



高知工科大学 経済・マネジメント学群

統計学 2

9. t 分布と母平均の推定

やない ゆう き
矢内 勇生



<https://yukiyanai.github.io>



yanai.yuki@kochi-tech.ac.jp



このトピックの目標

- 標本平均から母平均を推定する方法を理解する
 - ▶ もうやったのでは？？？
- t 分布を理解しよう！

母平均の推定：母分散が既知の場合（復習）

- 母平均の点推定：標本平均 = 点推定値
- 母平均の区間推定：
 - 標本平均の分布が正規分布に従うと考え、95%信頼区間を求めた
 - μ_x の95%信頼区間

$$\left[\bar{x} - 1.96 \frac{\sigma}{\sqrt{N}}, \bar{x} + 1.96 \frac{\sigma}{\sqrt{N}} \right] = [\bar{x} - 1.96 \cdot \text{SE}, \bar{x} + 1.96 \cdot \text{SE}]$$

- 母分散 σ^2 （あるいは母標準偏差 σ ）が既知と仮定

t 分布と区間推定

母分散が未知だったら？

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

- 母分散を知らない → 上の式が統計量にならない（標本から数値を得ることができないから）
- どうする？
 - ▶ 区間推定の方法を変える
 - ただし、点推定値は変わらない

母分散を知らない (かつ N が十分大きくない) ときの区間推定

- σ を知らないとき : σ の推定値として u を使う

▶ $\frac{\bar{x} - \mu_x}{\frac{u}{\sqrt{N}}}$ は標準正規分布に従わない

- ▶ 標準正規分布を使って求める信頼区間は使えない
 - 標準正規分布の代わりに **t 分布** を利用する

分散 (variance)

- 分散：データのばらつき（散らばり具合）を表す統計量
- 0以上の値をとる（データの値がすべて等しいとき0）
- データのばらつきが大きいほど、分散も大きくなる

分散の求め方（復習1）

- ・ 標本分散を表す記号： s^2 (σ^2 は母分散)

$$s^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2$$

標本分散の求め方（復習2）

1. 標本平均を求める
2. 偏差（= 観測値 - 標本平均）を各観測値について求める
3. それぞれの偏差を2乗する
4. 偏差の2乗をすべて足して、標本サイズ（ N ）で割る
（偏差²の平均値を求める）

標本分散の求め方（復習3）

- 右のデータの分散を求めてみる

$$s^2 = \frac{16 + 1 + 0 + 4 + 9}{5}$$
$$= \frac{30}{5} = 6$$

	x	偏差	偏差 ²
	1	-4	16
	4	-1	1
	5	0	0
	7	2	4
	8	3	9
計	25	0	30

標本分散の偏り

- 標本をいくつも抽出し、それぞれの標本について標本分散を求める：

標本分散の平均値 $<$ 母分散

➡ 標本分散を母分散の推定に使うと、（小さい方に）
偏ってしまう

★ 偏りが無い分散の求め方は？

不偏分散 (unbiased variance)

