Regresi Linier Sederhana dan Korelasi

(3 sesi)

Disusun oleh

Sigit Nugroho

Universitas Bengkulu

Pengertian

- **Regresi** merupakan teknik statistika yang digunakan untuk mempelajari hubungan fungsional dari satu atau beberapa peubah bebas (peubah yang mempengaruhi) terhadap satu peubah tak bebas (peubah yang dipengaruhi)

- **Korelasi** merupakan ukuran kekuatan hubungan dua peubah (tidak harus memiliki hubungan sebab akibat)

Regresi

- Dari derajat (pangkat) tiap peubah bebas
  - Linier (bila pangkatnya 1)
  - Non-linier (bila pangkatnya bukan 1)
- Dari banyaknya peubah bebas (yang mempengaruhi)
  - Sederhana (bila hanya ada satu peubah bebas)
  - Berganda (bila lebih dari satu peubah bebas)

Regresi Linier Sederhana

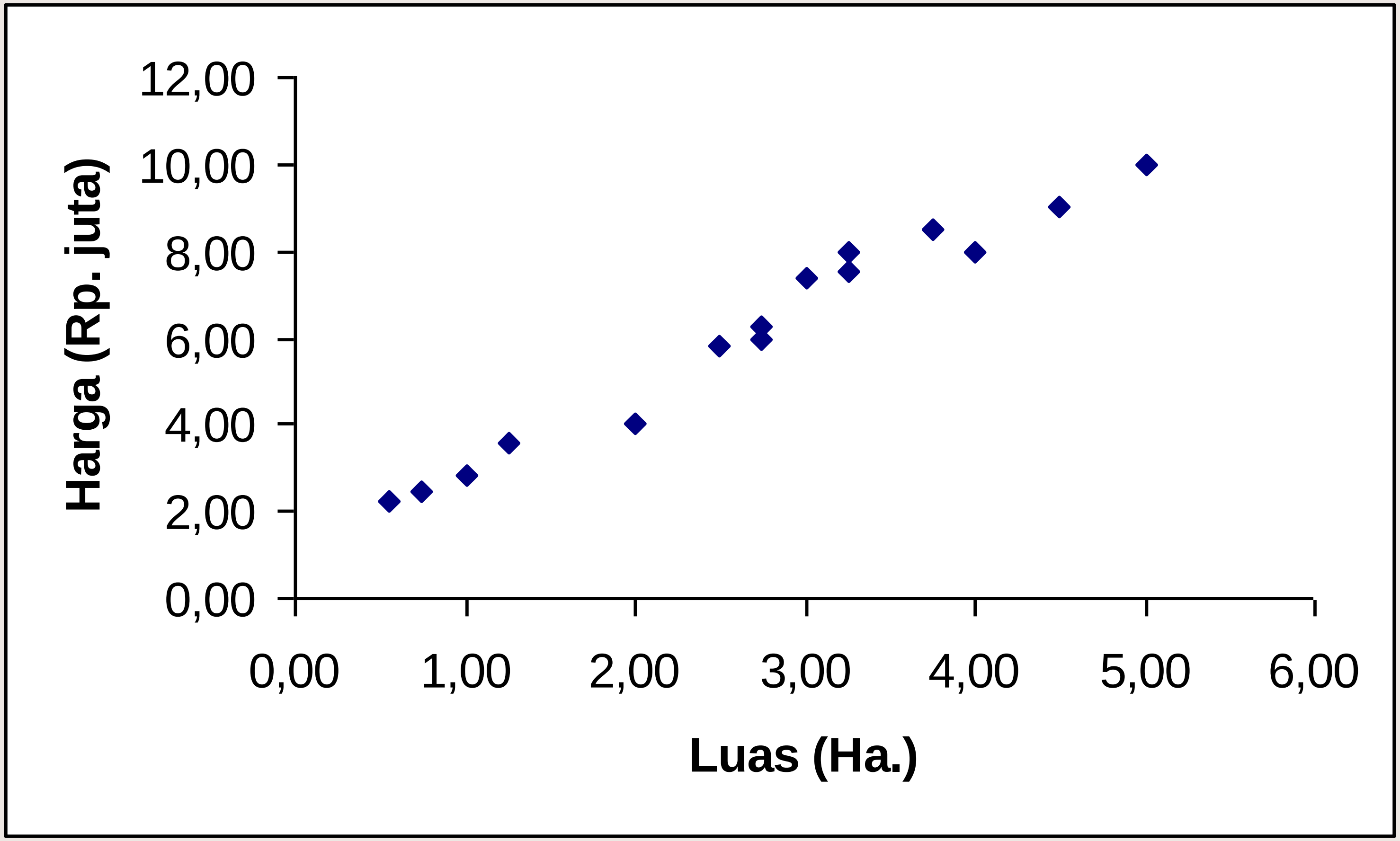
- **Model**
  - $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$
    - ❖ Y_i merupakan nilai pengamatan ke-i.
    - ❖ β_0 adalah parameter regresi (intersep)
    - ❖ β_1 adalah parameter regresi (slope)
    - ❖ ε_i kesalahan ke-i.
  - **Asumsi** :
    - peubah X terukur tanpa kesalahan; X tidak memiliki distribusi (bukan *random variable*)
    - kesalahan menyebar normal dengan rata-rata nol dengan simpangan baku σ_ε.

