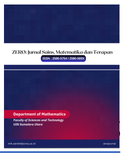
**Zero : Jurnal Sains, Matematika, dan Terapan**
E-ISSN : 2580-5754; P-ISSN : 2580-569X
Volume 9, Number 1, 2025
DOI: 10.30829/zero.v9i1.24581
Page: 26-36



# Zero Inflated Poisson Regression for Analyzing Excess Zeros in Job Transition Data: A Case Study of Tourism Workers in Malang Regency

[1] Sarni Maniar Berliana
Research Unit of Sustainable Development Goals, Politeknik Statistika STIS, Jakarta, Indonesia
[2] Rafidah Abd Karim
Academy of Language Studies, University Teknologi MARA Perak Branch Tapah Campus, Malaysia
[3] Rita Yuliana
Research Unit of Economic and Policy, Politeknik Statistika STIS, Jakarta, Indonesia
[4] Budyanra
Research Unit of Sustainable Development Goals, Politeknik Statistika STIS, Jakarta, Indonesia
[5] Achmad Prasetyo
Research Unit of Sustainable Development Goals, Politeknik Statistika STIS, Jakarta, Indonesia
[6] Arya Candra Kusuma
Directorate of Crops, Horticulture, and Plantation Statistics, BPS-Statistics Indonesia, Jakarta, Indonesia
[7] Priatmadani
Team of Production Statistics, BPS-Statistics of Majene Regency, West Sulawesi, Indonesia

**Article Info**

***Article history:***

Accepted, 28 May 2025

***Keywords:***

Count Data;
Excess Zeros;
Job Transitions;
Tourism Industry;
Zero-Inflated;
Poisson.

**ABSTRACT**

This study examines job transition patterns among 868 tourism workers in Malang, Indonesia, a key tourism hub, using 2023 Fieldwork Training data from Politeknik Statistika STIS. It aims to model transitions while addressing the high number of workers with no job changes using a Zero-Inflated Poisson (ZIP) model, which better captures these patterns than standard models. The ZIP model, including sex, age, relationship to the head of household, education, and foreign language proficiency, shows that proficient workers are more likely to remain in stable roles, while men and younger workers exhibit greater mobility, particularly when leveraging language skills. These findings support Indonesia's Sertifikasi Kompetensi SDM Pariwisata program by justifying targeted interventions: language training to enhance mobility for non-proficient workers, mentorship for female tour guides to address gender disparities, and digital skills programs for older workers to boost employability. These strategies align with government efforts to strengthen Malang's tourism workforce resilience.



***Corresponding Author:***

Sarni Maniar Berliana,
Research Unit of Sustainable Development Goals,
Politeknik Statistika STIS, Jakarta, Indonesia
Email: sarni@stis.ac.id

## 1. INTRODUCTION

The tourism industry is a major driver of economic growth in many countries worldwide [1], including Indonesia. Malang Regency, characterized by its abundant natural landscapes and varied attractions such as Coban Rondo, Mount Bromo, and Flora Wisata, is a popular tourist destination that attracts both local and international visitors [2]. As a labor-intensive sector, tourism requires workers with diverse skills to support services like accommodation, transportation, food and beverage, tour guiding, and recreation [3]. The high tourist influx in Malang Regency underscores the need for a substantial workforce, positioning the tourism industry as a vital contributor to job creation and economic development [2].

Given the sector's labor demands, analyzing labor dynamics, particularly job transitions, is crucial for understanding workforce mobility and its socioeconomic impacts. Labor market theory suggests that worker-firm interactions drive job transitions, with mobility reflecting economic opportunities or constraints [4]. In tourism, transitions redistribute skills (e.g., hotels hiring experienced guides), boosting local economies, but may disrupt worker stability, affecting earnings [5]. Factors like education, training, and socioeconomic conditions—such as age, gender, skills, and urban/rural residence—shape these transitions, influencing job quality and career paths [6, 7, 8, 9, 10, 11, 12]. Younger workers often transition to gain skills, while women face barriers to high-level roles, increasing mobility [6, 7, 8, 9]. Industry-specific skills enhance employability, and urban workers access more opportunities than rural ones [10, 11, 12]. By modeling these dynamics, stakeholders can design interventions, like skill development programs, to enhance workforce resilience and support sustainable tourism growth in Malang.

Job transitions are typically measured as count data, representing the number of job changes over a period. Such data often exhibit excess zeros, where many individuals report no transitions due to structural factors like job satisfaction, skill constraints, or market conditions [13]. Standard Poisson regression, which assumes equidispersion (mean equals variance), is often inadequate for such data, as excess zeros lead to overdispersion and biased estimates [14]. The Zero-Inflated Poisson (ZIP) model addresses this challenge by modeling two processes: a logistic component for structural zeros (e.g., workers who never transition due to specific characteristics) and a Poisson component for counts, including random zeros [15]. This dual-process approach makes the ZIP model well-suited for analyzing job transitions in Malang's tourism sector, where socioeconomic factors may create distinct patterns of labor mobility.

The ZIP model's applicability to count data with excess zeros is well-documented across various disciplines, highlighting its relevance for studying labor dynamics in tourism. For instance, in health economics, ZIP models have analyzed healthcare utilization, such as outpatient or doctor visits, where zeros reflect barriers like good health or lack of access [16, 17]. In health behavior research, ZIP has modeled counts of high-risk activities, with zeros indicating abstention [18], while in developmental studies, it has captured fine motor development counts, with zeros reflecting delays [19]. In economics, ZIP has been used to study consumer purchase frequencies, where zeros denote non-participation [20]. Within labor economics, ZIP has proven particularly valuable for modeling job mobility. Heitmüller [21] applied ZIP to examine job changes in the UK labor market, where excess zeros indicated stable employment due to regional or socioeconomic factors. Similarly, List [22] used ZIP to analyze academic job interviews for non-tenured economists in the U.S., with zeros linked to candidate characteristics like age or gender. Recent studies, such as Motalebi et al. [23], used ZIP to monitor social network interactions, while Truong et al. [24] applied it to social science data, and Argawu & Mekebo [25] analyzed under-five mortality in Ethiopia. These diverse applications underscore the ZIP model's ability to separate structural and random zeros, offering a robust framework for analyzing job transitions in Malang's tourism industry, where similar socioeconomic influences shape labor dynamics [26]. By integrating these insights, the ZIP model emerges as a powerful tool for capturing the complex interplay of factors driving workforce mobility in this context.

