p8130_hw1_ym2813

Anna Ma

9/26/2021

Problem 1

Please classify each of the following variables as qualitative (specify if binary, nominal, or ordinal) or quantitative (specify if discrete or continuous):

a)homework feedback, labeled as "poor", "fair", "good", "very good"

This is an ordinal qualitative variable.

b)homework feedback, labeled as "fail", "pass"

This is a binary qualitative variable.

c)country of birth

This is a nominal qualitative variable.

d)the quantity of grapes (in lbs)to make 3 liters of wine

This is a continuous quantitative variable.

e)number of TAs in the P8130 course

This is a discrete quantitative variable.

Problem 2

In a study of 133 individualswith a recent bike crash history,depression scores were measured using a standardized test.The depression scores for 14 of these individualsare as follows: 45, 39, 25, 47, 49, 5, 70, 99, 74, 37, 99, 35, 8, 59

a) Computethe following descriptive summariesof these data: mean, median, range, SD.

```
depression_score_2.1 <- c(45, 39, 25, 47, 49, 5, 70, 99, 74, 37, 99, 35, 8, 59)
mean(depression_score_2.1) # Data mean
```

```
## [1] 49.35714
```

```r
median(depression_score_2.1)  # Data median
```

```
## [1] 46
```

```r
max(depression_score_2.1, na.rm=TRUE) - min(depression_score_2.1, na.rm=TRUE)   # Data range
```

```
## [1] 94
```

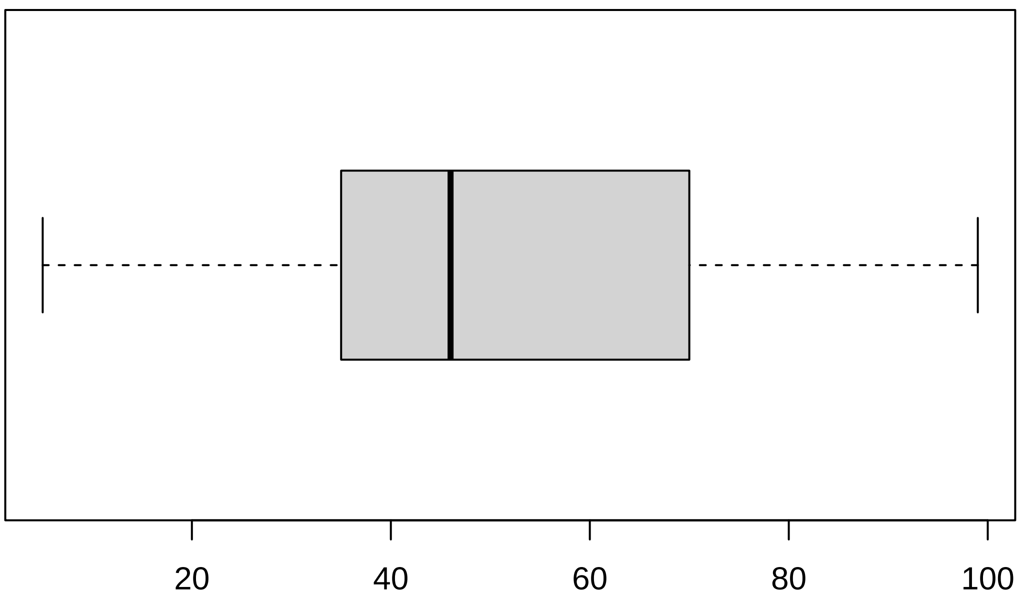
```r
sd(depression_score_2.1)  #Data standrd deviation
```

```
## [1] 28.84603
```

