BAB IX

ANALISIS REGRESI

1. Model Analisis Regresi-Linear

Analisis regresi-linear adalah metode statistic yang dapat digunakan untuk mempelajari hubungan antarsifat permasalahan yang sedang diselidiki. Model analisis regresi-linear dapat memodelkan hubungan antara 2 (dua) peubah atau lebih. Pada model ini terdapat peubah tidak bebas (y) yang mempunyai hubungan fungsional dengan satu atau lebih peubah bebas (x_i).

Dalam kasus yang paling sederhana, hubungan secara umum dapat dinyatakan dalam persamaan (1) berikut.

$$Y = A + BX \tag{1}$$

Y = peubah tidak bebas

X = peubah bebas

A = intersep atau konstanta regresi

B = koefisien regresi

Parameter A dan B dapat diperkirakan dengan menggunakan metode kuadrat terkecil yang meminimkan total kuadratis residual antara hasil model dengan hasil pengamatan. Nilai parameter A dan B bisa didapatkan dari persamaan (2)-(3) berikut.

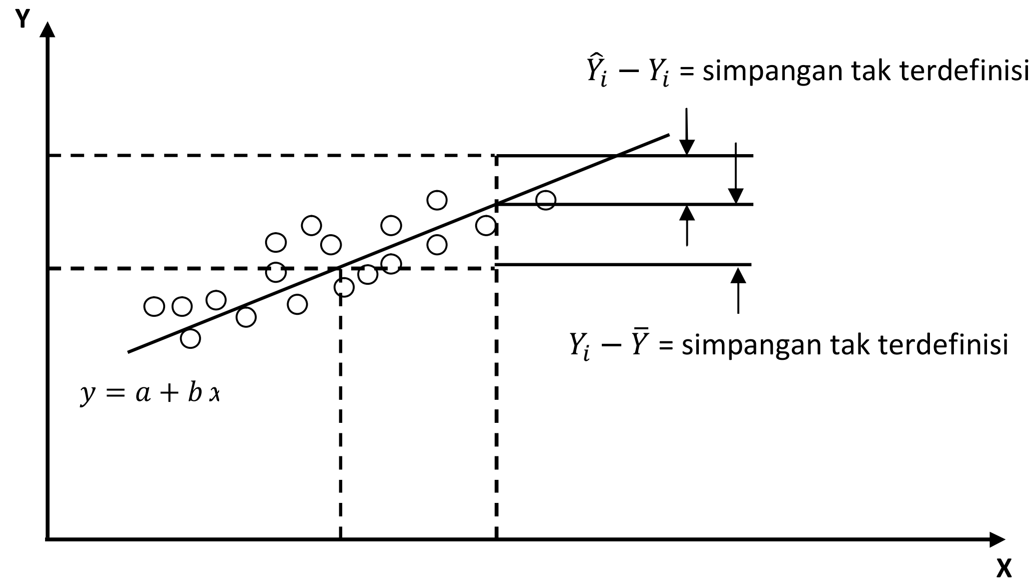
$$B = \frac{N\sum_{i=1}^{N}(X_iY_i) - \sum_{i=1}^{N}(X_i) \, . \sum_{i=1}^{N}(Y_i)}{N\sum_{i=1}^{N}(X_i^2) - \left(\sum_{i=1}^{N}(X_i)\right)^2} \tag{2}$$

$$A = \bar{Y} - B\bar{X} \tag{3}$$

$\bar{Y}$ dan $\bar{X}$ adalah nilai rata-rata dari Y_i dan X_i.

2. Koefisien Determinasi (R^2)

Gambar 1 memperlihatkan garis regresi dan beberapa data yang digunakan untuk mendapatkannya. Jika tidak terdapat nilai x, ramalan terbaik Y_i adalah $\bar{Y}$. Akan tetapi, gambar memperlihatkan bahwa untuk x_i, galat metode tersebut akan tinggi: $(\hat{Y}_i - \bar{Y})$. Jika x_i diketahui, ternyata ramalan terbaik Y_i dan hal ini memperkecil galat menjadi $\left(\hat{Y}_i - Y_i\right)$. $\hat{Y}_i = nilai\ hasil\ pemodelan$



Gambar 1 Beberapa jenis simpangan

Dari Gambar 1, kita dapatkan:

$$\underset{\text{Simpangan total}}{\left(\hat{Y}_i - \bar{Y}\right)} = \underset{\text{Simpangan terdefinisi}}{\left(Y_i - \bar{Y}\right)} + \underset{\text{Simpangan tidak terdefinisi}}{\left(\hat{Y}_i - Y_i\right)} \qquad (4)$$

