Stat 312: Lecture 26
Contingency Tables

Moo K. Chung
mchung@stat.wisc.edu

October 22, 2014

Concepts

1. Two-way contingency table: suppose there are I populations and each population is classified into J categories. Let n_{ij} be the number of *observed* elements in population i which fall into category j. We denote $n_{\bullet j} = \sum_{i=1}^{I} n_{ij}$ and $n_{i\bullet} = \sum_{j=1}^{J} n_{ij}$.

2. Testing for *homogeneity*. Let p_{ij} be the proportion of the elements in population i which fall into category j. Note that $\sum_{j=1}^{J} p_{ij} = 1$. We want to test if the proportions in the different categories are the same for all populations, i.e.

$$H_0 : p_{1j} = p_{2j} = \cdots = p_{Ij} \text{ for all } j.$$

3. The *expected* number of element $\mathbb{E}N_{ij}$ in population i which falls into category j. $\mathbb{E}N_{ij} = n_{i\bullet}p_{ij}$. Under the null hypothesis, $p_{ij} = \cdots = p_{Ij} = p_j$. So $\mathbb{E}N_{ij} = n_{i\bullet}p_j$. We estimate p_j by pooling I samples together. $\hat{p}_j = n_{\bullet j}/n$.

4. Test statistic:

$$\chi^2 = \sum_{i,j} \frac{(n_{ij} - \mathbb{E}N_{ij})^2}{\mathbb{E}N_{ij}} \sim \chi^2_{(I-1)(J-1)}.$$

In-class problems

Example. Suppose that 20 out of 50 females and 10 out of 40 males are depressed in a sample. Determine if the frequency of depression is related to sex.

Solution. Following the notations above, we are interested in testing $H_0 : p_{11} = p_{21}, p_{12} = p_{22}$. $n = 90$. Under H_0, we estimate $\hat{p}_1 = (20+10)/(20+10+30+30) = 1/3$ and $\hat{p}_2 = 1 - p_1 = 2/3$. The test statistic value is $\chi^2 =$

The cutoff value $\chi^2_{\alpha,1}$ for α-level test can be computed from $N(0,1)$.

$$\alpha = P(\chi^2_1 > \chi^2_{\alpha,1}) = P(Z^2 > \chi^2_{\alpha,1}) = 2P(Z > \chi_{\alpha,1}).$$

So $\chi^2_{\alpha,1} = Z^2_{\alpha/2}$. For $\alpha = 0.05$, $\chi^2_{\alpha,1} = 1.96^2$.

Self-study problems

Example 14.13.