

KONSUMSI ROKOK MASYARAKAT KOTA BANDUNG TAHUN 2015 DENGAN MODEL HURDLE NEGATIF BINOMIAL (HURDLE-NB)

Wulandari¹, Wida Tira Tedra², Irtania Muthia Rizki³, Dina Prariesa⁴

^{1,2,4}Badan Pusat Statistik, ³UIN Sulthan Thaha Saifuddin Jambi

Abstrak

Perilaku merokok memiliki resiko yang cukup tinggi terhadap kesehatan. Banyak dampak negatif yang ditimbulkan, baik untuk perokok aktif maupun perokok pasif. Selain itu, merokok juga berdampak terhadap ekonomi pelaku, selain pengeluaran untuk konsumsi rokok, juga biaya berobat akibat penyakit yang ditimbulkan oleh rokok. Di Kota Bandung, pengeluaran untuk rokok menempati urutan kedua setelah komoditi makanan jadi. Jumlah batang rokok yang dihisap setiap hari dipengaruhi oleh variabel demografi, variabel lingkungan sosial, variabel politik, serta variabel budaya. Penelitian ini bertujuan untuk mengidentifikasi faktor-faktor yang mempengaruhi jumlah rokok yang dihisap tiap hari masyarakat Kota Bandung. Variabel jenis kelamin, umur, status bekerja, dan pendidikan, akan dimodelkan dengan regresi Hurdle-NB. Hasil penelitian menunjukkan pada model Log, variabel umur, status bekerja, dan pendidikan berpengaruh terhadap rata-rata konsumsi rokok. Sedangkan pada model Logit, variabel jenis kelamin, umur, status bekerja, dan pendidikan berpengaruh terhadap kecenderungan seseorang untuk merokok atau tidak.

Kata kunci : rokok, zero excess, Hurdle-NB

Abstract

Lifestyle that high risk to health is smoking behavior. Smoking behavior has a negative impact on health, both for active smokers and passive smokers. In addition there are also negative impacts in terms of economy. In Bandung, spending on cigarettes ranks second after the food commodity. The number of cigarettes smoked each day is influenced by demographic variables, social environment variables, political variables, and cultural variables. In this study, the Bandung cigarette consumption and the factors suspected to affect, ie sex, age, work status, and education, will be modeled by Hurdle-NB regression. The results showed that Log model, age variable, work status, and education influence to the average of cigarette consumption. While on Logit model, gender variable, age, work status, and education have an effect on the tendency of someone to smoke or not.

Keywords : cigarette, zero excess, Hurdle-NB

PENDAHULUAN

Dalam kehidupan manusia, kesehatan merupakan masalah yang sangat penting. Namun, terdapat salah satu pola hidup yang berisiko tinggi terhadap kesehatan yaitu perilaku merokok. Perilaku merokok itu sendiri dapat diukur salah satunya melalui banyaknya jumlah rokok yang dihisap (Fitriani, 2012, dalam Hestiana, 2013). Perilaku merokok banyak memiliki dampak negatif untuk segi kesehatan. Kesehatan yang terancam akibat mengkonsumsi rokok tidak hanya didapatkan oleh orang yang merokok, melainkan juga didapatkan oleh orang disekitar perokok yang menghirup asap dari rokok (perokok pasif). Sebatang rokok mengandung 4.000 jenis senyawa kimia beracun yang berbahaya untuk tubuh, dimana 43 diantaranya bersifat karsiogenik (Aditama 2013, dalam Ambarwati, dkk 2014). Rokok memiliki faktor resiko berbagai macam penyakit, antara lain penyakit kardiovaskular, serebrovaskular, impotensi, berbagai jenis kanker (Alit, dkk dalam Santi, 2013).

Selain itu terdapat pula dampak negatif rokok dari segi ekonomi. Merokok menyebabkan dampak buruk bagi masyarakat karena kematian prematur, produktivitas yang hilang, dan beban keuangan yang ditanggung oleh perokok dan keluarga mereka, penyedia jasa keuangan, penyedia layanan asuransi, dan perusahaan pemberi kerja (theconversation.com). Kerugian makro ekonomi akibat konsumsi rokok di Indonesia pada 2015 mencapai hampir Rp 600 triliun, atau empat kali lipat dari jumlah cukai rokok pada tahun yang sama (sains.kompas.com).

Perilaku merokok turut berdampak pada sisi biaya belanja rokok, yang membuat orang tua seakan mengabaikan masa depan anak serta mempertinggi probabilitas angka rumah tangga miskin di masa mendatang (merdeka.com). Survei yang dilakukan oleh Dinas Kesehatan Kota Bandung menyatakan jika konsumsi masyarakat Kota Bandung dalam belanja rokok dan tembakau menempati urutan kedua setelah komoditi makanan jadi. Pengeluaran yang cukup tinggi untuk konsumsi rokok sebenarnya

dapat dialokasikan untuk kebutuhan lain yang lebih penting.

Mengingat konsumsi rokok yang tinggi di Kota Bandung serta faktor risiko yang terjadi akibat mengkonsumsi rokok, maka penelitian ini ingin mengetahui faktor-faktor yang mempengaruhi jumlah batang rokok yang dihisap. Dalam penelitian ini, konsumsi rokok per hari masyarakat Kota Bandung akan dimodelkan dengan Regresi *Zero Inflated Negative Binomial* (ZINB) dan *Hurdle Negative Binomial* (HNB). Kedua alat analisis tersebut sesuai untuk memodelkan data cacah dengan kondisi *excess zero*, yaitu kondisi dimana data respon banyak bernilai nol, dan dapat menanggapi overdispersi yang umumnya sering muncul pada analisis data cacah.

