

2025年度春学期

# 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



関西大学総合情報学部  
浅野 晃

「統計学」の後半は  
統計的推測

## 「統計的推測」とは

ここまで

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

最近はそうでもないのでは…(ぼそ)

その通りで、「ビッグデータ」という言葉がよく聞かれたこともあり、さらに今は「機械学習」もさかんになりました。

## 「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？  
※なぜ例が「男性」なのは、もう少し先で。

データを集める手間は劇的に減ったけれど

測定作業の手間や費用は変わらない

それに、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

2025年度春学期 統計学／関西大学総合情報学部 浅野 光 5 | 37

## 「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

2025年度春学期 統計学／関西大学総合情報学部 浅野 光 6 | 37

統計的推測の基本は  
「くじびき」🎯

## 偏った抽出をしてしまうと

統計的推測は、

集団のデータ全体を調べていないのに、  
集団全体のようすを調べようとする

結果が間違っている可能性がある

バレーボール🏐やバスケットボール🏀の選手ばかり選んでしまったら  
「日本人はすごく背が高い？」

2025年度春学期 統計学／関西大学総合情報学部 浅野 光 8 | 37

## 「まんべんなく」抽出できるか？

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

## くじ引きで選ぶと

集団からくじびきで選ぶと

偶然、 選手のような人ばかりを選んでしまって、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。

## 公平なくじ引きで選ぶ

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

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

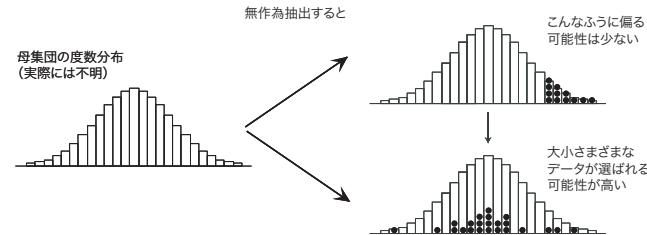
[無作為標本抽出]あるいは  
[無作為抽出]という

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)】

## 無作為抽出

集団からくじびきで選ぶと



★たくさんの人を抽出すると、偏らないか？

無作為抽出なら、そう期待できる。(今日の後半)

無作為抽出でなければ、必ずしもそうではない。

(ツイッターのTLは「鏡に映った自分の意見」)

2025年度春学期 統計学／関西大学総合情報学部 浅野 光 13 | 37

## 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、標本数やサンプル数とはいわない

家族(family)という言葉に似ている

※「母集団のサイズ」を母数とはいいません。母数は別の意味です(よくある誤り)

※「サンプル数」という誤った表記をよく見かけますが、標本サイズは数えられる  
程度の数であることが多いからでしょうか。

2025年度春学期 統計学／関西大学総合情報学部 浅野 光 14 | 37

## 度数分布と確率分布

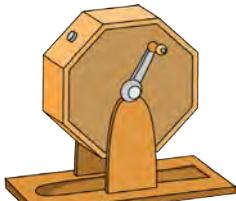
### 度数分布と確率分布

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

2025年度春学期 統計学／関西大学総合情報学部 浅野 光 16 | 37

## 「公平なくじびき」と当たり確率



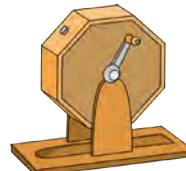
この中に入っている当たりくじの割合が  
**20%**とする

[https://illpop.com/png\\_season/dec01\\_a07.htm](https://illpop.com/png_season/dec01_a07.htm)

くじを1回ひいて、当たる確率は？ **20% 本当？**

2025年度春学期 統計学／関西大学総合情報学部 浅野 光 17 | 37

## 「公平なくじびき」と当たり確率



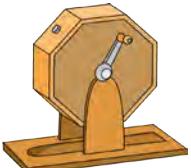
当たりくじの割合が**20%**なら、  
当たる確率も**20%**

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)

2025年度春学期 統計学／関西大学総合情報学部 浅野 光 18 | 37

## 「公平なくじびき」と当たり確率



どのくじも同じ確率で選ばれるのなら、  
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)

2025年度春学期 統計学／関西大学総合情報学部 浅野 光 19 | 37

## 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%
<b>172.5</b>	<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%**

2025年度春学期 統計学／関西大学総合情報学部 浅野 光 20 | 37

## 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

階級値	相対度数
162.5	15%
167.5	20%
172.5	20%
177.5	10%

### 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%

無作為抽出

2025年度春学期 統計学／関西大学総合情報学部 浅野 光 21 | 37

## 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

標本は、  
値がいくらになるかは決まっていない  
しかし確率分布が決まっている  
(知っているかどうかは別)

こういう数を[確率変数]という  
(中国語では隨機変数)

