

**STUDI SIMULASI EVALUASI KETEPATAN KLASIFIKASI
INDEKS MASSA TUBUH BERDASARKAN TABEL KLASIFIKASI
INDEKS MASSA TUBUH WHO DENGAN PENDEKATAN
REGRESI LOGISTIK ORDINAL DAN ANALISIS DISKRIMINAN
(Studi Kasus Klasifikasi Status Gizi Dewasa)**

Dewinta Dwinanda Puspitasari, S.Si

Jurusan Statistika Terapan Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Padjajaran
dewintadwinanda@yahoo.com

Septiadi Padmadisastra, Ph.D

Jurusan Statistika Terapan Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Padjajaran
s_padmadisastra@yahoo.com

Dr. Nussar Hajarisma, M.Si

Jurusan Statistika Terapan Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Padjajaran

Abstrak

Status gizi dewasa diklasifikasikan menjadi empat kelompok, yaitu *underweight*, normal, *overweight*, dan obesitas. Tujuan utama dari penelitian ini adalah evaluasi ketepatan klasifikasi indeks massa tubuh berdasarkan tabel klasifikasi indeks massa tubuh WHO dengan pendekatan regresi logistik ordinal dan analisis diskriminan. Selanjutnya, membandingkan ketepatan klasifikasi antara regresi logistik ordinal dan analisis diskriminan dalam pengklasifikasi status gizi dewasa. Studi kasus pengelompokkan status gizi dewasa berdasarkan variabel bebas yang terlibat yaitu, berat badan (kg) dan tinggi badan (cm) dengan studi simulasi. Berat badan dan tinggi badan bersifat kontinu dan berdistribusi normal multivariat.

Data simulasi menggunakan lima tingkatan ukuran sampel, yaitu $n_1=200$, $n_2=400$, $n_3=600$, $n_4=800$, dan $n_5=1000$ dengan pengulangan sebanyak 100 kali. Ukuran ketepatan klasifikasi yang digunakan adalah *Apparent Error Rate* (APER). Hasilnya adalah rata-rata kesalahan klasifikasi (APER) yang dihasilkan regresi logistik ordinal sebesar 28,56%, analisis diskriminan 28,49%, dan berdasarkan tabel klasifikasi indeks massa tubuh WHO sebesar 31,08%. Analisis Diskriminan menghasilkan nilai APER lebih kecil dibandingkan dengan Regresi

Logistik Ordinal. Klasifikasi indeks massa tubuh (IMT) dengan tabel klasifikasi IMT WHO menghasilkan APER paling besar dibandingkan dengan analisis diskriminan dan regresi logistik ordinal. Analisis diskriminan lebih baik diterapkan dalam pengklasifikasian status gizi dewasa dibandingkan dengan regresi logistik ordinal.

Kata Kunci: Tabel Klasifikasi Indeks Massa Tubuh, Regresi Logistik Ordinal, Analisis Diskriminan, Klasifikasi.

Abstract

Adult nutrient status was classified into four groups, underweight, normal, overweight, and obesity. The main purpose of this study was to evaluate the accuracy of classification body mass index (BMI) using simplified tables for BMI. Then compare the level of classification accuracy between classification using simplified tables for body mass index with Ordinal Logistic Regression and Discriminant Analysis using simulated data. The case study of this paper is about classification adult nutrient status whereas nutrient status standart WHO. Independent variables was used in this paper are weight (kg) and height (cm) which follow multivariat normal distribution.

Simulation data are used five levels of sample size $n_1=200, n_2=400, n_3=600, n_4=800,$ and $n_5=1000,$ every levels of sample size will be replicated as much as 100 times to see the consistency of each method in the classification. Apparent Error Rate is used to compare the level of classification accuracy. The result showed that the apparent error rate of ordinal logistic regression is 28,56%, disriminan analysis is 28,49%, and using body mass index simplified tables classification is 31,08%. Classification using simplified tables for body mass index has greatest APER than ordinal logistic regression and discriminan analysis. Discriminant analysis is better to classify adult nutrient status than ordinal logistic regression.

