Name (Last, First):	Student ID:
Lopez Sepulveda, Kevin	U5972827

Assignment 8

METCS544A3A4_F2024

Instructions:

4. For answering programming questions, please use Adobe Acrobat to edit the pdf file in two steps **[See Appendix: Example Question and Answer]**:
    a. Copy and paste your R code as text in the box provided (so that your teaching team can run your code);
    b. Screenshot your R console outputs, save them as a .PNG image file, and paste/insert them in the box provided.
    c. Show all work—credit will not be given for code without showing it in action, including a screenshot of R console outputs.
5. To answer non-programming questions, please type or handwrite your final answers clearly in the boxes. Show all work - credit will not be given for numerical solutions that appear without explanation in the space above the boxes. **You're encouraged to use R to graph/plot the data and produce numerical summaries; please append your code and screenshot of the outputs at the end of your PDF submission.**
6. **[Total 84 pts = 81 pts + 3 Extra Credit pts]**

Grading Rubric

Each question is worth 3 points and will be graded as follows:

3 points: Correct answer with work shown

2 points: Incorrect answer but attempt shows some understanding (work shown)

1 point: Incorrect answer but an attempt was made (work shown), or **correct answer without explanation (work not shown)**

0 points: Left blank or made little to no effort/work not shown

Reflective Journal [3 pts]

(Copy and paste the link to your live Google doc in the box below)

https://drive.google.com/drive/folders/1_8qcBjQVMfZggF42UYJuHQzMoBcyAy0Q?usp=drive_link

Part I. Collecting Data: Scope of Inference (15 pts)

Directions: For the study below, answer the questions that follow. Then, in the graphic organizer, made a check on the correct scope of the study.

Lack of sleep is associated with increased risk of cardiovascular disease, depression, and other health concerns. A new study now shows that lack of sleep also affects social interactions, making people less willing to help others (empathy). Scientists placed 24 healthy volunteers in a functional magnetic resonance imager (fMRI) to scan their brains after eight hours of sleep and then again after a night of no sleep. The order that they received the treatment of sleep vs no sleep was randomly selected with the flip of a coin. They found that areas of the brain that form the theory of mind network, which is engaged when people empathize with others or try to understand other people's wants and needs, were less active after a sleepless night. "Here, we found that a decrease in the quality of someone's sleep from one night to the next predicted a significant decrease in the desire to help other people from one subsequent day to the next." Source: https://www.sciencedaily.com/releases/2022/08/220823143827.htm

Is this an observational study or an experiment?Explain.
This is an experiment because they imposed the condition of no sleep onto the groups and measured the difference

Does this study involve random sampling or random assignment? Explain.
They did random assignment because the condition imposed was assigned by a flip of a coin

Can this study make inferences about the population or only the study participants?Explain.
This can only make inferences about the partipants as it does not say which population or which sample of population

Can this study make inferences about cause and effect or just an association between variables? Explain.
This can make an assosciation between variables but not cause and effect because there is a correlation but we may be missing factors such as socio economic factors

Which box of inference does this study fall under?

	Inference about cause and effect	No Inference about cause and effect
Inference about the Population		
No Inference about the Population	this one	

Part II. Probability Basics (36 pts) [Show your work, not just a number]

Write all probabilities as decimals and round to the nearest three decimal places when needed.

1) You pick 1 marble from a bag that contains 8 marbles (2 blue, 3 red, and 3 yellow). Find the following probabilities.

a) P(Blue)

2/8 or 1/4 or 25 percent or 0.25

b) P(Yellow)

3/8 or 37.5 percent or 0.375

c) P(Green)

3/8 or 37.5 percent or 0.375

d) P(Blue or Yellow)

5/8 or 62.5 percent or 0.625

e) P(not Blue)

6/8 or 3/4 or 75 percent or 0.75

f) P(Red, Blue, or Yellow)

8/8 or 100 percent or 1

2) Below are the probabilities of pulling out a particular color from an M&M bag.

a) What is the probability of pulling a blue M&M?

b) Describe the likelihood of this event.

Brown	Red	Yellow	Green	Orange	Blue
0.30	0.20	0.20	0.10	0.10	?

Answer: a) 0 or 0 perecent

Impossible

3) A survey of 324 people asked what their favorite food was. The results are shown below.

	Pizza	**Burgers**	**Fried Chicken**	**Other**	**Total**
Less than 18	60	23	5	34	122
18 and older	45	33	20	104	202
Total	105	56	25	137	324

If we randomly select a person from this sample,

a) What is the probability that a person likes fried chicken?

25/324 or 0.077

b) What is the probability that a person is less than 18 years old and likes burgers?

23/324 or 0.071

c) What is the probability that a person likes pizza or burgers?

161/324 or 0.497

d) What is the probability that a person is less than 18 or 18 and older?

324/324 or 1

Part III. Statistical Programming (30 pts)

Initialize the scores of 100 students as shown below (**the dataset scores.csv can be found and downloaded from Course_Materials/R_Materials/Datasets/**):

```
scores <- read.csv("scores.csv")
```

a) Show the default histogram of the student scores. Save the result of the histogram into a variable. Using only the **counts** and **breaks** property of this variable, write the R code to produce the following output. The code for the following output should not refer to the individual scores.

```
 3 students in range (35,40]
 4 students in range (40,45]
10 students in range (45,50]
13 students in range (50,55]
17 students in range (55,60]
27 students in range (60,65]
13 students in range (65,70]
 8 students in range (70,75]
 3 students in range (75,80]
 2 students in range (80,85]
```

b) Using the breaks option of the histogram, show the histogram and the custom output as shown below so that students in the range (70,90] get an A grade, (50,70] get a B grade, and (30-50] get a C grade. The code for the following output should not refer to the individual scores.

```
17 students in C grade range (30,50]
70 students in B grade range (50,70]
13 students in A grade range (70,90]
```

Answer: Copy and paste your R code in the box below (not an image but the text).

