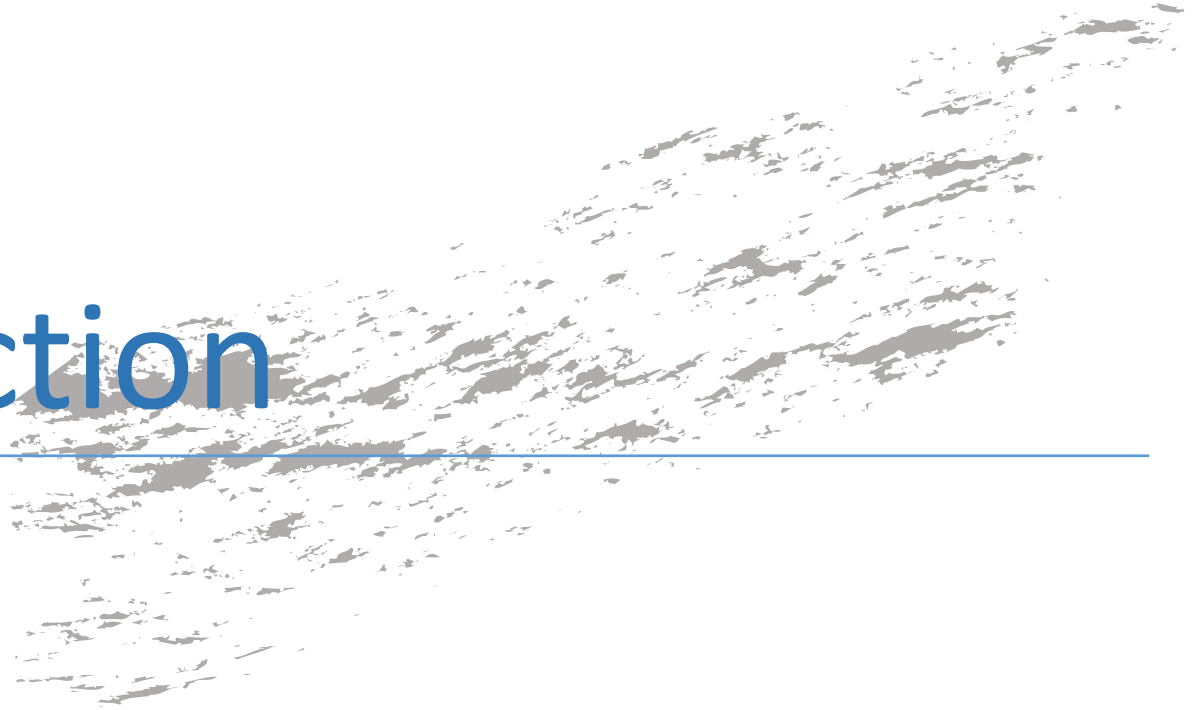
Factors influence the number of people living in a household

Team 2

Team member: Desheng Guo, Fengkai YU, Xunke DENG, Chengcheng Zheng, Panthakan Boonsuriyatham

