Lecture 11

Recap: estimating φ

Recap: Inference when φ is known

Inference when φ is unknown

Data

2015 Family Income and Expenditure Survey (FIES) on households in the Phillipines. Variables include

- `age`: age of the head of household
- `numLT5`: number in the household under 5 years old
- `total`: total number of people other than head of household
- `roof`: type of roof (stronger material can sometimes be used as a proxy for greater wealth)
- `location`: where the house is located (Central Luzon, Davao Region, Ilocos Region, Metro Manila, or Visayas)

Poisson regression model

Y_i = number of people in household other than head

$$Y_i \sim \text{Poisson}(\lambda_i)$$

$$\log(\lambda_i) = \beta_0 + \beta_1 \, \text{Age}_i$$

Model assumptions

Y_i = number of people in household other than head

$$Y_i \sim \text{Poisson}(\lambda_i)$$

$$\log(\lambda_i) = \beta_0 + \beta_1 \text{Age}_i$$

Question: What assumptions does this Poisson regression model make?

Model assumptions

Y_i = number of people in household other than head

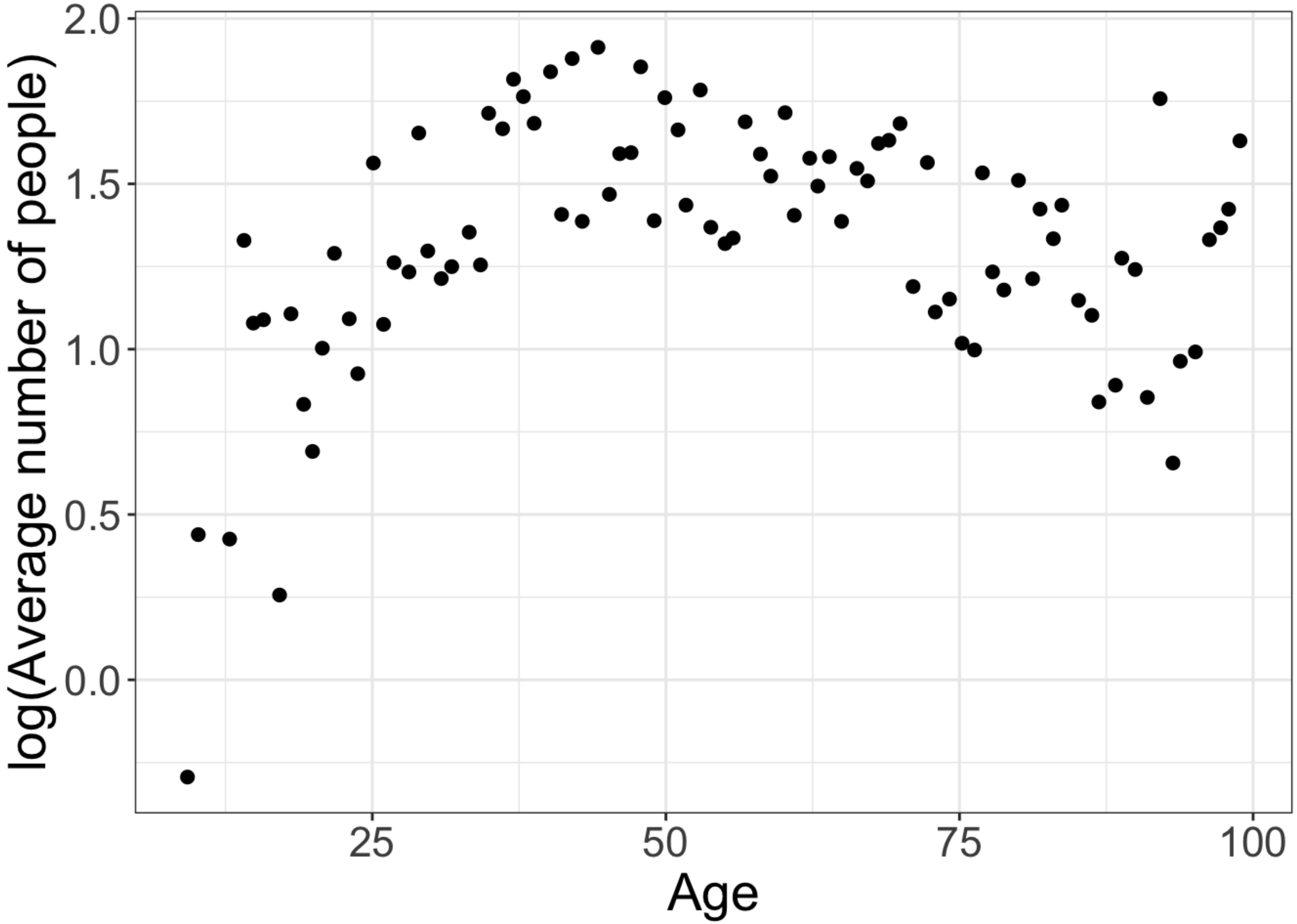
$$Y_i \sim \text{Poisson}(\lambda_i)$$

$$\log(\lambda_i) = \beta_0 + \beta_1 \text{Age}_i$$

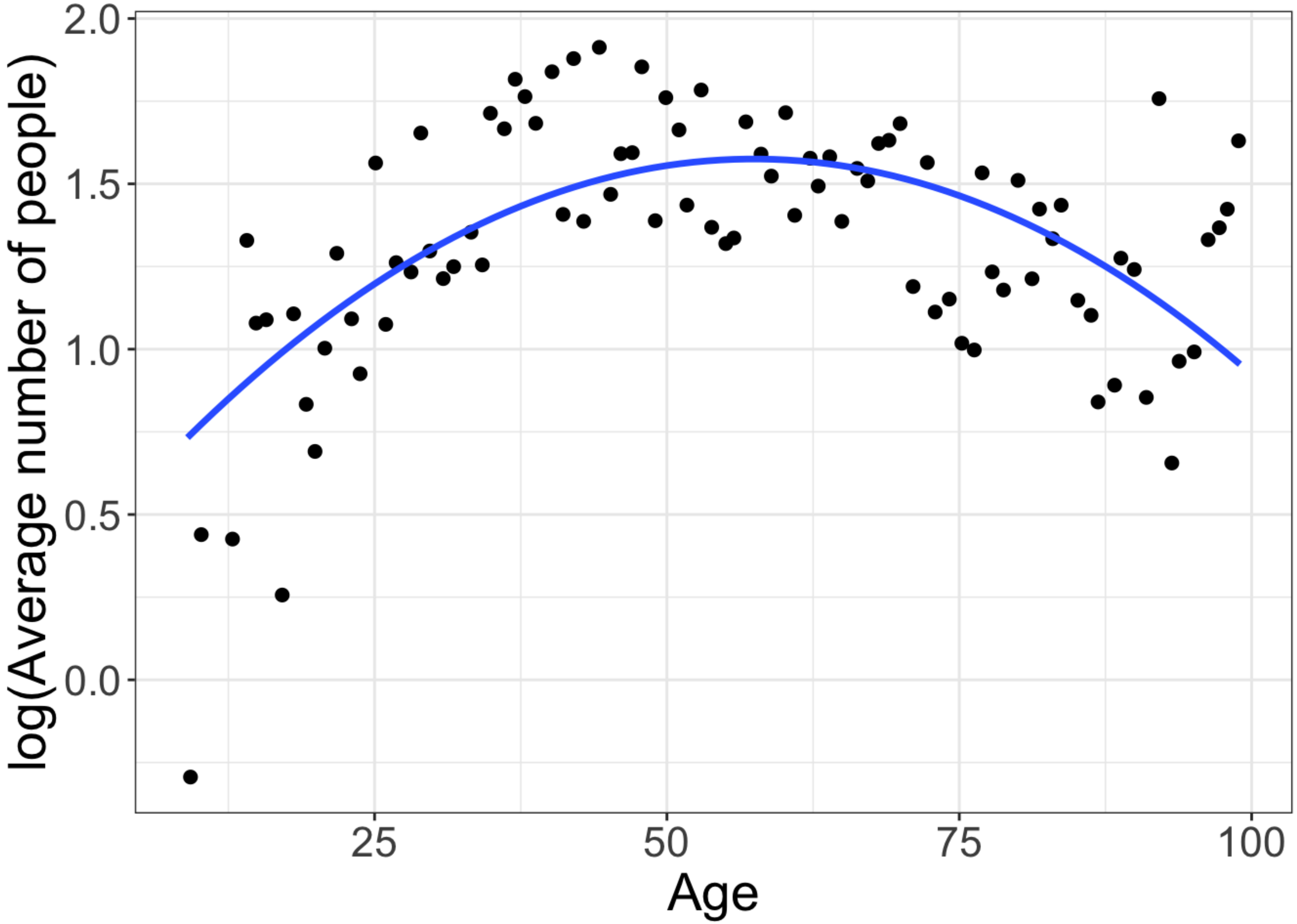
- **Shape:** The shape of the regression model is correct
- **Independence:** The observations are independent
- **Poisson distribution:** A Poisson distribution is a good choice for Y_i

