Model Regresi Probit Bivariat Pada Kasus Penderita HIV dan AIDS di Jawa Timur

Bella Yuliatin Puspita Sari, Farida Agustini Widjajati
Jurusan Matematika, Fakultas MIPA, Institut Teknologi Sepuluh Nopember (ITS)
Jl. Arief Rahman Hakim, Surabaya 60111 Indonesia
E-mail: farida@matematika.its.ac.id

***Abstrak*— Model regresi probit bivariat adalah model regresi probit yang menggunakan dua variabel respon yang bersifat kualitatif atau berkategori. Variabel prediktor yang digunakan adalah faktor-faktor yang mempengaruhi variabel respon. Metode estimasi parameter yang digunakan dalam probit bivariat adalah *Maximum Likelihood Estimation* (MLE). Setelah model awal regresi probit bivariat terbentuk, model diuji secara simultan untuk menguji bahwa keseluruhan variabel prediktor mempunyai pengaruh signifikan terhadap variabel respon dan uji parsial dilakukan untuk menguji signifikansi masing-masing variabel prediktor terhadap variabel respon. Setelah itu dibentuk model yang kemudian diidentifikasi kriteria kebaikan model menggunakan nilai *Akaike Information Criterion* (AIC). Data yang digunakan adalah kasus penderita HIV dan AIDS tiap kabupaten dan kota di Jawa Timur pada tahun 2013 sebagai variabel responnya dan menghasikan variabel prediktor persentase umur 25-49 tahun terhadap jumlah penduduk, persentase Askeskin atau Jamkesmas terhadap jumlah penduduk dan persentase jumlah sarana kesehatan terhadap jumlah penduduk sebagai faktor yang berpengaruh signifikan terhadap HIV dan AIDS.**

Kata Kunci*— *AIC, Probit Bivariat, HIV dan AIDS, MLE, Uji Simultan dan Parsial

I. PENDAHULUAN

KEMATIAN di Indonesia disebabkan banyak hal seperti kecelakaan, bermacam penyakit, dan lain-lain. Dalam hal ini angka kasus kematian yang disebabkan oleh berbagai penyakit cukup banyak. Salah satunya disebabkan oleh Penyakit Menular Seksual (PMS).

PMS adalah penyakit yang ditularkan dari satu orang ke orang lain melalui kontak hubungan seksual. Meskipun terutama ditularkan melalui hubungan seksual, namun dapat juga terjadi dari ibu kepada janin dalam kandungan atau saat kelahiran, melalui produk darah yang telah tercemar, kadang dapat ditularkan melalui alat kesehatan yang tidak steril [1]. Penyakit menular seksual ada yang dapat disembuhkan namun ada juga yang tidak dapat disembuhkan. HIV/AIDS termasuk dalam PMS yang tidak dapat disembuhkan dan paling berbahaya karena dapat merusak kekebalan tubuh seseorang dan berakibat kematian [2].

Di Jawa Timur sendiri angka kasus HIV dan AIDS kumulatif dari tahun 1987 sampai September 2014 masing-masing menempati urutan lima terbanyak di Indonesia [3]. Dengan demikian HIV dan AIDS perlu diperhatikan dalam menyumbangkan angka kasus kematian mengingat penyakit tersebut belum dapat disembuhkan. Jika seseorang terkena HIV maka sistem kekebalan tubuhnya semakin lama akan rusak. Dalam perkembangannya, orang yang telah terinfeksi virus HIV dapat mencapai stadium AIDS jika orang tersebut mengalami komplikasi gejala penyakit setelah sekian lama mengidap HIV.

Pada penelitian sebelumnya yang dilakukan oleh Chen dan Hamori tahun 2010 mengenai perbedaan antara pekerja laki-laki dan perempuan di Cina, dengan mempertimbangkan hubungan keputusan seorang perempuan untuk bekerja dan keputusan memperkerjakan seseorang pada sebuah organisasi diungkapkan untuk melihat hubungan antara dua variabel yang memiliki hubungan tersebut lebih baik menggunakan model probit bivariat [4]. Selain itu penelitian tentang HIV dan AIDS sebagai variabel respon pernah dilakukan oleh Dhewy dan Purhadi pada tahun 2014 menggunakan metode regresi poisson bivariat [5]. Penelitian menggunakan model probit bivariat pernah dilakukan oleh Romadhona juga menggunakan model regresi probit bivariat pada tahun 2015 untuk studi kasus pemberian imunisasi dasar dan ASI eksklusif [6].

Dalam tugas akhir ini digunakan model regresi probit bivariat untuk dapat mengidentifikasi dan menjelaskan faktor-faktor yang mempengaruhi HIV dan AIDS di Jawa Timur.

II. URAIAN PENELITIAN

A. Model Regresi Probit Bivariat

Model regresi probit bivariat adalah model regresi probit dengan menggunakan dua variabel respon. Model probit bivariat yang menggunakan dua kategori untuk masing-masing variabel responnya maka disebut model probit biner bivariat.

Model estimasi yang digunakan dalam model probit yaitu berasal dari CDF normal [7]. Dalam memodelkan regresi probit bivariat, digunakan dua variabel respon yang bersifat kualitatif dimana antar variabel respon tersebut memiliki hubungan [6].

Misalkan ada dua variabel respon Y_1dan Y_2 diasumsikan berasal dari variabel y_1^* dan y_2^* dengan

$$y_1^* = \beta_1^T x + \varepsilon_1$$
$$y_2^* = \beta_2^T x + \varepsilon_2$$

dimana

$\beta_1 = [\beta_{10} \quad \beta_{11} \quad \beta_{12} \cdots \quad \beta_{1q}]^T$
$\beta_2 = [\beta_{20} \quad \beta_{21} \quad \beta_{22} \cdots \quad \beta_{2q}]^T$
$x = [1 \quad x_1 \quad x_2 \cdots \quad x_q]^T$

dengan q adalah banyaknya variabel prediktor x, β_1 dan β_2 berukuran $(q+1) \times 1$ serta ε_1 dan ε_2 diasumsikan berdistribusi normal standar dengan $\mu = 0$ dan $\sigma^2 = 1$. Sehingga y_1^* serta y_2^* dinotasikan $y_1^* \sim N(\beta_1^T x, 1)$ dan $y_2^* \sim N(\beta_2^T x, 1)$. Jika ada dua variabel random yang berditribusi Normal yaitu y_1^* dan y_2^* maka menghasilkan distribusi Normal Bivariat. Pdf dari distribusi Normal Probit Bivariat adalah:

$$f(y_1^*, y_2^*) = \frac{1}{2\pi\sqrt{|\Sigma|}} exp\left(-\frac{1}{2}\begin{bmatrix} y_1^* - \beta_1^T x \\ y_2^* - \beta_2^T x \end{bmatrix} \Sigma^{-1} \begin{bmatrix} y_1^* - \beta_1^T x \\ y_2^* - \beta_2^T x \end{bmatrix}\right)$$

fungsi tersebut dinotasikan $(y_1^*, y_2^*) \sim N_2\left(\begin{bmatrix} \beta_1^T x \\ \beta_2^T x \end{bmatrix}, \Sigma\right)$

dimana

$$\Sigma = \begin{bmatrix} var(y_1^*) & cov(y_1^*, y_2^*) \\ cov(y_2^*, y_1^*) & var(y_2^*) \end{bmatrix}$$

dalam model probit bivariat terdapat beberapa asumsi yaitu:
1. $E(\varepsilon_1) = E(\varepsilon_2) = 0$
2. $Var(\varepsilon_1) = Var(\varepsilon_2) = 1$
3. $Cov(\varepsilon_1, \varepsilon_2) = \rho$

sehingga

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix} = \begin{bmatrix} 1 & \rho_{12} \\ \rho_{21} & 1 \end{bmatrix}$$

dan pdf menjadi Normal Standar Bivariat yaitu:

$$\phi(z_1, z_2) = \frac{1}{2\pi\sqrt{1-\rho^2}} exp\left(-\frac{1}{2}\frac{1}{(1-\rho^2)}(z_1^2 - 2\rho z_1 z_2 + z_2^2)\right)$$

dengan $\Phi(z_1, z_2)$ adalah CDF dari normal standar bivariat adalah:

$$\Phi(z_1, z_2) = \int_{-\infty}^{z_1}\int_{-\infty}^{z_2} \frac{1}{2\pi\sqrt{1-\rho^2}} exp\left(-\frac{1}{2}\frac{1}{(1-\rho^2)}(t_1^2 - 2\rho t_1 t_2 + t_2^2)\right) dt_1 dt_2$$

Dalam pembentukan kategori untuk masing-masing variabel respon pada model probit bivariat terlebih dahulu ditentukan *threshold* tertentu pada masing-masing variabel y_1^* dan y_2^* misal *r* dan *s*. Sehingga kategori yang terbentuk dari variabel $y_1^* = \beta_1^T x + \varepsilon_1$ adalah:

$Y = 0$ jika $y_1^* \leq r$ dan
$Y = 1$ jika $y_1^* > r$

sedangkan untuk variabel $y_2^* = \beta_2^T x + \varepsilon_2$ kategori yang terbentuk adalah:

$Y = 0$ jika $y_2^* \leq s$ dan
$Y = 1$ jika $y_2^* > s$

sehingga telah diperoleh asumsi *threshold* untuk masing-masing variabel respon dalam model probit biner bivariat [6].

Estimasi Parameter Model Probit Bivariat

Estimasi parameter dalam model probit bivariat menggunakan metode *Maximum Likelihood Estimation* (MLE). Dalam mendapatkan estimasi model probit bivariat didasarkan pada pembentukan tabel kontingensi yang dibuat yaitu tabel kontingensi dua arah antar dua variabel respon. Variabel respon pada tabel kontingensi dua arah menggunakan distribusi Multinomial. Dalam model probit bivariat parameter yang akan diestimasi adalah β dan ρ dimana $\beta = [\beta_1^T \quad \beta_2^T]^T$ dengan $\beta_1^T = [\beta_{10} \quad \beta_{11} \quad \beta_{12} \cdots \quad \beta_{1q}]$ dan $\beta_2^T = [\beta_{20} \quad \beta_{21} \quad \beta_{22} \cdots \quad \beta_{2q}]$ serta ρ adalah koefisien korelasi. Untuk mendapatkan hasil turunan terhadap parameter, diperlukan beberapa konsep dasar tentang turunan vektor menggunakan Lemma 2.1 yaitu [6]:

1. Jika diberikan vektor a berukuran $q \times 1$ dan w berukuran $q \times 1$ maka $\frac{\partial(a^T w)}{\partial a} = w$.
2. Jika $\Phi(a^T w)$ adalah distribusi normal kumulatif maka $\frac{\partial\Phi(a^T w)}{\partial w} = a\phi(a^T w)$ dimana $\phi(a^T w)$ adalah distribusi normal standar.
3. Jika $\phi(a^T w)$ adalah distribusi normal standar maka $\frac{\partial\phi(a^T w)}{\partial w} = -a(a^T w)\phi(a^T w)$

B. Multikolinieritas

Dalam pemodelan regresi diperlukan syarat bahwa tidak adanya korelasi antar variabel prediktor yang digunakan. Multikolinieritas merupakan kondisi dimana terdapat hubungan linier atau korelasi yang tinggi antar variabel prediktor di dalam model regresi. Jika besarnya korelasi melebihi 0,8 atau 80% maka korelasi berpasangan diantara dua variabel prediktor dikatakan tinggi [7].

Perhitungan koefisien korelasi antar dua variabel prediktor adalah sebagai berikut:

$$r_{x_i x_j} = \frac{n\sum_{i=1}^n \sum_{j=1}^n x_i x_j - (\sum_{i=1}^n x_i)(\sum_{j=1}^n x_j)}{\sqrt{n\sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}\sqrt{n\sum_{j=1}^n x_j^2 - (\sum_{j=1}^n x_j)^2}}$$

Untuk mengetahui korelasi antar variabel yang berupa data kategorik, dapat menggunakan uji dependensi dengan syarat tidak ada nilai frekuensi dari sel yang kurang dari satu dan jika ada sel yang berisi frekuensi kurang dari lima maka tidak boleh melebihi 20% dari jumlah total sel. Uji dependensi dilakukan sebagai berikut.

