

統計学II

早稲田大学政治経済学術院

西郷 浩

本日の目標

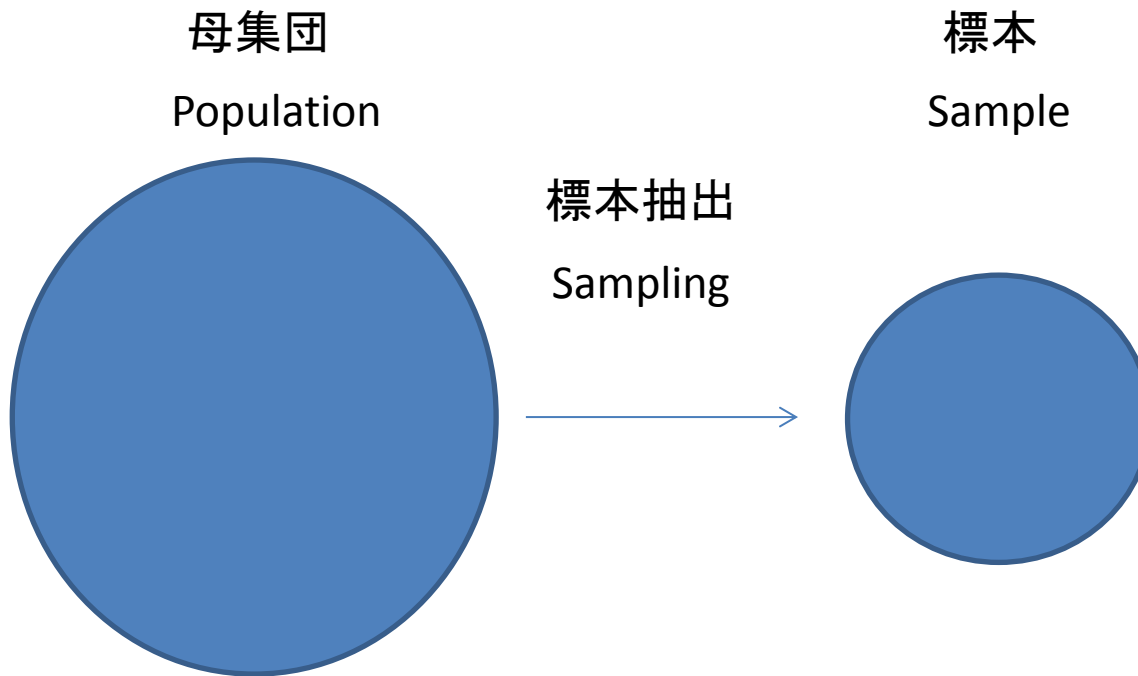
- 標本抽出
 - 母集団と標本
 - 確率標本抽出
 - 母数と統計量
 - 統計量の標本分布

母集団と標本(1)

- 母集団
 - 関心のある対象の全体
 - 有限母集団の例: ある選挙における有権者全体
 - 無限母集団の例: 「1つのサイコロを投げる」という試行を無限回繰り返す。
 - 「確率変数を発生させる仕組み」を母集団とする定義もある。
- 標本
 - 母集団の一部
 - 有限母集団から抽出される標本の例:
 - ある選挙における有権者の一部
 - 無限母集団から抽出される標本の例
 - 「1つのサイコロを投げる」という試行を10回繰り返す。
 - 「確率変数を発生させる仕組み」から発生したデータを標本とする定義もある。

母集団と標本(2)

