

Rコマンドを用いた統計解析 (7)

担当 繁樹 算男 教授

アシスタント 森 一将

1

心理学演習資料

前回の補足

- 正規乱数発生時のパラメータについて
 - 正規分布からの乱数発生時のパラメータの設定は以下のように行います
 - 平均
 - 期待値と読み替えてください
 - 一部の表記で期待値 = 母平均とみなして平均の用語を使っている場合もあります
 - 標準偏差
 - Rにおける関数の引数は標準偏差でした。訂正します。
 - 通常の表記の場合は分散を表します

2

心理学演習資料

今日の演習事項

- 確率分布と標本、無作為抽出 (復習)
 - 母集団と標本
 - 無作為抽出
- 統計的仮説検定
 - カイ二乗検定

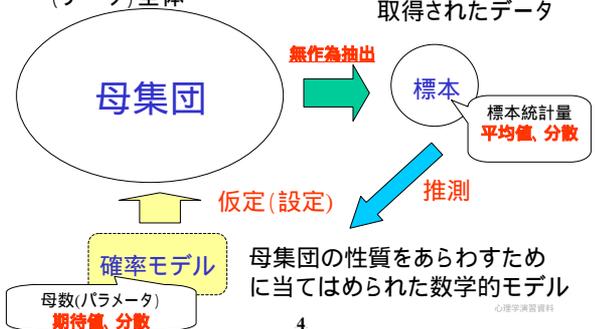
3

心理学演習資料

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



4

心理学演習資料

無作為抽出(ランダムサンプリング)

2009年7月20日

- 母集団のどの要素も**等しい確率**で抽出する
抽出回数を多くすると標本の持つ特性が母集団のものと等しくなる
母集団の特性についての推測が出来る

5

心理学演習資料

Rによる無作為抽出

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」
 - ある地域の高校男子1000名の身長データ(人工)
 - Excelのデータシートからの取得
母集団からのデータ抽出(おさらい)
母集団からの対象(被験者)抽出

6

心理学演習資料

母集団からのデータ抽出(1)

2009年7月20日

- すでにすべての対象に対して身長を測定した後、一部のデータを抽出し利用
- 配布の「rand_sample.xls」を利用
- Excelファイルからの取り込み
- 「full」セットを選択
- データセット名は「practice1」

7

心理学演習資料

母集団からのデータ抽出(2)

2009年7月20日

- 以下のコマンドを実行
 - practice1 #データセットの内容の確認
 - X<- practice1\$height #身長項目を変数に代入
 - Y<-Sample(x,100)
#100人分のデータを無作為抽出
 - Y #抽出したデータを表示

8

心理学演習資料

母集団からの対象(被験者)抽出(1)

2009年7月20日

- 母集団に無作為抽出を行い、抽出された対象に測定を実施
- 被験者のIDを抽出
- 配布の「rand_sample.xls」を利用
- Excelファイルからの取り込み
- 「id」セットを選択
- データセット名は「practice2」

9

心理学演習資料

母集団からの対象(被験者)抽出(2)

2009年7月20日

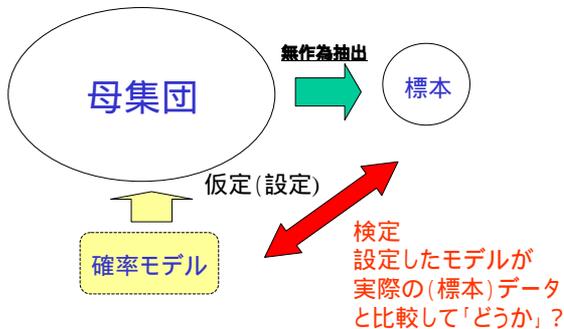
- 以下のコマンドを実行
 - practice2 #データセットの内容の確認
 - X<- practice1\$Sid #身長項目を変数に代入
 - Y<-Sample(x,100)
#100人分のデータを無作為抽出
 - Y #抽出したデータを表示

10

心理学演習資料

検定の基本的考え方

2009年7月20日



11

心理学演習資料

統計的仮説検定

2009年7月20日

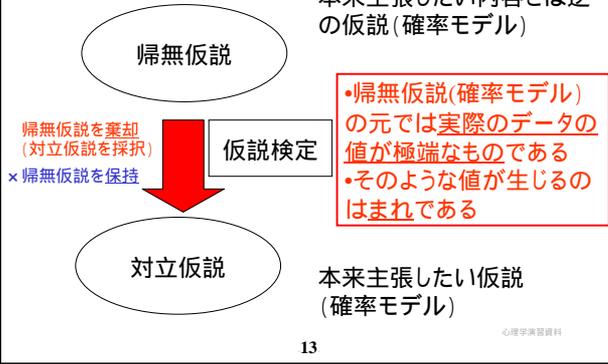
1. 帰無仮説、対立仮説を設定
2. 統計的仮説検定に用いられる標本統計量を選択
3. 棄却域、有意水準(判断の基準になる確率のしきい値)の設定
4. 実際のデータから標本統計量を計算
5. 検定統計量の値が...
 - 棄却域の値なら、帰無仮説を**棄却**(対立仮説を**採択**)する。
 - そうでなければ、帰無仮説を**保持**する。

12

心理学演習資料

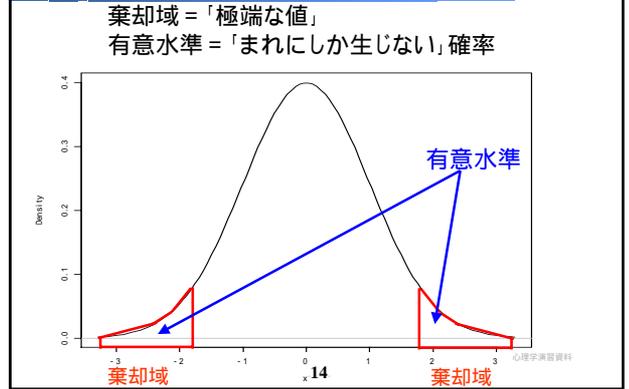
帰無仮説と対立仮説

2009年7月20日



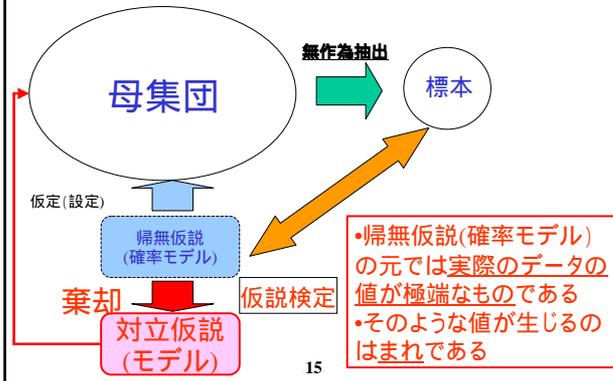
有意水準と棄却域

2009年7月20日



検定の基本的考え方

2009年7月20日



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

16

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

17

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

18

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」

19

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

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

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

理論値 = 周辺度数から計算

20

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

- よく用いられる有意水準:5%

21

心理学演習資料

4.実際のデータから標本統計量を計算

- 「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**

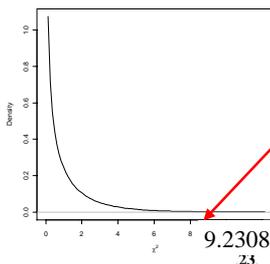
```
SPSS Syntax
* Data List.
DATA LIST FILE='practice1.sav'.
* Split.
SPLIT BY=area sex.
* Descriptives.
DESCRIPTIVES.
* Chi-Square.
CHISQ.
* Save.
SAVE OUTFILE='practice1_chi2.sav'.
END.
```

22

心理学演習資料

5検定統計量の値が...

```
SPSS Syntax
* Chi-Square.
CHISQ.
* Save.
SAVE OUTFILE='practice1_chi2.sav'.
END.
```



•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

23

心理学演習資料

結論

	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ない」
「A地区とB地区において男女の比率に差がある」

24

心理学演習資料

(再掲) 検定の基本的考え方

2009年7月20日

