HW7

109006206

Set Working Directories & Reading Files

```
setwd("/Users/olivia/Documents/Documents/Study/Semester 6/BACS/HW7")
require("ggplot2")
require("tidyverse")
require("FSA")
Media_1 <- read.csv("pls-media1.csv")
Media_2 <- read.csv("pls-media2.csv")
Media_3 <- read.csv("pls-media3.csv")
Media_4 <- read.csv("pls-media4.csv")
```

QUESTION 1

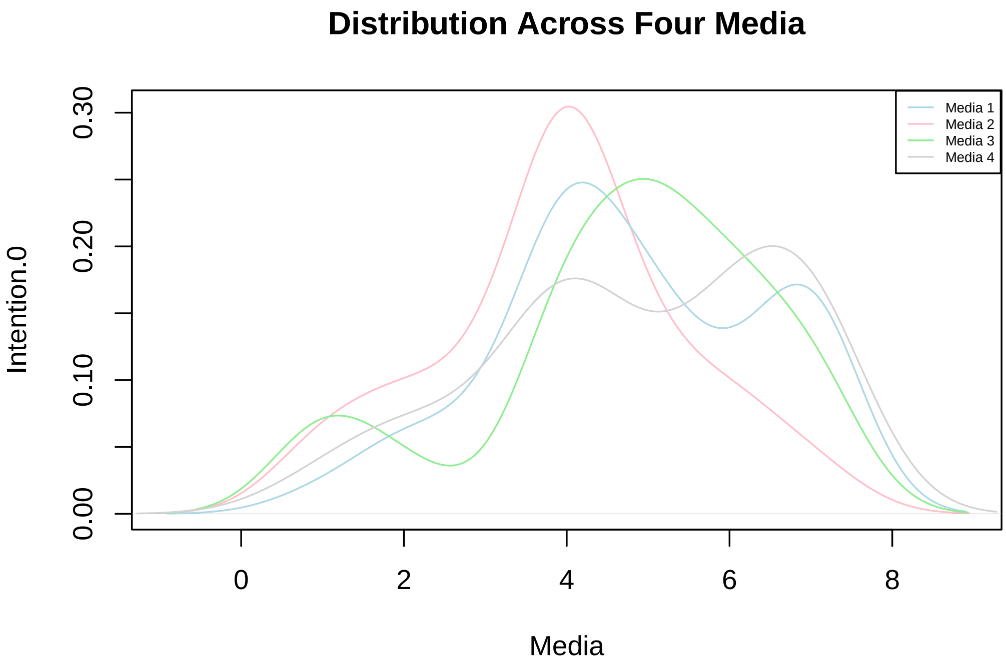
A) What are the means of viewers' intentions to share (INTEND.0) on each of the four media types?

```
m1_mean <- mean(Media_1$INTEND.0)
m2_mean <- mean(Media_2$INTEND.0)
m3_mean <- mean(Media_3$INTEND.0)
m4_mean <- mean(Media_4$INTEND.0)
```

m1_mean : 4.8095238
m2_mean : 3.9473684
m3_mean : 4.725
m4_mean : 4.8913043

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

```
m <-rbind(Media_1,Media_2,Media_3,Media_4)
plot(density(Media_2$INTEND.0),col = 'pink',main="Distribution Across Four Media",xlab="Media",ylab="In
lines(density(Media_1$INTEND.0),col = 'lightblue')
lines(density(Media_3$INTEND.0),col = 'lightgreen')
lines(density(Media_4$INTEND.0),col = 'lightgrey')
legend('topright',inset=0.001,c("Media 1","Media 2","Media 3","Media 4"),lty=1,col=c('lightblue','pink'
```



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

Answer : I think the media have some affect to the intention to share based on the visualization

QUESTION 2

A) State the null and alternative hypotheses when comparing INTEND.0 across four groups in ANOVA

Null Hypothesis: there is no difference in the means of INTEND.0 between the four groups
(H0: μ 1 = μ 2 = μ 3 = μ 4)

