

箱ひげ図：中央値と四分位値の可視化

(Tukeyの定義による) 四分位値：
中央値を境として分けた2群において、
それぞれの中央値に対応する。

「原種チューリップ」札の左の鉢

12.0cm 8.9cm 11.1cm 12.3cm
10.0cm 9.8cm 8.9cm 9.0cm

大きさ順に並べ替える

8.9 8.9 9.0 9.8 | 10.0 11.1 12.0 12.3 (cm)

中央値 = 9.9

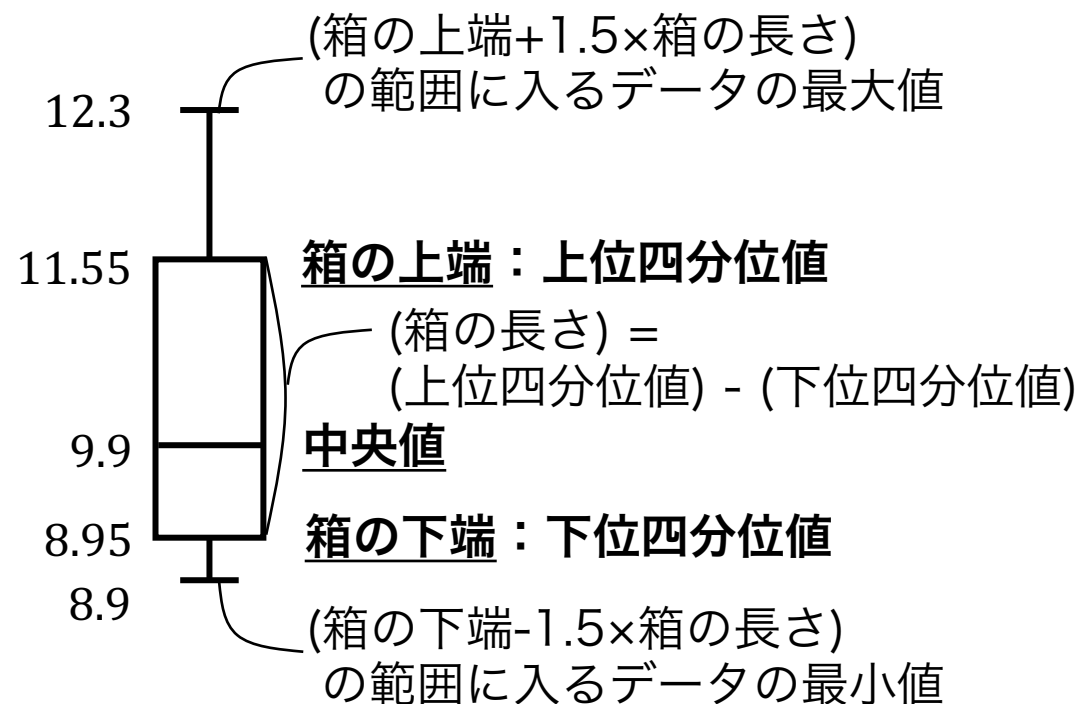
下位四分位値：

$$\frac{8.9 + 9.0}{2} = 8.95$$

上位四分位値：

$$\frac{11.1 + 12.0}{2} = 11.55$$

箱ひげ図

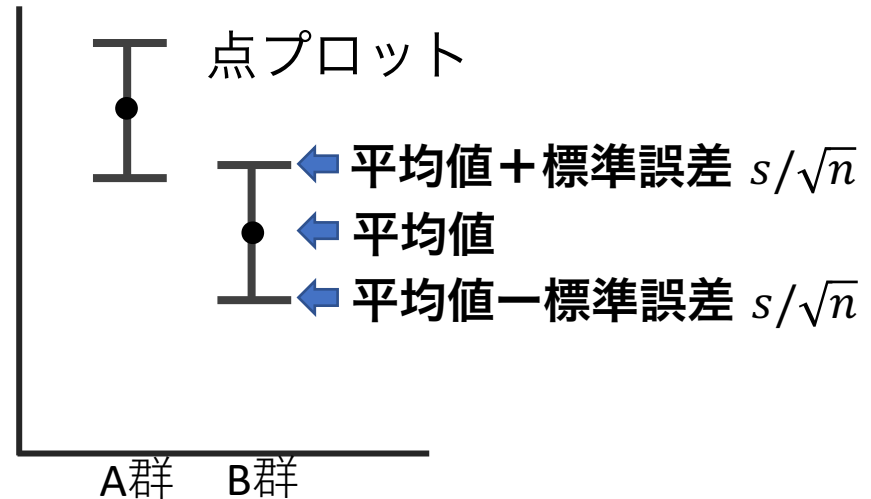
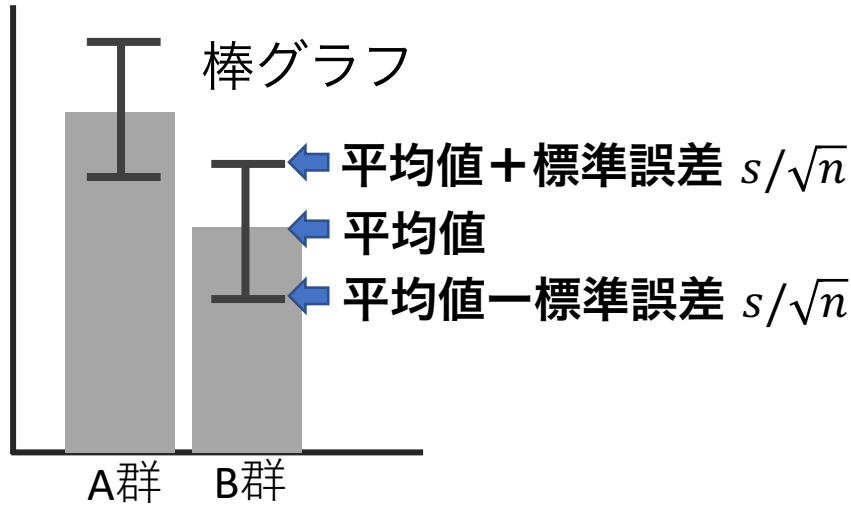