- 母分散の偏りのない推定値として使える
- 記号： u^2 で表す
- 分散を求める式の名分を少しだけ小さくする

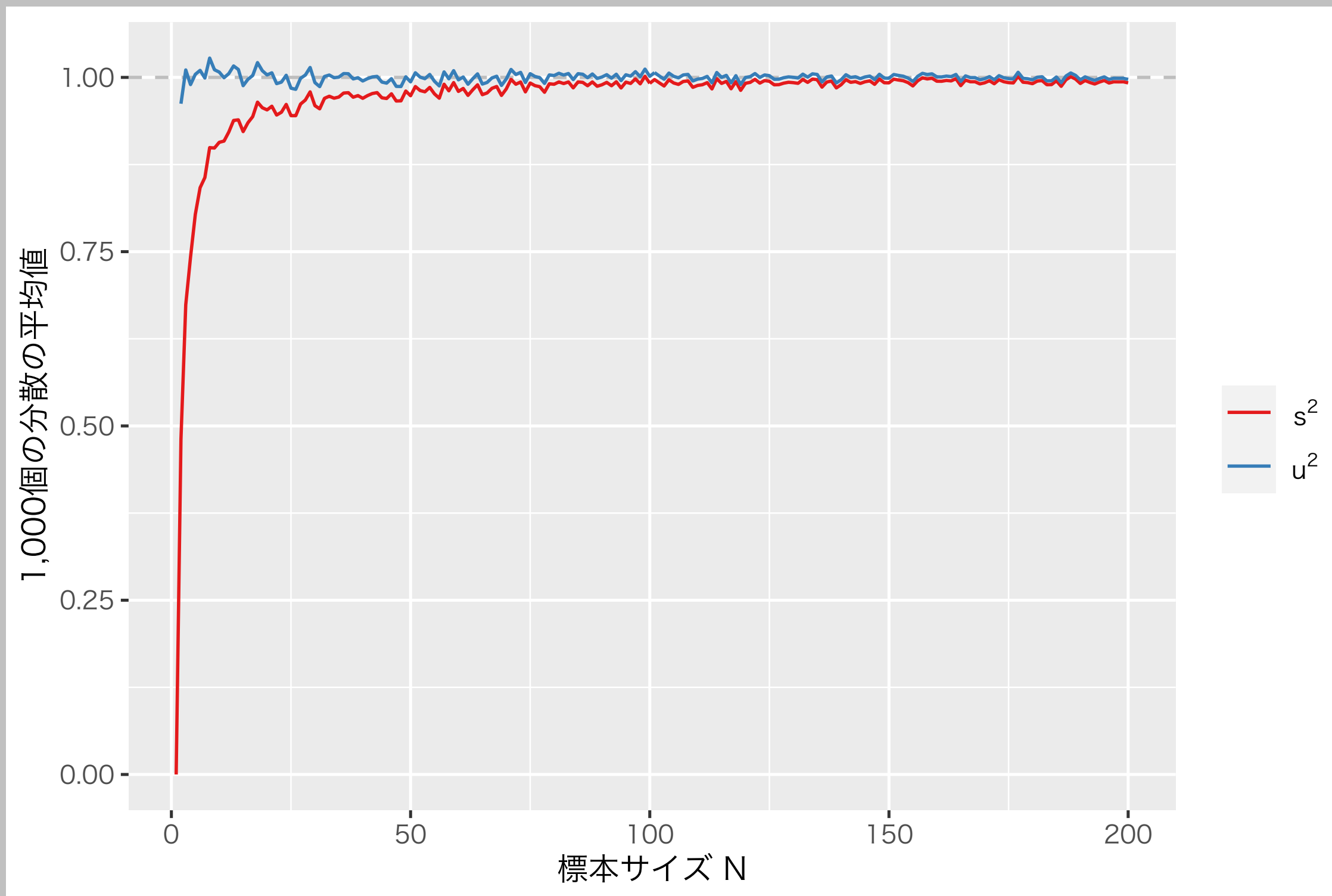
$$u^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2$$

不偏分散（続）

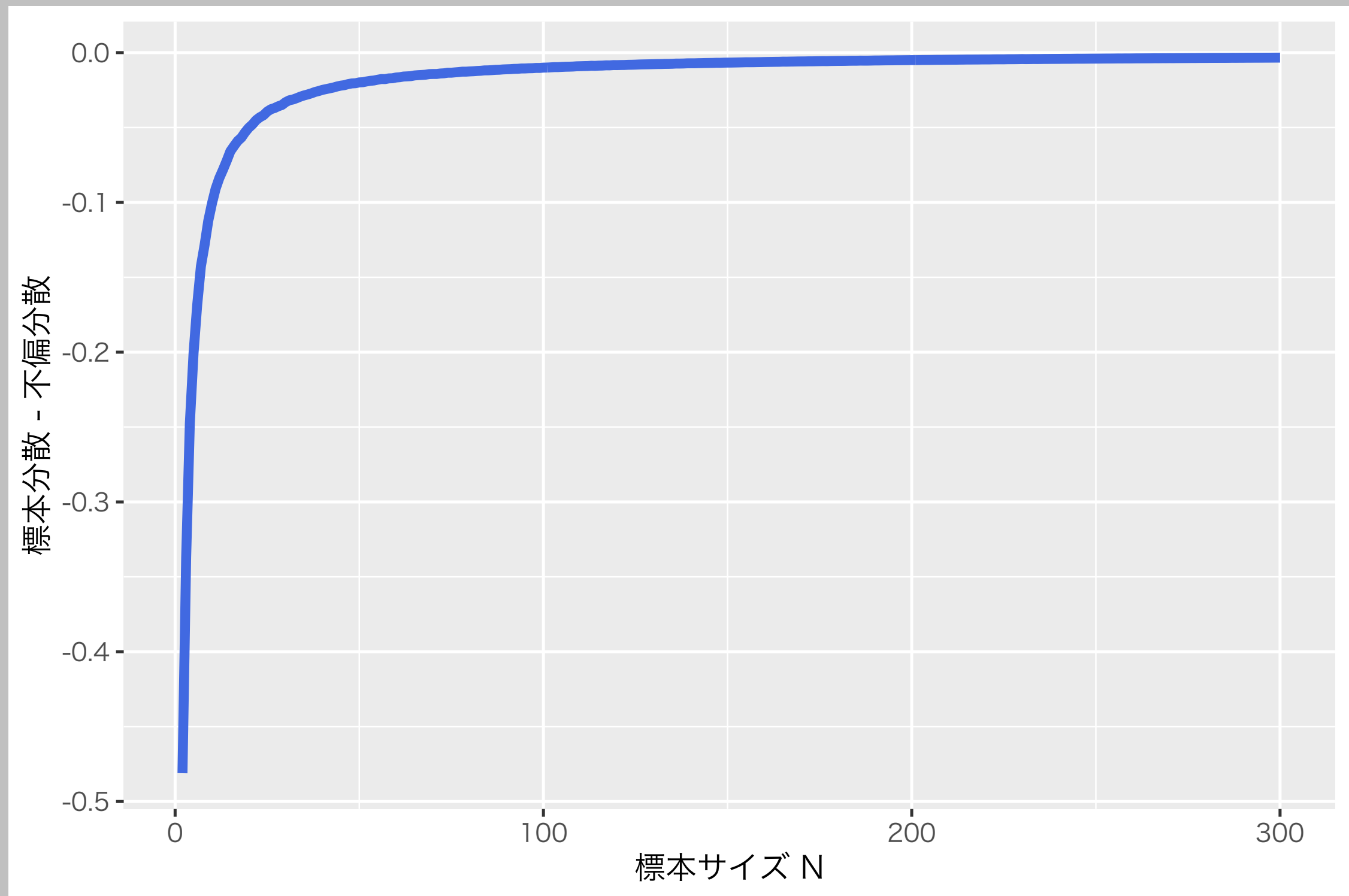
- 不偏分散：母分散の点推定値
 - ▶ Rでは `var()` で求める
- 標準偏差の点推定値は： u （不偏分散の平方根）
 - ▶ N が十分大きいときは、 N で割ってもよい（結果がほとんど同じなので）
 - ▶ 母平均を知っているときは、 N で割って分散を求める

