

# Rコマンダーを用いた統計解析 (6)

担当 繁樹 算男 教授

アシスタント 森 一将

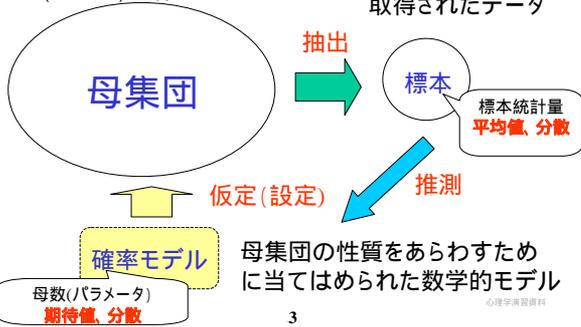
## 今日の演習事項

- 確率分布と標本、無作為抽出
  - 母集団と標本
  - 確率分布
  - 無作為抽出

## 母集団と標本と確率モデル

関心のある対象  
(データ)全体

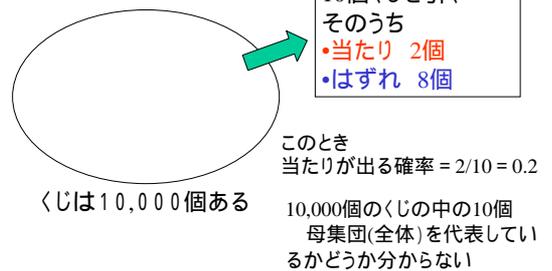
実験・調査などにより  
取得されたデータ



母集団の性質をあらわすために  
当てはめられた数学的モデル

## 確率分布(1)

- 確率変数とは？(頻度による定義)  
(例) 当たりくじ

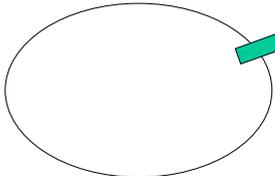


## 確率分布(2)

2009年7月20日

- 確率変数とは？(頻度による定義)

(例) 当たりくじ



くじは10,000個ある

100個くじを引く  
そのうち  
• 当たり 18個  
• はずれ 82個

このとき  
当たりが出る確率  
=  $18/100 = 0.18$

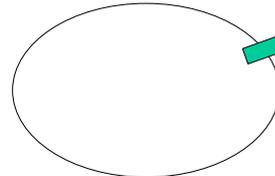
10,000個のくじの中の100個  
母集団(全体)を代表してい  
るかどうかまだ分からない

## 確率分布(3)

2009年7月20日

- 確率変数とは？(頻度による定義)

(例) 当たりくじ



くじは10,000個ある

1,000個くじを引く  
そのうち  
• 当たり 198個  
• はずれ 802個

このとき  
当たりが出る確率  
=  $198/100 = 0.198$

10,000個のくじの中の1,000個  
母集団(全体)を代表してい  
るように思える

## 確率分布(4)

2009年7月20日

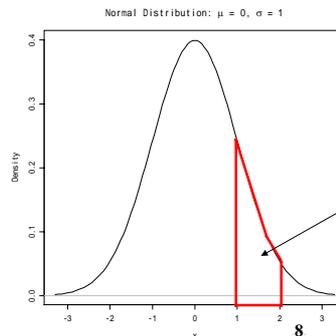
- 確率変数とは？(頻度による定義)
- N回の試行のうちある事象の起こった回数  $r$
- 事象の発生割合  $r/N$
- N(試行回数)を大きくすると  $r/N$  は母集団の持つ性質をあらわす **確率**

ただし、この考え方では定義できない確率もある(1度限りの勝負における確率など)

## 確率密度関数

2009年7月20日

- 確率を「面積」で表したものの



ある事象  
(例: 紙飛行機大会で  
が1~2mの成績を残す)  
確率は密度関数の  
面積で示される

## Rコマンダーによる確率分布の表示

2009年7月20日

- 「分布」「(各確率分布)」「(確率分布)の作図」
- 母数、繰り返し回数を設定
- 心理統計で利用される代表的な分布
  - 正規分布
  - t分布
  - カイ二乗分布
  - F分布

9

心理学演習資料

## 正規分布(1)

2009年7月20日

- 正規分布 N(期待値,分散)

標準正規分布 : N(0,1)

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$

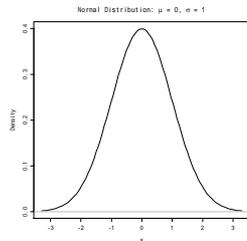
10

心理学演習資料

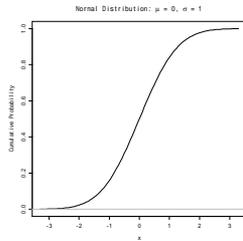
## 正規分布(2)

2009年7月20日

(確率密度関数)



(累積密度関数)



