

Assessment Comparison of Mathematical Model, (MLR) First and Second Order Prediction to Experimental Work Absenteeism Data According to Standard Residual Errors

A. Guardiola Mouhaffel^{1*}, F.J. Díaz Pérez³, R. Díaz Martín², A. Ibarra¹ and F. Morán Redondo¹

1. Camilo José Cela University, Madrid, Spain.

2. Distance university of Madrid (UDIMA), Madrid, Spain.

3. Institute of Environment. Faculty of Health Sciences. University San Jorge, Zaragoza, Spain.

*Author corresponded, PhD Professor Aggregate centre of Camilo José Cela University

Abstract

The present study analyses the absenteeism in production systems in series, concretely a factory of assembly of accessories for automotive in Catalonia that counts on 400 workers. The statistics of absenteeism can be analysed from several fronts understanding the direct influence on the worker of the system of shifts, since in this case in particular we will analyse the flexibility available to the workers in the number of hours per month as demanded by the company according to the Reduction of hours to avoid a record of employment regulation and to keep all workers. Data taken between 2011-2016 where the result sought to model will be the number of work casualties, taking as the main reference the average age and hours per shift worked, we will compare the methodology of simple linear regression with a double degree multiple with the objective of To create a predictive model of the risk of work termination.

Keywords: Assembly chain, Labour absenteeism, Predictive model, Production, Multiple lineal regression (MLR)

INTRODUCTION

Today, a large number of large companies choose production lines to increase their productivity. The division of labour in a production line simplifies the work. Untrained workers can substitute for absenteeism specialists. Another way to help simplify human work is the introduction of robust and automatic processes[1]. Replacing workers is an important activity on a production line because production can be stopped if absent workers cannot be replaced. In addition [3], absenteeism in assembly plants usually occurs at random, it is difficult to make adaptations depending on the difficulty of work and it is urgent to replace the worker [1].The production line is normally regulated by an internal structure. Workgroups on a production line are monitored by a supervisor, whose task is to control operations and product quality [3,4]. The work of each member of the group and the cohesion of the group as a unit are vital for achieving normal levels of quality when one of the workers in the production line has to be replaced by another worker [5,8].According to the scientific theory of labour, workers should receive

specialized training on their intervention in the production process before they are part of the production line. The specialization of workers on a production line is valuable because it contributes to the improvement of operations related to quality and efficiency. Nevertheless, There could be a significant reduction in the value of specialization due to the automation of routine tasks. There would be a loss of specialization when the automation of the production cycle requires little learning time[5,8,10]. The quality of the operations in the production line is extremely important because some defects of the products must be rectified as soon as possible; otherwise, costs will increase to improve defective products during processes after the production process. The trained worker contributes to the performance and quality of the product by carrying out its own task and alerting the company about possible deficiencies in the production process or about the deficiencies in the products themselves [10,11]. According to the theory, when a worker with a specialized training is replaced by a non-skilled worker, the probability of defects in the product increases and the likelihood of such a defect decreases. This description of the production process points to the conclusion that absenteeism on the assembly line. The objective of this study is to model work absenteeism in a small assembly line that is forced to flexibly reduce the hours of its 400 workers to avoid in the employment regulation file, **Figure 1** shows how the hours produced over the study period (2011-2015) are below the 64,000 monthly work hours of all its employees. RD Leg. 1/1995 [18] and its subsequent amendments, Art. 34 of the ET, relative to the Working Day, sets its limits of maximum duration to 40 weekly hours as average cash in the annual computation, although it opens the door to its daily concretion and its irregular distribution throughout the year to collective bargaining or to the agreement with the workers' representatives, but demands to respect the minimum rest periods established in the Law, which are of twelve hours between the end of one day and the beginning of the next, and fifteen minutes when the daily continuous day Exceeds six hours, considered as effective working time. **Figure 2** represents of workers hour distribution over the study period.The participants of this study are part of a production line of an original equipment manufacturer which is

represents in **figure 3**. This production line is made up of approximately 400 workers. These people provide eight hours of manual labour in a productive process, that is, that most of them perform mechanical and repetitive work that add value to the product. In no way can we consider that the production process is fully automated [15-17]. The production line employs the latest technology, sophisticated machinery and is characterized by a robust production process. Activities

include the production of metal parts, the production of other parts, painting and quality control. The system of organization is "just in time" [19, 20,21]. This factory applies international standards of product quality, process and management system. The level of production is constant. There is a two-year warranty period during which all defects have to be repaired by the company.

