BACS_Week8_106071041

106071041

2021/4/15

Question 1

```
# read data
experiment_data1 <- read.csv("pls-media1.csv")
experiment_data2 <- read.csv("pls-media2.csv")
experiment_data3 <- read.csv("pls-media3.csv")
experiment_data4 <- read.csv("pls-media4.csv")
```

```
# select INTEND0
INTEND0_1 <- experiment_data1$INTEND.0
INTEND0_2 <- experiment_data2$INTEND.0
INTEND0_3 <- experiment_data3$INTEND.0
INTEND0_4 <- experiment_data4$INTEND.0
```

a. 4 means

```
INTEND0_means <- sapply(list(INTEND0_1, INTEND0_2, INTEND0_3, INTEND0_4), mean)
```

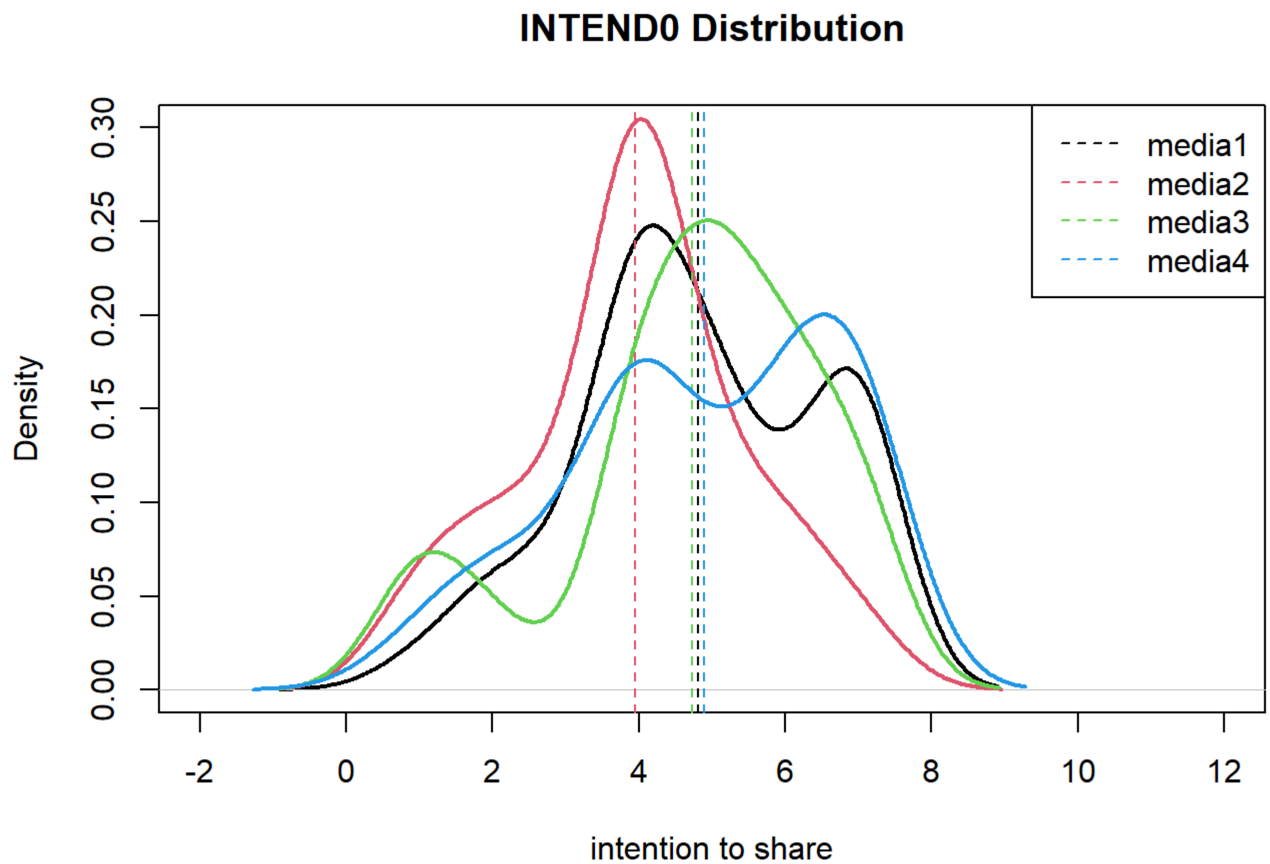
```
names(INTEND0_means) <- c("media1", "media2", "media3", "media4")
```

```
INTEND0_means
```

```
##   media1   media2   media3   media4
## 4.809524 3.947368 4.725000 4.891304
```

