Psychology/Neuroscience 201:
Statistics and Research Methods in Psychology

SPSS
Instructions

Last updated: August 2018

Table of Contents

SPSS: An Introduction

Welcome to your first experience with SPSS (Statistics Package for the Social Sciences). I hope you find this packet useful as you begin to navigate your way through this statistics software package. You will refer to these instructions throughout the semester, but I highly recommend hanging on to the packet for use in future classes and (if you are a Psychology or Neuroscience concentrator) for your Senior Project.

Overview

This section of the packet will walk you through the following steps:

1. Logging on to the Citrix server to access SPSS.
2. Downloading a data file to your computer and uploading it to your SSS (student storage server) space.
3. Opening a data file in SPSS and saving it to your computer or to your SSS space.
4. Creating a data file in SPSS.
5. Computing frequency tables and descriptive statistics and creating histograms/bar graphs using SPSS.
6. Doing some basic data manipulations using SPSS (recoding variables, computing new variables).
7. Ending your session with Citrix.

PART 1: Logging on to the Citrix server to access SPSS

SPSS for Windows is available via the Citrix server on any computer on campus, Macintosh or Windows. Consequently, you will be able to use SPSS on your own computer via the Citrix server. You will need to know your username and password for linking to the Citrix server, which should be the same as your e-mail username and password. The first time you use Citrix (and at the beginning of each academic year), you'll need to register (it's free). See https://www.hamilton.edu/offices/lits/rc/registering-and-logging-into-citrix if you need help.

1. Open your web browser and enter the following address: citrix.hamilton.edu. I recommend creating a Bookmark called "Citrix" so you can easily navigate to it whenever you need to. Once you've registered (see above), you will be prompted to download and install the Citrix receiver client. If you have any difficulty doing so, please see me.

2. Once the client is installed, the Citrix Metaframe window will open when you go to the above address, and you will be prompted for your username and password, which should be the same as your e-mail username and password (see Figure 1). After you type these, select the appropriate Domain: "students.hamilton.edu" and click on "Log On."

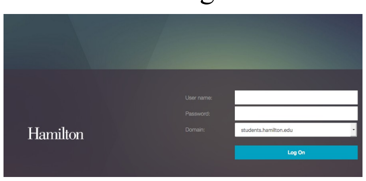


Figure 1

3. You should now see icons for the available software packages you can access through the Citrix server (see Figure 2). You will be using **SPSS**. Find the correct icon and click on it.

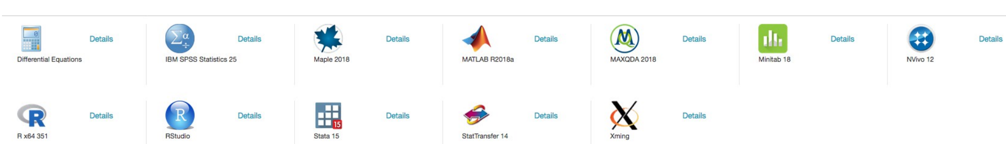


Figure 2

4. You will see a task progress window indicating that the application is starting. After a short lag (this could take up to 30 seconds or more), SPSS should be up and running and you can create a new data file or open an existing one. See instructions below in Part 3 for opening data files.

PART 2: Downloading the data file to the desktop and uploading it to your SSS space

1. Open your web browser to the "My Hamilton" web page at the following address: (http://my.hamilton.edu/myhamilton/login.cfm).

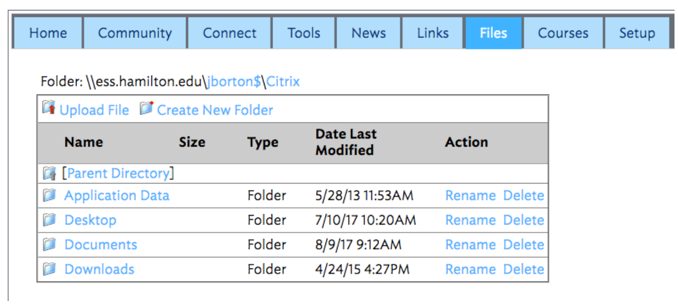
2. Log in using your email username and password.

3. Click on the "Courses" tab, and in the column under "My Blackboard Courses," click on the link to our course. The Blackboard page for our class will open.

4. Click on "Assignments" and you will see an item labeled "SPSS data file" with a link to a file called **GSS data 2016.sav**. (GSS = General Social Survey.) Click on the file and you will be prompted to save it. You have a couple of options here:

    a. Save the file to a folder for this course that you create on your laptop. Use this option if you plan to exclusively use your own laptop for data analysis in this course.

    b. Save the file to your SSS. This option is the most flexible, as you can then access it no matter what computer you happen to be using. The easiest way to accomplish this is to temporarily store the file on your local computer's desktop and then upload it to your SSS via the "Files" tab on your My Hamilton page. A folder called "Citrix" is in your SSS (it gets created the first time you open SPSS using Citrix) and there is a subfolder named "Documents" within that folder. Here is where you want to store all of your data files (see Figures 3 and 4 below). Once you've saved the file to your desktop, go to My Hamilton, choose the "Files" tab, click on "Citrix," then "Documents," and then choose "Upload File" to place the data file in the documents folder within the Citrix folder on your SSS.



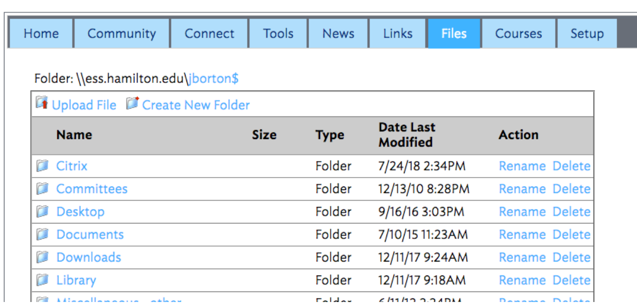


Figure 3

Figure 4

PART 3: Opening SPSS and saving your data file to your SSS space

Note that you cannot simply double-click on an SPSS (.sav) file to open it. You must be within the SPSS program in order to open files. When SPSS is first opened, it will prompt you with the following window:

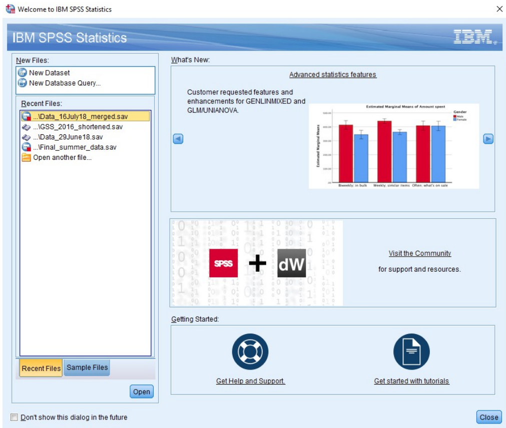


Figure 5

For the descriptive statistics project, you are using an existing data file, so select **"Open another file…"** in the "Recent Files:" box and then click "Open" at the bottom of the window (or just double-click "Open another file…"). A window will open to allow you to navigate to the pre-existing data file.

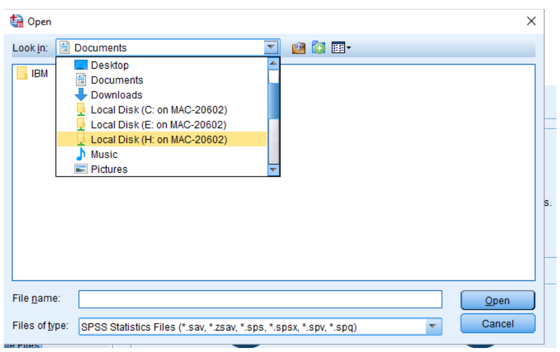


Figure 6

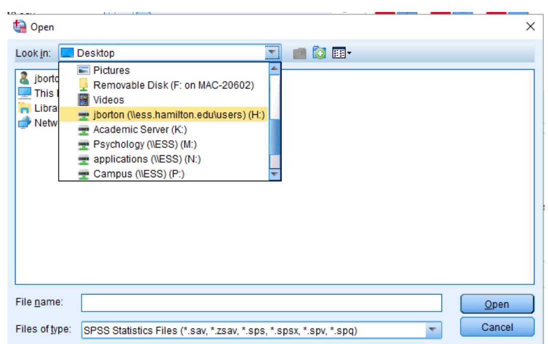


Figure 7

How to find your file:

1. If you saved it to a folder on your laptop (see Figure 6): Click the down arrow on the menu bar and navigate to Local Disk H (see above left). Note that Local Disk C is your hard drive and H is you, signed into your hard drive. It's faster to just click H rather than having to click C, then find your username, then get to H. Once you've clicked on H, you can navigate to the folder where you stored your file (or to your desktop, if you just dumped it there). Click on the file and then click "open". Your data file will open in SPSS.

2. If you saved it to your SSS (see Figure 7): Click the down arrow on the menu bar and navigate to your username followed by (\\sss.hamilton.edu/users). (Note that mine is ESS for "Employee Storage Space" rather than SSS.) If you click on that, you'll then see the Citrix folder and the documents folder into which you saved your file earlier. Click on the file and then click "open". Your data file will open in SPSS.

Saving files

Be sure to save frequently so you don't lose your work due to unforeseen circumstances. If you make changes to the file, you can save an updated version of it by clicking on **File** in the SPSS menu bar, then **Save As**. A window similar to that shown in Figure 6 or 7 will appear. Name your file with a meaningful, informative label, then click '**Save**.' A copy of your file will be saved in the folder you designate (either on your hard drive or on your SSS).

File types in SPSS

Note that different SPSS file types have different extensions:

.sav = data file
.spv = output file
.sps = syntax file (written code for SPSS to execute; we won't do much with syntax in this course, but you will become more familiar with it in subsequent courses. Note that any command you ask SPSS to run via the menus can be pasted into a syntax file and saved for later use. This is *very* handy when you're working on an actual research project.)

If you click on the folder icon in a data window, SPSS will automatically try to open data. If you click on the folder icon in an output window, SPSS will automatically try to open output. If you click on the folder icon in a syntax window, SPSS will automatically try to open syntax. If you are in a data window and want to open either an output or syntax file, you will need to go to **File**, **Open**, and then choose data, output, or syntax.

PART 4: Creating a data file in SPSS

When you open SPSS, one of your choices in the "New Files:" box is "**New Dataset**" (see Figure 5, above). If you click that option, you will be presented with an untitled data file (see Figure 8). Note that this file looks like a standard spreadsheet.

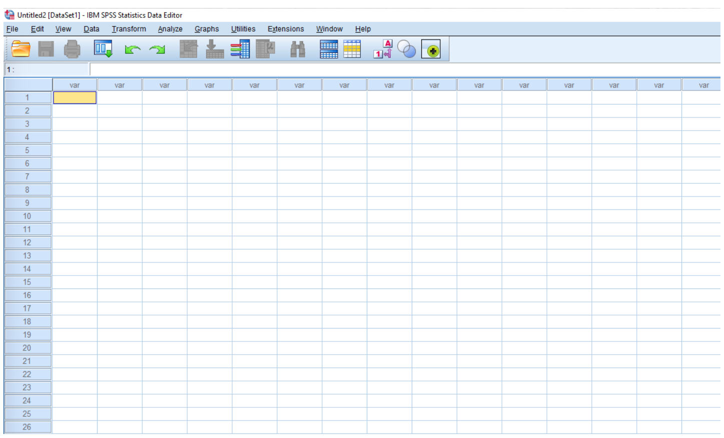


Figure 8

Before you can analyze your data, you must first input them into this file. Notice that the rows are already numbered and the columns are labeled "*var*." At this stage, they are shaded blue because they have not been activated with data or variable names.

As you get familiar with the format of this spreadsheet, you should recognize that each column represents a variable and each row represents a participant's data for each variable. So each row represents the scores from one particular participant.

Before entering your data into this data file window, you'll need to do a couple things. First, you need to figure out what you're going to call each variable, keeping in mind that SPSS won't let you begin a variable name with a number or use special characters, such as spaces. Next, you need to figure out what sort of information you'll enter for each variable. In some cases, both of these decisions might be easy; but in other cases, more thought might be required. For instance, if one of the variables that you want to enter is the gender of the participant, you can't enter the words "male" and "female;" instead, you need to develop a code (a way to quantify the variable). So, you might decide that for the variable "gender," you will use a "1" to represent male and a "0" to represent female. Of course, for some variables (e.g., age), you won't need a code, but will simply enter in the value the participant reports.

Define each variable

You can examine your variables/data in two different screens or “views”: **data view** and **variable view**. In data view, you see the basic spreadsheet depicted in Figure 8. In variable view, you can easily name and define your variables (see Figure 9). You can toggle back and forth between these two views by clicking on the appropriate button (see bottom left of the screen). You can also double-click on a “var” header in data view to automatically toggle to variable view.

I recommend that your first variable always be participant ID (an arbitrary number you assign to each participant for record-keeping purposes). Below is an illustration of the variable view with the first variable added. Notice the variable label (Participant ID Number) in the “label” column. It is important to get in the habit of labeling your variables. Labels help remind you what each variable is, and they also print out in the output so you know what you’re looking at. The variable labels for the data set for the descriptive statistics project are going to be very important in helping you understand what the survey items were.