11

心理学演習資料

## 正規分布(3)

2009年7月20日

- 分散を一定(1)にして  
期待値を大きく / 小さくすると確率密度関数はどのように変わるか？
- 期待値を一定(0)にして  
分散を大きく / 小さくすると確率密度関数はどのように変わるか？

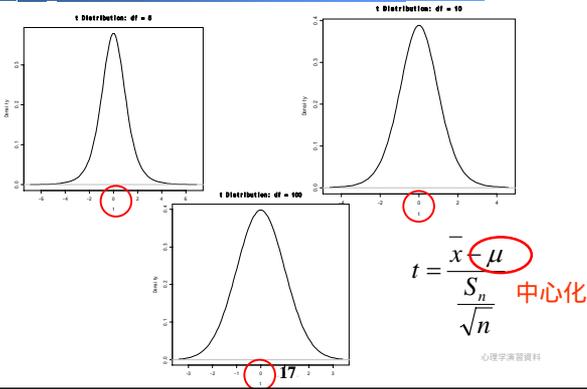
12

心理学演習資料



### t分布(3)

2009年7月20日



### カイ二乗分布

2009年7月20日

- カイ二乗分布  $\chi^2$ (自由度)

$x_1, x_2, \dots, x_n$  が正規分布  $N(\mu, \sigma^2)$  に従うとき

$$\chi^2 = \left(\frac{x_1 - \mu}{\sigma_1}\right)^2 + \left(\frac{x_2 - \mu}{\sigma_2}\right)^2 + \dots + \left(\frac{x_n - \mu}{\sigma_n}\right)^2$$

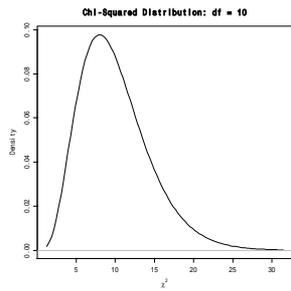
分散に似ている

は自由度  $n$  のカイ二乗分布に従う

### カイ二乗分布(2)

2009年7月20日

- 自由度10のカイ二乗分布を描いてみる



### F分布(1)

2009年7月20日

- F分布  $F$ (自由度1, 自由度2)
- 自由度  $n_1, n_2$  のカイ二乗分布の比

$$F = \frac{\frac{\chi_1^2}{n_1}}{\frac{\chi_2^2}{n_2}}$$

分散に似ている

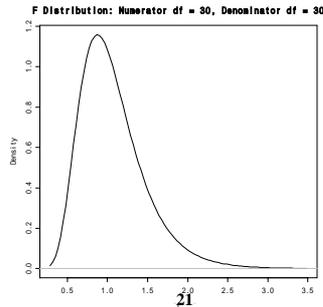
分散の比の検定  
分散分析

は自由度  $n_1, n_2$  のF分布に従う

## F分布(2)

2009年7月20日

- 自由度(30,30)のF分布を描いてみる



## 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

22

心理学演習資料

## Rによる無作為抽出(1)

2009年7月20日

- 桜データ: アメリカ桜に関するデータ
  - Girth(周囲の長さ)、Height(高さ)、Volume(体積)
- Rのbaseパッケージ内の「trees」データ
- コマンドラインから「trees」と入力

23

心理学演習資料

## Rによる無作為抽出(2)

2009年7月20日

- 以下のコマンドを実行
- `X<- trees$Height` #桜データの高さの部分を入力
- `X` #データ内容を確認
- `Y<-Sample(X,10)` #10個の標本を無作為抽出
- `Y` #データ内容を確認

キーボードの上下ヤジルシで過去の実行コマンドの履歴が出ます。

24

心理学演習資料

### Rによる無作為抽出(3)

2009年7月20日

- 母集団: 31個
  - 母集団の平均値を出してみる
- 標本(集団) その1: 3個
  - 標本の平均値を出してみる
  - 何回か無作為抽出を行ったとき結果は変わるか?
- 標本(集団) その2: 28個
  - 標本の平均値を出してみる
  - 何回か無作為抽出を行ったとき結果は変わるか?

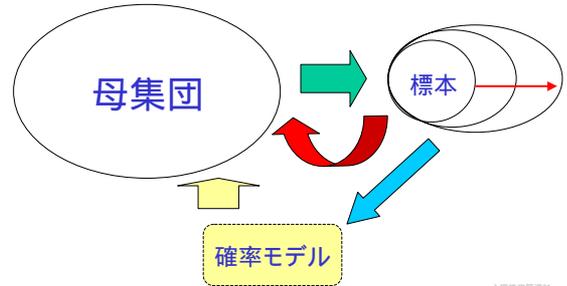
25

心理学演習資料

### 母集団と標本の関係

2009年7月20日

- 標本数が多いほど平均は母集団に近くなっていく
- 特性も似てくる



26

心理学演習資料

2009年7月20日

27

心理学演習資料