

## Permutation tests for contingency tables

A two-way contingency table displays the counts of observations that fall into each of the pairs of categories from two characteristics of interest.

The two types of sampling that we will consider are i) All  $n$  individuals selected at random, or ii) A fixed number of observations  $n_i$  is selected according to row characteristic  $i$ , and classified according to the column characteristic. These lead to the null hypotheses of independence, and homogeneity of row distributions, respectively, but they are equivalent and have the same chi-square test statistic:

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(n_{ij} - e_{ij})^2}{e_{ij}}.$$

When expected cell frequencies  $e_{ij}$  are  $\geq 5$ , then the chi-square statistic has an approximate chi-square distribution with  $(r-1)(c-1)$  degrees of freedom. When the expected counts do not satisfy this rule, then the chi-square approximation is not appropriate, but we can instead develop a permutation test.

### The Permutation Chi-Square Test

### Multiple Comparisons in Contingency Tables

The text describes a way to follow up after a rejection of the null hypothesis of homogeneity of row proportions, using a Tukey HSD-like procedure.