

A STUDY OF MULTIPLE REGRESSION ANALYSIS - A REVIEW

PAUL ANDREW BOURNE*

Overall this study by Uyanik and Guler (2013), published in the journal of Procedia-Social and Behavioral Sciences is a relatively good one, with some weaknesses in the areas of data analyses and methods. However, the article fulfills the following research prerequisites 1) identified a research problem; 2) stated an objective-the purpose of the study was to determine whether or not the five independent variables are significant predictors of KPSS-scores, 3) employed a method, 4) tested a hypothesis, 5) and generated discussion and conclusion. The authors succinctly present the use of multiple regression analysis by way of using 240 undergraduate students pursuing Psychological Counselling and Guidance at Sakarya University, for the academic year 2011-2012. The study relates to an employee selection examination that was conducted in 2012. The introduction of the paper is on substantiate issue of regression, highlighting the types (i.e., univariate and multivariate regressions), rationale for their usage (i.e. univariate, one dependent variable; multivariate, two or more dependent, and independent

variables), testing of assumptions of linearity (i.e. normality and linearity) and presents a purpose for the usage multiple regression analysis.

The weaknesses of the study are almost entirely in the method section of the paper. The researchers did not provide detailed information on the research design, operationalization, reliability and validity of variables, issues relating to ethics, and issues relating to conflict of interest. The findings are detailed, with some omissions (descriptive statistics-means, standard deviation, and confidence interval) and discussion was rich with appropriate conclusions that are generated from the research findings.

The findings section begins with a table that presents frequency for each of the variables, independent (i.e., measurement and evaluation; educational psychology; curriculum development; guidance and teaching methods) and dependent KPSS-scores, and missing values (Table 1). It can be deduced from this Table (Table 1) that the response rate is 100%.

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

Table 2 presents descriptive statistics (i.e. skewness and kurtosis) for all the variables in the model. Based on the model, it can be concluded that there errors are within the dataset. But that they are minimal because skewness values are less than ± 0.2 , which is far from ± 1 . However, one can disagree with the authors that there are no skewnesses in the data. In fact, minimal skewness does not mean no errors, it just means negligible

errors. Based on the kurtosis values (i.e., negative), the distribution is flattening and offers some insight that the variables are not in keeping with a normal distribution and that errors are present in all of the variables. As such, the minimal skewnesses of the variables allow for the usage of those constructs in bivariate correlations.

*Northern Caribbean University, Mandeville, Jamaica, W.I.
Correspondence E-mail Id: editor@eurekajournals.com

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

The Uyanik and Guler examine bivariate correlations by way of Ordinary Least Square (OLS) regression. This allowed for a single examination of all the independent variables and how they influence (or not the) the dependent variable (KPSS-scores). It can be deduced from Table 3 (OLS regression) that the level of significance is 5% (or a confidence interval of 95%).

A further extrapolation from Table 3 is that the bivariate correlations between each independent variable is significantly correlated with the dependent variable-using a level of significance of ≤ 0.05 . In fact, two of the five independent variables are negatively correlated with KPSS-scores (i.e. guidance and teaching methods), which can be seen from negative unstandardized B values. On the other hand, three are positively correlated with KPSS-scores.

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	.1642	4.491	.000	-.171	.283	.254	.124	1.821
Measur.	3.449	1.291	.889	2.671	.008	-.210	.173	.151	.229	4.678
Educ.Psyc.	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

The Uyanik and Guler speak of the non-strong correlation between each independent and the dependent variable (zero-order, $r_{xy} < 0.8$). They opine that “This case points that there is not any multiple relations between variables” (p. 238). While strong correlations may result in multicollinearity, this alone is not used to test for multicollinearity. It is this premise that the authors used to determine that multicollinearity does not exist and based this upon a VIF of 10. In fact, many scholars argued that a VIF in excess of 6 indicates a problem with multicollinearity, which is the case for teaching methods and guidance. The authors, however, fail to identify this problem because of the standard that they used. Moreover, Variance and Eigenvalues are

not used to determine multicollinearity, which was totally misinterpreted by the authors. They used Variance and Eigenvalues (in Table 4) to speak of multicollinearity, when Eigenvalues speak of weight of the variance for each independent value on the dependent variable. The Eigenvalue indicated that curriculum development had the greatest contribution to the variance followed by measurement, educational psychology, and lastly teaching methods and guidance. Furthermore, a STRONG correlation exists between curriculum development and guidance [0.7]; measurement and teaching methods [0.8]; and educational psychology and measurement [0.8] (Table 4).

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

Table 5 presents a multivariate regression analysis of the five independent variables (i.e., Curriculum; Measurement; Educational Psychology; Teaching Methods, and Guidance) on the dependent variable (i.e., KPSS-scores). The interpretation of this Table is well done including the β , ANOVA and the actual Unstandardized B for the actual model. The β value indicates the strength of each explanatory variable. The overall model is a good predictive one because the adjusted $R^2 = 0.87$, with the independent values being a good fit for the model ($F=306.5$, $P < 0.05$).

Overall, their final model (equation) fulfils a high statistical rigour; but the authors should have outlined that while collectively all 5 independent variables are significantly correlated with KPSS-scores, Curriculum and Guidance are not significantly related with KPSS-scores. Based on the value for the F-test and probability value, all the independent variables fit the linear model and so, they must be placed therein and I concur with them that they contribute the least to the model by way of the β values.

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

$R=0.932$ $R^2=0.87$ $F_{(5,232)}=306.534$ $p=0.00$

CONCLUSION

Overall, the paper fulfills all the requirements of good research as well as the critical issue of answering the research question. However, there are some errors embedded in the findings and these should have been corrected before the article was published. Despite the limitations of the paper, it is deserving of publication because

the weaknesses have not detracted from the overall quality of the research.

REFERENCE

- [1]. Uyanik, G.K. & Guler, N. (2013). A study of multiple regression analysis. *Procedia-Social and Behavioral Sciences*, 106, 234-240.