Jika kita kuadratkan total simpangan tersebut dan menjumlahkan semua nilai i, didapat:

$$\underset{\text{Simpangan total}}{\sum_{i=1}^{N}\left(\hat{Y}_i - \bar{Y}\right)^2} = \underset{\text{Simpangan terdefinisi}}{\sum_{i=1}^{N}\left(Y_i - \bar{Y}\right)^2} + \underset{\text{Simpangan tidak terdefinisi}}{\sum_{i=1}^{N}\left(\hat{Y}_i - Y_i\right)^2} \qquad (5)$$

$\left(\hat{Y}_i - \bar{Y}\right) = b\,x_i$ sehingga mudah dilihat bahwa variasi terdefinisi merupakan fungsi koefisien regresi b. Proses penggabungan total variasi disebut analisis variasi.

Koefisien determinasi didefinisikan sebagai nisbah antara variasi tidak terdefinisi dengan variasi total:

$$R^2 = 1 - \frac{\sum_{i=1}^{N}\left(\hat{Y}_i - Y_i\right)^2}{\sum_{i=1}^{N}\left(\hat{Y}_i - \bar{Y}\right)^2} \qquad (6)$$

Koefisien ini mempunyai batas limit sama dengan 1 (satu) (*perfect explanation*) dan 0 (nol) (*no explanation*). Nilai antara kedua batas limit ini ditafsirkan sebagai persentase total variasi yang dijelaskan oleh analisis regresi-linear.

3. Model Analisis Regresi-Linear-Berganda

Konsep ini merupakan pengembangan lanjut dari uraian di atas, khususnya pada kasus yang mempunyai lebih banyak peubah bebas dan parameter b. hal ini sangat diperlukan dalam realita yang menunjukkan bahwa beberapa peubah tata guna lahan secara simultan ternyata mempengaruhi bangkitan pergerakan.

Persamaan (7) memperlihatkan bentuk umum metode analisis regresi-linear-berganda.

$$Y = A + B_1X_1 + B_2X_2 + \cdots + B_ZX_Z \qquad (7)$$

Y = peubah tidak bebas
$X_1 .. X_z$ = peubah bebas
A = konstanta
$B_1 \ldots B_z$ = koefisien regresi

Analisis regresi-linear-berganda adalah suatu metode statistik. Untuk menggunakannya, terdapat beberapa asumsi yag perlu diperhatikan:

- Nilai peubah, khususnya peubah bebas, mempunyai nilai tertentu atau merupakan nilai yang didapat dari hasil survey tanpa kesalahan berarti;
- Peubah tidak bebas (*Y*) harus mempunyai hubungan korelasi linear dengan peubah bebas (X). jika hubungan tersebut tidak linear, transformasi linear harus dilakukan, meskipun batasan ini akan mempunyai implikasi lain dalam analisis residual;
- Efek peubah bebas pada peubah tidak bebas merupakan penjumlahan, dan harus tidak ada korelasi yang kuat antara sesama peubah bebas;
- Variansi peubah tidak bebas terhadap garis regresi harus sama untuk semua nilai peubah bebas;
- Nilai peubah tidak bebas harus tersebar normal atau minimal mendekati normal;
- Nilai peubah bebas sebaiknya merupakan besaran yang relative mudah diproyeksi.

Solusinya tetap sama, tetapi lebih kompleks sehingga beberapa hal baru harus dipertimbangkan, seperti:

a. **Multikolinear**. Hal ini terjadi karena adanya hubungan linear antarpeubah. Pada kasus ini, beberapa persamaan yang mengandung b tidak saling bebas dan tidak dapat dipecahkan secara unik.

b. **Parameter 'b' yang dibutuhkan**. Untuk memutuskan hal ini, beberapa factor harus dipertimbangkan:

- Apakah ada alasan teori yang kuat sehingga harus melibatkan peubah itu atau apakah peubah itu penting untuk proses uji dengan model tersebut?
- Apakah peubah itu signifikan dan apakah tanda koefisien parameter yang didapat sesuai dengan teori atau intuisi?

Jika diragukan, tetapkan salah satu cara, yaitu menghilangkan peubah itu dan melakukan proses regresi lagi untuk melihat efek dibuangnya peubah itu terhadap peubah lainnya yang masih digunakan oleh model tersebut. Jika ternyata tidak terlalu terpengaruh, peubah itu dibuang saja sehingga kita mendapatkan model yang lebih sederhana dan dapat ditaksir secara lebih tepat.

