

統計学 2

10. 母集団と標本

矢内 勇生

2019年5月20日

高知工科大学 経済・マネジメント学群

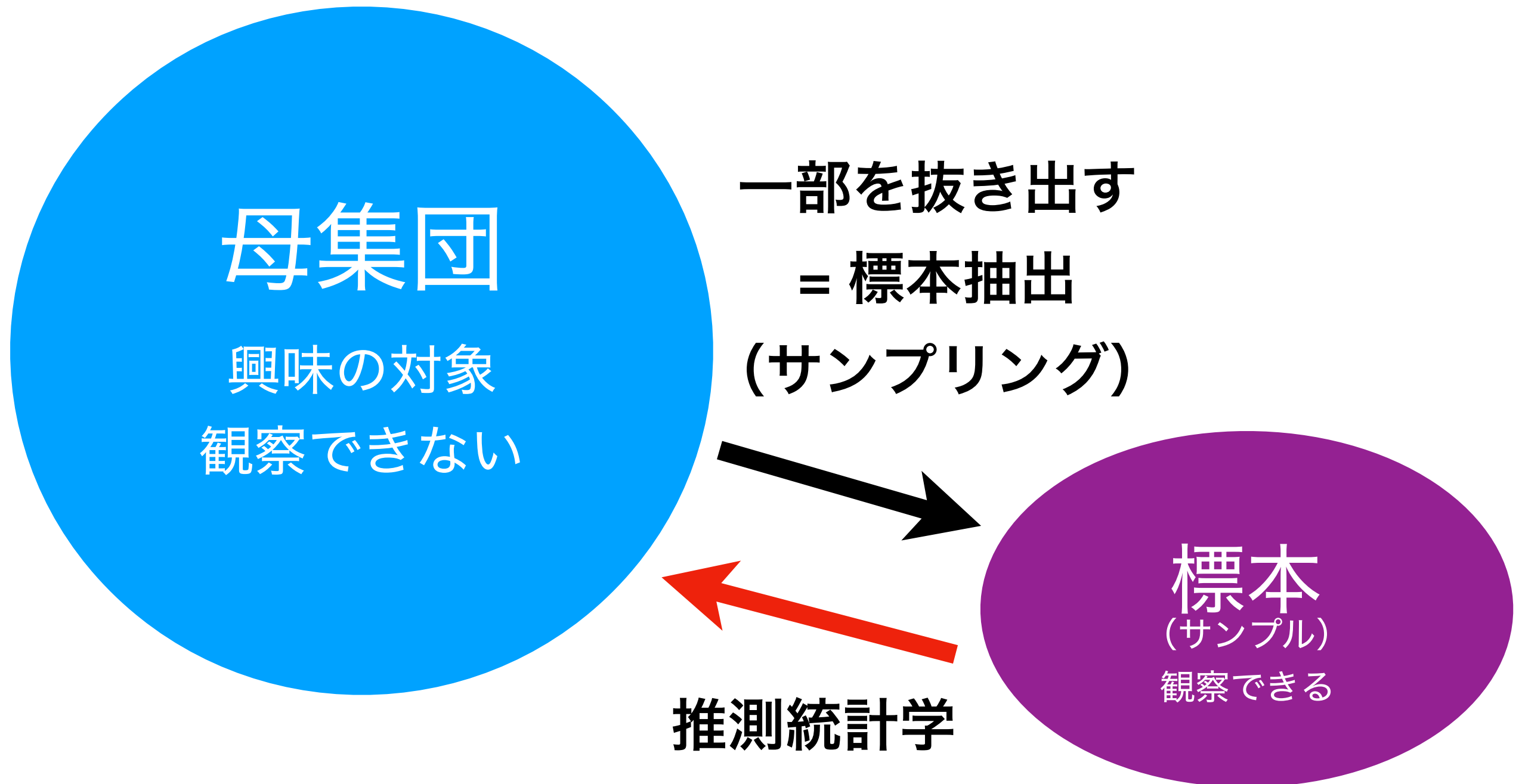
今日の目標

- 母集団と標本の違いを理解する

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）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；×でたらしめに）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**はつきもの

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2000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

