SOC 4015/5050: Lab-07 - Difference of Means Tests

Christopher Prener, Ph.D.

Fall 2018

Directions

Please complete all steps below. All work should be uploaded to your GitHub assignment repository by 4:15pm on Monday, October 29th, 2018.

Analysis Development: Create a Project Folder System

This initial section follows the project workflow that is available in the `lecture-03` repo!

a. Using RStudio, add an R Project to the *existing* directory in your assignments repository named `Lab-07`.

b. Add new folders as needed to you project.

c. Create a new text file for your `README.md` and include details on all files associated with yoru `Lab-07` project.

d. Create a new notebook with an expanded YAML heading.

e. Make sure your notebook has *completed* introductory, package loading, and data loading sections before proceeding with the parts below.

f. Be sure to "knit" your notebook at the end of the assignment!

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. Write the data into a `.csv` file in your `data/` subdirectory.[1]
3. Create a subset of your data that contains only under-5 mortality rate data for 1995 and 2015.[2] 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. Write the data into a `.csv` file in your `data/` subdirectory.

[1] *Hint:* Use the `dplyr::filter()` function.

[2] *Hint:* Use the `dplyr::filter()` function.

4. Create a copy of the 1995 and 2015 data set that is formatted as *wide* data. Write the data into a `.csv` file in your `data/` subdirectory.[3]

[3] *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 (μ) is 15. Is it possible that our sample is drawn from that population? Write the results of your test into a `.csv` file in your `results/` subdirectory.
7. Another study suggests the true population mean (μ) is 12. Is it possible that our sample is drawn from that population? Write the results of your test into a `.csv` file in your `results/` subdirectory.

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 used pooled variance estimates or not? Write the results of your test into a `.csv` file in your `results/` subdirectory.
9. Assess the other assumptions for the independent t test. Which assumptions are met or not met? Save any plots you create as `.png` files at `dpi = 300` into your `results/` subdirectory.
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. Write the results of your test into a `.csv` file in your `results/` subdirectory.
11. Calculate and interpret a Cohen's *d* effect size for the difference in neonatal mortality between African nations and all other countries. Write the results of your test into a `.csv` file in your `results/` subdirectory.
12. Create a stats plot of your choice of the mean neonatal mortality between African nations and all other countries. Save the plot as a `.png` file at `dpi = 300` into your `results/` subdirectory.

13. Create a violin plot of the mean neonatal mortality between African nations and all other countries. Provide a written interpretation of this plot. Save the plot as a `.png` file at `dpi = 300` into your `results/` subdirectory.

Part 4: Dependent T-test

14. Test the normality of the *difference* between under-5 mortality rates in 1995 and 2015. Do these data meet this assumption? Save any plots you create as `.png` files at `dpi = 300` into your `results/` subdirectory.
15. Assess the other assumptions for the dependent t test. Which assumptions are met or not met?
16. Calculate and interpret a dependent t test on these data - has there been a change in mean under-5 mortality rates between 1995 and 2015? Write the results of your test into a `.csv` file in your `results/` subdirectory.
17. Calculate and interpret a Cohen's d effect size for the difference in under-5 mortality rates in 1995 and 2015. Write the results of your test into a `.csv` file in your `results/` subdirectory.
18. Create a stats plot of your choice of the mean under-5 mortality between 1995 and 2015. Save the plot as a `.png` file at `dpi = 300` into your `results/` subdirectory.
19. Create a ridge plot of the mean under-5 mortality between 1995 and 2015. Provide a written interpretation of this plot. Save the plot as a `.png` file at `dpi = 300` into your `results/` subdirectory.

Part 5: Sample Size Estimates

20. 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.
21. 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.
22. 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.