STUDI KASUS PRAKTIKUM STATISTIKA REGRESI BAGIAN 1 – REGRESI LINIER SEDERHANA



Disusun oleh:

Chelsea Ayu Adhigiadany 21083010028 Statistika Regresi – B

Dosen Pengampu:

Trimono Pujiarto, S.Si, M.Si.

PROGRAM STUDI SAINS DATA FAKULTAS ILMU KOMPUTER UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN" JAWA TIMUR

2022

Daftar Isi

BAB 1	. PENDAHULUAN	3
1.1.	Tujuan Praktikum	3
1	1.1. Tujuan Instruksional Umum (TIU)	
1.	•	
1.2.	Permasalahan	3
BAB 2	. TINJAUAN PUSTAKA	2
2.1.	Regresi Linier Sederhana	2
2.2.	Uji Hipotesis	2
2.3.	Uji Asumsi	
2.4.	Koefisien Korelasi	5
BAB 3	. ANALISIS DAN PEMBAHASAN	7
BAB 4	. KESIMPULAN	11
DAFT	AR PUSTAKA	12
	PIRAN	
	Langkah-langkah analisis	
	Output Analisis pada SPSS	

BAB 1. PENDAHULUAN

1.1. Tujuan Praktikum

1.1.1. Tujuan Instruksional Umum (TIU)

Setelah mengikuti praktikum ini, mahasiswa dapat diharapkan melakukan pengolahan, analisis, dan membuat model regresi dari data/informasi hasil pengamatan serta dapat melakukan prediksi berdasarkan model yang dibangun dan dianalisis dengan menggunakan SPSS.

1.1.2. Tujuan Instruksional Khusus (TIK)

Setelah mengikuti praktikum ini, mahasiswa mahasiswa diharapkan mampu membangun model regresi linier sederhana dan membuat prediksi dari data pengamatan dengan menggunakan SPSS. Analisis Regresi Linear Sederhana digunakan untuk menggambarkan efek dari satu variabel independen terhadap satu variabel dependen.

1.2.Permasalahan

Diberikan data variable bebas *X* dan variable terikat *Y* sebagai berikut :

Y	X
78.5	7
74.3	1
104.3	11
87.6	11
95.9	7
109.2	11
102.7	3
72.5	1
93.1	2
115.9	21
83.8	1
113.3	11
109.4	10
74.3	1
93.1	2

Lakukan analisis regresi linier sederhana untuk memodelkan hubungan antara X dan Y. Analisis yang dilakukan meliputi :

- a. Model regresi
- b. Uji hipotesis
- c. Uji asumsi
- d. Koefisien korelasi
- e. Model akhir

BAB 2. TINJAUAN PUSTAKA

2.1.Regresi Linier Sederhana

Analisis regresi adalah suatu metode statistik yang mengamati hubungan antara variabel terikat Y dan serangkaian variabel bebas X_1, \ldots, X_p . Tujuan dari metode ini adalah untuk memprediksi nilai Y untuk nilai X yang diberikan. Model regresi linier sederhana adalah model regresi yang paling sederhana yang hanya memiliki satu variabel bebas X. Analisis regresi memiliki beberapa kegunaan, salah satunya untuk melakukan prediksi terhadap variabel terikat Y. Persamaan untuk model regresi linier sederhana adalah sebagai berikut.

$$Y = a + bX \tag{1}$$

Y adalah variabel terikat yang diramalkan, X adalah variabel bebas, a adalah intercep, yaitu nilai Y pada saat X = 0, dan b adalah slope, yaitu perubahan rata-rata Y terhadap perubahan satu unit X. Koefisien a dan b adalah koefisien regresi dimana nilai a dan b dapat dicari menggunakan persamaan berikut.

$$b = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2}$$

$$a = \frac{\sum y - b(\sum x)}{n}$$
(3)

Nilai α adalah slope, b adalah intercep dan n adalah banyaknya data yang digunakan dalam perhitungan.

