BAB II

REGRESI LINIER SEDERHANA

Analisis Regresi adalah merupakan salah satu cabang dari metode statistik, yaitu bentuk hubungan antara variabel Y (variabel dependen atau respons) dengan variabel X_1, X_2, X_3, ... ,X_k (variabel independen atau prediktor).

Tujuan pokok dalam penggunaan metode ini adalah untuk meramalkan ata u memperkirakan nilai dari satu variabel dalam hubungannya dengan variabel yang lain yang diketahui.

Hubungan yang didapat pada umumnya dinyatakan dalam bentuk persamaan yang menyatakan hubungan fungsional.

Dalam bab ini membatasi pada masalah sederhana, bagaimana meramalkan Y sebagai fungsi linier dari suatu variabel tunggal.

II.1. Model Probabilistik

$$Y = \beta_0 + \beta_1 X + \varepsilon \qquad \text{........................} \quad (\ 2:1\ )$$

Dimana:

- Y adalah variabel dependen
- X adalah variabel independen
- β_0 adalah titik potong (intercept) dengan sumbu Y
- β_1 adalah kemiringan (slope atau gradien)
- ε adalah kesalahan acak (random error)

Asumsi-asumsi untuk Model Probabilistik

Untuk setiap nilai X tertentu, Y memiliki suatu distribusi normal, dengan nilai rerata (mean) yang diberikan oleh persamaan,

$$E(Y|X) = \beta_0 + \beta_1 X$$

dan dengan suatu varian σ^2. Selanjutnya, setiap nilai Y adalah independen (bebas) terhadap setiap nilai Y lainnya.

Untuk menemukan garis "kecocokan terbaik" (best - fitting) untuk serangkaian data tertentu yaitu jika nilai yang diramalkan Y dengan $\hat{Y}$, persamaan prediksi adalah,

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (2:2)$$

Dimana $\hat{\beta}_0$ dan $\hat{\beta}_1$ merupakan estimasi untuk β_0 dan β_1 yang sebenarnya.

II.2. Metode Kuadrat Terkecil (The Method of Least Squares)

Dalam teori regresi maka garis yang paling mewakili ialah garis yang dibuat sedemikian rupa sehingga total e - rror yang mungkin timbul dapat ditekan sekecil mungkin. Metode yang terbaik digunakan untuk memperkecil (meminim-mumkan) jumlah kuadrat dari error ialah Metode Least Squa-re yaitu,

$$SSE = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (2:3)$$

$$= \sum_{i=1}^{n} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2 \quad \ldots\ldots\ldots\ldots \quad (2:3a)$$

Agar SSE (Sum Square Error) minimum (terkecil) maka dengan membuat turunan parsial terhadap $\hat{\beta}_0$ dan $\hat{\beta}_1$, dan menyama-kannya dengan nol,

$$\frac{\partial SSE}{\partial \hat{\beta}_0} = 0$$

$$2 \sum_{i=1}^{n} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)(-1) = 0$$

$$\sum_{i=1}^{n} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) = 0$$

$$\sum_{i=1}^{n} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) = 0$$

$$\sum_{i=1}^{n} Y_i - n\hat{\beta}_0 - \hat{\beta}_1 \sum_{i=1}^{n} X_i = 0$$

$$\sum_{i=1}^{n} Y_i = n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} X_i \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (1)$$

$$\frac{\partial\, SSE}{\partial \hat{\beta}_1} = 0$$

$$2\sum_{i=1}^{n} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)(-X_i) = 0$$

$$\sum_{i=1}^{n} (X_i Y_i - \hat{\beta}_0 X_i - \hat{\beta}_1 X_i^2) = 0$$

$$\sum_{i=1}^{n} X_i Y_i - \hat{\beta}_0 \sum_{i=1}^{n} X_i - \hat{\beta}_1 \sum_{i=1}^{n} X_i^2 = 0$$

$$\sum_{i=1}^{n} X_i Y_i = \hat{\beta}_0 \sum_{i=1}^{n} X_i + \hat{\beta}_1 \sum_{i=1}^{n} X_i^2 \quad \ldots\ldots\ldots \quad (2)$$

Persamaan (1) dibagi dengan n sehingga,

$$\frac{\sum_{i=1}^{n} Y_i}{n} = \frac{n\hat{\beta}_0}{n} + \frac{\hat{\beta}_1 \sum_{i=1}^{n} X_i}{n}$$

$$\bar{Y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{X}$$

dimana,

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

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

Jadi, $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$ .... (2:4)

Substitusi $\hat{\beta}_0$ ke persamaan (2) ,sehingga :

$$\sum_{i=1}^{n} X_i Y_i = (\bar{Y} - \hat{\beta}_1 \bar{X}) \sum_{i=1}^{n} X_i + \hat{\beta}_1 \sum_{i=1}^{n} X_i^2$$

$$\sum_{i=1}^{n} X_i Y_i = \bar{Y} \sum_{i=1}^{n} X_i - \hat{\beta}_1 \bar{X} \sum_{i=1}^{n} X_i + \hat{\beta}_1 \sum_{i=1}^{n} X_i^2$$

$$= \frac{1}{n} \sum_{i=1}^{n} Y_i \sum_{i=1}^{n} X_i - \hat{\beta}_1 \frac{1}{n} \sum_{i=1}^{n} X_i \sum_{i=1}^{n} X_i + \hat{\beta}_1 \sum_{i=1}^{n} X_i^2$$

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

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

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

Jadi,

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} X_i Y_i - \frac{1}{n} \sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{\sum_{i=1}^{n} X_i^2 - \frac{1}{n} \left(\sum_{i=1}^{n} X_i \right)^2} \quad \ldots\ldots\ldots\ldots\ldots \quad (2:5)$$

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

$$= \frac{\sum_{i=1}^{n} X_i Y_i - \bar{X} \sum_{i=1}^{n} Y_i - \bar{Y} \sum_{i=1}^{n} X_i + \bar{Y} \sum_{i=1}^{n} X_i}{\sum_{i=1}^{n} X_i^2 - \bar{X} \sum_{i=1}^{n} X_i - \bar{X} \sum_{i=1}^{n} X_i + \bar{X} \sum_{i=1}^{n} X_i}$$

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

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} Y_i(X_i-\bar{X})-\bar{Y}(X_i-\bar{X})}{\sum_{i=1}^{n} X_i(X_i-\bar{X})-\bar{X}(X_i-\bar{X})}$$

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

$$= \frac{\sum_{i=1}^{n} (X_i-\bar{X})(Y_i-\bar{Y})}{\sum_{i=1}^{n} (X_i-\bar{X})^2} \quad \text{.....................} \quad (2:5a)$$

$$= \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2} \quad \text{.....................} \quad (2:5b)$$

$$= \frac{SS_{xy}}{SS_{xx}} \quad \text{.....................} \quad (2:5c)$$

dimana,

$$\sum_{i=1}^{n} X_i = n\bar{X}$$

$$x_i = X_i - \bar{X}$$

$$y_i = Y_i - \bar{Y}$$

$$SS_{xy} = \sum_{i=1}^{n} X_i Y_i - \frac{1}{n} \sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i$$

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

$$= \sum_{i=1}^{n} x_i y_i$$

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

$$SS_{xx} = \sum_{i=1}^{n} (X_i - \overline{X})^2$$

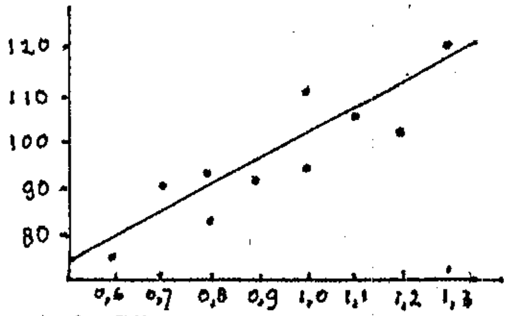
$$= \sum_{i=1}^{n} x_i^2$$

Sebagai ilustrasi, meninjau peramalan volume penjualan bruto bulanan, Y untuk suatu perusahaan yang volume penjualannya tidak banyak dipengaruhi oleh variasi musiman.

Tabel II.1. Pengeluaran untuk iklan dan volume penjualan untuk suatu perusahaan selama 10 bulan yang dipilih secara a cak (random)

Bulan	Pengeluaran untuk Iklan, X (xRp10.000.000)	Volume penjualan, Y (xRp10.000.000)
1	1,2	101
2	0,8	92
3	1,0	110
4	1,3	120
5	0,7	90
6	0,8	82
7	1,0	93
8	0,6	75
9	0,9	91
10	1,1	105

Pendekatan awal untuk melakukan analisis data adalah dengan memetakan data sebagai titik-titik (diagram pencar). Kemudian metode untuk memperoleh persamaan prediksi yang menghubungkan Y dengan X adalah dengan menempatkan sebuah garis pada diagram pencar sedemikian sehingga grafiknya merupakan kecocokan terbaik (best fit), yang ditunjukkan dalam Gambar II.1.



Contoh II.1.

Untuk memperoleh $\hat{\beta}_0$, $\hat{\beta}_1$ dan garis kuadrat terkecil, maka

Tabel II.2. Perhitungan dari data tabel II.1.

	Y_i	X_i	X_i^2	X_iY_i	Y_i^2
	101	1,2	1,44	121,2	10.201
	92	0,8	0,64	73,6	8.464
	110	1,0	1,00	110,0	12.100
	120	1,3	1,69	156,0	14.400
	90	0,7	0,49	63,0	8.100
	82	0,8	0,64	65,6	6.724
	93	1,0	1,00	93,0	8.649
	75	0,6	0,36	45,0	5.625
	91	0,9	0,81	81,9	8.281
	105	1,1	1,21	115,5	11.025
Jumlah	959	9,4	9,28	924,8	93.569

