

実務と統計(2)

—サンプルサイズの決め方—

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

10 検体を採取，100 人測定など任意にサンプルサイズ (n) を決めて有意差検定を行うケースがみられます。「n の大きさ」と「得られた差の大きさ」が検定結果に影響することから、「n が小さいと，差は大きいのに統計的に非有意，n が大きいと，差は小さいのに有意」となるケースがでてきます。

実務的には，「差があるのか？」ではなく，「どれほどの差があるのか？」が重要になる場合が多いと思います。あらかじめ検定対象とする「実務価値のある差」を定義し，この差を検出（統計有意）するのに必要なサンプルサイズを決める必要があります。

2 $\alpha_{\text{エラー}}$ ， $\beta_{\text{エラー}}$ ，検出力 ($1 - \beta_{\text{エラー}}$)

有意差検定は，差があるか否かを統計確率的に推論（判断）します。確率的推論である以上，判断ミス ($\alpha_{\text{エラー}}$ ， $\beta_{\text{エラー}}$) を犯す可能性（確率）が生じます。通常，あらゆる判断において，判断ミスは起こり得ます。

例 1：司法判断

$\alpha_{\text{エラー}}$ ：真実は犯人ではない→犯人（有罪）と判断
 $\beta_{\text{エラー}}$ ：真実は犯人→犯人ではない（無罪）と判断

例 2：行政判断

$\alpha_{\text{エラー}}$ ：真実は食中毒ではない→食中毒と判断
 $\beta_{\text{エラー}}$ ：真実は食中毒→食中毒ではないと判断

例 3：統計判断

$\alpha_{\text{エラー}}$ ：帰無仮説（差はない）が真→
帰無仮説を棄却（差がある）と判断
 $\beta_{\text{エラー}}$ ：対立仮説（差がある）が真→
帰無仮説を棄却しない（差があるとはいえない）と判断

統計検定における $\alpha_{\text{エラー}}$ は差の誤検出， $\beta_{\text{エラー}}$ は差の見逃しです。 $1 - \beta_{\text{エラー}}$ は差があるときの差の検出で，これを検出力 ($1 - \beta_{\text{エラー}}$) といいます（表）。

表 統計用語，内容及び意味

用語	内容	意味するところ
$\alpha_{\text{エラー}}$	差がないのに「差がある」と判断	差の誤検出（第一種の過誤）
$\beta_{\text{エラー}}$	差があるのに「差がない」と判断	差の見逃し（第二種の過誤）
検出力 ($1 - \beta_{\text{エラー}}$)	実際に差があるときに「差がある」と検出できる確率	確率が高いほど検出力が高い

なお， $\alpha_{\text{エラー}}$ の許容範囲を定めたのが，いわゆる有意水準で，有意水準 5% は，差の誤検出率 5% ($\alpha_{\text{エラー}} = 0.05$) を容認するという意味になります。

3 サンプルサイズ (n) の計算

「検出したい差の大きさ」と「検出条件 ($\alpha_{\text{エラー}}$ ，検出力 ($1 - \beta_{\text{エラー}}$))」を定めて n を求めます。ここではフリー統計ソフト EZR をダウンロードして，検定に必要な n を計算してみます。

例 1：既存薬の有効率が 50%（既知）である。新薬の有効率が 70%（比率差 +0.2）あれば，導入を検討したいとする。比率差 0.2 を片側検定するのに必要なサンプルサイズ (n) を求める。

統計解析より

→必要サンプルサイズの計算

→1 群の比率を既知の比率と比較するためのサンプルサイズの計算

比率の差 0.2 (0.5 と 0.7)， $\alpha_{\text{エラー}} = 0.01$ ，検出力 ($1 - \beta_{\text{エラー}}$) = 0.99 で条件設定すると，n = 125 が出力されます。

[†] 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）

1 群の比率を既知の比率と比較するためのサンプルサイズの計算

既知の比率 (0.0-1.0)	0.5
想定する比率 (0.0-1.0)	0.7
α エラー (0.0-1.0)	0.01
検出力 (1- β エラー) (0.0-1.0)	0.99
グループ 1 と 2 のサンプルサイズ比	1

方 法

- ☐ 両側
☒ one-sided

カイ 2 乗検定の連続性補正

- ☐ はい
☒ いいえ

SampleProportionSingleArm(0.5, 0.7, 0.05, 0.8, 1, 0)

	仮定
想定する比率	0.5
比較する比率	0.7
α エラー	0.05
	片側検定
検出力	0.8
必要サンプルサイズ	計算結果 37

例 2 : A 薬と B 薬の有効率に 20% の差 (比率差 0.2) があるのかを, α エラー=0.05, 検出力 0.8 で両側検定するのに必要な n を求める.