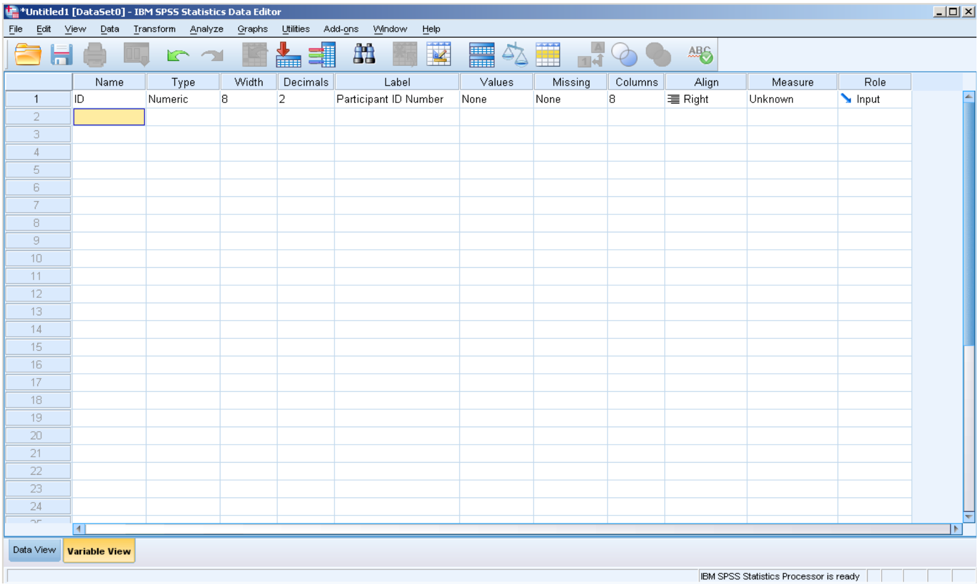


Figure 9

Let’s say that you wanted your second variable to be participant gender. As with participant ID, you need to provide a *variable name* (a short name, e.g., “gender”) and a *variable label* (a description of the variable to remind yourself of what the variable is; e.g., “participant gender”). For this variable, though, you’ll need to provide *value labels* to indicate what different levels of your variable mean. To do so, simply click in the appropriate cell in the “values” column (i.e., the cell that appears in the “gender” row). A lightly shaded blue box appears within the cell, and if you click on it, a large dialog box appears in the middle of the screen (see Figure 10). See how I’ve indicated that “1” represents male participants and “0” represents female participants? You can simply type the value in one box and the value label in the box beneath it. Then click on “Add.” Once you’ve added all your values and value labels, click “OK.”

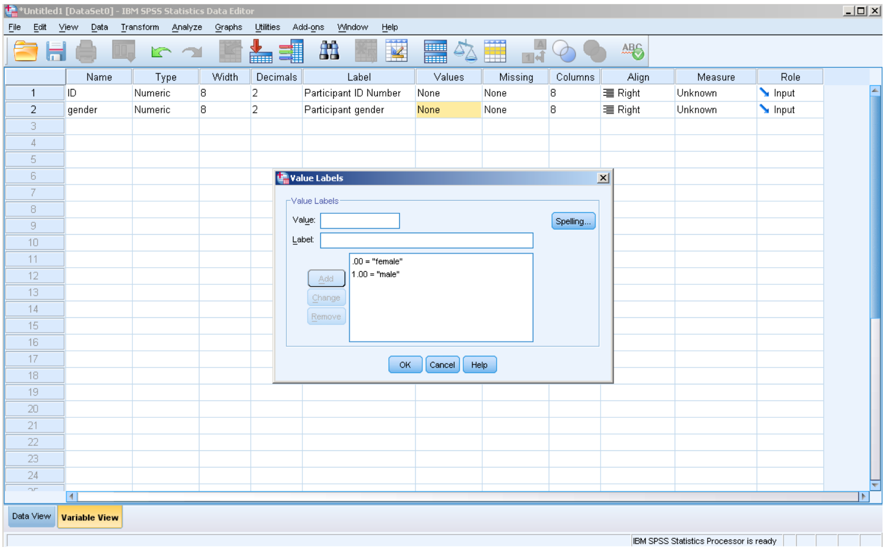


Figure 10

Determining the number of decimal places each variable has

In addition to giving your variable a label and defining its values, you may find it convenient to choose the number of decimal places you want displayed for each variable. For instance, if your variable is "gender," and the values are "1" and "0," it might annoy you to have SPSS display "1.00" and "0.00" all the way down the column (the default is 2 decimal places). To change the number of decimal places displayed, simply click on the appropriate cell (e.g., the one in the "gender" row and the "decimals" column). Up and down arrows appear in the right-hand part of the cell, and you can simply click on the down arrow until you have 0 decimal places (or you can type "0" directly).

Repeat this process of defining and labeling for all of your variables. NOTE: You're not entering any of the participants' data yet; you're only getting the data file set up by creating and labeling the variables.

Setting missing values

In SPSS, if you leave a cell blank, you will simply get a period (".") in that cell, and it will not be included in any of the data analyses. Sometimes it is useful to use a particular number to represent missing data (e.g., 99, as long as 99 is not a valid value for that variable), and then to alert SPSS that a 99 is a missing value. Note that in variable view, there is a column titled "Missing" that alerts you to all the missing values for each variable.

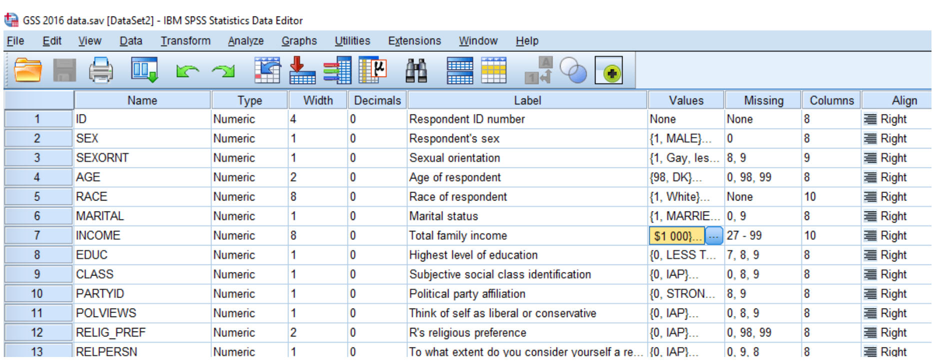


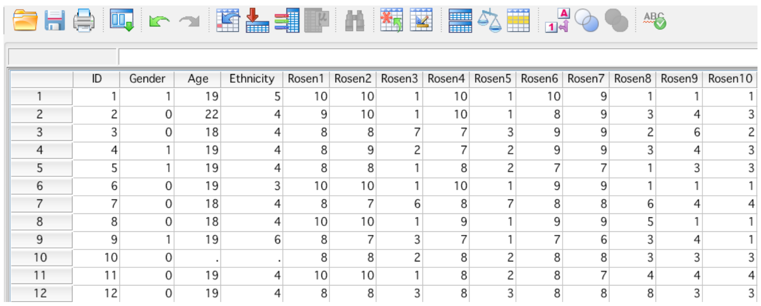
Figure 11

To define values as missing, click in the appropriate cell for that variable under the "Missing" column like you did for the value labels. You will get a dialog box into which you can input the appropriate missing values for that variable. Note that the General Social Survey data set includes quite a lot of missing values. These values are *not* incorporated into calculations involving that variable (e.g., the "99s" wouldn't get incorporated into a mean involving that variable).

Entering data

Once you've defined all the variables you need, you're ready to enter the actual data. Generally, it's easiest to enter all the data from a single participant before moving on to the next participant. That is, work row by row. I find it easiest to use the numerical keypad on the right side of an external keyboard (laptops don't have these, which is why I hate entering data from a laptop!).

Just for illustrative purposes, in Figure 12 you see the first portion of a sample questionnaire data file, showing 12 participants and their scores on the first several variables of a questionnaire.



	ID	Gender	Age	Ethnicity	Rosen1	Rosen2	Rosen3	Rosen4	Rosen5	Rosen6	Rosen7	Rosen8	Rosen9	Rosen10
1	1	1	19	5	10	10	1	10	1	10	9	1	1	1
2	2	0	22	4	9	10	1	10	1	8	9	3	4	3
3	3	0	18	4	8	8	7	7	3	9	9	2	6	2
4	4	1	19	4	8	9	2	7	2	9	9	3	4	3
5	5	1	19	4	8	8	1	8	2	7	7	1	3	3
6	6	0	19	3	10	10	1	10	1	9	9	1	1	1
7	7	0	18	4	8	7	6	8	7	8	8	6	4	4
8	8	0	18	4	10	10	1	9	1	9	9	5	1	1
9	9	1	19	6	8	7	3	7	1	7	6	3	4	1
10	10	0	.	.	8	8	2	8	2	8	8	3	3	3
11	11	0	19	4	10	10	1	8	2	8	7	4	4	4
12	12	0	19	4	8	8	3	8	3	8	8	8	3	3

Figure 12

Checking for errors
This is one of the most important steps of entering and analyzing data. If you made any errors inputting the numbers, it could wreak havoc with your analyses and your interpretation of the data. If the errors are big or numerous, it could mean the difference between supporting and rejecting a hypothesis! Be VERY CAREFUL whenever you enter data, and always double-check everything.

If you've conducted an online survey and have exported data from Qualtrics, you still need to double-check it for errors. For instance, I have sometimes discovered that a scale that was supposed to range from 1 to 6 in Qualtrics actually exported as bizarre numbers instead: 23, 24, 25, 26, 27, 28 instead of 1, 2, 3, 4, 5, 6. This can happen if you make a lot of edits to a survey question by deleting options and then adding them. You can avoid this by recoding the values in Qualtrics before exporting the data, but if you don't notice until later, you will need to recode them in SPSS. The important point is that you need to NOTICE the error in order to correct it, so examining your data is vital.

<u>PART 5: Exploring your data: Conducting basic statistical analyses using SPSS</u>

Now you're ready to start taking a look at your data. Below are described some of the descriptive statistics or graphing options you might be interested in using. Later in the semester, we'll discuss how to use more advanced techniques (e.g., *t* tests, ANOVAs, correlation, & regression).

> **TIP: Changing display preferences**
> Before running any analyses, click on the **Edit** menu, then scroll to the bottom and select **Options**. A window will appear where you will be able to tell SPSS that you want the Variable Lists in any analysis window to display the *names* of the variables (the default in SPSS is to display the labels, which can be cumbersome). Note in Figure 13 below how the variable names appear in the column, rather than long variable labels.
>
> Also: If you click on the Output tab in Options, you can specify that, in your output, you would like variables shown as names & labels, and variable values shown as values & labels. (Do so in both the "Outline Labeling" section and the "Pivot Table Labeling" section on the left of the dialog box.)

The “Frequencies” Command
If you want SPSS to print you a frequency distribution, go under the “Analyze” menu, scroll down to “Descriptive Statistics,” and then choose “Frequencies.” You’ll get a window like the one in Figure 13.

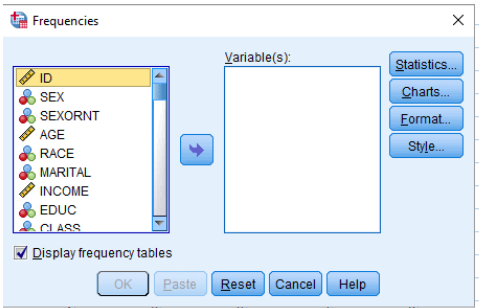


Figure 13

On the left side is a list of all the variables in your data file. Simply click on one you want to examine, then click the arrow to pop them into the “Variable(s)” box.

> **TIP: Selecting multiple variables at once**
> You can select multiple variables at the same time by holding down the **Control** key while clicking on the variable names with the mouse. Alternatively, if the variables you desire are listed in sequence in the data file, you may click on the first variable, hold down the **Shift** key and click on the last variable in the list that you want to examine.

Once you’ve selected the variables you’d like to examine, you need to decide what information you’d like. The default is just the frequency tables (notice the check mark in the box labeled “Display frequency tables”). If you don’t want frequency tables, you can click on the box to remove the checkmark. If you click on the “Statistics” button, you can choose which measures of central tendency & variability you want to use. REMEMBER: You need to figure out what level of measurement a particular variable represents before willy-nilly choosing statistics. For example, calculating a mean gender doesn’t make any sense, because gender is measured on a nominal scale. THE COMPUTER WILL DO WHATEVER YOU TELL IT TO DO, EVEN IF YOU TELL IT TO DO SOMETHING WRONG! So you have to make sure to choose your statistics wisely.

Histograms & Bar Graphs
If you want to look at your data graphically, you might choose to have SPSS draw you a histogram or a bar graph of a particular variable. Remember, histograms are used for quantitative variables (e.g., height), whereas bar graphs are used for qualitative variables (e.g., gender). Simply click on the “Charts” button within the frequencies command and choose which type of graph you want.

The “Descriptives” Command
Another way to get basic descriptive statistics is to go under “Analyze,” “Descriptive Statistics,” “Descriptives.” Here you’ll get many of the same options as with the “Frequencies” command, only you won’t get a frequency table. Use this command if you just want to get a quick mean and standard deviation for a bunch of variables at once.

Moving back and forth between the data window and the output window
Whenever you ask SPSS to calculate something for you, you automatically get bumped into the output window ("SPSS viewer") where the output from your commands (e.g., frequency tables, descriptive statistics, graphs) is displayed. To get back to the data window, simply go up to the "Window" menu, and scroll down until you see the name of the data file. If you want to edit a graph, double-click on it in the output window and you will be bumped into the "Chart Editor." Again, you can return to any other window by going up to the "Window" menu.

KEYBOARD SHORTCUTS FOR TOGGLING BETWEEN OPEN WINDOWS:

- Toggling between windows **within the same program**: On a Macintosh computer, selecting **Command** + ~ (the tilde symbol, top left of your keyboard) will toggle back and forth between open windows within the same program. (Sorry, Windows computers don't have this functionality!)

- Toggling **between *different* programs**: This is useful for, say, going back and forth between SPSS and Microsoft Word if you're writing up your homework. Hold down **Command+tab** on a Mac or **Alt+tab** on a PC.

Save, Save, and Save Again!

I strongly recommend saving your data file and output file every couple of minutes. You don't want to lose hours of work in case of a computer crash or power outage. Save frequently.

Printing Output

Unfortunately, SPSS output prints out in a ridiculously large font with very little information on each page. Printing output "as is" wastes a lot of trees. You can certainly save your output (.spv) files on your computer to look at on your screen, but whenever I ask you to print output for assignments, please do the following:

Take a screen grab copy of the relevant part of the output (**Ctrl+Command+Shift+4** on a Mac; for PCs, use the snipping tool) and paste it **(Command+v)** into your Microsoft Word document. You can then re-size the image to make it smaller to fit on your page.

