Homework7_BACS

109090035 helped by 109090046 109070028

3/29/2023

Question 1) Let’s explore and describe the data and develop some early intuitive thoughts:

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

```
media1 <- read_csv("/Users/user/Downloads/pls-media/pls-media1.csv")
```

```
Rows: 42 Columns: 20
── Column specification ────────────────────────────────────────────────────────────────────────────────
─────
Delimiter: ","
dbl (20): media, INTEND.0, INTEND.1, INTEND.2, ATT.0, ATT.1, ATT.2, Humor.0,...

ℹ Use `spec()` to retrieve the full column specification for this data.
ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
media2 <- read_csv("/Users/user/Downloads/pls-media/pls-media2.csv")
```

```
Rows: 38 Columns: 20
── Column specification ────────────────────────────────────────────────────────
──────
Delimiter: ","
dbl (20): media, INTEND.0, INTEND.1, INTEND.2, ATT.0, ATT.1, ATT.2, Humor.0,...

ℹ Use `spec()` to retrieve the full column specification for this data.
ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
media3 <- read_csv("/Users/user/Downloads/pls-media/pls-media3.csv")
```

```
Rows: 40 Columns: 20
── Column specification ────────────────────────────────────────────────────────
──────
Delimiter: ","
dbl (20): media, INTEND.0, INTEND.1, INTEND.2, ATT.0, ATT.1, ATT.2, Humor.0,...

ℹ Use `spec()` to retrieve the full column specification for this data.
ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
media4 <- read_csv("/Users/user/Downloads/pls-media/pls-media4.csv")
```

```
Rows: 46 Columns: 20
── Column specification ────────────────────────────────────────────────────────
──────
Delimiter: ","
dbl (20): media, INTEND.0, INTEND.1, INTEND.2, ATT.0, ATT.1, ATT.2, Humor.0,...

ℹ Use `spec()` to retrieve the full column specification for this data.
ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

lets add a media type column

```
media1$media_type <- "1 - Video (Animation + Audio)"
media2$media_type <- "2 - Video (Pictures + Audio)"
media3$media_type <- "3 - Webpage (Pictures + Text)"
media4$media_type <- "4 - Webpage (Text Only)"
```

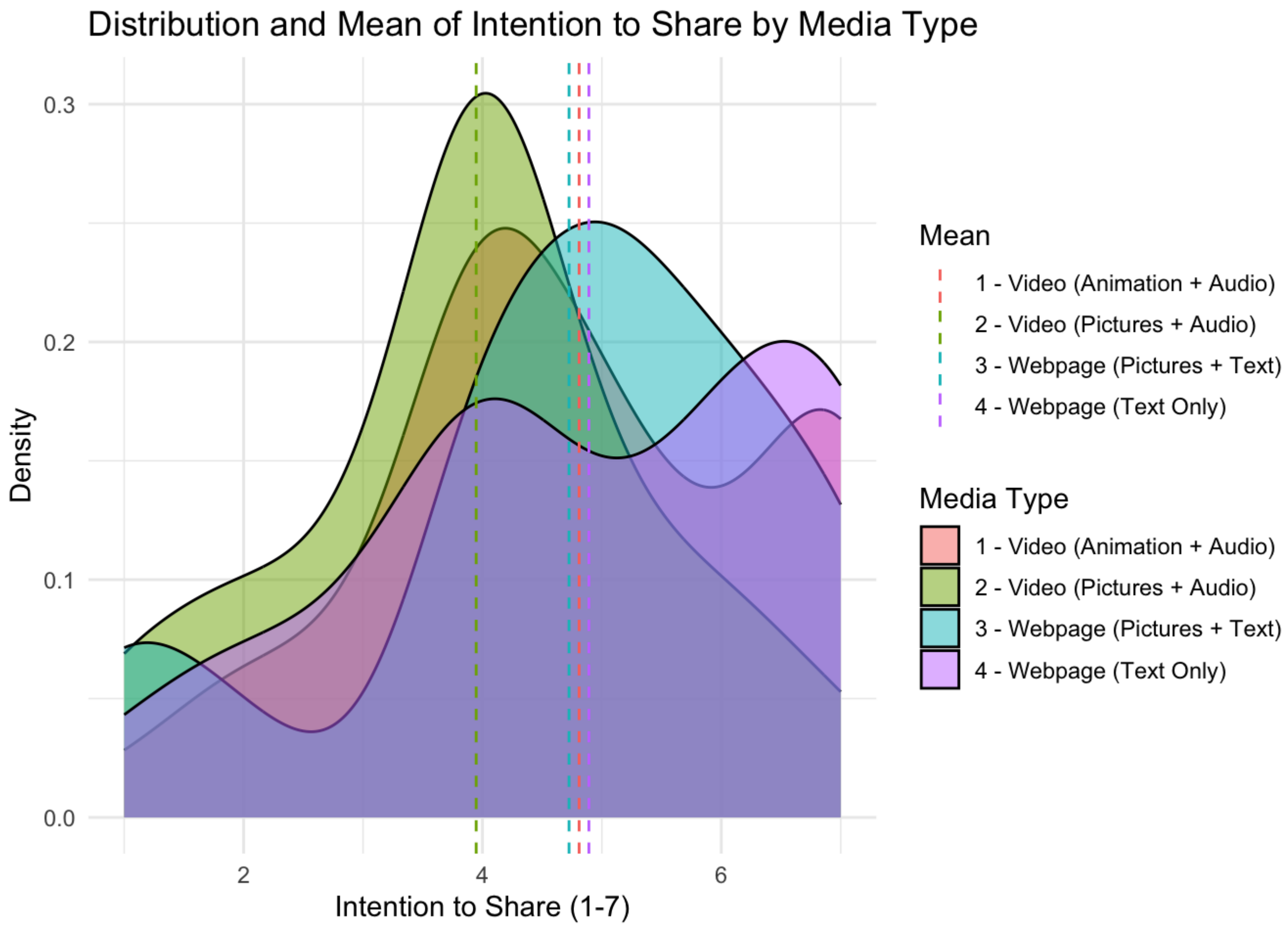
Combine the data and calculate the means for each media type

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

mean_intentions <- combined_data %>% group_by(media_type) %>% summarise(mean_intention = mean(INTEND.0, na.rm = T
RUE))

mean_intentions %>% ptable()
```

