hw6

Bijon Setyawan Raya

4/17/2021

Data Preparation

```
m1 <- read.csv("/home/johnbjohn/Documents/git_repos/bacs-hw/hw6/pls-media1.csv")
m2 <- read.csv("/home/johnbjohn/Documents/git_repos/bacs-hw/hw6/pls-media2.csv")
m3 <- read.csv("/home/johnbjohn/Documents/git_repos/bacs-hw/hw6/pls-media3.csv")
m4 <- read.csv("/home/johnbjohn/Documents/git_repos/bacs-hw/hw6/pls-media4.csv")
media <- rbind(m1,m2,m3,m4)
```

Question 1) Let's describe and visualize the data:

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

Here is the mean of `INTEND.0` in each data set: (starting from media1 all the way to media4)

1. `media1$INTEND.0` mean: 4.8095238
2. `media2$INTEND.0` mean: 3.9473684
3. `media3$INTEND.0` mean: 4.725
4. `media4$INTEND.0` mean: 4.8913043

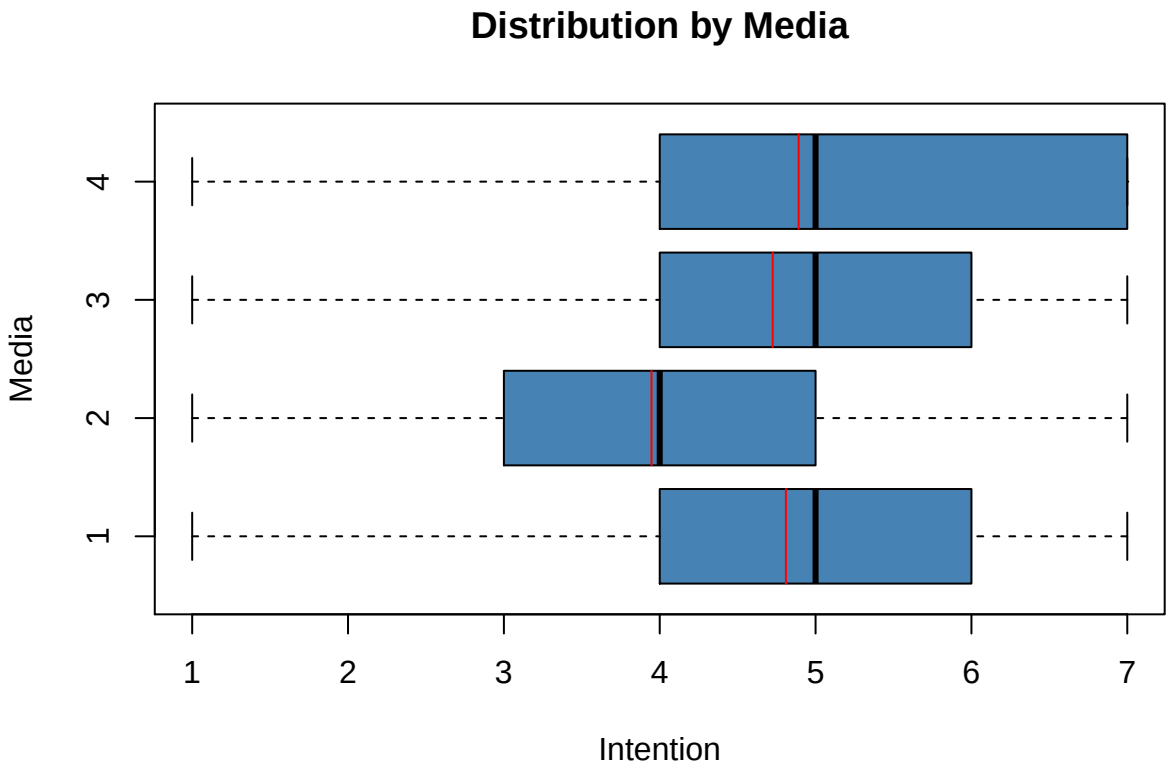
b. Visualize the distribution and mean of intention to share, across all four media.

```
test <-
  c("animation + audio", "pictures + audio", "pictures + text", "text only")

boxplot(
  INTEND.0 ~ media,
  data=media,
  main="Distribution by Media",
  xlab="Intention",
  ylab="Media",
  col="steelblue",
  border="black",
  horizontal = TRUE
```

```
)

x0s <- 1:4 - 0.4
x1s <- 1:4 + 0.4
y0s <- c(mean(m1$INTEND.0), mean(m2$INTEND.0), mean(m3$INTEND.0), mean(m4$INTEND.0))
segments(x0=y0s, y0=x0s, y1=x1s, col = "red")
```




```
# segments(y0=0.6, y1=1.4, x0=mean(m1$INTEND.0), col = "red")
# segments(y0=1.6, y1=2.4, x0=mean(m2$INTEND.0), col = "red")
# segments(y0=2.6, y1=3.4, x0=mean(m3$INTEND.0), col = "red")
# segments(y0=3.6, y1=4.4, x0=mean(m4$INTEND.0), col = "red")
```

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

In my opinion, the media won't affect the difference on intention to share.

Question 2) Let's try traditional one-way ANOVA:

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

Null hypothesis is the means of INTEND.0 in the four data sets are similar to each other. While the *Alternative hypothesis* is the means of INTEND.0 in the four data sets are not similar to each other.

b. Produce the traditional F-statistic for our test

```
grand_mean <- mean(media$INTEND.0)

intend0 <- tibble(
  media = c(1,2,3,4),
  instances = c(
    nrow(m1),
    nrow(m2),
    nrow(m3),
    nrow(m4)
  ),
  group_mean = c(
    mean(m1$INTEND.0),
    mean(m2$INTEND.0),
    mean(m3$INTEND.0),
    mean(m4$INTEND.0)
  ),
  var = c(
    var(m1$INTEND.0),
    var(m2$INTEND.0),
    var(m3$INTEND.0),
    var(m4$INTEND.0)
  ),
  sstr = c(
    nrow(m1) * (mean(m1$INTEND.0) - grand_mean)^2,
    nrow(m2) * (mean(m2$INTEND.0) - grand_mean)^2,
    nrow(m3) * (mean(m3$INTEND.0) - grand_mean)^2,
    nrow(m4) * (mean(m4$INTEND.0) - grand_mean)^2
  ),
  sse = c(
    (nrow(m1) - 1) * sd(m1$INTEND.0)^2,
    (nrow(m2) - 1) * sd(m2$INTEND.0)^2,
    (nrow(m3) - 1) * sd(m3$INTEND.0)^2,
    (nrow(m4) - 1) * sd(m4$INTEND.0)^2
  ),
)

intend0
```

```
## # A tibble: 4 x 6
##   media instances group_mean   var   sstr   sse
##   <dbl>     <int>      <dbl> <dbl>  <dbl> <dbl>
## 1     1        42       4.81  2.69  1.60  110.
## 2     2        38       3.95  2.32 16.9    85.9
## 3     3        40       4.72  3.08  0.489 120.
## 4     4        46       4.89  3.30  3.53  148.
```

```
n <- 4
sstr <- sum(intend0$sstr)
mstr <- sstr/(n-1)
sse <- sum(intend0$sse)
```

```r
mse <- sse/(nrow(media) - n)
f <- mstr / mse
f
```

```
## [1] 2.616669
```

c. What are the cut-off values of F for 95% and 99% confidence according the the null distribution of F?

d. According to the traditional ANOVA, do the four types of media produce the same mean intention to share, at 95% confidence? How about at 99% confidence?

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