Prestige Journal of Counselling Psychology, Vol. 2, No. 1, March 2019
A publication of the Association for the Promotion of Innovation in Education (APIE)
ISSN: 2651-5687 (Print)
ISSN: 2651-5709 (Online)

Comparative Robusity of Analysis of Variance and Regression Statistics using the Assumptions of Sample Size, Homogeneity of Variance and Normality of the Distribution

[1]**Aniyom Asuquo Obambon**

[1]**Isaac Ofem Ubi, Ph.D**
doctorubi08@yahoo.com
[1]*Department of Educational Foundations*
University of Calabar, Calabar
Cross River State, Nigeria

Abstract
The violation of statistical assumptions in testing hypotheses has been a problem confronting most researchers in different fields of endeavour. This study was principally carried out to compare the robustness of Analysis of Variance and Regression Statistics using the assumptions of sample size, homogeneity of variance, and normality of the distribution. The study area was Southern Cross River State, Nigeria. Cluster and simple random sampling techniques were adopted in selecting a sample of 640 students from a student population of 4,265. Two instruments were used for data collection. The first was a questionnaire called Students' Attitude and Test Anxiety Questionnaire (SAATAQ), while the second was scores from past examination scripts of WAEC May/June SSSCE for 2015/2016 (multiple choice test only). Cronbach Apha reliability estimate method was adopted to test the reliability of the instruments. The researchers tried, as much as they could, to meet the three assumptions of Analysis of Variance and Regression Statistics chosen for the study. Results of the study indicate that the F-ratios of both test statistics were robust as their p-values were .000 each. The study recommends strict adherence to the three assumptions by researchers when Analysis of variance and Regression statistics are proposed for a study.
Keywords: Comparative, ANOVA, Regression, Sample Size, Homogeneity, Variance, Normality

Introduction
The problem of knowing the right type of statistical tool to use in educational research, most especially among students, has been of great concern to many scholars. The problem has prompted many studies in the area of statistical significance (Akpan & Hay, 1993; Herr, 1986). The application of the right type of statistics in some research situations is of great concern to scholars. Regression analysis and analysis of variance are two powerful statistical tests that share similar assumptions. This is because they

belong to the family of the general linear model (multivariate or parametric statistics). For instance, homogeneity of variance in One-Way Analysis of Variance (One-way ANOVA), where the groups have the same variance on the outcome variable, is similar to Regression Analysis where a single continuous predictor, that is, the variance around the regression line, is the same at every point along the X axis.

Two-way Analysis of Variance (Two-way ANOVA) and multiple regression analysis have the same underlying principles of the general linear model. These are multivariate statistics with the following similar basic assumptions: the distribution of the dependent variable in the population from which the samples are drawn are normally distributed, the groups measured are independent of each other, the samples are randomly drawn from the population, the sample size should not be less than twenty, there must be homogeneity of variance among the treatment groups meaning that the populations from which the samples are drawn are equal, and so on. Two-way ANOVA is a multifactor statistical technique used when the researcher wishes to examine the combined effect of two independent variables, each categorized on a dependent variable. The F-Ratio is a test statistic for ANOVA proposed by Fisher in 1923. Conventionally, Two-Way ANOVA is believed to be from the same family with Multiple Regression. That is why when an analysis is done with regression, its output will also display ANOVA result, this signifies that both share certain similarities in common.

The three uniform assumptions of ANOVA and regression being considered here are firstly for ANOVA, it includes sample size, normality of distribution and homogeneity of variance; secondly for regression it is linear relationship, multivariate normality, and homescedacity. These assumptions can be paired one-on-one to mean the same thing for the two test statistics. In performing regression analysis, there is an ANOVA in its table, and also in performing ANOVA there is regression in the ANOVA table.
This study is centered on comparing the conventional two-way ANOVA (factorial design) with the normal Multiple Regression. It is very difficult to distinguish the differences between ANOVA and regression. This is because both statistical models have more similarities than differences. It can be said that ANOVA and regression are the two sides of the same coin. Analysis of variance (ANOVA) is similar to regression in that it is used to investigate and model the relationship between a response variable and one or more independent variables.

Regression analysis is a general procedure referring to the determination of statistical relationship between two or more variables. Multiple regression analysis, a type of regression analysis, is the study of how a dependent variable (y) is related to two or more independent variables. The regression coefficient (R) in regression can be converted to F in ANOVA, (Ukwuije, 2003). In ANOVA, significant F-ratio signifies

ISSN: 2651-5687 (Print)
ISSN: 2651-5709 (Online)

that the difference between the means is statistically significant; Ho: $x_I = x_2$. In regression a significant F-ratio implies that R^2 is significant (i.e relation between x and y, x is membership); Ho: $R^2 = 0$. Also in ANOVA, F is used to measure relationship, therefore $F^2 = R^2$.

On the basis of these similarities, comparison can be made on the strength of relationship/association as well as comparison between the variables involved in ANOVA and regression analysis. Most statistical instruments have the same underlying principles of the general linear model. Such statistics as analysis of variance and regression analysis often suffer abuse because of the assumption that each of them can solve the same problem. But a closer analysis will reveal that some may be more robust. Robust in this study means that the result of the test statistics is very significant, noticeable or which amongst the two test statistics produce a higher level of significance. Put in another way, which among the two test statistics is less resistant to violations of some of the assumptions.

The term robust statistics refers to procedures that are able to maintain the type 1 error rate of a test at its minimal level and also maintain the power of the test, even when data are non-normal and heteroscedastic (Wilcox, 2001). A Type 1 error occurs when the null hypothesis is falsely rejected (rejecting a null when the null is true). In other words, one concludes that a real effect exists when it does not. In contrast, a type II error occurs when the null hypothesis is not rejected even though it is false (accepting the Ho when it is not true). The power of a test is the probability that a type II error will not occur. The robustness of statistical techniques as indicated in their p-values, according to Kirk (2005), was a necessary part of a statistical significance testing in a research. Kirk added that the time had come to include practical significance in results of analysis.

Cumming (2007) conducted a review with large distribution of psychological variables, and discovered that, for robust parametric test to produce accurate results, the assumptions underlying them must be sufficiently satisfied. However, these assumptions are rarely met when analyzing real data, hence the need to employ robust statistical models. Cumming (2007) conducted an experimental study with 221 secondary school students on robustness of ANOVA and regression using attitude towards mathematics and discovered that these two statistical analyses are robust especially when assumptions are not violated. Robust statistical procedures exist that can solve the problems even when assumptions are violated (Wilcox, 2001).

Standard statistics tests indicate that the expected value of the F ratio is 1.0 [more precisely N/(N-2)] in a completely balanced fixed effects ANOVA, when the null hypothesis is true; even though some authors suggest that the null hypothesis is rarely true in practice. F-ratios less than 1.0 are reported quite frequently in literature. The research methodology literature in recent years has included a full frontal assault on statistical significance testing (Thompson, 2010). A level at which the test statistics should be rejected and is set a priori to conducting the test of data. A null hypothesis (Ho) and an alternative hypothesis (Ha) are stated, and if the value of the test statistic falls in the rejection region the null hypothesis is rejected (Thompson, 2010). It is because of different approaches to analyses and differences in philosophical beliefs that the issue of testing for statistical significance has arisen. Huberty's (2004) historical review of the importance of statistical significance testing confirmed this belief when he asserted that the research community has shifted from one perspective to another, within the same circle.

Onwuegbuzie and Levin (2002) posit that using robust methods in analyzing data is recommended instead of conducting classic parametric analyses on transformed data. He maintained that robust statistics are designed to perform well when classic assumptions are met, as well as when they are violated. Therefore, analyses conducted using robust methods should usually be trusted.

Turkey (2002) observed that the poor performance of classical statistics in the presence of small departure from normality has led some statisticians to warn that routine use of classical statistics is unsafe. He recommend that classical estimate of means, variances, F-values and correlations should only be used in conjunction with alternative methods that are robust with respect to departures from normality. Glass, Peckham and Sanders (2006) assert that in behavioural sciences, ANOVA is robust with respect to type 1 error (a type 1 error is rejecting a true null hypotheses) and non-normality. Zimmerman and Zumbo (2001) observed that although ANOVA has some moderate robustness properties with respect to type 1 error and non-normality, it is, in relation to type II error, (a type II error is accepting a false null hypothesis) very non-robust; this places a researcher in an unusual situation when interpreting ANOVA result.

Multiple Regression assumes that variables have normal distributions (Darlington, 2004; Osborne & Waters, 2002). This means that errors are normally distributed, and that a plot of the values of the residuals will approximate a normal curve (Keith, 2006). The assumption is based on the shape of normal distribution and gives the researcher knowledge about what values to expect (Keith, 2006). Once the sampling distribution of the mean is known, it is possible to make predictions for a new sample (Keith, 2006). Non-normally distributed variables can distort relationship and significance tests (Osborne & Waters, 2002). Outliers can influence both type I and Type II errors and the overall accuracy of results (Osborne & Waters, 2002). Multiple regression techniques gives researchers flexibility to address a wide variety of research questions

(Hoyt, 2006). Johnson (2013) in a study with 205 respondents found out that the smaller the p-value, the larger the significance because it tells the investigator that the hypothesis under consideration may not adequately explain the observation.

Statistical analyses play a major role in the work environment in areas such as business, science, finance, economics, engineering and education. This is centered on the fact that scholars in every sphere of human endeavour need one form of statistical analysis or the other. Researchers are often required to use the appropriate statistical tools, which when wrongly used will lead to inaccurate result and a faulty generalization of the findings. It has equally been observed that the problem of knowing the right type of statistical tool to use by educators, most especially among budding researchers, has been of great concern in recent times. Even when the right type of statistics is used, the validity of procedure adopted depends on certain assumptions it makes about various aspects of the problem. For instance, well-known linear methods such as Two-way Analysis of Variance (Two-way ANOVA) and Regression Analysis depend on assumptions. How well are researchers handling these assumptions is the reason for embarking on this study.

Research question: What is the difference in the level of robustness between ANOVA and Regression as determined by the values of F-Ratio?

Methodology

This study employed the descriptive survey design by employing a very large sample and using a questionnaire to collect data from the subjects. Data so collected was analyzed and its results used in describing the phenomenon under investigation. The sample size was 640 subjects drawn from a population of 4,265 Senior Secondary three (SS3) students in the seven Local Education Authorities (LEAs) in the Southern Education Zone of Cross River State, Nigeria. The procedure for sampling was multi-staged. The first stage involved clustering the students alongside schools and LEAs. The second involved selecting the subjects from each of the schools in the area. To do this, the researchers got the comprehensive list of the students in each school and used the ‘Hat and Draw’ method suggested by Denga and Ali (1998) in randomly selecting the number required.

Two instruments were used for data collection. The first was a questionnaire called Students Attitude and Test Anxiety Questionnaire (SATAQ), while the second was the West African Examination Council (WAEC) Mathematics Objective test for 2013/2014 academic session. The SATAQ was developed by the researchers while the WAEC test was adapted. The SATAQ was made up of two sections (A & B). Section A, which was on demographics, elicited data on the name of school, sex of the student and the class the student was as at the time of the survey. Information required in

Section A was basically nominal, so the subjects were required to only give the name of their school and tick 'Male' or 'Female' and 'SS1', 'SS2' or 'SS3' as it applied to them. Section B had 24 items, each of which was measured on a Likert-type scale of Strongly Agree (SA), Agree (A), Disagree (D), and Strongly Disagree (SD). Respondents were required to tick one of those options per item. 'SA' was scored four points, 'A' was scored three points, 'D' was scored two points, and 'SD' was scored one point for positively worded items. The scoring was reversed to one point, two points, three points and four points for 'SA', 'A', 'D', and 'SD' respectively for negatively worded items. The reliability estimate carried out showed that the instruments were highly reliable with coefficients of estimates of .79, .77 and .74 for Students' Attitude, Students' Test Anxiety and WAEC performance respectively. Data was collected personally by the researchers with skeletal assistance, in some school, from class teachers. The test statistics used for data analysis were Two-Way Analysis of Variance, and Multiple Regression Analysis. The two statistical techniques were deemed adequate for the analysis because they are both multivariate in nature and known to have the same assumptions. Their assumptions, though different in nomenclatures, mean the same thing when compared one-on-one.

