

Chapter 12 - Independence Tests and ANOVA

ANOVA, Analysis of Variance, is a test to determine if three or more different methods or treatments have the same effect on a population. For example, ANOVA testing might be used to determine if three different teaching methods produce the same test scores with a group of students. The measured output must be some type of group average such as average test score per group or average sales per group. ANOVA testing might also be used to determine if different combinations of product pricing and promotion have different effects in different markets.

Contents

- Basic Explanation of ANOVA
- ANOVA Tests the Null Hypothesis - That Nothing Is Different Between Groups
- Overview of ANOVA in Excel
 - Single Factor ANOVA
 - Two-Factor ANOVA Without Replication
 - Two-Factor ANOVA With Replication
- ANOVA:Single Factor Analysis
 - Problem: 3 Sales Closing Methods and Single Factor ANOVA
 - Problem Solving Steps
 - Analyze the Output
- ANOVA: Two-Factor Without Replication
 - Problem: 3 Sales Closing Methods, 5 Salespeople, and Two-Factor ANOVA Without Replication
 - Problem Solving Steps
 - Analyze the Output
- ANOVA:Two Factor With Replication
 - Problem: 3 Ad Headlines, 3 Ad Texts, their Interaction, and Two-Factor ANOVA With Replication
 - Problem Solving Steps
 - Analyze the Output
- ANOVA: Single Factor Analysis Calculated by Hand
 - Problem: 3 Closing Methods and Single Factor ANOVA Calculated By Hand

Basic Explanation of ANOVA

In summary, ANOVA testing is used to judge whether three or more groups have the same mean (for example, same test scores) after each group has had a **different treatment** applied to it (for example, a different teaching method applied to each group).

ANOVA Tests the Null Hypothesis - That Nothing Is Different Between Groups

The basic test of ANOVA is the Null Hypothesis that the different methods had no effect on the outcome that is being measured. Using the teaching method example, the Null Hypothesis in this case would be that the different teaching methods had no effect on the average test scores of student groups to which different treatments (teaching methods) were applied.

The Null Hypotheses, which is being tested, states that the average test score from each of the three groups or populations to which a different teaching method was applied should be the same.

Overview of ANOVA in Excel

The hand calculations for ANOVA are very tedious. Excel has a built-in ANOVA function that does a great job. The problems below are completed using Excel's built-in ANOVA function. The simplest example, known as Single Factor ANOVA, is completed using both Excel's built-in ANOVA function and also hand-calculated at the very end of this course module.

The ANOVA tests in Excel can be accessed by the drop-down menu:

Tools / Data Analysis.

Inside Excel's Data Analysis menu, there are three types of ANOVA analysis available:

- 1) Single Factor ANOVA**
- 2) Two-Factor ANOVA Without Replication**
- 3) Two-Factor ANOVA with Replication**

Each ANOVA test type is explained below:

Single Factor ANOVA

Single Factor ANOVA tests the effect of just one factor, in this case, the teaching method, on the measured outputs. The measured outputs are the mean test scores for the groups that had the different teaching methods applied to them. The Null Hypothesis for this one factor states that varying that factor has no effect on the outcome.

Two-Factor ANOVA Without Replication

Two-Factor ANOVA Without Replication - Allows testing of the original factor plus one other factor. For example, in addition to testing teaching methods, you could also test an additional factor, such as whether differences in teaching ability caused additional variation in the outcome of test average scores. Each factor has a Null Hypothesis which states that varying that factor had no effect on the outcome.

Two-Factor ANOVA With Replication

Two-Factor ANOVA With Replication allows for testing both factors as above. This method also allows us to test the effect of interaction between the factors upon the measured outcome. The test is replicated in two places. This allows for analysis of whether the interaction between the two factors has an effect on the measured outcome. The Null Hypothesis for this interaction test states that varying the interaction between the two factors has no effect on the measured outcome. Each of the other two factors being tested also has its own Null Hypothesis.

Examples of each type of ANOVA performed with Excel follow:

ANOVA: Single Factor Analysis

The Hand Calculation of this problem's ANOVA is performed at the end of this course module.

