ODU TIP DERGİSİ/ODU MEDICAL JOURNAL (ODU MED J)

August 2024;11(2):68-77

ARAŞTIRMA MAKALESİ/ RESEARCH ARTICLE

DOI: 10.56941/odutip.1492859

Usability of Ordinal Logistic Regression Analysis for Beck Depression Inventory

Canan Demir[1], Siddik Keskin[2], Hamit Mirtagioglu[3], Yildirim Demir[4]

[1]VanYüzüncü Yıl University, Van Health Services Vocational School, Health Technician Department, Van, Turkiye,
[2]Van Yüzüncü Yıl University, Faculty of Medicine, Department of Biostatistics, Van, Turkiye
[3]Bitlis Eren University, Faculty of Arts and Sciences, Department of Statistics, Bitlis, Turkiye
[4]Van Yüzüncü Yıl University, Faculty of Economics and Administrative Sciences, Department of Econometrics, Van, Turkiye

Received: 31 May 2024, Accepted: 14 August 2024, Published online: 31 August 2024

Abstract

Objective: In the study; For the Beck Depression Scale, the usability of the Ordinal logistic regression model was evaluated, taking into account the situation in which this scale was ordinal.

Materials and Methods: The study was conducted in the first and second-year students of the University of Van Yüzüncü Yıl, Health Services Vocational School. A total of 664 volunteer students who accepted to participate in the study were included in the study, and no sample was taken from the population. The students were asked 8 questions including socio-demographic characteristics, as well as questions of the 21-item Beck Depression Scale. Depression status with 4 order categories (Normal-Mild-Moderate-Severe) was taken as the dependent variable in the study, and the relationship of other socio-demographic variables with the depression status variable was examined.

Results: In the study, Nagelkerke pseudo R^2 value, one of the goodness of fit criteria, was found to be 0.062. In addition, the model fit criterion -2LL (Log-likelihood) test statistic ($p<0.05$) was found significant.

Conclusion: As a result, the usability of ordered logistic regression analysis in determining the relationships between the variables and depression was emphasized.

Keywords: Beck depression inventory, Statistical analysis, Ordinal logistic regression analysis

Beck Depresyon Ölçeği İçin Sıralı Lojistik Regresyon Analizinin Kullanılabilirliği

Özet

Amaç: Çalışmada; Beck Depresyon Ölçeği için, bu ölçeğin ordinal olduğu durum dikkate alınarak Ordinal lojistik regresyon modelinin kullanılabilirliği değerlendirilmiştir.

Materyal ve Metot: Çalışma, Van Yüzüncü Yıl Üniversitesi Sağlık Hizmetleri Meslek Yüksekokulu birinci ve ikinci sınıf öğrencileri ile gerçekleştirilmiştir. Çalışmaya katılmayı kabul eden toplam 664 gönüllü öğrenci çalışmaya dahil edilmiş, evrenden örneklem alınmamıştır. Öğrencilere sosyo-demografik özellikleri içeren 8 sorunun yanı sıra 21 maddelik Beck Depresyon Ölçeği soruları sorulmuştur. Çalışmada bağımlı değişken olarak 4 sıra kategorili (Normal-Hafif-Orta-Şiddetli) depresyon durumu alınmış ve diğer sosyo-demografik değişkenlerin depresyon durumu değişkeni ile ilişkisi incelenmiştir.

Bulgular: Çalışmada uyum iyiliği ölçütlerinden Nagelkerke yalancı R^2 değeri 0.062 olarak bulunmuştur. Ayrıca model uyum kriteri -2LL (Log-likelihood) test istatistiği ($p<0.05$) anlamlı bulunmuştur.

Sonuç: Sonuç olarak, değişkenler ile depresyon arasındaki ilişkilerin belirlenmesinde sıralı lojistik regresyon analizinin kullanılabilirliği vurgulanmıştır.

