

Lab 5: Introduction to the t-Test for Two Independent Samples

In the previous lab we worked on the PCAP 2013 data using descriptive analysis techniques. Now, we will turn our focus to investigate inferential analysis techniques. In this lab, we will focus on group differences between **two** groups of students.

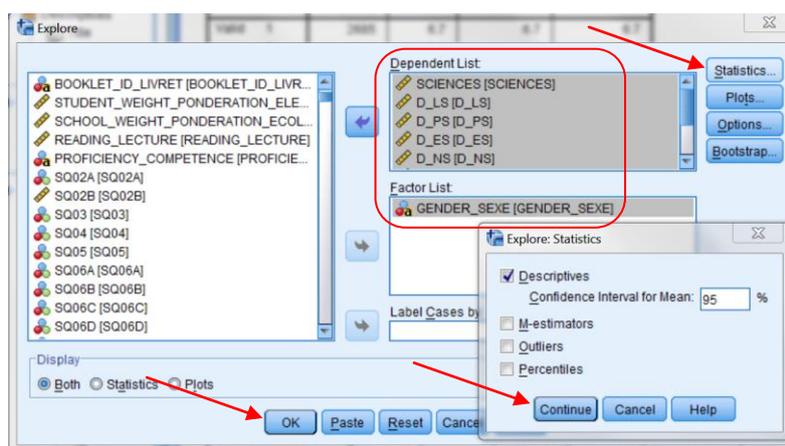
1. The t-test for independent samples – gender on performance

You are interested in whether there is a gender difference in terms of students' academic achievement in science, math, and reading (no subscores). The first step is to provide a descriptive table of the relevant variables, but you previously developed a descriptive table for these variables (during lab 2), so there is no need to re-develop a descriptive table again.

Question:

1. Please write the null hypothesis and alternative hypothesis for each of the continuous variables.

After explicitly identifying your hypotheses, it is important to provide some descriptives. Please develop a table to compare male and female students' academic achievement in science, math, and reading (excluding subscores). To investigate the gender groups, select Analyze → Descriptive Statistics → Explore → select all the variables you are interested in and move to the Dependent List box → place GENDER_SEXE into the Factor List box → select Statistics → ensure Descriptives is checked → select Continue → select OK.



The table below may be helpful as you develop your table.

Table 1

Title (please make a title for this table)

	Male				Female			
	Mean	Standard Deviation	Skewness	Kurtosis	Mean	Standard Deviation	Skewness	Kurtosis
Science								
Reading								
Math								

When you write a paper, you cannot just insert a table into the manuscript without an explanation. Please write a short paragraph to accompany the table you made, as if you were including the table into your published paper.

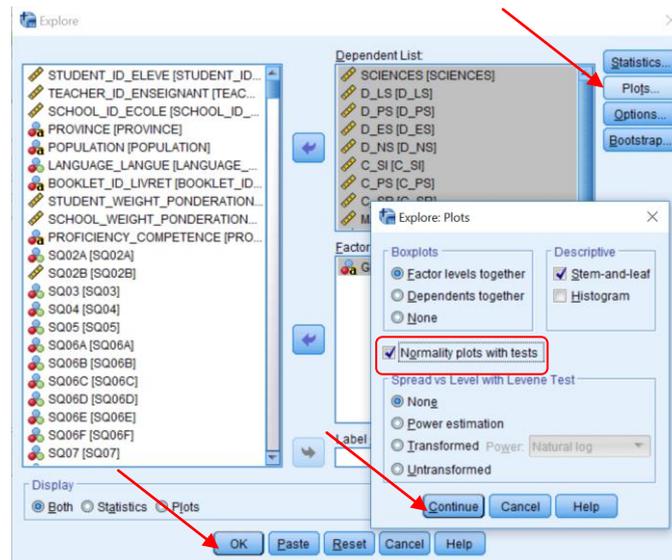
The next step is to test the three assumptions for your data: independence of sample, normality, and homogeneity of variance. When these assumptions are violated, the results of the analyses can be misleading or erroneous.

