Available online at www.sciencedirect.com

ScienceDirect

Procedia - Social and Behavioral Sciences 106 (2013) 234 – 240

Procedia
Social and Behavioral Sciences

4th International Conference on New Horizons in Education

A study on multiple linear regression analysis

Gülden Kaya Uyanık [i*], Neşe Güler [ii]

[i] Sakarya Unıversity, Hendek Education Faculty, Department of Educational Sciences, Sakarya, Turkey
[ii] Sakarya Unıversity, Hendek Education Faculty, Department of Educational Sciences, Sakarya, Turkey

Abstract

Regression analysis is a statistical technique for estimating the relationship among variables which have reason and result relation. Main focus of univariate regression is analyse the relationship between a dependent variable and one independent variable and formulates the linear relation equation between dependent and independent variable. Regression models with one dependent variable and more than one independent variables are called multilinear regression. In this study, data for multilinear regression analysis is occur from Sakarya University Education Faculty student's lesson (measurement and evaluation, educational psychology, program development, counseling and instructional techniques) scores and their 2012-KPSS score. Assumptions of multilinear regression analysis- normality, linearity, no extreme values- and missing value analysis were examined. The data that verify the assumptions were analyzed with multiple regression and lessons measurement and evaluation, instructional techniques, counseling, program development and educational psychology were estimate the KPSS respectively.

Keywords: Multiple Linear Regression educational sciences, KPSS

1. Introduction

Regression analysis is performed so as to determine the correlations between two or more variables having cause-effect relations, and to make predictions for the topic by using the relation.

*Tel: +905545856264, E-mail address: guldenk@sakarya.edu.tr

doi:10.1016/j.sbspro.2013.12.027

Answers are sought in this research to questions such as “are there any relations between dependent and independent variables?”, “if there are any relations, what is the power of the relation?”, “is it possible to make future-oriented predictions regarding the dependent variable?”, and “if certain conditions are controlled, what influences does a special variable or group of variables have over another variable or variables?” (Alpar, 2003).

The regression using one single independent variable is called univariate regression analysis while the analysis using more than one independent variable is called multivariate regression analysis (Tabachnick,1996, Büyüköztürk, 2002). Through univariate regression analysis, the relations between a dependent variable and an independent variable are analysed, and the equation representing the linear relations between the dependent and the independent variables is formulated. The regression models with one dependent variable and more than one independent variable, however, is known as multivariate regression analysis (Köksal, 1985; Tabachrick, 1996; Büyüköztürk, 2002).

In multivariate regression analysis, an attempt is made to account for the variation of the independent variables in the dependent variable synchronically (Ünver & Gamgam, 1999). Multivariate regression analysis model is formulated as in the following;

$$y = \beta_0 + \beta_1 x_1 + ... + \beta_n x_n + \varepsilon$$

y = dependent variable
X_i = independent variable
β_i = parameter
ε = error

The assumptions of multivariate regression analysis are normal distribution, linearity, freedom from extreme values and having no multiple ties between independent variables (Büyüköztürk, 2002).

The sample data used in the research were obtained from the end-of-the-term scores received from the measurement and evaluation, educational psychology, curriculum development, guidance, and teaching methods courses in the fall semester by 240 undergraduate students studying in the departments of Psychological Counselling and Guidance, Turkish Education, and Science Education of the Educational Faculty of Sakarya University in the 2011-2012 academic year as well as from the scores received from the KPSS exam (state employees selection examination) in 2012. Through the multiple linear regression analysis, it was checked whether or not the five independent variables in the standard model were significantly predictive of the KPSS score, which was the dependent variable, according to the ANOVA statistics.

2. Method

This is a descriptive study analysing whether or not the five independent variables in the standard model (namely, end-of-the-term scores received from the courses measurement and evaluation, educational psychology, curriculum development, guidance, and teaching methods) were significantly predictive of the KPSS score- the dependent variable- based on the ANOVA statistics.