b. plot distributions and means

```
plot(density(INTEND0_1), xlim = c(-2, 12), ylim = c(0, 0.3), main = "INTEND0 Distribution", lwd
 = 2, xlab = "intention to share")
abline(v = INTEND0_means[1], col = c(1), lty = "dashed")
lines(density(INTEND0_2), col = c(2), lwd = 2)
abline(v = INTEND0_means[2], col = c(2), lty = "dashed")
lines(density(INTEND0_3), col = c(3), lwd = 2)
abline(v = INTEND0_means[3], col = c(3), lty = "dashed")
lines(density(INTEND0_4), col = c(4), lwd = 2)
abline(v = INTEND0_means[4], col = c(4), lty = "dashed")
legend("topright", legend = c("media1", "media2", "media3", "media4"), col = c(1, 2, 3, 4), lty
 = c(2))
```



c. Do you feel that media type makes a difference on intention to share?

Yes, some media types got a lower score apparently(ex.media2, the red one)

Question 2

a. ANOVA Hypothesis

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

$$H_a : the\ means\ are\ not\ the\ same$$

b. F-statistic

```
F_statistic <- function(type1, type2){
  max(sd(type1), sd(type2))^2/min(sd(type1), sd(type2))^2
}
```

```
F_statistic(INTEND0_1, INTEND0_2)
```

```
## [1] 1.1607
```

```
F_statistic(INTEND0_1, INTEND0_3)
```

```
## [1] 1.141672
```

```
F_statistic(INTEND0_1, INTEND0_4)
```

```
## [1] 1.22434
```

```
F_statistic(INTEND0_2, INTEND0_3)
```

```
## [1] 1.325139
```

```
F_statistic(INTEND0_2, INTEND0_4)
```

```
## [1] 1.421091
```

```
F_statistic(INTEND0_3, INTEND0_4)
```

```
## [1] 1.072409
```

```
F_statistic_collection <- c(F_statistic(INTEND0_1, INTEND0_2),
F_statistic(INTEND0_1, INTEND0_3),
F_statistic(INTEND0_1, INTEND0_4),
F_statistic(INTEND0_2, INTEND0_3),
F_statistic(INTEND0_2, INTEND0_4),
F_statistic(INTEND0_3, INTEND0_4))
```

c. cut-off values for F

95% confidence

```
qf(p=0.95, df1 = length(INTEND0_1)-1, df2 = length(INTEND0_2)-1)
```

```
## [1] 1.712939
```

```
qf(p=0.95, df1 = length(INTEND0_1)-1, df2 = length(INTEND0_3)-1)
```

```
## [1] 1.696483
```

```
qf(p=0.95, df1 = length(INTEND0_1)-1, df2 = length(INTEND0_4)-1)
```

```
## [1] 1.655941
```

```
qf(p=0.95, df1 = length(INTEND0_2)-1, df2 = length(INTEND0_3)-1)
```

```
## [1] 1.713212
```

```
qf(p=0.95, df1 = length(INTEND0_2)-1, df2 = length(INTEND0_4)-1)
```

```
## [1] 1.673096
```

```
qf(p=0.95, df1 = length(INTEND0_3)-1, df2 = length(INTEND0_4)-1)
```

```
## [1] 1.66413
```

```
cutoff_95_collection <- c(qf(p=0.95, df1 = length(INTEND0_1)-1, df2 = length(INTEND0_2)-1),
qf(p=0.95, df1 = length(INTEND0_1)-1, df2 = length(INTEND0_3)-1),
qf(p=0.95, df1 = length(INTEND0_1)-1, df2 = length(INTEND0_4)-1),
qf(p=0.95, df1 = length(INTEND0_2)-1, df2 = length(INTEND0_3)-1),
qf(p=0.95, df1 = length(INTEND0_2)-1, df2 = length(INTEND0_4)-1),
qf(p=0.95, df1 = length(INTEND0_3)-1, df2 = length(INTEND0_4)-1))
```

