PENGANTAR MODEL LINEAR

Oleh: Suryana

Model linear menyangkut masalah statistik yang ketergantungannya terhadap parameter secara linear. Bentuk umum model linear adalah

$Y = \beta_0 + \beta_1 X_1 + \ . \ . \ . \ + \beta_p X_p + \varepsilon$,

dengan

Y = Variabel respon

X_i = Variabel prediktor; $i = 1, 2, . . . , p$

β_i = Parameter model linear;

ε = error.

A. Variabel

Variabel dalam model linear dilihat dari skala pengukuran, dapat berbentuk diskrit maupun kontinyu. Variabel diskrit terdiri dari skala nominal dan ordinal. Variabel kontinyu terdiri dari skala interval dan rasio. Definisi skala pengukuran dapat dibaca pada buku *"Mengolah Data Statistik dengan Mudah Menggunakan Minitab 14"* oleh Iriawan dan Astuti (2006) hal. 12-13 atau buku-buku pengalahan data lainnya.

Dilihat dari cara pengukuran variabel dapat berbentuk variabel tetap (*fixed variable*) atau variabel random (*random variable*). Variabel tetap diperoleh dari rancangan percobaan. Variabel ini didisain terlebih dahulu untuk mengontrol pengamatan. Biasanya level-levelnya ditetapkan terlebih dahulu. Sebagai contoh, dalam pengamatan produktivitas padi ditetapkan variabel kualitas benih (X_1) terdiri dari tidak berlabel, label merah jambu, dan labil biru dan metode pengolahan (X_2) terdiri dari intensifikasi dan non intensifikasi.

Berbeda dengan variabel tetap, variabel random tidak ditetapkan terlebih dahulu melainkan langsung diamati dan hasil pengukurannya dicatat. Sebagai contoh berat gabah kering panen (Y_1) dan berat gabah kering giling (Y_2) dalam penelitian produktivitas padi.

B. ANOVA, MANOVA, dan ANACOVA

Kombinasi variabel tetap dan variabel random dalam model linear menghasilkan metode analisis yang berbeda. ANOVA (*Analysis of Variance*) memerlukan prasyarat varibel prediktor merupakan variabel tetap dan variabel respon terdiri dari variabel random. Bila variabel respon lebih dari satu, metode analisisnya menggunakan MANOVA (*Multivariate Analysis of Variance*). Selanjutnya, bila variabel prediktor terdiri dari kombinasi antara variabel tetap dan variabel random maka metode analisis yang digunakan dinamakan ANACOVA.

C. Rancangan Percobaan dan Model Linear

Dalam penelitian produktivitas padi di atas, rancangan percobaan yang digunakan adalah 3*x*2 rancangan faktorial. Desainnya sebagai berikut:

Untuk Y_1

	X_{21}	X_{22}
X_{11}	$Y_{111}, Y_{112}, \ldots, Y_{11n_{11}}$	$Y_{211}, Y_{212}, \ldots, Y_{21n_{21}}$
X_{12}	$Y_{121}, Y_{122}, \ldots, Y_{12n_{12}}$	$Y_{221}, Y_{222}, \ldots, Y_{22n_{22}}$
X_{13}	$Y_{131}, Y_{132}, \ldots, Y_{13n_{13}}$	$Y_{321}, Y_{322}, \ldots, Y_{32n_{32}}$

Untuk Y_2

	X_{21}	X_{22}
X_{11}	$Y_{111}, Y_{112}, \ldots, Y_{11n_{11}}$	$Y_{211}, Y_{212}, \ldots, Y_{21n_{21}}$
X_{12}	$Y_{121}, Y_{122}, \ldots, Y_{12n_{12}}$	$Y_{221}, Y_{222}, \ldots, Y_{22n_{22}}$
X_{13}	$Y_{131}, Y_{132}, \ldots, Y_{13n_{13}}$	$Y_{321}, Y_{322}, \ldots, Y_{32n_{32}}$

Catatan:

a. Rancangan diatas menggunakan n_{ij} replikasi. Dengan rancangan ini, dihasilkan data tidak seimbang (*unbalanced data*). Bila n_{ij} = c maka akan dihasilkan data seimbang (*balanced data*).
b. Prinsip randomisasi harus dipertahankan. Gunakan angka random untuk melakukan perlakuan mana yang harus dilakukan lebih dahulu.

c. Fungsi replikasi untuk memperkecil kesalahan.

d. $n = n_{11} + n_{12} + \ . \ . \ . \ + n_{32}$.

