

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 medikritis (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: Pertama, 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

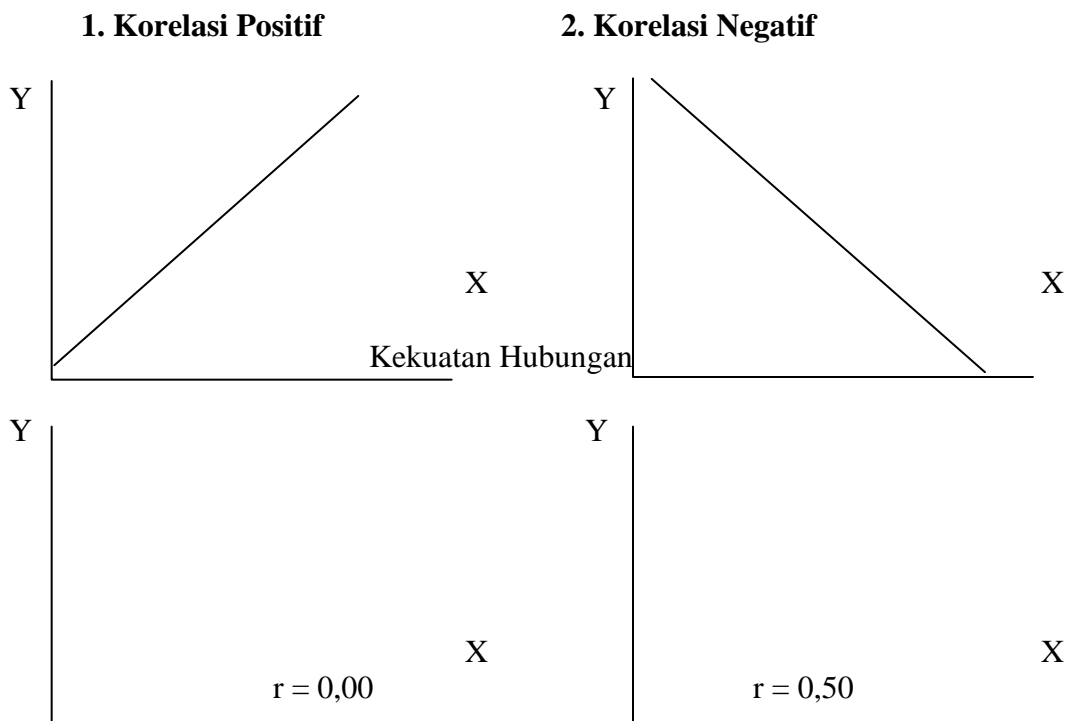
independen. Variabel dependen diasumsikan random/stokastik, yang berarti mempunyai distribusi probabilistik. Variabel independen/bebas diasumsikan memiliki nilai tetap (dalam pengambilan sampel yang berulang)

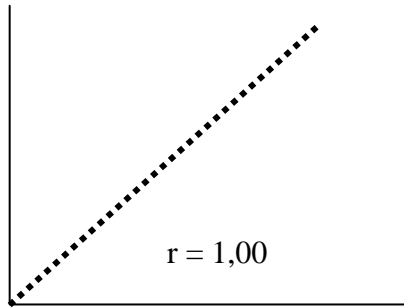
Teknik estimasi variabel dependen yang melandasi analisis regresi disebut Ordinary Least Squares (pangkat kuadrat terkecil biasa). Metode OLS diperkenalkan pertama kali oleh Carl Friedrich Gauss, seorang ahli matematika dari Jerman. Inti metode OLS adalah mengestimasi suatu garis regresi dengan jalan 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 penyebab. Karena itu hubungan korelasi belum dapat dikatakan sebagai hubungan sebab akibat.

Bentuk Hubungan





Keterangan :

1. Hubungan positif menyatakan hubungan semakin besar nilai pada variabel X, diikuti pula perubahan dengan semakin besar nilai pada variabel Y
2. Hubungan negatif menyatakan hubungan semakin besar nilai pada variabel X, diikuti pula perubahan dengan semakin kecil nilai pada variabel Y.
3. $r = 1,00$ menyatakan hubungan yang sempurna kuat; $r = 0,50$ menyatakan hubungan sedang; dan $0,00$ menyatakan tidak ada hubungan sama sekali (dua variabel tidak berhubungan).

Penggunaan Tehnik Korelasi

| No. | Tingkatan Skala Ukur | Teknik Korelasi yang sesuai |
|-----|----------------------|---|
| 1. | Nominal | 1. Koefisien Kontingensi |
| 2. | Ordinal | 1. Spearman Rank 2. Kendal τ (tau) |
| 3. | Interval dan Rasio | 1. Pearson Product Moment 2. Korelasi Ganda 3. Korelasi Parsial |

Korelasi Product Moment

RUMUS :

$$r = \frac{n (\sum xy) - (\sum x) \cdot (\sum y)}{\sqrt{n (\sum x^2) - (\sum x)^2} \sqrt{n (\sum y^2) - (\sum y)^2}}$$

Keterangan:

- r_{xy} = hubungan variabel X dengan Variabel Y
 X = Nilai variabel X
 Y = Nilai variabel Y

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 penyimpangan 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 $Y_i = b_1 + b_2 X_i + u_i$
- b. Nilai X diasumsikan non-stokastik, artinya nilai X dianggap tetap dalam sampel yang berulang.
- c. Nilai rata-rata kesalahan adalah nol, atau $E(u_i/X_i) = 0$.
- d. Homoskedastisitas, artinya variance kesalahan sama untuk setiap periode (Homo = sama, Skedastisitas = sebaran) dan dinyatakan dalam bentuk matematis $Var(u_i/X_i) = \sigma^2$.
- e. Tidak ada autokorelasi antar kesalahan (antara u_i dan u_j tidak ada korelasi) atau secara matematis $Cov(u_i, u_j/X_i, X_j) = 0$.
- f. Antara u_i dan X_i saling bebas, sehingga $Cov(u_i/X_i) = 0$.
- g. Jumlah observasi, n , harus lebih besar daripada jumlah parameter yang diestimasi (jumlah 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 H_0 ditolak). Sebaliknya disebut tidak signifikan bila nilai uji statistiknya berada dalam daerah dimana H_0 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 (crosssection) relatif rendah karena adanya variasi yang besar antara masing-masing 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-satunya 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 peduli 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 $R^2 = R^2 = 1$ sedangkan jika nilai $R^2 = 0$, maka adjusted $R^2 = (1 - k)/(n - k)$. Jika $k > 1$, maka adjusted R^2 akan bernilai negative.

b. Uji Signifikan Parameter Individual (Uji Statistik t)

Uji statistik t pada dasarnya menunjukkan seberapa jauh pengaruh satu variabel penjelas/independen secara individual dalam menerangkan variasi variabel dependen. Hipotesis nol (H_0) yang hendak diuji adalah apakah suatu parameter (β_i) sama dengan nol, atau

$$H_0: \beta_i = 0$$

Artinya apakah suatu variabel independen bukan merupakan penjelas yang signifikan terhadap variabel dependen. Hipotesis alternatifnya (H_A) parameter suatu variabel tidak sama dengan nol, atau

$$H_A: \beta_i \neq 0$$

Artinya, variabel tersebut merupakan penjelas yang signifikan terhadap variabel dependen.

Untuk menguji signifikansi pengaruh variabel x terhadap y digunakan uji t dengan rumus sebagai berikut :

- Rumus t hitung :

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

- Rumus t tabel : $t_{\alpha} df (n-2)$

dimana :

t = t hitung uji signifikansi

r = koefisien korelasi

n = jumlah periode

Dengan kriteria pengujian sebagai berikut :

H_0 diterima apabila $t_{test} \geq t_{tabel}$

H_0 ditolak apabila $t_{test} \leq t_{tabel}$.

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 determinasi
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 | 25.600 |
| 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

$$R = \frac{n (\sum xy) - (\sum x) \cdot (\sum y)}{\sqrt{n (\sum x^2) - (\sum x)^2} \sqrt{n (\sum y^2) - (\sum y)^2}}$$

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

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

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

$$R = 0,091$$

- Analisis Koefisien determinasi

$$\begin{aligned} R^2 &= R \times R \\ &= 0,091 \times 0,091 \\ &= 0,008 \end{aligned}$$

2. Persamaan regresi

$$Y = a + bx$$

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

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

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

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

$$b = \mathbf{0,215}$$

$$a = \frac{\sum y - b(\sum x)}{n}$$

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

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

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

$$a = \mathbf{149,114}$$

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.