

# 統計（医療統計）

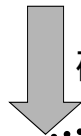
## 第8回 13章 母平均の区間推定

授業担当：徳永伸一

東京医科歯科大学教養部 数学講座

### Overview

- 確率（9章）
- 記述統計（10章） . . . . 情報の要約
  - 表やグラフで表す
  - 代表値（平均など）や散布度（分散など）を求める



確率モデル(11章)

- 推測統計(13章～)
  - 推定(点推定、区間推定) (13章)
  - 仮説検定(14章)

## 第11章 確率変数と確率分布

### I. 確率変数と確率分布の定義

- 1-確率変数の定義
  - ・・・ 離散型と連続型
- 2-離散型確率変数の確率分布
- 3-連続型確率変数の確率分布
  - ・・・ 確率密度関数

### II. 確率変数の特性値

- 1-期待値と分散・標準偏差の定義
- 2-確率変数の期待値と分散の性質
  - ・・・ 期待値の加法性
- 3-確率変数の標準化

## 第11章 確率変数と確率分布

### III. 確率変数の独立性

1. 確率変数の独立性の定義
2. 独立な確率変数の性質

### IV. 代表的な確率分布

- 2項分布, 正規分布など
- 正規分布の線形変換・標準化・再生性

### V. 中心極限定理と正規近似

1. 中心極限定理
2. 2項分布の正規近似・・・半整数補正

### VI. 標本分布

- 標本平均の分布

## 第13章 推定

- I. 母集団と標本
- II. 点推定
  - 不偏性, 不偏推定量
- III. 区間推定
- IV. 母平均の区間推定
  - 1. 母分散が既知のとき ←前期ここまで
  - 2. 母分散が未知のとき
- V. 母分散の区間推定
- VI. 母比率の区間推定 (←前期に先取り)

5

S.TOKUNAGA

[復習 (11章)] 標本平均の分布・まとめ (対比して再確認)

定理(正規分布の性質より)

$X_1, X_2, \dots, X_n$  を正規分布  $N(\mu, \sigma^2)$  に従う母集団から無作為抽出した標本とすると

$$\bar{X} \sim N(\mu, \sigma^2/n)$$

定理(中心極限定理の系)

$X_1, X_2, \dots, X_n$  を平均  $\mu$ , 分散  $\sigma^2$  である任意の母集団から無作為抽出した標本とすると,

標本サイズ  $n$  が十分大きければ, 近似的に

$$\bar{X} \sim N(\mu, \sigma^2/n)$$

となる.

★「3-その他の重要な標本分布」は後回し(13・14章にて).

6

## 19年度・20年度中間試験問題より

(19年度)

[3] (3) EXCELでRAND関数により10(列)×1000行の乱数表を作成し、さらに各行において10個の値の平均を求めた。この平均の値はどのような分布をとるか。理由と共にわかりやすい日本語で説明せよ。

(20年度)

[3] EXCELにおけるRAND関数は0以上1未満の値をランダムに取る。この値 $X$ を確率変数と見なしたとき、以下の問いに答えよ。

- (1)  $X$ が従う分布の確率密度関数 $f(x)$ はどのような関数となるか。適切に記述せよ。
- (2)  $X$ の期待値 $\mu$ と分散 $\sigma^2$ を、(連続型)確率変数の期待値と分散の定義に基づいて求めよ。
- (3) RAND関数で得た10個の値の平均を $Y$ とする。 $Y$ はどのような確率分布で近似できると考えられるか。理由と共に述べよ。

## 【重要】「標本数」と「標本サイズ」

「標本」とは、母集団から取り出したデータの値の「集合」である！

- 1回の標本抽出により、1つの標本が得られる。
- [標本に含まれるデータの値の個数]  
=「標本サイズ(標本の大きさ)」

★EXCEL実習3aの例では10列×1000行の値を標本サイズ10、1000標本(標本数1000)と見なして扱った。

★教科書p.105「取りだした標本の数を標本の大きさという」は不適切な記述。

## 第13章 推定

- I. 母集団と標本 ←前回ここから
- II. 点推定
  - 不偏性, 不偏推定量
- III. 区間推定
- IV. 母平均の区間推定
  - 1. 母分散が既知のとき ←前回ここまで
  - 2. 母分散が未知のとき
- V. 母分散の区間推定 . . . 省略
- VI. 母比率の区間推定

## I. 母集団と標本

### KEYWORDS :

- 母集団 ⇔ 標本 sample, 無作為標本
- 母平均, 母分散
- 層別, 層別抽出 stratified sampling
- 乱数

## [復習] II. 点推定 (1)

### KEYWORDS:

- 母数, 母集団パラメータ
  - 母平均, 母分散
- (標本)統計量, 統計値(統計量の実現値)
  - 標本平均, 標本分散, 標本比率
- 推定量(母数の推定に用いる統計量),  
推定値(推定量の実現値)

### 点推定と区間推定

- **点推定**:「無作為に選んだ標本から数値を得て, それから推定量の推定値を計算し, その推定値がイコール母数であるとする推定方法」
- **区間推定**:「推定値から, ”ある確率である数値の区間の中にある”とする推定方法」

11

S.TOKUNAGA

## [復習] II. 点推定 (2)

推定量が「不偏unbiasedである(偏りが無い)」とは:

- 「対応する母数より大きい(or小さい)値が得られやすい」といった傾向がない.
- その推定量を繰り返し実測し、得られた値(推定値)の平均値は、繰り返しの回数を増やすほど対応する母数に近づく.

厳密には:

- [定義]母数  $\theta$  の推定量  $\theta'$  に対し,

$$E(\theta') = \theta$$

のとき  $\theta'$  を  $\theta$  の不偏推定量(不偏性を持つ推定量)であるという.

12

S.TOKUNAGA

## [復習] II. 点推定 (3)

$X_1, X_2, \dots, X_n$  に対し

不偏分散  $U^2 := \{ \sum (X_k - \bar{X})^2 \} / (n-1)$

は, 母分散  $\sigma^2$  の不偏推定量.

▪ すなわち,  $E(U^2) = \sigma^2$  である (証明は割愛).

ということは

▪ 分散  $S^2 := \{ \sum (X_k - \bar{X})^2 \} / n$  については,

$$E(S^2) = E((n-1)/n U^2) = ((n-1)/n) \sigma^2$$

▪ つまり  $S^2$  の実測値は, 母分散より小さめの値をとる傾向にある (すなわち不偏でない).

## [復習] 不偏性に関する補足

▪ 標本平均  $\bar{X} := \sum X_k / n$  も母平均  $\mu$  の不偏推定量.

$$\because E(\bar{X}) = E(\sum X_k / n) = \sum E(X_k) / n = \mu$$

▪ たとえば  $n=3$  のとき  $X' = (X_1 + X_2 + 2X_3) / 4$  とおいても  $X'$  は  $\mu$  の不偏推定量 (確認せよ).

▪ 不偏分散の平方根  $U = \sqrt{U^2}$  は,  $\sigma$  の不偏推定量ではない!

$$\because E(U) = E(\sqrt{U^2}) \neq \sqrt{E(U^2)} = \sigma$$

## [復習] III. 区間推定

### 区間推定とは

- 「母数の値をズバリ推定するより, 母数が(高い確率で)存在する区間を推定する」という考え方.
- 「推定値が母数にどのくらい近い(誤差がどのくらいあるか)」も含めて推定する.

### 具体的には

- 「母数  $\theta$  が区間  $I$  に含まれる確率が  $O\%$ 」といった形の推定を行なう.
  - 「 $\theta$  の推定値  $\theta'$  の誤差が確率  $O\%$  で  $d$  以下」と考えても同じ.
  - $|\theta - \theta'| \leq d \Leftrightarrow \theta \in [\theta' - d, \theta' + d] (=I)$
- $\uparrow$  における
  - 区間 $\cdots(O\%)$ 信頼区間
  - 確率 $\cdots$ 信頼度(または信頼係数) $\cdots 95\%, 99\%$ など

## [復習] IV. 母平均の区間推定

### 母平均 $\mu$ の区間推定(母分散既知の場合)

問題設定:

- 標本サイズ  $n$ , 標本平均  $\bar{X}$  の実測値が与えられており, 母分散  $\sigma^2$  は既知とする.
- 母集団分布 $\cdots$ 正規分布なら好都合だが, 任意の分布でも  $n$  が大きければ, 標本平均の分布は中心極限定理により正規分布で近似可能.

以上の条件のもとで,

「 $\mu$  の  $100\gamma\%$  信頼区間」

を求める( $\gamma$  は信頼度 = 信頼係数で, 具体的には  $0.95, 0.99, 0.90$  など).

