Department of Statistics

COURSE STATS 330/762

Final exam 2008: Model answers for part B

1. (a) A model with interactions between all the factors and each continuous variable e.g. for factors A, B and continuous variables X,W we would fit the model A*B*X + A*B*W. There might not be enough data to fit such a model (requires enough points for each possible combination of the factor levels)

   (b) Stepwise regression, use of anova function, AIC.

   (c) All criteria choose model 6. Models 5 and 7 are not much worse on CV grounds

2. (a) The logistic model with response r/n, where r has a binomial distribution with parameters n and π, and logit (π) is a linear combination of the explanatory variables.

   (b) For continuous covariates, b is the change in log-odds corresponding to a unit increase in the covariates, assuming all other covariates are held constant. For factors, b is the change in the log-odds when the level is changed from the baseline (assuming main effects only). When there are interactions, there is no simple explanation, you must make a table of log-odds to assess the effect of changing the factor levels.

   (c) Note: the regression table was omitted – unfortunately I no longer have a copy.

   (d) A confidence interval for the log-odds is -2.272071 +/- 1.96* 0.5421498 or (-3.3347, -1.209457). Applying the function exp(x)/(1+exp(x)) to both ends of this interval we get (0.0344, 0.2298).

3. (a) Odds ratio is (assuming row facor in the 2x2 table is A and the column factor is B)

(odds of baseline of A at baseline level of) / (odds of baseline of A at other level of B)

   If the table is

		B	
		0	1
A	0	π_1	π_2
	1	π_3	π_4

then the OR is $(\pi_1 x\ \pi_4)/(\ \pi_2 x\ \pi_3)$.

The two factors are independent if the odds ratio is 1. To find a CI for the odds ratio we fit a Poisson regression count ~ A *B (the maximal model) and use the interaction line in the summary table to get the estimate and standard error of the log OR. An interval for the log OR is estimate +/- 1.96* std error. Exponentiating both ends gives a CI for the OR.

(b) The odds ratio s are the same when the 3-factor interaction is zero. To test if the 3-factor interaction is zero, fit the maximal model count~A*B*C and use the p-value in the 3-way interaction line in the summary table.

(c) The Defendant's race/Death penalty interaction is 0.3873 with a std error of 0.2785. The log OR confidence interval is 0.3873 +/- 1.96*0.2785 or (-0.1586 , 0.9332). Exponentiating gives (0.8534, 2.5425). Since this contains 1 the data are consistent with the OR being 1 (i.e. conditional independence of defendant's race and death penalty, given victim's race.)