



REGRESI LINIER

MOOC Unair 2022

Indah Lutfiya

CAPAIAN PEMBELAJARAN

Mahasiswa/ masyarakat umum mampu mengidentifikasi pemilihan uji statistik Regresi Linier menggunakan software SPSS dengan tepat serta menyajikan hasil informasi berdasarkan tutorial uji regresi Linier dengan benar

TUJUAN

Untuk mengetahui pengaruh variable bebas (independent variable)/ x terhadap variable terikat (y) (dependent variable).

Variabel bebas = variable yang mempengaruhi

Variabel terikat = variable yang dipengaruhi

$X \rightarrow y$ (uji regresi), tanda a

$X - y$ (uji korelasi)

Membuat pemodelan

JENIS

1. Regresi linier Sederhana (dua variable saja; 1 independent dan 1 dependent)
2. Regresi Linier Ganda (independent > 1 , dependent tetap 1)

ASUMSI (PRASYARAT) → TERUTAMA UNTUK REGRESI LINIER GANDA

1. **Data y (dependent) berskala minimal interval/ data berupa numerik**
2. **Data x berskala minimal nominal, tetapi umumnya interval-rasio**
3. Adanya linieritas yaitu pola hubungan variable dependent dan independent berbentuk linier.
4. Tidak adanya multikolinieritas antar variable (antar variable independent tidak saling berkorelasi) → Collinierity Statistic
5. Adanya Homoscedastisity (penyebaran) → *melihat scatterplot*
6. Sisaan (error) berdistribusi normal → *melihat chart histogram, melihat P-Plot*
7. Sisaan (error) saling bebas → pakai uji Durbin Watson

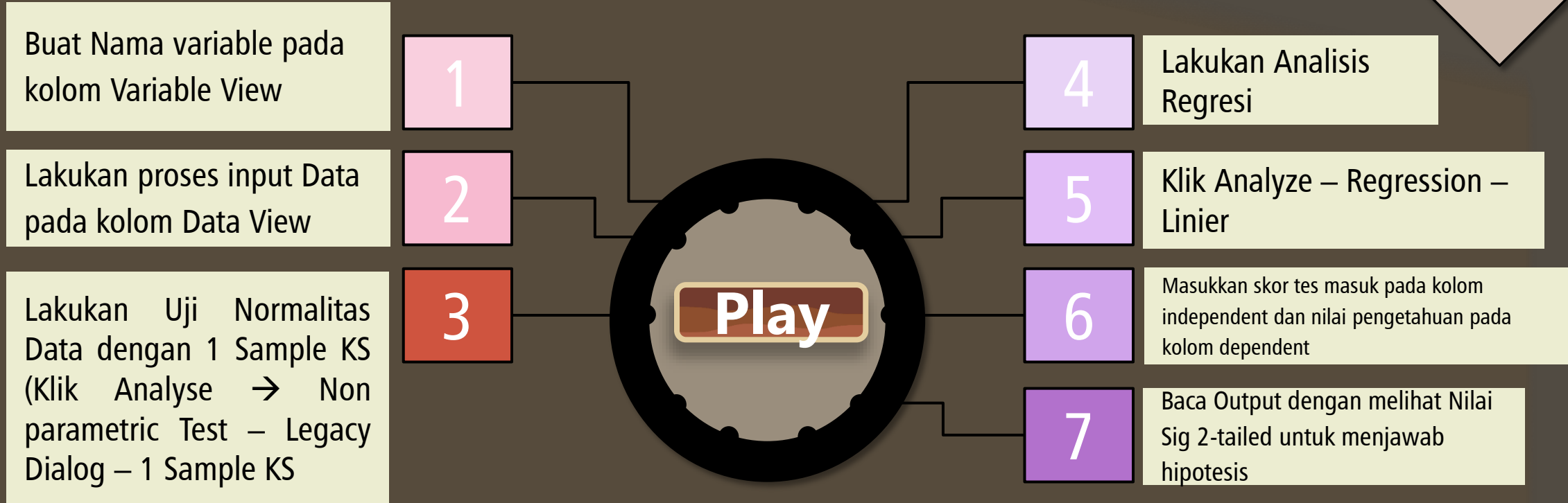
STUDI KASUS UJI REGRESI LINIER SEDERHANA

