

ANALISIS REGRESI LINEAR BERGANDA DENGAN SATU VARIABEL BONEKA (*DUMMY VARIABLE*)

Tanti Krisnawardhani, Nur Salam, dan Dewi Anggraini

Program Studi Matematika

Universitas Lambung Mangkurat

Jl. Jend. A. Yani km. 36 Kampus Unlam Banjarbaru

ABSTRAK

Untuk mengakomodasi adanya variabel bebas kualitatif ke dalam model regresi maka digunakan variabel boneka (*dummy variable*) dalam persamaan regresi yang dapat dituliskan sebagai berikut: $Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_i + \varepsilon_i$. Penelitian ini bertujuan untuk menentukan estimasi parameter dari model regresi linier berganda dengan satu variabel boneka menggunakan metode kuadrat terkecil, menentukan statistik uji kecocokan model dengan menggunakan metode *likelihood ratio* dan menerapkan model pada contoh kasus. Hasil penelitian menunjukkan bahwa estimasi parameter model regresinya adalah: $\hat{\beta} = (\mathbf{X}' \mathbf{X})^{-1} (\mathbf{X}' \mathbf{Y})$. Aplikasi dari penggunaan model regresi ini adalah pada contoh soal tentang perhitungan mengenai biaya kuliah (Y) berdasarkan nilai rata-rata tes bakat skolastik (*Scholastic Aptitude Test*) (X) dan tipe universitas (D) sehingga diperoleh model sebagai berikut: $\hat{Y} = -7263,56 + 19,52X + 8732,418D$.

Kata Kunci: *Regresi Linier Berganda, Variabel Boneka, Metode Kuadrat Terkecil dan Likelihood Ratio.*

1. PENDAHULUAN

Analisis regresi merupakan analisis yang digunakan untuk menjelaskan keterkaitan hubungan antara suatu variabel bebas (*independent*) terhadap variabel tak bebas (*dependent*) yang dapat dinyatakan sebagai bentuk model matematis. Dalam analisis regresi bukan hanya variabel-variabel bebas (X_i) kuantitatif yang mempengaruhi variabel tak bebas (Y_i), tetapi ada juga variabel-variabel bebas kualitatif yang juga ikut mempengaruhi, seperti jenis kelamin, musim, warna, pendidikan, dan lain sebagainya [1]. Variabel boneka ini mengambil bilangan biner seperti 1 atau 0. Kedua nilai yang diberikan tidak menunjukkan bilangan, tetapi hanya sebagai identifikasi kelas atau kategorinya. Sebagai contoh, 1 mungkin menunjukkan bahwa seseorang adalah perempuan dan 0 menunjukkan laki-laki atau 1 untuk menunjukkan bahwa seseorang merupakan lulusan perguruan tinggi dan 0 bukan lulusan perguruan tinggi. Model umum regresi linier berganda dengan variabel boneka adalah:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik} + \beta_{k+1} D_{i1} + \beta_{k+2} D_{i2} + \dots + \beta_{k+c} D_{ic} + \varepsilon_i$$

Penelitian ini akan dilakukan pengkajian mengenai bagaimana menentukan estimasi parameter pada model regresi linier berganda dengan satu variabel boneka menggunakan metode kuadrat terkecil, bagaimana menentukan statistik uji kecocokan model dalam regresi linier berganda dengan satu variabel boneka menggunakan metode *likelihood ratio* dan bagaimana menerapkan model pada contoh kasus dimana datanya berupa data kuantitatif dan kualitatif. Diharapkan

melalui penelitian ini dapat memberikan pemahaman yang mendalam mengenai analisis regresi linier berganda dengan satu variabel boneka.

2. METODOLOGI

Penelitian ini dilakukan dengan cara studi literatur dari berbagai sumber yang menunjang dan relevan dengan tinjauan yang dilakukan. Buku atau materi yang digunakan dalam penelitian ini adalah buku dan jurnal-jurnal yang terkait dengan materi tentang analisis regresi linier berganda dengan satu variabel boneka.

Langkah-langkah yang dilakukan dalam penelitian ini adalah:

1. Mempelajari analisis regresi linier berganda dengan satu variabel boneka.
2. Mengestimasi parameter model regresi linier berganda dengan satu variabel boneka menggunakan metode kuadrat terkecil.
 - a. Mengubah model linier menjadi eksplisit terhadap galat.
 - b. Mengkuadratkan galat yang diperoleh serta menjumlahkannya untuk seluruh pasangan data.
 - c. Mencari turunan parsialnya terhadap $\beta_0, \beta_1, \beta_2$ dan menyelesaikan turunan parsialnya.
3. Menentukan statistik uji kecocokan model untuk model regresi linier berganda dengan satu variabel boneka menggunakan metode *likelihood ratio*.
 - a. Menentukan hipotesis.
 - b. Membuat fungsi *likelihood* dan log *likelihood* dengan fungsi probabilitas dari distribusi normal.
 - c. Mencari turunan parsial terhadap fungsi log *likelihood* β_0, σ^2 sehingga didapatkan *likelihood ratio* dan GLR.
4. Menerapkan model regresi linier berganda dengan satu variabel boneka pada contoh soal.