- 標本抽出
 - 母集団から標本を抽出すること。



標本抽出法の区分

- 確率標本抽出法

- 標本の出現確率があらかじめ決められている標本抽出方法

- 例：単純無作為抽出法

- 推測統計学では確率標本抽出法を前提とする。

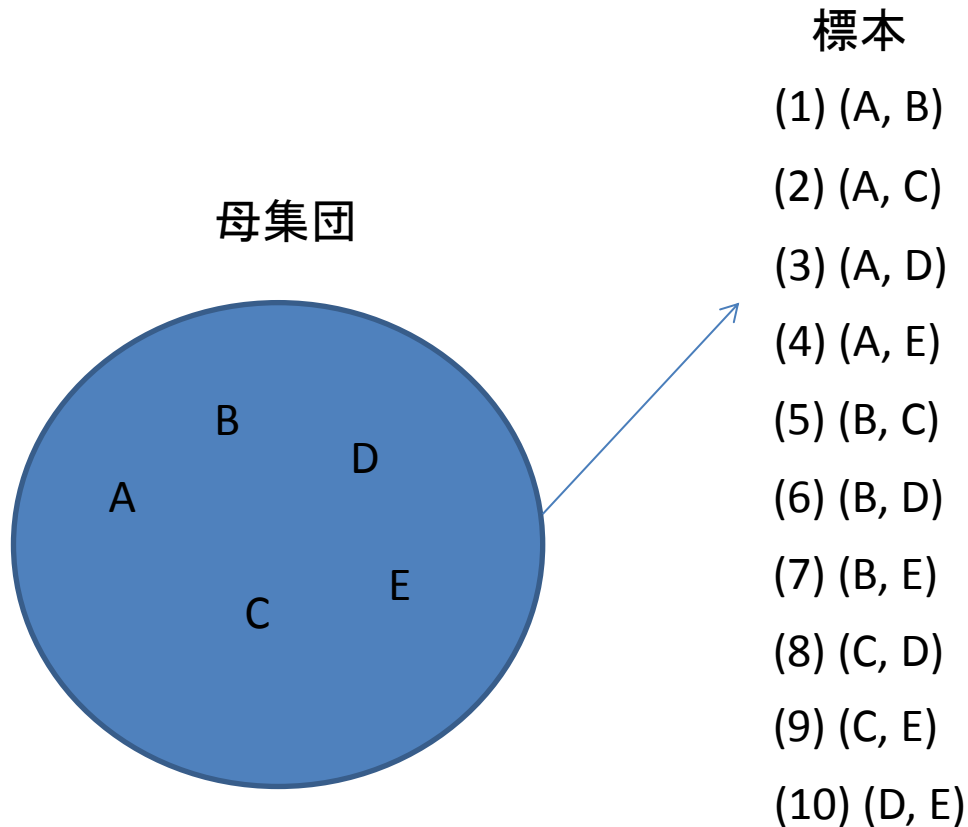
- 有意標本抽出法

- 標本の出現確率があらかじめ定められていない。

- 例：有識者による典型的な対象の選択。

単純無作為抽出法(1)

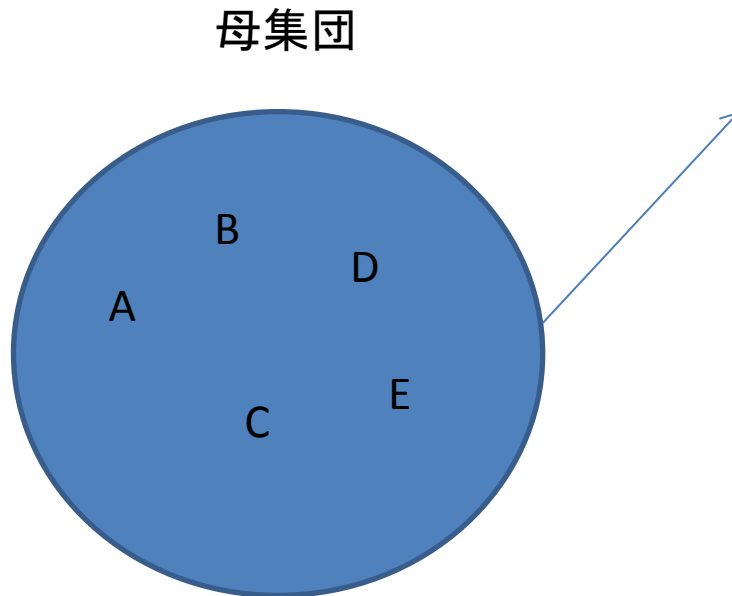
- 重複を認めない大きさ2の標本



単純無作為抽出法(2)

- 単純無作為抽出法

- 非復元抽出法で実行できる。



標本	確率
(1) (A, B)	1/10
(2) (A, C)	1/10
(3) (A, D)	1/10
(4) (A, E)	1/10
(5) (B, C)	1/10
(6) (B, D)	1/10
(7) (B, E)	1/10
(8) (C, D)	1/10
(9) (C, E)	1/10
(10) (D, E)	1/10

単純無作為抽出法(3)

