HW4-tk2886-2021

Tanvir Khan

Loading the data

```r
crash_df <-
  read.csv("Crash.csv")
```

Tidy the data

```r
crash_dfn <-
  crash_df %>%
  pivot_longer(everything(),
               names_to = "type_of_accidents",
               values_to = "Values")
```

Problem 2a

Generate Descriptive statistics for each group

```r
crash_df %>% summary()
```

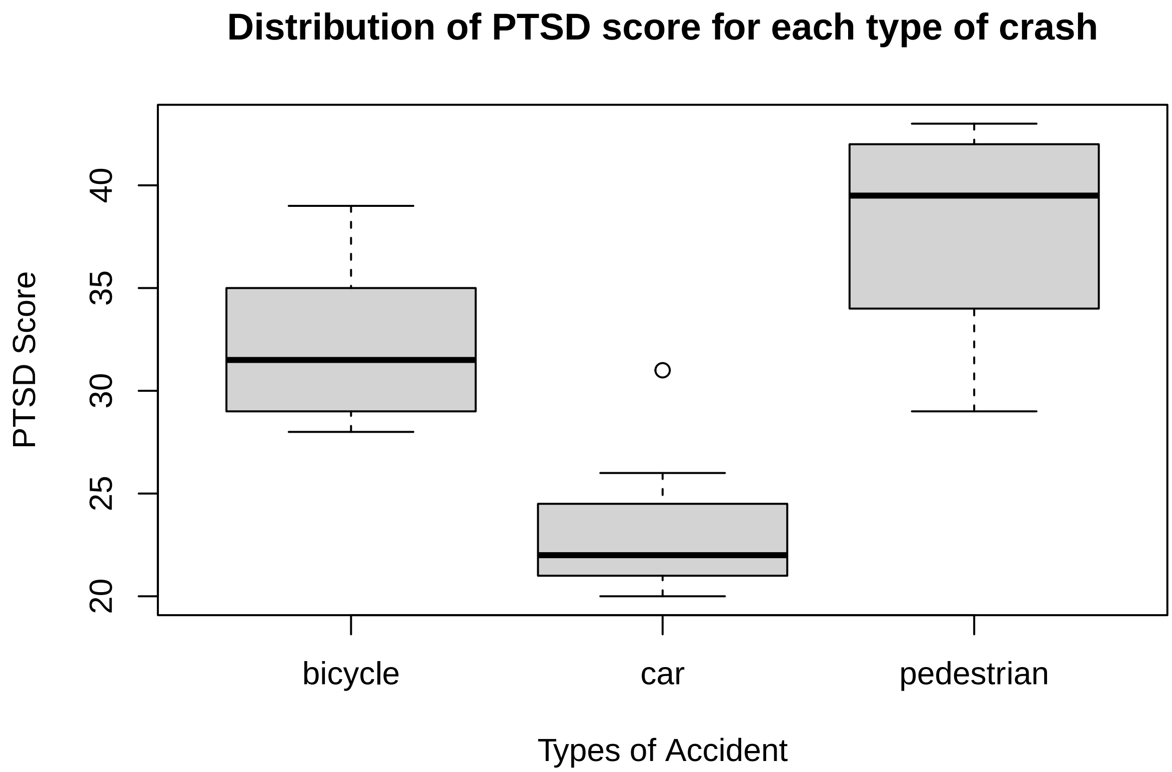
```
##    pedestrian       bicycle          car
##  Min.   :29.00   Min.   :28.0   Min.   :20.00
##  1st Qu.:36.00   1st Qu.:29.5   1st Qu.:21.00
##  Median :39.50   Median :31.5   Median :22.00
##  Mean   :37.88   Mean   :32.5   Mean   :23.43
##  3rd Qu.:42.00   3rd Qu.:34.5   3rd Qu.:24.50
##  Max.   :43.00   Max.   :39.0   Max.   :31.00
##  NA's   :2                      NA's   :3
```

```r
crash_df %>% summarize_if(is_numeric, sd, na.rm = T)
```

```
##   pedestrian  bicycle      car
## 1    5.43632 4.062019 3.866831
```

#boxplot

```r
boxplot(Values ~ type_of_accidents, data = crash_dfn,
        main = "Distribution of PTSD score for each type of crash",
        xlab = "Types of Accident",
        ylab = "PTSD Score")
```



Analysis n Differences Observed:

In the data set that was provided, the mean of the PTSD Score for Pedestrian Incidents is the largest among the three types of accidents. The mean of the PTSD Score of bicycle incidents is the second highest and the mean of the PTSD Scores for car incidents is the lowest. The standard deviation for PTSD score for pedestrian incident is 5.44, the standard deviation for PTSD score in bicycle incident is 4.06, and the standard deviation for the PTSD score for car crash is 3.87. Pedestrian has a larger standard deviation than bicycle and car. This indicates that the PTSD score for this pedestrian incidents is more spread out compared to the other two types if crash. When pedestrians is involved in an incident, their PTSD in generall is more varied compared to the other groups (bicycle and cars).