METODOLOGI

Analisis regresi umumnya digunakan untuk mengetahui hubungan suatu variabel respon dengan satu atau beberapa variabel prediktor. Jumlah rokok yang dihisap setiap hari merupakan data cacah nonnegatif. Untuk pemodelan variabel respon yang berupa data cacah, model yang biasa digunakan adalah regresi Poisson, regresi Binomial, dan regresi *Negative Binomial*. Konsumsi rokok dalam batang per hari merupakan salah satu kasus data cacahan dengan banyak nilai 0 (*excess zero*). Jumlah angka nol besar akan menyebabkan sebaran data tidak sesuai dengan distribusi standar. Model yang biasanya digunakan pada kasus *excess zero* adalah regresi *Zero Inflated Poisson* (ZIP) atau *Hurdle Poisson* (HP).

Permasalahan yang sering muncul pada kedua pemodelan tersebut adalah terjadinya *overdispersion*. Untuk mengatasi masalah *overdispersion* yang terjadi, salah satu cara adalah menggunakan *Zero Inflated Negative Binomial* (ZINB) atau *Hurdle Negative Binomial* (HNB). Kedua model tersebut digunakan untuk memodelkan data *count* dengan banyak nilai 0 pada respon dan terjadi *overdispersion*. Berdasarkan uraian di atas, untuk mengetahui faktor-faktor yang mempengaruhi jumlah batang rokok yang dihisap setiap hari di Kota Bandung pada tahun 2015 maka akan dimodelkan

konsumsi rokok dengan metode *Zero Inflated Negative Binomial* serta *Hurdle Negative Binomial*, kemudian memilih model mana yang lebih cocok digunakan untuk memodelkan konsumsi rokok di Kota Bandung.

Faktor Pemicu Kebiasaan Merokok

Terdapat banyak penelitian, baik dalam negeri maupun luar negeri berkenaan dengan perilaku merokok serta faktor-faktor yang mempengaruhi perilaku merokok. Smet (1999) dalam penelitian tentang penentu kebiasaan merokok di kalangan remaja di Semarang menyatakan bahwa kebiasaan merokok teman dekat adalah penentu utama seorang remaja untuk merokok. Moghimbeigi, dkk (2009) melakukan penelitian tentang kebiasaan pada remaja 15 – 20 tahun. Hasilnya remaja laki-laki lebih berisiko untuk merokok dibanding remaja perempuan, usia muda lebih berpeluang untuk memulai kebiasaan merokok dibanding usia yang lebih tua, dan tingkat pendidikan tidak berpengaruh terhadap konsumsi rokok. Hestiana (2013) dalam penelitiannya menyimpulkan bahwa jumlah batang rokok yang dihisap setiap hari dipengaruhi secara signifikan oleh jenis kelamin, umur, tingkat pendidikan, dan kegiatan utama yang dilakukan. Selain itu,

$$\Pr(X = x; r, p) = \begin{cases} \binom{x-1}{r-1} p^r (1-p)^{x-r}, & x = r, r+1, r+2, \dots \\ 0 & , x \text{ yang lain} \end{cases} \quad (2.2)$$

Pemeriksaan Multikolinieritas

Dalam analisis regresi, terdapat beberapa asumsi yang harus terpenuhi agar estimasi yang dihasilkan bersifat *unbiased* dan memiliki varians yang minimum. Salah satu asumsi yang harus terpenuhi adalah tidak terdapatnya multikolinieritas antar variabel prediktor. Multikolinieritas menunjukkan terdapat hubungan antara beberapa atau semua peubah prediktor pada analisis regresi. Adanya korelasi dalam model regresi menyebabkan taksiran parameter regresi menjadi tidak ada atau model yang dihasilkan memiliki varians yang besar (Gujarati, 2004). Salah satu deteksi

Wijayanti, dkk (2017) juga menyimpulkan bahwa jenis kelamin, usia, pengalaman, pengetahuan, dan sikap signifikan mempengaruhi perilaku merokok.

Data Diskrit

Data diskrit adalah data *count*, yaitu data yang nilainya non-negatif dan menyatakan banyaknya kejadian dalam interval waktu, ruang, atau volume tertentu. Beberapa distribusi yang mempunyai bentuk fungsi peluang dan nama tertentu dari peubah acak diskrit, yaitu:

a. Distribusi Poisson

Fungsi probabilitas variabel random diskrit yang berdistribusi Poisson dengan parameter λ (Cameron dan Trivedi, 1998) adalah sebagai berikut :

$$f_Y(y, \mu) = \begin{cases} \frac{\mu^y \cdot e^{-\mu}}{y!}, & y = 0, 1, 2, \dots \\ 0 & , y \text{ yang lain} \end{cases} \quad (2.1)$$

Distribusi Poisson memiliki rata-rata dan varians yang sama, yaitu $E(Y) = \mu$ dan $Var(Y) = \mu$

b. Distribusi Binomial Negatif

Distribusi binomial negatif mempunyai rata-rata $E(Y) = \frac{r(1-p)}{p}$ dan varian $Var(Y) = \frac{r(1-p)}{p}$.

