ANALISIS REGRESI LINIER SEDERHANA

Pertemuan 7

B.Wisnu Widagdo, S.T, M.Sc.IT

REGRESI DAN KORELASI

REGRESI

- Menganalisis besar pengaruh
- Bersifat kausal (sebab akibat)
- Harus ada variabel bebas dan variabel terikat
- Menghasilkan persamaan regresi

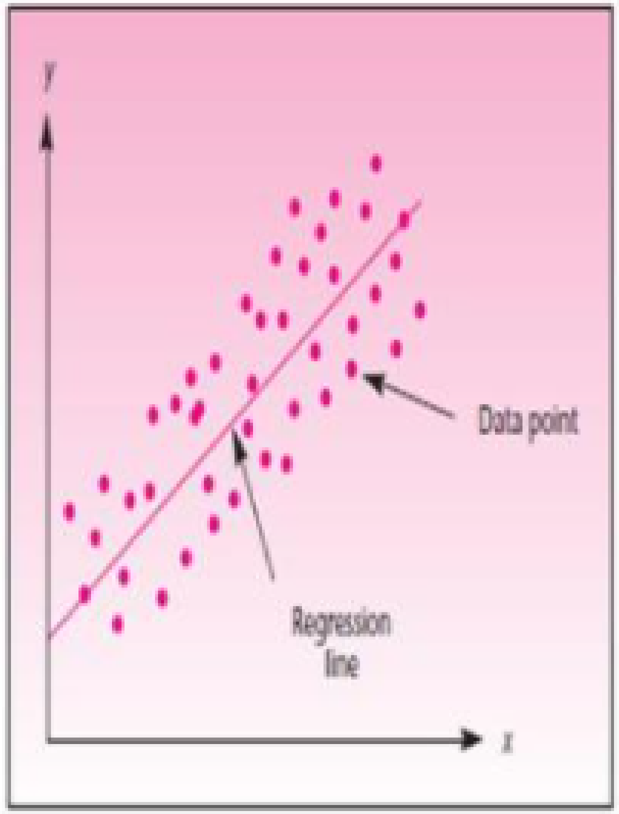
KORELASI

- Menganalisis tingkat hubungan
- Boleh tidak bersifat kausal (sebab akibat)
- Tidak harus ada variabel bebas dan variabel terikat
- Menghasilkan koefisien korelasi regresi

REGRESI SEDERHANA

- Hanya mengandung satu variabel bebas
- Bertujuan untuk menguji keberatian pengaruh dari variabel bebas terhadap variabel terikat
- Hubungan sebab akibat bersifat linier
- Model persamaan regresi :

  Y = bo + b1X + e

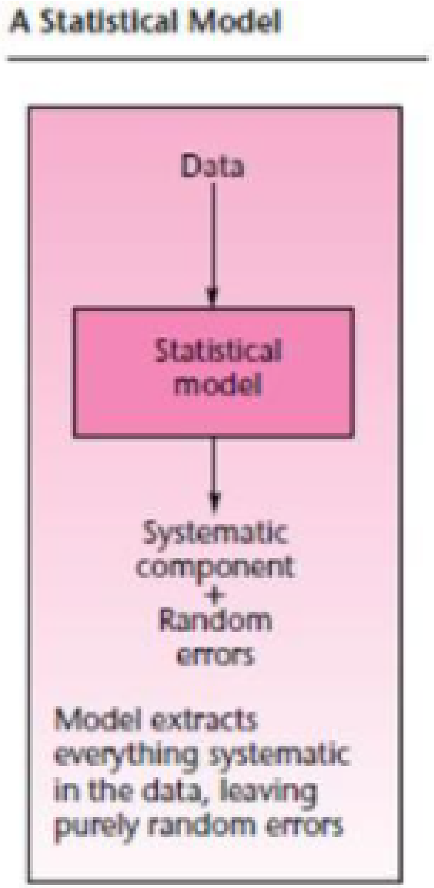


MODEL REGRESI

- Data bisa dimodelkan dengan unsur "systematic component" ditambah dengan "random errors"

- Dinyatakan dengan :

