The Great British Bake-Off

Kaatje Matthews-vanKoetsveld[1], Alie Hall[1], Olivia Kurani[1], Hailey Ernest[1]

[1]Department of Statistics, Cal Poly - SLO

May 31, 2024

Abstract

Throughout each episode of the series in 'The Great British Bake Off', there are varying viewership trends. Our analyses show evidence that the finale episode of each series has significantly more viewers than any other episode, with non-finale episodes having between 7.75 to 9.09 million viewers and finale episodes having 7.27 to 12.51 viewers. During series 8 of 'The Great British Bake Off', the popular cooking show underwent a few production changes that ultimately had a significant consequence on the viewership of that series and the ones to follow. Our analysis of series 8, 9, and 10 compared to the earlier seasons shows that there is a considerable decline in viewership compared to the earlier series. These results can give producers of the show an idea of what will best draw the attention of their audience and make adjustments to up the viewership for 'The Great British Bake Off'.

1 Introduction

Season after season, the "Great British Bake Off" is loved by its fans – but are all series and episodes of the show created equal? Or is there a secret ingredient to the show's success? "The Great British Bake Off" (GBBO) is a popular baking show that has run from 2010 to present, and features ten to twelve amateur bakers that hail from the Great Britain area. We examined two different trends within viewership of this show. The first investigates if there is a statistically significant difference in finale viewership compared to other episodes. The second discusses the series 8 changes within production of the show and consequences on viewership. In order to gain a better understanding of these questions, we used a t-test and linear models to look at datasets created by Alison Hill, Chester Ismay, and Richard Iannone, consisting of data collected from Wikipedia and PBS.

2 Data and Methods

We sourced our data on The Great British Bakeoff from TidyTuesday (Wickham et al. 2019), an online community that posts weekly datasets to a public repository for people to share their visualizations and analyses. Included in The Great British Bakeoff package (Hill, Ismay, and Iannone 2022), posted on October 25, 2022, are four datasets, consisting of data from Wikipedia and PBS collected by Alison Hill, Chester Ismay, and Richard Iannone. The data includes information on individual episode challenges, the bakers themselves, the episode ratings, and further information on the episodes. This project only utilizes the "Bakers'' and the "Ratings'' datasets. All of our data analysis will be done in R Version 4.3.2 (R Core Team 2023) The Bakers dataset includes each of the 120 bakers that have appeared on the show from seasons 1-10 (2010-2019) as an observation and contains some basic personal information such as age and occupation, and multiple other quantitative variables regarding their baking challenge scores from their season. The Ratings dataset has each of the 94 aired episodes from seasons 1-10 as observations and records the episode's viewers in the millions for the first 7 days and the first 28 days. Specifically, we are interested in the weekly episode viewership for each episode.

2.1 Data Cleaning and Manipulation

Although we did not truly manipulate any of our variables, we did add one very important indicator variable. This indicator variable was used to determine if an episode was a finale episode or not, which we determined by grouping our data by series, arranging by episode number and then mutating to say finale = TRUE.

In addition, at various points within our analysis, we subsetted different series of our data. To create our reduced linear model, for example, we subsetted the data to only include data from series 1-7.

2.2 Statistical Analysis

We used many different methods of analysis. Firstly, this included a t-test to determine if there was a statistically significant difference in viewership between finale and non-finale episodes. For this t-test, we examined a p-value and a confidence interval of the difference in means. Next, we applied two linear models to our data, one to a subset of our data from the first seven series, and a second one onto the full data. As part of this analysis, we looked at the comparative r-squared values for both of the linear models, and the p-value from a model utility test. We also compared the coefficients for the intercepts and slopes of the models.

3 Results

```
    Welch Two Sample t-test

data:  ratings_new$viewers_7day and ratings_new$finale
t = 25.036, df = 94.675, p-value < 2.2e-16
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 7.911036      Inf
sample estimates:
mean of x mean of y
 8.579606  0.106383
```
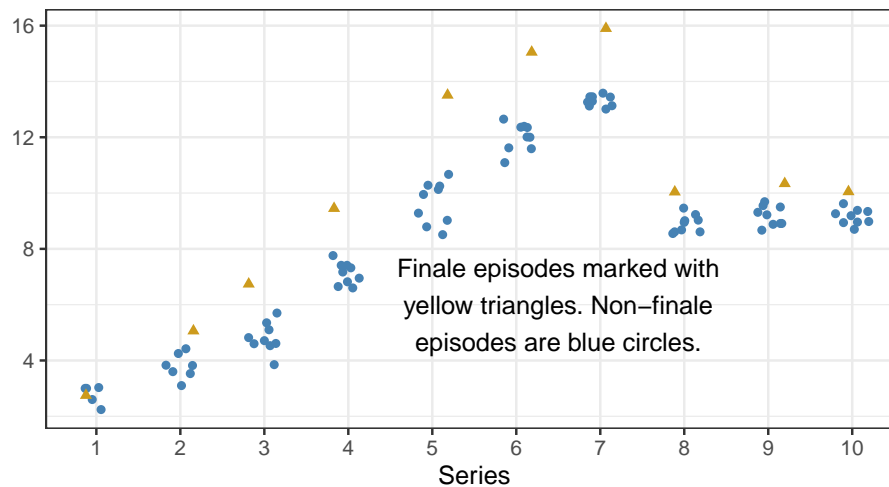

