Metode Statistika
Pertemuan XIII

Analisis Regresi dan Korelasi

Pengantar

- Apa itu analisis regresi?
- Apa bedanya dengan korelasi?

Analisis Regresi → Analisis statistika yang memanfaatkan hubungan antara dua atau lebih peubah kuantitatif sehingga salah satu peubah dapat diramalkan dari peubah lainnya.

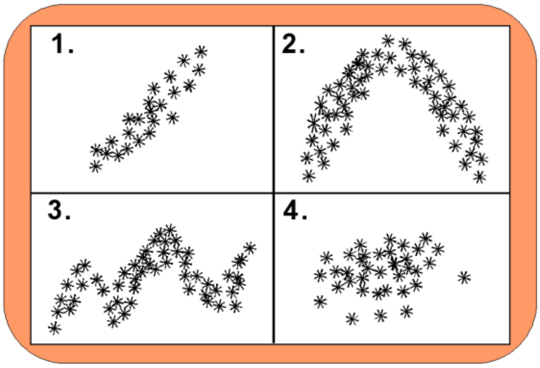
Korelasi → Analisis statistika yang mengukur derajat keeratan hubungan (linear) antara dua peubah kuantitatif.

REGRESI

Analisis statistika utk menentukan pola (kurva) hub antara 2 atau lebih peubah kuantitatif. Misal: Y = berat bdn, X = tinggi bdn

Hub antara Y dan X:
- Linear
- Kuadratik
- Kubik, dsb

Relationships between Continuous Variables



ANALISIS REGRESI

- Hubungan Antar Peubah:
  - Fungsional (deterministik) → Y=f(X) ; mis: Y=10X
  - Statistik (stokastik) → amatan tdk jatuh pas pd kurva

    Mis: IQ vs Prestasi, Berat vs Tinggi,

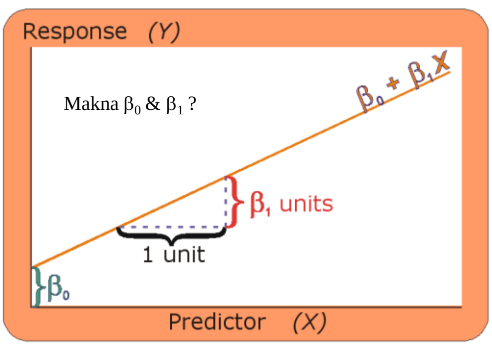
    Dosis Pupuk vs Produksi

Model regresi sederhana:

Populasi ➡ $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \;\; ; i = 1,2,\ldots,n$

Dugaan (sampel) ➡ $\hat{Y}_i = b_0 + b_1 X_i$

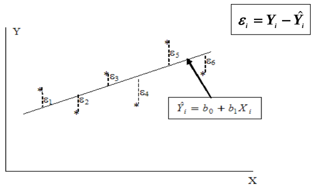
Simple Linear Regression Model



Pendugaan Parameter Regresi: β_0 & β_1

- Metode Kuadrat Terkecil:

  minimum JK Galat = min $\sum \varepsilon_i^2$



➢ b_0 penduga bagi β_0 dan b_1 penduga bagi β_1

$$\varepsilon_i = Y_i - \hat{Y}_i = Y_i - b_0 - b_1 X_i$$

$$\min \sum_{i=1}^{n} \varepsilon_i^2 \rightarrow \partial\{\sum_{i=1}^{n} (Y_i - b_0 - b_1 X_i)^2\} / \partial b_i = 0$$

$$\frac{\partial}{\partial b_0}\left[\sum_{i=1}^{n} (Y_i - b_0 - b_1 X_i)^2\right] = 0 \Rightarrow \; n b_0 + b_1 \sum_{i=1}^{n} X_i = \sum_{i=1}^{n} Y_i$$

$$\frac{\partial}{\partial b_1}\left[\sum_{i=1}^{n} (Y_i - b_0 - b_1 X_i)^2\right] = 0 \Rightarrow \; b_0 \sum_{i=1}^{n} X_i + b_1 \sum_{i=1}^{n} X_i^2 = \sum_{i=1}^{n} X_i Y_i$$

