高知工科大学 経済・マネジメント学群

統計学 2

7. 統計的検定と仮説検定の基礎

やない ゆうき
矢内 勇生

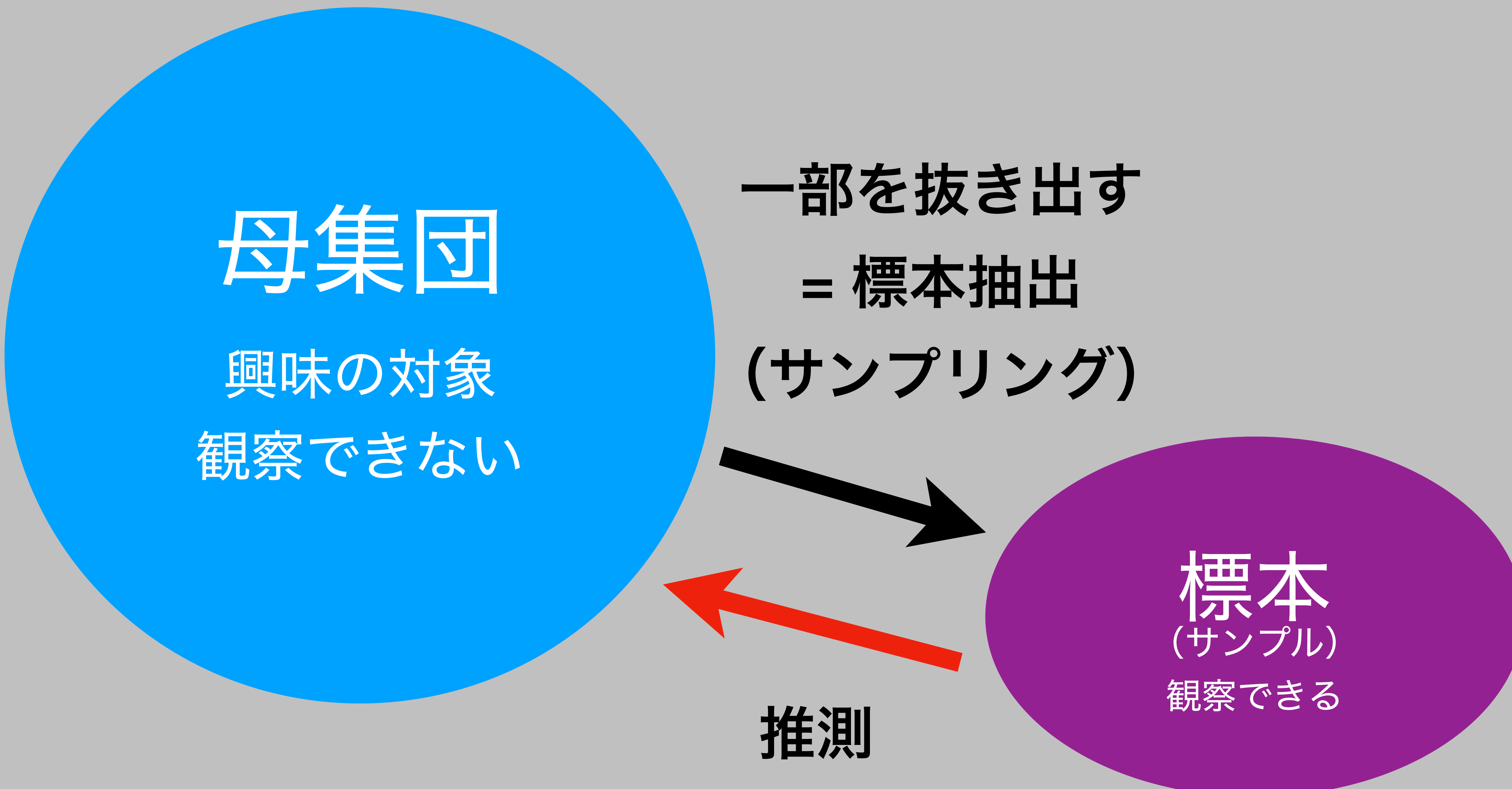
https://yukiyanai.github.io

yanai.yuki@kochi-tech.ac.jp

このトピックの目標

- 推測統計学の目的を理解する
  - 統計的検定とは？
  - 統計的推定とは？
- 母集団と標本の違いを理解する

母集団と標本



部分から全体を知る

- 通常、手に入るデータは「全体の一部」
  - 例）日本国民（母集団）が消費税増税に賛成かどうか知りたい
  - 2,000人（日本国民の一部）に賛成か反対か尋ねる
- 部分から得られる情報を使って全体（母集団）について考える
  - 例）2,000人の回答から日本人全体の賛否を推測する

★ **統計的推定** (statistical inference)

統計的検定の基礎

統計的検定の基礎

- 「公正な」（表が出る確率 $\theta = 0.5$ の）コインを N 回投げる
- 表が10回出た
  - 投げた回数 N はいくつか？
    - 仮説1：$N = 16$
    - 仮説2：$N = 36$

統計的検定の目的

- 標本から得られる情報を利用し、仮説（hypothesis）が正しいか正しくないか判断する

- 仮説が「正しくない」という証拠がない → 仮説を（とりあえず）保留にする
  - 「証拠がないこと」は「ないことの証拠」ではないので注意

- 仮説が「正しくない」と考える根拠がある → 仮説を棄却する

可能な仮説はたくさんある

N 枚の正しいコイン投げの例

- 10以上の整数であれば、仮説として成り立つ
- 問題は、それが妥当かどうか
  - 極端な例

仮説3：$N = 10$

仮説4：$N = 10000$

-統計的方法を使うまでもなく、妥当ではなさそう

どこまでが妥当か？

N 枚の公正なコイン投げの例

- 表が出る確率が 0.5 で、表が10枚出ているのだから、$N = 20$ と予測するのが最も妥当

  - $N = 19$ や $N = 21$ もそれほど悪くない仮説では
  - $N = 18$ や $N = 22$ もそれほど悪くない仮説では？
  - ・・・

