母集団と標本・標本分布

母集団と標本

- 母集団

対象となるすべての人，モノなどの集合

・無限に存在するモノ（コイントスの裏表）

・非常に多く存在するモノ（有権者）

- 母数（母平均，母比率...）

母集団を特徴づける値

・成人男性の身長の平均（母平均）

・コインの表の出現率，内閣支持率（母比率）

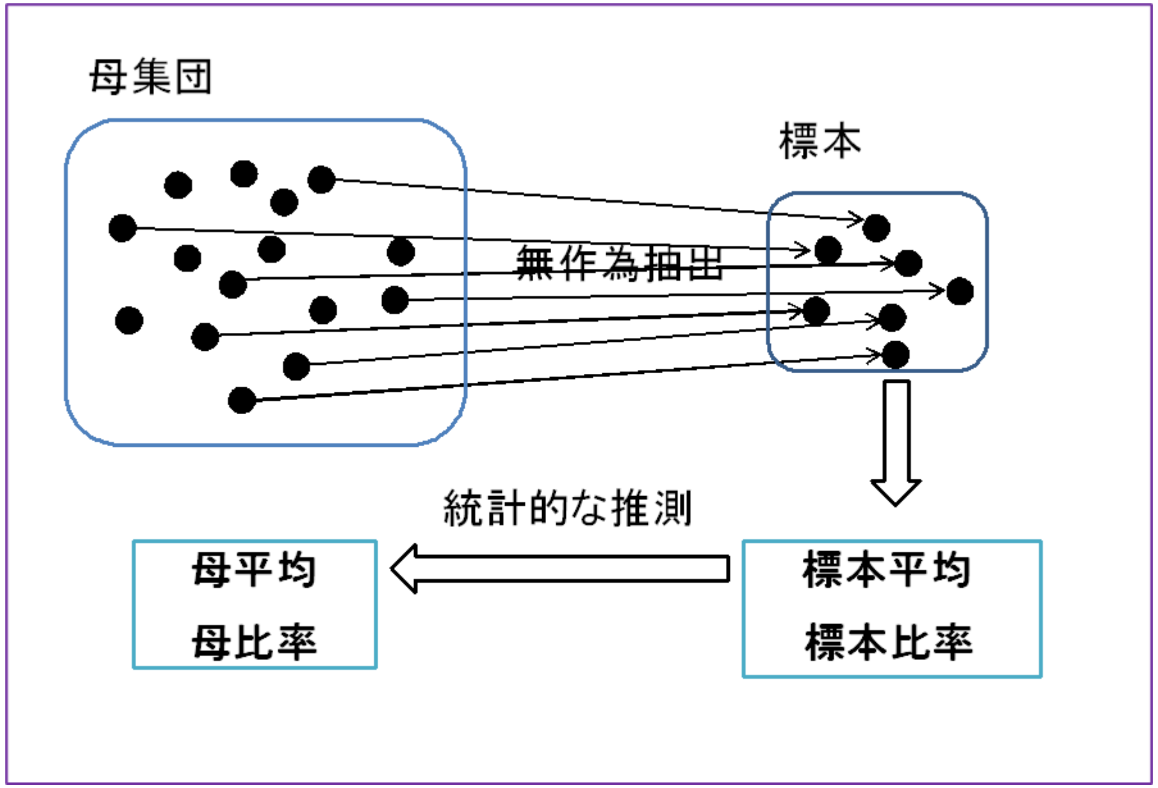
⇒ 母数に関することが知りたい

- 標本

母数の推測のため，母集団から選ばれた一部

標本を構成する個体の個数を標本の大きさ（標本サイズ）といい，一般に n で示す．

- 標本調査（サンプル調査）の考え方



- 標本から計算で求まる値

（標本平均，標本比率...）

標本分布と標準誤差（1）

- 母集団分布

  母集団が従う分布

  （正規分布を仮定することが多い）

  ※（確率）分布に対しては，期待値を平均と呼ぶことが多い．

- 標本

  母集団から大きさ n の標本として

  $X_1, X_2, \cdots, X_n$（大文字）

  を無作為抽出する ⇒ 確率変数

  実際に得られた値 $x_1, x_2, \cdots, x_n$（小文字）

  ※離散型変数を思い出しましょう $P(X = x_i)$

- 統計量

  標本 $X_1, X_2, \cdots, X_n$ のすべて，または一部の関数として表す量 ⇒ これも確率変数になる

- 標本平均（最もよく使われる統計量）

  $$\bar{X} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i$$

  母平均の推定に使われる．

  実測値 $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

  ※中央値も統計量

