STAT 659 — Solution to Exam 2 Summer 2014

1. (a) Do not reject $H_0 : \beta_{\texttt{may}} = \beta_{\texttt{had}} = 0$ at level $\alpha = 0.05$ since $G^2 = 98.539 - 91.726 = 6.812 < 12.59 = \chi^2_{6,0.05}$. There is insufficient evidence to warrant including `may` and `had` in the model, given that the other 9 predictors are included in the model.

   (b)
$$\log\left(\frac{\hat{\pi}_A}{\hat{\pi}_L}\right) = \log\left(\frac{\hat{\pi}_A}{\hat{\pi}_S}\right) - \log\left(\frac{\hat{\pi}_L}{\hat{\pi}_S}\right) = -0.9229 + 0.0263\texttt{the} - (-11.4740 + 0.1552\texttt{the}) = 10.5511 - 0.1289\texttt{the}$$
$$10.5511 - 0.1289\texttt{the} > 0 \Longleftrightarrow \texttt{the} < 81.855$$
   Thus, for $0 \leq \texttt{the} \leq 81$, $\hat{\pi}_A > \hat{\pi}_L$.

   (c) The odds of being Shakepeare rather than London for a given value of `the` is
$$\log\left(\frac{\hat{\pi}_S}{\hat{\pi}_L}\right) = -\log\left(\frac{\hat{\pi}_L}{\hat{\pi}_S}\right) = 11.4740 - 0.1552\texttt{the}.$$
   Then
$$\log(\widehat{OR}) = (11.4740 - 0.1552(\texttt{the} + 10)) - (11.4740 - 0.1552\texttt{the}) = -1.552.$$
   and $\widehat{OR} = e^{-1.552} = 0.2118$.

   (d)
$$\hat{\pi}_M = \frac{e^{-3.1494+100(0.0339)}}{1 + e^{-3.1494+100(0.0339)} + e^{-0.9929+100(0.0263)} + e^{-11.4740+100(0.1552)}} = 0.0197$$

2. (a) Using the saturated model, do not reject $H_0 : \beta_{G*D}$ since $X^2 = 1.4902 < 3.84 = \chi^2_{1,0.05}$ or $G^2 = 938.381 - 936.903 = 1.478 < 3.84 = \chi^2_{1,0.05}$. There is insufficient evidence to indicate that the odds ratios for `gender` and `depress` differ for the two levels of education.

   (b) Using the homogeneous association model, reject $H_0 : \beta_G = 0$ since $X^2 = 20.3695$ with a $P-$value < 0.0001. There is strong evidence of association between `depress` and `gender`, controlling for level of education.

   (c) i. $\text{logit}(\hat{\pi}(\texttt{high,female})) - \text{logit}(\hat{\pi}(\texttt{high,male})) = 0.7714$. Thus, $\widehat{OR} = e^{0.7714} = 2.1628$.

   ii. $\text{logit}(\hat{\pi}(\texttt{high,female})) - \text{logit}(\hat{\pi}(\texttt{high,male})) = 0.4818 + 0.4380 = 0.9198$. Thus, $\widehat{OR} = e^{0.9198} = 2.5088$.

3. (a) For Model A, the the two curves in the marginal model plot differ greatly. This indicates that the model that is linear in `age` is not appropriate and one or more nonlinear terms in `age` are needed. For Model B, the two plotted curves are nearly identical. This indicates that `age` is modelled appropriately in the linear predictor.

   (b) Reject $H_0 : \beta_{\texttt{age2}} = \beta_{\texttt{age3}} = \beta_{\texttt{age4}} = \beta_{\texttt{dur2}} = 0$ since $G^2 = 1749.965 - 1551.313 = 198.652 > 9.49 = \chi^2_{4,0.05}$. There is strong evidence that the additional polynomial terms improve the fit relative to that of Model B.

   (c) Model B has better predictive power than Model A since its ROC curve lies above the ROC curve for Model A. This indicates that for a given level of specificity, this model has a higher sensitivity. Also, the area under the curve (concordance index) is larger for Model B indicating greater predictive power.

   (d) The estimated coefficient for `univ` is $\hat{\beta}_{\texttt{univ}} = 1.6699$. Thus, the estimated odds of having at least one child are multiplied by $e^{1.6699} = 5.312$ if a woman has a university degree, keeping all the other variables constant.

   (e) Since the data are too sparse (997 unique profiles for 1761 observations), we use the Hosmer-Lemeshow Test for lack of fit. Since $X^2 = 8.1441$ with a $P-$value$= 0.4195$, we conclude that there is insufficient evidence to indicate a lack of fit for Model B.