BAB II
LANDASAN TEORI

2.1 Data Multivariat

Secara umum, analisis multivariat atau metode multivariat berhubungan dengan metode-metode statistik yang secara bersama-sama (simultan) melakukan analisis terhadap lebih dari dua variabel pada setiap objek atau orang. Jadi, bisa dikatakan analisis multivariat merupakan perluasan dari analisis univariat (seperti uji *t*) atau bivariat (seperti korelasi dan regresi sederhana).

Saat ini analisis multivariat mulai banyak digunakan dalam bidang ilmu, melengkapi analisis statistik univariat dan bivariat yang dahulunya menjadi 'satu-satunya' alternatif dalam analisis data. Bahkan dimasa mendatang, analisis multivariat dipastikan akan menggantikan peran pengolahan data dari analisis univariat dan bivariat, walaupun tidak dapat menggantikannya secara total. Namun demikian, tidak mudah mengartikan secara tepat apa itu statistik multivariat (Singgih, 2010).

Berdasarkan pengertian diatas dapat diambil kesimpulan bahwa data multivariat adalah data yang terdiri dari banyak variabel yang bisa dilihat pengaruh dan hubungan antara variabelnya.

2.2 Teknik Pengambilan Sampel

Sampel adalah sebagian dari populasi yang diambil sebagai sumber data dan dapat mewakili seluruh populasi (Riduwan, 2010). Ada dua macam teknik pengambilan sampel dalam penelitian yang umum dilakukan yaitu: *probabiliy sampling* dan *nonprobability sampling.*

a. Probability Sampling

Pada pengambilan sampel secara random, setiap unit populasi, mempunyai kesempatan yang sama untuk diambil sebagai sampel. Artinya jika elemen populasinya ada 100 dan yang akan dijadikan sampel adalah 25, maka setiap elemen tersebut mempunyai kemungkinan 25/100 untuk bisa dipilih menjadi sampel. Dengan cara random, pemilihan dapat diperkecil, sekecil mungkin. Ini merupakan salah satu usaha untuk mendapatkan sampel yang representatif. Keuntungan pengambilan sampel dengan *probability sampling* adalah sebagai berikut:

1. Derajat kepercayaan terhadap sampel dapat ditentukan.
2. Beda penaksiran parameter populasi dengan statistik sampel, dapat diperkirakan.
3. Besar sampel yang akan diambil dapat dihitung secara statistik.

Probability sampling digunakan untuk memberikan peluang yang sama pada setiap anggota populasi yang akan dipilih menjadi sampel penelitian. Yang tergolong ke dalam *probality sampling* sebagai berikut:

1. *Simple Random Sampling*

   Pengambilan sampel dengan cara *simple random sampling* digunakan tanpa memperhatikan tingkatan yang ada dalam anggota populasi. Hal ini dillakukan bila anggota populasi dianggap sejenis (homogen).

2. *Proportionate Stratifed Random Sampling*

   Proportionate stratifed random sampling merupakan pengambilan sampel dari anggota populasi secara acak dan berstrata menurut proporsionalnya. Digunakan apabila anggota populasi tak sejenis (heterogen).

3. *Disiproportionate Stratifed Random Sampling*

   Disiproportionate stratifed random sampling merupakan pengambilan sampel dari setiap strata jumlahnya sama tidak sebanding dengan jumlah populasi dengan proporsi sampel di setiap strata. Teknik pengambilan sampel ini digunakan apabila anggota populasi tak sejenis (heterogen).

4. *Cluster Sampling*

   Pengambilan sampel dilakukan terhadap sampling unit, dimana sampling unitnya terdiri dari satu kelompok (cluster). Tiap item (individu) di dalam kelompok yang terpilih akan diambil sebagai sampel. Cara ini dipakai bila populasi dapat dibagi dalam kelompok-kelompok dan setiap karakteristik yang dipelajari ada dalam setiap kelompok. Misalnya ingin meneliti gambaran karakteristik (umur, suku, pendidikan dan pekerjaan) orang tua mahasiswa FK SAINTEK UIN SUSKA RIAU. Mahasiswa FK dibagi dalam 6 tingkat (I s/d

VI). Pilih secara random salah satu tingkat (misal tingkat II). Maka orang tua semua mahasiswa yang berada pada tingkat II diambil sebagai sampel (cluster).

5. Sampel bertingkat (*Multi Stage Sampling*)

Multi stage sampling adalah metode teknik sampling dimana proses pengambilan sampelnya dilakukan secara bertingkat, baik bertingkat dua maupun lebih.

