International Journal of Social Science (IJSS)
Vol.1 Issue.3 October 2021, pp: 325-336
ISSN: 2798-3463 (Printed) | 2798-4079 (Online)
DOI: https://doi.org/10.53625/ijss.v1i3.5610

COMPARISON OF POISSON, ZIP, ZINB, HURDLE AND ZIGP REGRESSION ANALYSIS METHODS IN SCHOOL-AGED SMOKING CASE MODELING IN KUDUS DISTRICT, CENTRAL JAVA

By
Elok Fitriani Rafikasari[1], Suci Ismadyaliana[2], Dwilaksana Abdullah Rasyid[3], Setiawan[4]
[1]UIN Sayyid Ali Rahmatullah Tulungagung
[2,3,4]Institut Teknologi Sepuluh Nopember, Surabaya
Email: [1]elokfitriani@ymail.com, [2]suci@bps.go.id, [3]dwilaksana96@gmail.com, [4]setiawan@statistika.its.ac.id

Article Info
Article history:
Received July 11, 2021
Revised August 13, 2021
Received Sept 26, 2021

Keywords:
Cigarette Consumption, Poisson, ZIP, ZINB, Hurdle, ZIGP, Vuong

ABSTRACT
Smoking behavior among adolescents and children, especially boys, is increasing from time to time. This is very unfortunate considering the many harmful substances in cigarettes that can interfere with health. Modeling related to smoking cases for adolescents and school-age children is needed as a step to anticipate and deal with this problem. In this study we use the Poisson, ZIP, ZINB, Hurdle, and ZIGP regression methods to modeling the number of cigarettes consumed by adolescents and boys. The best model selection is done by the Vuong test. The results showed that the most suitable model was ZIGP with variables that had a significant effect on the amount of cigarette consumption in adolescents and children are age and education level

Corresponding Author:
Elok Fitriani Rafikasari
UIN Sayyid Ali Rahmatullah Tulungagung
Email: elokfitriani@ymail.com

1. INTRODUCTION

Smoking behavior has penetrated among teenagers and even children, especially boys. The average age of adolescents under 18 years is a senior high school student and below. Almost all schools apply disciplinary punishment to their students who are caught smoking. However, this punishment did not make them stop trying to smoke, they did it in secret. Results*Global Youth Tobacco Survey* (GYTS) in Indonesia in 2019 stated that 19.2 percent of students aged 13-15 years were found to be smoking and 22.6 percent of students aged 16-17 years were found to be smoking (Global Youth Tobacco Survey, 2020). According to the 2018 Riskesdas results, the prevalence of smoking aged 10-18 years was 9.1 percent, an increase compared to previous years (Health Research and Development Agency, 2018). BPS data shows that in 2018 there was an increase in the percentage of smoking among adolescents aged 10-18 years, but after that period it showed a decrease. The development of the percentage of adolescent smoking by age group can be seen in Figure 1.

International Journal of Social Science (IJSS)
Vol.1 Issue.3 October 2021, pp: 325-336
ISSN: 2798-3463 (Printed) | 2798-4079 (Online)

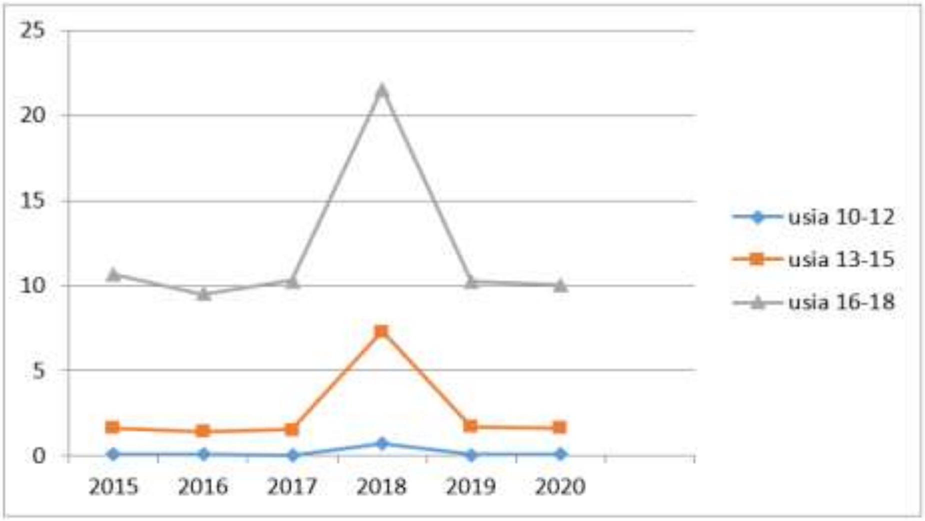

Figure 1. Percentage of Smoking in Population Age ≤ 18 Years
Source: Susenas KOR BPS (BPS, 2020) (processed)

