BAB I
PENDAHULUAN

SPSS (*Statistical Program for Social Science*) merupakan paket program aplikasi komputer untuk menganalisa data terutama untuk ilmu-ilmu sosial. Namun demikian, dengan SPSS kita bisa membuat laporan tabulasi, *chart* (grafik), plot (diagram), statistik deskriptif dan analisa statistik yang kompleks. Karena SPSS merupakan paket program untuk mengolah dan menganalisa data, maka untuk menjalankan program ini terlebih dahulu harus dipersiapkan data yang akan diolah dan dianalisa. SPSS for windows menggunakan 6 tipe window, yaitu *SPSS Data Editor, Output Window, Syntax Window, Chart Carousel, Chart Window,* dan *Help Window*.

Konsep Data dalam SPSS

- Data harus mempunyai struktur, format, dan jenis tertentu.
- Data yang disusun dari teks editor lain di luar SPSS, dapat dibaca SPSS : W*ordstar (Non Document), Side Kick, Notepad, Norton Commander, Kode ASCII* lainnya*, Lotus 123, Data Base.*

Struktur Data

- Data harus disusun dalam m baris dan n kolom
- Tiap baris data disebut *case* (kasus)
- Tiap kolom data mempunyai *heading* yang disebut variabel (*field*)
- Interaksi antara tiap variabel dan *case* disebut *value*

Format Data

- File data yang dibuat tidak lewat SPSS data editor punya 2 format, yaitu *fixed* (tentu) dan *free field* (bebas)
- File data yang dibuat lewat SPSS data editor punya format baku berorientasi pada *fixed.*
- Format *fixed* : setiap variabel menempati posisi dalam lokasi kolom tertentu dan setiap *case*-nya sama
- Format *free field* : menempatkan variabel tak perlu pada lokasi kolom tertentu. Dan untuk masing-masing *case*, lokasi tidak perlu harus sama. Dan dalam satu baris bisa lebih dari satu *case* (dipisahkan oleh spasi atau koma).
- Kelebihan format *fixed* : susunan rapi tapi menghabiskan memori
- Kelebihan format *free field* : menghemat memori tapi tidak rapi
- Tanda ‘ (apostrof) atau “ (petik) digunakan untuk menyatakan value lebih dari dua kata sebagai satu kesatuan

Missing Value

Missing value didefinisikan sebagai data hilang / tak lengkap. Hal ini diperhatikan karena data yang hilang akan sangat berpengaruh pada hasil pengolahan maupun analisa dari keseluruhan data. Misalnya, pada pengumpulan nilai tes IQ suatu kelas dengan 50 siswa, di mana 3 siswa yang tidak masuk pada salah satu hari tes dari hari yang dijadwalkan. Tentunya ketidakhadiran siswa tersebut akan mempengaruhi hasil analisa data kelas tersebut secara keseluruhan. Untuk mengatasi hal ini, nilai tes ketiga siswa tersebut harus diberi harga tertentu, misalnya 0 yang mendeklarasikan sebagai *missing value*. Dengan *value* ini *case* yang valid hanya 47 meski jumlah *case* 50.

Ada 2 jenis *missing value* yang dikenal oleh SPSS, yaitu :

1. *User missing value*, adalah *missing value* yang nilainya ditentukan oleh *user* (pemakai). Seperti pada pendataan nilai tes IQ di atas, ditentukan harga 0 sebagai *missing value*.
2. *Sistem missing value*, adalah *missing value* yang ditentukan secara otomatis oleh SPSS, yaitu bilamana dijumpai harga yang ilegal. Seperti didapatinya karakter alphabetik pada variabel numerik, atau perhitungan yang menghasilkan nilai tak terdefinisikan pada perintah transformasi data seperti pembagian dengan 0.

Konsep Variabel

Dilihat dari bentuknya, variabel terdiri dari :

1. Variabel Kualitatif. Variabel ini berbentuk klasifikasi atau kategori. Dalam tipologi dasar, variabel ini dibedakan menjadi 2 macam :
   - Tanpa peringkat, contoh : (Surabaya, Bandung, Jakarta), (Laki-laki, Perempuan), (Merah, Putih, Hitam, Hijau)
   - Dengan peringkat, contoh : (kurang, cukup, baik), (rendah,sedang, tinggi), (pesuruh, personalia, direktur).
2. Variabel Kuantitatif. Variabel ini merupakan skor yang berwujud numerik. Secara garis besar terdiri dari :
   - Diskrit, contoh : jumlah siswa, banyak kasus, jumlah penduduk
   - Kontinue, contoh : tinggi badan, berat badan, jarak lempar, besar kaki

Berdasarkan fungsinya, variabel dibedakan atas :

1. Variabel Independent
2. Variabel Dependent
3. Variabel Intervening
4. Variabel Sertaan / Moderator
5. Variabel Antensenden

Mendefinisikan Variabel

Pendefinisian variabel hanya dapat dilakukan bila SPSS data editor sedang aktif. Pada pendefinisisan variabel kita dapat melakukan pemberian nama variabel sekaligus menentukan format dari variabel tersebut. Ketentuan dalam memberikan nama variabel adalah sebagai berikut :

- Nama variabel harus diawali dengan huruf
- Tidak boleh diakhiri dengan tanda titik
- Panjang variabel maksimal 8 karakter
- Tidak boleh ada blank atau spasi dan karakter special seperti !, ?, ‘, dan *
- Harus unik, yaitu tidak boleh ada nama variabel yang sama
- Tidak membedakan huruf kecuali dengan huruf kapital
- Tidak boleh menggunakan istilah *reserved word* (istilah yang sudah ada pada SPSS)

Tombol Perintah

Tombol perintah umumnya terletak di sebelah kanan kotak dialog utama

OK : untuk menjalankan prosedur sekaligus menutup kotak dialog

Paste : menampilkan susunan perintah pada *syntax window* yang merupakan hasil transfer perintah-perintah yang diberikan pada kotak dialog

Reset : mengembalikan variabel terseleksi ke daftar variabel asal dan spesifikasinya default

Cancel : membatalkan seluruh perintah dan spesifikasinya dan sekaligus menutup kotak dialog

Help : melihat informasi yang berkenaan dengan kotak dialog yang aktif.

Tampilan Utama SPSS for Windows

1. ***File***. Digunakan untuk membuat file baru atau membuka file yang telah tersimpan atau membaca file *speadsheets* dan file data base yang dibuat dengan paket program lain
2. ***Edit***. Digunakan untuk memodifikasi, mengkopi, menghapus, mencari dan mengganti data atau teks dari output window maupun sintax window.
3. ***Data.*** Digunakan untuk membuat pilihan global dari file data SPSS seperti pendefinisian variabel, penggabungan file, transpose data, mengambil sebagian case, dll.
4. ***Transform.*** Digunakan untuk mentransformasi data, yaitu pembentukan variabel baru yang valuenya merupakan hasil transformai dari value variabel-variabel yang sudah ada. Atau memodifikasi variabel yang sudah ada berdasarkan variabel lain. Seperti transformasi dengan operator aritmatika, fungsi aritmatika, fungsi statistik, dan sebagainya
5. ***Analyze***. Digunakan untuk memilih berbagai prosedur statistika, seperti : korelasi, regresi, analisa varians, tes non parametrik, dan sebagainya

6. ***Graph***. Digunakan untuk mengaktualiasikan data ke dalam bar chart, pie chart, histogram, scatter plot dan bentuk grafik lainnya.
7. ***Utilities.*** Digunakan untuk mengubah fonts, mengakses data secara dinamik, menampilkan informasi isi file data SPSS, atau menampilkan indeks dari perintah-perintah SPSS.
8. ***Window.*** Digunakan untuk mengatur, memilih, dan mengontrol atribut- atribut window SPSS
9. ***Help.*** Digunakan untuk membuka window standard Microsoft Help yang memuat informasi bantuan bagaimana menggunakan bantuan dari fasilitas pada SPSS. Informasi bantuan ini juga bisa didapat lewat setiap kotak dialog.