C.1 Model 3*x*2 Rancangan Faktorial

Untuk variabel respon u dengan replikasi n_{ij} = 2, maka model 3*x*2 rancangan faktorial adalah:

$$Y_{ijku} = \mu_u + \tau_{iu} + \gamma_{ju} + (\tau\gamma)_{iju} + \varepsilon_{ijku}$$

Dengan

u = 1, 2; i = 1, 2, 3; j = 1, 2; k = 1, 2.

Untuk menyingkat, variabel respon dapat dinyatakan dalam bentuk vektor

$\mathbf{Y} = \begin{bmatrix} Y_1 & Y_2 \end{bmatrix}^t$.

Untuk u = 1,

$$\begin{aligned}
Y_{1111} &= \mu_1 + \tau_{11} + \gamma_{11} + (\tau\gamma)_{111} + \varepsilon_{1111} \\
Y_{1121} &= \mu_1 + \tau_{11} + \gamma_{11} + (\tau\gamma)_{111} + \varepsilon_{1121} \\
Y_{1211} &= \mu_1 + \tau_{11} + \gamma_{21} + (\tau\gamma)_{121} + \varepsilon_{1211} \\
Y_{1221} &= \mu_1 + \tau_{11} + \gamma_{21} + (\tau\gamma)_{121} + \varepsilon_{1221} \\
Y_{2111} &= \mu_1 + \tau_{21} + \gamma_{11} + (\tau\gamma)_{111} + \varepsilon_{2111} \\
Y_{2121} &= \mu_1 + \tau_{21} + \gamma_{11} + (\tau\gamma)_{111} + \varepsilon_{2121} \\
Y_{2211} &= \mu_1 + \tau_{21} + \gamma_{21} + (\tau\gamma)_{221} + \varepsilon_{2211} \\
Y_{2221} &= \mu_1 + \tau_{21} + \gamma_{21} + (\tau\gamma)_{221} + \varepsilon_{2221} \\
Y_{3111} &= \mu_1 + \tau_{31} + \gamma_{11} + (\tau\gamma)_{311} + \varepsilon_{3111} \\
Y_{3121} &= \mu_1 + \tau_{31} + \gamma_{11} + (\tau\gamma)_{311} + \varepsilon_{3121} \\
Y_{3211} &= \mu_1 + \tau_{31} + \gamma_{21} + (\tau\gamma)_{321} + \varepsilon_{3211} \\
Y_{3221} &= \mu_1 + \tau_{31} + \gamma_{21} + (\tau\gamma)_{321} + \varepsilon_{3221}
\end{aligned}$$

Dalam bentuk $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, diperoleh

$$\begin{bmatrix} Y_{111} \\ Y_{112} \\ Y_{121} \\ Y_{122} \\ Y_{211} \\ Y_{212} \\ Y_{221} \\ Y_{222} \\ Y_{311} \\ Y_{312} \\ Y_{321} \\ Y_{322} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \tau_3 \\ \gamma_1 \\ \gamma_2 \\ (\tau\gamma)_{11} \\ (\tau\gamma)_{12} \\ (\tau\gamma)_{21} \\ (\tau\gamma)_{22} \\ (\tau\gamma)_{31} \\ (\tau\gamma)_{32} \end{bmatrix} + \begin{bmatrix} \varepsilon_{111} \\ \varepsilon_{112} \\ \varepsilon_{121} \\ \varepsilon_{122} \\ \varepsilon_{211} \\ \varepsilon_{212} \\ \varepsilon_{221} \\ \varepsilon_{222} \\ \varepsilon_{311} \\ \varepsilon_{312} \\ \varepsilon_{321} \\ \varepsilon_{322} \end{bmatrix}$$

Ternyata matrik **X** yang dihasilkan singular. Vektor kolom matrik **X** bersifat dependen. Perhatikan vektor kolom-1 merupakan jumlah vektor kolom-2, kolom-3, dan kolom-4 atau jumlahan vektor kolom-5 dan kolom-6 dan lain-lain. Dengan demikian rank matrik **X** tidak penuh (*not full rank*). Akibatnya, inverse matrik **X** tidak ada dan estimasi parameter **β tidak dapat dihitung.**

