Prestige Journal of Education, Vol. 3, No. 2, December 2020 ISSN: 2645-324X (Print)
A publication of the Association for the Promotion of Innovation in Education (APIE) ISSN: 2645-3223 (Online)

Statistical Assumption of Homogeneity of Variance: A Query on the Robustness of Analysis of Variance and Regression Statistics

[1]**Isaac Ofem Ubi, Ph.D**
doctorubi08@yahoo.com

[1]**Anastecia William Osang**
osanganastecia@gmail.com

[1]**Aniyom Asuquo Obambon**
aniyomfavour1@gmail.com
[1]*Department of Educational Foundations*
University of Calabar, Calabar

Abstract
The study examined the robustness of analysis of variance (ANOVA) and regression statistics in the face of the homogeneity of variance assumption. The sample was made up of 640 students selected through stratified random sampling technique from secondary schools in Calabar Education zone of Cross River State, Nigeria. Two instruments – Students' Attitude and Test Anxiety Questionnaire (SATAQ), and WAEC May/June SSCE Mathematics objective test for 2017/2018 academic session were used for the study. Results of the study showed that Multiple Regression Statistical tool was more robust than the Two–Way ANOVA when the homogeneity of variance assumption is violated. The study recommends strict adherence to the assumption of homoscedasticity when ANOVA and Regression Statistics are proposed for a study.
Keywords: Robustness, homogeneity, variance, Analysis, Regression, homoscedasticity, heteroscedasticity.

Introduction

The basic problem of every researcher is how to infer properties of the population from the investigation of a sample. Statistical tests assist the researcher to establish empirical evidence for arriving at the strongest possible conclusion from limited amounts of data. They help to prove conclusions and inferences beyond reasonable doubt. This is made possible when the right type of statistical tool is applied in carrying out the test of significance. The more appropriate the tool used, the more likely will the result be meaningful. That is, a statistical tool may produce more robust result than

another, to the extent of its adequacy for the data it is used on. This underscores the importance of considering underlying assumptions for the use of statistical tools.

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, Obambon and Ubi (2019) said

> Homogeneity of variance in Analysis of Variance (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 (pp13).

This study is centred 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 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 'X'. The regression coefficient 'R' in regression can be converted to 'F' in ANOVA (Ukwuije, 2003). In ANOVA, significant F-ratio signifies that the difference between the mean scores is statistically significant; Ho: $X_I = X_2$. In regression a significant F-ratio implies that R^2 is significant, 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 produces higher level of significance. Putting this in another way, which among the two test statistics is less resistant to violation of the homogeneity assumption?

The F-test is computed by dividing the explained variance between groups (e.g., test anxieties) by the unexplained variance within the groups. Thus, statisticians frequently make the comment that analysis of variance (ANOVA) and linear regression are robust to the assumptions of homogeneity of variance (equal variance). This assumption provides strong back up to the robustness claim (Bradley, 1980). However, analysis of variance is simply a special case of linear regression. This is so because, analysis of variance and regression analysis are a particular form of statistical hypothesis testing heavily used in the analysis of experimental and predictive studies. This present study is an attempt at carrying out an analysis of variance test on a group of scores and also regressing those same scores. Each of the computations will result in an F-ratio, which will be compared for robustness.

Most statistical tests rely upon certain assumptions about the variables used in the analysis. Homogeneity of variance (homoscedasticity) means that the variance of errors is the same across all levels of the independent variables. When the variance of errors differs at different values of the independent variables, the opposite (heteroscedasticity) is indicated.

Berry and Feldman (1985) replicated by Tabachnick and Fidell (2019) assert that slight heteroscedasticity has little effect on significant tests in regression; however, when heteroscedasticity is marked it can lead to serious distortion of findings. This can seriously weaken the analysis thus increasing the possibility of a type 1 error - over-estimation or under-estimation of significance. Pedhazur (1997) noted that knowledge and understanding of the situations when violations of assumptions lead to serious biases, and when they are of little consequence, are essential to meaningful data analysis.

