REGRESI LINEAR SEDERHANA

MODUL 1

Dra. Sri Pangesti, S.U.

PENDAHULUAN

Analisis regresi merupakan analisis statistik yang mempelajari hubungan antara dua variabel atau lebih. Dalam analisis regresi linear diasumsikan berlakunya bentuk hubungan linear dalam parameter. Modul regresi linear yang paling sederhana adalah regresi linear dengan satu variabel bebas (*independent variable*).

Pokok bahasan dalam modul ini terdiri atas dua kegiatan belajar, *pertama*, tentang regresi linear dengan satu variabel bebas dan *kedua*, tentang inferensi dalam analisis regresi.

Pada Kegiatan Belajar 1, Anda akan mempelajari penaksiran (estimasi) fungsi regresi dan variansi suku sesatan $\left(\sigma^2\right)$ dengan distribusi suku sesatan belum ditentukan (diasumsikan). Pada Kegiatan Belajar 2, Anda akan mempelajari inferensi pada model regresi dengan suku sesatan berdistribusi normal.

Setelah mempelajari modul ini, secara umum Anda diharapkan dapat memahami dasar-dasar pemikiran dalam analisis regresi dan dapat melakukan inferensi model regresi linear sederhana secara tepat.

Secara khusus, Anda diharapkan dapat:

1. menentukan penaksir parameter model regresi dengan metode kuadrat terkecil,
2. menentukan penaksir parameter model regresi dengan metode maksimum likelihood,
3. menentukan selang kepercayaan untuk parameter model regresi,
4. menghitung koefisien korelasi dan determinasi,
5. melakukan analisis regresi dengan pendekatan analisis variansi.

Regresi Linear dengan Satu Variabel Bebas

Analisis regresi adalah analisis statistik yang mempelajari hubungan antara dua atau lebih variabel kuantitatif sehingga satu variabel dapat diramalkan (*predicted*) dari variabel lainnya.

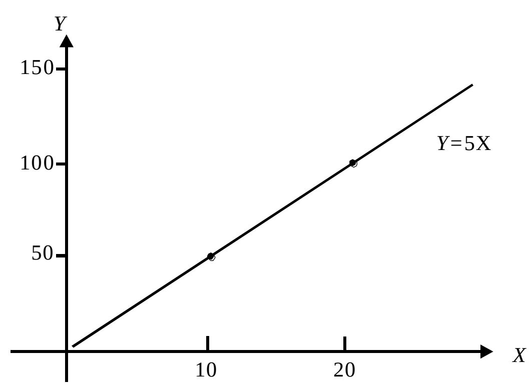
Hubungan antara dua variabel dapat dibedakan menjadi dua, yaitu *hubungan fungsional* dan *hubungan statistik*. Hubungan fungsional antara dua variabel dapat dinyatakan secara matematis; jika X variabel bebas (*indenpendent variable*) dan Y variabel tak bebas (*dependent variable*), hubungan fungsional ditulis dalam bentuk

$$Y = f\left(X\right)$$

Jika diketahui nilai X tertentu, fungsi f akan memberikan nilai Y yang bersesuaian.

Contoh 1.1

Misalkan Y = total nilai penjualan suatu produk (dalam ribuan rupiah) dan X = jumlah unit produk yang terjual. Jika harga jual Rp 5.000,00 per unit produk maka hubungan antara Y dan X dapat dinyatakan dalam Gambar 1.1.

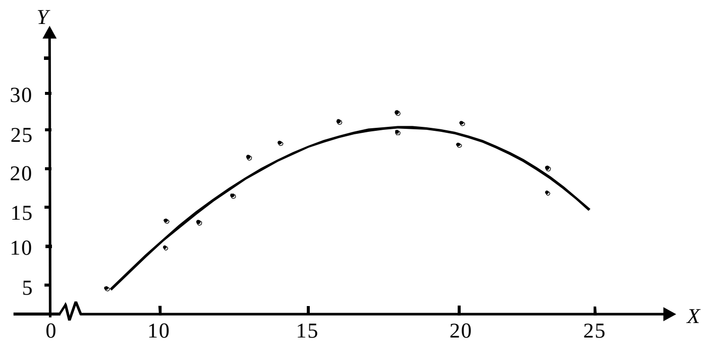


Gambar 1.1.

Hubungan statistik antara dua variabel tidak sempurna. Pada umumnya pada hubungan statistik, pengamatan tidak tepat jatuh pada kurve hubungan.

Contoh 1.2

Jika hubungan antara X yang menyatakan usia (dalam tahun) dan Y yang menyatakan tingkat steroid dinyatakan dalam Gambar 1.2.



Gambar 1.2.

Model regresi secara formal menyatakan dua hal tentang hubungan statistik, yaitu: kecenderungan variabel tak bebas (Y) berubah-ubah terhadap variabel bebas (X) dalam bentuk yang sistematik dan tersebarnya titik-titik di sekitar kurve hubungan statistik. Kedua hal tersebut dinyatakan dalam suatu model regresi yaitu: untuk setiap nilai X terdapat distribusi probabilitas dari Y dan mean dari distribusi probabilitas Y berubah-ubah secara sistematik terhadap perubahan nilai X.

Model Regresi Linear Sederhana

Model regresi dengan satu variabel bebas X dapat ditulis dalam bentuk

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \qquad , i = 1, 2, \ldots n$$

dengan

Y_i : nilai variabel tak bebas dalam trial ke-i,

β_0, β_1 : parameter,

X_i : konstanta yang diketahui nilainya, yakni nilai variabel bebas dalam trial ke-*i*,

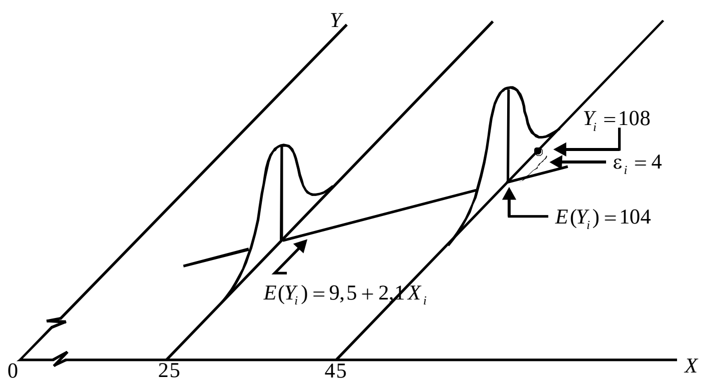
ε_i : suku sesatan random dengan $E(\varepsilon_i) = 0, \sigma(\varepsilon_i, \varepsilon_j) = \sigma^2$, ε_i dan ε_j tidak berkorelasi, kovariansi $\sigma(\varepsilon_i, \varepsilon_j) = 0$ untuk semua $i, j, i \neq j$

Sifat-sifat dari Model Regresi

1. Y_i merupakan jumlah dari dua komponen, yaitu suku konstan $\beta_0 + \beta_1 X_i$ dan suku random ε_i.
2. Karena $E(\varepsilon_i) = 0$ maka $E(Y_i) = E(\beta_0 + \beta_1 X_i + \varepsilon_i) = \beta_0 + \beta_1 X_i$. Hal ini berarti distribusi dari Y_i pada tingkat X dalam trial ke-*i* mempunyai mean $E(Y_i) = \beta_0 + \beta_1 X_i$
3. Nilai pengamatan Y pada trial ke-*i* jatuh pada jarak ε_i dari nilai fungsi regresinya $(E(Y_i))$ atau $Y_i - E(Y_i) = \varepsilon_i$.
4. Suku sesatan (ε_i) diasumsikan mempunyai variansi konstan (σ^2). Oleh karena itu, $\sigma^2(Y_i) = \sigma^2$.
5. Suku sesatan diasumsikan tidak berkorelasi. Karena ε_i dan ε_j tidak berkorelasi untuk $i \neq j$, maka Y_i dan Y_j tidak berkorelasi.

Contoh 1.3

Diketahui model regresi $Y_i = 9,5 + 2,1X + \varepsilon_i$. Misalkan dalam trial ke-*i* diperoleh nilai-nilai $X_i = 45$, $Y_i = 108$, dan suku sesatan $\varepsilon_i = 4$ maka $E(Y_i) = 9,5 + 2,1(45) = 104$ dan $Y_i = 104 + 4 = 108$. Visual model regresi tersebut dapat dilihat pada Gambar 1.3 berikut.



Gambar 1.3.

Model regresi linear sederhana $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ dapat ditulis dalam bentuk lain $Y_i = \beta_0 X_0 + \beta_1 X_i + \varepsilon_i$ dengan $X_0 = 1$. Model ini dapat kita ubah menjadi:

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 (X_i - \bar{X}) + \beta_1 \bar{X} + \varepsilon \\ &= (\beta_0 + \beta_1 \bar{X}) + \beta_1 (X_i - \bar{X}) + \varepsilon \\ &= \beta_0^* + \beta_1 (X_i - \bar{X}) + \varepsilon_i \end{aligned}$$

Jadi model regresi alternatif $Y_i = \beta_0^* + \beta_1 (X_i - \bar{X}) + \varepsilon_i$ dengan $\beta_0^* = \beta_0 + \beta_1 \bar{X}$

Penaksiran Fungsi Regresi

Untuk mendapatkan penaksir yang *baik* bagi parameter regresi $(\beta_0 \text{ dan } \beta_1)$ dapat digunakan metode kuadrat terkecil. Untuk setiap pasangan pengamatan $(X_i, Y_i),\ i = 1, 2, \ldots n$ pandang nilai simpangan Y_i terhadap $E(Y_i)$, yaitu:

$$Y_i - E(Y_i) = Y_i - (\beta_0 + \beta_0 X_i)$$

Jumlah kuadrat dari *n* simpangan ini ditulis dengan notasi *Q*, yakni

$$Q = \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_i)^2$$

Penaksir b_0 dan b_1 diperoleh dengan meminimumkan Q, yaitu dengan mendeferensialkan *Q* terhadap β_0 dan β_1.

$$\frac{\partial Q}{\partial \beta_0} = -2\sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_i)$$

$$\frac{\partial Q}{\partial \beta_1} = -2\sum_{i=1}^{n} X_i (Y_i - \beta_0 - \beta_1 X_i)$$

Selanjutnya masing-masing persamaan kita samakan dengan nol serta mengganti β_0 dengan b_0 dan β_1 dengan b_1.

$$\sum_{i=1}^{n} Y_i - nb_0 - b_1 \sum_{i=1}^{n} X_i = 0$$

$$\sum_{i=1}^{n} X_i Y_i - b_0 \sum_{i=1}^{n} X_i - b_1 \sum_{i=1}^{n} X_i^2 = 0$$

Kedua persamaan ini kita sebut sebagai persamaan normal. Penyelesaian persamaan normal untuk b_1 dan b_0 adalah

$$b_1 = \frac{\sum_{i=1}^{n} X_i Y_i - \dfrac{\sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{n}}{\sum_{i=1}^{n} X_i^2 - \dfrac{\left(\sum_{i=1}^{n} X_i\right)^2}{n}} = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n} (X_i - \bar{X})^2}$$

$$b_0 = \frac{1}{n}\left(\sum_{i=1}^{n} Y_i - b_1 \sum_{i=1}^{n} X_i\right) = \bar{Y} - b_1\bar{X}$$

Notasi-notasi berikut dapat dipergunakan untuk menyederhanakan penulisan rumus:

$$S_{YX} = \sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right) = \sum_{i=1}^{n}\left(X_i - \bar{X}\right)Y_i = \sum_{i=1}^{n} X_i\left(Y_i - \bar{Y}\right)$$

$$= \sum_{i=1}^{n} X_i Y_i - \frac{\sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{n} = \sum_{i=1}^{n} X_i Y_i - n\ \bar{X}\ \bar{Y}$$

$$S_{XX} = \sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 = \sum_{i=1}^{n}\left(X_i - \bar{X}\right)X_i$$

$$= \sum_{i=1}^{n} X_i^2 - \frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n} = \sum_{i=1}^{n} X_i^2 - n\ \bar{X}^2$$

$$S_{YY} = \sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2 = \sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)Y_i$$

$$= \sum_{i=1}^{n} Y_i^2 - \frac{\left(\sum_{i=1}^{n} Y_i\right)^2}{n} = \sum_{i=1}^{n} Y_i^2 - n\ \bar{Y}^2$$

Sehingga rumus untuk b_1 dapat ditulis dalam bentuk lebih sederhana $b_1 = \frac{S_{XY}}{S_{XX}}$ dan persamaan regresi taksiran $\hat{Y}_i = \bar{Y} + b_1\left(X_i - \bar{X}\right)$.

Contoh 1.4

Data berikut menunjukkan jumlah unit yang diproduksi $\left(X\right)$ dan jumlah jam kerja karyawan $\left(Y\right)$.

