母集団と標本
（教科書6, 8章）

北九州市立大学経済学部

齋藤 朗宏

今日の内容

- 母集団と標本
- 標本統計量

- 実習

標本抽出

- 調査対象となる集団のことを母集団（Universe, Population）と呼ぶ（ex. 日本人，大学生）．
- 母集団のすべての個体に対して行う調査を全数調査と呼ぶ.
- 母集団における統計量を母数と呼ぶ（母平均）．
- 母集団全体に対して調査を行うのは難しいので，母集団の中から一部の個体（標本）を選んで（抽出・サンプリング）調査（標本調査）・測定し，母数の推定値とする．サンプリングは，原則ランダムであるべき（無作為抽出）．
- 標本から計算された母数の推定値を標本平均などと呼ぶ．(pp.142-143)

母平均と標本平均

- 「日本人の身長」が知りたいとする．
  - 母集団は日本人全体となる．日本人全員の身長を測定し，その平均を求めると，その結果は母平均 μ_x となる．
  - 日本人の中からランダムに何人か選び，その身長の平均を求めると，得られる値は標本平均 $\bar{x}$ となる．

- 「九女生のTOEICの得点」が知りたいとする．
  - 母集団は九女生全体．九女生全員の得点を調べ，その平均を求めると，その結果は母平均 μ_y ．
  - 九女生の中からランダムに何人か選び，その得点の平均を求めると，得られる値は標本平均 $\bar{y}$ となる．

母平均・標本平均

- 母集団の総数をN，標本数をnとする．このとき，確率変数xの母平均μ_xは以下の通り．

$$\mu_x = \frac{1}{N}\sum_{i=1}^{N} x_i$$

- 標本平均$\bar{x}$は以下の通り．

$$\widehat{\mu}_x = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

母分散・標本の分散

- **母分散** σ_x^2 は以下の通り.

$$\sigma_x^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu_x)^2$$

- 標本の分散$\hat{\sigma}_x^2$は以下の通り.

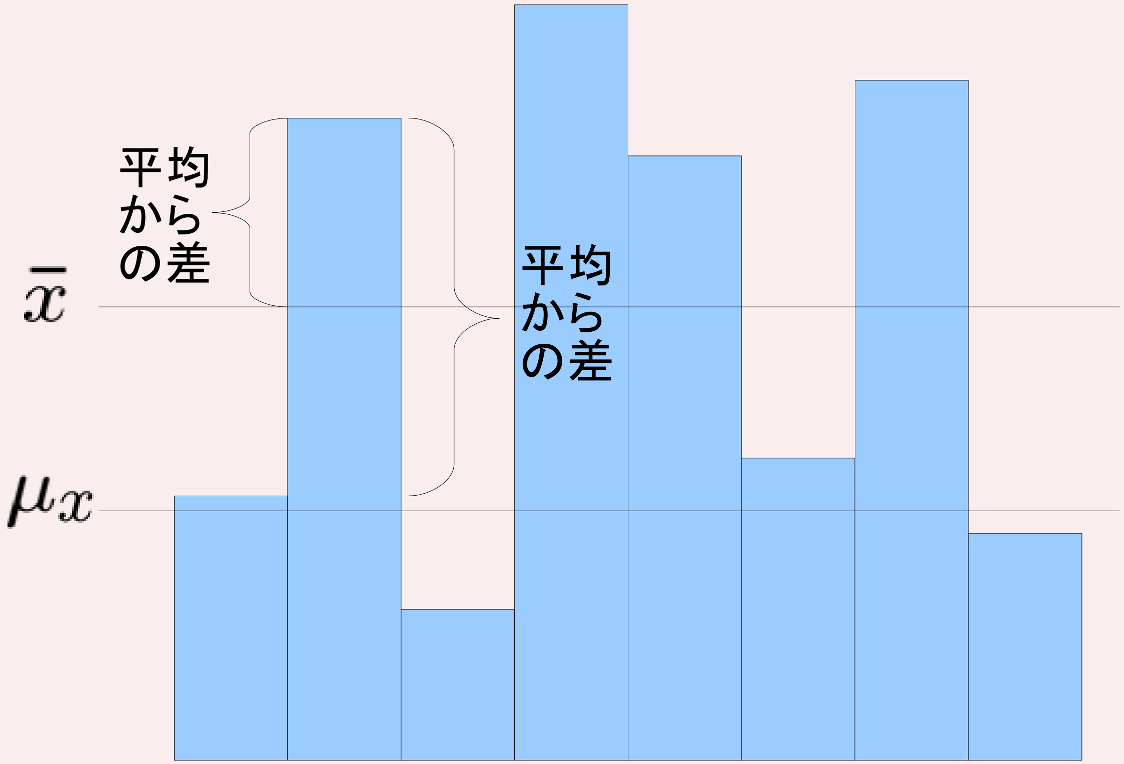
$$\hat{\sigma}_x^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \boxed{\mu_x})^2$$

