

ANALISIS REGRESI SEDERHANA

Dr. REFI ARIOEN, STP., MTA

Magiste Manajemen Teknologi

- ❑ Analisa regresi digunakan untuk mempelajari dan mengukur hubungan statistik yang terjadi antara dua atau lebih variabel. Dalam regresi sederhana dikaji dua variabel, sedangkan dalam regresi majemuk dikaji lebih dari dua variabel.
- ❑ Dalam analisa regresi suatu persamaan regresi hendak ditentukan dan digunakan untuk menggambarkan pola atau fungsi hubungan yang terdapat antar variabel.
- ❑ Variabel yang akan diestimasi nilainya disebut variabel terikat (*dependent variable* atau *response variable*) dan biasanya diplot pada sumbu tegak (sumbu-y). Sedangkan variabel bebas (*independent variable* atau *explanatory variable*) adalah variabel yang diasumsikan memberikan pengaruh terhadap variasi variabel terikat dan biasanya diplot pada sumbu datar (sumbu-x).

- **Tujuan Regresi** → menguji pengaruh antara satu variabel terhadap variabel lain.
- **Ada 2 variabel:**
 1. **Variabel Independent (variabel bebas)** → variabel yang mempengaruhi.
 2. **Variabel Dependent (variabel tergantung)** → variabel yang dipengaruhi.
- **Analisis Regresi** digunakan terutama untuk tujuan peramalan, dimana dalam model tersebut ada sebuah variabel dependent (terikat/tergantung) dan satu atau lebih variabel independent (bebas).
- **Dalam analisis regresi ini, akan dikembangkan sebuah *estimating equation* (persamaan regresi) yaitu suatu formula matematika yang mencari nilai variabel dependent dari hasil variabel independent yang diketahui.**

- **Pada analisis regresi yang dilakukan, dibedakan menjadi:**

- 1. Regresi Sederhana**

Persamaan regresi yang hanya melibatkan 1 variabel dependent dan 1 variabel independent.

- 2. Regresi Berganda**

Persamaan regresi yang melibatkan 1 variabel dependent dan lebih dari 1 variabel independent.

Pada regresi ada beberapa uji yang harus dilakukan di samping uji koefisien dan uji kelinearan. Uji tersebut adalah uji kolinearitas dan uji multikolinieritas.

a.Kolinearitas adalah keadaan dimana variabel-variabel independent dalam persamaan regresi memiliki hubungan kuat satu sama lain. Uji kolinearitas dapat dilakukan dengan Coolinearity Diagnostic. Hasilnya sebagai berikut:

- 1) Skor VIF (*Variabel inflation Factor*) $< 0,5$ tidak ada kolinieritas.**
 - 2) Skor VIF (*Variabel inflation Factor*) $> 0,5$ ada kolinieritas.**
 - 3) Skor VIF (*Variabel inflation Factor*) > 10 ada multikolinieritas serius.**
- Analisis dengan Coolinearity Diagnostic sebagai berikut:**
 - ➔ Dari baris menu pilih Analyze, kemudian pilih Regression.**
 - ➔ Dari serangkaian test, pilih Linear.**
 - ➔ Pindahkan variabel pada kotak Dependent dan ke kotak Independent.**
 - ➔ Pilih Statistic lalu check list Coolinearity Diagnostic.**

- b. Uji multikolinearitas bertujuan untuk menguji apakah model regresi ditemukan adanya korelasi antar lebih dari tiga variabel bebas (independent). Model regresi yang baik seharusnya tidak terjadi korelasi di antara variabel bebas. Jika variabel bebas saling berkorelasi, maka variabel-variabel ini tidak ortogonal. Variabel ortogonal adalah variabel bebas yang nilai korelasi antar sesama variabel bebas sama dengan nol.**

Beberapa teknik regresi:

- 1) Metode Entered yaitu penghitungan nilai variabel dalam satu langkah.**
- 2) Metode Backward yaitu variabel-variabel dimasukkan sekaligus, variabel bebas akan dieliminasi satu persatu dari variabel yang kolinieritasnya tinggi.**
- 3) Metode Forward yaitu variabel-variabel yang dimasukkan satu demi satu, jika ada kolinieritas maka variabel tersebut akan keluar.**
- 4) Metode Stepwise yaitu variabel-variabel dimasukkan satu demi satu, dimulai dari variabel yang memiliki hubungan paling kuat, namun variabel yang keluar tadi dapat masuk kembali.**

1. REGRESI SEDERHANA

- **Regresi sederhana dapat didefinisikan sebagai pengaruh antara 2 variabel saja, dimana terdiri dari 1 variabel independent/bebas dan 1 variabel dependent/terikat.**
- **Digunakan juga untuk membangun persamaan dan menggunakan persamaan tersebut untuk membuat perkiraan (*prediction*).**

