



STATISTIKA PENELITIAN

REGRESI LINIER

Analisis regresi merupakan salah satu teknik analisis yang paling populer di bidang penelitian sekarang ini. Analisis regresi digunakan untuk menjelaskan suatu variabel respon (variabel terikat / dependent / output) menggunakan satu atau lebih variabel input (variabel bebas, independent variable / eksogen). Jika variabel bebas terdiri dari 1 maka regresi sederhana yang digunakan, dan jika variabel input lebih dari 1, maka regresi ganda yang digunakan.

Persamaan regresi sederhana dinotasikan sebagai berikut : $Y = a + b X$ (Y adalah variabel respon, a = konstanta, b = parameter regresi)

Data : data yang digunakan untuk analisis regresi linier sederhana adalah minimal interval

Asumsi : asumsi yang harus dipenuhi dalam regresi sederhana adalah asumsi normalitas



Contoh Kasus

Seorang manajer penjualan salah satu agen sepeda motor ingin mengetahui pengaruh biaya promosi dengan jumlah unit motor yang terjual dalam beberapa tahun terakhir. Ia menggunakan data penjualan dan biaya promosi 3 tahun terakhir untuk meramalkan penjualan berdasarkan biaya promosi yang dikeluarkan setiap bulannya.

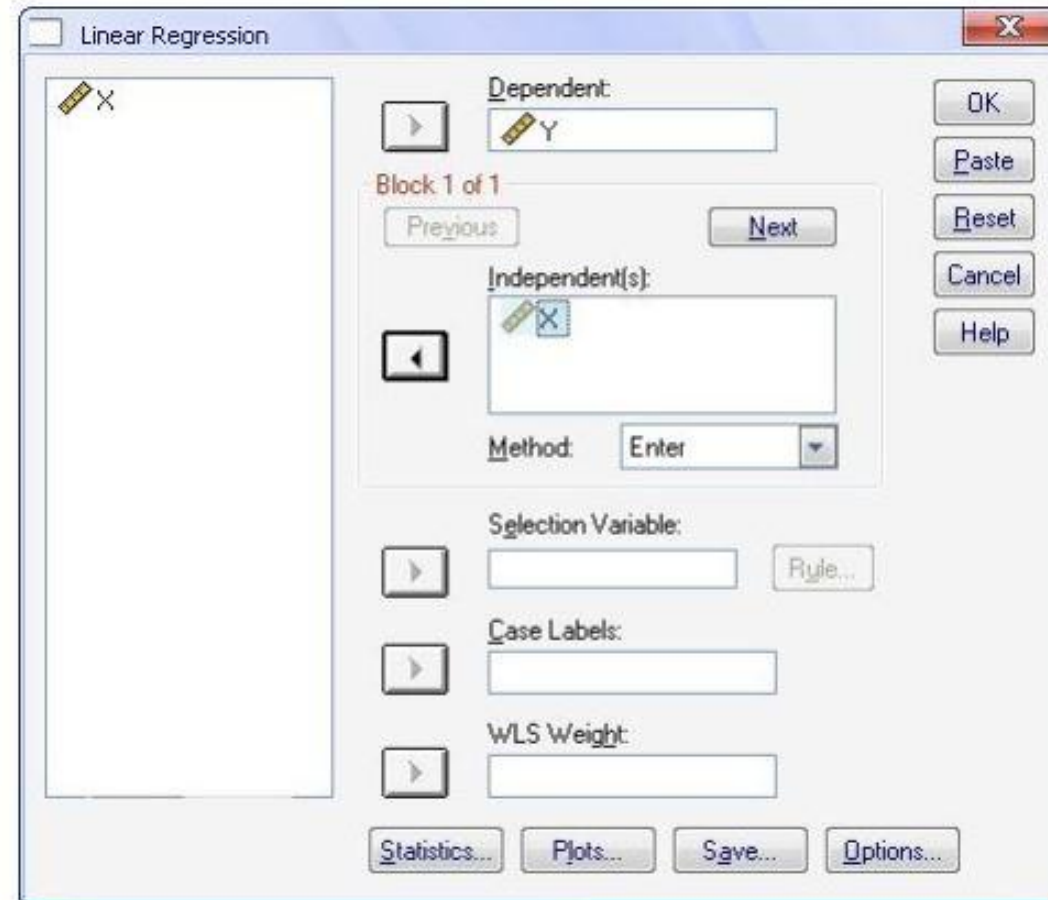
Data :

