

ANALISIS MULTIVARIAT

sesi-2 (multiple regression)

PROGRAM STUDI S1 PWK
UNIVERSITAS GADJAH MADA
2006



Multiple regression/Regresi berganda

Pengertian:

Suatu **variabel dependen** dipengaruhi oleh banyak **variabel independen**.

Jadi Y dipengaruhi oleh X_1, X_2, X_3, \dots dst (ini disebut juga *explanatory variable*).

Misal hanya dipengaruhi oleh X_1 dan X_2 saja, maka:

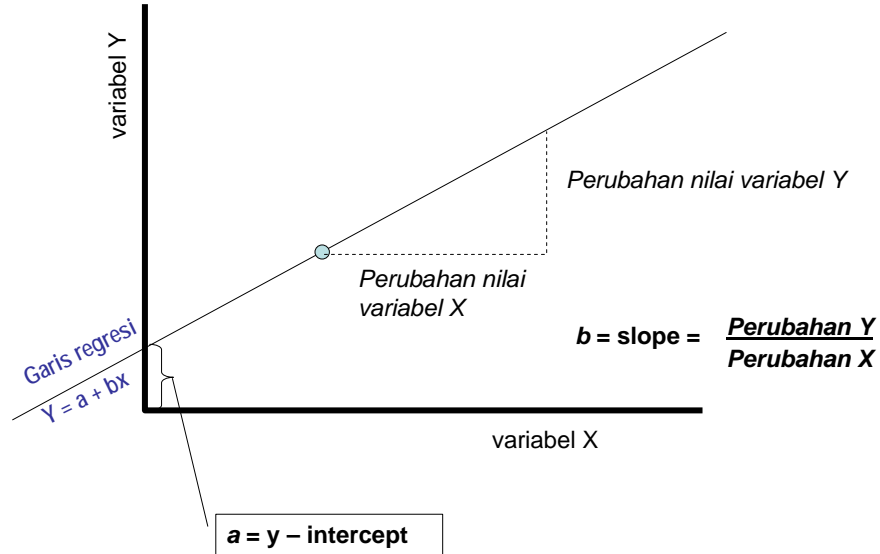
$$Y_i = \beta_0 + \beta_1 \cdot X_{1i} + \beta_2 \cdot X_{2i} + \varepsilon_i$$

constanta Koef..regresi 1 Koef..regresi 2 Stand.error

→ $Y_i = b_0 + b_1 \cdot X_{1i} + b_2 \cdot X_{2i} + e_i$

a ?

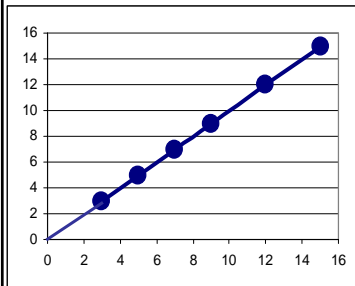
Penjelasan tentang a atau konstanta



reg

DATA:

Y	X
3	3
5	5
7	7
9	9
12	12
15	15



Regression

Variables Entered/Removed^a

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	1.000 ^a	1.000	1.000	.000

a. Predictors: (Constant), X

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig. ^a
1	Regression	99.500	1	99.500		
	Residual	.000	4	.000		
	Total	99.500	5			

