HW4-tk2886

Tanvir Khan

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.4     v dplyr   1.0.7
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   2.0.1     v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

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

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

```
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
```

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

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

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
qf(0.99, 2, 22)
```

```
## [1] 5.719022
```

Interpretation:

Problem 2c:

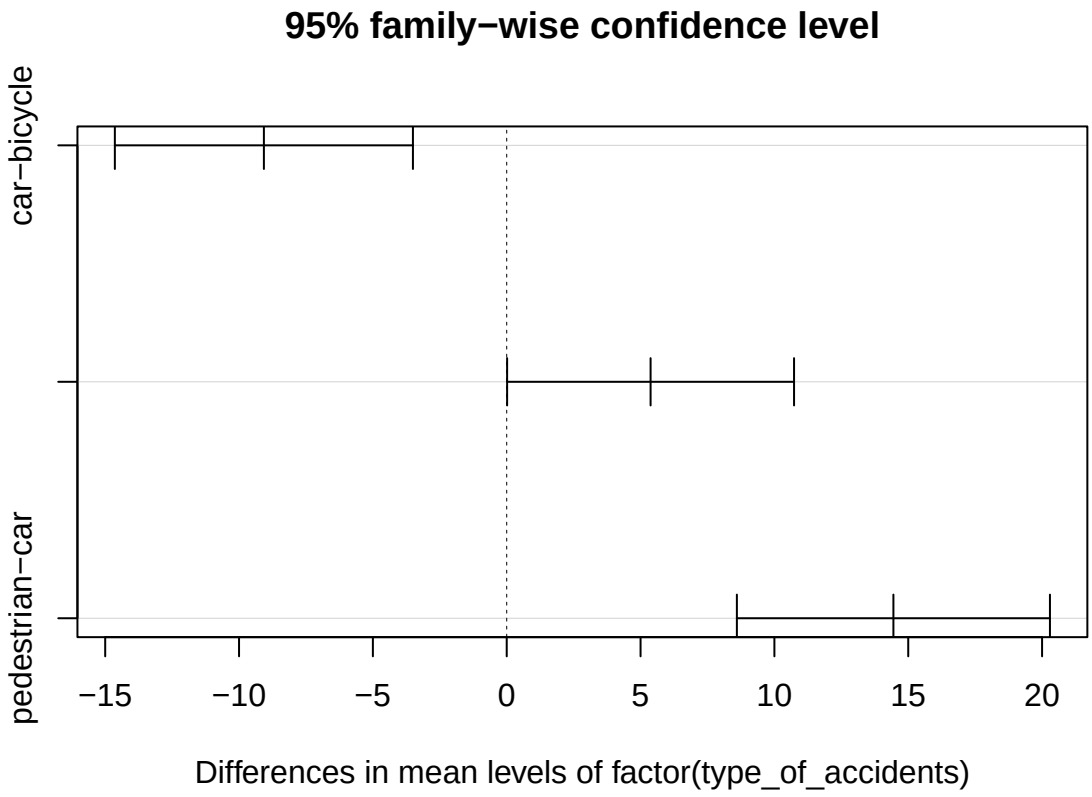
```r
# use pairwise.t.test() for Bonferroni, Holm, Benjamini-Hochberg
pairwise.t.test(crash_dfn$Values, crash_dfn$type_of_accidents, p.adj = 'bonferroni')
```

```
##
##  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: Ho: Relapse Status and types of anti-depressant are independent Ha: Relase Status and types of anti-depressant are associated/dependent

```
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 + (15
```

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