99% confidence

```
qf(p=0.99, df1 = length(INTEND0_1)-1, df2 = length(INTEND0_2)-1)
```

```
## [1] 2.151877
```

```
qf(p=0.99, df1 = length(INTEND0_1)-1, df2 = length(INTEND0_3)-1)
```

```
## [1] 2.121492
```

```
qf(p=0.99, df1 = length(INTEND0_1)-1, df2 = length(INTEND0_4)-1)
```

```
## [1] 2.047335
```

```
qf(p=0.99, df1 = length(INTEND0_2)-1, df2 = length(INTEND0_3)-1)
```

```
## [1] 2.1503
```

```
qf(p=0.99, df1 = length(INTEND0_2)-1, df2 = length(INTEND0_4)-1)
```

```
## [1] 2.076582
```

```
qf(p=0.99, df1 = length(INTEND0_3)-1, df2 = length(INTEND0_4)-1)
```

```
## [1] 2.061283
```

```
cutoff_99_collection <- c(qf(p=0.99, df1 = length(INTEND0_1)-1, df2 = length(INTEND0_2)-1),
qf(p=0.99, df1 = length(INTEND0_1)-1, df2 = length(INTEND0_3)-1),
qf(p=0.99, df1 = length(INTEND0_1)-1, df2 = length(INTEND0_4)-1),
qf(p=0.99, df1 = length(INTEND0_2)-1, df2 = length(INTEND0_3)-1),
qf(p=0.99, df1 = length(INTEND0_2)-1, df2 = length(INTEND0_4)-1),
qf(p=0.99, df1 = length(INTEND0_3)-1, df2 = length(INTEND0_4)-1))
```

d. ANOVA

they produce the same means both at 95% and 99% confidence

```
media1 <- data.frame(media = rep(1, length(INTEND0_1)), intention = INTEND0_1)
media2 <- data.frame(media = rep(2, length(INTEND0_2)), intention = INTEND0_2)
media3 <- data.frame(media = rep(3, length(INTEND0_3)), intention = INTEND0_3)
media4 <- data.frame(media = rep(4, length(INTEND0_4)), intention = INTEND0_4)
```

```
media <- rbind(media1, media2, media3, media4)
```

```
my_F <- oneway.test(media$intention ~ media$media, var.equal = TRUE, )
my_F
```

```
##
##  One-way analysis of means
##
## data:  media$intention and media$media
## F = 2.6167, num df = 3, denom df = 162, p-value = 0.05289
```

```
summary(aov(media$intention ~ factor(media$media)))
```