This study aims to investigate the influence of socioeconomic factors on job transitions within Malang Regency's tourism sector using the ZIP model, hypothesizing that explanatory variables, including foreign language proficiency, sex, age, education, and relationship to the head of household, significantly influence job transitions. By analyzing these transitions, the research seeks to deepen the understanding of labor dynamics and provide actionable insights for policymakers and industry stakeholders. The expected findings can integrate with tourism workforce planning systems, such as Indonesia's Sertifikasi Kompetensi SDM Pariwisata program, and digital dashboards, enabling data-driven strategies for skill development and workforce allocation to enhance job quality and promote sustainable growth in Malang's tourism industry.

The remainder of this paper is organized as follows: Section 2 outlines the research method, covering topics such as the distribution of count data, Poisson regression, and the Zero-Inflated Poisson (ZIP) regression model, including its parameter estimation and hypothesis testing, as well as overdispersion testing, model evaluation, and data description. Section 3 presents the results and analysis, exploring preliminary findings on job transition data, factors influencing job transitions among tourism workers in Malang Regency, and an evaluation of model performance. Section 4 concludes with policy recommendations and suggestions for future research.

## 2. RESEARCH METHOD

### 2.1 Distribution for Count Data

Count data, such as the number of job transitions over a fixed period, are characterized by non-negative integers (0, 1, 2, ...). The Poisson distribution is commonly used to model count data due to its simplicity and suitability for events occurring independently at a constant average rate. For a random variable $Y_i$ representing the count of job transitions for individual $i$, the Poisson probability mass function is:

$$P(Y_i = y_i) = \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!}, \quad y_i = 0,1,2,\ldots, \tag{1}$$

where $\lambda_i > 0$ is the expected count (mean and variance) for individual $i$ and $y_i!$ is the factorial of $y_i$ [13]. The Poisson distribution assumes equidispersion, where the mean equals the variance $(E(Y_i) = Var(Y_i) = \lambda_i)$. However, real-world count data, such as job transitions, often exhibit overdispersion (variance exceeds mean) or excess zeros, where the proportion of zero counts is higher than expected under a Poisson model [14].

### 2.2 Poisson Regression Model

To model the relationship between the expected count $\lambda_i$ and explanatory variables (e.g., age, education), Poisson regression is employed. The model assumes a log-linear relationship:

$$\log(\lambda_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip}, \tag{2}$$

where $x_{i1}, x_{i2}, \ldots, x_{ip}$ are explanatory variables for individual $i$ and $\beta_0, \beta_1, \ldots, \beta_p$ are regression coefficients [13]. The expected count is:

$$\lambda_i = \exp\left(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip}\right). \tag{3}$$

However, when data exhibit excess zeros, as observed in job transition data where many workers do not change jobs, the Poisson model may produce biased estimates due to its inability to account for the excess zero counts [14].

### 2.3 Zero-Inflated Poisson (ZIP) Regression Model

The Zero-Inflated Poisson (ZIP) model addresses excess zeros by combining two processes: a Poisson count process and a zero-inflation process [13]. This model combines a Poisson count process, which models the frequency of events for individuals with potential to experience the event, with a zero-inflation process that accounts for individuals who never experience the event due to specific characteristics [15]. Excess zeros arise from two sources: (1) structural zeros, where individuals are inherently unable to experience the event due to underlying factors, and (2) sampling zeros, where individuals could experience the event but record zero occurrences by chance, as predicted by the Poisson distribution [14]. For example, in a study of job transitions among tourism workers in Malang Regency (2020–2023), structural zeros may include workers proficient in a foreign language, such as English, who remain in stable, high-demand roles like international tour guides, reducing their likelihood of job changes. Sampling zeros may occur among younger workers who have the potential to switch jobs due to their age but do not transition during the observation period, possibly due to temporary market constraints post-COVID-19.

The ZIP model assumes that the observed count $Y_i$ is generated by a mixture of a logistic process (for structural zeros) and a Poisson process (for counts, including random zeros). The probability mass function is

$$P(Y_i = 0) = \pi_i + (1 - \pi_i)e^{-\lambda_i}, \tag{4}$$

$$P(Y_i = y_i) = (1 - \pi_i)\frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!}, \quad y_i = 1,2,\ldots, \tag{5}$$

where $\pi_i$ is the probability of a structural zero for individual $i$ and $\lambda_i$ is the expected count from the Poisson component [15]. The zero-inflation component models $\pi_i$ using a logistic regression

$$\text{logit}(\pi_i) = \log\left(\frac{\pi_i}{1 - \pi_i}\right) = \gamma_0 + \gamma_1 z_{i1} + \gamma_2 z_{i2} + \cdots + \gamma_m z_{im}, \tag{6}$$

where $z_{i1}, z_{i2}, \ldots, z_{im}$ are explanatory variables (which may overlap with those in the count model), and $\gamma_0, \gamma_1, \ldots, \gamma_m$ are coefficients. The count component follows the Poisson regression form

$$\log(\lambda_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip}. \tag{7}$$

The expected value of $Y_i$ in the ZIP model is

$$E(Y_i) = (1 - \pi_i)\lambda_i, \tag{8}$$

reflecting that only the non-structural zero group contributes to the expected count [15].

The explanatory variables for the zero-inflation and count components of the ZIP model need not be identical, as each component models distinct processes [15, 27]. The logistic component requires explanatory variables strongly associated with structural zeros, such as foreign language proficiency for stable tourism roles, while the Poisson component includes variables influencing event frequency, such as education and training. This flexibility enhances model efficiency by reducing overfitting and is particularly advantageous in moderate sample sizes, where parsimonious models improve estimation stability [27]. In this study, foreign language proficiency is included in both components, with additional socioeconomic variables (e.g., education, training) in the count component. Deepa et al. (2025) found a strong correlation between English language proficiency and career advancement in sectors where English is the primary communication medium, such as tourism, with proficient workers often securing stable, high-demand positions like international tour guides [28]. This makes foreign language proficiency a theoretically relevant variable for the zero-inflation component, which models structural zeros (workers unlikely to transition). This choice was validated using the Akaike Information Criterion (AIC), confirming improved model fit.

