

実務と統計(3)

—検定方法の選択—

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)



1 はじめに

実務的な観点から行う差の検定(有意差検定)は、一般に次の①～⑤の手順で進められます。

- ①検出したい差(実務価値のある差)の定義
- ②検出条件($\alpha_{エラー}$, 検出力)の設定

- ③必要サンプルサイズ(n)の算定(決定したnで無作為抽出)
- ④検定方法の選択
- ⑤結果解釈

⑤の結果解釈では、P値と信頼区間が重要となること(実務と統計(1))、①～③では、 $\alpha_{エラー}=0.05$ (有意水準5%)、検出力0.8程度を標準とするnの求め方(実務と統計(2))を述べたところです。

今回は、④の「検定方法の選択」について、連続変数の差の検定(平均差のt検定など)を例に、基本的考え方を整理します。

2 連続変数の差の検定方法

検定方法は、サンプルサイズ(n)の大きさやデータの正規性などから統計的手法に基づいて選択する場合と、意思決定上のリスクを重視して選択する場合に大別できます。

(1) 統計的手法からの検定選択

例えばt検定は、母集団が正規分布に従うことを前提としたパラメトリック検定です。その妥当性は、母分布の正規性の程度に依存します。ヒストグラムやQQプロットで分布を視覚的に確認し、正規性の検定も参考にして判断します。

〈正規性の検定(例)〉

フリー統計ソフトEZRをダウンロードして、次の例を検定してみます。

例：A, Bの2群から無作為に抽出したサンプルデータ(n=15)が次のとおりであったとき、各々の母集団の正規性を調べる(正規性の検定)。

EZRへのデータの取り込み

A群	B群
35	13
63	39
25	17
54	51
21	22
15	20
75	15
40	18
16	35
48	36
72	29
46	24
32	42
59	75
22	23

統計解析

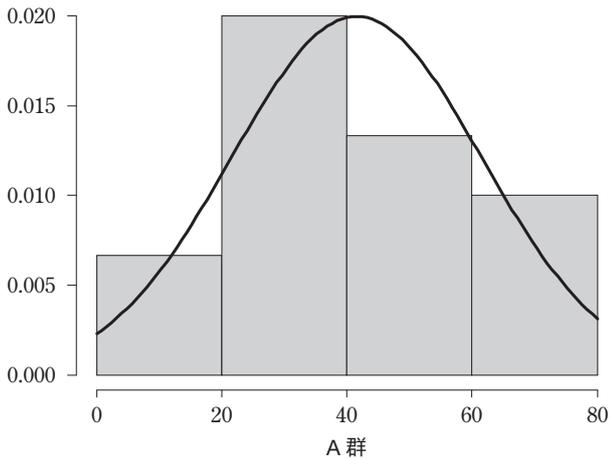
- 連続変数の解析
- 正規性の検定(Kolmogorov-Smimov検定)

検定条件

- 帰無仮説：データの母集団は正規分布に従う
- 対立仮説：データの母集団は正規分布に従わない
- 帰無仮説棄却： $P < 0.05$
- *nが5,000以下の場合、シャピロウィルクのP値を確認します。

[†] 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)

検定結果 (A 群)



Shapiro-Wilk normality test
 data: Dataset\$B 群
 W = 0.86727, p-value = 0.03077

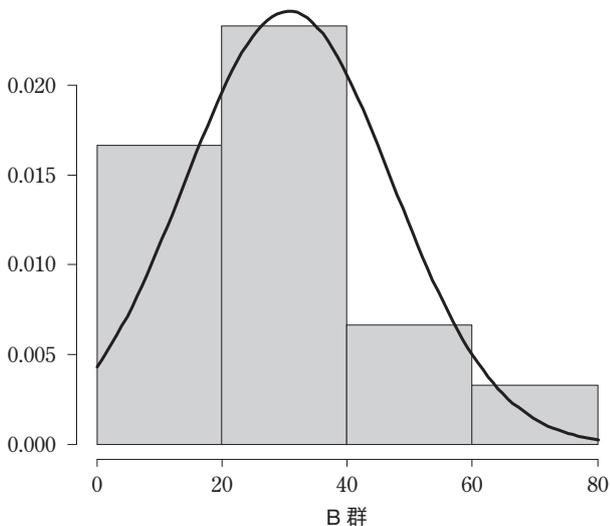
P=0.03077 (P<0.05) より帰無仮説は棄却され、「母集団は正規分布に従わない (非正規)」となります。この場合は、ヒストグラムや QQ プロットから「外れ値」がないかなどを確認した上で、母分布に特定分布を仮定しない「ノンパラメトリック検定 (Mann-Whitney の U 検定など)」の選択を考慮します。

