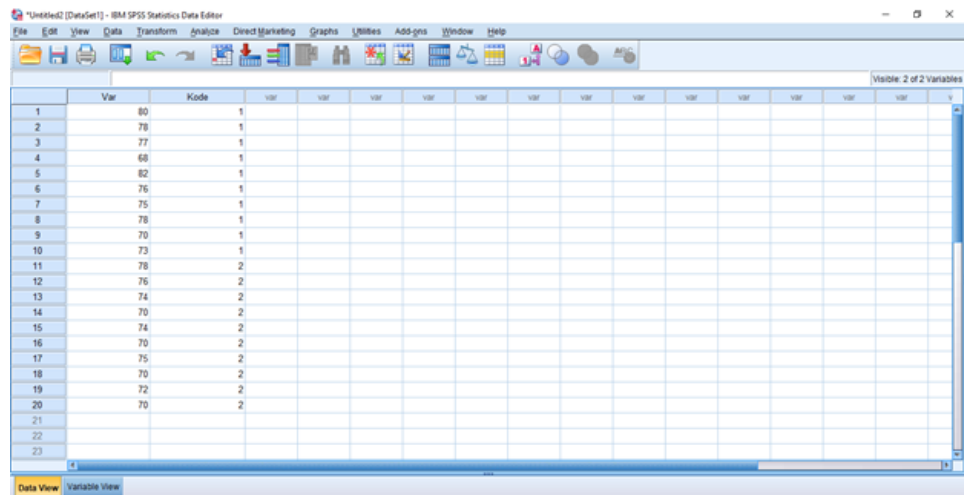
Seorang guru SMA Mercuri Buana ingin meneliti pengaruh les tambahan di sekolah terhadap prestasi belajar siswanya untuk mata pelajaran matematika. Dari 20 siswa akan dibagi menjadi 2 kelompok, yaitu mengikuti les tambahan (LT) dan tidak mengikuti les tambahan (TLT). Setelah selang beberapa bulan diadakan tes prestasi belajar matematika dan berikut hasil belajarnya :

LT	TLT
80	78
78	76
77	74
68	70
82	74
76	70
75	75
78	70
70	72
73	70

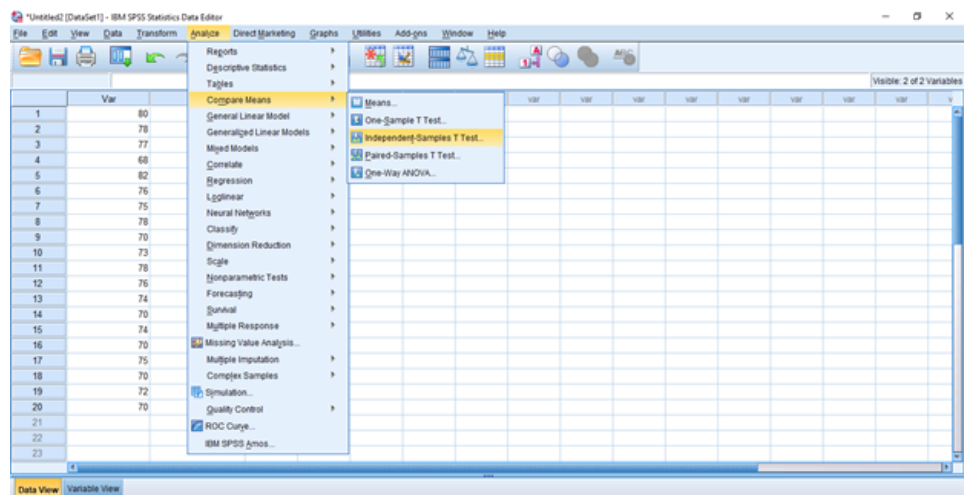
- **Hipotesis:**  
 $H_0$ : Tidak ada pengaruh les tambahan terhadap prestasi belajar siswa  
 $H_a$ : Ada pengaruh les tambahan terhadap prestasi belajar siswa



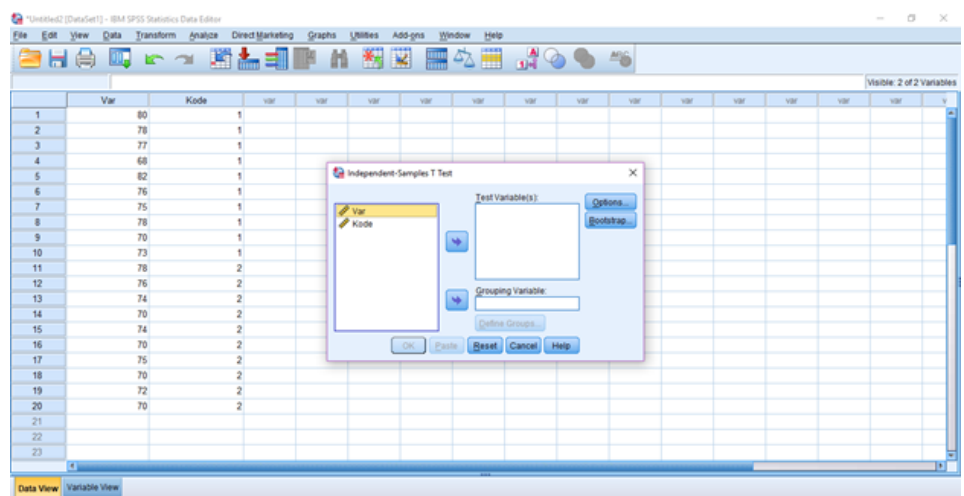
- Input data ke dalam Data View SPSS



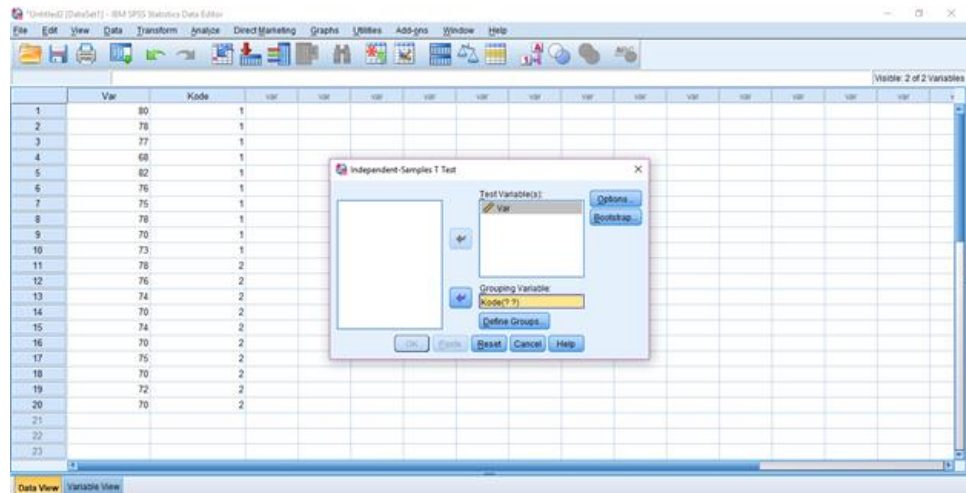
- Setelah itu, klik Analyze - Compare Means - Independent-Sample T Test.



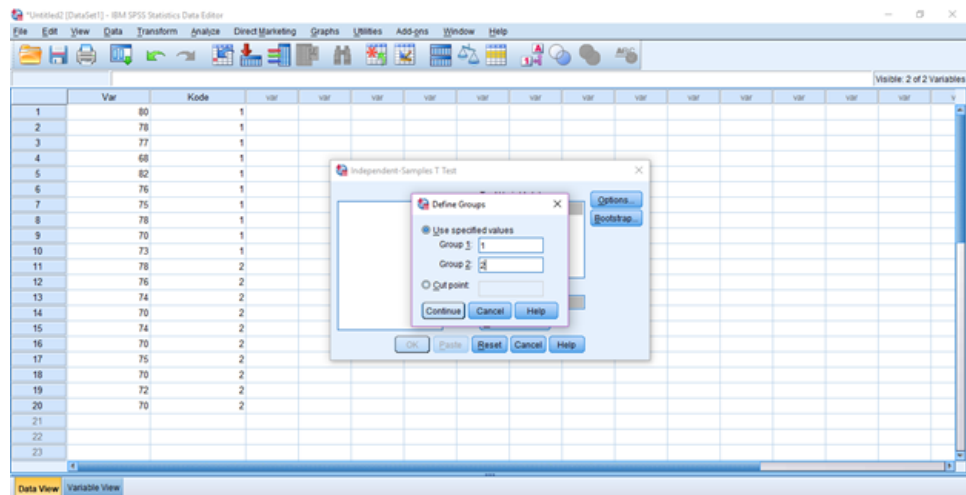
- Akan muncul kotak dialog seperti di bawah ini



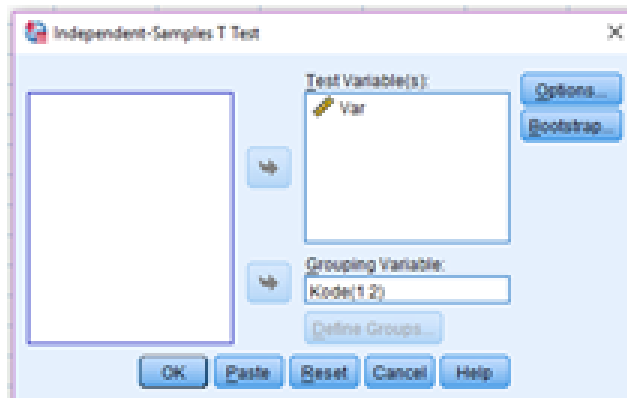
- Selanjutnya masukkan Var ke Test Variable(s) dan masukkan Kode ke Grouping Variable.



- Pada Define Groups akan muncul kotak dialog seperti ini, kemudian isikan pada Group 1 dengan 1 dan Group 2 dengan 2. Klik Continue



- Kemudian klik OK



- Diperoleh hasil output dengan SPSS sebagai berikut

Group Statistics					
	Kode	N	Mean	Std. Deviation	Std. Error Mean
Var.	LT	10	75.70	4.347	1.375
	TLT	10	72.90	2.923	.924

Independent Samples Test										
		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower		Upper
Var.	Equal variances assumed	.930	.348	1.690	18	.108	2.800	1.657	-680	6.280
	Equal			1.690	15.757	.111	2.800	1.657	-716	6.316

- **Output**

Hasil uji beda Independent-Sample t Test. Ada dua nilai t dan taraf signifikansinya yaitu Equal Variance assumed dan Equal Variance not assumed.

- Equal Variance assumed berarti kedua varian populasi adalah identik dan Equal Variance not assumed bahwa kedua varian populasi tidak identik. Untuk menentukannya digunakan uji F.
- Taraf sig. F adalah 0,930 artinya Ho diterima bahwa kedua varian populasi identik (Equal Variance assumed). Karena sig F mempunyai keputusan Equal Variance assumed maka t test menggunakan Equal Variance assumed.
- Nilai t Equal Variance assumed sebesar 1.690 dengan Sig. (2-tailed) adalah 0.108 atau lebih besar dari 0.05 sehingga diputuskan tidak ada perbedaan mengikuti les tambahan dengan tidak mengikuti les tambahan terhadap prestasi belajar mata pelajaran matematika siswa Mercuru buana.

## B. UJI BEDA UNTUK UJI NON PARAMETRIK

### 1. Uji Beda Berpasangan

#### a) Uji Mc Nemar

Uji Mc Nemar diperkenalkan oleh seorang ahli psikologi bernama Quinn Mc Nemar pada tahun 1947. Uji Mc Nemar digunakan untuk menentukan perubahan dalam proporsi bagi sampel-sampel yang berhubungan. Uji McNemar digunakan untuk penelitian yang membandingkan sebelum dan sesudah peristiwa dimana tiap objek digunakan pengontrol dirinya sendiri. Uji ini dilakukan pada 2 sampel yang berhubungan, skala pengukurannya berjenis nominal ataupun ordinal (binary respons) dan untuk crosstabulasi data 2 x 2.

Dengan uji Mc Nemar, dapat dianalisis perubahan untuk mengetahui apakah perubahan dalam kedua arah berkemungkinan sama atau tidak. Distribusi yang digunakan untuk menguji perubahan dalam respon tersebut digunakan distribusi chi-square sebagai panduan untuk menguji signifikansi setiap perubahan maka data perlu disusun ke dalam tabel segi empat ABCD

SEBELUM PERLAKUAN	SESUDAH PERLAKUAN	
	-	+
+	A	B
-	C	D

a: jika dia berubah dari positif ke negatif

d: jika dia berubah dari “negatif” ke “positif”.

b: tetap positif

c: tetap negative

- Karena  $a+d$  menunjukkan jumlah total individu yang berubah, maka harapan di bawah  $H_0$  adalah:
  - ✓  $\frac{1}{2} (a+d)$   $\bar{U}$  berubah dalam satu arah, dan
  - ✓  $\frac{1}{2} (a+d)$   $\bar{U}$  berubah dalam arah lain
  - ✓ Dengan kata lain  $\frac{1}{2} (a+d)$   $\bar{\delta}$  adalah frekuensi yang diharapkan di bawah  $H_0$ .
- Jika  $a$  = banyak kasus yang diobservasi dalam “sel a”

d = banyak kasus yang diobservasi dalam “sel d”

- $\frac{1}{2}(a+d)$  = banyak kasus yg diharapkan baik di “sel a” maupun di “sel d”,

### 1) Prosedur melakukan analisis :

- Letakkan frekuensi observasi dalam sel-sel tabel kontigensi yang telah disusun.
- Tentukan frekuensi yang diharapkan dalam sel A dan D.
- Frekuensi yang diharapkan ditentukan dengan:
- **$E = \frac{1}{2} ( A + D )$**
- Apabila frekuensi yang diharapkan (E) kurang dari 5 maka lakukan analisis dengan tes yang lain misalnya; tes binomial. Tetapi jika frekuensi yang diharapkan  $> 5$  maka lakukan analisis dan tentukan harga  $X^2$  nya.

$$X^2 = \frac{(|A - D| - 1)^2}{(A + D)}$$

- Tentukan titik kritis yang terletak pada tabel Chi Square dengan db = 1 pada alpha tertentu.
- Bandingkan harga  $X^2$  hasil analisis dengan  $X^2$  tabel atau titik kritisnya untuk memutuskan apakah  $H_0$  ditolak atau diterima.
- Kemukakan kesimpulan dari hasil analisis tersebut.  
Jika  $x_{hitung} < x_{tabel} = H_0$  diterima  
Jika  $x_{hitung} > x_{tabel} = H_0$  ditolak

### 2) Fungsi Uji Mc.Nemar

- Untuk uji signifikansi perubahan atau pergeseran yang dijelaskan melalui rancangan sebelum dan sesudah. Dengan demikian pengamatan dilakukan sebanyak 2 kali yaitu sebelum adanya perlakuan dan sesudahnya. Dalam hal ini individu juga berposisi sebagai pengontrol bagi dirinya sendiri.
- Dapat digunakan untuk uji efektivitas suatu perlakuan (*treatment*). Artinya jika perubahan terjadi maka perubahan itu terjadi semata hanya karena adanya perlakuan dimaksud.

### Contoh Kasus

- Diambil sampel 22 orang laki-laki dewasa. Akan dilakukan pengukuran terhadap tekanan nadi. Mereka diminta untuk melakukan olahraga ringan.
  - Kode 1 = tekanan nadi normal = 70- 80x/menit
  - Kode 2 = tekanan nadi tidak normal (>80).

Data di ambil sebelum dan sesudah dilakukan olahraga. Ingin diketahui apakah terdapat perbedaan atau perubahan tekanan nadi setelah dilakukan olahraga Bagaimana keputusan hipotesisnya jika menggunakan derajat kepercayaan 95 % dan derajat signifikansi 5 %

Data sebagai berikut :

Sampai ke	Pre exc	Post exc
1	1	2
2	2	2
3	1	2
4	2	2
5	1	2
6	1	2
7	1	2
8	2	1
9	2	2
10	1	2
11	1	2
12	1	1
13	1	2
14	2	1
15	2	2
16	1	1
17	1	2
18	1	2
19	1	2
20	2	2
21	1	2
22	1	2

### Pembahasan

- Penentuan hipotesis nol dan hipotesis alternative  
 $H_0$  : tidak ada perbedaan tentang tekanan nadi sebelum dan sesudah olahraga.

$H_1$  :ada perbedaan tentang tekanan nadi sebelum dan sesudah olahraga.

- Menentukan statistik uji = Uji Mc Nemar
- Tingkat signifikanm = (5%, n = 22)
- Distribusi sampling

Pre exc	Post exc		Total
	-	+	
+	13	2	15
-	5	2	7
Total	18	4	22

Keterangan:

+ = tekanan adi normal

- = tekanan nadi tidak normal

- Menentukan frekuensi yang diharapkan

$$E = \frac{1}{2} (A + D)$$

$$E = \frac{1}{2} (13 + 2) = 7,5$$

- Menentukan  $X^2$  menggunakan rumus Mc.Nemar

$$X^2 = \frac{(|A - D| - 1)^2}{(A + D)}$$

$$X^2 = \frac{(|13 - 2| - 1)^2}{(13 + 2)}$$

$$X^2 = \frac{100}{15} = 6,67$$

- Bandingkan nilai  $x^2$  hitung diatas nilai  $x^2$  pada tabel nilai chi square.  
Dengan  $cl= 95\%$ ,  $\alpha = 5\%$  (0,05) dan  $df = 1$ . Maka diperoleh nilai x tabel = 3,84
- Menentukan keputusan tolak  $H_0$   
Karena  $x^2$  hitung < nilai  $x^2$ tabel  $8,06 > 3,84$  maka  $H_0$  ditolak.
- Kesimpulan = ada perbedaan tentang tekanan nadi sebelum dan sesudah olahraga.

### Contoh Kasus

Penelitian diadakan oleh suatu perusahaan operator kartu seluler untuk mengetahui apakah kebijakan baru yang dijalankan oleh perusahaan tersebut berpengaruh terhadap perilaku konsumen dalam membeli produknya. Kebijakan tersebut adalah menaikkan harga voucher pulsa, akan tetapi dibarengi dengan pemberian bonus berupa sms gratis setiap hari. Terdapat sebanyak 300 pemilik handphone yang diamati. Sebanyak 80 konsumen berhenti berlangganan, 150 konsumen tetap berlangganan, 60 konsumen menjadi pelanggan baru, sedangkan 10 pemilik handphone tetap menggunakan kartu prabayarnya yang lama (tidak beralih menjadi pelanggan baru).

### Pembahasan

- Penentuan hipotesis nol dan hipotesis alternative  
 $H_0$  : Perilaku konsumen tidak berubah walaupun diberlakukan kebijakan baru  
 $H_1$  : Perilaku konsumen berubah akibat diberlakukannya kebijakan baru

- Distribusi silang

Kebijakan lama	Kebijakan baru		Total
	Tidak berlangganan	Berlangganan	
Berlangganan	80	150	230
Tidak Berlangganan	10	60	70
Total	90	210	300

- Menentukan frekuensi yang diharapkan

$$E = \frac{1}{2}(A + D)$$

$$E = \frac{1}{2}(80 + 60) = 70$$

- Menentukan  $X^2$  menggunakan rumus Mc.Nemar

$$X^2 = \frac{(|A - D| - 1)^2}{(A + D)}$$



$$X^2 = \frac{(|80 - 60| - 1)^2}{(80 + 60)}$$

$$X^2 = \frac{361}{140} = 2,58$$

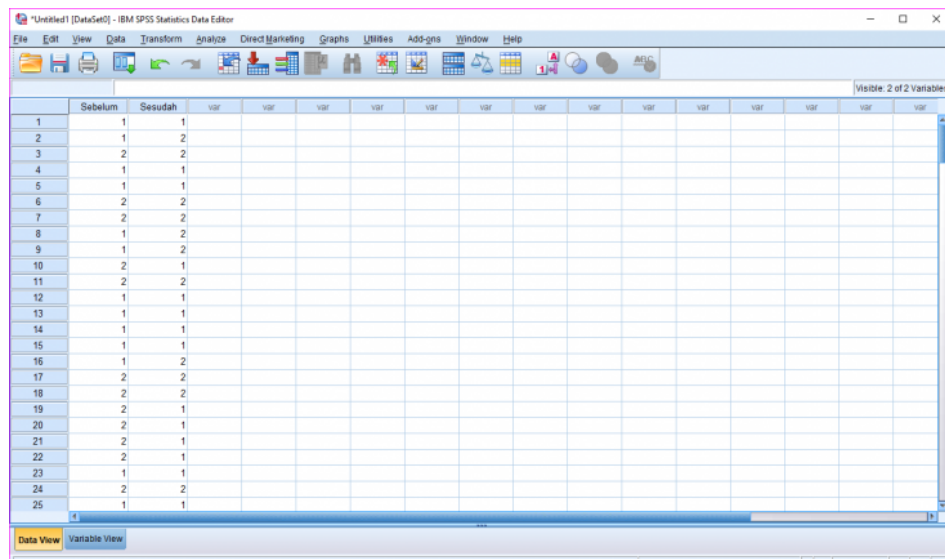
- Bandingkan nilai  $x^2$  hitung diatas nilai  $x^2$  pada tabel nilai chi square.

$$X^2 \text{ tabel} = 3,84$$

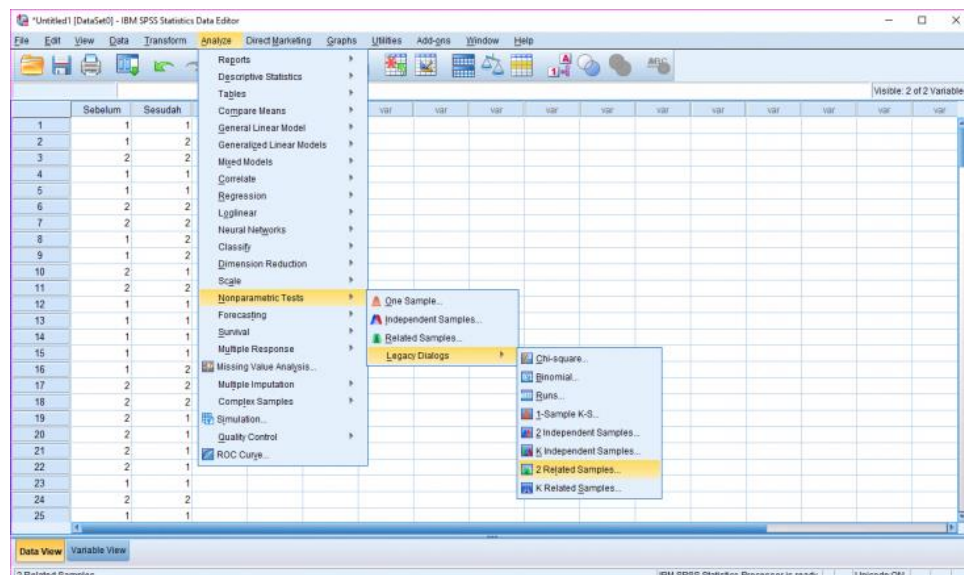
- Menentukan keputusan tolak  $H_0$ (Karena  $X^2$  hitung <  $X^2$  table,  $H_0$  diterima)
- Kesimpulan = Perilaku konsumen tidak berubah walaupun diberlakukan kebijakan baru. Walaupun diberlakukan kebijakan baru oleh perusahaan, perilaku konsumen akan tetap sama. Dengan demikian perusahaan dapat dengan aman menjalankan kebijakan baru tersebut tanpa harus kuatir ditinggalkan konsumen

### 3) Cara Analisis dan Interpretasi SPSS

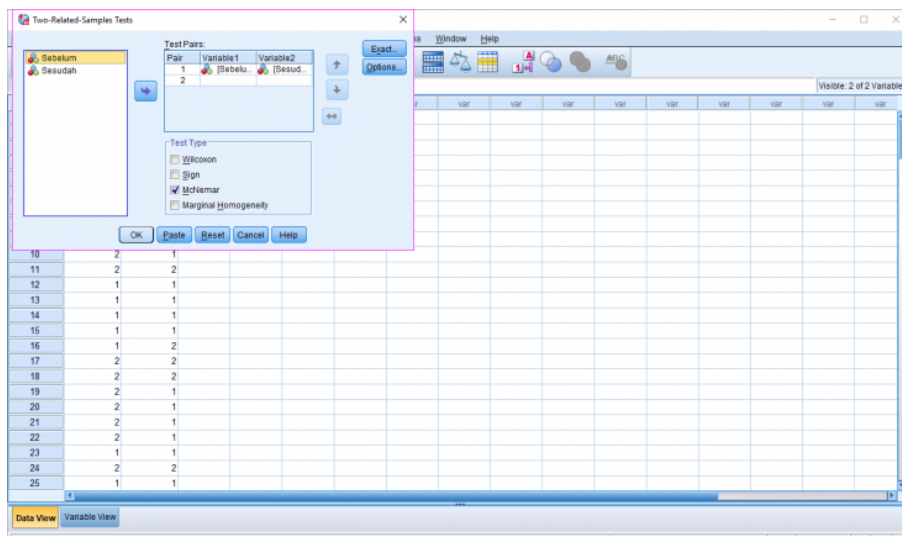
- **Input data ke dalam spss.**



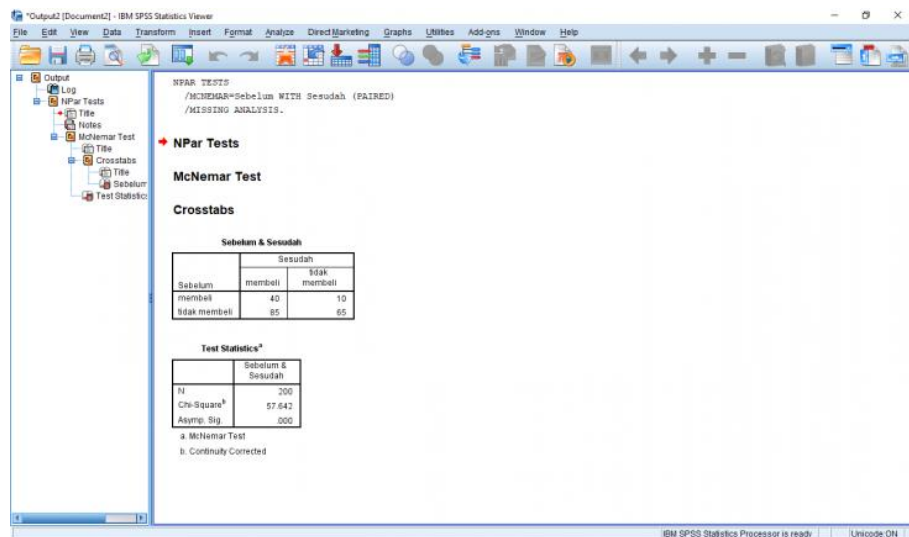
- Setelah itu memilih menu **Analyze** ⇒ **Nonparametric Tests** ⇒ **Legacy Dialogs** ⇒ **2 Related Sample**



- masukkan variabel sebelum kedalam kolom variable 1 dan variabel sesudah kedalam kolom variable 2, centang **Mc.Nemar** pada **Test Type** lalu klik **OK**



- **Output**



- Hasil dari pengujian menunjukkan bahwa nilai  $\chi^2_{obs} = 57,642$  dan  $p\text{-value}$  sebesar 0,000 dimana nilai ini lebih kecil dari nilai  $\alpha = 0,05$ , maka keputusannya adalah tolak  $H_0$ .

- **Kesimpulan**

- Uji Mc Nemar merupakan salah satu metode statistik uji non parametrik yang bisa digunakan apabila statistik uji parametrik tidak dapat digunakan.

## b) Uji Chocran

Uji Cochran Q diperkenalkan oleh William Gemmell Cochran (1909 – 1980). Uji Cochran Q digunakan untuk data yang bersifat nominal dengan k sampel yang bersifat dependent. Sebelum melangkah ke uji Cochran, kita kenal uji McNemar yang digunakan untuk menguji signifikansi dari perbedaan antara dua proporsi sampel yang saling berhubungan (data nominal dengan 2 sampel yang bersifat dependent). Cochran mengusulkan uji dimana sampel yang digunakan lebih dari 2 atau k sampel. Berikut ini adalah penjelasan uji Cochran Q

Uji Cochran digunakan untuk menganalisis secara statistik tingkat keberhasilan dari suatu data tertentu. Hipotesis yang diuji adalah hipotesis

terhadap beberapa variabel dikotomi yang memiliki arti sama. Variabel yang diukur berasal dari individu yang sama atau pada individu yang cocok/sesuai.

Uji ini menggunakan data nominal dengan sampel lebih dari dua dan data bersifat dependent. Data yang digunakan di dalam uji berbentuk binary, yaitu 1 untuk sukses dan 0 untuk gagal. Masing-masing dari perlakuan k secara independen diaplikasikan ke blok b dan masing-masing hasil diukur sebagai sukses (1) dan gagal (0)

	Treatment 1	Treatment 2	...	Treatment k
Block 1	X11	X12	...	X1k
Block 2	X21	X22	...	X2k
⋮	⋮	⋮	...	⋮
Block b	Xb1	Xb2	...	Xbk

▪ **Hipotesis:**

Ho: Treatments memiliki efektivitas yang sama

H1: Treatments memiliki perbedaan efektivitas

1) **Statistik Uji:**

$$T = k(k - 1) \sum_{j=1}^k \left( X_{.j} - \frac{N}{k} \right)^2 / \sum_{i=1}^b X_{i.} (k - X_{i.})$$

K : jumlah treatment

X.j : total kolom sampai treatment ke j

B : jumlah blok

Xi : total baris sampai blok ke i

N : total keseluruhan

Untuk tingkat signifikansi  $\alpha$ , daerah tolak berada pada

$$T > X^2_{1-\alpha, k-1}$$

Dimana  $X^2_{1-\alpha, k-1}$  adalah  $(1-\alpha)$  kuantil dari distribusi chi-square dengan derajat kebebasan k-1.