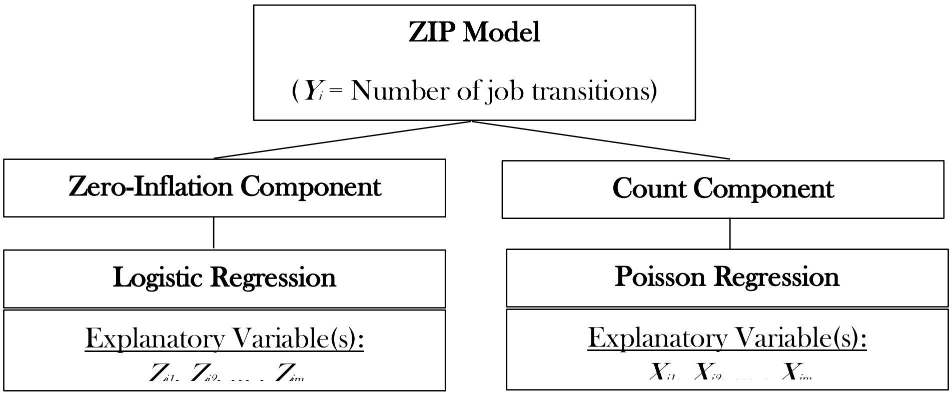



**Figure 1.** ZIP Model Structure with Explanatory Variables for Job Transitions

To illustrate the structure of the ZIP model and complement the selection of explanatory variables, Figure 1 presents a diagram detailing the zero-inflation and count components, including their respective dependent and explanatory variables, which combine to produce the observed job transition counts. The explanatory variables for the zero-inflation component, denoted as $Z_{i1}, Z_{i2}, \dots, Z_{im}$, range from 1 to $m$ while those for the Poisson component, denoted as $X_{i1}, X_{i2}, \dots, X_{ip}$, range from 1 to $p$ where $m$ need not equal $p$, reflecting the model's flexibility in accommodating distinct explanatory variables for each component.

### 2.3.1 Parameter Estimation

Parameters in the ZIP model, $\beta$ for the count component and $\gamma$ for the zero-inflation component, are estimated using maximum likelihood estimation (MLE). The log-likelihood function is maximized to obtain estimates of $\beta_0, \beta_1, \dots, \beta_p$ and $\gamma_0, \gamma_1, \dots, \gamma_m$. The likelihood contribution for individual $i$ is

$$L_i(\beta, \gamma) = \begin{cases} \pi_i + (1 - \pi_i)e^{-\lambda_i}, & \text{if } y_i = 0, \\ (1 - \pi_i)\dfrac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!}, & \text{if } y_i \geq 1. \end{cases} \tag{9}$$

To simplify, define the indicator $I(y_i = 0)$ which equals 1 if $y_i = 0$ and 0 otherwise. The likelihood can be written compactly as

$$L_i(\beta, \gamma) = \left(\pi_i + (1 - \pi_i)e^{-\lambda_i}\right)^{I(y_i=0))} \left[(1 - \pi_i)\frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!}\right]^{1-I(y_i=0)}. \tag{10}$$

The log-likelihood for a sample of $n$ independent observations is

$$\ell(\beta, \gamma) = \sum_{i=1}^{n} ln\left(L_i(\beta, \gamma)\right). \tag{11}$$

Substituting $L_i$, the log-likelihood is

$$\begin{aligned} \ell(\beta, \gamma) = \sum_{i=1}^{n} \{ & I(y_i = 0) ln\left(\pi_i + (1 - \pi_i)e^{-\lambda_i}\right) \\ & + \left(1 - I(y_i = 0)\right)\left(ln(1 - \pi_i) - \lambda_i + y_i ln(\lambda_i) - ln(y_i!)\right)\}, \end{aligned} \tag{12}$$

where $\pi_i = exp(\gamma_0 + \sum_{k=1}^{m} \gamma_k z_{ik})\left(1 + exp(\gamma_0 + \sum_{k=1}^{m} \gamma_k z_{ik})\right)^{-1}$ and $\lambda_i = exp\left(\beta_0 + \sum_{j=1}^{p} \beta_j x_{ij}\right)$. The parameters $\beta$ and $\gamma$ are estimated by maximizing $\ell(\beta, \gamma)$ in (12), typically using numerical optimization methods (e.g., Newton-Raphson) due to the non-linear nature of the function [14].

In the Poisson component, the exponentiated coefficient $e^{\beta_j}$ is the incidence rate ratio (IRR), representing the multiplicative effect on the expected count for a unit increase in $X_j$. In the zero-inflation component, $e^{\gamma_k}$ is the odds ratio (OR), indicating the change in odds of a structural zero for a unit increase in $Z_k$ [27].

### 2.3.2 Hypothesis Testing

Two hypothesis tests are employed to assess the significance of model parameters: a simultaneous test for all parameters and a partial test for individual parameters.

**Simultaneous Parameter Testing**

The likelihood ratio test (LRT) is used to evaluate whether all regression coefficients (excluding intercepts) in the count and zero-inflation components are zero, testing the overall model fit [13]. The hypothesis is

$$H_0: \beta_1 = \cdots = \beta_p = \gamma_1 = \cdots = \gamma_m = 0, \qquad (13)$$
$$H_1: \text{At least one } \beta_j \text{ or } \gamma_k \neq 0, \quad j = 1, \dots, p, \quad k = 1, \dots, m.$$

The LRT statistic is

$$\text{LRT} = -2[\ell_0 - \ell_1], \qquad (14)$$

where $\ell_0$ is the log-likelihood of the null model (intercepts only), and $\ell_1$ is the log-likelihood of the full ZIP model. The LRT statistic approximately follows a chi-square distribution with degrees of freedom (df) equal to the number of tested parameters ($\text{df} = p + m$). Reject $H_0$ if $\text{LRT} > \chi^2_{1-\alpha,p+m}$. Rejection of $H_0$ indicates that at least one explanatory variable significantly affects the count or zero-inflation process; otherwise, there is insufficient evidence to reject $H_0$ [13].