Fungsi probabilitas variabel random yang berdistribusi Binomial Negatif yaitu:

multikolinieritas adalah dengan menggunakan *variance inflation factor* (VIF). Rumus VIF adalah sebagai berikut (Kutner *et al*, 2005):

$$VIF_j = \frac{1}{1-R_j^2} \text{ dengan } j=1, 2, \dots, p \quad (2.3)$$

dimana R_j^2 merupakan koefisien determinasi dari x_j sebagai variabel dependen. Nilai R_j^2 berkisar antara 0 dan 1. *Rule of thumbs* yang umumnya digunakan untuk menentukan adanya multikolinieritas adalah nilai VIF yang lebih dari 10.

Model Zero Inflated Poisson (ZIP)

Regresi ZIP lebih tepat digunakan dibandingkan dengan regresi Poisson ketika data mengandung lebih banyak kejadian nol, terutama karena sangat mudah untuk memenuhi kesesuaian model dan tidak sulit untuk ditafsirkan. Akan tetapi, model ZIP tidak cocok untuk data yang *over/underdispersion*. ZIP mengasumsikan bahwa populasi terdiri dari dua tipe pengamatan yang berbeda. Salah satu didasarkan pada data cacah dengan distribusi Poisson, sedangkan yang kedua adalah data yang bernilai nilai nol. ZIP adalah model campuran yang sederhana untuk data diskrit dengan banyak peristiwa bernilai nol. Jika Y_i adalah variabel acak bebas yang mempunyai distribusi ZIP, maka observasi nol dapat dikembangkan dalam dua langkah. Langkah pertama terjadi dengan peluang p_i dan hanya menghasilkan data bernilai nol (*zero state*), langkah kedua terjadi dengan peluang $1-p_i$ dan menghasilkan nilai positif dan nilai nol dengan mengikuti distribusi Poisson. Fungsi kepadatan peluang adalah:

$$P(Y_i = y_i) = \begin{cases} p_i + (1 - p_i)e^{-\mu_i} & \text{untuk } y_i = 0 \\ (1 - p_i) \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!} & \text{untuk } y_i = 1, 2 \end{cases} \quad (2.4)$$

Nilai rata-rata dan varians dari ZIP adalah sebagai berikut:

$$E(Y_i) = (1 - p_i)\mu_i \quad (2.5)$$

$$\begin{aligned} \text{Var}(Y_i) &= (1 - p_i)[\mu_i^2 + \mu_i] \\ &\quad - (1 - p_i)^2 \mu_i^2 \quad (2.6) \\ &= E(Y_i)(1 + p_i \mu_i) \end{aligned}$$

Karena model ZIP tidak cocok digunakan pada data yang *over/underdispersion* maka setelah model ZIP terbentuk perlu dilakukan pengecekan *over/underdispersion* sebagai berikut:

$$S_\emptyset = \frac{\sum_{i=1}^n ((y_i - \mu_i)^2 - y_i)}{\sqrt{2 \sum_{i=1}^n \mu_i^2}} \quad (2.7)$$

Dimana μ_i adalah rata-rata yang diperoleh dari model ZIP. Jika nilai $S_\emptyset > z_\alpha$ menunjukkan terdapat kasus *overdispersion*

$$P(Y_i = y_i | \mathbf{x}_i, \mathbf{z}_i) = \begin{cases} \frac{\exp(\mathbf{z}'_i \boldsymbol{\gamma})}{1 + \exp(\mathbf{z}'_i \boldsymbol{\gamma})} + \frac{1}{1 + \exp(\mathbf{z}'_i \boldsymbol{\gamma})} (1 + \alpha \exp(\mathbf{x}'_j \boldsymbol{\beta}))^{-\alpha^{-1}}; y_i = 0 \\ \frac{1}{1 + \exp(\mathbf{z}'_i \boldsymbol{\gamma})} \frac{\Gamma(y_i + \alpha^{-1})}{y_i! \Gamma(\alpha^{-1})} \left(\frac{\alpha^{-1}}{\alpha^{-1} + \exp(\mathbf{x}'_j \boldsymbol{\beta})} \right)^{\alpha^{-1}}; y_i > 0 \end{cases} \quad (2.9)$$

Model Hurdle Poisson

Model *Hurdle Poisson* merupakan metode alternatif yang dapat digunakan untuk memodelkan data yang mengandung lebih banyak kejadian nol (*zero inflated*). Pada model *Hurdle Poisson* juga dilakukan dua pemodelan. Misalkan bahwa $P(y_i > 0) = \pi_i$ dan $P(y_i = 0) = 1 - \pi_i$ dan $\{y_i | y_i > 0\}$ mengikuti pmf dari *truncated-at-zero* $f(y_i; \mu_i)$. Maka distribusinya secara lengkap sebagai berikut (Agresti, 2015):

$$P(Y_i = y_i) = \begin{cases} 1 - \pi_i & ; y_i = 0 \\ \pi_i \frac{f(y_i; \mu_i)}{1 - f(0; \mu_i)} & ; y_i > 0 \end{cases} \quad (2.8)$$

Fungsi peluang dari model *hurdle* merupakan gabungan antara peluang pada model Logit dan model *truncated Poisson* (Cantoni dan Zedini, 2010). Penaksiran parameter dari model *hurdle Poisson* menggunakan metode *Maximum Likelihood Estimation* (MLE). Dalam mendapatkan estimasi dengan metode MLE fungsi yang dihasilkan tidak linier sehingga diselesaikan dengan algoritma *Fisher Scoring*.

