

2016年度秋学期 統計学 第10回

分布の推測とは —  
標本調査, 度数分布と確率分布

浅野 晃  
関西大学総合情報学部



## 「統計学」の後半は 統計的推測

### 「統計的推測」とは

ここまでは

データを度数分布で整理する  
度数分布を要約する (平均・分散)  
**記述統計学**

調べたいデータ全体を調べられるか？

### 「統計的推測」とは

調べたいデータ全体を調べられるか？

日本男性全員の身長を調べられるか？

データ全体の数値をすべて調べるのは、  
費用や時間がかかる

調べると、壊れてしまうものもある  
料理をすべて味見してしまったら、  
食べるものがなくなってしまう

## 「統計的推測」とは

調べたいデータ全体を調べられるか？

日本男性全員の身長を調べられるか？

データの一部を調べて

度数分布を推測する

いや、せめて平均や分散を推測する

**統計的推測**

2016年度秋学期 統計学

統計的推測の基本は  
「くじびき」

## 無作為抽出

統計的推測は、

集団のデータ全体を調べていないのに、  
集団全体のように調べようとする

結果が間違っている可能性がある

こんな人たちばかり

(バレーボール選手の写真)

選んでしまったら

「日本人はすごく背が高い？」

<http://www.asahi.com/articles/OSK201311200083.html>

2016年度秋学期 統計学

## 無作為抽出

わざわざ背の高い人ばかり選ぶことはない  
高低まんべんなく選べば、  
その平均は集団の平均とだいたい同じ

それはそうだけど

集団にどんな人がいるか何も知らないのに

選ばれた人が、集団のなかで  
背が高いか低いかなどわからない

2016年度秋学期 統計学

## 無作為抽出

集団にどんな人がいるか何も知らないのに  
選ばれた人が、集団のなかで  
背が高いか低いかなどわからない

「まんべんなく選ぶ」のは無理

なので

公平なくじ引きで選ぶ

## 無作為抽出

集団からくじびきで選ぶと

偶然こんな (バレーボール選手の写真) 人たちばかり

<http://www.asabi.com/articles/OSK201311200083.html>

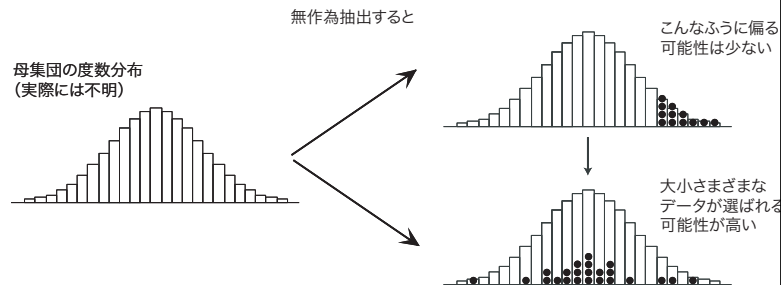
選んでしまっただけで、おかしな結果になる  
可能性がないわけではないけれど、

そうなる確率は小さい

その確率も  
計算できる。

## 無作為抽出

集団からくじびきで選ぶと



## 無作為抽出

データ全体から、いくつかの数値を  
公平なくじびきで選ぶ

【無作為標本抽出】という

調べたい (が全部を調べるの  
は無理な) 集団 【母集団】

調べられる程度のデータ  
【標本 (サンプル) 】

## 標本「サイズ」

「母集団」や「標本」という言葉は、  
「データ」と同様、数値の**集まり**をさす  
(1つ1つの数値ではない)

母集団も標本も、その中に含まれる  
数値の個数を**大きさ (サイズ)**という  
( [標本サイズ] とはいうが、  
標本数とはいわない)

家族(family)という言葉に似ている

2016年度秋学期 統計学

## 度数分布と確率分布

## 度数分布と確率分布

標本を無作為抽出するとき  
**ある数値が出てくる確率が**  
**どのくらいになるか**

さっきの

「偏った数値ばかり選んでしまう」確率を  
求めるのにも必要

2016年度秋学期 統計学

## 「公平なくじびき」と当たり確率

(回転抽選機の写真)

この中に入っている  
当たりくじの割合が  
**20%**とする

<http://epshop.net/epkyoto/7.1/15001/>

くじを1回ひいて、  
当たる確率は？ **20%** **本当？**

2016年度秋学期 統計学

## 「公平なくじびき」と当たり確率

(回転抽選機の写真)

当たりくじの割合が  
**20%**なら、  
 当たる確率も  
**20%**

<http://epshop.net/epkyoto/7.1/15001/>

これが本当であるためには、

- どのくじも同じ確率で選ばれる
- ある回のくじびきの結果が、他の回に影響しない (独立)

2016年度秋学期 統計学

## 「公平なくじびき」と当たり確率

(回転抽選機の写真)

