2018

REGRESI DAN KORELASI SEDERHANA

Team Dosen
Universitas Esa Unggul

Contents

Pendahuluan

Penentuan teknik analisis yang akan digunakan dalam suatu analisis -- biasanya dibahas dalam metodologi penelitian -- sangat tergantung dari kerangka konseptual yang menjadi dasar analisis dan tujuan dari pelaksanaan analisis. Agar hasil analisis dapat ditafsirkan secara tepat, maka harus dihindari penggunaan teknik analisis yang tidak sesuai dengan skala pengukuran dari variabel yang digunakan dalam analisis.

Dalam mengkaji hubungan antara dua variabel atau lebih, penting dipahami skala pengukuran dari setiap variabel tersebut, sehingga teknik analisis yang akan digunakan untuk menjelaskan hubungan atau kaitan antara variabel tersebut dapat dipilih yang paling sesuai. Apabila hal ini diabaikan, maka dapat menghasilkan interpretasi hasil yang kurang tepat (*misleading*).

Sebagai contoh, analisis regresi mensyaratkan bahwa variabel yang digunakan paling tidak diukur secara interval, sehingga kalau ada salah satu variabel yang diukur secara nominal atau ordinal maka hasil yang diperoleh tidak dapat ditafsirkan secara tepat, oleh karena asumsi dalam penggunaan teknik ini yang tidak dapat dilanggar adalah bahwa variabel yang digunakan mengikuti distribusi normal.

Pada modul ini teknik yang dibahas merupakan teknik analisis statistik inferensial, yaitu teknik yang mendasarkan pada pemanfaatan data yang diperoleh dari suatu sampel acak, sehingga hasilnya merupakan gambaran keadaan populasi dari mana sampel acak tersebut diambil. Teknik statistik semacam ini memberikan jaminan bahwa kesimpulan dan penafsiran dibuat dengan tingkat kesalahan yang rendah,[5] biasanya dipakai 0,05 (5 %).

Teknik analisis statistik yang dibahas dalam modul ini difokuskan hanya pada teknik yang dapat menjelaskan hubungan atau kaitan antara dua variabel (*bivariate*). Teknik analisis[6] statistik yang dibahas meliputi Analisis Regresi. 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.

Contoh lainnya, 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.

---

[5] Tingkat kesalahan tersebut dalam teori probabilita dan uji hipotesis merupakan besarnya probabilita menolak hipotesis benar, atau disebut kesalahan tipe-1. Istilah umum yang dipakai p-value, significancy of statistic (T, Z, χ^2, atau F), atau Pr>$|T|$, Pr>$|Z|$, Pr > χ^2, Pr > F tergantung statistik uji yang digunakan.

[6] Para pembaca dianggap sudah memahami prinsip dasar dari Analysis of Variance (ANOVA), karena semua uji statistik pada umumnya didasarkan pada dekomposisi dari total variance atas berbagai faktor penyebab perbedaan.

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. Sedangkan analisis regresi sederhana adalah analisis regresi antara satu variabel Y dan satu variabel X.

Pada materi ini kita hanya membahas persamaan regresi sederhana linier. 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

Bila dua variabel X dan Y berhubungan sebab akibat, dengan variabel X sebagai variabel independent (variabel bebas, variabel yang nilainya mempengaruhi nilai variabel tak bebas) dan variabel Y sebagai variabel dependent (variabel tak bebas, variabel yang nilainya dipengaruhi oleh variabel bebas), maka bila nilai variabel X diketahui, nilai tersebut dapat dipergunakan untuk memperkirakan nilai variabel Y jika bentuk hubungan kedua variabel tersebut diketahui.

