Solution for CDA2015

- **Solution to 3.6.**

Living arrangement	Stage of breast cancer Local	Advanced	Total
Alone	59	85	144
	(72)	(72)	32.58%
	(-2.638731)	(2.6387308)	
With spouse	109	100	209
	(104.5)	(104.5)	47.29%
	(0.8574376)	(-0.857438)	
With others	53	36	89
	(44.5)	(44.5)	20.14%
	(2.0164045)	(-2.016405)	
Total	221	221	442
	50%	50%	

From SAS output, $X^2 \approx 8.3292$, which is approximated by χ^2_{df} distribution with $df = 2$, and p-value $=0.0155$; Also $G^2 \approx 8.3752$, approximated by χ^2_{df} distribution with $df = 2$, and p-value$=0.0152$.

- **Solution to 3.18.**

Drug	Nervousness Yes	No	Total
Loratadine	4	184	188
	(2.4258)	(185.57)	30.57%
	(1.2186962)	(-1.218715)	
Placebo	2	260	262
	(3.3806)	(258.62)	42.26%
	(-0.994638)	(0.9946141)	
Chlorpheniramine	2	168	170
	(2.1935)	(167.81)	27.42%
	(-0.154409)	(0.1543916)	
Total	8	612	620
	1.29%	98.71%	

(a). No. $X^2 = 1.6234 \sim \chi^2(2)$ and $p = 0.4441$; $G^2 = 1.5528 \sim \chi^2(2)$ and $p = 0.4600$.

Alternative method. Since the number of observations in some cells are smaller than 5, we can also use Fisher exact test. By SAS output, the p-value for Fisher exact test is about 0.395, so that we can not reject the independence assumption, i.e. no inferential evidence that nervousness depends on drug.

Note: both methods are correct!

(b). (i) $\hat{\theta} = \frac{n_{11}n_{22}}{n_{12}n_{21}} \approx 2.826$, $\hat{\sigma}(\log\hat{\theta}) = \left(\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}\right)^{1/2} \approx 0.871$. For $\alpha = 0.05$, CI_α for $\log\theta$ is:

$$[\log(\hat{\theta}) - z_{\alpha/2}\hat{\sigma}(\log(\hat{\theta})), \log(\hat{\theta}) + z_{\alpha/2}\hat{\sigma}(\log(\hat{\theta}))] = [-0.668, 2.745]$$

and hence CI_α for θ is:

$$[e^{-0.668}, e^{2.745}] = [0.513, 15.565].$$

(ii) $\hat{\pi}_1 = 4/188$, $\hat{\pi}_2 = 2/262$, $\hat{\pi}_1 - \hat{\pi}_2 \approx 0.01364$, $\hat{\sigma}(\hat{\pi}_1 - \hat{\pi}_2) = \left(\frac{\hat{\pi}_1(1-\hat{\pi}_1)}{n_1} + \frac{\hat{\pi}_2(1-\hat{\pi}_2)}{n_2}\right)^{1/2} \approx 0.0118$. So CI_α for $\pi_1 - \pi_2$ is: $(-0.0094, 0.0368)$.

- **Solution to 3.26.** Recall that, as $n \to \infty$, $\sqrt{n}(\hat{\pi}_i - \pi_i) \xrightarrow{d} N(0, \pi_i(1-\pi_i))$ and $Cov(\hat{\pi}_i, \hat{\pi}_j) = -\pi_i\pi_j/n$ for $i \neq j$. Then by delta method,

$$\begin{aligned}
\sqrt{n}(g(\hat{\pi}) - g(\pi)) &= \sum_i \frac{\partial g(\pi)}{\partial \pi_i}\sqrt{n}(\hat{\pi}_i - \pi_i) + o_P(1) \\
&= \sum_i -\frac{\eta_i}{\delta^2}\sqrt{n}(\hat{\pi}_i - \pi_i) + o_P(1) \\
&\xrightarrow{d} N(0,\ \sigma^2),
\end{aligned}$$

where

$$\sigma^2 = \sum_i \frac{\eta_i^2}{\delta^4}\pi_i(1-\pi_i) - 2\sum\sum_{i<j}\frac{\eta_i\eta_j}{\delta^4}\pi_i\pi_j = \Big[\sum_i \pi_i\eta_i^2 - \big(\sum_i \pi_i\eta_i\big)^2\Big]/\delta^4.$$