media_type	mean_intention
1 - Video (Animation + Audio)	4.809524
2 - Video (Pictures + Audio)	3.947368
3 - Webpage (Pictures + Text)	4.725000
4 - Webpage (Text Only)	4.891304

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

```
ggplot(combined_data, aes(x = INTEND.0, fill = media_type)) +
  geom_density(alpha = 0.5) +
  geom_vline(data = mean_intentions, aes(xintercept = mean_intention, color = media_type), linetype = "dashed") +
  labs(title = "Distribution and Mean of Intention to Share by Media Type",
       x = "Intention to Share (1-7)",
       y = "Density",
       fill = "Media Type",
       color = "Mean") +
  theme_minimal()
```



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

It feels like media type 4 has less mean intention to share than other 3 types of media sharing , since the the mean and distribution of type 4 significant distinct from other 3 in mean and distribution.

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 (H0): There is no significant difference in the mean intention to share scores (INTEND.0) between the four media types.

Alternative hypothesis (Ha): There is a significant difference in the mean intention to share scores (INTEND.0) between at least two media types.

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

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

```
#overall mean
overall_mean <- mean(combined_data$INTEND.0, na.rm = TRUE)

#mean for each media type
group_means <- combined_data %>% group_by(media_type) %>% summarise(group_mean = mean(INTEND.0, na.rm = TRUE))

#Merge the group means back into the original data:
combined_data_with_means <- merge(combined_data, group_means, by = "media_type")

#Calculate the Sum of Squares for Treatments (SSTR) and Mean Sum of Squares for Treatments (MSTR):
SSTR <- sum((combined_data_with_means$group_mean - overall_mean)^2)
MSTR <- SSTR / (length(unique(combined_data$media_type)) - 1)

#Calculate the Sum of Squares for Error (SSE) and Mean Sum of Squares for Error (MSE):
SSE <- sum((combined_data_with_means$INTEND.0 - combined_data_with_means$group_mean)^2)
MSE <- SSE / (nrow(combined_data) - length(unique(combined_data$media_type)))
F_statistic <- MSTR / MSE
# Compute the p-value of F, from the null F-distribution:
df1 <- length(unique(combined_data$media_type)) - 1
df2 <- nrow(combined_data) - length(unique(combined_data$media_type))
p_value <- 1 - pf(F_statistic, df1, df2)


# Print MSTR
cat("MSTR:", MSTR, "\n")
```

```
MSTR: 7.507617
```

```
# Print MSE
cat("MSE:", MSE, "\n")
```

```
MSE: 2.869151
```

```
# Print F-statistic
cat("F-statistic:", F_statistic, "\n")
```

```
F-statistic: 2.616669
```

```
# Print p-value
cat("p-value:", p_value, "\n")
```

```
p-value: 0.05289015
```