Fungsi regresi linear dapat dinyatakan dalam hubungan matematis oleh: $Y = A + BX$. Sebagai misal $Y = 2 + 1,4X$, secara teoritis bila $X = 10$, maka $Y = 16$. Pada kenyataannya tidak demikian, sebab yang mempengaruhi Y bukan hanya X tetapi ada faktor lain yang tidak dimasukkan dalam persamaan, faktor tersebut secara keseluruhan disebut sebagai “kesalahan” (disturbance’s error). Adanya kesalahan ini menjadikan perkiraan menjadi tidak akurat, selalu ada resiko yang disebabkan oleh adanya kesalahan. Kesalahan ini tidak dapat dihilangkan sama sekali, maka resiko ini harus diperkecil sekecil mungkin dengan memperkecil kesalahan. Dengan memperhitungkan kesalahan, regresi linear dinyatakan sebagai $Y = A + BX + \epsilon$.

Asumsi yang digunakan dalam regresi linear adalah sebagai berikut:

- a. $E(\epsilon_i) = 0$
- b. $E(\epsilon_i^2) = \sigma^2$
- c. $E(\epsilon_i \epsilon_j) = \text{cov}(\epsilon_i, \epsilon_j) = 0$
- d. X_i konstan

Untuk memperkirakan A dan B dipergunakan metode kuadrat kesalahan terkecil, dimana

Model sebenarnya : $Y = A + BX + \epsilon$

Model perkiraan : $Y = a + bX + e$

a, b, dan e adalah penduga untuk A, B, dan ϵ

$$Y_i = a + bX_i + e_i \text{ atau } e_i = Y_i - (a + bX_i) \text{ dan } \sum_i e_i^2 = \sum (Y_i - (a + bX_i))^2.$$

penurunan parsial terhadap a dan b yang sederhana diperoleh

$$a = \bar{Y} - b\bar{X} = \frac{\sum_i Y_i \sum_i X_i^2 - \sum_i X_i \sum_i X_i Y_i}{n \sum_i X_i^2 - \left(\sum_i X_i \right)^2} \quad \text{dan}$$

$$b = \frac{n \sum_i X_i Y_i - \sum_i X_i \sum_i Y_i}{n \sum_i X_i^2 - \left(\sum_i X_i \right)^2}$$

Nilai variabel A dan B untuk populasi diberikan oleh

$$A = \bar{Y} - B\bar{X} \quad \text{dan}$$

$$B = \frac{E(XY) - E(X)E(Y)}{E(X^2) - \{E(X)\}^2} = \frac{\text{cov}(X, Y)}{\text{var}(X)} = \frac{\sigma_{xy}}{\sigma_x}$$

Bila $s_{xy} = \frac{1}{n} \left(\sum_i X_i Y_i - \sum_i X_i \sum_i Y_i / n \right)$ adalah penduga untuk σ_{xy} dan

$s_x^2 = \frac{1}{n} \left(\sum_i X_i^2 - \left(\sum_i X_i \right)^2 / n \right)$ adalah penduga untuk σ_x^2 , maka

$$b = \frac{s_{xy}}{s_x^2} = \frac{\sum_i x_i y_i}{\sum_i x_i^2}$$

$$\text{dimana } \sum_i x_i y_i = \left(\sum_i X_i Y_i - \sum_i X_i \sum_i Y_i / n \right) \text{ dan } \sum_i x_i^2 = \left(\sum_i X_i^2 - \left(\sum_i X_i \right)^2 / n \right)$$

$$\text{var}(b) = \sigma_b^2 = \frac{\sigma_e^2}{\sum_i x_i^2} \text{ dan } \text{var}(a) = \sigma_a^2 = \sigma_e^2 \left(\frac{1}{n} + \frac{\bar{X}^2}{\sum_i x_i^2} \right)$$

$$\text{cov}(a, b) = \sigma_{a,b} = \frac{-\bar{X}}{\sum_i x_i^2} \sigma_e^2$$

Contoh 1

Dari suatu praktikum fisika dasar diperoleh data yang menghubungkan variabel bebas x dan variabel terikat y seperti ditunjukkan dalam tabel berikut.

Uji ke-	x	y
1	6	30
2	9	49
3	3	18
4	8	42
5	7	39
6	5	25
7	8	41
8	10	52
Σ	56	296

Jika berdasarkan kajian teoritis dan sifat dari fenomena yang menghubungkan x dan y dapat diasumsikan mempunyai bentuk hubungan linier, maka persamaan garis regresinya dapat ditentukan sebagai berikut.