Cigarettes contain substances that are harmful in addition to damaging health as well as damaging the morale of teenagers. Smoking behavior among adolescents is a threat to the future of the nation. The government seriously responds to these problems through government regulations. The use of cigarettes has been regulated in RI Government Regulation Number 109 of 2012. Cigarette product packaging must include "prohibited from selling or giving to children under 18 years of age and pregnant women". Decline in the prevalence of smoking in children and adolescents
is also one of the targets of the RPJMN, in 2024 it is targeted that the prevalence of smoking at the age of 10-18 years is 8.7 percent of*baseline* 9.1 percent. Several developed countries such as the United States, California and Singapore have raised the minimum age limit for buying cigarettes to 21 years. By increasing the minimum limit, it is hoped that it will reduce young smokers.

One of the causes of smoking in adolescence is the desire to experiment. Adolescence is a period of searching for identity psychologically unstable. At this time the influence of peers is very dominant. For male adolescents smoking is a symbol of maturity, strength, leadership and attractiveness to the opposite sex (Bringham, 1991 in Komasari and Helmi, 2000). Peers and family are the parties that cause adolescents to smoke. Peers who smoke will attract others to join in smoking for reasons of cohesiveness. The family is an example or role model for children. If there is someone in the family who smokes, the child will follow the example. McGee's research in 2015 stated that parents, siblings and peers who smoke are significant risk factors for adolescents smoking (McGee et al., 2015). Basically smoking behavior comes from within oneself (internal) and environmental factors (external). Internal factors can be in the form of physical and mental conditions of adolescents.

This study will model the number of cigarettes consumed by male adolescents. Factors that influence adolescent smoking are grouped into internal and external factors. In this study the internal factors to be examined consisted of age, level of education completed, and activities last week. While the external factors are the area of residence and the presence of parents who smoke. In adolescence, as you get older, the prevalence of smoking also increases. BPS data for 2020 shows that the percentage of the population aged 10-12 years who smoke is 0.13 percent, the population aged 13-15 years is 1.64 percent, and the population aged 16-18 years is 10.07 percent. Previous research conducted by Utami (2020) also showed that the addition of teenage age has a significant effect on smoking behavior. There is a tendency that school level influences smoking behavior. The percentage of high school level adolescents who smoke is more than that of junior high school and even elementary school. The past week's activities were grouped into work, school and other in addition to these two. This grouping aims to capture working youth. Adolescents who have worked tend to smoke higher than adolescents who are still at school. Area of residence is considered as an environmental factor that influences smoking behavior, where the urban environment is worse than the rural environment. Meanwhile, the existence of a smoking family becomes a role model for adolescent smoking behavior. Utami's research (2020) mentions parents who

International Journal of Social Science (IJSS)
Vol.1 Issue.3 October 2021, pp: 325-336
ISSN: 2798-3463 (Printed) | 2798-4079 (Online)
DOI: https://doi.org/10.53625/ijss.v1i3.5610

Smoking has a significant effect on smoking behavior in adolescents. Kudus Regency is a district that is the largest cigarette producer in Indonesia, so it was used as a research locus.

The incidence of smoking at school age is a rare event and the number of cigarette consumption data in sticks is the data*count*, so that it can be modeled by Poisson regression. An important characteristic of the Poisson distribution is that it has the same mean and variance (*equidsipersion*). In practice these conditions are rarely met, the variance is often greater than the average or called*overdispersion*. The implication of not fulfilling equidisperson is that Poisson regression will produce biased parameter estimates. One way to overcome this is with a generalized Poisson model. Zeileis et al (2008) stated that one way to overcome overdispersion is with a negative binomial model, which can appear as a Poisson gamma mixture distribution. Through the negative binomial model approach, the variance may not equal the mean. Previously Consul and Shoukri submitted a model*Generalized Poisson* to overcome the problem of overdispersion (Shoukri, 1984).

Because the portion of adolescents who do not smoke is greater, the number of cigarettes consumed is mostly 0 so that it can be said*excess zero*. models with*excess zero* can be overcome by the Zero-Inflated model (Zeileis et al., 2008).*Zero-inflated* The model can be applied to count data models such as Poisson, Negative Binomial and Geometric.*Zero-inflated* earlier models introduced by (Mullahy, 1986) and (Lambert, 1992). Poisson model with overdispersion conditions and*excess zero* can be solved with*Zero Inflated Generalized Poisson* (ZIGP) and*Zero Inflated* Negative Binomial (ZINB). Another alternative can use the hurdle model to overcome these two conditions. The Hurdle model was first introduced by Cragg (1971). Hurdle models differ from*zero-inflated* models that contain components*mixture*, *zero* and*non zero*. The Hurdle model combines the left-truncated count data model with the zero hardle model (Zeileis et al., 2008). This study will look at comparisons between