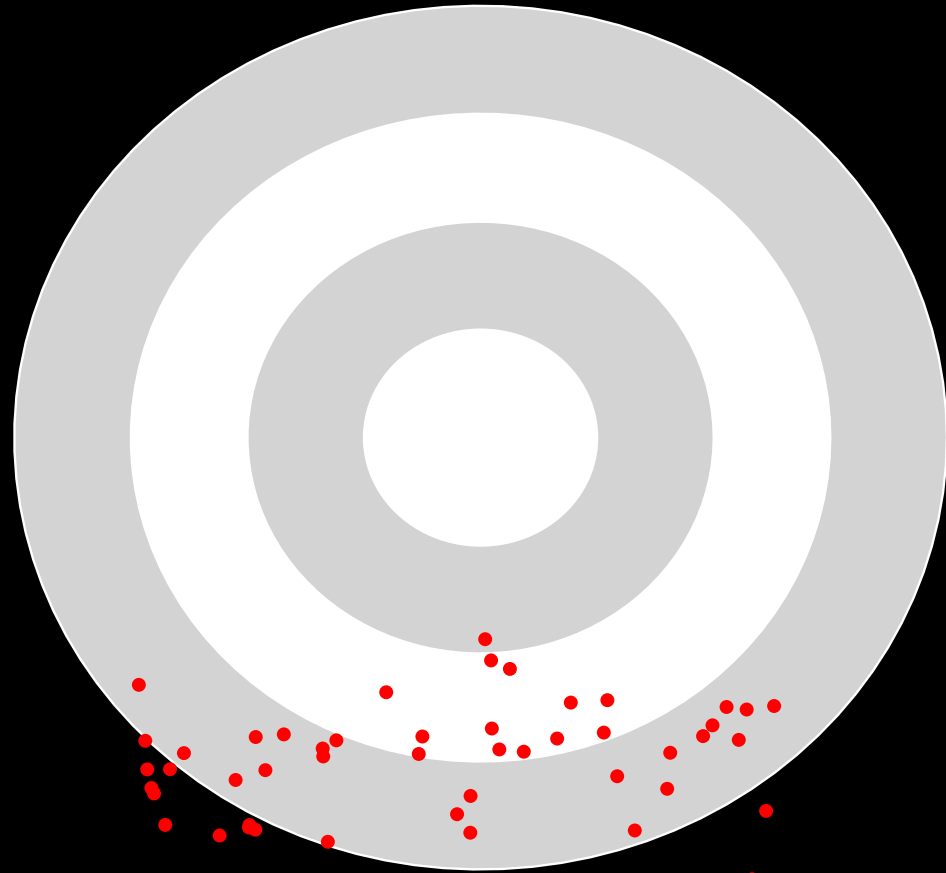
1. 誤差に偏り (bias) があるかどうか

- 偏りが無いもの (誤差の平均が0) が望ましい

