STAT 659 — Solution to Exam 2 Summer 2015

1. (a) Since the $P-$values of the Wald tests of $H_0 : \beta_{\texttt{k618}} = 0$ and $H_0 : \beta_{\texttt{hc}} = 0$ equal 0.3424 and 0.5877, respectively, we could consider eliminating these variables. To determine whether this is appropriate, we should carry out a simultaneous test of $H_0 : \beta_{\texttt{k618}} = \beta_{\texttt{hc}} = 0$.

   (b) Since there are the same number of unique profiles of the predictors as the number of observations, we need to use the Hosmer-Lemeshow test for goodness of fit. Since the H-L $X^2 = 24.7058$ with a $P-$value$= 0.0917$, we conclude that there is strong evidence of lack of fit for this model.

   (c) The odds of women who went to college participating in the labor force are estimated to be $e^{\hat{\beta}_{wc}} = e^{0.8072} = 2.24$ times the odds of a woman who did not go to college.

   (d)
   - Estimated sensitivity: $\hat{P}(\hat{y} = 1|y = 1) = 290/428 = 0.6776$
   - Estimated specificity: $\hat{P}(\hat{y} = 0|y = 0) = 220/325 = 0.6769$
   - Estimated probability of correct classification: $(290 + 220)/753 = 0.6773$

   (e) The marginal model plot for `age` has similar smooth lines indicating that the model linear in age is appropriate. The marginal model plot for `lwg` has a curve for the smoothed observations being highly curved whereas the smoothed predicted values are much more linear. This indicates that the model with a linear term in `lwg` is not appropriate and some nonlinear model in `lwg` should be used.

2. (a) The response variable `pubs` is ordered, so a cumulative odds model is reasonable. The test for the proportional odds assumption has $X^2 = 5.7017$ with a $P-$value$= 0.3363$ indicating that the proportional odds assumption is reasonable.

   (b)
   - Proportional odds model: Reject $H_0 : \beta_{\texttt{mar}} = 0$ at level 0.05 since $X^2 = 3.9481$ with a $P-$value $= 0.0469$ and conclude that `mar` is useful in the model, given that the other predictors are present.
   - Baseline category model: Do not reject $H_0 : \beta_{\texttt{mar},1} = \beta_{\texttt{mar},2} = 0$ at level 0.05 since $X^2 = 4.3792$ with a $P-$value $= 0.1120$ and conclude that `mar` is not useful in the model, given that the other predictors are present.

   (c) To make $P(\texttt{pubs} = 0)$ to be a large as possible, we need $\text{logit}(P(Y \leq 0))$ to be large. To make the linear predictor as large as possible, choose the largest value of predictors with positive coefficients and the smallest value of variables with negative coefficients. Thus, we would choose $\texttt{fem} = 1,\ \texttt{mar} = 0,\ \texttt{kid5} = 3,\ \texttt{phd} = 0.75,\ \texttt{ment} = 0$.

   (d) $\log(\hat{\pi}_2/\hat{\pi}_1) = -0.351 + 0.092x - (-0.531 + 0.062x) = 0.180 + 0.030x$. Now $\hat{\pi}_1 > \hat{\pi}_2$ iff $\log(\hat{\pi}_2/\hat{\pi}_1) = 0.18 + 0.03x < 0$. This occurs for $x < -0.18/0.03 = -6$. Since $x \geq 0$, $\hat{\pi}_1 < \hat{\pi}_2$ for all values of x.

3. (a) Since $G^2 = 9.0862$ with a $P-$value $= 0.0106$, we reject $H_0 : \beta_{P\times G} = 0$ and conclude that there is strong evidence that the ORs between `gender` and `accept` differ for the three programs.

   (b) Using the homogeneous association model, reject $H_0 : \beta_G = 0$ since the Wald $X^2 = 23.176$ with a $P-$value < 0.0001. There is strong evidence of partial association between `gender` and `accept`, controlling for `program`.

   (c)
   i. Homogeneous association model: $\text{logit}(\hat{\pi}(f, p)) - \text{logit}(\hat{\pi}(m, c)) = 0.5184 + 0.6612 + 0.00653 - (0.5184 + 0 - 3.5477) = 4.21543)$. Then $\widehat{OR} = e^{4.21543} = 67.72$.

   ii. Model with interaction: $\text{logit}(\hat{\pi}(f, p)) - \text{logit}(\hat{\pi}(m, c)) = 0.5339 + 0.2334 - 0.0429 + 0.8389 - (0.5339 + 0 - 3.3162 + 0) = 4.3456$. Then $\widehat{OR} = e^{4.3456} = 77.14$.