

Estimasi *Pseudo Poisson Maximum Likelihood* untuk Mengatasi Masalah dalam Model *Log-Linear* pada Kasus Kusta di Jawa Barat Tahun 2018

Ghina Fauziah*

Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Islam Bandung, Indonesia.

*ghinafauziah57@gmail.com

Abstract. Poisson regression is a statistical method used to analyze the relationship between the response variable and the predictor variable where the data from the response variable is in the form of count data and follows the Poisson distribution. Poisson regression is used to model rare or rare events, so that the response variable is very likely to have a value of zero. Poisson regression is a regression model whose response variable is non-negative. Usually, this model fits the linear regression applied to the log-transformed response variable. However, when the response variable data has a value of zero and is modeled using a log-linear model it will create a biased estimator as well as log-linear regression where heteroscedasticity occurs in the response variable will produce a biased parameter estimator. However, the pseudo poisson maximum likelihood (PPML) provides a natural way to deal with the problem. The purpose of this study is to determine the factors that influence leprosy cases in West Java in 2018 using PPML estimates. The results show that health facilities, healthy homes, and health insurance are factors that influence the number of leprosy cases in West Java in 2018. Using the AIC value, it shows that the use of PPML estimates produces better results than the log-linear model.

Keywords: Poisson Regression, Log-Linear, PPML, Leprosy.

Abstrak. Regresi poisson merupakan suatu metode statistika yang digunakan untuk menganalisa hubungan antara variabel respon dengan variabel prediktor dimana data dari variabel respon berbentuk data cacahan atau count data dan mengikuti distribusi poisson. Regresi poisson digunakan untuk memodelkan kejadian langka atau jarang terjadi, sehingga variabel respon sangat memungkinkan memiliki nilai nol. Regresi poisson merupakan model regresi yang variabel responnya bernilai non-negatif. Biasanya, model ini cocok dengan regresi linier yang diterapkan pada variabel respon yang ditransformasikan log. Namun, ketika data variabel respon memiliki nilai nol dan dimodelkan menggunakan model log-linear akan menciptakan suatu penaksir yang bias begitu juga regresi log-linear yang terjadi heteroskedastisitas pada variabel responnya akan menghasilkan suatu penaksir parameter yang bias. Namun, pseudo poisson maximum likelihood (PPML) menyediakan cara alami untuk menangani masalah tersebut. Tujuan penelitian ini yaitu untuk mengetahui faktor yang berpengaruh terhadap kasus kusta di Jawa Barat tahun 2018 menggunakan estimasi PPML. Hasil penelitian menunjukkan bahwa fasilitas kesehatan, rumah sehat, dan jaminan kesehatan menjadi faktor yang berpengaruh terhadap jumlah kasus kusta di Jawa Barat tahun 2018. Dengan menggunakan nilai AIC, menunjukkan bahwa penggunaan estimasi PPML menghasilkan hasil yang lebih baik dari pada model log-linear.

Kata Kunci: Regresi Poisson, Log-Linear, PPML, Kusta.

A. Pendahuluan

Regresi poisson merupakan suatu metode statistika yang digunakan untuk menganalisa hubungan antara variabel respon dengan variabel prediktor dimana data dari variabel respon berbentuk data cacahan atau count data, mengikuti distribusi poisson, dan bernilai non-negatif. Regresi poisson digunakan untuk memodelkan kejadian langka atau jarang terjadi, sehingga memungkinkan untuk variabel respon memiliki nilai nol. Biasanya model ini cocok dengan regresi linier yang diterapkan pada variabel respon yang ditransformasikan log. Karena data pada regresi poisson memungkinkan memiliki nilai nol dan ketika data tersebut dimodelkan menggunakan regresi log-linear maka akan menciptakan suatu penaksir yang bias. Begitu juga dengan regresi log-linear yang variabel responnya ditransformasikan-log dapat menciptakan bias ketika pada variabel sisa terdapat heteroskedastisitas. Sehingga karena adanya heteroskedastisitas maka parameter log-linear yang diperkirakan oleh Ordinary Least Square (OLS) sebagai elastisitas akan menyebabkan tidak konsisten. Namun, Pseudo Poisson Maximum Likelihood (PPML) menyediakan cara alami untuk menangani masalah suatu penaksir bias yang disebabkan karena adanya nilai nol pada variabel respon dan adanya heteroskedastisitas pada residual.