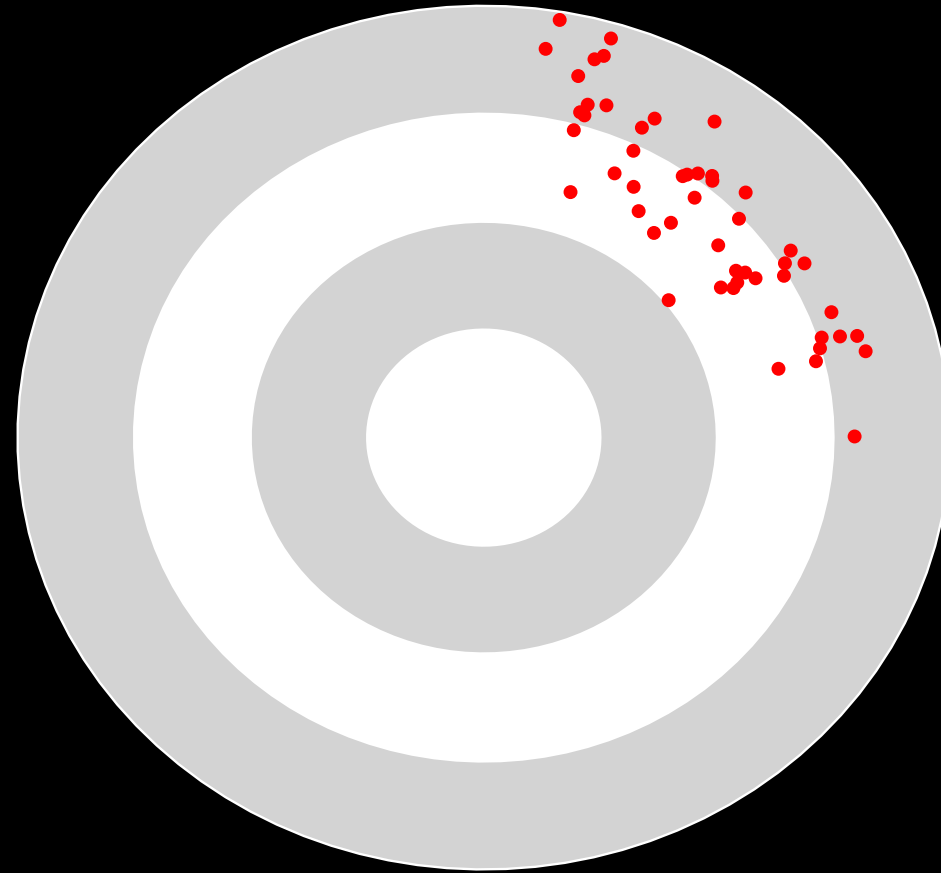
2. 誤差の大きさ

➡ 正確に推測するためには誤差が小さい方がよい

