BAB III

LATENT CLASS MODEL

3.1 MODEL

Latent class model adalah suatu model matematika yang menghubungkan probabilitas respon suatu individu untuk *K* variabel indikator kategorik, dengan suatu variabel laten *X* yang bersifat kategorik dengan *M* kelas. Misalkan $Y_1, Y_2, \ldots, Y_K$ adalah variabel indikator yang merupakan variabel biner.

Suatu individu akan memberikan jawaban (respon) untuk setiap variabel indikator. Misal y_{ki} adalah respon individu ke-i terhadap variabel indikator Y_k, i = 1,2,…,n ; k = 1,2,…,K. Sebut $\tilde{\boldsymbol{y}}_i = \left(y_{1i}, y_{2i}, \ldots, y_{Ki}\right)$ adalah respon individu ke-i untuk K variabel indikator.

Dimana:

$y_{ki} = 1$; jika individu i memberi jawaban "iya" untuk variabel indikator $\mathrm{Y}_k, k = 1,2,\ldots,K$

$y_{ki} = 0$; jika individu i memberi jawaban "tidak" untuk variabel indikator $\mathrm{Y}_k, k = 1,2,\ldots,K$

Sebut:

$$\Pr\left(\tilde{Y}_i = \tilde{\boldsymbol{y}}_i\right) = \Pr\left(Y_{1i} = y_{1i}, Y_{2i} = y_{2i}, \ldots, Y_{Ki} = y_{Ki}\right)$$

= probabilitas suatu individu memberikan respon y_{1i} untuk variabel Y_{1i}, y_{2i} untuk variabel Y_{2i}, …, y_{Ki} untuk variabel Y_{Ki}.

$$\Pr\left(\tilde{Y}_i = \tilde{\boldsymbol{y}}_{\boldsymbol{i}} \mid X = m\right) = \Pr\left(Y_{1i} = y_{1i}, Y_{2i} = y_{2i}, \ldots, Y_{Ki} = y_{Ki} \mid X = m\right)$$

= probabilitas suatu individu memberikan respon y_{1i} untuk variabel Y_{1i}, y_{2i} untuk variabel Y_{2i}, …, y_{Ki} untuk variabel Y_{Ki}, jika diketahui individu tersebut berada pada *latent class* ke-*m* dari variabel laten *X*.

$$\Pr\left(\tilde{Y}_i = \tilde{\boldsymbol{y}}_{\boldsymbol{i}}, X = m\right) = \Pr\left(Y_{1i} = y_{1i}, Y_{2i} = y_{2i}, \ldots, Y_{Ki} = y_{Ki}, X = m\right)$$

= probabilitas suatu individu *i* memberikan respon y_{1i} untuk variabel Y_{1i}, y_{2i} untuk variabel Y_{2i}, …, y_{Ki} untuk variabel Y_{Ki}, dan berada pada kelas laten ke-*m* dari variabel laten *X*.

π_m $= \Pr\left(X = m\right)$

= probabilitas suatu individu berada pada kelas ke-m dari variabel laten X ; m = 1,2,…

$$= \frac{\text{banyaknya individu di kelas ke-m}}{\text{banyak individu seluruhnya}}$$

$P_{k|m}$ $= \Pr\left(Y_{ki} = y_{ki} \mid X = m\right)$

= probabilitas menjawab “iya” untuk variable indicator Y_k, jika diketahui individu tersebut berada di kelas ke-m dari variable laten X

$$= \frac{\text{banyaknya individu di kelas ke-m yang menjawab "iya" untuk } Y_k}{\text{banyak individu di kelas ke-m}}$$

$Y_{ki|m}$, $k = 1,2,...,K$ adalah respon individu *i* untuk variabel indikator Y_k, jika diketahui individu tersebut berada di kelas ke-m dari variabel laten X.

$Y_{ki|m}$ berdistribusi binomial (1, $P_{k|m}$).

$$\Pr\left(Y_{ki} = y_{ki} \mid X = m\right) = P_{k|m}{}^{y_{ki}}\left(1 - P_{k|m}\right)^{1-y_{ki}}$$

k = 1, 2, …, K.

