

# Pemodelan Faktor-Faktor yang Mempengaruhi Jenis Kanker Payudara Menggunakan Regresi Logistik Biner (Kasus : Pasien Penderita Kanker Payudara di RSUP Dr. Wahidin Sudirohusodo tahun 2016)

Titi Kurnianti HR\*, M. Nadjib Bustan, & Ruliana

*Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Makassar, Indonesia*

**Keywords:** Breast Cancer, Logistic Regression, Binary Logistic Regression.

## **Abstract:**

Logistic regression is a method of statistical analysis that is applied to model the dependent variable which has two or more categories with one or more independent variables. Binary Logistic Regression is a statistical analysis that is used to analyze the relationship between one or more independent variables with variable binary or dichotomous responses. The free variables in logistic regression can be either categorical scale or continuous scale variables while the response variables are categorical scale variables. Binary Logistic Regression can be applied to health cases, especially in this study, namely breast cancer. In accordance with the description above, the author intends to study and conduct research on Modeling Factors Affecting Types of Breast Cancer Using Binary Logistic Regression (Case: Patients with Breast Cancer Patients at Dr. Wahidin Sudirohusodo Central General Hospital). From the results of the analysis it was found that the explanatory variables that significantly affected the type of cancer malignancy in patients with breast cancer were Chemotherapy variables ( $X_2$ ) and Metastase variables ( $X_5$ ), each of which had odds ratio values of 0.17 and 6.16.

## **1. Pendahuluan**

Regresi klasik adalah metode yang digunakan untuk mengetahui adanya pengaruh antara satu variabel bebas terhadap satu variabel terikat. Dengan demikian, berbagai kemungkinan kombinasi yang bisa terjadi, misalnya variabel bebas dan terikat keduanya kuantitatif, keduanya kualitatif atau salah satunya kuantitatif (Tiro, 2011). Apabila variabel terikatnya berskala kategorik maka akan ada beberapa permasalahan yang muncul yang tidak memungkinkan untuk tetap menggunakan regresi klasik sehingga harus menggunakan Regresi Logistik.

Menurut Hosmer dan Lemeshow (2000) regresi logistik adalah suatu metode analisis statistik yang diterapkan untuk memodelkan variabel dependen yang memiliki dua kategori atau lebih dengan satu atau lebih variabel independen. Regresi logistik tidak memerlukan asumsi normalitas, heteroskedasitas, dan autokorelasi. Untuk asumsi multikolinearitas karena hanya melibatkan peubah-peubah bebas, maka masih perlu dilakukan pengujian untuk pengujian multikolinearitas dapat digunakan pengujian *goodness of fit*, kemudian dilanjutkan dengan pengujian hipotesis guna melihat peubah-peubah bebas mana saja yang signifikan sehingga dapat tetap digunakan dalam penelitian. Selanjutnya, diantara peubah-peubah bebas yang signifikan, dapat dibentuk matriks korelasi, dan apabila tidak terdapat peubah-peubah bebas yang saling memiliki korelasi yang tinggi, maka dapat disimpulkan bahwa tidak terdapat gangguan multikolinearitas pada model penelitian (Hosmer dan Lemeshow, 2000).

\* Corresponding author.

*E-mail address:* titikurnianti14@gmail.com



Regresi Logistik biner merupakan suatu analisis statistika yang digunakan untuk menganalisis hubungan antara satu atau lebih peubah bebas dengan peubah respon yang bersifat biner atau *dichotomous*. Peubah bebas pada regresi logistik dapat berupa peubah skala kategorik maupun peubah yang skala kontinu sedangkan peubah respon berupa peubah berskala kategorik. Regresi logistik biner dapat diterapkan pada kasus kesehatan, khususnya pada penelitian ini yaitu mengenai kanker payudara.