Keywords: Simplified Tables for Body Mass Index, Ordinal Logistic Regression, Discriminant Analysis, Classification

1. PENDAHULUAN

Status gizi orang dewasa diklasifikasikan menjadi empat kelompok, yaitu *underweight*, *normoweight*, *overweight*, dan *obesity*. Pada umumnya ahli gizi melakukan pengklasifikasian status gizi dewasa dengan menggunakan metode indeks massa tubuh berdasarkan tabel klasifikasi indeks massa tubuh yang telah ada, yaitu tabel baku rujukan yang dikeluarkan oleh WHO.

Pengklasifikasian merupakan salah satu metode statistika dalam pengelompokan suatu data yang disusun secara sistematis. Pada penelitian ini membandingkan antara teknik klasifikasi regresi logistik, yaitu regresi logistik ordinal dengan analisis diskriminan yang bersifat multipel (analisis diskriminan multipel). Teknik klasifikasi yang digunakan regresi logistik ordinal karena status gizi dibedakan menjadi empat kelompok yang sifatnya menunjukkan tingkatan. Analisis diskriminan multipel digunakan karena status gizi dewasa dibedakan menjadi empat kelompok.

Kelebihan dari regresi logistik adalah memiliki odds rasio yang menunjukkan seberapa besar pengaruh variabel prediktor suatu kategori referensi pada suatu variabel respon dan dalam pemenuhan asumsi tidak memerlukan asumsi normalitas multivariat serta kesamaan matriks varians-kovarians sehingga metode ini cukup tahan (*robust*) untuk dapat diterapkan dalam berbagai keadaan data (Tatham *et. Al*, 1998). Kelemahan yang dihasilkan model regresi logistik berupa nilai probabilitas yang kurang praktis. Pada analisis diskriminan, memiliki kelebihan yaitu memberikan perhitungan yang lebih efisien (Sharma, 1996), sedangkan kelemahan dari analisis diskriminan adalah asumsi harus terpenuhi, dimana data harus memenuhi distribusi normal multivariat dan menghasilkan matriks varians-kovarians yang sama setiap kelompok (Johnson dan Wichern, 2002). Metode klasifikasi yang baik adalah metode yang menghasilkan ketepatan klasifikasi yang tinggi atau kesalahan klasifikasi yang minimal. Berdasarkan uraian di atas, pada penelitian ini akan dilakukan perbandingan ketepatan klasifikasi status gizi orang dewasa antara indeks massa tubuh berdasarkan tabel klasifikasi indeks massa tubuh, regresi logistik ordinal, dan analisis diskriminan.

2. TINJAUAN PUSTAKA

2.1 Indeks Massa Tubuh (IMT)

Indeks massa tubuh (IMT) merupakan alat atau cara yang sederhana untuk memantau status gizi orang dewasa. Pengukuran indeks massa tubuh hanya membutuhkan dua hal yaitu berat badan (kg) dan tinggi badan (cm). Untuk mengetahui nilai IMT, dihitung dengan rumus berikut :

$$IMT = \frac{BB(kg)}{[TB(m)]^2} \quad (2.1)$$

Pengklasifikasian status gizi dewasa berdasarkan tabel klasifikasi indeks massa tubuh sebagai berikut :

Tabel 2.1 Tabel Klasifikasi Status Gizi Dewasa

Klasifikasi	IMT(kg/m ²)
<i>Underweight</i>	IMT ≤ 18,5
Normal	18,5 < IMT < 25
<i>Overweight</i>	25 ≤ IMT < 30
Obesitas	IMT ≥ 30

(Sumber : WHO, 2000)

