



農場で使える統計講座

～ 第2回 平均とバラツキ ～

前回は、統計とはそもそもなんぞやという話にはじまり、乳検の例などについて書きました。農場は様々な数値化（グラフ化）できるデータがあり、それらをまとめ役に立つ情報を得ることこそ、まさに「統計」といえます。今回からは少し具体的な話に移り、数字を使いながら考えていきたいと思います。

平均の求め方

以下のような2つの農場を考えます。

	A 農場	B 農場
平均日乳量 (Kg/頭)	29.4	29.2

一頭当たりの平均日乳量はA農場で29.4Kg、B農場も29.2Kgとほぼ同じ値となっています。ここで考えたいのは、A農場とB農場は同じようなポテンシャルをもった農場なのかということです。

そもそも、平均とはどのような計算で求められるのでしょうか。

$$\text{平均} = \frac{\text{データの合計}}{\text{データの数}}$$

平均は上の式で求めることができます。エクセルだと *AVERAGE* という関数で計算できます。次にA農場とB農場の日乳量 (Kg) の内訳をみてみましょう。説明を簡単にするため、牛群は10頭しかいません！

	A農場の日乳量 (Kg)	B農場の日乳量 (Kg)
1 頭目	5	24
2 頭目	10	31
3 頭目	13	33
4 頭目	41	20
5 頭目	25	32
6 頭目	41	27
7 頭目	43	33
8 頭目	28	30
9 頭目	43	34
10 頭目	45	28

平均を計算するためには、データの合計が必要です。それぞれの農場ですべてを足すと、294 Kg (A農場)、292 Kg (B農場) となります。これを、データ数、つまり10でそれぞれ割ると、

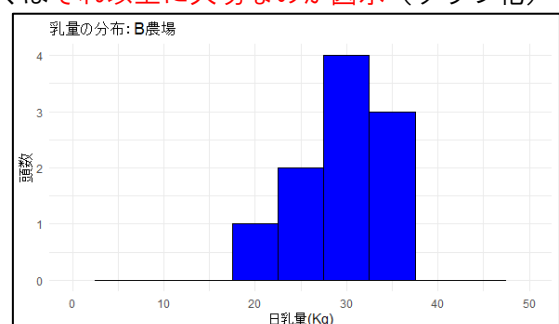
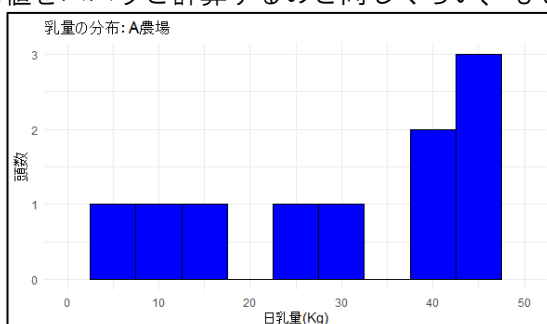
$$\text{A農場の平均日乳量} = 294 / 10 = 29.4 \text{ Kg}$$

$$\text{B農場の平均日乳量} = 292 / 10 = 29.2 \text{ Kg}$$

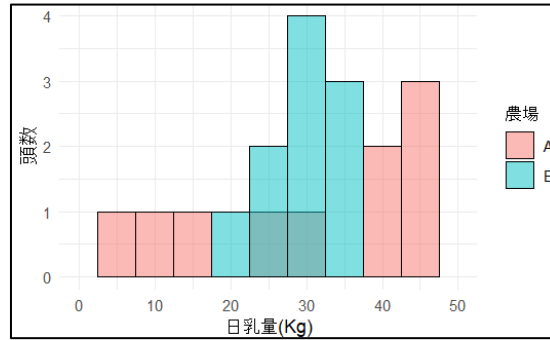
となり、最初の表の通りに平均を求めることができました。

グラフ化

実は、平均値をパッと計算するのと同じくらい、もしくはそれ以上に大切なのが図示（グラフ化）することです。



前のページではグラフの縦軸がそろっていないので、少し比較がしづらいです。一緒に表示してみます。



こうして表示すると、よりわかりやすくなったと思います！A農場は乳量が低いウシたちもいれば、40Kgを超えるウシも数頭いることがわかります。一方、B農場は乳量が比較的同じようなウシが多く、30Kg付近に多くのウシが存在していることがわかります。つまり、**A農場のほうがB農場に比べてバラツキが大きい**ことがわかります。それはなぜ？という理由はたくさんあると思いますが、ここでは深掘りしません。大切なのは、**平均値だけを見ているとこうしたバラツキはわからない**ということです。

バラツキを数値化する

そんな大切な情報であるバラツキも簡単に計算することができます。まずは「**分散**」とよばれるものです。**データそれぞれの値が、平均からどれだけ離れているか**に着目をします。エクセルでは *VAR.P* です。

$$\text{分散} = \frac{1}{\text{データの数}} \sum_{i=1}^{\text{データの数}} (i\text{番目のデータ} - \text{平均})^2$$

Σ（シグマ）とか出てきて少し難しそうに見えますが、（各データの値－平均）の2乗の合計を、データ数で割ったものです。例として、A農場だけ先ほどの各乳量データと平均乳量を使って計算してみます。

$$A\text{農場の分散} = \frac{1}{10} \{ (5 - 29.4)^2 + (10 - 29.4)^2 + (13 - 29.4)^2 + (25 - 29.4)^2 + (28 - 29.4)^2$$

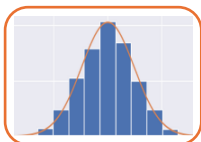
$$+ (41 - 29.4)^2 + (41 - 29.4)^2 + (43 - 29.4)^2 + (43 - 29.4)^2 + (45 - 29.4)^2 \} = 214.44$$

同様の計算をB農場でもすると、18.2 という値が得られました。離れている距離（平均より〇〇だけ大きい、もしくは△△小さい）なので、負の値にならないようにするためのテクニックで2乗をしています。ですが、Kg²って意味がわからないですね。そこで、分散のルートをとることにしました。これを「**標準偏差**」といい、そうすると単位もKgに戻りました。A農場とB農場で計算してみましょう。

$$A\text{農場の標準偏差} = \sqrt{214.44} = 14.64 \text{ Kg}$$

$$B\text{農場の標準偏差} = \sqrt{18.16} = 4.26 \text{ Kg}$$

平均値と一緒にまとめます。ちなみに標準偏差はエクセルでは *STDEV.P* という関数で計算可能です。



	平均値 (Kg)	標準偏差 (Kg)
A農場	29.4	14.6
B農場	29.2	4.3

このようにしてみると、A農場とB農場の平均値は同じですが、標準偏差が大きく違い、A農場のほうが平均値から遠い乳量のウシが多い、つまりバラツキが多い牛群であることがわかります。

まとめ

平均には落とし穴がありバラツキもみることが重要！図示は超有用で、多くの情報を与えてくれます！

かやの

※ ちなみに、平均値が頂点になるベル（釣鐘）のような形をした分布（正規分布）になっている場合、平均値±標準偏差の間にデータの約68%がおさまり、平均値±2×標準偏差で約95%のデータがその中に入ることが知られています。