2.2.Uji Hipotesis

Pengujian hipotesis dapat didasarkan dengan menggunakan dua hal, yaitu : tingkat signifikansi atau probabilitas (α) dan tingkat kepercayaan atau *confidence interval*. Didasarkan tingkat signifikansi mulai dari 0,01 sampai dengan 0,1. Yang dimaksud dengan tingkat signifikansi adalah probabilitas melakukan kesalahan tipe I, yaitu kesalahan menolak hipotesis Ketika hipotesis tersebut benar. Tingkat kepercayaan pada umumnuya ialah tingkat dimana sebesar 95% nilai sampel akan mewakili nilai populasi dimana sampel berasal. Dalam melakukan uji hipotesis terdapat dua hipotesis, yaitu H_0 (hipotesis nol) dan H_1 (hipotesis alternatif).

Contoh uji hipotesis misalnya rata-rata produktivitas pegawai sama dengan 10 ($\mu_x = 10$), maka bunyi hipotesisnya, yaitu :

- H_0 : rata-rata produktivitas pegawai sama dengan 10
- H_1 : rata-rata produktivitas pegawai tidak sama dengan 10

Hipotesis statistiknya:

- $H_0: \mu_x = 10$
- $H_1: \mu_x > 10$ untuk uji satu sisi (*one tailed*)
- $H_1: \mu_x < 10$
- $H_1: \mu_x \neq 10$ untuk uji dua sisi (*two tailed*)

Beberapa hal yang harus diperhatikan dalam uji hipotesis, yaitu :

- Untuk pengujian hipotesis, kita menggunakan data sampel
- Dalam pengujian akan menghasilkan dua kemungkinan, yaitu pengujian signifikan secara statistik jika kita menolak H_0 dan pengujian tidak signifikan secara statistik jika kita menerima H_0
- Jika kita menggunakan nilai t, maka nilai t yang semakin besar atau menjauhi 0, kita akan cenderung menolak H_0 ; sebaliknya jika nilai t semakin kecil atau mendekati 0 kita akan cenderung menerima H_0

2.3.Uji Asumsi

Uji asumsi klasik adalah persyaratan statistic yang harus dipenuhi pada analisis regresi linear berganda yang berbasis ordinary leas square (OLS). Untuk memastikan bahwa model regresi yang diperoleh merupakan model yang terbaik, dalam hal ketepatan estimasi, tidak bias, serta konsisten, maka perlu dilakukan pengujian asumsi klasik. Adapun pengujian asumsi klasik sebagai berikut:

1. Normalitas

Uji normalitas digunakan dalam model regresi untuk menguji apakah nilai residual yang dihasilkan terdistribusi secara normal atau tidak. Model regresi yang baik adalah model regresi yang memiliki nilai residual yang terdistribusi secara normal.

2. Linieritas

Uji linearitas bertujuan untuk mengetahui apakah dua variabel mempunyai hubungan yang linear atau tidak secara signifikan. Uji ini biasanya digunakan sebagai prasyarat dalam analisis korelasi atau regresi linear.

3. Homoskedastisitas

Uji ini digunakan untuk mengetahui ada atau tidaknya penyimpangan asumsi klasik heteroskedastisitas yaitu adanya ketidaksamaan varian dari residual untuk semua pengamatan pada model regresi. Prasyarat yang harus terpenuhi dalam model regresi adalah tidak adanya gejala heteroskedastisitas.

4. Non autokorelasi

Uji autokorelasi digunakan untuk mengetahui ada atau tidaknya penyimpangan asumsi klasik autokorelasi yaitu korelasi yang terjadi antara residual pada satu pengamatan dengan pengamatan lain pada model regresi. Prasyarat yang harus terpenuhi adalah tidak adanya autokorelasi dalam model regresi.

Metode pengujian yang sering digunakan adalah dengan uji Durbin-Watson (uji D-W) dengan ketentuan sebagai berikut :

- Jika d lebih kecil dari dL maka hipotesis nol ditolak, yang berarti terdapat autokorelasi positif.
- Jika d terletak antara dL dan dU maka tidak menghasilkan kesimpulan yang pasti.
- Jika d terletak antara dU dan (4-dU), maka hipotesis nol diterima, yang berarti tidak ada autokorelasi.
- Jika d terletak antara (4-dU) dan (4-dL) maka tidak menghasilkan kesimpulan yang pasti.
- Jika d lebih besar dari (4-dL) maka hipotesis nol ditolak, yang berarti terdapat autokorelasi negatif.

