Math 308 Assignment 7
Exercises 3.9

Nakul Joshi

March 5, 2014

4

The null hypothesis is that the difference in proportions is zero. However, performing the permutation test gave a p-value of 0.002, allowing us to reject the null at 1% confidence. Thus, the difference in proportions is statistically significant.

8

The p-value is 1, which does not let us reject the null hypothesis that the presence of competition has no value on the height change of the seedlings.

11

Null Hypothesis Voting preference is independent of age.

Alternative hypothesis Voting preference depends on age.

Age	Response		
	For	Against	All
18-29	172	52	224
30-49	313	103	416
50+	258	119	377
All	743	274	1017

Table 1: Observed values

Multiplying column marginal fractions by row marginal totals, we can get the expected values:

Age	Response	
	For	Against
18-29	164	60
30-49	304	112
50+	275	102

Table 2: Expected values

Then, we calculate the χ^2 test statistic:
$c = \sum_{i,j}^{\text{all cells}} \frac{(\text{observed}_{i,j} - \text{expected}_{i,j})^2}{\text{expected}_{i,j}} = 6.33$

Under the null, C follows a χ^2 distribution with $(3-1) \times (2-1) = 2$ degrees of freedom; i.e. $C \sim \chi^2_2$. So, the p-value is $P(C > c) = \int_c^\infty \frac{e^{-t/2}}{2}\,\mathrm{d}t \approx 0.042$.

Thus, we can reject the null at 5% significance, but not at 1% significance.

13

a)

We are testing for homogenity since we want to know whether the distribution of fin ray counts differs from lake to lake.

b)

Null hypothesis Fin ray distributions are the same from lake to lake.

Alternative hypothesis Fin ray distributions are different from lake to lake.

Habitat	Ray Count						
	36	35	34	33	32	31	All
Guadalupe	14	30	42	78	33	14	211
Cedro	11	28	53	66	27	9	194
San Clemente	10	17	61	53	22	10	173
All	71	110	190	230	114	64	779

Habitat	Ray Count					
	≥36	35	34	33	32	≤31
Guadalupe	19	30	51	62	31	17
Cedro	18	27	47	57	28	16
San Clemente	16	24	42	51	25	14

Table 3: Expected Values

$c = \sum_{i,j}^{\text{all cells}} \frac{(\text{observed}_{i,j} - \text{expected}_{i,j})^2}{\text{expected}_{i,j}} = 41.77$, where $C \sim \chi^2_{10}$. So, $p = P(C > c) = \int_c^\infty \frac{t^{10/2-1}e^{-t/2}}{2^{10/2}\Gamma(10/2)}\,\mathrm{d}t = \int_c^\infty \frac{t^4 e^{-t/2}}{768}\,\mathrm{d}t = 8 \times 10^{-6}$. So, we can reject the null.

17

a)

Happiness	Gender	
	Female	Male
Not too happy	109	61
Pretty happy	406	378
Very happy	205	210

Table 4: Observed happiness against gender data

b)

We get a p value of 0.004, allowing us to reject the null hypothesis of independence at a 1% significance level.

19

a)

Let the elements of the contingency table be $o_{i,j}$, and let their corresponding row and column totals be r_i and c_j respectively. Further, let the total number of observations $\sum_{i,j} o_{i,j} = n$. The corresponding expected values are then $e_{i,j} = \frac{r_i c_j}{n}$, which gives the test statistic $c = \sum_{i,j} \frac{(o_{i,j} - e_{i,j})^2}{e_{i,j}}$

However, if each element is multiplied by k, then $o_{i,j}$,r_i, c_j and n are each multiplied by k. So, the corresponding expected values become $e^*_{i,j} = \frac{k^2}{k} \times \frac{r_i c_j}{n} = k \times e_{i,j}$. The new test statistic $c^* = \sum_{i,j} \frac{(o^*_{i,j} - e^*_{i,j})^2}{e^*_{i,j}} = \sum_{i,j} \frac{k^2}{k} \times \frac{(o_{i,j} - e_{i,j})^2}{e_{i,j}} = k \times c$. Thus, the test statistic is also multiplied by k.