  $Y = \beta o + \beta 1X + \varepsilon$



KOMPONEN MODEL

- Parameter βo disebut dengan konstanta atau intersep, yaitu harga Y apabila nilai X berharga nol
- Parameter β1 disebut dengan koefisien regresi atau slope, yaitu besar pengaruh terhadap Y apabila harga X naik 1 satuan
- Nilai harapan pada Y adalah
  $E(Y) = \beta o + \beta 1X$
- Kompenen error (ε) adalah seliaih nilai antara nilai Y yang sesungguhnya dengan Y hasil model regresi

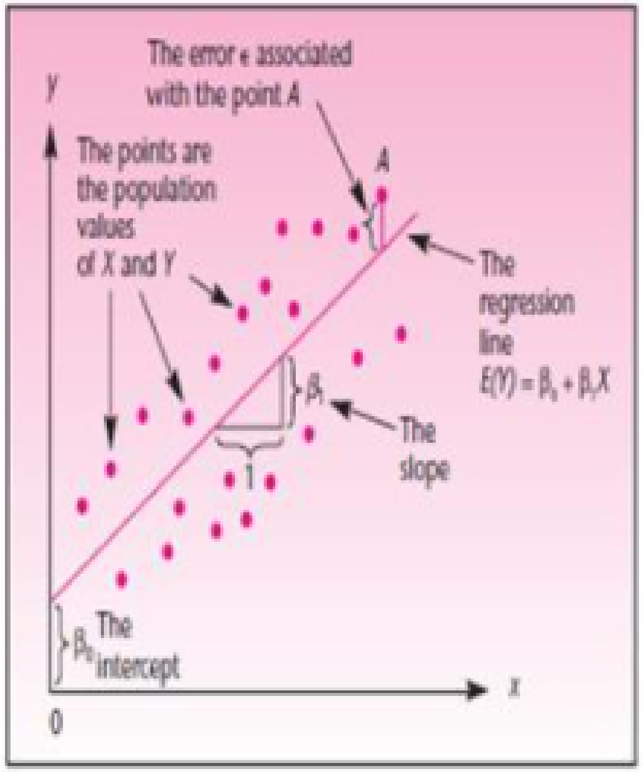
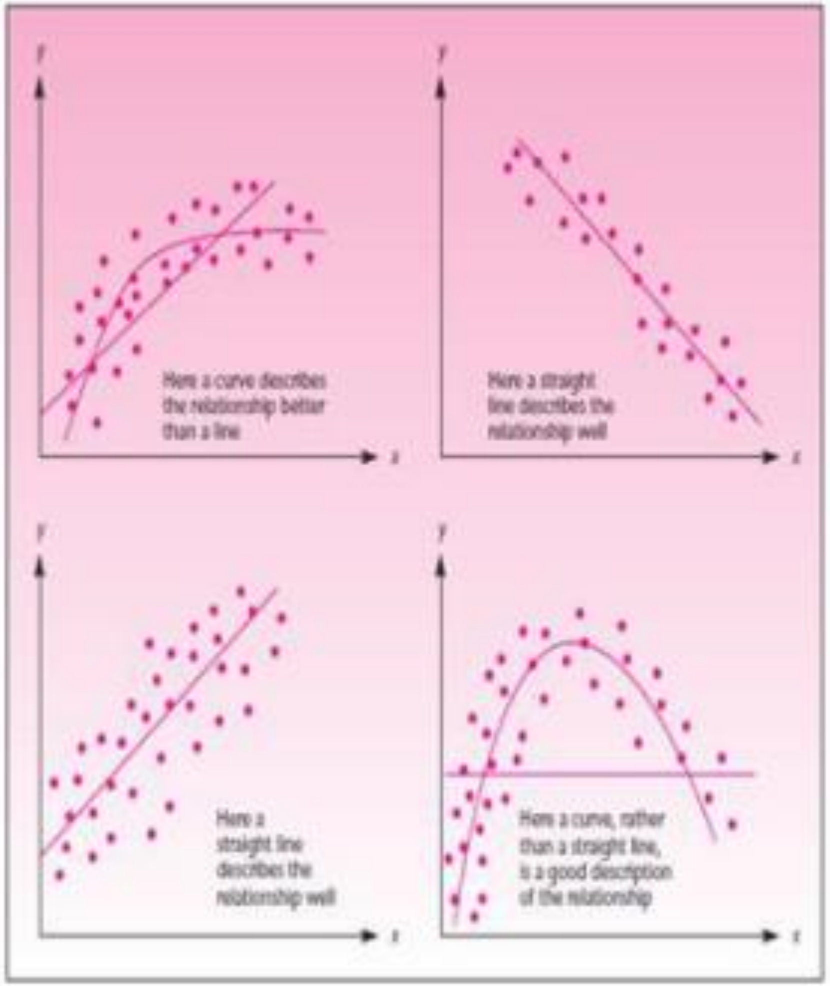


