Chapter 5: Statistics 1

The focus on this class will be on practical use and application of statistics to drawing inferences from data. It is assumed the student has background in basic statistics. Here the goal is to apply that basic understanding of statistics to (a) supporting inference from data in simple experimental designs and (b) reporting results in standard APA format. Scientific writing in psychology generally follows a format defined by the American Psychological Association (APA) and described in the Publication Manual. Explanation for how to prepare a full scientific report starts in the next chapter (Chapter 6) but in this chapter we will see how to go from basic quantitative description of the data to the pieces that are crucial for the Results section in an APA format manuscript.

# Learning Objectives

1. Distinguish between descriptive and inferential statistics
2. Identify the different kinds of descriptive statistics researchers use to summarize their data
3. Describe the purpose of inferential statistics and carry out a t-test
4. Reporting results in APA format
5. Preparing data visualizations (graphs) of results

# Descriptive Statistics

Descriptive statistics are used to organize or summarize a set of data. Examples include percentages, measures of central tendency (mean, median, mode), measures of dispersion (range, standard deviation, variance), and correlation coefficients.

Measures of central tendency are used to describe the typical, average and center of a distribution of scores. The mode is the most frequently occurring score in a distribution. The median is the midpoint of a distribution of scores. The mean is the average of a distribution of scores.

Measures of dispersion are also considered descriptive statistics. They are used to describe the degree of spread in a set of scores. So are all of the scores similar and clustered around the mean or is there a lot of variability in the scores? The range is a measure of dispersion that measures the distance between the highest and lowest scores in a distribution. The standard deviation is a more sophisticated measure of dispersion that measures the average distance of scores from the mean. The variance is just the standard deviation squared. So it also measures the distance of scores from the mean but in a different unit of measure.

Typically means and standard deviations are computed for experimental research studies in which an independent variable was manipulated to produce two or more groups and a dependent variable was measured quantitatively. The means from each experimental group or condition are calculated separately and are compared to see if they differ.

A critical descriptive statistic for reporting results that is not always thoroughly discussed in prerequisite statistics classes is the **standard error**, which is related to standard deviation (or other measures of variance) but reflects estimated error in the data estimate of the conditions means. This value is critical to evaluating whether there are reliable differences between the conditions means and is a standard element of reporting the descriptive statistics in APA format writing. The formula for calculating standard error, SE, is:

## Data Aggregation and Handling

To organize and evaluate the data from a research study, it is strongly recommended to use a spreadsheet software program such as Microsoft Excel. Standard practice for research data is to prepare a matrix in which the data for each participant is on a single row with values across columns in the sheet. An example is shown below in Figure 1.

In actual practice, getting experimental data organized includes several steps. Data from individual participants might be entered by hand (e.g., from scoring written answers from in-person data collection) or through an automated scoring program (for online data collection). Once all the data is accumulated in the same spreadsheet, there will be a review to ensure that any pilot data collection is not inadvertently included in the final results and any exclusion criteria for performance are applied. Exclusion might be done for participants who opted not to complete the study or who did not comply properly with the instructions for the task for any reason. The data provided here for practice analysis are already organized and non-compliant data are excluded.

Table

Description automatically generated

Figure . Sample Experiment 1 with one participant per row containing information about the condition, date completed, percent correct, trials missed and trivia performance.

## Practically Calculating Descriptive Statistics

When calculating the needed descriptive statistics to communicate the findings of a study, you will need to include the average (mean), the standard deviation of the sample (SD) and the standard error (SE).

To do this in Microsoft Excel or other spreadsheet-based software, you will need to be able to apply formulas to ranges of cells in the spreadsheet. Start with calculating the **mean (M)** performance across one of the groups, e.g., the ‘deep’ encoding condition. In the cell to the right of the word “Mean” (D38), use the =average() formula and select the values in the above column D. In the cell below (D39), use the =stdev.s() formula to obtain the **standard deviation (SD)** of the sample (corrected from standard deviation of the population) by applying that formula to the exact same range of data. Next use Excel to count the number of participants **(N)** in this group using the =count() function on the same data range. Finally, calculate the standard error from the SD and the N by entering a formula to calculate this from these values in cell D41.