**Partial Parameter Testing**

To test the significance of individual parameters (e.g., $\beta_j$ or $\gamma_k$, a Wald test is applied. The hypothesis (e.g., for $\beta_j$ is

$$H_0: \beta_j = 0, \qquad (15)$$
$$H_1: \beta_j \neq 0, \quad j = 1, \dots, p,$$

and similarly for $\gamma_k, k = 1, \dots, m$. The test statistic for hypothesis testing in (15) is

$$\text{Z} = \frac{\hat{\beta}_j}{\text{SE}(\hat{\beta}_j)}, \qquad (16)$$

where $\hat{\beta}_j$ is the estimated coefficient, and $\text{SE}(\hat{\beta}_j)$ is its standard error [29]. The same applies for $\hat{\gamma}_k$. The test statistic Z follows a standard normal distribution $\text{N}(0,1)$ under $\text{H}_0$. Reject $\text{H}_0$ if $|\text{Z}| > \text{z}_{\alpha/2}$. Rejection of $\text{H}_0$ indicates that the explanatory variable significantly affects the count or zero-inflation component; otherwise, there is insufficient evidence to reject $\text{H}_0$.

## 2.3 Overdispersion Test

Overdispersion occurs in count data when the variance exceeds the mean, violating the equidispersion assumption of the standard Poisson model, where $\text{E}(\text{Y}_i) = \text{Var}(\text{Y}_i) = \lambda_i$ for the count variable $\text{Y}_i$ and expected count $\lambda_i$ [13]. This phenomenon, common in datasets with excess zeros, can lead to biased estimates and underestimated standard errors in Poisson regression [14]. To detect overdispersion in count data, a statistical test evaluates the dispersion parameter, denoted $\phi$, which quantifies excess variability. The hypotheses are $\text{H}_0: \phi = 0$, indicating equidispersion ($\text{Var}(\text{Y}) = \lambda$), consistent with a Poisson model and the alternative hypothesis $\text{H}_1: \phi > 0$, indicating overdispersion ($\text{Var}(\text{Y}) = \lambda + \phi\lambda^2$), suggesting a Negative Binomial-like variance structure [30]. The test statistic is then computed as

$$\text{Z}_1 = \frac{\hat{\phi}}{\text{SE}(\hat{\phi})}, \qquad (17)$$

where $\text{SE}(\hat{\phi})$ is the standard error of $\hat{\phi}$, derived from the variance-covariance matrix of the estimator. This approach assumes a variance structure of $\text{Var}(\text{Y}) = \lambda + \phi\lambda^2$, often with a transformation to align with the Negative Binomial model [26]. Under the null hypothesis ($\phi = 0$), the test statistic $\text{Z}_1$ approximately follows a standard normal distribution for large sample sizes, due to the asymptotic properties of the maximum likelihood estimator of $\phi$. This enables the computation of a p-value to assess overdispersion. The null hypothesis is rejected if the p-value is less than the chosen significance level. A significant p-value indicates overdispersion, suggesting that the Poisson model is inadequate for the data [30].

## 2.4 Model Evaluation

To evaluate the goodness-of-fit and compare the ZIP model with alternative models (e.g., standard Poisson), the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) are used. These criteria balance model fit and complexity

$$\text{AIC} = -2\ell + 2\text{p}^*, \qquad (18)$$
$$\text{BIC} = -2\ell + \text{p}^*\log(\text{n}), \qquad (19)$$

where $\ell$ is the log-likelihood of the fitted model, $\text{p}^*$ is the total number of parameters, and $\text{n}$ is the sample size [31, 32]. Lower AIC or BIC values indicate a better trade-off between model fit and parsimony. AIC penalizes model complexity with a fixed term $2\text{p}^*$, while BIC imposes a stronger penalty proportional to the sample size $\text{p}^*\log(\text{n})$. Thus, BIC tends to favor simpler models, especially for larger datasets [30, 33]. In this study, both AIC and BIC are used to compare the ZIP model with the Poisson model to ensure robustness in model selection.

### 2.5 Data

This study utilizes data collected through a survey conducted by diploma four students of Politeknik Statistika STIS from January 23 to February 3, 2023, in Malang Regency, East Java, Indonesia. The survey targeted 868 tourism workers across three districts as shown in Figure 2—Bantur, Pujon, and Poncokusumo—using direct interviews and questionnaires to gather socio-demographic and economic data, including gender, age, education, household status, region of residence, and foreign language proficiency. Malang Regency was selected due to its significant tourism sector, which supports substantial workforce mobility. In 2021, Malang ranked among the top destinations for domestic tourist visits in East Java, alongside Surabaya and Banyuwangi, driven by diverse attractions such as natural sites, heritage destinations, and tourist villages [2]. The gradual recovery of the tourism sector post-COVID-19 has further amplified job mobility, making Malang an ideal setting to study job transition dynamics.

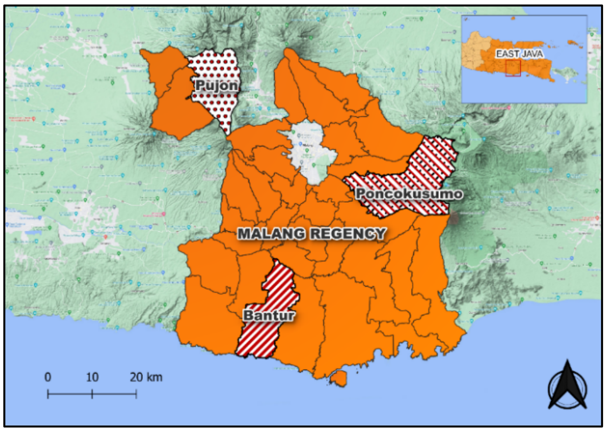



**Figure 2.** Selected Research Areas: Bantur, Pujon, and Poncokusumo

This study focuses on tourism workers aged 10 years or older in Malang Regency, East Java, Indonesia, sampled using a three-stage cluster sampling design. In the first stage, districts (Bantur, Pujon, and Poncokusumo) were selected from Malang Regency's district sampling frame using probability proportional to size (PPS) based on the number of tourist attractions and accommodations, sourced from BPS-Statistics of Malang Regency. In the second stage, census blocks within selected districts were sampled using PPS based on the number of tourism-related businesses. In the third stage, households were randomly selected within each census block, yielding 868 individuals employed in the tourism sector for the three years prior to 2023.