DIAGRAM PENCAR

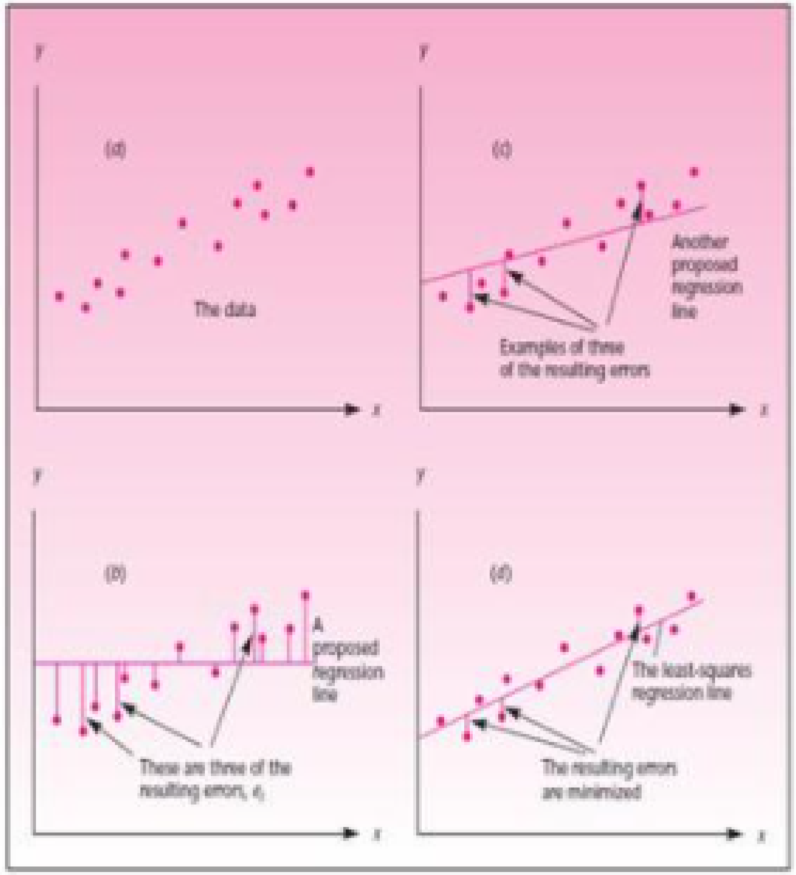
- Sifat hubungan antara variabel bebas dengan variabel terikat bisa dikenali dari hasil diagram pencar (*scatter plot*)

- Bila terlihat tidak bersifat linier (berbentuk kurva atau lengkungan), pemodelan bisa diselesaikan dengan model regresi non linier