Kanker payudara sudah menjadi suatu hal yang sangat menakutkan bagi kaum wanita baik di Indonesia maupun di seluruh dunia. Terutama wanita yang berumur antara 30-40 tahun, tingkat pasien kanker payudara meningkat secara drastis dari tahun ke tahun. Kasus kanker payudara sendiri masih berada di urutan pertama dibanding kanker rahim (serviks) yang terus mengancam kaum perempuan Makassar. Dari data Dinas Kesehatan Pemprov Sulsel tercatat 170 kasus kanker payudara di kota Makassar dan di Sulsel tercatat sekitar 317 kasus perempuan yang terdeteksi mengalami sakit kanker payudara (Tempo, 2015). Tanda awal dari kanker payudara biasanya dengan terdapatnya benjolan pada payudara yang pada mulanya hanya kecil kemudian semakin lama semakin membesar, dan ketika ditekan tidak terasa nyeri.

## 2. Tinjauan Pustaka

### 2.1. Regresi logistik

Regresi logistik biner digunakan untuk menganalisa suatu hubungan antara beberapa peubah bebas dengan suatu peubah terikat dikotomi atau biner. Dikatakan biner atau dikotomi adalah karena memiliki dua kategori yaitu 0 yang menyatakan ketidakterdapatnya suatu karakteristik dan 1 yang menyatakan keberadaan suatu karakteristik. Peubah respon ( $y$ ) mengikuti distribusi Bernoulli dengan fungsi probabilitas sebagai berikut (Agresti, 1990):

$$f(Y = y) = \pi(x)^y(1 - \pi(x))^{1-y}, \quad y = 0,1 \quad (2.1)$$

Kemudian Bentuk umum dari model peluang regresi logistik biner adalah sebagai berikut :

$$\pi(x) = \frac{\exp(g(x))}{1 + \exp(g(x))} \quad (2.2)$$

### 2.2. Pendugaan Parameter

Untuk model regresi logistik yaitu menggunakan *maximum likelihood*. Pendugaan parameternya diperoleh dengan cara memaksimalkan fungsi *likelihood*. Karena  $Y$  bersifat biner, maka distribusi *Bernoulli* dapat digunakan untuk menyatakan distribusi  $Y$ . oleh karena itu fungsi *likelihood* dinyatakan sebagai berikut, (Hosmer,2000).

$$l(\beta) = \prod_{i=1}^n \pi(x_i)^{y_i}(1 - \pi(x_i))^{1-y_i} \quad (2.3)$$

Dimana:

$$i = 1, 2, \dots, n$$

$y_i$  =respon pada pengamatan ke- $i$

$\pi(x)$  =peluang kejadian ke-  $i$

Fungsi *loglikelihood* [ $L(\beta)$ ], nilai parameter  $\beta$  yang memaksimalkan  $L(\beta)$ , dapat diperoleh dngan membuat turunan pertama dan kedua  $L(\beta)$  terhadap  $\beta$

$$L'(\beta) = \sum_{i=1}^n \{x_{ij}(y_i - n_i \cdot \pi_i)\} \quad (2.4)$$

Dan

$$L''(\beta) = -\sum_{i=1}^n n_i \cdot x_{ij} \pi_i \cdot (1 - \pi_i) x_{ik} \quad (2.5)$$

Persamaan (4) dan (5) merupakan nonlinier dalam nilai  $\beta$  dan memerlukan metode khusus yang bersifat *iterative* dengan bantuan perangkat lunak pemrograman. Metode *iterative* yang digunakan untuk menduga parameter  $\beta$  adalah metode *Fisher Scoring* yang merupakan modifikasi dari metode *Newton Raphson* untuk mengatasi kekonvergenan.

### 2.3. Pengujian Signifikansi Parameter

Setelah dilakukan pendugaan parameter maka langkah selanjutnya adalah menguji signifikansi parameter. Maka yang dilakukan yaitu menggunakan uji hipotesis statistik untuk menentukan apakah peubah bebas dalam model signifikan atau berpengaruh nyata terhadap peubah terikat. Pengujian signifikansi parameter dilakukan dengan menguji peubah bebas.