Presentation of results

Research question: What is the difference in the robustness between ANOVA and Regression as determined by the F-ratio?

Data used for analysis in this research question were from a survey on students' attitude, test anxiety, and mathematics performance in southern Cross River State, Nigeria. The study involved 635 students and data obtained from field indicated the three assumptions of the two test statistics explored were, to a great extent, met. The sample size was large (see Table 1), variances for the two independent variables and the one dependent variable were homogeneous (see Table 1), and the performance scores were normally distributed (see Figure 1). To answer the research question, a Univariate Analysis of Variance test and a Multiple Regression analysis were carried out using the respondents' attitude and test anxiety scores as independent variables and their Mathematics Performance test as the dependent variable. Results of the analysis are presented in tables 1 and 2.

ISSN: 2651-5687 (Print)
ISSN: 2651-5709 (Online)

Table 1: Summary of descriptive statistics showing the sample size and the variance for variables and sub-groups

S/No	Variable	Grouped	N	mean	SD	SD^2
1	Attitude	Positive	128	10.28	3.63	13.18
		Moderate	329	8.72	3.61	13.03
		Negative	178	9.31	3.80	14.44
		Total	635	9.20	3.76	14.14
2	Test Anxiety	Low	146	9.32	4.10	16.81
		Moderate	382	9.19	3.99	15.92
		High	107	9.07	4.01	16.08
		Total	635	9.20	3.76	14.14

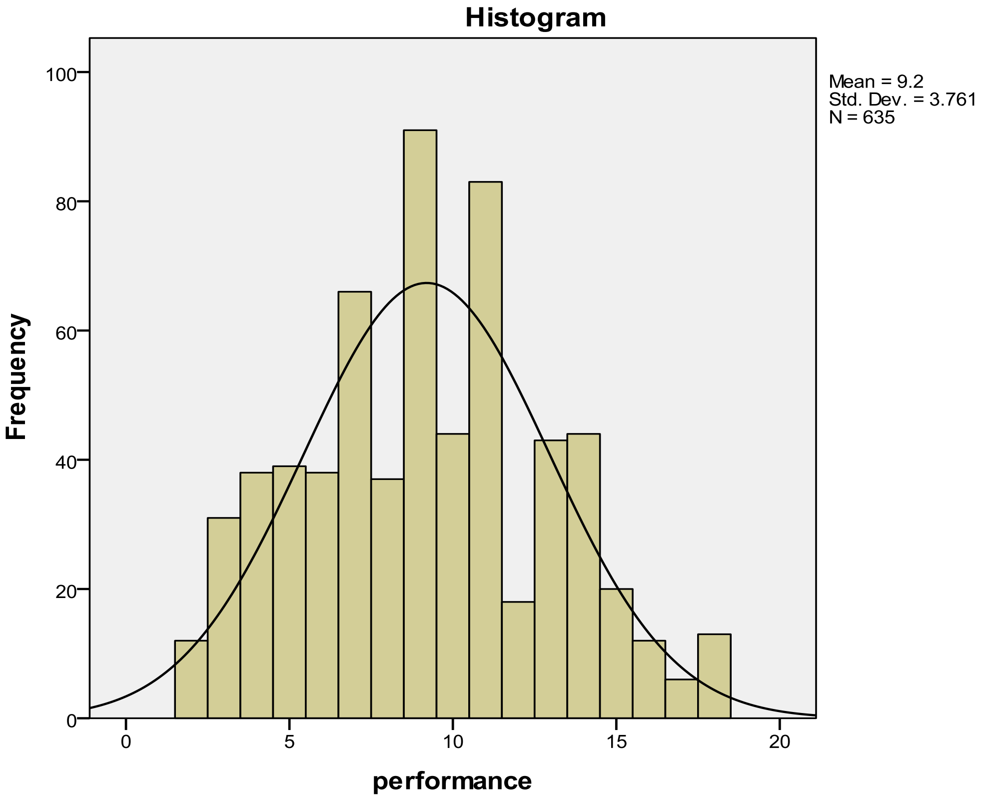

Figure 1: Histogram showing positively skewed distribution (sk = 0.162)