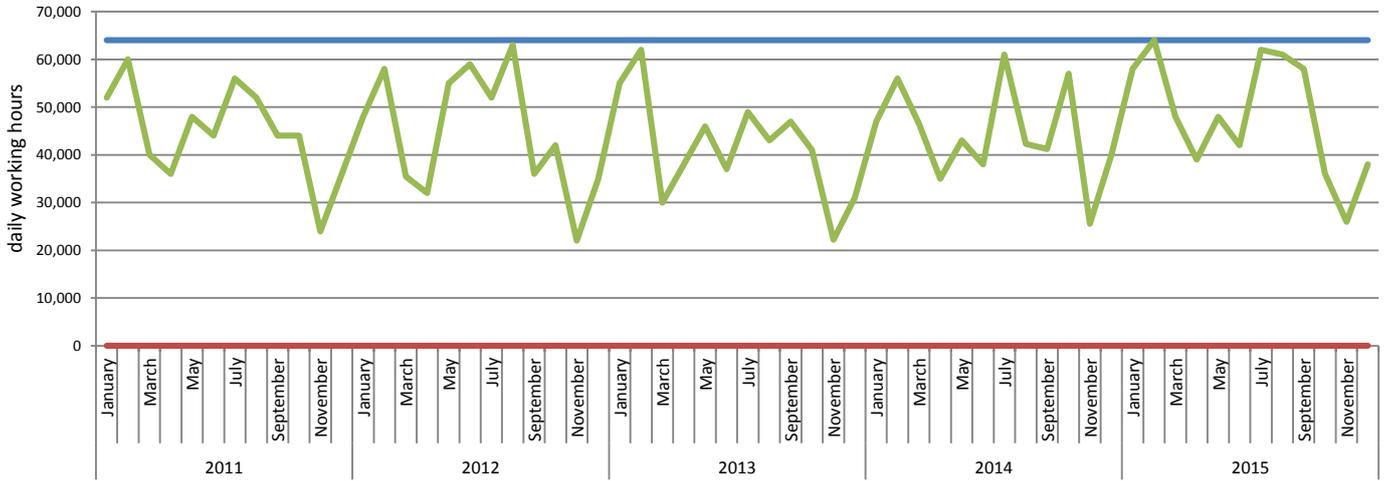


Figure 1. Production hours in the study period

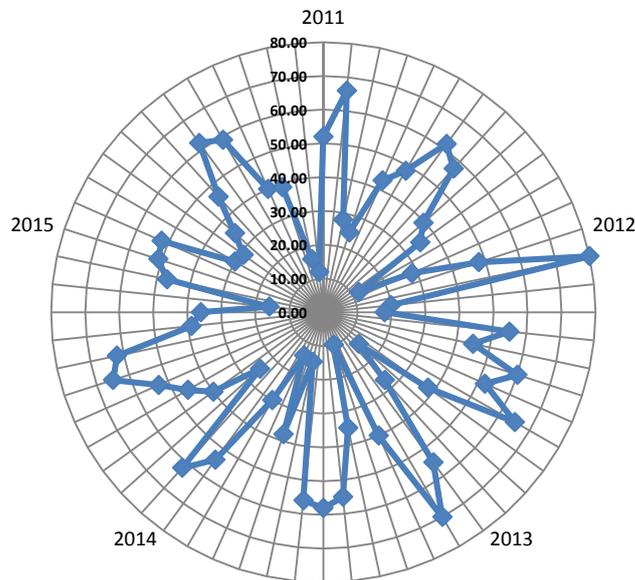


Figure 2: Number of hours absenteeism

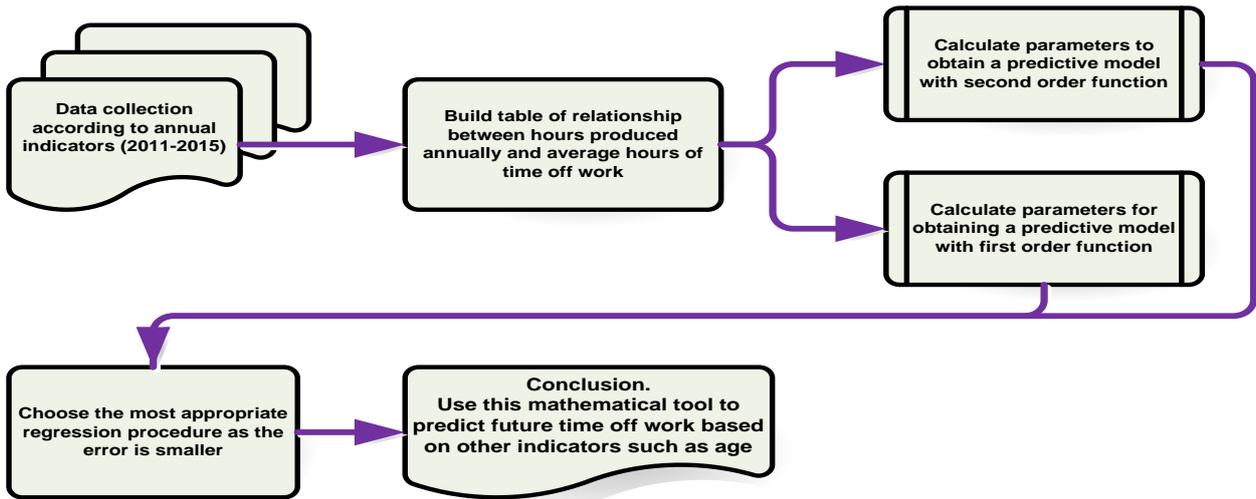


Figure 3. Schedule of research procedure

METHODOLOGY

Multiple linear regression analysis seeks to establish a relationship between a dependent variable in this case, labour absenteeism and two or more independent variables. Those are the employee absenteeism models,

$$Y = \beta_1 X_1 + \beta_2 X_2 + \epsilon \quad (1)$$

$$Y = \beta_1 \vartheta + \beta_2 \vartheta^2 + \beta_3 w + \beta_4 w^2 + \beta_5 \vartheta w + \epsilon \quad (2)$$

In a method based on a data record which is normally done based on the least squares method with Minimize the difference between the observed and estimated values [12-14]. The values obtained from the predictive model should have little or no correlation with the data from the baseline variables of the study to avoid problems caused by multicollinearity [9]. In this way we get some parameters first and a last independent term is called residual (or adjusted error).

Regression analysis is used differently to test the general significance (test F) of the equation Y (labour abstention) for the significance of each regression coefficient [7]. In order to obtain valid results of these tests AND distributed independently, with a mean of zero and a constant variance of r^2 , coefficient of determination, which not only indicates the goodness of the fit, but can also be interpreted as the amount of variation of the dependent variable explained by the regression equation [4, 5].

