

母平均 μ の区間推定

0、はじめに

- 今回の目標は図1のような「**95%信頼区間のエラーバーの表示**」です。
- 図1は以下の情報を示しています。
 - 標本サイズ $n=16$ (人) → 16 人の身長を測定した
 - 標本平均 $M=170\text{cm}$ → 16 人の平均身長が 170cm
 - 95%信頼区間 167~173cm
→ 母平均 μ は 95%の確率でこの区間にありそうだ

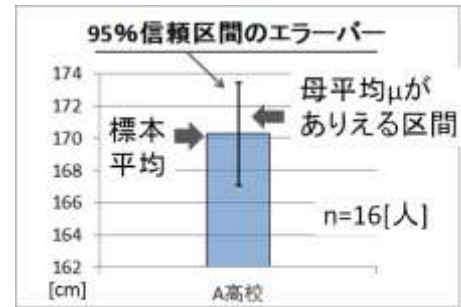


図1 A高校男子の平均身長

- 信頼区間のエラーバーには大きく 3つの利点があります。
 - 1、**母集団と標本** (サンプル) の違いを意識できる
 - 2、**標本サイズを大きくすると母平均に近づく**ことが直感的に理解できる
 - 3、**大小関係の判断ミス**を防ぐことができる

詳細は以下↓

1、母集団と標本 (サンプル) の違い

- 図2のように、抽出された一部を**標本** (サンプル) といい、その平均値を**標本平均 M** (※mean の頭文字) といいます。
- また、標本の発生源になる集団を**母集団**といい、その平均値を**母平均 μ** といいます。(※ μ は M に相当するギリシア文字)
- 本当に知りたい平均値は母平均 μ のはずですが、しかし、手に入る平均値は標本平均 M しかありません。
- そこで手元の標本平均 M を使って、**母平均 μ が存在しそうな区間を推定**します。これを区間推定といいます。
- 95%の確率で母平均 μ が存在しそうな区間を **95%信頼区間** といい、図1のようにエラーバーを用いて表示します。

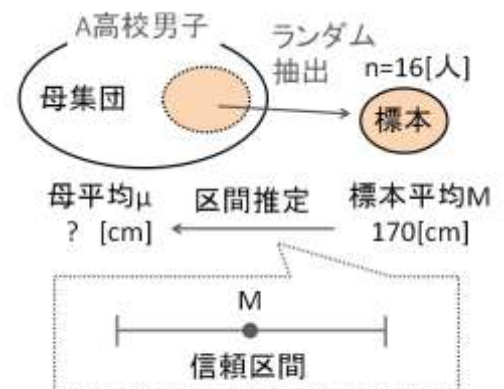


図2 母平均の区間推定のイメージ

2、標本サイズ n を大きくすると母平均 μ に近づく

- 図3からは次の2点が読み取れます。
 - ① 標本平均 M $n=16$ [人] → $M=170$ [cm]
 $n=100$ [人] → $M=171$ [cm]
 - ② 信頼区間のエラーバーがせまくなる
- これが意味することは以下です。

「**標本サイズ n を大きくすると、標本平均 M は母平均 μ に近づいていく**」 (←大数の法則という)
- 標本サイズを 16 人から 100 人にすることで、母平均 μ がありそうな区間を 171cm 付近まで絞り込めたわけです。
- 母平均 μ の区間をどのくらい絞り込むかは調査者による判断です。信頼区間のエラーバーはその判断を助ける重要なツールになります。(※必要な標本サイズを算出することも可能)

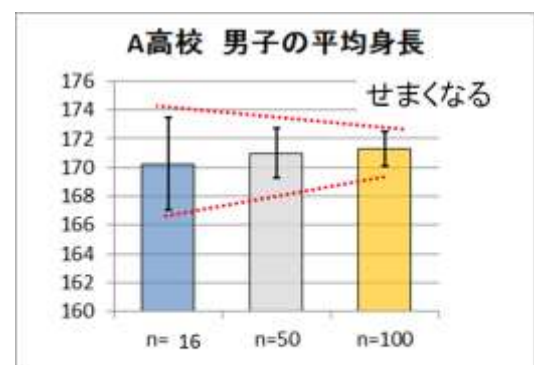


図3 標本サイズ n と信頼区間の関係
 n が大きいと、信頼区間がせまくなる

3、大小関係の判断ミスを防ぐ

- ・図4のように、3校の男子の平均身長があります。
A高校の男子は、B高校とC高校の男子よりも背が高いのでしょうか？低いのでしょうか？
- ・もし信頼区間のエラーバーがなければ
次のような判断をするかもしれません。