□

- **Solution to 3.30.** Recall the notation

$$\hat{\pi}_1 = y_1/n_1, \quad \hat{\pi}_2 = y_2/n_2, \quad \hat{\pi} = (y_1+y_2)/(n_1+n_2), \quad n = n_1 + n_2.$$

Then

$$z^2 = \frac{(y_1/n_1 - y_2/n_2)^2}{(\frac{y_1+y_2}{n_1+n_2})(1 - \frac{y_1+y_2}{n_1+n_2})(\frac{1}{n_1} + \frac{1}{n_2})} = \frac{(n_2y_1 - n_1y_2)^2(n_1+n_2)}{n_1n_2(y_1+y_2)(n_1+n_2-y_1-y_2)}.$$

For the table

y_1	$n_1 - y_1$
y_2	$n_2 - y_2$

,

$$X^2 = \frac{(y_1 - n_1\hat{\pi})^2}{n_1\hat{\pi}} + \frac{(n_1 - y_1 - n_1(1-\hat{\pi}))^2}{n_1(1-\hat{\pi})} + \frac{(y_2 - n_2\hat{\pi})^2}{n_2\hat{\pi}} + \frac{(n_2 - y_2 - n_2(1-\hat{\pi}))^2}{n_2(1-\hat{\pi})}.$$

Simple calculation implies that $X^2 = z^2$. □

- **Solution to 4.6.** (a). Note that $\log \mu_A = \alpha$ and $\log \mu_B = \alpha + \beta$. Then $\exp(\beta) = \mu_B/\mu_A$. If $\hat{\alpha}$ and $\hat{\beta}$ are the MLE of α and β, respectively, then $\hat{\mu}_A = \exp(\hat{\alpha})$ and $\hat{\mu}_B = \exp(\hat{\alpha} + \hat{\beta})$ are the MLE of μ_A and μ_B, respectively.

  (b). $L(\alpha, \beta) = n\bar{x}\alpha - ne^{\alpha} + m\bar{y}(\alpha + \beta) - me^{\alpha+\beta}$, where $\bar{x}$ and $\bar{y}$ are the means of the data for A and B, respectively, and their sample sizes are $n = 10$ and $m = 10$.

  So the MLE are $\hat{\alpha} = \log \bar{x} = \log 5 = 1.609438$, $\hat{\beta} = \log \bar{y} - \log \bar{x} = \log 9 - \log 5 = 0.5877867$. The information matrix is

$$\mathcal{J} = \begin{pmatrix} E\big(-\frac{\partial^2 L}{\partial \alpha^2}\big) & E\big(-\frac{\partial^2 L}{\partial \alpha \partial \beta}\big) \\ E\big(-\frac{\partial^2 L}{\partial \alpha \partial \beta}\big) & E\big(-\frac{\partial^2 L}{\partial \beta^2}\big) \end{pmatrix} = \begin{pmatrix} ne^{\alpha} + me^{\alpha+\beta} & me^{\alpha+\beta} \\ me^{\alpha+\beta} & me^{\alpha+\beta} \end{pmatrix} = \begin{pmatrix} 140 & 90 \\ 90 & 90 \end{pmatrix}.$$

  So, the covariance matrix of $(\hat{\alpha}, \hat{\beta})^T$ is

$$(\mathcal{J})^{-1} = \begin{pmatrix} 140 & 90 \\ 90 & 90 \end{pmatrix}^{-1} = \begin{pmatrix} 1/50 & -1/50 \\ -1/50 & 7/225 \end{pmatrix}.$$

  Hence, $\widehat{\mathrm{SE}}(\hat{\beta}) = \sqrt{7/225} \approx 0.1764$.

  Wald statistic $z_W = \hat{\beta}/\widehat{\mathrm{SE}}(\hat{\beta}) \approx 3.3326$ (or $Z_W^2 = 11.106$) and $p-$ value $\approx 8.6 \times 10^{-4}$.

