区間推定

母数の推定

- 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，内閣支持率

  ⇒ 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

  n ：標本の大きさ（標本サイズ）

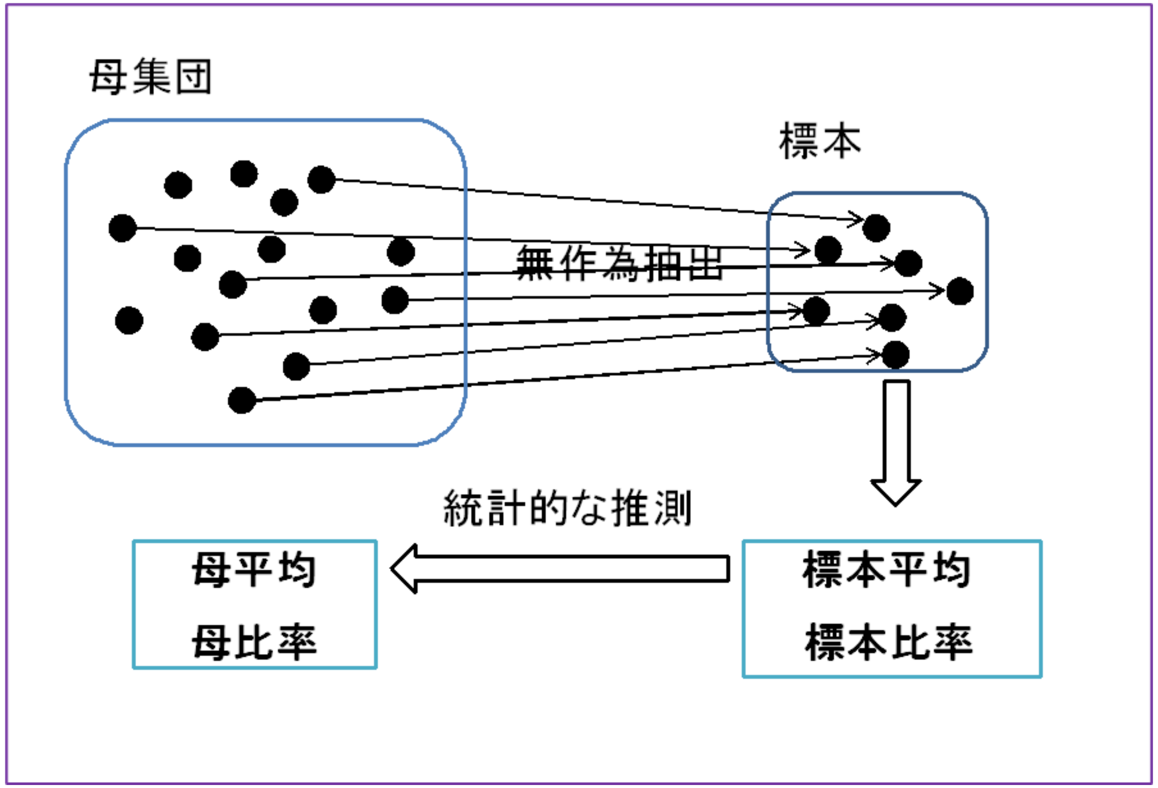
- 標本から母数（母平均，母比率）を推定

  標本平均　$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

  標本比率　$\hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i$

  成功（$X_i = 1$），失敗（$X_i = 0$）

- 標本調査（サンプル調査）の考え方



点推定と区間推定

- ここからの例

飲料水の内容量の母平均 μ を知りたい．

12本の製品を抜き取り中身を調べた．

363, 361, 363, 368, 366, 363

366, 365, 363, 365, 362, 363

標本平均 $\bar{x} = 364.0$

- 点推定

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

実測値 $\hat{p} = \frac{1}{n}\sum_{i=1}^{n} x_i$ ⇒ 点推定値

- 区間推定

母数 θ の推定を区間で示す．

$$[\hat{\theta} - \alpha,\ \hat{\theta} + \alpha]$$

この例では $[\bar{x} - \alpha,\ \bar{x} + \alpha]$ を考える．

- ＜重要＞ **母集団分布が正規分布**で

母平均 μ，母分散 σ^2（**既知**）

$X \sim N(\mu, \sigma^2) \Rightarrow \bar{X} \sim N(\mu, \sigma^2/n)$

n が大きいなら，正規分布の仮定は不要 ⇒ 正規近似ができる

（中心極限定理）

区間推定の考え方（1）

- 飲料水の内容量の仮定：

  内容量 X は,

  平均 μ ml，標準偏差2ml（分散4）の

  正規分布に従っているとする.

