

Analisa Regresi dan Korelasi Sederhana

Djayadi Nugroho, M.Kom
nugroho.stiemj.ac.id

Pendahuluan

- Dalam kondisi sehari-hari kita sering menjumpai adanya hubungan antara satu variabel dengan variabel lainnya. Sebagai contoh tingkat pendidikan seseorang berhubungan dengan gaji yang diperolehnya; dalam bidang pemasaran kita ketahui adanya hubungan antara volume penjualan dengan biaya advertensi dan lain-lain. Hubungan variabel diatas digambarkan adanya variabel bebas (X) dan tak bebas (Y).
- Hubungan antara dua atau lebih variabel tersebut ada dua macam, yaitu bentuk hubungan dan keeratan hubungan.
- Bila ingin diketahui bentuk hubungan, maka digunakan **analisis regresi**.
- Sedangkan bila yang ingin diketahui adalah keeratan hubungan, maka digunakan **analisis korelasi**.

- Analisis regresi adalah suatu proses melakukan estimasi untuk memperoleh suatu hubungan fungsional antara variabel acak Y dengan variabel X .
- Persamaan regresi digunakan untuk memprediksi nilai Y untuk nilai X tertentu.
- Analisis regresi sederhana adalah analisis regresi antara satu variabel Y dan satu variabel X .

- Alat lain untuk mempelajari hubungan antara dua variabel adalah analisis Korelasi.
- Analisis ini meliputi pengukuran arah dan kekuatan suatu hubungan linier antara dua variabel. Arah dan kekuatan hubungan ini dinyatakan dalam koefisien korelasi

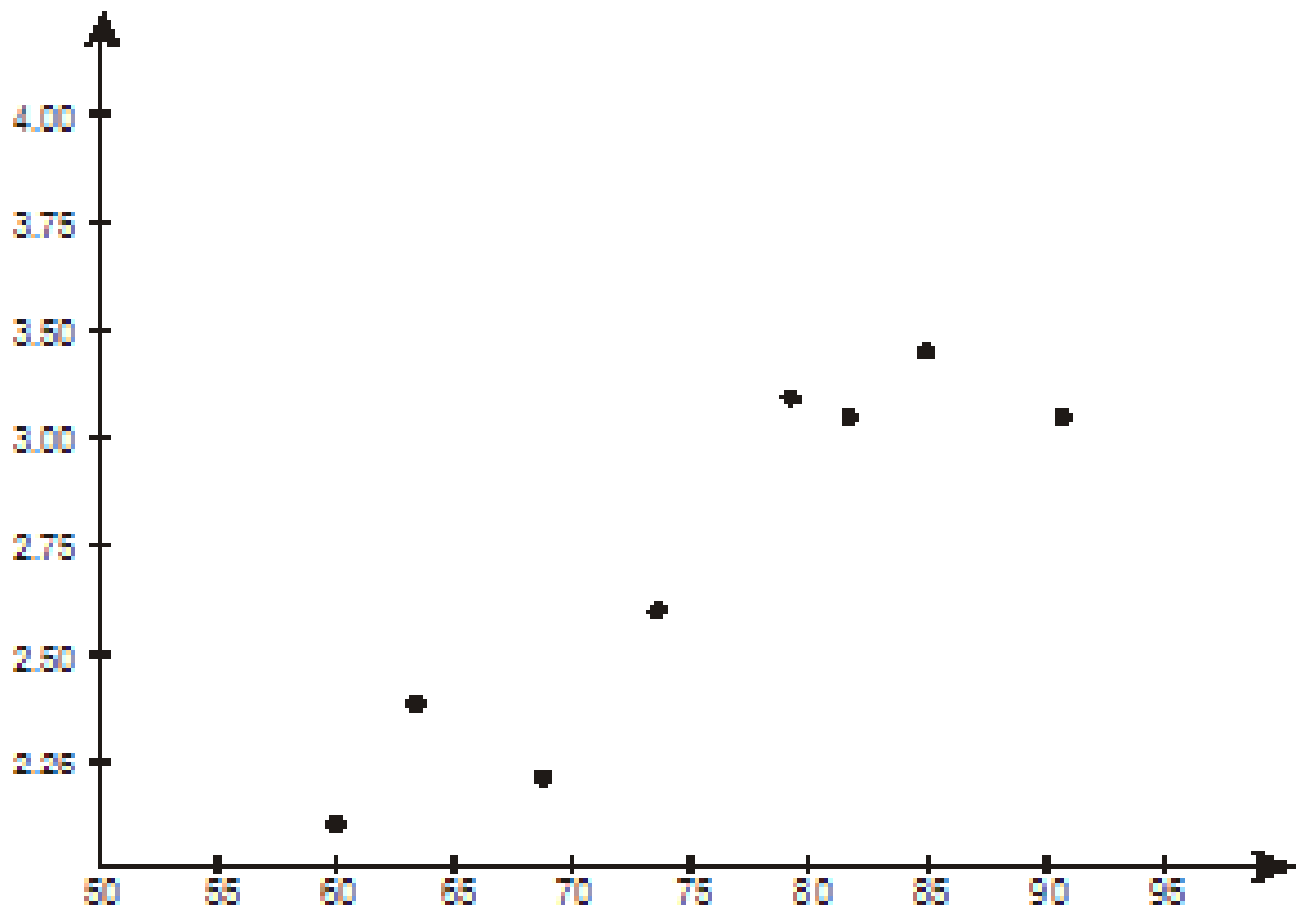
Scatter Diagram (diagram pencar)