Question:

2. Please explain whether or not your two groups satisfy the independence of sample assumption. If it satisfies the assumption, please explain why (i.e., provide qualitative evidence).

You need to provide empirical evidence of satisfying the normality and homogeneity of variance assumptions, as well. To test the normality assumption, we can physically look at the histograms of the data for the two groups, but human judgement is relatively subjective. Hence, we rely on objective, statistical tests, such the Kolmogorov-Smirnov and Shapiro-Wilk tests of normality. To run these tests, click Analyze → Descriptive Statistics → Explore → move the achievement scores into the Dependent List box → move gender into the Factor List box → click Plots → check off Normality plots with tests → click Continue → click OK.

Please ensure your student weights are used throughout the analyses.



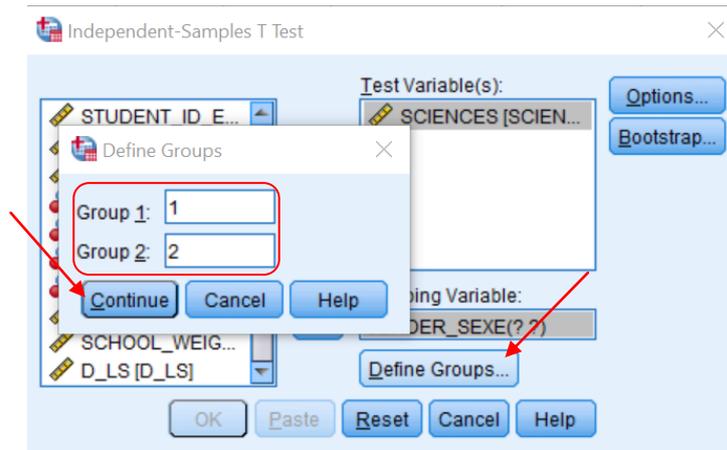
Now, analyze the tests of normality for each group of students. When the Kolmogorov-Smirnov and Shapiro-Wilk tests of normality are significant, it means the normality of the two groups' data are significantly different (i.e., the normality assumption is not met).

Questions:

3. What is the difference between the Kolmogorov-Smirnov and Shapiro-Wilk tests of normality?
4. Which test (i.e., Kolmogorov-Smirnov and Shapiro-Wilk) should you use and why?
5. Using the normality test that you chose in question 3, please explain whether or not your two groups satisfy the normality assumption. Please provide statistical evidence. Your response should follow the format “The data for this study [do/do not] satisfy the normality assumption at $p < 0.05$.”

To test the homogeneity of variance, we can, again, physically look at the histograms and variance values for the continuous variables, but it is difficult to tell if they are in fact different. Hence, we rely on statistical tests, such as Levene's test for equality of variances. Similar to the normality tests, when Levene's test is significant, it means the variances of the two groups' data are significantly different (i.e., the homogeneity of variance assumption is not met). Levene's test is automatically conducted with the t-test. To run a t-test for independent samples (e.g., gender) please click Analyze → compare Means → Independent-Samples T Test → move *all the relevant variables* into the Test Variable(s) box → move GENDER_SEXE into Grouping Variable → Define Groups... → use specified values → specify Group 1 = 1 and Group 2 = 2 → Continue → Options → check off “exclude

cases listwise" → Continue → OK



To read your output, you will need to look at the Independent Samples Test table. Use the ‘Sig’ columns to help you determine whether or not your data is statistically significantly different between the two groups. Significance values that are below 0.05 indicate a statistically significant difference between the two groups. For example, under the Sig. column under Levene’s Test for Equality of Variances, if we read across the READING_LECTURE variable row, the Sig. = 0.821, which indicates the variance between the two genders is not statistically significant, i.e., the two sets of data have equal variances assumed. Therefore, we read the top row of the reading variable. Now, we see that the Sig. (2-tailed) for the t-test is 0.000, which is less than 0.05, so the results indicate there is a statistically significant difference between the two genders in terms of reading achievement.

