Available online at https://journal.rescollacomm.com/index.php/ijbesd/index

International Journal of Business, Economics and Social Development

e-ISSN 2722-1156
p-ISSN 27722-1164

Vol. 6, No. 1, pp. 73-79, 2025

Calculation of Aggregate Accident Claims Based on the Number and Magnitude of Accident Losses in Indonesia from 2013 to 2022

Roy Donald Pangeran Siahaan[1*]

[1]*Universitas Padjajaran, Kabupaten Sumedang, Indonesia*

**Corresponding author email: roy21001@mail.unpad.ac.id*

Abstract

Accident losses are unavoidable when accidents occur. Data from the Indonesian Central Bureau of Statistics (BPS) show an increasing trend in accident claim amounts in recent years. This study aims to calculate aggregate accident claims based on the number and magnitude of accident losses in Indonesia from 2013 to 2022, enabling insurance companies or entities covering such losses to estimate the total liability required for all accident claims. The Generalized Extreme Value (GEV) distribution was applied to model claim amounts, while the negative binomial distribution was used for the number of accidents. The results show that the expected value of aggregate claims is IDR 236,768,463,076.93, with a variance of 7.28×10^{20}. The findings indicate that these distributions are suitable for modeling aggregate claims, providing insights for determining appropriate insurance premiums.

Keywords: accident losses, aggregate claims, GEV distribution, negative binomial distribution.

1. Introduction

Accidents are undesirable events that can cause material and non-material losses to victims. Therefore, insurance is necessary to cover the losses resulting from such accidents. According to data from the Indonesian Central Bureau of Statistics, the number of accidents and the magnitude of accident losses have shown an increasing trend, reaching a peak in 2022.

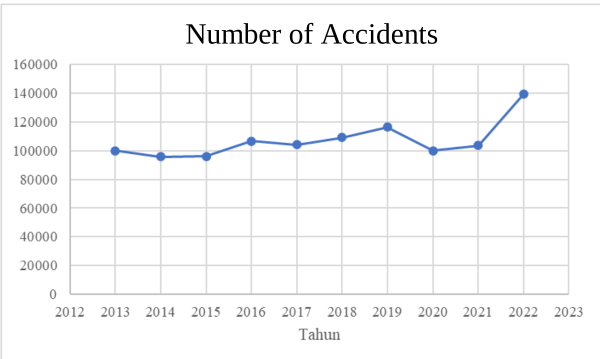

Figure 1: Scatter Plot of the Number of Accidents from 2013 to 2022

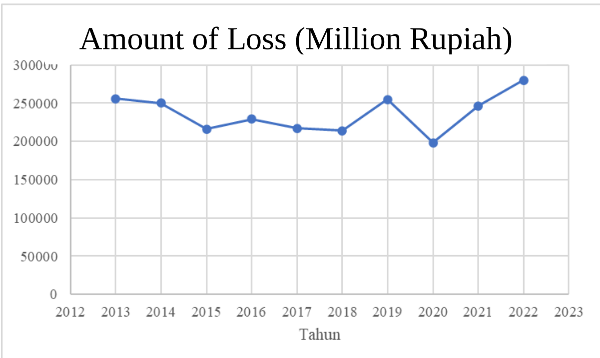

Figure 2: Scatter Plot of Accident Loss Magnitudes from 2013 to 2022

Given this trend, it is crucial for insurance companies or other entities covering accident claims to estimate the potential magnitude of accident losses to prepare and manage their finances effectively. This study aims to calculate aggregate accident claims based on the number and magnitude of accident losses in Indonesia from 2013 to 2022. The magnitude of losses and the number of claims will be modeled using appropriate distributions to calculate aggregate accident claims.

This study introduces a novel approach compared to previous research by employing the Generalized Extreme Value (GEV) distribution and the negative binomial distribution in its calculations.

Table 1: *Research Gap or Content Analysis*

Author	Title	GEV Distribution	Binomial Negative Distribution
Amaliah, Siswanah, Miasary (2019)	Analysis of Aggregate Claims with Negative Binomial and Discrete Uniform Distributions Using Convolution Method	No	Yes
Putra, Lesmana, Purnaba (2021)	Calculation of Motor Vehicle Insurance Premiums Using Generalized Linear Models with Tweedie Distribution	No	No
Yulita, Patricia, Hidayat (2024)	Determination of Pure Premiums from Four-Wheel Vehicle Insurance Claims with Comprehensive Coverage	No	Yes
Siahaan (2024)	Modeling Aggregate Accident Claims Based on the Number and Magnitude of Accident Losses in Indonesia from 2013 to 2022	Yes	Yes