ひげの上端・下端の範囲に
含まれない値は**外れ値**としてプロットする。
(左のデータでは該当なし)

※ ExcelやPythonライブラリの関数では
異なる定義に基づいて計算されるため、値が異なることがある

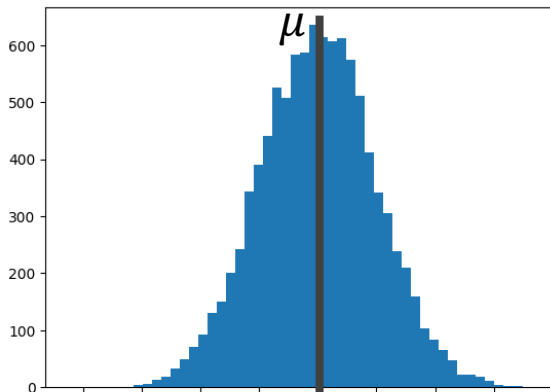
標準誤差を使ったエラーバー

平均値の差の検定をしたとき、推定した平均値の確からしさを示すため、標準誤差 (Standard Error; SEと書かれる) を使ったエラーバーを示す。



標準誤差 s/\sqrt{n} (SE)とも呼ぶ。

正規母集団から標本抽出した場合、または n が大きい（中心極限定理）場合は
標本平均 \bar{x} は正規分布に従う。



\bar{x} の平均： 母平均 μ の値

\bar{x} の標準偏差： σ/\sqrt{n}

母集団の標準偏差 σ の
代わりに不偏標準偏差 s を使い
標準誤差： s/\sqrt{n} とした。

エラーバーが短いほど…
→標準誤差が小さい
→標本平均の
ばらつきが小さい
→標本平均の精度が高い

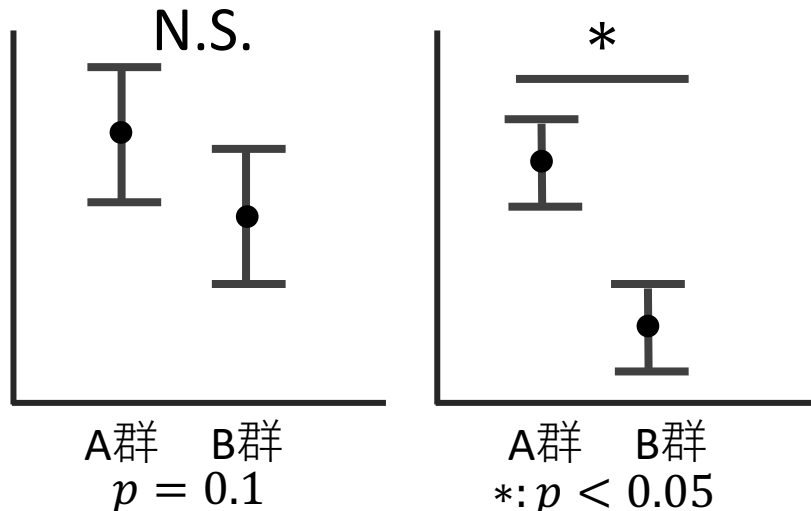
エラーバーによる平均値の差の目安

エラーバーだけでは統計的有意差の議論はできないが、
 p 値やアスタリスク* ($p < 0.05$ などを表す)とあわせてグラフに描かれる。

エラーバーが重なる時：

平均値に有意差はない、
という目安にされることがある。

(標準誤差のエラーバーは
平均値の推定精度を示す
ためのものなので、
必ず p 値の記載を確認する。)



(N.S.: 有意差なしを表す)

エラーバーには他にも種類があり
目的に応じて使い分ける：

1. **標準誤差(SE)**を表す
(平均値の推定精度を表すとき)
2. **標準偏差(SD)**を表す
(データのばらつきを示すとき)
3. **95%信頼区間**を表す
4. データの分布を表す

エラーバーが示すものを必ず記載する。
論文を読むときは必ず確認する。