

### 第3回 母集団と標本・確率分布

#### A. 母集団と標本

##### 1. 部分（標本）から全体（母集団）を推測

100個取ったデータからデータを取った集団より大きな集団の性質を推測できないか？

松江市にあるA小学校6年生の男子生徒100人について

- 1) 50m走のタイムを計った
- 2) 社会科の試験を行った
- 3) 父親の年収を調べた

どこまで推論を拡張してもよいだろうか？たとえば、来年、女子、中学生、日本全体、世界・・・

##### 2. 記述統計と推測統計

記述統計

推測統計

農学研究ではおもに推測統計に興味が置かれる。では記述統計と推測統計の違いは対象をどうとらえることによって生じるのか？

以下の①から⑤は記述統計か？推測統計か？

- ① 日本の人口は2004年現在、1億2千万人強である。
- ② 日本の人口は2050年には1億人になる。
- ③ 大山に2007年2月1日に29cmの降雪があった。
- ④ 大山には冬、雪が降る。
- ⑤ この風邪薬はよく効く。
- ⑥ この肥料をブロッコリーに施肥したところ、ビタミンCが50%増加した。
- ⑦ この肥料はブロッコリーの品質を向上させる。

記述統計とは標本をまとめたり、記述したりするためのものである。これに対して推測統計は、標本に基づき、より広い母集団に対して推定、推測を行うためのものである。

### 3. 母集団と標本とは？

母集団

標本

★ 国勢調査のように母集団と標本が一致する場合もごくまれにあるが、ほとんどのばあい母集団の一部を取り出して、標本にする。

### 4. 実在母集団と仮説的無限母集団

#### ① 実在母集団

#### ② 仮説的無限母集団

農学，生物学の対象である生物は無限母集団として，扱うのが普通である。

次の①から⑥について，それぞれ母集団は何か？標本は何か？母集団は実在母集団か，仮説的無限母集団かを考えよ。

① ある市の市長選で有権者 500 名を対象にだれに投票するかを調査した。

母集団 ( ) 標本 ( )

実在母集団・仮説的無限母集団

② あるテレビ局の番組の視聴率について，20 歳代 300 人に尋ねた。

母集団 ( ) 標本 ( )

実在母集団・仮説的無限母集団

③ ある食品工場では，工場排水の水質保証のために 1 ヶ月に 6 回の割合で放流口から採水して排水分析をしている。

母集団 ( ) 標本 ( )

実在母集団・仮説的無限母集団

④ 羊に与えるえさに混ぜると成長がよくなるという薬を 5 匹の羊に与え，調査した。

母集団 ( ) 標本 ( )

実在母集団・仮説的無限母集団

⑤ ある製薬会社は血圧を低下させる薬の候補であるAを20匹のラットに与え、効果があることを明らかにしたので人間への応用を考えている。

母集団 ( ) 標本 ( )

実在母集団・仮説的無限母集団

⑥ ある家電メーカーでは、エアコン製造ラインで種々の不良が発生している。そこで工場から200枚の基盤を調べたところ、11枚の不良品を発見した。

母集団 ( ) 標本 ( )

実在母集団・仮説的無限母集団

★ 統計解析とは標本を調べることによって母集団を推定するのが目的である。

実在母集団の場合：松江市にあるA小学校6年生の男子生徒100人について

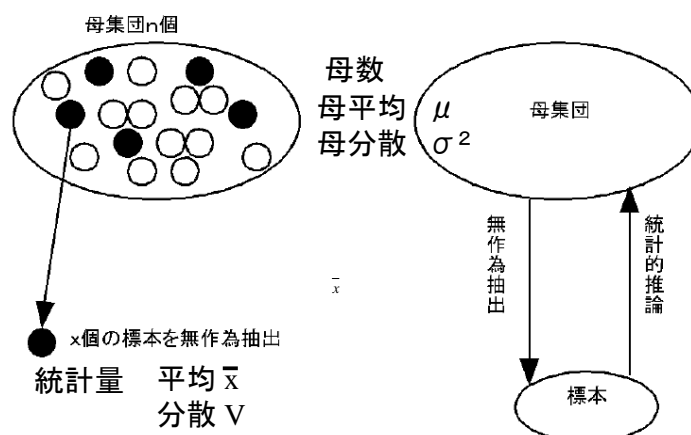
- 1) 50m 走のタイムを計った
- 2) 社会の試験を行った
- 3) 父親の年収を調べた
- 4) 風邪薬の試験をした

一方、農業や生物の研究対象は仮説的無限母集団であることが一般的である。

仮説的無限母集団の場合：ヘレフォード種のウシ（5歳）に成長のよくなる薬を混ぜた飼料を与えた。どのような集団にまで一般化してよいだろうか？

- 1) 年齢の違うヘレフォード種は？
- 2) 他の品種は？
- 3) 他の反芻動物は？

母集団の特徴を表す数値を母数 (parameter) という。母平均  $\mu$ ，母分散  $\sigma^2$  など、ギリシャ文字を使用する。これに対して標本のそれを統計量 statistic という。標本平均  $m$ ，標本分散  $V$  などローマ字を使用する。統計解析とは母数 (parameter) を推定するために集めた標本の統計量から母数を推定することである。わたしたちが知りたいのは母集団 (母数) である。しかし、実際に知ることができるのは標本 (統計量) である。



#### 4. 標本の抽出

★ 母集団を代表するように標本を抽出しなければならない。

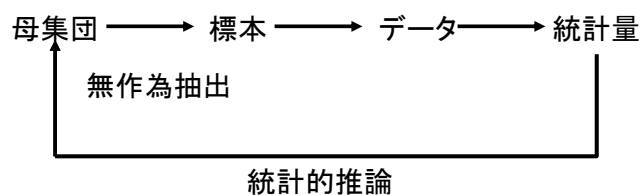
★ 統計学は標本を無作為に抽出することを要請する。

無作為 (at random) にサンプリングしなければならない 注意! でたらめとはちがう

無作為とは母集団のどの要素も等しい確率で抽出されることをいう

乱数表, さいころ, 乱数さい (正二十面体のさいころ), トランプ, コンピュータの疑似乱数

無作為のあり方も専門的知識が必要である。



★ 無作為抽出の例

① 松江市市長選を予想するために乱数表を使って, 500人の有権者を抽出した。

② K牧場の牛250頭のうち, 乱数さいで10頭を選んで, 血液検査した。

③ 瓶詰めの牛乳の脂肪価の検査で, よくかき混ぜた後ピペットを縦に入れ, 上層から下層までをサンプリングした。

④ ベルトコンベアで運ばれる製品をランダムな時間間隔でサンプリングした。

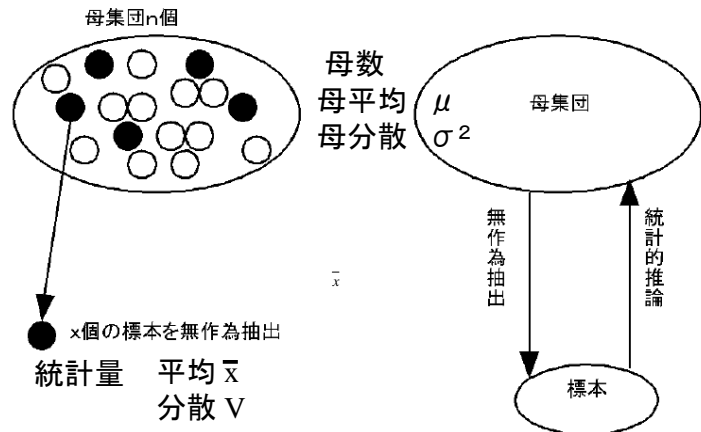
以上のように標本は母集団を推定するために調査する。標本の統計量から母集団の母数を推定するのが統計学である。しかし, まずここでは母集団がすでにわかっているものとして, 標本がどうなるかを考えよう。母集団がすでにわかっているとしたら, 標本がどうなるのかは確率的に論じることができる。例: 平均余命