The dependent variable, number of job transitions, is defined as the count of job changes within this period (e.g., a worker moving from taxi driver to tour guide, then to hotel employee, has two transitions). Independent variables include gender, age, household status, educational level, and foreign language proficiency, with the latter included in both the count and zero-inflated components of the Zero-Inflated Poisson model. Operational definitions and measurement scales are presented in Table 1.

**Table 1.** Operational Definition of Research Variables

| Variable | Operational Definition | Scale | Description |
|---|---|---|---|
| Number of job transitions | Count of job changes within or across organizations in the tourism sector (2020–2023). | Numerical | Count (0, 1, 2, ...) |
| Gender | Self-reported gender of the individual. | Categorical | 1 = Male<br>2 = Female* |
| Age | Age at the time of the survey, based on the last birthday. | Numerical | Years |
| Relationship to the head of household | Role of the individual within the household relative to the head. | Categorical | 1 = Head of the household<br>2 = Others* |
| Educational level | Highest level of education completed, verified by diploma or certificate. | Categorical | 1 = Lower than high school<br>2 = High school or higher* |
| Foreign language proficiency | Ability to communicate effectively in a non-Indonesian language, self-reported. | Categorical | 1 = Proficient<br>2 = Not proficient* |

Note: Categories marked with an asterisk (*) serve as the reference group in the Zero-Inflated Poisson regression model.

## 3. RESULT AND ANALYSIS

### 3.1 Preliminary Analysis

The study's sample comprises 868 tourism workers surveyed from January 23 to February 3, 2023, in three sub-districts of Malang Regency: Bantur, Pujon, and Poncokusumo. Of the respondents, 58.9% are male, 61.1% have a household status categorized as non-head (e.g., spouse or child), 56.2% have an educational level below high school, and 80.0% lack proficiency in a foreign language. The average age of respondents is 36.1 years.

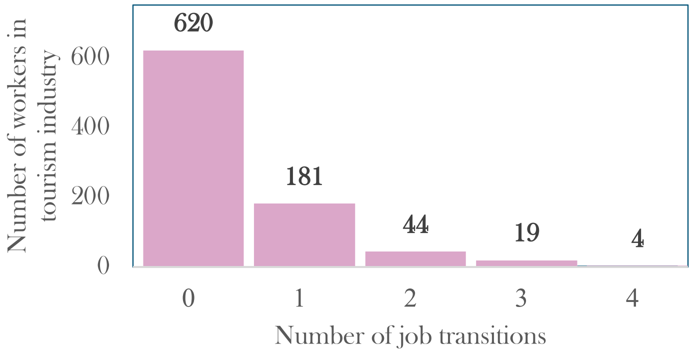



**Figure 3.** Number of Workers in the Tourism Industry by Number of Job Transitions, Malang Regency 2023

Figure illustrates the distribution of tourism workers in Malang Regency by number of job transitions from January 23, 2020, to February 3, 2023. Of the respondents, 71.4% had no job transitions, indicating excess zeros and suggesting high job stability in the sector [34]. Additionally, 20.9% of workers experienced one job transition and 5.1% experienced two, reflecting moderate mobility. Meanwhile, 2.7% experienced three or more transitions, indicating high mobility among a small group [35].

The job transition data revealed a mean of 0.394 and a variance of 0.527, yielding a variance-to-mean ratio of 1.339, suggesting potential overdispersion in the count data. To formally assess this, an overdispersion test was conducted on the Poisson model, resulting a test statistic $Z_1 = 3.4851$ (p-value= 0.000246) as defined in (17). These results confirm significant overdispersion, indicating that the standard Poisson model is inadequate for capturing the variability in the data.

### 3.2 ZIP Model Interpretation

Given the significant overdispersion in the job transition data, a Zero-Inflated Poisson (ZIP) model was utilized to address excess zeros and variability among tourism workers in Malang Regency. This section examines factors influencing job transitions, with foreign language proficiency as the sole explanatory variable in the zero-inflated component and multiple socioeconomic variables in the count component. Limiting the zero-inflated component to one variable enhances model parsimony and reduces overfitting, as supported by prior studies [13, 27]. Model selection was validated using the Akaike Information Criterion (AIC), where the ZIP model with only foreign language proficiency in the zero-inflated component (AIC = 1398.787, df = 8) outperformed the model including all explanatory variables (AIC = 1400.165, df = 12), indicating a better fit. Table 2 presents the regression coefficients for both components, with significance assessed at the 10% level ($p < 0.10$), offering insights into the drivers of job transitions in the tourism sector.

**Table 2.** ZIP Regression Coefficients for Zero-Inflated and Count Components of Job Transition Models, Malang Regency, 2023

| Component | Explanatory Variable | Coefficient | SE | p-value | OR/IRR |
|---|---|---|---|---|---|
| Count Component | | | | | |
| | (Intercept) | 0.291 | 0.259 | 0.262 | 1.338 |
| | Gender | | | | |
| | Male | 0.383 | 0.140 | 0.006 | 1.467 |
| | Female | - | | | |
| | Age | -0.035 | 0.006 | 0.000 | 0.966 |
| | Relationship to the head of household | | | | |
| | Head of household | -0.008 | 0.167 | 0.961 | 0.992 |
| | Others | | | | |

| | | | | | |
|---|---|---|---|---|---|
| | Educational level | | | | |
| | Lower than high school | - | | | |
| | High school or higher | -0.048 | 0.130 | 0.714 | 0.953 |
| | Foreign language proficiency | | | | |
| | Proficient | 0.351 | 0.214 | 0.102 | 1.420 |
| | Not proficient | - | | | |
| **Zero-Inflated Component** | | | | | |
| | (Intercept) | -1.081 | 0.403 | 0.007 | 0.339 |
| | Foreign language proficiency | | | | |
| | Proficient | 0.906 | 0.531 | 0.088 | 2.473 |
| | Not proficient | - | | | |
| Log-likelihood null model | = -721.88 | (df=2) | | | |
| Log-likelihood full model | = -691.39 | (df=8) | | | |
| LRT | = 60.981 | (p=0.0000) | | | |

