

ESTIMATION OF GENERALIZED METHOD OF MOMENT IN LOGISTIC REGRESSION MODEL

Muhammad Taurif R.¹, Bambang W. Otok², I Nyoman Iatra³
^{1,2,3} Jurusan Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
 Institut Teknologi Sepuluh Mopember Surabaya
¹E-mail: muhammadtaurifrasyyidin@gmail.com

Abstrak

Masalah pendugaan parameter dalam analisis regresi sering menjadi topik yang menarik dalam beberapa penelitian, terutama pada penelitian yang bertujuan untuk mengetahui kontribusi relatif dari setiap variabel bebas yang menjelaskan variabel tak bebasnya. Pengambilan sampel dari suatu populasi bertujuan untuk memperoleh informasi mengenai parameter populasi. Untuk mengetahui parameter populasi digunakan metode statistika inferensi, yaitu estimasi sehingga didapatkan suatu nilai dari penaksir parameter tersebut. Salah satu kesulitan yang sering ditemukan dalam analisis regresi adalah pada saat variabel respon (Y) bertipe data kategori, sedangkan variabel bebas atau prediktornya (X) bertipe data kategori maupun kontinu. Tidak terpenuhinya asumsi distribusi multivariat normal dikarenakan variabel bebas merupakan campuran antara variabel kontinyu (*metric*) dan kategorial (*non-metric*) misalnya, probabilitas bahwa orang yang menderita HIV/AIDS dapat diprediksi dari informasi usia, perilaku seks, jenis kelamin, dan lainnya. Estimasi yang baik adalah estimasi yang dapat menggambarkan kuantitas populasi melalui kuantitas sampel. *Generalized Method of Moments* (GMM) merupakan salah satu metode dalam estimasi parameter. Estimasi GMM digunakan untuk mengeksploitasi informasi bentuk kondisi momen populasi. Pada penelitian ini model regresi logistik akan diestimasi dengan metode GMM. Hasil penelitian ini diharapkan dapat memberikan gambaran tentang estimasi GMM pada model regresi logistik untuk kasus penderita HIV/AIDS di Surabaya.

Kata Kunci: *Estimasi, GMM, Kategori, Regresi Logistik.*

Abstract

Parameter estimation problem in the regression analysis is often a topic of interest in several studies, especially in studies that aim to determine the relative contribution of each independent variable that explains the independent variable. Sampling from a population is to obtain information about the population parameters. To determine the population parameters used methods of statistical inference, namely to obtain an estimate of the value of the parameter estimator. The difficulties that are often found in the regression analysis is when the response variable (Y) data type category, while the independent variable or the predictor (X) and continuous data type category. Non-fulfillment of the assumption of multivariate normal distribution because the independent variables are a mixture of continuous variables (metrics) and categorical (non-metric) for example, the probability that people who suffer from HIV/AIDS can be predicted from the information age, sexual behavior, gender, and others. Good estimate is the estimate of the quantity that can describe the population through the sample

quantity. *Generalized Method of Moments (GMM)* is one method in parameter estimation. GMM estimation is used to exploit the information form the population moment conditions. In this study, logistic regression models will be estimated by GMM method. The results of this study are expected to provide an overview of the GMM estimation in logistic regression models for the case of HIV/AIDS in Surabaya.

Keywords: Estimation, GMM, Categorical, Logistic Regression.

1 Pendahuluan

Analisis statistika adalah ilmu yang mempelajari perencanaan, pengumpulan, menganalisis, menginterpretasi, serta mempresentasikan data. Diantara beberapa metode dan analisis statistik yang biasa dipergunakan adalah analisis regresi. Analisis regresi membutuhkan suatu analisa untuk mengkaji estimasi parameter masing-masing fungsi yang terkait sesuai dengan syarat-syarat statistik tersebut dengan cara mengambil sampel dari suatu populasi.

Pengambilan sampel dari suatu populasi bertujuan untuk memperoleh informasi mengenai parameter populasi. Masalah pendugaan parameter dalam analisis regresi sering menjadi topik yang menarik, terutama pada penelitian yang bertujuan untuk mengetahui kontribusi relatif dari setiap variabel bebas yang menjelaskan variabel tak bebasnya. Salah satu kesulitan yang sering ditemukan dalam analisis regresi adalah pada saat variabel respon bertipe data kategori, sedangkan variabel bebas atau prediktornya bertipe data kategori maupun kontinu. Regresi logistik merupakan alternatif uji jika asumsi distribusi multivariat normal pada variabel bebasnya tidak bisa terpenuhi seperti dituliskan oleh Agresti [1].