Seorang HSE perusahaan ingin meneliti pengaruh nilai tes masuk pekerja dengan tingkat pengetahuan pekerja seputar APD (Alat Pelindung Diri). Dengan mengetahui hasil tersebut diharapkan manajemen perusahaan dapat memberikan intervensi yang sesuai, apakah perlu diadakan pelatihan khusus terkait APD atau hanya dijelaskan pada saat safety talk. Berikut adalah datanya:

Responden	NIILAI TES MASUK PEKERJA	PENGETAHUAN TERKAIT APD
1	67	70
2	60	75
3	70	75
4	77	70
5	81	80
6	83	80
7	78	85
8	76	80
9	70	60
10	65	55
11	60	55
12	67	60
13	68	60
14	72	65
15	74	65



Uji Regresi Linier Sederhana dengan SPSS



5

Bagaimana Mudah Bukan??
Ingin konsultasi statistik??, Anda bisa mengunjungi
<https://temanstatistik.wordpress.com/>
indah.lutfiya@vokasi.unair.ac.id

Latihan Uji regresi Linier Ganda (Multivariate)

Local Disk C → Program Files → IBM → SPSS → Statistic → 21 → samples → English → Employee data

Apakah ada pengaruh antara education, salbegin, jobtime, prevexp terhadap current salary (gaji saat ini) pada pegawai di PT X)?



1: Visible: 10 of 10 Variables

	id	gender	bdate	educ	jobcat	salary	salbegin	jobtime	prevexp	minority	var	var	var	var	var	var
1	1 ...		02/03/1952	15	Manager	\$57,000	\$27,000	98	144	No						
2	2 ...		05/23/1958	16	Clerical	\$40,200	\$18,750	98	36	No						
3	3 ...		07/26/1929	12	Clerical	\$21,450	\$12,000	98	381	No						
4	4 ...		04/15/1947	8	Clerical	\$21,900	\$13,200	98	190	No						
5	5 ...		02/09/1955	15	Clerical	\$45,000	\$21,000	98	138	No						
6	6 ...		08/22/1958	15	Clerical	\$32,100	\$13,500	98	67	No						
7	7 ...		04/26/1956	15	Clerical	\$36,000	\$18,750	98	114	No						
8	8 ...		05/06/1966	12	Clerical	\$21,900	\$9,750	98	missing	No						
9	9 ...		01/23/1946	15	Clerical	\$27,900	\$12,750	98	115	No						
10	10 ...		02/13/1946	12	Clerical	\$24,000	\$13,500	98	244	No						
11	11 ...		02/07/1950	16	Clerical	\$30,300	\$16,500	98	143	No						
12	12 ...		01/11/1966	8	Clerical	\$28,350	\$12,000	98	26	Yes						
13	13 ...		07/17/1960	15	Clerical	\$27,750	\$14,250	98	34	Yes						
14	14 ...		02/26/1949	15	Clerical	\$35,100	\$16,800	98	137	Yes						
15	15 ...		08/29/1962	12	Clerical	\$27,300	\$13,500	97	66	No						
16	16 ...		11/17/1964	12	Clerical	\$40,800	\$15,000	97	24	No						
17	17 ...		07/18/1962	15	Clerical	\$46,000	\$14,250	97	48	No						
18	18 ...		03/20/1956	16	Manager	\$103,750	\$27,510	97	70	No						
19	19 ...		08/19/1962	12	Clerical	\$42,300	\$14,250	97	103	No						
20	20 ...		01/23/1940	12	Clerical	\$26,250	\$11,550	97	48	No						
21	21 ...		02/19/1963	16	Clerical	\$38,850	\$15,000	97	17	No						

Data View Variable View



VARIABEL

1. Salary (gaji) \rightarrow variabel dependen = (y)
2. Education (lama pendidikan) = (x1)
3. Salbegin (salary begin); gaji pegawai pertama masuk kerja (x2)
4. Jobtime (lama kerja dalam 1 tahun) = (x3)
5. Prev experience (lama pekerjaan sebelumnya) (x4)



- Reports
- Descriptive Statistics
- Custom Tables
- Compare Means
- General Linear Model
- Generalized Linear Models
- Mixed Models
- Correlate
- Regression**
- Loglinear
- Neural Networks
- Classify
- Dimension Reduction
- Scale
- Nonparametric Tests
- Forecasting
- Survival
- Multiple Response
- Missing Value Analysis...
- Multiple Imputation
- Complex Samples
- Simulation...
- Quality Control
- ROC Curve...
- Spatial and Temporal Modeling...