Misalnya kita ingin meneliti berat badan dan tinggi badan murid SMA. Sesuai kondisi dan perhitungan, maka jumlah sampel yang diambil adalah ± 2000. Maka prosedur pengambilan sampelnya sebagai berikut:

(Indonesia)
↓
27 Provinsi
↓
Provinsi Riau
↓
Kabupaten Meranti
↓
Kecamatan Rangsang
↓
Ada 3 SMA (± 2000)

Cara ini digunakan apabila jumlah populasi sangat besar dan biaya penelitian kecil.

b. Non Probability sampling

Pemilihan sampel dengan cara ini tidak menghiraukan prinsip-prinsip *probability*. Pemilihan sampel tidak secara random. Hasil yang diharapkan hanya merupakan gambaran kasar tentang suatu keadaan. Cara ini dipergunakan bila biaya sangat sedikit, hasilnya diminta segera. Cara-cara yang dikenal adalah sebagai berikut:

1. Sampel dengan Maksud *(Purposive Samping)*

Pengambilan sampel dengan maksud ini dilakukan hanya atas dasar pertimbangan penelitinya saja yang menganggap unsur-unsur yang dikehendaki telah ada dalam anggota sampel yang akan diambil.

2. Sampel Tanpa Sengaja *(Accidental Sampling)*

Sampel diambil atas dasar seandainya saja, tanpa direncanakan lebih dahulu. Juga jumlah sampel yang di kehendaki tidak berdasarkan pertimbangan yang dapat dipertanggungjawabkan, asal memenuhi keperluan saja. Kesimpulan yang diperoleh bersifat kasar dan sementara saja.

3. Sampel Berjatah *(Quota Sampling)*

Pengambilan sampel hanya berdasarkan pertimbangan peneliti saja, hanya disini besar dan kriteria sampel telah ditentukan lebih dahulu. Misalnya, sampel yang akan diambil berjumlah 100 orang dengan perincian 50 laki-laki dan 50 perempuan yang berumur 15-40 tahun. Cara ini dipergunakan kalau peneliti mengenal betul daerah dan situasi daerah dimana penelitian akan dilakukan.

2.3 Penyebaran Kuisioner

Kuisioner merupakan beberapa pertanyaan yang akan diserahkan kepada responden untuk memperoleh suatu informasi. Pertanyaan tersebut di isi oleh responden tanpa bantuan dari pihak peneliti. Oleh karena itu, peneliti harus mampu membuat pertanyaan yang benar-benar jelas (valid) dan tidak meragukan responden. Untuk menghindari hal tersebut, maka perlu dilakukan uji coba kuisioner. Uji coba ini bertujuan untuk melihat pertanyaan-pertanyaan yang layak untuk dimuat dalam kuisioner. Adapun alat yang digunakan untuk melihat valid tidaknya suatu pertanyaan maka di lakukan uji validitas dan reliabilitas.

a. Validitas

Dalam penelitian kualitatif, validitas merupakan pengukuran derajat ketepatan antara data yang terdapat pada objek penelitian, yaitu menunjukkan sejauh mana suatu alat mampu bekerja dengan baik, sehingga dapat memperkecil tingkat kesalahan hasil penelitian.

Suatu *instrument* dalam kuisioner dinyatakan valid jika:

1. Koefisien kolerasi *product moment* melebihi 0,3 (soegiyono, 1999).
2. Jika koefisien kolerasi *product moment* $> r_{tabel}(\alpha; n-2)$, n = jumlah sampel.

Rumus yang dapat digunakan untuk menguji validitas menggunakan teknik kolerasi *product moment* adalah:

$$r = \frac{n\left(\sum_{i=1}^{n} XY\right) - \left(\sum_{i=1}^{n} X\right)\left(\sum_{i=1}^{n} Y\right)}{\sqrt{\left[n\left(\sum_{i=1}^{n} X^2\right) - \left(\sum_{i=1}^{n} X\right)^2\right]\left[n\left(\sum_{i=1}^{n} Y^2\right) - \left(\sum_{i=1}^{n} Y\right)^2\right]}}$$

keterangan :

n : Jumlah sampel

X : Skor variabel

Y : Total skor variabel

b. Reliabilitas

Reliabilitas adalah untuk mengetahui sejauh mana hasil pengukuran tetap konsisten, apa bila dilakukan pengukuran dua kali atau lebih dengan ketentuan alat ukur sama, respondennya sama namun dalam waktu yang berbeda. Teknik pengukuran reliabilitas suatu instrument penelitian tergantung dari skala yang digunakan. Teknik-tekniknya antara lain sebagai berikut:

1. Teknik Alfa cronbach (pilihan jawaban ≥3)
2. Teknik tes-retes (pengujian dua kali)
3. Teknik Spearman Brown(pilihan jawaban hanya 2)
4. Teknik kuder dan Richardson (K- R 20) (pilihan jawaban hanya 2)
5. Teknik kuder dan Richardson (K- R 21) (pilihan jawaban hanya 2)

Teknik yang digunakan pada penelitian ini adalah teknik Alfa Cronbach, teknik ini digunakan jika jawaban yang diberikan responden berbentuk skala seperti 1-3, 1-5, serta 1-7.

Misalnya responden memberikan jawaban sebagai berikut:

1. Baik = 3
2. Sedang = 2

3. Buruk = 1

Suatu instrument penelitian dikatakan reliabilitas dengan menggunakan teknik ini, bila koefisien reliabilitasnya > 0,6.

Tahapan pengujian reliabilitas dengan teknik Alfa Cronbach :

a. Menentukan nilai variansi setiap butir pertanyaan

$$\sigma_i^2 = \frac{\sum_{i=1}^{n} X_i^2 - \frac{(\sum_{i=1}^{n} X_i)^2}{n}}{n}$$

keterangan:

σ_i^2 : varians tiap butir soal

b. Menentukan nilai varians total

$$\sigma_t^2 = \frac{\sum_{i=1}^{n} X^2 - \frac{(\sum_{i=1}^{n} X)^2}{n}}{n}$$

keterangan:

σ_i^2 : vairans total

c. Menentukan reliabilitas instrumen

$$r_{11} = \left[\frac{k}{k-1}\right]\left[1 - \frac{\sum_{i=1}^{n} \sigma_b^2}{\sigma_t^2}\right]$$

Keterangan:

$\sum_{i=1}^{n} \sigma_b^2$: Jumlah varias butir soal

k : Jumlah butir pertanyaan

r_{11} : Koefisien riliabilitas

2.4 Regresi Logistik

Menurut *Hosmer* dan *Lemeshow* (2000) regresi logistik adalah metode analisis statistika yang mendeskripsikan hubungan antara peubah terikat yang memiliki dua kategori atau lebih dengan satu atau lebih peubah bebas berskala kategori atau kontinu. Pendugaan parameter yang digunakan dalam regresi logistik adalah peluang maksimum (*maximum likelihood*). Model regresi logistik terdiri atas regresi logistik dengan variabel terikat biner, ordinal, dan multinomial. Model regresi logistik dengan salah satu metode yang dapat digunakan untuk mencari hubungan variabel terikat yang bersifat dichotomous (berskala nominal atau ordinal dengan dua kategori) atau polycothomous (mempunyai skala nominal atau ordinal dengan lebih dari dua kategori) dengan satu atau lebih variabel bebas dan variabel terikat bersifat kontinyu atau kategorik. Menurut Hosmer dan Lemeshow Bentuk spesifik dari model regresi logistik dinyatakan sebagai berikut:

$$\pi(x) = \frac{\exp(\beta 0+\beta 1x)}{1+\exp(\beta 0+\beta 1x)}$$

Untuk mempermudah menaksirkan parameter dari persamaan tersebut maka dilakukan transpormasi dengan menggunakan transpormasi logistik, uraian transpormasi tersebut adalah sebagai berikut :

$$\pi(x) = \frac{\exp(\beta 0+\beta 1x)}{1+\exp(\beta 0+\beta 1x)}$$

$$\{\pi(x)\}\{1+\exp(\beta 0+\beta 1x)\} = \exp(\beta 0+\beta 1x)$$

$$\{\pi(x)\}+\{\pi(x)\exp(\beta 0+\beta 1x)\} = \exp(\beta 0+\beta 1x)$$

$$\pi(x) = \exp(\beta 0+\beta 1x) - \pi(x)\exp(\beta 0+\beta 1x)$$

$$\pi(x) = \{1-\pi(x)\}\exp(\beta 0+\beta 1x)$$

$$\frac{\pi(x)}{1-\pi(x)} = \exp(\beta 0+\beta 1x)$$

$$ln\left(\frac{\pi(x)}{1-\pi(x)}\right) = ln\exp(\beta 0+\beta 1x)$$

$$\mathrm{g(x)} = (\beta 0+\beta 1x)$$

dengan g(x) disebut dengan bentuk logit.

