Factorial Design Example SOLUTIONS

Dr Austin R Brown

Introduction

You work for a digital marketing agency. This agency wants to determine the effectiveness of two different types of Instagram ads on user engagement. They want to know if the ad format (image or video) and the ad content (funny or promotional) have an impact on the amount of time viewing the ad. To evaluate this, they recruit 100 participants and randomly assign them to one of the four conditions: image-funny, image-promotional, video-funny, or video-promotional. The participants are then asked to view the ad and the time they spend viewing the ad (in seconds) is recorded. The data are contained in the `Instagram Data.xlsx` file. With these data, I want you to:

1. Briefly define the objective of this experiment

The objective of this experiment is to determine if Instagram ad format and/or content is associated with differing viewing durations.

2. Specify the outcome variable

Time spent viewing (in seconds)

3. Specify the independent variables. What are some other possible lurking variables?

Format and Content. Age and educational attainment level may be lurking variables.

4. State all of the pairs of null and alternative hypotheses for this experiment.

$$H_0 : \alpha_{\text{Video}} = \alpha_{\text{Image}} = 0 \quad H_1 : \text{At least one } \alpha_i \neq 0$$

$$H_0 : \beta_{\text{Funny}} = \beta_{\text{Info}} = 0 \quad H_1 : \text{At least one } \beta_j \neq 0$$

$$H_0 : (\alpha\beta)_{VF} = (\alpha\beta)_{VI} = (\alpha\beta)_{IF} = (\alpha\beta)_{II} = 0 \quad H_1 : \text{At least one } (\alpha\beta)_{ij} \neq 0$$

5. Perform appropriate exploratory analysis, including data visualizations. Do the results of these analyses support the null or alternative hypotheses more strongly?

```
library(tidyverse)
library(readxl)
library(rstatix)
## Read in Data ##
insta <- read_excel("Instagram Data.xlsx")
## Check Means/SDs of Main Effects ##
## Format ##
insta |>
  group_by(Format) |>
  get_summary_stats(Time,type='mean_sd') |>
  select(Format,mean,sd)
```

```
# A tibble: 2 x 3
  Format  mean    sd
  <chr>  <dbl> <dbl>
1 Image   7.96  5.51
2 Video  27.0  11.1
```

```
## Content ##
insta |>
  group_by(Content) |>
  get_summary_stats(Time,type='mean_sd') |>
  select(Content,mean,sd)
```

```
# A tibble: 2 x 3
  Content        mean    sd
  <chr>         <dbl> <dbl>
1 Funny          24.0  14.1
2 Informational  10.9  7.30
```

```
## Interaction ##
insta |>
  group_by(Format,Content) |>
  get_summary_stats(Time,type='mean_sd') |>
  select(Format,Content,mean,sd)
```

```
# A tibble: 4 x 4
  Format Content         mean    sd
  <chr>  <chr>          <dbl> <dbl>
1 Image  Funny          11.2   5.51
2 Image  Informational  4.72   3.13
3 Video  Funny          36.8   5.60
4 Video  Informational 17.2    4.27
```

Examining Format, we can see that videos are viewed for quite a bit longer compared to images. Additionally, funny content seems to be viewed for meaningfully longer than informational content. We may also have evidence for an interaction effect since the difference between content within each format are a little different. We can observe this in our interaction plot and boxplot below.

```
## Boxplot - Interaction Plot Combo ##
insta |>
  ggplot(aes(x=Format)) +
  geom_boxplot(aes(y=Time,fill=Content)) +
  geom_point(data = insta |>
               group_by(Format,Content) |>
               get_summary_stats(Time,type='mean_sd'),
             aes(x=Format,y=mean,group=Content),
             shape=4,size=5) +
  geom_line(data = insta |>
               group_by(Format,Content) |>
               get_summary_stats(Time,type='mean_sd'),
             aes(x=Format,y=mean,group=Content,
                 color=Content)) +
  labs(title = "Boxplot/Interaction Combo Plot") +
  theme_classic() +
  theme(plot.title = element_text(hjust=0.5,face='bold'))
```