Maka probabilitas suatu individu yang berada pada kelas ke-*m* dari variabel laten *X* memberikan respon $\tilde{\boldsymbol{y}}_i$ dapat didefinisikan sebagai:

$$\begin{aligned}\Pr\left(\tilde{Y}_i = \tilde{\boldsymbol{y}}_i \mid X = m\right) &= \Pr\left(Y_{1i} = y_{1i}, Y_{2i} = y_{2i}, \ldots, Y_{Ki} = y_{Ki} \mid X = m\right) \\ &= \prod_{k=1}^{K} P_{k|m}{}^{y_{ki}}\left(1 - P_{k|m}\right)^{1-y_{ki}}\end{aligned} \tag{3.1.1}$$

Karena variabel-variabel $Y_{1i}, Y_{2i}, \ldots, Y_{Ki}$, dan *X* merupakan variabel kategorik, maka $\Pr\left(Y_{1i} = y_{1i}, Y_{2i} = y_{2i}, \ldots, Y_{Ki} = y_{Ki}, X = m\right)$ merupakan pdf bersama untuk $Y_{1i}, Y_{2i}, \ldots, Y_{Ki}, X$; $\Pr\left(Y_{1i} = y_{1i}, Y_{2i} = y_{2i}, \ldots, Y_{Ki} = y_{Ki}\right)$ merupakan pdf bersama dari variabel $Y_{1i}, Y_{2i}, \ldots, Y_{Ki}$ atau pdf marginal dari variabel $Y_{1i}, Y_{2i}, \ldots, Y_{Ki}$, dan *X*. Jadi,

$$\begin{aligned}\Pr\left(Y_{1i} = y_{1i}, Y_{2i} = y_{2i}, \ldots, Y_{Ki} = y_{Ki}\right) &= \sum_{m=1}^{M} \Pr\left(Y_{1i} = y_{1i}, Y_{2i} = y_{2i}, \ldots, Y_{Ki} = y_{Ki}, X = m\right) \\ &= \sum_{m=1}^{M} \Pr\left(\tilde{Y}_i = \tilde{y}_i, X = m\right)\end{aligned} \tag{3.1.2}$$

Dengan mensubstitusikan persamaan (3.1.1), didapatkan:

$$\begin{aligned}\Pr\left(\tilde{Y}_i = \tilde{y}_i, X = m\right) &= \Pr\left(X = m\right)\Pr\left(\tilde{Y}_i = \tilde{y}_i \mid X = m\right) \\ &= \pi_m \prod_{k=1}^{K} P_{k|m}{}^{y_{ki}} \left(1 - P_{k|m}\right)^{1-y_{ki}}\end{aligned} \tag{3.1.3}$$

Dengan mensubstitusikan persamaan (3.1.3) ke persamaan (3.1.2), maka dapat $\Pr\left(Y_{1i} = y_{1i}, Y_{2i} = y_{2i}, \ldots, Y_{Ki} = y_{Ki}\right)$ dituliskan sebagai:

$$\begin{aligned}\Pr\left(Y_{1i} = y_{1i}, Y_{2i} = y_{2i}, \ldots, Y_{Ki} = y_{Ki}\right) &= \sum_{m=1}^{M} \Pr\left(\tilde{Y}_i = \tilde{y}_i, X = m\right) \\ &= \sum_{m=1}^{M} \pi_m \Pr\left(\tilde{Y}_i = \tilde{y}_i \mid X = m\right) \\ &= \sum_{m=1}^{M} \pi_m \prod_{k=1}^{K} P_{k|m}{}^{y_{ki}} \left(1 - P_{k|m}\right)^{1-y_{ki}}\end{aligned}$$

Notasikan persamaan diatas dengan $= \Pi_{\tilde{y}_i}$

$$\Pi_{\tilde{y}_i} = \sum_{m=1}^{M} \pi_m \prod_{k=1}^{K} P_{k|m}{}^{y_{ki}} \left(1 - P_{k|m}\right)^{1-y_{ki}} \tag{3.1.4}$$

Persamaan di atas disebut *latent class model*.

Jadi, *latent class model* merepresentasikan probabilitas bersama dari variabel-variabel $Y_1, Y_2, \ldots, Y_K$, dalam bentuk parameter-parameter *latent class*, π_m dan $p_{k|m}$.