$$b_1 = \frac{\sum_{i=1}^{n} X_i Y_i - (\sum_{i=1}^{n} X_i)(\sum_{i=1}^{n} Y_i)/n}{\sum_{i=1}^{n} X_i^2 - (\sum_{i=1}^{n} X_i)^2/n} = \frac{JHK(XY)}{JK(X)}$$

$$b_0 = \bar{Y} - b_1 \bar{X}$$

Contoh Data

Percobaan dalam bidang lingkungan

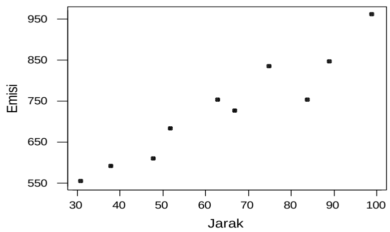
Apakah semakin tua mobil semakin besar juga emisi HC yang dihasilkan?

Diambil contoh 10 mobil secara acak, kemudian dicatat jarak tempuh yang sudah dijalani mobil (dalam ribu kilometer) dan diukur Emisi HC-nya (dalam ppm)

Jarak	Emisi
31	553
38	590
48	608
52	682
63	752
67	725
75	834
84	752
89	845
99	960

Analisis Regresi

Plot antara Emisi HC (ppm) dg Jarak Tempuh Mobil (ribu kilometer)



	Y	X	Y^2	X^2	XY
	553	31	305809	961	17143
	590	38	348100	1444	22420
	608	48	369664	2304	29184
	682	52	465124	2704	35464
	752	63	565504	3969	47376
	725	67	525625	4489	48575
	834	75	695556	5625	62550
	752	84	565504	7056	63168
	845	89	714025	7921	75205
	960	99	921600	9801	95040
Total =	7301	646	5476511	46274	496125
Rataan =	730,1	64,6			

$$n = 10; \sum X = 646; \sum Y = 7301; \sum XY = 496125$$
$$\sum X^2 = 46274; \sum Y^2 = 5476511$$

$$JHK(XY) = 496125 - (646)(7301)/10 = 24480,4$$
$$JK(X) = 46274 - (646)^2/10 = 4542,4$$

Kemiringan garis regresi (b_1):

$$b_1 = \frac{JHK(XY)}{JK(X)} = \frac{2448,4}{4542,4} = 5,39$$

Intersep (b_0):

$$b_0 = \bar{Y} - b_1\bar{X} = 730,1 - 5,39(64,6) = 381,95$$

Persamaan regresi:

$$\hat{Y} = 381{,}95 + 5{,}39X$$
$$atau$$
$$Emisi = 381{,}95 + 5{,}39 Jarak$$

- Seberapa layakkah persamaan di atas dpt digunakan utk dpt meramalkan besarnya emisi HC (Y) berdasarkan besarnya jarak yang ditempuh (X) oleh sebuah mobil?

Dua langkah yang perlu dilakukan:

1. Uji terhadap model regresi ??
   - bersama (model) → uji-F (Anova)
   - parsial (per koefisien) → uji-t

2. Hitung nilai kesesuaian model ??
   R^2

(Koef. Determinasi: % keragaman Y yang mampu dijelaskan oleh X)

Uji thd Model

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

(model tdk nyata) (model nyata)

ANOVA (Analysis of Variance) → Uji F

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

JK total = JK regresi + JK galat

Keragaman total = keragaman yang dapat dijelaskan oleh model + keragaman yang tidak dapat dijelaskan oleh model

Anova

Sumber	db	JK	KT	F
Regresi	1	JKR	KTR	KTR/KTG
Galat	n - 2	JKG	KTG	
Total	n - 1	JKT		

$F \sim F_{(1,n-2)}$

$$JKT = \sum_{i=1}^{n} Y_i^2 - (\sum_{i=1}^{n} Y_i)^2 / n$$
$$JKR = b_1 \times JHK(XY)$$
$$JKG = JKT - JKR$$
$$KT = JK / db \rightarrow KTR = JKR / 1 = JKR$$
$$KTG = JKG / (n-2)$$

$$JKT = 5476511 - (7301)^2 / 10 = 146050.9$$
$$JKR = 5.39(24480.4) = 131932.5$$
$$JKG = 146050.9 - 131932.5 = 14118.4$$
$$KTR = 131932.5$$
$$KTG = 14118.4 / 8 = 1764.8$$