Beberapa paket program telah menyediakan prosedur otomatis untuk menangani masalah ini (pendekatan langkah-demi-langkah atau *stepwise*). Akan tetapi, pendekatan ini masih mempunyai beberapa permasalahan.

c. **Koefisien determinasi.** Bentuknya sama dengan persamaan (6). Akan tetapi, pada kasus ini, tambahan peubah b biasanya meningkatkan nilai $\mathbf{R^2}$; untuk mengatasinya digunakan nilai $\mathbf{R^2}$ yang telah dikoreksi:

$$\overline{R^2} = \frac{\left[R^2 - \frac{K}{N-1}\right]}{\left[\frac{(N-1)}{(N-K-1)}\right]} \qquad (8)$$

N adalah ukuran sampel dan K adalah jumlah peubah b.

d. Koefisien korelasi. Koefisien korelasi ini digunakan untuk menentukan korelasi antara peubah tidak bebas dengan peubah bebas atau antara sesama peubah bebas. Koefisien korelasi ini dapat dihitung dengan berbagai cara yang salah satunya adalah (9) berikut.

$$r = \frac{N\sum_{i=1}^{N}(X_iY_i) - \sum_{i=1}^{N}(X_i)\,.\sum_{i=1}^{N}(Y_i)}{\sqrt{\left[N\sum_{i=1}^{N}(X_i^2) - \left(\sum_{i=1}^{N}(X_i)\right)^2\right]\left[N\sum_{i=1}^{N}(Y_i^2) - \left(\sum_{i=1}^{N}(Y_i)\right)^2\right]}} \quad (9)$$

Nilai r = 1 berarti bahwa korelasi antara peubah y dan x adalah positif (meningkatkannya nilai x akan mengakibatkan meningkatnya nilai y). sebaliknya, jika r = -1, berarti korelasi antara peubah y dan x adalah negative (meningkatkannya nilai x akan mengakibatkan menurunnya nilai y). Nilai r = 0 menyatakan tidak ada korelasi antarpeubah.

e. **Uji t-tes**. Uji t-tes dapat digunakan untuk 2 (dua) tujuan, yaitu untuk menguji signifikansi nilai koefisisen korelasi (r) dan untuk menguji signifikansi nilai koefisien regresi. Setiap peubah yang mempunyai koefisien regresi yang tidak signifikan secara statistic harus dibuang dari model.

3.1 Analisis dengan dua peubah bebas. Persamaan dengan 2 (dua) peubah bebas dapat dinyatakan dengan persamaan (10).

$$Y = b_0 + b_1X_1 + b_2X_2$$ di mana: (10)

Y = bangkitan atau tarikan (peubah tidak bebas)

X_1, X_2 = peubah bebas

b_0 = konstanta

b_1, b_2 = koefisien regresi

Nilai b_0, b_1, b_2 dapat dihitung dengan menggunakan analisis regresi-linear-berganda. Nilai b_0, b_1, b_2 bisa didapat dengan menyelesaikan 3 (tiga) buah persamaan linear simultan (11)-(13) berikut ini.

$$Nb_0 + b_1\sum_{i=1}^{N} X_{1i} + b_2\sum_{i=1}^{N} X_{2i} = \sum_{i=1}^{N} Y_i \quad (11)$$

$$b_0\sum_{i=1}^{N} X_{1i} + b_1\sum_{i=1}^{N}(X_{1i})^2 + b_2\sum_{i=1}^{N}(X_{1i}\,.X_{2i}) = \sum_{i=1}^{N}(X_{1i}\,.Y_i) \quad (12)$$

$$b_0\sum_{i=1}^{N} X_{2i} + b_1\sum_{i=1}^{N}(X_{1i}\,.X_{2i}) + b_2\sum_{i=1}^{N}(X_{2i})^2 = \sum_{i=1}^{N}(X_{2i}.Y_i) \quad (13)$$

Teknik eliminasi matriks *Gauss-Jordan* dapat digunakan untuk memecahkan permasalahan persamaan simultan (11)-(13).

