One-Way ANOVA

Explorable.com227.5K reads

A One-Way ANOVA (Analysis of Variance) is a statistical technique by which we can test if three or more means are equal. It tests if the value of a single variable differs significantly among three or more levels of a factor.

We can say we have a framework for one-way ANOVA when we have a single factor with three or more levels and multiple observations at each level.

In this kind of layout, we can calculate the mean of the observations within each level of our factor.

The concepts of factor, levels and multiple observations at each level can be best understood by an example.

Factor and Levels - An Example

Let us suppose that the Human Resources Department of a company desires to know if occupational stress varies according to age.

The variable of interest is therefore occupational stress as measured by a scale.

The factor being studied is age. There is just one factor (age) and hence a situation appropriate for one-way ANOVA.

Further suppose that the employees have been classified into three groups (levels):

- less than 40
- 40 to 55

- above 55

These three groups are the levels of factor age - there are three levels here. With this design, we shall have multiple observations in the form of scores on Occupational Stress from a number of employees belonging to the three levels of factor age. We are interested to know whether all the levels i.e. age groups have equal stress on the average.

Non-significance of the test statistic (F-statistic) associated with this technique would imply that age has no effect on stress experienced by employees in their respective occupations. On the other hand, significance would imply that stress afflicts different age groups differently.

Hypothesis Testing

Formally, the null hypothesis [1] to be tested [2] is of the form:

> H_0: All the age groups have equal stress on the average or $?_1 = ?_2 = ?_3$, where $?_1$, $?_2$, $?_3$ are mean stress scores for the three age groups.

The alternative hypothesis [3] is:

> H_1: The mean stress of at least one age group is significantly different.

One-way Anova and T-Test

The one-way ANOVA [4] is an extension of the independent two-sample t-test [5].

In the above example, if we considered only two age groups, say below 40 and above 40, then the independent samples t-test would have been enough although application of ANOVA would have also produced the same result.

In the example considered above, there were three age groups and hence it was necessary to use one-way ANOVA.

Often the interest is on acceptance or rejection of the null hypothesis. If it is rejected, this technique will not identify the level which is significantly different [6]. One has to perform t-tests [7] for this purpose.

This implies that if there exists difference between the means, we would have to carry out ${}^3C_2$ independent t-tests in order to locate the level which is significantly different. It would be ${}^kC_2$ number of t-tests in the general one-way ANOVA design with k levels.

Advantages

One of the principle advantages of this technique is that the number of observations need not be the same in each group.

Additionally, layout of the design and statistical analysis is simple.

Assumptions

For the validity of the results [8], some assumptions have been checked to hold before the technique is

applied. These are:

- Each level of the factor is applied to a sample. The population from which the sample was obtained must be normally distributed [9].
- The samples must be independent.
- The variances [10] of the population must be equal.

Replication and Randomization

In general, ANOVA experiments need to satisfy three principles - replication, randomization and local control [11].

Out of these three, only replication and randomization have to be satisfied while designing and implementing any one-way ANOVA experiment [12].

Replication refers to the application of each individual level of the factor to multiple subjects. In the above example, in order to apply the principle of replication, we had obtained occupational stress scores from more than one employee in each level (age group).

Randomization [13] refers to the random allocation [14] of the experimental units. In our example, employees were selected randomly for each of the age groups.

Links
[1] https://verify.explorable.com/null-hypothesis
[2] https://verify.explorable.com/hypothesis-testing
[3] https://verify.explorable.com/research-hypothesis
[4] http://www4.uwsp.edu/psych/stat/12/anova-1w.htm
[5] https://verify.explorable.com/independent-two-sample-t-test
[6] https://verify.explorable.com/significance-test-2
[7] https://verify.explorable.com/students-t-test-2
[8] https://verify.explorable.com/internal-validity
[9] https://verify.explorable.com/normal-probability-distribution
[10] https://verify.explorable.com/statistical-variance
[11] https://verify.explorable.com/controlled-variables
[12] https://verify.explorable.com/experimental-research
[13] https://verify.explorable.com/randomization
[14] https://verify.explorable.com/simple-random-sampling