

ANALISIS FAKTOR-FAKTOR YANG MEMPERNGARUHI KANKER LEHER RAHIM DI KOTA AMBON DENGAN MENGGUNAKAN REGRESI LOGISTIK BINER

(Studi kasus: Pasien di RSUD Dr. M. Haulussy Ambon)

Salmon Notje Aulele¹, H. W. M. Patty², Trisnawaty³

^{1, 2, 3}Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Pattimura
Jalan Ir. M. Putuhena, Kampus Unpatti, Poka, Ambon, Indonesia
e-mail: ¹salmon.aulele@yahoo.com

Abstrak

Kanker serviks atau kanker leher rahim adalah jenis penyakit kanker yang terjadi pada daerah leher rahim, yaitu bagian rahim yang terletak di bawah yang membuka ke arah liang vagina. Berawal dari leher rahim, apabila telah memasuki tahap lanjut, kanker ini bisa menyebar ke organ-organ lain di seluruh tubuh. Regresi logistik biner merupakan salah satu pendekatan model matematis yang digunakan untuk menganalisis hubungan beberapa faktor dengan sebuah variabel yang bersifat dikotomis (biner). Tujuan dari penelitian ini adalah menentukan faktor-faktor yang mempengaruhi penyebab kanker leher rahim di kota Ambon dengan menggunakan regresi logistik biner. Hasil penelitian menunjukkan bahwa faktor-faktor yang mempengaruhi kanker leher rahim di kota Ambon dengan menggunakan regresi logistik biner adalah usia (X_1) dan frekuensi menikah (X_4) dengan ketepatan pengklasifikasian penderita dan non penderita kanker leher rahim berturut-turut adalah 57,14% dan 66,67%. Secara keseluruhan, model regresi logistik yang telah diperoleh dapat mengklasifikasikan responden sebesar 61,9%.

Kata Kunci: kanker leher rahim, ketepatan klasifikasi, regresi logistik biner

ANALYSIS OF THE FACTORS THAT AFFECT CERVICAL CANCER IN THE CITY OF AMBON BY USING BINARY LOGISTIC REGRESSION

(Case Study: Patient in RSUD Dr. M. Haulussy Ambon)

Abstract

Cervical cancer is a type of cancer that occurs in the cervical region that is part of the uterus that is located under the opening to the vagina. Starting from the cervix if it has entered the stage, this cancer can spread to other organs throughout the body. Binary logistic regression is one approach used a mathematical model to analyze the relationship of several factors with a variable is dichotomous (binary). The purpose of this study is to determine the factors that influence the cause of cervical cancer in the city of Ambon by using binary logistic regression. The results showed that the factors that affect cervical cancer in the city of Ambon by using binary logistic regression are age (X_1) and frequency married (X_4) with the accuracy of the classification of patients and non-patients respectively was 57.14% and 66.67%. Overall, the logistic regression model that has been obtained can classify respondents of 61.9%.

Keywords: Accuracy classification, binary logistic regression, cervical cancer

1. Pendahuluan

Kanker serviks atau kanker leher rahim adalah jenis penyakit kanker yang terjadi pada daerah leher rahim yaitu, bagian rahim yang terletak di bawah yang membuka ke arah liang vagina. Berawal dari leher rahim, apabila telah memasuki tahap lanjut, kanker ini bisa menyebar ke organ-organ lain di seluruh tubuh. Menurut para ahli kanker, kanker leher rahim adalah salah satu jenis kanker yang paling dapat dicegah dan paling dapat disembuhkan dari semua kasus kanker.

