Quantitative Analysis Using SPSS 27

By Lindsay Demers, PhD

Director, Education Evaluation Core

Boston University School of Medicine

Caveat:

The overarching goal of this document is to provide clinician and scientist-educators with a basic road map for medical education research. The level of detail included here is not sufficient to gain expertise in these areas; rather, think of this as a document you can use for guidance when designing your research and making analytic choices. If you find a topic that is of substantial relevance to your work, please email me for additional resources on the topic, as only the bare minimum of detail is included here.

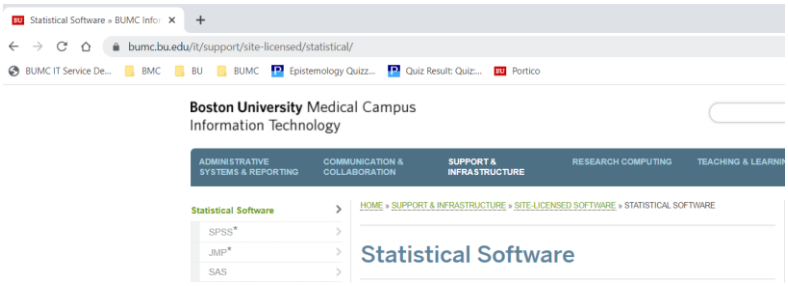
Further, I have only included here what seems to be most commonly used within the department of medicine. It is likely you will encounter instances that are not covered in this guidebook. If/when this occurs email the Education Evaluation Core for a consultation (if you're within the DoM or part of CELP) or email ldemers@bu.edu for additional resources.

This guide does not contain information on how to interpret the statistical output that you will produce. This omission was intentional, as I strongly encourage you to meet with someone who can interpret the output with you.

Using SPSS for Quantitative Analysis

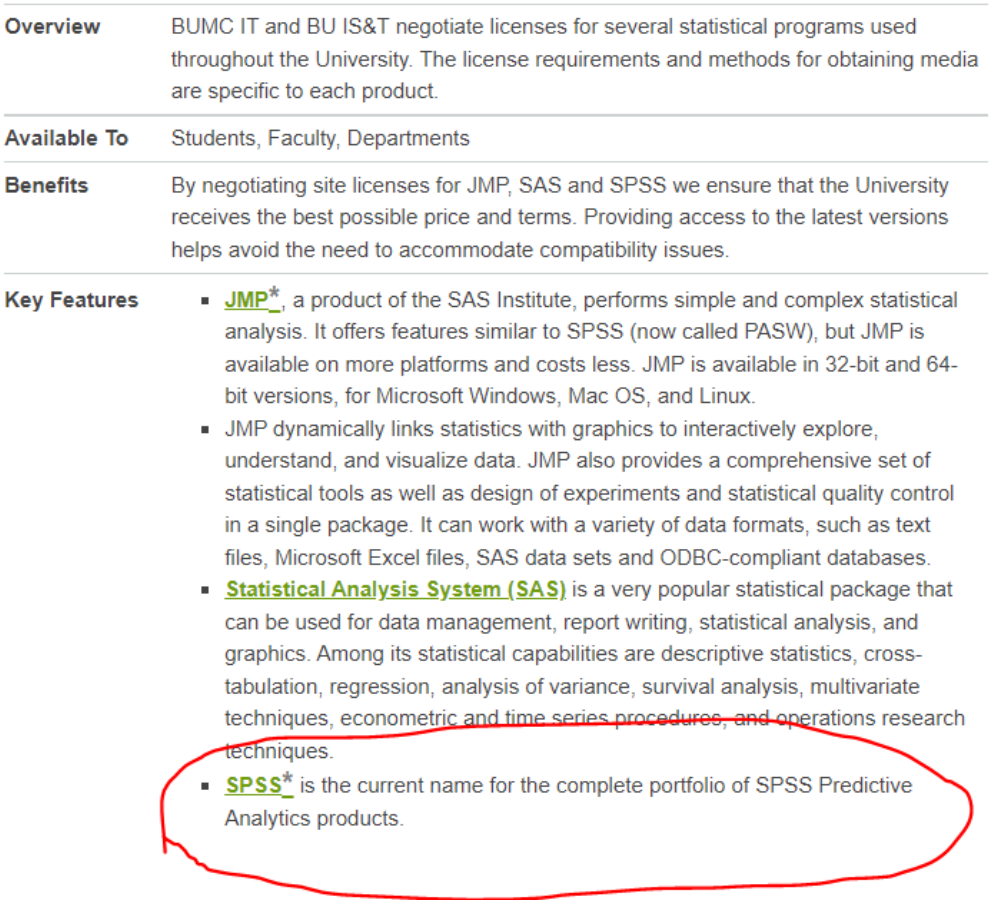
Obtaining SPSS from BUMC IT

In order to install SPSS onto your computer, you must have a BU/Kerberos login. If you want to install it onto a BMC computer, you will need to go through BMC IT to get the appropriate permissions. If it's any other machine, you can install it yourself right from the BUMC IT website:



You'll see below that there are other stats software options (JMP and SAS) available. I've never used JMP so do not have an opinion about it. SAS is a great program but has a steeper learning curve than SPSS and costs an additional $50 per year.

Click on SPSS (circled below) and a new window will open where you will enter your BU login information.



Overview	BUMC IT and BU IS&T negotiate licenses for several statistical programs used throughout the University. The license requirements and methods for obtaining media are specific to each product.
Available To	Students, Faculty, Departments
Benefits	By negotiating site licenses for JMP, SAS and SPSS we ensure that the University receives the best possible price and terms. Providing access to the latest versions helps avoid the need to accommodate compatibility issues.
Key Features	▪ JMP*, a product of the SAS Institute, performs simple and complex statistical analysis. It offers features similar to SPSS (now called PASW), but JMP is available on more platforms and costs less. JMP is available in 32-bit and 64-bit versions, for Microsoft Windows, Mac OS, and Linux. ▪ JMP dynamically links statistics with graphics to interactively explore, understand, and visualize data. JMP also provides a comprehensive set of statistical tools as well as design of experiments and statistical quality control in a single package. It can work with a variety of data formats, such as text files, Microsoft Excel files, SAS data sets and ODBC-compliant databases. ▪ Statistical Analysis System (SAS) is a very popular statistical package that can be used for data management, report writing, statistical analysis, and graphics. Among its statistical capabilities are descriptive statistics, cross-tabulation, regression, analysis of variance, survival analysis, multivariate techniques, econometric and time series procedures, and operations research techniques. ▪ SPSS* is the current name for the complete portfolio of SPSS Predictive Analytics products.

After you've logged in, you'll see a window asking for your role. Presumably if you're reading this you'll likely be clicking the first link – *Academic Students, Faculty and Staff Only*. Then you will be prompted to download and install the program.

Note also that through this page you can "extend" your SPSS license. Each year SPSS will stop working because you need to renew your license (you get numerous warnings before this happens). This annual renewal is what is meant by "extending" your SPSS. So, after your initial install, you will need to revisit this page each year to re-up the license on the copy you've downloaded.