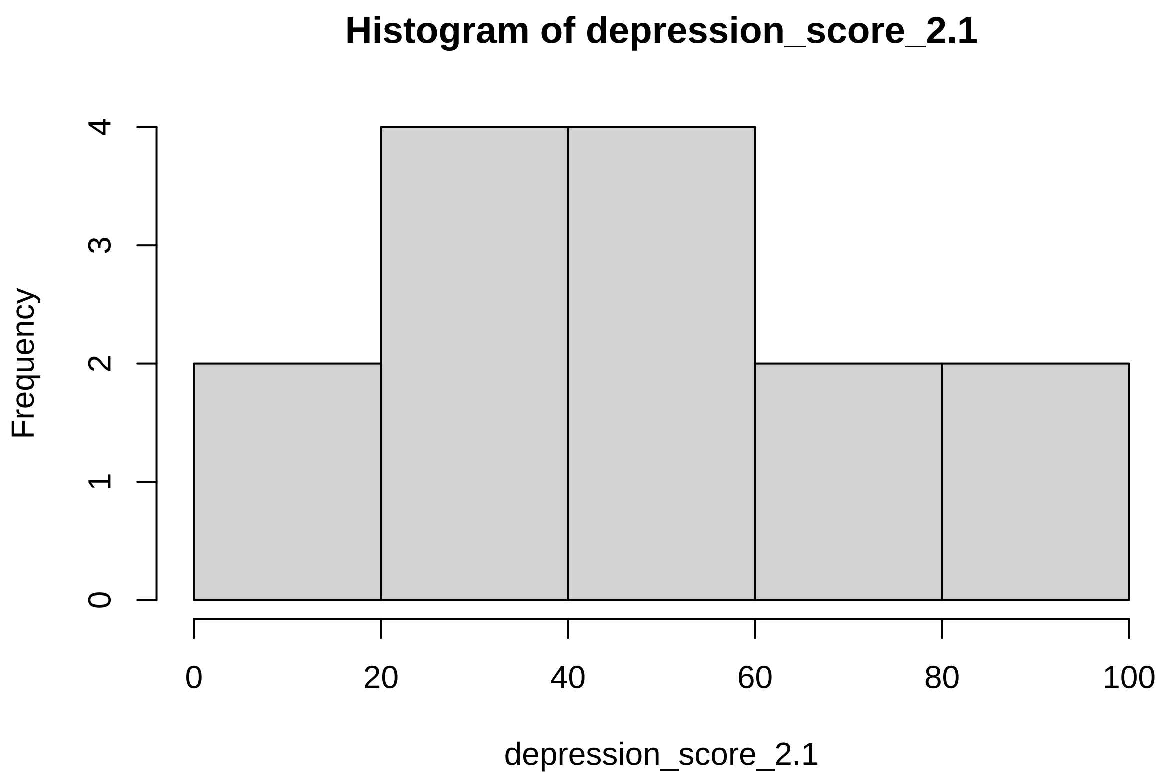
The mean of the depression scores for the 14 individuals is 49.3571429, the median of is 46, the range is 94, and the standard deviation is 28.8460293

b) Describe the box plot and the underlying distribution of the data. Use some of the following terms:left-skewed, right-skewed, symmetric, bimodal, unimodal distribution.

```r
boxplot(depression_score_2.1,horizontal = TRUE)
```




```r
hist(depression_score_2.1)
```

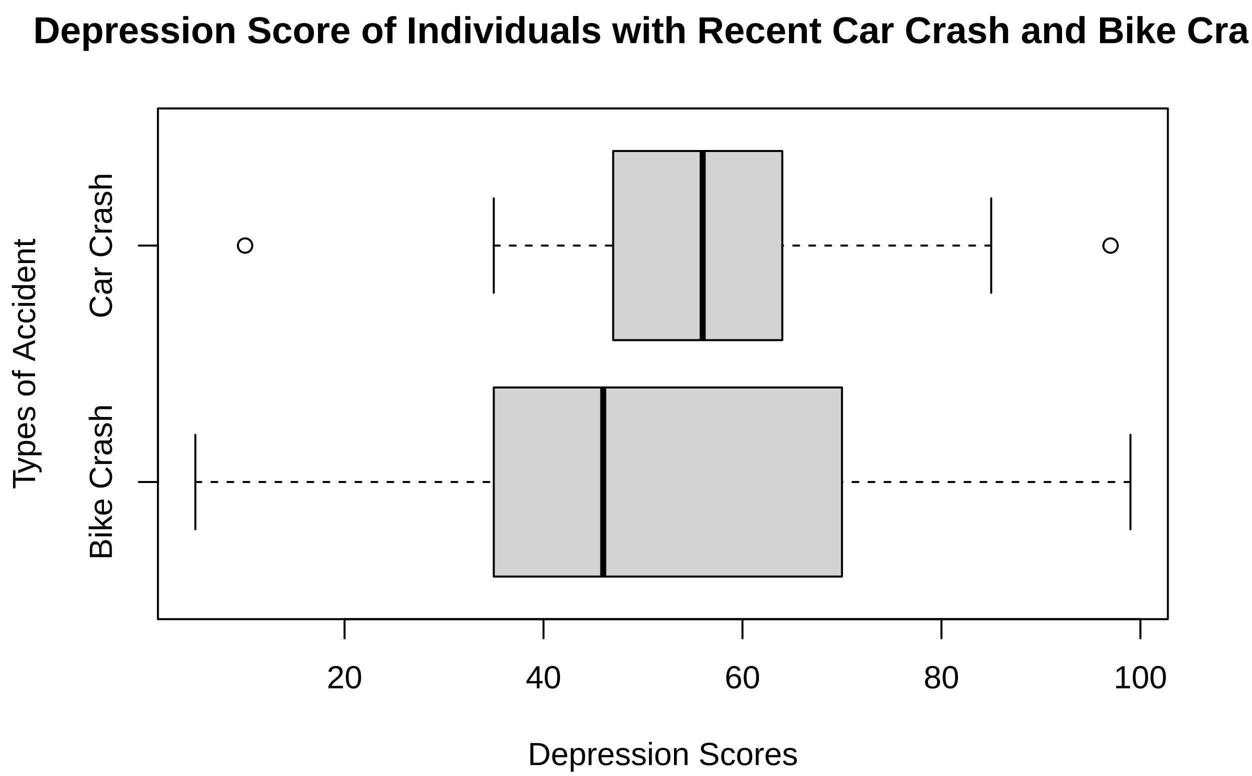


The box plot of the data is right skewed as the median is closer to the lower quartile, and it is also an unimodal distribution.