Alternative Hypothesis : there is a significant difference in the means of INTEND.0 (Not all means are equal)

B) Let's compute the F-statistic ourselves:

i) Show the code and results of computing MSTR, MSE, and F

```
mbind <- list(Media_1$INTEND.0,Media_2$INTEND.0,Media_3$INTEND.0,Media_4$INTEND.0)
k <- length(mbind)
n <- 4
grandmean<- mean(sapply(mbind,mean))
#MSTR
SSTR<-function(){
  (nrow(Media_1) * (mean(Media_1$INTEND.0) - grandmean)^2)+
  (nrow(Media_2) * (mean(Media_2$INTEND.0) - grandmean)^2)+
  (nrow(Media_3) * (mean(Media_3$INTEND.0) - grandmean)^2)+
  (nrow(Media_4) * (mean(Media_4$INTEND.0) - grandmean)^2)
}
df_mstr<-n-1
mstr<-SSTR()/df_mstr
mstr
```

```
## [1] 7.53239
```

```
#MSE
SSE<-function(){
    ((nrow(Media_1)-1) * (sd(Media_1$INTEND.0)^2))+
    ((nrow(Media_2)-1) * (sd(Media_2$INTEND.0)^2))+
    ((nrow(Media_3)-1) * (sd(Media_3$INTEND.0)^2))+
    ((nrow(Media_4)-1) * (sd(Media_4$INTEND.0)^2))
}
nt <- nrow(Media_1)+nrow(Media_2)+nrow(Media_3)+nrow(Media_4)
df_mse <- nt - k
mse <- SSE()/df_mse
mse
```

```
## [1] 2.869151
```

```
f_value <-mstr/mse
f_value
```

```
## [1] 2.625303
```

MSTR : 7.5323896
MSE : 2.8691509
F : 2.6253027

ii) Compute the p-value of F, from the null F-distribution; is the F-value significant? If so, state your conclusion for the hypotheses.

```
p_value<-pf(f_value, df_mstr, df_mse, lower.tail=FALSE)
```

Since the p-value is 0.0523069 so we can conclude that the F-value is not significant

C) Conduct the same one-way ANOVA using the aov() function in R – confirm that you got similar results.

```
m_new1 <- rbind(Media_1,Media_2,Media_3,Media_4)
m_new1 <- m_new1[1:2]
oneway.test(m_new1$INTEND.0 ~ factor(m_new1$media),var.equal = TRUE)
```

```
##
##  One-way analysis of means
##
## data:  m_new1$INTEND.0 and factor(m_new1$media)
## F = 2.6167, num df = 3, denom df = 162, p-value = 0.05289
```

```
anova_model <- aov(m_new1$INTEND.0 ~ factor(m_new1$media))
summary(anova_model)
```

```
##                       Df Sum Sq Mean Sq F value Pr(>F)
## factor(m_new1$media)   3   22.5   7.508   2.617 0.0529 .
## Residuals            162  464.8   2.869
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

D) Regardless of your conclusions, conduct a post-hoc Tukey test (feel free to use the TukeyHSD() function included in base R) to see if any pairs of media have significantly different means – what do you find?

```
TukeyHSD(anova_model, conf.level = 0.05)
```

```
##   Tukey multiple comparisons of means
##     5% family-wise confidence level
##
```

```
## Fit: aov(formula = m_new1$INTEND.0 ~ factor(m_new1$media))
##
## $`factor(m_new1$media)`
##            diff         lwr        upr     p adj
## 2-1 -0.86215539 -1.06562977 -0.6586810 0.1085727
## 3-1 -0.08452381 -0.28530983  0.1162622 0.9959223
## 4-1  0.08178054 -0.11218249  0.2757436 0.9959032
## 3-2  0.77763158  0.57175512  0.9835080 0.1825044
## 4-2  0.94393593  0.74470805  1.1431638 0.0573229
## 4-3  0.16630435 -0.03017708  0.3627858 0.9687417
```

Answer : As we can see from the data we could say there is no different means at $\alpha = .05$

E) Do you feel the classic requirements of one-way ANOVA were met?

1. The variance of the response variables is the same for all treatments/populations

Answer : As we can see from the data below we could conclude that there are some differences in the variance

