データの特徴を示す（pp9〜15）

Step1: ヒストグラム
Step2: 代表値と散布度

代表値と散布度

- 代表値：最も一般的な値
  平均値，中央値
  散布度：データのバラツキ度
  標準偏差，%点
- 左右対称：平均値と標準偏差
  非対称　：中央値と%点

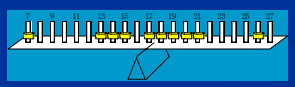
代表値はヒストグラムの位置

- 運動部男子生徒10人の身長
  157,170,167,169,168,171,176,164,165,163
  平均値は $\frac{157+170+\cdots+163}{10}=167$
- $x_1, x_2, \ldots, x_n$の平均値は

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

平均値の図解

- 7, 20, 17, 19, 18, 21, 26, 14, 15, 13
  (7+20+・・・・・・+13)/10=17



分散

- 要素から平均値を引いた値を偏差
  157-167,170-167,…,163-167
  ⇒ -10, 3, 0, 2, 1, 4, 9, -3,-2,-4
- 偏差の平均値は0
- 偏差の2乗の平均値を分散

$$\frac{(-10)^2+3^2+\cdots\cdots+(-2)^2+(-4)^2}{10}=24$$

- $x_1, x_2, \ldots, x_n$の平均値は

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

- 分散 $=\frac{(x_1-\bar{x})^2+\cdots\cdots+(x_n-\bar{x})^2}{n}$
- 分散の平方根を標準偏差(SD)

$$SD=\sqrt{\text{分散}}$$

$$SD=\sqrt{24}=4.90$$

標準偏差 SD (Standard Deviation)

- 個々の値が平均値から平均してどの程度離れているかを示す
  - （平均−2SD，平均＋2SD）
    に殆ど全ての要素が含まれる
  - 例 (167-9.8, 167+9.8)=(157.2, 176.8)
    に10人中9人含まれる
- 身長のように，左右対称で中央に集中する分布
  - 平均値とSDが 代表値と散布度に適する

変動係数

- 象とネズミの体重のバラツキを比較
  - 標準偏差の大きさで比較すれば，象がネズミより大きい（トンとグラム）
- 象を縮小してネズミの重さにして比較

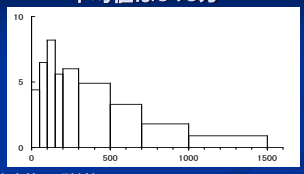
$$\text{変動係数}=\frac{\text{標準偏差}}{\text{平均}}$$

- 平均値が変化 ⇒ 標準偏差も変化
  - 標本の全ての要素を3倍すれば平均も標準偏差も3倍になる
  - 変動係数は不変

非対称分布の代表値 中央値：丁度中央の値

- 中央値の定義と計算
  1. 観測値を小さい順に並べる
  2. 標本サイズをnとする
     nが奇数：中央の値
     nが偶数：中央の2つの平均値
- 例
  3, 6, 7, 8, 120　の中央値は　7
  3, 5, 20, 21　の中央値は　(5+20)/2=12.5

貯蓄額の中央値は436万
平均値は646万



- 中央値の頑健性
  3, 6, 7, 8, 12において
  8が20になっても，3が6になっても中央値は7で不変

Excel関数

- データの個数　COUNT(範囲)
- 合計　SUM(範囲)
- 平均値　AVERAGE(範囲)
- 分散　VARP(範囲)
- 標準偏差　STDEVP(範囲)
- 中央値　MEDIAN(範囲)
- 最大値，最小値　MAX(範囲)，MIN(範囲)
- 四分位点　QUARTILE(範囲，戻り値)

注意

- 本節で扱った分散，標準偏差は
  第6節の標本分散（不偏分散），標本標準偏差
  とは結果が異なる

元のデータにaを足したときの平均値と標準偏差

$$\bar{X} = \frac{(x_1+a)+(x_2+a)+\cdots+(x_n+a)}{n}$$
$$= \frac{(x_1+x_2+\cdots+x_n)+na}{n}$$
$$= \bar{x}+a$$
$$SD = \sqrt{\frac{\{x_1+a-(\bar{x}+a)\}^2+\cdots+\{x_n+a-(\bar{x}+a)\}^2}{n}}$$
$$= \sqrt{\frac{(x_1-\bar{x})^2+\cdots+(x_n-\bar{x})^2}{n}}$$

元のデータをb倍したときの平均値と標準偏差

$$\bar{X} = \frac{bx_1+bx_2+\cdots+bx_n}{n}$$
$$= b\cdot\frac{x_1+x_2+\cdots+x_n}{n} = b\cdot\bar{x}$$
$$SD = \sqrt{\frac{(bx_1-b\bar{x})^2+(bx_2-b\bar{x})^2+\cdots+(bx_n-b\bar{x})^2}{n}}$$
$$= \sqrt{\frac{b^2(x_1-\bar{x})^2+b^2(x_2-\bar{x})^2+\cdots+b^2(x_n-\bar{x})^2}{n}}$$
$$= b\sqrt{\frac{(x_1-\bar{x})^2+(x_2-\bar{x})^2+\cdots+(x_n-\bar{x})^2}{n}}$$