# Using SPSS, Chapter 9:
# Hypothesis Testing - Two Samples

---
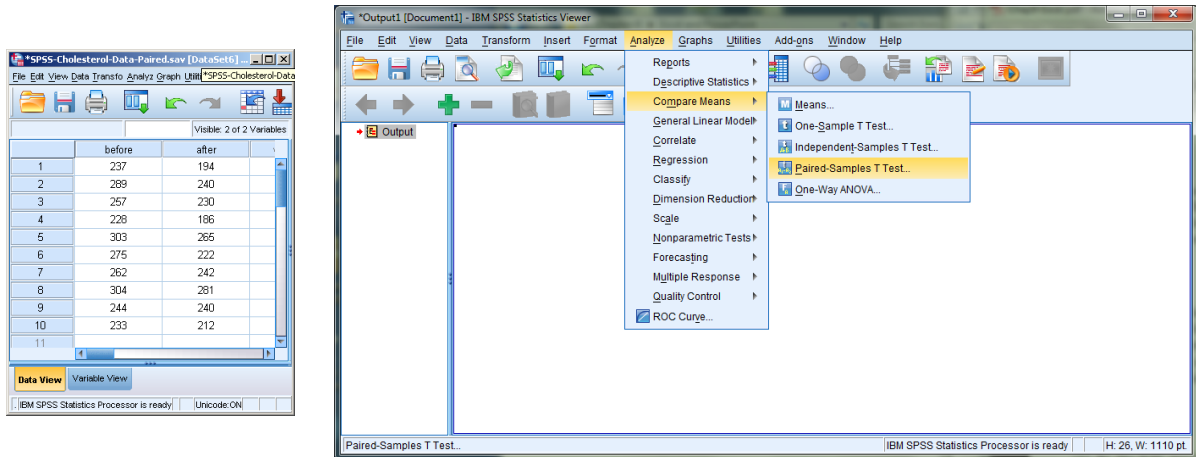
[1]Version 23 now can perform two-sample t-tests from summary statistics. The option is found under the `Compare Means` menu. I'll post instructions as soon as I get that version

# Chapter 9.1 - Hypothesis Tests for Mean Differences: Paired Data

- Create/Open the data.

- In the Statistics Viewer choose $\boxed{\texttt{Analyze}}$ → $\boxed{\text{Compare Means}}$ → $\boxed{\text{Paired-Samples T Test} \ldots}$



- This opens another dialogue box.
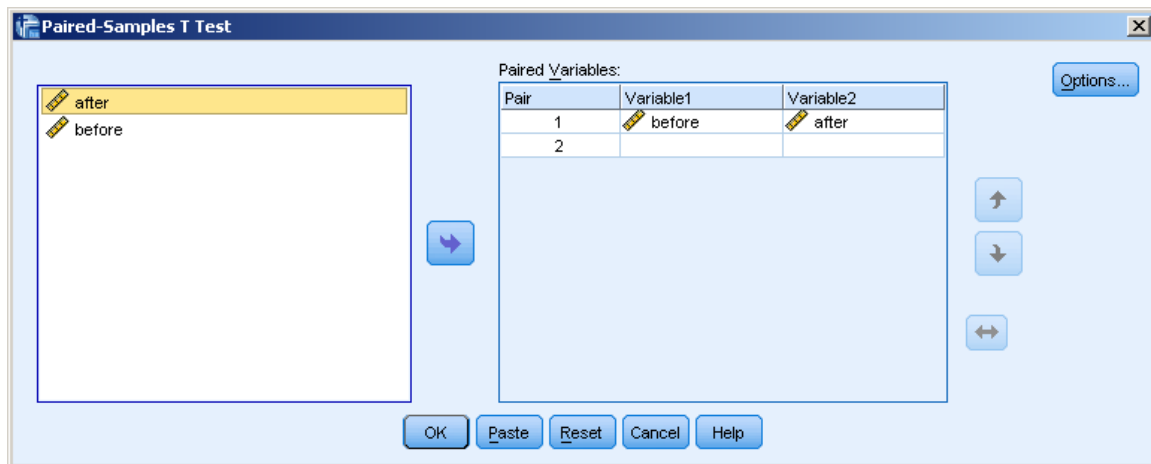


- Transfer one variable to `Variable 1` and the other to `Variable 2`.
  The test statistic is based on (`Variable 1`) - (`Variable 2`).
  You can do more than one set of pairs but we'll keep it simple here.

