Abstract

This paper explores ANOVA research where we are looking mainly on the ANOVA is a collection of a statistical model which is used in analyzing the differences among the given group's means and particularly their associated procedures which include the variation of a given set of data. In writing the paper, we need to come up with research which is concerning the ANOVA formulation and hot, it is collinear with other coefficients in the statistical study and also how the paper is based on the scientific study in general. The paper we start with first identifying the background study of the ANOVA, and then we focus on the statement of the problem where we will state the problem and also be able to justify the problem in general. Therefore we will look at the objectives of the study where we will discuss the two types of goals that are the general and also the specific in general. The paper also discusses the research methodology in general, and also the stated examples, and also the formulas which are used to define the ANOVA its relationship with the F- test.

Keywords: ANOVA, models, Unit-treatment additivity and sum of squares.

1. Introduction.

1.1 Background of the study.

Analysis of variance (ANOVA) is a collection of a statistical model which is used in analyzing the differences among the given group's means and particularly their associated procedures which include the variation of a given set of data. It was first developed by statistician and also evolutionary biologist Ronald Fisher. In the ANOVA setting, the variance is observed in a particular variance and is partitioned into components attributable to different sources of variation. In general, ANOVA provides a statistical test of whether and how the different data set is equal and through this, it generalizes the t-test and the f-test which is(F = variance between treatments/ variance within treatment) to more than two different data sets. ANOVA majorly primarily focuses on testing more than three data variables to prove for statistical significance. Originally analysis of variance included the following hypothesis testing which was first performed by Laplace in the 1770s. Later there was a development of the least square method by Laplace, and also Gauss in 1800 which included both observations and thus improved the hypothesis testing. Then on, Ronald Fisher introduced the word variance and in 1918 is when he proposed its formal analysis. W. Which gave the correlation between relatives on the supposition of the Mendelian inheritance. Fisher published the first application of this analysis of variance in 1921. The fixed model of analysis of variance usually applies to the situations in which the experimenter applies more than one treatments to the subject of the given set of data variables. And them observe if the response variable experiences can make any change. By this, there is a chance for the one to estimate the range of the response variables values which the estimation will produce in the whole population according to. It is usually used when the treatments are partially not fixed at all. And mainly occurs when the various factor levels are sampled from a population containing more than one data set. Hence their

levels are classified as random variables. It usually includes experimental elements of both the fixed and also the random models with only appropriately different interpretation and also analysis for the both fixed and also the mixed models. Repeated measures refer to the act of using ANOVA when the same subjects and also subjects are used for each treatment. ANOVA is in general considered to be a particular case of linear regression, which in turn is a particular case of the general linear model. The connection between the linear regression and ANOVA. First, one need to re-order the data so that the Kth observation is about the response o_k and the factors $\mathbf{Z}_{k,b}$ where b ϵ {1, 2,3..., B}. In one-way ANOVA B=1 and in two-way ANOVA B=2 and we assume the $\mathbf{b}_{oth}$ factor has $\mathbf{I}_b$ levels $\pi bb=1Ib$ (Heavey, 2015).

1.2 statement of a problem.

ANOVA are used in many ways in the quest of solving the statistical problems and also analysis of data which is collected from different samples either through sampling or even random selection. The biggest problem of our research is developing methods which can be used to establish the ANOVA and also improve its functionality to give out quality work while analyzing the data.

1.3 justification of the study.

ANOVA is a collection of a statistical model which is used in analyzing the differences among the given group's means and particularly their associated procedures which include the variation of a given set of data. Hence there is a need we develop a better ANOVA understanding to help most of the researchers and also analysis to be able to analyze the data efficiently and also accurately.

1.4 objectives.

These are the things which one or the organization is aimed at or which one is sought in general they refers to goals which the organization is planning to achieve over the stated period. Considering our case there are several objectives for our research paper.

1.4.1 General purpose.

The drive of this study is to develop a better understanding of ANOVA which is usually used for data analysis.

