

Model Regresi Tersensor Untuk *Mixture Data*

Defi Yusti Faidah, Gungum Darmawan, Resa Septiani Pontoh
Departemen Statistika FMIPA Universitas Padjadjaran
defi.yusti@unpad.ac.id

Abstrak— Seringkali pada kasus khusus terdapat nilai nol yang berlebih dan sisanya memiliki nilai yang cukup beragam. Data tersebut terdiri dari data yang berskala diskrit dan berskala kontinu. Karena memiliki struktur data yang berbeda maka disebut dengan *mixture data*. *Mixture data* biasa dikenal dengan data tersensor. Ciri lain dari data tersensor adalah sebagian nilai dari suatu rentang tertentu ditransformasikan sebagai suatu nilai tunggal. Dibutuhkan metode khusus untuk mengolah data tersebut. Penggunaan metode analisis regresi linier klasik untuk melihat hubungan variabel yang sifatnya tersensor dengan variabel prediktor menjadi kurang tepat. Untuk mengatasi hal tersebut, maka digunakan suatu model regresi untuk data tersensor yang dikenal dengan model regresi tersensor.

Kata kunci: *Mixture Data, Regresi Tersensor*

I. PENDAHULUAN

Analisis regresi merupakan suatu metode statistika yang digunakan untuk memodelkan hubungan antara variabel independen dan variabel dependen. Terdapat beberapa jenis model regresi jika dilihat dari struktur data variabel dependen yang dianalisis. Apabila variabel dependen memiliki skala pengukuran minimal interval maka model regresi yang digunakan adalah linear klasik. Sementara itu apabila variabel dependen memiliki skala pengukuran nominal maupun ordinal maka model regresi yang digunakan adalah regresi logistik atau probit. Akan tetapi terdapat data yang memiliki struktur campuran yaitu berskala diskrit dan kontinu. Karena memiliki struktur data yang berbeda maka disebut dengan *mixture data* [1]. *Mixture data* memiliki nilai nol pada sebagian observasi dan sisanya memiliki nilai yang bervariasi. Model regresi yang dapat memodelkan variabel dependen yang berupa *mixture data* adalah regresi tersensor.

Penggunaan regresi tersensor pada *mixture data* akan mengurangi efek bias jika dibandingkan dengan data yang diolah menggunakan regresi linier klasik. Hal ini dikarenakan data yang bernilai konstan dapat diolah secara bersama dengan data kontinu sehingga tidak akan kehilangan informasi yang berasal dari data diskrit [2]. Model regresi tersensor akan menghasilkan *standard error* yang lebih *robust* dibandingkan model regresi linier klasik dan juga nilai prediksi yang kuat [3]. Perbandingan antara regresi linier klasik bivariat dengan model regresi tersensor diperoleh kesimpulan model akan menghasilkan R^2 lebih besar daripada regresi linier klasik pada kasus *mixture data* [4].

II. METODE PENELITIAN

Berdasarkan paparan pada latar belakang maka metode yang digunakan pada penelitian ini adalah sebagai berikut.

A. Model Regresi Tersensor

Regresi tersensor merupakan analisis regresi yang digunakan untuk menggambarkan hubungan antara variabel dependen (Y) dan variabel independen (X) dimana variabel dependen tersebut berskala campuran. Model ini pertama kali dikemukakan oleh James Tobin pada 1958 yang digunakan untuk menganalisis pengeluaran para rumah tangga di Amerika Serikat untuk membeli mobil [4]. Pemodelan ini diawali dengan memperhatikan model sebagai berikut [2] :

$$y_i^* = \mathbf{x}_i^T \boldsymbol{\beta} + \varepsilon_i$$

dengan y_i^* adalah variabel laten dependen yang diobservasi untuk nilai yang lebih besar dari c dan tersensor untuk nilai lainnya, \mathbf{x}_i^T adalah vektor variabel bebas