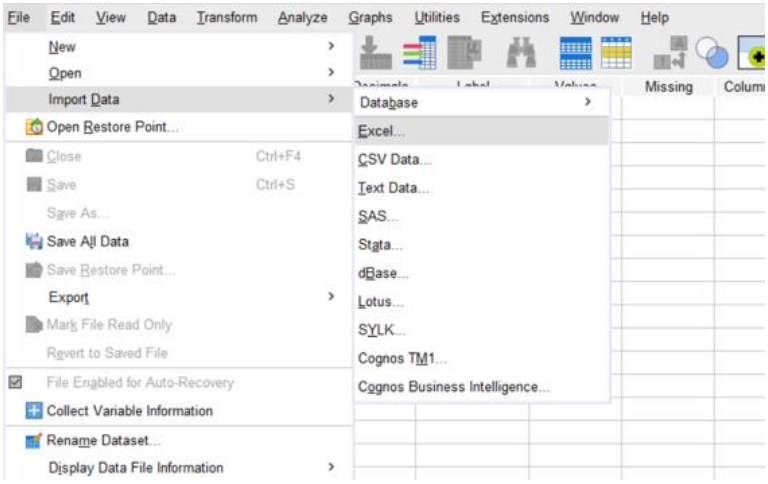
If you have any issues with downloading or installation, call BUMC IT: (617) 358-4540

Reading in Data from Excel

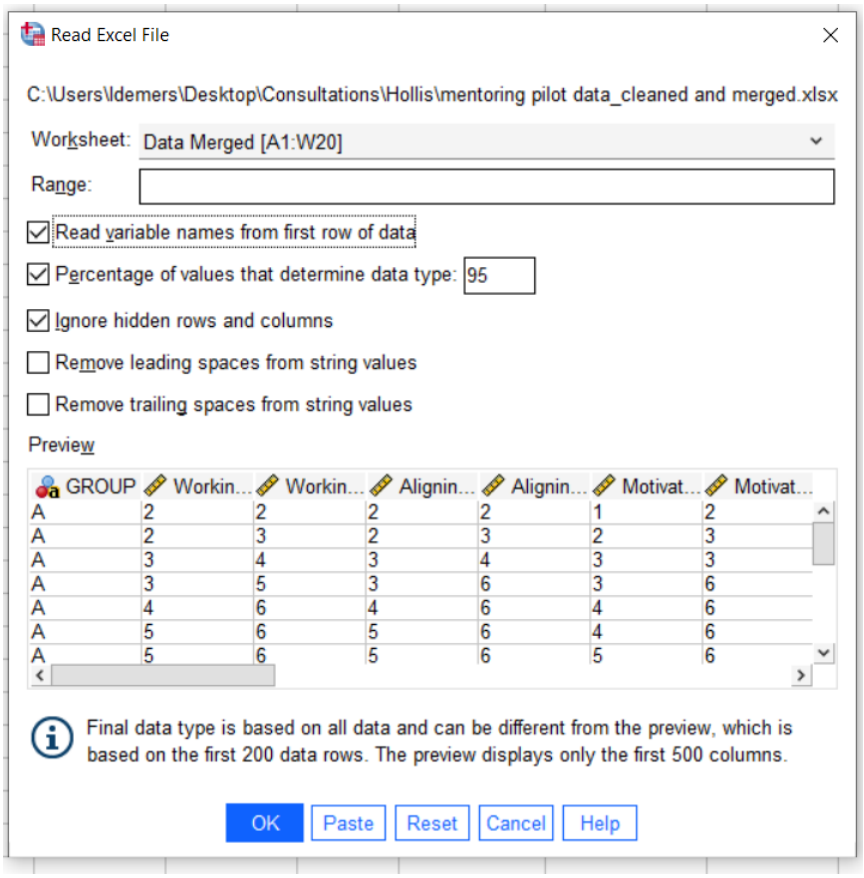
If you are pulling data out of Qualtrics, you have several options for the file type you'd like it to create. The two you'll most likely be using are either an Excel file (.xls) or an SPSS file (.sav). I tend to export my data directly into an SPSS file but I wanted to make sure you're all aware of the Excel option as well. If you are collecting data from some other source besides Qualtrics, you can choose to enter it directly into Excel or into SPSS by creating a new file in either program.

If you export your data as an SPSS file, you can simply double click the data file and it will open it up in SPSS.

To import data from an Excel spreadsheet into SPSS, open the program and select the following options in the SPSS dropdown menu:



Once you've navigated to the Excel file you want to import, the following window will pop up with a few options as well as a preview of what your data will look like once they've been imported into SPSS.



The first dropdown menu "Worksheet" is where you can tell SPSS which worksheet within the Excel file you want to be imported. If you only have one worksheet in your Excel file it will be selected by default.

Assuming your data are structured so that rows are individual cases and columns are your variables, you will want to make sure "Read variable names from first row" is checked. The second option determines how SPSS recognizes the data – e.g., if 95% of the values are words, it will assume that all the data in that column are of the same type. This is very easy to change once you've imported your data, but even so you shouldn't have a reason in most cases to uncheck this box (the default is for it to be checked). The remaining options are up to you to decide based on how your data are set up in Excel. More often than not you will not need to change the default settings for the three remaining options.

Once you're happy with how your data look in the preview window, simply hit "OK" and the file will be imported to SPSS.

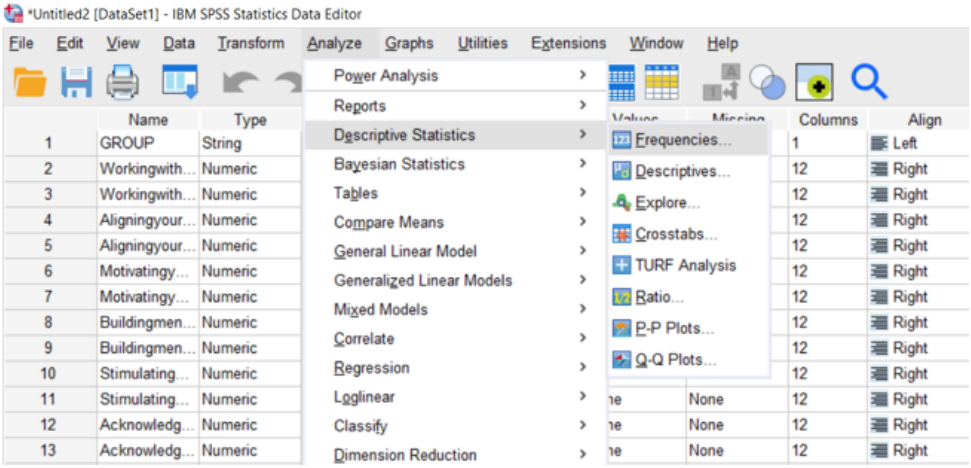
At this time you will want to save this SPSS file (.sav) so that you will not have to repeat the step to import the data from Excel.

Examining Data

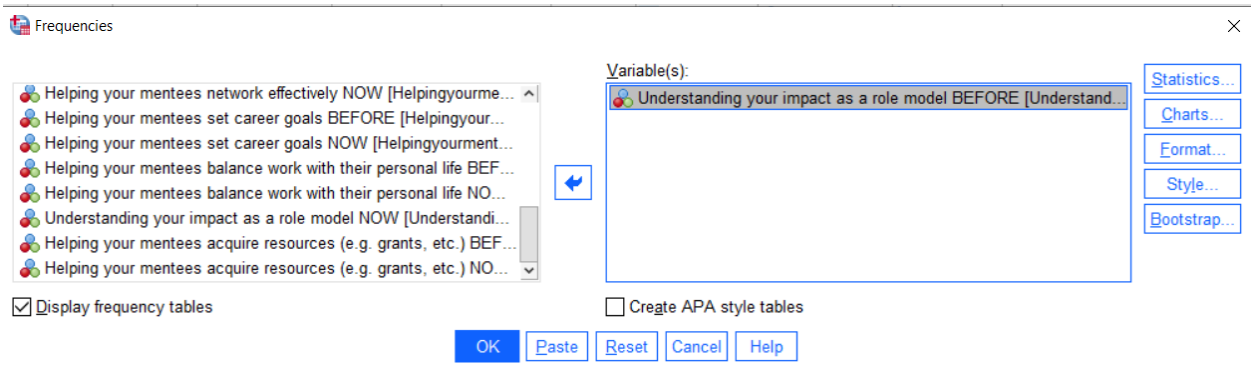
Although you will likely already know if your data are continuous based on what you measured and how you measured it, it is important to look at your data before undertaking any statistical analysis. Namely, you will want to look at the distribution of the data as well as the descriptive statistics. Further, for reporting purposes, you will likely need to generate some descriptive statistics around the demographics/make-up of your sample. The follow steps will allow you to produce this information for both purposes.