No	Promosi	Unit Terjual
1	70000	8000
2	68000	7800
3	84000	9800
4	85000	7800
5	68000	7900
6	76000	8100
7	70000	7800
8	71000	8000
9	70000	7800
10	67000	7600
11	72000	8600
12	77000	8100
13	68000	7600
14	71000	7500
15	77000	8700
16	82000	8300

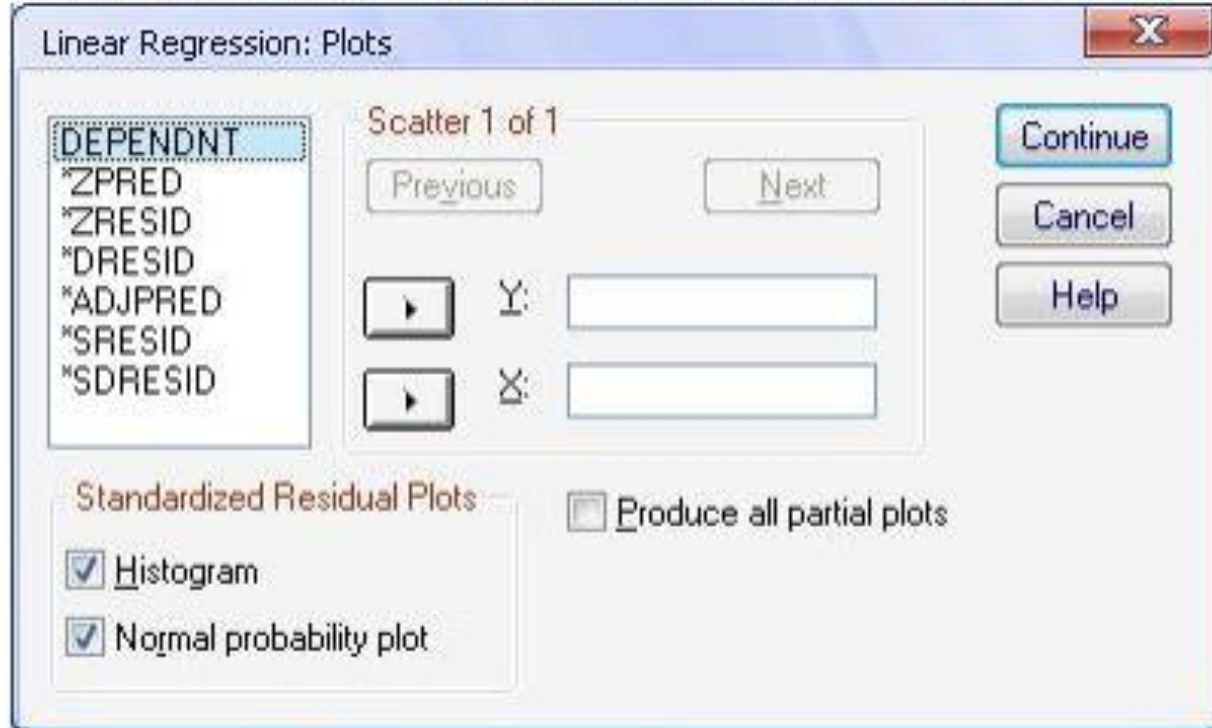
17	68000	7400
18	68000	8000
19	68000	7200
20	68000	7800
21	69000	7800
22	61000	7400
23	64000	7700
24	81000	9000
25	68000	7600
26	68000	7800
27	68000	7700
28	68000	7800
29	68000	7500
30	70000	7300
31	77000	7900
32	79000	7000
33	36000	4400
34	83000	8000
35	82000	7400
36	68000	7300

Data Penjualan dalam Ribuan
Data Unit Terjual (satuan)

Klik Analyze – Regression - Linier – lalu setting data seperti tampilan di bawah ini
Masukkan variabel X (biaya promosi) ke box independent, dan Y (unit terjual) ke box
“dependent”



Klik Plots, lalu Tick pada pilihan "Histogram" dan "Normal Probability Plot"