  推定量：推定のための統計量

  検定統計量：検定のための統計量

標本分布と標準誤差（2）

- 標本分布

  統計量は確率変数なので分布する．

  この分布のことを標本分布という．

  （例） 標本平均$\bar{X}$ の標本分布

✓ 母集団から大きさ 100 の標本を無作為抽出する

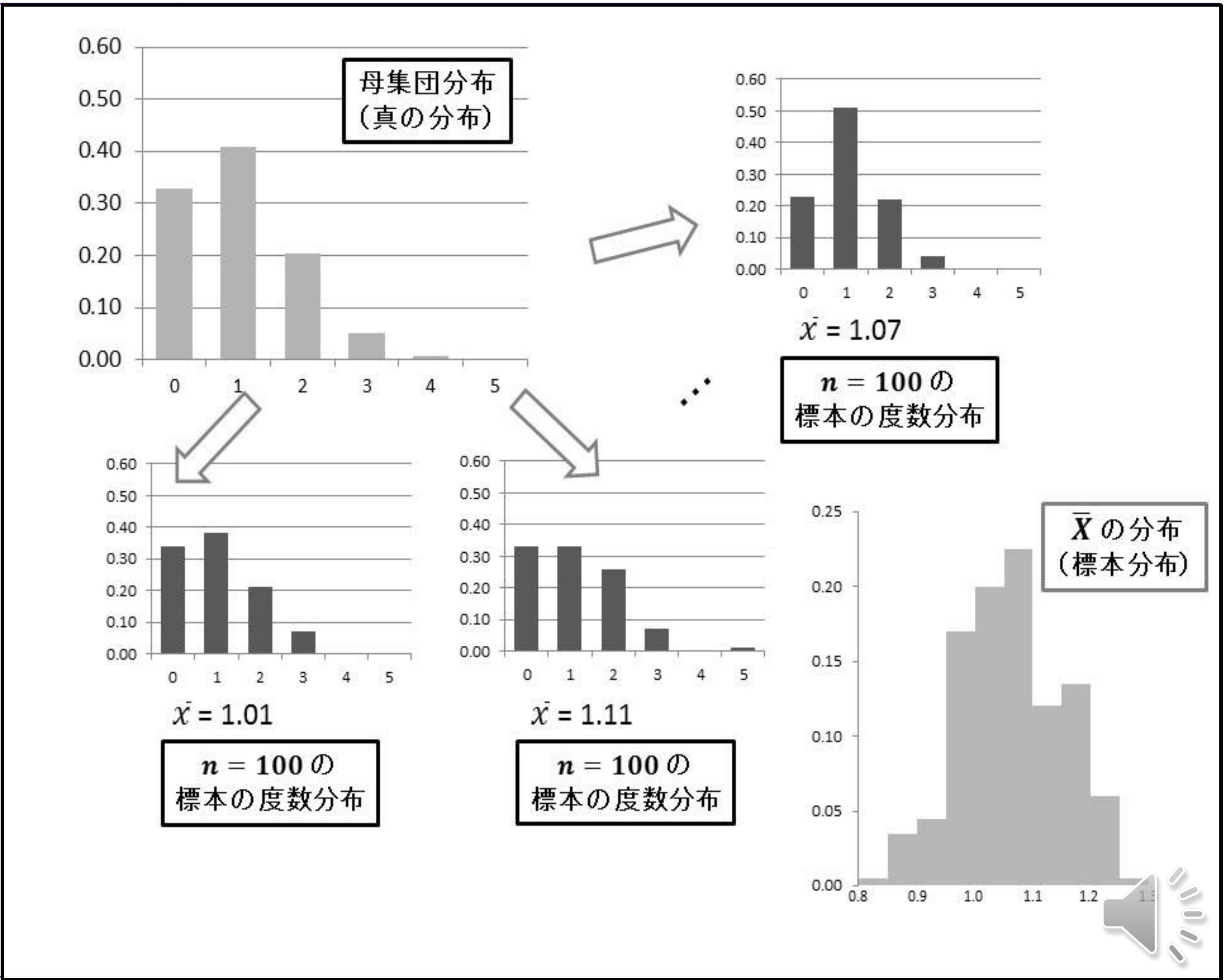
✓ ヒストグラム（度数分布図）を描く

✓ 実測値 $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$を計算する

  1.01, 1.11, … , 1.07

✓ $\bar{X}$ の分布（標本分布）が現れる

[シミュレーション標本分布.xlsm]



標本分布と標準誤差（3）

- 母集団分布の母平均 μ，母分散 σ^2

  母集団から大きさ n の標本として

  $X_1, X_2, \cdots, X_n$（大文字）を無作為抽出

  (母集団分布は正規分布でなくてもよい)

- 母平均 μ の推定量 ⇒ 標本平均 $\bar{X}$ を考える.

  重要な性質

  $\bar{X}$ の平均 $E[\bar{X}] = \mu$，$\bar{X}$ の分散 $V[\bar{X}] = \sigma^2/n$

  $\bar{X}$ の分布（標本分布）は n が大きいとき,

  正規分布 $N(\mu, \sigma^2/n)$ で近似できる.

  ※母集団分布が正規分布のときは<u>近似でなく</u>

  $\bar{X} \sim N(\mu, \sigma^2/n)$　“～：従う”

- 標準誤差

  <u>推定量$\hat{\theta}$</u> <u>の標準偏差</u>を標準誤差という.