Anova

Sumber	db	JK	KT	F
Regresi	1	131932.5	131932.5	74.76
Galat	8	14118.4	1764.8	
Total	9	146050.9		

$F_{(0.05,1,8)} = 5.32$; $F_{(0.01,1,8)} = 11.26$

Kesimpulan: H_0 ditolak, model regresi nyata pada taraf nyata 0.01.

Uji Parsial thd Koefisien β_0 dan β_1

Pengujian kemiringan garis (β_1):

(i) H_0: $\beta_1 \geq \beta_{10}$ vs H_1: $\beta_1 < \beta_{10}$;

(ii) H_0: $\beta_1 \leq \beta_{10}$ vs H_1: $\beta_1 > \beta_{10}$

(iii) H_0: $\beta_1 = \beta_{10}$ vs H_1: $\beta_1 \neq \beta_{10}$

- Statistik uji: $t_h = \dfrac{b_1 - \beta_{10}}{s_{b_1}}$ pada (n−2) derajat bebas

  dimana $s_{b_1} = \sqrt{\dfrac{KTG}{JKX}}$
- Khusus, untuk menguji apakah $\beta_1 = 0$, diperoleh

$$t_h = \frac{b_1}{\sqrt{\dfrac{KTG}{JKX}}}$$

Kriteria penolakan H_0 :

- $t_{hit} < -t_{\alpha\,(n-2)}$, untuk H_1: $\beta_1 < \beta_{10}$,
- $t_{hit} > t_{\alpha\,(n-2)}$, untuk H_1: $\beta_1 > \beta_{10}$, atau
- $|t_{hit}| > t_{\alpha/2\,(n-2)}$, untuk H_1: $\beta_1 \neq \beta_{10}$.

Pengujian Intersep (β_0)

- H_0: $\beta_0 = \beta_{00}$
- Statistik uji: $t_h = \dfrac{b_0 - \beta_{00}}{\sqrt{KTG\left(\dfrac{1}{n} + \dfrac{\bar{x}^2}{JKX}\right)}}$
- Kriteria penolakan H_0: sama dg uji β_1

Contoh:

a) Apakah laju perubahan emisi HC per 10000 km perubahan jarak lebih besar dari 5 ppm? Gunakan taraf nyata 5%.
b) Apakah tingkat emisi HC kendaraan baru (= 0 km jarak yg ditempuh) tidak berbeda dengan 250 ppm? Gunakan taraf nyata 5%.

c) H_0: $\beta_1 \le 5$ vs H_1: $\beta_1 > 5$ Jawab:

JK(X) = 4542.4, KTG = 1764.8 $\Rightarrow$ $s_{b_1} = \sqrt{\dfrac{1764.8}{4542.4}} = 0.623$

Statistik uji: $t_h = \dfrac{5.39 - 5}{0.623} = 0.626$

Daerah kritis: $t_h > t_{0.05(8)} \rightarrow t_h > 1.860$

Kesimpulan: terima H_0, laju perubahan emisi HC per 10000 km perubahan jarak paling besar sama dengan 5 ppm.

b) H_0: $\beta_0 = 250$ vs H_1: $\beta_0 \neq 250$

$$s_{b_0} = \sqrt{1764.8(1/10 + 64.6^2 / 4542.4)} = 42.4$$

Statistik uji: $t_h = \dfrac{b_0 - \beta_{00}}{s_{b_0}} = \dfrac{381.95 - 250}{42.4} = 3.112$

Daerah kritis: $|t_h| > t_{0.025(8)} = 2.306$

Kesimpulan: tolak H_0, tingkat emisi HC kendaraan baru (= 0 km jarak yg ditempuh) berbeda nyata dengan 250 ppm

Ukuran Kesesuaian Model

- Koefisien determinasi:

  R^2 = (JKR/JKT) x 100%

$\Rightarrow$ % keragaman Y yang mampu diterangkan oleh X

Untuk model emisi HC:

R^2 = (131932.5/146050.9)x100% = 90.3%

artinya: 90.3% keragaman dalam Emisi HC dpt diterangkan oleh Jarak sedangkan sisanya, 9.7%, diterangkan oleh komponen lain yg bersifat acak (galat).