X	30	20	60	80	40	50	60	30	70	60
Y	73	50	128	170	87	108	135	69	148	132

Dari data dihitung $\sum_{i=1}^{n} X_i = 500$; $\sum_{i=1}^{n} X_i^2 = 28.400$; $\sum_{i=1}^{n} X_i Y_i = 61.800$; $\sum_{i=1}^{n} Y_i = 1.100$; $\sum_{i=1}^{n} Y_i^2 = 134.660$; $n = 10$; $\bar{X} = 50$; $\bar{Y} = 110$. Dengan menggunakan rumus yang ada, diperoleh:

$$b_1 = \frac{\sum_{i=1}^{n} X_i Y_i - \frac{\sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{n}}{\sum_{i=1}^{n} X_i^2 - \frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n}} = \frac{61.800 - \frac{500(1.100)}{10}}{28.400 - \frac{(500)^2}{10}} = 2$$

$$b_0 = \frac{1}{n}\left(\sum_{i=1}^{n} Y_i - b_1 \sum_{i=1}^{n} X_i\right) = \frac{1}{10}\left(1.100 - 2(500)\right) = 10$$

Sehingga persamaan regresi taksirannya adalah $\hat{Y} = 10 + 2X$, artinya kita taksir rata-rata jam kerja bertambah dengan 2 jam untuk setiap pertambahan 1 unit produk.

Untuk model alternatif: $Y_i = \beta_0^* + \beta_1\left(X_i - \bar{X}\right) + \varepsilon_i$ dengan $\beta_0^* = \beta_0 + \beta_1 \bar{X}$, penaksir kuadrat terkecil untuk b_0 adalah $b_0^* = b_0 + b_1\bar{X} = \left(\bar{Y} - b_1\bar{X}\right) + b_1\bar{X} = \bar{Y}$. Sehingga persamaan regresi taksiran untuk model alternatif adalah $\hat{Y} = \bar{Y} + b_1\left(X - \bar{X}\right)$. Persamaan regresi taksiran untuk contoh di atas adalah $\hat{Y} = 110 + 2(X - 50)$.

Residual

Kesalahan *(residual)* ke-*i* adalah selisih antara nilai pengamatan Y_i dengan nilai taksirannya $\hat{Y}_i$, ditulis dengan notasi e_i.

$$e_i = Y_i - \hat{Y}_i = Y_i - b_0 - b_1 X_i$$

Kita perlu membedakan antara nilai suku sesatan $\varepsilon_i = Y_i - E(Y_i)$ dengan $e_i = Y_i - \hat{Y}_i$. Residual kita gunakan untuk mempelajari ketepatan model regresi untuk data.

Contoh 1.5

Dari Contoh 1.4 telah dihitung persamaan regresi $\hat{Y} = 10 + 2X$. Nilai taksiran $\hat{Y}_i$, e_i, dan e_i^2 dapat diperoleh dengan cara sebagai berikut. Misalkan untuk $X_1 = 30$ dan $Y_1 = 73$, maka diperoleh nilai-nilai $\hat{Y}_1 = 10 + 2(30) = 70$, $e_1 = 73 - 70 = 3$, dan $e_1^2 = 9$. Untuk nilai-nilai X_i dan Y_i lainnya, diperoleh nilai taksiran $\hat{Y}_i$, e_i, dan e_i^2 sebagai berikut.

No.	X_i	Y_i	$\hat{Y}_i$	e_i	e_i^2
1	30	73	70	3	9
2	20	50	50	0	0
3	60	128	130	-2	4
4	80	170	170	0	0
5	40	87	90	-3	9
6	50	108	110	-2	4
7	60	135	130	5	25
8	30	69	70	-1	1
9	70	148	150	-2	4
10	60	132	130	2	4
Total	500	1.100	1.100	0	60

Sifat-Sifat Garis Regresi Taksiran (*Fitted*)

Persamaan garis regresi yang dihitung dengan metode kuadrat terkecil memenuhi sifat-sifat berikut:

1. Jumlah residual sama dengan nol, yakni $\sum_{i=1}^{n} e_i = 0$

   Bukti:

$$\sum_{i=1}^{n} e_i = \sum_{i=1}^{n} \left(Y_i - b_0 - b_1 X_1\right)$$

$$= \sum_{i=1}^{n} Y_i - nb_0 - b_1 \sum_{i=1}^{n} X_i = 0 \text{ (dari persamaan normal)}$$

2. Jumlah kuadrat residual, $\sum_{i=1}^{n} e_i^2$ adalah minimum. Hal ini sesuai dengan syarat dalam penghitungan penaksir kuadrat terkecil untuk parameter regresi.

3. Jumlah nilai observasi Y_i sama dengan jumlah nilai taksiran $\hat{Y}_i$, yakni

$$\sum_{i=1}^{n} Y_i = \sum_{i=1}^{n} \hat{Y}_i$$

   Bukti:

   Dari persamaan normal diperoleh

$$\sum_{i=1}^{n} Y_i = nb_0 + b_1 \sum_{i=1}^{n} X_i = \sum_{i=1}^{n} b_0 + \sum_{i=1}^{n} b_1 X_i$$

$$= \sum_{i=1}^{n} \left(b_0 + b_1 X_i\right) = \sum_{i=1}^{n} \hat{Y}_i$$

4. Jumlah residual tertimbang sama dengan nol jika angka timbang adalah X_i, yakni

$$\sum_{i=1}^{n} X_i\, e_i = 0$$

Bukti:
Dari persamaan normal dapat diturunkan

$$\begin{aligned}\sum_{i=1}^{n} X_i\, e_i &= \sum_{i=1}^{n} X_i \left(Y_i - b_0 - b_1 X_i\right)\\ &= \sum_{i=1}^{n} X_i Y_i - b_0 \sum_{i=1}^{n} X_i - b_1 \sum_{i=1}^{n} X_i^2 = 0\end{aligned}$$

5. Jumlah residual tertimbang dengan angka timbang $\hat{Y}_i$ sama dengan nol, yaitu

$$\sum_{i=1}^{n} \hat{Y}_i\, e_i = 0$$

6. Garis regresi selalu melalui titik $\left(\overline{X}, \overline{Y}\right)$.

Bukti:
Untuk $X = \overline{X}$, diperoleh $\hat{Y} = \overline{Y} + b_1\left(X - \overline{X}\right) = \overline{Y} + b_1\left(\overline{X} - \overline{X}\right) = \overline{Y}$

Penaksiran Variansi Suku Sesatan

Variansi suku sesatan $\left(\sigma^2\right)$ kita taksir untuk mengetahui keragaman dari distribusi *Y*. Penaksir titik untuk σ^2 dapat dihitung dari residual e_i. Jumlah kuadrat dari e_i adalah

$$\begin{aligned}JKS &= \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n}\left(Y_i - \hat{Y}\right)^2 = \sum_{i=1}^{n}\left(Y_i - b_0 - b_1 X_i\right)^2\\ &= \sum_{i=1}^{n} Y_i^2 - b_0 \sum_{i=1}^{n} Y_i - b_1 \sum_{i=1}^{n} X_i Y_i\end{aligned}$$

Rumus lain untuk menghitung *JKS* adalah

$$JKS = S_{YY} - \frac{\left(S_{XY}\right)^2}{S_{XX}} = \sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2 - \frac{\left[\sum_{i=1}^{n}\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)\right]^2}{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}$$

atau

$$JKS = \left[\sum_{i=1}^{n} Y_i^2 - \frac{\left(\sum_{i=1}^{n} Y_i\right)^2}{n}\right] - \frac{\left[\sum_{i=1}^{n} X_i Y_i - \frac{\sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{n}\right]^2}{\left[\sum_{i=1}^{n} X_i^2 - \frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n}\right]}$$

JKS mempunyai derajat bebas $n-2$, dua derajat bebas hilang karena dalam perhitungan kita gunakan penaksir untuk β_0 dan β_1. Kuadrat rata-rata sesatan (*KRS*) dirumuskan sebagai:

$$KRS = \frac{JKS}{n-2} = \frac{\sum_{i=1}^{n} e_i^2}{n-2} = \frac{\sum_{i=1}^{n} \left(Y_i - \hat{Y}_i\right)^2}{n-2} = \frac{\sum_{i=1}^{n} \left(Y_i - b_0 - b_1 X_i\right)^2}{n-2}$$

KRS merupakan taksiran tak bias dari σ^2, yakni $E\left(KRS\right) = \sigma^2$

Contoh 1.5

Kita tinjau kembali Contoh 1.4

$$JKS = \left[\sum_{i=1}^{n} Y_i^2 - \frac{\left(\sum_{i=1}^{n} Y_i\right)^2}{n}\right] - \frac{\left[\sum_{i=1}^{n} X_i Y_i - \frac{\sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{n}\right]^2}{\left[\sum_{i=1}^{n} X_i^2 - \frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n}\right]}$$

$$= \left[134.660 - \frac{\left(1.100\right)^2}{10}\right] - \frac{\left[61.800 - \frac{\left(500\right)\left(1.100\right)}{10}\right]^2}{28.400 - \frac{\left(500\right)^2}{10}} = 13.660 - 13.600 = 60$$

dan $KRS = \frac{JKS}{n-2} = \frac{60}{8} = 7,5$

LATIHAN

Untuk memperdalam pemahaman Anda mengenai materi di atas, silakan Anda mengerjakan latihan berikut ini!

1) Suatu eksperimen dilakukan untuk melihat hubungan antara dosis pemupukan (X) dan hasil panen (Y). Dari hasil perhitungan didapat nilai-nilai

$$\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2=70,6; \quad \sum_{i=1}^{n}\left(Y_i-\bar{Y}\right)^2=98,5$$

$$\sum_{i=1}^{n}\left(X_i-\bar{X}\right)\left(Y_i-\bar{Y}\right)=68,3; \; n=15; \quad \bar{X}=10,8; \quad \bar{Y}=122,7$$

Jika hubungan antara Y dan X diasumsikan linear, maka

a) dengan metode kuadrat terkecil, hitunglah garis regresi $\hat{Y}=b_0+b_1X$

b) hitunglah nilai *JKS* dan selanjutnya hitung taksiran untuk σ^2

2) Diketahui observasi berpasangan (X,Y) sebagai berikut:

X	1	2	3	4	5
Y	0,9	2,1	2,5	3,3	3,8

a) Buat diagram titik

b) Hitunglah penduga kuadrat terkecil b_0 dan b_1

c) Hitunglah nilai taksiran dari σ^2

Petunjuk Jawaban Latihan

1) a) $$b_1=\frac{\sum_{i=}^{n}\left(X_i-\bar{X}\right)\left(Y_i-\bar{Y}\right)}{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2}=\frac{68,3}{70,6}=0,9674$$

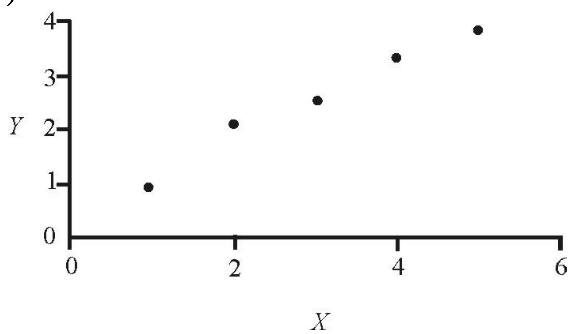
$$b_0=\bar{Y}-b_1\bar{X}=122,7-(0,9674)(10,8)=112,2518$$

sehingga garis regresi $\hat{Y}=112,2518+0,9674\,X$

b) $$JKS = \sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2 - \frac{\left[\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)\right]^2}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2} = 98,5 - \frac{\left(68,3\right)^2}{70,6} = 32,4251$$

$$\hat{\sigma}^2 = \frac{JKS}{n-2} = \frac{32,4251}{13} = 2,4942$$

2) a)



b) $$b_1 = \frac{\sum_{i=}^{n}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2} = \frac{44,8 - \frac{(15)(12,6)}{5}}{55 - \frac{(15)^2}{5}} = \frac{7}{10} = 0,7$$

$$b_0 = \bar{Y} - b_1\bar{X} = 2,52 - (0,7)(3) = 0,42$$

c) $$JKS = \sum_{i=1}^{n} Y_i^2 - b_0\sum_{i=1}^{n} Y_i - b_1\sum_{i=1}^{n} X_i Y_i$$

$$= 36,8 - (0,42)(12,6) - (0,7)(44,8) = 0,148$$

$$\hat{\sigma}^2 = \frac{JKS}{n-2} = \frac{0,148}{3} = 0,0493$$

RANGKUMAN

1. Dalam modul ini dipelajari hubungan statistik antara variabel bebas $\left(X\right)$ dan variabel tak bebas $\left(Y\right)$. Dalam analisis linear sederhana bentuk hubungannya adalah garis lurus dengan model $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$; $\varepsilon_i \sim N\left(0, \sigma^2\right)$ dan independen.

2. Untuk mencari estimasi dari β_0 dan β_1 digunakan metode kuadrat terkecil dan didapat penaksir titik:

$$b_1 = \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2} \text{ dan } b_0 = \bar{Y} - b_1\bar{X}$$

3. Selanjutnya didefinisikan:

$$JKS = \sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2 - \frac{\left[\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)\right]^2}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}$$

dan sebagai estimasi titik untuk σ^2 adalah:

$$KRS = \frac{JKS}{n-2} = s^2$$

TES FORMATIF 1

Pilih satu jawaban yang paling tepat dari beberapa alternatif jawaban yang disediakan!

