TUGAS AKHIR

MODEL REGRESI PROBIT BIVARIAT PADA KASUS PENDERITA HIV DAN AIDS DI JAWA TIMUR

BIVARIATE PROBIT REGRESSION MODEL AT THE CASE OF HIV AND AIDS PATIENTS IN EAST JAVA

Oleh:
Bella Yuliatin Puspita Sari
1211 100 025

Dosen Pembimbing:
Dra. Farida Agustini Widjajati, MS

JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA

2015

Abstrak

Model regresi probit bivariat adalah model regresi probit yang menggunakan dua variabel respon yang bersifat kualitatif atau berkategori. Empat ariabel prediktor yang digunakan sebagai faktor-faktor yang mempengaruhi variabel respon adalah bersifat kontinu. 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 jumlah penderita HIV dan AIDS tiap kabupaten dan kota di Jawa Timur pada tahun 2013 sebagai variabel responnya untuk membentuk model regresi probit bivariat dan menghasikan variabel prediktor persentase umur 25-49 tahun (x_1), persentase Askeskin atau Jamkesmas (x_2) dan persentase jumlah sarana kesehatan (x_4) sebagai faktor yang berpengaruh signifikan terhadap HIV dan AIDS.

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

Latar Belakang | Rumusan Masalah | Batasan Masalah | Tujuan | Manfaat

Latar Belakang

Berdasarkan penelitian sebelumnya

- Chen dan Hamori tahun 2010

Bivariate Probit Analysis of Differences Between Male and Female Formal Employment in Urban China

- Dhewy dan Purhadi 2014

Pemodelan Bivariate Poisson Regression dengan Kovarian Merupakan Fungsi dari Variabel Bebas dengan Studi Kasus Jumlah HIV dan AIDS di Jawa Timur pada tahun 2012

- Romadhona dan Ratnasari 2014

Pemodelan Pemberian Imunisasi Dasar dan ASI Eksklusif dengan Pendekatan Model Probit Biner Bivariat

Rumusan Masalah

1. Bagaimana cara menentukan estimasi parameter untuk model probit bivariat?
2. Bagaimana memodelkan regresi probit bivariat pada faktor-faktor yang mempengaruhi jumlah kasus HIV dan AIDS di Jawa Timur?

Batasan Masalah

1. Analisis estimasi parameter dalam penelitian ini dibatasi hanya untuk mengetahui langkah-langkah dalam mendapatkan parameter pada regresi probit bivariat.
2. Pembentukan model regresi probit bivariat menggunakan kasus HIV dan AIDS sebagai dua variabel respon dan menggunakan empat faktor yang mempengaruhi HIV dan AIDS berupa persentase usia, jaminan kesehatan, penyuluhan kesehatan, dan sarana kesehatan sebagai variabel prediktor.
3. Sumber data yang digunakan untuk membentuk model probit bivariat adalah berupa data sekunder Profil Kesehatan Jawa Timur tahun 2013 yang diperoleh dari Dinas Kesehatan Jawa Timur.
4. Software yang digunakan untuk menunjang tugas akhir ini adalah StataSE12, SPSS dan Matlab.

Tujuan

1. Mendapatkan langkah-langkah untuk mengestimasi parameter dalam model probit bivariat.
2. Memodelkan regresi probit bivariat untuk faktor-faktor yang mempengaruhi jumlah kasus HIV dan AIDS di Jawa Timur.

Manfaat

1. Memberikan bahan masukan untuk pemerintah khususnya pemerintah Jawa Timur dalam menangani kasus penyakit HIV dan AIDS agar jumlah penderita tidak semakin meningkat dengan melihat variabel mana yang berpengaruh signifikan dalam penyebab dan penyebaran penyakit HIV AIDS.
2. Dapat memberikan informasi dan wawasan tentang pembentukan regresi probit bivariat serta aplikasinya dalam bidang kesehatan khususnya pada kasus jumlah penderita penyakit HIV dan AIDS di Jawa Timur dengan menggunakan data kasus pada tahun 2013.

Analisis Regresi

Analisis regresi berkaitan dengan studi mengenai ketergantungan satu *variabel respon* terhadap satu atau lebih *variabel prediktor*, dengan tujuan untuk mengestimasi dan/ atau memperkirakan nilai rata-rata populasi variabel respon dari nilai yang diketahui pada variabel prediktor [7]. Persamaan regresi linear yaitu:

$$\hat{y} = a + bx$$

dengan a menyatakan konstanta, b adalah koefisien variabel prediktor x dan $\hat{y}$ digunakan untuk membedakan antara nilai estimasi yang dihasilkan garis regresi dan nilai pengamatan y yang sebenarnya sebagai variabel respon untuk suatu nilai x tertentu sebagai variabel prediktornya [8].

Pada umumnya model regresi biasa melibatkan variabel respon yang bersifat kuantitatif dan variabel prediktornya dapat bersifat kuantitatif maupun kualitatif ataupun gabungan dari keduanya. Namun sebuah variabel respon atau lebih dapat juga bersifat kualitatif seperti pemilihan keputusan “ya” atau “tidak”. Sehingga diperlukan pula model regresi dengan variabel respon yang bersifat kualitatif seperti model regresi probit [9].

