STATS 330 Advanced Statistical Modeling

Model answers to 2007 exam, part B

1 (a) One explanatory variable is a linear function of the others. Inflates the variance

(b) VIF = factor by which the variance is increased by collinearity. Calculated by $1/(1-R^2_j)$

or by the diagonal elements of the inverse of the correlation matrix.

(c) High correlation between them (0.930)

(d) Unrelated to the response and the other variables (correlation -0.399)

(e) Models 4 (best BIC, AIC), and 5 (best CV and adjusted R^2)

2. (a) Pearson : (r-np)/sqrt(np(1-p)), deviance – contribution to deviance due to a single observation .

(b) Grouped: deviance, size of residuals, Ungrouped : gam plots, HL statistic.

(c) Points 26 and 44 influential, but HL statistic OK. Probably OK

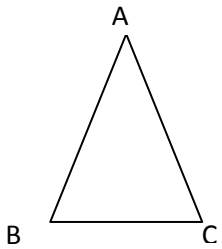
(d) B not significantly different from A, but P different from A, on the basis of p-values in the regression summary.

3. (a) Log m is a linear function of the covariates, response is Poisson with mean m, observations independent.

(b) (i) Fit model counts~A+B*C, check deviance p-value is not small, (ii) Fit model counts~A*C+B*C, check deviance p-value is not small.

(c) Homogeneous association model is OK (pvalue 0.2856)

(d)



(e) Log odds ratio for EducationHigh:PurchaseYes is +ve, so high education is associated with purchase . Log odds ratio for IncomeHigh:PurchaseYes is ive, so low income is associated with purchase.