Anahtar Kelimeler: Beck depresyon ölçeği, İstatistiksel analiz, Sıralı lojistik regresyon analizi

Suggested Citation: Demir C, Keskin S, Mirtagioglu H, Demir Y. Usability of Ordinal Logistic Regression Analysis for Beck Depression Inventory. ODU Med J, 2024;11(2): 68-77.

Address for correspondence/reprints:

Yildirim Demir

Telephone number: +90 (505) 452 33 29

E-mail: ydemir@yyu.edu.tr

INTRODUCTION

Depression is defined as a mental disorder affecting more than 264 million individuals worldwide (1). Depression is based on unwillingness and inability to enjoy daily activities, which they previously did willingly and fondly (2). In addition, there are negative self-concept symptoms such as fatigue, sleep and appetite disorders (3).

Depression can occur at any age and can affect all kinds of people, young, old, rich and poor (1). Current studies show that depression is caused by the interaction of genetic, biological, environmental and psychological factors (4). Life events such as childhood troubles, loss and unemployment contribute to the development of depression. The effects of depression can be reflected in the daily life and interpersonal relationships of the person and cause his/her performance in school and/or work life to decrease (5).

Depression is predicted to be a globally widespread disease that has a great burden on the population. The 21-item Beck Depression Inventory (BDI) was first proposed by Beck et al. and has been used in more than 7,000 studies to date. This scale is one of the most popular tools used worldwide to assess depressive symptoms (6).

Various statistical methods are used to evaluate the depression scale. However, depression scales tend to form asymmetric distributions. Therefore, commonly used traditional statistical methods such as t-test and linear regression may not be suitable. Therefore, in the study, For the Beck Depression Scale, the usability of the Ordinal logistic regression model was evaluated, taking into account the situation in which this scale was ordinal.

METHODS

After the approval of the Ethics Committee; the study was conducted in the first and second-year students of the University of Van Yüzüncü Yıl, Health Services Vocational School in the spring semester of the 2014-2015 academic year. A total of 664 volunteer students who accepted to participate in the study were included in the study, and no sample was taken from the population. The students were asked 8 questions (age, class, department, gender, marital status, number of siblings, residence and income status) including socio-demographic characteristics, as well as questions of the 21-item Beck Depression Scale (BDI). The categorical variables discussed in the study and the descriptive statistics (number and percentage) of these variables are summarized in Table 1.

Table 1. Variables considered in the study and descriptive statistics

Variable	Category (n=664)		%
Gender	Female	377	56.8
	Male	287	43.2
Class	1	441	66.4
	2	223	33.6
Marital status	Single	604	91.0
	Married	60	9.0
Place of residence	Dorm	239	36.0
	At home with friends	84	12.7
	At home with family	249	37.5
	At home with relatives	22	3.3
	Alone at home	70	10.5
Number of siblings	<3	124	18.7
	3-5	308	46.4
	6-10	193	29.1
	>10	39	5.8
Income	<1000	267	40.2
	1000-2000	215	32.4
	2000-3000	106	16.0
	>3000	76	11.4
Department	Anesthesia	69	10.4
	Dialysis	20	3.0
	First and Emergency Aid (Normal)	117	17.6
	First and Emergency Aid (Night)	115	17.3
	Perfusion	23	3.5
	Radiotherapy	26	3.9
	Medical Documentation and Secret	112	16.9
	Medical Laboratory	62	9.3
	Medical Imaging	41	6.2
	Elderly care	79	11.9

Logistic regression analysis shows that when the dependent variable is binomial (successful-unsuccessful, yes-absent, patient-healthy) or multinomial (Married-widowed-single, low-medium-high), it is an analysis method used to determine the relationship between the dependent variable (response variable) and explanatory (independent) variables (7,8).

In logistic regression analysis; instead of the value of the response variable, the probability of one of the values that this variable can take is estimated. Therefore, logistic regression analysis is mathematically based on probability, logarithm of odds and odds.

Odds: it is the ratio of the probability of occurrence (p) to the probability of absence (q).

Odds Ratio (OR): OR, which is a summary measure of the relationship between two variables, is the ratio of two probabilities to each other.