3.2 Analisis dengan tiga peubah bebas. Persamaan dengan 3 (tiga) peubah bebas dapat dinyatakan dengan persamaan (14).

$$Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 \qquad \text{di mana:} \tag{14}$$

$Y =$ peubah tidak bebas

$X_1, X_2, X_3 =$ peubah bebas

$b_0 =$ konstanta

$b_1, b_2, b_3 =$ koefisien regresi

Nilai b_0, b_1, b_2, b_3 dapat dihitung dengan menggunakan analisis regresi-linear- berganda. Nilai b_0, b_1, b_2, b_3 bisa didapat dengan menyelesaikan 4 (empat) buah persamaan linear simultan (15)-(18).

$$Nb_0 + b_1\sum_{i=1}^{N} X_{1i} + b_2\sum_{i=1}^{N} X_{2i} + b_3\sum_{i=1}^{N} X_{3i} = \sum_{i=1}^{N} Y_i \tag{15}$$

$$b_0\sum_{i=1}^{N} X_{1i} + b_1\sum_{i=1}^{N}(X_{1i})^2 + b_2\sum_{i=1}^{N}(X_{1i}\,.\,X_{2i}) + b_3\sum_{i=1}^{N}(X_{1i}\,.\,X_{3i}) = \sum_{i=1}^{N}(X_{1i}\,.\,Y_i) \tag{16}$$

$$b_0\sum_{i=1}^{N} X_{2i} + b_1\sum_{i=1}^{N}(X_{1i}\,.\,X_{2i}) + b_2\sum_{i=1}^{N}(X_{2i})^2 + b_3\sum_{i=1}^{N}(X_{2i}.\,X_{3i}) = \sum_{i=1}^{N}(X_{2i}.\,Y_i) \tag{17}$$

$$b_0\sum_{i=1}^{N} X_{3i} + b_1\sum_{i=1}^{N}(X_{1i}\,.\,X_{3i}) + b_2\sum_{i=1}^{N}(X_{2i}.\,X_{3i}) + b_3\sum_{i=1}^{N}(X_{3i})^2 = \sum_{i=1}^{N}(X_{3i}\,.\,Y_i) \tag{18}$$

Teknik eliminasi matriks *Gauss-Jordan* dapat digunakan untuk memecahkan permasalahan persamaan simultan (15)-(18).

4. Tahapan Uji Statistik dalam Model Analisis-Regresi

Dalam melakukan analisis-regresi, terdapat 4 (empat) tahap uji statistic yang mutlak harus dilakukan agar model yang dihasilkan dinyatakan abash. Ke-4 uji statistic tersebut akan diterangkan satu per satu sebagai berikut.

4.1 Uji Kecukupan Data

Uji statistic ini dilakukan untuk menentukan jumlah data minimum yang harus tersedia, baik untuk peubah bebas maupun peubah tidak bebas. Semakin tinggi tingkat akurasi yang diinginkan, semakin banyak data yang dibutuhkan. Jumlah data minimum dapat ditentukan dengan menggunakan persamaan (19) berikut:

$$N = \frac{CV^{\,2} Z_\alpha^2}{E^2} \tag{19}$$

CV = koefisien variasi

E = tingkat akurasi

Z_α = nilai variansi untuk tingkat kepercayaan α yang diinginkan.

Sebagai ilustrasi, diberikan contoh berikut ini. Tentukan berapa jumlah data minimum yang dibutuhkan untuk tingkat akurasi (E) 5% dengan tingkat kepercayaan (α) 95%. Untuk α = 95%, maka nilai Z_α adalah 1,645. Dengan mengasumsikan nilai CV = 1,0 didapatkan:

$$N = \frac{(1{,}0)(1{,}645)^2}{(0{,}05)^2} = 1082$$

Jadi, dibutuhkan jumlah data minimum sebanyak 1082 buah untuk tingkat akurasi 5% dengan tingkat kepercayaan 95%.

4.2 Uji Korelasi

Uji statistic ini harus dilakukan untuk memenuhi persyarakat model matematis: sesama peubah bebas tidak boleh saling berkorelasi, sedangkan antara peubah tidak bebas dengan peubah bebas harus ada korelasi yang kuat (baik positif maupun negatif).

Persamaan (9) merupakan persamaan uji korelasi yang mempunyai nilai r ($-1 \leq r \leq +1$). Nilai r yang mendekati -1 mempunyai arti bahwa kedua peubah tersebut saling berkorelasi negative (peningkatan nilai salah satu peubah akan menyebabkan penurunan nilai peubah lainnya).

Sebaliknya, jika nilai r yang mendekati +1 mempunyai arti bahwa kedua peubah tersebut saling berkorelasi positif (peningkatan nilai salah satu peubah akan menyebabkan peningkatan nilai peubah lainnya). Jika nilai r mendekati 0, tidak terdapat korelasi antara kedua peubah tersebut.

