Youth Receiving Intervention Serivices*

Hailey Jang

March 19, 2024

Table of contents

1 Introduction

In recent years, the dynamics of child intervention services have garnered significant attention, particularly in understanding the factors influencing the number of children and youth requiring such services. This study focuses on the province of Alberta, Toronto, analyzing the average monthly number of children and youth receiving child intervention services from 2013 to 2023. By employing a robust statistical framework with R Core Team (2023), we aim to uncover patterns and trends within this crucial social sector, providing insights into the efficacy and reach of child welfare interventions. Utilizing the Children and Family Services dataset, this paper explores the suitability of regression models, including negative binomial regressions,

*Code and data are available at: https://github.com/Hailey-Jang/Mini-essay10.git.

to accurately model and interpret the count data represented by the average monthly service recipients. The findings of this research are intended to contribute to the ongoing discourse on child welfare, offering evidence-based recommendations to enhance the effectiveness of child intervention services and thereby support the well-being of vulnerable populations.

The structure of the paper is organized as follows: Section 2 delves into the broader context of the dataset, emphasizing the crucial aspects of measurement relevant to our study. Section 3 elucidates the setup and justification of our chosen negative binomial regression. Lastly, Section 4 presents the study's contributions to our understanding of child intervention services in Alberta.

2 Data

Before delving into the statistical modeling of child intervention services data, a thorough examination of the dataset is imperative. This initial step involves scrutinizing the dataset's structure, encompassing the yearly distribution of data points, geographical details, the types of care provided, and the demographics of the children and youth served. Specifically, the dataset comprises entries from 2013 to 2023, detailing the average monthly number of children and youth in Alberta, Toronto, who are receiving child intervention services. Each record categorizes the data by year, type of care (such as Temporary Care, Permanent Care, etc.), and the recipients' gender. This exploration aims to identify any preliminary trends, data inconsistencies, or missing values that could influence the subsequent statistical analysis. Understanding these aspects is crucial for ensuring accurate data handling and for selecting the most appropriate regression model for the analysis.

3 Model

Given the nature of our dataset, which comprises count data representing the average monthly number of children and youth receiving child intervention services in Alberta, Toronto, the choice of an appropriate statistical model is paramount. Our response variable is overdispersed count data, evidenced by preliminary data analysis indicating that the variance significantly exceeds the mean. Such overdispersion is a common challenge in modeling count data, where the assumption of equal mean and variance (a key assumption of Poisson regression) does not hold. To address this, we employ Negative Binomial Regression (NBR), a robust alternative that accommodates overdispersion by introducing an extra parameter to model the variance independently of the mean.

```
      Year        Geography       Type of Care            Sex
 2013/2014:12   Alberta:120   Length:120          Female:40
 2014/2015:12                 Class :character    Male  :40
```

```
 2015/2016:12                  Mode  :character   Total :40
 2016/2017:12
 2017/2018:12
 2018/2019:12
 (Other)  :48
 Average Monthly Number                    TypeOfCare     Count
 Min.   :  818         All Child Intervention:30   Min.   :  818
 1st Qu.: 1492         Not In Care           :30   1st Qu.: 1492
 Median : 2646         Permanent Care        :30   Median : 2646
 Mean   : 3447         Temporary Care        :30   Mean   : 3447
 3rd Qu.: 5075                                     3rd Qu.: 5075
 Max.   :11062                                     Max.   :11062
```

3.1 Rationale for Negative Binomial Regression

Negative Binomial Regression is particularly advantageous for our analysis for several reasons:

1. **Flexibility in Modeling Overdispersed Counts:** The NBR model is ideal for data where the count variance exceeds the mean, allowing for a more accurate and reliable fit than Poisson regression, which assumes equal mean and variance.
2. **Incorporation of Explanatory Variables:** The model can incorporate multiple covariates, enabling us to examine the effects of different factors, such as the year, type of care, and other relevant variables, on the count of children and youth in intervention services.
3. **Interpretability of Results:** NBR provides coefficients that can be exponentiated to obtain incidence rate ratios, offering straightforward interpretability in the context of our study.