Boxplot/Interaction Combo Plot
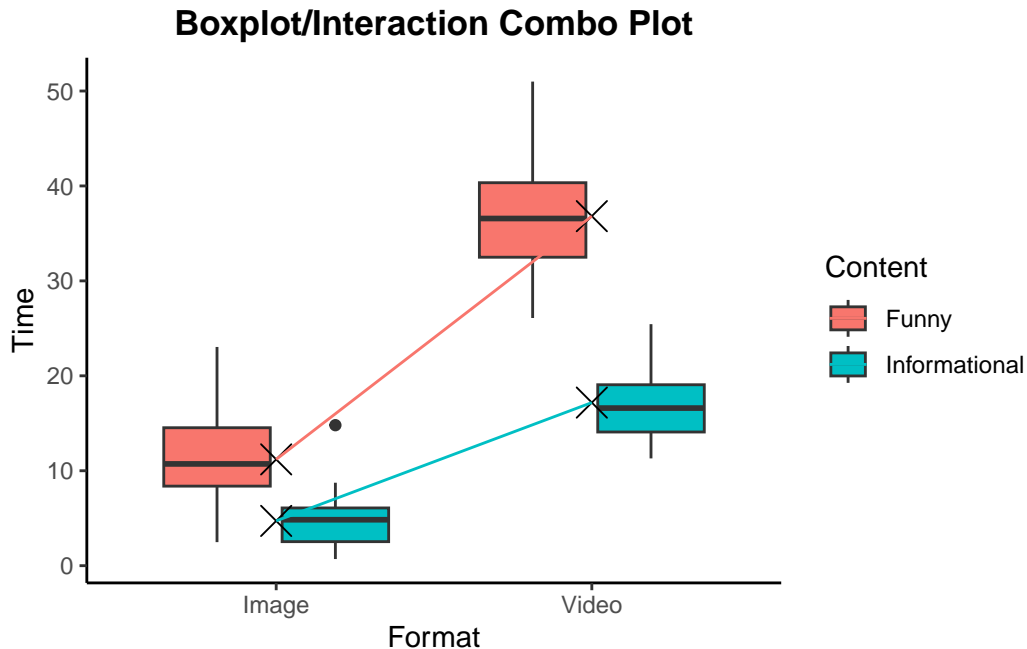
In this combo plot, we can clearly see a separation between format types, which would support the Format treatment's alternative hypothesis. Within formats, we can also see a separation between informational ads and funny ads which would provide support for the Content treatment's alternative hypothesis. Additionally, since the interaction lines are not parallel, this implies that the differences in content by format type may be meaningfully different. Thus, we may have evidence for a significant interaction effect.