Latihan !

1. Buatlah sebuah file data seperti berikut ini

	nama	sex	gol	ms_kerja	gajiawal	gajiskrg
1	Leibnitz	1	2	4	300000.00	439230.00
2	Legendre	1	3	3	500000.00	665500.00
3	Mergie Siegers	2	1	0	200000.00	200000.00
4	Lagrange	1	2	1	300000.00	330000.00
5	Lipschitz	1	3	5	500000.00	805255.00
6	Fourier	1	2	3	300000.00	399300.00
7	Laplace	1	4	5	750000.00	1207882.00
8	Chaucy	1	3	1	500000.00	5500000.00
9	Aljabar	1	3	2	500000.00	605000.00
10	Yani Dani Wijaya	2	1	2	200000.00	242000.00

Catatan :
Label Value :

- Sex : 1 = laki-laki
  2 = perempuan
- Gol : 1 = lulusan SMU
  2 = lulusan D3
  3 = lulusan S1
  4 = lulusan S2

Tampilkan pula label value pada data editor
Hapus Grid Line

2. Diberikan data sebagai berikut :

data1	data2	data3
8.14	11.23	8.06
5.00	5.96	5.49
8.18	11.16	8.71
4.97	5.93	5.58
6.00	10.00	6.50
6.50	9.80	6.70
7.00	9.90	6.90
7.20	10.50	7.00
6.90	11.00	5.90
7.00	10.70	5.80

a. Urutkan data1 dari yang kecil ke yang besar dan sebaliknya !
b. Bagaimanakah statistik deskriptif dari data1 – data3 ?
c. Buat histogram data1 !
d. Bagaimanakah boxplot data1?
e. Bagaimanakah boxplot data1 – data3 ? Gambarkan secara keseluruhan !
f. Dapatkan nilai ln dari data1 !
g. Dapatkan nilai log data1 !
h. Dapatkan nilai $\frac{1}{data2}$
i. Uji apakah ragam ketiga data tersebut sama atau tidak !
j. Bagaimanakah plot data1 terhadap data2 ?

BAB II
KORELASI

Analisa korelasi mempelajari apakah ada hubungan antara dua variabel atau lebih. Keeratan hubungan linier antara 2 variabel (koefisien korelasi) dilambangkan dengan **r**, dan dirumuskan :

$$r = \frac{n\sum x_i y_i - \left(\sum x_i\right)\left(\sum y_i\right)}{\sqrt{n\sum x_i^2 - \left(\sum x_i\right)^2}\sqrt{n\sum y_i^2 - \left(\sum y_i\right)^2}}$$

Beberapa sifat **r** adalah :

1. **r** dapat bernilai positif atau negatif.
2. Terletak antara batas –1 dan +1
3. Sifat dasarnya simetris, yaitu koefisien korelasi antara X dan Y ($\mathbf{r}_{xy}$) sama dengan koefisien korelasi antara Y dan X ($\mathbf{r}_{yx}$).
4. **r** hanyalah suatu ukuran hubungan linier atau ketergantungan linier saja, **r** tidak mempunyai arti untuk menggambarkan hubungan non linier.

Hipotesis :

H_o : Tidak ada hubungan (korelasi) antara dua variabel
H_1 : Ada hubungan (korelasi) antara dua variabel

Pengambilan keputusan :

Jika probabilitas > 0.05 maka H_o diterima.
Jika probabilitas < 0.05 maka H_o ditolak.

Catatan :
Keputusan dapat diambil dengan cepat dengan melihat nilai koefisien korelasinya, yaitu jika pada nilai koefisien korelasi bertanda (**) maka menyatakan ada hubungan pada tingkat signifikansi 1%.

Contoh :
Hitung nilai koefisien korelasi dari 4 variabel berikut, kemudian interpretasikan.

X1	X2	X3	X4
0.12	0	0.33	300
0.20	25	1.35	275
0.13	50	1.56	250
0.15	75	1.86	225
0.14	100	2.03	200
0.11	125	2.56	175
0.22	150	3.02	150
0.23	175	3.56	125
0.20	200	4.59	100
0.21	225	6.25	75
0.18	250	6.35	50
0.17	275	7.05	25
0.16	300	8.65	0

Untuk menghitung nilai koefisien korelasi, dari menu utama SPSS pilih ***Analyze*** → ***Correlate Bivariate ….***
Maka akan ditampilkan kotak dialog *bivariate correlation.*

Semua variabel numerik pada file data anda akan ditampilkan pada kotak daftar variabel. Pindahkan dua atau lebih data di variabel asal ke variabel terseleksi (variabel yang akan diolah oleh user) kemudian **OK.**

Hasil Analisa :

Correlations

		X1	X2	X3	X4
X1	Pearson Correlation	1,000	,385	,330	-,385
	Sig. (2-tailed)	,	,194	,272	,194
	N	13	13	13	13
X2	Pearson Correlation	,385	1,000	,972**	-1,000**
	Sig. (2-tailed)	,194	,	,000	,000
	N	13	13	13	13
X3	Pearson Correlation	,330	,972**	1,000	-,972**
	Sig. (2-tailed)	,272	,000	,	,000
	N	13	13	13	13
X4	Pearson Correlation	-,385	-1,000**	-,972**	1,000
	Sig. (2-tailed)	,194	,000	,000	,
	N	13	13	13	13

**. Correlation is significant at the 0.01 level (2-tailed).

Interpretasi :

1. Arti angka korelasi

   Ada dua hal dalam penafsiran korelasi, yaitu tanda + atau - yang berhubungan dengan arah korelasi serta kuat tidaknya korelasi. Contohnya, korelasi antara (X1,X2) dan (X1,X3) didapat angka +0.385 dan +0.330 (tanda + disertakan karena tidak ada tanda '-' pada output, jadi otomatis positif).

   Hal ini berarti :

   - Arah korelasi positif, atau semakin tinggi X2, X1 semakin besar dan sebaliknya . Demikian juga jika semakin tinggi X3, X1 juga makin besar dan sebaliknya. Namun demikian jika koefisien korelasi bertanda negatif seperti korelasi pada (X1,X4) berarti bahwa semakin tinggi X4 maka X1 makin turun dan sebaliknya.
   - Besar korelasi yang < 0.5, berarti korelasinya lemah, seperti pada korelasi antara (X1,X2), (X1,X3) dan (X1,X4). Sedangkan koefisien korelasi yang > 0.5 berarti korelasi/hubungannya kuat, seperti korelasi antara (X2,X3) dan (X3,X4).

2. Signifikansi hasil korelasi
   - Nilai probabilitas dari korelasi antara (X1,X2), (X1,X3) dan (X1,X4) berturut-turut adalah 0.194, 0.272 dan 0.194 yang lebih besar dari 0.05, berarti bahwa korelasi antara (X1,X2), (X1,X3) dan (X1,X4) tidak nyata secara statistika atau tidak ada hubungan (korelasi) antara (X1,X2), (X1,X3) dan (X1,X4)
   - Nilai probabilitas dari korelasi antara (X2,X3), (X2,X4) dan (X3,X4) semuanya adalah 0.000 yang lebih kecil dari 0.05 dan 0.01, berarti korelasi antara (X2,X3), (X2,X4) dan (X3,X4) nyata secara statistika atau ada hubungan (korelasi) yang kuat antara (X2,X3), (X2,X4) dan (X3,X4).

3. Jumlah data yang berkorelasi

   Karena tidak ada data yang hilang, maka data yang diproses N=13.

Latihan !

1. Kapan nilai koefisien korelasi (**r**) antara 2 variabel bernilai +1 ?

2. Kapan nilai koefisien korelasi (**r**) antara 2 variabel bernilai –1?

3. Berikut ini diberikan nilai 10 orang mahasiswa dalam ujian tengah semester dan ujian akhir semester mata kuliah statistik. Hitung koefisien korelasinya dan interpretasikan hasil analisanya.

UTS	63	79	50	91	69	83	77	56	58	68
UAS	77	82	59	81	70	79	85	59	63	80

4. Diperkirakan banyaknya kaleng yang rusak dalam suatu pengiriman dengan mobil pengangkut (Y) berhubungan dengan kecepatan mobil pengangkut (X). Tiga belas mobil pengangkut yang dipilih secara acak digunakan untuk memeriksa apakah perkiraan ini benar atau tidak. Data yang diperoleh adalah sebagai berikut :

X	4	3	5	8	4	3	3	4	3	5	7	3	8
Y	27	54	86	136	65	109	28	75	53	33	168	47	52

   a. Plot data di atas
   b. Hitung koefisien regresinya
   c. Apa kesimpulan anda ? (Gunakan α = 5%)

5. Ingin diketahui apakah ada korelasi (hubungan) antara variabel-variabel :
   Jumlah pelanggaran lalu lintas (tilang), jumlah kendaraan roda empat (mobil), kendaraan roda dua (sepeda motor), dan jumlah polisi. Untuk itu, diambil data mengenai variabel-variabel di atas pada sejumlah daerah pada waktu tertentu dengan hasil sebagai berikut :

Tilang	Mobil	Motor	Polisi
20	258	589	89
24	265	587	52
25	249	698	59
18	125	625	57
15	*	712	52
16	124	692	48
*	251	681	49
10	*	634	29
12	124	697	27
17	159	521	59

   Hitunglah koefisien korelasi di antara 4 variabel tersebut dan interpretasikan.
   Catatan :
   Perhatikan ada beberapa data yang diberi tanda *. Hal ini menunjukkan data tersebut "*missing*" atau tidak diketahui.

BAB III
REGRESI LINIER SEDERHANA

Analisis regresi dapat digunakan untuk mendapatkan persamaan regresi yang menunjukkan hubungan antar variabel dependen (variabel tak bebas) dengan satu atau beberapa variabel independen (variabel bebas). Jika variabel dependen dihubungkan dengan satu variabel independen saja, maka regresi yang dihasilkan adalah regresi linier sederhana. Nilai koefisien yang dihasilkan harus diuji secara statistik signifikan atau tidak. Apabila semua koefisien signifikan, persamaan regresi yang dihasilkan dapat digunakan untuk memprediksi nilai variabel dependen jika nilai variabel independen ditentukan. Seberapa besar pengaruh variabel independen terhadap terhadap variasi variabel dependen dapat diukur dengan besarnya nilai koefisien determinasi (**R^2**). Semakin besar nilai koefisien determinasi, semakin besar pula pengaruh variabel independen terhadap variasi variabel dependen.
Model regresi linier sederhana :

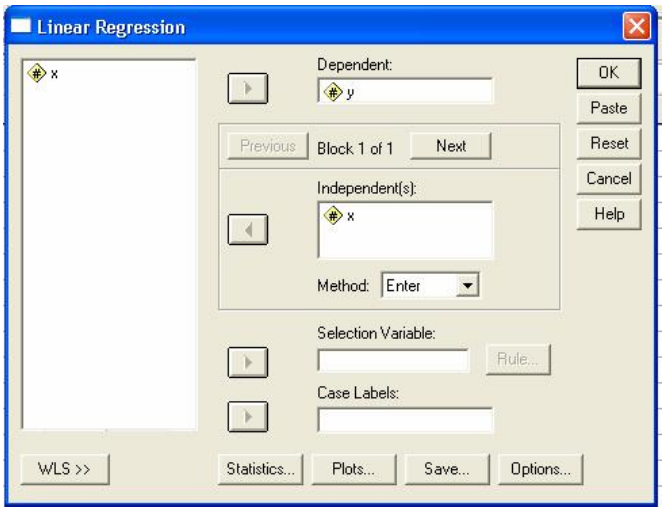
$$Y = \beta_0 + \beta_1 x$$

Contoh :
Data hipotetik (buatan) mengenai belanja konsumsi mingguan keluarga (Y) dan pendapatan mingguan keluarga (X) dalam dollar disajikan berikut ini.

Y	70	65	90	95	110	115	120	140	155	150
X	80	100	120	140	160	180	200	220	240	260

Dugalah model regresi dari data di atas!

Jawab :
Dari menu utama SPSS, pilih menu **Analyze** → **Regression** → **Linear….**
Tampak di layar tampilan sebagai berikut :



Dependent :Y
Independent(s) : X
Metod : Enter, OK

Hasil Analisa :

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	X[a]	,	Enter

a. All requested variables entered.

b. Dependent Variable: Y

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,981[a]	,962	,957	6,49

a. Predictors: (Constant), X

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	8552,727	1	8552,727	202,868	,000[a]
	Residual	337,273	8	42,159		
	Total	8890,000	9			

a. Predictors: (Constant), X

b. Dependent Variable: Y

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	24,455	6,414		3,813	,005
	X	,509	,036	,981	14,243	,000

a. Dependent Variable: Y

Interpretasi :

1. Model Summary
   - Angka R sebesar 0.981 menunjukkan bahwa korelasi / hubungan antara variabel tak bebas Y (belanja konsumsi mingguan keluarga) dengan variabel bebas X (pendapatan mingguan keluarga) adalah sangat kuat.
   - Angka R square atau koefisien determinasi adalah 0.962 (berasal dari 0.981 x 0.981). Namun jika jumlah variabel bebas lebih dari dua (misal dalam regresi berganda), lebih baik digunakan Adjusted Rsquare, yang bernilai 0.957 (yang selalu lebih kecil dari R square). Hal ini berarti bahwa 95.7% variasi dari Y = belanja konsumsi mingguan keluarga, bisa dijelaskan oleh variasi dari X = pendapatan mingguan keluarga. Sedangkan sisanya (100% - 95.7% = 4.3%) dijelaskan oleh sebab-sebab yang lain.
   - Standard Error of Estimate (SEE) adalah 6.49 atau $6.49 (satuan yang dipakai adalah variabel tak bebas). Makin kecil SEE akan membuat model regresi semakin tepat dalam memprediksi variabel dependen

2. ANOVA
   Dari uji ANOVA atau F-test, didapat F_{hitung} adalah 202.868 dengan tingkat signifikansi 0.000. Karena probabilitas (0.000) jauh lebih kecil dari 0.05, maka model regresi bisa dipakai untuk memprediksi Y.
3. Koefisien Regresi
   - Persamaan Regresi :
     Y = 24.455 + 0.509X
   - Konstanta sebesar 24.455 menyatakan bahwa jika tidak ada pendapatan mingguan keluarga, maka belanja konsumsi mingguan keluarga adalah $24.455
   - Koefisien regresi 0.509 menyatakan bahwa setiap penambahan (karena tanda +) $1 pendapatan mingguan keluarga akan meningkatkan belanja konsumsi keluarga sebesar $0.509
   - Uji t untuk menguji signifikansi konstanta dan variabel bebas
     Hipotesis :
     H_0 : Koefisien regresi tidak signifikan
     H_1 : Koefisien regresi signifikan
     Pengambilan keputusan (berdasarkan probabilitas)
     Jika probabilitas > 0.05 maka H_o diterima
     Jika probabilitas < 0.05 maka H_o ditolak
     Keputusan :
     Terlihat bahwa pada kolom *sig / significance* : variabel X mempunyai angka signifikan 0.000 (di bawah 0.05). Karena itu, X = pendapatan mingguan keluarga memang mempengaruhi Y= belanja konsumsi mingguan keluarga.