Selanjutnya, notasikan:

$$\begin{aligned}\Pi_{m|\tilde{\boldsymbol{y}}_i} &= \Pr\left(X = m \,|\, (\tilde{\boldsymbol{Y}}_i = \tilde{\boldsymbol{y}}_i\right) \\ &= \Pr\left(X = m \,|\, Y_{1i} = y_{1i}, Y_{2i} = y_{2i}, \ldots, Y_{Ki} = y_{Ki}\right)\end{aligned} \tag{3.1.5}$$

sebagai probabilitas bersyarat suatu individu akan berada pada kelas ke-*m* dari variabel laten X, jika diketahui individu tersebut memberikan respon $\left(y_{1i}, y_{2i}, \ldots, y_{Ki}\right)$ untuk variabel indikator $\left(Y_{1i}, Y_{2i}, \ldots, Y_{Ki}\right)$.

Dengan menggunakan teorema bayes, persamaan (3.1.5) menjadi:

$$\begin{aligned}\Pi_{m|\tilde{\boldsymbol{y}}_i} &= \frac{\Pr\left(X = m\right)\Pr\left(Y_{1i} = y_{1i}, Y_{2i} = y_{2i}, \ldots, Y_{Ki} = y_{Ki} \,|\, X = m\right)}{\sum_{s=1}^{M}\Pr\left(X = s\right)\Pr\left(Y_{1i} = y_{1i}, Y_{2i} = y_{2i}, \ldots, Y_{Ki} = y_{Ki} \,|\, X = s\right)} \\ &= \frac{\pi_m \prod_{k=1}^{K} P_{k|m}{}^{y_{ki}} \left(1 - P_{k|m}\right)^{1-y_{ki}}}{\sum_{s=1}^{M} \pi_s \prod_{k=1}^{K} P_{k|s}{}^{y_{ki}} \left(1 - P_{k|s}\right)^{1-y_{ki}}}\end{aligned} \tag{3.1.6}$$

Klasifikasi suatu individu pada kelas-kelas dari variabel laten *X* akan ditetapkan berdasarkan nilai probabilitas $\Pi_{m|\tilde{\boldsymbol{y}}_i}$ untuk setiap kelas-*m* dari individu tersebut. Kelas ke-*m* yang memberikan nilai terbesar untuk $\Pi_{m|\tilde{\boldsymbol{y}}_i}$ akan menjadi kelas dari variabel latent *X* untuk individu *i* tersebut yang memberikan jawaban $\tilde{\boldsymbol{y}}_i$.

3.2 PENAKSIRAN PARAMETER DALAM MODEL

Salah satu metode yang dapat digunakan untuk menaksir nilai parameter-parameter dalam *latent class model* adalah metode maksimum likelihood. Misalkan terdapat n individu yang diobservasi, maka didapatkan fungsi likelihood untuk *latent class model* adalah sebagai berikut:

$$L(\boldsymbol{\pi},\boldsymbol{P}) = \prod_{i=1}^{n}\left[\sum_{m=1}^{M}\pi_m \prod_{k=1}^{K} P_{k|m}{}^{y_{ki}}\left(1-P_{km}\right)^{1-y_{ki}}\right] \qquad (3.2.1)$$

dimana $\tilde{\boldsymbol{y}}_i = \left(y_{1i}, y_{2i}, \ldots, y_{Ki}\right)$, $\boldsymbol{\pi} = \left(\pi_1, \pi_2, \ldots, \pi_M\right); \boldsymbol{P} = \left(P_{11}, \ldots, P_{KM}\right)$.

Akan dicari nilai untuk π_m dan P_{km} ; $m = 1, 2, \ldots, M$ dan $k = 1, 2, \ldots, K$ yang memaksimumkan $L(\pi, \boldsymbol{P})$. Untuk mempermudah perhitungan, $L(\boldsymbol{\pi}, \boldsymbol{P})$ dapat dimodifikasi menjadi:

$$\begin{aligned} l = \ln\left(L(\boldsymbol{\pi},\boldsymbol{P})\right) &= \ln\left\{\prod_{i=1}^{n}\left[\sum_{m=1}^{M}\pi_m \prod_{k=1}^{K} P_{k|m}{}^{y_{ki}}\left(1-P_{k|m}\right)^{1-y_{ki}}\right]\right\} \\ &= \sum_{i=1}^{n}\ln\left[\sum_{m=1}^{M}\pi_m \prod_{k=1}^{K} P_{k|m}{}^{y_{ki}}\left(1-P_{k|m}\right)^{1-y_{ki}}\right] \end{aligned} \qquad (3.2.2)$$

