

Homework three

Yi Chen(yc3356)

February 12, 2018

Advanced Data Analysis

Homework Three

Yi Chen, yc3356

problem one

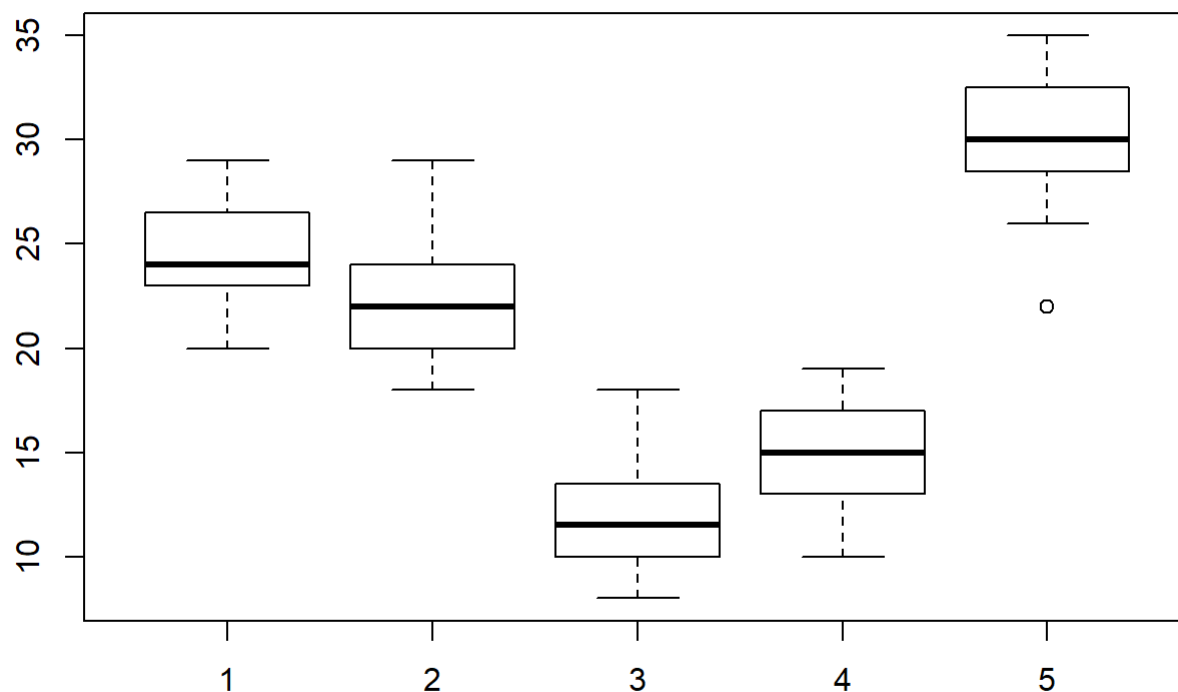
question (a)

Clearly, this is a one-way ANOVA question and the design is balanced (20 observations in each group)

```
# read the data
setwd("C:/Users/cheny/Desktop/study/second term/Advanced Data Analysis/homework/homework three")
data <- read.table("sofdrink.txt", header = TRUE)
```

first, let us analysis the data a little bit.

```
# boxplot
boxplot(data$Time_lapse~data$Agent)
```



Analysis

As we can see from the boxplot, there exist obvious difference among the means from different group. Thus, it is reasonable to assume that the factor agency do have some kind of effect on the time lapse.

