Journal of Engineering and Applied Sciences 12 (11): 2956-2962, 2017
ISSN: 1816-949X

Selection of the Best Regression Model to Explain the Variables that Influence Labor Accident Electrical Company Case

[1]Noel Varela Izquierdo, [2]Damaise Perez Fernandez,
[3]Omar Bonerge Pineda Lezama and [1]Amelec Viloria
[1]Faculty of Engineering, Universidad de la Costa (CUC), Barranquilla, Colombia
[2]Faculty of Business Economics, Universidad de Cienfuegos (UCF), Cienfueges, Cuba
[3]Faculty of Engineering,
Universidad Nacional Autonoma De Honduras En El Valle De Sula (Unah-vs),
San pedro sula, Honduras

Abstract: The present research proposes an alternative to select the best model that explains the relation of the variables that influence the labor accident in an electric power company. Among the techniques and tools used are those of occupational safety and health management, multivariate statistics, generalized linear models, the values of the deviation percentage explained and the adjusted percentage and the Akaike and Bayesian information criteria. The following variables were identified through the mentioned techniques, management commitment, compliance with legislation, prevention planning, training in prevention, updating of occupational risk management and policies that have a significant influence on work accident and through. The percentages and of the previously mentioned criteria were able to show that the logistic regression is the best model that explains the labor accident by presenting the highest percentage and the lowest values of the criteria when compared with the poisson regression and negative binomial models.

Key words: Labor accident, regression models, multivariate statistics, information criteria, highest percentage, negative binomial models

INTRODUCTION

Researcher such as Sanchez *et al.* (2011), Forastieri (2009) research accidents constitute a large source of cost generation. The economic costs of occupational and work-related injuries increase rapidly, according to a report by the International Labor Organization (Tarin and Galera, 2016). Although, it is impossible to set a value on human life, compensation figures indicate that the cost of illnesses and accidents at research represents about 4% of the world's gross domestic product (about $2.8 trillion or $2.6 trillion In the form of lost research time, production interruptions, research absenteeism, disease treatments, incapacity and survivor's benefits, the developed countries are affected by these Figures in spite of the high technological development (Izquierdo, 2015) which is based on the hypothesis that the hypothesis is that the hypothesis is that Rahmani *et al.* (2013), Unsar and Sut (2015).

At an international and national level, there are now alarming figures for the occurrence of occupational accidents. The most recent OIT calculations (2016) show that there are 2.3 million annual deaths and 317 million work-related accidents >5000 perday and for each fatal accident there are between 500 and 2000 injuries depending on the type of research (Hyung and Seung, 2016; Arquillos and Romero, 2016). Statistics such as these show the need to carry out scientific research that contributes to the reduction of these indicators, there by improving the working conditions (as these are the ones that favor the occurrence of these events) and physical, psychological and social well-being of the human factor that performs its functions in the reseach environments.

In recent times in which the application of mathematics to the modeling of various current phenomena has been expanded it is necessary to link this science with safety and health which has made it possible in many places to reduce problems the resaech accident.

In this last sense, the objective of this study is to describe some of the available statistical alternatives for the analysis and identification of the models that are related to labor accidents (variable counting) in turn to describe and compare the different models used for. This purpose to show their advantages and disadvantages and choose the one that best explains the research accident, all this in function of reducing the work accident at the level of companies.

Corresponding Author: Noel Varela Izquierdo, Faculty of Engineering, Universidad de la Costa (CUC), Barranquilla, Colombia

MATERIALS AND METHODS

Predictions of work accident current trends: Amelec and Carmen (2015a, b), Reyes *et al.* (2015) argue that when using a single statistical method to perform an estimation and assessment of occupational risk factors including in this aspect the investigation of accidents can not obtain optimal results in the control of these factors in the workplace. Future perspectives focus on the parallel application of deterministic and stochastic statistical estimation methods. Figure 1 shows the classification of these methods.

In the review of the scientific literature, the contributions given by the use of generalized linear models and/or variable transformations to analyze occupational accidents are increasingly evidenced. In this way, Tamura and Tanaka (2016) which state that the prediction of the labor accident entails a special problematic and show the contributions given to the prevention of accidents at work when using mathematical models.