TUJUAN MODEL REGRESI

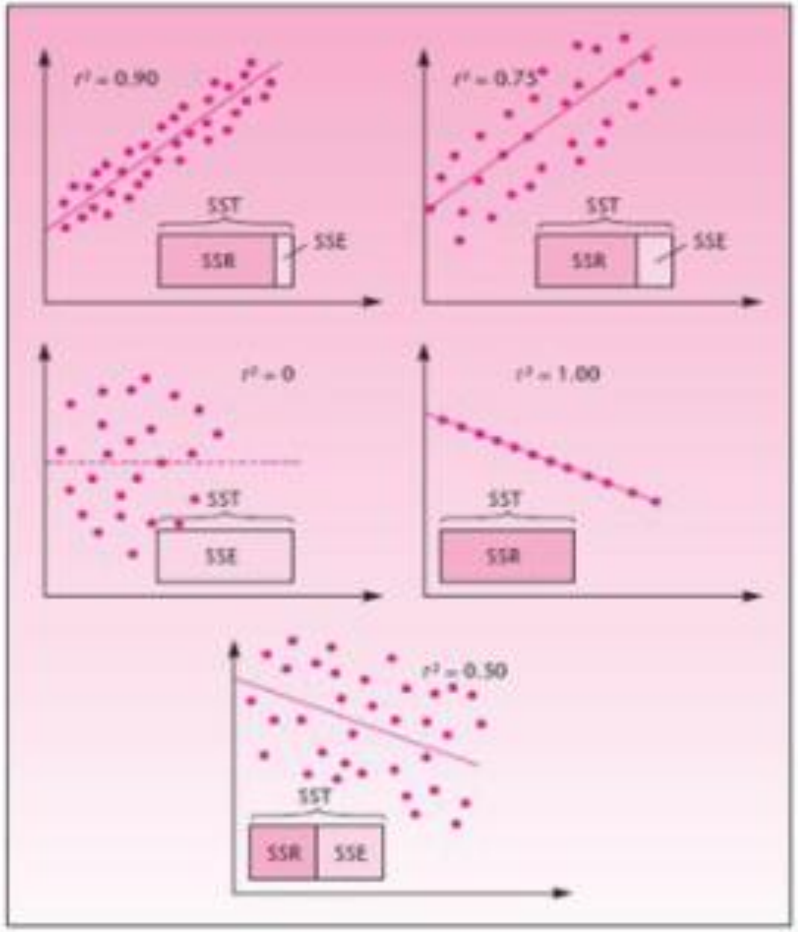
- Mencari persamaan terbaik yang paling dekat dengan data-data yang akan dianalisis
- (a) adalah scatter plot data yang dianalisis
- (b) Bila X sama sekali dianggap tidak berpengaruh pada Y (garis datar), error yang diperoleh sangat besar
- (c) dan (d) hasil analisis regresi dengan garis yang cukup dekat dengan data-data yang sesungguhnya



KOEFISIEN DETERMINASI

- Koefisien determinasi (R^2) adalah besaran statistik dalam model regresi yang dipergunakan untuk mengukur kontribusi variabel bebas dalam menjelaskan keragaman variabel terikat

- Pada gambar dijelaskan bahwa semakin sempurna hubungan variabel, maka semakin besar nilai R^2



ESTIMASI HARGA β_0 DAN β_1

- Persamaan regresi

  $Y = \beta_0 + \beta_1 X + \varepsilon$

- Estimasi persamaan adalah

  $Y = b_0 + b_1 X + e$

Definitions of sums of squares and cross-products useful in regression analysis:

$$SS_x = \sum (x - \bar{x})^2 = \sum x^2 - \frac{(\sum x)^2}{n}$$

$$SS_y = \sum (y - \bar{y})^2 = \sum y^2 - \frac{(\sum y)^2}{n}$$

$$SS_{xy} = \sum (x - \bar{x})(y - \bar{y}) = \sum xy - \frac{(\sum x)(\sum y)}{n} \tag{10-9}$$

The first definition in each case is the conceptual one using squared distances from the mean; the second part is a computational definition. Summations are over all data.

We now give the solutions of the normal equations, the least-squares estimators b_0 and b_1.

Least-squares regression estimators include the slope

$$b_1 = \frac{SS_{xy}}{SS_x}$$

and the intercept

$$b_0 = \bar{y} - b_1\bar{x} \tag{10-10}$$

ESTIMASI HARGA ERROR β_0 DAN β_1

The standard error of b_0 is

$$s(b_0) = \frac{s\sqrt{\sum x^2}}{\sqrt{nSS_X}} \qquad (10\text{–}14)$$

where $s = \sqrt{MSE}$.

The standard error of b_1 is

$$s(b_1) = \frac{s}{\sqrt{SS_X}} \qquad (10\text{–}15)$$

$$\text{df(error)} = n - 2$$

$$\begin{aligned} SSE &= \sum (Y - \hat{Y})^2 \\ &= SS_Y - \frac{(SS_{XY})^2}{SS_X} \\ &= SS_Y - b_1 SS_{XY} \end{aligned} \qquad (10\text{–}13)$$

An unbiased estimator of σ^2, denoted by S^2, is

$$MSE = \frac{SSE}{n - 2}$$

- Nilai βo dan β1 bersifat di estimasi, sehingga akan menghasilkan rentang nilai dengan simpangan sebesar s(b0) dan s(b1)

- Semakin kecil harga simpangan berarti semakin meyakinkan hasil estimasi tersebut