2.1. Research Data

The research data were composed of the end-of-the-term scores received from the measurement and evaluation, educational psychology, curriculum development, guidance, and teaching methods courses in the fall semester by 240 undergraduate students studying in the departments of Psychological Counselling and Guidance,

Turkish Education, and Science Education of the Educational Faculty of Sakarya University in the 2011-2012 academic year as well as the scores received from the KPSS exam (state employees selection examination) in 2012.

2.2. Data analysis

Firstly, the availability of the lost data was checked in the research through frequency analysis. Then, multivariate normality, linearity, freedom from extreme values, and multi-linear relation- the assumptions of multiple regression analysis- were analysed.

Prior to multivariate normality analyses, univariate normality assumption was analysed for each quantitative variable. In order to do this, the skewness coefficient and kurtosis coefficient, and histogram charts were examined.

For the multivariate linearity assumption, scatter diagram matrix was prepared.

In order to check the unidirectional extreme values, it was checked whether or not the Z scores of the variables were in the ±3 range. So as to check the multidirectional extreme values of the data, the Mahanolobis distances were used.

So as to see whether or not there were any multiple ties, the simple correlations, variance increase factors (VIFs), tolerance value and the condition index (CI) were examined.

Through the data which was analysed so as to see whether or not they satisfied the assumptions and which were cleared for this purpose, the linear regression analysis was performed.

3. Results

Firstly, the data were analysed for convenience for regression analysis. For this purpose, following the missing data analysis, multivariate normality, multivariate linearity, freedom from extreme values, and ties between independent variables were researched, respectively.

For the analysis of the lost data, the data were analysed descriptively, and it was found that there were no missing data. The related findings are shown in Table 1.

Table 1. Frequency Table for Missing Data

	Measurement	Edu.psyc.	Teach.Meth.	Guidance	Curric.Dev.	KPSS
Being	240	240	240	240	240	240
Missing	0	0	0	0	0	0

Prior to multivariate normality analyses, univariate normality was analysed for all of the quantitative variables. In order to examine the univariate normality assumptions, the skewness coefficient and kurtosis coefficient of the variables were analysed.

Table 2. Descriptive Statistics

	N	Skewness		Kurtosis	
		Statistic	Std. Error	Statistic	Std. Error
Measurement	238	.028	.158	-1.327	.314
Educ.Psych.	238	-.046	.158	-.746	.314
Inst.Methods	238	-.141	.158	-1.258	.314
Guidance	238	.059	.158	-1.396	.314
Curri.Develop.	238	.146	.158	-1.343	.314
KPSS	238	-.053	.158	-.728	.314

According to Table 2, since the skewness coefficients for all the variables are within the acceptable ±1 range, the variables may not be said to be skew. An examination of the kurtosis coefficient obtained for the variables in the table shows that the kurtosis coefficients calculated for the variables apart from educational psychology and KPSS are outside the ±1 range. In examining the normality premise, skewness values are more important (Büyüköztürk, 2002), and because the kurtosis coefficient does not differ greatly from the normal, this variable may be said to distribute normally. The normality premise of the variables could also be shown by drawing a histogram chart.

For the multivariate normality and linearity assumption, the scatter diagram is examined for each group. Figure 1 shows the scatter diagram matrix for all the groups.

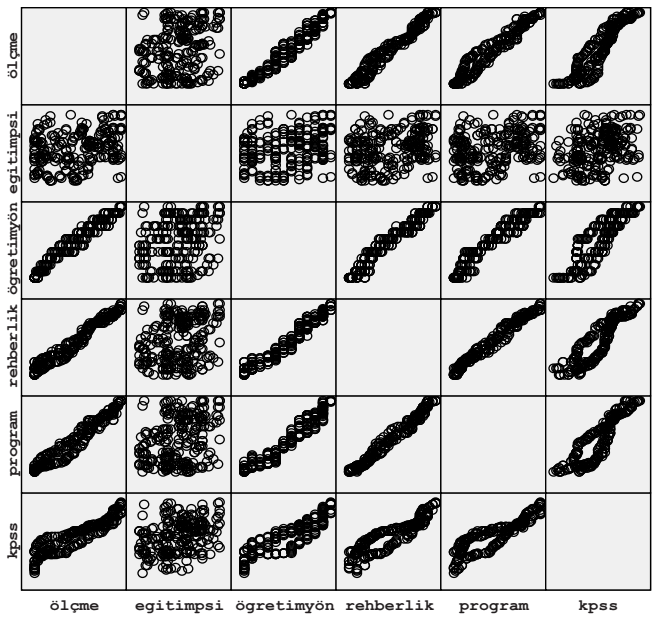

Figure 1. Scatter Diagram Matrix

On examining the scatter diagrams in Figure 1, it may be said that the diagrams are almost elyptic in shape.