Nilai dU dan dL dapat diperoleh dari tabel statistik Durbin Watson yang bergantung banyaknya observasi dan banyaknya variabel yang menjelaskan.

2.4. Koefisien Korelasi

Koefisien korelasi (r) adalah sebuah nilai yang dipergunakan untuk mengukur derajat keeratan hubungan antara dua variabel. Atau koefisien yang mengukur kuat tidaknya hubungan antara variabel variabel *X* dan *Y*.

Koefisien korelasi dapat dihitung menggunakan rumus koefisien korelasi person :

$$r = \frac{n \sum xy - \sum x \sum y}{\sqrt{n \sum x^2 - (\sum x)^2} \cdot \sqrt{n \sum y^2 - \sum y^2}}$$

Nilai koefisien korelasi r berkisar antara -1 dan 1 \rightarrow (-1 \leq r \leq 1) dengan table tingkat korelasi dan kekuatan hubungan sebagai berikut :

Nilai korelasi (r)	Tingkat hubungan
0	Tidak berkolerasi
0,01-0,20	Sangat rendah
0,21-0,40	Rendah
0,41-0,60	Agak rendah
0,61-0,80	Cukup
0,81-0,99	Tinggi
1	Sangat tinggi

BAB 3. ANALISIS DAN PEMBAHASAN

3.1 Model regresi

Coefficients^a

		Unstandardize		Standardized Coefficients		0:	Collinearity	
Mo	odel	ь	Std. Error	Beta	τ	Sig.	Tolerance	VIF
1	(Constant)	81.199	4.162		19.511	.000		
	X	1.899	.477	.742	3.985	.002	1.000	1.000

a. Dependent Variable: Y

Berdasarkan tabel di atas, dapat diketahui bahwa:

$$\beta_0 = 81,199$$

$$\beta_1 = 1,899$$

Maka dapat disimpulkan bahwa persamaan regresinya adalah Y = 81,199 + 1,899X

3.2 Uji hipotesis

Pada uji hipotesis dilakukan 2 macam uji yaitu untuk menguji kecocokan model dan juga menguji pengaruh variabel *X* dan *Y*.

3.2.1. Uji f (uji kecocokan model)

a. Menentukan hipotesis

$$H_0$$
: $\beta_0 = \beta_1 = 0$ (model regresi tidak sesuai)

$$H_1: \beta_1 \neq 0$$
 (model regresi sesuai)

b. Taraf signifikansi

$$\alpha = 5\% = 0.05$$

c. Uji statistik

ANOVA

	Model		Sum of Squares	df	Mean Square	F	Sig.
	1	Regression	1721.527	1	1721.527	15.882	.002 ^b
ı		Residual	1409.169	13	108.398		
I		Total	3130.696	14			

a. Dependent Variable: Y

Berdasarkan tabel anova dapat diketahui bahwa F = 15,882 dengan Sig = 0,002

d. Daerah kritis

Tolak H_0 jika nilai Sig $< \alpha$

e. Keputusan

Tolak H_0 karena nilai $Sig < \alpha \ (0.002 < 0.05)$

f. Kesimpulan

 H_0 ditolak karena nilai $Sig < \alpha$, yang berarti model regresi sesuai

3.2.2. Uji t (uji signifikansi)

a. Menentukan hipotesis

$$H_0: \beta_i = 0$$
 (variabel *X* tidak berpengaruh terhadap variabel *Y*)

$$H_1: \beta_i \neq 0$$
 (variabel *X* berpengaruh terhadap variabel *Y*)

b. Taraf signifikansi

$$\alpha = 5\% = 0.05$$

c. Uji statistik

Coefficients

		Unstandardize	d Coefficients Std. Error	Standardized Coefficients Beta		Sig.	Collinearity Tolerance	Statistics VIF
Model		ם	ota. Elloi	Dota	ι	oly.	Tolerance	VII
1	(Constant)	81.199	4.162		19.511	.000		
l	X	1.899	.477	.742	3.985	.002	1.000	1.000

a. Dependent Variable: \

Berdasarkan tabel coefficient, diketahui bahwa t = 3,985 dengan Sig = 0,002

d. Daerah kritis

b. Predictors: (Constant), X

Tolak H_0 jika nilai Sig $< \alpha$

e. Keputusan

Tolak H_0 karena nilai $Sig < \alpha (0.002 < 0.05)$

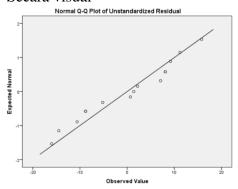
f. Kesimpulan

 H_0 ditolak karena nilai $Sig < \alpha$, yang berarti variabel X berpengaruh terhadap variabel Y