  $\mathbf{se}(\hat{\theta})$，**se**，**s.e**，**SE** などと表す.

- 分散σ^2が既知の時，標本平均 $\bar{X}$ の標準誤差は

  $se[\bar{X}] = \sqrt{V[\bar{X}]} = \sigma/\sqrt{n}$ となる.

- 標本平均 $\bar{X}$ の標準誤差は

  <u>標本の大きさ n の平方根に反比例する</u>

  n が大きくなるほど，標準誤差は小さくなり，そのスピードは$\sqrt{n}$である.

σ^2が未知の時は手続きが複雑

標本分布と標準誤差（シミュレーション）

- 母平均 $\mu = 390$，母分散 $\sigma^2 = 20^2$ の
  母集団から大きさ $n = 8$ の標本
  $X_1, X_2, \cdots, X_8$を無作為抽出

- 母平均 μ の推定量として標本平均 $\bar{X}$ を考える．

重要な性質

平均 $E[\bar{X}] = \mu = 390$

分散 $V[\bar{X}] = \sigma^2/n = 400/8 = 50 ≒ 7.07^2$

標本平均 $\bar{X}$ の標準誤差

$$se[\bar{X}] = \sqrt{V[\bar{X}]} = \sigma/\sqrt{n} ≒ 7.07$$

[シミュレーション標本分布.xlsm]

- 中心極限定理

  標本平均 $\bar{X}$ の分布（標本分布）は
  n が大きいとき，正規分布 $N(\mu, \sigma^2/n)$
  で近似できる．

[シミュレーション中心極限定理.xlsx]

- （まとめ）

標本平均 $\bar{X}$ は n が大きくなるほど，
・正規分布 $N(\mu, \sigma^2/n)$ に近づく．
・ 標準誤差はスピード$\sqrt{n}$で小さくなる．

二項分布の正規近似は中心極限定理
の特殊ケース