

『AI・データサイエンスのための数学入門』学習補助資料

信頼区間を正しく読み、使う

95% の意味・標準誤差・ t 分布・標本サイズ

1 動くのは母数ではなく、区間

母平均 μ は固定された未知の値です。一方、標本を取り直すと標本平均が変わるので、その周りに作る区間も動きます。95% という数値は、**区間を作る手順の被覆確率**を表します。

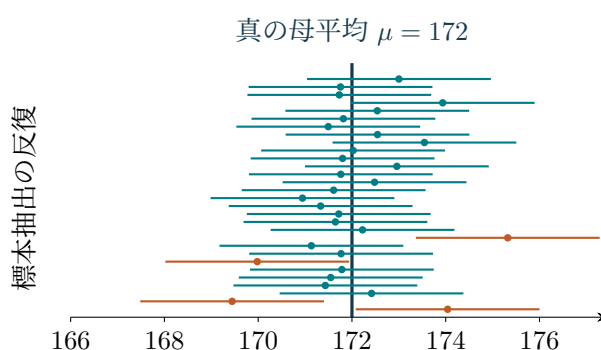


図 1: $\mu = 172$, $\sigma = 10$, $n = 100$ の正規モデルで、標本平均を 30 回生成した例。点は標本平均、横線は $\bar{x} \pm 1.96$ 。この例では 26 区間が母平均を含み、橙色の 4 区間は含みません（乱数シード 3）。

30 回なら必ず 95% になる、という意味ではない

95% は長期的な割合です。有限回の実験では含む区間の割合も変動します。独立に 20 回繰り返した場合、外れる回数の期待値は $20 \times 0.05 = 1$ ですが、必ず 1 回だけ外れるわけではありません。

観測前の確率と、観測後の区間を区別する

データを X 、区間の両端を $L(X), U(X)$ と書けば、正確な 95% 信頼区間の手順は

$$\Pr_{\mu}\{L(X) \leq \mu \leq U(X)\} = 0.95$$

を満たします。観測後の区間 $[L(x), U(x)]$ は固定され、母平均を含むか含まないかのどちらかで。この手順から、母平均自体の確率分布が与えられるわけではありません。

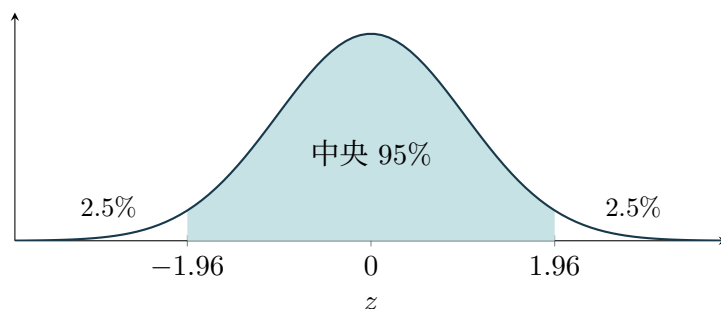
読み方：2～5 節で公式・図・注意点、6 節で本文の問題、7～8 節で追加問題と解答を扱います。基本モデルは独立同分布の標本です。

2 「推定値 ± 誤差マージン」の導き方

まず、 $X_1, \dots, X_n$ が独立に正規分布 $N(\mu, \sigma^2)$ に従い、母標準偏差 σ が既知とします。このとき

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right), \quad Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1).$$

標準正規分布では、中央の 95% に対応する境界が約 ± 1.96 です。



同じ事象を、 μ について書き直す

$c = z_{0.975} \approx 1.96$ と置きます。次の不等式はすべて同値です。

$$\begin{aligned} -c \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq c &\iff -c \frac{\sigma}{\sqrt{n}} \leq \bar{X} - \mu \leq c \frac{\sigma}{\sqrt{n}} \\ &\iff \bar{X} - c \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + c \frac{\sigma}{\sqrt{n}}. \end{aligned}$$

最初の事象の確率が 0.95 なので、最後の事象の確率も 0.95 です。観測後に $\bar{X}$ を $\bar{x}$ に置き換えると、具体的な信頼区間になります。

例：平均身長 の推定

$\bar{x} = 172.0$ cm、 $\sigma = 10$ cm、 $n = 100$ なら、標準誤差は 1 cm です。

$$172.0 \pm 1.96 \times 1 = [170.04, 173.96] \text{ cm}.$$

この区間は**母平均**の推定区間です。「個々の人の身長の 95% がこの区間に入る」という意味ではありません。

正確な公式と正規近似

正規母集団で σ が既知なら、この構成は正確です（分位点の丸めを除く）。非正規母集団では、独立性・有限分散などのもとで中心極限定理による近似になります。「 $n \geq 30$ なら常に十分」という一律の保証はありません。