Selain rumus di atas, rumus di bawah ini adalah sama, tetapi lebih mudah untuk penghitungannya:

$$Q = \frac{(k - 1) \left[ k \sum_{j=1}^k G_j^2 - \left( \sum_{j=1}^k G_j \right)^2 \right]}{k \sum_{i=1}^N L_i - \sum_{i=1}^N L_i^2}$$

**Keterangan:**

G<sub>j</sub>: jumlah keseluruhan sukses dalam kolom ke-j

L<sub>i</sub>: jumlah keseluruhan sukses dalam baris ke-i

**2) Keterbatasan Uji Cochran Q**

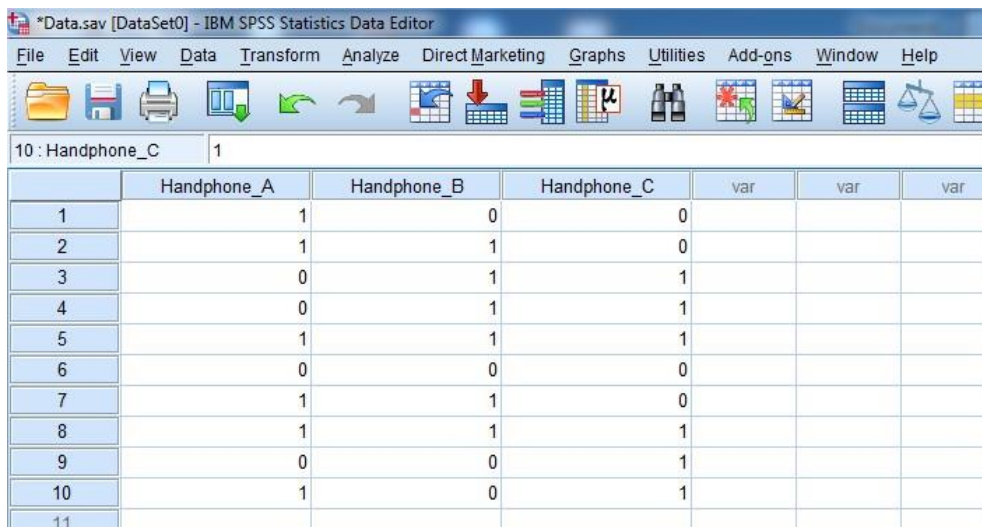
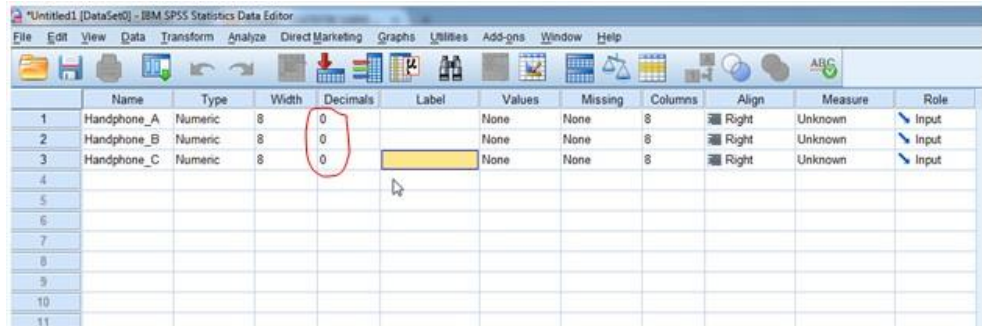
- a) Uji Cochran Q hanya menentukan terjadinya suatu perubahan tetapi mengevaluasi sejauh mana perubahan tersebut. Memungkinkan untuk melakukan beberapa uji McNemar, tetapi tidak ada efek interaksi yang dapat diukur.
- b) Uji ini dikenal lemah dalam mendeteksi true heterogeneity diantara studi yang dianggap signifikan. Meta-analisis seringkali hanya menyertakan sejumlah kecil studi, dan kekuatan uji dalam keadaan seperti itu rendah. Karena tes ini lemah dalam mendeteksi true heterogeneity, hasil non-signifikan tidak dapat diambil sebagai bukti homogenitas.
- c) Uji ini tidak mengakomodasi control terhadap kelompok, karena uji Cochran menggunakan observasi yang dependen.

**3) Cara Analisis dan Interpretasi SPSS**

Sebuah perusahaan IT memproduksi tiga buah handphone terbarunya, handphone tersebut sebelum dipasarkan harus dinilai kinerjanya. Untuk mendapat nilai kinerja setiap handphone maka digunakan pekerja sebanyak 10 orang untuk menilai ketiga handphone tersebut. Berikut data yang telah dikumpulkan.

Designer	Handphone A	Handphone B	Handphone C
A	1	0	0
B	1	1	0
C	0	1	1
D	0	1	1
E	1	1	1
F	0	0	0
G	1	1	0
H	1	1	1
I	0	0	1
J	1	0	1

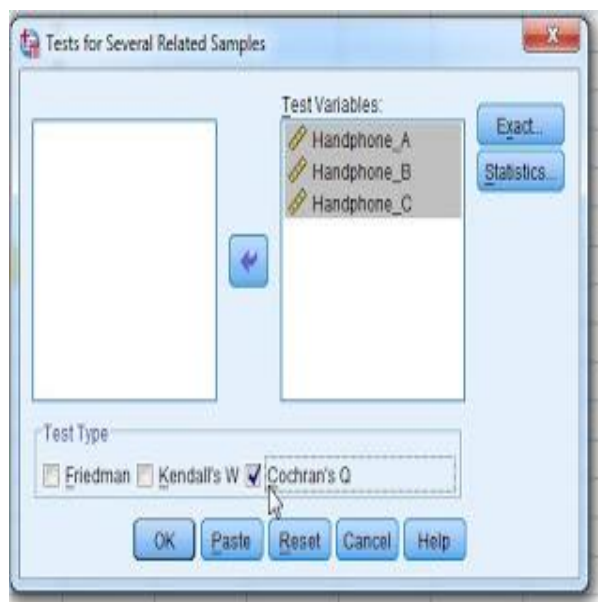
- Masukan data yang hendak di analisis di SPSS



- Pada toolbar paling atas, silahkan klik Analyze---> Nonparametric Tests---> Legacy Dialog---> k Related Samples... jika sudah terlihat seperti gambar berikut:



- Pindahkan Handphone\_A, Handphone\_B, Handphone\_C kedalam kota Test Variables:, lalu pada Test Type centang Cochran's Q, karena kita menggunakan uji Cochran's. Jika sudah di isi seperti gambar berikut:



- Output

**NPar Tests**

**Cochran Test**

**Frequencies**

	Value	
	0	1
Handphone_A	4	6
Handphone_B	4	6
Handphone_C	4	6

**Test Statistics**

N	10
Cochran's Q	.000 <sup>a</sup>
df	2
Asymp. Sig.	1.000

a. 1 is treated as a success.

- Analisis hasil  
Hipotesis yang digunakan dalam kasus ini adalah:  
 $H_0$  = Tidak ada perbedaan kinerja handphone A, B, dan C

H1 = Terdapat perbedaan kinerja handphone A, B, dan C

- Pengujian statistik dalam pengambilan keputusan:

Jika Q hitung < Q tabel dan nilai sig > 0.05 maka terima H0

Jika Q hitung > Q tabel dan nilai sig < 0.05 maka tolak H0

- Note: Q = Chi-square

Berdasarkan hasil yang diperoleh pada gambar Test Statistics, banyak anggota sampel (N=10) dan nilai Cochran's Q atau Q hitung 0.000, deviasi of freedom (df) 2, dan Asymp. Sig. 1.000. Sehingga dapat kita bandingkan Q hitung 0.000 dan Q tabel 0.1026 yaitu  $0.000 < 0.1026$  dan  $1.000 > 0.005$ . Jadi terima H0.

- Kesimpulan dari kasus perusahaan di atas adalah: Tidak ada perbedaan kinerja handphone A, handphone B, dan handphone C. atau dengan kata lain adalah sama.

#### e) Uji Wilcoxon

Uji tanda memanfaatkan hanya tanda-tanda 'plus' dan 'minus' yang diperoleh dari selisih antara nilai pengamatan dan median pembanding, tetapi mengabaikan besarnya selisih-selisih tersebut. Frank Wilcoxon (1945) memperkenalkan satu prosedur nonparametrik untuk menguji median yang memanfaatkan baik arah (tanda 'plus' dan 'minus') maupun besar arah itu. Uji ini dikenal dengan istilah uji peringkat bertanda Wilcoxon (*Wilcoxon signed-rank test*). Uji ini digunakan untuk menguji dua kelompok sampel terkait prosedur Non Parametrik

Uji ini digunakan untuk menguji kondisi (variabel) pada sampel yang berpasangan atau dapat juga untuk penelitian sebelum dan sesudah. Dalam uji ini ingin diketahui manakah yang lebih besar dari antara pasangan. Cara ini sekarang dinamakan uji Wilcoxon atau Uji Ranking Bertanda Wilcoxon. Merupakan penyempurnaan dari uji tanda. Uji Wilcoxon ini hampir sama dengan Uji Tanda tetapi besarnya selisih nilai angka antara positif dan negatif diperhitungkan, dan digunakan untuk menguji hipotesis komparatif 2 sampel berpasangan. Uji wilcoxon lebih peka daripada uji tanda dalam menentukan perbedaan antara rata-rata populasi dan karena itu akan dibahas secara mendalam. Jika sampel berpasangan lebih besar dari 25, maka



distribusinya dianggap akan mendekati distribusi normal. Untuk itu digunakan Z sebagai Uji Statistiknya.

Uji wilcoxon digunakan untuk menganalisis hasil-hasil pengamatan yang berpasangan dari dua data apakah berbeda atau tidak. Wilcoxon signed Rank test ini digunakan hanya untuk data bertipe interval atau ratio, namun datanya tidak mengikuti distribusi normal.

### 1) Uji hipotesis

$H_0 : d = 0$  (tidak ada perbedaan diantara dua perlakuan yang diberikan)

$H_1 : d \neq 0$  (ada perbedaan diantara dua perlakuan yang diberikan)

### 2) Asumsi-asumsi

- Menggunakan data berpasangan dan berasal dari populasi yang sama. ini sama dengan tujuan dari uji t berpasangan.
- Setiap pasangan dipilih secara acak dan independent. Maksudnya ini dalam pengambilan sampel tidak subjektif atau asal ambil. tapi pengambilan sampelnya secara acak.
- Skala pengukurannya minimal ordinal. dan tidak butuh asumsi normalitas. Inilah yang membedakan dengan uji t berpasangan. disini ada dua keadaan dalam menggunakan wilcoxon. Pertama, ketika data yang digunakan ordinal maka pakai wilcoxon. kasus kedua ketika datanya tuh interval atau rasio maka pertama kali lihat dulu apakah normal atau tidak. kalau normal pakai uji t berpasangan dan jika tidak normal baru pakai wilcoxon. untuk uji normalnya bisa lihat disini. uji normalitas. Beberapa peneliti juga mengatakan ketika data yang digunakan lebih dari 25, ada juga yang mengatakan lebih dari 30. maka pakai uji t berpasangan. alasannya dengan data yang 30 (dikatakan sampel besar) itu akan mendekati data normal. Jadi silahkan pilih dengan bijak.

### 3) Langkah- Langkah Pengujian

- Berikan jenjang (rank) untuk tiap beda dari pasangan pengamatan ( $y_i - x_i$ ) sesuai dengan besarnya, dari yang terkecil sampai terbesar tanpa memperhatikan tanda dari beda itu (nilai beda absolut).
- Bila ada dua atau lebih beda yang sama, maka jenjang untuk tiap-tiap beda itu adalah jenjang rata-rata

- Bubuhkan tanda positif atau negatif pada jenjang untuk tiap beda sesuai dengan tanda dari beda itu. Beda 0 tidak diperhatikan
- Jumlahkan semua jenjang bertanda positif atau negatif, tergantung dari mana yang memberikan jumlah yang lebih kecil setelah tandanya dihilangkan. Notasi jumlah jenjang yang lebih kecil ini dengan T
- Bandingkan nilai T yang diperoleh dengan nilai t uji wilcoxon

**4) Formula untuk data tidak memiliki nilai duplikat (no ties)**

$$Z = \frac{T - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}}$$

**Formula untuk data memiliki nilai duplikat (with ties)**

$$Z = \frac{T - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24} - \frac{\sum t^3 - \sum t}{48}}}$$

• **Contoh :**

Perusahaan garmen “MAJU” ingin mengukur peningkatan prestasi kerja karyawan diperusahaan setelah diberikan pelatihan. Untuk itu diambil sampel sebanyak 10 karyawan, datanya adalah sbb:

Karyawan	Nilai		Selisih X <sub>2</sub> - X <sub>1</sub>	Tanda	Peringkat
	Sebelum (X <sub>1</sub> )	Sesudah (X <sub>2</sub> )			
1	72	76	4	+	3.5
2	67	90	23	+	7
3	71	75	4	+	3.5
4	86	86	0		diabaikan
5	83	83	0		diabaikan
6	60	88	28	+	8
7	91	88	-3	-	2
8	70	82	12	+	6
9	69	67	-2	-	1
10	80	72	-8	-	5

### Nilai T-hitung

$$T(+) = 3.5 + 7 + 3.5 + 8 + 6 = 28$$

$$T(-) = 2 + 1 + 5 = 8$$

$$T' \text{ hitung} = \min(T+, T-) = 8$$

Karena ada 2 data yang nilainya nol, maka  $n = 8$

Tabel Wilcoxon  $n=8, \alpha=0.05$  adalah  $T\text{-Tabel}=3$

### Perhitungan Pendekatan Normal (with ties)

$$Z = \frac{8 - \frac{8(8+1)}{4}}{\sqrt{\frac{8(8+1)(2 \times 8 + 1)}{24} - \frac{6}{48}}} = -1.402$$

### 5) Cara Analisis dan Interpretasi SPSS

Seorang dokter ingin mengetahui pengaruh sebelum dan sesudah pemberian obat terhadap kenaikan berat badan pasien (berat badan dalam satuan kg). Berikut data yang diperoleh

No	Sebelum	Sesudah
1	60	62
2	56	60
3	70	70
4	68	67
5	65	66
6	54	57
7	55	54
8	66	68
9	76	77
10	78	77
11	45	45
12	56	57
13	65	68
14	64	65
15	67	66
16	72	74
17	71	75
18	74	76

Dari data di atas, pasien pertama berat badan sebelum minum obat 60 dan sesudah minum obat 62, dan seterusnya untuk pasien yang lain. Disini datanya sedikit dan kita anggap saja tidak diketahui distribusi dari data, oleh karenanya kita dapat menggunakan uji non-parametrik terhadap dua sampel yang berpasangan dari data di atas.

Pada kasus ini kita tidak perlu melihat perbedaan berat badan, Karena perbedaan di hitung dengan nilai  $(Y - X)$ , dimana  $Y =$  Sesudah, sedangkan  $X =$  Sebelum. Tetapi kita fokus pada naik atau turunnya berat badan secara signifikan setelah minum obat.

- Hipotesis

$H_0$ : Tidak terdapat kenaikan berat badan pasien yang bermakna antara sebelum dan sesudah minum obat

$H_1$ : Terdapat kenaikan berat badan pasien yang bermakna antara sebelum dan sesudah minum obat

- Pengujian Statistik:

Jika  $Z_{hitung} < Z_{tabel}$  atau nilai  $sig > 0.05$ , maka  $H_0$  diterima

Jika  $Z_{hitung} > Z_{tabel}$  atau nilai  $sig < 0.05$ , maka  $H_0$  di tolak

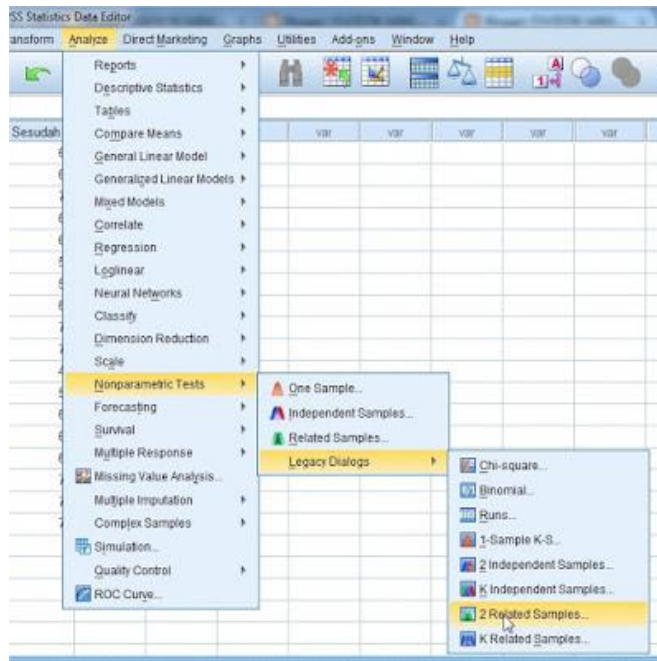
- Isi variabel view seperti gambar berikut:

Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role	
1	Sebelum	Numeric	8	0	Sebelum minum obat	None	None	8	Right	Unknown	Input
2	Sesudah	Numeric	8	0	Sesudah minum obat	None	None	8	Right	Unknown	Input

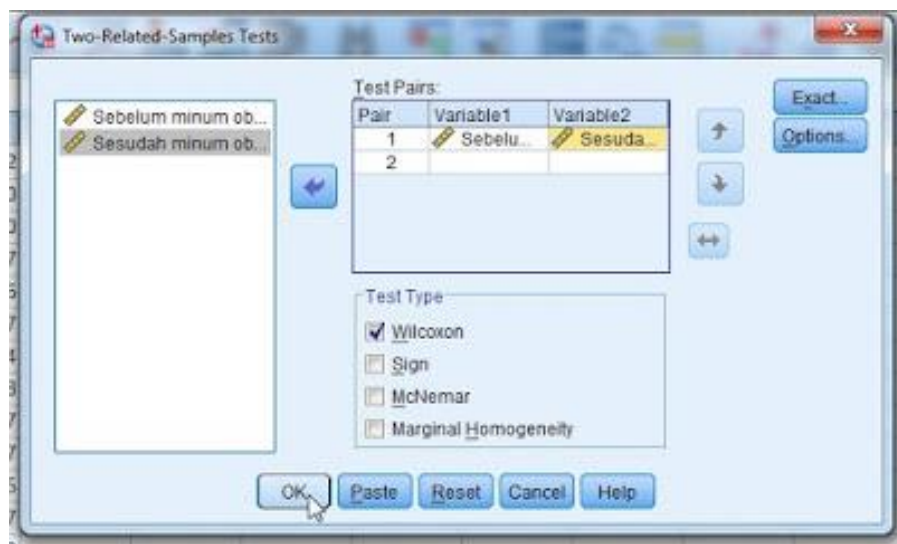
- Selanjutnya isikan data

	Sebelum	Sesudah	var
1	60	62	
2	56	60	
3	70	70	
4	68	67	
5	65	66	
6	54	57	
7	55	54	
8	66	68	
9	76	77	
10	78	77	
11	45	45	
12	56	57	
13	65	68	
14	64	65	
15	67	66	
16	72	74	
17	71	75	
18	74	76	
19			

- Perhatikan pada menu bar paling atas silahkan pilih Analyze - Nonparametric Test - Legacy Dialog - 2 Related Samples....



- Pada kotak Test Pairs: masukkan Sebelum minum obat ke dalam kolom Variable 1 dan Sesudah minum obat masukkan ke dalam kolom Variable 2, lalu pada Test Type centang Wilcoxon. atau terlihat seperti pada gambar berikut:



- **Output**

**NPar Tests**

**Wilcoxon Signed Ranks Test**

**Ranks**

		N	Mean Rank	Sum of Ranks
Sesudah minum obat - Sebelum minum obat	Negative Ranks	4 <sup>a</sup>	4.50	18.00
	Positive Ranks	12 <sup>b</sup>	9.83	118.00
	Ties	2 <sup>c</sup>		
	Total	18		

a. Sesudah minum obat < Sebelum minum obat  
 b. Sesudah minum obat > Sebelum minum obat  
 c. Sesudah minum obat = Sebelum minum obat

**Test Statistics<sup>a</sup>**

	Sesudah minum obat - Sebelum minum obat
Z	-2.628 <sup>b</sup>
Asymp. Sig. (2-tailed)	.009

a. Wilcoxon Signed Ranks Test  
 b. Based on negative ranks.

- **Interpretasi hasil**

- \* Tabel Ranks

Pada tabel Ranks kita dapat mengetahui nilai selisih (nilai beda +/-). Misalnya pada data baris pertama pada soal nilainya  $60-62 = -2$ ,  $56-60 = -4$  dan seterusnya. Setelah di hitung nilai selisih kemudian di beri ranking dari nilai kecil sampai ke yang besar. Sehingga di dapat pada kolom (N di tabel Ranks) diperoleh ranks negative ada 4 data dan ranks positive ada 12 data ( $4+12=16$ ), sedangkan Ties itu dikarenakan ada data yang sama pada baris ke 9 dan 10 (nilai 77 pada soal). Jadi jumlah keseluruhan data adalah  $4+12+2=18$  data. Kemudian kita dapat mengetahui Mean Rank dan Sum of Rank dari banyaknya data negative Ranks dan Positive Ranks.

- \* Tabel Test Statistics

Pada tabel Test Statistics di dapat nilai Z hitung adalah -2.628, sedangkan nilai Z tabel diperoleh dari tabel Z dengan alpha 5% atau 0.05 nilainya sekitar -

1.645 (tanda negatif disesuaikan saja tergantung output dari Z hitung). Sedangkan pada nilai Asymp. Sig (2-tailed) diperoleh 0.09.

- **Kesimpulan:**

Karena Z hitung > Z tabel yaitu  $-2.628 > -1.645$  atau nilai sig  $0.009 < 0.05$  sesuai dengan pengujian statistik yang kita gunakan maka  $H_0$  di tolak, Jadi terimalah  $H_1$ . Hal ini berarti terdapat kenaikan berat badan pasien yang bermakna antara sebelum dan sesudah minum obat

#### f) Uji Friedman

Metode uji Friedman digunakan untuk menguji kemaknaan pengaruh berbagai ( k ) perlakuan terhadap sejumlah kelompok subjek penelitian yang berhubungan, dengan mengontrol variabel ketiga yang mungkin berpengaruh dalam memperkirakan efek yang sesungguhnya dari perlakuan itu. Metode ini di kembangkan oleh Friedman pada tahun 1937. Data yang akan dianalisis minimal diukur dalam skala ordinal.

Karena yang dianalisis tidak hanya efek perlakuan terhadap subjek, tetapi juga pengaruh sebuah variabel luar yang menyebabkan variasi antar subjek, maka uji Friedman analog dengan metode analisis perametriik yang disebut ANOVA dua arah. Pengaruh variabel luar itu disebut blok ( b ).

Uji Friedman merupakan metode nonparametriik yang digunakan untuk rancangan acak kelompok lengkap. Tujuan uji Friedman adalah untuk melihat ada atau tidaknya perbedaan pengaruh antar perlakuan. Ketika pengaruh perlakuan-perlakuan memiliki pengaruh yang berbeda, respon dari subjek yang diberi suatu perlakuan akan memiliki median yang sama dengan respon dari subjek yang diberi perlakuan lainnya, setelah pengaruh pengelompokkan diubah dihilangkan. Sehingga, uji ini analog dengan dengan prosedur parametriik analisis ragam dua-arah. Rancangan data untuk uji Friedman ditampilkan dalam tabel di bawah, di mana baris mewakili kelompok dan kolom untuk perlakuan. Istilah perlakuan memiliki makna yang luas, misalnya status social ekonomi atau latar belakang pendidikan.

Kelompok	Perlakuan			
	1	2	.....	k
1	$X_{11}$	$X_{12}$	.....	$X_{1k}$
2	$X_{21}$	$X_{22}$	.....	$X_{2k}$
⋮	⋮	⋮	.....	⋮
⋮	⋮	⋮	.....	⋮
b	$X_{b1}$	$X_{b2}$	.....	$X_{bk}$
	$R_1$	$R_2$	.....	$R_k$

Ada perbedaan dalam hal pemeringkatan antara uji Kruskal-Wallis dan uji Friedman. Dalam uji Kruskal-Wallis pengamatan-pengamatan dari seluruh contoh yang telah digabungkan akan diperingkatkan relatif satu sama lain. Namun pada uji Friedman pengamatan-pengamatan dalam setiap kelompok diperingkatkan secara terpisah, sehingga setiap kelompok akan memiliki gugus data peringkat, dengan k adalah banyaknya perlakuan.

### 1) Asumsi

- Data terdiri dari b kelompok yang saling bebas dengan ukuran k perlakuan
- Peubah yang diamati bersifat kontinu
- Tidak ada interaksi antara kelompok dan perlakuan
- Pengamatan dalam setiap kelompok dapat diperingkatkan berdasarkan besarnya

### 2) Hipotesis

$H_0 : M_1 = M_2 = \dots = M_k$  atau k perlakuan memiliki median yang sama

$H_1 : \text{Ada minimal satu } M_i \neq M_j \text{ dimana } i \neq j \text{ dan } i, j = 1, 2, \dots, k$

### 3) Statistik Uji

Statistik uji Friedman dapat ditentukan melalui prosedur berikut :

- Urutkan pengamatan-pengamatan dalam setiap kelompok secara terpisah,
- Jika terdapat ties (nilai yang sama) dalam kelompok, beri peringkat tengah (midrank)
- Statistik uji Friedman dapat diperoleh melalui rumus :



$$\chi_r^2 = \frac{12}{bk(k+1)} \sum_{j=1}^k R_j^2 - 3b(k+1)$$

$$\text{atau } W = \frac{12 \sum_{j=1}^k R_j^2 - 3b^2k(k+1)^2}{b^2k(k^2-1)} \quad \text{di mana } W = \chi_r^2 / b(k-1)$$

$$\text{Apabila ada ties maka } W = \frac{12 \sum_{j=1}^k R_j - 3b^2k(k+1)^2}{b^2k(k^2-1) - b(\sum t^3 - \sum t)}$$

### Keterangan

- b : banyaknya kelompok  
k : banyaknya perlakuan  
R<sub>i</sub> : jumlah peringkat perlakuan ke-i  
t : banyaknya pengamatan yang bernilai sama (ties)

Uji W digunakan jika : k = 3 dan b ≤ 15

k = 4 dan b ≤ 8

k = 5 dan b ≤ 3

#### 4) Contoh Soal

Di bawah ini adalah data jumlah korosi berbagai jenis logam pada tiga jenis segel. Selidiki dengan Uji Friedman apakah ketiga jenis segel mempunyai kemampuan menahan korosi yang berbeda (gunakan taraf nyata 5%).

Logam	Segel		
	A	B	C
1	21	23	15
2	29	30	21
3	16	19	18
4	20	19	18
5	13	10	14
6	5	12	6
7	8	18	12
8	26	32	21
9	17	20	9
10	4	10	2

- Hipotesis :  
 H0 : Ketiga jenis segel mempunyai kemampuan menahan korosi yang sama  
 H1 : Minimal ada satu jenis segel yang mempunyai kemampuan menahan korosi yang berbeda dengan jenis segel lainnya
- Statistik Uji :

Logam	Segel					
	A		B		C	
1	21	2	23	3	15	1
2	29	2	30	3	21	1
3	16	1	19	3	18	2
4	20	3	19	2	18	1
5	13	2	10	1	14	3
6	5	1	12	3	6	2
7	8	1	18	3	12	2
8	26	2	32	3	21	1
9	17	2	20	3	9	1
10	4	2	10	3	2	1
	$R_A$	<b>18</b>	$R_B$	<b>27</b>	$R_C$	<b>15</b>

- Dengan menggunakan rumus di peroleh :

$$W = \frac{12 \sum_{j=1}^k R_j^2 - 3b^2k(k+1)^2}{b^2k(k^2-1)} = \frac{12(18^2 + 27^2 + 15^2) - 3(10)^2 3(3+1)^2}{(10)^2(3)(3^2-1)} = 0.39$$

k=3 dan b=10 menunjukkan peluang untuk mendapatkan sebuah nilai W yang sama besar atau lebih besar dari 0.39 ketika H0 benar adalah sebesar 0.018. Konsekuensinya, H0 ditolak pada taraf nyata 5% sehingga dapat disimpulkan bahwa tiga jenis segel tidak mempunyai kemampuan menahan korosi yang sama. Dengan menggunakan pendekatan khi-kuadrat, nilai 2

$$\chi_r^2 = 10(3-1)(0.39) = 7.80. \text{ Berdasarkan Tabel A.11, } \chi_{(\alpha=0.05, db=2)}^2 = 5.991.$$

Karena  $\chi_r^2 > \chi_{(\alpha=0.05, db=2)}^2$ , H0 ditolak.

- **Prosedur Perbandingan Berganda untuk Uji Friedman**

Ketika uji Friedman memberikan penolakan terhadap H0, yang artinya ada sepasang perlakuan yang mempunyai pengaruh berbeda terhadap respon, biasanya kita tertarik untuk menyelidiki lebih lanjut mengenai di mana perbedaan tersebut berada. Untuk itu diperlukanlah suatu prosedur

perbandingan berganda yang konsisten untuk dapat digunakan bersama dengan uji Friedman. Hipotesis yang diuji adalah :

$$H_0 : M_i = M_j$$

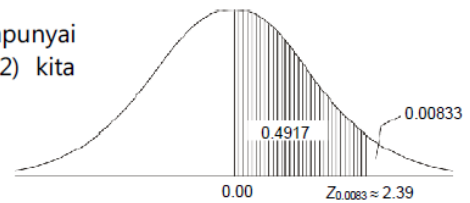
$$H_1 : M_i \neq M_j \text{ dimana } i \neq j$$

Ketika kita membandingkan semua kemungkinan pasangan perlakuan pada taraf nyata  $\alpha$ , dan banyaknya kelompok adalah besar, kita dapat menyatakan  $R_i$  dan  $R_j$  berbeda nyata apabila :

$$|R_i - R_j| \geq Z_{\frac{\alpha}{k(k-1)}} \sqrt{\frac{bk(k+1)}{6}}$$

Misalnya pada contoh di atas, kita mempunyai  $\alpha=0.05$ ,  $b=10$ ,  $k=3$ , dan dari tabel normal (A.2) kita peroleh  $Z_{0.00833} \approx 2.39$ . Sehingga kita peroleh :

$$2.39 \sqrt{\frac{10(3)(3+1)}{6}} = 10.688$$



Jumlah peringkat adalah  $RA=18$ ,  $RB=27$  dan  $RC=15$ , sehingga

$$|RA - RB| = 9 \quad |RA - RC| = 3 \quad |RB - RC| = 12$$

Dapat kita simpulkan bahwa segel jenis B dan segel jenis C mempunyai kemampuan menahan korosi yang berbeda ( $|RB - RC| = 12 > 10.688$ ), sedangkan pasangan jenis segel lainnya sama.

### 5) Cara Analisis dan Interpretasi SPSS

Suatu metode diet penurunan berat badan yaitu metode DASH (Dietary Approaches to Stop Hipertension) diuji coba terhadap 10 orang sebagai sampel. metode ini bertujuan menurunkan tekanan darah. Pelaku diet tidak berpantang terhadap makanan dan hanya memperbanyak sayuran dan buah-buahan. Untuk menguji apakah metode ini efektif menurunkan berat badan, dilakukan uji coba terhadap 10 orang. Pengukuran berat badan dilakukan sebelum program diet, 1 minggu melakukan program diet DASH dan 2 minggu kemudian. Apakah terdapat perbedaan antara ketiga kelompok sampel tersebut? Data sampel sebagai berikut :

Sebelum_diet	Minggu_1	Minggu_2
35.00	21.00	21.00
34.00	25.00	19.00
39.00	28.00	14.00
32.00	24.00	15.00
35.00	21.00	16.00
28.00	29.00	20.00
21.00	25.00	21.00
30.00	22.00	24.00
21.00	26.00	21.00
19.00	24.00	20.00

Data Sampel

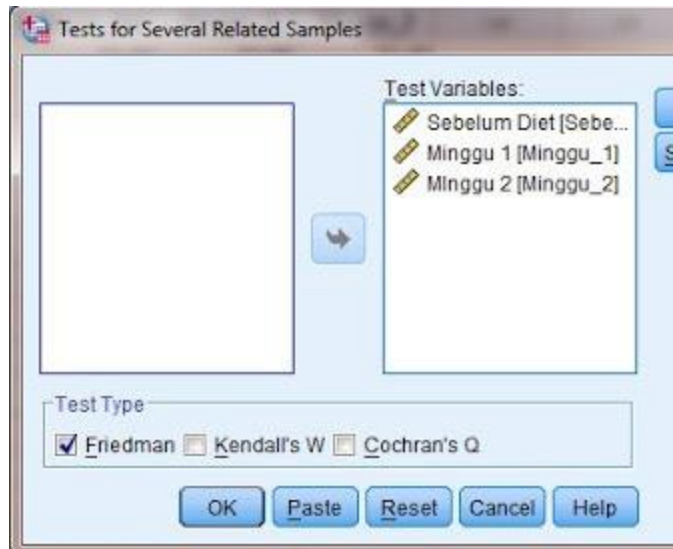
- Pada langkah pertama, akan diuji normalitas data dengan uji Kolmogorov Smirnov. Hasilnya sebagai berikut.

	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wi	
	Statistic	df	Sig.	Statistic	df
Sebelum Diet	.187	10	.200 <sup>*</sup>	.911	10
Minggu 1	.127	10	.200 <sup>*</sup>	.942	10
Minggu 2	.213	10	.200 <sup>*</sup>	.917	10

Uji Normalitas

Pada tabel uji Kolmogorov-Smirnov di atas, nilai signifikansi pada sampel sebelum diet sebesar (0.152), minggu 1 (0.002) dan minggu 2 (0.200). Hanya kelompok sampel sebelum diet dan sampel minggu 2 yang berdistribusi normal sedangkan kelompok sampel minggu 1 tidak berdistribusi normal. Oleh karena tidak terpenuhi asumsi normal pada semua kelompok sampel maka digunakan uji Friedman.

- Analyze Pilih Nonparametrics kemudian K Related Samples
- Masukkan variabel Sebelum diet, Minggu 1 dan Minggu 2 ke Test Variables
- Pada pilihan test type centang pilihan Friedman
- Pada menu Statistics, centang pilihan Descriptive, kemudian OK



Langkah Uji Friedman

- **Output**

	Mean Rank
Sebelum Diet	2.40
Minggu 1	2.25
Minggu 2	1.35

Nilai Mean Rank

Nilai rata-rata rank berat badan merupakan nilai bukan sebenarnya, tetapi dilakukan rangking terhadap data aktual. Nilai mean rank sebelum diet sebesar 2.40 pada minggu 1 nilai mean rank turun menjadi 2.25 sedangkan pada minggu 2 nilai mean rank turun lagi menjadi 1.35.

N	10
Chi-Square	6.973
df	2
Asymp. Sig.	.031

Nilai Chi-Square

Hasil uji Friedman, nilai chi-square sebesar 6,973. Nilai  $df=2$  ( $k-1$ ), dimana  $k$  adalah banyaknya kelompok sampel yaitu 3 sampel, sedangkan nilai

signifikansi p-value 0,031. Karena nilai p-value 0,031 lebih kecil dari 0,05 maka kesimpulannya adalah terdapat perbedaan nilai rata-rata rank antara sebelum diet, diet minggu 1 dan diet minggu 2. Untuk menguji atau membandingkan antara 2 kelompok, misalnya sampel sebelum diet dengan diet minggu 1, sebelum diet dan minggu 2 atau diet minggu 1 dengan minggu 2 dapat dilakukan dengan uji Post Hoc.

## 2. Uji Beda Independen

### a) Uji Mann Whitney

Uji Mann-Whitney atau lebih dikenal dengan u-test (juga disebut Mann-Whitney-Wilcoxon (MWW), Wilcoxon rank-sum test, or Wilcoxon-Mann-Whitney test). Uji ini dikembangkan oleh H.B Mann dan D.R. Whitney dalam tahun 1947. Uji Mann-Whitney ini digunakan sebagai alternatif lain dari uji T parametrik bila anggapan yang diperlukan bagi uji T tidak dijumpai. Teknik ini dipakai untuk mengetest signifikansi perbedaan antara dua populasi, dengan menggunakan sampel random yang ditarik dari populasi yang sama. Test ini berfungsi sebagai alternatif penggunaan uji-t bilamana persyaratan-persyaratan parametriknya tidak terpenuhi, dan bila datanya berskala ordinal. uji ini berbeda dengan uji wilcoxon karena uji wilcoxon untuk dua sampel yang berpasangan. sedangkan mann whitney khusus untuk dua sampel yang independent.

#### 1) Persyaratan

- Data berskala ordinal, interval atau rasio.
- Terdiri dari 2 kelompok yang independent atau saling bebas.
- Data kelompok I dan kelompok II tidak harus sama banyaknya harus sama banyaknya.
- Data tidak harus berdistribusi normal. sehingga tidak perlu uji normalitas

#### 2) Prosedur pengujian dapat dilakukan sebagai berikut :

- Susun kedua hasil Pengamatan menjadi satu kelompok sampel
- Hitung jenjang/ rangking untuk tiap – tiap nilai dalam sampel gabungan

- Jenjang atau rangking diberikan mulai dari nilai terkecil sampai terbesar
- Nilai beda sama diberi jenjang rata –rata
- Selanjutnya jumlahkan nilai jenjang untuk masing-masing sampel.
- Hitung Nilai statistik uji U.

Ada dua macam tehnik U-test ini, yaitu U-test untuk sampel-sampel kecil dimana  $n \leq 20$  dan U-test sampel besar bila  $n > 20$ . Oleh karena pada sampel besar bila  $n > 20$ , maka distribusi sampling U-nya mendekati distribusi normal, maka test signifikansi untuk uji hipotesis nihilnya disarankan menggunakan harga kritik Z pada tabel probabilitas normal. Sedangkan test signifikansi untuk sampel kecil digunakan harga kritik U . Adapun formula rumus Mann-Whitney Test. Berikut statistik uji yang digunakan dalam uji mann whitney:

### 3) Untuk sampel kecil ( $n_1$ atau $n_2 \leq 20$ )

Untuk sampel kecil dimana  $n_1$  atau  $n_2 \leq 20$ . maka digunakan rumus umum dari uji mann whitney. berikut statistik uji yang digunakan untuk sampel kecil.

$$U_1 = n_1 \cdot n_2 - U_2$$

$$U_2 = n_1 \cdot n_2 - U_1$$

Bisa menggunakan salah satu dari rumus di atas. Nah untuk mencari nilai  $U_1$  dan  $U_2$  seperti berikut.

$$U_1 = n_1 \cdot n_2 + \frac{n_2 (n_2 + 1)}{2} - \sum R_2$$

$$U_2 = n_1 \cdot n_2 + \frac{n_1 (n_1 + 1)}{2} - \sum R_1$$

#### **Keterangan:**

$U_1$  = Statistik uji  $U_1$

$U_2$  = Statistik uji  $U_2$

$R_1$  = jumlah rank sampel 1

$R_2$  = jumlah rank sampel 2

$n_1$  = banyaknya anggota sampel 1

$n_2$  = banyaknya anggota sampel 2

Setelah mendapatkan nilai statistik uji  $U_1$  dan  $U_2$ . kemudian mengambil nilai terkecil dari kedua nilai tersebut. Nilai terkecil yang diperoleh kemudian dibandingkan dengan tabel mann whitney.

#### 4) Untuk sampel besar ( $n_1$ atau $n_2 > 20$ )

Berbeda dengan kasus jumlah sampel kecil, jumlah sampel besar menggunakan statistik uji z karena jumlah sampel yang besar yaitu  $> 20$  setiap sampel. Cara ini tidak membutuhkan tabel mann whitney tapi menggunakan tabel z yang mungkin lebih populer. Caranya hampir sama untuk sampel kecil yaitu mencari  $U_1$  dan  $U_2$ . kemudian ada langkah tambahan untuk menentukan statistik uji z. Nantinya akan digunakan untuk membandingkan dengan tabel z. Berikut rumus yang digunakan.

$$Z = \frac{U - \frac{n_1 \cdot n_2}{2}}{\sqrt{\frac{n_1 \cdot n_2 \cdot (n_1 + n_2 + 1)}{12}}}$$

Rumus diatas digunakan apabila ada rangking yang berbeda. Sedangkan untuk ada rangking yang sama menggunakan rumus seperti berikut.

$$Z = \frac{U - \frac{n_1 \cdot n_2}{2}}{\sqrt{\left(\frac{n_1 \cdot n_2}{(n_1 + n_2) \cdot (n_1 + n_2 - 1)}\right) \left(\frac{(n_1 + n_2)^3 - (n_1 + n_2)}{12} - \sum \frac{t_i^3 - t_i}{12}\right)}}$$

#### 5) Contoh Kasus Untuk Sampel Kecil ( $U \leq 20$ )

Misalnya Tim Statistik Ceria penasaran ingin mengetahui apakah ada perbedaaan Denyut nadi pria dan denyut nadi wanita. kemudian dilakukan penarikan sampel untuk pria dan wanita dengan melihat denyut nadi masing-masing. Berikut hasil perhitungan masing-masing denyut nadi.



Denyut Nadi Pria	Denyut Nadi Wanita
90	79
89	82
82	85
89	88
91	85
86	80
85	80
86	
84	

- Pembahasan Untuk Sampel Kecil ( $U \leq 20$ )

Dari kasus di atas yang pertama kita lihat yaitu tujuannya. Dari tujuannya yaitu ada perbedaan antara denyut nadi pria dan wanita. dari tujuan itu ada tiga hal yang ditangkap yaitu analisis yang digunakan yaitu uji perbandingan dan sampel yang digunakan ada dua kelompok serta antar kelompok tersebut merupakan kelompok yang saling bebas atau independent. Bisa disimpulkan menggunakan uji beda dua rata-rata independent.

Sampai Tahap diatas masih berupa jenis metode yang digunakan yang tentunya masih umum. Sekarang kita menentukan metode yang digunakan. langkah selanjutnya melihat skala data yang digunakan. Pada ngerti kan. skala data ada 4 yaitu nominal, ordinal, interval dan rasio. untuk uji mann whitney minimal ordinal. artinya ordinal, interval dan rasio bisa digunakan untuk uji mann whitney. jika menggunakan data ordinal langsung pakai mann whitney. sedangkan apabila menggunakan data interval dan rasio harus diuji dulu apakah normal atau tidak. jika setelah diuji datanya normal menggunakan metode uji t beda dua rata-rata independent (parametrik). sedangkan apabila tidak normal menggunakan mann whitney (non parametrik).

Kembali ke contoh kasus. Dari tujuannya kita menggunakan analisis pebandingan dua rata-rata independent.

kemudian dari data yang digunakan yaitu interval. sehingga perlu uji normalitas terlebih dahulu untuk menentukan apakah menggunakan mann whitney atau uji t beda dua rata-rata independent. Dalam contoh ini kita anggap saja datanya tidak berdistribusi normal. Sehingga disini kita menggunakan uji Mann-Whitney.

- Hipotesis:

H0 : Denyut nadi wanita sama dengan denyut nadi pria

H1 : Denyut nadi wanita berbeda dengan denyut nadi pria

Susun kedua hasil Pengamatan menjadi satu kelompok sampel dan buat peringkat seperti berikut

Denyut Nadi	Rangking	Jenis Kelamin
79	1	Wanita
80	2,5	Wanita
80	2,5	Wanita
82	4,5	Pria
82	4,5	Wanita
84	6	Pria
85	8	Pria
85	8	Wanita
85	8	Wanita
86	10,5	Pria
86	10,5	Pria
88	12	Wanita
89	13,5	Pria
89	13,5	Pria
90	15	Pria
91	16	Pria

Selanjutnya jumlahkan nilai jenjang untuk masing-masing sampel

Denyut Nadi Pria	Rangking	Denyut Nadi Wanita	Rangking
90	15	79	1
89	13,5	82	4,5
82	4,5	85	8
89	13,5	88	12

91	16	85	8
86	10,5	80	2,5
85	8	80	2,5
86	10,5		
84	6		
Jumlah Rangkings	97,5		38,5

- Hitung Nilai statistik uji U

Setelah melalui langkah-langkah diatas. Sekarang saatnya untuk menghitung statistik uji U. Pertama yaitu dengan menghitung U1. Berikut perhitungannya.

$$U_1 = n_1 \cdot n_2 + \frac{n_2 (n_2 + 1)}{2} - \sum R_2$$

$$U_1 = 9 \cdot 7 + \frac{7(7+1)}{2} - 38,5 = 52,5$$

Sedangkan untuk menghitung U2. Bisa dengan menggunakan rumus.

$$U_2 = n_1 \cdot n_2 - U_1$$

$$U_2 = 9 \cdot 7 - 52,5$$

$$U_2 = 10,5$$

Kemudian dari kedua nilai tersebut diambil nilai terkecil yaitu 10,5 yang digunakan untuk membandingkan dengan tabel Mann Whitney. Untuk tabel mann whitney bisa didownload dilink berikut. [Tabel Mann Whitney](#).

- Cara membaca tabel mann whitney:

Pertama tentukan jumlah setiap sampel. Misalnya dalam contoh diatas yaitu  $n_1=9$  dan  $n_2 =7$ . Kemudian tentukan nilai titik kritis ( $\alpha$ ). dalam contoh ini menggunakan 0,05. Kemudian dihubungkan kolom  $n_1$  dan baris  $n_2$ . dan lihat titik kritis ( $\alpha$ ) yang digunakan yaitu 0,05. Hasilnya yaitu 12.

- Kesimpulan

Oleh karena nilai U statistik uji lebih kecil dari nilai U tabel Mann Whitney yaitu  $10,5 < 12$ . Sehingga Keputusan  $H_0$

ditolak, H1 diterima. Sehingga bisa disimpulkan ada perbedaan antara denyut nadi pria dan denyut nadi wanita.

#### 6) Contoh Kasus untuk Sampel Besar ( $U > 20$ )

Tim Statistik Ceria sedang mendapatkan kasus dalam penelitian mengenai kepadatan hunian rumah antara di daerah nelayan dengan daerah pertanian, Tim menggunakan  $\alpha = 0,05$ . Tim penasaran apakah ada perbedaan kepadatan hunian rumah antara daerah nelayan dengan daerah pertanian. didapatkan data seperti pada tabel di bawah. Disini sudah diranking caranya sama dengan contoh di atas.

Keadatan Rumah Nelayan	Rank	Keadatan Rumah Pertanian	Rank
4,25	37	1,75	1
3,1	21	2,35	8
3,25	25	3,22	23
3,05	19	3,4	29
2,41	10	2,67	13
2,15	6	4,01	33
2,25	7	1,9	3
3,52	31	2,48	11
2,03	5	3,33	27
1,85	2	3,26	26
4,19	36	2,89	17
2,86	15	3,35	28
4,02	34	2,87	16
3,83	32	2,55	12
1,92	4	3,46	30
		3,02	18
		3,23	24
		4,05	35
		3,21	2
		3,09	20
		2,83	14
		2,36	9
Jumlah Rank	284		419

- Pembahasan Untuk Sampel Besar ( $U > 20$ )

Dari kasus di atas yang pertama kita lihat yaitu tujuannya. Dari tujuannya yaitu ada perbedaan antara kepadatan rumah nelayan dengan petani. dari tujuan itu ada tiga hal yang ditangkap yaitu analisis yang digunakan yaitu uji perbandingan dan sampel yang digunakan ada dua kelompok serta antar kelompok tersebut merupakan kelompok yang saling bebas atau independent. Bisa disimpulkan menggunakan uji beda dua rata-rata independent. pemikirannya sama dengan cara di atas.

Sampai Tahap diatas masih berupa jenis metode yang digunakan yang tentunya masih umum. Sekarang kita menentukan metode yang digunakan. langkah selanjutnya melihat skala data yang digunakan. Pada ngerti kan. skala data ada 4 yaitu nominal, ordinal, interval dan rasio. untuk uji mann whitney minimal ordinal. artinya ordinal, interval dan rasio bisa digunakan untuk uji mann whitney. jika menggunakan data ordinal langsung pakai mann whitney. sedangkan apabila menggunakan data interval dan rasio harus diuji dulu apakah normal atau tidak. jika setelah diuji datanya normal menggunakan metode uji t beda dua rata-rata independent (parametrik). sedangkan apabila tidak normal menggunakan mann whitney (non parametrik).

Kembali ke contoh kasus. Dari tujuannya kita menggunakan analisis pebandingan dua rata-rata independent. kemudian dari data yang digunakan yaitu interval. sehingga perlu uji normalitas terlebih dahulu untuk menentukan apakah menggunakan mann whitney atau uji t beda dua rata-rata independent. Dalam contoh ini kita anggap saja datanya tidak berdistribusi normal.

- Hipotesis:

$H_0$  : Kepadatan rumah nelayan dan rumah petani sama

$H_1$  : Terdapat perbedaan kepadatan rumah nelayan dengan rumah petani

- Hitung Nilai statistik uji U

Sebelum melakukan perhitungan statistik uji, lakukan tahap seperti pada contoh sebelumnya yaitu mengurutkan data kemudian buat rank lalu dijumlahkan sehingga hasilnya seperti pada tabel di atas. Kemudian langsung ke perhitungannya saja. Pertama mencari U1.

$$U_1 = n_1 \cdot n_2 + \frac{n_2 (n_2 + 1)}{2} - \sum R_2$$

$$U_1 = 15 \cdot 22 + \frac{22(22+1)}{2} - 419 = \mathbf{164}$$

Kedua untuk menghitung U2. Bisa dengan menggunakan rumus.

$$U_2 = n_1 \cdot n_2 - U_1$$

$$U_2 = 15 \cdot 22 - 164$$

$$U_2 = 166$$

Berbeda dengan sampel kecil, untuk sampel besar menggunakan tabel Z sehingga perlu mencari nilai z dari nilai U yang telah diperoleh.

$$Z = \frac{U - \frac{n_1 \cdot n_2}{2}}{\sqrt{\frac{n_1 \cdot n_2 \cdot (n_1 + n_2 + 1)}{12}}}$$

$$Z = \frac{164 - \frac{15 \cdot 22}{2}}{\sqrt{\frac{15 \cdot 22 \cdot (15 + 22 + 1)}{12}}} = \mathbf{-0,0309}$$

Sedangkan apabila kita memasukkan nilai U2 maka hasilnya yaitu kebalikan dari nilai U1 yaitu +0,0309. Jadi tidak perlu dihitung lagi. Nah, kemudian yang diambil yaitu yang positif sehingga yang dibandingkan nanti yaitu 0,0309. Setelah memperoleh nilai Z maka langkah terakhir yaitu mencari nilai tabel Z. Nilai tabel pada tabel Z, Uji dua arah dengan  $\alpha = 5\%$ ,

yaitu 1, 96.

Nb: Jika ada yang butuh tabel z. silahkan download di link berikut. Tabel Z

- Kesimpulan

Oleh karena nilai statistik uji z lebih kecil dari nilai tabel Z yaitu  $0,0309 < 1,96$ . Sehingga Keputusan  $H_0$  diterima,  $H_1$  ditolak. Sehingga bisa disimpulkan tidak ada perbedaan kepadatan rumah nelayan dan petani.

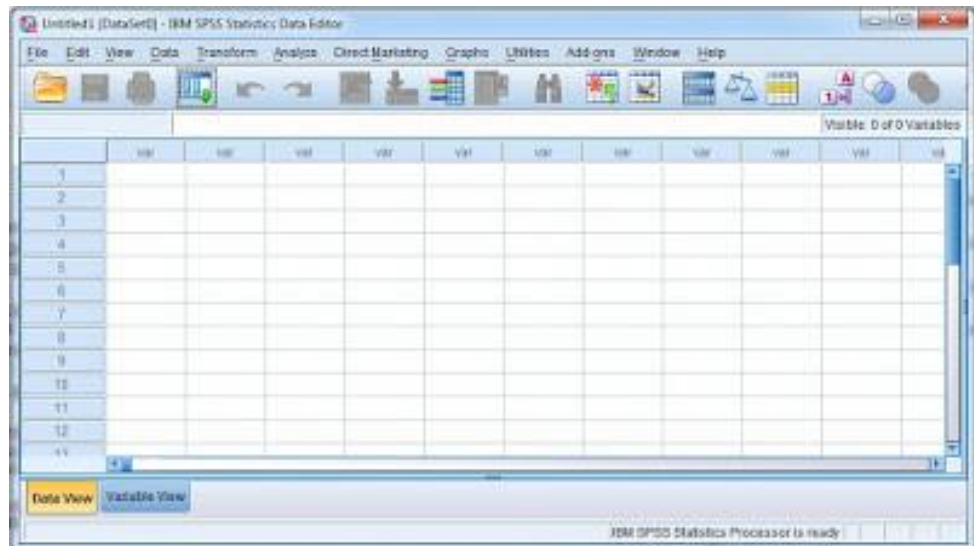
## 7) Cara Analisis dan Interpretasi SPSS

Hasil check up kadar gula darah normal pada orang dewasa yang berpuasa selama 8 jam sebelum di cek gula darah dan kadar gula darah orang yang tidak berpuasa.

No	Kadar Gula	Kelompok	No	Kadar Gula	Kelompok
1	140	Tidak Berpuasa	16	110	Berpuasa
2	165	Tidak Berpuasa	17	105	Berpuasa
3	135	Tidak Berpuasa	18	140	Berpuasa
4	173	Tidak Berpuasa	19	123	Berpuasa
5	181	Tidak Berpuasa	20	80	Berpuasa
6	134	Tidak Berpuasa	21	90	Berpuasa
7	155	Tidak Berpuasa	22	84	Berpuasa
8	160	Tidak Berpuasa	23	108	Berpuasa
9	130	Tidak Berpuasa	24	89	Berpuasa
10	170	Tidak Berpuasa	25	109	Berpuasa
11	161	Tidak Berpuasa			
12	123	Tidak Berpuasa			
13	190	Tidak Berpuasa			
14	100	Berpuasa			
15	95	Berpuasa			

Dari data di atas, seorang dokter ingin mengetahui apakah terdapat perbedaan secara signifikan kadar gula darah normal pada orang dewasa yang berpuasa dan kadar gula darah orang yang tidak berpuasa.

- Hipotesis  
 H0: Tidak berbeda secara bermakna antara kadar gula darah normal pada orang dewasa yang berpuasa dengan kadar gula darah orang yang tidak berpuasa.  
 H1: Berbeda secara bermakna antara kadar gula darah normal pada orang dewasa yang berpuasa dengan kadar gula darah orang yang tidak berpuasa.
- Pengujian statistik berdasarkan nilai Z dan probabilitas:  
 Jika  $Z_{hitung} < Z_{tabel}$  atau nilai  $sig > 0.05$ , maka H0 diterima dan sebaliknya
- Buka program SPSS sampai muncul seperti gambar berikut:

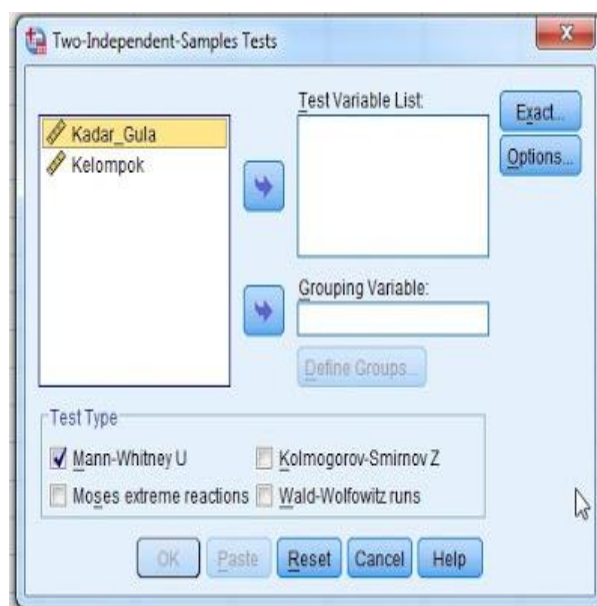




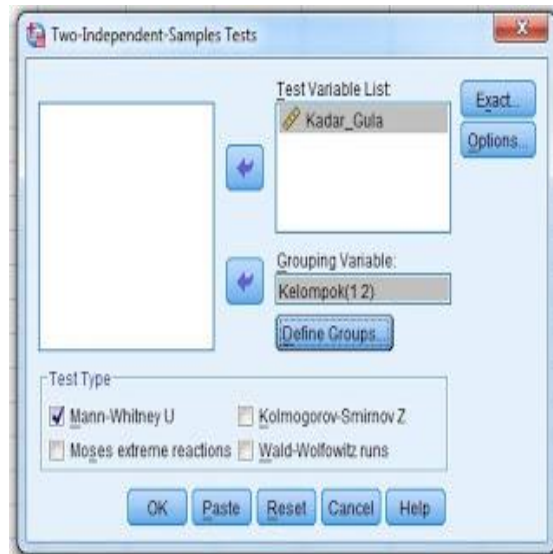
- Entry Data yang akan di analisis

	Kadar_Gula	Kelompok	var
1	140	1	
2	165	1	
3	135	1	
4	173	1	
5	181	1	
6	134	1	
7	155	1	
8	160	1	
9	130	1	
10	170	1	
11	161	1	
12	123	1	
13	190	1	
14	100	2	
15	95	2	
16	110	2	
17	105	2	
18	140	2	
19	123	2	
20	80	2	
21	90	2	
22	84	2	
23	108	2	

- Perhatikan menu toolbar paling atas, pilih menu Analyze ---> Nonparametric Test---> Legacy Dialogs---> 2 Independent Samples... jika sudah terlihat seperti gambar berikut:



- Pindahkan Kadar\_Gula kedalam kolom Test Variable List:, dan masukkan Kelompok kedalam kolom Grouping Variable: lalu klik Define Groups..., untuk kolom Group 1: ketik angka 1 yang artinya 1 ini untuk orang yang tidak berpuasa dan untuk kolom Group 2: ketik 2 yang artinya 2 ini untuk orang yang berpuasa, setelah pengisian selesai tekan Continue untuk melanjutkan ke tampilan sebelumnya. seperti gambar berikut:



- Perhatikan pada Test Type: Centang Mann-Whitney U seperti gambar di atas. jika sudah Klik OK. Maka keluar hasilnya seperti gambar berikut:

**NPar Tests**

**Mann-Whitney Test**

**Ranks**

Kelompok	N	Mean Rank	Sum of Ranks
Kadar_Gula tidak berpuasa	13	18.62	242.00
berpuasa	12	6.92	83.00
Total	25		

**Test Statistics<sup>a</sup>**

	Kadar_Gula
Mann-Whitney U	5.000
Wilcoxon W	83.000
Z	-3.972
Asymp. Sig. (2-tailed)	.000
Exact Sig. [2*(1-tailed Sig.)]	.000 <sup>b</sup>

a. Grouping Variable: Kelompok  
b. Not corrected for ties.

- Analisis hasil  
Berdasarkan pada tabel Test Statistics di atas, nilai Z hitung diperoleh -3.972 (positif dan negatif tidak diperhitungkan artinya sama saja). lalu nilai z hitung tersebut kita bandingkan dengan nilai Z tabel. Nilai Z tabel di peroleh -1.96.
- Pengambilan keputusan berdasarkan pengujian statistik yaitu:  
jika  $Z_{hitung} < Z_{tabel}$  dan nilai  $sig > 0.05$  maka  $H_0$  diterima  
jika  $Z_{hitung} > Z_{tabel}$  dan nilai  $sig < 0.05$  maka  $H_0$  di tolak dan terima  $H_1$
- Berdasarkan nilai Z dan nilai sig di atas yaitu dimana Z hitung  $-3.972 > Z_{tabel} -1.96$  dan nilai  $Asymp.Sig.(2-tailed) 0.000$  atau  $< 0.05$ . Sehingga dapat kita simpulkan tolak  $H_0$  dan terima  $H_1$ .
- Jadi: Berbeda secara bermakna antara kadar gula darah normal pada orang dewasa yang berpuasa dengan kadar gula darah orang yang tidak berpuasa

## b) Uji Kruskal Wallis

Analisis varian ranking satu arah Kruskal-Wallis atau biasa disebut Uji Kruskal-Wallis pertama kali diperkenalkan oleh William H. Kruskal dan W. Allen Wallis pada tahun 1952. Uji ini merupakan salah satu uji statistik nonparametrik dalam kasus k sampel independen. Uji Kruskal-Wallis digunakan untuk menguji apakah k sampel independen berasal dari populasi yang berbeda, dengan kata lain uji ini dapat digunakan untuk menguji hipotesis nol bahwa k sampel independen berasal dari populasi yang sama atau identik dalam hal harga rata-ratanya. Oleh karena itu, uji Kruskal-Wallis juga merupakan perluasan dari uji Mann-Whitney. Uji Kruskal Wallis adalah uji nonparametrik berbasis peringkat yang tujuannya untuk menentukan adakah perbedaan signifikan secara statistik antara dua atau lebih kelompok variabel independen pada variabel dependen yang berskala data numerik (interval/rasio) dan skala ordinal.

Uji ini identik dengan Uji One Way Anova pada pengujian parametris, sehingga uji ini merupakan alternatif bagi uji One Way ANOVA apabila tidak memenuhi asumsi misal asumsi normalitas. Selain sebagai uji alternatif,

kegunaan lain adalah sebagai perluasan dari uji Mann Whitney U Test, di mana kita ketahui bahwa uji tersebut hanya dapat digunakan pada 2 kelompok variabel dependen. Sedangkan Kruskal Wallis dapat digunakan pada lebih dari 2 kelompok misal 3, 4 atau lebih.

Oleh karena uji ini merupakan uji non parametris di mana asumsi normalitas boleh dilanggar, maka tidak perlu lagi ada uji normalitas misal uji shapiro wilk atau lilliefors. Sebagai ilustrasi adalah penelitian yang bertujuan untuk mengetahui adakah perbedaan pengaruh Metode Pembelajaran terhadap nilai ujian siswa. Di mana Metode pembelajaran sebagai variabel independen memiliki 3 kategori yaitu misal: metode A, metode B dan Metode C. Sedangkan nilai ujian sebagai variabel dependen berskala rasio yaitu berkisar antara 0 sd 100.

### 1) Dasar Pemikiran dan Metode

Data untuk pengujian Kruskal-Wallis pada umumnya dituangkan dalam tabel N baris dan k kolom. Banyaknya sampel yang terpilih dituliskan dalam tabel secara baris, sedangkan kelompok atau kategori yang tersedia dituliskan secara kolom.