Model Zero Inflated Negative Binomial (ZINB)

Model *Zero Inflated Negative Binomial* (ZINB) merupakan salah satu model yang digunakan untuk mengatasi overdispersi pada data cacah (*count*) dan mengatasi masalah dimana pada data respon terdapat banyak sekali nilai nol (*excess zeros*). Pada model ini, data cacahan diasumsikan dihasilkan melalui dua proses. Proses pertama, munculnya data *count* yang hanya berisi nol ditentukan dengan probabilitas p . Pada proses kedua, nilai nol dan nilai positif pada data cacahan, kedua-duanya dihasilkan oleh suatu proses yang mengikuti distribusi *Negative Binomial* (NB). (Sumber)

Parameter-parameter ZINB adalah $\mu_i = (\mu_1, \mu_2, \dots, \mu_n)'$, $F_i = (F_1, F_2, \dots, F_n)'$ dan α merupakan parameter dispersi. Bentuk μ_i dan F_i . model untuk ZINB yaitu:

Model Hurdle Negative Binomial (HNB)

Menurut Pontoh (2015) model *hurdle Negative Binomial* merupakan alternatif untuk data yang memiliki banyak nilai nol dan mengatasi terjadinya *overdispersi*. Misalkan bahwa Y_i merupakan variabel random diskrit dengan i adalah bilangan bulat non negatif ($i = 1, 2, \dots, n$) dan Y_i merupakan variabel respon dari model

$$P(Y_i = y_i | x_i, z_i) = \begin{cases} p_i & ; y_i = 0 \\ (1 - p_i) \frac{\Gamma(y_i + \alpha^{-1})}{\Gamma(y_i + 1)\Gamma(\alpha^{-1})} \left(\frac{(1 + \alpha\mu_i)^{-\alpha^{-1}} - y_i \alpha^{\gamma_i} \mu_i^{\gamma_i}}{1 - (1 + \alpha\mu_i)^{-\alpha^{-1}}} \right) & ; y_i > 0 \end{cases} \quad (2.10)$$

Pengujian Hipotesis Regresi

Pengujian hipotesis regresi dilakukan untuk menguji kelayakan model secara serentak dan parsial untuk melihat tingkat keberartian parameter model yang terbentuk.

a. Pengujian Kelayakan Model Secara Serentak (Overall)

Pengujian kelayakan model secara serentak dilakukan untuk menguji estimasi dari parameter-parameter Regresi ZINB. Dimana bentuk *Likelihood Ratio Test* dapat dituliskan dengan

$$G^2 = -2 \ln \left(\frac{L(\Omega_0)}{L(\Omega)} \right) \quad (2.11)$$

Tolak H_0 jika $G_{hitung}^2 > \chi_{db, \alpha}^2$ dengan db merupakan derajat bebas dan α yaitu tingkat signifikansi. Tolak hipotesis nol berarti terdapat minimal satu parameter dalam model yang signifikan.

b. Pengujian Parsial Koefisien

Pengujian koefisien secara parsial dilakukan untuk melihat signifikansi dari masing-masing parameter koefisien regresi. Pengujian parsial dilakukan pada parameter β_j dan γ_j masing-masing sebagai berikut:

Pada pengujian parameter β_j , dihitung dengan uji Statistik Wald

$$W_j = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \quad (2.12)$$

Tolak H_0 jika $|W_j| > t_{\alpha/2}$ atau p value $< \alpha$, dengan α adalah tingkat signifikan dan SE menyatakan *standard error*. Artinya bahwa parameter $\hat{\beta}_j$ signifikan dalam model. Sementara untuk pengujian parameter γ_j , dihitung dengan uji Statistik Wald

regresi HNB, maka nilai dari variabel respon, dimodelkan dalam dua keadaan. Keadaan pertama disebut *zero state* dan menghasilkan hanya pengamatan bernilai nol, sementara keadaan kedua disebut *negative binomial state* yang memiliki sebaran Binomial Negative. Fungsi peluang dari Y_i adalah :

$$W_j = \frac{\hat{\gamma}_j}{SE(\hat{\gamma}_j)} \quad (2.13)$$

Tolak H_0 jika $|W_j| > t_{\alpha/2}$ atau p value $< \alpha$, dengan α adalah tingkat signifikan. Artinya bahwa parameter $\hat{\gamma}_j$ signifikan dalam model.

Pemilihan Model Terbaik

Dalam memilih model terbaik dari beberapa model yang berbeda, maka dapat dilakukan dengan melihat masing-masing nilai *Akaike's Information Critrtion* (AIC) pada masing-masing model. Dimana nilai AIC adalah sebagai berikut:

$$AIC = -2 \ln L(\Omega) + 2k \quad (2.14)$$

dengan k adalah jumlah parameter dalam model

Sumber Data

Sumber data dalam penelitian ini berasal dari data Survei Sosial Ekonomi Nasional (SUSENAS) yang dilaksanakan oleh Badan Pusat Statistik pada bulan Maret 2015 di Kota Bandung. Objek penelitian adalah penduduk Kota Bandung berusia minimal 17 tahun dengan jumlah sampel sebanyak 2.589 responden. Variabel penelitian terdiri dari jumlah batang rokok yang dikonsumsi (Y), jenis kelamin (X_1), umur (X_2), status bekerja (X_3), dan pendidikan (X_4). Pengolahan dalam analisis dilakukan dengan menggunakan program R dan *Microsoft Excel*.

