Multiple logistic regression

STAT 340: Applied Regression

Interpretation: partial-regression coefficients

Recall from your reading that we can express the odds of "success" for individual i for some generic logistic regression model as follows:

$$\frac{\pi_i}{1-\pi_i} = \exp\{\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}\} = e^{\beta_0}(e^{\beta_1})^{x_{i1}} \cdots (e^{\beta_p})^{x_{ip}}.$$

e^{β_j} is the multiplicative effect of the odds of increasing x_j by 1, when we hold all other explanatory variables constant.

Exercise 1:

Consider the volunteer example from the lecture for today. For purposes of discussion, we will pretend that the assumptions for this are satisfied, despite the fact that we saw in the notes that the linear relationship between the log odd and neuroticism is *not* satisfied.

```
##
## 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
```

(a) Interpret the effect of sex on the odds of volunteering.

(b) Interpret the effect of extraversion on the odds of volunteering.

Exercise 2:

Next class, we will be discussing hypothesis testing and model selection. Using what you know about hypothesis testing for proportions (STAT 140, also beginning of STAT 242) and your reading in Section 14.1.4, I would like you to consider the reliability of the hypothesis tests conducted by default in R in the summary for the GLM above. As in linear regression, the default hypotheses are $H_0 : \beta_j = 0$ versus $H_A : \beta_j \neq 0$. The hypothesis tests are conducted using the Wald statistic:

$$Z_0 = \frac{\hat{\beta}_j - 0}{SE(\hat{\beta}_j)}.$$

The p-values that you see above come from the Normal(0,1) distribution. **Do you expect these p-values to be reasonable for assessing the strength of the relationship between the log odds of volunteering and each of the explanatory variables (and the interaction)?**