

Regresi Linier dan Korelasi

Program Sarjana Institut Teknologi Del

Outline

1. Konsep Regresi linier dan korelasi
2. Regresi linier sederhana
3. Sifat penaksiran metode kuadrat terkecil (MKT) / OLS (Ordinary Least Square)
4. Inferensi mengenai koefisien Regresi
5. Fitting data dan garis regresi
6. Pendekatan melalui analisis variansi
7. Korelasi

Konsep Regresi linier dan korelasi

- Regresi adalah model/ persamaan yang menggambarkan hubungan antara variabel prediktor (independen) terhadap variabel respon (dependen)
- Korelasi adalah sebuah ukuran yang menyatakan tingkat keeratan hubungan dua variabel atau lebih (nilai korelasi antara -1 s.d 1). Ingat mengenai kondisi variabel saat korelasi positif dan negatif.

Regresi linier sederhana

Dalam menggambarkan hubungan antara prediktor (X) dan respon (Y), kadang data membentuk pola tertentu. Model-model regresi antara lain:

1. Regresi Linier (dibagi menjadi regresi linier sederhana <prediktor hanya satu> dan regresi linier berganda/majemuk <prediktor lebih dari satu>)
2. Regresi Polinomial
3. Regresi eksponensial
4. dll

Model Regresi Linier Sederhana di Populasi

$$Y_i = \alpha + \beta X_i + \varepsilon_i$$

- Y_i = nilai ke-i dari variabel Y
- α = konstanta, rata-rata Y jika nilai semua X adalah nol.
- β = konstanta, koefisien regresi Y pada X
- ε_i = eror ke-i (pada model biasanya dianggap nol karena menyebar normal dengan rata-rata nol dan ragam tertentu)
- α diperoleh dari a pada sampel
- β diperoleh dari b pada sampel
 - ε_i diperoleh dari e_i

Garis Regresi Linier Sederhana

$$\hat{Y} = a + bX$$

- ❑ \hat{Y} adalah nilai prediksi, diprediksi dari nilai X tertentu
- ❑ Biasanya nilai prediksi dari Y tidak sama dengan nilai sebenarnya
- ❑ Perbedaan antara nilai sebenarnya dan nilai prediksi disebut eror dan dilambangkan dengan e pada sampel.
$$e_i = Y_i - \hat{Y}_i$$
- ❑ Error tidak ditulis lagi karena diasumsikan menyebar normal dengan rata-rata nol dan ragam tertentu

Penduga α dan β dengan Metode Kuadrat Terkecil (MKT/ OLS)

- Mencari a dan b , taksiran α dan β , sehingga meminimumkan jumlah kuadrat galat (JKG).

$$JKG = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - a - bx_i)^2$$

- Dengan menggunakan teorema dalam kalkulus, JKG akan minimum jika nilai a dan b adalah :

$$b = \frac{n(\sum XY) - (\sum X)(\sum Y)}{n\sum X^2 - (\sum X)^2}$$

Bagaimana
pendugaan
parameter dalam
regresi linier
berganda?

$$a = \frac{(\sum Y)(\sum X^2) - (\sum X)(\sum XY)}{n\sum X^2 - (\sum X)^2} = \frac{\sum Y}{n} - b \frac{\sum X}{n} = \bar{Y} - b\bar{X}$$

Penduga parameter regresi linier dengan matriks

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

dimana,

$$\mathbf{b} = \begin{bmatrix} b_0 \\ \vdots \\ b_k \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & x_{11} & \dots & x_{k1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & \dots & x_{kn} \end{bmatrix}, \mathbf{Y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

Dari model regresi $\mathbf{Y} = \mathbf{bX}$

Sifat penaksiran MKT / OLS adalah BLUE (Best Linier Unbias Efisien)

- Linier = membentuk model regresi linier
- Unbias = penaksir sama dengan parameter
- Efisien = varians minimum

Inferensi mengenai koefisien Regresi

- Uji Serentak/ Keseluruhan menggunakan uji F (ANOVA)
H0 : seluruh prediktor tidak berpengaruh ($\beta_I = 0$)
H1 : minimal satu prediktor berpengaruh ($\beta_I \neq 0$)
 $I = 1, 2, 3, \dots, k = \text{banyak prediktor}$
- Uji Parsial/ Masing-masing koefisien menggunakan uji t
H0 : prediktor ke-I tidak berpengaruh ($\beta_I = 0$)
H1 : prediktor ke-I berpengaruh ($\beta_I \neq 0$)
Pasangan hipotesis sebanyak k (sesuai dengan banyak prediktor)

Fitting data dan garis regresi

- Fitting data regresi menggunakan diagram pencar (diagram titik pasangan X dan Y)
- Garis regresi digambarkan berdasarkan persamaan yang terbentuk dengan a dan b hasil penaksiran MKT.

