



Rによる統計的検定: 独立な2群の比率差の検定

奥村泰之

国立精神・神経医療研究センター
精神保健研究所 社会精神保健研究部

於日本大学文理学部
図書館2階 メディア・ラボ 4(10:00~17:00)

Rによる統計的検定

- ◆ 比率差の**検定**の事例: 禁煙治療 (15 min)
- ◆ 禁煙治療データの解析 (30 min)
- ◆ 統計的検定の基礎 (60 min)
- ◆ 統計的検定の応用 (60 min)

統計的検定の応用

■ 学習目標

- 例数設計？
- 効果量？
- 95% CI？
- 比率差の検定の定義
- 無作為抽出の問題
- よくある誤解
- まとめ

統計的検定の応用

■ 例数設計 (検定力分析)

- 検定力をある値 (80%), 有意水準をある値 (5%) の精度で検定を行いたいときに, **最低限必要な標本サイズを求める方法**
- 検定力分析に必要な指標は,
(a) 検定力, (b) 有意水準, (c) 母集団効果量の3つ
- (a) 検定力と (b) 有意水準は概ね慣例で決まるが,
(c) **母集団効果量の見積り (予想) が最も重要で難しい**

統計的検定の応用

■ 効果量とは

- 帰無仮説が誤りである程度
- 母集団効果量: 母集団における効果量 (未知数)
- 標本効果量: 標本における効果量 (データから算出)

統計的検定の応用

■ 効果量の種類

- 基の単位の効果量...比率差, 平均値差など
 - 測定単位に意味があるので解釈がしやすい
 - e.g., 教授法Aと教授法Bの数学のテスト (100点満点) の平均得点の差は10点である
 - 測定単位が異なると研究間で比較ができない
 - e.g., 研究A: 100点満点の数学のテスト (10点差, 80点 vs 70点)
研究B: 50点満点の数学のテスト (10点差, 30点 vs 20点)
- 標準化された効果量...Cohenのh, Hedgesのg値など
 - 測定単位に意味がないので解釈が難しい
 - e.g., 教授法Aと教授法Bの数学のテスト (100点満点) の平均得点の差はHedgesのg値で0.5である
 - 測定単位が異なっても研究間で比較ができる (尺度不変の性質)
 - e.g., 研究A: 100点満点の数学のテスト (Hedgesのg = 0.5)
研究B: 50点満点の数学のテスト (Hedgesのg = 0.4)

統計的検定の応用

■ 例数設計のための母集団効果量の見積り方



1. 同様のリサーチ・クエスチョンのメタ分析または先行研究を確認して、標本効果量を確認する
 - 禁煙治療の事例 (Steinberg et al: Ann Intern Med 150: 447-454, 2009)
2. 予備研究を行い、標本効果量を算出する
3. 実質科学的、理論的な観点から、価値のある最低限の効果量の大きさを設定する
4. Jacob Cohenによる効果量の大きさの基準を使う

統計的検定の応用

■ 例数設計のための母集団効果量の見積り方

2. 予備研究を行い、標本効果量を算出する

- We calculated the sample size for the primary analysis on the basis of differences seen in a pilot study of 10 patients with newly diagnosed type 2 diabetes who received either the MED diet or the low-fat diet for 3 months. In this pilot study, HbA1c levels differed by 0.3%, with similar SDs that averaged 0.7. Assuming an α level of 0.05 and 80% power, the required number of patients for each group to observe a HbA1c difference of 0.25% is 87. To allow for a 25% drop-out rate, we randomly assigned 215 patients.

統計的検定の応用

■ 例数設計のための母集団効果量の見積り方

3. 実質科学的, 理論的な観点から, 価値のある最低限の効果量の大きさを設定する

- We asserted that a 25% absolute difference in the proportion of persons who correctly chose the superior drug would be clinically important (50% of the control group vs. 75% of the drug box group). With power set at 0.90 and a 2-sided P value of 0.05, we calculated that 85 people would be needed in each group. To allow for nonresponse, we aimed to recruit 120 people in each group.

統計的検定の応用

■ 例数設計のための母集団効果量の見積り方

4. Jacob Cohenによる効果量の大きさの基準を使う

- A four-arm randomized pre-test–post-test control group design with three-month follow-up was used to determine the effect of the intervention on participants' BMI and other related study measures among first year students (Campbell & Stanley, 1963). This study aimed to include at least 44 participants/arm, which was expected to yield power $\geq .80$, based on $\alpha \leq .05$ and **assuming a medium effect size (i.e., $f = .25$; Cohen, 1977).**

統計的検定の応用

■ Jacob Cohenによる効果量の大きさの基準

検定	母数	基準		
		小	中	大
独立な2群の比率差の検定	h	0.20	0.50	0.80
独立な2群の t 検定	d	0.20	0.50	0.80
分散分析	f	0.10	0.25	0.40
相関係数の検定	r	0.10	0.30	0.50
重回帰分析	R^2	0.02	0.13	0.26

