MODUL TSS UMUM 2022

MATERI 2

Metode Logistic Regression Dengan Software R

Statistics Center Undip

Tabel Kontingensi

Tabel kontingensi (*cross tabulation*) adalah salah satu metode statistik yang merepresenatsikan dua atau lebih variable secara bersama-sama (simultan) dan hasilnya berupa table yang merefleksikan distribusi bersama atau lebih variable dengan kategori terbatas (Lutfia Rohmi, 2017). Metode ini bertujuan untuk mengetahui hubungan antara dua atau lebih variable yang bersifat non kausalitas (bukan sebab akibat). Struktur table kontingensi adalah sebagai berikut :

		B				Total
		B_1	B_2	...	B_j	
A	A_1	n_{11}	n_{12}	...	n_{1j}	$n_{1.}$
	A_2	n_{21}	n_{22}	...	n_{2j}	$n_{2.}$
	⋮	⋮	⋮	...	⋮	⋮
	A_i	n_{i1}	n_{i2}	...	n_{ij}	$n_{i..}$
	Total	$n_{.1}$	$n_{.2}$	...	$n_{.j}$	$n_{...}$

Analisis Regresi Logistik

Regresi merupakan sebuah metode analisis yang digunakan untuk mengetahui hubungan antara satu atau lebih variabel prediktor (x) dengan variabel respon (y). Metode regresi logistik merupakan suatu analisis regresi yang digunakan untuk menggambarkan hubungan antara variabel respon berskala nominal atau ordinal dengan dua atau lebih kategori dan sekumpulan variabel prediktor bersifat kontinu atau kategorik (Agresti, 1990). Tujuan melakukan analisis data menggunakan regresi logistik adalah untuk mendapatkan model terbaik dan sederhana.

Terdapat beberapa macam regresi logistik, yaitu regresi logistik biner, regresi logistik multinomial, dan regresi logistik ordinal.

1. Regresi Logistik Biner

   Regresi logistik biner merupakan metode regresi logistik yang digunakan untuk menganalisis hubungan antara satu variabel respon dan beberapa variabel prediktor. Variabel respon yang digunakan berupa data kualitatif dikotomi, yaitu bernilai 1 untuk menyatakan keberadaan sebuah karakteristik dan bernilai 0 untuk menyatakan ketidakberadaan sebuah karakteristik.

2. Regresi Logistik Multinomial

   Regresi logistik multinomial digunakan untuk menganalisis hubungan antara satu variabel respon dan beberapa variabel bebas, dengan variabel responnya berupa data kualitatif yang mempunyai kategori lebih dari dua.

3. Regresi Logistik Ordinal

   Regresi logistik ordinal adalah metode regresi logistik untuk menganalisis hubungan antara beberapa variabel prediktor yang bersifat kategorik maupun kontinu dengan satu variabel respon yang terdiri dari dua kategori dan terdapat tingkatan dalam kategori tersebut.

   Penentuan estimasi parameter pada model regresi logistik ditentukan dengan metode maksimum *likelihood* kemudian dilanjutkan dengan iterasi Newton Raphson untuk menyelesaikan persamaan non linier, seperti persamaan *likelihood*. Untuk menguji signifikansi parameter pada ketiga metode tersebut digunakan uji rasio *Likelihood* dan uji Wald, sedangkan untuk menguji kecocokan model digunakan uji *Goodness of Fit*.

- Uji Rasio *Likelihood*
  Uji Rasio *Likelihood* adalah uji yang membandingkan model yang mengandung variabel bebas dan model yang tidak mengandung variabel bebas. Uji ini juga digunakan untuk menganalisis apakah model signifikan atau tidak.

- Uji Wald
  Uji Wald digunakan untuk mengetahui variabel-variabel bebas manakah yang mempunyai hubungan lebih kuat dengan variabel respon.

- Uji *Goodness of Fit*
  Uji *Goodness of Fit* digunakan untuk mengetahui ada tidaknya perbedaan antara prediksi dan hasil observasi.