→ 続いて信頼区間の求め方, でもその前に $\cdots$



[再確認] VI. 標本分布 (2)

標本平均の期待値と分散・標準偏差

$X_1, X_2, \dots, X_n$  を平均  $\mu$ , 分散  $\sigma^2$  である母集団から無作為抽出した標本とすると、  
 $X_1, X_2, \dots, X_n$  はそれぞれ、期待値  $\mu$ , 分散  $\sigma^2$  の互いに独立な確率変数と見なせる。

よって標本平均  $\bar{X}$  について

$$E(\bar{X}) = \mu \times n \times (1/n) = \mu$$

$$V(\bar{X}) = \sigma^2 \times n \times (1/n)^2 = \sigma^2/n$$

(期待値・分散の加法性↑) (↑積に関するE,Vの性質より)

$$\sigma(\bar{X}) = \sqrt{V(\bar{X})} = \sqrt{(\sigma^2/n)} = \sigma / \sqrt{n}$$

S. TOKUNAGA

17

[再確認] 標本平均の分布・まとめ (対比して再確認)

定理(正規分布の性質より)

$X_1, X_2, \dots, X_n$  を正規分布  $N(\mu, \sigma^2)$  に従う母集団から無作為抽出した標本とすると

$$\bar{X} \sim N(\mu, \sigma^2/n)$$

定理(中心極限定理の系)

$X_1, X_2, \dots, X_n$  を平均  $\mu$ , 分散  $\sigma^2$  である任意の母集団から無作為抽出した標本とすると、

標本サイズ  $n$  が十分大きければ、近似的に

$$\bar{X} \sim N(\mu, \sigma^2/n)$$

となる。

★以上を踏まえて区間推定の具体的な方法へ

18

## [復習] IV. 母平均の区間推定

### 母平均 $\mu$ の区間推定(母分散既知)の解法

- $\bar{X} \sim N(\mu, \sigma^2/n)$  と近似(中心極限定理による).
  - 注意: 正規母集団を仮定すれば厳密.
- $Z = (\bar{X} - \mu) / (\sigma / \sqrt{n})$  と標準化すると  $Z \sim N(0, 1)$
- $P(-z(\alpha/2) \leq Z \leq z(\alpha/2)) = \gamma$ 
  - ただし  $\gamma = 1 - \alpha$ .  $z(\alpha)$  は  $P(Z \geq z(\alpha)) = \alpha$  を満たす値.
  - たとえば  $\gamma = 0.95$  のとき  $\alpha/2 = 0.025$ ,  $z(\alpha/2) = z(0.025) = 1.96$
  - $z(\alpha/2)$  は「上側  $100(\alpha/2)\%$  点」と呼ばれる.
- $\uparrow$  を同値変形すると

$$P(\bar{X} - z(\alpha/2) \sigma / \sqrt{n} \leq \mu \leq \bar{X} + z(\alpha/2) \sigma / \sqrt{n}) = \gamma$$
よって「 $\mu$  の  $(100 \times \gamma)\%$  信頼区間」は

$$\left( \bar{X} - z(\alpha/2) \sigma / \sqrt{n}, \bar{X} + z(\alpha/2) \sigma / \sqrt{n} \right)$$

19

S.TOKUNAGA

## [復習] IV. 母平均の区間推定

### 教科書 p.107 例題1

コレステロールの平均値  $\mu$  の区間推定

- 母集団・・・成人男子  
(正規分布は p.106 で仮定)
  - 標本サイズ  $n = 36$  (人)
  - 標本平均  $\bar{X} = 63$  (mg/dl)・・・ $\mu$  の点推定値
  - 母標準偏差  $\sigma = 12$  (mg/dl)
- 以上の条件のもとで,  
「 $\mu$  の 95% 信頼区間を求めよ」という問題.  
(信頼度 95%)