PART 01
Introduction



Introduction

1. Research Question

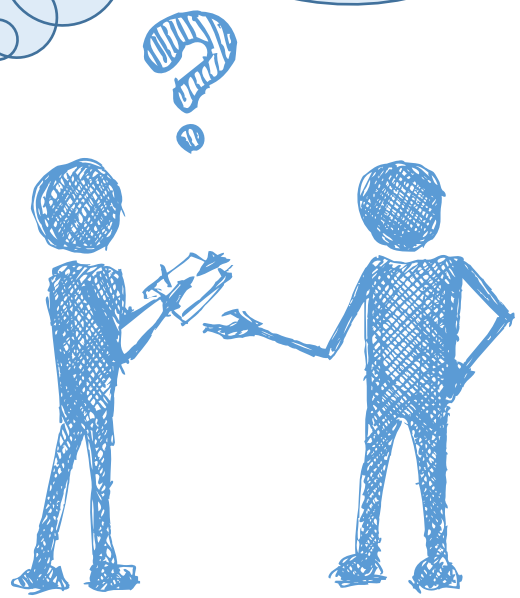
2. Research Approach

3. Data Collection

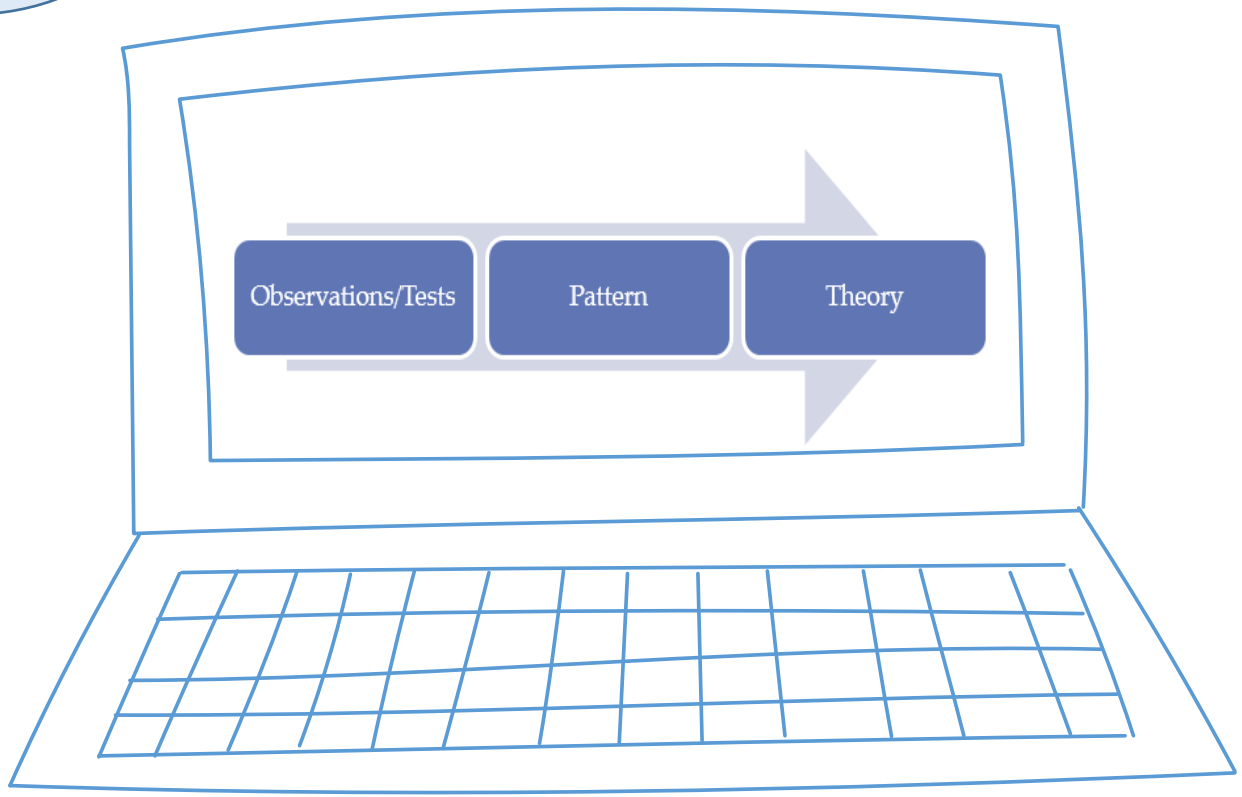
4. Data Analyze

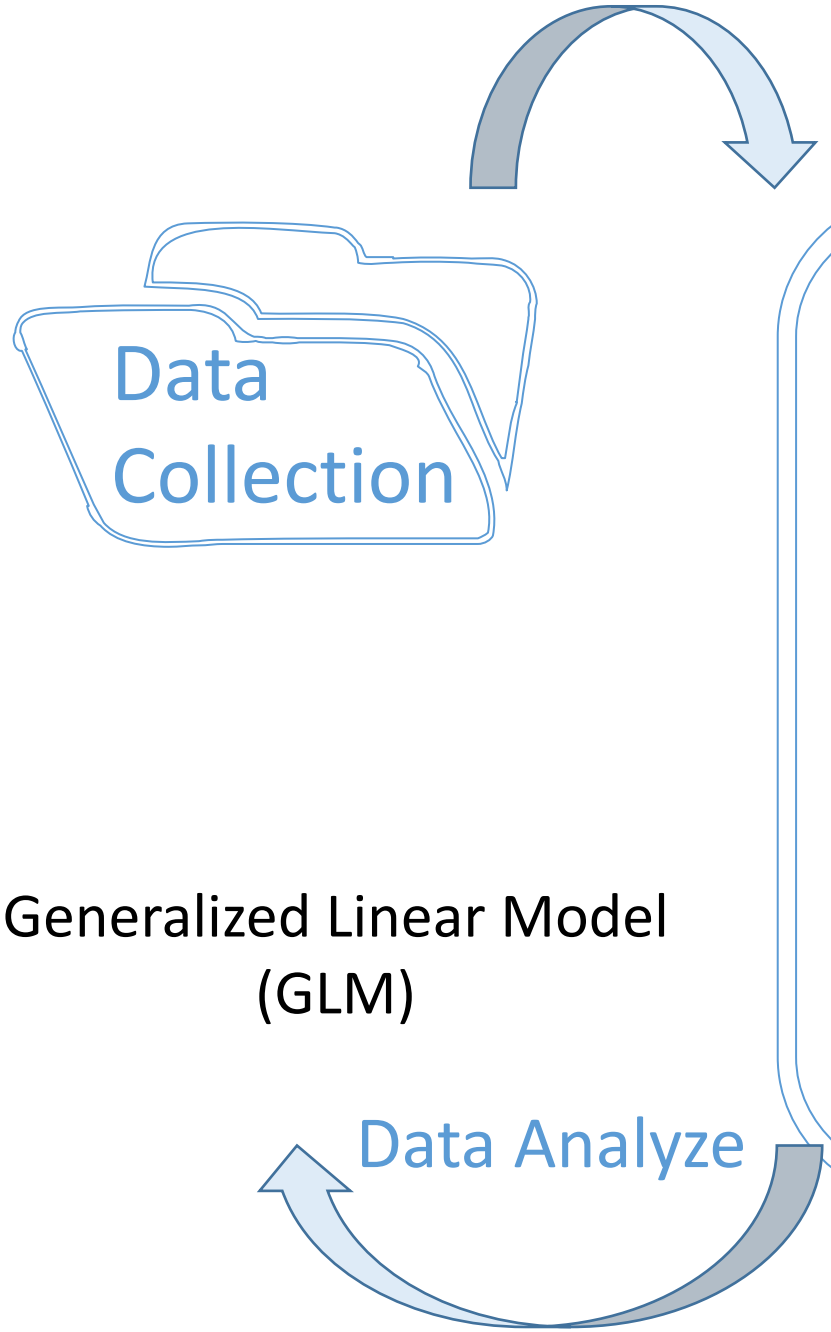
Research Question

Which household related variables influence
the number of people living in a household?