Hipotesis:
H_0: Variabel x_i dan x_j saling bebas
H_1: Variabel x_i dan x_j tidak saling bebas

Statistik Uji:

$$X^2 = \sum_{i=0}^I \sum_{j=0}^J \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

dimana
O_{ij} = banyaknya data yang termasuk dalam sel ke- i, j
E_{ij} = nilai harapan (taksiran) dari O_{ij}
$E_{ij} = \frac{n_i n_j}{n}$

Tolak H_0 jika nilai $X^2 > \chi^2_{\alpha,(I-1)(J-1)}$ [8].

C. Pengujian Signifikansi Parameter Model Probit

Uji signifikansi parameter model probit bivariat dilakukan dengan dua pengujian yaitu uji simultan untuk menguji apakah variabel prediktor mempunyai pengaruh signifikan terhadap variabel respon dan uji parsial untuk menguji apakah masing-masing parameter berpengaruh signifikan terhadap variabel respon Y_1 dan Y_2 [6].

a. Uji Simultan

Hipotesis:
$H_0: \beta_1 = \beta_2 = \cdots = \beta_q = 0$
H_1: Paling tidak ada satu $\beta_k \neq 0$ dengan $k = 1,2, \dots q$

Statistik uji:

$$G^2 = 2\left[\ln L(\hat{\Omega}) - \ln L(\hat{\omega})\right]$$

Dengan $L(\hat{\Omega})$ adalah fungsi *likelihood* di bawah populasi dan $L(\hat{\omega})$ adalah fungsi *likelihood* di bawah H_0. $G^2 \xrightarrow{d} W$ dan $W \sim \chi^2_{\alpha,df}$ apabila $n \to \infty$. Tolak H_0 jika $G^2 > \chi^2_{\alpha,df}$ dimana derajat bebas (df) yaitu banyaknya variabel prediktor dalam model.

b. Uji Parsial

Hipotesis:
$H_0: \beta_k = 0$
$H_1: \beta_k \neq 0$ dengan $k = 1,2, \dots q$

Statistik uji:

$$W = \frac{\hat{\beta}_k}{SE(\hat{\beta}_k)}$$

Statistik uji Wald mengikuti distribusi normal standar. Tolak H_0 jika $|W| > Z_{\alpha/2}$.

D. Pemilihan Model Terbaik

Untuk memilih model terbaik dari model probit bivariat dapat dilihat dari nilai *Akaike Information Criterion* (AIC). Semakin kecil nilai AIC maka model semakin baik [9]. Perhitungan AIC dirumuskan sebagai berikut:

$$AIC = -2lnL(\hat{\theta}) + 2p$$

dengan
$L(\hat{\theta})$ = nilai maksimum fungsi *likelihood*
p = banyak parameter dalam model

E. HIV dan AIDS

HIV atau *Humman Immunodeficiency Virus* adalah sejenis virus yang menyerang atau menginfeksi sel darah putih yang menyebabkan turunnya kekebalan tubuh manusia. Sementara AIDS atau *Acquired Immune Deficiency Syndrome* adalah sekumpulan gejala penyakit yang timbul karena turunnya kekebalan tubuh manusia yang disebabkan infeksi oleh HIV. Akibat menurunnya kekebalan tubuh maka orang tersebut sangat mudah terkena berbagai penyakit infeksi (infeksi oportunistik) yang sering berakibat fatal [10].

Seseorang yang telah terinfeksi HIV memerlukan adanya perawatan jangka panjang dalam pencegahan penyakit menjadi lebih parah mencapai stadium AIDS. Sehingga diperlukan pula jaminan kesehatan untuk masyarakat terutama masyarakat dengan ekonomi menengah kebawah. Dalam menangani kasus penyakit, sarana kesehatan dibutuhkan demi menunjang kesembuhan seorang pasien. Selain itu penting juga diadakannya penyuluhan kesehatan untuk masyarakat utamanya penyuluhan mengenai bahaya penyakit HIV/ AIDS.

III. HASIL DAN PEMBAHASAN

A. Analisis Data Penyakit HIV dan AIDS di Jawa Timur

Situasi kasus penyakit HIV dan AIDS tertinggi dari tahun 1989 sampai dengan bulan September 2014 untuk setiap kabupaten dan kota di Jawa Timur ditempati Kota Surabaya, Kota Malang dan Kabupaten Jember sebagai kasus tertinggi.

Pada penelitian ini diterapkan model regresi probit bivariat dengan menggunakan data kasus HIV dan AIDS untuk 38 Kabupaten dan Kota di Jawa Timur pada tahun 2013. Variabel penelitian yang digunakan adalah dua variabel respon yaitu kasus HIV (Y_1) dan kasus AIDS (Y_2) serta empat variabel prediktor yaitu persentase kelompok umur 25-49 (x_1), persentase Askeskin atau Jamkesmas (x_2), persentase penyuluhan kesehatan (x_3), dan persentase sarana kesehatan (x_4). Pengkategorian untuk HIV dan AIDS adalah:

Kode 0 = Kategori Rendah, jika *case rate* HIV tiap kabupaten/ kota dibawah *case rate provinsi* untuk HIV
Kode 1 = Kategori Tinggi, jika *case rate* HIV tiap kabupaten/ kota diatas *case rate provinsi* untuk HIV
Kode 0 = Kategori Rendah, jika *case rate* AIDS tiap kabupaten/ kota dibawah *case rate provinsi* untuk AIDS
Kode 1 = Kategori Tinggi, jika *case rate* AIDS tiap kabupaten/ kota diatas *case rate provinsi* untuk AIDS

B. Pemodelan Regresi Probit Bivariat

Untuk memodelkan suatu regresi diperlukan uji korelasi antar variabel untuk melihat ada tidaknya korelasi antar variabel.

1. Hubungan Antar Variabel

Uji untuk dua variabel respon sebagai berikut.
Hipotesis:
H_0: Variabel x_i dan x_j saling bebas (independen)
H_1: Variabel x_i dan x_j tidak saling bebas (dependen)
Statistik Uji:

$$X^2_{hitung} = \frac{(19-14{,}4)^2}{14{,}4} + \frac{(2-6{,}6)^2}{6{,}6} + \frac{(7-11{,}6)^2}{11{,}6} + \frac{(10-5{,}4)^2}{5{,}4}$$
$$= 10{,}568$$

Karena nilai $X^2_{hitung} > \chi^2_{0.05,1} = 3{,}84$, jadi tolak H_0. Dapat disimpulkan antar variabel respon tidak saling bebas atau terdapat hubungan antara variabel respon HIV dan AIDS.

Selanjutnya adalah mendeteksi ada tidaknya multikolinieritas antar variabel prediktor. Empat variabel prediktor dihitung koefisien korelasinya untuk mengetahui bahwa tidak ada korelasi antar variabel prediktor yang digunakan. Berdasarkan data yang telah diperoleh, perhitungan nilai koefisien korelasi untuk variabel prediktor x_1 dan x_2 yaitu:

$$r_{x_i x_j} = \frac{38(238{,}0156) - (702{,}65)(12{,}9981)}{\sqrt{38(13034{,}158) - 493713{,}12}\sqrt{38(5{,}258) - 168{,}95}}$$
$$= -0{,}400$$

Karena nilai korelasi tidak melebihi 0,8 atau 80%, dapat disimpulkan tidak terjadi multikolinieritas antar variabel prediktor x_1 dan x_2. Dengan cara yang sama dapat diperoleh nilai koefisien korelasi antar variabel prediktor pada Tabel 3. Masing-masing koefisien korelasi antar variabel prediktor tersebut tidak ada yang melebihi 80% sehingga dapat disimpulkan tidak terjadi multikolinieritas antar variabel prediktor. Apabila nilai korelasi bernilai negatif hal ini berarti kedua variabel memiliki hubungan yang tidak searah dimana jika x_1 bertambah maka x_2 berkurang.

Tabel 3.
Koefisien Korelasi Antar Variabel Prediktor

Korelasi	x_1	x_2	x_3	x_4
x_1	-	-0,400	-0,193	0,143
x_2	-0,400	-	-0,016	-0,377
x_3	-0,193	-0,016	-	0,378
x_4	0,143	-0,377	0,378	-

2. Estimasi Parameter Probit Bivariat

Tabel kontingensi frekuensi untuk variabel respon Y_1 dan Y_2 serta tabel kontingensi probabilitas adalah seperti berikut.

Tabel 1.
Tabel Kontingensi Frekuensi Variabel Y_1 dan Y_2

Y_1	Y_2	
	$Y_2 = 0$	$Y_2 = 1$
$Y_1 = 0$	Y_{00}	Y_{01}
$Y_1 = 1$	Y_{10}	Y_{11}

Tabel 2.
Tabel Kontingensi Probabilitas Variabel Y_1 dan Y_2

Y_1	Y_2		Total
	$Y_2 = 0$	$Y_2 = 1$	$p_{0.}(x) = 1 - p_1(x)$
$Y_1 = 0$	$p_{00}(x)$	$p_{01}(x)$	$p_{1.}(x) = p_1(x)$
$Y_1 = 1$	$p_{10}(x)$	$p_{11}(x)$	$p_{..}(x) = 1$
Total	$p_{.0}(x) = 1 - p_2(x)$	$p_{.1}(x) = p_2(x)$	$p_{0.}(x) = 1 - p_1(x)$

Sesuai dengan tabel kontingensi (2×2) yang dibentuk pada Tabel 1 dan Tabel 2, variabel respon berditribusi Multinomial dengan notasi $Y \sim MULT(1; p_{11}, p_{10}, p_{01})$ dengan $p_{00} = 1 - p_{11} - p_{10} - p_{01}$. Diketahui fungsi regresi probit bivariat yaitu:

$y_1^* = \beta_1^T x + \varepsilon_1$
$y_2^* = \beta_2^T x + \varepsilon_2$

dengan ε_1 dan ε_2 diasumsikan berdistribusi normal standar dengan $\mu = 0$ dan $\sigma^2 = 1$ dan pembentukan kategori dengan *threshold* untuk masing-masing variabel respon yaitu:

$Y_1 = 0$ jika $y_1^* \leq r$
$Y_1 = 1$ jika $y_1^* > r$ dan
$Y_2 = 0$ jika $y_2^* \leq s$
$Y_2 = 1$ jika $y_2^* > s$

dan probabilitas bersama untuk tiap sel pada Tabel 2 adalah:

$$p_{00}(x) = \Phi(z_1, z_2) \quad (1)$$

$$p_{01}(x) = \Phi(z_1) - \Phi(z_1, z_2) \quad (2)$$

$$p_{10}(x) = \Phi(z_2) - \Phi(z_1, z_2) \quad (3)$$

$$p_{11}(x) = 1 - \Phi(z_1) - \Phi(z_2) + \Phi(z_1, z_2) \quad (4)$$

dengan nilai marginal $p_1(x)$ dan $p_2(x)$ masing-masing yaitu:

$$p_1(x) = 1 - \Phi(r - \beta_1^T x) \quad (5)$$

$$p_2(x) = 1 - \Phi(s - \beta_2^T x) \quad (6)$$

Selanjutnya adalah membentuk fungsi *likelihood* dari variabel random bivariat yang berdistribusi Multinomial yaitu:

$$L(\beta, \rho) = \prod_{i=1}^{n} P(Y_{11} = y_{11}, Y_{10} = y_{10}, Y_{01} = y_{01})$$

$$= \prod_{i=1}^{n} p_{11}^{y_{11}}(x_i) p_{10}^{y_{10}}(x_i) p_{01}^{y_{01}}(x_i)[1 - p_{11}(x_i) - p_{10}(x_i) - p_{01}(x_i)]^{1 - y_{11} - y_{10} - y_{01}}$$

Dimisalkan $p_{gh}^{y_{gh}}(x_i) = p_{ghi}^{y_{ghi}}$ dimana $g = 0{,}1$ dan $h = 0{,}1$ kemudian subtitusi $y_{00i} = 1 - y_{11i} - y_{10i} - y_{01i}$ sehingga fungsi *likelihood* menjadi:

$$L(\beta, \rho) = \prod_{i=1}^{n} p_{11i}^{y_{11i}} p_{10i}^{y_{10i}} p_{01i}^{y_{01i}} [1 - p_{11i} - p_{10i} - p_{01i}]^{y_{00i}}$$