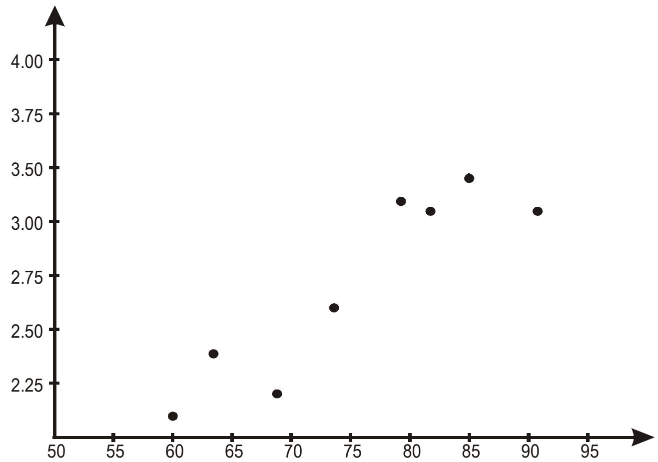
Untuk mengetahui pola hubungan yang mungkin terbentuk dari dua variabel X dan Y dapat dipergunakan Scatter diagram (diagram pencar). Scatter diagram adalah grafik yang menunjukkan titik-titik perpaduan nilai observasi dari 2 variabel (X & Y). Pada umumnya dalam grafik, variabel independent (X) diletakkan 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.

Tabel 1. 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 memrpediksikan 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. Sifat hubungan yang nonlinier (curva linier) justru banyak terjadi dalam masalah ekonomi, meskipun demikian pembahasan pada bab ini dibatasi hanya untuk hubungan yang linier.

Persamaan Regresi Linier

Analisis regresi merupakan alat yang dapat memberikan penjelasan hubungan antara dua jenis variabel yaitu hubungan antara variabel dependen atau variabel kriteria dengan variabel independen atau variabel prediktor[7]. Analisis hubungan antara dua variabel disebut sebagai

[7] Analisis regresi dapat dipandang sebagai alat yang menjelaskan hubungan linear dari satu variabel ke variabel lainnya didekomposisikan dan diuraikan (*descriptive tool*), dan sebagai alat untuk menginduksikan (*to infer*) hubungan antar variabel pada populasi. Namun demikian keduanya dipandang dari segi teknik statistik sangat erat kaitannya, tetapi keduanya dapat dipandang terpisah paling tidak pada tingkat konseptual. Lebih jelasnya lihat Kim, J. and F.J. Kohout dalam Nie et. al Statistical Package for Social Sciences. Second edition. pp 321-322. New York: McGraw Hill Book Company 1975.

analisis regresi sederhana jika hanya melibatkan satu variabel independen. Analisis disebut sebagai analisis regresi berganda jika melibatkan lebih dari satu variabel independen.

Hubungan antara variabel dependen (Y) dengan variabel independen (X) dituliskan dalam model linier umum:

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \ldots\ldots + \hat{\beta}_p X_p + e, \qquad e \sim N(0, \sigma^2), \quad \textit{Residual} \quad e = Y - \hat{Y}$$

di mana $\hat{\beta}_i$, i = 1,2,........p adalah koefisien regresi yang berarti besarnya perubahan pada $\hat{Y}$, jika X_i bertambah satu satuan dan variabel yang lain konstan, $\hat{\beta}_0$ adalah *intercept*. Residual e mengikuti distribusi normal dengan rata-rata 0 dan varians konstan sebesar σ^2. Asumsi dasar dalam analisis regresi adalah (i) setiap Y yang merupakan kombinasi linier atas X dan mengikuti distribusi normal, (ii) e tersebar secara acak dan tidak berpola mengikuti besarnya nilai X, (iii) tidak terdapat hubungan (korelasi) yang tinggi antar variabel X.

Analisis regresi sederhana hanya melibatkan satu variabel independen X, sehingga dalam persamaannya p=1, sehingga model liniernya adalah:

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X + e, \qquad e \sim N(0, \sigma^2), \quad \textit{Residual} \quad e = Y - \hat{Y}$$

Dengan model seperti pada persamaan diatas maka hipotesis yang diajukan untuk diuji adalah H_0: $\beta_1 = 0$ terhadap H_1: $\beta_1 \neq 0$. Untuk menolak H_0 harus dapat dibuktikan secara empirik bahwa $\beta_1 \neq 0$ atau β_1 bermakna (*significant*) atau dengan kata lain ada hubungan linier regresi antara Y dan X seperti pada persamaan kedua. Dalam literature lain persamaan regresi juga ditulis dalam bentuk:

Y’ = a + bX

Dimana:

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:

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

Contoh:

Berikut data hasil test karyawan dengan unit penjualan perminggu:

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

a) Tentukan persamaan regresi linier sederhana data diatas
b) Hitunglah nilai penjualan, apabila Salesperson memiliki hasil test sebesar 8

