HW3_solution

Problem 1 (a)

```
softdrink = read.table("sofdrink-1.txt", header = TRUE, fill = TRUE)
softdrink$Agent = as.factor(softdrink$Agent)
summary(softdrink)
```

```
##       Time       Agent   observation
##  Min.   : 8.00   1:20   Min.   : 1.00
##  1st Qu.:14.00   2:20   1st Qu.: 5.75
##  Median :21.00   3:20   Median :10.50
##  Mean   :20.75   4:20   Mean   :10.50
##  3rd Qu.:26.25   5:20   3rd Qu.:15.25
##  Max.   :35.00          Max.   :20.00
```

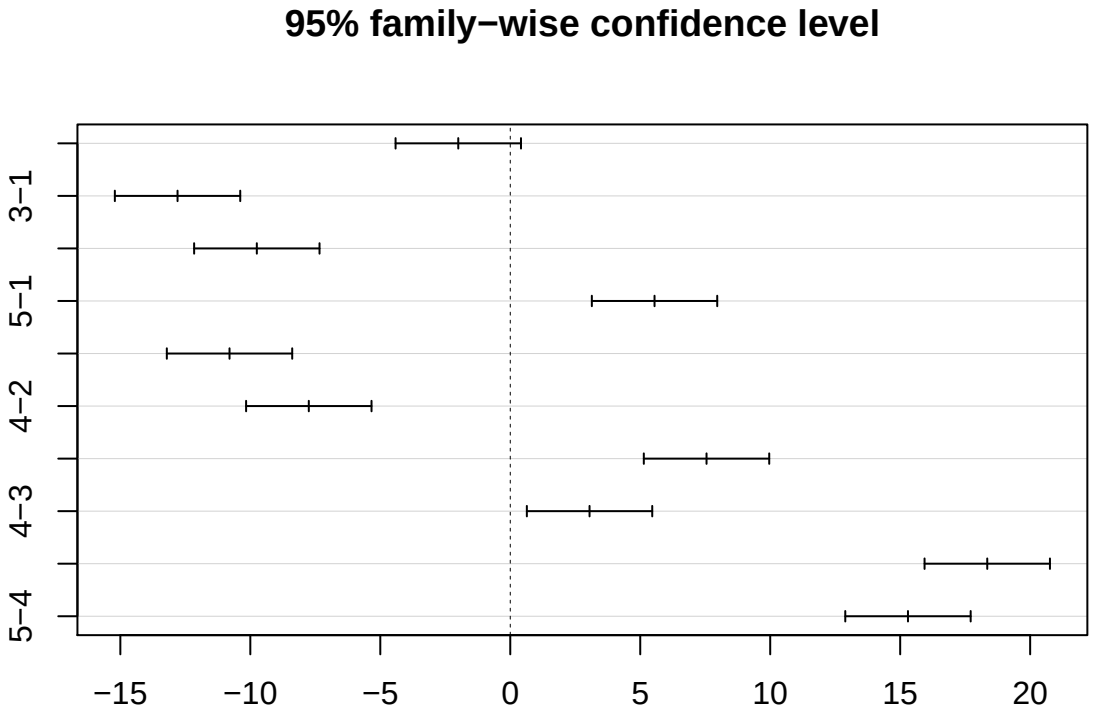
```
fit1 = aov(Time ~ Agent, data = softdrink)
summary(fit1)
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## Agent        4   4430  1107.5   147.2 <2e-16 ***
## Residuals   95    715     7.5
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Since the p-value is less than 0.05, we reject the null hypothesis that the mean time lapse is the same for all five agents.

(b)

```
fit2 = TukeyHSD(fit1,"Agent")
plot(fit2)
```




```
fit2
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = Time ~ Agent, data = softdrink)
##
## $Agent
##       diff         lwr         upr     p adj
## 2-1  -2.00  -4.4119302   0.4119302 0.1520498
## 3-1 -12.80 -15.2119302 -10.3880698 0.0000000
## 4-1  -9.75 -12.1619302  -7.3380698 0.0000000
## 5-1   5.55   3.1380698   7.9619302 0.0000001
## 3-2 -10.80 -13.2119302  -8.3880698 0.0000000
## 4-2  -7.75 -10.1619302  -5.3380698 0.0000000
## 5-2   7.55   5.1380698   9.9619302 0.0000000
## 4-3   3.05   0.6380698   5.4619302 0.0059245
## 5-3  18.35  15.9380698  20.7619302 0.0000000
## 5-4  15.30  12.8880698  17.7119302 0.0000000
```

(c)

```
new_softdrink = rbind(softdrink[softdrink$Agent==1,],softdrink[softdrink$Agent==3,],softdrink[softdrink$
fit3 = pairwise.t.test(new_softdrink$Time, new_softdrink$Agent, p.adj = "bonf")
fit3
```

