# 知能システム演習A
# ～統計−推定～

11月11日（月2）

B4　福田至

1. 母集団
2. 点推定
3. 区間推定

# 母集団と標本

- 統計的な調査には対象全体の資料を集めて調べる**全数調査**という方法がある．これに対して，調査の対象全体からその一部を抜き出して調べる**標本調査**という方法がある．
- 標本調査の場合，調査の対象全体を**母集団**という．母集団に属する対象を**個体**という．また，調査のために母集団から抜き出された個体の集合を**標本**といい，母集団から標本を抜き出すことを**抽出**という．
- 母集団の各個体を等しい確率で抽出する方法を**無作為抽出**といい，無作為抽出によって選ばれた標本を**無作為標本**という．

- 大きさ$N$の母集団において，変量$x$の取りうる異なる値を　$x_1$，$x_2$，⋯，$x_n$
とし，それぞれの値をとる個体の個数を
$f_1$，$f_2$，⋯，$f_n$とすると，この母集団における変量$x$の度数分布表は右の表になる．

| $x$ の値 | 度数 |
|---|---|
| $x_1$ | $f_1$ |
| $x_2$ | $f_2$ |
| ⋮ | ⋮ |
| $x_n$ | $f_n$ |
| 計 | $N$ |

- 変量$x$の値をXとするとき，Xは確率変数である．Xの確率分布表は下の表のようになる．
- このXの確率分布を**母集団分布**という．また，確率変数Xの期待値，標準偏差をそれぞれ**母平均**，**母標準偏差**といい，**m**，**σ**で表す．
- この母平均m，母標準偏差σは，母集団における変量$x$の平均値，標準偏差に，それぞれ一致する．

| X | | $x_1$ | $x_2$ | ⋯ | $x_n$ | 計 |
|---|---|---|---|---|---|---|
| P | | $\frac{f_1}{N}$ | $\frac{f_1}{N}$ | ⋯ | $\frac{f_1}{N}$ | 1 |

# 例題：1

- 数字1が書かれた札が10枚，数字2の札が20枚，数字3の札が30枚ある．この60枚の札を母集団とし，札の番号を変量と考える．
- このとき，母集団分布は以下の表のようになる．
- この時の母平均mと母標準偏差σを求めよ

| X | | 1 | 2 | 3 | 計 |
|---|---|---|---|---|---|
| P | | $\frac{1}{6}$ | $\frac{2}{6}$ | $\frac{3}{6}$ | 1 |

- m $= E(X) = \sum_{k=1}^{n} x_k p_k$
- $\sigma^2 = V(X) = E\{(X - E(X))^2\} = E(X^2) - \{E(X)^2\}$

# 練習：1

- 表は，あるクラス40人の5段階評価の成績を整理したものである．40人を母集団，成績を変量としたときの母集団分布を表に示せ．また、母平均，母標準偏差を求めよ．

| 成績 | | 1 | 2 | 3 | 4 | 5 | 計 |
|---|---|---|---|---|---|---|---|
| 人数 | | 2 | 6 | 24 | 6 | 2 | 40 |

# 点推定と区間推定

## ・推定

データを用いて未知である母数(母集団の分布を特徴付ける数値のこと)を定める作業

| 推定法 | 意味 |
| --- | --- |
| 点推定 | 未知母数をある一つの値で推定する方法.<br>例:母平均mを標本平均で推定する |
| 区間推定 | ある確率で未知母数が存在しうる範囲を推定する方法<br>「母平均mは確率0.95でAとBの間にある」など |

# 点推定

- n個のデータから成る母集団のバラツキ具合を表す指標として，次の**母分散**と標準偏差がある．

*1.* 母分散 $\sigma^2 = \frac{1}{n}\sum_{k=1}^{n}(x_k - \bar{x})^2 = \frac{1}{n}\sum_{k=1}^{n}(x_k^2 - \bar{x}^2)$

2. 標準偏差 $\sigma = \sqrt{\sigma^2}$

- 母平均と母分散を推定したいとき，これらの推定量（推定した推量）を$\tilde{m}$，$\widetilde{\sigma^2}$とおくことにする．これを不偏推定量とよぶ．
- 不偏推定量 $\tilde{m} = \bar{X} = \frac{1}{n}\sum_{k=1}^{n} x_k = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$であり，これを**標本平均**とよぶ．
- 不偏推定量 $\widetilde{\sigma^2} = \frac{1}{n-1}\sum_{k=1}^{n}(x_k - \bar{x})^2 = \frac{1}{n-1}(\sum_{k=1}^{n} x_k^2 - n\cdot\bar{x}^2)$であり，これを**標本分散**とよぶ．

# 点推定・練習：2

- 全国で1万人の学生がある数学の試験を受験した．この試験結果を母集団と見て，ここから無作為に抽出した10人の学生の試験結果を示す．

  45，76，10，81，24，65，59，90，37，63 （単位は点）

この標本を基に，母平均mと母分散$\sigma^2$の不偏推定量 $\tilde{m}$，$\widetilde{\sigma^2}$を少数第1位まで求めよ．

# 母平均の区間推定

- ～正規分布をしている母集団mの区間推定～

①一般に，母平均m，母標準偏差σをもつ母集団から抽出された大きさnの無作為標本の標本平均$\bar{x}$は，nが大きい時に正規分布$N(m, \frac{\sigma^2}{n})$に従う.

