### Student ID: 112077423

```r
library(ggplot2)
library(compstatslib)
library(tidyr)
library(FSA)
```

```
## Warning:     'FSA'          R     4.3.3
```

```r
alpha <- 0.05

df1 <- read.csv('pls-media1.csv')
df2 <- read.csv('pls-media2.csv')
df3 <- read.csv('pls-media3.csv')
df4 <- read.csv('pls-media4.csv')

tmp <- list(df1, df2, df3, df4)
```

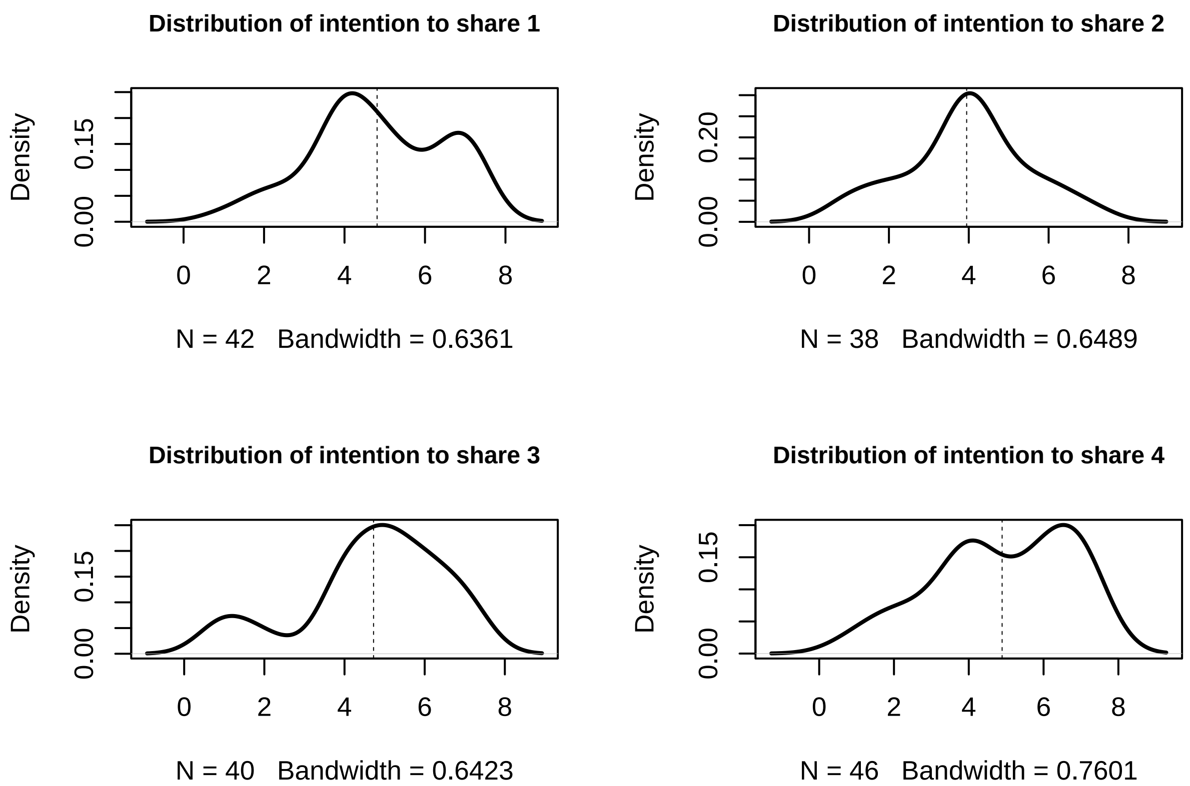
### Question 1

```r
par(mfrow=c(2,2))
means <- c()
cnt <- 1
for (i in tmp){
  means <- append(means, mean(i$INTEND.0))
  ttl <- paste('Distribution of intention to share', cnt)
  plot(density(i$INTEND.0), lwd=2, main=ttl, cex.main=0.9)
  abline(v=tail(means, n=1), lty="dashed",lw=0.5)
  out <- paste('Mean of intention to share', cnt, '->', lapply(tail(means, n=1), round, 2))
  print(out)
  cnt <- cnt + 1
}
```

```
## [1] "Mean of intention to share 1 -> 4.81"

## [1] "Mean of intention to share 2 -> 3.95"

## [1] "Mean of intention to share 3 -> 4.72"
```




```
## [1] "Mean of intention to share 4 -> 4.89"
```

Looking at the visualization, I believe media type does make a difference in the intention to share. I think the most successful is the 4th type.