Having examined the data so as to check the multidirectional extreme values by using the Mahanolobis distances (for df=4. p=0.01, Chi-square= 13.2767), person 31 and 21 were excluded from the analysis.

So as to see whether or not there were any multiple relations between variables, the simple correlations, variance increase factors (VIFs), tolerance value and the condition index (CI) were examined. The findings concerning these values are shown in Table 3 and Table 4.

Table 3. Multiple Relations Coefficients

	B	*Std. Error*	*Beta*	*t*	*p*	*Zero-order*	*Partial*	*Part*	*Tolerance*	*VIF*
Constant	*163.350*	*25.580*		*6.386*	*.000*					
Curr.Dev.	*6.606*	*1.471*	*1.642*	*4.491*	*.000*	*-.171*	*.283*	*.254*	*.124*	*1.821*
Measurem.	*3.449*	*1.291*	*.889*	*2.671*	*.008*	*-.210*	*.173*	*.151*	*.229*	*4.678*
Educ.Psys.	*1.174*	*.315*	*.226*	*3.730*	*.000*	*.187*	*.238*	*.211*	*.872*	*1.147*
Inst.Meth.	*-5.915*	*1.167*	*-1.479*	*-5.069*	*.000*	*-.263*	*-.316*	*-.287*	*.138*	*6.644*
Guidance	*-5.641*	*1.870*	*-1.340*	*-3.016*	*.003*	*-.219*	*-.194*	*-.171*	*.116*	*6.780*

When the simple corelations are examined, it is seen that none of the corelations coefficients are higher than .80. This case points that there is not any multiple relations between variables. If variance increase factors (VIF) equal or higher than 10, there is a multiple relations between variables and when looking at the Table 3, it can be seen that VIF values for all variables are smaller than 10. In additon to these, if tolerance values are higher than .10, no multiple relations between variables is decided. When examining the Talbe 3, all tolerance values are higher than 10. In Table 4, it can bee seen condition indexes (CI).

Table 4. Multiple Relations CI Values

Model	Factors	Eigenvalues	Condition Index	Variance Ratios					
				Constant	Curr.De.	Meas.	Edu.Psy.	Ins.Meth.	Guidance
1	1	5.912	1.000	.00	.00	.00	.00	.00	.00
	2	.066	5.457	.10	.00	.00	.18	.00	.00
	3	.018	8.293	.72	.00	.00	.78	.00	.00
	4	.002	12.053	.01	.26	.10	.02	.32	.04
	5	.001	19.992	.02	.03	.80	.00	.67	.00
	6	.001	22.271	.15	.70	.10	.02	.01	.96

In cases when the CI is bigger than 30, it is regarded that there are multiple relations between variables. Yet, as is clear from Table 4, all of the CI values calculated are smaller than 30. Thus, it was concluded that there were no multiple relations between variables.

With the data which were analysed whether or not they satisfied the assumptions and which were cleared accordingly, the regression analysis was done.

The findings obtained by doing the multiple linear regression analysis concerning whether or not the five independent variables in the standard model predicted significantly the KPSS score - the dependent variable- according to the ANOVA statistics, and the degree of emerging model's predicting the dependent variable in

consequence of the standard regression, and the degree of the model's explaining the variance in the dependent variable are shown in Table 5.