3.3 Uji asumsi

3.3.1. Normalitas

a. Secara visual



Pada Normal Q-Q Plot of Unstandardized Residual dapat dilihat bahwa plot-plot mengikuti garis lurus, sehingga dapat disimpulkan bahwa residual berdistribusi normal. Maka asumsi normalitas terpenuhi secara visual

b. Secara formal

• Menentukan hipotesis

 H_0 : residual berdistribusi normal

 H_1 : residual tidak berdistribusi normal

• Taraf signifikansi : $\alpha = 5\% = 0.05$

Uji statistik

Tests of Normality

	Kolm	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.	
Unstandardized Residual	.161	15	.200*	.942	15	.402	

^{*.} This is a lower bound of the true significance.

$$\sup |F0(X)-F(X)| = 0.942$$

$$Sig = 0.402$$

Daerah kritis

Tolak H_0 jika nilai Sig $< \alpha$

Keputusan

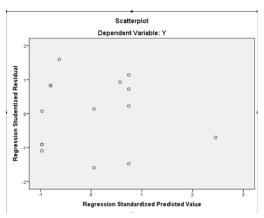
 H_0 diterima karena nilai $Sig > \alpha \ (0.402 < 0.05)$

Kesimpulan

 H_0 diterima karena nilai $Sig > \alpha$, yang berarti residual berdistribusi normal

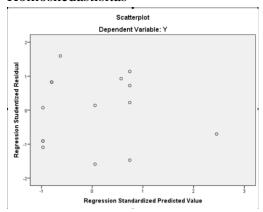
3.3.2. Linieritas

a. Lilliefors Significance Correction



Berdasarkan grafik zresid by zpred scatterplot dapat dilihat bahwa sebaran data acak atau tidak membentuk pola tertentu maka dapat disimpulkan bahwa uji linieritas terpenuhi.

3.3.3. Homoskedastisitas



Berdasarkan grafik sresid by zpred scatterplot dapat dilihat bahwa asumsi Heteroskedastisitas terpenuhi jika residual menyebar secara acak dan tidak membentuk pola.

3.3.4. Non autokorelasi

a. Hipotesis

 H_0 : tidak ada autokorelasi

 H_1 : ada autokorelasi

b. Taraf signifikansi : $\alpha = 5\% = 0.05$

c. Uji statistik

Model Summary^b

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin- Watson
1	.742ª	.550	.515	10.4114	1.983

a. Predictors: (Constant), X

Berdasarkan tabel Model Summary dapat diketahui : DW = 1,983

Dari tabel Durbin Watson $\alpha = 5\%$ dengan n = 15 dan k = 1 dapat diketahui :

	k=	1	k=:	2	k=	3	i
n	dL	dU	dL	dU	dL	dU	dl.
6	0.6102	1.4002					
7	0.6996	1.3564	0.4672	1.8964			
8	0.7629	1.3324	0.5591	1.7771	0.3674	2.2866	
9	0.8243	1.3199	0.6291	1.6993	0.4548	2.1282	0.2957
10	0.8791	1.3197	0.6972	1.6413	0.5253	2.0163	0.3760
11	0.9273	1.3241	0.7580	1.6044	0.5948	1.9280	0.4441
12	0.9708	1.3314	0.8122	1.5794	0.6577	1.8640	0.5120
13	1.0097	1.3404	0.8612	1.5621	0.7147	1.8159	0.5745
14	1.0450	1.3503	0.9054	1.5507	0.7667	1.7788	0.6321
15	1.0770	1.3605	0.9455	1.5432	0.8140	1.7501	0.6852

