Operations Manual 2

Data Analysis for Experiment 1

This chapter pairs with **Chapter 5** in the main text to provide a specific recipe for carrying out the t-test needed to analyze the data from Experiment 1 produced by students on the first day of class.

The first section provides a step-by-step walk through of installing and running the analysis program R and using a provided R *script* to do the t-test analysis on the data provided by the instructors.

The second section explains how to make a data visualization to present the results of the data as a bar plot. Instructions for making the graph are provided based on using Microsoft Excel. Note that one key element of the graph, the representation of the *standard error of the mean* on the average data, takes several steps and this option is not available on all software packages (e.g., Google Sheets and the online Microsoft Excel do not provide this option and cannot be used unfortunately). Chapter 6 in the text explains how to format the graph as a Figure to be included in an APA format manuscript.

Running a t-test in R

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

- Use the link below to go to R download page and choose the version that is compatible with your computer's operating system: 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 use this to 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. You will also need to have downloaded the provided data from instructors, usually as a file named something like “Exp1_data.csv”.

- 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 files and associated data from our experiments

- Open the file provided for data analysis. This is an R script for testing your installation and re-running the t-test analysis from our Experiment 1 data for the in-class 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 data file in the same folder as the analysis 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 Exp1_data.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 16 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 Recognition.score values for each

condition. Check that these numbers are identical to the descriptive statistics you calculated in your spreadsheet previously.

- 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.

It should look something like:

```
     Two Sample t-test

data:  Recognition by Condition
t = 6.005, df = 46, p-value = 2.846e-07
alternative hypothesis: true difference in means between
group deep and group shallow is not equal to 0
95 percent confidence interval:
 11.52491 23.14705
sample estimates:
   mean in group deep mean in group shallow
             91.66190              74.32593
```

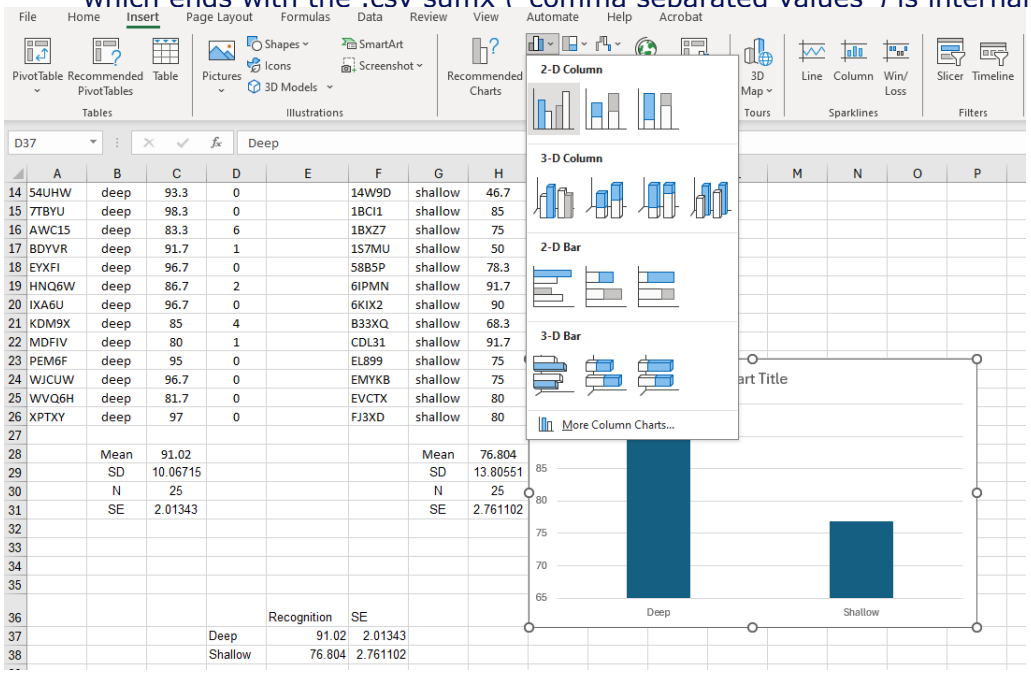
The output lines from the analysis will contain information about the calculated t-value and the associated p-value to assess the reliability of the effect. Refer to Chapter 5 for how to proceed with presenting this information in APA format for inclusion in a manuscript reporting the results of your experiment. The statement of inferential statistics should look like:

$t(46)=6.0, p<0.001$

Making a graph in Excel

Here we will illustrate some basic elements to include to accomplish this. For this we will be working with the same data in Excel format where formulas, graphs and othe features can be included with the data. The file used in R,

which ends with the .csv suffix ("comma separated values") is internally a



	A	B	C	D	E	F	G	H
14	54UHW	deep	93.3	0		14W9D	shallow	46.7
15	7TBYU	deep	98.3	0		1BCI1	shallow	85
16	AWC15	deep	83.3	6		1BXZ7	shallow	75
17	BDYVR	deep	91.7	1		1S7MU	shallow	50
18	EYXFI	deep	96.7	0		58B5P	shallow	78.3
19	HNQ6W	deep	86.7	2		6IPMN	shallow	91.7
20	IXA6U	deep	96.7	0		6KIX2	shallow	90
21	KDM9X	deep	85	4		B33XQ	shallow	68.3
22	MDFIV	deep	80	1		CDL31	shallow	91.7
23	PEM6F	deep	95	0		EL899	shallow	75
24	WJCUW	deep	96.7	0		EMYKB	shallow	75
25	WVQ6H	deep	81.7	0		EVCTX	shallow	80
26	XPTXY	deep	97	0		FJ3XD	shallow	80
27								
28		Mean	91.02				Mean	76.804
29		SD	10.06715				SD	13.80551
30		N	25				N	25
31		SE	2.01343				SE	2.761102
32								
33								
34								
35								
36					Recognition	SE		
37				Deep	91.02	2.01343		
38				Shallow	76.804	2.761102		

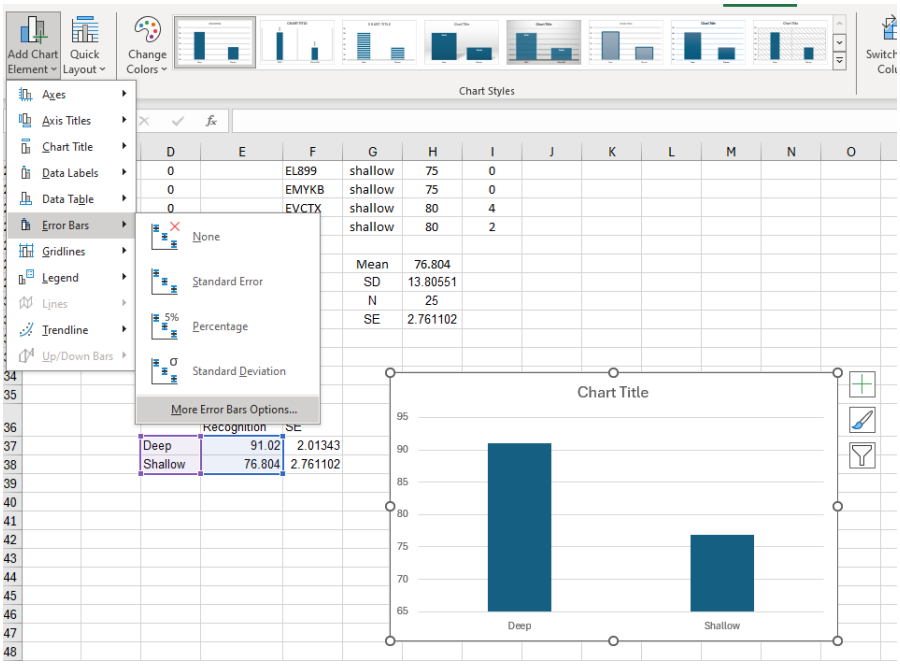
simple text file of information separated by commas. An Excel spreadsheet can contain a lot of additional information useful for data review and also provides facility for making data visualizations, specifically Figures to support presentation of experiment data.

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. For simplicity, we want these organized into a separate data table. In the figure below, you can see a labeled 2x2 table where the mean recognition memory percent correct scores are copied from the descriptive statistics to cells where they are adjacent at the bottom of the image.

In this image, those two cells and their labels are selected with the mouse and then in the Insert tab, the upper left option of Bar Charts is selected and then the upper left option of 2-D Columns is selected. As you do this, Excel already renders an initial image of what the starting bar plot is going to look like overlaid on the spreadsheet.

