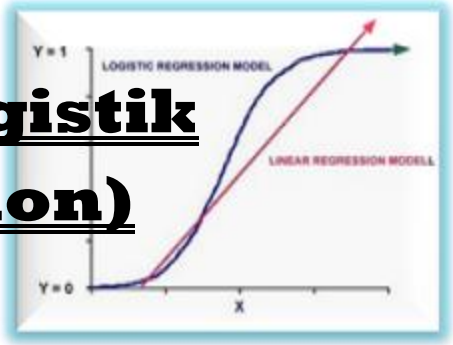


Analisis Regresi Logistik (Logistic Regression)

A. Teori Ringkas

Seringkali kita menemui situasi dimana variabel respon yang dimiliki berisifat kategorik, lebih khusus lagi bersifat dikotomi (memiliki dua buah nilai). Ini yang kita kenal sebagai regresi logistik biner (*binary logistic regression*).

Regresi logistik (*logistic regression*) sebenarnya sama dengan analisis regresi berganda, hanya variabel terikatnya merupakan variabel dummy (0 dan 1). Sebagai contoh, pengaruh beberapa rasio keuangan terhadap keterlambatan penyampaian laporan keuangan. Maka variabel terikatnya adalah 0 jika terlambat dan 1 jika tidak terlambat (tepat). Regresi logistik tidak memerlukan asumsi normalitas, meskipun *screening* data *outliers* tetap dapat dilakukan.

Model regresi logistik menggunakan transformasi logit. Pada model ini, yang diregrsikan adalah peluang variabel respon sama dengan 1. Model umum regresi logistik biner adalah :

$$P(Y=1) = \pi = \frac{e^{\beta_0+\beta_1 x_1+\ldots+\beta_k x_k}}{1+e^{\beta_0+\beta_1 x_1+\ldots+\beta_k x_k}}$$

Untuk memeriksa peranan variabel-variabel penielas (x) dalam model, dilakukan penguiian terhadap parameter model (β). Penguiian secara simultan dilakukan menggunakan uii *G*, sedangkan secara parsial menggunakan uji *Wald*. Statistik uji G adalah uji rasio kemungkinan (*likelihood ratio test*) yang digunakan untuk menguji peranan variabel penjelas di dalam model secara bersama-sama. Rumus umum uii G untuk menguii hipotesis :

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

H_1 : minimal ada satu β_1 yang tidak sama dengan 0

adalah :

$$G = 2\ln\left[\frac{likelihood \text{ tanpa peubah bebas}}{likelihood \text{ tanpa peubah bebas}}\right]$$

Created by : OLIS@STK08
E -Mail : olis.statistika08@gmail.com

Statistik G ini secara teoritis mengikuti sebaran χ^2 dengan derajat bebas k. criteria keputusan yang diambil yaitu menolak H_0 bila $G_{hitung} > \chi^2_{\alpha(k)}$

Sementara itu, uji Wald digunakan untuk menguji parameter β_i secara parsial. Hipotesis yang diuji adalah :

$H_0 : \beta_1 = 0$

$H_1 : \beta_1 \neq 0$

Formula untuk statistik *Wald* adalah :

$$W = \frac{\hat{\beta}_i}{SE\left(\hat{\beta}_i\right)}$$

Secara teori statistik W ini mengikuti sebaran normal baku jika H_0 benar. Kriteria keputusan adalah H_0 ditolak jika $|W| > Z_{\alpha/2}$.

Goodness of fit adalah suatu alat statistik yang digunakan untuk pengujian kebaikan atau kecocokan model yang dipostulatkan dibandingkan dengan data yang diamati. Pengepasan dikatakan baik jika ada kesesuaian antara model dengan data yang diamati. Metode yang digunakan untuk *goodness of fit* pada data kaegorik adalah metode *Pearson*, *Deviance* dan *Hosmer-Lemeshow*, dengan hipotesis uji:

H_0 : Model yang dipostutatkan layak

H_1 : Model yang dipostulatkan tidak layak.

Interpretasi regresi logistik menggunakan *odds ratio* (ψ) atau kemungkinan. yang menielaskan berapa lipat kenaikan atau penurunan peluang $Y = 1$, jika nilai variabel penjelas (X) berubah sebesar nilai tertentu. Nilai *odds ratio* selalu positif. Hubungan antara *odds ratio* (ψ) dan parameter model (β) adalah :

$$\psi_{ab} = e^{\beta(b-a)}$$

jika ψ_{ab} adalah *odds ratio* antara objek dengan nilai $X=b$ terhadap objek dengan nilai $X=a$.