$$dL = 1,0770$$

$$dU = 1,3605$$

b. Dependent Variable: Y

d. Daerah krisis

0 < DW < dL : Menolak H0, mengalami autokorelasi positif

dL < DW < dU : Ragu-ragu

dU < DW < 4-dU : Menerima H0, tidak ada autokorelasi

4-dU < DW < 4-dL: Ragu-ragu

4-dL < DW < 4 : Menolak H0, mengalami autokorelasi negatif

e. Keputusan

 H_0 diterima karena dU (1,3605) < DW (1,983) < 4-dU (4-1,3605=2,6395)

f. Kesimpulan

 H_0 diterima karena dU < DW < 4-dU, tidak ada autokorelasi

3.4 Koefisien korelasi

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin- Watson
1	.742ª	.550	.515	10.4114	1.983

a. Predictors: (Constant), X

b. Dependent Variable: Y

Melalui tabel Model Summary didapatkan nilai koefisien korelasi : R = 0,742 Nilai koefisien korelasi antara variabel X dan Y adalah 0,742, yang berarti hubungan korelasi X dan Y sangat kuat.

3.5 Model akhir

Berdasarkan uji F, model regresi yang dibuat cocok digunakan untuk analisis lebih lanjut dan berdasarkan uji t, koefisien parameter regresi X yaitu β 1 berpengaruh signifikan terhadap Y. Kesimpulannya model akhir sama dengan model awal yang telah dibuat, yaitu : Y = 81,199 + 1,899X

BAB 4. KESIMPULAN

Dari permasalahan yang terdapat pada bab 1, ditemukan hasil dari model regresi, uji hipotesis, uji asumsi, koefisien korelasi, dan juga model akhir dari sebuah tabel tersebut dengan bantuan SPSS dalam perhitungan adalah sebagai berikut:

- 1 Model regresi : Y = 81,199 + 1,899X
- 2 Dilakukan 2 uji pada tahapan uji hipotesis yaitu uji f dan uji t dengan hasil menolak H_0 karena nilai $Sig < \alpha$, yang berarti model regresi sesuai dan variabel X berpengaruh terhadap variabel Y
- Pada uji asumsi dilakukan 4 uji yaitu normalitas, linieritas, homoskedastisitas, dan juga non autokorelasi dengan hasil H_0 diterima karena nilai $Sig > \alpha$, yang berarti residual berdistribusi normal untuk uji normalitas, sebaran data yang acak menunjukkan bahwa uji linieritas dan homoskedastisitas terpenuhi. Sedangkan pada uji non autokorelasi diperoleh hasil H_0 diterima karena dU < DW < 4-dU, tidak ada autokorelasi
- 4 Koefisien korelasi antara variabel *X* dan *Y* adalah 0,742, yang berarti hubungan korelasi *X* dan *Y* sangat kuat.
- 5 Model akhir sama dengan model awal yang telah dibuat, yaitu : Y = 81,199 + 1,899X

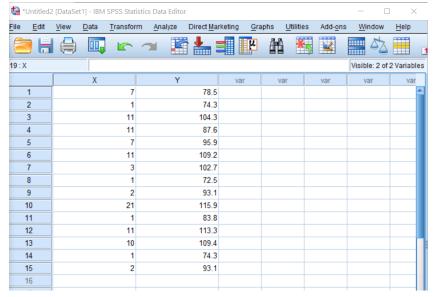
DAFTAR PUSTAKA

LAMPIRAN

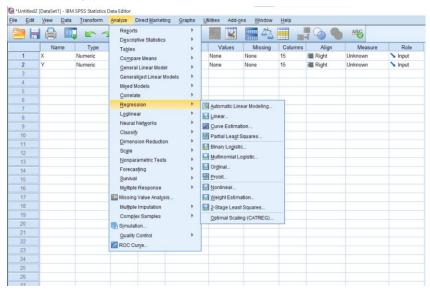
- 1. Langkah-langkah analisis
 - a. Buat file data. Sebelum memasukan data pada Data View, definisikan terlebih dahulu variabel respon Y dan variabel X pada Variabel View



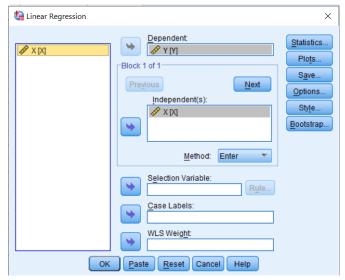
b. Isikan data yang tersedia pada sel-sel yang sesuai pada Data View.