1) Tabel berikut ini menunjukkan tinggi badan (X) dalam cm dan berat badan (Y) dalam kg dari 10 orang dewasa.

X	150	162	160	162	165	160	172	170	180	182
Y	55	67	60	70	65	79	79	76	89	90

Jika *X* sebagai variabel independen *Y* sebagai variabel dependen dalam model regresi $Y_i = \beta_0 + \beta_1 X + \varepsilon_i$. Dari data ini dapat dihitung nilai b_0 sama dengan ….
A. -103,63
B. -130,63
C. -103,36
D. -130,36

2) **Lihat soal nomor 1**, nilai b_1 sama dengan ….
A. 1,052
B. 1,521
C. 1,062
D. 1,620

3) **Lihat soal nomor 1**, untuk $X_k = 175$, nilai terhitung $\hat{Y}_k$ sama dengan ….
A. 81,28
B. 82,18
C. 82,84
D. 81,81

4) **Lihat soal nomor 1**, nilai variansi (s^2) sama dengan ….
A. 32,61
B. 31,16
C. 32,26
D. 31,59

5) Diketahui 10 observasi berpasangan (X, Y).

X	54,5	56,4	43,2	65,2	45,5	47,5	65,0	66,5	57,3	68,0
Y	61,5	61,2	32,0	52,5	31,5	22,5	53,0	56,8	34,8	52,7

Jika digunakan model regresi $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$; dari data di atas dihitung nilai b_0 sama dengan ….

A. -15,19
B. -15,20
C. -15,09
D. -15,92

6) **Lihat soal nomor 5**, nilai b_1 sama dengan ….
A. 1,073
B. 1,072
C. 1,721
D. 1,722

7) **Lihat soal nomor 5**, nilai estimasi dari variansi *Y*, yakni s^2 sama dengan ….
A. 931,46
B. 913,46
C. 116,37
D. 116,44

8) Diketahui: $n = 277$; $\bar{X} = 65$; $\bar{Y} = 72$; $\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 = 1600$;

$$\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right) = 2000\,;\ \sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2 = 3600$$

Jika digunakan model regresi $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$; nilai b_0 sama dengan ….
A. -9,27
B. -9,25
C. -9,15
D. -9,52

9) **Lihat soal nomor 8**, nilai b_1 sama dengan ….
A. -1,25
B. 1,25
C. -1,35
D. 1,53

10) **Lihat soal nomor 8**, nilai s^2 sama dengan ….
A. 3,99
B. 3,98
C. 3,97
D. 4,00

Cocokkanlah jawaban Anda dengan Kunci Jawaban Tes Formatif 1 yang terdapat di bagian akhir modul ini. Hitunglah jawaban Anda yang benar. Kemudian gunakan rumus di bawah ini untuk mengetahui tingkat penguasaan Anda terhadap materi Kegiatan Belajar 1.

Rumus:

$$\text{Tingkat penguasaan} = \frac{\text{Jumlah jawaban Anda yang benar}}{10} \times 100\%$$

Arti tingkat penguasaan yang Anda capai:

90 - 100%	=	baik sekali
80 - 89%	=	baik
70 - 79%	=	cukup
< 70%	=	kurang

Bila Anda mencapai tingkat penguasaan 80% atau lebih, Anda dapat meneruskan dengan Kegiatan Belajar 2. **Bagus!** Tetapi bila tingkat penguasaan Anda masih di bawah 80%, Anda harus mengulangi Kegiatan Belajar 1, terutama bagian yang belum Anda kuasai.

Inferensi dalam Analisis Regresi

Tanpa memperhatikan bentuk fungsional distribusi dari ε_i, penaksir kuadrat terkecil b_0 dan b_1 selalu bersifat tak bias dan mempunyai variansi minimum dibandingkan penaksir-penaksir tak bias linear lainnya. Untuk inferensi, kita memerlukan asumsi tentang bentuk fungsional distribusi dari ε_i. Salah satu asumsi baku yang diperlukan adalah $\varepsilon_i \sim N\left(0,\sigma^2\right)$. Secara umum model regresi dengan sesatan normal dinyatakan sebagai

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

dengan

Y_i : hasil observasi pada trial ke-i

X_i : konstanta yang diketahui nilainya, yakni tingkat variabel X dalam trial ke-i

β_0, β_1 : parameter

$\varepsilon_i \sim N\left(0,\sigma^2\right)$ independen; $i = 1, 2, \ldots n$

Salah satu alasan digunakannya asumsi $\varepsilon_i \sim N\left(0,\sigma^2\right)$ karena prosedur inferensi didasarkan pada distribusi t yang tidak peka terhadap penyimpangan normalitas yang tidak besar (*modurate*).

Penaksiran Parameter dengan Metode Kemungkinan Maksimum

Apabila bentuk fungsional dari distribusi probabilitas suku sesatan tertentu, penaksir untuk β_0, β_1, dan σ^2 dapat dihitung dengan metode kemungkinan maksimum (*maximum likelihood*). Fungsi likelihood untuk model regresi sesatan normal adalah:

$$L\left(\beta_0, \beta_1, \sigma^2\right) = \prod_{i=1}^{n} \frac{1}{\left(2\pi\sigma^2\right)^{\frac{1}{2}}} exp\left[-\frac{1}{2\sigma^2}\left(Y_i - \beta_0 - \beta_1 X_i\right)^2\right]$$

Nilai-nilai β_0, β_1, dan σ^2 yang memaksimumkan fungsi likelihood merupakan penaksir-penaksir *maximum likelihood* yang diperoleh dengan mendeferensialkan $L\left(\beta_0, \beta_1, \sigma^2\right)$ terhadap β_0, β_1, dan σ^2; kemudian masing-masing disamakan dengan nol. Untuk menyederhanakan perhitungan, kita dapat menggunakan $\ln L$ karena L dan $\ln L$ akan maksimum untuk nilai-nilai yang sama dari β_0, β_1, dan σ^2. Dengan mendeferensialkan $\ln L = -\frac{n}{2}\ln\left(2\pi\right) - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(Y_i - \beta_0 - \beta_1 X_i\right)^2$ terhadap β_0, β_1, dan σ^2 diperoleh:

$$\begin{aligned}
\frac{\partial \ln L}{\partial \beta_0} &= \frac{1}{\sigma^2}\sum_{i=1}^{n}\left(Y_i - \beta_0 - \beta_1 X_i\right) \\
\frac{\partial \ln L}{\partial \beta_1} &= \frac{1}{\sigma^2}\sum_{i=1}^{n}X_i\left(Y_i - \beta_0 - \beta_1 X_i\right) \\
\frac{\partial \ln L}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}\left(Y_i - \beta_0 - \beta_1 X_i\right)^2
\end{aligned}$$

Jika β_0, β_1, dan σ^2 masing-masing diganti dengan b_0, b_1 dan $\hat{\sigma}^2$ serta disamakan dengan nol, maka diperoleh persamaan

$$\sum_{i=1}^{n}\left(Y_i - b_0 - b_1 X_i\right) = 0$$

$$\sum_{i=1}^{n}\left(X_i Y_i - b_0 X_i - b_1 X_i^2\right) = 0$$

$$\frac{1}{n}\sum_{i=1}^{n}\left(Y_i - b_0 - b_1 X_i\right)^2 = \hat{\sigma}^2$$

Dua persamaan pertama, sama dengan persamaan normal, sehingga penaksir *maximum likelihood* untuk masing-masing parameter adalah sebagai berikut.

$$b_1 = \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}_i\right)\left(Y_i - \bar{Y}_i\right)}{\sum_{i=1}^{n}\left(X_i - \bar{X}_i\right)^2}$$

$$b_0 = \bar{Y} - b_1\bar{X}$$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2}{n}$$

Jadi penaksir *maksimum likelihood* untuk β_0 dan β_1 sama dengan penaksir kuadrat terkecil yakni b_0 dan b_1. Penaksir *maksimum likelihood* $\hat{\sigma}^2$ adalah bias sehingga untuk mendapatkan penaksir tak bias untuk σ^2 digunakan *KRS* (Kuadrat Rata-rata Sesatan).

Inferensi tentang β_1

Sering kali kita tertarik untuk melakukan inferensi tentang β_1, yaitu kemiringan (*slope*) dari garis regresi $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ dengan β_0 dan β_1 adalah parameter, X_i adalah konstanta yang diketahui, dan $\varepsilon_i \sim N\left(0, \sigma^2\right)$ adalah independen. Sebelum kita pelajari inferensi tentang β_1, kita bahas terlebih dahulu tentang distribusi sampling penaksir titik untuk β_1 yaitu

$$b_1 = \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{\sum_{i=1}^{n}\left(X_i - \bar{X}_i\right)^2}$$

Dari pengambilan sampel berulang pada nilai X tetap, diperoleh nilai b_1 yang bervariasi. Untuk model regresi sesatan normal akan dibuktikan bahwa

$$b_1 \sim N\left(\beta_1; \frac{\sigma^2}{\sum_{i=1}^{n}\left(X_i - \bar{X}_i\right)^2}\right)$$

Bukti:

b_1 dapat dinyatakan sebagai kombinasi linear dari Y_i yaitu:

$$b_1 = \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2} = \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)Y_i}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2} = \sum_{i=1}^{n} k_i Y_i$$

dengan

$$k_i = \frac{X_i - \bar{X}}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

Untuk X_i tetap, k_i merupakan kuantitas tetap. Oleh karena itu, b_1 merupakan kombinasi linear dari Y_i. Koefisien k_i mempunyai sifat-sifat sebagai berikut.

1. $\sum_{i=1}^{n} k_i = 0$

Bukti: $$\sum_{i=1}^{n} k_i = \sum_{i=1}^{n}\left(\frac{X_i - \bar{X}}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\right) = \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2} = \frac{0}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2} = 0$$

2. $\sum_{i=1}^{n} k_i X_i = 1$

Bukti: $$\sum_{i=1}^{n} k_i X_i = \sum_{i=1}^{n}\left(\frac{\left(X_i - \bar{X}\right)X_i}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\right) = \sum_{i=1}^{n}\left(\frac{\left(X_i - \bar{X}\right)^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\right) = \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2} = 1$$

3. $\sum_{i=1}^{n} k_i^2 = \dfrac{1}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$

Bukti: $$\sum_{i=1}^{n} k_i^2 = \sum_{i=1}^{n}\left(\frac{\left(X_i - \bar{X}\right)}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\right)^2$$

$$= \frac{1}{\left[\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2\right]^2}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 = \frac{1}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}$$

Kembali kita perhatikan distribusi sampling dari b_1 pada model regresi linear dengan $\varepsilon_i \sim N\left(0,\sigma^2\right)$. Distribusi sampling dari b_1 adalah normal, hal ini akibat dari b_1 merupakan kombinasi linear dari Y_i. Akan kita buktikan b_1 merupakan penaksir tak bias untuk β_1 atau $E\left(b_1\right) = \beta_1$.

Bukti:

$$E\left(b_1\right) = E\left(\sum_{i=1}^{n} k_i Y_i\right) = k_i E\left(Y_i\right)$$

$$= \sum_{i=1}^{n} k_i\left(\beta_0 + \beta_1 X_i\right) = \beta_0 \sum_{i=1}^{n} k_i + \beta_1 \sum_{i=1}^{n} k_i X_i = \beta_1$$

Selanjutnya variansi dari b_1 dijabarkan sebagai berikut.

$$\sigma^2(b_1) = \sigma^2\left(\sum_{i=1}^{n} k_i Y_i\right) = \sum_{i=1}^{n} k_i^2 \sigma^2(Y_i)$$
$$= \sum_{i=1}^{n} k_i^2 \sigma^2 = \sigma^2 \sum_{i=1}^{n} k_i^2 = \sigma^2 \frac{1}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}$$

Penaksir untuk $\sigma^2(b_1)$ diperoleh dengan mengganti σ^2 dengan *KRS* sehingga penaksir tak bias untuk $\sigma^2(b_1)$ adalah:

$$s^2(b_1) = \frac{KRS}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2} = \frac{KRS}{\sum_{i=1}^{n} X_i^2 - \frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n}}$$

Untuk membuktikan bahwa b_1 mempunyai variansi minimum di antara semua penaksir tak bias linear, kita nyatakan penaksir linear tak bias untuk β_1 dalam bentuk:

$$\tilde{\beta}_1 = \sum_{i=1}^{n} c_i Y_i \text{ dengan } c_i \text{ adalah konstanta sembarang.}$$

Jika $\tilde{\beta}_1$ penaksir tak bias maka:

$$E\left(\tilde{\beta}_1\right) = E\left(\sum_{i=1}^{n} c_i Y_i\right) = \sum_{i=1}^{n} c_i E(Y_i) = \beta_1.$$

Karena $E(Y_i) = \beta_0 + \beta_1 X_i$, maka syarat tak bias di atas menjadi:

$$E\left(\tilde{\beta}_1\right) = \sum_{i=1}^{n} c_i\left(\beta_0 + \beta_1 X_i\right) = \beta_0 \sum_{i=1}^{n} c_i + \beta_1 \sum_{i=1}^{n} c_i X_i = \beta_1.$$

Agar syarat tak bias berlaku maka haruslah $\sum_{i=1}^{n} c_i = 0$ dan $\sum_{i=1}^{n} c_i X_i = 1$.