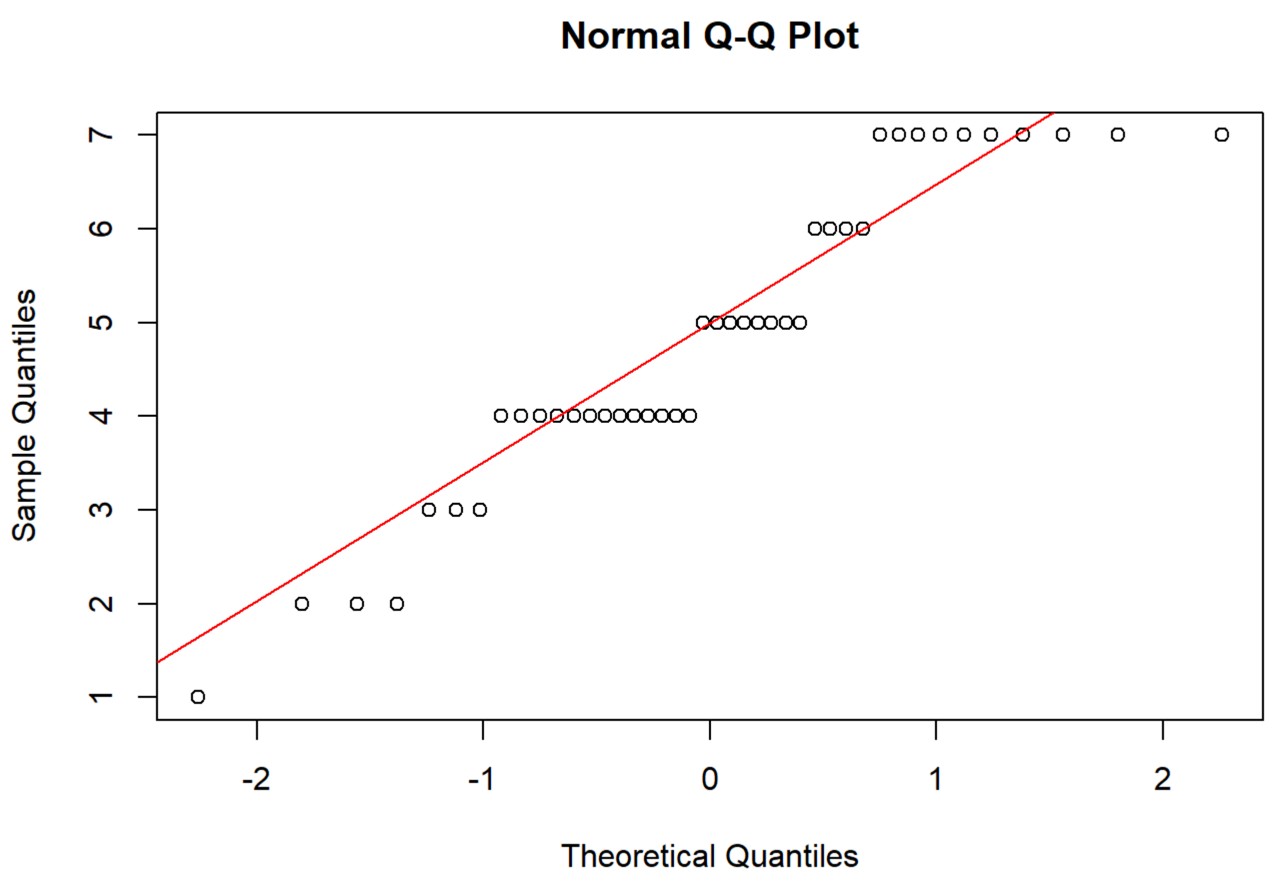
```
##                      Df Sum Sq Mean Sq F value Pr(>F)
## factor(media$media)   3   22.5   7.508   2.617 0.0529 .
## Residuals           162  464.8   2.869
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