Di Indonesia diperkirakan 15.000 kasus baru kanker serviks terjadi setiap tahunnya. Sedangkan angka kematiannya di perkirakan 7.500 kasus per tahun [1]. Sementara itu di Maluku sebelum tahun 2007 tercatat dari 10.000 perempuan di Maluku hanya dua orang yang meninggal akibat penyakit kanker serviks. Setelah

itu meningkat menjadi 10 perempuan per tahun. Oleh sebab itu dibutuhkan kerja sama antar berbagai komponen dan instansi teknis terkait. Terutama tim penggerak PKK untuk memberikan pemahaman dan advokasi kesehatan kepada kaum perempuan. Hal ini bertujuan untuk menekan peningkatan kasus penyakit tersebut serta menekan penyebaran penyakit ini, sehingga jumlah penderita dapat menurun drastis dalam lima tahun mendatang [2].

Sejumlah faktor yang diduga dapat mempermudah orang terinfeksi HPV (*Human Papilloma Virus*), diantaranya yaitu melakukan hubungan seks pada usia muda, sering berganti pasangan, dan memiliki sejarah merokok. Dari faktor-faktor tersebut, untuk melihat faktor-faktor risiko yang mempengaruhi seseorang terkena kanker leher rahim maka digunakan metode regresi logistik biner. Dalam [3] dikemukakan bahwa regresi logistik biner merupakan suatu metode analisis data yang digunakan untuk mencari hubungan antara variabel respon (y) yang bersifat biner atau dikotomis dengan variabel prediktor (x) yang bersifat polikotomis.

Regresi logistik biner merupakan salah satu pendekatan model matematis yang digunakan untuk menganalisis hubungan beberapa faktor dengan sebuah variabel yang bersifat dikotomis (biner). Pada regresi logistik jika variabel responnya terdiri dari dua kategori misalnya $Y = 1$ menyatakan hasil yang diperoleh “sukses” dan $Y = 0$ menyatakan hasil yang diperoleh “gagal” maka regresi logistik tersebut menggunakan regresi logistik biner. Secara umum model probabilitas regresi logistik dengan melibatkan beberapa variabel prediktor dapat diformulasikan sebagai berikut

$$E(y|x) = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k}} \quad (1)$$

dimana $E(y|x)$ merupakan penjumlahan dari $\pi(x)$. Fungsi $\pi(x)$ merupakan fungsi nonlinear sehingga perlu dilakukan transformasi logit untuk memperoleh fungsi yang linier. Dengan demikian dapat dilihat hubungan antara variabel respon (y) dengan variabel prediktornya (x). Bentuk logit $\pi(x)$ dari dinyatakan sebagai (x), yaitu:

$$g(x) = \left(\frac{\pi(x)}{1 - \pi(x)} \right) \quad (2)$$

Selanjutnya persamaan (1) dan (2) disubstitusikan sehingga diperoleh:

$$g(x) = \ln \left(\frac{\pi(x)}{1 - \pi(x)} \right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k$$

dengan $g(x)$ merupakan fungsi hubungan dari model regresi logistik yang disebut fungsi hubungan logit.

Diasumsikan sebuah sampel berukuran n dan terdiri atas pengamatan independen berpasangan (x_i, y_i) , $i = 1, 2, \dots, n$, dengan y_i menyatakan nilai variabel respon dan x_i adalah nilai variabel prediktor untuk subjek ke- i . Pada regresi linier, metode penaksiran parameter yang lazim digunakan adalah *least squares*, dengan konsep meminimumkan jumlah kuadrat residual. Metode estimasi yang mengarah pada fungsi *least squares* dalam model regresi linier (jika residual berdistribusi normal) disebut *maximum likelihood* [3] Maksimum \ln likelihood dapat diperoleh dengan cara mendiferensialkan $L(\beta)$ terhadap β dan menyamakannya dengan nol [4].

$$\frac{\partial L(\beta)}{\partial \beta_a} = \sum_{i=1}^n y_i x_{ia} - \sum_{i=1}^n x_{ia} \hat{\pi}(x_i)$$

$$0 = \sum_{i=1}^n y_i x_{ia} - \sum_{i=1}^n x_{ia} \hat{\pi}(x_i), a = 0, 1, \dots, k$$