2.2 Regresi Logistik Ordinal

Regresi logistik ordinal adalah suatu analisis regresi yang digunakan untuk menggambarkan hubungan antara variabel respon dengan sekumpulan variabel prediktor, dimana variabel respon bersifat ordinal, yaitu mempunyai lebih dari dua kategori ($k > 2$) dan setiap kategori dapat diperingkat. Variabel bebas dapat berupa data kategori atau kontinu yang terdiri atas dua variabel atau lebih. Metode yang sering digunakan untuk variabel respon berskala ordinal adalah dengan membentuk fungsi logit dari peluang kumulatif. Sehingga model logistik untuk respon ordinal sering disebut sebagai model logit kumulatif (Agresti, 2007). Peluang kumulatif, $P(Y \leq j | x_i)$ didefinisikan sebagai berikut :

$$P(Y \leq j | x_i) = \pi_j(x_i) = \frac{\exp\left(\beta_{0j} + \sum_{k=1}^p \beta_k x_k\right)}{1 + \exp\left(\beta_{0j} + \sum_{k=1}^p \beta_k x_k\right)} \quad (2.2)$$

Jika terdapat k kategori respon maka model regresi logistik ordinal yang terbentuk sebanyak $k - 1$. Model logistik kumulatif didefinisikan sebagai berikut:

$$\text{Logit}P(Y \leq j | x_i) = \beta_{0j} + \sum_{k=1}^p \beta_k x_k \quad j = 1, 2, \dots, k-1 \quad (2.3)$$

Jika terdapat empat kategori respon dimana $j = 1, 2, 3, 4$ maka nilai masing-masing peluang kategori respon ke- j diperoleh dengan persamaan :

$$\begin{aligned} \pi_1(x_i) &= P(Y \leq 1 | x_i) \\ \pi_2(x_i) &= P(Y \leq 2 | x_i) - P(Y \leq 1 | x_i) \\ \pi_3(x_i) &= P(Y \leq 3 | x_i) - P(Y \leq 2 | x_i) \\ \pi_4(x_i) &= 1 - P(Y \leq 3 | x_i) \end{aligned} \quad (2.4)$$

Untuk klasifikasi nilai $\pi_j(x_i)$ pada persamaan (2.4) akan dijadikan pedoman pengklasifikasian. Suatu pengamatan akan masuk dalam respon kategori j berdasarkan nilai $\pi_j(x_i)$ yang terbesar (Wibowo, 2002).

2.3 Analisis Diskriminan

Analisis diskriminan adalah teknik statistika yang dipergunakan untuk mengklasifikasikan suatu observasi ke dalam suatu kelompok berdasarkan sekumpulan variabel-variabel (Johnson & Wichern, 2002). Pemodelan antara variabel prediktor dan variabel respon dinyatakan oleh suatu fungsi yaitu fungsi diskriminan. Fungsi ini dibentuk dengan memaksimalkan jarak antar kelompok, sehingga memiliki kemampuan untuk membedakan antar kelompok. Fungsi diskriminan bersifat memaksimalkan rasio keragaman antar kelompok dengan keragaman dalam kelompok. Asumsi yang harus dipenuhi dalam analisis diskriminan adalah data berdistribusi normal multivariat dan adanya kesamaan struktur matriks varians-kovarians antar kelompok. Model umum analisis diskriminan merupakan suatu kombinasi linier bentuknya sebagai berikut:

$$Z_{ij} = a + W_1 X_{i1} + W_2 X_{i2} + \dots + W_p X_{ip} \quad (2.5)$$

Fungsi diskriminan linier (Johnson dan Wichern, 2002) adalah :

$$d_i(x) = \mu_i^t \Sigma^{-1} x - \frac{1}{2} \mu_i^t \Sigma^{-1} \mu_i + \ln p_i \quad i = 1, 2, \dots, k \quad (2.6)$$

Dengan μ_i adalah rata-rata populasi ke- i , Σ^{-1} adalah invers matriks kovariansi dan p_i adalah peluang *prior* pada populasi ke- i . Estimasi dari $d_i(x)$ adalah $\hat{d}_i(x)$

$$\hat{d}_i(x) = \bar{x}_i^t S_{gab}^{-1} x - \frac{1}{2} \bar{x}_i^t S_{gab}^{-1} \bar{x}_i + \ln p_i \quad (2.7)$$

dengan \bar{x}_i adalah rata-rata sampel kelompok ke- i , $\hat{d}_i(x)$ disebut juga sebagai fungsi diskriminan linier dengan parameter-parameter $\beta_i = \bar{x}_i^t S_{gab}^{-1}$ dan konstanta