20

S.TOKUNAGA

## [復習 (追記あり)] IV. 母平均の区間推定

### 例題1に関する補足

- 正規母集団の仮定がどこにも明記されていなければ、下記のいずれかの方針で考える。
  - 正規母集団に十分近い分布であると考えて、正規母集団の仮定を導入。
  - $n=36$ を十分大きいと見なし、標本平均の分布を正規分布で近似(中心極限定理より)。
  - 教科書では13章「1」の末尾(p.106)において正規母集団が仮定されている。  
★ただしカッコ内の説明はおかしいので無視してください。  
「母集団が小さければ正規母集団を仮定できない」は正しいが、  
「母集団が大きければ正規母集団を仮定できる」とは限らない。
- 公式  $(\bar{X} - z(\alpha/2)\sigma/\sqrt{n}, \bar{X} + z(\alpha/2)\sigma/\sqrt{n})$   
に値を代入すれば一応答えは出ます。
- 母標準偏差  $\sigma$  の値が既知というのは、かなり虫のいい仮定。
  - 現実的にはあまりないケースと思われるが、基本的な原理と手法を理解するためにあえて導入している。

21

S.TOKUNAGA

## [復習] IV. 母平均の区間推定

「 $\mu$ の $(100 \times \gamma)\%$ 信頼区間」:

$$(\bar{X} - z(\alpha/2)\sigma/\sqrt{n}, \bar{X} + z(\alpha/2)\sigma/\sqrt{n})$$

について:

- 標本平均(の実測値)を中心とする区間である。
- $\sigma/\sqrt{n}$  は標準誤差と呼ばれる。

[その他の重要な考察]

- $\gamma$ を大きくすると・・・
  - $\alpha = 1 - \gamma$ は小さくなる  $\rightarrow z(\alpha/2)$  は大きくなる。  
 $\rightarrow$ 信頼区間の幅が大きくなる。  
(外れる確率を減らすのだから、幅を大きく取る必要があるのは当然)
- $n$ を大きくすると  $\rightarrow$  信頼区間の幅が小さくなる。
  - 情報量が増えるのだから、誤差が減るのは当然

22

S.TOKUNAGA

## [復習] IV. 母平均の区間推定

「母分散既知の場合」のポイントをまとめると

- 母分散  $\sigma^2$  を用いて標本平均の分布が表せる.
  - 母集団分布が正規分布なら標本平均の分布も正規分布.
  - 正規母集団を仮定せずとも, 標本サイズ  $n$  が十分大きければ標本平均の分布は正規分布で近似でき, いずれにしても正規分布の問題に帰着できる.
  - だがその(標本平均が従う)正規分布  $N(\mu, \sigma^2/n)$  は, 母分散  $\sigma^2$  を用いて表されているのだから,  $\sigma^2$  の値がわからないと推測できない.
  - 現実には  $\sigma^2$  は未知のケースが多い!
- 「母分散未知」のケースへ.

23

S.  
TOKUN  
AGA

## IV. 母平均の区間推定 (母分散未知)

母平均  $\mu$  の区間推定 (母分散未知の場合)

母分散  $\sigma^2$  が未知であっても

標本の不偏分散  $U^2 := \{ \sum (x_i - \bar{x})^2 \} / (n-1)$

は, 常にわかる (標本データから計算できる).

そこで母分散未知の場合には:

- 方針1:  $U^2$  を  $\sigma^2$  の近似値 (推定値) として利用.
  - $U^2$  は  $\sigma^2$  の (不偏) 推定量ですからね.
  - 「近似」を認めてしまえば, またしても正規分布の問題に帰着. 推定の方法はほとんど同じ.  
( $U$  の値を  $\sigma$  のところに代入するだけ. 同じ公式が使える)
  - 標本サイズが大きければ, 良い (誤差の少ない) 近似値であることが期待できる.
  - だが **小標本の場合** は誤差が無視できないはず.  
→ 方針2へ

24

S.  
TOKUN  
AGA

#### IV. 母平均の区間推定 (母分散未知)

\*\*\*ここからまったく新しい内容\*\*\*

- 方針2:  $\sigma^2$ を用いずに表せる( $U^2$ を含む)統計量を導入する.  
→第11章 V

3-その他の重要な標本分布[2] t分布

以下の定理を利用:

定理(教科書p.100)