The investigations consulted always start from a descriptive study taking into account variables such as: age, occupation, sex, profession that many of them then include in the multivariate statistical analysis in this case are authors such as Dahlke (2015), Viloria and Parody (2016). There is a coincidence in the management of variables that affect labor accident such as: environmental conditions, risk assessment, aspects of worker behavior and supervisors, psychosocial aspects, physical conditions, work positions acquired by the researcher, inadequate training and ignorance of regulations concerning safety and health at research, duration of tasks. There are researches developed in this aspect shown by Tamura and Tanaka (2016), Lindsey (1999).

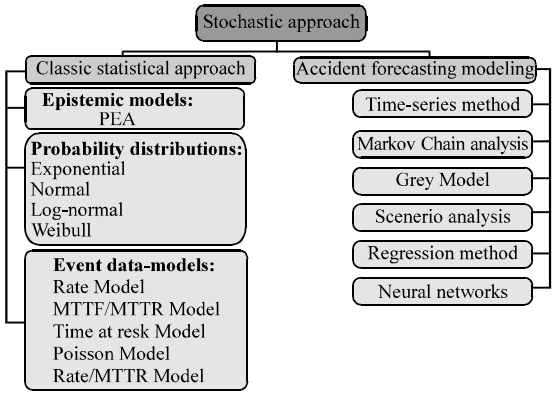

Fig. 1: Classification of the main risk analysis methodologies according to stochastic statistical estimation methods (Reyes *et al*., 2015)

All these reserachers coincide in proposing that regardless of the statistical method used with multivariate statistics as a tool in the investigation of occupational accidents, objective and useful information can be obtained which serves as an input to improve safety management processes and health at work.

Participating variables and questionnaires: In this step, we select those variables that are related to the dependent variable "labor accidents" we research with 3 large groups of variables they are shown in Fig. 2.

Methodology of regression models: The models that are commonly used in the analysis of the causal relationship of the labor accident variable and the associated independent variables are the so-called regression models for counting data, their nature being discrete. The main aspects of the regression models from the theoretical point of view are presented.

Poisson regression: The Poisson distribution is the reference model for counting data (Bakhtiyaria *et al*., 2012; Amelec and Carmen, 2015a, b). Phase count data with a low probability of occurrence (rare events) follow a known probability distribution, called the poisson distribution. The function used in this type of model is:

$$\log(\mu) = b_0 + b_1x_1 ... + b_nx_n$$
$$\mu = \exp(b_0 + b_1x_1 + ... + b_nx_n) \quad (1)$$

where, $\mu>0$ is the average parameter of the distribution which coincides with the value of the variance which defines the "equidispersion" property.

Logistic regression: Logistic regression, like linear regression is a tool that allows us to study the dependene

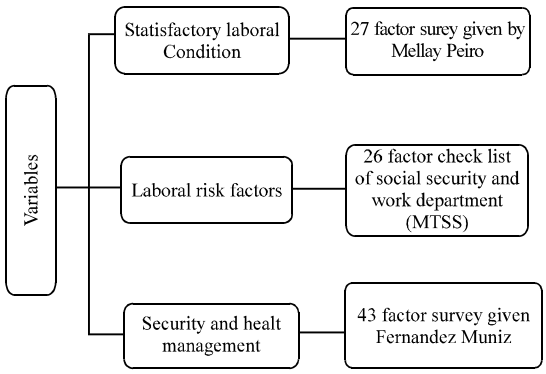

Fig. 2: Variables and questionnaires showing relationship to the dependent variable occupational accidents

between a dependent variable and a set of independent or predictive variables. An element that differentiates them is that the dependent variable is categorical and the predictive variables can be numerical or categorical that is of any nature. Let Y be the model response variable (take only two values 0 and 1):

$$Y = E(Y/X)+e \quad (2)$$

Where:

$$E = (Y/X)\pi(x) = \frac{\exp(\alpha+\beta X)}{1+\exp(\alpha+\beta X)} \quad (3)$$

e has Bernoulli distribution with parameter $\pi(X)$. It is interesting to highlight the transformation of $\pi(x)$ known as logit transformation, defined by the function:

$$g(X) = \ln\frac{\pi(x)}{1-n(x)} = \alpha+\beta x \quad (4)$$

The importance of this function lies in its characteristics. It is a linear function in the parameters, continuous and its values are found on the whole number line $-\infty<g(x)<\infty$.