Linear Regression: Plots

Scatter 1 of 1

DEPENDNT
*ZPRED
*ZRESID
*DRESID
*ADJPRED
*SRESID
*SDRESID

Previous Next

Y:

X:

Standardized Residual Plots

☒ Histogram

☒ Normal probability plot

☐ Produce all partial plots

Continue Cancel Help

Klik Continue, dan OK

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
1	.753 ^a	.568	.555	520.55480

a. Predictors: (Constant), X

b. Dependent Variable: Y

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	12092327	1	12092327.39	44.625	.000 ^a
	Residual	9213228	34	270977.299		
	Total	21305556	35			

a. Predictors: (Constant), X

b. Dependent Variable: Y

Coefficients^a

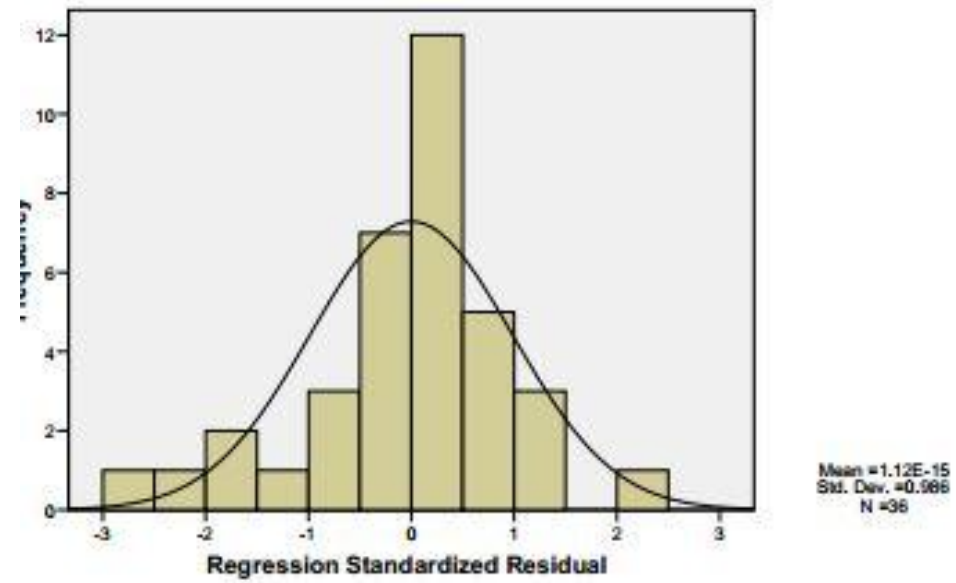
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	2881.296	735.625		3.917	.000
	X	.069	.010	.753	6.680	.000

a. Dependent Variable: Y

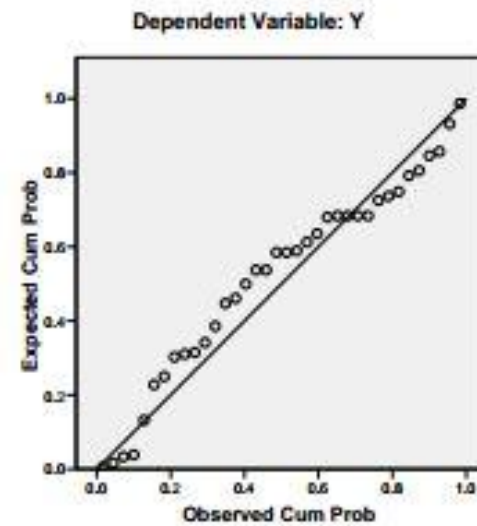
Residuals Statistics^a

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	5353.6338	8718.7607	7761.1111	587.78829	36
Residual	-1306.70	1149.916	.00000	513.06441	36
Std. Predicted Value	-4.096	1.629	.000	1.000	36
Std. Residual	-2.510	2.209	.000	.986	36

a. Dependent Variable: Y



Normal P-P Plot of Regression Standardized Residual



Sebelum memberikan interpretasi pada hasil regresi, dilakukan pengujian asumsi normalitas sebagai syarat regresi. Apabila berdistribusi normal maka analisis parametrik seperti analisis regresi dapat dilanjutkan, sebaliknya apabila tidak berdistribusi normal maka digunakan statistik non parametrik untuk menguji hipotesis. Pengujian normalitas ini menggunakan diagram histogram dan grafik p p-plot untuk memprediksi apakah data berdistribusi normal atau tidak.

