Pesquisa Operacional (2024) 44: e282287 p.1-25
doi: 10.1590/0101-7438.2023.043.00282287

Printed version ISSN 0101-7438 / Online version ISSN 1678-5142
www.scielo.br/pope
ARTICLES

OCCURRENCE OF RAILWAY ACCIDENTS RELATED TO SOME PERSONAL AND PROFESSIONAL TRAIN CONDUCTOR FACTORS: USE OF STATISTICAL MODELS IN THE PRESENCE OF EXCESS OF ZEROS

Felipe Mattos Tavares[1], Jorge Alberto Achcar[2], José Luíz Garcia Hermosilla[3] and Emerson Barili[4*]

Received January 15, 24 / Accepted May 21, 24

ABSTRACT. The main goal of this research was to identify personal and professional factors of train drivers which imply in the occurrence of accidents (denoted in this study as situations to non-compliance with procedures of the train driver consisting of unsafe acts that could cause major accidents) of a Brazilian logistics company related to accidentes in the period from 2014 to 2016. The research involved 348 drivers and some independent variables associated with each worker such as distances of removal (railway sections), driver's age, working time (time working in the company), marital status, number of trips (number of train trips as a driver); months of work (months of work in the company) and total hours driving trains. Responses of interest are accident rates per work month and accident counts. Under a Bayesian approach, Beta and Poisson regression models were assumed in the presence of excess zeros. For the accident rate/months of work response, the covariates number of trips, posting distances and age showed some evidence of significant effects. For the response accident count, the covariate hours driving trains shows some evidence that it is a significant covariate. These results may be of great interest to rail logistics company managers to improve rail safety using some Bayesian modeling approaches in the presence of excess zeros.

Keywords: railroad accidents, beta regression in the presence of excess of zeros, bayesian analysis.

1 INTRODUCTION

Trains usually are a safe way to travel. However, railway disasters could happen in many occasions. Each year there are thousands of accidents worldwide with injuries involv-

*Corresponding author

[1]University of Araraquara, Department of Production Engineering, Araraquara, SP, Brazil – E-mail: fmtavares260384@gmail.com – http://orcid.org/0000.0002.0921.7666

[2]University of São Paulo, Ribeirão Preto Medical School, Ribeirão Preto, SP, Brazil – E-mail: achcar@fmrp.usp.br – http://orcid.org/0000.0002.9868.9453

[3]University of Araraquara, Department of Production Engineering, Araraquara, SP, Brazil – E-mail: jlghermosilla@hotmail.com – http://orcid.org/0000.0002.5104.6725

[4]University of São Paulo, Ribeirão Preto Medical School, Ribeirão Preto, SP, Brazil – E-mail: ebarili2@gmail.com – http://orcid.org/0000.0001.8210.6746

ing trains, and hundreds of individuals are killed in these types of accidents. In general, some of the most common causes of train accidents (https://www.burge-law.com/what-are-the-causes-of-railway-accidents/) include:

- **Train operator error:** Human error could be an important factor in railway accidents (poor training, inexperience, reckless behavior, or a combination of these). As reckless behavior we could include operating the train too fast.
- **Track problems:** Track owners are responsible for keeping their tracks maintained and in good repair.
- **Lack of warning signals:** In some parts of the railway could be lack of warning signals where motorists, bicyclists, and pedestrians may not realize that a train is coming.
- **Warning signal defects:** The warning signals could have not been maintained properly or there are malfunctions.
- **Obstructed view of the railroad crossing:** Sometimes trees and other vegetation become overgrown, which could obstruct the view of the crossing.
- **Stalled vehicles:** Some train-vehicle crashes could happen because a vehicle gets stalled on top of the track, often due to a mechanical failure.
- **Distractions:** Some railway accidents are caused by distractions such as sending texts or other smartphone activity.
- **Faulty Equipment:** A train accident could happen due to some type of mechanical defect.

Other factors, such as organizational aspects, supervision characteristics, physical and technological factors, conditions of the operator (such as mental and physical state and limitations) and of the team Shappell & Wiegmann (2000) can cause train accidents.In this study, train accidents denote situations to non-compliance with procedures.Non-compliance with procedures, treated in the literature as violation of procedures, errors and unsafe acts. The violations of procedures can have different results, ranging from a violation without no consequences, going through violations that cause incidents and reaching violations that cause major accidents. Some studies (Tavares et al. (2021); Evans (2011); Kyriakidis et al. (2015)) relate unsafe acts by train drivers (unintentional failures in the mental or physical activities of individuals) and violations (disobedience of existing operational procedures in the organization) such as main cause of accidents in the railway sector. Many other studies on railway accidents and their causes are introduced in the literature. Wasnik (2010) introduced an analysis of railway fatalities in Central India; San Kim & Yoon (2013) considered an accident caution model for the railway industry with application of the model to 80 rail accident investigation reports from the UK; Shi et al. (2020) considered a correlation analysis of causes of railway accidents based on an mathematical algorithm; Wang et al. (2020) considered a study on correlation factors on railway accidents using association rule learning algorithms; Aher & Tiwari (2018) studied railway accidents in India considering impacts of causes, effects and management; Bala & Bhasin (2018) introduced a review on analysis of railway traffic accident using data mining technique; Dhaygude et al. (2019b,a) introduced different statistical analysis of railway accidents; Tavares et al. (2021) considered a study on the worker's profile and its relationship with the occurrence of unsafe acts assuming the case of train drivers of a logistics company in Brazil.

Hong et al. (2023) carried out a literature review analysis on the causality analysis of railway accidents, investigating the application of Natural Language Processing (NLP) to assist in the analysis; Rad et al. (2023) presented the lack of a comprehensive review of the literature on systemic modeling of railway accidents, analyzing accidents on railways from 2000 to 2022; Wang et al. (2023) proposed a modeling method to analyze the correlation of hazards in railway accidents based on graph knowledge theory identifying several important hazards; Liu et al. (2024) proposed a method called Comprehensive-Biased Random Walk with Different Restart (CBDRWR) in an analysis of the potential risk of the railway accident generation process; Yan et al. (2023) presented a railway accident prevention method based on the reinforcement learning model and multi-modal data to achieve active railway accident prevention strategies. In Brazil, the performance of this type of transport is worrying, as it is a country that has a railway network in operation of approximately 30,000 km (Murta et al., 2023), almost all are freight trains, and has a railway accident rate higher than the accident rate railways in the European Union, which has a railway network of 250,000 km. We observe that both personal, technical and structural investments in Brazilian railways are not sufficient to contain railway accidents. In this study, we address some important issues that concern the personal and professional characteristics of train conductors, such as understanding the level of stress, fatigue, decision-making capacity, as well as the emotional state that could affect the occurrence of non-compliance with procedures by the train conductors which could lead to accidents. Once the characteristics of the conductor are known, it is possible to establish training and management programs for such circumstances. Promoting safety in the railway sector requires a broad knowledge of the factors that contribute to a significant reduction in accidents on railway lines, so it is viable to promote a culture that values and encourages safety practices among railway professionals. Another important factor for safer rail travel would be the drafting and implementation of laws and regulations based on the personal characteristics of rail operators. Among the studies on Brazilian railways, we can highlight the study introduced by Georgiou (2009), who mapped the dynamics in the investigation of Brazilian railway development, identifying two problems: the misappropriation of public resources and a degenerative feedback system in decision-making; Keretch & De Paiva (2016) introduced a study on the number of accidents on railways due to different causes in Brazil. They observed that despite the increase in products transported, the number of accidents decreased from 1,638 in 2006 to 866 in 2013 according to a report from the Brazilian National Land Transport Agency (ANTT). In other study, Araújo & da Silva Sousa (2023) analyzed the reasons for the eradication of many railway branches in the last decades under a economy perspective. Souza et al. (2019) presented a study showing that the assembly of the wheelset, installation of the engine and traction assembly and wheel machining are the main failures that cause locomotives to derail; Souza et al. (2023) also studied important factors that could possibly contribute to the occurrence and severity of these accidents. It is known that the maintenance process is essential for whatever the means of transport to avoid possible accidents, however, among the railway maintenance proposals, we can highlight the study by Arruda et al. (2022) who evaluated a possible "Hot Box" failure, which is a failure that arises through temperature conditions in the bearings. de Almeida Eleutério & Rosa (2023) proposed a mathematical model to plan resource routes to meet the maintenance order, maximizing the number of maintenance orders fulfilled in the planned period.

