2016年度 森林統計学

第11回 7月4日 推定 (+第10回までのまとめ)

講義資料

高知大学農学部 森林科学コース主担当
鈴木保志

定理2(中心極限定理)

定理2. xが平均μ、標準偏差σのある分布に従うとき、大きさnの無作為標本に基づく標本平均 $\overline{x}$ は、nが無限に大きくなるとき、平均μ、標準偏差$\sigma/\sqrt{n}$ の正規分布に近づく。

- 要点: $\overline{x}$の分布は (nが大きければ正規分布に従い、)

① $\overline{\overline{x}} \fallingdotseq \mu$

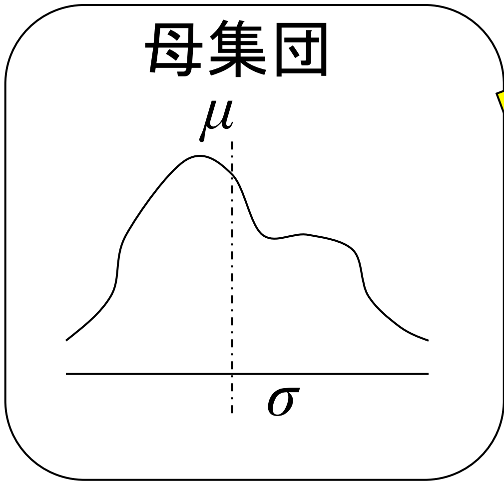
② $s_{\overline{x}} \approx \dfrac{\sigma}{\sqrt{n}}$

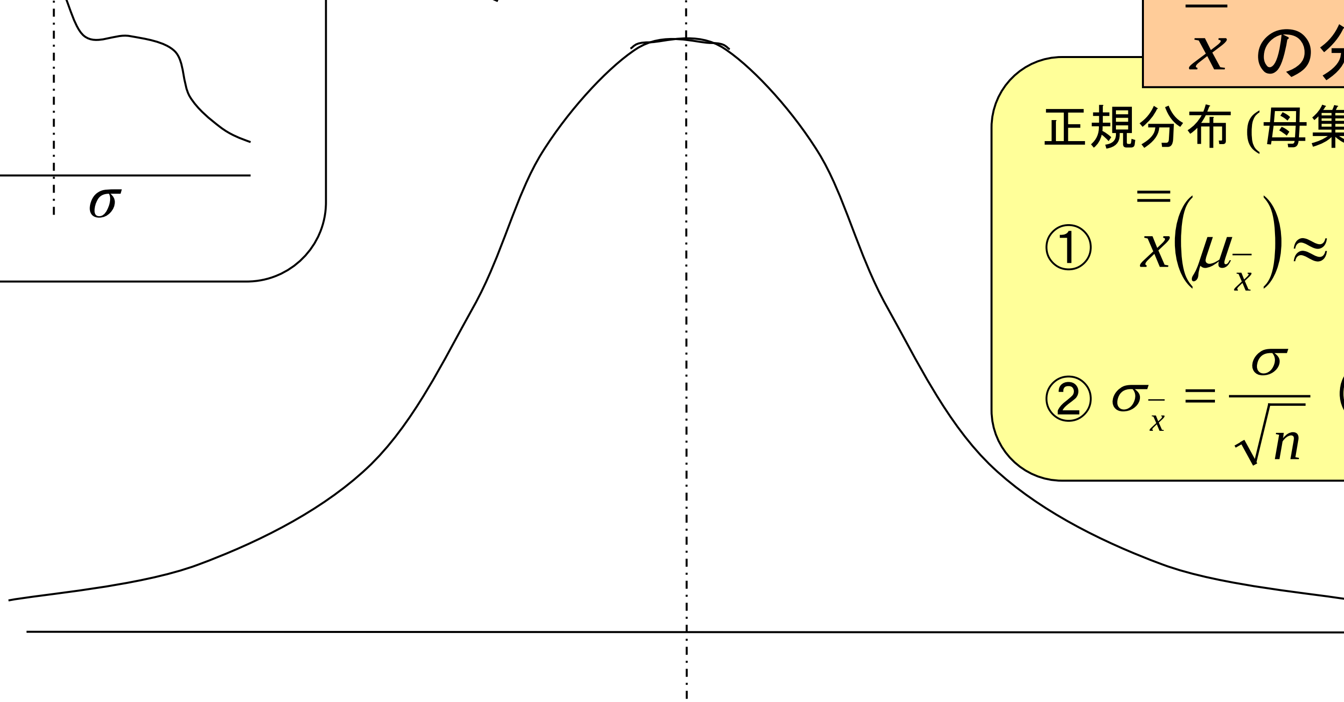
となる

- nが大きくなるほどμの推定精度が高くなる。
- 実は標本抽出の回数には関係ない(どの標本も同じ確度)