正規母集団 $N(\mu, \sigma^2)$ からの, 大きさ  $n$  の無作為標本  $\{X_1, X_2, \dots, X_n\}$  について, 標本平均  $X'$ , 不偏分散を  $U^2$  とする. このとき, 確率変数

$$T := (X' - \mu) / (U / \sqrt{n})$$

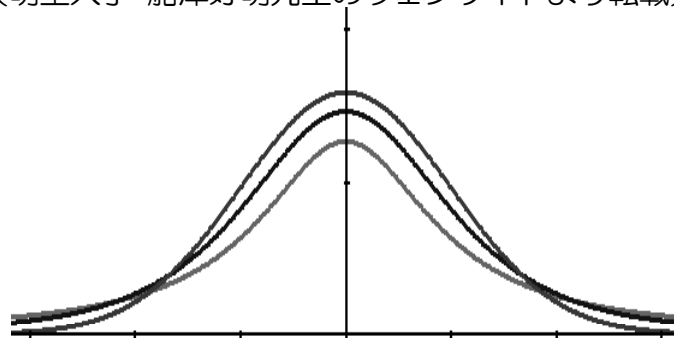
は自由度  $n-1$  の t 分布 (と呼ばれる分布) に従う.

25

S.  
TOKUN  
AGA

#### t 分布の密度関数のグラフ

(明星大学 船津好明先生のウェブサイトより転載)



青: 自由度3, 暗緑: 自由度1 (赤: 標準正規分布)

26

S.  
TOKUN  
AGA

## IV. 母平均の区間推定（母分散未知）

### t分布について

- 「スチューデントのt分布」とも呼ばれる。
  - 「スチューデントstudent」は発案者のペンネーム。
  - 定理中の $T = (X' - \mu) / (U / \sqrt{n})$  はスチューデント比とも呼ばれる。
- 厳密に定義するのは少々やっかいなので省略。ともかく理論的に「わかっている」分布として(巻末の数表などを用いて)利用する。
- 密度関数の概形は正規分布と似ている。対称性のある釣鐘型。
- 自由度と呼ばれるパラメータを持つ。
  - サイズ $n$ の標本から求めたスチューデント比 $T$ においては、自由度 $\nu$  (ギリシャ文字の「ニュー」) $= n - 1$ 。
  - 自由度とは大雑把に言えば「独立なパラメータの個数」。
  - $\nu$  大きいほど標準正規分布 $N(0,1)$ に近づき、 $\nu = \infty$ で完全に一致。
  - 自由度 $\nu$ のt分布の上側 $100\alpha\%$ 点を $t_{\nu}(\alpha)$ または $t(\nu, \alpha)$ で表す。
  - $t_{\nu}(\alpha)$ の値は教科書p.137の表から。

27

S.  
TOKUN  
AGA

## IV. 母平均の区間推定（母分散未知）

### 注意と補足

- 「正規母集団」の仮定はt分布を利用する上で本質的に重要。
- 標本平均 $X'$ の標準化変数:  
 $Z = (X' - \mu) / (\sigma / \sqrt{n})$ との類似に注意。
  - つまり $n$ が大きければ、 $T \doteq Z$ であるということ。  
(ただし、今考えているのは $n$ が小さいケースでした)

28

S.  
TOKUN  
AGA

## IV. 母平均の区間推定 (母分散未知)

### t分布を利用した $\mu$ の推定

- 母分散 $\sigma^2$ が未知の正規母集団 $N(\mu, \sigma^2)$ からの、大きさ $n$ の無作為標本 $\{x_1, x_2, \dots, x_n\}$ について、標本平均 $\bar{x}$ 、不偏分散を $U^2$ とする。

このとき

$$t = (\bar{x} - \mu) / (U / \sqrt{n}) \sim (\text{自由度 } n-1 \text{ の } t \text{ 分布})$$

だから

- 母平均 $\mu$ の信頼度 $\gamma = 1 - \alpha$ の信頼区間は

$$\left( \bar{x} - t_{n-1}(\alpha/2) U / \sqrt{n}, \bar{x} + t_{n-1}(\alpha/2) U / \sqrt{n} \right)$$

【注意】やっけることは母分散既知の場合と大差ない。自由度という新たなパラメータが出てきただけ。

[RECALL] 母分散既知の場合の $\mu$ の $100(1-\alpha)\%$ 信頼区間:

$$\left( \bar{X} - z(\alpha/2) \sigma / \sqrt{n}, \bar{X} + z(\alpha/2) \sigma / \sqrt{n} \right)$$

29

S.  
TOKUN  
AGA

## IV. 母平均の区間推定 (母分散未知)

### 例題2(教科書p.108)

- 母集団: 成人女性HDLコレステロール値  
(正規母集団はp.106で仮定)
- 標本サイズ $n=10$ , 標本平均 $\bar{x}' = 70$
- 標本の不偏分散 $U^2 = 14 \times 14 \times (10/9) \dots \star$   
(「標準偏差」の定義に注意!)
- 信頼度=0.95