Teladan Permasalahan

- Dari sebuah survai yang dilakukan di kampung Maju Makmur digunakan untuk mengetahui hubungan fungsional antara luas tanah (hektar) dan harganya (Rp. 00 Juta). Bila data berpasangan tentang luasan dan harga tanah diperoleh, bagaimana hubungan fungsionalnya ?

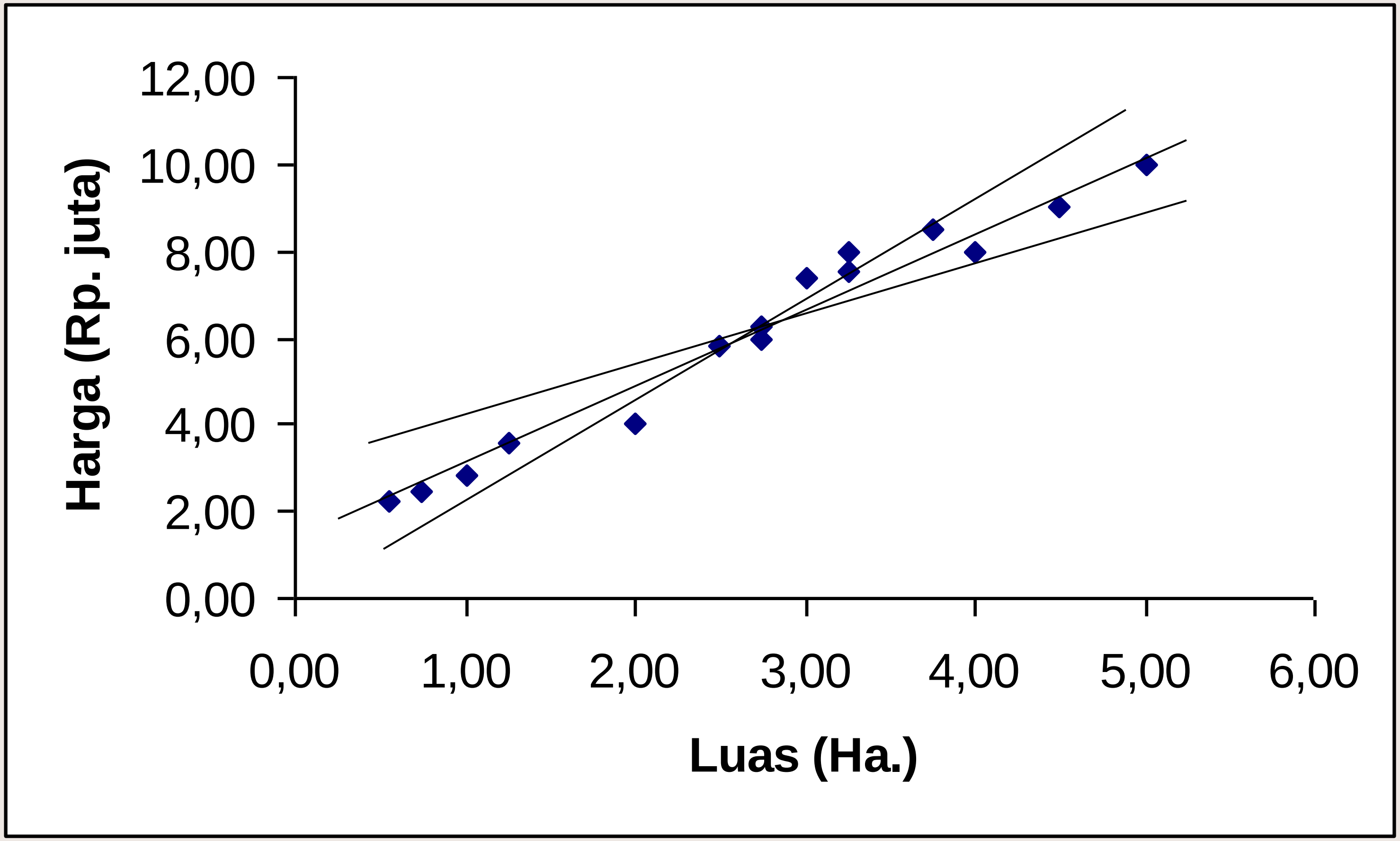
Luas	Harga
0,75	2,45
0,55	2,20
1,00	2,80
1,25	3,60
2,50	5,80
3,00	7,40
4,50	9,00
3,75	8,50
5,00	10,00
3,25	8,00
3,25	7,50
2,75	6,00
2,75	6,25
2,00	4,00
4,00	8,00