3 母分散が未知なら、なぜ t 分布なのか

実際には σ がわからないことが多いため、標本から

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2, \quad S = \sqrt{S^2}$$

を計算します。観測値を s^2, s と書きます。ここで使う s は、**不偏分散の平方根**です。 S 自体が母標準偏差の不偏推定量という意味ではありません。

分母も標本ごとに動く

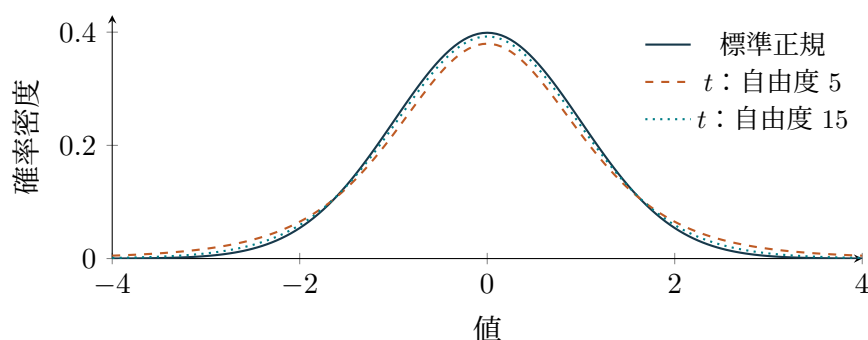
正規母集団からの独立標本では、次の事実が成り立ちます。

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1), \quad U = \frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2,$$

さらに Z と U は独立です。この性質を用いると

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} = \frac{Z}{\sqrt{U/(n-1)}} \sim t_{n-1}.$$

これは、正規変数を独立な分散推定の揺らぎで割った分布です。 S が小さく出ると比の絶対値が大きくなりやすいため、正規分布より裾が厚くなります。



両側 95% の分位点は、自由度 5 で約 2.571、自由度 15 で約 2.131、標準正規分布で約 1.960 です。自由度が大きくなると t 分布は標準正規分布に近づきます。

母平均の t 信頼区間

正規母集団・独立標本・ $n \geq 2$ のもとで、

$$\bar{x} \pm t_{n-1, 1-\alpha/2} \frac{s}{\sqrt{n}}$$

が正確な信頼区間になります。 $s/\sqrt{n}$ は**標準誤差の推定値**です。非正規母集団での使用は一般に近似であり、特に小標本や強い歪み・外れ値には注意が必要です。

4 区間の幅を決める 3 つの量

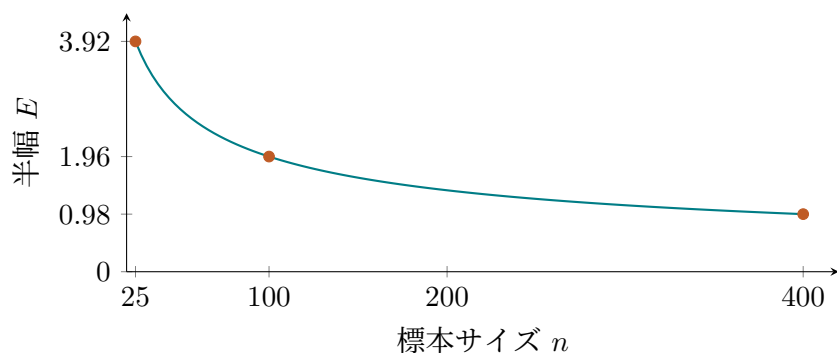
母標準偏差が既知のとき、誤差マージン（半幅）を E と書けば

$$E = z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}, \quad \text{区間全体の幅} = 2E.$$

信頼水準を上げると広く、ばらつきが大きいと広く、標本数を増やすと狭くなります。

信頼水準	分位点	半幅 E	全幅 $2E$
90%	1.645	1.645	3.290
95%	1.960	1.960	3.920
99%	2.576	2.576	5.152

表： $\sigma = 10$ 、 $n = 100$ を固定した比較。高い信頼水準と狭い区間幅は、そのままでは両立しません。



図：95% 信頼区間、 $\sigma = 10$ の場合。 $n = 25, 100, 400$ と 4 倍にするたび、半幅は 3.92, 1.96, 0.98 と半分になります。

必要な標本数を逆算する

半幅を E_0 以下にしたいなら、

$$z \frac{\sigma}{\sqrt{n}} \leq E_0 \iff n \geq \left(\frac{z\sigma}{E_0} \right)^2.$$

例えば 95%、 $\sigma = 10$ 、半幅 2 以下なら $n \geq (1.96 \cdot 10/2)^2 = 96.04$ なので、**切り上げて 97** が必要です。

平方根の法則の範囲

この比例関係は、信頼水準と σ を固定した既知分散の場合のものです。 t 区間では分位点も n に依存し、 s も変動するため、標本数を 4 倍にした実際の区間幅が必ず半分になるわけではありません。