If you are unfamiliar with doing calculations in Excel, select cell D41, type ‘=’ in the cell to start entering a formula, click on the cell above containing the SD, type ‘/’ to indicate that you wish to divide this and then type =sqrt() to invoke the square root formula and click on the cell containing the number of participants.

Now repeat this process for the participants in the ‘shallow’ condition. Note that in the sample workbook, the data from the shallow condition are in another matrix with the same basic format in columns K-Q. Data can be organized in many different ways in Excel but this approach is used to make the data from both conditions most easily visible.

# Inferential Statistics

As you learned in the section of this chapter on sampling, typically researchers sample from a population but ultimately they want to be able to generalize their results from the sample to a broader population. Researchers typically want to infer what the population is like based on the sample they studied. Inferential statistics are used for that purpose. Inferential statistics allow researchers to draw conclusions about a population based on data from a sample. Inferential statistics are crucial because the effects (i.e., the differences in the means or the correlation coefficient) that researchers find in a study may be due simply to random chance variability or they may be due to a real effect (i.e., they may reflect a real relationship between variables or a real effect of an independent variable on a dependent variable).

Researchers use inferential statistics to determine whether their effects are statistically significant. A statistically significant effect is one that is unlikely due to random chance and therefore likely represents a real effect in the population. More specifically results that have less than a 5% chance of being due to random error are typically considered statistically significant. When an effect is statistically significant it is appropriate to generalize the results from the sample to the population. In contrast, if inferential statistics reveal that there is more than a 5% chance that an effect could be due to chance error alone then the researcher must conclude that their result is not statistically significant.

It is important to keep in mind that statistics are probabilistic in nature. They allow researchers to determine whether the chances are low that their results are due to random error, but they don’t provide any absolute certainty. Hopefully, when we conclude that an effect is statistically significant it is a real effect that we would find if we tested the entire population. And hopefully when we conclude that an effect is not statistically significant there really is no effect and if we tested the entire population we would find no effect. And that 5% threshold is set at 5% to ensure that there is a high probability that we make a correct decision and that our determination of statistical significance is an accurate reflection of reality.

For our practical analysis of our sample Experiment 1 data, we will utilize a **t-test** analysis to evaluate whether there is a reliable difference between our experimental conditions.

## The t-Test

As we have seen throughout this book, many studies in psychology focus on the difference between two means. The most common null hypothesis test for this type of statistical relationship is the t- test. In this section, we look at three types of t tests that are used for slightly different research designs: the one-sample t-test, the dependent-samples t- test, and the independent-samples t- test.

## One-Sample t-Test

The one-sample t-test is used to compare a sample mean (M) with a hypothetical population mean (μ0) that provides some interesting standard of comparison. The null hypothesis is that the mean for the population (µ) is equal to the hypothetical population mean: μ = μ0, sometimes described as “chance performance.” The alternative hypothesis is that the mean for the population is different from the hypothetical population mean: μ ≠ μ0. To decide between these two hypotheses, we need to find the probability of obtaining the sample mean (or one more extreme) if the null hypothesis were true. But finding this p value requires first computing a test statistic called t. (A test statistic is a statistic that is computed only to help find the p value.) The formula for t is as follows:

Again, M is the sample mean and µ0 is the hypothetical population mean of interest. We can see the importance of the standard error statistic here that is calculated as described above.

Considering the formula for the one-sample t-test illustrates that the t-value is based on how different the observed data were from chance performance. For our Experiment 1, chance would be 50% correct on the recognition test. In many other simple designs chance performance might be zero. The size of the difference between the observed data M and chance is then adjusted by the size of the SE, the estimate of the variance in the mean. A good rough rule of thumb for reliability in a t-test is that the t-statistic should be >2.0 for a result to be statistically reliable (exact probabilities depend on the number of participants). Conceptually this means that the difference between the mean and chance should be roughly twice the size of the SE. This rule of thumb provides an intuitive way to get a quick sense of how robust the experimental effects are within a dataset and shows why calculation of the SE is an important component of the descriptive statistics.

