

Analysis of the Invariance and Generalizability of Multiple Linear Regression Model Results Obtained from Maslach Burnout Scale through Jackknife Method

Tolga Zaman, Kamil Alakus

Department of Statistics, Faculty of Science and Arts, Ondokuz Mayıs University, Samsun, Turkey

Email: tolga.zaman@omu.edu.tr

Received 19 October 2015; accepted 8 December 2015; published 11 December 2015

Copyright © 2015 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

The purpose of this study was to examine the burnout levels of research assistants in Ondokuz Mayıs University and to examine the results of multiple linear regression model based on the results obtained from Maslach Burnout Scale with Jackknife Method in terms of validity and generalizability. To do this, a questionnaire was given to 11 research assistants working at Ondokuz Mayıs University and the burnout scores of this questionnaire were taken as the dependent variable of the multiple linear regression model. The variable of burnout was explained with the variables of age, weekly hours of classes taught, monthly average credit card debt, numbers of published articles and reports, gender, marital status, number of children and the departments of the research assistants. Dummy variables were assigned to the variables of gender, marital status, number of children and the departments of the research assistants and thus, they were made quantitative. The significance of the model as a result of multiple linear regressions was examined through backward elimination method. After this, for the five explanatory variables which influenced the variable of burnout, standardized model coefficients and coefficients of determination, and 95% confidence intervals of these values were estimated through Jackknife Method and the generalizability of the parameter estimation results of these variables on population was researched.

Keywords

Jackknife Method, Invariance, Generalizability, Maslach Burnout Scale, Multiple Linear Regression, Backward Elimination Method

1. Introduction

Burnout is an important health problem for occupations which include personal relationships with people. Professional people living in different areas can give various reactions to stresses experienced at work. These reactions are: weariness related with work, stress and burnout [1]. Weariness related with work and stress can be seen in every occupation. However, burnout is an important health problem for occupations which include personal relationships with people because the most difficult thing is dealing with people. In this context, a person develops burnout syndrome. The symptoms of burnout syndrome, which is common in people working in the fields of health and education, can be grouped in three as emotional, physical and mental symptoms [2].

Burnout was defined by Maslach, Freudenberger and Pines at the beginning of 1970s and it had become a frequently studied issue recently. According to Freudenberger, emotional burnout means workers' not being able to carry out the necessities of their work as a result of being overworked [3]. Later, Maslach and Jackson developed this subject, examined the most accepted model of burnout and defined burnout as emotional exhaustion, depersonalization and a diminished sense of personal accomplishment [4].

In relation to burnout, an answer can be sought to the following question especially for people working in close contact with other people in health and education sectors: Is there an association between the workers' levels of burnout and job satisfaction and their personal characteristics such as age, gender, marital status, numbers of children, etc? One of the scales commonly used for the assessment of such situations was the Maslach Burnout Scale developed by Maslach and Jackson in 1981. This scale has a total of 22 items. 9 of these items are related to emotional exhaustion, while 5 are related to depersonalization and 8 are related to diminished sense of personal accomplishment. The items are graded in 7 different ways: 0 = never, ..., 6 = every day [5].

Scientific methods lead researchers to refute or prove a hypothesis through results based on experiment. Especially the studies with relatively little numbers of samples or subjects can be said to contain basic judgments based on the results of a special sample since it will be thought that they lack repetition even if they are designed very well or conducted meticulously. If the same values cannot be obtained from similar other studies, the findings of the research won't be accepted as the values of the general population. Thus, that sample will have a limited value [6].

The method applied to the data set in our study is Jackknife Method. The study aims to introduce the method and to show its applicability in case of need.

Generalizability of the results of a scientific study is very important as one of the scientific features [7]. This is because the results of especially experimental studies will shed light on similar future studies and contribute to the development of the related field. In addition, the objectivity and the reliability of the results will strengthen the organizational dimension of the field and it will protect the practicality, objectivity and validity of the results in the long run [8]. The results which cannot be generalized are specific to the sample and they cannot be reflected on the universe [9].

2. Jackknife Method and Theory

Jackknife is the first computer-based method for the estimation of bias and standard deviation. Jackknife resampling method was used by Maurice Quenouille in 1956 to remove the statistical bias. Later, it was expanded by John Tukey in 1958 in order to create hypothesis tests and confidence intervals and the name "Jackknife" was adapted [10]. Its expansion into linear regression was made by Miller (1964, 1974), Gray and Schucany (1972), Hinkley (1977), Reeds (1978) and Porr (1983, 1985).