  LR: when $\beta = 0$, $\hat{\alpha}_0 = \log \frac{1}{m+n}(n\bar{x} + m\bar{y}) = \log 14 - \log 2 \approx 1.946$. So $-2\log\Lambda = -2(L(\hat{\alpha}_0, 0) - L(\hat{\alpha}, \hat{\beta})) \approx 11.5894$. With $df = 1$, we get $p-$ value $\approx 6.6 \times 10^{-4}$.

  Remark on (b). Some students do as the following. Let $\hat{\mu} = \hat{\mu}_A - \hat{\mu}_B$, and test $H_0 : \mu = 0$ by using $\sigma^2(\hat{\mu}) = \sqrt{s_A^2/n + s_B^2/m}$ and $Z = (\hat{\mu} - 0)/\sigma(\hat{\mu}) \approx N(0, 1)$. The idea is smart but it is wrong for this case. Note that the populations of A and B are not normal distributed and the sample size $n = m = 10$ are small. So $Z = (\hat{\mu} - 0)/\sigma(\hat{\mu}) \approx N(0, 1)$ does not follows for small sample size.

  (c). Wald CI for β with confidence level 95%: $\hat{\beta} \pm 1.96\widehat{\mathrm{SE}}(\hat{\beta}) = (0.2420, 0.9335)$.
  By δ-method, Wald CI for μ_B/μ_A with confidence level 95%: $(\exp(0.2420), \exp(0.9335)) = (1.2738, 2.5434)$.

  (d). $n\bar{x} \sim \mathrm{Poi}(n\mu_A)$, $m\bar{y} \sim \mathrm{Poi}(m\mu_B)$, and $n\bar{x}$ and $m\bar{y}$ are independent. So, $n\bar{x}|(n\bar{x} + m\bar{y}) \sim \mathrm{Bin}(n\bar{x} + m\bar{y}, \pi = \frac{n\mu_A}{n\mu_A + m\mu_B})$.

  Then $L(\pi) = n\bar{x}\log\pi + m\bar{y}\log(1 - \pi) + \text{constant}$. So the MLE for π is $\hat{\pi} = \frac{n\bar{x}}{n\bar{x}+m\bar{y}} = 50/140 \approx 0.3571$, with $\widehat{SE}(\hat{\pi}) = \sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n\bar{x}+m\bar{y}}} \approx 0.0405$. So Wald statistic $z_W = \frac{\hat{\pi} - n/(n+m)}{\widehat{\mathrm{SE}}(\hat{\pi})} \approx -3.5277$, and $p-$ value $\approx 4.2 \times 10^{-4}$.

  LR: $LR = -2(L(n/(n + m)) - L(\hat{\pi})) \approx 11.5894$ and $p-$ value $= 6.6 \times 10^{-4}$.

  Remark on (d). Some students give $\widehat{SE}(\hat{\pi}) = \sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}$ which is wrong. Some students give $p = 2\Phi(-3.5277) = 0.000792$, which is wrong.

- **Solution to 4.13.** (a). The total number of made is 135, and the total number of attempts is 296. By maximizing the likelihood, the MLE of α is $\hat{\alpha} = 135/296 = 0.456$

and $\hat{SE}(\hat{\alpha}) = \sqrt{\hat{\alpha}(1-\hat{\alpha})/n} = 0.029$. O'Neal's estimated probability of making a free throw is 0.456, and a 95% confidence interval is (0.40, 0.51). Pearson test statistic

$$X^2 = \sum_{i=1}^{23} \Big(\frac{(y_i - n_i\hat{\alpha})^2}{n_i\hat{\alpha}} + \frac{(n_i - y_i - n_i(1-\hat{\alpha}))^2}{n_i(1-\hat{\alpha})} \Big) = 35.5, \quad \text{with } df = 22$$

provides evidence of lack of fit (since $p = 0.034$).

(b). Using quasi likelihood, $\sqrt{X^2/df} = 1.27$, so adjusted $\hat{SE}$ is 0.037 and the adjusted confidence interval for α is $(0.456 - 1.96 \times 0.037, 0.456 + 1.96 \times 0.037) = (0.38, 0.53)$, reflecting slight overdispersion.