Table 5. Multiple Linear Regression Analysis Results Related to KPSS Scores

Variables	B	SH_B	β	T	p	Paired-r	Partial
Constant	9.811	2.261	-	4.340	.000	-	-
Measure.	1.157	.114	1.421	10.136	.000	.924	.554
Edu.Psyc.	.090	.028	.082	3.230	.001	.348	.207
Ins.Meth.	-.339	.103	-.404	-3.291	.001	.884	-.211
Guidance	-.195	.165	-.221	-1.181	.239	.894	-.077
Curri.De.	.078	.130	.093	.604	.547	.891	.040
		R=0.932	R^2=0.87	$F_{(5.232)}$ =306.534	p= 0.00		

An examination of Table 5 makes it clear that the five independent variables in the standard model are significantly predictive of the dependent variable KPSS score according to the ANOVA statistics [F(5, 232)= 306.534, p< .05] . In consequence of the standard regression analysis, the model's degree of predicting the dependent variable was found to be R= .932. The model's degree of explaining the variance in the dependent variable was R^2=0.87. Looking at these coefficients, it may be said that the model predicts the dependent variable very well.

The absolute value of β (Beta) in Table 5 indicates the order of importance of the independent variables. The variable with the highest β value is relatively most important independent variable. On examining the contributions made by the independent variables in the model to the model, it was found that the end-of- the-term-scores received from the measurement and evaluation course made the biggest contribution with the value of (β= 1.421). It was followed by the scores received from teaching methods, guidance, curriculum development, and educational psychology courses, respectively. Although the contributions made by the scores received from guidance and from curriculum development were significant, they entered the model due to the property of the regression analysis, and they were found to make the smallest contributions to the model with their determination coefficients of .221 and .093, respectively.

Based on the regression analysis results, the regression equation was obtained as it is shown below:

KPSS= 9.811 + 1.157 Measurement + 0.090 Educational Psychology – 0.339 Teaching Methods – 0.195 Guidance + 0.078 Curriculum Development.

4. Conclusions

This study analyses whether or not the five independent variables in the standard model (namely, end-of-the-term scores received from the courses measurement and evaluation, educational psychology, curriculum development, guidance, and teaching methods) were significantly predictive of the KPSS score, the dependent variable, based on the ANOVA statistics. The primary aim of the research is to exemplify the multiple linear regression analysis with its stages.

For our purposes, the end-of-the-term scores received in the measurement and evaluation, educational psychology, curriculum development, guidance, and teaching methods courses in the fall semester by 240 undergraduate students studying in the departments of Psychological Counselling and Guidance, Turkish Education, and Science Education of the Educational Faculty of Sakarya University in the 2011-2012 academic

year were analysed for the power of predicting the KPSS scores. Firstly, the assumptions necessary for the multiple linear regression analysis were examined in the research, and the regression analysis was performed with the data which were thought to satisfy the assumptions. The standard model's degree of predicting the dependent model was found to be R=0.932. The model's degree of explaining the variance in the dependent variable was R^2=0.87. Looking at these coefficients, it may be said that the model predicts the dependent variable very well. On examining the contributions made by the independent variables in the model to the model, it was found that the end-of- the-term-scores received from the measurement and evaluation courses made the biggest contribution with the value of (β= 1.421); which is followed by by the scores received from teaching methods, guidance, curriculum development, and educational psychology courses, respectively.

In consequence, it might be recommended that similar research studies concerning determining the variables predictive of the scores received in KPSS - which is a hurdle that prospective teachers should overcome before starting their career after graduation from educational faculties – should be conducted with different variables and with larger samples.

REFERENCES

Alpar. R. (2003) *Uygulamalı çok değişkenli istatistiksel yöntemlere giriş 1* (ikinci baskı). Ankara: Nobel Yayınları

Büyüköztürk. Ş. (2002). *Sosyal bilimler için veri analizi el kitabı* . Ankara: Pegem Yayıncılık.

Köksal. B.A. (1985) *İstatistik Analiz Metodları*. Çağlayan Kitabevi. İstanbul.

Tabachnick. B.G. & S.L Fidell. (1996). *Using multivariate statistics. (3.Edition)*. Harper Collins College Publishers. New York.

Ünver. Ö.. Gamgam. H. (1999) *Uygulamalı İstatistik Yöntemleri*. Siyasal Kitabevi. Ankara.