Inductive Research Approach





FIES (Family Income and Expenditure Survey)
Recorded in MIMAROPA of the Philippines.

1249 samples with 11 variables

- **Total.Household.Income** is the Annual household income (in Philippine peso)
- **Region** is the region of the Philippines which a household is in
- **Total.Food.Expenditure** is the annual expenditure by the household on food (in Philippine peso)
- **Household.Head.Sex** is the head of the households sex
- **Household.Head.Age** is the head of the households age (in years)
- **Type.of.Household** is the relationship between the group of people living in the house
- **Total.Number.of.Family.members** is the number of people living in the house
- **House.Floor.Area** is the floor area of the house (in square meter)
- **House.Age** is the age of the building (in years)
- **Number.of.bedrooms** is the number of bedrooms in the house
- **Electricity** is the electricity status of the house (1=Yes, 0=No)

where “head of the household” is the person who is in charge of that house.

PART 02

Exploratory Data Analysis

Statistics Summary

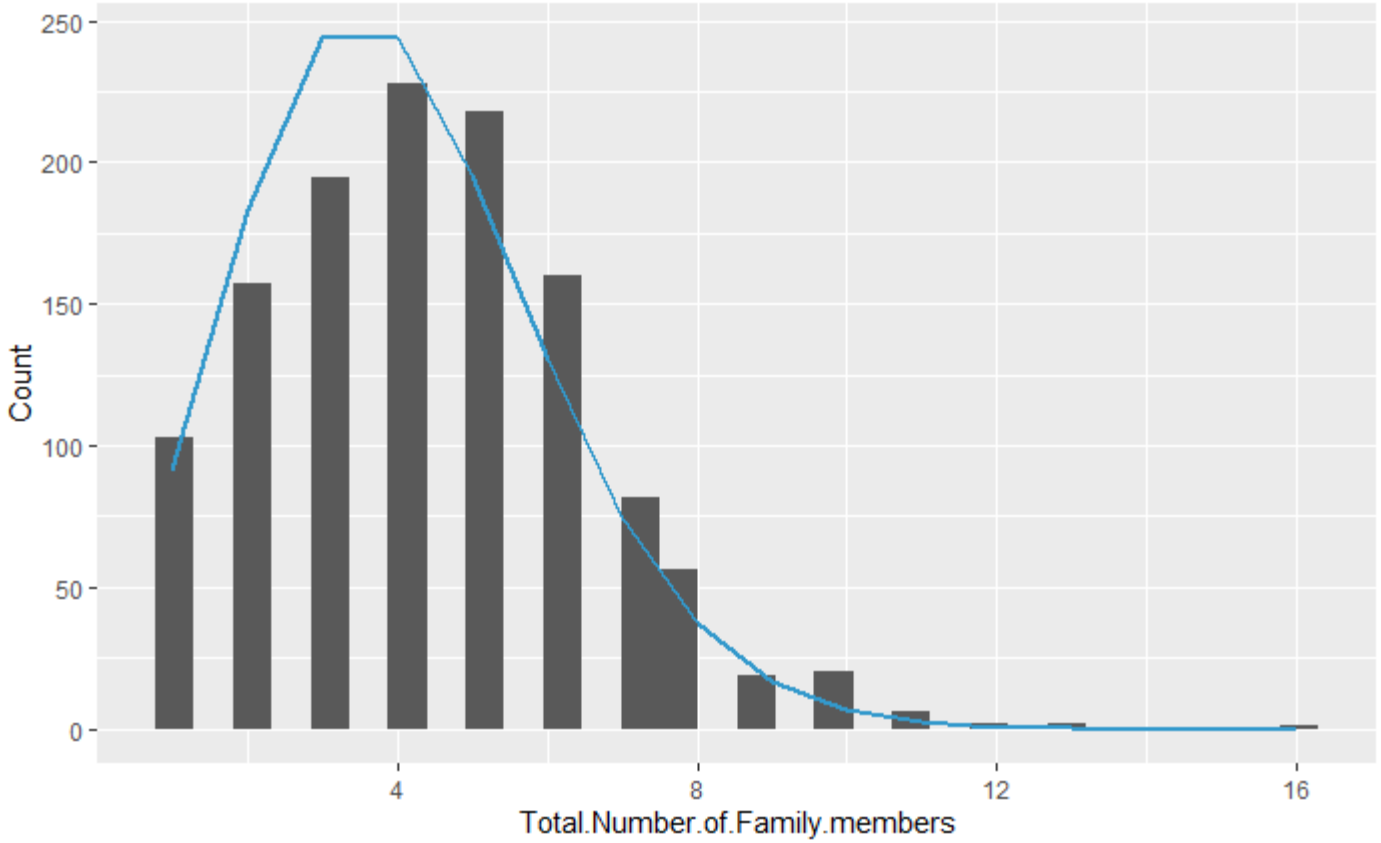
Summary statistics of continuous variables

