

Rコマンダーを用いた統計解析 (8)

担当 繁樹 算男 教授

アシスタント 森 一将

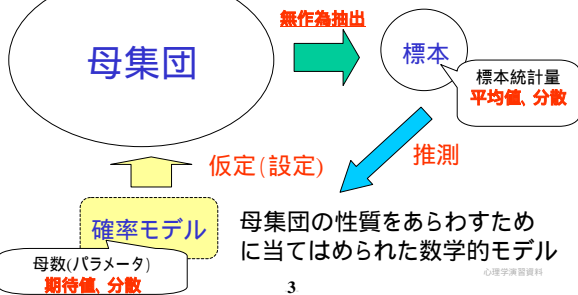
今日の演習事項

- 統計的仮説検定
 - クロス表の作成
 - カイ二乗検定

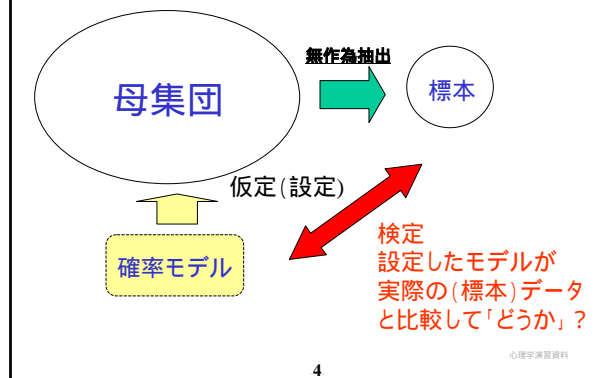
母集団と標本と確率モデル

関心のある対象
(データ)全体

実験・調査などにより
取得されたデータ



検定の基本的考え方



統計的仮説検定

2009年7月20日

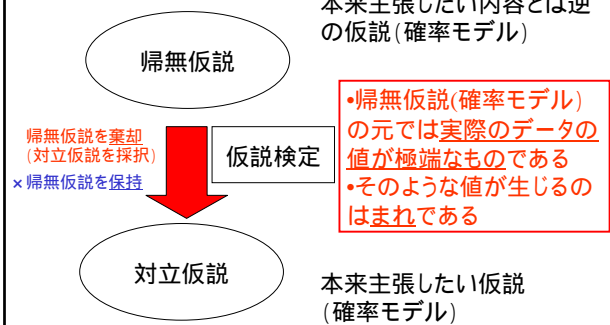
1. 帰無仮説、対立仮説を設定
2. 統計的仮説検定に用いられる標本統計量を選択
3. 棄却域、有意水準(判断の基準になる確率のしきい値)の設定
4. 実際のデータから標本統計量を計算
5. 検定統計量の値が...
 - 棄却域の値なら、帰無仮説を**棄却**(対立仮説を**採択**)する。
 - そうでなければ、帰無仮説を**保持**する。

5

心理学演習資料

帰無仮説と対立仮説

2009年7月20日



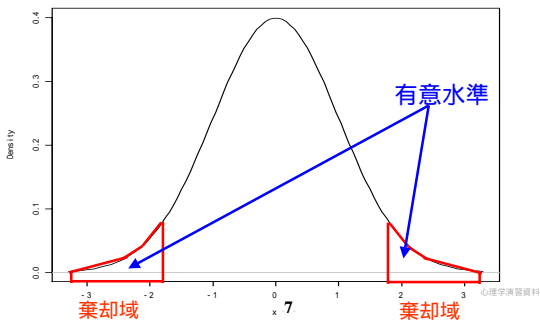
6

心理学演習資料

有意水準と棄却域

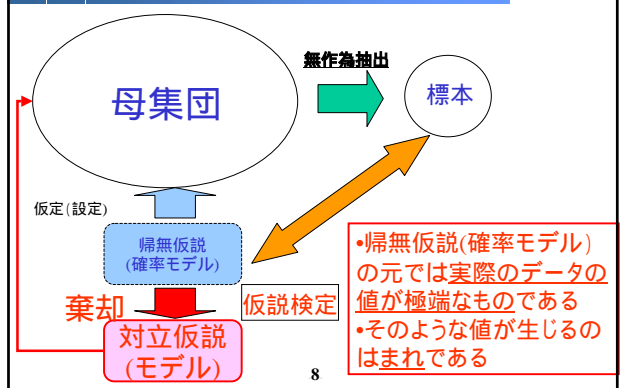
2009年7月20日

棄却域 = 「極端な値」
有意水準 = 「まれにしか生じない」確率



検定の基本的考え方

2009年7月20日



8

心理学演習資料

データ概要

2009年7月20日

- 実験データの取得を想定した練習
 - 「身長データ」
 - ある地域の高校生1000名を無作為抽出し、身長を測定
 - Excelのデータシートからの取得

9

心理学演習資料

カイ二乗検定

2009年7月20日

- 「クロス集計表」の検定
- クロス集計表
 - データに含まれる質的変数で頻度を集計
 - データ全体における2つ(以上)の質的変数の頻度を行列で表示

10

心理学演習資料

Rコマンドによるクロス集計表

2009年7月20日

- 「practice1」データセットの利用
- 「統計量」「分割表」「2元表」
- 「area」「sex」を選択
- 「独立性のカイ二乗検定」をはずす

| | 女 | 男 | 計 |
|-----|-----|-----|------|
| A地区 | 140 | 360 | 500 |
| B地区 | 185 | 315 | 500 |
| 計 | 325 | 675 | 1000 |