$c_i = -\frac{1}{2} \bar{x}_i^t S_{gab}^{-1} \bar{x}_i + \ln p_i$ untuk $i = 1, 2, \dots, k$. Alokasi x ke dalam kelompok k apabila

$$\hat{d}_k(x) = \max_{(t)} (\hat{d}_1(x), \hat{d}_2(x), \dots, \hat{d}_k(x)).$$

Pengklasifikasian obyek ke dalam populasi yang terdekat setara dengan pengklasifikasian obyek ke dalam populasi dengan peluang posterior yang paling besar. Besarnya peluang posterior tersebut yaitu (Johnson dan Wichern, 2002) :

$$P(t | x) = \frac{e^{-\frac{1}{2}d_t^2(x)}}{\sum_{i=1}^k e^{-\frac{1}{2}d_i^2(x)}} \quad t = 1, 2, \dots, k \quad (2.8)$$

3. PERBANDINGAN KETEPATAN KLASIFIKASI

Data yang digunakan adalah data simulasi dengan lima ukuran sampel $n_1 = 200$, $n_2 = 400$, $n_3 = 600$, $n_4 = 800$, dan $n_5 = 1000$. Semua ukuran sampel akan dilakukan replikasi sebanyak 100 kali untuk melihat kekonsistenan masing-masing metode dalam pengklasifikasian. Algoritma data simulasi dalam membandingkan kinerja setiap teknik klasifikasi dengan 100 replikasi data simulasi sebagai berikut :

1. Bangkitkan sampel acak berdistribusi normal multivariat yang berasal dari masing-masing kelompok dengan metode Monte Carlo, dimana ukuran sampel masing-masing kelompok sama, dengan skenario :

$$\mu = \begin{bmatrix} 53.93 \\ 153.39 \end{bmatrix}; \Sigma = \begin{bmatrix} 84.79 & 25.43 \\ 25.43 & 49.13 \end{bmatrix}$$

$$\mu_1 = \begin{bmatrix} 51.88 \\ 179.10 \end{bmatrix}; \Sigma_1 = \begin{bmatrix} 84.79 & 25.43 \\ 25.43 & 49.13 \end{bmatrix}$$

$$\mu_2 = \begin{bmatrix} 64.15 \\ 170.69 \end{bmatrix}; \Sigma_2 = \begin{bmatrix} 84.79 & 25.43 \\ 25.43 & 49.13 \end{bmatrix} .$$

$$\mu_3 = \begin{bmatrix} 74.26 \\ 164.66 \end{bmatrix}; \Sigma_3 = \begin{bmatrix} 84.79 & 25.43 \\ 25.43 & 49.13 \end{bmatrix}$$

$$\mu_4 = \begin{bmatrix} 81.38 \\ 154.38 \end{bmatrix}; \Sigma_4 = \begin{bmatrix} 84.79 & 25.43 \\ 25.43 & 49.13 \end{bmatrix}$$

2. Lakukan klasifikasi untuk masing-masing objek dengan menggunakan Regresi Logistik Ordinal, Analisis Diskriminan, dan Tabel Klasifikasi Indeks Massa Tubuh dengan bantuan *Software R 3.1.0* .
3. Hitung ketepatan klasifikasi untuk masing-masing metode klasifikasi, dengan metode *Apparent Error Rate (APER)* .
4. Ulangi langkah 1 sampai dengan 3 sebanyak 100 kali.

Evaluasi prosedur klasifikasi dengan nilai *APER* menyatakan nilai proporsi sampel yang salah diklasifikasikan (Johnson dan Wichern, 2002).