Langkah – langkah pengolahan data yang dilakukan adalah:

- a. Menentukan X (variabel bebas) yaitu rata-rata nilai tes bakat skolastik, D (variabel boneka) yaitu tipe universitas, serta Y (variabel tak bebas) yaitu biaya kuliah.
- b. Menganalisis data yaitu melakukan estimasi terhadap parameter model dengan menggunakan metode kuadrat terkecil.
- c. Melakukan statistik uji dimana tolak H_0 jika $F_{hit} > F_{\alpha, 2, n-3}$. Bila H_0 ditolak, ini berarti variabel bebas secara keseluruhan mempengaruhi variabel tak bebas atau dapat dikatakan bahwa paling tidak ada satu variabel bebas yang mempengaruhi variabel tak bebas sehingga model tersebut signifikan.
- d. Interpretasi parameter / model.

3. HASIL DAN PEMBAHASAN

3.1 Estimasi Parameter dengan Metode Kuadrat Terkecil

Metode kuadrat terkecil menggunakan pendekatan geometris, dimana secara geometris garis yang paling mewakili sebaran sampel adalah garis yang mempunyai simpangan minimum atau galat terkecil dengan pencaran data. Diketahui model untuk regresi berganda dengan satu variabel bebas (X) dan satu variabel boneka (D) adalah:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_i + \varepsilon_i \quad (1)$$

Langkah pertama yang dilakukan adalah mengubah model linier (1) menjadi eksplisit terhadap galat kemudian mengkuadratkan galat yang diperoleh serta menjumlahkannya untuk seluruh pasangan data sebagai berikut:

$$\sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_i - \beta_2 D_i)^2 \quad (2)$$

Dengan memisalkan $s = \sum_{i=1}^n \varepsilon_i^2$, kemudian melakukan penurunan parsial s terhadap $\beta_0, \beta_1, \beta_2$ kemudian meminimumkannya dan menyederhanakannya, sehingga diperoleh estimator β_0, β_1 dan β_2 adalah:

$$\hat{\beta}_0 = \frac{\sum_{i=1}^n Y_i - \hat{\beta}_1 \sum_{i=1}^n X_i - \hat{\beta}_2 \sum_{i=1}^n D_i}{n} \quad (3)$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n Y_i X_i - \hat{\beta}_0 \sum_{i=1}^n X_i - \hat{\beta}_2 \sum_{i=1}^n D_i X_i}{\sum_{i=1}^n X_i^2} \quad (4)$$

$$\hat{\beta}_2 = \frac{\sum_{i=1}^n Y_i D_i - \hat{\beta}_0 \sum_{i=1}^n D_i - \hat{\beta}_1 \sum_{i=1}^n X_i D_i}{\sum_{i=1}^n D_i^2} \quad (5)$$

3.2 Uji Kecocokan Model

Uji *Likelihood Ratio* atau pengujian secara simultan akan digunakan untuk menguji kecocokan model secara simultan. Statistik uji yang akan dicari adalah statistik uji untuk hipotesis berikut:

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

$$H_1 : \text{minimal ada satu } \beta_j \neq 0, j = 0, 1, 2.$$

Dengan $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{\beta}_2 D_i$ sedangkan parameter $\hat{\beta}_0$ dan $\hat{\sigma}^2$ tidak diketahui, sehingga fungsi *likelihood* dari Ω dan ω didefinisikan sebagai berikut:

$\ell(\Omega) = \prod_{i=1}^n f\left(y_i; \hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\sigma}^2\right)$ dimana $-\infty < \beta_j < \infty$, $\hat{\sigma}^2 = 0$ dan sedangkan

$$\ell(\omega) = \prod_{i=1}^n f\left(y_i; \hat{\beta}_{00}, \hat{\beta}_1, \hat{\beta}_2, \hat{\sigma}_0^2\right)$$

sehingga didapatkan fungsi *likelihoodnya* dengan fungsi probabilitas dari distribusi normal yaitu:

$$L = \prod_i \frac{1}{\sqrt{2\pi\sigma}} \exp\left[-\frac{1}{2\sigma^2}(y_i - \hat{y}_i)^2\right]$$

$$= (2\pi\sigma^2)^{-n/2} \exp\left[-\frac{1}{2\sigma^2} \sum_i \left(y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{\beta}_2 D_i)\right)^2\right]$$

