

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

第 1 回の授業では池にすむ魚の平均体重を調べるために、池の中の魚の一部を捕獲して、そこから推測することの是非について論じた。このように部分（統計学では標本という）から全体（母集団という）を推測することはよく行われることである。しかし、部分から全体への一般化はどれほど危険を伴うことだろうか？以下の例を考えてみよ。

例：クロールのタイムがほぼ同等であった日本人の 15 歳の男子 40 人を均等に 2 つのグループに分け、一方の 20 人に新しいクロールの教え方を適用したところ、従来の教え方を適用したグループ 20 人よりも 50m のタイムが向上した。

では、この結果は

- ① 日本人の 15 歳の男子全体, ② 日本人の 13～16 歳の男子全体, ③ 日本人の 15 歳男女とも, ④ 日本人男子全体, ⑤ アメリカ人の 15 歳の男子全体, ⑥ 砂漠に住む 15 歳の男子全体, ⑦ すべての人間・・・に適用できるだろうか？

このようにある事実あるいはある実験結果をどこまで推論を拡張してもよいだろうか？

2. 記述統計と推測統計

（ ）統計 観測値を要約し、記述するための方法

（ ）統計 観測値を推定や予測をするための基礎として用いる方法。
すなわちこれまでに観測されていない状況について推測する方法

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⑦は記述統計か？推測統計か？を考えてみよう．

- 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 () 母集団
理論的にはすべてを調査することのできる母集団
例：日本人全員, 地球上の海水すべて
- ② () 母集団
母集団に属する要素が無限にある母集団
例：工場で作られる製品, 医薬品の効果を調べる対象であるヒト

農学, 生物学の対象である生物は仮説的無限母集団として, 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。
しかし、母集団と標本の関係は理論では明確であっても現実にはそう簡単ではない。

サンプリングのパラドックス（新・涙なしの統計が P17 参照）ということばがある。標本はそれが母集団を代表していない限り、誤った理解を導くものとなる。しかし、その標本をとる前に、母集団がわかっているわけではなから、標本を正しく（無作為に）とったといえるだろうか？中身の分からない母集団から、適切に標本をとれるはずがあろうか？

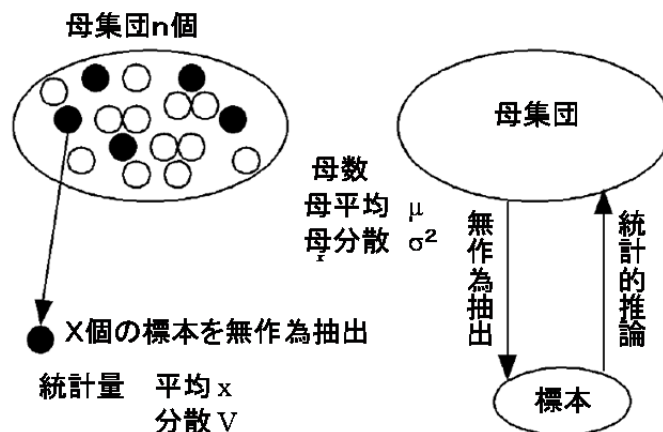
さらにある医者が高血圧の患者が何名かやってきた。そこから高血圧の患者全体への推論ができるだろうか？医者に来る患者に何らかの偏りがあるかもしれない。

質問 それぞれの医者の事例についてどんな偏りが考えられるだろうか？

- ① 高血圧専門の医者なら・・・
② 小児科医なら・・・

この医師は自分の標本（やってきた患者）にふさわしい母集団を考えることになる。

母集団の特徴を表す数値を母数 (parameter) という。母平均 μ ，母分散 σ^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

