My title*

My subtitle if needed

Tam Ly

April 13, 2024

First sentence. Second sentence. Third sentence. Fourth sentence.

Table of contents

1 Introduction **2**

2 Data **2**
2.1 Source . . . 2
2.2 Methodology . . . 2
2.3 Variables . . . 2
2.4 Measurement . . . 2

3 Model **2**
3.1 Model set-up . . . 2
3.2 Model justification . . . 3

4 Results **4**

5 Discussion **4**
5.1 First discussion point . . . 4
5.2 Second discussion point . . . 4
5.3 Third discussion point . . . 4
5.4 Weaknesses and next steps . . . 4

Appendix **5**

A Model details **5**
A.1 Posterior predictive check . . . 5

*Code and data are available at: https://github.com/atn-ly/beyonce

A.2 Diagnostics 5

B References **6**

1 Introduction

2 Data

2.1 Source

2.2 Methodology

2.3 Variables

2.4 Measurement

3 Model

The goal of our modeling strategy is to predict the number of Spotify streams for a Beyoncé song based on the number of weeks it spent on the *Billboard Hot 100*. We used a negative binomial regression model in a Bayesian framework. Negative binomial regression is a type of generalized linear model that is useful for modeling count data.

3.1 Model set-up

The model that we are interested in is:

$$
\begin{aligned}
y_i|\mu_i, r &\sim \text{NegBinom}(\mu_i, r) && (1)\\
\log(\mu_i) &= \alpha + \beta \times \text{Number of weeks}_i && (2)\\
\alpha &\sim \text{Normal}(0, 2.5) && (3)\\
\beta &\sim \text{Normal}(0, 2.5) && (4)\\
& && (5)
\end{aligned}
$$

Where:

- y_i is the outcome variable, representing the number of Spotify streams for song i,

- μ_i is a parameter for the negative binomial distribution, representing the probability of success in a single trial,
- r is a parameter for the negative binomial distribution, representing the number of successes,
- Number of weeks$_i$ is the predictor variable, representing the number of weeks spent on the Billboard Hot 100 for song i,
- α is a parameter, representing the intercept with a default prior probability distribution that is Normal with a mean of 0 and standard deviation of 2.5,
- β is a parameter, representing the slope coefficient with a default prior probability distribution that is Normal with a mean of 0 and standard deviation of 2.5.

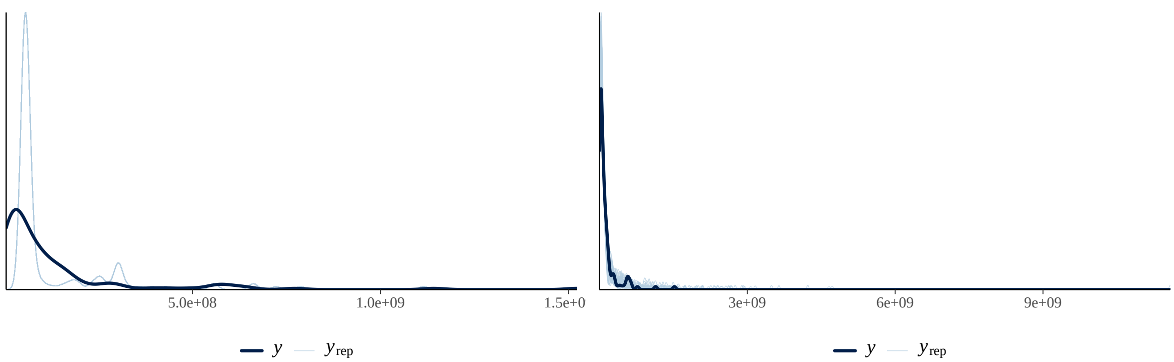
3.2 Model justification

We expect a positive relationship between the number of Spotify streams and weeks spent on the *Billboard Hot 100* based on the positive correlation we observed in the graph in Section 2.

Negative binomial regression operates under several assumptions. It assumes linearity between the outcome and predictor variables, independence of observations, and no multicollinearity.

We considered Poisson regression as an alternative model since it is also used for count data. However, one of the restrictions with Poisson regression is that it assumes equal mean and variance. Negative binomial regression relaxes this assumption to allow for over-dispersion. We fitted both and compared them using posterior predictive checks in Figure 1. We see that the negative binomial approach does a better job of fitting the data.

We considered additional model checking and diagnostic issues. Details and graphs can be found in Section A.