Dalam penghitungan uji Kruskal-Wallis ini, masing-masing nilai observasi diberi ranking secara keseluruhan dalam satu rangkaian. Pemberian ranking diurutkan dari nilai yang terkecil hingga nilai yang terbesar. Nilai yang terkecil diberi ranking 1 dan nilai yang terbesar diberi ranking N (dimana N adalah jumlah seluruh observasi). Apabila terdapat angka yang sama, maka ranking dari nilai-nilai tersebut adalah rata-rata ranking dari nilai-nilai observasi tersebut.

Jika seluruh nilai observasi telah diberi ranking, langkah selanjutnya adalah menghitung jumlah ranking dari masing-masing kolom ( $R_j$ ).

Sampel	Kelompok / Kategori						
	1	R	2	R	...	k	R
1	$X_{11}$	$R_{11}$	$X_{12}$	$R_{12}$	...	$X_{1k}$	$R_{1k}$
2	$X_{21}$	$R_{21}$	$X_{22}$	$R_{22}$	...	$X_{2k}$	$R_{2k}$

·							
·							
·							
n <sub>j</sub>	X <sub>n1</sub>	R <sub>n1</sub>	X <sub>n2</sub>	R <sub>n2</sub>	...	X <sub>nk</sub>	R <sub>nk</sub>
R <sub>j</sub>	-	R <sub>1</sub>	-	R <sub>2</sub>	...	-	R <sub>k</sub>

Selanjutnya, uji Kruskal-Wallis dapat didefinisikan dengan rumus:

$$H = \frac{12}{N(N+1)} \sum_{j=1}^k \frac{R_j^2}{n_j} - 3(N+1)$$

**Keterangan**

- H: nilai Kruskal-Wallis dari hasil penghitungan
- R<sub>j</sub>: jumlah rank dari kelompok/kategori ke-j
- n<sub>j</sub>: banyaknya kasus dalam sampel pada kategori ke-j
- k: banyaknya kelompok/kategori
- N: jumlah seluruh observasi (N=n<sub>1</sub>+n<sub>2</sub>+n<sub>3</sub>+.....+n<sub>k</sub>)

Jika ditemukan angka sama sebanyak lebih dari 25% nilai observasi sehingga mengakibatkan banyak nilai ranking yang sama, maka perlu adanya koreksi pada rumus penghitungan uji Kruskal-Wallis, dengan faktor koreksinya adalah:

$$1 - \frac{\sum T}{N^3 - N} = 1 - \frac{\sum (t^3 - t)}{N^3 - N}$$

**Keterangan**

- t : banyaknya nilai observasi tertentu yang sama pada serangkaian nilai observasi
- N : jumlah seluruh observasi (N=n<sub>1</sub>+n<sub>2</sub>+n<sub>3</sub>+.....+n<sub>k</sub>)

- Sehingga rumus uji Kruskal-Wallis dengan kasus angka sama berjumlah banyak adalah:

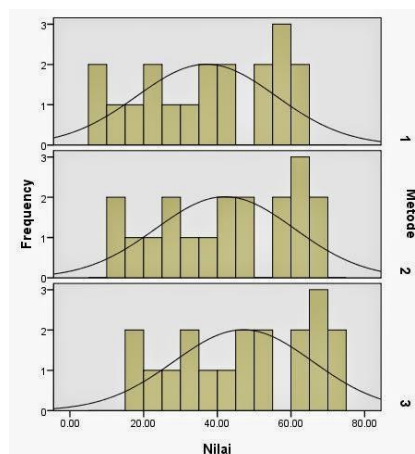
$$H = \frac{\frac{12}{N(N+1)} \sum_{j=1}^k \frac{R_j^2}{n_j} - 3(N+1)}{1 - \frac{\sum T}{N^3 - N}}$$

## 2) Asumsi Kruskal Wallis

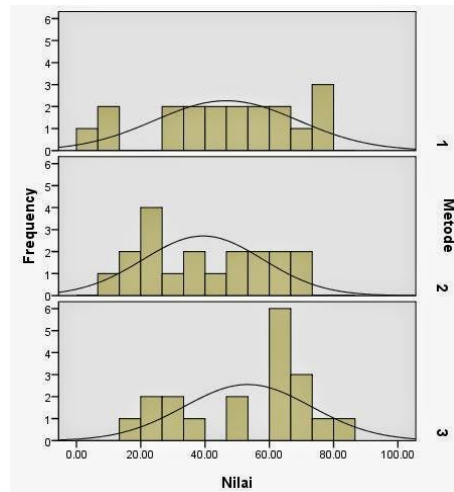
Perlu kami tekankan lagi, bahwa syarat atau asumsi uji ini adalah:

- Variabel independen berskala kategorik lebih dari 2 kategori.
- Variabel dependen berskala numeric (interval/rasio) atau skala ordinal.
- Independen artinya sampel ditiap kategori harus bebas satu sama lain, yaitu tidak boleh ada sampel yang berada pada 2 kategori atau lebih.
- Tiap kategori memiliki variabilitas yang sama, yaitu bentuk kurve histogram atau sebaran data yang sama (Lihat Histogram Variabilitas Sama). Apabila bentuk sebaran data sama, maka uji kruskall wallis dapat digunakan untuk menilai perbedaan Median antar kategori. Sedangkan jika bentuk sebaran tidak sama (Lihat Histogram Variabilitas Tidak Sama), maka uji ini tidak dapat digunakan untuk menilai perbedaan Median, jadi hanya untuk menilai perbedaan peringkat rata-rata.

- Histogram Variabilitas Sama



- Histogram Variabilitas Tidak Sama



### 3) Solusi Asumsi Kruskal Wallis

Solusi apabila Asumsi dilanggar adalah:

- Apabila kategori hanya ada maka gunakan uji Mann Whitney U Test.
- Apabila skala data di tiap variabel tidak sesuai, maka gunakan uji yang sesuai, misalkan skala data variabel independen dan dependen adalah nominal maka gunakan uji Chi-Square.
- Apabila Anggota sampel ditiap kategori sama, maka gunakan uji komparatif berpasangan untuk skala ordinal, yaitu uji Friedman Test.

### 4) Kesimpulan Hipotesis Kruskal Wallis

Hasil akhir dari uji Kruskall Wallis adalah nilai P value, yaitu apabila nilainya  $<$  batas kritis misalkan 0,05 maka kita dapat menarik kesimpulan statistik terhadap hipotesis yang diajukan yaitu: Ada pengaruh metode pembelajaran terhadap nilai ujian siswa atau yang berarti menerima H1 dan menolak H0.

- Post Hoc Kruskall Wallis

Selanjutnya jika menerima H1 maka bisa dilanjutkan dengan uji lanjut atau disebut juga uji post hoc. Uji post hoc setelah kruskall wallis salah satunya adalah uji mann whitney u test. Dengan uji tersebut kita bisa menilai antar kategori apakah yang ada perbedaan signifikan. Pada kasus di atas, maka uji post hoc yang dilakukan antara lain:

- Perbedaan Nilai Ujian Siswa antara metode A dan metode B
- Perbedaan Nilai Ujian Siswa antara metode A dan metode C
- Perbedaan Nilai Ujian Siswa antara metode B dan metode C.

### 5) Contoh Soal

Untuk membandingkan tingkat keefektifan dari 3 macam metode diet, maka sebanyak 22 orang mahasiswi yang dipilih dari suatu universitas dibagi ke dalam 3 kelompok yang mana masing-masing kelompok mengikuti program diet selama empat minggu sesuai dengan metode yang telah dibuat. Setelah program diet berakhir, maka diperoleh banyaknya berat badan yang hilang (dalam kg) dari mahasiswi-mahasiswi tersebut sebagai berikut:

Metode Diet 1		Metode Diet 2		Metode Diet 3	
Sampel	Berat Badan (BB) yg hilang	Sampel	Berat Badan (BB) yg hilang	Sampel	Berat Badan (BB) yg hilang
1	5,3	1	6,3	1	2,4
2	4,2	2	8,4	2	3,1
3	3,7	3	9,3	3	3,7
4	7,2	4	6,5	4	4,1
5	6,0	5	7,7	5	2,5
6	4,8	6	8,2	6	1,7
		7	9,5	7	5,3
				8	4,5
				9	1,3

Untuk menguji  $H_0$  yang menyatakan bahwa tingkat keefektifan dari ketiga metode diet di atas adalah sama, terhadap hipotesis alternatif yang menyatakan bahwa tingkat keefektifan ketiga metode di atas adalah tidak sama ( $\alpha = 5\%$ ).



**Jawaban :**

- Hipotesis

$H_0$  : tingkat keefektifan dari ketiga metode diet adalah sama

$H_1$  : tingkat keefektifan dari ketiga metode diet adalah tidak sama

- Tes Statistik : Kruskal-Wallis Test

- Tingkat Signifikansi :  $\alpha=5\%$ ,

- Distribusi sampling :

H mendekati distribusi Chi-Square dengan derajat bebas (k-1), sehingga wilayah kritis dapat ditentukan dengan menggunakan Tabel C.

- Penghitungan

$n_1=6 ; n_2=7 ; n_3=9 ; N= n_1 + n_2 + n_3 = 22$

Metode Diet 1		Metode Diet 2		Metode Diet 3	
BB yg hilang	Ranking	BB yg hilang	Ranking	BB yg hilang	Ranking
5,3	12,5	6,3	15	2,4	3
4,2	9	8,4	20	3,1	5
3,7	6,5	9,3	21	3,7	6,5
7,2	17	6,5	16	4,1	8
6,0	14	7,7	18	2,5	4
4,8	11	8,2	19	1,7	2
		9,5	22	5,3	12,5
				4,5	10
				1,3	1
	$R_1 = 70$		$R_2 = 131$		$R_3 = 52$

$$H = \frac{12}{N(N+1)} \sum_{j=1}^k \frac{R_j^2}{n_j} - 3(N+1)$$
$$= \frac{12}{22(22+1)} \left[ \frac{70^2}{6} + \frac{131^2}{7} + \frac{52^2}{9} \right] - 3(22+1)$$
$$= 15,633$$

- Daerah penolakan :  $H^3 c_{\alpha(k-1)}$  atau p-value  $\leq \alpha$
- Keputusan :  
 $c_{0,05(2)} = 5,991$   
 Karena  $15,633 > 5,991$        $H > c_{0,05(2)}$  , maka Tolak  $H_0$
- Kesimpulan : Dengan tingkat kepercayaan 95 %, belum cukup bukti untuk menyatakan bahwa tingkat keefektifan dari ketiga metode diet tersebut adalah sama.

# BAB 13 ANALYSIS OF VARIANCE (ANOVA)

---

## A. PENGERTIAN ANOVA

Uji ANOVA adalah bentuk khusus dari analisis statistik yang banyak digunakan dalam penelitian eksperimen. Metode analisis ini dikembangkan oleh **R.A Fisher**. Uji ANOVA juga adalah bentuk uji hipotesis statistik dimana kita mengambil kesimpulan berdasarkan data atau kelompok statistik inferentif. Hipotesis nol dari uji ANOVA adalah bahwa data adalah *simple random* dari populasi yang sama sehingga memiliki ekspektasi *mean* dan varians yang sama. Sebagai contoh penelitian perbedaan perlakuan terhadap sampel pasien yang sama. Hipotesis nol nya adalah semua perlakuan akan memiliki efek yang sama.

Meskipun uji t adalah statistik yang sering digunakan, hanya saja uji t dibatasi untuk menguji hipotesis dua kelompok. Uji ANOVA atau Analisis varians (ANOVA) dikembangkan untuk memungkinkan peneliti untuk menguji hipotesis perbandingan lebih dari dua kelompok. Dengan demikian, uji-t dan uji ANOVA adalah sama-sama metode statistik untuk perbandingan. Yang membedakan keduanya adalah hanya jumlah kelompok yang dibandingkan.

*Analysis of Variance* merupakan metode untuk menguji hubungan antara satu variabel dependen (skala metrik) dengan satu atau lebih variabel independen (skala nonmetrik atau kategorikal dengan kategori lebih dari dua). Misalkan kita ingin mengetahui apakah pengalaman kerja sebelumnya (variabel dependen) dipengaruhi oleh jabatan atau *job category* (variabel independen skala kategori).

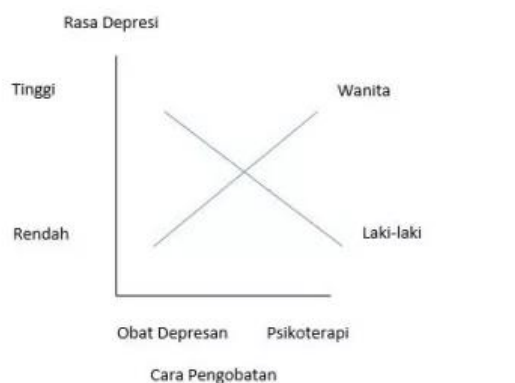
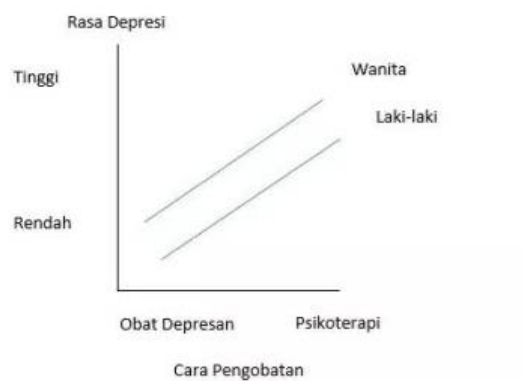
## B. KEGUNAAN ANOVA

ANOVA digunakan sebagai alat analisis untuk menguji hipotesis penelitian yang mana menilai adakah perbedaan rerata antara kelompok. Hasil akhir dari analisis ANOVA adalah nilai F test atau F hitung. Nilai F Hitung ini yang nantinya akan dibandingkan dengan nilai pada tabel f. Jika nilai f hitung lebih dari f tabel, maka dapat disimpulkan bahwa menerima H1 dan menolak H0 atau yang berarti ada perbedaan bermakna rerata pada semua kelompok.

Analisis ANOVA sering digunakan pada penelitian eksperimen dimana terdapat beberapa perlakuan. Peneliti ingin menguji, apakah ada perbedaan bermakna antar perlakuan tersebut.

ANOVA digunakan untuk mengetahui pengaruh utama (*main effect*) dan pengaruh interaksi (*interaction effect*) dari variabel independen kategorikal (sering disebut faktor) terhadap variabel dependen metrik. Pengaruh utama (*main effect*) adalah pengaruh langsung variabel independen terhadap variabel dependen. Sedangkan pengaruh interaksi adalah pengaruh bersama atau *joint effect* dua atau lebih variabel independen terhadap variabel dependen.

Sebagai misal kita ingin mengetahui pengaruh gender pasien yang berinteraksi dengan cara pengobatan yang diberikan untuk mengurangi rasa depresi pasien. Wanita akan memberi respon positif dengan cara pengobatan psikoterapi oleh karena mereka mempunyai kesempatan untuk berbincang-bincang tentang perasaannya, sedangkan pada laki-laki akan lebih senang dengan cara pengobatan meminum obat anti depresan. Dalam hal ini kita berharap cara pengobatan akan berinteraksi dengan gender untuk menurunkan rasa depresi. Berikut ini gambar grafik tanpa interaksi dan dengan interaksi.



Pada garis vertikal menunjukkan tingkat perbaikan depresi setelah dilakukan pengobatan (*treatment*), sedangkan garis horizontal merupakan dua cara pengobatan yang dilakukan yaitu dengan meminum obat antidepresan dan psikoterapi. Pengaruh variabel ketiga adalah gender atau jenis kelamin pasien yang digambarkan oleh dua garis paralel atau saling memotong. Pada Gambar 1 dimana garis gender paralel menunjukkan tidak ada interaksi antara cara pengobatan dengan gender. Wanita memiliki tingkat depresi yang tinggi dibandingkan laki-laki baik dengan cara pengobatan meminum obat atau lewat psikoterapi. Sedangkan pada Gambar 2, garis gender memotong yang berarti terdapat interaksi antara gender dengan cara pengobatan. Wanita akan memiliki depresi yang rendah jika meminum obat dan depresi tinggi jika lewat pengobatan psikoterapi. Sebaliknya laki-laki memiliki depresi yang rendah jika lewat psikoterapi dan akan tinggi depresinya jika meminum obat antidepresan.

Pengaruh gender dan cara pengobatan terhadap perbaikan tingkat depresi disebut pengaruh langsung (*main effect*). Pada *analysis of variance*, variabel independen kategorikal disebut dengan faktor dan jumlah kategori untuk masing-masing variabel independen disebut level. Sebagai misal variabel gender memiliki dua kategori wanita dan laki-laki maka disebut *two level faktor*. Variabel cara pengobatan juga mempunyai dua kategori yaitu meminum obat anti depresan atau psikoterapi. Studi yang ingin meneliti pengaruh dua atau lebih faktor sering disebut dengan *faktorial design*. Studi yang membandingkan dua level gender dan dua level cara pengobatan disebut dengan 2 x 2 faktorial design.

### C. CIRI-CIRI ANOVA

Ciri khasnya adalah adanya satu atau lebih variabel bebas sebagai faktor penyebab dan satu atau lebih variabel response sebagai akibat atau efek dari adanya faktor. Contoh penelitian yang dapat menggambarkan penjelasan ini: “Adakah pengaruh jenis bahan bakar terhadap umur thorax mesin.” Dari judul tersebut jelas sekali bahwa bahan bakar adalah faktor penyebab sedangkan umur thorax mesin adalah akibat atau efek dari adanya perlakuan faktor. Ciri lainnya adalah variabel response berskala data rasio atau interval (numerik atau kuantitatif).

ANOVA merupakan salah satu dari berbagai jenis uji parametris, karena mensyaratkan adanya distribusi normal pada variabel terikat per perlakuan atau

distribusi normal pada residual. Syarat normalitas ini mengasumsikan bahwa sample diambil secara acak dan dapat mewakili keseluruhan populasi agar hasil penelitian dapat digunakan sebagai generalisasi. Namun keunikannya, uji ini dapat dikatakan relatif robust atau kebal terhadap adanya asumsi tersebut.

#### D. JENIS ANOVA

Jenisnya adalah berdasarkan jumlah variabel faktor (independen variable atau variabel bebas) dan jumlah variabel responen (dependent variable atau variabel terikat). Pembagiannya adalah sebagai berikut:

##### 1. Univariat:

- Univariate One Way Analysis of Variance. Apabila variabel bebas dan variabel terikat jumlahnya satu.
- Univariate Two Way Analysis of Variance. Apabila variabel bebas ada 2, sedangkan variabel terikat ada satu.
- Univariate Multi way Analysis of Variance. Apabila variabel bebas ada  $> 2$ , sedangkan variabel terikat ada satu.

##### 2. Multivariat:

- Multivariate One Way Analysis of Variance. Apabila variabel bebas dan variabel terikat jumlahnya lebih dari satu.
- Multivariate Two Way Analysis of Variance. Apabila variabel bebas ada 2, sedangkan variabel terikat jumlahnya lebih dari satu.
- Multivariate Multi way Analysis of Variance. Apabila variabel bebas ada  $> 2$ , sedangkan variabel terikat jumlahnya lebih dari satu.

Jenis lain yang menggunakan prinsip ini adalah:

- Repeated Measure Analysis of variance.
- Analysis of Covariance (ANCOVA).
- Multivariate Analysis of covariance (MANCOVA).

#### E. ONE WAY ANOVA

ANOVA digunakan untuk melihat perbandingan rata-rata beberapa kelompok biasanya lebih dari dua kelompok. ANOVA satu arah digunakan pada kelompok yang digunakan berasal dari sampel yang berbeda tiap kelompok. Jadi, bisa

disimpulkan Pertama yang perlu dilihat tujuannya membandingkan rata-rata kelompok lebih dari dua. Kedua Sampel yang digunakan dari sampel yang berbeda per kelompok.

Disebut ANOVA satu arah (One Way Anova), karena pusat perhatian kita hanya satu, dalam hal ini bentuk kemasan suatu produk. Tetapi jika pusat perhatian kita, selain jenis kemasan, juga tertuju pada pengaruh aroma pada tingkat penjualan, maka digunakan ANOVA dua arah (Two Way Anova).

ANOVA satu jalur (One Way Anova) menggunakan prinsip perhitungan yang sangat sederhana, dalam analisis ini, variance total hanya dibagi atas: Variance antar perlakuan (between), dan variasi dalam perlakuan (within)/variance error.

Pada dasarnya ANOVA satu arah juga dapat digunakan untuk kasus yang diuji menggunakan ANOVA dua arah, namun kita harus melakukan pengujian satu persatu, sehingga jauh lebih efektif jika digunakan ANOVA dua arah.

### 1. Langkah-langkah dalam perhitungan ANOVA satu jalur:

- Tentukan k atau banyaknya perlakuan,
- Tentukan n atau banyaknya sampel,
- Hitung jumlah kuadrat total dengan rumus:

$$SS_T = \sum (X_{ij})^2 - \frac{(\sum T_j)^2}{n}$$

- Hitung jumlah kuadrat perlakuan dengan rumus:

$$SS_P = \sum \frac{\sum (T_j)^2}{n_j} - \frac{(\sum T_j)^2}{n}$$

- Cari harga F-Hitung dengan menggunakan rumus yang tertera pada tabel berikut,

Sumber Variasi	df	SS	MS	F-HITUNG
Antar Perlakuan	k-1	$SS_P$	$\frac{SS_P}{k-1}$	$\frac{MS_P}{MS_E}$
Dalam Perlakuan (error)	(n-1)-(k-1)	$SS_E = SS_T - SS_P$	$\frac{SS_E}{(n-1)-(k-1)}$	
Total	n-1	$SS_T$		

- Cari harga F tabel dengan mempertimbangkan (1) tingkat signifikansi ( $\alpha$ ), (2) df antar perlakuan, dan (3) df dalam perlakuan,

- g) Bandingkan harga F Hitung dengan F tabel,
- h) Bila F Hitung < F tabel, maka Ho diterima, yang berarti rata-rata kedua perlakuan tidak berbeda secara signifikan,
- i) Bila F Hitung > F tabel, maka Ho ditolak dan H1 diterima, yang berarti rata-rata kedua perlakuan berbeda secara signifikan.

## 2. Contoh Kasus ANOVA satu arah:

Terdapat 4 metode diet dan 3 golongan usia peserta program diet Berikut data rata-rata penurunan berat peserta keempat metode dalam tiga kelompok umur.

Sampel	Penurunan Berat Badan (Kg)			
	Metode 1	Metode 2	Metode 3	Metode 4
Sampel 1	4	8	7	6
Sampel 2	6	12	3	5
Sampel 3	4	-	-	5

### Solusi kasus ANOVA satu arah

- Merumuskan Hipotesis
  - H0 :  $\mu_1 = \mu_2 = \mu_3$  (Setiap metode memberikan rata-rata penurunan berat badan yang sama)
  - H1 : Ada suatu metode yang memberikan rata-rata penurunan berat badan yang tidak sama
- Identifikasi model.
 

Pertama. berdasarkan hipotesis yang digunakan yaitu membandingkan rata-rata lebih dari dua kelompok maka metode yang mungkin adalah Anova. kedua Sampel yang digunakan tiap kelompok berbeda perlakuan sehingga tipe ANOVA yang cocok adalah ANOVA satu arah.
- Memeriksa asumsi Anova.
 

Dalam metode ANOVA yang perlu diperhatikan ada empat seperti pada keterangan diatas. asumsi normal dan homogenitas antar varians kelompok harus terpenuhi. dalam contoh ini kita asumsikan asumsi terpenuhi karena kita fokus pada langkah-langkah ANOVA satu arah. kemudian kelompok yang dianalisis berasal dari kelompok saling bebas. dan data yang



digunakan merupakan data rasio. Setelah asumsi ini terpenuhi maka bisa lanjut ke perhitungan selanjutnya. kalau tidak ganti metode.

- Menghitung F hitung melalui Variabilitas, Derajat bebas dan Kuadrat tengah

a) Jumlah Kuadrat Total (JKT)

$$\mathbf{JKT = (4^2+6^2+4^2+8^2+12^2+7^2+3^2+6^2+5^2+5^2)-(60^2/10)=420-360=60}$$

b) Jumlah Kuadrat Kolom (JKK)

$$\mathbf{JKK=(14^2/3+20^2/2+10^2/2+16^2/3)-(60^2/10)=(65.33+200+50+85.33)-360 =40.67}$$

c) Jumlah Kuadrat Galat (JKG)

$$\mathbf{JKG = JKT - JKK = 60-40.67 = 19.33}$$

d) Kuadrat Tengah Kolom (KTK)

$$\mathbf{KTK = JKK / k-1 = 40.67/3 = 13.55}$$

e) Kuadrat Tengah Galat (KTG)

$$\mathbf{KTG = JKG / N - k = 19.33/6 = 3.22}$$

f) f hitung

$$\mathbf{f \ hitung =KTK / KTG = 13.55/3.22 = 4.21}$$

- Perhitungan Tabel anova

Agar mempermudah perhitungan kita menggunakan tabel berikut:

Sumber Keragaman (SK)	Jumlah Kuadrat (JK)	Derajat Bebas (db)	Kuadrat Tengah (KT)	F hitung
Kolom (K)	JKK = 40.67	db JKK = 4-1 = 3	KTK =13.55	F hitung = 4.21
Galat (G)	JKG = 19.33	db JKG= 10- 4=6	KTG =3.22	
Total (T)	JKT = 60	db JKT=10 -1 =9		

- Menghitung F<sub>tabel</sub>

F table pada  $\alpha = 0.05$  db1=3 dan dk2=6 adalah 4.76

- Kesimpulan :  
 Karena  $F_{hitung}$  ada di daerah penerimaan ( $F_{hitung} < F_{tabel}$ ) maka  $H_0$  terima, sehingga bisa disimpulkan setiap metode memberikan dampak rata-rata penurunan berat badan yang sama

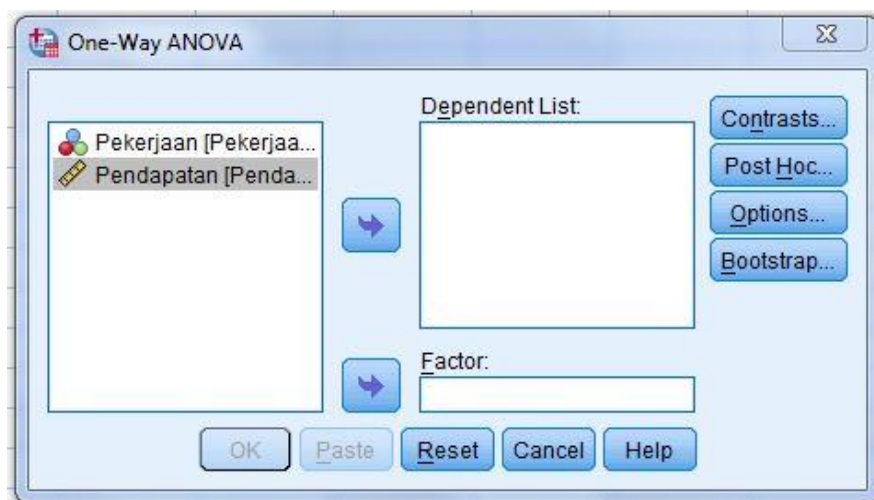
### 3. Cara Analisis dan Interpretasi SPSS

- Masukkan data yang akan di analisis

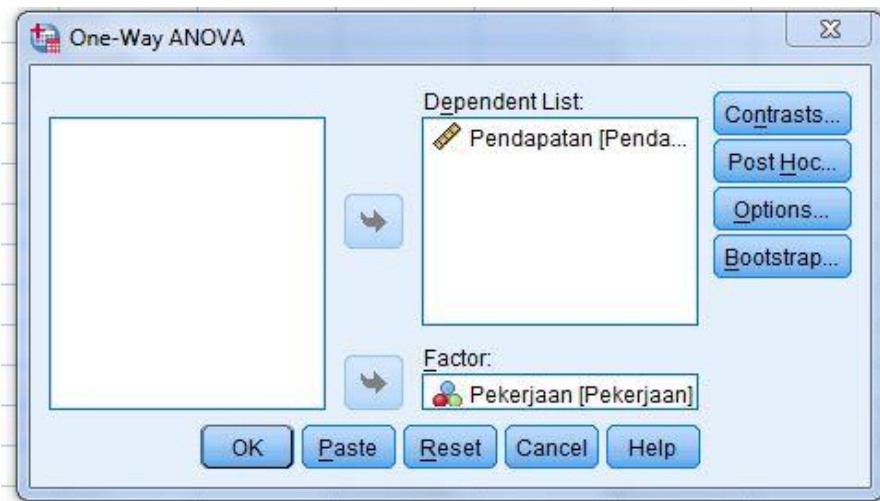
Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
Pekerjaan	Numeric	8	0	Pekerjaan	{1, Tani}...	None	8	Right	Nominal
Pendapatan	Numeric	8	0	Pendapatan	None	None	8	Right	Scale

	Pekerjaan	Pendapatan
1	1	122330
2	1	100000
3	1	252330
4	1	200000
5	1	192330
6	1	192330
7	1	252330
8	1	252330
9	2	122330
10	2	322330
11	2	122330
12	2	252330
13	2	222330
14	2	252330
15	2	363333
16	2	463333
17	3	310030
18	3	233333
19	3	310000
20	3	252330

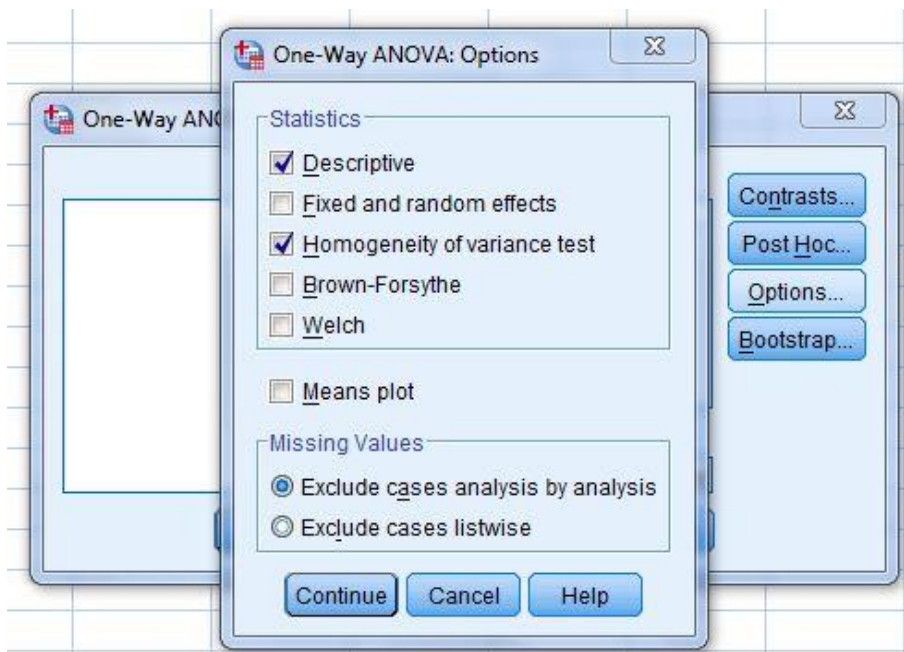
- Pada menu, pilih Analyze, Compare Means, One-Way ANOVA, sampai muncul jendela One-Way ANOVA seperti di bawah ini:



- Pilih variabel “Pendapatan” lalu masukkan ke kotak “Dependent List:” Kemudian pilih variabel “Pekerjaan” lalu masukkan ke kotak “Factor:” Sehingga nampak seperti di bawah ini:

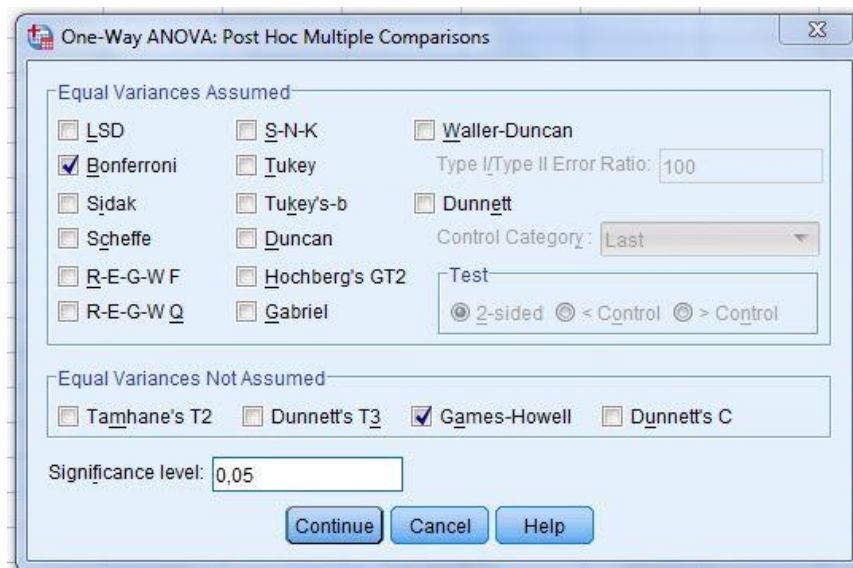


- Klik tombol Options, akan muncul jendela ini: Centang “Descriptive” dan “Homogeneity of variance test“



- Klik Continue

- Masih dijendela One Way ANOVA, klik tombol Post Hoc, sampai muncul jendela ini: Centang Bonferroni dan Games-Howell serta biarkan significance level = 0,05.