Variable	Mean	SD	Min	Median	Max	IQR
Total.Number.of.Family.members	4	2	1	4	16	2
Total.Household.Income	216,685	263,207	18,784	140,483	2,891,788	89,919
Total.Food.Expenditure	70,760	41,638	10,488	62,590	413,844	24,118
Household.Head.Age	51	14	15	51	87	10
House.Floor.Area	49	49	5	36	750	24
House.Age	16	13	0	14	105	8
Number.of.bedrooms	2	1	0	2	7	0

Summary statistics of discrete variables

Variable	Counts
Household.Head.Sex	Mal: 983, Fem: 266
Type.of.Household	Sin: 900, Ext: 344, Two: 5
Electricity	1: 1069, 0: 180

Distribution of Total.Number.of.Family.members



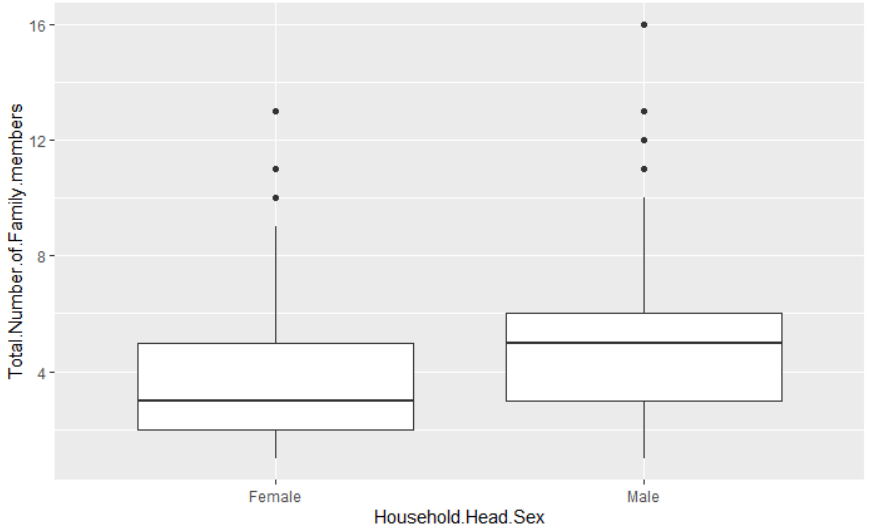
The distribution of number of family members follows Poisson distribution.

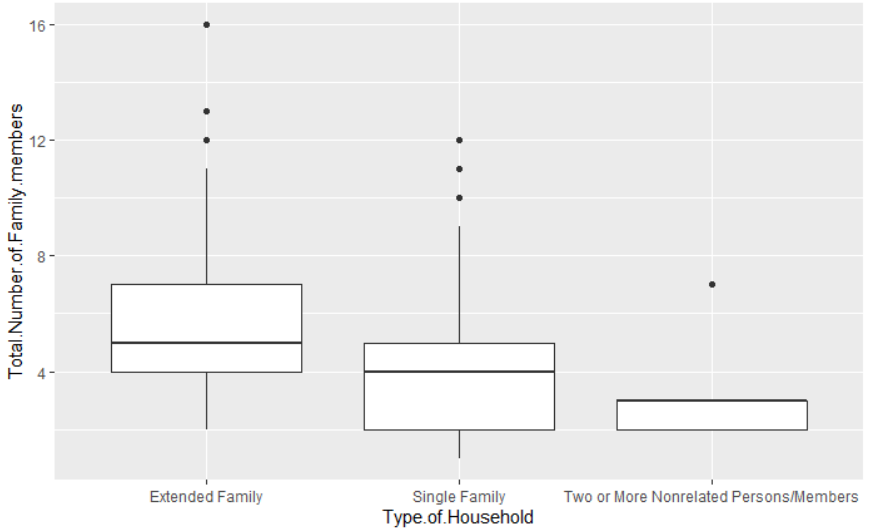
Note:

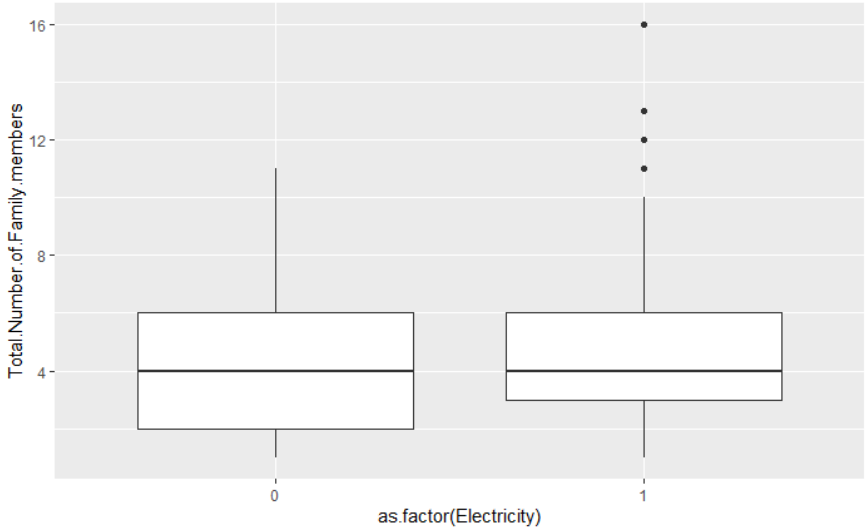
The blue line is the standard Poisson distribution whose mean equals 4.