1.4.2 Specific aim.

1. Partitioning of the sum of squares Logic of ANOVA.
2. Develop a randomization-based analysis.
3. Formulate a unit-treatment additivity.
4. Learn the F-test.

2. Research methodology, results and discussion

2.1 Randomization-based analysis

In a randomized controlled experiment, it involves the treatment which is randomly assigned to investigations units; this means following the temporary order. The primary objective of randomization is being directed before the experiment that carries them out. The main purpose of the randomization is that it is used to test the significance of the null hypothesis, which follows the ideas of C. S. Peirce and Ronald Fisher. The design is based on the analysis that was discussed and was developed by Francis J Anscombe. Kempthorne and his students made an

assumption of unit treatment additivity, which commonly discussed in books of Kempthorne (Dressler, 2015).

2.2 **Unit-treatment additivity**

This is the simplest form, the hypothesis of unit-treatment additivity that states what is observed the response of the j from the experiment unit I especially when receiving treatment j can be written as the sum of the group's response **y.** And the treatment assumption is the group additivity that implies that, for each treatment j the jth treatment that has the same effect to on each experiment unity $_{in}$ = yi+ J. The assumption also is critical in that its unit of therapy additivity usually cannot be directed falsified, that as per Cox and Kempthorne.

The results attained can be falsified for example in the in the randomized experiment, the assumption unit-treatment additivity demonstrates that the variance is constant for all treatments. Therefore, for the contraposition that ensures their the vital condition for the unit treatment additivity that is the variance which is constant. The use of unit treatment additivity and randomization that is similar to the design-based inference that is standard in finite-population survey sampling.

2.3 Partitioning of the sum of squares

The ANOVA uses the traditional standardized terminology; it defines the equation of sample variances, Where the divisor is known as the degrees of freedom (DF). The summation is known as the sum of squares (SS), which in turn results are called the mean square (MS) and the squared terms are the deviations from the sample mean. ANOVA approximates three samples variances which include the total differences that are based on all the observations deviations

from the grand mean; the second is the error variance that rests on the variations of treatment which means from appropriate treatments and the treatment difference (Cohen, 2016).

The fundamental technique is the dividing of the total sum of squares SS into the components related to the effects that are used in the model. For example the model for simplified ANOVA ~~that~~ is one type of treatment at different levels.

The number of degrees of freedom DF can also be divided in a similar means which includes one of the components that is commonly an error specifying a chi-squared distribution that describes the associated sum of squares while the same goes for the handling when there is no treatment effect (Cohen, 2016).

2.4 The F-test

The F-test is commonly used in comparing the factors of the total deviation. For example, ~~in one way~~ in one way, or single-factor ANOVA, statistical significance ~~that~~ is tested for by comparing the F-test statistics

$$F=MSTreatments/MSError$$

$$F= (SSTreatments/ (I – 1))/ (SSError/ (nT – I))$$

Where MS is mean square, I= number of treatments and Nt = total number of cases to the F-distribution with I – 1, n T – 1 degrees of freedom

3.0 Conclusion

Randomization is used in multiplying ways in the material, it involves the three roles in the application, a device for eliminating biases. For example from unobserved explanatory variables and the selection effects as the basis for estimating standard errors.

References (no bold, no underline)

Reference:

Welsch, R. E. (2015). The hat matrix in regression and ANOVA. The American Statistician, 32(1), 17-22.

Heavey, E. (2015). The continuing problem of false positives in repeated measures ANOVA in psychophysiology: A multivariate solution. Psychophysiology, 24(4), 479-486.

Cohen, J. (2016). Eta-squared and partial eta-squared in fixed factor ANOVA designs. Educational and psychological measurement.

Dressler, W. W. (2015). The five things you need to know about statistics: Quantification in ethnographic research.

Heavey, E. (2015). Statistics for Nursing: A practical approach.

Crawley, M. J. (2015). Statistics: An Introduction using R.

Sande, A. ., & Byvelds, C. (2015). Statistics for social justice: A structural perspective

indent this. Called a "hanging indent."