Now we plot the one way anova table here:

```
data$Agent <- as.factor(data$Agent)
# anova table
oneway_anove <- aov(formula = Time_lapse~Agent, data = data)
summary(oneway_anove)
```

```
##           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
```

Analysis

According to the result in the table. We can take a F test:

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$$

As we can see, SSR=4430, SSE=715, MSR=1107.5, MSE=7.5, F=147.2

If we take $\alpha = 0.05$, we have $F(0.05, 4, 95) = 2.467494$. Besides here $\Pr(>F) < 0.05$. Thus, we can reject H_0

Conclusion: mean time lapse differ for the five agents.

question (b)

```
# use the Tukey procedure
tk <- TukeyHSD(oneway_anove)
tk
```

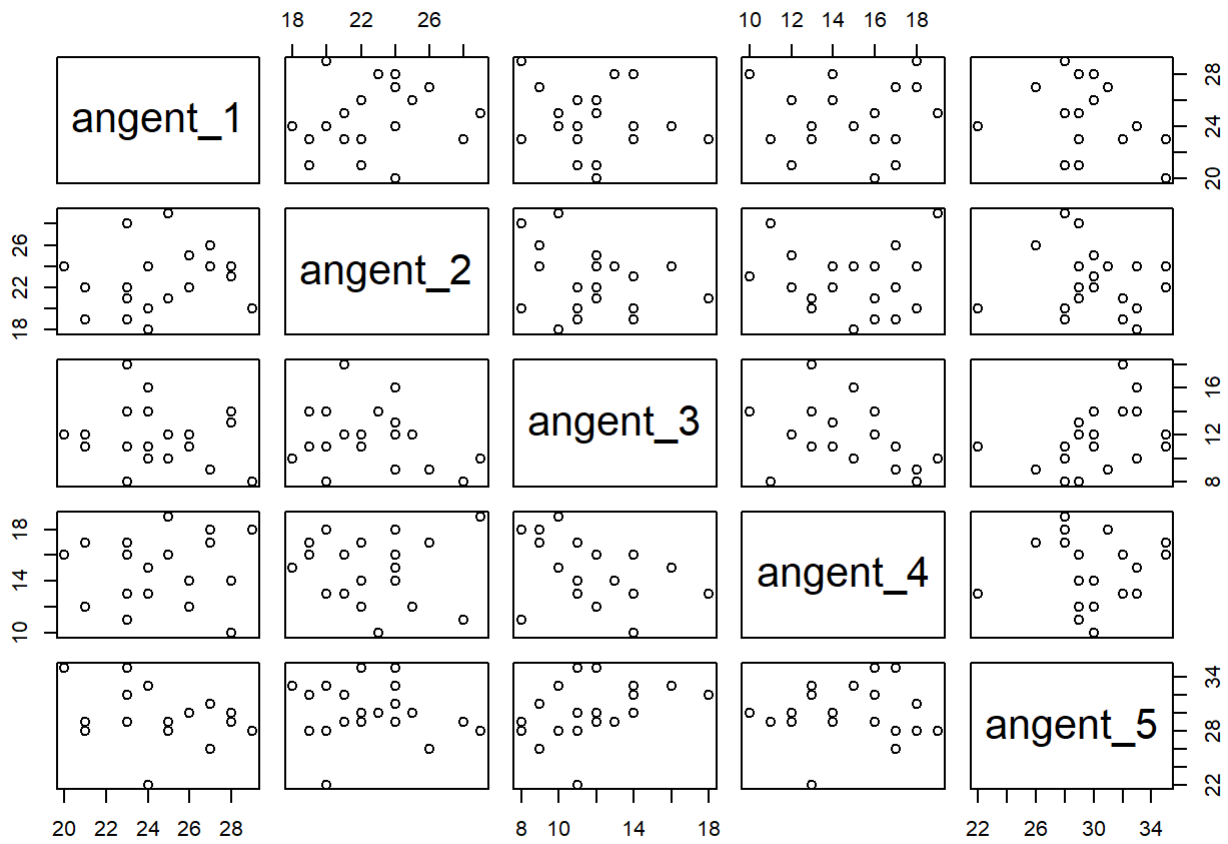
```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = Time_lapse ~ Agent, data = data)
##
## $Agent
##      diff      lwr      upr    p adj
## 2-1  -2.00 -4.4119302  0.4119302 0.1520498
## 3-1 -12.80 -15.2119302 -10.3880698 0.0000000
## 4-1  -9.75 -12.1619302  -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
```

