STAT 3014/ 3914

Semester 2	Applied Statistics	2011

Week 6 Solutions

Tutorial Exercise.

1. Estimated relative risk $= \frac{\hat{p}_1}{\hat{p}_2}$.
   For the drug group: $\hat{p}_1 = \frac{22}{82} = 0.2683$.
   For the placebo group: $\hat{p}_2 = \frac{28}{82} = 0.3415$.
   Relative risk $= \frac{0.2683}{0.3415} = 0.7857$.

   Test $H_0 : p_1 = p_2$ *vs.* $H_1 : p_1 \neq p_2$.

	No effect	Helped/harmed	
Drug	22 (25)	60 (57)	82
Placebo	28 (25)	54 (57)	82
	50	114	164

   $$\chi^2 = \sum_{ij} \frac{o_{ij}^2}{e_{ij}} - 164 = 1.0358$$

   p-value $= P(\chi_1^2 \geq 1.0358) = 0.309$.

   Thus the data are consistent with the drug and placebo having the same effect.

2. Consider the 2×2 table

a	b	(a+b)
c	d	(c+d)
(a+c)	(b+d)	n

   Show that $(a - (a+b)(a+c)/n) = (ad - bc)/n$. Hence show that the Pearson statistic can be written as
   $$X^2 = \frac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}.$$

   $$\chi^2 = \sum \frac{(o_{ij} - e_{ij})^2}{e_{ij}}$$

   Note:

   $$a - \frac{(a+b)(a+c)}{n} = \frac{a(a+b+c+d) - (a^2+ab+ac+bc)}{n} = \frac{(ad-bc)}{n}$$

$$b - \frac{(a+b)(b+d)}{n} = \frac{(bc-ad)}{n}$$
$$c - \frac{(a+c)(c+d)}{n} = \frac{(bc-ad)}{n}$$
$$d - \frac{(c+d)(b+d)}{n} = \frac{(ad-bc)}{n}$$

Thus

$$\begin{aligned}
\chi^2 &= \sum \frac{(o_{ij}-e_{ij})^2}{e_{ij}} \\
&= \frac{(ad-bc)^2)}{n^2}\left[\frac{n}{(a+b)(a+c)} + \frac{n}{(a+b)(b+d)} + \frac{n}{(a+c)(c+d)} + \frac{n}{(c+d)(b+d)}\right] \\
&= \frac{(ad-bc)^2}{n}\left[\frac{(b+d)+(a+c)}{(a+b)(a+c)(b+d)} + \frac{(b+d)+(a+c)}{(a+c)(b+d)(c+d)}\right] \\
&= \frac{(ad-bc)^2}{(a+b)(a+c)(b+d)(c+d)}[c+d+a+b] \\
&= \frac{n(ad-bc)^2}{(a+b)(a+c)(b+d)(c+d)}
\end{aligned}$$

3. Consider the following table from a study of women with breast cancer (Effery et al, 1986).

Factor	Cases	Controls	
Yes	4 (5.43)	11 (9.57)	15
No	63 (64.57)	107 (108.93)	170
	67	118	185

(a) Test $H_0 : p_{cases} = p_{controls}$
$\chi^2 = \sum \frac{o_{ij}^2}{e_{ij}} - 185 = 0.644$
p-value $= P(\chi_1^2 \geq 0.644) = 0.422$.
Thus there does not appear to be a relationship between early oral contraceptive use and breast cancer.

(b) Estimate the odds ratio for this situation.
Estimated relative risk $= \frac{4/15}{63/170} = 0.7195 = \frac{1}{1.39}$
Estimated odds ratio $\hat{\theta} = \frac{4 \times 107}{11 \times 63} = 0.6176$.

An approximate 95% confidence interval for the $log(\hat{\theta})$ is

$$\begin{aligned} log(\hat{\theta}) &\pm 1.96\sqrt{\frac{1}{4}+\frac{1}{11}+\frac{1}{63}+\frac{1}{107}} \\ -0.4819 &\pm 1.96\sqrt{0.366} \\ -0.4819 &\pm 1.186 \end{aligned}$$

i.e. $(-1.668, 0.704)$
Thus an approximate 95% CI for θ is $(e^{-1.668}, e^{0.704}) = (0.1886, 2.022)$
Note the interval is NOT centered at 0.6176.

4. Since the observed frequencies for poor growth and development are small, we use Fisher's exact test.

		Complete family	Single parent family	
Growth and	Good	26	33	59
Development	Poor	1	8	9
		27	41	68

Let p_c denote the proportion of children from complete families with poor growth and development and let p_s be the corresponding figure for single parent families. Test $H_0 : p_c = p_s$ vs $H_1 : p_s > p_c$.

Focus on the $(2, 1)$ cell. Small values in these cell support H_1. Therefore pvalue for the test is $P(X \leq 1|$ marginal totals).

$$\frac{\binom{9}{1}\binom{59}{26}+\binom{9}{0}\binom{59}{27}}{\binom{68}{27}}$$

$$= \frac{\binom{27}{26}\binom{41}{33}+\binom{27}{27}\binom{41}{32}}{\binom{68}{59}} = 0.0523 + 0.0071 = 0.0594$$

Thus the result is marginal but there is not strong evidence to support the claim that children from single parent families are disadvantaged in growth and development.

5. Let the proportion agreeing a the two terms be p_1 and p_2 respectively.Test $H_0 : p_1 = p_2$ vs $H_1 : p_1 \neq p_2$.

Using McNemar's test for paired data, the statistic value is

$$T = \frac{(18-6)^2}{18+6} = 6$$

$$p - value = P(\chi_1^2 \geq 6) = P(|Z| \geq 2.449) = 0.014.$$

Thus there is evidence of a shift in attitude over the one month period.