

実務と統計(7)

—母平均推定と1標本t検定におけるnの設計—

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)



サンプルサイズの決め方(実務と統計(2)78巻12号)で差の検定におけるn設計の基本的考え方を、(4)(5)(79巻の2号と3号)で2群平均差のt検定におけるnの算定方法について述べました。

ここでは、母平均の推定と検定(1標本t検定)におけるnの設計について述べます。

例：A町では、40代男性の生活習慣病対策の一環として最高血圧130 mmHgを基準(基準値 $\mu_0=130$ mmHg)に健康づくりを進めています。

この町に住む40代男性の血圧(母平均 μ)について、以下を調べることにします。

- ① 母平均 μ の推定
- ② μ の基準値($\mu_0=130$ mmHg)との差の検定(1標本t検定, 両側)

必要サンプルサイズの設計

本研究の目的は、母平均 μ の①推定と②基準値との差の検定、の両方です。そのため①②のそれぞれで必要サンプルサイズ(n)を算定し、最終的に大きい方を採用します。

① μ の推定におけるn設計(信頼区間重視)

母平均 μ をどの程度の精度(誤差の程度)で推定したいかを考えます。ここでは、母集団の標準偏差を10 mmHg(先行研究などから想定)、母平均の95%信頼区間を ± 5 mmHgに設定することとします。

これは、母平均の95%信頼区間が125～135 mmHg程度の幅(誤差 ± 5 mmHg)に収まる精度で推定したい、という意味になります。

[nの計算]

EZRで計算してみます。

統計解析

→必要サンプルサイズの計算

→1群の平均値の信頼区間をある幅におさめるためのサンプルサイズの計算

1群の平均値の信頼区間をある幅におさめるためのサンプルサイズの計算

想定する標準偏差	10
信頼区間の幅(上限と下限の差)	10
信頼水準	95

EZRでは、信頼区間の入力とは半幅ではなく全幅(上限－下限)を入力する仕様です。今回の設定は ± 5 mmHgなので信頼区間の全幅は10 mmHg、これを入力します。 $n=16$ の結果が得られます。

必要サンプルサイズ

計算結果

16

推定精度とサンプルサイズ(n)の関係

精密な推定(信頼区間が狭い)を行うには、大きなnが必要になります。例えば、信頼区間を ± 1 mmHg(信頼区間全幅2 mmHg)に設定すると $n=385$ で、このnでは研究実施が困難かも知れません。推定精度と実行可能性のバランスを考えて研究設計(n決定)する必要があります。

なお、標準偏差もnに関係します。先行研究調査などから妥当な標準偏差を想定することも重要となります。

② 差の検定におけるn設計(検出力重視)

次に、基準値との差の検定を考えます。ここでは1標本t検定(両側検定)を行うとします。検出したい差(実務的に重要な差)を10 mmHgとします。

[†] 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)

〈設定条件〉

基準値： $\mu_0 = 130$ mmHg

標準偏差：10 mmHg (①推定と同じ想定)

検出したい差 (実務的に重要な差)：10 mmHg

効果量 (one-sample Cohen's d) を計算すると、 $d = \text{検出したい差} / \text{標準偏差} = 1.0$ です。

Cohen の d の目安、0.2 (小)、0.5 (中)、0.8 (大) に照らすと、 $d = 1.0$ は、かなり大きな効果量といえます。

〈検出条件〉

$\alpha_{\text{エラー}} = 0.05$

検出力 $1 - \beta_{\text{エラー}} = 0.8$

通常は α 、 β で表記されますが、本資料中では「検定エラーの確率」であることを明確にするため、エラーを省略せず表記しています。

検出力 $1 - \beta_{\text{エラー}} = 0.8$ とは、「真の平均血圧が基準値 130 mmHg より 10 mmHg 高い (または低い) と仮定した場合に、その差を 80% の確率で検出でき、20% の確率で検出できない ($\beta_{\text{エラー}} = 0.2$)」を意味します。

一方、 $\alpha_{\text{エラー}} = 0.05$ は、「真の平均血圧が基準値と差がない場合に、差があると誤って判定する確率 (差の誤検出確率)」が 5% であることを意味します。

これらの設定条件と検出条件を満たすサンプルサイズ (n) を計算します。

〔n の計算〕

n の設計では計算を簡単にするため、通常は正規分布 (Z 値) を用いて近似計算します。実際の解析では t 検定を用いますが、一般には、この近似で計算した n で実用上は支障がないと考えていいかと思います。