Table 2: Summary of two-way ANOVA for the composite effect of attitude and test anxiety on students' mathematics performance as measured by the F-ratio

Source of Variation	Sum of Squares	df	Mean Square	F	p-value
Corrected Model	615.532	8	76.942	546.791	.000
Intercept	33043.397	1	33043.397	2476.831	.000
Attitude	223.448	2	111.724	8.374	.000
Anxiety	109.994	2	54.972	4.121	.017
Attitude*Anxiety	362.378	4	90.595	6.791	.000
Error	8351.466	626	13.341		
Total	62695.000	635			
Corrected Total	8966.998	634			

Table 3: Summary of Multiple Regression analysis for the composite effect of attitude and test anxiety on students' mathematics performance as measured by the F-ratio

Source of Variation	Sum of Squares	df	Mean Square	F	p-value
Regression	263.720	2	131.860	9.575	.000
Residual	8703.279	632	13.771		
Total	8966.998	634			

Results of analysis in Table 1 show that there were 635 subjects in the sample with a distribution of 128, 329, 178 for positive, moderate, negative attitudes respectively, and 146, 382, 107 for low, moderate, high levels of test anxiety respectively. This shows that the sample size is large thus meeting the assumption of Sample size. The variances of the groups were 13.18, 13.03, 14.44 for positive, moderate, negative attitudes respectively, and 16.81, 15.92, 16.08 for low, moderate, high levels of test anxiety respectively. This variances were close, thus shows that the scores met the assumption of homogeneity of variance. Result in Figure 1 is a normal curve on the same data super-imposed on the histogram. This shows the skewed level of the distribution. Statistically, the mode was 0.162 standard deviations less than the median. This index is low, thus the study concludes that, more of the scores obtained by the respondents in the mathematics test were at the bell-shape of the curve. This means that, the distribution was, to some great extent, normal. It was based on this that the researchers assumed that the data also met the normality assumption.

Result in Table 2 shows that the F-ratio was 6.791 with a p-value of 0.000 when a two-way ANOVA was carried out using attitude and test anxiety scores on students' performance in mathematics. To carry out the analysis, the respondents were categorized into those with positive, moderate and negative attitudes, and those with low, moderate and high test anxiety. Result in Table 3 shows that the F-ratio was 9.575 also with a p-value of 0.000 when a multiple regression analysis was carried out using attitude and test anxiety scores. This time the scores were used as continuous. The comparative analysis of the two test statistics showed that, the p-values obtained using the two were each 0.000. This means that none was significantly more robust than the other when the three assumptions are met, especially when the p-value is used as the basis of argument. It was however noticed that there was a slight difference in the F-ratios obtained with Multiple Regression Statistic (F = 9.575) being higher than ANOVA Statistic (F = 6.791). In more precise terms, therefore, this probably indicates that the Regression Statistic is more robust when consideration is based on their F-ratios.

Discussion of findings

The finding obtained in this study showed that when the three assumptions considered in this study were all met, the result of the data analysis using ANOVA was found to have an F-ratio of 6.791 with a P-value of 0.000; similarly, the result of the data analysis using multiple regression also showed an F-ratio of 9.575 with a p-value of 0.000. The comparative analysis of the two test statistics showed that, the P-values obtained using the two tests statistics were each 0.000, this means that none was significantly more robust than the other when the three assumptions were met. This result is supported by Kirk (2005) on the robustness of statistical techniques as indicated in their p-values that statistical significance testing was a necessary part of statistical analysis testing. This result is also in agreement with the fining of Onwuegbuzie and Levin (2002) who posited that using robust methods in analyzing data is recommended instead of conducting analyses on transformed data. He maintained that robust statistics are designed to perform well even when assumptions are met, as well as when they are violated. Therefore, analyses conducted using robust methods should usually be trusted.

