Practice Exam

Name:_____

Directions

- You **do not** need to fill all the space provided. In many cases, large amounts of white space only exist to correct for general spacing of the exam
- You **do not** need to write in complete sentences for all questions: if I desire you to answer in a complete sentence I will indicate this
- The practice exam is shorter than the actual exam
- The actual exam will have at least one question about linear regression with a few sub-questions
- Have fun (please)

Question 1

Answer the following questions in 1-2 sentences

Part A Briefly describe the *statistical framework*. In particular, what is a *parameter*, what is a *statistic*, and how are these two related to one another?

Part B What two things do a *distribution* tell us about a variable?

Part C Explain why most times it is not practical to conduct a census.

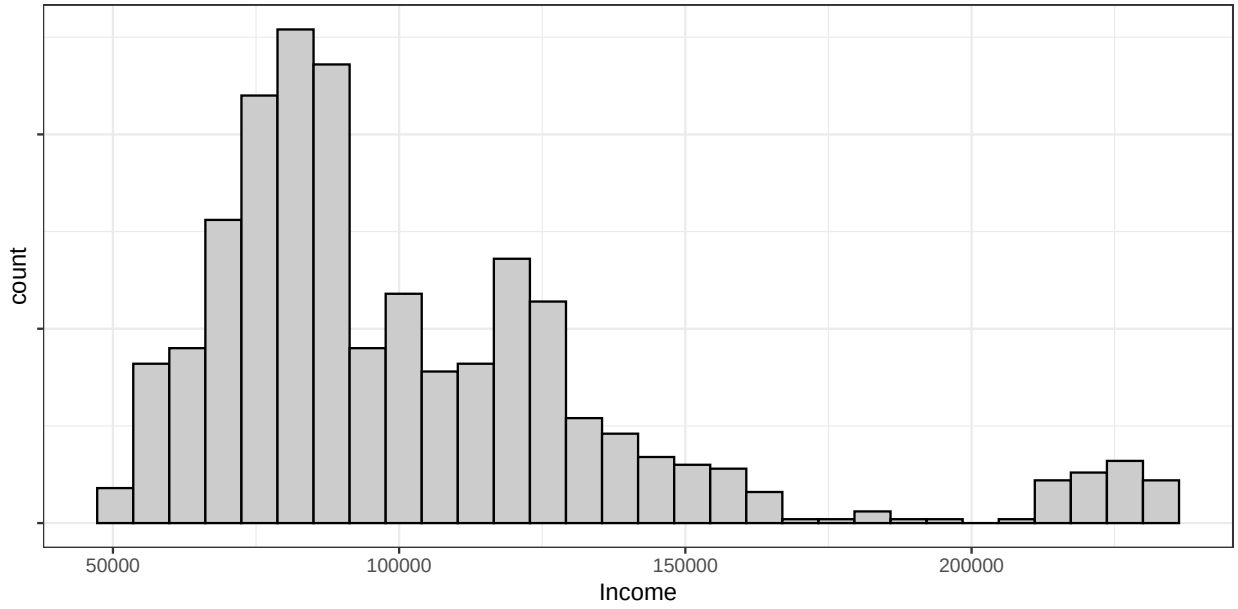
Part D Explain the difference between random sampling and random assignment. What type of claim does using random assignment allow us to make?

Part E Explain what makes an experiment different from an observational study.

Part F: Explain *why* random assignment allows us to make causal claims.

Question 2

A survey of incomes for 1000 individuals with graduate degrees in business was conducted five years following graduation, the distribution of which is demonstrated in the box plot below.




```
# 5 number summary and mean
df %>% pull(x) %>% summary
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   51662   76283   89047  102449  120362  234359
```

```
# std.dev
df %>% pull(x) %>% sd()
```

```
## [1] 38490.08
```

Part A: Use the histogram to describe this variable in a complete sentence as we learned in class. Make sure to include context.

Part B: Explain the concept of *robustness* as it applies to using statistics for describing a variable.

Question 3

The table below presents the results from the 2006 General Social Survey conducted by the National Opinion Research Center polling 1,009 respondents on their age and level of job satisfaction. Row totals have been added in the last column.

For the questions below, you do not need complete sentences but you must show the calculations used to derive your answer

Age	Job Satisfaction			Total
	Low	Medium	High	Sum
<30	34	53	88	175
30-50	80	174	304	558
>50	29	75	172	276

Part A: What percentage of all respondents were between 30-50 years old and indicated that they had *low* job satisfaction?

Part B: Of those responsdents who were between 30-50 years old, what proportion indicated they had *low* job satisfaction.

Part C: Of those responsdents who had *low* job satisfaction, what proportion were between 30-50 years old?

Part D: Which age demographic had the highest proportion of respondents indicate that they had *high* job satisfaction?

Part E: Sketch a graphic/chart to display your results for Part D.

Question 4

A study is designed to test the effect of light level and noise level on exam performance of college students. The light treatments considered are fluorescent overhead lighting, yellow overhead lighting, no overhead lighting (only desk lamps). The noise treatments considered are no noise, construction noise, and human chatter noise.

The researcher contacted a local college and 200 students volunteered to participate in the study. These students were randomly assigned to the treatments.

Part A: Is this an experiment or an observational study? Explain your answer.

Part B: What is an experimental unit in this study?

Part C: How many factors are considered in this study? Identify them, and describe their levels.

Part D: Does this study employ random sampling? Explain your answer.

Part E: Are the results of this study generalizable to all college students? Explain your answer.

Part F: Can we draw causal conclusions from the results of this study? Explain your answer.