Application of logistic regression in industrial maintenance management

Aplicación de la regresión logística en la gestión de mantenimiento industrial

HERRERA-SÁNCHEZ, Gustavo†*, SILVA-JUÁREZ, Alejandro, MORÁN-BRAVO, Luz del Carmen and DESAMPEDRO-POBLANO, Héctor Manuel

Universidad Tecnológica de Puebla

ID 1st Author: *Gustavo, Herrera-Sánchez* / **ORC ID**: 0000-0001-5276-5062, **Researcher ID Thomson:** F-6595-2018, **arXiv Author ID:** herreragh, **CVU CONAHCYT ID**: 459805

ID 1st Co-author: *Alejandro, Silva-Juárez* / **ORC ID**: 0000-001-8473-9803, **Researcher ID Thomson:** F-6969-2018, **arXiv Author ID**: alejandrosilva1 y CVU: 637028

ID 2nd Co-author: *Luz Del Carmen, MORÁN-BRAVO* / **ORC ID**: 0000-0002-7096-2075, **Researcher ID Thomson:** G-2686-2018, **arXiv Author ID**: XVRU3-JP9XUY, **CVU CONAHCYT ID** 75419

ID 3rd Co-author: *Héctor Manuel, Desampedro-Poblano* / **ORC ID**: 0000-0002-9235-9007, **Researcher ID Thomson:** Q-8174-2018, **arXiv Author ID**: V9HG4K-7ZVVF3, **CVU CONAHCYT ID** 548477

DOI: 10.35429/JEDT.2023.12.7.1.7 Received January 30, 2023; Accepted June 30, 2023

Abstract

In the area of industrial maintenance, the application of statistical methods is essential, in that sense, the purpose of this analysis is to explore logistic regression as an element of industrial maintenance management. By means of logistic regression, a predictor equation for the response variable, machine failure, is obtained by correlating it with categorical and continuous predictor variables. The continuous explanatory variables are machine age, mean time between failures, mean time to repair and the categorical ones are application of preventive and corrective maintenance. The results obtained indicate that only the explanatory variable preventive maintenance is significant to the response variable by applying the Wald test and this result was also validated with goodness-of-fit tests. Logistic regression is more used in other areas, such as health, however, in maintenance categorical variables are used such as machine with autonomous maintenance whose result is yes/no, therefore, it is important to incorporate a regression model that considers different types of independent variables, in addition to the use of emerging technologies of Industry 4.0 such as Machine Learning for the prediction of scenarios for efficient maintenance management.

Correlating, Logistic Regression, Predictor, Preventive, Corrective

Resumen

En el área de mantenimiento industrial es primordial la aplicación de métodos estadísticos, en ese sentido, el propósito de este análisis es explorar la regresión logística como un elemento de la gestión del mantenimiento industrial. Mediante la regresión logística se obtiene una ecuación predictora para la variable de respuesta, máquina falla correlacionándola con variables predictoras categóricas y continuas. Las variables explicativas continuas son edad de la máquina, tiempo medio entre fallas, tiempo medio para reparar y las categóricas son aplicación de mantenimiento preventivo y correctivo. Los resultados obtenidos indican que únicamente la variable explicativa mantenimiento preventivo es significativa a la variable de respuesta mediante la prueba de Wald y también se validó este resultado con pruebas de bondad de ajuste. La regresión logística es más utilizada en el otras área, como de la salud, sin embargo, en mantenimiento se utilizan varias variables categóricas como máquina con mantenimiento autónomo cuyo resultado es si/no, por ello, es importante incorporar un modelo de regresión que considera a diferentes tipos de variables independientes, además de la utilización de las tecnologías emergentes de la Industria 4.0 como Machine Learning para la predicción de escenarios para una eficiente gestión del mantenimiento.

Correlación, Regresión Logística, Predictor, Preventivo, Correctivo

Citation: HERRERA-SÁNCHEZ, Gustavo, SILVA-JUÁREZ, Alejandro, MORÁN-BRAVO, Luz del Carmen and DESAMPEDRO-POBLANO, Héctor Manuel. Application of logistic regression in industrial maintenance management. Journal-Economic Development Technological Chance and Growth. 2023. 7-12: 1-7

† Researcher contributing first author.

