HW1_report

Wentong

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"

An ordinal qualitative variable.

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

A binary qualitative variable.

c)country of birth

A nominal qualitative variable

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

A continuous quantitative variable

e)number of TAs in the P8130 course

A discrete quantitative variable

Problem 2

In a study of 133 individuals with 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)Compute the following descriptive summaries of these data: mean, median, range, SD.

```r
problem_2.1 = c(45, 39, 25, 47, 49, 5, 70, 99, 74, 37, 99, 35, 8, 59)
mean(problem_2.1)
```

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

```r
median(problem_2.1)
```

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

```r
max(problem_2.1) - min(problem_2.1)
```

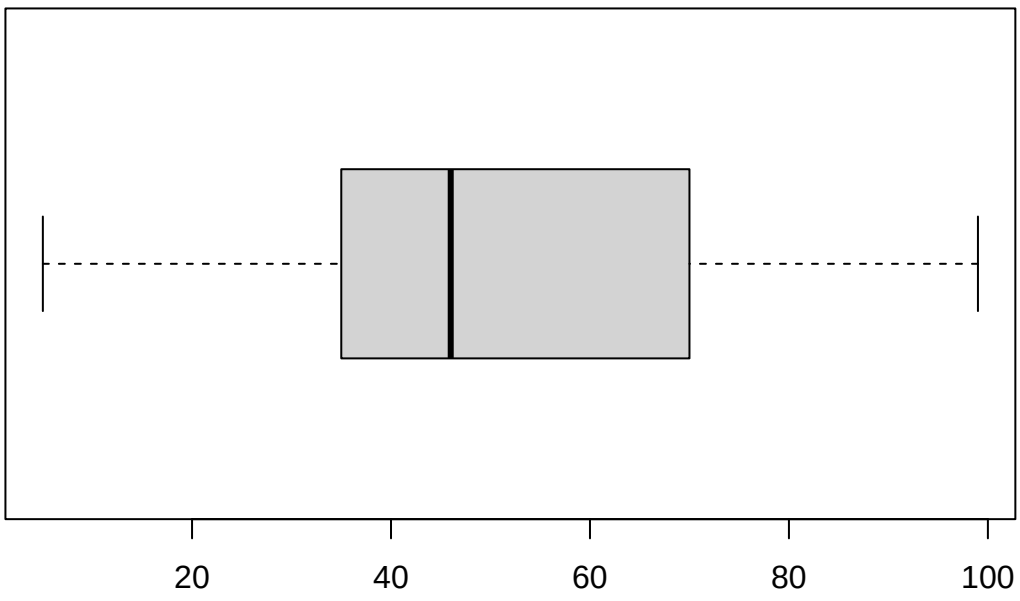
```
## [1] 94
```

```r
sd(problem_2.1)
```

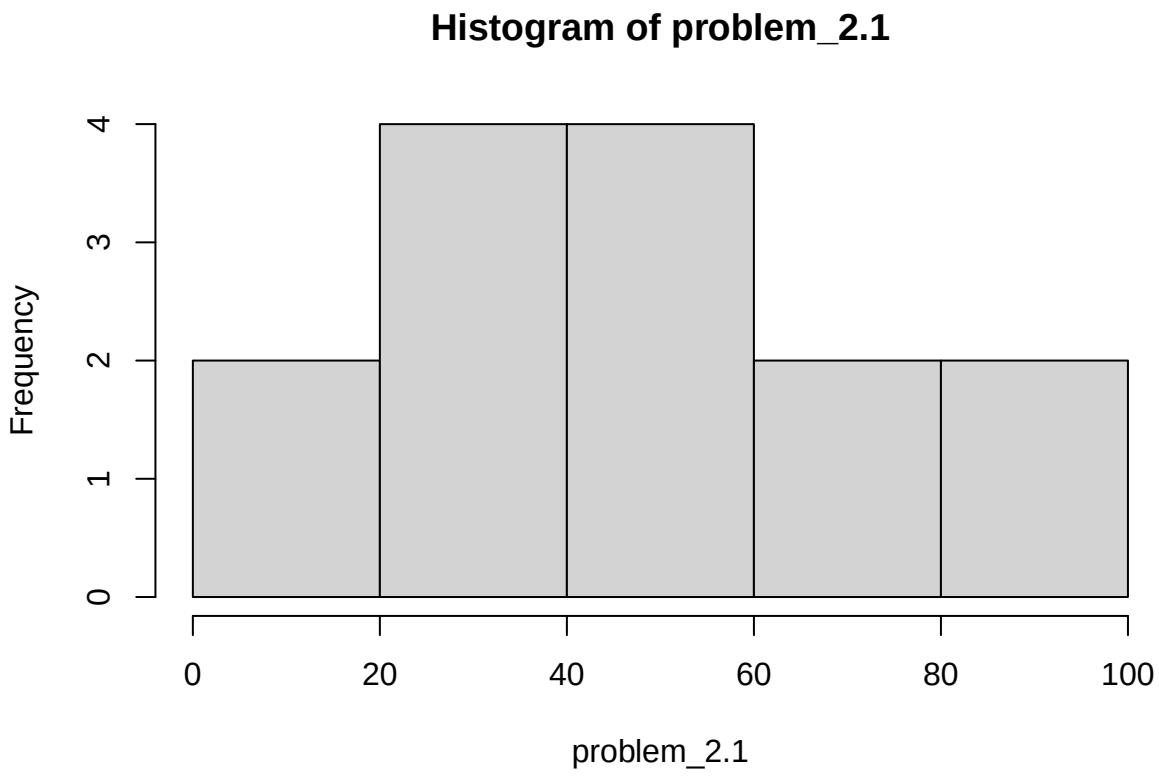
```
## [1] 28.84603
```

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(problem_2.1,horizontal = TRUE)
```



```
hist(problem_2.1)
```



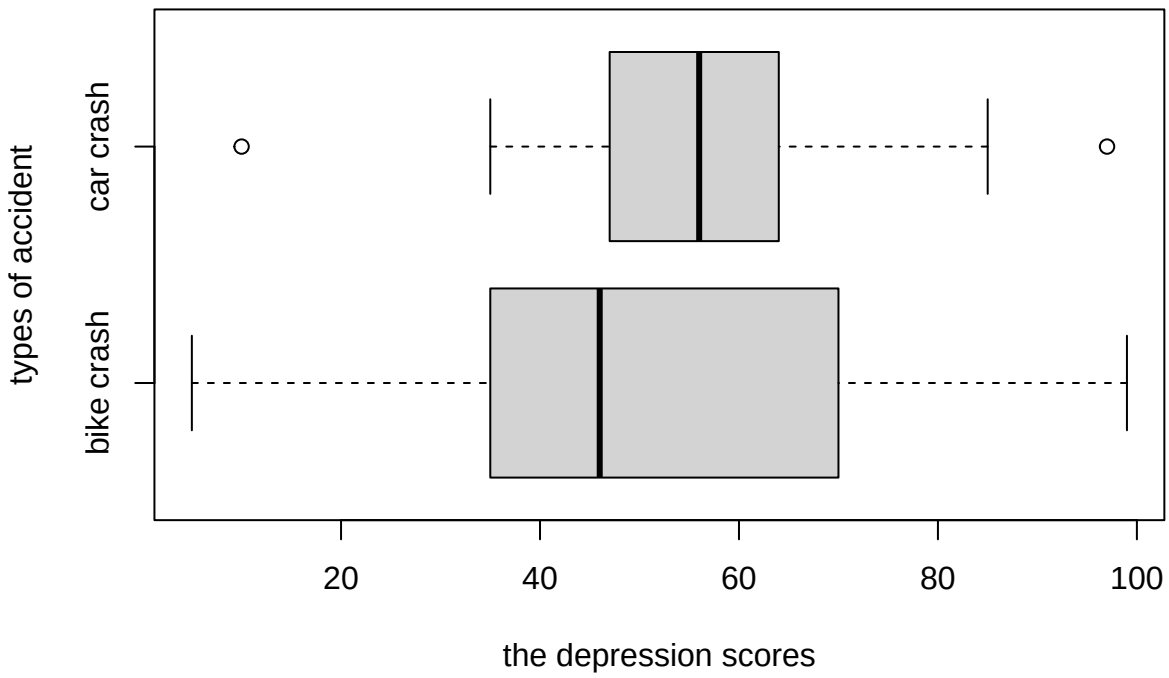
According to the plots, the data of the bike crash is right-skewed and an unimodal distribution.

Additionally, 140 individuals with a recent car crash history also participated 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

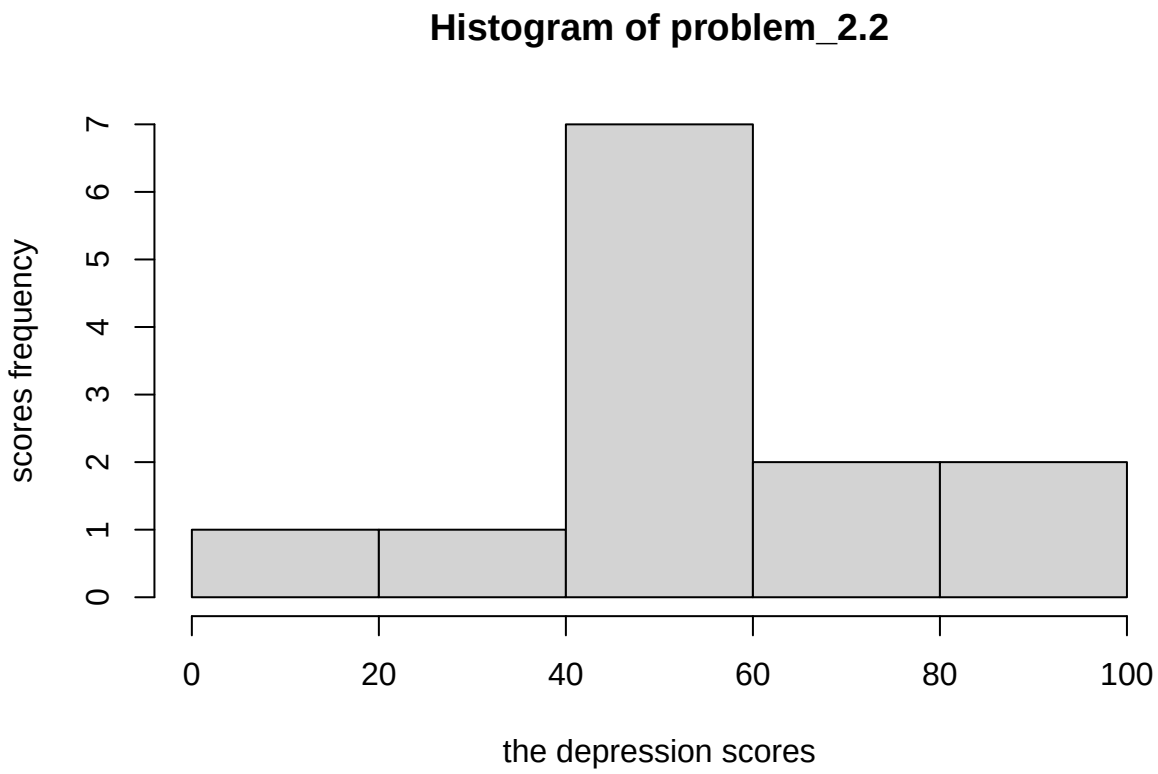
a)Using R, make a side-by-side box plot of the depression scores stratified by type of accident. Make sure you label your figure appropriately.

```
problem_2.2 = c(67, 50, 85, 43, 64, 35, 47, 97, 58, 58, 10, 56, 50)
types <- c("bike crash","car crash")
boxplot(problem_2.1, problem_2.2, names = types, horizontal = TRUE,
        xlab = "the 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(problem_2.2, xlab = "the depression scores", ylab = "scores frequency")
```



According to the plots, the data of the bike crash is right-skewed and an unimodal distribution and the data of the car crash is an unimodal distribution.

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

The plots show that the bike crash group 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?

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

c)Compute P(B U A).

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

Problem 4

5% of women above age of 75 have dementia. Among women (75+ years old) with dementia, 80% have positive findings on their CT scan. Among women (75+ years old) who don't have dementia, 10% will have a positive CT scan findings. A randomly-selected woman (75+ years old) had a positive CT scan findings. What is the probability that she actually has dementia? Compute by hand and show the key steps. The answer can be hand written.