

## Fox Module 22: Generalized linear models, discrete and continuous data

(The attached PDF file has better formatting.)

### *Homework assignment: Education and Auto Accidents*

The homework assignment follows the discussion forum reading for this module.

We fit a linear model to three groups of drivers:

<i>Exposures</i>	<i>Years of Schooling</i>	<i>Auto Accidents per 100 Drivers</i>
1,000	8	15
1,000	12	8
1,000	16	3

- The X value is the years of schooling.
- The Y value is the number of auto accidents per 100 drivers.

The table shows that drivers with

- 8 years of schooling (elementary school) have claim frequencies of 15%.
- 12 years of schooling (high school) have claim frequencies of 8%.
- 16 years of schooling (college) have claim frequencies of 3%.

We compare GLMs with different distributions of the error term.

- Normal distribution with a constant variance.
- Poisson distribution.

Assume each year of schooling has the *same linear effect* on claim frequency.

- We fit a straight line to the three points.
  - The variance of the error term depends on the GLM.
- A. Which model gives the higher claim frequency for drivers with eight years of schooling?  
B. Which model gives the higher claim frequency for college educated drivers?  
C. Why might a linear model not be proper for these data? How does decreasing marginal utility affects the slopes? If a driver with 9 years of schooling has an expected claim frequency 1 percentage point less than a driver with 8 years of schooling, should the difference from 12 to 13 years of schooling be more or less than 1 percentage point?  
D. How do actuaries treat class dimensions like years of schooling? Do actuaries treat this as a quantitative or qualitative class dimension?