## Question 2

**(a)**

H0: the means of intention to share for given 4 datasets are the same

Ha: the means of intention to share for given 4 datasets are the different

**(b)**

```
media1 <- df1$INTEND.0
media2 <- df2$INTEND.0
media3 <- df3$INTEND.0
media4 <- df4$INTEND.0

mean_of_means <- mean(means)
stds <- c(sd(media1), sd(media2), sd(media3), sd(media4))
n <- c(length(media1), length(media2), length(media3), length(media4))

mstr <- (sum(n * (means - mean_of_means)^2)) / 3
mse <- (sum((n-1)*stds^2)) / (sum(n) - 4)
```

```r
F_stat <- mstr / mse

out1 <- paste('MSTR:', round(mstr,1))
out2 <- paste('MSE', round(mse,1))
out3 <- paste('F:', round(F_stat,2))
cat(out1, out2, out3, sep='\n')
```

```
## MSTR: 7.5
## MSE 2.9
## F: 2.63
```

```r
critical_val <- qf(p=0.95, df1=3, df2=(sum(n) - 4))
p_value <- pf(F_stat, 3, (sum(n) - 4), lower.tail=FALSE)
out1 <- paste('p_value:', p_value)
out2 <- paste('critical value:', round(critical_val,2))
cat(out1, out2, sep='\n')
```

```
## p_value: 0.0523068574612358
## critical value: 2.66
```

Since the p-value (0.052) is larger than the level of significance (0.05), we fail to reject the null hypothesis. Besides, since the F statistic is not higher than the critical value, then the difference among groups is not statistically significant

**(c)**

```r
# to avoid recycling
length(media1) <- 46
length(media2) <- 46
length(media3) <- 46
# transform into long format
df <- data.frame(cbind(media1, media2, media3, media4))
df_long <- gather(df, key = 'type', value = 'score', na.rm = TRUE)
dim(df_long)
```

```
## [1] 166   2
```

```r
anova_model <- aov(df_long$score ~ factor(df_long$type))
summary(anova_model)
```

```
##                       Df Sum Sq Mean Sq F value Pr(>F)
## factor(df_long$type)   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
```

The results are similar to the above.

**(d)**

```r
TukeyHSD(anova_model, conf.level = 0.95)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = df_long$score ~ factor(df_long$type))
##
## $`factor(df_long$type)`
##                      diff         lwr       upr     p adj
## media2-media1 -0.86215539 -1.84660332 0.1222925 0.1085727
## media3-media1 -0.08452381 -1.05596494 0.8869173 0.9959223
## media4-media1  0.08178054 -0.85664966 1.0202107 0.9959032
## media3-media2  0.77763158 -0.21843807 1.7737012 0.1825044
## media4-media2  0.94393593 -0.01996662 1.9078385 0.0573229
## media4-media3  0.16630435 -0.78431033 1.1169190 0.9687417
```

We can see from the output that there is no statistically significant difference between the mean of intention to share for each media type at the 0.05 significance level.