2.4.1 Regresi Logistik Multinomial

Regresi logistik multinomial merupakan regresi logistik dimana variabel terikatnya bersifat *polychotomous* yaitu nilai variabel terikatnya lebih dari dua kategori. Regresi logistik multinomial digunakan untuk melihat pengaruh satu atau beberapa variabel bebas terhadap variabel terikat yang memiliki lebih dari dua kategori. Misalkan $P(y=j|x) = \pi_j(x)$ dimana j = 0, 1, 2, 3, . . ., k-1 adalah peubah nominal yang digunakan dalam model. Berikut ini adalah persamaan umum yang digunakan untuk menyatakan peluang bersyarat bagi setiap kategori (*Hosmer* dan *Lemeshow*, 2000) :

$$P(y=j|x) = \frac{\exp\left(\mathrm{g_j(x)}\right)}{1+\sum_{j=0}^{m-1}\exp\left(\mathrm{g_k(x)}\right)}$$

Keterangan :

$P(y = j|x)$: peluang bersyarat dari variabel terikat pada variabel x

$\pi_j(x)$: persamaan regresi logistik untuk variabel terikat j

$g_j(x)$: fungsi logit pada variabel terikat j, dengan j = 0, 1, 2

$g_k(x)$: fungsi logit pada variabel terikat k, dengan k = 0, 1, 2

Misalkan ada tiga kategori pada variabel terikat (Y) yang diberi kode 0, 1, 2 maka terdapat tiga fungsi peluang bersyarat untuk regresi logistik multinomial ini, fungsi logit untuk peluang bersyarat tersebur adalah :

$$P(y = 0|x) = \pi_0(x) = \frac{e^{g0(x)}}{e^{g0(x)} + e^{g1(x)} + e^{g2(x)}}$$

$$= \frac{1}{1 + e^{g1(x)} + e^{g2(x)}}$$

$$P(y = 1|x) = \pi_1(x) = \frac{e^{g1(x)}}{e^{g0(x)} + e^{g1(x)} + e^{g2(x)}}$$

$$P(y = 2|x) = \pi_2(x) = \frac{e^{g2(x)}}{e^{g0(x)} + e^{g1(x)} + e^{g2(x)}}$$

Dengan $g_0(x) = 0$, probabilitas terikat π_0, π_1, π_2 dan $\sum_{j=0}^{2} \pi j = 1$ (*Hosmer,* 2000)

Kemudian untuk mendapatkan fungsi logit dilakukan transformasi logit , model regresi logistik untuk *p* variabel bebas adalah :

$$\pi(x) = \frac{\exp(\beta0+\beta1x1+\beta2x2+\beta3x3+\cdots+\beta pxp)}{1+\exp(\beta0+\beta1x1+\beta2x2+\beta3x3+\cdots+\beta pxp)}$$

$$\{\pi(x)\}\{1 + \exp(\beta0 + \beta1x1 + \beta2x2 + \beta3x3 + \cdots + \beta pxp)\} = \exp(\beta0 + \beta1x1 + \beta2x2 + \beta3x3 + \cdots + \beta pxp)$$

$$\{\pi(x)\} + \{\pi(x)\exp(\beta0 + \beta1x1 + \beta2x2 + \beta3x3 + \cdots + \beta pxp)\} = \exp(\beta0 + \beta1x1 + \beta2x2 + \beta3x3 + \cdots + \beta pxp)$$

$$\pi(x) = \exp(\beta0 + \beta1x1 + \beta2x2 + \beta3x3 + \cdots + \beta pxp) - \pi(x)\exp(\beta0 + \beta1x1 + \beta2x2 + \beta3x3 + \cdots + \beta pxp)$$

$$\pi(x) = \{1 - \pi(x)\}\exp(\beta0 + \beta1x1 + \beta2x2 + \beta3x3 + \cdots + \beta pxp)$$

$$\frac{\pi(x)}{1-\pi(x)} = \exp(\beta0 + \beta1x1 + \beta2x2 + \beta3x3 + \cdots + \beta pxp)$$

$$ln\left(\frac{\pi(x)}{1-\pi(x)}\right) = ln\exp(\beta0 + \beta1x1 + \beta2x2 + \beta3x3 + \cdots + \beta pxp)$$

$$g(x) = (\beta0 + \beta1x1 + \beta2x2 + \beta3x3 + \cdots + \beta pxp)$$

Menurut *Hosmer*, untuk kasus regresi logistik multinomial dengan 3 kategori kita membutuhkan dua fugsi logit, sehingga bisa dituliskan sebagai berikut:

$$g1(x) = \beta10 + \beta11x1 + \beta12x2 + \beta13x3 + \cdots + \beta1pxp$$
$$g2(x) = \beta20 + \beta21x1 + \beta22x2 + \beta23x3 + \cdots + \beta2pxp$$

