Assignment 1

Jaimie Chin

2023-02-01

Question 1

A researcher collects the ages of students in a graduate program to get a sense if how old people are in the program.

a. Create a data set in R using the above data (using data.frame() would help)

```r
# Create a dataframe of participants and their age

ages = data.frame(Participants = c(1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15),
                  Participant_Age = c(28, 23, 28, 30, 24, 30, 20, 25, 29, 24, 24, 24, 20, 28, 29))

ages
```

```
##    Participants Participant_Age
## 1             1              28
## 2             2              23
## 3             3              28
## 4             4              30
## 5             5              24
## 6             6              30
## 7             7              20
## 8             8              25
## 9             9              29
## 10           10              24
## 11           11              24
## 12           12              24
## 13           13              20
## 14           14              28
## 15           15              29
```

b. What is the mean? What is the median? What is the mode?

```r
# Create a function to find the mode of a set of data
getMode = function(v) {
  uniq = unique(v)
  uniq[which.max(tabulate(match(v, uniq)))]
}
```

```r
# Find the mean of ages in the sample
mean = mean(ages$Participant_Age)
sprintf("The mean age in this sample is: %s", mean)
```

```
## [1] "The mean age in this sample is: 25.7333333333333"
```

```
# Find the median of ages in the sample
median = median(ages$Participant_Age)
sprintf("The median age in this sample is: %s", median)
```

```
## [1] "The median age in this sample is: 25"
```

```
# Find the mode of ages in the sample
mode = getMode(ages$Participant_Age)
sprintf("The mode age in this sample is: %s", mode)
```

```
## [1] "The mode age in this sample is: 24"
```

c. What can you say about the three measures of central tendency?

Question 2

Create a random selection of 10 numbers ranging from 1 to 100 (using the sample() function with replace set to TRUE). You'll need to store this sample into an object (i.e., create a variable) so that you can do calculations with the sample.

```
# Create a sample following the parameters
sample = sample(1:100, 10, replace=TRUE)
sample
```

```
##  [1] 21 61 54 84 61 59 57 35 58 23
```

a. What is the median? What is the mean? Is there a difference between the two? Why?

```
# Find the mean of the sample
mean = mean(sample)
sprintf("The mean in this sample is: %s", mean)
```

```
## [1] "The mean in this sample is: 51.3"
```

```
# Find the median of the sample
median = median(sample)
sprintf("The median in this sample is: %s", median)
```

```
## [1] "The median in this sample is: 57.5"
```

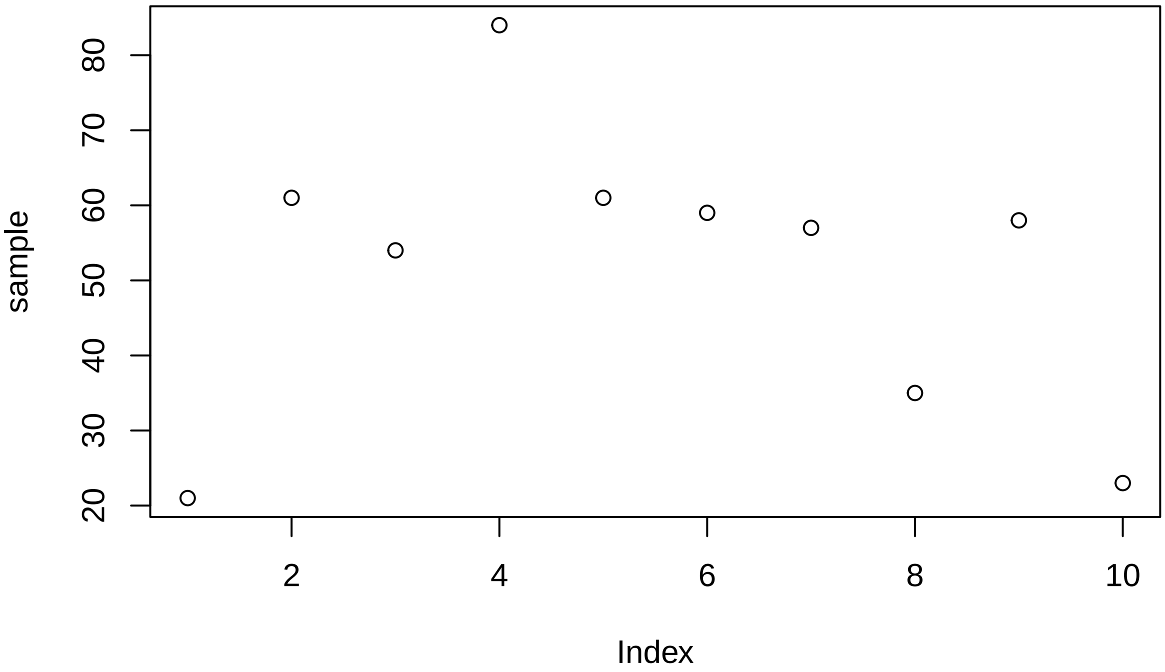
b. Compute the standard deviation

```
# Compute the standard deviation of the sample
sd = sd(sample)
sprintf("The standard deviation of this sample is: %s", sd)
```

```
## [1] "The standard deviation of this sample is: 19.3967351204612"
```

c. Create a plot to visualize these data

```
# Create a scatter plot of the sample values
plot(sample)
```