Introduction

Logistic regression is a technique that allows us to establish a relationship between a discrete, mainly dichotomous variable with possible outcomes of accepted or not accepted and predictor variables, which can be quantitative or qualitative in a probability of occurrence of the particular phenomenon. Logistic regression has been used in predictive maintenance for machine failure analysis (Battifarano et al, 2019; Yongyi et al, 2019; Yugapriya et al, 2022), in machine efficiency assessment through Overall Equipment Effetiveness (OEE) (Borucka, Grzelak, 2019), its main use has been in the areas of medicine and psychology (Alzen et al, 2018; Oyekale, 2022; Zabor et al, 2022).

The purpose of this research is to apply a logistic regression model of the state of failure and state of function of machines for maintenance management using logistic regression, thus, to analyse and evaluate how dichotomous independent variables influence the probability of occurrence of the states of a machine, i.e., in the state of operation (State of Functioning, SoFu) and in the state of failure (State of Failure, SoFa).

Materials and methods

To predict when a machine is in SoFu or SoFa, there are different regression methods such as simple linear (Le, T at al, 2014; Teng, et al, 2016), multiple linear regression (Bicharra et al, 2014), non-linear regression (Mosallam et al, 2011), Monte Carlo simulation (Srivastava et al, 2020), determination of reliability, maintainability and availability indicators using continuous probability distributions (Mora, 2009; Ramesh, Krishman, 2017).

However, there are situations where the random variables of study are discrete and dichotomous in the area of maintenance, for example, if the machine failed, if the machine was applied preventive maintenance, if the machine has an autonomous maintenance routine.

Therefore, the statistical regression models mentioned above are not suitable for the prediction of the operating status as a mathematical model involving discrete variables and continuous variables is required.

Logistic regression does not require a linear relationship between the response and predictor variables nor does it emphasise the assumptions of linearity, normality, homoscedasticity and level of measurement.

Logistic Regression Model

The mathematical model of a logistic regression allows to observe the possible correlations between independent variable $X_1, X_2, X_3, \ldots, X_n$, which can be continuous or discrete with a discrete and dichotomous dependent variable Y, usually taking the values of 0 and 1. Also the predictor variables can be qualitative or quantitative. The logistic regression model is based on the logistic or sigmoid function, equation 1.:

$$f(x) = \frac{e^x}{1+e^x} = \frac{1}{1+e^{-x'}} \quad (1)$$

The sigmoid function fulfils $\lim_{x\to\infty} f(x) = 0$ y $\lim_{x\to\infty} f(x) = 1$, the transformation from logistic regression to linear regression $Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_n X_n$, substituting in the sigmoid function the *x* by the linear function $\beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n$, is obtained (equation 2):

$$P(Y) = \frac{1}{1+e^{-(\beta_0+\beta_1 x_1+\cdots+\beta_n x_n)}} = \frac{e^{\beta_0+\beta_1 x_1+\cdots+\beta_n x_n}}{1+e^{\beta_0+\beta_1 x_1+\cdots+\beta_n x_n}} \quad (2)$$

Through this equation, the possible values of *Y* are 1 and 0, which is interpreted as the probability of occurrence of *Y*, the closer the value of P(*Y*) is to 1, the more likely it is that *Y* will occur.

On the other hand, if we have a dependent variable *Y* which is dichotomous with values of 1 or 0 and n independent variables $X=(X_1, X_2, X_3, \ldots, X_n)'$, the generalised model for the logistic regression is (equation 3):

$$P(Y=1|X=x) = px = \frac{e^{\beta_0+\sum_{i=1}^{n} x_i}}{1+e^{\beta_0+\sum_{i=1}^{n} x_i}} \quad \mathbf{(3)}$$

For maintenance management (Eagle Technology, 2020; MaintainX, 2021; Sellitto, 2020), logistic regression is considered appropriate, since we have continuous and discrete predictor variables, for example, continuous variables can be the cost of maintenance.