Variansi dari $\tilde{\beta}_1$ adalah:

$$\sigma^2\left(\tilde{\beta}_1\right) = \sum_{i=1}^{n} c_i^2 \sigma^2(Y_i) = \sigma^2 \sum_{i=1}^{n} c_i^2 .$$

Kita definisikan:

$c_i = k_i + d_i$; $k_i = \dfrac{X_i - \bar{X}}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}$; d_i adalah konstanta sembarang,

sehingga

$$\begin{aligned}
\sigma^2\left(\tilde{\beta}_1\right) &= \sigma^2 \sum_{i=1}^{n} c_i^2 \\
&= \sigma^2 \sum_{i=1}^{n} \left(k_i + d_i\right)^2 \\
&= \sigma^2 \left(\sum_{i=1}^{n} k_i^2 + \sum_{i=1}^{n} d_i^2 + 2\sum_{i=1}^{n} k_i d_i \right)
\end{aligned}$$

Telah kita buktikan bahwa

$$\sum_{i=1}^{n} k_i^2 = \frac{1}{\sum_{i=1}^{n}(X_i - \bar{X})^2} \quad \text{dan } \sigma^2\left(b_1\right) = \sigma^2 \sum_{i=1}^{n} k_i^2$$

sehingga

$$\begin{aligned}
\sum_{i=1}^{n} k_i\, d_i &= \sum_{i=1}^{n} k_i\left(c_i - k_i\right) = \sum_{i=1}^{n} c_i\, k_i - \sum_{i=1}^{n} k_i^2 \\
&= \sum_{i=1}^{n} c_i \left(\frac{X_i - \bar{X}}{\sum_{i=1}^{n}(X_i - \bar{X})^2} \right) - \frac{1}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2} \\
&= \frac{\sum_{i=1}^{n} c_i\, X_i - \bar{X}\sum_{i=1}^{n} c_i}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2} - \frac{1}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2} = 0
\end{aligned}$$

Kita peroleh $\sigma^2\left(\tilde{\beta}_1\right) = \sigma^2\left(b_1\right) + \sigma^2 \sum_{i=1}^{n} d_i^2$, sehingga terbukti $\sigma^2\left(\tilde{\beta}_1\right) \geq \sigma^2\left(b_1\right)$. Nilai terkecil diperoleh jika $\sum_{i=1}^{n} d_i^2 = 0$ atau semua $\sum_{i=1}^{n} d_i = 0$ sehingga $c_i = k_i$. Jadi, $\sigma^2\left(b_1\right)$ merupakan nilai terkecil untuk $\sigma^2\left(\tilde{\beta}_1\right)$.

Distribusi Sampling dari $\dfrac{(b_1 - \beta_1)}{s(b_1)}$

Diketahui $b_1 \sim N\left(\beta_1, \sigma^2(b_1)\right)$ dengan $\sigma^2(b_1) = \dfrac{\sigma^2}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}$, sehingga statistik $\dfrac{b_1 - \beta_1}{\sigma(b_1)} \sim N(0, 1)$. Pada umumnya $\sigma^2(b_1)$ tidak diketahui, diganti dengan $s(b_1)$. Statistik $\dfrac{b_1 - \beta_1}{s(b_1)}$ dapat ditulis dalam bentuk $\dfrac{b_1 - \beta_1}{\sigma(b_1)} \Big/ \dfrac{s(b_1)}{\sigma(b_1)}$. Diketahui $\dfrac{b_1 - \beta_1}{\sigma(b_1)} \sim z$ dan

$$\frac{s^2(b_1)}{\sigma^2(b_1)} = \frac{\dfrac{KRS}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}}{\dfrac{\sigma^2}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}} = \frac{KRS}{\sigma^2} = \frac{\dfrac{JKS}{n-2}}{\sigma^2} = \frac{JKS}{\sigma^2(n-2)} \sim \frac{\chi^2_{(n-2)}}{n-2} \text{ sehingga } \frac{b_1 - \beta_1}{s(b_1)} \sim \frac{z}{\sqrt{\dfrac{\chi^2_{(n-2)}}{n-2}}}$$

Karena b_1 dan $\dfrac{JKS}{\sigma^2}$ independen maka z dan χ^2 independen, sehingga $\dfrac{b_1 - \beta_1}{s(b_1)} \sim t_{(n-2)}$.

Selang Kepercayaan untuk β_1

Karena $\dfrac{b_1 - \beta_1}{s(b_1)} \sim t_{(n-2)}$ maka kita dapat menyatakan

$$P\left[t_{\left(\frac{\alpha}{2};n-2\right)} \le \frac{b_1 - \beta_1}{s(b_1)} \le t_{\left(1-\frac{\alpha}{2};n-2\right)}\right] = 1 - \alpha$$

Karena distribusi *t* simetri, maka ketidaksamaan di atas dapat kita tulis sebagai berikut.

$$P\left[b_1 - t_{\left(\frac{\alpha}{2};n-2\right)}\, s(b_1) \le \beta_1 \le b_1 + t_{\left(\frac{\alpha}{2};n-2\right)}\, s(b_1)\right] = 1 - \alpha$$

Karena probabilitas ini berlaku untuk semua nilai yang mungkin dari β_1 maka batas kepercayaan untuk β_1 adalah:

$$b_1 \pm t_{\left(\frac{\alpha}{2};n-2\right)}\, s(b_1)$$

Contoh 1.6

Diketahui data sebagai berikut.

X_i	1,5	1,8	2,4	3,0	3,5	3,9	4,4	4,8	5,0
Y_i	4,8	5,7	7,0	8,3	10,9	12,4	13,1	13,6	15,3

Jika digunakan model regresi linear $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ dengan $\varepsilon_i \sim N\left(0, \sigma^2\right)$, independen. Untuk menaksir β_1 dengan selang kepercayaan 95%, dari data diperoleh:

$$n = 9;\ \sum_{i=1}^{n} X_i = 30,3;\ \bar{X} = 3,3667;\ \sum_{i=1}^{n} Y_i = 91,1;\ \bar{Y} = 10,1222;$$

$$\sum_{i=1}^{n} X_i^2 = 115,11;\ \sum_{i=1}^{n} Y_i^2 = 1036,65;\ \sum_{i=1}^{n} X_i Y_i = 345,09$$

$$b_1 = \frac{\sum_{i=1}^{n} X_i Y_i - \dfrac{\sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{n}}{\sum_{i=1}^{n} X_i^2 - \dfrac{\left(\sum_{i=1}^{n} X_i\right)^2}{n}} = \frac{345,09 - \dfrac{(30,3)(91,1)}{9}}{115,11 - \dfrac{(30,3)^2}{9}} = 2,9303$$

$$JKS = \left[\sum_{i=1}^{n} Y_i^2 - \frac{\left(\sum_{i=1}^{n} Y_i\right)^2}{n}\right] - \frac{\left[\sum_{i=1}^{n} X_i Y_i - \dfrac{\sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{n}\right]^2}{\left[\sum_{i=1}^{n} X_i^2 - \dfrac{\left(\sum_{i=1}^{n} X_i\right)^2}{n}\right]}$$

$$= \left[1036,65 - \frac{(91,1)^2}{9}\right] - \frac{\left[345,09 - \dfrac{(30,3)(91,1)}{9}\right]^2}{115,11 - \dfrac{(30,3)^2}{9}} = 2,026$$

$$KRS = \frac{JKS}{n-2} = \frac{2,026}{7} = 0,289$$

Selanjutnya kita hitung

$$s^2(b_1) = \frac{KRS}{\left(\sum_{i=1}^{n} X_i^2 - \frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n}\right)} = \frac{KRS}{S_{XX}} = \frac{0,289}{115,11 - \frac{(30,3)^2}{9}} = \frac{0,289}{13,1} = 0,02206$$

$$s(b_1) = \sqrt{0,02206} = 0,1485 \text{ dan } t_{(0,025;\,7)} = 2,365$$

Selang kepercayaan 95% untuk β_1 adalah

$$2,9303 - 2,365(0,1485) \le \beta_1 \le 2,9303 + 2,365(0,1485) \text{ atau } 2,579 \le \beta_1 \le 3,282$$

Uji tentang β_1

Karena $\frac{(b_1 - \beta_1)}{s(b_1)} \sim t_{(n-2)}$, uji tentang β_1 dapat kita lakukan berdasarkan distribusi *t*.

Contoh 1.7

Jika kita hanya ingin menguji apakah ada (atau tidak) hubungan linear antara *Y* dan *X* maka hipotesis kita rumuskan sebagai:

$$H_0 : \beta_1 = 0 \text{ dan } H_1 : \beta_1 \ne 0$$

Untuk $\alpha = 0,05$, kesimpulan tentang H_0 dapat diperoleh dengan menggunakan selang kepercayaan 95% untuk β_1. Untuk Contoh 1.6, selang kepercayaan 95% untuk β_1 adalah $2,579 \le \beta_1 \le 3,282$ tidak memuat 0 (batas kepercayaan kiri dan kanan positif) sehingga diperoleh kesimpulan menolak H_0.

Uji yang lebih tegas dapat dilakukan berdasarkan statistik uji $t^* = \frac{b_1}{s(b_1)}$. Aturan pengambilan kesimpulan untuk taraf nyata α adalah menerima H_0 jika $|t^*| \le t_{\left(\frac{\alpha}{2};\, n-2\right)}$ dan menolak H_0 jika $|t^*| > t_{\left(\frac{\alpha}{2};\, n-2\right)}$.

Contoh 1.8

Dari Contoh 1.6, hipotesis kita tulis sebagai $H_0 : \beta_1 = 0$ dan $H_1 : \beta_1 \ne 0$. Selanjutnya kita hitung statistik uji $t^* = \frac{b_1}{s(b_1)} = \frac{2,9303}{0,1485} = 19,73$. Untuk taraf nyata $\alpha = 0,05$, diperoleh nilai $t_{(0,025;7)} = 2,365$. Karena $|t^*| > 2,365$ maka kita menolak $H_0 : \beta_1 = 0$, berarti ada hubungan linear antara *Y* dan *X*.

Sering kali kita ingin menguji apakah β_1 sama dengan nilai tertentu, hipotesis kita rumuskan sebagai:

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

Statistik uji:

$$t^* = \frac{b_1 - \beta_1}{s(b_1)}$$

Jika kita ingin menguji apakah β_1 positif, hipotesis dinyatakan sebagai:

$$H_0 : \beta_1 \leq 0 \quad \text{dan} \quad H_1 : \beta_1 > 0$$

Kesimpulan untuk taraf nyata α adalah menerima H_0 jika $t^* \leq t_{(\alpha;\, n-2)}$ dan menolak H_0 jika $t^* > t_{(\alpha;\, n-2)}$.

Contoh 1.9

Dari Contoh 1.8 diperoleh $t^* = \dfrac{b_1 - \beta_1}{s(b_1)} = 19,73$. Untuk $\alpha = 0,05$, diperoleh nilai $t_{(0,05;7)} = 1,895$. Karena $t^* > 1,895$, maka H_0 ditolak, berarti β_1 positif.

Inferensi tentang β_0

Sebelum kita melakukan inferensi untuk β_0, kita pelajari terlebih dahulu tentang distribusi sampling dari b_0. Penaksir titik untuk β_0 adalah:

$$b_0 = \bar{Y} - b_1 \bar{X}$$

dapat dinyatakan sebagai kombinasi linear dari Y_i yaitu:

$$b_0 = \sum_{i=1}^{n} l_i \, Y_i \text{ dengan } l_i = \frac{1}{n} - \bar{X} \, k_i$$

dan

$$k_i = \frac{X_i - \bar{X}}{\sum_{i=1}^{n} \left(X_i - \bar{X} \right)^2}$$

Distribusi sampling untuk b_0 adalah normal dan b_0 merupakan penaksir tak bias dari β_0 atau $E(b_0) = \beta_0$.

Bukti:

$$b_0 = \bar{Y} - b_1 \bar{X}$$
$$E(b_0) = E(\bar{Y}) - E(b_1 \bar{X}) = \beta_0 + \beta_1 \bar{X} - \beta_1 \bar{X} = \beta_0$$

Variansi dari b_0 kita peroleh sebagi berikut.

$$b_0 = \sum_{i=1}^{n} l_i Y_i$$
$$\sigma^2(b_0) = \sum_{i=1}^{n} l_i^2 \sigma^2(Y_i) = \sigma^2 \sum_{i=1}^{n} l_i^2 = \sigma^2 \left(\frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^{n} (X_i - \bar{X})^2} \right)$$

Sehingga untuk model regresi linear dengan sesatan normal, diperoleh:

$$b_0 \sim N\left[\beta_0, \sigma^2(b_0)\right]$$

dengan

$$\sigma^2(b_0) = \sigma^2 \left(\frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^{n} (X_i - \bar{X})^2} \right)$$

dan distribusi sampling untuk $\dfrac{b_0 - \beta_0}{\sigma(b_0)}$ adalah normal standar atau $\dfrac{b_0 - \beta_0}{\sigma(b_0)} \sim N(0,1)$.