Ordinal Logistic Regression: The dependent variable Y is categorised as 0, 1 or 2. Assuming that the ordinal dependent variable Y categorised as 0, 1, 2, ..., K can take K +1 values, it is necessary to decide which category to use as the reference value. An extension of this is to take $Y = 0$ as the reference or baseline outcome and construct logit functions that compare other categories with it (9). The two logit functions are expressed by two models given in Equations (1) and (2).

$$g_1(x) = ln\left[\frac{P_r(Y=1|x)}{P_r(Y=0|x)}\right]$$

$$= \beta_{10} + \beta_{11}x_1 + \beta_{12}x_2 + \cdots + \beta_{1p}x_p$$

$$= x'^{\beta_1} \quad (1)$$

$$g_2(x) = ln\left[\frac{P_r(Y=2|x)}{P_r(Y=0|x)}\right]$$

$$= \beta_{20} + \beta_{21}x_1 + \beta_{22}x_2 + \cdots + \beta_{2p}x_p$$

$$= x'^{\beta_2} \quad (2)$$

Given the covariate vector, the conditional probabilities of each outcome category are as follows

$$P_r(Y = 0|x) = \frac{1}{1 + e^{g_1(x)} + e^{g_2(x)}} \quad (3)$$

$$P_r(Y = 1|x) = \frac{e^{g_1(x)}}{1 + e^{g_1(x)} + e^{g_2(x)}} \quad (4)$$

$$P_r(Y = 2|x) = \frac{e^{g_2(x)}}{1 + e^{g_1(x)} + e^{g_2(x)}} \quad (5)$$

A general expression for conditional probability in the categorical model is given by Equation (6).

$$\pi_j(x) = P_r(Y = j|x) = \frac{e^{g_j(x)}}{\sum_{k=0}^{2} e^{g_k(x)}} \quad (6)$$

Where both vector β_0 and $g_0(x)$ are equal to 0.

We denote a general expression for the probability that the outcome is equal to k conditional on a vector, x, of p covariates as $Pr[Y = k|x] = \emptyset_k(x)$. Logits are

$$g_k(x) = ln\left[\frac{\pi_k(x)}{\pi_0(x)}\right] = \beta_{k0} + x'\beta_k \quad (7)$$

for $k = 1, 2, \ldots, \mathrm{K}$

The basic procedure consists of the following steps:

Expressions defining model-specific logits are used to construct an equation describing $\emptyset_k(x)$ as a function of the unknown parameters. The values of K +1 dimensional multinomial result values, $z' = (z_0, z_1, \ldots, z_k)$, are occurred the ordinal output as $z_k = 1$ if $y = k$ and $z_k = 0$ otherwise. It follows that only one value of z is equal to 1. The general form of the likelihood for a sample of n independent observations, (y_i, x_i), $i = 1, 2, \ldots, n$ is

$$l(\beta) = \prod_{i=1}^{n} [\emptyset_0(x_i)^{z_{0i}} \emptyset_1(x_i)^{z_{1i}} \times \ldots \times \emptyset_K(x_i)^{z_{Ki}}] (8)$$

Where we use "β" somewhat imprecisely to denote both the p slope coefficients and the K model-specific intercept coefficients. It follows that the log-likelihood function is

$$L(\beta) = \prod_{i=1}^{n} z_{0i} ln[\emptyset_0(x_i)] + z_{1i}[\emptyset_1(x_i)] + \cdots + z_{Ki} ln[\emptyset_K(x_i)] \quad (9)$$

An estimator of the covariance matrix of the estimated coefficients can be obtained in the usual way by taking the inverse of the negative of the matrix of second-order partial derivatives with respect to $\hat{\beta}$.

The log-likelihood function is used to compare predicted and observed values. Using the likelihood functions, estimated values can be expressed as follows.

$$D = 2\, ln \frac{Likelihood\ of\ current\ model}{Likelihood\ of\ saturated\ model} \quad (10)$$

The D (Likelihood Ratio or Deviance) statistic, which corresponds to the sum of squares error in linear regression, plays an important role in judging the goodness of fit (9, 10).