11

心理学演習資料

カイ二乗検定

2009年7月20日

| | 女 | 男 | 計 |
|-----|-----|-----|------|
| A地区 | 140 | 360 | 500 |
| B地区 | 185 | 315 | 500 |
| 計 | 325 | 675 | 1000 |

- A地区とB地区において男女の比率に差があることを示したい

12

心理学演習資料

1. 帰無仮説、対立仮説を設定

2009年7月20日

- A地区とB地区において男女の比率に差があることを示したい
- 帰無仮説: **本来主張したい内容とは逆のもの**
「A地区とB地区において男女の比率に差がない」
- 対立仮説: **本来主張したい内容**
「A地区とB地区において男女の比率に差がある」

心理学演習資料

13

2. 統計的仮説検定に用いられる標本統計量を選択

2009年7月20日

- クロス表における検定 = χ^2 (カイ二乗) 統計量を選択

$$\chi^2 = \frac{(\text{実現値} - \text{理論値})^2}{\text{理論値}} \text{ の総和}$$

| | 女 | 男 | 計 |
|-----|-----|-----|------|
| A地区 | 360 | 140 | 500 |
| B地区 | 165 | 335 | 500 |
| 計 | 325 | 675 | 1000 |

理論値 = 周辺度数から計算

心理学演習資料

14

2. 統計的仮説検定に用いられる標本統計量を選択

2009年7月20日

- 理論値の算出方法
タテ・ヨコの合計値をとる
タテ・ヨコの割合を計算
(全体の標本数) × (タテの割合) × (ヨコの割合)

| | 女 | 男 | 計 |
|-----|-------|-------|------|
| A地区 | 162.5 | 337.5 | 500 |
| B地区 | 162.5 | 337.5 | 500 |
| 計 | 325 | 675 | 1000 |

0.325 0.675

心理学演習資料

15

3. 棄却域、有意水準 (判断の基準になる確率のしきい値) の設定

2009年7月20日

- よく用いられる有意水準: 5%

心理学演習資料

16

4.実際のデータから標本統計量を計算

2009年7月20日

- 「practice1」データセットの利用
- 「統計量」「分割表」「2元表」
- 「area」「sex」を選択
- 「独立性のカイ二乗検定」を**はずさない**

```

>> table(xtabs(practice1[,c("area", "sex")]))
      area sex
      --- --
      A  1  1
      A  1  2
      A  2  1
      A  2  2
  
```

17

心理学演習資料

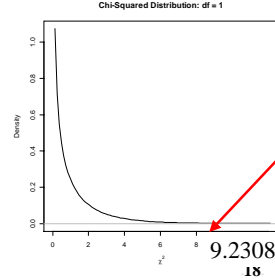
5検定統計量の値が...

2009年7月20日

```

FREQ:2007.5 CHI-SQUARED TEST
Area1 .White
CHI-SQUARED = 9.2308, DF = 1, P-VALUE = 0.00238
  
```

自由度 = (ヨコの数-1) × (タテの数-1)



• 帰無仮説(確率モデル)の元では**実際のデータの値が極端なもの**である
• そのような値が生じるのは**まれ**である

帰無仮説は**正しくない**(棄却する)
対立仮説が**正しい**

18

結論

2009年7月20日

| | 女 | 男 | 計 |
|-----|-----|-----|------|
| A地区 | 140 | 360 | 500 |
| B地区 | 185 | 315 | 500 |
| 計 | 325 | 675 | 1000 |

仮説検定により有意差が見られた

× 「A地区とB地区において男女の比率に**差がない**」
「A地区とB地区において男女の比率に**差がある**」

19

心理学演習資料

Rコマンドによるクロス集計表(2)

2009年7月20日

- 2 × 3のクロス表の作成

| ID | height | area | sex |
|----|--------|------|-----|
| 1 | 147.1 | A地区 | 男 |
| 2 | 170.6 | A地区 | 女 |
| 3 | 167.2 | A地区 | 男 |
| 4 | 158.5 | A地区 | 女 |
| 5 | 175.9 | A地区 | 男 |
| 6 | 157.4 | A地区 | 男 |
| 7 | 150.8 | A地区 | 女 |
| 8 | 174.0 | A地区 | 男 |
| 9 | 174.3 | A地区 | 女 |
| 10 | 156.5 | A地区 | 女 |

身長によって高校生を3つの群に分けてみる

- 高身長群
- 中身長群
- 低身長群

20

心理学演習資料

Rコマンドーによるクロス集計表(2)

2009年7月20日

- 「データ」「アクティブデータセット内の変数の管理」「数値変数を区間で区分」
- 「height」変数を選択
- 新しい変数名 「gun」(群)
- 区関数 3
- 「名前を指定」
- 「等間隔の区間」