表1: 非復元抽出のもとでの無作為抽出

1 × 2	A	B	C	D	E
A	0	$1/5 \times 1/4$	$1/5 \times 1/4$	$1/5 \times 1/4$	$1/5 \times 1/4$
B	$1/5 \times 1/4$	0	$1/5 \times 1/4$	$1/5 \times 1/4$	$1/5 \times 1/4$
C	$1/5 \times 1/4$	$1/5 \times 1/4$	0	$1/5 \times 1/4$	$1/5 \times 1/4$
D	$1/5 \times 1/4$	$1/5 \times 1/4$	$1/5 \times 1/4$	0	$1/5 \times 1/4$
E	$1/5 \times 1/4$	$1/5 \times 1/4$	$1/5 \times 1/4$	$1/5 \times 1/4$	0

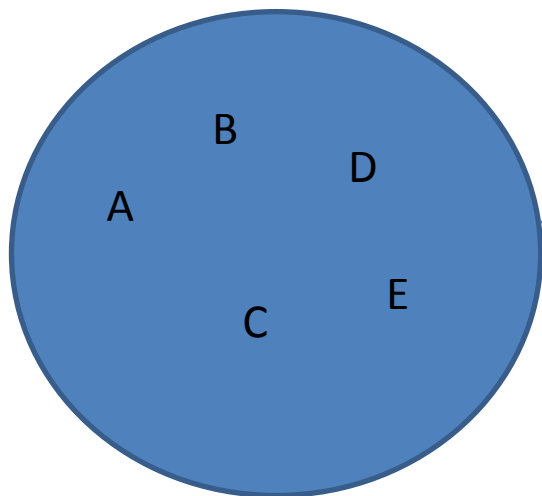
1回目で、A, B, C, D, E のうちの1つを等確率で抜き取り、2回目で、残りの4つのうち1つを等確率で抜き取る。

復元抽出法の場合(1)

- 復元抽出法

- 無限母集団からの抽出に対応

母集団



抜き取り	確率	標本	確率
(A, A)	1/25	→ (1) (A, A)	1/25
(A, B)	1/25	→ (2) (A, B)	2/25
(A, C)	1/25		
(A, D)	1/25		
(A, E)	1/25	...	
(B, A)	1/25		
(B, B)	1/25		
(B, C)	1/25	(15) (E, E)	1/25
...			
(E, E)	1/25		

復元抽出法の場合(2)

表2: 復元抽出のもとでの無作為抽出

1 × 2	A	B	C	D	E
A	$1/5 \times 1/5$	$1/5 \times 1/5$	$1/5 \times 1/5$	$1/5 \times 1/5$	$1/5 \times 1/5$
B	$1/5 \times 1/5$	$1/5 \times 1/5$	$1/5 \times 1/5$	$1/5 \times 1/5$	$1/5 \times 1/5$
C	$1/5 \times 1/5$	$1/5 \times 1/5$	$1/5 \times 1/5$	$1/5 \times 1/5$	$1/5 \times 1/5$
D	$1/5 \times 1/5$	$1/5 \times 1/5$	$1/5 \times 1/5$	$1/5 \times 1/5$	$1/5 \times 1/5$
E	$1/5 \times 1/5$	$1/5 \times 1/5$	$1/5 \times 1/5$	$1/5 \times 1/5$	$1/5 \times 1/5$

おのおのの回で、A, B, C, D, E のうちの1つを等確率で抜き取る。

母数と統計量(1)

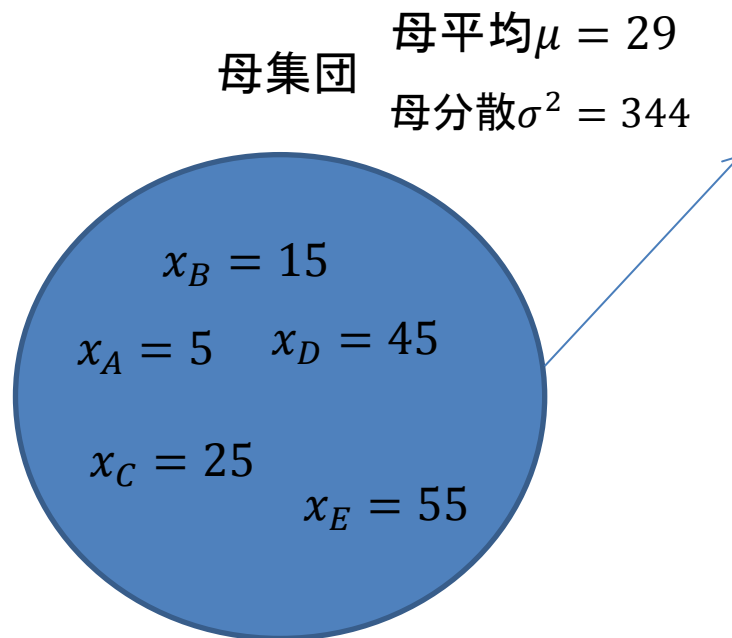
- 母数
 - 母集団情報を使って計算した値
 - 推測統計では、未知であることが前提となる。
 - 例：
 - 母平均：母集団の平均値
 - 母分散：母集団の分散