定理2 のイメージ

n 個の標本$(x_1, x_2, \ldots, x_n)$抽出 $\rightarrow$ $\bar{x}$ を算出





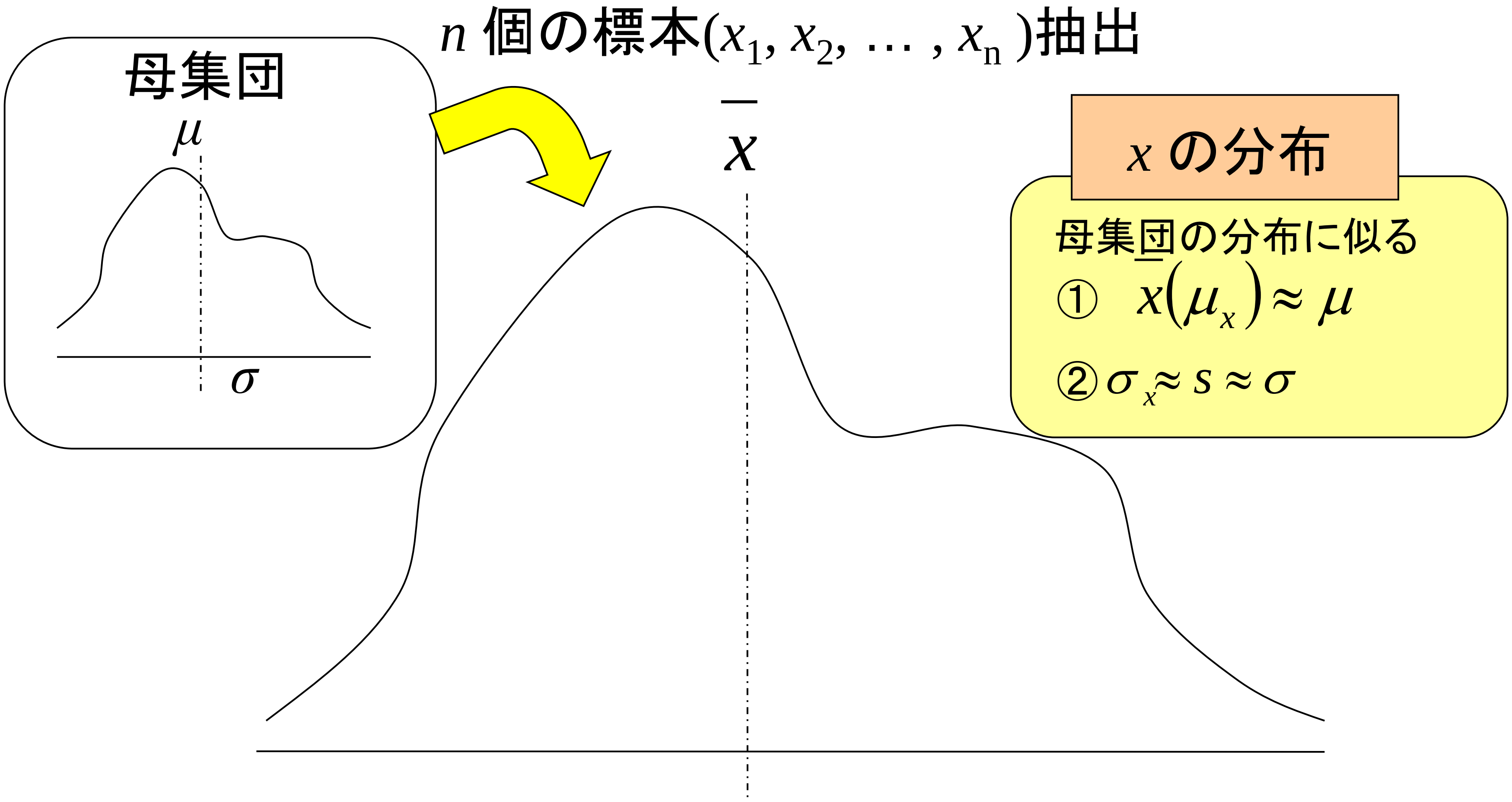
$\bar{x}$ の分布

正規分布 (母集団に関わらず)