Penaksir untuk $\sigma^2(b_0)$ adalah:

$$s^2(b_0) = KRS \left[\frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2} \right] = KRS \left[\frac{\sum_{i=1}^{n} X_i^2}{n \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2} \right]$$

dan

$$\frac{b_0 - \beta_0}{s(b_0)} \sim t_{(n-2)}$$

Selang kepercayaan untuk β_0 pada taraf kepercayaan $(1-\alpha)$ adalah:

$$b_0 - t_{\left(\frac{\alpha}{2};\, n-2\right)} s(b_0) \le \beta_0 \le b_0 + t_{\left(\frac{\alpha}{2};\, n-2\right)} s(b_0)$$

Contoh 1.10

Dari data pada Contoh 1.6 diperoleh:

$\sum_{i=1}^{n} X_i = 30,3$; $\sum_{i=1}^{n} X_i^2 = 115,11$; $\bar{X} = 3,3667$; $\bar{Y} = 10,1222$; $n = 9$; $b_1 = 2,9303$; dan $KRS = 0,289$. Selanjutnya kita hitung:

$$b_0 = \bar{Y} - b_1\bar{X} = 10,1222 - 2,9303(3,3667) = 0,2568$$

$$s^2(b_0) = KRS\left[\frac{\sum_{i=1}^{n} X_i^2}{n\sum_{i=1}^{n}(X_i - \bar{X})^2}\right] = 0,289\left[\frac{115,11}{9(13,1)}\right] = 0,282$$

$$s(b_0) = \sqrt{0,282} = 0,531$$

Untuk taraf kepercayaan 90%, diperoleh $t_{(0,05;\,7)} = 1,895$ sehingga selang kepercayaan untuk β_0 adalah:

$$0,2568 - (1,895)(0,531) \le \beta_0 \le 0,2568 + (1,895)(0,531) \text{ atau } -0,749 \le \beta_0 \le 1,263$$

Penaksir Selang untuk $E(Y_h)$

Salah satu tujuan utama analisis regresi adalah untuk menaksir mean dari satu atau lebih distribusi probabilitas dari *Y*. Misalkan X_h adalah taraf dari *X* di mana kita ingin menaksir mean dari *Y* dan X_h merupakan salah satu nilai sampel atau suatu nilai lain dari *X*, maka untuk $X = X_h$ diperoleh mean $\hat{Y}_h = E(Y_h)$.

Untuk model regresi dengan sesatan normal, distribusi sampling dari $\hat{Y}_h = b_0 + b_1 X_h$ adalah normal dengan mean dan variansi masing-masing adalah:

$$E(\hat{Y}_h) = E(Y_h)$$

dan

$$\sigma^2(\hat{Y}_h) = \sigma^2\left[\frac{1}{n} + \frac{(X_h - \bar{X})^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\right]$$

$\hat{Y}_h$ merupakan penaksir titik tak bias dari $E(Y_h)$ yang ditunjukkan sebagai berikut.

$$\begin{aligned} E(\hat{Y}_h) &= E(b_0 + b_1 X_h) \\ &= E(b_0) + X_h E(b_1) \\ &= \beta_0 + \beta_1 X_h = E(Y_h) \end{aligned}$$

Keragaman dari distribusi sampling $\hat{Y}_h$ dipengaruhi oleh besarnya simpangan X_h dari $\bar{X}$. Makin jauh X_h dari $\bar{X}$, nilai $\sigma^2\left(\hat{Y}_h\right)$ makin besar. Untuk menentukan $\sigma^2\left(\hat{Y}_h\right)$, kita buktikan dulu bahwa b_1 dan $\bar{Y}$ tidak berkorelasi.

Bukti:

$$\bar{Y} = \sum_{i=1}^{n} \frac{1}{n} Y_i \ ; \ b_1 = \sum_{i=1}^{n} k_i Y_i$$

dengan

$$k_i = \frac{X_i - \bar{X}}{\sum_{i=1}^{n} (X_i - \bar{X})^2} \text{ dan } Y_i \text{ independen.}$$

$$\sigma^2\left(\bar{Y}, b_1\right) = \sum_{i=1}^{n} \left(\frac{1}{n}\right) k_i \,\sigma^2\left(Y_i\right) = \frac{\sigma^2}{n} \sum_{i=1}^{n} k_i = 0, \text{ karena } \sum_{i=1}^{n} k_i = 0$$

Sekarang kita hitung variansi $\hat{Y}_h$, yaitu:

$$\sigma^2\left(\hat{Y}_h\right) = \sigma^2\left(\bar{Y} + b_1\left(X_h - \bar{X}\right)\right)$$

Karena $\bar{Y}$ dan b_1 independen serta X_h dan $\bar{X}$ konstan maka diperoleh:

$$\begin{aligned}
\sigma^2\left(\hat{Y}_h\right) &= \sigma^2\left(\bar{Y}\right) + \left(X_h - \bar{X}\right)^2 \sigma^2\left(b_1\right) \\
&= \frac{\sigma^2\left(Y_i\right)}{n} + \left(X_h - \bar{X}\right)^2 \sigma^2\left(b_1\right) \\
&= \frac{\sigma^2}{n} + \left(X_h - \bar{X}\right)^2 \frac{\sigma^2}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2} \\
&= \sigma^2 \left[\frac{1}{n} + \frac{\left(X_h - \bar{X}\right)^2}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}\right]
\end{aligned}$$

Penaksir untuk $\sigma^2\left(\hat{Y}_h\right)$ adalah:

$$s^2\left(\hat{Y}_h\right) = KRS\left[\frac{1}{n} + \frac{\left(X_h - \bar{X}\right)^2}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}\right]$$

Untuk model regresi dengan sesatan normal, distribusi sampling $\dfrac{\hat{Y}_h - E\left(Y_h\right)}{s\left(\hat{Y}_h\right)}$ adalah distribusi $t_{(n-2)}$, sehingga selang kepercayaan $\left(1-\alpha\right)$ 100% untuk $E\left(Y_h\right)$ adalah:

$$\hat{Y}_h - t_{\left(\frac{\alpha}{2};\, n-2\right)} s\left(\hat{Y}_h\right) \le E\left(Y_h\right) \le \hat{Y}_h + t_{\left(\frac{\alpha}{2};\, n-2\right)} s\left(\hat{Y}_h\right)$$

Contoh 1.11

Dari data pada Contoh 1.6 diperoleh:

$b_0 = 0,2568$; $b_1 = 2,93$; $\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 = 13,1$; $KRS = 0,289$; $n = 9$; $\bar{X} = 3,3667$

Untuk $X_h = 4$ diperoleh:

$$\hat{Y}_h = 0,2568 + \left(2,93\right)\left(4\right) = 11,977$$

$$s^2\left(\hat{Y}_h\right) = 0,289\left[\frac{1}{9} + \frac{\left(4-3,3667\right)^2}{13,1}\right] = 0,04096 \text{ atau } s\left(\hat{Y}_h\right) = 0,202$$

$$t_{(0,025;\, 7)} = 2,365$$

Selang kepercayaan 95% untuk $E\left(Y_h\right)$ adalah:

$$11,977 - 2,365\left(0,202\right) \le E\left(Y_h\right) \le 11,977 + 2,365\left(0,202\right)$$

$$11,499 \le E\left(Y_h\right) \le 12,455$$

Peramalan Pengamatan Baru

Pengamatan baru $Y_{h(baru)}$ dipandang sebagi hasil dari suatu trial baru yang bebas terhadap trial lain. Jika parameter β_0 dan β_1 diketahui maka selang peramalan $\left(1-\alpha\right)$100% untuk $Y_{h(baru)}$ adalah:

$$E\left(Y_h\right) - z_{\frac{\alpha}{2}}\,\sigma \le Y_{h(baru)} \le E\left(Y_h\right) + z_{\frac{\alpha}{2}}\,\sigma$$

Contoh 1.12

Misalkan diketahui $\beta_0 = 9,5$; $\beta_1 = 2,1$; $\sigma^2 = 10$.

Untuk $X_h = 40$ didapat $E\left(Y_{40}\right) = 9,5 + \left(2,1\right)\left(40\right) = 93,5$. Selang peramalan 95% untuk Y_{40} adalah:

$$93,5 - 1,96\sqrt{10} \le Y_{40} \le 93,5 + 1,96\sqrt{10} \text{ atau } 87,30 \le Y_{40} \le 99,70$$

Jika parameter β_0 dan β_1 tidak diketahui, kita gunakan penaksir b_0 dan b_1. Selang peramalan di atas tidak dapat kita gunakan karena kita hanya menggunakan taksiran dari $E(Y_h)$. Dalam penaksiran untuk $E(Y_h)$, kita gunakan selang kepercayaan sehingga letak distribusi dari Y tidak dapat ditentukan dengan tepat.

Dengan menganggap pengamatan baru independen terhadap sampel yang digunakan untuk menaksir garis regresi, maka variansi dari $Y_{h(baru)}$ terdiri dari variansi distribusi sampling $\hat{Y}_h$ dan variansi distribusi Y pada $X = X_h$, yaitu:

$$\sigma^2\left(Y_{h(baru)}\right) = \sigma^2\left(\hat{Y}_h - Y\right) = \sigma^2\left(\hat{Y}_h\right) + \sigma^2\left(Y\right)$$

Penaksir tak bias dari $\sigma^2\left(Y_{h(baru)}\right)$ adalah:

$$s^2\left(Y_{h(baru)}\right) = s^2\left(\hat{Y}_h\right) + KRS = KRS\left[1 + \frac{1}{n} + \frac{\left(X_h - \bar{X}\right)^2}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}\right]$$

Contoh 1.13

Dari data pada Contoh 1.6 diperoleh:

$\bar{X} = 3,3667$; $\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 = 13,1$; $KRS = 0,289$; $n = 9$; $b_0 = 0,2568$; $b_1 = 2,93$

Sehingga untuk $X_h = 2$ diperoleh $\hat{Y}_h = 0,2568 + 2,93(2) = 6,117$, $t_{(0,025;\,7)} = 2,365$; dan $s^2\left(\hat{Y}_2\right) = 0,289\left[1 + \frac{1}{9} + \frac{(2 - 3,3667)^2}{13,1}\right] = 0,362$.

Selang peramalan 95% untuk $Y_{h(baru)}$ adalah:

$$6,117 - 2,365\sqrt{0,362} \le Y_{h(baru)} \le 6,117 + 2,365\sqrt{0,362} \text{ atau } 4,693 \le Y_{h(baru)} \le 7,541$$

Ramalan Mean dari *m* Pengamatan Baru

Sering kali diinginkan meramal mean dari m pengamatan baru untuk taraf variabel bebas, X tertentu. Nilai ramalan mean dari Y ditulis sebagai $\bar{Y}_{h(baru)}$. Batas ramalan $(1-\alpha)100\%$ untuk $\bar{Y}_{h(baru)}$ adalah:

$$\hat{Y}_h \pm t_{\left(\frac{\alpha}{2};\, n-2\right)} s\left(\bar{Y}_{h(baru)}\right)$$

dengan

$$s^2\left(\bar{Y}_{h(baru)}\right) = s^2\left(\hat{Y}_h\right) + \frac{KRS}{m}$$

atau ekuivalen dengan

$$s^2\left(\bar{Y}_{h(baru)}\right) = KRS\left[\frac{1}{m}+\frac{1}{n}+\frac{\left(X_h-\bar{X}\right)^2}{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2}\right]$$

Dari rumus diketahui bahwa $s^2\left(\bar{Y}_{h(baru)}\right)$ mempunyai dua komponen yaitu variansi distribusi sampling dari $\bar{Y}_h$ dan variansi mean *m* pengamatan dari distribusi probabilitas *Y* pada $X = X_h$.

Contoh 1.14

Kita tinjau kembali Contoh 1.6. Jika $m=3$, untuk $X_h = 2$ akan dihitung selang ramalan 90% untuk $\bar{Y}_{h(baru)}$. Dari data yang ada telah dihitung:

$\hat{Y}_h = 6{,}117$; $s^2\left(\hat{Y}_h\right) = 0{,}362$; $KRS = 0{,}289$; $s^2\left(\bar{Y}_{h(baru)}\right) = 0{,}362 + \frac{0{,}289}{3} = 0{,}458$ atau $s\left(\bar{Y}_{h(baru)}\right) = 0{,}677$ dan $t_{(0{,}05;\,7)} = 1{,}895$.

