

## 対応を持つ2つの群の平均値の検定（等分散を前提）

対応を持つ2つの群のデータを

第1群  $\{x_i\}_{i=1}^N$ 、第2群  $\{y_i\}_{i=1}^N$

とした時、2つの群の平均値が等しいかどうかを検定する。

対応を持つとは、例えば、ダイエット薬の効果を検査する為に、N人の被験者の

- ・ 服用前の体重データを  $\{x_i\}$
- ・ 服用後の体重データを  $\{y_i\}$

として同一対象の変化を調査するときのデータの対応のことである。

但し 以下を前提とする。

- (1) 2つの群のデータは正規分布に従う。
- (2) 2つの群のデータの分散は等しいとみなす。(等分散性を仮定)

$\mu_x$  を第1群の母平均、 $\mu_y$  を第2群の母平均とする。

更に 差のデータを  $d_i = x_i - y_i$  とする。

差のデータの母平均を  $\mu_d$  とする。(  $\mu_d$  は未知 )

また  $\bar{d} = \frac{1}{N} \sum_{i=1}^N d_i$  ,  $U_d^2 = \frac{1}{N-1} \sum_{i=1}^N (d_i - \bar{d})^2$  として

$$T = \frac{\frac{\bar{d} - \mu_d}{\frac{U_d}{\sqrt{N}}}}{\frac{U_d}{\sqrt{N}}} = \frac{\bar{d} - \mu_d}{U_d} \sqrt{N} \quad \text{で計算される } T \text{ は 自由度 } N-1 \text{ の } t \text{ 分布に従う。}$$

検定についての考え方には3通りあり、帰無仮説はいずれも

帰無仮説 : 2群間の平均値には差がない。つまり  $\mu_x = \mu_y$

であるが、対立仮説は

対立仮説(1) : 2群間の平均値には差がないとはいえない(両側検定)

対立仮説(2) :  $\mu_x < \mu_y$  (左片側検定)

対立仮説(3) :  $\mu_x > \mu_y$  (右片側検定)

の3通りとなる。

ここでは  $\mu_d = 0$  を仮定するので、実際には 以下を計算し

$$T = \frac{\bar{d}}{\frac{U_d}{\sqrt{N}}} = \frac{\bar{d}}{U_d} \sqrt{N}$$

Tが自由度 N-1 の T分布の棄却域に入る場合帰無仮説を棄却し、そうでない場合帰無仮説を採択する。

なお、サンプルサイズ N が大きいと、統計量 Tの式の一部の  $\sqrt{N}$  が大きくなり、Tの絶対値が大きくなり、p値が小さくなり、帰無仮説が棄却されることが多くなる。

その為、サンプルサイズの大小に影響されにくい 効果量  $d$  を同時に表示する。

$$\text{効果量 } d = \frac{\bar{d}}{\sqrt{U_x^2 + U_y^2 - 2U_{xy}}}$$

$$\text{ただし } \bar{X} = \frac{\sum_{i=1}^N x_i}{N}, \bar{Y} = \frac{\sum_{i=1}^N y_i}{N}, U_x^2 = \frac{\sum_{i=1}^N (x_i - \bar{X})^2}{N-1}, U_y^2 = \frac{\sum_{i=1}^N (y_i - \bar{Y})^2}{N-1}, U_{xy} = \frac{\sum_{i=1}^N (x_i - \bar{X})(y_i - \bar{Y})}{N-1}$$