.ANOVA

Analysis of Variance is a method of testing the equality of three or more population means by analyzing sample variance.

One-Way ANOVA

The one-way ANOVA is used to compare three or more population means when there is one factor of interest.

Requirements

- The populations have distributions that are approximately normal.
- The populations have the same variance.
- The samples are simple random samples of quantitative data.
- The samples are independent of each other.
- The different samples are from populations that are categorized in only one. way

One-Way ANOVA is a hypothesis test. There are seven steps for a hypothesis test.

Example

A professor at a local University believes there is a relationship between head size and the major of the students in her biostatistics classes. She takes a random sample of 20 students from each of three classes and records their major and head circumference. The data are shown in the following table.

Step 1: State the null hypothesis.

Mean 1 equals mean 2 equals mean 3 equals mean 4.

Step 2: State the Alternative hypothesis.

At least one mean is different.

Step 3: State the Level of Significance.

The level of significance is 0.05.

Step 4: State the test statistic.

$$F = \frac{\text{variance between samples}}{\text{variance within samples}}$$

The test statistic follows the F distribution which has two degrees of freedom, one for the numerator and one for the denominator.

The calculations for the test statistic are complicated, so a software program is generally used for the calculations. We will be using Microsoft Excel for this example.

Step 5: Calculate

The calculations are done in Microsoft Excel using the data analysis toolpak. Enter the data into the spread sheet as shown here. Click on data and the data analysis tookpak button is on the right.

When you click on the button a dialogue box appears.

Choose ANOVA One Factor. Then another dialogue box appears.

Input range is where the data is in the table. Be sure to put a check in the box for labels in the first row and then pick a cell for the results to be displayed in.

We need to be able to interpret these results.

Step 6: Statistical Conclusion

The summary output from Excel gives the observed test statistic as well as the critical value.

The null hypothesis is rejected if the observed test statistic is greater than the critical value.

Since F equals 0.816124 and is less than 2.769431 which is the F critical value, then we fail to reject the null hypothesis.

Step 7: Experimental Conclusion

There is not sufficient evidence to indicate that there is a statistically significant difference between the mean head size of different majors in the biostatistics classes of this professor at a level of significance of 0.05.

Two-Way ANOVA

There is an interaction between two factors if the effect of one of the factors changes for different categories of the other factor.

Requirements

- For each cell, the sample values come from a population with a distribution that is normal.
- For each cell, the sample values come from population having the same variance.
- The samples are simple random samples.
- The samples are independent of each other.
- The sample values are categorized two ways.
- All of the cells have the same number of sample values.

Example

A professor at a local University believes there is a relationship between head size, the major of the students, and the gender of students in her biostatistics classes. She takes a random sample from her three classes. The data is in the following table.

Notice that the sample size for each set of categories is the same. (i.e., female and premed had 4 data values as does male and premed etc…)

A two-way ANOVA essentially does three different hypothesis tests. The first test is for interaction effect, then effect from each of the two factors. All the test statistics are calculated at once in Microsoft Excel (data analysis toolpak).

Start by entering the data in Excel as follows then select the data analysis toolpak under the data tab.

Then select ANOVA: Two factor with replication (since there is more than one sample per category duo)

Select the data that is to be used and the number of data values in each category due. That is 4 for this example. The default alpha is 0.05. Then select the cell to have the results output to.

The results give a lot of information but we are concerned with the last table displayed.

The ANOVA table gives the F test statistic for each of the three tests for consideration.

Start by looking at the interaction test statistic.

We could look at either the critical value or the p-value to determine if there is an effect due to interaction. The p-value is the probability of a more extreme value than the observed test statistic. If the p-value is less than the critical value 0.05, then reject the null hypothesis that there is no effect from the interaction of the factors. Since $p-value = 0.087227 > 0.05 = \alpha$, then we fail to reject the null hypothesis. So there is sufficient evidence to indicate an effect due to the interaction of gender and major.

We can look at the effect of gender and major individually also.

We can again look at the p-value. If p-value is less than the level of significance then we reject the null hypothesis that there is no effect from the factor of major.

Since $p-value = 0.173639 > 0.05 = \alpha$, then we fail to reject the null hypothesis that there is no effect due to the factor of major.

We can again look at the p-value. If p-value is less than the level of significance then we reject the null hypothesis that there is no effect from the factor of gender.

Since $p-value = 0.012084 < 0.05 = \alpha$, then we reject the null hypothesis that there is no effect due to the factor of gender.

Putting all the statistical conclusions together we can see that there is no effect from the interaction of gender and major, on the head circumference and there is no effect on head circumference due to major but there is an effect due to gender on head circumference at a statistically significant level of 0.05.