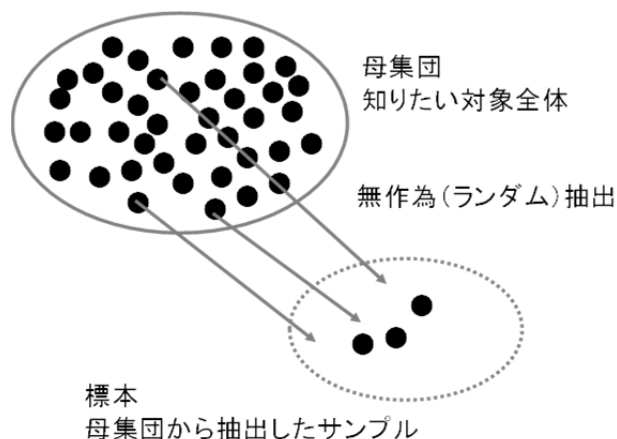
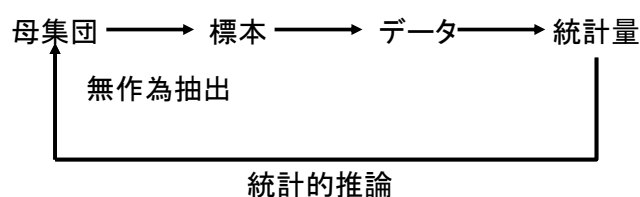
★ 母集団を代表するように標本を抽出しなければならない。

★ 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。

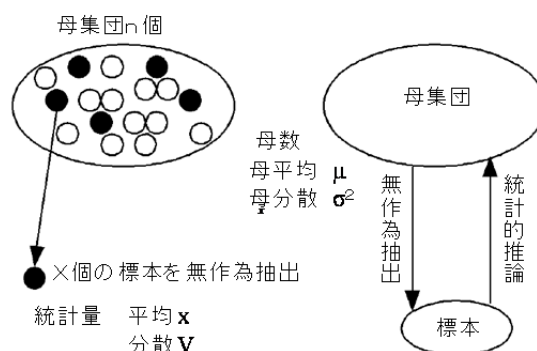
以上のように標本は母集団を推定するために調査する。標本の統計量から母集団の母数を推定するのが統計学である。しかし、まずここでは母集団がすでにわかっているものとして、標本がどうなるかを考えよう。母集団がすでにわかっているとしたら、標本がどうなるのかは確率的に論じることができる。

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

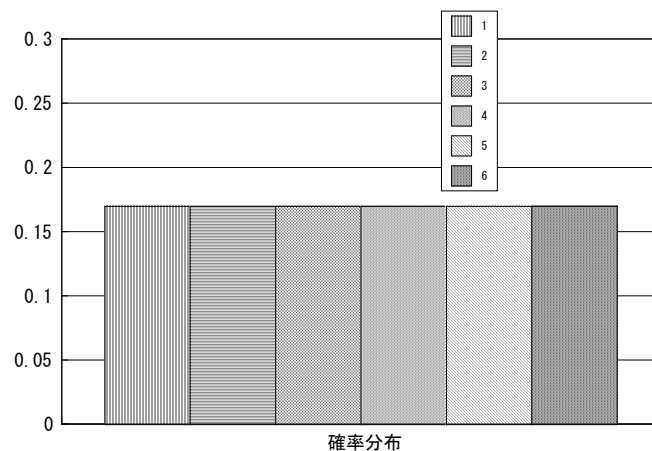
天気予報の確率変数は

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．

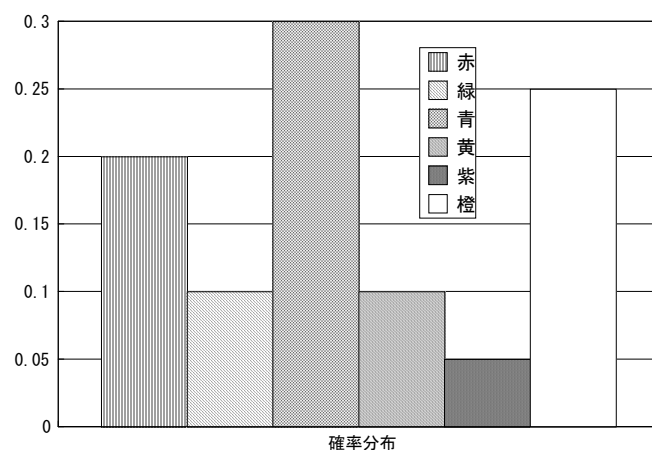
★ さいころを振るときの確率分布 1, 2, 3, 4, 5, 6 の 6 つの確率変数を取る．それぞれの確率変数の起こる確率はどれも $1/6$ である．