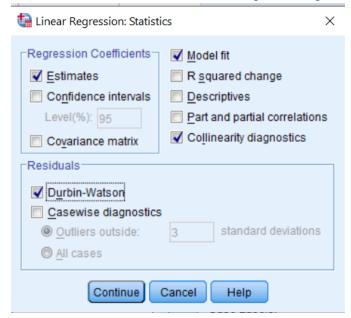
c. Klik tombol Analyze – Regresion – Linear pada menu utama SPSS pada Data View



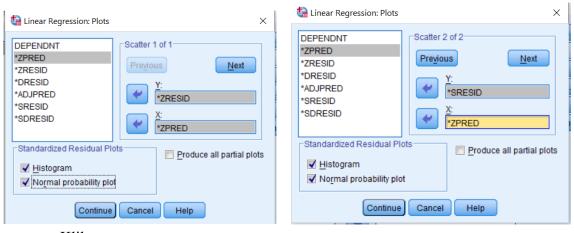
d. Isilah kotak Dependent dengan variabel Y (respon) dan kotak Independent dengan variabel X (predictor)



e. Klik statistics, sesuaikan dengan settingan di bawah



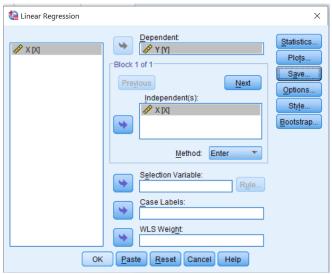
f. Klik plot, sesuaikan dengan setting berikut, tekan next untuk beralih ke scatter selanjutnya



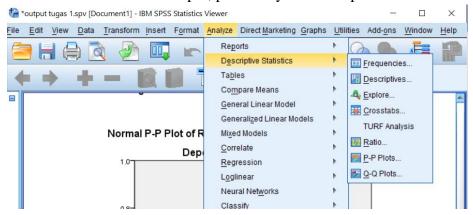
g. Klik save



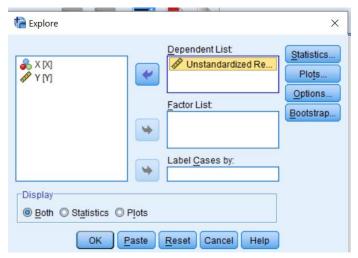
h. Klik continue kemudian klik ok ketika tampilan sudah berubah seperti dibawah



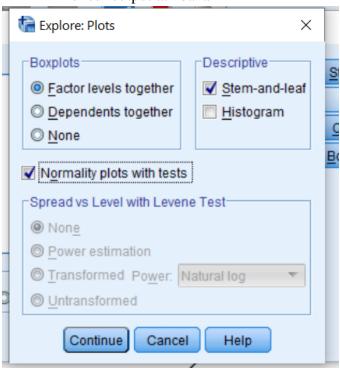
i. Setelah muncul output, pilih analyze -> secriptive statistics -> explore



j. Pilih kolom tambahan yang muncul ke dalam dependent list



k. Sesuaikan dengan setting berikut dan klik continue, maka pada bagian output akan muncul output tambahan



2. Output Analisis pada SPSS

NEW FILE.

DATASET NAME DataSetl WINDOW=FRONT.

REGRESSION

/MISSING LISTWISE

/STATISTICS COEFF OUTS R ANOVA COLLIN TOL

/CRITERIA=PIN(.05) POUT(.10)

/NOORIGIN

/DEPENDENT Y

/METHOD=ENTER X

/SCATTERFLOT=(*ZRESID ,*ZFRED) (*SRESID ,*ZFRED)

/RESIDUALS DURBIN HISTOGRAM(ZRESID) NORMPROB(ZRESID)

Regression

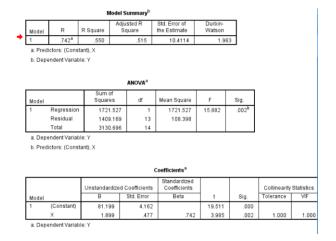
[DataSet1]

Variables Entered/Removed

Model	Variables Entered	Variables Removed	Method
1	Xp		Enter

a. Dependent Variable: Y

b. All requested variables entered.



