STAT 659 — Solution to Exam 1
Summer 2014

1. (a) Since the proportion of patients suffering from cancer was determined by the study design, we cannot use these data to estimate the proportion of cancer patients among those subjects exposed to passive smoke or those not exposed to passive smoke, and hence, we cannot estimate the relative risk.

   (b) $\widehat{OR} = \frac{120\times155}{115\times80} = 2.022$ and $SE = \sqrt{1/120 + 1/155 + 1/115 + 1/80} = 0.1897$. Thus, a 95% confidence interval for OR is $\exp(\log(2.022) \pm 1.96(0.1897))$ or $(1.394, 2.933)$.

2. (a) The marginal odds-ratio for gender and depression is 2.250. The partial odds-ratios are 1.619 for `educ=low` and 2.509 for `educ=high`. Since these are in the same direction and of comparable magnitudes, Simpson's paradox is not present for these data.

   (b)

	(i) the test of equal odds ratios	(ii) the test of partial association
Value of test statistic	$BD = 1.4942$	$CMH = 20.7447$
$P-$value	0.2216	$< .0001$
Conclusion	Fail to reject $H_0 : \theta_1 = \theta_2$. There is insufficient evidence to indicate that the odds ratios differ for the two levels of education.	Reject $H_0 : \theta_1 = \theta_2 = 1$. There is strong evidence of association between gender and depression, controlling for level of education.

   (c) The 95% confidence interval for the common odds ratio is $(1.5442, 3.0139)$ (Mantel-Haenszel) or $(1.5399, 3.0106)$ (logit). Since we did not find evidence that the odds ratios differed using the Breslow-Day test, it is appropriate to use a confidence interval for a common odds ratio.

3. To test $H_0 : \pi = 0.6$ versus $H_a : \pi > 0.6$, we use the rejection region $Y \geq c$. For the observation $Y = 10$, the $P-$ value equals $P[Y \geq 10] = 0.064 + 0.017 + 0.002 = 0.083$. The mid $P-$value equals $0.5P[Y = 10] + P[Y \geq 11] = 0.5(0.064) + 0.017 + 0.002 = 0.051$. Both indicate that we fail to reject H_0 at level 0.05 and that there is insufficient evidence to conclude that the new treatment is more effective in relieving joint pain.

4. (a) The estimated expected frequency is the first cell is $\hat{E}_i = 1761(0.2074) = 365.23$, and the contribution of the first cell to the chi-squared statistic is $(398 - 365.23)^2/365.23 = 2.940$. Since $X^2 = 70.3774 > 9.49 = \chi^2_{6-1-1,0.05}$, we reject H_0 : "The Poisson distribution fits the data" and conclude that the Poisson distribution is not an adequate fit for the number of children.

   (b) Since $G^2 = 1915.0 - 1718.6 = 196.4 > 9.49 = \chi^2_{4,0.05}$, reject $H_0 : \beta_{\texttt{age2}} = \beta_{\texttt{age3}} = \beta_{\texttt{age4}} = \beta_{\texttt{dur2}} = 0$ and conclude that the additional polynomial terms improve the fit of the model.

   (c) Keeping all other predictors constant, $\log(\hat{\mu}_{\texttt{univ}}) - \log(\hat{\mu}_{\texttt{nouniv}}) = \hat{\beta}_{\texttt{univ}} = 0.5858$. Thus, $\hat{\mu}_{\texttt{univ}}/\hat{\mu}_{\texttt{nouniv}} = e^{0.5858} = 1.796$. Thus, women who attend university have an increase of around 80% in the estimated mean number of children over those who not attend university.

   (d) The model with the smallest AIC_C is model 3. If we compare it to model 2, $G^2 = 1725.6 - 1718.8 = 6.8 < 11.05 = \chi^2_{5,0.05}$. Thus, the extra terms in model 2 do not significantly improve upon model 3. On the other hand, when we compare model 4 to model 3, $G^2 = 1733.8 - 1725.6 = 6.8 < 8.2 > 3.84 = \chi^2_{1,0.05}$. Thus, model 3 significantly improves upon model 2. We chose model 3 as the most appropriate model.

5. (a) First, $\hat{\pi} = 911/2955 = 0.3083$. Thus, a 95% confidence interval for the proportion of *very happy* Americans is

$$0.3083 \pm 1.96\sqrt{\frac{(0.3083)(1-0.3083)}{2955}} = 0.3083 \pm 0.0167 \quad \text{or} \quad (0.292, 0.325).$$

(b)

	(i) general alternative	(ii) ordered alternative
Value of test statistic	$X^2 = 172.2637$	$CMH = 146.5501$
$P-$value	$< .0001$	$< .0001$
Conclusion	Reject the null hypothesis of independence of income and happiness. There is strong evidence of association between income and happiness.	Reject the null hypothesis of independence of income and happiness. There is strong evidence of a linear trend in the association of income and happiness.