The assumption of homogeneity of variance requires equal population variances per group in ANOVA, and equal population variances for every value of the independent variable for regression (Kashy & Russel, 2009). Although researchers might be tempted to think that most statistical procedures are relatively robust against these violations, several studies have shown that this is not often the case, and that in the case of ANOVA, unequal group sizes can have a negative impact on the technique's robustness (Lix & Wilcox, 2008). This creates a situation where there is a rich literature in education and social sciences, but the present study is forced to call into question the validity of many of these results, conclusions, and assertions, as there is no idea whether the assumptions of the statistical tests were met.

In analyzing data with multiple regression model, if the variance of the 'y' is not constant, then the error variance will not be constant (Green & Salkind, 2011). Lorenzen and Anderson (2004) conducted an experimental study with scores that were

heteroscedastic, using the independent variable X and the dependent variable Y; it was discovered that the common form of such heteroscedasticity in Y is that the variance of Y may increase as the mean of Y increases, for data with positive X and decrease as the mean of Y decreases, for data with negative X and Y. They further discovered that "unless the heteroscedasicity of the Y is high enough, its effect on the P-value will not be severe; the least square estimates will still be unbiased."

Historically, studies on homogeneity of variance have been somehow conflicting. For instance, the study by Lorenzen and Anderson (2004) opined that the assumption thought to be the most critical among the assumptions of ANOVA and Regression Analysis was the homogeneity of variance assumption. This, however, negates previous findings by Box (1954), which demonstrated that the 'F' test in ANOVA was most robust for alpha (α) while working with a fixed model having equal sample sizes. The study showed that for relatively large (one variance up to nine times larger than another) departures from homogeneity, the alpha (α) level may only change from .05 to about .06. This is not considered to be of any practical importance. It should be pointed out that the only time an alpha (α) level can increase dramatically is when the sample size is negatively correlated with the size of the variance.

The assumption of homogeneity of variance requires equal population variances per group in ANOVA, and equal population variances for every value of the independent variable for regression. Although researchers might be tempted to think that most statistical procedures are relatively robust against violation, several studies have shown that this is not often the case, and that in the case of ANOVA, unequal group sizes can have a negative impact on robustness. Nagin (2005) observed that the strength of the regression model is its ability to identify groups or classes of individuals with dissimilarities in relationship.

Huberty (2004) carried out an experiment to determine the resistance nature of multivariate techniques to a little departure from homogeneity of variance assumption and discovered that multiple regressions are insensitive to a little deviation from homogeneity of variance assumption. Based on this finding, he interpreted robustness as insensitive to small deviations from the assumption the model imposes on the data.

Common exercise in empirical studies is a "robustness check," where the researcher examines how regression coefficient estimate behaves when the regression specification is modified in some way, typically by adding or removing repressors. Kirk (2005) advocated investigations of this sort, arguing that "fragility of regression coefficient estimates is indicative of specification error, and that sensitivity analyses (i.e. robustness checks) should be routinely conducted to help diagnose misspecification in statistical data before analyses."

Bivariate/multivariate data cleaning can also be important in Multiple Regression. Most regression or multivariate statistics texts (e.g, Pedhazur, 1997) discuss the examination of standardized or student Z residuals, or indices of leverage. Analysis by Osborne and Waters (2002) showed that removal of univariate and bivariate outliers can reduce the probability of Type I and Type II errors, and improve accuracy of estimates. Outlier removal is straightforward in most statistical software.

The assumptions of equality of variances also called homogeneity of variance, or homescedasticity can be framed in terms of a variance ratio (VR). If two populations have similar variances, their VR will be close to 1:1. Large variance ratios have also been found in reviews of studies published in clinical and experimental psychology journals. Nagin (2005) observed that "the strength of the regression model is its ability to identify groups or classes of individuals with similarities in relationship."

Research questions

1. What is the comparative robusity of ANOVA and Regression statistics when the assumption of homogeneity of variance is met?
2. What is the comparative robusity of ANOVA and Regression statistics when the assumption of homogeneity of variance is not met?