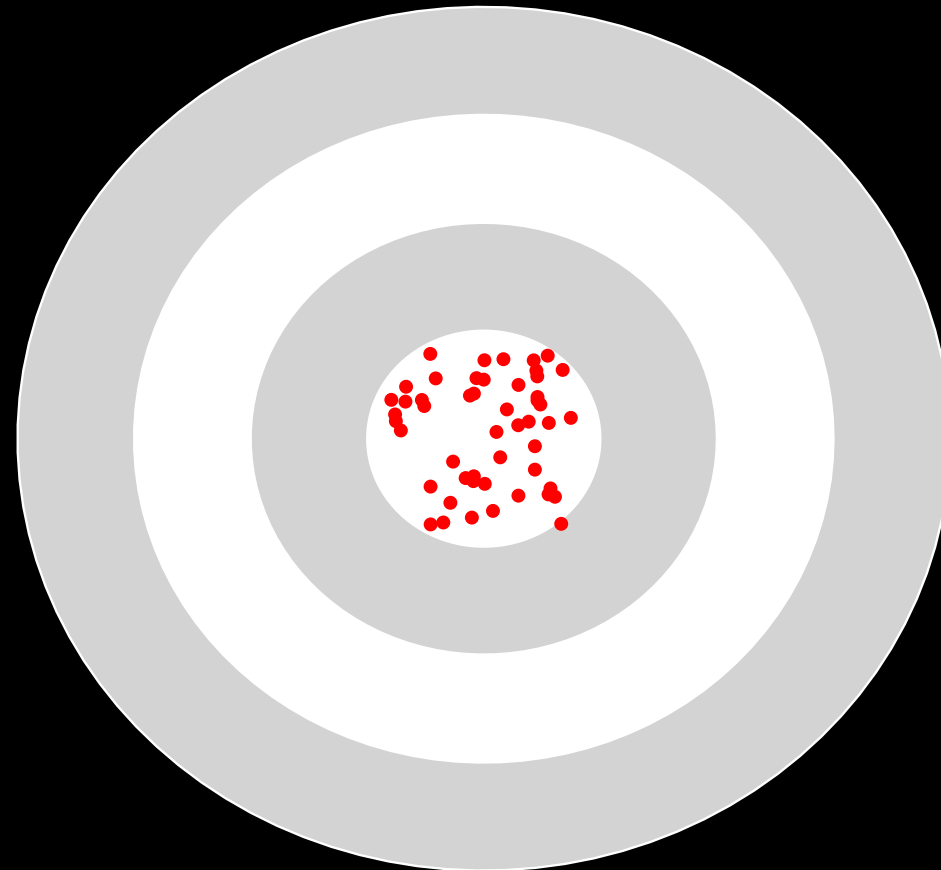
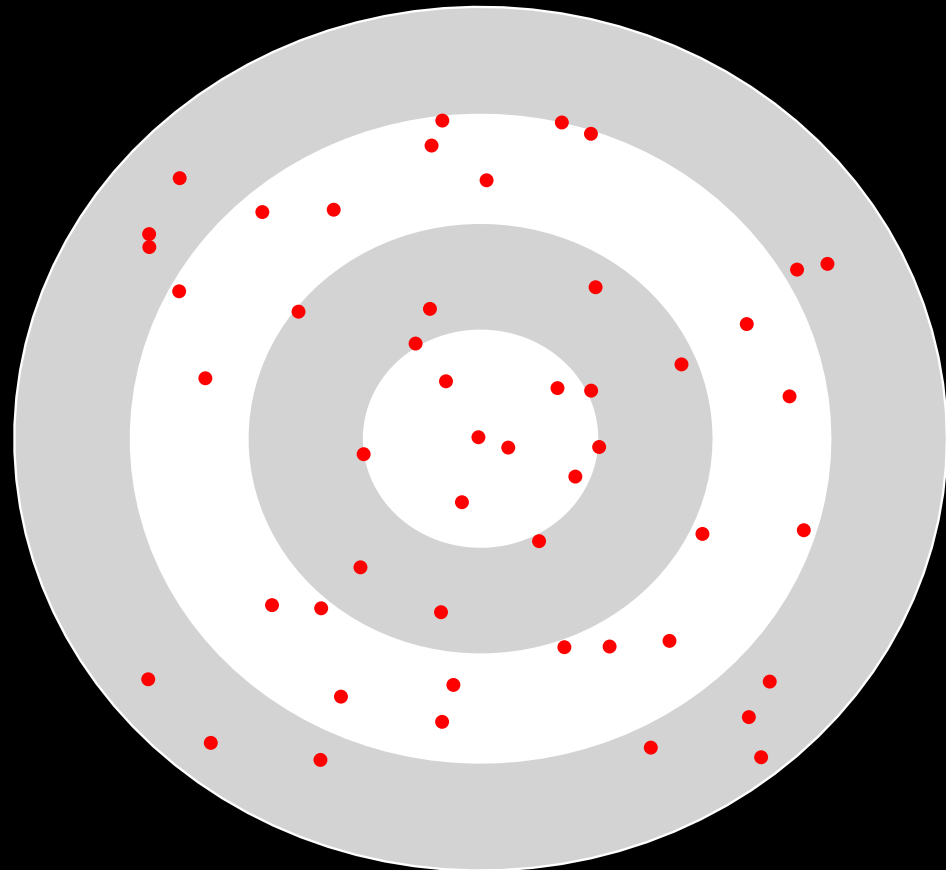
(a) 偏りがある