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

$$= 9,28 - \frac{(9,4)^2}{10}$$

$$= 0,444$$

$$SS_{xy} = \sum_{i=1}^{n} X_i Y_i - \frac{(\sum_{i=1}^{n} X_i)(\sum_{i=1}^{n} Y_i)}{n}$$

$$= 924,8 - \frac{(9,4)(959)}{10}$$

$$= 23,34$$

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

$$= \frac{959}{10} \qquad = \frac{9,4}{10}$$

$$= 95,9 \qquad = 0,94$$

Jadi,

$$\hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}}$$

$$= \frac{23,34}{0,444}$$

$$= 52,5676$$

$$\approx 52,57$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

$$= 95,9 - (52,5676)(0,94)$$

$$\approx 46,49$$

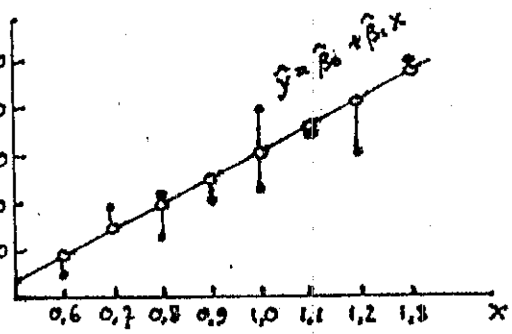
Dan sesuai dengan prinsip kuadrat terkecil, garis lurus dengan kecocokan terbaik (atau garis regresi) yang menghubungkan pengeluaran iklan dengan volume penjualan adalah,

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$$

$$= 46,49 + 52,57\ X$$

Sehingga dapat diramalkan Y untuk nilai X tertentu dengan menggunakan grafik gambar II.2. atau dengan mensubstitusikan nilai X kedalam persamaan prediksi.

Gambar II.2. Persamaan prediksi linear



II.3. Menghitung s^2 ,sebagai penaksir (estimator) untuk σ^2

Variabilitas dari kesalahan acak, yang diukur dengan σ^2 memainkan peranan penting sewaktu mengestimasi atau meramalkan dengan menggunakan garis kuadrat terkecil.

Rumus yang menghasilkan penaksir (estimator) terbaik untuk σ^2 , yang akan bersifat tidak bias dan didasarkan atas n - 2 derajad bebas ialah,

$$\hat{\sigma}^2 = s^2 = \frac{SSE}{n-2}$$

Dan adapun untuk menghitung SSE,

$$SSE = SS_{yy} - \hat{\beta}_1 SS_{xy}$$

dimana ,

$$SS_{yy} = \sum_{i=1}^{n} (Y_i - \bar{Y})^2$$

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

$$SS_{xy} = \sum_{i=1}^{n} X_i Y_i - \frac{(\sum_{i=1}^{n} X_i)(\sum_{i=1}^{n} Y_i)}{n}$$

Contoh II.2.

Dari tabel II.2. Hitung dan taksirlah σ^2

Penyelesaian : menghitung

$$SS_{yy} = 93.569 - \frac{(959)^2}{10}$$

$$= 1.600,9$$

$$SS_{xy} = 23,34$$

$$\hat{\beta}_1 = 52,5676$$

Dengan mensubstitusikan nilai SS_{yy}, nilai $\hat{\beta}_1$ dan nilai SS_{xy} kedalam rumus untuk SSE, diperoleh

$$SSE = 1.600,9 - (52,5676)(23,34)$$
$$= 373,97$$

Jadi,

$$s^2 = \frac{373,97}{10 - 2}$$
$$= 46,75$$

Dan $s = \sqrt{46,75} = 6,84$ yang merupakan taksiran untuk σ

Interpretasi praktis yang dapat diberikan untuk s akhirnya terletak pada arti dari σ. Karena σ mengukur penyebaran nilai-nilai Y disekitar garis yang menghubungkan rerata $E(Y|X) = \beta_0 + \beta_1 X$,dan akan diharapkan (atas dasar Hukum Empiris) kira-kira 95% dari nilai Y berada dalam selang (interval) 2σ dari garis. Karena tidak mengetahui besarnya σ ,maka 2s memberikan nilai yang mendekati setengah lebar interval.

II.4. Asumsi-asumsi dalam Analisa Regresi.

Asumsi yang diberikan /diperlukan dalam estimasi model regresi $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ $\quad i = 1, 2, \ldots, n$ adalah :

1. ε_i merupakan variabel random dengan rerata (mean) = 0 dan varian = σ^2 ,yaitu: $E(\varepsilon_i) = 0$; $Var(\varepsilon_i) = \sigma^2$
2. Tidak ada korelasi antara ε_i dan ε_j untuk $i \neq j$, yaitu $Cov(\varepsilon_i, \varepsilon_j) = 0$,yang berarti $E(Y_i) = \beta_0 + \beta_1 X_i$ dengan $Var(Y_i) = \sigma^2$ dan tidak ada korelasi antara Y_i

dan Y_j untuk $i \neq j$

3. ε_i mengikuti distribusi normal dengan mean = 0 dan var σ^2, atau jika dinyatakan secara ringkas :

$\varepsilon_i \backsim N(0, \sigma^2)$

II.5, Inferensi mengenai kemiringan β_1 dari suatu garis.