Beck Depression Scale (BDS): It is a 21-item self-assessment scale that was developed by Aron T. Beck et al. in 1961 and measures the risk of depression and symptoms of depression in

adults (11). The Turkish reliability and validity study of the scale was conducted by Hisli, and it was stated that it could be used to measure depression symptoms in university students (12).

Scoring and evaluation of the Beck depression scale: Beck depression scale is an easy test to score and evaluate. First of all, all the numbers marked in the sentence groups consisting of four items are added together and the total score obtained is determined. Then this score is found in the rubric below.

- 0-9: Shows normal level
- 10-18: Shows signs of mild depression
- 19-29: Indicates moderate depression
- 30-63: Shows signs of severe depression.

Depression status with 4 order categories (Normal-Mild-Moderate-Severe) was taken as the dependent variable in the study, and relationship of other socio-demographic variables with the depression status variable was examined.

RESULTS

Ordered Logistic Regression Analysis Results

In the study, the last categories of categorical variables included in the model were taken as reference categories.

Model fit criteria for sequential logistic regression analysis are given in Table 2. As seen in Table 2; -2LL (Log-likelihood) test statistic was found to be significant ($p<0.05$).

Table 2. Model fitting analysis for ordinal logistic regression

Model	Chi-Square	-2 Log Likelihood	df	p
Intercept Only		1425.130		
Final	36.618	1388.512	24	0.048
False R^2 values				
McFadden = 0.022, Nagelkerke = 0.062, Cox and Snell = 0.057				

The Nagelkerke false R^2 value, one of the measures of goodness of fit, was found to be 0.062 in the study. In R^2 statistics; the log-likelihood value of the model, which includes only the model constant term, is accepted as the general sum of squares, and the likelihood value for the entire model is accepted as the sum of error squares. Thus, the probability ratio expresses the success of the complete model compared to the model containing only the fixed term. The likelihood is between 0 and 1. Thus the logarithm of the likelihood is less than or equal to zero. If the model contains a very low likelihood ratio value, the logarithm of likelihood will be large. Thus, the small value of the logarithm of the likelihood indicates that the complete model is better than the model containing only the constant term (13).

When Table 3 was examined, the effect of the variables of department, class, age, marital status, number of siblings, place of residence and income on the level of depression was not found statistically significant, while the effect of the gender factor was significant ($p <0.05$). According to this, assuming that other variables are kept constant, it can be said that female

students are 1,559 times more likely to be depressed than male students. However, although it was not found statistically significant; Students studying in the primary and emergency aid (evening education) department were approximately 1.8 times more likely to be depressed than those studying in the dialysis department, and students studying in the dialysis department were approximately 1/0,178 = 5.6 times more likely to be depressed than those studying in the medical imaging department. Similarly, students under 25 tend to be 1.2 times more likely to be depressed than students over 25, and students staying home with friends tend to be 2.06 times more likely to be depressed than other students. When the tendency to depression is examined by income level; it can be said that students with an income of less than 1000 Lira tend to be 1.4 times more depressed than students with an income of over 3000 Lira.