★どこまでが妥当？ $\rightarrow$ **統計的検定**で決める

正規分布の性質を利用した統計的検定

- 正規分布では、平均 ± 2標準偏差の範囲にデータの95% が含まれる
  - より正確には**平均 ± 1.96標準偏差 (sd)**
  - この区間を検定に利用する！

統計的検定の方法

- **ある仮説が正しいと仮定して**、平均 ±1.96sd の区間に観測されたデータが含まれるかどうか確かめる
  - 含まれる → データが95%の一部、すなわち「ありがちな値」なので、仮説は「妥当でないとはいえない」→ 仮説を棄却せず保留する
  - 含まれない → 5%しか起こらないはずの値をデータとして観測してしまった → 「起こりにくい」はずのデータが現に手元にある → 仮定がおかしいのでは？→ 仮説を棄却する

N 回コイン投げの仮説検定

- 表が 0.5 ($\theta = 0.5$) の確率で出るコインをN 回投げ、10 回表が出た
- 仮説1：$N = 16$
- 仮説2：$N = 36$

✤ コイン投げをN 回行う → 二項分布

- 二項分布の平均 = $N\theta$
- 二項分布の分散 = $N\theta(1 - \theta)$

仮説1の検証

仮説1（$N = 16$）が正しいとすると、

- 平均 $= N\theta = 16 \cdot 0.5 = 8$

- 分散 $= N\theta(1 - \theta) = 16 \cdot 0.5 \cdot (1 - 0.5) = 4$

- 標準偏差 $= \sqrt{\text{分散}} = 2$

  - 平均$- 1.96 \cdot \mathrm{sd} = 8 - 1.96 \cdot 2 = 4.08$

  - 平均$+ 1.96 \cdot \mathrm{sd} = 8 + 1.96 \cdot 2 = 11.92$

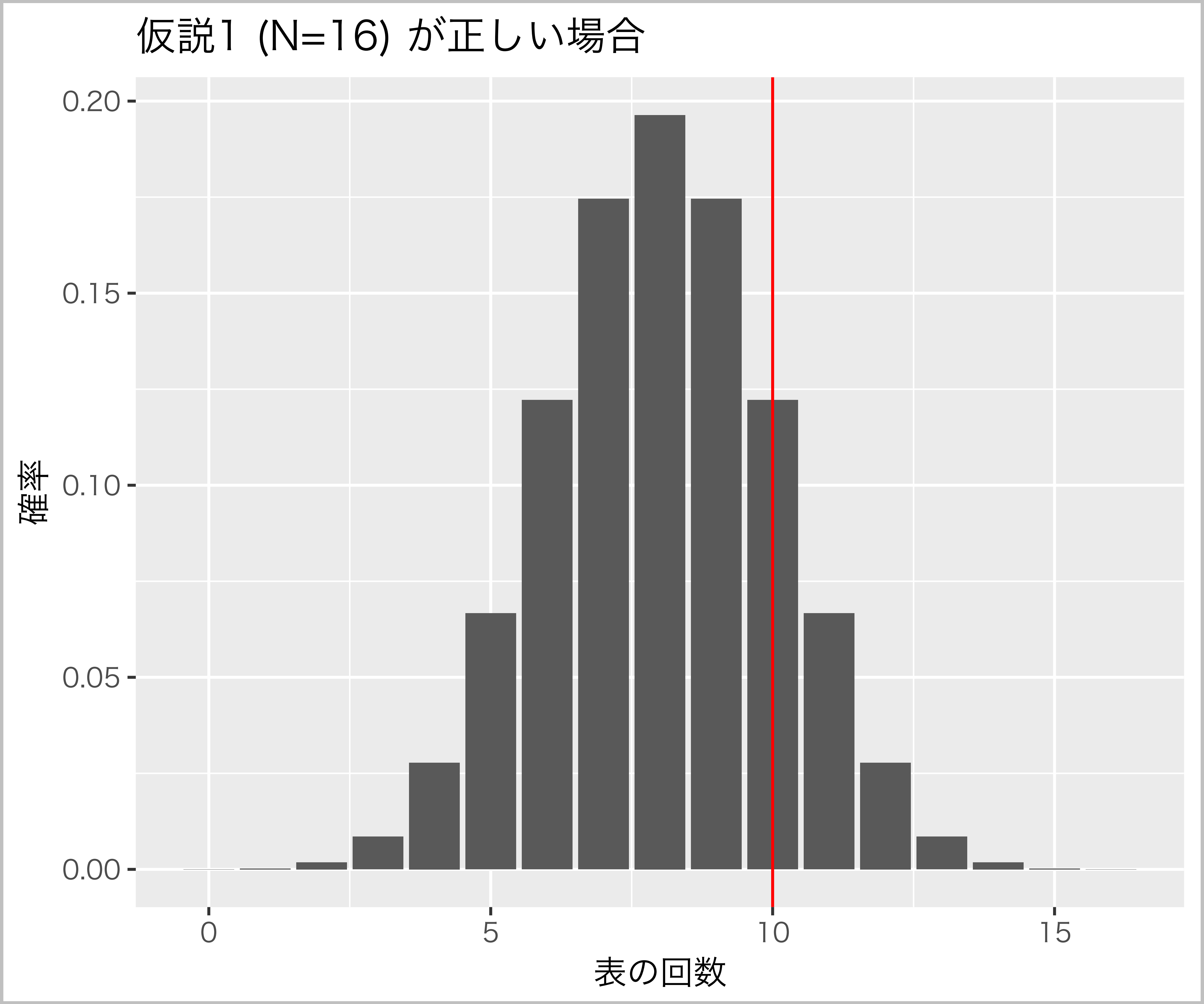