ISSN-2524-2024

The age of the machine, the Mean Time Between Failures (MTBF), Mean Time To Failure (MTTF), Mean Time To Repair (MTTR), among others. For discrete and dichotomous variables, one can start by asking what happens if corrective maintenance, preventive maintenance, predictive maintenance, the technician who performed the maintenance, etc., was applied.

The data for this test are shown in table 1. Where the response variable is and machine failure takes the value of yes and no when the machine is available for its function, with the following predictor variables being: machine age (ME), MTBF, MTTR, preventive maintenance (PM) and corrective maintenance (CM).

Machine failure	EM	MTBF	MTTR	MP	MC
Si	7	9652	9	0	0
No	3	612	3	1	0
Si	6	6769	6	0	0
No	4	7344	8	0	0
Si	3	1198	8	0	1
No	4	9572	7	1	0
Si	2	8212	6	0	1
No	7	8498	7	1	1
No	5	5475	7	0	0
No	9	1272	8	0	1
No	6	9790	7	1	1
No	7	2469	4	0	0
No	3	7132	2	1	0
Si	10	6254	2	0	0
Si	7	6054	3	0	0
Si	4	9815	2	0	1
Si	2	9367	9	0	0
Si	5	408	4	1	1
Si	7	883	8	1	0
No	9	1534	6	1	1
No	11	4745	11	1	0
No	13	899	15	0	1
Si	3	3752	20	0	0
Si	10	5194	17	0	0
No	9	939	10	1	1
No	20	2657	4	0	1
Si	17	9918	15	1	1
Si	11	7324	17	0	1
No	18	6689	8	0	1
Si	15	743	8	0	1
Si	2	1200	2	0	1
No	5	4001	1	1	0
No	3	2500	5	1	0
No	8	2356	6	1	0

Table 1 Data for logistic regression
Source: Own Elaboration

Results and discussion

Using minitab software, the regression is solved by the logistic regression method with the Logit link function, see equation 4.

$$Y' = 0.15 - 0.125EM + 0.00008MTBF + 0.1015MTTF - 1.928MP_{si} + 0.597MC_{si} \quad (4)$$

The equations for the discrete variables are shown in table 2. The positive coefficients of the predictor equations indicate that the machine is likely to fail as the value of the predictor increases, on the other hand, the negative coefficients indicate that the event of machine failure is less likely as the value of the predictor increases.

MP.	MC.	Ecuation
No	No	$' = 0.147 - 0.1246EM + 0.00008MTBF +$ $1015\,MTTR$
No	Yes	$Y' = 0.743 - 0.1246EM + 0.00008MTBF + 0.1015MTTR$
Yes	No	$Y' = -1.781 - 0.1246EM + 0.00008MTBF + 0.1015MTTR$
Yes	Yes	$Y' = -1.184 - 0.1246EM + 0.00008MTBF + 0.1015MTTR$

Table 2 Model equations
Source: Own Elaboration based on minitab

Analysing the variance table, see table 3, it is observed that the explanatory variable of predictive maintenance is statistically significant to the response variable with a confidence level of 95%, therefore, this categorical factor significantly influences the dependent variable and the machine failure.

Source	GL	Chi square	Value - p
Regression	5	7.43	0.190
Machine age	1	1.65	0.199
MTBF	1	0.42	0.516
MTTR	1	1.06	0.302
M. Preventive	1	5.22	0.022
M. Corrective	1	0.46	0.499

Table 3 Analysis of Variance
Source: Own elaboration based on minitab: Own elaboration based on minitab

The other p-values of the predictor variables are not statistically significant, indicating that the regression model can be reduced without these terms.

Verifying that the logistic regression predictor equation fits the data, the following goodness-of-fit tests were performed:

– Deviance with value - p = 0.112.
– Pearson's test p = 0.232
– Hosmer - Lemeshow test p = 0.873

With these results, the p-values are greater than the significance level of the investigation, therefore, there is no significant statistical evidence to conclude that the model does not fit the data.