```
##
```

```
##  Pairwise comparisons using t tests with pooled SD
##
## data:  new_softdrink$Time and new_softdrink$Agent
##
##   1       3
## 3 < 2e-16 -
## 5 8.4e-08 < 2e-16
##
## P value adjustment method: bonferroni
```

(d) The 95% CI is $\hat{L} \pm t_{n-k}(0.5/2) SE(\hat{L}) = 10.275 \pm t_{95}(0.025) \cdot (\sqrt{7.5}\sqrt{\frac{(1/2)^2}{20} \cdot 4}) = 10.275 \pm 1.985 \cdot \sqrt{7.5/20} = [10.275 \pm 1.216]$.

(e) L1: $2 \pm 2.437(\sqrt{7.5}\sqrt{\frac{(1)^2}{20} \cdot 2})$

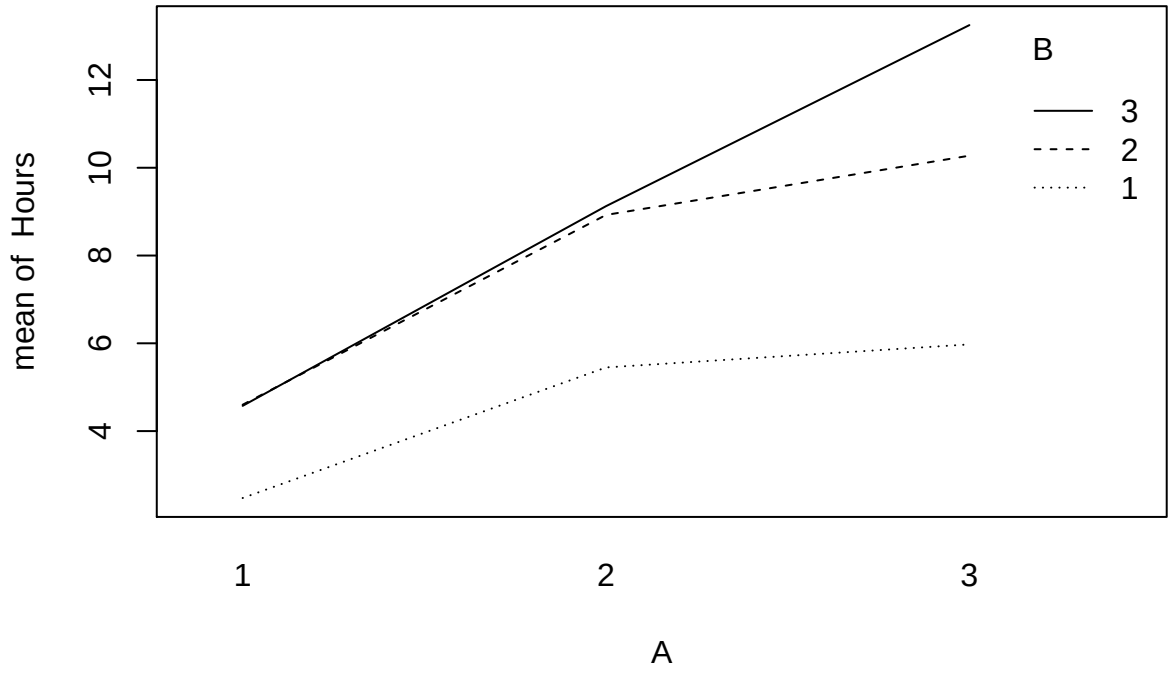
L2: $-6.55 \pm 2.437(\sqrt{7.5}\sqrt{\frac{(1/2)^2}{20} \cdot 2 + \frac{(-1)^2}{20}})$

L3: $-16.825 \pm 2.437(\sqrt{7.5}\sqrt{\frac{(1/2)^2}{20} \cdot 2 + \frac{(-1)^2}{20}})$

(f) $20.75 \pm 1.985(\sqrt{7.5}\sqrt{\frac{(1/4)^2}{20} + \frac{(1/5)^2}{20} \cdot 3 + \frac{(17/20)^2}{20}})$

Problem 2 (a)

```
hayfever = read.table("hayfever.csv", header = TRUE, fill = TRUE, sep=",")
hayfever = hayfever[,1:4]
hayfever$A = as.factor(hayfever$A)
hayfever$B = as.factor(hayfever$B)
attach(hayfever)
interaction.plot(A,B,Hours)
```



```r
summary(aov(Hours~A*B))
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## A            2 220.02  110.01  1827.9 <2e-16 ***
## B            2 123.66   61.83  1027.3 <2e-16 ***
## A:B          4  29.43    7.36   122.2 <2e-16 ***
## Residuals   27   1.63    0.06
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The p-value suggests interaction between A and B.

(b)

```r
summary(hayfever)
```

```
##      Hours         A      B       Observation
##  Min.   : 2.300   1:12   1:12   Min.   :1.00
##  1st Qu.: 4.675   2:12   2:12   1st Qu.:1.75
##  Median : 6.000   3:12   3:12   Median :2.50
##  Mean   : 7.183                 Mean   :2.50
##  3rd Qu.: 9.325                 3rd Qu.:3.25
##  Max.   :13.500                 Max.   :4.00
```

```r
fit4 = aov(Hours~A+B)
summary(fit4)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## A            2 220.02  110.01  109.83 8.51e-15 ***
## B            2 123.66   61.83   61.73 1.55e-11 ***
## Residuals   31  31.05    1.00
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The effects are still present without interaction.

(c)

```r
fit5 = aov(Hours~A+B+A*B)
summary(fit5)
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## A            2 220.02  110.01  1827.9 <2e-16 ***
## B            2 123.66   61.83  1027.3 <2e-16 ***
## A:B          4  29.43    7.36   122.2 <2e-16 ***
## Residuals   27   1.63    0.06
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$9.125 \pm t_{27}(0.025)\sqrt{0.06} = 9.125 \pm 2.05\sqrt{0.06}$

(d) $(4.6 - 2.475) \pm 2.05\sqrt{0.06}\sqrt{2}$