

## MODEL REGRESI *HURDLE NEGATIVE BINOMIAL* (HNB) UNTUK PEMODELAN KONSUMSI ROKOK DI PROVINSI SULAWESI TENGAH

Drajat Indra Purnama<sup>1</sup>

<sup>1</sup>Badan Pusat Statistik (BPS) Kabupaten Parigi Moutong

<sup>1</sup>drajatindrapurnama@bps.go.id

### ABSTRACT

The average expenditure on cigarettes per capita in Sulawesi Tengah Province has increased in 2020. There are several factors that can affect a person's cigarette consumption including gender, age, education and health. To model cigarette consumption with several influencing factors can be use the poisson regression model or the Zero Inflated Poisson (ZIP) model. However, the two regression models cannot solve the excess zero and overdispersion problems so use the Hurdle Negative Binomial (HNB) regression model. The results of the analysis of cigarette consumption data in Central Sulawesi Province using the HNB model provide the best modeling results compared to the poisson regression model and the ZIP model because it has the smallest Akaike's Information Criterion (AIC) value. The results of testing the factors that significantly influence cigarette consumption in Central Sulawesi Province in the HNB regression model, namely the count model are gender, age and health. Whereas in the zerohurdle model, it is gender, age and education.

**Keywords** : Cigarette, Excess Zero, Hurdle Negative Binomial, Overdispersion, Sulawesi Tengah

### ABSTRAK

Rata-rata pengeluaran perkapita untuk rokok di Provinsi Sulawesi Tengah mengalami kenaikan pada Tahun 2020. Terdapat beberapa faktor yang dapat mempengaruhi konsumsi rokok seseorang antara lain jenis kelamin, umur, pendidikan dan kesehatan. Untuk memodelkan konsumsi rokok dengan beberapa faktor yang mempengaruhi tersebut dapat menggunakan model regresi poisson ataupun model *Zero Inflated Poisson* (ZIP). Akan tetapi kedua model regresi tersebut tidak dapat mengatasi permasalahan *excess zero* dan overdispersi sehingga digunakan model regresi *Hurdle Negative Binomial* (HNB). Hasil analisis data konsumsi rokok Provinsi Sulawesi Tengah menggunakan model HNB memberikan hasil pemodelan terbaik dibandingkan model regresi poisson dan model ZIP dikarenakan memiliki nilai *Akaike's Information Criterion* (AIC) terkecil. Hasil pengujian faktor yang mempengaruhi konsumsi rokok di Provinsi Sulawesi Tengah secara signifikan pada model regresi HNB yaitu model *count* adalah jenis kelamin, umur dan kesehatan. Sedangkan pada pada model *zero hurdle* adalah jenis kelamin, umur dan pendidikan.

**Kata kunci** : Rokok, *Excess Zero*, *Hurdle Negative Binomial*, Overdispersi, Sulawesi Tengah

## I. PENDAHULUAN

Rokok adalah lintingan atau gulungan tembakau yang digulung dengan kertas atau kulit jagung. Rokok bagi sebagian orang merupakan sebuah benda yang harus dihindari karena pengaruh buruknya bagi kesehatan, tetapi bagi sebagian lainnya rokok telah menjadi sebuah kebutuhan. Sebagai gambaran, rata-rata pengeluaran per kapita sebulan untuk komoditas makanan di Provinsi Sulawesi Tengah sebesar 516.839 rupiah (BPS, 2021). Dari jumlah pengeluaran tersebut sebesar 77.566 rupiah atau 15 persen merupakan pengeluaran untuk rokok. Angka tersebut memang masih kalah dibandingkan dengan jumlah pengeluaran untuk makanan dan minuman jadi (135.774 rupiah) tetapi lebih tinggi dibandingkan pengeluaran untuk padi-padian yang hanya 76.268 rupiah. Jika dibandingkan dengan tahun 2019 terjadi peningkatan jumlah pengeluaran untuk rokok karena di tahun 2019 hanya 75.472 rupiah. Hal ini menunjukkan adanya peningkatan perilaku merokok di Provinsi Sulawesi Tengah.