Kasus penyakit kusta merupakan kejadian langka atau jarang terjadi dengan data diskrit yang merupakan data cacahan sehingga dalam pemodelannya dapat menggunakan regresi poisson. Salah satu daerah tropis di Indonesia yang masih ditemukan penyakit kusta adalah Jawa Barat. Berdasarkan data dari Badan Pusat Statistika Jawa Barat bahwa telah ditemukan penderita penyakit kusta sebanyak 2.042 orang pada tahun 2018.

Untuk mengestimasi model regresi poisson dengan menyertakan satu buah variabel respon dan k buah variabel prediktor. Sehingga untuk melakukan estimasi poisson dilakukan implementasi Pseudo Poisson Maximum Likelihood (PPML). Selanjutnya untuk memperbaiki masalah dalam model log-linear diatas, maka digunakan estimasi PPML untuk mengatasi masalah dalam model log-linear pada kasus kusta di Jawa Barat tahun 2018. Berdasarkan uraian diatas, tujuan yang ingin dicapai dari penelitian ini adalah:

1. Untuk mengetahui pemodelan log-linear dan regresi poisson menggunakan estimasi pseudo poisson maximum likelihood pada kasus kusta di Jawa Barat pada tahun 2018.
2. Untuk Mengetahui faktor-faktor apa saja yang berpengaruh terhadap kasus kusta di Jawa Barat pada tahun 2018.

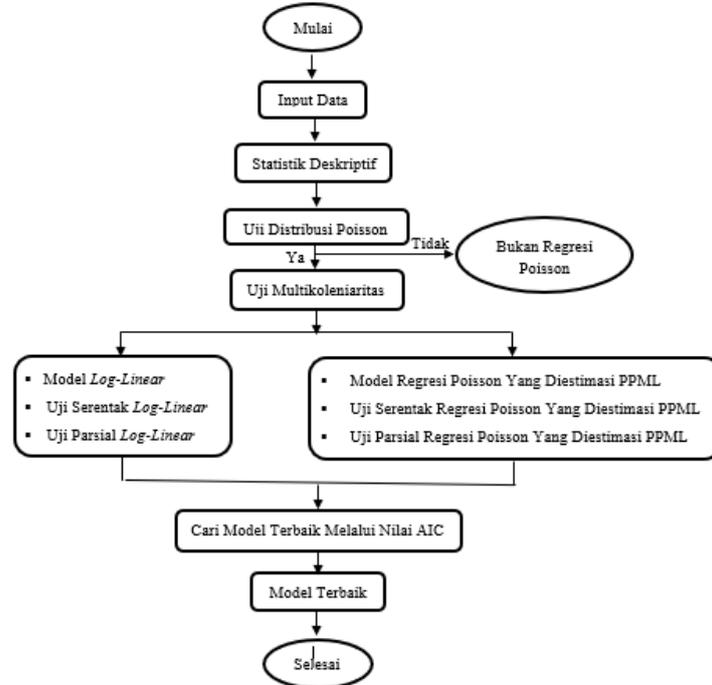
B. Metodologi

Data yang digunakan untuk penelitian ini adalah data sekunder yang didapatkan dari *website* Badan Pusat Statistika Jawa Barat. Periode yang diambil berupa data kasus penderita penyakit kusta di kabupaten/kota di Jawa Barat yang terdiri dari 2.042 kasus yang terjadi di 18 kabupaten dan 9 kota di Jawa Barat pada tahun 2018. Data penelitian ini terdiri dari satu buah variabel respon dan 5 buah variabel prediktor.