統計的検定の応用

■ Cohenの基準の特徴

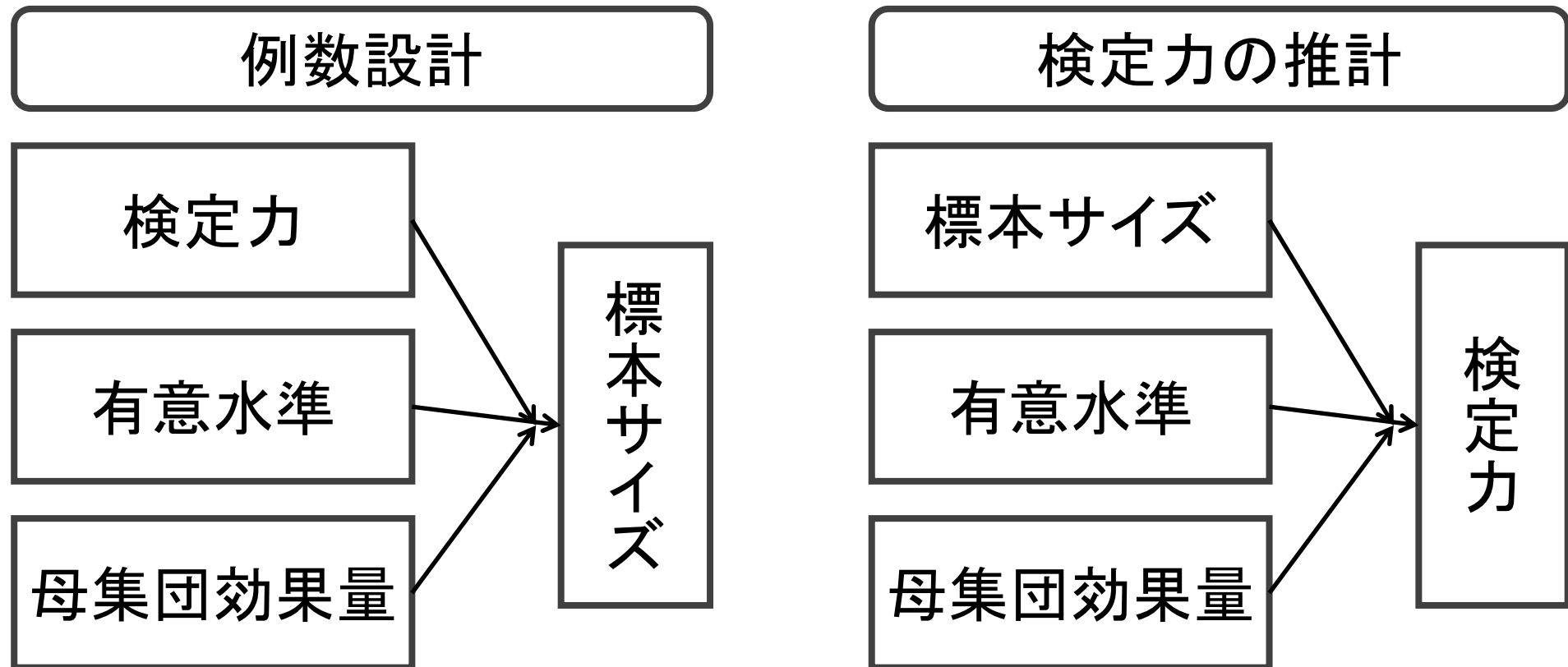
- 行動科学の領域におけるCohenの主観により決められた効果の大きさのベンチマーク
- 研究領域/仮説により、**実質的に意味のある効果の大きさは異なる**ので、この基準を絶対視してはならない

■ 効果の大きさの基準の設計意図

- 中...注意深い観察者ならば裸眼で認識できる程度の差異
- 小...中よりも明らかに小さいが、取るに足らない差異とは言い難い差異
- 大...中と小の差を、中に足した大きさ

統計的検定の応用

■ 統計的検定における4つ指標の関連



統計的検定の応用

■ 演習

1. `power.prop.test` を使い、以下の条件の例数設計をすること。母集団における第1群の比率は30%である。母集団における第2群の比率が10%, 20%, 28%の3つの条件を想定し、検定力80%, 有意水準5%の基で、必要な標本サイズを推計すること。
2. `power.prop.test` を使い、以下の条件の検定力を推計すること。母集団における第1群の比率は30%, 第2群の比率は20%である。有意水準5%の基で、各群50名, 100名, 300名の3つの条件を想定し、検定力を推計すること。

統計的検定の応用

■ 検定力の推計の関数

- `power.prop.test`(n =標本サイズ, $p1$ =群1の比率, $p2$ =群2の比率, $sig.level$ =有意水準)

Two-sample comparison of proportions power calculation

```
power = 0.2087730
```

検定力

NOTE: n is number in *each* group

統計的検定の応用

■ 例数設計の要点

- 母集団効果量が大きな現象であれば、小さい標本サイズでも、良い精度 (検定力80%, 有意水準5%) の検定を実施できる。

■ 検定力の推計の要点

- 検定力が50%を切るような標本サイズしか確保できないならば、研究をしない方が良い
 - 帰無仮説が誤っている場合に検定力が50%を切ることは、「研究を行わずにコインを投げて裏が出たら帰無仮説を棄却するよりも、正しく判断する確率が低い」ことを意味する
 - 本当は存在するはずの重要な差異を見逃す確率が非常に高い条件下で行う研究は、貴重な研究資源を使用することの倫理的な問題がある