Before selecting a regression model, it is advisable to corroborate the matrix of variables both dependent (labour absenteeism) and predictive base variables (total hours produced and age of workers) to see if the linear regression is appropriate and, if so, what model should be adopted. Here a linear equation might be sufficient, but to take care of any curvature a second order term was incorporated. Both dispersion diagram matrices also indicate a strong correlation between independent variables models [5-7].

The actual regression analysis was performed iteratively with the IBM SPSS Statistics software package, version 21 [9, 11]. During these iterations, outliers were removed from the residual conditions, as calculated during the linear regression analysis were Based on a distribution of these error terms. The extent to which these terms were removed from the standard error terms.

RESULTS

With the actual values of the data sets and it shows that both equations perform satisfactorily. This figure can also be interpreted as the display of the R^2 value given above.

FIRST-ORDER PREDICTIVE MODEL :

Table 1. Correlation Matrix first-order

| | ABSENTEEISM | X1 |
|-------------|-------------|-------------|
| ABSENTEEISM | 1 | |
| X1 | -0,04759468 | 1 |
| X2 | 0,967856069 | 0,011051467 |

$$Y = -23.00 - 0.192X_1 + 0.0016X_2 \quad (3)$$

$R^2 = 0,920 \quad R^2_{Adj} = 0,921 \quad F = 70.67$
Durbin-Watson = 2.02

SECOND-ORDER PREDICTIVE MODEL :

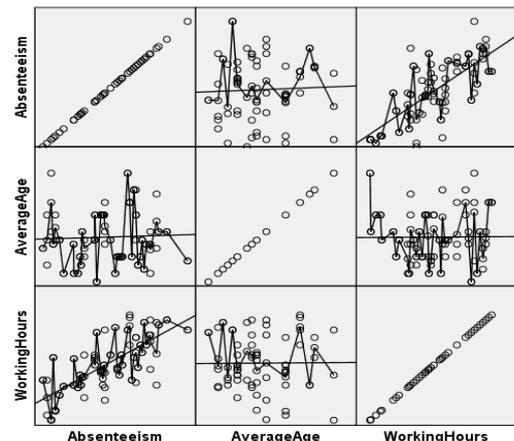


Figure 4. Regression of Intersection of input data.

