

# SOC 4930/5050: Lab-08 - Difference of Means Tests

Christopher Prener, Ph.D.

October 16<sup>th</sup>, 2017

## Directions

Please complete all steps below. Your well-formatted R Notebook source (the .Rmd file) and html output along with your L<sup>A</sup>T<sub>E</sub>X pdf output should be uploaded to your GitHub assignment repository by 4:15pm on Monday, October 30<sup>th</sup>, 2017. You will need to have the updated testDriveR package installed (see the Wiki/Jotter).

## Part 1: Data Preparation

1. Using the data table childMortality in the testDriveR, create a binary variable that is TRUE when continent == "Africa" and is otherwise FALSE.
2. Create a subset of your data that contains only neonatal mortality rate data for 2015.<sup>1</sup>
3. Create a subset of your data that contains only under-5 mortality rate data for 1995 and 2015.<sup>2</sup> These data should be formatted as *long* data, and the year variable should be converted to character using base::as.character() as part of a dplyr::mutate() call.
4. Create a copy of the 1995 and 2015 data set that is formatted as *wide* data.<sup>3</sup>

<sup>1</sup> Hint: Use the dplyr::filter() function.

<sup>2</sup> Hint: Use the dplyr::filter() function.

<sup>3</sup> Hint: Be sure to remove all variables except countryName, year, and estimate.

## Part 2: One-sample T Test

5. What is the mean number of deaths per 1,000 live births in the 2015 sample of neonatal mortality?
6. A study suggests that the true population mean ( $\mu$ ) is 15. Is it possible that our sample is drawn from that population?
7. Another study suggests the true population mean ( $\mu$ ) is 12. Is it possible that our sample is drawn from that population?

### *Part 3: Independent T Test*

8. Test the homogeneity of variance assumption for neonatal mortality between African nations and all other countries. Do these data meet that assumption? Should we use pooled variance estimates or not?
9. Assess the other assumptions for the independent t test. Which assumptions are met or not met?
10. Calculate and interpret the appropriate version of the independent t test, and provide an interpretation that includes an assessment of the validity of your findings given the assumption analyses above.
11. Calculate and interpret a Cohen's  $d$  effect size for the difference in neonatal mortality between African nations and all other countries.
12. Create a box plot and a violin plot of the mean neonatal mortality between African nations and all other countries. Provide a written interpretation of both plots.

### *Part 4: Dependent T-test*

13. Test the normality of the *difference* between under-5 mortality rates in 1995 and 2005. Do these data meet this assumption?
14. Assess the other assumptions for the dependent t test. Which assumptions are met or not met?
15. Calculate and interpret a dependent t test on these data - has there been a change in mean under-5 mortality rates between 1995 and 2005?
16. Calculate and interpret a Cohen's  $d$  effect size for the difference in under-5 mortality rates in 1995 and 2005.
17. Create a violin plot and a ridge plot of the mean neonatal mortality between 1995 and 2005. Provide a written interpretation of both plots, and compare them - which one seems more effective?

### *Part 5: Sample Size Estimates*

18. What is the sample size needed to detect a small effect ( $d = .2$ ) at  $\alpha = .05$  when  $\beta = .90$  for a two-tailed independent t test.

19. What is the sample size needed to detect a moderate effect ( $d = .5$ ) at  $\alpha = .05$  when  $\beta = .80$  for a two-tailed dependent t test.
20. What is the sample size needed to detect a large effect ( $d = .8$ ) at  $\alpha = .05$  when  $\beta = .75$  for a two-tailed independent t test.

*Part 6: A Simple L<sup>A</sup>T<sub>E</sub>X Document*

21. Using R, produce a descriptive statistics table for the neonatal mortality data from 2015. Copy-and-paste this table into a new L<sup>A</sup>T<sub>E</sub>X project. Your document should have a title with an author and a date, and should have a single section with two subsections. In the first subsection, quickly describe how the data were created by you. In the second subsection, provide the table. Use `typewriter` text for the variable names in the table (you will have to edit the L<sup>A</sup>T<sub>E</sub>X output you added to your project).