Sampling method is the process of choosing sample from the population. Thus, it is one of the essential stages of a scientific research. The most important feature of a sample is providing the smallest sampling error in order to reach valid and consistent estimations about the population. However, parametric estimation methods do not give reliable results when the number of samples is small and also they can disrupt the hypotheses of the parametric methods.

Recently, resampling methods have begun to be used when the data cannot be assessed with parametric methods. One of these is Jackknife Method. Jackknife Method has been developed to minimize the sampling error by obtaining small confidence intervals in the estimation of population parameters. This method is also called Pocket Knife method. Pocket Knife can be used for various things. This method, also, is a method that can be used for various practices. Thus, the similarity is very appropriate. In addition, one of the things that can be done with a pocket knife is cutting. A procedure similar to cutting is done in Jackknife Method, too.

The basic logic of the method depends on omitting each estimation value of the data set once and calculating the sample statistics from the remaining observations. This way, only n number of different samples can be formed from n numbers of observations.

Let's have the sample $X = (x_1, x_2, \dots, x_n)$ and $\hat{\theta} = s(X)$ as the estimator. According to Jackknife Method, the sample when the i . observation is omitted is

$$x_{(i)} = (x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_n); \quad i = 1, 2, \dots, n$$

Thus, the estimator is

$$\hat{\theta}_{(i)} = s(x_{(i)})$$

Jackknife estimation of the bias is as follows:

$$\widehat{Bias}_{Jackk} = (n-1)(\hat{\theta}_{(.)} - \hat{\theta})$$

Here, $\hat{\theta}_{(.)}$ is the Jackknife estimation of θ and it is calculated through the equation $\hat{\theta}_{(.)} = \frac{\sum_{i=1}^n \hat{\theta}_{(i)}}{n}$. Jackknife estimation of standard error is obtained by using the equation:

$$\widehat{se}_{Jackk} = \frac{\left\{ \frac{\left[\sum_{i=1}^n (\hat{\theta}_{(i)} - \hat{\theta}_{(.)})^2 \right]}{n-1} \right\}^{1/2}}{\sqrt{n}}$$

Pseudo values are calculated with the following equation in Jackknife estimation [11]-[13]:

$$\text{Pseudo value}_i = n\hat{\theta} - (n-1)\hat{\theta}_i$$

3. Practice

This study examined the burnout levels of research assistants of Ondokuz Mayıs University based on Jackknife Method by using Maslach Burnout Scale. A questionnaire was given to 11 research assistants working at Ondokuz Mayıs University for this purpose. The number of samples was specifically kept small. This is because Jackknife Method is a statistical method which tests the generalizability of the parameter values obtained from statistics calculated from small numbers of samples to the general population. Of the 22 items in the Maslach Burnout Scale, 9 of these items are related to emotional exhaustion, while 5 are related to depersonalization and 8 are related to diminished sense of personal accomplishment. These items are graded in 7 different ways as 0 = never, ..., 6 = every day [5]. The burnout score which is found by adding up the responses given to these items is expressed as our dependent variable in the regression model. The independent variables with which we tried to explain this dependent variable are Age (X_1), Weekly hours of classes taught (X_2), Monthly average credit card statement (X_3), Numbers of published articles and reports (X_4), Gender (X_5), Marital status (X_6), Number of children (X_7), Department (X_8). The values of these variables were obtained from the questionnaire. Dummy variables were assigned to the variables of gender, marital status, number of children and department and they were made quantitative: Gender: 1 (male), 0 (female); Marital status: 1 (married), 0 (single); Number of children: 1 (yes), 0 (none); Department: 1 (statistic), 0 (other). Regression model was constructed as linearly because there are linear relationships between independent variables and response variable. Thus, the regression model built was expressed as:

$$Y_i : \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7 + \beta_8 X_8 + \varepsilon_j$$

The variables obtained as a result of the questionnaire are given in **Table 1**.

The multiple linear regression analyses of this data were analyzed by using SPSS 20.0 package program and the independent variables significant for the model were chosen through backward elimination method. The results are summarized in **Table 2**.

