106073401

# BACS HW (Week 6)

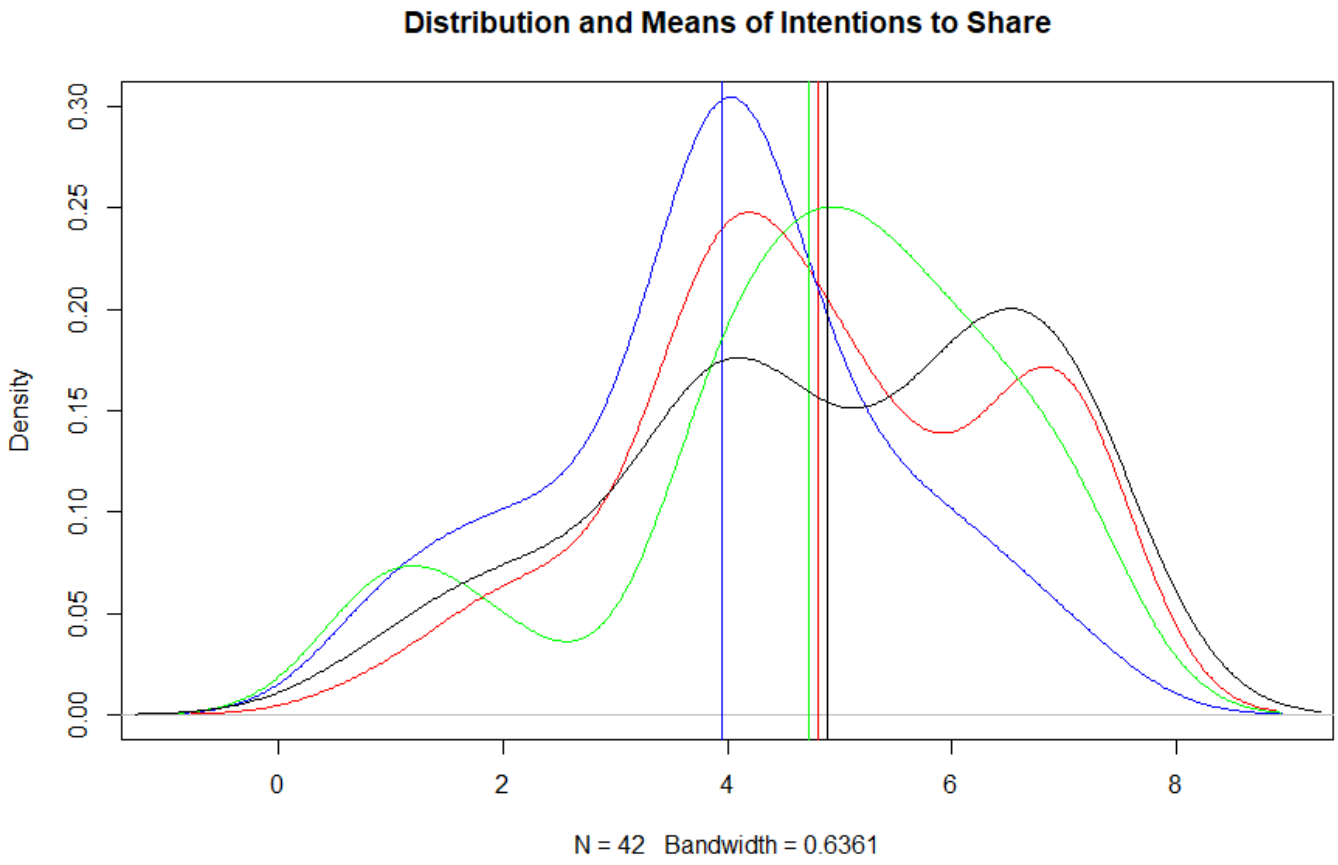
1. Describe and visualize data.

```
> intend_1<-read.csv("pls-media1.csv",header=TRUE)[,2]
> intend_2<-read.csv("pls-media2.csv",header=TRUE)[,2]
> intend_3<-read.csv("pls-media3.csv",header=TRUE)[,2]
> intend_4<-read.csv("pls-media4.csv",header=TRUE)[,2]
```