Pada penelitian ini akan dilakukan analisis statistik deskriptif dan estimasi *pseudo poisson maximum likelihood* untuk mengatasi masalah dalam pemodelan *log-linear* pada kasus kusta di Jawa Barat tahun 2018. Analisis statistik deskriptif digunakan untuk melihat gambaran jumlah kasus kusta di kabupaten/kota di Jawa Barat pada tahun 2018 dan pemodelan *log-linear* yang diestimasi menggunakan *pseudo poisson maximum likelihood* digunakan untuk memperoleh faktor-faktor yang berpengaruh terhadap jumlah kasus kusta di setiap kabupaten/kota di Jawa Barat pada tahun 2018. Dalam penelitian ini menggunakan *software* Stata 14 dengan tahapan sebagai berikut:

1. Melakukan pemeriksaan distribusi poisson pada variabel respon yaitu jumlah kasus kusta.
2. Mendeteksi multikolinieritas antar variabel prediktor menggunakan nilai VIF.
3. Menganalisis model *log-linear* dan regresi poisson yang diestimasi *pseudo poisson maximum likelihood* menggunakan langkah-langkah sebagai berikut: .
 - a. Melakukan penaksiran parameter model *log-linear* dan regresi poisson yang diestimasi menggunakan *pseudo poisson maximum likelihood*.

- b. Menguji signifikansi parameter secara serentak (simultan).
 - c. Menguji signifikansi parameter secara parsial (individu) untuk mengetahui variabel prediktor yang menjadi faktor-faktor berpengaruh terhadap jumlah kasus kusta di Jawa Barat tahun 2018.
4. Menentukan model terbaik dengan nilai AIC.



Gambar 1. Flow Chart Analisis Penelitian

C. Hasil dan Pembahasan

Untuk mengatasi masalah suatu penaksir bias yang disebabkan karena adanya nilai nol pada variabel respon dan adanya heteroskedastisitas pada residual, maka digunakan estimasi *pseudo poisson maximum likelihood* untuk mengatasi masalah pada model *log-linear* pada kasus kusta di Jawa Barat tahun 2018. Proses analisis yang pertama yaitu melakukan deskripsi data.

Tabel 1. Deskripsi Data

	Mean	Std. Dev.	Skewness	Kurtosis	Median	Observations
KUSTA	75.62963	105.5466	1.365318	3.54024	18	28

Berdasarkan tabel diatas, dapat dilihat bahwa data penyakit kusta sangat condong ke kanan karena nilai *mean* sebesar 75.63 lebih besar dibandingkan nilai median sebesar 18, berarti menunjukkan bahwa data tersebut memiliki lebih banyak nilai observasi berada dibawah nilai *mean*. Selain itu, nilai standar deviasi hampir mendekati 1,5 kali rata-rata. Sehingga dapat dikatakan layak untuk melakukan transformasi log pada variabel respon untuk mendapatkan model *log-linear* yang lebih baik dalam pemodelan penyakit kusta. Berikut adalah deskripsi data kusta dengan transformasi-log.

Tabel 2. Deskripsi Data Kusta dengan Transformasi-log

	Mean	Std. Dev.	Skewness	Kurtosis	Median	Observations
KUSTA	3.16118	1.781775	-.043486	1.90671	3.07423	28

Berdasarkan tabel diatas, dapat dilihat bahwa data dengan menggunakan transformasi log dapat mengurangi bahkan menghilangkan kemiringan data. Dapat dilihat bahwa nilai *mean* sebesar 3.16 mendekati nilai median sebesar 3.07, begitu juga dengan nilai statistik kemiringan mengalami penurunan dari 1.37 menjadi -0.04. keruncingannya yaitu sebesar 1.91, mendekati

nilai normal 2. Sehingga dapat disimpulkan bahwa dengan menggunakan transformasi log dapat mengurangi kemiringan pada data penyakit kusta.

