Psych 6136: GLMs for Count Data

School absenteeism

This exercise examines the fitting of various count data GLM models to data about school absenteeism in rural New South Wales, Australia. The data are contained in the data frame `quine` in the MASS package, and can be loaded using `data(quine, package="MASS")`

```
> str(quine)
'data.frame':   146 obs. of  5 variables:
 $ Eth : Factor w/ 2 levels "A","N": 1 1 1 1 1 1 1 1 1 1 ...
 $ Sex : Factor w/ 2 levels "F","M": 2 2 2 2 2 2 2 2 2 2 ...
 $ Age : Factor w/ 4 levels "F0","F1","F2",..: 1 1 1 1 1 1 1 1 2 2 ...
 $ Lrn : Factor w/ 2 levels "AL","SL": 2 2 2 1 1 1 1 1 2 2 ...
 $ Days: int  2 11 14 5 5 13 20 22 6 6 ...
```

`Eth` is ethnic background: Aboriginal or Not, ("A" or "N"); `Age` is a factor representing age group: Primary ("F0"), or forms "F1," "F2" or "F3" (it probably should be an ordered factor); `Lrn` is learner status: factor with levels Average or Slow learner, ("AL" or "SL"). The response is `Days`, number of days absent from school.

Some questions are:

- Which factors affect the number of Days absent?
- Is the Poisson model reasonable here?
- Is there evidence for any interactions among the predictors?

If you get stuck, an R script is available, https://friendly.github.io/psy6136/R/quine.R

Load the packages we will use here:

```
library(car)
library(lmtest)
library(effects)
library(AER)
```

1. All the predictors are factors, but it will be convenient to make `Age` an ordered factor, and use more meaningful labels for `Lrn`. Examine the sample sizes in the 4-way table. Is there anything unusual here? (You might also make a mosaic plot of `quine.tab`)

```
quine$Age <- ordered(quine$Age)
levels(quine$Lrn) <- c("Average", "Slow")
quine.tab <- xtabs(~ Lrn + Age + Sex + Eth, data=quine)
ftable(Age + Sex ~ Lrn + Eth, data=quine.tab)
```

count-data

2. Fit a main effects Poisson model with `glm()`, predicting Days from all predictors. Note that it is necessary to specify `family=poisson` for a count response. What do you conclude? Are there any terms that should be dropped according to these tests?

```
quine$Age <- ordered(quine$Age)
quine.mod1 <- glm(Days ~ ., data=quine, family=poisson)
summary(quine.mod1)
car::Anova(quine.mod1)
```

3. Test this model for overdispersion, using `AER::dispersiontest()`. What do you conclude?
   `dispersiontest(quine.mod1)`

4. Re-fit this model as a quasi-Poisson model. Do the standard tests for the model terms. Does this make a difference in conclusions? Make an effect plot of the predicted values for model interpretation.

5. As a screening device, fit the model with all two-way interactions, using `update()` on the quasi-Poisson model from the previous step. Do the standard tests for the model terms. What do you conclude so far?

6. Try adding what seem to be important interactions to the one-way quasi-Poisson model.

7. Compare the models you have fit using `anova()` What do you conclude is the best model so far?

8. Make and interpret an effect plot for your final model.