Aktual	Prediksi				Total
	$\hat{\pi}_1$	$\hat{\pi}_2$	$\hat{\pi}_3$	$\hat{\pi}_4$	
π_1	n ₁₁	n ₁₂	n ₁₃	n ₁₄	A
π_2	n ₂₁	n ₂₂	n ₂₃	n ₂₄	B
π_3	n ₃₁	n ₃₂	n ₃₃	n ₃₄	C
π_4	n ₄₁	n ₄₂	n ₄₃	n ₄₄	D
Total	E	F	G	H	I

Rumus ketepatan klasifikasi tersebut adalah sebagai berikut :

Melihat ketepatan klasifikasi secara keseluruhan nilai APER adalah

$$APER = \left(\frac{n_{12} + n_{13} + n_{14} + n_{21} + n_{23} + n_{24} + \dots + n_{43}}{I} \right) \times 100\%$$

4. HASIL DAN PEMBAHASAN

Hasil kesalahan klasifikasi (APER) dengan regresi logistik ordinal, analisis diskriminan, dan berdasarkan pengklasifian tabel klasifikasi indeks massa tubuh untuk seluruh tingkatan ukuran sampel dengan replikasi sebanyak 100 kali ditunjukkan pada tabel 4.1.

Tabel 4.1. Hasil Persentase Kesalahan Klasifikasi (APER) Regresi Logistik Ordinal, Analisis Diskriminan, dan Tabel Klasifikasi Indeks Massa Tubuh

n	APER Regresi Logistik Ordinal	APER Analisis Diskriminan	APER Indeks Massa Tubuh
200	28,08	28,17	30,8
400	28,63	28,54	31,12
600	28,8368	28,6801	31,3433
800	28,64625	28,5225	31,06625
1000	28,642	28,564	31,118
Rata-rata kesalahan klasifikasi	28,56701	28,49532	31,08951

Rata-rata kesalahan klasifikasi (APER) yang dihasilkan regresi logistik ordinal sebesar 28,56%, analisis diskriminan 28,49%, dan berdasarkan tabel klasifikasi indeks massa tubuh sebesar 31,08%. Dari Tabel 4.1 dapat terlihat bahwa APER yang dihasilkan oleh analisis diskriminan untuk semua tingkat ukuran sampel dengan replikasi sebanyak 100 kali menghasilkan kesalahan klasifikasi paling kecil dibandingkan dengan regresi logistik ordinal dan tabel klasifikasi indeks

massa tubuh. Pengklasifikasian status gizi dewasa yang selama ini dilakukan berdasarkan tabel klasifikasi indeks massa tubuh menghasilkan tingkat kesalahan klasifikasi paling besar dibandingkan dengan regresi logistik ordinal dan analisis diskriminan.

5. KESIMPULAN DAN SARAN

5.1 Kesimpulan

1. Untuk semua tingkatan ukuran sampel dengan replikasi sebanyak 100 kali dan memenuhi asumsi normal multivariat serta matriks varians-kovarians untuk setiap kelompok sama dalam pengklasifikasian status gizi dewasa menghasilkan analisis diskriminan memberikan nilai APER yang lebih kecil dibandingkan dengan regresi logistik ordinal. Analisis diskriminan lebih baik diterapkan untuk klasifikasi status gizi dewasa dibandingkan dengan regresi logistik ordinal.
2. Pengklasifikasian status gizi dewasa dengan indeks massa tubuh berdasarkan tabel klasifikasi indeks massa tubuh menghasilkan nilai APER yang paling besar dibandingkan dengan teknik klasifikasi statistika, yaitu analisis diskriminan dan regresi logistik ordinal.
3. Variabel yang digunakan dalam penelitian ini, yaitu berat badan dan tinggi badan yang bersifat kontinu dan bivariat normal sehingga analisis diskriminan menghasilkan ketepatan klasifikasi yang lebih baik dibandingkan dengan regresi logistik ordinal.

5.2 Saran

1. Dalam penelitian ini hanya melibatkan data simulasi normal bivariat, maka disarankan untuk penelitian selanjutnya digunakan data normal multivariat.
2. Perlu adanya evaluasi atau penelitian mengenai tabel klasifikasi indeks massa tubuh agar menghasilkan ketepatan klasifikasi yang lebih baik lagi

3. Metode Analisis Regresi Logistik Ordinal dan Analisis Diskriminan perlu dibandingkan dengan metode lain agar diperoleh suatu metode klasifikasi yang lebih baik lagi.