Selanjutnya akan dilakukan pemeriksaan distribusi poisson sebagai salah satu syarat asumsi yang harus terpenuhi dalam regresi poisson. Berikut adalah hasil dari pemeriksaan distribusi poisson:

Tabel 3. Pemeriksaan Distribusi Poisson

Variabel	P-Value
Kusta	0.308

Pada tabel diatas menunjukkan nilai signifikansi dari pengujian distribusi poisson menggunakan statistik *Kolmogorov-Smirnov* sebesar 0,308, nilai tersebut lebih besar dari taraf nyata $\alpha = 0,05$ maka H_0 diterima. Sehingga dapat disimpulkan bahwa variabel respon untuk jumlah kasus kusta di provinsi Jawa Barat tahun 2018 berdistribusi poisson.

Setelah melakukan pemeriksaan distribusi poisson, selanjutnya terdapat asumsi yang harus dipenuhi dalam regresi poisson yaitu pengujian multikoleniaritas yang dilakukan untuk mengetahui apakah terdapat suatu variabel *independent* yang termasuk kombinasi linier dari variabel-variabel *independent* yang lain. Berikut adalah hasil pengujian dari multikoleniaritas:

Tabel 4. Pengujian Multikoleniaritas

Variabel	VIF
Kepadatan Penduduk	2.13
Fasilitas Kesehatan	1.83
Rumah Sehat	1.54
PHBS	1.51
Jaminan Kesehatan	1.27

Berdasarkan tabel diatas dapat dilihat bahwa nilai VIF untuk setiap variabel independent kurang dari 10. Hal tersebut menunjukkan bahwa tidak terdapat multikoleniaritas antar variabel independent. Sehingga kelima variabel independent diatas dapat digunakan dalam pemodelan regresi poisson dengan efek tetap berdimensi tinggi.

Penaksiran parameter model regresi *log-linear* menggunakan metode *Maximum Likelihood Estimation* (MLE). Berikut merupakan hasil penaksiran parameter model regresi *log-linear* akan disajikan pada tabel berikut:

Tabel 5. Penaksiran Parameter Model Regresi Log-Linear

lnKusta	Coef.	Std. Err.	P > t
Kepadatan Penduduk	-.2821716	06411999	0.665
Fasilitas Kesehatan	-.0000312	.0000756	0.684
Rumah Sehat	4.82e ⁻⁰⁶	1.68e ⁻⁰⁶	0.010
PHBS	-.1005116	07788031	0.899
Jaminan Kesehatan	-5.98e ⁻⁰⁷	3.84e ⁻⁰⁷	0.135
Konstanta	2.93723	2.014646	0.160

Berdasarkan hasil penaksiran pada tabel diatas, diperoleh model regresi *log-linear* yaitu:
 $\log Y_i = 2.94 - 0.28x_1 - 0.00003x_2 + 0.000004x_3 - 0.10x_4 - 0.000006x_5$

Selanjutnya akan dilakukan pengujian parameter pada model regresi *log-linear* secara serentak menggunakan statistik uji *G*. Berikut merupakan hasil perhitungan statistik uji *G*, nilai kritis, nilai signifikansi, dan keputusan uji:

Tabel 6. Statistik Uji

Uji G	$\chi^2_{(0,05;5)}$	P-value	Keputusan
75.304	11.0705	0.000	H_0 ditolak

Berdasarkan nilai statistik uji *G*, *chi-kuadrat*, dan nilai signifikansi pada tabel diatas,

didapatkan keputusan bahwa H_0 ditolak pada taraf nyata $\alpha = 5\%$, karena nilai statistik uji $G = 75.304 > \chi^2_{(0,05;5)} = 11.0705$ atau bisa juga dilihat dari nilai $p\text{-value} = 0.000 < \alpha = 0,05$. Sehingga dapat disimpulkan bahwa tingkat kepadatan penduduk, fasilitas kesehatan, rumah sehat, rumah tangga berperilaku hidup bersih dan sehat, dan jumlah pengguna jaminan kesehatan berpengaruh secara serentak terhadap jumlah kasus kusta di Jawa Barat tahun 2018.