PENGUJIAN HIPOTESIS

- Pengujian dilakukan terhadap koefisien regresi

$$H_0: \beta_1 = 0$$
$$H_1: \beta_1 \neq 0$$

- Statistik uji yang dihitung adalah thitung yang dibandingkan dengan nilai kritis t pada derajat bebas (n-2)

$$t_{(n-2)} = \frac{b_1}{s(b_1)}$$

KORELASI DAN REGRESI LINIER SEDERHANA

1. Pendahuluan

Istilah "regresi" pertama kali diperkenalkan oleh Sir Francis Galton pada tahun 1886. Galton menemukan adanya tendensi bahwa orang tua yang memiliki tubuh tinggi, memiliki anak-anak yang tinggi pula dan orang tua yang pendek memiliki anak-anak yang pendek pula. Kendati demikian, ia mengamati ada kecenderungan bahwa tinggi anak bergerak menuju rata-rata tinggi populasi secara keseluruhan. Dengan kata lain ketinggian anak yang amat tinggi atau orang tua yang amat pendek cenderung bergerak ke arah rata-rata tinggi populasi. Inilah yang disebut hukum Galton mengenai regresi universal. Dalam bahasa Galton ia menyebutnya sebagai regresi menuju medikritas (Maddala, 1992).

Interpretasi modern mengenai regresi agak berlainan dengan regresi versi Galton. Secara umum. analisis regresi pada dasarnya adalah studi mengenai ketergantungan variabel dependen (terikat) dengan satu atau lebih variabel independen (variabel penjelas/bebas), dengan tujuan untuk mengestimasi dan/atau memprediksi rata-rata populasi atau nilai rata-rata variabel dependen berdasarkan nilai variabel independen yang diketahui (Gujarati, 2003).

Hasil analisis regresi adalah berupa koefisien untuk masing-masing variabel independen. Koefisien ini diperoleh dengan cara memprediksi nilai variabel dependen dengan suatu persamaan; Koefisien regresi dihitung dengan dua tujuan sekaligus: Fertama, meminimumkan penyimpangan antara nilai aktual dan nilai estimasi variabel dependen berdasarkan data yang ada (Tabachnick, 1996).

2. Regresi vs Korelasi

Analisis korelasi bertujuan untuk mengukur kekuatan asosiasi (hubungan) linear antara dua variabel. Korelasi tidak menunjukkan hubungan fungsional atau dengan kata lain, analisis korelasi tidak membedakan antara variabel dependen dengan variabel independen.

Dalam analisis regresi, selain mengukur kekuatan hubungan antara dua variabel atau lebih, juga menunjukkan arah hubungan antara variabel dependen dengan variabel independen. Variabel dependen diasumsikan random/stokastik, yang berarti mempunyai distribusi probabilistik. Variabel independen/bebas diasumsikan memiliki nilai tetap (dalam pengambilan sampel yang berulang)

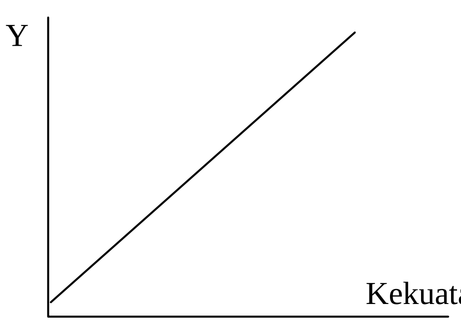
Teknik estimasi variabel dependen yang melandasi analisis regresi disebut Ordinary Least Squares (pangkat kuadrat terkecil biasa). Metode OLS diperkenalkan pertama kali oleh Carl Friedrich Gauss, seorang ahli matematika dari Jerman. Inti metode OLS adalah mengestimasi suatu garis regresi dengan jalan meminimalkan jumlah dari kuadrat kesalahan setiap observasi terhadap garis tersebut.

4. KORELASI

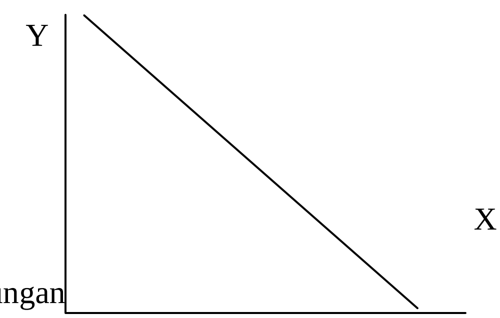
Korelasi menyatakan derajat hubungan antara dua variabel tanpa memperhatikan variabel mana yang menjadi peubah. Karena itu hubugan korelasi belum dapat dikatakan sebagai hubungan sebab akibat.