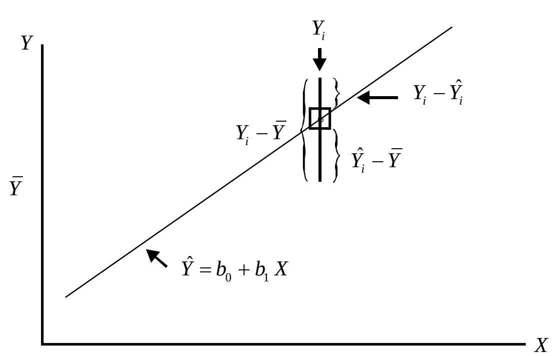
Sehingga selang ramalan 90% untuk $\bar{Y}_{h(baru)}$ adalah:

$$6{,}117 - 1{,}895(0{,}677) \le \bar{Y}_{h(baru)} \le 6{,}117 + 1{,}895(0{,}677) \text{ atau } 4{,}834 \le \bar{Y}_{h(baru)} \le 7{,}399$$

Selang ramalan $\bar{Y}_{h(baru)}$ lebih sempit dibandingkan dengan selang ramalan untuk $Y_{h(baru)}$ karena selang ramalan untuk $Y_{h(baru)}$ memuat $\bar{Y}_{h(baru)}$.

Pendekatan Analisis Variansi untuk Analisis Regresi

Dasar pendekatan analisis variansi adalah pemecahan jumlah kuadrat dan derajat bebas yang berkaitan dengan *Y*. Pandang simpangan total $\left(Y_i - \bar{Y}\right)$ pada Gambar 1.4 berikut ini.



Gambar 1.4.

Pemecahan simpangan:

$$\underbrace{Y_i - \bar{Y}}_{\text{simpangan total}} = \underbrace{\hat{Y}_i - \bar{Y}}_{\substack{\text{simpangan nilai regresi}\\ \text{taksiran sekitar mean}}} + \underbrace{Y_i - \hat{Y}_i}_{\substack{\text{simpangan sekitar}\\ \text{garis regresi taksiran}}}$$

sehingga pemecahan *JK* menjadi:

$$\underbrace{\sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2}_{\substack{\text{jumlah kuadrat}\\ \text{sekitar mean}}} = \underbrace{\sum_{i=1}^{n}\left(\hat{Y}_i - \bar{Y}\right)^2}_{\substack{\text{jumlah kuadrat}\\ \text{karena regresi}}} + \underbrace{\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2}_{\substack{\text{jumlah kuadrat}\\ \text{sekitar regresi}}}$$

atau

$$JK = JKR + JKS$$

JK adalah jumlah kuadrat total, $JK = 0$ berarti semua pengamatan sama. Makin besar nilai *JK* makin besar variasi (keragaman) nilai-nilai Y_i. *JKR* adalah jumlah kuadrat regresi merupakan ukuran keragaman pengamatan yang diterangkan oleh garis regresi. *JKS* adalah jumlah kuadrat sesatan merupakan ukuran keragaman pengamatan terhadap garis regresinya. Jika *JKS* = 0, semua pengamatan jatuh pada garis regresi. Rumus-rumus untuk menghitung jumlah kuadrat adalah:

$$JK = \sum_{i=1}^{n} Y_i^2 - \frac{\left(\sum_{i=1}^{n} Y_i\right)^2}{n} = \sum_{i=1}^{n} Y_i^2 - n\bar{Y}^2 = S_{YY}$$

$$JKR = b_1\left(\sum_{i=1}^{n} X_i Y_i - \frac{\sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{n}\right) = b_1 S_{XY}$$

$$= b_1^2 \sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 = b_1^2\, S_{XX}$$

$$JKS = \left(\sum_{i=1}^{n} Y_i^2 - \frac{\left(\sum_{i=1}^{n} Y_i\right)^2}{n}\right) - \frac{\left(\sum_{i=1}^{n} X_i Y_i - \frac{\sum_{i=1}^{n} X_i Y_i}{n}\right)^2}{\left(\sum_{i=1}^{n} X_i^2 - \frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n}\right)} = S_{YY} - \frac{\left(S_{XY}\right)^2}{S_{XX}}$$

Pemecahan derajat bebas sesuai dengan pemecahan jumlah kuadrat total. Derajat bebas untuk *JK* sama dengan $n-1$. Satu derajat bebas hilang karena simpangan $Y_i - \bar{Y}$ tidak independen, yakni $\sum_{i=1}^{n}\left(Y_i - \bar{Y}\right) = 0$ atau ekuivalen satu derajat bebas hilang karena mean sampel $\bar{Y}$ digunakan untuk menaksir mean pupolasi μ. *JKS* mempunyai derajat bebas $n-2$ atau dua derajat bebas hilang karena dalam menghitung nilai $\hat{Y}_i$ digunakan taksiran untuk parameter β_0 dan β_1. *JKR* mempunyai derajat bebas satu. Ada dua parameter dalam persamaan regresi, tetapi simpangan $Y_i - \bar{Y}$ tidak bebas, yakni $\sum_{i=1}^{n}\left(\hat{Y}_i - \bar{Y}\right) = 0$ sehingga satu derajat bebas hilang. Secara singkat, derajat bebas dari masing-masing jumlah kuadrat adalah $n-1=1+(n-2)$ dan tabel analisis variansi (ANAVA) untuk regresi linear sederhana diberikan oleh Tabel 1.1 berikut.

Tabel 1.1. ANAVA untuk Regresi Linear Sederhana

Sumber Variasi	Derajat Bebas (db)	Jumlah Kuadrat (JK)	Kuadrat Rata-rata (KR)
Regresi	1	*JKR*	*KRR*
Sesatan	$n-2$	*JKS*	*KRS*
Total	$n-1$	*JK*	

Harapan Kuadrat Rata-rata

Agar dapat melakukan inferensi berdasarkan pendekatan analisis variansi, kita perlu mengetahui nilai harapan dari kuadrat rata-rata, yaitu:

$$E(KRS) = \sigma^2$$

$$E(KRR) = \sigma^2 + \beta_1^2 \sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$$

Bukti:

Untuk model regresi dengan sesatan normal $\frac{JKS}{\sigma^2} \sim \chi^2_{(n-2)}$, maka

$$E\left(\frac{JKS}{\sigma^2}\right) = n-2$$

atau

$$E\left(\frac{JKS}{n-2}\right) = E(KRS) = \sigma^2$$

Untuk mendapatkan $E(KRR)$, kita tulis *JKR* dalam bentuk sebagai berikut.

$$JKR = b_1^2 \sum_{i=1}^{n} \left(X_i - \bar{X} \right)^2$$

Karena

$$\begin{aligned} \sigma^2(b_1) &= E\left[b_1 - E(b_1) \right]^2 \\ &= E\left(b_1^2 \right) - \left[E(b_1) \right]^2 \\ &= E\left(b_1^2 \right) - \beta_1^2 \end{aligned}$$

atau

$$E\left(b_1^2 \right) = \sigma^2(b_1) + \beta_1^2 = \frac{\sigma^2}{\sum_{i=1}^{n} \left(X_i - \bar{X} \right)^2} + \beta_1^2$$

maka

$$\begin{aligned} E(JKR) &= E\left[b_1^2 \sum_{i=1}^{n} \left(X_i - \bar{X} \right)^2 \right] \\ &= E\left(b_1^2 \right) \sum_{i=1}^{n} \left(X_i - \bar{X} \right)^2 \\ &= \left[\frac{\sigma^2}{\sum_{i=1}^{n} \left(X_i - \bar{X} \right)^2} + \beta_1^2 \right] \sum_{i=1}^{n} \left(X_i - \bar{X} \right)^2 \\ &= \sigma^2 + \beta_1^2 \sum_{i=1}^{n} \left(X_i - \bar{X} \right)^2 \end{aligned}$$

Jadi

$$E(KRR) = E\left(\frac{JKR}{1} \right) = \sigma^2 + \beta_1^2 \sum_{i=1}^{n} \left(X_i - \bar{X} \right)^2$$

Uji Hipotesis

Pendekatan analisis variansi dengan uji *F* dapat dilakukan untuk menguji hipotesis:

$$H_0 : \beta_1 = 0 \text{ terhadap } H_1 : \beta_1 \neq 0$$

Pada regresi linear sederhana dan statistik ujinya adalah:

$$F^* = \frac{KRR}{KRS}$$

Jika $\beta_1 = 0$ maka semua Y_i mempunyai mean sama yaitu $\mu = \beta_0$ dan variansi sama yaitu σ^2. Variabel $\frac{JKS}{\sigma^2}$ dan $\frac{JKR}{\sigma^2}$ adalah independen dan berdistribusi χ^2 sehingga, statistik uji F^* dapat ditulis dalam bentuk:

$$F^* = \frac{\frac{JKR}{\sigma^2} \Big/ 1}{\frac{JKS}{\sigma^2} \Big/ n-2} = \frac{KRR}{KRS}$$

Jika H_0 benar, maka

$$F^* \sim \frac{\chi^2_{(1)} \big/ 1}{\chi^2_{(n-2)} \big/ n-2}$$

Karena $\chi^2_{(1)}$ dan $\chi^2_{(n-2)}$ independen sehingga $F^* \sim F_{(1,\, n-2)}$. Kesimpulan yang diperoleh yakni menerima H_0 jika $F^* \leq F_{(\alpha;\, 1;\, n-2)}$ dan menolak H_0 jika $F^* > F_{(\alpha;\, 1;\, n-2)}$ dengan $F_{(\alpha;\, 1;\, n-2)}$ adalah persentase $(1-\alpha)100$ dari distribusi *F*.

Contoh 1.15

Diketahui data sebagai berikut:

X	1	1	2	3	4	4	5	6	6	7
Y	2,1	2,5	3,1	3	3,8	3,2	4,3	3,9	4,4	4,8

Jika digunakan model regresi linear sederhana $Y = \beta_0 + \beta_1 X + \varepsilon$ dengan sesatan normal $\varepsilon_i \sim N(0,\, \sigma^2)$ independen, maka untuk menguji $H_0 : \beta_1 = 0$ terhadap $H_1 : \beta_1 \neq 0$ dengan uji *F*, dari data kita hitung:

$$\sum_{i=1}^{n} X_i^2 - \frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n} = 193 - \frac{(39)^2}{10} = 40,9$$

$$\sum_{i=1}^{n} Y_i^2 - \frac{\left(\sum_{i=1}^{n} Y_i\right)^2}{n} = 130,05 - \frac{(35,1)^2}{10} = 6,85$$

$$\sum_{i=1}^{n} X_i Y_i - \frac{\sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{n} = 152,7 - \frac{(39)(35,1)}{10} = 15,81$$

sehingga

$$JK = S_{YY} = \sum_{i=1}^{n} Y_i^2 - \frac{\left(\sum_{i=1}^{n} Y_i\right)^2}{n} = 6,85$$

$$b_1 = \frac{S_{XY}}{S_{XX}} = \frac{\sum_{i=1}^{n} X_i Y_i - \frac{\sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{n}}{\sum_{i=1}^{n} X_i^2 - \frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n}} = \frac{15,81}{40,9} = 0,387$$

$$JKR = b_1^2 S_{XX} = b_1^2 \left(\sum_{i=1}^{n} X_i^2 - \frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n} \right) = (0,387)^2 (40,9) = 6,13$$

$$JKS = JK - JKR = 6,85 - 6,13 = 0,72$$

Selanjutnya hasil hitungan, kita susun dalam tabel ANAVA sebagai berikut.

Tabel 1.2. ANAVA

Sumber Variasi	db	JK	KR	F
Regresi Sesatan	1 8	6,13 0,72	6,13 0,09	$F^* = \frac{6,13}{0,09} = 68,1$
Total	9	6,85		

Untuk $\alpha = 0,05$, dari tabel distribusi F diperoleh nilai $F_{(0,05;\,1,8)} = 5,32$, sehingga diperoleh kesimpulan menolak H_0 karena $F^* > 5,32$, artinya ada hubungan linear antara Y dan X atau garis regresi nyata.

Kesetaraan Uji *F* dan Uji *t*

Pada taraf nyata α, uji F untuk hipotesis $H_0 : \beta_0 = 0$ terhadap $H_1 : \beta_1 \neq 0$ secara aljabar setara dengan uji t dua sisi. Hal ini dapat kita tunjukkan sebagai berikut.

$$F^* = \frac{JKR/1}{JKS/n-2} = \frac{b_1^2 \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2}{KRS}$$

Karena diketahui $s^2(b_1) = \dfrac{KRS}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$, maka kita peroleh $F^* = \dfrac{b_1^2}{s^2(b_1)} = \left[\dfrac{b_1}{s(b_1)}\right]^2$

Statistik t^* untuk uji $H_0: \beta_1 = 0$ terhadap $H_1: \beta_1 \neq 0$ adalah:

$$t^* = \frac{b_1}{s(b_1)}$$

atau

$$\left(t^*\right)^2 = \left(\frac{b_1}{s(b_1)}\right)^2 = F^*$$

Dari hubungan antara t^* dan F^*, kita peroleh hubungan antara persentil distribusi *t* dan *F* sebagai

$$\left(t_{\left(\frac{\alpha}{2};\, n-2\right)}\right)^2 = F_{(\alpha;\, 1;\, n-2)}$$

Contoh 1.16

Kita tinjau kembali Contoh 1.15. Dari data yang ada diperoleh $F^* = 68{,}1$; $b_1 = 0{,}387$; $s^2(b_1) = \dfrac{0{,}09}{40{,}9} = 0{,}0022$; $s(b_1) = 0{,}047$ sehingga $\left(t^*\right)^2 = \left(\dfrac{0{,}387}{0{,}047}\right)^2 = 68{,}06 \approx 68{,}1$ dan $\left(t_{(0{,}025;\,8)}\right)^2 = (2{,}306)^2 = 5{,}32 = F_{(0{,}05;\,1;8)}$.