Perilaku merokok seseorang dipengaruhi oleh beberapa faktor. Penelitian Timban (2018) menunjukkan bahwa terdapat hubungan umur, jenis kelamin dan tingkat pendidikan dengan status merokok. Penelitian tersebut memberikan hasil bahwa terdapat hubungan antara umur dengan merokok dan paling banyak terjadi pada umur yang lebih tua. Jenis kelamin mempengaruhi perilaku merokok dimana lebih banyak perokok laki-laki dibandingkan perokok perempuan. Sedangkan dilihat dari pendidikan, pendidikan seseorang berpengaruh dengan perilaku merokok karena seseorang dengan pendidikan rendah cenderung merokok lebih banyak dari pada seseorang yang berpendidikan tinggi. Selain itu Susanna (2003) dalam Riska Rosita (2012) menyimpulkan bahwa perokok akan semakin mudah untuk berhenti merokok ketika kesehatannya terganggu akibat semakin banyaknya jumlah rokok yang dihisap. Untuk melihat adanya pengaruh jenis kelamin, umur, pendidikan dan kesehatan terhadap jumlah konsumsi rokok diperlukan sebuah metode analisis yang salah satu menggunakan analisis regresi.

Analisis regresi adalah analisis statistika yang digunakan untuk memodelkan hubungan antara variabel dependen dengan satu atau lebih variabel independen (Hosmer, 2020). Persamaan regresi digunakan untuk menyatakan hubungan antara variabel dependen dan variabel independen. Variabel dependen pada analisis regresi umumnya berupa data kontinu. Akan tetapi tidak menutup kemungkinan variabel dependen adalah berupa data diskrit. Model regresi yang dapat digunakan untuk menganalisis hubungan antara variabel independen berupa data kontinu, diskrit atau campuran dan variabel dependen berupa data diskrit salah satunya adalah model regresi Poisson.

Model regresi Poisson memiliki beberapa asumsi yang harus dipenuhi. Asumsi tersebut antara lain variabel dependen berdistribusi Poisson, variabel independen terbebas dari multikolinearitas dan *equidispersi*. *Equidispersi* merupakan kondisi dimana variabel dependen memiliki nilai mean dan varians sama (Myers, 2010). Akan tetapi terkadang terjadi suatu keadaan yang disebut *overdispersi* dimana nilai varian lebih besar dari nilai rata-rata. Salah satu penyebab kondisi *overdispersi* pada data *count* adalah *excess zero* yang merupakan kondisi dimana proporsi

nilai nol pada data variabel dependen lebih besar dari nilai lainnya. Pada dasarnya kondisi *excess zero* pada model regresi poisson ini dapat diatasi dengan model regresi *Zero Inflated Poisson* (ZIP). Permasalahannya adalah ketika *excess zero* dan *overdispersi* terjadi bersamaan maka model regresi *Zero Inflated Poisson* (ZIP) kurang sesuai digunakan untuk memodelkan data dengan kondisi *excess zero* dan *overdispersi* karena model yang terbentuk menyebabkan estimasi yang bias. Menurut Zhang (2019), salah satu model yang mampu mengatasi permasalahan overdispersi akibat banyaknya data yang bernilai nol (*excess zeros*) adalah model regresi *Hurdle Negative Binomial* (HNB).

Berdasarkan penjelasan diatas maka pada penelitian ini dilakukan penerapan model regresi *Hurdle Negative Binomial* (HNB) pada data konsumsi rokok di Provinsi Sulawesi Tengah. Hasil pemodelan data konsumsi rokok di Provinsi Sulawesi Tengah dengan menggunakan model regresi *Hurdle Negative Binomial* (HNB) diharapkan mampu memberikan hasil yang lebih baik daripada model regresi poisson dan model regresi *Zero Inflated Poisson* (ZIP) dalam mengatasi masalah *excess zero* dan overdispersi.

## II. METODE PENELITIAN

### 2.1. Data dan Variabel Penelitian

Penelitian ini menggunakan data sekunder yang diperoleh dari Badan Pusat Statistik (BPS) Provinsi Sulawesi Tengah. Data tersebut merupakan data hasil pendataan Survei Sosial Ekonomi Nasional (Susenas) Maret 2020. Variabel penelitian terdiri dari variabel dependen yaitu jumlah konsumsi rokok (Y) dan variabel independen antara lain Jenis Kelamin ( $X_1$ ), Umur ( $X_2$ ), Pendidikan ( $X_3$ ) dan Kesehatan ( $X_4$ ). Jumlah pengamatan pada penelitian ini adalah 20.254 orang yang merupakan orang berusia di atas 15 tahun yang selama seminggu terakhir melakukan salah satu kegiatan bekerja, sekolah, mengurus rumah tangga atau kegiatan lain selain kegiatan pribadi. Variabel yang digunakan pada penelitian ini disajikan pada Tabel 1.

Tabel 1 : Variabel Penelitian

Variabel	Deskripsi	Keterangan
Y	Jumlah Konsumsi Rokok	Batang/Minggu
$X_1$	Jenis Kelamin	1 = Laki-laki 2 = Perempuan
$X_2$	Umur	Data rasio
$X_3$	Pendidikan	1 = SD/Sederajat 2 = SMP/Sederajat 3 = SMA/Sederajat 4 = PT/Sederajat
$X_4$	Kesehatan	1 = Ada Keluhan Kesehatan 2 = Tidak ada keluhan kesehatan

## 2.2. Metode Analisis Data

Metode analisis yang digunakan dalam penelitian ini adalah sebagai berikut :

1. Melakukan analisis deskriptif
2. Melakukan identifikasi multikolinearitas pada variabel independen
3. Membentuk Model Regresi Poisson
  - a. Melakukan estimasi parameter model
  - b. Melakukan pengujian kesesuaian model
  - c. Melakukan pengujian signifikansi parameter model
  - d. Melakukan pengujian overdispersi
4. Membentuk Model Regresi *Zero Inflated Poisson* (ZIP)
  - a. Melakukan estimasi parameter model
  - b. Melakukan pengujian kesesuaian model
  - c. Melakukan pengujian signifikansi parameter model
  - d. Melakukan pengujian overdispersi
5. Membentuk Model Regresi *Hurdle Negative Binomial* (HNB)
  - a. Melakukan estimasi parameter model
  - b. Melakukan pengujian kesesuaian model
  - c. Melakukan pengujian signifikansi parameter model
  - d. Melakukan pengujian overdispersi
6. Menentukan model terbaik menggunakan kriteria AIC
7. Menginterpretasi model terbaik
8. Pengolahan data pada penelitian ini menggunakan bantuan software R seri 3.6.1.

## III. HASIL DAN PEMBAHASAN

### 3.1. Analisis Deskriptif

Hasil analisis deskriptif variabel penelitian yang digunakan pada penelitian ini disajikan pada Tabel 2.

Tabel 2 : Statistik Deskriptif Variabel Penelitian

Keterangan	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	Y
Minimum	1	1	1	1	0
Rata-rata	1,5	2,084	2,183	1,751	29,25
Maksimum	2	4	4	2	560
Standar Deviasi	0,50001	0,96767	1,09643	0,43238	54,26021

### 3.2. Identifikasi Multikolinearitas

Nilai *Variance Inflation Factor* (VIF) digunakan untuk mengidentifikasi adanya multikolinieritas pada variabel independen (Gujarati, 2009). Nilai VIF  $\geq 10$  menunjukkan bahwa terdapat multikolinieritas. Nilai VIF dari variabel independen pada Tabel 3

menunjukkan nilai kurang dari 10 sehingga dapat disimpulkan bahwa tidak terjadi multikolinieritas.

Tabel 3 : Nilai VIF Variabel Independen

Variabel Independen	VIF
X <sub>1</sub>	1,00248
X <sub>2</sub>	1,16522
X <sub>3</sub>	1,12972
X <sub>4</sub>	1,04458

### 3.3. Model Regresi *Poisson*

Model regresi Poisson merupakan penerapan dari *Generalized Linear Model* (GLM). GLM adalah perluasan dari model regresi yang menggambarkan hubungan antara variabel dependen dengan variabel independen dimana variabel dependennya memiliki sebaran eksponensial. Pada regresi Poisson variabel dependen merupakan data cacah atau *count* (Hilbe, 2011). Model regresi Poisson digunakan untuk memodelkan banyaknya kemunculan suatu kejadian pada interval waktu tertentu.

Pada Tabel 4 disajikan estimasi parameter dan uji signifikansi parameter model regresi poisson.

Tabel 4 : Estimasi Parameter Model Regresi Poisson

Variabel	Estimasi	Standar Error	<i>p-value</i>
Intercept	7,930785	0,012212	<2e-16
X <sub>1</sub>	-3,878029	0,009200	<2e-16
X <sub>2</sub>	0,074165	0,001401	<2e-16
X <sub>3</sub>	-0,147878	0,001289	<2e-16
X <sub>4</sub>	0,084047	0,003165	<2e-16

Berdasarkan pada Tabel 4 dilakukan pengujian parameter model regresi poisson secara simultan menggunakan nilai *Likelihood Ratio Test* (LRT) dengan  $H_0$  adalah variabel independen tidak berpengaruh signifikan terhadap variabel dependen. Hasil perhitungan LRT menggunakan *Software R 3.6.1* diperoleh nilai LRT sebesar 724.036 lebih besar dari  $\chi^2_{0,05;5} = 11,07$  sehingga dapat disimpulkan bahwa  $H_0$  ditolak. Artinya bahwa terdapat salah satu variabel independen yang berpengaruh signifikan terhadap variabel dependen. Selanjutnya dilakukan pengujian secara parsial dengan  $\alpha = 0,05$  yang menunjukkan bahwa semua parameter model poisson memiliki nilai *p-value* <  $\alpha = 0,05$ . Hal ini dapat disimpulkan bahwa semua variabel independen berpengaruh secara signifikan terhadap Jumlah Konsumsi Rokok (Y).

Menurut Payne (2018), salah satu cara untuk mendeteksi overdispersi adalah dengan membagi *Pearson Chi-Square* dengan derajat bebas. Jika hasil pembagian melebihi angka 1, maka dikatakan terjadi overdispersi. Rumus pembagian *Pearson Chi-Square* dengan derajat bebas diperoleh

$$\hat{\phi} = \frac{882.858,9}{20.248} = 43,60228$$

Nilai rasio dispersi  $\hat{\phi} = 43,60228 > 1$  yang berarti bahwa pada model regresi poisson terjadi *overdispersi*

### 3.4. Model Regresi *Zero Inflated Poisson* (ZIP)

Model regresi ZIP adalah model campuran yang sederhana untuk data diskret dengan banyak peristiwa nol (Jansakul dan Hinde, 2002). Model regresi *Zero Inflated Poisson* (ZIP) terdiri dari model *count* atau model log dan model *zero inflation* atau model logit. Estimasi dan uji signifikansi parameter model regresi ZIP disajikan pada Tabel 5.

Tabel 5 : Estimasi Parameter Model Regresi ZIP

Variabel	Estimasi	Standar Error	<i>p-value</i>
Model <i>Count</i> atau Model Log			
Intercept	4,8093143	0,0124499	<2e-16
X <sub>1</sub>	-0,4266555	0,0092568	<2e-16
X <sub>2</sub>	0,0295158	0,0014835	<2e-16
X <sub>3</sub>	-0,0005425	0,0012715	0,67
X <sub>4</sub>	0,0739746	0,0031534	<2e-16
Model <i>Zero Inflation</i> atau Model Logit			
Intercept	-5,345001	0,146152	< 2e-16
X <sub>1</sub>	4,440991	0,076841	< 2e-16
X <sub>2</sub>	-0,126111	0,021854	7,9e-09
X <sub>3</sub>	0,355931	0,019348	< 2e-16
X <sub>4</sub>	0,002533	0,048243	0,958

Berdasarkan pada Tabel 5 dilakukan pengujian kesesuaian model regresi ZIP dengan  $H_0$  adalah variabel independen tidak berpengaruh signifikan terhadap variabel dependen. Pengujian kesesuaian model regresi ZIP dilakukan menggunakan *Likelihood Ratio Test* (LRT). Berdasarkan perhitungan LRT menggunakan *Software R 3.6.1* diperoleh nilai LRT sebesar 13.116 lebih besar dari  $\chi^2_{0,05;10} = 18,307$  sehingga disimpulkan  $H_0$  ditolak. Hal ini berarti bahwa terdapat minimal satu variabel independen yang berpengaruh signifikan terhadap variabel dependen. Sedangkan pengujian masing-masing variabel independen secara parsial dengan  $\alpha = 0,05$  menunjukkan bahwa :

1. Pada model *count* nilai *p-value*  $< \alpha = 0,05$  adalah pada variabel Jenis Kelamin (X<sub>1</sub>), variabel Umur (X<sub>2</sub>) dan variabel Kesehatan (X<sub>4</sub>) yang berarti bahwa ketiga variabel

tersebut berpengaruh signifikan terhadap Jumlah Konsumsi Rokok (Y). Sedangkan untuk variabel Pendidikan (X<sub>3</sub>) memiliki nilai *p-value* >  $\alpha = 0,05$  yang berarti bahwa variabel Pendidikan (X<sub>3</sub>) tidak berpengaruh secara signifikan terhadap Jumlah Konsumsi Rokok (Y).

2. Pada model *zero inflation* dengan  $\alpha = 0,05$  nilai *p-value* <  $\alpha = 0,05$  adalah pada variabel Jenis Kelamin (X<sub>1</sub>), variabel Umur (X<sub>2</sub>) dan variabel Pendidikan (X<sub>3</sub>) yang berarti bahwa ketiga variabel tersebut berpengaruh signifikan terhadap Jumlah Konsumsi Rokok (Y). Sedangkan untuk variabel Kesehatan (X<sub>4</sub>) memiliki nilai *p-value* >  $\alpha = 0,05$  yang berarti bahwa variabel Kesehatan (X<sub>4</sub>) tidak berpengaruh secara signifikan terhadap Jumlah Konsumsi Rokok (Y).

Uji asumsi overdispersi pada model regresi ZIP dilakukan dengan membagi nilai *Pearson Chi-Square* dengan derajat bebas. Jika hasil pembagian melebihi angka 1, maka dikatakan terjadi overdispersi. Rumus pembagian *Pearson Chi-Square* dengan derajat bebas diperoleh

$$\hat{\phi} = \frac{33.812,46}{20.243} = 1,67033$$

Nilai rasio dispersi  $\hat{\phi} = 1,67033 > 1$  yang berarti bahwa pada regresi ZIP terjadi overdispersi sehingga untuk mengatasinya digunakan model regresi *Hurdle Negative Binomial* (HNB).

### 3.5. Model Regresi *Hurdle Negative Binomial* (HNB)

Model *Hurdle Negative Binomial* (HNB) digunakan untuk memodelkan data *count* yang memiliki kelebihan nol (*excess zero*) dan overdispersi (Andika, 2019). Model regresi HNB menangani *outcome* dari variabel dependen secara terpisah. Ketika *outcome* dari variabel dependen positif, maka digunakan distribusi yang terpenggal (*truncated*). Model regresi *Hurdle Negative Binomial* (HNB) terdiri dari model *count* atau model *truncated negative binomial* dan model *zero hurdle* atau model logit. Estimasi dan uji signifikansi parameter model regresi HNB disajikan pada Tabel 6.

Tabel 6 : Estimasi Parameter Model Regresi HNB

Variabel	Estimasi	Standar Error	<i>p-value</i>
<i>Model Count</i> atau Model <i>Truncated Negative Binomial</i>			
Intercept	4,8049848	0,0661885	< 2e-16
X <sub>1</sub>	-0,4282909	0,0452697	< 2e-16
X <sub>2</sub>	0,0313608	0,0091048	0,000572
X <sub>3</sub>	-0,0007638	0,0075054	0,918940
X <sub>4</sub>	0,0753190	0,0184172	4,32e-05
Log (theta)	1,0281307	0,0178695	< 2e-16

Model <i>Zero Hurdle</i> atau Model Logit			
Intercept	5,345001	0,146152	< 2e-16
X <sub>1</sub>	-4,440991	0,076841	< 2e-16
X <sub>2</sub>	0,126111	0,021854	7,9e-09
X <sub>3</sub>	-0,355931	0,019348	< 2e-16
X <sub>4</sub>	-0,002533	0,048243	0,958

Berdasarkan pada Tabel 6 dilakukan pengujian kesesuaian model regresi HNB dengan  $H_0$  adalah variabel independen tidak berpengaruh signifikan terhadap variabel dependen. Pengujian kesesuaian model regresi HNB dilakukan menggunakan *Likelihood Ratio Test* (LRT). Berdasarkan perhitungan LRT menggunakan Software R 3.6.1 diperoleh nilai LRT sebesar 9.858 lebih besar dari  $\chi^2_{0,05;11} = 19,675$  sehingga disimpulkan  $H_0$  ditolak. Hal ini berarti bahwa terdapat minimal satu variabel independen yang berpengaruh signifikan terhadap variabel dependen. Sedangkan pengujian masing-masing variabel independen secara parsial dengan  $\alpha = 0,05$  menunjukkan bahwa :

1. Pada model *count* nilai  $p\text{-value} < \alpha = 0,05$  adalah pada variabel Jenis Kelamin ( $X_1$ ), variabel Umur ( $X_2$ ) dan variabel Kesehatan ( $X_4$ ) yang berarti bahwa ketiga variabel tersebut berpengaruh signifikan terhadap Jumlah Konsumsi Rokok (Y). Sedangkan untuk variabel Pendidikan ( $X_3$ ) memiliki nilai  $p\text{-value} > \alpha = 0,05$  yang berarti bahwa variabel Pendidikan ( $X_3$ ) tidak berpengaruh secara signifikan terhadap Jumlah Konsumsi Rokok (Y).
2. Pada model *zero Hurdle* dengan  $\alpha = 0,05$  nilai  $p\text{-value} < \alpha = 0,05$  adalah pada variabel Jenis Kelamin ( $X_1$ ), variabel Umur ( $X_2$ ) dan variabel Pendidikan ( $X_3$ ) yang berarti bahwa ketiga variabel tersebut berpengaruh signifikan terhadap Jumlah Konsumsi Rokok (Y). Sedangkan untuk variabel Kesehatan ( $X_4$ ) memiliki nilai  $p\text{-value} > \alpha = 0,05$  yang berarti bahwa variabel Kesehatan ( $X_4$ ) tidak berpengaruh secara signifikan terhadap jumlah konsumsi rokok (Y).

Uji asumsi overdispersi pada model regresi HNB dilakukan dengan membagi nilai *Pearson Chi-Square* dengan derajat bebas. Jika hasil pembagian melebihi angka 1, maka dikatakan terjadi overdispersi. Rumus pembagian *Pearson Chi-Square* dengan derajat bebas diperoleh

$$\hat{\phi} = \frac{20.949,31}{20.243} = 1,03489$$

Nilai rasio dispersi  $\hat{\phi} = 1,03489 > 1$  tetapi angka tersebut sangat mendekati nilai 1 maka dapat dikatakan bahwa model HNB mampu mengatasi masalah overdispersi jika dibandingkan model regresi Poisson dan model regresi ZIP.



### 3.6. Penentuan Model Terbaik

Menurut Sharma dan Landge (2013), nilai *Akaike's Information Criterion* (AIC) digunakan sebagai kriteria pemilihan model terbaik. Model yang terbaik yaitu dengan memilih model yang mempunyai nilai AIC terkecil. Perbandingan nilai AIC dari Model Regresi Poisson, Model ZIP dan Model HNB disajikan pada Tabel 7.

Tabel 7 : Perbandingan Nilai AIC

Model	AIC
Regresi Poisson	921.098,7
Regresi ZIP	246.998,3
Regresi HNB	81.122,9

Pada Tabel 7 dapat dilihat bahwa nilai AIC model regresi HNB lebih kecil dibandingkan nilai AIC model regresi poisson dan model regresi ZIP. Hal ini dapat disimpulkan bahwa model HNB lebih baik digunakan dalam memodelkan data jumlah konsumsi rokok di Provinsi Sulawesi Tengah.

### 3.7. Interpretasi model regresi HNB

Interpretasi model regresi HNB adalah :

1. Model *count* menjelaskan bahwa jumlah konsumsi rokok dipengaruhi oleh variabel independen yang signifikan.
  - a. Pada variabel Jenis Kelamin ( $X_1$ ) dapat diinterpretasikan bahwa perempuan memiliki jumlah konsumsi rokok sebesar  $\exp(-0,4282909) = 0,01380$  kali lebih banyak atau dapat dikatakan bahwa jumlah konsumsi rokok laki-laki adalah 72,45089 kali lebih banyak dibandingkan jumlah konsumsi rokok perempuan.
  - b. Pada variabel Umur ( $X_2$ ) dapat diinterpretasikan bahwa ketika umur seseorang meningkat satu jenjang umur maka jumlah konsumsi rokok meningkat  $\exp(0,0313608) = 1,03186$  kali.
  - c. Pada variabel Kesehatan ( $X_4$ ) dapat diinterpretasikan bahwa seseorang yang tidak mempunyai keluhan kesehatan memiliki jumlah konsumsi rokok  $\exp(0,0753190) = 1,07823$  lebih banyak dibandingkan seseorang yang mempunyai keluhan kesehatan.
2. Model *Zero Hurdle* menjelaskan bahwa peluang jumlah konsumsi rokok bernilai nol dipengaruhi oleh variabel independen yang berpengaruh secara signifikan terhadap Jumlah Konsumsi Rokok ( $Y$ ).
  - a. Pada variabel Jenis Kelamin ( $X_1$ ), perempuan memiliki peluang untuk merokok sebesar  $\exp(-4,440991) = 0,01178$  kali lebih besar dibanding laki-laki atau dapat dikatakan bahwa peluang perempuan untuk tidak merokok adalah 84,85900 kali lebih besar dibandingkan laki-laki.