3.2 Model Setup

We specify our Negative Binomial Regression model as follows:

$\log(\mu_i) = \beta_0 + \beta_1 \text{Year}_i + \beta_2 \text{Type of Care}_i + ...$

where μ_i represents the expected count for the i -th observation, β_0 is the intercept, and β_1, β_2, β_3 are the coefficients of the explanatory variables. The subscript i indexes the data points in our dataset.

To implement this model in R, we use the `glm.nb` function from the MASS package, which fits the regression model via maximum likelihood estimation. The following R script demonstrates this process:

```
      Year  Geography TypeOfCare        Sex 
        10          1          4          3 


Call:
glm.nb(formula = Count ~ Year + TypeOfCare + Sex, data = data, 
    init.theta = 83.94802403, link = log)

Coefficients:
                          Estimate Std. Error z value Pr(>|z|)    
(Intercept)               8.587432   0.039178 219.192  < 2e-16 ***
Year2014/2015            -0.122439   0.045385  -2.698  0.00698 ** 
Year2015/2016            -0.147934   0.045396  -3.259  0.00112 ** 
Year2016/2017            -0.063027   0.045360  -1.389  0.16468    
Year2017/2018            -0.048628   0.045354  -1.072  0.28364    
Year2018/2019            -0.002813   0.045336  -0.062  0.95053    
Year2019/2020             0.021211   0.045327   0.468  0.63981    
Year2020/2021            -0.038960   0.045350  -0.859  0.39029    
Year2021/2022            -0.110686   0.045380  -2.439  0.01472 *  
Year2022/2023            -0.183093   0.045411  -4.032 5.53e-05 ***
TypeOfCareNot In Care    -1.333378   0.028634 -46.566  < 2e-16 ***
TypeOfCarePermanent Care -0.660830   0.028460 -23.220  < 2e-16 ***
TypeOfCareTemporary Care -1.514057   0.028704 -52.746  < 2e-16 ***
SexMale                   0.062449   0.024953   2.503  0.01233 *  
SexTotal                  0.724835   0.024827  29.196  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(83.948) family taken to be 1)

    Null deviance: 4864.25  on 119  degrees of freedom
Residual deviance:  122.05  on 105  degrees of freedom
AIC: 1743

Number of Fisher Scoring iterations: 1


              Theta:  83.9 
          Std. Err.:  11.4 
```

```
 2 x log-likelihood:  -1710.986
```

The `summary` function provides a comprehensive overview of the model's coefficients, their statistical significance, and diagnostic measures to assess the model's fit. Interpreting these results allows us to understand the influence of each explanatory variable on the average monthly number of children and youth receiving child intervention services.

3.3 Justification of the Model

The selection of Negative Binomial Regression is justified by the data's inherent characteristics and the model's alignment with our research objectives. By accommodating the overdispersion in our count data, the NBR model ensures robust, reliable estimates that enhance our understanding of the factors influencing child intervention services in Alberta. This approach not only enhances the precision of our estimates but also contributes to the methodological rigor of our analysis, ensuring that our findings can reliably inform policy and practice in child welfare services.

4 Results

The Negative Binomial Regression model was employed to analyze the data on the average monthly number of children and youth receiving child intervention services in Alberta, Toronto. This section presents the findings from the model, interpreting the estimated coefficients, the model's goodness-of-fit, and their implications for understanding the dynamics of child intervention services.

4.1 Model Interpretation

The estimated coefficients from the Negative Binomial Regression model provide insights into the factors that are significantly associated with the variations in the average monthly number of children and youth in care.

4.2 Implications

The findings from this study have important implications for policymakers and stakeholders involved in child welfare. Understanding the factors influencing the average monthly number of children and youth in intervention services allows for more informed decision-making and strategic planning. For example, the trend over the years can indicate the effectiveness of existing policies or highlight the need for new strategies to address the changing dynamics in child intervention needs.

References

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