Tabel perhitungan:

Uji ke-	x	y	xy	x^2	y^2
1	6	30	180	36	900
2	9	49	441	81	2401
3	3	18	54	9	324
4	8	42	336	64	1764
5	7	39	273	49	1521
6	5	25	125	25	625
7	8	41	328	64	1681
8	10	52	520	100	2704
Σ	56	296	2257	428	12920
$\bar{x} = \frac{\sum x}{n} = \frac{56}{8} = 7 \qquad \bar{y} = \frac{\sum y}{n} = \frac{296}{8} = 37$					

Kolom y^2 ditambahkan pada tabel meskipun belum digunakan untuk perhitungan persamaan garis regresi. Nilai tersebut akan digunakan kemudian. Jadi dengan menggunakan hasil pada tabel, nilai dari konstanta a dan b dapat ditentukan:

$$b = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2} = \frac{8(2257) - (56)(296)}{8(428) - (56)^2} = \frac{1480}{288} = 5,1389$$

$$a = \bar{y} - b\bar{x} = 37 - (5,1389)(7) = 1,0277$$

Jadi persamaan garis regresi linier yang menggambarkan hubungan antara variabel x dan y dari data sampel pada percobaan/praktikum di atas adalah:

$$\hat{y} = a + bx = 1,0277 + 5,1389x$$

Dengan menggunakan persamaan garis regresi yang diperoleh, maka dapat diperkirakan hasil yang akan diperoleh (nilai y) untuk suatu nilai x tertentu. Misalnya untuk $x = 4$ maka dapat diperkirakan bahwa y akan bernilai:

$$\hat{y} = a + bx = 1,0277 + 5,1389x = 1,0277 + 5,1389(4) = 21,583$$

❖ **Kasus**

PT Cemerlang dalam beberapa bulan gencar mempromosikan sejumlah peralatan elektronik dengan membuka outlet-outlet di berbagai daerah. Berikut data mengenai Penjualan dan Biaya Promosi yang dikeluarkan di 15 daerah di Indonesia.

Daerah	Sales (Juta Rupiah)	Promosi (Juta Rupiah)
Jakarta	205	26
Tangerang	206	28
Bekasi	254	35
Bogor	246	31
Bandung	201	21
Semarang	291	49
Solo	234	30
Yogya	209	30
Surabaya	204	24
Purwokerto	216	31
Madiun	245	32
Tuban	286	47
Malang	312	54
Kudus	265	40
Pekalongan	322	42

❖ **Penyelesaian:**

Disini karena akan diketahui besar hubungan atau seberapa jauh biaya promosi berpengaruh terhadap Penjualan PT. CEMERLANG, maka akan dilakukan uji regresi, dengan variabel dependent adalah Sales/Penjualan, dan variabel independent adalah Biaya Promosi.

➤ **Analyze → Regression → Linear**

➤ **Pengisian kotak dialog:**

- Ⓜ **Dependent;** masukkan variabel **Sales** pada kotak **Dependent**.
- Ⓜ **Independent (s);** masukkan variabel **promosi** ke kotak **Independent**.
- Ⓜ **Case Labels** atau keterangan pada kasus. Oleh karena kasus didasarkan pada daerah-daerah, maka masukkan variabel **daerah** ke kotak **Case Labels**.
- Ⓜ **Method;** untuk keseragaman pilih **Enter**, yaitu prosedur pemilihan variabel dimana semua variabel dalam blok dimasukkan dalam perhitungan "*single step*".

❖ **Penyelesaian (lanjutan)**

- ④ **Pilih kolom Options dengan klik mouse pada pilihan tersebut.**

Pengisian:

- ▣ **Untuk Stepping Method Criteria, digunakan uji F yang mengambil standar angka probabilitas sebesar 5%. Oleh karena itu, angka Entry .05 atau 5% dipilih.**
- ▣ **Pilihan default Include constant in equation atau menyertakan persamaan regresi tetap dipilih.**
- ▣ **Missing Value; gunakan default dari SPSS, yaitu Exclude cases listwise.**
- ▣ **Continue**

❖ **Penyelesaian (lanjutan)**

Ⓢ **Pilih kolom Statistics**

Pengisian:

- ▣ **Regression Coefficient** atau perlakuan koefisien regresi, tetap aktifkan pilihan **Estimate**.
- ▣ Pada pilihan **Descriptive** pada kolom sebelah kanan, tetap aktifkan **Model Fit**.
- ▣ **Residuals** → **Casewise Diagnostics** → **All Cases** untuk melihat pengaruh regresi terhadap semua daerah
- ▣ **Continue**

❖ **Penyelesaian (lanjutan)**

- ④ **Pilih kolom Plots atau berhubungan dengan gambar/grafik untuk regresi. Direncanakan ada 3 plot (gambar) sehubungan dengan analisis regresi.**

Pengisian:

- ▣ **Klik pilihan SDRESID dan masukkan ke pilihan Y, lalu klik ZPRED dan masukkan ke pilihan X. Klik Next untuk melanjutkan pengisian plot kedua.**
- ▣ **Klik ZPRED dan masukkan ke pilihan Y, lalu klik DEPENDNT dan masukkan ke pilihan X. Klik Next untuk melanjutkan pengisi plot ketiga.**
- ▣ **Untuk plot ketiga, pada pilihan Standardized Residual Plots, klik Normal Probability Plot.**
- ▣ **Continue**
- ▣ **Ok**

Terima Kasih