どのくじも同じ確率で  
 選ばれるのなら、

<http://epshop.net/epkyoto/7.1/15001/>

くじの総数のうち20%が当たり  
 →当たりが出る確率は20%

(ラプラスの確率の定義)

2016年度秋学期 統計学

## 度数分布で考えると

母集団の度数分布

階級値	相対度数
⋮	
162.5	15%
167.5	20%
172.5	<b>20%</b>
177.5	10%
⋮	

無作為抽出 →

階級値  
 172.5の  
 人が選ば  
 れる確率  
 は  
**20%**

2016年度秋学期 統計学

## 度数分布で考えると

どの階級についても同じだから

母集団の度数分布

標本の [確率分布]

階級値	相対度数	無作為抽出 →	階級値	選ばれる確率
162.5	15%		162.5	15%
167.5	20%		167.5	20%
172.5	20%		172.5	20%
177.5	10%		177.5	10%

2016年度秋学期 統計学

## 確率分布と確率変数

つまり

母集団の度数分布  
(母集団分布)

=

標本の確率分布

階級値	選ばれる確率
162.5	15%
167.5	20%
172.5	20%
177.5	10%

いくらとは  
決まってい  
ないが、

確率分布が  
決まっている

【確率変数】  
という

2016年度秋学期 統計学

## 何が知りたいのか

母集団の度数分布が  
知りたい

標本の確率分布を推定すればよい

標本の確率分布,  
推定できる？

くじを1本だけひいても、  
当たり確率はわからない  
どうする？

2016年度秋学期 統計学

## 復元抽出と非復元抽出

- どのくじも同じ確率で選ばれる
- ある回のくじびきの結果が、  
他の回に影響しない (独立)

こうであるためには、

出たくじをすぐに箱に戻す  
【復元抽出】

(回転抽選機の写真)

<http://epshop.net/epkyoto/7.1/15001/>

実際にはあまりやらない…

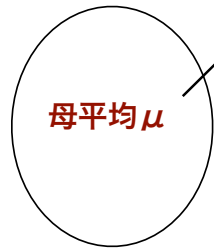
2016年度秋学期 統計学

## 標本平均と母平均

A.Asano, Kansai Univ.

## 母平均の推定

母集団  
(日本男性全体)



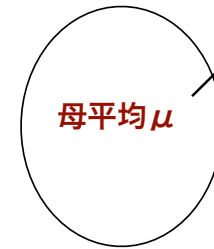
標本として数値を  
いくつか取り出して、  
それらの平均  
[標本平均]

標本平均は母平均に  
近い値になるか？

母平均が知りたい  
が、日本男性全員は調べられない

## 母平均の推定

母集団  
(日本男性全体)



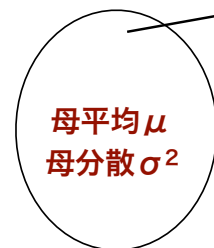
[標本平均]