Maka,

$$\frac{\partial l}{\partial \pi_m} = \sum_{i=1}^{n} \frac{\prod_{k=1}^{K} P_{k|m}{}^{y_{ki}}\left(1-P_{k|m}\right)^{1-y_{ki}}}{\sum_{s=1}^{M}\pi_s \prod_{j=1}^{K} P_{j|s}{}^{y_{ji}}\left(1-P_{j|s}\right)^{1-y_{ji}}}$$

untuk $m = 1, 2, \ldots, M$.

dan

$$\frac{\partial l}{\partial P_{km}} = \sum_{i=1}^{n} \left\{ \frac{\left[y_{ki} P_{k|m}{}^{y_{ki}-1} \left(1-P_{k|m}\right)^{1-y_{ki}} - \left(1-y_{ki}\right)\left(1-P_{k|m}\right)^{-y_{ki}} P_{k|m}{}^{y_{ki}} \right] \left[\pi_m \prod_{\substack{j=1 \\ j\neq k}}^{K} P_{j|m}{}^{y_{ji}} \left(1-P_{j|m}\right)^{1-y_{ji}} \right]}{\sum_{s=1}^{M} \pi_s \prod_{j=1}^{K} P_{j|s}{}^{y_{ji}} \left(1-P_{j|s}\right)^{1-y_{ji}}} \right\}$$

Untuk k = 1, 2, …, K.

Taksiran maksimum likelihood dari 3.2.1 sulit diselesaikan secara analitis dan memuat variable laten X. Oleh karena itu akan digunakan algoritma EM (*Expectation-Maximization*) untuk memperoleh taksiran dari parameter-parameter pada *latent class model*.

Algoritma EM adalah suatu proses iteratif untuk menghitung taksiran maksimum likelihood yang dilakukan dengan 2 tahap, tahapan E-*step* dan M-*step*.

a) E-*step* (langkah ekspektasi)

Seperti telah dijelaskan dalam BAB II, dalam E-*step* akan dicari $E\left[\log\left[p(x,y,\theta_t)\right] \middle| y,\hat{\theta}_{t-1}\right] = \sum_x \log p(x,y,\theta_t)\, p\left(x \mid y,\hat{\theta}_{t-1}\right)$. Dalam *latent class model*, tahapan E-step dilakukan untuk mencari ekspektasi dari $\log\left(\prod_{i=1}^{n} \Pr\left(\tilde{Y}_i = \tilde{y}_i, X = m\right)\right)$ untuk setiap kelas-*m* dari variabel laten *X*, dengan diberikan data terobservasi Y_i dan nilai asumsi untuk π_m dan $P_{k|m}$,

$m = 1,2,\ldots,M; k = 1,2,\ldots,K$.

Berdasarkan persamaan (3.1.3), $\log\left(\prod_{i=1}^{n}\Pr\left(\tilde{Y}_i = \tilde{\boldsymbol{y}}_i, X = m\right)\right)$ dapat dituliskan sebagai berikut:

$$\log\left(\prod_{i=1}^{n}\Pr\left(\tilde{Y}_i = \tilde{\boldsymbol{y}}_i, X = m\right)\right) = \log\left(\prod_{i=1}^{n}\left[\pi_m \prod_{k=1}^{K} P_{k|m}{}^{y_{ki}}\left(1 - P_{k|m}\right)^{1-y_{ki}}\right]\right)$$

E-*step* dilakukan dengan mencari

$$E\left[\log\left(\prod_{i=1}^{n}\left[\boldsymbol{\pi}_m{}^{(t)}\prod_{k=1}^{K}P_{k|m}^{(t)\,y_{ki}}\left(1-P_{k|m}^{(t)}\right)^{1-y_{ki}}\right]\right)\middle| y_i, \hat{\boldsymbol{\pi}}_m{}^{(t-1)}, \hat{P}_{k|m}{}^{(t-1)}\right]$$

dimana:

$\boldsymbol{\pi}_m{}^{(t)}$ adalah nilai π_m pada iterasi ke-t

$P_{k|m}^{(t)}$ adalah nilai $P_{k|m}$ pada iterasi ke-t

$$\begin{aligned}
&E\left(\log\left(\prod_{i=1}^{n}\Pr\left(\tilde{Y}_i = \tilde{\boldsymbol{y}}_i, X = m\right)\right) \mid y_i, \hat{\pi}_m{}^{(t-1)}, \hat{P}_{k|m}^{(t-1)}\right)\\
&= \sum_{m=1}^{M}\log\left(\prod_{i=1}^{n}\Pr\left(\tilde{Y}_i = \tilde{\boldsymbol{y}}_i, X = m\right)\right)\Pr\left(\tilde{\boldsymbol{y}}_i, x \mid \tilde{\boldsymbol{y}}_i, \hat{\pi}_m{}^{(t-1)}, \hat{P}_{k|m}^{(t-1)}\right)\\
&= \sum_{m=1}^{M}\sum_{i=1}^{n}\log\left(\Pr\left(\tilde{Y}_i = \tilde{\boldsymbol{y}}_i, X = m\right)\right)\Pr\left(x \mid y_i, \hat{\pi}_m{}^{(t-1)}, \hat{P}_{k|m}^{(t-1)}\right)\\
&= \sum_{m=1}^{M}\sum_{i=1}^{n}\log\left(\Pr\left(\tilde{Y}_i = \tilde{\boldsymbol{y}}_i, X = m\right)\right)\Pi_{m \mid y_i, \hat{\pi}_m^{(t-1)}, \hat{P}_{k|m}^{(t-1)}}\\
&= \sum_{m=1}^{M}\sum_{i=1}^{n}\log\left(\pi_m{}^{t}\prod_{k=1}^{K}P_{k|m}^{(t)y_{ki}}\left(1-P_{k|m}^{(t)}\right)^{1-y_{ki}}\right)\left(\frac{\hat{\pi}_m{}^{(t-1)}\prod_{k=1}^{K}\hat{P}_{k|m}^{(t-1)y_{ki}}\left(1-\hat{P}_{k|m}^{(t-1)}\right)^{1-y_{ki}}}{\sum_{m=1}^{M}\hat{\pi}_m{}^{(t-1)}\prod_{k=1}^{K}\hat{P}_{k|m}^{(t-1)y_{ki}}\left(1-\hat{P}_{k|m}^{(t-1)}\right)^{1-y_{ki}}}\right)
\end{aligned}$$

t = 1,2,...

b) M-*step* (langkah maksimisasi)

Setelah melakukan E-*step*, langkah selanjutnya dengan melakukan M-*step*, dimana pada proses ini akan dicari nilai taksiran untuk $\pi_m^{(t)}$ dan $P_{k|m}^{(t)}$ yang memaksimumkan

$$E\left(\log\left(\prod_{i=1}^{n}\Pr\left(\tilde{Y}_i = \tilde{\boldsymbol{y}}_i, X = m\right)\right) \mid \tilde{\boldsymbol{y}}_i, \hat{\pi}_m^{(t-1)}, \hat{P}_{k|m}^{(t-1)}\right)$$

$$= \sum_{m=1}^{M}\sum_{i=1}^{n}\log\left(\pi_m^{\ t}\prod_{k=1}^{K}P_{k|m}^{(t)y_{ki}}\left(1-P_{k|m}^{(t)}\right)^{1-y_{ki}}\right)\left(\frac{\hat{\pi}_m^{(t-1)}\prod_{k=1}^{K}\hat{P}_{k|m}^{(t-1)y_{ki}}\left(1-\hat{P}_{k|m}^{(t-1)}\right)^{1-y_{ki}}}{\sum_{m=1}^{M}\hat{\pi}_m^{(t-1)}\prod_{k=1}^{K}\hat{P}_{k|m}^{(t-1)y_{ki}}\left(1-\hat{P}_{k|m}^{(t-1)}\right)^{1-y_{ki}}}\right)$$

yang didapat pada E-step.

Proses E-*step* dan M-*step* ini akan dilakukan terus secara iteratif sampai didapatkan suatu taksiran untuk π_m danP_{km} yang konvergen atau didapatkan $\left|\hat{\pi}_m^{(t)} - \hat{\pi}_m^{(t-1)}\right|$ dan $\left|\hat{P}_{km}^{(t)} - \hat{P}_{km}^{(t-1)}\right|$, k = 1, 2, …, K ; m = 1, 2, …, M, yang cukup kecil.

3.3 UJI KECOCOKAN MODEL

Dalam pembentukan suatu model diperlukan uji kecocokan model. Dalam pembentukan *latent class model*, uji kecocokan model dilakukan untuk setiap banyak kelas $M \geq 1$, kemudian menentukan banyak kelas optimal yang memberikan model yang cocok . Uji kecocokan model pada tugas akhir ini akan dilakukan dengan uji rasio likelihood.