## The Dependent-Samples t–Test

The dependent-samples t-test, also called the paired-samples t-test, is used to compare two means for the same sample tested at two different times or under two different conditions. This comparison is appropriate for pretest-posttest designs or within-participants experiments. The null hypothesis is that the means at the two times or under the two conditions are the same in the population. The alternative hypothesis is that they are not the same. This test can also be one-tailed if the researcher has good reason to expect the difference goes in a particular direction.

It helps to think of the dependent-samples t-test as a special case of the one-sample t-test. However, the first step in the dependent-samples t-test is to reduce the two scores for each participant to a single difference score by taking the difference between them. At this point, the dependent-samples t-test becomes a one-sample t-test on the difference scores. The hypothetical population mean (µ0) of interest is 0 because this is what the mean difference score would be if there were no difference on average between the two times or two conditions. We can now think of the null hypothesis as being that the mean difference score in the population is 0 (µ0 = 0) and the alternative hypothesis as being that the mean difference score in the population is not 0 (µ0 ≠ 0). We will return to the question of how to conceptualize the dependent-samples t-test as a one-sample t-test in Chapter 7.

## The Independent-Samples t-Test

The independent-samples t-test is used to compare the means of two separate samples (M1 and M2). The two samples might have been tested under different conditions in a between-participants experiment like our Experiment 1, or they could be pre-existing groups in a cross-sectional design (e.g., women and men, extraverts and introverts). The null hypothesis is that the means of the two populations are the same: µ1 = µ2. The alternative hypothesis is that they are not the same: µ1 ≠ µ2. Again, the test can be one-tailed if the researcher has good reason to expect the difference goes in a particular direction.

The t statistic here is a bit more complicated because it must take into account two sample means, two standard deviations, and two sample sizes. The formula for calculating this statistic, as taught in basic statistics classes is:

Notice that this formula includes squared standard deviations (the variances) that appear inside the square root symbol, but these are calculated slightly differently than the SE of each group. Here the calculation of the variance is based on pooled variance, which reflects an assumption of equal variances across the experimental conditions.

When first introduced to statistics, it is common to review some of the critical assumptions about the underlying data distributions and to consider the question of whether the experimental hypothesis is “directional.” For our purposes here, we will focus on how to carry out the analysis under some simplifying assumptions.

First, we will always assume that we should do a “two-tailed” test to evaluate whether our results are statistically reliable. The reference to tail here refers to the distribution of t scores that is evaluated together with the **degrees of freedom (df)** often in the context of looking up t-values on a probability table. We will not be using probability tables here, we will use software to calculate exact probability estimates from our data. Technically, if we had a strongly directional hypothesis, we could use a “one-tailed” test to evaluate whether our hypothesis is robust. However, this approach is known to use a much less stringent threshold for reliability, which greatly increase the risk of a Type 1 error in inference, which we want to avoid. In most practical cases in experimental psychological science, if we have data that does not meet the more stringent two-tailed criteria, we would prefer to reconsider the questions of our experimental design and operational definitions and try to do a better experiment. Thus we will always assume/prefer to do a two-tailed test and avoid the complex question of whether a one-tailed test would be appropriate.

## Degrees of Freedom (df)

To evaluate the probability with which we can reject the null hypothesis, we will need both the t-statistic and the **degrees of freedom (df)** for the analysis. For a two independent sample t-statistic under the assumption of equal variances, the df will always be the total number of participants minus 2. Conceptually, the more df there are, the more participants are included in the study and the more confident we can be of our conclusions. Because it is intrinsic to the statistical evaluation of data, the df must always be included in properly formatted reports of experimental results.

In some conditions, we might observe conditions where the df is not exactly equal to the total N-2 because of a correction made by the analysis software for ‘unequal variances.’ However, if you did not hypothesize in advance that the variances would be unequal, you should generally not be using a t-test if you have observe robustly unequal variances.