Table 1. Data set ($n = 11, p = 8$).

Number	Y (Burnout)	X ₁ (Age)	X ₂ (Weekly hours of classes)	X ₃ (Average credit card statement)	X ₄ (Number of published articles)	X ₅ (Gender)	X ₆ (Marital status)	X ₇ (Number of children)	X ₈ (Department)
1	71	26	3	317	2	1	0	0	0
2	31	32	5	600	4	1	0	0	0
3	27	27	6	500	2	0	0	0	1
4	20	27	9	1500	1	1	0	0	1
5	27	28	10	2000	13	1	1	0	1
6	43	26	10	750	8	1	0	0	1
7	45	26	6	1200	6	1	0	0	1
8	45	29	5	2500	10	1	1	1	1
9	52	27	7	2000	15	1	1	1	1
10	47	36	7	750	2	0	1	1	1
11	28	33	14	350	6	1	1	1	0

Table 2. Multiple linear regression results.

Variable	Non-standardized coefficients		Standardized coefficients		t-test	Significance probability
	B	Std. Error	Beta			
Fixed	43.632	7.141			6.110	0.002
X ₂	-7.134	0.907	-1.489		-7.865	0.001
X ₃	-0.044	0.008	-2.259		-5.794	0.002
X ₅	56.505	9.866	1.550		5.727	0.002
X ₆	42.111	6.933	1.491		6.074	0.002
X ₈	46.099	9.587	1.460		4.809	0.005

The Model's Coefficient of Determination value: 0.927; Significance test of the model is $F = 12.779$; its significance is 0.007.

Table 2 gives the results of explanatory variables significant for the regression model based on the backward elimination method. When **Table 2** is examined, it can be seen that explanatory variables of X_2, X_3, X_5, X_6 and X_8 are statistically significant for the model ($p < 0.05$). According to t-tests statistics results and significance values, selected variables have a significant effect on response variable for 95% confidence level. The multiple linear regression model formed with the help of these variables was also found to be statistically significant ($p = 0.007 < 0.05$). Determination coefficient was calculated as $R^2 = 0.927$ for the model. This ratio is pretty high and it shows the explanation capability of independent variables in the model which were statistically significant (X_2, X_3, X_5, X_6, X_8).

Table 3 gives the standardized beta coefficients which are obtained as a result of the regression analysis conducted by leaving out one observation each time with Jackknife Method.

By using the standardized beta estimation values in **Table 3**, pseudo values of the explanatory variables X_2, X_3, X_5, X_6 and X_8 in **Table 4** were obtained

In terms of absolute value, $|t_{\text{calculation}}| > 2.228$ shows the reliability and the invariance of the estimated regression parameters. This expression shows that parameter estimation values of the explanatory variables X_2, X_3, X_5, X_6 show determination with original values. This means that the beta coefficients obtained with Jackknife Method do not belong to this sample only, but they are estimations which can also be generalized to the population.

When **Table 5** is analyzed, it is understood from the expression $|t_{\text{calculation}}| = |-3.135| > 2.228$ that the average weekly hours of classes taught estimation value $\hat{\beta}_{2\text{jackk}} = -1.439$ obtained through Jackknife Method confirms the first original value $\hat{\beta}_{2\text{orj}} = -1.489$ calculated and it was seen that the Jackknife parameter value was within the 95% confidence interval. From this result, we can say that the variable of weekly hours of classes taught which is an explanatory variable in the model is not valid only for this sample, but it can also be generalized for

Table 3. Beta coefficients obtained with Jackknife Method.

Omitted observation	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_5$	$\hat{\beta}_6$	$\hat{\beta}_8$	R^2	$\hat{\sigma}$
None omitted	-1.489	-2.259	1.550	1.491	1.460	0.927	5.618
Omit the 1st observation	-1.499	-2.525	1.831	1.790	1.665	0.949	3.722
Omit the 2nd observation	-1.415	-1.924	1.377	1.249	0.954	0.974	3.673
Omit the 3rd observation	-1.591	-2.395	1.330	1.604	1.583	0.927	6.022
Omit the 4th observation	-1.780	2.752	1.841	1.803	1.731	0.927	5.662
Omit the 5th observation	-1.489	-2.139	1.585	1.455	1.477	0.925	6.129
Omit the 6th observation	-1.430	-2.158	1.483	1.416	1.369	0.928	6.232
Omit the 7th observation	-1.568	-2.471	1.707	1.487	1.636	0.951	5.126
Omit the 8th observation	-1.433	-1.893	1.568	1.410	1.482	0.930	6.122
Omit the 9th observation	-1.530	-2.156	1.571	1.429	1.467	0.923	6.229
Omit the 10th observation	-1.564	-2.389	1.161	1.528	1.539	0.931	6.022
Omit the 11th observation	-1.155	-2.192	1.599	1.421	1.359	0.925	6.157