Dimana $\hat{\pi}(x_i) = \exp(\sum_{j=0}^k \hat{\beta}_j x_{ij}) / 1 + \exp(\sum_{j=0}^k \hat{\beta}_j x_{ij})$ menyatakan estimasi dari $\pi(x_i)$ dengan menggunakan metode *maximum likelihood*. Dari hasil penurunan pertama pada persamaan tersebut, nilai β diestimasi dengan metode numerik karena persamaannya bersifat nonlinier. Sedangkan metode untuk mengestimasi varians dan kovarians dari taksiran β dikembangkan menurut teori MLE (*Maximum Likelihood Estimation*) yang menyatakan bahwa estimasi varians dan kovariansi diperoleh dari turunan kedua fungsi \ln Likelihood.

Setelah mengestimasi koefisien-koefisien model regresi logistik, maka perlu dilakukan pengujian untuk mengetahui variabel prediktor mana yang berpengaruh secara signifikan terhadap variabel respon.

Pengujian keberartian parameter model dengan satu variabel prediktor dilakukan untuk mengetahui ada atau tidaknya hubungan antara suatu variabel prediktor dan variabel respon dengan menggunakan statistik uji Wald (W). Rasio yang dihasilkan dari persamaan akan mengikuti distribusi normal baku [3]. Sehingga untuk memperoleh keputusan, nilai statistik uji dibandingkan dengan distribusi normal baku (Z). Kriteria penolakan H_0 adalah jika $|W| > Z_{\alpha/2}$ dan jika $p\text{-value} \leq 0,05$. Sedangkan uji signifikansi parameter pada model multivariat dilakukan sebagai upaya memeriksa peranan masing-masing variabel prediktor dalam model secara bersama-sama. Dibawah H_0 , statistik uji G akan mengikuti distribusi *Chi-square* dengan derajat bebas k . Sehingga untuk memperoleh keputusan, nilai statistik uji G dibandingkan dengan nilai $\chi^2_{(\alpha,k)}$. H_0 ditolak jika $G > \chi^2_{(\alpha,k)}$.

Uji kesesuaian model (*goodness-of-fit*) dihitung berdasarkan nilai $\hat{\pi}$ yang tergantung pada susunan variabel-variabel prediktor dalam model, bukan pada jumlah variabel prediktor. Berikut ini adalah prosedur pengujian kesesuaian model.

H_0 : Model sesuai

H_1 : Model tidak sesuai

Statistik Uji [3]:

$$\hat{C}(\text{Hosmer-Lemeshow}) = \sum_{k=1}^g \frac{(o_k - n_k' \bar{\pi}_k)^2}{n_k' \bar{\pi}_k (1 - \bar{\pi}_k)}$$

dengan:

g = Banyaknya grup

n_k' = Jumlah subjek pada grup ke- k

$o_k = \sum_{j=1}^{c_k} y_j$, jumlah nilai variabel respon pada c_k kombinasi variabel prediktor

$\bar{\pi}_k = \frac{\sum_{j=1}^{c_k} m_j \hat{\pi}(x_j)}{n_k'}$, rata-rata taksiran probabilitas dimana m_j adalah banyaknya subjek pada c_k kombinasi variabel prediktor

Jika H_0 benar, maka distribusi statistik uji \hat{C} mengikuti distribusi *Chi-square* dengan derajat bebas $g-2$ [3]. Daerah penolakan H_0 adalah $\hat{C} > \chi^2_{(g-2,\alpha)}$.

Pada pemodelan regresi logistik biner, interpretasi parameter bertujuan untuk mengetahui arti dari nilai taksiran parameter pada variabel prediktor. Terdapat dua jenis variabel prediktor, yaitu variabel yang bersifat kategorik dan variabel kontinu. Estimasi koefisien dari variabel prediktor menyatakan *slope* atau nilai perubahan variabel respon untuk setiap perubahan satu unit variabel prediktor. Interpretasi meliputi penentuan hubungan fungsional antara variabel respon dan variabel prediktor serta mendefinisikan unit perubahan variabel respon yang disebabkan oleh variabel prediktor.