Analysis

As we can see from the result of Tukey procedure. We can conclude that there are groups: Group one: agent 1 , agent 2 Group two: agent 3 Group three: agent 5 Group four: agent 4

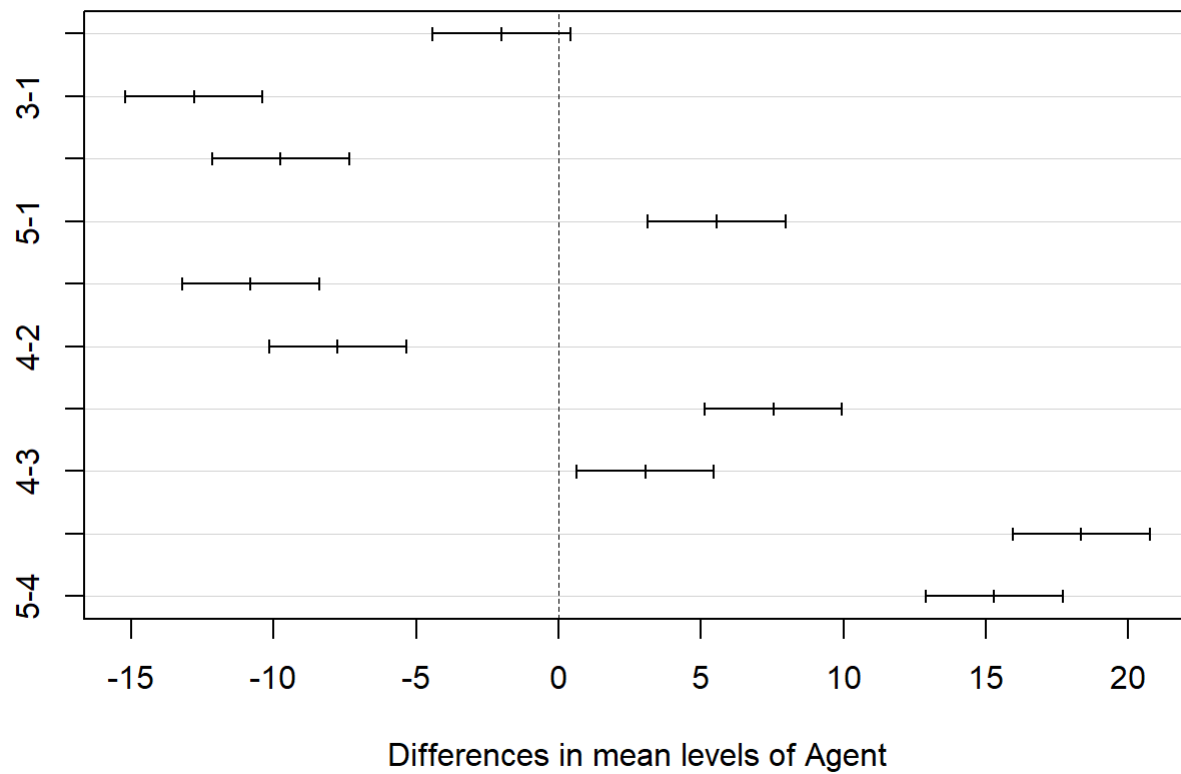
If we just reduce out accuracy a little bit. We can also take agent 3 and agent 4 as a group

```
agent_1 <- data$Time_lapse[data$Agent==1]
agent_2 <- data$Time_lapse[data$Agent==2]
agent_3 <- data$Time_lapse[data$Agent==3]
agent_4 <- data$Time_lapse[data$Agent==4]
agent_5 <- data$Time_lapse[data$Agent==5]
pair_compare <- as.data.frame(cbind(agent_1, agent_2, agent_3, agent_4, agent_5))
pairs(pair_compare[1:5], pch = 21)
```



```
plot(tk)
```