「標本は、確率変数(の一種)である」

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

2025年度春学期 統計学／関西大学総合情報学部 浅野 光 22 | 37

## 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？

それは、「くじの結果から当たり確率を推定する」と同じ

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？

2025年度春学期 統計学／関西大学総合情報学部 浅野 光 23 | 37

## 標本平均と母平均



## 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  
  
母平均  $\mu$   
母分散  $\sigma^2$

サイズ  $n$  の標本1セット 標本平均  
 $X_1 \ X_2 \ \dots \ X_n$   $\bar{X}$

仮に、何度も標本を抽出したとしたら？

## $X_1$ の期待値と分散は

母集団  
  
母平均  $\mu$   
母分散  $\sigma^2$

サイズ  $n$  の標本1セット 標本平均  
 $X_1 \ X_2 \ \dots \ X_n$   $\bar{X}$

$X_1 \ X_2 \ \dots \ X_n$   $\bar{X}$

$X_1 \ X_2 \ \dots \ X_n$   $\bar{X}$

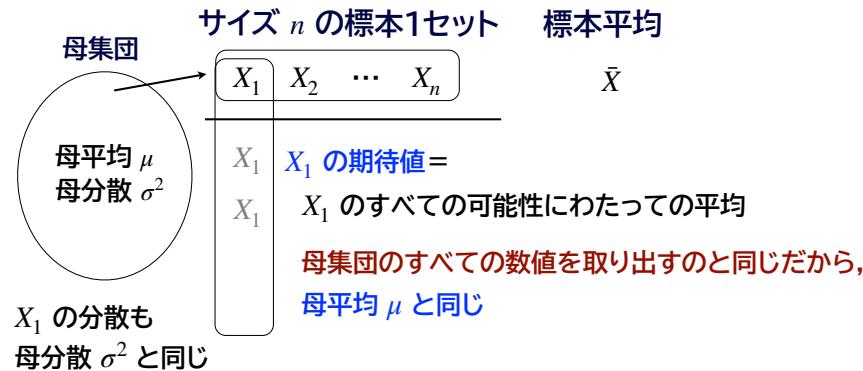
$\vdots$   
[期待値]  $\mu$   
分散  $\sigma^2$

$X_1$  のさまざまな可能性  
その平均を、

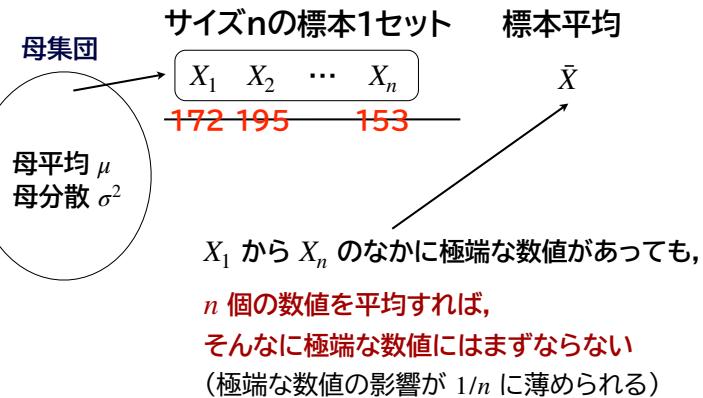
期待値とは？

## 期待値とは？

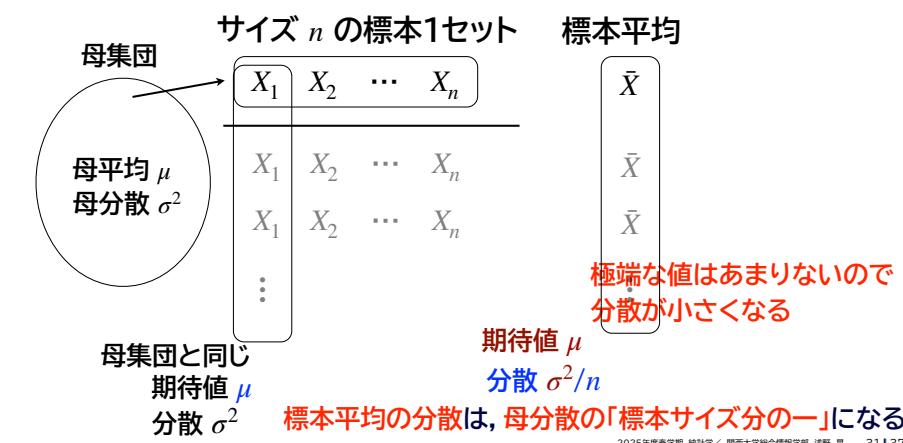
期待値は平均の一種で「すべての可能性にわたっての平均」



## 標本平均の期待値と分散は



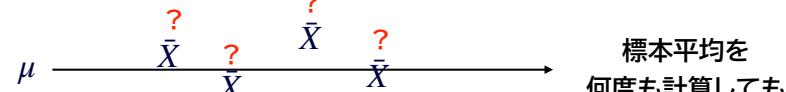
## 標本平均の期待値と分散は



## 母平均の推定

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



いつ計算しても、たいていそれほど変わらない

いま1回だけ計算した標本平均は、上のどれなのかわからないが  
たいてい、ほぼ母平均に近い値だろう

## 母平均の推定へ

いま1回だけ計算した標本平均は、  
「たいてい、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？

このあたりから次回へ

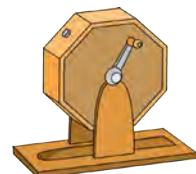
理想的な無作為抽出とは

## 復元抽出と非復元抽出

理想的な無作為抽出は、「公平なくじびき」

- ・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]

## 「標本の大きさ」の意味

母分散が  $\sigma^2$  のとき、標本平均の分散が  $\sigma^2/n$

標本平均の分散に関係しているのは  
**標本の大きさ**であって、母集団の大きさは関係ない

推測の確かさに影響するのは  
**標本の大きさ**であって、  
**標本の大きさの、母集団の大きさに対する割合 ではない**

## 標本の大きさとは

「10人からなる標本」の意味は、  
1,000人からなる母集団でも100,000人からなる母集団でも同じ

🤔...



理想的な無作為抽出では、復元抽出を行う  
標本サイズは、  
「取り出された数値の個数」というよりも  
「同一の母集団から数値ひとつひとつを取り出す回数」  
→ 母集団の大きさに対する割合は無関係

(非復元抽出をした場合は、計算で補正する方法がある)