Problem 2b:

```r
res1 = aov(Values ~ factor(type_of_accidents), data = crash_dfn)
summary(res1)
```

```
##                            Df Sum Sq Mean Sq F value   Pr(>F)    
## factor(type_of_accidents)  2  790.4   395.2   19.53 1.33e-05 ***
## Residuals                 22  445.1    20.2                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 5 observations deleted due to missingness
```

```r
critical_value = qf(0.99, 2, 22)
```

Interpretation:

Hypothesis: Ho: mu1 = mu2 = mu3

Ha: at least two means are not equal

Test-Statistics: F-Value: 19.53

Critical Value: Critical Value: 5.7190219

Decision Intercepted: Our F-statistics (19.53) is bigger than our critical value (5.72), we reject the null hypothesis. At 0.01 significance level, we reject the null hypothesis and conclude that at least two mean PTSD score from the three type of crash groups are different.

Problem 2c:

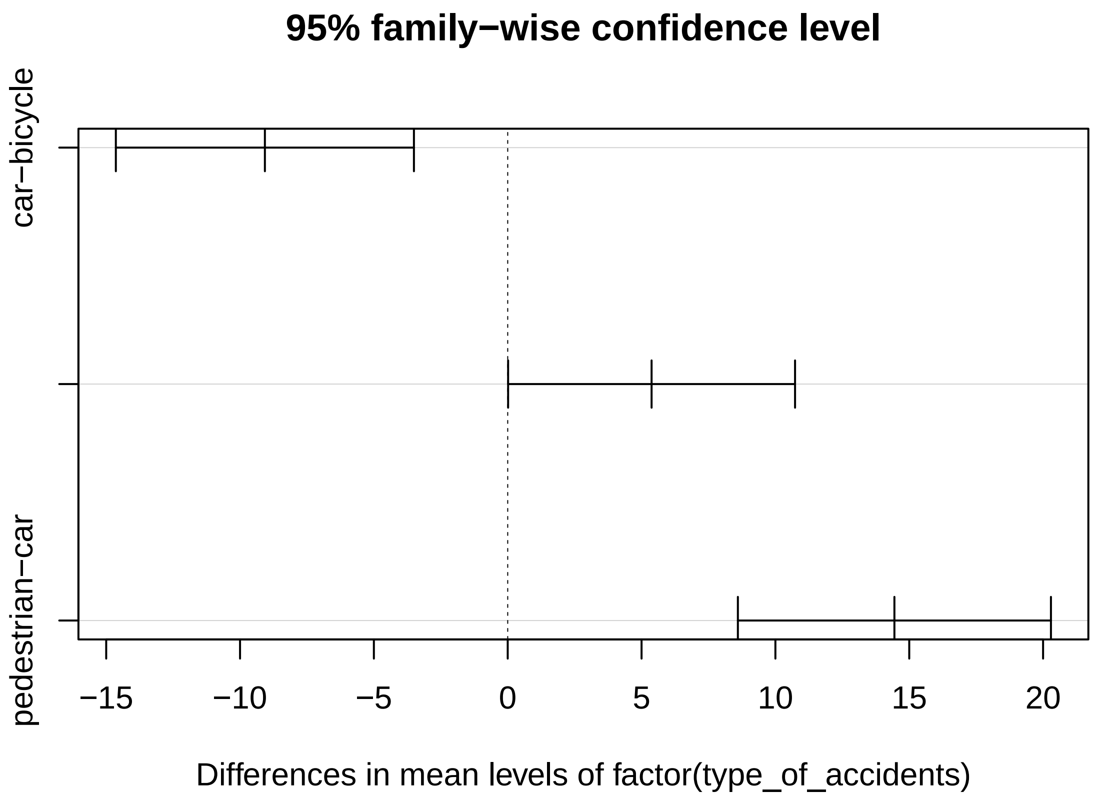
```r
# use pairwise.t.test() for Bonferroni, Holm, Benjamini-Hochberg
pairwise.t.test(crash_dfn$Values, crash_dfn$type_of_accidents, p.adj = 'bonferroni')
```

```
## 
##  Pairwise comparisons using t tests with pooled SD 
## 
## data:  crash_dfn$Values and crash_dfn$type_of_accidents 
## 
##            bicycle car    
## car        0.0014  -      
## pedestrian 0.0586  9.1e-06
## 
## P value adjustment method: bonferroni
```

```r
# use Tukey
Tukey_comp = TukeyHSD(res1)
Tukey_comp
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
## 
## Fit: aov(formula = Values ~ factor(type_of_accidents), data = crash_dfn)
## 
## $`factor(type_of_accidents)`
##                         diff         lwr       upr     p adj
## car-bicycle        -9.071429 -14.63967214 -3.503185 0.0013441
## pedestrian-bicycle  5.375000   0.01537946 10.734621 0.0492580
## pedestrian-car     14.446429   8.59860314 20.294254 0.0000088
```

```r
plot(Tukey_comp)
```



Problem 3b:

```r
test_statistcs = (15-17.67)^2/17.67 + (18-17.67)^2/17.67 + (20-17.67)^2/17.67 + (18-15.33)^2/15.33 + (1
```

```r
critical_value = qchisq(.95, 1)
```

Problem 3c:

Ho: Relapse Status and types of anti-depressant are independent

Ha: Relase Status and types of anti-depressant are associated/dependent