2. METHODS

Model Regresi Poisson

The Poisson regression model is a regression model in which the response variable has a Poisson distribution and is in the form of discrete data (Cahyandari, 2014). Example $Y \sim P(\mu)$ for *probability density function* for Y is:

$$f(y|\mu) = \begin{cases} \frac{\mu^y \, exp(-\mu)}{y!}, & y = 0,1,2,\cdots; \mu > 0 \\ 0 \quad , & \text{untuk yang lainnya} \end{cases}$$

$$E(Y) = \mu$$

$$Var(Y) = \mu$$

The model for Poisson regression is:

$$g(\mu_i) = \ln \mu_i = x_i^T \beta$$

$$\mu_i = g^{-1}(x_i^T \beta) = e^{\beta_0 + \sum_{j=1}^{k} \beta_j x_{ij}}$$

$$\mu_i = e^{x_i^T \beta}$$

Regression Models*Generalized Poisson* (GPR)

Generalized Poisson Regression is a Poisson regression model which assumes the random component is a Generalized Poisson distribution or it can be interpreted that the GPR is a Poisson regression model without assuming*equidisperse*(Ismail & Jemain, 2007). For example $Y_i \sim GP(\mu_i, \omega)$; $i = 1,2,3,\cdots,n$ for *probability density function* to Y_i is:

$$f(y_i|\mu_i,\omega) = \left[\frac{\mu_i}{1+\omega\mu_i}\right]^{y_i} \frac{(1+\omega y_i)^{y_i-1}}{y_i!} exp\left(-\frac{\mu_i(1+\omega y_i)}{1+\omega\mu_i}\right)$$

$$E(Y_i) = \mu_i$$

$$Var(Y_i) = \mu_i(1+\omega\mu_i)^2$$

with $\mu_i = e^{x_i^T \boldsymbol{\beta}}$; $\boldsymbol{x}_i^T$ is the predictor variable vector and $\boldsymbol{\beta}$ is the regression parameter vector. So the model for GPR is:

$ln(\mu_i) = \boldsymbol{x}_i^T \boldsymbol{\beta} = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki}$

$\mu_i = exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki})$ (Girik Allo et al., 2019).

Regression Models*Zero Inflated Poisson* (ZIP)

Regression models*Zero Inflated Poisson* (ZIP) is used to perform analysis on observations that are mostly zero. Example y_i independent random variable ZIP distribution, the zero value is thought to appear because*zero state* with

International Journal of Social Science (IJSS)
Vol.1 Issue.3 October 2021, pp: 325-336
ISSN: 2798-3463 (Printed) | 2798-4079 (Online)

probability ω_i and only produces zero-value observations. In addition, the zero value is also thought to appear because*Poisson state* with probability $(1-\omega_i)$ and distribute*Poisson* with average μ_i. Pdf for y_i is:

$$P(Y=y_i)=\begin{cases}\omega_i+(1-\omega_i)e^{-\mu_i};\ y_i=0\\ \dfrac{(1-\omega_i)e^{-\mu_i}\mu_i^{y_i}}{y_i!};\ y_i>0; 0\le\omega_i\le 1\end{cases}$$

or can be denoted by $Y_i \sim ZIP(\mu,\omega)$ (Jansakul & Hinde, 2002). Relationship between μ and ω is $\log(\mu)=X\beta$ and $logit(\omega)=\log\left(\frac{\omega}{1-\omega}\right)=X\gamma$ dimana X is a variable matrix consists of a set of experimental factors related to the probability at*zero state* and*mean Poisson* on *Poisson state* β and γ is the regression parameter to be estimated.

Regression Models*Zero Inflated Generalized Poisson* (ZIGP)

The introduction of the ZIGP regression model has been done by Famoye and Singh (2003). Distribution on three parameters*m*, $\phi$and*oh*can be written*AND~ ZIGP(*m ϕ o, *,)* with*probability density function* (pdf) as follows:

$$P(Y=y|\mu,\varphi,\omega)=\begin{cases}\omega+(1-\omega)e^{-\frac{\mu}{\varphi}}, \text{for } y=0\\ (1-\omega)\dfrac{\mu(\mu+(\varphi-1)y)^{y-1}}{y!}\varphi^{-y}e^{-\frac{1}{\varphi}(\mu+(\varphi-1)y)},\ \text{for } y>0\end{cases}$$

Dimana $\varphi>\max\left(\frac{1}{2},1-\frac{\mu}{m}\right)$, *m* adalah nilai maksimum dengan $\mu+m(\varphi-1)>0$ jika Rata-rata dan variansi dari model ZIGP adalah sebagai berikut (Famoye & Singh, 2021):

$$E(Y_i \mid X=x_i)=(1-\omega)\mu_i$$

$$Var(Y_i \mid X=x_i)=E(Y_i \mid X=x_i)(\varphi^2+\mu_i\omega)$$

The model introduced by Famoye and Singh (2003) was later developed by Czado (2007) to become the ZIGP regression model (μ_i,φ_i,ω_i) which has three parameters, the parameter mean (m), parameter overdispersion (φ) and parameters*zero-inflation* (ω) (Czado et al., 2007). Then each parameter in the ZIGP distribution has a linear relationship to the ZIGP regression parameters in the following equation:

1. Average

$$E(Y_i \mid \beta)=\mu_i(\beta):=E_ie^{x_i^t\beta}=E_ie^{x_i^t\beta+\log(E_i)}>0$$

2. Overdispersi

$$\varphi_i(\alpha):=1+e^{\omega_i^t\alpha}>1$$

3. Zero-inflation

$$\omega_i(\gamma):=\frac{e^{z_i^t\gamma}}{1+e^{z_i^t\gamma}}\in(0,1)$$

Regression Models*Zero Inflated Negative Binomial* (ZINB)

Regression models*Zero Inflated Negative Binomial* (ZINB) is a model formed from a mixed Poisson Gamma distribution. This model can be used to model discrete data with many zero values in the response variable (*zero inflation*) and overdispersion occurs (Garay et al., 2011). if y_i is random variable with $i=1,2,3,\cdots,n$ then the value of the response variable occurs in two situations. The first condition is mentioned*zero state* and generates only zero-valued observations, while the second state is called*negative binomial state* which has Negative Binomial distribution.*Probability density function* (pdf) ZINB is as follows**:**

International Journal of Social Science (IJSS)
Vol.1 Issue.3 October 2021, pp: 325-336
ISSN: 2798-3463 (Printed) | 2798-4079 (Online)
DOI: https://doi.org/10.53625/ijss.v1i3.5610

$$P(Y_i = y_i) = \begin{cases} \pi_i + (1-\pi_i)\left(\frac{1}{1+\kappa\mu_i}\right)^{\frac{1}{k}}, \text{ for } y_i{=}0 \\ (1-\pi_i)\frac{\Gamma\left(y_i+\frac{1}{\kappa}\right)}{\Gamma\left(\frac{1}{\kappa}\right)y!}\left(\frac{1}{1+\kappa\mu_i}\right)^{\frac{1}{k}}\left(\frac{\kappa\mu_i}{1+\kappa\mu_i}\right)^{y_i}, \text{ for } y_i{>}0 \end{cases}$$

Which $0 \le \pi_i \le 1, \mu_i \ge 0$, κ is the dispersion parameter, $\Gamma(.)$ is the gamma function. When $\pi_i = 0$, random variable y_i have a spread *Negative Binomial* with average μ_i and dispersion parameters κ. The ZINB regression model can be expressed in the model*negative binomial state* and*zero inflation*. Models for*negative binomial state* $\hat{\mu}_i$ is $\ln \hat{\mu}_i = \hat{\beta}_0 + \sum_{j=0}^{p} \hat{\beta}_j x_{ij}$, $i = 1,2,...,n$ dan $j = 1,2,..., p$. Models for*zero inflation* $\hat{\pi}_i$ is $\text{logit } \hat{\pi}_i = \hat{\gamma}_0 + \sum_{j=1}^{p} \hat{\gamma}_j x_{ij}$, $i = 1,2,...,n$ dan $j = 1,2,..., p$.

Model *Hurdle*

Model *Hurdle* is a model introduced by John G. Cragg in which a random variable is modeled using two parts, namely those containing zero probabilities and non-zero values (*non zero*) or can be written as:

$$P(x = 0) = \theta$$
$$P(x \neq 0) = p_{x\neq0}(x)$$

dimana $p_{x\neq0}(x)$ is a function*truncated probability distribution* at 0. Value*non zero* modeled using the Normal model and zero values modeled with the Probit model (Cragg, 1971). Next models*Hurdle* developed for*count data* with Poisson, Geometric, and Negative Binomial models for*non-zero count* (Mullahy, 1986).

Overdispersion Test

The poisson regression model requires*equidisperse*, namely the condition in which the mean and variance of the response variable are the same. However, sometimes phenomena occur*overdispersion*in data modeled by Poisson distribution. Condition*overdispersion*is a condition where the avariance is greater than the mean which indicates that the model is not suitable for the data. The procedure to be carried out for the test is as follows.

1. Formulation of the hypothesis
   H_0 : There is no overdispersion
   H_1 : There is overdispersion
2. Test Statistic
   $$S = \frac{\sum_{i=1}^{n}\{(y_i-\mu_i)^2-\mu_i\}}{\{2\sum_{i=1}^{n}\mu_i^2\}^{\frac{1}{2}}}$$
3. Test Criteria
   The significance level is α, then reject H_0 if $|S| > Z_{\alpha/2}$
4. Conclusion
   IF H_0 is Rejected, then there is overdispersion in the model, so it can be said that the model is not appropriate (Cahyandari, 2014).