In this study, we consider as the main objective of research, to discover possible relationships between personal and professional characteristics of train drivers and some factors related to the structure of a railroad located in southern Brazil with the occurrence of accidents. The quantitative research involved 348 train conductors related to the occurrence of accidents in the period of years 2014/2016 and some independent variables (factors) associated with each train driver. The responses of interest studied are the accident rates per month of work (number of accidents/months of work) and total accident count for each train conductor. The dataset shows excess of zeros, that is, many workers had no accidents in the period, which is common in the railway area where accidents are not so frequent. As the accident rates per month of work (number of accidents/months of work) are defined in the interval (0,1), we assume beta regression models adapted for the presence of excess of zeros for data analysis (zero-inflated Beta or ZIB model). For the accident count response by each worker, we assume a Poisson regression model also in the presence of excess of zeros (zero-inflated Poisson or ZIP model). Given the difficulty to obtain usual classical inferences (maximum likelihood estimators), we use a Bayesian inference approach and MCMC (Markov Chain Monte Carlo) simulation methods to simulate samples from the joint posterior distributions of interest.

The article is organized from here as follows: Section 2 presents the data and a preliminary statistical analysis; Section 3 introduces the proposed methodology; Section 4 presents the obtained results; finally, Section 5 presents some concluding remarks.

2 DATA AND PRELIMINARY STATISTICAL ANALYSIS

The data set has information on 348 train conductors related to the occurrence of accidents and information on some personal and professional covariates associated with each worker (independent variables) as detachment distances (railway sections), train conductor age, job time (time working in the company), marital status (1: married; 2: single), trip count (number of train rides as conductor); work months (months working in the company) and total hours conducting trains (Appendix 1). The responses of interest studied are the accident rates per month of work (number of accidents/months of work) and total accident count for each driver.

As a preliminary analysis of the data, we initially consider the binary responses (occurrence or not of accidents) denoting as success the non-occurrence of accidents for the train conductor (success denoted as $Y = 1$) and as failure the occurrence of one or more accidents for the train conductor (failure denotes as $Y = 0$) considering the total data set ($n = 348$ observations) where the binary random variable Y has a Bernoulli distribution with probability of success p given by $P(Y = y) = p^y(1-p)^{n-y}$, $y = 0$ or 1 and a logistic regression model (see for example, Montgomery & Runger (2010)) given by,

$$\begin{aligned} \text{logit}(p_i) =& \log[p_i/(1-p_i)] = \beta_0 + \beta_1 \text{detachment.distances}_i + \beta_2 \text{age}_i + \beta_3 \text{job.time}_i + \\ & \beta_4 \text{marital.status}_i + \beta_5 \text{trip.count}_i + \beta_6 \text{work.months}_i + \\ & \beta_7 \text{hours.conducting.trains}_i \end{aligned} \tag{1}$$

where, $i = 1, 2, \ldots, 348$.

From the data set, we observed 135 failures (train conductors with accidents) and 213 successes (train conductors without accidents). Table 1 shows the results (maximum likelihood estimators-MLE) of the regression parameters associated with the covariates detachment distances, train conductor age, job time, marital status, trip count, work months and total hours conducting trains (use of the Minitab software®).

From the results in Table 1, we observe that the covariates detachment distances, trip count and work months show significant effects on the probabilities of no occurrence of accidents since the associated p-values are smaller than 0.05. From the signals of the MLE in each case, we conclude that larger detachment distances (negative MLE estimator) implies in smaller probabilities p of no occurrence of accidents (possible long distances with few stops increase the speed of the trains implying in higher chance for accidents), showing an increase on the probability $1 - p$ to have accidents; also larger trip counts (negative MLE estimator) implies in smaller probabilities of a train conductor do not have accidents (increasing the probability to have accidents associated to great number of trips, possibly leading to stress of the train conductor). In the contrary, increasing work months (months working in the company) increases (positive estimator) the probability of a train conductor do not have accidents (more experience of the conductor increases the probability of success, that is, the probability of no occurrence of accidents).

Table 1 – MLE, standard errors (SE) and p-values (logistic regresssion).

covariate	Coef	SE Coef	p-value
detachment distances	-0.0019	0.00085	0.0230
age	0.0072	0.02400	0.7650
job time	-0.0027	0.05910	0.9640
marital status	-0.1240	0.29200	0.6700
trip count	-0.0034	0.00106	0.0010
work months	0.1411	0.02600	< 0.0010
hours conducting trains	-0.0470	0.16500	0.7760

Since the main goal of this study is related to the responses to accident rates per month of work (number of accidents/months of work) and total accident count for each driver associated with the covariates detachment distances, train conductor age, job time, marital status, trip count, work months and total hours conducting trains, we need to use more elaborated statistical models related to rates and count of accidents. In this way, we assume beta regression models for the rates and Poisson regression models for the accident counts for each train conductor.

As the dataset shows excess of zero responses, we need to use existing zero-inflated statistical models which consider the data as a mixture of observations with one component consisting of zero responses and another component consisting of non-zero responses, where we need to check the possible dependences between the responses (rates or count) associated with each covariate. Figure 1 shows the scatter plots of the two responses (accident rates per month of work and total accident count for each driver) associated with each covariate only considering the observations with non-zero count of accidents.

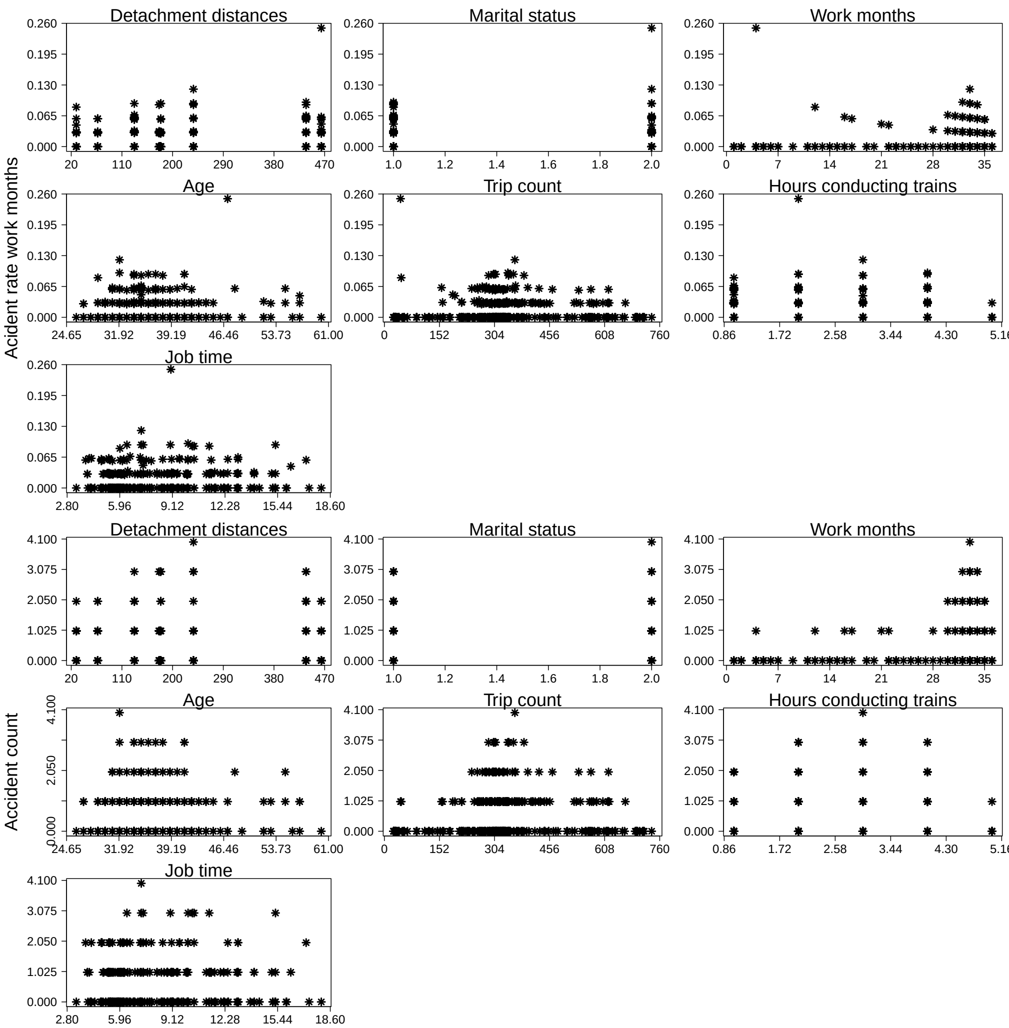

Figure 1 – Scatter plots (accident rates by months worked and total accident counts for each conductor) associated to each covariate.