Akan diuji hipotesis sebagai berikut:

H_0 : Model cocok untuk suatu M tertentu ; M = 1, 2, 3, ….

H_1 : tidak demikian.

Pada pengujian hipotesis tersebut, ingin dilihat, apakah

$\hat{\Pi}_{\tilde{\boldsymbol{y}}_i} = P_{\tilde{\boldsymbol{y}}_l}$; untuk $\tilde{\boldsymbol{y}}_i = \tilde{\boldsymbol{y}}_l$.

$\hat{\Pi}_{\tilde{\boldsymbol{y}}_i}$ adalah taksiran dari model dan $P_{\hat{\boldsymbol{y}}_l}$ didapat dari data.

$$\hat{\Pi}_{\tilde{y}_i} = \sum_{m=1}^{M} \hat{\pi}_m \prod_{k=1}^{K} \hat{P}_{k|m}{}^{y_{ki}} \left(1-\hat{P}_{k|m}\right)^{1-y_{ki}}$$

$$P_{\tilde{y}_l} = \frac{\text{banyak individu yang memberi jawaban } \tilde{\boldsymbol{y}}_l}{\text{banyak individu seluruhnya}} = \frac{f_{\tilde{y}_l}}{n}$$

Definisikan fungsi likelihood sebagai berikut:

- $L_0 = \prod_{i=1}^{n} \left[\sum_{m=1}^{M} \hat{\pi}_m \prod_{k=1}^{K} \hat{P}_{k|m}{}^{y_{ki}} \left(1-\hat{P}_{k|m}\right)^{1-y_{ki}} \right] = \prod_{i=1}^{n} \left[\hat{\Pi}_{\tilde{y}_i} \right]$

  Setiap individu *i* akan memberikan jawaban (respon) untuk setiap variabel indikator. Karena terdapat K variabel indikator biner, maka aka nada 2^K kemungkinan jawaban dari suatu individu, yaitu

(0,0,…,0),…,(1,1,…,1). Notasikan dengan $\tilde{y}_l$. Dan $f_{\tilde{y}_l}$ adalah banyaknya individu yang memberi respon $\tilde{y}_l$. Oleh karena itu L_0 dapat dituliskan menjadi:

$$\begin{aligned} L_0 &= \prod_{i=1}^{n}\left[\sum_{m=1}^{M}\hat{\pi}_m \prod_{k=1}^{K}\hat{P}_{k|m}{}^{y_{ki}}\left(1-\hat{P}_{k|m}\right)^{1-y_{ki}}\right] \\ &= \prod_{i=1}^{n}\left[\hat{\Pi}_{\tilde{y}_i}\right] \\ &= \prod_{\tilde{\boldsymbol{y}}_l=(0,0,\ldots,0)}^{(1,1,\ldots,1)}\left[\hat{\Pi}_{\tilde{y}_l}\right]^{f_{\tilde{y}_l}} \\ &= \prod_{\tilde{\boldsymbol{y}}_l=(0,0,\ldots,0)}^{(1,1,\ldots,1)}\left[\hat{\Pi}_{\tilde{y}_l}\right]^{n.P_{\tilde{y}_l}} \end{aligned}$$

- $$\boldsymbol{L_1} = \prod_{\tilde{\boldsymbol{y}}_l=(0,0,\ldots,0)}^{(1,1,\ldots,1)}\left(\frac{f_{\tilde{\boldsymbol{y}}_l}}{n}\right)^{f_{\tilde{y}_l}}$$

$$= \prod_{\tilde{\boldsymbol{y}}_l=(0,0,\ldots,0)}^{(1,1,\ldots,1)}\left(P_{\tilde{\boldsymbol{y}}_l}\right)^{f_{\tilde{y}_l}} = \prod_{\tilde{\boldsymbol{y}}_l=(0,0,\ldots,0)}^{(1,1,\ldots,1)}\left(P_{\tilde{\boldsymbol{y}}_l}\right)^{n.P_{\tilde{y}_l}}$$

