STAT 659 — Solution to Exam 1 **Summer 2015**

1. (a) To test $H_0 : \pi = 0.5$ versus $H_a : \pi > 0.5$, the $P-$value= $P[X \geq 10] = 0.016 + 0.003 = 0.019$ and the mid-$P-$value= $0.5P[X = 10] + P[X \geq 11] = (0.5)(0.016) + 0.003 = 0.011$. Both these are less than 0.05, so we can reject H_0 and conclude that more than half the individuals that eat the fiber-enriched crackers experience some bloating.

   (b) Using the Agresti-Coull interval, we use $p = (10+2)/(12+4) = 0.75$ with $SE = \sqrt{(0.75)(0.25)/16} = 0.108$ resulting in $0.75 \pm (1.96)(0.108) = 0.75 \pm 0.212$. Thus, $(0.538, 0.962)$ is a 95% confidence interval for π, the proportion who suffer some bloating.

   (c) $\begin{array}{c|c} 11 & 1 \\ \hline 5 & 7 \end{array} \qquad \begin{array}{c|c} 12 & 0 \\ \hline 4 & 8 \end{array}$

   (d) Since the sample size is small, we use Fisher's exact test for $H_0 : \theta = 1$ versus $H_a : \theta > 1$. The reported right-sided $P-$value= 0.0965 and the mid-$P-$value= $0.0965 - (0.5)(0.0829) = 0.0551$. Thus, we fail to reject H_0 and conclude that there is insufficient evidence to conclude that there is a positive association between eating fiber-enriched crackers and experiencing bloating.

2. (a) The marginal $\widehat{OR}$ is 2.6328 whereas the partial $\widehat{OR}$s are 0.3422, 0.7919, and 0.7964. Since the association is in a different direction in the partial tables from that in the marginal table, Simpson's paradox is present.

   (b)

	(i)test of equal odds ratios	(ii) the test of partial association
Value of test statistic	9.0481	23.8134
$P-$value	0.0108	$< .0001$
Conclusion	Reject H_0 and conclude that the ORs differ for the 3 programs.	Reject H_0 and conclude that there is an association between `gender` and `accept`, controlling for `program`.

   (c) i. Common odds ratio: (0.3903,0.6712) or (0.4037,0.7029)

   ii. Marginal odds ratio: $(2.2145, 3.1300)$

   iii. Comment on appropriateness: Since Simpson's paradox is present, the marginal OR does not summarize the nature of the association in the partial tables. Since the BD test led us to conclude that the ORs differ, we cannot use the confidence interval for the common OR.

3. (a) The estimated expected for the zero cell is $E_0 = (915)(0.1840) = 168.36$ and its contribution to the chi-squared statistic is $(275 - 168.36)^2/168.36 = 67.546$. Since $X^2 = 246.14 > 11.07 = \chi^2_{5,0.05}$, we can reject H_0 : "The Poisson distribution fits the data" and conclude that fit by the Poisson distribution is not adequate.

   (b) We reject $H_0 : \beta_{\texttt{phd}} = \beta_{\texttt{mar}} = 0$ since $G^2 = 1640.85 - 1634.37 = 6.48 > 5.99 = \chi^2_{2,0.05}$. Thus, we cannot simulaneously omit both `phd` and `mar` from the model.

   (c) Keeping all other variables constant, $\log(\hat{\mu}_{\texttt{unmar,fem}}) - \log(\hat{\mu}_{\texttt{mar,male}}) = -0.2246 - 0.1552 = -0.3798$. Thus, $\hat{\mu}_{\texttt{unmar,fem}}/\hat{\mu}_{\texttt{mar,male}} = e^{-0.3798} = 0.684$. A unmarried female's estimated number of publications is 68.4% of that of a married male, keeping other variables constant.

   (d) The model with the lowest AIC_C is Model 2. To determine if Model 1 improves on Model 2, we fail to reject $H_0 : \beta_{\texttt{phd}} = 0$ since $G^2 = 1634.6 - 1634.4 = 0.2 < 3.84 = \chi^2_{1,0.05}$. To see whether we can omit `mar` from Model 2, we reject $H_0 : \beta_{\texttt{mar}} = 0$ since $G^2 = 1640.9 - 1634.6 = 6.3 > 3.84 = \chi^2_{1,0.05}$ and conclude that `mar` is needed in the model. Thus, we choose Model 2.

   (e) The deviance/df for the Poisson model is 1.798 indicating a possible lack of fit, whereas this ratio was 1.105 for the negative binomial model. Also, the confidence interval for the dispersion parameter in the negative binomial model was (0.349,0.559) which does not include zero. Thus, there is evidence of overdispersion in the data relative to the Poisson model.