Bentuk Hubungan

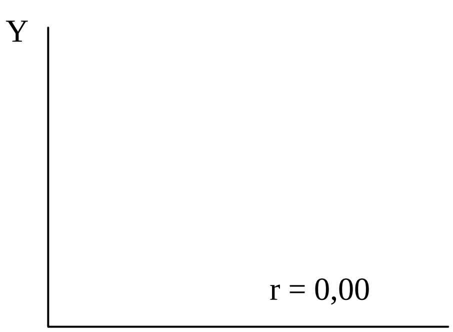
1. Korelasi Positif

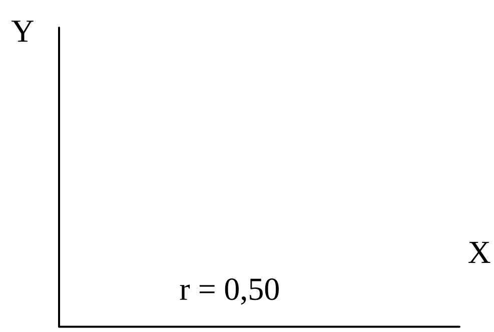


2. Korelasi Negatif



Kekuatan Hubungan





5. REGRESI LINIER SEDERHANA

Tujuan utama materi ini adalah bagaimana menghitung suatu perkiraan atau persamaan regresi yang akan menjelaskan hubungan antara dua variabel.

Regresi merupakan suatu alat ukur yang juga digunakan untuk mengukur ada atau tidaknya korelasi antarvariabelnya. Istilah regresi itu sendiri berarti ramalan atau taksiran.Persamaan yang digunakan untuk mendapatkan garis regresi pada data diagram pencar disebut persamaan regresi.

Untuk menempatkan garis regresi pada data yang diperoleh maka digunakan metode kuadrat terkecil, sehingga bentuk persamaan regresi adalah sebagai berikut:

$$Y' = a + b X$$

Kesamaan di antara garis regresi dan garis trend tidak dapat berakhir dengan persamaan garis lurus. Garis regresi (seperti garis trend dan nilai tengah aritmatika) memiliki dua sifat matematis berikut : $\Sigma(Y - Y') = 0$ dan $\Sigma(Y - Y')^2$ = nilai terkecil atau terendah. Dengan perkataan lain, garis regresi akan ditempatkan pada data dalam diagram sedemikian rupa sehingga penyimpangan (perbedaan) positif titik-titik terhadap titik-titik pencar di atas garis akan mengimbangi penyimpangan negatif titik-titik pencar yang terletak di bawah garis, sehingga hasil pinyimpangan keseluruhan titik-titik terhadap garis lurus adalah nol.

Untuk tujuan diatas, perhitungan analisis regresi dapat dipermudah dengan menggunakan rumus sebagai berikut :

$$Y = a + bx$$

$$b = \frac{n\,(\Sigma xy) - (\Sigma x).(\Sigma y)}{n\,(\Sigma x^2) - (\Sigma x)^2}$$

$$a = \frac{\Sigma y - b.\,(\Sigma x)}{n}$$

6. Asumsi Ordinary Least Squares

Menurut Gujarati (2003) asumsi utama yang mendasari model regresi linear klasik dengan menggunakan model OLS adalah:

a. Model regresi linear, artinya linear dalam parameter seperti dalam persamaan di bawah ini Yi=bl+b2Xi+ui
b. Nilai X diasumsikan non-stokastik, artinya nilai X dianggap tetap dalam sampel yang berulang.
c. Nilai rata-rata kesalahan adalah nol, atau E(ui/Xi) = 0.
d. Homoskedastisitas, artinya variance kesalahan sama untuk setiap periode (Homo = sama, Skedastisitas = sebaran) dan dinyatakan dalam bentuk matematis Var (ui/Xi) = 6^2.
e. Tidak ada autokorelasi antar kesalahan (antara ui dan uj tidak ada korelasi) atau secara matematis Cov (ui,uj/Xi,Xj)= 0.
f. Antara ui dan Xi saling bebas, sehingga Cov (ui/Xi) = 0.
g. Jumlah observasi, n, harus lebih besar daripada jumlah parameter yang diestimasi (jumah variabel bebas).
h. Adanya variabilitas dalam nilai X, artinya nilai X harus berbeda.
i. Model regresi telah dispesifikasi secara benar. Dengan kata lain tidak ada bias (kesalahan) spesifikasi dalam model yang digunakan dalam analisis empirik.
j. Tidak ada multikolinearitas yang sempurna antar variabel bebas.