Remark on (a). Alternative way to calculate X^2 for testing independence, i.e.

$$\tilde{X}^2 = \sum_{i=1}^{23}\sum_{j=1}^{2} \frac{(\tilde{n}_{ij} - \tilde{\mu}_{ij})^2}{\tilde{\mu}_{ij}}, \quad \text{with } \tilde{\mu}_{ij} = \tilde{n}_{i+}\tilde{n}_{+j}/n.$$

Note that in two notation settings, $\tilde{n}_{i+} = n_i$, $\tilde{n}_{i1} = y_i$, $\tilde{n}_{i2} = n_i - y_i$, $\tilde{n}_{+1}/n = \hat{\alpha}$ and $n_{+2}/n = (1-\hat{\alpha})$. So, it is easy to see that $X^2 = \tilde{X}^2$.

Remark on (b). Alternative way to calculate X^2 for overdispersion, i.e.

$$X^2 = \sum_i \frac{(y_i - \hat{\mu})^2}{\nu^*(\hat{\mu})} = \sum_{i=1}^{23} \frac{(y_i - n_i\hat{\alpha})^2}{n_i\hat{\alpha}(1-\hat{\alpha})}.$$

This is the same X^2 in (a) since

$$\frac{(y_i - n_i\hat{\alpha})^2}{n_i\hat{\alpha}} + \frac{(n_i - y_i - n_i(1-\hat{\alpha}))^2}{n_i(1-\hat{\alpha})} = \frac{(y_i - n_i\hat{\alpha})^2}{n_i\hat{\alpha}(1-\hat{\alpha})}.$$

- **Solution to 4.26.** By conditional density function, we have

$$E[P(T \geq t_o|S)] = E\left[\int_{t_o}^{+\infty} f(t|s)\mathrm{d}t\right] = \int_{t_o}^{+\infty} E[f(t|s)]\mathrm{d}t = \int_{t_o}^{+\infty} f(t)\mathrm{d}t = P(\beta).$$

So, $P(T \geq t_o|S)$ is an unbiased estiamtor of $P(\beta)$.

Let $h(S) = P(T \geq t_o|S)$. It is an unbiased estiamtor of $P(\beta)$. Suppose $\phi(X)$ is another unbiased estimator of $P(\beta)$. Then

$$\begin{aligned} Var_\beta[\phi(X)] &= E_\beta[\phi(X) - P(\beta)]^2 \\ &= E_\beta[(\phi(X) - h(S)) + (h(S) - P(\beta))]^2 \\ &= E_\beta[\phi(X) - h(S)]^2 + 2E_\beta[(\phi(X) - h(S))(h(S) - P(\beta))] + E_\beta[h(S) - P(\beta)]^2. \end{aligned}$$

Since S is a sufficient statistic, it follows that by conditional expectation,

$$\begin{aligned} E_\beta[(\phi(X) - h(S))(h(S) - P(\beta))] &= E_\beta[E[(\phi(X) - h(S))(h(S) - P(\beta))|S]] \\ &= E_\beta[(\phi(X) - h(S))E[(h(S) - P(\beta))|S]] \\ &= 0. \end{aligned}$$

So for all $\forall\beta$, it follows that

$$Var_\beta[\phi(X)] = E_\beta[\phi(X) - h(S)]^2 + Var_\beta[h(S))] \geq Var_\beta[h(S))].$$

Hence $h(S) = P(T \geq t_o|S)$ is $P(\beta)$'s UMVUE.

- **Solution to 4.29.** (a). Since ϕ is symmetric, $\Phi(0) = 0.5$. Setting $\alpha + \beta x = 0$ gives $x = -\alpha/\beta$.

  (b). The derivative of Φ at $x = -\alpha/\beta$ is $\beta\phi(\alpha + \beta(-\alpha/\beta)) = \beta\phi(0)$. The logistic pdf has $\phi(x) = e^x/(1 + e^x)^2$ which equals 0.25 at $x = 0$; the standard normal pdf equals $1/\sqrt{2\pi}$ at $x = 0$.

  (c). Note that

  $$\pi(x) = \Phi(\alpha + \beta x) = \Phi\Big(\frac{x - (-\alpha/\beta)}{1/\beta}\Big).$$

  So, the probit regression curve has the shape of a normal cdf with mean $-\alpha/\beta$ and standard deviation $1/|\beta|$. $\square$