Dengan melakukan penurunan parsial L terhadap $\hat{\beta}_0$ dan σ^2 kemudian meminimumkan dan menyederhanakannya, sehingga diperoleh:

$$\lambda(y) = \frac{\ell(\hat{\omega})}{\ell(\hat{\Omega})} = \left[\frac{\sum (\hat{y}_i - \bar{y})^2 + \sum (y_i - \hat{y}_i)^2}{\sum (y_i - \hat{y}_i)^2} \right]^{-n/2} = \left[1 + \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \hat{y}_i)^2} \right]^{-n/2}$$

$$\frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \hat{y}_i)^2}$$

Sehingga statistik ujiinya $F_{hit} = \frac{2}{\frac{\sum (y_i - \hat{y}_i)^2}{n-3}} \sim F_{(\alpha, 2, n-3)}$

Tolak H_0 jika $F_{hit} > F_{\alpha, 2, n-3}$.

Bila H_0 ditolak, artinya variabel bebas secara keseluruhan mempengaruhi variabel respon atau dapat dikatakan bahwa paling tidak ada satu variabel bebas yang mempengaruhi variabel respon sehingga model tersebut signifikan pada tingkat signifikansi α .

3.3 Aplikasi

Sumber data yang digunakan pada aplikasi adalah contoh data dari sejumlah universitas berbeda dengan pertimbangan bahwa rata-rata nilai tes bakat skolastik pada suatu universitas dan tipe universitas mempengaruhi biaya kuliah per tahun. Penelitian dilakukan dengan mengambil sebuah sampel acak sederhana dari 80 universitas berbeda dengan tipenya yang digolongkan menjadi universitas negeri dan swasta serta rata-rata nilai tes bakat skolastik yang berbeda dari tiap universitas tersebut. Dimana biaya kuliah dijadikan sebagai variabel tak bebas (Y_i), rata-rata nilai tes bakat skolastik dijadikan variabel bebas (X_i) dan tipe

universitas dijadikan variabel boneka (D_i). menggunakan estimasi kuadrat terkecil yang diselesaikan dengan eliminasi biasa didapatkan solusi bagi $\hat{\beta}_0$, $\hat{\beta}_1$ dan $\hat{\beta}_2$ sebagai berikut: $\hat{\beta}_0 = -7263,56$; $\hat{\beta}_1 = 19,52$; $\hat{\beta}_2 = 8732,418$. Sehingga model regresi linier berganda dengan satu variabel boneka adalah sebagai berikut:

$$\hat{Y} = -7263,56 + 19,52X + 8732,418D \quad (6)$$

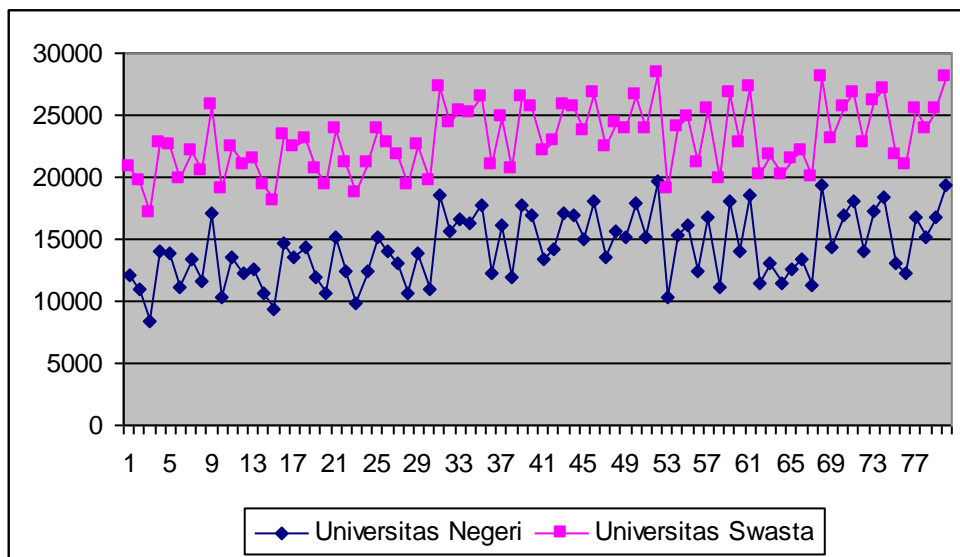
Dengan $\hat{\beta}_2 = 8732,418$ untuk $D = 0$ sebagai variabel boneka untuk universitas negeri modelnya adalah:

$$\hat{Y} = -7263,56 + 19,52X \quad (7)$$

Dimana setiap kenaikan nilai rata-rata tes bakat skolastik sebesar 1 satuan maka biaya kuliahnya menjadi \$-7244,04. Dan untuk $D = 1$ sebagai variabel boneka untuk universitas swasta, modelnya menjadi:

$$\hat{Y} = 1468,858 + 19,52X \quad (8)$$

Dimana setiap kenaikan nilai rata-rata tes bakat skolastik sebesar 1 satuan maka biaya kuliahnya menjadi \$1488,378.