- Scatter diagram adalah grafik yang menunjukkan titik-titik perpaduan nilai observasi dari 2 variabel (X & Y).
- Pada umumnya dalam grafik, variabel independent (X) diletakkan pada garis horisontal, sedangkan variabel dependent (Y) pada garis vertikal.
- Dari scatter diagram dapat diperoleh informasi tentang bentuk hubungan antara dua variabel X dan Y dengan melihat macam pola yang terbentuk.
- Selain memberikan informasi tentang bentuk hubungan dari kedua variabel, pola yang terbentuk juga dapat menggambarkan keeratan hubungan dari kedua variabel tersebut.

- Nilai test masuk dan IP Mahasiswa

Mahasiswa	A	B	C	D	E	F	G	H
Nilai test masuk (X)	74	69	85	63	82	60	79	91
IP (Y)	2.6	2.2	3.4	2.3	3.1	2.1	3.2	3.1

- Dari informasi tersebut jika nilai test masuk digunakan untuk memprediksikan keberhasilan studi mahasiswa, maka test masuk merupakan variabel independent, sedangkan IP sebagai variabel dependent.
- Bila dibuat plot atas pasangan nilai diatas, akan diperoleh scatter diagram berikut :



- Dari scatter diagram yang terbentuk dapat diberikan beberapa penjelasan sebagai berikut :
 1. Hubungan kedua variabel tersebut adalah positif karena peningkatan nilai X juga diikuti peningkatan nilai Y (searah)
 2. Derajat atau tingkat hubungan kedua variabel X dan Y sangat erat (titik-titik yang menunjukkan pertemuan nilai X dan Y mendekati garis lurus)
 3. Hubungan kedua variabel adalah linier, karena titik-titik yang menunjukkan pertemuan nilai X dan Y tersebut dapat menggambarkan garis lurus.
- Berdasarkan pola hubungan antara X dan Y yang diperoleh dari scatter diagram maka secara garis besar sifat hubungan antara variabel independent (X) dan variabel dependent (Y) dapat diklasifikasikan sebagai hubungan linier dan hubungan nonlinier.

Persamaan Regresi Linier

- $Y' = a + bX$
- $Y' =$ nilai Y prediksi $Y =$ Variabel terikat
- $a =$ nilai rata-rata Y prediksi jika $X = 0$
- $b =$ rata-rata perubahan pada Y jika X berubah 1 satuan
- $X =$ Variabel bebas
- Untuk menghitung koefisien a dan b pada persamaan diatas digunakan rumus :

- $$a = \frac{\sum Y - b \sum X}{n}$$

$$b = \frac{n \sum XY - \sum X \sum Y}{n(\sum X^2) - (\sum X)^2}$$

- Contoh :
- Berikut data hasil test karyawan dengan unit penjualan perminggu :

Salesman	Hasil Test (X)	Penjualan (Y)
A	4	5
B	7	12
C	3	4
D	6	8
E	10	11

- Tentukan persamaan regresi linier sederhana data diatas
- Hitunglah nilai penjualan, apabila salesman memiliki hasil test sebesar 8