★ 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．

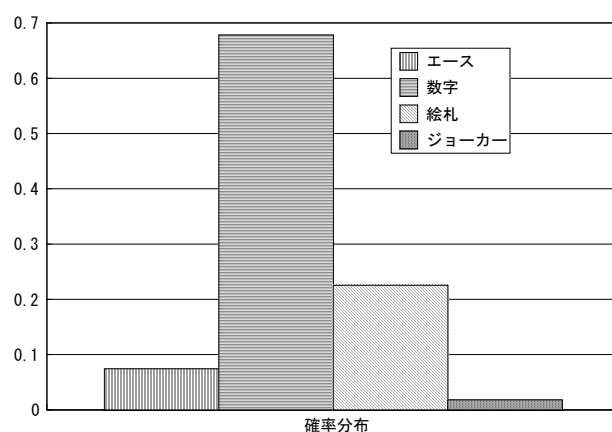
赤い花が咲く確率は？

緑の花が咲く確率は？



トランプでエース，数字，絵札，ジョーカーを引く確率分布

数字を引く確率は？



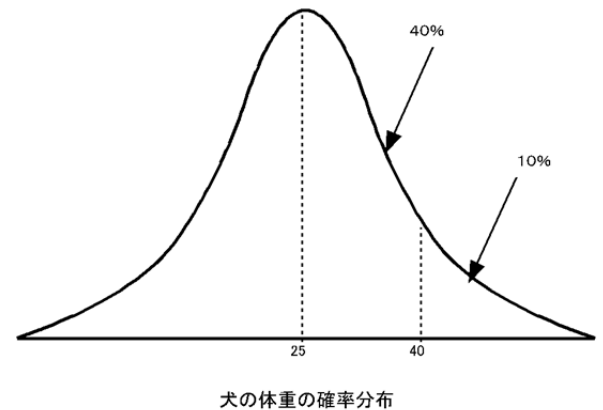
★ 確率分布は母集団について考えること．標本はあくまでも母集団の確率分布を予想するために調査する．

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

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．



この血統の犬を無作為に選んだとき，体重がちょうど 25kg である確率は（ ）である．体重がちょうど 40kg である確率は（ ）である．体重が 25～40kg の間である確率は（ ）である．体重が 40kg 以上である確率は（ ）である．

C. 標本から母集団の確率分布を予想する

1. 標本の数が多いときには標本のヒストグラムから母集団の確率分布を想像できる

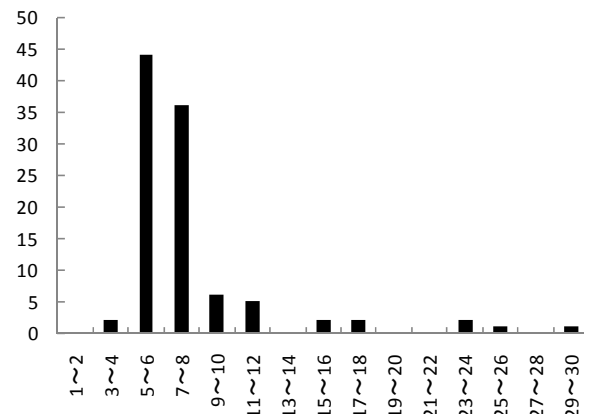
以下のヒストグラム（標本）をみて，もとの母集団を想像しながら，ある値あるいはある範囲の値を取る確率を読みとってみよう．