手元には標本しかない．よって母平均は未知

$$\hat{\sigma}_x^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \boxed{\bar{x}})^2$$

これにも問題がある

何故不偏分散か？



- $\bar{x}$ と μ_x とは多くの場合一致しない, μ_x の真の値はどこだかわからないが，それが少しでも$\bar{x}$とずれていた場合，標本分散では過小推定.

不偏分散

- 母分散 σ_x^2 は以下の通り.

$$\sigma_x^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu_x)^2$$

- 不偏分散 s_x^2 は，母集団数 N が十分大きい時以下の通り.

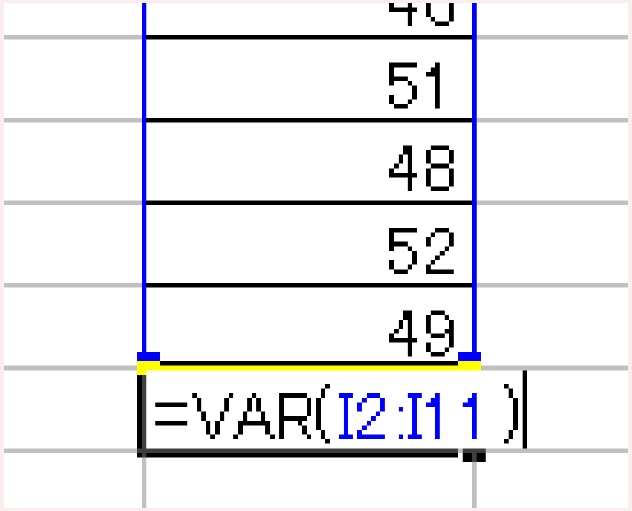
$$\hat{\sigma}_x^2 = s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

p.193

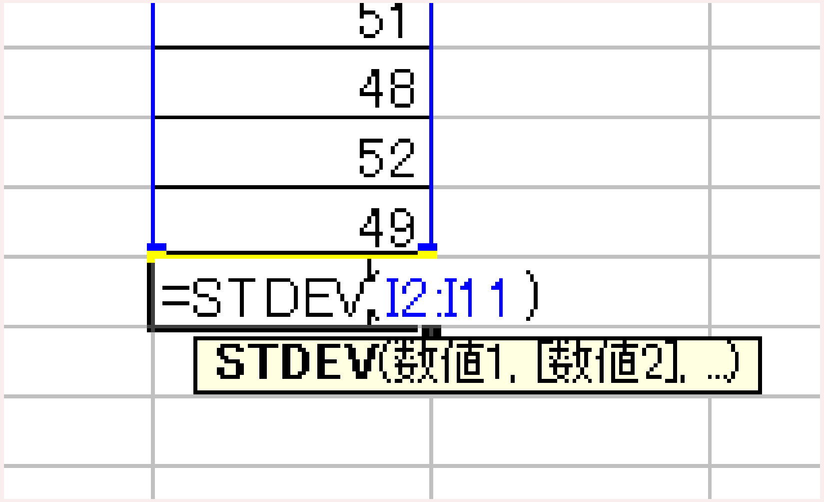
不偏分散

- 母分散と同じ式で**標本分散**を計算すると，分散を過小推定する（$\bar{x}$ は分散を最小にするμ_xの推定値である）ため補正された不偏分散を用いる．尚，不偏分散を標本分散と呼ぶこともあるので混同に注意すること．
- 従って，不偏分散の式を使用する，即ち$n-1$で割るのは，母平均μ_x が未知で，代わりに標本平均$\bar{x}$ を使う場合のみである．たとえ標本の分散であったとしても，何らかの事情で μ_x が既知であるならば，nで割る点に注意が必要である．

不偏分散



分散



標準偏差

実習

- 「お年玉」について，そのデータを母集団であると見なしたときと，母集団からの標本であると見なしたとき，それぞれについて，分散，標準偏差を求め比較せよ.