From Figure 1, we observe that it is difficult to conclude which covariates affect the responses (accident rates per month worked and total accident counts for each train conductor). Possibly trip count and detachment distances affect the response accident rates per month worked and total accident count for each train conductor affect the response count of accidents, but it is needed a good statistical model to discover possible covariates affecting the two responses.

The main goals of this study are:

- To verify statistically if some covariate affects the response given by the accident rates per month worked assuming a beta regression model adapted for the presence of excess zeros for data analysis (zero-inflated Beta or ZIB model).
- To verify statistically if some covariate affects the response given by the number of accidents assuming a Poisson regression model adapted for the presence of excess zeros for data analysis (zero-inflated Poisson or ZIP model).

3 METHODS

In this section, we present the statistical models used in the data analysis.

3.1 The zero-inflated Poisson (ZIP) model

An important assumption of the Poisson distribution is that the variance of the count outcome is equal to the mean. In practical work this assumption could be not verified, that is, we have 'overdispersion'. The zero-inflated Poisson (ZIP) is an alternative to deal with this problem. This model assumes that there are two different types of individuals in the data:

- Individuals with zero count (no occurrence of accidents) with a probability p (0 - group).
- Individuals with counts (number of accidents different of zero) that could be predicted by the standard Poisson distribution (not 0 - group).

We could have zero count from each one of the two groups: if the zero is from the 0- group, it indicates that the observation is free from the probability of having a positive outcome Scott Long (1997); Hall (2000). The overall model is a mixture of the probabilities from the two groups, which allows for both the overdispersion and excess zeros that cannot be predicted by the standard Poisson model.

The binary outcome to be in the 0 - group could be modeled by a binary Bernoulli distribution with success probability p. The probability of outcome not be in the 0 - group is given by $1-p$. For an observation not belonging to the 0 - group, we could assume a standard Poisson distribution with mass probability function given by

$$f(y) = P(Y = y) = \frac{e^{-\mu}\mu^{y}}{y!}, \quad y = 1, 2, 3, \ldots \tag{2}$$

where μ is the conditional mean given the outcome belong to the not 0 - group.

In this way, the mixed probabilities for ZIP are expressed as follows:

- Zero counts in 0 - group: $P(Y=0)=p$
- Non zero counts in not 0 - group: $P(Y=y)=(1-p)e^{-\mu}\mu^y/y!$
- Overall, we have,

$$P(Y=y)=p \text{ if } y=0 \text{ and } P(Y=y)=(1-p)e^{-\mu}\mu^y/y!\,, \text{ if } y>0 \tag{3}$$

Since $0\leq p\leq 1$, the overall mean of the ZIP given by $E(Y)=\mu(1-p)$ is smaller than the conditional mean μ. The ZIP structure also shows overdispersion, since the overal variance is given by $var(Y)=\mu(1-p)(1+\mu p)$ (see Erdman et al. (2008)).

Assuming an indicator variable $\delta=1$ if $Y=0$ and $\delta=0$ if $Y>0$, the contribution of one observation to the likelihood function for μ and p is given by,

$$L(\mu,p)=(p)^{\delta}[(1-p)e^{-\mu}\mu^y/y!]^{1-\delta} \tag{4}$$

In presence of a vector of p covariates $\mathbf{x}=(x_1,x_2,\ldots,x_p)$, we assume the regression model, $\mu=\exp(x'\beta)$, where $x'\beta=\beta_0+\beta_1x_1+\beta_2x_2+\cdots+\beta_px_p$.

3.2 The zero-inflated Beta (ZIB) model

Following the same arguments presented in Section (3.1) for the ZIP model, we now assume a Beta distribution for the rates (accident rates by months of work) that is a continuous random variable defined in the interval (0,1) . The probability density function for the Y (rate) assuming a Beta distribution, is given by,

$$f(y)=cy^{a-1}(1-y)^{b-1}, \quad 0<y<1 \tag{5}$$

where c is the Beta function, given by $c=B(a,b)=\Gamma(a+b)/\Gamma(a)\Gamma(b)$ and the conditional mean given the outcome belong to the not 0 - group is given by $\mu=E(Y)=a/(a+b)$. The conditional variance is given by, $var(Y)=ab/[(a+b)^2(a+b+1)$.

In this way, the mixed probabilities for ZIB are expressed as follows:

- For the zero counts in 0 - group we have the probabilities:

$$P(Y=0)=p \quad \text{and} \quad P(Y>0)=1-p$$

- For the non zero counts in not always-0 group, we have the probability density function:

$$f_1(y)=(1-p)\Gamma(a+b)/\Gamma(a)\Gamma(b)y^{a-1}(1-y)^{b-1}$$

- Overall

$$P(Y = y) = p \quad \text{if} \quad y = 0 \quad \text{and} \quad P(Y = y) = (1 - p)cy^{a-1}(1 - y)^{b-1} \quad \text{if} \quad y > 0 \tag{6}$$

Since $0 \leq y \leq 1$, the overall mean of the ZIB model is given by $E(Y) = \mu(1 - p)$, where $\mu = a/(a+b)$.

Assuming an indicator variable $\delta = 1$ if $Y = 0$ and $\delta = 0$ if $Y > 0$, the contribution of one observation to the likelihood function for a, b and p is given by,

$$L(a,b,p) = (p)^{\delta}[(1 - p)cy^{a-1}(1 - y)^{b-1}]^{1-\delta} \tag{7}$$

In presence of a vector of covariates associated to each unit, it is assumed a regression model considering a reparametrized form for the beta distribution with density (5) given by, $\mu = a/(a + b)$ and $\Phi = a + b$ (Ferrari & Cribari-Neto (2004), Jørgensen (1997), da Silva et al. (2021)). In this way, we have, $a = \Phi\mu, b = (1 - \mu)\Phi$, $E(Y) = \mu$ and $var(Y) = V(\mu)/(1 + \Phi)$ where $V(\mu) = \mu(1 - \mu)$, so that μ is the mean of the response variable and Φ can be interpreted as a precision parameter in the sense that, for fixed μ, the larger the value of Φ, the smaller the variance of Y. The probability density function of the random variable Y can be written, in the new parameterization, as,

$$f(y/\mu,\Phi) = \Gamma(\Phi)\Gamma(\Phi\mu)\Gamma[(1 - \mu)\Phi]y^{\Phi\mu-1}(1 - y)^{(1-\mu)\Phi-1} \tag{8}$$

where $0 < \mu < 1$ and $\Phi > 0$.

Assuming the presence of a covariate vector $\mathbf{x} = (x_1, x_2, \ldots, x_p)$' with p covariates associated to each observation, it is assumed the following regression model for the mean Cepeda-Cuervo et al. (2014),

$$\text{logit}(\mu) = \log[\mu/(1 - \mu)] = \beta' x = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p \tag{9}$$

where $\beta = (\beta_0, \beta_1, \beta_2, \ldots, \beta_p)$' is a vector of regression parameters.

We assume a Bayesian analysis for the data assuming both classes of assumed models (accident rates and accident counts). Combining the joint prior distribution for the parameters of each assumed model, the joint posterior distribution for the parameters of the model is obtained using the Bayes formula Box & Tiao (1973). The posterior summaries of interest are obtained using Markov Chain Monte Carlo (MCMC) simulation methods as the popular Gibbs sampling algorithm or the Metropolis-Hastings algorithm (Gelfand & Smith (1990); Chib & Greenberg (1995)) using the free existing OpenBugs software (Lunn et al. (2000)).