Alternatively, you can export your output as "rich text format," which will open in Microsoft Word, but this is still fairly cumbersome and uses more trees than necessary. If you can't get the screen grab to work, though, you can do this: From within the output window, go to the "File" menu and choose "Export." You will see the window below (Figure 14). Choose "Word/RTF(*.doc)" as the type (which is the default), click "browse" to decide where to save the file, give the file a descriptive name, and click OK in the Export window. You can now open the saved file in Microsoft Word, reformat the tables if you wish, delete any unnecessary information, and then print your output.

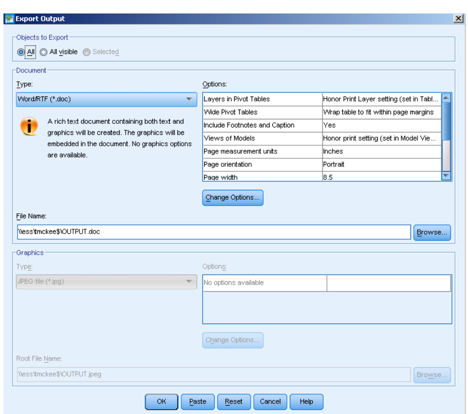


Figure 14

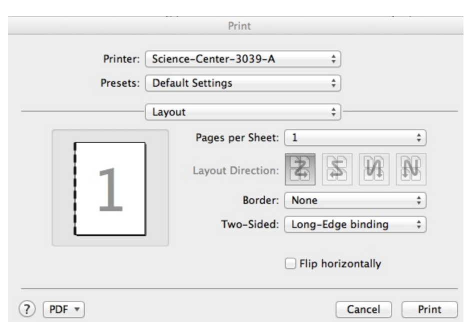


Figure 15

Part 6: Performing basic data manipulations and computing new variables

Recoding Variables

Sometimes it's useful to be able to combine existing categories of responses into broader categories. For example, in the 2016 General Social Survey, the income variable has very narrow categories (1 = under $1,000; 2 = $1,000-$2,999; 3 = $3,000-$3,999; 4 = $4,000 – 4,999, etc…through 26 = $170,000 or over). Perhaps you don't need 26 different income categories, but would rather group them into more meaningful chunks. You could decide that you want the following categories instead:

1 = under $20,000 (this would encompass values 1 – 12 in the old variable)
2 = between $20,000 and $40,000 (values 13 – 17 in the old variable)
3 = between $40,000 and $60,000 (values 18 – 19 in the old variable)
4 = between $60,000 and $90,000 (values 20 – 21 in the old variable; note that there's unfortunately not an even break at $80,000 in the data set)
5 = between $90,000 and $130,000 (values 22 – 23 in the old variable)
6 = over $130,000 (values 24 – 26 in the old variable)

To do this, you would need to recode the original values into new ones. The original values 1-12 would be recoded as a 1 (under $20,000) in the new variable. The original 13-17 would become 2, and so on. To accomplish this in SPSS, go up to the **Transform** menu and scroll down to **Recode** (see Figure 16).

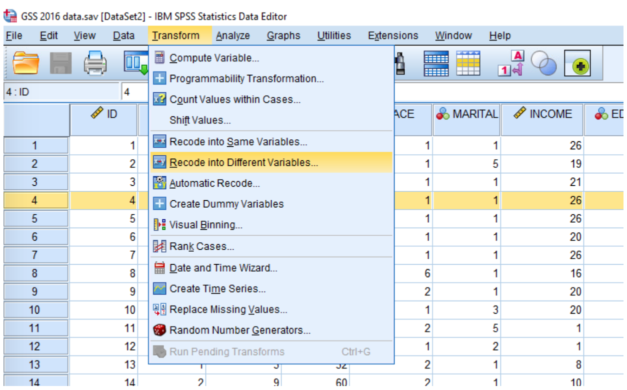


Figure 16

You get two options: "Recode into same variables" and "Recode into different variables." The former changes the existing variable and the latter creates a new variable with the recoded values. I recommend always creating a new (different) variable and leaving the original intact. If you choose "Into different variables," you will get the window you see below:

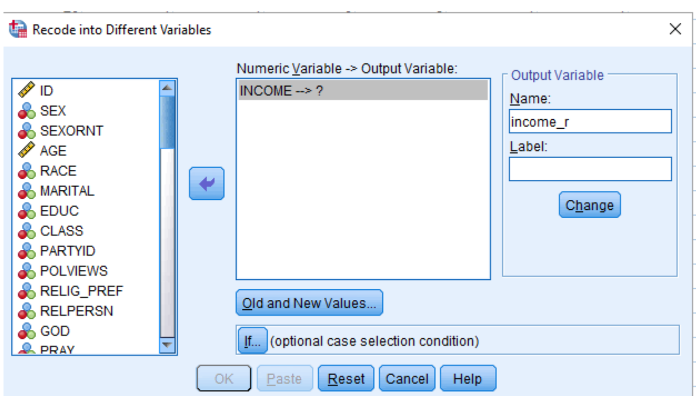


Figure 17

The names of all your variables will appear in the list on the left part of the window. Scroll down until you find the variable(s) you're looking for. In this case, I want to recode the INCOME variable. Simply click on the variable you want, then click on the arrow (which will be pointing the opposite of the way it's pointing in the figure). Now you need to tell SPSS what you want the new (recoded) variables to be called. To make life easier, I usually just add "r" to the original variable name, to remind myself that the variable has been recoded. In this case, income becomes income_r. Type this new name in the "output variable" box on the right of the window and click on the "Change" button.

Now you're ready to tell SPSS <u>how</u> to recode the variable. To do so, first click on the "Old and New Values" box. You'll get the screen you see below. In this case, we want the first 12 income categories of the old variable to become a "1" in the new, recoded variable. Notice that in the dialog box below, I've defined a range from 1-12 for "Old Value". On the right, I've entered a 1 under "New Value". You would then click on "Add," and keep going with the rest of the recodes (e.g., 13-17 becomes a 2, 18-19 becomes a 3, etc.). When you've added all the recodes to the Old→New box, click Continue.

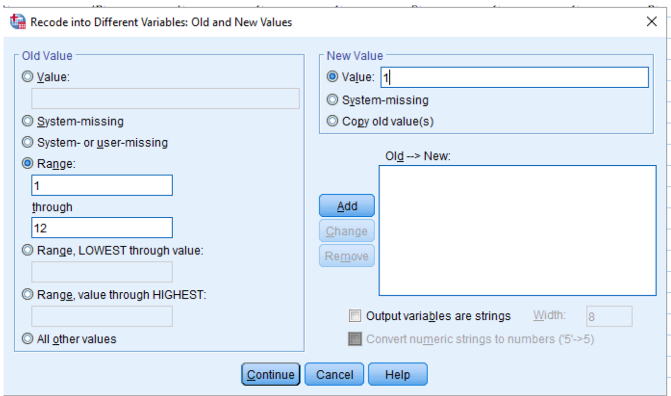


Figure 18

This will take you back to the window in Figure 17, and you can simply click "OK." If you now scroll to the end of your data file, you'll find a new column there called income_r. You can run a frequency on it (Analyze, Descriptive Statistics, Frequencies) to double-check that your categories look good.

Recoding is useful in other instances as well. For instance, questionnaires are often constructed such that approximately half of the questions are framed positively and the other half are framed negatively. For example, the 2016 General Social Survey included 6 items from the Life Orientation Test – Revised, a measure of optimism (see below). These items are scored on a 5-point Likert-type scale, from 1 (strongly disagree) to 5 (strongly agree). Note that items 2, 4, and 5 are reverse-worded, such that high numbers actually mean greater pessimism rather than greater optimism. Researchers generally want to combine items such as these into a single measure for analysis (i.e., a single optimism score that represents the mean of each respondent's rating for all six questions). To do so, we would first have to reverse-score the two reverse-worded items, so that higher numbers always mean higher optimism.

	Name	Type	Width	Decimals	Label	Values	Missing
24	LOTR1	Numeric	8	0	(Life Orientation Test - Revised; measure of Optimism). In uncertain times I usually expect the best.	{0, IAP}...	0, 8, 9
25	LOTR2	Numeric	8	0	If something can go wrong for me it will (R)	{0, IAP}...	0, 8, 9
26	LOTR3	Numeric	8	0	I'm always optimistic about my future	{0, IAP}...	0, 8, 9
27	LOTR4	Numeric	8	0	I hardly ever expect things to go my way (R)	{0, IAP}...	0, 8, 9
28	LOTR5	Numeric	8	0	I rarely count on good things happening to me	{0, IAP}...	0, 8, 9
29	LOTR6	Numeric	8	0	I expect more good things to happen to me than bad	{0, IAP}...	0, 8, 9

Figure 19

We could do this in a similar way to how we recoded the income variable, above. To reverse the response scale for items LOTR2, LOTR4, and LOTR5, we need to recode the values such that 1=5, 2=4, 3=3, 4=2, and 5=1. Since both variables are being recoded in the same way, we can put both of them in the box shown in Figure 17. When we click "Old and New Values," we'll get the following:

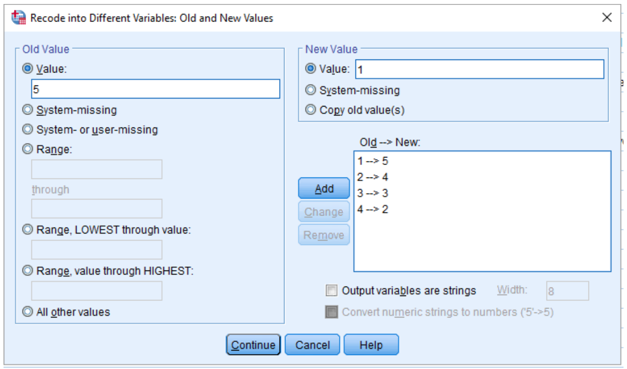


Figure 20

Note that this time we're recoding individual values rather than a range. Note how I've already recoded 1 into 5, 2 into 4, 3 into 3 (probably unnecessary, but I do it just in case….), and 4 into 2, and am in the process of changing 5 into 1.

Computing a new variable from existing ones

Once you have recoded any reverse-worded items, you can compute what is known as a "composite variable" for the questionnaire. That is, you will compute a variable that represents each participant's mean of all the questionnaire items. (Note: Technically you would only do this if you had ascertained that all of the variables were highly intercorrelated with one another, but that's a lesson for another course! Given that this is an established optimism scale, we'll go ahead and compute the mean.)

There are 6 items on the optimism scale, so if I wanted to compute each participant's mean optimism score based on those 6 items, I could use a compute statement. Don't forget that we need to use the recoded items instead of the originals for questions 2, 4, and 5. To create our compute statement, we would go up to the Transform menu again and scroll down to "Compute." You should then see a window that looks like the one below:

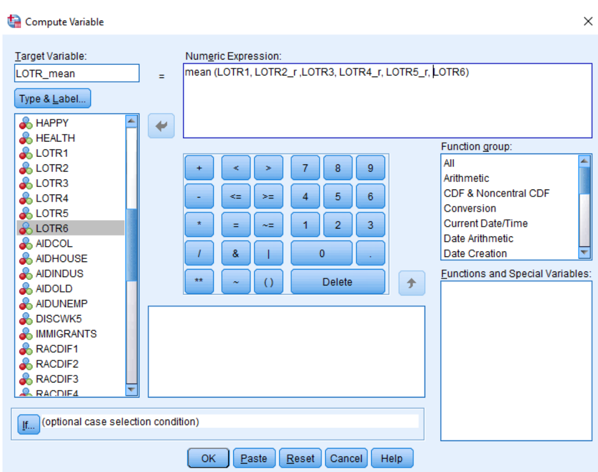


Figure 21

Note how I've typed "LOTR_mean" in the "Target Variable" box. That means I want SPSS to create a new column labeled "LOTR_mean" that will represent each participant's mean on the optimism scale. In the "Numeric Expression" box, I've indicated that I want this new variable to equal the mean of the items on the optimism scale, substituting the recoded items for the originals where appropriate. When you're all set, click "OK." If you scroll to the end of the data file, you'll see a new column labeled "LOTR_mean." This variable would represent each participant's average optimism score based on the 6-item questionnaire.

NOTE: You can use the compute command to perform any of a number of functions. I've chosen to calculate a mean.

TIP: Pasting syntax. If you wanted to save this formula to use again later, you could click "Paste," and the formula would get pasted into the Syntax window, where you could run the command by highlighting it and clicking the green "go" button, and could also save it for later use.

Play around! Explore! Have fun!

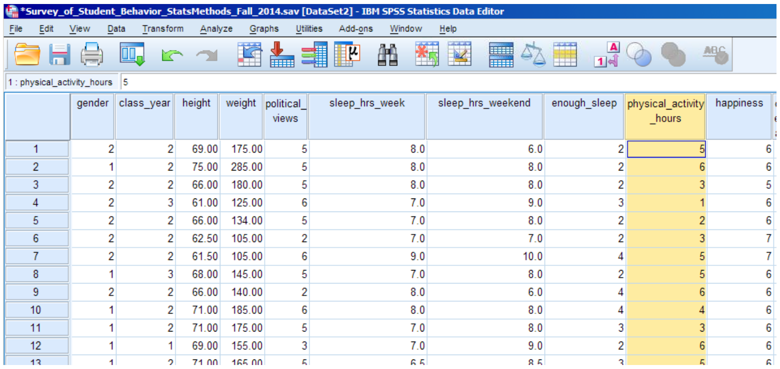
SPSS is a fairly user-friendly program. If you get lost, simply scroll around in the menus until you find what you're looking for. If you're really lost, you can probably Google for help. Remember, though: SPSS will blindly do whatever you tell it to do, regardless of whether or not it's a sensible command. So be smart about what you tell it to do. Now you're ready to explore your data. Enjoy!

Part 7: Ending your session with Citrix

When you are done working, remember to **save** your files. From within SPSS, choose **File**, then **Exit**. This will automatically log you off of Citrix as well. (If you forget to do this and instead quit Citrix without first quitting SPSS, it is very important to select log off rather than disconnect so you don't have multiple Citrix sessions open.)

