NTU JOURNAL OF PURE SCIENCES
EISSN: 2789-1097

RESEARCH ARTICLE
Open Access

Diagnosis of The Most Important Factors Affecting The Annual Performance Evaluation of The Teaching Cadres Using The Technique of Binary -Response Logistic Regression

Dhafer Myasar Jabur
Email: thafer.allela@ntu.edu.iq
Department of Community Health, Al Mosul Technical Institute, Northern Technical University, Mosul, Iraq
Corresponding author: Dhafer Myasar Jabur , Email: thafer.allela@ntu.edu.iq

Received: 14-08-2022, **Accepted:** 26-09-2022, **Published online:** 20-12-2022

Abstract. The aim of the research is to identify the main factors affecting the evaluation of the annual performance of the teaching staff using the Binary Logistic Regression technique by applying to a random sample of size(217) teaching and teaching staff from four different disciplines (technology, administrative, agricultural, The results of the analysis showed that the completion and publication of scientific research was the most important factor affecting the evaluation of the performance of the teaching staff, followed by the teaching use of modern teaching methods and its incorporation of technology into education.

Keywords: logistic regression ، Bernoulli Distribution, maximum likelihood function, Wald test

1-Introduction

The research aims to find a binary logistic regression model by estimating the values of the parameters of each independent variable that will appear in the model, and to test the significance of the estimated model as well as the significance of the transactions of the independent variables.

The sample was drawn from one of the formations of the Ministry of Higher Education and Scientific Research and it included (217) singles, and based on the questionnaire form that was distributed to the teaching staff, the data necessary for the analysis were obtained, as the questionnaire form contained (31) questions, these questions were taken from an evaluation form Annual performance approved by the Ministry of Higher Education and Scientific Research for the academic year 2017-2018.

2-Concept of Logistic Regression

Logistic regression is defined as that type of regression used to predict the values of the dependent variables (qualitative or categorical) depending on a group of mixed independent variables) [4]

It can also be defined as a statistical method for examining the relationship between the qualitative dependent variable and one or more independent variables .[1]

The statistical method used to examine and reconcile the relationship between the qualitative dependent variable and one of the independent variables of any type, in this case, it is called binary logistic regression (Binary Logistic Regression), the relationship between the qualitative two-value dependent variable and several independent variables (quantitative or qualitative) is called a "multiple logistic regression). [3]

The logistic regression model is based on the basic assumption that the dependent variable, that we are interested in studying is a variable that follows the Bernoulli Distribution that takes the value (1) with probability (p) and the value (0) with probability (q = 1- p). [9]

If we have two variables, one of which is double-response (y), the simple linear regression is not appropriate because

$$E(y/x) = p\ (\ y=1) = \acute{p} \qquad (1)$$

The value confined between the two numbers (0,1) and thus the model is not applicable by applying the regression, to solve this problem perform the appropriate mathematical transformation on the dependent variable (y).

The probability value is confined between ($0 \le p \le 1$) and then the ratio ((p / (1-p)) is a positive amount between ($0, \infty$) in other words that ($0 \le p \le 1$) and taking the natural logarithm of the conversion value (p / q) the field of its value is confined between ($-\infty \le$ LOG (p / q) $\le \infty$), and accordingly.

the regression model can be written in the case of one independent variable as follows:

$$Log\ (\ p / q\ ) = \hat{b}0 + \hat{b}\ x \qquad (2)$$

If there is more than one independent variable, the model becomes as follows:

$$Log\ (\ p / q\ ) = b0 + \sum_{j=1}^{k} \hat{b}j\ Xij \qquad i = 1\ ,\ 2\ ,\ldots.,\ n \qquad j = 1\ ,\ 2\ ,\ \ldots.\ ,\ k \qquad (3)$$

The previous equation can be converted to the following form

$$p = \frac{1}{1+\exp\left[-(b0+\sum\sum \hat{b}j\ Xij)\ \right]} \qquad (4)$$

Where (exp) is the inverse of the natural logarithm. [6] [8] [12]