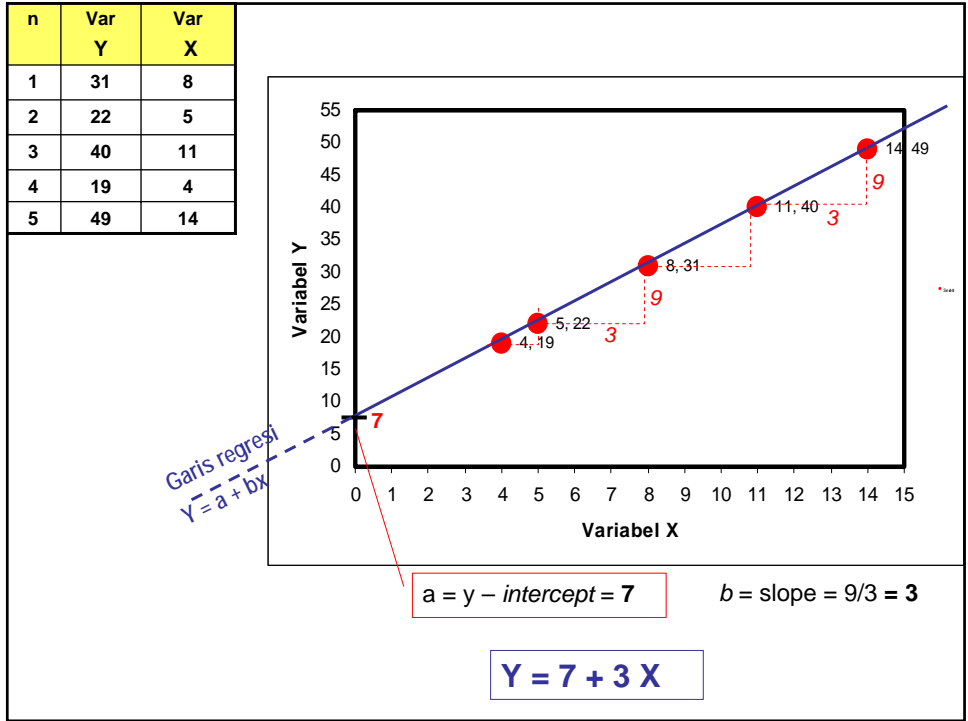
a. Predictors: (Constant), X

b. Dependent Variable: Y

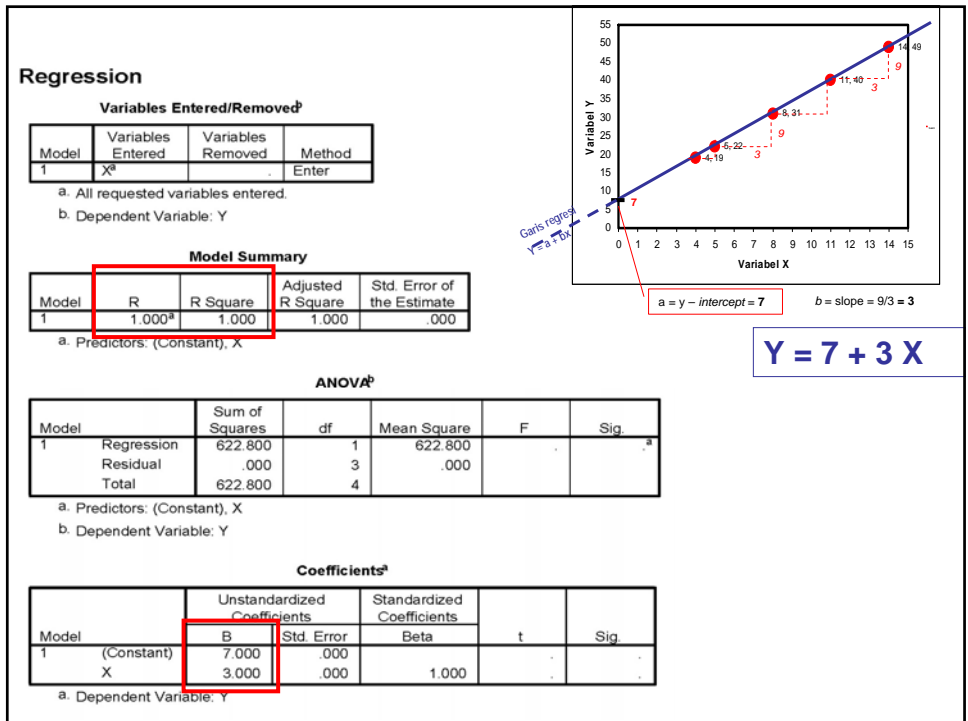
Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.000	.000			
	X	1.000	.000	1.000		

a. Dependent Variable: Y



regresi



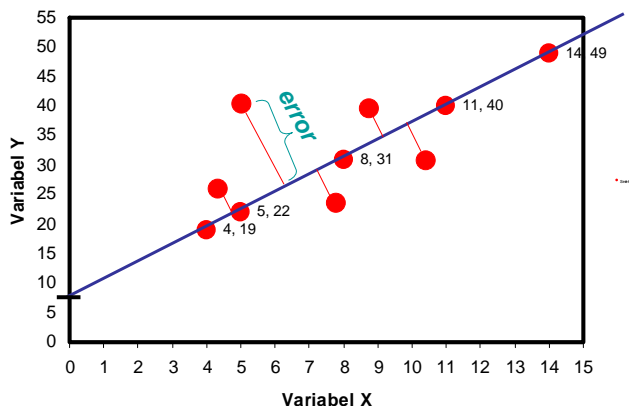
Pada kenyataannya, banyak kasus yang tidak terletak pada satu garis (regresi). Ini kemudian memunculkan **error**.