Selanjutnya akan dilakukan pengujian parameter pada model regresi *log-linear* secara parsial menggunakan statistik uji t yang dapat dilihat pada tabel diatas. Dan dapat disimpulkan bahwa hanya variabel rumah sehat yang berpengaruh secara individu terhadap jumlah kasus kusta di kabupaten/kota di Jawa Barat tahun 2018. Hal ini ditunjukkan dengan nilai mutlak statistik t dari variabel rumah sehat yang memiliki nilai lebih besar dari nilai $t_{\text{tabel}} = 2,074$ atau dapat dilihat nilai $p\text{-value} = 0,010$ dari variabel rumah sehat lebih kecil dari nilai $\alpha = 0,05$, sedangkan variabel kepadatan penduduk, fasilitas kesehatan, rumah tangga yang memiliki perilaku hidup bersih dan sehat, dan pengguna jaminan kesehatan tidak berpengaruh terhadap jumlah kasus kusta karena memiliki nilai mutlak statistik t kurang dari nilai $t_{\text{tabel}} = 2,074$ atau nilai $p\text{-value}$ dari ke empat variabel tersebut lebih besar dari nilai $\alpha = 0,05$.

Penaksiran parameter model regresi *poisson* menggunakan metode oleh *Pseudo Poisson Maximum Likelihood Estimation* (PPML). Berikut merupakan hasil penaksiran parameter model regresi *poisson* yang diestimasi oleh PPML :

Tabel 7. Hasil Penaksiran Parameter Model Regresi Poisson

Kusta	Coef.	Std. Err.	P > z
Kepadatan Penduduk	.575754	.4343109	0.185
Fasilitas Kesehatan	-.0000662	.0000306	0.030
Rumah Sehat	5.02e ⁻⁰⁶	1.25e ⁻⁰⁶	0.000
PHBS	-.5570536	.5076484	0.273
Jaminan Kesehatan	-7.44e ⁻⁰⁷	2.44e ⁻⁰⁷	0.002
Konstanta	2.493253	1.25507	0.047

Berdasarkan hasil penaksiran pada tabel diatas, diperoleh model regresi poisson dengan estimasi PPML yaitu:

$$\hat{\mu}_i = \exp(2.4933 + 0.57658 - 0.00007X_{i2} + 0.000005X_{i3} - 0.557X_{i4} - 0.0000007X_{i5})$$

Selanjutnya akan dilakukan pengujian parameter pada model regresi poisson dengan estimasi PPML secara serentak menggunakan statistik uji G . Berikut merupakan hasil perhitungan statistik uji G , nilai kritis, nilai signifikansi, dan keputusan uji:

Tabel 8. Statistik Uji

Uji G	$\chi^2_{(0,05;5)}$	$P\text{-value}$	Keputusan
43.02	11.0705	0.000	H_0 ditolak

Berdasarkan nilai statistik uji G , *chi-kuadrat*, dan nilai signifikansi pada tabel diatas, didapatkan keputusan bahwa H_0 ditolak pada taraf nyata $\alpha = 5\%$, karena nilai statistik uji $G = 43.02 > \chi^2_{(0,05;5)} = 11.0705$ atau bisa juga dilihat dari nilai $p\text{-value} = 0.000 < \alpha = 0,05$. Sehingga dapat disimpulkan bahwa tingkat kepadatan penduduk, fasilitas kesehatan, rumah sehat, rumah tangga berperilaku hidup bersih dan sehat, dan jumlah pengguna jaminan kesehatan berpengaruh secara serentak terhadap jumlah kasus kusta di Jawa Barat tahun 2018.