Negative binomial regression: The assumed statistical model for the data is that the values of the dependent variable Y follow a negative binomial distribution of the form:

$$p(Y) = \frac{\Gamma(Y+\alpha^{-1})}{\Gamma(Y+1)\,\Gamma(\alpha^{-1})}\left[\frac{\alpha^{-1}}{\alpha^{-1}+\mu}\right]^{\alpha^{-1}}\left[\frac{\mu}{\alpha^{-1}+\mu}\right]^{Y},\ \mu>0,\ \alpha\geq 0 \quad (5)$$

where, the mean is the product of λ, the rate at which events occur and the sampling period t according to:

$$E(Y) = \mu = \lambda t \quad (6)$$

The variance of Y is given by:

$$Var(Y) = \mu + a\mu^2 \quad (7)$$

If $\alpha = 0$, negative binomial distribution is reduced to the Poisson distribution. It also assumes that the rate is related to the predictor variables through a log-linear function of the form:

$$\log(\lambda) = \beta_0+\beta_1x_1+\beta_2x_2+\ldots+\beta_kx_k \quad (8)$$

Goodness of fit for regression models

Akaike Information Criteria (AIC): Gualdron argue that the Akaike information criterion proposed in 1974 and used as an unbiased estimator, specifies that the lowest AIC probability model is the one that is selected as the best at that the data is adjusted. The function is given by the maximization of the logarithm of the maximum likelihood denoted as (ln L) and K is the number of parameters of the probability function (parameters in the model):

$$AIC = 2K-2\ln(L) \quad (9)$$

The structure of the AIC is composed of the maximization of the logarithm of likelihood that is to say as a component of the lack of fit of the model and K as the number of estimated parameters.

Bayesian Information Criterion (BIC): BIC serves to select the model between a finite set of models is closely related to the AIC criterion and is based in part on the probability function. To improve the inconsistency of the AIC criterion (Sanchez *et al*., 2011) presented a model selection criterion from the Bayesian perspective. Schwarz stated that the Bayesian solution is to select the model with a high probability a posteriori. The Bayesian Information Criterion (BIC) is defined as Cebador *et al*. (2015):

$$BIC = 2\ln(L)+K\ln(n) \quad (10)$$

Purpose of study (electric power company): Many researchers are exposed to electrical hazards in the development of their activities. Specifically, researchers in the electricity sector are exposed to risks such as: electrocution (fatal) electric shock, burns and other damages caused by contact with electric power. Electricity ranks fifth in work-related injuries resulting in fatal accidents. According to the Occupational Safety and Health Administration (OSHA) the average number of accidents resulting from the risk of electrocution is 12,976 per day and an average of 86 fatal accidents are recorded annually in processes related to the generation, transmission and distribution by Cameron and Trivedi (1990), Lindsey (1999).

In order to illustrate the application of the models described so far, the company chosen for this study is one of the organizations considered of great importance to the country which is engaged in the activities related to the maintenance and construction of electricity networks (Sector is located in the province of Holguin, Cuba). This company has a total of 1493 researchers, of whom 71% are exposed to high risks. In recent years there have been 52 accidents of which 4 were fatal, 23 severe and 25 minor. Of the 52 accidents that occurred in the job with the highest incidence, the lineman, mainly due to electrical contact

which have caused health damage and even death to several workers. All of the above indicates the need for studies to reduce this indicator.

RESULTS

The following is a summary of the methods used and the results are analyzed and discussed. Identification of the variables in the model.

Expert method:. The expert method is used to know the variables that become part of the factor analysis, eight experts give their individual judgment, among them are specialists in health and safety, researchers with vast experience as well as professors who investigate in the Thematic areas belonging to the University of Cienfuegos with the aim of reducing the variables to be used in the factor analysis. The judgment of the experts is consistent for both cases and uses the statistical package SPSS version 19.0. Concluding that 15 and 19 factors related to occupational satisfaction and occupational safety and Health Management can be eliminated, respectively. Finally, there are 12 factors related to work satisfaction and 24 belonging to occupational health and safety management with which the rest of the corresponding analyzes are carried out in the present investigation.

Optimal scaling: The optimum scaling is performed by the need to work with quantitative variables that guarantee a better fit of the model so that the variables of the categorical type are transformed to metric, using the statistical program SPSS version 19.0. In both cases the results obtained (Cronbach's alpha coefficient and percentage of total variance explained) are acceptable (>0.8). The following is a summary of the methods used and the results are analyzed and discussed.

