Edgar Campos

Data Assignment 2

Case: Webflicks

Report

To find out if a certain release schedule or genre of show impacts views, we employ power analysis to figure out how many shows we'll need to allocate to each combination of genre and release schedule. Power analysis also helps to interpret our results, as it brings into context practically significant differences, such as a certain release having a practically significant increase in views compared to shows with different release schedules. Designing our experiment, we believe that a Factorial Design will likely be the best, as we have two factors we're testing, Genre and Release Schedule, and we have a response variable to test them, each show's respective Views. We are also interested in figuring out if there is an interaction effect, meaning that there may be a certain combination of genre and release schedule for a show that affects views. To put this into context, "Do Mystery shows see higher views with a Drop schedule as viewers are intrigued to figure out what the mystery is?", or other related questions can be answered with interaction.

After our power analysis, we found that we'll need 240 shows for a properly powered experiment (see Figure 1 of the Appendix). This is considerably higher than the 170 shows that are within our budget, and we should consider the potential increase in revenue gained from understanding which genre of shows see higher views with different release schedules, and the costs of increasing our sample of shows by 70.

For this experiment, we will continue with our budgeting constraints as higher power is necessary for clinical studies to detect meaningful differences, for example, but as this experiment is more about social sciences, a lower power will likely be sufficient. To implement our design of planned random assignment, we'll need 168 shows, 14 shows for each combination of genre and release schedule. It should be noted that this design does not consider other external factors that may affect view count, such as demographic information, popularity of individual shows/actors, or even quality of shows (can be measured by reviews). These other factors can be studied in a future analysis.

After running our experiment, we use Two-Way ANOVA with Interaction to interpret our data (see assumption checking in Figure 2). At the standard 5% significance level, we find no significant evidence of an interaction effect between Genre and Release Schedule or evidence of Genre affecting views (see Figure 3). However, we do see evidence of Release Schedule impacting views on shows. For determining which release schedule creates the highest views, we employ post-hoc analysis conducting more Two-Way ANOVA analyses with subsets of the data, correcting for multiple testing problem with a simple Bonferroni correction. After correcting for multiple tests, at the 5% significance level we find that Dual Drop is most different from the other two release schedules (see Figure 4). We now use some simple Exploratory Data Analysis to see in what way Dual Drop is different from the other two (Figure 5).

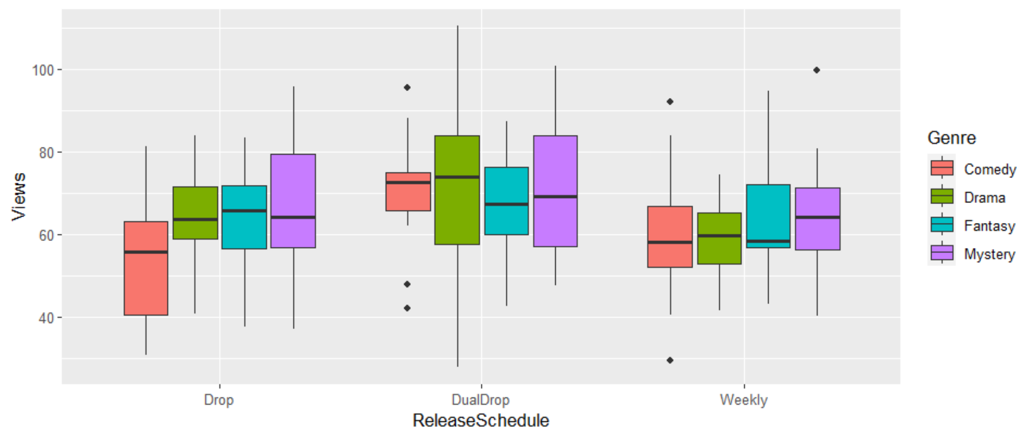

Figure 5: Boxplot of shows by their views, grouped by release schedule and genre.

As seen in the boxplot, Dual Drop sees higher views among the different genres of shows compared to both Drop and Weekly releases. We conclude that Webflicks should implement a Dual Drop release schedule for their shows to increase their views.

Appendix

This appendix contains additional figures detailing data summarization and calculations. The graphs in Figure 2 are all for checking assumptions of our test. All output provided is produced from RStudio.

```
> pwr.f2.test(u = (3-1)*(4-1), v = NULL, f2= (.25)^2,
+             sig.level = 0.05, power = 0.8)

     Multiple regression power calculation 

              u = 6
              v = 217.3637
             f2 = 0.0625
      sig.level = 0.05
          power = 0.8
```

Figure 1: Results of power analysis using pwr.f2.test in the pwr library.

Figure 1 shows the results of our power analysis, using the standard effect size of .25 in ANOVA, standard significance level of 5% and power of 80%. With our denominator degrees of freedom being 217.36, we find n to be 19.11, meaning that for our analysis we will need 20 subjects per combination. As we have 12 combinations, with two factors of 3 and 4 factor levels respectively, we find that we'll need a total sample size of 240. As our maximum total sample size is 170, we'll have to use 14 per combination, for a total of 168, which will lower the power of our analysis. We can calculate the power of this experiment, but retrospective power analyses are generally frowned upon. While they have their place in evaluating the studies of others, a retrospective power analysis will not help us here.

For the assumption check of our Two-Way ANOVA with analysis, we need to ensure:

1. Independence of Observations
2. Normality of Residuals
3. Constant standard deviation

For our first assumption, we allocated each show randomly using planned random assignment (as shown in Figure 2A), so there is no reason to think that the independence assumption is violated, thus concluding that the assumption is likely reasonable.

```
#planned random assignment
Assignments <-data.table(A = rep(c("Drop", "Weekly", "DualDrop"),
                                each = 168/3, times = 1),
                         B = rep(c("Comedy","Drama","Mystery", "Fantasy"),
                                each = 168/(4*3), times = 3))
Assignments <- Assignments[sample(1:168)]
Webflicks$ReleaseSchedule <- Assignments[,A]
Webflicks$Genre <- Assignments[,B]
```

Figure 2A: Implementation of Planned Random Assignment for experiment

For our normality assumption, we examine the Quantile-Quantile plots of the residuals from our model fit. Looking at our QQplot below (Figure 2B), we see a fairly straight line so we can conclude the normality assumption is likely reasonable.

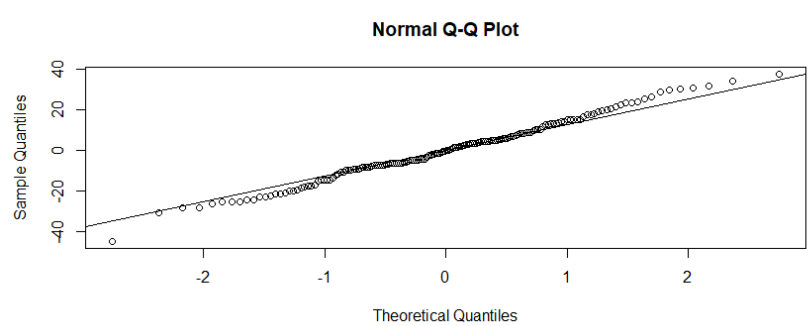

Figure 2B: Quantile-Quantile plot of residuals

We evaluate homogeneity with a residual vs. predicted values plot. Looking at our plot below (Figure 2C), we see that the spread is fairly constant, so the assumption is likely reasonable.

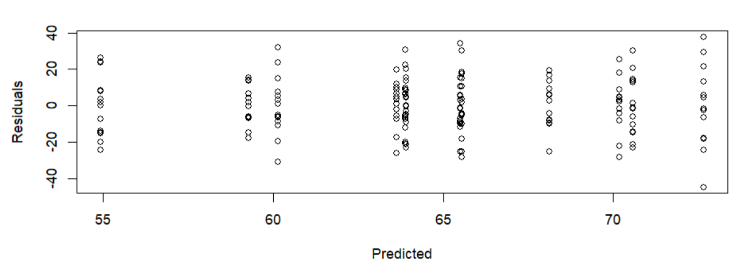

Figure 2C: Predicted vs Residual plot

To carry out our Two-Way ANOVA with interaction analysis in R, we use the aov() and summary() functions to give us the p-values for the interaction effect and the two different factors. At the 5% significance level, we find that there is no statistically significant interaction effect between Release Schedule and Genre (p = 0.73). We also find no statistically significant difference in views between the Genres after accounting for Release Schedule (p = 0.43). However, we find a statistically significant difference in views with the different Release Schedules.

```
> fit <- aov(Views~ReleaseSchedule * Genre,
+            data=Webflicks)
> summary(fit)
                       Df Sum Sq Mean Sq F value  Pr(>F)   
ReleaseSchedule         2   2561  1280.5   5.509 0.00488 **
Genre                   3    650   216.6   0.932 0.42700   
ReleaseSchedule:Genre   6    830   138.3   0.595 0.73404   
Residuals             156  36264   232.5                   
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 3: Output of Two-Way ANOVA with interaction analysis

To carry out our post-Hoc analysis, as we found not a significant interaction effect, we run Two-Way ANOVA analyses with subsets of the data to figure out which factor level is different from the others. We do this in R using %in% with our data table. To mitigate the multiple testing problem, we use a simple Bonferroni correction to adjust our p-values. From our analysis in Figure 4, we find that Drop and Weekly release schedules are not statistically different, while Dual Drop is different from both at the 5% significance level.

```
Webflicks<- as.data.table(Webflicks)
summary(aov(Views ~ ReleaseSchedule * Genre,
            data = Webflicks[ReleaseSchedule %in% c("DualDrop","Drop")]))
#DualDrop-Weekly uncorrected p = 0.00597 (corrected = 0.01791)
#DualDrop-Drop uncorrected p = 0.00582 (corrected = 0.01746)
#Drop-Weekly uncorrected p = 0.944 (corrected = 1)
```

Figure 4: Post hoc analysis using %in%