

## PEMODELAN ZERO INFLATED NEGATIVE BINOMIAL (ZINB) PADA KASUS JUMLAH BEPERGIAN PENDUDUK PROVINSI SULAWESI TENGAH

Iut Tri Utami<sup>1</sup>

<sup>1</sup>Departemen Statistika, Universitas Diponegoro  
Jl. Prof. Soedarto, SH, Kampus UNDIP Tembalang  
<sup>1</sup>triutami.iut@gmail.com

### ABSTRACT

Traveling is something that has been done by every resident, especially the people of Central Sulawesi province. The people of Central Sulawesi province who travel have a certain purpose either because of their own desires or following others. The number of people traveling from Central Sulawesi Province can be analyzed using the zero inflated negative binomial regression method (ZINB). ZINB regression is a method used to model calculated or discrete data with many zero values on the response variable (zero inflation) and overdispersion occurs. The result shows that the factors affecting the number of people traveling population from the Central Sulawesi Province are age ( $X_1$ ), the people who have jobs in mining and quarrying ( $X_{32}$ ), the people who have jobs in electricity and gas ( $X_{34}$ ). The zero inflated negative binomial (ZINB) regression model is better at modeling cases of the number of people traveling in Central Sulawesi Province in 2016 compared to the Poisson regression model because it has the smallest AIC value.

**Keywords** : Amount Traveled, Overdispersion, Zero Inflated Negative Binomial Regression

### ABSTRAK

Bepergian merupakan hal yang pernah dilakukan oleh setiap penduduk khususnya penduduk Provinsi Sulawesi Tengah. Penduduk Provinsi Sulawesi Tengah yang melakukan bepergian memiliki tujuan tertentu baik karena keinginan sendiri atau mengikuti orang lain. Jumlah penduduk provinsi Sulawesi Tengah yang bepergian dapat dianalisis menggunakan metode regresi *zero inflated negative binomial* (ZINB). Hasil penelitian menunjukkan bahwa faktor-faktor mempengaruhi jumlah penduduk yang berpergian di Kota Palu adalah umur ( $X_1$ ), penduduk yang memiliki pekerjaan di pertambangan dan penggalian ( $X_{32}$ ) dan penduduk yang memiliki pekerjaan di bidang listrik dan gas ( $X_{34}$ ). Model regresi *zero inflated negative binomial* (ZINB) lebih baik dalam memodelkan kasus jumlah bepergian penduduk Provinsi Sulawesi Tengah tahun 2016 dibandingkan dengan model regresi poisson karena memiliki nilai AIC terkecil.

**Kata kunci** : Jumlah Bepergian, Overdispersi, Zero Inflated Negative Binomial Regression

### I. PENDAHULUAN

Pertumbuhan ekonomi di Indonesia berkembang sangat pesat. Tren konsumsi masyarakat yang semakin meningkat mendorong pertumbuhan ekonomi yang berkelanjutan. Pertumbuhan ekonomi yang tinggi disertai peningkatan daya beli masyarakat mendorong masyarakat melakukan perjalanan wisata. Perjalanan Wisata (Perjalanan/Bepergian) adalah bepergian/perjalanan yang dilakukan oleh penduduk Indonesia dari tempat tinggal biasanya ke tempat di luar lingkungan sehari-hari di wilayah geografis Indonesia secara sukarela selama kurang dari 6 bulan dan bukan bertujuan untuk bekerja (memperoleh upah/gaji dari penduduk ditempat yang dituju) atau sekolah secara rutin. Tempat diluar lingkungan sehari-hari dimaksud

meliputi obyek wisata komersial, menginap di akomodasi komersial, dan jarak perjalanan 100 km pp atau lebih (Statistik Profil Wisatawan Nusantara, 2016).

Masyarakat di Provinsi Sulawesi Tengah sering melakukan bepergian keluar kota. Tujuan mereka dalam bepergian atau melakukan perjalanan diantaranya adalah berlibur, rekreasi, olahraga, bisnis, mengunjungi teman dan keluarga, menghadiri pertemuan, berziarah/keagamaan, dan kunjungan dengan alasan kesehatan. Dalam buku Statistik Profil Wisatawan Nusantara (2014), sebagian besar masyarakat Provinsi Sulawesi Tengah melakukan perjalanan dengan tujuan untuk mengunjungi teman atau keluarga. Proporsi masyarakat Provinsi Sulawesi Tengah yang mengunjungi teman atau keluarga mencapai sekitar 59,18%. Sedangkan proporsi masyarakat provinsi Sulawesi Tengah dengan tujuan utama bepergian paling sedikit adalah olahraga/kesenian yaitu sekitar 0,31%.

