STAT-E 150

Chapter 10: Logistic Regression (continued)

Individual Logistic Regression Hypothesis:

H_0: β_k=0
H_A: $\beta_k \neq 0$ where k= 1, 2,3,…

The Wald statistic is used to determine if an individual logistic regression coefficient is significantly different from zero

Sample conclusions:
Due to a small p-value, we reject the null hypothesis that the coefficient (β) is zero. There is sufficient evidence to indicate that the predictor is statistically different from zero.

Overall Model and Multiple Logistic Regression:
H_0: $\beta_1 = \beta_2 = \beta_3 = 0$
H_a: at least one beta is not equal to zero

The Omnibus test with the Chi-statistic is used to determine if the overall multiple logistic regression coefficients are significantly different from zero.

Since $p<.001$, we can reject the null hypothesis. This suggests that predictors together are useful in predicing the dependent variable.

$\log(\text{odds}) = \log(p/1\text{-}p) = \beta_0 + \beta_1 x + \beta_2 x + \beta 3\, x$
E(y) = p, the probability of success
We transform the odds using the natural logarithm, and we use the term **log odds** or **logit** for this transformation.

Other things to remember:
If log(odds) changes by b_1 then odds increases by $e^{\beta 1}$
In other words, the change in odds associated with a unit change in x is $e^{\beta 1}$, which can be denoted as $\text{Exp}(\beta_1)$ (found in the rightmost column of the Variables in the Equation output).
Another way to say it, the odds of "success" for a one unit increase in x_k changes by a multiplicative factor of $\text{Exp}(\beta_1)$.

Finding the best model:

2-Log likelihood test - can be used to assess how well a model would fit the data, this doesn't tell us much by itself but it is critical for comparisons.

Assumptions/Conditions

- Linearity: Box-Tidwell Test
- Independence: No pairing or clustering of the data in space or me
- Random: Need to make sure a 'spinner model' is valid – super important!!

a 3

Sample questions:

Which predictors increase the odds of being an honors student?

This data was collected on 200 high schools students and are scores on various tests, including science, reading, and the presence (or absence) of a social studies course for 1 semester (**ses(1)**) or for 2 semesters (**ses(2)**).[1]

Here is the output from the full model:

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	81.781[a]	.419	.532

a. Estimation terminated at iteration number 6 because parameter estimates changed by less than .001.

Here is the output from the reduced model:

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	87.336[a]	.365	.497

a. Estimation terminated at iteration number 5 because parameter estimates changed by less than .001.

Variables in the Equation

		B [j]	S.E. [k]	Wald [l]	df [m]	Sig. [l]	Exp(B) [n]
Step 1[a]	read	.098	.025	15.199	1	.000	1.103
	science	.066	.027	5.867	1	.015	1.068
	ses			6.690	2	.035	
	ses(1)	.058	.532	.012	1	.913	1.060
	ses(2)	-1.013	.444	5.212	1	.022	.363
	Constant	-9.561	1.662	33.112	1	.000	.000

a. Variable(s) entered on step 1: read, science, ses.

[2]

1. What is the overall model including all predictors based on the output above?
   $\log(odds) = -9.561 + 0.098*x_{read} + 0.066*x_{science} + 0.058*x_{ses(1)} - 1.013*x_{ses(2)}$

---

[1] http://www.ats.ucla.edu/stat/spss/output/logistic.htm
[2] Ibid.

2. What is the model including predictors at a 0.05 significance level based on the output above (reduced model)?
   log(odds) = -9.561 + 0.098*x_{read} + 0.066*$x_{science}$ - 1.013*$x_{ses(2)}$

3. Which model is more useful?
   According to the model summary, the full model is more useful.

   Test stastistic: 87.336 – 81.781 = 5.555

   DF = 1 since we removed 1 term

   http://www.fourmilab.ch/rpkp/experiments/analysis/chiCalc.html

4. What is the probability of being an honors student given 2 semesters of social science and test scores in reading of 95% and science 75%?

$$\pi = \frac{(e^{-9.561 + 0.098*xread + 0.066*xscience + 0.058*xses(1) - 1.013*xses(2)})}{1 + (e^{-9.561 + 0.098*xread + 0.066*xscience + 0.058*xses(1) - 1.013*xses(2)})}$$

$$\pi = \frac{(e^{-9.561 + 0.098*(.95) + 0.066*(.75) + 0.058*(0) - 1.013*(1)})}{1 + (e^{-9.561 + 0.098*(.95) + 0.066*(.75) + 0.058*(0) - 1.013*(1)})}$$