Using the Normit link function instead of Logit, the goodness of fit test results do not show a significant change:

– Deviation with value - p = 0.114.
– Pearson's test p = 0.238
– Hosmer - Lemeshow test p = 0.733

With the Normit link function, the Wald test, no significant change is observed in the ANOVA table, the preventive maintenance variable remains statistically significant with p = 0.017 at the 95% significance level.

In relation to multiconeality, the results obtained from the coefficients of the predictor equation are not severe, see table 4. The variance inflation factors are close to unity (Del Valle and Guerra, 2012), which indicates that there is no correlation between the predictor variables, therefore, the model is reliable to forecast

Coefficient			
Term	Coef	EE coef	FIV
Constant	0.15	1.20	
Machine age	-0.1246	0.0971	1.33
MTBF	0.00008	0.000124	1.05
MTTR	0.1015	0.0984	1.14
M. Prev_Si	-1.928	0.844	1.03
M. Corr_Si	0.597	0.884	1.22

Table 4 Variance inflation factors
Source: Elaboration based on minitab

In table 5, the odds ratio (ODDS) for continuous predictors reveals that the variable age of the machine has a value of 0.8828, which indicates that the event the machine fails is unlikely to occur because there is a negative association. If we perform the inverse operation 1/0.8828 = 1.13 it indicates that there is a probability of 1.13 times that the machine will work. For MTTR and MTBF values are greater than unity (Salas, 1996), i.e. there is a positive association between the event and the machine is less likely to fail, however, for the explanatory variable MTBF its ODD is practically 1 showing no association.

The MTTR denotes a positive association, since, there is a 1.1 times probability of machine failure considering all other constant values.

	Odds ratio	IC de 95%
Machine age	0.8828	(0.7299, 1.0679)
MTBF	1.0001	(0.9998, 1.0003)
MTTR	1.1069	(0.9127, 1.3423)

Table 5 Likelihood ratios for continuous predictors
Source: Elaboration based on minitab

With respect to the odds ratio of the categorical predictors, table 6, the preventive maintenance variable is 0.1454 times more likely that the machine fails, if we obtain its inverse 1/0.1454 = 6.88 times that the machine does not fail if we apply preventive maintenance keeping the other explanatory variables constant. The association is positive for corrective maintenance, when corrective maintenance increases, the machine is more likely to fail.

Level A	Level B	Odds	IC de 95%
Preventive maintenance			
Yes	No	0.1454	(0.0278, 0.7599)
Corrective maintenance			
Yes	No	1.8171	(0.3215,10.2691)

Table 6 Odds ratios for categorical predictors
Source: Elaboration based on minitab

Figure 1 shows an area under the curve of 0.8021, according to Sweest's (1988) criteria, the area under the curve, known as ROC, is in the range of 0.7 to 0.9 and this indicates that it has an acceptable discriminant capacity for when the machine fails or does not fail. On the other hand, according to Hilbe (2015), values from 0.5 to 0.65 have low predictive ability, values from 0.65 to 0.80 have moderate ability, values between 0.8 and 0.90 indicate strong predictive ability and values greater than 0.9 indicate high predictive ability, but this last relationship almost never happens. It is desirable that it is greater than 0.9 in order to have a higher sensitivity and specificity and thus obtain few erroneous results in the variable y'.

The most usual is to find a curve between 0.7 and 0.9 (del Valle, n.d.) with an overlap between the sensitivity or true positive rate, TPR, and the specificity or false positive rate, FPR. For our study it means that the sensitivity has an acceptable discrimination in predicting when the machine fails and when it does not fail.

Figure 1 ROC curve for the study

On the other hand, according to the results of the analysis of variance in table 3, the logistic regression is solved with the machine failure versus preventive maintenance, resulting in the following equation:

$$y' = 0.619 - 1.918\ MP_{si} \tag{5}$$

Comparing equations 4 and 5, there is no big difference in the coefficient of the preventive maintenance variable, as well as in the goodness of fit tests: for deviation, the p-value is 0.184 and for Pearson the p-value is 0.371.