Agar estimasi parameter **β** dapat hitung, perlu dilakukan reparameterisasi dengan memberikan batasan (*constrain*):

$$\sum_{i=1}^{3} \tau_i = \sum_{j=1}^{2} \gamma_j = 0;$$

$$\sum_{i=1}^{3} (\tau\gamma)_{ij} = \sum_{j=1}^{2} (\tau\gamma)_{ij} = 0$$

Akibatnya,

$$\tau_3 = -\tau_1 - \tau_2$$

$$\gamma_2 = -\gamma_1$$

$$(\tau\gamma)_{12} = -(\tau\gamma)_{11}$$

$$(\tau\gamma)_{22} = -(\tau\gamma)_{21}$$

$$(\tau\gamma)_{32} = -(\tau\gamma)_{31} = (\tau\gamma)_{11} + (\tau\gamma)_{21}$$

$$(\tau\gamma)_{31} = -(\tau\gamma)_{11} - (\tau\gamma)_{21}$$

Sehingga $\mathbf{Y} = \mathbf{X}^* \boldsymbol{\beta}^* + \boldsymbol{\varepsilon}$ menjadi

$$\begin{bmatrix} Y_{111} \\ Y_{112} \\ Y_{121} \\ Y_{122} \\ Y_{211} \\ Y_{212} \\ Y_{221} \\ Y_{222} \\ Y_{311} \\ Y_{312} \\ Y_{321} \\ Y_{322} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & -1 & -1 & 0 \\ 1 & 1 & 0 & -1 & -1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & -1 & 0 & -1 \\ 1 & 0 & 1 & -1 & 0 & -1 \\ 1 & -1 & -1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & -1 & -1 \\ 1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & -1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \gamma_1 \\ (\tau\gamma)_{11} \\ (\tau\gamma)_{21} \end{bmatrix} + \begin{bmatrix} \varepsilon_{111} \\ \varepsilon_{112} \\ \varepsilon_{121} \\ \varepsilon_{122} \\ \varepsilon_{211} \\ \varepsilon_{212} \\ \varepsilon_{221} \\ \varepsilon_{222} \\ \varepsilon_{311} \\ \varepsilon_{312} \\ \varepsilon_{321} \\ \varepsilon_{322} \end{bmatrix}$$

Dengan

$$\mathbf{X}^* = \begin{bmatrix} 1 & 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & -1 & -1 & 0 \\ 1 & 1 & 0 & -1 & -1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & -1 & 0 & -1 \\ 1 & 0 & 1 & -1 & 0 & -1 \\ 1 & -1 & -1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & -1 & -1 \\ 1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & -1 & 1 & 1 \end{bmatrix} \text{ dan } \boldsymbol{\beta}^* = \begin{bmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \gamma_1 \\ (\tau\gamma)_{11} \\ (\tau\gamma)_{21} \end{bmatrix}$$

Dalam SPSS, matrik $\mathbf{X}^*$ **dinamakan matrik konfigurasi.**

C.2 Estimasi Parameter β.

Ada tiga metode mencari parameter **β** yaitu:

a. Ordinary Least Square (OLS),
b. Maximum Likelihood (ML),
c. Weighted Least Square (WLS).

Baik OLS maupun ML menghasilkan :

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^{*\prime}\mathbf{X}^{*}\right)^{-1}\mathbf{X}^{*\prime}\mathbf{Y}$$

Dengan bantuan Minitab 14,

$$\mathbf{X}^{*\prime}\mathbf{X}^{*} = \begin{bmatrix} 12 & 0 & 0 & 0 & 0 & 0 \\ 0 & 8 & 4 & 0 & 0 & 0 \\ 0 & 4 & 8 & 0 & 0 & 0 \\ 0 & 0 & 0 & 12 & 0 & 0 \\ 0 & 0 & 0 & 0 & 8 & 4 \\ 0 & 0 & 0 & 0 & 4 & 8 \end{bmatrix}$$

Dan

$$\left(\mathbf{X}^{*\prime}\mathbf{X}^{*}\right)^{-1} = \begin{bmatrix} 0.0833 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.1667 & -0.083 & 0 & 0 & 0 \\ 0 & -0.083 & 0.1667 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.0833 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.1667 & -0.083 \\ 0 & 0 & 0 & 0 & -0.083 & 0.1667 \end{bmatrix}$$

Catatan:

a. Partisi Matrik

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}$$

$$|\mathbf{A}| = |\mathbf{A}_{11}|\left|\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\right|$$

$$\mathbf{A}^{-1} = \begin{bmatrix} \mathbf{A}_{11}^{-1} + \mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{B}^{-1}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & -\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{B}^{-1} \\ -\mathbf{B}^{-1}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & \mathbf{B}^{-1} \end{bmatrix}$$

b. Penyebab matrik tidak *full rank*.

   a. Jumlah data kurang dari jumlah variabel,

   b. Minimum dibutuhkan data $n > p + 1$,

   c. Terdapat vektor kolom yang dependen.

c. Kesimpulan: model $3x2$ rancangan faktorial dapat dinyatakan sebagai model linear.