Diagram Pencar
(Scatter Plot)



Mana pendekatan yang baik ?

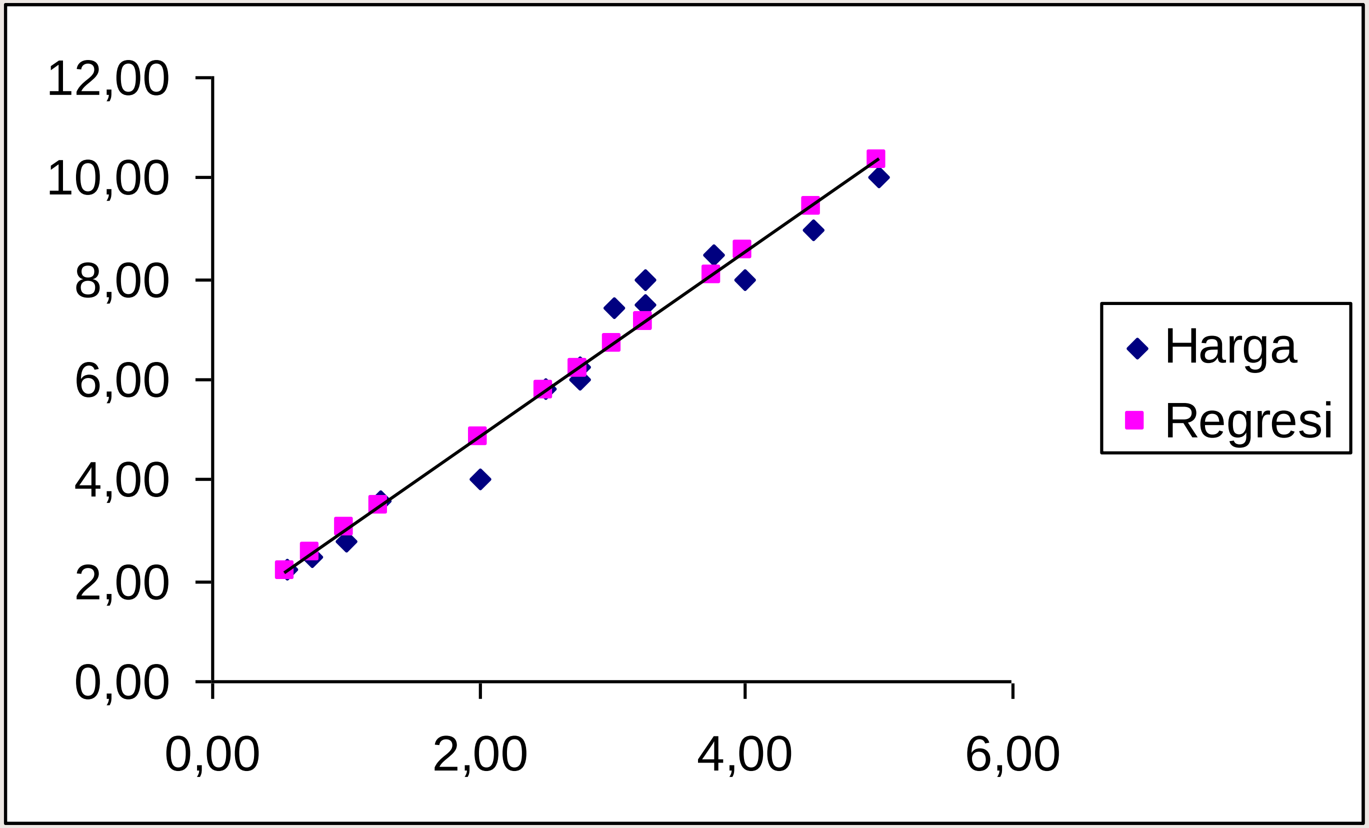
Garis lurus yang sedemikian rupa sehingga melewati seluruh titik (data) pada diagram pencar → yang mendekati