For the ODD with respect to the preventive maintenance variable there is no significant difference, see table 6 and 7.

Level A	Level B	Odds	IC de 95%
Preventive maintenance			
Yes	No	0.1469	(0.0305, 0.7099)

Table 7 Probability relationships for machine failure vs. Preventive Maintenance

Acknowledgements

Sincere thanks to the Universidad Tecnológica de Puebla for their support in providing the financial resources for this research.

Conclusions

With logistic regression applied to maintenance management, we have a statistical tool that allows us to make predictions of whether a machine is in its functional state or in a state of failure using categorical variables, not only with continuous variables. Goodness-of-fit tests were performed to test the suitability of the model in different scenarios with the Logit and Normit link functions, with no significant differences.

On the other hand, two predictor equations were obtained, one model with all the variables of the initial analysis and another one only with the significant variable without having a relevant difference. With respect to the ROC curve, the model is adequate for the prediction of the response variable.

Like any process, it can be improved, as other predictor variables can be incorporated, such as the cost of total maintenance, corrective, preventive, predictive, maintenance policies with respect to the machines, criticality, all with the aim of strengthening the prediction model for a relevant industrial maintenance management.

Finally, incorporating the disruptive technologies of Industry 4.0, since companies tend to be cyber-physical systems that have vertical and horizontal integration, and data collection, in accordance with the ISA - 95 standard, with sensors on the machines will lead to the use of data in the cloud, the analysis of big data through Machine Learning and real-time decision making. This leads to a new maintenance management 4.0. Logistic regression will then use sensitive data to predict machine failure through the use of Machine Learnig with real-time data collection for increased productivity and competitiveness in industrial maintenance.

References

Alzen, J., Langdon, L., & Otero, V. (2018). A logistic regression investigation of the relationship betwen the Learning Assistant model and failure rates in introductory. *STEM course. International Journal of STEM Education*, 5, 56, https://doi.org/10.1186/s40594-018-0152-1

Battifarano, M., DeSmet, D., Madabhushi, A. & Nabar, P. (2018). Predicting Future Machine Failure from Machine State Using Logistic Regression. *arXiv preprint* arXiv:1804.06022, 1-6.

Bicharra, A., Ferraz, I., Viterbo, J., & de Pavia, D. (2014). Applying Multiple Regression Analysis to adjust Operational Limits in Condition-Based Maintenance. *Bazzan, A. Pichara, K. (eds) Advances in Artificial Intelligence – IBERAMIA 2014. IBERAMIA 2014. Lecture Notes in Computer Science()*, vol 8864. Springer, Cham, 754-764 https://doi.org/10.1007/978-3-319-12027-0_61.

Borucka, A., & Grzelack, M. (2019, 9(22), 4770). Application of Logistic Regression for Production Machinary Efficiency Evaluation. *Applied Science*, https://doi.org/10.3390/app9224770.

del Valle, A.R. (s.f.) Curvas ROC (Receiver-Operating-Characteristic) y sus aplicaciones. Sevilla https://idus.us.es/bitstream/handle/11441/63201/Valle%20Benavides%20Ana%20Roc%C3%ADo%20del%20TFG.pdf: Universidad de Sevilla.

Del Valle, J., & Guerra, C. (2012). La multicolinealidad en modelos de Regresión Lineal Múltiple. *Revista Ciencias Técnicas Agropecurias Vol 21 No. 4* https://scielo.sld.cu/scielo.php?script=sci_arttex&pid=S2071-00542012000400013.

