

# 統計学 2

## 12. 標本平均と母平均 (2)

矢内 勇生

2019年5月27日

高知工科大学 経済・マネジメント学群

# 今日の内容

- 標本平均を利用した母平均の推定をマスターする！
  - 点推定
  - 区間推定、信頼区間

# 標本平均から母平均を 推測する：点推定

- 点推定：母数を1つの値で推定する
- 母平均の点推定値 (point estimate) = 標本平均
  - 標本平均の平均（標本分布の中心）は母平均に一致する
  - 標本平均は母平均の不偏推定量

# 不偏性 (unbiasedness)

- 標本平均 ( $\bar{x}$ ) には誤差がある :  $\bar{x} \neq \mu_x$
- しかし、標本平均の平均は母平均 ( $\mu_x$ ) に一致する

$$E(\bar{x}) = \mu_x$$

- この性質を不偏性と呼ぶ : 推定値に望まれる性質の1つ

# 標本平均から母平均を 推測する：区間推定

- 区間推定：母数を区間で推定する

= 推定に幅を持たせる

- 区間推定に用いる区間：信頼区間 (confidence interval: CI)

# 標準正規分布の特徴を 利用して推測する

- 標準正規分布の特徴： $[-1.96, 1.96]$  の区間にデータの95%が収まる
- 正規分布に従う変数を標準化することで、標準正規分布を使える
- ★ 標本サイズ ( $n$ ) が大きくなれば、誤差の分布は正規分布に近づく（中心極限定理）

# 標本平均を標準化する

- 標本平均の平均 = 母平均  $\mu$
- 標本平均の標準偏差 = 標準誤差 SE

➡ 標本平均の z 値は、

$$z = \frac{\bar{x} - \mu}{\text{SE}} = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

# z 値の95%が $[-1.96, 1.96]$ にある

- 標本平均のz 値のうち、95%は 区間  $[-1.96, 1.96]$  に収まるはず
- つまり、たくさんある標本の95%について、次の式が成り立つ：

$$-1.96 \leq z \leq 1.96$$

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



# 95%信頼区間を求める (1)

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

- 既知のもの：  $n, \bar{x}$  ( $\sigma$  も知っているとする)
- 推定の対象：  $\mu$
- 上の不等式を  $\mu$  について解けば、  $\mu$  (母平均) の95%信頼区間が得られる

# 95%信頼区間を求める (2)

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

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

➡  $\mu$  の95%信頼区間は

$$\left[ \bar{x} - 1.96 \frac{\sigma}{\sqrt{n}}, \bar{x} + 1.96 \frac{\sigma}{\sqrt{n}} \right]$$

$$= [\bar{x} - 1.96 \cdot \text{SE}, \bar{x} + 1.96 \cdot \text{SE}]$$

# 95%以外の信頼区間を求める

- 求めるパーセントに応じて、1.96 の代わりに適切な数字を選ぶ
  - 標準正規分布表を使う
  - R では `qnorm()` で求める)

(例)

- 50% : 標本平均  $\pm 0.67SE$
- 99% : 標本平均  $\pm 2.58SE$
- 99.9% : 標本平均  $\pm 3.29SE$

# qnorm() の使い方

- 標準正規分布で分布の中央部分95%の区間がどこにあるか求める方法
  - 正規分布は左右対称：中央95%を得るには、両側2.5% (0.025) ずつ除外すればよい
  - つまり、標準正規分布で、左端から2.5%の点と97.5%の点を見つければよい
  - Rでは

```
qnorm(c(0.025, 0.975))
```