- $\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}$

＊$\bar{x}$は$N(m, \frac{\sigma^2}{n})$

＊標本平均$\bar{x}$が平均m，分散$\frac{\sigma^2}{n}$の正規分布に従う

②変数の基準化を行う

- $z = \frac{\bar{x} - m}{(\frac{\sigma}{\sqrt{n}})}$

＊zは$N(0, 1)$の標準正規分布に従う

# 母平均の区間推定

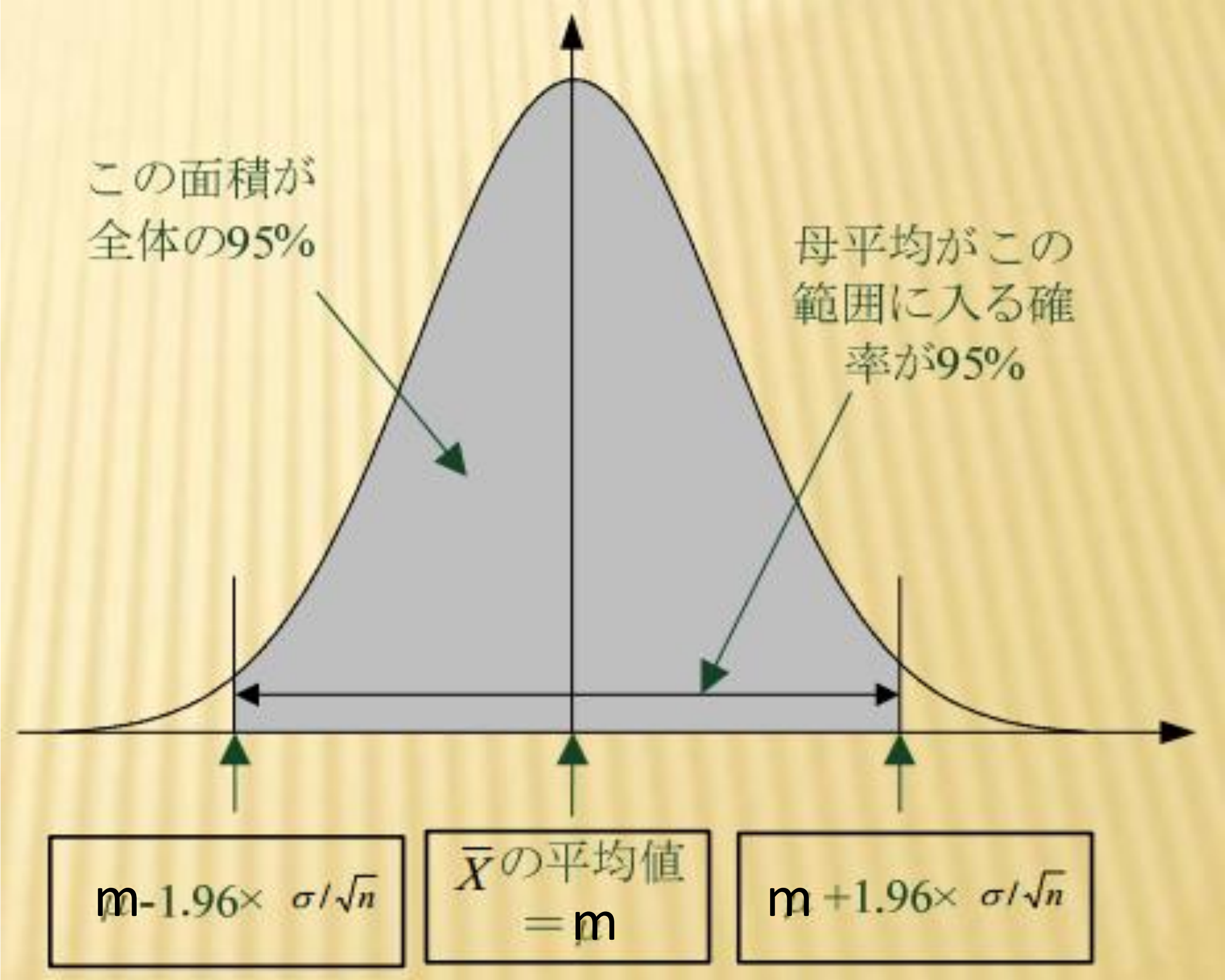
③基準化された$\bar{x}$の標準正規分布において，$\bar{x}$はmを中心として95%の確率で次の区間に現れ，かつmを含む．

よって，この$\bar{x}$の信頼度95%の信頼区間は

$$\bar{X} - 1.96 \cdot \frac{\sigma}{\sqrt{n}} \leqq \mathrm{m} \leqq \bar{X} + 1.96 \cdot \frac{\sigma}{\sqrt{n}}$$

になる．信頼区間が与えられた区間推定は，上記のような式のみで導ける．

＊よく使用される信頼区間として90%，95%，99%がある



# 区間推定・例題：2

- 大量に生産されたある一種類の製品の中から，400個（=n）を無作為に抽出して重さを図ったら，平均値（$\bar{X}$）98.8g，標準偏差（σ）2.0gを得た．この製品の平均重量mgに対して，信頼度95%の信頼区間を求めよ．

# 区間推定・練習：3

- 砂糖の袋の山から100袋を無作為に抽出して，重さの平均値300.4gを得た。重さの母標準偏差を7.5gとして，袋の山に存在する1袋あたりの重さ（平均）を信頼度95%で推定せよ.

# 参考WEBページ

- http://web.agr.ehime-u.ac.jp/~nkakko/image/Text7.pdf
- http://statistics.co.jp/reference/Toukeigaku_Nyumon/nyumon4.pdf

# 課題:

- 信頼度90%と98%のときの,例題2と練習3の問題を解く。