HW8_Report

107023058

Brief Description

The researcher runs an experiment where each of these four alternative media is shown to a different panel of randomly assigned people. Afterwards, viewers were surveyed about their thoughts, including a question (labeled INTEND.0) about their intention to share what they had seen with others

```
# Pre-processing the data
media_1 <- read.csv('D:/Retro/NTHU/課程講義/大三/計算統計於商業分析之應用/
HW8/pls-media/pls-media1.csv',header = T)$INTEND.0
media_2 <- read.csv('D:/Retro/NTHU/課程講義/大三/計算統計於商業分析之應用/
HW8/pls-media/pls-media2.csv',header = T)$INTEND.0
media_3 <- read.csv('D:/Retro/NTHU/課程講義/大三/計算統計於商業分析之應用/
HW8/pls-media/pls-media3.csv',header = T)$INTEND.0
media_4 <- read.csv('D:/Retro/NTHU/課程講義/大三/計算統計於商業分析之應用/
HW8/pls-media/pls-media4.csv',header = T)$INTEND.0

media <- c(media_1,media_2,media_3,media_4)
pls_media <- list(m1 = media_1,m2 = media_2,m3 = media_3,m4 = media_4)
```

Question 1. Describe and visualize the data

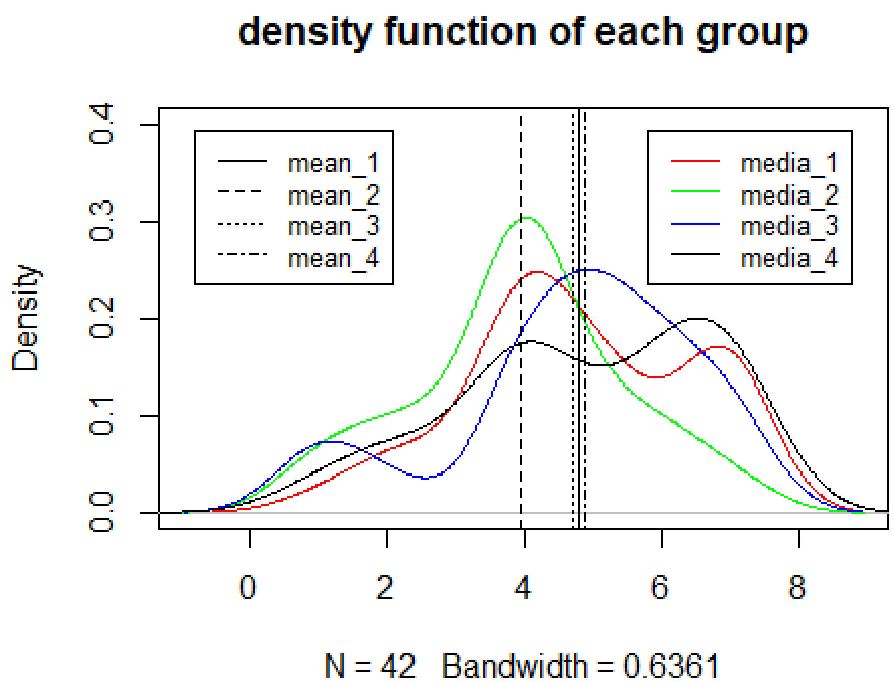
(a). means of viewers intentions to share (INTEND.0) for each media type

```
mean_1 <- mean(media_1)
mean_2 <- mean(media_2)
mean_3 <- mean(media_3)
mean_4 <- mean(media_4)
```

(b). Visualize the distribution and mean of intention to share, across all four media.

```
# first using distribution to visualize them
plot(density(media_1),col = 'red',ylim = c(0,0.4),'density function of
each group')
lines(density(media_2),col = 'green')
lines(density(media_3),col = 'blue')
lines(density(media_4))
# add lines of their means
abline(v = mean(media_1),lty = 1)
abline(v = mean(media_2),lty = 2)
abline(v = mean(media_3),lty = 3)
abline(v = mean(media_4),lty = 4)
```