The log (p / q) or Ln (p / q) transformation is called the Logit Transformation, and the logistic function is a continuous function that takes values (0, 1) and approaches (y) to zero whenever the right-hand side of the logistic function approaches ($- \infty$) and (y) approaches one whenever the right-hand side of this function approaches (∞), and that this function is equal to zero when the right-hand side of this function is equal to zero and the ratio (p / q) is called the preference ratio or the success preference (Odds of Success) or the preference ratio of the desired event, that the ratio (p / q) is called the (Odds of Failure) and the amount Log (p / q) is called the logarithm of the (Log Odds Ratio) or log. [10] [15]

The logistic regression model is simply a "logarithmic conversion of linear regression, so it is appropriate to use the characteristics of the Logistic Distribution, which limits the estimated probabilities, making them confined between (zero and one)". [11]

The estimation of the parameters of the log model is done by the method (Maximum Likelihood), maximum likelihood function (M.L) measures the probabilities observed for a number (n) of the independent variables (P1, P2, ..., Pn) that falls in the sample and the product of these probabilities represents the greatest probability function. [2] [3]

3-The importance of logistic regression coefficients (Wald test)

the formula for the test statistic used in the Wald test:

$$W_n = ng(\hat{\theta}_n)^{\top}\left[J_g(\hat{\theta}_n)\hat{V}_n J_g(\hat{\theta}_n)^{\top}\right]^{-1} g(\hat{\theta}_n) \qquad (5)$$

where n is the sample size, and $\hat{\nabla}_n$ is a consistent estimate of the asymptotic covariance matrix of $\hat{\theta}_n$ (Wald) test is used To demonstrate the importance of the coefficients of the logistic regression model, which has a chi-square distribution ($\chi 2$), [5].

4-Quality of fit of the logistic regression model (Hosmer And Lemeshow test)

Hosmer and Lemeshow test is used, as it is based on calculating the chi-square statistic for the difference between the observed values and the expected values (Expected), and was suggested by Hosmer And Lemeshow using a Chi-square distribution to detect the deviations of the logistic model. The count of this test consists of an observed part not based on a theoretical model and the other calculated from the estimates of the logistic model. [13] [14]

The practical side

1- Data collection

For the purpose of obtaining data for the research, a questionnaire form was designed, consisting of two axes, which are as follows:

1- The focus of teaching and knowledge acquisition

2- The center of scientific and research activity, which include four paragraphs:

a - Scientific research published internationally.

b - Scientific research published in the Arab world or locally.

c - International conferences, training courses and symposia.

d - Arab or local conferences, training courses and scientific symposia.

The questionnaire form was distributed to the teaching staff of one of the formations in the Ministry of Higher Education and Scientific Research with (217) forms and no form was lost, and as shown in the table (1) below

Table 1. shows the number of cases covered by the search

		N	Percent
Selected Cases	Included in Analysis	217	100.0
	Missing Cases	0	0.0
	Total	217	100.0
Total		217	100.0

2-Data modeling and model quality testing

Using the statistical analysis program SPSS. version 25, estimate the parameters of the logistic regression model by adopting the greatest probability function (or greatest probability) to reach the best estimate of the coefficients, which is at the lowest value of minus double the logarithm of the greatest possibility function.

In the seventeenth session, we obtained the lowest value of the derivative of minus twice the greatest probability function and stopped at this cycle because the change in the estimated parameters became very slow with slight differences, and it was considered the best result that can be obtained for the parameters.

Table 2. Model parameters, standard error, Wald-statistic, degrees of freedom, and significant parameters

Variables in the Equation							
		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	1.286	.165	60.745	1	.000	3.617
Step 1	2-1	-36.406	2.15	286.632	1	.000	.075
	2-1-3	36.406	2.32	246.246	1	.000	.300
	2-2	72.812	2.41	912.792	1	.000	.142
	2-2-4	36.406	2.11	297.701	1	.000	3.12
	4-1	-36.406	2.27	257.213	1	.268	.233
	2-4-2	72.812	2.31	993.532	1	.000	.142

Table (2) above shows the parameters of the final logistic regression model that were obtained in the seventeenth session, as well as the standard error of each parameter, the number of degrees of freedom, the (Wald) statistic, and the significance of the parameters.