Generating a Histogram

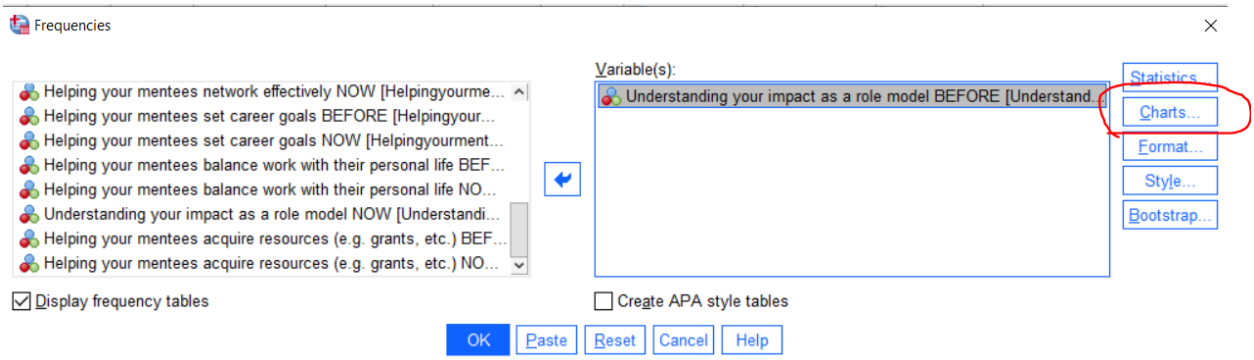
It is important to look at a histogram of your data (especially your outcome variables of interest) to see how they are distributed. If the data look fairly normally distributed (like a bell curve) you are likely ok to move ahead with using parametric statistical tests. If they are not, you will need to use nonparametric statistical tests. A list of what tests to use when is provided in the Quantitative Methods guide previously distributed.



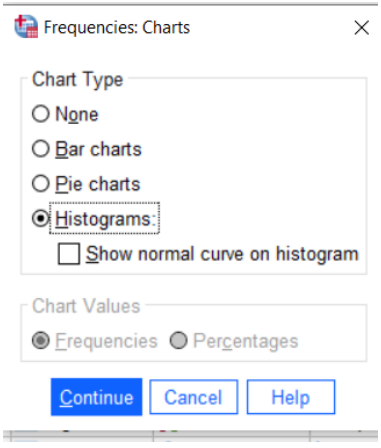
Once the Frequencies menu opens, you select the variables for which you'd like to produce a histogram and move them over to the empty box on the right ("Variable(s)").



Next, click “Charts” on the righthand side of the window.



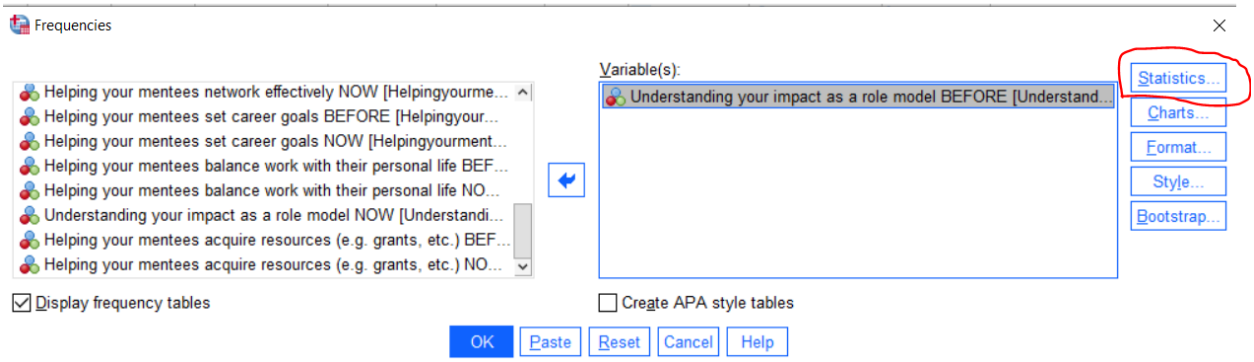
A new window will pop up (see image below). From this window select “Histograms.” You can choose to have SPSS superimpose a normal distribution over your data if you would like by checking the box below the Histograms option. Hit “Continue” and then “Ok”

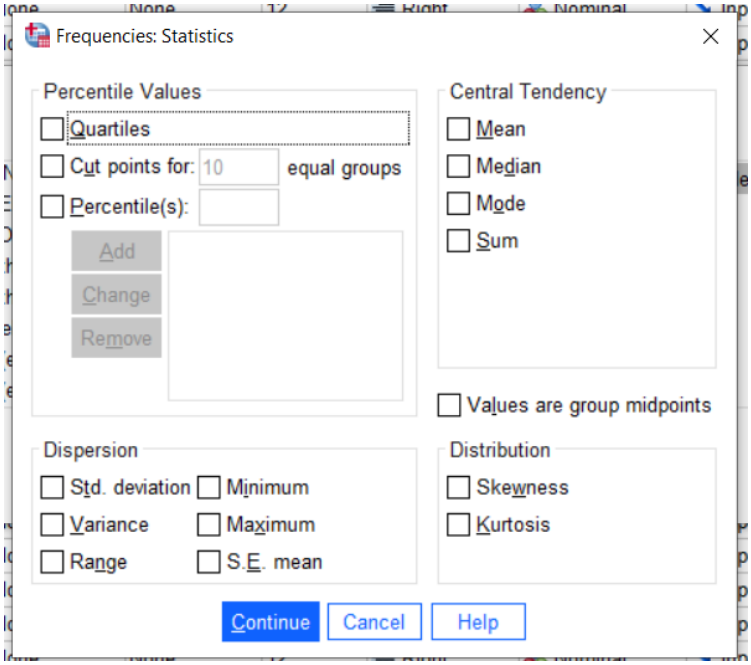


An output window will open and will include a table with the frequencies for each value of the variables you selected (the number of times each value appears in the data set). Below that will be the histogram(s) that SPSS generated.

Generating Descriptive Statistics

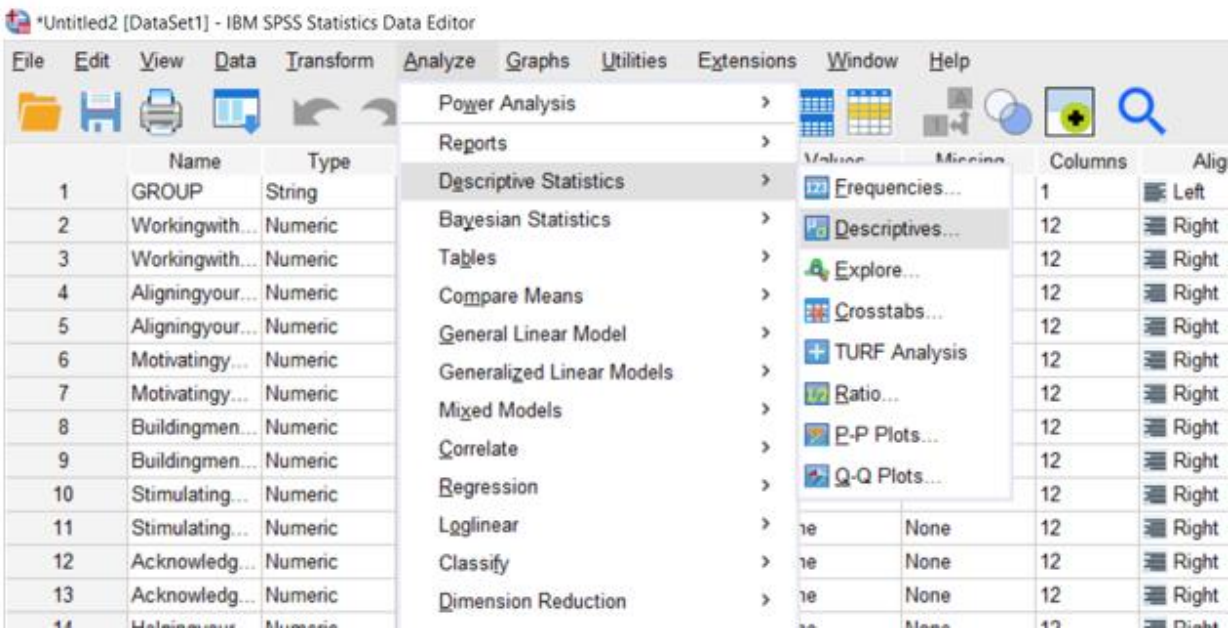
There are two ways you can generate the mean, standard deviation, median, etc. of your data. One way is to do it at the same time as when you're generating your histogram. By clicking on the "statistics" button within the "Frequencies" menu, you are given the option of asking SPSS to produce various descriptive statistics.