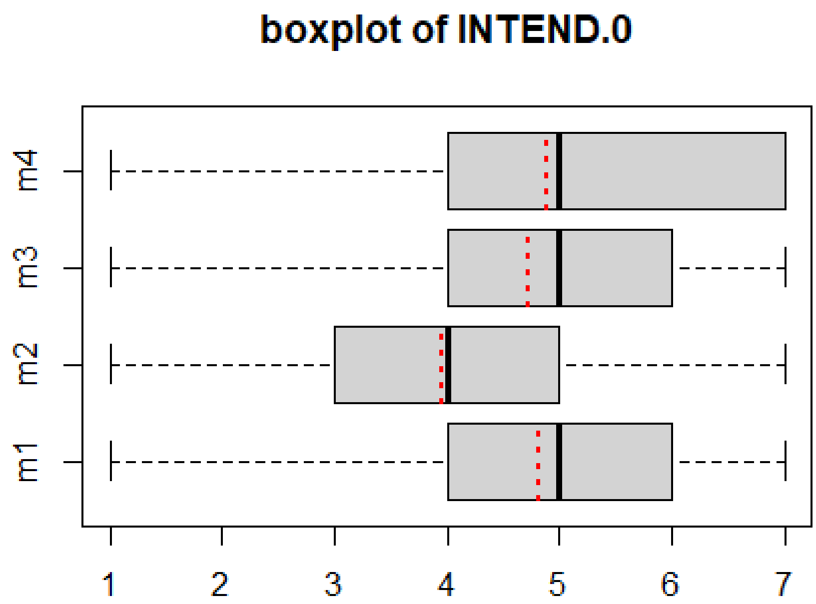
```
# add the labels
legend('topleft',inset=.05,c('mean_1','mean_2','mean_3','mean_4'),lty =
 c(1,2,3,4)
,cex = 0.8)
legend('topright',inset=.05,c('media_1','media_2','media_3','media_4'),
lty = c(1,1,1,1),col=c("red","green","blue","black"),cex = 0.8)
```




```
# then using another method(boxplot) to visualize them
boxplot(pls_media,horizontal = T,main = 'boxplot of INTEND.0')

Q1_mean_segment <- function(n = 1, dataset){
  mean <- mean(dataset)
  segments(x0 = mean, y0 = n-0.4, x1 = mean, y1 = n+0.4, col = rgb(1,0,
0), lwd = 2, lty = "dotted")
}

# add their means onto the graph
Q1_mean_segment(1, media_1)
Q1_mean_segment(2, media_2)
Q1_mean_segment(3, media_3)
Q1_mean_segment(4, media_4)
```



p.s. 黃川(Sorry I can't find out his student ID.) shared his code in the discussion area and it helped me a lot in the boxplot part of my code.

(c). From the visualization alone, do you feel that media type makes a difference on intention to share?

Ans:

In my opinion, the data are quiet the same, only the type of pls_media2 is a little bit different.

Question 2. Traditional way of ANOVA

(a). State the null and alternative hypothesis

Null: the mean of four treatment are the same

Alternative : the mean of four treatment are not the same

(b). Produce the traditional F-statistic for our test

```
all_mean<-mean(media) # overall mean

# sum of squares due to treatments
sstr <- (length(media_1)*(mean(media_1)-all_mean)^2
      + length(media_2)*(mean(media_2)-all_mean)^2
      + length(media_3)*(mean(media_3)-all_mean)^2
      + length(media_4)*(mean(media_4)-all_mean)^2)
```

```r
# sum of squares due to error
sse  <- ((length(media_1)-1)*var(media_1)
      + (length(media_2)-1)*var(media_2)
      + (length(media_3)-1)*var(media_3)
      + (length(media_4)-1)*var(media_4))

# mean square due to treatments
mstr <- sstr/3

# mean square due to treatments
mse <- sse/(length(media)-4)

f_value <- mstr/mse
f_value

## [1] 2.616669
```

(c). Find the 95% and 99% cut-off value of the NULL distribution of F

```r
qf(p = 0.95, df1 = 3,df2 = length(media)-4) # 95% cut off

## [1] 2.660406

qf(p = 0.99, df1 = 3,df2 = length(media)-4) # 99% cut off

## [1] 3.904807
```

(d). According to the traditional ANOVA, do the four types of media produce the same mean intention to share, at 95% confidence? How about at 99% confidence?

As the result above, we can see that the f_value is 2.616669, while the cut-off value of 95% is 2.660406 and 99% is 3.904807. So we can say that under 95% or 99% confidence level, we don't have enough evidence to reject the null hypothesis.

(e). Do you feel the classic requirements of one-way ANOVA are met?

I used the oneway.test() function to examine whether I have calculated is true or not:

```r
# convert the medias data into data frame
media1 <- data.frame(type = rep(1,length(media_1)),intend = media_1)
media2 <- data.frame(type = rep(2,length(media_2)),intend = media_2)
media3 <- data.frame(type = rep(3,length(media_3)),intend = media_3)
media4 <- data.frame(type = rep(4,length(media_4)),intend = media_4)
```

```r
media_all <- rbind(media1,media2,media3,media4)
oneway.test(media_all$intend~factor(media_all$type),var.equal = T)
```

```
##
##  One-way analysis of means
##
## data:  media_all$intend and factor(media_all$type)
## F = 2.6167, num df = 3, denom df = 162, p-value = 0.05289
```

The result is the same as part(b) , so I think it may be true. We still can not reject the null hypothesis.