#### 2.3.1. Uji keseluruhan

Uji keseluruhan atau biasa disebut uji model *chi-square*, dilakukan sebagai upaya memeriksa peranan peubah bebas dalam model secara simultan atau secara bersamaan. Pengujian ini dilakukan dengan menggunakan uji *Likelihood Ratio* dengan hipotesis sebagai berikut:

$$H_0: \beta_1 = \beta_2 = \beta_3 = \dots = \beta_p = 0$$

$$H_1: \text{paling sedikit ada satu } \beta_i \neq 0; i = 1, 2, \dots, p$$

Statistik Uji

$$G^2 = -2 \ln \left( \frac{L_1}{L_0} \right)$$

$$G^2 = -2 \ln \left( \frac{\binom{n_1}{n} n_1 \binom{n_0}{n} n_0}{\prod_{i=1}^n \hat{\pi}_i^{y_i} (1 - \hat{\pi}_i)^{(1 - y_i)}} \right) \quad (2.6)$$

Keterangan:

$n_1$  = banyaknya observasi yang berkategori 1

$n_0$  = banyaknya observasi yang berkategori 0

$n$  = banyaknya observasi ( $n_1 + n_0$ )

$L_1$  = Likelihood tanpa peubah bebas tertentu

$L_0$  = likelihood dengan peubah bebas tertentu

Statistik uji  $G^2$  mengikuti distribusi chi-square, sehingga untuk memperoleh keputusan dilakukan perbandingan dengan nilai  $X^2$  tabel.

Kriteria penolakan (tolak  $H_0$ ) jika nilai  $G^2 > X^2_{(\alpha, db)}$  (Hosmer & Lemeshow, 2000).

#### 2.3.2. Uji Parsial (Uji Wald)

Adapun hipotesis dalam uji parsial atau uji satu-satu adalah sebagai berikut:

$$H_0: \beta_i = 0; i = 1, \dots, i$$

$$H_0: \beta_i \neq 0$$

Statistik uji untuk menguji kesignifikanan tiap-tiap parameter peubah bebas (koefisien) adalah dengan menggunakan statistik Uji Wald.

$$W_i = \left[ \frac{\hat{\beta}_i}{SE(\hat{\beta}_i)} \right]; i = 1, 2, \dots, k \quad (2.7)$$

Dimana untuk statistik uji wald, kriteria penolakan  $H_0$  adalah tolak  $H_0$  jika  $|W| > Z_{1-\alpha/2}$  (Hosmer & Lemeshow, 2000).

### 2.4. Uji Kesesuaian Model

Uji ini digunakan untuk menilai apakah model sesuai atau tidak. Pada pengujian ini, uji yang digunakan adalah uji Hosmer dan Lemeshow dengan hipotesis yang digunakan:

$H_0$  : model sesuai (tidak ada perbedaan antara data sampel yang digunakan dengan model yang digunakan)  
 $H_1$  : model tidak sesuai (ada perbedaan antara data sampel yang digunakan dengan model yang digunakan)  
 Statistik Uji ;

$$\bar{c} = \sum_{p=1}^g \frac{(o_p - n'_p \bar{\pi}_p)^2}{n'_p \bar{\pi}_p (1 - \bar{\pi}_p)} \quad (2.8)$$

Keterangan :

$g$  = banyaknya grup

$n'_p$  = banyaknya observasi dalam grup ke-p

$o_p$  = jumlah nilai Y pada grup ke-p

$\bar{\pi}_p$  = rata-rata dari  $\bar{\pi}$  untuk grup ke-p

Statistik  $\bar{C}$  akan mengikuti sebaran  $X^2$  dengan derajat bebas  $g - 2$ .

Kriteria pengambilan keputusan yang diambil yaitu  $H_0$  jika  $\bar{c}_{hitung} > X^2_{\alpha(g-2)}$ .