母数と統計量(2)

- 統計量
 - 標本情報から計算した値
 - 推測統計では、既知であることが前提となる。
 - 例：
 - 標本平均：標本の平均
 - 標本分散：標本の分散
 - 計算式が複数あることに注意する。

統計量の標本分布(1)

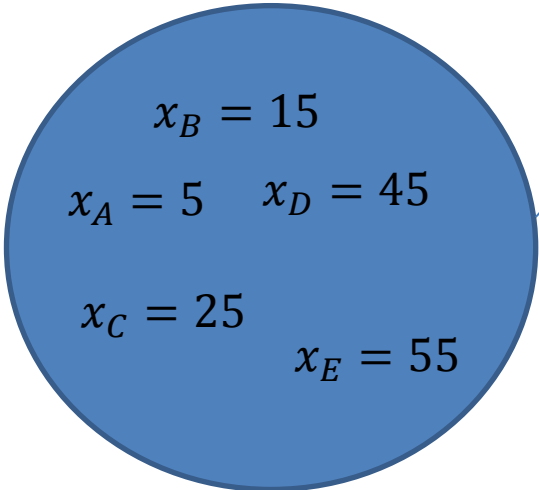
- 単純無作為抽出法



標本	標本平均
(1) (A, B)	10
(2) (A, C)	15
(3) (A, D)	25
(4) (A, E)	30
(5) (B, C)	20
(6) (B, D)	30
(7) (B, E)	35
(8) (C, D)	35
(9) (C, E)	40
(10) (D, E)	50

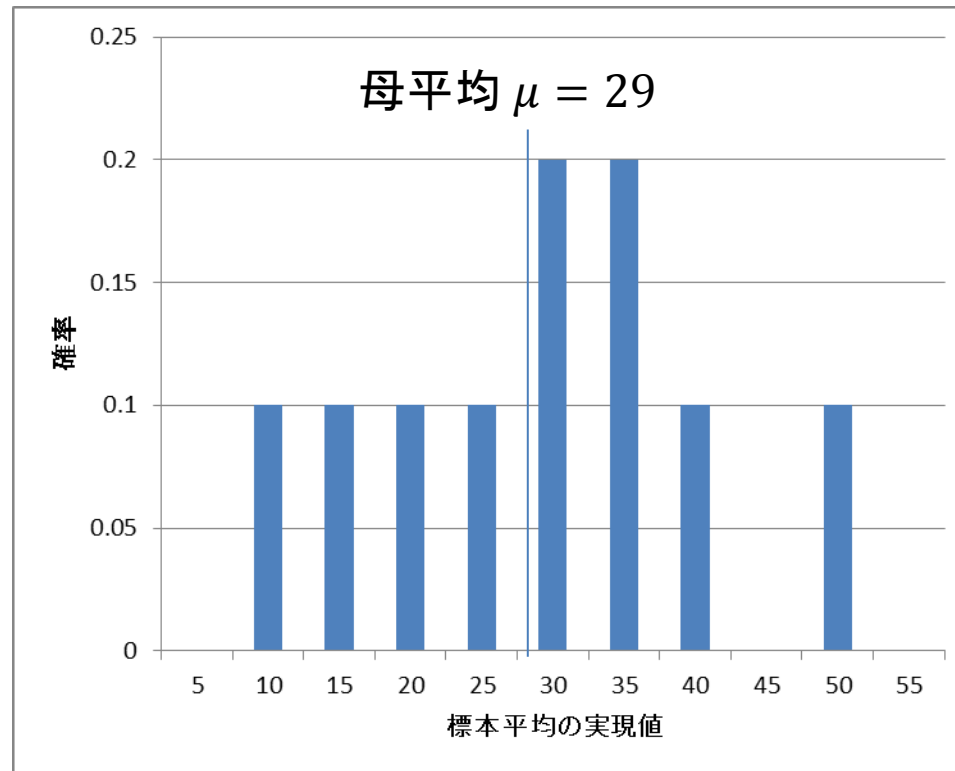
統計量の標本分布(2)

