Multinomial logistic regression

- We use the `multinom` function from the **nnet** package to estimate a multinomial logistic regression model.
- Remark - There are other functions in other R packages capable of multinomial regression, such as the **mlogit** package.
- The `multinom` function does not require the data to be reshaped (as the mlogit package does)
- (Similar format to example code found in Hilbe's *Logistic Regression Models*).

Multinomial logistic regression

- We must choose the level of our outcome that we wish to use as our **baseline** and specify this in the `relevel` function.
  Let's choose "academic".
- Then, we run our model using *multinom*.
- The `multinom` command does not include p-value calculation for the regression coefficients.
- (We can calculate p-values using Wald tests or z-tests).

Multinomial Logistic Regression with R

```
ml$prog2 <- relevel(ml$prog, ref = "academic")
test <- multinom(prog2 ~ ses + write, data = ml)

 # weights:  15 (8 variable)
 initial  value 219.722458
 iter  10 value 179.982880
 final    value 179.981726
 converged
```

Multinomial Logistic Regression with R

```
summary(test)

 Call:
 multinom(formula = prog2 ~ ses + write,
             data = ml)

 Coefficients:
          (Intercept) sesmiddle seshigh    write
 general        2.852   -0.5333 -1.1628 -0.05793
 vocation       5.218    0.2914 -0.9827 -0.11360
```

Multinomial Logistic Regression with R

```
....
 Std. Errors:
          (Intercept) sesmiddle seshigh   write
 general        1.166    0.4437  0.5142 0.02141
 vocation       1.164    0.4764  0.5956 0.02222

 Residual Deviance: 360
 AIC: 376
```

Multinomial Logistic Regression with R

Wald Test

- The Wald test in the context of logistic regression is used to determine whether a certain predictor variable X is significant or not. It rejects the null hypothesis of the corresponding coefficient being zero.
- The test consists of dividing the value of the coefficient by standard error

Multinomial Logistic Regression with R

```
# Coefficients Divided by Standard Errors
# Then Compute p-values.
# 2-tailed z test
# p.values

          (Intercept) sesmiddle seshigh     write
 general    1.448e-02    0.2294 0.02374 6.819e-03
 vocation   7.299e-06    0.5408 0.09895 3.176e-07
```

Multinomial Regression with R

- Remark: Some output is generated by running the model, even though we are assigning the model to a new R object.
- This model-running output includes some iteration history and includes the final **negative log-likelihood** (+ 179.981726).
- This value multiplied by two is then seen in the model summary as the **Residual Deviance** and it can be used in comparisons of nested models (360).

Multinomial Regression with R

- As with many summary outputs, the output contains a column of coefficients and a column of standard errors.
- Each of these blocks has one row of values corresponding to a model equation.
- Focusing on the block of coefficients, we can look at the first row comparing `prog = "general"` to our baseline `prog = "academic"` and the second row comparing `prog = "vocation"` to our baseline `prog = "academic"`.

Multinomial Regression with R

- If we consider our coefficients from the first row to be b_1 and our coefficients from the second row to be b_2, we can write our model equations:

$$ln\left(\frac{P(prog = gen.)}{P(prog = acad.)}\right) = b_{10} + b_{11}(ses = 2) + b_{12}(ses = 3) + b_{13}write$$

$$ln\left(\frac{P(prog = voc.)}{P(prog = acad.)}\right) = b_{20} + b_{21}(ses = 2) + b_{22}(ses = 3) + b_{23}write$$

Multinomial Regression with R

- A one-unit increase in the variable **write** is associated with the decrease in the log odds of being in general program vs. academic program in the amount of 0.058 (b_{13}).
- A one-unit increase in the variable **write** is associated with the decrease in the log odds of being in vocation program vs. academic program in the amount of 0.1136 (b_{23}).

Multinomial Regression with R

- The log odds of being in general program vs. in academic program will decrease by 1.163 if moving from ses="low" to ses="high" b_{12}.
- The log odds of being in general program vs. in academic program will decrease by 0.533 if moving from ses="low" to ses="middle" b_{11}, although this coefficient is not significant.

Multinomial Regression with R

- The log odds of being in vocation program vs. in academic program will decrease by 0.983 if moving from ses="low" to ses="high"(b_{22}).
- The log odds of being in vocation program vs. in academic program will increase by 0.291 if moving from ses="low" to ses="middle"(b_{21}), although this coefficient is not signficant.