**(e)**

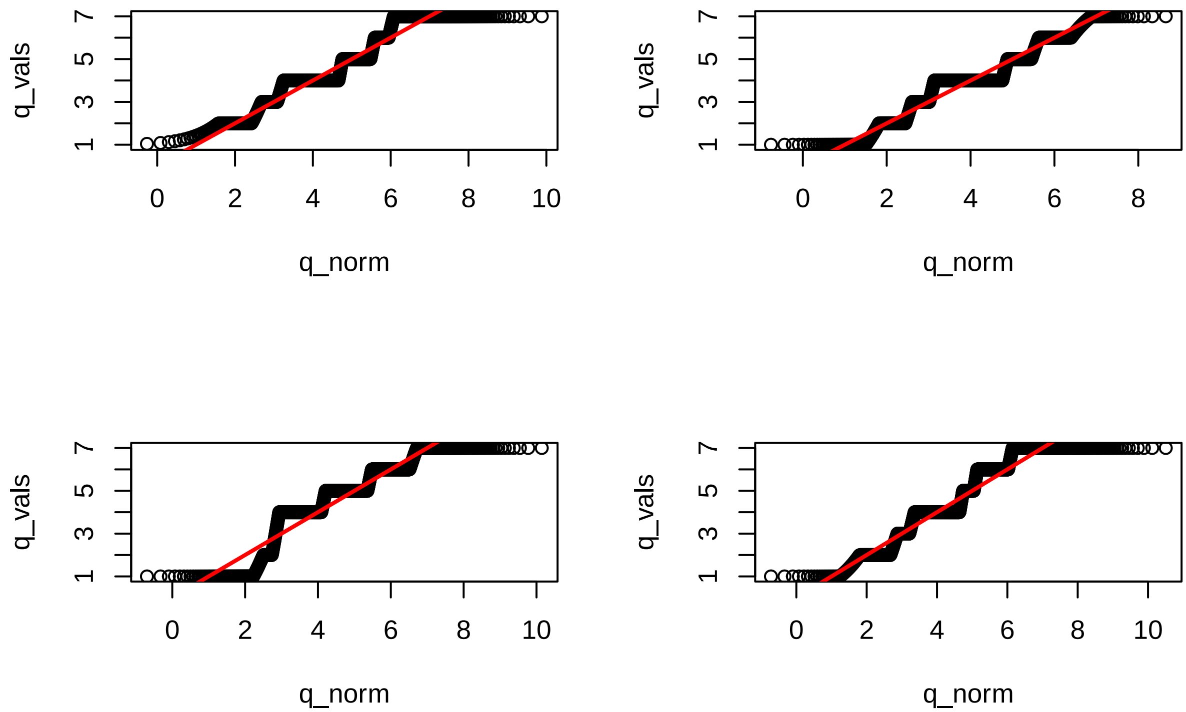
```r
variance <- c(var(df1$INTEND.0), var(df2$INTEND.0), var(df3$INTEND.0), var(df4$INTEND.0))
cat(variance)
```

```
## 2.694541 2.321479 3.076282 3.299034
```

```r
par(mfrow=c(2,2))

norm_qq_plot <- function(values) {
  probs1000 <- seq(0, 1, 0.001)
  q_vals <- quantile(values, probs1000)
  q_norm <- qnorm(probs1000, mean=mean(values), sd=sd(values))
  plot(q_norm, q_vals)
  abline(a=0, b=1, col="red", lwd=2)
}

norm_qq_plot(df1$INTEND.0)
norm_qq_plot(df2$INTEND.0)
norm_qq_plot(df3$INTEND.0)
norm_qq_plot(df4$INTEND.0)
```




As we can see, each intention to share a variable is not perfectly normally distributed. Also, the variance of the variables is not really the same for all media types. The observations are probably independent. However, ANOVA is robust to minor violations of normality and variances, so classic requirements of one-way ANOVA were met.

## Question 3

**(a)**

H0: All groups would give similar value if randomly drawn from them

Ha: At least one group would give a larger value than another if randomly drawn

**(b)**

```r
ranks <- rank(df_long$score)
groups <- split(ranks, as.factor(df_long$type))
sums <- sapply(groups, sum)

H <- (12/(sum(n)*(sum(n)+1))) * (sum(sums^2/n)) - 3 * (sum(n) + 1)
out1 <- paste('H value:', round(H, 2))
k <- 4
kw_p <- 1 - pchisq(H, df = k-1)
out2 <- paste('p value:', kw_p)
cat(out1, out2, sep='\n')
```

```
## H value: 8.45
```

```
## p value: 0.037492918119218
```

Since the p-value (0.037) is smaller than the level of significance (0.05), H value is significant, we reject the null hypothesis.

**(c)**

```r
kruskal.test(score ~ type, data = df_long)
```

```
## 
##  Kruskal-Wallis rank sum test
## 
## data:  score by type
## Kruskal-Wallis chi-squared = 8.8283, df = 3, p-value = 0.03166
```

The results are similar to the above.

**(d)**

```r
dunnTest(score ~ as.factor(type), data = df_long, method = "bonferroni")
```

```
##        Comparison           Z     P.unadj      P.adj
## 1 media1 - media2  2.30087819 0.021398517 0.12839110
## 2 media1 - media3 -0.09233644 0.926430736 1.00000000
## 3 media2 - media3 -2.36408588 0.018074622 0.10844773
## 4 media1 - media4 -0.31452459 0.753122646 1.00000000
## 5 media2 - media4 -2.65613380 0.007904225 0.04742535
## 6 media3 - media4 -0.21613379 0.828883460 1.00000000
```

At alpha = .05, media2 and media4 are the only two medias that are statistically significantly different from each other.