However, the marginal probabilities are unchanged since the k's cancel on the row and overall totals. Further, the degrees of freedom are unaffected since they only depend upon r and c.

b)

Originally, the p-value was $\int_c^\infty f(t;k)\,\mathrm{d}t$, but the new p-value becomes $\int_{kc}^\infty f(t;k)\,\mathrm{d}t$. Thus, the p value reduces, since $k > 1 \implies kc > c$.

22

a)

p	q
0.2	16.1
0.4	20.2
0.6	23.8
0.8	27.9

b)

Interval	Counts	
	Obs.	Exp.
<16.11	16	10
16.11–20.23	13	10
20.23–23.77	9	10
23.77–27.89	9	10
>27.89	3	10

c)

We get a p-value of 0.048, which does not let us reject the hypothesis that the data was drawn from a $N(22, 7^2)$ distribution.

25

We get a p-value of 0.78, which does not let us reject the hypothesis that the numbers are uniformly distributed.

29

a)

Let N_{ij}be the element at row i and column j. Then, since there are only two rows and two columns, let the remaining row and column indices be $\bar{\imath}$ and $\bar{\jmath}$ respectively. Then, the row total is $R_i = N_{ij} + N_{i\bar{\jmath}}$ and the column total is $C_j = N_{ij} + N_{\bar{\imath}j}$. So, the expected value at i, j is

$$\begin{aligned}\breve{E}[N_{ij}] &= \frac{R_i \times C_j}{n} \\ &= \frac{(N_{ij} + N_{i\bar{\jmath}})(N_{ij} + N_{\bar{\imath}j})}{n}\end{aligned}$$

This gives:

$$\begin{aligned}&N_{ij} - \breve{E}[N_{ij}] \\ &= N_{ij} - \frac{(N_{ij} + N_{i\bar{\jmath}})(N_{ij} + N_{\bar{\imath}j})}{n} \\ &= \frac{N_{ij}(\cancel{N_{ij}} + \cancel{N_{i\bar{\jmath}}} + \cancel{N_{\bar{\imath}j}} + N_{\overline{\imath\jmath}}) - (\cancel{N_{ij}^2} + \cancel{N_{ij}N_{i\bar{\jmath}}} + \cancel{N_{ij}N_{\bar{\imath}j}} + N_{i\bar{\jmath}}N_{\bar{\imath}j})}{n} \\ &= \frac{N_{ij}N_{\overline{\imath\jmath}} - N_{i\bar{\jmath}}N_{\bar{\imath}j}}{n}\end{aligned}$$

Thus:

$$(N_{ij} - \breve{E}[N_{ij}])^2 = \left(\frac{N_{ij}N_{\overline{\imath\jmath}} - N_{i\bar{\jmath}}N_{\bar{\imath}j}}{n}\right)^2$$

This expression is symmetrical; i.e., it does not change by swapping i and $\bar{\imath}$ or j and $\bar{\jmath}$. So, it is equivalent to:

$$\left(\frac{N_{11}N_{22} - N_{21}N_{12}}{n}\right)^2$$

which is independent of i, j $\square$.

b)

Call the previously obtained expression k. Then,

$$\begin{aligned}C &= kn/R_1C_1 + kn/R_1C_2 + kn/R_2C_1 + kn/R_2C_2 \\ &= kn\frac{R_1C_1 + R_1C_2 + R_2C_1 + R_2C_2}{R_1R_2C_1C_2} \\ &= kn\frac{R_1(C_1 + C_2) + R_2(C_1 + C_2)}{R_1R_2C_1C_2} \\ &= kn\frac{n(R_1 + R_2)}{R_1R_2C_1C_2} = k\frac{n^3}{R_1R_2C_1C_2} \\ &= \frac{(N_{11}N_{22} - N_{21}N_{12})^2}{n^2}\frac{n^3}{R_1R_2C_1C_2} \\ &= n(N_{11}N_{22} - N_{21}N_{12})^2/R_1R_2C_1C_2 \quad \square\end{aligned}$$

c)

Via R, the expression is verified since both methods yield $C \approx 0.0234$