Pendekatan Uji Linear Umum

Analisis variansi untuk menguji $H_0: \beta_1 = 0$ terhadap $H_1: \beta_1 \neq 0$ merupakan pengujian untuk model statistik linear. Akan kita pelajari pendekatan uji (linear) secara umum, yang dapat kita gunakan untuk pengujian model-model linear. Tiga langkah dalam pengujian ini adalah:

1. Menentukan model yang dipandang sesuai untuk data. Model ini kita sebut *model lengkap*. Untuk regresi linear sederhana, model lengkap:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

   Penaksir parameter dihitung dengan metode kuadrat terkecil. Kemudian kita hitung jumlah kuadrat sesatan untuk model lengkap, yang ditulis dengan notasi $JKS(L)$ sebagai:

$$JKS(L) = \sum_{i=1}^{n}\left(Y_i - (b_0 + b_1 X_i)\right)^2 = \sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2$$

$JKS(L)$ mengukur keragaman pengamatan Y_i sekitar garis regresi.

2. Perhatikan hipotesis $H_0 : \beta_1 = 0$ dan $H_1 : \beta_1 \neq 0$ model di bawah H_0 disebut *model tereduksi*. Maksudnya jika $\beta_1 = 0$, model lengkap tereduksi menjadi $Y_i = \beta_0 + \varepsilon_i$. Penaksir parameter untuk model tereduksi dihitung dengan metode kuadrat terkecil, didapat $b_0 = \overline{Y}$ dan jumlah kuadrat sesatan untuk model tereduksi, adalah:

$$JKS(R) = \sum_{i=1}^{n}\left(Y_i - b_0\right)^2 = \sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2 = JK$$

3. Sekarang kita bandingkan $JKS(L)$ dengan $JKS(R)$. Tampak bahwa $JKS(L) \leq JKS(R)$. Apabila $JKS(R) - JKS(L)$ terkecil maka kita cenderung untuk menerima H_0 dan jika sebaliknya maka H_0 kita tolak. Dengan menggunakan statistik uji:

$$F^* = \frac{\left(JKS(R) - JKS(L)\right)/\left(db_R - db_L\right)}{JKS(L)/db_L}$$

Diperoleh kesimpulan menerima H_0 jika $F^* \leq F_{(\alpha;\, db_R - db_L,\, db_L)}$ dan menolak H_0 jika $F^* \leq F_{(\alpha;\, db_R - db_L,\, db_L)}$.

Untuk pengujian hipotesis $H_0 : \beta_1 = 0$ terhadap $H_1 : \beta_1 \neq 0$, kita peroleh $JKS(R) = JK$ dengan $d.b = n-1$; $JKS(L) = JKS$ dengan $d.b = n-2$; dan statistik uji

$$F^* = \frac{(JK - JKS)/\left((n-1)-(n-2)\right)}{JKS/(n-2)} = \frac{JKR/1}{JKS/(n-2)} = \frac{KRR}{KRS}$$

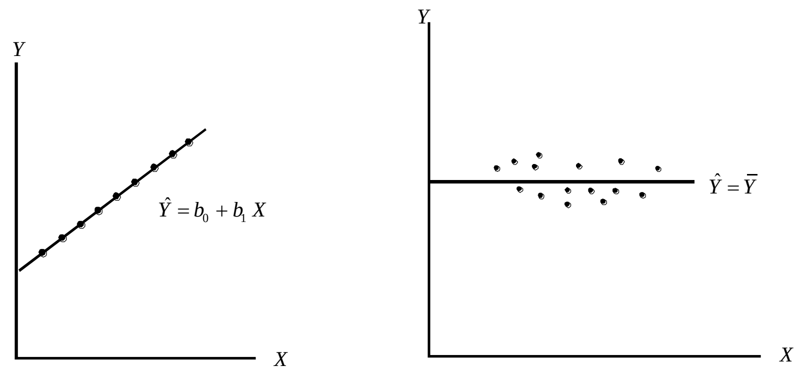
Jadi F^* sama dengan statistik uji analisis variansi.

Ukuran Deskriptif untuk Hubungan antara *X* dan *Y*

Dalam taksiran dan ramalan, untuk mengukur ketepatan hanya panjang interval yang diperhatikan, tetapi derajat hubungan variabel *X* dan variabel *Y* tidak diperhatikan. Sebagai ukuran deskriptif, dapat digunakan koefisien determinasi dan koefisien korelasi. Koefisien determinasi didefinisikan sebagai

$$r^2 = \frac{JK - JKS}{JK} = \frac{JKR}{JK} = 1 - \frac{JKS}{JK}$$

Karena $0 \le JKS \le JK$ maka $0 \le r^2 \le 1$. Makin besar harga r^2, berarti makin besar pengaruh variabel *X* terhadap variasi harga *Y*.



(a) $r^2 = 1$ (b) $r^2 = 0$

Gambar 1.5.

Apabila semua observasi terletak pada garis regresinya, maka $JKS = 0$ sehingga didapat $r^2 = 1 - \frac{0}{JK} = 1$. Apabila $b_1 = 0$ berarti $\hat{Y}_i = \bar{Y}$ dan $JKR = \sum_{i=1}^{n}\left(\hat{Y}_i - \bar{Y}\right)^2 = 0$ atau $JKS = JK$ sehingga didapat $r^2 = 0$. Koefisien korelasi didefinisikan sebagai $r = \pm\sqrt{r^2}$. Tanda dari koefisien korelasi sesuai dengan tanda dari b_1. Nilai dari r selalu terletak dalam $[-1, 1]$ atau $-1 \le r \le 1$. Untuk menghitung koefisien korelasi digunakan rumus sebagai berikut.

$$r = \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{\left[\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 \sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2\right]^{\frac{1}{2}}}$$

$$= \frac{\left[\sum_{i=1}^{n} X_i Y_i - \frac{\sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{n}\right]}{\left[\left(\sum_{i=1}^{n} X_i^2 - \frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n}\right)\left(\sum_{i=1}^{n} Y_i - \frac{\left(\sum_{i=1}^{n} Y_i\right)^2}{n}\right)\right]^{\frac{1}{2}}}$$

Hubungan antara b_1 dan r adalah:

$$b_1 = \left[\frac{\sum_{i=1}^{n} \left(Y_i - \bar{Y} \right)^2}{\sum_{i=1}^{n} \left(X_i - \bar{X} \right)^2} \right]^{\frac{1}{2}} r$$

Jika $r = 0$ maka $b_1 = 0$ dan sebaliknya jika nilai r dan r^2 mendekati 1 maka dapat dilakukan inferensi yang cukup tepat terhadap *Y*.

Contoh 1.17

Diketahui data pengamatan sampel sebagai berikut.

X	1,5	1,8	2,4	3,0	3,5	3,9	4,4	4,8	5,0
Y	4,8	5,7	7,0	8,3	10,9	12,4	13,1	13,6	15,3

Akan dihitung nilai r^2 dan r dari data tersebut dan diperoleh:

$$\sum_{i=1}^{n} X_i^2 - \frac{\left(\sum_{i=1}^{n} X_i \right)^2}{n} = 115{,}11 - \frac{(30{,}3)^2}{9} = 13{,}10$$

$$\sum_{i=1}^{n} Y_i^2 - \frac{\left(\sum_{i=1}^{n} Y_i \right)^2}{n} = 1036{,}65 - \frac{(91{,}1)^2}{9} = 114{,}52$$

$$\sum_{i=1}^{n} X_i Y_i - \frac{\sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{n} = 345{,}09 - \frac{(30{,}3)(91{,}1)}{9} = 38{,}39$$

Selanjutnya kita hitung

$$b_1 = \frac{\sum_{i=1}^{n} \left(X_i - \bar{Y} \right)\left(Y_i - \bar{X} \right)}{\sum_{i=1}^{n} \left(X_i - \bar{X} \right)^2} = \frac{38{,}39}{13{,}10} = 2{,}9303$$

$$JK = 114{,}52$$

$$JKS = 114{,}52 - (2{,}9303)(38{,}39) = 2{,}026$$

$$r^2 = \frac{JK - JKS}{JK} = \frac{114{,}52 - 2{,}026}{114{,}52} = 0{,}9823$$

Hal ini berarti 98,23 persen variasi-variasi variabel *Y* disebabkan oleh hubungannya dengan variabel *X*. Sedangkan koefisien korelasi $r = \sqrt{r^2} = \sqrt{0{,}9823} = 0{,}99$ (tanda +

sesuai dengan tanda b_1) atau dengan menggunakan rumus untuk koefisien korelasi didapat $r = \dfrac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{\left[\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 \sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2\right]^{\frac{1}{2}}} = \dfrac{38,39}{\left[(13,10)(114,52)\right]^{\frac{1}{2}}} = 0,99$.

LATIHAN

Untuk memperdalam pemahaman Anda mengenai materi di atas, silakan Anda mengerjakan latihan berikut ini!

1) Jika penduga kuadrat terkecil b_1 merupakan linear dari Y_i, yakni:

$$b_1 = \sum_{i=1}^{n} k_i Y_i \quad \text{dengan} \quad k_i = \frac{X_i - \bar{X}}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

Tunjukkan bahwa $\sum_{i=1}^{n} k_i Y_i = 1$ dan $\sum_{i=1}^{n} k_i^2 = \dfrac{1}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}$

2) Data dalam tabel berikut menunjukkan periode waktu yang digunakan dalam suatu percobaan pengeringan bahan (X) dalam jam dan berkurangnya berat bahan (Y) dalam gram.

X	4,4	4,5	4,8	5,5	5,7	5,9	6,3	6,9	7,5	7,8
Y	13,1	11,5	10,8	13,8	15,1	12,7	12,7	13,8	17,6	18,8

Jika digunakan model regresi linear dengan sesatan normal,

a) carilah persamaan regresi taksiran
b) carilah selang kepercayaan 90% untuk β_1
c) ujilah $H_0 : \beta_1 = 0$ terhadap $H_1 : \beta_1 \neq 0$ pada tingkat signifikansi $\alpha = 10\%$
d) carilah selang kepercayaan 95% untuk $E\left(Y_7\right)$
e) untuk $X_h = 8$, carilah selang ramalan 90% untuk Y_8
f) untuk $X_h = 8$, m = 5, carilah selang ramalan 90% mean dari m observasi baru, yakni $\bar{Y}_{8(baru)}$
g) hitunglah koefisien determinasi r^2

3) Diketahu tabel ANAVA sebagai berikut.

Sumber Variasi	db	JK	KR
Regresi	1	40	-
Sesatan	50	-	-
Total	51	240	

a) Lengkapilah tabel ANAVA tersebut.
b) Uji $H_0 : \beta_1 = 0$ terhadap $H_1 : \beta_1 \neq 0$ pada tingkat signifikansi $\alpha = 0{,}05$
c) Hitunglah nilai koefisien determinasi dan kesimpulan apa yang diperoleh dari hasil perhitungan tersebut.

Petunjuk Jawaban Latihan

1) Lihat penjelasan sifat-sifat koefisien k_i .