Dengan $g(x)$: fungsi logit regresi logistik multinomial

β : parameter fungsi regresi logistik multionomial

x : variabel bebas fungsi regresi logistik multionomial

Secara umum langkah-langkah dalam analisis regresi logistik multinomial adalah melakukan pendugaan parameter untuk mencari nilai estimatornya, melakukan pengujian secara simultan untuk mengetahui apakah ada variabel bebas yang berpengaruh terhadap variabel terikatnya, melakukan pengujian parameter secara parsial untuk mengetahui variabel bebas yang paling berpengaruh terhadap variabel terikat tersebut, membentuk model logistik serta menginterpretasikan dengan melihat probabilitas dari regresi logistik multinomial.

1. Pendugaan parameter

   Dalam model regresi logistik, nilai harapan antar variabel terikat tidak linier serta memiliki varian-varian yang tidak sama sehingga penduga parameter β diperoleh melalui metode *maximum likelihood*. Untuk memecahkan masalah sistem persamaan yang tidak linier tersebut, solusi yang digunakan adalah dengan mengestimasi β melalui proses iterasi *newton raphson*. Jika variabel terikat Y mempunyai lebih dari dua hasil mungkin, maka variabel terikat Y berdistribusi multinomial. Untuk membuat fungsi likelihood pada model dengan kategori respon lebih dari 2 (misalnya 3 kategori) hal pertama yang dilakukan adalah membuat 3 variabel biner yang dikode 0 atau 1 untuk menandai *group membership* (anggota kelompok) pada suatu observasi.

   Dalam kasus ini 3 variabel biner tersebut hanya untuk menjelaskan fungsi likelihood dan bukan digunakan untuk analisis regresi logistik multinomial yang sebenarnya. 3 variabel biner tersebut diberi kode sebagai berikut :

   Jika Y= 0 maka $Y_0 = 1$, $Y_1 = 0$, dan $Y_2 = 0$; jika Y= 1 maka $Y_0 = 0$, $Y_1 = 1$, dan $Y_2 = 0$; jika Y= 2 maka $Y_0 = 0$, $Y_1 = 0$, dan $Y_2 = 1$. Jumlah variabel tersebut adalah $\sum_{j=0}^{2} Y_j = 1$.

   Fungsi *likelihood* untuk sebuah sampel dari n pengamatan :

$$l(\beta)= \prod_{i=1}^{n}[\pi_0\ (x_i)^{y_{0i}}\pi_1(x_i)^{y_{1i}}\pi_2(x_i)^{y_{2i}}]$$

$$= \prod_{i=1}^{n}\left(\frac{1}{1+e^{g1(xi)}+e^{g2(xi)}}\right)^{y_{0i}}\left(\frac{e^{g1(x)}}{1+e^{g1(xi)}+e^{g2(xi)}}\right)^{y_{1i}}\left(\frac{e^{g1(x)}}{1+e^{g1(xi)}+e^{g2(xi)}}\right)^{y_{2i}}$$

$$= \prod_{i=1}^{n}[e^{g1(xi)}]^{y_{1i}}\ [e^{g2(xi)}]^{y_{2i}}\left(\frac{1}{1+e^{g1(xi)}+e^{g2(xi)}}\right)^{y_{0i}+y_{1i}+y_{2i}} \tag{1}$$

Dengan mengambil logaritma natural persamaan (1) dan $\sum_{i=0}^{n} y_{ji} = 1$ untuk setiap i : 1,2,3,…, n maka diperoleh fungsi *log likelihood* sebagai berikut :

$$L(\beta) = \ln[l(\beta)]$$

$$= \ln\prod_{i=1}^{n}[e^{g1(xi)}]^{y_{1i}}\ [e^{g2(xi)}]^{y_{2i}}\left(\frac{1}{1+e^{g1(xi)}+e^{g2(xi)}}\right)^{y_{0i}+y_{1i}+y_{2i}}$$

