TUTORIAL EXERCISE PACKAGE - 2013

MATH1015 - BIOSTATISTICS

Semester 1 **Tutorial Week 11** 2013

Given observed frequencies $x_1, \ldots, x_g$, the χ^2 test of *model fit* between x_i and the model under $H_0 : p_i = p_{i0},\ i = 1, \ldots, g$ is $\chi^2_{\text{obs}} = \sum_{i=1}^{g} \frac{(x_i - np_{i0})^2}{np_{i0}} \sim \chi^2_{g-1}$ where $E_i = np_{i0} \geq 5$.

Given observed frequencies $x_{11}, \ldots, x_{rc}$ in the $r \times c$ contingency table, the χ^2 test of *independence* between the row and column variables under $H_0 : p_{ij} = p_i p_j,\ i = 1, \ldots, r; j = 1, \ldots, c$, is

$$\chi^2_{\text{obs}} = \sum_{i=1}^{r}\sum_{j=1}^{c} \frac{(x_{ij} - \frac{r_i \times c_j}{n})^2}{\frac{r_i \times c_j}{n}} \sim \chi^2_{(r-1)(c-1)} \text{ where } E_{ij} = \frac{r_i \times c_j}{n} \geq 5 \text{ and}$$

r_i and c_j are the i row total and j column total respectively.

1. Use the χ^2 tables to answer the following.

(a) If $P(\chi^2_5 > a) = 0.10$, find a.

(b) If $P(\chi^2_{10} \geq a) = 0.05$, find a.

(c) Range for $P(\chi^2_{25} \geq 38.5)$.

(d) Range for $P(\chi^2_{12} \geq 22.1)$.

2. **Multiple choice** The following table gives the observed frequencies of genotypes A, B, and C of 100 plants:

Genotype	A	B	C	Total
Observed frequency, O_i	18	55	27	100

Under the null hypothesis that A, B, and C are in the ratio of 1:2:1, the expected frequencies, E_i's (respectively) are:

(a) 25, 25, 75 (b) 25, 25, 50 (c) 50, 25, 25 (d) 25, 50, 25 (e) none of these.

3. **Multiple choice** The observed value of $\chi^2_{\text{obs}} = \sum_{i=1}^{3} \frac{(O_i - E_i)^2}{E_i}$ for the data in Q2 is:

(a) 2.62 (b) 102.62 (c) 100 (d) 0.262 (e) 1.62

4. Using the information from Q2 and Q3, test whether the model 1:2:1 fits the data (in Q2) well.

5. The two main causes of mortality are diseases of the heart and cancer. Ten years ago, the mortality rates of diseases of the heart, cancer and other were 45%, 40% and 15% respectively. To determine if the mortality rates have changed over the past ten years, 200 records of death are randomly chosen and 102 are caused by diseases of the heart and 82 by cancer. Can the analyst infer at the 5% level of significance that the mortality rates have changed over the past ten year?

6. The personnel manager of a consumer product company asked a random sample of employees how they felt about the work they were doing. The following table gives a breakdown of their responses by gender.

Gender	Very interesting	Fairly interesting	Not interesting	Total
Male	70	41	9	120
Female	35	34	11	80
Total	105	75	20	200

Do the data provide sufficient evidence to conclude that the level of job satisfaction is related to gender? Use $\alpha = 0.10$.

Use R to answer Q7 to Q10

7. Check your answers to Q1 (c) and (d) using R. **Hint:** `pchisq(q,df)` gives $P(\chi^2_{\text{df}} < q)$.

8. Do Q5 using R.

Hint: Execute `chisq.test(c(102,82,16),p=c(0.45,0.4,0.15))`. Note that the the default value of the parameter `p` refers to equal proportions. Because the hypothesized proportions `p` are not all equal, we have to specify them.

9. Do Q6 using R.

Hint: Create a vector, say `x`, with the data 70, 35, 41, 34, 9, 11 from Q6. Transform the vector `x` into a 2 by 3 matrix `x.mat` using `x.mat=matrix(x,2,3)`. Then execute the command `chisq.test(x.mat)`.

10. The file `lead-IQ.txt` contains two columns of observations from 124 children. The first column stores levels of `distance` (1: 0-1 miles; 2: 1-2.5 miles; 3: 2.5-4.1 miles) of their homes to a smelter. The second column stores their levels of `IQ` (1: below 100; 2 above 100).

(a) Read the data using the command below:

```
dat=read.table(file=url("http://www.maths.usyd.edu.au/math1015/r/lead-IQ.txt"))
```

Set `x=dat[,1]` and `y=dat[,2]`.

(b) Generate a 3×2 contingency table using `tab=table(x,y)`. Type `tab` again to view the table.

(c) Perform a Chi-square test to test the dependence between the factors of `distance` and `IQ` using the command `chisq.test(tab)`. Report the test statistic, degrees of freedom and P-value. What is your conclusion?

Semester 1 | Problem Sheet Week 11 | 2013

---

1. Consumer panel preferences for three proposed fast food restaurants are as follow:

Restaurant A	Restaurant B	Restaurant C
48	62	40

Test if there is a preference among the three restaurants.

2. Book P.182, Q10.13. Note that the test for heterogeneity of the proportions is equivalent to the test of dependence between the two variables.

3. Book P.182-183, Q10.20