e. the classic requirements of one-way ANOVA are met?

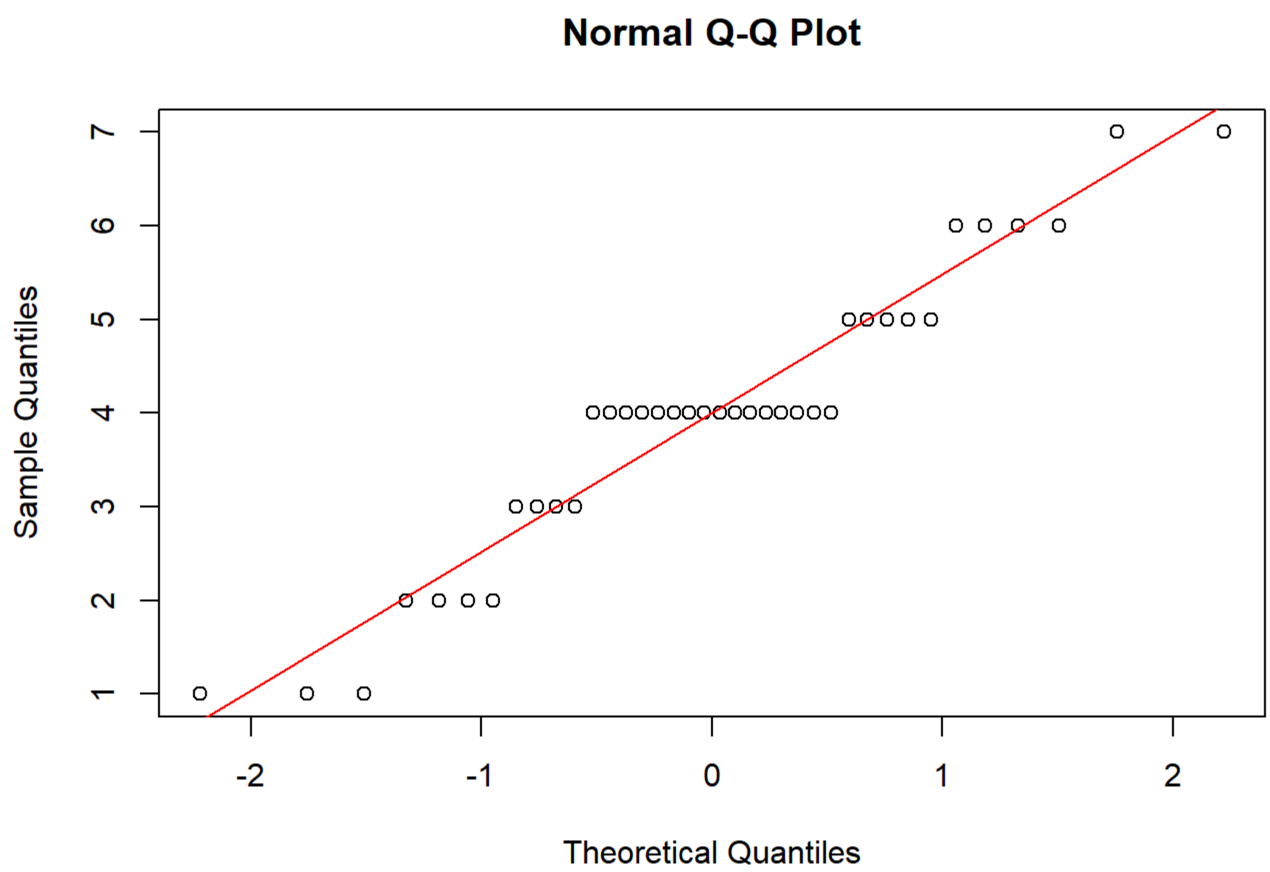
Except for requirement 1, normally distributed, requirement 2 and requirement 3 are met, respectively same variances and independent observations.