5 母比率：0・1 データの平均を推定する

購入した人を $Y_i = 1$ 、購入しなかった人を $Y_i = 0$ とすると、独立なベルヌーイ標本では

$$E[Y_i] = p, \quad V[Y_i] = p(1-p), \quad \hat{p} = \frac{1}{n} \sum_i Y_i.$$

したがって、標本平均と同じ計算から

$$E[\hat{p}] = p, \quad V[\hat{p}] = \frac{p(1-p)}{n}.$$

真の標準誤差は $\sqrt{p(1-p)/n}$ ですが、 p は未知です。 p を $\hat{p}$ に置き換えて得るのが、**推定標準誤差** $\sqrt{\hat{p}(1-\hat{p})/n}$ です。

本文の購入率の例を、途中で丸めずに計算する

$n = 1000$ 、購入者 120 人なら $\hat{p} = 0.12$ です。

$$\widehat{SE} = \sqrt{\frac{0.12 \cdot 0.88}{1000}} \approx 0.010276,$$

$$E = 1.96 \times 0.010276 \approx 0.020141, \quad CI \approx [0.09986, 0.14014].$$

百分率では約 [9.99%, 14.01%]、本文の丸め方では [10.0%, 14.0%] です。半幅は約 **2.01 パーセントポイント**で、「購入率の 2.01%」ではありません。

正規近似の区間には弱点がある

この式は Wald（ワルド）区間と呼ばれます。成功数・失敗数が少ないと、被覆確率が 95% から大きくずれることがあります。本文の「両方 5 以上」は目安であり、正確さの保証ではありません。

例えば 20 人中 0 人なら、式は $\hat{p} = 0$ から $[0, 0]$ を返します。しかし、標本で購入者が 0 人だっただけで、母比率が 0 と確定するわけではありません。推定式が不確実性を捉え損なっています。

発展：Wilson 区間という補い方

$z = z_{1-\alpha/2}$ として、次の区間がよく使われます。

$$\frac{\hat{p} + z^2/(2n) \pm z\sqrt{\hat{p}(1-\hat{p})/n + z^2/(4n^2)}}{1 + z^2/n}.$$

未知の p を含む不等式 $(\hat{p} - p)^2 \leq z^2 p(1-p)/n$ を、 p について解くと得られます。0 人の場合でも幅が 0 になりません。ただし、これも被覆確率が常に厳密に 95% となる方法ではありません。

抽出方法も前提の一部：広告を見た 1000 人が、推定したい対象をどう代表しているかを確認します。選択バイアス、無回答、同じ人の反復観測などによる影響は、この区間を計算するだけでは解消されません。

6 本文の練習問題と検定との対応

問 1：母分散未知の場合の信頼区間

正規母集団からの独立標本を仮定します。 $n = 16$ 、 $\bar{x} = 50$ 、 $s = 8$ なので、推定標準誤差は

$$\frac{s}{\sqrt{n}} = \frac{8}{\sqrt{16}} = 2.$$

自由度は $n - 1 = 15$ 。指定の分位点 2.13 を使うと

$$50 \pm 2.13 \times 2 = 50 \pm 4.26 = [45.74, 54.26].$$

母分散が既知なら $\sigma/\sqrt{n}$ と標準正規分位点を使います。未知なら $s/\sqrt{n}$ と t 分位点を使います。未知分散の小標本で、単純に 1.96 を使うと、分散推定の不確実性を十分に織り込めません。

問 2：信頼区間と両側検定の対応

$[46.08, 53.92]$ の中心は $(46.08 + 53.92)/2 = 50$ 、半幅は 3.92、全幅は 7.84 です。「誤差幅」という言葉は曖昧なので、半幅か全幅かを明示するとよいでしょう。

同じモデル・標準誤差・分位点を用いる対応する 5% 両側検定では、

- $H_0: \mu = 55:55$ は区間外なので、**棄却します**。
- $H_0: \mu = 50:50$ は区間内なので、**棄却しません**。

棄却しないことは、母平均が 50 であると証明したことではありません。なお、問 1 と問 2 の区間は数値が異なり、同一の区間ではありません。

区間と検定が対応する理由

既知分散の場合、帰無仮説の値を μ_0 とすると、両側検定で棄却しない条件は

$$\left| \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \right| \leq z_{0.975}.$$

これを μ_0 について解くと

$$\bar{x} - z_{0.975} \frac{\sigma}{\sqrt{n}} \leq \mu_0 \leq \bar{x} + z_{0.975} \frac{\sigma}{\sqrt{n}}.$$