Berdasarkan hasil uji di atas terlihat bahwa menyebar merata ke kanan dan kekiri bagian kurva normal, dan membentuk kurva normal, sehingga dapat disimpulkan residual memenuhi asumsi normalitas. Hasil pengujian dengan memperhatikan grafik p p-plot juga menunjukkan kesimpulan serupa dengan histogram. Dari tampilan di atas terlihat bahwa data data menyebar di sekitar garis diagonal dan mengikuti arah garis diagonal, sehingga dapat dinyatakan normal.

Korelasi antara biaya promosi dengan penjualan (unit terjual) adalah sebesar 0.753, dengan koefisien determinasi 0.568. Dengan demikian dapat dinyatakan bahwa variasi penjualan mampu dijelaskan oleh biaya promosi sebesar 56.80%, dan sisanya dipengaruhi faktor lain selain biaya promosi

Coefficients^a

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
1 (Constant)	2881.296	735.625		3.917	.000
X	.069	.010	.753	6.680	.000

a. Dependent Variable: Y

Persamaan regresi : $\text{Penjualan} = 2881.296 + 0.069 (\text{biaya promosi})$, Persamaan regresi tersebut mempunyai makna sebagai berikut: Konstanta sebesar 2881.296 berarti bahwa tanpa adanya biaya yang dikeluarkan untuk promosi, maka penjualan sepeda motor adalah sebesar 2,881 satuan. Jika variabel biaya promosi naik (satu juta) maka akan menyebabkan kenaikan (karena tanda positif) sebesar 0.069 pada penjualan sepeda motor.

Pengujian Hipotesis

Ho : Tidak ada pengaruh X terhadap Y

Ha : Ada pengaruh positif dan signifikan X terhadap Y

Pengambilan keputusan (berdasarkan probabilitas) :

Jika probabilitas $> 0,05$ maka Ho diterima, sedangkan jika probabilitas $< 0,05$ maka Ho ditolak. Dari hasil uji signifikansi terlihat bahwa nilai probabilitas adalah sebesar 0,00 ($< 0,01$) sehingga Ho ditolak. Artinya, pengaruh biaya promosi terhadap penjualan signifikan sehingga hipotesis alternatif (Ha) diterima.

Hasil uji melalui probabilitas ini juga relevan dengan pengujian melalui statistik t. Nilai t hitung adalah sebesar 6.680, sementara t tabel diperoleh dari $dk = n - 2 = 36 - 2 = 34$ (dalam uji ini, diambil $dk = 30$) dan taraf signifikansi 1% adalah sebesar 2.704 . Karena t hitung $>$ t tabel ($6.680 > 2.704$) maka Ho ditolak, artinya pengaruh X terhadap Y adalah positif dan terbukti signifikan berdasarkan pengujian statistik.

Analisis Regresi dapat digunakan untuk menemukan persamaan regresi yang menunjukkan hubungan antara variabel dependen (variabel respon) dengan satu atau beberapa variabel independen (variabel prediktor). Jika variabel dependen dihubungkan dengan satu variabel independen saja, maka variabel regresi yang dihasilkan adalah regresi linier sederhana, dan jika variabel independennya lebih dari satu maka yang dihasilkan adalah persamaan regresi linier berganda (*multiple linier regression*). Nilai koefisien regresi 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.



Sebesar berapa pengaruh variabel independen 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. Harga koefisien determinasi akan berharga 1 jika seluruh observasi jatuh pada garis regresi, dan akan berharga 0 jika tidak ada hubungan linear antara variabel dependen dengan variabel independen.



Regresi linear adalah hubungan secara linear antara variabel dependen dengan variabel independen yang digunakan untuk memprediksi atau meramalkan suatu nilai variabel dependen berdasarkan variabel independen.