Problem: 3 Sales Closing Methods and Single Factor ANOVA

Problem: Three different sale closing methods were used. Three groups of four salespeople were randomly chosen. Each group was instructed to use only one of the closing methods for all of their sales. Sales totals of each salesperson over the next two weeks were collected. Determine with a 95% level of certainty whether there is a difference in the effectiveness of the closing methods. Following are sales results for all 12 salespeople on the next page:

| Sales Group 1 | Closing Method 1 |
|---------------|------------------|
| Salesperson 1 | 16 |
| Salesperson 2 | 21 |
| Salesperson 3 | 18 |
| Salesperson 4 | 13 |

| Sales Group 2 | Closing Method 2 |
|---------------|------------------|
| Salesperson 5 | 19 |
| Salesperson 6 | 20 |
| Salesperson 7 | 21 |
| Salesperson 8 | 20 |

| Sales Group 3 | Closing Method 3 |
|----------------|------------------|
| Salesperson 9 | 24 |
| Salesperson 10 | 21 |
| Salesperson 11 | 22 |
| Salesperson 12 | 25 |

The above information must be arranged as below for input into Excel:

| Sales Group ----> | Group 1 | Group 2 | Group 3 |
|-------------------|----------|----------|----------|
| Treatment ----> | Method 1 | Method 2 | Method 3 |
| Salesperson A | 16 | 19 | 24 |
| Salesperson B | 21 | 20 | 21 |
| Salesperson C | 18 | 21 | 22 |
| Salesperson D | 13 | 20 | 25 |

Problem Solving Steps

This is a **Single Factor ANOVA test** because we are testing only whether different variations of a single factor (Closing Method) have an effect on measured outcome (sales of each salesperson) using a different method. Nothing else is entered into the test that might have an effect on the measured outcome. **The abilities of all salespeople are assumed to be similar. Each individual salesperson will use only one of the closing methods.**

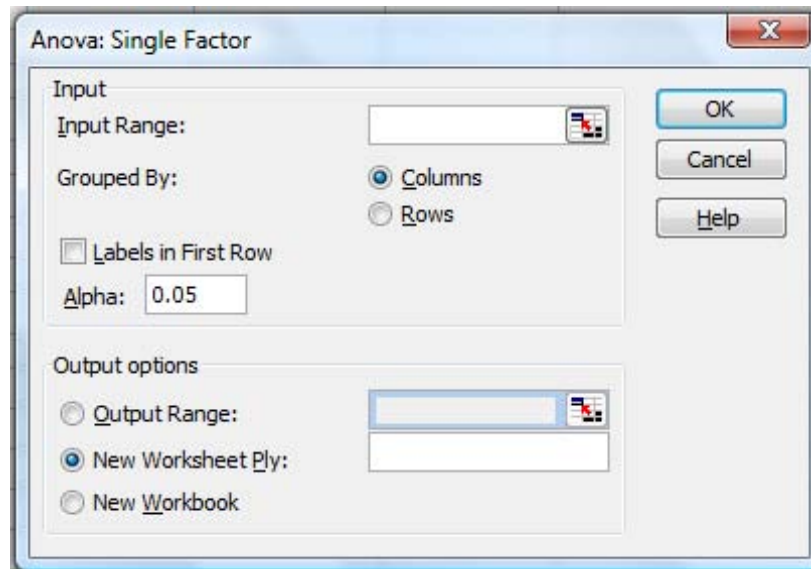
The **Null Hypothesis** for this test states that the closing methods used will have no effect on the measured output (sales).

Level of Certainty = 95% = $1 - \alpha$

Level of Significance = Alpha = $\alpha = 0.05$

Excel Instructions:

Tools / Data Analysis / Anova: Single Factor



You will be required to provide the following 5 inputs:

- 1) Input Range:** Highlight the data and labels as is highlighted in yellow previously. Note that your data must be initially arranged in columns as above or similarly in rows. Note that only the Method labels and not the Salesperson labels are highlighted because Single Factor ANOVA tests the effects of only one factor, the training method. If you also wanted to test the additional factor of whether the individual abilities of each salesperson were also a factor in the outcome, the salesperson labels would be highlighted and Two-Factor ANOVA would be used. The next example in this module performs this test.