The bar is the distribution of Total.Number.of.Family.members.

Analysis of Features

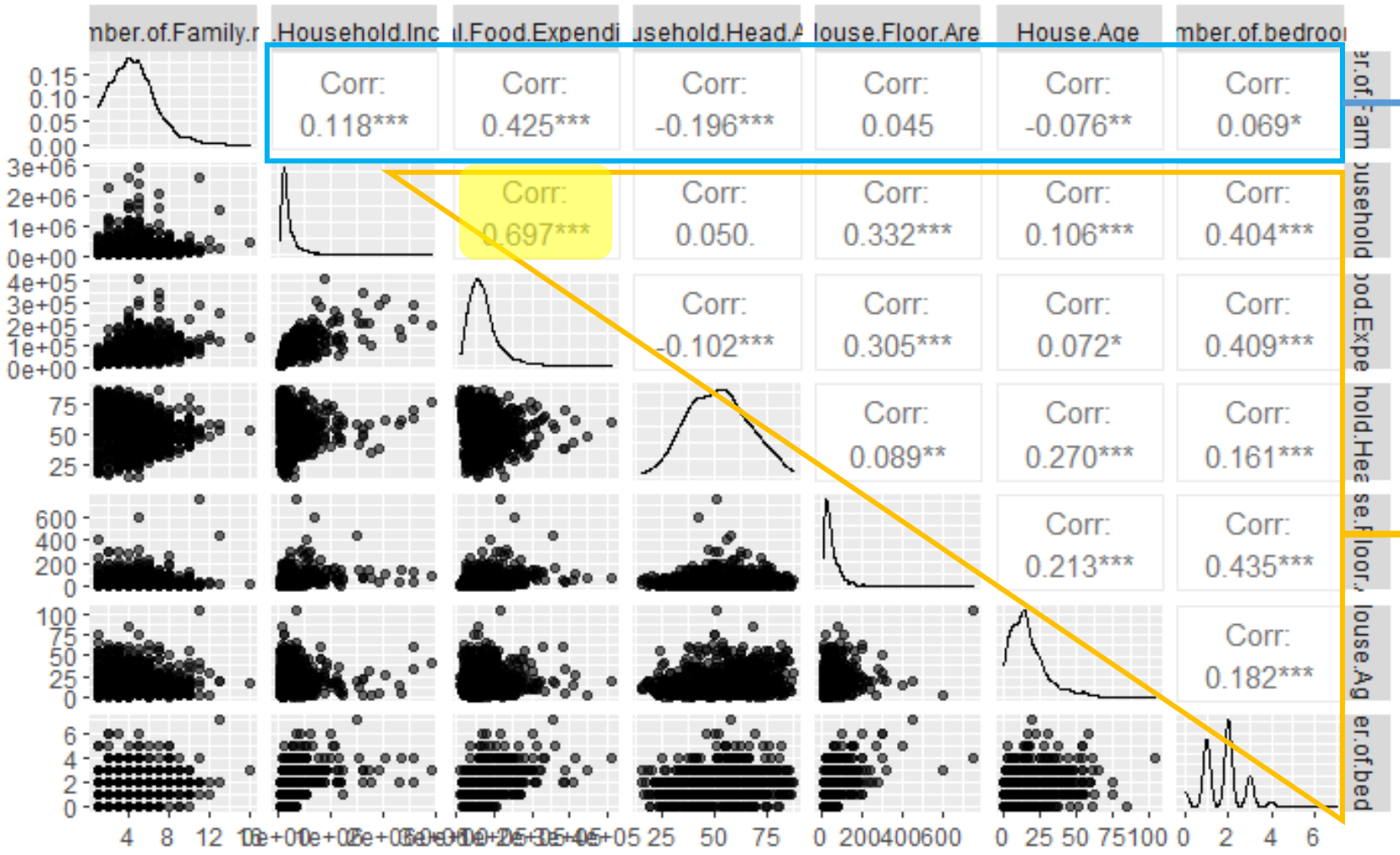






1.On average, male head families have more members than female.
2. Extended Families have more members than single families and single families have more members than nonrelated persons families.
3. Electricity has no impact on the numbers of members in each family.