The count component of the ZIP model in Table 2 provides insight into the variables influencing the frequency of job transitions, with the relationship with the head of household (coefficient = -0.008, p = 0.961, IRR = 0.992) and educational level (coefficient = -0.048, p = 0.714, IRR = 0.953 for high school or higher) demonstrate no statistically significant impact on job transitions. The lack of significance for the relationship with the head of household (p = 0.961) likely arises from its limited relevance to job mobility, as household status—whether a worker is the head of the household or a dependent—does not directly shape labor market dynamics or career opportunities [36]. This suggests that in the tourism sector, job mobility is driven more by individual attributes than by family roles or responsibilities, indicating that household dynamics play a minimal role in influencing workers' decisions or opportunities to change jobs. Similarly, the non-significant effect of educational level (p = 0.714) can be attributed to the broad categorization of education (lower than high school vs. high school or higher), which may obscure the nuanced effects of specific qualifications, skills, or training that are more directly tied to job transitions [37]. This finding implies that in the tourism industry, formal education alone does not strongly dictate job mobility; instead, practical skills, hands-on experience, and other specialized qualifications appear to be more critical drivers of transitions between roles [35, 34].

Gender shows significant effect on job transitions, with males exhibiting a higher rate of transitions compared to females (coefficient = 0.383, p = 0.006, IRR = 1.467). This indicates that males have a 46.7% higher incidence rate of job transitions, a finding that is statistically significant at the 1% level (p = 0.006). The significance of gender may be attributed to societal and occupational factors, such as greater occupational flexibility for men, who may face fewer barriers in pursuing diverse roles or industries, particularly in sectors like tourism where physical demands or cultural norms might favor male mobility. Additionally, men may be more likely to take risks in career changes or have access to broader networks that facilitate job transitions. This aligns with prior research, such as England [38], which found that gender roles and societal expectations often result in men experiencing greater job mobility, particularly in industries with flexible or non-traditional career paths. Similarly, Eagly and Carli [39] highlight that men often navigate career "labyrinths" with fewer constraints than women, who may face family-related responsibilities or workplace biases that limit mobility.

Examination of the count component in the ZIP model, as shown in Table 2, highlights a significant effect of age on job transitions, with a negative relationship observed (coefficient = -0.035, p = 0.000, IRR = 0.966). This result, significant at the 1% level (p = 0.000), indicates that for each additional year of age, the incidence rate of job transitions decreases by 3.4%, suggesting older workers are less likely to change jobs. This significant effect may be driven by several factors, including a preference for stability among older workers, who may prioritize job security, established routines, or accumulated benefits over the uncertainty of new roles. Additionally, older workers might face reduced opportunities for mobility due to age-related biases, declining physical capacity, or fewer vacancies suited to their experience, particularly in dynamic sectors like tourism [40, 41, 42]. This finding aligns with prior research, such as Ng and Feldman [43], which demonstrates that older employees tend to exhibit lower job mobility, often due to stronger ties to current roles and a reduced inclination to take career risks. Similarly, Mincer and Jovanovic [44] argue that job tenure increases with age, as workers invest more in firm-specific human capital, deterring transitions.

As shown in Table 2, the ZIP model reveals a nuanced role of foreign language proficiency in job transitions among tourism workers in Malang Regency. In the count component, the coefficient (0.351, p = 0.102, IRR = 1.420) suggests proficient workers experience a 42% higher transition rate than non-proficient peers, though the effect is statistically insignificant (p > 0.10). In contrast, the zero-inflated component indicates a marginally significant effect (coefficient = 0.906, p = 0.088, OR = 2.473), implying proficient workers are 2.473 times more likely to have zero transitions. This dual pattern suggests that multilingual workers may either remain in stable, high-demand roles (e.g., tourism guides) or leverage language skills for increased mobility in specific contexts [28, 45, 46, 47]. The limited impact on transition frequency may reflect variability in language skill demand across industries, where proficiency is valuable in niche roles like international tour guides but less influential when broader factors like experience or job availability dominate [48, 49].

### 3.3 Model Comparison

In Table 3, the Zero-Inflated Poisson (ZIP) model was evaluated against the standard Poisson model to assess their fit for job transition data in Malang, accounting for excess zeros and overdispersion. The ZIP model demonstrates a better fit, with an AIC of 1398.787 compared to 1414.506 for the Poisson model, and a BIC of 1436.917 versus 1443.103, indicating superior performance in balancing goodness of fit and model complexity. The ZIP model's higher log-likelihood (-691.39 vs. -701.25) further suggests improved predictive accuracy, as it better captures the observed data's structure, particularly for zero transitions.

**Table 3.** Comparison of Model Fit for ZIP and Poisson Models in Job Transitions Analysis

| Model | Degrees of Freedom | AIC | BIC | Log-Likelihood |
|---|---|---|---|---|
| ZIP Model | 8 | 1398.787 | 1436.917 | -691.39 |
| Poisson Model | 6 | 1414.506 | 1443.103 | -701.25 |

Overdispersion is evident in the data as stated previously, with a variance-to-mean ratio of 1.339 and a significant overdispersion test result ($p=0.000246$), confirming the presence of excess variability beyond the Poisson assumption. The ZIP model effectively addresses this by incorporating a zero-inflated component, capturing the high proportion of workers with no job transitions, and a count component, modeling the frequency of transitions.