DAFTAR PUSTAKA

- Agresti A. 2007. *An Introduction to Categorical Data Analysis*. New York. John Wiley and Sons
- Badan Penelitian dan Pengembangan Kesehatan. 2010. *Riset Kesehatan Dasar (Riskesdas 2010)*. Kementerian Kesehatan Republik Indonesia, Jakarta.
- Centre for Obesity Research and Education. 2007. *Body Mass Index: BMI Calculator*. Didapat dari <http://www.core.monash.org/bmi.html> [Agustus 2014].
- Depkes, RI. 2000. *Pedoman Praktis Untuk Mempertahankan Berat Badan Normal Berdasarkan Indeks Massa Tubuh (IMT) dengan Gizi Seimbang*. Jakarta: Depkes RI.
- Dillon, W.R. dan Goldstein M. 1984. *Multivariate Analysis: Methods and Applications*. New York: John Wiley & Sons, Inc..
- Grafik Indeks Massa Tubuh Anak. http://en.wikipedia.org/wiki/Body_mass_index
- Hajarisman, N. 2008. *Statistika Multivariat*, UNISBA Bandung
- Hosmer, David W., Lemeshow S. 2000. *Applied Logistic Regression*. New York. John Wiley and Sons.
- Johnson, R.A dan Wichern, D.W. 2002, *Applied Multivariate Statistical Analysis, fifth edition*, Prantice Hall, New Jersey.
- Kleinbaum, David G, Mitchel Klein. (2010). *Logistic Regression*. Springer Science-Bussiness Media. New York.
- Krzanowski, W.J. (1975), "Discrimination and Classification using Both Binary and Continuous Variable", *Journal of the American Statisitical Association*, 70;782-352.
- Manel, S, Marie, DJ, dan Ormerod, SJ. (1999). *Comparison disrciminant analysis, neural networks, and logistic regression for predisting species distributions: a case study with Himalayan river bird*, *Ecological Modelling*, 120;337-347.

- Pohar, M., Blas M., dan Turk, S. (2004). *Comparison of Logistic Regression and Linear Discriminant Analysis : A Simulation Study*, Metodološki zvezki, Vol.1, No.1, hal.143-161.
- Press, S.J. dan Wilson, S. (1978), "Choosing Between Logistic Regression and Discrimination Analysis", *Journal of the American Statistical Association*, 73;699-364.
- Sharma, S. (1996), "Applied Multivariate Techniques", John Wiley, New York.
- Sugondo, S., 2006. *Obesitas. Dalam: Sudoyono, W.A., Setiyohadi, B., Alwi, I., Simadibrata, M., & Setiati, S. Buku Ajar Ilmu Penyakit Dalam. Jilid III. Edisi 4. Jakarta: Pusat Penerbitan Departemen Ilmu Penyakit Dalam FKUI, 1919-1924.*
- Supranto, J. 2004. *Ekonometri Jilid 2*. Jakarta: Penerbit Ghalia Indonesia.
- Tatham, R.L., Hair, J.F, Anderson, R.E., dan Black, W.C., (1998), "*Multivariate Data Analysis*", Prentice Hall, New Jersey.
- Yulia I.P.D. (1997), "Analisis Statistika Terhadap Kerusakan Wall Tile di PD Sarana Bangunan Unit Pabrik Keramik dan Pengolahan Bahan Tulungagung", Tugas Akhir, Statistika ITS.
- Wibowo, W. (2002), "Perbandingan Hasil Klasifikasi Analisis Diskriminan dan Regresi Logistik pada Pengklasifikasian Data Respon Biner", *KAPPA*, Vol. 3, No. 1, hal. 36-45.
- WHO. 2000. *Body Mass Index (BMI) = Indeks Massa Tubuh*. [http://www.obesitas.web.id/indonesia/bmi\(i\).htm](http://www.obesitas.web.id/indonesia/bmi(i).htm) [Juli 2014].