Regresi Linear Berganda

Pertemuan 4

Data terdiri dari n amatan pada peubah respon Y dan k peubah penjelas, $X_1, X_2, \dots, X_k$. Hubungan antara Y dan $X_1, X_2, \dots, X_k$ diformulasikan sebagai model linear sebagai berikut.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon$$

Dimana $\beta_0, \beta_1, \beta_2, \dots, \beta_p$ adalah parameter regresi yang bersifat tetap dan ε adalah peubah acak dari sisaan/galat. Peubah respon Y adalah fungsi linear dari X dan ε memuat informasi untuk menjelaskan Y yang tidak dapat dijelaskan oleh X

Nomor Observasi	Peubah Respon Y	Peubah Penjelas			
		X_1	X_2	$\dots$	X_k
1	y_1	x_{11}	x_{12}	$\dots$	x_{1k}
2	y_2	x_{21}	x_{22}	$\dots$	x_{2k}
3	y_3	x_{31}	x_{32}	$\dots$	x_{3k}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	y_n	x_{n1}	x_{n2}	$\dots$	x_{nk}

$$y_i = \beta_0 + \beta_1 x i_1 + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i$$
$$i = 1,2,3, \dots, n$$

y_i menunjukkan nilai rataan respon ke-i dari peubah respon Y, $x_{i1}, x_{i2}, \dots, x_{ik}$ menunjukkan nilai peubah penjelas unit ke-I dan ε_i menunjukkan galat/sisaan

Model regresi linear berganda dalam notasi matriks dapat dituliskan sebagai

$$Y = X\beta + \varepsilon$$
$$Y_{n\times 1} = X_{n\times(k+1)}\beta_{(k+1)\times 1} + \varepsilon_{n\times 1}$$

$$y_{n\times 1} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, x_{n\times(k+1)} = \begin{bmatrix} 1 & X_{11} & \dots & X_{k1} \\ 1 & X_{12} & \dots & X_{k2} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & X_{1n} & \dots & X_{kn} \end{bmatrix}, \beta_{(k+1)\times n} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix},$$

$$\varepsilon_{n\times 1} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Asumsi Regresi Linear Berganda

1. Kondisi Gauss-Marcov

- $E[\varepsilon_i] = 0$, nilai harapan/rataan galat sama dengan nol
- $E\left[\varepsilon_i^2\right] = var[\varepsilon] = \sigma^2 I$, ragam sisaan homogen untuk setiap x (homoskedastisitas)
- $E\left[\varepsilon_i \varepsilon_j\right] = 0$, sisaan saling bebas/tidak ada autokorelasi

1. Galat menyebar normal
2. Galat bebas terhadap peubah bebas, $cov(x_i, \varepsilon_i) = 0$
3. Tidak ada multikolineritas pada peubah bebas

Pendugaan Parameter (MKT)

$$\sum_{i=1}^{n} e_i^2 = e'e = (y - Xb)'(y - Xb)$$

$$e'e = y'y - y'Xb - b'X'y + b'X'Xb$$

Karena $b'X'y$ adalah matriks berukuran 1x1 maka $b'X'y = (b'X'y)' = y'Xb$, sehingga

$$e'e = y'y - 2y'Xb + b'(X'X)b$$

$$= y'y - 2b'X'y + b'(X'X)b$$

Pendugaan Parameter (MKT)

Penduga parameter b didapatkan menggunakan teknik turunan jumlah kuadrat galat terhadap parameter dengan menyamadengankan solusi menjadi nol.

$$\frac{\partial e'e}{\partial b} = -2X'y + 2(X'X)b = 0$$

$$(X'X)b = X'y$$

Sehingga penduga bagi parameter regresi berganda

$$b_{(k+1)\times 1} = (X'X)^{-1}_{(k+1)\times(k+1)} X'_{(k+1)\times n} y_{n\times 1}$$