- Klik Continue - OK
  - Output
- ➔ **Oneway**

[DataSet1] C:\Users\Anwar\Documents\One Way Anova.sav

#### Descriptives

Pendapatan

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Tani	8	195497,50	58801,595	20789,503	146338,14	244656,86	100000	252330
Buruh	8	265080,75	116584,291	41218,772	167613,84	362547,66	122330	463333
Lainnya	8	326423,25	97331,637	34411,930	245051,97	407794,53	233333	552330
Total	24	262333,83	105153,627	21464,394	217931,35	306736,32	100000	552330

#### Test of Homogeneity of Variances

Pendapatan

Levene Statistic	df1	df2	Sig.
1,083	2	21	,357

Pendapatan					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	68656754666	2	34328377333	3,883	,037
Within Groups	1,857E+11	21	8840990696		
Total	2,543E+11	23			

### Post Hoc Tests

#### Multiple Comparisons

Dependent Variable: Pendapatan

	(I) Pekerjaan	(J) Pekerjaan	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Bonferroni	Tani	Buruh	-69583,250	47013,271	,461	-191881,22	52714,72
		Lainnya	-130925,750*	47013,271	,033	-253223,72	-8627,78
	Buruh	Tani	69583,250	47013,271	,461	-52714,72	191881,22
		Lainnya	-61342,500	47013,271	,618	-183640,47	60955,47
	Lainnya	Tani	130925,750*	47013,271	,033	8627,78	253223,72
		Buruh	61342,500	47013,271	,618	-60955,47	183640,47
Games-Howell	Tani	Buruh	-69583,250	46164,820	,327	-195444,84	56278,34
		Lainnya	-130925,750*	40204,283	,018	-238804,51	-23046,99
	Buruh	Tani	69583,250	46164,820	,327	-56278,34	195444,84
		Lainnya	-61342,500	53695,140	,506	-202390,99	79705,99
	Lainnya	Tani	130925,750*	40204,283	,018	23046,99	238804,51
		Buruh	61342,500	53695,140	,506	-79705,99	202390,99

\*. The mean difference is significant at the 0.05 level.

- **Interprestasi Uji ANOVA**

- Dari tabel Descriptives nampak bahwa responden yang bekerja sebagai Tani rata-rata berpendapatan sebesar 195497,50, Buruh rata-rata berpendapatan sebesar 265080,75 dan Lainnya rata-rata berpendapatan 326423,25. Selanjutnya untuk melihat uji kita lihat di tabel ANOVA.
- Sebelum melanjutkan uji perlu diingat bahwa salah satu asumsi ANOVA adalah variansnya sama. Dari tabel Test of Homegeneity of Variances terlihat bahwa hasil uji menunjukkan bahwa varian ketiga kelompok tersebut sama (P-value = 0,357), sehingga uji ANOVA valid untuk menguji hubungan ini.
- Selanjutnya untuk melihat apakah ada perbedaan pendapatan dari ketiga kelompok pekerja tersebut. Kita lihat tabel ANOVA, dari tabel itu pada kolom Sig. diperoleh nilai P (P-value) = 0,037. Dengan demikian pada taraf nyata = 0,05 kita menolak Ho, sehingga kesimpulan yang didapatkan adalah ada perbedaan yang bermakna rata-rata pendapatan berdasarkan ketiga kelompok pekerjaan tersebut.

- Interpretasi Uji ANOVA: Post Hoc
  - Jika hasil uji menunjukkan  $H_0$  gagal ditolak (tidak ada perbedaan), maka uji lanjut (Post Hoc Test) tidak dilakukan. Sebaliknya jika hasil uji menunjukkan  $H_0$  ditolak (ada perbedaan), maka uji lanjut (Post Hoc Test) harus dilakukan.
  - Karena hasil uji ANOVA menunjukkan adanya perbedaan yang bermakna, maka uji selanjutnya adalah melihat kelompok mana saja yang berbeda.
  - Untuk menentukan uji lanjut mana yang digunakan, maka kembali kita lihat tabel Test of Homogeneity of Variances, bila hasil tes menunjukkan varian sama, maka uji lanjut yang digunakan adalah uji Bonferroni. Namun bila hasil tes menunjukkan varian tidak sama, maka uji lanjut yang digunakan adalah uji Games-Howell.
  - Dari Test of Homogeneity menghasilkan bahwa varian ketiga kelompok tersebut sama, maka uji lanjut (Post Hoc Test) yang digunakan adalah Uji Bonferroni.
  - Dari tabel Post Hoc Test di atas memperlihatkan bahwa kelompok yang menunjukkan adanya perbedaan rata-rata pendapatan (ditandai dengan tanda bintang “\*”) adalah Kelompok “Tani” dan “Lainnya”.

## F. TWO-WAY ANOVA

ANOVA digunakan untuk melihat perbandingan rata-rata beberapa kelompok biasanya lebih dari dua kelompok. ANOVA dua arah digunakan pada kelompok yang digunakan berasal dari sampel yang sama tiap kelompok. Sama disini diartikan berasal dari kategori yang sama. Jadi, bisa disimpulkan Pertama yang perlu dilihat tujuannya membandingkan rata-rata kelompok lebih dari dua. Kedua Sampel yang digunakan merupakan sampel yang sudah dikategorikan per kelompok sama. Konsep ketiga yang perlu dimengerti adalah setiap kelompok tersebut dilakukan pengulangan pengujian. ini seperti menggabung ANOVA satu arah dan ANOVA dua arah tanpa interkasi.

Two way ANOVA digunakan untuk menguji hipotesis komparatif rata-rata k sampel bila peneliti melakukan kategorisasi terhadap sampel kedalam beberapa blok, sehingga bila variabilitas atau sumber keragaman pada uji One Way ANOVA berasal dari perlakuan dan galat, maka pada two way ANOVA sumber keragaman tidak hanya berasal dari perlakuan dan galat, tapi juga berasal dari blok.

## 1. Langkah-langkah dalam perhitungan ANOVA dua jalur (two way ANOVA):

- Identifikasi nilai: t (jumlah perlakuan), r (jumlah blok),
- hitung jumlah pengamatan total (n), yaitu:  $n = r \times t$ ,
- Hitung jumlah kuadrat total dengan rumus:

$$SS_T = \sum (X_{ij})^2 - \frac{(\sum T_j)^2}{n}$$

- Hitung jumlah kuadrat perlakuan dengan rumus:

$$SS_P = \sum \frac{\sum (P_1)^2}{r} - \frac{(\sum T_j)^2}{n}$$

- Hitung jumlah kuadrat antar blok dengan rumus:

$$SS_B = \sum \frac{\sum (B_1)^2}{t} - \frac{(\sum T_j)^2}{n}$$

- Cari harga F-Hitung dengan menggunakan rumus yang tertera pada tabel.

Sumber Variasi	df	SS	MS	F-HITUNG
Antar Blok	r-1	$SS_B$	$MS_B = \frac{SS_B}{r-1}$	$\frac{MS_B}{MS_E}$
Antar Perlakuan	t-1	$SS_P$	$MS_P = \frac{SS_P}{t-1}$	$\frac{MS_P}{MS_E}$
Dalam Perlakuan (error)	$(n-1)-(k-1)-(t-1)$	$SS_E = SS_T - SS_B - SS_P$	$\frac{SS_E}{(r-1)-(t-1)}$	
Total	n-1	$SS_T$		

- Cari harga F tabel dengan mempertimbangkan (1) tingkat signifikansi ( $\alpha$ ), (2) df1 yaitu df dari MS terbesar, dan (3) df2 yaitu df dari MS terkecil.
- Bandingkan harga F Hitung dengan F tabel.
  - Bila F Hitung < F tabel, maka  $H_0$  diterima, yang berarti rata-rata kedua perlakuan tidak berbeda secara signifikan,
  - Bila F Hitung > F tabel, maka  $H_0$  ditolak dan  $H_1$  diterima, yang berarti rata-rata kedua perlakuan berbeda secara signifikan.

## 2. Hipotesis dalam ANOVA (analysis of variance) dengan interaksi:

Dalam analysis of variance dua arah dengan interaksi terdapat tiga hipotesis yang digunakan yaitu apakah ada perbedaan rata-rata antar kategori baik kategori berdasarkan baris maupun kolom. hipotesis tambahan satu lagi yaitu apakah ada interaksi antara kategori baris dan kolom. Berikut hipotesis dalam ANOVA dua arah dengan interaksi.

- Hipotesis ANOVA kolom  
H0:  $a_1 = a_2 = \dots = a_k$ , Tidak ada perbedaan yang nyata antara rata-rata hitung dari kategori kolom  
H1:  $a_1 \neq a_2 \neq \dots \neq a_k$ , Ada perbedaan yang nyata antara rata-rata hitung dari kategori kolom
- Hipotesis ANOVA baris  
H0:  $b_1 = b_2 = \dots = b_j$ , Tidak ada perbedaan yang nyata antara rata-rata hitung dari kategori baris  
H1:  $b_1 \neq b_2 \neq \dots \neq b_j$ , Ada perbedaan yang nyata antara rata-rata hitung dari kategori baris
- Hipotesis interaksi  
H0:  $(ab)_{11} = (ab)_{12} = \dots = (ab)_{kj}$ , Tidak ada interaksi antara variabel baris dan kolom  
H1:  $(ab)_{11} \neq (ab)_{12} \neq \dots \neq (ab)_{kj}$ , ada interaksi antara variabel baris dan kolom

## 3. Langkah-langkah melakukan uji hipotesis dengan ANOVA

- a) Kelompokkan berdasarkan kategori tertentu. Untuk memudahkan pengelompokkan dan perhitungan, buat tabel data sesuai dengan kategori berisi sampel dan kuadrat dari sampel tersebut. Hitung pula total dari sampel dan kuadrat sampel tiap kelompok. Selain itu, tentukan pula hipotesis nol (H0) dan hipotesis alternatif (H1).
- b) Menentukan tipe ANOVA untuk menentukan tipe anova. terlebih dahulu bertanya apakah dari hipotesis tersebut cocok untuk anova? jika tujuannya membandingkan rata-rata tiga kelompok atau lebih maka boleh pakai Anova. Pertanyaan kedua apakah sampel tiap kelompok diambil dari sampel yang sudah dikategorikan? jika berasal dari sampel yang sudah dikategorikan maka menggunakan ANOVA dua arah/two way. Pertanyaan



ketiga apakah dalam sampel yang dikategorikan tadi terjadi pengulangan atau tidak? jika terjadi pengulangan maka menggunakan ANOVA dua arah dengan interaksi.

c) Memeriksa apakah sudah memenuhi asumsi-asumsi sehingga bisa digunakan anova

- Normalitas,  
adalah Menguji apakah data tiap kelompok memiliki distribusi normal. hal ini bisa dilakukan dengan uji kolmogorov smirnov, shapira wilk.
- Homogenitas  
adalah Menguji apakah varians tiap kelompok sama. Dalam menghitung homogenitas bisa digunakan uji bartlett dan uji levene.
- Saling bebas  
Menunjukkan bahwa setiap kelompok tidak saling berhubungan. Biasanya yang digunakan logika apakah saling bebas atau tidak.
- Aditif (Saling menjumlahkan).  
Artinya data yang dianalisis merupakan data interval/rasio

d) Menghitung variabilitas dari seluruh sampel. Pengukuran total variabilitas atas data dapat dikelompokkan menjadi tiga bagian, berikut rumus dalam Anova:

- Total of sum squares (SSt) – jumlah kuadrat total (jkt). Merupakan jumlah kuadrat selisih antara skor individual dengan rata-rata totalnya.

$$JKT = \sum_{i=1}^r \sum_{j=1}^k \sum_{m=1}^n x_{ijm}^2 - \frac{T^{***2}}{rkn}$$

**Keterangan:**

k = banyaknya kolom

r = Banyaknya baris

n = banyak ulangan

x<sub>ijm</sub> = data pada baris ke-i, kolom ke-j dan ulangan ke-m

T<sup>\*\*\*</sup> = Total (jumlah) seluruh pengamatan

- Sum Square Between column – jumlah kuadrat kolom (jkk). Variansi rata-rata kelompok sampel terhadap rata-rata keseluruhannya. Variansi di sini lebih terpengaruh karena adanya perbedaan perlakuan antar kelompok.

$$JKK = \sum_{j=1}^k \frac{T_{.j}^2}{rn} - \frac{T_{...}^2}{rkn}$$

**Keterangan**

T\*j\* = Total (jumlah) ulangan pada kolom ke-j

- Sum Square Between row – jumlah kuadrat baris (jkb). Variansi rata-rata kelompok sampel terhadap rata-rata keseluruhannya. Variansi di sini lebih terpengaruh karena adanya perbedaan perlakuan antar kelompok.

$$JKB = \sum_{i=1}^r \frac{T_{i.}^2}{kn} - \frac{T_{...}^2}{rkn}$$

**Keterangan**

Ti\*\* = Total (jumlah) ulangan pada baris ke-i

- Interaksi JK[BK]  
Variansi rata-rata kelompok interaksi baris dan kolom terhadap rata-rata keseluruhannya. Variansi di sini lebih terpengaruh karena adanya perbedaan perlakuan antar kelompok.

$$JK[BK] = \frac{\sum_{i=1}^r \sum_{j=1}^k T_{ij}^2}{n} - \frac{\sum_{i=1}^r T_{i.}^2}{kn} - \frac{\sum_{j=1}^k T_{.j}^2}{rn} + \frac{T_{...}^2}{rkn}$$

- Sum Square within(SSw) – jumlah kuadrat galat (jkg). Variansi yang ada dalam masing-masing kelompok. Banyaknya variansi akan tergantung pada banyaknya kelompok, dan variansi di sini tidak terpengaruh / tergantung oleh perbedaan perlakuan antar kelompok.

$$JKG = JKT - JKK - JKB - JK[BK]$$

e) Menghitung derajat kebebasan (degree of freedom). Derajat kebebasan atau degree of freedom (dilambangkan dengan v, dof, atau db) dalam ANOVA akan sebanyak variabilitas. Oleh karena itu, ada tiga macam derajat kebebasan yang akan kita hitung:

- Derajat kebebasan untuk JKT merupakan derajat kebebasan dari Jumlah kuadrat total (JKT) ini akan kita lambangkan dengan dof JKT.

$$\mathbf{db\ JKT = rkn - 1}$$

- Derajat kebebasan untuk JKK merupakan derajat kebebasan dari Jumlah kuadrat kolom (JKK) ini akan kita lambangkan dengan dof JKK.

$$\mathbf{db\ JKK = k-1}$$

- Derajat kebebasan untuk JKB merupakan derajat kebebasan dari Jumlah kuadrat baris (JKB) ini akan kita lambangkan dengan dof JKB.

$$\mathbf{db\ JKB = r-1}$$

- Derajat kebebasan untuk JK[BK] merupakan derajat kebebasan dari Jumlah kuadrat interaksi baris dan kolom JK[BK] ini akan kita lambangkan dengan dof JK[BK].

$$\mathbf{db\ JK[BK] = [r-1][k-1]}$$

- Derajat kebebasan untuk JKG Merupakan derajat kebebasan dari Jumlah kuadrat galat (JKG) ini akan kita lambangkan dengan dof JKG

$$\mathbf{db\ JKG = rk[n- 1]}$$

f) Menghitung variance antar kelompok dan variance dalam kelompok. Variance dalam ANOVA, baik untuk antar kelompok maupun dalam kelompok sering disebut dengan kuadrat tengah atau deviasi rata-rata kuadrat (mean squared deviation) dan dilambangkan dengan MS atau KT. Dengan demikian, maka mean squared deviation masing-masing dapat dicari dengan rumus sebagai berikut:

- $\mathbf{KTK = JKK / db\ JKK}$
- $\mathbf{KTB = JKB / db\ JKB}$

- $KTG = JKG / db\ JKG$
- $KT[BK] = JK[BK] / db\ JK[BK]$

g) Menghitung F hitung menghitung nilai distribusi F (Fhitung) berdasarkan perbandingan variance antar kelompok dan variance dalam kelompok. Fhitung ada tiga karena hipotesis ada tiga, sehingga setiap f hitung menjawab hipotesis.

- $F_{hitung} (kolom) = KTK/KTG$
- $F_{hitung} (baris) = KTB/KTG$
- $F_{hitung} (interaksi) = KT[BK]/JKG$

h) Menghitung F tabel selain itu, F berdasarkan tabel (Ftabel) juga dihitung, berdasarkan nilai derajat kebebasan (langkah ke-4) menggunakan tabel distribusi-F. Jangan lupa untuk mencantumkan gambar posisi Fhitung dan Ftabel dalam grafik distribusi-F.

i) Membandingkan F hitung dengan F tabel :

- Jika  $F_{hitung} > F_{tabel}$  : tolak  $H_0$
- Jika  $F_{hitung} \leq F_{tabel}$  : terima  $H_0$

j) Buat kesimpulan, sesuai dengan kasus awal yang ditanyakan. Simpulkan, apakah perlakuan (treatment) memiliki efek yang signifikan pada sampel data atau tidak. Jika hasil tidak signifikan, berarti seluruh rata-rata sampel adalah sama. Jika perlakuan menghasilkan efek yang signifikan, setidaknya satu dari rata-rata sampel berbeda dari rata-rata sampel yang lain.

#### 4. Contoh 1 Kasus ANOVA dua arah dengan interaksi:

Terdapat 4 metode diet, 3 kelompok umur dan 3 ulangan. Berikut adalah data rata-rata penurunan berat badan setelah 1 bulan melakukan diet. Ujilah apakah penurunan berat badan sama untuk setiap metode diet, kelompok umur dan interaksi dengan taraf uji 5 %?

Umur	Penurunan Berat Badan (Kg)			
	Metode 1	Metode 2	Metode 3	Metode 4
< 20 tahun	5	0	3	4

#1	4	2	4	2
#2	5	1	8	2
#3				
20-40 tahun				
#1	5	4	2	5
#2	6	2	2	3
#3	2	1	4	2
> 40 tahun				
#1	4	5	2	6
#2	4	5	1	4
#3	5	0	2	4

### **Solusi kasus ANOVAdua arah dengan interaksi**

#### a) Merumuskan Hipotesis

- Hipotesis ANOVAkolom

$H_0: a_1 = a_2 = \dots = a_k$ , Tidak ada perbedaan yang nyata antara rata-rata hitung dari kategori Metode(kolom)

$H_1: a_1 \neq a_2 \neq \dots \neq a_k$ , Ada perbedaan yang nyata antara rata-rata hitung dari kategori Metode (kolom)

- Hipotesis ANOVAbaris

$H_0: b_1 = b_2 = \dots = b_j$ , Tidak ada perbedaan yang nyata antara rata-rata hitung dari kategori umur (baris)

$H_1: b_1 \neq b_2 \neq \dots \neq b_j$ , Ada perbedaan yang nyata antara rata-rata hitung dari kategori umur (baris)

- Hipotesis interaksi

$H_0: (ab)_{11} = (ab)_{12} = \dots = (ab)_{kj}$ , Tidak ada interaksi antara variabel metode dan umur

$H_1: (ab)_{11} \neq (ab)_{12} \neq \dots \neq (ab)_{kj}$ , ada interaksi antara variabel metode dan umur

#### b) Identifikasi model.

Pertama. berdasarkan hipotesis yang digunakan yaitu membandingkan rata-rata lebih dari dua kelompok maka metode yang mungkin adalah Anova. kedua Sampel yang digunakan tiap kelompok sudah dikategorikan sehingga tipe ANOVA yang cocok adalah ANOVAdua arah. kemudian dari tiap kategori tersebut dilakukan pengulangan sehingga kita menggunakan ANOVAdua arah dengan interaksi.

- Memeriksa asumsi Anova.

Dalam metode ANOVA yang perlu diperhatikan ada empat seperti pada keterangan diatas. asumsi normal dan homogenitas antar varians kelompok harus terpenuhi. dalam contoh ini kita asumsikan asumsi terpenuhi karena kita fokus pada langkah-langkah ANOVA dua arah dengan interaksi. kemudian kelompok yang dianalisis berasal dari kelompok saling bebas. dan data yang digunakan merupakan data rasio. Setelah asumsi ini terpenuhi maka bisa lanjut ke perhitungan selanjutnya. kalau tidak ganti metode.

- c) Menyusun/mengkategorikan tabel data agar lebih mudah menghitungnya. Penghitungannya agak berbeda dengan jenis ANOVA yang lain. perhitungannya terpisah seperti berikut:

Umur	Penurunan Berat Badan (Kg)				Total Baris
	Metode 1	Metode 2	Metode 3	Metode 4	
< 20 tahun					
#1	5	0	3	4	T1** = 40
#2	4	2	4	2	
#3	5	1	8	2	
	T11* = 14	T12* = 3	T13* = 15	T14* = 8	
20-40 tahun					
#1	5	4	2	5	T2**
#2	6	2	2	3	
#3	2	1	4	2	
	T21* = 13	T22* = 7	T23* = 8	T24* = 10	
> 40 tahun					
#1	4	5	2	6	T3**=42
#2	4	5	1	4	
#3	5	0	2	4	
	T31* = 13	T32* = 10	T33* = 5	T34* = 14	
Total Kolom	T*1* = 40	T*2* = 20	T*3* = 28	T*4* = 32	Total T***=120

d) Perhitungan Tabel anova

Agar mempermudah perhitungan kita menggunakan tabel berikut,

Sumber Keragaman (SK)	Jumlah Kuadrat (JK)	Derajat Bebas (db)	Kuadrat Tengah (KT)	F hitung
Kolom (K)	JKK = 23,11	db JKK = 4-1 = 3	KTK = 7,70	F hitung = 3,04
Baris (B)	JKB = 0,67	db JKB = 3-1 = 2	KTB = 0,085	F hitung = 0,13
Interaksi (BK)	JK[BK] = 31,56	db JK[BK] = 2x3 = 6	KT[BK] = 5,26	F hitung = 0,28
Galat (G)	JKG = 60,67	db JKG = 3x4x2 = 24	KTG = 2,53	
Total (T)	JKT = 116	db JKT = [3x4x3] - 1 = 35		

e) Menghitung F tabel

- F table Kolom pada  $\alpha = 0.05$  db JKK=3 dan db JKG=4 adalah 3,01
- F table Baris pada  $\alpha = 0.05$  db JKB=2 dan db JKG=24 adalah 3,40
- F table Interaksi pada  $\alpha = 0.05$  db JK[BK]=6 dan db JKG=24 adalah 2,51

f) Kesimpulan :

Perhitungan menunjukkan bahwa rata-rata penurunan berat badan pada Baris [Kel. Umur] dan Interaksi tidak berbeda [masih dianggap sama] hal ini terlihat dari f tabel untuk baris dan interaksi lebih kecil dari f hitung sedangkan rata-rata penurunan berat badan dalam Kolom [metode diet] dapat dikatakan berbeda karena f tabel lebih besar dari f hitung.

## 5. Contoh 1 Kasus ANOVA dua arah dengan interaksi:

Seorang konsultan mesin dari perusahaan penyalur atau DEALER kendaraan diminta untuk mengkaji apakah ada perbedaan rata-rata efisiensi pemakaian BBM (kilometer/liter) antara tiga merek mobil. Disamping itu, ia diminta juga untuk mengkaji apakah ada perbedaan rata-rata efisiensi pemakaian BBM yang disebabkan oleh kapasitas mesin. Dari hasil pengumpulan data yang dilakukan konsultan tersebut diperoleh data sebagai berikut :

Kapasitas (ml)	Merek Mobil			Jumlah (Baris)
	A-1	A-2	A-3	
1300 ml	10	11	11	32
1500 ml	11	12	11	34
Jumlah (Kolom)	21	23	22	66

a) Uji Asumsi Data.

Dalam hal ini kita anggap sudah memenuhi asumsi, karena langkah uji asumsi data sama dengan sebelumnya. Namun uji asumsi ini wajib dilakukan agar analisis ANOVA yang diperoleh memiliki keakuratan yang baik.

b) Tabel Pengamatan

Kapasitas (ml)	Merek Mobil			Jumlah (Baris)
	A-1	A-2	A-3	
1300 ml	10	11	11	32
1500 ml	11	12	11	34
Jumlah (Kolom)	21	23	22	66

c) Melakukan Perhitungan

- Jumlah baris ( $r$ ) = 2
- Jumlah kolom ( $k$ ) = 3
- $T_{..} = 66$   
 $T_{1.} = 32$   
 $T_{2.} = 34$   
 $T_{.1} = 21$   
 $T_{.2} = 23$   
 $T_{.3} = 22$

$$JK_B = \sum_{i=1}^r \left( \frac{T_{i.}^2}{k} \right) - \left( \frac{T_{..}^2}{rk} \right) = \left( \frac{32^2}{3} \right) + \left( \frac{34^2}{3} \right) - \left( \frac{66^2}{6} \right) = \frac{1024}{3} + \frac{1156}{3} - \frac{2178}{3} = \frac{2180}{3} - \frac{2178}{3} = \frac{2}{3}$$

$$JK_T = \sum_{i=1}^r \sum_{j=1}^k x_{ij}^2 - \left( \frac{T_{..}^2}{rk} \right) = \sum_{i=1}^2 \sum_{j=1}^3 x_{ij}^2 - \left( \frac{T_{..}^2}{2 \cdot 3} \right) = (10^2 + 11^2 + 11^2 + 11^2 + 12^2 + 11^2) - \frac{66^2}{6}$$

$$= 728 - 726 = 2$$



$$JKK = \sum_{j=1}^k \left( \frac{T_{*j}^2}{r} \right) - \left( \frac{T_{**}^2}{rk} \right) = \left( \frac{21^2}{2} \right) + \left( \frac{23^2}{2} \right) + \left( \frac{22^2}{2} \right) - \left( \frac{66^2}{6} \right) = \frac{441}{2} + \frac{529}{2} + \frac{484}{2} - \frac{2178}{3}$$

$$= \frac{1454}{2} - \frac{2178}{3} = 1$$

$$JKB = JKT - JKK - JKB = 2 - 1 - \frac{2}{3} = \frac{1}{3}$$

- df (baris) = r-1 = 2-1 = 1
- df (kolom) = k-1 = 3-1 = 2
- df (galat) = (r-1)(k-1) = 1\*2 = 2
- df (total) = (2\*3 - 1) = 5

d) Merumuskan Hipotesis

- Hipotesis Uji untuk Kolom :

Ho : Rata-rata efisiensi pemakaian BBM ketiga merek mobil tersebut adalah sama

Ha : Rata-rata efisiensi pemakaian BBM ketiga merek mobil tersebut adalah berbeda

- Hipotesis Uji untuk Baris :

Ho : Rata-rata efisiensi pemakaian BBM kedua kapasitas mesin mobil tersebut adalah sama

Ha : Rata-rata efisiensi pemakaian BBM kedua kapasitas mesin mobil tersebut adalah berbeda

e) Menentukan Tarf Signifikansi

Kita pilih nilai signifikansi alpha 5%.

f) Membuat Hasil Perhitungan kedalam Tabel ANOVAdan Menentukan F-Tabel

Sumber Variasi	JK	df	MK	F
Antar Baris	JKB = 2/3	1	MKB = JKB/df = 2/3 : 1 = 2/3	$F_{\text{baris}} = \text{MKB}/\text{MKG} = 2/3 : 1/6 = 4$
Antar Kolom	JKK = 1	2	MKK = JKK/df = 1/2	$F_{\text{kolom}} = \text{MKK}/\text{MKG} = 1/2 : 1/6 = 3$
Error	JKG = 1/3	2	MKG = JKG/df = 1/3 : 2 = 1/6	
Total	JKT = 2	5	MKT = JKT/df = 2/5	

g) Menentukan Wilayah Kritis atau Kriteria Pengujian

Mencari nilai F-tabel untuk:

Baris : F-tabel =  $F(5\%; 1; 2) = 18,513$

Kolom : F-tabel =  $F(5\%; 2; 2) = 19,000$

h) Keputusan

Baris : F-hitung =  $4 < F\text{-tabel} = 18,513$ . Ho diterima

Kolom : F-Hitung =  $3 < F\text{-tabel} = 19,000$ . Ho diterima

i) Kesimpulan

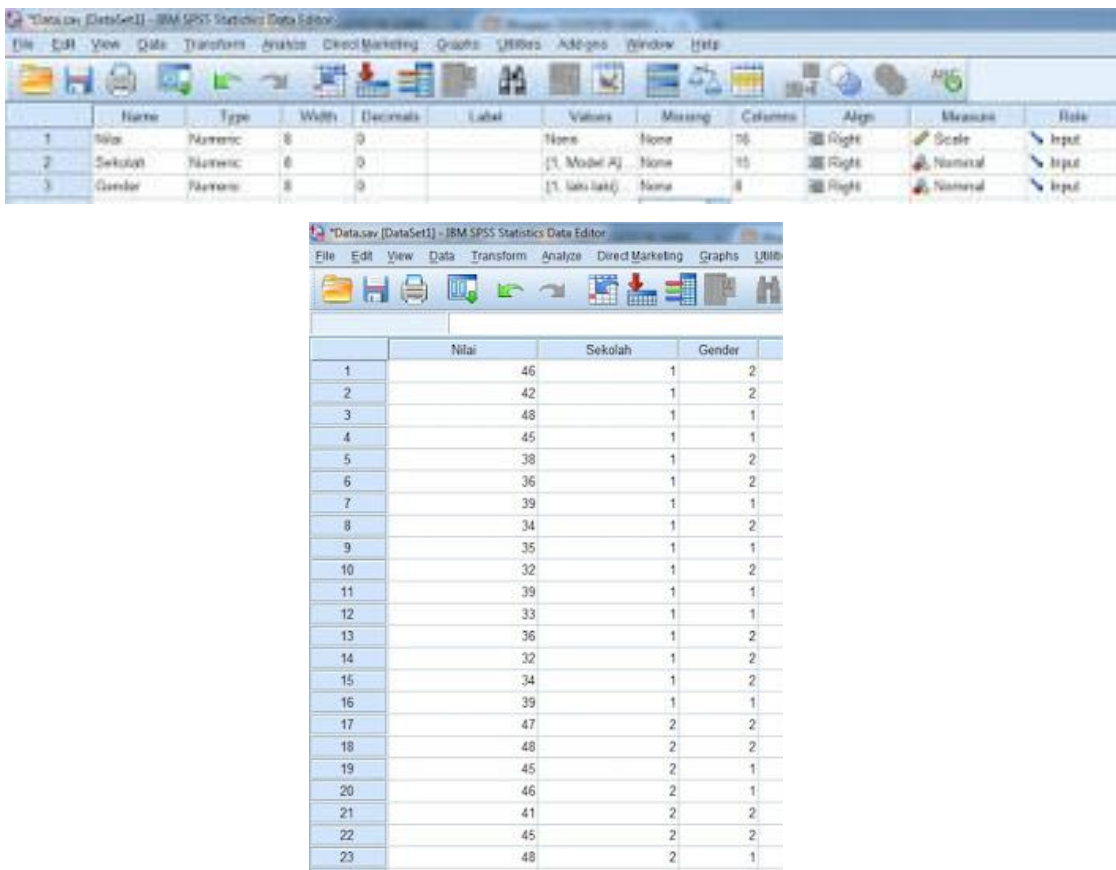
- Tidak ada perbedaan nyata rata-rata efisiensi pemakaian BBM terhadap ketiga merek mobil tersebut.
- Tidak ada perbedaan nyata rata-rata efisiensi pemakaian BBM terhadap kedua kapasitas mesin tersebut.

