Categorical Data Analysis
Assignment III
Solution Set

3.2

Since the study at hand was a case-control study, it is not possible to estimate the difference in proportions or the relative risk directly from the data.

An appropriate measure of association is the odds ratio.

$$\hat{\theta} = \frac{(688)(59)}{(650)(21)} = 2.9738$$

$$\log \hat{\theta} = 1.0898$$

$$\hat{\sigma}(\log \hat{\theta}) = \left(\frac{1}{688} + \frac{1}{650} + \frac{1}{21} + \frac{1}{59} \right)^{1/2} = 0.2599$$

95% CI for $\log \theta$:

$$\log \hat{\theta} \pm z_{\alpha/2} (\hat{\sigma}(\log \hat{\theta}))$$

$$(1.0898) \pm (1.96)(0.2599)$$

$$(0.5804, 1.5992)$$

95% CI for θ:

$$(\exp(0.5804), \exp(1.5992))$$

$$(1.7867, 4.9493)$$

The odds ratio estimate indicates that the odds of developing lung cancer are 2.9738 times higher for a smoker than for a non-smoker.

The confidence interval for the odds ratio suggests that the true odds ratio is at least 1.7867 and at most 4.9493.

3.3

	Second Shot	
First Shot	Made	Missed
Made	251	34
Missed	48	5

We will conduct the Pearson chi-squared test for independence.

$$X^2 = \Sigma_i \Sigma_j \frac{(n_{ij} - \hat{\mu}_{ij})^2}{\hat{\mu}_{ij}}$$

$$= 0.0049 + 0.0378 + 0.0265 + 0.2034$$

$$= 0.2727$$

P = 0.6015. There is insufficient evidence to indicate a dependence between the outcomes of successive free throws.

3.7

SAS

	MI		
	Fatal	Non-Fatal	No Attack
Placebo	18	171	10,845
Aspirin	5	99	10,933

$G^2 = 27.5893$ (2 df) ; $P < .0001$

	MI	
	Fatal	Non-Fatal
Placebo	18	171
Aspirin	5	99

$G_1^2 = 2.2173$ (1 df) ; $P = .1365$

	MI	
	Attack	No Attack
Placebo	189	10,845
Aspirin	104	10,933

$G_2^2 = 25.3720$ (1 df) ; $P < .0001$

Note: $G^2 = G_1^2 + G_2^2$.

The result of the initial test (on the 2×3 table) indicates that myocardial infarction status is dependent on aspirin intake.

The results of the partitioning support the following conclusions:

(1) for those experiencing a heart attack, whether the attack is fatal or non-fatal is <u>NOT</u> dependent on aspirin intake,

(2) whether one experiences a heart attack (either fatal or non-fatal) <u>IS</u> dependent on aspirin intake.

3.11

SAS

(a) $X^2 = 8.8709$ (6 df); $P = 0.1810$

$G^2 = 8.9165$ (6 df); $P = 0.1783$

The Pearson chi-squared and likelihood-ratio chi-squared tests do not provide compelling evidence of a dependence between family income and educational aspirations.

These tests are not optimal for the data at hand because the tests are designed for nominal variables, and yet the variables in the present application are ordinal.

(c) $M^2 = 0.6316$ (1 df); $P = 0.4268$.

The CMH linear trend test is more appropriate for the data at hand because it is designed for ordinal variables.

However, in the present application, the CMH test is even less powerful than the Pearson chi-squared or likelihood-ratio chi-squared test, yielding a larger p-value. Thus, the CMH test also fails to provide evidence of a dependence between family income and educational aspirations.

3.12

SAS

$\hat{\gamma} = 0.3873$

$\hat{\sigma}(\hat{\gamma}) = 0.0366$ (ASE in SAS)

The Wald CI for γ is of the form
$\hat{\gamma} \pm z_{\alpha/2}(\hat{\sigma}(\hat{\gamma}))$.

95% CI for γ:

$$\hat{\gamma} \pm z_{\alpha/2}(\hat{\sigma}(\hat{\gamma}))$$

$$(0.3873) \pm (1.96)(0.0366)$$

$$(0.3156, 0.4590)$$

Based on the point estimate of γ and the CI for γ, it appears that there is a positive association of moderate strength between schooling and attitude on abortion.

3.22

From the class notes from §1.4, we have the following result for the large-sample distribution of $\hat{\pi}$:

$$(\hat{\pi} - \pi) \,\dot{\sim}\, N\left(0, \frac{\pi(1-\pi)}{n}\right).$$

(Here, the variance of $\hat{\pi}$ is given by the reciprocal of $I(\pi) = n/(\pi(1-\pi))$.)

For application of the delta method, we have

$$g(\pi) = \log(\pi/(1-\pi)),$$

$$\begin{aligned} d(\pi) &= \frac{\partial}{\partial \pi}\left[\log\left(\frac{\pi}{(1-\pi)}\right)\right] \\ &= \frac{(1-\pi)}{\pi}\left[\frac{(1-\pi)(1) - (\pi)(-1)}{(1-\pi)^2}\right] \\ &= \frac{(1-\pi)}{\pi}\left[\frac{1}{(1-\pi)^2}\right] \\ &= \frac{1}{\pi(1-\pi)}, \end{aligned}$$

$$\Sigma(\pi) = \frac{\pi(1-\pi)}{n} \qquad \text{(from above)}.$$

Thus,

$$\begin{aligned} (d(\pi))^2\,\Sigma(\pi) &= \left(\frac{1}{\pi(1-\pi)}\right)^2\left(\frac{\pi(1-\pi)}{n}\right) \\ &= \frac{1}{n\pi(1-\pi)}. \end{aligned}$$

Therefore, we have the following result for the large-sample distribution of $\log(\hat{\pi}/(1-\hat{\pi}))$:

$$\left(\log(\hat{\pi}/(1-\hat{\pi})) - \log(\pi/(1-\pi))\right) \mathrel{\dot\sim} N\left(0, \frac{1}{m\pi(1-\pi)}\right).$$

The preceding leads to the Wald CI for $\log(\pi/(1-\pi))$ provided in the problem statement.

To use the Wald CI to obtain a CI for π, we employ the inverse transformation corresponding to $g(\pi)$: i.e., $g^{-1}(x) = e^x/(1+e^x)$. If (a,b) denotes the Wald CI for $\log(\pi/(1-\pi))$, the associated CI for π is given by $(e^a/(1+e^a), e^b/(1+e^b))$.

3.34 (a)
$$\begin{aligned}
X^2 &= \Sigma_i \frac{(m_i-\hat{\mu}_i)^2}{\hat{\mu}_i} \\
&= \Sigma_i \left(\frac{m_i^2 - 2m_i\hat{\mu}_i + \hat{\mu}_i^2}{\hat{\mu}_i}\right) \\
&= \Sigma_i \left(\frac{m_i^2}{\hat{\mu}_i} - 2m_i + \hat{\mu}_i\right) \\
&= \Sigma_i \left(\frac{m_i^2}{\hat{\mu}_i}\right) + \Sigma_i(\hat{\mu}_i - 2m_i) \\
&= \Sigma_i \left(\frac{m_i^2}{\hat{\mu}_i}\right) + (-m) \\
&= \Sigma_i \left(\frac{m_i^2}{\hat{\mu}_i}\right) + (-\Sigma_i m_i) \\
&= \Sigma_i m_i \left(\frac{m_i}{\hat{\mu}_i} - 1\right)
\end{aligned}$$

$$= \frac{2}{(1)(1+1)} \Sigma_i m_i \left(\left(\frac{m_i}{\hat{\mu}_i} \right)^{(1)} - 1 \right)$$

(b) $$\lim_{\lambda \to 0} \frac{2}{\lambda(\lambda+1)} \Sigma_i m_i \left(\left(\frac{m_i}{\hat{\mu}_i} \right)^{\lambda} - 1 \right)$$

$$= \left(\lim_{\lambda \to 0} \frac{2}{(\lambda+1)} \right) \left\{ \Sigma_i m_i \left[\lim_{\lambda \to 0} \left(\frac{\left(\frac{m_i}{\hat{\mu}_i}\right)^{\lambda} - 1}{\lambda} \right) \right] \right\}$$

$$= (2) \left\{ \Sigma_i m_i \left[\log \left(\frac{m_i}{\hat{\mu}_i} \right) \right] \right\}$$

$$= G^2$$

(c) $$\lim_{\lambda \to -1} \frac{2}{\lambda(\lambda+1)} \Sigma_i m_i \left(\left(\frac{m_i}{\hat{\mu}_i} \right)^{\lambda} - 1 \right)$$

$$= \lim_{\lambda \to -1} \frac{2}{\lambda(\lambda+1)} \Sigma_i m_i \left(\left(\frac{\hat{\mu}_i}{m_i} \right) \left(\frac{m_i}{\hat{\mu}_i} \right)^{\lambda+1} - 1 \right)$$

$$= \lim_{\lambda \to -1} \frac{2}{\lambda(\lambda+1)} \Sigma_i \left(\hat{\mu}_i \left(\frac{m_i}{\hat{\mu}_i} \right)^{\lambda+1} - m_i \right)$$

$$= \lim_{\lambda \to -1} \frac{2}{\lambda(\lambda+1)} \left[\Sigma_i \left(\hat{\mu}_i \left(\frac{m_i}{\hat{\mu}_i} \right)^{\lambda+1} \right) - \Sigma_i m_i \right]$$

$$= \lim_{\lambda \to -1} \frac{2}{\lambda(\lambda+1)} \left[\Sigma_i \left(\hat{\mu}_i \left(\frac{m_i}{\hat{\mu}_i} \right)^{\lambda+1} \right) - \Sigma_i \hat{\mu}_i \right]$$

$$= \lim_{\lambda \to -1} \frac{2}{\lambda(\lambda+1)} \left[\Sigma_i \left(\hat{\mu}_i \left(\left(\frac{m_i}{\hat{\mu}_i} \right)^{\lambda+1} - 1 \right) \right) \right]$$

$$= \left(\lim_{\lambda \to -1} \frac{2}{\lambda} \right) \left\{ \Sigma_i \hat{\mu}_i \left[\lim_{\lambda \to -1} \left(\frac{\left(\frac{m_i}{\hat{\mu}_i}\right)^{\lambda+1} - 1}{\lambda+1} \right) \right] \right\}$$

$$= (-2) \left\{ \Sigma_i \hat{\mu}_i \left[\log \left(\frac{m_i}{\hat{\mu}_i} \right) \right] \right\}$$

$$= 2 \Sigma_i \hat{\mu}_i \log \left(\frac{\hat{\mu}_i}{m_i} \right)$$

(d) $$\Sigma_i \frac{(m_i - \hat{\mu}_i)^2}{m_i}$$

$$= \Sigma_i \left(\frac{m_i^2 - 2m_i\hat{\mu}_i + \hat{\mu}_i^2}{m_i} \right)$$

$$= \Sigma_i \left(\frac{\hat{\mu}_i^2}{m_i} - 2\hat{\mu}_i + m_i \right)$$

$$= \Sigma_i \left(\frac{\hat{\mu}_i^2}{m_i} \right) + \Sigma_i (m_i - 2\hat{\mu}_i)$$

$$= \Sigma_i \left(\frac{\hat{\mu}_i^2}{m_i} \right) + (-m)$$

$$= \Sigma_i \left(\frac{\hat{\mu}_i^2}{m_i} \right) + (-\Sigma_i m_i)$$

$$= \Sigma_i m_i \left(\left(\frac{\hat{\mu}_i}{m_i} \right)^2 - 1 \right)$$

$$= \frac{2}{(-2)((-2)+1)} \Sigma_i m_i \left(\left(\frac{m_i}{\hat{\mu}_i} \right)^{-2} - 1 \right)$$

(e) $$4 \Sigma_i (\sqrt{m_i} - \sqrt{\hat{\mu}_i})^2$$

$$= 4 \Sigma_i (m_i + \hat{\mu}_i - 2\sqrt{m_i\hat{\mu}_i})$$

$$= 4 \left[(-2 \Sigma_i \sqrt{m_i\hat{\mu}_i}) + (\Sigma_i (m_i + \hat{\mu}_i)) \right]$$

$$= 4 \left[(-2 \Sigma_i \sqrt{m_i\hat{\mu}_i}) + (2m) \right]$$

$$= 4 \left[(-2 \Sigma_i \sqrt{m_i\hat{\mu}_i}) + (2 \Sigma_i m_i) \right]$$

$$= -8\left(\Sigma_i\left(\sqrt{m_i\hat{\mu}_i} - m_i\right)\right)$$

$$= \frac{2}{\left(-\frac{1}{2}\right)\left(\left(-\frac{1}{2}\right)+1\right)} \Sigma_i\, m_i \left(\left(\frac{m_i}{\hat{\mu}_i}\right)^{\left(-\frac{1}{2}\right)} - 1\right)$$

3.39 $$\hat{u} = -\frac{\Sigma_i \Sigma_j\, p_{ij} \log\left(p_{ij}/(p_{i+}\, p_{+j})\right)}{\Sigma_j\, p_{+j} \log p_{+j}}$$

$$= -\frac{\Sigma_i \Sigma_j\, (m_{ij}/m) \log\left(\frac{m_{ij}/m}{(m_{i+}/m)\,(m_{+j}/m)}\right)}{\Sigma_j\, p_{+j} \log p_{+j}}$$

$$= -\frac{2\,\Sigma_i \Sigma_j\, m_{ij} \log\left(\frac{m_{ij}}{(m_{i+}\, m_{+j})/m}\right)}{2m\, \Sigma_j\, p_{+j} \log p_{+j}}$$

$$= -\frac{\left(2\,\Sigma_i \Sigma_j\, m_{ij} \log\left(\frac{m_{ij}}{\hat{\mu}_{ij}}\right)\right)}{2m\, \Sigma_j\, p_{+j} \log p_{+j}}$$

$$= -\frac{G^2}{2m\, \Sigma_j\, p_{+j} \log p_{+j}}$$