Untuk regresi logistik dimana variabel prediktor bersifat dikotomis, nilai dikategorikan 0 atau 1. Pada model ini, ada dua nilai $\pi(x)$ dan dua nilai $1 - \pi(x)$.

Tabel 1. Nilai-nilai $\pi(x)$ dan $1 - \pi(x)$ Untuk Variabel Prediktor Dikotomis

| Variabel respon | Variabel Prediktor | |
|-----------------|--|--|
| | x = 1 | x = 0 |
| y = 1 | $\pi(1) = \frac{\exp(\beta_0 + \beta_1)}{1 + \exp(\beta_0 + \beta_1)}$ | $\pi(0) = \frac{\exp(\beta_0)}{1 + \exp(\beta_0)}$ |
| y = 0 | $1 - \pi(1) = \frac{1}{1 + \exp(\beta_0 + \beta_1)}$ | $1 - \pi(0) = \frac{1}{1 + \exp(\beta_0)}$ |
| Total | 1.0 | 0.1 |

Nilai odds dari variabel respon diantara pengamatan dengan $x = 1$ adalah $\frac{\pi(1)}{1-\pi(1)}$, sedangkan jika $x = 0$, nilai odds $\frac{\pi(0)}{1-\pi(0)}$ (Hosmer and Lemeshow, 1989). Ln odds, sebagaimana didefinisikan sebelumnya sebagai logit, adalah

$$g(1) = \ln \left[\frac{\pi(1)}{1-\pi(1)} \right] \text{ dan } g(0) = \ln \left[\frac{\pi(0)}{1-\pi(0)} \right]$$

Odds rasio, dinotasikan ψ , didefinisikan sebagai perbandingan odds untuk $x = 1$ terhadap odds untuk $x = 0$, sebagaimana dituliskan pada Persamaan (3) [4].

$$\psi = \frac{\pi(1)/[1-\pi(1)]}{\pi(0)/[1-\pi(0)]}$$

Berdasarkan Tabel 1, nilai odds rasio adalah

$$\begin{aligned} \psi &= \frac{\left(\frac{\exp(\beta_0 + \beta_1)}{1 + \exp(\beta_0 + \beta_1)} \right) \left(\frac{1}{1 + \exp(\beta_0)} \right)}{\left(\frac{\exp(\beta_0)}{1 + \exp(\beta_0)} \right) \left(\frac{1}{1 + \exp(\beta_0 + \beta_1)} \right)} \\ &= \exp(\beta_1) \end{aligned}$$

Odds rasio adalah ukuran asosiasi yang dapat diartikan secara luas, terutama dalam epidemiologi. Dari Persamaan (3), odds rasio berarti rata-rata besarnya kecenderungan variabel respon bernilai tertentu jika $x = 1$ dibandingkan jika $x = 0$.

Evaluasi prosedur klasifikasi adalah suatu evaluasi yang melihat peluang kesalahan klasifikasi yang dilakukan oleh suatu fungsi klasifikasi (Johnson and Wichern, 1992). Ukuran yang dipakai adalah *apparent error rate* (APER). Nilai APER menyatakan nilai proporsi sampel yang salah diklasifikasikan oleh fungsi klasifikasi. Penentuan kesalahan klasifikasi dapat dilihat dari Tabel 2, jika subjek hanya diklasifikasikan menjadi dua kelompok, yakni y_1 dan y_2 .

Tabel 2. Tabel Klasifikasi

| Hasil Observasi | Taksiran | |
|-----------------|----------|----------|
| | y_1 | y_2 |
| y_1 | n_{11} | n_{12} |
| y_2 | n_{21} | n_{22} |

Keterangan :

n_{11} = Jumlah subjek dari y_1 tepat diklasifikasikan sebagai y_1

n_{12} = Jumlah subjek dari y_1 salah diklasifikasikan sebagai y_2

n_{21} = Jumlah subjek dari y_2 salah diklasifikasikan sebagai y_1

n_{22} = Jumlah subjek dari y_2 tepat diklasifikasikan sebagai y_2

$$\text{APER (\%)} = \frac{n_{12} + n_{21}}{n_{11} + n_{12} + n_{21} + n_{22}}$$

2. Metode Penelitian

Sumber data dari penelitian ini berasal dari informasi pasien yang pernah menjalani perawatan di IRNA (Instalasi Rawat Inap) RSUD dr. M. Haulussy Ambon sepanjang tahun 2015. Responden terbagi menjadi dua kelompok, yakni penderita (*case*) dan non penderita (*control*) kanker leher rahim. Data untuk kelompok *case* merupakan data sekunder yang diperoleh dari rekam medik pasien yang terdiagnosis menderita kanker leher rahim pada tahun 2015 dan *control* merupakan data primer yang diperoleh menggunakan media komunikasi berupa kuesioner. Responden yang digolongkan sebagai kelompok *control*

pada penelitian ini tidak memiliki hubungan kekerabatan dengan responden dari kelompok *case*. Hal ini dimaksudkan untuk memenuhi asumsi keindependenan antara responden kelompok *case* dan *control*. Banyaknya responden yang diambil sebagai sampel pada kelompok *control* disesuaikan dengan kelompok *case* dengan asumsi kondisi fisik awal seluruh responden adalah sama.

Model regresi logistik biner digunakan untuk menganalisa hubungan antara satu variabel respon dan beberapa variabel prediktor, dengan variabel responnya berupa data kualitatif dikotomi yaitu bernilai 1 untuk menyatakan keberadaan sebuah karakteristik dan bernilai 0 untuk menyatakan ketidakterdapatnya sebuah karakteristik. Variabel-variabel yang digunakan pada penelitian ini adalah variabel respon dan variabel prediktor. Variabel respon terdiri atas dua kategori yaitu penderita kanker leher rahim (diberi kode 1) dan non penderita kanker leher rahim (diberi kode 0).

Terdapat enam variabel prediktor yang digunakan dalam penelitian ini. Variabel prediktor pertama adalah usia (X_1). Variabel prediktor kedua adalah status (X_2), yang dikategorikan menjadi menikah (diberi kode 1) dan janda (diberi kode 0). Variabel prediktor ketiga adalah usia menikah (X_3) atau usia pertama kali menikah bagi responden yang pernah menikah lebih dari satu kali, yang dikategorikan menjadi ≤ 17 tahun (diberi kode 1) dan > 17 tahun (diberi kode 0). Variabel prediktor keempat adalah frekuensi menikah (X_4), yang dikategorikan menjadi satu kali (diberi kode 1) dan lebih dari satu kali (diberi kode 2). Variabel prediktor kelima adalah jumlah anak (X_5), dikategorikan menjadi 0 – 2 orang (diberi kode 1), 0 – 2 orang (diberi kode 2), dan >5 orang (diberi kode 3). Variabel prediktor keenam yaitu penggunaan KB (X_6), yang dikategorikan menjadi tidak (diberi kode 1) dan ya (diberi kode 0). Keenam variabel prediktor yang dipilih untuk mengetahui karakteristik responden apakah responden masih aktif dalam hubungan seksual atau tidak, pernah berganti pasangan seksual atau tidak, frekuensi persalinan yang pernah dijalani, dan apakah responden ekseptor (orang yang mengikuti program KB) atau tidak.

Jika variasi karakteristik responden telah diketahui, maka model regresi logistik univariat antara variabel respon dengan masing-masing variabel prediktor dapat ditentukan. Selanjutnya dilakukan uji signifikansi parameter dari setiap model regresi logistik univariat untuk mengetahui variabel-variabel prediktor mana yang berpengaruh secara signifikan terhadap variabel respon, sehingga model regresi logistik multivariat antara variabel respon dengan variabel-variabel prediktor yang signifikan dapat ditentukan pula. Model regresi logistik kemudian diuji kesesuaian modelnya (*goodness-of-fit*), yakni model yang seluruh variabel prediktornya signifikan.