Analysis of Features



Weak or very weak correlation

↓

Weak or even no linear relationship

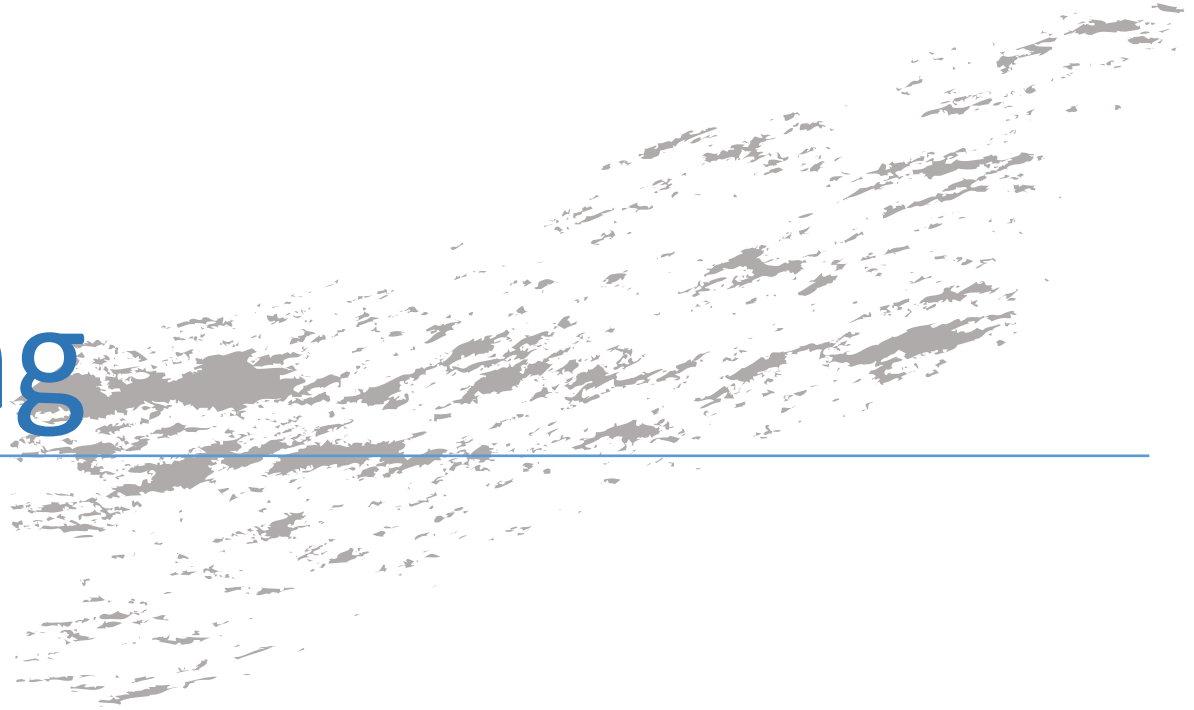
Except income and expenditure (moderate positive correlation), other features only have weak or very weak correlation.

↓

No multicollinearity

PART 03

Modeling



Model Optimization Process

Preliminary model

We use the Poisson regression model in GLM to preliminarily fit all variables.

Step wise

"House.Floor.Area", "Electricity" are dropped out. AIC decreases from 4932 to 4930.

Drop Outlier

One outlier is detected whose "Total.Food.Expenditure" is much higher than other samples. AIC decreases from 4930 to 4901.

Logarithm Transform

Because "Total.Household.Income" and "Total.Food.Expenditure" are too large, their coefficients are too close to 1. So we use logarithm transform, which makes us see the influence of these two variables more clearly.

Step wise & Drop Outlier

After we use logarithm transformation, we repeat step wise and dropping outlier again. Finally, we get our best model whose AIC is 4795.

Best Model

Best Model & Goodness of Fit of Model

Best Model:

$$\log(\hat{Y}_i) = \log(\hat{\mu}_i) = 0.0430769 + 0.7757138 \cdot \log(Total.Household.Income) + 2.0700685 \cdot \log(Total.Food.Expenditure) + 1.2118807 \cdot \mathbb{I}_{\text{Male}}(x) + 0.9965963 \cdot Household.Head.Age + 0.7225769 \cdot \mathbb{I}_{\text{Single Family}}(x) + 0.997657 \cdot HouseAge$$

where

- `Total.Household.Income` is the Annual household income (in Philippine peso);
- `Total.Food.Expenditure` is the annual expenditure by the household on food (in Philippine peso);
- `Household.Head.Age` is the head of the households age (in years);
- `House.Age` is the age of the building (in years)
- $\mathbb{I}_{\text{Male}}(x)$ is an indicator function such that

$$\mathbb{I}_{\text{Male}}(x) = \begin{cases} 1 & \text{if Sex of } x\text{th observation is Male,} \\ 0 & \text{Otherwise.} \end{cases}$$

- $\mathbb{I}_{\text{Single Family}}(x)$ is an indicator function such that

$$\mathbb{I}_{\text{Single Family}}(x) = \begin{cases} 1 & \text{if type of family of } x\text{th observation is Single Family,} \\ 0 & \text{Otherwise.} \end{cases}$$

