

STAT 341: Assignment 3 - Spring 2020

Name

62 Marks, Due: Friday, July 10 at 10:00am

NOTES

Your assignment must be submitted by the due date listed at the top of this document, and it must be submitted electronically in .pdf format via Crowdmark/LEARN. This means that your responses for different questions should be in separate .pdf files. Your .pdf solution files must have been generated by R Markdown unless otherwise specified. Additionally:

- For mathematical questions: your solutions must be produced by LaTeX (from within R Markdown). Handwritten and scanned/photographed solutions will not be accepted and you will receive zero points.
- For computational questions: R code should always be included in your solution (via code chunks in R Markdown). If code is required and you provide none, you will receive zero points.
 - **Exception** any functions used in the notes or function glossary can be loaded using `echo=FALSE` but any other code chunks should have `echo=TRUE`. e.g. the code chunk loading `gradientDescent` can use `echo=FALSE` but chunks that call `gradientDescent` should have `echo=TRUE`.
- For interpretation question: plain text (within R Markdown) is fine.

Organization and comprehensibility is part of a full solution. Consequently, points will be deducted for solutions that are not organized and incomprehensible.

- You will submit your solutions in the form of one pdf file per question through LEARN. For example, for Q1 you should submit one pdf file containing the solution to the first question only. Failing to follow the formatting instructions may result in your whole paper or individual questions receiving a grade of 0%.

Question 1 - 28 Marks - Estimating Totals

For this question you will need the Titanic data which can be found in the `carData` package. Here we will focus on the female passengers.

```
library(carData)
data(TitanicSurvival)
Titanic = na.omit(TitanicSurvival)
```

```
Titanic = Titanic[Titanic$sex == "female", ]
Titanic$survived1 = as.numeric(Titanic$survived == "yes")
head(Titanic)
```

```
##               survived    sex age passengerClass survived1
## Allen, Miss. Elisabeth Walton      yes female   29         1st         1
## Allison, Miss. Helen Loraine       no female    2         1st         0
## Allison, Mrs. Hudson J C (Bessi    no female   25         1st         0
## Andrews, Miss. Kornelia Theodos   yes female   63         1st         1
## Appleton, Mrs. Edward Dale (Cha    yes female   53         1st         1
## Astor, Mrs. John Jacob (Madelei    yes female   18         1st         1
```

Use the sample below from the Titanic data to answer the following questions.

```
# set.seed(341) TitanicSample = sample(388, 25) TitanicSample

TitanicSample <- c(57, 349, 147, 21, 310, 293, 62, 96, 304, 31, 258, 5,
  381, 295, 180, 354, 346, 245, 88, 1, 383, 188, 333, 216, 351)
```

to answer the questions below.

- [4 marks] Calculate the Horvitz-Thompson estimate of the average age and provide the standard error of the estimate.
- [4 marks] Calculate the Horvitz-Thompson estimate of the proportion of age less than or equal to 25 and provide the standard error of the estimate.
- [4 marks] Calculate the Horvitz-Thompson estimate of the proportion of age less than or equal to 45 and provide the standard error of the estimate.
- [4 marks] In two separate graphs, plot the Horvitz-Thompson estimate of the cumulative distribution function of age and the standard error of the estimate.
 - Note** Similar to the cdf, the standard error is also a function of age.
- [4 marks] Plot the Horvitz-Thompson estimate of the cdf of age, and overlay the lines of ± 2 times the standard error.
- [4 marks] Calculate the Horvitz-Thompson estimate of the proportion of females who survived and provide the standard error of the estimate.
- [4 marks] Calculate the Horvitz-Thompson estimate of the proportion of females in 1st class and provide the standard error of the estimate.

Question 2 - 16 Marks - Comparing Estimators

Using a population, compare two (or more) estimators of some population attribute. Your answer should include:

- description of the population and the context of the problem,
- the effect of changes in some parameters such as the sample size,
- numerical and graphical comparison and their descriptions,
- a conclusion.

- You answer should be limited to 1 to 3 pages.

Your solution should be **in your own words**, but as motivating examples, see from the Samples exercises:

- 1.6 Comparing the standard deviation and the Median Absolute Deviation,
- 1.7 Comparing two sampling designs,
- 1.10 Comparing measures of dispersion,
- 1.13 Comparing the Mean and Median and
- 1.17 Estimating the median for Radar Data.

Rubric

| Criteria | Descriptor | Marks |
|-----------------------|---|-------|
| Population/Estimators | Description and Difficulty | /4 |
| Format | Clarity, Organization and LaTeX | /4 |
| Comparison | Description, Results and Graphic | /4 |
| Discussion/Summary | Justification and Relevant Terminology used | /4 |

Question 3 - 8 Marks - Test Question

Using the material from Sections 3.1, 3.2, 4.1 and 4.2, construct one exam question worth approximately 2 marks. Notes;

- There is a one page limit.
- Provide a full solution with mark allocations.
- The question can be about a single topic or two disjoint topics.
- **Handwritten** questions and answers are acceptable, but it must be legible.
- TRUE/FALSE or multiple choice questions are not allowed.
- Note: Do not expect full marks for reproducing examples or questions from the exercises, assignments, lecture notes or webpages.

However, you may copy and paste from the notes the definitions, or generic R codes.

Rubric

| Criteria | Descriptor | Marks |
|----------|--|-------|
| Question | Concept, Difficulty, Clarity and Creativity | /4 |
| Solution | Explanation, Correct Justification and Clarity | /4 |

Question 4 - 10 Marks - Summarize

In your own words summarize the subsection **4.2.1-Anatomy_of_a_Significance_level**

- You are limited to 1 to 2 pages.
- You are recommended to use a combination of formulas, full sentences an example.

Rubric

| Criteria | Descriptor | Marks |
|----------|--|-------|
| Format | Organization | /3 |
| Writing | Clarity & Grammar | /2 |
| Content | Coverage, Depth, Relevant Terminology used and Example | /5 |