21

Rコマンドーによるクロス集計表(2)

2009年7月20日

- 区間名
- 高群、中群、低群



身長による群分け

22

Rコマンドーによるクロス集計表(2)

2009年7月20日

- 「統計量」「分割表」「2元表」
- 「area」「gun」を選択
- 「独立性のカイ二乗検定」をはずす

| | 高群 | 中群 | 低群 | 計 |
|-----|-----|-----|----|------|
| A地区 | 77 | 393 | 30 | 500 |
| B地区 | 76 | 373 | 51 | 500 |
| 計 | 153 | 766 | 81 | 1000 |

23

心理学演習資料

カイ二乗検定

2009年7月20日

| | 高群 | 中群 | 低群 | 計 |
|-----|-----|-----|----|------|
| A地区 | 77 | 393 | 30 | 500 |
| B地区 | 76 | 373 | 51 | 500 |
| 計 | 153 | 766 | 81 | 1000 |

- A地区とB地区において身長群の比率に差があることを示したい

24

心理学演習資料

1. 帰無仮説、対立仮説を設定

2009年7月20日

- A地区とB地区において身長群の比率に差があることを示したい
- 帰無仮説: 本来主張したい内容とは逆のもの
「A地区とB地区において身長群の比率に差がない」
- 対立仮説: 本来主張したい内容
「A地区とB地区において身長群の比率に差がある」

25

心理学演習資料

2. 統計的仮説検定に用いられる標本統計量を選択

2009年7月20日

- クロス表における検定 = χ^2 (カイ二乗) 統計量を選択

$$\chi^2 = \frac{(\text{実現値} - \text{理論値})^2}{\text{理論値}} \text{の総和}$$

26

心理学演習資料

2. 統計的仮説検定に用いられる標本統計量を選択

2009年7月20日

- 理論値の算出方法
タテ・ヨコの合計値をとる
タテ・ヨコの割合を計算
(全体の標本数) × (タテの割合) × (ヨコの割合)

| | 高群 | 中群 | 低群 | 計 | |
|-----|-------|-------|-------|------|-----|
| A地区 | 76.5 | 383 | 40.5 | 500 | 0.5 |
| B地区 | 76.5 | 383 | 40.5 | 500 | 0.5 |
| 計 | 153 | 766 | 81 | 1000 | |
| | 0.153 | 0.766 | 0.081 | | |

27

心理学演習資料

3. 棄却域、有意水準 (判断の基準になる確率のしきい値) の設定

2009年7月20日

- よく用いられる有意水準: 5%

28

心理学演習資料

4.実際のデータから標本統計量を計算

2009年7月20日

- 「practice1」データセットの利用
- 「統計量」「分割表」「2元表」
- 「area」「gun」を選択
- 「独立性のカイ二乗検定」を**はずさない**

```

>>> table
      area  gun
      ----  --
A地区  77  393
B地区  76  373
計     153 766
    
```

```

>>> test
      Pearson Chi-Square test
data:  table
Chi-squared = 5.9732, df = 1, p-value = 0.01844
    
```

心理学演習資料

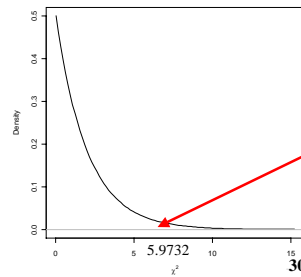
5検定統計量の値が...

2009年7月20日

PEARSON'S CHI-SQUARE TEST

自由度 = (ヨコの数-1) × (タテの数-1)

data: table
Chi-squared = 5.9732, df = 1, p-value = 0.01844



• 帰無仮説(確率モデル)の元では実際のデータの値が極端なものである
• そのような値が生じるのはまれである

帰無仮説は正しくないとはいえない(保持する)
対立仮説については何もいえない

心理学演習資料

結論

2009年7月20日

| | 高群 | 中群 | 低群 | 計 |
|-----|-----|-----|----|------|
| A地区 | 77 | 393 | 30 | 500 |
| B地区 | 76 | 373 | 51 | 500 |
| 計 | 153 | 766 | 81 | 1000 |

仮説検定により有意差は見られなかった

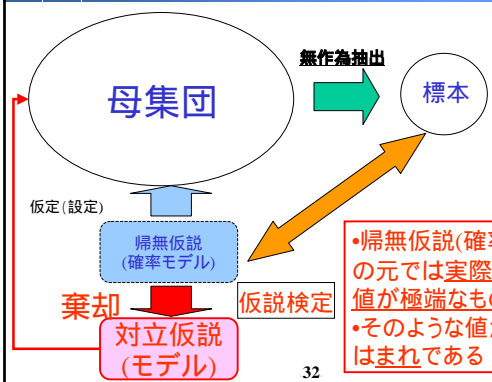
「A地区とB地区において身長群の比率に差があるとはいえない」

「比率が同じ」かどうかは仮説検定では分からない

31

(再掲) 検定の基本的考え方

2009年7月20日



• 帰無仮説(確率モデル)の元では実際のデータの値が極端なものである
• そのような値が生じるのはまれである

32