Equivalence of the Chi-Squared Tests

The Chi-square test for homogeneity

Consider Situation (2) in the bottom half of Fig. 11.2.7 in the text where we wish to test a null hypothesis of identical row distributions. Once more we need an estimate of the counts that we would expect to see if the hypothesis was true. Under H_0, the probability of any individual falling into $B = j$ is the same, p_j say, regardless of the sample or group that the individual belongs to. Therefore, of the R_i individuals in group $A = i$, we would expect $R_i p_j$ to fall into $B = j$. This is the expected value of the count in cell (i, j) when H_0 is true. We can't use this expression for the expected counts because we don't know the p_j's. However, we do have $n = \Sigma R_i$ individuals in the data set, and when H_0 is true, each individual has probability p_j of falling into $B = j$. Our best estimate of p_j in this situation is simply the proportion of all n individuals who fall into $B = j$, namely

$$\widehat{p}_j = \frac{C_j}{n}.$$

The corresponding estimate of the expected count in cell (i, j) is, therefore,

$$\widehat{E}_{ij} = R_i \widehat{p}_j = \frac{R_i C_j}{n}$$

which is the same as the formula given in Section 11.2.2 in the text.

The Chi-square test for independence

To customize the Chi-square test for this situation, we have to estimate the counts that we would expect to see in the table if the independence hypothesis was true. As depicted in Situation (1) of Fig. 11.2.7 in the text, we have n individuals falling into the table and the probability that any individual falls into cell (i, j) is $\text{pr}(A = i \text{ and } B = j)$. The expected number falling into cell (i, j) is thus

$$\begin{aligned} E_{ij} &= n \, \text{pr}(A = i \text{ and } B = j) \\ &= n \, \text{pr}(A = i) \, \text{pr}(B = j) \quad \text{if } H_0 \text{ is true.} \end{aligned}$$

We can estimate $\text{pr}(A = i)$ by the proportion of the n individuals that fall into the ith row, namely R_i/n. Similarly, we can estimate $\text{pr}(B = j)$ by the proportion falling into the jth column, namely C_j/n. Substituting these estimates for $\text{pr}(A = i)$ and $\text{pr}(B = j)$ in the equation above gives

$$\widehat{E}_{ij} = n \frac{R_i}{n} \frac{C_j}{n} = \frac{R_i C_j}{n},$$

which is the same formula as the one used for the chi-squared test of homogeneity. We can motivate the degrees of freedom formula $df = (I-1)(J-1)$ as follows. For Chi-square tests in which we have k categories but have to estimate q independent parameters from the data in order to estimate the expected counts, the degrees of freedom are

$$df = k - 1 - q.$$

We have n individuals falling into IJ categories. The distribution for A involves I probabilities which add to 1. The number of independent parameters is thus $I-1$. Similarly, the distribution of B involves $J-1$ independent parameters. Putting it all together, we get $q = (I-1)+(J-1)$ and

$$df \;=\; IJ - 1 - ((I-1)+(J-1)) \;=\; (I-1)(J-1).$$

Equivalence of independence and homogeneity tests in Situation (1)

The probability distribution for the jth column is given by each frequency entry in the column divided by by the column sum of the frequencies, or in terms of probabilities,

$$\frac{\mathrm{pr}(A=i \text{ and } B=j)}{\sum_i^I \mathrm{pr}(A=i \text{ and } B=j)} = \frac{\mathrm{pr}(A=i \text{ and } B=j)}{\mathrm{pr}(B=j)} = \mathrm{pr}(A=i \,|\, B=j).$$

If the independence hypothesis is true, it follows from the definition of independence of two events that, for each column j,

$$\mathrm{pr}(A=i \,|\, B=j) = \mathrm{pr}(A=i) \quad \text{for every } i.$$

The right hand side of the above equation tells us that the jth column distribution does not depend on j so that the jth column distribution is the same for every j. Independence is therefore equivalent to equal column distributions (column homogeneity). Furthermore, independence also implies

$$\mathrm{pr}(B=j \,|\, A=i) = \mathrm{pr}(B=j) \quad \text{for every } i,$$

and by a similar argument, independence is equivalent to equal row distributions (row homogeneity).