Final Exam Practice Problems

Fall 2018

Course topics

The final exam will include many of the topics below. Part I of the exam will ask you to select the appropriate descriptive and inferential statistics in several situations, conduct analyses, and interpret results. You will be able to use materials from the course, the internet, etc (no humans other than the instructor, though) to do this part of the work. Part II of the exam will be multiple choice; no materials other than a calculator are permitted. *You will not need to know how to do anything in R for the multiple choice, it is on statistical concepts only.*

Course concepts to review:

- variable types (continuous vs. categorical)
- central tendency (mean, median, mode)
- spread (standard deviation, variance, range, IQR, standard error)
- skew
- frequencies and percentages
- confidence intervals
- the normal distribution
- z-scores
- correlation coefficients
- chi-squared
- one-sample and independent-samples t-tests
- sign test, Mann-Whitney U test, K-S test
- ANOVA
- Kruskal-Wallis test
- linear and logistic regression
- data management (recoding variables, missing values)

Practice data

Load the smokers data set. The data was saved as a CSV file at this URL: http://tinyurl.com/z2m3cgq

Smokers codebook

The data are from a survey of Missouri smokers.

- id: unique identifier for each participant
- numcigs: number of cigarettes smoked per day
- ageonset: age started smoking
- yearssmoking: years as a smoker so far
- packyears: number of packs per day * years smoking
- age: age in years
- vage_g: age in years in categories (18-24, 25-34, 35-44, 45-54, 55-64, 65+)
- marital_status: marital status in categories (1-Married, 2-Divorced, 3-Widowed, etc.)
- employment: employment status in categories
- Income: income in categories
- sex: participant sex in categories (1-Male, 2-Female)
- VBMI4: body mass index
- VBMI4CAT: body mass index in categories
- vracegr: race in categories

- educa: education level in categories
- income2: income in fewer categories
- sexorien: sexual orientation in categories
- rucacat: rurality in categories
- livesmoke: lives with a smoker (1-Yes, 2-No)
- expsmoke: exposed to secondhand smoke (1-Yes, 2-No)
- smokehome: smoking allowed at home (Allowed, Not allowed)
- expsmokecar: exposed to secondhand smoke in a car (1-Yes, 2-No)
- nosmokecar: smoking allowed in your car (Allowed, Not allowed)
- expwork: exposed to secondhand smoke at work (1-Yes, 2-No)
- protctwrk: protected from secondhand smoke at work (1-Yes, 2-No)
- protctpub: protected from secondhand smoke in public places (1-Yes, 2-No)
- strongerpol: should be stronger policies protecting from secondhand smoke (1-Yes, 2-No)
- final_wt_scale: survey weights

Practice problems

Be sure to check assumptions as you are selecting statistical tests. For any ANOVA or Kruskal-Wallis significant results, use a post-hoc test to find which groups are different from each other. For any chi-squared significant results, review standardized residuals to determine if any groups are significantly larger or smaller than expected. Note that these tasks may not cover every single test from class, but are examples of the sorts of questions you will likely encounter on Part I of the final. The smokers data has a good mix of continuous and categorical variables, so you should be able to practice all of the tests from class using the variables in this data set. Here are a few things to try:

- Compute the appropriate descriptive statistics for 10 of the variables in this data set and produce a table OR a paragraph that uses these descriptive statistics to describe who is in the sample
- Choose, conduct, and interpret the appropriate statistical test to examine the relationship between age of initiation and living with a smoker
- Choose, conduct, and interpret the appropriate statistical test to examine the relationship between livesmoke and exposed to secondhand smoke
- Choose, conduct, and interpret the appropriate statistical test to examine age by VBMI4CAT
- Choose, conduct, and interpret the appropriate statistical test for examine the relationship between motivation to quit and percentage of friends who smoke.
- Predict livesmoke based on age and educa
- Predict packyears based on age