Note that our guidelines here of strongly assuming equal variance and always use two-tailed tests reflect an approach of simplifying the core approach to statistical inference in a way that is consistent with most simple experimental design. Our goal in this class is to provide hands-on practice with the process and tools of psychological research in an accessible and tractable way. More complex analysis, including reconsideration of these assumptions, is common in psychological research and there are a wide range of more sophisticated tools for analysis that are applied in cutting-edge research. These advanced topics in research methods and analysis are outside the scope of our introductory class here.

# Running a t-test in R

Start by installing the R program and the RStudio suite (in 2 steps)

* Click the link below to go to R download page and choose the version that is compatible with your operating system (i.e. Mac High Sierra, Mac Catalina, Windows 10, etc.): <https://cran.r-project.org>
* Once R has downloaded, install it on your computer.
  + It requires permissions.
  + Accept the license.
  + Install all the default components.
  + Don’t customize startup options.
  + Default additional tasks are fine.
* After R has been installed click the link below and download the RStudio version that is compatible with your operating system:
  + <https://rstudio.com/products/rstudio/download/#download>
  + If you are coming through the RStudio site, go to products, then RStudio Desktop. Use the Open Source Edition (Free).
  + Download will adjust to your OS. The Windows download is 171M, so be aware of bandwidth constraints and speed.
  + Current version is 1.3.
  + MS Windows complains to me that it isn’t a Microsoft verified app. However, it is safe to install.
* Once RStudio has downloaded, install it on your computer.
  + Note: You will not be able to install/run RStudio until R has been installed.

Use the RStudio program to start an analysis session

* Launch RStudio. You should see a screen with 4 panels. We will be primarily working with the left 2 panels.
  + The top left panel will have lines of code, a ‘script’ for carrying out the steps required for an analysis.
  + The bottom left panel will have the output results of executing those steps, including error messages if something goes wrong.
* Use File -> Open and navigate to the folder on your computer where you’ve installed the 205 files and associated data from our experiments
  + Open the file “205Exp1\_Fall2022.R” This is an R script for testing your installation and re-running the t-test analysis from our Experiment 1 data for the Inclass experiment.
* On a fresh install, this will produce a warning that there are required packages that are not installed. The option to install them is provided. You can also install them by working through the script analysis steps.
* Set the “working directory” to where your data are stored on your computer. If you have put the “exp1\_data.csv” in the same folder as the “205\_Exp1\_Fall2022.R” file, navigate to the Session menu, then to Set Working Directory and select the top option “To Source File location.”
* To run a single step of the analysis press the “🡪 Run” button that is in the upper right part of the top-left panel. This carries out the step in the script on which the cursor is currently. If you didn’t do the installation of the ‘psych’ and ‘ez’ packages above, put the cursor on line 2 and Run. Then put the cursor on line 3 and Run.
* The installation process will also download and install a series of other packages needed (called dependencies). The process should only take a few minutes to run.
* Now move down to line 6, “library(psych)” and press Run. This loads a set of routines for data analysis for psychology experiment data that are helpful.
* The cursor moves down to the next line after each Run. Press it again to load the library on line 7, 8, and 9 (‘psychTools’, ‘tidyr’, and ‘ez’).
* With luck you are not getting error messages in the bottom left panel. If you are, something may have gone wrong with the above steps.
* The next step, line 12 will start loading our actual data. If everything is working you should see: “Data from the .csv file Inclass\_Exp1\_data\_R.csv has been loaded.” In red in bottom left panel.
* Run on line 13 will cause the data table to be printed in another tab. It should look a lot like what the source data file looks like if you open it in Excel or another spreadsheet program.
* Run on line 17 to see the output of the describeBy function, which provides descriptive statistics for our data. You may notice that this needs to be unpacked a bit to find the key numbers, which are the Test\_score values for each condition. Check that these numbers are identical to the descriptive statistics you calculated in Excel above.
* Run on line 19 to carry out the two independent samples t-test for the data.

If everything works up to this point, then congratulations! You have just run your first formal analysis of experimental psychological data.