4 RESULTS

From the data of 348 train conductors, we observed 213 observations (train conductors) with zero occurrences of accidents, that is, workers without accidents. Thus, the proportion of values equal to zero is 0.6121 (61.21%) and the proportion of individuals with accidents is given by $1 - p = 0.3879(38.79\%)$.

4.1 First response of interest: accident rate per months worked

As all observed accident rates per months worked are given in the interval (0,1), that is, the number of accidents per worker (a rare event) is always less than the months worked in the company, we assume the rates assumed by a logit transformation, that is, the responses are given by y = log[rate/(1-rate)].

Initially we assume the beta regression model defined in Section 3.2 without assuming the presence of covariates, defined by (5) and (6). For a Bayesian analysis, we assume uniform independent prior distributions, that is, $a \sim U(0.100)$, $b \sim U(0.1000)$ and $p \sim U(0.1)$ where $U(\alpha, \beta)$ denotes a uniform distribution in the interval (α, β). Thus we are assuming non-informative prior distributions for the parameters a, b and p. Using the Openbugs software, we initially generated 11,000 Gibbs samples, discarded to eliminate the effect of initial values in the iterative procedure of simulating samples of the joint posterior distribution for a, b and p. Next, we simulated another 10,000 samples by choosing each 10th generated sample, totaling 1,000 samples to be used to get the posterior summaries of interest. The convergence of the simulation algorithm was verified from graphs of the samples generated for each parameter. Table 2 shows the posterior summaries of interest (posterior means, posterior standard deviations and 95% credibility intervals for each parameter). The posterior means are the estimators of the parameters obtained by assuming a quadratic loss function.

The conditional mean (only for workers where accidents are observed) of the beta distribution (4) is given by $a/(a+b) = 4.571/(4.571+98.22) = 0.04446887$. The conditional sample mean obtained from the data is given by, 5.99722/153 = 0.04442385. The proportion of zeros estimated by the model is given by 0.6113. As the proportion of zeros in the sample is given by 0.6121, we conclude that the proposed model is well fitted by the data.

Table 2 – Posterior summaries (accident rates/months of work without covariates).

	Mean	S.D.	Lower 95% c.i	Upper 95% c.i
a	4.571	0.517	3.632	5.711
b	98.220	11.480	76.640	124.100
c	0.6113	0.0249	0.5642	0.6572

The non-conditional sample mean is given from the data by 0.0172334 and the non-conditional mean estimated by the model is given by, $\mu(1-p) = 0.0172850$ where $\mu = a/(a+b) = 4.571/(4.571+98.22) = 0.04446887$, again indicating the excellent fit of the model to the data.

With the presence of the covariates detachment distances, conductor age, conductor company time, marital status, amount of train rides, sum of work months and hours conducting trains, we assume the regression model defined by (5), (6) , (8)) and (9), that is,

$$\begin{aligned}\text{logit}(\mu_i) = &\log[\mu_i/(1-\mu_i)] = \beta_0 + \beta_1 \text{detachment.distances}_i + \beta_2 \text{age}_i + \\ &\beta_3 \text{job.time}_i + \beta_4 \text{marital.status}_i + \beta_5 \text{trip.count}_i + \beta_6 \text{work.months}_i + \\ &\beta_7 \text{hours.conducting.trains}_i\end{aligned} \tag{10}$$

where $i = 1, \ldots, 348$. For a Bayesian analysis, we assume normal independent prior distributions for the regression parameters, that is, $\beta_0 \sim N(0,1)$, $\beta_j \sim N(0,0.1)$, $j = 1, \ldots, 7$ where $N(0,1)$ denotes a normal distribution with a mean equal to zero and variance equal to one and uniform prior distributions $p \sim U(0,1)$ and $\Phi \sim U(1,10)$ for the parameters p and Φ. Also using the Openbugs software and the same previously used simulation scheme (11,000 as a burn-in sample and 1,000 additional samples chosen from a total of 50,000 samples chosen from 50 out of 50), we obtain the posterior summaries of interest. Table 3 shows the posterior summaries of interest.

Table 3 – Posterior summaries (accident rates/months of work).

	Mean	S.D.	Lower 90% c.i	Upper 90% c.i
β_0	0.2055	1.0050	-1.7160	2.2150
β_1	-0.1393	0.1263	-0.4384	0.0981
β_2	0.2481	0.1872	-0.0868	0.6272
β_3	0.0518	0.3078	-0.5428	0.7244
β_4	-0.0049	0.3145	-0.6457	0.5987
β_5	-0.2892	0.1439	-0.6222	-0.0664
β_6	0.1124	0.3115	-0.5109	0.7006
β_7	0.0186	0.4550	-0.9163	0.8822
p	0.6135	0.0267	0.5611	0.6653
Φ	1.1780	0.1823	1.0050	1.7000

From the results of Table 3, we can conclude that the covariate amount of train rides (trip count) shows evidence that it is a significant covariate in the response accident rates/months of work because zero is not included in the 95% credibility interval for the regression parameter $\beta_5(-0.6222;-0.0664)$. The β_5 regression parameter estimator is negative (-0.2892). Thus, with the increase in train rides, there is a decrease in the accident rate per months worked. The covariates detachment distances and age also show some evidence of significant effects on the response on accident rates/work months as zero is almost not included in the 95% credibility intervals for the regression parameters $\beta_1(-0.4384;0.0981)$ and $\beta_2(-0.0868;0.6272)$. We observed a negative effect (β_1 is estimated by a negative value) of the covariate detachment distances on the response accident rates/work months, that is, increasing the distances between stops decreases the accident/work month rate and a positive effect (β_2 is estimated by a positive value) of the covariate age on the response accident rates/months of work, that is, increasing the age of the driver increases the rate of accidents/months worked. All other covariates show no significant effects on the response as the zero value is included and well centered in the corresponding 95% credibility intervals for each regression parameter.

4.2 Second response of interest: number of accidents per train conductor

Considering now the counts of accidents per train conductor, where the value zero is frequent (no accidents), we consider in the analysis of the count data the zero-inflated Poisson (ZIP) model introduced in Section 3.1.

Initially we assume the zero-inflated Poisson (ZIP) regression model defined in Section 3.1 without assuming the presence of covariates, defined by (2) and (3). For a Bayesian analysis, we assume uniform independent prior distributions, that is, $\mu \sim G(0.1, 0.1)$ and $p \sim U(0,1)$ where $G(\alpha, \beta)$ denotes a gamma distribution with mean α/β and variance α/β_2. Thus, we are assuming non-informative prior distributions for the parameters μ and p. Using the Openbugs software, we initially generated 1,000 Gibbs samples, discarded to eliminate the effect of the initial values in the iterative procedure of simulating samples of the joint posterior distribution for μ and p. Next, we simulate another 10,000 samples by choosing each 10th generated sample, totaling 1,000 samples used to find the posterior summaries of interest. The convergence of the simulation algorithm was verified from graphs of the samples generated for each parameter. Table 4 presents the posterior summaries of interest.

Table 4 – Posterior summaries (accident counts without covariates).

	Mean	S.D.	Lower 95% c.i.	Upper 95% c.i
μ	1.3840	0.0987	1.1980	1.5820
p	0.6115	0.0262	0.5622	0.6618

The conditional mean μ (only for workers where accidents are observed) of the Poisson distribution (2) is estimated to be 1.384. The conditional sample mean obtained from the data is given by, 187/135 = 1.38519. The proportion of zeros estimated by the model is given by 0.6115. As the sample proportion of zeros is given by 0.6121, we conclude that the model is well fitted by the data.