To fully test the efficiency and fit of the model, the highest probability ratio that follows a distribution is used ($\chi 2$ chi – square), where the results appeared in Table (3) :

Table 3. Omnibus Tests of Model Coefficients

Omnibus Tests of Model Coefficients				
		Chi-square	df	Sig.
Step 1	Step	226.790	12	.000
	Block	226.790	12	.000
	Model	226.790	12	.000

The value of ($\chi 2$=226.790) is significant at a level of significance less than 0.001, which confirms the significance of the fully successful model.

To reveal the deviations of the logistic model, Hosmer-Lemeshow proposed an nonparametric test based on calculating the value of the (-2) chi-square statistic for the difference between the observed values and the expected values (Expected), since the chi-square count consists of two parts: scenes not based on a model. [5]

Table 4. Hosmer and Lemeshow test

Hosmer and Lemeshow Test			
Step	Chi-square	df	Sig.
1	.999	6	.000

From Table (4), we note that the value of the statistic = .999 ($\chi 2$) and that sig = .000, we accept the null hypothesis (there are no deviations in the observed and expected values) and this confirms the good fit for the model.

The following table shows the percentage of correct classification.

Table 5. percentages of the classification

Classification Tablea					
	Observed		Predicted		
			Evaluation		Percentage
			0	1	Correct
Step 1	Evaluation	0	47	0	100.0
		1	0	170	100.0
	Overall Percentage				100.0
a. Constant is included in the model.					

And the values of these percentages are that the factors (independent variables) influenced 100% of the failure cases (47 cases) and also 100% affected the success cases (170 cases).

3-Interpretation of the binary logistic regression model

We note from Table (6) that we got estimates for the coefficients of independent variables in Log-odds units, as well as the fixed term of the model, and accordingly the binary logistic regression model can be written as follows:

$$\text{Log}\left(\frac{\hat{\rho}}{1-\hat{\rho}}\right)= 1.286 -36.4X1+36.406X2+72.812X3+36.406X4-36.406X5+72.812X6 \quad (7)$$

-The estimated value of the fixed term in the model was (1.286).

-The independent variables and their importance in the estimated model, the analytical results showed:

-The first two variables (modern teaching methods and demonstration methods are not used, and modern teaching technology X1 (1-2)).

- The fifth (does not supervise graduate studies for undergraduate or postgraduate students X5 (1-4)).

They have the same significance with an estimated regression parameter whose value is (b1, b5 = -36.406), and this means that if the answer of the questioned instructor in his opinion on these two questions was yes (X1,5 = 1), this would lead to a decrease in his annual evaluation score by exp(-36,406)But if the answer of the questioned teacher has no opinion on these two questions (X1,5 = 0), then this will lead to the removal of these two variables from the estimated model.

-The second two variables (the ability to keep pace with the scientific progress of research published in international journals X2 (2-1-3)).

The fourth (your preference for the Arabic language in research writing X4 (2-2-4)),
they have the same significance with an estimated regression parameter whose value is (b2, b4 = + 36.4062), and this means that if the answer of the questioned instructor on these two questions is yes (X2,4 = 1), this will lead to an increase in his annual evaluation score by (36.406), But if the answer of the questioned teacher has no opinion on these two questions (X2,4 = 0), then this will lead to the removal of these two variables from the estimated model.

-The third variable (you have scientific research acceptable for publication in the Arab world or locally X3 (2-2))
And the sixth (you have a participation in scientific research in a local or Arab conference X6 (2-4-2))
They have the same only significance with an estimated regression parameter whose value is (b3, b6 = 72.812), and this means that if the answer of the questioned instructor on these two questions was yes (X3,6 = 1), this would lead to a decrease in his annual evaluation score by (72.812), but if The answer of the teaching respondent, his opinion, to these two questions was "No" (X3,6 = 0). This will lead to the removal of these two variables from the estimated model.