仮説1の検証（続）

仮説1（$N = 16$）が正しいとすると、

- データの95%は4.08 と11.92 の間の値をとる
- 実際に観測したデータは10であり、この区間に含まれる

➡仮説1が「おかしい」という証拠はない

➡仮説1を保留する（棄てずにとっておく）


仮説1 (N=16) が正しい場合
0.20
0.15
0.10
0.05
0.00
確率
0
5
10
15
表の回数

仮説2の検証

仮説2（$N = 36$）が 正しいとすると、

- 平均 $= N\theta = 36 \cdot 0.5 = 18$
- 分散 $= N\theta(1 - \theta) = 36 \cdot 0.5 \cdot (1 - 0.5) = 9$
- 標準偏差 $= \sqrt{\text{分散}} = 3$
    - 平均$- 1.96 \cdot \text{sd} = 18 - 1.96 \cdot 3 = 12.12$
    - 平均$+ 1.96 \cdot \text{sd} = 18 + 1.96 \cdot 3 = 23.88$

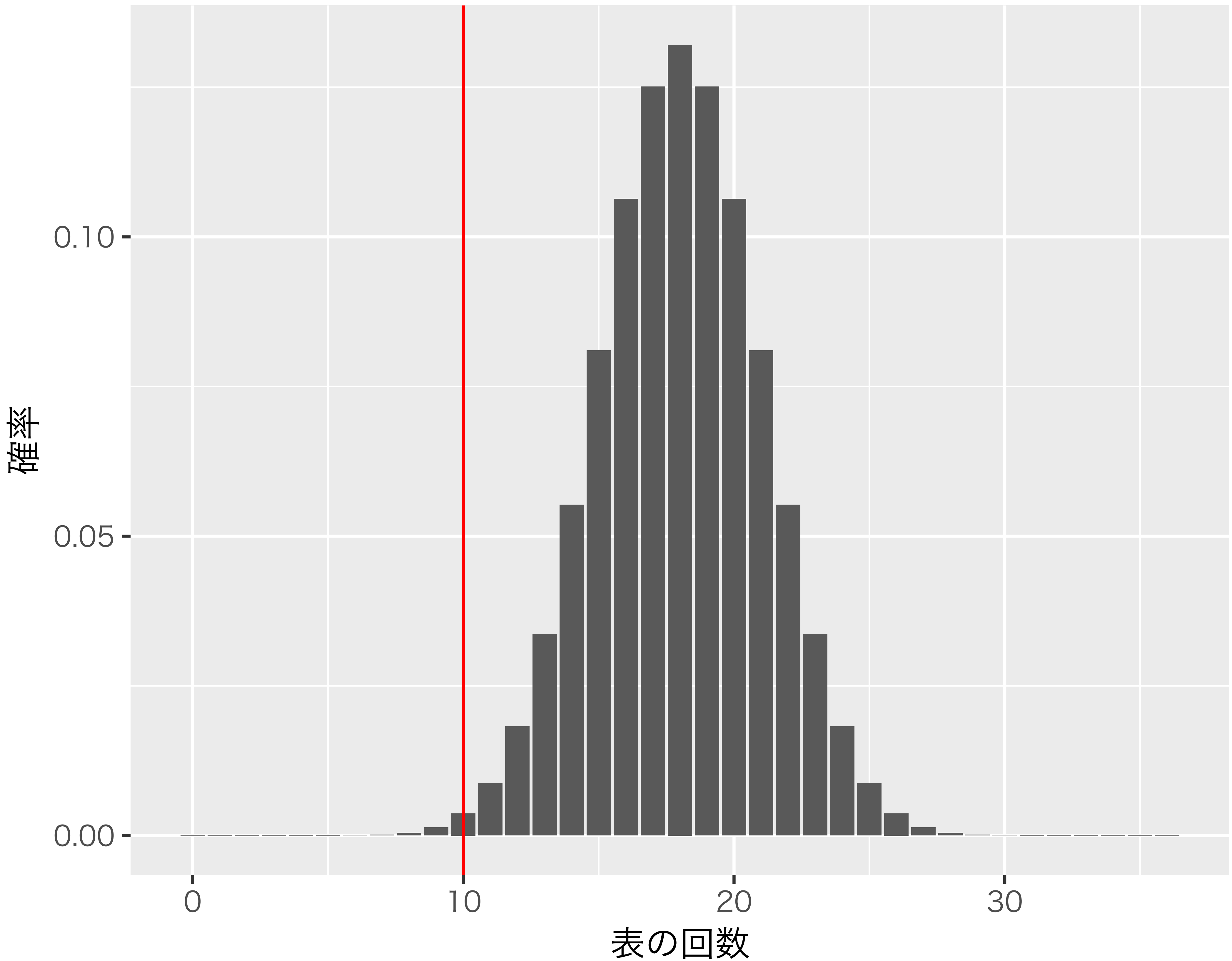
仮説2の検証（続）

仮説2（$N = 36$）が 正しいとすると、

- データの95%は12.12と23.88の間の値を取る
- 実際に観測したデータは10であり、この区間に含まれない

➡仮説2を棄却 (reject)する（仮説2は妥当でない）

仮説2 (N=36) が正しい場合



例題の結論

- 仮説1は妥当だが、仮説2は妥当とはいえない

▶ 「妥当＝真実」**ではない**！

- $N = 20$ や $N = 19$ という仮説も、受け容れられるかも
- しかし、真実の N はただ1つ存在する
  - 仮説検定では、**正しくないものははっきりわかる**（棄却できる）が、**保留した仮説が正しいとは限らない**（単に「ありそう（妥当）」というだけ）

統計的推定の基礎

統計的推定の基礎

同じ例題で考える

- 正しいコインを N 回投げる
- 表が10回出た
- 投げた回数 N は何回だと考えられるか？

1. 1つの値を答える：点推定
2. 予測に幅をもたせる：区間推定

点推定の例

- N 回コイン（表が出る確率 $\theta = 0.5$）を投げたところ、表が10回出た
- 投げた回数 N はいくつだと考えられるか？
  - 二項分布の平均は $N\theta$
  - 手持ちのデータは10 $\rightarrow$ 平均は10
  - $N\theta = 10 \Rightarrow N = 10/\theta = 10/0.5 = 20$：**点推定値**

区間推定の例

- コイン（表が出る確率 $\theta = 0.5$）を N 回投げて表が10回出た
- 投げた回数 N は何回から何回の間だと考えられるか？
  - 統計的検定により、16枚は妥当な仮説だが36枚は妥当な仮説でないことがわかっている
  - 他にも妥当な仮説はあるはず（例: $N = 20$）
    - 妥当な仮説全体を、推定値として使う

区間推定の例 (続)

- コイン（表が出る確率 $\theta = 0.5$）を N 回投げて表が10回出た
  - 平均 $\mu = N\theta = N/2$, 分散 $\sigma^2 = N\theta(1-\theta) = N/4$
- このとき、どんな仮説が棄却され、どんな仮説が保留される？
  - $-1.96 \le z \le 1.96$ となる z を与える N は保留される（棄却されない)
  - z は、

$$z = \frac{x - \mu}{\sigma} = \frac{10 - \frac{N}{2}}{\sqrt{\frac{N}{4}}}$$

区間推定の例（続）

- $N = 13$ から $N = 30$ までの仮説はどれも棄却できない
- $N < 13, N > 30$ は棄却
  - 区間推定：$13 \leq N \leq 30$

★「±1.96」は標準正規分布の95%が収まる最短区間

  - 求めた区間を「95%信頼区間」と呼ぶ

N	z
12	2.309
13	1.942
14	1.604
15	1.291
16	1
17	0.728
18	0.471
19	0.229
20	0
21	-0.218
22	-0.426
23	-0.626
24	-0.817
25	-1
26	-1.177
27	-1.347
28	-1.512
29	-1.671
30	-1.826
31	-1.976

ここまでのまとめ

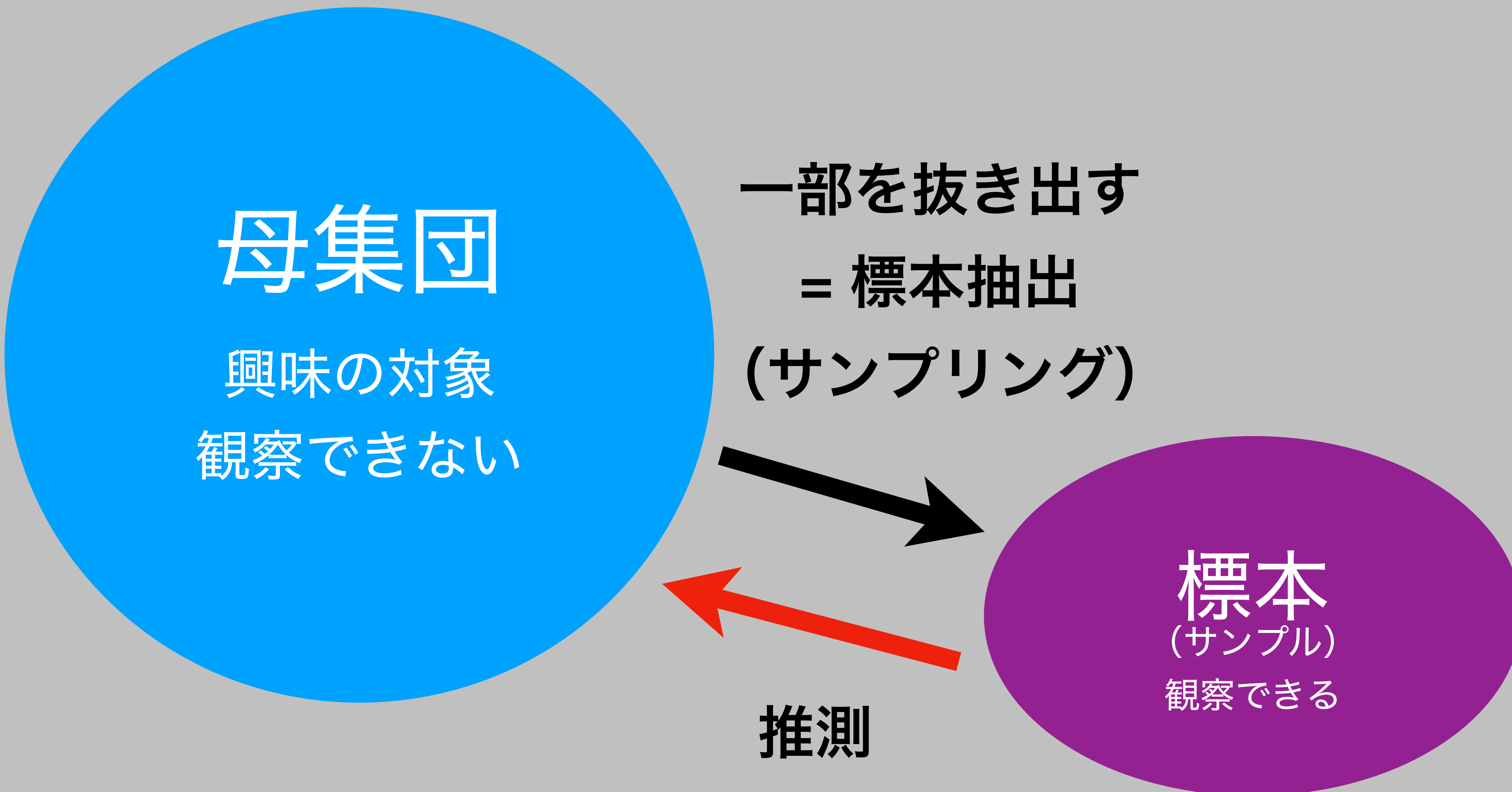
- 推測統計学とは、部分（標本、データ）から全体（母集団、興味の対象）を知るための方法
  - 統計的検定：仮説を保留 or 棄却？
  - 統計的推定：点推定と区間推定
- 実習：
  - https://yukiyanai.github.io/jp/classes/stat2/contents/R/introduction-to-inference.html

母集団と標本

調査の方法

- 全数調査（悉皆(しっかい)調査）：興味がある集団そのもの（母集団全体）を調べる調査
  - 例：国勢調査
- 標本調査：母集団の特徴を知るために、その一部（標本）を取り出して調べる
  - 例：世論調査など多くの調査

母集団と標本



標本調査の必要性

- 興味の対象が大きいとき、すべてを調べるのは大変（例：日本人全体が母集団の場合）
  - 時間・金・人手がかかる（2010年国勢調査の経費は約670億円）
- すべてを調べられない場合もある
  - 例：製品の耐性テスト、料理の味見

母数と標本の統計量（1）

- 母数（パラメタ, parameter）：母集団が持っている特徴
  - 母平均、母分散、母比率など
- 統計量 (statistic)：標本から知ることができる特徴
  - 標本平均、標本分散、標本比率など

母数と標本の統計量（2）

	母数（母集団）	統計量（標本）
標準偏差	母標準偏差	標本標準偏差
分散	母分散	標本分散
比率	母比率	標本比率
平均	母平均	標本平均

推測統計学

- 統計量（statistics）を使って**母数（パラメタ, parameters）**を推測する！

文字の使い分け

- 母集団：ギリシャ文字
- 標本：アルファベット

★ ただし、この使い方は絶対ではない

	母数	統計量
標準偏差	σ（sigma）	s
分散	σ^2	s^2
比率	π (pi)	p
平均	μ (mu)	*変数名にバーを付ける

標本の選び方

- 標本の選び方は様々
- 明らかにダメな例：

★日本の有権者全体に興味があるとき、

- 女性だけ選ぶ
- 高齢者だけ選ぶ
- 東京都民だけ選ぶ

✦ これらはどれも偏っている（**バイアス** [bias] がある）

単純無作為抽出 (simple random sampling; SRS)

- 母集団から標本をランダムに（◯確率的に; ×でたらめに）選ぶこと
- 母集団を構成するそれぞれの個体が選ばれる確率が等しい
  - 無作為抽出で選び出された標本は、母集団の**偏りのない**縮図であるとみなすことができる
  - ただし、**誤差 (error)** は必ずある

標本の選び方と調べ方

- 単純無作為抽出以外のサンプリング法や調査の実施方法（面接調査、郵送調査など）については「社会調査」の文献を参照
  - 廣瀬雅代ほか『サンプリングって何だろう』（2018年、岩波書店）
  - 大谷信介ほか『社会調査へのアプローチ　第2版』（2005年：ミネルヴァ書房）
  - 神林博史・三輪哲『社会調査のための統計学』（2011年：技術評論社）

標本の数 ≠ 標本サイズ

- 標本の数：母集団から取り出した集団の数（通常は1つの標本しか手に入らない）
- 標本サイズ（N）：**1つの標本**に含まれる個体の数

例）日本の有権者から2,000人の標本を2回抽出した

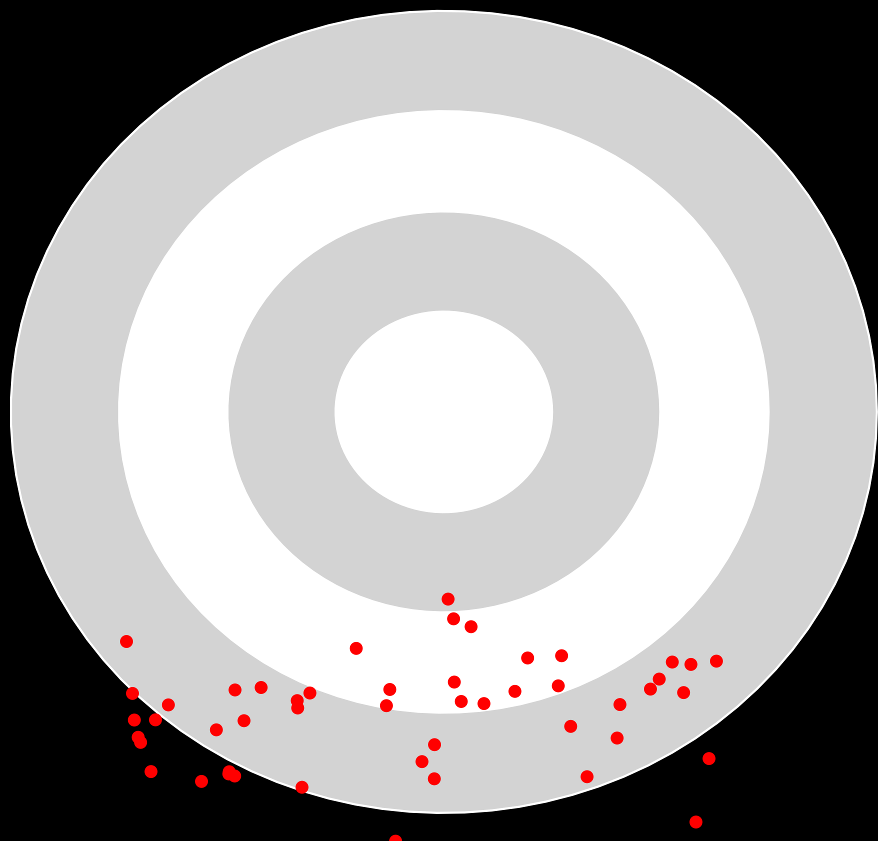
- 標本の数 = 2
- 標本サイズ N = 2000

標本には誤差がある

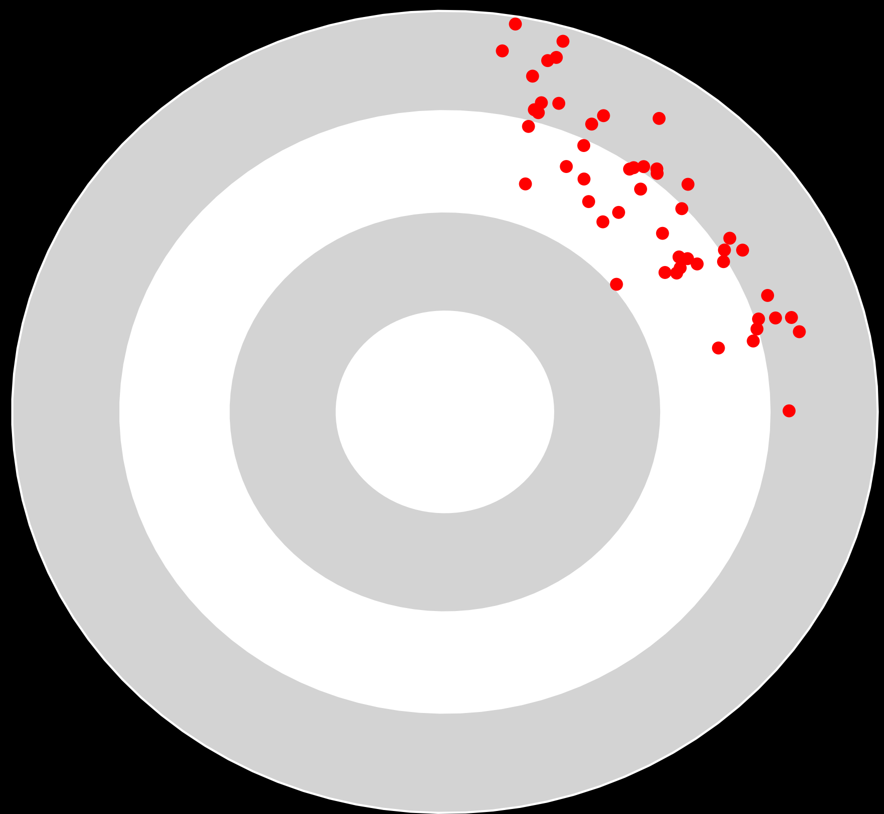
- 標本から得られる統計量が母数にぴったり一致するとは限らない！
  - 誤差（error）がある
- 問題は

1. 誤差に偏り（bias）があるかどうか
   - 偏りがないもの（誤差の平均が0）が望ましい
2. 誤差の大きさ
   - 正確に推測するためには誤差が小さい方がよい

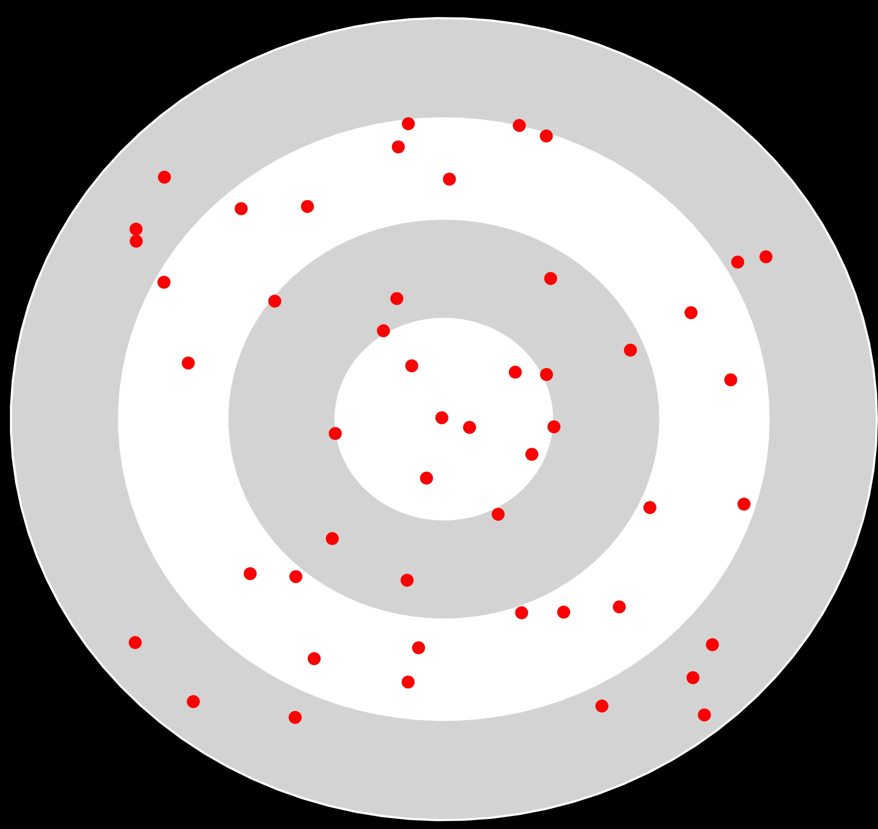
(a) 偏りがある



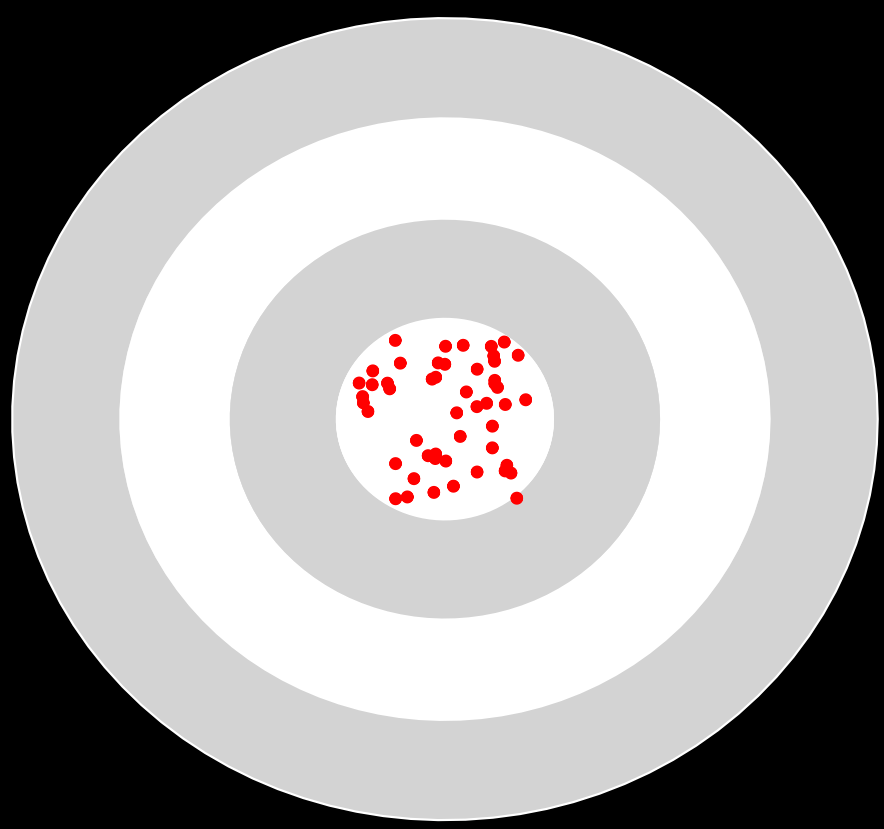
(b) 偏りがある



(c) 偏りがなく、ばらつきが大きい



(d) 偏りがなく、ばらつきが小さい



(d) がベスト！

1万人から100人を抽出する（1）

例）女性4600人（母比率 $\pi = 0.46$）、男性5400人（$1 - \pi = 0.54$）の計1万人からなる母集団から100人を単純無作為抽出で選ぶ

- 標本1：女性比率= 50/100人 = 0.5 > 0.46
- 標本2：女性比率 = 44/100人 = 0.44 < 0.46
- 標本3：女性比率 = 46/100人 = 0.46
- 他の標本　：女性比率 = ？

1万人から100人を抽出する（2）

- 1万人から100人を選ぶ方法は全部で約6.5x10^{241}通り → 全部の組み合わせを試すのは難しい（実践的には不可能）

- コンピュータ・シミュレーションで $N = 100$ のサンプルを1万個抽出してみる（標本サイズ=100, 標本の数=10,000）

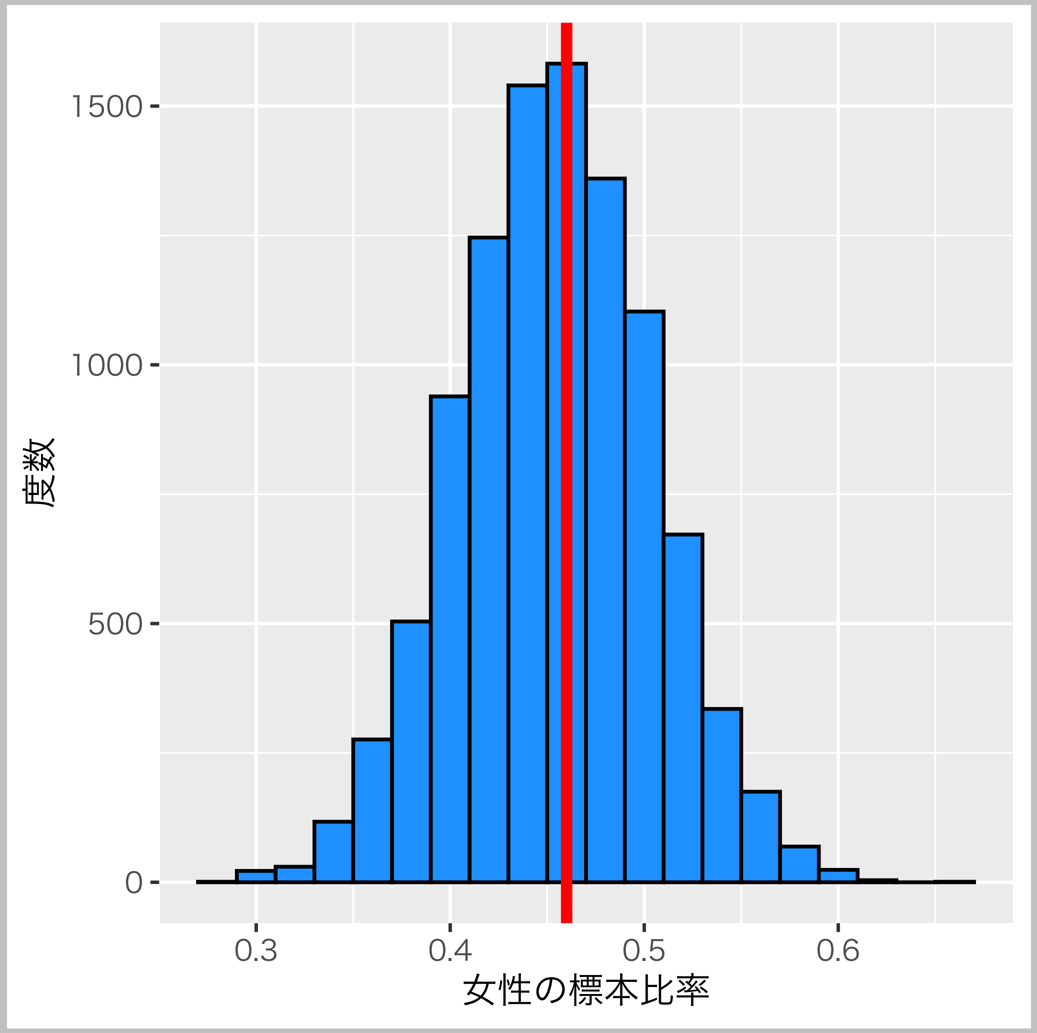
1万人から100人を抽出する（3）

- 女性の標本比率は、母比率より大きかったり小さかったりする
- 標本比率のばらつきの中心は母比率

★偏りがない：「**平均すれば**」知りたいことがわかる

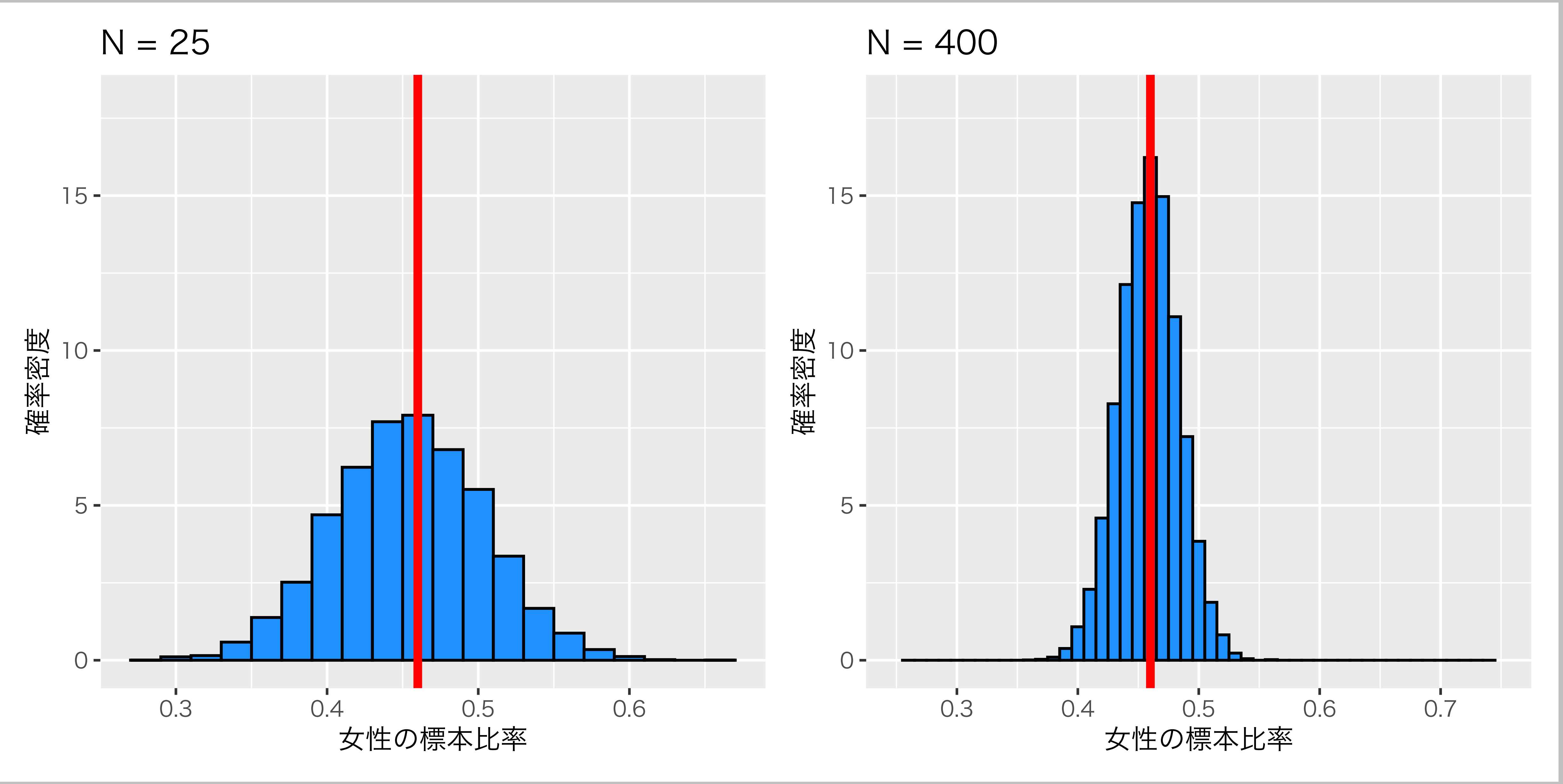
- 統計量の標本間でのばらつきは**標準誤差**で測られる

標本サイズ100の標本を
1万個抽出した結果



標本サイズを変えてみる

標本サイズ N の標本を1万個抽出した結果



「母集団と標本」のまとめ

- 母集団から標本を抽出する
- 標本には誤差がつきもの
  - 標本分布と標準誤差（次回の内容）
  - 標本サイズが重要な気がする（今後の注目ポイント）
- 実習：
  - https://yukiyanai.github.io/jp/classes/stat2/contents/R/pop-n-samples.html

次回予告

8. 標本平均と母平均