Homework: Inference

due 2019-11-29 at 5:00 pm

Goals

This homework serves two goals

- for you to use R locally.
- make inference for difference of two proportions and single mean

Assignment

Type: Individual

Tasks:

For this homework you will be using RStudio on a local machine, either your own personal computer or a school computer. Note that RStudio should be available on most computers across campus.

All this time, we have been working with R Markdown files and knitting them to pdf documents on the Cloud. Whenever you click File > New File > R Markdown it gives you three different file types as a choice, HTML, PDF, and Word. I normally use HTML and teach with HTML. However, the reason we have been using PDF all this time was because Gradescope only accepts pdf file uploads!

For this homework we will be using HTML and one the homework is completed we will make it a pdf. The reason is because pdf files requires extra installation on your computer. However, if you are interested in making these installations such as installing TinyTex you are free to work with pdf. You can find instructions **here (https://bookdown.org/yihui/rmarkdown/installation.html)** on how to download TinyTex.

Make sure to watch the video on how to create R projects.

Homework Questions

1. OpenIntro 6.22 Sleep deprivation, CA vs. OR Part I
2. OpenIntro 6.24 Sleep deprivation, CA vs. OR Part II
3. Given the following information, calculate p-value and decide whether to the null hypothesis would be rejected at $\alpha = 0.05$. Assume conditions have been met.

a. n = 135, T = 4.32, $H_A : p \neq 0.5$

b. n = 9, T = -1.67, $H_A : p < 0.3$

c. n = 24, T = -5.64, $H_A : p > 0.2$

5. For this question we will be working with [Breast Cancer Coimbra Data Set]**https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Coimbra (https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Coimbra)**). Click on the Data Folder to download the data in `csv` format.

Make sure to familiarize yourself with the data by reading about the variables on the website. Note that the data comes this **study (https://bmccancer.biomedcentral.com/articles/10.1186/s12885-017-3877-1)**. It is reported in the study that the blood draws were taken when participants were fasting. According to [Mayo Clibic] a fasting blood sugar level less than 100 is mg/dL is normal.

a. In the data provided to you, in other words in the sample, do breast cancer patients have normal glucose level on average? Calculate sample statistic. Make sure to share your code and comment on your finding.

b. Estimate 95% confidence interval for average glucose level of breast cancer patients.

c. Set up and test hypotheses to determine whether breast cancer patients have normal glucose levels or not.

Scoring: 25 points

Submission

In the Files pane in the lower right part of RStudio Cloud window check the `hw6-prop-inference.pdf` file. Click `More`, then `Export` and then `Download`.

Upload the downloaded homework on Gradescope.

Come to our **office hours (https://mdogucu.ics.uci.edu/teaching/stats67-fa19/office.html)** if you need help with questions from the book, R, or how to submit your homework.