Table 2. Matrix correlation linear model of second order.

| | ABSEN | g | g2 | w | w2 |
|-------|--------------|-------------|-------------|-------------|-------------|
| ABSEN | 1 | | | | |
| g | -0,047594682 | 1 | | | |
| g2 | -0,055919477 | 0,99872585 | 1 | | |
| w | 0,967856069 | 0,011051467 | 0,000442601 | 1 | |
| w2 | 0,969821658 | 0,048909868 | 0,036358695 | 0,990201653 | 1 |
| gw | 0,802818139 | 0,495685835 | 0,488404946 | 0,869048297 | 0,872307017 |

$$Y = 62.50 - 6.989 + 0.13 g^2 + 3,27E-03w + 8,79E-09w^2 - 6.17E-5g w \quad (4)$$

$$R^2 = 0,968 \quad R^2_{Adj} = 0,952 \quad F = 36,64$$

Durbin-Watson = 2.70

EVALUATION DATA :

Figure 4 shows the combined regressions between the input data, which gives us an indication of the correlative index between said data. R^2 expresses the proportion of variance of the dependent variable which is explained by the independent variable, which is best in the second order model; adjusted R^2 is an improvement of R^2 according to the number of variables [7-9].

The data of the Durbin-Watson statistic is a data that determines independences depending data, where values less than 2 indicate positive self-correlation and those higher than 2 self-negative correlations. Thus the Durbin-Watson coefficient is better adjusted in the linear regression model than in the second order model. As both the values of the two models are between 1.5 and 2.5, we can assume that depending data are independent. The RSME (mean square root deviation) of the first and second grade models 0.81 and 2.34 suggesting that the estimation equation presented above works reasonably well during the study period.

In summary the error was distributed with a constant variance. The Durbin-Watson statistic suggested that the errors were not auto correlated for both models [19-21]

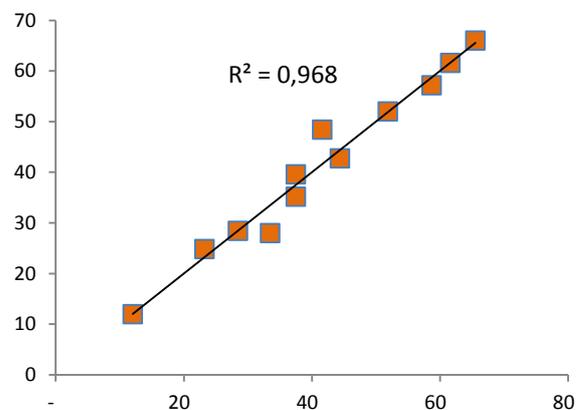


Figure 6: Fitting data predicting Second-order

The results of the show of the predictive data with respect to those of the independent variable are shown in Figures 5 and 6 where we check the reliability of the sampling according to the best fit of the regression line.

