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

1. (a) Reject $H_0 : \pi_{1+} = \pi_{+1}$ at level 0.05 since

$$X^2 = \frac{(150-20)^2}{150+20} = 99.4 > 3.84 = \chi^2_{1,0.05}.$$

There is strong evidence that the proportion who are proud of America's economic progress differs from the proportion who are proud of America's science and technology progress.

(b)
- Marginal odds ratio: $\widehat{OR} = (912/171)/(1042/41) = 0.2099$
- Conditional odds ratio: $20/150 = 0.1333$

(c) The independence, symmetry, and marginal homogeneity model all have very poor fit. The quasi-independence ($G^2 = 10.5 < 11.07 = \chi^2_{5,0.05}$), ordinal quasi-symmetry ($G^2 = 8.89 < 11.07 = \chi^2_{5,0.05}$), and quasi-symmetry ($G^2 = 3.42 < 7.81 = \chi^2_{3,0.05}$). Since both the QI and OQS models have 5 df, we choose the OQS model with a smaller deviance. We then compare the OQS and QS models using $G^2 = 8.89 - 3.42 = 5.47 < 5.99 = \chi^2_{2,0.05}$). Thus, there is not significant evidence that the more complex QS model improves on the OQS model. We will choose the OQS model.

(d)
- Since $G^2 = 197.9 > 7.81 = \chi^2_{3,0.05}$), we reject the fit of the marginal homogeneity model and conclude that there is strong evidence that the marginal probabilities for the different opinions on America's economic progress and America's science and technology progress differ.
- Since $G^2 = 218.0 - 3.42 = 214.58 > 7.81 = \chi^2_{3,0.05}$), we reject the symmetry model in favor of the QS model. This implies that the marginal probabilities for the different opinions on America's economic progress and America's science and technology progress differ

2. (a) The model with the lowest AIC is the main effects model. To see whether the model that also included two-factor interactions improves on this model, we find $G^2 = 11.15 - 3.74 = 7.68 < 12.59 = \chi^2_{6,0.05}$ and conclude that the main effects model is adequate.

(b) To compare the new model to the main effects model, we find $G^2 = 11.15 - 5.23 = 5.92 < 5.99 = \chi^2_{2,0.05}$ indicating that the new model does not significantly improve on the main effects model. To compare the new model to the main effects and two-factor interactions model, we find $G^2 = 5.23 - 3.74 = 1.49 < 9.49 = \chi^2_{4,0.05}$ indicating including the 4 additional interactions does not significantly improve on our new model. In part (a) we determined that the main effects model was preferable to the model that also include the two-factor interactions. Thus, we would use the main-effects model.

(c) (i) $(YE, YJ, YS, YT, EJST)$ (ii) $(YEJ, YST, EJST)$

(d) The estimated log-odds-ratio is $\text{logit}(\hat{\pi}(ESTJ)) - \text{logit}(\hat{\pi}(INTJ)) = 0.5550 - 0.4292 = 0.1258$. The estimated odds-ratio is $e^{0.1258} = 1.134$.

(e) The smallest logit corresponds to the smallest value of the linear predictor. Since the coefficents of EI and TF are positive, we choose I and F. Since the coefficients of SN and JP are negative, we choose S and J. Thus, the $ISFJ$ personality type has the lowest estimated probability of drinking alcohol frequently.

(f) i. $\widehat{OR} = \frac{49\times 602}{48\times 351} = 1.751$

ii. $\widehat{OR} = e^{0.6873} = 1.988$

3. (a) Since $G^2((DG, DS, GS)|(DGS)) = 0.77 < 3.84 = \chi^2_{1,0.05}$, do not reject H_0 : All $\lambda^{DGS}_{ijk} = 0$ and conclude that there is insufficient evidence to indicate that the odds rations between `smoke` and `depress` are different for males and females.

(b) Since $G^2((DG, GS)|(DG, DS, GS)) = 33.0 - 0.77 = 32.23 > 3.84 = \chi^2_{1,0.05}$, reject H_0 : All $\lambda^{DS}_{ij} = 0$ and conclude that there is partial association between `smoke` and `depress`, controlling for gender.

(c) The only models with adequate fit are the homogeneous association model and the saturated model. In part (a), we found the the saturated model does not improve on the homogeneous association model. Thus, we should use the homogeneous association model.

- Homogeneous association model (let the subscripts represent in order `smoke,gender,depress`)

$$
\begin{aligned}
\log(\widehat{OR}) &= \log\left(\frac{\hat{\mu}_{111}\hat{\mu}_{100}}{\hat{\mu}_{110}\hat{\mu}_{101}}\right) \\
&= (6.04 + 0.75 + 0.73 - 3.97 - 0.78 + 0.92 + 0.94) + (6.04 + 0.75) \\
&\quad -(6.04 + 0.75 + 0.73 - 0.78) - (6.04 + 0.75 - 3.97 + 0.92) = 0.94
\end{aligned}
$$

  Thus, $\widehat{OR} = e^{0.9369} = 2.552$.

- Saturated model

$$
\begin{aligned}
\log(\widehat{OR}) &= \log\left(\frac{\hat{\mu}_{111}\hat{\mu}_{100}}{\hat{\mu}_{110}\hat{\mu}_{101}}\right) \\
&= (6.03 + 0.76 + 0.74 - 3.73 - 0.80 + 0.63 + 0.65 + 0.36) + (6.03 + 0.76) \\
&\quad -(6.03 + 0.76 + 0.74 - 0.80) - (6.03 + 0.76 - 3.73 + 0.63) = 0.65 + 0.36 = 1.01
\end{aligned}
$$

Thus, $\widehat{OR} = e^{1.0122} = 2.752$.