Identification of the variables in the model

Expert method: The expert method is used to know the variables that become part of the factor analysis, eight experts give their individual judgment among them are specialists in health and safety, workers with vast experience as well as professors who investigate in the Thematic areas belonging to the University of Cienfuegos with the aim of reducing the variables to be used in the factor analysis. The judgment of the experts is consistent for both cases and uses the statistical package SPSS version 19.0. Concluding that 15 and 19 factors related to occupational satisfaction and occupational safety and health management can be eliminated, respectively. Finally, there are 12 factors related to research satisfaction and 24 belonging to occupational health and safety management with which the rest of the corresponding analyzes are carried out in the present investigation.

Table 1: Summary of the results obtained in the factorial analysis

Characteristic factors	Results
Coefficient of adequacy (KMO)	Acceptance range (>0.50)
Bartlett's sphericity test	Test is not an identity matrix
Anti-image correlation matrix	Very low values
Sampling Adequacy Measure (MSA)	High values on your diagonal

Optimal scaling: The optimum scaling is performed by the need to research with quantitative variables that guarantee a better fit of the model so that the variables of the categorical type are transformed to metric, using the statistical program SPSS version 19.0. In both cases, the results obtained (Cronbach's alpha coefficient and percentage of total variance explained) are acceptable (>0.8).

Factor analysis: The factor analysis is performed with the 12 variables related to research satisfaction and 24 pertaining to occupational safety and health management derived from the expert method in order to find a way to summarize the information contained in a series of original factors in a smaller series of composite dimensions with a minimal loss of information. All the variables involved are metrics and form an appropriate homogeneous set for factor analysis. Table 1 shows a summary of the results obtained in the factorial analysis related to the variables associated to the work satisfaction and the variables associated with occupational safety and health management and it was concluded that the factorial procedure that is performed may provide satisfactory conclusions.

The principal component method is used when the fundamental interest is focused on predicting or reducing the number of factors necessary to justify the maximum portion of the variance represented in the original set of variables. After performing the processing we observe the commonalities and we conclude that all variables are above 0.5 so that they become part of the study. The matrix of rotated factor weights shows that all factors saturate in some component (according to VARIMAX) obtaining three components with the variables associated to work satisfaction and five to occupational safety and health management in Table 2.

Selection and interpretation of the model: For the choice of the model that best explains the work accident, logistic regression, poisson regression and negative binomial regression are used for the reasons presented previously. Next, the data processing for the 3 selected models is shown, using the statgraphics centurion 15 in which the model is adjusted taking into account the maximum

Table 2: Components of the variables to be used in the mathematical model

Variables	Discriptions
Components associated with job satisfaction	
Management commitment	Aspects related to: training, supervision, medical services, how the company complies with safety regulations and laws
Work conditions	Aspects related to the lighting, ventilation and temperature of the workplace
Supervision	Aspects related to the proximity and frequency with which it is supervised as well as in which they judge thier work
Components associated with occupational safety and health management	
Compliance with legislation	Aspects related to the training of workers, instruction manuals or working procedures, systems for risk assessment, dissemination and implementation of prevention plans, effectiveness of the emergency plan, in addition to notification, research, analysis and registration accidents and incidents
Prevention planning	Aspects related to the information system, prevention plans, identification of corrective actions and verification of assigned objectives
Training in prevention	Aspects related to the training needs, norms of action or working procedures, emergency plans and systematic inspections of the functioning of the system
Updating of occupational risk management	Aspects related to the information that is provided to the worker, systems to identify risks and review of prevention plans
Policy	Aspects related to the concern of management and principles to be followed by all members of the organization

Table 3: Mathematical models adjusted from the regressions used

Variables	Final fit model
MRL	Occupational Accidents = exp(ε)/(1+exp(ε)) where ε = -9,82066-3,60431*SL1-2,56811*GSS1-7,12273*GSS3-3,02833*GSS4-5,35337*GSS5
MRP	Occupational accidents = exp(-5,71197+1,65921*GSS2-4,06726*GSS3-3,89328*GSS4)
MRBN	Occupational accidents = exp(-5,71197+1,65921*GSS2-4,06726*GSS3-3,89328*GSS4)

likelihood which allows to estimate the parameters of a probabilistic model so that they are the more likely from the data obtained.

The variables for the models are: SL1: Management commitment, GSS1: Compliance with Legislation, GSS2: Prevention Planning, GSS3: prevention training, GSS4: Occupational Risk Management Update, GSS5: Policy.

Comparing the statistical significance of the effects according to the Logistic Regression (MRL) Model, the Poisson Regression Model (MRP) and the Negative Binomial Regression (MRBN) Model we obtain the same results in the case of the last two because there is no overdispersion in the data. The components as well as the final model of each regression used are shown in Table 3.