This result is supported by the finding of Cumming (2007) who conducted an experimental study with 221 secondary school students on robustness of ANOVA and regression and discovered that these two test statistics are violated. This result is also in agreement with the findings of Johnson (2013) who in a study with 205 respondents found out that the smaller the P-value, the larger the significance because it tells the investigator that the hypothesis under consideration may not adequately explain the observation.

Conclusion and recommendations

This study was a kind of litmus test on three assumptions of ANOVA and Regression statistics. The two statistical tools share these three assumptions, so a fair attempt was made at meeting the assumptions in the process of handling data during their use for analysis. Results of the survey, as were noticed, showed that the p-values obtained using the two were each 0.000. This means that none was significantly more robust than the other when the three assumptions are met, especially when the p-value is used as the basis of argument. It was however noticed that there was a slight difference in the F-ratios obtained with Multiple Regression Statistic ($F = 9.575$) being higher than ANOVA Statistic ($F = 6.791$). In more precise terms, therefore, this probably indicates that the Regression Statistic is more robust when consideration is based on their F-ratios. These results are inconclusive. More research work should be carried out in the area and such research should attempt to violate the assumptions to see if by such violations any one of the statistical tools can yield a lower p-value than the other. Such a result will imply that the statistic with the lower p-value is more robust. Until then, this present study recommends that researchers should always endeavour to meet the basic assumptions about a distribution when applying ANOVA and Regression Statistics.

References

Akpan, S. M. & Hay, H. O. (1993). *Research methods and statistics in Education.* Calabar: Matrine publishing Press.

Cumming, J. H. (2007). *Statistical models: Theory and practice*. Cambridge: Cambridge University press.

Darlington, R. (2004). Multiple regression psychological research and practice. *Psychological Bulletin*, 69(3),161-182.

Denga, D. I. & Ali, A. (1998). *An introduction to research method and statistics in education and social sciences*. Calabar: Rapid Education Publishers Limited.

Glass, G. V., Peckham, P. D. & Sanders, J. R. (2006). Consequences of failure to meet the assumptions underlying the fixed effects analysis of variance and covariance. *Reviews in Educational Research*, 42, 237-248.

Herr, D. G. (1986). On the history of ANOVA in unbalanced factorial designs. The first 30 years. *The American Statistician, 40,* 265-270.

Hoyt, W. (2006). Analysis and Interpretation of findings of using multiple regression technique. *Rehabilitation Counselling Bulletin*, 49(4), 223-233.

Huberty, C. J. (2004). Historical Original of Statistical testing practices. The treatment of fisher versus Neyman-Pearson Views in text books. *The Journal of Experimental Education*, 61(4), 317-333.

Johnson, V. (2013). Revise Standards for statistical evidence. Proceedings of the National Academy of Science. USA.

Keith, T. (2006). *Multiple regression and beyond*. Pearson, NY: Allyn & Bacon.

Kirk, R. E. (2005). Statistical significance testing from three perspectives. *The Journal of Experimental research*, *7*(4), 19-33.

Onwuegbuzie, A. J. & Levin, J. R. (2002). Without Supporting statistical evidence, where would reported measures of substantive importance lead us? To no good effect. *Journal of modern Applied Statistical methods*, 2, 131-151.

Osborn, J. & Waters, E. (2002). Four assumptions of regression that researcher should always test. *Practical Assessment, Research and Evaluation*, *8*(2), 32-41.

Thompson, B. (2010). The use and misuses of Statistical Significance Testing. *The Journal of Experimental Education, 6*(4), 26-31.

Turkey, J. W. (2002). *A survey of sampling from contaminated distributions*. Stanford, CA: Stanford University Press.

Ukwuije, R. P. I. (2003). *Peanuts educational statistics.* Port-Harcourt: Celwil Nigeria Limited.

Wilcox, R. R. (2001). *Introduction to robust estimation and hypothesis testing (2nd ed)*. San Diego, CA: Academic Press.

Zimmerman, D. W. & Zumbo, I. (2001). Statistical Significance level of non parametric test groups. *Journal of General Psychology, 127, 354-364.*