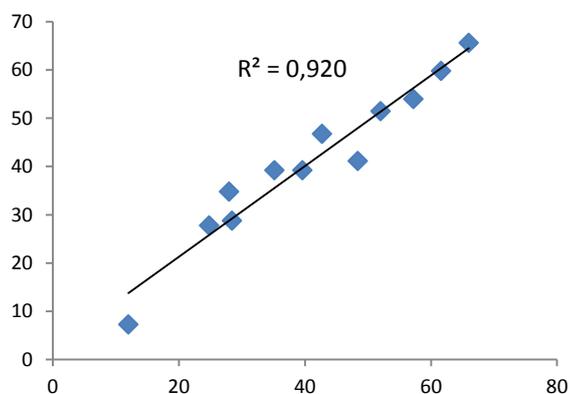


Figure 5: Fitting data predicting first-order

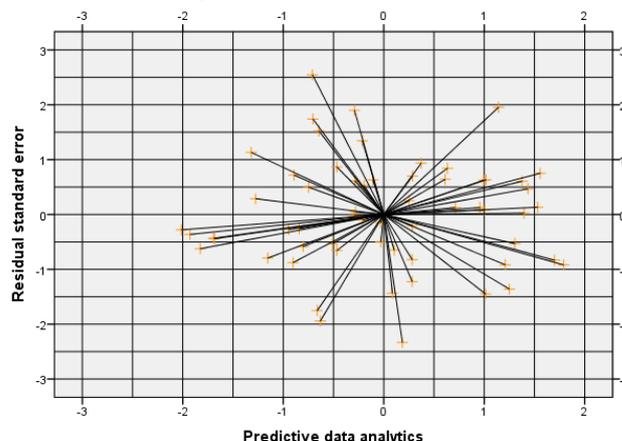


Figure 7.Residuals regression standard error

CONCLUSIONS AND DISCUSSIONS

In the present study, work absenteeism was modeled in an assembly factory, which is under 10% with a regression coefficient R^2 of 0.922 and 0.968 for the first and second order models, the first being more accurate. Otherwise the error in the fit as shown in **Figure 7** is much higher in the first-order model, with practically twice as much as the difference between the two in the third quarter.

Results from the study confirms that the second-order model EMPLOYEE ABSENTEEISM AND PSYCHOSOCIAL CONSEQUENCES chain model gives a reasonable approximation in modelling the relationship between staffing level and organization performance in the case where workers may be leave from the work place. The model does seem to overestimate the probability of accepting work and the utilization. Reasons for this are being investigated. However, the error is quite small.

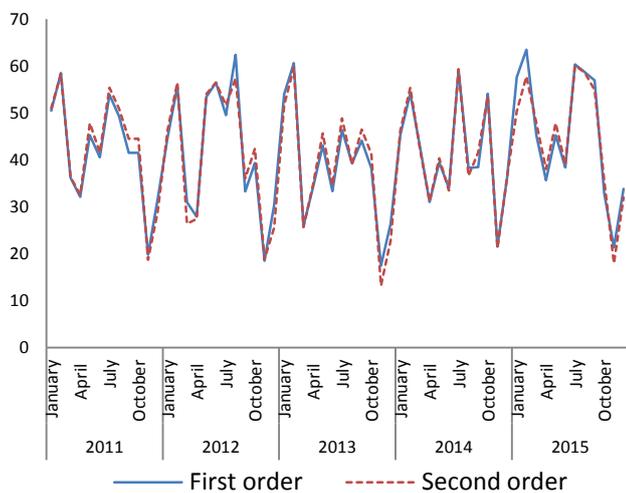


Figure 8. Predicting fittings models

Figure 8 represents the fitting models predictions that ten percent of absenteeism is very high for any production line and could be considered a valid level for many production lines. It is important to emphasize that this research has been done on thousands of observations of a specific case and that new studies should be analysed to confirm its validity. However, we can conclude that the lack of relationship between these variables is due to the transformation of the worker into a mechanical part of the process, where the relevant physical capacities are relevant. We also highlight the enormous capacity of this organization to reduce customer complaints during the observed period [16,17]. This indicates that workers and engineers have managed the problems successfully and that continuous improvement has been developed and focused to improve customer satisfaction. In this case, we can see that mechanical work is compatible with creative work and that, for skilled workers, the creative part is what differentiates them from non-experts [19].

The aim of this research is to impact on the design of business strategies and, above all, on the design of union strategies in promoting activities that allow workers to generate value with

their skills and experience. The possibility of forecasting the Absenteeism by means of a regression and can parameterize the values of this regression according to production hour's values and age average workers done the opportunity to forecasting the cost and production efficiency in confection process [20,21]. Thus the proposed model provides a useful tool for managers in measuring the relationship between staffing level and performance of workers in a group for three aspects: the probability of a group in carrying out work, the availability of workers, and the group utilization. The model can be used in setting up the workforce size of a newly formed group when the arrival rate of work, the processing rate of workers, and the average time of on leave by workers are estimated.