Inferensi awal yang dikehendaki dalam mempelajari hubungan antara Y dan X adalah mengenai adanya eksistensi hubungan tadi. Yaitu, apakah data itu menyajikan bukti yang cukup untuk menunjukkan bahwa X menyumbangkan (memberikan) informasi untuk peramalan Y dalam lingkup pengamatan ?.Atau cukup besarkah kemungkinan bahwa jika Y dan X sama sekali tidak mempunyai hubungan. Estimator (penduga) $\hat{\beta}_1$ adalah sangat berguna dalam menyusun statistik uji untuk hipotesis

Dari asumsi normalitas dapat ditunjukkan $\hat{\beta}_1$ dalam pengambilan sampel secara berulang dan bahwa nilai yang diharapkan (nilai rerata) dan varians dari $\hat{\beta}_1$ adalah,

$$E(\hat{\beta}_1) = \beta_1 \quad \text{dan} \quad \sigma^2_{\hat{\beta}_1} = \frac{\sigma^2}{SS_{xx}}$$

Dari rumus (2:5b) diketahui,

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}$$

Dan karena,

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

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

$$= \sum_{i=1}^{n} X_i - \sum_{i=1}^{n} X_i$$

$$= 0$$

$$\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i (X_i - \overline{X})$$

$$= \sum_{i=1}^{n} x_i X_i - \overline{X} \sum_{i=1}^{n} x_i$$

$$= \sum_{i=1}^{n} x_i X_i \quad \text{,maka } x_i = X_i$$

Dengan cara yang sama berlaku, $y_i = Y_i$

Sehingga,

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i Y_i}{\sum_{i=1}^{n} x_i^2} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (2:5d)$$

Jika,

$$t_i = \frac{x_i}{\sum_{i=1}^{n} x_i^2}$$

$$\sum_{i=1}^{n} t_i = \frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} x_i^2}$$

$$= 0$$

$$\sum_{i=1}^{n} t_i X_i = \frac{\sum_{i=1}^{n} x_i X_i}{\sum_{i=1}^{n} x_i^2}$$

$$= \frac{\sum_{i=1}^{n} x_i x_i}{\sum_{i=1}^{n} x_i^2}$$

$$= \frac{\sum_{i=1}^{n} x_i^2}{\sum_{i=1}^{n} x_i^2}$$

$$= 1$$

Maka,

$$\hat{\beta}_1 = \sum_{i=1}^{n} t_i Y_i \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (2:5e)$$

merupakan fungsi linier dari Y_i

$$= \sum_{i=1}^{n} t_i(\beta_0 + \beta_1 X_i + \varepsilon_i)$$

$$= \beta_0 \sum_{i=1}^{n} t_i + \beta_1 \sum_{i=1}^{n} t_i X_i + \sum_{i=1}^{n} t_i \varepsilon_i$$

$$= \beta_0(0) + \beta_1(1) + \sum_{i=1}^{n} t_i \varepsilon_i$$

$$= \beta_1 + \sum_{i=1}^{n} t_i \varepsilon_i$$,diambil nilai harapan (expected value)

$$E(\hat{\beta}_1) = \beta_1 + \sum_{i=1}^{n} t_i \, E(\varepsilon_i),$$ karena $\sum_{i=1}^{n} t_i$ konstan

$$= \beta_1 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (2:6)$$

Hal ini menunjukkan bahwa $\hat{\beta}_1$ merupakan estimator (penduga) tidak berbias dari nilai β_1 yang sebenarnya.

II.5.1. Standard Deviasi (error) koefisien regresi $\hat{\beta}_1$

Varians $\hat{\beta}_1$ didefinisikan ,

$$\sigma^2_{\hat{\beta}_1} = E[\hat{\beta}_1 - E(\hat{\beta}_1)]^2$$

$$= E(\hat{\beta}_1 - \beta_1)^2$$

dimana,

$$\hat{\beta}_1 = \beta_1 + \sum_{i=1}^{n} t_i \varepsilon_i$$

$$\hat{\beta}_1 - \beta_1 = \sum_{i=1}^{n} t_i \varepsilon_i$$

maka,

$$\sigma^2_{\hat{\beta}_1} = E\left(\sum_{i=1}^{n} t_i \varepsilon_i\right)^2$$

$$= E\left[(t_1\varepsilon_1)^2 + (t_2\varepsilon_2)^2 + \ldots + (t_n\varepsilon_n)^2 + \right.$$

$$+2(t_1t_2\varepsilon_1\varepsilon_2)+2(t_1t_3\varepsilon_1\varepsilon_3)+\ \ldots$$

$$+2(t_1t_n\varepsilon_1\varepsilon_n)+\ \ldots\ +2(t_{n-1}t_n\varepsilon_{n-1}\varepsilon_n)\Big]$$

$$\sigma^2_{\hat{\beta}_1}=E\Big[\sum_{i=1}^{n} t_i^2\varepsilon_i^2 + 2\sum_{i<j}\sum t_i t_j \varepsilon_i\varepsilon_j\Big]$$

$$=\sum_{i=1}^{n} t_i^2\ E(\varepsilon_i^2)+\ 2\sum_{i<j}\sum t_i t_j\ E(\varepsilon_i\varepsilon_j)$$