## 6. Cara Analisis dan Interpretasi SPSS

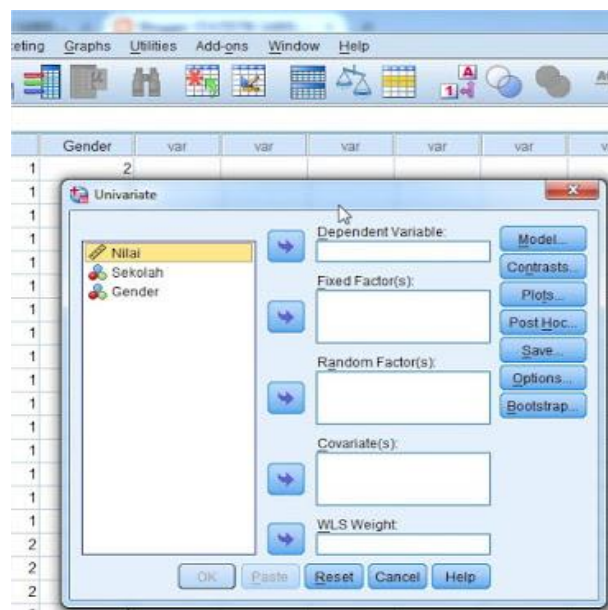
Seorang guru ingin mengetahui prestasi belajar siswa berdasarkan gender (laki-laki dan perempuan) setelah menerapkan model pembelajaran kooperatif pada materi yang sama. Sampel yang diambil 3 sekolah, dimana ketiga sekolah tersebut diterapkan model pembelajaran yang berbeda. Berikut hasil tes belajarnya:

Sekolah 1 (Model A)			Sekolah 2 (Model B)			Sekolah 3 (Model C)		
Nomor Urut	Nilai	Gender	Nomor Urut	Nilai	Gender	Nomor Urut	Nilai	Gender
1	46	Perempuan	1	47	Perempuan	1	58	Perempuan
2	42	Perempuan	2	48	Perempuan	2	25	Perempuan
3	48	Laki-laki	3	45	Laki-laki	3	34	Laki-laki
4	45	Laki-laki	4	46	Laki-laki	4	26	Laki-laki
5	38	Perempuan	5	41	Perempuan	5	62	Perempuan
6	36	Perempuan	6	45	Perempuan	6	44	Perempuan
7	39	Laki-laki	7	48	Laki-laki	7	43	Laki-laki
8	34	Perempuan	8	42	Perempuan	8	39	Perempuan
9	35	Laki-laki	9	41	Laki-laki	9	39	Laki-laki
10	32	Perempuan	10	45	Perempuan	10	44	Perempuan
11	39	Laki-laki	11	48	Laki-laki	11	32	Laki-laki
12	33	Laki-laki	12	33	Laki-laki	12	38	Laki-laki
13	36	Perempuan	13	39	Perempuan	13	42	Perempuan
14	32	Perempuan	14	47	Perempuan	14	39	Perempuan
15	34	Perempuan	15	42	Perempuan	15	41	Perempuan
16	39	Laki-laki	16	41	Laki-laki	16	33	Laki-laki

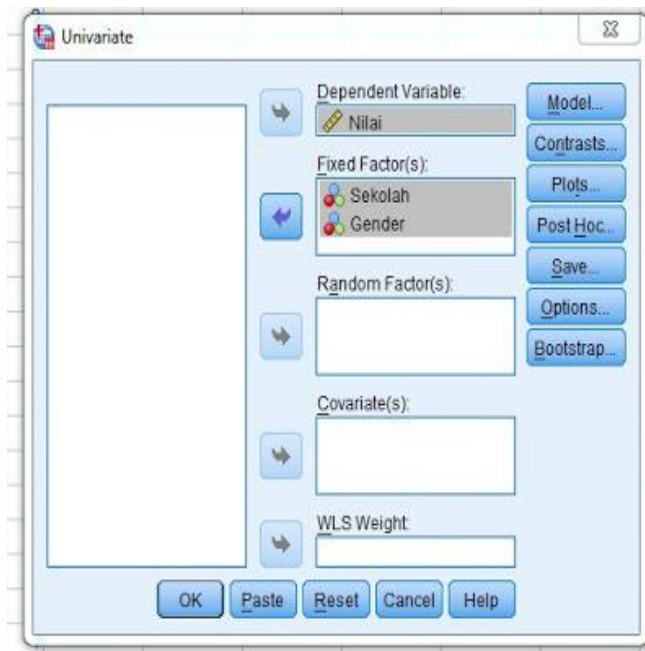
- Buka program SPSS dan masukan data yang akan di analisis



- Uji asumsi data berdistribusi normal atau tidak.
- Bila data sudah berdistribusi normal, langkah selanjutnya kita lanjut uji two way ANOVA yaitu: silahkan klik menu Analyze ---> General Linear Model ---> Univariate. Jika benar terlihat seperti pada gambar berikut:



- Masukkan Nilai kedalam kotak Dependent Variabel, sedangkan Sekolah dan Gender masukkan kedalam kotak Fixed Factor(s) terlihat seperti gambar berikut:



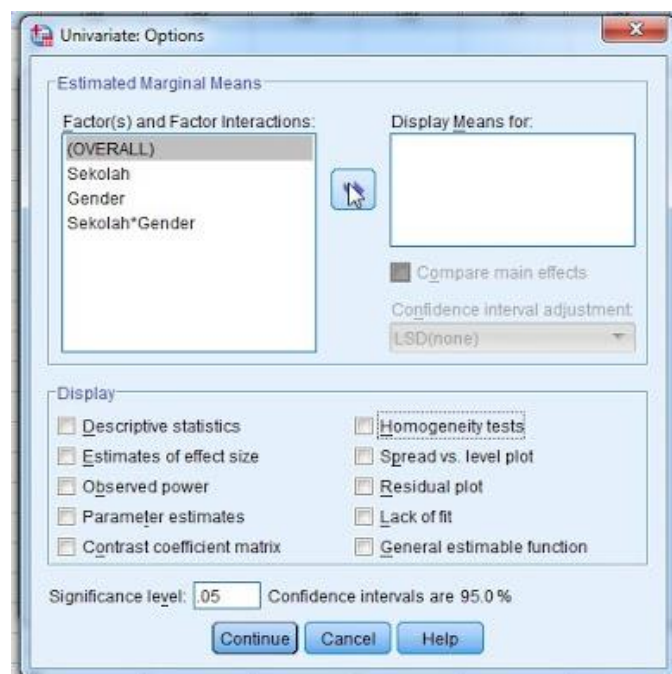
- Kemudian klik Plots, maka muncul kotak Univariate: Profile Plots seperti gambar dibawah, lalu masukkan variabel Gender kedalam kotak Horizontal Axis, sedangkan variabel Sekolah masukkan kedalam kotak Separate Lines. Seperti gambar berikut:



- Selanjutnya klik Add, maka akan muncul di dalam kotak Plots "Gender\*Sekolah" seperti pada gambar berikut:

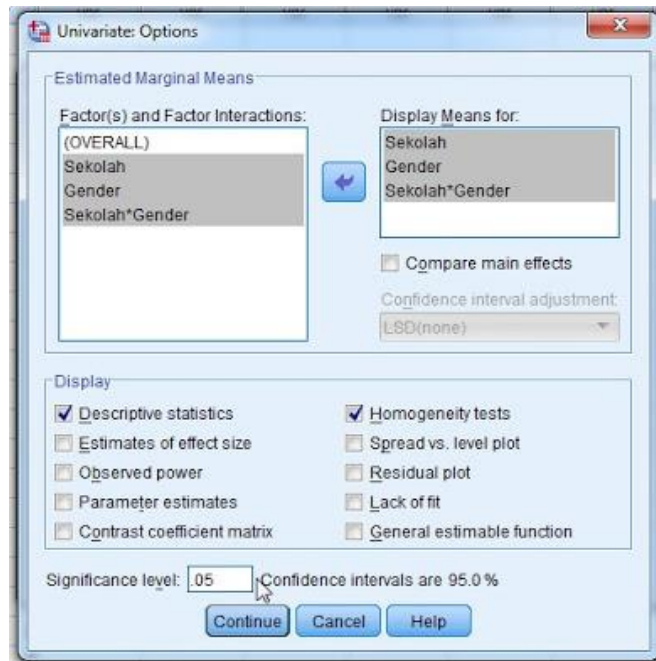


- Selanjutnya klik Options, maka muncul kotak Univariate: Options seperti gambar berikut:



- Masukkan Variabel Sekolah, Gender, Sekolah\*Gender kedalam kotak Display Means for. pada kotak Display centang Descriptive statistics dan Homogeneity

test, lalu perhatikan kotak Significance level: disini saya menggunakan taraf alpha 0.05. Nilai alpha ini di isi sesuai dengan nilai kepercayaan dari peneliti masing-masing dan tergantung jenis penelitiannya. jika sudah di isi maka terlihat seperti gambar berikut:



- **Output**

		Value Label	N
Sekolah	1	Model A	16
	2	Model B	16
	3	Model C	16
Gender	1	laki-laki	21
	2	perempuan	27

Dari gambar Between-subjects Factor. Kita dapat mengetahui jumlah sekolah ada 3 dan ketiga sekolah menggunakan model yang berbeda yaitu sekolah 1 menggunakan model A yang ada dalam kolom Value Label, begitu juga untuk sekolah 2 dan 3. sedangkan banyak siswa setiap sekolah sama yaitu 16 siswa yang terdiri dari laki-laki dan perempuan. Jadi jumlah keseluruhan ada 48 siswa

didapat dari 21 siswa laki-laki dan 27 siswa perempuan. Berikut Analisis selanjutnya

**Descriptive Statistics**

Dependent Variable: Nilai

Sekolah	Gender	Mean	Std. Deviation	N
Model A	laki-laki	39.71	5.251	7
	perempuan	36.67	4.690	9
	Total	38.00	5.020	16
Model B	laki-laki	43.14	5.336	7
	perempuan	44.00	3.122	9
	Total	43.63	4.097	16
Model C	laki-laki	35.00	5.538	7
	perempuan	43.78	10.860	9
	Total	39.94	9.767	16
Total	laki-laki	39.29	6.141	21
	perempuan	41.48	7.623	27
	Total	40.52	7.029	48

Dari gambar Descriptive Statistics di atas, kita dapat mengetahui deskripsi hasil dari nilai prestasi belajar siswa dari setiap sekolah berdasarkan gender. Misalnya: Sekolah 1 menerapkan model A diperoleh nilai rata-rata (mean) untuk siswa laki-laki sebesar 39,71, standar deviasi 5,251 dan jumlah siswa laki-laki ada 7. sedangkan siswa perempuan diperoleh nilai rata-rata 36,67, standar deviasi 4.690 dan jumlah siswa perempuan 9. Begitu juga untuk sekolah 2 dan 3. Dari sini kita belum bisa mengambil kesimpulan prestasi belajar dari setiap sekolah, karena hanya deskripsi data saja. Berikut Analisis selanjutnya

**Levene's Test of Equality of Error Variances<sup>a</sup>**

Dependent Variable: Nilai

F	df1	df2	Sig.
1.330	5	42	.270

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + Sekolah + Gender + Sekolah \* Gender

Dari Gambar Levene's Test of Equality of Error Variances (homogenitas), diperoleh nilai Sig 0.270 atau nilai sig ini  $> 0.05$  (syarat homogenitas), artinya ketiga sampel mempunyai varians yang sama (sudah memenuhi syarat uji Anova). Berikut analisis selanjutnya

**Tests of Between-Subjects Effects**

Dependent Variable: Nilai

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	604.138 <sup>a</sup>	5	120.828	2.954	.023
Intercept	77056.953	1	77056.953	1883.988	.000
Sekolah	251.394	2	125.697	3.073	.057
Gender	56.953	1	56.953	1.392	.245
Sekolah * Gender	285.894	2	142.947	3.495	.039
Error	1717.841	42	40.901		
Total	81135.000	48			
Corrected Total	2321.979	47			

a. R Squared = .260 (Adjusted R Squared = .172)

Dari gambar Test of Between-Subjects Effects di atas merupakan hasil akhir dari uji two way ANOVA yang kita butuhkan. Berikut penjelasannya:

**\* Corrected Model**

Dari koreksi model ini kita dapat mengetahui seberapa besar pengaruh variabel independent (variabel bebas) terhadap variabel dependent (variabel terikat). Dalam hal ini variabel independent yaitu faktor-faktor yang akan diukur oleh peneliti (sekolah, Gender dan sekolah\*Gender) untuk menentukan hubungan antara variabel dependent (nilai yang akan di amati). Dari tabel di atas bisa kita lihat berdasarkan nilai (sig), bila nilai sig  $< 0.05$  yaitu ( $0.023 < 0.05$ ) berarti model yang diperoleh valid.

**\* Intercept**

Nilai intercept dalam hal ini merupakan nilai siswa pada variabel nilai yang berkontribusi pada nilai itu sendiri tanpa dipengaruhi oleh variabel independent, artinya berubah nilai pada variabel dependent tidak ada pengaruh sedikit pun oleh variabel independent. Dari tabel di atas bisa kita lihat berdasarkan nilai (sig), bila nilai sig  $< 0.05$  yaitu ( $0.000 < 0.05$ ) berarti intercept ini berkontribusi secara signifikan



**\* Sekolah**

Berpengaruh atau tidaknya sekolah terhadap hasil belajar siswa ditandai dari nilai signifikan, dari tabel di atas nilai sig 0.057 atau nilai ( $0.57 > 0.05$ ) dalam kasus ini berarti sekolah tidak terlalu berpengaruh signifikan terhadap hasil belajar siswa.

**\* Gender**

Berpengaruh atau tidaknya gender terhadap hasil belajar ditandai dari nilai signifikan, dari tabel di atas nilai sig 0.245 atau nilai ( $0.245 > 0.05$ ) dalam kasus ini berarti gender tidak terlalu berpengaruh signifikan terhadap hasil belajarnya.

**\* Sekolah\*Gender**

Uji ini bertujuan untuk mengetahui apakah ada hubungan yang signifikan antara 2 faktor, dalam kasus ini kita akan menguji ada atau tidaknya interaksi antara model pembelajaran dengan kelompok gender.

- **Hipotesis:**

H0: tidak ada interaksi antara model pembelajaran terhadap kelompok gender

H1: ada interaksi antara model pembelajaran terhadap kelompok gender

- **Pengambilan keputusan:**

jika  $F_{hitung} < F_{tabel}$  atau nilai sig  $> 0.05$ , maka H0 diterima

jika  $F_{hitung} > F_{tabel}$  atau nilai sig  $< 0.05$ , maka H0 di tolak, jadi terimalah H1

- **Keputusan:**

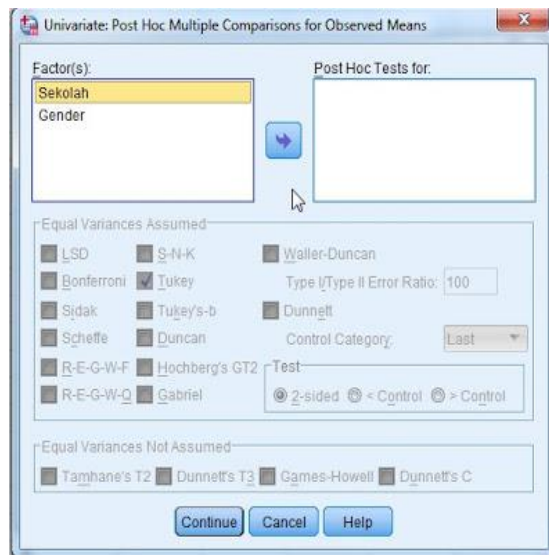
Dari tabel di atas terlihat bahwa nilai  $F_{hitung}$  3.495 dan  $F_{tabel}$  3.19 ( $F_{tabel}$  lihat tabel distribusi F untuk anova) atau nilai signifikan yang di peroleh dari tabel di atas sebesar 0.039. Jadi dapat disimpulkan: ada interaksi antara model pembelajaran terhadap kelompok gender. Karena ada interaksi antara model pembelajaran terhadap kelompok gender dari setiap sekolah, maka disini kita perlu uji lanjut Post Hoc. Berikut cara uji lanjut Post Hoc ANOVA2 arah

- **Uji lanjut**

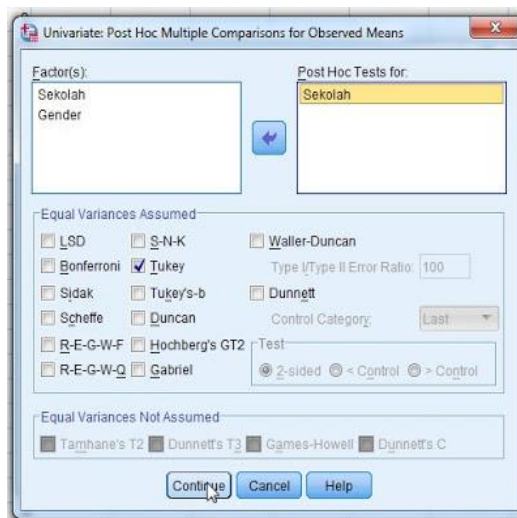
Kembali lagi ke langkah 1 sampai dengan langkah ke 19 di atas, setelah klik Continue pada langkah ke 19. maka muncul seperti gambar berikut:



- Klik Post Hoc pada gambar, maka muncul seperti gambar berikut:



- Pada kotak Factor(s) pindahkan sekolah ke dalam kotak Post Hoc Test for, disini kita menggunakan uji Tukey. Seperti terlihat pada gambar berikut:



- Setelah itu Klik Continue, lalu klik OK. Silahkan lihat hasilnya pada Post Hoc Test (ada dibawah pada hasil output) atau seperti pada gambar berikut:

**Post Hoc Tests**

**Sekolah**

**Multiple Comparisons**

Dependent Variable: Nilai  
Tukey HSD

(I) Sekolah	(J) Sekolah	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Model A	Model B	-5.63*	2.261	.044	-11.12	-.13
	Model C	-1.94	2.261	.670	-7.43	3.56
Model B	Model A	5.63*	2.261	.044	.13	11.12
	Model C	3.69	2.261	.244	-1.81	9.18
Model C	Model A	1.94	2.261	.670	-3.56	7.43
	Model B	-3.69	2.261	.244	-9.18	1.81

Based on observed means.  
The error term is Mean Square(Error) = 40.901.  
\*. The mean difference is significant at the .05 level.

- Analisis hasil uji lanjut Post Hoc Tukey  
Dari tabel Multiple Comparisons, coba perhatikan dalam kolom Mean Difference (I-J) itu ada yang berbintang (\*) dan ada yang tidak berbintang (\*). Dimana yang bertanda \* itu artinya model tersebut berbeda secara signifikan dibandingkan dengan model-model yang lain.
- Kesimpulan: Perbedaan secara signifikan yaitu pada sekolah 1 model A dan sekolah 2 model B

## G. MULTIVARIATE WAY ANOVA

Salah satu teknik dalam analisis multivariat yaitu Multivariate Analysis of Variance (MANOVA) atau Analisis Variansi Multivariat. MANOVA mempunyai pengertian sebagai suatu teknik statistik yang digunakan untuk menghitung pengujian signifikansi perbedaan rata-rata secara bersamaan antara kelompok untuk dua atau lebih variabel terikat. MANOVA adalah generalisasi dari ANOVA untuk situasi dimana terdapat beberapa variabel terikat. MANOVA memiliki kelebihan bila dibandingkan ANOVA.

Penggunaan MANOVA memiliki keunggulan yaitu mampu menganalisis semua variabel terikat secara simultan, sehingga dapat memperkecil kesalahan tipe I ( $\alpha$ ) dalam pengambilan keputusan uji statistik. MANOVA merupakan uji yang kuat dan salah satu cara menghindari peningkatan alpha sebagai "filter" untuk ANOVA. MANOVA mampu mendeteksi dan mengungkapkan perbedaan yang tidak ditampilkan ANOVA-ANOVA pada masing-masing variabel terikat secara terpisah. MANOVA juga mampu mengoreksi hasil ANOVA palsu yang disebabkan peningkatan alpha saat melakukan beberapa tes ANOVA pada masing-masing variabel terikat. Hal ini berarti dengan menggunakan MANOVA, peneliti dapat meningkatkan kesempatan untuk menemukan perubahan sebagai akibat dari perlakuan yang berbeda dan interaksinya. Dengan demikian, temuan-temuan hasil penelitian akan semakin kaya dan sangat berguna bagi perkembangan ilmu pengetahuan.

Pearson dan Fisher adalah orang pertama yang memperkenalkan analisis multivariat. Mereka merupakan orang yang memiliki pemahaman paling mendasar dalam aplikasi statistika. Salah satu teknik dalam analisis multivariat yaitu Multivariate Analysis of Variance (MANOVA) atau Analisis Variansi Multivariat. MANOVA mempunyai pengertian sebagai suatu teknik statistik yang digunakan untuk menghitung pengujian signifikansi perbedaan rata-rata secara bersamaan antara kelompok untuk dua atau lebih variabel terikat. MANOVA adalah generalisasi dari ANOVA untuk situasi dimana terdapat beberapa variabel terikat. Menurut Field, MANOVA memiliki kemiripan asumsi dengan ANOVA tetapi diperluas untuk kasus multivariat. Adapun asumsi yang harus dipenuhi pada MANOVA yaitu:

1. Independen: Pengamatan harus independen secara statistik. Dipenuhinya persyaratan ini dimaksudkan agar perlakuan yang diberikan kepada setiap sampel, independen antara satu dengan lainnya.
2. Sampel acak: Dalam statistika untuk hal pengambilan sampel harus dilakukan secara random (acak) dari populasinya atau dengan kata lain menggunakan teknik probabilitas. Selain itu, data yang diukur (variabel terikat) dalam penelitian berskala interval.
3. Normalitas multivariat: Dalam ANOVA, diasumsikan bahwa variabel terikat berdistribusi normal di dalam masing-masing kelompok, sedangkan dalam kasus MANOVA diasumsikan bahwa variabel terikat (secara bersama) berdistribusi normal multivariat di dalam kelompok.

4. Homogenitas matriks kovariansi: Dalam ANOVA, diasumsikan bahwa variansi pada setiap kelompok sama (homogenitas variansi). Sedangkan dalam MANOVA, diasumsikan benar untuk setiap variabel terikat memiliki variansi yang sama pada setiap kelompok, selain itu diasumsikan juga bahwa korelasi antara manapun variabel terikat adalah sama dalam semua kelompok. Asumsi ini diuji dengan pengujian apakah matriks kovariansi populasi dari kelompok yang berbeda adalah sama.

Berdasarkan hal tersebut, maka perlu dilakukan uji persyaratan sebelum digunakan MANOVA. Uji persyaratan ini pada prinsipnya bertujuan untuk memastikan bahwa MANOVA bisa digunakan dan hasil pengujiannya dapat diinterpretasikan dengan tepat. Namun dalam praktiknya, asumsi yang harus diuji yaitu populasi-populasi berdistribusi normal multivariat dan matriks kovariansi populasi-populasi sama, sedangkan asumsi lainnya tidak perlu dilakukan pengujian.

Terdapat beberapa statistik uji MANOVA yaitu Wilks' Lambda, Pillai, Lawley-Hotelling, dan Roy's Largest Root. Banyak software statistik menyajikan perhitungan keempat statistik uji MANOVA tersebut, dan biasanya keempat statistik uji tersebut menghasilkan kesimpulan yang sama. Dalam kasus ketika keempat statistik uji tersebut menghasilkan kesimpulan yang berbeda dalam hal menerima dan menolak hipotesis, cara yang dapat dilakukan yaitu menguji nilai eigen dan matriks kovariansi serta mengevaluasi permasalahan kesimpulan dalam karakteristik statistik uji. Hal ini dengan pertimbangan bahwa secara historis, Wilks' Lambda telah memainkan peran dominan dalam uji-uji signifikansi dalam MANOVA. Wilks' Lambda merupakan statistik uji MANOVA yang pertama kali diturunkan dan memiliki pendekatan  $\chi^2$  dan F yang terkenal.

Sebelum melakukan perhitungan statistik uji MANOVA, terlebih dahulu dihitung matriks Sum-of-Squares and Cross-Products (SSCP).