Jawab:

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 + b X

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

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

$\therefore$ Y = 1.202 + 1.133

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:

Biaya Perawatran Kendaraan

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
	$\Sigma X = 30$	$\Sigma Y = 18$	$\Sigma XY = 100$	$\Sigma X^2 = 200$	$\Sigma Y^2 = 56.42$

Y = a + bX

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

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

$\therefore$ Y = 2 + 0.2X

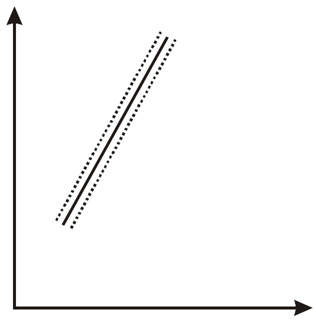
Jika X = 8

Y = 2 + 0.2 (8) = 3.6

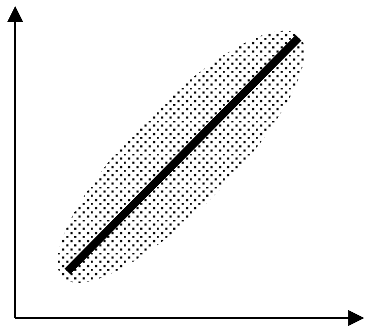
Standard Error Estimasi

Proses selanjutnya dalam mempelajari analisis regresi adalah mengukur ketepatan persamaan estimasi. Ukuran keteparan 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 pregresi.



(a)
Garis Regresi ini lebih tepat sebagai estimatr dari hubungan X dan Y



(b)
Garis Regresi ini kurang sebagai estimatr dari hubungan X dan Y

Formula dari Standard Error Estimasi (Se) adalah:

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

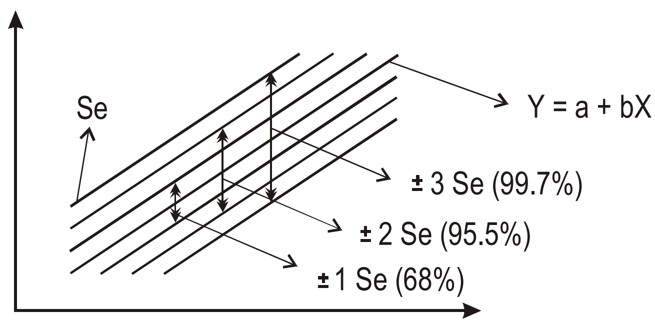
Dengan asumsi bahwa semua nilai observasi yang berada di sekitar garis regresi mengikuti distribusi normal maka:

68% Nilai observasi berada dalam jarak ± 1 Se

95% Nilai observasi berada dalam jarak ± 2 Se

99.7% Nilai observasi berada dalam jarak ± 3 Se

± 1 Se, ± 2 Se, dan ± 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 meghitung 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 poasitif 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:

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

dimana:

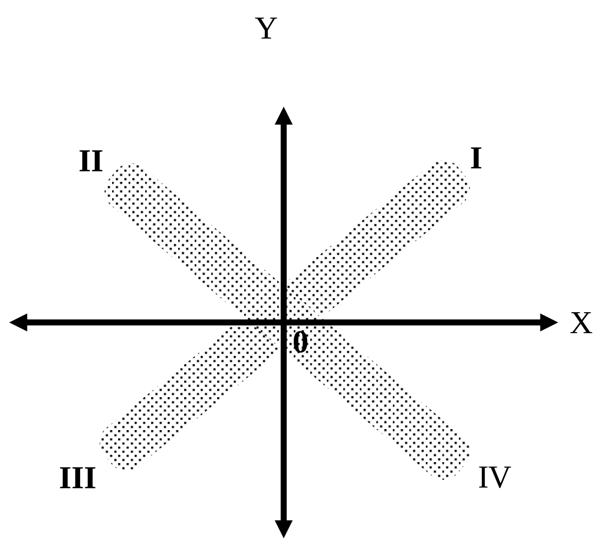
n = jumlah pasangan data

X = variabel bebas

Y = variabel terikat

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:

II	**I**
0	+
III	IV



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

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

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

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

Pendugaan dan Pengujian Parameter A, B dan Koefisien Korelasi

Persamaan regresi dan koefisien korelasi pada umumnya dihitung dengan menggunakan data sampel. Persamaan regresi sampel ini digunakan untuk menduga persamaan regresi populasi yang tidak diketahui. Untuk mengetahui apakah persamaan regresi sampel harus diuji validitasnya terlebih dahulu. Hal yang sama berlaku untuk koefisien korelasi.

