Homework 11 answers

S320/520

Upload your answers as a PDF file or Word document through the Assignments tab on Canvas by 4pm, Thursday 3rd December.

Answers must be in your own words. Show working and include all graphs you are asked to draw (in R.)

1. *Trosset chapter 13.4 exercise 1*

   The null hypothesis is $p_1 = \cdots = p_8 = 1/8$. The expected cell counts are $e_j = 144/8 = 18$ for $j = 1, \ldots, 8$. Find the G^2, and compare it to a chi-square distribution with 7 degrees of freedom:

```
ob = c(29, 19, 18, 25, 17, 10, 15, 11)
ex = rep(18, 8)
G2 = 2 * sum(ob * log(ob/ex))
1 - pchisq(G2, df=7)
```

   This gives $G^2 = 16.14$ and a P-value of 0.024. This is small, so we have some evidence that the horse's starting position affects its chance of winning. (Pearson's chi-square is 16.33, giving a P-value of 0.022 and the same conclusion.)

2. *Trosset chapter 13.4 exercise 3*

   (a) Let $p_j = P(E_j)$. Then

$$\begin{aligned} p_1 &= \frac{3}{4} \times \frac{3}{4} = \frac{9}{16} \\ p_2 &= \frac{3}{4} \times \frac{1}{4} = \frac{3}{16} \\ p_3 &= \frac{1}{4} \times \frac{3}{4} = \frac{3}{16} \\ p_4 &= \frac{1}{4} \times \frac{1}{4} = \frac{1}{16} \end{aligned}$$

   (b) The expected cell counts, $e_j = np_j$, are $e_1 = 906, e_2 = 302, e_3 = 302, e_4 = 101$. The maximum likelihood chi-squared statistic is

$$G^2 = 2 \sum_{j=1}^{4} o_j \log(o_j/e_j) = 1.4776.$$

Comparing this to a chi-square distribution with $4 - 0 - 1 = 3$ degrees of freedom, the P-value is `1 - pchisq(1.4776)`, or 0.687. This is not small, so we have no evidence that the probabilities in (a) are wrong. (Pearson's chi-square statistic is 1.4687, giving a P-value of 0.690 and the same conclusion.)

3. *Trosset chapter 13.4 exercise 10*

   Of the 521 fireflies, 298 were male and 223 were female, while 323 were caught at 3 feet and 198 were caught are 35 ft. Given these proportions, if sex and height were independent, we would expect the following counts:

   - Male at 3 ft: $298 \times 323/521 = 185$
   - Male at 35 ft: $298 \times 198/521 = 113$
   - Female at 3 ft: $223 \times 323/521 = 138$
   - Female at 35 ft: $223 \times 198/521 = 85$

   Find the G^2, and compare it to a chi-square distribution with $(2-1)(2-1) = 1$ degrees of freedom:

```
ob = c(173, 125, 150, 73)
ex = c(298*323, 298*198, 223*323, 223*198) / 521
G2 = 2 * sum(ob * log(ob/ex))
1 - pchisq(G2, df=1)
```

   This gives $G^2 = 4.62$ and a P-value of 0.032. This is small, so we have some evidence that the sex ratio of Panamanian sandflies varies with height above ground. (Pearson's chi-square is 4.59, giving the same P-value to 2 dp.)

4. *Trosset chapter 13.4 exercise 11*

   The o_{ij} are given and $n = 538$. Under the null hypothesis that patient response is independent of histological type, the $e_{ij} = o_{i+}o_{+j}/n$ are as follows: For each of the 12 cells, we compute

	Positive	Partial	None
LP	60.7	18.9	24.4
NS	56.0	17.5	22.5
MC	155.2	48.5	62.3
LD	42.0	13.1	16.9

   $o_{ij} \log(o_{ij}/e_{ij})$, then sum and double to obtain $G^2 = 68.3$. There are $(4-1)(3-1) = 6$ degrees of freedom, so the P-value is `1 - pchisq(68.3, df=6)`, or 9.1×10^{-13}. We reject the null hypothesis of independence: The response to treatment does vary by histological type. (Pearson's chi-square is 75.9, giving a P-value of 2.5×10^{-14} and the same conclusion.)

5. (Trosset chapter 14.6 exercise 3.)

   (a) Looking at a QQ plot, x seems consistent with a normal distribution.

(b) Looking at a QQ plot, y seems consistent with a normal distribution.

(c) The data is not bivariate normal — it's not an ellipse but an X.

(d) It looks like the data is a mixture of two lines.

6. *(Compulsory for S520, optional for half-credit for S320.) Trosset chapter 13.4 exercise 6. NOTE: From the book errata, the correct formula for the Poisson probability mass function is*

$$P(X = x) = \frac{\lambda^x \exp(-\lambda)}{x!}$$

This can be calculated in R as `dpois(x, lambda)`.

(a)
```
freq = c(57, 203, 383, 525, 532, 408, 273, 139, 45, 27, 10, 4, 0, 1, 1)
xbar = sum(freq * (0:14)) / 2608
```
$\bar{x}$ is 3.87. This is also a good estimate of λ.

(b) Combining the high categories into "10 or more":

```
ob = c(57, 203, 383, 525, 532, 408, 273, 139, 45, 27, 16)
ex = dpois(0:9, xbar) * 2608
ex[11] = 2608 - sum(ex[1:10])
G2 = 2 * sum(ob * log(ob/ex))
1 - pchisq(G2, df=9)
```

Using 11 categories, G^2 is 13.916. Comparing this to a chi-square distribution with $11 - 1 - 1 = 9$ degrees of freedom (the extra "-1" is because we estimated $\hat{\lambda} = \bar{x}$ from the data) gives a P-value of 0.125. The alpha-particle data is consistent with a Poisson distribution. (Pearson's chi-square is 12.873, giving a P-value of 0.168 and the same conclusion.)