Lucasfilm vs. Disney: Who Did Star Wars Better?

Aaron Brown, Aly Milne, and andrew Bargeron

2023-12-08

Introduction

Who shot first, Han or Greedo? Which is better, the Prequels or the Sequels? Obviously, there are plenty of controversies relating to the Star Wars franchise within its community of fans. However, amongst all these debates, we aim to answer the most disputed and relevant issue at this time, which is whether Star Wars has improved since Disney acquired it from Lucasfilm. We look to explore the differences in the movies and show ratings of the Star Wars franchise before and after the ownership was passed on to Disney. We will divide the data into sub-populations: one being all the films and shows released prior to 2013, and the other being all movies and shows from 2013 onwards. We will look at the mean and variance of the two sub-populations. The parameters we are most interested in are the mean ratings of movies and shows while the franchise was still under the influence of George Lucas' creative mind and Lucasfilm, denoted as μ_L, the mean ratings of movies and shows after Disney became the new owner, denoted as μ_D, their standard deviations, and the difference parameter.

Methods

We found most of the data already pulled from IMDb on data.world. However, some of the newer shows (Ahoksa, Book of Boba Fett) were not included, so we added those manually by checking the ratings for each episode, weighting them by the number of reviews, and aggregating to the total average rating for the season. We also aggregated the episodes in the pre-pulled dataset up to their seasons as well (for TV series). Lastly, we updated the years for each season as they were left as the total range of the entire series (ex. Clone Wars was startYear=2008 and endYear=2020 across all seasons). After bringing the cleaned data into R and splitting it by years (LucasFilm is up to 2012, Disney is 2013 and beyond), we constructed histograms and Q-Q plots of the data to confirm normality. While the histogram of pre-2013 (LucasFilm) was a little bimodal, the Q-Q plot ensured reasonable normality. Both histogram and Q-Q plot for post-2012 (Disney) confirmed normality as good as we could expect for a small dataset.

We used the Normal distribution for our likelihood because the average ratings are continuous and roughly normally distributed. We used the Normal distribution for the prior of the average ratings and Inverse Gamma distribution for the prior of variance in those ratings. Normal was appropriate since we expected the quality would be consistent and centered around a specific value for both groups. Inverse gamma was appropriate because it is the conjugate prior and it handles the positive-only nature of variance well. For our priors on the mean for both datasets, we used λ=8 and τ^2=20 for a Normal distribution, both twice the priors from our Goodreads activity since these ratings were on a scale twice as large (1-5 vs. 1-10). We selected a large value for τ^2 since there is a fair amount of uncertainty in this data. Additionally, we used in inverse gamma distribution with γ=2.01 and ϕ=0.66 as the priors for both datasets to model variance. Thus, for the two populations (LucasFIlm and Disney), we have unknown parameters μ (average rating) and σ^2 (variance of the ratings).

$\pi(\mu)$ $\pi(\sigma^2)$
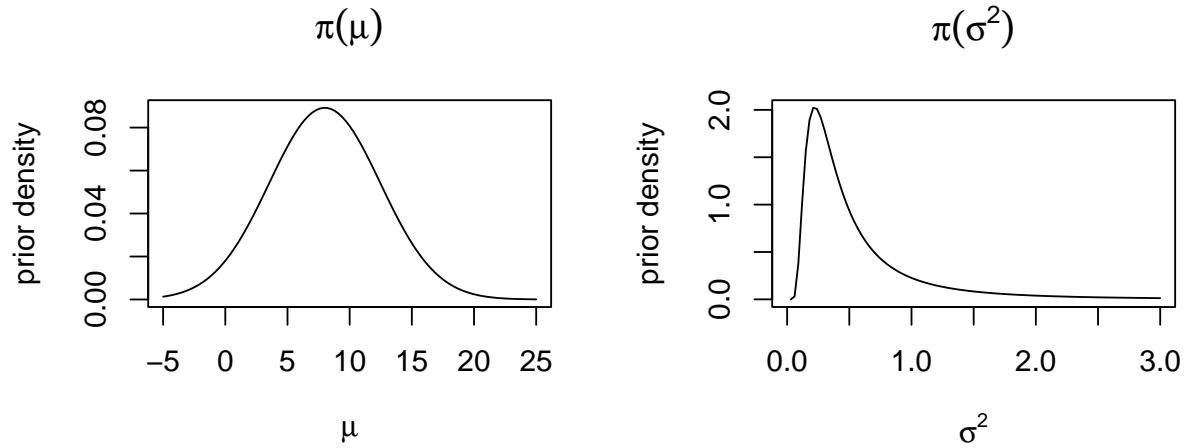
We used Gibbs sampling to estimate the posterior distributions of the mean and variance of both groups. We initialized with the sample mean and variance, then iteratively updated these parameters for 10,000 iterations. We specifically sampled new values for the mean and variance in each iteration based on their respective full conditional distributions, progressively refining these estimates to reflect the underlying data distribution.

Results

The average rating for LucasFilm content was 7.3326 and the variance was 2.2514.

$\pi(\mu_L \mid \text{data})$ $\pi(\sigma_L^2 \mid \text{data})$
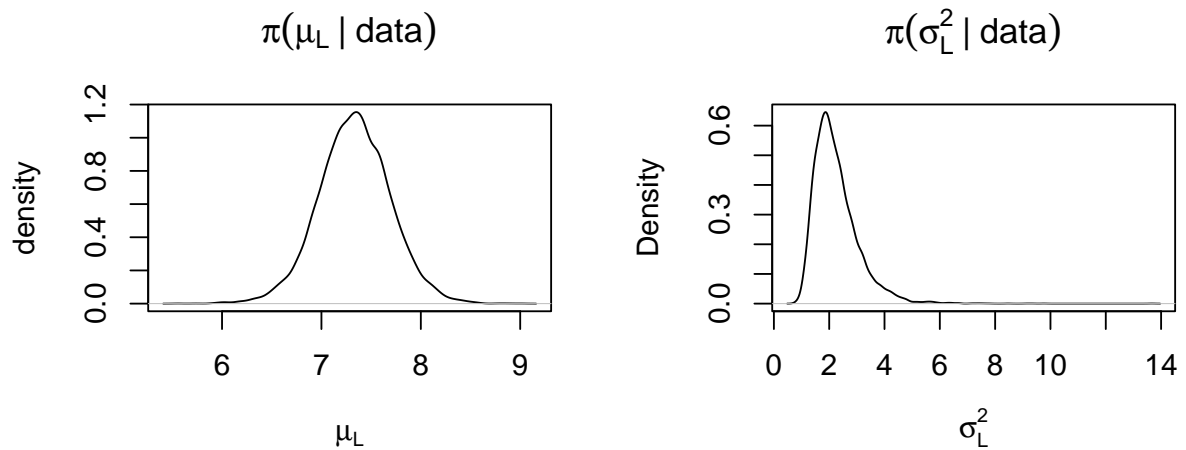
For Disney, these values were 7.7391 and 0.5629.

$\pi(\mu_D \mid \text{data})$

$\pi(\sigma_D^2 \mid \text{data})$