MATH 338

FINAL EXAM

SOFTWARE PORTION

DUE: WED, MAY 16, 2018 at 9:30 PM

Exam Rules

You may refer to your textbook, any notes/code you wrote, and anything on Titanium. You may refer to other books, other Sapling resources, or the wider Internet, but if you use these resources to help you answer questions, you must cite them properly.

You may ask Dr. Wynne to clarify what a question is asking for, or to help you troubleshoot RGuroo errors and/or debug your R code. You may not ask any other people for help (on- or offline).

For full credit, include all R code (if using RStudio), graphs, and output. Save your answers as a .docx or .pdf file and upload the file to Titanium.

Honor Statement

I certify that all work on this exam is my own, and that I have neither given nor received unauthorized help on the exam.

Name: _____

Date: _____

Exam Instructions

This portion of the exam consists of three problems. Read the entire text of each problem and import the associated dataset into your software of choice. Then, for each problem:

A) [1 pt] Perform background analysis to determine the most appropriate inferential procedure to use to answer the question. Write the procedure you will use. Be as specific as possible.

B) [1.5 pts] Perform background and exploratory data analysis to determine whether the assumptions of the procedure you identified in part (A) have been met.

C) [0.5 pts] State explicitly whether all assumptions have been met or at least not terribly violated.

For **one** of the problems, there is a major violation of at least one assumption. **Do not perform any inference for this problem. Simply state that the suggested inferential procedure is not appropriate and explain why.**

For the remaining **two** problems:

D) [3 pts] Using a default 95% confidence level/5% significance level, perform the inferential procedure from part (A) in the statistical software, and answer the question.

You will receive 1 point for pasting relevant code and output (including error messages if you cannot get the code/dialog to work properly). The remaining 2 points will be earned for correctly restating the relevant part(s) of the output (for example, test statistic and p-value) and writing an appropriate interpretation/conclusion in the context of the question.

Exam Problem 1

By convention, a study should have a statistical power of at least 80% (0.8) to be considered a "good study." However, it is not uncommon for studies to be published despite having much lower power.

The file rheumatoid_arthritis_power.csv contains the following variables for 34 studies of risk factors for rheumatoid arthritis:

- Marker: the genetic or lifestyle risk factor being studied
- Year: the year the meta-analysis reporting the study was published
- Author: the first author of the meta-analysis reporting the study
- Power: the estimated power of the study investigating the marker (reported as a proportion)

You can consider Year and Author to jointly describe the "source" of the information used to estimate the power for each study.

Assuming that this sample is representative of all studies published about risk factors for rheumatoid arthritis, do rheumatoid arthritis studies, on average, achieve 80% power?

Important Note: This question does not ask you to compute power. Power is the variable of interest.

A) What inferential procedure would you use to answer this question?

B) Perform background and exploratory data analysis to determine whether the assumptions of the procedure have been met. Write and paste your analysis below.

C) Are all assumptions met (or at least not terribly violated)? If not, explain the violation(s).

D) If assumptions are violated, stop here. If assumptions are okay, perform the inference:

Exam Problem 2

Some biologists are interested in the relationship between the length and width of an animal's mouth, because that relationship can influence the animal's diet. Although the length and width are obviously both dependent on the animal's size, it is reasonable to assume that the relationship does not change in mature animals of the same species.

The file crocs.csv contains the following variables, measured on 25 American crocodiles (*Crocodylus acutus*):

- Length: the length of the mouth, in mm
- Width: the width of the mouth, in mm

Assuming that this sample is representative of all American crocodiles, estimate the population mean mouth width of all American crocodiles whose mouths are 600 mm in length.

A) What inferential procedure would you use to answer this question?

B) Perform background and exploratory data analysis to determine whether the assumptions of the procedure have been met. Write and paste your analysis below.

C) Are all assumptions met (or at least not terribly violated)? If not, explain the violation(s).

D) If assumptions are violated, stop here. If assumptions are okay, perform the inference:

Exam Problem 3

Maximum oxygen uptake volume (VO_{2max}, measured in mL per kg per minute) is a key indicator of aerobic fitness. A 2013 study suggested that a VO_{2max} of 60 mL/kg/min is the minimum threshold necessary to play men's professional soccer at an elite level. The VO_{2max} of severely injured athletes often decreases following injury.

The file ACL.csv contains the following variables, for 20 professional soccer players who underwent anterior cruciate ligament (ACL) reconstruction surgery:

- ID: the study ID number for the player
- Pre: the player's VO_{2max} measured after ACL injury but before the surgery
- Post: the player's VO_{2max} measured six months after the surgery

Assuming that this sample is representative of all players with ACL injuries requiring reconstruction surgery, estimate the amount by which an injured player's VO_{2max} increases due to the surgery and six-month rehabilitation period.

A) What inferential procedure would you use to answer this question?

B) Perform background and exploratory data analysis to determine whether the assumptions of the procedure have been met. Write and paste your analysis below.

C) Are all assumptions met (or at least not terribly violated)? If not, explain the violation(s).

D) If assumptions are violated, stop here. If assumptions are okay, perform the inference: