

ROUGH DRAFT: Predicting Measles and Rubella cases simultaneously using Bayesian Multilevel Models

Andrew Ross

June 27th, 2025

Introduction

Pediatricians often recommend measles-mumps-rubella vaccines in children. However, some countries do not have access to the vaccine. Because of this measles. In this paper, we utilize the `brms` package in R to implement a Bayesian multilevel Poisson model to predict the number of measles cases for 2025 based on different variables.

Data

The data was received by Jen Richmond from the World Health Organization and sent to TidyTuesday for analysis. TidyTuesday is an open source data repository and collaboration where analysts can practice their data skills [Community, 2024]. There are two datasets: one has information based on a given year and the other has similar data but broken down into months. For this analysis, we use the first dataset. The data dictionary can be found here: <https://github.com/rfordatascience/tidytuesday/blob/main/data/2025/2025-06-24/readme.md>. For completeness, we remove all missing information.

Our response variable is the number measles, and figure 1 shows the histogram. We can see that the number of measles have a large amount of 0 values and the values are heavily skewed to the right. This suggests we could model this with a zero-inflated count model such as Poisson or negative binomial.

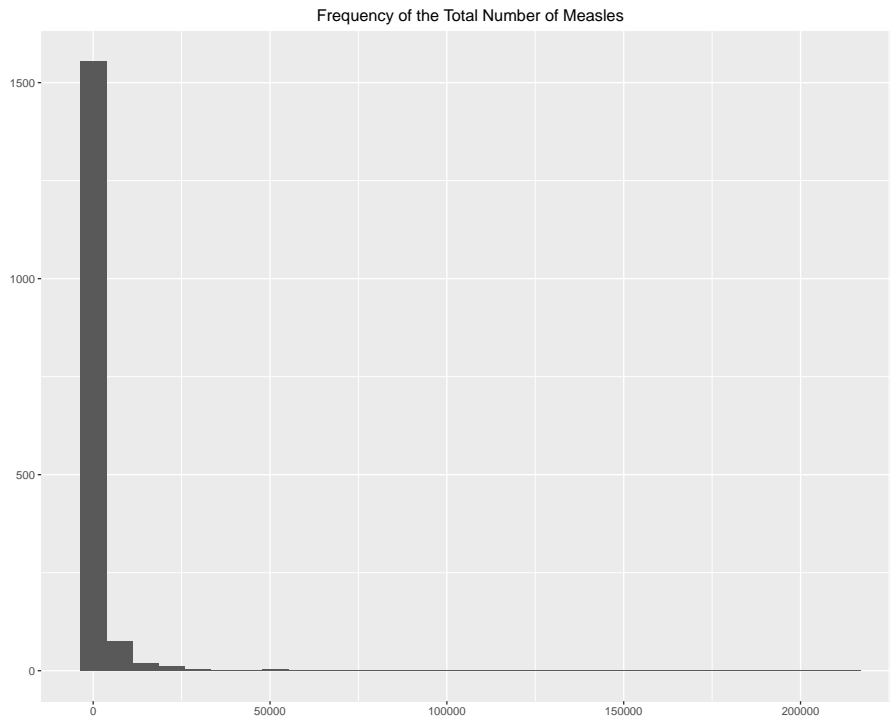


Figure 1: Enter Caption

Figure 2 shows the number of measles cases across time with respect to each region and country. For this visual, we present the past 5 years.

Figure 2: Enter Caption

Figure 3 shows the frequency of our variables.

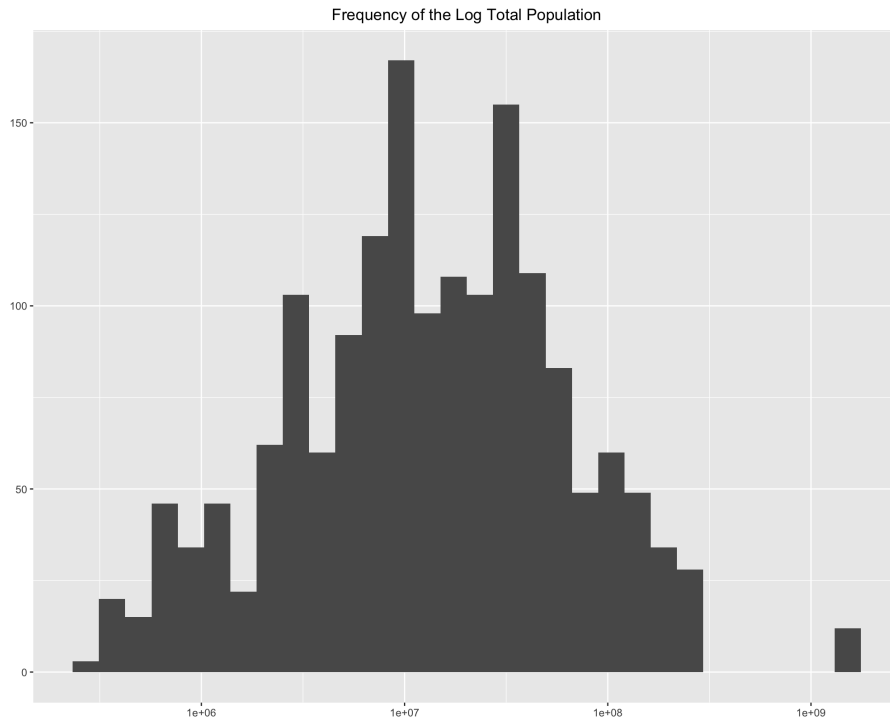


Figure 3: Enter Caption

Figure 4 shows the correlation matrix of each variable in the dataset. While total population is low to moderately correlated with the total number of measles cases, we feel it makes sense to include it in the model. We would expect the countries with more people will have a higher measles cases.

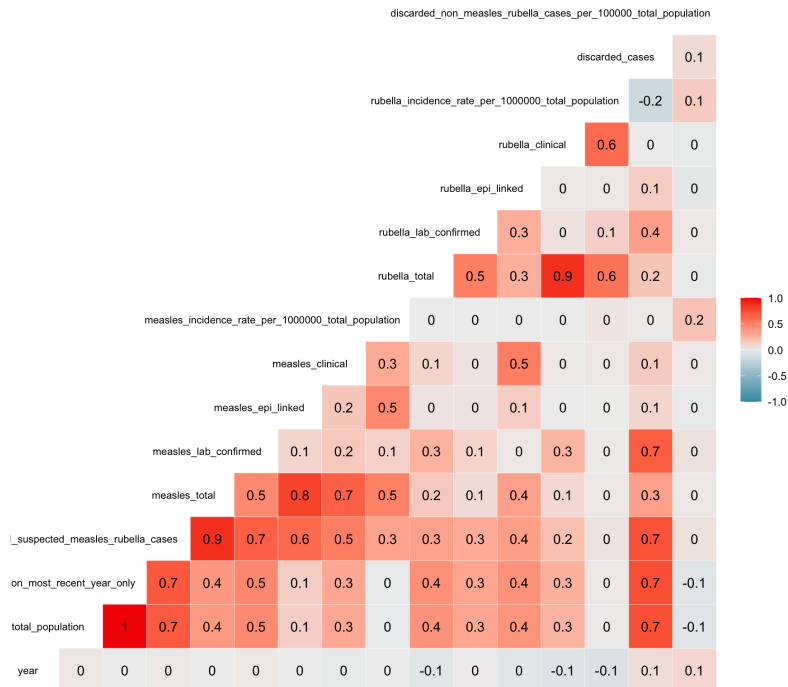


Figure 4: Enter Caption

Model

We define our model below.

Zero-Inflated Poisson Model Specification

Let y_i be the observed count of measles cases for observation i , with the following structure:

$$y_i \sim \begin{cases} 0 & \text{with probability } \pi_i + (1 - \pi_i) \cdot \text{Poi}(0 \mid \lambda_i) \\ \text{Poi}(y_i \mid \lambda_i) & \text{with probability } (1 - \pi_i) \end{cases}$$

where:

- λ_i is the Poisson mean (rate parameter) for count data,
- π_i is the zero-inflation probability.

Count Model (Poisson Mean)

The Poisson mean is modeled on the log scale:

$$\log(\lambda_i) = \beta_0 + \beta_1 \log(\text{Population}_i) + u_{\text{year}[i]} + u_{\text{region}[i]} + u_{\text{country}[i]}$$

where:

- β_0 : intercept,
- β_1 : effect of log population size,
- $u_{\text{year}[i]} \sim \mathcal{N}(0, \sigma_{\text{year}}^2)$,
- $u_{\text{region}[i]} \sim \mathcal{N}(0, \sigma_{\text{region}}^2)$,
- $u_{\text{country}[i]} \sim \mathcal{N}(0, \sigma_{\text{country}}^2)$,
- Random effects are nested: country \subset region \subset year.

Zero-Inflation Model

The zero-inflation probability π_i is modeled via a logistic link:

$$\text{logit}(\pi_i) = \gamma_0 + \gamma_1 \log(\text{Population}_i)$$

where:

- γ_0 : intercept for zero-inflation,
- γ_1 : effect of log population size on zero-inflation probability.

Summary

The full model is a zero-inflated Poisson:

$$y_i \sim \text{ZIPoisson}(\lambda_i, \pi_i)$$

with

$$\log(\lambda_i) = \beta_0 + \beta_1 \log(\text{Population}_i) + u_{\text{year}[i]} + u_{\text{region}[i]} + u_{\text{country}[i]}$$

$$\text{logit}(\pi_i) = \gamma_0 + \gamma_1 \log(\text{Population}_i)$$

Priors

We define the priors follows:

- $\beta_0 \sim N(0, 10)$
- $\beta_1 \sim N(0, 1)$
- $\gamma_0 \sim N(0, 10)$
- $\gamma_1 \sim N(0, 10)$

Results

The RMSE for the model was around 300 measles cases.

Considerations

We considered different combinations of the random effects, but all produced a higher root mean squared error. We considered a zero inflated negative binomial model, which produced a higher root mean squared error.

Future Work

For future analysis, we would like to utilize more covariates in the model. We would like to use a bivariate Poisson model to predict total measles and rubella cases simultaneously.

References

[Community, 2024] Community, D. S. L. (2024). Tidy tuesday: A weekly social data project.