# LEARNING OBJECTIVES

The purpose of this exercise is to understand how you should state the null hypothesis and alternative hypothesis, explain the level of significance and how to conduct a couple of example statistical methods that can be used to determine significance.

In this exercise you will:

* Clearly state a testable hypothesis and identify the difference between type I and type II errors.
* Conduct a chi-square test to demonstrate the significance between two variables in our data set.
* Conduct a one-way analysis of variance (ANOVA) test to examine differences among three or more samples.

To complete the exercise, you will need to **download and use the Hypothesis-data.xls file**. Submit your workbook with **sheets for your chi-square analysis and ANOVA analysis**. Include an **answer sheet** where you record your answers.

## Hypothesis Testing

Hypothesis testing using statistical methods is a way of testing that the results of our survey or other data we have collected (i.e. through AB testing) yield meaningful or valid results and that the results we observe didn’t happen by chance.

**Alpha Value.** Before you run your hypothesis test, you need to decide on the alpha level, also called the significance level. In statistical terms the alpha level is the probability of rejecting the null hypothesis when the null hypothesis is true. i.e. the probability of making the wrong decision.

### Type I and Type II Errors

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Type of Error** | | | **To avoid this**  **type of error,**  **you should** | **To do this in a**  **chi-square test,**  **you should** | **What does this do?** |
| Type I Error | **False Positive** | Rejecting the  null hypothesis when it is true | Make it harder  to reject the  null hypothesis | Choose a lower  alpha value  (p < .01) | This increases the size of the critical value, making it harder to beat. |
| Type II Error | **False Negative** | Accepting the  null hypothesis when it is false | Make it harder  to accept the  null hypothesis | Choose a higher  alpha value  (p < .05) | This decreases the size of the critical value, making it easier to beat. |

# Directions for Chi-Square

Chi-square is a widely used test of significance and is particularly useful in tests involving categorical/nominal data. Using this technique, we test for significant differences between the observed distribution of data among categories and the expected distribution based on our sample (Cooper).

## Our Hypothesis

**Null Hypothesis (H0)**: Gender does not have an association with whether an individual perceives the Internet as good or bad for society in general.

**Alternate Hypothesis (H1)**: Gender does have an association with whether an individual perceives the Internet as good or bad for society in general.

To test our hypothesis, we have to decide an alpha value

**For this exercise use an Alpha level of 0.05**

## Creating expected and observed frequencies

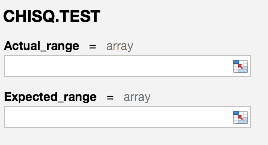
In this example we will be using the Society Pivot data, looking at the relationship between the respondents reported Sex and their perception of the value of the Internet on our society. Our first step is to create a contingency or frequency table of **GOODFORSOCIETY** by **SEX.**

Create the expected values for this table.

1. First, highlight the Observed pivot table (provided in the Excel doc), and copy it to a blank area below the original table. Right-click, choose **Paste Special,** and then select **Values**. (Had you simply chosen “Paste”, you would have difficulty later in changing the content of cells in your new table.)
2. Change the heading of your new table to “Expected Results” so that you remember what your new table will represent. You are not going to need the individual cell frequencies in the center of the table, so highlight and delete these. However, be careful not to delete the row and column totals – you will need these.
3. Now enter a formula to calculate the expected results in each cell of the table. The expected value in each cell **= Row Total \* Column Total / Grand Total.**
4. Be sure to set the decimal point to zero.
5. When you have finished, make sure that the figures in the body of your table add up to the correct column and row totals.

We can now calculate chi-square. Excel has a function called CHISQ.TEST that returns the test for independence: the value from the chi-squared distribution for the statistic and the appropriate degrees of freedom.

1. Under your expected results table enter the CHISQ.TEST function.



1. In the dialog enter your actual and observed range of data. Do not include the row or column headers in the array – only the Observed and Expected Results (the numbers in the cells). Note that the CHISQ.TEST function automatically accounts for the degrees of freedom.
2. Most likely you will see a small number, and you may see a strange result - something like 2.81637E-10 for example. This is the scientific notation for "ten to the power of..." If you see this type of result, format the results of the CHISQ.TEST cell as number, with at least 6 decimal places.
3. Label this cell as “**p value**.”
4. Compare this to the Alpha value we selected (0.05) to determine if you accept or reject the null hypothesis.

Now repeat the chi-square test comparing:

1. Party Preference. Use the **Party Pivot** sheet as your data source for this calculation.

## Questions:

1. Based on the results of your chi square test, would you accept or reject the null hypothesis regarding **SEX** and **GOODFORSOCIETY**?
2. How would you express this finding using natural language instead of variable labels?
3. What would our results look like if we had chosen a different Alpha value?
4. Based on your results would you accept or reject a null hypothesis that states “party preference does not have an association with whether an individual perceives the Internet as good or bad for society in general.”

# Directions for ANOVA

Analysis of Variance (ANOVA) is a method for determining if association exists between two variables when one variable is nominal and the second variable is ratio. In this example, we will examine three variables from the PEW study to address the following question:

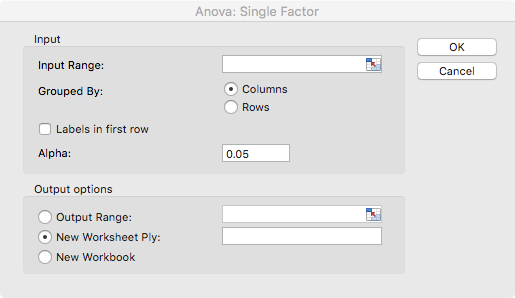
**Does a person’s age have some association with their perception of the social benefit of online environments?**

|  |  |
| --- | --- |
| **TOGETHER ME?**  (pial8) | Have you ever experienced any of the following things online? Have you ever seen an online group come together to help a person or a community solve a problem?     1. Yes 2. No |
| **GOODFORSOCIETY**  (pial11) | Overall, when you add up all the advantages and disadvantages of the internet, would you say the internet has mostly been a GOOD thing or a BAD thing for society?     1. Good thing 2. Bad thing 3. **(VOL.)** Some of both |
| **AGE** | Record respondent’s age in years |

Typically, the ratio variable is the outcome variable, and most tests are structured as to whether or not group membership has an influence on the outcome variable. However, in this example, the ratio variable would be the independent variable.

Open the workbook to the tab labeled ANOVA Sheet. Notice that that data is arranged differently than you see in the Working Sheet. For each variable, the indicators are arranged as columns. In each column are the responses to AGE for all the subjects for each indicator. In essence, the responses have been grouped according to the indicators for **GOOD FOR SOCIETY**, and then again for **TOGETHER**. Each indicator does not have the same level of age indicators

1. To calculate an ANOVA, go to the Data tab and select Data Analysis. Find the test for single factor ANOVA.



1. Using the menu, enter the range for the data (be sure to indicate that your data has labels).
2. Enter an Alpha value. How confident do you want to be in your results? Do you want to avoid Type I or Type II error? For this exercise, use an Alpha value of 0.05.
3. Compare the F, F-Critical and P values.

## Questions

Run the ANOVA test for **AGE** and **GOOD FOR SOCIETY**.

1. Why is it that AGE cannot logically serve as the dependent variable in this example?
2. Write a null hypothesis and alternate hypothesis for AGE and GOOD FOR SOCIETY.
3. Do you accept or reject the null hypothesis? Is there an association between these two variables? Express your findings in natural language.

Run the ANOVA test for **AGE** and **TOGETHER**.

1. Write a null hypothesis and alternate hypothesis for AGE and TOGETHER.
2. Do you accept or reject the null hypothesis? Is there an association between these two variables? Express your findings in natural language.
3. Choose one other variable to test for an association with AGE. Copy the data from the working sheet and configure the data in two columns so you can conduct your test.
4. Write your null and alternate hypotheses.
5. After doing the test, do you accept or reject the null hypothesis? Is there an association between these two variables? Express your findings in natural language.
6. Many people associate themselves with a gender other than male or female. What would you need to do to analyze the data if there were more than two responses for gender?

# Reference & Further Reading

O'Leary, Zina; Hunt J, Jennifer S. (2016). Workplace Research: Conducting small-scale research in organizations (Page 204). SAGE Publications.

Cooper, D. R., & Schindler, P. S. (2011). Business research methods. New York: McGraw- Hill Higher Education.