$$= \ln\Big[\{\exp(\textstyle\sum_{i=1}^{n} g_1(x_i)\}^{\sum_{\mathrm{i}=1}^{\mathrm{n}} y_{1i}}\{\exp(\sum_{i=1}^{n} g_2(x_i)\}^{\sum_{\mathrm{i}=1}^{\mathrm{n}} y_{2i}}\{1+$$
$$\exp(\textstyle\sum_{i=1}^{n} g_1(x_i)+\sum_{i=1}^{n} g_2(x_i)\}^{-\sum_{\mathrm{i}=1}^{\mathrm{n}}(y_{0i}+y_{1i}+y_{2i}}\Big]$$

$$= \ln\{\exp(\textstyle\sum_{i=1}^{n} g_1(x_i)\}^{\sum_{\mathrm{i}=1}^{\mathrm{n}} y_{1i}}+\ln\{\exp(\sum_{i=1}^{n} g_2(x_i)\}^{\sum_{\mathrm{i}=1}^{\mathrm{n}} y_{2i}}+$$
$$\ln\{1+\exp(\textstyle\sum_{i=1}^{n} g_1(x_i)+\sum_{i=1}^{n} g_2(x_i)\}^{-\sum_{\mathrm{i}=1}^{\mathrm{n}}(y_{0i}+y_{1i}+y_{2i})}$$

$$= \textstyle\sum_{\mathrm{i}=1}^{\mathrm{n}} y_{1i}\ln\{\exp(g_1(x_i))\}+\sum_{\mathrm{i}=1}^{\mathrm{n}} y_{2i}\ln\{\exp(g_2(x_i))\}-$$
$$\textstyle\sum_{\mathrm{i}=1}^{\mathrm{n}}(y_{0i}+y_{1i}+y_{2i})\ln\big[1+exp\big(g_1(x_i)\big)+exp\big(g_2(x_i)\big)\big]$$

$$L(\beta) = \sum_{i=1}^{n}[y_{1i}g_1(x_i)+y_{2i}g_1(x_i)] - ln\big(1+e^{g1(xi)}+e^{g2(xi)}\big) \tag{2}$$

Agar nilai fungsi (β_{jk}) mencapai maksimum maka dilakukan diferensiasi terhadap (β_{jk}), dengan syarat $\frac{\partial L(\beta)}{\partial \beta_{jk}} = 0$. Untuk menyederhanakan notasi, $\pi_{ji} = \pi_{ji}(xi)$. Pada persamaan (2) dimisalkan :

$$a = \sum_{i=1}^{n}[y_{1i}g_1(x_i)+y_{2i}g_1(x_i)]$$
$$b = -ln\big(1+e^{g1(xi)}+e^{g2(xi)}\big)$$

Turunan pertama dari a

$$\frac{\partial L(\beta)}{\partial \beta_{jk}} = \sum_{i=1}^{n}\Big(y_{1i}\big(x_{0i}+x_{1i}+x_{2i}+\cdots+x_{pi}\big)\Big)+\Big(y_{2i}\big(x_{0i}+x_{1i}+x_{2i}+$$
$$\cdots+x_{pi}\big)\Big)$$

$$\frac{\partial L(\beta)}{\partial \beta_{jk}} = \sum_{i=1}^{n} x_{ki} y_{ji}$$

Turunan pertama dari b

$$\frac{\partial L(\beta)}{\partial \beta_{jk}} = -ln\left(1 + e^{g1(xi)} + e^{g2(xi)}\right)$$

$$= -\frac{1}{\left(1+e^{g1(xi)}+e^{g2(xi)}\right)}\left[1 + e^{g1(xi)} + e^{g2(xi)}\right]^{('\beta_{jk})}$$

$$= -\frac{1}{\left(1+e^{g1(xi)}+e^{g2(xi)}\right)}\left[e^{g1(xi)}(x_{0i} + x_{1i} + x_{2i} + \cdots + x_{pi}) + e^{g2(xi)}(x_{0i} + x_{1i} + x_{2i} + \cdots + x_{pi})\right]$$

$$= -\sum_{i=1}^{n} x_{ki}\left[\frac{e^{g1(xi)}+e^{g2(xi)}}{\left(1+e^{g1(xi)}+e^{g2(xi)}\right)}\right]$$

$$= -\sum_{i=1}^{n} x_{ki}\,\pi_{ji}$$

Jadi,

$$\frac{\partial L(\beta)}{\partial \beta_{jk}} = \sum_{i=1}^{n} x_{ki} y_{ji} - \sum_{i=1}^{n} x_{ki}\,\pi_{ji}$$

$$\frac{\partial L(\beta)}{\partial \beta_{jk}} = \sum_{i=1}^{n} x_{ki}(y_{ji} - \pi_{ji})$$