The non-conditional sample mean is given from the data by 0.5374 and the non-conditional mean estimated by the model is given by, $E(Y) = \mu(1-p) = 1.384(1-0.6115) = 0.537684$ again indicating the excellent fit of the model to the data

With the presence of the covariates detachment distances, conductor age, conductor company time, marital status, amount of train rides, sum of work months and hours conducting trains, we assume the regression model defined by (2), (3) and $\mu = \exp(x'\beta)$, where $x'\beta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p$, that is,

$$\begin{aligned}\log(\mu_i) = \beta_0 + \beta_1 \text{detachment.distances}_i + \beta_2 \text{age}_i + \beta_3 \text{job.time}_i + \beta_4 \text{marital.status}_i + \\ \beta_5 \text{trip.count}_i + \beta_6 \text{work.months}_i + \beta_7 \text{hours.conducting.trains}_i\end{aligned} \tag{11}$$

where $i = 1, \ldots, 348$. For a Bayesian analysis, we assume normal independent prior distributions for the regression parameters, that is, $\beta_0 \sim N(0,1)$, $\beta_j \sim N(0,0.1)$, $j = 1, \ldots, 7$ and a uniform uniform prior distribution, $p \sim U(0,1)$ for the parameter p. Also using the Openbugs software and the same simulation scheme used earlier (311,000 as a burn-in sample and 1,000 additional samples chosen from a total of 400,000 samples chosen from 100 out of 100), we obtain the posterior summaries of interest. Table 5 shows the posterior summaries of interest.

Table 5 – Posterior summaries (accident counts with covariates).

	Mean	S.D.	Lower 80% c.i	Upper 80% c.i.
β_0	-0.1548	0.7054	-1.0260	0.7819
β_1	0.0002	0.0005	-0.0005	0.0009
β_2	-0.0049	0.0140	-0.0224	0.0129
β_3	-0.0330	0.0374	-0.0806	0.0147
β_4	0.1149	0.1575	-0.0867	0.3163
β_5	-0.0005	0.0009	-0.0016	0.0063
β_6	0.0197	0.0198	-0.0042	0.0459
β_7	0.1209	0.0959	-0.0006	0.2461
p	0.6138	0.0266	0.5791	0.6475

We can conclude that the covariate hours conducting trains shows some evidence that it is a significant covariate in the accident count response (zero is almost not included (upper limit -0.0006 close to zero) in the 80% credibility interval for the regression parameter β_7. The Bayesian estimator of the regression parameter β_7 is positive. Thus, with the increasing of hours conducting trains, there is a decrease in the accident count. This shows that with more experience of the conductor, there is a decrease in the number of accidents on the railway. All other covariates do not show significant effects on the response as the zero value is included and well centered in the corresponding 80% credibility intervals for each regression parameter.

5 CONCLUDING REMARKS

This study identified some personal and professional factors of train conductors of a railway logistics company with the occurrence of accidents in the period of years ranging from 2014 to 2016 where the responses of interest are given by the accident rates per month of work (number of accidents/months of work) and accident count. From a statistical analysis of the dataset using Beta and Poisson regression models in the presence of excess of zeros under a Bayesian approach using MCMC simulation methods, it was possible to identify some important results:

- For the response accident rates/months of work the covariate amount of train rides (trip count) shows evidence that it is a significant covariate in the response. The covariates detachment distances and conductor age also show some evidence of significant effects on the response on accident rates/work months.
- For the response accident count, the covariate hours conducting trains shows some evidence that it is a significant covariate in the accident count.

These results could be of great interest to the managers of the railway logistics company, to improve the railway safety.

The use of the proposed Beta and Poisson regression models also could be assumed for other situations, especially in industrial and transport accidents where there are many zero values (no occurrence of accidents) related to workers. It is important to point out, that the use of standard usual statistical models is not appropriate in the statistical analysis of this type of data.

Under a Bayesian approach using MCMC methods it is possible to get accurate inferences for the assumed mixture models. The use of the existing free software Openbug simplifies the simulation of samples of the joint posterior distribution, not requiring great computational knowledge. Other advantage of the Bayesian approach in applications: possibility to have informative prior distributions elicited from experts in the railway sector leading to more accurate results.

References

AHER SB & TIWARI DR. 2018. Railway disasters in India: causes, effects and management. *Int J Rev Res Soc Sci*, **6**(2): 125–32.

ARAÚJO MCM & DA SILVA SOUSA MA. 2023. Ferrovias brasileiras: Histórico e processo de estagnação. *Engenharia, Gestão e Inovação*, **8**: 42–51.

ARRUDA J, FRANCO L & VASCONCELOS L. 2022. Redução de custo com inspeção e manutenção em rodeiros de vagões ferroviários: Um estudo de casos nos rolamentos de rodeiros ferroviários. In: *Gestão e Manutenção Industrial e Mineração*, vol. 2. Poisson.

BALA M & BHASIN A. 2018. A Review on Analysis of Railway Traffic Accident with Data Mining Techniques. *International Journal of Computer Sciences and Engineering*, **6**(6): 1251–1256.

BOX GE & TIAO GC. 1973. *Bayesian inference in statistical analysis*. Reading, MA: Addison-Wesley.

CEPEDA-CUERVO E, ACHCAR JA & LOPERA LG. 2014. Bivariate beta regression models: joint modeling of the mean, dispersion and association parameters. *Journal of Applied Statistics*, **41**(3): 677–687.

CHIB S & GREENBERG E. 1995. Understanding the metropolis-hastings algorithm. *The American Statistician*, **49**(4): 327–335.

DA SILVA LRL, ACHCAR JA & HERMOSILLA JLG. 2021. Longitudinal work ability index (wai): a case study with workers in an agricultural research company. *Independent Journal of Management & Production*, **12**(2): 470–485.

DE ALMEIDA ELEUTÉRIO G & ROSA RA. 2023. Planejamento das rotas dos recursos ferroviários para realização da manutenção ferroviária considerando sincronismo, precedência e prioridade. *Transportes*, **31**(2): e2644–e2644.

DHAYGUDE A, DEOKAR Y, BASTE N, DENGALE R & DARADE M. 2019a. Statistical Analysis of Railway Accidents. *International Research Journal of Engineering and Technology (IRJET)*, **6**: 4539–4544.

DHAYGUDE A, DEOKAR Y, BASTE N, DENGALE R & DARADE M. 2019b. Study of railway accidents. *International Journal of Scientific & Engineering Research (IJSER)*, **10**(4): 223–230.

ERDMAN D, JACKSON L, SINKO A ET AL. 2008. Zero-inflated Poisson and zero-inflated negative binomial models using the COUNTREG procedure. *Sas Global Forum*, **2008**: 1–11.

EVANS AW. 2011. Fatal train accidents on Europe's railways: 1980–2009. *Accident Analysis & Prevention*, **43**(1): 391–401.

FERRARI S & CRIBARI-NETO F. 2004. Beta regression for modelling rates and proportions. *Journal of Applied Statistics*, **31**(7): 799–815.

GELFAND AE & SMITH AF. 1990. Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, **85**(410): 398–409.

GEORGIOU I. 2009. Mapping railway development prospects in Brazil. *Transport Reviews*, **29**(6): 685–714.

HALL DB. 2000. Zero-inflated Poisson and binomial regression with random effects: a case study. *Biometrics*, **56**(4): 1030–1039.

HONG WT, CLIFTON G & NELSON JD. 2023. Railway accident causation analysis: current approaches, challenges and potential solutions. *Accident Analysis & Prevention*, **186**: 107049.

JØRGENSEN B. 1997. Proper dispersion models. *Brazilian Journal of Probability and Statistics*, **11**(2): 89–128.

KERETCH EM & DE PAIVA CEL. 2016. Rail Accidents Caused by Failure on the Permanent Way. *Journal of Traffic and Logistics Engineering*, **4**(1). Available at: https://api.semanticscholar.org/CorpusID:115020659.

KYRIAKIDIS M, PAK KT & MAJUMDAR A. 2015. Railway accidents caused by human error: historic analysis of UK railways, 1945 to 2012. *Transportation Research Record*, **2476**(1): 126–136.

LIU Y, LI K & YAN D. 2024. Quantification analysis of potential risk in railway accidents: A new random walk based approach. *Reliability Engineering & System Safety*, **242**: 109778.

LUNN DJ, THOMAS A, BEST N & SPIEGELHALTER D. 2000. WinBUGS-a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*, **10**: 325–337.

MONTGOMERY DC & RUNGER GC. 2010. *Applied Statistics and Probability for Engineers*. John Wiley & Sons.