標本分散と不偏分散の平均値

標準正規分布 から各 N について1000個の標本を抽出



標本分散と不偏分散の差



母分散を知らない（かつ N が十分大きくない）ときの区間推定

- σ を知らないとき： σ の推定値として u を使う

▶ $\frac{\bar{x} - \mu_x}{\frac{u}{\sqrt{N}}}$ は標準正規分布に従わない

- ▶ 標準正規分布を使って求める信頼区間は使えない
 - 標準正規分布の代わりに **t 分布** を利用する

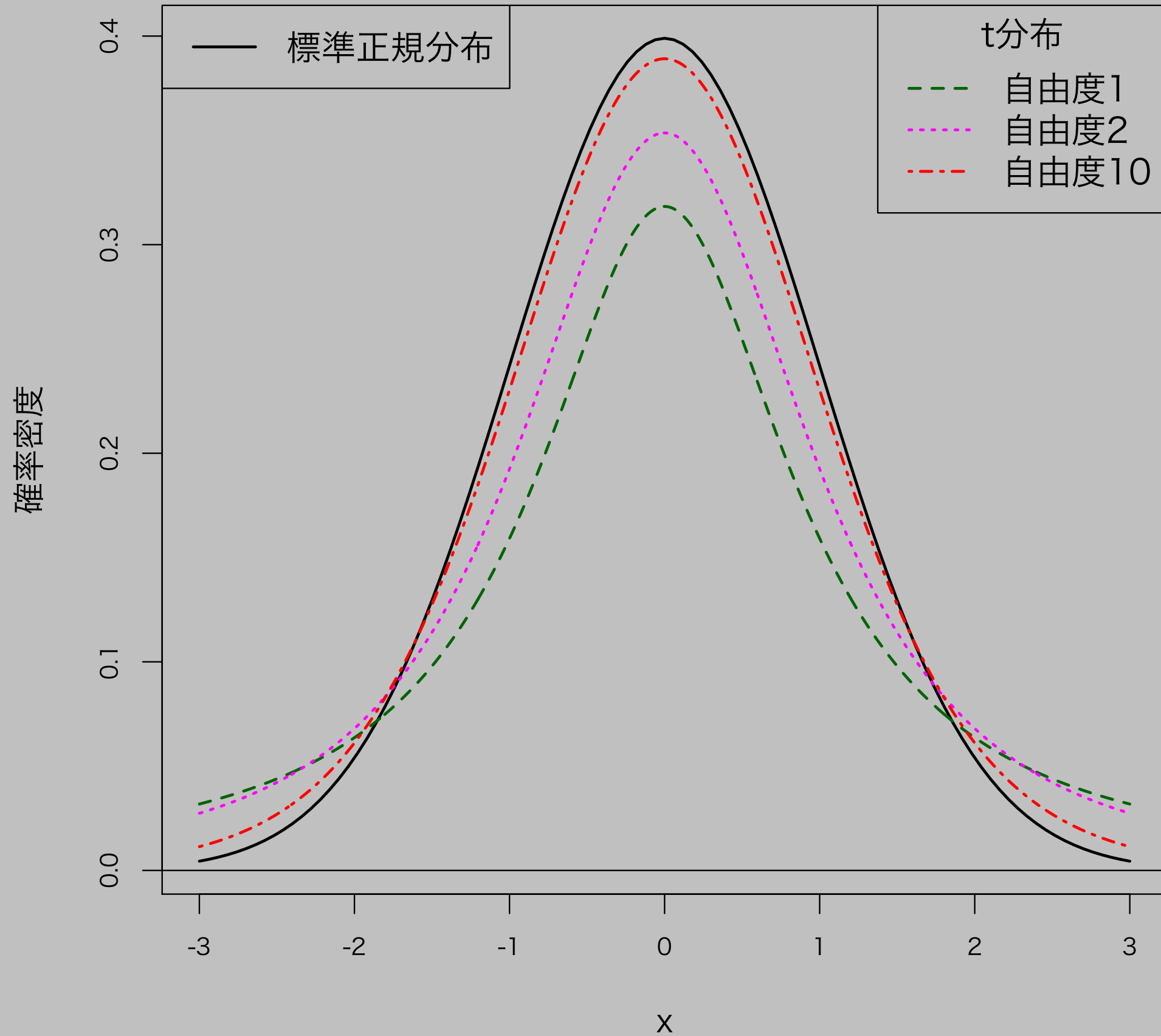
t 分布 (Student's t distribution)