Pendugaan dan pengujian A

Pendugaan dan pengujian hipotesis parameter A dapat menggunakan distribusi t.

$t = \frac{a - Ao}{Sa}$ dengan df = n – 2

dengan interval keyakinannya sebesar 95%, pendugaan parameter A dapat diberikan secara umum sebagai:

P (a – t(½ α, df) Sa < A < a + t(½ α, df) Sa) = 1 – α

P (a – t(0,025; n - 2) Sa < A < a + t(0,025; n - 2) = 0,95

Contoh:

Uji dan dugalah A dengan menggunakan penduga a hasil persamaan Y = 29,993 + 0,0566 X. dalam soal persamaan diketahui sebagai berikut:

N = 11

$\bar{X}$ = 1.223,9

$\bar{Y}$ = 99,27

Σ X2 = 26.855.704,50

Σ Y2 = 142.467,16

Σ XY = 1.923.540,133

Σ Y = 1.091,97

Σ (X – $\bar{X}$)2 = 10.378.461,18

a = 29,9973

b = 0,0566

Prosedur pendugaan A:

α = 1 – 0,95 = 5% = 0,05

½ α = 0,025 ➔ t (½ α ; df) ➔ t (0,025, 9) = 2,262

P (a – t(0,025; 11 – 2) Sa < A < a + t (0,025; 11 – 2) Sa) = 0,95.

$$S_a^2 = \frac{\frac{1}{n-2}(\sum Y^2 - a(\sum Y) - b(\sum XY)(\sum X^2)}{n\sum(X-\overline{X})^2}$$

$$= \frac{\frac{(142.467{,}16 - 29{,}997(1.091{,}97) - 0{,}0566(1.923.540{,}133)}{11-2} x(26.855.704{,}50)}{11(10.378.461{,}18)}$$

$$= \frac{\frac{(142.467{,}16 - 32.755{,}8241 - 108.872{,}372}{9} x(26.855.704{,}50)}{114.163.073}$$

$$= \frac{93{,}22(26.855.704{,}50)}{114.163.073} = 21{,}929$$

$$Sa = \sqrt{21{,}929} = 4{,}682$$

P (29,997 – 2,262 (4,682) < A < 29,997 + 2,262 (4,682)) = 0,95

P (29,997 – 10,591 < A < 29,997 + 10,591) = 0,95

P (19,406 < A < 40,588) = 0,95

Jelas, batas atas pendugaan A ialah 40,588 dan batas bawahnya menjadi 19,405.

Prosedur pengujian Ao = 0, langkah demi langkah adalah sebagai berikut:

1. $H_0 : A = 0$ $H_1 : A \neq 0$
2. $\alpha = 0{,}05$
3. Statistik uji ➔ $t = \frac{a - Ao}{Sa} = \frac{29{,}997 - 0}{4{,}682} = 6{,}4069$
4. Daerah kritis, karena pengujian terhadap 2 arah, maka daerah kritisnya adalah ➔ t (½ α ; n-2) ➔ t (0,025; 9) = 2,262
5. Kesimpulan
   Karena daerah kritis lebih kecil dari statistik uji (2,262 < 6,4069), maka H_0 ditolak artinya bahwa parameter A tidak sama dengan nol

Pendugaan dan Pengujian B

Pendugaan dan pengujian parameter B prosesnya juga sama dengan parameter A, termasuk pengujiannya juga dapat menggunakan distribusi t.

