ANSWERS PTO →

MA 2611-DL01 Applied Statistics I
D-Term Spring 2023 - Homework 01

Due: R 03/23 by 11.59 p.m.

I. Show all work as described in class. Partial credits will be given. Submit your solutions in a pdf format to canvas. Please write your name.

II. For the questions that are based on R programming, please copy, and paste the R plots and codes into a word file and then submit or submit the R markdown file.

1. Write the sample unit, target population, sample frame, and sample design for the following examples:
   (a) A cable television company wants to measure the satisfaction of its customers. To do so, it mails a questionnaire to 800 selected households. Assume that all 800 households return the questionnaire. For this survey, we will focus on the question, Overall, are you satisfied with the cable service we provide?"

   (b) An auto analyst is conducting a satisfaction survey, sampling from a list of 10,000 new car buyers. The list includes 2,500 Ford buyers, 2,500 GM buyers, 2,500 Honda buyers, and 2,500 Toyota buyers. The analyst selects a sample of 400 car buyers, by randomly sampling 100 buyers of each brand.

2. A plastics molding factory that must treat its waste before discharge. The average amount of pollutant discharged (lb per day) depends on three factors: the chemical compound added (choose either chemical P or chemical Q); the treatment temperature (72 °F or 100 °F); and the stirring speed (200 rpm or 400 rpm). Determine the experimental unit, response, factor(s), factor levels, at least three nuisance factors, and treatments for this study.

3. Imagine that you are asked to determine students' opinions at your school about a potential change in library hours. Describe how you could go about getting a sample of each of the following types: random sample, sample of convenience, stratified sample, single stage cluster sample, and systematic sample.

1). SAMPLE UNIT, TARGET POPULATION, SAMPLE FRAME, SAMPLE DESIGN.

a). SAMPLE UNIT : HOUSE HOLDS

TARGET POPULATION : CUSTOMERS OF THE TELEVISION COMPANY

SAMPLE FRAME : A SELECTION OF 500 HOUSEHOLDS FROM THE LIST OF CUSTOMERS

SAMPLE DESIGN : SIMPLE RANDOM SAMPLING

b). SAMPLE UNIT : CAR BUYERS

TARGET POPULATION : NEW CAR BUYERS

SAMPLE FRAME : LIST OF 10,000 NEW CAR BUYERS (2,500 CARS FROM 4 DIFFERENT COMPANIES

SAMPLE DESIGN : STRATIFIED SAMPLING (400 NEW CAR BUYERS ; RANDOM SAMPLING 100 BUYERS OF EACH BRAND)

2). EXPERIMENTAL UNIT : * A batch of waste from the yoghurt modelling factory

RESPONSE : * Average amount of pollutant discharged

FACTORS : * Chemical component P (or) Q ; * Treatment temperature 72°F (or) 100°F ; * Stirring Speed 200 RPM (or) 400 RPM

FACTOR LEVEL : * Amount of chemical component ; * Treatment temperature ; * Speed of stirring

NUISANCE FACTORS : * Batch size ; * Time the batch was processed ; * Weather outside when batch was processed.

TREATMENTS : There are 8 treatments in this study ; By combining the level of 3 factors.

3). RANDOM SAMPLE : From the student ID numbers we have ; we can use a random number generator to select a particular number of students according to the sample size set to give the students a fair opportunity.

SAMPLE OF CONVENIENCE : Ask for this research ; we can interview and collect data from the students who are present at the library ; but could potentially lead to a biased outcome.

STRATIFIED SAMPLING : From the data we have, we can category the students based on what year they are in and what major they take and their gender ; and the sample could be selected from each group proportionally to create fair sample set as a whole, but more importantly to get fair results.

SINGLE STAGE CLUSTER : Randomly selecting a group of students from a big population. In this case randomly selecting a few classes from each grade level or department

SYSTEMATIC SAMPLING : From the list of students, selecting every n^{th} student until the desired sample size is considered to be systematic sampling. Therefore selecting every 5th or 5th student from the list enables us to get the data we need.

