農場で使える統計講座
～ 第7回 区間推定その1 ～

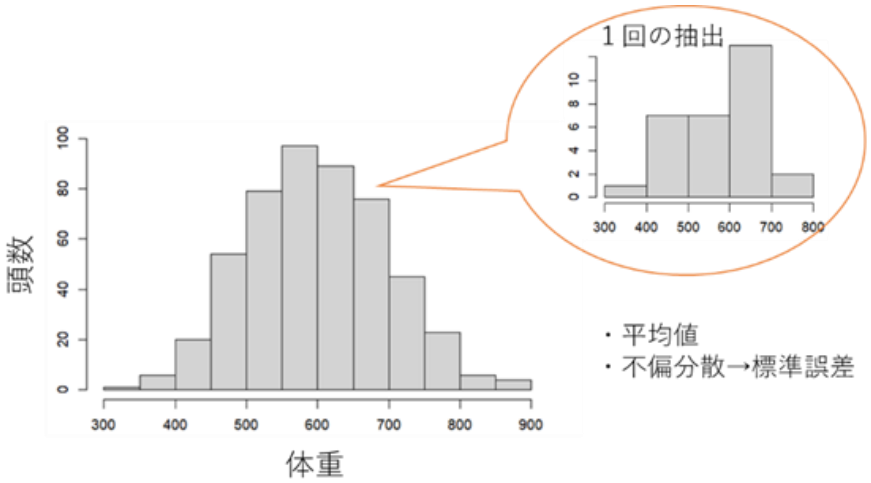
かやの

前回は平均値のバラツキをとらえる標準誤差などについて説明しました。たくさんの頭数がいる中で、その一部をサンプルして乳量や病原菌の有無などをチェックし「推定」するわけですが、そこから出した平均値なども当然のことながらバラツクわけです。そのバラツキを表したのが標準誤差でした。今回は１点という平均値（点推定）から一歩踏み込み、A から B までというような区間推定について説明します。

区間推定（信頼区間）に必要な材料

区間推定といわれてピンと来ない人は、信頼区間を思い浮かべてください。９５%信頼区間などといわれるアレです。信頼区間の推定は、調べたい対象がなんらかの分布に従うと仮定して計算するものから、ベイズ推定と呼ばれるものまで様々な手法があります。このシリーズでは、よりシンプルな方法として（一例として）t 分布および正規分布を利用した区間推定をご紹介します。今回は t 分布を利用して、区間推定（信頼区間）を推定しようと思います！

前回、５００頭の牛群から３０頭をサンプルして、群の平均体重を推定するということをやりました。



さらに、その平均値のバラツキを調べるために標準誤差というものを求めました。それが重要になってくるのですが、t 分布を利用した平均値の区間推定に必要な材料を以下に書きます。

標本平均と標準誤差

この２つが分かれば、ＯＫです。標準誤差は不偏分散と標本数（データ数）で求められるので、

標本平均と不偏分散とデータ数

となり、必要な数値はこの３つと言い換えることもできます（標準誤差などは第６回で説明しました！）。

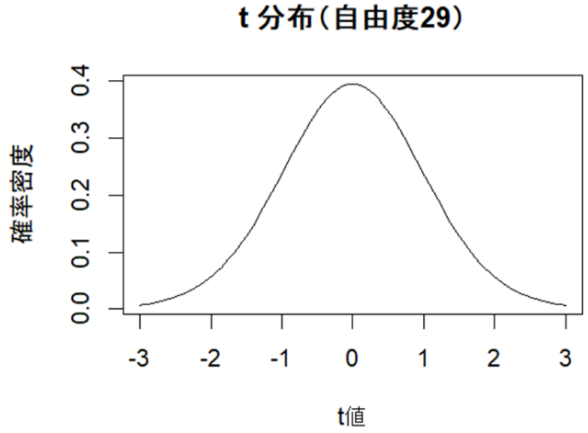
t 統計量と t 分布

この辺りは少し難しい話になるので、興味を持てない人は飛ばしてください。いきなりですが、t 統計量というものを考えます。

$$t = \frac{標本平均 - 母平均}{標準誤差} = \frac{標本平均 - 母平均}{\sqrt{\frac{不偏分散}{標本数}}}$$

なにかよくわからないけれど、とりあえず手持ちの数値で t 統計量というものが求められそうだということだけでも感じてください。次に分布について考えてみますが、統計量 t というものは、自由度 n-1 の

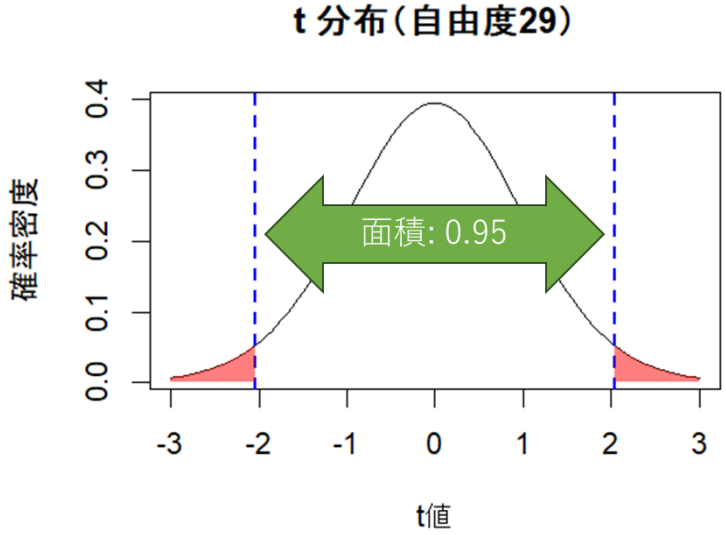
ｔ分布という確率分布に従うということが知られています。というかそういうルールです。数学のポイントは、理由はよくわからないけど公式やルールを受け入れることです！自由度 n-1 とは、ここでは簡単に標本数-1 が自由度であると理解してください。いま、サンプル数（標本数）は３０ですので、自由度は３０－１＝２９となります。それを図にしたものが下に示すものです。



これが自由度２９のｔ分布です。確率分布とは曲線下の面積を全部足したら１になるということです。

95％信頼区間（t分布を利用した場合）

では次に、９５％信頼区間とはどういうことかというと、下の図が直感的にわかりやすいです。



ｔ値（ｔ統計量）が分布面積の９５％をカバーするような範囲にあればいいということになります。上の図でいえば、サンプル数が３０のｔ分布の場合（自由度２９）、だいたいですが、―２から２の間にｔがあればＯＫということです。厳密には、自由度２９のｔ分布で９５％信頼区間（有意水準５％ともいう）を示すようなｔ統計量はｔ分布表というもので確認すると（Google 先生に聞いたらすぐに出てきます）、2.045 ということがわかります。ということは、ｔ分布は０を挟んで左右対称なので、９５％信頼区間を表すｔ統計量の値は、以下のようになります。

$$-2.045 \leq t \leq 2.045$$

小休止

中途半端ですが今月はここで終わります。区間推定は、A から B までの範囲の中に求める数字があることを教えてくれます。点推定より実用性は高いかもしれません。計算に必要な推定値はこれまで触れてきたものですが、ｔ分布や信頼区間の考え方は少し時間をかけて理解する必要があるかもしれません。