4 -Conclusions

1-The teaching's completion of scientific research and submitting it for publication locally or in the Arab world was one of the most important factors affecting the annual performance evaluation. Equally important is the teaching's participation in a local or Arab scientific conference.

2-On the second level of importance, we found that the teacher's ability to keep pace with global scientific progress and his ability to publish in international scientific journals is an influential factor, which is naturally related to the teacher's ability to carry out scientific research in the English language (the fact that most international scientific journals with a high impact factor publish written research In English).

3-The two factors were the teacher's use of modern teaching methods, his introduction of technology into education, and his supervision of the research of students 'graduation for primary or higher studies, with an adverse effect, as their lack of presence among the teaching could be considered one of the factors that severely affect his annual performance evaluation.

5-Recommendations

1. Using the logistic regression technique to study a multi-response dependent variable (triple, quadruple, or more).
2. Motivating the teaching cadres to pay attention to the factors affecting the annual performance evaluation whose importance appeared in search.
3. To publish the results of the annual evaluation of the teaching cadres and not regard them as confidential and to honor the distinguished teaching staff Motivates the rest of the teaching staff to improve their performance.

6 –References

[1] Anderson, J.A. (1982), Logistic Discrimination, John Wiley, New York, London.
[2] Binary Logistic Regression with PASW/SPSS ,available at : http://core.ecu.edu/psyc/wuenschk/MV/Multreg/Logistic-SPSS.PDF .
[3] Carpov, N. Gama, M. Georgieva, and J. R. Troncoso-Pastoriza, "Privacypreserving semi-parallel logistic regression training with fully homomorphic encryption," BMC Medical Genomics, vol. 13, no. 7, pp. 1–10, 2020.
[4] D.; Cook ,P. and others , (2001) , " Binary Response and Logistic Regression Analysis " , Part of the lowa state University NSE/ILI project Traditional Statistical Methods , www.public.iastate.edu.
[5] Dallal, Gerard E. (2001). Logistic Regression. Available at: www.tufts.edu/~gdallal/logistic.htm
[6] David, W. Hosmer Jr., Stanley Lemeshow, Rodney X.(2013);" Applied Logistic Regression". Sturdivant chaper 9 Logistic Regression models for the analysis of correlated data.
[7] Hosmer,D.W,Lemeshow,S.and Klor,J.(1988):Applied Regression Analysis and other Multibariable methods DWS-KENT publishing company,a division of wadsworth.
[8] J. H. Cheon, D. Kim, Y. Kim, and Y. Song, "Ensemble method for privacypreserving logistic regression based on homomorphic encryption," IEEE Access, vol. 6, pp. 46 938–46 948, 2018.
[9] M. Kim, Y. Song, S. Wang, Y. Xia, and X. Jiang, "Secure logistic regression based on homomorphic encryption: Design and evaluation," JMIR medical informatics, vol. 6, no. 2, p. e19, 2018.
[10] Oluwafadekemi Areo and Obindah Gershon. "Personal Income Tax Compliance in Nigeria: A Generalised Ordered Logistic Regression". In: Research in World Economy 11 (June 2020), p. 261.
[11] Tomasz Rymarczyk et al. "Logistic Regression for Machine Learning in Process Tomography". In: Sensors 19 (Aug. 2019), p. 3400.
[12] Walker, J. (1996) "Methodology Application: Logistic Regression Using the codes Data" Developed For Department of Transportation, National Highway Traffic Safety Administration (NHTS), Washington DC, April 30, 1996.

[13] Wuensch , Karl L. , (2014) ," Binary Logistic Regression with SPSS".
www.care.ecu.edu/psyc/wunesch/MV/Multerg/logistic_SPSS.

[14] Wang, J. and Park, E. Active learning for penalized logistic regression via sequential experimental design. Neurocomputing, 222:183–190, 2017.

[15] Zuhdai , 2017, " parameters estimation of geographically weighted oridinal logistic regression (gwolr) model ", journal of phsics:conference series , pp 2