Contoh data nilai matematika mempengaruhi nilai fisika:

Mat (X)	Fisika (Y)	XY	X ²	Y ²
60	80	4800	3600	6400
45	69	3105	2025	4761
50	71	3550	2500	5041
60	85	5100	3600	7225
50	80	4000	2500	6400
65	82	5330	4225	6724
60	89	5340	3600	7921
65	93	6045	4225	8649
50	76	3800	2500	5776
65	86	5590	4225	7396
45	71	3195	2025	5041
50	69	3450	2500	4761
$\Sigma X=665$	$\Sigma Y=951$	$\Sigma XY=53305$	$\Sigma X^2=37525$	$\Sigma Y^2=76095$

Menduga garis regresi dengan MKT:

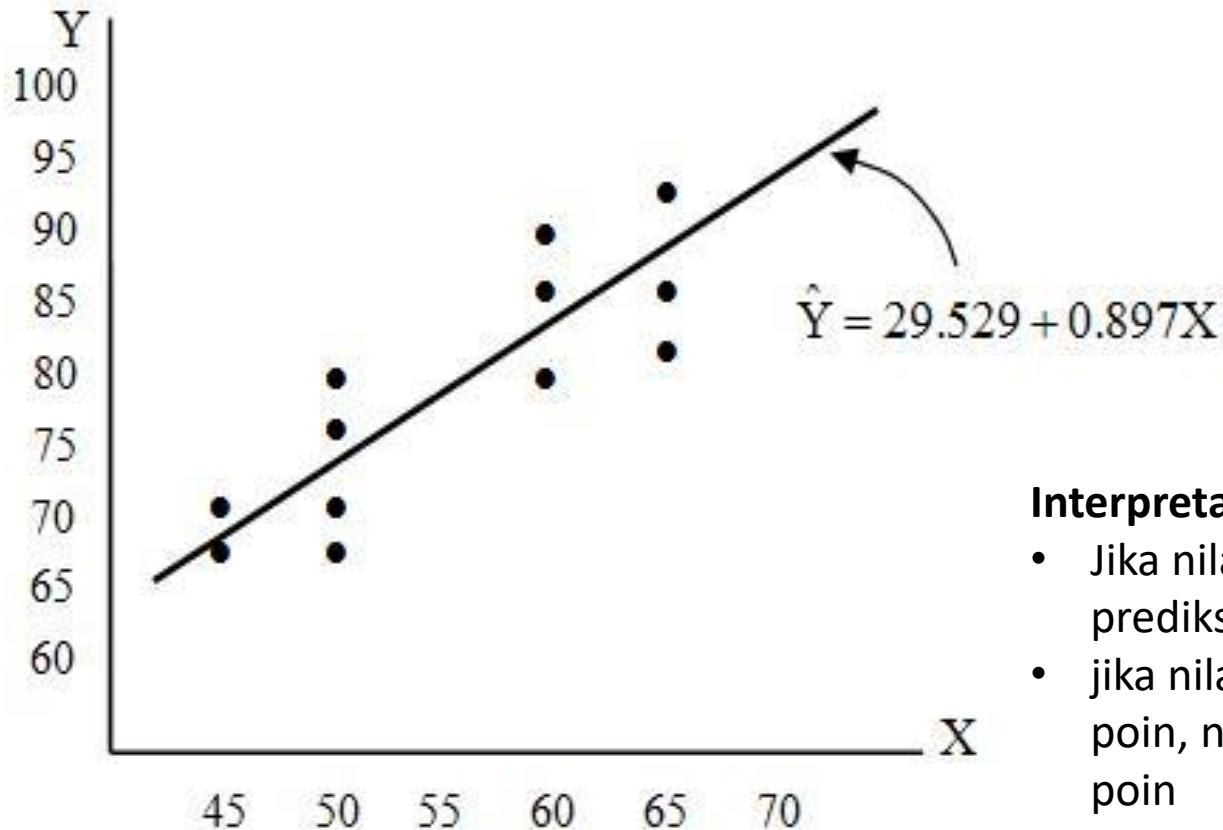
$$b = \frac{(12)(53305) - (665)(951)}{(12)(37525) - (665)^2} = \frac{7245}{8075} \\ = 0.897213622$$

$$a = \frac{951}{12} - 0.897 \left(\frac{665}{12} \right) = 29.55 \quad \text{atau}$$

$$a = \frac{(951)(37525) - (665)(53305)}{(12)(37525) - (665)^2} = \frac{238450}{8075} \\ = 29.52941176$$

Sehingga Garis Regresi : $\hat{Y} = 29.529 + 0.897X$

Diagram Pencar



Interpretasi:

- Jika nilai matematika nol, prediksi nilai fisika adalah 29,5.
- jika nilai matematika naik satu poin, nilai fisika akan naik 0,897 poin

Pendekatan melalui analisis variansi (ANOVA)

- Analisis variansi/ Analysis of Varians (ANOVA) digunakan untuk menguji hipotesis rata-rata lebih dari 2 populasi dengan konsep membandingkan ragam (uji F)

Sumber variasi	Jumlah kuadrat	Derajat kebebasan	Rataan kuadrat	f hitungan
Regresi	JKR	1	JKR	JKR/s^2
Galat	JKG	$n - 2$	$s^2 = \frac{JKG}{n - 2}$	
Total	JKT	$n - 1$		

- Pada regresi, ANOVA digunakan untuk menguji apakah koefisien regresi secara serentak berpengaruh dalam model.