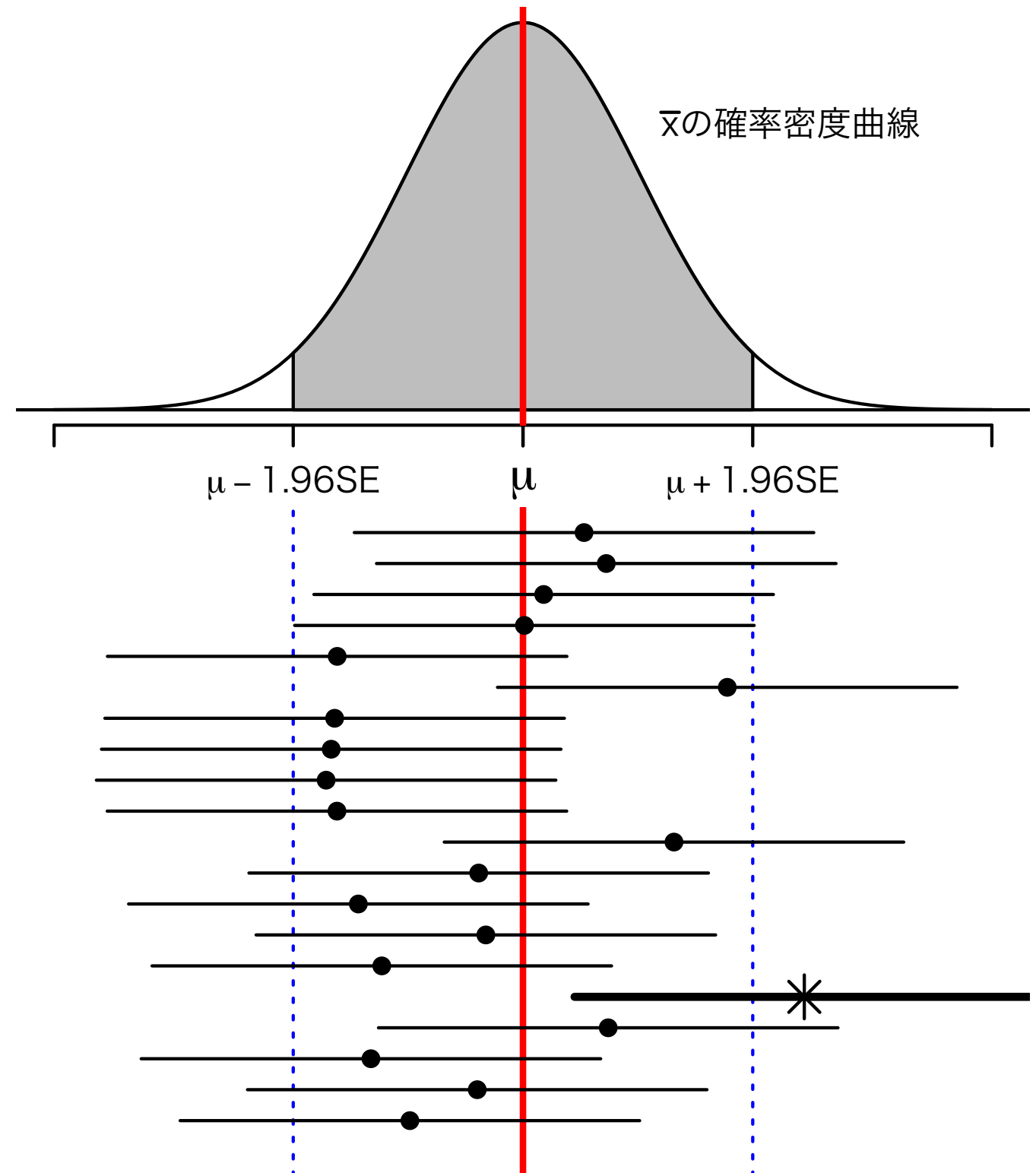
とすると、-1.96 と 1.96 という結果が得られる

# 信頼区間についての注意

- 95%信頼区間  $\neq$  母平均がその区間にある確率が95%
- 95%信頼区間 = 標本をたくさん抽出し、同じ手順でそれぞれの標本から信頼区間を求めたとき、母平均を含んでいる区間を得る確率が95%
- よって、1つの標本から得た95%信頼区間が真の平均を含んでいる確率は1か0（つまり、含んでいるか含んでいないかのどちらか）

# 様々な標本の95%信頼区間

- 標本によって、信頼区間は変わる
- 通常、手元には1つの標本しかない
- 手元にある標本の信頼区間は、「母数を含む」か「母数を含まない」のどちらか一方



# 練習問題

- ある人の血圧の計測値を母集団とすると、現在の実際の血圧 $\mu$ を母平均として、母標準偏差が10の正規分布をしている。

(1) 1回だけ血圧を測ったら、計測値は130だった。95%信頼区間は？

(2) 4回血圧を測ったら、計測値は{131, 135, 140, 138} だった。95%信頼区間は？

# 今日のまとめ

- 標本平均から母平均を推測する
  - 母平均の点推定値 = 標本平均 (不偏推定量)
  - 母平均の95%信頼区間 = [標本平均  $\pm$  1.96SE]