SPSS Instructions for a One-Sample *t* Test

We will use data from a class questionnaire distributed in a previous semester of Psych Stats to illustrate how to conduct a one-sample *t* test. We will examine the number of hours of sleep per week that students get, and whether this number differs significantly from the recommended 8 hours. The first several cases of the file appear as follows:



	gender	class_year	height	weight	political_views	sleep_hrs_week	sleep_hrs_weekend	enough_sleep	physical_activity_hours	happiness
1	2	2	69.00	175.00	5	8.0	6.0	2	5	6
2	1	2	75.00	285.00	5	8.0	8.0	2	6	6
3	2	2	66.00	180.00	5	8.0	8.0	2	3	5
4	2	3	61.00	125.00	6	7.0	9.0	3	1	6
5	2	2	66.00	134.00	5	7.0	8.0	2	2	6
6	2	2	62.50	105.00	2	7.0	7.0	2	3	7
7	2	2	61.50	105.00	6	9.0	10.0	4	5	7
8	1	3	68.00	145.00	5	7.0	8.0	2	5	6
9	2	2	66.00	140.00	2	8.0	6.0	4	6	6
10	1	2	71.00	185.00	6	8.0	8.0	4	4	6
11	1	2	71.00	175.00	5	7.0	8.0	3	3	6
12	1	1	69.00	155.00	3	7.0	9.0	2	6	6

Figure 1

To conduct a one-sample *t* test in SPSS, go to the "Analyze" menu and click on "Compare means" and then "One-Sample T Test." Click on the DV you want to analyze (in this case, "sleep_hrs_week") and click on the arrow to put it in the "Test Variable(s)" box (see Figure 2).

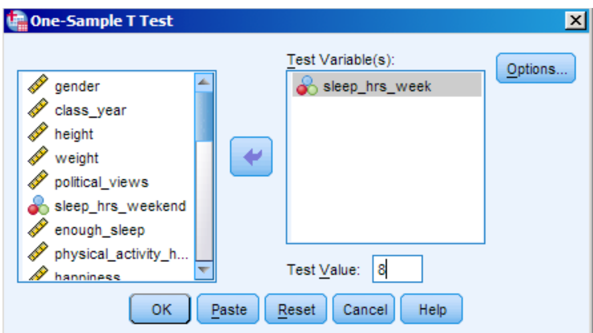


Figure 2

Next, you need to decide what your test value is going to be. This is the value to which you will compare the overall mean of the distribution. For our example, the test value is 8 hours, as we want to know whether the mean number of hours of sleep in our sample differs from 8. Type "8" in the "Test Value" box (see Figure 2). You will get the output shown on the next page:

T-Test

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
How many hours of sleep do you get on an average week night (Sunday through Thursday)?	128	7.328	.9525	.0842

One-Sample Test

	Test Value = 8					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
How many hours of sleep do you get on an average week night (Sunday through Thursday)?	-7.981	127	.000	-.6719	-.838	-.505

The top box gives you the overall *M* and *SD* of the distribution. Here, we see that students in this sample get an average of 7.33 hours of sleep per weeknight. Is this number significantly different from the recommended 8 hours of sleep? To find out, look in the second box. You will find the *t* value (-7.98; notice that it is negative because it is *less* than the test value of 8), the df (N – 1), and the "Sig. (2-tailed)" value, which represents the *p* value. It does indeed appear that the students in this sample get significantly less sleep during the week than 8 hours per night.

The 95% confidence interval of the difference is also shown; this interval does not contain 0 (i.e., no difference between observed mean and test value). If you wanted to compute the 95% confidence interval around the observed mean, you could do so with the information present in the table as well as the *t* table in your Stats text. The lower confidence limit would be 7.33 – $t_{.05}$(.08), where the $t_{.05}$ represents the critical value of *t* from the *t* table in your book for a non-directional test with an alpha of .05 and $N = 120$ (the last case before *N* approaches infinity). This critical value is 1.98. So, the lower confidence limit (rounded to 2 decimal places) would be 7.33 – 1.98 (.08) = 7.17. The upper confidence limit (rounded to 2 decimal places) would be 7.33 + 1.98 (.08) = 7.49. We can therefore be 95% certain that the true value of the population mean falls between 7.17 and 7.49, an interval that does NOT contain our test value of 8 hours.

Sample Results Section:

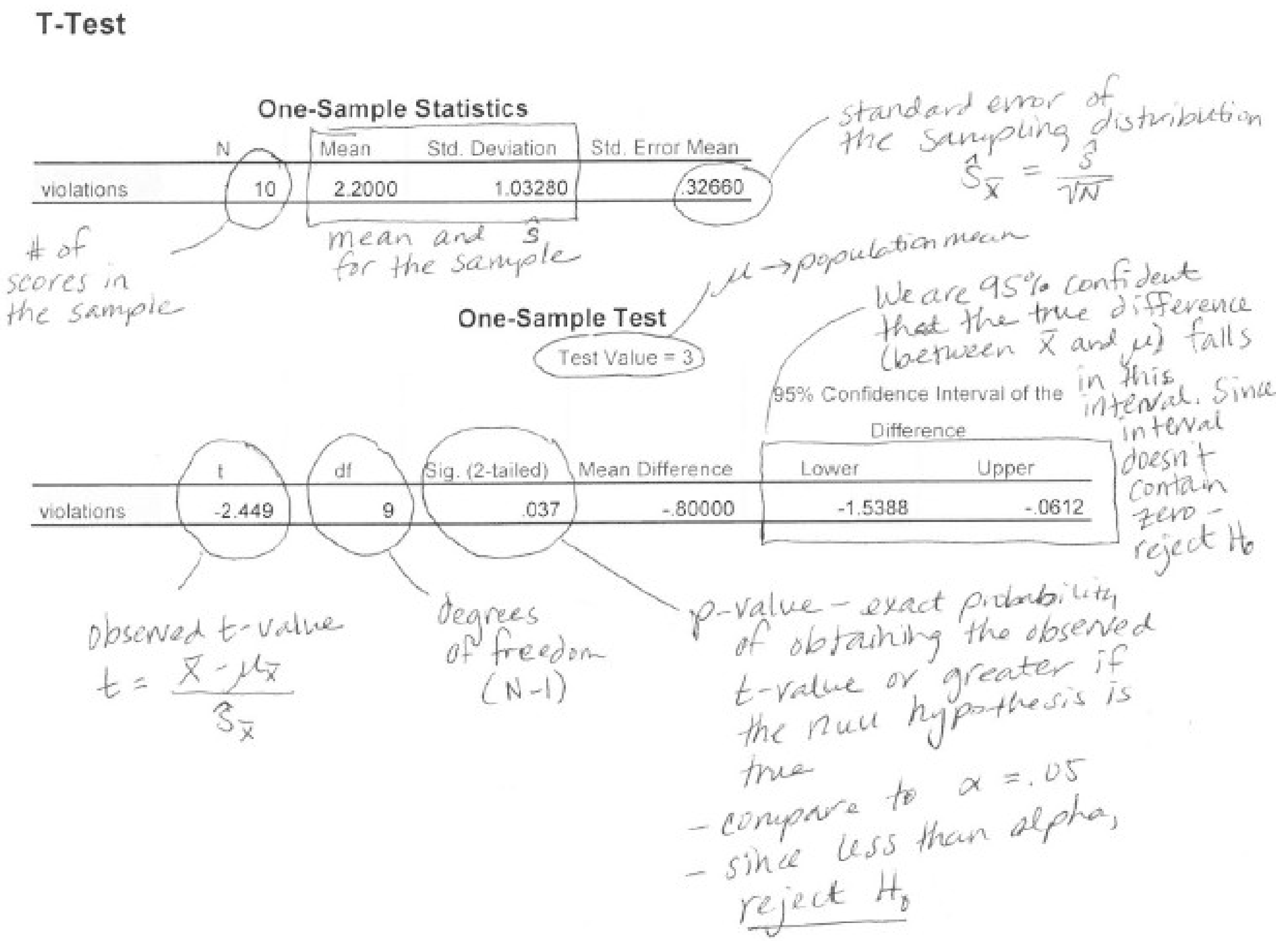
A one-sample *t* test was conducted to determine whether Hamilton students get more or less than the recommended 8 hours of sleep per night during the week. The mean number of hours of sleep per weeknight reported by the sample ($M = 7.33$, $SD = 0.95$) was statistically significantly lower than 8 hours, $t(127) = -7.98$, $p < .001$; 95% CI [7.17, 7.49].

Important details for writing a Results section involving a one-sample *t* test:

- You should italicize *M*, *SD*, *t*, and *p*.
- Current APA Style is to report the exact *p* value (e.g., $p = .004$) unless it is < .001.
- When $p > .05$, report that the difference was NONsignificant, not INsignificant. Report the exact *p* value even if the test is not significant.
- Round everything to 2 decimal places, except the *p* value, which can be up to 3 decimal places.
- Be sure to state whether the observed mean was higher or lower than the test value for any significant test.
- Don't forget to report the 95% confidence interval.

Sample SPSS Annotation

When you are asked to annotate your SPSS output for your homework assignments, you should do so in a manner similar to the example shown below.



T-Test

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
violations	10	2.2000	1.03280	.32660

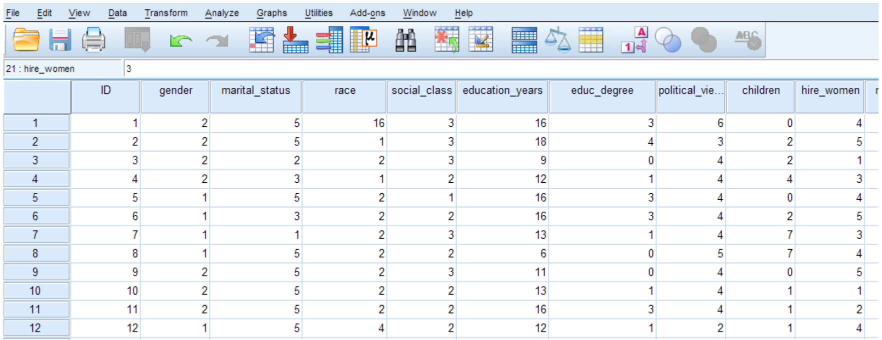
One-Sample Test

Test Value = 3

	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
violations	-2.449	9	.037	-.80000	-1.5388	-.0612

SPSS Instructions for Independent Samples *t* Test

This handout uses data from the National Opinion Research Center General Social Survey to illustrate how to conduct an independent samples *t* test. We will examine gender differences in people's belief that companies should hire and promote women. The first several cases of the file appear as follows:



	ID	gender	marital_status	race	social_class	education_years	educ_degree	political_vie...	children	hire_women
1	1	2	5	16	3	16	3	6	0	4
2	2	2	5	1	3	18	4	3	2	5
3	3	2	2	2	3	9	0	4	2	1
4	4	2	3	1	2	12	1	4	4	3
5	5	1	5	2	1	16	3	4	0	4
6	6	1	3	2	2	16	3	4	2	5
7	7	1	1	2	3	13	1	4	7	3
8	8	1	5	2	2	6	0	5	7	4
9	9	2	5	2	3	11	0	4	0	5
10	10	2	5	2	2	13	1	4	1	1
11	11	2	5	2	2	16	3	4	1	2
12	12	1	5	4	2	12	1	2	1	4

Figure 1

To conduct an independent samples *t* test, go to the "Analyze" menu and click on "Compare means" and then "Independent-Samples T Test." Click on the DV you want to analyze (in this case, "hire_women") and click on the arrow to put it in the "Test Variable(s)" box. Click on your IV (gender) and put it in the "Grouping Variable" box (see Figure 2).

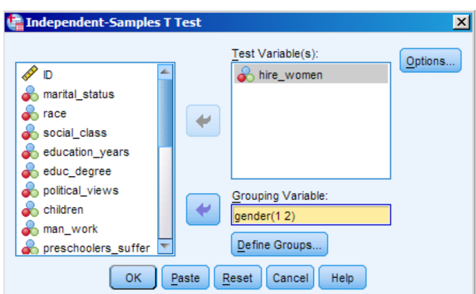


Figure 2

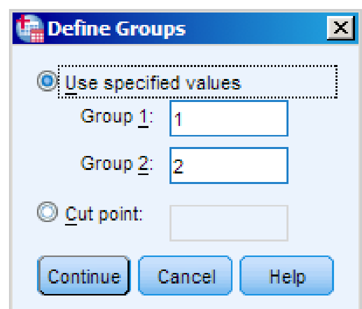


Figure 3

Notice the question marks next to gender in the Grouping Variable box. Click on "Define Groups" (you'll see the box in Figure 3) to tell SPSS what numbers are used to represent each of the two groups (e.g., 1 = male, 2 = female). Once you've done so, click "Continue" and then "OK." SPSS will bump you to the Output window where you can examine the results of the *t* test (see output on next page).

Group Statistics

	RESPONDENT'S GENDER	N	Mean
SHOULD HIRE AND PROMOTE WOMEN	MALE	373	3.43
	FEMALE	294	3.74

Notice that SPSS computes the means and standard deviations for the two levels of your independent variable.

Before you examine the results of the actual *t* test, you should check the results for **Levene's Test for Equality of Variances**, which assesses whether the variances of the two groups are significantly different from each other (i.e., whether the homogeneity of variance assumption has been violated). If the *p* value for the Levene's test (in the "Sig." column) is greater than .05, then the variances are *not* significantly different from one another (i.e., the homogeneity of variance assumption has been satisfied), and you may use the *t* value and degrees of freedom in the row marked "equal variances assumed." If the significance value for Levene's Test is less than .05, you will instead need to use the values reported in the row labeled "equal variances not assumed." Even if Levene's Test is significant, it is acceptable to use the df from the "equal variances assumed" row.
In the output shown here, the variances are significantly different, so we need to use the *t* test results from the row labeled "equal variances not assumed." We see that the observed *t* is -3.69, and the *p* value is .000 (which should be reported as $p < .001$). Since $p < .05$, this test is statistically significant. We'll now look at the means to determine which group had the higher mean. It looks like female participants agree that companies should hire and promote women to a significantly greater extent ($M = 3.74$, $SD = 1.05$) than do male participants ($M = 3.43$, $SD = 1.10$). When you're writing the results of an independent samples *t* test, be sure to include the *t* (rounded to two decimal places), df, *p* value, the *M* and *SD* for each group, and the strength of the relationship.