$t = \frac{b - Bo}{Sb}$ dengan df = n – 2

dengan interval keyakinan sebesar tertentu, penduganya parameter B dapat diberikan secara umum sebagai:

P (b – t(½ α, df) Sb < B < b + t(½ α, df) Sb) = 1 – α

$$S_b^2 = \frac{\frac{1}{n-2}(\sum Y^2 - a(\sum Y) - b(\sum XY)}{\sum (X - \overline{X})^2}$$

Contoh :

Uji dan dugalah B dengan interval keyakinan 95% dengan menggunakan penduga b hasil persamaan Y = 29,9973 + 0,0566 X

Dalam soal persamaan tersebut diketahui data sebagai berikut:

N = 11	$\overline{X}$ = 1.223,9	$\overline{Y}$ = 99,27
Σ X2 = 26.855.704,50		Σ Y2 = 1.091,97
Σ XY = 1.923.540,133		Σ Y2 = 142.467,16
Σ (X – $\overline{X}$)2 = 10.378.461,18		b = 0,0566

Prosedur pendugaan B:

α = 1 – 0,95 = 5% = 0,05

½ α = 0,025 ➔ t (½ α ; df) ➔ t (0,025, 9) = 2,262

$$S_b^2 = \frac{\frac{1}{n-2}(\sum Y^2 - a(\sum Y) - b(\sum XY)}{\sum (X - \overline{X})^2}$$

$$= \frac{\frac{(142.467{,}16 - 29{,}997(1.091{,}97) - 0{,}0566(1.923.540{,}133)}{11-2}}{10.378.461{,}18}$$

$$= \frac{93,22}{10.378.461,18} = 0,00000898$$

$$Sb = \sqrt{0,00000898} = 0,003$$

P (b – t(½ α, df) Sb < B < b + t(½ α, df) Sb) = 1 – α

P (0,0566 – t (0,025; 11-2)0,003 < B < 0,0566 + t (0,025; 11-2)) = 1 – α

P (0,0566 – 0,0068 < B < 0,0566 + 0,0068 = 0,95

P (0,049814 < B < 0,063414) = 0,95

Prosedur pengujian B0 = 0, langkah demi langkah adalah sebagai berikut:

1. H 0 : B = 0 H1 : B > 0
2. α = 0,05
3. Statistik uji secara searah dengan df = n – 2

   $$t = \frac{b - Bo}{Sb} = \frac{0,0566 - 0 - 0}{0,003} = 18,866$$

4. Daerah kritis, ➔ t (α ; n-2) ➔ t (0,05; 9) = 1,833
5. Kesimpulan

   Karena statistik uji lebih besar dari daerah kritis, H_0 ditolak (H1 diterima)

Sebagian besar statistisi melakukan pengujian terhadap parameter B dengan menggunakan statistik uji F:

$F = \dfrac{b^2 \sum (X - \overline{X})^2}{S^2_{Y.X}}$ dengan df bagi pembilangnya sebesar V1 = 1 dan df penyebutnya V2 = n – 2.

Pengujian Koefisien Korelasi Populasi: ρ (Baca 'Rho')

Langkah-langkah:

1. Menentukan H0 dan H1

H0 : $\rho = 0$

H1 : $\rho \neq 0$

2. Menentukan daerah penerimaan H0 dengan menggunakan statistik t dengan df = n – 2

   n = jumlah pasangan observasi

   H0 diterima jika statistik uji berada diantara –t (½ α; df) dan t (½ α; df)

   Jika tidak, H0 ditolak dan H1 diterima

3. Menghitung nilai statistik uji

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

4. Membandingkan statistik uji dengan t tabel untuk mengambil kesimpulan.

<u>Kasus soal</u>

Ada keyakinan bahwa antara penjualan dan keuntungan ada hubungannya. Berikut adalah nilai penjualan dan keuntungan dari 11 pedagang yang ada di Jl. Mawar Kota XYZ pada tahun 20xx.

Pedagang	Penjualan (juta)	Keuntungan (juta)
Hermanto	89,2	4,9
Usman	18,6	4,4
Marwan	18,2	1,3
Syamsul	71,7	8,0
Kadam	58,6	6,6
Herman	46,8	4,1
Iskandar	17,5	2,6
Hamid	11,9	1,7
Agus	19,6	3,5
Bahrul Alam	51,2	8,2
Suyanto	28,6	6,1

Diminta:

a. tentukan persamaan regresi linier pengaruh penjualan terhadap keuntungan! Dan tentukan nilai prediksi keuntungan jika penjualan sebesar 45
b. Tentukan nilai koefisien korelasinya. Dan berikan interpretasinya?
c. Ujilah koefisien korelasi dengan menggunakan alfa 5%