STAT 340: Multiple Logistic Regression

GLMs for Binary Response Data

Reading

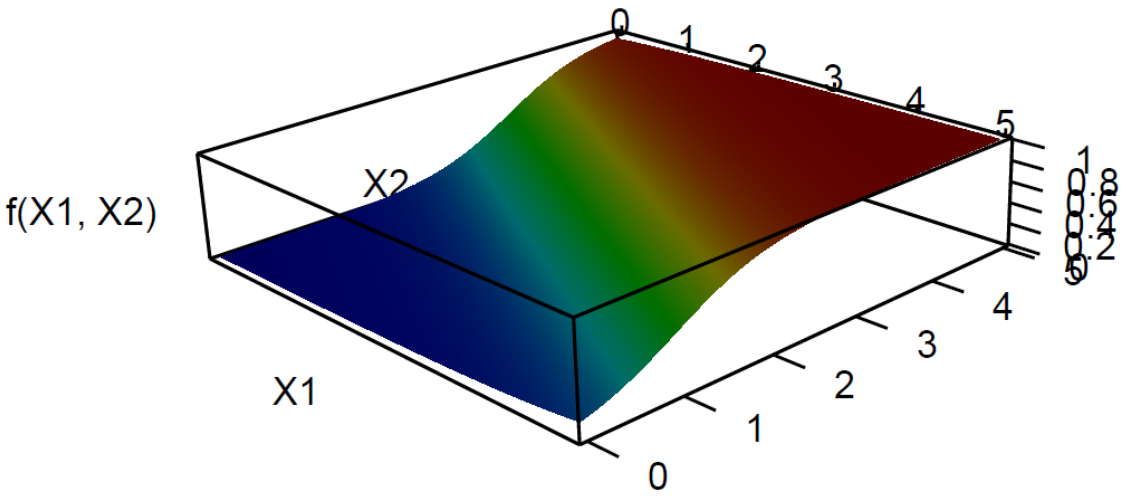
J. Fox, 3rd Edition, pages 380-388

Logistic Regression with Multiple Explanatory Variables

We will now extend logistic regression to allow for p explanatory variables which may be either quantitative or categorical.

$$P(Y_i = 1|X_{i1}, \ldots, X_{ip}) = p(X_{i1}, \ldots, X_{ip}) = \frac{e^{\beta_0+\beta_1 X_{i1}+\cdots+\beta_p X_{ip}}}{1 + e^{\beta_0+\beta_1 X_{i1}+\cdots+\beta_p X_{ip}}}$$

Illustration with $p = 2$ explanatory variables:



Example: Volunteering for a Psychological Experiment

In 1987, Cowles and Davis collected data on students in an introductory psychology class to assess their willingness to participate in a psychological experiment. These data are available in the `Cowles` data set in the **carData** package. The variables in the data set are as follows:

- `neuroticism` = a personality dimension, a numeric variable with integer scores on a scale potentially ranging from zero to 24.
- `extraversion` = a personality dimension, a numeric variable with a potential range of zero to 24.
- `sex` = factor variable with two levels, "male" and "female"
- `volunteer` = factor variable with levels "yes" and "no"; 597 of the 1421 students volunteered.

The goal is to estimate the probability of volunteering based on the two quantitative and one categorial explanatory variables.

```
glm_volunteer <- glm(volunteer ~ sex+neuroticism*extraversion,
                     data=Cowles, family=binomial)
summary(glm_volunteer)
```

```
##
## Call:
## glm(formula = volunteer ~ sex + neuroticism * extraversion, family = binomial,
##     data = Cowles)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.4749  -1.0602  -0.8934   1.2609   1.9978
##
## Coefficients:
##                            Estimate Std. Error z value Pr(>|z|)
## (Intercept)              -2.358207   0.501320  -4.704 2.55e-06 ***
## sexmale                  -0.247152   0.111631  -2.214  0.02683 *
## neuroticism               0.110777   0.037648   2.942  0.00326 **
## extraversion              0.166816   0.037719   4.423 9.75e-06 ***
## neuroticism:extraversion -0.008552   0.002934  -2.915  0.00355 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1933.5  on 1420  degrees of freedom
## Residual deviance: 1897.4  on 1416  degrees of freedom
## AIC: 1907.4
##
## Number of Fisher Scoring iterations: 4
```