(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) が理想！

1万人から100人を抽出する (1)

例) 男性5400人 (比率は0.54)、女性4600人 (0.46) 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

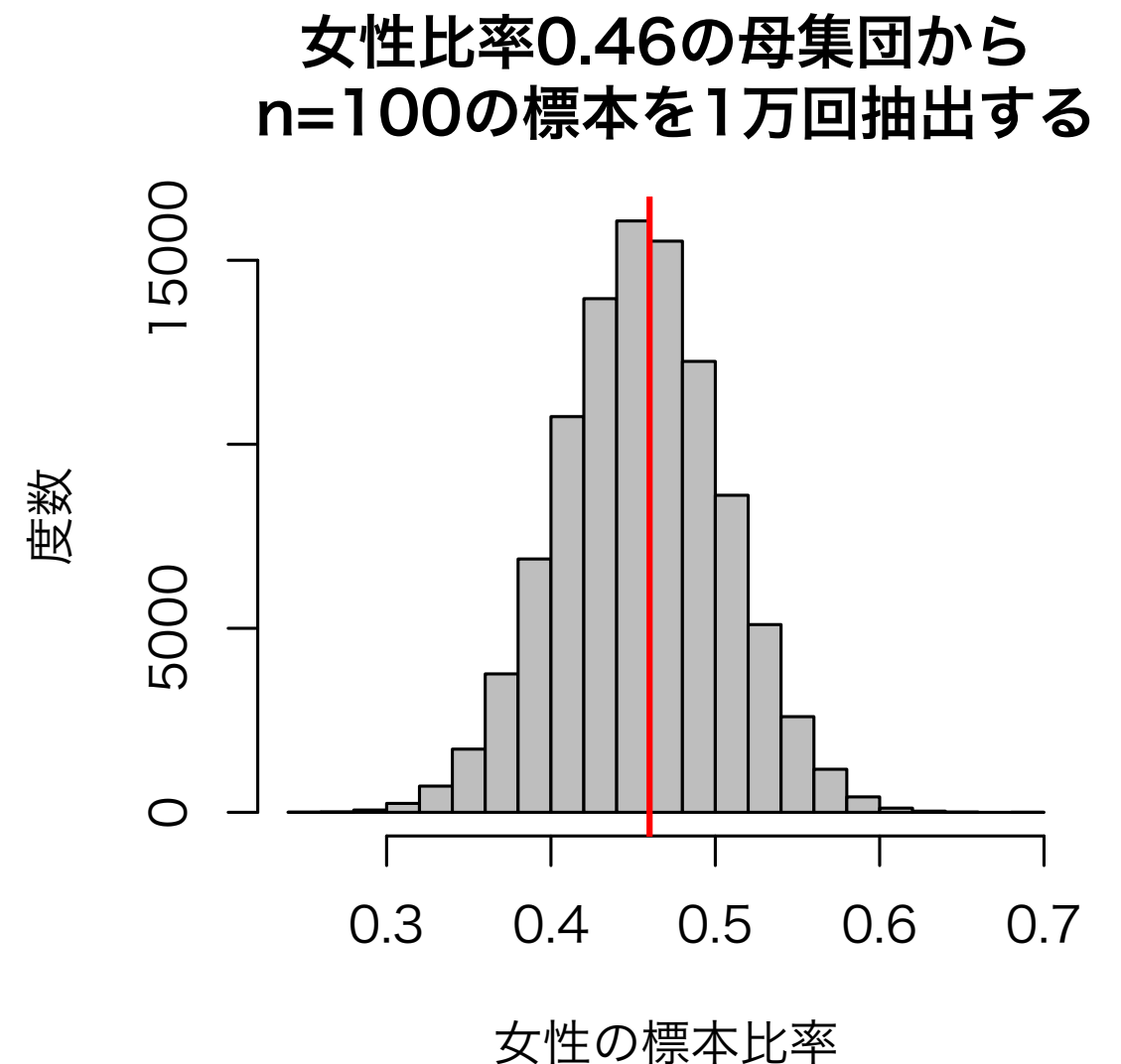
- ▶ 標本1 : 女性比率 = $50/100人 = 0.5 > 母比率 0.46$
- ▶ 標本2 : 女性比率 = $44/100人 = 0.44 < 0.46$
- ▶ 標本3 : 女性比率 = $46/100人 = 0.46$
- ▶ 他の標本 : 女性比率 = ?

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 6.5×10^{241} 通り
→ 全部の組み合わせを試すのは難しい
- 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=10000)

