BACS HW – Week 7

10807023

Question 1) Let's **develop some intuition** about the data and results:

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

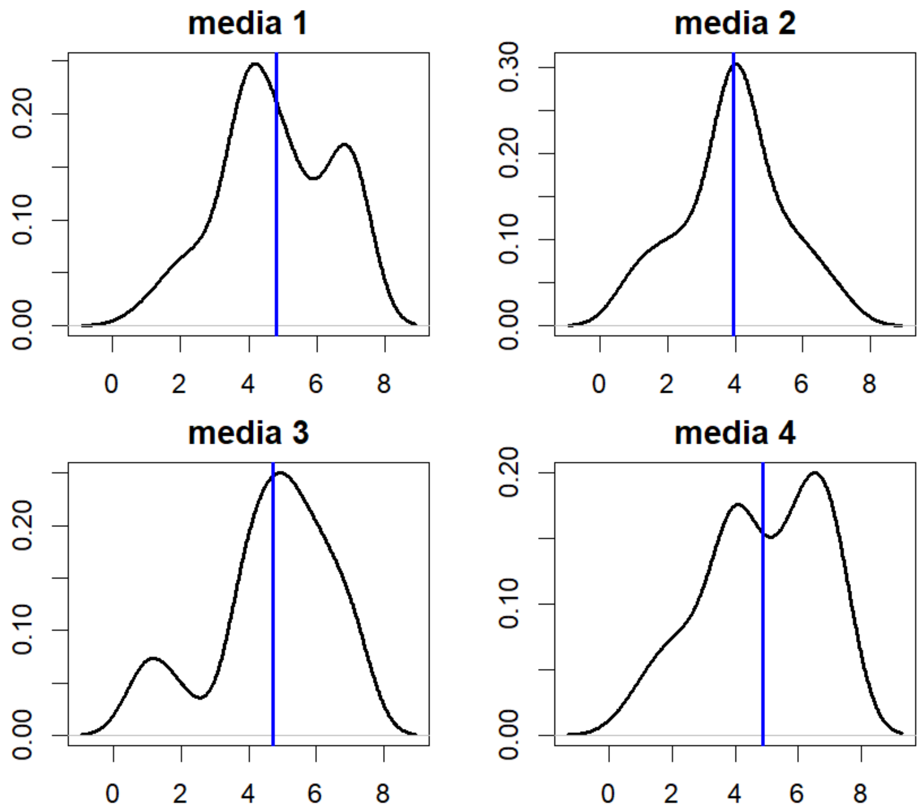
```
> #1-(a)
> m1<-read.csv("pls-media1.csv")$INTEND.0
> m2<-read.csv("pls-media2.csv")$INTEND.0
> m3<-read.csv("pls-media3.csv")$INTEND.0
> m4<-read.csv("pls-media4.csv")$INTEND.0
> mean1<-mean(m1)
> mean2<-mean(m2)
> mean3<-mean(m3)
> mean4<-mean(m4)
> mean_grand<-mean(c(mean1,mean2,mean3,mean4))

> mean1
[1] 4.809524
> mean2
[1] 3.947368
> mean3
[1] 4.725
> mean4
[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)

```
> #1-(b)
> par(mar = c(2, 2, 2, 2))
> par(mfrow=c(2,2))    # set the plotting area into a 1*2 array
> plot(density(m1),main ="media 1",lwd=2)
> abline(v=mean(m1),col="blue",lwd=2)
> plot(density(m2),main ="media 2",lwd=2)
> abline(v=mean(m2),col="blue",lwd=2)
> plot(density(m3),main ="media 3",lwd=2)
```

```
> abline(v=mean(m3),col="blue",lwd=2)
> plot(density(m4),main ="media 4",lwd=2)
> abline(v=mean(m4),col="blue",lwd=2)
```



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

Ans: From the figure above, we can see that means of these media are quite similar,

while the distributions look quite different. Therefore, I will conclude that media do make 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

Ans:

Null hypothesis: the means of four media are the same

Alternative hypothesis: the means of four media are not the same

b. Let's compute the F-statistic ourselves:

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