Eagle Technology, Inc. (20 de abril de 2023): *MAINTENANCE KEY PERFOMANCE INDICATORS E-book 2020*. Obtenido de https://www.caba.org/wp-content/uploads/2021/07/IS-2021-186.pdf

Regression. Boca Raton, Florida: CRC Press Taylor & Francis Group. https://ftp.idu.ac.id/wp-content/uploads/ebook/ip/REGRESI%20LOGISTIK/Practical%20Guide%20to%20Logistic%20Regression%20(%20PDFDrive%20).pdf

Le, T., Luo, M., Zhou, J., & Chan, H. (2014). Predictive maintenance decisión using statical linear regression and kernel methods. *Proceedings of the 2014 IEEE Emerging Technology and Factory Automation (ETFA), 1-6*, doi: 10.1109/ETFA.2014.7005357

MaintainX. (15 de abril de 2023). *Understanding Maintenance Metrics and KPIs 2021*. Obtenido de https://f.hubspotusercontent10.net/hubfs/9125465/downloadables/Understanting%20Maintenance%20Metrics%20and%20KPIs%20PDF%20Guide%20CLE.pdf

Mora. A. (2009). *Mantenimiento, Planeación, ejecución y control*. México: Alfaomega. https://elvisjgblog.files.wordpress.com/2019/11/mantenimiento-planeacic3b3n-ejecucic3b3n-y-control-alberto-mora-gutic3a9rrez.pdf

Mosallam, A., Byttner, S., Svensson, M., & Rögnvaldsson, T. (2011). *Nonlinear Relation Mining for Maintenance Prediction. 2011 Aerospace Conference (págs. 1-9)* doi: 10.1109/AERO.2011.5747581. Big Sky, MT, USA: IEEE.

Oyekale, A. (2022) Factors Influencing Willingness to Be Vaccinated against COVID-19 in Nigeria. *International Journal of Environment Research and Public Health, 6816*; https://doi.org/10.3390/ijerph19116816.

Ramesh, K., & Krishman, V. (2017) Reliability Availability and Maintainability Analysis of Systems. *Journal of Mathematics and Informatics Vol 11, 131-141* https://doi.org/10.22457/jmi.v11a17

Salas, M. (18 de mayo de 2023). *Estadística Española* Vol. 38, Num. 14, 1996. Obtenido de https://metodos-avanzados.sociales.uba.ar/wp-content/uploads/sites/216/2014/03/Regres_log_SALAS-Velasco.pdf.

Sellitto, M. A. (2020). Análisis de políticas de mantenimiento soportada por simulación en una célula de fabricación flexible. *Ingeniare. Revista chilena de ingeniería, vol*. 28 No. 2, 293-303 https://www.scielo.cl/pdf/Ingeniare/v28n2/0718-3305-ingeniare-28-02-293.pdf

Srivastava, A., KUmar, G., & Gupta, P. (2020). Estimating maintenance Budget using Monte Carlo simulation. *Life Cycle Reliability and safety Engineering Vol 9,* 77-89 https://doi.org/10.1007/s41872-020-00110-7.

Sweets, J. (1988). Measuring the Accuracy of Diagnostic System. *SCIENCE Vol 240*, Issue 4857, 1285-1293, DOI: 10.1126/science.3287615

Teng, H., Hagood, M., Yatheepan, Y., Fu, Y., & Li, H. (2016). The delopment of Regression Models to Estimate Routine Maintenance Costs for State Highway Infrastructure. *Journal of Transportation Technology > Vol. 6* No. 5, 339-359 DOI: 10.4236/jtts.2016.65030.

Yongyi, R., Xin, Z., Pengfeng, L., Yoggang, W., &Ruilong, D. (2019). A Survey of Predictive Maintenance: Systems, Purpose and Approaches. *IEEE Communications Surveys & Tutorial, 20,* 1-36. https://doi.org/10.48550/arXiv.1912.07383

Yugapriya, M., Judeson, A., & Jayanthy, S. (2022). Predictive Maintenance of Hydraulic System using Machine Learning Algorithms. *International Conference on Electronics and Renewable Systems (ICEARS). (*pags.1208-1214). Tuticorin, India: IEEE doi: 10.1109/ICEARS53579.2022.9751840

Zabor, E., Reddy, C., Tendulkar, R., & Patil, S. (2022). Logistic Regression in Clinical Studies. *International Journal of Radiation Oncology Biology Physics,* 271-277 https://doi.org/10.1016/j.ijrobp.2021.08.007.

ISSN-2524-2024