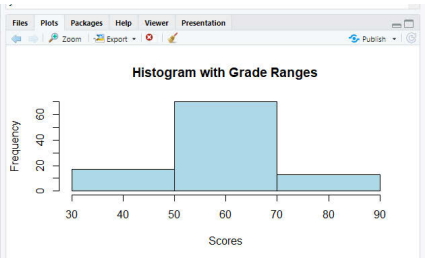
```
# Define custom breaks for the grade ranges
grade_breaks <- c(30, 50, 70, 90)

# Generate the histogram using the custom breaks
grade_hist <- hist(scores$Score, breaks = grade_breaks, right = TRUE,
                   main = "Histogram with Grade Ranges",
                   xlab = "Scores", ylab = "Frequency", col = "lightblue")

# Extract counts and breaks
counts <- grade_hist$counts
breaks <- grade_hist$breaks
cat("Counts per grade range:\n")
for (i in seq_along(counts)) {
  grade <- switch(i,
                  "1" = "C grade",
                  "2" = "B grade",
                  "3" = "A grade")
  cat(paste0(counts[i], " students in ", grade,
             " range (", breaks[i], ", ", breaks[i + 1], "]\n"))
}
```

Screenshot of your R console outputs and paste the image in the box below

```
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
> View(scores)
> hist_result <- hist(scores$Score, main = "Default Histogram of Scores",
+                     xlab = "Scores", ylab = "Frequency")
> counts <- hist_result$counts
> breaks <- hist_result$breaks
> cat("Counts per sub-range:\n")
Counts per sub-range:
> for (i in seq_along(counts)) {
+   cat(paste0(counts[i], " students in range (",
+              breaks[i], ", ", breaks[i + 1], "]\n"))
+ }
3 students in range (35, 40]
4 students in range (40, 45]
10 students in range (45, 50]
13 students in range (50, 55]
17 students in range (55, 60]
27 students in range (60, 65]
13 students in range (65, 70]
8 students in range (70, 75]
3 students in range (75, 80]
```

```
3 students in range (35, 40]
4 students in range (40, 45]
10 students in range (45, 50]
13 students in range (50, 55]
17 students in range (55, 60]
27 students in range (60, 65]
13 students in range (65, 70]
8 students in range (70, 75]
3 students in range (75, 80]
2 students in range (80, 85]
> grade_breaks <- c(30, 50, 70, 90)
> grade_hist <- hist(scores$Score, breaks = grade_breaks, right = TRUE,
+                    main = "Histogram with Grade Ranges",
+                    xlab = "Scores", ylab = "Frequency", col = "lightblue")
> counts <- grade_hist$counts
> breaks <- grade_hist$breaks
>
> cat("Counts per grade range:\n")
Counts per grade range:
> for (i in seq_along(counts)) {
```

```
> grade_hist <- hist(scores$Score, breaks = grade_breaks, right = TRUE,
+                    main = "Histogram with Grade Ranges",
+                    xlab = "Scores", ylab = "Frequency", col = "lightblue")
> counts <- grade_hist$counts
> breaks <- grade_hist$breaks
>
> cat("Counts per grade range:\n")
Counts per grade range:
> for (i in seq_along(counts)) {
+   grade <- switch(i,
+                   "1" = "C grade",
+                   "2" = "B grade",
+                   "3" = "A grade")
+   cat(paste0(counts[i], " students in ", grade,
+              " range (", breaks[i], ", ", breaks[i + 1], "]\n"))
+ }
17 students in C grade range (30, 50]
70 students in B grade range (50, 70]
13 students in A grade range (70, 90]
>
```



Appendix: Example Question and Answer for R programming questions:

Calculate the sum $\sum_{j=0}^{n} r^j$, where r has been assigned the value 1.08, and compare with $(1-r^{n+1})/(1-r)$, for $n = 10, 20, 30, 40$.

Answer: Copy and paste your R code in the box below (not an image but the text).

```
r <- 1.08
n <- c(10, 20, 30, 40)
sum1 <- c()
for(i in n){
  x <- 0:i
  sum1 <- c(sum1, sum(r^x))
}
sum1   # This gives the calculated sums for n = 10, 20, 30, 40.

sum2 <- (1 - r^(n + 1)) / (1 - r)
sum2

sum2 - sum1   # The formula works.
```

Screenshot of your R console outputs and paste the image in the box below

```
> r <- 1.08
> n <- c(10, 20, 30, 40)
> sum1 <- c()
> for(i in n){
+   x <- 0:i
+   sum1 <- c(sum1, sum(r^x))
+ }
> sum1   # This gives the calculated sums for n = 10, 20, 30, 40.
[1]  16.64549  50.42292 123.34587 280.78104
> sum2 <- (1 - r^(n + 1)) / (1 - r)
> sum2
[1]  16.64549  50.42292 123.34587 280.78104
> sum2 - sum1   # The formula works.
[1] 0 0 0 0
```

THE END