経済統計学講義ノート No.7

推定

蛭川雅之

2024 年 11 月 7 日

目　次

1　推定とは

- 推測統計は推定と検定を含む。
  1. 推定：標本から母集団分布の未知母数（例：平均、標準偏差. . . ）の値を推測する。
  2. 検定：分析者が想定する母集団に関する仮説が正しいかを標本に照らしあわせて客観的に判断する。
- 推定は以下の 2 種類に大別される。
  1. 点推定：未知母数を唯一の値で推定する。
  2. 区間推定：未知母数をある区間で推定する。

2　点推定

2.1　点推定の一例

- 母集団の平均 μ を大きさ n の標本 $X_1, \ldots, X_n$ を用いて推定したい。
- 既に述べた通り、標本平均

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

  が候補の一つとなる。
  - このとき、標本平均 $\bar{X}$ を母集団の平均 μ の推定量であるという。

2.2 幾つかの疑問

1. 推定量と推定値の違いは何か？
   - 推定量：
     - 推定の方法・構造を指す。
     - 標本の関数であり、確率変数である。
   - 推定値：
     - 推定量から算出された数値を指す。
2. 未知母数 (例えば、母集団の平均 μ) に対する推定量はただ一つか？
   - 一つに限定されない。
   - モーメント法による推定（＝標本平均 $\bar{X}$）のほか、最尤法による推定も考えられる。

2.3 推定量の望ましい性質

- 理想的には、以下の性質を同時に満たす推定量が望ましい。
  1. 不偏性
  2. 一致性
  3. 有効性
- 実際には、状況に応じていずれかの基準を優先する。
- 以下、推定したい未知母数を θ、推定量を $\hat{\theta}$ と表記する。

2.3.1　不偏性

定義 1 推定量 $\hat{\theta}$ が

$$E\left(\hat{\theta}\right) = \theta$$

を満たすとき、$\hat{\theta}$ は θ の不偏推定量であるという。

- 「$\hat{\theta}$ は不偏性を持つ」ともいう。
- 推定量 $\hat{\theta}$ が θ の不偏推定量であるとは、推定量 $\hat{\theta}$ のバイアス

$$b\left(\hat{\theta}\right) = E\left(\hat{\theta}\right) - \theta$$

  がゼロであることに他ならない。

2.3.2　一致性

定義 2 推定量 $\hat{\theta}$ が $n \to \infty$ に対し

$$\hat{\theta} \xrightarrow{p} \theta$$

を満たすとき、$\hat{\theta}$ は θ の一致推定量であるという。

- $E\left(\hat{\theta}\right) = \theta$ かつ $Var\left(\hat{\theta}\right) = 0$ が理想であるが、現実には成立しそうにない。
  - 十分大きな n に対し、これらが近似的に成立することは依然として望ましい。
- 一致性は推定量が最小限満たすべき性質である。
  - 一致性を持たない $\Rightarrow$ 大標本を利用する意味はあるか？

2.3.3　有効性

定義 3 θ に関して不偏性を持つ 2 つの推定量 $\hat{\theta}$ と $\hat{\theta}^*$ に対し

$$Var\left(\hat{\theta}\right) \leq Var\left(\hat{\theta}^*\right)$$

が成り立つとき、$\hat{\theta}$ は $\hat{\theta}^*$ より有効であるという。

定義 4 θ に関して不偏性を持つ推定量 $\hat{\theta}$ がどのような不偏推定量 $\hat{\theta}^*$ に対しても

$$Var\left(\hat{\theta}\right) \leq Var\left(\hat{\theta}^*\right)$$

を満たすとき、$\hat{\theta}$ を最小分散不偏推定量あるいは有効推定量であるという。

2.3.4 平均 2 乗誤差

- 以下のような推定量を比較する場合、どちらが望ましいのか？
  1. 小さいバイアスと小さい分散を持つ推定量。
  2. 不偏であるが大きい分散を持つ推定量。

定義 5 θ の推定量 $\hat{\theta}$ に関する平均 2 乗誤差 (Mean Squared Error, "MSE") は

$$MSE\left(\hat{\theta}\right) = E\left(\hat{\theta} - \theta\right)^2$$

と定義される。

- 平均 2 乗誤差は次のように分解できる。

$$MSE\left(\hat{\theta}\right) = \left\{b\left(\hat{\theta}\right)\right\}^2 + Var\left(\hat{\theta}\right) = \left(\hat{\theta}\text{のバイアスの 2 乗}\right) + \left(\hat{\theta}\text{の分散}\right)$$

