

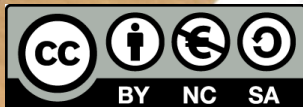
**rd-5. t 検定**

**データサイエンス演習**

**(R システムを使用)**

<https://www.kkaneko.jp/de/rd/index.html>

**金子邦彦**



# アウトライン



5-1. 母集団と標本

5-2. t 検定

# 5-1 母集団と標本

# 母集団



**母集団**は、調査や研究の対象となる全体の集団のこと

- 母集団の把握と理解が重要

(例) 人類全体、20歳以上の人類全体

# サンプリングと標本



- 母集団全体を調べるのが困難な場合、**サンプリング**を適切に行う

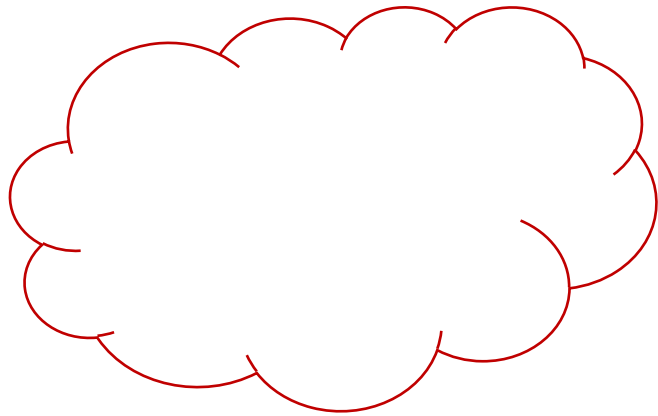
(例) 1000名をランダムに選ぶ

- **サンプリング**は、母集団から一部を選ぶこと。
- 母集団全体を調べるのではなく、一部を調べることになる。
- **標本**は、サンプリングで選ばれたもののこと。



# サンプリングと標本

母集団



あるときの標本

128
104
124
85
120

平均

112.2

別の標本

118
110
96
85
109

平均

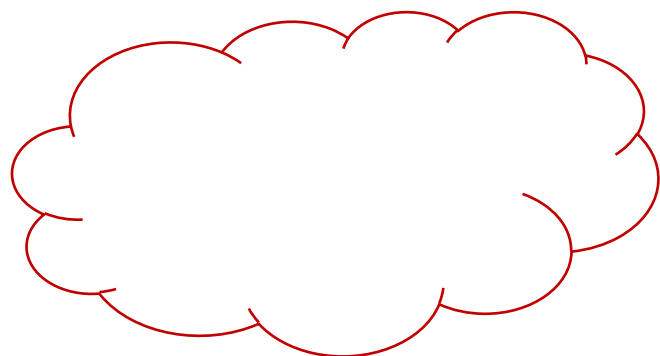
103.6

選ばれた標本によっては、  
**値が違い、平均なども異なってくる**

# 十分な数の標本が必要

あるときの標本

母集団



128
104
124
85
120

- 標本の大きさが小さいと、結果の信頼性が下がる
- 十分な数の標本を得ることが重要
- 標本の大きさの決定は簡単に決めることができない
- 母集団の特徴、調査や研究の目的によって、適切な標本の大きさは変わることに注意しよう

# まとめ

- **母集団**：調査や研究の対象となる**全体の集団**

- **サンプリング**：

母集団全体を調べることが困難な場合、母集団から一部を選ぶサンプリングを行う。

母集団の特徴や性質を**推測**することが可能となる。

- **標本**：

**標本**は、母集団からサンプリングで選ばれた母集団の一部。

標本から得られたデータを分析し、母集団全体の性質や傾向を推測可能。

【注意点】十分な標本サイズの確保が必要。ランダムに選択するなどの考慮が重要。



# 5-2 t 検定

# t 検定



t 検定は、2つの標本の平均値が統計的に有意に異なるかどうかを判断するための統計手法

## 【注意点】

- 標本が正規分布に従っていること
- 外れ値が存在する場合は、取り除いたり、適切に修正すること
- 十分な標本サイズを確保すること。小さな標本サイズでは、結果の信頼性が下がる可能性がある