- スチューデント (Student) の t 分布
- 確率密度関数：

$$f(x | k) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\sqrt{k\pi}\Gamma\left(\frac{k}{2}\right)} \left(1 + \frac{x^2}{k}\right)^{-\frac{k+1}{2}}$$

- ▶ ただし、 k は x の自由度、 $\Gamma(\cdot)$ はガンマ関数

標準正規分布とt分布



t 分布の特徴

- 自由度によって形が決まる
- 概形は標準正規分布に似ている
 - 分布の中心は0
 - 標準正規分布より山の頂上が低い
 - 標準正規分布より裾が厚い
- 自由度が大きくなるにつれ、標準正規分布に近づく

ウィリアム・ゴセット (William Sealy Gosset : 1876-1937)

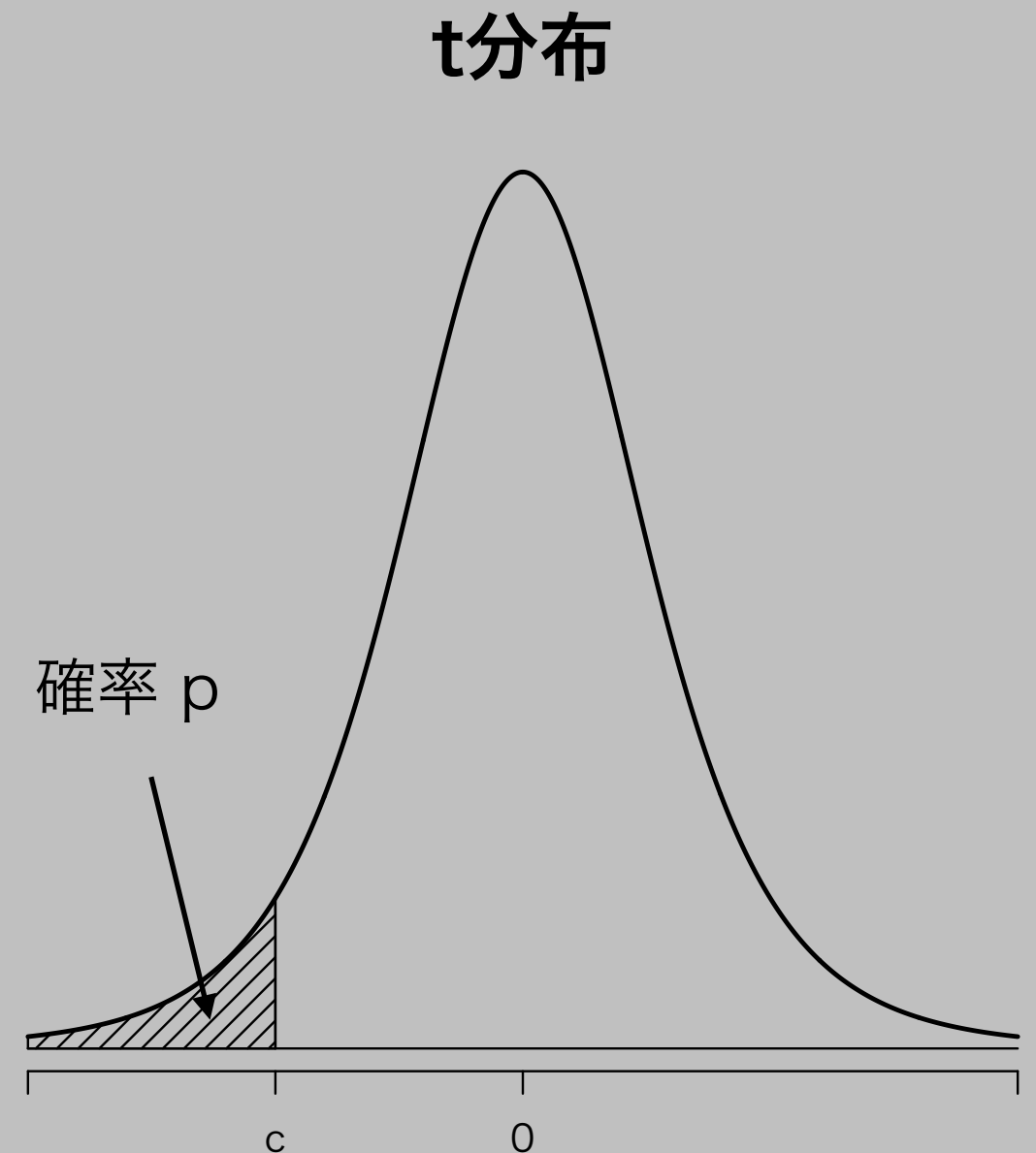
- イギリスの統計学者・醸造技術者（ギネス社に勤務）
- 推測統計学の確立に貢献
- ペンネーム：Student
- t 分布を発見した

