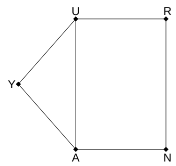
STAT 659 — Solution to Final Exam Summer 2014

1. (a) Since $G^2 = 46.13 > 9.49 = \chi^2_{4,0.05}$ and $G^2 = 14.71 > 7.81 = \chi^2_{3,0.05}$, the independence and symmetry models have a definite lack of fit. Since $G^2 = 3.20 < 5.99 = \chi^2_{2,0.05}$ and $G^2 = 0.24 < 3.84 = \chi^2_{1,0.05}$, both the ordinal quasi-symmetry and quasi-symmetry models are fit well by the data. Since $G^2 = 3.20 - 0.24 = 2.96 < 3.84 = \chi^2_{1,0.05}$, the QS model does not improve significantly on the OQS model. Thus, the OQS model is the most parsimonious model that adequately fits the data.

   (b) Since $G^2 = 14.71 - 0.24 = 14.47 > 5.99 = \chi^2_{2,0.05}$, the QS model significantly improves upon the symmetry model. This implies that marginal homogeneity does not hold for this table.

2. (a)
   - $(YU, YN, YR, YA, UNRA)$
   - $(YUA, YN, YR, UNRA)$

   (b) Y is conditionally independent of $\{R, N\}$ given $\{U, A\}$. Thus, Y is also conditionally independent of $\{R\}$ given $\{U, A\}$ and of $\{N\}$ given $\{U, A\}$.



3. (a) All of the following tables are more extreme:

22	1
18	14

23	0
17	15

   (b) We test $H_0 : \theta = 1$ versus $H_a : \theta > 1$ using Fisher's exact test. The $P-$value= 0.0083 and the mid-$P-$value= $0.0083 - 0.0074/2 = 0.0046$. Both the $P-$value and the mid-$P-$value provide strong evidence that a greater proportion of the mice exposed to smoke had tumors than did the control mice.

4. (a) Do not reject $H_0 : \lambda^{DGE} = 0$ in the saturated model since $G^2 = 1.48 - 0 = 1.48 < 3.84 = \chi^2_{1,0.05}$. There is insufficient evidence to indicate that the odds ratio between `gender` and `depress` differ for the two levels of education.

   (b) Using the homogeneous association model, reject $H_0 : \lambda^{DG} = 0$ since $G^2 = 22.82 - 1.48 = 21.34 > 3.84 = \chi^2_{1,0.05}$. There is strong evidence of association between `gender` and `depress`, controlling for the two levels of education.

   (c) All the models except (DG, GE), (DG, DE, GE), and (DGE) have strong evidence of a lack of fit. In part (a), we determined that the three-way interaction was not significant, so we would prefer (DG, DE, GE) to the saturated model. Since $G^2 = 5.39 - 1.48 = 3.91 > 3.84 = \chi^2_{1,0.05}$, we find that the DE interaction is needed. Thus, we recommend the homogeneous association model, (DG, DE, GE).

5. (a) Table I is more appropriate since the data are paired with each individual receiving both allergy tests. To test $H_0 : \pi_{1+} = \pi_{+i}$, we use McNemar's test. Since $X^2 = (12-28)^2/(12+28) => 3.84 = \chi^2_{1,0.05}$, we reject H_0 and conclude that the proportion of individuals suffering allergic reactions differs for the two types of penicillin.

   (b) First, $\widehat{OR} = 28/12 = 2.333$. Then a 95% confidence interval for $\log(OR)$ is $\log(28/12) \pm 1.96\sqrt{1/28 + 1/12} = 0.847 \pm 0.676$, or $(0.171, 1.523)$. Exponentiate the endpoints to obtain the 95% confidence interval for OR: $(1.186, 4.589)$. The odds of an individual being allergic to type BT penicillin are from 1.2 to 4.6 times the odds of an individual being allergic to Type G penicillin.

6. (a) The 95% confidence interval for $\beta_{\texttt{Tr}}$ is given by $0.9379 \pm (1.96)(0.3310)$ or $(0.289, 1.587)$. We exponentiate the end points to obtain a 95% confidence interval for the OR for the cumulative logit of the treatment group relative to the control group: $(1.335, 4.887)$. Thus, treatment group has a larger cumulative probability $P(\texttt{y} \leq j)$ for a fixed age. This implies that the treatment group has a higher probability for low values of `y` which implies less pain for individuals in this group. Thus, the treatment appears effective.

   (b)

$$P[\texttt{y} \leq 4] = \frac{e^{-3.2271+0.9379}}{1+e^{-3.2271+0.9379}} = 0.0902 \Longrightarrow P[\texttt{y} = 5] = 1 - P[\texttt{y} \leq 4] = 1 - 0.0902 = 0.9098$$

   Note: Due to a data entry error, the intercepts and the coefficient for `Age1` were not correct on the output. Intercept 4 should be 2.3825 resulting in $P[\texttt{y} \leq 4] = 0.965$ and $P[\texttt{y} = 5] = 0.035$. This matches up well with the observed 2 of 64 treatment patients with severe pain. The output used to answer parts (a) and (c) of this problem was not affected.

   (c) Since the response is ordered, we can consider a proportional odds model. However, the test for the POM has $X^2 = 21.3978$ with a $P-$value$= 0.0110$. This indicates that the POM model may not fit the data well.