HASIL DAN PEMBAHASAN

Karakteristik Perokok di Kota Bandung

Kebiasaan merokok merupakan salah satu indikator perilaku masyarakat yang berpengaruh negatif terhadap derajat kesehatan (Profil Kesehatan Indonesia, 2016). Kebiasaan merokok dapat diukur melalui umur mulai merokok, intensitas merokok (jumlah batang rokok yang dihisap setiap hari dan waktu yang dibutuhkan seseorang untuk segera merokok setelah bangun pagi), lama merokok, dan jenis rokok (Fitriani, 2012 dalam Hestiana, 2013). Informasi tentang perilaku perokok setiap hari dapat membantu memprediksi gambaran beban penyakit tidak menular yang akan datang seperti penyakit kardiovaskuler, diabetes, penyakit paru obstruktif kronis, dan kanker tertentu (Riskesdas, 2010).

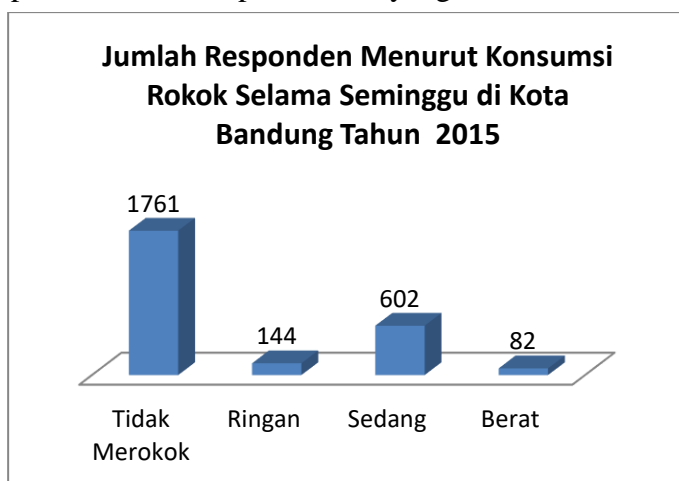
Penduduk berumur 17 tahun ke atas yang dianalisis sebanyak 2.589 responden. Dari 2.589 responden diketahui bahwa sejumlah 1.761 orang atau sebesar 68 persen responden adalah penduduk yang tidak

merokok selama satu minggu satu minggu terakhir, sedangkan 828 orang atau sebesar 32 persen responden adalah penduduk yang merokok setiap hari.

Menurut Smet (1994) dalam Hestiana (2013) mengklasifikasikan perokok berdasarkan banyaknya rokok yang dihisap, yaitu:

- Perokok berat yang menghisap lebih dari 100 batang rokok per minggu.
- Perokok sedang yang menghisap kurang lebih 25-99 batang rokok per minggu.
- Perokok ringan yang menghisap kurang lebih 1-24 batang rokok per minggu.

Dari penduduk yang dikategorikan merokok, sebagian besar responden rata-rata menghabiskan 25 – 99 rokok per minggunya atau sebanyak 23 persen. Jumlah responden kategori perokok ringan yaitu sebanyak 6 persen, sedangkan responden termasuk kategori perokok berat sebanyak 82 responden atau 3 persen.

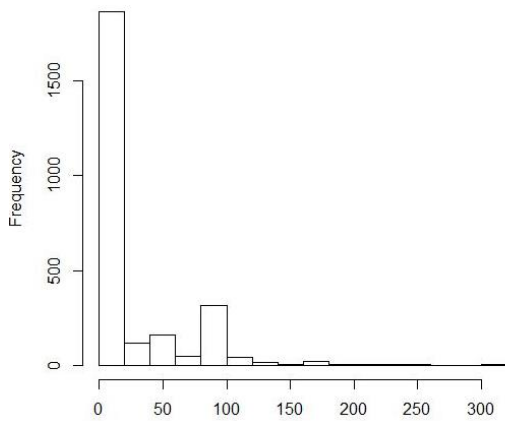


Statistik Deskriptif	
Mean	64.31
Standard Error	1.33
Median	72
Mode	84
Std. Deviation	38.39
Range	307
Minimal	2
Maksimal	309
Jumlah	828

Gambar 1. Gambaran Konsumsi Rokok Masyarakat Bandung Berdasarkan Statistik Deskriptif

Berdasarkan hasil SUSENAS, masyarakat Bandung yang merokok rata-rata menghabiskan 64 – 65 batang rokok perminggunya atau 9 – 10 batang rokok perhari. Apabila diasumsikan harga satu pak rokok Rp 15.000,00, maka rata-rata masyarakat Bandung menghabiskan Rp 450.000,00 per bulannya untuk rokok. Pengeluaran sebesar itu setara dengan 25 kilogram beras, 5 kilogram telur ayam dan 6 liter susu cair.

Presifikasi Model



Gambar 2. Histogram Frekuensi Responden Berdasarkan Konsumsi Rokok Seminggu yang Lalu

Regresi Poisson merupakan regresi non linier untuk memodelkan kejadian acak yang peluang terjadinya relative kecil dalam selang waktu tertentu. Gambar di atas merupakan histogram dari jumlah responden berdasarkan jumlah batang rokok yang dikonsumsi selama seminggu yang lalu. Pada data konsumsi rokok, terdapat banyak amatan yang bernilai nol yaitu sebesar 62 persen. Berdasarkan fenomena di atas maka model awal yang digunakan adalah *Zero Inflated Poisson*.

Multikolinieritas

Salah satu hal yang harus dipastikan dalam regresi adalah tidak adanya multikolinieritas antar variabel prediktor. Hal ini untuk memastikan bahwa kita taksiran parameter akan *exist*. Pemeriksaan multikolinieritas dilakukan dengan melihat nilai VIF. Berikut adalah nilai VIF keempat variabel prediktor.

Tabel 1. Nilai VIF Variabel-variabel Prediktor

Variabel	X1	X2	X3	X4
VIF	1.095	1.106	1.091	1.111

Dari tabel di atas, dapat dilihat bahwa nilai VIF dari masing-masing variabel prediktor tidak lebih besar dari 10, sehingga dapat disimpulkan bahwa tidak terdapat kolineritas pada variabel prediktor.

Pemodelan Regresi *Zero Inflated Poisson* (ZIP) dan *Hurdle-Poisson*

Model ZIP maupun *Hurdle-NB* terbagi menjadi 2 model yaitu model Log untuk respon $y > 0$ dan model Logit untuk respon $y = 0$.

Berikut adalah model ZIP untuk konsumsi rokok penduduk Kota Bandung Tahun 2015.

Model Log :

$$\log(\hat{\mu}) = 3,45 + 0,16X_1 + 0,004X_2 + 0,28X_3 + 0,08X_4$$

$$\hat{\mu} = \exp(3,45 + 0,16X_1 + 0,09X_2 + 0,26X_3 + 0,08X_4)$$

Model Logit:

$$\text{logit}(\hat{\pi}) = 3,13 - 4,07X_1 + 0,007X_2 - 1,18X_3 + 0,61X_4$$

$$\frac{\hat{\pi}}{1 - \hat{\pi}} = \exp(3,13 - 4,07X_1 + 0,007X_2 - 1,18X_3 + 0,61X_4)$$

Sedangkan model konsumsi rokok dengan menggunakan regresi *Hurdle Poisson* adalah:

Model Log :

$$\log(\hat{\mu}) = 3,45 + 0,16X_1 + 0,004X_2 + 0,28X_3 + 0,08X_4$$

$$\hat{\mu} = \exp(3,45 + 0,16X_1 + 0,09X_2 + 0,26X_3 + 0,08X_4)$$

Model Logit :

$$\text{logit}(\hat{\pi}) = -3,13 + 4,07X_1 - 0,007X_2 + 1,18X_3 - 0,61X_4$$

$$\frac{\hat{\pi}}{1 - \hat{\pi}} = \exp(-3,13 + 4,07X_1 - 0,007X_2 + 1,18X_3 - 0,61X_4)$$

Pengecekan Overdispersi.

Pengecekan overdispersi dilakukan dengan melakukan uji skor (Yang, dkk, 2009). Nilai uji skor yang lebih besar dari $z_{0,10} = 1,65$ menunjukkan bahwa model terjadi overdispersi. Nilai uji skor untuk model ZIP adalah 30,95696 sedangkan model *Hurdle Poisson* adalah 30,95699. Karena nilai uji skor model ZIP maupun *Hurdle Poisson* lebih dari $z_{0,10} = 1,65$, maka dapat disimpulkan bahwa model ZIP maupun model *Hurdle-Poisson* terjadi overdispersi.

Pemodelan Regresi Zero Inflated Negatif Binomial dan Hurdle Negatif Binomial

Karena model ZIP dan Hurdle Poisson terjadi overdispersi, maka model yang lebih tepat digunakan adalah ZINB dan *Hurdle-NB*. Sama dengan model ZIP dan Hurdle Poisson, kedua model inipun terdiri dari dua model, yaitu model Log dan Model Logit

Berikut adalah model ZIP untuk konsumsi rokok penduduk Kota Bandung Tahun 2015.

Model Log :

$$\log(\hat{\mu}) = 3,45 + 0,16X_1 + 0,004X_2 + 0,28X_3 + 0,08X_4$$

$$\hat{\mu} = \exp(3,45 + 0,16X_1 + 0,09X_2 + 0,26X_3 + 0,08X_4)$$

Model Logit :

$$\text{logit}(\hat{\pi}) = 3,13 - 4,07X_1 + 0,007X_2 - 1,18X_3 + 0,61X_4$$

$$\frac{\hat{\pi}}{1 - \hat{\pi}} = \exp(3,13 - 4,07X_1 + 0,007X_2 - 1,18X_3 + 0,61X_4)$$

Sedangkan model konsumsi rokok dengan menggunakan regresi Hurdle Poisson adalah:

Model Log :

$$\log(\hat{\mu}) = 3,45 + 0,16X_1 + 0,004X_2 + 0,28X_3 + 0,08X_4$$

$$\hat{\mu} = \exp(3,45 + 0,16X_1 + 0,09X_2 + 0,26X_3 + 0,08X_4)$$

Model Logit :

$$\text{logit}(\hat{\pi}) = -3,13 + 4,07X_1 - 0,007X_2 + 1,18X_3 - 0,61X_4$$

$$\frac{\hat{\pi}}{1 - \hat{\pi}} = \exp(-3,13 + 4,07X_1 - 0,007X_2 + 1,18X_3 - 0,61X_4)$$

Pemilihan model

Pemilihan model dilakukan untuk melihat model mana yang lebih baik di antara model

ZINB dan Hurdle-NB. Statistik yang digunakan adalah AIC, dimana nilai AIC model yang lebih kecil menunjukkan model lebih baik dibanding model dengan AIC yang lebih besar. Berikut adalah nilai AIC kedua model.

Tabel 2. AIC Model ZINB dan Hurdle-NB

Model	AIC
ZINB	10.170,46
Hurdle-NB	10.170,44

Dari tabel di atas, model yang memiliki AIC lebih kecil adalah Hurdle-NB, sehingga model ini yang akan digunakan dalam pemodelan konsumsi rokok Penduduk Bandung Tahun 2015.