```
Variance<- c(var(Media_1$INTEND.0),
var(Media_2$INTEND.0),
var(Media_3$INTEND.0),
var(Media_4$INTEND.0))
Variance
```

```
## [1] 2.694541 2.321479 3.076282 3.299034
```

2. Each treatment/population's response variable is normally distributed

```
shapiro.test(Media_1$INTEND.0)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  Media_1$INTEND.0
## W = 0.91279, p-value = 0.003557
```

```
shapiro.test(Media_2$INTEND.0)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  Media_2$INTEND.0
## W = 0.92974, p-value = 0.01969
```

```
shapiro.test(Media_3$INTEND.0)
```

```
##
##  Shapiro-Wilk normality test
```

```
##
## data:  Media_3$INTEND.0
## W = 0.88247, p-value = 0.0006139
```

```
shapiro.test(Media_4$INTEND.0)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  Media_4$INTEND.0
## W = 0.89611, p-value = 0.0006242
```

Answer : As we can see from the result the second data p-value is bigger than 0.05 which indicates that the data is normally distributed.

In conclusion, the classic requirements of one-way ANOVA were not met.

QUESTION 3

A) State the null and alternative hypotheses

Null Hypothesis : All media would give similar value if randomly drawn from them
(H0: μ 1 = μ 2 = μ 3 = μ 4)
Alternative Hypothesis : At least one media would give a larger value than another if randomly drawn

B) Let's compute (an approximate) Kruskal Wallis H ourselves (use the formula we saw in class or another formula might have found at a reputable website/book):

i) Show the code and results of computing H

```
mkw <- m[1:2]
ranks <- rank(mkw$INTEND.0)
N<-length(ranks)
group_ranks <- split(ranks,f = mkw$media)

sum_ranks <-sapply(group_ranks,sum)
length_ranks<- (sapply(group_ranks,length))
Total<- sum((sum_ranks^2)/length_ranks)

H<- (((12/(N*(N+1))) * Total) - 3*(N+1))
H
```

```
## [1] 8.45466
```

H : 8.4546598

ii) Compute the p-value of H, from the null chi-square distribution; is the H value significant? If so, state your conclusion of the hypotheses.

```
kw_p<-1 -pchisq(H, df=k-1)
kw_p
```

```
## [1] 0.03749292
```

P-Value of H : 0.0374929
Answer : In conclusion, we should reject the null hypothesis since the p-value < 0.05

C) Conduct the same test using the kruskal.wallis() function in R – confirm that you got similar results.

```
kruskal.test(mkw$INTEND.0 ~ mkw$media, mkw)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  mkw$INTEND.0 by mkw$media
## Kruskal-Wallis chi-squared = 8.8283, df = 3, p-value = 0.03166
```

D) Regardless of your conclusions, conduct a post-hoc Dunn test to see if the values of any pairs of media are significantly different – what are your conclusions?

```
dunnTest(INTEND.0~media, data = mkw , method = "bonferroni")
```

```
## Warning: media was coerced to a factor.
```

```
##   Comparison           Z     P.unadj      P.adj
## 1      1 - 2  2.30087819 0.021398517 0.12839110
## 2      1 - 3 -0.09233644 0.926430736 1.00000000
## 3      2 - 3 -2.36408588 0.018074622 0.10844773
## 4      1 - 4 -0.31452459 0.753122646 1.00000000
## 5      2 - 4 -2.65613380 0.007904225 0.04742535
## 6      3 - 4 -0.21613379 0.828883460 1.00000000
```

Answer : At $\alpha = .05$, we can conclude that media 2 and media 4 are statistically significantly different from each other (P.adj = 0.0474)