Dalam analisis regresi, estimasi parameter merupakan tujuan penting untuk mendapatkan estimator. Estimasi yang baik adalah estimasi yang dapat menggambarkan kuantitas populasi melalui kuantitas sampel. *Generalized Method of Moments (GMM)* merupakan salah satu metode dalam estimasi parameter. Metode GMM diperkenalkan oleh Hansen sebagai estimasi parameter yang meminimalkan bentuk kuadrat dari kondisi momen sampel yang terboboti.

Keunggulan dari estimasi GMM ialah memungkinkan untuk melakukan estimasi secara lebih detail pada data penelitian yang memiliki masalah ketidakpastian parameter seperti saat variabel dependen memiliki parameter yang tidak diketahui dan harus diestimasi seperti yang dipaparkan oleh Bontemps, dkk [2].

Dengan pesatnya perkembangan penelitian, diperlukan penyesuaian berdasarkan kepentingan terhadap masing-masing bidang penelitian dengan tujuan untuk meningkatkan pengetahuan atau pengembangan ilmu mengenai kesehatan maupun bidang penelitian lainnya. Tujuan dari penelitian ini adalah menerapkan dan mendapatkan bentuk estimasi GMM pada model regresi logistik.

2 Tinjauan Pustaka

2.1 Analisis Regresi

Analisis regresi dalam statistika merupakan salah satu metode yang bertujuan menentukan hubungan sebab-akibat antara satu variabel dengan variabel-variabel

lainnya. Penerapan analisis regresi dapat dijumpai secara luas pada berbagai bidang seperti kesehatan, ekonomi, teknik, manajemen, ilmu pertanian, sosial, dan sebagainya. Analisis regresi berguna untuk mempresentasikan hubungan dari dua variabel atau lebih, menelusuri pola hubungan pada model yang belum diketahui secara sempurna sebagai alat statistik secara eksploratif.

Bentuk umum model regresi untuk j variabel bebas dengan notasi sebagai berikut:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_j x_j + \varepsilon \quad (1)$$

Dimana ε adalah *error*, $\beta_0, \beta_1, \beta_2, \dots, \beta_j$ adalah parameter, Y adalah variabel respon dan x_1, x_2, \dots, x_j adalah variabel bebas yang berbeda. Dalam matrik ditulis sebagai berikut:

$$y = X\beta + \varepsilon, \text{ dimana } \varepsilon \sim N(0, \sigma^2 I) \quad (2)$$

2.2 Analisis Regresi Logistik

Regresi logistik merupakan suatu metode analisis data yang digunakan untuk mencari hubungan antara variabel respon yang bersifat biner atau dikotomous dengan variabel prediktor yang bersifat polikotomous. *Outcome* dari variabel respon (y) terdiri dari 2 kategori yaitu “sukses” dan “gagal” yang dinotasikan $y = 1$ (sukses) dan $y = 0$ (gagal) mengacu pada Hosmer dan Lemeshow [3] dengan fungsi probabilitas untuk setiap observasi dinotasikan:

$$f(y) = \pi^y (1 - \pi)^{1-y}; y = 0, 1 \quad (3)$$

dimana jika $y = 0$ maka $f(y) = 1 - \pi$ dan jika $y = 1$ maka $f(y) = \pi$. Fungsi regresi logistiknya dapat ditulis:

$$f(z) = \frac{e^z}{1 + e^z} \quad (4)$$

dengan $z = \beta_0 + \beta_1 x_1 + \dots + \beta_j x_j$

Nilai z antara $-\infty$ dan $+\infty$ sehingga nilai $f(z)$ terletak antara 0 dan 1 untuk setiap nilai z yang diberikan. Hal tersebut menunjukkan bahwa model regresi logistik sebenarnya menggambarkan probabilitas atau resiko dari suatu obyek. Model regresi logistik sebagai berikut:

$$\pi(x) = \frac{e^{(\beta_0 + \beta_1 x_1 + \dots + \beta_j x_j)}}{1 + e^{(\beta_0 + \beta_1 x_1 + \dots + \beta_j x_j)}} \quad (5)$$

2.3 Estimasi Parameter

Sampel digunakan untuk mendapatkan informasi mengenai parameter populasi. Statistik inferensia merupakan alat penduga parameter populasi. Terdapat dua cara dalam menduga parameter populasi yaitu pengestimasi dan pengujian hipotesis. Estimasi dapat diartikan sebagai penentuan nilai-nilai yang diperoleh dari data sampel dan dapat digunakan sebagai pengganti nilai parameter yang tidak diketahui. Estimasi parameter terdapat dua macam yakni, estimasi titik dan estimasi interval. Estimasi titik merupakan penentuan sebuah nilai yang diperoleh dari data sampel dan dapat digunakan sebagai pengganti nilai dari parameter yang tidak diketahui. Metode momen, MLE, dan

metode kuadrat terkecil merupakan metode estimasi titik. Estimasi interval merupakan penentuan nilai-nilai yang berbentuk interval yang diperoleh dari data sampel dan dapat digunakan sebagai pengganti nilai dari parameter populasi yang tidak diketahui.

2.4 Method of Moment (MM)

Method of moment (MM) berasal dari estimasi momen pertama distribusi dengan mengasumsikan distribusi suatu populasi memiliki nilai rata-rata dan variansi sama dengan 1. *Method of moment* merupakan salah satu metode yang digunakan untuk memperoleh suatu estimator untuk parameter θ dengan ide dasar berupa menyamakan antara momen-momen populasi dengan momen-momen sampel.

Salah satu contoh penerapan *Method of moment* pada populasi berdistribusi $N(\mu, \sigma^2)$ dengan demikian fungsi probabilitas dari X diberikan oleh :

$$f(x, \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}, -\infty < x < \infty \quad (6)$$

berdasarkan fungsi probabilitas tersebut dapat dihitung momen-momen teoritis yakni:

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} xf(x; \mu, \sigma^2) dx = \mu \\ E(X^2) &= \int_{-\infty}^{\infty} x^2 f(x; \mu, \sigma^2) dx \\ &= \text{Var}(x) + [E(x)]^2 \\ &= \sigma^2 + \mu^2 \end{aligned} \quad (7)$$

dari hasil diatas dibentuk persamaan-persamaan:

$$(1). \mu_1(\hat{\mu}, \hat{\sigma}^2) = M_1 \Rightarrow \hat{\mu} = \frac{1}{n} \sum_{i=1}^n X_i \quad (8)$$

$$(2). \mu_2(\hat{\mu}, \hat{\sigma}^2) = M_2 \Rightarrow \hat{\sigma}^2 + \hat{\mu}^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 \quad (9)$$

dengan menyelesaikan kedua persamaan secara simultan didapatkan:

$$\begin{aligned} \hat{\mu}_{MM} &= \frac{1}{n} \sum_{i=1}^n X_i = \bar{X} \\ \hat{\sigma}^2_{MM} &= \frac{1}{n} \sum_{i=1}^n X_i^2 - \hat{\mu}_{MM}^2 \\ &= \frac{1}{n} \sum_{i=1}^n X_i^2 - \left(\frac{1}{n} \sum_{i=1}^n X_i \right)^2 \\ &= \frac{1}{n} \left(\sum_{i=1}^n X_i^2 - \frac{\left(\sum_{i=1}^n X_i \right)^2}{n} \right) \\ &= \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \end{aligned} \quad (10)$$

2.5 Generalized Method of Moment (GMM)

Generalized Method of Moment (GMM) merupakan metode penaksiran parameter perluasan dari metode momen. Metode momen tidak dapat digunakan apabila banyaknya variabel instrumen lebih besar dibandingkan dengan jumlah parameter yang akan ditaksir. GMM menyamakan momen kondisi dari populasi dengan momen kondisi dari sampel. Metode GMM merupakan salah satu metode yang dapat mengatasi kondisi data dengan pelanggaran asumsi-asumsi pada analisis regresi. GMM didapat dengan cara meminimumkan jumlah kuadrat terboboti dari momen kondisi sampel. Matrik

bobot \hat{W} merupakan suatu matrik simetris definit positif berukuran $L \times L$ yang bukan fungsi dari β dengan notasi sebagai berikut:

$$J(\hat{\beta}) = \left\| \bar{g}(\hat{\beta}) \right\|_{\hat{W}}^2 = \bar{g}(\hat{\beta})' \hat{W} \bar{g}(\hat{\beta}) \tag{11}$$

maka,

$$J(\hat{\beta}) = \bar{g}(\hat{\beta})' \hat{W} \bar{g}(\hat{\beta}) \tag{12}$$

dimana mengacu pada model regresi

$$\mathbf{y} = \mathbf{X}_i \beta + \varepsilon_i \text{ dengan } i = 1, 2, 3, \dots, N \tag{13}$$

dan momen kondisi dari sampel adalah sebagai berikut:

$$\bar{g}(\hat{\beta}) = N^{-1} \sum_{i=1}^N \mathbf{Z}_i' (y_i - \mathbf{X}_i \hat{\beta}) \tag{14}$$