2. Literature Review

2.1. Aggregate Claims

Aggregate claims refer to the total claims from a risk over a specific period (Dickson, 2006). Aggregate claims are the sum of all individual claims. If N represents the number of claims for a risk over a period and X represents the random variable for claim amounts, aggregate claims can be calculated using the following equation (1)

$$S = \sum_{i=1}^{N} X_i \tag{1}$$

Significant measures in determining aggregate claims include the expected value and variance. The expected value of aggregate claims can be calculated using equation (2) (Yulita et al., 2024).

$$E(S) = E(X) \cdot E(N) \tag{2}$$

where,
$E(S)$: Expected value of aggregate claims S,
$E(X)$: Expected value of claim amounts,
$E(N)$: Expected value of claim frequency.

The variance of aggregate claims can be calculated using equation (3)

$$V(S) = E(N) \cdot V(X) + V(N) \cdot E(X)^2 \tag{3}$$

where,
$V(S)$: Variance of aggregate claims S,
$V(X)$: Variance of claim amounts,
$V(N)$: Variance of claim frequency.

2.2. Generalized Extreme Value Distribution

Generalized Extreme Value (GEV) distribution is a continuous distribution that combines the Gumbel, Fréchet, and Weibull distributions. It was first introduced by Jenkinson (1955). The cumulative distribution function (CDF) of the GEV distribution is given by equation (4) (Rahayu, 2012).

$$F(x;\mu,\sigma,\xi) = \begin{cases} exp\left\{-\left(1+\xi\left(\frac{x-\mu}{\sigma}\right)\right)^{-\frac{1}{\xi}}\right\}, -\infty < x < \infty, \xi \neq 0, 1+\xi\left(\frac{x-\mu}{\sigma}\right) > 1 \\ exp\left\{-exp\left(\frac{x-\mu}{\sigma}\right)\right\}-, -\infty < x < \infty, \xi = 0 \end{cases} \tag{4}$$

where:
μ : location parameter,
σ : scale parameter,
ξ : shape parameter,

The probability density function (PDF) of the GEV distribution can be derived from the CDF and is given by equation (5).

$$f(x;\mu,\sigma,\xi) = \begin{cases} \frac{1}{\sigma}\left\{\left(1+\xi\left(\frac{x-\mu}{\sigma}\right)\right)^{-\frac{1}{\xi}} exp\left\{-\left(1+\xi\left(\frac{x-\mu}{\sigma}\right)\right)^{-\frac{1}{\xi}}\right\}\right\}, \xi \neq 0 \\ \frac{1}{\sigma}\left\{exp\left(\frac{x-\mu}{\sigma}\right)\right\} exp\left\{-exp\left(\frac{x-\mu}{\sigma}\right)\right\}, \xi = 0 \end{cases} \tag{5}$$

The expected value of the GEV distribution is given by equation (6) (Novaes, 2022).

$$E(X) = \begin{cases} \mu + \sigma\left(\frac{\Gamma(1-\xi)-1}{\xi}\right), \xi \neq 0, \\ \mu - \sigma\gamma, \xi = 0 \end{cases} \tag{6}$$

where γ is the Euler-Mascheroni constant.

The variance of the GEV distribution is given by equation (7)

$$Var(X) = \sigma^2[\Gamma(1-2\xi) - \Gamma(1-\xi)^2] \tag{7}$$

2.3. Distribusi Binomial Negatif

The negative binomial distribution calculates the probability of xx successes in a sequence of independent trials with success probability p before r failures occur (Novaes, 2022). The probability mass function (PMF) is given by equation (8).

$$f(x) = \binom{x+r-1}{x} p^x (1-p)^r \tag{8}$$

if r is not an integer, the PMF becomes equation (9)

$$f(x) = \frac{\Gamma(x + r)}{\Gamma(r)\Gamma(x + 1)} p^x (1 - p)^r \tag{9}$$

The expected value of the negative binomial distribution is given by equation (10)

$$E(X) = \frac{r}{p} \tag{10}$$

The variance of the negative binomial distribution is given by equation (11).

$$E(X) = \frac{r(1 - p)}{p^2} \tag{11}$$

2.4. Kolmogorov-Smirnov Test

The Kolmogorov-Smirnov test is a goodness-of-fit test, where the degree of conformity between the data and a specific theoretical distribution is the main focus of this test (Nuryadi *et al*., 2017). In this test, the cumulative frequency distribution occurring under the theoretical distribution is calculated. This test also compares the frequency distribution with the cumulative frequency distribution from the observed data.