```
> #2-(b)-i
> SSTR<-length(m1)*((mean1-mean_grand)^2)+length(m2)*((mean2-
mean_grand)^2)+length(m3)*((mean3-mean_grand)^2)+length(m4)*((mean4-
mean_grand)^2)
> df_MSTR<-4-1
> MSTR<-SSTR/df_MSTR
> MSTR
[1] 7.53239
> SSE<-(length(m1)-1)*var(m1)+(length(m2)-1)*var(m2)+(length(m3)-
1)*var(m3)+(length(m4)-1)*var(m4)
> df_MSE<-length(m1)+length(m2)+length(m3)+length(m4)-4
> MSE<-SSE/df_MSE
> MSE
[1] 2.869151
> F<-MSTR/MSE
> F
[1] 2.625303
```

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.

```
> #2-(b)-ii
> qf(p=0.95,df_MSTR,df_MSE)
[1] 2.660406
> p_val<-pf(F,df_MSTR,df_MSE,lower.tail = FALSE)
> p_val
[1] 0.05230686
```

Ans: From the above result, the p-value of anova is 0.052, which is quite close to the confidence level, but we still not have the power to reject the null hypothesis.

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

```
> #2-(c)
> m1<-data.frame(m1)
> m2<-data.frame(m2)
> m3<-data.frame(m3)
> m4<-data.frame(m4)
> library(reshape2)
> media<-
melt(c(m1,m2,m3,m4),id.vars=NULL,variable.name="media",value.name="score")
```

```
> anova_model<-aov(media$score~factor(media$L1))
> summary(anova_model)
                  Df Sum Sq Mean Sq F value Pr(>F)
factor(media$L1)   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 in R) to see if any *pairs of media have significantly different means* – what do you find?

```
> #2-(d)
> TukeyHSD(anova_model,conf.level = 0.05)
  Tukey multiple comparisons of means
    5% family-wise confidence level

Fit: aov(formula = media$score ~ factor(media$L1))

$`factor(media$L1)`
             diff         lwr        upr     p adj
m2-m1 -0.86215539 -1.06562977 -0.6586810 0.1085727
m3-m1 -0.08452381 -0.28530983  0.1162622 0.9959223
m4-m1  0.08178054 -0.11218249  0.2757436 0.9959032
m3-m2  0.77763158  0.57175512  0.9835080 0.1825044
m4-m2  0.94393593  0.74470805  1.1431638 0.0573229
m4-m3  0.16630435 -0.03017708  0.3627858 0.9687417
```

Ans: By above result, we can see that all of the p value are bigger than 0.05,so we can’t reject any H0.Therefore, we can’t say any pair of groups are different, which correspond to the result in 2-(b) and (c).

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

(Feel free to use any combination of methods we saw in class or any analysis we haven’t covered)

```
> c(var(m1),var(m2),var(m3),var(m4))
[1] 2.694541 2.321479 3.076282 3.299034
```

Ans: No, by the results above and 1-a, we can clearly see that the variance of the four group is quite different. Also, by the visualized figure, the curve in the destribution of media 1,3,4 do not follow the normal distribution.

Question 3) Let's use the **non-parametric Kruskal Wallis** test:

a. State the null and alternative hypotheses (in terms of distribution or difference of mean ranks)

Ans: Null hypothesis: numbers of for group are the same

Alternative hypothesis: numbers of for group are not the same

b. Let's compute (an approximate) Kruskal Wallis H ourselves:

i.Show the code and results of computing H

```
> #3-(b)-i
> c(m1,m2,m3,m4)
  [1] 3 5 4 5 5 4 4 5 4 1 7 3 4 7 7 5 4 6 4 6 4 2 7 7 6 4 2 3 5
 [30] 7 4 5 4 4 7 7 4 2 5 6 7 7 4 6 4 4 5 4 4 4 4 7 4 4 2 4 4 5
 [59] 4 1 7 3 4 3 6 4 2 5 5 3 1 2 5 2 6 3 1 4 4 6 1 4 1 5 6 6 5
 [88] 7 5 5 7 5 4 2 5 4 6 5 4 7 5 4 5 6 6 1 4 2 7 4 4 1 4 7 6 6
[117] 5 6 5 7 3 4 4 2 7 7 5 7 5 6 4 7 6 4 4 1 2 1 6 7 2 7 4 6 5
[146] 5 6 4 3 2 6 4 6 6 4 7 6 7 3 4 7 4 7 7 7 4
> score_rank<-rank(media$score)
> score_rank
  [1]  28.5  97.5  58.5  97.5  97.5  58.5  58.5  97.5  58.5
 [10]   5.5 151.5  28.5  58.5 151.5 151.5  97.5  58.5 124.0
 [19]  58.5 124.0  58.5  17.0 151.5 151.5 124.0  58.5  17.0
 [28]  28.5  97.5 151.5  58.5  97.5  58.5  58.5 151.5 151.5
 [37]  58.5  17.0  97.5 124.0 151.5 151.5  58.5 124.0  58.5
 [46]  58.5  97.5  58.5  58.5  58.5  58.5 151.5  58.5  58.5
 [55]  17.0  58.5  58.5  97.5  58.5   5.5 151.5  28.5  58.5
 [64]  28.5 124.0  58.5  17.0  97.5  97.5  28.5   5.5  17.0
 [73]  97.5  17.0 124.0  28.5   5.5  58.5  58.5 124.0   5.5
 [82]  58.5   5.5  97.5 124.0 124.0  97.5 151.5  97.5  97.5
 [91] 151.5  97.5  58.5  17.0  97.5  58.5 124.0  97.5  58.5
[100] 151.5  97.5  58.5  97.5 124.0 124.0   5.5  58.5  17.0
[109] 151.5  58.5  58.5   5.5  58.5 151.5 124.0 124.0  97.5
[118] 124.0  97.5 151.5  28.5  58.5  58.5  17.0 151.5 151.5
[127]  97.5 151.5  97.5 124.0  58.5 151.5 124.0  58.5  58.5
[136]   5.5  17.0   5.5 124.0 151.5  17.0 151.5  58.5 124.0
[145]  97.5  97.5 124.0  58.5  28.5  17.0 124.0  58.5 124.0
[154] 124.0  58.5 151.5 124.0 151.5  28.5  58.5 151.5  58.5
[163] 151.5 151.5 151.5  58.5
> ranked_groups<-split(score_rank,media$L1)
```

```
> R1<-sum(ranked_groups$m1)
> R2<-sum(ranked_groups$m2)
> R3<-sum(ranked_groups$m3)
> R4<-sum(ranked_groups$m4)
> length(m1)
[1] 42
> N<-length(m1)+length(m2)+length(m3)+length(m4)
> N
[1] 166
> H<-
(12*((R1^2/length(m1))+(R2^2/length(m2))+(R3^2/length(m3))+(R4^2/length(m4)
))/(N*(N+1)))-3*(N+1)
> H
[1] 8.45466
```

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.

```
> #3-(b)-ii
> kw_p<-1-pchisq(H,df=3)
> kw_p
[1] 0.03749292
```

Ans: Since p value<0.05, we can reject H0. Therefore, numbers in the four groups are not the same

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

```
> #3-(c)
> kruskal.test(score~factor(L1),data=media)


        Kruskal-Wallis rank sum test


data:  score by factor(L1)
Kruskal-Wallis chi-squ
ared = 8.8283, df = 3, p-value =
0.03166
```

Ans: Since p value<0.05, we can reject H0. Therefore,numbers in the four groups are not the same. And the result is similar with 3-b.

d. Regardless of your conclusions, conduct a post-hoc Dunn test (feel free to use the dunnTest() function from the FSA package) to see if any *pairs of media are significantly different* – what do you find?

```
> #3-(d)
> #install.packages("FSA")
> library(FSA)
> dunnTest(score~factor(L1),data=media)
Dunn (1964) Kruskal-Wallis multiple comparison
  p-values adjusted with the Holm method.

  Comparison           Z      P.unadj      P.adj
1    m1 - m2  2.30087819 0.021398517 0.08559407
2    m1 - m3 -0.09233644 0.926430736 0.92643074
3    m2 - m3 -2.36408588 0.018074622 0.09037311
4    m1 - m4 -0.31452459 0.753122646 1.00000000
5    m2 - m4 -2.65613380 0.007904225 0.04742535
6    m3 - m4 -0.21613379 0.828883460 1.00000000
```

Ans: The p value of m2-m4 is smaller than 0.05, so we'll conclude that m2 and m4 are different.