Visible: 10 of 10 Variables

	id	gender	birthdate	salary	salbegin	jobtime	prevexp	minority	var	var	var	var	var	var
1	1	...	02	\$57,000	\$27,000	98	144	No						
2	2	...	05	\$40,200	\$18,750	98	36	No						
3	3	...	07	\$21,450	\$12,000	98	381	No						
4	4	...	04				190	No						
5	5	...	02				138	No						
6	6	...	08				67	No						
7	7	...	04				114	No						
8	8	...	04				missing	No						
9	9	...	01				115	No						
10	10	...	02				244	No						
11	11	...	02				143	No						
12	12	...	01				26	Yes						
13	13	...	07				34	Yes						
14	14	...	02				137	Yes						
15	15	...	08				66	No						
16	16	...	11				24	No						
17	17	...	07				48	No						
18	18	...	03	103,750	\$27,510	97	70	No						
19	19	...	08	\$42,300	\$14,250	97	103	No						
20	20	...	01	\$26,250	\$11,550	97	48	No						
21	21	...	02	\$38,850	\$15,000	97	17	No						

Data View Variable View



1: Visible: 10 of 10 Variables

	id	gender	bdate	educ	jobcat	salary	salbegin	jobtime	prevexp	minority	var	var	var	var	var
1	1	...	02/03/1952	15											
2	2	...	05/23/1958	16											
3	3	...	07/26/1929	12											
4	4	...	04/15/1947	8											
5	5	...	02/09/1955	15											
6	6	...	08/22/1958	15											
7	7	...	04/26/1956	15											
8	8	...	05/06/1966	12											
9	9	...	01/23/1946	15											
10	10	...	02/13/1946	12											
11	11	...	02/07/1950	16											
12	12	...	01/11/1966	8											
13	13	...	07/17/1960	15											
14	14	...	02/26/1949	15											
15	15	...	08/29/1962	12											
16	16	...	11/17/1964	12											
17	17	...	07/18/1962	15											
18	18	...	03/20/1956	16											
19	19	...	08/19/1962	12	Clerical	\$42,300	\$14,250	97	103	No					
20	20	...	01/23/1940	12	Clerical	\$26,250	\$11,550	97	48	No					
21	21	...	02/19/1963	16	Clerical	\$38,850	\$15,000	97	17	No					

Linear Regression

Dependent: Current Salary [salary]

Block 1 of 1

Independent(s): Educational Level (years) ..., Beginning Salary [salbegin], Months since Hire [jobtime]

Method: Enter

Selection Variable: [] Rule...

Case Labels: []

WLS Weight: []

OK Paste Reset Cancel Help

Statistics... Plots... Save... Options... Style... Bootstrap...

Data View Variable View



- Output
 - Log
 - Regression
 - Title
 - Notes
 - Active Dataset
 - Variables Entered
 - Model Summary
 - ANOVA
 - Coefficients
 - Log
 - Regression
 - Title
 - Notes
 - Active Dataset
 - Variables Entered
 - Model Summary
 - ANOVA
 - Coefficients
 - Log
 - Regression
 - Title
 - Notes
 - Active Dataset
 - Variables Entered
 - Model Summary
 - ANOVA
 - Coefficients
 - Residuals Statistics
 - Charts
 - Title
 - *sresid by *

Model

1	(Constant)
	Educational Level (years)
	Beginning Salary
	Months since Hire
	Previous Experience (months)

a. Dependent Variable: Current Salary

Predicted Value					
Std. Predicted Value					
Standard Error of Predicted Value					
Adjusted Predicted Value					
Residual					
Std. Residual					
Deleted Residual	-\$31,485.213	\$46,621.117	-\$5.882	\$7,553.608	474
Stud. Deleted Residual	-4.160	6.474	.002	1.016	474
Mahal. Distance	.300	85.439	3.992	5.306	474
Cook's Distance	.000	.223	.003	.016	474
Centered Leverage Value	.001	.181	.008	.011	474

a. Dependent Variable: Current Salary

Linear Regression

Dependent

Linear Regression: Statistics

Regression Coefficients

- Estimates
- Confidence intervals
- Level(%): 95
- Covariance matrix

Model fit

- Model fit
- R squared change
- Descriptives
- Part and partial correlations
- Collinearity diagnostics

Residuals

- Durbin-Watson
- Casewise diagnostics
- Outliers outside: 3 standard deviations
- All cases