Now create a random sample of 100 numbers ranging from 1 to 100 (again, using the sample() function with replace set to TRUE).

```
# Create a sample following the parameters
sample = sample(1:100, 100, replace=TRUE)
sample
```

```
##   [1]  8 18 34 45 62 87 75 99 44 61 21 37 95 98 96 66 63 84 75 17 48 97 83 36 16
##  [26] 81 11  1 65 18 29 12  4 73 77 74  8 42 52  7 44 20 25 75 18 90 33 22 66 72
##  [51] 92 52 35 44  1 74 47 15 22 97 57 65 28  6 48 54 68 55 33 81 61 11 83 41 69
##  [76]  1 25 60 30 35 37 34 13  3 59  6 16 13 85 33 85  7 91 60 91 33 21 11 89 90
```

a. What is the median, mean, and standard deviation of this sample?

```
# Find the mean of the sample
mean = mean(sample)
sprintf("The mean in this sample is: %s", mean)
```

```
## [1] "The mean in this sample is: 47.51"
```

```
# Find the median of the sample
median = median(sample)
sprintf("The median in this sample is: %s", median)
```

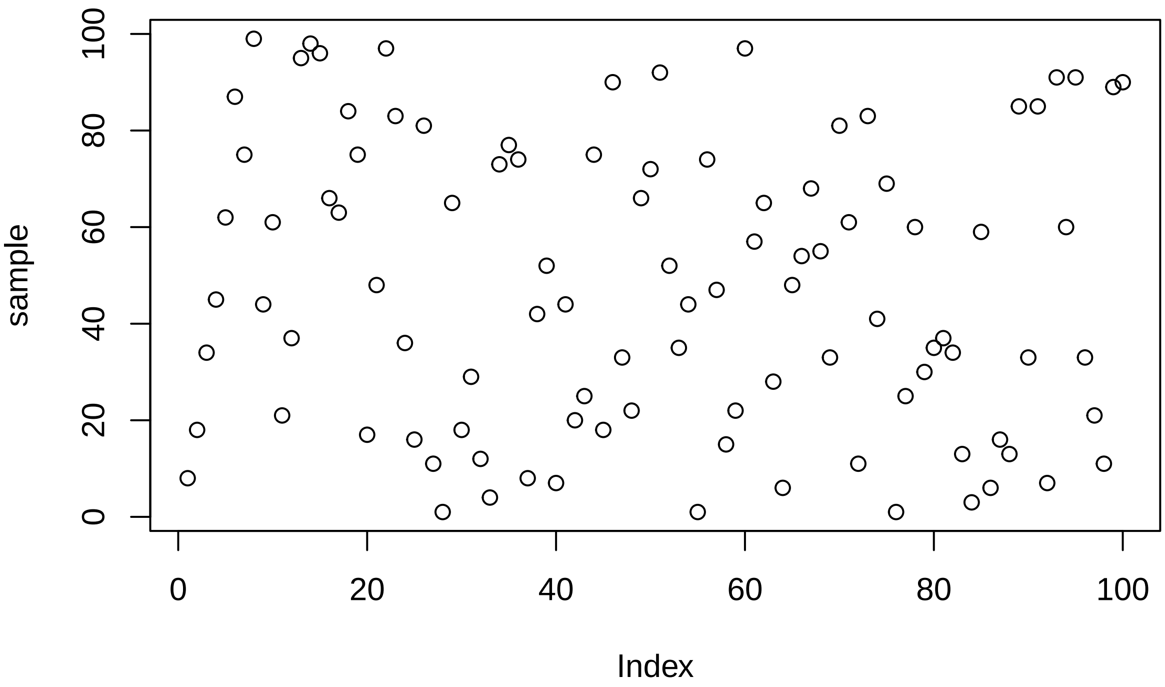
```
## [1] "The median in this sample is: 44.5"
```

```
# Compute the standard deviation of the sample
sd = sd(sample)
sprintf("The standard deviation of this sample is: %s", sd)
```

```
## [1] "The standard deviation of this sample is: 29.7560773862104"
```

b. Create a plot to visualize these data

```
# Create a scatter plot of the sample values
plot(sample)
```



c. How do these values from your sample of 100 compare to your first sample of 10? Are they similar/different? Why? [If you expected them to be similar and they are not (or vice versa), why aren't they similar (or different)?]