Korelasi

Analisis korelasi : mengukur seberapa eratnya hubungan antara dua peubah acak dengan menggunakan suatu bilangan (koefisien korelasi).

Korelasi berdasarkan skala data:

1. Nominal/ ordinal menggunakan korelasi Rank Spearman
2. Interval/ ratio menggunakan korelasi Pearson/ Product Moment

Korelasi Pearson yang dikuadratkan membentuk KOEFISIEN DETERMINASI yang digunakan untuk mengukur seberapa baik model regresi (R^2 besar model semakin baik)

Korelasi berdasarkan arah hubungannya dapat dibedakan, jadi berapa ?

1. Korelasi Positif

Jika arah hubungannya searah

2. Korelasi Negatif

Jika arah hubungannya berlawanan arah

3. Korelasi Nihil

Jika perubahan kadang searah tetapi kadang berlawanan arah.

KORELASI PEARSON/PRODUCT MOMENT

- Digunakan untuk menentukan besarnya koefisien korelasi jika data yang digunakan berskala interval atau rasio.
- Rumus yang digunakan:

$$r_{xy} = \frac{n \sum x_i y_i - (\sum x_i)(\sum y_i)}{\sqrt{\{n \sum x_i^2 - (\sum x_i)^2\}} \sqrt{\{n \sum y_i^2 - (\sum y_i)^2\}}}$$

Contoh Kasus:

Seorang manager melakukan survai untuk meneliti apakah ada korelasi antara biaya produksi dan pendapatan di perusahaan X.

Untuk menjawab permasalahan tersebut diambil sampel sebanyak 10 anak perusahaan.

Biaya produksi	2	4	6	6	8	8	9	8	9	10
Pendapatan	10	20	50	55	60	65	75	70	81	85

Analisis Data Korelasi Pearson

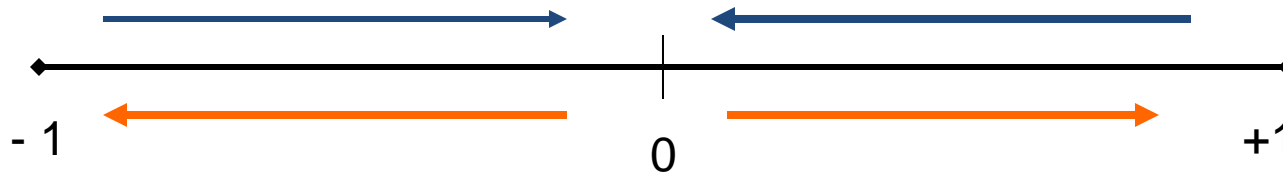
N	Xi	Yi	Xi^2	Yi^2	XY
1	2	10	4	100	20
2	4	20	16	400	80
3	6	50	36	2500	300
4	6	55	36	3025	330
5	8	60	64	3600	480
6	8	65	64	4225	520
7	9	75	81	5625	675
8	8	70	64	4900	560
9	9	81	81	6561	729
10	10	85	100	7225	850
Jlh	70	571	546	38161	4544

$$r_{xy} = \frac{10(4544) - (70)(571)}{\sqrt{\{10(546) - (70)^2\}} \sqrt{\{10(38161) - (571)^2\}}} = 0,981$$

Berapa Nilai Koefesien Korelasi ?

- Koefesien korelasi akan selalu sebesar :

$$- 1 \leq r \leq + 1$$



Berikut adalah pedoman untuk memberikan interpretasi serta analisis bagi koefisien korelasi menurut Prof. Dr. Sugiyono.

Dalam buku Metode Penelitian Administasi (Bandung : Alfabeta) adalah sebagai berikut

0.00 - 0,199 = sangat rendah

0,20 - 0,3999 = rendah

0,40 - 0,5999 = sedang

0,60 - 0,799 = kuat

0,80 - 1,000 = sangat kuat

Korelasi By Deborah Rumsey

- **Exactly -1 .** A perfect downhill (negative) linear relationship
- **-0.70 .** A strong downhill (negative) linear relationship
- **-0.50 .** A moderate downhill (negative) relationship
- **-0.30 .** A weak downhill (negative) linear relationship
- **0 .** No linear relationship
- **$+0.30$.** A weak uphill (positive) linear relationship
- **$+0.50$.** A moderate uphill (positive) relationship
- **$+0.70$.** A strong uphill (positive) linear relationship
- **Exactly $+1$.** A perfect uphill (positive) linear relationship

<http://www.dummies.com/education/math/statistics/how-to-interpret-a-correlation-coefficient-r/>