95% family-wise confidence level

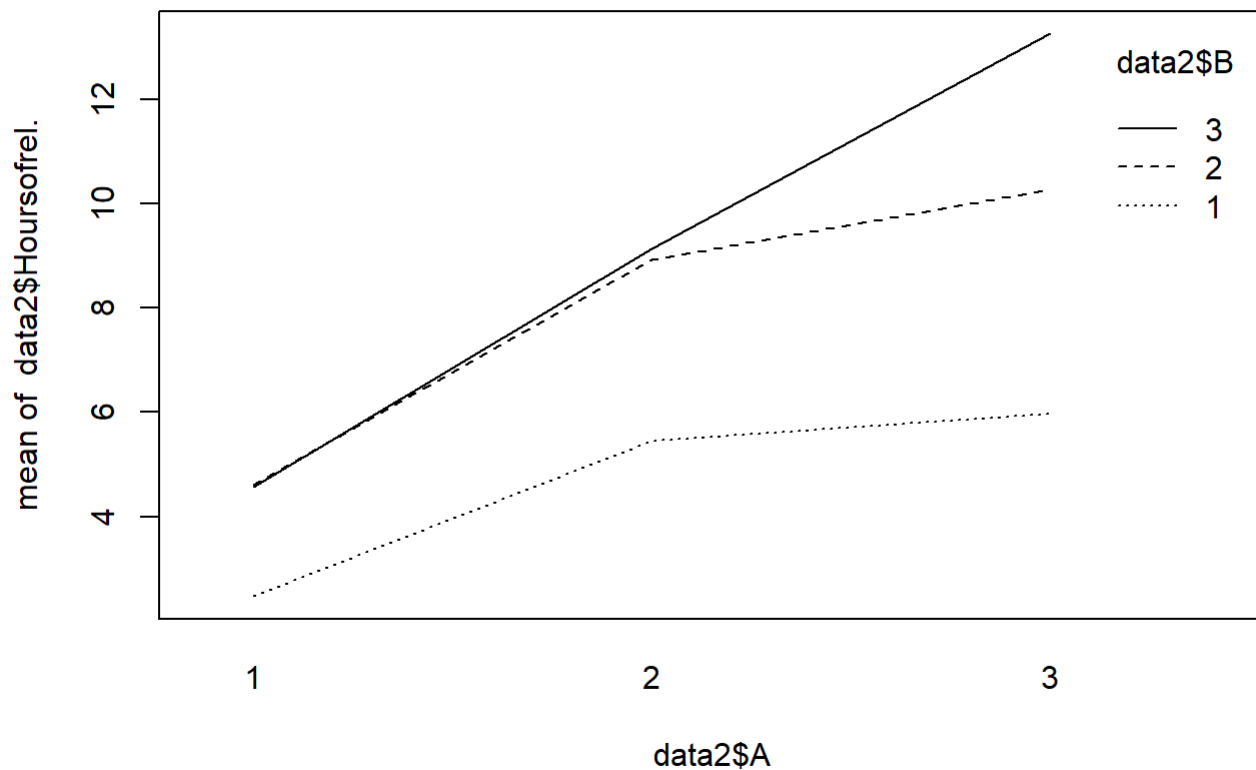


Problem Two

question one

```
data2 <- read.table("Hayfever.txt", header = TRUE)
data2$A <- as.factor(data2$A)
data2$B <- as.factor(data2$B)
```

```
# interaction plot
interaction.plot(data2$A, data2$B, data2$Hoursofrel.)
```



Analysis There is some kind of interaction between factor A and factor B.

```
summary(aov(data2$Hoursofrel. ~ data2$A*data2$B))
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## data2$A      2  220.02   110.01  1827.9 <2e-16 ***
## data2$B      2   123.66    61.83   1027.3 <2e-16 ***
## data2$A:data2$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
```

Analysis

As we can see from the table. The F value for interaction term of factor A and factor B is 122.2. The p value is smaller than 2e-16. Which means that the interaction term indeed effect.