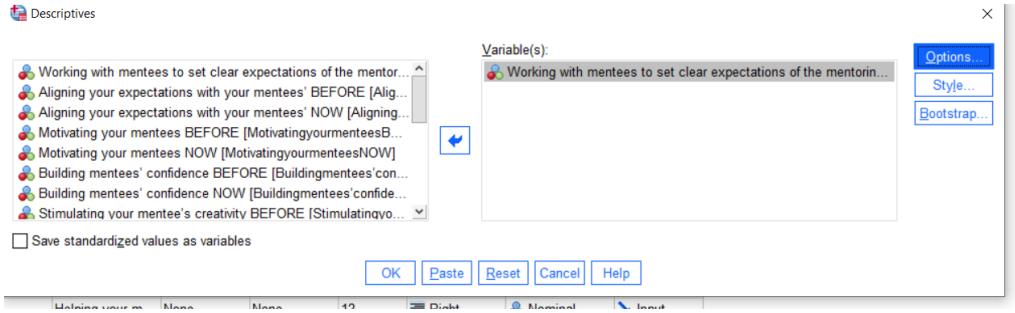


If you choose to generate descriptive statistics in this way, they will be provided in a table at the beginning of the output, before the frequency tables and histograms.

Alternatively, you can go to the "Descriptives" menu as shown below.



Similar to what you did to produce the histogram, you will want to select the variables for which you want to generate descriptive statistics and move them over from the left to the empty window on the right. Then, hit "Ok."



An output window will open that includes the information for the variables you have requested.

Analyzing Data

Once you have examined your data, you will have an idea of what statistical test to run based on the decision tree in the guide on Quantitative Methods. As noted in the beginning of this guide, I do not include information for interpreting the statistical output because it is important that you look at the output with someone who understands statistics and can interpret the output with you.

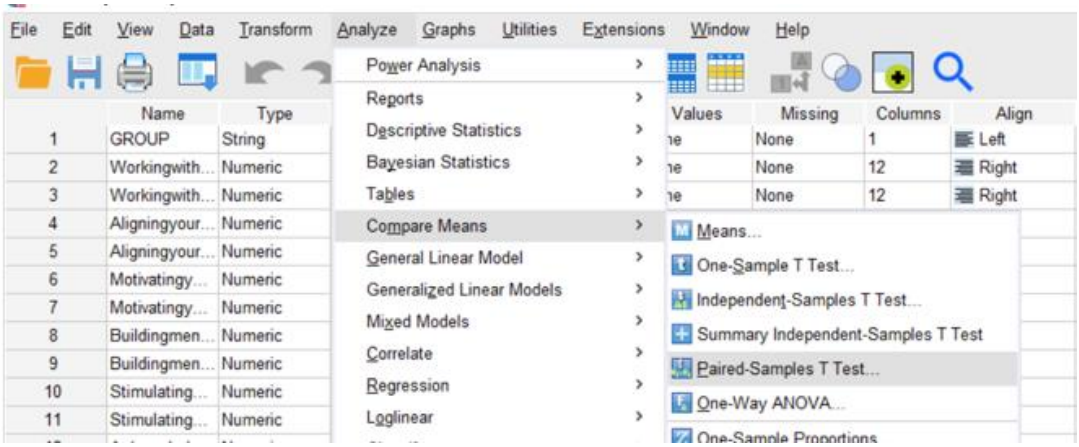
Parametric Tests: t test and ANOVA

There are two types of t test and two types of ANOVA covered in this guide. To start, we will focus on the ones used for looking at change within-subjects (i.e., paired sample t tests and repeated measures ANOVA). Second, we will focus on those tests that look at change between-subjects (e.g., a treatment and comparison group).

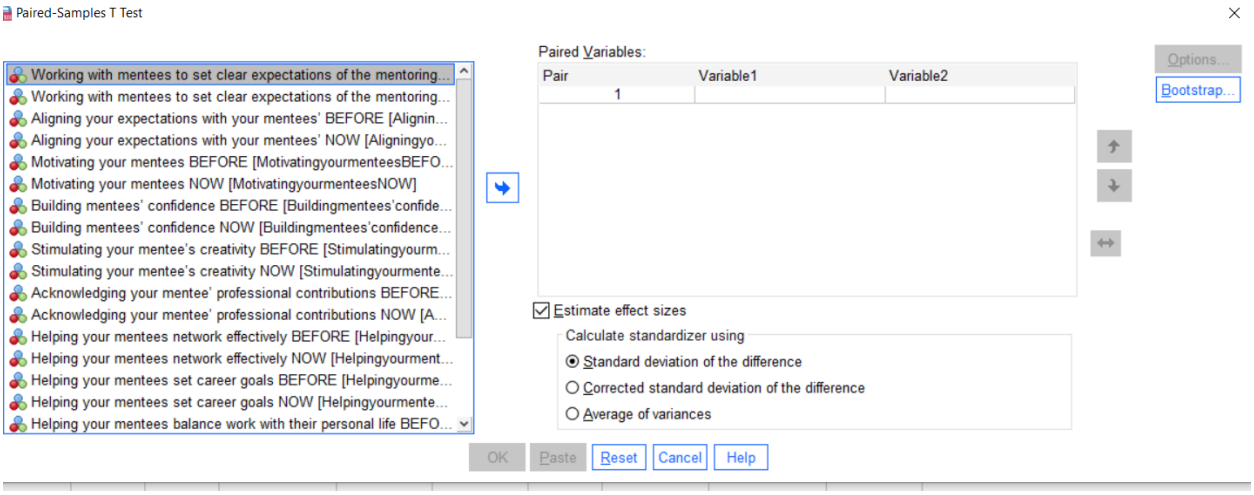
Within-Subjects (Pre-/post- or Pre-/Post-/Delayed post- designs)

Paired sample t test

You should use a paired sample t test when you are looking at two time points nested within an individual (usually this is pre-/post-). To run this test, follow the dropdown menu commands in the image below.



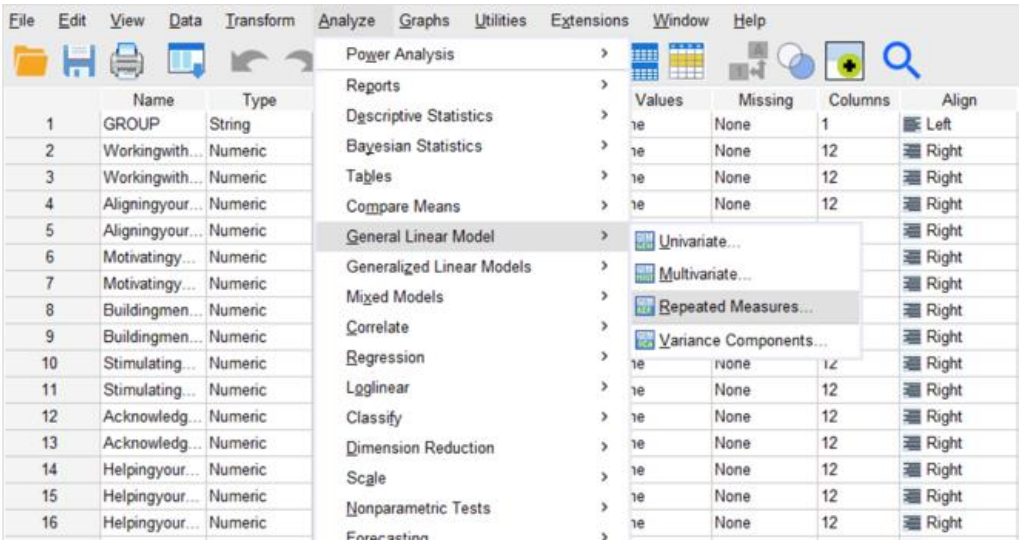
Once you've selected "Paired-samples t test" the window below will appear.