Spesifikasi minimum untuk prosedur ini adalah:

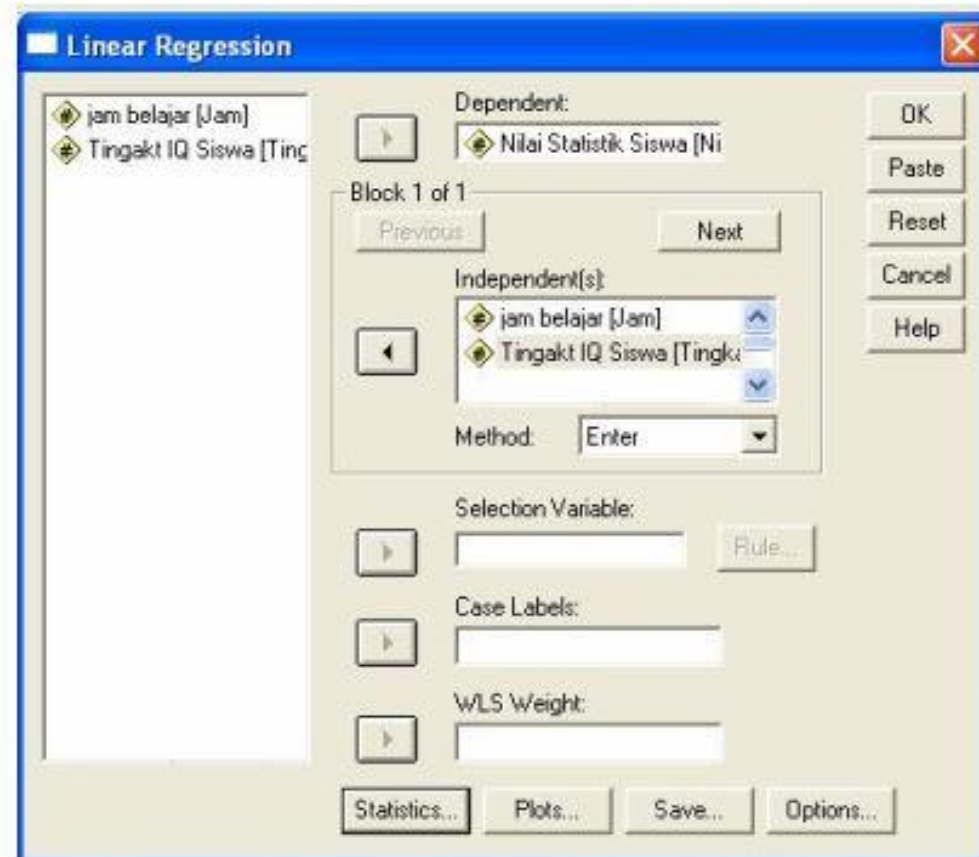
- Satu variabel dependen
- Satu atau beberapa variabel independen

Gunakan data pada **analisis korelasi** untuk menentukan persamaan regresinya, dan kemudian gunakan untuk memprediksikan nilai ujian statistik jika diketahui lama jam belajar dan tingkat IQ-nya.

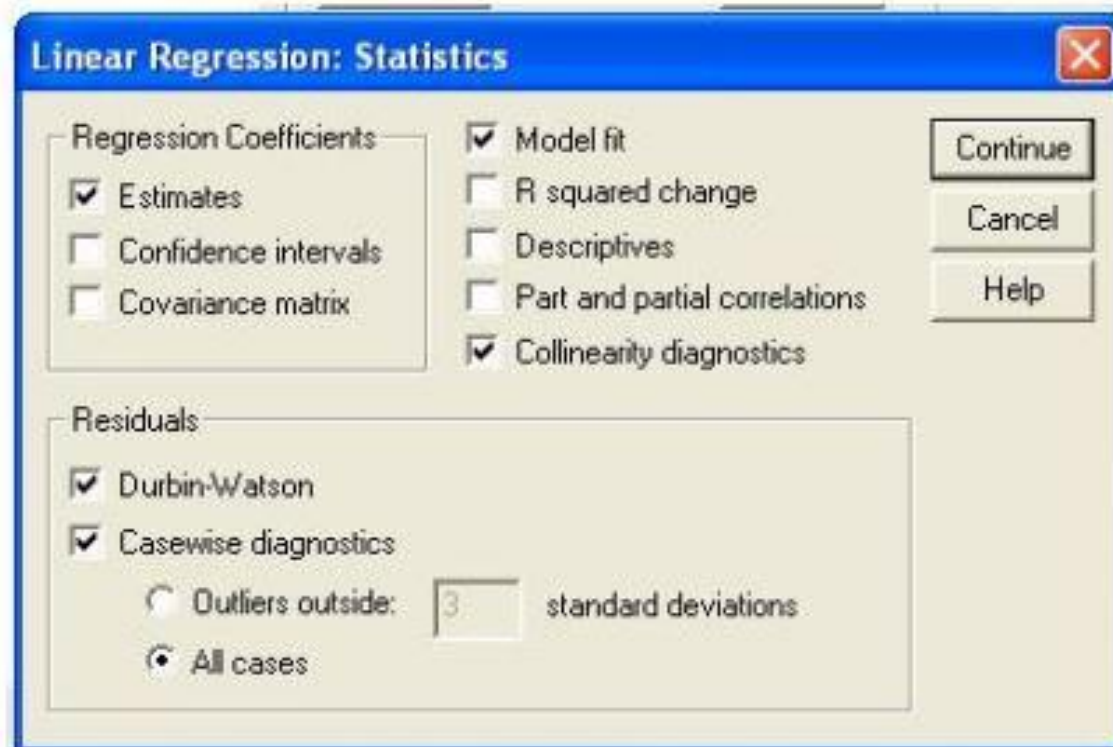
Jam Belajar (jam)	1	0,5	1	1,5	0,5	2	2	1	1,5	0,5
Tingkat IQ	115	120	112	107	119	100	98	110	105	124
Nilai Statistik	70	80	75	65	85	60	60	70	80	95