Requirement 1: normally distributed (not met)

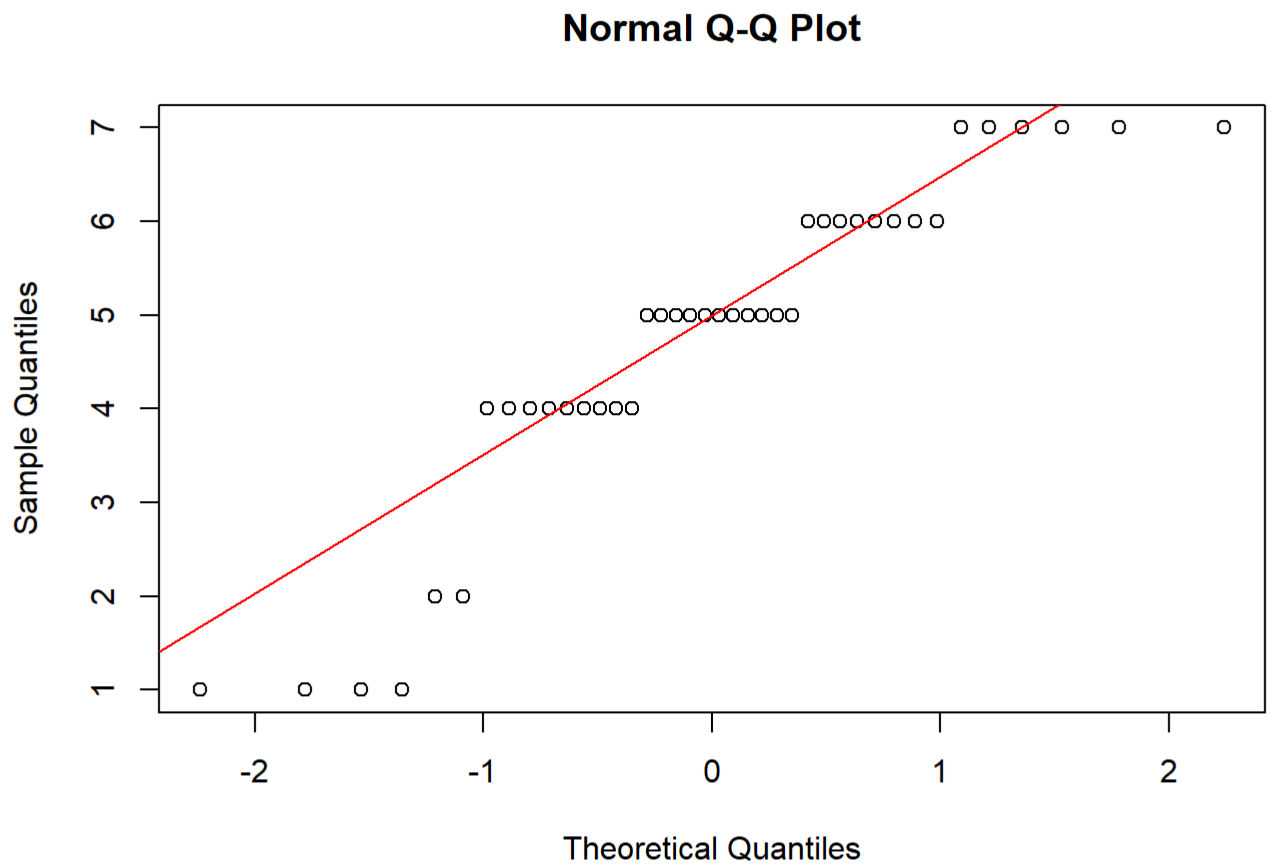
```
# Requirement 1: normally distributed
qqnorm(INTEND0_1)
qqline(INTEND0_1, col = "red")
```



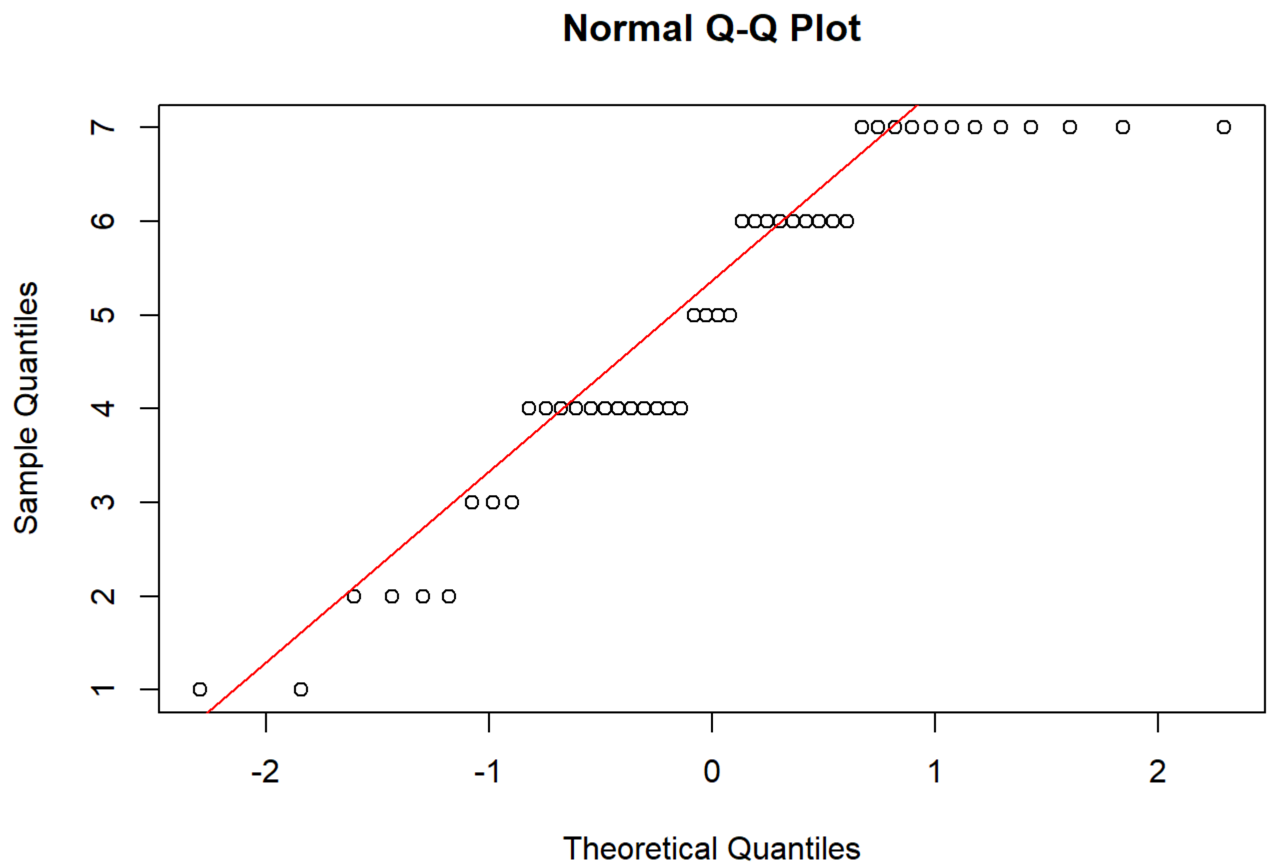

```
qqnorm(INTEND0_2)
qqline(INTEND0_2, col = "red")
```




```
qqnorm(INTEND0_3)
qqline(INTEND0_3, col = "red")
```




```
qqnorm(INTEND0_4)
qqline(INTEND0_4, col = "red")
```



Requirement 2: The variances are the same (met)

```
# 95%
ifelse(all(F_statistic_collection < cutoff_95_collection)==T, "same variances at 95% confidence"
,  "different variances at 95% confidence")
```

```
## [1] "same variances at 95% confidence"
```

```
# 99%
ifelse(all(F_statistic_collection < cutoff_99_collection)==T,  "same variances at 99% confidenc
e",  "different variances at 99% confidence")
```

```
## [1] "same variances at 99% confidence"
```

Requirement 3: The observations are independent(met)
Intuitively, people will feedback as they want since there's no extra interference that can possibly occur in the experiment.

Question 3

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

```
boot_anova <- function(m1, m2, m3, m4, media_num){
  # Null: resample from mean-centered treatments
  null_m1 = sample(m1 - mean(m1), replace = TRUE)
  null_m2 = sample(m2 - mean(m2), replace = TRUE)
  null_m3 = sample(m3 - mean(m3), replace = TRUE)
  null_m4 = sample(m4 - mean(m4), replace = TRUE)
  null_values = c(null_m1, null_m2, null_m3, null_m4)

  # Alternative: resample from the actual treatments
  alt_m1  = sample(m1, replace = TRUE)
  alt_m2  = sample(m2, replace = TRUE)
  alt_m3  = sample(m3, replace = TRUE)
  alt_m4  = sample(m4, replace = TRUE)
  alt_values  = c(alt_m1, alt_m2, alt_m3, alt_m4)

  # ANOVA for null and alternative hypothesis
  c(oneway.test(null_values~ media_num, var.equal = TRUE)$statistic, oneway.test(alt_values~ med
ia_num, var.equal = TRUE)$statistic)
}
```

```
f_values <- replicate(5000, boot_anova(INTEND0_1, INTEND0_2, INTEND0_3, INTEND0_4, media$media))
f_nulls <- f_values[1,]
f_alts <- f_values[2,]
```

b. From the bootstrapped null values of F, What are the cutoff values for 95% and 99% confidence?