Question 3. Bootstrapping ANOVA

(a). Bootstrap the null values of F and also the alternative values of the F-statistic.

```r
boot_anova <- function(m1,m2,m3,m4,num){
  null_m1<-sample(m1-mean(m1),replace = T)
  null_m2<-sample(m2-mean(m2),replace = T)
  null_m3<-sample(m3-mean(m3),replace = T)
  null_m4<-sample(m4-mean(m4),replace = T)
  null_m <- c(null_m1,null_m2,null_m3,null_m4)

  alt_m1 <- sample(m1,replace = T)
  alt_m2 <- sample(m2,replace = T)
  alt_m3 <- sample(m3,replace = T)
  alt_m4 <- sample(m4,replace = T)
  alt_m <- c(alt_m1,alt_m2,alt_m3,alt_m4)

  return(c(oneway.test(null_m ~ num,var.equal = T)$statistic,
           oneway.test(alt_m ~ num,var.equal = T)$statistic))
}

boot_f_value <- replicate(5000,boot_anova(media_1,media_2,media_3,media
_4,media_all$type))
boot_f_null <- boot_f_value[1,] # the bootstrapping null f-value
boot_f_alt <- boot_f_value[2,] # the bootstrapping alternative f-value
```

(b). The cutoff values for 95% and 99% confidence

```r
quantile(boot_f_null,0.95) # 95% cut-off
```

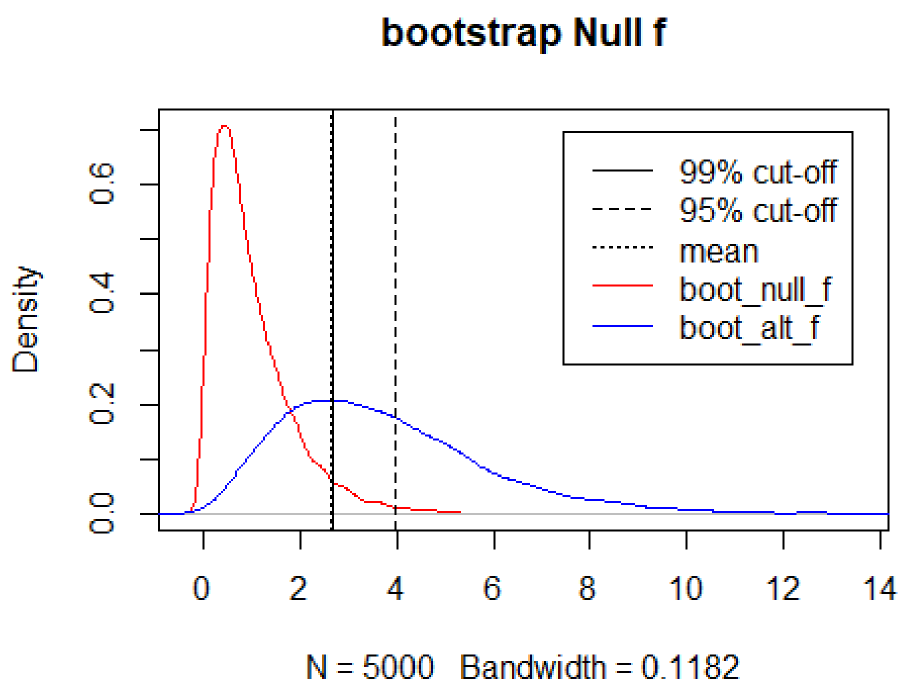
```
##      95%
## 2.654323
```

```r
quantile(boot_f_null,0.99) # 99% cut-off
```

```
##      99%
## 3.984813
```

(c). Visualize the distribution of bootstrapped null values of F, the 95% and 99% cutoff values of F , and also the original F-value from bootstrapped alternative values.

```
plot(density(boot_f_null),col = 'red',main = 'bootstrap Null f')
lines(density(boot_f_alt),col = 'blue')
# add the cut-off lines and f-value
abline(v = quantile(boot_f_null,0.95),lty = 1)
abline(v = quantile(boot_f_null,0.99),lty = 2)
abline(v = f_value,lty = 3)
#add the label
legend('topright',inset=.05,c('99% cut-off','95% cut-off','mean','boot_null_f','boot_alt_f'),lty = c(1,2,3,1,1),col = c('black','black','black','red','blue'))
```



(d). According to the bootstrap, do the four types of media produce the same mean intention to share, at 95% confidence? How about at 99% confidence?

Ans:

From the density distribution above, we can find that the mean of alternative is less than the 95% cut-off line,but larger than the 99% cut-off line. As a result , I think that we can not reject the null hypothesis, which means the four types of media produce the same mean intention to share under 95% confidence level.But we will reject it under 99% confidence level.