

分割表の 2 属性の独立性検定

分割表（クロス集計表）の 2 つの属性が独立であることを検定する。

2 つの属性を持つ N 個の標本を考え、

- ・ 属性 A は R 個のカテゴリ $A_1 \dots A_R$ に分類される。
- ・ 属性 B は C 個のカテゴリ $B_1 \dots B_C$ に分類される。

とする。

カテゴリ A_i 、 B_j の観測度数を f_{ij} として、以下のクロス集計表を作成する。

	B_1	B_2	\dots	B_C	計
A_1	$f_{1,1}$	$f_{1,2}$	\dots	$f_{1,C}$	a_1
A_2	$f_{2,1}$	$f_{2,2}$	\dots	$f_{2,C}$	a_2
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
A_R	$f_{R,1}$	$f_{R,2}$	\dots	$f_{R,C}$	a_R
計	b_1	b_2	\dots	b_C	N

上記の分割表（クロス集計表）が、属性 A、B について独立であるとは、 A_i かつ B_j の確率 $P(A_i \text{ かつ } B_j)$ について

$$P(A_i \text{ かつ } B_j) = P(A_i) \times P(B_j)$$

$$P(A_i) = a_i / N$$

$$P(B_j) = b_j / N$$

が成り立つことなので、属性 A_i 、 B_j の期待度数 E_{ij} は以下となる。

$$E_{ij} = P(A_i) \times P(B_j) \times N = a_i \times b_j / N$$

2 つの属性が独立に近ければ、 f_{ij} は E_{ij} に近く、 $(f_{ij} - E_{ij})^2$ は小さいはず。

従って、 $(f_{ij} - E_{ij})^2$ の合計が 小さければ独立に近い。

一般に統計量 $\chi^2 = \sum_{i=1}^R \sum_{j=1}^C \frac{(f_{ij} - E_{ij})^2}{E_{ij}}$ は自由度 $(R-1) \times (C-1)$ の

χ^2 分布に従う。

つまり χ^2 値が小さければ独立に近く、大きければ独立ではなくなる。

これを利用して、独立性の検定が行われる。上記の χ^2 値を計算し、

帰無仮説 : 2 属性は独立である (χ^2 値の有意確率 p 値が有意水準以上)

対立仮説 : 2 属性は独立でない (χ^2 値の有意確率 p 値が有意水準以下)
を検討する。