3. Hasil dan Pembahasan

Hasil pengolahan data menunjukkan sebagian besar responden berstatus menikah, yakni 76,2% dan sisanya 23,8% responden adalah janda. Dengan kata lain, terjadi perbedaan frekuensi yang cukup mencolok antara kategori pada variabel status (X_2). Berdasarkan usia menikah, 61,9% responden menikah pada usia diatas 20 tahun. Hal ini dapat menjadi salah satu indikasi bahwa masyarakat sudah mempunyai kesadaran yang tinggi untuk tidak menikah di usia muda. Presentasi responden yang menikah satu kali lebih banyak dari pada responden yang menikah lebih dari satu kali, yakni 77,38% responden. Berdasarkan jumlah anak, responden yang memiliki jumlah anak 0-2 orang dan 3-5 orang adalah hampir sama, masing-masing sebesar 40,46 % dan 45,24% dari jumlah responden keseluruhan. Sedangkan responden yang memiliki anak lebih dari lima yaitu sebanyak 12 orang atau 14,30% responden. responden yang mengikuti program KB lebih banyak dibandingkan dengan yang tidak mengikuti program KB, yaitu 54,76% responden mengikuti program KB sedangkan sisanya 45,24% responden tidak mengikuti program KB.

Untuk mengetahui pola hubungan antara faktor-faktor risiko penyebab kanker leher rahim dengan peluang pasien terdiagnosis sebagai penderita kanker leher rahim, dapat digunakan analisis regresi logistik. Selain itu, dapat diketahui besarnya pengaruh setiap faktor dalam menentukan peluang seseorang untuk menderita kanker leher rahim. Pada penelitian ini, dilakukan regresi logistik biner dengan satu variabel prediktor dan lebih dari satu variabel prediktor.

3.1. Regresi Logistik Biner dengan Satu Variabel Prediktor

Pengujian signifikansi parameter sebagai koefisien dari variabel prediktor pada masing-masing model univariat perlu dilakukan. Hipotesis yang digunakan adalah sebagai berikut:

$H_0 : \beta_j = 0$ (variabel X_j tidak signifikan mempengaruhi Y)

$H_1 : \beta_j \neq 0$ (variabel X_j signifikan mempengaruhi Y)

Dengan menggunakan *software* SPSS, diperoleh hasil sebagai berikut:

Tabel 3. Pengujian Signifikansi Parameter Model Regresi Logistik biner Univariat

| Variabel | $\hat{\beta}$ | Wald | <i>P-value</i> | Keputusan |
|--------------------------------|---------------|-------|----------------|--------------|
| Usia (x_1) | 0,063 | 7,654 | 0,006 | Tolak H_0 |
| Status (x_2) (1) | 0,263 | 0,262 | 0,609 | Terima H_0 |
| Usia menikah (x_3) (1) | -0,202 | 0,202 | 0,653 | Terima H_0 |
| Frekuensi menikah (x_4)(1) | 1,664 | 7,282 | 0,007 | Tolak H_0 |
| Jumlah anak (x_5)(1) | -1,212 | 3,031 | 0,082 | Terima H_0 |
| Jumlah anak (x_5)(2) | 0,317 | 0,219 | 0,640 | Terima H_0 |
| Penggunaan KB (x_6)(1) | -0,192 | 0,192 | 0,661 | Terima H_0 |

Berdasarkan hasil pengujian tersebut, faktor-faktor yang berpengaruh secara individu terhadap penyakit kanker leher rahim adalah usia (X_1) dan frekuensi menikah (X_4) karena nilai *p-value* < 0,05. Sehingga untuk pemodelan regresi logistik biner multivariat, hanya menggunakan variabel X_1 dan X_4 .