3　正規母集団の平均の区間推定

3.1　区間推定とは

- 母数がどのような値をとるかを一点で求めるのでなく、特定の確実さでどのような区間に入っているかを求める。
  - 特定の確実さ = 信頼係数(例:$100\,(1-\alpha)\,\% = 90\%, 95\%, 99\%$)
  - 区間 = 信頼区間
  - 信頼係数 $\uparrow(\downarrow) \Rightarrow$ 誤っている「確率」$\downarrow(\uparrow)$・区間幅 $\uparrow(\downarrow)$
- 最もよく使われる「母集団の平均に関する区間推定」に特化する。
  - 正規母集団 $\Rightarrow$ 正規分布（分散既知）・t 分布（分散未知）
  - 未知の母集団 $\Rightarrow$ 正規分布（中心極限定理に基づく正規近似）
- 同一の信頼係数を持つ区間の作り方は無限にある。
  - 区間幅を最短にする観点から、信頼区間を左右対称にとる。

3.2 正規母集団の分散が既知の場合

- 正規母集団 $N\left(\mu, \sigma^2\right)$ から大きさ n の標本 $X_1, \ldots, X_n$ を無作為に抽出するとき、標本平均 $\bar{X}$ の確率分布は正規分布 $N\left(\mu, \sigma^2/n\right)$ である。
  - $\bar{X}$ を標準化したものの確率分布は標準正規分布である。

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

- 信頼係数を 95% にとると、

$$0.95 = \Pr\left\{-1.96 \leq \frac{\sqrt{n}\left(\bar{X} - \mu\right)}{\sigma} \leq 1.96\right\}$$

である。

- 標本平均 $\bar{X}$ の実現値を $\bar{x}$ と表記し、連立不等式

  $$-1.96 \leq \frac{\sqrt{n}\,(\bar{x}-\mu)}{\sigma} \leq 1.96$$

  を μ について解くことにより、分散既知の場合の μ に関する 95% 信頼区間

  $$\bar{x} - 1.96\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}$$

  が得られる。

問題 6 A 社で生産される乾電池の寿命は標準偏差 200 時間の正規分布に従うことが経験的にわかっているとする。いま 16 個の乾電池を無作為に抽出し検査したところ、その平均寿命（＝標本平均）は 3098 時間であった。同社で生産される乾電池全体の平均寿命に関する 95% 信頼区間を求めよ。

3.3 正規母集団の分散が未知の場合

- 実のところ、正規母集団の分散が既知であるとは考えにくい。
  - 母集団の分散が未知の場合、標本分散 $\hat{\sigma}^2$ もしくは不偏分散 s^2 で代用する。
- 標本平均 $\bar{X}$ と標本分散

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$$

を用いて '標準化' した t 統計量の確率分布は、正規分布でなく自由度 $(n-1)$ の t 分布になる。

$$t = \frac{\sqrt{n-1}\left(\bar{X} - \mu\right)}{\hat{\sigma}} \sim t\left(n-1\right)$$

– 標本分散 $\hat{\sigma}^2$ の代わりに不偏分散

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$$

を用いると、t 統計量は以下のように書き改められる。

$$t = \frac{\sqrt{n}\left(\bar{X} - \mu\right)}{s} \sim t\left(n-1\right)$$

- 信頼係数を 95% にとり、自由度 $(n-1)$ の t 分布の上側確率が 2.5% となる分位点を $t_{0.025}\left(= t_{0.025}\left(n-1\right)\right)$ と表記すると、

$$0.95 = \Pr\left\{-t_{0.025} \leq \frac{\sqrt{n-1}\left(\bar{X} - \mu\right)}{\hat{\sigma}} \leq t_{0.025}\right\}$$