- 2) Grouped By:** Select Columns or Rows, depending on how your data is arranged. In this case, the data is arranged in columns

- 3) Data Labeling:** Select whether your data is labeled in the first column (if the data is grouped in rows) or labeled in the first row (if your data is grouped in rows). In this case, the data is labeled in the first row.

- 4) Alpha:** You will type in the level of significance here. The level of significance (alpha) is the max level of error you are willing to tolerate. If you can only tolerate a possibility of 1% of error, the level of significance (alpha) here would be 0.01. Prefilled in this input box is 0.05. This represents a maximum error of 5%. Another way of looking at this would be to say that 5% maximum error represents being at least 95% certain of the correctness of the output.

- 5) Output:** Here you would enter (or select) the upper-left corner of where you want the output printed on your worksheet. You will need to make sure that the place

where you have specified the output to be printed has at least 7 free columns and 15 free rows.

Hit "OK" and Excel will perform the Single Factor ANOVA analysis and output the following result where you have specified:

| Anova: Single Factor | | | | | | |
|-----------------------------|--------------|------------|----------------|-----------------|--------------------|---------------|
| SUMMARY | | | | | | |
| <i>Groups</i> | <i>Count</i> | <i>Sum</i> | <i>Average</i> | <i>Variance</i> | | |
| Method 1 | 4 | 68 | 17 | 11.33333333 | | |
| Method 2 | 4 | 80 | 20 | 0.66666667 | | |
| Method 3 | 4 | 92 | 23 | 3.33333333 | | |
| ANOVA | | | | | | |
| <i>Source of Variation</i> | <i>SS</i> | <i>df</i> | <i>MS</i> | <i>F</i> | <i>P-value</i> | <i>F crit</i> |
| Between Groups | 72 | 2 | 36 | 7.043478261 | 0.014419201 | 4.256494729 |
| Within Groups | 46 | 9 | 5.111111111 | | | |
| Total | 118 | 11 | | | | |

P-value = 0.014419201
Alpha = 0.05

Analyze the Output

The P-value

Without going into too much statistical detail, the cell in the output entitled P-value will tell you whether the three different closing methods have an effect on the measured output (sales of salespeople using the different training methods). Highlighted in Yellow on the Excel output is the P-value for Between Groups variation.

General Rule:

If P-value < alpha ----> The different treatments affected the output

If P-value > alpha ----> The different treatments did not affect the output

If this P-value (0.01441920) is less than the level of significance (Alpha = 0.05) that was specified as one of the Excel inputs, we can say with 95% certainty (having a maximum chance of error of 5%) that the closing methods produced different effects on the measured output (sales)

In this case, the P-Value for between Group Variation (0.01449201) is less the Alpha (0.05) so we can say that the closing method employed did make a difference in the sales with no more than a 5% chance of being wrong. The alpha can be thought of as the maximum chance of being wrong.

Hand calculations of this entire problem are shown at the end of this module. After taking a quick look at that, you will probably be happy to let Excel do all of the calculation work. The calculation of the P value and the associated F Value are time-consuming and rigorous. No need to do them by hand if Excel will do it for you.

ANOVA: Two Factor Without Replication

Problem: 3 Sales Closing Methods, 5 Salespeople, and Two-Factor ANOVA Without Replication

Problem: A sales manager was experimenting to determine the effectiveness of 3 different types of closing methods. He instructed each of his 5 salespeople to try all 3 methods. Each salesperson would use only one of the closing methods for one full week. Every week each salesperson had to switch to another of the 3 closing methods that he had not previously used. At the end of 3 weeks, the sales results were provided below. Determine with 99% certainty whether the closing methods used made a difference in the sales results. Also determine with 99% certainty whether the individual abilities of each salesperson made a difference in the sales results.

| | Method 1 | Method 2 | Method 3 |
|---------------|----------|----------|----------|
| Salesperson 1 | 51 | 57 | 72 |
| Salesperson 2 | 109 | 112 | 117 |
| Salesperson 3 | 47 | 43 | 51 |
| Salesperson 4 | 98 | 98 | 107 |
| Salesperson 5 | 70 | 69 | 77 |

Note that this is different than the test run for Single factor ANOVA above. The single factor ANOVA test used 12 salespeople who each used only one closing method. This two-factor ANOVA uses only five salespeople. Each salesperson here must use each of the three closing methods for one week.