① $\bar{\bar{x}}\left(\mu_{\bar{x}}\right) \approx \mu$

② $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$（標準誤差）

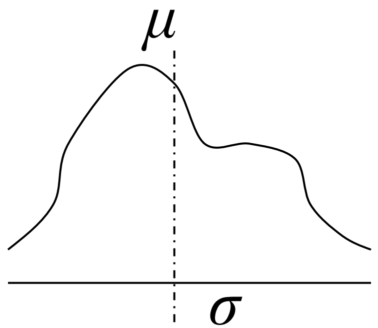
単なる x の分布だと…



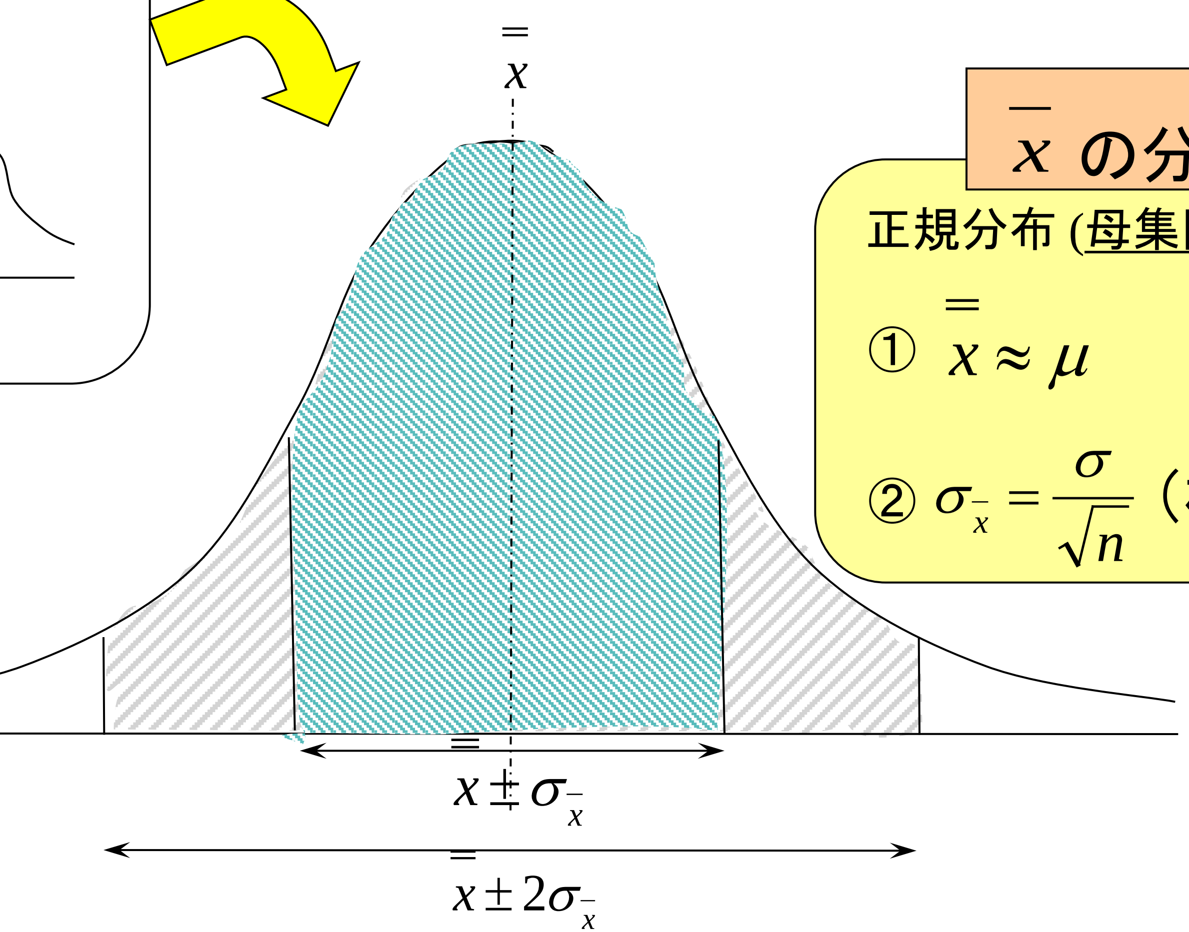
- 課題4 だと、抽出した標本で作ったヒストグラムがこれに相当する

定理2 のイメージ

母集団



n 個の標本($x_1, x_2, \ldots , x_n$)抽出→ $\bar{x}$ を算出



$\bar{x}$ の分布

正規分布 (母集団に関わらず)

① $\bar{\bar{x}} \approx \mu$

② $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$ （標準誤差）

- 正規分布なので、平均値±標準偏差に入る確率は正規分布表に従う→推定に正規分布表を利用できる

推定の要点

- 母平均の信頼区間を求める
  - 母平均の点推定値は標本平均 (信頼区間の中心)

$$\bar{x} \doteqdot \mu$$

  - 信頼区間 e は 標準誤差 $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$ の z 倍

$$\bar{x} \pm e \quad \left(e = z \cdot \sigma_{\bar{x}}\right)$$

まとめて書くと

$$\bar{x} \pm z \cdot \frac{\sigma}{\sqrt{n}}$$

    - 標準誤差 $\sigma/\sqrt{n}$ は標本平均 $\bar{x}$ の標準偏差
    - z は約2 (標準偏差の±2倍の範囲内に約9割が入る)
      - 正確には信頼確率95%とすると z = 1.96 (表IVから)

正規分布表の見方

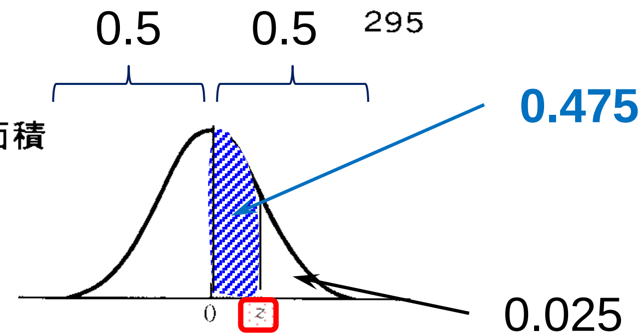


表 IV　標準正規分布の面積

表の中の数字は $z=0$ から z の正値までの曲線下の部分の面積である．z の負値に対する面積は対称性を利用して求めればよい．

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0										
0.1										
0.2										
0.3										
0.4										
0.5										
0.6										
0.7										
0.8										
0.9										
1.0										
1.1										
1.2										
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857

- 両側の棄却域が5%となるzは?
- 片側は2.5% (0.025)
- 正規分布は左右対称
- の部分は 0.5 − 0.025 = 0.475
- z は 1.96

信頼区間の推定手法

母集団: 平均μ, 標準偏差σ
標本: 標本数 n, 標本平均 $\bar{x}$, 標本標準偏差 s

「推定」手法の判断

母集団は正規分布?

Yes

定理1: $\bar{x}$は正規分布 (z; 表IV)

$$\bar{x} \fallingdotseq \mu, \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

95%信頼区間なら
$z = 1.96$

No

nは大きい?
$n \geqq 25$?

Yes

定理2: σをsで代替可(z; 表IV)

$$\bar{x} \fallingdotseq \mu, \sigma_{\bar{x}} = \frac{s}{\sqrt{n}}$$

95%信頼区間なら
$z = 1.96$

教科書
p.143の例

No

t分布: zでなくt (表V)を使う

$$\bar{x} \fallingdotseq \mu, \quad \sigma_{\bar{x}} = \frac{s}{\sqrt{n}}$$

t: 自由度$\nu = n - 1$
により異なる
(zよりは大きい)

信頼区間: $\bar{x} \pm e$

信頼幅 $e = (z$または$t) \times \sigma_{\bar{x}}$

標準誤差

z (またはt): 信頼確率により決まる

推定の要点

- 普通は、教科書p.150 例1. の(d)のパターン [推定フローチャートの一番下のパターン] が多い
  - 母平均μ・母標準偏差 σ 不明で標本数も多くない
  - σ をs (標本標準偏差) で置きかえた定理2を適用
  - z (正規分布) でなく t 分布を使う
- 母平均の点推定値は標本平均
- 信頼区間は

信頼幅 e

$$\boxed{\bar{x}} \pm \boxed{t \cdot \frac{s}{\sqrt{n}}}$$

  - t は t 分布表から求める (自由度$\nu = n - 1$)