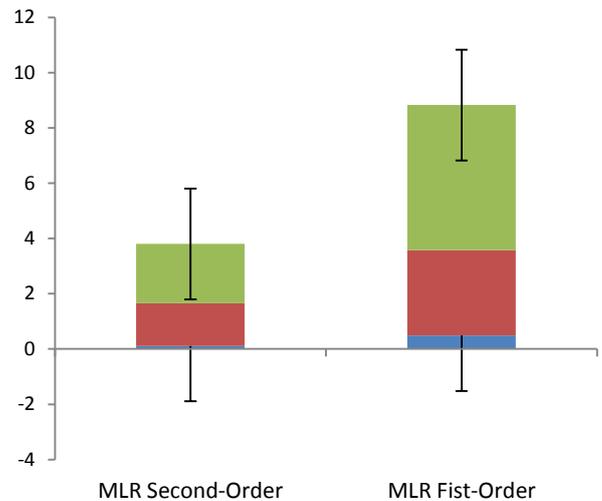


Figure 9: Box-Plot error between two models

REFERENCES

- [1] M.J.Azofra, (2000). Methodological Papers Center for Sociological Research. ISBN: 978-8474762723
- [2] García Fernando M., (2003). The analysis of social reality. Research Methods and Techniques 3rd Ed.
- [3] Segarra, A. y Callejón, M. (2002). New firms' survival and market turbulence: new evidence from Spain, Review of Industrial Organization, Volume 20, Issue 1, pp 1-14.
- [4] Agarwal, B.L., (1991). Basic Statistics, Second Edition, New Delhi, Wiley Eastern Ltd.
- [5] Chatfield, C., (1994). The Analysis of Time Series-An Introduction, Fourth Edition, London, Chapman and Hall.
- [6] Holmström, L., Koistinen, P., Laaksonen, J. and Oja, E., (1997). Neural and Statistical Classifiers: Taxonomy and Two Case Studies, Jan. 1997, IEEE Trans. Neural Networks, Volume 8, Issue 8, pp 5-15.
- [7] Montgomery, D. C. and Lynwood, A. J., (1996). Forecasting and Time Series Analysis, New York, McGraw-Hill.

- [8] D.C. Montgomery, E.A. Peck, G.G. Vining (2006). Introduction to linear regression analysis John Wiley & Sons, Hoboken
- [9] Mathur, S., (1998). Stock Market Forecasting Using Neural Networks-An MBA Project Report Submitted to School of Management Studies, IGNOU, New Delhi.
- [10] Mathur, S., Shukla, A.K. and Pant, R.P., Agra, Sept. 22-23, 2001, International Conference on Optimization Techniques and its Applications in Engineering and Technology, A Comparative Study of Neural Network and Regression Models for Estimating Stock Prices.
- [11] Roadknight, C.M., Balls, G.R., Mills, G.E. and Palmer-Brown, D., Modeling (1997). Complex Environmental Data, July 1997, IEEE Trans. Neural Networks, Volume 8, Issue 4, pp852-861.
- [12] Shukla, M.C., and Gulshan, S.S., (1980). Statistics, Fourth Edition, New Delhi, S.Chand and Company.
- [13] J.C. Lam, K.K.W. Wan, D. Liu, C.L. Tsang, (2012). Multiple regression models for energy use in air-conditioned office buildings in different climates Energy Convers Manage, Volume 51, Issue 12, pp. 2692–2697
- [14] M.G. Lawrence, (2005). The relationship between relative humidity and the dew point temperature in moist air: a simple conversion and applications Bull Am Meteorological Soc, Volume 86, pp. 225–233
- [15] Park S.J., Lee Y.L. (2015). Numerical analysis on performance variation of linear compressor equipped with economizer. International Journal of Applied Engineering Research Volume 13, Issue 3, pp 5411-5420
- [16] Ricardo Díaz Martín, GastonSanglier Contreras and Adib Guardiola Mouhaffel, (2015). Calculation of the Equilibrium Moisture Content of Pellets and Olive Stones using the Henderson Mathematical Model. International Journal of Applied Engineering Research, Volume 12, Issue 10, pp. 31681-31690
- [17] Díaz Martín, Contreras, Adib Guardiola Mouhaffel, C. Martínez Domínguez, B. Arcones, F. Morán Redonda, (2016). Using Multiple Regression Analysis Lineal to Predict Occupation Market Work in Occupational Hazard Prevention Services. International Journal of Applied Engineering Research, Volume 3, Issue 12, pp. 283-288
- [18] Government of Spain. (1995) Royal Legislative Decree 1/1995, dated March 24, approving the revised text of the Workers' Statute Act (effective until November 13, 2015).
- [19] Christopher, M., McKinnon, A., Sharp, J., Wilding, R., Peck, H., Chapman, P., Juttner, U. and Bolumole, Y. (2002). Supply chain vulnerability. Cranfield University. Final Report for the Department for Transport, Local Government and Regions.
- [20] Burtonshaw-Gunn, S. A. (2002).Examining risk and supply chain collaborative working in the UK construction industry. In 2nd International Research Seminar on Risk and the Supply Chain
- [21] Mentzer, J. T., Dewitt, W., Keebler, J. S., Min, S., Nix, N. W., Smith, C. D. and Zacharia, Z. G. (2001). Supply Chain Management, Thousand Oaks, CA: Sage.