4. Menilai Goodness of Fit Suatu Model

Ketepatan fungsi regresi sampel dalam menaksir nilai aktual dapat diukur dari Goodness of fitnya. Secara statistik, setidaknya ini dapat diukur dari nilai koefisien determinasi, nilai statistik F dan nilai statistik t. Perhitungan statistik disebut signifikan secara statistik apabila nilai uji statistiknya berada dalam daerah kritis (daerah dimana Ho ditolak). Sebaliknya disebut tidak signifikan bila nilai uji statistiknya berada dalam daerah dimana Ho diterima.

a. **Koefisien Determinasi**

Koefisien determinasi (R^2) pada intinya mengukur seberapa jauh kemampuan model dalam menerangkan variasi variabel dependen. Nilai koefisien determinasi adalah antara nol dan satu. Nilai R^2 = yang kecil berarti kemampuan variabel-variabel independen dalam menjelaskan variasi variabel dependen amat terbatas. Nilai yang mendekati satu berarti variabel-variabel independen memberikan hampir semua informasi yang dibutuhkan untuk memprediksi variasi variabel dependen. Secara umum koefisien determinasi untuk data silang (crossection) relatif rendah karena adanya variasi yang besar antara masing-rnasing pengamatan, sedangkan untuk data runtun waktu (time series) biasanya mempunyai nilai koefisien determinasi yang tinggi.

Satu hal yang perlu dicatat adalah masalah regresi lancung (spurious regression). Insukindro (1998) menekankan bahwa koefisien determinasi hanyalah salah satu dan bukan satu-satunva kriteria memilih model yang baik. Alasannya bila suatu estimasi regresi linear menghasilkan koefisien determinasi yang tinggi, tetapi tidak konsisten dengan teori ekonomika yang dipilih oleh peneliti, atau tidak lolos dari uji asumsi klasik, maka model tersebut bukanlah model penaksir *yang* baik dan seharusnya tidak dipilih menjadi model empirik.

Kelemahan mendasar penggunaan koefisien determinasi adalah bias terhadap jumlah variabel independen yang dimasukkan kedalam model. Setiap tambahan satu variabel independen, maka R^2 pasti meningkat tidak perduli apakah variabel tersebut berpengaruh secara signifikan terhadap variabel dependen. Oleh karena itu banyak peneliti menganjurkan untuk menggunakan nilai Adjusted R^2 pada saat mengevaluasi mana model regresi terbaik. Tidak seperti R^2, nilai Adjusted R^2 dapat naik atau turun apabila satu variabel independen ditambahkan kedalam model.

Dalam kenyataan nilai adjusted R^2 dapat bernilai negatif, walaupun yang dikehendaki harus bernilai positif. Menurut Gujarati (2003) jika dalam uji empiris didapat nilai adjusted R^2 negatif, maka nilai adjusted R^2 dianggap bernilai nol. Secara matematis jika nilai $R^2 = 1$, maka Adjusted $R^2 = R^2 = I$ sedangkan jika nilai $R^2 = 0$, maka adjusted $R^2 = (1 - k)/(n - k)$. Jika $k > 1$, maka adjusted R= akan bernilai negative.

CONTOH SOAL

ANALISIS REGRESI LINIER SEDERHANA :

Tabulasi hasil penelitian variabel kualitas layanan dengan volume penjualan sabun cuci diperoleh data sebagai berikut :

n	X	Y
1	45	120
2	48	173
3	63	149
4	46	166
5	56	170
6	52	174
7	56	156
8	47	158
9	56	150
10	55	160
11	52	157
12	50	177
13	60	166
14	55	160
15	45	155
16	47	159
17	53	159
18	49	172
19	57	168
20	58	159

Hitunglah :

1. Nilai korelasi dan determiasi
2. Persamaan regresi linier sederhana
3. Kesimpulan apa yang diperoleh dari hubungan dua variabel tersebut

