



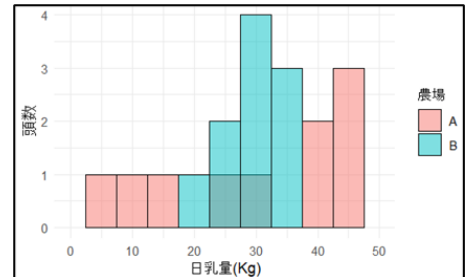
## 農場で使える統計講座 ～ 第3回 中央値と箱ひげ図 ～

先月は、平均やバラツキ（分散・標準偏差）の求め方とその重要性について書きました。平均値だけで考えるのではなく、まずは図示して目で**分布を確認**することがとても大切です。今回は、平均値と同じような代表的な指標である中央値や雑誌などで見ることもある箱ひげ図などについて説明します。

### 前回のおさらい(平均値や標準偏差)

引き続き、A農場とB農場の日乳量データを例にして考えていきます。まず、前回チェックした乳量の分布と平均値の表を思い出してください。

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



平均値は同じような値ですが、**バラツキが大きく違う**ということがわかりました。

### 中央値など

平均値だけではなく、データをまとめるために用いられる指標がいくつかあります。それらは日ごろ目にしている乳検や農協の資料などにも使われています。まずは、「**中央値**」と呼ばれるものです。その名の通り、データを小さい順から並べて、ちょうど真ん中（中央）にある数値のことをいいます。A農場とB農場の乳量データを低い順から並べてみます。

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

今回の例のようにデータの数が偶数個の場合（A農場・B農場ともに10頭）、真ん中にくる数値が1つではないため、真ん中にある2つの数値の平均をとります。**赤枠に注目！**

$$A農場の中央値 = (28+41) \div 2 = 34.5 \text{ Kg}$$

$$B農場の中央値 = (30+31) \div 2 = 30.5 \text{ Kg}$$

さらにこのまま「**最頻値**」という指標についても説明します。最頻値もその名が示す通り、最も出現頻度の多い値です。**青枠に注目！**

A農場の最頻値は、41 Kg と 43 Kg

B農場の最頻値は 33 Kg

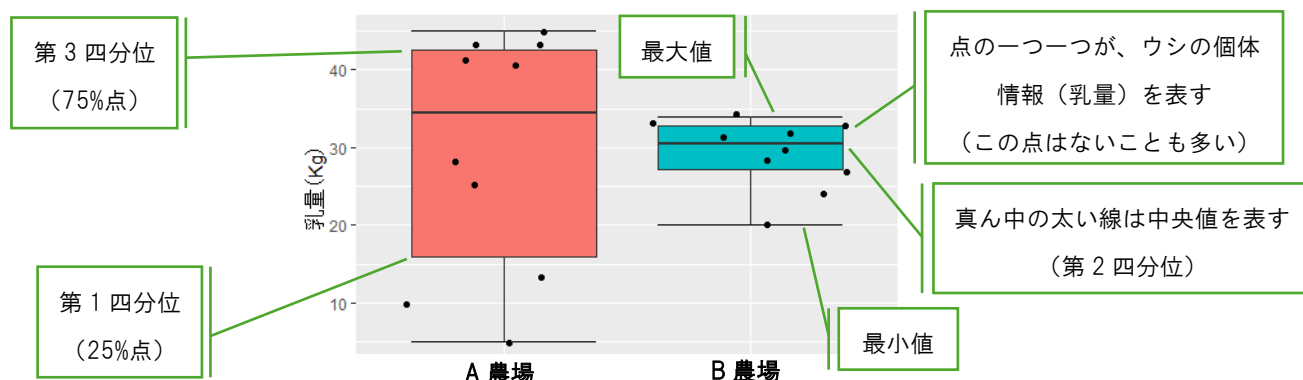
続いて、日乳量の平均値と合わせていま求めた中央値と最頻値を一緒にみてみましょう。

|      | 平均値 (Kg) | 中央値 (Kg) | 最頻値 (Kg) |
|------|----------|----------|----------|
| A 農場 | 29.4     | 34.5     | 41 と 43  |
| B 農場 | 29.2     | 30.5     | 33       |

平均値だけに着目すると、A 農場・B 農場同じような日乳量となっていますが、他の指標では差が開いているのがわかると思います。こうした違いの原因は、最初のグラフの通り乳量の**分布が異なる**からです。

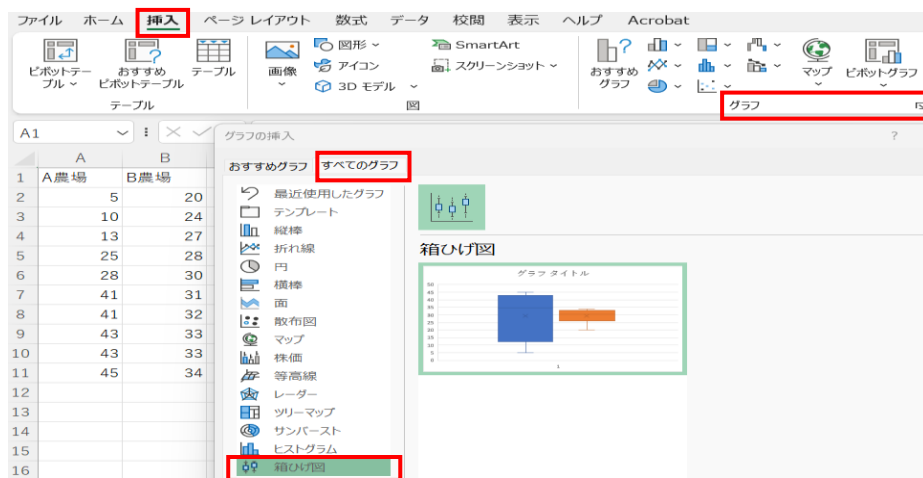
## 箱ひげ図

「**箱ひげ図**」という名前を聞いたことはありませんか？箱ひげ図はデータの分布がどのようなになっているのか**大きさ順に並べてわかりやすくまとめられた図**のことです。前頁で出したような棒グラフよりも多くの情報が直感的に理解できるようなものになっています。下の図が箱ひげ図と呼ばれるものです。



箱の真ん中にある太線が中央値を示しています。小さい順から並べて下の四角には最初から数えて 25～50%の数が入り、上の四角には 50～75%の数が入ります。データを等分に 4 つに区切っているこの範囲を四分位範囲と呼びます。**2 つの四角を合わせた箱にはデータの 50%、つまり半分が収まっていること**になります。この図からも、中央値は A 農場のほうが B よりも高くなっていますが、データに幅があり、B 農場は中央値こそ A よりも低いですが、多くのウシが似通った乳量であり、バラツキが小さいことがわかります。また、A 農場では乳量が最も低いウシは 5Kg 程度ですが、B は 20Kg 近くもあります。単にデータだけを並べたよりも、視覚的かつ直感的に牛“群”の状態を把握しやすいことがわかりますね！

ちなみにエクセルでは、挿入>グラフ>すべてのグラフ>箱ひげ図 で描くことができます。



## まとめ

平均以外の「集団・群」を表現する指標として中央値や最頻値、そして便利な図である箱ひげ図を紹介しました。データがあればエクセルなどで簡単にチェックできます。興味のある方はお試しあれ！

かやの