Selanjutnya fungsi *likelihood* tersebut di *ln*-kan sehingga didapatkan fungsi:

$$\ln L(\beta, \rho) = \sum_{i=1}^{n} y_{11i} \ln p_{11i} + y_{10i} \ln p_{10i} + y_{01i} \ln p_{01i} + y_{00i} \ln(1 - p_{11i} - p_{10i} - p_{01i}) \quad (7)$$

Persamaan (7) mengandung parameter β dan ρ dimana $\beta = [\beta_1^T \quad \beta_2^T]^T$ dengan $\beta_1^T = [\beta_{10} \quad \beta_{11} \quad \beta_{12} \cdots \quad \beta_{1q}]$ dan $\beta_2^T = [\beta_{20} \quad \beta_{21} \quad \beta_{22} \cdots \quad \beta_{2q}]$. Untuk memudahkan penurunan, subtitusi $p_{11i} = p_{2i} - p_{01i}$ dan $p_{10i} = p_{1i} - p_{2i} - p_{01i}$ sehingga persamaan (7) menjadi:

$$\ln L(\beta, \rho) = \sum_{i=1}^{n} [y_{11i} \ln(p_{2i} - p_{01i}) + y_{10i} \ln(p_{1i} - p_{2i} + p_{01i}) + y_{01i} \ln p_{01i} + y_{00i} \ln(1 - p_{1i} - p_{01i})]$$

Fungsi turunan pertama $\ln L(\beta, \rho)$ terhadap β_1 adalah sebagai berikut:

$$\frac{\partial \ln L(\beta, \rho)}{\partial \beta_1} = \sum_{i=1}^{n} \left[y_{11i} \frac{1}{p_{2i} - p_{01i}} \left(-\frac{\partial p_{01i}}{\partial \beta_1} \right) + y_{10i} \frac{1}{p_{1i} - p_{2i} + p_{01i}} \left(\frac{\partial p_{1i}}{\partial \beta_1} + \frac{\partial p_{01i}}{\partial \beta_1} \right) + y_{01i} \frac{1}{p_{01i}} \frac{\partial p_{01i}}{\partial \beta_1} + y_{00i} \frac{1}{1 - p_{1i} - p_{01i}} \left(-\frac{\partial p_{1i}}{\partial \beta_1} - \frac{\partial p_{01i}}{\partial \beta_1} \right) \right] \quad (8)$$

Selanjutnya pada persamaan (8) dapat dimisalkan:

$a_i = \frac{1}{p_{2i} - p_{01i}}$, $b_i = \frac{1}{p_{1i} - p_{2i} - p_{01i}}$, $c_i = \frac{1}{p_{01i}}$, $d_i = \frac{1}{1 - p_{1i} - p_{01i}}$

Sehingga persamaan (8) disederhanakan menjadi:

$$\frac{\partial \ln L(\beta, \rho)}{\partial \beta_1} = \sum_{i=1}^{n} \left[(-a_i y_{11i} + b_i y_{10i} + c_i y_{01i} - d_i y_{00i}) \frac{\partial p_{01i}}{\partial \beta_1} + (b_i y_{10i} - d_i y_{00i}) \frac{\partial p_{1i}}{\partial \beta_1} \right] \quad (9)$$

Berdasarkan persamaan (2), untuk mencari turunan dari $\frac{\partial p_{01i}}{\partial \beta_1}$ adalah sebagai berikut:

$$\frac{\partial p_{01i}}{\partial \beta_1} = \frac{\partial \left(\Phi(z_{1i}) - \Phi(z_{1i}, z_{2i}) \right)}{\partial \beta_1} = \frac{\partial \Phi(z_{1i})}{\partial \beta_1} - \frac{\partial \Phi(z_{1i}, z_{2i})}{\partial \beta_1} \quad (10)$$

Dari persamaan (10) perlu dicari turunan dari $\frac{\partial \Phi(z_{1i})}{\partial \beta_1}$ dan $\frac{\partial \Phi(z_{1i}, z_{2i})}{\partial \beta_1}$ sebagai berikut:

$$\frac{\partial \Phi(z_{1i})}{\partial \beta_1} = \frac{\partial \left(\Phi(r - \beta_1^T x_i) \right)}{\partial \beta_1} = \frac{\partial \left(\Phi(r - \beta_1^T x_i) \right)}{\partial \left(\Phi(r - \beta_1^T x_i) \right)} \frac{\partial \left((r - \beta_1^T x_i) \right)}{\partial \beta_1} = \emptyset(r - \beta_1^T x_i)(-x_i) = -x_i \emptyset(z_{1i}) \quad (11)$$

$$\frac{\partial \Phi(z_{1i}, z_{2i})}{\partial \beta_1} = -x_i \frac{1}{2} \emptyset(z_{1i}) \left[1 + erf\left(\frac{z_{2i} - z_{1i}\rho}{\sqrt{2(1 - \rho^2)}} \right) \right] = -x_i \varphi_{1i} \quad (12)$$

dimana

$$\mathrm{erf}(x) = \frac{2}{\pi^2} \left[\int_{t=0}^{x} e^{t^2} dt \right]$$

$$\varphi_{1i} = \frac{1}{2} \emptyset(z_{1i}) \left[1 + erf\left(\frac{z_{2i} - z_{1i}\rho}{\sqrt{2(1 - \rho^2)}} \right) \right]$$

dengan mensubtitusikan persamaan (11) dan (12) ke persamaan (10) didapatkan:

$$\frac{\partial p_{01i}}{\partial \beta_1} = -x_i \emptyset(z_{1i}) + x_i \varphi_{1i} \quad (13)$$

Selanjutnya dari persamaan (5) turunan dari $\frac{\partial p_{1i}}{\partial \beta_1}$ adalah:

$$\frac{\partial p_{1i}}{\partial \beta_1} = \frac{\partial \left(1 - \Phi(z_{1i}) \right)}{\partial \beta_1} = \frac{\partial (1)}{\partial \beta_1} - \frac{\partial \left(\Phi(r - \beta_1^T x_i) \right)}{\partial \beta_1} = 0 - \left(\frac{\partial \left(\Phi(r - \beta_1^T x_i) \right)}{\partial \left(\Phi(r - \beta_1^T x_i) \right)} \frac{\partial \left((r - \beta_1^T x_i) \right)}{\partial \beta_1} \right) = -\emptyset(r - \beta_1^T x_i)(-x_i) = x_i \emptyset(z_{1i}) \quad (14)$$