## 2.5. Odds Rasio

Odds rasio merupakan ukuran risiko atau kecenderungan untuk mengalami kejadian 'sukses' antara suatu kategori dengan kategori lainnya. Odds rasio adalah ukuran yang memperkirakan berapa besar kecenderungan variabel-variabel terikat (Hosmer dan Lemeshow, 2000). Odds rasio merupakan ukuran untuk mengetahui risiko kecenderungan yang mengalami suatu kejadian tertentu antara kategori yang satu dengan yang lain dalam suatu variabel yang dinotasikan dengan  $\varphi$  dan didefinisikan sebagai rasio dari odds untuk  $x = 1$  dan  $x = 0$ , sehingga dapat dituliskan pada persamaan berikut:

$$\varphi = \frac{\left( \frac{e^{\beta_0 + \beta_1}}{1 + e^{\beta_0 + \beta_1}} \right) \left( \frac{1}{1 + e^{\beta_0}} \right)}{\left( \frac{e^{\beta_0}}{1 + e^{\beta_0}} \right) \left( \frac{1}{1 + e^{\beta_0 + \beta_1}} \right)} = \frac{e^{\beta_0 + \beta_1}}{e^{\beta_0}} = e^{\beta_1} \quad (2.9)$$

Menurut Hosmer dan Lemeshow (2000), sifat-sifat Odds Rasio yaitu:

- $\varphi = 1$  tidak terdapat hubungan antara kedua variabel
- $\varphi < 1$  terdapat hubungan negatif antara kedua variabel
- $\varphi > 1$  terdapat hubungan positif antara kedua variabel

## 2.6. Kanker Payudara

Kanker payudara adalah suatu penyakit *neoplasma* ganas yang merupakan suatu pertumbuhan jaringan payudara abnormal yang berbeda dengan jaringan disekitarnya. Pertumbuhan sel yang abnormal pada jaringan payudara seseorang. Payudara wanita terdiri dari *lobulus* (kelenjar susu), *duktus* (saluran susu), lemak dan jaringan ikat, pembuluh darah dan *limfe*. Sebagian besar *breast cancer* bermula pada sel-sel yang melapisi *duktus* (kanker duktal), beberapa bermula di *lobulus* (kanker lobular), serta sebagian kecil bermula di jaringan lain (Bustan, 2015). Sesuai dengan urutan kanker leher Rahim (serviks) merupakan kanker pembunuh wanita nomor satu di Indonesia dan peringkat kedua di dunia setelah *breast cancer* (Messwati, 2005).

Kanker juga merupakan salah satu penyakit tidak menular yang menjadi masalah kesehatan masyarakat, baik di dunia maupun di Indonesia. Pada *breast cancer* dimana kondisi sel telah kehilangan pengendalian dan mekanisme normalnya, sehingga mengalami pertumbuhan yang tidak normal, cepat dan tidak terkendali.

### 3. Metode Penelitian

#### 3.1. Sumber Data

Data yang digunakan dalam penelitian ini berupa data sekunder atau dokumentasi tertulis berupa data rekam medis tentang penderita Penyakit Kanker Payudara di RSUP Dr. Wahidin Sudirohusodo Makassar tahun 2016. Adapun peubah penjelas yang digunakan adalah usia pasien, kemoterapi, status pernikahan, letak kanker, dan status metastase.

#### 3.2. Defenisi Operasional Peubah

Variabel yang digunakan dalam penelitian ini terdiri atas variabel (Y) dari dua kategori yaitu kategori 0 adalah penderita jenis kanker payudara jinak dan kategori 1 adalah jenis kanker payudara ganas. Adapun variabel prediktor (X) yang digunakan dalam penelitian ini ditampilkan pada Tabel 3.1 berikut:

**Table 3.1** Peubah-Peubah Faktor Penyakit Kanker Payudara

Peubah	Keterangan	Jenis	Kategori
Y	Penderita Kanker Payudara	Kategori	1 = Ganas 0 = Jinak
X <sub>1</sub>	Usia pasien	Kategori	1 = >40 tahun 0 = ≤40 tahun
X <sub>2</sub>	Kemoterapi	Kategori	1 = Tidak 0 = Ya
X <sub>3</sub>	Status Pernikahan	Kategori	1 = Belum Menikah 0=Menikah
X <sub>4</sub>	Letak Kanker	Kategori	1 = Kiri 0 = Kanan
X <sub>5</sub>	Status Metastase	Kategori	1 = Ya 0 = Tidak

