MATH 532 Project Presentation

Mpox Surveillance and Reporting in Nigeria: Insights from Count Data Modeling

Aliu A. Adebiyi

2025-05-06

Introduction

- Monkeypox is a zoonotic disease caused by infection with the Monkeypox Virus (MPXV), an orthopoxvirus with symptoms similar but less severe than smallpox (Precious et al., 2023).
- Its burden is more pronounced in developing countries, making the outbreak a global concern (Precious et al., 2023).
- In Nigeria, preparedness and timely response are vital, hence the recent effort by the Nigeria Centre for Disease Control (NCDC) to collate reported cases across states.
- This project analyzes Mpox data (2017–2024) using count regression models to identify factors influencing reported cases across Nigerian states.

Study Design and Data Sources

- This is an ecological study of Mpox cases reported in Nigeria between 2017 and 2024.
- Data on reported cases were obtained from the NCDC (NCDC, 2024).
- Explanatory variables—population density, GDP per capita, literacy, temperature, and region—were sourced from the Nigerian Bureau of Statistics (NBS, 2021).
- Variable selection was based on prior studies on COVID-19 distribution in Nigeria (Daniel & Adejumo, 2021).

Count Regression Modelling

Let Y_i be the number of reported cases. Since the response variable is count data, we use a **Poisson regression model**(Faraway, 2016):

$$Y_i \sim \text{Poisson}(\mu_i), \quad \text{where} \quad E(Y_i) = \mu_i, \quad \text{Var}(Y_i) = \mu_i$$

A **log link function** relates the mean to the predictors:

$$\log(\mu_i) = \mathbf{X}_i^T \boldsymbol{\beta} \quad \Rightarrow \quad \mu_i = \exp(\mathbf{X}_i^T \boldsymbol{\beta})$$

Here, $\mathbf{X}_i$ is a vector of predictors and $\boldsymbol{\beta}$ is the vector of coefficients.

Procedure for Count Modelling Analysis

- Plotted the response variable against one or more predictors.
- Fitted a Poisson model using all predictors.
- Checked deviance and residual plots for model fit.
- Evaluated the dispersion parameter to detect overdispersion.
- Fitted a Negative Binomial model to handle overdispersion and checked diagnostics.
- Compared the full model with the null (intercept-only) model.
- Used stepwise selection to identify significant predictors and fitted a reduced model.
- Interpreted the final model and drew conclusions.

Data Exploration

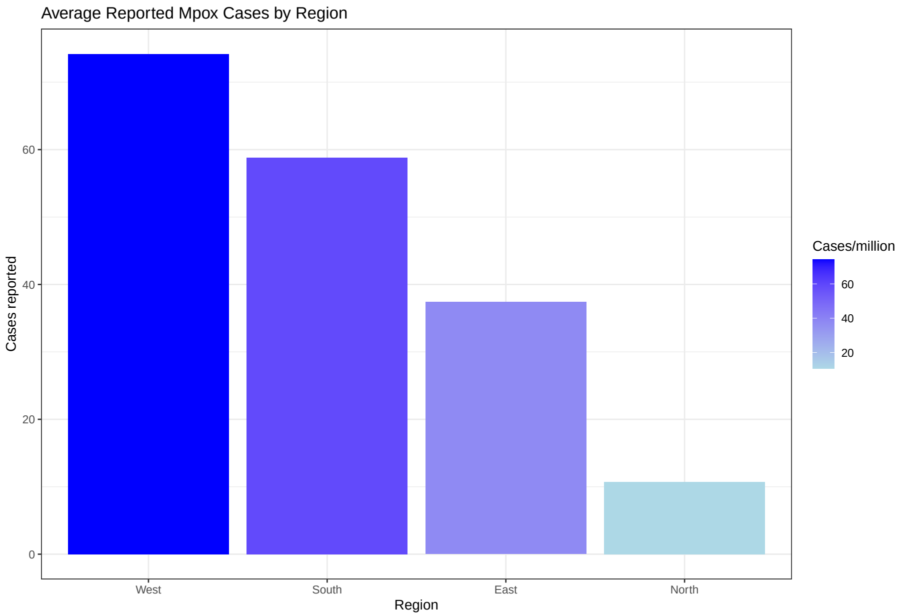
```r
head(mpox)
```

```
## # A tibble: 6 x 7
##   state      cases pop_density gdp_per_capital literacy avg_
##   <chr>      <dbl>       <dbl>           <dbl>    <dbl>    <
## 1 Abia          68       0.761           3.00      85.1
## 2 Adamawa       16       0.110           1.42      57.7
## 3 Akwa Ibom     26       0.795           2.78      89.5
## 4 Anambra       29       1.14            1.62      82.1
## 5 Bauchi         1       0.133           0.983     39.8
## 6 Bayelsa       95       0.251           2.67      74.9
```

