

社会調査 第1回・第2回

早稲田大学政治経済学部

2018年度秋学期

西郷浩(前半)・土屋隆裕(後半)

本日の目標

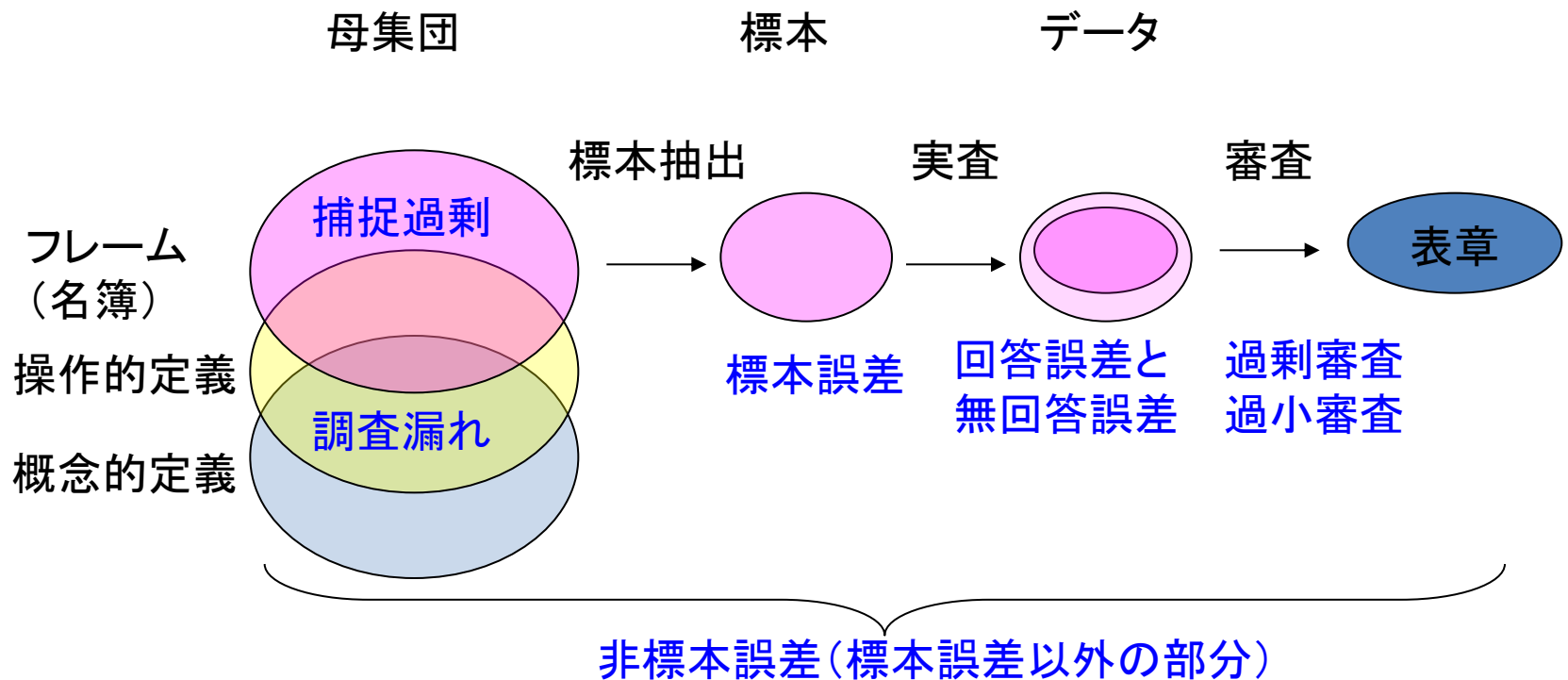
- 第1回授業
 - この講義のルール
 - 統計調査における誤差
- 第2回授業
 - 単純無作為抽出法
 - 定義
 - 性質
 - 標本平均の標本分布

第1回：この講義のルール

- 担当者2名の役割分担
 - 西郷浩(前半15回)
 - 社会調査の理論について解説
 - 土屋隆裕(後半15回)
 - 社会調査の実際について解説
- 講義の進め方
 - 月曜日4時限・5時限

第1回 調査における誤差(1)

■ 図1: 調査の段階と誤差



調査における誤差(2)

- 母集団を規定する際に発生する誤差
 - 概念的母集団：漠然とした調査対象
 - 例：日本に常住している人
 - 操作的に定義された母集団（目標母集団）
 - 例：日本に3ヶ月以上住んでいる（または住む予定である）人
 - 母集団名簿（抽出枠）
 - 作成（更新）時点と

調査における誤差(3)

- 標本を抽出する際に発生する誤差
 - 標本誤差
 - 一部(標本)の情報にもとづく計算結果
全部(母集団)の情報にもとづく計算結果
とのズレ
- 実査において発生する誤差
 - 無回答誤差: 全部 or 一部の項目が無回答
 - 回答誤差: 回答そのものにふくまれる誤差

調査における誤差(4)

- 審査の段階で発生する誤差
 - 過小審査
 - 修正すべきものを修正しなかった。
 - 過剰審査
 - 修正すべきでないものを修正してしまった。
- 公表の段階で発生する誤差
 - 利用者の誤用
 - 理解しやすい解説の必要性

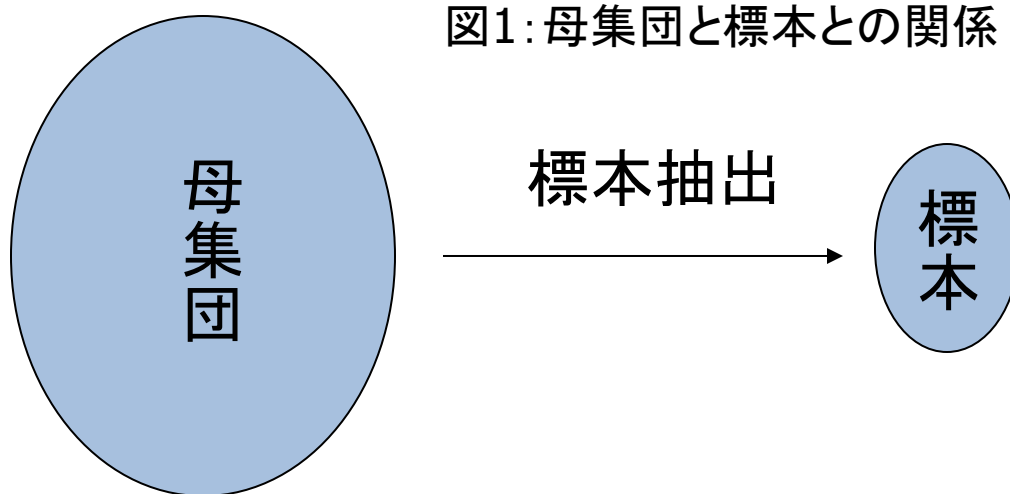
調査における誤差(5)

- 標本誤差と非標本誤差
 - 標本誤差
 - 理論的に考察しやすい。
 - 本日の講義の中心
 - 非標本誤差（標本誤差以外の誤差全部）
 - 影響が大きい。しかし、理論的に考察しにくい。
 - 実際の調査では、非標本誤差の管理が大切。

第2回単純無作為抽出法

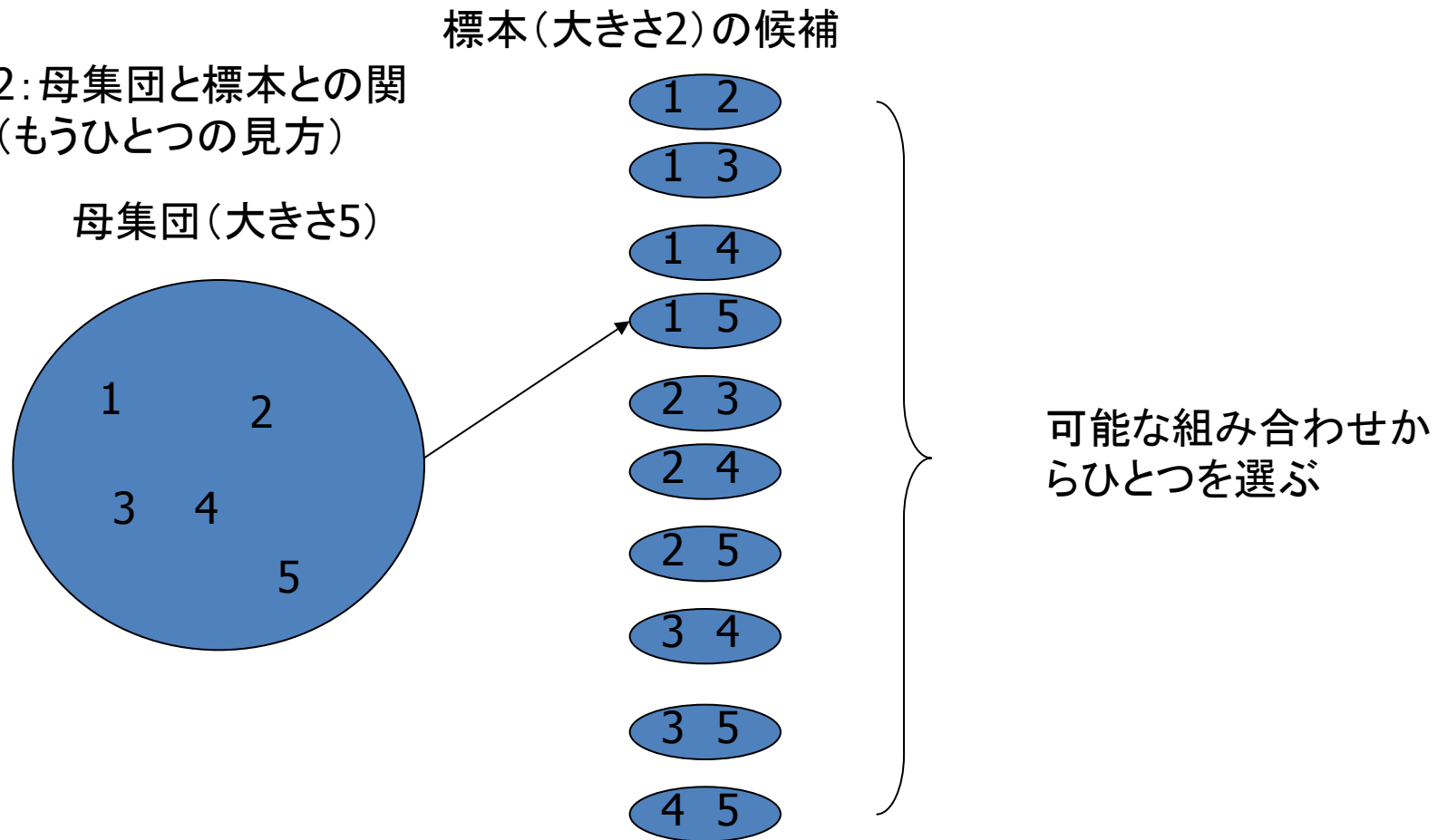
標本の選び方(1)

- 母集団：対象とする集団の全体（大きさ N ）
- 標本：母集団の一部（大きさ n ）



標本の選び方(2)

図2: 母集団と標本との関係(もうひとつの見方)



標本の選び方(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)
 - 層別無作為抽出(後述)

- 有意標本抽出

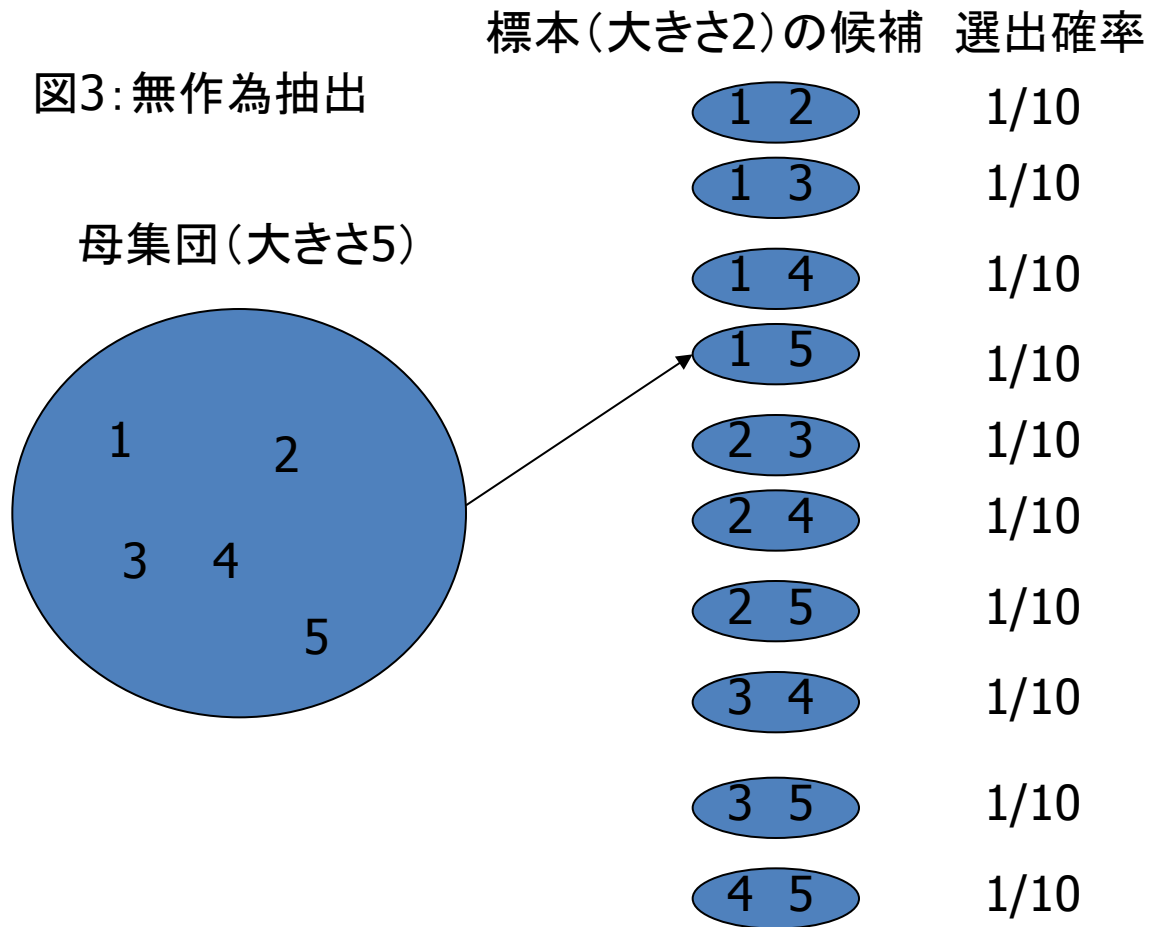
- ひとつの標本が抽出される確率があらかじめ決められていない。

標本の選び方(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。
 - 政府統計調査では、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。

標本誤差の見積もり(1)

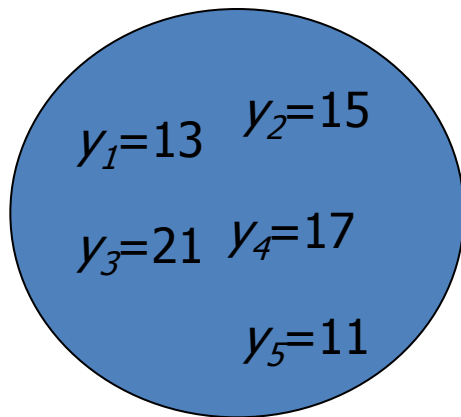
図3: 無作為抽出



標本誤差の見積もり(2)

図4: 標本平均の値

母集団(大きさ5)



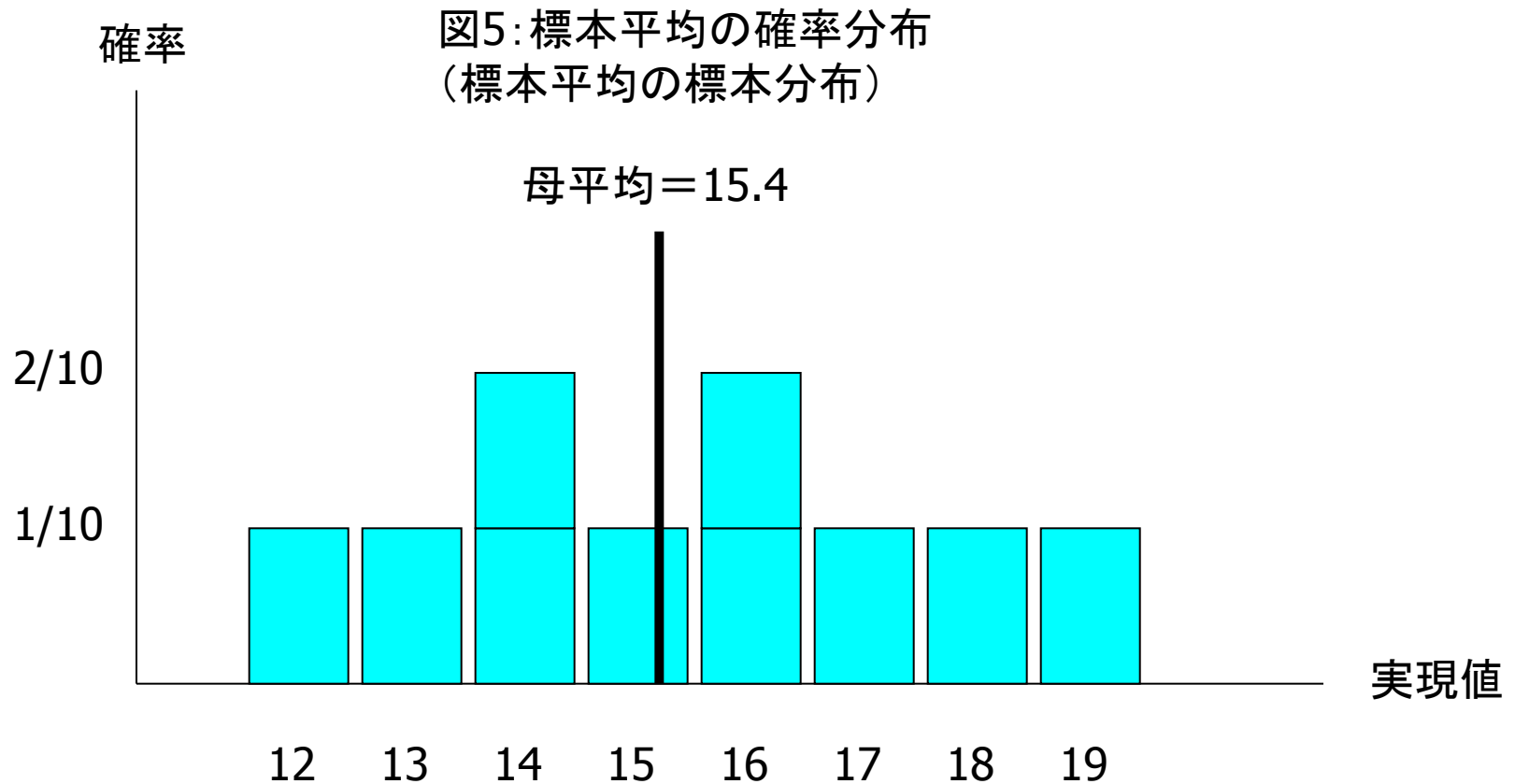
母平均(母集団における平均)
 $= (13+15+21+17+11)/5$
 $= 15.4$

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	$(17+11)/2=14$

標本誤差の見積もり(3)

- 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)

標本誤差の見積もり(4)



標本誤差の見積もり(5)

- 標本平均(確率変数)と母平均との関係
 - 標本平均の出方をあらわした確率分布の中心の位置に母平均が位置している。
 - 標本平均の期待値
$$\begin{aligned} &= (1/10) \times 12 + (1/10) \times 13 + (2/10) \times 14 \\ &\quad + (1/10) \times 15 + (2/10) \times 16 + (1/10) \times 17 \\ &\quad + (1/10) \times 18 + (1/10) \times 19 \\ &= 15.4 (\text{母平均}) \end{aligned}$$
 - 標本平均の出方全体(確率分布)と母平均には対応関係が見出せる。

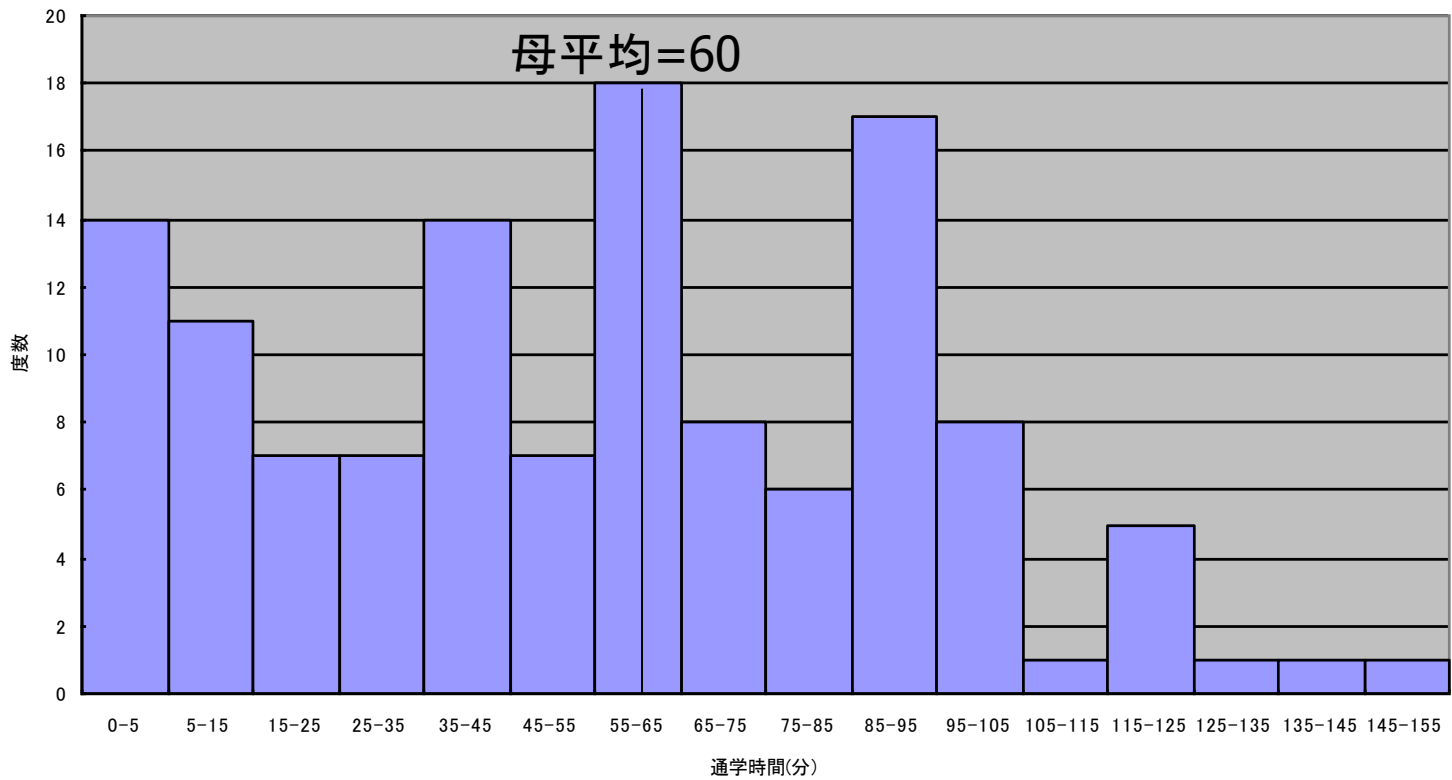
標本誤差の見積もり(6)

■ 実験

- 母集団:
2002年X月X日の「統計理論」出席者
- y : 通学時間(分)
- 母集団における通学時間の分布
 - ところどころにヒープがある。
 - 注: 記憶にもとづく回答の典型。
 - 標本平均(確率変数)の出方と母平均とに対応関係が見出せるか?