Selanjutnya akan dilakukan pengujian parameter pada model regresi poisson dengan estimasi PPML secara parsial menggunakan statistik uji Z yang dapat dilihat pada tabel diatas. Dan dapat disimpulkan bahwa variabel fasilitas kesehatan, jumlah rumah sehat, dan penduduk pengguna jaminan kesehatan masing-masing berpengaruh secara individu terhadap jumlah kasus kusta di kabupaten/kota di Jawa Barat tahun 2018. Hal ini ditunjukkan dengan nilai mutlak statistik Z dari masing-masing ketiga variabel tersebut memiliki nilai lebih besar dari nilai $Z_{\text{tabel}} = 1,96$ atau dapat dilihat nilai $p\text{-value}$ dari masing-masing ketiga variabel tersebut lebih

kecil dari nilai $\alpha = 0,05$, sedangkan tingkat kepadatan penduduk dan rumah tangga yang memiliki perilaku hidup bersih dan sehat tidak berpengaruh terhadap jumlah kasus kusta karena memiliki nilai mutlak statistik Z kurang dari nilai $Z_{\text{tabel}} = 1,96$ atau nilai p -value lebih besar dari nilai $\alpha = 0,05$.

Setelah dilakukan estimasi dan pengujian parameter, selanjutnya untuk mengetahui perbandingan mana model regresi yang lebih baik antara model regresi *log-linear* dan regresi poisson dengan estimasi PPML dapat dilihat dari masing-masing nilai AIC kedua regresi tersebut. Berikut merupakan hasil nilai AIC:

Tabel 9. Hasil Nilai AIC

Kriteria	Regresi <i>Log-Linear</i>	Regresi Poisson Dengan PPML
AIC	1701.379	100.649

Berdasarkan nilai AIC yang terdapat pada tabel diatas, dapat disimpulkan bahwa regresi poisson yang diestimasi PPML memiliki nilai AIC lebih kecil dibandingkan dengan regresi *log-linear*. Sehingga pada penelitian ini menunjukkan bahwa regresi poisson yang diestimasi PPML lebih baik dari pada model *log-linear*.

D. Kesimpulan

Berdasarkan pembahasan dalam penelitian ini, peneliti menyimpulkan beberapa hasil penelitian sebagai berikut:

1. Model regresi *log-linear* yaitu dengan persamaan sebagai berikut:

$$\log Y_i = 2.94 - 0.28x_1 - 0.00003x_2 + 0.000004x_3 - 0.10x_4 - 0.0000006x_5$$
 Model regresi poisson dengan estimasi PPML yaitu dengan persamaan sebagai berikut:

$$\hat{\mu}_i = \exp(2.493 + 0.577 - 0.00007X_{i2} + 0.000005X_{i3} - 0.557X_{i4} - 0.0000007X_{i5})$$
2. Faktor yang mempengaruhi penyakit kusta di seluruh kabupaten/kota yang ada di Jawa Barat tahun 2018 menggunakan regresi *log-linear* adalah variabel rumah sehat.

Acknowledge

Peneliti mengucapkan terima kasih kepada Dinas Kesehatan dan Badan Pusat Statistika Jawa Barat yang telah memberikan data untuk membantu terlaksananya penelitian ini. Tak lupa juga penulis ucapkan terima kasih kepada para dosen Statistika Unisba, khususnya kepada ibu Siti Sunendiari, Dra., M.Si. yang selalu memberikan arahan dan saran kepada penulis hingga akhirnya penelitian ini dapat terselesaikan dengan baik.

Daftar Pustaka

- [1] Cameron, A. d. (2013). *Regression Analysis of Count Data*. New York: Cambridge University Press.
- [2] Fundacao Getulio Vargas, S. P. (2019). Estimating Poisson Pseudo-Maximum-Likelihood Rather Than Log-Linear Model Of A Log-Transformed Dependent Variable. *Journal of RAUSP Management*, 3-4.
- [3] GM, W. (1998). The Logged Dependent Variable, Heteroscedasticity, And The Retransformation Problem. *Journal Of Health Economics*, 2-4.
- [4] Kharisma, M. S. (2017). Regresi Poisson Dalam Memodelkan Jumlah Penderita Kusta Di Provinsi Bali. *Jurnal Matematika*, 2-3.
- [5] Motta, V. (2019). Estimating Poisson Pseudo Maximum Likelihood Rather Than Log-Linear Model Of A Log Transformed Dependent Variable. *The Stata Journal*, 1-8.