$$Y_i = a + b X_i + \varepsilon_i \quad i = 1, 2, \dots, N \text{ (jumlah sampel)}$$

$$\varepsilon_i = Y_i - a - b X_i$$

Karena banyak, maka diperlukan "the total sum of squared errors" (SSE):

$$SSE = \sum_{i=1}^N \varepsilon_i^2 \rightarrow \text{semakin kecil semakin baik (minimalkan!!)}$$



Dalam output SPSS disebut RESIDUAL



Asumsi yang mendasari analisis multiple regression ini:

1. Variable pengganggu e adalah variabel random.
2. Variabel random e punya nilai rata-rata untuk masing-masing $X_i = 0$
3. Homoskedastisitas= varians masing-masing e adalah sama untuk setiap X_i
4. Normalitas, yakni variabel pengganggu e berdistribusi normal.
5. Variabel pengganggu e independen terhadap variabel penjelasan X_i .
6. Tidak ada kesalahan pengukuran pada variabel penjelas (explanatory variable).
7. Tidak ada multikolinieritas sempurna antar variabel independen.
8. Fungsi yang ditaksir adalah hubungan yang identified.
9. Spesifikasi modelnya benar
10. Explanatory variable adalah nonstochastic (bisa diprediksi).
11. Nonautocorrelation.

Silakan pelajari sendiri dari berbagai pustaka

Contoh:

Suatu penelitian tentang keterkaitan banyak variabel didapat data seperti di samping ini.

Tujuan penelitian tersebut adalah mencari besarnya pengaruh independen (explanatory) variabel yakni X1, X2, X3, X4 terhadap dependen variabelnya yakni Y.

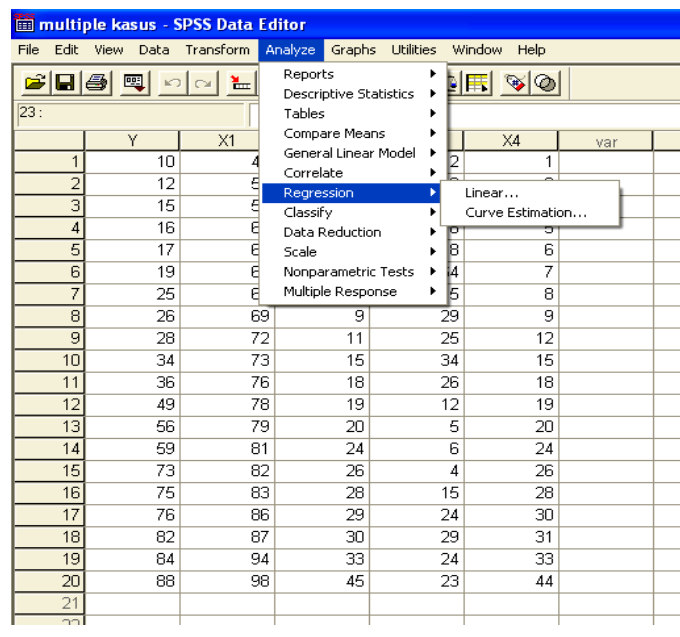
Penelitian tersebut juga mencari berapa jumlah Y yang diperlukan bila masing-masing explanatory variable menjadi 100.

No	Y	X1	X2	X3	X4
1	10	45	1	12	1
2	12	58	2	16	2
3	15	59	4	5	4
4	16	61	5	8	5
5	17	66	6	8	6
6	19	67	7	64	7
7	25	68	8	5	8
8	26	69	9	29	9
9	28	72	11	25	12
10	34	73	15	34	15
11	36	76	18	26	18
12	49	78	19	12	19
13	56	79	20	5	20
14	59	81	24	6	24
15	73	82	26	4	26
16	75	83	28	15	28
17	76	86	29	24	30
18	82	87	30	29	31
19	84	94	33	24	33
20	88	98	45	23	44

Agam

Gunakan software SPSS.

