

Paper 2 Assignment

GLMs for Count Data

Data was collected on doctor visits from a sample of 5,190 people in the 1977/1978 Australian Health Survey. Cameron and Trivedi (1986) sought to explain the variation in doctor visits using one or more explanatory variables. The data can be found in an R data set from `library(AER)` accessible with the command `data("DoctorVisits")`. Variable descriptions can be found under `help("DoctorVisits")`.

You can start looking at these models using Poisson regression, but then you will want to explore the zero-inflated Poisson (ZIP) model for these data (see Section 15.2.1, Zero-Inflated Poisson Regression in your textbook, or video lecture for Thursday, 9/10). In R, the ZIP model can be fit using the function `zeroinfl()` in the `pscl` package. If you want to get a head start on your paper and don't feel comfortable with a ZIP model without some discussion, consider writing up the Introduction and Methods section, as far as you can go. You can also write up any investigation of the assumptions for a Poisson regression model, and make appropriate plots/tables.

See the class syllabus and homework rubric for grading details.