Midterm

Instructions

- Read all of these instructions closely.
- The midterm is due Monday, March 6, 2023 at 4:15pm EST.
- Submit files electronically to me via:
  - Github or
  - Slack or
  - Email
- You may use class resources and online resources during the exam. You may not use ChatGPT or any other software the write the code for you. You also may not message each other during the exam.

Question 1

For question 1, we'll use the model output from a simple regression.

```
X <- c(4.17,5.58,5.18,6.11,4.50,4.61,5.17,4.53,5.33,5.14)
Y <- c(4.81,4.17,4.41,3.59,5.87,3.83,6.03,4.89,4.32,4.69)
lm_output <- lm(Y ~ X)
```

1a

`lm_output` is a list, containing regression model output. Print only the first element of this list.

Answer:

```
print(lm_output[1:3])
```

```
## $coefficients
## (Intercept)           X
##   7.7957139  -0.6229559
##
## $residuals
##           1           2           3           4           5           6
## -0.38798795 -0.14962019 -0.15880253 -0.39945359  0.87758748 -1.09388737
##           7           8           9          10
##  1.45496791 -0.08372384 -0.15535916  0.09627923
##
## $effects
##  (Intercept)            X
## -14.73937617   1.08972057  -0.05758951  -0.27729458   0.96348516  -1.00551222
##
##   1.55595570   0.00284951  -0.05076775   0.19659135
```

1b

In words, what is the **class,** or type of data, that you printed in 1a? It is not required, but you can also use code to justify your answer.

```
class <- lm_output[1:3]
is.list(class)
```

```
## [1] TRUE
```

The type of data is Answer: A list

1c

In words, what is the **data structure** that you printed in 1a? It is not required, but you can also use code to justify your answer.

```
class <- lm_output[1:3]
str(class)
```

```
## List of 3
##  $ coefficients: Named num [1:2] 7.796 -0.623
##   ..- attr(*, "names")= chr [1:2] "(Intercept)" "X"
##  $ residuals   : Named num [1:10] -0.388 -0.15 -0.159 -0.399 0.878 ...
##   ..- attr(*, "names")= chr [1:10] "1" "2" "3" "4" ...
##  $ effects     : Named num [1:10] -14.7394 1.0897 -0.0576 -0.2773 0.9635 ...
##   ..- attr(*, "names")= chr [1:10] "(Intercept)" "X" "" "" ...
```

Answer: The data structure of this list is made up of a series of arrays

Question 2

In words, describe what the following code is doing. Note the `cars` data has 50 observations and two columns: `speed` which is the stopping speed in mph, and `dist` which is the stopping distance in feet. The data were recorded in the 1920s so the cars were pretty slow!

```
data(cars)
for(i in 1:nrow(cars)){
  if(cars$speed[i] == min(cars$speed)){
    print("Stopping distance (in feet) for slowest car(s) in the dataset")
    print(cars$dist[i])
  }
}
```

```
## [1] "Stopping distance (in feet) for slowest car(s) in the dataset"
## [1] 2
## [1] "Stopping distance (in feet) for slowest car(s) in the dataset"
## [1] 10
```

Answer: 'cars' is a dataframe of 50 observations with varying degrees of speed and distance traveled. I assume the measurements are all for the same car. The for loop here is saying that for each observation (each interval or measurement), if the car's speed is equal to the minimum speed in that variable (4 in this case) then display (print) "Stopping distance (in feet) for slowest car(s) in the dataset" as well as the distance traveled below it. This happened twice at observations 2 and 10.

Question 3

3a

Write a function that will take a vector `vec` as the only input and returns the number of elements that are `NA` in `vec`.

Answer:

```
myfunction <- function(data){
length(which(is.na(vec)))
}
```

3b

Use your function to print the number of elements that are NA in the following vector.

Answer:

```
vec <- c(1, 10, NA, NA, 12)
myfunction(vec)
```

```
## [1] 2
```