The summary will show the Mean Sum of Squares for Treatments (MSTR), Mean Sum of Squares for Error (MSE), and the F-statistic.

2. Compute the p-value of F, from the null F-distribution; is the F-value significant?

```
# Extract p-value
print(p_value)
```

```
[1] 0.05289015
```

If the p-value is less than the significance level (e.g., 0.05), it indicates that the F-value is significant, and we reject the null hypothesis in favor of the alternative hypothesis. This means there is a significant difference in the mean intention to share scores (INTEND.0) between at least two media types.

If so, state your conclusion for the hypotheses.

Since 0.05289015 is higher than significance level, we conclude that the F-value is not significant.

and we cannot reject the null hypothesis in favor of the alternative hypothesis. This means there is no significant difference in the mean intention to share scores (INTEND.0) between at least two media types.

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

```
# One-way ANOVA
anova_results <- aov(INTEND.0 ~ media_type, data = combined_data)
summary(anova_results)
```

```
             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
```

d. Regardless of your conclusions, conduct a post-hoc Tukey test (feel free to use the TukeyHSD() function included in base R) to see if any pairs of media have significantly different means – what do you find?

```
# Tukey post-hoc test
tukey_results <- TukeyHSD(anova_results)
print(tukey_results)
```

```
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = INTEND.0 ~ media_type, data = combined_data)

$media_type
                                                                   diff
2 - Video (Pictures + Audio)-1 - Video (Animation + Audio)  -0.86215539
3 - Webpage (Pictures + Text)-1 - Video (Animation + Audio) -0.08452381
4 - Webpage (Text Only)-1 - Video (Animation + Audio)        0.08178054
3 - Webpage (Pictures + Text)-2 - Video (Pictures + Audio)   0.77763158
4 - Webpage (Text Only)-2 - Video (Pictures + Audio)         0.94393593
4 - Webpage (Text Only)-3 - Webpage (Pictures + Text)        0.16630435
                                                                    lwr
2 - Video (Pictures + Audio)-1 - Video (Animation + Audio)  -1.84660332
3 - Webpage (Pictures + Text)-1 - Video (Animation + Audio) -1.05596494
4 - Webpage (Text Only)-1 - Video (Animation + Audio)       -0.85664966
3 - Webpage (Pictures + Text)-2 - Video (Pictures + Audio)  -0.21843807
4 - Webpage (Text Only)-2 - Video (Pictures + Audio)        -0.01996662
4 - Webpage (Text Only)-3 - Webpage (Pictures + Text)       -0.78431033
                                                                  upr     p adj
2 - Video (Pictures + Audio)-1 - Video (Animation + Audio)  0.1222925 0.1085727
3 - Webpage (Pictures + Text)-1 - Video (Animation + Audio) 0.8869173 0.9959223
4 - Webpage (Text Only)-1 - Video (Animation + Audio)       1.0202107 0.9959032
3 - Webpage (Pictures + Text)-2 - Video (Pictures + Audio)  1.7737012 0.1825044
4 - Webpage (Text Only)-2 - Video (Pictures + Audio)        1.9078385 0.0573229
4 - Webpage (Text Only)-3 - Webpage (Pictures + Text)       1.1169190 0.9687417
```

It turns out that the type 4 - Webpage (Text Only)- and type 2 - Video (Pictures + Audio) might closely to be reject (there mean are not the same) due to their p-value = 0.0573229, close to the significant alpha (e.g 0.05)