MURTA ALS, MURTA MDPA, DE FREITAS MS & DAS NEVES FMDA. 2023. CO2 reduction in brazilian road and rail transport. *Revista Valore*, **8**: 8015. Available at: https://api.semanticscholar.org/CorpusID:258751889.

RAD MA, LEFSRUD LM & HENDRY MT. 2023. Application of systems thinking accident analysis methods: A review for railways. *Safety Science*, **160**: 106066.

SAN KIM D & YOON WC. 2013. An accident causation model for the railway industry: Application of the model to 80 rail accident investigation reports from the UK. *Safety Science*, **60**: 57–68.

SCOTT LONG J. 1997. Regression Models for Categorical and Limited Dependent Variables. *Advanced Quantitative Techniques in the Social Sciences*, **7**.

SHAPPELL SA & WIEGMANN DA. 2000. The human factors analysis and classification system-HFACS, US Department of Transportation, 1–15, .

SHI J, WANG Y & ZHENG W. 2020. Correlation Analysis of Causes of Railway Accidents Based on Improved Apriori Algorithm. In: *2020 13th International Symposium on Computational Intelligence and Design (ISCID)*. pp. 274–277. IEEE.

SOUZA APD, TAVARES FM, BONIKOWSKI RTR & PEREIRA VHM. 2019. Reduction of number of railroad accidents with locomotive as the main cause. Final Project (Rail Transportation Management) - Deutsche Bahn; Instituto de Transporte e Logística, Brasília, .

SOUZA E, BITTENCOURT T, AMES I, RIBEIRO D & CARVALHO H. 2023. Drive-by damage detection methodology for high-speed railway bridges applying Mel-frequency cepstral coefficients. In: *Life-Cycle of Structures and Infrastructure Systems*. pp. 252–259. CRC Press.

TAVARES FM, HERMOSILLA JLG, ACHCAR JA & DA SILVA E. 2021. O perfil do trabalhador e sua relação com a ocorrência de atos inseguros: o caso de maquinistas de trens de uma empresa de logística. *Conjecturas*, **21**(3): 98–121.

WANG N, YANG X, CHEN J, WANG H & WU J. 2023. Hazards correlation analysis of railway accidents: A real-world case study based on the decade-long UK railway accident data. *Safety Science*, **166**: 106238.

WANG Y, ZHENG W, DONG H & GAO P. 2020. Factors correlation mining on railway accidents using association rule learning algorithm. In: *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*. pp. 1–6. IEEE.

WASNIK R. 2010. Analysis of railway fatalities in central India. *Governing Council 2010-2012*, **32**: 311.

YAN D, LI K, ZHU Q & LIU Y. 2023. A railway accident prevention method based on reinforcement learning–Active preventive strategy by multi-modal data. *Reliability Engineering & System Safety*, **234**: 109136.

How to cite

TAVARES FM, ACHCAR JA, HERMOSILLA JLG & BARILI E. 2024. Occurrence of railway accidents related to some personal and professional train conductor factors: use of statistical models in the presence of excess of zeros. *Pesquisa Operacional*, **44**: e282287. doi: 10.1590/0101-7438.2023.043.00282287.

APPENDIX 1

detachment distances	age	job time	marital status	trip count	work months	hours conducting trains	delta	accident rate/work months	accident count
464	38	3,35	1	127	14	1	1	0	0
132	37	3,91	1	614	34	1	0	0,0588	2
437	35	4,01	1	315	34	1	0	0,0294	1
179	32	4,02	1	277	34	1	0	0,0294	1
237	46	4,05	1	311	34	2	1	0	0
132	34	4,13	1	138	16	1	0	0,0625	1
437	29	4,18	1	279	33	1	1	0	0
437	31	4,21	1	285	33	1	1	0	0
437	34	4,23	2	295	33	2	1	0	0
437	31	4,25	1	240	32	1	0	0,0625	2
132	35	4,25	2	654	33	2	1	0	0
437	35	4,25	1	269	30	1	1	0	0
29	30	4,46	2	510	31	1	1	0	0
179	33	4,81	1	714	34	2	1	0	0
29	40	4,81	1	557	34	1	1	0	0
132	34	4,85	1	286	33	1	0	0,0606	2
132	36	4,85	1	314	35	2	0	0,0571	2
464	33	4,87	1	126	12	1	1	0	0
464	39	4,87	1	26	6	1	1	0	0
132	32	4,9	1	347	34	1	0	0,0588	2
179	30	4,95	1	313	34	1	0	0,0294	1
132	32	4,96	1	294	32	2	1	0	0
132	30	5	1	301	34	1	0	0,0294	1
464	28	5,12	2	195	24	2	1	0	0
464	31	5,12	1	245	22	2	1	0	0
176	35	5,21	1	399	32	1	0	0,0313	1
237	33	5,27	1	301	35	2	1	0	0
29	34	5,27	1	534	34	1	0	0,0294	1
29	36	5,27	1	516	33	1	0	0,0303	1
237	38	5,27	1	310	35	1	1	0	0
29	40	5,27	1	574	33	1	1	0	0
29	36	5,28	2	566	33	1	0	0,0303	1
464	32	5,29	1	569	33	2	1	0	0
464	38	5,29	1	375	33	2	0	0,0303	1
67	33	5,3	1	277	34	2	0	0,0294	1
132	35	5,3	2	384	32	2	0	0,0625	2
29	33	5,31	1	580	34	1	0	0,0294	1
29	35	5,31	1	561	33	2	1	0	0
132	35	5,31	1	495	33	1	1	0	0
132	37	5,31	1	292	33	1	0	0,0606	2
237	39	5,31	1	369	32	1	0	0,0313	1
29	41	5,32	1	2	2	2	1	0	0
237	34	5,33	2	311	33	1	1	0	0
464	26	5,35	2	352	33	1	1	0	0
29	33	5,35	2	564	34	1	1	0	0
179	37	5,36	1	270	32	1	1	0	0
179	37	5,36	1	289	35	1	0	0,0286	1

detachment distances	age	job time	marital status	trip count	work months	hours conducting trains	delta	accident rate/work months	accident count
29	31	5,38	1	598	33	2	0	0,0303	1
132	39	5,38	1	454	34	2	0	0,0588	2
237	35	5,39	1	354	34	2	1	0	0
29	36	5,4	1	518	33	1	1	0	0
29	37	5,4	1	610	34	1	1	0	0
29	42	5,4	1	608	33	2	1	0	0
237	33	5,41	1	134	16	1	1	0	0
29	47	5,41	1	632	34	2	1	0	0
237	33	5,42	1	231	23	1	1	0	0
237	35	5,42	1	381	34	1	1	0	0
237	37	5,42	1	321	32	2	1	0	0
237	38	5,42	1	346	34	1	0	0,0294	1
437	32	5,43	1	322	33	1	1	0	0
437	38	5,43	1	307	33	2	1	0	0
437	39	5,43	1	263	31	1	1	0	0
179	30	5,48	1	297	34	2	0	0,0294	1
67	40	5,5	1	307	31	1	1	0	0
29	34	5,51	1	587	33	2	1	0	0
179	35	5,52	1	283	35	1	0	0,0571	2
29	27	5,54	2	571	35	1	0	0,0286	1
29	28	5,54	2	548	32	1	1	0	0
464	33	5,54	1	252	22	2	1	0	0
464	29	5,55	1	41	4	2	1	0	0
237	30	5,58	1	247	17	1	1	0	0
237	42	5,58	1	320	34	1	0	0,0294	1
237	31	5,65	1	16	2	1	1	0	0
179	29	5,68	2	690	34	2	1	0	0
464	33	5,68	1	256	32	1	0	0,0313	1
237	39	5,75	1	313	33	1	1	0	0
437	30	5,77	1	274	33	1	1	0	0
464	27	5,81	2	310	35	1	1	0	0
464	31	5,81	2	261	32	1	1	0	0
464	31	5,81	1	601	33	1	1	0	0
464	33	5,81	1	278	34	1	0	0,0294	1
464	34	5,81	2	607	32	1	1	0	0
237	37	5,83	1	318	33	1	1	0	0
237	37	5,83	1	317	32	2	1	0	0
437	29	5,87	1	296	33	1	1	0	0
67	30	5,87	1	253	34	2	1	0	0
132	32	5,87	1	430	32	1	1	0	0
176	35	5,87	1	329	33	1	1	0	0
437	45	5,87	1	302	33	3	0	0,0303	1
464	28	5,9	1	2	2	1	1	0	0
464	29	5,9	2	1	1	2	1	0	0
464	31	5,9	1	346	33	1	0	0,0303	1
464	44	5,9	1	1	1	2	1	0	0
464	27	5,92	1	200	11	1	1	0	0
464	31	5,92	1	189	22	1	1	0	0
464	38	5,92	1	260	28	1	1	0	0