4.3 Uji Linearitas

Uji statistic ini perlu dilakukan untuk memastikan apakah model dapat didekati dengan model analisis-regresi-linear atau model analisis-regresi-tidak-linear.

4.4 Uji Kesesuaian

Uji statistic ini harus dilakukan untuk menentukan model terbaik. Terdapat beberapa model yang dapat digunakan, yaitu model analisis-regresi, model kemiripan-maksimum, dan model entropi-maksimum. Pada umumnya, uji ini didasarkan atas kedekatan atau kesesuaian hasil model dengan hasil observasi.

Dua uji kesesuaian yang paling sering digunakan adalah (a) model analisis-regresi dan (b) model kemiripan-maksimum. Model analisis-regresi mengasumsikan bahwa model terbaik adalah model yang mempunyai total kuadratis residual antara hasil model dengan hasil pengamatan (observasi) yang paling minimum.

$$\text{Meminimumkan } S = \sum_{i=1}^{N}\left(Y_i - \widehat{Y}_i\right)^2 \qquad (1.84)$$

Model kemiripan-maksimum mengasumsikan bahwa model terbaik adalah model yang mempunyai total perkalian peluang antara hasil model dengan hasil pengamatan (observasi) yang paling maskimum (mendekati 1).

$$\text{Memaksimumkan } L = \prod_{i=1}^{N}\left(\frac{Y_i}{\widehat{Y}_i}\right) \qquad (1.85)$$

5. Indikator Uji Kesesuaian Matriks

5.1 Root Mean Square Error (RMSE) dan Standar Deviasi (SD)

Indicator uji kesesuaian RMSE adalah suatu indicator kesalahan yang didasarkan pada total kuadratis dari simpangan antara hasil model dengan hasil observasi yang dapat didefinisikan sebagai persamaan (1.86):

$$RMSE = \sqrt{\sum_{i=1}^{N}\sum_{d=1}^{N}\left[\frac{\left(\widehat{T}_{id} - T_{id}\right)^2}{N.(N-1)}\right]} \qquad \text{untuk } i \neq d \qquad (1.86)$$

$N =$ jumlah baris atau kolom matriks

$T_{id}\,\widehat{T}_{id} =$ nilai sel matriks hasil model dan hasil observasi beberapa penelitian menggunakan standar deviasi dari simpanagan yang dapat didefinisikan sebagai persamaan (1.87):

$$SD = \sqrt{\sum_{i=1}^{N}\sum_{d=1}^{N}\left[\frac{\left(\hat{T}_{id} - T_{id}\right)^2}{N.(N-1)-1}\right]} \qquad \text{untuk } i \neq d \qquad (1.87)$$

Dari persamaan (1.86)-(1.87) terlihat bahwa semakin besar nilai N maka nilai RMSE kira-kira akan sama dengan nilai SD. Indicator %RMSE digunakan untuk membandingkan 2 buah MAT yang mempunyai jumlah sel yang berbeda.

$$\%\mathrm{RMSE} = \left(\frac{\mathrm{RMSE}}{\mathrm{T}_1}\right) \mathrm{x} 100\% \qquad (1.88)$$

$$\mathrm{T}_1 = \frac{1}{\mathrm{N}(\mathrm{N}-1)}\sum_{i=1}^{N}\sum_{d=1}^{N}\hat{T}_{id} \qquad (1.89)$$

Semakin besar nilai RMSE, %RMSE, dan SD maka semakin tidak akurat MAT hasil penaksiran dibandingkan MAT hasil pengamatan.

5.2 Mean Absolute Error (MAE)

MAE adalah bentuk ukuran simpangan paling sederhana yang dapat didefinisikan sebagai persamaan (1.90).

$$MAE = \sum_{i=1}^{N}\sum_{d=1}^{N}\left|\frac{\left(\hat{T}_{id} - T_{id}\right)^2}{N.(N-1)}\right| \qquad \text{untuk } i \neq d \qquad (1.90)$$

Dari persamaan (1.90) terlihat bahwa nilai MAE kurang sensitive terhadap nilai mutlak kesalahan yang besar dibandingkan dengan RMSE. Semakin besar nilai MAE maka semakin tidak akurat MAT hasil penaksiran dibandingkan MAT hasil pengamatan.