Problem Solving Steps

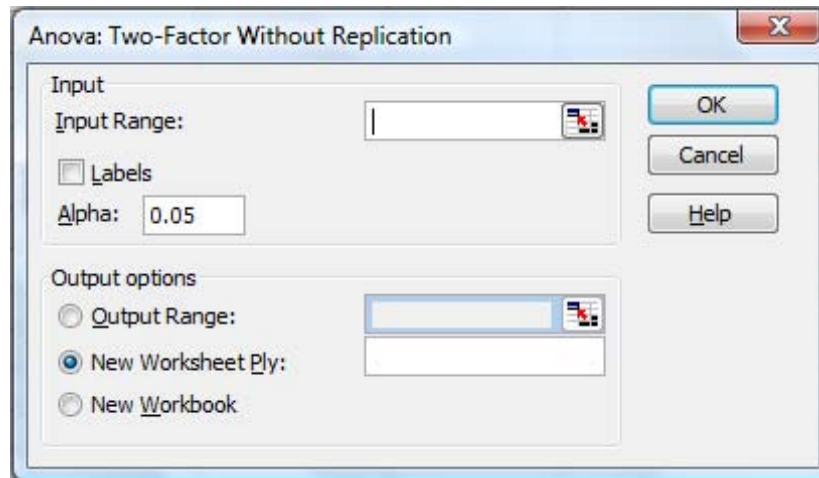
This is a **Two Factor ANOVA test** because there are two variables that could effect the measured outcome (sales). The two given variables are the selling ability of each salesperson and the closing method used. Nothing else is entered into this test that might have an effect on the measured output (sales).

This is a **test Without Replication** because there is no way to test whether the two variables interact with each other to affect the measured outcome. An example of a test With Replication is in the ANOVA example that follows this one.

The two **Null Hypotheses** for this test are as follows. Each of the two variables (the salesperson's ability and the closing method used) have their own Null Hypothesis. The Null Hypothesis for salesperson's ability states that there is not enough variation in the abilities of the salespeople to influence sales. In other words, the measured output (sales) will not be affected by which salesperson does the selling. The Null Hypothesis for the closing method used states that the choice of closing method will have no affect on sales.

Excel Instructions:

Tools / Data Analysis / Anova: Two-Factor Without Replication



You will be required to provide the following 5 inputs:

- 1) Input Range:** Highlight the data and labels as is highlighted in Yellow previously. Note that your data must be initially arranged in columns as above or similarly in rows. Note that both the Method labels and the Salesperson labels are highlighted because they are the two factors being analyzed.
- 2) Grouped By:** Select Columns or Rows, depending on how your data is arranged. In this case the data is arranged in columns
- 3) Data Labeling:** Select whether your data is labeled in the first column (if the data is grouped in rows) or labeled in the first row (if your data is grouped in columns). In this case, the data is labeled in the first row.

4) Alpha: You will type in the level of significance here. The level of significance (alpha) is the max level of error you are willing to tolerate. If you can only tolerate a possibility of 1% of error, the level of significance (alpha) here would be 0.01. Prefilled in this input box is 0.05. This represents a maximum error of 5%. Another way of looking at this would be to say that 5% maximum error represents being at least 95% certain of the correctness of the output. Type in 0.01 for Alpha.

5) Output: Here you would enter (or select) the upper-left corner of where you want the output printed on your worksheet. You will need to make sure that the place that you have specified the output to be printed has at least 7 free columns and 20 free rows.

Hit "OK" and Excel will perform the Two-Factor ANOVA Without Replication analysis and output the following result on the next page to where you have specified:

| Anova: Two-Factor Without Replication | | | | | | |
|--|--------------|------------|----------------|-----------------|----------------|---------------|
| <i>SUMMARY</i> | <i>Count</i> | <i>Sum</i> | <i>Average</i> | <i>Variance</i> | | |
| Salesperson 1 | 3 | 180 | 60 | 117 | | |
| Salesperson 2 | 3 | 338 | 112.6666667 | 16.33333333 | | |
| Salesperson 3 | 3 | 141 | 47 | 16 | | |
| Salesperson 4 | 3 | 303 | 101 | 27 | | |
| Salesperson 5 | 3 | 216 | 72 | 19 | | |
| Method 1 | 5 | 375 | 75 | 767.5 | | |
| Method 2 | 5 | 379 | 75.8 | 819.7 | | |
| Method 3 | 5 | 424 | 84.8 | 724.2 | | |
| ANOVA | | | | | | |
| <i>Source of Variation</i> | <i>SS</i> | <i>df</i> | <i>MS</i> | <i>F</i> | <i>P-value</i> | <i>F crit</i> |
| Rows | 9151.066667 | 4 | 2287.766667 | 193.6050776 | 5.42004E-08 | 7.006076623 |
| Columns | 296.1333333 | 2 | 148.0666667 | 12.5303244 | 0.003428581 | 8.649110641 |
| Error | 94.53333333 | 8 | 11.81666667 | | | |
| Total | 9541.733333 | 14 | | | | |

Alpha = 0.01

The p-Value for the Rows (5.42004E-08) is much less than the level of significance (0.05) so the individual selling ability of each salesperson made a difference in sales results.

The p-Value for columns (0.003428581) is much less than the level of significance (0.05) so the closing method used made a difference in sales results

Analyze the Output

The P-value

Without going into too much statistical detail, the cells in the output entitled P-value will tell you whether the three different closing methods and individual selling abilities have an effect on the measured output of sales

General Rule:

If P-value < alpha ----> The different treatments affected the output

If P-value > alpha ----> The different treatments did not affect the output

The data interpretation for Two-Factor ANOVA is very similar to Single Factor ANOVA. The only difference is that now two variables (choice of salesperson and choice of closing method) are being tested to determine whether either had an effect on the measured output (sales). If the P-value associated with a variable is less than alpha (0.01), then we can state that this variable had an effect on sales.

Data for each salesperson is arranged in rows. The P-value associated with choice of salesperson is highlighted in Blue. This is the P-value for the source of variation in rows. This P-value is 5.42004×10^{-8} and is much less than Alpha (0.01) so we can state that the choice of salesperson almost certainly had an effect on sales.

Data for each closing method is arranged in columns. The P-value associated with choice of closing method is highlighted in Yellow. This is the P-value for the source of variation in columns. This P-value is .003428581 and is much less than alpha (0.01) so we can state that the choice of closing method almost certainly had an effect on sales.

Anova: Two Factor With Replication

Problem: 3 Headlines, 3 Ad Texts, their Interaction, and Two-Factor ANOVA With Replication

Problem: A company was conducting pay-per-click advertising on both Google and Yahoo. The Internet marketing manager was testing headlines and ad text for ads he had written for one product. The manager created 3 headlines and 3 sets of ad text. Altogether there were 9 possible ad combinations of headline and ad text. The manager then ran all 9 ads for an equal number of times on both the Google and Yahoo paid search networks. Below is the resulting Click-Through Rate (CTR) for each combination of Headline / Ad Text / Search Engine after the ads ran for 2 weeks. Use a 95% confidence level to determine whether 1) Ad Text, 2) Headline, and 3) Interaction between Ad Text and Headline had an effect on Click-Through Rate.

| | Headline 1 | Headline 2 | Headline 3 | |
|-----------|------------|------------|------------|--------|
| Ad Text 1 | 2.80 | 2.04 | 1.58 | Google |
| | 2.73 | 1.33 | 1.26 | Yahoo |
| Ad Text 2 | 3.29 | 1.50 | 1.00 | Google |
| | 2.68 | 1.40 | 1.82 | Yahoo |
| Ad Text 3 | 2.54 | 3.15 | 1.92 | Google |
| | 2.59 | 2.88 | 1.33 | Yahoo |

Problem Solving Steps

This is a **Two Factor ANOVA** test because there are two variables that could effect the measured outcome (Click-Through Rate on each search engine) The two given variables are the Ad Text and Headline.