Pandang :

$$\begin{aligned} \lambda &= \frac{L_0}{\boldsymbol{L_1}} \\ &= \frac{\prod_{\tilde{\boldsymbol{y}}_l=(0,0,\ldots,0)}^{(1,1,\ldots,1)}\left[\hat{\Pi}_{\tilde{y}_l}\right]^{n.P_{\tilde{y}_l}}}{\prod_{\tilde{\boldsymbol{y}}_l=(0,0,\ldots,0)}^{(1,1,\ldots,1)}\left(P_{\tilde{\boldsymbol{y}}_l}\right)^{n.P_{\tilde{y}_l}}} \\ &= \prod_{\tilde{\boldsymbol{y}}_l=(0,0,\ldots,0)}^{(1,1,\ldots,1)}\frac{\left[\hat{\Pi}_{\tilde{y}_l}\right]^{n.P_{\tilde{y}_l}}}{\left(P_{\tilde{\boldsymbol{y}}_l}\right)^{n.P_{\tilde{y}_l}}} = \prod_{\tilde{\boldsymbol{y}}_l=(0,0,\ldots,0)}^{(1,1,\ldots,1)}\left[\frac{\hat{\Pi}_{\tilde{y}_l}}{P_{\tilde{\boldsymbol{y}}_l}}\right]^{n.P_{\tilde{y}_l}} \end{aligned}$$

Statistik uji yang digunakan dalam uji kecocokan *latent class model* yaitu

$$L^2 = -2\ln\lambda = -2\ln\left[\prod_{\tilde{\boldsymbol{y}}_l=(0,0,...,0)}^{(1,1,...,1)}\left[\frac{\hat{\Pi}_{\tilde{y}_l}}{P_{\tilde{\boldsymbol{y}}_l}}\right]^{n.P_{\tilde{y}_l}}\right]$$

$$= 2\ln\left[\prod_{\tilde{\boldsymbol{y}}_l=(0,0,...,0)}^{(1,1,...,1)}\left[\frac{\hat{\Pi}_{\tilde{y}_l}}{P_{\tilde{\boldsymbol{y}}_l}}\right]^{n.P_{\tilde{y}_l}}\right]^{-1}$$

$$= 2\ln\left[\prod_{\tilde{\boldsymbol{y}}_l=(0,0,...,0)}^{(1,1,...,1)}\left[\frac{P_{\tilde{\boldsymbol{y}}_l}}{\hat{\Pi}_{\tilde{y}_l}}\right]^{n.P_{\tilde{y}_l}}\right]$$

$$= 2\sum_{\tilde{\boldsymbol{y}}_l=(0,0,...,0)}^{(1,1,...,1)}\ln\left[\frac{P_{\tilde{\boldsymbol{y}}_l}}{\hat{\Pi}_{\tilde{y}_l}}\right]^{n.P_{\tilde{y}_l}}$$

$$= 2n\sum_{\tilde{\boldsymbol{y}}_l=(0,0,...,0)}^{(1,1,...,1)}P_{\tilde{\boldsymbol{y}}_l}\ln\left[\frac{P_{\tilde{\boldsymbol{y}}_l}}{\hat{\Pi}_{\tilde{y}_l}}\right]$$

L^2 akan berdistribusi $\chi^2{}_r$, dimana $r = 2^K - 1 - (M - 1 + MK)$. H_0 akan ditolak jika $L^2 > \chi^2{}_r$.

Pada kasus dimana $P_{(y_l)} = \hat{\Pi}_{(y_l)}$, untuk setiap y_l ; $y_l = (0,0,...,0),...,$, $(1,1,...,1)$ kecocokan model dianggap sempurna dan L^2 akan bernilai 0. Pada kasus $L^2 > 0$, L^2 menghitung kekurang cocokan model. Jadi, dapat disimpulkan bahwa L^2 semakin kecil (L^2 mendekati 0), model semakin cocok.

Untuk memilih *latent class model* terbaik, diantara yang cocok, atau dengan kata lain menentukan banyak kelas optimal yang cukup menjelaskan hubungan diantara variabel-variabel indikatornya adalah dengan melihat

%reduksi dari L^2. %reduksi L^2 dihitung dengan menganggap model dengan M=1 sebagai *baseline* (M_0), dan membandingkan $L^2\left(M_0\right)$ dengan L^2 yang didapat untuk M > 1.

$$\%\text{reduksi} L^2 = \frac{L^2\left(M_0\right) - L^2\left(M_{(M-1)}\right)}{L^2\left(M_0\right)} \times 100\% ;\quad \text{m} = 2, 3, \ldots$$

Model yang dipilih adalah model dengan banyak kelas M, dimana terdapat perbedaan nilai % reduksi L^2 model dengan banyak kelas M dengan nilai % reduksi L^2 model dengan banyak kelas (M -1) cukup signifikan.