Exact one-sample Kolmogorov-Smirnov test
 data: Dataset\$A 群
 D = 0.12913, p-value = 0.9363
 alternative hypothesis: two-sided

> # サンプル数が 5000 以下の場合のみ Shapiro-Wilk 検定の結果も表示されます。(サンプル数=15)
 > shapiro.test (Dataset\$A 群)
 Shapiro-Wilk normality test
 data: Dataset\$A 群
 W = 0.9427, p-value = 0.4175

P 値=0.4175 (P ≥ 0.05) より帰無仮説は棄却されません。「得られたデータの母集団は非正規 (正規分布に従わない) とは言えない」となります。この場合は、一般的に、母分布が正規分布に従うとみなして「パラメトリック検定 (t 検定など)」を選択します。

検定結果 (B 群)



正規性の検定結果 (シャピローウィルク検定)

群	P 値	判定	解釈	検定法の選択方向
A 群	0.4175 (P ≥ 0.05)	帰無仮説は棄却されない	母集団は非正規とはいえない (正規性ありとみなす)	パラメトリック検定 (t 検定など) を適用
B 群	0.03077 (P < 0.05)	帰無仮説は棄却される	母集団は正規分布に従わない (非正規)	ノンパラメトリック検定 (Mann-Whitney の U 検定など) を検討

〈正規性検定の注意点〉

検定結果は、データサイズ (n) が大きい場合は、わずかなずれでも「非正規」(帰無仮説棄却)、逆に n が小さいと、ずれが比較的大きくても「非正規とはいえない」(帰無仮説非棄却) と判定される場合がでてきます。したがって、差の検定方法を選択する目的で正規性の検定を行う場合は、概ね n=10~30 程度を目安とし、外れ値の影響にも注意しながら、ヒストグラムや QQ プロットによる視覚的な分布確認を併用することが重要です。

なお、n が比較的大きい場合 (概ね n ≥ 30 程度) は、中心極限定理により、母分布の形状にかかわらず平均の分布が正規分布に近づくため、パラメトリック検定 (t 検定など) を適用しても大きな問題は生じにくいと考えられます。

一方、n が特に小さい場合 (概ね n < 10 程度) は、情報量が少なく正規性の診断が困難です。パラメトリック検定とノンパラメトリック検定 (Mann-Whitney の U 検定など) を併用し、総合的に判断することが望まれます。両者の結果に大きな乖離がなければ、結論の妥当性が高いと判断できます。

(2) 意思決定上のリスクからの検定選択

一般に、正規性の前提が妥当な場合のパラメトリック

検定は $\beta_{\text{エラー}}$ （差の見逃し）を起こしにくい傾向があり、逆に、ノンパラメトリック検定では $\alpha_{\text{エラー}}$ （差の誤検出）を起こしにくい反面、 $\beta_{\text{エラー}}$ が大きくなりやすい傾向があります。

「誤検定に伴う意思決定が及ぼす不利益の方向と重要性」を考慮して検定方法を選択するのも、合理的な考え方といえます。比喩的にいうと、疑わしきは罰せず、という刑事訴訟法の理念に照らせば「冤罪（ $\alpha_{\text{エラー}}$ ）」の回避が重要であり、ノンパラメトリック検定を選択する、一方、医薬品の安全性評価のように「副作用の見逃し（ $\beta_{\text{エラー}}$ ）」を小さくすることが重視される場合や、食中毒（疑い）事案での「食中毒の見逃し（ $\beta_{\text{エラー}}$ ）」による消費者危険の回避を優先すべき場面では、パラメトリック検定を選択する、といった考え方です。

3 ま と め

連続変数の差の解析では、中心極限定理、ヒストグラ

ム、QQプロット、正規性検定といった統計的手法を用いて母分布の特性を把握し、適切な検定方法（パラメトリック検定またはノンパラメトリック検定）を選択します。場合によっては、両者を併用して結果の頑健性を確認することも有用です。

一方、意思決定上のリスクを重視して検定方法を選択する考え方もあります。 $\alpha_{\text{エラー}}$ の回避を優先する場合はノンパラメトリック検定、 $\beta_{\text{エラー}}$ の回避が重要ならパラメトリック検定を選ぶ、といった判断です。業務内容によっては、こちらの考え方の方が自然な判断基準となる場合もあります。特に公権力の行使や社会的影響が大きいケースでは、その重要性はいっそう高まります。

いずれにしても、単に「t検定で平均差を調べた」とだけ記載するのでは、検定選択の恣意性を疑われかねません。その検定を採用した理由を明確に示すことが重要です。報告書などでは、検定方法の選択理由を明記しておくことが望まれます。