sehingga,

$$\begin{aligned} J(\hat{\beta}) &= \left[N^{-1} \sum_{i=1}^N \mathbf{Z}_i' (y_i - \mathbf{X}_i \hat{\beta}) \right]' \hat{W} \left[N^{-1} \sum_{i=1}^N \mathbf{Z}_i' (y_i - \mathbf{X}_i \hat{\beta}) \right] \\ &= \left(N^{-1} \sum_{i=1}^N y_i \mathbf{Z}_i' \right)' \hat{W} \left(N^{-1} \sum_{i=1}^N \mathbf{Z}_i' y_i \right) - \left(N^{-1} \sum_{i=1}^N y_i \mathbf{Z}_i' \right)' \hat{W} \left(N^{-1} \sum_{i=1}^N \mathbf{Z}_i' \mathbf{X}_i \hat{\beta} \right) \\ &\quad - \left(N^{-1} \sum_{i=1}^N \hat{\beta}' \mathbf{X}_i \mathbf{Z}_i' \right)' \hat{W} \left(N^{-1} \sum_{i=1}^N \mathbf{Z}_i' y_i \right) + \left(N^{-1} \sum_{i=1}^N \hat{\beta}' \mathbf{X}_i \mathbf{Z}_i' \right)' \hat{W} \left(N^{-1} \sum_{i=1}^N \mathbf{Z}_i' \mathbf{X}_i \hat{\beta} \right) \\ &\tag{15} \\ &= \left(N^{-1} \sum_{i=1}^N y_i \mathbf{Z}_i' \right)' \hat{W} \left(N^{-1} \sum_{i=1}^N \mathbf{Z}_i' y_i \right) - 2 \left(N^{-1} \sum_{i=1}^N \hat{\beta}' \mathbf{X}_i \mathbf{Z}_i' \right)' \hat{W} \left(N^{-1} \sum_{i=1}^N \mathbf{Z}_i' y_i \right) \\ &\quad + \left(N^{-1} \sum_{i=1}^N \hat{\beta}' \mathbf{X}_i \mathbf{Z}_i' \right)' \hat{W} \left(N^{-1} \sum_{i=1}^N \mathbf{Z}_i' \mathbf{X}_i \hat{\beta} \right) \end{aligned} \tag{16}$$

dimana,

$$\begin{aligned} \frac{\partial J(\hat{\beta})}{\partial \hat{\beta}} &= -2 \left(N^{-1} \sum_{i=1}^N \hat{\beta}' \mathbf{X}_i \mathbf{Z}_i' \right)' \hat{W} \left(N^{-1} \sum_{i=1}^N \mathbf{Z}_i' y_i \right) \\ &\quad + 2 \left(N^{-1} \sum_{i=1}^N \hat{\beta}' \mathbf{X}_i \mathbf{Z}_i' \right)' \hat{W} \left(N^{-1} \sum_{i=1}^N \mathbf{Z}_i' \mathbf{X}_i \hat{\beta} \right) \end{aligned} \tag{17}$$

2.6 Interpretasi Koefisien Parameter

Interpretasi terhadap koefisien parameter ini dilakukan untuk menentukan kecenderungan atau hubungan fungsional antar variabel prediktor dengan variabel respon serta menunjukkan pengaruh perubahan nilai pada variabel yang bersangkutan. Dalam hal ini digunakan besaran *Odds ratio* atau e^β dan dinyatakan dengan ψ . Untuk regresi logistik dimana variabel prediktor bersifat dikotomis, nilai x dikategorikan 0 atau 1. Pada model ini, ada dua nilai $\pi(x)$ dan dua nilai $1 - \pi(x)$.

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

$$g(1) = \ln \left[\frac{\pi(1)}{1 - \pi(1)} \right] \text{ dan } g(0) = \ln \left[\frac{\pi(0)}{1 - \pi(0)} \right] \quad (18)$$

Odds ratio diartikan sebagai kecenderungan variabel respon memiliki suatu nilai tertentu jika diberikan $x = 1$ dan dibandingkan pada $x = 0$, yang dapat dituliskan dalam persamaan berikut seperti dituliskan oleh Hosmer dan Lemeshow [3].