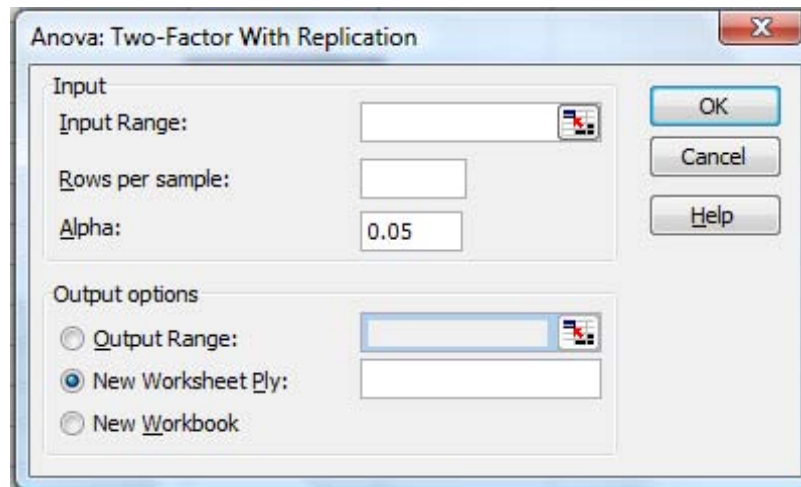
This is a **test With Replication** because the same tests are performed in different places to determine if there is interaction between the variables (Ad Text and Headline) that might affect the measured output (Click-Through Rate on each search engine).

Null Hypotheses for all variables - Each of the two variables (Ad Text and Headline) has its own Null Hypothesis. The Null Hypothesis for Ad Text states that there is no variation in Click-Through Rate on each search engine as a result of different Ad Text. In other words, the measured output (CTR in each search engine) will not be affected by the Ad Text used in the ads. The Null Hypothesis for Headline states that the choice of Headline for ads in each search engine will not affect Click-Through Rate for that search engine.

The replication of the tests in different search engines provides the chance to test whether interaction between the two variables affects the measured outcome. The Null Hypothesis associated with this facet of the test states that interaction between the two variables will not have an affect on the output.

Excel Instructions:

Tools / Data Analysis / Anova: Two-Factor With Replication



You will be required to provide the following 4 inputs:

- 1) Input Range:** Highlight the data and labels as is highlighted in Yellow previously. Note that your data must be initially arranged in columns as above or similarly in rows. Note that both the Headline labels and the Ad Text labels are highlighted because they are the two factors being analyzed. Also note that data from duplicate tests run in different territories is arranged in pairs of rows, not columns. Excel requires this.

You do not need to state whether your columns and rows are labeled. One caution - if you use labels, place labels on both the columns and rows, as is the case in this example. Both columns and rows need to be labeled, or do not use labels at all.

2) Rows per Sample - State how many times each test will be duplicated in different locations. In this case, Each test was duplicated in one additional territory, so there are 2 Rows per Sample in this example.

3) Alpha: You will type in the level of significance here. The level of significance (Alpha) is the max level of error you are willing to tolerate. If you can only tolerate a possibility of 1% of error, the level of significance (alpha) here would be 0.01. Prefilled in this input box is 0.05. This represents a maximum error of 5%. Another way of looking at this would be to say that 5% maximum error represents being at least 95% certain in the correctness of the output.

4) Output: Here you would enter (or select) the upper-left corner of where you want the output printed on your worksheet. You will need to make sure that the place where you have specified the output to be printed has at least 7 free columns and 36 free rows.

Hit "OK" and Excel will perform the Two Factor ANOVA with Replication analysis and output the following result wherever you have specified.

Anova: Two-Factor With Replication

| SUMMARY | Headline 1 | Headline 2 | Headline 3 | Total |
|------------------|-------------|------------|------------|-------------|
| <i>Ad Text 1</i> | | | | |
| Count | 2 | 2 | 2 | 6 |
| Sum | 5.53 | 3.37 | 2.84 | 11.74 |
| Average | 2.765 | 1.685 | 1.42 | 1.956666667 |
| Variance | 0.00245 | 0.25205 | 0.0512 | 0.467226667 |
| <i>Ad Text 2</i> | | | | |
| Count | 2 | 2 | 2 | 6 |
| Sum | 5.97 | 2.9 | 2.82 | 11.69 |
| Average | 2.985 | 1.45 | 1.41 | 1.948333333 |
| Variance | 0.18605 | 0.005 | 0.3362 | 0.750576667 |
| <i>Ad Text 3</i> | | | | |
| Count | 2 | 2 | 2 | 6 |
| Sum | 5.13 | 6.03 | 3.25 | 14.41 |
| Average | 2.565 | 3.015 | 1.625 | 2.401666667 |
| Variance | 0.00125 | 0.03645 | 0.17405 | 0.444776667 |
| <i>Total</i> | | | | |
| Count | 6 | 6 | 6 | |
| Sum | 16.63 | 12.3 | 8.91 | |
| Average | 2.771666667 | 2.05 | 1.485 | |
| Variance | 0.073256667 | 0.62848 | 0.12407 | |

