R Lab: Poisson Regression

STAT 340: Applied Regression

In this lab, you will continue exploring the Philipinnes household data set from the class notes. This will give you a chance to (a) identify potentially more reasonable models than those that were presented in the lecture itself and (b) address the original questions posed in the lab, which are listed below (*).

```
## Read in data from csv file
fHH1 <- read_csv("https://raw.githubusercontent.com/proback/BeyondMLR/master/data/fHH1.csv")
```

```
## Warning: Missing column names filled in: 'X1' [1]
```

```
## Parsed with column specification:
## cols(
##   X1 = col_double(),
##   location = col_character(),
##   age = col_double(),
##   total = col_double(),
##   numLT5 = col_double(),
##   roof = col_character()
## )
```

```
fHH1 <- fHH1[,-1]
```

Poisson Regression (no overdispersion)

Examine the relationship between the age of heads of households in the Phillipines and number of people in the household. Be sure to write out the model you use, check your assumptions, and adjust for any potential confounding variables.

Interpret the results of your model that you chose above in the context of the problem. Justify your model choice by some appropriate formal statistical criteria.

(*) At what age are heads of households in the Philippines most likely to find the largest number of people in their household? How would you go about solving this problem? Hint: this is a maximization problem based on your model. For what value of age is the likelihood maximized? It may be useful to consider graphical representations of this problem, but it is not required. The R function for density for a Poisson random variable is `dpois()`.

(*) Is this association (the relationship between age of heads of households and number of people in the household) similar for less affluent households (measured by the presence of a roof made from predominantly light/salvaged materials)?

Quasi-Poisson Regression

One way to check for overdispersion (which means that the variance is larger than the mean in Poisson regression) is to fit a quasi-Poisson regression model to your data. The quasi-Poisson model introduces a dispersion parameter, so now $Var(Y_i|\eta_i) = \phi\lambda_i$, for $\phi > 1$, rather than $Var(Y_i|\eta_i) = \lambda_i$, which is the

assumption we make in Poisson regression. We cannot write down a likelihood for this model, but we can find quasi-likelihood estimators of the regression coefficients.

Using the model you identified for the relationship between head of household age and number of people in the household, check to see if there is evidence of overdispersion by fitting a quasi-Poisson model (change `family=quasipoisson(link="log")`) and checking the dispersion parameter estimate. If there is no evidence of overdispersion, $\phi = 1$.