#### 3.3. Teknik Analisis Data

Adapun teknik analisis data berikut ini agar dapat mencapai tujuan penelitian adalah sebagai berikut:

1. Melakukan Eksplorasi Data untuk mengetahui gambaran umum angka pasien penderita kanker payudara di RSUP Dr. Wahidin Sudirohusodo Makassar
2. Menentukan peubah-peubah penjelas yang akan digunakan dalam pendugaan model
3. Melakukan pendugaan parameter model regresi logistik
4. Melakukan uji signifikansi parameter model regresi logistik
5. Melakukan uji kesesuaian model regresi logistik
6. Interpretasi model regresi logistik menggunakan *Odds rasio*
7. Kesimpulan

## 4. Hasil dan Pembahasan

### 4.1. Model Regresi logistik

Model regresi logistik digunakan untuk melihat hubungan antara jenis keganasan kanker dengan faktor-faktor yang berpengaruh terhadap pasien penderita kanker payudara. Kemudian dilakukan pengujian secara serentak dan parsial untuk mengetahui peubah-peubah yang signifikan. Peubah respon yang digunakan adalah jenis keganasan (Y) yang terdiri dari dua kategori, yaitu ganas dan jinak.

Model regresi logistik dibentuk menggunakan lima peubah prediktor secara bersamaan. Nilai statistik uji  $G$  yang dihasilkan pada model regresi logistik pada tabel 1.

**Tabel 1** Uji Signifikansi Parameter Secara Serentak

	$G^2$	$Db$	$\chi^2_{(a,db)}$
Model	14,38	5	11,07

Tabel 1 menunjukkan bahwa nilai statistik uji  $G$  yang dihasilkan pada model regresi logistik adalah 14.38 jika dibandingkan dengan nilai  $\chi^2_{(0,05;5)} = 11,07$  maka nilai statistik uji  $G$  lebih besar daripada  $\chi^2_{(0,05;5)}$ . Kesimpulan pengujian model regresi logistik secara serentak adalah tolak  $H_0$ , berarti minimal terdapat satu peubah penjelas yang berpengaruh terhadap jenis keganasan kanker terhadap pasien penderita kanker payudara. Dengan demikian pengujian dilanjutkan untuk mencari peubah mana yang berpengaruh signifikan secara parsial.

Penduga parameter dan hasil uji hipotesis secara parsial dari model regresi logistik ditunjukkan pada Tabel 2.

**Tabel 2** Pendugaan Parameter Model Regresi Logistik

Parameter	Dugaan	Galat baku	p-value	Odds Rasio
Intercept ( $\beta_0$ )	-1,18	1,08	0,28	
Usia Pasien ( $X_1$ )	0,25	0,88	0,78	1,29
Kemoterapi ( $X_2$ )	-1,75	0,85	0,04	0,17
Status Keluarga ( $X_3$ )	-1,74	0,98	0,07	0,17
Letak Kanker ( $X_4$ )	0,44	0,81	0,58	1,57
Status Metastase ( $X_5$ )	1,82	0,84	0,03	6,16

Tabel 2 menjelaskan bahwa peubah Kemoterapi ( $X_2$ ) dan peubah Metastase ( $X_5$ ) memiliki masing-masing nilai p-value  $x_2 = 0,04$  &  $x_5 = 0,03$  yang lebih kecil dari taraf kepercayaan ( $\alpha$ ) sebesar 5% maka dapat disimpulkan bahwa peubah penjelas yang berpengaruh nyata terhadap jenis keganasan kanker terhadap pasien penderita kanker payudara adalah persentase kemoterapi kanker payudara ( $X_2$ ) dan persentase metastase pasien kanker payudara ( $X_5$ ).