Uji Vuong

Vuong's test aims to find out which model is better. For example*($y_i|x_i$)* is the probability predicted from the fourth observation*i*, so it can be defined as follows (Vuong, 1989).

$$m_i = \log\left(\frac{f_1(y|x)}{f_2(y|x)}\right)$$

International Journal of Social Science (IJSS)
Vol.1 Issue.3 October 2021, pp: 325-336
ISSN: 2798-3463 (Printed) | 2798-4079 (Online)

With the following hypothesis:
H_0 : $E(m_i) = 0$ (Model 1 and model 2 are equally good)
H_1 : $E(m_i) > 0$ (Model 1 is better than model 2)
The test statistics used are as follows:

$$V = \frac{\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} m_i\right)}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(m_i - m\right)^2}} \sim N(0,1)$$

Reject H_0 if value $V > Z_\alpha$ or value $V < p\text{-}value$, which shows that model 1 is better than model 2.

Data Sources and Variables

Data Sources and Variables

Data sourced from the 2020 National Socioeconomic Survey (Susenas) KOR BPS Kudus Regency. The research sample is a subset of the Susenas KOR sample for the Regency Kudus, with criteria of age 13-18 years and male gender. The variables used are summarized in Table 1.

Table 1. Research Variables

Variabel	Notasi	Definisi	Skala
Dependen	Tottaly cigarete	Tottaly consumed during a week	Ratio 0,1, 2, …
Independen	Age	Age Respondent's (13-18 years)	Rasio 13, 14, …, 18
	P	Education level terminated	Catagorical 0 : to elementary school equivalent 1 : junior high school equivalent 2 : High school equivalent
	K	Most activities during last week	Catagorical 0 : Other 1 : Work 2 : School
	TT	Status of residence	Catagorical 0 : Village 1 : City
	Parents Smoking	Condition of parents at home	Catagorical 0 : do not smoke1 1 : smoking

3. RESULTS AND DISCUSSION

Data Expolorer

The number of research respondents was 114 people, with smoking conditions as many as 16 people and 98 people who did not smoke. Respondents aged 13-15 years were 57 people, and those aged 16-18 years were also 57 people. Of the respondents who smoked, there was 1 person aged 13-15 years, and the rest aged 16-18 years. The amount of cigarette consumption by age of the respondents can be seen in Table 2.

International Journal of Social Science (IJSS)
Vol.1 Issue.3 October 2021, pp: 325-336
ISSN: 2798-3463 (Printed) | 2798-4079 (Online)
DOI: https://doi.org/10.53625/ijss.v1i3.5610

Table 2. Number of Cigarette Consumption by Age

Tottaly of Cigarette	Age		Totally
	13-15	16-18	
0	56	42	98
1-25	1	3	4
26-50	0	6	6
>50	0	6	6
Total	57	57	114

Source: Susenas KOR, BPS Kabupaten Kudus (processed)

At the age of adolescents, the higher the education level, the more cigarette consumption is suspected. Respondents with the highest elementary school diploma were 47 people, 58 people with the highest junior high school diploma, and 9 people with the highest high school diploma. Respondents with the highest elementary school diploma means that if they do not drop out of school, the respondent is at the junior high school level. Respondents with the highest junior high school diploma, meaning that if they did not drop out of school, the respondent was at the high school level. Respondents with the highest high school diploma mean that they have completed high school education. From the research sample, it was found that there were 4 smokers with the highest elementary school diploma, 7 smokers with the highest junior high school diploma, and 6 smokers with the highest high school diploma. The amount of cigarette consumption according to the highest level of education completed can be seen in Table 3.

Tabel 3. Jumlah Konsumsi Rokok menurut Pendidikan Tertinggi yang Ditamatkan

Jumlah Batang Rokok	Pendidikan			Total
	sd. SD	SMP	SMA	
0	43	51	4	98
1-25	2	1	1	4
26-50	0	3	3	6
>50	2	3	1	6
Total	47	58	9	114

Sumber: Susenas KOR, BPS Kabupaten Kudus (diolah)

Respondents with activities a week ago were working as many as 15 people, activities a week ago going to school were 89 people, and other activities besides work and school were 10 people. Respondents who worked and smoked were 9 people, who attended school and smoked were 2 people, and in other activities (not working & not going to school) and smoking were 5 people. The amount of cigarette consumption according to activity a week ago can be seen in Table 4.

Table 4. Number of Cigarette Consumption by Activities in the Last Week

Tottally Cigarette	Last week's activities			Total
	Work	School	Other	
0	6	87	5	98
1-25	1	0	3	4
26-50	2	2	2	6
>50	6	0	0	6
Total	15	89	10	114

Sumber: Susenas KOR, BPS Kabupaten Kudus (diolah)

Living environment is thought to influence smoking behavior in adolescents. There were 18 respondents who lived in villages and 96 people who lived in cities. From the research sample obtained 5 people who live in villages

International Journal of Social Science (IJSS)
Vol.1 Issue.3 October 2021, pp: 325-336
ISSN: 2798-3463 (Printed) | 2798-4079 (Online)

and smoke, and 11 people who live in cities and smoke. The amount of cigarette consumption by residence status can be seen in Table 5.

