Bio-instrumentation & Biostatistics

Dr. Bibha Verma
PG Department of Zoology
Maharaja College Ara

PG Sem II, CC-6, Unit -IV

4.3 (c) Analysis of Variance (ANOVA)

Analysis of variance (**ANOVA**) is a collection of statistical models and their associated estimation procedures (such as the "variation" among and between groups) used to analyze the differences among group means in a sample. ANOVA was developed by the statistician Ronald Fisher. The ANOVA is based on the law of total variance, where the observed variance in a particular variable is partitioned into components attributable to different sources of variation. In its simplest form, ANOVA provides a statistical test of whether two or more population means are equal, and therefore generalizes the *t*-test beyond two means.

One of the attributes of ANOVA that ensured its early popularity was computational elegance. The structure of the additive model allows solution for the additive coefficients by simple algebra rather than by matrix calculations. In the era of mechanical calculators this simplicity was critical. The determination of statistical significance also required access to tables of the F function which were supplied by early statistics texts.

Analysts use the ANOVA test to determine the influence that independent variables have on the dependent variable in a regression study.

The t- and z-test methods developed in the 20th century were used for statistical analysis until 1918, when Ronald Fisher created the analysis of variance method. ANOVA is also called the Fisher analysis of variance, and it is the extension of the t- and z-tests. The term became well-known in 1925, after appearing in Fisher's book, "Statistical Methods for Research Workers." It was employed in experimental psychology and later expanded to subjects that were more complex.

The Formula for ANOVA is:

$$F = \frac{MST}{MSE}$$

where:

$F = $ ANOVA coefficient

$MST = $ Mean sum of squares due to treatment

$MSE = $ Mean sum of squares due to error

What Does the Analysis of Variance Reveal?

The ANOVA test is the initial step in analyzing factors that affect a given data set. Once the test is finished, an analyst performs additional testing on the methodical factors that measurably contribute to the data set's inconsistency. The analyst utilizes the ANOVA test results in an f-test to generate additional data that aligns with the proposed regression models.

The ANOVA test allows a comparison of more than two groups at the same time to determine whether a relationship exists between them. The result of the ANOVA formula, the F statistic (also called the F-ratio), allows for the analysis of multiple groups of data to determine the variability between samples and within samples.

If no real difference exists between the tested groups, which is called the null hypothesis, the result of the ANOVA's F-ratio statistic will be close to 1. Fluctuations in its sampling will likely follow the Fisher F distribution. This is actually a group of distribution functions, with two characteristic numbers, called the numerator degrees of freedom and the denominator degrees of freedom.

KEY TAKEAWAYS

Analysis of variance, or ANOVA, is a statistical method that separates observed variance data into different components to use for additional tests.

A one-way ANOVA is used for three or more groups of data, to gain information about the relationship between the dependent and independent variables.

If no true variance exists between the groups, the ANOVA's F-ratio should equal close to 1.

Example of How to Use ANOVA

A researcher might, for example, test students from multiple colleges to see if students from one of the colleges consistently outperform students from the other colleges. In a business application, an R&D researcher might test two different processes of creating a product to see if one process is better than the other in terms of cost efficiency.

The type of ANOVA test used depends on a number of factors. It is applied when data needs to be experimental. Analysis of variance is employed if there is no access to statistical software resulting in computing ANOVA by hand. It is simple to use and best suited for small samples. With many experimental designs, the sample sizes have to be the same for the various factor level combinations.

ANOVA is helpful for testing three or more variables. It is similar to multiple two-sample t-tests. However, it results in fewer type I errors and is appropriate for a range of issues. ANOVA groups differences by comparing the means of each group and includes spreading out the variance into diverse sources. It is employed with subjects, test groups, between groups and within groups.

One-Way ANOVA Versus Two-Way ANOVA

There are two types of ANOVA: one-way (or unidirectional) and two-way. One-way or two-way refers to the number of independent variables in your analysis of variance test. A one-way ANOVA evaluates the impact of a sole factor on a sole response variable. It determines whether all the samples are the same. The one-way ANOVA is used to determine whether there are any statistically significant differences between the means of three or more independent (unrelated) groups.

A two-way ANOVA is an extension of the one-way ANOVA. With a one-way, you have one independent variable affecting a dependent variable. With a two-way ANOVA, there are two independents. For example, a two-way ANOVA allows a company to compare worker productivity based on two independent variables, such as salary and skill set. It is utilized to observe the interaction between the two factors and tests the effect of two factors at the same time.

Characteristics

ANOVA is used in the analysis of comparative experiments, those in which only the difference in outcomes is of interest. The statistical significance of the experiment is determined by a ratio of two variances. This ratio is independent of several possible alterations to the experimental observations: Adding a constant to all observations does not alter significance. Multiplying all observations by a constant does not alter significance. So ANOVA statistical significance result is independent of constant bias and scaling errors as well as the units used in expressing observations. In the era of mechanical calculation it was common to subtract a constant from all observations (when equivalent to dropping leading digits) to simplify data entry.[38][39] This is an example of data coding.

References:

https://sphweb.bumc.bu.edu/otlt/MPH-Modules/BS/BS704_HypothesisTesting-ANOVA/BS704_HypothesisTesting-Anova_print.html

https://www.ebpi.uzh.ch/dam/jcr:ffffffff-c1f2-5119-0000-000078458c66/slides-anova.pdf

https://www.investopedia.com/terms/a/anova.asp

http://onlinestatbook.com/2/analysis_of_variance/intro.html

https://en.wikipedia.org/wiki/Analysis_of_variance