統計解析

→ 必要サンプルサイズの計算

→ 2 群の比率の比較のためのサンプルサイズの計算

SampleProportionSingleArm(0.5, 0.7, 0.01, 0.99, 1, 0)

	仮定
想定する比率	0.5
比較する比率	0.7
α エラー	0.01

	片側検定
検出力	0.99

必要サンプルサイズ	計算結果 125
-----------	-------------

このように, 誤判定を減らすために, α エラーを小さく (0.01), 検出力を高く (0.99) 設定しすぎると, 多くのデータ (n) が必要となり, 実行するのが難しくなってきます. そこで通常は, α エラー=0.05, 検出力を 0.8 前後で標準設定します. EZR では α エラー=0.05, 検出力 0.8 がデフォルト設定されているので, これに従い計算すると, n=37 となります.

1 群の比率を既知の比率と比較するためのサンプルサイズの計算

既知の比率 (0.0-1.0)	0.5
想定する比率 (0.0-1.0)	0.7
α エラー (0.0-1.0)	0.05
検出力 (1- β エラー) (0.0-1.0)	0.8
グループ 1 と 2 のサンプルサイズ比	1

方 法

- ☐ 両側
☒ one-sided

カイ 2 乗検定の連続性補正

- ☐ はい
☒ いいえ

2 群の比率の比較のためのサンプルサイズの計算

グループ 1 の比率 (0.0-1.0)	0.5
グループ 2 の比率 (0.0-1.0)	0.3
α エラー (0.0-1.0)	0.05
検出力 (1- β エラー) (0.0-1.0)	0.8
グループ 1 と 2 のサンプルサイズ比	1

方 法

- ☒ 両側
☐ one-sided

カイ 2 乗検定の連続性補正

- ☐ はい (あるいは Fisher 正確性検定)
☒ いいえ

SampleProportion (0.5, 0.3, 0.05, 0.8, 2, 1, 0)

	仮定
P1	0.5
P2	0.3
α エラー	0.05

	両側検定
検出力	0.8
N2 と N1 のサンプルサイズの比	1

必要サンプルサイズ	計算結果
N1	93
N2	93

〈グループ 1 と 2 の比率差 0.2 の設定方法〉

類似文献や先行研究があれば参考にし, ない場合は最も保守的な値 (比率 (P) = 0.5) を基準にするのが一般的

です。グループ1と2の比率をそれぞれ0.5と0.3（または0.5と0.7）と置いて、比率差0.2を設定します。P=0.5のときに分散（ $P(1-P)$ ）が最大で、最も多くのサンプルが必要になる前提となり、サンプルnが足りないリスクを最小化できます。

例3：2群の平均血圧に5 mmHgの差があるのかを、 $\alpha_{\text{エラー}}=0.05$ 、検出力0.8で両側検定するのに必要なnを求める。

統計解析

→必要サンプルサイズの計算

→2群の平均値の比較のためのサンプルサイズの計算

2群共通の標準偏差（SD）は、類似文献や先行研究の値などを参考に設定します。ここではSD=6として計算してみます。

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	5
2群共通の標準偏差（SD）	6
α エラー（0.0–1.0）	0.05
検出力（ $1-\beta$ エラー）（0.0–1.0）	0.8
グループ1と2のサンプルサイズ比	1

方 法

- ☒ 両側
☐ one-sided

SampleMean (5, 6, 0.05, 0.8, 2, 1)

	仮定
2群間の平均値の差	5
標準偏差	6
α エラー	0.05

	両側検定
検出力	0.8
N2とN1のサンプルサイズの比	1

必要サンプルサイズ	計算結果
N1	23
N2	23

各群23人（計46人）の血圧を測る必要があります。

4 ま と め

「どの程度の差を実務的に意味があるかとみなすか」を明確にすることで、差の過大・過小評価を防ぎ、適切なサンプルサイズを設定することができます。実務的に意味のある差を定義し、標準的な検出条件（ $\alpha_{\text{エラー}}=0.05$ 、検出力0.8）を用いて必要なnを計算します。このようにして得られたnが実行困難性を伴う場合には、調査研究の目的や方向性を踏まえ、 $\alpha_{\text{エラー}}$ や検出力の設定を見直します。

なお、実務的に意味のある差の設定は主観に左右されやすいので、できるだけ関係者間での合意形成を得ることが望まれます。