3.2. Regresi Logistik Biner dengan Lebih Dari Satu Variabel Prediktor

Setelah diperoleh faktor-faktor yang berpengaruh dengan satu variabel prediktor, dilakukan regresi logistik biner dengan memasukkan variabel X_1 dan X_4 secara bersama-sama untuk memeriksa ada atau tidaknya hubungan antara variabel tersebut. Untuk mengetahui apakah parameter-parameter model telah signifikan atau tidak maka dilakukan langkah-langkah pengujian dengan hipotesis yang digunakan yaitu:

$H_0 : \beta = 0$ (variabel X_1 dan X_4 secara bersama-sama tidak mempengaruhi Y)

$H_1 : \beta \neq 0$ (minimal ada satu variabel bebas yang mempengaruhi Y)

Dengan menggunakan *software* SPSS, diperoleh statistik uji G (*likelihood ratio test*) yaitu 13,862. Dengan menggunakan α sebesar 5% maka diperoleh $\chi^2_{(\alpha; k)} = \chi^2_{(0,05; 2)} = 5,99$. Karena nilai $G = 13,862 > \chi^2_{(0,05; 2)}$ maka Tolak H_0 , sehingga minimal ada satu variabel bebas yang mempengaruhi variabel respon (Y). Dan untuk mengetahui variabel mana yang berpengaruh maka dilakukan pengujian signifikansi parameter secara parsial. Hipotesis yang digunakan adalah sebagai berikut :

$H_0 : \beta_j = 0$ (variabel X_j tidak signifikan mempengaruhi Y)

$H_1 : \beta_j \neq 0$ (variabel X_j signifikan mempengaruhi Y)

Dengan menggunakan *software* SPSS, diperoleh hasil sebagai berikut :

Tabel 4. Pengujian Signifikansi Parameter Secara Parsial

| Variabel | β | Wald | <i>P-value</i> | Keputusan |
|-----------------------------|---------|--------|----------------|-------------|
| Konstanta | -4,045 | 10,235 | 0,001 | Tolak H_0 |
| Usia X_1 | 0,052 | 4,747 | 0,029 | Tolak H_0 |
| Frekuensi menikah X_4 (1) | 1,349 | 4,461 | 0,035 | Tolak H_0 |

Berdasarkan Tabel 4, terlihat bahwa faktor-faktor yang berpengaruh secara signifikan terhadap penyakit kanker leher rahim adalah usia (X_1) dan frekuensi menikah (X_4) karena nilai *p-value* < 0,05. Sehingga model regresi logistik biner yang diperoleh adalah :

$$\hat{\pi} = \frac{\exp(-4,045 + 0,052 x_1 + 1,349 x_4)}{1 + \exp(-4,045 + 0,052 x_1 + 1,349 x_4)}$$

3.3. Uji Kesesuaian Model

Karena ada sejumlah responden yang memiliki karakteristik sama dalam hal usia dan frekuensi menikah, maka perlu dilakukan pengujian untuk mengetahui apakah model regresi logistik biner yang didapatkan telah sesuai atau tidak. Hipotesis yang digunakan adalah sebagai berikut :

H_0 : Model Sesuai

H_1 : Model Tidak Sesuai

Dengan menggunakan *software* SPSS, diperoleh statistik uji \hat{C} (Hosmer – Lemeshow) adalah 8,519 dan nilai *p-value* 0,384. Karena nilai *p-value* = 0,384 > α = 0,05 maka terima H_0 , sehingga model regresi logistik biner yang diperoleh telah sesuai. Dengan kata lain model dapat digunakan untuk menjelaskan seberapa besar peluang pasien untuk menderita kanker leher rahim berdasarkan variabel usia (X_1) dan variabel frekuensi menikah (X_4).