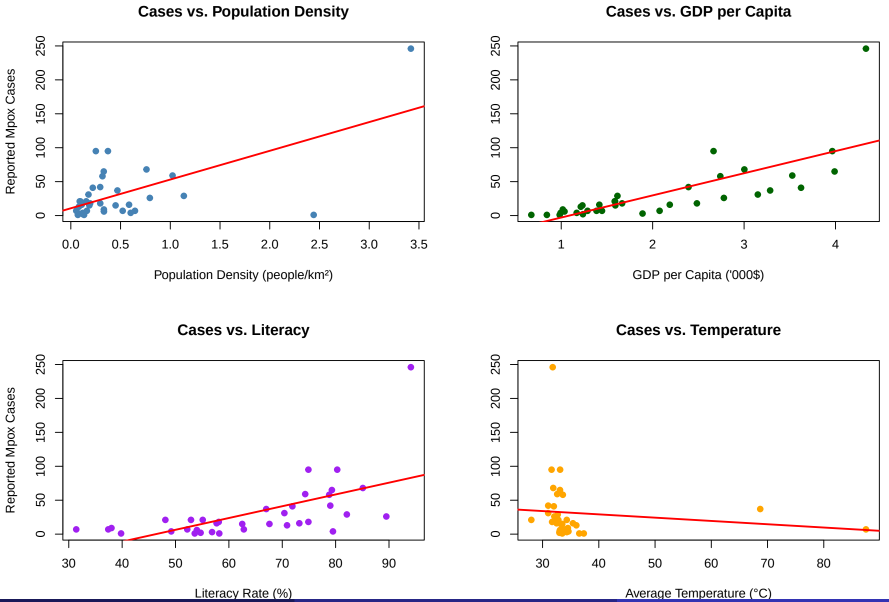
```r
dim(mpox)
```

```
## [1] 36  7
```

Data Exploration



Analysis - Scatter Plots

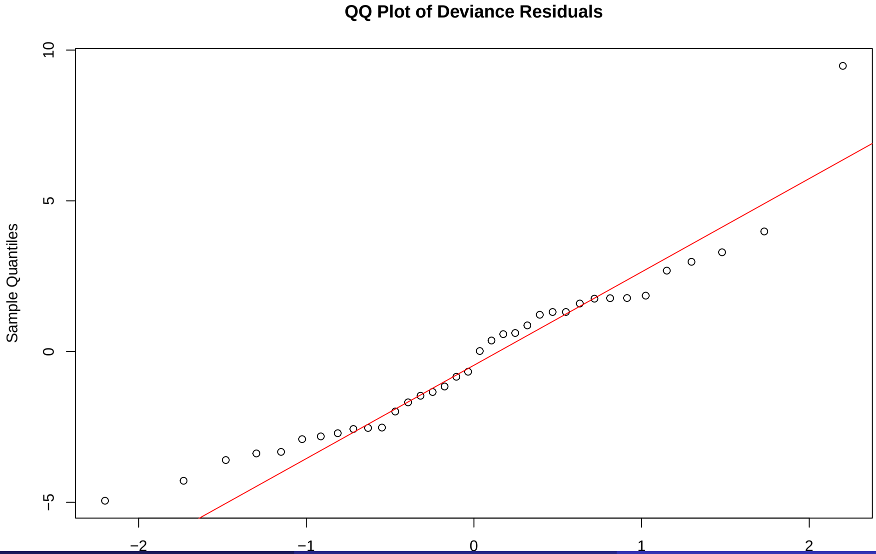


Analysis - Poisson Model.

```r
pois_mod <- glm(cases~.-state, data=mpox, family="poisson")
sumary(pois_mod)
```

```
##                  Estimate Std. Error z value  Pr(>|z|)
## (Intercept)     0.6334710  0.3346825  1.8928   0.05839
## pop_density     0.1058964  0.0492778  2.1490   0.03164
## gdp_per_capital  0.6290814  0.0578554 10.8733 < 2.2e-16
## literacy        0.0174884  0.0044443  3.9350 8.319e-05
## avg_temp       -0.0065861  0.0044081 -1.4941   0.13515
## regionEast      0.1632751  0.1264885  1.2908   0.19676
## regionSouth     0.0832690  0.1503148  0.5540   0.57960
## regionWest      0.3106380  0.1330863  2.3341   0.01959
##
## n = 36 p = 8
## Deviance = 219.04070 Null Deviance = 1426.45447 (Differenc
```