Continue Cancel Help

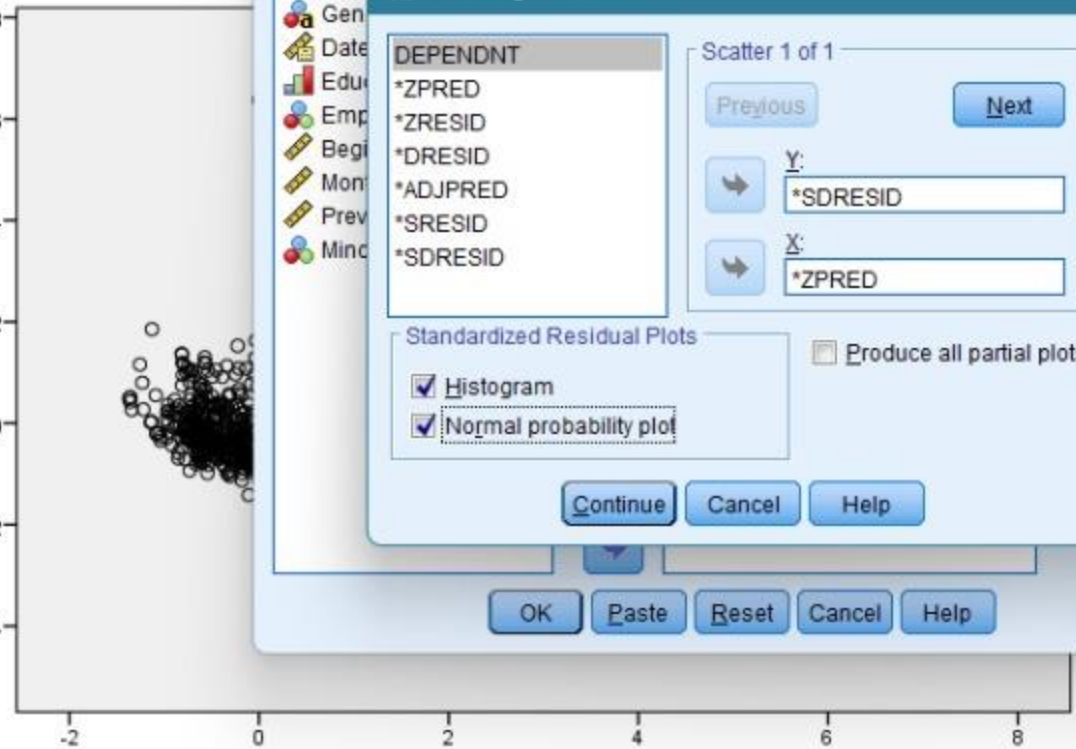
OK Paste Reset Cancel Help



- Log
- Regression
 - Coefficients
 - Log
 - Regression
 - Title
 - Notes
 - Active Dataset
 - Variables Entered
 - Model Summary
 - ANOVA
 - Coefficients
 - Log
 - Regression
 - Title
 - Notes
 - Variables Entered
 - Model Summary
 - ANOVA
 - Coefficients
 - Residuals Statistics
 - Charts
 - Title
 - *sdresid b
- Log
- Regression
 - Title
 - Notes
 - Variables Entered
 - Model Summary
 - ANOVA
 - Coefficients
 - Collinearity Diagnostics
 - Residuals Statistics
- Charts
 - Title
 - *sdresid b