Metode Jumlah Kuadrat Galat Terkecil

(Least Squares Method)

merupakan salah satu kriteria yang memenuhi, agar apabila kuadrat dari kesalahan itu dijumlahkan akan se minimum mungkin.



Persamaan Regresi

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$$

dimana

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i y_i - \frac{\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n}}{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Teladan Hitungan

Luas X	Harga Y	XY	X^2	Y^2
0,75	2,45	1,8375	0,5625	6,0025
0,55	2,20	1,2100	0,3025	4,8400
1,00	2,80	2,8000	1,0000	7,8400
1,25	3,60	4,5000	1,5625	12,9600
2,50	5,80	14,5000	6,2500	33,6400
3,00	7,40	22,2000	9,0000	54,7600
4,50	9,00	40,5000	20,2500	81,0000
3,75	8,50	31,8750	14,0625	72,2500
5,00	10,00	50,0000	25,0000	100,0000
3,25	8,00	26,0000	10,5625	64,0000
3,25	7,50	24,3750	10,5625	56,2500
2,75	6,00	16,5000	7,5625	36,0000
2,75	6,25	17,1875	7,5625	39,0625
2,00	4,00	8,0000	4,0000	16,0000
4,00	8,00	32,0000	16,0000	64,0000
40,30	91,50	293,4850	134,2400	648,6050
2,69	6,10			
			slope	**1,835**
			intersep	**1,169**

Persamaan Regresi

serta penjelasannya

$$\hat{Y}_i = 1{,}169 + 1{,}835 \ X_i$$

Slope bernilai 1,835. Artinya : dua luasan tanah yang berbeda seluas **satu** hektar, tanah yang lebih luas akan memiliki perkiraan harga Rp. 1,835 juta lebih tinggi.

JANGAN diartikan sbb: bila luas tanah meningkat satu hektar, maka harga tanah akan meningkat Rp. 1,835 juta.

Persaman Regresi

serta penjelasannya

$$\hat{Y}_i = 1{,}169 + 1{,}835 \ X_i$$

Slope bernilai 1,169. Untuk teladan ini nilai intersep tidak memiliki arti.