Jika $\beta > 0$ maka odds-ratio akan $\psi > 1$

Jika $\beta = 0$ maka odds-ratio akan $\psi = 1$

Jika $\beta < 0$ maka odds-ratio akan $\psi < 1$

Andaikan diperoleh β bentuk pendapatan per bulan (satuan juta rupiah) sebesar 0,2, maka *odds ratio* antara orang yang berpenghasilan 2 juta terhadap orang yang berpenghasilan 1 juta adalah :

$$\psi_{ab} = e^{0,2(2-1)} = 1,35$$

Artinya, orang yang berpenghasilan 2 juta rupiah memiliki kecenderungan 1,35 kali lipat lebih berhasil daripada yang berpenghasilan 1 juta.

B. Contoh Kasus

Berikut ini diberikan contoh penggunaaan regresi logistik untuk mengkaji hubungan antara kebiasaan mendengkur saat tidur malam dengan status memiliki penyakit jantung. Untuk maksud tersebut diperoleh data:

KEBIASAAN MENDENGKUR	PENYAKIT JANTUNG		TOTAL
	YA	TIDAK	
Tidak pernah	24	1355	1379
Kadang-kadang	35	603	638
Hamper setiap malam	21	192	213
Setiap malam	30	224	254

Skor 0, 2, 4, 5 digunakan untuk mengkategorikan kebiasaan mendengkur (X) secara berturut-turut yaitu tidak pernah, kadang-kadang, hampir setiap malam, setiap malam. Variabel respon Y merupakan status kepemilikan penyakit jantung berskala biner dengan Y = 0 iika tidak dan Y = 1 jika ya.

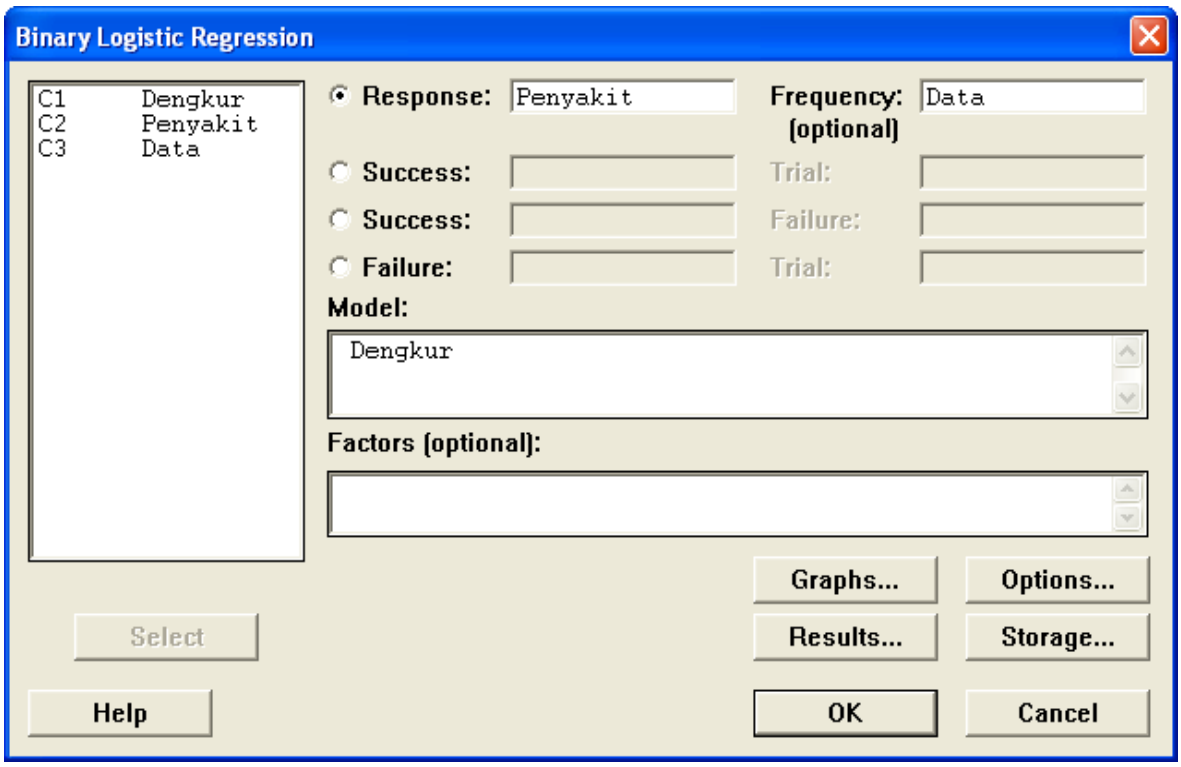
C. Prosedur Praktikum

1. Input data kebiasaan mendengkur saat tidur malam pada *worksheet Minitab* dengan format seperti di bawah ini.

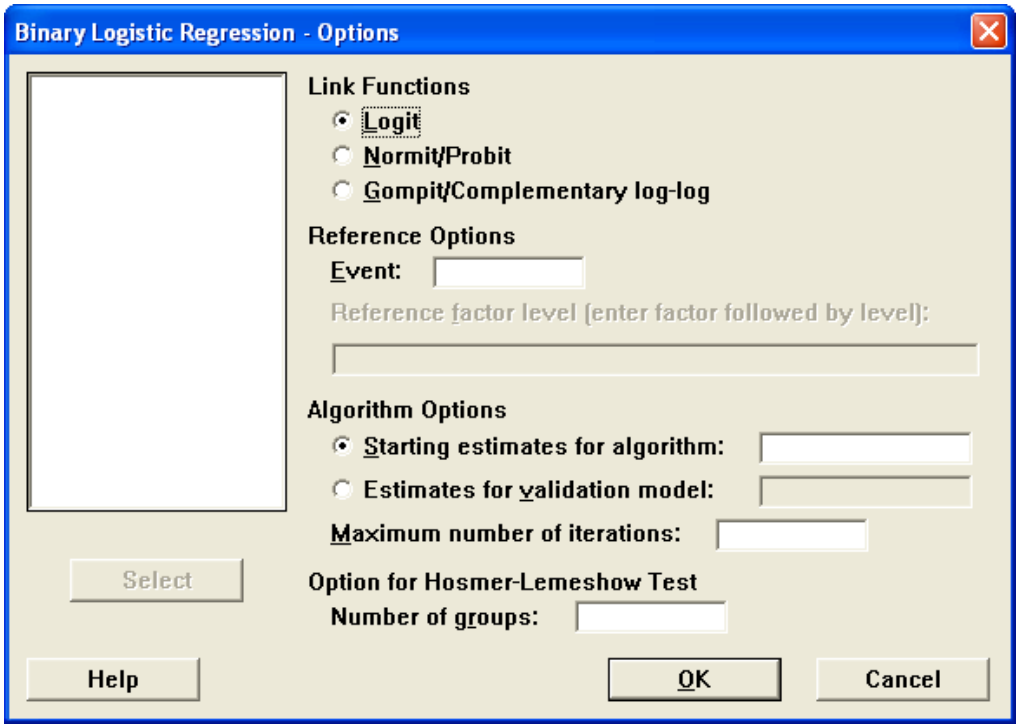
Worksheet 1 ***

↓	C1	C2	C3
	Dengkur	Penyakit	Data
1	0	1	24
2	2	1	35
3	4	1	21
4	5	1	30
5	0	0	1355
6	2	0	603
7	4	0	192
8	5	0	224
9			
10			

2. Lakukan analisis data dengan mengklik *Stat* → *Regression* → *Binary Logistic* → *Regression* sehingga muncul kotak dialog berikut.



3. Pada kotak dialog ini, masukkan variabel respon yaitu penyakit ke *Response*, lalu masukkan Data ke *Frequency*. Sedangkan pada *Model* masukkan variabel penjelas yaitu Dengkur.
4. Kemudian klik *Options* untuk memilih *link function* yang digunakan, seperti tampak pada gambar di bawah, kemudian pilih *Logit*.



5. Setanjutnya, klik **OK**.
6. Lalu **OK** lagi

D. Hasil Analisis dan Pembahasan

❖ Hasil Output :

Binary Logistic Regression: Penyakit versus Dengkur

```
Link Function: Logit

Response Information

Variable  Value   Count
Penyakit  1         110   (Event)
          0        2374
          Total    2484

Frequency: Data


Logistic Regression Table

                                                  Odds     95% CI
Predictor       Coef     SE Coef        Z       P  Ratio  Lower  Upper
Constant    -3,86625    0,166214   -23,26   0,000
Dengkur     0,397337   0,0500106     7,95   0,000   1,49   1,35   1,64


Log-Likelihood = -418,866
Test that all slopes are zero: G = 63,096, DF = 1, P-Value = 0,000
```