2) Dari data yang ada diperoleh: $n = 10; \sum_{i=1}^{n} X_i = 59{,}3; \sum_{i=1}^{n} Y_i = 139{,}9; \sum_{i=1}^{n} X_i^2 = 364{,}59;$

$\sum_{i=1}^{n} Y_i^2 = 2015{,}17; \sum_{i=1}^{n} X_i Y_i = 852$

a) $$b_1 = \frac{n\sum_{i=1}^{n} X_i Y_i - \sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{n\sum_{i=1}^{n} X_i^2 - \left(\sum_{i=1}^{n} X_i\right)^2} = 1{,}7304 \quad \text{dan } b_0 = \bar{Y} - b_1\bar{X} = 3{,}7288$$

Persamaan regresi taksiran $\hat{Y}_i = 3{,}7288 + 1{,}7304 X_i$

b) $$JKS = \sum_{i=1}^{n} Y_i^2 - b_0 \sum_{i=1}^{n} Y_i - b_1 \sum_{i=1}^{n} X_i Y_i = 19{,}2203$$

$$s^2 = \frac{JKS}{n-2} = \frac{19{,}2203}{8} = 2{,}4025$$

$$s^2(b_1) = \frac{s^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = 0{,}1857 \text{ atau } s(b_1) = \sqrt{0{,}1857} = 0{,}4309$$

$$t_{\left(\frac{\alpha}{2};\, n-2\right)} = t_{(0{,}05;8)} = 1{,}860$$

Selang kepercayaan 90% untuk β_1 adalah:

$$1{,}7304 - (1{,}860)(0{,}4309) \leq \beta_1 \leq 1{,}7304 + (1{,}860)(0{,}4309) \text{ atau}$$

$$0{,}929 \leq \beta_1 \leq 2{,}532$$

c) Untuk uji $H_0 : \beta_1 = 0$ terhadap $H_1 : \beta_1 \neq 0$ diperoleh statistik uji:

$$t^* = \frac{b_1}{s(b_1)} = \frac{1,7304}{0,4309} = 4,0158$$

Karena $\left| t^* \right| > t_{\left(\frac{\alpha}{2};n-2\right)}$ maka H_0 ditolak

d) $\hat{Y}_7 = 3,7288 + 1,7304(7) = 15,8415$

$$s^2\left(\hat{Y}_7\right) = s^2\left[\frac{1}{n} + \frac{\left(7-\bar{X}\right)^2}{\sum_{i=1}^{n}(X_i-\bar{X})^2}\right] = (2,4025)\left[\frac{1}{10} + \frac{(7-5,93)^2}{12,941}\right] = 0,4528\,;$$

$s\left(\hat{Y}_7\right) = 0,6729$ dan $t_{(0,025;8)} = 2,306$

Selang kepercayaan 95% untuk $E(Y_7)$ adalah:

$$15,8415 - (2,306)(0,6729) \le E(Y_7) \le 15,8415 + (2,306)(0,6729)$$

$$14,29 \le E(Y_7) \le 17,39$$

e) $\hat{Y}_8 = 3,7288 + 1,7304(8) = 17,5719$

$$s^2\left(Y_{8(baru)}\right) = s^2\left[1 + \frac{1}{n} + \frac{\left(8-\bar{X}\right)^2}{\sum_{i=1}^{n}(X_i-\bar{X})^2}\right] = (2,4025)\left[1 + \frac{1}{10} + \frac{(8-5,93)^2}{12,941}\right] = 3,4382$$

$s\left(Y_{8(baru)}\right) = 1,8542$

Selang ramalan 90% untuk $Y_{8(baru)}$ adalah:

$$17,5719 - (1,86)(1,8542) \le Y_{8(baru)} \le 17,5719 + (1,86)(1,8542)$$

$$14,123 \le Y_{8(baru)} \le 21,021$$

f) $$s^2\left(\bar{Y}_{8(baru)}\right) = s^2\left[\frac{1}{m} + \frac{1}{n} + \frac{\left(8-\bar{X}\right)^2}{\sum_{i=1}^{n}(X_i-\bar{X})^2}\right] = (2,4025)\left[\frac{1}{5} + \frac{1}{10} + \frac{(8-5,93)^2}{12,941}\right] = 1,5162$$

$s\left(\bar{Y}_{8(baru)}\right) = 1,2314$

Selang ramalan 90% untuk $\bar{Y}_{8(baru)}$ adalah:

$$17,5719 - (1,86)(1,2314) \le \bar{Y}_{8(baru)} \le 17,5719 + (1,86)(1,2314) \text{ atau}$$

$$15,2816 \le \bar{Y}_{8(baru)} \le 19,8622$$

g) Koefisien determinasi $r^2 = b_1^2 \dfrac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2} = 0,668$

3) a) Tabel ANAVA

Sumber Variasi	db	JK	KR
Regresi	1	40	40
Sesatan	50	200	4
Total	51	240	

b) Statistik Uji: $F^* = \dfrac{KRR}{KRS} = \dfrac{40}{4} = 10$

Dari tabel diperoleh $F_{(1;50;0,05)} = 4,04$, karena $F^* > 4,04$ maka diperoleh kesimpulan tolak H_0.

c) Koefisien determinasi $r^2 = \dfrac{JKR}{JK} = \dfrac{40}{240} = 0,167$

Artinya 16,7% variasi variabel *Y* disebabkan oleh hubungannya dengan variabel *X*.

RANGKUMAN

1. Untuk dapat melakukan inferensi dalam analisis regresi, terutama jika sampel yang digunakan kecil diperlukan asumsi $\varepsilon_i \sim N\left(0, \sigma^2\right)$.

2. Inferensi untuk β_1 dan β_0 dilakukan dengan menghitung selang kepercayaan $(1-\alpha)100\%$ dari β_0 dan β_1, yakni:

$$b_1 - t_{\left(\frac{\alpha}{2};n-2\right)}\, s\left(b_1\right) \le \beta_1 \le b_1 - t_{\left(\frac{\alpha}{2};n-2\right)}\, s\left(b_1\right)$$

$$b_0 - t_{\left(\frac{\alpha}{2};n-2\right)}\, s\left(b_0\right) \le \beta_0 \le b_0 + t_{\left(\frac{\alpha}{2};n-2\right)}\, s\left(b_0\right)$$

dengan

$$s(b_1) = \frac{KRS}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \frac{s^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

$$s(b_0) = KRS\left[\frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\right]$$

3. Untuk $X = X_h$, inferensi untuk mean respons $E(Y_h)$ dilakukan dengan menghitung selang kepercayaan $(1-\alpha)100\%$, yakni

$$\hat{Y}_h - t_{\left(\frac{\alpha}{2};n-2\right)}\, s(\hat{Y}_h) \le E(Y_h) \le \hat{Y}_h - t_{\left(\frac{\alpha}{2};n-2\right)}\, s(\hat{Y}_h)$$

dengan

$$\hat{Y}_h = b_0 + b_1 X_h \text{ dan } s(\hat{Y}_h) = KRS\left[\frac{1}{n} + \frac{(X_h - \bar{X})^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\right]$$

4. Dalam peramalan observasi baru Y_h, jika parameter β_0, β_1, σ diketahui dan $E(Y_h) = \beta_0 + \beta_1 X_h$, maka selang ramalan $(1-\alpha)100\%$ untuk Y_h adalah:

$$E(Y_h) - z_{\frac{\alpha}{2}}\,\sigma \le Y_{h(baru)} \le E(Y_h) + z_{\frac{\alpha}{2}}\,\sigma,.$$

Jika parameter-parameter β_0, β_1 dan σ tidak diketahui, digunakan selang ramalan $(1-\alpha)100\%$:

$$\hat{Y}_h - t_{\left(\frac{\alpha}{2};n-2\right)}\, s(Y_{h(baru)}) \le \hat{Y}_{h(baru)} \le Y_h + t_{\left(\frac{\alpha}{2};n-2\right)}\, s(Y_{h(baru)})$$

dengan

$$\hat{Y}_h = b_0 + b_1 X_h \text{ dan } s^2(Y_{h(baru)}) = KRS\left[1 + \frac{1}{n} + \frac{(X_h - \bar{X})^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\right]$$

Jika ada *m* pengamatan baru, digunakan selang ramalan $(1-\alpha)100\%$

$$\hat{Y}_h - t_{\left(\frac{\alpha}{2};n-2\right)}\, s(\bar{Y}_{h(baru)}) \le \bar{Y}_{h(baru)} \le \hat{Y} + t_{\left(\frac{\alpha}{2};n-2\right)}\, s(\bar{Y}_{h(baru)}).$$

dengan

$$s^2\left(\bar{Y}_{h(baru)}\right) = KRS\left[\frac{1}{m} + \frac{1}{n} + \frac{\left(X_h - \bar{X}\right)^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\right]$$

5. Sebagai ukuran deskriptif, didefinisikan koefisien determinasi

$$r^2 = \frac{JK - JKS}{JK} = \frac{JKR}{JK} = 1 - \frac{JKS}{JK}; \quad 0 \le r^2 \le 1$$

yang mengukur besarnya pengaruh variabel X terhadap variabel Y.
Ukuran lain adalah koefisien korelasi, yakni:

$$r = \frac{\sum_{i=1}^{n} X_i Y_i - \dfrac{\sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{n}}{\sqrt{\left[\sum_{i=1}^{n} X_i^2 - \dfrac{\left(\sum_{i=1}^{n} X_i\right)^2}{n}\right]\left[\sum_{i=1}^{n} Y_i^2 - \dfrac{\left(\sum_{i=1}^{n} Y_i\right)^2}{n}\right]}} \quad ; -1 \le r \le 1$$

Tanda dari r sesuai dengan tanda dari b_1.

TES FORMATIF 2

Pilih satu jawaban yang paling tepat dari beberapa alternatif jawaban yang disediakan!

1) Dari pasangan data X (variabel independen) dan Y (variabel respons), dihitung nilai sebagai berikut.

   $$\sum_{i=1}^{n} X_i^2 = 121, \ \sum_{i=1}^{n} X_i = 20; \ \sum_{i=1}^{n} X_i Y_i = -82; \ \sum_{i=1}^{n} Y_i^2 = 516; \ \sum_{i=1}^{n} Y_i = 40; \ n = 40$$

   Jika digunakan model regresi linear sederhana, maka nilai variansi s^2 sama dengan ….
   A. 4,48
   B. 4,00
   C. 4,49
   D. 3,89

2) **Lihat soal nomor 1**, nilai $s^2(b_1)$ sama dengan ….
   A. 0,04938
   B. 0,05531
   C. 0,04310
   D. 0,04792

3) **Lihat soal nomor 1**, selang kepercayaan 95% untuk β_1 adalah ….
   A. $-2,495 \leq \beta_1 \leq -1,505$
   B. $-2,512 \leq \beta_1 \leq -1,488$
   C. $-2,437 \leq \beta_1 \leq -1,563$
   D. $-2,402 \leq \beta_1 \leq -1,598$

4) Data berikut menunjukkan jumlah senyawa kimia yang larut dalam 100 gram air (*Y*) pada berbagai temperatur (*X*) dalam $\left({}^{0}C\right)$.

X	0	15	30	45	50	75
Y	8	12	24	33	39	44

   Dari data dapat dihitung nilai s^2 sama dengan ….
   A. 4,38
   B. 5,48
   C. 3,64
   D. 4,84

5) **Lihat soal nomor 4**, nilai $s^2(b_1)$ sama dengan ….
   A. 0,0011
   B. 0,0093
   C. 0,0014
   D. 0,0012

6) **Lihat soal nomor 4**, untuk uji $H_0 : \beta_1 = 0$ terhadap $H_1 : \beta_1 \neq 0,$ statistik penguji t^* sama dengan ….
   A. 17,934
   B. 13,794
   C. 12,794
   D. 4,794

7) **Lihat soal nomor 4,** untuk uji $H_0 : \beta_1 = 0$ terhadap $H_1 : \beta_1 \neq 0$ dengan analisis variansi dihitung statistik penguji F^* sama dengan ….
   A. 285,84
   B. 190,17
   C. 0,9794
   D. 0,4948

8) **Lihat soal nomor 4,** selang kepercayaan 95% untuk $E(Y_k)$ yang bersesuai dengan $X_k = 35$ adalah ….
   A. $23,46 \leq E(Y_k) \leq 27,30$
   B. $23,71 \leq E(Y_k) \leq 27,05$
   C. $23,55 \leq E(Y_k) \leq 27,21$
   D. $23,33 \leq E(Y_k) \leq 27,43$

9) **Lihat soal nomor 4**, koefisien determinasi r^2 sama dengan ….
   A. 0,98
   B. 0,90
   C. 0,92
   D. 0,95

Cocokkanlah jawaban Anda dengan Kunci Jawaban Tes Formatif 2 yang terdapat di bagian akhir modul ini. Hitunglah jawaban Anda yang benar. Kemudian gunakan rumus di bawah ini untuk mengetahui tingkat penguasaan Anda terhadap materi Kegiatan Belajar 2.

Rumus:

$$\text{Tingkat penguasaan} = \frac{\text{Jumlah jawaban Anda yang benar}}{9} \times 100\%$$

Arti tingkat penguasaan yang Anda capai:
90 - 100% = baik sekali
80 - 89% = baik
70 - 79% = cukup
< 70% = kurang

Bila Anda mencapai tingkat penguasaan 80% atau lebih, Anda dapat meneruskan dengan modul selanjutnya. **Bagus!** Tetapi bila tingkat penguasaan Anda masih di bawah 80%, Anda harus mengulangi Kegiatan Belajar 2, terutama bagian yang belum Anda kuasai.

Kunci Jawaban Tes Formatif

Tes Formatif 1

1) A
2) C
3) B untuk $X_k = 175,\ \hat{Y}_k = b_0 + b_1 X_k$
4) D $JKS = \sum_{i=1}^{n}(Y_i - \bar{Y})^2 - \dfrac{\left[\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)\right]^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$ dan $s^2 = KRS = \dfrac{JKS}{n-2}$
5) B
6) A
7) C
8) B
9) B $b_1 = \dfrac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$
10) D

Tes Formatif 2

1) B hitung dengan rumus $s^2 = \dfrac{JKS}{n-2}$
2) A gunakan rumus $s^2\left(b_1\right) = \dfrac{s^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$
3) B
4) B
5) C

6) B statistik penguji $t^* = \dfrac{b_1}{s(b_1)}$

7) B statistik penguji $F^* = \dfrac{KRR}{KRS}$

8) D

9) A gunakan rumus $r^2 = \dfrac{JK - JKS}{JK} = \dfrac{JKR}{JK}$

Daftar Pustaka

Draper, N. & Smith, H. (1998). *Applied Regression Analysis*. New York: Wiley.

Montgomery, D.C. & Peck, E. A. (1992). *Introduction to Linear Regression Analysis.* New York: Wiley.

Neter, J., Wasserman, W. & Kutner, M. H. (1990). *Applied Linear Statistical Models*. Irwin.