1 Chapter 1

1.1 Inference Rules

We are given a fancy little chart for these rules in the annotated questions:

		Random Sample?		
		Yes	No	
Random Allocation?	Yes	Population - Yes Causal - Yes	Population - No Causal - Yes	Randomized Experiment	A
	No	Population - Yes Causal - No	Population - No Causal - No	Observational Study

random sample is where the study units that we chose from the population are chosen randomly. A random allocation is when we have random assignment to the treatment groups of our study. So in question 1 we see that there is a random sample taken from the participants of the study, but there is no random allocation of the participants to the study group.

1.2 Criteria for an Experimental Study

- Must manipulate the explanatory variable
- Must have random sampling
- Must control the experiment environment

1.3 Criteria for Population Inferences

- We can make population inferences on the population studied if we can obtain a random sample of participants from the study

1.4 Criteria for Causal Inferences

- We can make causal inferences if the original study groups were obtained through random sample

2 Chapter 2

2.1 T-Test Procedure

Process for performing a hypothesis test using the t-distribution:

1. State the null and alternative hypotheses:
   - H0: There is no effect or difference between the population means.
   - Ha: There is an effect or difference between the population means.

2. Determine the significance level ():
   - Common levels: 0.05, 0.01.

3. Calculate the t-statistic:

- t = (sample mean - population mean) / (standard deviation of the sample / sqrt(sample size))

4. Determine the degrees of freedom (df):
   - df = sample size - 1.

5. Look up the critical value in a t-table:
   - Use df and to find the critical value.

6. Make a decision:
   - If t is greater than the critical value, reject H0.
   - If t is less than the critical value, fail to reject H0.

7. Draw a conclusion:
   - Based on the decision, either reject or fail to reject H0, and state conclusion in terms of problem being studied.

3 Appendix A - Questions and Answers

3.1 Chapter 1

Preamble 1. *In Orangetown in 2014, middle-aged people between the ages of 40-50 years old joined two different groups, an exercise group and a group that watched movies together. The participants had made their own decision about which group to join. In 2019, researchers selected random samples from each of these two groups to investigate any possible effects of activity group (exercise group versus movie group) on their mental and physical health. They took records of their systolic blood pressure (SBP) readings (in mm Hg), their Body Mass Index (BMI), that is body weight/(height)2 (in kg/cm2), and their feelings of wellbeing (on a scale of 1-5).*

Question 1.   *a Identify the study units*

   b What is/are the explanatory variable(s) or factor(s)? State the type of data scale used (e.g., categorical, ordinal, etc.) for recording each explanatory variable.

   c What is/are the response variable(s)? State the type of data scale used (e.g., categorical, ordinal, etc.) for recording each response variable.

   d Give details of where the study was conducted.

   e Give details of regarding temporal aspects of the study.

   f Is this an observational or experimental study? Very briefly explain your answer.

   g Based on this study, will it be possible to make population inferences? Briefly explain your answer.

   h Based on this study, will it be possible to make causal inferences? Briefly explain your answer.

Answer 1.   *a The study units are the atomic entities being studied, so in this case it is the middle-aged people between ages 40-50 in both the movie group and the exercise group.*

b The explanatory variable is the categorical value of the group that the individual was a part of. This is a categorical variable with two categories.

c The response variables for the data are:

- ***SBP** - continuous, quantitative*
- ***BMI** - continuous, quantitative*
- ***Wellbeing** - ordinal, 5 categories*

d The study was conducted in Orangetown.

e The study was started in 2014 and data collection was performed in 2019.

*f This is an observational study since the researchers **did not**:*

- *Manipulate the explanatory variable*
- *Control the environment variables*
- *There was no random assignment to groups (the participants decided themselves)*

g Since the researchers took a random sample from the two groups, we can make population inferences for the two groups specifically.

h We cannot make causal inferences based on this study since there was no random assignment to these groups.

Preamble 2. *In 2016, a researcher in a health product company in Toronto wanted to test whether a new health drink he had developed is effective in improving the general health of their customers. He sent emails to all their customers asking for volunteers to participate in the study. It took one month to organize the study. A total of 100 customers wanted to participate. From this group, he asked for volunteers who wanted to join the treatment group and consume the new health drink daily, with 65 customers responding that they wanted to try the drink. The remaining 35 customers formed the control group and were not given the health drink. After six months, the researcher measured heartbeat rates (in beats per minute), cholesterol levels (in milligrams per deciliter), and fitness level (on a scale of 1-15, based on performance in several exercise tests) of all participants.*

Question 2. *Identify the study units*

Answer 2. *The study units in this case are the 100 voluntary customers that chose to participate in the study.*

Question 3. *What is the population of interest?*

Answer 3. *The population of interest is the customers of the health drink.*

Question 4. *What is/are the explanatory variables? What types of data are used?*

Answer 4. *The explanatory variable is a whether or not the user consumes the health drink. This is a categorical variable with two categories.*

Question 5. *What is/are the responding variables? What type of data is used?*

Answer 5. *The responding variables are:*

- **Heartrate**: *Discrete numerical value*

- **Cholesterol**: *Continuous numerical value*

- **Fitness Level**: *Ordinal*

Question 6. *Give details of where and when the study was conducted*

Answer 6. *The study was conducted in Toronto.*
The study began in 2017, with data collection occuring 6 months later.

Question 7. *Was this an observational or experimental study? Justify.*

Answer 7. *This was an observational study since the researcher did not modify the explanatory variable, did not control the environment variables and did not have random sampling.*

Question 8. *Based on this study, will it be possible to make either population inferences or causal inferences or both? Briefly explain your answer.*

Answer 8. *We are not able to make population inferences on the 100 individuals studied since there is no random selection in the data collection phase.*
We will not be able to make causal inferences based on the study since there was no random assignment to the groups.

Question 9. *Based on the description of the study given above, what are the possible weaknesses of this study. There are at least six weaknesses.*

Answer 9. *1. The groups are not of equal size*

2. The study will have a bias towards people who already have a healthy lifestyle

3.

Question 10. *According to a 2002 article in a national newspaper, female students attending private schools in Canada scored significantly better on a standardized aptitude test than males. The results were from a random sample of female and male students attending private schools across Canada. Can we make population inferences? Can we make causal inferences?*

Answer 10. *We can make population inferences for the population of males and females attending Canadian private schools, but we cannot make further generalizations.*
We cannot make causal inferences because assignment to the groups is not random.

Question 11. *An experiment was conducted to explore the effects of two teaching styles for Alberta students in grade 12 mathematics. The individuals in the experiment were a random sample of grade 12 math students from a particular high school in southern Alberta. The students were randomly assigned to the two treatment groups, each group taught by the same teacher. After a three-week course, all students took an identical standardized exam. The response was the students score on the exam. Can we make population inferences? Can we make causal inferences?*

3.2 Chapter 2

3.3 Chapter 3