Table 5. Number of Cigarette Consumption by Residence Status

Totally Cigarette	Status of Residence		
	Village	City	Total
0	13	85	98
1-25	0	4	4
26-50	1	5	6
>50	4	2	6
Total	18	96	114

Source: Susenas KOR, BPS Kabupaten Kudus (processed)

The condition of parents who smoke is considered as one of the triggers for adolescents to smoke. The research sample showed that 63 people with parental conditions did not smoke, and 51 people with smoking parents. From the group whose parents did not smoke, 5 of them were teenagers who smoked. Meanwhile, from the group whose parents smoked, 11 of them also smoked. The amount of cigarette consumption according to the condition of the parents can be seen in Table 6.

Table 6. Number of Cigarette Consumption according to Parents' Conditions

Tottally cigarette	Parents condition		
	No Smooking	Smooking	Total
0	58	40	98
1-25	1	3	4
26-50	3	3	6
>50	1	5	6
Total	63	51	114

Data on the number of cigarette consumption for a week is in the range of 0 - 168 cigarettes. The smallest number of respondents who smoked cigarettes was 6 cigarettes a week, while the most was 168 cigarettes a week. The distribution of data on the number of respondents' cigarette consumption can be seen in the following figure.

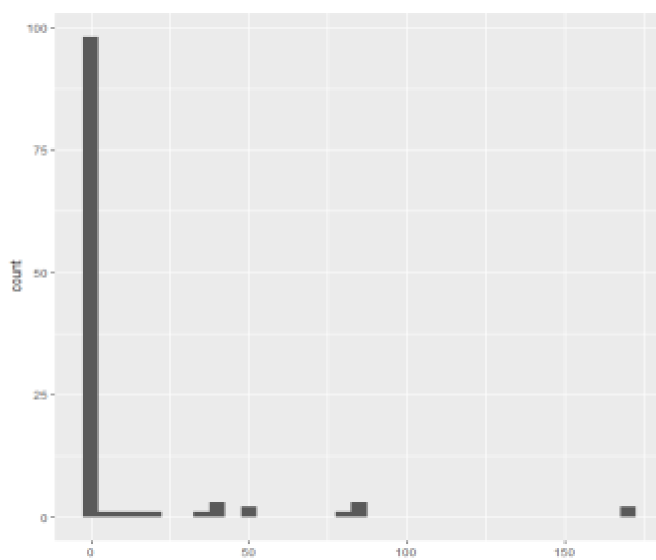

Figure 1. Distribution of Respondents' Total Cigarette Consumption

Source: Susenas KOR, BPS Kabupaten Kudus (processed)

International Journal of Social Science (IJSS)
Vol.1 Issue.3 October 2021, pp: 325-336
ISSN: 2798-3463 (Printed) | 2798-4079 (Online)
DOI: https://doi.org/10.53625/ijss.v1i3.5610

Cigarette Consumption Models

Before modeling, it is first tested whether the Poisson model complies*equidispertion* or not. If not meet*equidispertion*, then the resulting parameter estimates will be biased. The overdispersion test results obtained a Z value of 1.7048 with a p-value of 0.04411, so it can be concluded that there is overdispersion in the Poisson model. Furthermore, modeling is carried out that can overcome the conditions*excess zero* and overdispersion namely the ZINB model,*Hurdle* and ZIGP. Comparison between models can be seen in Table 7 below.

Table 7. Comparison between Models

Independen	Poisson		ZIP		ZINB		Hurdle		ZIGP	
	Estimate	S.E	Estimate	S.E	Estimate	S.E	Estimate	S.E	Estimate	S.E
Interc	-7.063**	0.769	0.658	0.893	-1.350	3.655	-1.284	3.602	-5.564**	1.673
Umur	0.615**	0.047	0.185**	0.053	0.312	0.221	0.308	0.218	0.639**	0.094
P1	-1.487**	0.089	-0.945**	0.086	-0.731*	0.384	-0.728*	0.382	-1.216**	0.172
P2	-1.631**	0.106	-0.867**	0.105	-0.687	0.522	-0.686	0.520	-1.959**	0.205
K1	1.240**	0.107	1.274**	0.116	1.229**	0.485	1.223**	0.481	-0.165	0.173
K2	-1.537**	0.158	1.192**	0.163	1.370**	0.634	1.359**	0.627	1.727	1.866
TT	-0.694**	0.074	-0.234**	0.086	-0.315	0.384	-0.316	0.383	-0.167	0.172
Ortu rokok	0.968**	0.073	0.055**	0.085	-0.312	0.385	-0.310	0.384	-	-
Log Likelihood			-156.3		-93.64		-93.66		-88.4	
MSE	23.61355		217.4405		296.5412		295.6325		13864.73	
AIC	1406.6		344.6438		221.2888		221.3153		209	

*significant at the 10% level
**significant at the 5% level

The Poisson model yields all significant independent variables, the lowest MSE value, and the highest AIC value among other models. However, this Poisson model produces biased estimates. ZIP models accommodate*excess zero*, but does not accommodate overdispersion. The ZIP model also produces all significant independent variables except*intercept*, and smaller AIC values compared to the Poisson model. but larger than the ZINB model,*Hurdle* and ZIGP.