JANGAN diartikan sbb: bila luas tanah (x) = 0 hektar, maka harga tanah adalah Rp. 1,169 juta. Pengartian seperti ini TIDAK benar. Kenapa ???

Persamaan Regresi

serta penjelasannya

$$Y_{x=3} = 1{,}169 + 1{,}835\ (3) = 6{,}675$$

$$Y_{x=2} = 1{,}169 + 1{,}835\ (2) = 4{,}840$$

Tanah yang luasnya 3 ha memiliki perkiraan harga Rp. 1,835 juta lebih tinggi dari yang 2 ha

Menguji Koeffisien Regresi

$$H_0 : \beta_1 = \beta_{10} \text{ vs } H_1 : \beta_1 \neq \beta_{10}$$

Statistik Uji

$$t_{hit} = \frac{\hat{\beta}_1 - \beta_{10}}{s_{\hat{\beta}_1}}$$

dimana

$$s_{\hat{\beta}_1} = \sqrt{\frac{1}{n-2}\left[\frac{\left(\sum y_i^2 - \left(\sum y_i\right)^2 \big/ n\right) - \hat{\beta}_1\left(\sum x_i y_i - \left(\sum x_i\right)\left(\sum y_i\right) \big/ n\right)}{\left(\sum x_i^2 - \left(\sum x_i\right)^2 \big/ n\right)}\right]} = \sqrt{\frac{1}{n-2}\left(\frac{JK(Y) - \hat{\beta}_1 JHK(XY)}{JK(X)}\right)}$$

Kriteria Penolakan: Tolak hipotesis nol jika $t_{hit} < -t_{\alpha/2;n-2}$ atau $t_{hit} > t_{\alpha/2;n-2}$

Menguji Koeffisien Regresi

Jika kita misalkan berikut ini adalah simpangan baku galat, yang dinotasikan dengan

$$s_\varepsilon = \sqrt{\left[\frac{\left(\sum y_i^2 - {\left(\sum y_i\right)^2}/{n}\right) - \hat{\beta}_1\left(\sum x_i y_i - {\left(\sum x_i\right)\left(\sum y_i\right)}/{n}\right)}{n-2}\right]} = \sqrt{\frac{JK(Y) - \hat{\beta}_1 JHK(XY)}{n-2}}$$

Maka simpangan baku bagi penduga slope β_1 dapat dituliskan sebagai berikut

$$s_{\hat{\beta}_1} = \sqrt{s_\varepsilon^2\left(\frac{1}{JK(X)}\right)} \qquad JK(X) = \sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}$$

Menguji Koeffisien Regresi

$$H_0 : \beta_0 = \beta_{10} \text{ vs } H_1 : \beta_0 \neq \beta_{00}$$

Statistik Uji

$$t_{hit} = \frac{\hat{\beta}_0 - \beta_{00}}{s_{\hat{\beta}_0}}$$

dimana

$$s_{\hat{\beta}_0} = \sqrt{s_\varepsilon^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{JK(X)} \right)}$$

Kriteria Penolakan: Tolak hipotesis nol jika $t_{hit} < -t_{\alpha/2;n-2}$ atau $t_{hit} > t_{\alpha/2;n-2}$

Nilai Dugaan dan Simpangan Bakunya

Apabila dilakukan sampling yang berulang untuk nilai $X = x$ tertentu dari salah satu nilai x yang kita gunakan, maka nilai dugaan modelnya adalah

$$\hat{y}_x = \hat{\beta}_o + \hat{\beta}_1 x$$

Dengan simpangan baku

$$s_{\hat{y}_{\bar{x}}} = \sqrt{s_\varepsilon^2\left(\frac{1}{n} + \frac{(x-\bar{x})^2}{JK(X)}\right)}$$

Nilai Dugaan dan Simpangan Bakunya

Apabila kasus baru didapat untuk nilai $X = x$ yaitu x dari nilai yang ada diluar amatan kita

$$\hat{y}_x = \hat{\beta}_o + \hat{\beta}_1 x$$

Dengan simpangan baku

$$s_{\hat{y}_{\tilde{x}}} = \sqrt{s_\varepsilon^2 \left(1 + \frac{1}{n} + \frac{(\tilde{x} - \bar{x})^2}{JK(X)}\right)}$$