Latihan

1. Biaya pemeliharaan traktor pengangkat barang terus meningkat dengan bertambahnya umur traktor. Berikut ini adalah datanya :

Umur (tahun)	Biaya ($)	Umur (tahun)	Biaya ($)
4.5	619	5.0	1194
4.5	1049	0.5	163
4.5	1033	0.5	182
4.0	495	6.0	764
4.0	723	6.0	1373
4.0	681	1.0	978
5.0	890	1.0	466
5.0	1522	1.0	549
5.5	987		

   a. Tentukan mana sebagai variabel bebas dan mana variabel tak bebas. Beri alasan.
   b. Dugalah model regresi dari data tersebut. Apa kesimpulannya.

2. Tinggi busa sabun dalam bak cuci piring sangat penting bagi produsen sabun. Sebuah percobaan diadakan dengan mengubah-ubah banyaknya sabun dan kemudian mengukur tinggi busa dalam bak cuci piring yang standar setelah dilakukan pengadukan tertentu. Data yang diperoleh adalah sebagai berikut :

Gram Produk (X)	Tinggi Busa (Y)
4.0	33
4.5	42
5.0	45
5.5	51
6.0	53
6.5	61
7.0	62

Misalkan model yang digunakan adalah :

$$Y = \beta_0 + \beta_1 x + \varepsilon$$

a. Tentukan persamaaan regresinya
b. Adakan pengujian terhadap koefisisen regresinya.
c. Hitunglah sisaan-sisaannya dan periksa apakah ada petunjuk perlunya digunakan model yang lebih rumit.

3. Dalam suatu percobaan yang serupa dengan soal nomor 2, penelitinya mengatakan bahwa model $Y = \beta_0 + \beta_1 x + \varepsilon$ adalah "tidak masuk akal kecuali $\beta_0 = 0$, sebab semua orang tahu kalau anda tidak memasukkan sabun ke dalam bak cuci piring, maka busa tidak akan terbentuk." Jadi, ia berkeras menggunakan model $Y = \beta_1 x + \varepsilon$. Data yang diperoleh adalah :

Gram Produk (X)	Tinggi Busa (Y)
3.5	24.4
4.0	32.1
4.5	37.1
5.0	40.4
5.5	43.3
6.0	51.4
6.5	61.9
7.0	66.1
7.5	77.2
8.0	79.2

a. Bila model itu diterima, tentukan nilai dugaan bagi β_1
b. Dengan menggunakan persamaan regresinya, dugalah $\hat{Y}$ untuk setiap X.
c. Periksalah sisaan-sisaannya
d. Apa kesimpulan anda dan berikan saran bagi peneliti tersebut.

BAB IV

REGRESI LINIER BERGANDA

Analisa regresi merupakan metode statistika yang mencari hubungan antara 2 peubah, yaitu peubah bebas (X) dan peubah tak bebas (Y). Jika variabel tak bebas (Y) dihubungkan lebih dari satu variabel bebas (X), maka didapat persamaan regresi linier berganda (multiple Linear Regression)
Model :

Tahapan penyusunan model regresi berganda meliputi :

- Menentukan mana variabel bebas (independen) dan mana variabel tak bebas (dependen)
- Menentukan metode pembuatan model regresi (enter, Stepwise, Forward, Backward)
- Melihat ada tidaknya data yang outlier (ekstrim)
- Menguji asumsi-asumsi pada regresi berganda, sepert normalitas, linieritas, heteroskedastisitas, multikolinieritas, autokorelasi dan lainnya.
- Menguji signifikansi model (uji t, uji F dan sebagainya)
- Interpretasi model regresi berganda

Untuk SPSS, prosedur regresi linier menyediakan 5 macam metode untuk menyusun persamaan regresi, yaitu pemilihan Enter, Stepwise, Forward, Eliminasi Backward dan removal. Karena regresi berganda cukup kompleks dan bervariasi, pada modul ini akan dijelaskan prosedur pembuatan model regresi berganda dan uji signifikansinya dengan metode enter.

Contoh :
Suhu plat pembungkus dan jarak plat pembungkus dalam mesin pembungkus sabun mempengaruhi persentase sabun terbungkus yang lolos inspeksi. Beberapa data tentang peubah itu talah dikumpulkan, yaitu sebagai berikut :

a. Gunakan moodel linier dan tentukan nilai dugaan bagi
b. Apakah secar keseluruhan regresi itu nyata ? (guanakan =0.05)
c. Apakah salah satu variabel itu lebih bermanfaat daripada yang lain didalam meramalkan persentase sabun yang terbungkus rapi ?
d. Apa saran anda berkaitan dengan bekerjanya mesin pembungkus ini ?

Jawab :
Dari menu utama SPSS, pilih menu **Analyze** ▸ **Regression** ▸ **Linear**
Dependent : Y
Independent (s) : X1, X2
Method : Enter
OK

Hasil Analisa :
Interpretasi :

4. Model Summary
   - Angka R sebesar 0.865 menunjukkan bahwa korelasi / hubungan antara variabel tak bebas Y (persentase sabun terbungkus) dengan 2 variabel bebas X (jarak dan suhu plat pembungkus) adalah sangat kuat.
   - Angka R square atau koefisien determinasi adalah 0.747 (berasal dari 0.865 x 0.865). Namun karena jumlah variabel bebas lebih dari dua maka lebih baik digunakan Adjusted Rsquare, yang bernilai 0.709 (yang selalu lebih kecil dari R square). Hal ini berarti bahwa 70.9% variasi dari Y = persentase sabun terbungkus, bisa dijelaskan

oleh variasi dari X = pendapatan mingguan keluarga. Sedangkan sisanya (100% - 95.7% = 4.3%) dijelaskan oleh sebab-sebab yang lain.

- Standard Error of Estimate (SEE) adalah 6.49 atau $6.49 (satuan yang dipakai adalah variabel tak bebas). Makin kecil SEE akan membuat model regresi semakin tepat dalam memprediksi variabel dependen

5. ANOVA

Dari uji ANOVA atau F-test, didapat F_{hitung} adalah 202.868 dengan tingkat signifikansi 0.000. Karena probabilitas (0.000) jauh lebih kecil dari 0.05, maka model regresi bisa dipakai untuk memprediksi Y.

6. Koefisien Regresi
   - Persamaan Regresi :

   Y = 24.455 + 0.509X
   - Konstanta sebesar 24.455 menyatakan bahwa jika tidak ada pendapatan mingguan keluarga, maka belanja konsumsi mingguan keluarga adalah $24.455
   - Koefisien regresi 0.509 menyatakan bahwa setiap penambahan (karena tanda +) $1 pendapatan mingguan keluarga akan meningkatkan belanja konsumsi keluarga sebesar $0.509
   - Uji t untuk menguji signifikansi konstanta dan variabel bebas

   Hipotesis :

   H_0 : Koefisien regresi tidak signifikan

   H_1 : Koefisien regresi signifikan

   Pengambilan keputusan (berdasarkan probabilitas)

   Jika probabilitas > 0.05 maka H_o diterima

   Jika probabilitas < 0.05 maka H_o ditolak

   Keputusan :

   Terlihat bahwa pada kolom *sig / significance* : variabel X mempunyai angka signifikan 0.000 (di bawah 0.05). Karena itu, X = pendapatan mingguan keluarga memang mempengaruhi Y= belanja konsumsi mingguan keluarga.

Latihan !

1. Dugalah model dengan menggunakan data di bawah ini
2. Dugalah model dengan menggunakan data di bawah ini. Susunlah tabel analisa ragamnya, dan lakukan uji F parsial untuk menguji H melawan untuk I = 1,2
3. Data

a. Dugalah semua dalam model
b. Bauat tabel analisa ragamnya
c. Dengan menggunakan =0.05, ujilah apakah regresi keseluruhannya secara statistik nyata atau tidak
d. Hitunglah kuadarat koefisien korelasi gandanya, yaitu R2. Berapa proporsi keragaman total di sekitar yang dapat dijelaskan oleh kedua peubah tersebut ?
e. Seberapa bermanfaatkah regresi ini jika hanya menggunakan X1 saja? Berapa sumbangan X2 jika X2 sudah ada dalam regresi?
f. Seberapa bermanfaatkah regresi ini jika hanya menggunakan X2 saja ? Berapa sumbangan X1 jika X2 sudah ada dalam regresi ?
g. Apa kesimpulan anda ?