Model ZINB, *Hurdle*, and ZIGP has accommodated the conditions*excess zero* and overdispersion. From the comparison of the three models, the ZIGP model has the smallest AIC value and the highest likelihood value. Even so, the ZIGP model MSE is the largest of the three. To ensure the selection of the best model, the Vuong test is used. Vuong test results as follows.

Table 8. Vuong Test Results

Test Model	Hipotesis alternative	p-value	Conclusion
ZINB/ZIP	Model ZINB is better than ZIP	0.0052	Model ZINB is better than ZIP
Hurdle/ZIP	Model Hurdle is better than ZIP	0.0052	Model Hurdle is better than ZIP
ZIGP/ZIP	Model ZIGP is better than ZIP	0.0373	Model ZIGP is better than ZIP
ZINB/Hurdle	Model ZINB is better than Hurdle	0.365	ZINB dan Hurdle are equally good
ZIGP/ZINB	Model ZIGP is better than ZINB	0.0000	ZIGP is better than ZINB

The Vuong test results above show that the ZINB, Hurdle, and ZIGP models are better than the ZIP models. A comparison between the ZINB and Hurdle models shows that both models are equally good, so a comparison with the ZIGP model can use only one. Comparison between the ZIGP and ZINB models shows that the ZIGP model is better. So it can be concluded that based on the results of the Vuong test the best model is the ZIGP model.

Model Interpretation

Modeling the amount of cigarette consumption in adolescents in Kudus Regency with conditions*excess zero* and the most appropriate overdispersion is to use the ZIGP model. The results of the estimation of the average model parameters, the zero inflation model and the overdispersion model can be seen in Table 9.

Table 9. ZIGP Model Estimation Results

		Estimate	Std. Error	z value	Pr(>|z|)	
MU REGRESSION						
b0	Intercept	-5.56372	1.67301	-3.3256	0.0009	***
b1	GLM$Umur	0.63859	0.09378	6.8096	9.79e-12	***
b2	GLM$P1	-1.21614	0.17175	-7.0809	1.43e-12	***
b3	GLM$P2	-1.95907	0.20545	-9.5356	<2e-16	***
b4	GLM$K1	-0.16461	0.17340	-0.9493	0.3425	
b5	GLM$K2	1.72741	1.86611	0.9257	0.3546	
b6	GLM$TT	-0.16748	0.17177	-0.9750	0.3295	
PHI REGRESSION						
a0	Intercept	-14.60171	388334.43152	-0.00004	1.0000	
a1	GLM$K1	-2.56892	0.67133	-3.82658	0.0001	***
a2	GLM$K2	4.30253	2.11162	2.03755	0.0416	*
a3	GLM$`ortu merokok`	18.37599	388334.43152	0.00005	1.0000	
OMEGA REGRESSION						
g0	Intercept	-10.21769	11.88198	-0.85993	0.3898	
g1	GLM$Umur	0.56750	0.65777	0.86277	0.3883	
g2	GLM$K1	1.25459	1.42633	0.87959	0.3791	
g3	GLM$K2	11.86699	19.98074	0.59392	0.5526	
g4	GLM$`ortu merokok`	-20.38340	8392.47502	-0.00243	0.9981	

The results of testing the model parameters show that age and education level have a significant effect on the amount of cigarette consumption. The higher the age of adolescents, the tendency to consume cigarettes will be more and more. For every additional age of 1 year, the tendency to consume cigarettes will increase by 1.89 times. Teenagers with junior high school graduates tend to smoke less than those with elementary school graduates. Teenagers with high school graduation tended to smoke less than those with elementary school graduates.*Covariate* what causes the overdispersion of the model is the activity during the past week. While the covariate in the ZIGP zero inflation model is*covariate* which is significant in the model*zero inflation* ZINB and*Hurdle*. on models*zero inflation* ZIGP *covariate* it is not significant.

4. CONCLUSIONS AND RECOMMENDATIONS

From the results of the study it can be concluded that the best model for modeling the amount of cigarette consumption among adolescents in Kudus Regency is the ZIGP model. The results of the model parameter test show that age and education level have a significant effect on the amount of cigarette consumption.*Covariate* what causes the overdispersion of the model is the activity during the past week.*Covariate* on models*zero inflation* ZIGP nothing significant, even on the ZINB and*Hurdle* significant. Suggestions for further research are to test which ZIGP type I and II models provide the best model, as well as add independent variables that are thought to have a strong influence on adolescent smoking behavior.