- b. Pada variabel Umur ( $X_2$ ), setiap kenaikan satu jenjang umur, maka peluang seseorang untuk tidak merokok adalah  $\exp(0,126111) = 1,13441$  kali lebih besar dibandingkan jenjang umur dibawahnya.
- c. Pada variabel Pendidikan ( $X_3$ ), setiap kenaikan satu jenjang pendidikan, maka peluang seseorang untuk tidak merokok adalah  $\exp(-0,355931) = 0,70053$  kali lebih lebih besar dibandingkan jenjang pendidikan dibawahnya. Hal ini dapat dikatakan bahwa setiap kenaikan satu jenjang pendidikan, peluang seseorang untuk merokok adalah 1,42751 kali lebih besar dibandingkan jenjang pendidikan dibawahnya.

#### IV. KESIMPULAN

Berdasarkan hasil analisis data konsumsi rokok di Provinsi Sulawesi Tengah diperoleh kesimpulan bahwa :

1. perbandingan nilai AIC model regresi poisson, model regresi ZIP dan model regresi HNB menunjukkan bahwa nilai AIC model regresi HNB yang paling kecil. Hal ini dapat disimpulkan bahwa model regresi HNB lebih baik digunakan untuk memodelkan data konsumsi rokok di Provinsi Sulawesi Tengah yang mengandung *excess zero* dan overdispersi.
2. Hasil pengujian faktor yang mempengaruhi konsumsi rokok di Provinsi Sulawesi Tengah secara signifikan pada model regresi HNB yaitu model *count* adalah jenis kelamin, umur dan kesehatan. Sedangkan pada pada model *zero hurdle* adalah jenis kelamin, umur dan pendidikan.

#### DAFTAR PUSTAKA

- [1]. Andika, A., Abdullah, A.S., dan Nurrohmah, S, Hurdle Negative Binomial Regression Model, Proceeding of ICSA 2019, 57-68.
- [2]. Badan Pusat Statistik Provinsi Sulawesi Tengah, *Provinsi Sulawesi Tengah Dalam Angka 2021*, Palu, 2021, Badan Pusat Statistik.
- [3]. Gujarati, D., *Dasar-Dasar Ekonometrika Jilid 2*, Jakarta, 2009, Erlangga.
- [4]. Hilbe, J.M., *Negative Binomial Regression, Second Edition*, New York, 2011, Cambrigde University Press.
- [5]. Hosmer, D.W., and Lemeshow, S, *Applied Logistic Regression*. New York, 2000, John Wiley & Sons, Inc.
- [6]. Jansakul, N., and Hinde, J.P, Score Tests for Zero-Inflated Poisson Models, *Computational Statistics & Data Analysis* Vol. 40 No.1, 2002, 75-96.

- [7]. Myers, R.H., Montgomery, D.C., Vining, G.G., *et al*, *Generalized Linear Models with Applications in Engineering and The Sciences*, Second edition, New Jersey, 2010, John Wiley and Sons.
- [8]. Payne, E.H, Gebregziabher, M., Hardind, J.W., *et al*, An empirical approach to determine a threshold for assessing overdispersion in Poisson and negative binomial models for count data, *Commun Stat Simul Comput* 47(6), 2018, 1722–1738.
- [9]. Rosita, R., Suswardanya, D.L, dan Abidin, Z, Penentu Keberhasilan Berhenti Merokok Pada Mahasiswa, *KEMAS* 8(1), 2012, 1-9.
- [10]. Sharma, A.K., and Landge, V.S, Zero Inflated Negative Binomial for Modelling Heavy Vehicle Crash Rate on Indian Rural Highway, *International Journal of Advances in Engineering and Technology* 5(2), 2013, 292-301.
- [11]. Timban, I., Langi, F.F.L.G., dan Kaunang, W.P.J, Determinan Merokok di Indonesia Analisis SURvei Demografi dan Kesehatan Indonesia Tahun 2012, *Jurnal KESMAS*, 2018, Vol. 7 No. 5.
- [12]. Zhang, C., *Statistical Modeling of Count Data with Over-Dispersion or Zero-Inflation Problems*, Theses, Dissertations and Culminating Projects, 2019.