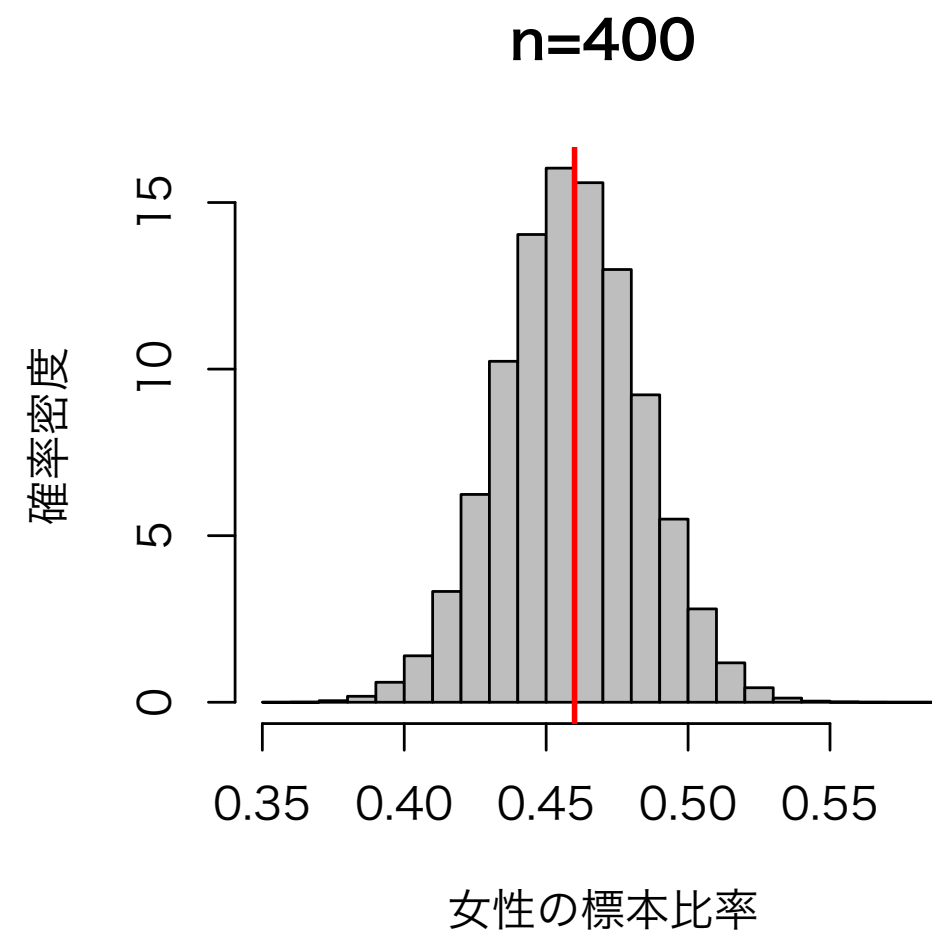
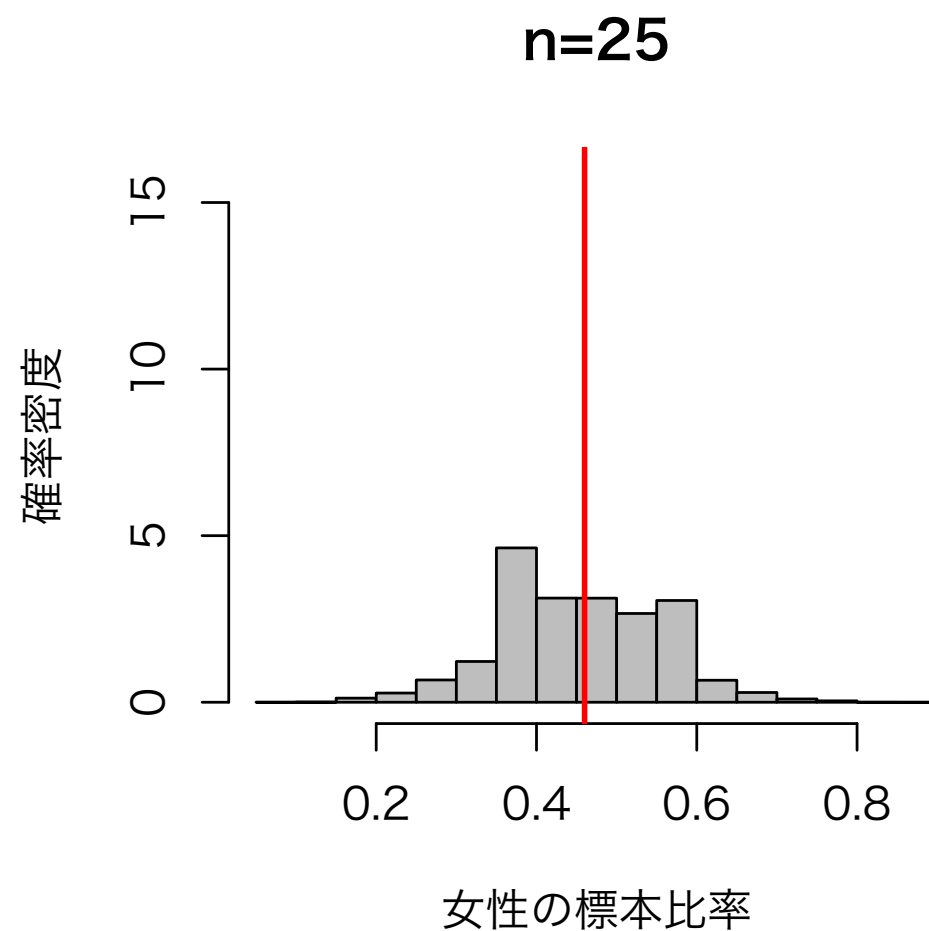
1万人から100人を抽出する (3)

- 女性の数は多すぎたり少なすぎたりする
 - 標本比率のばらつきの中心は母比率
- ➡ 偏りが無い（平均的には、知りたいことがわかる）



標本サイズを変えてみる

女性の割合が0.46 の母集団から
標本サイズ n の標本を1万個抽出する



今日のまとめ

- 母集団から単純無作為抽出で標本をとる
- 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）