

CONTINGENCY TABLES

Why

We want to summarize the interaction between two binary traits.

Definition

The contingency table of a population $\{1,\ldots,n\}$ with respect to binary traits $a,b:\{1,2,\ldots,n\}\to\{0,1\}$ is the array $A\in \mathbf{N}^{2\times 2}$ of natural numbers defined by

$$A = \begin{bmatrix} |a^{-1}(0) \cap b^{-1}(0)| & |a^{-1}(0) \cap b^{-1}(1)| \\ |a^{-1}(1) \cap b^{-1}(0)| & |a^{-1}(0) \cap b^{-1}(1)| \end{bmatrix}.$$

We interpret A_{11} as the number of individuals which have neither trait, A_{12} as the individuals which have trait b but not trait a, and so on.

Normalization

These four sets partition $\{1, 2, ..., n\}$, so that if we divide the elements by n, we obtain four numbers which sum to 1, the 2 by 2 table with these entries is called the *normalized contingency table*.

Contingency arrays

In general, we have k binary traits, each of which an individual may or may not have. We encode these traits using k functions

$$a_1, \ldots, a_k : \{1, 2, \ldots, n\} \to \{0, 1\}.$$

The contingency array is k-dimensional array A, with

$$A_x = \bigcap_{j=1}^k a_j^{-1}(x_j),$$

where $x \in \{0, 1\}^k$.