「標本平均 $\bar{x}$ ±信頼幅　e 」という形式はこれまでと同じ

信頼幅に使う記号が z でなく t なのは正規分布でなくt分布を使うため

信頼幅 e は 分布表からの値×標準誤差 (s / $\sqrt{n}$) でやはり他の推定と同形式

例題 (p.150) で練習

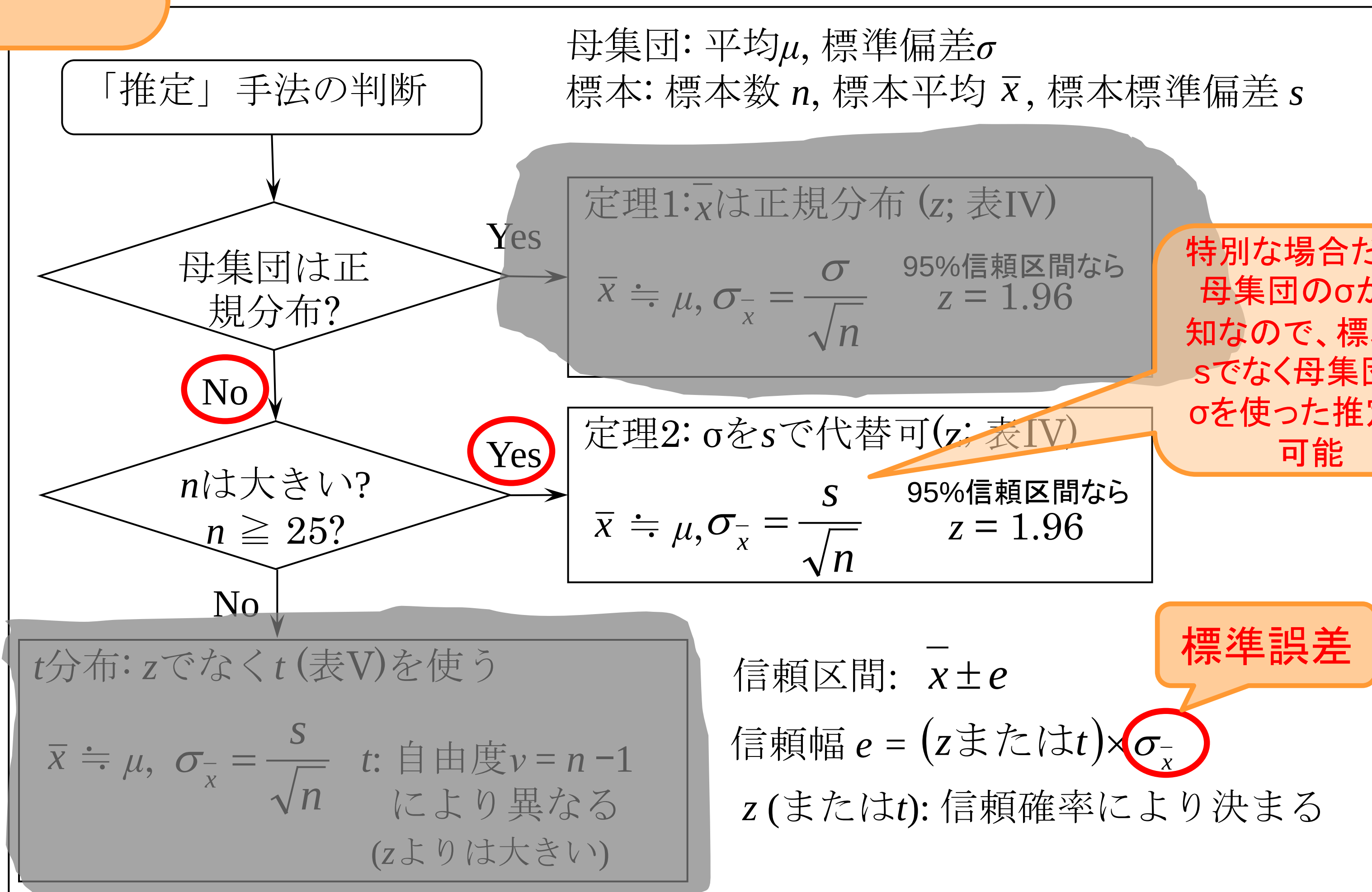
- 車の燃費
  - 母集団(これまでの車): μ = 18.0, σ =2.0
  - 標本(添加物実験): n = 25, $\bar{x}$ = 18.5, s = 2.2

(a), (c) n = 25 の標本平均の信頼区間は?

- 母集団から標本をとることを考えて区間推定
- その中に 18.5 が入っているなら添加物の効果はない
- 各自解いてみる (この間に未提出課題を回収)

信頼区間の推定手法

例題の説明から、
・母集団の分布型は不明
・ただし母集団のσは既知(=2.0)
・ n = 25



特別な場合だが、母集団のσが既知なので、標本のsでなく母集団のσを使った推定が可能

標準誤差

例題 (p.150) で練習

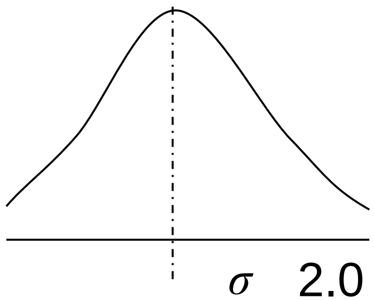
- 車の燃費
  - 母集団(これまでの車): μ = 18.0, σ =2.0
  - 標本(添加物実験): n = 25, $\bar{x}$ = 18.5, s = 2.2

(a), (c) n = 25 の標本平均の信頼区間は?

