Statistika dan Probabilitas

Analisis Regresi

Deri Siswara

deri.siswara@perbanas.id

ABFI Institute Perbanas

November 26, 2025

Analisis Regresi

Apa itu Analisis Regresi?

Definisi

Analisis Regresi adalah metode statistik untuk memodelkan dan menganalisis hubungan antara satu variabel dependen (Y) dengan satu atau lebih variabel independen (X).

Tujuan:

- Memahami hubungan antar variabel
- Memprediksi nilai variabel dependen berdasarkan variabel independen
- Mengidentifikasi variabel yang paling berpengaruh

Jenis-jenis Regresi

Jenis Regresi	Karakteristik	Contoh
Regresi Linear Sederhana	1 variabel independen	Hubungan antara jam belajar dan nilai ujian
Regresi Linear Berganda	≥ 2 variabel independen	Hubungan antara pengalaman, pendidikan, dan gaji

Dengan regresi, kita dapat menjawab pertanyaan seperti: "Bagaimana perubahan jam belajar mempengaruhi nilai ujian?" dan membuktikan apakah hubungan tersebut *statistically significant*.

Regresi Linear Sederhana

Konsep Dasar

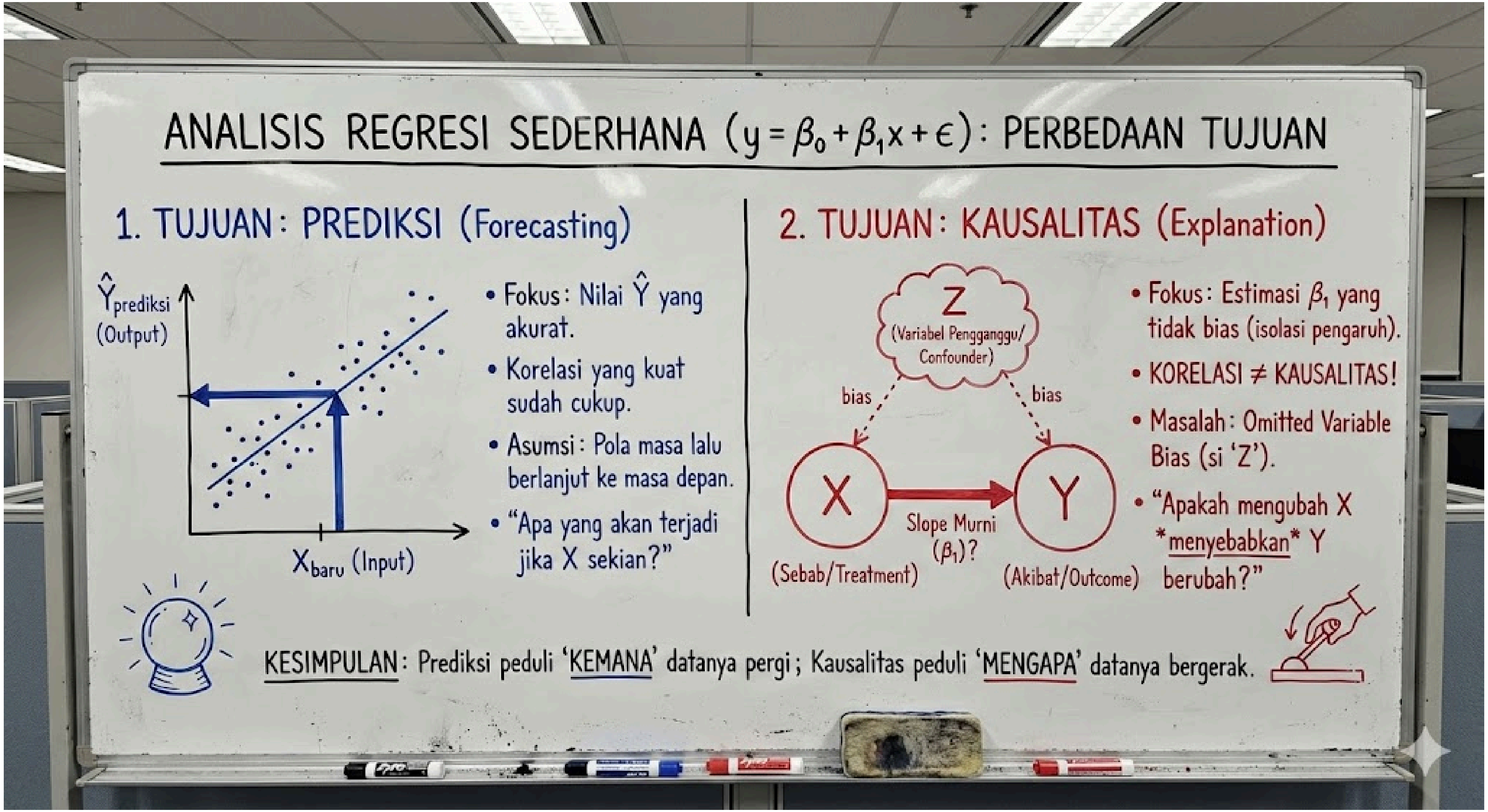
Model Regresi Linear Sederhana:

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

Keterangan:

- Y = Variabel dependen (variabel yang diprediksi)
- X = Variabel independen (variabel prediktor)
- β_0 = Intersep (konstanta)
- β_1 = Kemiringan (slope/koefisien regresi)
- ϵ = Error (galat)

Prediksi vs Kausalitas



Metode Kuadrat Terkecil (OLS)

Ordinary Least Squares (OLS):

Metode untuk menemukan garis regresi terbaik dengan **meminimalkan jumlah kuadrat residual**.

$$\text{Minimize: } \sum_{i=1}^{n} \left(Y_i - \hat{Y}_i\right)^2 = \sum_{i=1}^{n} e_i^2$$

Di mana:

- $e_i = Y_i - \hat{Y}_i$ adalah residual
- Residual = Selisih antara nilai aktual dan nilai prediksi

Persamaan Regresi Linear

Persamaan Regresi Sample:

$$\hat{Y} = b_0 + b_1 X$$

Di mana:

- $\hat{Y}$ = Nilai prediksi Y
- b_0 = Estimasi intersep
- b_1 = Estimasi slope/koefisien regresi

Rumus Koefisien:

$$b_1 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \frac{S_{XY}}{S_{XX}}$$

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

Interpretasi Koefisien Regresi

Intersep (b_0)

- Nilai Y ketika X = 0
- Titik potong garis dengan sumbu Y
- **Contoh:** Jika $b_0 = 50$, nilai ujian adalah 50 ketika jam belajar = 0

Slope (b_1)

- Perubahan Y untuk setiap perubahan 1 unit X
- **Contoh:** Jika $b_1 = 5$, setiap tambahan 1 jam belajar meningkatkan nilai ujian sebesar 5 poin

Evaluasi Model Regresi

Koefisien Determinasi (R^2)

R-squared (R^2):

$$R^2 = \frac{\mathrm{SSR}}{\mathrm{SST}} = 1 - \frac{\mathrm{SSE}}{\mathrm{SST}}$$

Di mana:

- **SST** (Total Sum of Squares) = $\sum \left(Y_i - \bar{Y}\right)^2$
- **SSR** (Regression Sum of Squares) = $\sum \left(\hat{Y}_i - \bar{Y}\right)^2$
- **SSE** (Error Sum of Squares) = $\sum \left(Y_i - \hat{Y}_i\right)^2$

Interpretasi:

- R^2 = Proporsi variansi Y yang dijelaskan oleh X
- Rentang: $0 \leq R^2 \leq 1$
- $R^2 = 0.75 \longrightarrow 75\%$ variansi Y dijelaskan oleh model

Standard Error of Estimate

Standard Error (S_e):

$$S_e = \sqrt{\frac{\sum \left(Y_i - \hat{Y}_i\right)^2}{n-2}} = \sqrt{\frac{\mathrm{SSE}}{n-2}}$$

Interpretasi:

- Mengukur rata-rata jarak data dari garis regresi
- Semakin kecil S_e, semakin baik model
- Satuan sama dengan variabel dependen (Y)

Uji Signifikansi Koefisien Regresi

Uji t untuk Slope (b_1):

Hipotesis:

- $H_0 : \beta_1 = 0$ (tidak ada hubungan linear)
- $H_1 : \beta_1 \neq 0$ (ada hubungan linear)

Statistik Uji:

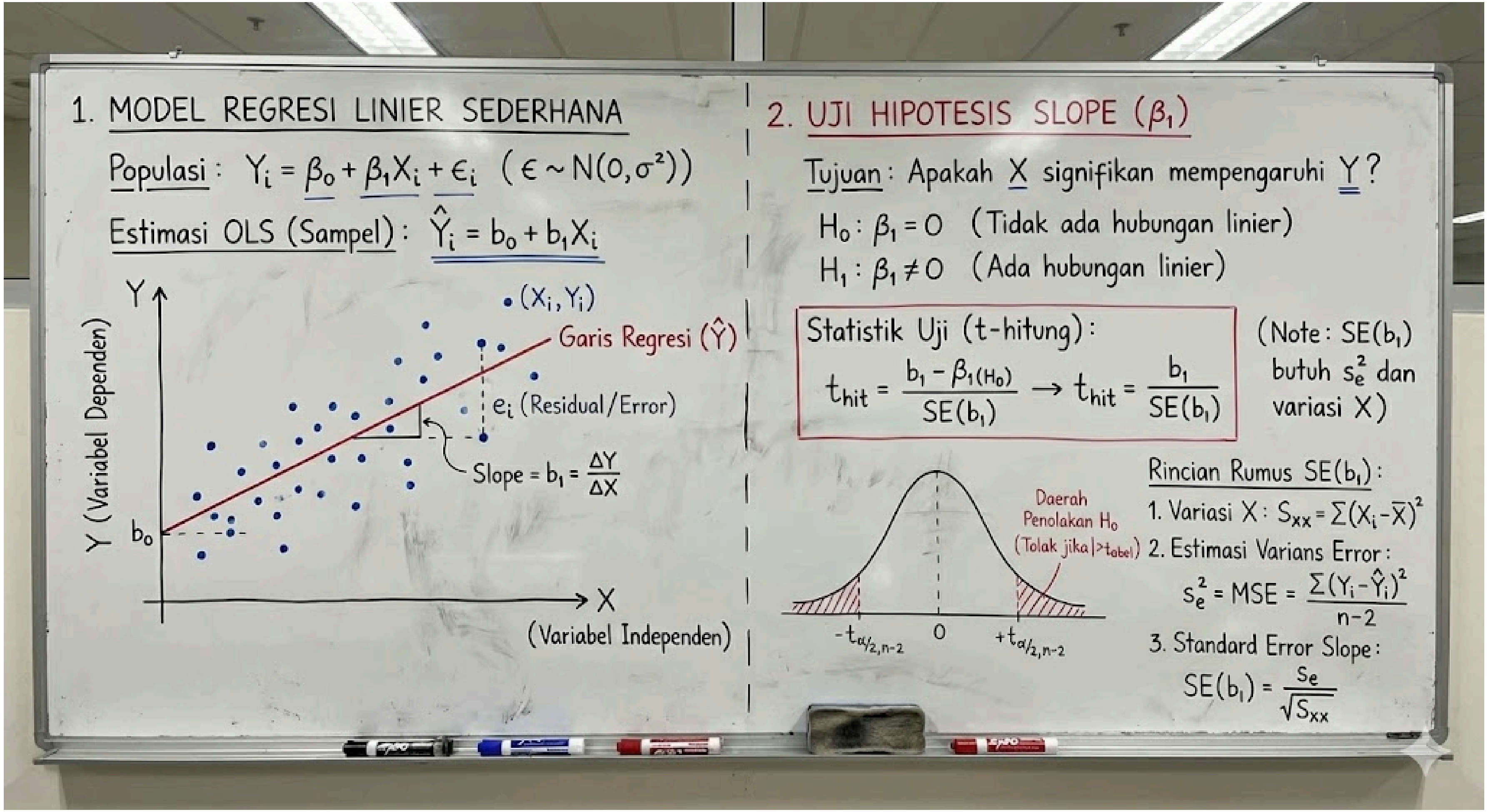
$$t = \frac{b_1}{SE_{b_1}} = \frac{b_1}{S_e/\sqrt{S_{XX}}}$$

dengan df = n - 2, di mana SE_{b_1} adalah standard error dari b_1. $S_{XX} = \sum \left(X_i - \bar{X}\right)^2$

Keputusan:

- Tolak H_0 jika $\left|t_{hitung}\right| > t_{\alpha/2,n-2}$ atau p-value < α

Uji Signifikansi Koefisien Regresi



Contoh Kasus Regresi Linear Sederhana

Contoh 1: Hubungan Jam Belajar dan Nilai Ujian

Data

Mahasiswa	Jam Belajar (X)	Nilai Ujian (Y)
1	2	60
2	3	68
3	4	74
4	5	82
5	6	88
6	7	92

Pertanyaan: Buatlah model regresi untuk memprediksi nilai ujian berdasarkan jam belajar!