Analyze → Regression → Linear

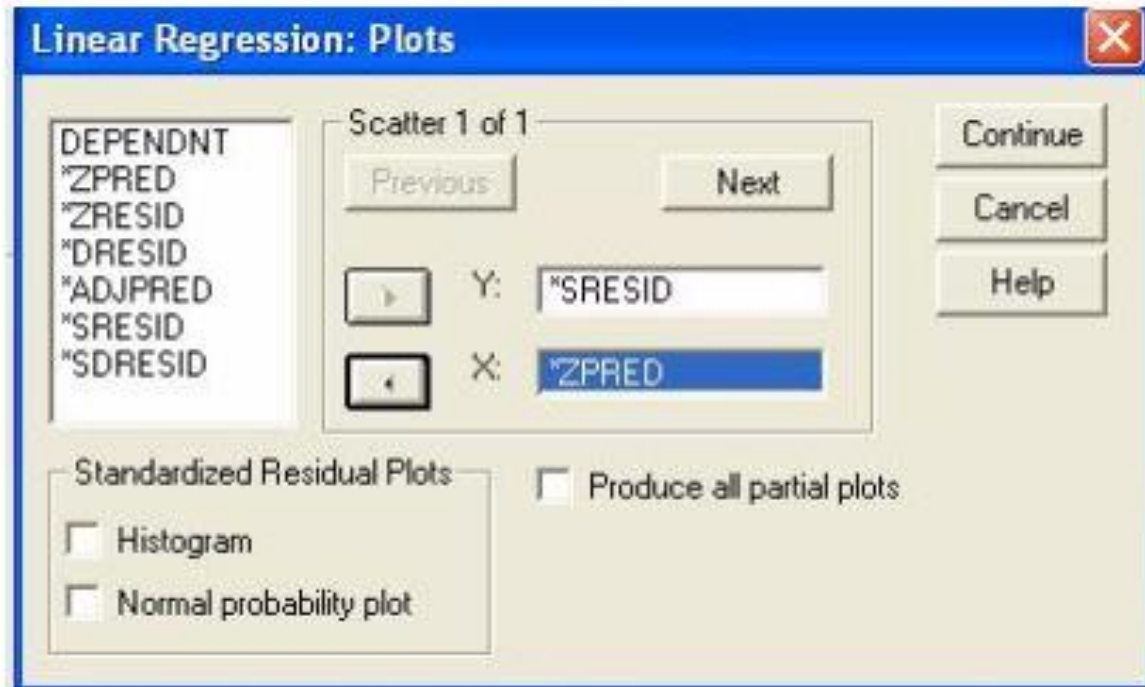
Maka akan ditampilkan kotak dialog **Linear regression**.