Faktor-faktor yang mempengaruhi jumlah bepergian penduduk dapat di analisis dengan menggunakan regresi *zero inflated negative binomial* (ZINB). Data jumlah bepergian penduduk memiliki nilai nol yang banyak sehingga terjadi overdispersi yaitu nilai varians lebih besar dari nilai mean pada data respon. Salah satu metode yang dapat mengatasi overdispersi pada data diskrit adalah *zero inflated negative binomial* (ZINB). Model regresi ZINB memiliki dua keadaan, yaitu keadaan pertama disebut *zero state* terjadi dengan probabilitas  $p_i$  dan menghasilkan hanya observasi bernilai nol. Sementara keadaan kedua disebut *negative binomial state* terjadi dengan probabilitas  $(1 - p_i)$  dan berdistribusi binomial negatif dengan rata-rata  $\mu$ , dengan  $0 < p_i < 1$ .

Penelitian ini membahas tentang analisis faktor-faktor yang mempengaruhi jumlah bepergian penduduk di Provinsi Sulawesi Tengah menggunakan model *zero inflated negative binomial* (ZINB). Variabel prediktor yang digunakan adalah umur, jenis kelamin, pekerjaan, dan jumlah anggota rumah tangga. Hasil yang diperoleh dari penelitian ini diharapkan dapat menjadi evaluasi bagi pemerintah dalam menyusun kebijakan pariwisata yang lebih efektif.

## II. METODE PENELITIAN

### 2.1. Sumber Data

Data yang digunakan pada penelitian ini yaitu data penduduk yang bepergian dari Provinsi Sulawesi Tengah pada tahun 2016 yaitu sebanyak 234 orang. Data diambil dari publikasi Statistik Profil Wisatawan Nusantara . Dalam penelitian ini terdapat dua variabel penelitian yang digunakan yaitu variabel respon (Y) dan empat variabel prediktor (X), Berikut ini merupakan variabel-variabel yang digunakan sebagai variabel respon dan variabel-variabel prediktor yang digunakan dalam penelitian.

Tabel 1 : Variabel penelitian dan definisi operasional

Simbol	Variabel	Definisi Operasional
Y	Bepergian Penduduk	Data diskrit
X <sub>1</sub>	Umur	Data yang digunakan adalah data rasio
X <sub>2</sub>	Jenis Kelamin	1. Laki-laki 2. Perempuan
X <sub>3</sub>	Pekerjaan	0: Lainnya

		1: Pertanian
		2: Pertambangan dan penggalian
		3: Industri pengolahan
		4: Listrik dan gas
		5. Perdagangan, hotel, dan rumah makan
$X_4$	Jumlah Anggota Rumah Tangga	Data yang digunakan adalah data rasio

## 2.2. Metode Analisis Data

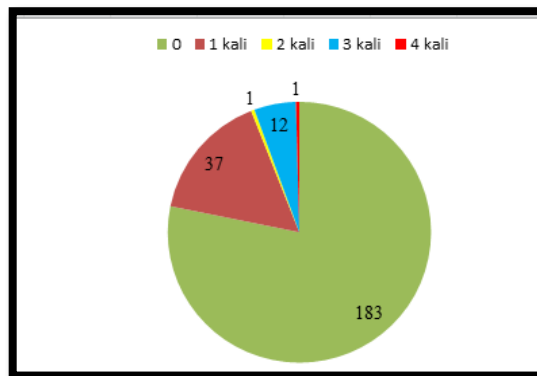
Analisis dalam penelitian ini menggunakan regresi *zero inflated negative binomial* (ZINB), dengan menggunakan *software* R 3.5.2. Berikut ini adalah tahapan-tahapan yang digunakan dalam analisis:

1. Melakukan pengambilan data.
2. Memeriksa variabel respon apakah berdistribusi poisson atau tidak dengan menggunakan uji Kolmogorov-Smirnov.
3. Melakukan uji multikolinearitas pada variabel prediktor dengan melihat nilai *variance inflation factors* (VIF).
4. Menentukan model regresi poisson.
5. Menguji kesesuaian model regresi Poisson menggunakan statistik uji D atau *Deviance*.
6. Memeriksa overdispersi dengan melihat nilai *Deviance* dan nilai *Pearson Chi square*.
7. Menentukan model Regresi Zero Inflated Negatif Binomial (ZINB)
8. Melakukan pengujian parameter model menggunakan uji G secara simultan dan uji wald secara parsial.
9. Menguji kebaikan model dengan menentukan nilai AIC
10. Membuat kesimpulan.

## III. HASIL DAN PEMBAHASAN

### 3.1. Analisis Deskriptif

Analisis deskriptif merupakan analisis yang digunakan untuk melihat karakteristik profil dari variabel respon dan variabel prediktor. Adapun variabel respon yang digunakan dalam penelitian ini adalah jumlah bepergian penduduk Provinsi Sulawesi Tengah (Y). Variabel prediktor yang digunakan adalah umur ( $X_1$ ), jenis kelamin ( $X_2$ ), pekerjaan ( $X_3$ ), jumlah anggota rumah tangga ( $X_4$ ). Statistik deskriptif untuk variabel jumlah bepergian penduduk (Y).



Gambar 1 : Diagram Jumlah Penduduk Provinsi Sulawesi Tengah yang Bepergian.

Pada Gambar 1 menunjukkan bahwa penduduk Provinsi Sulawesi Tengah tidak melakukan bepergian yaitu sebanyak 183 orang, yang melakukan bepergian sebanyak 1 kali terdapat 37 orang. penduduk yang melakukan bepergian sebanyak 2 kali terdapat 1 orang, bepergian sebanyak 3 kali terdapat 12 orang dan yang bepergian sebanyak 4 kali terdapat 1 orang.

### 3.2. Uji Distribusi Variabel Respon

Uji distribusi variabel respon dilakukan untuk mengetahui apakah variabel respon mengikuti sebaran poisson atau tidak. Pemeriksaan sebaran variabel respon dilakukan menggunakan uji Kolmogorov-Smirnov (Nurmaleni dan Rahayu, 2017). Pemeriksaan distribusi variabel respon dilakukan pada 234 pengamatan, Hipotesis yang digunakan adalah :

$H_0$  : Variabel respon berdistribusi poisson

$H_1$  : Variabel respon tidak berdistribusi poisson

Tabel 2 : Uji Distribusi Poisson Variabel Respon

<i>Asymp.sig. (2 - tailed)</i>	Taraf signifikansi ( $\alpha$ )
0,221	0,05

Berdasarkan Tabel 2 dapat diketahui bahwa nilai *asymp.sig. (2 - tailed)* sebesar 0,221 lebih besar dari  $\alpha (0,05)$ , yang berarti bahwa  $H_0$  diterima artinya variabel respon berdistribusi poisson.

### 3.3. Uji Multikolinearitas pada Variabel Prediktor

Setelah melakukan uji distribusi poisson pada variabel respon maka dilakukan uji multikolinearitas pada variabel prediktor menggunakan nilai VIF (Hocking, 1996). Uji ini digunakan untuk melihat variabel mana yang mengalami multikolinearitas, diperoleh nilai VIF yang di tunjukkan pada Tabel 3.

Tabel 3 : Nilai VIF

Variabel	VIF
$X_1$	1,235
$X_2$	1,041

$X_3$	1,207
$X_4$	1,039

Berdasarkan Tabel 3 dapat diketahui bahwa nilai VIF dari variabel prediktor  $X_1$  sampai  $X_4$  menunjukkan bahwa tidak terjadi multikolinearitas karena nilai VIF tersebut kurang dari 10 (Fotheringham and Oshan, 2016).

### 3.4. Pemodelan Regresi Poisson

Model regresi poisson digunakan untuk memodelkan data dengan variabel respon yang berdistribusi poisson. Persamaan model regresi poisson yang terbentuk dalam penelitian ini adalah sebagai berikut :

$$y_i = \exp(-1,715 - 0,524X_{22} - 0,733X_{34} + 1,358X_{35} + 0,182X_4) \quad (1)$$

### 3.5. Uji Kesesuaian Model Regresi Poisson

Uji kesesuaian model digunakan untuk mengetahui bahwa model sesuai atau tidak terdapat perbedaan yang signifikan antara hasil pengamatan dengan kemungkinan hasil prediksi model (Kleinbaum *et al.*, 1978).

Hipotesis yang digunakan dalam penelitian ini adalah sebagai berikut:

$H_0$  : Model sesuai (tidak terdapat perbedaan yang signifikan antara hasil pengamatan dengan kemungkinan hasil prediksi model).

$H_1$  : Model tidak sesuai (terdapat perbedaan yang signifikan antara hasil pengamatan dengan kemungkinan hasil prediksi model).

Setelah dilakukan pengolahan data menggunakan *software* R, diperoleh nilai *deviance* dan *pearson chi-square* seperti tampak pada Tabel 4.

Tabel 4 : Uji Kesesuaian Model

<i>Deviance</i>	<i>Chi-square</i>
256,318	265,301

Pada Tabel 4 diketahui bahwa nilai *deviance* = 256,318 < *chi-square* (265,301) maka  $H_0$  diterima, artinya model sesuai atau tidak terdapat perbedaan yang signifikan antara hasil pengamatan dengan hasil prediksi model.

### 3.6. Uji Overdispersion

*Overdispersion* adalah kondisi dimana nilai varians dari variabel respon (Y) lebih besar dari nilai rata-rata (Hilbe, 2011). *Overdispersion* dapat dideteksi dengan menggunakan nilai *deviance* dan nilai *pearson chi-square* dibagi dengan derajat bebas bernilai lebih dari 1 (Agresti, 2002). Hasil uji *overdispersion* dengan menggunakan nilai *deviance* dan nilai *pearson chi-square* ditunjukkan pada Tabel 5.

Tabel 5. Hasil Uji Overdispersion Regresi Poisson

Kriteria	Nilai	Db	Db/ Nilai
Deviance	256,318	229	1,119
Pearson Chi-Square	369,560	229	1,614

Pada Tabel 5 menunjukkan bahwa nilai *deviance* dan nilai *pearson chi-square* dibagi dengan derajat bebas masing-masing adalah 1,119 dan 1,614. Hasil tersebut

menunjukkan nilai pembagian lebih dari satu sehingga variabel respon mengalami *overdispersion*.

### 3.7. Model Zero Inflated Negative Binomial (ZINB)

Model *zero inflated negative binomial* (ZINB) terbentuk dari dua keadaan yaitu *zero state* dan *negative binomial state* (Garay *et al.*, 2011). Persamaan model ZINB yang terbentuk adalah sebagai berikut :

- a. Model *negative binomial state* untuk  $\hat{\mu}$

Model *negative binomial state* untuk penduga rata-rata ( $\hat{\mu}$ ) adalah sebagai berikut:

$$\ln(\hat{\mu}) = (-1,104 + 0,064 X_1 - 2,344 X_{32} - 2,220 X_{34})$$

(2)

- b. Model zero inflation untuk  $\hat{\pi}$

Model zero inflation yang didapatkan untuk penduga phi ( $\hat{\pi}$ ) adalah sebagai berikut :

$$\text{logit}(\hat{\pi}) = (-0,387 + 0,121 X_1)$$

(3)

### 3.8. Pengujian Secara Simultan dan parsial Regresi Poisson

Pengujian signifikansi parameter secara simultan dapat dilihat dari statistik uji G (Hosmer dan Lemeshow, 2000). Uji signifikansi parameter secara parsial berfungsi untuk mengetahui variabel yang berpengaruh secara signifikan terhadap variabel respon secara individu. Hasil pengujian signifikansi secara simultan dan parsial terlihat pada Tabel 6.

Tabel 6 : Uji Parsial pada Variabel Prediktor Regresi Poisson

Parameter	Estimate	Std.Error	z value	Pr(> z )
$\hat{\beta}_0$	1,7155509	0.5331310	3.218	0.00129
$\hat{\beta}_1$	-0.0001542	0.0075628	0.020	0.98373
$\hat{\beta}_{22}$	-0.5244650	0.2480237	-2.115	0.03447
$\hat{\beta}_{31}$	0.7759026	0.5022737	1.545	0.12240
$\hat{\beta}_{32}$	1.1119720	0.6711806	1.657	0.09757
$\hat{\beta}_{33}$	0.6273319	1.0325207	0.608	0.54347
$\hat{\beta}_{34}$	-0.7333339	0.3412932	-2.149	0.03166
$\hat{\beta}_{36}$	1.3588825	0.5441146	2.497	0.01251
$\hat{\beta}_4$	0.1823721	0.0913974	1.995	0.04600
Statistik uji G =372.1				

Berdasarkan Tabel 6 dapat diketahui bahwa nilai statistik uji G = 372.1 dibandingkan dengan nilai *chi-square* (265,301). Nilai statistik uji G > *chi-square*, hal ini menunjukkan bahwa ada pengaruh variabel prediktor secara simultan terhadap variabel respon atau  $H_0$  ditolak.

Adapaun estimasi parameter regresi secara parsial terdapat empat variabel prediktor yang memiliki nilai *p-value* <  $\alpha$  yaitu 0,05. Hal ini menunjukkan bahwa

variabel prediktor yang berpengaruh secara parsial pada model regresi poisson adalah jenis kelamin perempuan, jenis pekerjaan bidang listrik dan gas, perdagangan, hotel, rumah makan, serta jumlah anggota rumah tangga.

### 3.9. Pengujian Secara Simultan dan parsial Regresi ZINB

Parameter yang diuji pada pengujian serentak ini mencakup seluruh parameter  $\beta$  dan secara bersama-sama. Hasil pengujian secara simultan dan parsial ditunjukkan pada Tabel 7.

Tabel 7 : Uji Parsial pada Variabel Prediktor Regresi ZINB

Parameter	Estimate	Std.Error	z value	Pr(>  z )
$\hat{\beta}_0$	-1.10484	0.87510	-1.263	0.2068
$\hat{\beta}_1$	0.06446	0.01483	4.347	1.38e-05
$\hat{\beta}_{22}$	-0.71392	0.37965	-1.880	0.0600
$\hat{\beta}_{31}$	-0.81053	0.78639	-1.031	0.3027
$\hat{\beta}_{32}$	-2.34437	0.93784	-2.500	0.0124
$\hat{\beta}_{33}$	-1.25290	1.20740	-1.038	0.2994
$\hat{\beta}_{34}$	-2.22067	0.52866	-4.201	2.66e-05
$\hat{\beta}_{36}$	-0.93239	0.74990	-1.243	0.2137
$\hat{\beta}_4$	-0.02601	0.14031	-0.185	0.8529
$\hat{\gamma}_0$	-0.38721	2.03333	-0.190	0.8490
$\hat{\gamma}_1$	0.12121	0.04984	2.432	0.0150
$\hat{\gamma}_{22}$	-0.51254	0.89005	-0.576	0.5647
$\hat{\gamma}_{31}$	-2.76756	1.58443	-1.747	0.0807
$\hat{\gamma}_{32}$	-18.5215	7601.950	-0.002	0.9981
$\hat{\gamma}_{33}$	-15.9531	1467.341	-0.011	0.9913
$\hat{\gamma}_{34}$	-2.6787	1.50865	-1.776	0.0758
$\hat{\gamma}_{36}$	-22.0058	7768.263	-0.003	0.9977
$\hat{\gamma}_4$	-0.49486	0.36024	-1.374	0.1695
Statistik uji G = 311,2				

Berdasarkan Tabel 7 dapat diketahui bahwa pada estimasi parameter regresi secara simultan dengan melihat nilai statistik uji G dibandingkan dengan nilai *chi-square* (30,143). Nilai statistik uji G > *chi-square*, hal ini menunjukkan bahwa ada pengaruh variabel prediktor secara simultan terhadap variabel respon atau  $H_0$  ditolak.