Pengujian Signifikansi Parameter Regresi Hurdle-Negatif Binomial (H-NB)

Pengujian signifikansi parameter digunakan untuk mengetahui variabel-variabel yang signifikan berpengaruh pada model Log ataupun pada model Logit. Pengujian signifikansi parameter dilakukan secara simultan maupun secara parsial.

Dengan menggunakan *Likelihood Ratio test*, didapat nilai G^2 sebesar 1373,3. Dengan tingkat signifikansi 0,05, didapat nilai $\chi^2_{0,05;8} = 15,51$. Karena nilai G^2 lebih besar dari nilai $\chi^2_{0,05;8}$ maka kita dapat menolak hipotesis nol. Sehingga dapat disimpulkan bahwa minimal terdapat satu variabel prediktor yang berpengaruh terhadap variabel respon.

Selanjutnya, karena secara simultan terdapat minimal satu variabel yang signifikan, selanjutnya ingin dilihat, variabel mana saja yang signifikan berpengaruh terdapat konsumsi batang rokok.

Tabel 3. Taksiran, *p-value*, dan Signifikansi Parameter-parameter Model Log dan Logit H-NB

Model	Parameter	Taksiran	<i>p-value</i>	Signifikansi
Log	β_0	3,47	2.e-16	Signifikan
	β_1 (jenis kelamin)	0,16	0,17	Tidak Signifikan
	β_2 (umur)	0,004	0,0104	Signifikan
	β_3 (status bekerja)	0,27	1.01e-05	Signifikan
	β_4 (pendidikan)	0,08	0,07	Signifikan
Logit	γ_0	-3,43	< 2.e-16	Signifikan
	γ_1 (jenis kelamin)	4,07	< 2.e-16	Signifikan
	γ_2 (umur)	-0,007	0,0598	Signifikan
	γ_3 (status bekerja)	1,18	< 2.e-16	Signifikan
	γ_4 (pendidikan)	-0,61	2.19e-05	Signifikan

Berdasarkan pengujian parsial, pada model Log Hurdle-NB variabel yang signifikan pada α 0,10 adalah umur, status bekerja, dan pendidikan, sedangkan jenis kelamin tidak signifikan berpengaruh terhadap banyak konsumsi rokok. Sedangkan pada model Logit Hurdle-NB, keempat variabel prediktor dalam model, yaitu jenis kelamin, umur, status bekerja, dan pendidikan, signifikan berpengaruh terhadap konsumsi rokok.

Interpretasi Parameter Regresi Hurdle-Negatif Binomial

a. Model Log Hurdle-NB

Model Log pada Hurdle-NB menjelaskan banyak konsumsi rokok yang dipengaruhi oleh variabel-variabel prediktor signifikan. Untuk variabel umur, setiap kenaikan umur satu tahun, maka konsumsi rokok setiap minggunya akan bertambah sebanyak 1,004 batang rokok. Artinya semakin meningkat usia seseorang, maka konsumsinya juga akan meningkat. Hal ini sejalan dengan analisis deskriptif sebelumnya.

Untuk variabel status bekerja, ketika status seseorang bekerja maka jumlah konsumsi rokok akan meningkat sebesar $\exp(0,27)=1,31$ batang. Artinya konsumsi rokok orang yang bekerja lebih banyak daripada konsumsi rokok orang yang tidak bekerja.

Untuk variabel pendidikan, ketika pendidikan seseorang minimal SMA sederajat, maka konsumsi rokok yang dihabiskan akan meningkat sebesar $\exp(0,08)=1,08$ batang. Artinya semakin tinggi tingkat pendidikan seseorang, malah menyebabkan konsumsi rokok semakin meningkat. Hal ini mungkin dikarenakan karena semakin tinggi pendidikan seseorang, maka penghasilannya juga cenderung meningkat, sehingga dengan meningkatnya pendapatan, seseorang lebih mungkin untuk mengkonsumsi rokok yang lebih banyak.

b. Model Logit Hurdle-NB

Model Logit Hurdle-NB menjelaskan peluang respon Y_i bernilai nol dipengaruhi oleh kelima variabel yang signifikan tersebut di atas. Untuk variabel jenis kelamin, peluang laki-laki untuk merokok adalah $\exp(4,07)=58,56$ kali lebih mungkin

dibanding tidak merokok. Dengan kata lain laki-laki lebih besar peluangnya untuk merokok daripada tidak merokok.

Untuk variabel umur, ketika umur seseorang naik satu tahun, maka peluang untuk merokok adalah $\exp(-0,007)=0,99$ kali lebih mungkin dibanding tidak merokok. Artinya kecenderungan seseorang untuk merokok atau tidak merokok lebih mungkin dimulai ketika usia yang lebih muda dibanding usia yang lebih tua. Hal ini sesuai fenomena sekarang yang mana kecenderungan kebiasaan untuk merokok atau tidak merokok dimulai saat usia muda, bukan pada usia yang lebih tua.

Untuk variabel status bekerja, ketika seseorang bekerja, maka peluangnya untuk tidak merokok adalah $\exp(-0,007)=0,99$ lebih mungkin dibanding untuk tidak. Dengan kata lain, seseorang yang bekerja lebih besar peluangnya untuk merokok daripada tidak merokok.

Untuk variabel pendidikan, ketika pendidikan seseorang minimal SMA sederajat, maka peluangnya untuk merokok adalah $\exp(-0,61)=0,54$ dibanding tidak merokok. Dengan kata lain, peluang seseorang yang memiliki jenjang pendidikan lebih tinggi peluang untuk tidak merokok lebih besar daripada merokok.