BAB V
BEBERAPA BENTUK MODEL DENGAN EKONOMETRIK

Dalam bab sebelumnya kita telah belajar regresi linier ssederhana dan regresi linier berganda. Model regresi majemuk yang paling sederhana adalah regresi tiga variabel dengan satu variabel tak bebas dan dua variabel bebas. Jika pada bab sebelumnya kita membahas model regresi majemuk yang linier dalam parameter dan linier dalam variabel. Pada bab ini kita akan belajar model regresi majemuk yang linier dalam parameter tetapi tidak linier dalam variabel. Dalam contoh ekonometrika misalnya, diasumsikan secara implicit bahwa hanya pendapatan Y yang mempengaruhi X. tetapi teori ekonomi jarang yang sesederhana itu, karena disamping pendapatan, sejumlah variabel lain nampaknya juga mempengaruhi belanja konsumsi. Sebagai contoh, permintaan untuk barang dagangan nampaknya tidak hanya tergantung pada harganya sendiri, tetapi juga pada harga barang yang lain yang bersaing. Pada ekonometrika, kita mengenal fungsi produksi Cobb-Douglas sbb:

$$Y = \beta_0 X_1^{\beta_1} \; X_2^{\beta_2}$$

Model regresi di atas merupakan model regresi majemuk yang linier dalam parameter tetapi tidak linier dalam variabel. Persamaan di atas bisa juga dinyatakan sebagai :

$$\text{Ln } \gamma = \alpha + \beta_1 \text{ ln } X_1 + \beta_2 \text{ ln } X_2$$

Dimana $\alpha = \text{ln } \beta_0$

Contoh :
Dari data berikut ini, buatlah sebuah model regresi dari fungsi Cobb-Douglas !

Y	X_1	X_2
16607.7	275.5	17803.7
17511.3	274.4	18096.8
20171.2	269.7	18271.8
20932.9	267.0	19167.3
20406.0	267.8	19647.6
20837.6	275.0	20803.5
24806.3	283.0	22076.6
26465.8	300.7	23445.2
27403.0	307.5	24939.0
28628.7	303.7	26713.7
29904.5	304.7	29957.8
27508.2	298.6	31585.9
29035.5	295.5	33474.5
29281.5	299.0	34821.8
31535.8	288.1	41794.3

JAWAB :

Terlebih dahulu, kita mencari nilai ln dari masing-masing variable. Didapat data sebagai berikut:

Ln Y	Ln X_1	Ln X_2
9.72	5.62	9.79
9.77	5.61	9.80
9.91	5.60	9.81
9.95	5.59	9.86
9.92	5.59	9.89
9.94	5.62	9.94

10.12	5.65	10.00
10.18	5.71	10.06
10.22	5.73	10.12
10.26	5.72	10.19
10.31	5.72	10.31
10.22	5.70	10.36
10.28	5.69	10.42
10.28	5.70	10.46
10.36	5.66	10.64

Dari menu utama SPSS pilih menu **Analyze** ➤ **Regression** ➤ **Linier…….**
Dependent : Ln Y
Independent (s) : Ln X1, Ln X2
Method : Enter
OK

Hasil Analisa :
Regression

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	LN_X2, LN_X1[a]	.	Enter

a. All requested variables entered.

b. Dependent Variable: LN_Y

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.946[a]	.895	.878	7.278E-02

a. Predictors: (Constant), LN_X2, LN_X1

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	.544	2	.272	51.328	.000[a]
	Residual	6.357E-02	12	5.297E-03		
	Total	.607	14			

a. Predictors: (Constant), LN_X2, LN_X1

b. Dependent Variable: LN_Y

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-3.557	2.331		-1.526	.153
	LN_X1	1.540	.509	.385	3.027	.011
	LN_X2	.488	.097	.642	5.051	.000

a. Dependent Variable: LN_Y

BAB VI
HETEROSKEDASTISITAS

Model regresi menghendaki tidak adanya masalah heteroskedastisitas. Jika varian dari galat suatu pengamatan ke pengamatan yang lain tetap, maka disebut homoskedastisitas. Jika varian berbeda, disebut heteroskedastisitas. Heteroskedastisitas tidak merusak sifat ketidakbiasan dan konsistensi dari penaksir Metode Kuadrat Terkecil. Tetapi, penaksir ini tidak lagi mempunyai varian minimum atau efisien. Dengan kata lain, mereka tidak lagi BLUE (*Best Linear Unbiased Estimator* / Penaksir tak bias linier terbaik). Penaksir yang BLUE diberikan olaeh Metode Kuadrat Terkecil Tertimbang. Contoh masalah heteroskedastisitas adalah orang kaya akan bervaraiasi dalam membelanjakan uangnya, sedangkan orang miskin hanya bisa sedikit bervariasi dalam berbelanja. Hal ini menunjukkan variasi yang tidak sama antara kedua golongan tersebut, yang berarti timbul masalah heteroskedastisitas.

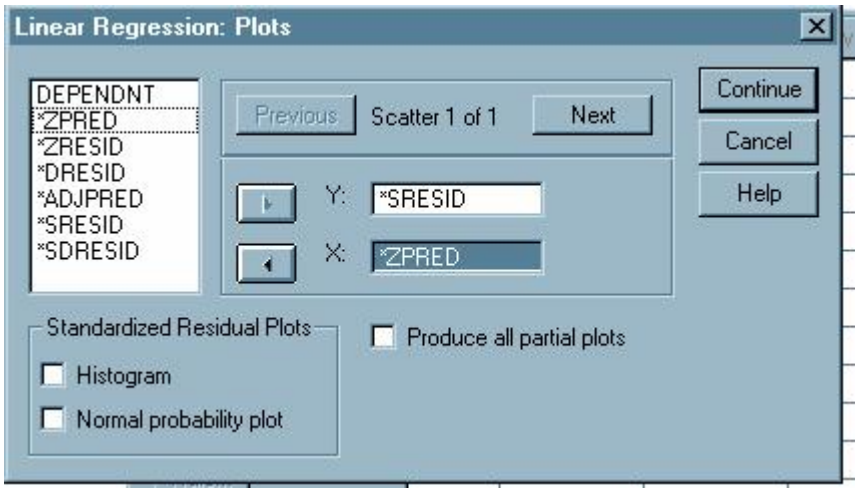
Contoh :
Data berikut menyajikan rata-rata produktivitas tenaga kerja dan standar deviasi dari produktivitas tenaga kerja untuk sembilan kelas jumlah karyawan

Jumlah karyawan	Rata-rata Produktivitas	Standar Deviasi Produktivitas
1 – 4	9355	2487
5 – 9	8584	2642
10 -19	7962	3055
20 – 49	8275	2706
50 – 99	8389	3119
100 – 249	9418	4493
250 – 499	9795	4910
500 – 999	10281	5893
1000 - 2499	11750	5550

Apa yang dapat anda katakan mengenai sifat heteroskedastisitas, jika ada dalam data tadi ?