Question 3

A researcher wants to study the impact of the presence (in the room) of a sweet snack on task completion. 30 participants are given 5 logic problems to solve. Half of the participants are randomly assigned to desks that have only a pencil and the word problems. The remaining participants are assigned to desks with a pencil, the same word problems, and a candy dispenser. Participants are timed and the completion times are recorded. These are the times (in seconds)

```
# Create a dataframe of Candy_Absent group

candy_absent = data.frame(Participant = c(1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15),
                  Reaction_Time = c(501, 536, 659, 317, 530, 523, 381, 573, 535, 509, 604, 704, 370, 44

candy_absent
```

```
##    Participant Reaction_Time
## 1            1           501
## 2            2           536
## 3            3           659
## 4            4           317
## 5            5           530
## 6            6           523
## 7            7           381
## 8             8           573
## 9             9           535
## 10           10           509
## 11           11           604
## 12           12           704
## 13           13           370
## 14           14           440
## 15           15           404
```

```r
# Create a dataframe of Candy_Present group

candy_present = data.frame(Participant = c(16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30),
                  Reaction_Time = c(690, 691, 510, 586, 675, 470,  533, 693, 440, 614, 475, 374, 500, 4

candy_present
```

```
##    Participant Reaction_Time
## 1           16           690
## 2           17           691
## 3           18           510
## 4           19           586
## 5           20           675
## 6           21           470
## 7           22           533
## 8           23           693
## 9           24           440
## 10          25           614
## 11          26           475
## 12          27           374
## 13          28           500
## 14          29           478
## 15          30           664
```

a. Compute the mean and median for both groups. What do you think about the results you've computed?

```r
# Find the mean of reaction times in the Candy Absent group
mean = mean(candy_absent$Reaction_Time)
sprintf("The mean reaction time in the Candy-Absent group is: %s", mean)
```

```
## [1] "The mean reaction time in the Candy-Absent group is: 505.733333333333"
```

```r
# Find the median of reaction time in the Candy Absent group
median = median(candy_absent$Reaction_Time)
sprintf("The median reaction time in the Candy-Absent group: %s", median)
```

```
## [1] "The median reaction time in the Candy-Absent group: 523"
```

```r
# Find the mean of reaction times in the Candy Present group
mean = mean(candy_present$Reaction_Time)
sprintf("The mean reaction time in the Candy-Present group is: %s", mean)
```

```
## [1] "The mean reaction time in the Candy-Present group is: 559.533333333333"
```

```r
# Find the median of reaction time in the Candy Present group
median = median(candy_present$Reaction_Time)
sprintf("The median reaction time in the Candy-Present group: %s", median)
```

```
## [1] "The median reaction time in the Candy-Present group: 533"
```

b. If you changed the highest score in the Candy-Absent group to be 10 times the original value, what would happen to the mean? What about median?

```r
# Find the index of the max reaction time in the Candy Absent group
max_time = which.max(candy_absent$Reaction_Time)

# Create a copy of the candy_absent data frame and change the value of the highest score
candy_absent_copy = cbind(candy_absent)
candy_absent_copy = replace(candy_absent_copy$Reaction_Time, candy_absent_copy$Participant == max_time,

# Find the mean of reaction times in the altered Candy Absent group
mean = mean(candy_absent_copy)
sprintf("The mean reaction time in the altered Candy-Absent group is: %s", mean)
```

```
## [1] "The mean reaction time in the altered Candy-Absent group is: 928.133333333333"
```

```r
# Find the median of reaction time in the altered Candy Absent group
median = median(candy_absent_copy)
sprintf("The median reaction time in the altered Candy-Absent group: %s", median)
```

```
## [1] "The median reaction time in the altered Candy-Absent group: 523"
```

c. If you changed the highest score in the Candy-Present group to be one tenth is original value, what would happen to the mean? What about the median?

```r
# Find the index of the max reaction time in the Candy Present group
max_time = which.max(candy_present$Reaction_Time)
max_time = candy_present$Participant[[max_time]]

# Create a copy of the candy_present data frame and change the value of the highest score
candy_present_copy = cbind(candy_present)
candy_present_copy = replace(candy_present_copy$Reaction_Time, candy_present_copy$Participant == max_ti

# Find the mean of reaction times in the altered Candy Present group
mean = mean(candy_present_copy)
sprintf("The mean reaction time in the altered Candy-Present group is: %s", mean)
```

```
## [1] "The mean reaction time in the altered Candy-Present group is: 975.333333333333"
```

```r
# Find the median of reaction time in the altered Candy Present group
median = median(candy_present_copy)
sprintf("The median reaction time in the altered Candy-Present group: %s", median)
```

```
## [1] "The median reaction time in the altered Candy-Present group: 533"
```