Karena,

$$t_i = \frac{x_i}{\sum_{i=1}^{n} x_i^2}$$, dijumlah kuadratkan ,akan didapat

$$\sum_{i=1}^{n} t_i^2 = \frac{\sum_{i=1}^{n} x_i^2}{\left(\sum_{i=1}^{n} x_i^2\right)^2}$$

$$= \frac{1}{\sum_{i=1}^{n} x_i^2}$$

Sehingga,

$$\sigma^2_{\hat{\beta}_1} = \frac{1}{\sum_{i=1}^{n} x_i^2}\,\sigma^2 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (\ 2:7\ )$$

$$= \frac{1}{\sum_{i=1}^{n} (X_i-\bar{X})^2}\,\sigma^2 \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (\ 2:7a)$$

$$= \frac{1}{SS_{xx}}\,\sigma^2 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (\ 2:7b)$$

Jadi Standard Deviasi :

$$\sigma_{\hat{\beta}_1} = \frac{1}{\sqrt{SS_{xx}}}\,\sigma \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (\ 2:8\ )$$

Besarnya σ dapat diestimasi dari besarnya s ,sehingga :

$$s_{\hat{\beta}_1} = \frac{s}{\sqrt{SS_{xx}}} \quad \dots\dots\dots\dots\dots\dots\dots\dots\dots \quad (2:8a)$$

Bila diasumsikan bahwa variasi disekitar garis regresi tersebut mengikuti distribusi normal, artinya error ε_i berasal dari distribusi normal yang sama, $N(0, \sigma^2)$ maka besarnya batas kepercayaan(selang keyakinan) untuk $\hat{\beta}_1$:

$$\hat{\beta}_1 \pm \frac{t_{\alpha/2}\, s}{\sqrt{SS_{xx}}} \quad \dots\dots\dots\dots\dots\dots\dots\dots\dots \quad (2:9)$$

Sedangkan bentuk test hipotesanya :

H_o : $\beta_1 = 0$

H_a : $\beta_1 \neq 0$

Statistik test (uji) :

$$t_{hitung} = \frac{\hat{\beta}_1 - 0}{s} \sqrt{SS_{xx}}$$

$$= \frac{\hat{\beta}_1}{s} \sqrt{SS_{xx}} \quad \dots\dots\dots\dots\dots\dots\dots\dots\dots \quad (2:10)$$

Daerah penolakan : lihat nilai kritis untuk t, Tabel 2 dalam Lampiran (Appendix), untuk derajad kebebasan (n-2)

Contoh II.3.

Untuk data volume penjualan dan biaya iklan pada tabel II.2.
Statistik uji adalah,

$$t_{hitung} = \frac{\hat{\beta}_1}{s} \sqrt{SS_{xx}}$$

Dengan mensubstitusikan nilai-nilai yang telah dihitung dalam contoh II.1. dan contoh II.2. maka diperoleh :

$$t_{hitung} = \frac{52,57}{6,84} \sqrt{0,444}$$

$$= 5,12$$

Dan jika dipilih $\alpha = 0,05$,akan menolak H_0 jika $t > 2,306$ atau $t < -2,306$. Nilai kritis t diperoleh dari tabel t , dengan menggunakan derajat kebebasan (n-2), yaitu (10-2)=8 Melihat bahwa statistik uji (t_{hitung}) melampui nilai kritis ,$t > 2,306$ maka menolak hipotesis nol $\beta_1 = 0$ dan menyimpulkan bahwa ada bukti yang menunjukkan bahwa pengeluaran biaya iklan memberikan informasi untuk digunakan dalam peramalan volume penjualan bruto bulanan.

Contoh II.4.

Dapatkan selang keyakinan 95% untuk β_1 dengan menggunakan data pada tabel II.1.

Penyelesaian: Selang keyakinan 95% untuk β_1, dengan menggunakan nilai-nilai yang telah dihitung sebelumnya, adalah

$$\hat{\beta}_1 \pm \frac{t_{0,025}\, s}{\sqrt{SS_{xx}}}$$

Dengan mensubstitusikan nilai-nilai yang telah dihitung, maka diperoleh :

$$52,57 \pm \frac{(2,306)(6,84)}{\sqrt{0,444}}$$

atau $52,57 \pm 23,67$

Ini berarti bahwa kenaikan volume penjualan bulanan untuk kenaikan setiap unit (Rp 10.000.000) pengeluaran iklan berada dalam selang dari 28,90 sampai 76,24 atau dalam unit yang asli untuk Y, dari Rp 289.000.000 sampai Rp 762.400.000.

II.6. Mengestimasi (menduga) E(Y|X), Nilai yang diharapkan (rerata) Y untuk suatu nilai X tertentu.