X	Y	X ²	XY	Y ²
4	5	16	20	25
7	12	49	84	144
3	4	9	12	16
6	8	36	48	64
10	11	100	110	121
30	40	210	274	370

a. $Y' = a + bX$

$$b = \frac{n \sum XY - \sum X \sum Y}{n(\sum X^2) - (\sum X)^2} = \frac{5(274) - (30)(40)}{5(210) - (30)^2} = \frac{1370 - 1200}{1050 - 900} = 1.133$$

$$a = \frac{\sum Y - b \sum X}{n} = \frac{40 - (1.133)(30)}{5} = 1.202$$

$$\therefore Y = 1.202 + 1.133X$$

b. Jika $X = 8$ $Y = 1.202 + 1.133(8)$
 $= 1.202 + 9.1$
 $= 10.302 \approx 10$

- Contoh :
- Seorang pengusaha usaha transportasi ingin mengetahui hubungan antara umur kendaraan dengan biaya perawatannya. Setelah dilakukan pengamatan, diketahui hubungan antara umur kendaraan dengan biaya perawatan sebagai berikut :

Nomor Kendaraan	Umur Kendaraan (tahun) (X)	Biaya Reparasi (Rp. 000.000) (Y)
H 101 CC	5	3.1
H 104 CC	11	4
H 207 CC	4	3
H 532 CC	5	3.4
H 227 CC	3	2.5
H 438 CC	2	2

- Berdasarkan informasi tersebut diatas dapat dilakukan, estimasi garis regresi berdasarkan metode kuadrat terkecil sebagai berikut :

No.	Umur (X)	Biaya Perawatan (Y)	XY	X ²	Y ²
1	5	3.1	15.5	25	9.61
2	11	4	44.0	121	16
3	4	3	12.0	16	9
4	5	3.4	17.0	25	11.56
5	3	2.5	7.5	9	6.25
6	2	2	4	4	4
	ΣX = 30	ΣY = 18	ΣXY = 100	ΣX ² = 200	ΣY ² = 56.42

$$Y = a + bX$$

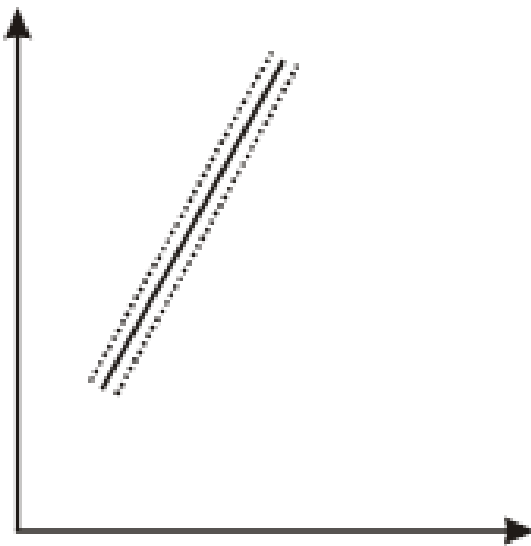
$$b = \frac{n \sum XY - \sum X \sum Y}{n(\sum X^2) - (\sum X)^2} = \frac{6(100) - (30)(18)}{6(200) - (30)^2} = \frac{600 - 540}{1200 - 900} = \frac{60}{300} = 0.2$$

$$a = \frac{\sum Y - b \sum X}{n} = \frac{18 - (0.2)(30)}{6} = \frac{18 - 6}{6} = 2$$