Penduga Interval bagi Koeffisien Regresi

Selang Kepercayaan 100(1-α)% bagi β_1 adalah

$$\hat{\beta}_1 - t_{\frac{\alpha}{2};n-2} s_{\hat{\beta}_1} < \beta_1 < \hat{\beta}_1 + t_{\frac{\alpha}{2};n-2} s_{\hat{\beta}_1}$$

Selang Kepercayaan 100(1-α)% bagi β_0 adalah

$$\hat{\beta}_0 - t_{\frac{\alpha}{2};n-2} s_{\hat{\beta}_0} < \beta_0 < \hat{\beta}_0 + t_{\frac{\alpha}{2};n-2} s_{\hat{\beta}_0}$$

Koeffisien Korelasi

- **Mengukur keeratan hubungan dua peubah** (tidak harus memiliki hubungan sebab akibat). Dinotasikan dengan ρ_{xy} atau singkatnya ρ saja.
- Nilainya $-1 \leq \rho_{xy} \leq +1$
  - Jika $\rho_{xy} \rightarrow -1$ kedua peubah berhubungan kuat tapi berlawanan arah
  - Jika $\rho_{xy} \rightarrow +1$ kedua peubah berhubungan kuat dan searah
  - Jika $\rho_{xy} \rightarrow 0$ kedua peubah tidak memiliki hubungan
- Koeffisien korelasi contoh (bila tidak seluruh anggota populasi diamati) dinotasikan dengan r_{xy} atau r saja
- Tanda +/- dari koeffisien korelasi sama dengan tanda dari slope

Koeffisien Korelasi

$$r_{xy} = \frac{\sum_{i=1}^{n} x_i y_i - \frac{\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n}}{\sqrt{\left[\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}\right]\left[\sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}\right]}} = \frac{JHK(XY)}{\sqrt{JK(X)JK(Y)}}$$

$$r_{xy} = \frac{47,66}{\sqrt{(25,97)(90,46)}} = 0,983$$

Penjelasan

arti koeffisien korelasi

$$r_{xy} = \frac{47{,}66}{\sqrt{(25{,}97)(90{,}46)}} = 0{,}983$$

Dari data yang kita miliki terlihat bahwa terdapat hubungan yang cukup kuat antara luas tanah dan harganya. Karena tandanya +, maka semakin luas tanah, semakin tinggi harganya

Menguji Koeffisien Korelasi

$$H_0 : \rho = \rho_0 \;\; \text{vs} \;\; H_1 : \rho \neq \rho_0$$

Statistik uji

$$z_{hit} = \frac{\sqrt{n-3}}{2} \ln\left[\left(\frac{1+r}{1-r}\right)\left(\frac{1-\rho_0}{1+\rho_0}\right)\right]$$

Kriteria Penolakan Hipotesis Nol: Tolak Hipotesis Nol jika $z_{hit} < z_{\alpha/2}$ atau $z_{hit} > z_{1-\alpha/2}$

Menguji Koeffisien Korelasi

$$H_0 : \rho = 0 \quad \text{vs} \quad H_1 : \rho \neq 0$$

Statistik uji (n > 30)

$$z_{hit} = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

Kriteria Penolakan Hipotesis Nol: Tolak Hipotesis Nol jika $z_{hit} < z_{\alpha/2}$ atau $z_{hit} > z_{1-\alpha/2}$

Menguji Koeffisien Korelasi

$$H_0 : \rho = 0 \;\; \text{vs} \;\; H_1 : \rho \neq 0$$

Statistik uji ($n \leq 30$)

$$t_{hit} = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

Kriteria Penolakan Hipotesis Nol: Tolak Hipotesis Nol jika $t_{hit} < -t_{\alpha/2;n-2}$ atau $t_{hit} > t_{\alpha/2;n-2}$