標本平均は母平均に  
近い値になるか？

もし偏った標本が得られて  
いたら、標本平均は母平均  
と大きく食い違うことに

## 母平均の推定

母集団

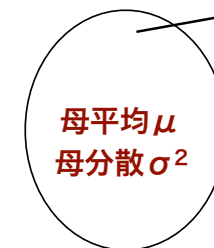


サイズnの標本1セット      標本平均  
X1 X2 ... Xn                       $\bar{X}$

仮に、何度も標本を  
抽出したとしたら？

## 母平均の推定

母集団



サイズnの標本1セット      標本平均  
X1 X2 ... Xn                       $\bar{X}$

X1	X2	...	Xn	$\bar{X}$
X1	X2	...	Xn	$\bar{X}$
⋮	X1のさまざまな可能性			⋮

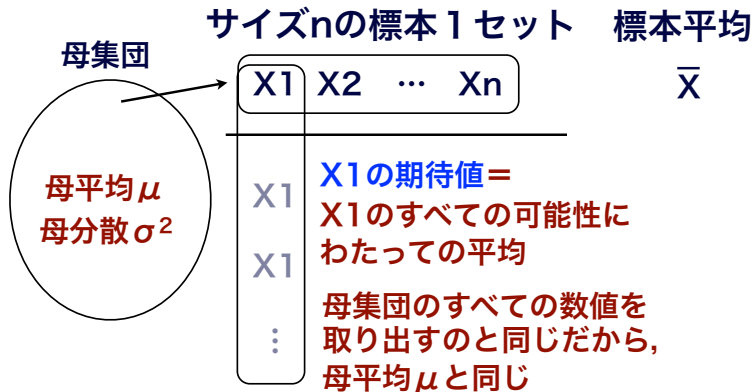
[期待値]  $\mu$   
分散  $\sigma^2$

期待値？

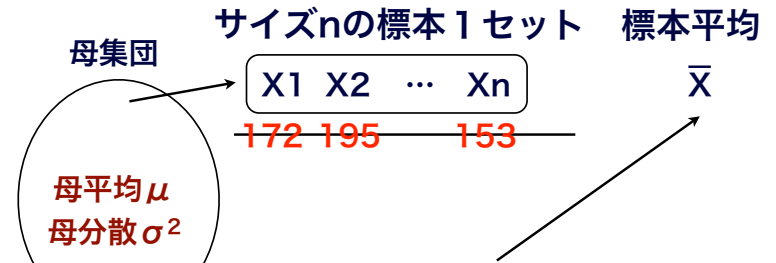
# 期待値とは

平均の一種

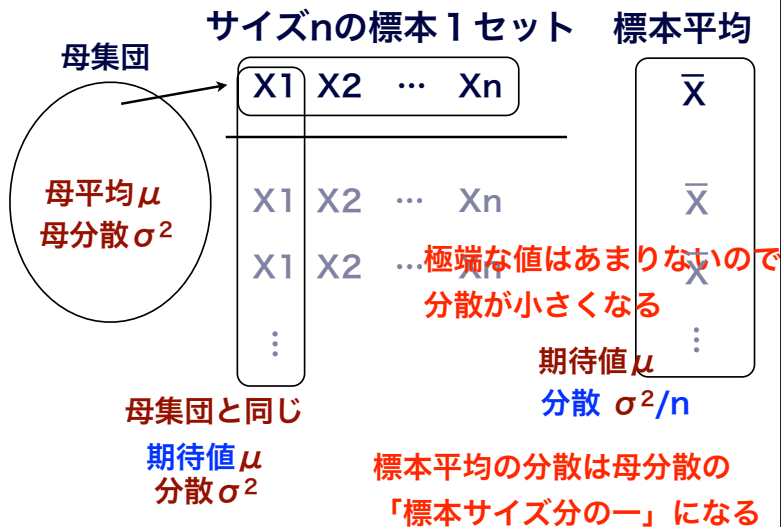
「すべての可能性にわたっての平均」



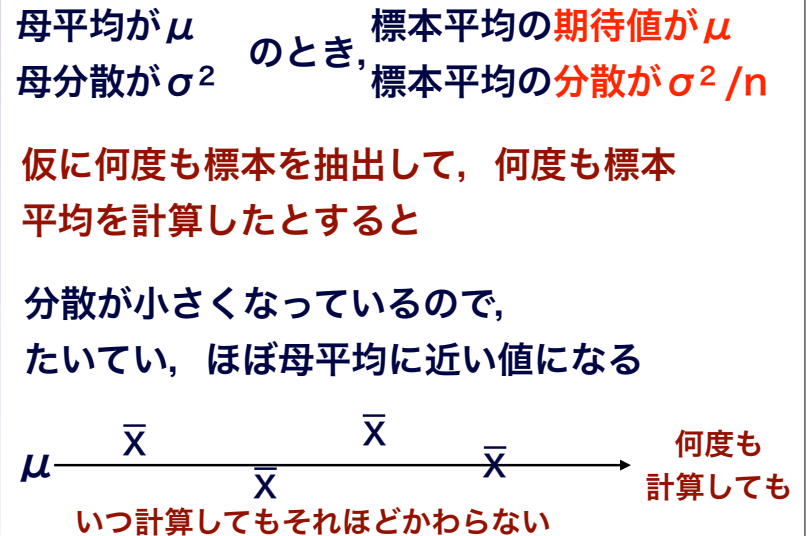
# 標本平均は



# 母平均の推定



# 母平均の推定





## 母平均の推定

母平均が  $\mu$  のとき、  
母分散が  $\sigma^2$  のとき、  
標本平均の期待値が  $\mu$   
標本平均の分散が  $\sigma^2/n$

仮に何度も標本を抽出して、  
何度も標本平均を計算したとすると

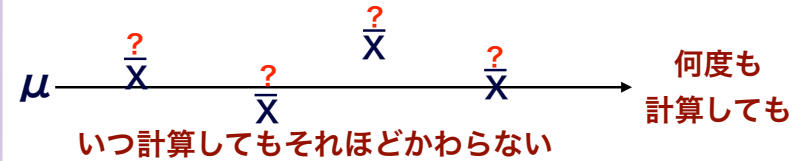
分散が小さくなっているの  
で、たいてい、ほぼ母平均に  
近い値になる

いま1回だけ計算した標本平均も、  
おそらく、ほぼ母平均に近い値  
だろう

2016年度秋学期 統計学

## 母平均の推定

いま1回だけ計算した標本平均も、  
おそらく、ほぼ母平均に近い値  
だろう



いま計算した標本平均は  
上のどれにあたるかはわからないが、  
いずれにせよあまりかわらない

2016年度秋学期 統計学

## 母平均の推定

いま1回だけ計算した標本平均も、  
おそらく、ほぼ母平均に近い値  
だろう

どのくらい近い？

どのくらいの確率で？  
はずれる確率は？

このあたりは  
次回。

2016年度秋学期 統計学