

## bacs\_hw6

110071010

2024-03-25

### Question 1 : Data Exploration & Description

*# Reading in the files*

```
pls_media1 <- read.csv("pls-media1.csv")
pls_media2 <- read.csv("pls-media2.csv")
pls_media3 <- read.csv("pls-media3.csv")
pls_media4 <- read.csv("pls-media4.csv")
```

```
media1 <- pls_media1$INTEND.0
media2 <- pls_media2$INTEND.0
media3 <- pls_media3$INTEND.0
media4 <- pls_media4$INTEND.0
```

*# Combining four individual datasets into dataframe*

*# NA are filled in due to different lengths between datasets*

```
maxLength <- max(length(media1), length(media2), length(media3), length(media4))
```

```
media1 <- ifelse(1:maxLength <= length(media1), media1, NA)
media2 <- ifelse(1:maxLength <= length(media2), media2, NA)
media3 <- ifelse(1:maxLength <= length(media3), media3, NA)
media4 <- ifelse(1:maxLength <= length(media4), media4, NA)
```

```
df_media <- data.frame(ani_aud = media1, pic_aud = media2, pic_txt = media3, txt_only = media4)
```

```
head(df_media)
```

```
##   ani_aud pic_aud pic_txt txt_only
## 1      3      4      1      3
## 2      5      6      4      4
## 3      4      4      1      4
## 4      5      4      5      2
## 5      5      5      6      7
## 6      4      4      6      7
```

1a

Question

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

**Answer**

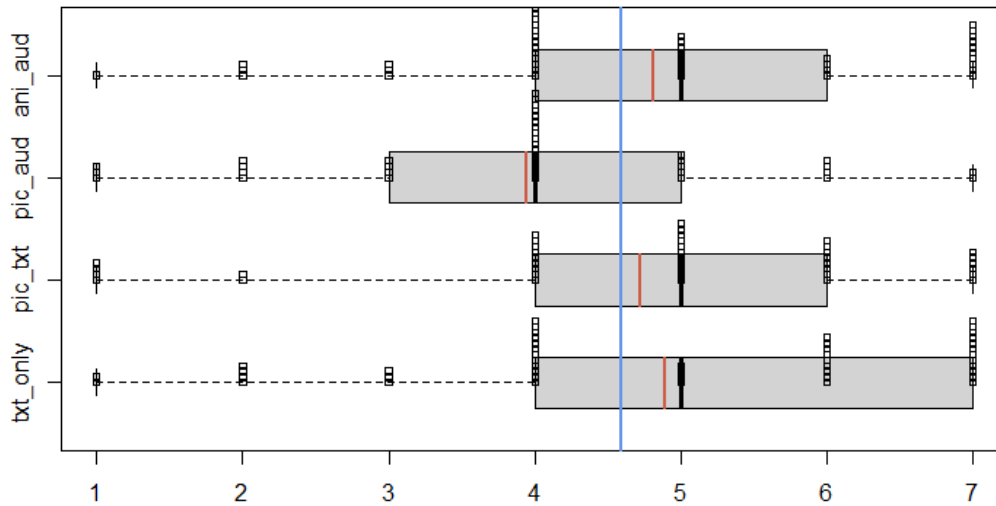
```
sapply(df_media, function(x) mean(x, na.rm = TRUE))  
  
## ani_aud pic_aud pic_txt txt_only  
## 4.809524 3.947368 4.725000 4.891304
```

**1b**

**Instruction**

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

```
boxplot(rev(df_media), horizontal = TRUE, boxlwd = 1, outwex = 0.5, boxwex = 0.5)  
stripchart(rev(df_media), method = "stack", add = TRUE, cex = 0.6)  
  
# Adding the general mean(blue)  
abline(v=mean(sapply(df_media, function(x) mean(x, na.rm = TRUE))), col = 'cornflowerblue', lwd = 2)  
  
# Adding mean(coral) to each medium's boxplot  
means <- colMeans(rev(df_media), na.rm = TRUE)  
positions <- 1:ncol(df_media)  
  
for (i in seq_along(means)){  
  segments(x0 = means[i], x1 = means[i], y0 = positions[i] - 0.25, y1 = positions[i] + 0.25, col = 'coral3', lwd = 2)  
}
```



1c

### Question

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

```
sapply(df_media, function(x) sd(x, na.rm = TRUE))
```

```
## ani_aud pic_aud pic_txt txt_only
## 1.641506 1.523640 1.753933 1.816324
```

### Answer

Yes. With the box representing “pictures and audio” clearly drifting toward the left, the result might hint at generally lower intentions to share among respondents for this media form. In contrast, the larger width of “text only” box might suggest that participant opinions on sharing “Text Only” media are more polarized, as evidenced by its larger standard deviation in the output above.