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

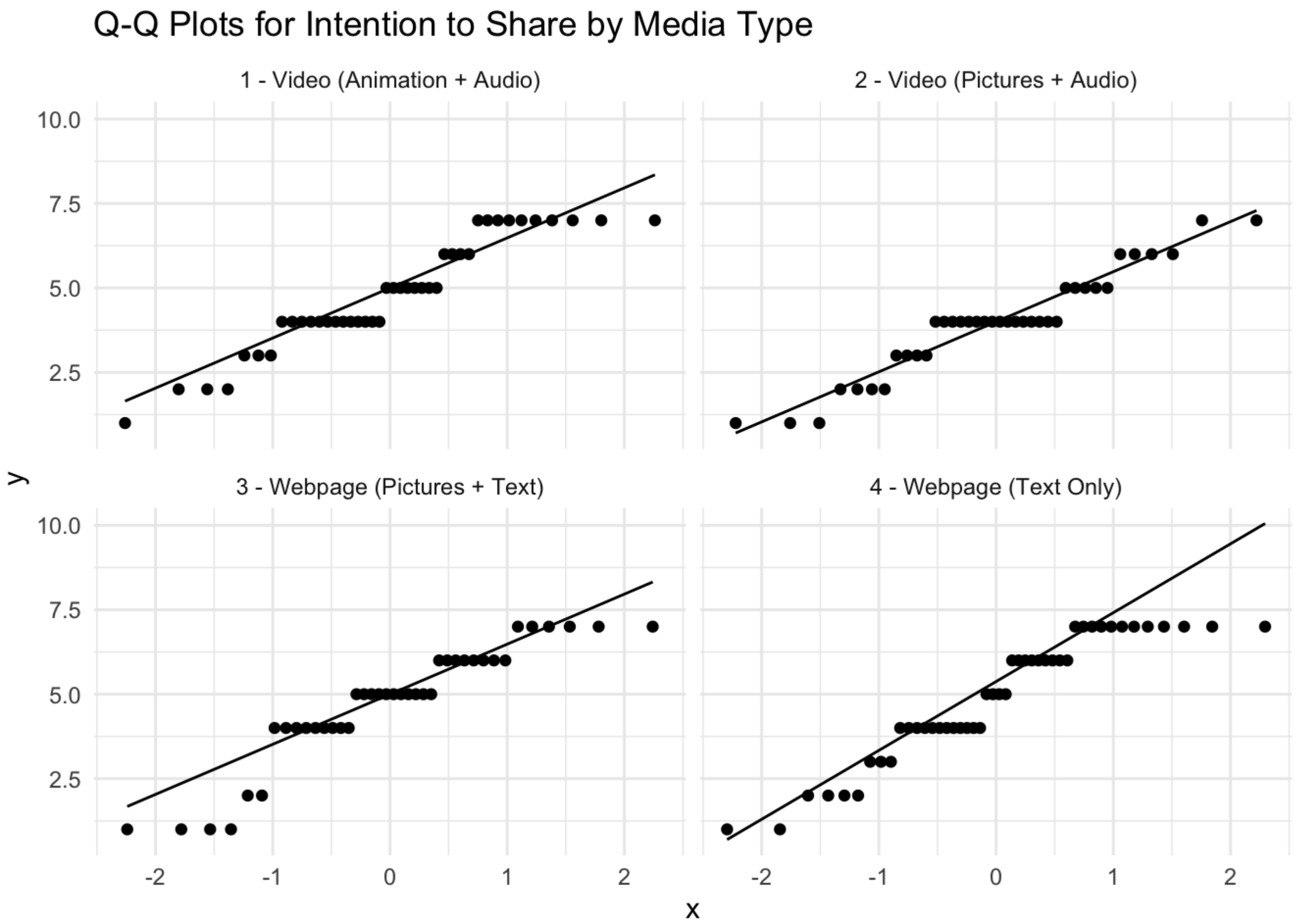
To test the assumptions that need to be met for one-way ANOVA:

1. Independence: Observations within each group should be independent of each other.

Independence: This assumption is usually met by the study design. In this case, the researcher randomly assigned people to different groups, which should ensure independence.

2. Normality: The response variable (INTEND.0) should be approximately normally distributed within each group.

```
## QQ plot for checking Normality
ggplot(combined_data, aes(sample = INTEND.0)) +
  geom_qq() +
  geom_qq_line() +
  facet_wrap(~ media_type) +
  theme_minimal() +
  labs(title = "Q-Q Plots for Intention to Share by Media Type")
```



We can see most of them meet Normality !

