

Model Regresi Logistik Biner pada Tingkat Pengangguran Terbuka di Provinsi Sulawesi Barat Tahun 2017

Alfiah Safitri¹, Sudarmin¹, Muhammad Nusrang¹

¹Prodi Statistika FMIPA UNM Makassar

e-mail: alfiahsafitri15@gmail.com

Abstrak Regresi logistik merupakan model regresi yang sering digunakan untuk pemodelan data kategori. Pada penelitian ini dilakukan pemodelan regresi logistik dan penerapannya pada Tingkat Pengangguran Terbuka di Provinsi Sulawesi Barat. Pendugaan parameter yang digunakan adalah Maximum Likelihood Estimation (MLE). Metode uji G simultan dan uji parsial digunakan untuk signifikansi peubah. Tujuan dari penelitian ini yaitu untuk mencari model terbaik yang terbentuk dari faktor-faktor yang berpengaruh terhadap Tingkat Pengangguran Terbuka di Provinsi Sulawesi Barat. Hasil dari penelitian ini menunjukkan bahwa peubah bebas yang signifikan terhadap Tingkat Pengangguran Terbuka di Provinsi Sulawesi Barat adalah persentase jenis kelamin dengan kategori laki-laki dan persentase lapangan usaha dengan kategori pertanian dimana pada peubah jenis kelamin memiliki nilai p-value $0,017 < 5\%$ dan pada peubah lapangan usaha memiliki nilai p-value $0,006 < 5\%$ dapat di artikan bahwa kedua peubah tersebut berpengaruh terhadap Tingkat Pengangguran Terbuka.

Kata Kunci: Regresi Logistik, Kategor, Tingkat Pengangguran Terbuka

Abstract Logistic Regression is a regression model that is often used for category data modeling. In this research carried out logistic regression modeling and its application to the Open Unemployment Rate in West Sulawesi Province. Parameter estimate that is used the Maximum Likelihood Estimate (MLE). G test method and partial test that is used for variable significance. The purpose of this research is to look for the best model formed from factors that influence of Open Unemployment Rate in West Sulawesi Province. The results from this research showed that independent variable is sex percentage with male category and business field percentage with agriculture category which of sex variable have p-value $0,017 < 5\%$ percentage and in the business field Variable have p-value $0,006 < 5\%$ percentage can be interpreted that both of variable are influential of the Open Unemployment Rate.

Keywords: logistic Regression, category, Open Unemployment Rate.

1. PENDAHULUAN

Salah satu permasalahan yang ada di Provinsi Sulawesi Barat tidak jauh berbeda dengan permasalahan nasional, yakni masih tingginya angka pengangguran jika dibandingkan dengan beberapa provinsi lain di Pulau Sulawesi. Oleh karena itu, pengangguran menjadi tanggung jawab bersama terutama pemerintah. Pemerintah sebagai penyangga proses perbaikan kehidupan masyarakat dalam sebuah pemerintahan perlu mencari jalan keluar dengan merumuskan langkah-langkah yang sistematis dan strategis sebagai upaya penanganan permasalahan pengangguran.

Masalah pengangguran memang selalu menjadi suatu persoalan dipecahkan dalam perekonomian Indonesia. Jumlah penduduk yang bertambah setiap tahun mengakibatkan bertambahnya jumlah angkatan kerja dan tentunya akan memberikan makna bahwa jumlah orang yang mencari pekerjaan akan meningkat, seiring dengan itu tenaga kerja juga akan bertambah.

Mengatasi permasalahan tersebut diperlukan untuk mencari tahu penyebab timbulnya tingkat pengangguran terbuka. Untuk melihat faktor-faktor yang mempengaruhi tingkat pengangguran maka digunakan regresi logistik dimana peubah terikat (Y) adalah 0 untuk bekerja dan 1 untuk tidak bekerja (pengangguran) dengan mengikuti distribusi Bernoulli.

Regresi logistik adalah suatu metode analisis statistika yang mendeskripsikan hubungan antara peubah terikat yang memiliki dua kategori atau lebih peubah bebas berskala kategori atau interval (Hosmer dan Lemeshow, 1989). Regresi logistik merupakan regresi nonlinear, digunakan untuk menjelaskan hubungan antara

X dan Y yang bersifat tidak linear, tidak normalan sebaran Y, keragaman respon tidak konstan yang tidak dapat dijelaskan dengan model linear biasa (Agresti, 1996). Apabila dilakukan uji regresi logistik dengan asumsi regresi klasik maka akan terjadi bias dari penduga semula.