  $$X \sim N(\mu, 2^2)$$

  $n = 12$

- 利用する推定量（推定のための統計量）

  標本平均 $\bar{X}$ は,

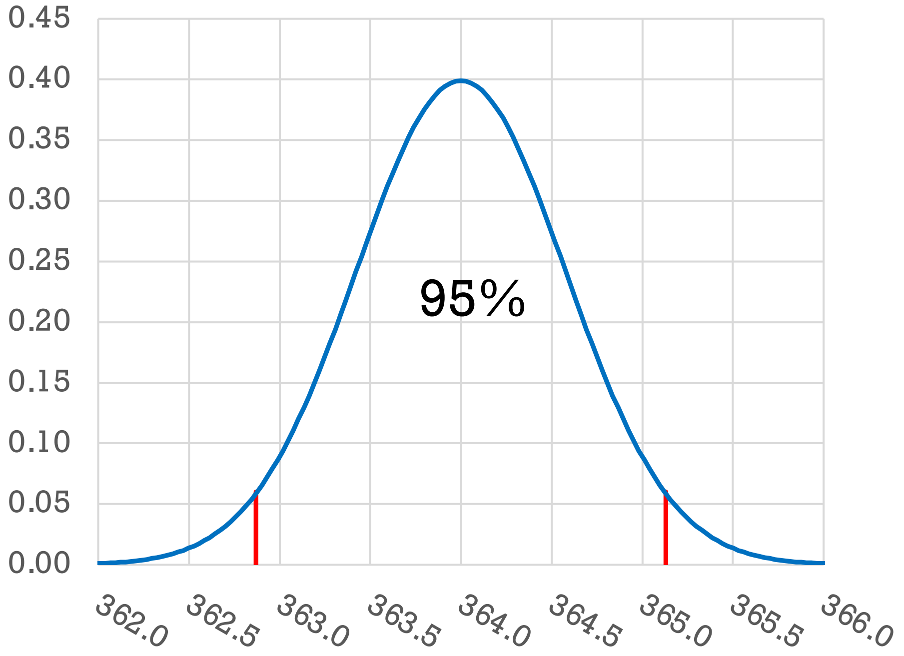
  $$\bar{X} \sim N\left(\mu, \frac{2^2}{12}\right)$$

  $\bar{x} = 364.0$，標準誤差 $= \sqrt{\frac{2^2}{12}}$

- 標準正規分布 $N(0,1)$表から片側2.5%（両側5%）になるところを探す ⇒ 1.96

  $$364.0 \pm 1.96\sqrt{\frac{2^2}{12}} \Rightarrow$$

  $$[364.0 - 1.1,\ 364.0 + 1.1] = [362.9,\ 365.1]$$



区間推定の考え方（2）

- 信頼係数（信頼度）と信頼区間

標準正規分布 $N(0,1)$表

✓ 95%信頼区間

⇒ 片側2.5%になるところ ⇒ 1.96

$$364.0 \pm 1.96\sqrt{\frac{2^2}{12}} \Rightarrow [364.0-1.1,\ \ 364.0+1.1]$$

$$= [362.9,\ 365.1]$$

✓ 90%信頼区間

⇒ 片側5%になるところ ⇒ 1.645

$$364.0 \pm 1.645\sqrt{\frac{2^2}{12}} \Rightarrow [364.0-0.9,\ \ 364.0+0.9]$$

$$= [363.1,\ 364.9]$$

- 信頼区間の性質

① 信頼係数が大きくなる

⇒ 信頼区間は広くなる

② 標本の大きさが大きくなる

⇒ 信頼区間は狭くなる

例　12個 ⇒ 48個（4倍）

$$1.96\sqrt{\frac{2^2}{12}} = 1.132,\quad 1.96\sqrt{\frac{2^2}{48}} = 0.566\ (1/2倍)$$

信頼区間の幅をどのくらいにしたいかで標本の大きさ n を決めることができる．

信頼区間の意味

- 信頼区間のシミュレーション

母集団分布　正規分布 $N(50, 10^2)$

標本の大きさ　$n = 9$

標準正規分布 $N(0, 1)$表

✓ 95%信頼区間

⇒ 片側2.5%になるところ ⇒ 1.96

$$\bar{x} \pm 1.96\sqrt{\frac{10^2}{9}} \Rightarrow [\bar{x} - 6.5, \bar{x} + 6.5]$$

たとえば，$\bar{x} = 52.0$ のとき

95%信頼区間は [45.5, 58.5] となり，

真値の 50 を含んでいる．

シミュレーション信頼区間.xls

✓ 100回の実験のうち

95回　くらいは真値を含む

5回　くらいは真値を含まない

⇒ これが 信頼係数95% の意味

×この区間に95%の確率で真値があるのではない．

対応がある場合の区間推定

- 対応のある例

ある学校では生徒10人に補講をした

次の表は，ほぼ同じ程度の試験の
　補講後（1行目）と補講前（2行目）
の試験結果と，補講後から補講前の点数を
引いた点数（3行目）である．

$\Rightarrow$ 前後に対応があるので点数差を考察

✓ 点数差 X は，母平均 μ，母分散$\sigma^2 = 4.5$(既知)

$$X \sim N(\mu, 4.5)$$

の正規分布に従っているとする．

✓ 90%信頼区間

$$1.8 \pm 1.645\sqrt{\frac{4.5}{10}}$$

$$\Rightarrow [1.8 - 1.1,\ \ 1.8 + 1.1] = [0.7,\ \ 2.9]$$

補講後	84	70	67	65	65	75	80	75	63	77
補講前	78	72	63	63	60	80	75	70	70	72
差	6	-2	4	2	5	-5	5	5	-7	5

✓ データから　$\bar{x} = 1.8$

母比率の区間推定

- 標本分布

  標本比率　$\hat{\boldsymbol{p}} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{X_i}$

  成功（$X_i = 1$），失敗（$X_i = 0$）

- 母比率の点推定

  実測値　$\hat{p} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x_i}$　$\Rightarrow$　点推定値

- （例）支持率

  無作為抽出で選ばれた 3000人 の内 300人 が支持していると答えた.

  $\Rightarrow \ \hat{p} = 0.1$

- 二項分布 $B(n,p)$

  母平均 np，母分散 $\sigma^2 = np(1-p)$

  標本の大きさ n が大きい ⇒ 正規近似

  $B(3000,p) \ \rightarrow N(3000p, 3000p(1-p))$

  $\rightarrow \hat{\boldsymbol{p}} \sim N(p, \ \frac{p(1-p)}{3000})$

- 95%信頼区間

  $0.1 \pm 1.96\sqrt{\frac{0.1\times0.9}{3000}}$

  $\Rightarrow \ [0.1 - 0.01, \ 0.1 + 0.01] = [0.09, \ 0.11]$

ひとつひとつ手順を踏むとわかるようになります