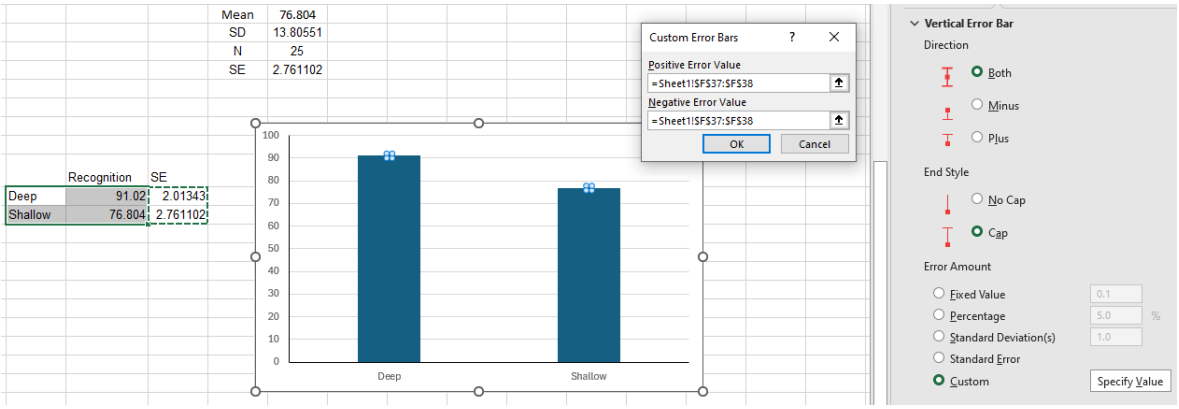
From this initial draft, we need to do some editing to the layout of this chart to make it effective and in approximately standard format. First, the Chart Title can be cut as we generally do not include titles on figures in manuscripts. Titles are used to help describe data in presentation formats, 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 correctly the size of the calculated 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. We will need to specify both the positive and negative sizes of the error bars, above and below the mean. For our data, these are the same sizes. For both the Positive Error Value and Negative Error Value choose the range where the SE values have been copied (F37:F38 above). Because they are next to each other, we can select both values and these will be correctly applied to both bars. The figure below shows roughly what this will look like. When done. select Ok.



This feature of setting the error bars flexibly to specific values for each group allows for correct presentation of both the means of the observed data, shown in the height of the bars, and the variance, shown in the SE bars. This puts many of the key descriptive statistics into the Figure visually. In addition, a useful trick is to look at the range implied by the error bars. For an independent samples experiment, if the error bars to not overlap (touch), then you most likely have a reliable difference between the groups. That means that the figure is also carrying some implicit information about the inferential statistics. You should always check or carefully include the actual statistical test in the reported text, but a well-made figure acts as a very effective overview of the results.

You can further edit the graph you have made to improve the style and

effectiveness of the visualization. Always remember that the goal of a Figure in a research report is to effectively communicate results. For example, Excel currently defaults to drawing the axes in a nearly invisible light gray color. You can select these and change to a more visible black. Excel also defaults to including horizontal grid lines which you will rarely see in published papers and you can choose to remove this element of the graph if you wish.

The graph can be copy/pasted into your manuscript document from here, but note that you should refer to Chapter 6 for correct formatting of a Figure in an APA manuscipt and a Figure Caption must be included with the graph.

Results for Experiment 1

With the output of the R script for analysis of the Experiment 1 data, the descriptive statistics (from either Excel or R) and the graph visualizing the performance on the Recognition Memory dependent variable, you are now ready to write the Results section of your experiment report for the first study. Your report should follow standard APA scientific writing style, which is reviewed in Chapter 6 and is also available through the Publication Manual of the American Psychological Association. Note that a lot of the numbers you have worked through and looked at here do not end up appearing in the final report. The report presents specifically the descriptive (means, sandarad deviations and standard errors) and inferential statistics (t-test, p-value) in standard formatting. The figure is accompanied by a note or caption explaining what the visualziation is intended to communicate.

The data matrix of individual results is not included in a standard experiment report, nor is the full output of the statistical program. These are sometimes included as Supplemental Materials in modern publishing protocols, but are not considered standard parts of the report. Because the quantitative reports are often essential to supporting conclusions drawn from a research study, it is particularly important that the numbers be accurately reported in the manuscript. Typographic errors or mis-formatting the numerical report can lead to catastrophic mistakes in communcating your findings. Proofreading those sections multiple times is highly recommended.