Math 332 A - Mathematical Statistics

Ethan Jensen

April 10, 2020

HW p.370 #8, p.379 #74, p.380 #78

1 p. 370 #8

Show that the two formulas for χ^2 on pages 368 and 369 are equivalent.

We first define the following terms:
$f_{i1} = x_i$, and $f_{i2} = n_i - x_i$
$e_{i1} = n_i\hat{\theta}_i$, and $e_{i2} = n_i(1-\hat{\theta}_i)$

Let us begin.

$$\sum_{i=1}^{k}\sum_{j=1}^{2}\frac{(f_{ij}-e_{ij})^2}{e_{ij}} = \sum_{i=1}^{k}\sum_{j=1}^{2}\frac{f_{ij}^2 - 2e_{ij}f_{ij} + e_{ij}^2}{e_{ij}} = \sum_{i=1}^{k}\sum_{j=1}^{2}\frac{f_{ij}^2}{e_{ij}} - 2f_{ij} + e_{ij}$$

$$\sum_{i=1}^{k}\sum_{j=1}^{2}\frac{(f_{ij}-e_{ij})^2}{e_{ij}} = \sum_{i=1}^{k}\frac{f_{i1}^2}{e_{i1}} + \frac{f_{i2}^2}{e_{i2}} - 2(f_{i1}+f_{i2}) + (e_{i1}+e_{i2})$$

$$\sum_{i=1}^{k}\sum_{j=1}^{2}\frac{(f_{ij}-e_{ij})^2}{e_{ij}} = \sum_{i=1}^{k}\frac{x_i^2}{n_i\hat{\theta}_i} + \frac{(n_i-x_i)^2}{n_i(1-\hat{\theta}_i)} - 2n_i + n_i$$

$$\sum_{i=1}^{k}\sum_{j=1}^{2}\frac{(f_{ij}-e_{ij})^2}{e_{ij}} = \sum_{i=1}^{k}\frac{x_i^2(1-\hat{\theta}_i) + (n_i-x_i)^2\hat{\theta}_i - n_i^2\hat{\theta}_i(1-\hat{\theta}_i)}{n_i\hat{\theta}_i(1-\hat{\theta}_i)}$$

$$\sum_{i=1}^{k}\sum_{j=1}^{2}\frac{(f_{ij}-e_{ij})^2}{e_{ij}} = \sum_{i=1}^{k}\frac{x_i^2 - x_i^2\hat{\theta}_i + n_i^2\hat{\theta}_i - 2n_ix_i\hat{\theta}_i + x_i^2\hat{\theta}_i - n_i^2\hat{\theta}_i + n_i^2\hat{\theta}_i^2}{n_i\hat{\theta}_i(1-\hat{\theta}_i)}$$

$$\sum_{i=1}^{k}\sum_{j=1}^{2}\frac{(f_{ij}-e_{ij})^2}{e_{ij}} = \sum_{i=1}^{k}\frac{x_i^2 - 2n_ix_i\hat{\theta}_i + n_i^2\hat{\theta}_i^2}{n_i\hat{\theta}_i(1-\hat{\theta}_i)} = \sum_{i=1}^{k}\frac{(x_i-n_i\hat{\theta}_i)^2}{n_i\hat{\theta}_i(1-\hat{\theta}_i)}$$

$$\boxed{\chi^2 = \sum_{i=1}^{k}\sum_{j=1}^{2}\frac{(f_{ij}-e_{ij})^2}{e_{ij}} = \sum_{i=1}^{k}\frac{(x_i-n_i\hat{\theta}_i)^2}{n_i\hat{\theta}_i(1-\hat{\theta}_i)}}$$

■

2 p. 379 #74

In random samples of 250 people with low incolmes, 200 people with average incomes, and 150 people with high incomes, there were, respectively, 155, 118, and 87 who favor a certain piece of legislation. Use the 0.05 level of significance to test the null hypothesis $\theta_1 = \theta_2 = \theta_3$ (that the proportion of the people favoring the legislation is the same for all three income groups) against the alternative hypothesis that the three θ's are not all equal.

Constructing our table, we have

Data (f_{ij})	**Favors**	**Does not Favor**	**Totals**
Low Income	155	95	250
Mid Income	118	82	200
High Income	87	63	150
Totals	360	240	600

$H_0:\ \theta_1 = \theta_2 = \theta_3$
$H_1:\ \theta_1, \theta_2$, and θ_3 are not all equal.
$\alpha = 0.05$. The pooled estimate of θ is $\theta = \frac{360}{600} = 0.6$.
With our pooled estimate, we can construct the estimate table.

Expected Frequencies (e_{ij})	**Favors**	**Does not Favor**	**Totals**
Low Income	150	100	250
Mid Income	120	80	200
High Income	90	60	150
Totals	360	240	600

From no particular theorem, the following sum is chi-square with k-1 degrees of freedom.

$$\sum_{i=1}^{k}\sum_{j=1}^{2}\frac{(f_{ij}-e_{ij})^2}{e_{ij}}$$

So we do the following test with 0.05 level of significance. We check:

$$\sum_{i=1}^{3}\sum_{j=1}^{2}\frac{(f_{ij}-e_{ij})^2}{e_{ij}} < \chi^2_{0.05,2}$$

.

$$\chi^2_{0.05,2} = 5.991$$

$$\sum_{i=1}^{3}\sum_{j=1}^{2}\frac{(f_{ij}-e_{ij})^2}{e_{ij}} = 0.75(\text{very small!})$$

Since $0.75 < 5.991$, we must accept the null hypothesis.

$\boxed{\text{Income level does not affect whether people favor this piece of legislation}}$

$\boxed{\text{the p-value is 0.687289. (big!)}}$

3 p. 380 #78

The following sample data pertain to the shipments received by a large firm from three different vendors:

Data (f_{ij})	**Number rejected**	**Number imperfect but acceptable**	**Number perfect**	**Totals**
Vendor A	12	23	89	124
Vendor B	8	12	62	82
Vendor C	21	30	119	170
Totals	41	65	270	376

Test at the 0.01 level of significance whether the three vendors ship products of equal quality.

I am not quite sure how this works, but I assume the population is multinomial. We break the cases into three categories into two.

$\theta_1 = 41/376$, $\theta_2 = 65/376$, $\theta_3 = 270/376$

We construct the table of expected values for each cell using the 3 θ values.

Expected Frequencies (e_{ij})	**Number rejected**	**Number imperfect but acceptable**	**Number perfect**	**Totals**
Vendor A	12	23	89	124
Vendor B	8	12	62	82
Vendor C	21	30	119	170
Totals	41	65	270	376

From no particular theorem, the following sum is chi-square with $(k-1)(c-1)$ degrees of freedom, where k and c are the category sizes.

$$\sum_{i=1}^{k}\sum_{j=1}^{c}\frac{(f_{ij}-e_{ij})^2}{e_{ij}}$$

So we do the following test with 0.01 level of significance. We check:

$$\sum_{i=1}^{3}\sum_{j=1}^{3}\frac{(f_{ij}-e_{ij})^2}{e_{ij}} < \chi^2_{0.01,4}$$

$$\text{From Table 5, } \chi^2_{0.01,4} = 13.277$$

$$\sum_{i=1}^{3}\sum_{j=1}^{3}\frac{(f_{ij}-e_{ij})^2}{e_{ij}} = 1.300642$$

$\boxed{\text{Since } 1.300642 < 13.277\text{, we accept the null hypothesis.}}$

The p-value of 1.300642 is 0.86126, which is greater than 0.01.

■