講義にて配布した資料の抜粋・縮小版です



平成22年度
統計工学　講義資料（2）

～平均と分散Ｉ～

(C) Sadami SUZUKI



## 代表値としての平均値

### 平均値

- 対象となる集団のある性質（特性）に関するデータを“平らに均した”**代表値**

- 算術平均…Arithmetic Mean
- 幾何平均…Geometric Mean
- 調和平均…Harmonic Mean
- 絶対平均…Mean Absolute Value
- RMS…Root Mean Square

(C) Sadami SUZUKI



## その他の主たる基本的な統計値

- 先の身長データに戻って考えよう

| 番号 | 身長(cm) |
|---|---|
| 1 | 180.7 |
| 2 | 178.3 |
| 3 | 182.6 |
| 4 | 173.0 |
| 5 | 177.5 |
| 6 | 175.1 |
| 7 | 164.9 |
| 8 | 175.5 |
| 9 | 185.4 |

大きい順に並び替え

| 番号 | 身長(cm) | |
|---|---|---|
| 9 | 185.4 | 最大値 |
| 3 | 182.6 | |
| 1 | 180.7 | |
| 2 | 178.3 | |
| 5 | 177.5 | 中央値（メディアン：Median） |
| 8 | 175.5 | |
| 6 | 175.1 | |
| 4 | 173.0 | |
| 7 | 164.9 | 最小値 |

範囲

(C) Sadami SUZUKI



## 算術平均の重要な性質

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

① 算術平均はデータの1次変換を保持する

$$y_i = ax_i + b \quad \Rightarrow \quad \bar{y} = a\bar{x} + b \qquad (a, b \text{ は定数})$$

② 算術平均からの偏差の和は常に0（ゼロ）である

$$\sum_{i=1}^{n}\left(x_i - \bar{x}\right) \equiv 0$$

③ 算術平均からの偏差の平方和は他のいかなる一定値からの偏差平方和よりも小さい

$$\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 \le \sum_{i=1}^{n}\left(x_i - a\right)^2 \qquad (a \text{ は定数})$$

(C) Sadami SUZUKI



## 「平均値」だけで当該集団の特徴を表せるだろうか？

1. 「平均値」から得られる情報は画一的？
   - 考える対象（母集団）によって標本観測に基づく統計値としての「平均値」から得られる情報は異なる
   - **標本抽出法（無作為抽出）の重要性**
2. 二つの集団を比べる際は？

(C) Sadami SUZUKI



## データ分布状況の視覚化

**度数分布（ヒストグラム）**
**データのバラツキ具合に着目**

(C) Sadami SUZUKI

講義にて配布した資料の抜粋・縮小版です

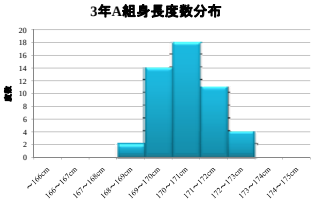


## 度数分布を調べる

**度数分布** ・・・データを大きさによっていくつかの区間(級)に分け，各級に入るデータの数(度数)を明らかにしたもの

**3年A組の身長に関する度数分布表とヒストグラム**

| 階級 | 度数 |
|---|---|
| ～166cm | 0 |
| 166～167cm | 0 |
| 167～168cm | 0 |
| 168～169cm | 2 |
| 169～170cm | 14 |
| 170～171cm | 18 |
| 171～172cm | 11 |
| 172～173cm | 4 |
| 173～174cm | 0 |
| 174～175cm | 0 |
| 合計 | 49 |





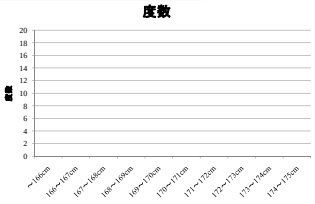



## 度数分布を調べる

【確認テスト】　3年B組の身長データから度数分布表およびヒストグラムを作成せよ

**3年B組の身長に関する度数分布表とヒストグラム**

| 階級 | 度数 |
|---|---|
| ～166cm | |
| 166～167cm | |
| 167～168cm | |
| 168～169cm | |
| 169～170cm | |
| 170～171cm | |
| 171～172cm | |
| 172～173cm | |
| 173～174cm | |
| 174～175cm | |
| 合計 | |





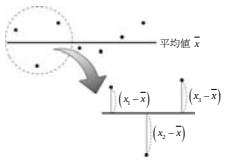



# 製品の品質に置き換えて考えてみよう・・・





## データのバラツキを数値化しよう




- バラツキの数値化(1)・・・ $\sum_{i=1}^{n}(x_i-\bar{x})$
  - (観測値－平均値)の総和
  - 算術平均からの偏差の和は常にゼロ
- バラツキの数値化(2)・・・ $\sum_{i=1}^{n}\left|(x_i-\bar{x})\right|$
  - |(観測値－平均値)|の総和
  - 計算上の処理が煩雑化
- バラツキの数値化(3)・・・ $\sum_{i=1}^{n}(x_i-\bar{x})^2$
  - ((観測値－平均値)$^2$)の総和
  - データ数の増加に伴い大きくなってしまう
- バラツキの数値化(4)・・・ $\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$
  - ((観測値－平均値)$^2$)の総和÷データ数





## データのバラツキ具合を表す統計値

- **"分散"** ・・・((観測値－平均値)$^2$)の平均値

$$s_x^2=\frac{(x_1-\bar{x})^2+(x_2-\bar{x})^2+\cdots+(x_n-\bar{x})^2}{n}=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

- **"標準偏差"** ・・・もとのデータと次元を揃えた分散の平方根
  ※ 平均値からの偏差のRMSに相当し，偏差の大きさの平均なるイメージの値

$$s_x=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2}$$





## 分散(標準偏差)の重要な性質

分散：$s_x^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$　標準偏差：$s_x=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2}$

① データの1次変換と分散(標準偏差)

$$y_i=ax_i+b \quad \Rightarrow \quad s_y^2=a^2s_x^2\ ,\ s_y=|a|s_x \qquad (a,b\ \text{は定数})$$

② 分散の計算式・・・「平均値からの偏差の2乗の平均」⇔「2乗の平均値から平均値の2乗を引く」

$$s_x^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2=\frac{1}{n}\sum_{i=1}^{n}x_i^2-\bar{x}^2$$

③ 平均値に加え，分散(標準偏差)の値が分かれば，どのような範囲にどの程度のデータが含まれるかといった情報の概略を把握することができる