×

- ・A高校の男子 > B高校の男子
- ・A高校の男子 < C高校の男子

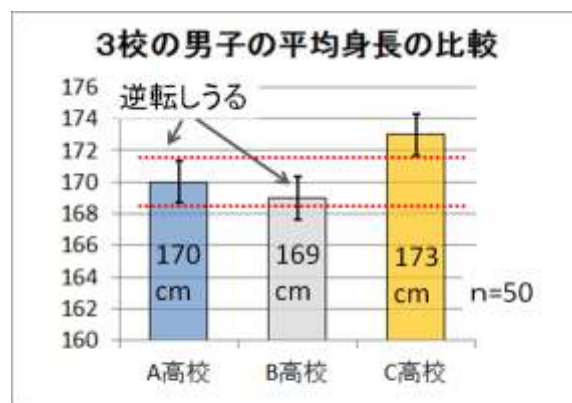


図4 大小関係の判断と信頼区間

A校とB校は母平均が逆転する可能性あり

- ・しかし、信頼区間のエラーバーがあると以下のように判断ができます。

○

- ・A高校の男子 > B高校の男子 とは**判断できない**
- ・A高校の男子 < C高校の男子 といえる（高確率で）

- ・A高校とB高校は**信頼区間**の大部分が重なっています。これは「A高校の母平均 μ_A とB高校の母平均 μ_B が**逆転する可能性がある**」ということを意味します。よって大小関係の判断ができません。
つまり、A高校とB高校の1cmの差は“**意味のある差**”ではなかったということです。
判断するには標本サイズを大きくして、信頼区間をせまくすることが考えられます。

- ・対してA高校とC高校の信頼区間は重なっていません。これは「各母平均が逆転する可能性が低い」ことを意味します。よってA高校はC高校よりも平均身長が小さいと判断できます。
つまり、A高校とC高校の3cmの差は“**意味のある差**”だといえます。これを**有意差**（ゆういさ）といいます。
- ・このように信頼区間を表示すると、大小関係を**より正しく判断できる**ようになります。
（もっと正しく判断するには“検定”という統計的手法を使う。母平均の比較ならt検定を用いる）

4、95%信頼区間の求め方（エクセルの方法）

- ・「エクセルの基本統計量から求める方法」を紹介します。

信頼区間の求め方

- ① データ → データ分析 → 基本統計量をクリック
- ② 入力範囲と出力先を指定
→ 統計情報と信頼区間にチェック
- ③ OK をクリック → 図5のような表が出る
- ④ 図5のように平均と信頼区間から求める

エラーバーの表示

- ① グラフをつくる
- ② グラフツールのレイアウト → 誤差範囲 → その他の誤差範囲オプション
- ③ 縦軸誤差範囲 → ユーザー設定 → 図5の信頼区間の値（3.194）を選択

以上です。

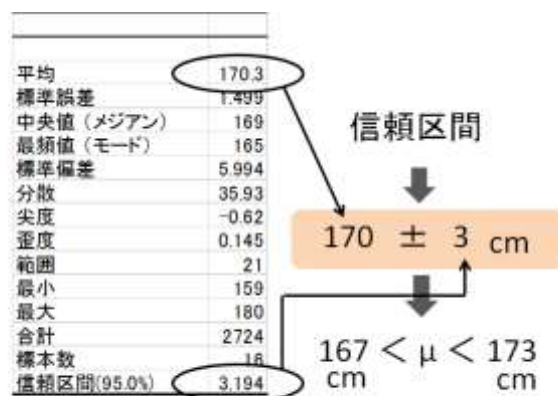


図5 基本統計量から信頼区間を求める

- ・他にも「**エクセルの関数を使う方法**」もあります。しかし数学的な知識が必要なので少し難しいです。
WEBサイトの「教材」に「**母平均の区間推定**」という Excel ファイルを作りました。参考にしてください。