



統計学入門 第14回

早稲田大学政治経済学部
西郷 浩



本日の目標

- 全数調査と標本調査
 - 全数調査の長短
 - 標本調査の長短
- 標本抽出
 - 確率標本抽出と有意標本抽出
- 確率変数としての標本平均
 - 標本平均(統計量の一つ)の標本分布



全数調査と標本調査(1)

■ 統計調査

- 調査対象を決め、調査の日時・期間・方法を定めて、調査対象の特徴をあらわす調査項目について調査対象から回答をえること。

- 例:

- わが国の人口に関する調査
- わが国の世帯の所得に関する調査
- わが国の事業所の売上高、従業員数に関する調査
- わが国の企業の活動に関する調査



全数調査と標本調査(2)

■ 全数調査

■ 調査対象全部を調査する。

■ 国勢調査(総務省)

- わが国の常住人口全部を調査する。

■ 経済センサス(総務省・経済産業省)

- わが国で操業する事業所・企業全部を調査する。

■ 商業統計調査(経済産業省)

- わが国の卸売業・小売業事業所全部を調査する。

■ 工業統計調査(経済産業省)

- わが国の製造業事業所全部を調査する。



全数調査と標本調査(3)

■ 標本調査

- 調査対象の一部(標本)を調査する。
 - 家計調査(総務省)
 - 家計の支出・貯蓄等を調査する。
 - 商業動態統計調査(経済産業省)
 - 商業の売上高を調査する。
 - 毎月勤労統計調査(厚生労働省)
 - 事業所の雇用や給与、労働時間を調査する。
 - 作物統計調査(農林水産省)
 - 農業経営体の収穫量にを調査する。



全数調査と標本調査(4)

■ 全数調査の長短

■ 長所

- 標本誤差がない。
 - 標本誤差:一部しか調べないことによって発生する誤差
- 小地域についての集計が可能である。

■ 短所

- 費用・時間がかかる。
- 詳細な質問項目が設けにくい。
- 非標本誤差(標本誤差以外の誤差)が大きい。



全数調査と標本調査(5)

■ 標本調査の長短

■ 長所

- 費用・時間が(全数調査ほど)かからない。
- 詳細な項目が調べられる。
- 非標本誤差が小さい。

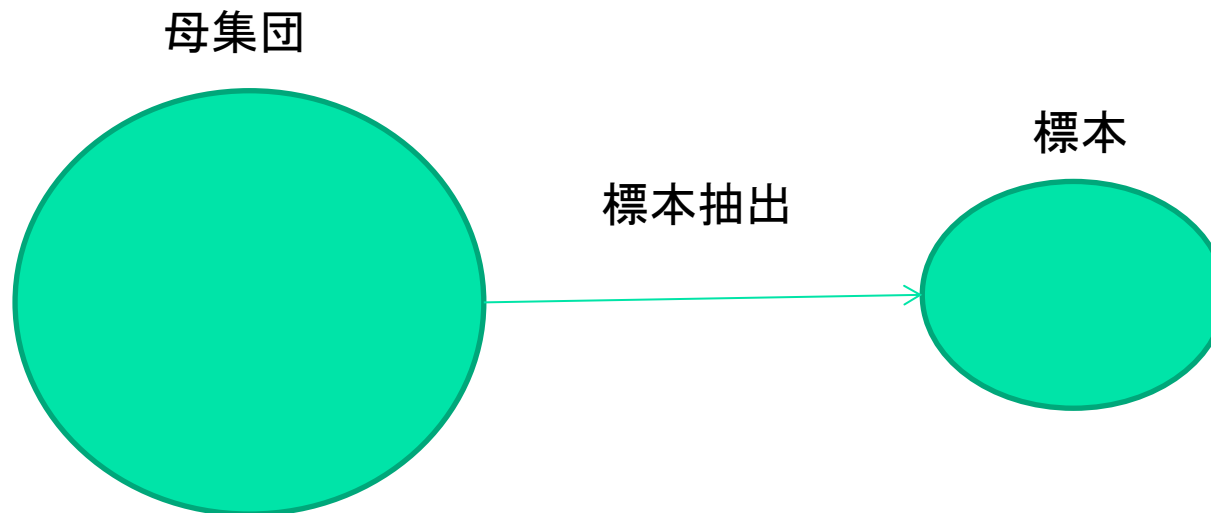
■ 短所

- 標本誤差が発生する。
 - ただし、確率的に制御可能である。

標本抽出(1)

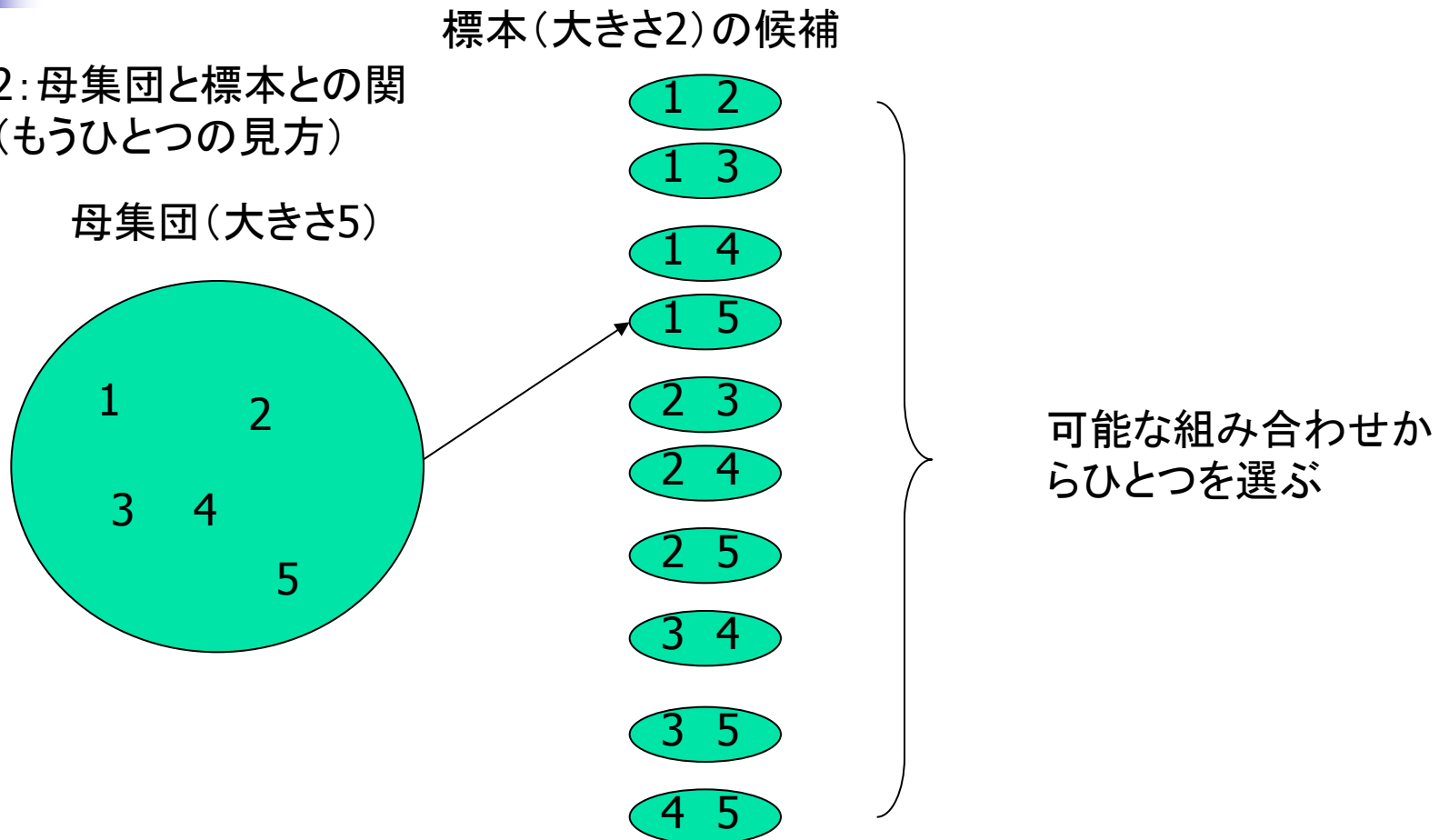
- 母集団と標本

図1: 母集団からの標本抽出



標本抽出(2)

図2: 母集団と標本との関係(もうひとつの見方)





標本抽出(3)

- 標本抽出

- 確率標本抽出

- ひとつの標本が、あらかじめ決められた確率で抽出される。
 - 単純無作為抽出(無作為抽出ともよぶ)

- 有意標本抽出

- ひとつの標本が抽出される確率があらかじめ決められていない。

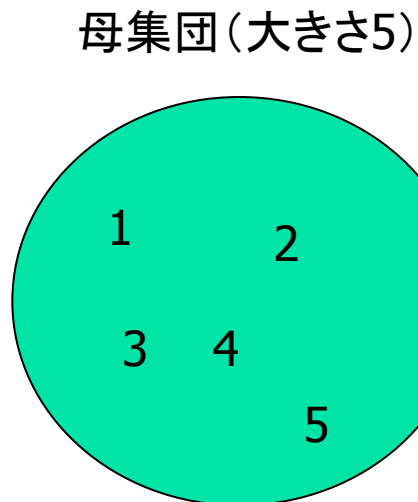


標本抽出(4)

- なぜ確率標本抽出が必要か？
 - 一部(標本)の情報から全部(母集団)の情報を復元できるようにするため。
 - 公的統計調査では、個別の事例よりも、全体がどのようなになっているかを明らかにすることが重要になる。
 - 実際の調査では...。
 - 確率標本抽出が額面どおりに実現するとはかぎらない。

標本抽出(5)

図3: 無作為抽出



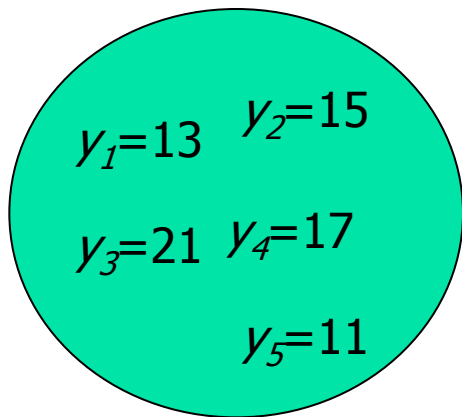
標本(大きさ2)の候補 選出確率

1 2	1/10
1 3	1/10
1 4	1/10
1 5	1/10
2 3	1/10
2 4	1/10
2 5	1/10
3 4	1/10
3 5	1/10
4 5	1/10

標本抽出(6)

図4: 標本平均の値

母集団(大きさ5)



母平均(母集団における平均)
 $= (13+15+21+17+11)/5$
 $= 15.4$

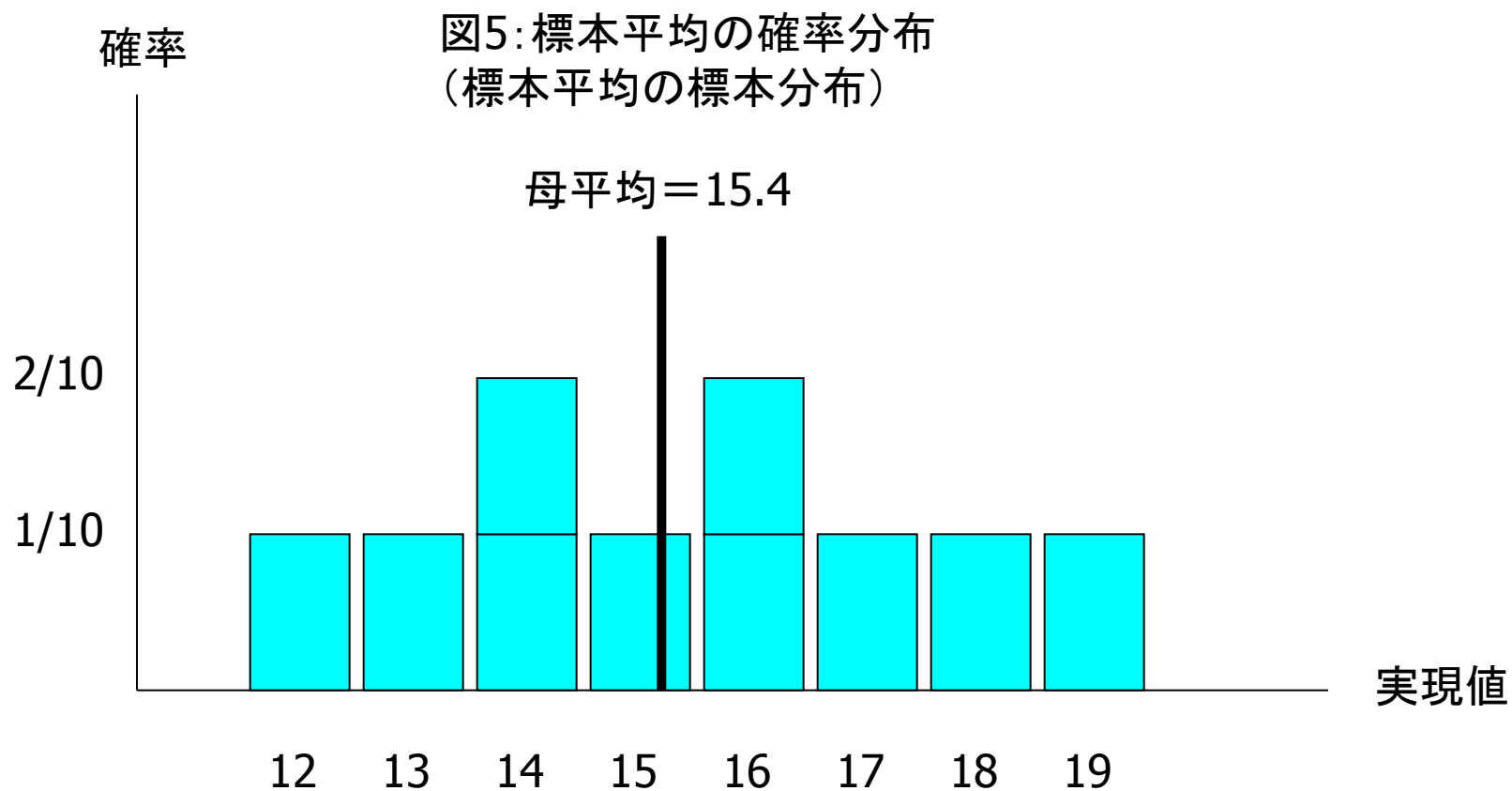
標本(大きさ2)の候補	選出確率	標本平均
1 2	1/10	$(13+15)/2=14$
1 3	1/10	$(13+21)/2=17$
1 4	1/10	$(13+17)/2=15$
1 5	1/10	$(13+11)/2=12$
2 3	1/10	$(15+21)/2=18$
2 4	1/10	$(15+17)/2=16$
2 5	1/10	$(15+11)/2=13$
3 4	1/10	$(21+17)/2=19$
3 5	1/10	$(21+11)/2=16$
4 5	1/10	$(17+11)/2=14$



標本抽出(7)

- 観察できること
 - どの標本平均も母平均と異なる。
 - 標本誤差がある。
 - ひとつひとつの標本平均と母平均とをばらばらに比べているかぎり、「標本誤差がある」ということ以外はなにもいえない。
- 見方を変えて...
 - 標本平均は確率的に変化するもの(確率変数)

標本抽出(8)





標本抽出(9)

- 標本平均

- $\bar{X} = \frac{1}{2}(X_1 + X_2)$

- 標本平均の期待値

- $$E(\bar{X}) = \frac{1}{10} \times 12 + \frac{1}{10} \times 13 + \frac{2}{10} \times 14 + \frac{1}{10} \times 15 + \frac{2}{10} \times 16 + \frac{1}{10} \times 17 + \frac{1}{10} \times 18 + \frac{1}{10} \times 19 = 15.4$$



標本平均の標本分布(1)

■ 実験1

■ 1セット:

- サイコロを10回投げて平均を計算する。

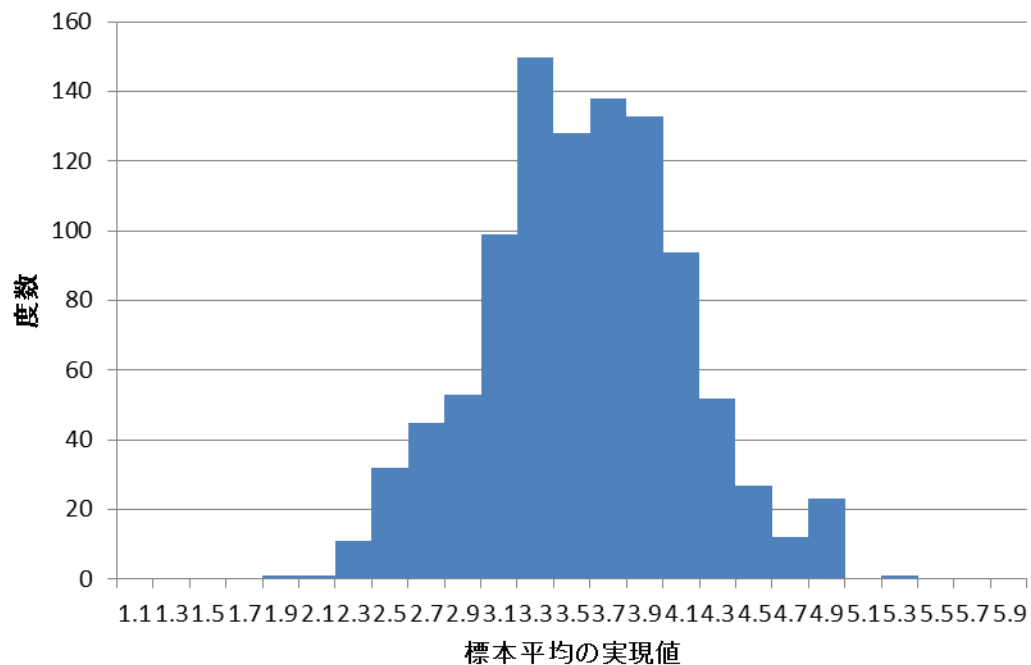
- Or: 1から6までの数値を10個(復元)無作為抽出して平均を計算する。

■ 1000セット繰り返してヒストグラムを作成する。

- シミュレーションによる標本平均の標本分布の近似。

標本平均の標本分布(2)

図6: 10個のサイコロの出目の標本平均の
標本分布の近似





標本平均の標本分布(3)

■ 実験2

■ 1セット:

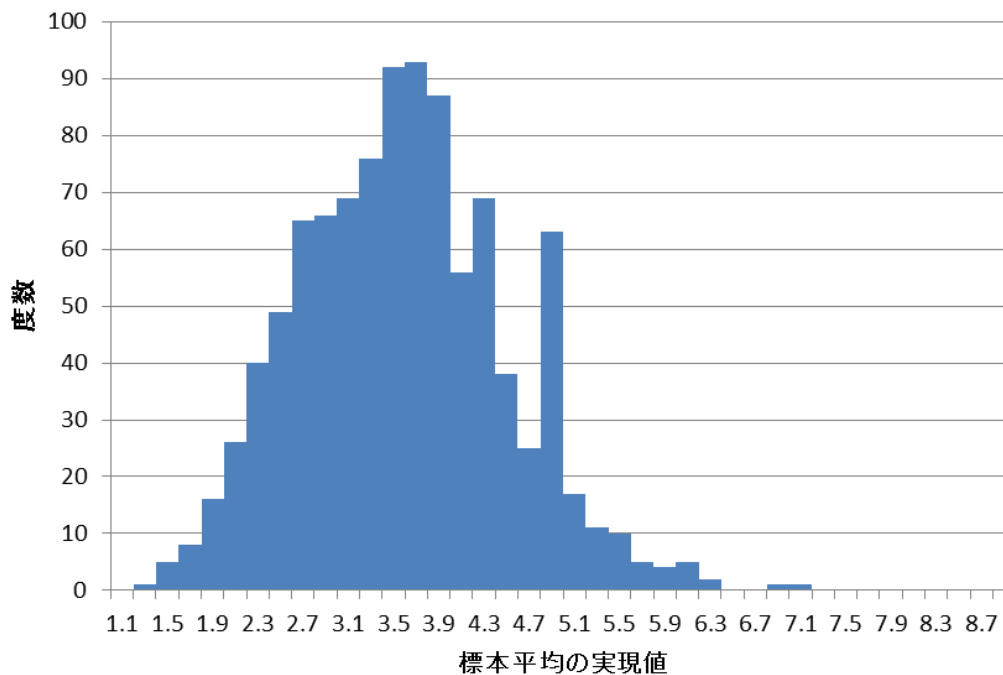
- [1, 1, 2, 2, 6, 9]から数値を10個復元無作為抽出して平均を計算する。

■ 1000セット繰り返してヒストグラムを作成する。

- シミュレーションによる標本平均の標本分布の近似。

標本平均の標本分布(4)

図7:10個の値からの無作為標本の標本平均の標本分布の近似





標本平均の標本分布(5)

- 標本平均の標本分布
 - 標本平均を確率変数として見たときの確率分布
 - 標本抽出に伴って発生する確率分布なので標本分布と呼ばれる。
 - The sampling distribution of the sample mean
 - 標本平均の期待値は、ひとつひとつの確率変数の期待値に等しい。
 - 標本平均の形はベル型に見える。



標本平均の標本分布(6)

- 推測統計では
 - 母平均の値を標本から推測する。
 - 母平均と標本平均との確率的関係を利用する。
 - 標本平均の期待値が母平均に等しい。
 - 標本平均の確率的な出方(標本分布)がベル型である。
 - ベル型の分布は、ある理論分布(正規分布)に近いことが知られている。
 - このことを活用して推測が可能となる。