ANOVA

| Source of Variation | SS | df | MS | F | P-value | F crit |
|---------------------|-------------|----|-------------|-------------|-------------|-------------|
| Sample | 0.807211111 | 2 | 0.403605556 | 3.477026898 | 0.076062669 | 4.256494729 |
| Columns | 4.991077778 | 2 | 2.495538889 | 21.49885134 | 0.00037339 | 4.256494729 |
| Interaction | 2.277122222 | 4 | 0.569280556 | 4.904302671 | 0.022409688 | 3.633088512 |
| Within | 1.0447 | 9 | 0.116077778 | | | |
| Total | 9.120111111 | 17 | | | | |

Alpha = 0.05

The p-Value for Sample (0.076062) is more than the level of significance (0.05). We cannot reject the NULL Hypothesis that states that the Ad Text does not affect Click-Through Rate.

The p-value for Columns (0.00037339) is less than the level of significance (0.05). This indicates that choice of Headline affects Click-Through Rate. This is a rejection of the Null Hypothesis

The p-Value for Interaction (0.022409) is less than the level of significance. This indicates that different combinations of interactions (Ad Text / Headline) affects CTR.

Analyze the Output

The P-value

Interpreting the Output: Without going into too much statistical detail, the cells in the output entitled P-value will tell you whether the different ad text, headlines, and interactions between them had an effect on the measured output (Click-Through Rate on the two search engines). Highlighted in blue, yellow, and green on the Excel output is the P-value for each of these variables.

General Rule:

If P-value < alpha ----> The different treatments affected the output

If P-value > alpha ----> The different treatments did not affect the output

The data interpretation for Two-Factor ANOVA is very similar to Single Factor ANOVA. The only difference is that now two variables (choice of ad text and choice of headline)

along with interaction between the two variables are being tested to determine whether any of these has an effect on the measured output (Click-Through Rate). If the P-value associated with any of these variable is less than alpha (0.05), then we can state that this variable had an effect on the measured output (CTR).

Data for each Ad Text is arranged in rows. Excel has labeled this "Samples" in the output. The P-value associated with Ad Text is highlighted in Light Blue. This is the P-value for the source of variation in rows. This P-value is 0.076062 and is GREATER than alpha (0.05) so we can state with no more than 5% chance of error that Ad Text did not affect measured output (Click-Through Rate in each search engine). In other words, we cannot disprove the Null Hypothesis associated with this variable.

Data for each Headline is arranged in columns. The P-value associated with choice of Headline is highlighted in Yellow. This is the P-value for the source of variation in columns. This P-value is .00037339 and is much less than Alpha (0.05) so we can state that the choice of Headline almost certainly had an effect on the measured output (Click-Through Rate in each search engine).

Data and the P-Value associated with possible interaction between the two variables is highlighted in Green. This P-value is 0.022409688 and is less then the alpha (0.05) so we can state with less than 5% chance of error that interaction between the variables (Ad Text and Headline) did have an . effect on Click-Through Rate

ANOVA: Single Factor - Single Factor Analysis Calculated by Hand

(Excel calculation of Single Factor ANOVA is shown at the top of this Worksheet)

Problem 3: Closing Methods and Single Factor ANOVA Calculated by Hand

Problem: Three different sale closing methods were used. Three groups of four salespeople were randomly chosen. Each group was instructed to use only one of the closing methods for all of their sales. Sales totals of each salesperson over the next two weeks were collected. Determine with a 95% level of certainty whether there is a difference in the effectiveness of the closing methods. Below are the sales results for all 12 salespeople. All salespeople were assumed to have equal closing abilities.