$$\psi = \frac{\pi(1) / [1 - \pi(1)]}{\pi(0) / [1 - \pi(0)]} \quad (19)$$

nilai *odds ratio* adalah sebagai berikut

$$\begin{aligned} \psi &= \frac{\left(\frac{\exp(\beta_0 + \beta_1)}{1 + \exp(\beta_0 + \beta_1)} \right) \left(\frac{1}{1 + \exp(\beta_0)} \right)}{\left(\frac{\exp(\beta_0)}{1 + \exp(\beta_0)} \right) \left(\frac{1}{1 + \exp(\beta_0 + \beta_1)} \right)} \\ &= \frac{\exp(\beta_0 + \beta_1)}{\exp(\beta_0)} \\ &= \exp(\beta_1) \end{aligned} \quad (20)$$

dengan kriteria, jika nilai $(\psi) < 1$, maka antara variabel prediktor dan variabel respon terdapat hubungan negatif setiap kali perubahan nilai variabel bebas (x) dan jika *Odds ratio* $(\psi) > 1$ maka antara variabel prediktor dengan variabel respon terdapat hubungan positif setiap kali perubahan nilai variabel bebas (x). Keputusan tidak terdapat hubungan antara variabel prediktor dengan variabel respon diambil jika nilai *Odds ratio* $(\psi) = 1$.

3 Metode Penelitian

3.1 Data dan Variabel Penelitian

Data yang akan digunakan dalam penelitian ini adalah data sekunder berupa data kasus penderita HIV/AIDS di Surabaya pada tahun 2013 berdasarkan kuesioner LSM ORBIT dengan variabel penelitian yang digunakan adalah variabel klien, pola perilaku, dan riwayat penggunaan jarum suntik.

3.2 Langkah-langkah Penelitian

Secara lebih rinci langkah-langkah dalam penelitian ini adalah sebagai berikut:

1. Pengumpulan studi literatur
2. Perumusan masalah penelitian
3. Pengidentifikasian data penelitian
4. Melakukan pendeskripsian data
5. Memodelkan regresi logistik biner
6. Mengestimasi parameter β dengan menggunakan metode GMM
7. Melakukan pengujian kesesuaian model dengan kriteria uji \hat{C} mengikuti distribusi *chi-square* dengan derajat bebas $g-2$ mengacu pada paparan Hosmer and Lemeshow [3].
8. Melakukan pengujian estimasi parameter secara parsial dengan kriteria uji *Wald* yang mengikuti distribusi normal sehingga H_0 ditolak jika $|W| > Z_{\alpha/2}$ dan secara serentak dengan kriteria uji *G* dimana nilai *G* mengikuti distribusi *Chi Square*
9. Interpretasi koefisien parameter melalui nilai *odds ratio*.
10. Kesimpulan.

4 Hasil Penelitian

Estimasi GMM pada regresi logistik biner didapatkan bentuk model sebagai berikut :

$$\hat{\beta} = \left[\left(\sum_{i=1}^N \mathbf{x}_i' z_i \right) \hat{W} \left(\sum_{i=1}^N z_i \mathbf{x}_i \right) \right]^{-1} \left[\left(\sum_{i=1}^N \mathbf{x}_i' z_i \right) \hat{W} \left(\sum_{i=1}^N z_i y_i \right) \right] \quad (21)$$

5 Kesimpulan dan Saran

5.1 Kesimpulan

Adapun kesimpulan yang didapat pada penelitian ialah sebagai berikut:

1. Hasil estimasi GMM untuk regresi logistik biner ditentukan pada kriteria nilai *Odds ratio*.
2. Penerapan estimasi GMM lebih efisien untuk mengatasi masalah heteroskedastisitas.

5.2 Saran

Pada penerapan estimasi GMM selanjutnya diharapkan lebih mengeksplorasi bentuk regresi logistik multinomial, ordinal dan bentuk lain.

Ucapan Terima Kasih

Penulis mengucapkan terima kasih kepada Universitas Jember sebagai penyelenggara Seminar Nasional Matematika yang telah memberikan izin terhadap peneliti sebagai pemakalah beserta Direktorat Pendidikan Tinggi, yang telah memberikan dukungan

finansial melalui Beasiswa Pascasarjana Dalam Negeri tahun 2013-2015. Penulis juga menyampaikan ucapan terima kasih kepada pihak Pascasarjana Institut Teknologi Sepuluh Nopember khususnya FMIPA Statistika serta pembimbing tesis Dr. Bambang Widjanarko Otok, M.Si. dan Dr. I Nyoman Latra, M.Si. yang telah memberikan bimbingan guna kelancaran dalam menyusun proposal penelitian ini.

Daftar Pustaka

- [1] Agresi, Alan. *Categorical Data Analysis*. John Wiley and Sons, Inc: New York, 1990.
- [2] Bontems, Christian, Meddahi, Nour. *Testing Normality: A GMM Approach. Scientifics Series*. Montreal. Cirano, 2002.
- [3] Hosmer, D. W., and Stanley Lemeshow. *Applied Logistic Regression*. Edition 2. John Wiley and Sons, Inc: New York, Book, 2000.