$\mathbf{x}_i^T = [1 \ X_{i1} \ \dots \ X_{ip}]$, $\boldsymbol{\beta}$ adalah vektor parameter koefisien, dengan $\boldsymbol{\beta} = [\beta_0 \ \beta_1 \ \dots \ \beta_k]^T$ dan ε_i adalah *error* yang diasumsikan berdistribusi $N(0, \sigma^2)$. Nilai observasi y_i , $i = 1, 2, \dots, n$ diperoleh dari persamaan (1) sehingga untuk persamaan data tersensor adalah

$$y_i = \begin{cases} c & \text{untuk } y_i^* \leq c \\ y_i^* = \mathbf{x}_i^T \boldsymbol{\beta} + \varepsilon_i & \text{untuk lainnya} \end{cases}$$

dengan y_i^* berdistribusi normal dengan mean $\mathbf{x}_i^T \boldsymbol{\beta}$ dan varians σ^2

Persamaan yang mengandung beberapa variabel prediktor dan berpengaruh terhadap variabel respon dapat dilakukan pengujian dengan *likelihood ratio test* [5],[6]. *Likelihood ratio test* digunakan untuk menguji estimasi parameter secara serentak, sedangkan uji *wald* digunakan untuk pengujian secara individu.

B. Sumber Data

Data yang digunakan adalah data sekunder yang yang diperoleh dari Badan Pusat Statistik (BPS) yaitu data Survei Sosial Ekonomi Nasional (Susenas) dan Survei Angkatan Kerja Nasional (Sakernas) tahun 2010. Pada penelitian ini yang dijadikan unit observasi adalah 118 kabupaten/kota di Pulau Jawa.

C. Variabel Penelitian

Variabel respon dalam penelitian ini adalah TPT perempuan dimana nilai c yang digunakan adalah 7,14. Hal ini karena pada umumnya nilai TPT perempuan Indonesia pada tahun 2010 mencapai 7,14. Terdapat sembilan variabel prediktor yang digunakan yaitu persentase penduduk yang tinggal di daerah perkotaan (X_1), angka pertumbuhan penduduk (X_2), seks rasio (X_3), persentase penduduk yang berpendidikan di atas SLTP – Perguruan Tinggi (X_4), persentase penduduk yang mampu membaca dan menulis (X_5), tingkat pertumbuhan ekonomi (X_6), dan partisipasi angkatan kerja (X_7).

D. Metode Analisis

Langkah awal untuk memodelkan TPT perempuan adalah membentuk model tersensor. adalah meregresikan variabel y terhadap semua variabel prediktor. Mencari nilai estimasi parameter dengan metode *MLE*. Melakukan pengujian terhadap estimasi parameter yang telah didapat pada langkah b dengan menggunakan *LR test* untuk keseluruhan model dan uji *wald* untuk menguji secara individu setiap estimasi parameter. Langkah terakhir dilakukan uji kebaikan model dengan menghitung nilai R^2 model.

III. HASIL DAN PEMBAHASAN

Hasil dari pemodelan TPT Perempuan dengan ketujuh variabel prediktor disajikan pada Tabel 1.

Tabel 1. Penaksiran Parameter Model Regresi Tersensor

Variabel	Estimasi	SE	Z
Konstan	-51,91	18,13	-2,86*
X_1	0,0751	0,0267	2,81*
X_2	0,0593	0,5279	0,11
X_3	0,7461	0,1677	4,45*
X_4	0,0212	0,0638	0,33
X_5	0,0186	0,1258	0,15

X_6	-1,0022	0,4848	-2,07*
X_7	-0,1499	0,1017	- 1,47
LR = 167,7799 $\chi^2_{(0,05;7)} = 14,067$ $R^2 = 65,60 \%$			

Hasil statistik uji LR pada model kedua adalah 167,78 yang lebih besar dari 14,067 sehingga dengan menggunakan taraf $\alpha = 5\%$, paling sedikit terdapat satu $\beta_k \neq 0$ Variabel yang berpengaruh secara signifikan pada taraf $\alpha = 5\%$ adalah persentase penduduk yang tinggal di perkotaan (X_1), seks rasio (X_3) dan tingkat pertumbuhan ekonomi (X_6). Terdapat beberapa variabel prediktor yang tidak signifikan. Oleh karena itu diperlukan tahapan seleksi variabel prediktor untuk mendapatkan model terbaik. Seleksi variabel dilakukan dengan menggunakan metode *Backward Elimination*. Hasil pengolahan disajikan pada Tabel 3.

