ISBN: 978-81-971650-3-0

11. Analysis of Variance and Co-Variance

Dr. Ch. G. Gupta

Asst. Professor of Botany,
Dept. Of Botany, SR and BGNR Govt. Arts and Science College (A).
Khammam, Telangana.

Abstract:

When analyzing data gathered using experimental approaches, statistical procedures such as analysis of variance (ANOVA) and analysis of covariance (ANACOVA) are most appropriate. This has led to their increased usage in the domains of psychology and health and decreased use in sociological studies where survey methods are the norm. an explanation of the statistical techniques for Analysis of Covariance (ANCOVA) and Analysis of Variance (ANOVA), both one-way and two-way.

The chapter includes an explanation of the fundamental statistics (the mean and the variance) that underpin one-way ANOVA, as well as the elements of the ANOVA summary table, estimation techniques, necessary presumptions, planned and post-hoc comparisons, and instances of one-way ANOVA's application in social work research. We will talk about in this essay. Variance and Covariance Analysis..

Keywords:

Variance, Co-Variance, Statistical Techniques, Data Collected, Experimental Methods, Summary Table, Assumptions, Parametric, F-Test, T-Test, Linear Models.

11.1 Introduction:

In probability theory and statistics, the mathematical concepts of variance and covariance are commonly utilized. A covariance is a measure of the directional relationship between two random variables, whereas variance is the spread of a data set around its mean value.

11.2 History of ANOVA:

Before Ronald Fisher invented the analysis of variance method in 1918, statistical analysis was conducted using the t- and z-test techniques that were developed in the 20th century. QIMR Berghofer Medical Research Institute: Genetic Epidemiology, Translational Neurogenomics, Psychiatric Genetics, and Statistical Genetics. "The Correlation Between Relatives on the Supposition of Mendelian Inheritance." [1]

The ANOVA is an extension of the t- and z-tests and is also referred to as the Fisher analysis of variance. The term sprang to notoriety in 1925 with its appearance in Fisher's book, "Statistical Methods for Research Workers." It was first applied to more complicated subjects in experimental psychology.

Analyzing the factors that influence a particular data set starts with the ANOVA test. Following the test, an analyst tests the methodological aspects that quantifiably lead to the inconsistency of the data set again. To produce more data that supports the suggested regression models, the analyst applies the f-test results from the ANOVA test.

Analysis of Variance (ANOVA) is defined as "Separation of variance ascribable to one group of causes from the variance ascribable to other group" by Professor R. A. Fisher. Thus, using this method, the overall variation in the data is separated into two parts: variation resulting from assignable causes (differences across groups) and variation resulting from random causes (differences within groups).

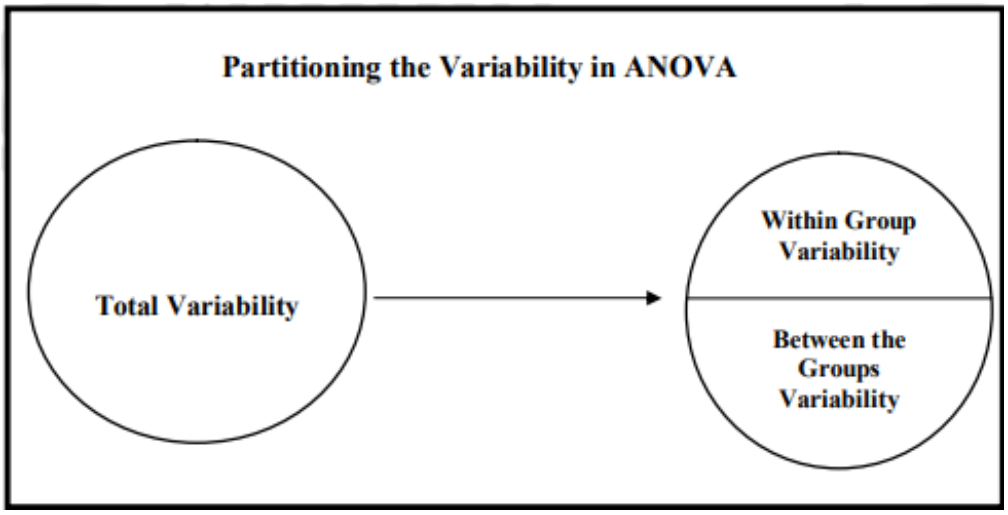

Figure 11.1: Partitioning the Variability in ANOVA [2]

With regard to the estimation problem, the analysis of variance technique resolves the issues of estimating and testing to ascertain whether to infer the presence of a true difference among "treatment" means, among variety means, and under specific circumstances among other means.