## 4. CONCLUSION

This study on job transitions among tourism workers in Malang Regency highlights the complex interplay of factors shaping worker mobility and stability. By employing a Zero-Inflated Poisson (ZIP) model, the analysis effectively accounted for excess zeros in the data, revealing patterns that a standard Poisson or linear regression model could not capture. The ZIP model distinguished workers with no transitions (structural zeros), primarily driven by foreign language proficiency promoting stability, from those with transitions, influenced by gender and age. Specifically, proficient workers tend to remain in stable roles, while men and younger workers are more likely to change jobs. In contrast, relationship to the head of household and educational attainment were not significant explanatory variables of job transitions, underscoring the dominance of individual attributes and tourism-specific demands in shaping labor market dynamics.

These findings offer actionable implications for workforce planners through Indonesia's Sertifikasi Kompetensi SDM Pariwisata program. To address gender disparities, training programs should include mentorship and leadership workshops tailored for female workers, particularly in roles like tour guiding or hospitality management. For foreign language proficiency, vocational centers should provide subsidized, role-specific language courses (e.g., English for international tour guides) to enhance mobility for non-proficient workers. To support older workers, retraining initiatives focusing on digital skills or emerging tourism niches (e.g., eco-tourism) could boost employability. These strategies align with government efforts to strengthen the resilience of Malang's tourism labor market.

Despite its contributions, this study has limitations. The cross-sectional data limits causal inferences about job transitions, and findings may not fully generalize beyond Malang's tourism sector. Future research could employ longitudinal data to track mobility patterns over time or explore subsectors (e.g., hospitality vs. adventure tourism) to uncover sector-specific dynamics. Testing alternative models, such as Zero-Inflated Negative Binomial or hurdle models, could further refine the understanding of transition complexity. By addressing these gaps, future studies can build on this work to inform more robust labor market policies.

## 5. REFERENCES


[1] World Tourism Organization, UNWTO Tourism Highlights: 2017 Edition, Madrid: UNWTO, 2017.
[2] BPS-Statistics of Malang Regency, Statistik Daerah Kabupaten Malang 2023 [Regional Statistics of Malang Regency 2023], BPS-Statistics of Malang Regency, 2023.
[3] H. Dewhurst, P. Dewhurst dan R. Livesey, “Tourism and Hospitality SME Training Needs and Provision: A Sub-Regional Analysis,” vol. 7, no. 2, pp. 131-143, 2007, doi: 10.1057/palgrave.thr.6050037.
[4] N. G. Mankiw, Macroeconomics, 11 penyunt., Worth Publishers, 2022.
[5] World Economic Forum, The Future of Jobs Report 2023, Geneva: World Economic Forum, 2023.
[6] Y. E. Akgündüz, A. Aldan, Y. K. Bağır dan H. Torun, “Job mobility in Turkey,” Central Bank Review, vol. 19, no. 3, pp. 83-91, 2019, doi: 10.1016/j.cbrev.2019.08.002.
[7] G. J. Borjas, Labor Economics, 8 penyunt., McGraw-Hill Education, 2019.
[8] A. Doris, D. O'Neill dan O. Sweetman, “Why Do the Earnings of Male and Female Graduates Diverge? The Role of Motherhood and Job Dynamics,” IZA Discussion Papers No. 15805, 2022.
[9] K. Hutchings, C.-l. Moyle, A. Chai, N. Garofano dan S. Moore, “Segregation of women in tourism employment in the APEC region,” Tourism Management Perspectives, vol. 34, pp. 1-15, 2020, doi: 10.1016/j.tmp.2020.100655.
[10] S. Rana dan R. Shaikh, “The Role of English Speaking- Skills in Career Progression: A Case Study among Sudanese Undergraduate EFL Students,” World Journal of English Language, vol. 14, no. 2, pp. 349-357, 2024, doi: 10.5430/wjel.v14n2p349.
[11] M. M. Roshid dan R. Chowdhury, “English language proficiency and employment: A case study of Bangladeshi graduates in Australian employment market,” Mevlana International Journal of Education, vol. 3, no. 1, pp. 68-81, 2013, doi: 10.13054/mije.13.06.3.1.
[12] I. Booyens, “Education and skills in tourism: Implications for youth employment in South Africa,” Development Southern Africa, vol. 37, no. 5, p. 825–839, 2020, doi: 10.1080/0376835X.2020.1725447.
[13] A. C. Cameron dan P. K. Trivedi, Regression analysis of count data, 2nd penyunt., New York: Cambridge University Press., 2013.
[14] J. M. Hilbe, Modeling Count Data, Cambridge University Press, 2014.
[15] D. Lambert, “Zero-Inflated Poisson Regression, with an Application to Defects in Manufacturing,” Technometrics, vol. 34, no. 1, pp. 1-14, 1992, doi: 10.2307/1269547.
[16] D. D. Le, R. L. Gonzalez dan J. U. Matola, “Modeling count data for health care utilization: an empirical study of outpatient visits among Vietnamese older people,” BMC Med Inform Decis Mak., vol. 21, no. 265, pp. 1-14, 2021, doi: 10.1186/s12911-021-01619-2.
[17] K. K. W. Yau dan A. H. Lee, “Zero-inflated Poisson regression with random effects to evaluate an occupational injury prevention programme,” Statistics in Medicine, vol. 20, pp. 2907-2920, 2001, doi: 10.1002/sim.860.
[18] D. C. Heilbron, “Zero-Altered and other Regression Models for Count Data with Added Zeros,” Biometrical Journal, vol. 36, no. 5, pp. 531-547, 1994, doi: 10.1002/bimj.4710360505.
[19] Y. B. Cheung, “Zero-inflated models for regression analysis of count data: a study of growth and development,” Statistics in Medicine, vol. 21, no. 10, pp. 1461-1469, 2002, doi: 10.1002/sim.1088.
[20] W. H. Greene, Econometric Analysis, 5 penyunt., New Jersey: Pearson Education, Inc., 2003.
[21] A. Heitmueller, “Job Mobility in Britain: Are the Scots different? Evidence from the BHPS,” IZA Discussion Paper No. 773, pp. 1-31, 2003, doi: 10.1111/j.0036-9292.2004.00309.x.
[22] J. A. List, “Determinants of securing academic interviews after tenure denial: evidence from a zero-inflated Poisson model,” Applied Economics, vol. 33, no. 11, pp. 1423-1431, 2001, doi: 10.1080/00036840010009856.
[23] N. Motalebi, M. S. Owlia, A. Amiri dan M. S. Fallahnezhad, “Monitoring social networks based on Zero-inflated Poisson regression model,” Communications in Statistics - Theory and Methods, vol. 52, no. 7, pp. 2099-2115, 2023, doi: 10.1080/03610926.2021.1945103.
[24] B.-C. Truong, P. K. Hung, C.-C. Dinh dan M. J. McAleer, “Zero-inflated poisson regression models: Applications in the sciences and social sciences,” Annals of Financial EconomicsVol. 16, No. 02, 2150006 (2021), vol. 16, no. 2, pp. 1-19, 2021, doi: 10.1142/S2010495221500068.
[25] A. S. Argawu dan G. G. Mekebo, “Zero-inflated Poisson regression analysis of factors associated with under-five mortality in Ethiopia using 2019 Ethiopian mini demographic and health survey data,” PLoS ONE, vol. 18, no. 11, pp. 1-15, 2023, doi: 10.1371/journal.pone.0291426.
[26] A. Zeileis, C. Kleiber dan S. Jackman, “Regression Models for Count Data in R,” Journal of Statistical Software, vol. 27, no. 8, pp. 1-25, 2008, doi: 10.18637/jss.v027.i08.
[27] J. S. Long, Regression Models for Categorical and Limited Dependent Variables, Sage Publications, Inc, 1997.
[28] M. Deepa, T. Golda, J. A. Meeran, V. Sivalanka dan P. S. Priya, “The Role of English Language Proficiency in Career Advancement: A Review of Empirical Studies,” Journal of Harbin Engineering University, vol. 44, no. 8, pp. 1-8, 2023.