Distribusi Bernoulli

Jika dalam sebuah percobaan hanya menghasilkan dua kejadian yaitu E adalah kejadian "sukses" dan komplemennya E' adalah kejadian "gagal" maka E terjadi dengan peluang $p = P(E)$ dan peluang terjadinya E' adalah $q = P(E') = 1 - p$. Sebuah variabel random Y yang diasumsikan hanya bernilai 0 atau 1 disebut sebagai *variabel Bernoulli* dan hasil yang didapatkan dari percobaan hanya dua tipe disebut *percobaan Bernoulli*.

$$Y(e) = \begin{cases} 1 \; jika \; e \in E \\ 0 \; jika \; e \in E' \end{cases}$$

Probability density function atau pdf dari Y adalah

$$f(y) = p^y q^{1-y} \qquad y = 0,1 \qquad (2.1)$$

dengan mean p dan varians pq. Sehingga pdf dari $f(0) = q$ dan $f(1) = p$

Distribusi Bernoulli merupakan proses dari distribusi binomial [10].

Distribusi Binomial

Distribusi binomial merupakan distribusi dengan variabel acak diskrit untuk n kali percobaan Bernoulli yang saling independen atau saling bebas dimana hasil dari suatu percobaan tidak mempengaruhi percobaan yang lainnya. Pdf dari suatu variabel acak Y berdistribusi binomial adalah

$$b(y;n,p) = \binom{n}{y} p^y (1-p)^{n-y} \qquad y = 0,1,\dots,n$$

dengan y adalah jumlah sukses dalam n percobaan dan p yaitu peluang sukses untuk setiap percobaan.

Distribusi binomial dinotasikan sebagai $Y \sim BIN(n,p)$ dengan parameter n dan p. Untuk *Cumulative Distribution Function* (CDF) dari distribusi binomial yaitu

$$B(y;n,p) = \sum_{k=0}^{y} b(k;n,p) \qquad y = 0,1,2,\dots,n$$

Mean dan varian untuk distribusi binomial masing-masing adalah np dan $np(1-p)$ [10].

Distribusi Multinomial

Untuk $E_1, E_2, \dots E_k, E_{k+1}$ kejadian dan diberikan $p = P(E_i)$ untuk $i = 1,2, \dots, k+1$ pada percobaan yang saling independen sehingga diberikan Y_i adalah jumlah kejadian yang telah terjadi pada E_i berdistribusi multinomial dengan pdf

$$f(y_1, y_2, \dots, y_k) = \frac{n!}{y_1! y_2! \dots y_{k+1}!} p_1^{y_1} p_2^{y_2} \dots p_{k+1}^{y_{k+1}}$$

dimana

$$y_{k+1} = n - \sum_{i=1}^{k} y_i$$

$$p_{k+1} = 1 - \sum_{i=1}^{k} p_i$$

Mean untuk distribusi multinomial adalah $\mu = np_i$ dan varian $Var(Y_i) = np_i(1 - p_i)$ [10].

Distribusi Normal

Jika y suatu variabel random normal, maka pdf dari y adalah:

$$f(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2} ; -\infty < y < \infty$$

Parameter yang digunakan dalam distribusi normal adalah mean μ (dengan $-\infty < \mu < \infty$) dan varians σ^2 (dengan $0 < \sigma^2 < \infty$) dengan notasi $y \sim N(\mu : \sigma^2)$. Sedangkan untuk CDF dari y didefinisikan sebagai probabilitas bahwa variabel random normal y kurang atau sama dengan suatu nilai a yang dirumuskan:

$$P\{y \leq a\} = F(a) = \int_{-\infty}^{a} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2} dy$$

Distribusi normal standar adalah distribusi normal dengan parameter $\mu = 0$ dan $\sigma^2 = 1$. Sehingga pdf dari distribusi normal standar yaitu:

$$\emptyset(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z)^2} ; -\infty < z < \infty$$

dengan $z = \frac{y-\mu}{\sigma}$ sehingga CDF dari distribusi normal standar adalah [10]

$$\Phi(z) = \int_{-\infty}^{z} \emptyset(t)\, dt = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(t)^2} dt$$

Uji Tidak Multikolinieritas

Keadaan multikolinieritas dapat dideteksi dengan uji korelasi antar variabel prediktor.

Hipotesis:

$H_0: \rho_{x_1 x_2} \neq 0$

$H_1: \rho_{x_1 x_2} = 0$

Statistik Uji:

$$z = \frac{1}{2} \ln \left| \frac{(1+r)}{(1-r)} \right|$$

Dimana rumus r sebagai koefisien korelasi adalah sebagai berikut:

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

Tolak H_0 jika $z < z_{\propto/2} = 1{,}96$. Jika besarnya korelasi melebihi 0,8 atau 80% maka korelasi berpasangan diantara dua variabel prediktor dikatakan tinggi.

Uji Dependensi

Untuk mengetahui hubungan korelasi antar variabel yang berupa data kategorik, dapat menggunakan uji dependensi. Syarat uji dependesi adalah 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:

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

dimana

O_{ij} = banyaknya data yang termasuk dalam sel ke- i, j

E_{ij} = nilai harapan (taksiran) dari O_{ij} dengan rumus

$$E_{ij} = \frac{n_i n_j}{n}$$

Tolak H_0 jika nilai $X^2 > X^2_{\propto,(I-1)(J-1)}$ [11].

Probit Univariat

Misalkan ada variabel respon Y yang diasumsikan berasal dari variabel y^* sebagai model regresi probit univariat yang ditulis

$$y^* = \beta^T x + \varepsilon$$

dimana

$\beta = [\beta_0 \quad \beta_1 \cdots \quad \beta_q]^T$

$x = [1 \quad x_1 \quad x_2 \cdots \quad x_q]^T$

q = banyaknya variabel prediktor

dengan x adalah vektor dari variabel prediktor dan parameter β adalah vektor parameter koefisien dari variabel prediktor yang masing-masing x dan β berukuran $(q+1) \times 1$. Serta ε adalah faktor galat yang diasumsikan berdistribusi normal standar dengan $\mu = 0$ dan $\sigma^2 = 1$ sehingga dinotasikan $y^* \sim N(\beta^T x, 1)$ dengan mean $\beta^T x$ dan varians satu [6]. Pdf dari variabel y^* berdistribusi normal adalah:

$$\emptyset(y^*) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y^* - \beta^T x)^2}$$

dan CDF-nya yaitu:

$$\Phi(y^*) = \int_{-\infty}^{y^*} \emptyset(t)\, dt = \int_{-\infty}^{y^*} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(t - \beta^T x)^2}\, dt$$

Pembentukan kategori pada variabel respon Y diberikan *threshold* atau ambang batas tertentu misal r sehingga dapat ditulis sebagai berikut:

$Y = 0$ jika $y^* \leq r$

$Y = 1$ jika $y^* > r$

dapat dicari probabilitas untuk $Y = 0$ atau probabilitas dari kejadian gagal $(q(x))$ dengan mensubtitusi nilai y^* sebagai berikut:

$$\begin{aligned} P(Y = 0) &= P(y^* \leq r) \\ &= P(\beta^T x + \varepsilon \leq r) \\ &= P(\varepsilon \leq r - \beta^T x) \\ &= \Phi(r - \beta^T x) = q(x) \end{aligned} \tag{2.2}$$

dan probabilitas untuk $Y = 1$ atau probabilitas dari kejadian sukses $(p(x))$ adalah:

$$\begin{aligned} P(Y = 0) &= P(y^* > r) \\ &= 1 - P(y^* \leq r) \\ &= 1 - P(\varepsilon \leq r - \beta^T x) \\ &= 1 - \Phi(r - \beta^T x) = p(x) \end{aligned} \tag{2.3}$$

Dari persamaan (2.1) dan (2.2) diperoleh probabilitas untuk kejadian sukses adalah $p(x) = 1 - \Phi(r - \beta^T x)$ [6].

Estimasi Pararameter Probit Univariat

Untuk melakukan estimasi parameter dalam regresi probit digunakan metode *Maximum Likelihood Estimation* (MLE) karena model probit telah diketahui distribusinya. Langkah-langkah estimasi parameter untuk model probit biner univariat adalah:

1. Menentukan *n* sampel random misal $Y_1, Y_2, \dots, Y_n$ yang saling IID (*Independent Identically Distribution*).
2. Membentuk fungsi *likelihood* berdistribusi *Bernoulli* yaitu $L(\beta) = \prod_{i=1}^{n} \left(p(x_i)\right)^{y_i} \left(q(x_i)\right)^{1-y_i}$ kemudian melakukan transformasi *ln* terhadap fungsi tersebut.
3. Memaksimumkan fungsi *ln likelihood* dengan menurunkan terhadap parameter β kemudian disamadengankan nol.
4. Jika hasil yang didapatkan tidak *closed form* maka digunakan digunakan metode numerik menggunakan iterasi Newton Raphson.
5. Iterasi akan berhenti jika $\left|\beta^{(m)} - \beta^{(m-1)}\right| \leq \Theta$ dengan Θ adalah bilangan yang sangat kecil.

Probit Bivariat

Misalkan ada dua variabel respon Y_1 dan 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:

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

Dalam pembentukan kategori untuk masing-masing variabel respon pada model probit bivariat tidak berbeda dengan proses pembentukan kategori untuk variabel respon model regresi probit univariat. Terlebih dahulu ditentukan *threshold* tertentu pada masing-masing variabel yang tidak teramati yaitu 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 *threshold* untuk masing-masing variabel respon dalam model probit biner bivariat [6].

Estimasi Pararameter Probit Bivariat

Untuk membentuk model probit bivariat dilakukan estimasi parameter dengan menggunakan metode *Maximum Likelihood Estimation* (MLE). Dalam mendapatkan estimasi model probit bivariat didasarkan pada pembentukan tabel kontingensi terlebih dahulu. Tabel kontingensi yang dibuat adalah 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 yang dicari, diperlukan beberapa konsep dasar tentang turunan vektor menggunakan Lemma 2.1 yaitu [6]:

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

Uji Signifikansi

Uji Serentak

Hipotesis:

$H_0: \beta_1 = \beta_2 = \cdots = \beta_q = 0$

H_1: Paling tidak ada satu $\beta_i \neq 0$ dengan $i = 1,2, \dots q$

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 $W \sim X^2_{\propto,df}$ apabila $n \to \infty$. Tolak H_0 jika $G^2 > X^2_{\propto,df}$ atau *p-value* $<\propto$ dimana derajat bebas (df) yaitu banyaknya variabel prediktor dalam model.

Uji Parsial

Hipotesis:

$H_0: \beta_i = 0$

$H_1: \beta_i \neq 0$ dengan $i = 1,2, \dots q$

Statistik uji:

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

Statistik uji Wald mengikuti distribusi normal standar. Tolak H_0 jika $|W| > Z_{\alpha/2}$ atau jika *p-value* $<\propto$.