- 単純無作為抽出法

母集団		母平均 $\mu = 29$	母分散 $\sigma^2 = 344$	標本	標本平均	確率
				(1) (A, B)	10	1/10
				(2) (A, C)	15	1/10
				(3) (A, D)	25	1/10
				(4) (A, E)	30	1/10
				(5) (B, C)	20	1/10
				(6) (B, D)	30	1/10
				(7) (B, E)	35	1/10
				(8) (C, D)	35	1/10
				(9) (C, E)	40	1/10
				(10) (D, E)	50	1/10

統計量の標本分布(3)

図1: 単純無作為抽出法のもとでの標本平均の標本分布



統計量の標本分布 (4)

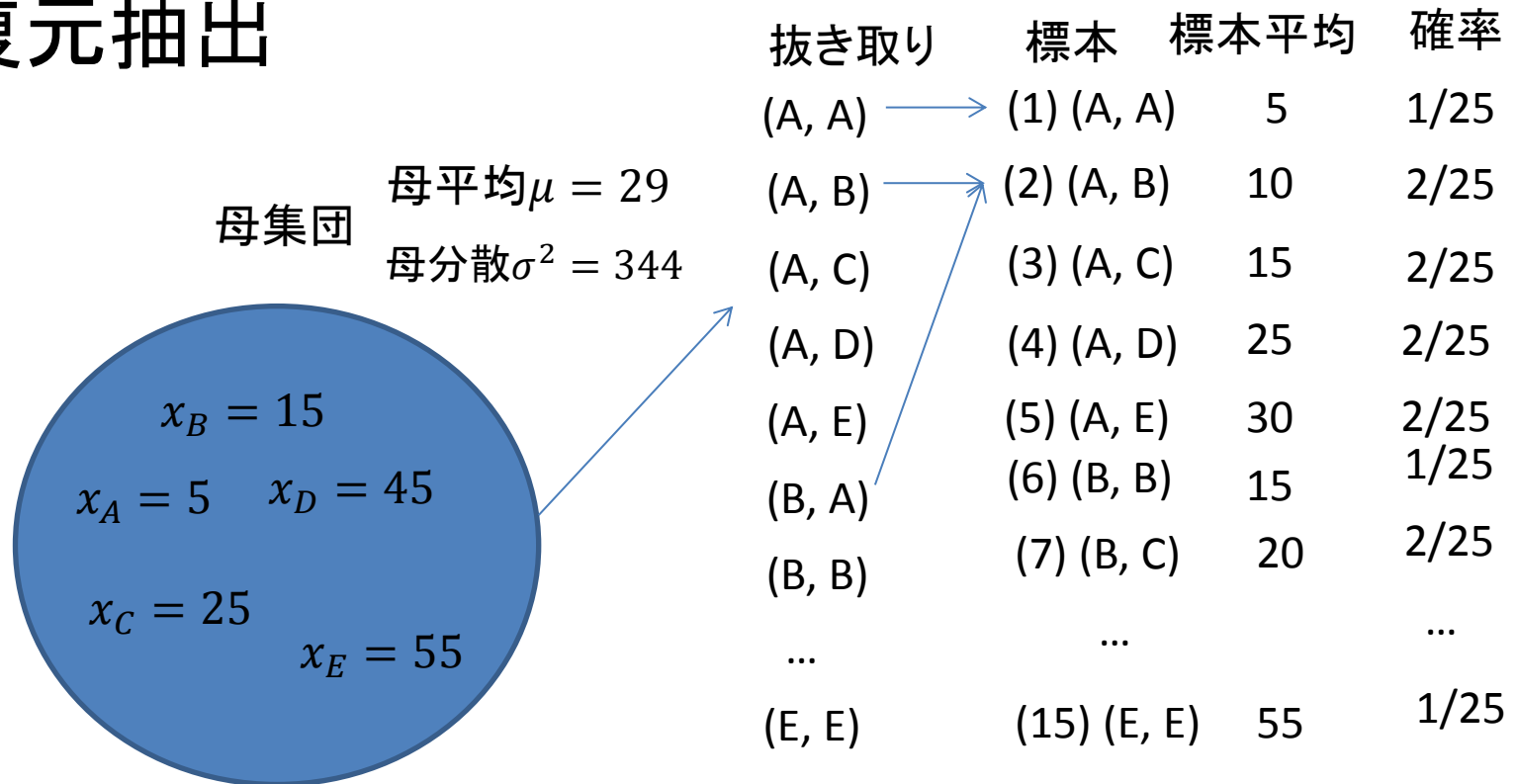
- 確率変数としての標本平均の期待値と分散

$$\begin{aligned} - E(\bar{X}) &= \frac{1}{10} \times 10 + \frac{1}{10} \times 15 + \frac{1}{10} \times 20 + \frac{1}{10} \times 25 \\ &\quad + \frac{2}{10} \times 30 + \frac{2}{10} \times 35 + \frac{1}{10} \times 40 + \frac{1}{10} \times 50 \\ &= 29 (= \mu) \end{aligned}$$

