AGLM Project Presentation

Peter Shewmaker

April 21, 2019

Nested Models

- Two models are called *nested* if one is a special case of the other.
- For example, if:

$$Y_1 = \alpha_1 + \beta_2 X_2$$

$$Y_2 = \alpha_1 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$$

then the first model is nested within the second model. This is because $Y_1 = Y_2$ if both X_1 and X_3 are held at 0.

- If two models are nested, we can perform a Likelihood Ratio Test to determine which of the models may be a better fit.

Likelihood Ratio Test

- Null hypothesis: the simpler model is the better one.
- The test statistic is a constant multiple of the logarithm of the likelihood ratio of the simpler (smaller) model to the log-likelihood of the larger model, that is:

$$LRT = -2 \log \left(\frac{\mathcal{L}_s(\theta)}{\mathcal{L}_g(\theta)} \right) = -2[\log(\mathcal{L}_s(\theta) - \log(\mathcal{L}_g(\theta))]$$

  where $\mathcal{L}_s(\theta)$ is the likelihood of the smaller model, and $\mathcal{L}_g(\theta)$ is the likelihood of the larger model.
- The test statistic approximates a chi-squared random variable, with degrees of freedom equal to the difference of the number of parameters between the two models.
- The function "lrtest" in the package "lmtest" performs the likelihood ratio test for two nested (generalized) linear models.

Predicting memory score with a linear model

- The variable "log_mem_imm" is a measure of the patient's memory. The following models use some demographic information as well as performance on cognitive tests to predict the patient's score on this memory test.

```
m1 <- lm(log_mem_imm ~ age_visit + mmse + cdr_global,
         data = alzheimers_data_filtered)
m2 <- lm(log_mem_imm ~ age_visit + mmse + cdr_global
         + sex + ethnicity + educ,
         data = alzheimers_data_filtered)
```

Performing a Likelihood Ratio Test with "lrtest"

```
lrtest(m1, m2)
```

```
## Likelihood ratio test
##
## Model 1: log_mem_imm ~ age_visit + mmse + cdr_global
## Model 2: log_mem_imm ~ age_visit + mmse + cdr_global + sex + ethnicity +
##     educ
##   #Df LogLik Df  Chisq Pr(>Chisq)
## 1   5 -21722
## 2   8 -21657  3 130.36  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- Since the test statistic is sufficiently large, we can reject the null hypothesis and conclude that the larger model is more likely than the smaller one in this case.