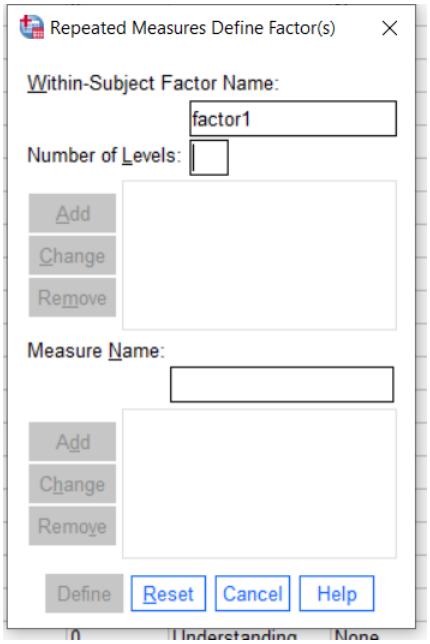
Move the two variables you want to compare from the window on the left to the window on the right. Then, hit "OK" and an output window will open up with the results.

Repeated Measures ANOVA

When you have more than two time points within an individual (e.g., pre-, post- and delayed post-) you will use a repeated measures ANOVA. Whereas the paired samples t test only allows for two time points, the repeated measures ANOVA allows for three or more. In order to run the repeated measures ANOVA, follow the dropdown menu commands in the image below.

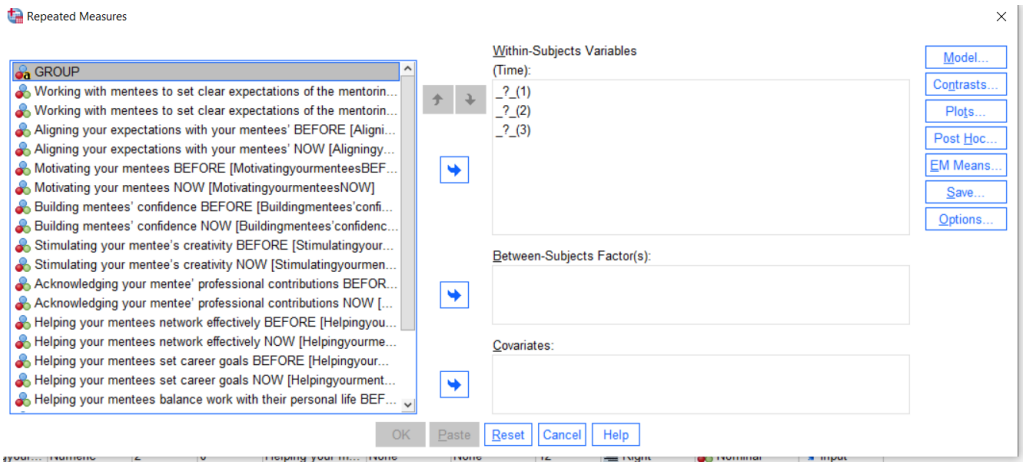


Once you've selected "Repeated Measures ANOVA" the window below will pop up.

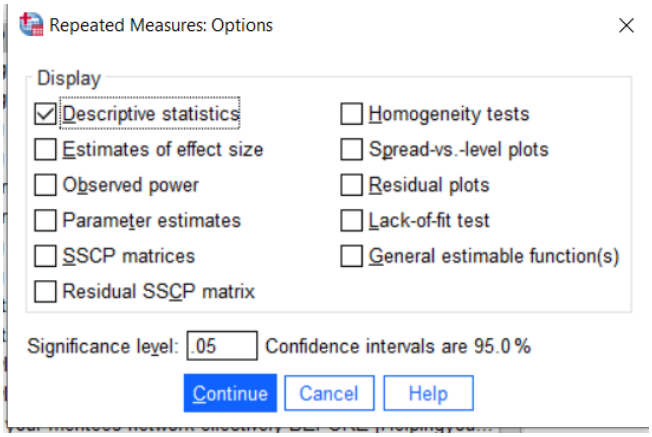


The within-subjects factor name is called factor1 by default, but you can change it to whatever you want. For example, if you have a pre-, post- and delayed post-, you might rename factor1 to "time." Then, for "Number of Levels" you put in the number of time points you have. So if you have a pre-, post- and delayed post-, you would have 3 levels and input the number 3. Once you've renamed the within-subjects factor if you want to, and added in the number of levels you have, click "Add." You won't need to add or change anything else in this window. To move to the next step, hit "Define."

After you've hit "Define" the window below will appear.



In this example I've specified three levels of our within-subjects factor (pre-, post- and delayed post-), so on the first window on the right ("Within-Subjects Variables") there are spaces for each of the three time points. From the variable list on the left, add in your pre- variable to the (1) space, the post-variable to the (2) space and the delayed post- variable to the (3) space. If you have more than three time points you'd continue in this same way. Once you've moved over the appropriate variables, hit "Options" on the right.



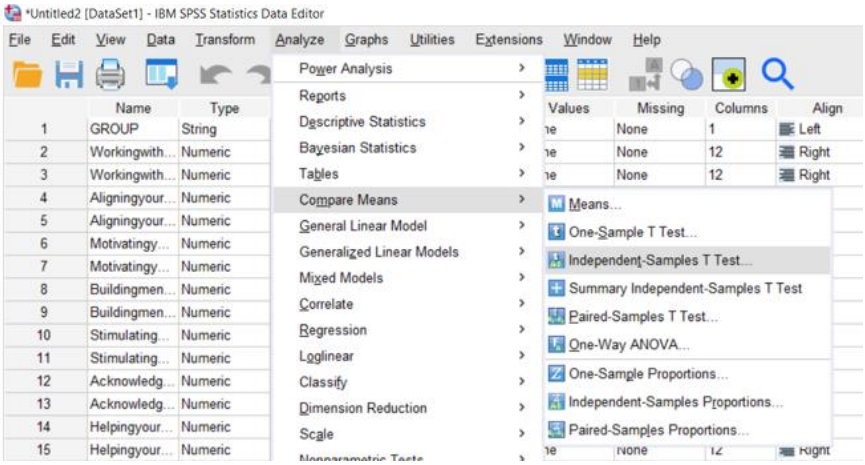
Check off "descriptive statistics" within the Options menu, then hit "Continue" to return to the previous menu.

Finally, hit "OK" and the results will be generated in the output window.

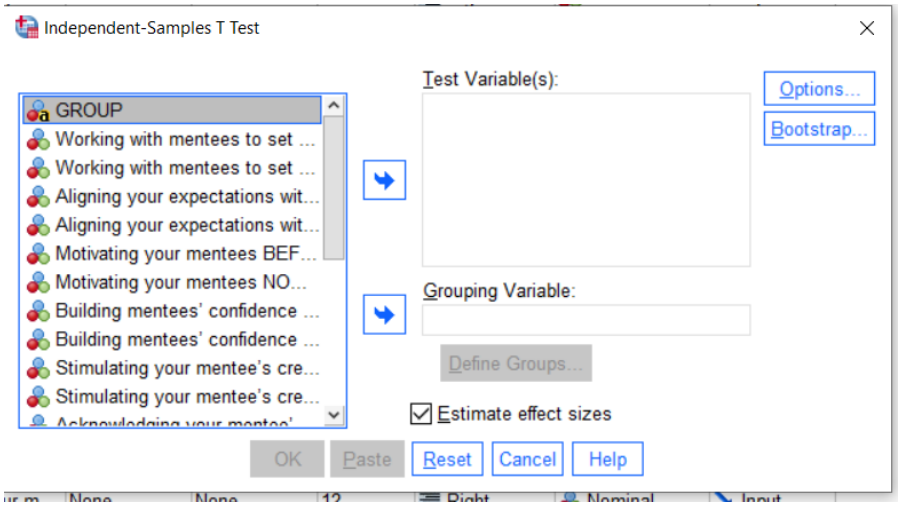
Between-Subjects (quasi-experimental or experimental designs)