$$\begin{aligned} - V(\bar{X}) &= \frac{1}{10} \times (10 - 29)^2 + \frac{1}{10} \times (15 - 29)^2 + \frac{1}{10} \times (20 - 29)^2 \\ &\quad + \frac{1}{10} \times (25 - 29)^2 + \frac{2}{10} \times (30 - 29)^2 + \frac{2}{10} \times (35 - 29)^2 \\ &\quad + \frac{1}{10} \times (40 - 29)^2 + \frac{1}{10} \times (50 - 29)^2 \\ &= 129 \left(= \frac{\sigma^2}{n} \frac{N - n}{N - 1} \right) \end{aligned}$$

統計量の標本分布：復元抽出(1)

- 復元抽出



統計量の標本分布：復元抽出(2)

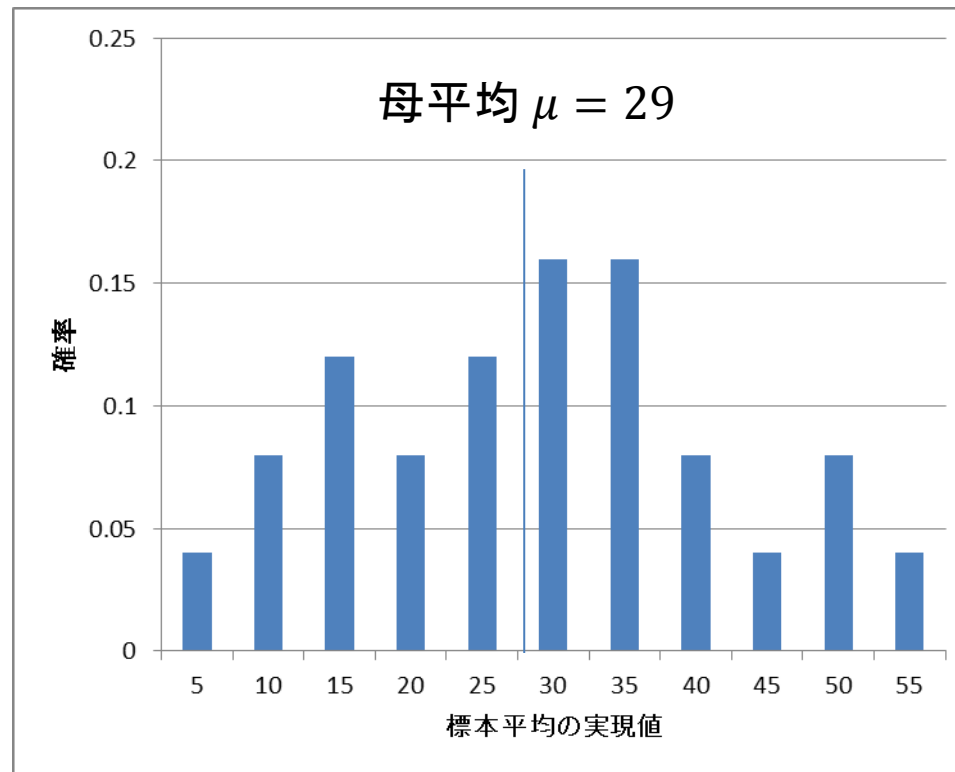
表3: 復元抽出のもとでの標本平均

1 × 2	A 5	B 15	C 25	D 45	E 55
A 5	5	10	15	25	30
B 15	10	15	20	30	35
C 25	15	20	25	35	40
D 45	25	30	35	45	50
E 55	30	35	40	50	55