Independent Samples Test										
		Levene's Test for Equality of Variances				t-test for Equality of Means				
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
SCIENCES	Equal variances assumed	.285	.593	-1.438	2991	.150	-5.27193	3.66556	-12.45919	1.91534
	Equal variances not assumed			-1.439	2984.146	.150	-5.27193	3.66478	-12.45769	1.91383
MATH	Equal variances assumed	3.598	.058	-1.572	2991	.116	-4.64839	2.95657	-10.44551	1.14873
	Equal variances not assumed			-1.570	2960.402	.116	-4.64839	2.96037	-10.45298	1.15620
READING_LECTURE	Equal variances assumed	.051	.821	-10.367	2991	.000	-31.37857	3.02687	-37.31353	-25.44361
	Equal variances not assumed			-10.368	2983.491	.000	-31.37857	3.02642	-37.31265	-25.44449

Question:

- Using Levene’s test, please explain whether or not your two groups satisfy the homogeneity of variance assumption. Please provide statistical evidence. Your response should follow the format “The data for this study [do/do not] satisfy the homogeneity of variance assumption, $p[</>]0.05$.”

Since the tests for the assumptions are not the main focus of our research, we usually do not

need to provide too many pieces of evidence (e.g., Kolmogorov-Smirnov, Shapiro-Wilk, and Levene's F statistics) to support our claims that our data satisfy or violate these assumptions. We generally only need to clearly indicate whether we statistically significantly satisfy or violate these assumptions.

Questions:

7. Does your data satisfy all three assumptions?
8. It has been said that "the t-test is robust against violations of normality." What does this mean? (You may need to use Google for this item.)
9. Would you recommend continuing with the analyses by using a t-test to analyze your data?

Regardless of your response to Question 9, we will continue to analyze the results of the t-test.

Questions:

10. For each of the three variables, please indicate the t-value, degrees of freedom, and observed p-value. (**Note:** if the homogeneity of variance assumption is not met, you will need to use the data from the second line.)
11. Looking at the p-values, which (if any) of your variables land in the critical region?
12. Are there gender differences between students' achievement in science, math, and reading? Answer this question using the format "There was a statistical [significant/non-significant] gender difference on [variable name] in terms of students' achievement, $t(df)=[t\text{-value}]$, $p[</>]0.05$." Please ensure you follow APA rules when writing this sentence.
13. Using the mean difference* and standard deviation from your SPSS output, please manually calculate Cohen's d and percentage of variance explained. (Hint: consider using an online calculator for these values – Google is helpful for this item). What does each of these values indicate (i.e., what does the effect size tell us)? **You need to calculate effect sizes only for variables that show significant differences between the two groups. **

***Note:** the mean difference is the difference between the two means you reported in your descriptive table. In the SPSS output it is simply called *mean*, but in your

publications you will need to indicate *mean difference* to show the difference between the mean reported as part of your descriptive analyses and the mean difference reported as part of your inferential t-test analyses.

2. The t-test for independent samples – birth location on performance

You recently read some literature that suggests students in Canada who were born in Canada perform better on achievement tests than students in Canada who were born elsewhere. Please note: this item focuses on birth place (i.e., born in Canada) and does not necessarily indicate “Canadian students” (i.e., nationality). You would like to test this theory. Please follow the steps above to answer the research question: Do students who are born in Canada, as self-identified by the students (i.e., SQ03), perform better on achievement tests in science, math, and reading?

Questions:

14. Please provide a table of descriptives (e.g., mean, standard deviation, standard error mean) for this new grouping. Note: only include statistics that are relevant to this study.
15. Following the steps above (i.e., section 1. **The t-test for independent samples – gender on performance**), please write a **brief** paragraph (no more than five sentences) explaining your results. In your paragraph, you should (a) clearly state whether your data satisfy the three statistical assumptions (you may need to explain each one separately), (b) answer the research question, (c) discuss the effect size of your inferential tests.
16. SPSS provides you with the p-value for a 2-tailed test. What would the p-value be for a 1-tailed test? (You may need to use Google for this question.)

You have now completed the lab.