## Writing up statistical reports

However, we are not done yet with our statistical process. The use of statistics in experimental research is to support conclusions about our research study and the hypothesis about the underlying constructs. The final step in handling our inferential statistics is to format the output to follow the standard reporting format for an APA publication. This reporting format will include the t-statistic, degrees of freedom and the p-value that indicates the probability of the data occurring as observed under the null hypothesis. Any program to support calculation of statistical inference will provide that information, but every program tends to have its own unique way of formatting the output. Look through the output to identify those three key numbers and then format the results within the following basic frame:

t(DF) = X.XX, p<.YY

In your report of the results, replace the DF with the reported degrees of freedom, replace the X’s with the reported t-statistic and replace the Y’s with the reported p-value. In general, you will want to round the numbers to 2 significant digits no matter how many digits your output contains. Remember that the goal at this point is to be able to communicate the results of the experiment. With our rule of thumb that we expect t>2 for reliable results, we do not need to include a lot of additional numbers after the decimal point to make our case. Similarly, the point of reporting the p-value is to establish that it is <.05, the standard criterion for psychological science reporting. Being a lot smaller than .05 does not contribute to understanding the validity of the experiment.

# Data Visualization

A particularly useful way to communicate the results of a research study is to prepare a figure that illustrates the difference in experimental conditions. As an example, we will make a Figure for our Experiment 1 data using MS Excel (note that Excel calls data visualizations ‘charts’ but they are formally referred to as Figures in APA format).

For a simple 2-group design like our Experiment 1, a bar plot is a highly effective method for communicating results. Later we will consider the use of line graphs for illustrating more complex designs. In those more complex designs, we can compare the value of graphing data as bars or lines. There is no hard rule on which is better and the choice of how to present data is up to the author of the manuscript reporting the results. The goal is to use a data visualization figure to help the reader understand the results, so it should be foremost in your plan for visualization design to ensure that it communicates effectively. Here we will illustrate some basic elements to include to accomplish this.

The first step towards creating a figure is generally to create a separate, labeled table of the key numbers that will contribute to the graph. The numbers we need for the graph will be the mean and the SE for each of the two conditions in our study: deep and shallow. You can format them as following (note that the data numbers are for illustration and do not reflect our actual data):

Table, calendar

Description automatically generated

Making the initial bar graph is as simple as selecting the grid from A2 to B3 and then Insert Column or Bar Chart from the MS ribbon. That will produce a preliminary chart that looks like:

Chart, bar chart

Description automatically generated

We need to do some editing to the layout of this chart to make it effective and in approximately standard format. First, the Chart Title can be cut as we generally do not include titles on figures in manuscripts. Titles are used to help describe data in presentation format, but APA reporting format requires that Figures be accompanied by a figure caption which is where the description of the illustration should be included. In addition, both the x- and y- axis should be drawn in black to ensure visibility of the axes. You will want to label the y-axis by adding the Chart Element, Axis Title -> Primary Vertical and then change the text label to Recognition Score. You may also optionally choose to remove the horizontal lines (these are chart elements called gridlines accessed through the Chart Design menu) or even change the color of the bars. You

Once the basic layout is set, the last element to be added is brackets reflecting the SE of mean. We kept these numbers near our means above, but note that we did not select those numbers when making the graph (if you have 4 bars in your graph, you may have selected them accidentally).

To add error bars scaled to the SE, click on the graph and specifically one of the two bars. Then in the Add Chart Element menu, select the Error Bars option and the bottom choice, More Error Bars Options from there. In the Format Error Bars pane, choose Custom for your error bar size (bottom option) and select Specify Value. For both the Positive Error Value and Negative Error Value choose the range where you put the SE values (C2:C3 above), then select Ok. Your figure should look something like this now.

Chart, box and whisker chart

Description automatically generated

Note that in this example the size of the error bars is slightly different and should line up with the data from those cells. In this sample data, the Shallow data has slightly larger SE than Deep.

The figure can now be copied from Excel directly into a document where you are preparing the report of your research. You can now write the Figure Caption to accompany the figure which explains what you are trying to communicate with your data visualization. In general in the caption, you will want to explain the axes (the DV is shown on the y-axis) the conditions, the direction of the results and note that “brackets reflect the standard error of the means.”

Review and discussion of how to properly report your results in APA standard format will continue in Chapter 6.