BACS HW (Week8)

106070038

2021-04-18

Special thanks to 106002103 for teaching me how to draw a plot by "ggplot2".

Question 1

a.What are the means of viewers intentions to share (INTEND.0) for each media type? (report four means)

```
# install.packages('readr')
library(readr)
media1 <- read_csv("pls-media1.csv")
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   .default = col_double()
## )
## i Use `spec()` for the full column specifications.
```

```
# sapply(media1, mean)
media1_intend0 <- media1$INTEND.0
mean(media1_intend0)
```

```
## [1] 4.809524
```

```
media2 <- read_csv("pls-media2.csv")
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   .default = col_double()
## )
## i Use `spec()` for the full column specifications.
```

```
# sapply(media2, mean)
media2_intend0 <- media2$INTEND.0
mean(media2_intend0)
```

```
## [1] 3.947368
```

```
media3 <- read_csv("pls-media3.csv")
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   .default = col_double()
## )
## i Use `spec()` for the full column specifications.
```

```
# sapply(media3, mean)
media3_intend0 <- media3$INTEND.0
mean(media3_intend0)
```

```
## [1] 4.725
```

```
media4 <- read_csv("pls-media4.csv")
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   .default = col_double()
## )
## i Use `spec()` for the full column specifications.
```

```
# sapply(media4, mean)
media4_intend0 <- media4$INTEND.0
mean(media4_intend0)
```

```
## [1] 4.891304
```

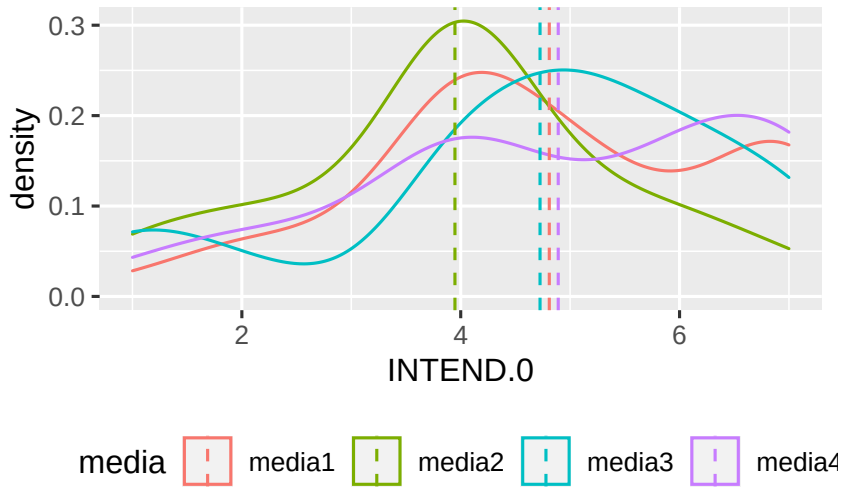
b. Visualize the distribution and mean of intention to share, across all four media.

(Your choice of data visualization; Try to put them all on the same plot and make it look sensible)

```
media1_ <- data.frame(media = rep("media1", length(media1$INTEND.0)), INTEND.0 = media1$INTEND.0)
media2_ <- data.frame(media = rep("media2", length(media2$INTEND.0)), INTEND.0 = media2$INTEND.0)
media3_ <- data.frame(media = rep("media3", length(media3$INTEND.0)), INTEND.0 = media3$INTEND.0)
media4_ <- data.frame(media = rep("media4", length(media4$INTEND.0)), INTEND.0 = media4$INTEND.0)

library(ggplot2)
library(plyr)
media_df <- rbind(media1_, media2_, media3_, media4_)
mu <- ddply(media_df, "media", summarise, grp.mean = mean(INTEND.0))
p <- ggplot(media_df, aes(x=INTEND.0, color = media)) +
            geom_density() +
            geom_vline(data = mu, aes(xintercept = grp.mean, color = as.factor(media)),
            linetype = "dashed") +
```

```
  theme(legend.position = "bottom")
p
```



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

- Yes, the distribution of media1, media2, media3, media4 look different.

Question 2 one-way ANOVA

a.Hnull, Halt

- $Hnull : \mu 1 = \mu 2 = \mu 3 = \mu 4$
- $Halt : themeansarenotsame$

b. traditional F-statistic

```
media_c <- cbind(media1$INTEND.0, media2$INTEND.0, media3$INTEND.0, media4$INTEND.0)
```

```
## Warning in cbind(media1$INTEND.0, media2$INTEND.0, media3$INTEND.0,
## media4$INTEND.0): number of rows of result is not a multiple of vector length
## (arg 1)
```

```
media_df <- data.frame(media1 = media_c[,1], media2 = media_c[,2], media3 = media_c[,3], media4 = media
```

```
n <- nrow(media_df)
sstr <- n*sum((sapply(media_df, mean) - mean(sapply(media_df, mean)))^2)
df_mstr <- 4-1
mstr <- sstr/df_mstr
```

```
sse <- sum((n-1)*sapply(media_df,var))
```

```r
df_mse <- 46*4 - 4
mse <- sse/df_mse

f_value <- mstr/mse
qf(p=0.95, df1=df_mstr, df2=df_mse)
```

```
## [1] 2.654792
```

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

- sstr = 20.32065, df_mstr = 3, mse = 504.7609, df_mse = 180
- f_value = 2.415479
- qf(p=0.95, df1=df_mstr, df2=df_mse) = 2.654792
- Reject H0.

c. cut-off values of F for 95% and 99% confidence

```r
qf(p=0.95, df1=df_mstr, df2=df_mse)
```

```
## [1] 2.654792
```

```r
qf(p=0.99, df1=df_mstr, df2=df_mse)
```

```
## [1] 3.892266
```

- 95%: 2.654792
- 99%: 3.892266

d. traditional ANOVA

```r
id1 <- data.frame(strategy=rep(1,length(media1$INTEND.0)), intend=media1$INTEND.0)
id2 <- data.frame(strategy=rep(2,length(media2$INTEND.0)), intend=media2$INTEND.0)
id3 <- data.frame(strategy=rep(3,length(media3$INTEND.0)), intend=media3$INTEND.0)
id4 <- data.frame(strategy=rep(4,length(media4$INTEND.0)), intend=media4$INTEND.0)
ids <- rbind(id1, id2, id3, id4)
oneway.test(ids$intend ~ ids$strategy, var.equal=TRUE)
```

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

- One-way analysis of means => F = 2.6167
- 95%: Reject.
- 99%: Not reject.

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

- classic requirements:
  - 1. Each treatment/population's response variable is normally distributed => NO
  - 2. The variance (s2) of the response variables is the same for all treatments/populations => NO
  - 3. The observations are independent: the response variables are not related => YES

Question 3 bootstrapping ANOVA

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

```
boot_anova <- function(t1, t2, t3, t4, treat_nums) {
  null_grp1 = sample(t1 - mean(t1), replace=TRUE)
  null_grp2 = sample(t2 - mean(t2), replace=TRUE)
  null_grp3 = sample(t3 - mean(t3), replace=TRUE)
  null_grp4 = sample(t4 - mean(t4), replace=TRUE)
  null_values = c(null_grp1, null_grp2, null_grp3, null_grp4)

    alt_grp1 = sample(t1, replace=TRUE)
    alt_grp2 = sample(t2, replace=TRUE)
    alt_grp3 = sample(t3, replace=TRUE)
    alt_grp4 = sample(t4, replace=TRUE)

    alt_values = c(alt_grp1, alt_grp2, alt_grp3, alt_grp4)

  c(oneway.test(null_values ~ treat_nums, var.equal=TRUE)$statistic,
    oneway.test(alt_values ~ treat_nums, var.equal=TRUE)$statistic)
}
f_values <- replicate(5000, boot_anova(id1$intend, id2$intend, id3$intend, id4$intend, ids$strategy))


f_nulls <- f_values[1,]
f_alts <- f_values[2,]

mean(f_nulls)
```

```
## [1] 1.011835
```

```
quantile(f_nulls, 0.95)
```

```
##     95%
## 2.66349
```

```r
mean(f_alts)
```

```
## [1] 3.780794
```

b. the cutoff values for 95% and 99% confidence

```r
qf(p=0.95, df1=df_mstr, df2=df_mse)
```

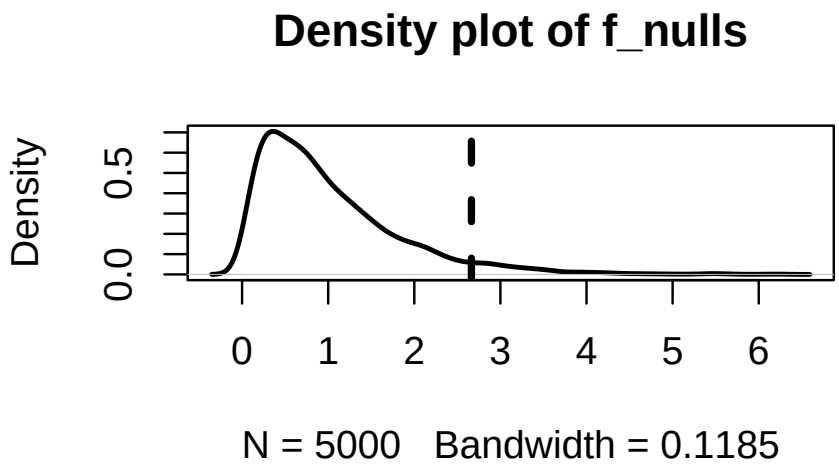
```
## [1] 2.654792
```

```r
qf(p=0.99, df1=df_mstr, df2=df_mse)
```

```
## [1] 3.892266
```

c. Visualize the distribution of bootstrapped null values of F

```r
plot(density(f_nulls), lwd = 2, main = "Density plot of f_nulls") + abline(v = quantile(f_nulls, 0.95),
```




```
## integer(0)
```

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?

- 95%: Reject, the four types of media does not produce the same mean.
- 99%: Not reject, the four types of media produce the same mean.