STATS 3001 / STATS 4104 / STATS 7054
Statistical Modelling III
Practical 5 - Poisson regression

Week 9

GOAL

Perform a Poisson regression to predict the number of people in a househouse based on the age of the head of the household.

DATA

The Philippine Statistics Authority (PSA) spearheads the Family Income and Expenditure Survey (FIES) nationwide. The survey, which is undertaken every three years, is aimed at providing data on family income and expenditure, including levels of consumption by item of expenditure. The data, from the 2015 FIES, is a subset of 1500 of the 40,000 observations (Philippine Statistics Authority 2015). The data set focuses on five regions: Central Luzon, Metro Manila, Ilocos, Davao, and Visayas.

The data is in the file `fHH1.csv`. Each row is a household, and the follow variables are recorded:

- `location`: where the house is located (Central Luzon, Davao Region, Ilocos Region, Metro Manila, or Visayas)
- `age`: the age of the head of household
- `total`: the number of people in the household other than the head
- `numLT5`: the number in the household under 5 years of age
- `roof`: the type of roof in the household (either Predominantly Light/Salvaged Material, or Predominantly Strong Material.

STEPS

1. Read in the dataset.
2. Produce a bar-chart of `total`
3. Produce a scatter-plot of `total` against `age` - add a smoothing line.
4. Fit the Poisson regression
$$total \sim age$$
5. Interpret the coefficient of age.
6. Obtain the Pearson residuals. Plot these against `age`. Is the model adequate?
7. Fit the Poisson regression
$$total \sim age + age^2$$
8. Repeat the residual plots for the new model.
9. Compare the models using a likelihood ratio test, and AIC.
10. Calculate the predicted values for model M2. What is the age of the head of the household associated with the largest fitted value?