Estimasi parameter secara parsial untuk keadaan *negative binomial state* terdapat dua variabel prediktor dan pada estimasi parameter untuk *zero state* terdapat satu variabel prediktor yang memiliki nilai  $p\text{-value} < \alpha$  (0,05). Hal ini menunjukkan bahwa variabel prediktor yang berpengaruh secara parsial pada model *negative binomial state* adalah umur ( $X_1$ ) dan pekerjaan ( $X_3$ ) di bidang listrik dan gas. Adapun variabel prediktor yang berpengaruh secara signifikan pada model *zero state* adalah umur ( $X_1$ ).

### 3.10. Ukuran Keباikan Model

Penentuan kebaikan model dapat dilihat dari nilai *akaike information criterion* (AIC) (Akaike, 1974). Pada penelitian ini dilakukan perbandingan antara nilai AIC model regresi poisson dan nilai AIC model *zero inflated negative binomial* (ZINB) untuk melihat model terbaik. Berikut ini adalah nilai AIC model regresi poisson dan model regresi ZINB :

Tabel 8 : Nilai AIC Model Regresi

Model Regresi	Nilai AIC
Poisson	374,31
ZINB	349,13

Berdasarkan Tabel 8 menunjukkan bahwa nilai AIC untuk model regresi ZINB lebih kecil dibandingkan dengan nilai AIC untuk model regresi poisson, yaitu  $349,13 < 374,31$ . Hal ini dapat disimpulkan bahwa model regresi *zero inflated negative binomial* (ZINB) lebih baik untuk memodelkan kasus jumlah bepergian penduduk Provinsi Sulawesi Tengah tahun 2016 dibandingkan dengan model regresi poisson.

## IV. KESIMPULAN

Berdasarkan hasil dan pembahasan yang telah dilakukan sebelumnya maka diperoleh beberapa kesimpulan sebagai berikut:

1. Model regresi *zero inflated negative binomial* (ZINB) lebih baik dalam memodelkan kasus jumlah bepergian penduduk Provinsi Sulawesi Tengah tahun 2016 dibandingkan dengan model regresi poisson karena memiliki nilai AIC terkecil.
2. Faktor yang berpengaruh terhadap jumlah bepergian penduduk Provinsi Sulawesi Tengah adalah variabel umur ( $X_1$ ), penduduk yang memiliki pekerjaan di bidang pertambangan dan penggalian ( $X_{32}$ ) dan penduduk yang memiliki pekerjaan di bidang listrik dan gas ( $X_{34}$ ).

## DAFTAR PUSTAKA

- [1] Agresti, A., *Categorical Data Analysis, 2th ed*, New York, 2002, John Wiley & Sons, Inc.



- [2] Akaike H., *A New Look At The Statistical Model Identification*. IEEE Trans Autom Control, 19, 1974, 716–723.
- [3] Badan Pusat Statistik dan Kementrian Pariwisata, *Statistik Profil Wisatawan Nusantara 2016*, Jakarta, 2016, BPS dan Kementrian Pariwisata.
- [4] Fotheringham, A.S., Oshan, T.M., Geographically Weighted Regression And Multicollinearity: Dispelling The Myth, *Journal of Geographical Systems*, 18, 2016, 303–329.
- [5] Garay, A. M., Hashimoto, E. M., Ortega, E. M. M., & Lachos, V. H., *On Estimation And Influence Diagnostics for Zero Inflated Negative Binomial Regression Model*, Computational Statistics and Analysis, 55, 2011, 1304-1318.
- [6] Hilbe, J. M., *Negative Binomial Regression*, New York, 2011, Cambridge University Press.
- [7] Hocking, R., *Methods and Application of Linier Models*, New York, 1996, John Wiley and Sons.
- [8] Hosmer, D. W., & Lemeshow, S., *Applied Logistic Regression*, New York, 2000, John Wiley and Sons.
- [9] Kleinbaum, D.G., L.L. Kupeer., A. Nizam., dan K.E. Muller., *Applied Regression Analysis and Other Multivariabel*, Fourth Editon, Canada, 1978, Thomson Brooks Cole.
- [10] Nurmaleni dan Rahayu, N.F., *Faktor-Faktor Yang Mempengaruhi Angka Penderita Gizi Buruk Pada Balita Di Papua Tahun 2015 Dengan Metode Regresi Zero Inflated Poisson (ZIP)*, Jurnal Logika, 7(1), 2017, 1-14.