Independent Samples t test

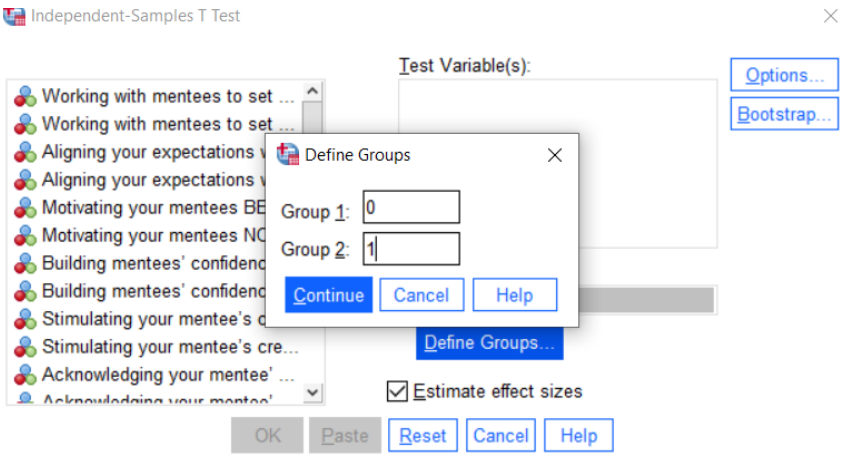
When you have two groups (e.g., a treatment and a comparison group) you can analyze your data using an independent samples t test. To run an independent samples t test, follow the dropdown menu commands shown below.



Once you have selected the independent samples t test, the following window will appear.



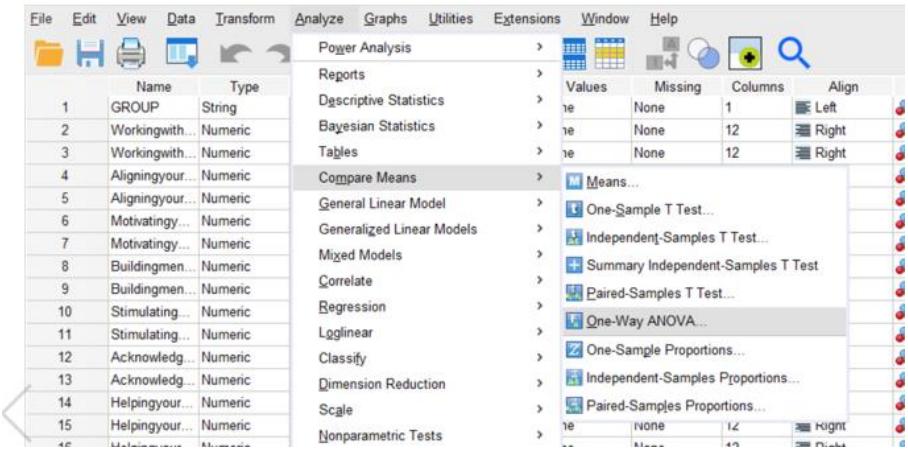
You'll start by putting your between-subjects variable ("Grouping Variable") into the second window on the right. Then, you'll click "Define Groups" and you will input the two numeric values of your group variable. For example, you may have the Group variable be 0 for comparison group and 1 for treatment group.



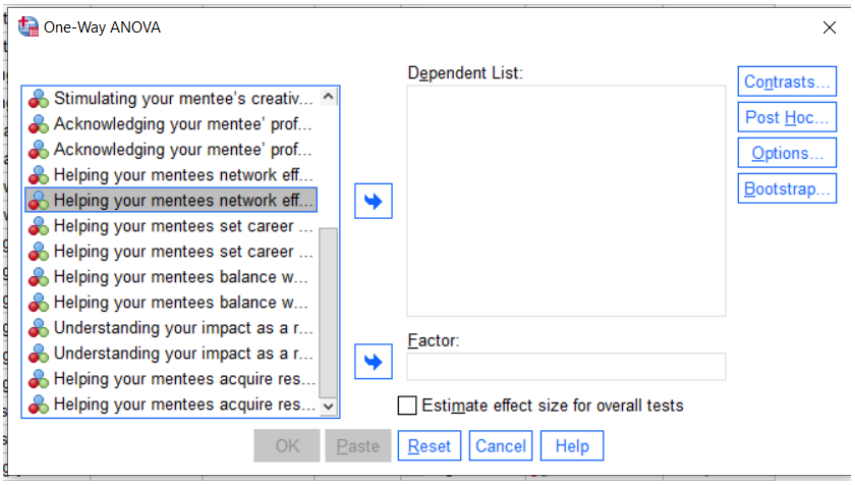
Next, hit “Continue” to close the “Define Groups” pop up window. Then you should move your outcomes of interest (the variables for which you’re looking for between-group differences) into the top window on the right (“Test Variables”). Then hit “OK” and your statistical output will be generated in the output window.

One-way ANOVA

If you have more than two groups, you will use a one-way ANOVA. To run a one-way ANOVA follow the dropdown menu commands shown in the image below.



Once you’ve selected One-Way ANOVA, the following window will pop up.



Move your between-subjects variable (e.g., group membership) to the “Factor” window on the bottom right. Next, move your outcome variables (s) of interest to the “Dependent List” window.

Hit “OK” and the output for your One-Way ANOVA will be generated in the output window.

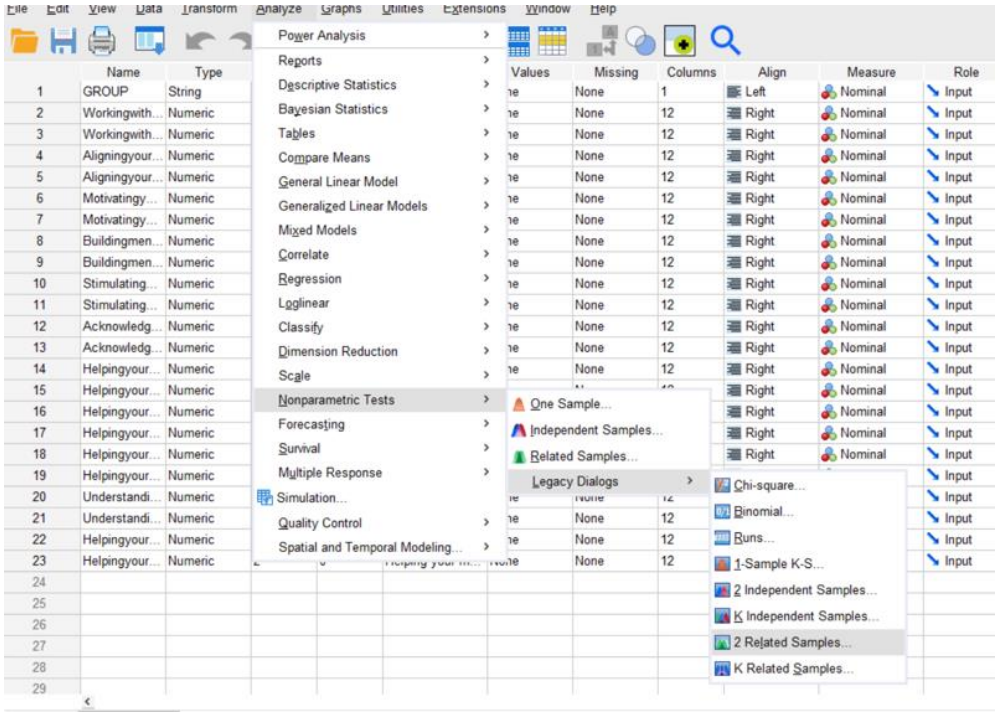
Nonparametric Tests

In this part of the guide we cover the nonparametric equivalents to the parametric tests described above.