[29] A. Agresti, An Introduction to Categorical Data Analysis, 3 penyunt., Hoboken, NJ: John Wiley & Sons, Inc., 2019.
[30] J. Zhang, Y. Yang dan J. Ding, "Information criteria for model selection," WIREs Computational Statistics, vol. 15, no. 5, pp. 1-27, 2023, doi: 10.1002/wics.1607.
[31] H. Akaike, "Information theory and an extension of the maximum likelihood principle," dalam Second International Symposium on Information Theory, Budapest, 1973.
[32] G. Schwarz, "Estimating the Dimension of a Model," Annals of Statistics, vol. 6, no. 2, p. 461–464, 1978, doi: 10.1214/aos/1176344136.
[33] A. Chakrabarti dan J. K. Ghosh, "AIC, BIC and Recent Advances in Model Selection," Philosophy of Statistics, vol. 7, pp. 583-605, 2011, doi: 10.1016/B978-0-444-51862-0.50018-6.
[34] M. F. Rahmatika, M. Pudjihardjo dan Sasongko, "Tourism Employment Changes in Indonesia," International Journal of Business, Economics and Law, vol. 21, no. 5, pp. 173-177, 2020.
[35] A. Grobelna dan K. Skrzeszewska, "Seasonality: Is it a Problem or Challenge Facing Future Tourism Employment? Implications for Management," Journal of Entrepreneurship, Management and Innovation, vol. 15, no. 1, pp. 205-230, 2019, doi: 10.7341/20191518.
[36] P. M. Blau dan O. D. Duncan, The American Occupational Structure, New York: Free Press, 1967.
[37] G. S. Becker, Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education, University of Chicago Press, 1993.
[38] P. England, "The Gender Revolution: Uneven and Stalled," Gender and Society, vol. 24, no. 2, pp. 149-166, 2010, doi: 10.1177/0891243210361475.
[39] A. H. Eagly dan L. L. Carli, Through the labyrinth: the truth about how women become leaders, Boston: Harvard Business Review Press, 2007.
[40] A. H. D. Lange, B. V. d. Heijden, T. V. Vuuren, T. Furunes, C. D. Lange dan J. Dikkers, "Employable as We Age? A Systematic Review of Relationships Between Age Conceptualizations and Employability," Frontiers in Psychology, vol. 11, pp. 1-14, 2021, doi: 10.3389/fpsyg.2020.605684.
[41] A. Tansel dan Z. A. Ozdemir, "Transitions across labor market states including formal/informal division in Egypt," Review of Development Economics, vol. 23, no. 4, pp. 1-22, 2019, doi: 10.1111/rode.12620.
[42] L. Sugiharti, N. Aditina dan M. A. Esquivias, "Worker Transition Across Formal and Informal Sectors: A Panel Data Analysis in Indonesia," Asian Economic and Financial Review, vol. 12, no. 11, p. 923–937, 2022, doi: 10.55493/5002.v12i11.4635.
[43] T. W. H. Ng dan D. C. Feldman, "Age, work experience, and the psychological contract," Journal of Organizational Behavior, vol. 30, no. 8, p. 1053–1075, 2009, doi: 10.1002/job.599.
[44] J. Mincer dan B. Jovanovic, "Labor Mobility and Wages," dalam Studies in Labor Markets, National Bureau of Economic Research, Inc., 1981, pp. 21-64.
[45] B. R. Chiswick dan P. W. Miller, The Economics of Language: International Analyses, London: Routledge, 2007.
[46] C. Dustmann dan F. Fabbri, "Language Proficiency and Labour Market Performance of Immigrants in the UK," The Economic Journal, vol. 113, no. 489, pp. 695-717, 2003.
[47] H. G. Bloemen, "Language proficiency of migrants: the relation with job satisfaction and skill matching," Empirical Economics, vol. 65, p. 2899–2952, 2023, doi: 10.1007/s00181-023-02445-3.
[48] A. Ladkin, "Exploring tourism labor," Annals of Tourism Research, vol. 38, no. 3, pp. 1135-1155, 2011, doi: 10.1016/j.annals.2011.03.010.
[49] A. Ladkin, S. Mooney, D. Solnet, T. Baum, R. Robinson dan H. Yan, "A review of research into tourism work and employment: Launching the Annals of Tourism Research curated collection on tourism work and employment," Annals of Tourism Research, vol. 100, 2023, doi: 10.1016/j.annals.2023.103554.