Analisis Regresi Logistik Multinomial

Analisis regresi logistik multinomial adalah salah satu bentuk analisis regresi logistik untuk menggambarkan hubungan beberapa variabel independen dengan suatu variabel dependen multinomial. Analisis ini digunakan jika variabel respon atau dependen berskala multinomial yaitu berskala nominal dengan lebih dari dua kategori (Hosmer dan Lemeshow, 2000). Data dengan skala nominal adalah data angka yang tidak menunjukkan adanya tingkatan. Apabila terdapat *k* kategorik pada variabel independen maka jumlah model logistik yang terbentuk adalah *k-1*. Bentuk umum dari regresi logistik multinomial adalah

$$\pi = \frac{e^{g(x)}}{1 + e^{g(x)}}$$

Dengan $g(x) = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$

Secara umum, bentuk dari fungsi logistic dengan variable respon atau dependen yang terdiri dari tiga kategori adalah sebagai berikut.

$$g_j(x) = \beta_{j0} + \beta_{j1} x_1 + \cdots + \beta_{jp} x_p$$

Suatu variable respon dengan tiga kategori akan menghasilkan dua persamaan logit, dimana masing-masing persamaan membentuk model regresi logistic multinomial yang membandingkan antar kelompok kategori terhadap pembanding, seperti dibawah.

$$g_1(x) = log\frac{P(Y = 2|x)}{P(Y = 1|x)}$$

$$= log\frac{\pi_2(x)}{\pi_1(x)} = \beta_{10} + \beta_{11} x_1 + \beta_{12} x_2 + \cdots + \beta_{1p} x_p$$

$$g_2(x) = log\frac{P(Y = 3|x)}{P(Y = 1|x)}$$

$$= log\frac{\pi_3(x)}{\pi_1(x)} = \beta_{20} + \beta_{21} x_1 + \beta_{22} x_2 + \cdots + \beta_{2p} x_p$$

Peluang dari masing-masing kategori respon sebagai berikut.

$$P(x) = \pi_1(x) = \frac{expg_1(x)}{1 + expg_1(x) + expg_2(x)}$$

$$P(x) = \pi_2(x) = \frac{expg_2(x)}{1 + expg_1(x) + expg_2(x)}$$

$$P(x) = \pi_3(x) = \frac{expg_3(x)}{1 + expg_1(x) + expg_2(x)}$$

a. Estimasi Parameter

Metode estimasi yang mengarah kepada metode kuadrat terkecil (*least squares*) dalam regresi linear disebut metode kemungkinan nilai maksimum atau *maximum likelihood* *estimator* (Hosmer dan Lemeshow, 1989). Metode ini digunakan untuk menduga parameter-parameter model regresi logistic dengan memberikan nilai estimasi β dengan memaksimumkan fungsi *likelihood* (Agresti,2002). Berikut ini fungsi *likelihood* untuk n sample random.

$l(\beta) = \prod_{i=1}^{n} \quad [\pi_0(x_i)^{y_{0i}}\pi_1(x_i)^{y_{1i}}\pi_2(x_i)^{y_{2i}}$, dengan i = 1,2,….,J

Dari persamaan di atas didapat fungsi *ln likelihood*, seperti dibawah

$$L(\beta) = \sum_{i=1}^{n} \quad y_{0_i} ln\,[\pi_0(x_i)] + y_{1_i} ln\,[\pi_1(x_i)] + y_{2_i} ln\,[\pi_2(x_i)]$$

Fungsi maksimum ln-*likelihood* diperoleh dengan mendiferensialkan $L(\beta)$ terhadap β dan disama dengankan dengan 0. *Maximum Likelihood Estimation* (MLE) adalah metode yang digunakan untuk menaksir varians dan kovarian dari taksiran β yang didapat dari turunan kedua fungsi ln-*likelihood*. Digunakan metode Newton Raphson untuk memaksimumkan fungsi *likelihood*. Langkah-langkah iterasi Newton Raphson adalah

1. Menentukan nilai penduga awal estimasi parameter $\hat{\beta}^{(0)}$. Penduga atau taksiran yang digunakan sama seperti pada metode regresi linear berganda dengan matriks sebagai berikut

   $$X = [1\; x_{11}\; x_{12}\; \dots\; x_{1p}\;\; 1\; x_{21}\; x_{22}\; \dots\; x_{2p}\;\; \vdots\; 1\;\; \vdots\; x_{n1}\;\; \vdots\; \dots\; \vdots\;\; x_{n2}\; \dots\; x_{np}\;\;]$$

2. Membentuk vector gradien

   $$g^{(t)}\left(g^{(t)}\right) = \left(\frac{\partial l(\beta)}{\partial \beta_0}\right), \frac{\partial l(\beta)}{\partial \beta_1}, \dots, \frac{\partial l(\beta)}{\partial \beta_p}$$