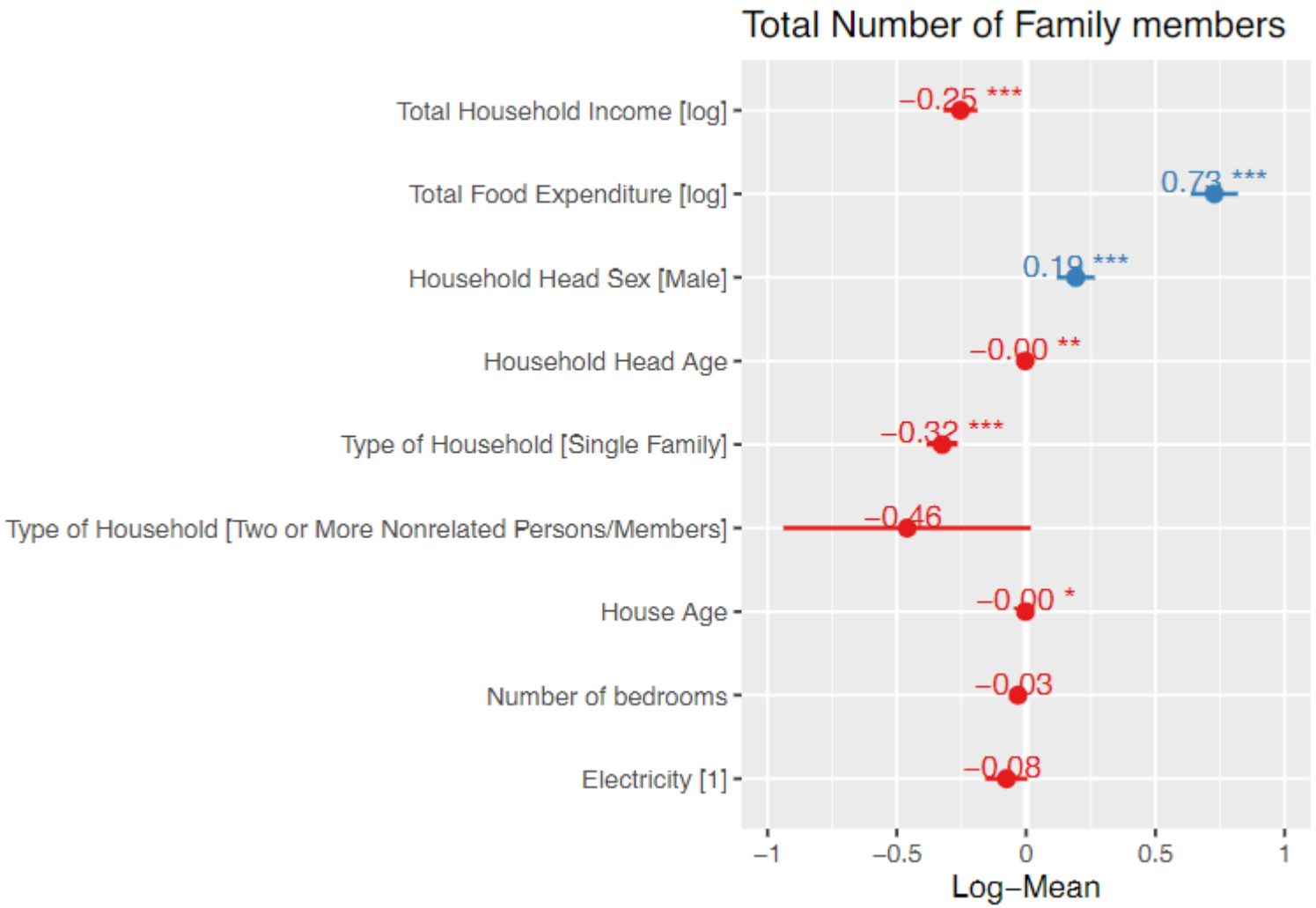
Goodness of Fit of Poisson Model:

chisq	df	p.value
749.7502	1238	1

PART 04

Results and Conclusion

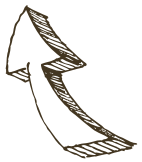
Coefficient and Interpretation



Exp(coef)

Parameter	Exp(coef)
Intercept	0.043
log(Total.Household.Income)	0.776
log(Total.Food.Expenditure)	2.070
Household.Head.SexMale	1.211
Household.Head.Age	0.997
Type.of.HouseholdSingle Family	0.723
Type.of.HouseholdTwo or More Nonrelated Persons/Members	0.631
House.Age	0.998
Number.of.bedrooms	0.969
Electricity1	0.927

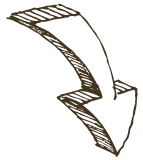
Conclusions



Positive Impactions

log(Total.Food.Expenditure)
(the biggest impaction)

Household.Head.SexMale



Negative Impactions

log(Total.Household.Income)