  - 母集団から標本をとることを考えて区間推定
  - その中に 18.5 が入っているなら添加物の効果はない

母集団 (正規分布)

μ 18.0

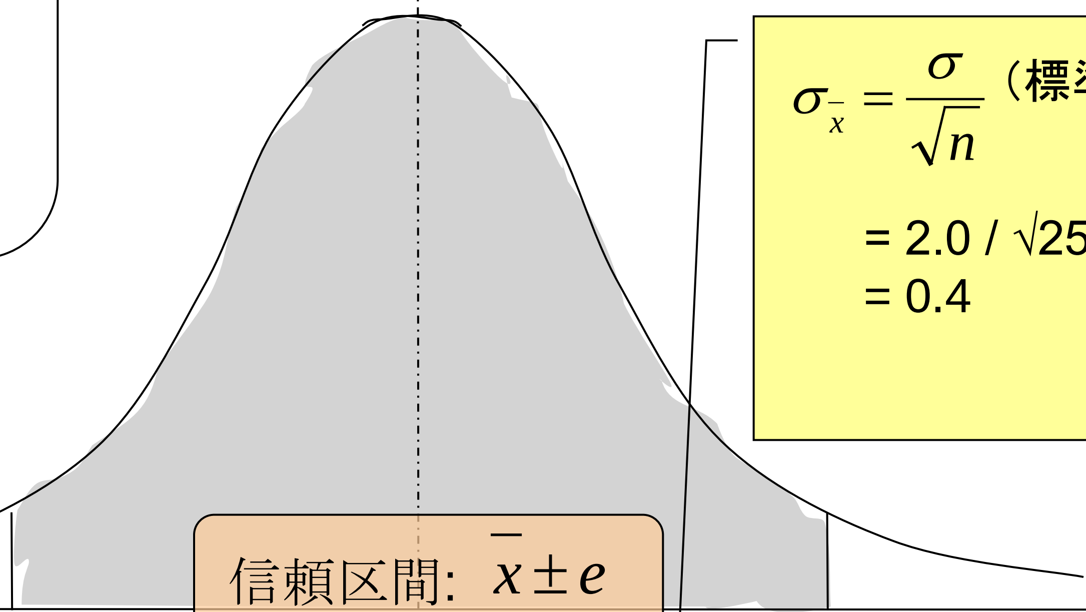


σ 2.0

標本 n = 25

$\bar{x}$

信頼区間と$\bar{x}$の分布



信頼区間: $\bar{x} \pm e$

$\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$ （標準誤差）

= 2.0 / √25

= 0.4

信頼幅 $e = (z\,\text{または}\,t) \times \sigma_{\bar{x}}$ = 1.96 ×0.4 = 0.78

<u>信頼確率</u> 0.95 → z = 1.96

95%

標本の 18.5 はこの範囲内

∴添加物の効果がなくても、よくあることといえる

例題 (p.150) で練習

- 車の燃費
  - 母集団(これまでの車): μ = 18.0, σ=2.0
  - 標本(添加物実験): n = 25, $\bar{x}$ = 18.5, s = 2.2

(b) 信頼幅が0.5 (信頼区間が1.0) より狭くなるようにするには、n をいくつにすればよいか?

- 現状(n = 25) の信頼幅は0.78; もっと精度よく、添加物実験車の母平均(μ')を推定したい

(b) 信頼幅 $e < 0.5$ となる n？

- 信頼幅 e が 0.5 以下になるためには…

$$e = z \cdot \sigma_{\bar{x}} = z \cdot \frac{\sigma}{\sqrt{n}} \qquad e < z \cdot \frac{\sigma}{\sqrt{n}}$$

- e (0.5), σ(2.0), z (1.96) はわかっていて、n が不明

$$n > \left(\frac{z \cdot \sigma}{e} \right)^2 \qquad n > \left(\frac{1.96 \cdot 2.0}{0.5} \right)^2 = 61.5$$