$$\therefore Y = 2 + 0.2X$$

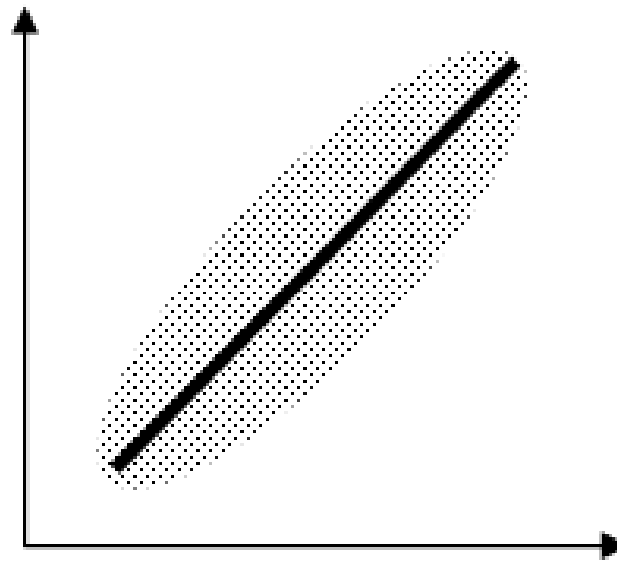
STANDARD ERROR ESTIMASI

- Proses selanjutnya dalam mempelajari analisis regresi adalah mengukur ketepatan persamaan estimasi. Ukuran ketepatan persamaan-persamaan estimasi tersebut disebut standard error estimasi yang dilambangkan dengan Se .
- Standard Error Estimasi adalah standard deviasi yang digunakan untuk mengukur penyebaran nilai observasi di sekitar garis regresi.
- Standard error estimasi, mendekati sama dengan standard deviasi, keduanya merupakan ukuran penyebaran. Standard deviasi digunakan untuk mengukur penyebaran dari kumpulan nilai observasi dengan bertitik tolak pada mean, sedangkan standard error estimasi bertitik tolak pada garis regresi.



(a)

Garis Regresi ini lebih tepat sebagai estimasi dari hubungan X dan Y



(b)

Garis Regresi ini kurang sebagai estimasi dari hubungan X dan Y

Formula dari Standard Error Estimasi (Se) adalah :

$$Se = \sqrt{\frac{\sum(Y - Y')^2}{n - 2}} \text{ atau}$$

$$Se = \sqrt{\frac{\sum Y^2 - a(\sum Y) - b(\sum XY)}{n - 2}}$$

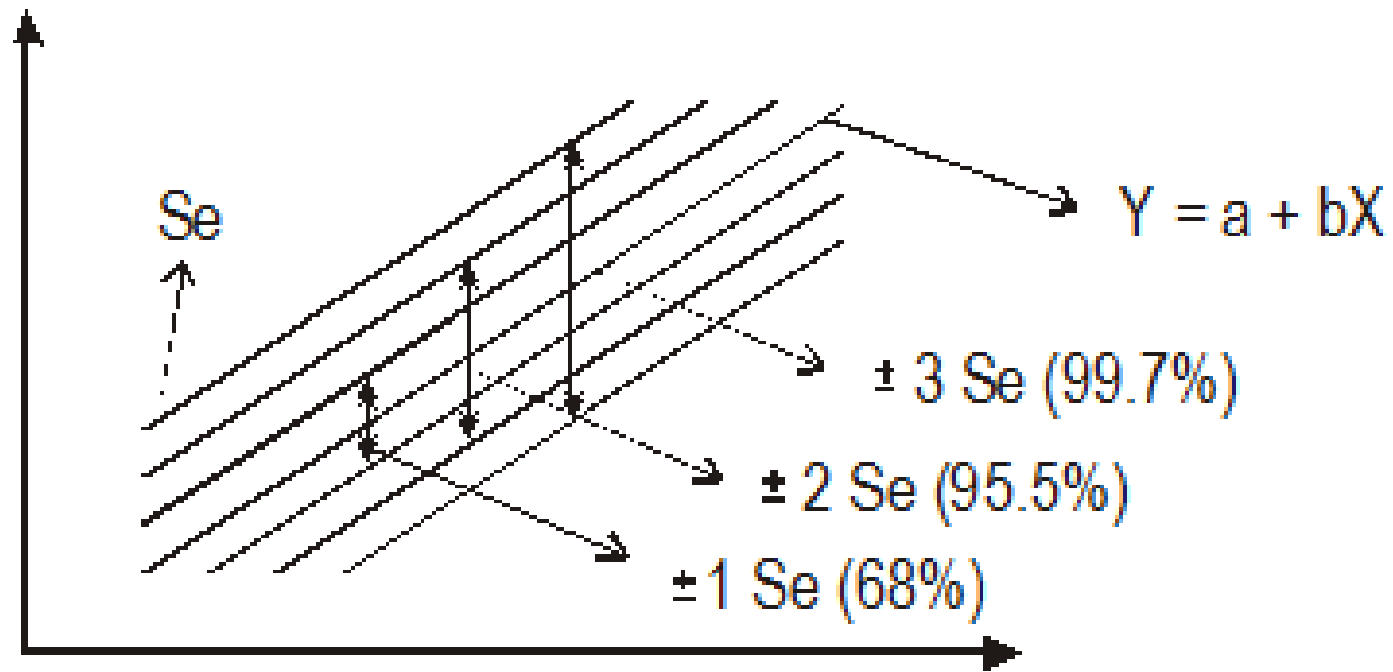
- Contoh :
- Kembali pada pemilik usaha angkutan yang berupaya mengadakan prediksi terhadap biaya perawatan tiap mobil dengan melihat masa pakainya, telah ditemukan persamaan estimasi :

$$Y' = 2 + 0.2X$$

$$Se = \sqrt{\frac{\sum Y^2 - a(\sum Y) - b(\sum XY)}{n-2}}$$

$$Se = \sqrt{\frac{56.42 - 2(18) - 0.2(100)}{6-2}} = \sqrt{\frac{56.42 - 36 - 20}{4}} = \sqrt{\frac{0.42}{4}} = 0.324$$

