

OPIM 5603-B14 — Statistics in Business Analytics

Fall 2019, University of Connecticut

Homework 9 - v1

Instructions: Please complete the following questions and submit them as an RNotebook (as an Rmd file) via the submission link on HuskyCT. You must submit the assignment by the time and due date listed on the course syllabus. Failure to submit a file by the deadline will result in a score of 0 on the assignment.

Set the heading of the RNotebook as an `html_document`, with a table of contents and without numbered sections. Add your name and a date to the header as well. The solution to each problem should be a separate section (specified by `#`), and each subproblem should be set as a subsection (specified as `##`). For example, for Problem 2, you should have a section titled Problem 2, specified by:

```
# Problem 2
```

in your RNotebook. Also, for subproblem b in Problem 2, you should have a subsection, specified by:

```
## Problem 2b
```

As with all course material, the problems appearing in this homework assignment are taken from the instructor's real-world experiences, from other courses taught at the University of Connecticut, and from the sources listed in the course syllabus.

Note that R code submitted should work independent of the data that sits in the data structure. For example, suppose there was a vector `r_vec` with the values (1, 2, 6) and the problem asks for you to create R code to create a vector `answer` which doubles each element of `r_vec`. The answer

```
answer <- c(2, 4, 12)
```

would be given no credit. The answer

```
answer <- 2*r_vec
```

would be an appropriate answer.

You must show all steps in your solution. For example, if a problem asks for the expected value of a random variable that is binomially distributed with $n = 10$ and $\pi = 0.3$, and you simply write

3,

this will be given no credit. However,

$10 * 0.3$

would be given credit.

If you have any questions, please submit them via email to the instructor and/or the teaching assistant prior to submitting your solution.

Problem 1 (35 points)

One concern in Major League Baseball is that the time to play games has become too long. There have been recent efforts to enforce rules that will speed up the game. In the data set **Baseball Games** we have data reporting a set of games that were played on a given day, where the columns, in order, report the game index, the teams playing the game, the league (either AL or NL), the margin of victory, the number of pitchers, the attendance at the game, and the amount of time that the game took, in minutes. We will conduct a statistical test to determine if the average length of games is less than 213 minutes at the 99% confidence level.

- a. State the null hypothesis.
- b. State the alternative hypothesis.
- c. Can we use a parametric test? Explain in detail.
- d. **Independent from the answer given in the previous part, run a parametric statistical test.** What is the p -value for the statistical test? Show all steps.
- e. At a confidence level of 99%, would you reject or fail to reject the null hypothesis?
- f. If you conducted a t -test, use the `wBoot` package to conduct the same hypothesis test. If you conducted a bootstrap test using `wBoot`, do a t -test. Compare the results. How do they differ?
- g. The League is mostly concerned that at least 80% of games will conclude within 4 hours. They would like to make the statement ‘80% of our games conclude within 4 hours.’ Based on this data set, conduct a hypothesis test to evaluate if this claim is true at 90% and 98% confidence.

Problem 2 (35 points)

Two managers at a global finance firm are competing for a promotion. The CEO has asked each to record the time it took their respective teams to complete the projects assigned to them over the last week. File `ProjectCompletionTimes.xlsx` reports the results for both managers.

- a. Is there statistical evidence to conclude that the mean completion times for projects for the teams of each of the managers is different?
- b. Is there statistical evidence to conclude that the median completion times for projects for the teams of each of the managers is different?
- c. Is there statistical evidence to conclude that the ratio of the variance of the completion times for projects for the teams of each of the employees is different?

Problem 3 (30 points)

I want you to run a statistical test for the difference between the means of some population parameters that you are most interested in! For example, suppose you are interested in daily fantasy sports, and you want to know if quarterbacks of the visiting teams have on average more fantasy points than quarterbacks on home teams. You should download real data and make a statistical test!

Be creative! Are you interested in cooking? Are you interested in sports? Are you interested in the stock market and in investing? Find an interesting question to ask from data that you can access. You will be graded on the precision of the question you are seeking to find an answer to, the validity of the hypotheses, and the correctness of the statistical test you employ.