Tabel 1. Rangkuman Matriks SSCP MANOVA Dua Jalan Sel Tak Sama

Sumber Variasi	Matriks SSCP	Derajat Kebebasan
Faktor A	$H_A = (\hat{A}\hat{M})' [A(W'W)^{-1}A']^{-1} \hat{A}\hat{M}$	$v_{HA} = a - 1$
Faktor B	$H_B = (\hat{B}\hat{M})' [B(W'W)^{-1}B']^{-1} \hat{B}\hat{M}$	$v_{HB} = b - 1$
Interaksi	$H_{AB} = (\hat{C}\hat{M})' [C(W'W)^{-1}C']^{-1} \hat{C}\hat{M}$	$v_{HAB} = (a - 1)(b - 1)$
Galat	$E = (X - W\hat{M})'(X - W\hat{M})$ atau	$v_E = N - ab$

$$E = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{n_{ij}} (x_{ijk} - \bar{x}_{ij}) (x_{ijk} - \bar{x}_{ij})'$$

$$\text{Total } T = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{n_{ij}} (x_{ijk} - \bar{x}) (x_{ijk} - \bar{x})' \quad N - 1$$

**Keterangan:**

a : banyaknya baris (klasifikasi faktor A)

b : banyaknya kolom (klasifikasi faktor B)

N : banyaknya seluruh data amatan = n11 + n12 + ... + nij

W : rank penuh sebesar ab merupakan matriks ukuran N×(ab)

W'W : matriks diagonal ukuran (ab)×(ab) dengan entri nij = diag(n11, n12, ..., nab)

X : matriks ukuran N×p dengan entri seluruh data amatan : matriks ukuran (ab)×p dengan entri  $M_{ij} = \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} x_{ijk}$

A : matriks kontras untuk efek utama baris (faktor A)

B : matriks kontras untuk efek utama kolom (faktor B)

C : matriks kontras untuk efek sederhana (interaksi)

Sebagai contoh perhitungan matriks kontras, ditinjau kasus pada pada desain faktorial 2×3 dengan 2 buah variabel terikat. Perhitungan matriks kontras pada kasus tersebut sebagai berikut:

Telah diketahui

$$M = (\mu_{(1)}, \mu_{(2)}) = (\mu_{11}, \mu_{12}, \mu_{13}, \mu_{21}, \mu_{22}, \mu_{23})'$$

Faktor A memiliki 2 nilai yang bersesuaian dengan 3 kolom faktor B. Untuk 2 nilai pada faktor A mempunyai 1 derajat kebebasan, sehingga memerlukan 1 kontras. Efek utama faktor A diperoleh dengan membandingkan baris pertama dengan baris kedua. Ini dapat dilakukan dengan satu kontras ortogonal.

$$\begin{aligned} a' M &= (\mu_{11} + \mu_{12} + \mu_{13}) - (\mu_{21} + \mu_{22} + \mu_{23}) \\ &= \mu_{11} + \mu_{12} + \mu_{13} - \mu_{21} - \mu_{22} - \mu_{23} \\ &= (1 \ 1 \ 1 \ -1 \ -1 \ -1) M \\ A &= a' = (1 \ 1 \ 1 \ -1 \ -1 \ -1) \end{aligned}$$

Faktor B memiliki 3 nilai yang bersesuaian dengan 2 baris faktor A. Untuk 3 nilai pada faktor B mempunyai 2 derajat kebebasan, sehingga memerlukan 2 kontras. Efek utama faktor B diperoleh dengan membandingkan kolom pertama dengan dua kolom lainnya, dan kemudian membandingkan kolom kedua dengan kolom ketiga. Ini dapat dilakukan dengan dua kontras ortogonal berikut.

$$\begin{aligned}
 b_1' M &= 2(\mu_{11} + \mu_{21}) - (\mu_{12} + \mu_{22}) - (\mu_{13} + \mu_{23}) & b_2' M &= (\mu_{12} + \mu_{22}) - (\mu_{13} + \mu_{23}) \\
 &= 2\mu_{11} - \mu_{12} - \mu_{13} + 2\mu_{21} - \mu_{22} - \mu_{23} & &= \mu_{12} - \mu_{13} + \mu_{22} - \mu_{23} \\
 &= (2 \quad -1 \quad -1 \quad 2 \quad -1 \quad -1)M & &= (0 \quad 1 \quad -1 \quad 0 \quad 1 \quad -1)M
 \end{aligned}$$

$$\mathbf{B} = \begin{pmatrix} 2 & -1 & -1 & 2 & -1 & -1 \\ 0 & 1 & -1 & 0 & 1 & -1 \end{pmatrix}$$

Kontras  $a'M$ ,  $b_1'M$ , dan  $b_2'M$  membandingkan unsur-unsur di dalam rerata vektor  $M$ . Interaksi AB mempunyai 2 derajat kebebasan, sehingga memerlukan 2 kontras berhubungan yang diperoleh dengan mengalikan unsur-unsur bersesuaian dari  $a_1$  dan  $b_1$  serta  $a_1$  dan  $b_2$ . Perkalian menurut unsur dari dua vektor tersebut dinamakan Hadamard product.  $ij \mu$

$$\begin{aligned}
 c_1' &= (2 \quad -1 \quad -1 \quad -2 \quad 1 \quad 1) & c_2' &= (0 \quad 1 \quad -1 \quad 0 \quad -1 \quad 1) \\
 \mathbf{C} &= \begin{pmatrix} 2 & -1 & -1 & -2 & 1 & 1 \\ 0 & 1 & -1 & 0 & -1 & 1 \end{pmatrix}
 \end{aligned}$$

Untuk kasus lainnya dapat menyesuaikan prosedur perhitungan matriks kontras sebagaimana yang telah dijelaskan. Selanjutnya, perhitungan Wilks' untuk pengujian hipotesis terkait interaksi dan efek utama sebagai berikut  $\square$

$$\Lambda_A = \frac{|\mathbf{E}|}{|\mathbf{E} + \mathbf{H}_A|} \quad \Lambda_B = \frac{|\mathbf{E}|}{|\mathbf{E} + \mathbf{H}_B|} \quad \Lambda_{AB} = \frac{|\mathbf{E}|}{|\mathbf{E} + \mathbf{H}_{AB}|}$$

## 1. Prosedur MANOVA

Setelah mengetahui statistik uji MANOVA khususnya Wilks' Lambda, berikut prosedur pengujian MANOVA dua jalan dengan sel tak sama

a) Menetapkan Hipotesis

- Hipotesis perbedaan efek antarbaris

$$H_{0A} : \mathbf{AM} = \mathbf{O} \text{ (arti } H_{0A}: \begin{bmatrix} \mu_{11\bullet} \\ \mu_{21\bullet} \\ \vdots \\ \mu_{p1\bullet} \end{bmatrix} = \begin{bmatrix} \mu_{12\bullet} \\ \mu_{22\bullet} \\ \vdots \\ \mu_{p2\bullet} \end{bmatrix} = \dots = \begin{bmatrix} \mu_{1a\bullet} \\ \mu_{2a\bullet} \\ \vdots \\ \mu_{pa\bullet} \end{bmatrix} )$$

$H_{1A} : \mathbf{AM} \neq \mathbf{O}$  (arti  $H_{1A}$ : paling sedikit ada dua matriks rerata baris yang tidak sama)

- Hipotesis perbedaan efek antarkolom

$$H_{0B} : \mathbf{BM} = \mathbf{O} \text{ (arti } H_{0B}: \begin{bmatrix} \mu_{1\bullet 1} \\ \mu_{2\bullet 1} \\ \vdots \\ \mu_{p\bullet 1} \end{bmatrix} = \begin{bmatrix} \mu_{1\bullet 2} \\ \mu_{2\bullet 2} \\ \vdots \\ \mu_{p\bullet 2} \end{bmatrix} = \dots = \begin{bmatrix} \mu_{1\bullet b} \\ \mu_{2\bullet b} \\ \vdots \\ \mu_{p\bullet b} \end{bmatrix} )$$

$H_{1B} : \mathbf{BM} \neq \mathbf{O}$  (arti  $H_{1B}$ : paling sedikit ada dua matriks rerata kolom yang tidak sama)

- Hipotesis interaksi

$H_{0AB} : \mathbf{CM} = \mathbf{O}$  (arti  $H_{0AB}$ : tidak ada interaksi baris dan kolom terhadap variabel-variabel terikat)

$H_{1AB} : \mathbf{CM} \neq \mathbf{O}$  (arti  $H_{1AB}$ : ada interaksi baris dan kolom terhadap variabel-variabel terikat)

b) Taraf Signifikansi ( $\alpha$ )

c) Menentukan Statistik Uji

Untuk melakukan uji hipotesis menggunakan MANOVA dua jalan mengikuti formula statistik uji Wilks' sebagaimana disajikan pada Tabel 2. Langkah awal dalam menguji hipotesis ini adalah menentukan statistik uji yang tepat sesuai dengan model analisis. Hal yang perlu diperhatikan dalam memilih formula statistik uji Wilks' yaitu banyaknya variabel terikat dan desain faktorial yang digunakan. Sebagai contoh penentuan formula statistik uji Wilks', ditinjau kasus pada desain faktorial  $2 \times 3$  dengan 2 buah variabel terikat. Pada contoh ini terdapat dua variabel terikat ( $p = 2$ ). Ketika untuk  $p = 2$  dan sebarang  $v_H$ , maka formula statistik uji Wilks' yang tepat adalah:

$$F = \left( \frac{1 - \sqrt{\Lambda}}{\sqrt{\Lambda}} \right) \left( \frac{v_E - 1}{v_H} \right) \sim F(2v_H, 2(v_E - 1))$$



- ini perlu adanya modifikasi pada nilai-nilai dan H v E

1) Pengaruh Faktor A

Karena  $v_{HA} = a - 1$  dan  $v_E = N - ab$ , maka

$$F_A = \left( \frac{1 - \sqrt{\Lambda_A}}{\sqrt{\Lambda_A}} \right) \left( \frac{N - ab - 1}{a - 1} \right) \sim F(2(a - 1), 2(N - ab - 1))$$

2) Pengaruh Faktor B

Karena  $v_{HB} = b - 1$  dan  $v_E = N - ab$ , maka

$$F_B = \left( \frac{1 - \sqrt{\Lambda_B}}{\sqrt{\Lambda_B}} \right) \left( \frac{N - ab - 1}{b - 1} \right) \sim F(2(b - 1), 2(N - ab - 1))$$

3) Interaksi Antara Faktor A dan Faktor B

Karena  $v_{HAB} = (a - 1)(b - 1)$  dan  $v_E = N - ab$ , maka

$$F_{AB} = \left( \frac{1 - \sqrt{\Lambda_{AB}}}{\sqrt{\Lambda_{AB}}} \right) \left( \frac{N - ab - 1}{(a - 1)(b - 1)} \right) \sim F(2(a - 1)(b - 1), 2(N - ab - 1))$$

**Keterangan**

a : banyaknya baris

b : banyaknya kolom

N : banyaknya seluruh data amatan

: derajat kebebasan HA = a - 1 HA v

: derajat kebebasan HB = b - 1 HB v

: derajat kebebasan HAB = (a - 1)(b - 1) HAB v

: derajat kebebasan E = N - ab E v

Tabel 2. Tabel MANOVA

Sumber Keragaman	Matriks Jumlah Kuadrat dan Hasil Kali	Derajat Bebas
Perlakuan	$\sum_{i=1}^t r_i (\bar{y}_{i.} - \bar{y}_{..})(\bar{y}_{i.} - \bar{y}_{..})'$	$t - 1$
Galat	$\sum_{i=1}^t \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i.})(y_{ij} - \bar{y}_{i.})'$	$\sum_{i=1}^t n_i - t$
Total (terkoreksi dengan nilai tengah)	$\sum_{i=1}^t \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{..})(y_{ij} - \bar{y}_{..})'$	$\sum_{i=1}^t n_i - 1$

d) Daerah Kritis

Untuk masing-masing nilai F di atas, daerah kritisnya adalah:

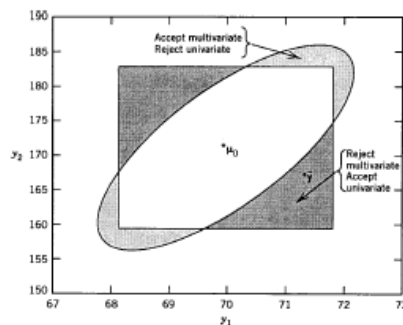
1) Daerah kritis  $F_A$  adalah  $DK = \{F | F > F_{\alpha; 2(a-1); 2(N-ab-1)}\}$

2) Daerah kritis  $F_B$  adalah  $DK = \{F | F > F_{\alpha; 2(b-1); 2(N-ab-1)}\}$

3) Daerah kritis  $F_{AB}$  adalah  $DK = \{F | F > F_{\alpha; 2(a-1)(b-1); 2(N-ab-1)}\}$

- e) Keputusan Uji:  $H_0$  ditolak jika  $F_{obs} \in DK_{obs}$ .
- f) Menentukan kesimpulan dari keputusan uji yang ada.

Jika uji multivariat yang didasarkan pada  $\lambda$  menolak  $H_0$ , ini dapat diikuti oleh uji F pada setiap p individual. Kita dapat merumuskan hipotesis perbandingan rerata menyilang k kelompok untuk setiap variabel, yaitu,  $H_{0r} : \mu_{1r} = \mu_{2r} = \dots = \mu_{kr}$ ,  $r = 1, 2, \dots, p$ . Itu tidak selalu berarti bahwa setiap uji F pada p individual variabel akan menolak  $H_{0r}$  yang bersesuaian. Sebaliknya, ada kemungkinan bahwa salah satu atau lebih dari F akan menolak  $H_{0r}$  ketika uji  $\lambda$  menerima  $H_0$ . Dalam kedua kasus, dimana uji multivariat dan uji univariat berbeda, kita menggunakan hasil uji multivariat daripada hasil uji univariat. Hal ini mirip hubungan antara uji Z2 dan uji z yang ditunjukkan pada gambar berikut.



Gambar 3. Daerah Penerimaan dan Penolakan Uji Univariat dan Uji Multivariat

## 2. Tahapan dalam MANOVA:

- Tahap 1: Tujuan MANOVA

Untuk menganalisis hubungan dependensi yang diwakili sebagai perbedaan dalam seperangkat ukuran dependen lintas seri kelompok yang dibentuk oleh satu atau lebih ukuran independent kategorial.

Untuk menyediakan pengertian yang mendalam mengenai sifat dan kekuatan sifat prediksi dari ukuran independen seperti juga keterkaitan dan perbedaan dalam ukuran multiple dependen. Terdapat 3 kategori permasalahan klasifikasi multivariate dilihat dari aspek penerapn MANOVA, yaitu :

- Multiple Univariate Questions.

Peneliti yang mempelajari pertanyaan-pertanyaan multiple univariate mengidentifikasi sejumlah variabel dependen secara terpisah yang akan dianalisis secara terpisah tetapi memerlukan beberapa kontrol terhadap kesalahan eksperimen. Dalam kejadian ini, MANOVA digunakan untuk

menguji apakah perbedaan menyeluruh ditemukan di antara kelompok, dan kemudian pengujian univariate yang terpisah dilakukan untuk mendapatkan setiap isu untuk masing-masing variabel dependen.

- Structured Multivariate Questions.

Peneliti berurusan dengan pertanyaan multivariat terstruktur yang dikumpulkan dari dua atau lebih ukuran dependen yang saling memiliki hubungan khusus. Situasi yang umum dalam kategori ini adalah ukuran yang berulang dimana ada multiple respon dari setiap subyek yang mungkin selama atau setelah pre test-post test memberikan stimulus, misalnya seperti iklan. Disini MANOVA menyediakan metode terstruktur untuk menspesifikasi perbandingan perbedaan antar kelompok terhadap seperangkat ukuran dependen dalam menjaga efisiensi secara statistik.

- Intrinsically Multivariate Questions.

Pertanyaan ini meliputi sejumlah ukuran dependen dimana mempertimbangkan prinsip bagaimana mereka berbeda sebagai keseluruhan lintas kelompok. Perbedaan ukuran dependen secara individual kurang menarik dibandingkan pengaruh kolektif. Keunggulan MANOVA tidak hanya dapat menguji perbedaan keseluruhan tetapi juga perbedaan antara kombinasi dari ukuran dependen yang tidak dapat muncul. Jenis pertanyaan ini bisa diselesaikan oleh MANOVA karena kemampuannya mendeteksi perbedaan multivariate walaupun pengujian univariate tunggal tidak menunjukkan perbedaan.

• Tahap 2: Disain Riset MANOVA

Membutuhkan ukuran sampel yang lebih besar dibandingkan univariate ANOVA dan harus melebihi batasan khusus pada setiap sel (kelompok). Analisis yang direkomendasikan adalah minimal ukuran sel 20 observasi. Sampel pada setiap sel harus lebih besar dari jumlah variabel dependen.

- Desain Faktorial – Analisis dengan dua treatment atau lebih.

Pemilihan Treatments, penggunaan yang umum dari disain faktorial meliputi pertanyaan riset yang berhubungan dengan dua atau lebih variabel independen nonmetrik terhadap seperangkat variabel dependen. Dalam masalah ini variabel independen dispesifikasikan dalam desain eksperimen

atau termasuk dalam desain eksperimen lapangan atau survey menggunakan kuesioner. Namun dalam beberapa hal treatment ditambahkan setelah desain analisis.

Treatment tambahan yang sering digunakan adalah blocking factor dimana karakteristik nonmetrik digunakan post hoc untuk segmentasi responden untuk mendapatkan homogenitas kelompok yang lebih besar dan mengurangi sumber varians MSW. Dengan demikian kemampuan uji statistik untuk mengidentifikasi perbedaan dapat ditingkatkan. Sebagai contoh misalnya pada iklan, sebelumnya laki-laki dianggap akan mempunyai reaksi yang berbeda dengan perempuan terhadap tampilan iklan dan jika gender ditambahkan sebagai blocking factor, perbedaan setiap pesan menjadi lebih jelas, sedangkan perbedaan akan menjadi samar jika laki-laki dan perempuan di asumsikan akan bereaksi dengan berbeda dan tidak terpisah. Pengaruh jenis pesan dan gender kemudian di evaluasi secara terpisah, menyediakan pengujian yang lebih tepat terhadap pengaruh individual.

- Menggunakan covariate-ANCOVA dan MANCOVA

#### **ANCOVA**

Covariate metrik umumnya dimasukkan dalam desain eksperimental untuk menghilangkan pengaruh extraneous dari variabel dependen sehingga dapat meningkatkan varian dalam kelompok (MSW). Hal ini sama dengan menggunakan blocking factor hanya saja variabelnya metrik. Prosedurnya sama dengan regresi linier yang digunakan untuk menghilangkan variasi dalam variabel dependen yang berhubungan dengan satu atau lebih covariate, kemudian ANOVA konvensional dilakukan pada variabel dependen yang disesuaikan.

#### **MANCOVA**

Merupakan perluasan dari prinsip ANCOVA untuk multivariate analisis (variabel dependen berganda) yaitu MANCOVA dipandang sebagai MANOVA terhadap regresi residual atau varian dalam variabel dependen yang tidak dapat dijelaskan oleh covariate

- Tahap 3: Asumsi ANOVA dan MANOVA

Untuk prosedur pengujian multivariat menggunakan MANOVA menjadi valid, maka harus memenuhi 3 asumsi yaitu :

- Observasi harus independen
- Matriks varians-kovarians harus sama (atau dapat diperbandingkan) untuk setiap kelompok treatment.
- Variabel dependen harus memiliki distribusi normal multivariat.
- Normalitas multivariat dapat diasumsikan tetapi sulit dalam pengujian. Normalitas univariat tidak menjamin normalitas multivariat, namun jika seluruh variabel memenuhi normalitas univariat maka kemencengan dari normalitas multivariat tidak konsekuensial. Kriteria lainnya adalah Linearitas dan multikolinearitas diantara variabel dependen serta sensitivitas terhadap outliers.

- Tahap 4: Estimasi Model MANOVA dan Menguji Overall Fit

Saat analisis MANOVA telah diformulasikan dan uji asumsi telah terpenuhi, pengujian terhadap signifikansi perbedaan antar kelompok yang dibentuk dengan treatment dapat dilakukan. Dalam melakukan pengujian ini peneliti harus memilih pengujian statistik yang tepat untuk tujuan penelitiannya. Namun dalam situasi tertentu khususnya jika analisis menjadi lebih kompleks, peneliti harus mengevaluasi kekuatan dari pengujian statistik untuk menyediakan perspektif yang lebih informatif dan memperoleh hasil yang diharapkan.

- Tahap 5: Interpretasi Hasil MANOVA

Jika covariate termasuk dalam model GLM, maka lakukan analisis model baik dengan maupun tanpa covariates. Jika covariates tidak meningkatkan kekuatan statistik atau tidak mempengaruhi secara signifikan terhadap pengaruh treatment, maka dapat didrop dari analisis final.

Jika dua atau lebih variabel independen (treatments) termasuk dalam analisis, interaksi harus diuji sebelum mengambil kesimpulan mengenai pengaruh utama dari variabel independen :

- Jika interaksi tidak signifikan secara statistik, maka pengaruh utama dapat diinterpretasi langsung karena perbedaan antar treatments dipertimbangkan konstan antar level kombinasi.
  - Jika interaksi signifikan secara statistik, dan perbedaan tidak konstan antar level kombinasi, maka interaksi harus ditentukan apakah ordinal atau disordinal:
  - Jika variabel independen memiliki dua atau lebih kelompok, dua jenis prosedur dapat digunakan untuk mengisolasi sumber perbedaan yaitu :
    - Post Hoc Methods
 

Post-hoc tests, menguji perbedaan potensial secara statistik antar seluruh mean kombinasi yang mungkin. Post-hoc tests memiliki kekuatan yang terbatas dan sangat cocok untuk mengidentifikasi pengaruh yang besar. Post Hoc Methods meliputi : Scheffe, Tukey's honestly significant difference (HSD), Tukey's extension of the Fisher least significant difference (LSD), Duncan's multiple-range test, Newman-Kuels test.
    - A Priori or Planned comparisons
 

Adalah prosedur yang tepat digunakan jika a priori theoretical reasons menyatakan bahwa kelompok tertentu akan berbeda dari satu kelompok dengan kelompok lainnya. Kesalahan Type I akan bertambah sebagaimana jumlah planned comparisons meningkat.
- Tahap 6: Validasi terhadap Hasil
 

Teknik analisis varian (ANOVA dan MANOVA) dikembangkan untuk situasi eksperimen, dengan replikasi sebagai pengertian utama terhadap validasi. Ketegasan perlakuan treatment eksperimen memungkinkan penggunaan yang luas dari eksperimen yang sama pada multiple populasi untuk menguji generalizability dari hasil.

Dalam penelitian ilmu pengetahuan sosial dan bisnis, bagaimanapun, eksperimen kerap kali digantikan dengan uji statistik dengan keadaan nonexperimental seperti survey penelitian. Kemampuan untuk memvalidasi hasil dalam keadaan ini adalah berdasarkan kemampuan replikasi dari treatment.

Tabel 3. Tabel Bartlett

Banyaknya Variabel Pengamatan	Banyaknya Perlakuan	Sebaran Percontohan Untuk Data Normal Ganda
$p = 1$	$t \geq 2$	$\left( \frac{\sum_{i=1}^t r_i - t}{t-1} \right) \left( \frac{1-\Lambda}{\Lambda} \right) \sim F_{t-1, \sum_{i=1}^t r_i}$
$p = 2$	$t \geq 2$	$\left( \frac{\sum_{i=1}^t r_i - t - 1}{t-1} \right) \left( \frac{1-\sqrt{\Lambda}}{\sqrt{\Lambda}} \right) \sim F_{2(t-1), 2\left(\sum_{i=1}^t r_i - t - 1\right)}$
$p \geq 1$	$t = 2$	$\left( \frac{\sum_{i=1}^t r_i - p - 1}{p} \right) \left( \frac{1-\Lambda}{\Lambda} \right) \sim F_{p, \sum_{i=1}^t r_i - p - 1}$
$p \geq 1$	$t = 3$	$\left( \frac{\sum_{i=1}^t r_i - p - 2}{p} \right) \left( \frac{1-\sqrt{\Lambda}}{\sqrt{\Lambda}} \right) \sim F_{2p, 2\left(\sum_{i=1}^t r_i - p - 2\right)}$
Untuk $p$ dan $t$ selain empat kategori di atas		$\left[ (n-1) - \left( \frac{p+t}{2} \right) \right] \ln \Lambda \sim \chi_{p(t-1)}^2$

Ketika MANOVA menunjukkan bahwa hipotesis nol ditolak, maka dilanjutkan dengan ANOVA pada masing-masing variabel terikat. prosedur ini akan menjaga taraf kesalahan  $\alpha$  sepanjang uji ANOVA dilaksanakan hanya jika uji MANOVA menolak hipotesis nol.

Komparasi ganda adalah tindak lanjut dari ANOVA apabila hasil analisis tersebut menunjukkan bahwa hipotesis nol ditolak dan nilai dari variabel penelitian lebih dari dua. Metode Scheffé dipilih karena menghasilkan cacah beda rerata signifikan paling sedikit dibandingkan metode lainnya. Metode ini terdapat empat macam komparasi, yaitu komparasi ganda rerata antar baris, komparasi ganda rerata antar kolom, komparasi ganda antar sel pada baris yang sama, dan komparasi ganda rerata antar sel pada kolom yang sama. Uji komparasi rerata antar sel untuk faktor interaksi tidak dapat dilakukan secara langsung menggunakan program

General Linear Model (GLM) pada paket software SPSS. Fasilitas yang dapat digunakan untuk melihat adanya interaksi secara langsung adalah profil efek. Jika profil variabel bebas pertama dan profil variabel bebas kedua tidak berpotongan, maka kecenderungannya tidak ada interaksi di antara mereka. Sebaliknya, jika profil variabel bebas pertama berpotongan dengan profil variabel bebas kedua, maka

kecenderungannya ada interaksi di antara keduanya. Namun, profil efek tidak dapat mengindikasikan perbedaan rerata. Untuk mengidentifikasi perbedaan secara spesifik tetap membutuhkan uji pasca ANOVA (disebut juga uji lanjut atau komparasi ganda). Misalnya kasus pada artikel ini menggunakan desain faktorial  $2 \times 3$  dengan 2 buah variabel bebas. Variabel bebas pertama memiliki 2 nilai dan variabel bebas kedua memiliki 3 nilai, sehingga terdapat 6 sel (kondisi eksperimentasi). Selain program GLM, uji komparasi ganda pada paket software SPSS dapat dilakukan melalui program One-Way ANOVA. Prosedur yang dapat dilakukan untuk keperluan tersebut adalah memanipulasi data dengan merubah 6 kondisi eksperimentasi menjadi 6 nilai (dianggap satu variabel bebas).

MANOVA memiliki asumsi yang mirip dengan ANOVA tetapi diperluas untuk kasus multivariat. Adapun asumsi yang harus dipenuhi pada MANOVA yaitu independen, sampel acak, normalitas multivariat, dan homogenitas matriks kovariansi. Uji persyaratan MANOVA meliputi uji normalitas multivariat yang dapat menggunakan uji Mardia dan uji homogenitas matriks kovariansi yang dapat menggunakan uji Box's M. Kesulitan penerapan analisis multivariat pada perhitungannya yang terlalu rumit, sudah terpecahkan dengan tersedianya fitur software statistik yang semakin canggih.

Terdapat beberapa statistik uji MANOVA yaitu Wilks' Lambda, Pillai, Lawley-Hotelling, dan Roy's Largest Root. Banyak software statistik menyajikan perhitungan keempat statistik uji MANOVA tersebut, dan biasanya keempat statistik uji tersebut menghasilkan kesimpulan yang sama. Ketika MANOVA menunjukkan bahwa hipotesis nol ditolak, maka dilanjutkan dengan ANOVA pada masing-masing variabel terikat. Prosedur ini akan menjaga taraf kesalahan  $\alpha$  sepanjang uji ANOVA dilaksanakan hanya jika uji MANOVA menolak hipotesis nol. Apabila ANOVA menunjukkan bahwa hipotesis nol ditolak, maka dilakukan uji post hoc menggunakan metode Scheffe'. Metode ini dilakukan ketika variabel bebas memiliki lebih dari dua nilai. Jika suatu variabel bebas hanya memiliki dua nilai, maka untuk melihat perbedaan antara keduanya dapat langsung dilakukan perbandingan pada masing-masing rerata marginalnya.



### 3. Cara Analisis dan Interpretasi SPSS

Kita akan melakukan penelitian yang berjudul “Pengaruh Pekerjaan Orang Tua Terhadap Nilai Ujian Matematika, Fisika dan Biologi Siswa Kelas 6 SD A”.

Perhatikan judul penelitian di atas, semuanya ada 4 variabel yang diteliti:

1. Pekerjaan Orang Tua
  2. Nilai Ujian Matematika
  3. Nilai Ujian Fisika
  4. Nilai Ujian Biologi
- Pekerjaan merupakan variabel independen, yang bertipe kategorik atau skala data nominal atau kualitatif. Terdiri dari 3 kategori: Tani, Buruh dan PNS.
  - Nilai Ujian semuanya variabel dependen yang bertipe numerik atau kuantitatif atau skala data interval/Rasio.

Berdasar contoh di atas, maka jelas harus anda pahami kembali bahwa uji MANOVA harus terdiri dari 1 variabel independen berskala kualitatif dan lebih dari 1 variabel dependen berskala data kuantitatif berdistribusi normal.

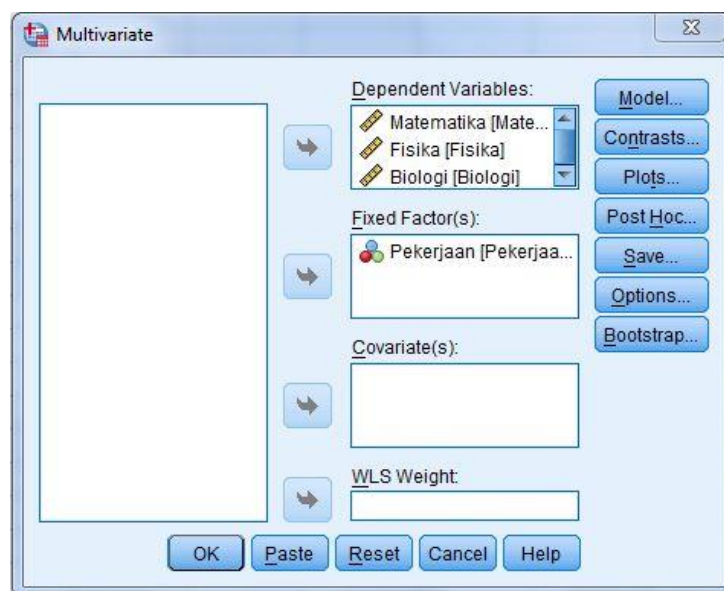
- Dataset Manova

Responden	Pekerjaan	Matematika	Fisika	Biologi
1	1	35	36	38
2	2	54	58	60
3	1	31	33	34
4	2	58	62	64
5	1	35	37	38
6	2	62	66	68
7	1	39	41	42
8	1	41	43	44
9	2	68	72	74
10	1	45	47	48
11	1	47	49	50
12	2	74	78	80
13	1	51	53	54
14	2	78	82	84
15	3	75	81	87
16	1	57	59	60
17	3	79	85	91
18	1	61	63	64
19	2	88	92	76
20	3	85	91	97
21	3	87	93	99
22	2	94	98	82
23	1	71	73	74
24	2	98	82	86

- Masukan data yang akan di analisis pada SPSS

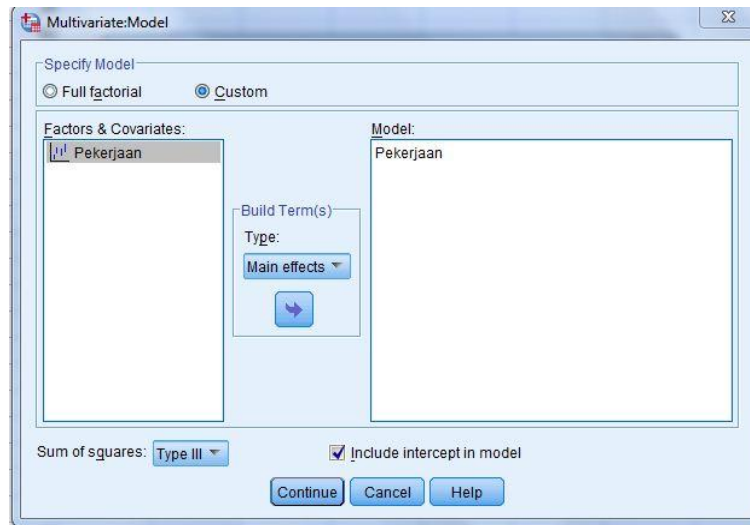
	Pekerjaan	Matematika	Fisika	Biologi
3	1	31	33	34
4	2	58	62	64
5	1	35	37	38
6	2	62	66	68
7	1	39	41	42
8	1	41	43	44
9	2	68	72	74
10	1	45	47	48
11	1	47	49	50
12	2	74	78	80
13	1	51	53	54
14	2	78	82	84
15	3	75	81	87
16	1	57	59	60
17	3	79	85	91
18	1	61	63	64
19	2	88	92	76
20	3	85	91	97
21	3	87	93	99
22	2	94	98	82
23	1	71	73	74
24	2	98	82	86

- Langkah berikutnya adalah: pada menu di SPSS, klik Analyze, General Linear Model, Multivariate: Lihat Tabel di bawah ini: Lalu masukkan variabel Matematika, Fisika dan Biologi ke kotak “Dependent Variables”. Masukkan Variabel Pekerjaan ke dalam kotak Fixed Factor (s).

