

仮説と検定

稲積 泰宏 (いなづみ やすひろ)

今日の内容

- 仮説検定とは
- 帰無仮説と対立仮説
- 検定統計量
- 有意水準と p 値
- 第1種の誤りと第2種の誤り

仮説検定とは

定義

データに基づいて、母集団に関する仮説が正しいかどうかを統計的に判断する方法

区間推定との違い

- 区間推定：母数の範囲を推定する
- 仮説検定：母数に関する主張の正否を判定する

仮説検定の具体例

例1：新薬の効果

この新薬には効果があるか？

例2：製品の品質管理

不良品率は基準値以下か？

例3：教育プログラムの効果

新しい学習方法は従来より優れているか？

帰無仮説と対立仮説

帰無仮説

記号： H_0

検証したい主張に対する「否定的な立場」または「現状維持」の仮説

対立仮説

記号： H_1

検証したい主張、帰無仮説が棄却されたときに採択される仮説

仮説の設定例

例：新薬の効果検証

従来の治療法での治癒率は60%

- H_0 : 新薬の治癒率 $p = 0.6$ (効果なし)
- H_1 : 新薬の治癒率 $p > 0.6$ (効果あり)

重要な原則

帰無仮説は「差がない」「効果がない」という保守的な立場
対立仮説が証明したい主張

片側検定と両側検定

片側検定

- 右側検定 : $H_1 : p > p_0$
- 左側検定 : $H_1 : p < p_0$

両側検定

- $H_1 : p \neq p_0$

使い分け

- 特定の方向の変化に関心がある → 片側検定
- 変化の有無に関心がある → 両側検定

仮説検定の手順

1. 帰無仮説 H_0 と対立仮説 H_1 を設定
2. 有意水準 α を決める（通常 0.05 または 0.01）
3. 検定統計量を計算
4. p 値を計算
5. 判定を下す

検定統計量

定義

仮説の真偽を判断するために、標本から計算する統計量

役割

検定統計量の値が、帰無仮説のもとでどれくらい極端かを評価することで、仮説の妥当性を判断する

有意水準

定義

帰無仮説が正しいにもかかわらず、それを棄却してしまう確率の上限

記号： α

よく使われる値

- $\alpha = 0.05$ (5%)
- $\alpha = 0.01$ (1%)

有意水準が小さいほど、判定は保守的（帰無仮説を棄却しにくい）

p 値（有意確率）

定義

帰無仮説のもとで、観測された検定統計量以上に極端な値が得られる確率

判定方法

- $p\text{値} < \alpha \rightarrow$ 帰無仮説を棄却する
- $p\text{値} \geq \alpha \rightarrow$ 帰無仮説を受容する

p 値の解釈

- p 値が小さいほど、帰無仮説に反する強い証拠
- p 値が大きいほど、帰無仮説と矛盾しない

注意：「帰無仮説を採択する」とは言わない

p 値の計算（両側検定の場合）

検定統計量を z とすると

$$p\text{値} = 2 \times P(Z > |z|)$$

ここで $Z \sim N(0, 1)$

（説明の簡略化のためZ検定を用いる。詳細は次回以降。）

p 値の計算（片側検定の場合）

右側検定 ($H_1 : p > p_0$)

$$p\text{値} = P(Z > z)$$

左側検定 ($H_1 : p < p_0$)

$$p\text{値} = P(Z < z)$$

「ロボットAとロボットBの強さは等しい」

問1)

検定における2種類の誤り

第1種の誤り

- 帰無仮説が真であるのに、それを棄却してしまう誤り
- 確率： α （有意水準）

例）実際には新薬に効果がないのに、偶然のばらつきで「効果あり」と判断してしまう。

第2種の誤り

- 帰無仮説が偽であるのに、それを棄却しない誤り
- 確率： β

例）本当は治癒率が上がっているのに、標本数が少ないため差を検出できず「効果なし」と結論づけてしまう。

誤りの関係

	H_0 が真	H_0 が偽
H_0 を受容	正しい判定	第2種の誤り (確率 β)
H_0 を棄却	第1種の誤り (確率 α)	正しい判定

トレードオフ

α を小さくすると β は大きくなる
通常は第1種の誤りを制御 (α を固定)

検出力

定義

対立仮説が真のとき、正しく帰無仮説を棄却する確率

$$\text{検出力} = 1 - \beta$$

検出力を高める方法

1. 標本サイズを大きくする
2. 有意水準を大きくする（ただし第1種の誤りが増える）
3. 効果が大きい場合（真の p と p_0 の差が大きい）

仮説検定の基本

- 帰無仮説 H_0 と対立仮説 H_1 を設定
- 有意水準 α で第1種の誤りを制御
- 検定統計量を計算し、 p 値で判定

感想を会議のチャット欄へ