

# Homework Assignment 1

Helinda He

October 14, 2022

---

## Question 1 (Exercise 1.2 on Page 11):

Write your answer for question 1 here. Since the question has four parts, use:

a)

78%

b)

80%

c)

the control group has higher percentage of patients have a good experience in the treatment.

d)

the difference between the percentage is very small which means that people cannot conclude that the treatment is useful or not.

## Question 2 (Exercise 1.3 on Page 19):

- 1) will the air pollution influence length of gestation date.
- 2) the pregnant women between 1989 and 1993 the number of the subject is 143,196
- 3) Length of gestation - numerical, continuous levels of carbon monoxide - numerical, continuous levels of nitrogen dioxide - numerical, continuous levels of ozone - numerical, continuous

## Question 3 (Exercise 1.19 on Page 30):

- 1) observational studies
- 2) I think we can apply stratified sampling method since there are four groups in the situation

**Question 4 (Exercise 1.21 on Page 30):**

- 1) the relationship between life expectancy and percentage of internet users has a strong relationship. They are positively associated.
- 2) observational study
- 3) the health system can be a great factor that can influence the result of estimated life expectancy. If a country has a good health system, the estimated life expectancy will be longer. Otherwise, it can be shorter if a country has a poor health system.

**Question 5 (Exercise 1.27 on Page 31):**

- 1) Simple Random Sample - some people might forget to bring the survey back and they might not willing to tell others their performance at school
- 2) Convenience Sample - his friend cannot represent the whole population of the group
- 3) Convenience Sample - this method cannot represent the whole population and people who does not take the class might complete the survey which will influence the result of the survey
- 4) Multistage Sample - maybe some of the 5 classes have a bad instructor of certain course which will somehow influence the performance at school.

**Question 6 (Exercise 1.38 on Page 36):**

- 1) simple random sampling pro: the sample can have a better representation of the population. cons: sometimes it is difficult to collect data since they are living in different neighborhoods.
- 2) stratified sampling pro: the sample can have a good representation of the population cons: it might be hard to divide the whole population in to 20 groups
- 3) cluster sampling pro: it is more convenient in collecting data compare to simple random sampling and stratified cons: it cannot be a good representation of the population
- 4) multistage sampling pro: it is easier to collect data cons: it will have bias if there is a neighborhood has worse environment than others.
- 5) convenience sampling pro: very easy to collect data cons: it will have a lot bias since they are at the same location which is near to the city council offices.

**Question 7 (Exercise 1.42 on Page 38):**

a)

observational study

b)

the explanatory variables will be the screen time, child's sex and age and on mother's education, ethnicity, psychological distress, and employment.

c)

the response variables will be psychological well-being score.

d)

I do not think that the result of the study can represent the whole population because the study did not use the proper sampling method, like simple random sampling, which might have bias.

e)

I do not think that the result can be used to establish some relationship between the explanatory variables and response variables because there are too many of the explanatory variables and people cannot tell which of them cause the result of the psychological well-being score.

### **Question 8 (Exercise 1.43 on Page 38):**

a)

County, state, driver's race, number of stops per year, percentage of stopped cars searched, and percentage of stopped drivers arrested.

b)

County - categorical, regular categorical State - categorical, regular categorical Driver's race - categorical, regular categorical Number of stops per year - numerical, discrete Percentage of stopped cars searched - numerical, continuous Percentage of stopped drivers arrested - numerical, continuous

c)

In this situation, the explanatory variable will be the driver's race and the response variable will be the vehicle search rates.

### **Question 9:**

- 1) Sampling is to find a proper group of representation of the population to collect the data in the study.
- 2) Sampling is important because it can help to save many time and money for the study. A good sampling can be a good representation of the whole population.
- 3)
  - Simple Random Sampling
  - Systematic Sampling
  - Stratified Sampling
  - Cluster Sampling
  - Multistage Sampling
  - Convenience Sampling

**Question 10:**

- 1) Dependent variable is changing based on the other variables. Independent variable is variables that are not associated.
- 2) Observational study is based on the observation and measurements. Experimental study is based on the experiment which researchers set up experimental and control group and then observe characteristics and take measurements.
- 3) Stratified sampling is dividing the population in to groups and randomly select certain amount population from each group. On the other hand, the cluster sampling is dividing the population into groups and randomly select certain amount groups.