## Question 2 : One-way ANOVA

2a

### Instruction

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

### Solution

- **Null Hypothesis (H0):** There is no difference in the mean intention to share (INTEND.0) across the four media types
- **Alternative Hypothesis (H1):** At least one media type has a different mean intention to share (INTEND.0) compared to the others

### 2b

Let's compute the F-statistic ourselves.

#### (i)

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

```
# Calculate overall mean
overall_mean <- mean(unlist(df_media), na.rm = TRUE)

# Calculate group means and number of observations per group
group_means <- colMeans(df_media, na.rm = TRUE)
n <- colSums(!is.na(df_media))

# SSTR: Sum of Squares for Treatment
sstr <- sum(n * (group_means - overall_mean)^2)
df_mstr <- length(group_means) - 1
mstr <- sstr / df_mstr

# SSE: Sum of Squares for Error
sse <- sum(sapply(df_media, function(x) sum((x - mean(x, na.rm = TRUE))^2, na.rm = TRUE)))
df_mse <- sum(n) - length(group_means)
mse <- sse / df_mse

# F-value
f_value <- mstr / mse

# P-value
p_value <- pf(f_value, df_mstr, df_mse, lower.tail = FALSE)

# Displaying results
cat(" MSTR =", mstr, "\n",
    "MSE =", mse, "\n",
    "F-value =", f_value, "\n")

## MSTR = 7.507617
## MSE = 2.869151
## F-value = 2.616669
```

(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.

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

# Printing results
cat(" p_value =", p_value)

## p_value = 0.05289015
```

### **Conclusion**

Since p-value > 0.05, we do not reject the null hypothesis.

2c

### **Instruction**

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

```
# Transforming df_media to long format suitable for ANOVA
df_long <- stack(df_media)
names(df_long) <- c("INTEND.0", "media_type")

# Ensure media_type is treated as a factor
df_long$media_type <- factor(df_long$media_type)

# Performing ANOVA with aov()
anova_model <- aov(INTEND.0 ~ media_type, data = df_long)

# Displaying the detailed summary of the model
summary(anova_model)

##              Df Sum Sq Mean Sq F value Pr(>F)
## media_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
```

### **What's Observed**

The result aligns with 2b.

2d

### **Question**

Regardless of your conclusions, conduct a post-hoc Tukey test to see if any pairs of media have significantly different means – what do you find?

### **Solution**

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

##    Tukey multiple comparisons of means
##      95% family-wise confidence level
##
## Fit: aov(formula = INTEND.0 ~ media_type, data = df_long)
##
## $media_type
##              diff          lwr          upr      p adj
## pic_aud-ani_aud -0.86215539 -1.84660332  0.1222925 0.1085727
## pic_txt-ani_aud -0.08452381 -1.05596494  0.8869173 0.9959223
## txt_only-ani_aud  0.08178054 -0.85664966  1.0202107 0.9959032
## pic_txt-pic_aud   0.77763158 -0.21843807  1.7737012 0.1825044
## txt_only-pic_aud  0.94393593 -0.01996662  1.9078385 0.0573229
## txt_only-pic_txt  0.16630435 -0.78431033  1.1169190 0.9687417
```

### **What's Observed**

All comparisons get their p-value > 0.05, and that between txt\_only and pic\_aud displays the biggest difference.

## **2e**

### **Question**

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

### **Answer**

#### **1) Independence (YES)**

By design, independence is guaranteed.