Question: How could I assess these assumptions?

Shape: log empirical means plot

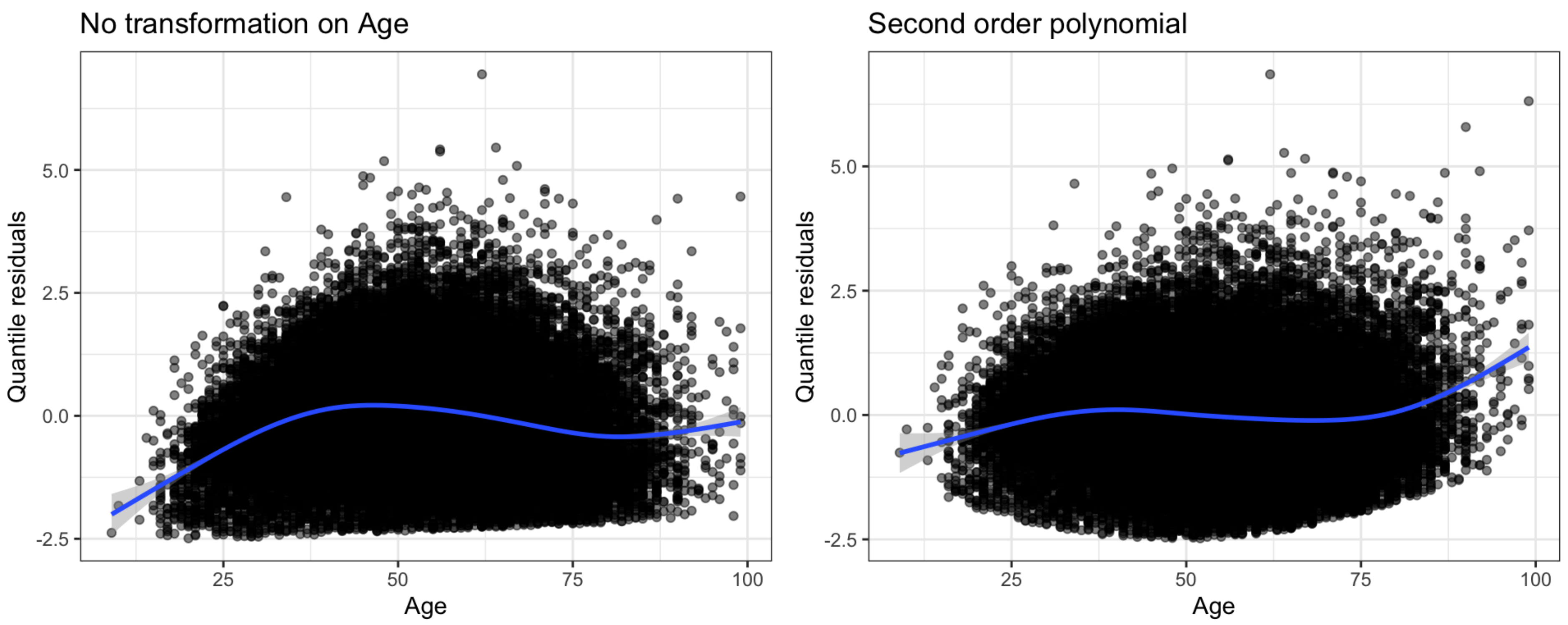


Shape: log empirical means plot



Shape: quantile residual plot

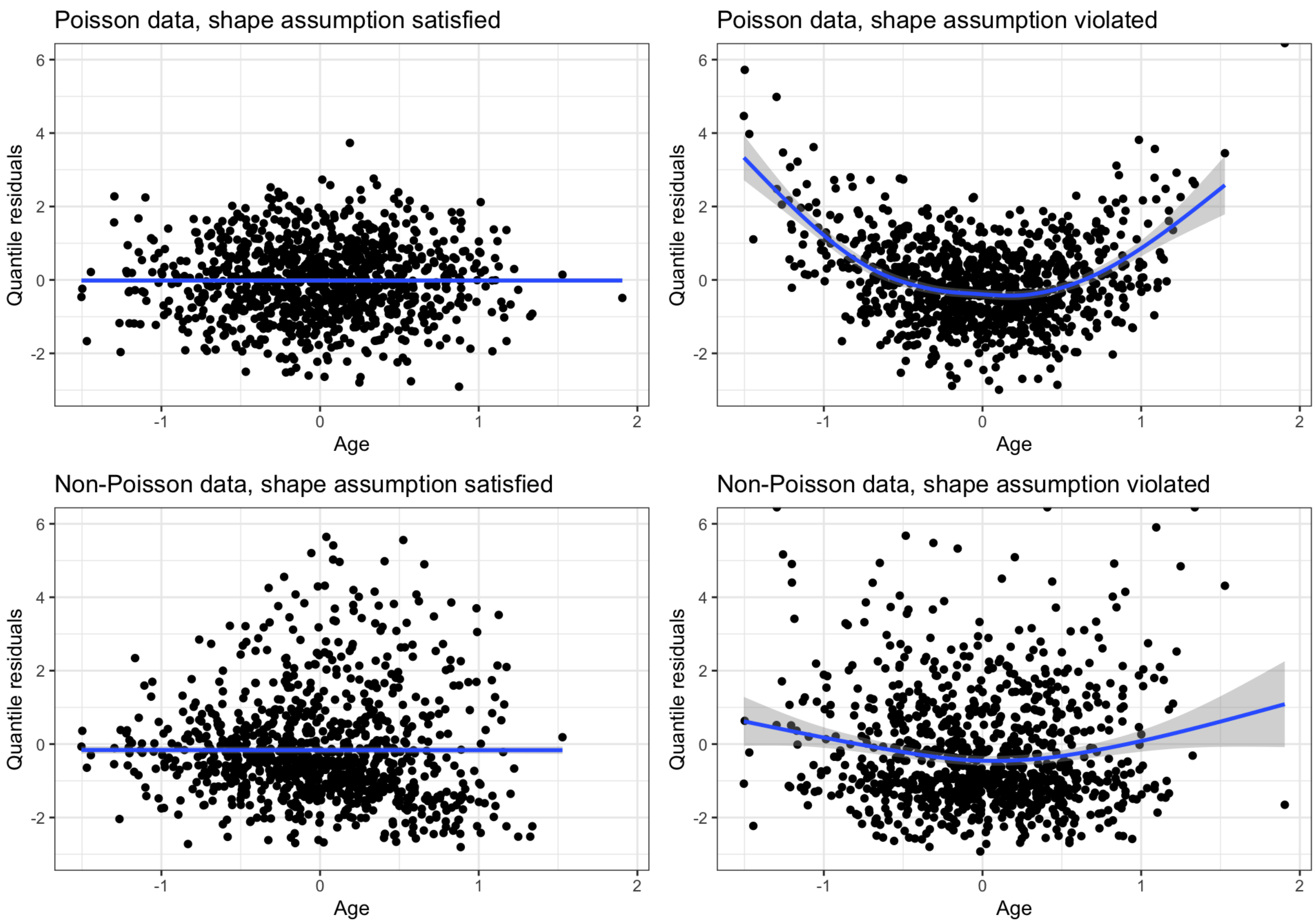
```r
m1 <- glm(total ~ age,
          data = fies, family = poisson)
m2 <- glm(total ~ poly(age, 2),
          data = fies, family = poisson)
```



Class activity

https://sta712-f23.github.io/class_activities/ca_lecture_11.html

Class activity



Using quantile residual plots

We can use the quantile residual plot to assess the shape and distribution assumptions: