STA 114: Statistics

HW 10

Due Wed Nov 30 2011

1. A researchers surveys n college students and counts how many support, how many oppose and how many are undecided about a recently introduced federal policy. Letting X_1, X_2, X_3 denote these counts, she models $X = (X_1, X_2, X_3)$ as $X \sim \mathsf{Multinomial}(n, p)$, $p \in \Delta_3$.

   (a) Give the p-value for testing $H_0 : p = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ against $p_0 \neq (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ based on Pearson's chi-square tests for observed counts $X_1 = 140$, $X_2 = 165$, $X_3 = 195$.

   (b) The researcher wants to test whether the actual proportions of supporters and opposers in the entire college are equal. Give a mathematical formulation of this null hypothesis.

   (c) Find the restricted MLE of p under the null hypothesis in part (b) [i..e., maximize the likelihood function only over the null set].

   (d) Give the p-value for testing the null hypothesis in part (b) based on Pearson's chi-square tests, with same observed counts as in part (a).

2. A total of 309 wafer defects were recorded and the defects were classified as being one of four types, A, B, C, or D. At the same time each wafer was identified according to the production shift in which it was manufactured, 1, 2, or 3. These counts are presented in the following table.

Shift	Type of Defect A	B	C	D	Total
1	15	21	45	13	94
2	26	31	34	5	96
3	33	17	49	20	119
Total	74	69	128	38	309

   Give the p-value (based on Pearson's chi-squared tests) for testing independence between type of defect and production shift.

3. Consider a 2-category count data $X = (X_1, X_2)$ modeled as $X \sim \mathsf{Multinomial}(n, p)$, $p \in \Delta_2$. We want to test $H_0 : p = p_0$ against $H_1 : p \neq p_0$ for some $p_0 = (p_{01}, p_{02}) \in \Delta_2$. We could use Pearson's chi-square tests for this. As an alternative, we could just look at X_1 (because $X_2 = n - X_1$ is completely determined by X_2) as our data and use the model $X_1 \sim \mathsf{Binomial}(n, p_1)$ and test for $H_0 : p_1 = p_{01}$ against $H_1 : p_1 \neq p_{01}$ based on the approximate ML tests for this model. Are the two tests equivalent, i..e, would they result in identical p-values for any observation $x = (x_1, x_2)$?

4. Twenty two Duke students reported their weekly food expenditures to be

```
125 140 200 200 190 100 140 250 125 180 110
125 120 130 140 150 120 100  95 195  95 130
```

(a) Test if these data are from a normal distribution [use bins with equal expected counts not lower than 5]. Report the p-value range.

(b) Test if the log-food expenditures are from a normal distribution [same criteria for bins]. Report the p-value range.