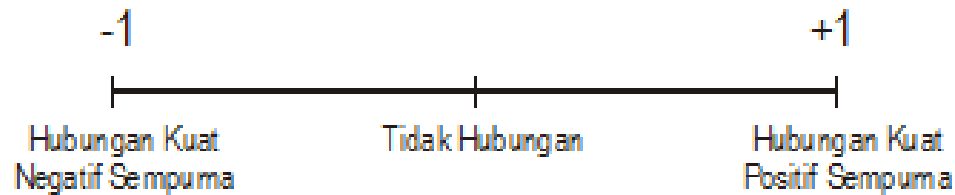
- Sebenarnya standard error estimasi dapat diinterpretasikan seperti halnya standard deviasi terhadap nilai mean.
- Semakin besar nilai Se, semakin tersebar nilai observasi yang berada di sekitar garis regresi atau sebaliknya semakin kecil nilai Se, maka penyebaran nilai observasi akan mendekati garis regresi.
- Apabila $Se = 0$ berarti tidak ada penyebaran atau semua nilai observasi terletak pada garis regresi sehingga garis regresi yang terbentuk dapat digunakan secara sempurna untuk mengadakan prediksi nilai variabel dependent.
- Dengan asumsi bahwa semua nilai observasi yang berada di sekitar garis regresi mengikuti distribusi normal maka :
 - 68% Nilai observasi berada dalam jarak $\pm 1 Se$
 - 95% Nilai observasi berada dalam jarak $\pm 2 Se$
 - 99.7% Nilai observasi berada dalam jarak $\pm 3 Se$



$\pm 1 Se$, $\pm 2 Se$, dan $\pm 3 Se$ terletak disekitar garis regresi

KOEFISIEN KORELASI LINIER SEDERHANA

- Bila analisis regresi berusaha memprediksi bentuk hubungan antara variabel Y dan X agar dapat memprediksi variabel Y untuk variabel X tertentu, analisis korelasi berusaha menghitung arah dan kekuatan hubungan antara variabel Y dan variabel X.
- Perbedaan utama regresi dengan korelasi adalah jika pada analisis regresi terdapat hubungan sebab akibat, pada analisis korelasi hubungan semacam ini tidak ada. Artinya korelasi antara Y dengan X akan sama dengan korelasi antara X dengan Y.
- Kekuatan dan arah hubungan antara 2 variabel diukur dengan koefisien korelasi. Koefisien korelasi bertanda + (positif) atau – (negatif), dengan angka yang berkisar dari -1 hingga $+1$.



- Semakin mendekati +1, koefisien korelasi menunjukkan adanya hubungan yang positif dan kuat. Koefisien korelasi yang mendekati -1 menunjukkan hubungan yang negatif dan kuat. Jika koefisien korelasi mendekati 0, memberikan indikasi bahwa ke 2 variabel tidak memiliki hubungan.
- Untuk mencari koefisien korelasi linier sederhana digunakan rumus sebagai berikut :

$$r = \frac{n \cdot \sum XY - \sum X \cdot \sum Y}{\sqrt{n \cdot \sum X^2 - (\sum X)^2} \sqrt{n \cdot \sum Y^2 - (\sum Y)^2}}$$

dimana :