Pada kasus penelitian yang terdapat peubah terikat (Y) dan peubah bebas (X) yang masing-masing peubahnya berupa data kategori, ini akan sulit di analisis dengan menggunakan metode regresi linear biasa. Dengan demikian, untuk mengatasi hal tersebut maka digunakan Regresi logistik peubah dummy.

Metode pendugaan parameter yang sering digunakan dalam model regresi logistik yaitu menggunakan penduga parameter maksimum *likelihood* (MLE).

2. TINJAUAN PUSTAKA

2.1. Regresi Logistik

Regresi logistik adalah suatu metode analisis statistika untuk mendeskripsikan hubungan antara peubah respon yang memiliki dua atau lebih kategori dengan satu atau lebih peubah penjelas yang berskala kategori atau interval (Hosmer & Lemeshow, 2000).

Analisis regresi logistik menghasilkan peubah yang bersifat biner dan dikotomi. Model regresi logistik biner digunakan jika peubah terikatnya menghasilkan dua kategori yang bernilai 0 dan 1, sehingga mengikuti distribusi *Bernoulli* sebagai berikut (Agresti, 1990):

$$f(Y = y) = \pi(x)^y (1 - \pi(x)^{1-y}), \quad y = 0, 1 \quad (2.1)$$

Dimana jika $y = 0$ maka $P(Y = 0) = 1 - \pi$ dan jika $y = 1$ maka $P(Y = 1) = \pi$.

Model regresi logistik dengan nilai peubah terikat (Y) yang dikotomi dan peubah bebas (X) untuk subjek ke- i . di asumsikan peubah Y diberi kode satu yang menunjukkan “bekerja” dan diberi kode nol untuk “ tidak bekerja ”, yang merupakan suatu hasil dikotomi.

Regresi logistik tidak memodelkan secara langsung peubah terikat (Y) dengan peubah bebas (X), melainkan melalui transformasi peubah terikat ke variabel logit yang merupakan natural log dari *odds ratio*. Model dengan p peubah bebas. Model regresi logistik dengan k buah peubah penjelas dibentuk dengan nilai sdinotasikan sebagai berikut:

$$\pi(x) = \frac{\exp(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)}{1 + \exp(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)} \quad (2.2)$$

Dimana :

$\pi(x)$:peluang kejadian sukses dengan nilai probabilitas $0 \leq \pi(x) \leq 1$

β_0 : *Intersept (bilangan konstan)*

β_1, \dots, β_p : parameter regresi logistik

X_1, \dots, X_p : Nilai peubah bebas

2.2 Penduga Parameter

Pada penelitian ini kita menggunakan metode maksimum *likelihood*. Pada dasarnya metode maksimum *likelihood* merupakan metode kuadrat terkecil tertimbang dengan beberapa proses iterasi.

Metode ini akan memberikan landasan untuk pendekatan kami dengan regresi logistik. Pengertian paling umum metode maksimum *likelihood* menghasilkan nilai untuk parameter yang tidak diketahui yang memaksimalkan kemungkinan memperoleh kumpulan data teramati. Demikian, penaksir yang

dihasilkan adalah mereka yang setuju paling mendekati dengan data yang diamati.

Jika peubah terikat (Y) dikodekan sebagai 0 atau 1 maka pernyataan diberikan untuk persamaan (2.8) memberikan probabilitas bersyarat bahwa Y sama dengan 1 diberikan x . Ini dilambangkan sebagai $P(Y = 1|x)$. Maka kuantitas $1 - \pi(x)$ memberikan probabilitas bersyarat bahwa Y sama dengan nol yang diberi x , $P(Y = 0|x)$. Kemudian, untuk pasangan (x_i, y_i) , dimana $y_i = 1$, kontribusi untuk fungsi *likelihood* adalah $\pi(x_i)$, dan untuk pasangan itu dimana $y_i=0$, kontribusi untuk fungsi *likelihood* adalah $1 - \pi(x_i)$, dimana kuantitas $\pi(x_i)$ menunjukkan nilai $\pi(x)$ dihitung pada x_i . Maka fungsi maksimum *likelihood* adalah:

$$l(\beta) = \prod_{i=1}^n \pi(x_i)^{y_i} [1 - \pi(x_i)]^{1-y_i}, i= 1, 2, \dots, p \quad (2.3)$$

Dengan :

y_i = pengamatan pada peubah penjelas ke- i

$\pi(x_i)$ = peluang untuk peubah penjelas ke- i

Parameter β_i diduga dengan memaksimumkan persamaan diatas untuk mempermudah perhitungan maka dilakukan pendekatan logaritma sehingga fungsi *loglikelihood* sebagai berikut:

$$L(\beta) = \ln[l(\beta)] = \sum_{i=1}^n \{y_i \ln[\pi(x_i)] + (1 - y_i) \ln[1 - \pi(x_i)]\}$$

Nilai dugaan β dapat diperoleh dengan membuat turunan pertama dari $L(\beta)$ terhadap β yang hasilnya sama dengan nol dengan $i = 1, 2, \dots, p$. Dari nilai tersebut dapat diketahui penduga $\pi(x)$ dimana $\hat{g}(x)$ adalah penduga logit sebagai fungsi linier dari peubah penjelas (Hosmer dan Lemeshow, 2000).

3. METODE PENELITIAN

3.1 Sumber Data

Data yang digunakan pada penelitian ini adalah data sekunder yang berupa row data yang bersumber dari Badan Pusat Statistik Provinsi Sulawesi Barat. Row data tersebut diolah lebih lanjut untuk dijadikan bahan penelitian penyusunan skripsi yang berjudul “Aplikasi Regresi Logistik Untuk Melihat Faktor Yang Mempengaruhi Tingkat Pengangguran Terbuka di Provinsi Sulawesi Barat.

Adapun peubah penjelas dalam penelitian ini antara lain yaitu jenis kelamin, tingkat pendidikan, usia, daerah tempat tinggal dan lapangan usaha. Sedangkan peubah terikat yang akan digunakan di sini adalah tingkat pengangguran di Provinsi Sulawesi Barat.

3.2 Definisi Operasional Peubah

Peubah yang digunakan dalam penelitian ini terdiri atas peubah (Y) dari dua kategori yaitu kategori 0 adalah tidak bekerja dan kategori 1 adalah memiliki pekerjaan (bekerja). Adapun peubah penjelas (X) yang digunakan dalam penelitian ini ditampilkan pada Tabel 3.1 berikut:

1. Pengangguran meliputi penduduk yang tidak bekerja tetapi sedang mencari pekerjaan, atau mempersiapkan suatu usaha, atau merasa tidak mungkin mendapatkan pekerjaan (putus asa), atau

sudah diterima bekerja tetapi belum memulai (BPS, 2017).

2. Usia kerja adalah penduduk usia kerja adalah penduduk yang berusia 15 tahun ke atas, sesuai dengan ketentuan dalam Undang-Undang Ketenagakerjaan No. 13 Tahun 2003 (BPS, 2016).
3. Tingkat Pendidikan adalah Tingkat pendidikan dalam hal ini merupakan tingkat pendidikan terakhir yang ditamatkan oleh responden dengan ijazah / STTB tertinggi yang dimiliki.
4. Jenis kelamin adalah jenis kelamin responden yaitu laki-laki atau perempuan
5. Daerah tempat tinggal adalah domisili responden di daerah pedesaan atau di perkotaan.
6. Lapangan usaha/jenis pekerjaan adalah bidang kegiatan dari pekerjaan/usaha/perusahaan/kantor tempat seseorang bekerja, atau yang dihasilkan oleh perusahaan/kantor tempat responden bekerja. lapangan usaha tersebut di kelompokkan menjadi:
 - a. Pertanian meliputi : pertanian, kehutanan, perburuan dan perikanan.
 - b. Industri antara lain: lapangan usaha pertambangan; industri pengolahan; listrik, gas dan air; serta bangunan.
 - c. Jasa-jasa antara lain: Perdagangan besar dan eceran, rumah makan dan hotel; angkutan, pergudangan, komunikasi; keuangan, asuransi, usaha persewaan bangunan, tanah dan jasa perusahaan; jasa kemasyarakatan.

Tabel 3.1 Peubah-peubah faktor tingkat pengangguran

Peubah	Keterangan	Jenis	Keterangan
Y	Tingkat pengangguran	Kategori	1 = bekerja 0 = tidak bekerja
X ₁	Usia	Kontinu	-
X ₂	Tingkat pendidikan	Kategori	2 = ≤ SLTP 1 = SLTA 0 = Diploma / S1/S2
X ₃	Jenis kelamin	Kategori	1 = laki-laki 0 = perempuan
X ₄	Daerah tempat tinggal	Kategori	1 = perkotaan 0 = perdesaan
X ₅	Lapangan Usaha	Kategori	2 = Jasa 1 = pertanian 0 = industry

3.3 Teknik Analisis Data

Adapun teknik analisis data pada penelitian ini yaitu:

1. Melakukan eksplorasi data untuk mengetahui gambaran dari Tingkat Pengangguran Terbuka di Provinsi Sulawesi Barat.
2. Menentukan peubah-peubah penjelas yang akan digunakan
3. Melakukan pendugaan parameter regresi logistik
4. Melakukan uji signifikan model
5. Melakukan uji kesesuaian model
6. Interpretasi dengan menggunakan *odds ratio*
7. Kesimpulan dan saran

4. HASIL DAN PEMBAHASAN

4.1 Model Regresi Logistik

Analisis regresi logistik merupakan salah satu analisis statistika yang dapat digunakan untuk mengetahui faktor-faktor yang mempengaruhi tingkat pengangguran yang ada di Provinsi Sulawesi Barat. Kemudian kita melakukan pengujian dengan uji simultan dan uji parsial untuk mengetahui peubah-peubah yang signifikan. Dengan peubah terikat (Y) yang digunakan adalah Tingka Pengangguran Terbuka yang dikategorikan sebagai “bekerja” dan “tidak bekerja”.

Model regresi logistik biner pada data penelitian ini menggunakan lima peubah bebas secara bersamaan. Hasil nilai uji G pada model regresi logistik adalah:

Tabel 4.6 Uji signifikansi parameter secara serentak

	chisq	Df	$\chi^2_{(a,db)}$
Model	42,858	7	14,067

Berdasarkan Tabel 4.6 dapat dilihat bahwa nilai statistik uji G yang dihasilkan pada model regresi logistik adalah $42,858 > \chi^2_{(0,05;7)} = 14,067$. Dapat disimpulkan pengujian model regresi logistik pada uji simultan adalah tolak H₀, yang berarti minimal terdapat satu peubah penjelas yang berpengaruh terhadap tingkat pengangguran Provinsi Sulawesi Barat. Selanjutnya dilakukan uji parsial untuk mengetahui peubah mana yang berpengaruh signifikan. Penduga parameter dan hasil uji hipotesis secara parsial dari model regresi logistik ditunjukkan pada Tabel 4.7.

Tabel 4.7 Pendugaan parameter model regresi logistik

Parameter	estimasi	Standar Error	p-value	Odds Ratio
	0,382	1,844	0,835	
X ₁	-0,037	0,032	0,252	9,363
X ₂₁	-0,776	1,814	0,669	4,602
X ₂₂	-0,087	1,572	0,956	9,168
X ₃₁	1,865	0,779	0,017	6,457
X ₄₁	-0,485	0,958	0,613	6,159
X ₅₁	3,218	1,182	0,006	2,499
X ₅₂	20,577	2203,000	0,993	8,640

Berdasarkan Tabel 4.7 dapat dilihat bahwa peubah yang signifikan yaitu ditujukan pada peubah jenis kelamin (X₃₁) dengan nilai *p-value* = 0,017 < 5% dan pada peubah lapangan usaha (X₅₁) dengan nilai *p-value* = 0,006 < 5% berpengaruh terhadap tingkat pengangguran di Provinsi Sulawesi Barat.

Sehingga model regresi logistik yang terbentuk adalah dapat dilihat pada persamaan di bawah ini

$$\pi(x) = \frac{\exp(0,38 + 1,86x_{31} + 3,21x_{51})}{1 + \exp(0,38 + 1,86x_{31} + 3,21x_{51})} \tag{4.1}$$

Interpretasi berdasarkan nilai *odds ratio* maka dapat dijelaskan bahwa hasil penelitian yang menunjukkan adanya hubungan peubah yang berpengaruh terhadap Tingkat Pengangguran Terbuka di Provinsi Sulawesi Barat yaitu peubah jenis kelamin (X₃₁) adalah positif dengan nilai *odds ratio* sebesar 6,46 yang menjelaskan bahwa peubah jenis kelamin (X₃₁) peluang untuk tidak bekerja sebesar 6,46 kali. Sedangkan hasil untuk peubah lapangan usaha (X₅₁) menunjukkan adanya pengaruh signifikan terhadap Tingkat pengangguran Terbuka memiliki peluang untuk tidak bekerja dengan nilai *odds ratio* sebesar 2,49 kali.

4.2 Uji Kesesuaian Regresi Logistik

Uji kesesuaian model regresi logistik digunakan untuk mengevaluasi apakah model sesuai atau tidak sesuai dengan data, nilai observasi yang diperoleh sama atau mendekati dengan yang diharapkan dalam model. Untuk melihat hasil sesuai atau tidaknya data dalam model regresi logistik dapat dilihat dengan

menggunakan uji Hosmer dan Lemeshow, hasil yang diperoleh dilihat pada tabel di bawah ini:

Tabel 4.8 Hasil uji kesesuaian model

<i>X-Square</i>	<i>df</i>	<i>p-value</i>
6,933	8	0,544

Berdasarkan pada Tabel 4.8 menunjukkan hasil yang diperoleh sebesar 6,933 dan $p\text{-value} = 0,544$. Dapat diartikan bahwa gagal tolak H_0 karena $p\text{-value} > \alpha$ dengan $\alpha = 5\%$. Maka dapat disimpulkan bahwa model sesuai dengan observasi data yang diharapkan atau tidak terdapat perbedaan yang nyata antara observasi dengan prediksi model.

5. KESIMPULAN DAN SARAN

5.1 Kesimpulan

Dari hasil penelitian mengenai pemodelan regresi logistik terhadap tingkat pengangguran terbuka di Provinsi Sulawesi Barat.

1. Dari hasil analisis regresi logistik biner tersebut maka diperoleh beberapa peubah yang berpengaruh signifikan terhadap tingkat pengangguran di Provinsi Sulawesi Barat yaitu peubah jenis kelamin (X_3) dan lapangan usaha (X_5) dimana *odds ratio* untuk peubah jenis kelamin (X_3) sebesar 6,46 yang berarti peubah jenis kelamin memiliki kecenderungan sebesar 6,46 kali (dibandingkan dengan responden yang berjenis kelamin perempuan) untuk peubah tidak bekerja (dibandingkan dengan yang bekerja). Sedangkan *odds ratio* untuk peubah lapangan usaha (X_5) sebesar 2,49 yang berarti peubah lapangan usaha memiliki kecenderungan sebesar 2,49 kali (dibandingkan dengan lapangan usaha jasa dan industri).

5.2 Saran

1. Diharapkan untuk penelitian selanjutnya agar mengkaji lebih mendalam lagi mengenai model regresi logistik ordinal terutama dalam pengembangannya untuk 2 kategori atau lebih.
2. Analisis yang dilakukan pada penelitian ini merupakan analisis regresi logistik dengan menggunakan data biner, untuk itu diharapkan pada peneitian-penelitian berikutnya dilakukan dengan menggunakan regresi logistik multinom.
3. Bagi para peneliti yang ingin melakukan penelitian yang menggunakan metode regresi logistik biner dapat menggunakan hasil analisis ini sebagai bahan perbandingan, dan diharapkan bersedia menindaklanjuti jika terdapat hal-hal yang dianggap tidak sesuai.

Badan Pusat Statistik. (2015). *Statistik Angkatan Kerja Sulawesi Barat 2015*. Mamuju: Badan Pusat Statistik.

Badan Pusat Statistika. *Pedoman Pencacah Survei Angkatan Kerja 2017*.

De Muth, J. E. (2014). *Basic Statistics and Pharmaceutical Statical Applications (third edition)*. New York: Taylor & Francis Group, LLC.

Draper, N. R. & H. Smith. (1998). *Analisis Regresi Terapan*. New York : John Wiley and Sons, Inc

Hendayana, R. (2012). *Application Method of Logistic Regression Analyze the Agricultural Technology Adaption*.

Hosmer, D. W., & Lemeshow, S. (2000). *Applied Logistic Regression (Second Edition)*. New York: John. Wiley & Sond, INC.Juanda, B (2009). *EKONOMETRIKA: Pemodelan dan Pendugaan*. IPB Press

Kleinbaum, D. G. & Klein, M. (2010). *Logistic Regression (third edition)*. New York: Springer

Ningtias, I. P., & Rahayu, S. P. (2017). Pemodelan Faktor-faktor yang Mempengaruhi Tingkat Pengangguran Terbuka di Provinsi Jawa Timur Tahun 2015 Menggunakan Regresi Spasial. *jurnal SAINS DAN SENI ITS*, 212-218

Oke & Oyeka, I. C. A. (2013). Estimating The Fisher's Scoring Matrix Formula from Logistic Model. *American Journal Of Theoretical and Applied Statistics*. 2013

Tiro, M. A. (2004). *Analisis Regresi dengan Data Kategori(second edition)*. Makassar: Andira Publisher.

Tiro, M. A., Sukarna., & Aswi. (2008). *Pengantar Teori Peluang*. Makassar: Andira Publisher

Walpole, R. E. (2017). *Pengantar Statistika (third edition)*. Jakarta: PT Gramedia

DAFTAR PUSTAKA

Agresti, A. (1990). *Categorical Data Analysis*. New York: John Willey & Sons.

Agresti, A. (2002). *Categorical Data Analysis (second edition)*. New York: John Willey & Sons.