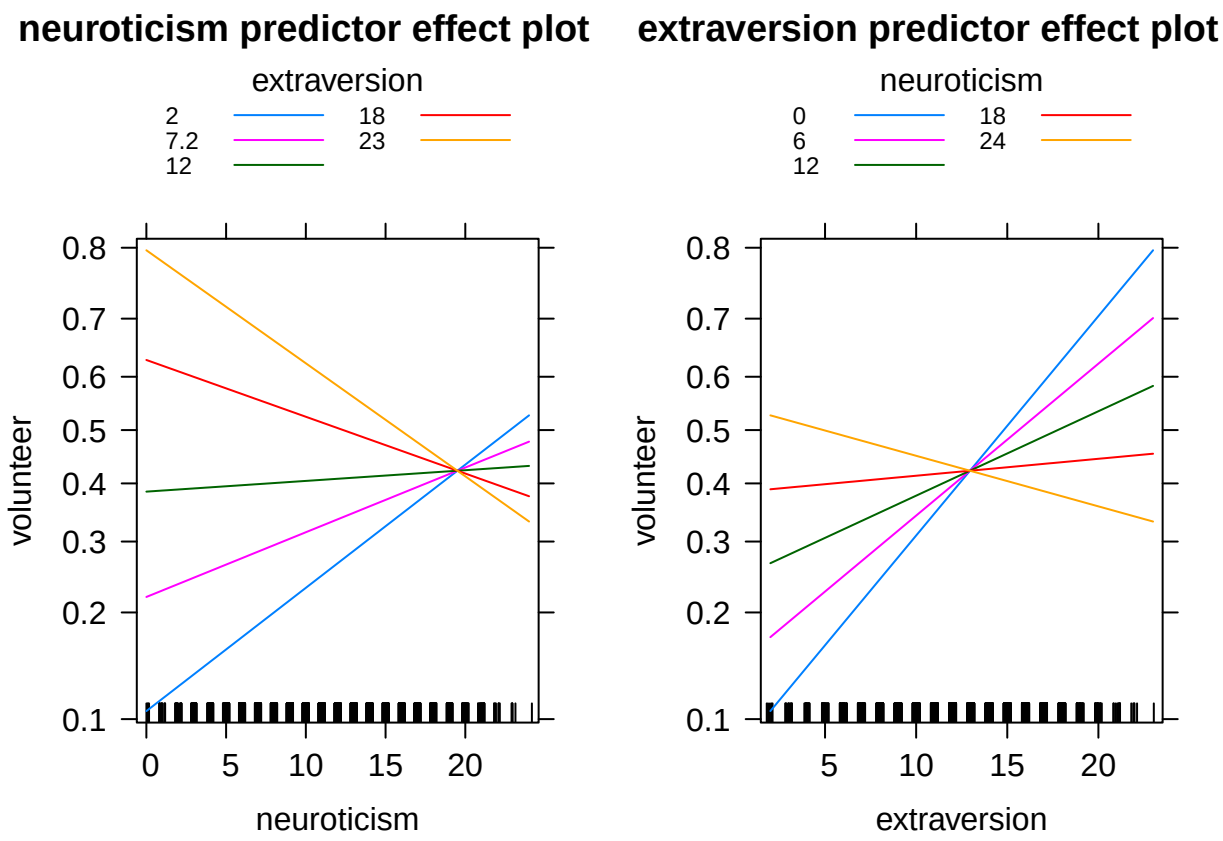
What is the estimated equation for this model? Express it matrix form.

Effect Display Plots

```
library(effects)
```

```
## Use the command
##     lattice::trellis.par.set(effectsTheme())
##   to customize lattice options for effects plots.
## See ?effectsTheme for details.
```

```
plot(predictorEffects(glm_volunteer, ~ neuroticism + extraversion),
     xlevels=list(neuroticism=seq(0,24,by=8),
                  extraversion=seq(0,24,by=8)),
     lines=list(multiline=TRUE))
```



Interpret the effect display plots.

References:

M. Cowles and C. Davis. (1987). The subject matter of psychology: Volunteers. *British Journal of Social Psychology*, 26, 97-102.

J. Fox. 2016. *Applied Regression Analysis and Generalized Linear Models*, 3rd Edition. Sage.

J. Fox. (2003). Effect Displays in R for Generalised Linear Models. *Journal of Statistical Software*, 8(15).

J. Fox and S. Weisberg. 2019. *An R Companion to Applied Regression*, 3rd Edition. Sage.

F. Ramsey and D. Schafer. 2013. *The Statistical Sleuth: A Course in Methods of Data Analysis*, 3rd Edition. Cengage.