

- 独立性の検定 (1)
 - 分割表 (クロス表) (1)
 - 分割表とは (1)
 - 期待度数 (2)
 - 独立性の検定 (2×2より大きい表の場合: 自由度 $df > 1$) (3)
 - 帰無仮説と対立仮説 (3)
 - 検定統計量の算出 (3)
 - 仮説の判定 (両側検定) (4)
 - 独立性の検定 (2×2表の場合: 自由度 $df = 1$) (4)
 - 帰無仮説と対立仮説 (5)
 - 2×2分割表 (5)
 - 検定統計量の算出 (5)
 - 仮説の判定 (両側検定) (6)

独立性の検定

- 2つの変数に関連性があるか、つまり2つの変数の独立性を検定する。
- アンケートの結果の分析などに利用できる、基本的な手法のひとつ。

▲ ▼

分割表 (クロス表)

▲ ▼

分割表とは

- 観測された2つの変数 (要因と結果など) を組み合わせた表を、「分割表 (クロス表)」という
 - クロス集計表ともいう
 - Excelでは「ピボットテーブル」の機能で作ることができる
- k 行列の表からなる分割表を、「 $k \times l$ 分割表」という

	B_1	B_2	...	B_l	計
A_1	n_{11}	n_{12}	...	n_{1l}	$n_{1\cdot}$
A_2	n_{21}	n_{22}	...	n_{2l}	$n_{2\cdot}$
...	...	...	...	...	...
A_k	n_{k1}	n_{k2}	...	n_{kl}	$n_{k\cdot}$
計	$n_{\cdot 1}$	$n_{\cdot 2}$	...	$n_{\cdot l}$	n

- なお、周辺分布（右端の列や最下行の値）は、次のような意味になる。

- 標本数：

$$n_{\cdot j}$$

- 第 i 行の標本数：

$$n_{i\cdot}$$

：

- 第 j 列の標本数：

$$n_{\cdot j}$$

：

$$n_{i\cdot} = \sum_{j=1}^l n_{ij}$$

$$n_{\cdot j} = \sum_{i=1}^k n_{ij}$$

$$n = \sum_{i=1}^k \sum_{j=1}^l n_{ij}$$

▲ ▼

期待度数

- 分割表の各セルの期待値は、周辺分布の値から、次のように計算する。

- i 行 j 列のセルの期待値：

$$e_{ij}$$

$$e_{ij} = n \times \frac{n_{i.}}{n} \times \frac{n_{.j}}{n}$$

$$= \frac{n_{i.} n_{.j}}{n}$$

▲ ▼

独立性の検定（2×2より大きい表の場合：自由度 $df > 1$ ）

- 2行2列より大きい分割表の場合は、カイ二乗（

$$\chi^2$$

）分布を利用して検定する

▲ ▼

帰無仮説と対立仮説

2つの変数が独立であるか（関連がないか）を調べる。

- 帰無仮説

$$H_0$$

は「2つの変数は独立である（関連がない）」

- 対立仮説

$$H_1$$

は「2つの変数は独立ではない（関連がある）」

▲ ▼

検定統計量の算出

- 自由度

$$(k-1) \times (l-1)$$

のカイ二乗（

$$\chi^2$$

）分布にしたがう、検定統計量

$$\chi_0^2$$

を次の式から算出する

$$\chi_0^2 = \sum_{i=1}^k \sum_{j=1}^l \frac{(n_{ij} - e_{ij})^2}{e_{ij}}$$



仮説の判定（両側検定）

- 検定統計量

$$\chi_0^2$$

と、自由度

$$df = (k-1) \times (l-1)$$

、有意水準

$$\alpha$$

の有意点の値（カイ二乗分布表などから求める）を使って、判定をする

- 帰無仮説

$$H_0$$

を棄却：

$$|\chi_0^2| > \chi^2$$

- 「有意に差がある」「検定の結果、有意である」

- 帰無仮説

$$H_0$$

を採択：

$$|\chi_0^2| < \chi^2$$

- 「有意に差はない」「検定の結果、有意でない」「差があるとはいえない」



独立性の検定（2×2表の場合：自由度 $df=1$ ）

- 2行2列の分割表の場合は、直接確率を計算するか、カイ二乗（

$$\chi^2$$

）分布に近似した検定統計量で検定する

- フィッシャー(Fisher)の直接確率法

- 標本数が20未満、または標本数が40未満で最小期待値が5未満の場合

- イエーツ(Yates)の連続補正

- 標本数が40未満で、フィッシャーの直接確率法の条件を満たさない場合

- ここでは、Yatesの連続補正について説明する



帰無仮説と対立仮説

2つの変数が独立であるか（関連がないか）を調べるを調べる。

- 帰無仮説

$$H_0$$

は「2つの変数は独立である（関連がない）」

- 対立仮説

$$H_1$$

は「2つの変数は独立ではない（関連がある）」



2×2分割表

- 観測値による分割表を、次のようにあらわす

	教育後		計
	はい	いいえ	
はい	a	b	$a+b$
いいえ	c	d	$c+d$
計	$a+c$	$b+d$	$a+b+c+d=n$

- 期待値による分割表は、次のような表になる

	要因1	要因2	計
結果A	$(a+b) \times \frac{a+c}{n}$	$(a+b) \times \frac{b+d}{n}$	$a+b$
結果B	$(c+d) \times \frac{a+c}{n}$	$(c+d) \times \frac{b+d}{n}$	$c+d$
計	$a+c$	$b+d$	$a+b+c+d=n$



検定統計量の算出

- 2×2分割表では、次の式のような簡便な方法から、自由度

$$(2-1) \times (2-1) = 1$$

のカイ二乗（

$$\chi^2$$

) 分布にしたがう、検定統計量

$$\chi_0^2$$

を次の式から算出できる

$$\chi_o^2 = \frac{(ad-bc)^2 n}{(a+b)(c+d)(a+c)(b+d)}$$

- しかし、この方法では、計算した値が実際の

$$\chi^2$$

分布とずれてしまうことがわかっている

- 理由は、

$$\chi^2$$

分布は連続的にもかかわらず、計算した検定統計量は離散的だから

- そこで、Yatesの連続補正を使って、検定統計量を補正する
 - 原則として、2×2分割表ではYatesの連続補正を使うと考えてよい

$$\chi_o^2 = \frac{(|ad-bc| - \frac{n}{2})^2 n}{(a+b)(c+d)(a+c)(b+d)}$$

▲ ▼

仮説の判定（両側検定）

- 検定統計量

$$\chi_0^2$$

と、自由度

$$df = (k-1) \times (l-1)$$

、有意水準

$$\alpha$$

の有意点の値（カイ二乗分布表などから求める）を使って、判定をする

- 帰無仮説

$$H_0$$

を棄却：

$$|\chi_0^2| > \chi^2$$

- 「有意に差がある」「検定の結果、有意である」

○帰無仮説

$$H_0$$

を採択：

$$|x_0|^2 < \chi^2$$

- 「有意に差はない」「検定の結果、有意でない」「差があるとはいえない」