Collinearity Diagnostics^a

			Condition	Variance Pr	oportions
Model	Dimension	Eigenvalue	Index	(Constant)	X
1	1	1.763	1.000	.12	.12
l	2	.237	2.730	.88	.88

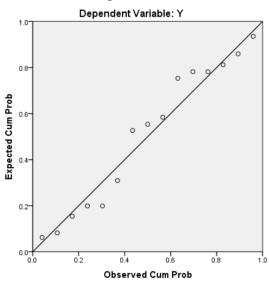
a. Dependent Variable: Y

Residuals Statistics^a

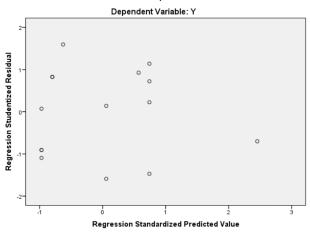
	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	83.098	121.080	93.860	11.0890	15
Std. Predicted Value	970	2.455	.000	1.000	15
Standard Error of Predicted Value	2.693	7.340	3.656	1.081	15
Adjusted Predicted Value	82.990	126.200	94.199	11.9209	15
Residual	-15.9930	15.8033	.0000	10.0327	15
Std. Residual	-1.536	1.518	.000	.964	15
Stud. Residual	-1.590	1.595	013	1.024	15
Deleted Residual	-17.1397	17.4590	3390	11.4030	15
Stud. Deleted Residual	-1.702	1.709	020	1.056	15
Mahal. Distance	.003	6.026	.933	1.441	15
Cook's Distance	.000	.243	.070	.063	15
Centered Leverage Value	.000	.430	.067	.103	15

a. Dependent Variable: Y

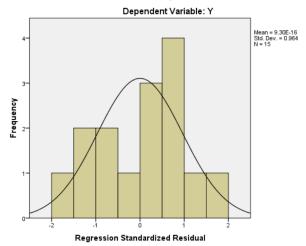
Normal P-P Plot of Regression Standardized Residual

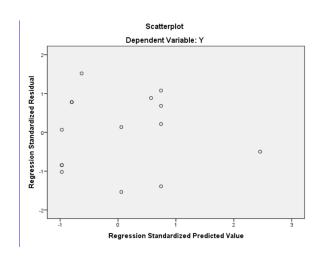


Scatterplot



Histogram





Explore

[DataSetl] D:\semester 3\tugas 1\tugas 1.sav

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Unstandardized Residual	15	100.0%	0	0.0%	15	100.0%

Descriptive

			Statistic	Std. Error
Unstandardized Residual	Mean		.0000000	2.59043014
	95% Confidence Interval for Mean	Lower Bound	-5.5559201	
		Upper Bound	5.5559201	
	5% Trimmed Mean	.0105385		
	Median	1.4069693		
	Variance	100.655		
	Std. Deviation	10.03269277		
	Minimum	-15.99303		
	Maximum	15.80334		
	Range	31.79637		
	Interquartile Range	16.90091		
	Skewness	184	.580	
	Kurtosis	-1.259	1.121	

Tests of Normality

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Unstandardized Residual	.161	15	.200*	.942	15	.402

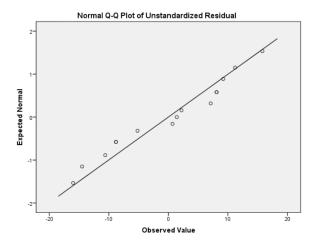
- *. This is a lower bound of the true significance.
- a. Lilliefors Significance Correction

Unstandardized Residual

Unstandardized Residual Stem-and-Leaf Plot

Frequency	Stem &	Leaf
3,00	-1 .	045
3,00	-0 .	588
7,00	0.	0127889
2,00	1 .	15

Stem width: 10,00000 Each leaf: 1 case(s)



Observed Value