Pemilihan Model Terbaik

Untuk memilih model terbaik dari model probit bivariat yang telah terbentuk, dapat dilihat dari nilai *Akaike Information Criterion* (AIC). Semakin kecil nilai AIC maka model semakin baik [12]. Perhitungan AIC dirumuskan sebagai berikut:

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

dengan

$L(\hat{\theta})$ = nilai maksimum fungsi *likelihood*

p = banyak parameter dalam model

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 *AcquiredImmune 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. Pengidap HIV memerlukan pengobatan dengan *Antiretroviral* (ARV) untuk menurunkan jumlah virus HIV di dalam tubuh agar tidak masuk ke stadium AIDS. Sedangkan pengidap AIDS memerlukan pengobatan ARV untuk mencegah terjadinya infeksi oportunistik dengan berbagai komplikasinya [13].

Studi Literatur

Pengumpulan dan Pengolahan Data

Estimasi Parameter dan Pembentukan Model Regresi Probit Bivariat

Pengambilan Kesimpulan

Konsultasi dan Penyusunan Buku

Studi Literatur

Tahap ini dilakukan untuk menentukan permasalahan dan tujuan yang diambil dalam tugas akhir. Kemudian perlu dilakukan studi literatur untuk penyelesaian tugas akhir terkait masalah dan tujuan yang telah dirumuskan. Studi literatur diperoleh melalui jurnal ilmiah, paper, buku-buku serta artikel yang ada di internet.

Pengumpulan dan Pengolahan Data (1)

Data yang digunakan untuk menyelesaikan tugas akhir ini berupa data sekunder dari Profil Kesehatan Jawa Timur tahun 2013 yang berasal dari Dinas Kesehatan Jawa Timur.

Variabel yang digunakan dalam penelitian ini terdiri dari dua variabel respon dan empat variabel prediktor. Penjelasan dari masing-masing variabel adalah sebagai berikut:

Variabel Respon

Dua variabel respon yang digunakan adalah HIV dan AIDS dengan kategori sebagai berikut:

Kategori HIV

Kode 0 = Kategori Rendah, jika *case rate* tiap kabupaten/kota berada dibawah *case rate* provinsi untuk HIV

Kode 1 = Kategori Tinggi, jika *case rate* tiap kabupaten/kota berada diatas *case rate* provinsi untuk HIV

Kategori AIDS

Kode 0 = Kategori Rendah, jika *case rate* tiap kabupaten/kota berada dibawah *case rate* provinsi untuk AIDS

Kode 1 = Kategori Tinggi, jika *case rate* tiap kabupaten/kota berada diatas *case rate* provinsi untuk AIDS

Pengumpulan dan Pengolahan Data (2)

Variabel Prediktor

Faktor-faktor yang mempengaruhi penyebaran penyakit HIV dan AIDS digunakan sebagai variabel prediktor, dengan penjelasan berikut:

1. Persentase jumlah kelompok umur 25-49 tahun terhadap jumlah penduduk (x_1).
2. Persentase jumlah Askeskin atau Jamkesmas terhadap jumlah penduduk (x_2).
3. Persentase jumlah penyuluhan kesehatan terhadap jumlah penduduk (x_3).
4. Persentase jumlah sarana kesehatan terhadap jumlah penduduk (x_4).

Estimasi Parameter dan Pembentukan Model

Dalam tahap ini dilakukan pembahasan mengenai cara mendapatkan estimasi parameter menggunakan metode *Maximum Likelihood Estimation* (MLE) dan pembentukan model regresi probit bivariat.

Pengambilan Kesimpulan

Jika dalam langkah menganalisis dan memodelkan regresi probit bivariat telah selesai dilakukan maka diambil kesimpulan yang sesuai dengan tujuan dari tugas akhir.

Konsultasi dan Penyusunan Buku Tugas Akhir

Konsultasi dilakukan dengan dosen pembimbing guna memperoleh saran dan perbaikan dalam proses awal pengerjaan sampai penyelesaian tugas akhir dan kemudian dilakukan penyusunan buku untuk menjelaskan hasil dari tugas akhir yang telah dikerjakan.

Hasil dan Pembahasan

Analisis Data Penyakit HIV dan AIDS di Jawa Timur (1)

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.

Kategori	Frekuensi	Persen
HIV di bawah case rate	21	55,3
HIV di atas case rate	17	44,7
Total	38	100

Kategori	Frekuensi	Persen
AIDS di bawah case rate	26	68,4
AIDS di atas case rate	12	31,6
Total	38	100

Kasus HIV	Kasus AIDS		Total
	Rendah	Tinggi	
Rendah	19	2	21
Tinggi	7	10	17
Total	26	12	38

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 ($_1$), persentase Askeskin atau Jamkesmas (X_2), persentase penyuluhan kesehatan (X_3), dan persentase sarana kesehatan (X_4).

Analisis Data Penyakit HIV dan AIDS di Jawa Timur (2)

Statistik deskriptif untuk tiap variabel respon dan variabel prediktor tersebut dapat dilihat pada tabel berikut:

Variabel	Mean	Standar Deviasi	Min	Max
Y_1	0,447	0,504	0	1
Y_2	0,316	0,471	0	1
X_1	18,49	1,06	15,68	21,05
X_2	0,34206	0,14817	0,00038	0,6178
X_3	1,171	1,126	0,003	4.95
X_4	0,004	0,002	0,002	0,010

Pemodelan Regresi Probit Bivariat (1)

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

Hubungan Antar Variabel Respon

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.

Pemodelan Regresi Probit Bivariat (2)

Uji Multikolinieritas Antar Variabel Prediktor

Empat variabel prediktor harus memenuhi syarat bahwa tidak ada korelasi antar variabel prediktor yang digunakan sebagai berikut.