Sample Results Section:

An independent groups *t* test compared male and female participants' agreement with the belief that companies should hire and promote women. The test was statistically significant, $t\ (665) = -3.69$, $p < .001$, indicating that female participants more strongly agreed with the statement ($M = 3.74$, $SD = 1.05$) than male participants did ($M = 3.43$, $SD = 1.10$); 95% CI [-0.47, -0.14]. The strength of this relationship, as indexed by eta^2, was small at .02.

Important details for writing a Results section involving an independent samples *t* test:

- You should italicize *M*, *SD*, *t*, and *p*.
- Current APA Style is to report the exact *p* value (e.g., $p = .004$) unless it is < .001.
- When $p > .05$, report that the difference was NONsignificant, not INsignificant. Report the exact *p* value even if the test is not significant.
- Round everything to 2 decimal places, except the *p* value, which can be up to 3 decimal places.
- Be sure to state the nature of any significant effect (which mean is higher).
- Don't forget to report eta-squared (effect size) for any significant effect.
- State the 95% confidence interval.
-

SPSS Instructions for a One-way ANOVA

Imagine we want to know whether providing a jury with information about a defendant's prior criminal record influences the extent to which they believe the defendant is guilty. Participants heard a recording of a case about a woman accused of writing bad checks, and were randomly assigned to one of three groups: Group 1 was told she had previously bounced checks (i.e., she had a criminal record); Group 2 was told that this was her first offense (i.e., she had a clean record); and Group 3 was given no information (control). The DV was a 1-10 rating of how guilty participants thought the defendant was. Because the DV is at least interval and the IV is between-subjects with more than 2 groups, a one-way ANOVA is appropriate.

To run a one-way ANOVA in SPSS, go under the "Analyze" menu to "Compare means" to "One-way ANOVA…" You will get the following window:

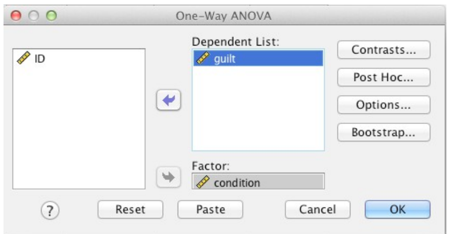


Figure 1

Click on the variable you want to use as your IV ("condition" in the above example) and move it to the "Factor" box by clicking on the appropriate arrow. Next, click on the variable you want to use as your DV ("guilt") and put it in the box labeled "Dependent List." If you wanted, you could put a whole bunch of DVs in the box at the same time, and SPSS would run a separate ANOVA for each DV.

Now you need to tell SPSS which post-hoc multiple comparison procedure(s) you want to use (to compare pairs of means if the overall *F* is significant). Click on the box labeled "Post Hoc." SPSS will give you the window in Figure 2 below.

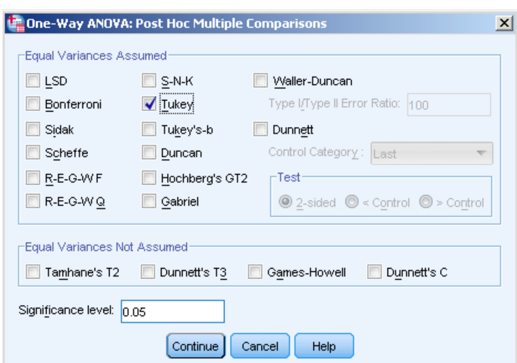


Figure 2

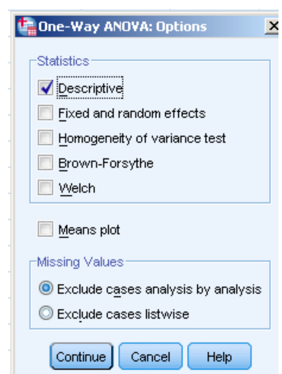


Figure 3

You are given the option to choose among various post-hoc tests. Tukey's HSD is the most widely-used procedure, so go ahead and click on the box next to "Tukey" (NOT "Tukey-b"). Click "Continue" to get back to the ANOVA window.

Next, you want to make sure that you get descriptive information printed out (means and standard deviations for each condition), as well as the labels of each condition. To do that, from the "One-way ANOVA" window, click on "Options." When the window comes up, click in the box labeled "Descriptive" (See Figure 3 above) and then click "Continue." When you get back to the main ANOVA window, click "OK". Examine your output.

The first part of your output provides the Ms and SDs for each condition. Note also that you are given the standard errors for each condition, which are equal to the SD divided by the square root of n (in this case, 5). The confidence intervals here are around each mean, NOT around the difference between means (those are in a later piece of the output).

Descriptives

rating of defendant's guilt

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
criminal record	5	8.00	2.121	.949	5.37	10.63	5	10
clean record	5	4.00	2.236	1.000	1.22	6.78	1	7
no information	5	5.00	2.550	1.140	1.83	8.17	3	9
Total	15	5.67	2.769	.715	4.13	7.20	1	10

Next, you get your basic ANOVA source table, as shown below. You would write F (2, 12) = 4.06, p = .045. Note that you can use the source table to compute eta^2 if your finding is significant (as it is here, because p < .05). Eta^2 = 43.333/107.333 = .40 (a large effect).

ANOVA

rating of defendant's guilt

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	43.333	2	21.667	4.063	.045
Within Groups	64.000	12	5.333		
Total	107.333	14			

Given that the overall F is significant, you would go ahead and examine the post-hoc test output to determine which means are significantly different from one another. SPSS provides this information two different ways; the first appears below:

Post Hoc Tests

Multiple Comparisons

Dependent Variable: rating of defendant's guilt

Tukey HSD

(I) experimental condition	(J) experimental condition	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
criminal record	clean record	4.000*	1.461	.044	.10	7.90
	no information	3.000	1.461	.142	-.90	6.90
clean record	criminal record	-4.000*	1.461	.044	-7.90	-.10
	no information	-1.000	1.461	.777	-4.90	2.90
no information	criminal record	-3.000	1.461	.142	-6.90	.90
	clean record	1.000	1.461	.777	-2.90	4.90

*. The mean difference is significant at the 0.05 level.

In the above output, you can examine the size and significance of the three pairwise comparisons:
Criminal vs. clean (p = .044)
Criminal vs. no info (p = .142)
Clean vs. no info (p = .777)

Note that the same information is provided twice (e.g., criminal vs. clean, then clean vs. criminal), so you can ignore the duplicates. You can also use this table to find the 95% confidence interval around the difference between each pair of means (e.g., the 95% CI around the difference between criminal and clean is .10 – 7.90).

SPSS also provides another way of examining the differences between means:

Homogeneous Subsets

rating of defendant's guilt

Tukey HSD[a]

experimental condition	N	Subset for alpha = 0.05	
		1	2
clean record	5	4.00	
no information	5	5.00	5.00
criminal record	5		8.00
Sig.		.777	.142

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 5.000.

Means in different subsets differ at *p* < .05, whereas means in the same subset do not. In the table above, clean record (*M* = 4.00) is in Subset 1 and criminal record (*M* = 8.00) is in Subset 2, so we know that they significantly differ. However, no information (*M* = 5.00) is in Subsets 1 AND 2, which means that it does not differ from either clean or criminal record. I find this way of interpreting the differences easier in terms of explaining the overall pattern of results in the simplest way.

Sample Results Section:

A one-way ANOVA on guilt ratings as a function of prior information about a defendant was statistically significant, *F* (2, 12) = 4.06, *p* = .045. The strength of the relationship, as indexed by eta^2, was .40, indicating a strong effect. A Tukey HSD test indicated that mean guilt ratings were significantly higher for participants told a defendant had a prior criminal record (*M* = 8.00, *SD* = 2.12) than for those told she had a clean record (*M* = 4.00, *SD* = 2.24); 95% CI [.10, 7.90]. Guilt ratings for those not given information about a prior criminal record (*M* = 5.00, *SD* = 2.55) did not differ from ratings in the criminal record and clean record conditions; 95% CIs [-.90, 6.90] and [-4.90, 2.90], respectively.

Important details for writing a Results section involving a one-way ANOVA:

- You should italicize *M*, *SD*, *F*, and *p*.
- Don't forget to write the degrees of freedom, between and within, in parentheses.
- Round *M*, *SD*, and *F* to 2 decimal places. You should report *p* to 3 decimal places. If SPSS notes that *p* = .000, then report *p* < .001. Otherwise, provide the exact *p* value, even for a non-significant *F*.
- First, indicate whether the overall *F* was significant, and then, discuss the results of the post hoc test (be sure to state which post hoc test you used). In describing the pattern of results, be sure to state the *M*s and *SD*s of each group in parentheses, and don't forget to report the 95% confidence intervals for the differences between group means.
- There is no need to provide the *M* and *SD* for the same group twice (i.e., if you are comparing that group to two different groups, you only need to provide the descriptive statistics the first time you mention that group).
- Make sure to include information about the strength of the relationship if you indeed found a significant *F*. [You need to calculate eta^2 by hand using numbers from the source table.]

SPSS Instructions for a Correlated Groups *t* Test

Example: Imagine that a researcher was interested in the effect of sleep deprivation on motor skills performance. Five participants were tested on a motor-skills task after 24 hours of sleep deprivation and again after 36 hours. The DV is the number of errors made on the motor skills task.

The data for this example appear below. Note how these data are entered differently from the way data are entered for an independent groups *t* test. In the repeated-measures case, each level of the IV is represented as a separate column.

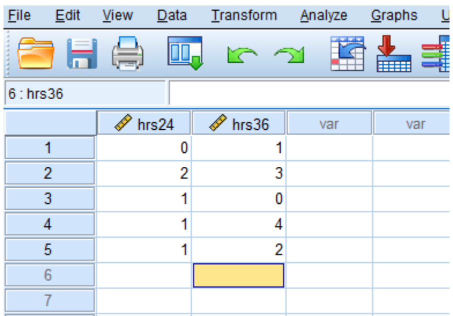


Figure 1

To conduct a correlated groups *t* tests in SPSS, go under the “Analyze” menu to “Compare Means” to “Paired-Samples T Test.” You will get a window similar to the following:

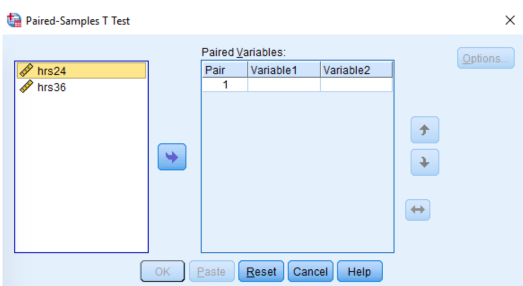


Figure 2

You then want to choose which variables to compare by clicking on them and putting them in the “Paired Variables” box (if you have multiple comparisons to make, you can create multiple pairs of variables). If you shift+click on both, you can add them both at once. Then click “OK” and you will get bumped to the output window where you can examine your findings (see below).

Descriptive statistics for the two levels of your IV:

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	hrs24	1.00	5	.707	.316
	hrs36	2.00	5	1.581	.707

Ignore the next table. It represents the correlation (*r)* between the two variables.

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	hrs24 & hrs36	5	.447	.450

This table provides you with the difference between the two means (-1), the observed *t*, df, *p*-value. And 95% CI.

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1	hrs24 - hrs36	-1.000	1.414	.632	-2.756	.756	-1.581	4	.189

We see that the observed *t*, with 4 df, is -1.58, and the *p* value is .189. Because *p* >.05, this test is not statistically significant. If we *did* find a statistically significant *t*-value, then we would look at the means for the two conditions to determine the nature of the relationship and report strength of the relationship.

Sample Results Section:

A correlated groups *t* test that compared average number of effors on a motor skills task after 24 hours (*M* = 1.00, *SD* = .71) and 36 hours (*M* = 2.00, *SD* = 1.58) of sleep deprivation was statistically nonsignificant, *t* (4) = -1.58, *p* = .189; 95% CI [-2.76, 0.76].

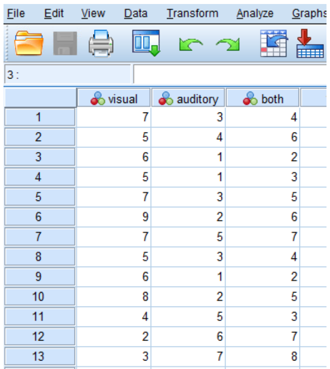
Important details for writing a Results section involving a correlated groups *t* test:

- You should italicize *M*, *SD*, *t*, and *p*.
- Current APA Style is to report the exact *p* value (e.g., *p* = .004) unless it is < .001.
- When *p* > .05, report that the difference was NONsignificant, not INsignificant.
- Round everything to 2 decimal places, except the *p* value, which can be up to 3 decimal places.
- Be sure to state the nature of any significant effect (which mean was higher).
- Be sure to report eta-squared (strength of the effect) for any significant effect.
- Don't forget to report the 95% confidence interval for the mean difference.

SPSS Instructions for a Repeated Measures ANOVA

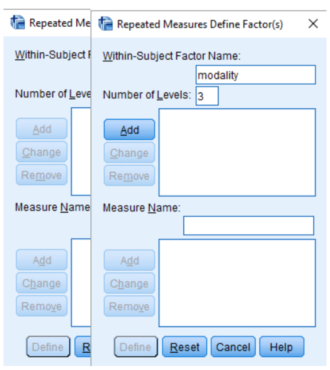
Example: Imagine that a researcher was interested in the effect of presentation modality on word recall. Words were presented to participants in all three of the following modalities: visual only, auditory only, both visual and auditory. The DV was the number of words recalled in each modality.