Masukkan data lalu lakukan Analyze, pilih Regression dan kemudian pilih Linear...



multiple kasus - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window Help

23:

	Y	X1	X2	X3	X4
1	10	45	1	12	1
2	12	58	2	16	2
3	15	59	4	5	4
4	16	61	5	8	5
5	17	66	6	8	6
6	19	67	7	64	7
7	25	68	8	5	8
8	26	69	9	29	9
9	28	72	11	25	12
10	34	73	15	34	15
11	36	76	18	26	18
12	49	78	19	12	19
13	56	79	20	5	20
14	59	81	24	6	24
15	73	82	26	4	26
16	75	83	28	15	28
17	76	86	29	24	30
18	82	87	30	29	31
19	84	94	33	24	33
20	88	98	45	23	44
21					

Linear Regression

Dependent: Y

Independent(s): X1, X2, X3

Method: Enter

Selection: Enter, Stepwise, Remove, Backward, Forward

Case Labels:

WLS Weight:

Statistics... Plots... Save... Options...

Masukkan variabel Y ke cell Dependent, dan masukan seluruh X1, X2, X3, X4 ke cell Independent(s).

Method ada 5 pilihan : Enter, Stepwise, Remove, Backward, dan Forward

Regmm

Regression - ENTER
Variables Entered/Removed(b)

Model	Variables Entered	Variables Removed	Method
1	X4, X3, X1, X2(a)		Enter

a All requested variables entered.
b Dependent Variable: Y

Model Summary
b Dependent Variable: Y

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,984(a)	,969	,960	5,493

a Predictors: (Constant), X4, X3, X1, X2

ANOVA(b)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	14011,342	4	3502,835	116,076	,000(a)
	Residual	452,658	15	30,177		
	Total	14464,000	19			

a Predictors: (Constant), X4, X3, X1, X2
b Dependent Variable: Y

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	10,006	19,697		,508	,619
	X1	-,038	,352	-,018	-,108	,915
	X2	-6,439	2,867	-2,842	-2,246	,040
	X3	-,151	,091	-,079	-1,664	,117
	X4	8,719	2,934	3,839	2,972	,009

a Dependent Variable: Y

Method ENTER: semua variabel independen dimasukkan kedalam persamaan dan dicari r^2 nya (t-value atau F-value).

Regression - Forward
Variables Entered/Removed(a)

Model	Variables Entered	Variables Removed	Method
1	X4	.	Forward (Criterion: Probability-of-F-to-enter <= .050)

a Dependent Variable: Y

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,977(a)	,954	,951	6,080

a Predictors: (Constant), X4

ANOVA(b)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	13798,559	1	13798,559	373,247	,000(a)
	Residual	665,441	18	36,969		
	Total	14464,000	19			

a Predictors: (Constant), X4
b Dependent Variable: Y

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	6,065	2,388		2,539	,021
	X4	2,218	,115	,977	19,320	,000

a Dependent Variable: Y

Excluded Variables(b)

Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics Tolerance
1	X1	-,033(a)	-,183	,857	-,044	,081
	X2	-2,508(a)	-1,957	,067	-,429	,001
	X3	-,065(a)	-1,300	,211	-,301	,996

a Predictors in the Model: (Constant), X4
b Dependent Variable: Y

Method FOREWARD (inclusion): Mengkaitkan lebih dulu antara Y dengan X.. yang r^2 paling besar, kemudian step berikutnya ditambahkan lagi X..(lain) yang korelasi parsialnya paling besar dan akan berhenti bila ditambahkan lagi X..lain yang tidak menambah r^2 -nya (t-value atau F-value).

regmm

Regression - Backward (1)
Variables Entered/Removed(b)

Model	Variables Entered	Variables Removed	Method
1	X4, X3, X1, X2(a)	.	Enter
2	.	X1	Backward (criterion: Probability of F-to-remove >= ,100).

a All requested variables entered.
b Dependent Variable: Y

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,984(a)	,969	,960	5,493
2	,984(b)	,969	,963	5,321

a Predictors: (Constant), X4, X3, X1, X2
b Predictors: (Constant), X4, X3, X2

ANOVA(c)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	14011,342	4	3502,835	116,076	,000(a)
	Residual	452,658	15	30,177		
	Total	14464,000	19			
2	Regression	14010,988	3	4670,329	164,952	,000(b)
	Residual	453,012	16	28,313		
	Total	14464,000	19			

a Predictors: (Constant), X4, X3, X1, X2
b Predictors: (Constant), X4, X3, X2
c Dependent Variable: Y

Method BACKWARD (elimination): Memasukkan semua explanasi variabel, kemudian step berikutnya mengurangi peran X..yang yang paling kecil perannya menentukan r^2 -nya. Penghentian akan dilakukan bila mencapai nilai r^2 yang berkurang secara berarti.