n = jumlah pasangan data

X = variabel bebas

Y = variabel terikat

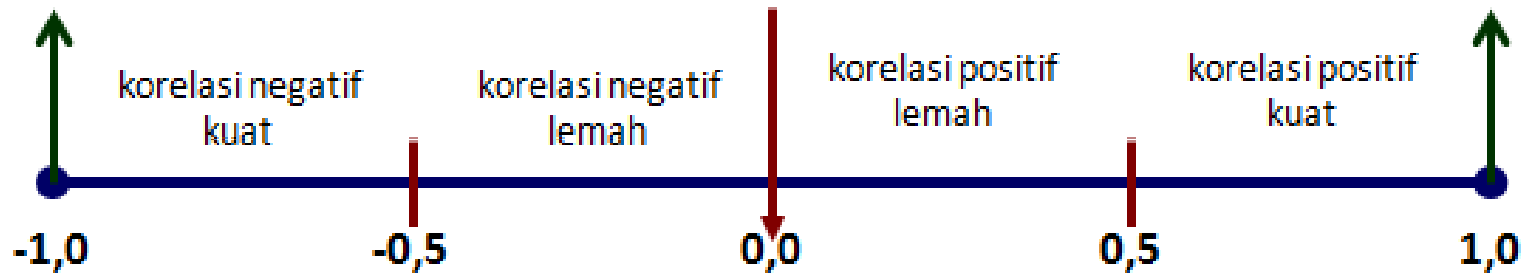
korelasi negatif sempurna

korelasi negatif sedang

tidak ada korelasi

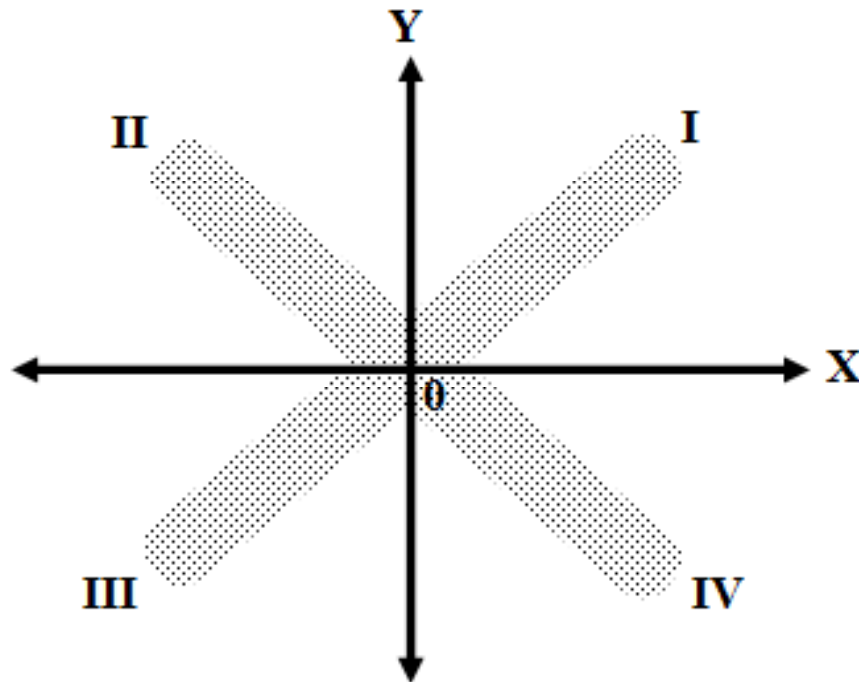
korelasi positif sedang

korelasi positif kuat



← korelasi negatif korelasi positif →

- Apa yang dimaksud dengan korelasi positif dan negatif? Jika 2 variabel berkorelasi positif, kenaikan variabel satu akan diikuti kenaikan variabel lain dan penurunan variabel satu diikuti dengan penurunan variabel lain.
- Sedangkan korelasi negatif menunjukkan jika satu variabel naik, variabel lain akan turun. Perhatikan gambar berikut :



- Kuandran I : Jika X naik, Y naik (korelasi positif)
- Kuandran III : Jika X turun, Y turun (korelasi positif)
- Kuandran II : Jika X turun, Y naik (korelasi negatif)
- Kuandran IV : Jika X naik, Y turun (korelasi negatif)

Contoh :

Mencari koefisien korelasi antara variabel penjualan dengan variabel hasil test.

Salinan	Hasil Test (X)	Penjualan (Y)	X ²	XY	Y ²
A	4	5	16	20	25
B	7	12	49	84	144
C	3	4	9	12	16
D	6	8	36	48	64
E	10	11	100	110	120
Σ	30	40	210	274	370

$$r = \frac{n \cdot \sum XY - \sum X \cdot \sum Y}{\sqrt{n \cdot \sum X^2 - (\sum X)^2} \cdot \sqrt{n \cdot \sum Y^2 - (\sum Y)^2}}$$

$$r = \frac{5(274) - (30)(40)}{\sqrt{5(210) - (30)^2} \cdot \sqrt{5(370) - (40)^2}} = 0.87$$

artinya antara hasil test dengan penjualan memiliki hubungan yang positif dan cukup kuat.

RUMUS UJI t UNTUK UJI KORELASI

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

- atau

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

di mana:

t : Nilai t-hitung

r : Nilai koefisien korelasi

n : Jumlah data pengamatan