Table 4. Pseudo values.

Omitted observation	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_5$	$\hat{\beta}_6$	$\hat{\beta}_8$	R^2	$\hat{\sigma}$
None omitted	-1.489	-2.259	1.550	1.491	1.460	0.927	5.618
Omit the 1st observation	-1.389	0.401	-1.260	-1.499	-0.59	0.707	24.578
Omit the 2nd observation	-2.229	-5.509	3.280	3.911	6.520	0.457	25.068
Omit the 3rd observation	-0.469	-0.899	3.750	0.361	0.230	0.927	1.578
Omit the 4th observation	1.421	2.671	-1.360	-1.629	-1.25	0.927	5.178
Omit the 5th observation	-1.689	-3.459	1.20	1.851	1.290	0.947	0.508
Omit the 6th observation	-2.079	-3.269	2.220	2.241	2.370	0.917	-0.522
Omit the 7th observation	-0.699	-0.139	-0.020	1.531	-0.30	0.687	10.538
Omit the 8th observation	-2.049	-5.919	1.370	2.301	1.240	0.897	0.678
Omit the 9th observation	-1.079	-3.289	1.340	2.111	1.39	0.967	-0.492
Omit the 10th observation	-0.739	-0.959	5.440	1.121	0.670	0.887	1.578
Omit the 11th observation	-4.829	-2.929	1.060	2.191	2.470	0.947	0.228
Average pseudo values obtained by Jackknife tech.	-1.439	-2.127	1.547	1.317	1.276	0.842	6.255
Average pseudo standard error	0.459	0.785	0.623	0.503	0.632	0.048	2.931
Calculated t value	-3.135*	-2.710*	2.483*	2.618*	2.019	17.54*	2.134
Table value, $df = 10$;	2.228	2.228	2.228	2.228	2.228	2.228	2.228

*shows the reliability and invariance of the estimated parameter values.

Table 5. 95% Confidence Interval (CI) of parameter estimation values calculated by Jackknife technique.

	X_2	X_3	X_5	X_6	X_8	R^2	$\hat{\sigma}$
Original coefficient	-1.489**	-2.259**	1.550**	1.491**	1.460**	0.927**	5.618**
Average pseudo values obtained by Jackk. Tech.	-1.439	-2.127	1.547	1.317	1.276	0.842	6.255
Average Pseudo Standard Error	0.459	0.785	0.623	0.503	0.632	0.048	2.931
Lower Limit 95% CI	-2.462	-3.876	0.159	0.196	-0.132	0.735	-0.275
Upper Limit 95% CI	-0.416	-0.378	2.935	2.437	2.684	0.948	12.785

** shows that original estimation values are within 95% confidence intervals obtained with Jackknife technique.

the population. In terms of other explanatory variables of average monthly credit card statement, gender and marital status, it can be seen in **Table 5** that Jackknife estimation values confirm the first original $\hat{\beta}$ values

calculated and likewise, Jackknife parameter values of the explanatory variables average monthly credit card statement, gender and marital status are within the 95% confidence interval and these variables can also be generalized for the population.

Table 4 shows that the explanatory variable of department $|t_{\text{calculation}}| = |2.019| < 2.228$, the effect of which is examined on the dependent variable of burnout, is not significant. In this case, it can be thought that the effect of the variable of department is sample-specific.

Lastly, $R^2 = 0.842$ calculated by Jackknife Method and the standard error value is 0.048. A t -value which is higher than the critical value shows that R^2 can also be generalized. That is, the specificity coefficient of the model is not sample specific; it can be generalized to population. That is, it can be said that in similar studies conducted in different times, the percentage of explaining the dependent variable with independent variables will be the same or similar to the values obtained from this study. Similarly, it can be said that due to their generalizability to the population, the explanatory variables X_2 , X_3 , X_5 , X_6 can be statistically significant variables in different studies.

4. Conclusion and Discussion

In the study, which used multiple linear regression method for the data set with a small sample, Jackknife Method was used to obtain real-like, strong and valid estimations. Maslach Burnout Scale was used in the study. With this scale, multiple linear regression models were formed, parameter estimations of explanatory variables important for the model were made and the confidence intervals of these estimations were formed and it was concluded that original beta estimation values were within these confidence intervals. In order to serve the purposes of Jackknife Method, the sample size was chosen as 11. For this reason, instead of repeating the study with a new sample, we could conclude that the results obtained with Jackknife Method were strong in terms of invariance, validity and reliability. In addition, for five explanatory variables which influenced the dependent variable, standardized beta coefficients and determination coefficient, and the 95% confidence intervals of these values were calculated. Of these explanatory variables, the variables of weekly hours of classes taught, monthly average credit card debt, gender and marital status were within the confidence interval of Jackknife parameter estimator and it was concluded that these variables could be generalized. However, it was concluded that the variable of department was found to be specific to only this sample. This study exhibited a concise investigation of the Jackknife Method for linear regression analysis with an application. The results obtained with Jackknife technique presented the replicability and generalizability property of the method. In further studies, other kind of resampling techniques could be examined such as bootstrap for regression context. We planned to adopt different resampling techniques and explored the results for several regression models.

References

- [1] Dolunay, A.B. (2002) Keçiören İlçesi “Genel Liseler ve Teknik—Ticaret ve Meslek Liselerinde görevli Öğretmenlerde Tükenmişlik Durumu” Araştırması. *Ankara Üniversitesi Tıp Fakültesi*, **55**, S. 51-62.
- [2] Terzi, Y. and Sağlam, V. (2008) Araştırma Görevlilerinin Mesleki Tükenmişlik Durumu. *Natural and Applied Sciences Statistics. E Journal of New World Sciences Academy*, **3**, Article ID: A0049.
- [3] Freudenberg, N.J. (1974) Staff Burnout. *Jouranal of Social*, **30**, 159-165.
- [4] Maslach, J. and Jackson, S.E. (1981) The Measurement of Experienced Burnout. *Journal of Occupational Behavior*, **2**, 99-113. <http://dx.doi.org/10.1002/job.4030020205>
- [5] Sucuoğlu, B. and Kuloğlu, N. (1996) Özürlü Çocuklarla Çalışan Öğretmenelerde Tükenmişliğin Değerlendirilmesi. Vol. 10, *Türk Psikoloji Dergisi*. Türk Psikologlar Derneği Yayınları. Cilt.
- [6] Bekiroğlu, N., Konyalıoğlu, R. and Karahan, N. (2013) Çoklu Doğrusal Regresyon Sonuçlarının Jackknife Tekniği ile Tekrarlanabilirliğinin Değerlendirilmesi. *Marmara Medical Journal*, **26**, 63-67.
- [7] Borg, W.R. and Gall, D. (1983) Educational Research. Longman, Newyork.
- [8] Kayri, M. and Büyüköztürk, Ş. (2009) Nicel Bilimsel Bulgulara Ait Genellebilirliğin Jackknief Yöntemi ile İncelenmesi: Uygulamalı Bir Çalışma. Kuram ve Uygulamada Eğitim Bilimleri. *Educational Science: Theory & Practice*, **9**, 1751-1780.
- [9] Thompson, S.K. (1992) Sampling. John Willey & Sons, Inc., Newyork.
- [10] Tukey, J.W. (1958) Bias and Confidence in Not-Quite Large Sample. *Annals of Mathematical Statistics*, **29**, 614.
- [11] Bradley, E. and Tibshirani, R.J. (1993) Monograph on Statistics and Applied Probability 57.

- [12] Fenwick, I. (1979) Techniques in Market Measurement: The Jackknife. *Journal of Marketing Research*, **163**, 410-414.
- [13] Abdi, H. and Williams, J. L. (2010) Jackknife in Neil Salkind (Ed.). Encyclopedia of Research Design.
<http://dx.doi.org/10.4135/9781412961288.n202>