である。

- 標本平均 $\bar{X}$ の実現値を $\bar{x}$ と表記し、連立不等式

$$-t_{0.025} \leq \frac{\sqrt{n-1}\,(\bar{x}-\mu)}{\hat{\sigma}} \leq t_{0.025}$$

を μ について解くことにより、分散未知の場合の μ に関する 95% 信頼区間

$$\bar{x} - t_{0.025}\frac{\hat{\sigma}}{\sqrt{n-1}} \leq \mu \leq \bar{x} + t_{0.025}\frac{\hat{\sigma}}{\sqrt{n-1}}$$

が得られる。

問題 7 あるスーパーマーケットで無作為に選んだ 9 日にイチゴの販売量 (単位:パック) を調べたところ、66　75　73　62　70　65　77　64　78 であった。このスーパーマーケットにおけるイチゴの 1 日当たり販売量が正規分布に従うものと仮定し、販売量の平均に関する 95% 信頼区間を求めよ。【ヒント：Excel の関数 `tinv` が t 分布の $100q$% 分位点を与える。】

4　未知母集団の平均の区間推定

4.1　一般ケース

- 現実には、母集団が正規分布であることすら適切でない場合が多い。
  - 母集団の分布は特定しないが、母集団の平均 μ および分散 σ^2 が存在するとの仮定は維持する。
- まず、分散が既知であるとする。
  - 中心極限定理により、十分大きい n に対し標本平均 $\bar{X}$ の確率分布は $N\left(\mu, \sigma^2/n\right)$ で近似できる。
  - 正規母集団で分散が既知の場合と同様に、次のような母集団の平均 μ に関する近似的 95% 信頼区間が得られる。

$$\bar{x} - 1.96\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}$$

- 分散が未知の場合はどうするか？
  - $\hat{\sigma}^2$ および s^2 はともに σ^2 の一致推定量であるから、上の信頼区間にある σ を $\hat{\sigma}$ もしくは s で置き換えればよい。
  - 例えば $\hat{\sigma}$ を使用する場合、母集団の平均 μ に関する近似的 95% 信頼区間は

$$\bar{x} - 1.96\frac{\hat{\sigma}}{\sqrt{n}} \leq \mu \leq \bar{x} + 1.96\frac{\hat{\sigma}}{\sqrt{n}}$$

となる。

問題 8 中学受験を目指す小学 6 年生の父母 400 名を対象に、1 か月の補習教育費（＝塾・家庭教師等の費用）を調査したところ、平均は 32,000 円、また標準偏差は 5,500 円であった。母集団の分布形は未知として、中学受験を目指す小学 6 年生の父母が 1 か月に負担する補習教育費の平均に関する 95% 信頼区間を求めよ。

4.2　母集団比率の区間推定

- 母集団比率 p を区間推定する問題を考える。
  - 母集団に成功確率 p のベルヌーイ分布を仮定する。
- 中心極限定理により、十分大きい n に対し標本平均 $\bar{X}$（＝標本の成功比率 $\hat{p}$）の分布は正規分布 $N\left(p, p\left(1-p\right)/n\right)$ で近似できる。
  - 信頼係数が 95% の場合、

$$0.95 \approx \Pr\left\{-1.96 \leq \frac{\sqrt{n}\left(\hat{p}-p\right)}{\sqrt{p\left(1-p\right)}} \leq 1.96\right\}$$

であるから、以下のような母集団比率 p に関する近似的 95%「信頼区間」が得られる。

$$\hat{p}-1.96\sqrt{\frac{p\left(1-p\right)}{n}} \leq p \leq \hat{p}+1.96\sqrt{\frac{p\left(1-p\right)}{n}}$$

- $\sqrt{p(1-p)/n}$ に未知母数 p が含まれているが、どうするか？
  - この部分の p を一致推定量 $\hat{p}$ に置き換え、最終的に、母集団比率 p に関する近似的 95% 信頼区間は

$$\hat{p} - 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

となる。

問題 9 あるスポーツのテレビ中継の視聴率を 900 人に対して調査したところ、324 人が「見た」と答えた。このテレビ中継の視聴率（母集団比率）に関する 95% 信頼区間を求めよ。