Using the values of the deviation percentage explained and the adjusted percentage: The statgraphics centurion XV program was used to find the percentage values of explained deviation and the adjusted percentage. The results of the goodness of fit analysis show in Fig. 3.

From the previous Fig. 3, the first model mentioned above is finally selected to describe the relationship between work-related accidents and independent variables (management commitment, compliance with legislation, training in prevention, updating of occupational risk management and Prevention) is the logistic regression. This presents a greater percentage of the explained deviation and adjusted percentage of 68 and 26%, respectively.

Use of the akaike and bayes information criterion: To apply the AIC (Akaike Information Criterion) criterion and

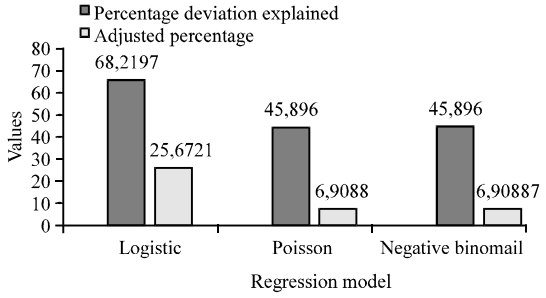

Fig. 3: Percentage of deviation explained and percentage adjusted

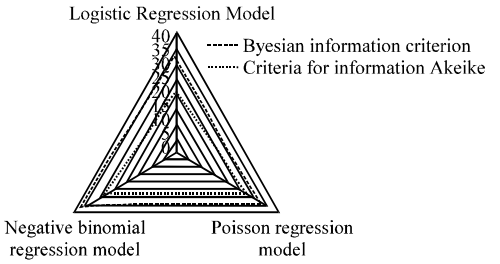

Fig. 4: Values of the Akeike and Bayesian information criteria

the BIC criterion (Bayesian Criterion of Information) the statistical software package SPSS V.20 was used. The results of the criteria are shown in Fig. 4.

From the Fig. 4, the selected model is the logistic regression model. This presents the lowest aic values of 20.96 and BIC of 32.67 compared to poisson which

are: 27.10 and 34.9 and those of the negative binomial regression model whose values are 28.59 and 36.39, respectively.

DISCUSSION

The most relevant results submitted for discussion in this section revolve on two axes. The first is the significant aspects that have been addressed by different theorists of the subject and that emerge from the empirical results of the research. The second methodological considerations on mathematical models and their use in studies of work accident whose contributions are visualized in the results obtained and shown in this study.

Regarding the significant aspects, it should be emphasized that the variables with the greatest statistical significance in the occurrence of occupational accidents obtained in the chosen mathematical model coincide with the aspects identified as weak points in the diagnosis of the process of prevention of occupational risks and with the analysis of claims made in the company. In this way, a congruence and a validation between what is identified with the tools of occupational risk management and the use of generalized linear models in the studies of labor accident, an issue widely used today, is achieved. As a further way of checking this congruence, it is worth clarifying that in the various instruments used to obtain the information, the actors in each case did not coincide because the diagnosis of the prevention process involved specialists in the health and safety of the company and In the application of the surveys that were later used for the use of the mathematical analysis, contributed researcher with experience in the job studied.

In the light of the results obtained in this research, a chain reaction or domino effect can be visualized (as it raises the literature on the subject) in the (variable) causes that with greater statistical significance are currently causing occupational accidents In the object of practical study studied.

From a mathematical statistical point of view, similar results are obtained in terms of statistical significance, since the same variables have a statistically significant effect on the explanation of accidents when adjusting one model or another. Finally, although the percentages of variance explained appear relatively small, the practical importance of any small change in terms of accident occurrence allows us to place ourselves in a situation of small statistical effect but a large social effect.

As for the second line of discussion on methodological issues, a contribution is made by proposing steps and tools that allow the study of safety and health with an approach to process and an integration of mathematical statistics to the solution of problems related to accidents labor.

CONCLUSION

At the end of the present investigation, it is concluded: the research carried out gives a practical contribution to the theoretical assumptions and fieldwork that precede it as it is verified that the use of multivariate statistics and regression analysis, coupled with the application of techniques of security management and occupational health, identify objectively, variables and models that have influenced the occurrence of occupational accidents.