Peramalan (Prediksi)

Prediksi (peramalan) nilai peubah terikat (Y) berdasarkan nilai peubah bebas (X), X = x_0.

x_0 hrs ada dlm kisaran X hasil pengamatan.

Penduga titik bagi Y: $\hat{Y}_{X=x_0} = b_0 + b_1 x_0$

Penduga ragam bagi $\hat{Y}$:

$$s^2(\hat{Y}) = KTG\left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{JK(X)}\right)$$

Selang kepercayaan (1-α)100% bagi $\hat{Y}_{X=x_0}$:

$$(b_0 + b_1 x_0) \pm t_{\alpha/2(n-2)} \sqrt{KTG\left\{\ 1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{JK(X)}\ \right\}}$$

- Berapa besar emisi HC dihasilkan bila jarak = 50000 km? Tentukan selang kepercayaan 95% bagi tingkat emisi HC tsb.

$$\hat{Y}_{X=50} = 381.95 + 5.39(50) = 651.45$$

$$s^2(\hat{Y}) = 1764.8\left\{1 + 1/10 + (50 - 64.6)^2 / 4542.4\right\} = 2024.103$$

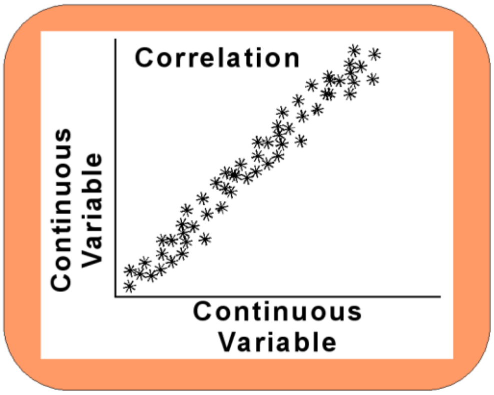
Selang kepercayaan 95% bagi emisi HC adalah:

$$651.45 - 2.306\sqrt{2024.103} < \hat{Y}_{X=50} < 651.45 + 2.306\sqrt{2024.103}$$
$$547.7 < \hat{Y}_{X=50} < 755.2$$

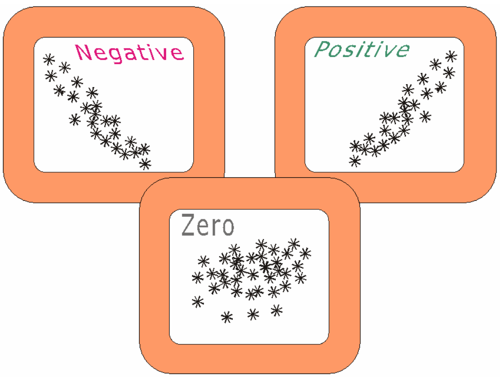
Regression Analysis: Emisi versus Jarak

```
The regression equation is
Emisi = 382 + 5.39 Jarak

Predictor    Coef  SE Coef     T      P
Constant   381.95    42.40  9.01  0.000
Jarak      5.3893   0.6233  8.65  0.000

S = 42.0096   R-Sq = 90.3%   R-Sq(adj) = 89.1%

Analysis of Variance

Source          DF      SS      MS      F      P
Regression       1  131932  131932  74.76  0.000
Residual Error   8   14118    1765
Total            9  146051

Predicted Values for New Observations

New
Obs    Fit  SE Fit      95% CI          95% PI
  1  651.4    16.1  (614.3, 688.5)  (547.7, 755.2)

New
Obs  Jarak
  1   50.0
```