Charts

Regression Studentized Deleted (Press) Residual



Regression Standardized Predicted Value

Linear Regression

Dependent

Linear Regression: Plots

DEPENDENT

- *ZPRED
- *ZRESID
- *DRESID
- *ADJPRED
- *SRESID
- *SDRESID

Scatter 1 of 1

Previous Next

Y: *SDRESID

X: *ZPRED

Standardized Residual Plots

- Histogram
- Normal probability plot
- Produce all partial plots

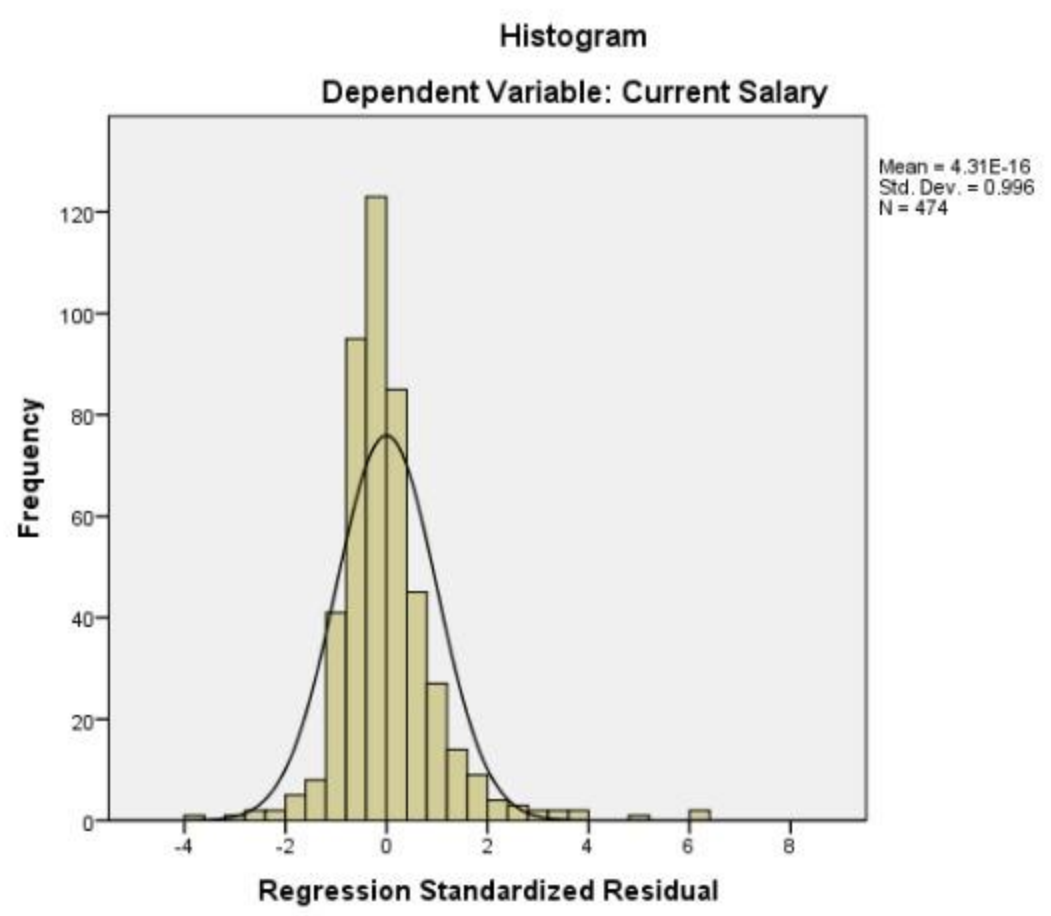
Continue Cancel Help

OK Paste Reset Cancel Help



- Model Summary
- ANOVA
- Coefficients
- Residuals Stat
- Charts
 - Title
 - *sresid b
- Log
- Regression
 - Title
 - Notes
 - Variables Enter
 - Model Summar
 - ANOVA
 - Coefficients
 - Collinearity Dia
 - Residuals Stati
 - Charts
 - Title
 - *sresid b
- Log
- Regression
 - Title
 - Notes
 - Variables Enter
 - Model Summar
 - ANOVA
 - Coefficients
 - Collinearity Dia
 - Residuals Stati
 - Charts
 - Title
 - *zresid His
 - *zresid No
 - *sresid b