〈計算方法〉

$\alpha_{\text{エラー}} = 0.05$ (両側) なので片側は 0.025、正規分布表からこれに相当する Z 値 = 1.96、検出力 $1 - \beta_{\text{エラー}} = 0.8$ なので $\beta_{\text{エラー}} = 0.2$ 、相当する Z 値 = 0.84 です。

$n = ((1.96 + 0.84) \times 10 / 10)^2 = 7.84$ 、切り上げて $n = 8$ です。効果量 ($d = 1.0$) が大きいので、比較的小さな n での差の検出が可能です。

なお、EZR には 1 標本 t 検定における n 計算が標準ではないので、他ソフトや AI で計算すると、同様に $n = 8$ (ただし、t 分布を用いた厳密計算仕様だと n は 9 ~ 10 程度) が出力されます。

サンプルサイズ n の最終決定

①推定 → $n = 16$ 、②検定 → $n = 8$ でした。より大きい方のサンプルサイズ (① $n = 16$) を採用します。

$n = 16$ を用いることで、① μ の推定 (母平均血圧を

± 5 mmHg の精度で推定する)、②差の検定 (基準値 130 mmHg との差 10 mmHg を 80% 確率で検出する) の両方を満たす研究計画となります。

データの収集 (仮想データ)

被験者 16 人 ($n = 16$) の血圧測定の結果が、次のとおりであったとします。EZR にデータを取り込みます (実務と統計 (4) 79 巻 2 号参照)。

最高血圧 (mmHg)
132
122
118
128
133
132
112
112
126
120
115
142
146
130
135
109

データから母分布の正規性を確認します。t 検定は母集団がおおむね正規分布 (大きなズレはない) であることを前提とします。ただし、サンプルサイズが小さい場合、正規性検定だけで判断するのは難しいため、ヒストグラムや外れ値の有無もあわせて確認します。

統計解析

→ 連続変数の解析

→ 正規性の検定 (Kolmogorov-Smirnov 検定)

結 果

n が小さい場合は K-S 検定より Shapiro-Wilk 検定を参照します。P 値 = 0.7479 で帰無仮説非棄却 (実務と統計 (3) 79 巻 1 号参照) で、ヒストグラムにも強い歪みを認めず、正規分布から大きく逸脱しているとはいえない。また、顕著な外れ値も見られないので、t 検定の適用に大きな支障はないと判断されます。

1 標本の平均値の t 検定

統計解析

→ 連続変数の解析

→ 1 標本の平均値の t 検定

仮 説

帰無仮説：母平均 $\mu = 130$ mmHg

対立仮説：母平均 $\mu \neq 130$ mmHg

帰無仮説の棄却条件： $P < 0.05$ （両側検定）

1 標本の t 検定	
変数（1 つ選択）	
最高血圧 .mmHg.	
対立仮説	
☒ 母平均 $\mu \neq \mu_0$	
☐ 母平均 $\mu < \mu_0$	
☐ 母平均 $\mu > \mu_0$	
帰無仮説： $\mu = \mu_0$: $\mu_0 =$ 130	
信頼水準： 0.95	

結 果

平均=125.75, 95%信頼区間 119.9209-131.5791, P 値=0.141

サンプル平均は 125.75 mmHg, 母平均の 95%信頼区間は 120.0 ~ 131.6 mmHg で, 基準値の 130 mmHg を

含んでいます。1 標本 t 検定の P 値は $P = 0.141$ で帰無仮説は棄却されません（非有意）。

結 論

A 町の 40 代男性の平均血圧の 95%信頼区間は 120.0 ~ 131.6 mmHg と推定された。1 標本 t 検定（両側）の結果, 母平均が 130 mmHg と異なるという, 統計学的有意差は認められなかった。ただし, これは 130 mmHg と差がないことを証明するものではなく, 今回のデータからは差があるという十分な証拠が得られなかったことを意味します。

ま と め

サンプルサイズ (n) の大きさは, 母平均 (μ) の推定や検定の結果に影響します。“なぜ, その n なのか”, が問われます。研究目的を明確にし, 最適 n を設計することが推奨されます。n 設計の意義を一言でいうなら, 労力対効果の最適化です。

μ の推定研究では, 信頼区間重視で n を算定します。“どの程度の推定精度が得られればいいのか”, 信頼区間の範囲を決めて必要 n を計算します。差の検定では, 検出力を中心に組み立てます。“どの程度の差をどの程度の確率で見つけるのか”, を決めて n を計算します。

実務では, 推定と検定の両方を研究目的とする場合もあります。この場合は, 大きい方の n を採用することで, 両研究の目的を満たせることになります。