Vektor nilai dugaan $\hat{\boldsymbol{y}}$ $\quad \hat{\boldsymbol{y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{H}\boldsymbol{y}$ $\quad \boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ adalah matriks hat atau matriks proyeksi

Vektor sisaan dapat dituliskan sebagai $\boxed{\boldsymbol{e} = \boldsymbol{y} - \hat{\boldsymbol{y}}}$ $\boxed{\hat{\boldsymbol{y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{H}\boldsymbol{y}}$ $\Rightarrow$ $\boldsymbol{e} = \boldsymbol{y} - \boldsymbol{H}\boldsymbol{y}$, $\boldsymbol{e} = (\boldsymbol{I}_n - \boldsymbol{H})\boldsymbol{y}$

Sifat Sifat penduga MKT

1. $\hat{\beta}$ merupakan penduga tak bias bagi β
2. $Var(\hat{\beta}) = \sigma^2 (X'X)^{-1}$, misalkan $C = (X'X)^{-1}$ maka $Var(\hat{\beta}) = \sigma^2 C$
3. Penduga $\hat{\beta}_j$ menyebar normal dengan nilai tengah β_j dan ragam $\sigma^2 c_{jj}$
4. Dugaan bagi ragam y atau ragam $\varepsilon(\sigma^2)$

$$\hat{\sigma}^2 = \frac{SSResidual}{n-p} = \frac{y'y - \hat{\beta}x'y}{n-p}$$

Interpretasi Parameter

β_0 : Intersep (Nilai rataan Y ketika $X_1 = X_2 = \cdots = X_k = 0$)

Koefisien regresi β_j

Dugaan perubahan rataan nilai y ketika X_j berubah satu satuan dan peubah penjelas lainnya dianggap tetap

Selang kepercayaan parameter

Selang kepercayaan $(1-\alpha)100\%$ bagi parameter β_j:

$$\boldsymbol{b_j \pm t_{\frac{\alpha}{2},(n-p)} s \sqrt{c_{jj}}} \qquad \boldsymbol{,}\ j = 0,1,2,\dots,k$$

$$\boldsymbol{s = \sqrt{\frac{SSRes}{n-p}} = \sqrt{\frac{e'e}{n-(k+1)}} = \sqrt{\frac{\sum e_i^2}{n-(k+1)}}}$$

$\boldsymbol{s}$ adalah dugaan simpangan baku σ. $\boldsymbol{n}$ adalah banyaknya amatan. $\boldsymbol{p}$ adalah banyaknya parameter. $\boldsymbol{k}$ adalah banyaknya peubah penjelas. $\boldsymbol{c_{jj}}$ adalah nilai dari **baris ke-j** dan **kolom ke-j** pada matriks

$$(\boldsymbol{X'X})^{-1} = \begin{bmatrix} \boldsymbol{c_{00}} & \boldsymbol{c_{01}} & \cdots & \boldsymbol{c_{0k}} \\ \boldsymbol{c_{10}} & \boldsymbol{c_{11}} & \cdots & \boldsymbol{c_{1k}} \\ \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{c_{ko}} & \boldsymbol{c_{k1}} & \cdots & \boldsymbol{c_{kk}} \end{bmatrix}.$$

Uji Hipotesis Parameter Secara Simultan (Uji F Simultan)

Digunakan untuk mengetahui apakah peubah-peubah penjelas yang ada dalam model berpengaruh secara serempak atau tidak terhadap respon

Model regresi:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 + \cdots + \beta_k X_k + \varepsilon$$

$H_0 : \beta_1 = \beta_2 = \cdots = \beta_k = 0$

H_1 : minimal ada satu $\beta_j \neq 0$, $j = 1, 2, \ldots, k$

Uji Hipotesis Parameter Secara Simultan (Uji F Simultan)

Tabel Sidik Ragam

Sumber Keragaman	**Derajat Bebas**	**Jumlah Kuadrat**	**Kuadrat Tengah**	**F**
Regresi	k	$b'X'Y - \frac{Y'11Y'}{n}$	JKR/k	KTR/KTG
Galat	n-(k+1)	$Y'Y - b'X'Y$	JKG/n-(k+1)	
Total	n-1	$Y'Y - \frac{Y'11Y'}{n}$		

Tolak H_0 jika $F > F_{k,n-k-1,\alpha}$ atau $p - value < \alpha$

Uji Hipotesis Parameter Secara Parsial (Uji parsial)

Uji parsial memberikan informasi mengenai pentingnya sebuah peubah penjelas tertentu dalam model yang melibatkan peubah penjelas lainnya.