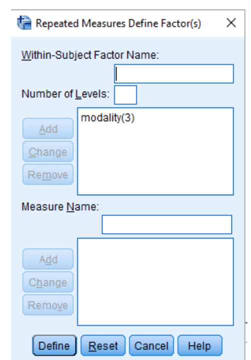
The data for this example appear below. Note how these data are entered differently from the way data are entered for a one-way between-subjects ANOVA. In the repeated-measures case, each level of the IV is represented as a separate column.



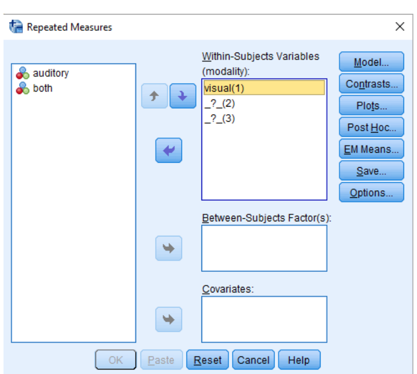
	visual	auditory	both
1	7	3	4
2	5	4	6
3	6	1	2
4	5	1	3
5	7	3	5
6	9	2	6
7	7	5	7
8	5	3	4
9	6	1	2
10	8	2	5
11	4	5	3
12	2	6	7
13	3	7	8

Run a one-way repeated-measures ANOVA in SPSS by going under the "Analyze" menu to "General Linear Model" then "Repeated Measures…" You will get the first window below:

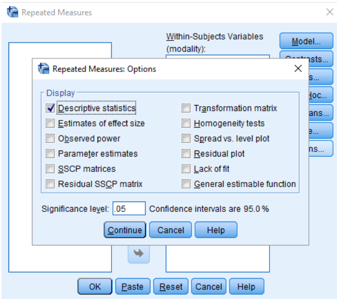




As noted above, in your data window, there is not one single column in which your DV can be found, but rather, three columns (visual, auditory, both). Thus, you will need to tell SPSS what the name of your within-subjects variable (factor) is. In the middle window above, I have named this factor "modality" and noted that it has 3 levels. Next, I clicked on "Add" (see window on right above). Then I selected "Define" to define my within-subjects variable:



Simply click on the name of the level of the IV in the left column, then click on the arrow, and it will bump into the right-hand box. Do this for all three levels of the IV (notice in the window that I have done it only for the first one so far). If you click on the first one, then shift+click on the last one, you can select all three at once. Next, click on “Options” so you can ask SPSS for descriptive statistics (means and *SD*s for each condition):



Click on the box below labeled “Descriptive statistics.” Then click “Continue.” Now click on “OK” in the “Repeated Measures” box. You should now get bumped to the output window, where you can examine your findings. SPSS produces a lot of output for the repeated-measures ANOVA, and you can ignore much of it. Note that you will be using the *F* value from the “Tests of Within-Subjects Effects” box (NOT the “Multivariate Tests” box).

As always, you get Ms and SDs for each condition. Note that, because this was a within-subjects design, the total sample size is 13 (i.e., the same 13 people were in each condition). Also note that the SD for the “both” condition would be rounded to 2.04 (it ends in exactly 5, so you would round to make it an even number).

Descriptive Statistics

	Mean	Std. Deviation	N
visual	5.92	2.060	13
auditory	3.15	1.994	13
both	4.85	2.035	13

You can ignore the following table:

Multivariate Tests[a]

Effect		Value	F	Hypothesis df	Error df	Sig.
modality	Pillai's Trace	.567	7.207[b]	2.000	11.000	.010
	Wilks' Lambda	.433	7.207[b]	2.000	11.000	.010
	Hotelling's Trace	1.310	7.207[b]	2.000	11.000	.010
	Roy's Largest Root	1.310	7.207[b]	2.000	11.000	.010

a. Design: Intercept
Within Subjects Design: modality

b. Exact statistic

If Mauchly’s Test is significant (*p*-value in the Sig. column is less than .05), then the sphericity assumption has been violated. Be sure to report in your APA-style write-up if the sphericity assumption was violated.

Mauchly's Test of Sphericity[a]

Measure: MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[b] Greenhouse-Geisser	Huynh-Feldt	Lower-bound
modality	.404	9.983	2	.007	.626	.665	.500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept
Within Subjects Design: modality

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Here is your ANOVA source table:

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
modality	Sphericity Assumed	50.667	2	25.333	6.247	.007
	Greenhouse-Geisser	50.667	1.253	40.444	6.247	.019
	Huynh-Feldt	50.667	1.329	38.117	6.247	.017
	Lower-bound	50.667	1.000	50.667	6.247	.028
Error(modality)	Sphericity Assumed	97.333	24	4.056		
	Greenhouse-Geisser	97.333	15.033	6.475		
	Huynh-Feldt	97.333	15.951	6.102		
	Lower-bound	97.333	12.000	8.111		

To determine if your *F* is significant, examine the appropriate *p* value in the source table depending on whether or not sphericity is violated (i.e., if sphericity is violated, then use the *p* value in the row labeled "Greenhouse-Geisser" or "Huynh-Feldt"; if sphericity is not violated, then use the *p* value in the row labeled "Sphericity Assumed"). In your APA-style write-up, you can report the degrees of freedom associated with the "Sphericity Assumed" row even if you are using one of the correction factors. You simply need to report the appropriate *p* value that corresponds to the correction factor you chose. For this class, we will use the Huynh-Feldt correction.

Note that SPSS does not bother printing out the total SS or total df, so the table looks a little different than when you compute the ANOVA by hand. Remember that $eta^2 = SS_{IV}/SS_{IV} + SS_{total}$. So in this case, $eta^2 =$ 50.667/50.667+97.333 = .34, a large effect.

All remaining SPSS output can be ignored!

Whenever you conduct a repeated-measures ANOVA in SPSS, you will examine the post-hoc differences between means by conducting a series of correlated groups *t* tests in SPSS, one for each pairwise comparison you wish to make. You would then use a modified Bonferroni procedure (see pp. 432-3 of your textbook) to control Type I error rate.

To conduct the correlated groups *t* tests in SPSS, go under the "Analyze" menu to "Compare Means" then "Paired-Samples t Test" (refer to the earlier section in this guide if you've forgotten). Remember, you can conduct all 3 correlated groups *t* tests at once in SPSS. In this example, you would conduct three correlated groups *t* tests: comparing visual to auditory, visual to both, and auditory to both. On a Macintosh keyboard, holding down the Command key while selecting non-sequential variables will allow you to select them both at the same time.

Click on "OK" once all of your pairs are in the "Paired Variables" part of the box and you will get output that provides you with a table containing the *t* values and *p* values (see below).

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1	visual - auditory	2.769	3.632	1.007	.574	4.964	2.749	12	.018
Pair 2	visual - both	1.077	2.957	.820	-.710	2.864	1.313	12	.214
Pair 3	auditory - both	-1.692	1.548	.429	-2.628	-.757	-3.941	12	.002

You then place the comparisons in order from the highest to lowest *t* value (in absolute value terms), and perform the modified Bonferroni procedure as follows:

Null hyp. tested	Absolute value of *t*	*p* value	Critical alpha	Null hyp. rejected?
$\mu_a = \mu_b$	3.941	.002	.05/3 = .017	Yes
$\mu_v = \mu_a$	2.749	.018	.05/2 = .025	Yes
$\mu_v = \mu_b$	1.313	.214	.05/1 = .050	No

If the observed *p* value is smaller than the critical alpha, then you can reject the null hypothesis. Thus, it appears that average recall was significantly better when words were presented in both modalities than when they were presented in the auditory modality, but that average recall when words were presented in the visual modality was not significantly different from average recall for words presented in the auditory modality or in both modalities.

Note that if you had a study with four conditions instead of just three, you would need to conduct six different pairwise comparisons. Your critical alphas would need to be adjusted accordingly (.05/6, .05/5, .05/4, .05/3, .05/2, .05).

Sample Results Section:

A one-way repeated measures ANOVA was conducted comparing word recall across modality of presentation (visual only, auditory only, or both visual and auditory). Because the sphericity assumption was violated, we used the Huynh-Feldt correction, which indicated a significant effect, F (2, 24) = 6.25, p = .007. The strength of the relationship, as indexed by eta^2, was .34, a strong effect. A series of correlated groups t tests using a modified Bonferroni procedure to control for the overall alpha level indicated that average recall was significantly worse when words were presented auditorially (M = 3.15, SD = 1.99) than when they were presented either visually (M = 5.92, SD = 2.06) or both auditorially and visually (M = 4.85, SD = 2.04); 95% CIs [0.57, 4.96] and [-2.63, -0.76], respectively. Recall did not differ between words presented just visually or in both modalities combined; 95% CI [-0.71, 2.86].

SPSS Instructions for Correlation and Regression

We will use data from a previous semester's Statistics class questionnaire to illustrate how to conduct a correlation analysis and a regression analysis. We will examine the correlation between number of hours of TV watched per week and cumulative GPA.

Correlation

Before you conduct a Pearson's *r*, it is very important to examine a scatterplot of the two variables to get an idea about whether they are linearly related to one another. In SPSS, go under the "Graphs" menu to "Legacy Dialogs" and choose "Scatter/Dot…" You will see the window in Figure 1. Choose "Simple Scatter" (it should already be selected) and click on "Define." You will now see a window like the one in Figure 2.

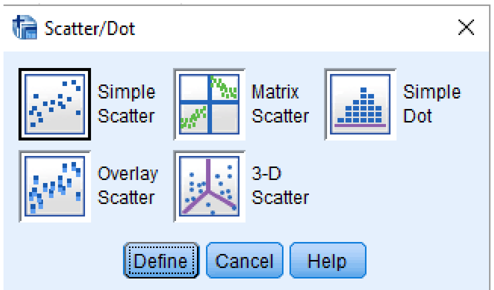


Figure 1

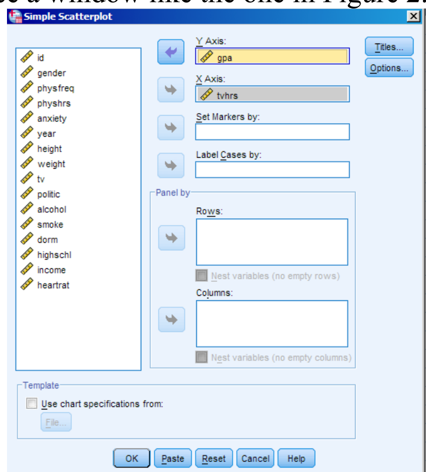


Figure 2

Scroll down the list of variables on the left until you find the variable you want on the Y axis (the criterion variable; in this case, "gpa"). Click on this variable, then click the top arrow. The variable name should now appear in the box labeled "Y Axis." Now find the variable you want on the X axis (the predictor variable; in this case, "tvhrs"). Click on this variable, then click on the arrow next to the "X axis" box. Click OK. SPSS will bump you into the output window, where you'll see your scatterplot. Take a look at it. Are there any outliers? Does the relationship appear linear? (If it's curvilinear, a Pearson's *r* is inappropriate to use.) Double-click on the graph, and when SPSS bumps you into the chart editor, go under the "Elements" menu and choose "Fit line at total." [You can close the "Properties" box if it pops up.] SPSS will draw in the least-squares line, which will better help you see whether the data are linearly related, and whether the relationship is positive or negative.

If the scatterplot looks OK, you are ready to proceed to calculating the correlation. To do so, go to the "Analyze" menu, scroll down to "Correlate" and choose "Bivariate." You will see a window like the one below:

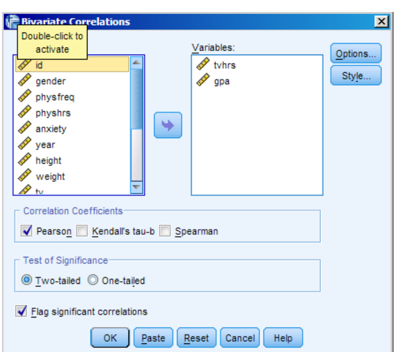


Figure 3

Find the two variables you used in the scatterplot. Select each and place them both in the "Variables" box. Where it says, "Correlation Coefficients," make sure that "Pearson" is selected (this is the default). Under "Options," choose to have SPSS compute the means and standard deviations for the two variables. Now click OK. SPSS will bump you into the output window and produce a correlation matrix (see Figure 4).

Correlations

Correlations

		Number of hours/week watching TV	Cumulative Hamilton GPA
Number of hours/week watching TV	Pearson Correlation	1	-.465*
	Sig. (2-tailed)		.017
	N	26	26
Cumulative Hamilton GPA	Pearson Correlation	-.465*	1
	Sig. (2-tailed)	.017	
	N	26	26

*. Correlation is significant at the 0.05 level (2-tailed).

Figure 4

You'll note that this matrix produces redundant information, so you only need to look at half of it. In each box the top value is Pearson's *r* (the correlation). Below that, the "Sig (2-tailed)" value is the *p* value, or the probability of getting a correlation that big or bigger if there really was no relationship between the two variables in the population. Finally, the bottom number is *N* (the sample size). Note that you could put as many variables into a correlation matrix as you like (not just two).

Sample Results Section:

A Pearson correlation indicated the number of hours per week of television that participants watched (*M* = 3.86, *SD* = 3.08) was significantly negatively correlated with their cumulative GPA (*M* = 86.62, *SD* = 6.09), *r* (24) = -.47, *p* = .017, such that higher levels of television viewing were associated with a lower GPA. The strength of this correlation, as indexed by r^2, was .22, indicating a strong relationship.

Note that for a correlation, df = N – 2. When you're writing the results of a correlation, be sure to include the *r* (rounded to two decimal places), df, *p* value, and the *M* and *SD* for each variable.

Regression

If two variables are significantly correlated, then scores on one variable can be used to predict scores on the other. To conduct a linear regression in SPSS, go under "Analyze," "Regression," "Linear" and you'll see a window like the one in Figure 5.

