Homework 3

蒋翌坤 20307100013

4.6

a

$\log \mu_A = \alpha$ and $\log \mu_B = \alpha + \beta$, so $\exp(\beta) = \frac{\mu_B}{\mu_A}$. Let $\hat{\alpha}$ and $\hat{\beta}$ be the estimate, then the interpretation can be that $\hat{\mu}_A = \exp(\hat{\alpha})$ and $\hat{\mu}_B = \exp(\hat{\alpha} + \hat{\beta})$

b

Let x_i and y_i be the samples from A and B, we have

$$L(\alpha, \beta) = \sum_{i=1}^{10} (x_i \alpha - e^{\alpha} - \log x_i!) + \sum_{i=1}^{10} (y_i(\alpha + \beta) - e^{\alpha+\beta} - \log y_i!)$$

The MLE of α and β are $\hat{\alpha} = \log \bar{x} = \log 5$ and $\hat{\beta} = \log \bar{y} - \log \bar{x} = \log 9 - \log 5$.

$\mathbb{E}\left(-\frac{\partial^2 L}{\partial \alpha^2}\right) = 10e^{\alpha} + 10e^{\alpha+\beta} = 10 \times 5 + 10 \times 9 = 140$;

$\mathbb{E}\left(-\frac{\partial^2 L}{\partial \alpha \partial \beta}\right) = \mathbb{E}\Big((\sum x_i + \sum y_i - 10e^{\alpha} - 10e^{\alpha+\beta})(\sum y_i - 10e^{\alpha+\beta})\Big) = \mathbb{E}\Big(\sum x_i \sum y_i + (\sum y_i)^2 - 10e^{\alpha} \sum y_i - 10e^{\alpha+\beta} \sum y_i\Big) = 50 \times 90 + 90^2 + 90 - 10 \times 5 \times 90 - 10 \times 9 \times 90 = 90$;

$\mathbb{E}\left(-\frac{\partial^2 L}{\partial \beta^2}\right) = 10e^{\alpha+\beta} = 90$

Information matrix $\mathcal{J} = \begin{pmatrix} \mathbb{E}\left(-\frac{\partial^2 L}{\partial \alpha^2}\right) & \mathbb{E}\left(-\frac{\partial^2 L}{\partial \alpha \partial \beta}\right) \\ \mathbb{E}\left(-\frac{\partial^2 L}{\partial \alpha \partial \beta}\right) & \mathbb{E}\left(-\frac{\partial^2 L}{\partial \beta^2}\right) \end{pmatrix} = \begin{pmatrix} 140 & 90 \\ 90 & 90 \end{pmatrix}$; $\mathcal{J}^{-1} = \begin{pmatrix} \frac{1}{50} & -\frac{1}{50} \\ -\frac{1}{50} & \frac{7}{225} \end{pmatrix}$, so $\widehat{\mathrm{SE}}(\hat{\beta}) = \sqrt{\frac{7}{225}}$

Wald statistic: $z_w = \frac{\hat{\beta}}{\widehat{\mathrm{SE}}(\hat{\beta})} = \frac{\log 9 - \log 5}{\sqrt{\frac{7}{225}}} \approx 18.89$. p-value < 0.001. So we reject the null hypothesis that $\mu_A = \mu_B$ and $\beta = 0$. That is there is a significant difference between the two treatments.

c

Using $\hat{\beta}$ and $\widehat{\mathrm{SE}}(\hat{\beta})$, we can construct a 95% confidence interval for β: $\hat{\beta} \pm 1.96\widehat{\mathrm{SE}}(\hat{\beta}) = (\log 9 - \log 5) \pm 1.96\sqrt{\frac{7}{225}} = (0.24, 0.93)$. So the 95% confidence interval for $\frac{\mu_B}{\mu_A} = \exp(\beta)$ is $(e^{0.24}, e^{0.93}) = (1.27, 2.54)$

d

We know $\sum x_i \sim \mathrm{Poi}(10\mu_A)$ and $\sum y_i \sim \mathrm{Poi}(10\mu_B)$, $\sum x_i$ and $\sum y_i$ are independent, so $\sum x_i | \sum x_i + \sum y_i \sim \mathrm{Bin}(\sum x_i + \sum y_i, \frac{\mu_A}{\mu_A + \mu_B})$

Under null hypothesis that $\mu_A = \mu_B$, we should see $\pi = \frac{\mu_A}{\mu_A + \mu_B} = \frac{1}{2}$.

$\hat{\pi} = \frac{\sum x_i}{\sum x_i + \sum y_i} = \frac{5}{14}$, $\widehat{\mathrm{SE}}(\hat{\pi}) = \sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{\sum x_i + \sum y_i}} = \sqrt{\frac{\frac{5}{14}(1 - \frac{5}{14})}{140}} \approx 0.04$. So the wald statistic is $\frac{\hat{\pi} - \pi}{\widehat{\mathrm{SE}}(\pi)} = \frac{\frac{5}{14} - \frac{1}{2}}{0.04} \approx -3.53$. p-value < 0.001. So we reject the null hypothesis that $\mu_A = \mu_B$.

4.13

a

Total number made is 145 and total attempts is 296. So $\hat{\alpha} = \frac{145}{296} \approx 0.49$ and $\widehat{\mathrm{SE}}(\hat{\alpha}) = \sqrt{\frac{\hat{\alpha}(1-\hat{\alpha})}{296}} \approx 0.029$. The interpretation is that the estimated percentage that O'Neal's free throw is made is 49% and the standard error is 2.9%.

Let x_i be the number made and n_i be the number of attempts in each game. $X^2 = \sum_{i=1}^{23} \frac{(x_i - n_i\hat{\alpha})^2}{n_i\hat{\alpha}} + \frac{(n_i - x_i - n_i(1-\hat{\alpha}))^2}{n_i(1-\hat{\alpha})} = 35.5$ with $df = 22$. p-value is 0.03. So the model does not appear to fit adequately.

b

$\sqrt{\frac{X^2}{22}} \approx 1.27$, so we adjust SE by a factor of 1.27. The adjusted SE is 0.037. The original confidence inter of α is $0.49 \pm 1.96 \times 0.029 = (0.40, 0.51)$. The adjusted confidence interval is $0.49 \pm 1.96 \times 0.037 = (0.38, 0.53)$. We can see that the adjusted confidence interval is wider, reflecting overdispersion.

4.29

a

Since the PDF is symmetric around 0, we have $\Phi(0) = 0.5 \Rightarrow \alpha + \beta x = 0 \Rightarrow x = -\alpha/\beta$

b

$$\frac{d\pi(x)}{dx} = \frac{d}{dx}\Phi(\alpha + \beta x) = \phi(\alpha + \beta x)\beta, \text{ when } \pi(x) = 0.5, x = -\alpha/\beta, \frac{d\pi(x)}{dx} = \beta\phi(0)$$

For logit link, $\phi(x) = \frac{e^{-x}}{(1+e^{-x})^2}$, $\beta\phi(0) = 0.25\beta$; For probit link, $\phi(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$, $\beta\phi(0) = \frac{\beta}{\sqrt{2\pi}}$

c

$\pi(x)$ is from the class of normal CDFs, suppose its mean and variance is μ and σ^2, we have $\pi(\mu) = 0.5 \Rightarrow \mu = -\alpha/\beta$

$\frac{d}{dx}\pi(x) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, when $x = \mu$, $\frac{d}{dx}\pi(x) = \frac{1}{\sqrt{2\pi\sigma^2}} = \frac{\beta}{\sqrt{2\pi}} \Rightarrow \sigma = \frac{1}{|\beta|}$