Regression – Backward (2)

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	10,006	19,697		,508	,619
	X1	-,038	,352	-,018	-,108	,915
	X2	-6,439	2,867	-,2842	-2,246	,040
	X3	-,151	,091	-,079	-1,664	,117
	X4	8,719	2,934	3,839	2,972	,009
2	(Constant)	7,894	2,566		3,076	,007
	X2	-6,407	2,763	-,2828	-2,319	,034
	X3	-,153	,086	-,080	-1,784	,093
	X4	8,649	2,771	3,808	3,121	,007

a Dependent Variable: Y

Excluded Variables(b)

Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics
						Tolerance
2	X1	-,018(a)	-,108	,915	-,028	,076

a Predictors in the Model: (Constant), X4, X3, X2

b Dependent Variable: Y

reg mm

Regression - Stepwise

Variables Entered/Removed(a)

Model	Variables Entered	Variables Removed	Method
1	X4		Stepwise (Criteria: Probability-of-F-to-enter <= ,050, Probability-of-F-to-remove >= ,100).

a Dependent Variable: Y

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,977(a)	,954	,951	6,080

a Predictors: (Constant), X4

ANOVA(b)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	13798,559	1	13798,559	373,247	,000(a)
	Residual	665,441	18	36,969		
	Total	14464,000	19			

a Predictors: (Constant), X4

b Dependent Variable: Y

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	6,065	2,388		2,539	,021
	X4	2,218	,115	,977	19,320	,000

a Dependent Variable: Y

Excluded Variables(b)

Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics
						Tolerance
1	X1	-,033(a)	-,183	,857	-,044	,081
	X2	-2,508(a)	-1,957	,067	-,429	,001
	X3	-,065(a)	-1,300	,211	-,301	,996

a Predictors in the Model: (Constant), X4

b Dependent Variable: Y

Method STEPWISE:

Menggabungkan method FOREWARD dan BACKWARD untuk menghasilkan kombinasi variabel yang mempunyai r^2 paling besar.

Regression - Remove

Warnings

Invalid REGRESSION METHOD subcommand specification--REMOVE cannot be used as the first method when building an equation. REGRESSION has inserted ENTER as the first method; REMOVE is now the second method.

Variables Entered/Removed(c)

Model	Variables Entered	Variables Removed	Method
1	X4, X3, X1, X2(a)	.	Enter
2	.(a)	X4, X1, X3, X2(b)	Remove

- a All requested variables entered.
 b All requested variables removed.
 c Dependent Variable: Y

Method REMOVE (?): semua variabel independen TIDAK dimasukkan kedalam persamaan. Biasanya hal ini berkaitan dengan method BACKWARD atau FOREWARD.



Regression - ENTER

Variables Entered/Removed(b)

Model	Variables Entered	Variables Removed	Method
1	X4, X3, X1, X2(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	,984(a)	,969	,960	5,493

- a Predictors: (Constant), X4, X3, X1, X2

ANOVA(b)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	14011,342	4	3502,835	116,076	,000(a)
	Residual	452,658	15	30,177		
	Total	14464,000	19			

- a Predictors: (Constant), X4, X3, X1, X2
 b Dependent Variable: Y

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	10,006	19,697		,508	,619
	X1	-,038	,352	-,018	-,108	,915
	X2	-6,439	2,867	-2,842	-2,246	,040
	X3	-,151	,091	-,079	-1,664	,117
	X4	8,719	2,934	3,839	2,972	,009

- a Dependent Variable: Y

Adjusted R Square : digunakan sebagai koefisien determinasi. Semakin tinggi nilai adjusted R Square berarti semakin baik model regresinya.

Standard Error of the Estimate: Semakin kecil nilainya akan semakin tepat untuk memprediksi variabel dependen.

Dalam kasus ini = 0,484 → termasuk kecil.