Plots of Residuals, Poisson Model



Analysis - Diagnostic

Over-dispersion Parameter

```
dp = pois_mod$deviance/pois_mod$df.residual
dp
```

```
## [1] 7.822882
```

This necessitate the use of count regression modeling that account for over dispersion parameters as an alternative.

Analysis - Negative Binomial Regression

```
nb_mod <-glm.nb(cases ~ . -state, data = mpox)
nb_mod
```

```
##
## Call:  glm.nb(formula = cases ~ . - state, data = mpox, ini
##     link = log)
##
## Coefficients:
##     (Intercept)      pop_density  gdp_per_capital          I
##        0.990309        -0.072892         0.834065          (
##        avg_temp       regionEast      regionSouth        reg
##       -0.014483         0.271380        -0.006992          (
##
## Degrees of Freedom: 35 Total (i.e. Null);  28 Residual
## Null Deviance:       166.3
## Residual Deviance: 38.91     AIC: 283.8
```

Analysis - AIC criterion for significant predictors

```r
step(nb_mod, trace = 0)
```

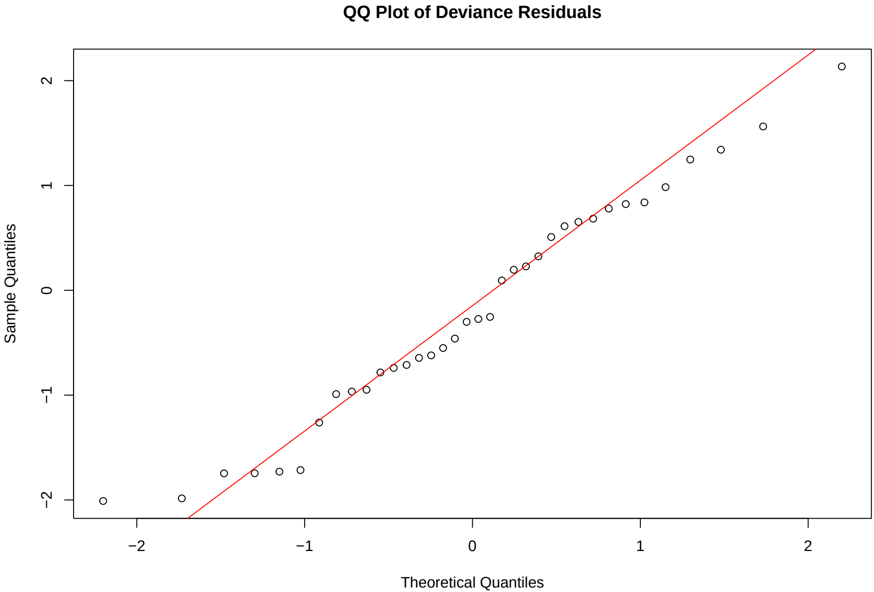
```
##
## Call:  glm.nb(formula = cases ~ gdp_per_capital + literacy,
##     init.theta = 3.688143768, link = log)
##
## Coefficients:
##     (Intercept)  gdp_per_capital         literacy
##         0.12174          0.74750          0.01987
##
## Degrees of Freedom: 35 Total (i.e. Null);  33 Residual
## Null Deviance:       157.8
## Residual Deviance: 39.51     AIC: 276.2
```

