Psych 6136: Logistic regression

Volunteering for a psychology experiment

This exercise examines the fitting of various logistic regression models to data about personality characteristics (neuroticism and extraversion) that might be predictive of whether students volunteer to participate in a Psychology experiment. The data are contained in the data frame `Cowles` in the car package, and can be loaded using `data(Cowles, package="carData")`

```
> str(Cowles)
'data.frame':   1421 obs. of  4 variables:
 $ neuroticism : int  16 8 5 8 9 6 8 12 15 18 ...
 $ extraversion: int  13 14 16 20 19 15 10 11 16 7 ...
 $ sex         : Factor w/ 2 levels "female","male": 1 2 2 1 2 2 1 2 2 2 ...
 $ volunteer   : Factor w/ 2 levels "no","yes": 1 1 1 1 1 1 1 1 1 1 ...
```

Some questions are:

- Do males and females volunteer equally often?
- How does the probability of volunteering vary with neuroticism?
- Is there evidence for any interactions among the predictors?

If you get stuck, you can find a script for this exercise at
https://friendly.github.io/psy6136/tutorials/logistic-tutorial.R

1. Simple numerical summaries and plots are often useful, but standard methods are not particularly helpful when the response is binary. Try the following, and try to understand the results.

```
summary(Cowles)
plot(Cowles)
car::scatterplotMatrix(Cowles)
```

2. Fit a **main effects model** with `glm()`, predicting volunteer from `sex, neuroticism` and `extraversion`. Note that it is necessary to specify `family=binomial` for a binary response. What do you conclude?

```
mod.cowles0 <- glm(volunteer ~ sex + neuroticism + extraversion,
    data=Cowles, family=binomial)
summary(mod.cowles0)
Anova(mod.cowles0)
```

3. Continue the analysis, but now fit a model, `mod.cowles1`, containing main effects and **all two-way interactions** of the predictors. Use `summary()` and `Anova()` as shown above. What do you conclude about the various two-way interactions?

logistic

4. Fit a model containing the main effects and the **one two-way term** that seems to be important in your `cowles.mod1`. Call this model `cowles.mod`.

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

6. **Effect plots** are a great way to visualize the predicted model results for any linear or generalized linear model. They show the high-order terms in the model, averaging over the predictors not shown in a given plot. Try the following on your `cowles.mod`

```
library(effects)
eff.cowles <- allEffects(mod.cowles,
                 xlevels = list(extraversion = seq(0, 24, 8)))
plot(eff.cowles, multiline=TRUE)
```