#### **2) Data follows ND (YES)**

The data we look into are means, so the fact that all lengths in each column (media type) are over 30 has ensured that CLT holds, as demonstrated in code outputs below.

```
sum(!is.na(df_media$ani_aud))

## [1] 42

sum(!is.na(df_media$pic_aud))

## [1] 38

sum(!is.na(df_media$pic_txt))
```

```
## [1] 40
sum(!is.na(df_media$txt_only))
## [1] 46
```

### 3) Homogeneity of Variances (YES)

Checking with Levene's test result, one can conclude that the assumption is not violated.

```
# Install and load the car package for Levene's test
# install.packages("car")
library(car)

# Levene's test for homogeneity of variances
leveneTest(INTEND.0 ~ media_type, data=df_long)

## Levene's Test for Homogeneity of Variance (center = median)
##           Df F value Pr(>F)
## group      3  1.5403 0.2061
##           162
```

## Question 3 : Non-parametric Kruskal Wallis Test

### 3a

#### Instruction

State the null and alternative hypotheses

#### Solution

**Null Hypothesis (H0):** The median values of viewers' intention to share are equal across all four media types.

**Alternative Hypothesis (H1):** At least one of the media types has a median value of viewers' intention to share that is significantly different from the others.

### 3b

#### Instruction

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

#### (i)

Show the code and results of computing H

```
# Rank all the combined values across groups
all_values <- unlist(df_media) # Combine all group values into a single
vector
```

```

sales_ranks <- rank(all_values) # Rank all combined values

# Group the ranks into original groups
group_ranks <- split(sales_ranks, df_long$media_type)

# Sum the ranks for each group
rank_sums <- sapply(group_ranks, sum)

# Apply the Kruskal Wallis formula to sum the squared ranks
n <- length(all_values) # total number of observations
k <- length(rank_sums) # number of groups
H <- (12 / (n*(n+1))) * sum(rank_sums^2 / colSums(!is.na(df_media))) -
3*(n+1)

# Printing H
cat(" Kruskal Wallis H:", H, "\n")

## Kruskal Wallis H: 64.99589

```

(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.

```

# Find p-value of H from  $\chi^2$  distribution
kw_p <- 1 - pchisq(H, df = k-1)
cat(" p-value:", kw_p)

## p-value: 5.02931e-14

```

### Conclusion

No. Since p-value < 0.05, we reject null hypothesis.

## 3c

### Instruction

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

```

# Perform Kruskal-Wallis test using kruskal.test()
kruskal.test(INTEND.0 ~ media_type, data=df_long)

##
## Kruskal-Wallis rank sum test
##
## data: INTEND.0 by media_type
## Kruskal-Wallis chi-squared = 8.8283, df = 3, p-value = 0.03166

```

### Conclusion



Both tests (self-built one in 3b & built-in one in 3c) give the same conclusion that null hypothesis is rejected.

### 3d

#### Question

Regardless of your conclusions, conduct a post-hoc Dunn test to see if the values of any pairs of media are significantly different – what are your conclusions?

```
# Import FSA
# install.packages("FSA")
library(FSA)

# Conducting Dunn's test for post-hoc analysis
dunnTest(INTEND.0 ~ media_type, data = df_long, method = "bonferroni")

## Warning: Some rows deleted from 'x' and 'g' because missing data.
## Dunn (1964) Kruskal-Wallis multiple comparison
##   p-values adjusted with the Bonferroni method.
##
```

	Comparison	Z	P.unadj	P.adj
## 1	ani_aud - pic_aud	2.30087819	0.021398517	0.12839110
## 2	ani_aud - pic_txt	-0.09233644	0.926430736	1.00000000
## 3	pic_aud - pic_txt	-2.36408588	0.018074622	0.10844773
## 4	ani_aud - txt_only	-0.31452459	0.753122646	1.00000000
## 5	pic_aud - txt_only	-2.65613380	0.007904225	0.04742535
## 6	pic_txt - txt_only	-0.21613379	0.828883460	1.00000000

#### Conclusion

Among the pairwise comparisons in the post-hoc Dunn test, only comparison between pictures and audio (**pic\_aud**) versus text only (**txt\_only**) stands out. This finding implies that the way information is presented (with pictures and audio versus just text) can significantly affect viewers' intentions to share that information.