Additionally, 140 individuals with a recent car crash history alsoparticipated in the study.The depression scores for 13 of these individualsare given below:67, 50, 85, 43, 64, 35, 47, 97, 58, 58, 10, 56, 50

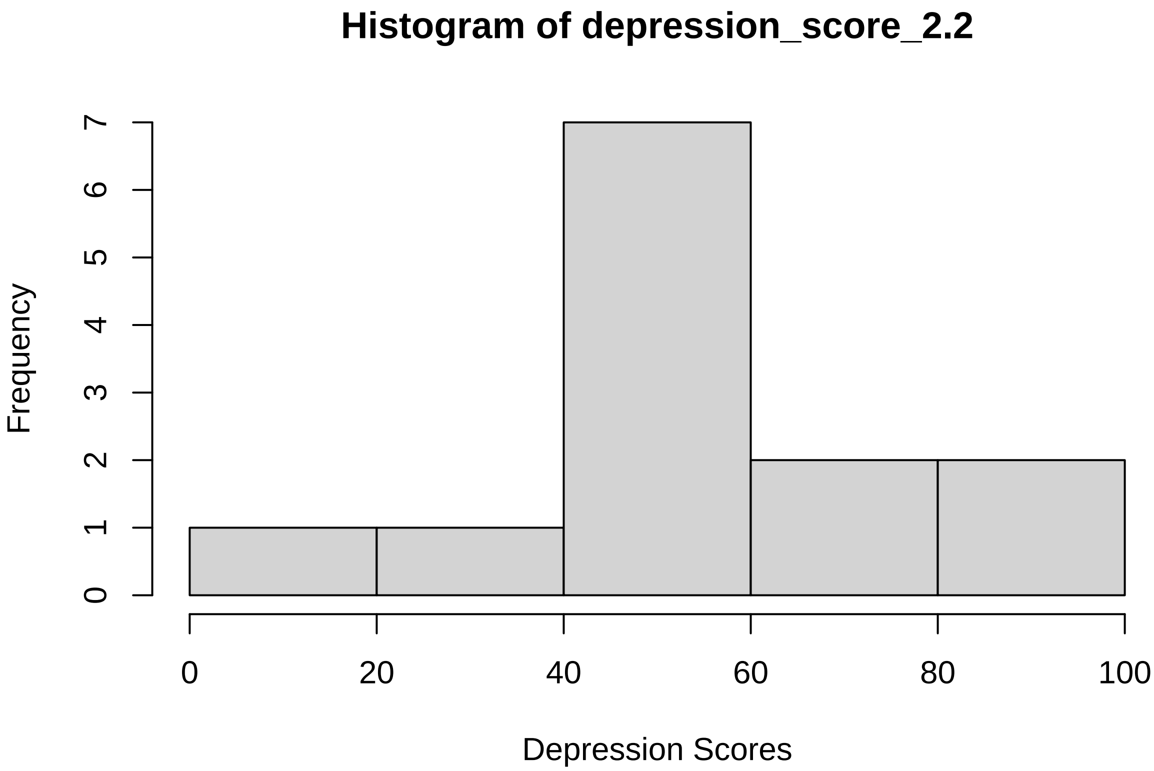
a) Using R, make a side-by-side box plot of the depression scores stratified by type of accident.Make sure you label your figure appropriately.

```
depression_score_2.2 <- c(67, 50, 85, 43, 64, 35, 47, 97, 58, 58, 10, 56, 50)
types <- c("Bike Crash","Car Crash")
boxplot(depression_score_2.1,
        depression_score_2.2,
        names = types, horizontal = TRUE,
        main = "Depression Score of Individuals with Recent Car Crash and Bike Crash",
        xlab = "Depression Scores",
        ylab = "Types of Accident")
```



b) Describe each of the box plots and the underlying distribution of the data. Use some of the following terms:left-skewed, right-skewed, symmetric, bimodal, unimodal distribution.

```
hist(depression_score_2.2,xlab = "Depression Scores")
```



The box plot of the depression score of those who had bike crash is right skewed and unimodal. The box plot for those who had car crash has an unimodal distribution

c) Comparing the 2 box plots,which group appears to have a lower typical depression score?

It appears that the group who had bike crash have a lower typical depression score.

Problem 3

Suppose we toss one fair 12-sided die:

a)Let's define the event A as "an even number appears". What is the probability of the event A?

The probability of even A is

$$P = \frac{A}{All\ Possible\ Numbers} = \frac{6}{12} = \frac{1}{2}$$

b)Let's define the event B as "number 10 appears". What is the probability of the event B?

The probability of even A is

$$P = \frac{B}{All\ Possible\ Numbers} = \frac{1}{12}$$

c)Compute P(B U A).

$$P(B \cup A) = \frac{1}{2}$$

d) Are events A and B independent? Why?Prove your answer.

Event A and B is not independent:

proof : By definition, A and B are independent events if and only if P(A|B) = P(A). Given that B has already occurred such that number 10 has appeared, then the probability of an even number appears, which is event A, should be 1. However, P(A) is $\frac{1}{2}$ as shown in a). Therefore,P(A|B) is not equal to P(A). Thus, A and B are not independent. □