つまり、信頼区間は「この両側検定で棄却されない母平均の値全体」です。 t 区間でも対応する t 検定について同じ関係が成り立ちます。

誤解しやすい 3 点

95% 信頼区間は、個々の観測値の 95% を含む範囲でも、推定誤差の絶対的な上限でもありません。また、標本数が多くて区間が狭くても、抽出の偏りまで小さいとは限りません。

7 追加の練習問題

必要に応じて $z_{0.975} = 1.96$ 、 $z_{0.995} = 2.576$ を使いなさい。母平均の問題では、正規母集団からの独立標本を仮定します。

問 3：信頼水準の解釈

次の説明のどこが誤っているか、修正せよ。

- (1) 95% 信頼区間には、標本データの 95% が含まれる。
- (2) 95% 信頼区間を 20 回作ると、必ず 19 回は母平均を含む。
- (3) ある 95% 信頼区間が得られたので、母平均自体が 95% の確率でこの区間内にあると頻度論的にいえる。

問 4：母分散既知の計算

$n = 100$ 、 $\bar{x} = 80$ 、 $\sigma = 15$ とする。母平均の 95% 信頼区間、半幅、全幅を求めよ。

問 5：信頼水準を上げるとどうなるか

問 4 のデータから 99% 信頼区間を求めよ。95% 信頼区間と比較して、幅が変わる理由を説明せよ。

問 6：必要標本数

$\sigma = 12$ が既知で、95% 信頼区間の半幅を 2 以下にしたい。必要な標本数の最小値を求めよ。半幅を 1 以下にする場合も求めよ。

問 7： t 信頼区間

$n = 25$ 、 $\bar{x} = 100$ 、 $s = 10$ 、 $t_{24,0.975} = 2.064$ とする。推定標準誤差と 95% 信頼区間を求めよ。

問 8：母比率の推定

独立な 400 件の観測で、成功が 100 件あった。母比率の Wald 型の約 95% 信頼区間を求め、百分率でも示せ。正規近似の目安も確認せよ。

問 9：信頼区間と検定

ある母平均の 95% 信頼区間が $[8.2, 11.8]$ だった。対応する 5% 両側検定で、 $H_0 : \mu = 12$ と $H_0 : \mu = 10$ をそれぞれどう判断するか。後者を棄却しないことは何を意味しないか。

発展：問 10

95% 信頼区間を独立に 20 回構成する。少なくとも 1 回は母数を含まない確率を求めよ。ただし、各区間の被覆確率は正確に 0.95 とする。

8 追加問題の解答と考え方

問 3

(1) 信頼区間が推定するのは母数であり、個々のデータが入る割合ではありません。(2) 19 回というのは含む回数の期待値です。有限回の包含回数は変動します。(3) 95% は標本抽出を繰り返したときの区間構成手順の被覆確率です。観測後の固定区間に対して、固定された母数の確率をこの手順だけから割り当てることはできません。

問 4

標準誤差は $15/\sqrt{100} = 1.5$ 。半幅は $1.96 \times 1.5 = 2.94$ 、全幅は 5.88 です。区間は

$$80 \pm 2.94 = [77.06, 82.94].$$

問 5

半幅は $2.576 \times 1.5 = 3.864$ なので、 $[76.136, 83.864]$ 。全幅は 7.728 です。同じ標本数・ばらつきのもとで、より高い包含率を得るには区間を広げる必要があります。

問 6

半幅 2 以下では $n \geq (1.96 \cdot 12/2)^2 = 138.2976$ なので **139**。半幅 1 以下では $n \geq (1.96 \cdot 12)^2 = 553.1904$ なので **554** です。丸め前の必要数は 4 倍ですが、整数への切り上げ後は正確に 4 倍とは限りません。

問 7

推定標準誤差は $10/\sqrt{25} = 2$ 。半幅は $2.064 \times 2 = 4.128$ なので、区間は $[95.872, 104.128]$ です。

問 8

$\hat{p} = 100/400 = 0.25$ 、推定標準誤差は

$$\sqrt{\frac{0.25 \cdot 0.75}{400}} \approx 0.021651.$$

半幅は約 0.042435 なので、区間は約 $[0.2076, 0.2924]$ 、百分率では約 $[20.76\%, 29.24\%]$ です。成功数 100、失敗数 300 はどちらも 5 以上ですが、この目安だけで正確な 95% 被覆が保証されるわけではありません。

問 9

12 は区間外なので棄却、10 は区間内なので棄却しません。後者は「母平均が 10 だと証明された」「帰無仮説が 95% の確率で正しい」という意味ではありません。

問 10

独立性より、20 区間すべてが母数を含む確率は 0.95^{20} 。したがって、求める確率は

$$1 - 0.95^{20} \approx 0.6415 \quad (\text{約 } 64.2\%).$$

1 回の不包含確率が 5% でも、20 回のうち少なくとも 1 回外れることは珍しくありません。「外れる回数の期待値が 1」と「1 回以上外れる確率」は異なる量です。