1. Masukkan variabel Nilai Statistik ke kotak *Dependent*, sedangkan variabel jam belajar dan tingkat IQ ke kotak *Independent (s)*.
2. Klik tab **Statistics**, maka akan ditampilkan kotak dialog **Linear regression Statistics**



3. Berilah centang pada **colinearity diagnostics**, **Durbin-Watson** dan **Casewise diagnostics** kemudian pilih **All cases**. Setelah itu, klik **continue**.
4. Klik tab **Plots**, maka akan ditampilkan kotak dialog **Linear Regression Plots**





5. Kliklah ***SRESID** (*Studentized Residual*), kemudian masukkan ke kotak **Y**. Selanjutnya, kliklah ***ZPRED** (*Standardized Predicted Value*) kemudian masukkan ke kotak **X**. Setelah itu, klik **continue**
6. Kliklah OK, maka hasil output akan ditampilkan sebagai berikut:

Variables Entered/Removed^b

Model	Variables Entered	Variables Removed	Method
1	tingkat_iq, jam_bel	.	Enter

a. All requested variables entered.

b. Dependent Variable: nilai_statistik

Model Summary^b

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	,844 ^a	,713	,631	6,838	1,228

a. Predictors: (Constant), Tingkat IQ, Jam Belajar

b. Dependent Variable: Nilai Statistik

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	812.703	2	406.351	8.691	.013 ^a
	Residual	327.297	7	46.757		
	Total	1140.000	9			

a. Predictors: (Constant), tingkat_iq, jam_bel

b. Dependent Variable: nilai_statistik

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	-37.338	149.868		-.249	.810		
	Jam Belajar	-1.254	17.505	-.065	-.072	.945	.050	19.824
	Tingkat IQ Siswa	1.016	1.173	.781	.866	.415	.050	19.824

a. Dependent Variable: Nilai Statistik

Collinearity Diagnostics^a

Model	Dimension	Eigenvalue	Condition Index	Variance Proportions		
				(Constant)	Jam Belajar	Tingkat IQ Siswa
1	1	2.846	1.000	.00	.00	.00
	2	.153	4.308	.00	.04	.00
	3	.000	155.977	1.00	.96	1.00

a. Dependent Variable: Nilai Statistik

Casewise Diagnostics^a

Case Number	Std. Residual	Nilai Statistik	Predicted Value	Residual
1	-1.207	70	78.25	-8.252
2	-.579	80	83.96	-3.960
3	-.030	75	75.20	-.204
4	-.658	65	69.50	-4.497
5	.301	85	82.94	2.056
6	-.257	60	61.76	-1.757
7	.040	60	59.73	.275
8	-.464	70	73.17	-3.172
9	1.833	80	67.46	12.535
10	1.020	95	88.02	6.976

a. Dependent Variable: Nilai Statistik

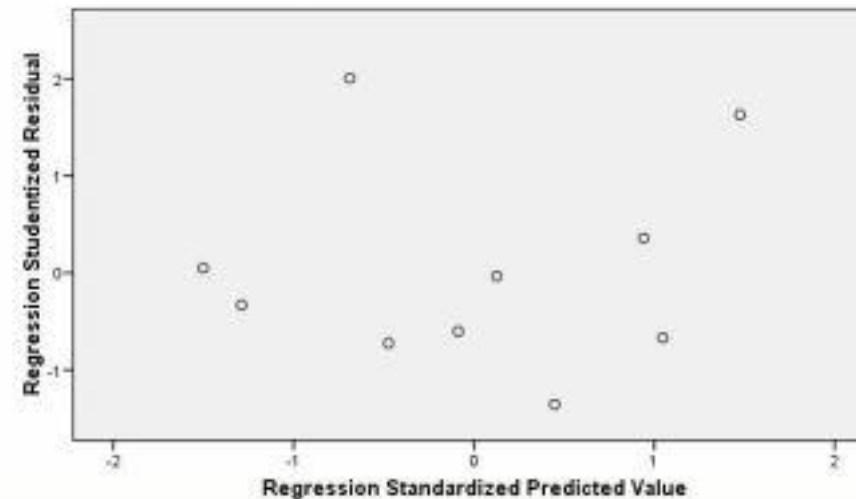
Residuals Statistics^a

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	59.73	88.02	74.00	9.503	10
Std. Predicted Value	-1.502	1.476	.000	1.000	10
Standard Error of Predicted Value	2.635	5.338	3.655	.864	10
Adjusted Predicted Value	59.58	85.25	73.33	8.601	10
Residual	-8.252	12.535	.000	6.030	10
Std. Residual	-1.207	1.833	.000	.882	10
Stud. Residual	-1.353	2.006	.034	1.058	10
Deleted Residual	-10.376	17.859	.669	9.132	10
Stud. Deleted Residual	-1.458	2.848	.145	1.297	10
Mahal. Distance	.437	4.584	1.800	1.297	10
Cook's Distance	.000	1.386	.202	.424	10
Centered Leverage Value	.049	.509	.200	.144	10