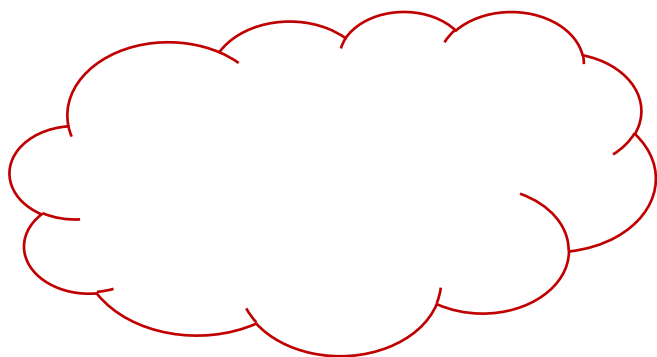
# 複数の母集団



## • 母集団が複数あるという考え方は重要

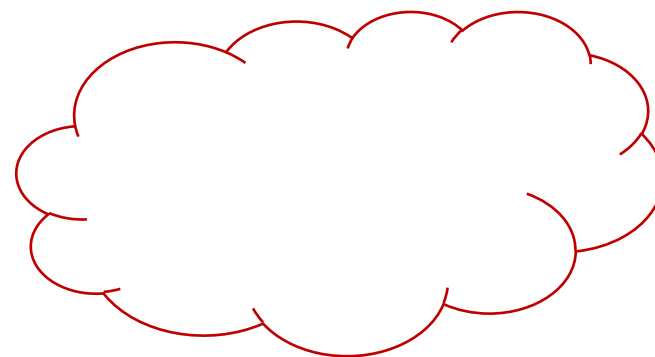
あなたは大学生です。授業Aを受けた人と、授業Aを受けていない人の調査し、比較してみたいと考えました

母集団



授業Aを受けた人

母集団

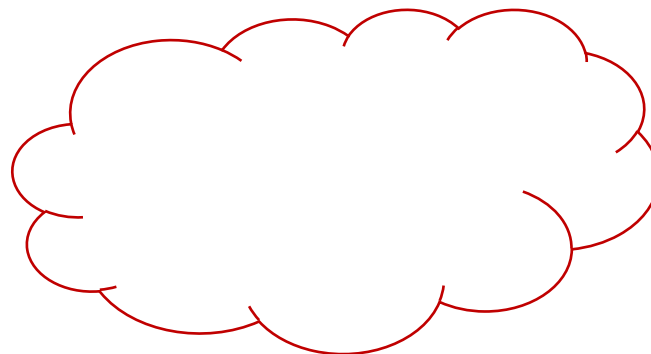


授業Aを受けていない人

母集団が2つ

# 2つの母集団と2つの標本

母集団

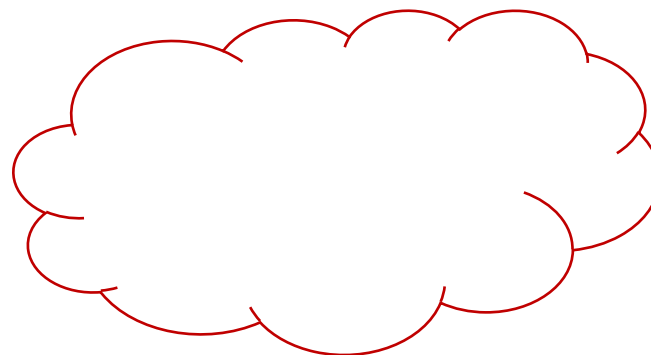


サンプリング

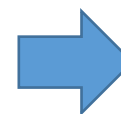


標本

別の母集団



サンプリング



標本

# t 検定と p 値



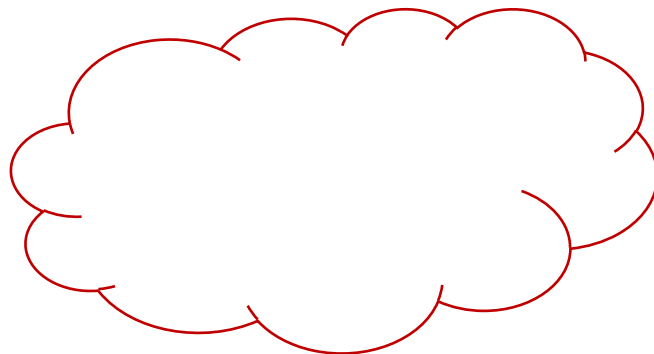
- t 検定は、2つの標本の平均値が統計的に有意に異なるかどうかを判断するための統計手法
- p 値は、2つの標本の差が偶然による（有意でない）確率を示す

