

## 実務と統計(5)

### —対応ある2群の平均値比較 (paired t検定)—

山内俊平<sup>†</sup> (豊橋市立看護専門学校 非常勤講師,  
豊橋市食肉衛生検査所 元所長)



今回は、対応のある2群間の平均値の比較 (paired t検定) について、EZRを用いて統計的結論を導く練習をしてみます。

**例：paired t検定 (主解析)**  
〈研究目的〉ある筋トレメニューを実施すると、体温が変化するのかを調べる。

なお、この設定における効果量 (Cohenのd) は、 $d=0.5/0.3 \approx 1.67$  となります。非常に大きな効果量を仮定したことになり、本調査は探索的研究としての位置づけとなります。

#### (結果)

paired t検定に必要なサンプルサイズは  $n=6$  と出力されます。効果量を大きく設定しているため、 $n=6$  の小標本でも検出可能と判断されたこととなります。

#### 〈主解析の条件設定〉

対応のある2群間の平均値の比較 (paired t検定)  
有意水準： $\alpha_{エラー} = 0.05$  (両側, 各側 0.025)  
検出力： $1 - \beta_{エラー} = 0.8$

#### 〈データ収集 (仮想データ)〉

6名の被験者を得て、筋トレ前後の体温を測定した結果、以下であったとします。2群のデータをEZRに取り込みます (実務と統計(4)参照)。

筋トレ前	筋トレ後
36.1	36.5
36.2	36.9
36.5	36.8
36.5	37.1
36.8	37.2
37.2	37.5

#### 〈サンプルサイズ (n) の計算〉

##### 統計解析

- 必要サンプルサイズの計算
- 対応のある2群の平均値の比較のためのサンプルサイズの計算

多くの被験者を得ることは難しそうです。そこで、大きい効果量で  $n$  を設計することとします。筋トレ前後の体温差の平均を  $0.5^{\circ}\text{C}$  とし、類似の先行研究を参考に筋トレ前後の体温差の標準偏差 (SD) を  $0.3^{\circ}\text{C}$  で仮定します。EZRによる「対応のある2群の平均値の比較のためのサンプルサイズの計算」では、差の標準偏差を「2群共通の標準偏差 (SD)」として入力する仕様であるため、この値を同欄に入力します。

#### 主解析 (paired t検定)

##### 統計解析

- 連続変数の解析
- 対応のある2群間の平均値の比較 (paired t検定)

対応のある2群の平均値の比較のためのサンプルサイズの計算	
2群間の平均値の差	0.5
2群共通の標準偏差 (SD)	0.3
$\alpha$ エラー (0.0-1.0)	0.05
検出力 ( $1 - \beta$ エラー) (0.0-1.0)	0.8
方法	
<input checked="" type="radio"/> 両側	
<input type="radio"/> one-sided	

対応のある2群間の平均値の比較 (paired t検定)	
第1の変数 (1つ選択)	第2の変数 (1つ選択)
筋トレ後 筋トレ前	筋トレ後 筋トレ前
対立仮説	
<input checked="" type="radio"/> 両側	
<input type="radio"/> 差 < 0	
<input type="radio"/> 差 > 0	
信頼水準	
0.95	

<sup>†</sup> 連絡責任者：山内俊平 (豊橋市立看護専門学校 非常勤講師, 豊橋市食肉衛生検査所 元所長)

## 仮説

帰無仮説：筋トレ前後の体温（平均）に差はない  
対立仮説：筋トレ前後の体温（平均）に差がある  
帰無仮説の棄却条件： $P < 0.05$

## 検定結果

summary.ttest	平均	標準偏差	P 値
筋トレ前	36.55	0.4037326	0.00111
筋トレ後	37.00	0.3464102	

## 結果解釈

$P < 0.05$  より帰無仮説は棄却されます。「筋トレ後に体温が有意に上昇した」と結論づけられます。

なお、因果関係を分析したわけではないので、「筋トレにより体温が上昇した」と書くと、「書きすぎ」となるので注意が必要です。

## 感度分析（Wilcoxon 符号付順位検定）

### （感度分析を行う理由）

この研究設計ではサンプルサイズが  $n=6$  と極めて小さいため、正規性の検定や分布診断の信頼性が低い（診断困難）という制約があります。そのため、分布仮定に依存しないノンパラメトリック検定を感度分析として併用し、主解析の結果が分布仮定に強く依存していないかを確認することとします。

## 統計解析

→ノンパラメトリック検定

→対応のある 2 群間の比較（Wilcoxon 符号付順位検定）

対応のある 2 群間の比較（Wilcoxon 符号付順位検定）	
第 1 の変数（1 つ選択）	第 2 の変数（1 つ選択）
<input type="checkbox"/> 筋トレ後 <input checked="" type="checkbox"/> 筋トレ前	<input type="checkbox"/> 筋トレ後 <input checked="" type="checkbox"/> 筋トレ前
対立仮説	
<input checked="" type="radio"/> 両側	
<input type="radio"/> 差 $< 0$	
<input type="radio"/> 差 $> 0$	
検定のタイプ	
<input checked="" type="radio"/> デフォルト	
<input type="radio"/> 正規	
<input type="radio"/> 正規近似	
<input type="radio"/> 連続修正を用いた正規近似	

## 仮説

帰無仮説：筋トレ前後の体温（中央値）に差はない  
対立仮説：筋トレ前後の体温（中央値）に差がある  
帰無仮説の棄却条件： $P < 0.05$

## 検定結果

対応のある 2 群間の比較（Wilcoxon 符号付順位検定）
P 値=0.0355

## 結果解釈

$P < 0.05$  より帰無仮説は棄却されます（有意）。

## 結果の統合的解釈（頑健性の確認）

- ・主解析（paired t 検定）→  $P = 0.00111$
  - ・感度分析（Wilcoxon 符号付順位検定）→  $P = 0.0355$
- 両検定ともに P 値が有意水準の 5% を下回っており、解析手法や分布仮定が異なっても結果（統計的に有意）は一致しています。

## ま と め

本例では、実行可能性（被験者を得にくい）を考慮して、大きい効果量を前提に小標本での研究設計（対応のある 2 群間の平均差の比較）を行い、主解析として paired t 検定を用いました。同検定は、対応のある群間の差の分布が概ね正規であることを前提とするパラメトリック検定ですが、本例のように極端に小さい  $n$  では、正規性の診断が困難です。そこで、分布仮定の影響を受けにくい Wilcoxon 符号付順位検定（ノンパラメトリック検定）を感度分析として併用しました。

その結果、いずれも統計的に有意な結果が得られたため、統計的結論（筋トレ後に体温が上昇）は比較的頑健であると判断できます。

なお、仮にパラメトリック検定とノンパラメトリック検定で結論が一致しなかった場合には、結果が分布仮定や外れ値に敏感である可能性が示唆されます。解釈には、より慎重な姿勢が求められます。