Bentuk umum persamaan *likelihood*nya adalah :

$$\frac{\partial L(\beta)}{\partial \beta_{jk}} = 0 \rightarrow \sum_{i=1}^{n} x_{ki}(y_{ji} - \pi_{ji}) = 0$$

$$\sum_{i=1}^{n} x_{ki} y_{ji} = \sum_{i=1}^{n} x_{ki}\pi_{ji}(xi)$$

$$\sum_{i=1}^{n} x_{ki}\,\frac{e^{gj(xi)}}{1+\sum_{i=1}^{n} e^{gj(xi)}}) = \sum_{i=1}^{n} x_{ki} y_{ji} \tag{3}$$

Untuk $j = 1,2,3, \dots, m-1$ dan $k = 1,2,3, \dots, p$

Turunan kedua $L(\beta)$ diperoleh :

$$\frac{\partial^2 L(\beta)}{(\partial \beta_{jk})^2} = \sum_{i=1}^{n}({x_{ji}}^2)\pi_{(xi)}\left[1 - \pi_{(xi)}\right] \tag{4}$$

Nilai β_{jk} dapat ditentukan, tetapi sangat sulit menghitung nilai β_{jk} karena persamaannya rumit. Oleh karena itu, untuk membantu penghitungan estimasi dari β_{jk} digunakan suatu metode khusus untuk menghitung estimator *maximum likelihood* β_{jk}, misalnya metode *Newthon-Raphson*. Metode *Newton-Raphson* melibatkan turunan pertama dan turunan kedua dari fungsi *log likelihood*.

Prosedur iterasi *Newton-Raphson* untuk menghitung estimator maksimum *Likelihood* $\hat{\beta}_{jk}$ adalah sebagai berikut:

a. Pilih taksiran awal untuk β_{jkh}, dengan $h = 1,2,3, \ldots$, misal $\beta_{jk1} = 0$

b. Pada setiap iterasi ke- $(h+1)$ hitung taksiran baru :

$$0 = \{M^T(y-\pi)_h\} + \left[\left(\hat{\beta}_{jkh+1} - \hat{\beta}_{jkh}\right).\{-I(\beta)_h\}\right]$$

$$\hat{\beta}_{jkh+1} = \hat{\beta}_{jkh} + \{I(\beta)_h\}^{-1}.\{M^T(y-\pi)_h\}$$

Dengan $I(\beta)$: matriks informasi

$$\hat{I}(\hat{\beta}) =$$

$$\begin{bmatrix} \hat{I}(\hat{\beta})_{11} & \hat{I}(\hat{\beta})_{12} & \cdots & \hat{I}(\hat{\beta})_{1(m-1)} \\ \hat{I}(\hat{\beta})_{21} & \hat{I}(\hat{\beta})_{22} & \cdots & \hat{I}(\hat{\beta})_{1(m-1)} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{I}(\hat{\beta})_{(m-1)1} & \hat{I}(\hat{\beta})_{(m-2)2} & \cdots & \hat{I}(\hat{\beta})_{(m-2)(m-1)} \\ \hat{I}(\hat{\beta})_{(m-1)1} & \hat{I}(\hat{\beta})_{(m-1)2} & \cdots & \hat{I}(\hat{\beta})_{(m-2)(m-1)} \end{bmatrix}_{(p+1)(m-1)x(p+1)(m-1)}$$

M^T: matriks transpose dari variabel bebas (x)

c. Iterasi berakhir jika diperoleh $\hat{\beta}_{jk(h+1)} \approx \hat{\beta}_{jkh}$

2. Pengujian parameter

Pengujian terhadap parameter model regresi logistik multinomial ini dilakukan sebagai upaya memeriksa peranan variabel bebas terhadap variabel terikat. Uji yang dilakukan ada dua yaitu pengujian secara serentak (simultan) dan pengujian secara parsial.