3.Homogeneity of variance: The variances of the response variable should be approximately equal across groups.

To check if these assumptions are met, you can perform the following tests and visualizations:

```
#  perform Levene's test to check for equal variances:
leveneTest(INTEND.0 ~ media_type, data = combined_data)
```

```
Warning in leveneTest.default(y = y, group = group, ...): group coerced to
factor.
```

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

Since the p-value from Levene’s test is greater than the chosen significance level (e.g., 0.05), we can assume that the variances are equal across groups.

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

a. State the null and alternative hypotheses

Null hypothesis (H0): There is no significant difference in the distribution of intention to share scores (INTEND.0) between the four media types.

Alternative hypothesis (Ha): There is a significant difference in the distribution of intention to share scores (INTEND.0) between at least two media types.

b. Let’s compute (an approximate) Kruskal Wallis H ourselves (use the formula we saw in class or another formula might have found at a reputable website/book):

1. Show the code and results of computing H

Let’s compute (an approximate) Kruskal Wallis H ourselves:

```
# Rank the data
ranked_data <- combined_data %>% mutate(rank = rank(INTEND.0))

# Calculate the sum of ranks for each group
sum_of_ranks <- ranked_data %>% group_by(media_type) %>% summarise(sum_rank = sum(rank))

# Calculate the number of observations in each group
group_sizes <- ranked_data %>% group_by(media_type) %>% summarise(group_size = n())

# Compute H
n_total <- nrow(ranked_data)
H <- (12 / (n_total * (n_total + 1))) * sum(((sum_of_ranks$sum_rank)^2) / group_sizes$group_size) - 3 * (n_total
+ 1)

cat("H-Statistics:", H, "\n")
```

```
H-Statistics: 8.45466
```

2. 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.

```
# Calculate degrees of freedom
df <- length(unique(ranked_data$media_type)) - 1

# Compute the p-value
p_value <- 1 - pchisq(H, df)

cat("p-value:", p_value, "\n")
```

```
p-value: 0.03749292
```

Check if the H value is significant:

The p-value is less than the significance level, it indicates that the H value is significant, and we can reject the null hypothesis in favor of the alternative hypothesis. This means there is a significant difference in the distribution of intention to share scores (INTEND.0) between at least two media types.

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

```
# Kruskal-Wallis test
kruskal_test <- kruskal.test(INTEND.0 ~ media_type, data = combined_data)
print(kruskal_test)
```

```


    Kruskal-Wallis rank sum test

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

Compare the H and p-value to previous answer, we can see they are similar.

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 the values of any pairs of media are significantly different – what are your conclusions?

```
# Dunn post-hoc test
dunn_test <- dunnTest(INTEND.0 ~ media_type, data = combined_data, method = "bonferroni")
```

```
Warning: media_type was coerced to a factor.
```

```
print(dunn_test)
```

```
Dunn (1964) Kruskal-Wallis multiple comparison
```

```
p-values adjusted with the Bonferroni method.
```

```
                                                     Comparison           Z
1  1 - Video (Animation + Audio) - 2 - Video (Pictures + Audio)  2.30087819
2 1 - Video (Animation + Audio) - 3 - Webpage (Pictures + Text) -0.09233644
3  2 - Video (Pictures + Audio) - 3 - Webpage (Pictures + Text) -2.36408588
4       1 - Video (Animation + Audio) - 4 - Webpage (Text Only) -0.31452459
5        2 - Video (Pictures + Audio) - 4 - Webpage (Text Only) -2.65613380
6       3 - Webpage (Pictures + Text) - 4 - Webpage (Text Only) -0.21613379
      P.unadj      P.adj
1 0.021398517 0.12839110
2 0.926430736 1.00000000
3 0.018074622 0.10844773
4 0.753122646 1.00000000
5 0.007904225 0.04742535
6 0.828883460 1.00000000
```

Looking at the adjusted p-values (P.adj), we can see that only the comparison between media type 2 (Video with Pictures + Audio) and media type 4 (Webpage with Text Only) has a significant difference (p-value = 0.04742535) at a 95% confidence level.

The other comparisons do not show significant differences between the intention to share scores for the respective media types, as their adjusted p-values are greater than 0.05.