Hipotesis

$\boldsymbol{H_0}: \boldsymbol{\beta_j} = \mathbf{0}$ $\quad$ *(peubah penjelas x_j tidak berpengaruh terhadap peubah respon)*

$\boldsymbol{H_1}: \boldsymbol{\beta_j} \neq \mathbf{0}$ $\quad$ *(peubah penjelas x_j berpengaruh terhadap peubah respon, setelah peubah penjelas lainnya ada dalam model)*

Statistik Uji

$$t = \frac{b_j}{se(b_j)} = \frac{b_j}{\sqrt{s^2 c_{jj}}} \qquad , s^2 = KTG$$

Tolak $\boldsymbol{H_0}$ jika $|t| > t_{\frac{\alpha}{2}, n-p}$ atau $p - value < \alpha$

Uji Hipotesis Parameter Secara Parsial (Uji parsial)

Uji parsial memberikan informasi mengenai pentingnya sebuah peubah penjelas tertentu dalam model yang melibatkan peubah penjelas lainnya.

Hipotesis

$\boldsymbol{H_0}: \boldsymbol{\beta_j} = \mathbf{0}$ $\quad$ *(peubah penjelas x_j tidak berpengaruh terhadap peubah respon)*

$\boldsymbol{H_1}: \boldsymbol{\beta_j} \neq \mathbf{0}$ $\quad$ *(peubah penjelas x_j berpengaruh terhadap peubah respon, setelah peubah penjelas lainnya ada dalam model)*

Statistik Uji

$$t = \frac{b_j}{se(b_j)} = \frac{b_j}{\sqrt{s^2 c_{jj}}} \qquad , s^2 = KTG$$

Tolak $\boldsymbol{H_0}$ jika $|t| > t_{\frac{\alpha}{2}, n-p}$ atau $p - value < \alpha$

Ukuran Kebaikan Model

Nilai R^2 digunakan untuk mengukur kesesuaian model linier dengan data yang ada. Nilai R^2 yang tinggi atau mendekati satu tidak selalu berarti bahwa model tersebut cocok dengan data, karna dalam regresi linear berganda, semakin banyak peubah penjelas maka akan semakin tinggi nilai R^2. Untuk mengatasi hal tersebut, $adj\ \mathrm{R}^2$ $(\mathrm{R}_\mathrm{a}^2)$ juga digunakan untuk menilai kebaikan model

$$R_a^2 = 1 - \frac{JKG/(n-(k+1))}{JKT/(n-1)}$$

$$R_a^2 = 1 - \frac{n-1}{n-(k+1)}(1-R^2)$$

Penduga Selang Kepercayaan bagi Y

Selang kepercayaan $(1-\alpha)100\%$ bagi nilai rataan/nilai harapan y pada nilai x_* adalah

$$\boldsymbol{x_*}'\boldsymbol{b} \pm \boldsymbol{t}_{n-(k+1),\frac{\alpha}{2}}\, \boldsymbol{s}\sqrt{\boldsymbol{x}_*'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{x}_*}$$

Selang kepercayaan $(1-\alpha)100\%$ bagi nilai y pada nilai x_* adalah

$$\boldsymbol{x_*}'\boldsymbol{b} \pm \boldsymbol{t}_{n-(k+1),\frac{\alpha}{2}}\, \boldsymbol{s}\sqrt{1+\boldsymbol{x}_*'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{x}_*}$$