a) The mean of each data

```
> mean(intend_1)
[1] 4.809524
> mean(intend_2)
[1] 3.947368
> mean(intend_3)
[1] 4.725
> mean(intend_4)
[1] 4.891304
```

b) Visualize the distribution and mean




c) From the graph, the mean of each types of media are very close, it seems that media type do not make difference on intention to share.

106073401

2. One-way ANOVA

a) H0: The differences between the means of each types of media = 0
H1: The differences between the means of each types of media ≠ 0

b) F-statistic = 2.6167

```
> intend<-c(intend_1,intend_2,intend_3,intend_4)
> SSTR<-length(intend_1)*(mean(intend_1)-mean(intend))^2 +
+          length(intend_2)*(mean(intend_2)-mean(intend))^2 +
+          length(intend_3)*(mean(intend_3)-mean(intend))^2 +
+          length(intend_4)*(mean(intend_4)-mean(intend))^2
> MSTR<-SSTR/3
> SSE <- (length(intend_1)-1)*sd(intend_1)^2 +
+     (length(intend_2)-1)*sd(intend_2)^2 +
+     (length(intend_3)-1)*sd(intend_3)^2 +
+     (length(intend_4)-1)*sd(intend_4)^2
> df_MSE<- length(intend)-4
> MSE<-SSE/df_MSE
> f_stat<-MSTR/MSE
> f_stat
[1] 2.616669
```

c) 95% cut-off value=2.6604, 99% cut-off value=3.9048

```
> qf(p=0.95, df1=3, df2=df_MSE)
[1] 2.660406
> qf(p=0.99, df1=3, df2=df_MSE)
[1] 3.904807
```

d) According to traditional ANOVA test, as F is less than both the 95% and 99% cut-off value, we cannot reject the null hypothesis.

```
#Change the data format
> intend_1<-data.frame(media=rep(1,length(intend_1)),intend=intend_1)
> intend_2<-data.frame(media=rep(2,length(intend_2)),intend=intend_2)
> intend_3<-data.frame(media=rep(3,length(intend_3)),intend=intend_3)
> intend_4<-data.frame(media=rep(4,length(intend_4)),intend=intend_4)
intend<-rbind(intend_1,intend_2,intend_3,intend_4)
> oneway.test(intend$intend ~ intend$media,var.equal = TRUE)
        One-way analysis of means

data:  intend$intend and intend$media
F = 2.6167, num df = 3, denom df = 162, p-value = 0.05289
```

e) As the data are not normal distributed, it doesn't meet the requirement of one-way ANOVA.

106073401

3. Bootstrapping ANOVA

   a) Bootstrap the null and alternative hypotheses

```
#Reformat the data as integer
> intend_1<-intend_1$intend
> intend_2<-intend_2$intend
> intend_3<-intend_3$intend
> intend_4<-intend_4$intend
#Bootstrap function
> boot_anova <- function(t1, t2, t3, t4, media_nums) {
+     size1 = length(t1)
+     size2 = length(t2)
+     size3 = length(t3)
+     size4 = length(t4)
+
+     null_grp1 = sample(t1 - mean(t1), size1, replace=TRUE)
+     null_grp2 = sample(t2 - mean(t2), size2, replace=TRUE)
+     null_grp3 = sample(t3 - mean(t3), size3, replace=TRUE)
+     null_grp4 = sample(t4 - mean(t4), size4, replace=TRUE)
+     null_values = c(null_grp1, null_grp2, null_grp3, null_grp4)
+
+     alt_grp1 = sample(t1, size1, replace=TRUE)
+     alt_grp2 = sample(t2, size2, replace=TRUE)
+     alt_grp3 = sample(t3, size3, replace=TRUE)
+     alt_grp4 = sample(t4, size4, replace=TRUE)
+     alt_values = c(alt_grp1, alt_grp2, alt_grp3, alt_grp4)
+
+     c(oneway.test(null_values ~ media_nums,var.equal=TRUE)$statistic,
+       oneway.test(alt_values ~ media_nums, var.equal=TRUE)$statistic)
+ }
> set.seed(99)
> f_values <- replicate(3000, boot_anova(intend_1, intend_2, intend_3, intend_4,
intend$media))
> f_nulls <- f_values[1,]
> f_alts <- f_values[2,]
> mean(f_nulls)
[1] 1.016004
> mean(f_alts)
[1] 3.775775
```

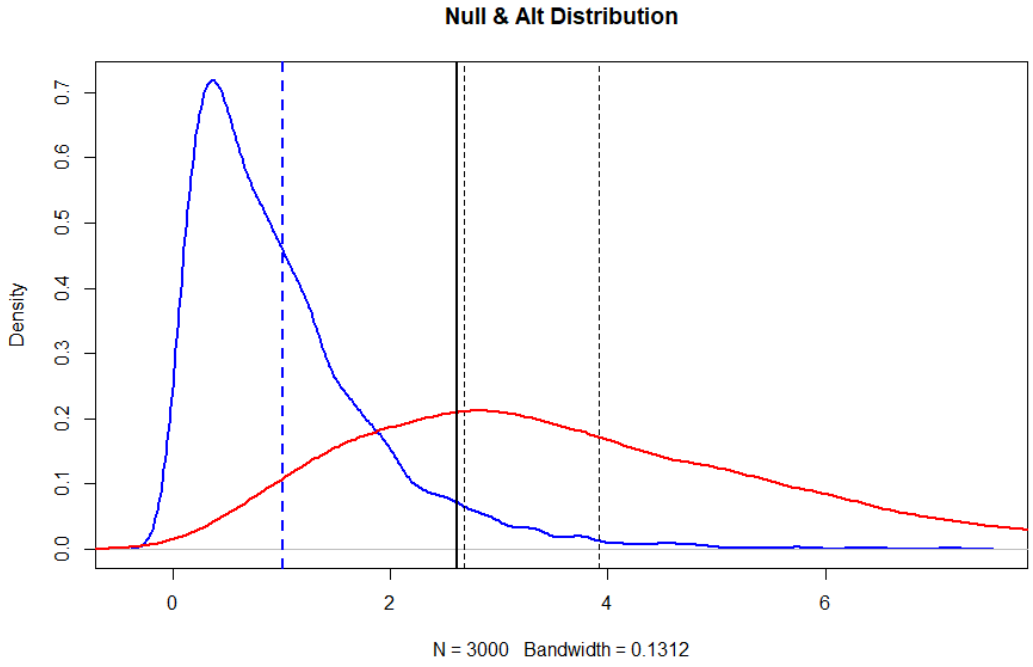
106073401

b) The 95% cut-off value=2.6604, the 99% cut-off value=3.9048.

```
> qf(p=0.95, df1=3, df2=df_MSE)
[1] 2.660406
> qf(p=0.99, df1=3, df2=df_MSE)
[1] 3.904807
```

c) Visualizing the null value, alternative value of F and 95%, 99% cut-off value.

```
> plot(density(f_nulls), col = "blue", lwd=2,main = "Null & Alt Distribution")
> abline(v=quantile(f_nulls,0.95), lty = "dashed")
> abline(v=quantile(f_nulls,0.99), lty = "dashed")
> abline(v=mean(f_nulls),lty="dashed", lwd=2, col="blue")
> lines(density(f_alts), lwd=2, col="red")
> anova <- oneway.test(intend$intend ~ intend$media, var.equal=TRUE)
> abline(v=anova$statistic,lwd=2)
```




d) Due to the result of bootstrapping, the four types of media produce the same mean intention to share at both 95% and 99% confidence.