B. 確率分布

1. 統計と確率

① 先験的確率

実験を行わないでも求められる確率



② 経験的確率

大数の法則に基づいて，統計から定義される確率である。

母集団がわかれば，無作為に取り出した標本がどのような値を取るかを予想できる．それを与えるのが確率分布である．

2. 確率変数と確率分布の概念

確率変数 どのような値を取るかが確率で決まっている変数 必ずしも数字とは限らない。

例 さいころを振るときの確率変数は

トランプを引くときの確率変数は

天気予報の確率変数は

確率分布 確率変数がどのような値になるかという分布を示したもの

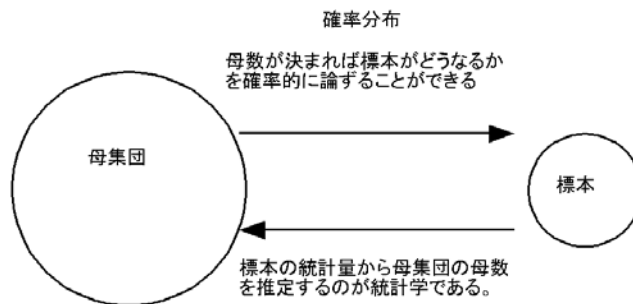
3. 離散分布と連続分布

① 離散分布 確率変数がとびとびの値（離散量）を取る。

例 さいころを振るときの確率分布 1, 2, 3, 4, 5, 6 の6つの確率変数を取る。

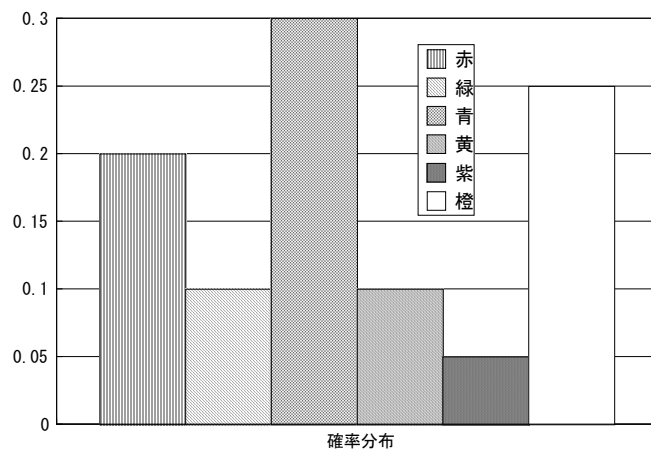
② 連続分布 確率変数が連続的な値（連続量）を取る.

例 犬の体重, 人の身長



③ 離散分布の例

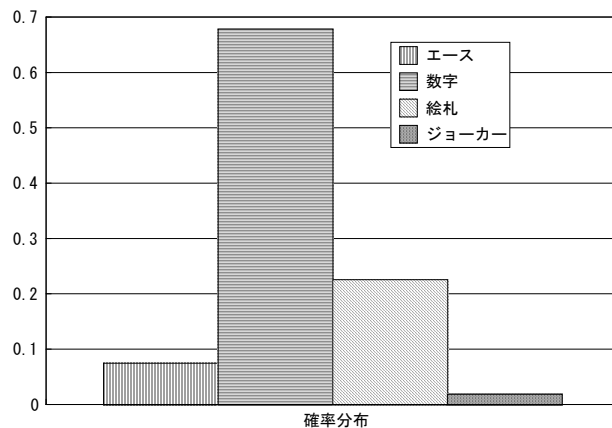
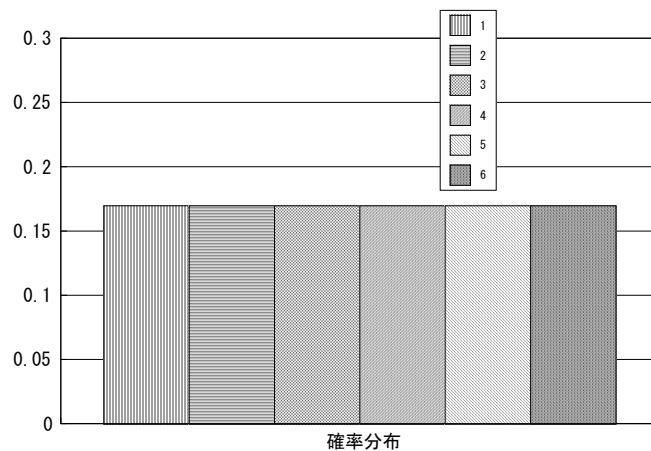
★ ある花は遺伝子の組成から図のような確率で花の色が決まっている. 花の色が確率変数であり, それぞれの確率を示した確率分布が右図である.