Berdasarkan data yang telah diperoleh, dapat dihitung 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 diperoleh nilai koefisien korelasi antar variabel prediktor pada tabel di bawah. Masing-masing koefisien korelasi antar variabel prediktor tersebut tidak ada yang melebihi 80% sehingga dapat disimpulkan tidak terjadi multikolinieritas antar variabel prediktor.

Korelasi	X1	X2	X3	X4
X1		-0,400	-0,193	0,143
X2	-0,400		-0,016	-0,377
X3	-0,193	-0,016		0,378
X4	0,143	-0,377	0,378	

Estimasi Parameter Probit Bivariat (1)

Untuk mengestimasi parameter probit bivariat dapat dilakukan dengan metode *Maximum Likelihood Estimation* (MLE). Terlebih dahulu dibentuk tabel kontingensi (2×2) yang melibatkan variabel respon Y_1 dan Y_2 dan kemudian ditentukan probabilitas untuk tiap sel.

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

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

Estimasi Parameter Probit Bivariat (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:

$$p_1(x) = 1 - \Phi(r - \beta_1^T x) \quad (5)$$
$$p_2(x) = 1 - \Phi(s - \beta_2^T x) \quad (6)$$

Estimasi Parameter Probit Bivariat (3)

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:

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

Estimasi Parameter Probit Bivariat (4)

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:

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

Estimasi Parameter Probit Bivariat (5)

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

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

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:

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

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

dimana

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

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

Estimasi Parameter Probit Bivariat (6)

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

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

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

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

$$= -\emptyset(r - \beta_1^T x_i)(-x_i) = x_i \emptyset(z_{1i}) \qquad (14)$$

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

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

Persamaan (15) dapat disederhanakan menjadi:

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

Estimasi Parameter Probit Bivariat (7)

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:

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

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

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 turunan kedua sebagai elemen dari matriks Hessian.

Pemodelan Regresi Probit Bivariat (4)

Model Regresi Probit Bivariat

Model probit bivariat dibentuk dengan menggunakan data kasus penyakit HIV dan AIDS di Jawa Timur pada tahun 2013 sebagai dua variabel respon dengan melibatkan empat variabel prediktor yang diduga berpengaruh terhadap variabel respon. Kemudian model yang telah dibentuk diuji secara serentak dan parsial.

Dengan menggunakan bantuan *software* StataSE12 diperoleh model awal probit bivariat sebagai berikut:

$\hat{y}_1^* = -10{,}422 + 0{,}591x_1 - 0{,}402x_2 - 0{,}397x_3 - 22{,}343x_4$

$\hat{y}_2^* = 4{,}238 - 0{,}075x_1 - 3{,}458x_2 - 0{,}396x_3 - 490{,}929x_4$

Uji Serentak

Hipotesis:

$H_0: \beta_{11} = \beta_{12} = \beta_{13} = \beta_{14} = 0$ dan $\beta_{21} = \beta_{22} = \beta_{23} = \beta_{24} = 0$

H_1: Paling tidak ada satu $\beta_{kl} \neq 0$; k = 1,2 dan l = 1,2,3,4

Statistik Uji:

$G^2 = 2\left[\ln L(\hat{\Omega}) - \ln L(\hat{\omega})\right] = 2[-32{,}743502 - (-44{,}250417)] = 23{,}104$

Nilai G^2 adalah sebesar 23,014 yang kemudian dibandingkan dengan nilai $X^2_{0,05;8} = 15{,}51$ ($G^2 > X^2_{0,05;8}$) dengan keputusan tolak H_0. Jadi kesimpulan yang dapat diambil adalah paling tidak ada satu variabel yang signifikan terhadap variabel respon.

Pemodelan Regresi Probit Bivariat (5)

Uji Parsial

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

Hipotesis:

$H_0: \beta_{11} = 0$

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

Statistik uji:

$$W = \frac{\hat{\beta}_{11}}{SE(\hat{\beta}_{11})} = \frac{0{,}591}{0{,}252} = 2{,}34$$

Dengan cara yang sama diperoleh statistik uji untuk masing-masing parameter pada berikut:

Parameter	SE	W
$\hat{\beta}_{12} = -0{,}402$	1,573	−0,26
$\hat{\beta}_{13} = -0{,}397$	0,274	−1,45
$\hat{\beta}_{14} = -22{,}343$	170,786	−0,13
$\hat{\beta}_{21} = -0{,}075$	0,238	−0,32
$\hat{\beta}_{22} = -3{,}458$	1,781	−1,94
$\hat{\beta}_{23} = -0{,}396$	0,282	−1,41
$\hat{\beta}_{24} = -490{,}929$	224,114	−2,19

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

Pemodelan Regresi Probit Bivariat (6)

Pemilihan Model Terbaik

Untuk mendapatkan model terbaik dilakukan dengan menggunakan metode *backward* yaitu dengan cara mengkombinasikan semua kemungkinan model yaitu sebanyak $2^q - 1$ dimana q adalah banyaknya variabel prediktor kemudian diambil model dengan nilai AIC terkecil.