Figure 1: This graph displays the viewership by series of the Great British Bake Off, with blue dots representing the viewership for non-finale episodes and yellow triangles viewership of finale episodes.

Table 1: This table displays summary statisitcs for non-finale episodes and finale episodes. These summary statistics are as follows: mean, standard deviation, median, quartile 1 and quartile 3.

Table 1: Summary Stats for Non Finale and Finale Episodes

type	mean	sd	median	q1	q3
Non Finale Episodes	8.423726	3.129113	8.945	6.3750	10.1600
Finale Episodes	9.889000	4.223286	10.045	7.4175	12.7175

Viewership after 7 days of Episode Release by Series with Linear

Note: Our Linear Models have been projected over this graph for contrast
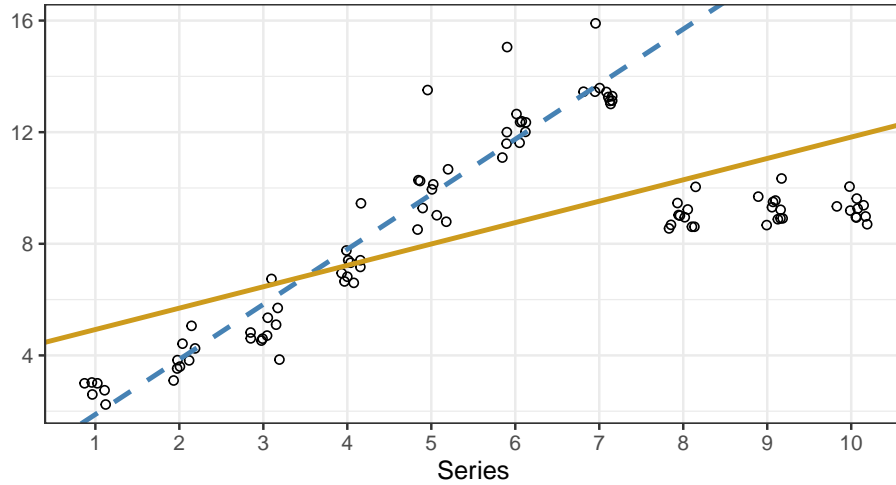

Figure 2: Graph of viewership for every episode with two lines projected onto it. Blue line is projected onto the graph showing the subsetted linear model. Yellow line is projected onto the line showing the full linear model. This graph shows how the subsetting linear model fits the first seven series extremely well, showing that series 8, 9, and 10 have a different trend.

r.squared	p.value	type
0.4218684	0	full
0.9306152	0	reduced

4

4 Discussion

There is statistically significant evidence that the mean viewership 7 days after an episode airs is statistically different for finale and non-finale episodes. However, the confidence interval overlaps, practically making the mean viewership between the finale and non-finale insignificantly different. We can conclude that there tends to be a correlation but not causation between mean viewership of finale and non-finale episodes.

The contrast in viewership in series 8 and beyond compared to preceding series is also a critical detail that producers of 'The Great British Bake Off' should take into consideration. Given that the producers make the decision of who the MC's and judges of the show should be, they should be aware of the consequence these decisions have on viewership. This statistically significant difference in viewership in these series can be translated to a more practical significance as well. If viewership continues to stay at this decreased levels, the producers of the show need to know how much this will affect the show's revenue. Because this is an analysis based on observations, we cannot fully conclude causation between the production changes and decline in viewership, however we can see that there is a strong correlation between the two.

5 Conclusion

All in all, our analyses have shown that finale episodes of series have a significantly higher viewership than other episodes, and that since the production changes during series 8, the overall viewership of 'The Great British Bake Off' has declined. These differences in viewership have real consequences for the overall revenue of the show and how budgets are created for future series. Producers can take extra care when planning finale episodes, since that is when the largest audience is watching. Additionally, producers can try to negotiate for better airing times in an effort to bring back up viewership throughout the series. In the future, it would be intriguing to investigate the ratings based on seasons or baker backgrounds. Researching the baker's background affects the show's success will provide insight into the show's winners.

References

Hill, Alison, Chester Ismay, and Richard Iannone. 2022. *Bakeoff: Data from "the Great British Bake Off"*. https://CRAN.R-project.org/package=bakeoff.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019.

"Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.