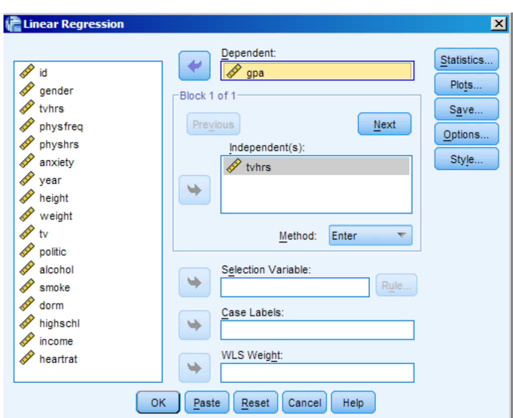


Figure 5

Put the criterion (Y) variable in the box labeled "Dependent" and the predictor (X) variable in the box labeled "Independent(s)." Click OK. You'll see the output below.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.465[a]	.216	.183	5.50744

a. Predictors: (Constant), Number of hours/week watching TV

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	90.174	1.755		51.376	.000
	Number of hours/week watching TV	-.921	.358	-.465	-2.572	.017

a. Dependent Variable: Cumulative Hamilton GPA

In the “Model Summary” table, note that R is provided, which represents the correlation between predicted Y values and actual Y values (it can range from 0 to 1). The R Square is the coefficient of determination, or strength of the relationship between X and Y. [Note that if you had multiple predictors in the model, this number would represent the total variability in Y explained by all of the predictors.] The Std. Error of the Estimate is an *estimate* of the standard error of estimate in the population, as discussed in chapter 14 of your book.

In the “Coefficients” table you’ll see numbers listed under the column “B” for your predictor variable and for (Constant). The number under B next to your predictor variable is the SLOPE of the regression line. The number under B next to (Constant) is the Y INTERCEPT of the regression line. You can compute a regression equation from this output.

Sample Results Section (to be added to correlation results):

The regression equation for predicting cumulative GPA from number of hours per week watching television was $\hat{Y} = 90.17 - .92X$, and the estimated standard error of estimate was 5.51.

SPSS Instructions for a Chi-Square

The MacArthur "Genius" award winner Elizabeth (Betsy) Levy Paluck's research on bullying has demonstrated that an effective way to reduce bullying in middle schools is to identify students with a great deal of social influence among their peers and provide them with an intervention encouraging them to take a stance against bad behavior in their school. Imagine that one school received this social influence intervention, one received a traditional anti-bullying education intervention, and one received nothing (control). The behavior of 100 students at each of the three schools was monitored for 8 months afterwards to see if any bullying occurred. [Note: because one of the variables has marginal frequencies that are fixed—100 people in each of the three conditions—the test would be a chi-square test of homogeneity.] To enter these data in SPSS you would have one column for each variable. For the condition variable, 1 = social influence intervention, 2 = traditional intervention, and 3 = control. For the bullying variable, 1 = had bullied in past 8 months and 2 = had not bullied in past 8 months. Each student would then have one number in each column to indicate his/her standing on the two variables.

Once the data are entered, go to the "Analyze" menu, scroll down to "Descriptive Statistics," and then select "Crosstabs." You will see a window like the one in Figure 1 below. Select one variable to be in rows and the other variable to be in columns.

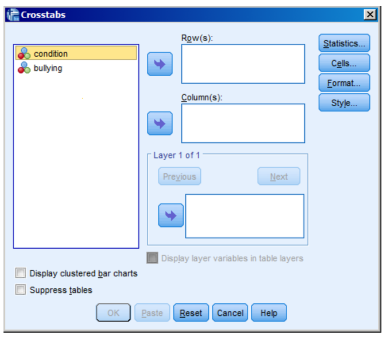


Figure 1

Click on the "Statistics" button and you will see a box like the one in Figure 2 below. Select "Chi-Square" as well as "Phi and Cramer's V" under "Nominal" and then click "Continue." Then click on the "Cells" button and you will see a window like the one in Figure 3 below. "Observed" is already selected under "Counts," but you also need to select "Adjusted standardized" under "Residuals" and then click "Continue."

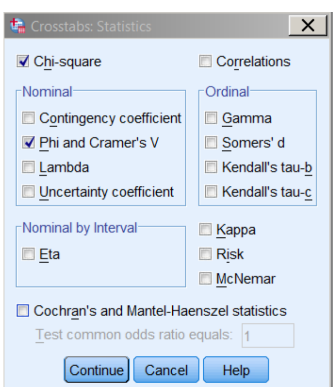


Figure 2

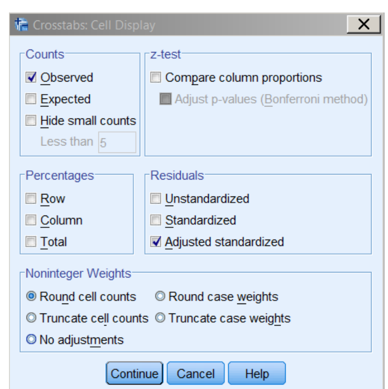


Figure 3

Click the "OK" button and you will get bumped to the output window where your output will look like the following:

bullied or not * experimental condition Crosstabulation

			experimental condition			
			social network intervention	traditional educational intervention	control	Total
bullied or not	bullied	Count	4	15	17	36
		Adjusted Residual	-3.0	1.1	1.9	
	didn't bully	Count	96	85	83	264
		Adjusted Residual	3.0	-1.1	-1.9	
Total		Count	100	100	100	300

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	9.280[a]	2	.010
Likelihood Ratio	10.847	2	.004
Linear-by-Linear Association	7.975	1	.005
N of Valid Cases	300		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 12.00.

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	.176	.010
	Cramer's V	.176	.010
N of Valid Cases		300	

The first table provides the observed counts for your cells as well as the adjusted standardized residuals, which you use to determine the nature of the relationship if the chi-square is significant. More on that in a moment.

The second table provides the chi-square statistic, df, and *p*-value in the first row and the total N in the last row. The other two rows can be ignored. In this case, it appears that the chi-square statistic is significant, as $p = .010$. Now we can go back to the first table and determine the nature of the effect by interpreting the adjusted standardized residuals.

If the value for the adjusted standardized residual for a cell is greater than or equal to the absolute value of 1.96 (i.e., the z-score that cuts off 5% of the distribution), then the observed frequency significantly differs from the expected frequency for that cell—i.e., the cell is contributing to the nature of the effect.

In this case, it appears that the number of people who bullied (vs. didn't) in the social network condition is different from what would be expected by chance. However, the residuals indicate that the counts for the traditional intervention and control conditions don't differ from chance. You could interpret this pattern as demonstrating that there was significantly less bullying than expected following the social influence intervention than following either the traditional intervention or no intervention.

The final table provides Cramer's Phi and V statistics, which you use to interpret the strength of a significant relationship (use Phi for 2 x 2 tables and V for tables larger than 2 x 2). Both vary between 0 and 1. The following norms can be used to interpret Cramer's V statistic where L is the number of levels of the variable with the fewest levels.

Effect Size	**When L – 1 = 1**	**When L – 1 = 2**	**When L – 1 = 3**
Small	0.10	0.07	0.06
Medium	0.30	0.21	0.17
Large	0.50	0.35	0.29

Sample Results section (Note: should be double-spaced):

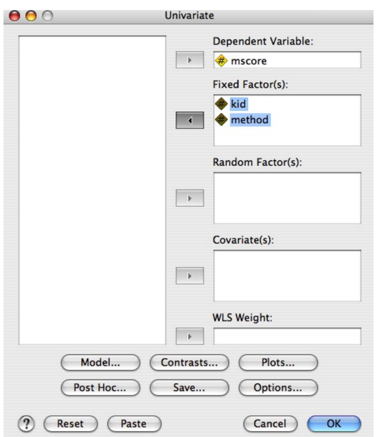
A chi-square test of the relationship between intervention condition and the presence or absence of bullying was statistically significant, χ^2 (2, $N = 300$) = 9.28, $p = .010$. As indexed by Cramer's V statistic, the strength of the relationship was small at .18. This finding reflects the fact that students were less likely to bully than expected in the social influence intervention condition, but not in the traditional intervention or control conditions.

SPSS Instructions for a 2 x 2 Factorial ANOVA

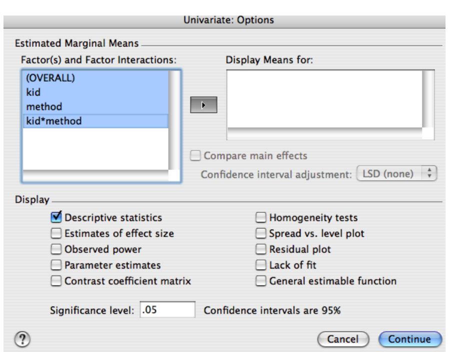
Imagine that an educational psychologist was interested in assessing the effects of two different ways of teaching math (Method A and Method B). She assumed that Method B would be better than Method A, but only for kids who were struggling in math. For kids who typically do well in math, both methods should be equally effective. Her data appear below (numbers indicate scores on the last math test):

	Method A		Method B	
Kids who struggle	75	65	85	86
	70	55	82	90
	80	82	79	86
	68	73	77	80
	64	53	70	91
Kids who do well	88	85	92	96
	84	90	84	82
	88	96	97	80
	94	91	79	95
	85	81	87	89

The first thing you need to do is to figure out how to get these data into SPSS. You have 2 IVs and one DV, so you'll need 3 columns of data: kid type (1 = struggle, 2 = does well), method (1 = A, 2 = B), and math score. Next, you need to do the overall factorial ANOVA. Go up to the Analyze menu to General Linear Model. Choose Univariate (for one dependent variable). When the window (pictured below) comes up, put the dependent variable (which I've called "mscore") into the appropriate box. Next, enter each factor (IV) into the "Fixed Factor(s)" box.



Next, click on the "Options" box to instruct SPSS to give you descriptive statistics (means and standard deviations) for each condition. You will get the window you see below:



Select each factor and the interaction, then click on the right arrow to place these into the box labeled "Display Means for." Next, click on the "Descriptive statistics" box below the "Display" label, and then click "Continue." You will get bumped back to the previous window. [Note: If any of your IVs had three or more levels, you would next need to choose "Post Hoc" from the initial ANOVA window in order to get SPSS to run a Tukey HSD test to follow up on any significant main effects for that IV. Since our two IVs have only two levels each, we can skip this step, because we can determine the nature of a significant main effect by simply examining the means.]

You should get the output pictured below:

Descriptive Statistics

Dependent Variable: MSCORE

KID	METHOD	Mean	Std. Deviation	N
kids who struggle	Method A	68.50	9.629	10
	Method B	82.60	6.363	10
	Total	75.55	10.743	20
kids who do well	Method A	88.20	4.662	10
	Method B	88.10	6.740	10
	Total	88.15	5.641	20
Total	Method A	78.35	12.504	20
	Method B	85.35	6.976	20
	Total	81.85	10.604	40

Tests of Between-Subjects Effects

Dependent Variable: MSCORE

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	2581.700[a]	3	860.567	17.179	.000
Intercept	267976.900	1	267976.900	5349.434	.000
KID	1587.600	1	1587.600	31.692	.000
METHOD	490.000	1	490.000	9.782	.003
KID * METHOD	504.100	1	504.100	10.063	.003
Error	1803.400	36	50.094		
Total	272362.000	40			
Corrected Total	4385.100	39			

a. R Squared = .589 (Adjusted R Squared = .554)

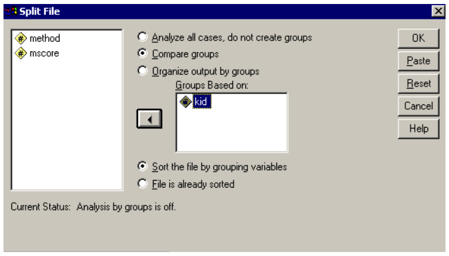
Looking first at the source table, disregard the first two rows, and focus only on the rows for kid, method, kid * method, error, and corrected total. It appears that there's a significant main effect for kid, a significant main effect for method, and a significant interaction between the two. Note that the interaction may qualify the main effects, so we definitely want to take a look at the nature of it.

Understanding the Nature of the Effects

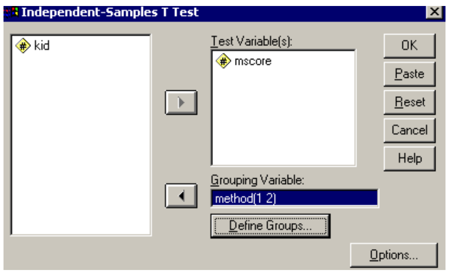
You can examine the means and standard deviations in the output box labeled "Descriptive Statistics." First, looking at the main effect for kid, it appears that the mean for kids who typically do well (88.15) is higher than the mean for kids who typically struggle (75.55). This finding isn't too surprising. Next, looking at the main effect for method, you can see that Method B (M = 85.35) results in higher math scores than Method A (M = 78.35). Note that because each of these factors has just two levels, you can understand the nature of each significant main effect simply by examining the means.

Now let's try to understand the interaction. Just looking at the means, it looks like there's a big difference in math scores by method for the kids who struggle, but not much of a difference for kids who typically do well. Of course, you know that we can't just *assume* this pattern; we need to actually check it out by examining the simple effects. We could do this in one of two ways: (1) We could look at Method within kid type (Method A vs. Method B for kids who do well; Method A vs. Method B for kids who struggle) or (2) We could look at kid type within Method (kids who do well vs. kids who struggle for Method A; kids who do well vs. kids who struggle for Method B). Either comparison is OK, but you shouldn't do both. Let's look at method within kid type.