Dengan demikian untuk mendapatkan model terbaik probit bivariat, pemodelan dilakukan sebanyak 15 kali sebagai berikut:

No	Variabel Prediktor	Nilai AIC	Eliminasi
1.	x_1, x_2, x_3, x_4	87,487	
2.	x_2, x_3, x_4	93,448	x_1
3.	x_1, x_3, x_4	91,286	x_2
4.	$\mathbf{x_1, x_2, x_4}$	**86,152**	$\mathbf{x_3}$
5.	x_1, x_2, x_3	90,261	x_4
6.	x_3, x_4	94,070	x_1, x_2
7.	x_2, x_4	92,280	x_1, x_3
8.	x_2, x_3	94,007	x_1, x_4
9.	x_1, x_4	90,743	x_2, x_3
10.	x_1, x_3	90,078	x_2, x_4
11.	x_1, x_2	92,930	x_3, x_4
12.	x_4	94,187	x_1, x_2, x_3
13.	x_3	92,549	x_1, x_2, x_4
14.	x_2	95,827	x_1, x_3, x_4
15.	x_1	91,813	x_2, x_3, x_4

Pemodelan Regresi Probit Bivariat (7)

Model terbaik dipilih dari nilai AIC terkecil adalah dengan tiga variabel prediktor. Perhitungan AIC untuk model terbaik adalah sebagai berikut:

$$AIC = -2lnL(\hat{\theta}) + 2p = -2(-34.076215) + 2(9) = 68{,}15243 + 18 = 86{,}152$$

Hasil model terbaik probit bivariat yaitu:

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

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

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 ($x_4 = 0{,}004$), 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 $s = 0$, didapatkan hasil dari probabilitas untuk setiap kategori menggunakan Matlab 2013 sebagai berikut:

$\hat{p}_{00}^*(x) = \Phi(z_1, z_2) = \Phi(-1{,}453; 0{,}001) = 0{,}0567$

$\hat{p}_{01}^*(x) = \Phi(z_1) - \Phi(z_1, z_2) = \Phi(-1{,}453) - \Phi(-1{,}453; 0{,}001) = 0{,}0187$

$\hat{p}_{10}^*(x) = \Phi(z_2) - \Phi(z_1, z_2) = \Phi(-0{,}001) - \Phi(-1{,}453; 0{,}001) = 0{,}4435$

$\hat{p}_{11}^*(x) = 1 - \Phi(z_1) - \Phi(z_2) + \Phi(z_1, z_2) = 1 - \Phi(1{,}453) - \Phi(-0{,}001) + \Phi(-1{,}453; 0{,}001) = 0{,}4812$

Berdasarkan nilai probabilitas dari tiap kategori yang telah didapatkan, dapat disimpulkan bahwa Kota Surabaya mempunyai probabilitas cukup tinggi yaitu sebesar 0,4812 untuk berada dalam kategori HIV yang tinggi dan kategori AIDS yang juga tinggi.

Pemodelan Regresi Probit Bivariat (8)

Uji Serentak

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

Statistisk uji:

$G^2 = 2\left[\ln L(\widehat{\Omega}) - \ln L(\widehat{\omega})\right] = 2[-34{,}076 - (-44{,}250)] = 20{,}348$

Nilai G^2 sebesar 20,348 kemudian dibandingkan dengan nilai $\chi^2_{0{,}05;6} = 12{,}59$ ($G^2 > \chi^2_{0{,}05;6}$) dengan keputusan tolak H_0.

Jadi kesimpulan yang dapat diambil adalah paling tidak ada satu variabel yang signifikan terhadap variabel respon.

Pemodelan Regresi Probit Bivariat (9)

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 berikut:

Parameter	SE	W
$\hat{\beta}_{12} = -0{,}664$	1,601	−0,41
$\hat{\beta}_{14} = -136{,}298$	147,188	−0,93
$\hat{\beta}_{21} = -0{,}015$	0,221	−0,07
$\hat{\beta}_{22} = -3{,}656$	1,755	−2,08
$\hat{\beta}_{24} = -557{,}638$	242,198	−2,30

Tolak H_0 jika $|W| > Z_{\alpha/2}$. Berdasarkan perhitungan Uji Wald tersebut diperoleh hasil uji parsial untuk tiap variabel prediktor pada variabel respon Y_1 dan Y_2 dengan taraf signifikan $\propto = 5\%$ yaitu variabel kelompok umur 25-49 tahun (x_1) berpengaruh signifikan terhadap HIV. Sementara variabel Askeskin atau Jamkesmas (x_2) dan variabel sarana kesehatan (x_4) berpengaruh signifikan terhadap AIDS.

Pemodelan Regresi Probit Bivariat (10)

Efek Marginal

Efek marginal digunakan untuk melihat besarnya pengaruh perubahan suatu variabel prediktor terhadap variabel respon dengan asumsi variabel lainnya konstan. Dalam hal ini diambil sampel untuk data kota Surabaya. Dengan menggunakan Matlab 2013 diperoleh hasil efek marginal untuk tiap variabel prediktor di Kota Surabaya. Penjelasannya sebagai berikut:

Persentase Umur 25-49 tahun (x_1)

Besarnya efek marginal variabel persentase umur 25-49 tahun (x_1) dengan persentase kelompok umur 25-49 tahun $(x_1 = 20{,}87)$, persentase Askeskin atau Jamkesmas $(x_2 = 0{,}13753)$ dan persentase jumlah sarana kesehatan $(x_4 = 0{,}004)$ adalah sebagai berikut:

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

$$\frac{\partial \hat{p}_{01}^*}{\partial x_1} = \frac{\partial \big(\Phi(z_1) - \Phi(z_1, z_2)\big)}{\partial x_1} = -\hat{\beta}_{1,1}^* \phi(z_1) + \hat{\beta}_{1,1}^* \varphi_1 + \hat{\beta}_{2,1}^* \varphi_2 = -0{,}612\phi(z_1) + 0{,}612\varphi_1 + (-0{,}015)\varphi_2 = -0{,}0439$$

$$\frac{\partial \hat{p}_{10}^*}{\partial x_1} = \frac{\partial \big(\Phi(z_2) - \Phi(z_1, z_2)\big)}{\partial x_1} = -\hat{\beta}_{2,1}^* \phi(z_2) + \hat{\beta}_{1,1}^* \varphi_1 + \hat{\beta}_{2,1}^* \varphi_2 = -(-0{,}015)\phi(z_1) + 0{,}612\varphi_1 + (-0{,}015)\varphi_2 = 0{,}0490$$

$$\frac{\partial \hat{p}_{11}^*}{\partial x_1} = \frac{\partial (1 - \Phi(z_1) - \Phi(z_2) + \Phi(z_1, z_2))}{\partial x_1} = \hat{\beta}_{1,1}^* \phi(z_1) + \hat{\beta}_{2,1}^* \phi(z_2) - \hat{\beta}_{1,1}^* \varphi_1 - \hat{\beta}_{2,1}^* \varphi_2 = 0{,}612\phi(z_1) + (-0{,}015)\phi(z_2) - 0{,}612\varphi_1 + (-0{,}015)\varphi_2 = 0{,}0379$$

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.

Pemodelan Regresi Probit Bivariat (11)

Persentase Askeskin atau Jamkesmas (x_2)

Besarnya efek marginal variabel persentase Askeskin atau Jamkesmas (x_2) dengan persentase kelompok umur 25-49 tahun $(x_1 = 20{,}87)$, persentase Askeskin atau Jamkesmas $(x_2 = 0{,}13753)$ dan persentase jumlah sarana kesehatan $(x_4 = 0{,}004)$ adalah sebagai berikut:

$$\frac{\partial \hat{p}_{00}^*}{\partial x_2} = \frac{\partial \Phi(z_1, z_2)}{\partial x_2} = -\hat{\beta}_{1,2}^* \varphi_1 - \hat{\beta}_{2,2}^* \varphi_2 = -(-0{,}664)\varphi_1 - (-3{,}656)\varphi_2 = 0{,}1571$$

$$\frac{\partial \hat{p}_{01}^*}{\partial x_2} = \frac{\partial \big(\Phi(z_1) - \Phi(z_1, z_2)\big)}{\partial x_2} = -\hat{\beta}_{1,2}^* \phi(z_1) + \hat{\beta}_{1,2}^* \varphi_1 + \hat{\beta}_{2,2}^* \varphi_2 = -(-0{,}664)\phi(z_1) + (-0{,}664)\varphi_1 + (-3{,}656)\varphi_2 = -0{,}0628$$

$$\frac{\partial \hat{p}_{10}^*}{\partial x_2} = \frac{\partial \big(\Phi(z_2) - \Phi(z_1, z_2)\big)}{\partial x_2} = -\hat{\beta}_{2,2}^* \phi(z_2) + \hat{\beta}_{1,2}^* \varphi_1 + \hat{\beta}_{2,2}^* \varphi_2 = -(-0{,}664)\phi(z_1) + (-0{,}664)\varphi_1 + (-0{,}664)\varphi_2 = 1{,}3014$$

$$\frac{\partial \hat{p}_{11}^*}{\partial x_2} = \frac{\partial (1 - \Phi(z_1) - \Phi(z_2) + \Phi(z_1, z_2))}{\partial x_2} = \hat{\beta}_{1,2}^* \phi(z_1) + \hat{\beta}_{2,2}^* \phi(z_2) - \hat{\beta}_{1,2}^* \varphi_1 - \hat{\beta}_{2,2}^* \varphi_2 = (-0{,}664)\phi(z_1) + (-0{,}664)\phi(z_2) - (-0{,}664)\varphi_1 + (-0{,}664)\varphi_2 = -1{,}3958$$

Efek marginal variabel persentase Askeskin atau Jamkesmas (x_2) terhadap $\hat{p}_{11}^*$ adalah sebesar -1,3958 yang berarti bahwa perubahan persentase Askeskin atau Jamkesmas (x_2) sebesar satu satuan akan menurunkan sebesar 1,3958 terhadap probabilitas HIV tinggi dan AIDS yang juga tinggi. Variabel persentase Askeskin atau Jamkesmas hanya signifikan terhadap variabel Y_2 yaitu kasus AIDS.

Pemodelan Regresi Probit Bivariat (12)

Persentase Jumlah Sarana Kesehatan (x_4)

Besarnya efek marginal variabel persentase jumlah sarana kesehatan (x_4) dengan persentase kelompok umur 25-49 tahun $(x_1 = 20{,}87)$, persentase Askeskin atau Jamkesmas $(x_2 = 0{,}13753)$ dan persentase jumlah sarana kesehatan $(x_4 = 0{,}004)$ adalah sebagai berikut:

$$\frac{\partial \hat{p}_{00}^*}{\partial x_4} = \frac{\partial \Phi(z_1, z_2)}{\partial x_4} = -\hat{\beta}_{1,4}^* \varphi_1 - \hat{\beta}_{2,4}^* \varphi_2 = -(-136{,}298)\varphi_1 - (-557{,}638)\varphi_2 = 26{,}4533$$

$$\frac{\partial \hat{p}_{01}^*}{\partial x_4} = \frac{\partial\big(\Phi(z_1) - \Phi(z_1, z_2)\big)}{\partial x_4} = -\hat{\beta}_{1,4}^* \phi(z_1) + \hat{\beta}_{1,4}^* \varphi_1 + \hat{\beta}_{2,4}^* \varphi_2 = -(-136{,}298)\phi(z_1) + (-136{,}298)\varphi_1 + (-557{,}638)\varphi_2 = -7{,}0872$$