Methodology

The study area was Calabar Education Zone of Cross River State, Nigeria. The design adopted for the study was survey. A sample of 640 respondents was selected from the population of 4,265 students of secondary schools in the study area. The selection was done through stratified random sampling technique. The basis of stratification was the seven Local Education Authorities of the zone. From these seven authorities, 11 secondary schools were selected and the sample drawn through simple random sampling technique.

Two instruments were used for data collection. The first was named Student Attitude and Test Anxiety Questionnaire (SATAQ), while the second was an adopted 2017/2018 WAEC past Mathematics Objective Test (WAEC-Mat). The SATAQ was developed by the researchers and consists of two sections. Section A is made up of students' personal data while section B is made up of a four point Likert scale responses of Strongly Agree (SA), Agree (A), Disagree (D) and Strongly Disagree (SD). The section B had a 24 item information on students' attitude and test anxiety ranging from Strongly Agree (SA) to Strongly Disagree (SD). Responses were scored such that SA=4 points, A=3 points, D=2 points and SD=1 point for all positively worded items, and reversed to SA=1 point, A= 2 points D=3 points and SD= 4 points for negatively worded items.

To ascertain the validity of the instruments, SATAQ was subjected to face validity, while content validity through a test blueprint was conducted on WAEC-Mat. The reliability estimates using Cronbach Alpha method for SATAQ and Split Half method for WAEC-Mat were .78 and .74 respectively. After administering the instruments, the data that was obtained from these instruments was analyzed and interpreted accordingly. The data was analysed using Two-Way ANOVA and Multiple Regression.

Presentation of results

Research question one: What is the comparative robusity of ANOVA and Regression statistics when the assumption of homogeneity of variance is met?

Data used for answering this research question were from responses to the two instruments, SATAQ and WAEC-Mat. Data obtained from field indicated that the assumption of homogeneity of variance was, to a great extent, met. The variances for the two independent variables and the one dependent variable were homogeneous (see Table 1). To answer the research question, a Two-Way Analysis of Variance test and a Multiple Regression analysis were carried out using the respondents' attitudinal 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, 2 and 3.

Table 1: Summary of descriptive statistics showing the sample size and the variances across groups when the assumption of homogeneity is met.

Variable	Groups	n	Mean	SD	SD^2
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
Test Anxiety	Positive	146	9.32	4.10	16.81
	Moderate	382	9.19	3.99	15.92
	Negative	107	9.07	4.01	16.08
	Total	635	9.20	3.76	14.14

ISSN: 2645-324X (Print)
ISSN: 2645-3223 (Online)

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 when the assumption of homogeneity is met

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 when the assumption of homogeneity is met

Source of Variation	Sum of Squares	df	Mean Square	F	p-value
Regression	563.720	2	131.860		
Residual	8703.279	632	13.771	9.575	.000
Total	8966.998	634			

Results in table 1 show that 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. These variances were close, thus show that the scores met, to a high extent, the assumption of homogeneity of variance.

Results in Table 2 show that the F-ratio was 6.791 with a p-value of 0.000 when a two-way ANOVA was carried. Result in Table 3 show that the F-ratio was 9.575 also with a p-value of 0.000 when a multiple regression analysis was carried out using the same data used for the ANOVA test. 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 assumption of homogeneity of variance was reasonably met.

Research question two: What is the comparative robusity of ANOVA and Regression statistics when the assumption of homogeneity of variance is not met?

Data used for analysis in this research question were adapted from the same study used in research question one. However, to violate the assumption, the researchers deliberately altered the performance scores of some of the subjects in their categories (sub-groups) to extremely high and extremely low scores thus causing the group variance to be non-homogeneous. To do this, each time a score was increased, the corresponding figure is reduced from another score in the same group. The variances so obtained are presented in Table 4. This created some abnormally less uniform variability of scores among the groups. The data was then subjected to two-way analysis of variance and multiple regression analysis. Results of the analysis are presented in tables 4, 5 and 6.