6. Build a two-way ANOVA model including the two-way interaction. Test the assumption of normality. Do the results of the normality test(s) support the assumption of normality?

```r
## Build Two-Way ANOVA with Interaction ##
mod <- aov(Time~Format*Content,data=insta)
## Assess Normality Assumption ##
## QQ-Plot ##
insta |>
  ggplot(aes(sample=resid(mod))) +
  geom_qq() +
  geom_qq_line() +
  theme_classic()
```
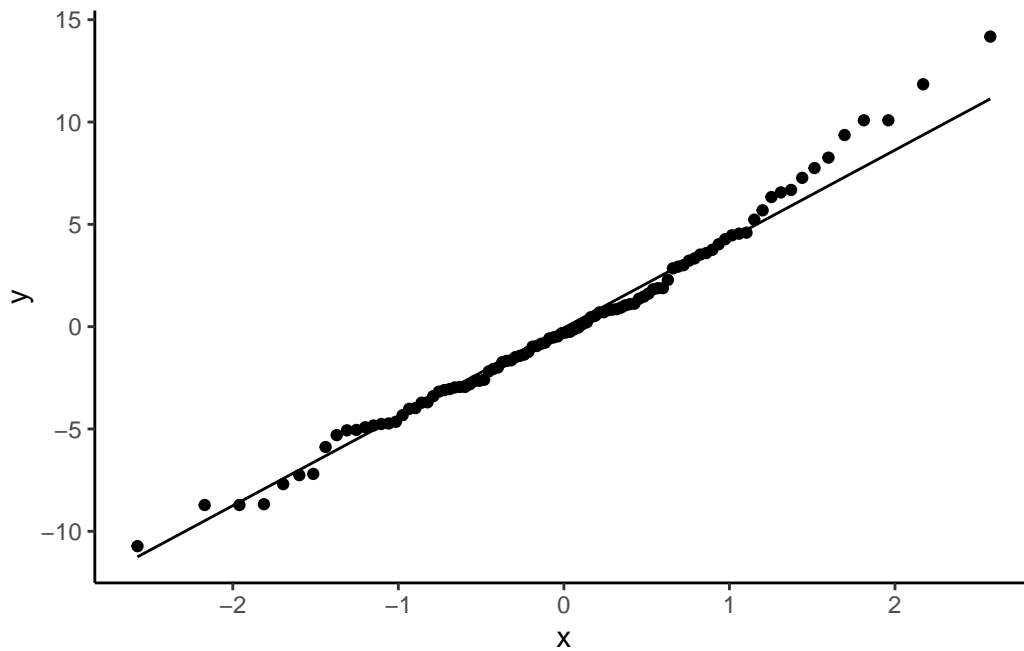
```
## Shapiro-Wilk Test ##
mod |>
  resid() |>
  shapiro_test()
```

```
# A tibble: 1 x 3
  variable    statistic p.value
  <chr>           <dbl>   <dbl>
1 resid(mod)      0.984   0.250
```

Since most of the points on the QQ-plot fall along the diagonal line, and since the p-value associated with the S-W test gives stronger evidence in favor of the null hypothesis, we can reasonably conclude the normality assumption is met.

7. Test the assumption of homogeneity of variance. Do the results of the test(s) support the assumption of homogeneity of variance?

```
## Constant Variance ##
## Fitted vs Standardized Residuals ##
insta |>
  ggplot(aes(x=fitted(mod),y=rstandard(mod))) +
  geom_point() +
  geom_hline(yintercept=0,color='blue') +
  geom_hline(yintercept=3,color='red') +
  geom_hline(yintercept=-3,color='red') +
  theme_classic()
```
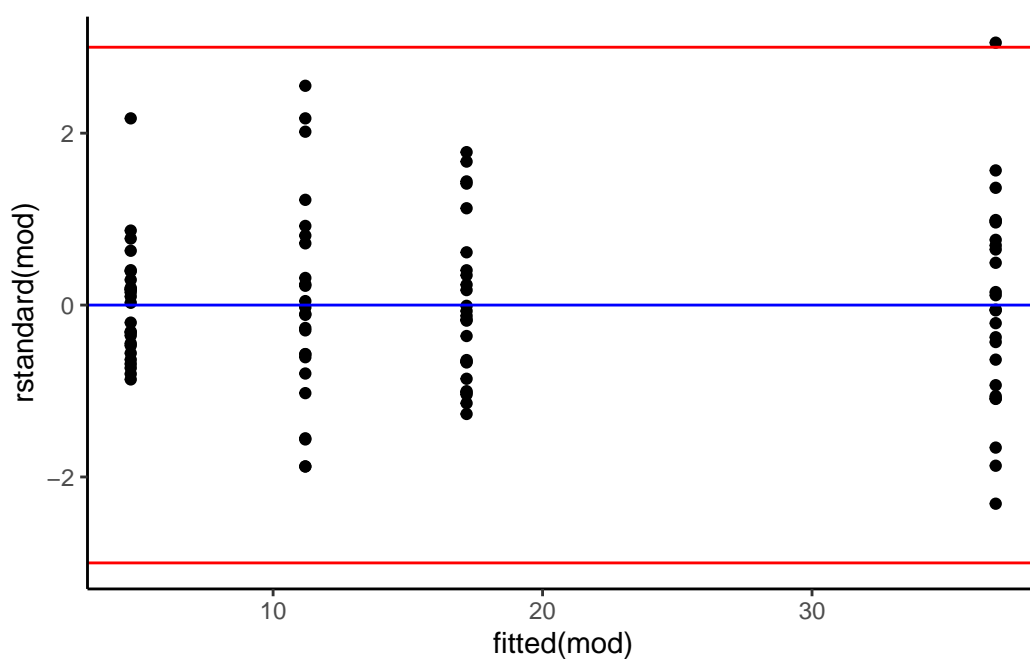


```
## BP Test ##
library(broom)
library(lmtest)
mod |>
  bptest() |>
  tidy()
```

```
# A tibble: 1 x 4
  statistic p.value parameter method
```

```
       <dbl>    <dbl>      <dbl> <chr>
1      6.68   0.0830          3 studentized Breusch-Pagan test
```