t 分布の使い方

- ・ 自分が求めたい確率 p と自由度 df の組み合わせを決める
- ・ Rで c を求める

$qt(p, df)$

- ・ t 分布は左右対称なので、片側（負の側、左側）の c がわかれば、反対側（正の側、右側） $-c$ もわかる



標本平均と t 分布

- $\frac{\bar{x} - \mu_x}{\frac{u}{\sqrt{N}}}$ は、自由度 $N - 1$ の t 分布に従う
- 標本サイズ $N = 10$ のとき
 - ▶ 自由度は9 : $p = 0.025$ とすると、 $c = -2.2622$
 - ▶ 標本の95%について

$$-2.2622 \leq \frac{\bar{x} - \mu_x}{\frac{u}{\sqrt{N}}} \leq 2.2622$$

一般的な場合

- 一般的に、自由度 k と 確率 p によって決まる t 分布の $-c$ を $t_{k,p}$ と書くことにすると、標本の $100(1 - 2p)\%$ について、

$$-t_{N-1,p} \leq \frac{\bar{x} - \mu_x}{\frac{u}{\sqrt{N}}} \leq t_{N-1,p}$$

となる

- $p = 0.025$ なら、 $(1 - 2 \cdot 0.025) = 0.95$ (つまり95%) について、

$$-t_{N-1,0.025} \leq \frac{\bar{x} - \mu_x}{\frac{u}{\sqrt{N}}} \leq t_{N-1,0.025}$$

母平均の95%信頼区間

$$-t_{N-1,0.025} \leq \frac{\bar{x} - \mu_x}{\frac{u}{\sqrt{N}}} \leq t_{N-1,0.025}$$

- 上の式のうち、標本平均 \bar{x} 、不偏分散の平方根 u 、標本サイズ N は知っていて、母平均 μ_x を推定したい
- ▶ 上の不等式を μ_x について解けばよい

$$\bar{x} - t_{N-1,0.025} \frac{u}{\sqrt{N}} \leq \mu_x \leq \bar{x} + t_{N-1,0.025} \frac{u}{\sqrt{N}}$$

このトピックのまとめ

- 母平均の点推定値は標本平均
- 母平均の信頼区間の求め方は、状況によって異なる
 - ▶ 母分散が既知 ~~(または標本サイズ N が十分大きいとき)~~
 - 標準正規分布を使った信頼区間
 - ▶ 母分散が未知
 - t 分布を使った信頼区間
- 実習：
 - <https://yukiyanai.github.io/jp/classes/stat2/contents/R/t-distribution.html>

次回予告

10. 2つの平均値を比較する