5.3 Koefisien Determinasi (R^2)

Indikator uji kesesuaian R^2 dapat didefinisikan sebagai persamaan (1.91):

$$\mathrm{R}^2 = 1 - \frac{\sum_{i=1}^{N}\sum_{d=1}^{N}\left(\hat{T}_{id} - T_{id}\right)^2}{\sum_{i=1}^{N}\sum_{d=1}^{N}\left(\hat{T}_{id} - T_1\right)^2} \qquad \text{untuk } i \neq d \qquad (1.91)$$

Indicator uji statistic R^2 ini merupakan suatu uji statistic yang paling sering digunakan. Indicator ini akan memberikan bobot sangat tinggi untuk kesalahan absolut besar.

Oleh karena itu, nilai R^2 yang tinggi tidak dapat diperoleh dari matriks berjumlah sel besar dengan kesalahan kecil, tetapi sangat jelek pada nilai sel yang kecil.

5.4 Normalized Mean Absolute Error (NMAE)

Beberapa indicator uji kesesuaian yang telah diuraikan di atas seperti RMSE, SD, %RMSE, MAE, dan R^2 tidak dapat digunakan untuk membandingkan MAT jika diterapkan pada daerah kajian yang berbeda, karena nilai MAT sangat tergantung pada kondisi local seperti ukuran matriks dan lainnya.

Untuk tujuan ini, disarankan menggunakan indicator uji statistic NMAE yang didefinisikan sebagai persamaan (1.92).

$$NMAE = \left(\frac{MAE}{T_1} \right) x\,100 \qquad (1.92)$$

6. Kumpulan Soal

1. Asumsi terdapat satu persamaan dengan 2 (dua) buah peubah bebas sebagaimana tertera pada persamaan (1.93) berikut.

$$Y = b_0 + b_1X_1 + b_2X_2 \qquad \text{di mana:} \qquad (1.93)$$

Y = bangkitan atau tarikan (peubah tidak bebas)

X_1, X_2 = peubah bebas

b_0 = konstanta

b_1, b_2 = koefisien regresi

Jika diketahui beberapa data nilai peubah bebas (Y) dan peubah tidak bebas (X_1) dan (X_2), saudara diminta untuk menentukan nilai konstanta (b_0) dan koefisien regresi (b_1, b_2) dari persamaan (1.93)?

No	Y	X_1	X_2	No	Y	X_1	X_2
1	200	40	50	6	110	50	60
2	420	30	100	7	340	60	80
3	620	30	20	8	260	70	50

4	820	70	60	9	480	80	40
5	920	80	90	10	900	40	70

2. Asumsi terdapat satu persamaan dengan 3 (tiga) buah peubah bebas sebagaimana tertera pada persamaan (1.94) berikut.

$$Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 \qquad \text{di mana:} \qquad (1.94)$$

$Y =$ bangkitan atau tarikan (peubah tidak bebas)

$X_1, X_2, X_3 =$ peubah bebas

$b_0 =$ konstanta

$b_1, b_2, b_3 =$ koefisien regresi

Jika diketahui beberapa data nilai peubah bebas (Y) dan peubah tidak bebas (X_1), (X_2), dan (X_3), saudara diminta untuk menentukan nilai konstanta (b_0) dan koefisien regresi (b_1, b_2, b_3) dari persamaan (1.94)?

No	Y	X_1	X_2	X_3	No	Y	X_1	X_2	X_3
1	200	40	50	50	6	110	50	60	50
2	420	30	100	100	7	340	60	80	100
3	620	30	20	20	8	260	70	50	20
4	820	70	60	60	9	480	80	40	60
5	920	80	90	90	10	900	40	70	90

3. Saudara diminta untuk menghitung nilai indicator uji kesesuaian matriks RMSE, SD, %RMSE, MAE, R^2, dan NMAE antara 2 (dua) buah matriks berikut ini.

Tabel 1.2: Matriks I *(T_{id})*

No	1	2	3	4	5	6	7	8
1	20	40	50	60	80	50	60	80
2	40	30	100	50	80	100	50	80
3	60	30	20	90	150	20	90	150
4	80	70	60	40	200	60	40	200
5	100	80	90	80	50	90	80	50
6	60	30	20	90	150	20	90	150

7	80	70	60	40	200	60	40	200
8	100	80	90	80	50	90	80	50

Tabel 1.3: Matriks II *(T_{id})*

No	1	2	3	4	5	6	7	8
1	22	41	54	64	82	54	62	81
2	41	32	103	54	83	104	52	82
3	63	33	22	92	155	23	93	153
4	79	74	61	43	201	63	44	204
5	98	85	89	81	49	92	85	47
6	64	25	19	85	148	19	89	147
7	75	68	58	38	197	57	38	198
8	95	77	85	76	46	88	77	48