The hypotheses used in this test are as follows:
H_0 : $F(x) = F^*(x)$ [the data follows the specified distribution]
H_1 : $F(x) \neq F^*(x)$ [the data does not follow the specified distribution]

The test statistic used in this test can be calculated using equation (12) as follows.

$$D = sup\,|F^*(x) - S(x)| \tag{12}$$

where,
D : maximum difference,
$F^*(x)$: theoretical cumulative distribution function,
$S(x)$:eEmpirical cumulative distribution function.

In this test, if the test statistic is greater than the critical value, then H_0 is rejected, indicating that the data follows a different distribution. Conversely, if the test statistic is smaller than the critical value, then H_0 is accepted, indicating that the data follows the tested distribution.

2.5. Anderson-Darling Test

The Anderson-Darling test is one of the model fit tests used for continuous distributions. In this test, a more flexible approach is taken based on the theoretical distribution being used (Yulita et al., 2024).

The hypotheses used in this test are as follows:
H_0 : the data follows the specified distribution
H_1 : the data does not follow the specified distribution

The test statistic used in this test can be calculated using Equation (13) as follows

$$A^2 = -n - \frac{1}{n}\sum_{i=1}^{n} \left[(2n - i)\left[ln\big(F(x_i)\big) + ln\big(1 - F(x_{n-i+1})\big)\right]\right] \tag{13}$$

where,
A^2 : anderson-Darling test statistic,
$F(x)$: theoretical cumulative frequency distribution,
n : number of data points.

In this test, if the test statistic is greater than the critical value, then H_0 is rejected, meaning the data follows another distribution. Conversely, if the test statistic is smaller than the critical value, then H_0 is accepted, meaning the data follows the tested distribution.

2.6. Chi-Square Test

The Chi-Square test is a non-parametric model fit test that can be used for discrete distributions. This test can be applied to discrete distributions such as the Poisson distribution (Yulita et al., 2024). The Chi-Square test can be considered a proportion test for two or more events, making it discrete (Fitri et al., 2023).

The hypotheses used in this test are as follows:
H_0 : the data follows the specified distribution,
H_1 : the data follows another distribution.

To calculate the test statistic for the Chi-Square test, the degrees of freedom (df). must be determined. The degrees of freedom represent the total number of observations minus the number of independent constraints used in the study. The degrees of freedom can be calculated using Equation (14).

$$df = k - 1 - p \tag{14}$$

where,
df : degrees of freedom,
k : number of data points,
p : number of parameters.

The test statistic used in this test can be calculated using Equation (15).

$$X^2 = \sum_{j=1}^{n} \frac{\left(E_j - O_j\right)^2}{E_j} \tag{15}$$

where,
X^2 : Chi-Square test statistic,
E_j : Expected value for the j-th category,
O_j : Observed value for the j-th category.

Additionally, another test statistic is obtained from the Chi-Square table, where the test statistic can be found using the table with Equation (16).

$$X^2_{tabel} = X^2_{df,\alpha} \tag{16}$$

where,
X^2_{tabel} : Chi-Square value from the table
α : confidence level

If $X^2 \geq X^2_{tabel}$, then reject H_0 meaning the data follows another distribution. Otherwise, accept H_0 meaning the data follows the specified distribution.

3. Materials and Methods

3.1. Materials

The object of this study is the frequency and magnitude of accident losses in Indonesia from 2013 to 2022. This data was obtained from the Central Bureau of Statistics. The dataset is presented in Table 1 below:

Table 1: Frequency and Magnitude of Accident Losses

Year	Frequency	Magnitude (Million IDR)
2013	100106	255864
2014	95906	250021
2015	96233	215892
2016	106644	229137
2017	104237	217031
2018	109215	213866
2019	116411	254779
2020	100028	198456

2021	103645	246653
2022	139258	280009

3.2. Methods

This study follows the following steps:

1. Calculating the magnitude of loss per accident using Equation (17).

$$Magnitude\ per\ accident = \frac{Magnitude}{Frekuency} \tag{17}$$

2. Performing distribution fitting using the Generalized Extreme Value (GEV) distribution for accident loss magnitude and the Negative Binomial distribution for accident frequency. Parameter estimation is conducted using EasyFit software.
3. Conducting distribution tests, including the Kolmogorov-Smirnov test, Anderson-Darling test, and Chi-Square test for both distributions.
4. Calculating the expectation and variance of the aggregate claim.

