

SOC 4930/5050: Lab-07 - Difference of Means Tests by Hand

Christopher Prener, Ph.D.

October 9th, 2016

Directions

Please complete all steps below. Your your work “by hand” as well as your well-formatted R Notebook source (the .Rmd file) and html output should be uploaded to your GitHub assignment repository by 4:15pm on Monday, October 16th, 2017. The following data represent standardized test scores on a variety of tasks.

Part 1: One-sample T-test

Variable	Obs	Mean	Std. Dev.	Min	Max
-----+-----					
math	200	52.645	9.368448	33	75

1. Using the above data, test to see whether the sample data comes from a population where the average score on the math portion of a standardized test is 52. Be sure to provide a complete interpretation of the results.
2. Test to see whether the sample data comes from a population where the average score on the math portion of a standardized test is 54. Be sure to provide a complete interpretation of the results.

Part 2: Independent T-test

Writing Scores by Gender

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
-----+-----						
male	91	50.12088	1.080274	10.30516	47.97473	52.26703
female	109	54.99083	.7790686	8.133715	53.44658	56.53507
-----+-----						
combined	200	52.775	.6702372	9.478586	51.45332	54.09668
-----+-----						

3. Assuming *equal* variances, test to see whether there is a significant difference in writing scores between men and women in this sample. Be sure to provide a complete interpretation of the results.
4. Based on your answer to question 3, calculate and interpret the appropriate effect size.
5. Assuming *unequal* variances, test to see whether there is a significant difference in writing scores between men and women in this sample. Be sure to provide a complete interpretation of the results.
6. Based on your answer to question 5, calculate and interpret the appropriate effect size.

Part 3: Dependent T-test

Variable	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
math	200	52.645	.6624493	9.368448	51.33868	53.95132
science	200	51.85	.7000987	9.900891	50.46944	53.23056
diff	200	.795	.5864593	8.293787	-.3614723	1.951472

7. Since there is overlap between math and science skills, it is possible that these two scores are not independent. Test to see whether there is a significant difference in math and science scores in this sample. Be sure to provide a complete interpretation of the results.
8. Based on your answer to question 7, calculate and interpret the appropriate effect size.

Part 4: Reshaping Data

The following data include Gini coefficients at two different time periods for three of the four so-called “BRIC” countries (Brazil, Russia, India, and China), which represent major developing countries. Gini coefficients range from 0 (complete income equality) to 1 (complete income inequality).

country	period	gini
Brazil	2008	.544
Brazil	2012	.527
China	2008	.428
China	2012	.422
Russia	2008	.414
Russia	2012	.416

9. If we wanted to reshape these data, which verb is most appropriate? Why?
10. What is the key?
11. What is the value?
12. Draw out a reshaped data table with new variable names and values filled in.

The following data include population counts for three cities in the United States at two different time periods.

country	pop1900	pop2000
Los Angeles	102479	3695364
New York	3437202	8008278
St. Louis	575328	346904

13. If we wanted to reshape these data, which verb is most appropriate? Why?
14. What could we name the key?
15. What could we name the value?
16. Which variables (i.e. columns) will contribute to the values?
17. Draw out a reshaped data table with new variable names and values filled in.