Table 4: Summary of descriptive statistics showing the sample size and the variance for variable and sub-groups when the assumption of homogeneity is violated

Variable	Groups	n	Mean	SD	SD^2
Attitude	Positive	128	10.28	5.88	34.57
	Moderate	329	8.72	2.65	7.02
	Negative	178	9.31	7.52	56.55
	Total	635	9.20	6.21	38.56
Test Anxiety	Positive	146	9.32	7.13	50.84
	Moderate	382	9.19	4.48	20.07
	Negative	107	9.07	3.95	15.60
	Total	635	9.20	6.57	43.16

Table 5: Summary of two-way ANOVA for the composite effect of attitude and test anxiety on students' mathematics performance as measured by the F-ratio after violating homogeneity of variance assumption

Source of Variation	Sum of Squares	df	Mean Square	F	p-value
Corrected Model	7.568	8	0.946	3.969	.000
Intercept	84855.074	1	84855.074	356032.500	.000
Attitude	3.305	2	1.652	6.933	.001
Anxiety	1.703	2	0.851	3.572	.029
Attitude*Anxiety	0.615	4	0.154	0.645	.631
Error	149.198	626	0.238		
Total	134697.000	635			
Corrected Total	156.765	634			

ISSN: 2645-324X (Print)
ISSN: 2645-3223 (Online)

Table 6: Summary of Multiple Regression analysis for the composite effect of attitude and test anxiety on students' mathematics performance as measured by the F-ratio after violating homogeneity of variance assumption

Source of Variation	Sum of Squares	df	Mean Square	F	p-value
Regression	563.720	2	131.860		
Residual	8703.279	632	13.771	9.575	.000
Total	8966.998	634			

Results in table 4 show that the group variances for positive, moderate, and negative attitudes were 34.57, 7.02 and 56.55 respectively, while 50.84, 20.07, and 15.60 for low moderate and high test anxieties respectively. This implies that variances within groups were not homogeneous, thus the distribution failed to meet the homogeneity of variance assumption. Results of the Two-Way ANOVA and multiple regression analysis show that the F-ratios were 0.645 and 9.049 with p-values of 0.631 and 0.000 respectively. Since the p-value for the Two-Way ANOVA is greater than the p-value for the multiple regression analysis using the same data, it could be concluded that, Multiple Regression statistic is more robust than Analysis of Variance statistic when a distribution does not meet the assumption of homogeneity of variance.

Discussion of the findings

The finding obtained in this study showed that the result of the two-way ANOVA F-ratio was 0.645 and a P-value of 0.631 while that of the multiple regression has an F-ratio of 9.049 with a corresponding P-value of 0.000 respectively. Since the P-value for the two-way ANOVA is greater than the P-value for the multiple regression analysis using the same set of data, it was concluded that multiple regression statistic is more robust than analysis of variance statistic when distribution does not meet the assumption of homogeneity of variance. This result is supported by the findings of Lorenzen and Anderson (2004) who conducted an experimental study using regression analysis with scores that were heteroscedastic using the independent variable X and the dependent variable Y. It was discovered in the study that the common form of such heteroscedasticity in Y is that the variance of Y increases as the mean of Y increases for data with positive X and increases as the mean of Y increases, for data with X and Y; they further discovered that unless the heteroscedasticity of the Y is pronounced, its effect on the P-value will not be severe, the least square estimate will be unbiased.

This result is also in agreement with the findings of Nagin (2005), who discovered that the strength of the regression model is its ability to identify groups or classes of individual with dissimilarities in relationship, observations which do not follow the pattern of other observations is termed outlier. One instance in which robust regression

should be considered is when there is a strong suspicion of heteroscedasticity. Also, this result is supported by the findings of Obambon and Ubi (2019), which states that software packages usually default to homoscedastic models, eventhough such model may be less accurate than heteroscedastic model; one simple approach is to apply the regression least square percentage errors as this will reduce the influence of the larger values of the dependent variable. This result is supported still by the findings of Obambon and Ubi (2019) which stated further that robust regression methods are designed to be not overly affected by this assumption violation.