- n を 62以上にすればよい
  - 今 n = 25 なので、標本を37個追加すればよい

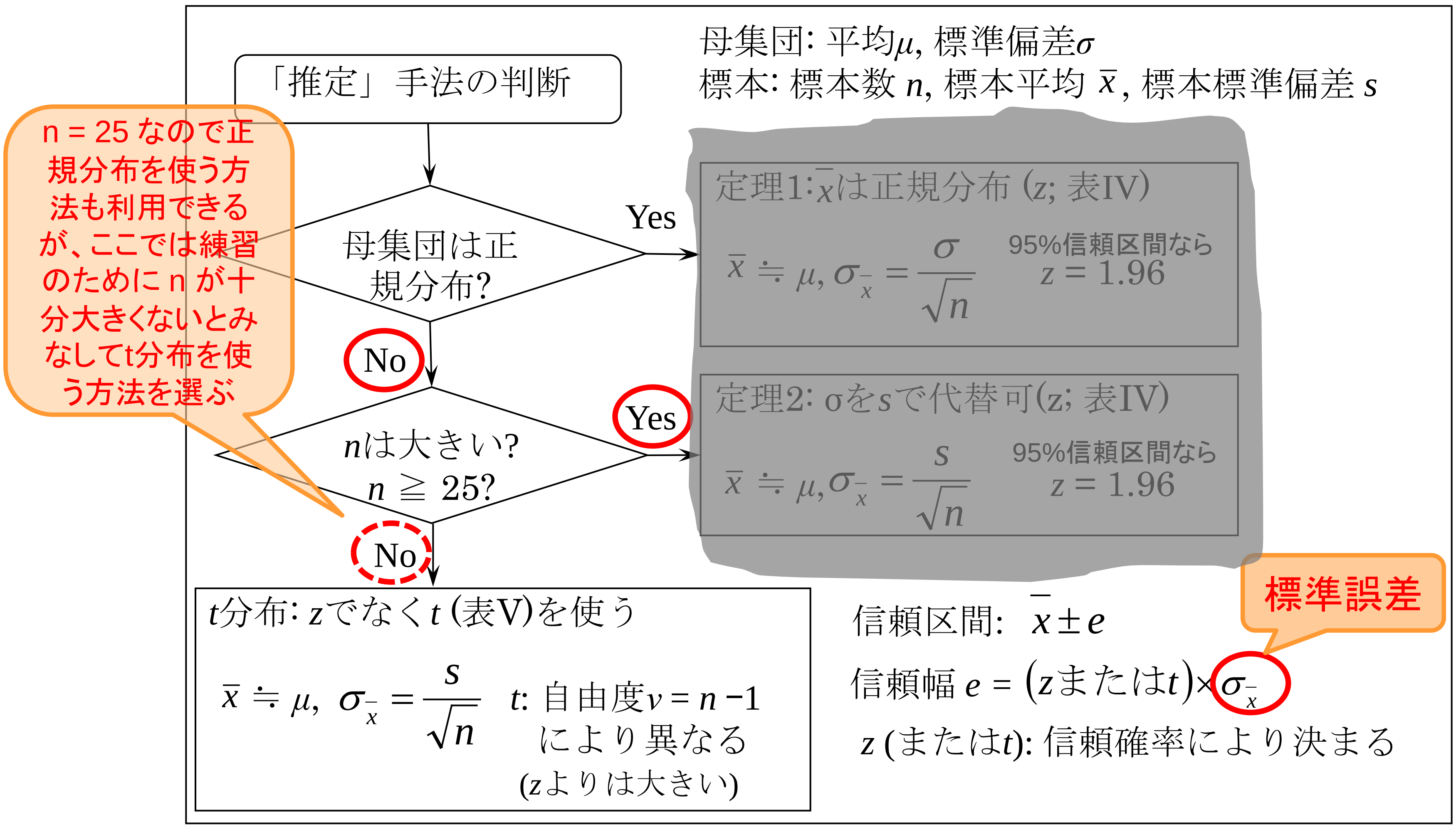
例題 (p.150) で練習

- 車の燃費
  - 母集団(これまでの車): μ = 18.0, σ=2.0
  - 標本(添加物実験): n = 25, $\overline{x}$ = 18.5, s = 2.2

(d) 添加物実験車の母平均の信頼区間は?

- 添加物車はこれまでの車と違うとして、母集団の値は知らないものとして実験車の n, $\overline{x}$, s からμ'を推定してみる
  - その区間にこれまでの車のμがなければ、μとμ'は異なるということになる
  - n = 25 だが小標本法として t 分布を使用する

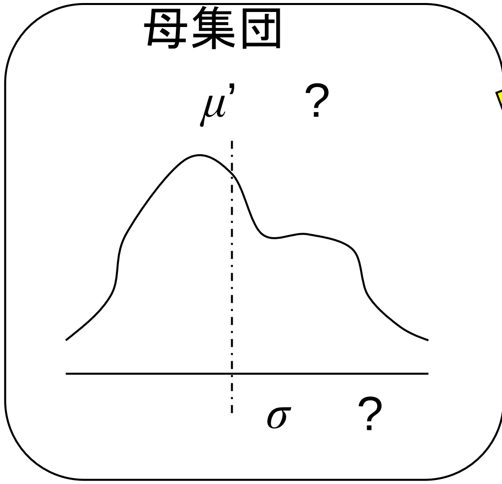
信頼区間の推定手法


母集団: 平均μ, 標準偏差σ
標本: 標本数 n, 標本平均 $\bar{x}$, 標本標準偏差 s
「推定」手法の判断
母集団は正規分布?
Yes
No
nは大きい? $n \geqq 25$?
Yes
No
定理1: $\bar{x}$は正規分布 (z; 表IV)
$\bar{x} \fallingdotseq \mu, \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$
95%信頼区間なら $z = 1.96$
定理2: σをsで代替可(z; 表IV)
$\bar{x} \fallingdotseq \mu, \sigma_{\bar{x}} = \frac{s}{\sqrt{n}}$
95%信頼区間なら $z = 1.96$
t分布: zでなくt (表V)を使う
$\bar{x} \fallingdotseq \mu, \ \sigma_{\bar{x}} = \frac{s}{\sqrt{n}}$
t: 自由度$\nu = n - 1$ により異なる
(zよりは大きい)
n = 25 なので正規分布を使う方法も利用できるが、ここでは練習のために n が十分大きくないとみなしてt分布を使う方法を選ぶ
信頼区間: $\bar{x} \pm e$
信頼幅 $e = (z$または$t) \times \sigma_{\bar{x}}$
標準誤差
z (またはt): 信頼確率により決まる

例題 (p.150) で練習

- 車の燃費
  - 母集団(これまでの車): μ = 18.0, σ=2.0
  - 標本(添加物実験): n = 25, $\bar{x}$ = 18.5, s = 2.2