3. Membentuk matriks *Hessian* H

$$H^{(t)}\left(\beta^{(t)}\right) = \left[\frac{\partial^2 lnL(\beta)}{\partial \beta_0^2}\; \frac{\partial^2 lnL(\beta)}{\partial \beta_0 \partial \beta_1}\; \dots\; \frac{\partial^2 lnL(\beta)}{\partial \beta_0 \partial \beta_p}\; \frac{\partial^2 lnL(\beta)}{\partial \beta_0 \partial \beta_1}\; \frac{\partial^2 lnL(\beta)}{\partial \beta_1^2}\; \dots\; \dots\; \frac{\partial^2 lnL(\beta)}{\partial \beta_1 \partial \beta_p}\; \dots\; \frac{\partial^2 lnL(\beta)}{\partial \beta_0 \partial \beta_p}\; \dots\; \frac{\partial^2 lnL(\beta)}{\partial \beta_1 \partial \beta_p}\; \dots\; \partial^2\right.$$

4. Mensubstitusikan nilai $\hat{\beta}^{(0)}$ pada elemen vector g dan matriks H sehingga didapatkan vector $g^{(t)}\left(\beta^{(t)}\right)$ dan matriks $H^{(t)}\left(\beta^{(t)}\right)$
5. Iterasi dimulai dari t=0 pada persamaan dibawah ini

   $$\beta^{(t+1)} = \beta^{(t)} - (H^{(t)}\left(\beta^{(t)}\right))^{-1} g^{(t)}\left(\beta^{(t)}\right)$$

6. Apabila estimator parameter belum konvergen, maka Kembali ke Langkah sebelumnya sampai iterasi ke *t=t+1*. Proses iterasi dihentikan apabila $\left|\beta^{(t+1)} - \beta^{(t)}\right| < \varepsilon$. Hasil estimasi yang diperoleh pada iterasi terakhir adalah $\beta^{(t+1)}$.

b. Pengujian Parameter

Menurut Hosmer dan Lemeshow (2000), perlu dilakukan pengujian terhadap parameter model untuk memeriksa pengaruh variable bebas atau predictor terhadap model. Terdapat dua pengujian pada analisis regresi logistic multinomial yaitu :

1. Pengujian Parameter secara Simultan (Uji G)

   Pengujian ini dilakukan untuk mengetahui pengaruh variable bebas atau predictor terhadap model secara bersama-sama.

   - Hipotesis :

     H₀ : $\beta_1 = \beta_2 = \cdots = \beta_p = 0$ (Tidak terdapat pengaruh variable bebas terhadap model)

     H₁ : Minimal terdapat satu $\beta_k \neq 0,\ k = 1,2,\ldots,p$

   - Statistik uji :

$$G = -2\ ln\left[\frac{(\frac{n_1}{n})^{n_1}(\frac{n_2}{n})^{n_2}(\frac{n_3}{n})^{n_3}}{\prod_{i=1}^{n}\quad [\pi_1(x_i)^{y_{1i}}\pi_2(x_i)^{y_{2i}}\pi_3(x_i)^{y_{3i}}]}\right]$$

     Dengan

     $n_1 = \sum_{i=1}^{n} \quad y_{1i},\ n_2 = \sum_{i=1}^{n} \quad y_{2i},\ n_3 = \sum_{i=1}^{n} \quad y_{3i}$ dan $n = n_1 + n_2 + n_3$

     n adalah banyaknya nilai observasi

   - Kriteria Penolakan

     Tolak H₀ jika nilai $G^2 > X^2_{\alpha,db\ (banyaknya\ variabel\ bebas)}$

2. Pengujian Parameter secara Parsial (Uji Wald)

   - Hipotesis :

     H₀ : $\beta_k = 0$

     H₁ : $\beta_k \neq 0,\ k = 1,2,\ldots,p$

   - Statistik uji :

$$W_k = \left[\frac{\widehat{\beta_k}}{S\hat{E}(\widehat{\beta_k})}\right]$$

   - Kriteria Penolakan

     Tolak H₀ jika nilai $|W_k| > Z_{\alpha/2}$

c. Uji Kesesuaian Model (Goodness of Fit)

Uji kesesuaian model dilakukan untuk mengetahui model yang diperoleh sudah sesuai atau tidak.