- ① もとの母集団の確率分布をヒストグラムの上に書き込む．
- ② 授業で指定した値あるいはある範囲の値をとる確率を読み取る．

A. マッチが燃え尽きるまでにかかる時間

マッチが燃え尽きるまでにかかる時間が 10 秒以上である確率

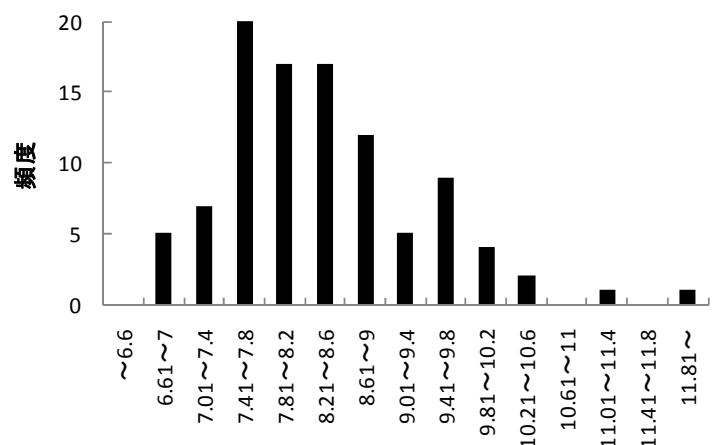
5～6 秒である確率



B. 中学生男子の 50m 走のタイム

ある中学生男子を無作為に選んだときにタイムが 8 秒以内である確率

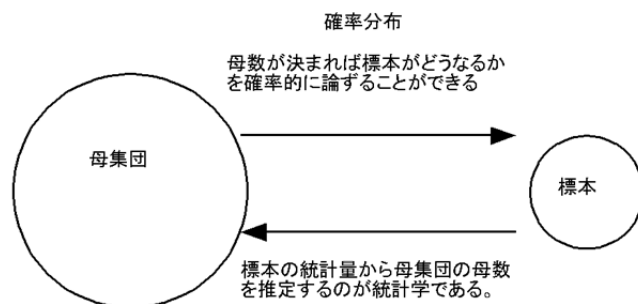
7～8 秒である確率



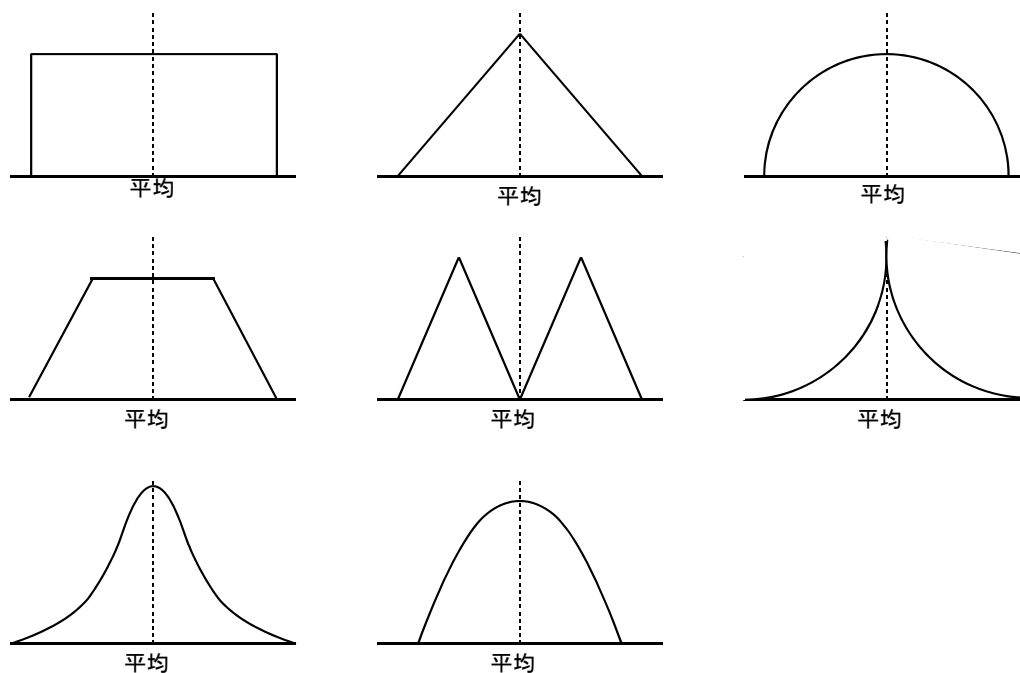
2. 確率分布を使った母集団と標本の関係

以上のように標本を多数集めれば、母集団の確率分布がわかり、その結果、標本がどのような確率で得られるかがわかる。

しかし、実際に実験や調査をするときに 100 以上の標本を集めることは労力、コストなどがかかり、簡単ではない。もし確率分布によくあるパターンがあるなら、標本が少数しかなくても、その確率分布を適用すればよいのではないだろうか？では確率分布によく出現するパターンはあるのだろうか？



質問 以下の確率分布のうち、魚の体重、トマトの収量、ウシの餌を食べる量など生物や農業に関わるデータについて、いちばんありそうな分布はどれかを考えてみよう。



D. 宿題

宿題は <https://moodle.cerd.shimane-u.ac.jp/moodle/> をご覧ください。