4). From the following scenario, we can determine this portrays an example of a selection bias. This occurred when the sample selected for the study is not a representative of the population being studied, leading to a biased estimate of the population parameter. The interviews are using a systematic sampling to collect the data about the use of the IV drug. Therefore, the results obtained from this study may not accurately reflect the incidence of IV drug use in the entire city. For example, individuals who are not selected may have different characteristics from those who are selected, leading to an inaccurate estimate.

5). a). Stratified Random Sampling (SRS)

b). Simple Random Sampling.

c). Stratified Random Sampling (SRS)

d). Cluster Sampling

6). a). NUMERICAL & DISCRETE ; RATIO SCALE

b). NUMERICAL & CONTINUOUS ; RATIO SCALE

c). NUMERICAL & DISCRETE ; RATIO SCALE

d). NUMERICAL CONTINUOUS ; RATIO SCALE

e). CATEGORICAL ; NOMINAL SCALE

f). CATEGORICAL ; NOMINAL SCALE

g). NUMERICAL & DISCRETE ; RATIO SCALE

too general the ordered here here selected with bias

4. The following is an example of one or more selection bias, nonresponse bias or response bias. Which one(s) is it an example of? Defend your choice(s).

   "In a study to ascertain the incidence of IV drug use, interviewers at several shopping malls in a certain city select every tenth shopper and ask them if they are an IV drug user" ??

5. For the following examples, write down the sampling method used in each situation.

   (a) A large company surveys 100 employees by taking random samples of 10 managers and 90 non-managerial employees. SRS

   (b) Each student at a school has a student identification number. Counselors have a computer that generates 50 random identification numbers, and the students associated with those numbers are asked to take a survey. RANDOM

   (c) A student council surveys 100 students by taking random samples of 25 freshmen, 25 sophomores, 25 juniors, and 25 seniors. SRS

   (d) An airline company wants to survey its customers one day, so they randomly select 5 flights that day and survey every passenger on those flights. RANDOM

6. For each of the following variables, determine whether the variable is categorical or numerical and determine its measurement scale. If the variable is numerical determine whether the variable is discrete or continuous.

   (a) Number of cellphones in the household. Numerical, Discrete

   (b) Monthly data usage (in MB). Numerical, Continuous

   (c) Number of text messages exchanged per month. Numerical, Discrete

   (d) Voice usage per month (in minutes). Numerical, Continuous

   (e) Whether the cellphone is used for email. Categorical

   (f) Academic major. Categorical

   (g) Number of textbooks purchased. Numerical, Discrete

7. **Use R** when ever it is relevant. The following data set shows the mid-term grade of a Biology class of 30 students.

78	45	67	56	88	26	89	91	75	80
58	68	77	82	79	80	93	46	55	62
78	86	89	78	66	78	81	63	90	83

(a) Make a frequency distribution for the data. Use a class width of 10.

(b) Draw a histogram for the data. What can you say about the shape of the distribution?

(c) Complete the table in categorizing data as below (note that interval ranges are different from part (a)).

Range	Grade	Frequency
85-100	A	
75-84	B	
65-74	C	
50-64	D	
0-49	F	

(d) Draw a pie chart to display the distribution of grades in part (c).

```
# Author Udaya Vijay Anand
```

Q7

```
# input the data
grades <- c(78, 45, 67, 56, 88, 26, 89, 91, 75, 80,
            58, 68, 77, 82, 79, 80, 93, 46, 55, 62,
            78, 86, 89, 78, 66, 78, 81, 63, 90, 83)
```

#(a) Make a frequency distribution for the data. Use a class width of 10.