To do so, we will need to split the file on "kid type" and conduct two *t* tests comparing Method A to Method B: one for kids who do well and one for kids who struggle. Go to the "Data" menu and choose "Split file". You will see the window below:



Click on "compare groups" and select "kid," then the right arrow to put it into the box labeled "Groups Based on:." Then click OK. SPSS will now conduct all analyses on kids who do well and kids who struggle separately. Next, you'll need to conduct an independent samples *t* test (Go to "Analyze," "Compare Means," "Independent Samples T Test") on Method:



You should now see the following output :

T-Test

Group Statistics

kid		method	N	Mean	Std. Deviation	Std. Error Mean
kids who struggle	mscore	Method A	10	68.50	9.629	3.045
		Method B	10	82.60	6.363	2.012
kids who do well	mscore	Method A	10	88.20	4.662	1.474
		Method B	10	88.10	6.740	2.132

Independent Samples Test

			Levene's Test for Equality of Variances		t-test for Equality of Means						
										95% Confidence Interval of the Difference	
kid			F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
kids who struggle	mscore	Equal variances assumed	1.456	.243	-3.863	18	.001	-14.100	3.650	-21.768	-6.432
		Equal variances not assumed			-3.863	15.601	.001	-14.100	3.650	-21.853	-6.347
kids who do well	mscore	Equal variances assumed	2.596	.125	.039	18	.970	.100	2.592	-5.345	5.545
		Equal variances not assumed			.039	16.007	.970	.100	2.592	-5.394	5.594

As you can see, the *t* value is significant for kids who struggle, but not for kids who do well. By looking at the means, we see that Method B is superior to Method A for kids who struggle, and the methods are equally helpful for kids who do well. It would appear that this interaction qualifies the main effect for Method: That is, although it appeared that, overall, Method B was more effective than Method A, we know from examining the interaction that Method B is better only for kids who struggle. Thus, we wouldn't want to make any general statement about Method B being better for everyone. (This is what it means for an interaction to *qualify* a main effect.) It's probably still safe to interpret the main effect that kids who typically do well do better than kids who typically struggle, regardless of teaching method.

Sample APA-Style Results Section:

A 2 (teaching method: Method A vs. Method B) x 2 (student skill level: struggling vs. doing well) between-subjects factorial ANOVA was conducted on math exam scores. There was a significant main effect of teaching method, F (1,36) = 9.78, p = .003, such that students who were instructed with Method B (M = 85.35, SD = 6.98) performed better on the math test than did students who were instructed with Method A (M = 78.35, SD = 12.50). The strength of the effect, as indexed by eta^2, was .11, indicating a medium effect. As predicted, the main effect for type of student was significant, F (1,36) = 31.69, p < .001, such that students who struggle with math (M = 75.55, SD = 10.74) performed worse on the math exam than did students who do well with math (M = 88.15, SD = 5.64). The strength of the effect, as indexed by eta^2, was .36, a strong effect.

The main effect of type of instruction was qualified by a significant interaction between teaching method and type of student, F (1,36) = 10.06, p = .003. The strength of the interaction, as indexed by eta^2, was .11, a medium effect. To understand the nature of the interaction, two independent-groups *t* tests, comparing Method A to Method B, were conducted: one for students who do well in math and one for students who struggle. As seen in Figure 1, for students who do well, there was no significant difference in math exam scores for those instructed with Method A (M = 88.20, SD = 4.66) and those instructed with Method B (M = 88.10, SD = 6.74), t (18) = .04, p = .97; 95% CI [-5.35, 5.55]. However, students who struggle in math performed significantly better on the math exam when instructed with Method B (M = 82.60, SD = 6.36) than with Method A (M = 68.50, SD = 9.63), t (18) = -3.86, p = .001; 95% CI [-21.77, -6.43].

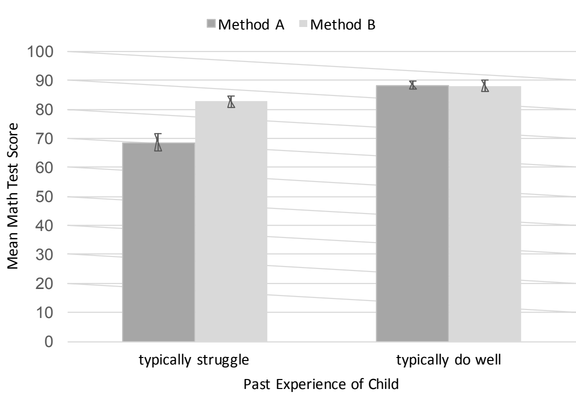


Figure 1. Mean math test score as a function of teaching method and child's typical experience with math. Bars represent one standard error above and below the mean.

Understanding SPSS Output for a 2 x 3 Factorial ANOVA

Class example: 15 control participants and 15 participants with a temporal lobe lesion were given a list of words to memorize. Participants in each group were randomly assigned to receive an auditory cue, a visual cue, or no cue upon recall. Number of words recalled was recorded for each participant.

Univariate Analysis of Variance

Between-Subjects Factors

		Value Label	N
lesion condition	1	control	15
	2	temporal lobe lesion	15
recall cue condition	1	free recall	10
	2	auditory cue	10
	3	visual cue	10

Descriptive Statistics

Dependent Variable: recall score (# of items recalled)

lesion condition	recall cue condition	Mean	Std. Deviation	N
control	free recall	15.20	2.683	5
	auditory cue	24.80	7.396	5
	visual cue	21.80	6.140	5
	Total	20.60	6.759	15
temporal lobe lesion	free recall	5.00	2.739	5
	auditory cue	4.60	3.050	5
	visual cue	25.60	6.025	5
	Total	11.73	10.872	15
Total	free recall	10.10	5.953	10
	auditory cue	14.70	11.908	10
	visual cue	23.70	6.075	10
	Total	16.17	9.973	30

Tests of Between-Subjects Effects

Dependent Variable: recall score (# of items recalled)

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	2273.367[a]	5	454.673	17.865	.000
Intercept	7840.833	1	7840.833	308.088	.000
lesion	589.633	1	589.633	23.168	.000
cue	957.067	2	478.533	18.803	.000
lesion * cue	726.667	2	363.333	14.276	.000
Error	610.800	24	25.450		
Total	10725.000	30			
Corrected Total	2884.167	29			

a. R Squared = .788 (Adjusted R Squared = .744)

So, there's a significant main effect for lesion, a significant main effect for cue, and a significant lesion x cue interaction. There is no need to look at the Estimated Marginal Means tables yet.

Post Hoc Tests (Tukey test for recall cue condition main effect)

recall cue condition

Multiple Comparisons

Dependent Variable: recall score (# of items recalled)
Tukey HSD

(I) recall cue condition	(J) recall cue condition	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
free recall	auditory cue	-4.60	2.256	.125	-10.23	1.03
	visual cue	-13.60*	2.256	.000	-19.23	-7.97
auditory cue	free recall	4.60	2.256	.125	-1.03	10.23
	visual cue	-9.00*	2.256	.002	-14.63	-3.37
visual cue	free recall	13.60*	2.256	.000	7.97	19.23
	auditory cue	9.00*	2.256	.002	3.37	14.63

Based on observed means.
The error term is Mean Square(Error) = 25.450.
*. The mean difference is significant at the .05 level.

Homogeneous Subsets

recall score (# of items recalled)

Tukey HSD[a,b]

recall cue condition	N	Subset	
		1	2
free recall	10	10.10	
auditory cue	10	14.70	
visual cue	10		23.70
Sig.		.125	1.000

Means for groups in homogeneous subsets are displayed.
Based on observed means.
The error term is Mean Square(Error) = 25.450.
a. Uses Harmonic Mean Sample Size = 10.000.
b. Alpha = .05.

The nature of the main effect for cue is such that participants provided with a visual cue performed significantly better than those provided with an auditory cue or no cue, who did not differ from each other. Now we need to interpret the nature of the significant interaction. We split the file on the lesion variable and conduct two one-way ANOVAs on the recall cue variable.

Oneway

Descriptives

recall score (# of items recalled)

lesion condition		N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
						Lower Bound	Upper Bound		
control	free recall	5	15.20	2.683	1.200	11.87	18.53	11	18
	auditory cue	5	24.80	7.396	3.308	15.62	33.98	17	34
	visual cue	5	21.80	6.140	2.746	14.18	29.42	17	32
	Total	15	20.60	6.759	1.745	16.86	24.34	11	34
temporal lobe lesion	free recall	5	5.00	2.739	1.225	1.60	8.40	2	9
	auditory cue	5	4.60	3.050	1.364	.81	8.39	1	9
	visual cue	5	25.60	6.025	2.694	18.12	33.08	21	34
	Total	15	11.73	10.872	2.807	5.71	17.75	1	34

ANOVA

recall score (# of items recalled)

lesion condition		Sum of Squares	df	Mean Square	F	Sig.
control	Between Groups	241.200	2	120.600	3.633	.058
	Within Groups	398.400	12	33.200		
	Total	639.600	14			
temporal lobe lesion	Between Groups	1442.533	2	721.267	40.750	.000
	Within Groups	212.400	12	17.700		
	Total	1654.933	14			

For the control participants, there was no significant effect of type of cue on recall. For the participants with lesions, there was a significant effect of type of cue on recall. We need to examine the Tukey test to interpret the nature of the effect of cue type on recall for the participants with lesions.

Post Hoc Tests

Multiple Comparisons

Dependent Variable: recall score (# of items recalled)
Tukey HSD

lesion condition	(I) recall cue condition	(J) recall cue condition	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	Upper Bound
control	free recall	auditory cue	-9.600	3.644	.053	-19.32	.12
		visual cue	-6.600	3.644	.208	-16.32	3.12
	auditory cue	free recall	9.600	3.644	.053	-.12	19.32
		visual cue	3.000	3.644	.696	-6.72	12.72
	visual cue	free recall	6.600	3.644	.208	-3.12	16.32
		auditory cue	-3.000	3.644	.696	-12.72	6.72
temporal lobe lesion	free recall	auditory cue	.400	2.661	.988	-6.70	7.50
		visual cue	-20.600*	2.661	.000	-27.70	-13.50
	auditory cue	free recall	-.400	2.661	.988	-7.50	6.70
		visual cue	-21.000*	2.661	.000	-28.10	-13.90
	visual cue	free recall	20.600*	2.661	.000	13.50	27.70
		auditory cue	21.000*	2.661	.000	13.90	28.10

*. The mean difference is significant at the 0.05 level.

The ANOVA was not significant for the control participants, so this post-hoc test does not need to be interpreted (notice that SPSS prints it out anyway; you have to know better not to look at it!)

Homogeneous Subsets

recall score (# of items recalled)

lesion condition=control

Tukey HSD[a]

recall cue condition	N	Subset for alpha = 0.05
		1
free recall	5	15.20
visual cue	5	21.80
auditory cue	5	24.80
Sig.		.053

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 5.000.

lesion condition=temporal lobe lesion

Tukey HSD[a]

recall cue condition	N	Subset for alpha = 0.05	
		1	2
auditory cue	5	4.60	
free recall	5	5.00	
visual cue	5		25.60
Sig.		.988	1.000

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 5.000.

For the participants with lesions, providing visual cues resulted in significantly better recall than providing either auditory cues or no cues, which did not differ from each other. Thus, the main effect of type of cue was qualified by the significant interaction. That pattern holds only for the participants with a temporal lobe lesion.

Sample Results Section:

A 2 (condition: control vs. participants with a temporal lobe lesion) x 3 (type of recall cue: auditory, visual, none) between-subjects factorial ANOVA was conducted on number of words recalled from a list. As predicted, there was a significant main effect of condition, F (1,24) = 23.17, p < .001, such that participants with a temporal lobe lesion recalled significantly fewer words (M = 11.73, SD = 10.87) than did participants in the control condition (M = 20.60, SD = 6.76). The strength of the effect, as indexed by eta^2, was .20, indicating a strong effect. The main effect for type of cue was also significant, F (2,24) = 18.80, p < .001. A Tukey HSD test indicated that participants provided with a visual cue (*M* = 23.70, *SD* = 6.08) performed significantly better than those provided with an auditory cue (*M* = 14.70, *SD* = 11.91; 95% CI [3.37, 14.63]) or no cue (*M* = 10.10, *SD* = 5.95; 95% CI [7.97, 19.23]), who did not differ from each other; 95% CI [-1.03, 10.23]. The strength of the effect, as indexed by eta^2, was .33 a strong effect.

The main effect of type of cue was qualified by a significant interaction between condition and cue type, F (2,24) = 14.28, $p < .001$. The strength of the interaction, as indexed by eta^2, was .25, a large effect. To understand the nature of the interaction, two one-way between subjects ANOVAs comparing cue type were conducted: one for participants with a temporal lobe lesion and one for participants in the control group. As seen in Figure 1, for control participants, a oneway ANOVA indicated no significant difference in number of words recalled for those provided with no cue (*M* = 15.20, *SD* = 2.68) as compared to an auditory cue (*M* = 24.80, *SD* = 7.40; 95% CI [-19.32, .12]) or a visual cue (*M* = 21.80, *SD* = 6.14; 95% CI [-16.32, 3.12]), nor for those provided with an auditory cue as compared to a visual cue, 95% CI [-6.72, 12.72], F (2,12) = 3.63, p = .058. However, the oneway ANOVA for participants with a temporal lobe lesion was significant, F (2,12) = 40.75, $p < .001$. A Tukey HSD test indicated that participants with a temporal lobe lesion performed significantly better on the recall task when provided with a visual cue (*M* = 25.60, *SD* = 6.03) than those provided with either an auditory cue (*M* = 4.60, *SD* = 3.05; 95% CI [13.90, 28.10]) or no cue (*M* = 5.00, *SD* = 2.74; 95% CI [13.50, 27.70]), who did not differ from each other, 95% CI [-7.50, 6.70].

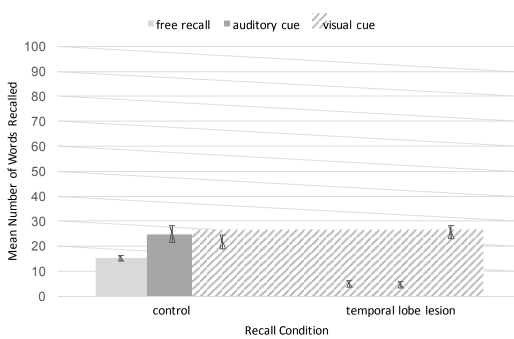


Figure 1. Mean number of words recalled as a function of recall condition and participant status. Bars represent one standard error above and below the mean.