



農場で使える統計講座

～ 第5回 点推定その1 ～

これまで、平均や中央値など基本的な統計指標についてかんたんな例を用いて説明してきました。今回は一歩踏み込んで、「推定」ということについて考えていきたいと思います。

推定とは

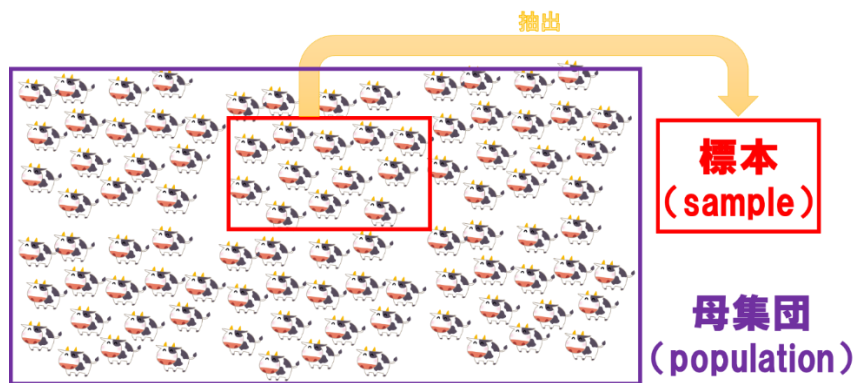
いきなりですが、こんなことを考えたことはありませんか？

- 農場における親牛の平均体重ってどれくらいなんだろう（全頭調べるのは大変だけど）
- 農場における搾乳牛のDMI（乾物摂取量）ってだいたいどれくらい？（全頭調べるのは大変だけど）
- 農場のTPI（Total Performance Index）ってどれくらい？（全頭調べるのは大変だけど）

そんなことを一度は考えたことのあるアナタ！もうそれは「**推定**」をしようとしています。これまで説明してきた例では、牛群に10頭しかいないような農場を想定して平均やら標準偏差やらを計算してきました。しかし、実際には何十頭、何百頭という状況で色々な項目をすべてのウシに対して調べていられません（乳検などをしていなければ）。そこで、その中から何頭か（もしくは何十頭とか）を抽出して、注目している項目に関して調べることになります。それを推定と言います。

推定のイメージ 1

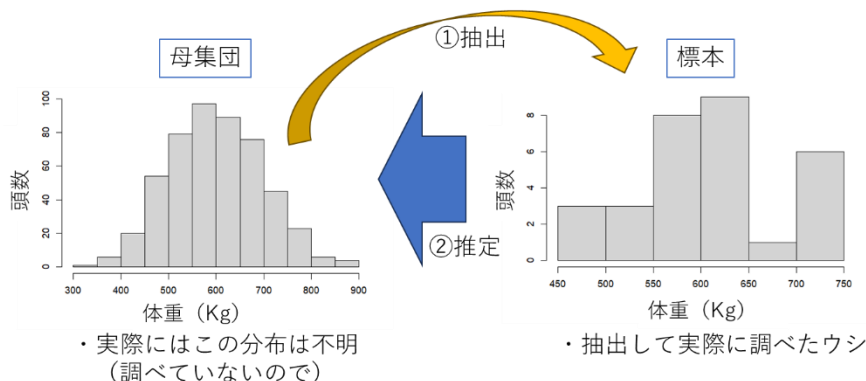
ずばり、推定のイメージを図示すると下のようになります。



紫色の線で囲まれたウシたちが牛群全体だとします。それを**母集団**（population）と言います。そこから何頭かピックアップします（**抽出**）。その抽出したウシたちを**標本**（sample）と呼びます。母集団を全頭調べるのは難しいけれど、抽出した標本だけに着目して調べるのは可能そうですね。つまり、標本のウシを調べて、牛群全体がどんな特徴をもっているのか明らかにすることが「推定」です。

推定のイメージ 2

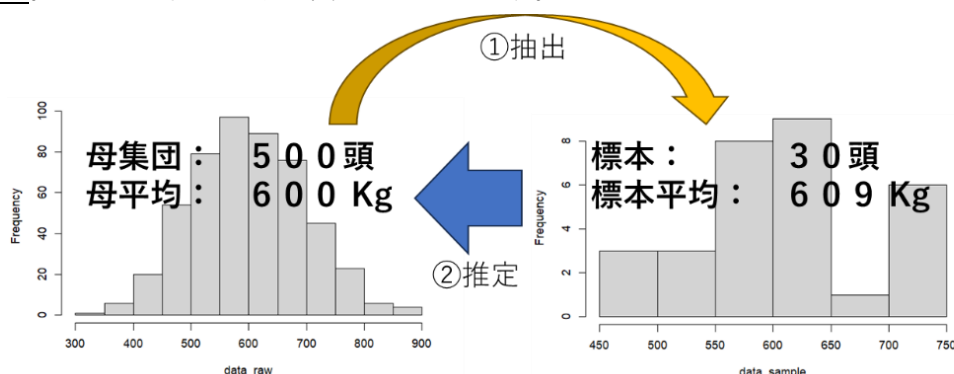
もう少し、推定について具体的に掘り下げてみましょう。



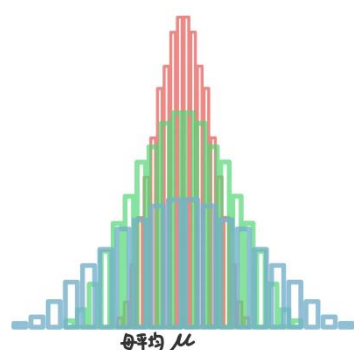
いまテキトーに作った 500 頭牛群の体重データがあります（前のページ下左図）。しかし、実際にはこの数値・分布は不明です。成牛全頭の体重を調べることはかなり難しいからです。そこで、そこから 30 頭の標本を抽出します（①）。30 頭くらいなら頑張れば実際に測定することが可能ですよね。そうしてわかった体重の数値・分布が右図になります。そして、今度は**標本の数値を用いて“不明な母集団”の体重を調べてみよう**ということを行います。これが推定です（②）。

点推定

この 30 頭のサンプルを使って、500 頭の牛群全体の平均体重・分布を知ることができる！これは確かにそうなのですが、実は問題もあります。抽出したウシたちが必ずしも牛群全体を正しく反映しているとは限りません。さっきと同じような図を下に出します。



今回の例では、テキトーに 500 頭分の体重データをつくったといいましたが、平均をピッタリ 600Kg にしてね、というルールを設定しました（なので 600 にピークがある）。しかし、実際に抽出した 30 頭分のデータをみると、平均は 609Kg となっており、両者に違いが出ています。これは、直感的にもそりゃそうだろうと思う方もいるかもしれません。ベストなのは全頭調べることなのですが、それができないから抽出して標本（サンプル）から牛群全体の体重を推定するという、いわば妥協案なわけです。というわけで、ある程度の差は仕方ないということになります。その違いを縮めるために、牛群全体からできるだけランダムに抽出されるようにしたり、標本の数を増やしたりといったことが必要な場合があります。



サンプルサイズが
大きければ（小→中→大）
標本平均は母平均に
寄っていく（精度が上がり）ます



“クマの数学日記” から図引用

(<https://math-journal.bear-fruits.online/ln-n-clt/>)

大数の法則というものがあって、少ない標本数だと母集団（本来知りたい対象の牛群全体）との差が大きい場合がありますが、**標本数を増やしていくことで母集団をより正確に捉えられるよ！**というものです（上にあるウェブから引用した図がわかりやすい）。なので、まずは「できる範囲でサンプルをして、その後、標本数を増やしていく」というのも一つの手です。ちなみに、調べたいけど、調べることのできない集団（今回の例では牛群全体）のことは母集団と説明しましたが、その平均値を母平均といいます（この値は実際にはわかりません）。標本から求めた平均値を標本平均といいます。（次回に続く）

かやの