[解答]

自由度 $=n-1=10-1=9$

教科書p.137の表より $t_9(0.025) = 2.262$ を読み取る。

95%信頼区間は

$$\left( \bar{x}' - t_{n-1}(0.025) U / \sqrt{n}, \bar{x}' + t_{n-1}(0.025) U / \sqrt{n} \right)$$

に該当する値を代入して

$$= ( \quad , \quad )$$

30

S.  
TOKUN  
AGA

## VI. 母比率の区間推定

### 「母集団比率に関する推測」とは

- ベルヌーイ母集団に関する推測.
  - 2項分布の応用.
  - ある程度大きな標本を扱うケースがほとんどなので、たいていは「2項分布の正規近似」を利用する.
- 「世論調査の類」. 支持率調査など、身近に興味深い例が多い.
  - 「統計的な理解を深めるよいチャンス」.

31

S.  
TOKUN  
AGA

## VI. 母比率の区間推定

### 2項分布の正規近似RECALL

中心極限定理により、 $n$ が十分大きいとき、 $B(n,p)$ は $N(np, np(1-p))$ で近似できる。

- ∴  $B(n,p)$ に従う確率変数は、 $B(1,p)$ (という同一の分布)に従う独立な $n$ 個の確率変数の和と見なせるから。
- 従って、標本比率 $P=X/n$ の分布も、正規分布  $N(p, p(1-p)/n)$  で近似できる.
  - さらに $P$ の標準化変数:
$$Z = (P-p) / \sqrt{p(1-p)/n}$$
は近似的に標準正規分布に従う.
  - 分布が決まれば、あとはこれまでと同じ考え方で進めればよいはず(?)

32

S.  
TOKUN  
AGA



## VI. 母比率の区間推定

とりあえずやってみる.

(以下母比率を $p$ , 標本比率 $P=X/n$ の実現値を $P_0$ とする)

$$Z = (P-p) / \sqrt{p(1-p)/n}$$

が近似的に標準正規分布に従うので

$P(-z(\alpha/2) \leq (P_0-p) / \sqrt{p(1-p)/n} \leq z(\alpha/2)) = 1-\alpha$   
左辺のカッコ内を同値変形すると

$$P_0 - z(\alpha/2)\sqrt{p(1-p)/n} \leq p \leq P_0 + z(\alpha/2)\sqrt{p(1-p)/n}$$

すなわち、区間:

$$(P_0 - z(\alpha/2)\sqrt{p(1-p)/n}, P_0 + z(\alpha/2)\sqrt{p(1-p)/n}) \quad \dots (*)$$

に母比率 $p$ が含まれる確率が $1-\alpha$ .

ところが(\*)は未知数である $p$ そのものを含んでいるので、これをそのまま信頼区間とすることはできない!

そこで...

## VI. 母比率の区間推定

**推定**を行う際は,

$$(P_0 - z(\alpha/2)\sqrt{p(1-p)/n}, P_0 + z(\alpha/2)\sqrt{p(1-p)/n})$$

において $p$ を近似値(推定値) $P_0$ で置き換える.

すなわち信頼度 $\gamma = 1 - \alpha$ の信頼区間は:

$$(P_0 - z(\alpha/2)\sqrt{P_0(1-P_0)/n}, P_0 + z(\alpha/2)\sqrt{P_0(1-P_0)/n})$$

★ただし、誤差の最大値を見積もりたいときは $p=0.5$ を採用.

- $p(1-p)$ は $p=0.5$ のとき最大値0.25を取ることに注意  
(2次関数のグラフを思い出せ!).
- 「誤差の最大値を見積もりたいとき」の例:  
→誤差が一定値以下となるような標本サイズを決定する問題など.  
(標本サイズを決定する時点では $P_0$ の値は得られていない!)

★(14章でやる)「母比率の検定」との違いに注意!

## 第13章 推定

- I. 母集団と標本
- II. 点推定
  - 不偏性, 不偏推定量
- III. 区間推定
- IV. 母平均の区間推定
  - 1. 母分散が既知のとき・・・今回ここから
  - 2. 母分散が未知のとき
- V. 母分散の区間推定・・・省略
- VI. 母比率の区間推定
  - \*\*\*今回ここまで\*\*\*