STAT 659 — Solution to Final Exam **Summer 2015**

1. (a) $\hat{P}(\texttt{pubs} = 0) = 1/(1 + e^{-0.25+10(-.06)-0.07} + e^{-0.58+10(0.09)+0.09}) = 0.2581.$
   (b) $\text{logit}(\hat{P}(Y \leq 1) = 1.04 + 10(-.06) - 0.09 =).3795$. Then $\hat{P}(Y = 2) = 1 - e^{0.3795}/(1 + e^{0.3795}) = 0.4062$.

2. (a) Since $X^2 = (449 - 528)^2/(449 + 528) = 6.388 > 3.84 = \chi^2_{1,0.05}$, we reject $H_0 : \pi_{1+} = \pi_{+1}$ and conclude that there is strong evidence that the proportions of fathers in the two status categories differs from the proportions of sons in the two categories.
   (b)
   - Marginal odds ratio: $\widehat{OR} = (2420/895)/(2341/974) = 1.125$
   - Conditional odds ratio: $\widehat{OR} = 528/449 = 1.176$

   (c) All the models show strong lack of fit except for the quasi-symmetry model ($G^2 = 2.3 < 7.81$). Since the QS model fits well, it is the most appropriate model.
   (d)
   - Since $G^2 = 27.8 > 7.81 = \chi^2_{3,0.05}$, there is strong evidence that the marginal homogeneity model does not fit these data.
   - Since $G^2(Symm|QS) = 30.1 - 2.3 = 27.8 > 7.81 = \chi^2_{3,0.05}$, there is strong evidence that the symmetry model does not hold, given quasi-symmetry. This implies that the QS model does not have marginal homogeneity.

3. (a) i. General alternative: Since $G^2 = 11.03 < 13.51 = \chi^2_{8,0.05}$, there is insufficient evidence to reject the independence model in favor of the most general model, the saturated model.
   ii. Ordered alternative: Since $G^2 = 11.03 - 6.09 = 4.94 < 3.84 = \chi^2_{1,0.05}$, there is sufficient evidence to reject the independence model in favor of the linear by linear model.
   (b) All the models fit the data (compare the deviances to the chi-square percentile). Since the linear by linear model improves upon simplest model (independence) and is not improved upon by the row effects model ($G^2 = 6.09 - 2.43 = 3.66 < 7.81$) or the column effects model ($G^2 = 6.09 - 4.88 = 1.21 < 3, 84$), we choose the linear by linear model.

4. (a) Since $G^2 = 9.1 > 5.99 = \chi^2_{2,0.05}$, we reject H_0 : All $\lambda^{AGP} = 0$ in the saturated model. This implies that there are differences in the `gender` by `accept` ORs for the three programs.
   (b) Since $G^2 = 34.0 - 9.1 = 24.9 > 3.84 = \chi^2_{1,0.05}$, we reject H_0 : All $\lambda^{AG} = 0$ in the homogeneous association model. This implies that there is strong evidence of partial association between `gender` and `accept`, controlling for `program`.
   (c) All models except for the saturated model exhibit a lack of fit. The next best model is the homogeneous association model which we rejected in part (a). We select the saturated model, (AGP).
   (d)
   - Homogeneous association model: $\widehat{OR} = e^{0.6614} = 1.937$
   - Saturated model: $\widehat{OR} = e^{0.2334+0.8389} = 2.9221$

5. (a) The model with smallest AIC is Model 4. Since $G^2 = 754 - 752.9 = 1.1 < 3.84$, the more complex Model 3 does not improve upon Model 4. Since $G^2 = 770.2 - 754 = 16.2 > 3.84$, Model 4 improves upon the simpler Model 5. Thus, we select Model 4.
   (b)
$$\hat{\pi} = \frac{e^{3.03-1.53-30(0.07)}}{1 + e^{3.03-1.53-30(0.07)}} = 0.3401.$$