THE UNIVERSITY OF SYDNEY
MATH1905 STATISTICS ADVANCED

Semester 2	**Tutorial Week 13 – Solutions**	2011

Tutorial solutions.

1. (a) $P < 0.05$ is significant
(b) $P > 0.25$ is not significant
(c) $P > 0.20$ is not significant
(d) $P > 0.10$ is not significant.

2. Answer (e) standard deviation and correlation coefficient.

3. Answer (e) $(1.3, 13.7)$.

4. Answer (c) $\mathrm{P}(X \leq 3) = 0.0106$.

5. (a) $a = 9.236$ (3dp)
(b) `qchisq(0.95,10) ==> 18.307`
(c) with `1-pchisq(38.5,25) ==> 0.0413`
(d) with `1-pchisq(22.1,12) ==> 0.0364`

6. $9 + 3 + 3 + 1 = 16$, the probabilities are $\frac{9}{16}, \frac{3}{16}, \frac{3}{16}, \frac{1}{16}$. The corresponding expected frequencies are.

$$E_1 = 556 \times \frac{9}{16} = 312.75,\ E_2 = 104.25,\ E_3 = 104.25,\ E_4 = 34.75.$$

The goodness of fit statistic is $X^2 = \sum_{i=1}^{4} \frac{O_i^2}{E_i} - 556 = 0.47$. Thus, the P-value $= \mathrm{P}(\chi_3^2 \geq 0.47) = 0.92$ so the data are consistent with the model.

7. (a) Total $n = 400$. With expected frequencies under the 'no linkage model' we obtain

$$X^2 = \sum_{i=1}^{4} \frac{O_i^2}{E_i} - 400 = 18$$

and corresponding P-value of $\mathrm{P}(X_3^2 \geq 18) < 0.0005$. We conclude that the 'no linkage' model does not fit the data well.

(b) Estimating p gives $\hat{p} = \frac{86 + 74}{400} = 0.4$. The expected frequencies under the linkage model are

$$E_1 = 120,\ E_2 = 80,\ E_3 = 80,\ E_4 = 120.$$

The observed Pearson X^2 statistic is 1.97. Thus to 3dp,

$$P\text{-value} = \mathrm{P}(\chi_2^2 \geq 1.97) = 0.373.$$

The data are consistent with the linkage model. Note the degrees of freedom are 2 as p has been estimated from the (same) data used to assess the goodness of fit.

8. (a) Assuming a normal model we calculate the expected frequencies using $\bar{x} = 18.85$ as the estimate for μ and $s = 5.55$ as the estimate for σ, i.e. $X \sim \mathcal{N}(18.85, 5.55^2)$:

$$\begin{aligned}
\mathrm{P}(X \leq 12.95) = \mathrm{P}(Z \leq -1.06) = 0.1446 \quad &\Rightarrow \quad E_1 = 80 \times 0.1446 = 11.568,\\
\mathrm{P}(12.95 \leq X \geq 16.95) = \mathrm{P}(-1.06 \leq Z \geq -0.34) = 0.2223 \quad &\Rightarrow \quad E_2 = 80 \times 0.2223 = 17.784,\\
\mathrm{P}(16.95 \leq X \geq 20.95) = \mathrm{P}(-0.34 \leq Z \geq 0.38) = 0.2811 \quad &\Rightarrow \quad E_3 = 80 \times 0.2811 = 22.488,\\
\mathrm{P}(20.95 \leq X \geq 24.95) = \mathrm{P}(0.38 \leq Z \geq 1.10) = 0.2163 \quad &\Rightarrow \quad E_4 = 80 \times 0.2163 = 17.304,\\
\mathrm{P}(X \geq 24.95) = \mathrm{P}(Z \geq 1.10) = 0.1357 \quad &\Rightarrow \quad E_5 = 80 \times 0.1357 = 10.856.
\end{aligned}$$

(b) The goodness of fit statistic is

$$X^2 = \sum_{i=1}^{5} \frac{O_i^2}{E_i} - n = 1.27.$$

The P-value is $\mathrm{P}(\chi^2_{5-2-1} \geq 1.27) =$ `pchisq(1.27,2,lower.tail=FALSE)` $= 0.53$ (2dp) so the data are consistent with the normal model.