- If you click on $\boxed{\text{Options...}}$, you can choose the confidence level for the confidence interval given in the results.

- The results of this test are displayed in the Statistics Viewer.

The 95% confidence interval for the mean difference.

The Sample Means

The Mean Difference

The Test Statistic
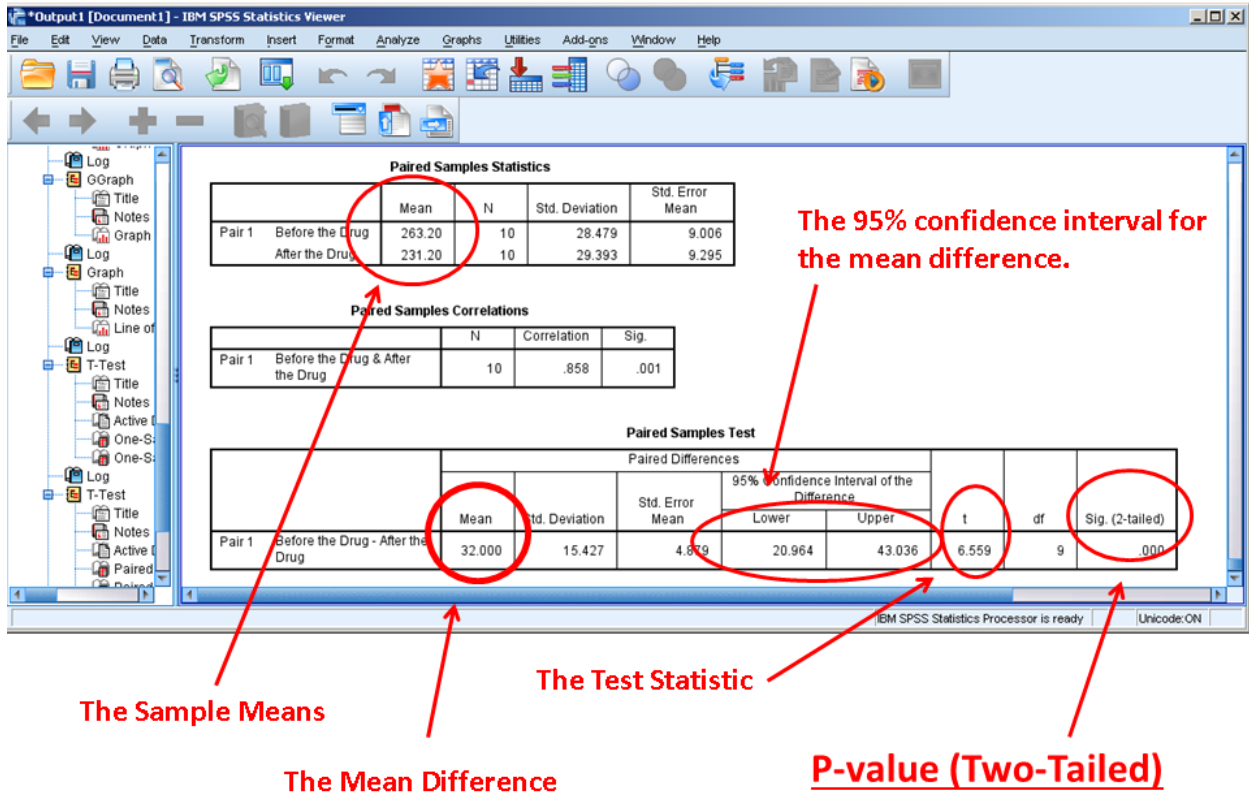
P-value (Two-Tailed)

- Here, the two-tailed $P$-value is given as 0.000 (it is actually 0.0001). If you were doing a one-tailed test you would use half of this as the $P$-value. Regardless of whether we were performing a one or two-tailed test, we would reject the null hypothesis that the mean difference is zero. Using the confidence interval approach, we are 95% confident that the mean difference is between 21 and 43.

