Homework 2

蒋翌坤 20307100013

3.6

We can construct the following contingency table with expected frequency under independence in the bracket:

	Living arrangement			
Stage of breast cancer	Living alone	Living with spouse	Living with others	Total
Local	85	100	36	221
	(72)	(104.5)	(44.5)	
Advanced	59	109	53	221
	(72)	(104.5)	(44.5)	
Total	144	209	89	442

We can then calculate the Pearson's chi-squared statistic and the likelihood-ratio chi-squared statistic: $X^2 = 8.33, G^2 = 8.38$ with $df = 2$. Their p-value is 0.016 and 0.015 respectively.

3.18

a

We can construct the following contingency table with expected frequency under independence in the bracket:

	Drug			
Nervousness	Loratadine	Placebo	ChorophenIramine	Total
Yes	4	2	2	8
	(2.43)	(3.38)	(2.19)	
No	184	260	168	612
	(185.57)	(258.62)	(167.81)	
Total	188	262	170	620

We can calculate that $X^2 = 1.62$ with $df = 2$. P-value is 0.44. Therefore, there is no inferential evidence that nervousness depends on drug.

b

(i) $\hat{\theta} = \frac{4\times 260}{2\times 184} = 2.83$, $\hat{\sigma}(\log\hat{\theta}) = \left(\frac{1}{4} + \frac{1}{2} + \frac{1}{260} + \frac{1}{184}\right)^{\frac{1}{2}} = 0.87$. Therefore, under $\alpha = 0.05$, the confidence interval for odds ratio is $[\hat{\theta}\exp(\pm 1.96\hat{\sigma}(\log\hat{\theta}))] = [0.512, 15.592]$

(ii) $\hat{\pi}_1 - \hat{\pi}_2 = \frac{4}{184} - \frac{2}{260} = 0.014$, $\hat{\sigma}(\hat{\pi}_1 - \hat{\pi}_2) = \left(\frac{\frac{4}{184}(1-\frac{4}{184})}{188} + \frac{\frac{2}{260}(1-\frac{2}{260})}{262}\right)^{\frac{1}{2}} = 0.012$. Therefore, under $\alpha = 0.05$, the confidence interval for the difference of proportions suffering nervousness is $[\hat{\pi}_1 - \hat{\pi}_2 \pm 1.96\hat{\sigma}(\hat{\pi}_1 - \hat{\pi}_2)] = [-0.0093, 0.0374]$

3.26

We know that $\sqrt{n}(g(\hat{\boldsymbol{\pi}}) - g(\boldsymbol{\pi})) \xrightarrow{d} N(0, \sigma^2)$ and $\mathbb{E}(\hat{\pi}_i) = \pi_i, \mathrm{Var}(\hat{\pi}_i) = \pi_i(1-\pi_i)/n$, $\mathrm{Cov}(\hat{\pi}_i) = -\pi_i\pi_j/n$. Therefore,

$$g(\hat{\boldsymbol{\pi}}) - g(\boldsymbol{\pi}) \approx \sum_i \frac{\partial g(\boldsymbol{\pi})}{\partial \pi_i}(\hat{\pi}_i - \pi_i) = \sum_i \frac{\partial \nu/\delta}{\partial \pi_i}(\hat{\pi}_i - \pi_i) = \sum_i \frac{\frac{\partial \nu}{\partial \pi_i}\delta - \nu\frac{\partial \delta}{\partial \pi_i}}{\delta^2}(\hat{\pi}_i - \pi_i) = \sum_i \frac{\eta_i}{\delta^2}(\hat{\pi}_i - \pi_i)$$

$$\sigma^2 = \mathrm{Var}(\sqrt{n}(g(\hat{\boldsymbol{\pi}}) - g(\boldsymbol{\pi}))) \approx n\Big(\sum_i \frac{\eta_i^2}{\delta^4}\frac{\pi_i(1-\pi_i)}{n} - \sum_{i\neq j}\frac{\eta_i\eta_j}{\delta^4}\frac{\pi_i\pi_j}{n}\Big) = \frac{\sum_i \eta_i^2\pi_i - \Big(\sum_i \eta_i\pi_i\Big)^2}{\delta^4}$$

3.30

$$z^2 = \frac{(\frac{y_1}{n_1} - \frac{y_2}{n_2})^2}{\frac{y_1+y_2}{n_1+n_2}\frac{n_1+n_2-y_1-y_2}{n_1+n_2}\frac{n_1+n_2}{n_1n_2}} = \frac{(n_1+n_2)(y_1n_2 - y_2n_1)^2}{(y_1+y_2)(n_1+n_2-y_1-y_2)(n_1n_2)}$$

For a 2×2 contingency table, consider $n_{11} = y_1, n_{12} = n_1 - y_1, n_{21} = y_2, n_{21} = n_2 - y_2$. We have $\mu_{11} = n_1\hat{\pi}, \mu_{12} = n_1(1-\hat{\pi}), \mu_{21} = n_2\hat{\pi}, \mu_{22} = n_2(1-\hat{\pi})$

$$\begin{aligned}
X^2 &= \sum \frac{(n_{ij} - \mu_{ij})^2}{\mu_{ij}} \\
&= \frac{(y_1 - n_1\hat{\pi})^2}{n_1\hat{\pi}} + \frac{(n_1 - y_1 - n_1(1-\hat{\pi}))^2}{n_1(1-\hat{\pi})} + \frac{(y_2 - n_2\hat{\pi})^2}{n_2\hat{\pi}} + \frac{(n_2 - y_2 - n_2(1-\hat{\pi}))^2}{n_2(1-\hat{\pi})} \\
&= \frac{y_1^2 + n_1^2\hat{\pi}^2 + 4n_1y_1\hat{\pi}^2 - 2n_1y_1\hat{\pi}}{n_1\hat{\pi}(1-\hat{\pi})} + \frac{y_2^2 + n_2^2\hat{\pi}^2 + 4n_2y_2\hat{\pi}^2 - 2n_2y_2\hat{\pi}}{n_2\hat{\pi}(1-\hat{\pi})} \\
&= \frac{\frac{y_1^2}{n_1} + \frac{y_2^2}{n_2} + (n_1+n_2)\hat{\pi} + 4(y_1+y_2)\hat{\pi} - 2(y_1+y_2)}{\hat{\pi}(1-\hat{\pi})} \\
&= \frac{(n_1+n_2)^2(n_2y_1^2 + n_1y_2^2 + n_1n_2(y_1+y_2) + 4\frac{n_1n_2(y_1+y_2)^2}{n_1+n_2} - 2n_1n_2(y_1+y_2))}{(y_1+y_2)(n_1+n_2-y_1-y_2)(n_1n_2)} \\
&= \frac{(n_1+n_2)(y_1n_2 - y_2n_1)^2}{(y_1+y_2)(n_1+n_2-y_1-y_2)(n_1n_2)} \\
&= z^2
\end{aligned}$$