a. Dependent Variable: Nilai Statistik

Scatterplot

Dependent Variable: Nilai Statistik



Dari output yang kedua pada kolom variabel Entered menunjukkan tidak ada variabel yang dikeluarkan (removed), atau dengan kata lain kedua variabel bebas dimasukkan dalam perhitungan regresi.

R square = 0,713 . Hal ini berarti 71,3% variabel dependent **Nilai Ujian Statistik** dijelaskan oleh variabel independent **Lama jam belajar** dan **Tingkat IQ**.

- **UJI ANOVA**

Hipotesis:

H_0 : Tidak ada pengaruh antara jam belajar dan tingkat IQ terhadap nilai statistic

H_1 : Ada pengaruh antara jam belajar dan tingkat IQ terhadap nilai statistic

Dari uji ANOVA didapatkan $F_{hitung} = 8,691$ dengan tingkat signifikansi 0,013 karena probabilitas $< 0,05$ maka model regresi bisa digunakan untuk memprediksi nilai ujian statistik.

- **UJI T** (untuk menguji signifikansi koefisien-koefisien dari variabel-variabel independen)

Syarat Persamaan Regresi :

Regresi Linier : $Y = B_0 + B_1X_1$

Regresi Berganda : $Y = B_0 + B_1X_1 + B_2X_2 + \dots + B_nX_n$

Dari hasil perhitungan diatas dapat dibuat (sementara) persamaan regresi estimasi:

$$Y = -37,338 + 1,016 X_1 - 1,254 X_2$$

Dimana :

Y = Nilai ujian statistik ; X_1 = Tingkat IQ ; X_2 = Lama jam belajar

Hipotesis

- H_0 : Koefisien regresi tidak signifikan
- H_1 : Koefisien regresi signifikan

Misalkan bentuk umum persamaan regresi: $Y = B_0 + B_1X_1 + B_2X_2 + B_3X_3$

Pengambilan Keputusan

a. Berdasarkan perbandingan t_{hitung} dengan t_{tabel}

Syarat :

- H_0 diterima : Jika t_{hitung} berada diantara nilai $- t_{tabel}$ dan $+ t_{tabel}$.
- H_0 ditolak : Jika t_{hitung} tidak berada diantara nilai $- t_{tabel}$ dan $+ t_{tabel}$.

Nilai t_{hitung} masing-masing koefisien regresi berturut-turut:

$t_1 = 0,866$ (t_{hitung} untuk variabel independen Tingkat IQ)

$t_2 = -0,072$ (t_{hitung} untuk variabel independen Lama jam belajar)

Dari tabel, dengan Tingkat signifikansi (α) 5% dengan df (derajat kebebasan = 7, maka didapatkan $t_{tabel} = \pm 2,26$. Karena nilai t_1 dan t_2 berada diantara t_{tabel} maka H_0 diterima. Dari persamaan tersebut menunjukkan setiap penambahan 1 tingkat IQ akan meningkatkan nilai ujian sebesar 1,016 dan setiap penambahan -1 lama jam belajar mengurangi nilai ujian statistik sebesar 1,254.

Berdasarkan probabilitas

Syarat :

- Jika probabilitas $> 0,05$, maka H_0 diterima
- Jika probabilitas $< 0,05$, maka H_0 ditolak

Karena nilai probabilitas untuk $t_1 = 0,415 > 0,05$ maka H_0 diterima dan probabilitas $t_2 = 0,945 > 0,05$ maka H_0 diterima, dengan kesimpulan yang sama dengan perbandingan diatas.



REFLEKSI



- 1. Informasi penting hari ini**
- 2. Manfaat penting dari informasi penting hari ini**
- 3. Tindak lanjut yang dapat saudara lakukan**



Thank you!
Any questions?