4. Results and Discussion

First, the magnitude of loss per accident was calculated using Equation (15), yielding the results presented in Table 2.

Table 2: Magnitude of Loss Per Accident

Year	Magnitude of Loss Per Accident
2013	2555930.713
2014	2606938.044
2015	2243430.008
2016	2148615.956
2017	2080295.609
2018	1958210.868
2019	2188616.196
2020	1984004.479
2021	2379786.772
2022	2010721.108

Next, the claim frequency data was fitted to the Negative Binomial distribution, and the accident loss magnitude was fitted to the GEV distribution using EasyFit software. After fitting, the parameters of each distribution were estimated. The claim frequency follows a Negative Binomial distribution with $n = 77$ and $p = 0.00072057$. The loss magnitude follows a GEV distribution with $k = -0.28916$, $\sigma = 182670$ and $\mu = 2096600$.

Then, both distributions were tested using the Kolmogorov-Smirnov test, the Anderson-Darling test, and the Chi-Square test for both distributions using EasyFit, and the results obtained are as follows:

Table 3: Negative Binomial Distribution Test Results

α	Kolmogorov-Smirnov Test	Anderson Darling Test
0.01	Reject H_0	Reject H_0
0.02	Reject H_0	Reject H_0
0.05	Reject H_0	Reject H_0
0.1	Reject H_0	Reject H_0
0.2	Reject H_0	Reject H_0

Table 4: GEV Distribution Test Results

α	Kolmogorov-Smirnov Test	Anderson Darling Test	Chi-Square Test
0.01	Reject H_0	Reject H_0	Reject H_0
0.02	Reject H_0	Reject H_0	Reject H_0
0.05	Reject H_0	Reject H_0	Reject H_0
0.1	Reject H_0	Reject H_0	Reject H_0
0.2	Reject H_0	Reject H_0	Reject H_0

From the overall distribution test results, it can be concluded that both distributions fit the claim frequency and accident loss data. Subsequently, the mathematical expectation value of claim frequency was calculated using equation (9), and the mathematical expectation value of accident loss was calculated using equation (6). Using Python software, the expected claim frequency was obtained as 106859.84706551758 and the expected loss per accident was IDR 1,481,034.9326895317. Consequently, using equation (2), the mathematical expectation of aggregate claims for accident losses was obtained as IDR 236,768,463,076.92972 with a variance of 7.275183534685082e+20 for the Aggregate Claims Model based on the Number and Size of Accident Losses in Indonesia from 2013-2022.

5. Conclussion

Based on the research results, it was found that the negative binomial distribution and the GEV distribution are suitable for the obtained accident loss data. Subsequently, the aggregate claim results yielded an expected value of IDR 236,768,463,076.93 with a variance of 7.275183534685082e+20. From these results, it can be concluded that insurance companies covering accident losses should have funds amounting to the expected value obtained, which is IDR 236,768,463,076.93. Additionally, due to the high variance, it is recommended to have a reserve fund to anticipate potential variations.

References

Dickson, D. C. M. (2006). Insurance Risk and Ruin. *Journal of the American Statistical Association*, *101*(475), 1316–1316. https://doi.org/10.1198/jasa.2006.s131

Fitri, A., Rahim, R., Nurhayati, Aziz, Pagiling, S. L., Natsir, I., Munfarikhatin, A., Simanjuntak, D. N., HUatgaol, K., & Anugrah, N. E. (2023). Statistical Basics for Research. In *Yayasan Kita Menulis*. https://repository.unugiri.ac.id:8443/id/eprint/4882/1/Anisa %2C Book Basics of Statistics for Research.pdf

Novaes, M. (2022). Probability distributions. *Revista Brasileira de Ensino de Fisica*, *44*. https://doi.org/10.1590/1806-9126-RBEF-2021-0424

Nuryadi, Astuti, T. D., Utami, E. S., & Budiantara, M. (2017). Textbook of Fundamentals of Research Statistics. In *Sibuku Media*.

Rahayu, A. (2012). *ESTIMATION OF GENERALIZED EXTREME VALUE (GEV) DISTRIBUTION PARAMETERS (Case Study: Identification of Climate Change in Jakarta)*.

Yulita, T., Patricia, M., & Hidayat, A. S. E. (2024). Determination of Pure Premiums from Four-Wheeled Vehicle Insurance Claim Data with Comprehensive Protection Type. *VARIANCE: Journal of Statistics and Its Applications*, *6*(1), 75–86. https://doi.org/10.30598/variancevol6iss1page75-86