どの実現値も、母平均 $\mu = 29$ と等しくない。

統計量の標本分布：復元抽出(3)

図2：復元抽出のもとでの標本平均の標本分布



統計量の標本分布：復元抽出(4)

- 確率変数としての標本平均の期待値と分散

$$\begin{aligned} - E(\bar{X}) &= \frac{1}{25} \times 5 + \frac{2}{25} \times 10 + \frac{3}{25} \times 15 + \frac{2}{25} \times 20 + \frac{3}{25} \times 25 + \frac{4}{25} \times 30 \\ &\quad + \frac{4}{25} \times 35 + \frac{2}{25} \times 40 + \frac{1}{25} \times 45 + \frac{2}{25} \times 50 + \frac{1}{25} \times 55 \\ &= 29 (= \mu) \end{aligned}$$

$$\begin{aligned} - V(\bar{X}) &= \frac{1}{25} \times (5 - 29)^2 + \frac{2}{25} \times (10 - 29)^2 + \frac{3}{25} \times (15 - 29)^2 \\ &\quad + \frac{2}{25} \times (20 - 29)^2 + \frac{3}{25} \times (25 - 29)^2 + \frac{4}{25} \times (30 - 29)^2 \\ &\quad + \frac{4}{25} \times (35 - 29)^2 + \frac{2}{25} \times (40 - 29)^2 + \frac{1}{25} \times (45 - 29)^2 \\ &\quad + \frac{2}{25} \times (50 - 29)^2 + \frac{1}{25} \times (55 - 29)^2 \\ &= 172 \left(= \frac{\sigma^2}{n} \right) \end{aligned}$$

統計量の標本分布：復元抽出(5)

- もうひとつの見方

- i 回目の抜き取りの結果をあらわす確率変数

- $X_i = \begin{cases} 5 & (1/5) \\ 15 & (1/5) \\ 25 & (1/5) \\ 45 & (1/5) \\ 55 & (1/5) \end{cases}, E(X_i) = 29, V(X_i) = 344.$

- X_1 と X_2 は相互に独立である。

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

- $E(\bar{X}) = E\left\{\frac{1}{2} \times (X_1 + X_2)\right\} = \frac{1}{2}E(X_1 + X_2) = 29 (= E(X_i) = \mu)$

- $V(\bar{X}) = V\left\{\frac{1}{2} \times (X_1 + X_2)\right\} = \left(\frac{1}{2}\right)^2 V(X_1 + X_2) = \frac{344}{2} \left(= \frac{V(X_i)}{2} = \frac{\sigma^2}{2}\right)$

統計量の標本分布：復元抽出(6)

- 拡張

- $X_i \sim iid(\mu, \sigma^2), i = 1, 2, \dots, n$

- $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$

- $E(\bar{X}) = E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} E\left(\sum_{i=1}^n X_i\right) =$
 $\frac{1}{n} \sum_{i=1}^n E(X_i) = \mu$

- $V(\bar{X}) = V\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \left(\frac{1}{n}\right)^2 V\left(\sum_{i=1}^n X_i\right) =$
 $\frac{1}{n^2} \sum_{i=1}^n V(X_i) = \frac{\sigma^2}{n}$

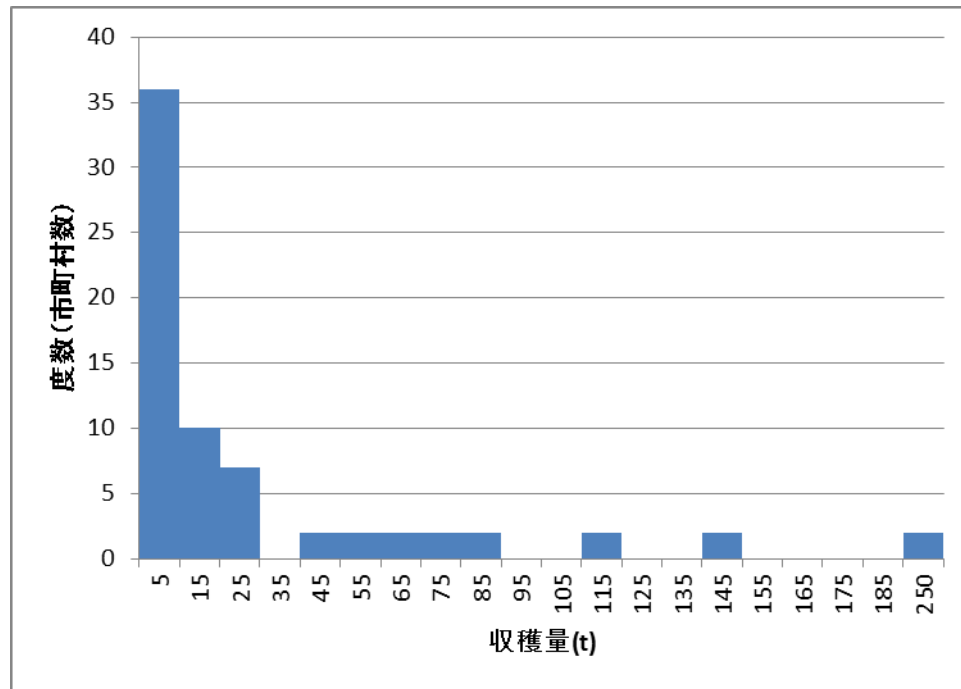
実験(1)