Kebaikan Model

Dengan menggunakan korelasi Pearson, diperoleh nilai koefisien determinasi R^2 model sebesar 0,33. Artinya 33 persen keragaman konsumsi rokok penduduk Kota Bandung dapat dijelaskan oleh variabel-variabel prediktor dalam model, sedangkan sisanya yaitu 0,67 persen dijelaskan oleh variabel-variabel lain di luar model. Nilai R^2 sebesar 0,33 mengindikasikan bahwa model Hurdle-NB dengan keempat variabel prediktor dalam model belum dapat menjelaskan konsumsi rokok dengan baik.

KESIMPULAN DAN SARAN

Berdasarkan hasil pemodelan diperoleh bahwa untuk model Logit variabel yang berpengaruh terhadap kecenderungan seseorang untuk merokok atau tidak yaitu variabel jenis kelamin, umur, status bekerja dan pendidikan. Sementara untuk model Log, variabel umur, status bekerja, dan

pendidikan berpengaruh terhadap rata-rata konsumsi rokok. Untuk penelitian selanjutnya, penelitian dengan menggunakan data primer lebih disarankan untuk dilakukan karena dapat memperoleh data-data individual lain terkait faktor penyebab perilaku merokok pada seseorang.

DAFTAR PUSTAKA

- Agresti, Alan. 2015. *Foundation of Linier and Generalized Linear Models*. Canada: John Wiley& Sons, Inc.
- Ambarwati, dkk. 2014. Media Leaflet, Video, dan Pengetahuan Siswa SD Tentang Bahaya Merokok (Studi pada Siswa SDN 78 Sabrang Lor Mojosongo Surakarta). *Jurnal Kesehatan Masyarakat* 10 (1): 7 – 13.
- Santi. 2013. Hubungan Pengetahuan Tentang Rokok dengan Sikap terhadap Bahaya Merokok pada Siswa SMK 1 Batik Surakarta. Fakultas Kedokteran Universitas Muhammadiyah Surakarta. Skripsi. Surakarta.
- Moghimbeigi, A, dkk. (2009). A Score Test for Zero-Inflation in Multilevel Count Data. *Computational Statistics and Data Analysis*, 53: 1239 – 1248.
- Wijayanti, Erlina, dkk. 2017. Faktor-faktor yang Berhubungan dengan Perilaku Merokok pada Remaja Kampung Bojong Rawalele, Jatimakmur, Bekasi. *Global Medical and Health Communication*: 5(3): 194 – 8.
- Cameron, A. Colin, & Trivedi, Pravin K. 1998. *Regression Analysis of Count Data*. New York: Cambridge University Press.
- Bain, L.J. dan Engelhardt, M. (1992). *Introduction to Probability and Mathematical Statistics Second Edition*. California: Duxbury Press
- Cantoni, E., and A. Zedini. 2010. A Robust Version of the Hurdle Model. *Journal of Statistical Planning and Inference*, Vol.141(3), pp:1214-1223.
- Gujarati, Damodar. 2004. *Basic Econometrics, Fourth Edition*. The McGraw-Hill Companies.
- Hestiana, Fida. 2013. Pemodelan *Mixture Count Regression* Pada Data *Zero Inflated* dengan Pendekatan Bayesian (Studi Kasus: Jumlah Batang Rokok Yang Dihisap Setiap Hari di Provinsi Sumatera Selatan Tahun 2010). Jakarta: Sekolah Tinggi Ilmu Statistik.
- Kristiono, Cahyo. 2010. Regresi ZINB untuk Pemodelan Data Respon *Count* dengan *Excess Zeros* (Studi Kasus: Konsumsi Miras pada Remaja). Bandung: Universitas Padjadjaran.
- Kutner, M. H., Nachtsheim, C. J., Neter, J. dan Li, W., 2005. *Applied Linear Regression Models, Fifth Edition*. Mc Graw Hill Companies, Inc.
- McCullagh, P. dan Nelder, J.A. (1989). *Generalized Linear Models*, 2nd Ed. New York: Chapman and Hall
- Pontoh, Septiani R dan Faidah, Defi Y. 2015. Penerapan Hurdle Negative Binomial pada Data Tersensor. Yogyakarta: UNY.
- Smet, Bart, dkk. 1999. Determinants of Smoking Behaviour Among Adolescent in Semarang, Indonesia. *Tabacco Control*; 8: 186 – 191.
- Widayanti, Citra Yanuar. 2017. Pemodelan Banyak Kematian Bayi dan Ibu di Provinsi Jawa Barat Menggunakan Regresi Binomial Negatif Bivariat. Bandung: *Tesis*. Bandung: Universitas Padjadjaran.
- Kementerian Kesehatan. 2010. Riset Kesehatan Dasar Riskesdas 2010. Jakarta: Badan Penelitian dan Pengembangan Kesehatan Kementerian Kesehatan RI.
- Kemeterian Kesehatan. 2017. Profil Kesehatan Indonesia Tahun 2016. Jakarta: Kemeterian Kesehatan RI.
- <https://sains.kompas.com/read/2018/01/05/070500823/kerugian-ekonomi-dari-konsumsi-rokok-indonesia-hampir-rp-600-triliun> (Diakses 24 Oktober 2018)
- <http://theconversation.com/riset-terbaru-kerugian-ekonomi-di-balik-konsumsi-rokok-di-indonesia-hampir-rp600-triliun-89089> (Diakses 24 Oktober 2018)
- <https://www.merdeka.com/uang/penelitian-pengeluaran-masyarakat-indonesia-untuk-beli-rokok-terus-naik.html> (Diakses 24 Oktober 2018)