* Please read the instructions carefully.
* For all hypothesis tests use a level of significance α=5%, unless otherwise indicated.
* Print and sign your name and date on the front page.
* Resources: You may consult the course web site, your notes, and the course textbook. You may also use any publicly available material you wish, including books, the internet, etc. But you may NOT submit questions to internet discussion groups.
* Non-collaboration: **NO** collaboration is permitted on this exam. It is trusted that you will not discuss this exam or related course material with any other person (classmate, TA, etc.). You **must abide by** the **Brown University Academic Code** concerning examinations, quizzes, and tests (see [http://goo.gl/mQtfSa).](http://goo.gl/mQtfSa))
* Questions: You may consult **only** with the PHP-1510-2510 Instructor (Dr. Sullivan) for clarifications. Please direct message the professor in slack with your questions. The instructor may also post clarifications on slack. No help for code or data analysis will be given.
* Submission: Print this instruction sheet, sign your name and submission date, scan the signed page and attach it as the front page of your submission. Submit the complete package electronically through Canvas.
  + This project is **due** on Thursday, December 19, 2019, before 11:59pm.
  + **NO** late submission will be accepted.

Name (print)

Signature

Date



**STUDY DESCRIPTION**

In healthcare environment there is a widely recognized need for lowering hospitalization costs while maintaining high quality of health services. One of the key factors contributing to increased hospitalization costs is the length of stay (“los”). It is believed that some insurance companies have been more successful than others at minimizing hospital length of stay.

A study has been conducted to evaluate differences between two of the largest insurance companies in the US. The data for this particular case study comprise information abstracted from the hospital medical records of pneumonia patients (children and adolescents) between the ages of 2 and 18 years old.

A sample of 392 pneumonia cases insured from Insurer A, and 395 pneumonia cases insured from Insurer B, was randomly selected from 29 metropolitan hospitals in the US. The dataset “***hospital.csv***” includes information abstracted from each patient’s medical record regarding patient and hospital characteristics ([**Table 1**](#_bookmark0)).

The primary research question of this study is to test whether one of the two insurance companies has been more successful at minimizing the length of hospitalization (and hence the respective healthcare cost), taking into account other important factors that may affect this comparison.

|  |  |
| --- | --- |
| **Table 1. Information from the study included in the “hospital.csv” dataset.** | |
| **Variable name** | **Description** |
| **Patient characteristics:** | |
| “id” | Patient identification number. |
| “age” | Age (in years). |
| “sex” | Sex (boy or girl). |
| “race” | White, Hispanic, African-American, Asian, or other. |
| “los” | Length of stay in hospital (in days). |
| “complic” | If there were any treatment complication. |
| **Hospital characteristics:** | |
| “hosp.id” | Hospital identification number. |
| “beds” | Total number of beds. |
| “type” | The hospital owner (public, or private). |
| “company” | Insurance company (Insurer A or Insurer B) |

# INSTRUCTIONS:

1. Perform exploratory data analysis (EDA) to describe all the information (variables) collected in this study. [Note: A complete EDA should include both suitable descriptive statistics, and plots.]
2. Conduct statistical analyses to identify significant differences between the two insurance companies regarding both patient- and hospital-related characteristics.
3. Choose the most appropriate statistical approach to answer the primary research question.
4. Summarize your findings from parts 1 and 2 in **one** well-organized (meaningful and easy to understand structure of columns and rows), self-explanatory (properly labeled, etc.) table.
5. Create **one** table to summarize the results from the analysis you conducted in part 3 to answer the primary research question of this study. Again, make sure that this table is well-organized, self-explanatory.

# SUBMISSION:

1. **Final Project Write-up:**

Prepare a report summarizing all the analysis and conclusions about this project. The write-up should:

* 1. be in the form of a pdf or a word document.
  2. not exceed three pages.
  3. approximate what might be submitted as a manuscript for peer review to a biomedical journal.
  4. be organized in four main sections, namely the “Introduction”, “Methods”, “Results”, and “Discussion”.
     1. Introduction: Description of the study and the primary objective (research) question.
     2. Methods: Description of the statistical analyses that you have performed, proceeding from more general and exploratory to more complex techniques.
     3. Results: Overall description of the sample. Summarizing results from EDA and other statistical analyses performed for the purposes of this project.
     4. Discussion: Summarizing main findings from your analysis with emphasis on the research questions. Comment on any weaknesses, or limitations.

# Appendix:

Include **all** tables, plots and any further information that you deem useful for describing your results in an Appendix. Make sure to properly label each table and plot, so they are self-explanatory and easy to read. Submit this appendix as a pdf or word document.

# R-code:

Prepare a text file with all the R-code you used for your analysis. Please make sure the code is properly annotated so as to help the reader understand which part of the code corresponds to which part of the data analysis.

# GRADING:

1. The final project will be evaluated based on:
   1. correctness of the content and approaches used for answering the research questions.
   2. completeness. All research questions must be addressed. Final submission should include **three** files (write-up, appendix, and R-code). Failure to submit **all** these three files will result in a **5-points** penalty.
   3. neatness. The **three** final outcomes of this project (write-up, appendix, and R-code) should follow the specific formats stipulated in the instructions. Severe readability issues will result in the **deduction of additional points** from the final grade. You will **NOT** receive full credit by directly copying and pasting any output from the software.