- Hipotesis :

H$_0$: Model regresi logistic sesuai (tidak terdapat perbedaan nyata antara hasil observasi dan prediksi)

H$_1$: Model regresi logistic tidak sesuai (terdapat perbedaan nyata antara hasil observasi dan prediksi)

- Statistik uji :

$$\hat{C} = \sum_{k=1}^{g} \frac{(o_k - n_k \underline{\pi}_k)^2}{n_k \underline{\pi}_k (1 - \underline{\pi}_k)}$$

dengan : o_k : observasi pada kelompok ke-k

$\underline{\pi}_k$: rata-rata kejadian kelompok-k

n_k : banyak observasi pada kelompok ke-k

- Kriteria Penerimaan

Terima H$_0$ jika nilai $pvalue > \alpha$ atau nilai $\hat{C} \leq X^2$ (Hosmer dan Lemeshow, 2000)

SYNTAX R

Langkah-langkah analisis regresi logistik multinomial pada *software* R adalah sebagai berikut :

1. Input data

```
Data=read.csv("E:DataMultinomial.csv",sep=";",header=TRUE)

names(Data)<-c("Y", "X1", "X2", "X3")

head(Data)
```

2. Eksplorasi data

```
str(Data)
```

3. Install package

```
install.package("foreign")

install.package("nnet")

install.package("caret")

install.package("e1071")
```

4. Membentuk model multinominal

```
model<-multinom(Data2$'Keputusan'~.,data=Data)
```

```
summary(model)
```

5. Menghitung nilai Uji Wald dan P-Value

```
z<-summary(model)$coefficients/summary(model)$standard.
errorsz

p<-(1-pnorm(abs(z),0,1))*2
```

6. Melakukan uji kelayakan model atau Goodness of Fit

```
chisq.test(Data2$'Keputusan',predict(model))
chisq.test()
```

7. Melakukan interpretasi koefisien

   Odds Ratio:

```
(OR=exp(coef(model)))
```

STUDI KASUS

Disajikan sebuah data Universitas X yang terdiri dari data Kelas, Gender, Nilai Statistika, Nilai Matematika, dan Nilai Bahasa Indonesia. Variabel dependen yang digunakan yaitu Kelas dan variabel independen yaitu Gender, Nilai Statistika, Nilai Matematika, dan Nilai Bahasa. Diperoleh data sebagai berikut :

Kelas	Gender	Statistika	Matematika	Bahasa Indonesia
3	1	1.36	1.64	1.16
1	0	1.36	1.64	1.44
3	0	1.56	1.76	1.04
3	0	1.48	1.68	1.32
3	0	1.56	1.60	1.56
1	1	1.68	1.68	1.24
3	0	1.24	1.84	1.56
3	0	2.00	1.60	1.36
3	1	1.56	1.32	1.68
3	0	1.36	1.84	1.56
3	1	1.36	1.60	1.56
2	0	1.56	1.52	1.88
3	0	1.88	1.76	1.44
3	1	1.44	1.48	1.68
3	1	1.40	1.60	2.04
1	0	1.76	1.56	1.36
1	1	1.12	1.72	1.76
3	1	1.68	1.52	1.44
2	0	1.36	1.80	1.56
3	0	1.36	1.56	1.68
1	1	1.68	1.68	1.68
3	0	1.48	1.80	1.56

3	1	1.64	1.60	1.56
3	1	1.76	1.60	1.60
2	1	1.88	1.72	1.68
2	0	1.76	1.96	1.56
1	1	1.76	1.76	1.40
1	0	1.76	1.84	1.56
2	1	1.88	1.84	1.60
2	1	1.76	1.64	1.56
1	1	1.88	1.68	1.68
3	0	1.88	1.88	1.68
2	0	1.88	1.72	1.80
2	1	1.88	1.96	1.32
3	0	1.76	1.60	2.00
3	0	1.68	1.56	2.24
3	0	2.00	1.96	1.76
2	1	1.64	1.68	1.36
1	0	1.68	1.72	2.00
2	0	1.88	1.64	1.68
1	0	1.68	2.28	1.88
2	0	2.00	2.00	1.44
2	1	1.80	1.76	1.36
3	0	2.00	1.56	1.96
2	1	1.56	2.08	1.76
3	1	1.68	1.64	1.68
3	1	1.88	1.56	1.88
1	1	1.56	1.56	1.88
3	1	1.88	2.08	1.56
2	1	1.76	1.92	1.56
2	1	2.00	1.80	1.76

3	0	2.20	1.60	1.76
1	1	1.76	1.84	1.88
2	0	1.80	2.00	1.24
2	1	1.76	1.72	1.76
1	0	1.68	1.72	2.00
2	0	1.88	2.16	1.68
1	1	1.44	1.68	2.00
3	1	1.88	1.80	1.36
1	0	1.76	2.16	2.32
3	1	1.68	1.88	1.88
2	0	2.00	1.80	2.20
2	1	1.88	1.72	1.92
3	1	1.72	1.60	2.00
3	0	1.92	2.08	1.76
3	1	1.88	2.08	1.92
1	1	1.68	1.80	2.00
2	1	2.08	1.96	1.76
2	1	2.00	2.12	1.56
1	1	1.56	2.16	2.00
2	1	1.68	2.00	2.00
2	1	1.88	1.92	1.76
2	0	2.08	2.04	2.12
1	0	2.28	1.64	1.88
3	1	1.68	2.04	1.88
1	1	1.88	1.84	2.00
3	1	1.84	2.20	1.76
1	1	2.00	1.84	2.00
1	0	1.76	2.44	2.00
2	0	2.00	1.80	2.32

3	0	1.44	2.16	2.44
2	1	1.56	2.16	2.12
2	0	1.88	2.04	2.00
1	0	2.08	1.96	2.20
3	1	2.00	2.24	1.88
2	1	2.00	2.12	2.12
1	1	1.88	1.96	2.12
1	0	2.20	2.28	2.12
1	0	2.00	1.68	2.12
3	0	2.20	1.84	2.32
3	1	2.00	1.80	2.32
1	1	1.72	2.20	2.20
3	0	2.52	1.88	2.12
2	1	2.28	2.00	2.04
3	1	2.00	1.64	2.20
1	1	2.20	2.08	1.68
2	0	1.88	2.28	2.12
3	1	2.00	2.12	2.20
3	0	2.40	2.04	2.12
1	1	2.28	2.28	2.20
2	0	1.76	2.04	2.52
3	1	2.28	1.60	2.44
3	1	2.08	2.00	2.16
3	0	1.68	2.28	2.88
1	0	2.52	1.40	2.64
3	0	1.88	2.28	2.32
2	0	2.20	2.16	1.96
2	1	2.52	2.16	2.00
2	1	2.28	2.16	2.00

2	1	2.40	1.96	2.00
3	1	2.28	2.04	2.12
2	1	2.08	2.32	2.12
2	1	1.88	2.40	2.00
2	1	1.88	2.04	2.52
2	0	1.88	2.52	2.12
2	0	2.08	2.00	2.44
2	1	2.00	2.12	2.44
2	1	2.08	2.12	2.44
2	1	2.20	2.20	2.00
2	1	2.20	2.24	2.32
2	0	2.52	1.96	2.20
2	0	2.20	2.44	2.44
1	0	2.52	1.96	2.64
2	1	2.60	2.28	1.84
2	0	2.08	2.64	1.88
1	1	1.88	2.24	2.64
2	0	2.40	2.40	2.20
2	0	2.16	2.28	2.20
2	0	2.52	2.36	2.60
2	0	2.40	2.32	2.32
2	0	2.40	2.28	2.44
2	1	2.08	2.56	2.32
1	0	2.28	2.40	2.32
1	0	2.52	2.16	2.32
2	0	2.12	2.44	2.28
1	1	2.20	2.32	2.32
1	1	2.28	2.24	2.32
1	1	2.08	2.40	2.24

2	0	2.52	2.20	2.32
2	0	2.52	2.28	2.20
2	0	2.20	2.48	2.32
2	1	2.52	2.24	2.20
2	0	2.28	2.04	2.52
3	1	2.72	2.24	2.00
1	1	2.00	2.44	2.52
2	1	2.28	2.52	2.20
2	1	2.60	2.44	2.32
2	1	2.40	2.68	2.00
2	0	2.52	1.92	2.52
2	0	1.88	2.44	2.76
1	1	2.08	2.28	2.52
1	0	2.60	1.92	2.52
2	1	2.52	2.60	2.16
2	1	2.40	2.48	2.44
2	1	2.72	2.44	2.20
3	1	2.72	2.12	2.52
1	0	2.40	2.32	2.44
2	0	2.20	2.56	2.52
2	1	2.52	2.60	2.12
1	0	2.72	2.24	2.52
2	1	2.64	2.68	2.44
2	1	2.72	2.56	2.76
2	0	2.60	2.52	2.64
2	0	2.60	2.72	2.64
2	0	2.72	3.00	2.64
2	1	2.84	2.76	2.32
2	1	2.72	2.60	2.76

2	0	2.72	2.84	2.64
2	0	2.92	2.84	2.44
2	1	3.04	2.40	2.68
2	1	2.52	2.84	2.76
3	0	2.52	3.00	2.88
2	0	2.92	2.84	2.52
2	1	2.84	2.88	2.64
2	0	2.92	2.92	2.76

Syntax R

```
rm(list=ls())
library(foreign)
library(nnet)
library(ggplot2)
library(reshape2)
library(caret)
library(performance)
library(pscl)  #penggunaan pr2

data <- multinomial
str(data)
data$Kelas <- relevel(data$Kelas, ref = "1")
data$Kelas<- as.factor(data$Kelas)
m <- multinom(Kelas ~ ., data = data)
summary(m)
z <- summary(m)$coefficients/summary(m)$standard.errors
z
p <- (1 - pnorm(abs(z), 0, 1)) * 2
p
multicollinearity(m)
model_performance(m,metrics = "all")
pR2(m)
pkel = predict(m,data[,-1])
```

```
tableses<- table(data$Kelas, pkel)
confusionMatrix(tableses)

#Melakukan interpretasi koefisien
(OR=exp(coef(m)))
#Uji Goodness of Fit
chisq.test(data$'Kelas',predict(m))
```

Interpretasi

MODEL AWAL

```
Call:
multinom(formula = Kelas ~ ., data = data)

Coefficients:
  (Intercept)           Gender  Statistika  Matematika  `Bahasa
Indonesia`
2    -3.953925 -0.05968374 1.46593497   2.5206817          -
1.7233962
3     3.889219 -0.14211024 0.05463102 -0.9656046          -
0.9661121

Std. Errors:
  (Intercept)   Gender Statistika Matematika `Bahasa Indonesia`
2    1.493740 0.4138783  0.7334823  0.8522933         0.7272094
3    1.617364 0.4390754  0.7937716  0.9031826         0.7421060

Residual Deviance: 313.9794
AIC: 333.9794
```

$$\pi_1 = \frac{1}{1 + e^{-3.953925 - 0.05968374 Gender1 + 1.46593497 Stat + 2.5206817 Mat - 1.7233962 B.Ind} + e^{3.889219 - 0.14211024 Gender2 +}}$$

$$\pi_2 = \frac{e^{-3.953925 - 0.05968374 Gender1 + 1.46593497 Stat + 2.5206817 Mat - 1.7233962 B.}}{1 + e^{-3.953925 - 0.05968374 Gender1 + 1.46593497 Stat + 2.5206817 Mat - 1.7233962 B.Ind} + e^{3.889219 - 0.14211024 Gender2 +}}$$

$$\pi_3 = \frac{e^{3.889219-0.14211024Gender2+0.05463102Stat-0.9656046Mat-0.9661121B.In}}{1+e^{-3.953925\ -0.05968374Gender1+1.46593497Stat+2.5206817Mat-1.7233962B.Ind}+e^{3.889219-0.14211024Gender2+}}$$

UJI RASIO LIKELIHOOD

- Hipotesis

  $H_0 : \beta_{1p} = \beta_{2p} = \ldots = \beta_{(r-1)p} = 0$

  H_1: salah satu dari $\beta_{jk} \neq 0$ dengan j = 1, 2 ; k = 1, 2
- Taraf signifikansi

  α = 5%
- Statistik Uji

$$G = -2\ ln\ ln \left(\frac{likelihood\ tanpa\ variabel\ bebas}{likelihood\ dengan\ variabel\ bebas}\right)$$

$$= 192{,}257151 - 156{,}989703 = 35{,}267448$$

- Kriteria Uji

  H_0 ditolak jika $G > \chi^2_{(\alpha,v)}$ dimana $v = (c-1)\sum_{k=1}^{p} \quad (J_k - 1) = (3-1)[(2-1)] = 2$ dan $\chi^2_{(0{,}05;2)} = 5.991465$
- Keputusan

