

# Lab 5

Group Member Names - here

## Lab Overview

All students attending class in the group can turn in a single document with each participants name. Students not attending class will need to complete their own lab.

Being a huge fan of the movie *Titanic*, you have decided to put your newfound statistical sampling skills to a test. You are curious about the survival rate of passengers on the Titanic and based on your movie viewing think that controlling for different strata (e.g. the Jack stratum (Leo DiCaprio) and the Rose stratum (Kate Winslet)) might be a good idea.

### 1. (5 points)

You have the ability to take 150 samples of passengers on the Titanic to determine whether they survived the shipwreck. Assume the cost of sampling each passenger is the same. Using the information below pertaining to six strata, allocate the 150 samples across the strata to minimize the variance. You do not know the population variance for each stratum, but you have found estimates of  $p_h$  from another study. Calculate how to allocate the 150 samples across the strata and fill in the last column in the table below.

| Stratum          | $N_h$ | $\hat{p}$ | $n_h$ |
|------------------|-------|-----------|-------|
| 1st Class Male   | 122   | 0.4       |       |
| 2nd Class Male   | 108   | 0.2       |       |
| 3rd Class Male   | 347   | 0.1       |       |
| 1st Class Female | 94    | 0.95      |       |
| 2nd Class Female | 76    | 0.90      |       |
| 3rd Class Female | 144   | 0.5       |       |

### 2. (5 points)

Using your allocation from above, take a stratified random sample from the dataset `titanic.csv`, and compute a point estimate for the survival probability of the entire population,  $\hat{p}_{str}$ . (Include your R code or hand written computations).

```
library(readr)
library(dplyr)
titanic <- read_csv('http://math.montana.edu/ahoegh/teaching/stat446/titanic.csv')
```

```
## Parsed with column specification:
## cols(
##   PassengerId = col_double(),
##   Survived = col_double(),
##   Pclass = col_double(),
##   Name = col_character(),
##   Sex = col_character(),
##   Age = col_double(),
##   SibSp = col_double(),
##   Parch = col_double(),
##   Ticket = col_character(),
##   Fare = col_double(),
```

```
## Cabin = col_character(),
## Embarked = col_character()
## )

titanic_filtered <- titanic %>% select(Pclass, Sex, Survived)

first_male <- titanic_filtered %>% filter(Sex == 'male' & Pclass == 1)
second_male <- titanic_filtered %>% filter(Sex == 'male' & Pclass == 2)
third_male <- titanic_filtered %>% filter(Sex == 'male' & Pclass == 3)
first_female <- titanic_filtered %>% filter(Sex == 'female' & Pclass == 1)
second_female <- titanic_filtered %>% filter(Sex == 'female' & Pclass == 2)
third_female <- titanic_filtered %>% filter(Sex == 'female' & Pclass == 3)
```

### 3. (5 points)

Compute a confidence interval for  $\hat{p}_{str}$ .

### 3. (4 points)

You remember crying after Jack dies in the movie, but given the different survival rates by stratum, does the ending of the movie seem to be an accurate historical reflection?