$$\frac{\partial \hat{p}_{10}^*}{\partial x_4} = \frac{\partial\big(\Phi(z_2) - \Phi(z_1, z_2)\big)}{\partial x_4} = -\hat{\beta}_{2,4}^* \phi(z_2) + \hat{\beta}_{1,4}^* \varphi_1 + \hat{\beta}_{2,4}^* \varphi_2 = -(-557{,}638)\phi(z_1) + (-136{,}298)\varphi_1 + (-557{,}638)\varphi_2 = 196{,}0121$$

$$\frac{\partial \hat{p}_{11}^*}{\partial x_4} = \frac{\partial\big(1 - \Phi(z_1) - \Phi(z_2) + \Phi(z_1, z_2)\big)}{\partial x_4} = \hat{\beta}_{1,4}^* \phi(z_1) + \hat{\beta}_{2,4}^* \phi(z_2) - \hat{\beta}_{1,4}^* \varphi_1 - \hat{\beta}_{2,4}^* \varphi_2 = (-136{,}298)\phi(z_1) + (-557{,}638)\phi(z_2) - (-136{,}298)\varphi_1 + (-557{,}638)\varphi_2 = -215{,}3782$$

Efek marginal variabel persentase jumlah sarana kesehatan (x_4) terhadap $\hat{p}_{11}^*$ adalah sebesar -215,3782 yang berarti bahwa perubahan persentase jumlah sarana kesehatan (x_4) sebesar satu satuan akan menurunkan sebesar 215,3782 terhadap probabilitas HIV tinggi dan AIDS yang juga tinggi. Variabel persentase jumlah sarana kesehatan hanya signifikan terhadap variabel Y_2 yaitu kasus AIDS.

Pemodelan Regresi Probit Bivariat (13)

Ketepatan Klasifikasi Model Probit Bivariat

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 4.12. 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.

		Prediksi				Total
		Y_{00}	Y_{01}	Y_{10}	Y_{11}	
Aktual	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.

Kesimpulan

Kesimpulan (1)

Berdasarkan hasil pembahasan yang telah didapatkan, kesimpulan yang dapat diambil adalah:

1. Cara untuk mengestimasi parameter dalam model probit bivariat menggunakan *Maximum Likelihood Estimation* (MLE) pertama adalah membuat tabel kontingensi dua arah antar dua variabel respon dan mengetahui persamaan dari probabilitas tiap kategori. Kemudian mendapatkan fungsi *likelihood* dengan distribusi multinomial dan selanjutnya mendapatkan turunan pertama dari fungsi *ln likelihood* terhadap parameter β_1, β_2 dan ρ. Hasil dari turunan pertama yang didapatkan tidak *closed form*, sehingga dilakukan analisis numerik dengan mengunakan iterasi Newton Raphson dengan rumus $\beta^{(m)} = \beta^{(m-1)} - \left[H\left(\beta^{(m-1)}\right)\right]^{-1} g\left(\beta^{(m-1)}\right)$. 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.

Kesimpulan (2)

2. Model regresi probit bivariat untuk faktor-faktor yang mempengaruhi jumlah kasus HIV dan AIDS di Jawa Timur adalah sebagai berikut:

$$\hat{y}_1^* = -10{,}699 + 0{,}612x_1 - 0{,}664x_2 - 136{,}298x_4$$
$$\hat{y}_2^* = 3{,}046 - 0{,}015x_1 - 3{,}656x_2 - 557{,}638x_4$$

Model tersebut merupakan model terbaik dari regresi probit bivariat untuk kasus HIV dan AIDS di Jawa Timur dengan nilai AIC terkecil yaitu sebesar 86,152. Variabel prediktor yang signifikan terhadap HIV (Y_1) adalah persentase kelompok umur 25-49 tahun terhadap jumlah penduduk (x_1). Sedangkan 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).

Diambil interpretasi untuk satu sampel yaitu Kota Surabaya adalah perubahan persentase umur (x_1) sebesar satuan akan meningkatkan 0,0379 terhadap probabilitas HIV berkategori tinggi dan AIDS berkategori tinggi, perubahan persentase jumlah Askeskin atau Jamkesmas (x_2) sebesar satuan akan menurunkan sebesar 1,3958 terhadap probabilitas HIV berkategori tinggi dan AIDS berkategori tinggi, dan perubahan persentase jumlah sarana kesehatan (x_4) sebesar satuan akan menurunkan sebesar 215,3782 terhadap probabilitas HIV berkategori tinggi dan AIDS berkategori tinggi.

Saran

Berdasarkan hasil penelitian yang telah dilakukan peneliti memberikan saran kepada Dinas Kesehatan Provinsi Jawa Timur khususnya untuk lebih memperhatikan aspek kelompok umur 25-49 tahun, jumlah Askeskin atau Jamkesmas serta jumlah sarana kesehatan dengan harapana dapat mengurangi junlah kasus HIV danAIDS di Jawa Timur.

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 Satu. Salemba Empat. Jakarta.

[8] Walpole, R. E. 1982. *"Introduction to Statistics"*. 3th ed. Mc Millan. New York. Terjemahan B. Sumantri. (1995). *"Pengantar Statistika"*. Edisi Ketiga. PT. Gramedia Pustaka Utama. Jakarta.

[9] 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.

[10] Bain, L. J. dan Engelhardt, M. 1992. *"Introduction to Probability and Mathematical Statistics"*. 2nd ed. Duxburry Press. Belmont, California.

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

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

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