Contoh 1: Perhitungan Koefisien

Langkah 1: Hitung rata-rata

$$\bar{X} = \frac{2+3+4+5+6+7}{6} = 4.5$$

$$\bar{Y} = \frac{60+68+74+82+88+92}{6} = 77.33$$

Contoh 1: Perhitungan Koefisien

Langkah 2: Hitung S_{XY} dan S_{XX}

X_i	Y_i	$X_i - \bar{X}$	$Y_i - \bar{Y}$	$(X_i - \bar{X})(Y_i - \bar{Y})$	$(X_i - \bar{X})^2$
2	60	−2.5	−17.33	43.33	6.25
3	68	−1.5	−9.33	14.00	2.25
4	74	−0.5	−3.33	1.67	0.25
5	82	0.5	4.67	2.33	0.25
6	88	1.5	10.67	16.00	2.25
7	92	2.5	14.67	36.67	6.25

$$S_{XX} = \sum (X_i - \bar{X})^2 = 17.5$$

$$S_{XY} = \sum (X_i - \bar{X})(Y_i - \bar{Y}) = 114$$

Contoh 1: Model Regresi

Langkah 3: Hitung koefisien

$$b_1 = \frac{S_{XY}}{S_{XX}} = \frac{114}{17.5} = 6.514$$

$$b_0 = \bar{Y} - b_1\bar{X} = 77.33 - 6.514(4.5) = 48.02$$

Model Regresi:

$$\hat{Y} = 48.02 + 6.514X$$

Interpretasi:

- Setiap tambahan 1 jam belajar meningkatkan nilai ujian sebesar 6.514 poin
- Jika tidak belajar (X=0), nilai prediksi adalah 48.02

Contoh 1: Prediksi

Prediksi Nilai untuk Jam Belajar Tertentu:

- **X = 8 jam:** $\hat{Y} = 48.02 + 6.514(8) = 100.13$
- **X = 10 jam:** $\hat{Y} = 48.02 + 6.514(10) = 113.16$

⚠ Peringatan Ekstrapolasi

Prediksi di luar rentang data (X < 2 atau X > 7) bisa tidak akurat karena kita tidak tahu apakah hubungan linear tetap berlaku.

Contoh 1: Evaluasi Model

Langkah 1: Hitung Nilai Prediksi dan Residual

X_i	Y_i	$\hat{Y}_i$	$Y_i - \bar{Y}$	$\hat{Y}_i - \bar{Y}$	$Y_i - \hat{Y}_i$	$\left(Y_i - \bar{Y}\right)^2$	$\left(\hat{Y}_i - \bar{Y}\right)^2$	$\left(Y_i - \hat{Y}_i\right)^2$
2	60	61.05	−17.33	−16.28	−1.05	300.33	265.04	1.10
3	68	67.56	−9.33	−9.77	0.44	87.05	95.45	0.19
4	74	74.08	−3.33	−3.25	−0.08	11.09	10.56	0.01
5	82	80.59	4.67	3.26	1.41	21.81	10.63	1.99
6	88	87.10	10.67	9.77	0.90	113.85	95.45	0.81
7	92	93.62	14.67	16.29	−1.62	215.21	265.36	2.62

Langkah 2: Hitung SST, SSR, dan SSE

$$\mathrm{SST} = \sum \left(Y_i - \bar{Y}\right)^2 = 749.34$$

$$\mathrm{SSR} = \sum \left(\hat{Y}_i - \bar{Y}\right)^2 = 742.49$$

Contoh 1: Evaluasi Model

$$\mathrm{SSE} = \sum \left(Y_i - \hat{Y}_i\right)^2 = 6.72$$

Langkah 3: Hitung R^2

$$R^2 = \frac{\mathrm{SSR}}{\mathrm{SST}} = \frac{742.49}{749.34} = 0.991$$

Interpretasi: Model menjelaskan 99.1% variansi nilai ujian. Ini menunjukkan hubungan yang sangat kuat antara jam belajar dan nilai ujian.

Contoh 1: Standard Error

Hitung Standard Error of Estimate:

$$S_e = \sqrt{\frac{\text{SSE}}{n-2}} = \sqrt{\frac{6.72}{6-2}} = \sqrt{1.68} = 1.30$$

Interpretasi: Rata-rata, nilai aktual berbeda sekitar 1.30 poin dari nilai prediksi model.