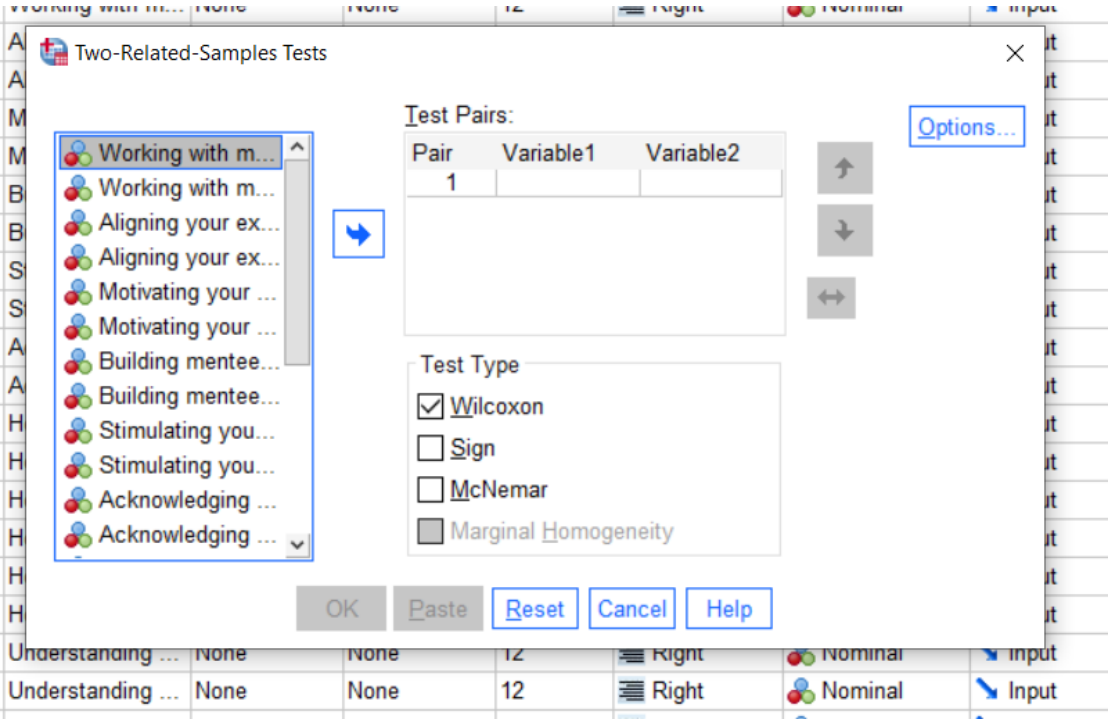
Within-Subjects (Pre-/post- or Pre-/Post-/Delayed post- designs)

Sign Test or Wilcoxon

When you have two datapoints within individuals that you want to compare (e.g., pre- and post-), you can use a Sign test or a Wilcoxon test; both will answer the question of whether the datapoints significantly differ albeit in somewhat different ways. To run either of these tests follow the commands from the dropdown menu as shown below.



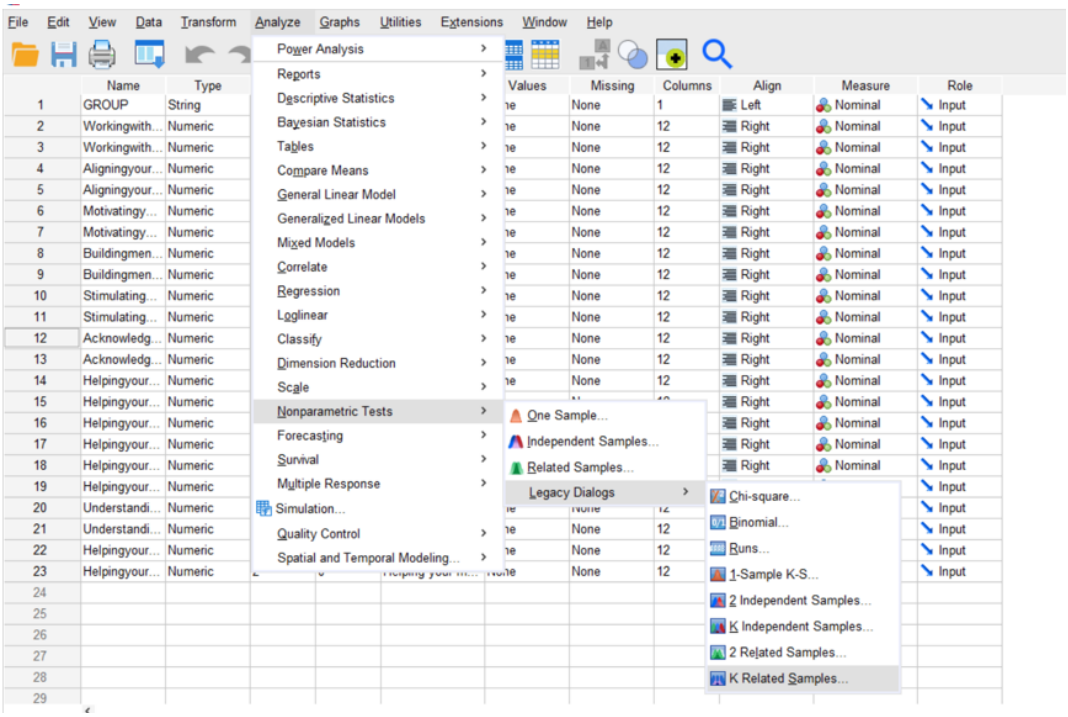
Once you've selected "2 Related Samples" the following window will pop up. Note that by default, SPSS selects Wilcoxon. If you wish to run a Sign test instead, you can simply check that box and uncheck the Wilcoxon box.



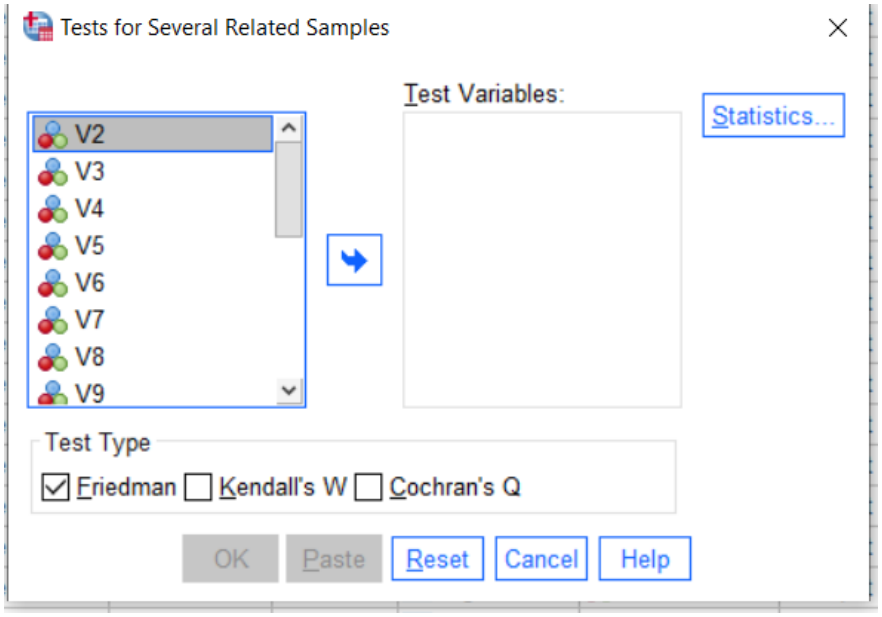
As with the paired samples t test, for this test you will move each paired variable (e.g., pre- and post-) over to the "Test Pairs" window on the right. Once you've done that, hit "OK" and your results will be generated in the output window.

Friedman

When you have three or more datapoints within individuals that you want to compare, you can use the Friedman test by navigating the dropdown menus as shown below.



Once you've selected "K Related Samples" the following window will pop-up. Note that the Friedman test is the default selection.