Jawab :
Dari menu utama SPSS, pilih menu **Analyze** ➤ **Regression** ➤ **Linear….**
Dependent : Y
Independent(s) : X
Method : Enter

Tekan tombol **Plot …**, hingga tampak di layar tampilan seperti pada Gambar 6.1 berikut ini



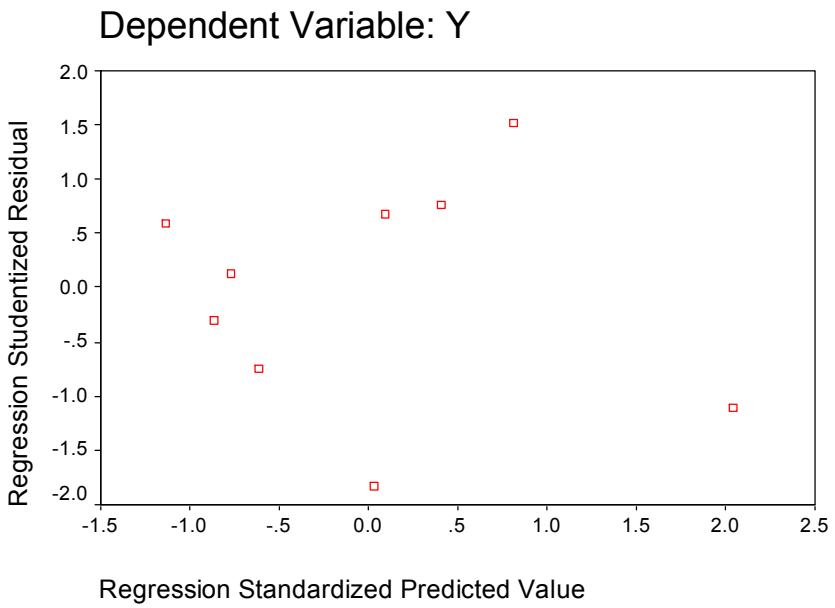
Gambar 6.1 Kotak Dialog Statistik

Masukkan variabel SRESID pada sumbu (pilihan) Y
Masukkan variabel ZPRED pada sumbu (pilihan) X
Continue, OK

Catatan :
Deteksi adanya heteroskedastisitas
Deteksi dengan melihat ada tidaknya pola tertentu pada grafik yang dihasilkan, dimana sumbu X adalah Y yang telah diprediksi, dan sumbu X adalah residual (Y prediksi – Y sesungguhnya) yang telah distudentized.
Dasar pengambilan keputusan :

- Jika ada pola tertentu, seperti titik-titik (point-point) yang ada membentuk suatu pola tertentu yang teratur (bergelombang, melebar kemudian menyempit), maka telah terjadi heteroskedastisitas
- Jika tidak ada pola yang jelas, serta titik-titik menyebar di atas dan di bawah angka 0 pada sumbu Y , maka tidak terjadi heteroskedastisitas.

Hasil Analisa :
(Hanya tampilan grafik saja)



Interpretasi :

Dari grafik di atas, terlihat titik-titk menyebar dengan membentuk pola tertentu. Hal ini berarti terjadi heteroskedastisitas pada model regresi, sehingga model regresi tidak layak dipakai untuk prediksi.

Langkah untuk menghilangkan heteroskedastisitas.

Transformasi Log seringkali akan mengurangi heteroskedastisitas. Hal ini disebabkan karena transformasi yang memampatkan skala untuk pengukuran variabel, menguragi perbedaan antara kedua nilai tadi dari sepuluh kali lipat menjadi perbedaan dua kali lipat. Misal, angka 80 adalah 10 kali angka 8, tetapi ln 80 (= 4.3820) hanya dua kali besarnya ln 8 (= 2.0794).

Perhatikan dua model regresi berikut :

$Y_i = \beta_0 + \beta_1 X_i + u_i$ (persamaan 6.1)

$\text{Ln } Y_i = \beta_0 + \beta_1 \ln X_i + u_i$ (persamaan 6.2)

Manfaat tambahan dari transformasi log bahwa koefisien kemiringan β_1 mengukur elastisitas dari Y terhadap X, yaitu persentase perubahan dalam Y untuk persentase perubahan dalam X. Misalnya, jika Y adalah konsumsi dan X adalah pendapatan, maka β_1 (dalam persamaan 6.2) akan mengukur elastisitas pendapatan, sedangkan dalam model regresi asli (persamaan 6.1) β_1, hanya mengukur tingkat rata-rata pertumbuhan konsumsi untuk satu unit perubahan dalam pendapatan.

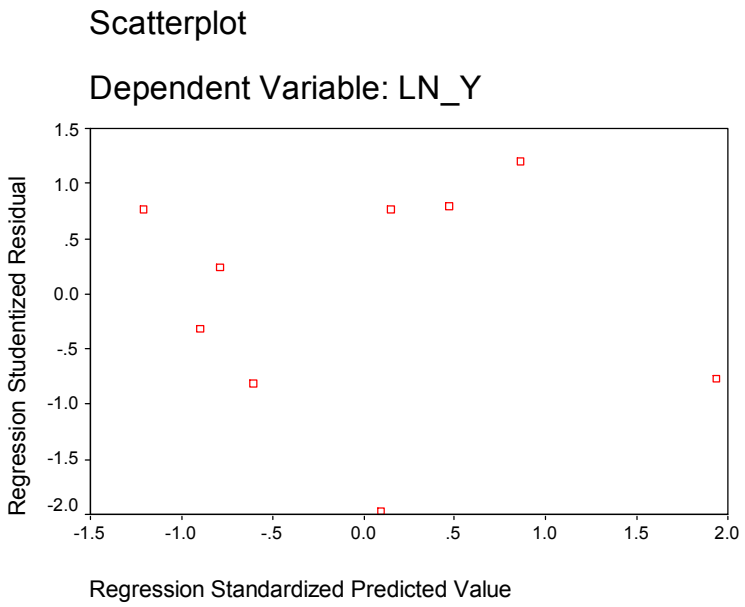
Dari data yang sama di atas, kita akan hitung nilai Ln X dan Ln Y, didapat :

X	Y	Ln X	Ln Y
9355	2487	9.14	7.82
8584	2642	9.06	7.88
7962	3055	8.98	8.02
8275	2706	9.02	7.90
8389	3119	9.03	8.05
9418	4493	9.15	8.41
9795	4910	9.19	8.50

10281	5893	9.24	8.68
11750	5550	9.37	8.62

Dari menu utama SPSS, pilih menu **Analyze** ➤ **Regression** ➤ **Linear….**
Dependent : Ln Y
Independent(S) : Ln X
Method : Enter
Tekan tombol **Plot …,** dan masukkan variabel SRESID pada sumbu (pilihan) Y dan masukkan variabel ZPRED pada sumbu (pilihan) X.
Continue, OK

Hasil Analisa :



Interpretasi :

Dari grafik di atas, terlihat titik-titk menyebar secara acak dan tidak berpola. Hal ini berarti tidak terjadi problem heteroskedastisitas pada model regresi, sehingga model regresi layak dipakai untuk prediksi.

BAB VII

AUTOKORELASI

Dalam analisis regresi diasumsikan bahwa galat acaknya tidak saling berkorelasi satu sama lain. Kalau asumsi ini tidak terpenuhi berarti ada korelasi diantara galat acaknya, maka estimasi koefisiennya tetap tidak bebrbias tetapi variasi dari koefisien itu tidak minimal lagi. Autokorelasi didefinisikan sebagi terjadinya korelasi diantara data-data pengamatan, atau dengan kata lain munculnya suatu data dipengaruhi data sebelumnya. Artinya, ada korelasi antar anggota sampel yang diurutkan berdasar waktu. Tentu saja model regresi yang baik adalah regresi yang bebas dari autokorelasi. Autukorelasi pada sebagian besar kasus ditemukan pada regresi yang datanya adalah *time series*, atau berdasarkan waktu berkala, seperti bulanan, tahunan, dan sebagainya.

Contoh :
Y (nilai penjualan) : 64 61 84 70 88 92 72 77
X (biaya promosi) : 20 16 34 23 27 32 18 22
Buktikan apakah terdapat autokorelasi dari data di atas ?