Gambar 1. Perbedaan titik-titik regresi biaya kuliah antara universitas negeri dan swasta

4. KESIMPULAN

Kesimpulan dari penelitian yaitu:

1. Estimasi parameter terhadap β dengan metode kuadrat terkecil diperoleh tiga persamaan normal sebagai berikut:

$$\hat{\beta}_0 n + \hat{\beta}_1 \sum_{i=1}^n X_i + \hat{\beta}_2 \sum_{i=1}^n D_i = \sum_{i=1}^n Y_i$$

$$\hat{\beta}_0 \sum_{i=1}^n X_i + \hat{\beta}_1 \sum_{i=1}^n X_i^2 + \hat{\beta}_2 \sum_{i=1}^n D_i X_i = \sum_{i=1}^n Y_i X_i$$

$$\hat{\beta}_0 \sum_{i=1}^n D_i + \hat{\beta}_1 \sum_{i=1}^n X_i D_i + \hat{\beta}_2 \sum_{i=1}^n D_i^2 = \sum_{i=1}^n Y_i D_i$$

sehingga jika diselesaikan akan menghasilkan koefisien $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$ dan

secara matrik dapat dituliskan $\hat{\beta} = (\mathbf{X}' \mathbf{X})^{-1} (\mathbf{X}' \mathbf{Y})$.

2. Statistik uji dari regresi linier berganda dengan satu variabel boneka untuk uji kecocokan model dengan metode *likelihood ratio* untuk menguji hipotesis $H_0 : \beta_1 = \beta_2 = 0$ melawan hipotesis H_1 : minimal ada satu $\beta_j \neq 0$,

$$F_{hit} = \frac{\sum \left(\hat{y}_i - \bar{y} \right)^2}{\frac{\sum \left(y_i - \hat{y}_i \right)^2}{n-3}} \sim F_{(\alpha, 2, n-3)}$$

Tolak H_0 jika $F_{hit} > F_{\alpha, 2, n-3}$.

Bila H_0 ditolak, berarti variabel bebas secara keseluruhan mempengaruhi variabel tak bebas atau dapat dikatakan bahwa paling tidak ada satu variabel bebas yang mempengaruhi variabel tak bebas sehingga model tersebut signifikan pada tingkat signifikansi α .

3. Hasil aplikasi yang berkaitan dengan regresi linier berganda dengan satu variabel boneka adalah sebagai berikut:

$$\hat{Y} = -7263.56 + 19.52X + 8732.418D.$$

Jika universitas tersebut merupakan universitas negeri maka biaya kuliah menjadi \$-7244,04 setiap kenaikan nilai rata-rata tes bakat skolastik sebesar 1 satuan sedangkan pada universitas swasta, setiap kenaikan nilai rata-rata tes bakat skolastik sebesar 1 satuan, biaya kuliah menjadi \$1488,378. Sehingga dapat disimpulkan bahwa biaya kuliah di universitas swasta cenderung lebih mahal dibandingkan biaya kuliah di universitas negeri.

5. DAFTAR PUSTAKA

- [1]. Gasperz, V. 1991. *Ekonometrika Terapan I*. Penerbit Tarsito. Bandung.
- [2]. Bain, L. J., & Engelhardt, M. 1992. *Introduction to Probability and Mathematical Statistics, Edisi pertama*. PT Belmont company, California.
- [3]. Gujarati, D. 1995. *Ekonometrika Dasar*. Terjemahan Sumarmo Zain. Erlangga, Jakarta.
- [4]. Indah, M., & Tirta, I. M 2004. *Peubah Boneka (Dummy variable) Dalam Analisis Regresi Linier*. Fakultas MIPA Universitas Jember.
- [5]. Levine, D. M. 2005. *Statistics For Managers*. Prentice Hall Inc, a division of pearson Education.
- [6]. Montgomery, D. C., & Peck E. A. 1991. *Introduction to Linear Regression Analysis*. 2nd edition. New York. John Wiley and Sons.

- [7]. Nachrowi, D. N., & Usman, H. 2002. *Penggunaan Teknik Ekonometri*. Raja Grafindo Perkasa, Jakarta.
- [8]. Supranto, J. 1989. *Statistik Teori dan Aplikasinya, edisi keenam*. Erlangga.
- [9]. Tirta, M. I. 2000. *Pemodelan Matematika*. Diktat Kuliah Jurusan Matematika Fakultas MIPA Universitas Jember.
- [10]. Walpole, E. R., & Myers, R. H. 2002. *Probability and Statistics For Engineers and Scientist*. Prentice Hall. New Jersey.