detachment distances	age	job time	marital status	trip count	work months	hours conducting trains	delta	accident rate/work months	accident count
29	53	5,92	1	497	33	3	1	0	0
29	30	5,96	1	560	33	2	0	0,0303	1
29	29	5,97	1	23	12	1	0	0,0833	1
132	29	5,98	1	444	34	1	1	0	0
464	30	5,99	1	365	26	1	1	0	0
67	39	5,99	1	294	34	1	0	0,0588	2
179	38	6	1	313	34	2	0	0,0588	2
67	36	6,01	1	233	35	2	1	0	0
132	41	6,01	2	420	33	2	0	0,0303	1
464	27	6,04	1	274	35	1	0	0,0286	1
464	28	6,04	1	334	33	1	1	0	0
464	31	6,04	1	432	23	1	1	0	0
132	29	6,06	2	322	33	1	0	0,0303	1
437	37	6,12	1	291	35	1	0	0,0571	2
464	38	6,12	1	302	33	1	1	0	0
464	30	6,13	1	257	34	1	0	0,0294	1
179	33	6,13	2	292	34	1	1	0	0
437	31	6,17	2	289	34	1	1	0	0
179	37	6,17	1	697	34	2	1	0	0
179	41	6,17	1	286	33	1	1	0	0
179	41	6,19	1	692	34	2	1	0	0
179	43	6,19	1	737	34	2	1	0	0
437	32	6,2	1	265	33	1	0	0,0303	1
437	32	6,2	1	301	33	3	0	0,0606	2
437	33	6,2	1	265	33	1	0	0,0303	1
437	35	6,2	1	291	34	2	1	0	0
437	38	6,2	1	290	33	2	0	0,0303	1
437	40	6,2	1	303	33	2	0	0,0303	1
237	29	6,22	1	30	5	1	1	0	0
237	35	6,22	1	352	33	2	0	0,0303	1
464	33	6,25	1	252	36	1	0	0,0278	1
464	30	6,33	1	113	14	2	1	0	0
29	31	6,39	2	564	34	1	0	0,0588	2
179	32	6,39	1	313	33	2	1	0	0
179	34	6,39	1	288	33	2	0	0,0909	3
67	33	6,4	2	284	35	2	1	0	0
179	36	6,41	1	66	4	2	1	0	0
464	35	6,44	2	245	28	1	0	0,0357	1
464	35	6,52	1	276	33	1	1	0	0
29	32	6,55	1	612	33	2	0	0,0303	1
29	32	6,55	2	556	36	2	1	0	0
132	35	6,58	1	349	30	1	0	0,0667	2
132	35	6,61	1	286	33	2	1	0	0
179	37	6,77	2	614	31	2	1	0	0
179	35	6,78	2	694	34	3	1	0	0
464	29	6,81	2	141	23	2	1	0	0
464	30	6,82	1	457	32	1	1	0	0
464	30	6,82	1	212	32	2	1	0	0
464	37	6,82	1	329	22	2	1	0	0

detachment distances	age	job time	marital status	trip count	work months	hours conducting trains	delta	accident rate/work months	accident count
464	39	6,82	1	273	33	1	1	0	0
132	32	6,83	2	305	33	2	1	0	0
132	33	6,87	1	415	34	3	0	0,0294	1
464	36	6,87	2	315	23	2	1	0	0
176	36	7,03	1	318	34	2	0	0,0294	1
132	43	7,05	2	6	5	2	1	0	0
464	33	7,15	1	292	32	2	1	0	0
237	34	7,18	2	249	22	1	1	0	0
464	30	7,19	2	299	32	1	1	0	0
464	31	7,19	1	368	33	1	1	0	0
237	32	7,19	1	318	34	1	1	0	0
464	33	7,19	1	264	34	2	1	0	0
132	34	7,19	2	306	33	2	1	0	0
437	35	7,19	1	266	31	2	0	0,0645	2
132	36	7,19	2	404	34	1	1	0	0
464	36	7,19	1	288	35	2	1	0	0
237	37	7,19	1	330	35	2	1	0	0
132	42	7,19	1	339	33	2	0	0,0303	1
29	33	7,24	2	560	33	2	1	0	0
437	35	7,24	1	287	34	2	0	0,0588	2
176	37	7,24	2	319	34	1	1	0	0
237	32	7,25	2	347	33	3	0	0,1212	4
237	33	7,25	2	334	33	1	1	0	0
29	34	7,25	1	12	5	2	1	0	0
237	35	7,25	1	326	34	3	0	0,0294	1
237	36	7,25	1	330	33	2	0	0,0909	3
237	41	7,25	2	214	14	1	1	0	0
29	38	7,32	1	1	1	2	1	0	0
464	32	7,36	1	291	34	3	0	0,0294	1
464	34	7,36	1	313	35	1	0	0,0571	2
464	35	7,36	1	170	21	1	0	0,0476	1
132	41	7,36	2	292	33	2	0	0,0909	3
179	37	7,38	1	330	33	1	0	0,0303	1
29	40	7,45	1	3	2	1	1	0	0
67	31	7,55	1	262	34	2	0	0,0588	2
179	34	7,61	1	711	33	3	1	0	0
67	36	7,63	2	326	33	3	0	0,0303	1
464	32	7,74	1	261	34	1	1	0	0
176	35	7,76	1	257	33	1	1	0	0
132	42	7,76	2	298	33	1	0	0,0303	1
132	33	7,85	1	529	35	2	0	0,0571	2
464	30	7,87	1	275	33	2	1	0	0
464	32	7,87	2	229	31	2	1	0	0
464	36	7,87	1	6	6	2	1	0	0
176	37	8,19	1	196	31	1	0	0,0323	1
176	38	8,19	1	144	25	1	1	0	0
67	34	8,2	1	356	35	1	0	0,0286	1
464	32	8,46	1	358	31	1	1	0	0
464	34	8,46	1	583	35	1	1	0	0

detachment distances	age	job time	marital status	trip count	work months	hours conducting trains	delta	accident rate/work months	accident count
437	36	8,46	1	331	33	3	1	0	0
464	37	8,46	1	257	34	2	0	0,0294	1
464	41	8,46	1	565	33	2	1	0	0
179	32	8,52	1	538	32	3	1	0	0
132	55	8,54	1	284	33	3	0	0,0606	2
237	38	8,62	1	326	34	3	1	0	0
237	39	8,62	2	328	34	2	1	0,0000	0
237	40	8,62	1	397	33	2	0	0,0303	1
132	40	8,62	1	200	34	1	1	0,0000	0
67	32	8,79	1	248	33	2	0	0,0303	1
437	36	8,81	1	333	34	3	1	0,0000	0
237	37	8,81	1	405	33	2	0	0,0303	1
176	39	8,81	1	447	34	2	0	0,0294	1
237	40	8,81	1	216	16	1	1	0,0000	0
132	30	8,83	1	304	33	2	0	0,0303	1
132	35	8,83	2	281	33	1	0	0,0303	1
464	42	8,84	1	8	6	3	1	0,0000	0
237	45	8,84	1	7	5	3	1	0,0000	0
132	35	8,99	1	241	26	2	1	0,0000	0
237	37	9	2	343	33	2	0	0,0909	3
464	31	9,03	2	197	33	3	1	0	0
464	47	9,03	2	21	4	2	0	0,2500	1
176	57	9,03	1	344	33	2	1	0	0
237	30	9,07	1	322	31	2	0	0,0323	1
176	34	9,07	1	431	35	2	0	0,0286	1
437	34	9,07	2	90	11	2	1	0,0000	0
237	35	9,07	1	160	12	2	1	0,0000	0
437	40	9,07	1	303	33	3	0	0,0606	2
464	35	9,11	1	276	33	2	1	0,0000	0
67	29	9,12	1	248	34	3	1	0,0000	0
29	31	9,12	1	522	32	1	0	0,0313	1
29	34	9,12	2	6	4	3	1	0	0
179	32	9,14	1	695	34	3	1	0	0
29	32	9,14	1	387	19	3	1	0	0
179	36	9,14	1	713	33	3	1	0	0
176	34	9,15	1	299	36	2	1	0	0
176	35	9,15	2	297	34	2	1	0	0
437	35	9,15	1	101	20	2	1	0	0
176	35	9,15	1	432	33	1	1	0	0
176	35	9,15	1	325	34	3	1	0,0000	0
176	39	9,15	1	352	35	1	1	0,0000	0
437	29	9,18	1	281	32	1	0	0,0313	1
176	42	9,18	2	102	14	1	1	0,0000	0
464	41	9,38	1	240	34	1	0	0,0294	1
179	39	9,4	1	310	33	3	1	0,0000	0
179	43	9,44	1	348	32	1	0	0,0313	1
237	32	9,45	1	296	19	2	1	0,0000	0
237	35	9,45	2	31	4	2	1	0,0000	0
29	33	9,46	1	10	5	1	1	0,0000	0