1

2

```
Goodness-of-Fit Tests

Method            Chi-Square  DF       P
Pearson              2,87432   2   0,238
Deviance             2,80891   2   0,246
Hosmer-Lemeshow      2,47090   1   0,116


Table of Observed and Expected Frequencies:
(See Hosmer-Lemeshow Test for the Pearson Chi-Square Statistic)

Value            1                    0
Group   Observed   Expected   Observed   Expected   Total
    1         24      28,3       1355     1350,7    1379
    2         35      28,3        603      609,7     638
    3         51      53,5        416      413,5     467


Measures of Association:
(Between the Response Variable and Predicted Probabilities)

Pairs        Number  Percent  Summary Measures
Concordant   153043     58,6  Somers' D                0,42
Discordant    43720     16,7  Goodman-Kruskal Gamma    0,56
Ties          64377     24,7  Kendall's Tau-a          0,04
Total        261140    100,0
```

3

4

5

- **Hipotesis :**

  Output (3) :

  $H_0 : \beta_X = 0$

  $H_1 : \beta_X \neq 0$

- **Interpretasi :**

Output tersebut diawali dengan *Response Information*. Output (1) ini menampilkan informasi yang berkaitan dengan kategori pada variabel respon, dimana untuk kasus ini variabel Perryakit Jantung yang memiliki dua kategori yaitu Ya sebanyak 110 orang dan kategori Tidak sebanyak 2374 orang. Pada kategori Ya terdapat keterangan (*event*) yang menandakan bahwa kategori ini dijadikan acuan dalam analisis regresi ini.

Output (2) adalah merupakan tampilan dari *Logistic Regression Table*. Bagian ini memuat informasi dugaan koefisien model, besarnya galat baku dugaan ini, nilai Z_{hitung} dan nilai p beserta nilai *odds ratio* dan selang kepercayaan 95% untuk *odds ratio* ini. Dari output bagian ini terlihat bahwa untuk variabel dengkur (Z_{hitung} = 7,95 dan nilai p = 0,000) memiliki nilai p yang kurang dari 0,05 yang menandakan bahwa parameter koefisien model tidak sama dengan nol pada α = 5%. Koefisien 0,39734 pada variabel

dengkur merupakan dugaan besarnya perubahan pada $\log \frac{P(\text{Ya})}{P(\text{Tidak})}$ ketika seseorang mendengkur pada saat tidur malam dibandingkan dengan yang tidak mendengkur.

Selaniutnya masih pada output (2) ditampilkan pula besarnya nilai Log-Likelihood beserta statistik G. Statistik G ini digunakan untuk menguji H_0 bahwa semua koefisien yang berkaitan dengan variabel X adalah sama dengan nol, lawan hipotesis H_1 bahwa tidak semua koefisien sama dengan nol. Pada kasus ini, nilai statistik G sebesar 63,096 dengan nilai *P* sebesar 0,000 yang menandakan bahwa terdapat cukup bukti untuk menyatakan bahwa paling tidak satu dari koefisien yang ada tidak sama dengan nol (pada α = 5%).

Catatan : Karena variabel penietas yang digunakan hanya satu, yaitu variabel dengkur, maka hasilnya akan sama dengan uji *Wald*. Nilai *odds-ratio* dari output *Minitab* diatas tidak bermakna karena *Minitab* hanya mampu memberikan hasil *odds-ratio* yang bermakna untuk variabel bebas yang mempunyai 2 kategori saja.

Hasil output (3) menampilkan ***Goodness of Fit Tests*** atau penguiian kebaikan model. Untuk penguiian ini ditampilkan tiga macam uji yaitu *Pearson*, *Deviance* dan *Hosmer-Lemeshow*. Ketiga pengujian ini menghasilkan nilai *P* yang berkisar antara 0,116 dan 0,245 yang lebih besar daripada α = 5%. Sehingga kita simpulkan bahwa tidak terdapat cukup bukti untuk menyatakan bahwa model yang dihasilkan tidak cukup baik dalam menerangkan data.

Hasil output (4) adalah *Table of Observed and Expected Frequencies* yang merupakan nilai amatan dan nilai harapan frekuensi atau setiap sel. Kemudian hasil yang terakhir pada output (5) yaitu *Measures of Association* untuk melihat sejauh mana keragaman variabel respon *Y* dapat dijelaskan oleh variabel penjelas *X* dengan melihat nilai persen dari Concordant. Pada kasus ini diperoleh nilainya sebesar 58,6 % yang artinya hanya sebesar 58,6 % keragaman variabel respon (status kepemilikan penyakit jantung) yang mampu dijelaskan oleh variabel penjelas (kebiasaan mendengkur).