Tabel 2. Penaksiran Parameter Model Regresi Tersensor Hasil *Backward Elimination*

Variabel	Estimasi	SE	Z
Konstan	-69,72	12,41	-5,62*
X_1	0,0903	0,0242	3,72*
X_3	0,7868	0,1388	5,67*
X_4	0,0519	0,0190	2,72*
X_5	0,0277	0,0102	2,74*
X_6	-0,8898	0,3776	-2,36*
LR = 140,813 $\chi^2_{(0,05;7)} = 11,075$ $R^2 = 75,87 \%$			

Berdasarkan hasil *Backward Elimination* pada Tabel dapat diketahui bahwa semua variabel prediktor dalam model sudah signifikan. Model regresi tersensor untuk TPT Perempuan adalah sebagai berikut

$$\hat{y} = -69,72 + 0,0903X_1 + 0,7868X_3 + 0,0519X_4 + 0,0277X_5 - 0,8898X_6$$

Hasil statistik uji LR adalah 140,813 yang lebih besar dari 11,0705 sehingga dengan menggunakan taraf $\alpha = 5\%$ paling sedikit terdapat satu $\beta_k \neq 0$ Variabel yang berpengaruh secara signifikan pada taraf $\alpha = 5\%$ adalah persentase penduduk yang tinggal di perkotaan (X_1), seks rasio (X_3), persentase penduduk yang yang berpendidikan di atas SLTP (X_4), persentase penduduk yang mampu membaca dan menulis (X_5) dan dan tingkat pertumbuhan ekonomi (X_6). Variabel yang bertanda positif yaitu persentase penduduk yang tinggal di daerah perkotaan, seks rasio, persentase penduduk yang berpendidikan di atas SLTP, persentase penduduk yang bisa membaca dan menulis, akan mengakibatkan meningkatnya TPT Perempuan sedangkan yang bertanda negatif yaitu pertumbuhan ekonomi akan menurunkan TPT Perempuan. Koefisien determinasi (R^2) yang dihasilkan adalah 75,87 persen. Hal ini berarti bahwa model regresi tersensor mampu menjelaskan variasi dari TPT perempuan sebesar 75,87 persen, sedangkan sisanya dijelaskan oleh variabel lain diluar model.

IV. SIMPULAN DAN SARAN



Faktor-faktor yang mempengaruhi TPT Perempuan di Pulau Jawa adalah persentase penduduk yang tinggal di daerah perkotaan, seks rasio, persentase penduduk yang berpendidikan di atas SLTP, persentase penduduk yang bisa membaca dan menulis, serta pertumbuhan ekonomi. Penelitian ini belum memperhatikan adanya keterkaitan antar wilayah dalam pemodelan. Perlu dilakukan pemodelan yang melibatkan aspek spasial untuk mengatasi adanya keterkaitan antar wilayah yang saling berdekatan.

DAFTAR PUSTAKA

- [1] BPS, "Booklet Agustus 2012", Jakarta: BPS RI, 2013.
- [2] W.H. Greene, "Econometrics Analysis, 6th edition," New Jersey: Prentice Hall, 2008.
- [3] S.H. Cox, and Y. Linn, "Annuity Lapse Rate Modeling: Tobit Or Not Tobit", [Journal of economic and social measurement](#), vol.38, pp: 432, 2006.
- [4] J. Tobin, "Estimation of Relationships for Limited Dependent Variables", *Econometrica*, vol.26(1); pp: 24-36, 1958.
- [5] I.Y. Suhardi, and R. Llewlyn, "Penggunaan Model Regresi Tobit untuk Menganalisa Faktor-Faktor yang Berpengaruh terhadap Kepuasan Konsumen untuk Jasa Pengangkutan Barang", *Jurnal Manajemen & Kewirausahaan*, vol.3(2), pp: 106-112, 2001.
- [6] M.D. Hadad, W. Santoso, and A. Alisjahbana, "Model dan Estimasi Permintaan dan Penawaran Kredit Konsumsi Rumah Tangga di Indonesia", *Jurnal Bank Indonesia*, pp: 1-25, 2004.
- [7] B. Karli, & A. Bilgic, "Factors Affecting Meat And Meat Products Consumption Quantities In Sanliurfa Province", *Mediterranean Agricultural Sciences*, vol.20(1), pp: 127-136, 2007
- [8] D. Purnomo, "Fenomena Migrasi Tenaga Kerja dan Perannya Bagi Pembangunan Daerah Asal: Studi Empiris di Kabupaten Wonogiri", *Jurnal Ekonomi Pembangunan*, vol.10(1), p: 84-102, 2009.