- If you want to test for a mean difference greater than zero,

$$H_1 : \mu_1 - \mu_2 > \delta,$$

you will have to create another variable in your data set which is (Variable 1) - (Variable 2), then use a `One-Sample T Test` on this variable of differences as described in Chapter 8.3. SPSS will allow a non-zero hypothesized value with a one-sample test but not with a paired difference test.

# Chapter 9.2 - Hypothesis Tests for Two Means: Independent Data

- Create/Open the sample data. Be sure that each row corresponds to a single case. This means your different sample data will all be in one column where another column (variable) will determine the sample from which the measurement came.

- In the Statistics Viewer choose $\boxed{\texttt{Analyze}} \rightarrow \boxed{\text{Compare Means}} \rightarrow \boxed{\text{Independent-Samples T Test} \dots}$
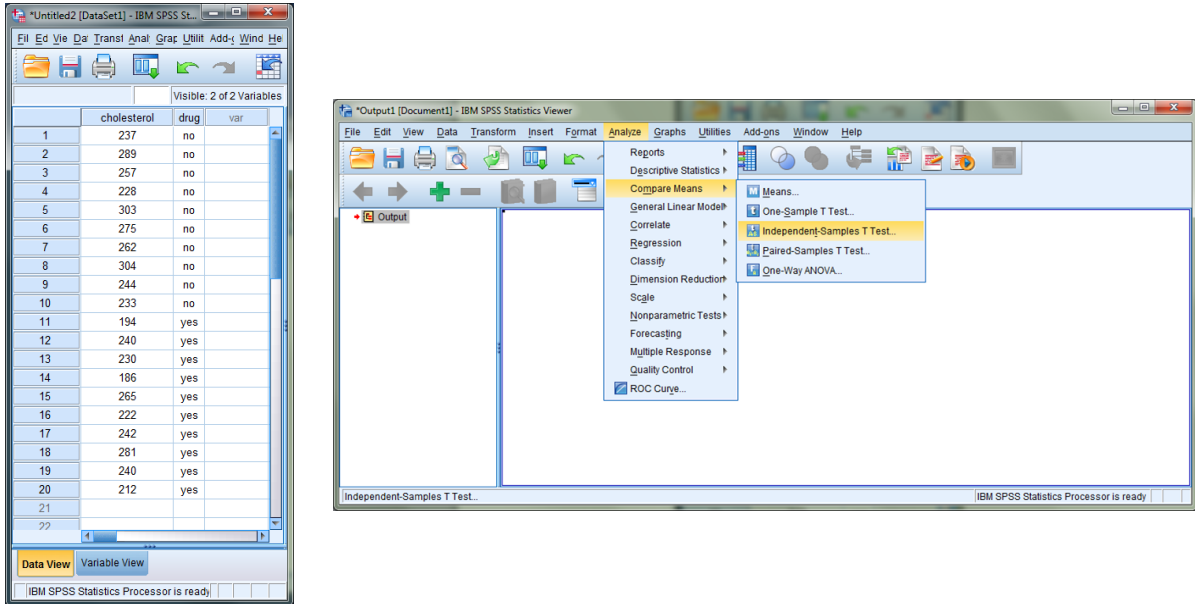


- This opens another dialogue box.
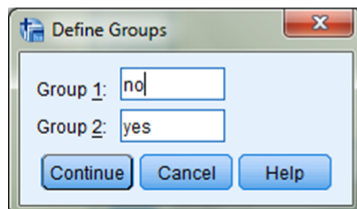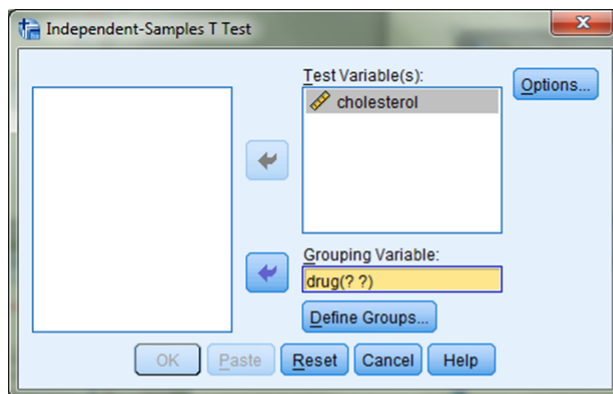


1) Transfer the variable you are testing (cholesterol)
2) Transfer the grouping (drug).
3) Define Groups (opens a new window)
4) Group 1: no to the drug.
5) Group 2: yes to the drug.
6) Continue then OK.

- Click $\boxed{\text{Options...}}$ to choose the confidence level for the confidence interval given in the results.

- The results of this test are displayed in the Statistics Viewer.
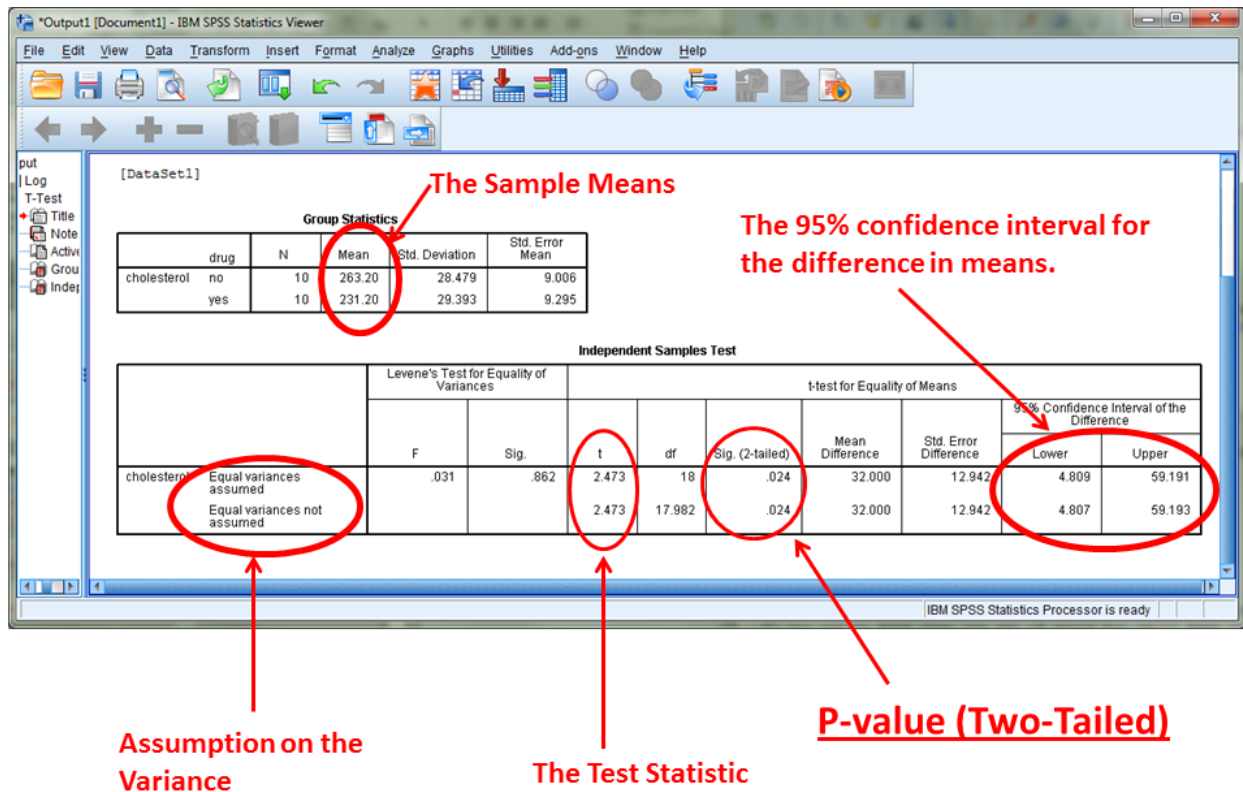


- Here, the two-tailed *P*-value is given as 0.024. Depending on your significance level you may or may not reject the null hypothesis of equal means. If you were doing a one-tailed test, the *P*-value would be $0.024/2 = 0.012$. Using the confidence interval approach, we are 95% confident that the difference in population means is between 5 and 59. That's a big interval. Notice, it doesn't matter whether we assume the variances are equal or unequal.