p 値 = 0.99 のとき、「偶然による確率は 99 %」  
⇒ 有意であるとも有意でないともいえない

P 値 = 0.0005 のとき、「偶然による確率は 0.05 %」  
⇒ おそらく 有意である

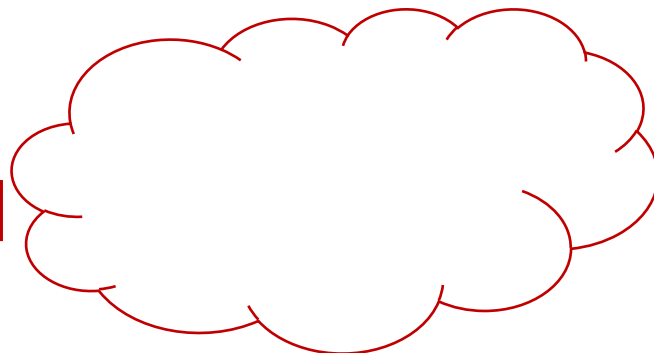
# 2つの母集団と2つの標本

母集団



128  
104  
124  
85  
120

別の母集団



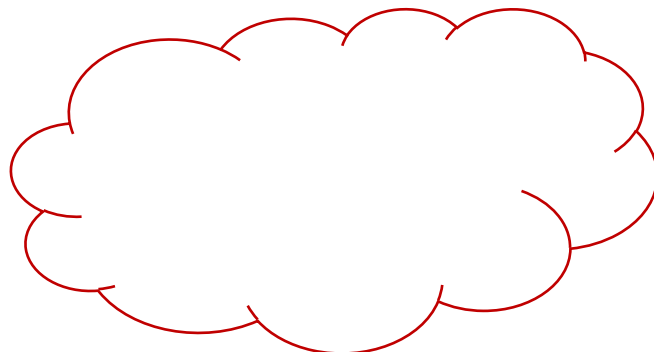
180  
191  
189  
131  
130  
150

2つの標本から t 検定の  
p 値を算出 : 0.006908

2つの標本の差が偶然による  
(有意でない) 確率が低い 14

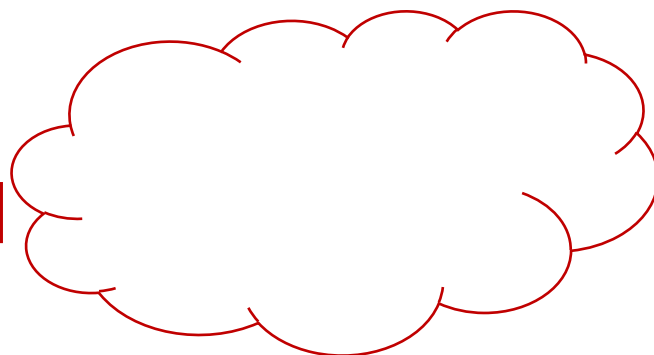
## 2つの母集団と2つの標本

母集団



128  
104  
124  
85  
120

別の母集団



100  
106  
89  
89  
105

2つの標本から t 検定の  
p 値を算出 : 0.1541

有意であるとも有意でないとも言えない

# p 値と有意性



- t 検定の p 値は、2つの標本の差が偶然による確率を示す
- p値が小さいとき「とても偶然とは思えず、有意である」と考える
- p値が大ききときは「偶然であるとも、偶然でないとも言えない」と考える



## • t検定

- **t検定は、2つの標本の平均値の統計的な有意性を判断する統計手法**
- **標本が正規分布に従い、外れ値を適切に扱い、十分な標本サイズを確保することが重要**

## • t検定のp値

- **t検定のp値は、2つの標本の差が偶然である確率**
- **p値が低いとき、差が統計的に有意である可能性が高まる**

## • p値の解釈

- **p値が小さいとき、「差は統計的に有意であり、偶然とは考えにくい」と考える**
- **p値が大ききときは、「差は統計的に有意であるとは言いきれない。偶然であるとも、偶然でないとも言えない」と考える**

# Rでのt検定



R システム : `t.test(<標本 1>, <標本 2>, var.equal=F)`

```
> t.test( c(128, 104, 124, 85, 120), c(100, 106, 89, 89, 105), var.equal=F )  
  
welch Two Sample t-test  
  
data:  c(128, 104, 124, 85, 120) and c(100, 106, 89, 89, 105)  
t = 1.6434, df = 5.6914, p-value = 0.1541  
alternative hypothesis: true difference in means is not equal to 0  
95 percent confidence interval:  
 -7.325827 36.125827  
sample estimates:  
mean of x mean of y  
 112.2      97.8
```

p値が表示される

```
t.test( c(128, 104, 124, 85, 120), c(100, 106, 89, 89,  
105), var.equal=F )
```

# t 検定



```
> t.test( c(128, 104, 124, 85, 120), c(100, 106, 89, 89, 105), var.equal=F )  
  
welch Two Sample t-test  
  
data: c(128, 104, 124, 85, 120) and c(100, 106, 89, 89, 105)  
t = 1.6434, df = 5.6914, p-value = 0.1541  
alternative hypothesis: true difference in means is not equal to 0  
95 percent confidence interval:  
-7.325827 36.125827  
sample estimates:  
mean of x mean of y  
112.2 97.8
```

```
> t.test( c(128, 104, 124, 85, 120), c(180, 190, 189, 131, 130, 150), var.equal=F )  
  
welch Two Sample t-test  
  
data: c(128, 104, 124, 85, 120) and c(180, 190, 189, 131, 130, 150)  
t = -3.5422, df = 8.4866, p-value = 0.006908  
alternative hypothesis: true difference in means is not equal to 0  
95 percent confidence interval:  
-81.35114 -17.58219  
sample estimates:  
mean of x mean of y  
112.2000 161.6667
```

■ p 値 < 0.05 が、判断の分かれ目の目安という考え方も

# 演習の例



1. サイズが 5 以上の数値データを, 2個準備しなさい

データ 1	
データ 2	

2. 1のデータについて t 検定を行い, その p 値を求めなさい

<p値>