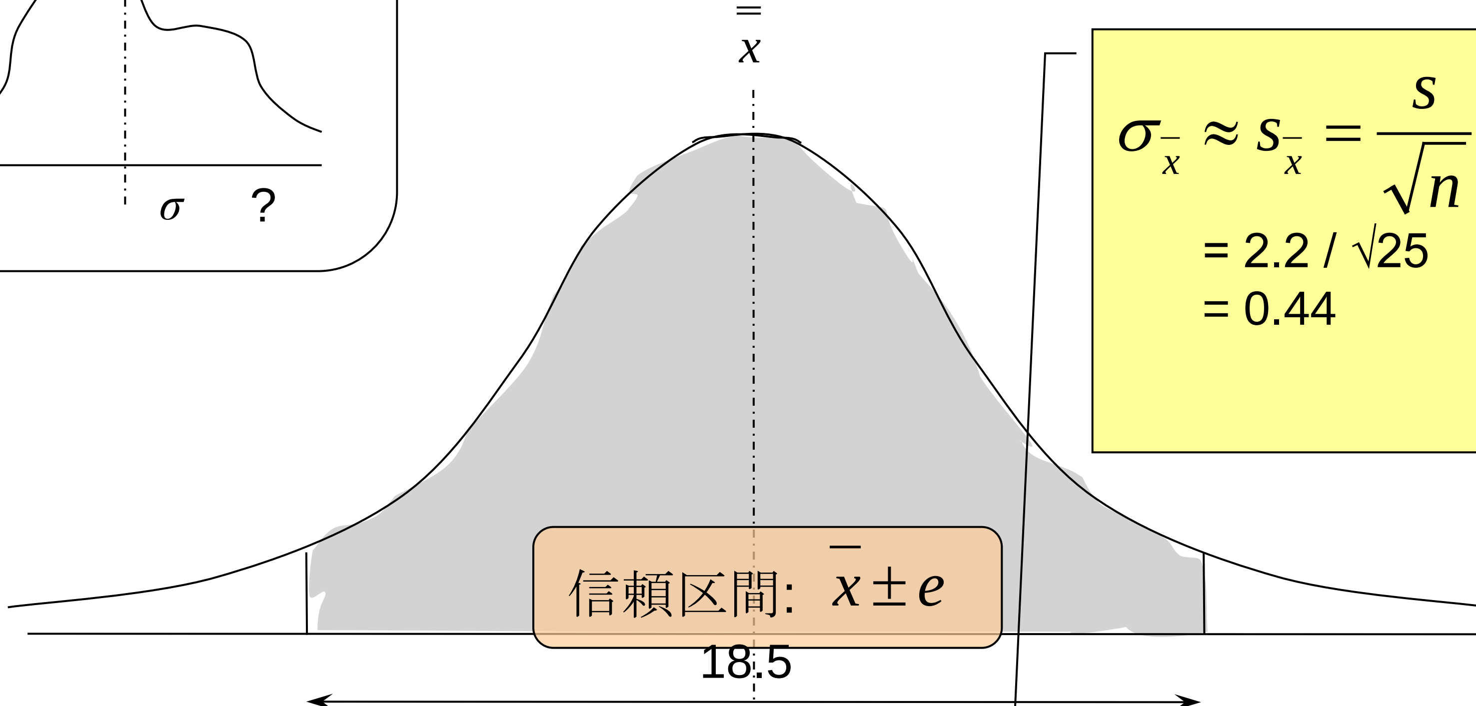
(d) 添加物実験車の母平均の信頼区間は?
  - 添加物車はこれまでの車と違うとして、母集団の値は知らないものとして実験車の n, $\bar{x}$, s からμ'を推定してみる
    - その区間にこれまでの車のμがなければ、μとμ'は異なるということになる
    - <u>n = 25 だが小標本法として t 分布を使用する</u>



標本 $n = 25,\ \bar{x} = 18.5,\ s = 2.2$

小標本法 (t分布を使用)

$$\sigma_{\bar{x}} \approx s_{\bar{x}} = \frac{s}{\sqrt{n}}$$
$= 2.2 / \sqrt{25}$
$= 0.44$



信頼幅 $e = (z\text{または}t) \times \sigma_{\bar{x}}$

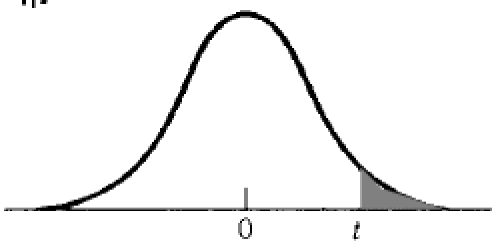
信頼確率 0.95 $\rightarrow$ $t = 2.064$

自由度 $\nu = 25 - 1 = 24$

危険率 P = (1 – 信頼率) / 2
信頼率: 0.95

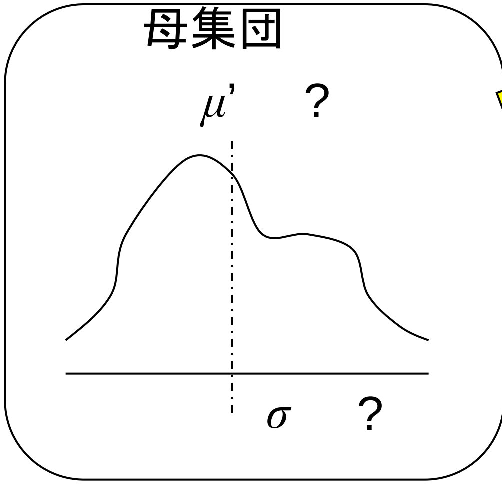
表 V スチューデントの t 分布

表側の数字は自由度 (ν) を表わす．表頭の数字は t が表の中の数値を超える確率 (P) を表わす．負の t の値に対しては分布の対称性を利用すればよい．



ν \ P	.10	.05	.025	.01	.005
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
⋮	⋮	⋮	⋮	⋮	⋮
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771

