Handout: Multicategory Logistic Regression

Example 1

Contraceptive use in dependency on age (from G. Rodriguez, 2007, online notes)
The data has been taken of the report on the Demographic and Health Survey conducted in El Salvador in 1985 (FESAL-1985). The table shows 3165 currently married women classified by age, grouped in five-year intervals, and current use of contraception, classified as sterilization, other methods, and no method.

age * contra Crosstabulation

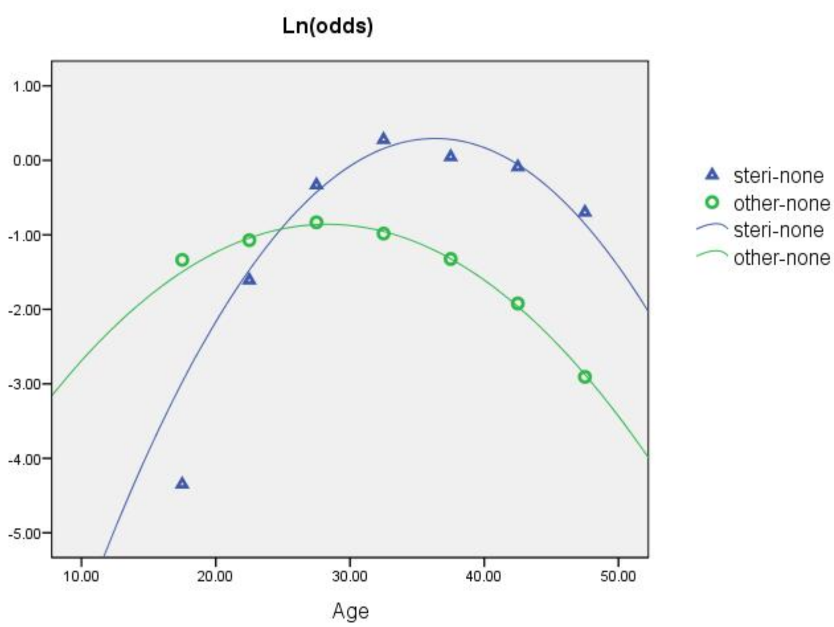
Count

		contra			Total
		n	o	s	
age	17.50	232	61	3	296
	22.50	400	137	80	617
	27.50	301	131	216	648
	32.50	203	76	268	547
	37.50	188	50	197	435
	42.50	164	24	150	338
	47.50	183	10	91	284
Total		1671	489	1005	3165

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	430.028[a]	12	.000
Likelihood Ratio	521.103	12	.000
N of Valid Cases	3165		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 43.88.



The graph indicates that there is a quadratic (non linear) relationship between the log odds (for sterilization versus none and other versus none) and age.
First we will fit a linear model for the logits and age, a multicategory logistic regression model with predictor age. The model equations are:

$$\begin{aligned} log(\pi_s/\pi_n) &= \alpha_s + \beta_s age, \\ log(\pi_o/\pi_n) &= \alpha_o + \beta_o age, \end{aligned}$$

Second we will fit a quadratic model for the logits and age, a multicategory logistic regression model with predictor age and age^2. The model equations are:

$$\begin{aligned} log(\pi_s/\pi_n) &= \alpha_s + \beta_{1s} age + \beta_{2s} age^2, \\ log(\pi_o/\pi_n) &= \alpha_o + \beta_{1o} age + \beta_{2o} age^2, \end{aligned}$$

The SPSS output can be found in file "CONTRACEPTIVE.PDF"

odds-ratio	variable	χ^2	*df*	*P*
sterilization - none	age	243.22	1	< 0.001
	age^2	218.25	1	< 0.001
other - none	age	31.47	1	< 0.001
	age^2	39.23	1	< 0.001
sterilization - other	age	54.79	1	< 0.001
	age^2	28.24	1	< 0.001

For getting the comparison between sterilization I reran the analysis with baseline category "other". The estimated model equations are:

$$\begin{aligned} log(\pi_s/\pi_n) &= -12.6 + .710\ age - 0.010\ age^2, \\ log(\pi_o/\pi_n) &= -4.5 + .264\ age + -0.005\ age^2, \end{aligned}$$

These are the curves shown in the graph above.

Example 2

$$\hat{\pi}_s = \frac{e^{-12.6+.710\ age-0.010\ age^2}}{e^{-12.6+.710\ age-0.010\ age^2} + e^{-4.5+.264\ age+-0.005\ age^2} + 1}$$

$$\hat{\pi}_o = \frac{e^{-4.5+.264\ age+-0.005\ age^2}}{e^{-12.6+.710\ age-0.010\ age^2} + e^{-4.5+.264\ age+-0.005\ age^2} + 1}$$

and

$$\hat{\pi}_n = \frac{1}{e^{-12.6+.710\ age-0.010\ age^2} + e^{-4.5+.264\ age+-0.005\ age^2} + 1}$$

Using these equations for age = 22.5:

$$\hat{\pi}_s = .13,\ \hat{\pi}_o = .23,\ \hat{\pi}_n = .64$$

To Based on the contingency tables we would estimate
for s: 80/617=0.129, o: 137/617=0.222, and n: 400/617=0.648.
This model has 6 parameters, the contingency table has $(3-1)\times(7-1) = 12$ degrees of freedom. Therefore the model needs a much smaller number of parameters to almost perfectly replicates the estimated probabilities.