Jawab :

n	x	y	xy	x^2	y^2
1	45	120	5.400	2.025	14.400
2	48	173	8.304	2.304	29.929
3	63	149	9.387	3.969	22.201
4	46	166	7.636	2.116	27.556
5	56	170	9.520	3.136	28.900
6	52	174	9.048	2.704	30.276
7	56	156	8.736	3.136	24.336
8	47	158	7.426	2.209	24.964
9	56	150	8.400	3.136	22.500
10	55	160	8.800	3.025	25600
11	52	157	8.164	2.704	24.649
12	50	177	8.850	2.500	31.329
13	60	166	9.960	3.600	27.556
14	55	160	8.800	3.025	25.600
15	45	155	6.975	2.025	24.025
16	47	159	7.473	2.209	25.281
17	53	159	8.427	2.809	25.281
18	49	172	8.428	2.401	29.584
19	57	168	9.576	3.249	28.224
20	58	159	9.222	3.364	25.281
	1.050	**3.208**	**168.532**	**55.646**	**517.472**

1. - Analisa Korelasi

$$R = \frac{n\,(\Sigma xy) - (\Sigma x).(\Sigma y)}{\sqrt{n\,(\Sigma x^2) - (\Sigma x)^2}\,\sqrt{n\,(\Sigma y^2) - (\Sigma y)^2}}$$

$$R = \frac{20\,(168.532) - (1.050).(3.208)}{\sqrt{20\,(55.646) - 1.102.500}\quad\sqrt{20\,(517.472) - 10.291.264}}$$

$$R = \frac{3.370.640 - 3.368.400}{\sqrt{10.420 \;\; x \;\; 58.176}}$$

$$R = \frac{2240}{24621{,}01}$$

$$R = 0{,}091$$

- **Analisis Koefisien determinasi**

$R^2 = R \times R$
$= 0{,}091 \times 0{,}091$
$= 0{,}008$

2. **Persamaan regresi**

$Y = a + bx$

$$b = \frac{n\,(\Sigma xy) - (\Sigma x).(\Sigma y)}{n\,(\Sigma x^2) - (\Sigma x)^2}$$

$$b = \frac{20\,(168.532) - (1.050).(3.208)}{20\,(55.646) - 1.102.500}$$

$$b = \frac{3.370.640 \;-\; 3.368.400}{1.112.920 - 1.102.500}$$

$$b = \frac{2.240}{10.420}$$

$b =$ **0,215**

$$a = \frac{\Sigma y - b.\,(\Sigma x)}{n}$$

$$a = \frac{3.208 - \;0{,}215 \times 1.050}{20}$$

$$a = \frac{3.208 - \;225{,}72}{20}$$

$$a = \frac{2982{,}28}{20}$$

$a =$ **149,114**

Persamaan regresinya Y = 149,114 + 0,215 X

3. Kesimpulan

a. Nilai koefisien korelasi diperoleh sebesar 0,091. Hal ini berarti adanya hubungan positif antara kualitas layanan dengan rata-rata penjualan, namun jika dilihat dari nilai korelasi hubungan variabel tersebut termasuk kategori rendah. Dengan demikian berarti kualitas layanan memiliki hubungan rendah terhadap kenaikan rata-rata penjualan.

   Nilai koefisien determinasi sebesar 0,008. Hal ini menunjukkan kemampuan variabel kualitas layanan dalam mempengaruhi variabel rata-rata penjualan barang sebesar 0,8%, sedangkan sisanya sebesar 99,2% dipengaruhi oleh faktor lain.

b. Y = 149,114 + 0,215 X

- Nilai konstanta (a) = 149,114

  Nilai konstanta (a) sebesar 149,114, menunjukkan besarnya variabel rata-rata penjualan barang yang tidak dipengaruhi oleh kualitas layanan atau dapat diartikan pada saat nilai kualitas layanan sebesar 0, maka rata-rata penjualan sebesar 149.114.
- Koefisien regresi sebesar 0,215, berarti kualitas layanan mempunyai hubungan positif atau searah dengan rata-rata penjualan, karena koefisien regresi bernilai positif. Setiap peningkatan 1 satuan kualitas layanan maka akan berpengaruh terhadap peningkatan rata-rata penjualan sebesar 0,215 satuan. Begitu juga sebaliknya setiap penurunan kualitas layanan sebesar 1 satuan akan berpengaruh terhadap penurunan rata-rata penjualan sebesar 0,215 satuan.