

第8章 検定

1. カイ自乗検定の方法

χ^2 検定は、カテゴリーデータによく用いられ、分布（特定の項目）が大きい、または小さいなどの差があると言えるか検定します。

その方法は、各調査項目（カテゴリー）が何個あるかの度数 F_i と理論的に期待される理論的度数との差を求め、これを理論的度数で割ります。各項目について計算し、全体の和がカイ自乗となります。

$$\chi^2 = \frac{\sum_{i=1}^r (F_i - NP)^2}{NP}$$

理論的に期待される度数は、全体の数を N 、カテゴリーの生起率を P とすると NP です。

たとえば、10円玉の表裏の確率 P は 0.5 ($\frac{1}{2}$) です。500回投げたとすると表の出る期待値は、

$$NP = 500 \times 0.5 = 250$$

になります。

		B	
		0	1
A	0	F_{00}	F_{01}
	1	F_{10}	F_{11}

		B	
		0	1
A	0	10	20
	1	20	50

$N=100$

各 F_i の期待値は、いくつになるでしょう。 P はいくつか考えて下

$N=200$ の時の期待値はいくつですか。

$$\begin{aligned} \chi^2 &= \frac{\sum_{i=0}^1 \sum_{j=0}^1 (F_{ij} - NP)^2}{NP} \\ &= \frac{(F_{00} - NP)^2}{NP} + \frac{(F_{01} - NP)^2}{NP} + \frac{(F_{10} - NP)^2}{NP} + \frac{(F_{11} - NP)^2}{NP} \\ &= \end{aligned}$$

(次の表を使って、計算してみよう)

前の事例を使って計算してみてください。

このときカイ自乗は、データが取り得る自由度があります。 $F=r-1$ です。これはカテゴリーの数から1を引いた数になります。なぜでしょうか。

このときのカテゴリー数は、Aが2です。Bも2です。自由度はカテゴリー数-1です。

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

になります。

これを使って、 x^2 の計算した値と自由度 f を使って、 x^2 の表から値を見出します。

※表は別紙

表の使い方と見出した数値の意味を考えましょう。

(注) まず、どの程度の有意となる Pr を求めるか考えて下さい。

自分の結論を $Pr=0.05$ としてみましょう。もし自由度が1であれば表を見ると1と $Pr=0.05$ のクロス点は3.841になっていますね。この値より x^2 の値が大きければ $Pr=0.05$ より Pr は小さいことになります。・・・ Pr が小さいほど有意の差が大きいと言えます。(より差がある。)

もし自由度が3で $Pr=0.05$ とすると7.814以上あれば $Pr=0.05$ より小さくなります。

(注) 自由度 f は、なぜ「カテゴリーの数-1」なのでしょうか。

10円玉の表・裏はカテゴリー2ですが、選べるのは1つです。では、問題の回答が①、②、③、④あったとすると、自由に選べる数は $4-1$ です。