The classification of analysis of variance techniques is as follows:

A. Parametric ANOVA.
B. Non-Parametric ANOVA.

If only one response variable is taken into consideration, a parametric ANOVA might be categorized as simply ANOVA. Multivariate analysis of variance is the term used when more than one response variable is being examined (MANOVA).

One-way ANOVA is used when there is only one independent variable taken into account that influences the dependent or response variable. It is referred to as an n-way ANOVA if there are more than one independent variable or explanatory variable, or n (say). The ANOVA is referred to as a two-way categorized ANOVA if n equals 2.

When an experimenter wants to examine the interactions between the explanatory variables, they utilize Factorial ANOVA. When the same participants (experimental units) are utilized for every treatment (level of explanatory variable), repeated measure ANOVA is employed. When there is more than one response variable, multivariate analysis of variance, or MANOVA, is utilized.

Under the normalcy assumption, the F test in ANOVA has been applied. However, in certain instances, the response variable's transformation and distribution do not follow a normal distribution. Non-parametric ANOVA techniques are required in situations like these, when the variable Y_{ij} is nonnormal and the necessary data to make it normal is unknown. Rankings can be used to set the Kruskal Wallis One-way ANOVA. Friedman test is a type of one-way repeated measures ANOVA used to analyze non-parametric data among explanatory factors. Various statistical linear models, such as the Fixed effect, Random effect, and Mixed effect models, can make use of this technique. [3]

The ANOVA technique is useful in any situation where we wish to compare more than two populations, such as when comparing the crop yield from various seed varieties, the fuel efficiency of four different automobiles, the smoking habits of five different university student groups, and so forth.

Under such circumstances, it is usually not desirable to take into account every potential combination of two populations at once, as that would necessitate a large number of tests before an assessment could be made. And even then, certain relationships (especially the interaction effects) might go undiscovered because this would take a significant amount of time and money. As a result, the ANOVA technique is typically used to compare the means of all the populations at the same time and explore any discrepancies.

11.3 Analysis of Variance:

Examination of Variance A statistical method for determining which types of effects are significant and estimating the effects (factors) is the analysis of variance. It involves examining measurements and quantitative findings of observations and experiments that depend on multiple types of effects functioning concurrently. It seems sense that an appropriate theory to assess measurements, observations, and experiment findings would be helpful for designing experiments.

Sir Ronald A. Fisher created the original methodology for analysis of variance, which focuses on selecting appropriate experiments to investigate factors that are likely to have substantial impacts. The primary strategy is an isolating process of the variances related to various parameters or specified sources. This partitioning of the total sum of squares is an acceptable partitioning of the total sample variance mathematically.

The theory that there is no difference between two or more population means can be tested using the analysis of variance (ANOVA) method.

One can use the z-test or t-test to determine the significance of the mean difference between the two samples.

Multiple t-tests can be used to compare the means for each pair when there are more than two means.

11.4 Classification of ANOVA:

One-way ANOVA: One factor examines the factor that differs between the given categories, with n possible values.

Two-way ANOVA: To assess the analysis of the influencing variable, two components are simultaneously included in the study process. [4]

One-Way vs. Two-Way ANOVA:

The effect of a single factor on a single response variable is assessed using a one-way ANOVA. It ascertains if every sample is the same. To find out if there are any statistically significant differences between the means of three or more independent groups, one-way ANOVA is utilized.

