

GLMs for Count Data Exercise

```
knitr::opts_chunk$set(message = FALSE)
knitr::opts_chunk$set(warning = FALSE)
```

```
# install.packages("AER")
# install.packages("pscl")
library(AER)
library(pscl)
```

Use the **AER** package to load the **Medicaid1986** dataset. Information on the included data can be found using the `?Medicaid1986` command.

Suppose we are interested in modeling the number of doctor visits per year for each individual in the dataset, conditional on the information provided in all other given variables. What type of data is this response variable, and which distribution would be most applicable in this case?

Part I

First, try using a log-linear model to fit the data. Use model diagnostics to determine whether the necessary assumptions hold, and provide an interpretation for the model coefficients. Do you think this model is a good fit?

Part II

Next, try fitting a hurdle model for the same outcome. Why might this approach be useful in this case? How do the predictions differ between the two models?