Persamaan (13) dan (14) selanjutnya disubtitusikan ke persamaan (9) dengan hasil sebagai berikut:

$$\frac{\partial \ln L(\beta, \rho)}{\partial \beta_1} = \sum_{i=1}^{n} \left[\left((-a_i y_{11i} + b_i y_{10i} + c_i y_{01i} - d_i y_{00i})(-x_i \emptyset(z_{1i}) + x_i \varphi_{1i}) \right) + \left((b_i y_{10i} - d_i y_{00i})(x_i \emptyset(z_{1i})) \right) \right] \quad (15)$$

Persamaan (15) dapat disederhanakan menjadi:

$$\frac{\partial \ln L(\beta, \rho)}{\partial \beta_1} = \sum_{i=1}^{n} [(a_i y_{11i} - c_i y_{01i}) x_i \emptyset(z_{1i}) + (-a_i y_{11i} + b_i y_{10i} + c_i y_{01i} - d_i y_{00i}) x_i \varphi_{1i}]$$

Sehingga didapatkan turunan pertama $\ln L(\beta, \rho)$ terhadap β_1 adalah:

$$\frac{\partial \ln L(\beta, \rho)}{\partial \beta_1} = \sum_{i=1}^{n} x_i [(a_i y_{11i} - c_i y_{01i}) \emptyset(z_{1i}) + (-a_i y_{11i} + b_i y_{10i} + c_i y_{01i} - d_i y_{00i}) \varphi_{1i}] \quad (16)$$

Dengan cara yang sama didapatkan turunan pertama $\ln L(\beta, \rho)$ terhadap β_2 dan ρ seperti berikut:

$$\frac{\partial \ln L(\beta,\rho)}{\partial \beta_2} = \sum_{i=1}^{n} \left[(a_i y_{11i} - b_i y_{10i})\left(x_i \emptyset(z_{2i})\right) + (-a_i y_{11i} + b_i y_{10i} + c_i y_{01i} - d_i y_{00i})(x_i \varphi_{2i}) \right] \quad (17)$$

$$\frac{\partial \ln L(\beta,\rho)}{\partial \rho} = \sum_{i=1}^{n} \left[(a_i y_{11i} - b_i y_{10i} - c_i y_{01i} + d_i y_{00i})\varphi_i \right] \quad (18)$$

Turunan pertama fungsi *ln likelihood* terhadap masing-masing parameter yang diperoleh pada persamaan (16), (17) dan (18) tidak *closed form* sehingga dilakukan analisis numerik mengunakan iterasi *Newton Raphson* dengan rumus:

$$\beta^{(m)} = \beta^{(m-1)} - \left[H\left(\beta^{(m-1)}\right)\right]^{-1} g\left(\beta^{(m-1)}\right)$$

dimana

$\beta^{(m)}$ = parameter β dengan iterasi ke-*m*

$\beta^{(m-1)}$ = parameter β dengan iterasi ke- $(m-1)$

$H\left(\beta^{(m-1)}\right)$ = matriks dari turunan kedua fungsi *ln likelihood* (matriks *Hessian*)

$g\left(\beta^{(m-1)}\right)$ = vektor dari turunan pertama fungsi *ln likelihood*

Elemen dari vektor $g(\beta)$ masing-masing diberikan oleh persamaan (16), (17) dan (18) sebagai turunan pertama dari fungsi $\ln L(\beta,\rho)$ terhadap parameter β_1, β_2 dan ρ. Selain itu dengan cara yang sama digunakan elemen dari matriks Hessian sebagai turunan kedua dari masing-masing parameter.

3. Pemilihan Model Terbaik

Mendapatkan model terbaik dilakukan dengan mengkombinasikan semua kemungkinan model yaitu sebanyak $2^q - 1$ dimana q adalah banyaknya variabel prediktor kemudian diambil model dengan nilai AIC terkecil. Sehingga untuk mendapatkan model terbaik probit bivariat, pemodelan dilakukan sebanyak 15 kali dan dipilih nilai AIC terkecil adalah dengan tiga variabel prediktor yaitu x_1, x_2, x_4. Perhitungan AIC untuk model terbaik adalah sebagai berikut:

$AIC = -2lnL(\hat{\theta}) + 2p = -2(-34.076215) + 2(9) = 86{,}152$

Dengan menggunakan bantuan StataSE12 model terbaik probit bivariat yaitu:

$$\hat{y}_1^* = -10{,}699 + 0{,}612x_1 - 0{,}664x_2 - 136{,}298x_4 \quad (19)$$

$$\hat{y}_2^* = 3{,}046 - 0{,}015x_1 - 3{,}656x_2 - 557{,}638x_4 \quad (20)$$

Dapat diambil contoh untuk Kota Surabaya, dengan persentase kelompok umur 25-49 tahun $(x_1 = 20{,}87)$ dan persentase Askeskin atau Jamkesmas $(x_2 = 0{,}13753)$ serta persentase jumlah sarana kesehatan $(x_4 = 0{,}004)$, berdasarkan model terbaik dari $\hat{y}_1^*$ dan $\hat{y}_2^*$ pada persamaan (19) dan (20) didapatkan nilai $\hat{y}_1^*$ dan $\hat{y}_2^*$ sebagai berikut:

$$\hat{y}_1^* = -10{,}699 + 0{,}612(20{,}87) - 0{,}664(0{,}13753) - 136{,}298(0{,}004) = 1{,}453$$

$$\hat{y}_2^* = 3{,}046 - 0{,}015(20{,}87) - 3{,}656(0{,}13753) - 557{,}638(0{,}004) = -0{,}001$$

Karena $z_1 = -y_1^*$ dan $z_2 = -y_2^*$ dimana $r = 0$ dan $s = 0$ dan berdasarkan persamaan (1) sampai dengan (4) didapatkan hasil dari probabilitas untuk setiap kategori adalah sebagai berikut:

$\hat{p}_{00}^*(x)\Phi(-1{,}453; 0{,}001) = 0{,}0567$

$\hat{p}_{01}^*(x) = \Phi(-1{,}453) - \Phi(-1{,}453; 0{,}001) = 0{,}0187$

$\hat{p}_{10}^*(x) = \Phi(-0{,}001) - \Phi(-1{,}453; 0{,}001) = 0{,}4435$

$\hat{p}_{11}^*(x) = 1 - \Phi(1{,}453) - \Phi(-0{,}001) + \Phi(-1{,}453; 0{,}001) = 0{,}4812$

Berdasarkan nilai probabilitas dari tiap kategori yang telah didapatkan, dapat disimpulkan bahwa Kota Surabaya mempunyai probabilitas sebesar 0,4812 atau 48,12% untuk berada dalam kategori HIV tinggi dan kategori AIDS tinggi sehingga penanganan kasus penyakit HIV dan AIDS di Kota Surabaya masih perlu diperhatikan.

a. Uji Simultan

Dengan bantuan *software* StataSE12 didapatkan nilai log *likelihood* masing-masing model sehingga didapatkan statistik uji untuk G^2 sebagai berikut.

Hipotesis:

$H_0: \beta_{11} = \beta_{12} = \beta_{14} = 0$ dan $\beta_{21} = \beta_{22} = \beta_{24} = 0$

H_1: Paling tidak ada satu $\beta_{kl} \neq 0$; k = 1,2 dan l = 1,2,4

Statistik Uji:

$G^2 = 2[-34{,}076 - (-44{,}250)] = 20{,}348$

Nilai G^2 sebesar 20,348 kemudian dibandingkan dengan nilai $X^2_{0,05;6} = 12{,}59$ $(G^2 > \chi^2_{0,05;6})$ dengan keputusan tolak H_0. Jadi kesimpulan yang dapat diambil adalah paling tidak ada satu variabel yang signifikan terhadap variabel respon.

b. Uji Parsial

Untuk parameter $\hat{\beta}_{11}$

Hipotesis:

$H_0: \beta_{11} = 0$

$H_1: \beta_{11} \neq 0$

Statistik uji:

$$W = \frac{\hat{\beta}_{11}}{SE(\hat{\beta}_{11})} = \frac{0{,}612}{0{,}249} = 2{,}46$$

Dengan cara yang sama diperoleh statistik uji untuk parameter yang lain pada Tabel 4 sebagai berikut:

Tabel 4.
Nilai Koefisien dan Standar Error masing-masing Variabel Prediktor

Untuk Y_1	$\hat{\beta}_i$	$SE(\hat{\beta}_i)$	W
x_2	-0,664	1,601	-0,41
x_4	-136,298	147,188	-0,93
Untuk Y_2			
x_1	-0,015	0,221	-0,07
x_2	-3,656	1,755	-2,08
x_4	-557,638	242,198	-2,30

Tolak H_0 jika $|W| > Z_{\alpha/2}$ dengan taraf signifikan $\alpha = 5\%$. Sehingga variabel persentase kelompok umur 25-49 tahun (x_1) berpengaruh signifikan terhadap HIV. Sedangkan variabel persentase Askeskin atau Jamkesmas (x_2) dan variabel persentase sarana kesehatan (x_4) berpengaruh signifikan terhadap AIDS.

4. Interprestasi Model Regresi Probit Bivariat

Interpretasi pada model regresi probit bivariat dilihat dari besarnya efek marginal tiap variabel prediktor yang signifikan. Efek marginal digunakan untuk melihat besarnya pengaruh perubahan suatu variabel prediktor terhadap variabel respon dengan asumsi variabel lainnya konstan. Diambil contoh efek marginal variabel persentase umur 25-49 tahun $(x_1 = 20{,}87)$ di Kota Surabaya dengan menggunakan Matlab sebagai berikut:

$$\frac{\partial \hat{p}_{00}^*}{\partial x_1} = \frac{\partial \Phi(z_1, z_2)}{\partial x_1} = -\hat{\beta}_{1,1}^*\varphi_1 - \hat{\beta}_{2,1}^*\varphi_2 = -0{,}612\varphi_1 - (-0{,}015)\varphi_2 = -0{,}0430$$

$$\frac{\partial \hat{p}_{01}^*}{\partial x_1} = \frac{\partial\left(\Phi(z_1) - \Phi(z_1, z_2)\right)}{\partial x_1} = -\hat{\beta}_{1,1}^*\phi(z_1) + \hat{\beta}_{1,1}^*\varphi_1 + \hat{\beta}_{2,1}^*\varphi_2$$

$= -0{,}612\phi(z_1) + 0{,}612\varphi_1 + (-0{,}015)\varphi_2 = -0{,}0439$

$$\frac{\partial \hat{p}_{10}^*}{\partial x_1} = \frac{\partial(\Phi(z_2) - \Phi(z_1, z_2))}{\partial x_1} = -\hat{\beta}_{2,1}^*\phi(z_2) + \hat{\beta}_{1,1}^*\varphi_1 + \hat{\beta}_{2,1}^*\varphi_2$$

$= -(-0{,}015)\phi(z_1) + 0{,}612\varphi_1 + (-0{,}015)\varphi_2 = 0{,}0490$

$$\frac{\partial \hat{p}_{11}^*}{\partial x_1} = \frac{\partial(1 - \Phi(z_1) - \Phi(z_2) + \Phi(z_1, z_2))}{\partial x_1}$$

$= \hat{\beta}_{1,1}^*\phi(z_1) + \hat{\beta}_{2,1}^*\phi(z_2) - \hat{\beta}_{1,1}^*\varphi_1 - \hat{\beta}_{2,1}^*\varphi_2$

$= 0{,}612\phi(z_1) + (-0{,}015)\phi(z_2) - 0{,}612\varphi_1 + (-0{,}015)\varphi_2$

$= 0{,}0379$

Efek marginal variabel persentase umur 25-49 tahun (x_1) terhadap $\hat{p}_{11}^*$ adalah sebesar 0,0379 yang berarti bahwa perubahan persentase umur 25-49 tahun (x_1) sebesar satu satuan akan meningkatkan 0,0379 terhadap probabilitas HIV tinggi dan AIDS yang juga tinggi. Variabel persentase umur hanya signifikan terhadap variabel Y_1 yaitu kasus HIV.

5. Ketepatan Klasifikasi Model

Ketepatan klasifikasi digunakan untuk melihat seberapa besar ketepatan antara data aktual dengan hasil prediksinya. Tabel kontingensi dari ketepatan klasifikasi antara data aktual dengan hasil prediksi dapat dilihat pada Tabel 5. Dikatakan klasifikasi akan tepat jika hasil prediksi yaitu nilai probabilitas untuk masing-masing data diambil yang terbesar kemudian dibandingkan dengan data aktual yang telah didapatkan.

Tabel 5.
Tabel Kontingensi Ketepatan Klasifikasi Model Terbaik

	Prediksi				
Aktual	Y_{00}	Y_{01}	Y_{10}	Y_{11}	Total
Y_{00}	**0**	6	13	0	19
Y_{01}	0	**1**	0	1	2
Y_{10}	0	0	**6**	1	7
Y_{11}	0	3	4	**3**	10
Total	0	10	23	5	38

Berdasarkan model terbaik yang telah dibentuk, ketepatan klasifikasi untuk memprediksi yaitu sebesar 26,32%. Dari model yang dihasilkan, tidak ada variabel prediktor yang signifikan terhadap semua variabel respon sehingga didapatkan besarnya ketepatan klasifikasi yang masih kecil.

IV. KESIMPULAN

Berdasarkan hasil dan pembahasan yang diperoleh, kesimpulan yang dapat diambil adalah:

a. Langkah-langkah untuk mengestimasi parameter dalam model probit bivariat adalah sebagai berikut:
   1. Membuat tabel kontingensi dua arah antar dua variabel respon kemudian mengetahui persamaan dari probabilitas.
   2. Mendapatkan fungsi *likelihood* berdistribusi multinomial kemudian mendapatkan fungsi *ln likelihood*.
   3. Mendapatkan turunan pertama dari fungsi *ln likelihood* terhadap parameter β_1, β_2 dan ρ.
   4. Hasil dari turunan pertama yang didapatkan tidak *closed form*, sehingga dilakukan analisis numerik dengan mengunakan iterasi Newton Raphson dengan rumus $\beta^{(m)} = \beta^{(m-1)} - \left[H\left(\beta^{(m-1)}\right)\right]^{-1} g\left(\beta^{(m-1)}\right)$.
   5. Mendapatkan turunan kedua dari fungsi *ln likelihood* terhadap parameter β_1, β_2 dan ρ.
   6. Menentukan nilai awal parameter β dan ρ untuk iterasi yaitu sama dengan nol. Iterasi akan berhenti dan didapatkan nilai estimasi jika nilai selisih antara $\beta^{(m)}$ dan $\beta^{(m-1)}$ sangat kecil.

b. Model regresi probit bivariat untuk faktor-faktor yang mempengaruhi jumlah kasus HIV dan AIDS di Jawa Timur adalah sebagai berikut:
   $\hat{y}_1^* = -10{,}699 + 0{,}612x_1 - 0{,}664x_2 - 136{,}298x_4$
   $\hat{y}_2^* = 3{,}046 - 0{,}015x_1 - 3{,}656x_2 - 557{,}638x_4$
   Variabel prediktor yang signifikan terhadap HIV (Y_1) adalah persentase kelompok umur 25-49 tahun terhadap jumlah penduduk (x_1) dan variabel yang signifikan terhadap AIDS (Y_1) adalah persentase jumlah Askeskin atau Jamkesmas terhadap jumlah penduduk (x_2) dan persentase jumlah sarana kesehatan terhadap jumlah penduduk (x_4). Model tersebut merupakan model terbaik dari regresi probit bivariat untuk kasus HIV dan AIDS di Jawa Timur dengan nilai AIC terkecil sebesar 86,152.

DAFTAR PUSTAKA

[1] Hospital, Siloam. 2014. *"Bisakah IMS Dicegah?"*. Diakses tanggal 11 Februari 2015 dari http://www.siloamhospitals.com.

[2] Abiasa, Himpunan. 2013. *"Informasi Dasar Tentang IMS (Infeksi Menular Seksual)"*. Diakses tanggal 11 Februari 2015 dari http://www.himpunan-abiasa.com.

[3] Dinas Kesehatan Jawa Timur. 2014. *"Analisa Situasi dan Kebijakan HIV AIDS Jatim"*. Seksi P2 Dinkes Provinsi Jatim. Surabaya.

[4] Chen, G. dan Hamori, S. 2010. *"Bivariate Probit Analysis of Differences Between Male and Female Formal Employment in Urban China"*. **Journal of Asian Economics 21**, Hal. 494-501.

[5] Dhewy, R. C. dan Purhadi. 2014. *"Pemodelan Bivariate Poisson Regression dengan Kovarian Merupakan Fungsi dari Variabel Bebas. Studi Kasus: Jumlah HIV dan AIDS di Propinsi Jawa Timur Tahun 2012"*. **Tesis**. Program Magister Statistika Institut Teknologi Sepuluh Nopember. Surabaya.

[6] Romadhona, M. N. 2015. *"Pemodelan Pemberian Imunisasi Dasar dan ASI Eksklusif dengan Pendekatan Model Probit Biner Bivariat"*. **Tesis**. Program Magister Statistika Institut Teknologi Sepuluh Nopember. Surabaya

[7] Gujarati, D. N. dan D. C Porter. 2009. *"Basic Econometrics"*. 5th ed. Mc Graw Hill Education. USA. Terjemahan E. Mardanugraha, S. Wardhani dan C. Mangunsong. (2013). *"Dasar-dasar Ekonometrika"*. Edisi Kelima. Buku Dua. Salemba Empat. Jakarta.

[8] Ramachandran, K.M. dan Tsokos, C. P. 2009. *"Mathematical Statistics with Applications"*. Elsefier Academic Press. USA.

[9] Al-Jammal, Z, Y. 2010. *"Multiple Regression Model Seletion by Information Criteria"*. **Iraqi Journal of Statistical Science**, Hal. 1-12.

[10] Kementrian Kesehatan RI. 2014. *"Situasi dan Analisis HIV AIDS"*. Pusat Data dan Informasi Kementrian Kesehatan RI. Jakarta.