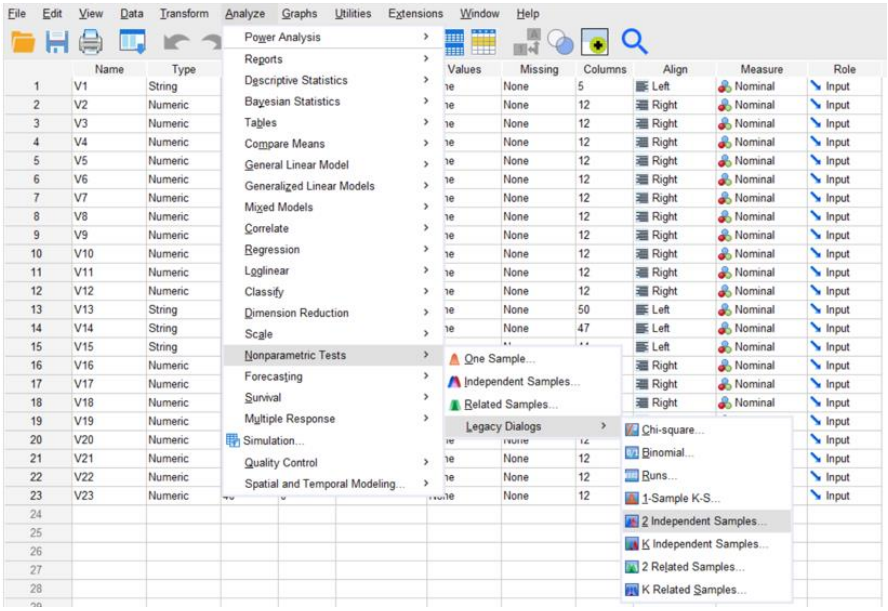


Move the three data points you wish to compare (note that you can have three or more, but in this example I'm using three) into the "Test Variables" window. In an example where you have pre-, post- and delayed post- data, you'd put those three variables in this box. Next, hit "OK" and your results will be generated in the output window.

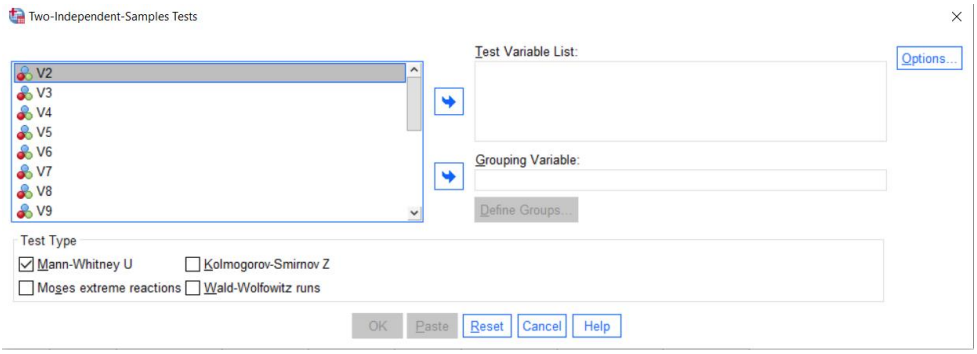
Between-Subjects

Mann-Whitney U

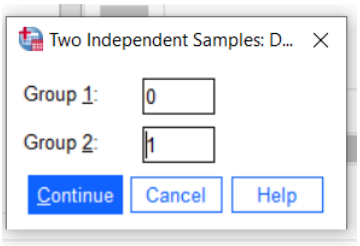
When you have two groups you wish to compare (e.g., a treatment and comparison group) you can use the Mann-Whitney U test by following the dropdown menu commands below.



Once you've selected "2 Independent Samples" the following window will pop up. Note that the Mann-Whitney U test is selected by default.



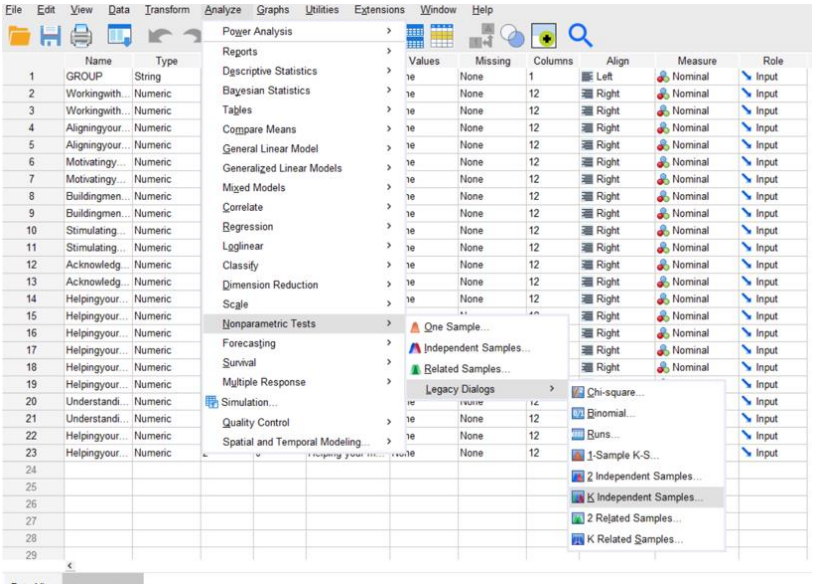
You'll start by putting your between-subjects variable ("Grouping Variable") into the second window on the right. Then, you'll click "Define Groups" and you will input the two numeric values of your group variable. For example, you may have the Group variable be 0 for comparison group and 1 for treatment group



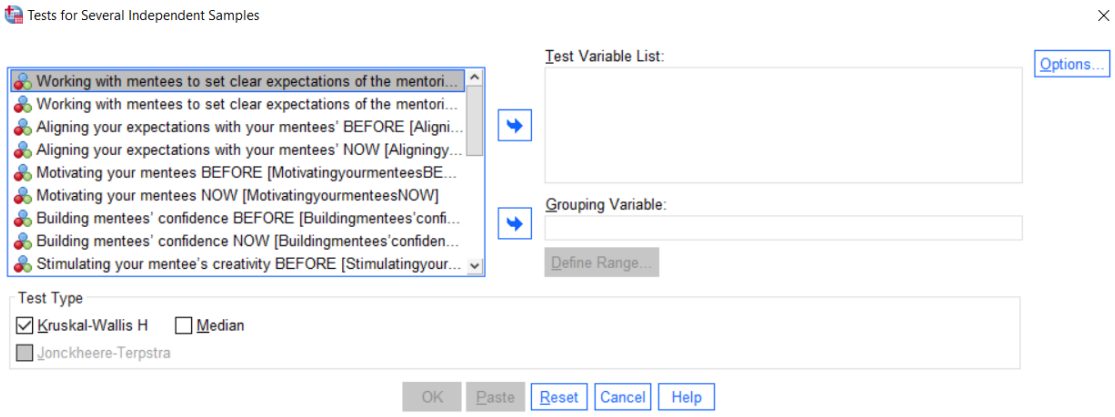
Next, hit "Continue" to close the "Define Groups" pop up window. Then you should move your outcomes of interest (the variables for which you're looking for between-group differences) into the top window on the right ("Test Variable List"). Then hit "OK" and your results will be generated in the output window.

Kruskal-Wallis H

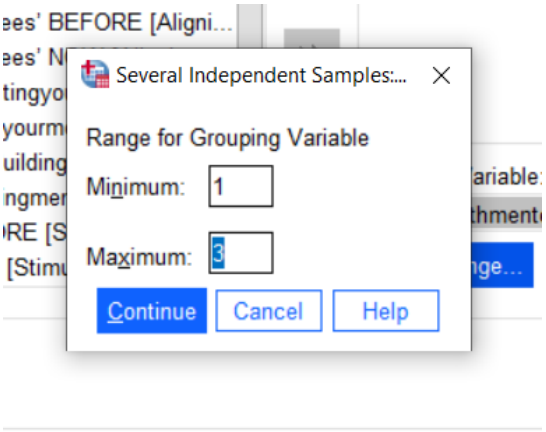
When you have three or more groups you wish to compare you can use the Kruskal-Wallis H test by using the commands from the dropdown menu as shown below.



Once you've selected "K Independent Samples" the following window will pop up. Note that the Kruskal-Walls H test is selected by default.



You'll start by putting your between-subjects variable into the second window on the right ("Grouping Variable"). Then, you'll click "Define Groups" and you will input the minimum and maximum values of your grouping variable. For example, say you had three groups in your study (e.g., a comparison group, Treatment version #1 group and Treatment version #2 group) and the grouping variable was coded such that comparison group = 1, Treatment version #1 = 2, and Treatment version #2 = 3. In this example the minimum and maximum values are 1 and 3.



Hit “Continue” to close the “Define Groups” pop up window. Then, move your outcomes of interest (the variables for which you’re looking for between-group differences) into the top window on the right (“Test Variable List”). Then hit “OK” and your statistical output will be generated in the output window.