Homework 1

蒋翌坤 20307100013

1.6

a

$$G^2 = -2\log\Lambda = 2\Big(y\log\frac{y}{n\pi_0} + (n-y)\log\frac{n-y}{n-n\pi_0}\Big)$$
$$= 2\Big(25\log\frac{25}{25-25\times 0.5}\Big) = 2\Big[25\log(\frac{25}{12.5})\Big] = 34.7$$

b

$$\chi_s^2 = z_s^2 = \frac{(y-n\pi_0)^2}{n\pi_0(1-\pi_0)} = \frac{(0-25\times 0.5)^2}{25\times 0.5\times 0.5} = 25$$

c

$\hat{\pi} = 0 \Rightarrow z_w = \dfrac{\hat{\pi}-\pi_0}{\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}} = \dfrac{-0.5}{0} \Rightarrow$ The wald or chi-squared statistic is infinite.

1.8

The experiment has a binomial parameter π. A successful observation is when the seedlings are green. The hypothesis is therefore $H_0 : \pi = 0.75$. The observed value is $y = 854, n = 1103$.

We use the wald statistic to test the hypothesis. $z_w = \dfrac{\hat{\pi}-\pi_0}{\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}} = \dfrac{854/1103 - 0.75}{\sqrt{\frac{854/1103\times(1-854/1103)}{1103}}} = 1.93$

P-value is 0.054. It is larger than 0.05, which means that under 5% confidence level, we do not reject the null hypothesis that the true ratio of green to yellow is 3:1.

1.30

a

$$L(\theta) = \sum_{j=1}^{3} n_j \log p_j = n_1\log\theta^2 + n_2\log(2\theta(1-\theta)) + n_3\log(1-\theta)^2$$

$$\frac{\partial L(\theta)}{\partial\theta} = \frac{2n_1}{\theta} + \frac{n_2}{\theta} - \frac{n_2}{1-\theta} - \frac{2n_3}{1-\theta} = 0 \Rightarrow \hat{\theta} = \frac{2n_1+n_2}{2n_1+2n_2+2n_3}$$

b

$$-\frac{\partial^2 L(\theta)}{\partial \theta^2} = -\frac{\partial(\frac{2n_1}{\theta} + \frac{n_2}{\theta} - \frac{n_2}{1-\theta} - \frac{2n_3}{1-\theta})}{\partial \theta} = \frac{2n_1}{\theta^2} + \frac{n_2}{\theta^2} + \frac{n_2}{(1-\theta)^2} + \frac{2n_3}{(1-\theta)^2} = \frac{2n_1+n_2}{\theta^2} + \frac{n_2+2n_3}{(1-\theta)^2}$$

$$\mathbb{E}(-\frac{\partial^2 L(\theta)}{\partial \theta^2}) = \frac{2\theta^2 n + 2\theta(1-\theta)n}{\theta^2} + \frac{2\theta(1-\theta)n + 2(1-\theta)^2 n}{(1-\theta)^2} = \frac{2n}{\theta} + \frac{2n}{1-\theta} = \frac{2n}{\theta(1-\theta)}$$

$$\mathrm{SE}(\hat{\theta}) = \frac{1}{\sqrt{\mathbb{E}(-\frac{\partial^2 L(\hat{\theta})}{\partial \hat{\theta}^2})}} = \sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{2n}}$$

c

The observed counts expects to be $\hat{\theta}^2 n, 2\hat{\theta}(1-\hat{\theta})n, (1-\hat{\theta})^2 n$. Then we can use log-likelihood ratio test to test whether the probabilities truly have this pattern. Specifically, we test the hypothesis $H_0 : \pi_1 = \hat{\theta}^2, \pi_2 = 2\hat{\theta}(1-\hat{\theta}), \pi_3 = (1-\hat{\theta})^2$ with $\mathrm{df} = 3 - 1 - 1 = 1$.

2.8

a

The correct interpretation is that the probability female survive and male do not survive was 11.4 times higher than the probability that female do not survive and male survive.

