生物科学研究所 研究報告 2024 年 12 月 25 日

比率の差: z 検定とカイ二乗検定は同等である

井口豊*
*生物科学研究所，長野県岡谷市

DOI: https://doi.org/10.5281/zenodo.14554454

1. はじめに

2 × 2 分割表に集計されたデータを考えるとき，母比率の差の検定として，カイ二乗検定と z 検定とがある。前者は 2 種類の属性の独立性を検定するノンパラメトリック検定であり，後者は文字通り 2 群（サンプル数 2）の比率（母比率）の差を検定するパラメトリック検定である。

これらは，検定の名称も異なり，目的も異なるように思えるが，実は，全く同等な検定である。しかし，そのような解説は，意外と少ない。

ここでは次の表 1 のように，属性 A と B が，それぞれ 2 群に分かれ，度数（データ数）が，a，b，c，d，全度数が n ($= a + b + c + d$) であると考える。これは，2 行 2 列（2 × 2）クロス集計表とも呼ばれる。

表 1

	A_1	A_2	合計
B_1	a	b	$a + b$
B_2	c	d	$c + d$
合計	$a + c$	$b + d$	n

例えば，A を性別（A_1: 男，A_2: 女），B を賛否（B_1: はい，B_2: いいえ）のように考えれば良い。

これから，カイ二乗検定と z 検定の計算式の同等性を説明していくが，連続データとするための，いわゆるイェーツ補正（Yate's continuity correction）は本質的問題ではないので，ここでは使わないことにする。

2. カイ二乗検定

表 1 のデータに対して，「帰無仮説 H_0: A と B は独立」を考え，独立性の検定としてカイ二乗検定を行なうと，その統計量（カイ二乗値）は，以下のように計算される。

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

ここで検定自由度は
(2 − 1) * (2 − 1) = 1
である。

3. z 検定

一方で，性別（A_1: 男， A_2: 女）で，はい（B_1）と答えた比率を考え，「帰無仮説 H_0: A_1 と A_2 における B_1 の母比率は等しい」として， z 検定を行なうと，その統計量（z 値）は，以下のように計算される。

$$z = \frac{p_1 - p_2}{\sqrt{p(1-p)\left(\frac{1}{a+c} + \frac{1}{b+d}\right)}}$$

ここで，z は標準正規分布に従う確率変数である。さらに， p_1, p_2, p は，以下のように定義される。

$$p_1 = \frac{a}{a+c}$$

$$p_2 = \frac{b}{b+d}$$

$$p = \frac{a+b}{a+b+c+d}$$

最後の p は統合比率（pooled proportion）と呼ばれる。なぜ，統合（プールすること， pooling）が必要なのかは，井口（2024）を参考にしてほしい。

統計学の定理より，標準正規分布に従う確率変数 z の二乗 z^2 は，自由度 1 のカイ二乗分布の確率変数でもある。したがって，前述の計算式を使って，次の等式を証明すれば，カイ二乗検定が z 検定と同等なものであることが言える。

$$\chi^2 = z^2$$

これは，次のようにも書き換えられる。

$$\chi^2 - z^2 = 0$$

式を変形していけば筆算でも可能（学生の試験向き）だが，ここでは，フリーソフトでもある数式処理システム Maxima を使って，これを証明してみる。

```
/* スクリプト開始 */

n1: a+c$
n2: b+d$

n: n1+n2$

/* カイ二乗値 */
chisq: n*(a*d-b*c)^2/((a+b)*(c+d)*(a+c)*(b+d))$

p1: a/n1$
p2: b/n2$

p: (a+b)/n$

/* z 値 */
z: (p1-p2)/sqrt(p*(1-p)*(1/n1+1/n2))$

/* カイ二乗値 - z 値の二乗*/
ratsimp(chisq-z^2);

/* スクリプト終了 */
```

結果は，以下の図 1 のとおり， 0 となり， 2 × 2 クロス集計データに対して，
カイ二乗値 = z 値の二乗
であることが証明された。

```
(%i9)  n1: a+c$
       n2: b+d$
       n: n1+n2$
       /* カイ二乗値 */
       chisq: n·(a·d－b·c)^2/((a+b)·(c+d)·(a+c)·(b+d))$
       p1: a/n1$
       p2: b/n2$
       p: (a+b)/n$
       /* z 値 */
       z: (p1－p2)/sqrt(p·(1－p)·(1/n1+1/n2))$
       /* カイ二乗値 - z 値の二乗*/
       ratsimp(chisq－z^2);
(%o9)  0
```

図 1. Maxima によるカイ二乗値と z 値の差

ノンパラメトリック検定であるカイ二乗検定とパラメトリック検定である z 検定が全く同じ式であることにも注意したい。母数（パラメータ）に注目するかどうかで呼称が変わるケースとも言える。

参考文献

井口豊（2024）比率の差 *Z* 検定の注意点：統合比率を使う理由. 生物科学研究所研究報告 2024 年 11 月 11 日. https://doi.org/10.5281/zenodo.14064942