Question 4

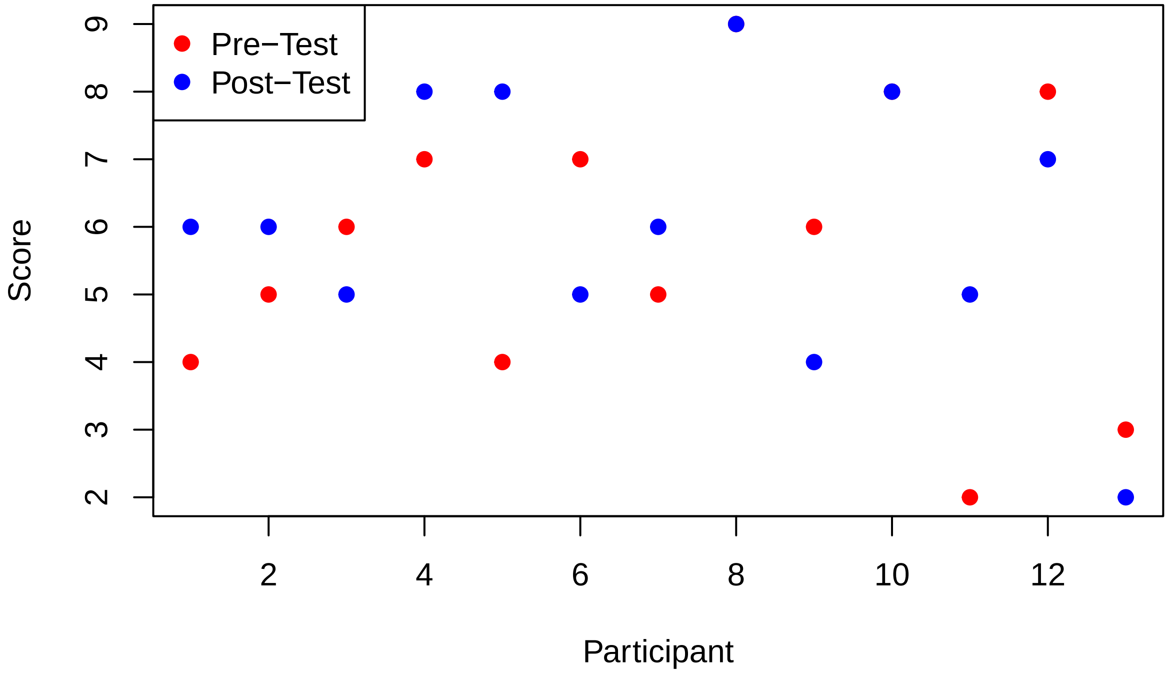
A teacher developed a new studying technique and was interested in seeing if it worked. To test this, they had students take a test before using the studying technique (to get a baseline score) and then had students take a similar test one week later (after using the studying technique).

```r
# Create a dataframe of scores

scores = data.frame(Participant = c(1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13),
                    Pre_Test = c(4, 5, 6, 7, 4, 7, 5, 9, 6, 8, 2, 8, 3),
                    Post_Test = c(6, 6, 5, 8, 8, 5, 6, 9, 4, 8, 5, 7, 2))

scores
```

```
##    Participant Pre_Test Post_Test
## 1            1        4         6
## 2            2        5         6
## 3            3        6         5
## 4            4        7         8
## 5            5        4         8
## 6            6        7         5
## 7            7        5         6
## 8            8        9         9
## 9            9        6         4
## 10          10        8         8
## 11          11        2         5
## 12          12        8         7
## 13          13        3         2
```

a. Plot participants' pre- and post-test scores in a single plot. What do you see? Does there appear to be a relationship between their scores before and after using the studying technique?

```r
# Plot Pre-Test & Post-Test
plot(scores$Participant, scores$Pre_Test, pch=19, col="red", xlab = "Participant", ylab = "Score")
points(scores$Participant, scores$Post_Test, pch=19, col = "blue")
legend("topleft", legend = c("Pre-Test", "Post-Test"),
       pch = 19, col = c("red", "blue"))
```



b. Now check to see if there is a correlation between participants’ pre- and post-test scores. What is the correlation? You can use the cor() function in R.

```r
# Find the correlation between participants’ pre- and post-test scores
corr = cor(scores$Pre_Test, scores$Post_Test)
sprintf("The correlation between participants' pre- and post-test scores is: %s", corr)
```

```
## [1] "The correlation between participants' pre- and post-test scores is: 0.581405769696434"
```

c. What do you think of this teacher’s experiment to test their studying technique? Do you think they did a good experiment to yield appropriate data to test the effectiveness of their studying technique?

If yes, why? Be concrete about what you think the teacher did right.

If no, give one suggestion (besides running with a larger sample size) the teacher could do to improve their experiment?