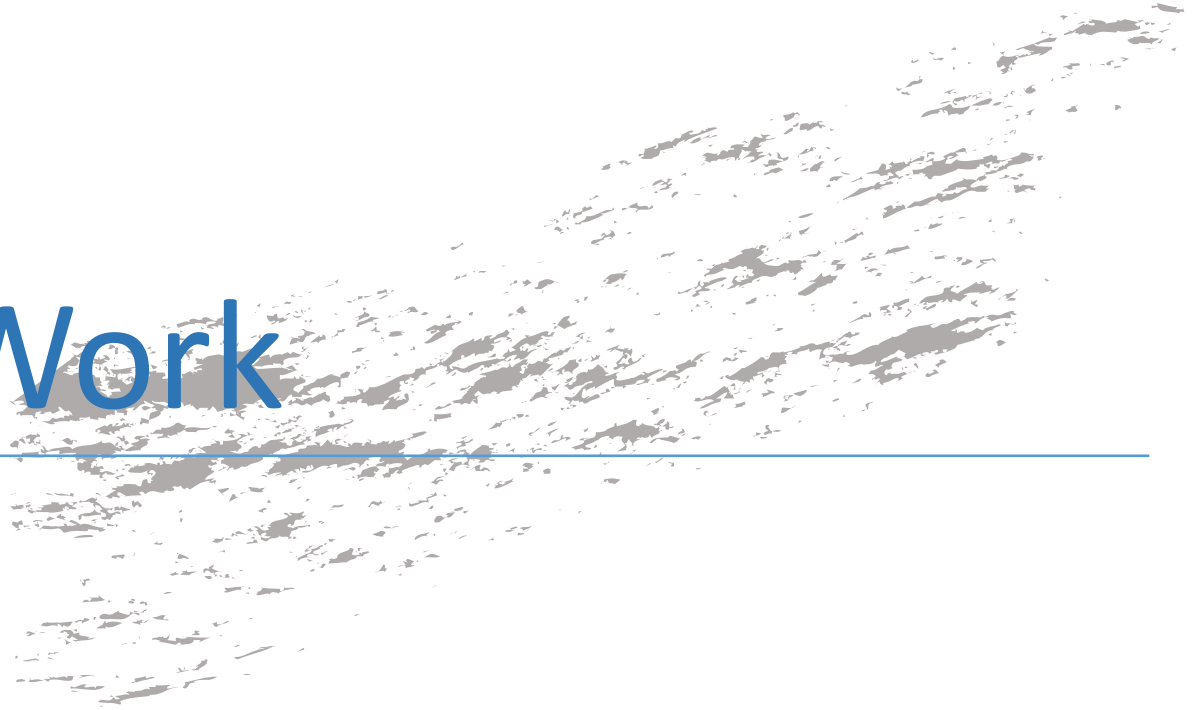
Type.of.HouseholdSingle Family

House.Age

Household.Head.Age

PART 05

Future Work



Future Work

Larger Sample Size

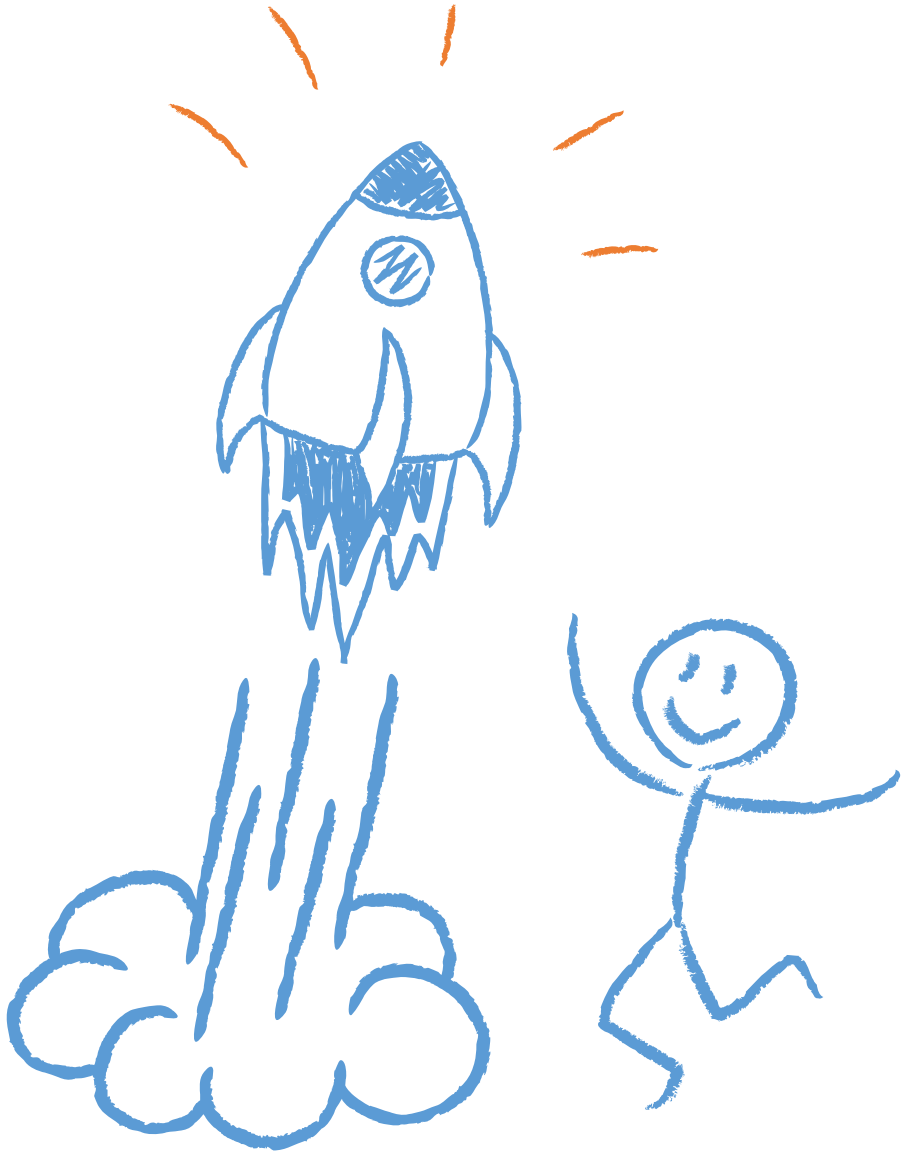
Sampling in MIMAROPA

Sampling among Philippine

Shorter Sample Period

Sampling each three year

Sampling each year



THANK YOU