- シミュレーション
 - 母集団
 - 抽出単位: 長野県内のそばを生産する69市町村
 - 変数: 抽出単位内のそばの収穫量
 - 推定対象:
 - 長野県内のそばの収穫量の一市町村当たりの平均値
 - 69倍すれば長野県内のそばの収穫量の合計になる。
 - サンプルサイズ: $n = 10$ (市町村)
 - 抽出方法: 単純無作為抽出法

実験(2)

- 母集団の分布

図3: 母集団における収穫量の分布

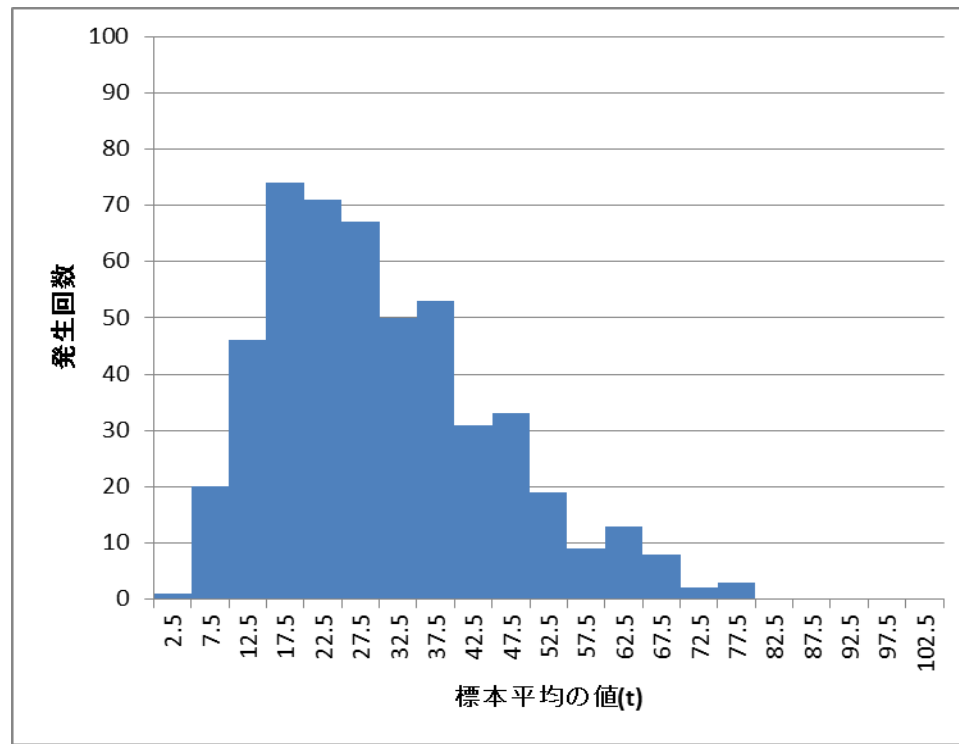


資料: 農林水産省「作物統計調査」平成29年調査

実験(3)

- 標本平均の標本分布

図4: 標本平均の標本分布



注: 図3からの抽出実験に基づく結果

実験(4)

- 実験結果の概観
 - 母集団の分布と標本平均の標本分布
 - 分布の形がだいぶ異なる。
 - 標本平均の標本分布には、母集団とは異なる支配原理が作用しているようだ。
 - » 次回講義で詳しく調べる。