The plot of the fitted values versus the standardized residuals shows us that our four groups are relatively similar in length and centered about 0 with the exception of the last group which has potential evidence of an outlier (above +3). The B-P test gives us evidence in favor of the null hypothesis meaning that we can reasonably assume constant/equal variance between our groups.

8. Report the F-statistic for the interaction effect and its associated p-value from the two-way ANOVA model. Which of our two hypotheses is more strongly supported? Why?

9. Report the F-statistics for the main effects and their associated p-values from the two-way ANOVA model. Which of our two hypotheses is more strongly supported? Why?

```
mod |>
  tidy()
```

```
# A tibble: 4 x 6
  term              df sumsq meansq statistic   p.value
  <chr>          <dbl> <dbl>  <dbl>     <dbl>     <dbl>
1 Format             1 9065. 9065.       404.  3.62e-36
2 Content            1 4263. 4263.       190.  1.76e-24
3 Format:Content     1 1082. 1082.      48.2   4.55e-10
4 Residuals         96 2155.   22.4       NA    NA
```

Starting with the interaction effect, $F(1, 96) = 48.1989$ with an associated p-value far less than the typically used 0.05 threshold. This is a statistical confirmation of what we saw in the interaction plot. The same can be said of the test for Content ($F(1, 96) = 189.91$, $p < 0.05$) and Format ($F(1, 96) = 403.87$, $p < 0.05$). In all cases, the alternative hypothesis is supported.

10. If the data more strongly support the alternative hypothesis in any of the above F-tests, perform Tukey's HSD post-hoc test to determine which groups significantly different from each other. If the data more strongly support the null hypothesis, explain why a post-hoc test would not be appropriate.

```
TukeyHSD(mod) |>
  tidy() |>
  select(contrast,estimate,adj.p.value)
```

```
# A tibble: 8 x 3
  contrast                              estimate adj.p.value
  <chr>                                    <dbl>       <dbl>
1 Video-Image                               19.0     4.28e-10
2 Informational-Funny                      -13.1     4.28e-10
3 Video:Funny-Image:Funny                   25.6     4.28e-10
4 Image:Informational-Image:Funny          -6.48     2.97e- 5
5 Video:Informational-Image:Funny           5.98     1.27e- 4
6 Image:Informational-Video:Funny         -32.1     4.28e-10
7 Video:Informational-Video:Funny         -19.6     4.28e-10
8 Video:Informational-Image:Informational   12.5     4.28e-10
```

Here we can see that:
- Main Effect: Videos were viewed for significantly longer than Images
- Main Effect: Funny content was viewed for significantly longer than Informational content
- Interaction Effect: Funny videos were viewed for significantly longer than Funny images
- Interaction Effect: Funny images were viewed for significantly longer than Informational images
- Interaction Effect: Informational videos were viewed for significantly longer than Funny images (interesting!)
- Interaction Effect: Funny videos were viewed for significantly longer than Informational images
- Interaction Effect: Funny videos were viewed for significantly longer than Informational videos
- Interaction Effect: Informational videos were viewed for significantly longer than Informational images.

11. Write a brief conclusion summarizing the results of your analyses.

To conclude, in our experiment, we found that videos and funny content tended to be viewed for longer than images and informational content. However, the combination of a video also being funny in content led to the longest viewed ads, indicating a significant interaction between Content and Format. This result obviously doesn't account for potential lurking variables related to characteristics of the viewer, but does offer insight into ads that receive the most attention.