例 花の色の確率分布

さいころを振るときの確率分布

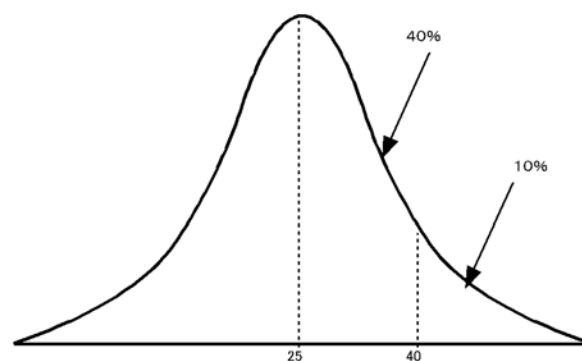
トランプでエース, 数字, 絵札, ジョーカーを引く確率分布



★ 確率分布は母集団について考えること。標本はあくまでも母集団の確率分布を予想するために調査する。

④ 連続分布の例

連続分布の場合の確率は面積の割合で表示される。  
したがって、ある値を取る確率はゼロである。  
右の例はある血統の犬の体重の確率分布である。



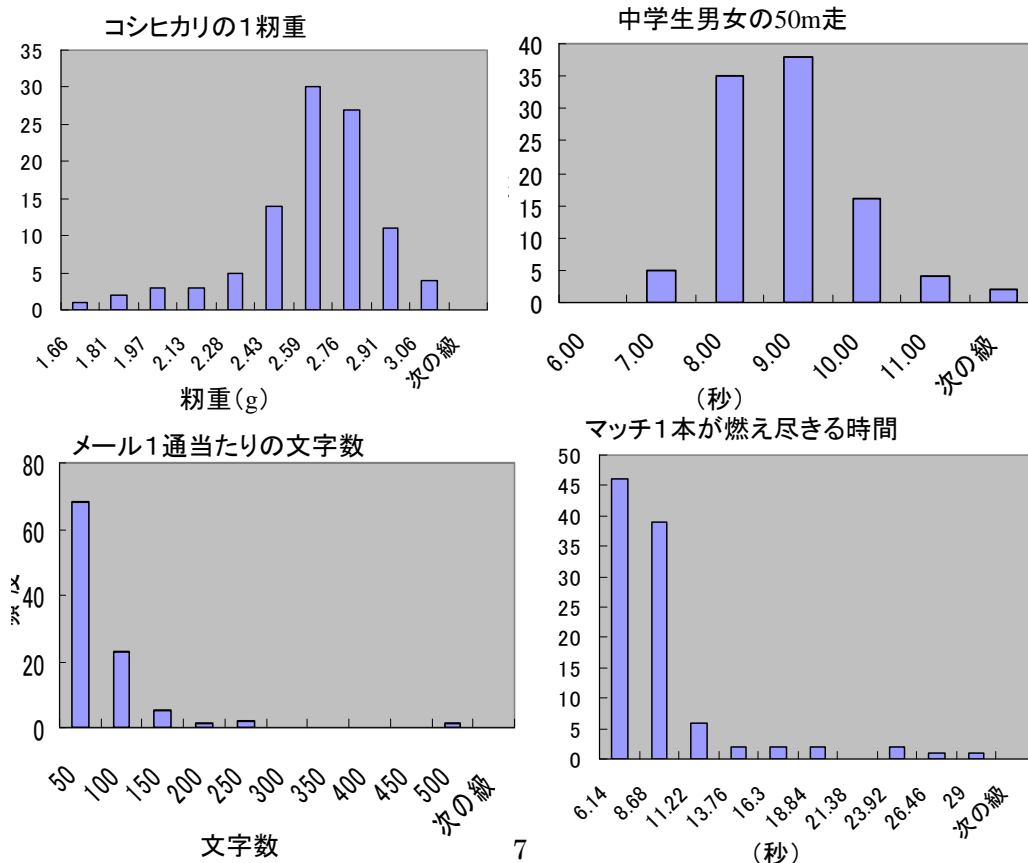
犬の体重の確率分布

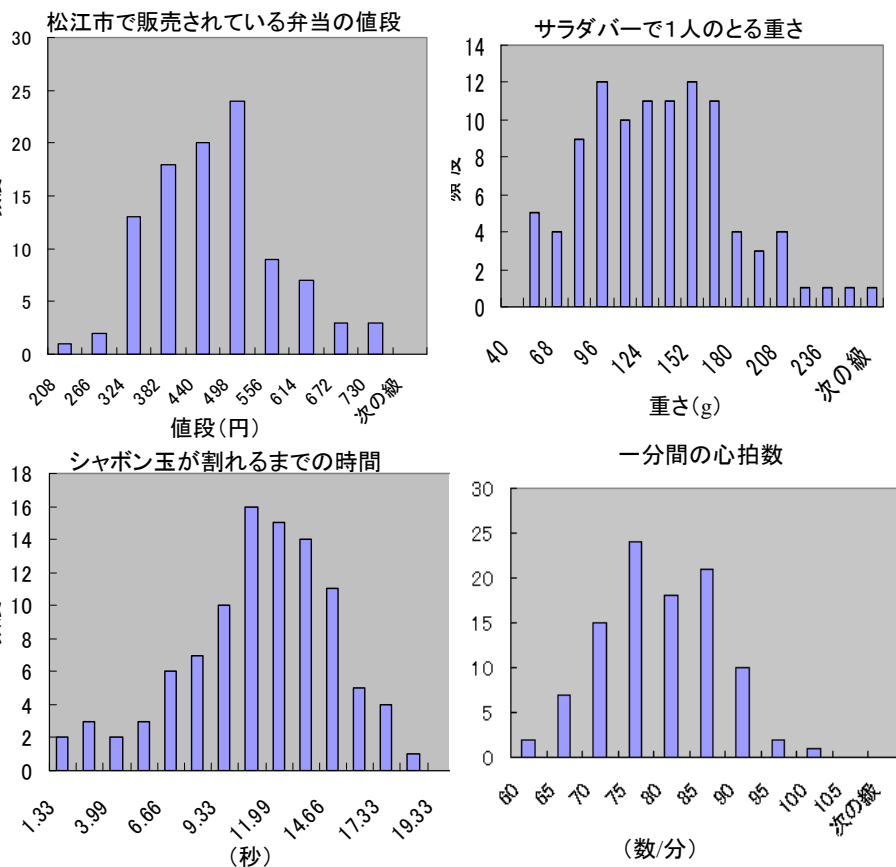
確率分布がわかれば標本がどのような確率で得られるかがわかる。逆に標本と確率分布のパターンがわかれば確率分布を再現できるのではないか。では確率分布にパターンはあるのだろうか？

C. 分布の分類

1. 昨年までの学生の宿題で集まったデータから作ったヒストグラムをみて、分類してみよう

- ① 元の母集団は実在母集団か、仮説的無限母集団か。さらにその確率分布は離散分布か連続分布か？
- ② 形に注目 左右対称、山が一つかそれ以上かなど
- ③ 母集団を想像しながら、分布の分類 ここからある値を取る確率を読みとってみよう。