自由度
ν(ニュー)
は $n-1$



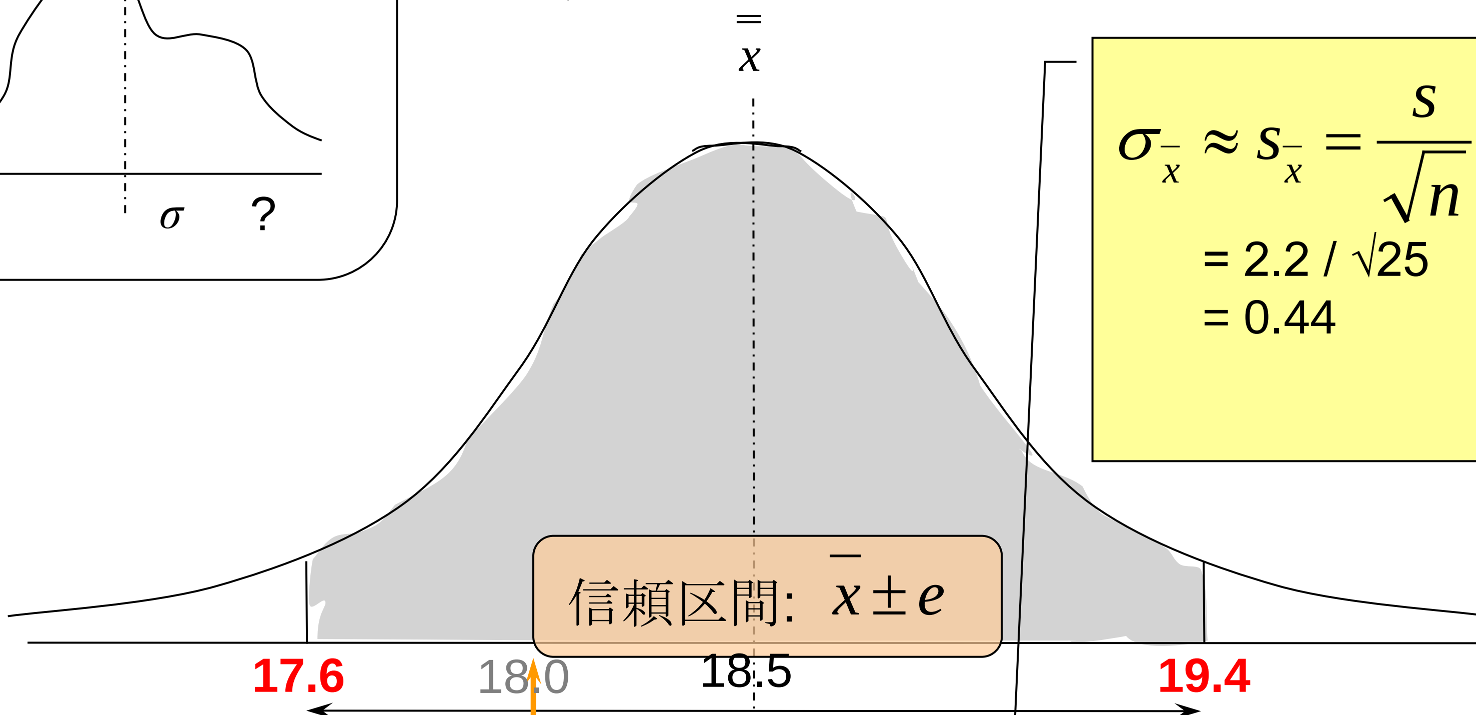
標本 $n = 25,\ \bar{x} = 18.5,\ s = 2.2$

小標本法 (t分布を使用)

$$\sigma_{\bar{x}} \approx s_{\bar{x}} = \frac{s}{\sqrt{n}}$$

= 2.2 / √25

= 0.44



信頼区間: $\bar{x} \pm e$

$$\text{信頼幅 } e = (z\text{または}t) \times \sigma_{\bar{x}}$$ = 2.06 × 0.44 = 0.90

信頼確率 0.95 → t = 2.0639

自由度 ν = 25 – 1 = 24

無添加母集団のμ= 18.0 はこの範囲内

∴添加物の効果がなくても、よくあることといえる

第7章の問題の推奨問題

- 2節(母平均 μ の推定): 1., 3., 4.
- 3節(大標本法 – σ の不偏推定値sの利用): 9., 12., 13.
- 5節(小標本法 – t 分布): 25, 26, 28., 29.
  - 問題25: 小標本法による推定(標準的問題)
  - 問題26: 大標本法と小標本法の比較
    - 小標本法(t 分布)は精度が低くなる(信頼区間の幅が広くなる)ので、可能ならばnを大きくして大標本法(正規分布)が使えるようにしたほうがよい
- 一般問題: 30.

［課題4］のコメント

- 標本抽出の方法、$\bar{x}$ や s の計算はおおむねできていた
  - が s の計算ができていない人もいる: 要注意
    - $\bar{x}$ の標準偏差 $s_{\bar{x}}$ と s の平均 $\bar{s}$ の混同もあった
- 母集団の集計、ヒストグラムの作成などの復習もおおむねOK
- 適切なケタどりに注意
  - 元データの計測精度 m に対し、$\bar{x}$ や s は $m/10$ 程度が目安
- 考察して欲しかった要点は…

[課題4]の考察の要点

- 考察 (1)：「$\bar{x}$ の分布は定理1･2のようになっているか?」
  - $\bar{\bar{x}} \fallingdotseq \mu$ ($\bar{x}$は 母平均 μ の不偏推定値)
  - $\bar{x}$ のバラツキ($s_{\bar{x}}$)は n が大きい方が小さい
    - → その程度は「定理2」に従う（$s_{\bar{x}} \fallingdotseq \dfrac{\sigma}{\sqrt{n}}$）
- 考察(2)：「s はσの不偏推定値になっているか?」
  - 個々の s はバラツキはあっても σ くらいの値
    - なので σ を s で置き換えることができる

もういちど自分のデータで確認してみよう

各自課題4のデータで復習

- いずれかの標本を選んで、母平均を推定してみる(信頼区間を求めてみる)
  - 標本平均 $\bar{x}$ 、標本標準偏差 s、標本数 n
  - 母標準偏差 σ は不明だが、s はσの不偏推定値なので σ の代わりに s を使う
    - 1. **<u>小標本法</u>**: 信頼確率95%として、t分布表(教科書表V)から所定のt値を読み取る (P = 0.025, $\nu = n - 1$)

$$\bar{x} \pm t \cdot \frac{\sigma}{\sqrt{n}} \qquad \left(\bar{x} \pm t \cdot \frac{s}{\sqrt{n}} \right)$$

    - 2.簡便法: t の代わりに 2 を用いる
  - この範囲に、母平均 μ は入っているか?