detachment distances	age	job time	marital status	trip count	work months	hours conducting trains	delta	accident rate/work months	accident count
67	29	9,5	2	254	33	2	1	0,0000	0
464	35	9,53	2	266	32	2	0	0,0625	2
437	37	9,56	2	224	33	1	0	0,0606	2
176	49	9,66	1	470	35	1	1	0,0000	0
464	35	9,76	2	205	32	2	1	0,0000	0
464	38	9,76	1	293	35	2	1	0,0000	0
237	31	9,79	2	218	11	3	1	0,0000	0
67	55	9,94	1	290	33	4	0	0,0303	1
67	32	9,99	1	302	35	2	0	0,0286	1
67	36	9,99	1	281	35	1	0	0,0286	1
179	33	10,01	2	282	36	1	0	0,0278	1
179	38	10,01	1	210	35	3	1	0,0000	0
179	44	10,01	1	703	34	3	1	0	0
237	56	10,02	2	307	34	3	1	0	0
437	30	10,05	1	284	34	3	1	0	0
437	31	10,05	2	326	33	4	0	0,0606	2
237	36	10,05	2	313	33	3	0	0,0303	1
132	31	10,06	1	32	2	2	1	0,0000	0
237	32	10,06	1	373	34	3	1	0,0000	0
437	32	10,06	1	327	32	4	0	0,0938	3
179	36	10,06	1	284	25	3	1	0,0000	0
179	37	10,06	1	316	34	3	0	0,0588	2
237	38	10,06	1	303	34	2	0	0,0294	1
179	40	10,06	1	418	33	3	1	0,0000	0
237	41	10,06	1	282	30	2	1	0,0000	0
176	42	10,06	1	318	29	3	1	0,0000	0
179	43	10,06	1	550	33	3	1	0	0
132	43	10,06	1	314	33	3	1	0	0
176	38	10,32	1	373	34	3	0	0,0882	3
464	33	10,42	1	315	34	2	1	0,0000	0
237	34	10,42	1	329	34	3	0	0,0882	3
237	35	10,42	1	416	33	4	0	0,0606	2
464	39	10,42	1	195	13	4	1	0,0000	0
132	36	11,15	2	16	7	2	1	0,0000	0
132	40	11,15	1	307	34	4	0	0,0294	1
237	35	11,33	2	327	33	3	0	0,0303	1
437	35	11,33	1	272	34	3	0	0,0882	3
237	38	11,33	1	613	33	3	0	0,0303	1
176	41	11,33	1	443	35	2	1	0	0
176	47	11,33	1	67	7	4	1	0	0
176	53	11,33	1	360	34	4	0	0,0294	1
67	38	11,39	1	268	32	3	0	0,0313	1
67	32	11,45	1	195	32	4	0	0,0313	1
67	34	11,45	1	264	33	4	0	0,0303	1
132	35	11,45	2	420	34	3	0	0,0294	1
464	35	11,45	1	269	34	3	1	0,0000	0
179	37	11,45	1	716	34	4	1	0,0000	0
437	37	11,45	1	302	34	3	0	0,0294	1
237	42	11,45	1	242	17	3	0	0,0588	1

detachment distances	age	job time	marital status	trip count	work months	hours conducting trains	delta	accident rate/work months	accident count
132	34	11,66	1	324	30	3	0	0,0333	1
67	40	11,66	1	282	34	4	1	0	0
67	41	11,66	2	303	33	2	1	0	0
237	36	11,72	2	625	33	4	1	0	0
132	38	12	1	303	33	3	0	0,0303	1
132	41	12,03	1	163	33	3	1	0,0000	0
176	35	12,2	1	314	33	3	1	0,0000	0
179	36	12,2	2	393	33	4	0	0,0303	1
132	38	12,2	1	327	33	3	0	0,0303	1
29	42	12,2	2	4	4	4	1	0	0
437	43	12,2	1	295	33	3	1	0	0
176	37	12,21	1	433	34	4	1	0	0
176	42	12,21	1	17	4	3	1	0	0
464	37	12,25	1	287	33	4	1	0	0
132	40	12,45	1	315	23	3	1	0	0
464	42	12,45	1	257	32	1	0	0,0313	1
464	48	12,45	1	273	33	4	0	0,0606	2
179	34	12,48	1	369	32	4	1	0,0000	0
67	38	12,48	1	291	36	3	1	0,0000	0
237	39	12,99	1	294	33	1	1	0,0000	0
437	41	12,99	1	324	33	3	0	0,0303	1
179	34	13,06	1	200	27	5	1	0,0000	0
132	35	13,06	1	340	33	4	0	0,0606	2
132	35	13,06	1	616	33	3	0	0,0303	1
132	38	13,06	1	68	9	3	1	0	0
67	41	13,06	1	305	33	4	1	0,0000	0
437	41	13,06	2	254	31	4	0	0,0645	2
132	42	13,06	1	296	33	3	0	0,0303	1
176	43	13,06	1	366	33	4	0	0,0303	1
132	44	13,06	1	141	32	3	0	0,0313	1
132	47	13,06	2	103	31	3	1	0	0
237	45	13,08	1	301	30	4	1	0	0
179	35	13,79	1	464	31	5	1	0	0
132	52	14,02	2	231	30	4	0	0,0333	1
464	34	14,03	1	314	32	4	1	0,0000	0
437	35	14,03	1	437	34	4	0	0,0294	1
132	35	14,03	1	612	32	4	1	0,0000	0
132	37	14,03	1	634	34	4	1	0,0000	0
132	39	14,03	1	662	33	4	0	0,0303	1
179	52	14,34	1	193	30	5	1	0	0
179	60	14,34	1	90	15	5	1	0	0
29	57	15,08	1	580	33	3	0	0,0303	1
437	34	15,15	1	293	33	3	1	0	0
67	37	15,15	1	226	34	5	1	0	0
132	42	15,28	1	189	15	5	1	0	0
437	37	15,31	1	228	26	5	1	0	0
132	38	15,31	2	1	1	4	1	0	0
179	41	15,31	2	286	33	4	0	0,0909	3
132	44	15,31	2	298	33	5	0	0,0303	1

detachment distances	age	job time	marital status	trip count	work months	hours conducting trains	delta	accident rate/work months	accident count
179	43	15,95	1	693	32	4	1	0,0000	0
237	39	15,98	1	660	32	5	1	0,0000	0
29	57	16,23	2	176	22	3	0	0,0455	1
67	38	17,15	1	274	34	3	0	0,0588	2
67	38	17,33	1	291	33	5	1	0	0
237	52	18,06	1	320	33	5	1	0	0
67	56	29,72	1	199	34	5	1	0	0