Charts

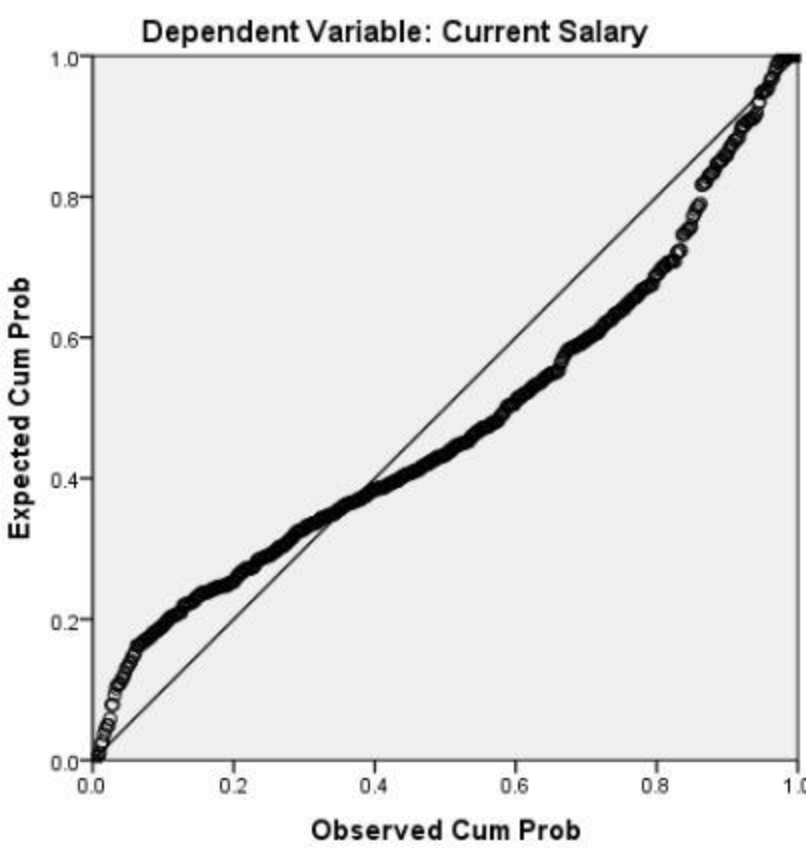


Sisaan berdistribusi normal, Kurva sisaan mendekati Kurva Normal



- Model Summary
- ANOVA
- Coefficients
- Residuals Statistics
- Charts
 - Title
 - *sresid b
- Log
- Regression
 - Title
 - Notes
 - Variables Entered/Removed
 - Model Summary
 - ANOVA
 - Coefficients
 - Collinearity Diagnostics
 - Residuals Statistics
 - Charts
 - Title
 - *sresid b
- Log
- Regression
 - Title
 - Notes
 - Variables Entered/Removed
 - Model Summary
 - ANOVA
 - Coefficients
 - Collinearity Diagnostics
 - Residuals Statistics
 - Charts
 - Title
 - *zresid Histogram
 - *zresid Normal Q-Q Plot
 - *sresid b

Normal P-P Plot of Regression Standardized Residual

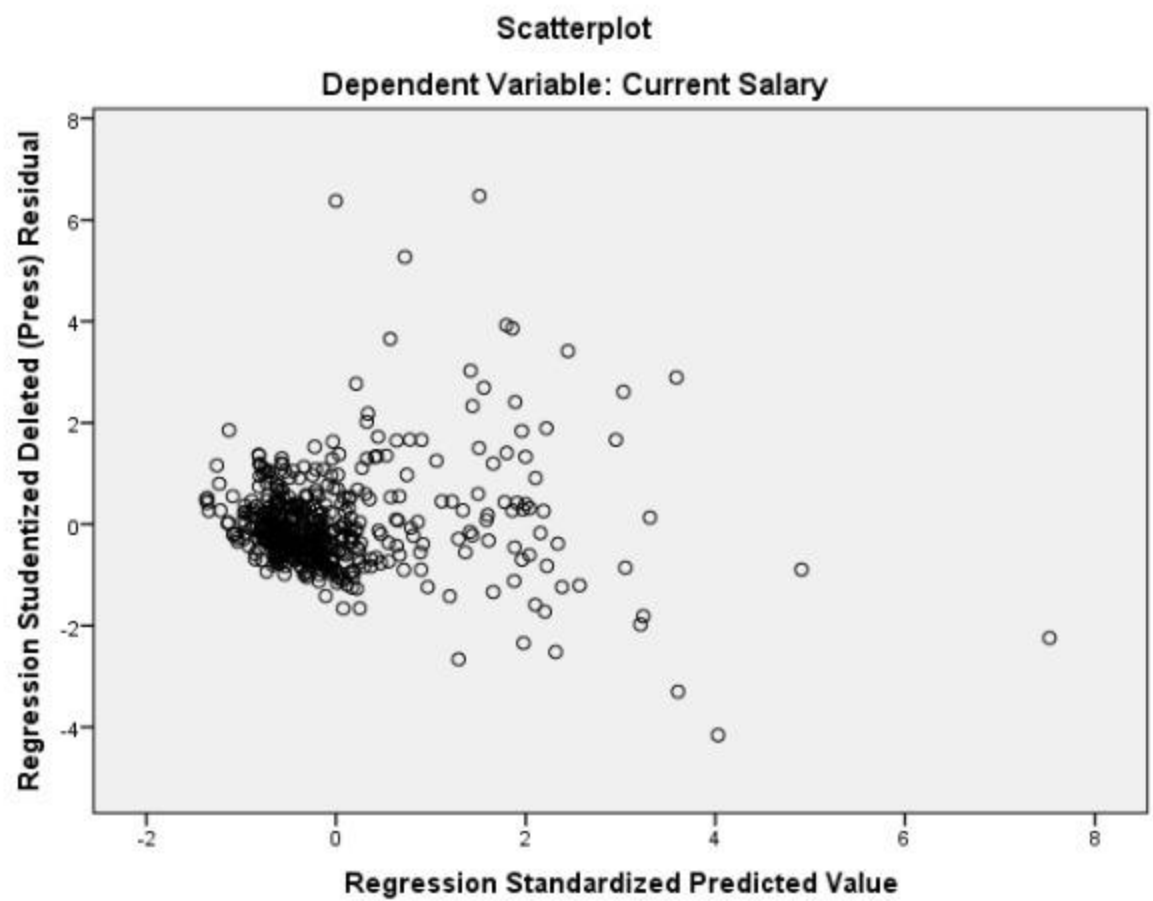


Sisaan berdistribusi normal,
Apabila titik-titik mendekati 45,
berarti Kurva Normal

Scatterplot

- Output
 - Log
 - Regression
 - Title
 - Notes
 - Active Dataset
 - Variables Entered
 - Model Summary
 - ANOVA
 - Coefficients
 - Log
 - Regression
 - Title
 - Notes
 - Active Dataset
 - Variables Entered
 - Model Summary
 - ANOVA
 - Coefficients
 - Log
 - Regression
 - Title
 - Notes
 - Variables Entered
 - Model Summary
 - ANOVA
 - Coefficients
 - Residuals Statisti
 - Charts
 - Title
 - *sdrsid by *2

Charts



Gambar tidak menunjukkan pola gelombang, melengkung atau mengerucut – asumsi Homoscedastisity terpenuhi

PEMBAHASAN

1. Nilai $R = 0,900$ artinya adanya keeratan hubungan antara variable independent (x_1, x_2, x_3, x_4) terhadap dependent (y). Semakin mendekati 1 maka hubungan semakin kuat. Sedangkan semakin mendekati 0 maka hubungan semakin lemah. nilai 0,9 menunjukkan hubungan kuat.
2. Nilai R^2 (RSquare) atau biasa disebut Koefisien Determinasi sebesar 0,81 artinya keragaman variable independent (x_1, x_2, x_3, x_4) dapat menjelaskan 81% keragaman variable dependent. Model termasuk sangat baik karena mendekati 100%. Sisanya sebanyak 19% dapat dijelaskan oleh variable lain yang tidak sedang diteliti.
3. $R \text{ Adjusted} = 0,809$
4. Nilai Durbin Watson di Output = 1,921

DURBIN WATSON

Hipotesis=

H_0 = sisaan saling bebas (tidak ada korelasi antar sisaan)

H_1 = sisaan tidak saling bebas (terdapat korelasi antar sisaan)

Cara pengambilan Keputusan:

- a) **Jika Nilai D antara D_U dan $4 - D_U$, maka H_0 diterima**
- b) Jika Nilai $D < D_L$ maka H_0 ditolak (Autokorelasi positif)
- c) Jika Nilai $D > 4 - D_L$ maka H_0 ditolak (Autokorelasi Negatif)
- d) Jika Nilai D antara D_U dan D_L atau antara $4 - D_U$ dan $4 - D_L$ maka hasilnya tidak dapat disimpulkan

MULTIKOLINIERITAS

Lihat kolom Collinierity Statistic akan ada Tolerance dan VIF

Nilai VIF artinya $1 / \text{tolerance}$.

Jika nilai Tolerance dibawah 0,1 atau Nilai VIF diatas 10 maka artinya terdapat korelasi pada variable independent

Apabila nilai Tolerance **TIDAK** dibawah 0,1 atau Nilai VIF **TIDAK** diatas 10 maka artinya **TIDAK** terdapat korelasi pada variable independent (Tidak terdapat multikolinieritas pada variable independent) → **ASUMSI terpenuhi**

MODEL YANG DIDAPATKAN

Y (current salary) = $-16149,671 + 669,914$ (education level) $+1,768$ (beginning salary) $+ 161,486$ (months since hire) $- 17,303$ (previous experience)

TERIMA KASIH