

# Variables influencing the number of members in a household domiciled in SOCCSKSARGEN, a region of the Philippines

Author: Urszula Wolanowska, School of Mathematics & Statistics, University of Glasgow, Glasgow, United Kingdom

E-mail: 2322312w@student.gla.ac.uk

## Introduction

- It is of interest to the government to investigate variables that affect the number of people living in one household in the Philippines.
- Data on 2122 households is obtained from the **Family Income and Expenditure Survey (FIES)** conducted by the Philippine Statistics Authority.
- My analysis covers the SOCCSKSARGEN region of the Philippines (Figure 1).



Figure 1: Map of the Philippines with the SOCCSKSARGEN region highlighted in red

**Response (y):** Number of Household Members

**Explanatory variables (X):** Household Income (in Philippine peso ₱), Food Expenditure (in ₱), Head' of Household Gender, Head' of Household Age, Type of Household, House Floor Area, House Age, Number of Bedrooms, Electricity Availability

## Approach and Methodology

- In Figure 2, it can be seen that the number of people living in one household, our response variable  $y$ , follows a Poisson distribution such that:  $y \sim Po(\mu)$ .
- Here, the parameter  $\mu$  corresponds to the average number of people living in a household.
- This is equivalent to saying that the probability distribution function of  $y$  is:  $f(y) = \frac{\mu^y e^{-\mu}}{y!}$ .

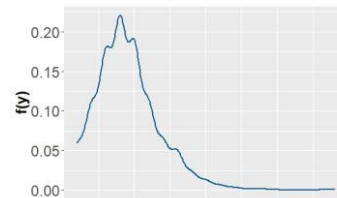


Figure 2: Kernel estimation of the density function of the total number of household members ( $y$ )

The relationship between the number of people living in a household  $y$  and the predictor variables  $X$  is modelled using Poisson, Quasi-Poisson and Negative Binomial Regression.

## Results

The best model fit was achieved with the **Quasi-Poisson regression with log-transformed covariates**:

$$(1) \quad E(y) = \log \hat{\mu} = \log(x^T \hat{\beta}),$$

and  $Var(y) = \phi \hat{\mu}$  where

$\phi = \frac{\text{Pearson's chi-squared statistics}}{\text{Residual degrees of freedom}} = \frac{X^2}{n-p}$  is the dispersion parameter for this model.

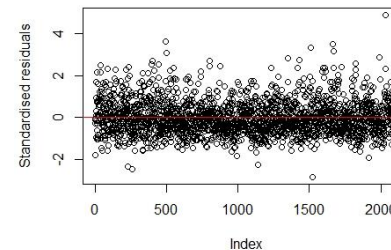


Figure 3: Standardised Residuals plot for model (1)

- Parameter estimates  $\hat{\beta}$  are all significant, meaning **there is a significant relationship between all the explanatory variables and the response**.
- Nonetheless, there is still some variation unexplained by this model. This can be seen in Figure 3, where plenty of standardised residuals are above 2.

## Conclusion and discussion

- All the explanatory variables provided have a significant impact on the number of household members.
- However, **more predictors** could be included to account for more variation in the response variable. Potential covariates could include Head's of Household Marital Status, Employment Status, interactions and many others.

## References

Figure 1, PhilAtlas, Map data © OpenStreetMap contributors. 2020. SOCCSKSARGEN (Region XII). [Online]. [Accessed 29 October 2020]. Available from: <https://www.philatlas.com/mindanao/r12.html>