Jawab :
Dari menu utama SPSS, pilih menu **Analyze** ► **Regression** ► **Linear**
Dependent : Y
Independent (s) : X
Method : Enter
Tekan tombol, **Statistics …,** dan aktifkan pilihan Durbin-Watson pada bagian residuals
Continue, OK

Catatan :
Deteksi adanya autokorelasi
Ada tidaknya autokorelasi dapat dilihat dari statistik d Durbin-Watson. Nilai statistik d terletak antara 0 dan 4. Bila :

D-W	Kesimpulan
< 1.10	Ada autokorelasi positif
1.10 – 1.54	Tanpa Kesimpulan
1.55 – 2.46	Tidak Ada Autokorelasi
2.47 – 2.90	Tanpa Kesimpulan
> 2.91	Ada Autokorelasi Negatif

Jika ada masalah autokorelasi, maka model regresi yang seharusnya signifikan (lihat angka F dan signifikansinya), menjadi tidak layak untuk dipaklai. Autokorelasi bisa diatasi dengan beberapa cara antara lain :

- Melakukan teransformasi data
- Menambah data observasi

Hasil Analisa :

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	X[a]	.	Enter

a. All requested variables entered.

b. Dependent Variable: Y

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.862[a]	.743	.700	6.16	.830

a. Predictors: (Constant), X

b. Dependent Variable: Y

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	658.503	1	658.503	17.367	.006[a]
	Residual	227.497	6	37.916		
	Total	886.000	7			

a. Predictors: (Constant), X

b. Dependent Variable: Y

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	40.082	8.890		4.509	.004
	X	1.497	.359	.862	4.167	.006

a. Dependent Variable: Y

Interpretasi :
Pada bagian model summary, terlihat angka Durbin-Watson sebesar +0.830. Hal ini berarti model regresi di atas terdapat masalah autokorelasi positif, sehingga model regresi yang didapat menjadi tidak layak untuk dipakai.

Langkah untuk menghilangkan autokorelasi :
Autokorelasi dapat timbul karena berbagai alasan. Jika terjadi autokorelasi, maka diperlukan tindakan perbaikan. Jika koefisien autokorelasi diketahui, masalah autokorelasi dapat diatasi dengan transformasi data mengikuti prosedur persamaan perbedaan yang digeneralisasikan. Namun tidak selalu dengan transformasi, dapat menyelesaikan masalah autokorelasi. Untuk data kita di atas, dicoba dengan transformasi ln.

Y	X	Ln Y	Ln X

64	20	4.16	3.00
61	16	4.11	2.77
84	34	4.43	3.53
70	23	4.25	3.14
88	27	4.48	3.30
92	32	4.52	3.47
72	18	4.28	2.89
77	22	4.34	3.09

Dari menu utama SPSS, pilih menu **Analyze ➔ Regression ➔ Linear ….**
Dependent : Y
Independent (s) : X2
Method : Enter
Tekan tombol **Statistics…,** dan aktifkan pilihan Durbin- Watson pada bagian residuals.
Continue, OK.

Hasil Analisa :

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	LN_X[a]	.	Enter

a. All requested variables entered.

b. Dependent Variable: LN_Y

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.873[a]	.763	.723	7.832E-02	.876

a. Predictors: (Constant), LN_X

b. Dependent Variable: LN_Y

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	.118	1	.118	19.270	.005[a]
	Residual	3.680E-02	6	6.134E-03		
	Total	.155	7			

a. Predictors: (Constant), LN_X

b. Dependent Variable: LN_Y

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	2.791	.350		7.982	.000
	LN_X	.486	.111	.873	4.390	.005

a. Dependent Variable: LN_Y

Interpretasi :
Meskipun telah dicoba untuk melakukan transformasi data dengan transformasi ln, namun angka Durbin-Watson (0.876) tetap menunjukkan adanya autokorelasi positif. Hal ini berarti model regresi di atas terdapat masalah autokorelasi, sehingga model regresi yang didapat menjadi tidak layak untuk dipakai. Perlu dicoba cara lain untuk menghilangkan masalah autokorelasi misalnya dengan menambah data observasi.

BAB VIII

MULTIKOLINIERITAS

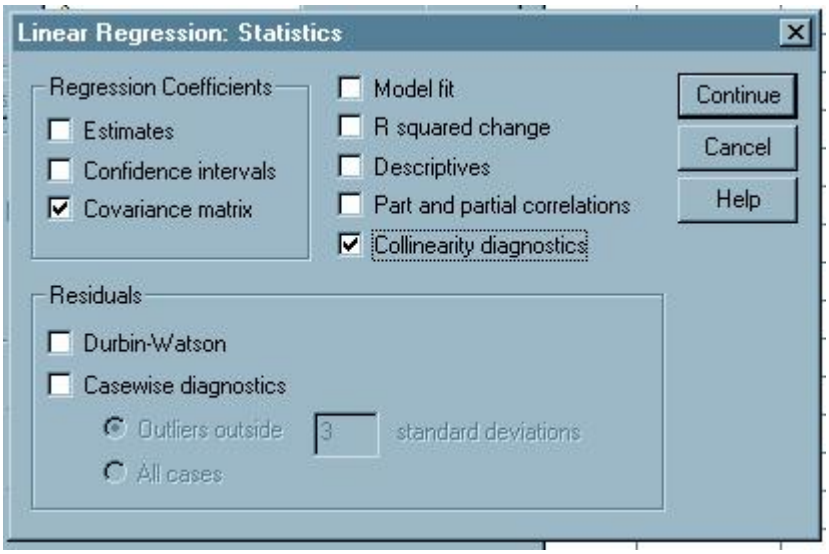
Satu asumsi dari analisis regresi adalah tidak adanya korelasi antar variabel independen (variabel bebas). Jika terjadi korelasi, maka dinamakan terdapat problem multikolinieritas. Tanda yang paling jelas dari multikolinieritas adalah adalah ketika $\mathbf{R^2}$ sangat tinggi (0.7 – 1), tetapi tidak satu pun koefisien regresi yang signifikan (nyata) secara statistik. Salah satu cara untuk mengatasi multikolinieritas adalah dengan transformasi data dan mengeluarkan satu dari variabel yang berkolinier.

Contoh :
Data hipotetik mengenai belanja konsumsi Y, pendapatan X2 dan kekayaan X3.

Y	X2	X3
70	80	810
65	100	1009
90	120	1273
95	140	1425
110	160	1633
115	180	1876
120	200	2052
140	220	2201
155	240	2435
150	260	2686

Uji apakah terdapat multikolinieritas dari data di atas !

Jawab :
Dari menu utama SPSS, pilih menu **Analyze** ➤ **Regression** ➤ **Linear ….**
Dependent : Y
Independent (s) : X2, X3
Method : Enter
Tekan tombol **Statistic …,** hingga tampak di layar tampilan seperti pada Gambar 8.1 berikut ini.



Gambar 8.1 Kotak Dialog Statistik

Non aktifkan pilihan Estimates dan Model Fit
Aktifkan pilihan Covariance matrix dan Collinearity diagnostics
Continue, OK.

Catatan :
Deteksi adanya multikolinieritas

a. Besaran VIF (Variance Inflation Factor) dan Tolerance
   Pedoman suatu model regresi yang bebas multikolinieritas adalah :
   - Mempunyai nilai VIF di sekitar angka 1
   - Mempunyai angka TOLERANCE mendekati 1.

b. Besaran korelasi antar variabel independent
   Pedoman suatu model regresi yang bebas multikolinieritas adalah koefisien korelasi antar variabel independen haruslah lemah (di bawah 0.5). Jika korelasi kuat, maka terjadi problem multikolinieritas.

c. Pengujian pada eigen value, jika eigen value mendekati nol maka terjadi multikolinieritas.