Overview



Korelasi



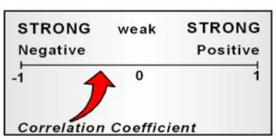
Koefisien Korelasi Pearson (r)

$$r_{xy} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\{\sum(x_i - \bar{x})^2\}\{\sum(y_i - \bar{y})^2\}}} = \frac{JHK(XY)}{\sqrt{JK(X)JK(Y)}}$$

$-1 \le r \le 1$

r sebagai penduga bagi ρ (koefisien korelasi populasi)

$|r| = \sqrt{R^2}$



Uji Korelasi:

H_0: $\rho = \rho_0$ vs H_1: $\rho < \rho_0$, $\rho > \rho_0$, atau $\rho \neq \rho_0$

Statistiik uji:

$$z_h = \frac{\sqrt{n-3}}{2}\left[\ln\left(\frac{1+r}{1-r}\right) - \ln\left(\frac{1+\rho_0}{1-\rho_0}\right)\right] = \frac{\sqrt{n-3}}{2}\ln\left[\frac{(1+r)(1-\rho_0)}{(1-r)(1+\rho_0)}\right]$$

Khusus utk H_0: $\rho = 0$ vs H_1: $\rho \neq 0$, statistik ujinya:

$$t_h = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \text{ pada (n-2) derajat bebas,}$$

Korelasi antara Emisi (Y) dan Jarak (X) adalah:

JHK(XY) = 24480.4; JK(X) = 4542.4;
JK(Y) = 146050.9

$$\Rightarrow \quad r = \frac{24480.4}{\sqrt{(4542.4)(146050.9)}} = 0.95$$

Uji thd koefisien korelasi: H_0: $\rho = 0$ vs H_1: $\rho \neq 0$

$$t_h = \frac{0.95\sqrt{8}}{\sqrt{1-0.95^2}} = 8.605$$

Daerah kritis pada $\alpha = 0.05$ adalah $|t_h| > t_{0.025(8)} = 2.306$

Kesimpulan: Tolak H_0, artinya ada korelasi linear positif yang kuat dan nyata antara emisi dan jarak dengan tingkat kepercayaan 95%.

Latihan:

1. Untuk data emisi HC, ujilah hipotesis bahwa H_0 : ρ = 0,90 vs H_1: ρ > 0,90 pada taraf nyata 5%.
2. X = curah hujan (0.01 cm)
   Y = debu yang terbawa (μg/m³)

Observasi	X	Y
1	4.3	126
2	4.5	121
3	5.9	116
4	5.6	118
5	6.1	114
6	5.2	118
7	3.8	132
8	2.1	141
9	7.5	108

Pertanyaan:

a. Tentukan persamaan garis regresinya
b. Ujilah model regresi tsb. Apakah model tsb nyata? Bila ya, pada taraf nyata berapa?
c. Hitunglah R^2. Apa artinya?
d. Buatlah SK 95% utk menduga banyaknya debu yg terbawa bila curah hujannya 0.048 cm.
e. Hitunglah r.
f. Ujilah H0: ρ = -0.5 lawan H1: ρ < -0.5 pada taraf nyata 5%.

X	Y	X^2	XY	Y^2
4.3	126	18.49	541.8	15876
4.5	121	20.25	544.5	14641
5.9	116	34.81	684.4	13456
5.6	118	31.36	660.8	13924
6.1	114	37.21	695.4	12996
5.2	118	27.04	613.6	13924
3.8	132	14.44	501.6	17424
2.1	141	4.41	296.1	19881
7.5	108	56.25	810	11664
45	1094	244.26	5348.2	133786

JKX = 19.26　　b1 = -6.3239875

JHK(XY) = -121.8　　b0 = 153.17549

JKY = 804.2222

JKT = 804.2222　　R^2 = 95.77722

JKR = 770.2617

Sumber	db	JK	KT	F
Regresi	1	770.2617	770.26168	158.7676
Galat	7	33.96054	4.8515057	
Total	8	804.2222		

$F_{(0.05,1,7)}$ = 5.591　　$F_{(0.01,1,7)}$ = 12.246

r = -0.97866