- Again, SPSS will only test that the difference in means is not equal to zero. You can't have a non-zero test value like you can for the one-sample t-test. I don't know why. There is a way around this but it entails subtracting the hypothesized difference in means from all of the cases of the variable with the larger mean.

# Chapter 9.3 - Hypothesis Tests for Two Proportions

SPSS doesn't do this test. There are plenty of online calculators that will. Generally you only have to input the proportion (or number) of successes and the sample size for each sample and hit a calculate button somewhere. It will usually give you a test statistic ($z$) and the $P$-value.

I like the collection of statistical calculators found at Social Science Statistics
http://www.socscistatistics.com/tests/Default.aspx.

**Usage**

- Click on the link above.
  Make a mental note of the variety of available statistical calculators. It's a useful collection.

- Choose $Z$-test calculator for two population proportions.

- Click on the $\boxed{\text{Take me to the calculator!}}$ button.

- Enter the number (or proportion in decimal) of successes ($x$) in the first sample.

- Enter the size of the first sample.

- Do the same for the second sample.

- Choose your significance level.

- Choose one or two-tailed.

- Click on the $\boxed{\text{Calculate Z}}$ button.

- It gives you the test statistic ($Z$-score) and the $P$-value in blue at the bottom of the page.

- Easy Peasy Lemon Squeezy.

**In case this calculator disappears, there are others at the following fine locations.**

- in-silico: http://in-silico.net/tools/statistics/ztest. Use `Equal` Variance.

- AusVet: http://epitools.ausvet.com.au/content.php?page=z-test-2.

- VassarStats: http://vassarstats.net/propdiff_ind.html.

# Creating and Importing Data

- There are two ways to get data into SPSS.

  - You can enter the data by typing it directly into the data editor.
  - You can open an existing data file by selecting the $\boxed{\texttt{File}}$ tab, then $\boxed{\texttt{Open}}$, then $\boxed{\texttt{Data...}}$. Then select the type of file from the list of options. If it is not already an SPSS (`.sav`) data file, you will be prompted to answer some questions. For example, if you open an Excel file it may ask which worksheet and whether or not the first row contains labels.

- Make sure your data is formatted as described below.

  - <u>Rows = Cases</u>
    Each row represents a case such as each respondent to a questionnaire.
  - <u>Columns = Variables</u>
    Each column represents a variable being tracked or measured. For example, the answers to a specific question on a questionnaire defines it's own variable (column). As such, each row represents an individual case for all variables.
  - <u>Cells contain values</u>
    Each cell contains a single value of a variable for a case.

  It is possible to enter data in the form of a frequency table but then you must do some alterations before analyzing such data.

- Once you have the data opened in the data editor, click the $\boxed{\texttt{Variable View}}$ tab at the bottom of the data editor. In this view, each variable is now a row and you must make sure all your variables are defined appropriately. The most important distinctions are

  - $\boxed{\texttt{TYPE}}$: The most common types are
    * `Numeric`: Used for quantitative data. These are numbers with no commas and a period delimiting the decimal places. SPSS will not allow you to enter non-numeric characters into a cell of numeric type.
    * `Date`: Used for dates or times from a menu of formats.
    * `String`: Used for qualitative data. Avoid symbols such as *, -, +, ?, etc.
  - $\boxed{\texttt{Measure}}$: There are three levels of measurement.
    * `Scale` is for ratio or interval levels of measurement.
    * `Ordinal` is for ordinal or ranked data.
    * `Nominal` is for qualitative data.
  - $\boxed{\texttt{Values}}$: If you have numeric values representing qualitative data such a 1=male and 0=female, you will probably want this to be labelled accordingly in graphs and outputs. Click on the cell in the `Values` column for that variable and assign labels for each value.