Hasil Analisa :

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	X3, X2[a]	.	Enter

a. All requested variables entered.

b. Dependent Variable: Y

Model Summary[a]

a. Dependent Variable: Y

Coefficients[a]

Model		Collinearity Statistics	
		Tolerance	VIF
1	X2	.002	482.128
	X3	.002	482.128

a. Dependent Variable: Y

Coefficient Correlations[a]

Model			X3	X2
1	Correlations	X3	1.000	-.999
		X2	-.999	1.000
	Covariances	X3	6.507E-03	-6.63E-02
		X2	-6.63E-02	.677

a. Dependent Variable: Y

Collinearity Diagnostics[a]

Model	Dimension	Eigenvalue	Condition Index	Variance Proportions (Constant)	X2	X3
1	1	2.930	1.000	.01	.00	.00
	2	6.971E-02	6.483	.98	.00	.00
	3	1.060E-04	166.245	.00	1.00	1.00

a. Dependent Variable: Y

Interpretasi :
Pada bagian COEFFICIENT terlihat angka untuk kedua variabel bebas angka VIF = 482.128 yang jauh lebih besar dari 1. Demikian juga nilai TOLERANCE = 0.002 yang jauh lebih kecil dari 1. Dengan demikian dapat disimpulkan bahwa pada model regresi tersebut terdapat problem multikolinieritas. Koefisien korelasi antar variabel bebas (X2 dan X3) sangat kuat, yaitu sebesar –0.999. Hal ini menunjukkan adanya problem multikolinieritas dalam model regresi di atas.

Langkah untuk menghilangkan multikolinieritas.

Salah satu cara untuk menghilangkan problem multikolinieritas adalah dengan mengeluarkan salah satu variabel dan melakukan transformasi data. Berikut ini dicoba untuk menghilangkan problem multikolinieritas dengan mengeluarkan salah satu variable. Dari soal di atas, dicoba untuk membuat model regresi baru tanpa melibatkan X3 (kekayaan).

Dari menu utama SPSS, pilih menu **Analyze** ▶ **Regression** ▶ **Linear ….**
Dependent : Y
Independent (s) : X2Method : Enter
Tekan tombol **Statistics…,** non aktifkan pilihan Estimates dan Model Fit dan aktifkan pilihan Covariance matrix dan Collinearity diagnostics
Continue, OK.

Hasil Analisa :

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	X2[a]	.	Enter

a. All requested variables entered.

b. Dependent Variable: Y

Coefficients[a]

Model		Collinearity Statistics Tolerance	VIF
1	X2	1.000	1.000

a. Dependent Variable: Y

Coefficient Correlations[a]

Model			X2
1	Correlations	X2	1.000
	Covariances	X2	1.278E-03

a. Dependent Variable: Y

Collinearity Diagnostics[a]

Model	Dimension	Eigenvalue	Condition Index	Variance Proportions	
				(Constant)	X2
1	1	1.947	1.000	.03	.03
	2	5.263E-02	6.083	.97	.97

a. Dependent Variable: Y

Interpretasi :

Pada bagian COEFFICIENT terlihat angka VIF dan TOLERANCE untuk variable bebas X2 = 1.000. Dengan demikian dapat disimpulkan bahwa pada model regresi tanpa melibatkan X3 tidak terdapat problem multikolinieritas, sehingga model tersebut layak untuk digunakan untuk menduga Y = belanja konsumsi.

BAB IX

REGRESI ATAS VARIABEL DUMMY

Variabel dummy adalah variabel yang digunakan untuk membuat kategori data yang bersifat kualitatif (nominal). Seperti gender (jenis kelamin) yang terdiri atas pria dan wanita. Dalam model regresi yang tidak mengenal data berupa huruf, maka pria dan wanita tersebut perlu diubah menjadi kode tertentu, seperti 1 untuk pria dan 0 untuk wanita. Variabel pria dan wanita inilah yang disebut variabel dummy.

Model regresi variabel dummy : $Y_i = B_1 + B_2D_i + \mu_i$

di mana :

D_i = 0, gagal jika sampel tidak punya sifat yang dimaksud.
1, sukses jika sampel punya sifat yang dimaksud.

Contoh :

Data :

Y = Gaji awal

X = 1, jika lulusan perguruan tinggi
= 0, jika lulusan selaian perguruan tinggi.

X	1	0	0	1	1	0	1	0	0	1
Y	21.2	17.5	17	20.5	21	18.5	21.7	18	19	22

Buatlah model regresi dengan variabel dummy dari data diatas!

Jawab :

Untuk membuat model regresi dengan variabel dummy, dari menu utama SPSS, pilih menu **Analyze → Regression → Linear….**

Dependent : Y

Independent (s) : X

Method : Enter

Continue, OK

Hasil Analisa :

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	X[a]	.	Enter

a. All requested variables entered.

b. Dependent Variable: Y

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.935[a]	.874	.858	.697

a. Predictors: (Constant), X

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	26.896	1	26.896	55.342	.000[a]
	Residual	3.888	8	.486		
	Total	30.784	9			

a. Predictors: (Constant), X

b. Dependent Variable: Y

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	18.000	.312		57.735	.000
	X	3.280	.441	.935	7.439	.000

a. Dependent Variable: Y

Interpretasi :

1. Model Summary
   Angka R sebesar 0.935 menunjukkan angka korelasi yang sangat kuat antara gaji awal dan lulusan pegawai
2. Anova
   Dari uji Anova, didapat F hitung adalah 55.342 dengan tingkat signifikansi 0.000 yang masih di bawah 0.05. Hal ini menunjukkan pengaruh variable independent (lulusan) signifikan terhadap gaji awal yang diterima.
3. Koefisien Regresi
   Persamaan regresi yang didapat :
   Gaji awal = 18.000 + 3.280D_i
   Uji t untuk menguji signifikansi konstanta dan variable independent.
   Terlihat bahwa pada kolom sig / significance , variable independent dan konstanta mempunyai tingkat signifikansi di bawah 0.05 (0.000 untuk konstanta dan lulusan). Hal ini berarti model regresi sudah layak untuk memprediksi pengaruh lulusan pada gaji awal yang diterima.

Tafsiran :

- Konstanta 18.000 menunjukkan gaji awal yang diterima :
  - Lulusan perguruan tinggi (D=1) adalah :
    Gaji awal = 18.000 + 3.28 (1) = 21.280
    atau $21.280/bulan
  - Lulusan non perguruan tinggi (D=0) adalah :
    Gaji awal = 18.000 + 3.28 (0) = 18.000
    atau $18.000

  Jadi di sini jelas ada perbedaan antara gaji lulusan perguruan tinggi dan non perguruan tinggi.
- Koefisien 3.280 menunjukkan bahwa lulusan perguruan tinggi (D=1) mempunyai rata-rata lebih tinggi $3.280/bulan dibanding lulusan non perguruan tinggi.

Latihan :

1. Seorang konsultan Sumber Daya Manusia ingin mengetahui apakah ada pengaruh yang signifikan (nyata) antara gaji yang diberikan kepada salesman selama ini (Rupiah/bulan) berdasarkan jenis kelamin salesman. Data yang diperoleh adalah sebagai berikut :

Gaji	Jenis Kelamin	Pendidikan	Usia
210000	1	1	25
195000	0	1	22
185000	0	1	26
21000	1	1	23
170000	0	0	25
210000	1	0	35
200000	1	0	33
160000	0	0	36
180000	0	0	39
210000	1	0	37
225000	1	0	30
200000	1	0	33
213000	0	1	31
215000	0	1	37
245000	1	1	38

   Catatan : Jenis Kelamin → 0 = wanita
   1 = pria
   Bantulah konsultan tersebut untuk menginterpretasikan hasil analisa yang diperoleh !

2. Untuk data yang sama dengan nomor 1 (pada kolom pendidikan, 0 = SMU dan 1 = Pendidikan Tinggi), analisa apakah ada pengaruh yang nyata antara gaji yang diberikan kepada salesman selama ini (Rupiah/bulan) dengan jenis kelamin, tingkat pendidikan dan usia salesman. Interpretasikan hasil analisa yang anda peroleh !