  H_0 ditolak karena $G = 35{,}267448 > \chi^2_{(0{,}05;2)}(5.991465)$
- Kesimpulan

  Pada taraf signifikansi 5%, H_0 ditolak, sehingga dapat disimpulkan bahwa model signifikan dan secara bersama-sama variabel bebas mempengaruhi model

UJI WALD

- Hipotesis :

  $H_0 : \beta_k = 0$

  $H_1 : \beta_k \neq 0,\ k = 1,2,\ldots,p$
- Statistik uji :

$$\mathrm{W}j = \left\{\frac{\hat{\beta}_j}{Se(\hat{\beta}_j)}\right\}$$

- Kriteria Penolakan

  Tolak H_0 jika nilai $|W_k| > Z_{\alpha/2}$

- Keputusan

  Model 2

  H_0 diterima untuk variabel Gender karena nilai $p - value$ $(0{,}8853378) > \alpha(0{,}05)$

  H_0 ditolak untuk variabel Statistika karena nilai $p - value(0{,}04565207) > \alpha(0{,}05)$

  H_0 ditolak untuk variabel Statistika karena nilai $p - value(0{,}00310116) > \alpha(0{,}05)$

  H_0 ditolak untuk variabel Statistika karena nilai $p - value(0{,}01779405 > \alpha(0{,}05)$

  Model 3

  H_0 diterima untuk variabel Gender karena nilai $p - value$ $(0{,}7461970) > \alpha(0{,}05)$

  H_0 diterima untuk variabel Statistika karena nilai $p - value(0{,}9451923) > \alpha(0{,}05)$

  H_0 diterima untuk variabel Statistika karena nilai $p - value(0{,}28501856) > \alpha(0{,}05)$

  H_0 diterima untuk variabel Statistika karena nilai $p - value(0{,}19296701 > \alpha(0{,}05)$

- Kesimpulan

  Pada model 2 variabel gender tidak signifikan terhadap model dan variabel statistika, matematika, dan Bahasa Indonesia signifikan terhadap model. Sedangkan pada model 3 variabel gender, statistika, matematika, dan Bahasa Indonesia tidak signifikan terhadap model. Sehingga dapat disimpulkan bahwa variabel statistika, matematika, dan Bahasa Indonesia tetap dimasukkan dalam model.

UJI GOODNESS OF FIT

- Hipotesis :

  H_0 : Model regresi logistic sesuai (tidak terdapat perbedaan nyata antara hasil observasi dan prediksi)

  H_1 : Model regresi logistic tidak sesuai (terdapat perbedaan nyata antara hasil observasi dan prediksi)

- Statistik uji :

```
> chisq.test(data$'Kelas',predict(m))


        Pearson's Chi-squared test


data:  data$Kelas and predict(m)
X-squared = 32.481, df = 4, p-value = 1.525e-06
```

- Keputusan

  H_0 ditolak karena $G = 32.481 > \chi^2_{(0,05;,4)}(5.991465)$

- Kesimpulan

  Pada taraf signifikansi 5%, H_0 ditolak, sehingga dapat disimpulkan bahwa model terdapat perbedaan nyata antara hasil observasi dan prediksi)

ODDS RATIO

```
> (OR=exp(coef(m)))
  (Intercept)    Gender Statistika Matematika `Bahasa
Indonesia`
2  0.01917927 0.9420624   4.331591 12.4370716
0.1784590
3 48.87267899 0.8675256   1.056151  0.3807529
0.3805597
```

Variabel Gender

Artinya bahwa peluang mahasiswa masuk kelas kedua adalah sebesar 0.9420624 kali lebih kecil dari peluang kelas pertama dan ketiga.

Variabel Statistika

Artinya bahwa semakin baik nilai statistika maka peluang mahasiswa masuk kelas kedua adalah sebesar 4.331591 kali lebih besar dari peluang kelas pertama dan ketiga.

Variabel Matematika

Artinya bahwa semakin baik nilai matematika maka peluang mahasiswa masuk kelas kedua adalah sebesar 12.4370716 kali lebih besar dari peluang kelas pertama dan ketiga.

Variabel Bahasa Indonesia

Artinya bahwa semakin baik nilai Bahasa Indonesia maka peluang mahasiswa masuk kelas kedua adalah sebesar 0.1784590 kali lebih kecil dari peluang kelas pertama dan ketiga.