Homework1

Kangyu Xu (kx2224)

2024-09-20

Question1

(a) Qualitative: Ordinal
(b) Qualitative: Binary
(c) Qualitative: Nominal
(d) Quatitative: Continuous
(e) Quatitative: Discrete

Question2

```r
# Basic info
bike_scores <- c(45, 39, 25, 47, 49, 5, 70, 99, 74, 37, 99, 35, 8, 59)
car_scores <- c(67, 50, 85, 43, 64, 35, 47, 97, 58, 58, 10, 56, 50)
```

a) Compute Mean, Median, Range, SD

```r
mean_bike = mean(bike_scores)
median_bike = median(bike_scores)
range_bike = range(bike_scores)
sd_bike = sd(bike_scores)
cat("Mean (bike crash group):", mean_bike, "\n")
```

```
## Mean (bike crash group): 49.35714
```

```r
cat("Median (bike crash group):", median_bike, "\n")
```

```
## Median (bike crash group): 46
```

```r
cat("Range (bike crash group):", range_bike, "\n")
```

```
## Range (bike crash group): 5 99
```

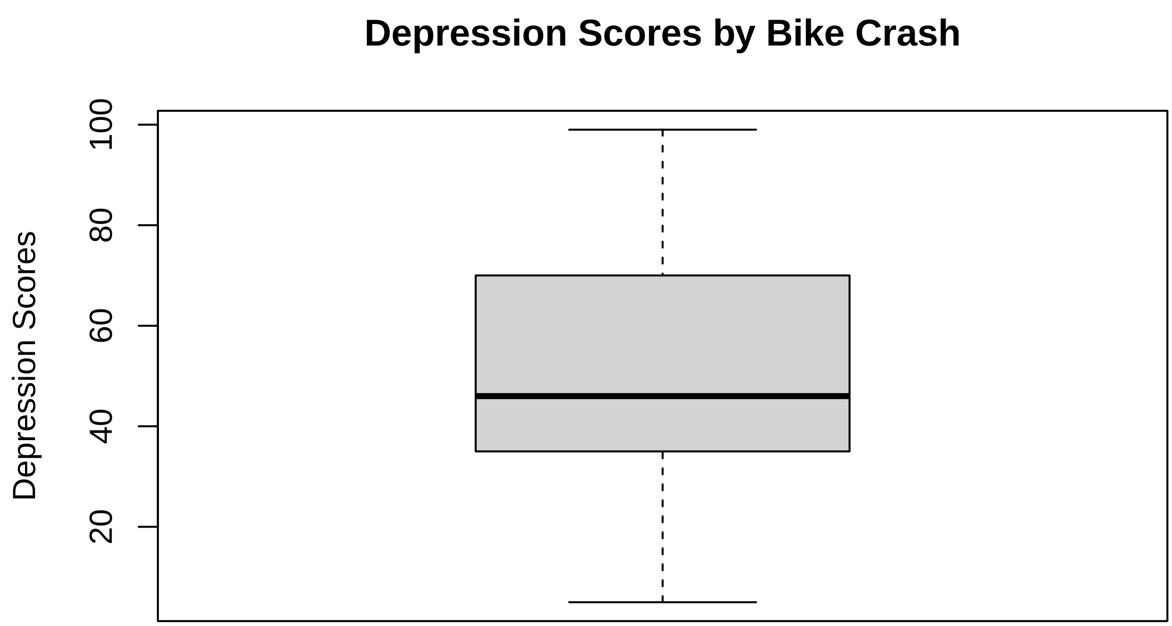
```r
cat("Standard Deviation (bike crash group):", sd_bike, "\n")
```

```
## Standard Deviation (bike crash group): 28.84603
```

b) Description

```r
# Box plot
accident_type <- factor(c(rep("Bike", length(bike_scores))))
depression_scores <- c(bike_scores)

boxplot(depression_scores,
        main = "Depression Scores by Bike Crash",
        ylab = "Depression Scores",
        border = "black")
```



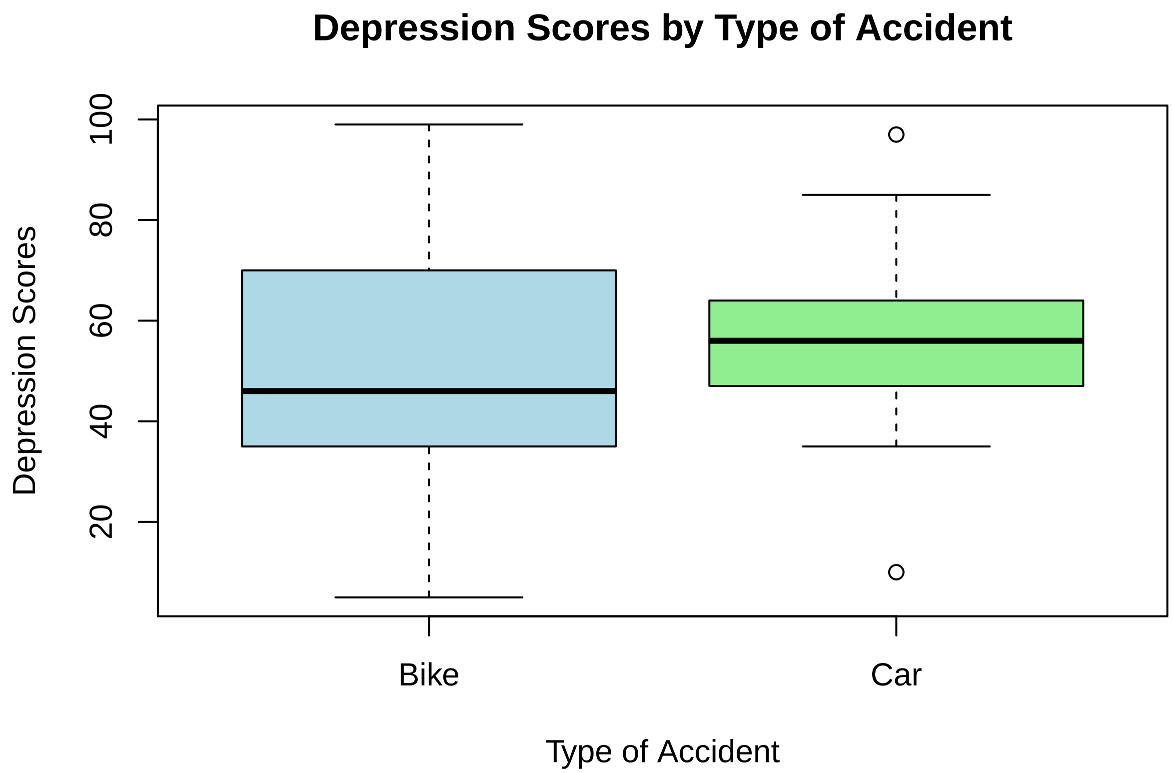
Therefore the boxplot shows that the median of this dataset is 46. And the shape of box shows that the underlying distribution seems like right-skewed and unimodal.

c) Plot

```r
# Box plot
accident_type <- factor(c(rep("Bike", length(bike_scores)), rep("Car", length(car_scores))))
depression_scores <- c(bike_scores, car_scores)

boxplot(depression_scores ~ accident_type,
        main = "Depression Scores by Type of Accident",
        xlab = "Type of Accident",
```

```
        ylab = "Depression Scores",
        col = c("lightblue", "lightgreen"))
```



d) description

The boxplot of bike-crash dataset shows that the median of this dataset is 46. And the shape of box shows that the underlying distribution seems like right-skewed and unimodal. The boxplot of car-crash dataset shows that the median of this dataset is 58. And the shape of box shows that the underlying distribution seems like symmetric and unimodal.

e) Comparison

The bike crash group has a lower median depression score (around 46) compared to the car crash group (with a median around 58). This suggests that the typical depression score is lower in bike crash group.

Question3

a)

$$A = \text{"an even number appears"}$$
$$P(A) = \frac{1}{2}$$

b)

$$B = \text{"number 10 appears"}$$
$$P(B) = \tfrac{1}{12}$$

c)

$$\because B \subset A$$
$$\therefore P(B \cup A) = P(A) = \tfrac{1}{2}$$

d)

A and B are not independent. This is because if A does not happen, B cannot happen neither.

Question4

$$
\begin{aligned}
&P(dementia) = 5\% \\
&P(positive|dementia) = 80\% \\
&P(positive|\neg dementia) = 10\% \\
&------------------------------ \\
&\therefore P(dementia|positive) = \frac{P(dementia, positive)}{P(positive)} \\
&= \frac{P(positive|dementia) \times P(dementia)}{\sum_{s \subset \{dementia, \neg dementia\}} P(positive|s) \times P(s)} \\
&= \frac{0.8 * 0.05}{0.8 * 0.05 + 0.1 * 0.95} \\
&= \frac{8}{27}
\end{aligned}
$$

So the probability that she actually has dementia is about 0.2962