```
# 95%
boot_95_cutoff <- quantile(f_nulls, 0.95)
boot_95_cutoff
```

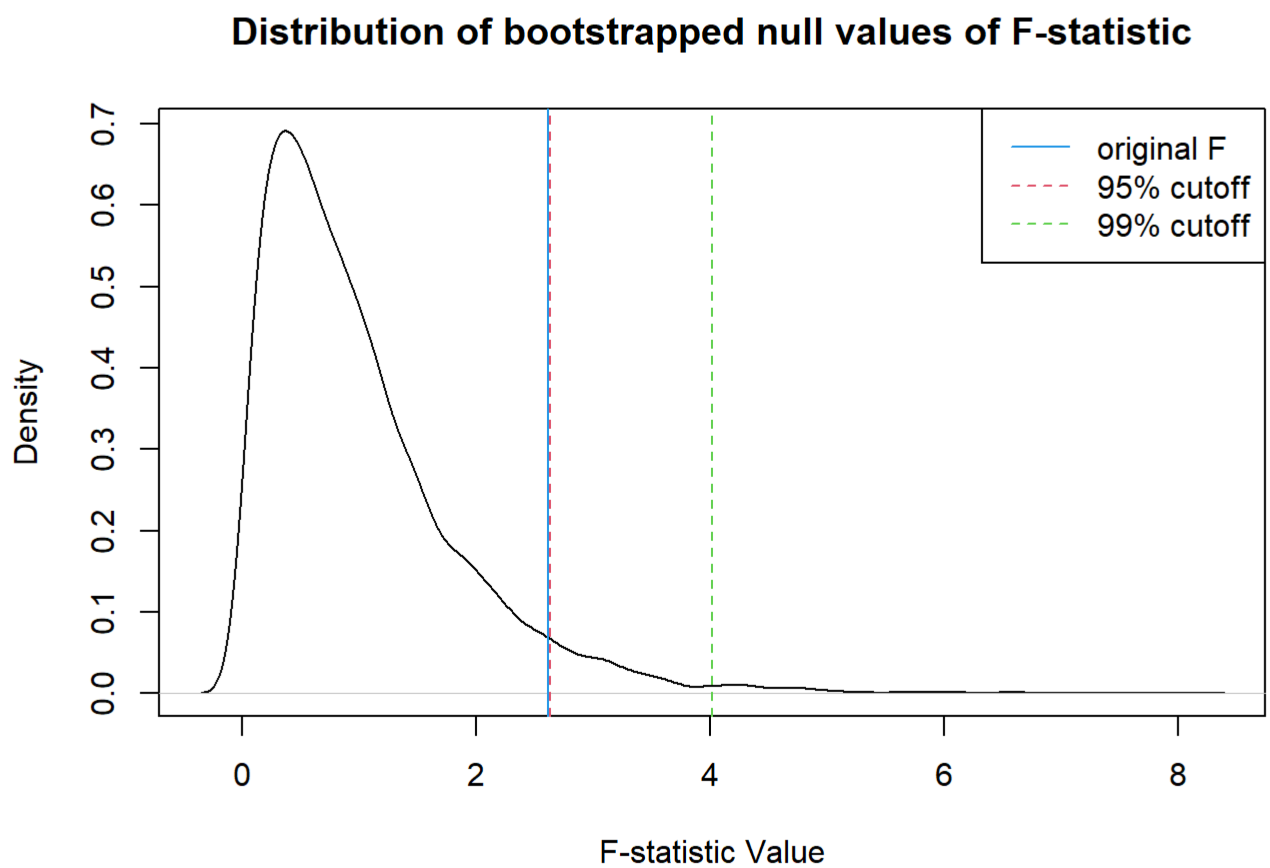
```
##      95%
## 2.636648
```

```
# 99%
boot_99_cutoff <- quantile(f_nulls, 0.99)
boot_99_cutoff
```

```
##      99%
## 4.014838
```

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(f_nulls), main = "Distribution of bootstrapped null values of F-statistic", xlab =
"F-statistic Value")
abline(v = boot_95_cutoff, col = 2, lty = 2)
abline(v = boot_99_cutoff, col = 3, lty = 2)
abline(v = my_F$statistic, col = 4, lty = 1)
legend("topright", legend = c("original F","95% cutoff", "99% cutoff"), lty = c(1, 2, 2), col =
 c(4, 2, 3))
```



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?

At both confidence level, they have the same intention to share.