REFERENCES

[1] Badan Litbang Kesehatan, K. K. R. (2018). Laporan_Nasional_RKD2018_FINAL.pdf. In *Badan Penelitian dan Pengembangan Kesehatan* (p. 198). http://labdata.litbang.kemkes.go.id/images/download/laporan/RKD/2018/Laporan_Nasional_RKD2018_FINAL.pdf

[2] BPS. (2020). *Susenas Kor 2020*. 26.

[3] Cahyandari, R. (2014). Pengujian Overdispersi pada Model Regresi Poisson. *Statistika*, *14*(2), 69–76. https://ejournal.unisba.ac.id/index.php/statistika/article/view/1204

[4] Cragg, J. G. (1971). *This content downloaded from 77.103.16.28 on Wed*. *39*(5), 829–844.

[5] Czado, C., Erhardt, V., Min, A., & Wagner, S. (2007). Zero-inflated generalized Poisson models with regression effects on the mean, dispersion and zero-inflation level applied to patent outsourcing rates. *Statistical Modelling*, *7*(2), 125–153. https://doi.org/10.1177/1471082X0700700202

[6] Famoye, F., & Singh, K. P. (2021). Zero-Inflated Generalized Poisson Regression Model with an Application to Domestic Violence Data. *Journal of Data Science*, *4*(1), 117–130. https://doi.org/10.6339/jds.2006.04(1).257

[7] Garay, A. M., Hashimoto, E. M., Ortega, E. M. M., & Lachos, V. H. (2011). On estimation and influence diagnostics for zero-inflated negative binomial regression models. *Computational Statistics and Data Analysis*, *55*(3), 1304–1318. https://doi.org/10.1016/j.csda.2010.09.019

[8] Girik Allo, C. B., Otok, B. W., & Purhadi. (2019). Estimation Parameter of Generalized Poisson Regression Model Using Generalized Method of Moments and Its Application. *IOP Conference Series: Materials Science and Engineering*, *546*(5). https://doi.org/10.1088/1757-899X/546/5/052050

[9] Global Youth Tobacco Survey. (2020). Lembar Informasi Indonesia 2019 (Global Youth Tobacco Survei). *World Health Organization*, 1–2. https://www.who.int/indonesia/news/events/world-no-tobacco-day-2020

[10] Ismail, N., & Jemain, A. A. (2007). Handling Overdispersion with Negative Binomial and Generalized Poisson Regression Models Casualty Actuarial Society Forum , Winter 2007. *Society*, 103–158. http://www.casualtyactuaries.com/pubs/forum/07wforum/07w109.pdf

[11] Jansakul, N., & Hinde, J. P. (2002). Score test for zero-inflated poisson models. *Computational Statistics and Data Analysis*, *40*(1), 75–96. https://doi.org/10.1016/S0167-9473(01)00104-9

[12] Lambert, D. (1992). Zero-inflated poisson regression, with an application to defects in manufacturing. *Technometrics*, *34*(1), 1–14. https://doi.org/10.1080/00401706.1992.10485228

[13] McGee, C. E., Trigwell, J., Fairclough, S. J., Murphy, R. C., Porcellato, L., Ussher, M., & Foweather, L. (2015). Influence of family and friend smoking on intentions to smoke and smoking-related attitudes and refusal self-efficacy among 9-10 year old children from deprived neighbourhoods: A cross-sectional study. *BMC Public Health*, *15*(1), 15–21. https://doi.org/10.1186/s12889-015-1513-z

[14] Mullahy, J. (1986). Specification and testing of some modified count data models. *Journal of Econometrics*, *33*(3), 341–365. https://doi.org/10.1016/0304-4076(86)90002-3

[15] Utami, N. (2020). Pengaruh Kebiasaan Merokok Orang Tua terhadap Perilaku Merokok Remaja Indonesia. Media Kesehatan Indonesia, 16(3), 327-335. DOI : 10.30597/mkmi.v16i3.9801

[16] Shoukri, M. M. (1984). Maximum likelihood estimation for the generalized poisson distribution. *Communications in Statistics - Theory and Methods*, *13*(12), 1533–1547. https://doi.org/10.1080/03610928408828776

[17] Zeileis, A., Kleiber, C., & Jackman, S. (2008). Regression models for count data in R. *Journal of Statistical Software*, *27*(8), 1–25. https://doi.org/10.18637/jss.v027.i08

International Journal of Social Science (IJSS)
Vol.1 Issue.3 October 2021, pp: 325-336
ISSN: 2798-3463 (Printed) | 2798-4079 (Online)

THIS PAGE HAS INTENTIONALLY BEEN LEFT BLANK