Setelah dilakukan pengujian signifikansi parameter secara simultan dan parsial maka akan didapatkan model logistik. Adapun model logistik yang terbentuk dari peubah penjelas ditunjukkan pada Persamaan 4.10

$$\pi(x) = \frac{\exp(-1,75x_2 + 1,82x_5)}{1 + \exp(-1,75x_2 + 1,82x_5)}$$

Hasil dari nilai *Odds rasio* dapat dijelaskan bahwa:

1. **Kemoterapi**  
Hubungan antara persentase kemoterapi ( $X_2$ ) dengan jenis keganasan kanker adalah negatif dengan nilai *odds rasio* sebesar 0,17 yang artinya pasien penderita kanker payudara yang tidak kemoterapi memiliki peluang untuk terkena kanker payudara ganas sebesar 0,17 kali lebih besar pada pasien yang kemoterapi.
2. **Metastase**  
Hubungan antara metastase ( $X_5$ ) dengan jenis keganasan kanker adalah positif, dengan nilai *odds rasio* sebesar 6,16 yang menjelaskan bahwa pasien penderita kanker payudara yang telah metastase memiliki peluang terkena kanker payudara ganas sebesar 6,16 kali lebih besar dari pada pasien yang belum metastase.

#### 4.2. Uji Kesesuaian Model Regresi Logistik

Uji kesesuaian model regresi logistik digunakan untuk mengetahui apakah model sesuai atau tidak terdapat perbedaan yang nyata antara hasil observasi dengan prediksi model, pada uji ini statistik uji yang digunakan adalah uji Hosmer dan Lemeshow seperti yang ditunjukkan pada Tabel 3.

**Tabel 3** Hasil Uji Kesesuaian Model

<i>Chi-Square</i>	<i>df</i>	p-value
7,81	3	0,29

Nilai *Chi-Square* pada Tabel 4.4 yang diperoleh sebesar 7,81 dan p-value = 0,29. Dapat diartikan bahwa gagal tolak  $H_0$  karena p-value >  $\alpha$  dengan  $\alpha = 5\%$ . Maka dapat disimpulkan bahwa model sesuai atau tidak terdapat perbedaan yang nyata antara data sampel pasien kanker yang digunakan dengan model data kanker yang digunakan.

### 5. Kesimpulan

Berdasarkan dari hasil penelitian yang telah dilakukan, maka dapat disimpulkan bahwa:

1. Pendugaan parameter untuk regresi logistik biner adalah:

$$\widehat{\beta}_0 = \sum_{i=1}^n Y_i - \sum_{i=1}^n \pi(x_i) = \sum_{i=1}^n \{x_i(y_i + \pi_i)\}$$

$$\widehat{\beta}_1 = \sum_{i=1}^n Y_i X_i - \sum_{i=1}^n X_i \pi(x_i) = \sum_{i=1}^n \{x_i \pi_i (1 - \pi_i) x_i\}$$

2. Model yang terbentuk untuk faktor-faktor yang mempengaruhi jenis keganasan kanker payudara adalah:

$$\pi(x) = \frac{\exp(-1,75x_2 + 1,82x_5)}{1 + \exp(-1,75x_2 + 1,82x_5)}$$

dimana;

Faktor-faktor yang berpengaruh anantara jenis keganasan kanker terhadap pasien penderita kanker payudara adalah peubah kemoterapi dan peubah metastase yang masing-masing memiliki nilai odds rasio sebesar 0,17 dan 6,16.

### References

- Agresti, A. (1990). *Categorical Data Analysis*. New York: John Willey & Sons.
- Bustan, M.N. (2015). *Manajemen penendalian penyakit tidak menular*. Jakarta: Rineka cipta.
- Hosmer, D. W., & Lemeshow, S. (2000). *Applied Logistic Regression (Second Edition)*. New York: John. Wiley & Sond, INC.
- Kleinbaum, D. G. & Klein, M. (2010). *Logistic Regression (third edition)*. New York: Springer.

Tiro, M. A. (2011). *Analisis Regresi dengan Data Kategori. Edisi ketiga*. Makassar: Andira Publisher.

Tiro, M. A., Sukarna., & Aswi. (2008). *Pengantar Teori Peluang*. Makassar: Andira Publisher