Analysis - Reduced Model

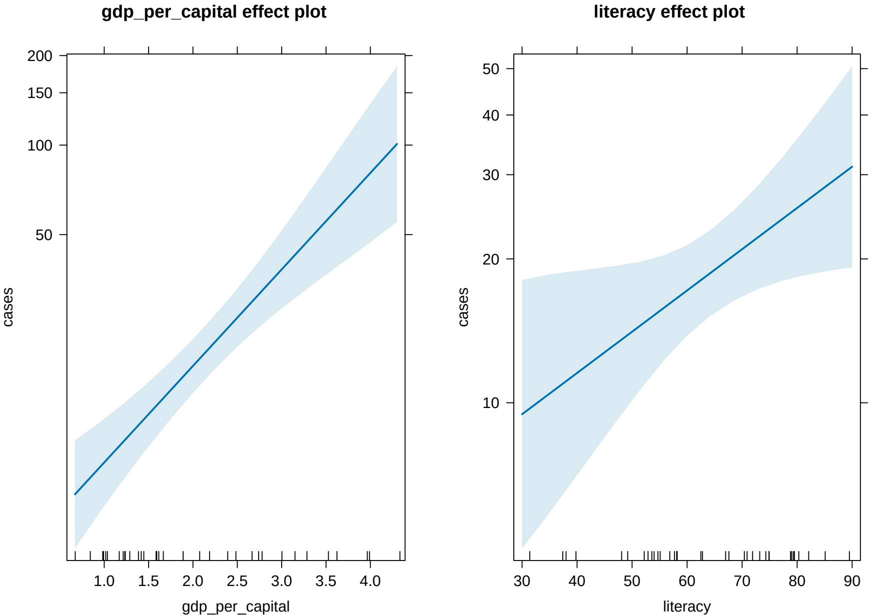
```
reduced_mod <- glm.nb(cases~gdp_per_capital+literacy -state, d
summary(reduced_mod)
```

```
##
## Call:
## glm.nb(formula = cases ~ gdp_per_capital + literacy - state
##     data = mpox, init.theta = 3.688143766, link = log)
##
## Coefficients:
##                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)     0.121737   0.451513   0.270   0.7875
## gdp_per_capital 0.747496   0.127013   5.885 3.98e-09 ***
## literacy        0.019873   0.008656   2.296   0.0217 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '
##
## (Dispersion parameter for Negative Binomial(3.6881) family
```

Analysis - Diagnostic



Result and Interpretation



Result and Interpretation

```r
exp(coef(reduced_mod))
```

```
##     (Intercept) gdp_per_capital        literacy
##        1.129457        2.111707        1.020072
```

The coefficient of GDP per capita gives $exp(0.747) = 2.11$ indicating that for every unit increase in gdp_per_capital, the expected number of Mpox cases increases by a factor of 2.11 holding literacy constant. Also, Literacy has a coefficient of $exp(0.019) = 1.02$ meaning that for every 1% increase in literacy rate, the expected number of Mpox cases increases by about 2%, holding GDP constant.

The implication of this result is that higher reported cases in wealthier and more literate regions may reflect better surveillance, health infrastructure, and awareness, rather than higher actual transmission.

Conclusion

- A Negative Binomial model was used to examine factors influencing reported Mpox cases in Nigeria.
- After adjusting for over-dispersion, GDP per capita and literacy rate emerged as key predictors.
- Regions with higher GDP per capita were over 2 times more likely to report Mpox cases, suggesting stronger surveillance and reporting systems.
- Literacy showed a modest positive effect, indicating the role of awareness and information access in disease detection.
- Results underline the importance of socioeconomic and educational factors in public health surveillance across regions.

References

- Daniel, O. J., & Adejumo, O. A. (2021). *Spatial distribution of COVID-19 in Nigeria.* West African Journal of Medicine, 38(8), 732–737.

- Faraway, J. J. (2016). *Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models* (2nd ed.). Chapman and Hall/CRC.

- Nigerian Centre for Disease Control (NCDC). (2024). *Update on Mpox (MPX) in Nigeria.* Retrieved from www.ncdc.gov.ng

- Nigerian Data Portal. (2021). *State population 2006.* Retrieved from https://nigeria.opendataforafrica.org/ifpbxbd

- Precious, N. D., Agboola, P., Oluwatimilehin, O., Olakunle, O. K., Olaniyi, P., Adiatu, A. I., Olusogo, A. P., Obiwulu, D. J., Adeola, O. A., Ebubechukwu, E. S., Oluwakayode, A. M., Akano, O. S., & Kolawole, Q. O. (2023). Re-emergence of monkeypox virus outbreak in Nigeria: Epidemic preparedness and response

THANK YOU

Download the code and the complete Rmarkdown file from my github