標本誤差の見積もり(7)

図6: 母集団における通学時間の分布



標本誤差の見積もり(8)

－実験の内容：

- ・「 n 人の出席者を無作為抽出して、標本平均を求める」を300回繰り返す。

－ $n = 2, 5, 10$.

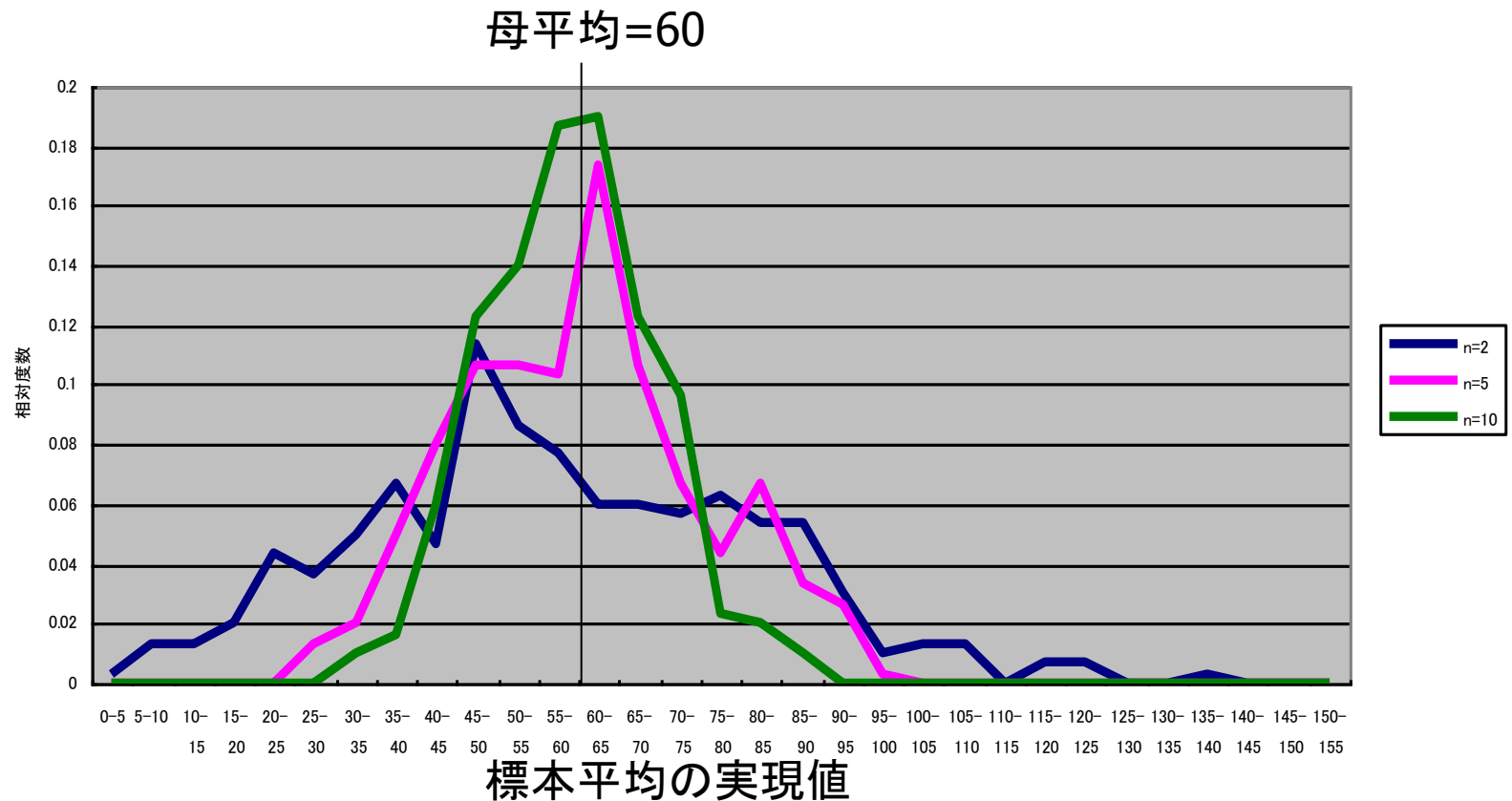
- ・300回求めた標本平均の度数分布(経験的な標本の出方)を描く。

－実験においてみるべきところ

- ・標本平均の出方(標本分布)と母平均との対応関係。

標本誤差の見積もり(9)

図7: 標本平均の標本分布と母平均との関係



標本誤差の見積もり(10)

- 観察できること
 - 母平均は、標本平均の標本分布(標本平均の確率的な出方)の中心に位置している。
 - 標本平均の期待値は母平均に等しい。
 - 「偏りがない」と表現する。
 - n が大きくなるにつれて、標本平均の出方のバラツキは小さくなる。
 - 標本平均の分散 = 母分散/ n

標本誤差の見積もり(11)

- 標本平均の確率的な出方(標本平均の標本分布)が左右対称のベル型の分布になる。
 - 標本平均の標本分布が正規分布(よく知られている理論分布)によって近似できる。
 - 中心極限定理とよばれる。
- 3つの性質を総合して:
 - 標本平均の確率的な出方(標本平均の標本分布)は、期待値が母平均、分散が 母分散/ n の正規分布で近似できる。

標本誤差の見積もり(12)

- 母分散とは
 - 母集団における変数のバラツキの尺度
 - 図4のデータの場合
 - 母分散 = $\{(13-15.4)^2 + (15-15.4)^2 + (21-15.4)^2 + (17-15.4)^2 + (11-15.4)^2\}/5$
= 11.84
 - 標本分散によって推定できる。
 - 図4のデータで標本(1,2)を使った場合:
 - 標本分散 = $\{(13-14)^2 + (15-14)^2\}/2 = 1$
 - » いくつかの推定方法が考案されている。

標本誤差の見積もり(13)

- 正規分布の性質の利用
 - 「標本平均が
母平均 $-2(\text{母分散}/n)^{1/2}$, 母平均 $+2(\text{母分散}/n)^{1/2}$
の範囲に出現する」確率は約0.95となる。
 - 「母平均が
標本平均 $-2(\text{母分散}/n)^{1/2}$, 標本平均 $+2(\text{母分散}/n)^{1/2}$
の範囲に挟まれる」確率は約0.95となる。
 - 「母平均が
標本平均 $-2(\text{標本分散}/n)^{1/2}$, 標本平均 $+2(\text{標本分散}/n)^{1/2}$
の範囲に挟まれる」確率は約0.95となる。

標本誤差の見積もり(14)

- 図4のデータで計算してみると...

表1: 図4のデータによる母平均の区間推定

標本の候補	標本情報		標本平均	標本分散	下限	上限	捕捉？
(1,2)	13	15	14	1	12.6	15.4	○
(1,3)	13	21	17	16	11.3	22.7	○
(1,4)	13	17	15	4	12.2	17.8	○
(1,5)	13	11	12	1	10.6	13.4	×
(2,3)	15	21	18	9	13.6	22.2	○
(2,4)	15	17	16	1	14.6	17.4	○
(2,5)	15	11	13	4	10.2	15.8	○
(3,4)	21	17	19	4	16.2	21.8	×
(3,5)	21	11	16	25	8.9	23.1	○
(4,5)	17	11	14	9	9.6	18.2	○

単純無作為抽出法：例

- MS Excel のファイル
 - 農林水産省「作物統計調査」
 - 北海道の水稲