| Sales Group 1 | Closing Method 1 |
|---------------|------------------|
| Salesperson 1 | 16 |
| Salesperson 2 | 21 |
| Salesperson 3 | 18 |
| Salesperson 4 | 13 |

| Sales Group 2 | Closing Method 2 |
|---------------|------------------|
| Salesperson 5 | 19 |
| Salesperson 6 | 20 |
| Salesperson 7 | 21 |
| Salesperson 8 | 20 |

| Sales Group 3 | Closing Method 3 |
|----------------|------------------|
| Salesperson 9 | 24 |
| Salesperson 10 | 21 |
| Salesperson 11 | 22 |
| Salesperson 12 | 25 |

Arrange the data as below to facilitate calculations:

| Sales Group ----> | Group 1 | Group 2 | Group 3 |
|-------------------|----------|----------|----------|
| Treatment ----> | Method 1 | Method 2 | Method 3 |
| Salesperson A | 16 | 19 | 24 |
| Salesperson B | 21 | 20 | 21 |
| Salesperson C | 18 | 21 | 22 |
| Salesperson D | 13 | 20 | 25 |

Here over the next several pages are the Hand-Calculations for this Single Factor ANOVA problem:

| | | | |
|--|----|----|----|
| Column Total | 68 | 80 | 92 |
| Column Mean | 17 | 20 | 23 |
| Grand Mean = $(17 + 20 + 23) / 3$ | | | |
| Grand Mean = | 20 | | |
| Column Mean - Grand Mean | -3 | 0 | 3 |
| $(\text{Column Mean} - \text{Grand Mean})^2$ | 9 | 0 | 9 |
| # Rows * $[(\text{Column Mean} - \text{Grand Mean})^2]$ | 36 | 0 | 36 |
| Sum of Squares Between Groups = $36 + 0 + 36 = 72$ | | | |

| | | |
|--|----------|----------|
| Method 1 | Method 2 | Method 3 |
| 16 | 19 | 24 |
| 21 | 20 | 21 |
| 18 | 21 | 22 |
| 13 | 20 | 25 |
| 68 | 80 | 92 |
| 17 | 20 | 23 |
| Method 1 | Method 2 | Method 3 |
| 16 - 17 | 19 - 20 | 24 - 23 |
| 21 - 17 | 20 - 20 | 21 - 23 |
| 18 - 17 | 21 - 20 | 22 - 23 |
| 13 - 17 | 20 - 20 | 25 - 23 |
| Method 1 | Method 2 | Method 3 |
| -1 | -1 | 1 |
| 4 | 0 | -2 |
| 1 | 1 | -1 |
| -4 | 0 | 2 |
| Square each | | |
| Method 1 | Method 2 | Method 3 |
| 1 | 1 | 1 |
| 16 | 0 | 4 |
| 1 | 1 | 1 |
| 16 | 0 | 4 |
| 34 | 2 | 10 |
| Sum of Squares Within Treatments = 34 + 2 + 10 = 46 | | |

Degrees of Freedom

Between Groups DOF = # groups - 1 = c - 1 = 3 - 1 = 2

Within Groups DOF = C(r-1) = 3 (4 - 1) = 9

Total Degrees of Freedom = 9 + 2 = 11

Sum of Squares

Between Groups Sum of the Squares 72

Sum of Squares Within Groups 46

Total Sum of the Squares 118

Mean Squares

MS = Mean Square = Sum of Squares / degrees of freedom

| | SS | df | MS |
|--|----|----|-------------|
| | 72 | 2 | 36 |
| | 46 | 9 | 5.111111111 |

F Statistic

F Statistic = (MS Between Group) / (MS Within Groups)

F Statistic = 36 / 5.111111 = 7.043478261

p Value

p-Value = FDIST(F Statistic,DOF Between Groups,DOF Within Groups) =

p-Value = FDIST(7.043478,2,9) =
0.0144192029269803

The p-Value represents the proportion of area under the F Distribution curve to the right of the given F value. If this p-Value is less than the stated level of significance, this demonstrates that there is a difference in the objects or process being analyzed. - in other words, there is a difference in the variances.

The p-value of 0.014419 is less than the designated level of significance of 0.05. This indicates that there is less than a 5% chance that this result could have occurred if there was no difference in effectiveness between the methods. Therefore, there is at least 95% certainty that there is a real difference in effectiveness of the methods. The Null Hypothesis which was rejected states that choice of closing method does not affect sales.