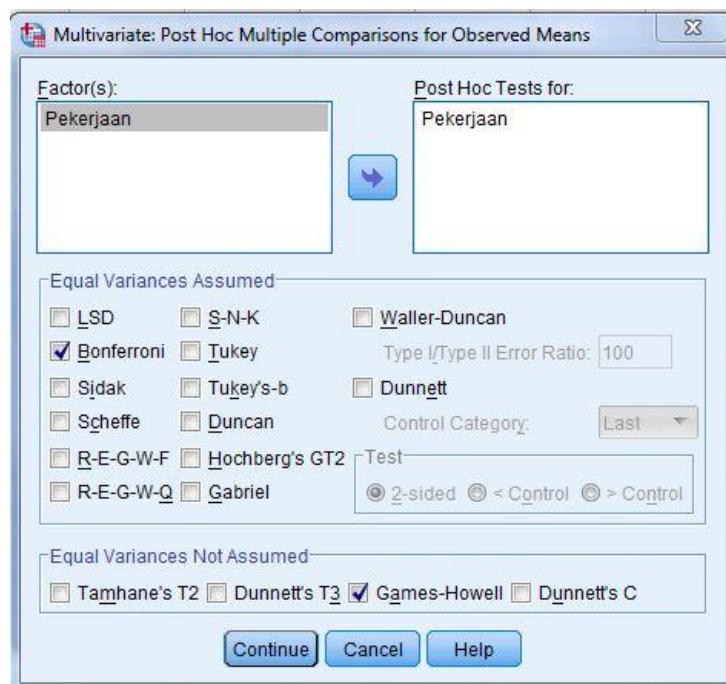


- Klik Tombol Model. Anda bisa menggunakan nilai bawaan (default) yaitu Full Factorial atau menggunakan nilai Custom, yaitu dengan memilih Custom

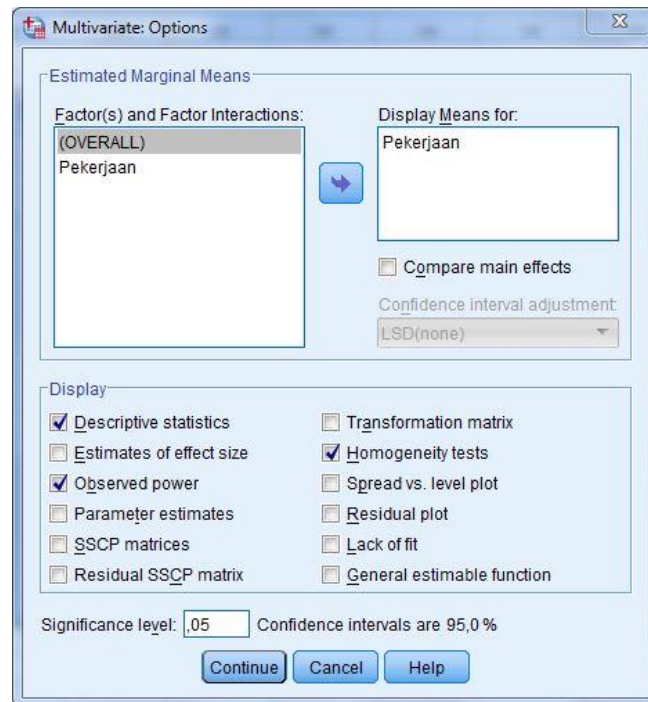
dan memasukkan Pekerjaan ke dalam kotak Model dan mengubah Type ke Main Effects.



- Klik Tombol Post Hoc. Maka akan muncul jendela seperti di bawah ini: Lalu masukkan Factor Pekerjaan ke kotak Post Hoc Test For, pada Equal Variances Assumed centang Bonferroni dan pada Equal Variances Not Assumed centang Games-Howell.
- Uji Bonferroni dipakai untuk membedakan kategori mana yang berbeda pada variabel independen apabila hasil uji homogenitas test pada Levene's Test menunjukkan memiliki varians yang berbeda dengan nilai sig. > 0,05. Sedangkan Games-Howell dipakai apabila sig. <0,05).



- Kemudian anda tekan tombol options. Lalu masukkan factor Pekerjaan ke dalam kotak Display Means for. Pada display, centang Descriptive Statistics, Observed Power dan Homogeneity Tests. Dan biarkan Significance Level 0,05.



- **Output**

Tabel Deskriptive

Descriptive Statistics				
	Pekerjaan	Mean	Std. Deviation	N
Matematika	Tani	46,64	12,388	11
	Buruh	74,89	15,878	9
	PNS	81,50	5,508	4
	Total	63,04	20,031	24
Fisika	Tani	48,55	12,485	11
	Buruh	76,67	13,454	9
	PNS	87,50	5,508	4
	Total	65,58	20,128	24
Biologi	Tani	49,64	12,388	11
	Buruh	74,89	9,171	9
	PNS	93,50	5,508	4
	Total	66,42	19,755	24

Tabel di atas, menunjukkan hasil uji deskriptif. Contoh dari tabel di atas: Responden dengan Pekerjaan Tani, rata-rata (mean) nilai matematika sebesar 46,64 dan jumlahnya ada 11 orang. Sedangkan Buruh rata-rata nilai fisika

sebesar 76,67 dan jumlahnya 9 orang, begitu pula yang lain cara membacanya sama. Agar lebih jelas, baca artikel kami yang berjudul “Analisis Deskriptif dengan Excel”.

### Output Multivariat ANOVA(MANOVA)

**Multivariate Tests<sup>a</sup>**

Effect		Value	F	Hypothesis df	Error df	Sig.	Noncent. Parameter	Observed Power <sup>d</sup>
Intercept	Pillai's Trace	,982	352,811 <sup>b</sup>	3,000	19,000	,000	1058,433	1,000
	Wilks' Lambda	,018	352,811 <sup>b</sup>	3,000	19,000	,000	1058,433	1,000
	Hotelling's Trace	55,707	352,811 <sup>b</sup>	3,000	19,000	,000	1058,433	1,000
	Roy's Largest Root	55,707	352,811 <sup>b</sup>	3,000	19,000	,000	1058,433	1,000
Pekerjaan	Pillai's Trace	1,092	8,015	6,000	40,000	,000	48,087	1,000
	Wilks' Lambda	,144	10,379 <sup>b</sup>	6,000	38,000	,000	62,276	1,000
	Hotelling's Trace	4,324	12,972	6,000	36,000	,000	77,834	1,000
	Roy's Largest Root	3,904	26,028 <sup>c</sup>	3,000	20,000	,000	78,085	1,000

a. Design: Intercept + Pekerjaan

b. Exact statistic

c. The statistic is an upper bound on F that yields a lower bound on the significance level.

d. Computed using alpha = ,05

Seperti biasa untuk hasil uji F, kita mengabaikan bagian yang diberi label “Intercept.” Pada baris yang di bawah pada tabel di atas menunjukkan 4 nomor memberikan nilai P value untuk empat uji multivariat yang berbeda. Hasil tersebut memberitahukan pada kita jika ada pengaruh yang signifikan dari variabel independen pada semua variabel dependen. Jika Anda ditanya “Secara keseluruhan, apakah ada pengaruh yang signifikan dari variabel independen pada satu set kelompok variabel dependen”, “maka Anda akan menjalankan MANOVA dan melihat hasil uji multivariat ini sebagai kesimpulan Anda”. “Artinya, jika 4 nilai p-value menunjukkan <0,05, maka signifikan pada level kepercayaan 95%”.

- Ingat bahwa berdasar tabel di atas, tidak ada satu uji multivariat tunggal, yang ada adalah empat jenis uji yang berbeda.

## Tabel Uji Asumsi Levene Test Pada MANOVA

**Levene's Test of Equality of Error Variances<sup>a</sup>**

	F	df1	df2	Sig.
Matematika	2,138	2	21	,143
Fisika	1,368	2	21	,276
Biologi	1,371	2	21	,276

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + Pekerjaan

Tabel di atas, menunjukkan hasil uji homogenitas yaitu uji Levene. Dikatakan semua variabel memiliki varian yang sama apabila nilai sig. > 0,05. Nilai ini nantinya akan mempengaruhi pilihan uji Post Hoc apa yang digunakan. Apabila Sig. >0,05 maka uji Post Hoc menggunakan Uji Benferroni, sedangkan jika <0,05 maka menggunakan Games-Howell. Hasil di atas menunjukkan semua variabel dependen memiliki varian yang sama sebab Sig. >0,05 sehingga uji Post Hoc yang digunakan nantinya adalah Benferroni.

## Output Uji MANOVA

**Tests of Between-Subjects Effects**

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Noncent. Parameter	Observed Power <sup>d</sup>
Corrected Model	Matematika	5586,524 <sup>a</sup>	2	2793,262	16,104	,000	32,208	,998
	Fisika	6220,106 <sup>b</sup>	2	3110,053	21,084	,000	42,167	1,000
	Biologi	6677,399 <sup>c</sup>	2	3338,699	30,505	,000	61,009	1,000
Intercept	Matematika	91188,962	1	91188,962	525,739	,000	525,739	1,000
	Fisika	100098,284	1	100098,284	678,583	,000	678,583	1,000
	Biologi	105161,253	1	105161,253	960,822	,000	960,822	1,000
Pekerjaan	Matematika	5586,524	2	2793,262	16,104	,000	32,208	,998
	Fisika	6220,106	2	3110,053	21,084	,000	42,167	1,000
	Biologi	6677,399	2	3338,699	30,505	,000	61,009	1,000
Error	Matematika	3642,434	21	173,449				
	Fisika	3097,727	21	147,511				
	Biologi	2298,434	21	109,449				
Total	Matematika	104611,000	24					
	Fisika	112546,000	24					
	Biologi	114844,000	24					
Corrected Total	Matematika	9228,958	23					
	Fisika	9317,833	23					
	Biologi	8975,833	23					

a. R Squared = ,605 (Adjusted R Squared = ,568)

b. R Squared = ,668 (Adjusted R Squared = ,636)

c. R Squared = ,744 (Adjusted R Squared = ,720)

d. Computed using alpha = ,05

Tabel di atas menunjukkan nilai uji Manova. Seperti biasanya pada uji F, anda akan disuguhkan dengan beberapa nilai: Corrected Model, Intercept, Pekerjaan, Error dan Total. Dalam bahasan kali ini, kami menganggap semua pembaca telah membaca artikel-artikel sebelumnya, sehingga memahami maksud dari nilai-nilai tersebut.

### Output Univariat ANOVA Di Dalam MANOVA

Langsung saja kita lihat baris “Pekerjaan” pada tabel di atas. Pada baris tersebut ada 3 baris lagi, yaitu “Matematika”, “Fisika” dan “Biologi”. Maksud dari hal tersebut adalah tiap baris menunjukkan hasil uji pengaruh satu variabel independen yaitu pekerjaan terhadap masing-masing variabel dependen. Dari hasil di atas, lihat nilai pada kolom “Sig.”. Dikatakan Signifikan apabila nilai Sig. < 0,05.

Contoh di atas ketiga nilai menunjukkan 0,000 di mana < 0,05, sehingga kesimpulan dan jawaban hipotesis adalah:

- Pekerjaan secara bermakna mempengaruhi nilai ujian matematika dengan P Value 0,000 yang artinya H0 Ditolak atau H1 Diterima.
- Pekerjaan secara bermakna mempengaruhi nilai ujian Fisika dengan P Value 0,000 yang artinya H0 Ditolak atau H1 Diterima.
- Pekerjaan secara bermakna mempengaruhi nilai ujian Biologi dengan P Value 0,000 yang artinya H0 Ditolak atau H1 Diterima.

### Output Uji Post Hoc Setelah MANOVA

#### Post Hoc Tests

#### Pekerjaan

Multiple Comparisons								
Dependent Variable		(I) Pekerjaan	(J) Pekerjaan	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
							Lower Bound	Upper Bound
Matematika	Bonferroni	Tani	Buruh	-28,25*	5,919	,000	-43,65	-12,85
			PNS	-34,86*	7,690	,001	-54,87	-14,86
		Buruh	Tani	28,25*	5,919	,000	12,85	43,65
			PNS	-6,61	7,914	1,000	-27,20	13,98
		PNS	Tani	34,86*	7,690	,001	14,86	54,87
			Buruh	6,61	7,914	1,000	-13,98	27,20
	Games-Howell	Tani	Buruh	-28,25*	6,478	,002	-45,08	-11,42
			PNS	-34,86*	4,640	,000	-47,24	-22,48
		Buruh	Tani	28,25*	6,478	,002	11,42	45,08
			PNS	-6,61	5,966	,529	-22,77	9,55
		PNS	Tani	34,86*	4,640	,000	22,48	47,24
			Buruh	6,61	5,966	,529	-9,55	22,77

Fisika	Bonferroni	Tani	Buruh	-28,12*	5,459	,000	-42,32	-13,92
			PNS	-38,95*	7,091	,000	-57,40	-20,51
		Buruh	Tani	28,12*	5,459	,000	13,92	42,32
			PNS	-10,83	7,298	,458	-29,82	8,15
		PNS	Tani	38,95*	7,091	,000	20,51	57,40
			Buruh	10,83	7,298	,458	-8,15	29,82
	Games-Howell	Tani	Buruh	-28,12*	5,855	,000	-43,17	-13,07
			PNS	-38,95*	4,664	,000	-51,39	-26,52
		Buruh	Tani	28,12*	5,855	,000	13,07	43,17
			PNS	-10,83	5,263	,144	-25,05	3,38
		PNS	Tani	38,95*	4,664	,000	26,52	51,39
			Buruh	10,83	5,263	,144	-3,38	25,05
Biologi	Bonferroni	Tani	Buruh	-25,25*	4,702	,000	-37,48	-13,02
			PNS	-43,86*	6,108	,000	-59,75	-27,97
		Buruh	Tani	25,25*	4,702	,000	13,02	37,48
			PNS	-18,61*	6,287	,022	-34,97	-2,26
		PNS	Tani	43,86*	6,108	,000	27,97	59,75
			Buruh	18,61*	6,287	,022	2,26	34,97
	Games-Howell	Tani	Buruh	-25,25*	4,827	,000	-37,58	-12,93
			PNS	-43,86*	4,640	,000	-56,24	-31,48
		Buruh	Tani	25,25*	4,827	,000	12,93	37,58
			PNS	-18,61*	4,114	,003	-29,98	-7,24
		PNS	Tani	43,86*	4,640	,000	31,48	56,24
			Buruh	18,61*	4,114	,003	7,24	29,98

Based on observed means.  
The error term is Mean Square(Error) = 109,449.  
\*. The mean difference is significant at the ,05 level.

### Post Hoc MANOVA

Tabel di atas menunjukkan hasil Uji Post Hoc. Karena nilai uji homogenitas menunjukkan Sig. >0,05 pada semua variabel, maka masing-masing pengaruh variabel independen terhadap variabel dependen menggunakan uji Benferroni. Dikatakan ada perbedaan variabel dependen yaitu nilai ujian berdasarkan variabel independen yaitu Pekerjaan apabila memiliki tanda bintang.

### Kesimpulan Interpretasi Uji MANOVA

Dalam Contoh di atas dapat disimpulkan sebagai berikut:

- Untuk perbedaan nilai ujian matematika berdasarkan pekerjaan, yang memiliki perbedaan adalah Tani dengan Buruh dan Tani dengan PNS.
- Untuk perbedaan nilai ujian fisika berdasarkan pekerjaan, yang memiliki perbedaan adalah Tani dengan Buruh dan Tani dengan PNS.
- Untuk perbedaan nilai ujian biologi berdasarkan pekerjaan, yang memiliki perbedaan adalah Tani dengan Buruh, Tani dengan PNS dan Buruh dengan PNS.



# DAFTAR PUSTAKA

---

- A.S. Fotheringham, C. Brunson, and M. Charlton and Sandt Damanik Witwicky. (2002) *Geographically weighted regression: the analysis of spatially varying relationships*. Wiley.
- Agresti, Alan. (2002). *Categorical Data Analysis*. New York: Wiley-Interscience. ISBN 0-471-36093-7.
- Alhusin, S., 2002. *Aplikasi Statistik Praktis Dengan SPSS For Windows*, J & J Learning, Yogyakarta.
- Alhusin, Syahri, “*Aplikasi Statistik Praktis dengan Menggunakan SPSS 10 for Windows*”, Edisi Kedua, Yogyakarta: Graha Ilmu, 2003.
- Amemiya, T. (1985). *Advanced Econometrics*. Harvard University Press. ISBN 0-674-00560-0.
- Anderson, T. W. 2003. *An Introduction to Multivariate Statistical Analysis*. New Jersey: John Wiley & Sons, Inc.
- Anonim. 2003. *Analisis Peubah Ganda*. IPB. Bandung.
- Arikunto, Suharsimi, “*Prosedur Penelitian : Suatu Pendekatan Praktek*”, Edisi Revisi V, Jakarta: PT Rineka Cipta, 2002.
- Balakrishnan, N. (1991). *Handbook of the Logistic Distribution*. Marcel Dekker, Inc. ISBN 978-0824785871.
- Bhratara Karya Aksara. Jakarta. Birkes, David and Yadolah Dodge, *Alternative Methods of Regression*. ISBN 0-471-56881-3
- Bryce, G. R. 1975. The One-Way Model. *American Statistician* 29, 69-70.
- Bryce, G. R., Carter, M. W., and Reader, M. W. 1976. *Nonsingular and singular transformations in the fixed model*. Presented at the Annual Meeting of the American Statistical Association, Boston, MA, August 1976.
- Bryce, G. R., Scott, D. T, and Carter, M. W. 1980. *Estimation and Hypothesis Testing in Linear Models—A Reparameterization Approach*. Communications in Statistics—Series A, Theory and Methods 9, 131-150.
- Chatfield, C. (1993) "Calculating Interval Forecasts," *Journal of Business and Economic Statistics*, 11. pp. 121–135.
- Christensen. 1996. *Analysis Of Variance, Design, and Regression*. Chapman and Hall. London. French and Friends. *Multivariate Analysis Of Variance (MANOVA)*.
- Chu, Wei; Keerthi, S. Sathiya (2007). + Msgstr "Dukung regresi ordinal vektor". *Komputasi Saraf*. 19 (3): 792 – 815. Cite Seer X 10.1.1.297.3637 . doi : 10.1162 neco.2007.19.3.792 . PMID 17298234 .
- Cochran and Cox. 1957. *Experimental Designs*. John Willey and Sons. New York.
- Corder, G.W. and Foreman, D.I. (2009). *Nonparametric Statistics for Non-Statisticians: A Step-by-Step Approach* Wiley, ISBN 978-0-470-45461-9
- Crammer, Koby; Singer, Yoram (2001). *Pranking dengan peringkat* . NIPS.

- Draper, N.R. and Smith, H. (1998). *Applied Regression Analysis* Wiley Series in Probability and Statistics
- Field, A. 2009. *Discovering Statistics Using SPSS*. London: Sage Publications.
- Fikri Lukiastusi dan Muliawan Hamdani. 2012. *Statistika Non Parametris : Aplikasinya dalam Bidang Ekonomi dan Bisnis*. Yogyakarta : CAPS.
- Fox, J. (1997). *Applied Regression Analysis, Linear Models and Related Methods*. Sage
- Gani, J. 1982. *The Making of Statisticians*. Springer- Verlag. New York.
- Ghazali, Imam, 2016, "Aplikasi Analisis Multivariat Dengan Program IBM SPSS 23", Cetakan kedelapan, Universitas Diponegoro, Semarang.
- Ghozali, I., 2005. *Aplikasi Analisis Multivariat Dengan program SPSS*, Badan Penerbit Universitas Diponegoro, Semarang.
- Ghozali, I., dan Castellan J., 2002. *Statistik Non-Parametrik Teori dan Aplikasi Dengan Program SPSS*, Badan Penerbit universitas Diponegoro, Semarang.
- Greene, William H. (2003). *Econometric Analysis, fifth edition*. Prentice Hall. ISBN 0-13-066189-9.
- Gutiérrez, PA; Pérez-Ortiz, M .; Sánchez-Monedero, J .; Fernández-Navarro, F .; Hervás-Martínez, C. (Januari 2016). "Metode Regresi Ordinal: Survei dan Studi Eksperimental". *Transaksi IEEE pada Pengetahuan dan Rekayasa Data* .
- Hanafiah, K. A. 1991. *Rancangan Percobaan Teori dan Aplikasi*. Universitas Sriwijaya. Palembang.
- Hardle, W. dan Simar, L. 2007. *Applied Multivariate Statistical Analysis*. Berlin: Springer-Verlag.
- Hardle, W., *Applied Nonparametric Regression* (1990), ISBN 0-521-42950-1
- Harlow, L. 2005. *The Essence of Multivariate Thinking Basic Themes and Methods*. Lawrence Erlbaum Associates. London.
- Herbrich, Ralf; Graepel, Thore; Obermayer, Klaus (2000). "Batas Batas Peringkat Besar untuk Regresi Ordinal". *Uang Muka dalam Klasifikasi Margin Besar* . MIT Press. hlm. 115–132.
- Hilbe, Joseph M. (2009). *Logistic Regression Models*. Chapman & Hall/CRC Press. ISBN 978-1-4200-7575-5.
- Hocking, R. R. 1985. *The Analysis of Linear Models*. Monterey, CA: Brooks/Cole.
- Hocking, R. R. 1996. *Methods and Applications of Linear Models*. New York: Wiley.
- Hocking, R. R. and Speed, F M. 1975. A Full Rank Analysis of Some Linear Model Problems. *Journal of the American Statistical Association* 70, 706-712.
- Hosmer, David W. (2000). *Applied Logistic Regression, 2nd ed*. New York; Chichester, Wiley. ISBN 0-471-35632-8.
- Iriawan, N., dan Astuti S.P, 2006. *Mengolah Data Statistik Dengan Mudah Menggunakan Minitab 14*, Andi, Yogyakarta.
- Ian Bryman and Duncan Cramer. 2005. *Quantitatif Data Analysis with Minitab : A guide for Sosial Scientists*. London : Taylor and Francis e-library.

- Lentner, M and Thomas. B. 1986. *Experimental Design and Analysis*. Valey Book Company. Blacksburg.
- Leps, J. dan Smilauer, P. 1999. *Multivariate Analysis of Ecological Data*. Ceske Budejovice: Faculty of Biological Science, University of South Bohemia.
- Lindley, D.V.(1987). "Regression and correlation analysis," New Palgrave: A Dictionary of Economics, v. 4, pp. 120–23.
- Mallows, C. (1998). 1997, Fisher Memorial Lecture. The Zeroth Problem. *ASA* 52(1) : 1-9.
- Mattjik, A. A. dan Sumertajaya, I. M. 2011. *Sidik Peubah Ganda dengan Menggunakan SAS*. Bogor: IPB Press.
- McCullagh, Peter (1980). + Msgstr "Model regresi untuk data ordinal". *Jurnal Masyarakat Statistik Kerajaan . Seri B (Metodologi)*. 42 (2): 109–142.
- Meade, N. and T. Islam (1995) "Prediction Intervals for Growth Curve Forecasts," *Journal of Forecasting*, 14, pp. 413–430.
- Morrison, D. F. 1990. *Multivariate Statistical Methods*. New York: McGraw-Hill, Inc.
- Morrison. 1978. *Multivariate Statistical Method*. Mcgraw-Hill International Book Company. Auckland.
- N. Cressie (1996) Change of Support and the Modifiable Areal Unit Problem. *Geographical Systems* 3:159–180.
- Nasoetion, A. H. dan Rambe, A. 1984. *Teori Statistika untuk ilmu-ilmu Kuantitatif*.
- Nawari. 2010. *Analisis Regresi dengan MS Excel 2007 dan SPSS 17*. Jakarta: PT Elex Media Komputindo.
- Neil. J Salkind. 2007. *Encyclopedia Measurement and Statistics*. Volume 1. Sage Publications.
- Nelder, J. A. 1974. Letter to The Editor. *Journal of the Royal Statistical Society (Series C)*, 23, 232.
- Oja, H. 2010. *Multivariate Nonparametric Methods with R: An Approach Based on Spatial Signs and Ranks*. New York: Springer Science+Business Media.
- orca: *Regresi Ordinal dan Algoritma Klasifikasi*, *AYRNA*, 2017-11-21, diambil 2017-11-21
- Parzen. Emanuel. 2002. Data Mining, Statistical Methods Mining and History of Statistics". (Department of statistics Texas A & M University College)
- Peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia Nomor 53 Tahun 2015 tentang *Penilaian Hasil Belajar oleh Pendidik dan Satuan Pendidikan pada Pendidikan Dasar dan Pendidikan Menengah*.
- Pratista, A., 2003. *Aplikasi SPSS 10.05 Dalam Statistik dan Rancangan Percobaan*, Alfabeta, Bandung.
- Priyatno, Duwi, "Mandiri Belajar SPSS", Cetakan Ketiga, Yogyakarta: Media Kom, 2008.
- Priyatno, Duwi, 2013, "Mandiri Belajar Analisis Data Dengan SPSS", Yogyakarta: Media Kom.

- Priyatno, Duwi, 2014, “*SPSS 22 Pengolah Data Terpraktis*”, Edisi 1, Yogyakarta: ANDI.
- Priyatno, Duwi. 2010. *Paham Analisa Statistik Data dengan SPSS. Mediakom. Yogyakarta.*
- Rao, C. R. and Szekely, G. J. 2000. *Statistics for The 21<sup>st</sup> Century. Methodologies for Applications of the future.* Marcell Dekker. New York.
- Rencher, A. 1998. *Multivariate Statistical Inference and Applications.* Jhon Willey & Sons Inc. Canada.
- Rencher, A. 2002. *Methods of Multivariate Analysis.* Jhon Willey & Sons Inc. Canada.
- Rencher, A. C. 1998. *Multivariate Statistical Inference and Applications.* Kanada: John Wiley and Sons, Inc.
- Rencher, A. C. 2002. *Methods of Multivariate Analysis.* Kanada: John Wiley and Sons, Inc.
- Rennie, Jason DM; Srebro, Nathan (2005). *Kehilangan Fungsi untuk Tingkat Preferensi: Regresi dengan Label Urutan Diskrit(PDF)*. Proc Lokakarya Multidisiplin IJCAI tentang Kemajuan dalam Penanganan Preferensi.
- Richard and Dean. 2002. *Applied Multivariate Statistical Analysis.* Prentice Hall. USA.
- Riduwan. 2006. *Dasar-Dasar Statistika.* Alfabeta. Bandung.
- Saefuddin, A. 2000. Profil Jurusan Statistika. FMIPA-IPB.
- Santoso, S., 2000. *SPSS Mengolah Data Statistik Secara Profesional Versi 7.5,* Elex Media Komputindo, Jakarta.
- Santoso, S., 2000. *Buku Latihan SPSS Statistik Parametrik,* Media Elex Komputindo, Jakarta.
- Santoso, S., dan Tjiptono F., 2001. Riset Pemasaran Konsep dan aplikasi Dengan SPSS, Media Elex Komputindo, Jakarta.
- Santoso, Singgih, 2000, *Buku Latihan SPSS Statistik Parametrik.* PT.ELEK Media Komputindo. Jakarta
- Santoso, Singgih, 2013, “*Menguasai SPSS 21 di Era Informasi*”. PT.ELEK Media Komputindo. Jakarta.
- Sastrosupadi, A. 2000. *Rancangan Percobaan Praktis Bidang Pertanian.* Kanisius. Malang.
- Searle, S. R. 1977. *Analysis of Variance of Unbalanced Data From 3-Way and Higher-Order Classifications.* Technical Report BU-606-M, Cornell University, Biometrics Units.
- Searle, S. R., Speed, F. M., and Henderson, H. V. 1981. Some computational and model equivalencies in analysis of variance of unequal-subclass-numbers data. *American Statistician* 35, 16-33.
- Sekaran, Uma, 2000. “*Research Methods for Business, A Skill Building Approach*”, New York: John Wiley n Sons
- Sekaran, Uma. 2006. *Research Methods For Business: Metodologi Penelitian untuk Bisnis,* Penerbit Salemba Empat.

- Shashua, Amnon; Levin, Anat (2002). *Pemeringkatan dengan prinsip margin besar: Dua pendekatan*. NIPS.
- Simon, H. 2007. *Metode Inventore Hutan*. Pustaka Pelajar. Yogyakarta.
- Speed, F M., Hocking, R. R., and Hackney, O. P. 1978. Methods of analysis of linear models with unbalanced data. *Journal of the American Statistical Association* 73, 105-112.
- Speed, F. M. 1969. *A New Approach to the Analysis of Linear Models*. Technical Report, National Aeronautics and Space Administration, Houston, TX. NASA Technical memo, NASA TM X-58030.
- Spiegel, Murray .R. 1961. *Theory and Problem of Statistics*. Mc Graw-Hill. New York.
- Steel dan James. 1991. *Terjemahan Prinsip dan Prosedur Statistika*. Gramedia Pustaka Utama. Jakarta.
- Steven, J. 2002. *Applied Multivariate Statistics for the Social Sciences*. New Jersey: Lawrence Erlbaum Associates.
- Steven, S.S., 1946. *On The Theory of Scales Measurement*, Science.
- Sudjana, 1996. *Metoda Statistik*, Tarsito, Bandung.
- Sugiyono, "Metode Penelitian Bisnis", Bandung: CV. Alfabeta, 2007.
- Sugiyono. 2005. *Statistika untuk Penelitian*. Alfabeta. Bandung.
- Sugiyono. 2008. *Metode Penelitian Bisnis*. Alfabeta. Bandung.
- Sugiyono. 2008. *Statistik Nonparametris untuk Penelitian Cetakan Keenam* Alfabeta. Bandung.
- Sugiyono. 2009. *Metode Penelitian Kuantitatif Kualitatif dan R & B*. Bandung.
- Sumber : Irawan, N & Astuti S.P. 2006. *Mengolah Data Statistik dengan Mudah Menggunakan Minitab 14*. Yogyakarta :Penerbit Andi.
- Supranto, J. 2004. *Analisis Multivariat : Arti dan Interpretasi*. Rineka Cipta. Jakarta.
- Tabachnick, B. G. dan Fidell, L. S. 2007. *Using Multivariate Statistics*. Boston: Pearson Education, Inc.
- Timm, N. H. 2002. *Applied Multivariate Analysis*. New York: Springer-Verlag.
- Trihendradi, C., 2005. *Step by Step SPSS 13 Analisis Data Statistik*, Andi, Yogyakarta.
- Urquhart, N. S., Weeks, D. L., dan Henderson, C. R. 1973. Estimation Associated with Linear Models: A Revisitation. *Communications in Statistics* 1, 303-330.
- Walpole dan Raymond. 1995. *Terjemahan Ilmu Peluang dan Statistika Untuk Insinyur dan Ilmuwan*. ITB. Bandung.
- Walpole, Ronald E. 1992. *Pengantar Statistika Edisi ke-3*. PT Gramedia Pustaka Utama. Jakarta.
- Walpole. R. 1990. *Terjemahan Pengantar Statistika*. PT. Gramedia. Jakarta.
- Wijaya, 2000. *Statistika Non Parametrik Aplikasi Program SPSS*, Alfabeta, Bandung.
- William H. Kruskal and Judith M. Tanur, ed. (1978), "Linear Hypotheses," *International Encyclopedia of Statistics*. Free Press, v. 1,

- Winship, Christopher; Mare, Robert D. (1984). "Model Regresi dengan Variabel Ordinal" *Ulasan Sosiologis Amerika* .49 (4): 512–525. doi : 10.2307 / 2095465 . JSTOR 2095465 .
- Wooldridge, Jeffrey M. (2010). *Analisis Ekonometrik Potongan Lintang dan Data Panel* . MIT Press. hlm. 655-657. ISBN 9780262232586 .
- Yang, K and Jayant. T. 2004. *Multivariate Statistical Method in Quality Management*. McGraw-Hill International Book Company. Auckland.
- Zuhri, S. 1996. *Beberapa Perlakuan Fisik terhadap Perkecambahan Benih Kemiri (Aleurites moluccana Wild.)*. Skripsi Fakultas Pertanian UNIB. Bengkulu.

# LAMPIRAN-LAMPIRAN

---