I. Pengujian parameter dengan uji *likelihood ratio* (uji simultan atau uji G)

Statistik uji G, yaitu uji yang digunakan untuk menguji apakah ada pengaruh variabel bebas terhadap variabel terikat secara bersama-sama, sekurang-kurangnya ada satu variabel bebas yang berpengaruh. Adapun pengujian hipotesis yang dilakukan adalah :

H_0 : $\beta_1 = \beta_2 = \cdots = \beta_p = 0$ yang artinya tidak ada pengaruh antara sekumpulan variabel bebas dengan variabel terikat

H_1 : minimal ada satu $\beta_j \neq 0$ yang artinya minimal ada satu variabel bebas yang berpengaruh terhadap variabel terikat. Dengan uji statistiknya :

$$G = -2ln\left(\frac{l_0}{l_k}\right)$$

Dimana l_0 adalah *likelihood* tanpa variabel bebas dan l_k adalah *likelihood* dengan variabel bebas. Statistik uji G ini mengikuti sebaran chi-square bila n mendekati tak terhingga dengan derajat bebas *p* dimana *p* = (r-1)(c-1), *r* dan *c* masing-masing adalah banyaknya kategori pada variabel bebas dan variabel terikat. H_0 akan ditolak pada tingkat signifikansi α apabila nilai $G > X^2{}_{p;a}$ atau $(p - value) < a$, dengan kesimpulan bahwa variabel bebas secara bersama-sama atau keseluruhan mempengaruhi variabel terikat, dapat juga dikatakan bahwa paling sedikit ada satu koefisien $\beta_j \neq 0$. Untuk mengetahui β_j mana yang berpengaruh signifikan, dapat dilakukan uji parameter β_j secara parsial dengan uji wald.

II. Pengujian parameter dengan uji *wald* (uji parsial)

Pengujian parameter satu per satu menggunakan statistik uji wald. Uji ini dilakukan dengan membandingkan model terbaik yang dihasilkan oleh uji simultan terhadap model tanpa variabel bebas didalam model terbaik. Uji parsial ini digunakan untuk melihat variabel bebas yang paling berpengaruh terhadap variabel terikat. Hipotesis yang akan diuji adalah sebagai berikut:

H_o : $\beta_j = 0$, yang artinya tidak ada pengaruh antara variabel bebas ke-j terhadap variabel terikat

H_1 : $\beta_j \neq 0$, yang artinya ada pengaruh variabel bebas terhadap variabel terikat. Statistik ujinya adalah :

$w = \left[\frac{\hat{\beta}_j}{s_e(\hat{\beta}_j)}\right]^2$, dimana $j = 1, 2, \dots, p$ dengan $\hat{\beta}_j$ adalah penduga dari β_j dan s_e adalah penduga galat baku dari β_j. Nilai $s_e(\hat{\beta}_j)$

diperoleh dari nilai diagonal utama matrik varian kovarian.

$var\ cov(\hat{\beta}_j) = [X'VX]^{-1}$

$$V = \begin{bmatrix} \hat{\pi}_1(1-\hat{\pi}_1) & 0 & \dots & 0 \\ 0 & \hat{\pi}_2(1-\hat{\pi}_2) & \dots & \vdots \\ \vdots & \vdots & \ddots & \hat{\pi}_n(1-\hat{\pi}_n) \\ 0 & 0 & \dots & \end{bmatrix}$$

$var(\hat{\beta}_j) = elemen\ ke - j\ matik\ diagonal\ [X'VX]^{-1}$

$s_e(\hat{\beta}_j) = \sqrt{var(\hat{\beta}_j}$

W diasumsikan mengikuti sebaran *chi-square* dengan derajat kebebasan 1. H_0 akan ditolak jika nilai $W > X^2{}_{(1;a)}$ atau $(p-value) < a$, jika H_0 ditolak maka dapat disimpulkan bahwa β_j signifikan. Dengan kata lain, variabel bebas X secara parsial berpengaruh signifikan terhadap variabel terikat.

3. Probabilitas regresi logistik multinomial

Misalkan **X** adalah variabel bebas yang berukuran $(p+1)$ dan variabel terikat **Y** (kategori) mempunyai kategori $j = 0,1,2$ dengan probabilitas respon π_0, π_1, π_2 dan $\sum_{j=0}^{2} \pi_j = 1$. Probabilitas bersyarat $P(y = j|x) = \pi_j, j = 0,1,2$

Jadi probabilitas bersyarat $P(y = j|x) = \pi_j,\ \ j = 0,1,2$ dapat ditulis sebagai berikut :

$\pi_0(x) = \frac{1}{1+e^{g1(x)}+e^{g2(x)}}$

$\pi_1(x) = \frac{e^{g1(x)}}{1+e^{g1(x)}+e^{g2(x)}}$

$\pi_2(x) = \frac{e^{g2(x)}}{1+e^{g1(x)}+e^{g2(x)}}$

Dengan :

$g1(x) = \beta_{10} + \beta_{11}X_1 + \beta_{12}X_2$

$g2(x) = \beta_{20} + \beta_{21}X_1 + \beta_{22}X_2$