The one-way ANOVA is expanded upon by the two-way ANOVA. A one-way relationship has one independent variable influencing one dependent variable. There are two independents in a two-way ANOVA. For instance, a business can compare employee productivity depending on two independent factors, like skill set and salary, using a two-way ANOVA. Its purpose is to examine the simultaneous effects of two elements and observe how they interact. [5]

11.5 Calculation of Variance:

By squaring each number in a data collection and comparing it to the mean, one can find variance by dividing the total squared by the total number of values in the data set.

The established variance formula is:

$\sigma 2 = \Sigma (x - \prod) N$

In the above-given formula,

'x' represents an individual data point.

'$\prod$' represents the data point's mean value.

'N' represents the total data point numbers.

11.6 Analysis of Co-Variance:

An approach that combines linear regression and analysis of variance (ANOVA) is called analysis of covariance (ANCOVA).

A response (the dependent variable) and two or more predictor variables (referred to as covariates) with at least one continuous (quantitative, scaled) and one categorical (nominal, non-scaled) variable are included in the ANCOVA analysis of grouped data.

With distinct groupings having different coefficients of the line, analysts can use the ANCOVA technique to model a variable's response as a linear function of predictor(s). In a nutshell, the overall concept involves adding more variables, or covariates, to the model as a statistical control to decrease error variation, explain variance in the dependent variable, and boost the underlying design's statistical power (sensitivity).

Therefore, it is not the same as the analysis of variance (ANOVA), which is employed to ascertain if differences between test samples could be attributable to random variation. [6]

There are two types of Covariance:

The two variables are related and stay constant when there is a positive covariance.

The variables are said to be negatively related or to move in opposing directions if there is a negative covariance.

The formula for determining covariance

$$COV\ (x, y) = \frac{\sum_{i=1}^{n}(x_i - \tilde{x})\,(y_i - \tilde{y})}{n - 1}$$

In the above-given formula-

'x' represents an independent variable.

'y' represents a dependent variable.

'N' represents the number of the data points.

'$\bar{x}$' represents the mean value of 'x'.

'$\bar{y}$' represents the mean value of 'y'.

Key Concepts and Terms:

Covariate:

An independent variable that is continuous at the interval level. ANOVA must be used in place of ANCOVA if there are no variables, and ANCOVA should be used in place of ANOVA if there are covariates. Control factors are frequently employed, such as covariates. One way to account for starting group differences in math ability, or whatever is being measured in the ANCOVA study, is to utilize the baseline pre-test score as a covariate.

In other words, after controlling for the effects of interval covariates, we use ANCOVA to examine the effects of categorical independents on an interval dependent variable, such as the answer.

(This is comparable to regression, in which the control impact of these independents is reflected by the beta weights of categorical independents entered after interval independents as dummy variables). [7]

F-test:

In the case of a single interval dependent and many (>2) groups established by a categorical independent, the F-test of significance is employed to examine each main and interaction effect. F is the variance within groups divided by the variance between groups.

If the calculated p-value is low, then there are meaningful connections. ANCOVA output often includes adjusted means, which are analyzed if the F-test indicates the presence of significant connections.

Understanding the function of the variables can be gained by comparing the original and adjusted group means.

The adjustment indicates how these k means were changed to account for the covariates for k groups made up of categories of the categorical independents and measured on the dependent variable.

This adjustment is usually made using linear regression of the following kind: Ymean – b*(Xith.mean-Xmean) = Yadj.mean, where X is the covariate, i is one of the k groups, b is the regression coefficient, and Y is the interval dependent.

When Y is normalized, there is no constant. Of course, the equation also contains extra similar X terms for different covariates. [8]

The **t-test** evaluates the significance of a single interval dependent's mean difference between two groups that are each defined by a categorical independent.

Differences between Analysis of Variance and Co-Variance:

An analysis of variance, or ANOVA, is a statistical method for determining whether the means of various groups or levels are equal. To ascertain whether there are noteworthy variations between the means of two or more groups, it is employed.

An analysis of covariance, or ANCOVA, is a statistical method that accounts for the effects of one or more covariates when examining the connection between a dependent variable and one or more independent variables.