In the analysis of different mathematical models to explain the work accident, it is possible to identify that the logistic regression model yields results with a lower AIC (Akaike information criterion) value of 20,963 and a BIC (Bayesian information criterion) of 32.67, taking into account the percentage of deviation explained and the percentage adjusted with values of 68.2197 and 25.6721 respectively. In turn when applying the AIC criterion, this same mathematical model shows the lowest value in the analysis of goodness of fit 20.963 with a AAIC = 0 which makes it possible to select the aforementioned model and as variables with a significant influence on labor accident the commitment of the direction, the compliance of the legislation, the training in prevention, the update of the management of labor risks and the politics, proposing preventive measures that make possible to extend actions to reduce the labor accident in the studied company.

REFERENCES

Amelec, V. and V. Carmen, 2015a. Design of a model of evaluation of productivity for microfinance institutions. Adv. Sci. Lett., 21: 1529-1533.

Amelec, V. and V. Carmen, 2015b. Validation of a model for productivity evaluation for microfinance institutions. Adv. Sci. Lett., 21: 1610-1614.

Arquillos, A.L. and J.C.R. Romero, 2016. Analysis of workplace accidents in automotive repair workshops in Spain. Saf. Health Work, 7: 231-236.

Bakhtiyari, M., A. Delpisheh, S.M. Riahi, A. Latifi and F. Zayeri *et al*., 2012. Epidemiology of occupational accidents among Iranian insured workers. Saf. Sci., 50: 1480-1484.

Cameron, A.C. and P.K. Trivedi, 1990. Regression-based tests for over dispersion in the Poisson model. J. Econ., 46: 347-364.

Cebador, M.S., J.C.R. Romero, J.A.C. Castrillo and A.L. Arquillos, 2015. A decade of occupational accidents in Andalusian (Spain) public universities. Saf. Sci., 80: 23-32.

Dahlke, G., 2015. Ergonomic criteria in the investigation of indirect causes of accidents. Proceedings of the 6th International Conference on Applied Human Factors and Ergonomics (AHFE) and the Affiliated Conferences, July 26-30, 2015, Elsevier, Las Vegas, Nevada, pp: 4868-4875.

Forastieri, V., 2009. [The time lost due to accidents at work (In Spanish)]. Secur. Environ. Mag., 115: 6-15.

Hyung, Y.K. and S.L. Seung, 2016. A policy intervention study to identify high-risk groups to prevent industrial accidents in republic of Korea. Saf. Health Work, 7: 213-217.

Izquierdo, N.V., 2015. [Improvement of the process of management of safety and health in the work: Example of the energy sector (In Spanish)]. Innovare J. Sci. Technol., 3: 1-10.

Lindsey, J.K., 1999. Introductory Statistics: A Modelling Approach. Oxford University Press, New York, USA.,.

Rahmani, A., M. Khadem, E. Madreseh, H.A. Aghaei and M. Raei *et al*., 2013. Descriptive study of occupational accidents and their causes among electricity distribution company workers at an eight-year period in Iran. Saf. Health Work, 4: 160-165.

Reyes, R.M., J.D.L. Riva, A. Maldonado and A. Woocay, 2015. Association between human error and occupational accidents' contributing factors for hand injuries in the automotive manufacturing industry. Proceedings of the 6th International Conference on Applied Human Factors and Ergonomics (AHFE) and the Affiliated Conferences Vol. 3, July 26-30, 2015, Elsevier, Las Vegas, Nevada, pp: 6498-6504.

Sanchez, A.S., P.R. Fernandez, F.S. Lasheras, F.J.D.C. Juez and P.G. Nieto, 2011. Prediction of work-related accidents according to working conditions using support vector machines. Appl. Math. Comput., 218: 3539-3552.

Tamura, N. and T. Tanaka, 2016. Japan's recent tendencies of accidents in building facilities and workers' accidents in the environment of extreme temperature. Procedia Eng., 146: 278-287.

Tarin, J. and A. Galera, 2016. [OSH management system and work accidents in construction: Empirical evidence of 23 million hours of work in Argentina, Chile, Peru, Mexico and Florida USA (In Spanish)]. Intl. J. ORP. Found., 7: 11-31.

Unsar, A.S. and N. Sut, 2015. Occupational accidents in the energy sector: Analysis of occupational accidents that occurred in thermal and hydroelectric centrals between 2002 and 2010 in Turkey. Procedia Social Behav. Sci., 181: 388-397.

Viloria, A. and A. Parody, 2016. Methodology for obtaining a predictive model academic performance of students from first partial note and percentage of absence. Indian J. Sci. Technol., Vol. 9, 10.17485/ijst/2016/v9i46/107369.