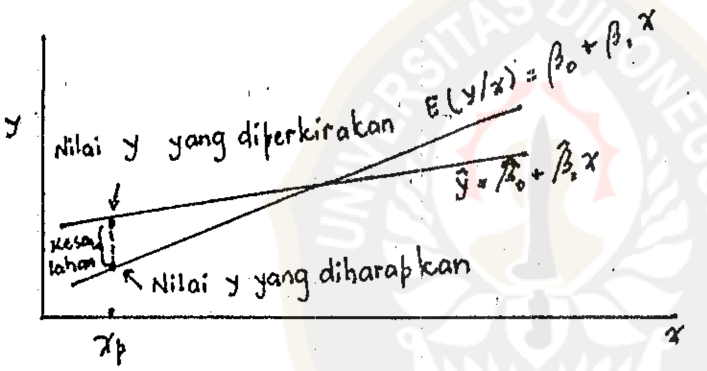
Garis regresi (garis suai=fitted line),

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$$

mencoba untuk mengestimasi garis yang menghubungkan rerata /mean $E(Y|X)$, (jadi mengestimasi β_0 dan β_1). Jadi $\hat{Y}$ dapat digunakan untuk mengestimasi nilai yang diharapkan dari $\overline{Y}$ dan juga untuk memperkirakan beberapa nilai Y yang mungkin diamati dimasa yang akan datang.

Dalam bagian ini akan membahas estimasi dari nilai yang diharapkan untuk Y, untuk suatu nilai X tertentu.

Dari gambar II.3. Nilai Y yang diharapkan dan yang diperkirakan



Dapat diamati, garis pertama menyatakan garis yang menghubungkan rerata (mean)

$$E(Y|X) = \beta_0 + \beta_1 X$$

dan garis kedua adalah persamaan prediksi yang diperoleh dari pencocokkan

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$$

Dari gambar terlihat bahwa kesalahan dalam mengestimasi nilai yang diharapkan untuk Y pada saat $X = X_p$ adalah penyimpangan (deviation) antara kedua garis diatas titik X_p.

Dapat ditunjukkan bahwa nilai yang diharapkan

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$$

merupakan penduga yang tak bias untuk $E(Y|X)$, yaitu

$E(\hat{Y}) = \beta_0 + \beta_1 X$,dan $\hat{Y}$ didistribusikan secara normal, dengan varians,

$$\sigma^2_{\hat{Y}} = \sigma^2\left[\frac{1}{n} + \frac{(X_p - \bar{X})^2}{SS_{xx}}\right] \quad \ldots\ldots\ldots\ldots\ldots (2:11)$$

Varians yang diperkirakan berpadanan untuk Y diperoleh dengan menggantikan s^2 untuk σ^2.

Hasil yang digambarkan diatas dapat digunakan untuk menguji hipotesis mengenai rata-rata atau nilai yang diharapkan untuk Y, untuk suatu nilai X tertentu, misal X_p.

Hipotesis nol adalah

$$H_0 : E(Y|X = X_p) = E_0$$

dimana E_0 adalah nilai numerik yang dihipotesiskan untuk E(Y) pada saat $X=X_p$. Dapat ditunjukkan bahwa besaran,

$$t = \frac{\hat{Y} - E_0}{\text{yang diperkirakan } \sigma_{\hat{Y}}}$$

$$= \frac{\hat{Y} - E_0}{s\sqrt{\frac{1}{n} + \frac{(X_p - \bar{X})^2}{SS_{xx}}}} \quad \ldots\ldots\ldots\ldots\ldots (2:12)$$

Daerah penolakan : lihat nilai kritis untuk t, pada Tabel 7. dalam Lampiran (Appendix), untuk derajad kebebasan (n - 2)

Selang keyakinan yang bersesuaian, dengan koefisien keyakinan $(1 - \alpha)$, untuk nilai Y yang diharapkan untuk $X=X_p$ sebagai berikut,

Selang Keyakinan untuk E(Y|X),

$$\hat{Y} \pm t_{\alpha/2}\, s\sqrt{\frac{1}{n} + \frac{(X_p-\bar{X})^2}{SS_{xx}}} \quad \ldots\ldots\ldots\ldots\ldots (2:13)$$

Contoh II.5.

Dengan menggunakan data pada tabel II.1. dapatkan volume penjualan bulanan yang diharapkan dengan selang keyakinan 95 % untuk pengeluaran iklan X= 1,0 (Rp 10.000.000)

Penyelesaian : Untuk mengestimasi (menduga) rata-rata penjualan bulanan untuk pengeluaran iklan $X_p = 1,0$

Dengan menggunakan

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_p$$

untuk menghitung $\hat{Y}$,yaitu nilai yang diperlukan untuk $E(Y|X) = 1,0$. Kemudian dengan menggunakan nilai-nilai yang telah dihitung dalam contoh sebelumnya, diperoleh

$$\hat{Y} = 46,49 + (52,57)(1,0)$$

$$= 99,06$$

atau Rp 990.600.000

Rumus untuk selang keyakinan 95% adalah

$$\hat{Y} \pm t_{0,025}\, s \sqrt{\frac{1}{n} + \frac{(X_p - \bar{X})^2}{SS_{xx}}}$$

Dengan mensubstitusikan kedalam bentuk ini, maka didapatkan bahwa selang keyakinan 95% untuk volume penjualan yang diharapkan (rata-rata), untuk pengeluaran iklan sebesar 1,0 adalah

$$99,06 \pm (2,306)(6,84) \sqrt{\frac{1}{10} + \frac{(1,0 - 0,94)^2}{0,444}}$$

$$99,06 \pm 5,19$$

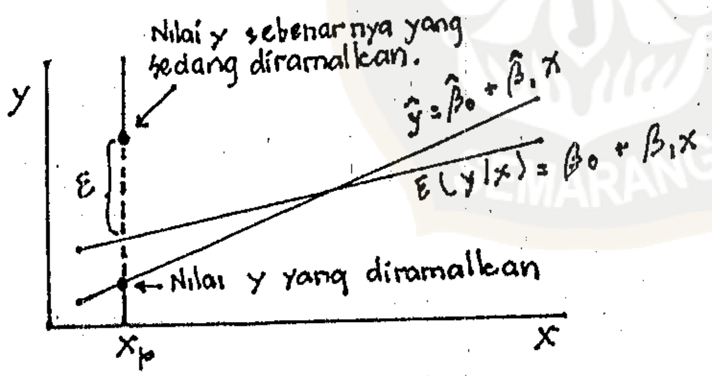
atau dari 93,87 sampai 104,25

Karena tiap unit volume penjualan Rp 10000.000 ,jadi perkiraan bahwa rata-rata volume penjualan bulanan jika perusahaan mengestimasikan Rp 10.000.000 untuk iklan berada dalam selang antara Rp 938.700.000 sampai dengan Rp 1.042.500.000

II.7. Meramalkan nilai Y tertentu untuk nilai X tertentu

Andaikan persamaan prediksi yang diperoleh dari sepuluh pengukuran pada tabel II.1. ,berdasarkan pada data yang diamati untuk meramalkan volume penjuala n untuk suatu bulan pada saat perusahaan sedang atau telah beroperasi. Kalau pengeluaran untuk iklan yang dilakukan perusahaan selama bulan yang diteliti adalah X_p, secara intuisi dapat dil ihat bahwa kesalahan dal am peramalan (penyimpangan antara Y dan volume penjualan sebenarnya Y yang terjadi selama bulan itu) terdiri dari dua unsur $(Y-\hat{Y})$ sama d engan penyimpangan antara $\hat{Y}$ dengan nilai Y yang diha rapkan seperti yang diuraikan dalam bagian II.6. (ditunjukan dalam gambar II.3.) ditambah kesalahan acak ε yang menyatakan perbedaan antara nilai Y yang sebenarnya dan nilai Y yang diharapkan, dan yang ditunjukan dalam gambar II.4.

Gambar II.4. Kesalahan dalam meramalkan suatu nilai Y tertentu



Varians kesalahan $(Y-\hat{Y})$ dalam meramal kan suatu nilai Y tertentu untuk $X=X_p$ adalah

$$\sigma^2_{(Y-\hat{Y})} = \sigma^2 \left[1 + \frac{1}{n} + \frac{(X_p - \overline{X})^2}{SS_{xx}} \right] \quad \ldots\ldots\ldots\ldots \quad (2:14)$$

Jika n sangat besar, unsur kedua dan ketiga di dalam kurung menjadi kecil dan varians kesalahan peramalan mendekayi σ^2. Hasil ini bisa digunakan untuk menyusun selang peramalan untuk Y, untuk $X=X_p$. Koefisien keyakinan untuk selang peramalan (prediction interval) adalah $(1 - \alpha)$

Selang Peramalan $(1 - \alpha)100\%$ untuk Y

$$\hat{Y} \pm t_{\alpha/2}\, s \sqrt{1 + \frac{1}{n} + \frac{(X_p - \bar{X})^2}{SS_{xx}}} \quad \ldots\ldots\ldots \quad (2:15)$$

Contoh II.6.

Dapatkan selang peramalan 95% untuk penjualan perusahaan bulan yang akan datang kalau pengeluaran untuk iklan adalah Rp 10.000.000 dengan asumsi bahwa kondisi ekonomi lainnya kira-kira tetap sama dengan bulan-bulan yang termasuk dalam tabel II.1.

Penyelesaian : Kalau dalam suatu bulan tertentu pengeluaran untuk iklan adalah Rp 10.000.000, maka $X_p = 1{,}0$ dan akan diramalkan bahwa volume penjualan akan,

$$99{,}06 \pm (2{,}306)(6{,}84)\sqrt{1 + \frac{1}{10} + \frac{(1{,}0-0{,}94)^2}{0{,}444}}$$

$$\therefore \quad 99{,}06 \pm 16{,}60$$

atau dari 82,46 sampai 115,66 dimana bahwa setiap unit volume penjualan menyatakan Rp 10.000.000. Jadi selang peramalan 95% untuk volume penjualan bulan yang akan datang adalah Rp 990.600.000 ± Rp 166.000.000 atau dari Rp 824.600.000 sampai Rp 1.156.600.000.