Contoh 1: Uji Hipotesis

Uji Signifikansi Koefisien Regresi (Uji-t untuk b_1)

Langkah 1: Tentukan Hipotesis

- $H_0 : \beta_1 = 0$ (jam belajar tidak berpengaruh terhadap nilai ujian)
- $H_1 : \beta_1 \neq 0$ (jam belajar berpengaruh terhadap nilai ujian)

Langkah 2: Tentukan Tingkat Signifikansi

$$\alpha = 0.05$$

Langkah 3: Hitung Statistik Uji

$$SE_{b_1} = \frac{S_e}{\sqrt{S_{XX}}} = \frac{1.30}{\sqrt{17.5}} = \frac{1.30}{4.183} = 0.311$$

Contoh 1: Uji Hipotesis

$$t_{hitung} = \frac{b_1}{SE_{b_1}} = \frac{6.514}{0.311} = 20.95$$

Langkah 4: Tentukan Nilai Kritis

Dengan $df = n - 2 = 6 - 2 = 4$ dan $\alpha/2 = 0.025$:

$$t_{\alpha/2,4} = t_{0.025,4} = 2.776$$

Langkah 5: Keputusan

Karena $\left|t_{hitung}\right| = 20.95 > t_{0.025,4} = 2.776$, maka **tolak $\boldsymbol{H_0}$**.

Kesimpulan: Terdapat cukup bukti pada tingkat signifikansi 5% untuk menyimpulkan bahwa jam belajar berpengaruh signifikan terhadap nilai ujian. Dengan kata lain, koefisien regresi (β_1) berbeda secara signifikan dari nol.

Contoh 1: Uji Hipotesis

Interval Kepercayaan 95% untuk $\boldsymbol{\beta_1}$:

$$b_1 \pm t_{\alpha/2,n-2} \times SE_{b_1}$$

$$6.514 \pm 2.776 \times 0.311$$

$$6.514 \pm 0.863$$

$$[5.651, 7.377]$$

Interpretasi: Kita 95% yakin bahwa setiap tambahan 1 jam belajar akan meningkatkan nilai ujian antara 5.65 hingga 7.38 poin.

Latihan Soal Regresi Linear Sederhana

Latihan 1: Hubungan Pengeluaran Iklan dan Penjualan

Data

Sebuah perusahaan ingin mengetahui hubungan antara pengeluaran iklan (dalam juta rupiah) dengan penjualan produk (dalam unit). Data yang dikumpulkan selama 8 minggu adalah sebagai berikut:

Minggu	Pengeluaran Iklan (X)	Penjualan (Y)
1	5	120
2	7	145
3	10	180
4	8	160
5	12	195
6	15	230
7	9	165
8	11	185

Latihan 1: Hubungan Pengeluaran Iklan dan Penjualan

Pertanyaan:

a. Buatlah model regresi linear sederhana untuk memprediksi penjualan berdasarkan pengeluaran iklan!
b. Hitung koefisien determinasi (R^2) dan interpretasikan hasilnya!
c. Uji apakah pengeluaran iklan berpengaruh signifikan terhadap penjualan pada tingkat signifikansi 5%!
d. Jika perusahaan mengalokasikan 13 juta rupiah untuk iklan, berapa prediksi penjualannya?
e. Buatlah interval kepercayaan 95% untuk koefisien regresi (β_1)!

Latihan 2: Hubungan Suhu dan Konsumsi Listrik

Data

Sebuah apartemen ingin menganalisis hubungan antara suhu udara (°C) dengan konsumsi listrik harian (kWh). Data yang dikumpulkan selama 7 hari adalah:

Hari	Suhu (X)	Konsumsi Listrik (Y)
1	28	85
2	30	95
3	32	102
4	35	115
5	33	108
6	31	98
7	29	88

Latihan 2: Hubungan Suhu dan Konsumsi Listrik

Pertanyaan:

a. Buatlah model regresi linear sederhana untuk memprediksi konsumsi listrik berdasarkan suhu!
b. Interpretasikan nilai intersep (b_0) dan slope (b_1) dalam konteks masalah!
c. Hitung standard error of estimate (S_e) dan jelaskan artinya!
d. Uji apakah suhu berpengaruh signifikan terhadap konsumsi listrik pada tingkat signifikansi 1%!
e. Jika suhu mencapai 36°C, berapa prediksi konsumsi listrik?