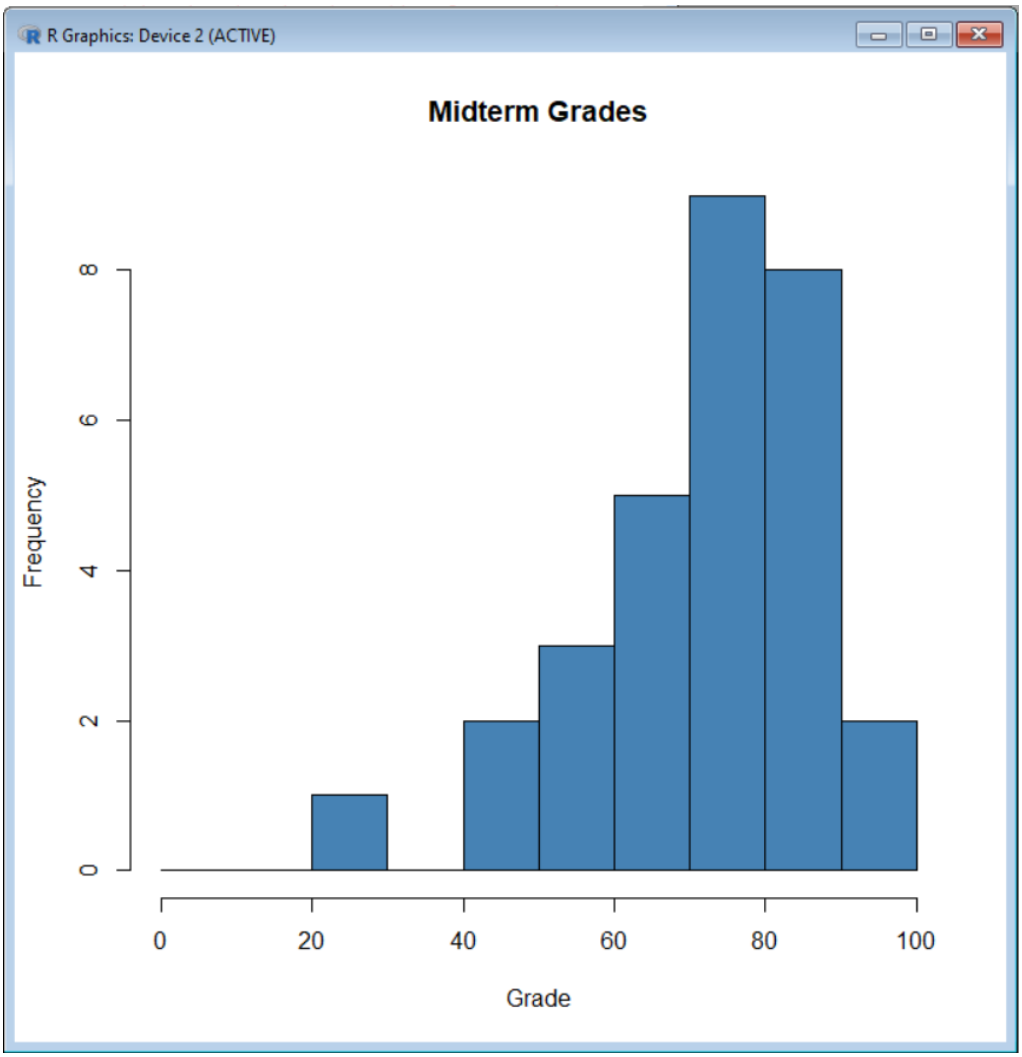
```
# create frequency distribution
freq_table <- table(cut(grades, breaks=seq(0, 100, by=10),
right=FALSE))
freq_table
```

```
> freq_table
grade_ranges
  [0,50)  [50,65)  [65,75)  [75,85) [85,100)
       3        5        3       12        7
```

```
#-----------------------------------------------------------------------
```

 # (b) Draw a histogram for the data. What can you say about the shape of the distribution?

```
# plot histogram
hist(grades, breaks=seq(0, 100, by=10), col="steelblue", main="Midterm
Grades", xlab="Grade")
```



#----------

 # (c) Complete the table in categorizing data as below (note that interval ranges are different from part (a)).

```
# create grade ranges
grade_ranges <- cut(grades, breaks=c(0, 50, 65, 75, 85, 100),
right=FALSE)

# create frequency table
freq_table <- table(grade_ranges)
freq_table
```

```
> freq_table
grade_ranges
  [0,50)   [50,65)   [65,75)   [75,85)  [85,100)
       3         5         3        12         7
```

```
#-----------------------------------------------------------------------
```

#(d) Draw a pie chart to display the distribution of grades in part (c).

```
# create pie chart

pie(freq_table, main="Distribution of Grades",
labels=names(freq_table), col=rainbow(length(freq_table)))
```