II.8. Koefisien Korelasi

Suatu ukuran (indikator) kekuatan hubungan linear antara dua variabel Y dan X yang bebas dari skala pengukuran, maka ukuran tersebut dinamakan korelasi linear antara Y dan X. Ukuran korelasi linear yang biasa digunakan dalam statistik dikenal sebagai koefisien korelasi product-moment dari Pearson antara Y dan X. Besarannya dinyatakan dengan simbol r,

$$r = \frac{SS_{xy}}{\sqrt{SS_{xx}SS_{yy}}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (2:16)$$

Koefisien korelasi untuk sampel r merupakan pendugaan (estimator) bagi koefisien korelasi untuk populasi ρ ,yang akan diperoleh jika koefisien korelasi dihitung dengan menggunakan semua titik dalam populasi.

Contoh II.7.

Hitunglah koefisien korelasi untuk data pengeluaran iklan dan volume penjualan pada tabel II.1.

Penyelesaian :

Koefisien korelasi untuk data pengeluaran iklan volume penjualan pada tabel II.1. dapat diperoleh dengan menggunakan rumus untuk r dan besaran,

$$SS_{xy} = 23,34 \qquad SS_{xx} = 0,444 \qquad SS_{yy} = 1.600,9$$

yang telah dihitung sebelumnya.

Maka,

$$r = \frac{23,34}{\sqrt{0,444(1.600,9)}}$$

$$= 0,88$$

Secara umum bahwa r mengukur hubungan (asosiasi) linear antara dua variabel Y dan X. Bila r = 1 atau -1, semua titik berada pada satu garis lurus ;jika r = 0 ,titik-titik menyebar dan tidak memberikan bukti adanya hubungan linear.

II.8.1. Koefisien Determinasi

Koefisien determinasi didefinisikan,

$$r^2 = \frac{SS_{yy} - SSE}{SS_{yy}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2:17)$$

$$= 1 - \frac{SSE}{SS_{yy}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2:17a)$$

dimana,

$$0 \leq r^2 \leq 1$$

Koefisien determinasi ($=r^2$) akan memberikan interpretasi yang lebih berarti mengenai kekuatan hubungan antara Y dan X daripada koefisien korelasi (=r).

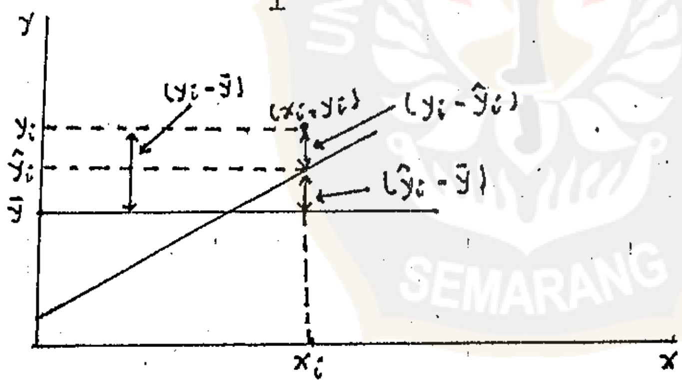
II.9. Dapat Dijumlahkannya Jumlah Kuadrat.

Suatu sifat penting dari analisis regresi adalah ke - cenderungan untuk membagi total jumlah kuadrat penyimpangan,

$$SS_{yy} = \sum_{i=1}^{n} (Y_i - \overline{Y})^2$$

kedalam dua bagian. Pembagian $(Y_i - \overline{Y})$ terlihat dalam ga - mbar II.5.

Gambar II.5. Pemisahan $(Y_i - \overline{Y})$ kedalam $(Y_i - \hat{Y}_i)$ dan $(\hat{Y}_i - \overline{Y})$



Satu bagian yang dapat dihubungkan dengan jumlah kuadrat pe- nyimpangan nilai Y disekitar garis regresi yang dicocokkan, SSE. Bagian lain menyatakan pengurangan dalam jumlah kuadrat penyimpangan sebagai akibat dari informasi yang diberikan oleh penambahan variabel X.

Untuk memahami dipisahkannya jumlah kuadrat, maka de- ngan memperhatikan bahwa analisis regresi cenderung untuk memisahkan tiap pengukuran dari reratanya, $(Y_i - \overline{Y})$, kedalam dua bagian. Jadi,

$$(Y_i - \overline{Y}) = (Y_i - \hat{Y}_i) + (\hat{Y}_i - \overline{Y})$$

Dengan mengambil jumlah kuadrat penyimpangan untuk seluruh pengamatan untuk tiap unsur dalam pemisahan ($Y_i - \bar{Y}$) ,dapat ditunjukkan,

$$SS_{yy} = \sum_{i=1}^{n} (Y_i - \bar{Y})^2$$

$$= \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2 + \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 \quad \text{.........} \quad (2:18)$$

Jadi, total jumlah kuadrat dari Y, yang dinyatakan dengan simbol SS_{yy} (dinamakan Total SS), dapat dipisahkan kedalam dua komponen.

Pemisahan Total Jumlah Kuadrat Y,

Total SS = SSR + SSE

dimana,

$$SSR = \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2$$

= jumlah kuadrat karena regresi

yang berhubungan dengan regresi yang jumlah variasi total yang dijelaskan oleh variabel tambahan X, dan

$$SSE = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$$

= jumlah kuadrat kesalahan

yang merupakan jumlah total variasi yang tidak dijelaskan oleh variabel tambahan X.