ANOVA(b)						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	14011,342	4	3502,835	116,076	,000(a)
	Residual	452,658	15	30,177		
	Total	14464,000	19			

a Predictors: (Constant), X4, X3, X1, X2
 b Dependent Variable: Y

Coefficients(a)						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	10,006	19,697		,508	,619
	X1	-,038	,352	-,018	-,108	,915
	X2	-6,439	2,867	-2,842	-2,246	,040
	X3	-,151	,091	-,079	-1,664	,117
	X4	8,719	2,934	3,839	2,972	,009

a Dependent Variable: Y

F – test berguna untuk menguji persamaan multiple regresi untuk total model.

$$F_{hitung} = 116,076 > F_{tabel} (k, n-k-1)\alpha .$$

$k =$ jumlah explanatory variable, $n =$ jumlah sampel. Jadi $F (4, 20-4-1)_{0,05} = 3,01$.

Artinya: ditolak H_0 , jadi persamaan multiple regresinya signifikan. Dengan demikian persamaan tersebut bisa menjadi alat prediksi dan estimasi.

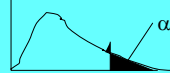
t – test berguna untuk menguji persamaan regresi untuk tiap variabel.

Koefisien dari X1 : $t = -0,108$ cek dengan t tabel !!! ($n =$ jumlah sampel)

Koefisien dari X2 : $t = -2,246$ cek dengan t tabel !!! ($n =$ jumlah sampel). dst. X3, X4

1. Cari dulu α -nya
2. Cari k
3. Cari n-k-1
4. Hubungkan v_1 dan v_2

F-distribution critical points ($\alpha = \dots\dots\dots$)



v_1	1	2	3	4	5	6
v_2					
1						
2						
3						
4						
5						
,						
,						

$F_{tabel} (k, n-k-1)\alpha .$



No	Y	X1	X2	X3	X4
1	10	45	1	12	1
2	12	58	2	16	2
⋮					
⋮					
20	88	98	45	23	44

$k =$ jumlah explanatory variable = 4
 $n =$ jumlah sampel = 20

Jadi $F (4, 20-4-1) \alpha = \dots\dots\dots$

Langkah-langkah analisis:

1. **Cek dulu r dan r² nya.** r = 0,984 positif dan kuat keterkaitannya. r² = 0,969 kuat pengaruh independen variabel terhadap dependen variabelnya. *Tentunya cek pula adjusted R square dan std.error of the estimate.*

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,984(a)	,969	,960	5,493

a Predictors: (Constant), X4, X3, X1, X2

2. **Uji dulu F-test.** Bila hasil F hitung menunjukkan lebih besar daripada F-tabel maka persamaan multiple regresi tersebut bisa digunakan untuk prediksi dan estimasi.

ANOVA(b)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	14011,342	4	3502,835	116,076	,000(a)
	Residual	452,658	15	30,177		
	Total	14464,000	19			

a Predictors: (Constant), X4, X3, X1, X2

b Dependent Variable: Y

$F_{hitung} = 116,076 > F_{tabel} (4, 15) 0,05 = 3,01$. Artinya: ditolak Ho, jadi persamaan multiple regresinya signifikan → bisa menjadi alat prediksi dan estimasi.

3. Persamaan untuk prediksi:

$$Y = a + b_1 \cdot X_1 + b_2 \cdot X_2 + b_3 \cdot X_3 + b_4 \cdot X_4$$

$$Y = 10,006 - 0,038 \cdot X_1 - 6,439 \cdot X_2 - 0,151 \cdot X_3 + 8,719 \cdot X_4$$

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	10,006	19,697		,508	,619
	X1	-,038	,352	-,018	-,108	,915
	X2	-6,439	2,867	-2,842	-2,246	,040
	X3	-,151	,091	-,079	-1,664	,117
	X4	8,719	2,934	3,839	2,972	,009

a Dependent Variable: Y

Standard error dihitung untuk masing-masing variabel tsb.

Penelitian tersebut juga mencari berapa jumlah Y yang diperlukan bila masing-masing explanatory variabel menjadi 100.

$$Y = 10,006 - 0,038 \cdot X_1 - 6,439 \cdot X_2 - 0,151 \cdot X_3 + 8,719 \cdot X_4$$

$(t=-0,108)$ $(t=-2,842)$ $(t=-0,079)$ $(t=-3,839)$ $(F=116,076)$

$$Y = 10,006 - 0,038 \cdot 100 - 6,439 \cdot 100 - 0,151 \cdot 100 + 8,719 \cdot 100$$

$$\text{Jadi } Y = 219,106$$