3.4. Interpretasi Model Regresi Logistik Biner

Setelah mendapatkan model regresi logistik biner yang sesuai dimana semua variabel prediktor didalamnya telah berpengaruh secara signifikan terhadap variabel respon, maka langkah selanjutnya adalah menginterpretasikan model tersebut. Jika model regresi logistik biner yang terbaik ditulis dalam bentuk logit, maka menjadi

$$\hat{g}(x) = -4,045 + 0,052x_1 + 1,349x_4(1)$$

Untuk menginterpretasikan model tersebut dapat digunakan nilai odds rasion. Nilai odds rasio yang diperoleh disajikan dalam tabel sebagai berikut:

Tabel 5. Nilai Odds Rasio

| Variabel | Exp ($\hat{\beta}$) |
|--------------------------------|-----------------------|
| Usia (X_1) | 1,053 |
| Frekuensi menikah (X_4)(1) | 3,853 |

Berdasarkan nilai odds rasio, dapat dijelaskan bahwa wanita yang frekuensi menikah satu kali akan berisiko menderita kanker leher rahim 3,853 kali lebih tinggi dibandingkan dengan wanita yang frekuensi menikah lebih dari satu kali.

3.5. Ketepatan Pengklasifikasian Responden

Selanjutnya menentukan ketepatan klasifikasi dari model regresi logistik biner yang dihasilkan. Dengan menggunakan *software* SPSS, diperoleh hasil sebagai berikut :

Tabel 6. Pengklasifikasian Penderita dan Non Penderita Kanker Leher Rahim

| Observasi | Taksiran | | Ketepatan klasifikasi |
|------------------------|---------------|-----------|-----------------------|
| | Non penderita | Penderita | |
| Non penderita | 28 | 14 | 66,67% |
| Penderita | 18 | 24 | 57,14% |
| Persentase keseluruhan | | | 61,90% |

Berdasarkan Tabel 6, terlihat bahwa besarnya ketepatan pengklasifikasian penderita dan non penderita kanker leher rahim untuk kelompok non penderita adalah 66,67% dan 57,14% untuk kelompok penderita. Secara keseluruhan, model regresi logistik biner yang telah diperoleh dapat mengklasifikasikan responden dengan benar sebanyak 52 orang diantara 84 responden atau 61,9%. Besarnya missklasifikasi yaitu 38,1%. Kesalahan klasifikasi dari model regresi logistik biner ini masih cukup besar. Hal tersebut dimungkinkan karena sedikitnya variabel prediktor yang masuk kedalam model.

4. Kesimpulan

Kesimpulan yang diambil dari penelitian ini yaitu :

1. Faktor-faktor yang signifikan mempengaruhi kanker leher rahim di kota Ambon adalah usia (X_1) dan frekuensi menikah (X_4). Sehingga model regresi logistik biner yang diperoleh adalah :

$$\hat{\pi} = \frac{\exp(-4,045 + 0,052 x_1 + 1,349 x_4)}{1 + \exp(-4,045 + 0,052 x_1 + 1,349 x_4)}$$

2. Besarnya ketepatan pengklasifikasian penderita dan non penderita kanker leher rahim dari model regresi logistik biner yang diperoleh adalah 61,90%. Dengan rincian, tingkat ketepatan pengklasifikasian untuk kelompok *control* adalah 66,67% dan 57,14% untuk kelompok *case*.

Untuk penelitian lanjutan, sebaiknya terlebih dahulu berkonsultasi dengan dokter ahli untuk menentukan kelompok *control* dengan lebih tepat. Selain itu, dapat pula mencoba untuk menggunakan metode regresi logistik yang melibatkan interaksi antar variabel prediktor atau metode MARS (*Multivariate Adaptive Regression Splines*) yang lebih kompleks.

Daftar Pustaka

- [1] O. Emilia, Bebas Ancaman Kanker Serviks (Fakta, Pencegahan, dan Penanganan Dini Terhadap Serangan Kanker Serviks), Yogyakarta: Media Pressindo, 2010.
- [2] A. Maluku, "10 Wanita Maluku Meninggal Karena Kanker Serviks," Antara Maluku, Ambon, 2015.
- [3] D. d. L. S. Hosmer, Applied Logistic Regression, Second Ed, Singapura: John Wiley & Sons. Inc, 1989.
- [4] A. Agresti, Categorical Data Analysis, New York: John Wiley & Sons Inc., 1990.