If the female and male percentage of those who did not survive are the same, then the quoted interpretation is approximately correct.

b

$$\pi_{11} = \frac{2.9}{2.9+1} \approx 74.36\% \qquad \frac{\pi_{11}(1-\pi_{21})}{(1-\pi_{11})\pi_{21}} = 11.4 \Rightarrow \pi_{21} = \frac{29}{143} \approx 20.28\%$$

Therefore, for female, the proportion who survived is 74.36%, for male, the proportion who survived is 20.28%.

2.12

$\theta_{AG(A)} = 0.35 \quad \theta_{AG(B)} = 0.80 \quad \theta_{AG(C)} = 1.13 \quad \theta_{AG(D)} = 0.92 \quad \theta_{AG(E)} = 1.22 \quad \theta_{AG(F)} = 0.83$

Marginal odds ratio: $\theta_{AG} = 1.84$

For Department A, the conditional odds ratio is 0.35, which means that female have a higher chance of admission. For the rest of the Department, the conditional odds ratio is close to 1, which means that the admission chance for male and female are similar. The marginal odds ratio is 1.84, which indicates male have a higher chance of admission overall. The reason for the different indications is that male applies more to Department A and B, which have a higher chance of admission, while female applies more to Department C, D, E, F, which have a lower chance of admission.

2.18

a

$$\Omega_0 = \frac{7}{61} \quad \Omega_1 = \frac{55}{129} \quad \Omega_2 = \frac{489}{570} \quad \Omega_3 = \frac{475}{431} \quad \Omega_4 = \frac{293}{154} \quad \Omega_5 = \frac{38}{12}$$

$$\theta_{01} = 0.27 \quad \theta_{02} = 0.13 \quad \theta_{03} = 0.10 \quad \theta_{04} = 0.06 \quad \theta_{05} = 0.04$$

As smoking increases, the odds of lung cancer increases, the odds ratios that pair the level of smoking with no smoking decreases. This means that the more you smoke, the more likely you will get lung cancer.

b

$$\theta_{ij} = \frac{\pi_{ij}\pi_{i+1,j+1}}{\pi_{i,j+1}\pi_{i+1,j}} = \frac{\Omega_i}{\Omega_{i+1}} = \exp\{\log\Omega_i - \log\Omega_{i+1}\} = \exp\{(\alpha+\beta i) - (\alpha+\beta(i+1))\} = \exp\{-\beta\}$$

Therefore, the local odds ratios are identical. It is a constant equal to $\exp\{-\beta\}$.

c

Yes. We can just take the observed probability of lung cancer at each level of smoking to estimate.

The estimated odds ratios in part (a) are meaningful. If we treat the odds ratios as the true odds ratios, then we can use the data of those who do not smoke to estimate the probability of lung cancer at each level of smoking.

d

Let $F_{j|i} = \pi_{0|i} + \cdots + \pi_{j|i}, j = 0,1,2,3,4,5, i = 1,2$, we have

$$\pi_{0|1} = 0.005 \quad \pi_{1|1} = 0.041 \quad \pi_{2|1} = 0.360 \quad \pi_{3|1} = 0.350 \quad \pi_{4|1} = 0.216 \quad \pi_{5|1} = 0.028$$
$$\pi_{0|2} = 0.045 \quad \pi_{1|2} = 0.095 \quad \pi_{2|2} = 0.042 \quad \pi_{3|2} = 0.318 \quad \pi_{4|2} = 0.113 \quad \pi_{5|2} = 0.009$$

$$F_{0|1} = 0.005 \quad F_{1|1} = 0.046 \quad F_{2|1} = 0.406 \quad F_{3|1} = 0.756 \quad F_{4|1} = 0.972 \quad F_{5|1} = 1.000$$
$$F_{0|2} = 0.045 \quad F_{1|2} = 0.140 \quad F_{2|2} = 0.560 \quad F_{3|2} = 0.878 \quad F_{4|2} = 0.991 \quad F_{5|2} = 1.000$$

So, $F_{j|1} \leq F_{j|2}$ for all j. Therefore, the disease groups are stochastically ordered with respect to their distributions on smoking level. To be specific, the distribution of control patients are stochastically smaller than the distribution of patients with lung cancer. This means the probability of a control patient being at a higher level of smoking is smaller than the probability of a patient with lung cancer being at a higher level of smoking. This implies the popular belief that smoking increases the chance of lung cancer.