This result is also in line with the findings of Huberty (2004), who discovered that multiple regression is insensitive to a little deviation from homoscedasticity assumption. Similarly, the researcher discovered that multiple regression statistical tool is a robust statistics which is not unduly affected by deviation from extreme variation in scores within the distribution; in the face of heteroscedasticity in error terms, one will still have unbiased parameter estimate.

The nature of the findings and their replication of earlier study findings are not unexpected. The statistical assumption of homogeneity of variance is long standing and accepted globally by statisticians. Any research into it is to try control of extraneous variables like sample size, location of subjects, personal characteristics of subjects, etc. Violation of the assumption will normally minimize results to indicate that an injury occurred on data collected. In all, the F-ratios obtained in the present study do appear bigger for Regression Analysis than ANOVA. The study thus suggests that Regression statistic is more robust in data handling than Analysis of variance. This conclusion, however, awaits more confirmations by future authors.

Conclusion

This study was an exploration of one of the assumptions of ANOVA and Regression statistics. These two statistical tools share three assumptions one of which is the assumption of homogeneity of variance. It was this homogeneity of variance assumption that was explored to assess the robustness of the F-ratios that may be obtained when the assumption is violated. Two approaches were used in carrying out the assessment. First, the researchers used data that met, to a reasonable extent, the homogeneity assumption and noticed that the p-values obtained from the test of significance using the two statistics were each 0.000. This meant that none was significantly more robust than the other. Secondly, the researchers altered the performance scores of some of the subjects in their categories (sub-groups) to extremely high and extremely low scores thus causing the group variance to be non-homogeneous. These altered scores yielded p-values of 0.631 and 0.000 for ANOVA and Regression respectively. The study concluded that Regression statistic is more

robust than ANOVA statistic with regards to the assumption of homogeneity of variance.

Recommendations

The study thus recommends that, though the assumption of homogeneity of variance is important for both AVOVA and Regression statistics, users of ANOVA statistic in data analysis should take precaution when collecting data to ensure homogeneity of variance. The study also recommends that more research work be carried out in the area using higher samples.

References

Berry, W. D. & Feldman, S. (1985). Multiple Regression in practice. University Paper series on quantitative application in social science series no 07-0501, New Bury Park California.

Box, G. E. P. (1954). Some theorem of quadratic forms applied in the study of analysis of variance problems. Effects of inequality of variance and correlation between errors in the two-way classification. *Annals of Mathematical Statistics, 25, 484-498.*

Bradley, J. V. (1980). Nonrobustness in Z, t and f test at large sample size. *Bulletin of the pschonomics society, 16*(5), *333-336.*

Green, S. B. & Salkind, N.J. (2011). *Using SPSS for windows and macintoshi analyzing and understanding data* (6th ed.). Upper saddle River, New Jersey: Prentice Hall.

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

Kashy, D. A. & Russel, D. W. (2009). *Practice principles and pragmatics*. New York: John Wiley.

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

Lix, L. M. & Wilcox, R. R. (2008). *A generally robust approach for testing hypotheses, and setting confidence interval for effect size*. NJ: Hoboken, John Wiley and Sons.

Lorenzen, T. J. & Anderson, V. L. (2004). *Design of Experiments: A No- name Approach*. New York: Mercel Dekker.

Nagin, D. S. (2005). *Applied Multiple Regression Correlation Analysis for the Behavioural Science* (3rd ed). Mahwah, NJ: Lawrence Erlbaum Associates.

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

Obambon, A. & Ubi, I. O. (2019). Comparative robusity of analysis of variance and regression statistics using the assumption of sample size, homogeneity of variance and normality of the distribution. *Prestige Journal of Counselling Psychology, 2*(1), 12-22. Retrieved online from http://openaccessglobal.com/images/pjcpvol2no1/Comparative_Robusity_of_Analysis_of_Variance_and_Regression_Statistics.pdf

Pedhazur, E. J. (1997). *Multiple Regression in Behavioural Research: Explanation and prediction* (3rd ed). Fort Wort, Tx: Harcourt Brace College Publishers.

Tabachnick, B. G. & Fidell, L. S. (2019). *Using multivariate statistics* (7th ed.). Retrieved from www.pearson.com

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