It allows you to test for mean differences while accounting for the effects of other factors, so you may think of it as a hybrid of regression analysis and ANOVA. [9]

Table 11.1: Differences between Analysis of Variance and Co-Variance:

	ANOVA	ANCOVA
Purpose	Compare means of two or more groups	Compare means of two or more groups while controlling for the effects of one or more covariates.
Assumptions	Normality and equality of variances	Normality, equality of variances, and linearity between the dependent and independent variables are assumed.
Test statistic	F-value is used as test statistic	F-value or partial F-value are used as test statistic
Null hypothesis	The means of all groups are equal	The means of all groups are equal after adjusting for the effects of the covariates
Alternative hypothesis	Means of at least one group is different	Means of at least one group is different after adjusting for the effects of the covariates
Post-hoc tests	Tukey's HSD, Scheffé's test, etc. can be used for post-hoc analysis	Tukey's HSD, Scheffé's test, etc., or tests specific to the covariates used can be used for post-hoc analysis.

Similarities between ANOVA and ANCOVA:

a. Statistical techniques such as ANOVA and ANCOVA are employed to examine the correlation between an outcome variable and one or more predictor factors.
b. When examining group mean differences or similarities, ANOVA and ANCOVA are also utilized.
c. The F-distribution is used by both ANOVA and ANCOVA to assess the importance of the findings.
d. Data from observational and experimental studies can both be analyzed using ANOVA and ANCOVA.
e. One-way and multi-way designs can be employed with both ANOVA and ANCOVA.
f. The inclusion of interaction effects in ANOVA and ANCOVA refers to the situation in which the impact of one predictor variable on the result variable is contingent upon the level of another predictor variable. [10]

11.7 Conclusion:

Many mathematical concepts make extensive use of variance and covariance. This subject serves as a foundation for derivative problems in statistical mathematics and analytical

studies in addition to mathematics. Finding all or most of the differentiable combinations and outcomes that a random variable can have inside a given range is made incredibly simple with the assistance of variance. One's proficiency with numbers will undoubtedly increase if they can solve, derive, and advance any numerical problem based on covariance and variances. If you understand the basics well, solving problems based on variance and covariance is quite simple and takes less time.

11.8 References:

1. Kshirsagar A.M. (1983). A Course in Linear Models. New York: Marcel Dekker, Inc. [This is a text book, which describes the main aspects of general linear models of statistics including ANOVA (models I and II) as well as analysis of covariance].
2. Silknitter, K. O., Wisnowski, J. W., & Montgomery, D. (1999). Analysis of covariance: A useful technique for analyzing quality improvement experiments. Quality and Reliability Engineering International, 15(4), 303-316.
3. R. A. Fisher, Statistical Methods for Research Workers. Edinburgh: Oliver & Boyd, 1925.
4. R. C. Littel, W. W. Stroup, and R. J. Freund, SAS for Linear Models. Cary, NC: SAS Institute Inc., 2002.
5. Wilke, M., 'Health reports in vogue again for drug advertisers', Advertising Age 68(33) (18 August 1997), 31; Iyer, E.S., 'The influence of verbal content and relative newness on the effectiveness of comparative advertising', Journal of Advertising 17, (1988), 15–21.
6. Tacq, J., Multivariate Analysis Techniques in Social Science Research (Thousand Oaks, CA: Sage, 1997); Daniel, W.W. and Terrell, J.C., Business Statistics, 7th edn (Boston, MA: Houghton Mifflin, 1995).
7. A.S.AL-Mouel, and J.M. Jassim, "twoway multivariate Repeated measurements Analysis of variance model" J. Basrah Researches (Sciences) Vol.32. par 1. -17- 31, (2006).
8. G.W. Snedecor, and W. G. Cochran, "Statistical Methods ", edition, Ames, Iowa State University press, Iowa, (1967).
9. Field, A. (2009) Discovering Statistics using SPSS (And sex and drugs and rock 'n' roll), 3rd ed., London: SAGE.
10. Tabachnick, B. G.; Fidell, L. S. (2007). Using Multivariate Statistics (5th ed.). Boston: Pearson Education.