(a) Poisson model

(b) Negative binomial model

Figure 1: Comparing posterior prediction checks for Poisson and negative binomial models

4 Results

5 Discussion

5.1 First discussion point

5.2 Second discussion point

5.3 Third discussion point

5.4 Weaknesses and next steps

Appendix

A Model details

A.1 Posterior predictive check

In Figure 2 we compare the posterior with the prior. It shows that estimates of our intercept changed once the data was taken into account. This suggests that there is an issue with the default prior we specified for α.

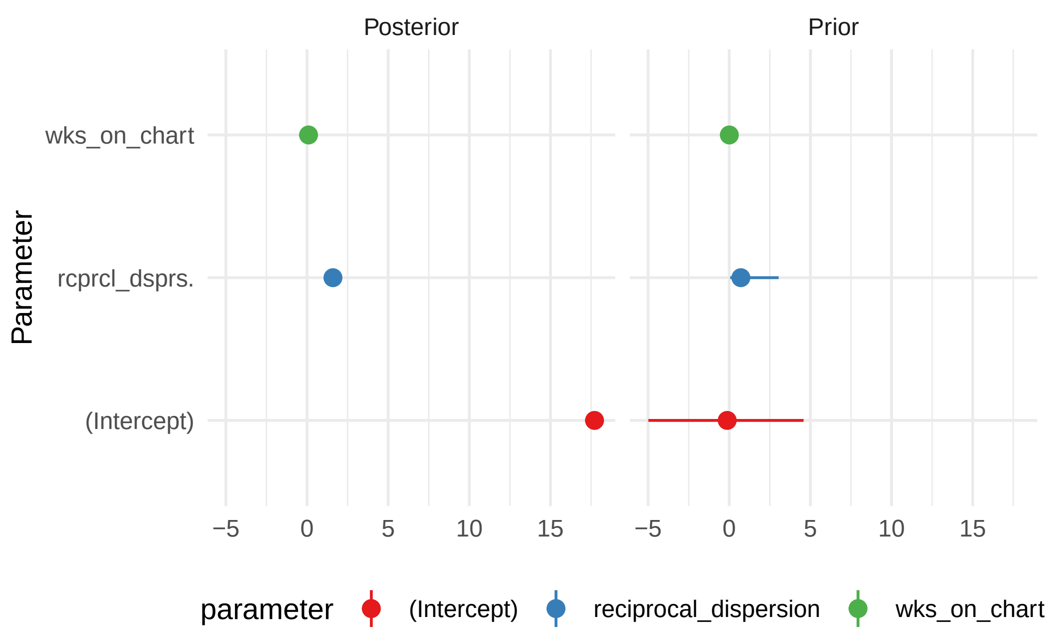
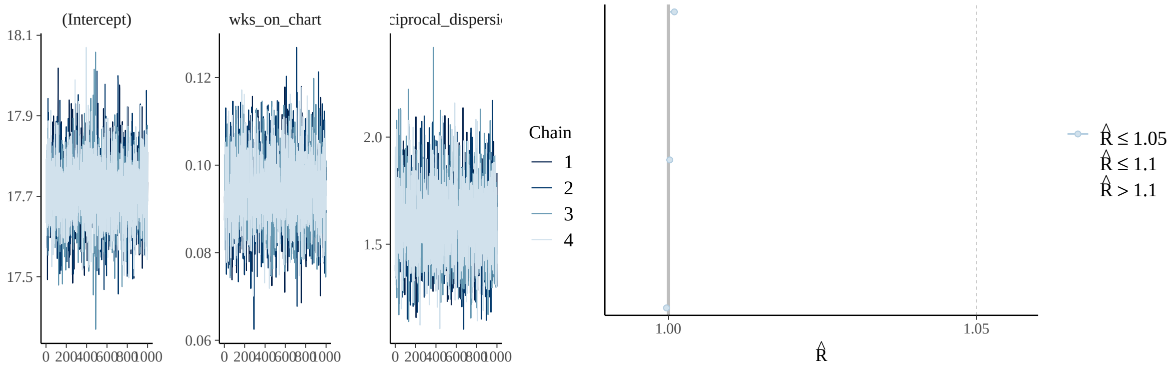


Figure 2: Comparing the posterior with the prior

A.2 Diagnostics

Figure 3a is a trace plot. It shows that there are no horizontal lines that appear to bounce around and have a nice overlap between the chains. This does not suggest anything out of the ordinary.

Figure 3b is a Rhat plot. It shows that everything is close to 1 and no more than 1.1. This does not suggest any problems.



(a) Trace plot (b) Rhat plot

Figure 3: Checking the convergence of the MCMC algorithm

B References