Table 3. Ordinal logistic regression analysis summary results

Variable	Category	Exp(β)	P	95% Confidence Interval	
				Lower B.	Upper B.
Department	First and Emergency Aid (Night)	1.824	0.190	0.741	3.891
	Medical Documentation and Secret	1.061	0.901	0.419	4.054
	Medical Imaging	0.528	0.236	0.183	4.933
	Anesthesia	1.182	0.729	0.459	4.17
	Elderly care	1.094	0.853	0.42	4.233
	Medical Laboratory	1.292	0.608	0.484	4.396
	Perfusion	1.110	0.859	0.356	5.533
	Radiotherapy	1.315	0.637	0.421	5.566
	First and Emergency Aid (Normal)	1.071	0.880	0.435	3.902
	Dialysis				
Class	1	1.048	0.772	0.761	1.621
	2				
Age	17-25	1.201	0.612	0.593	2.899
	>25				
Gender	Female	1.559	0.006	1.136	1.611
	Male				
Marital status	Married	0.899	0.767	0.445	2.888
	Single				
Number of siblings	<3	0.552	0.165	0.238	3.555
	3-5	0.758	0.494	0.341	3.327
	6-9	0.659	0.311	0.294	3.378
	>10				
Place of residence	Dorm	1.504	0.147	0.866	2.298
	At home with friends	1.545	0.364	0.604	4.127
	At home with family	1.171	0.636	0.609	2.679
	At home with relatives	2.061	0.136	0.796	4.204
	Alone at home	1.520	0.151	0.858	2.371
	Other				
Income	<1000	1.451	0.209	0.81	2.415
	1000-2000	1.183	0.571	0.661	2.409
	2000-3000	1.267	0.463	0.672	2.603
	>3000				

DISCUSSION

In the study, descriptive information about ordinal logistic regression method was given and the usability of the method in determining the relationship between depression and some demographic characteristics in students was evaluated. The likelihood ratio statistic, which shows the overall significance of the model with the variables included in the model, was found to be significant ($p < 0.05$). Among the variables, the relationship between the gender factor and the level of depression was found to be significant. Also, although it was not found statistically significant; department, age, place of residence, and income level variables were observed to be potential risk factors for depression level.

In a study by (14); The relationship between the socio-demographic characteristics of individuals and their happiness levels was analyzed with sequential logistic regression analysis and it was emphasized that happiness levels may differ in terms of socio-demographic characteristics by years, but still basically similar results. In a study by (15), ordered logistic regression analysis was used to determine the factors affecting students' happiness. The goodness of fit of the model was evaluated with Cox and Snell, Nagelkerke, McFadden tests and it was emphasized that the factors affecting the happiness of the students were “age”, “satisfaction with the department”, “peace in the family” and “the opinion that the city of Sivas is suitable for students”. In a study by (16), ordered logistic regression analysis was preferred to determine the reasons for student absenteeism. It was emphasized that GPA and weekly course hours are significantly associated with absenteeism tendency. On the other hand, used the sequential logistic regression model to determine the factors that may have an impact on the overall satisfaction level of students studying at Atatürk University (17).

Another method used in many different fields such as binary and multinomial logistic regression analyses is the ordinal logistic regression method. The method is also known as the proportional probability model because it includes log probabilities and transformations used during prediction. The use of the ordinal logistic regression model relies on the presence of large dependent variables and the fulfillment of model assumptions (18, 19). On the other hand, ordered logistic regression analysis can be considered as an extension of the simple (binary) logistic regression model. In simple logistic regression, the logarithm of odds for the occurrence or occurrence of the event in the variable taken as the response variable is modeled as a linear combination of independent or explanatory variables. However, this modeling approach ignores the order of the categorically dependent variable. The ordered logistic regression model overcomes this problem by

using the cumulative categories of the dependent variable in calculating the logarithm of odds. Thus, interpretation of the coefficients calculated for ordinal logistic regression differs from multinomial and binary logistic regression (20).

Ordinal logistic regression includes the assumption of proportional odds, as well as the assumption that the categories of the response variable are ordered and that there is no multiple correlations between the explanatory variables. The proportional odds assumption expresses that the effect of explanatory variables is the same or equal in computing the logarithm of each odds. In this case, a single model will be sufficient for the coefficients. If this assumption is not satisfied, different models are needed to describe the relationship between each pair of outcomes. However, this assumption does not apply to the constant term, as the constant term takes different values for each equation.

CONCLUSION

As a result, as in depression levels, the response variable may be ordered in many areas, and the relationships between these variable and explanatory variables can be linear or nonlinear. In such cases, ordinal logistic regression analysis can be used as an appropriate approach to determine the relationship between variables.

Ethics Committee Approval: Ethic committee approval report was obtained from Yüzüncü Yıl University, Faculty of Medicine, Non-Interventional Clinical Research Ethics Committee (Date and number: 10.03.2015 / 21)

Peer-review: Externally peer-reviewed.

Author Contributions: Idea- Equal; Design - Equal; Control - Equal; Data Collection and/or Processing - Equal; Analysis and/or Interpretation - Equal; Literature Review - Equal; Spelling - Equal; Critical Review – Equal

Conflict of Interest: The authors declared no conflict of interest.

Financial Disclosure: The authors declared that this study has not received no financial support.

REFERENCES

1. Anonymous 1. World Health Organization. Depression fact sheets. https://www.who.int/health-topics/depression#tab=tab_1. Accessed 05 April 2020.
2. Bedaso A, Kediro G, Yeneabat T. Factors associated with depression among prisoners in southern Ethiopia: a cross-sectional study. BMC Research Notes 2018; 11: 637-6.
3. Palmer GA, Happe MC, Paxson JM, Jurek BK, Graca JJ, Olson SA. Psychometric properties of the Beck Depression Inventory-II for OEF/OIF veterans in a polytrauma sample. Military Medicine 2014; 179: 879-884.
4. Anonymous 2. National Institute of Mental Health, Depression, Transforming the

understanding and treatment of mental illness. https://www.nimh.nih.gov/health/topics/depression/index.shtml#part_145397. Accessed 02 April 2020.

5. Taycan O, Kutlu L, Çimen S, Aydın N. Relation between sociodemographic characteristics depression and burnout levels of nurse working in university hospital. Anatolian Journal of Psychiatry 2006; 7: 100-108.
6. Wang YP, Gorenstein C. Psychometric properties of the Beck Depression Inventory-II: a comprehensive review. Brazilian Journal of Psychiatry 2013; 35: 416-431.
7. Gok AC, Özdemir A. Forecasting the sector portions of banks by logistic regression analysis. Dokuz Eylül University Journal of Faculty of Business 2011; 1: 43-51.
8. Alpar R. Applied multivariate statistical methods. Ankara, Delta Yayıncılık; 2013.
9. Liu X. Applied Ordinal Logistic Regression Using Stata: From Single-Level to Multilevel Modeling. London, Sage Publications; 2016.
10. Hosmer DW, Lemeshow S, Sturdivant RX. Applied Logistic Regression. New York, John Wiley & Sons; 2013.
11. Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. Archives of General Psychiatry 1961; 4: 561-571.
12. Hisli N. The reliability and validity of Beck depression inventory for university students. Turkish Journal of Psychology. 1989; 7:3-13.
13. Akın HB, Şentürk E. Analysing levels of happiness of individuals with ordinal logistic analysis. Oneri 2012; 10: 183-193.
14. Keskin S. Overview of pseudo r-square measures. Güler A (Editor) Current Academic Studies in Health Sciences, Cetinje, IVPE, 2018. p.1-9.
15. Goktolga ZG, Karakıs E. Determination of the Relationship between Students' Economic Status and Their Happiness by Ordinal Logistic Regression Method: The Case of Cumhuriyet University. EconWorld2017@Paris Proceedings, Paris, France; 2017, July 25-27.
16. Ugurlu CT, Usta HG, Şimşek AS. Absence trend of undergraduates: logistic prediction. Kastamonu Education Journal 2018; 26: 345-349.
17. Orcanlı K, Oktay E. Pirim M. Investigation of Factors Affecting Satisfaction with Time-Use: Example of Atatürk University Students. Journal of Graduate School of Social Sciences 2019; 23: 797-825.
18. Eygu H, Gulluce A. Determination of customer satisfaction in conservative concept hotels by ordinal logistic regression analysis. Journal of Financial Risk Management 2017; 6: 269-284.
19. Eygu H, Kılınç A. The research of factors that affects on life satisfaction: the sample of

Kayseri province. OPUS International Journal of Society Researches 2020; 16: 3591-3618.

20. Turgut V. Investigation of sequential logistic regression model. Van Yuzuncu Yil University, Institute of Health Sciences 2020.