#### D. 宿題

- 第1回の授業の宿題で調べたデータの母集団は何か？また標本は何か？さらにこの母集団は実在母集団か，仮説的無限母集団か？この第1回のデータについて，どのくらい大きなレベルの母集団にまで利用できるかを考えてみよ。
- 第1回の授業の宿題で調べたデータの母集団は離散分布か，連続分布か？その分布の形は左右対称か，あるいはどちらかにゆがんでいるか？ひと山かふた山か？さらに確率がおおよそ0.2 およびおおよそ0.5になる確率変数（範囲でもよい）を，いちばん階級数の多いヒストグラム（15階級程度に分類したヒストグラム）について考えてみよ．ホームページに掲載された過去の宿題のヒストグラムの中から，自分のデータともっとも分布の形が異なると思ったデータおよびもっとも分布の形が似ていると思ったデータのヒストグラムをコピーして載せよ。
- 来週は統計学においてもっとも基本的な連続分布である正規分布を学ぶ．正規分布に従うと考えられる事象はたくさんある．以下のものは正規分布にほぼ従う．以下のものあるいはそれに類するものについて，100個以上のサンプルを調査し，その平均，分散を求め，ヒストグラム（階級の数は7~10程度）を書け．ノートには元となったデータもすべてきちんと載せること．データの有効数字は3つ以上とする．
  - ★ 卵の重さ，イネの草丈，トマトの酸度，ブドウの糖度
  - ★ 個装したチョコレート菓子，1個1個の重さ（mgまで精密にはかる）

なお数人のグループで調べてかまわない。