< 資料の整理 1 >

n 個の実数値データ $\{x_1, x_2, \cdots, x_n\}$ に対し

$$\boxed{m = \frac{1}{n}\{x_1 + x_2 + \cdots + x_n\} = \frac{1}{n}\sum_{k=1}^{n} x_k} \quad \text{(平均)}$$

を**平均**という。また平均 m に対し

$$\boxed{v = \frac{1}{n}\{(x_1 - m)^2 + (x_2 - m)^2 + \cdots + (x_n - m)^2\} = \frac{1}{n}\sum_{k=1}^{n}(x_k - m)^2} \quad \text{(分散)}$$

を**分散**といい，その正の平方根 $\sqrt{v}$ を**標準偏差**といい

$$\boxed{\sigma = \sqrt{v} = \sqrt{\frac{1}{n}\sum_{k=1}^{n}(x_k - m)^2}} \quad \text{(標準偏差)}$$

で表す。

例 1 7 個数値データ $\{0,\ 2,\ 4,\ 6,\ 8,\ 10,\ 12\}$ の平均 m は

$$m = \frac{1}{7}\{0 + 2 + 4 + 6 + 8 + 10 + 12\} = 6$$

分散は

$$v = \frac{1}{7}\{(0-6)^2 + (2-6)^2 + (4-6)^2 + (6-6)^2 + (8-6)^2 + (10-6)^2 + (12-6)^2\} = 16$$

より標準偏差は $\sigma = \sqrt{v} = \sqrt{16} = 4$

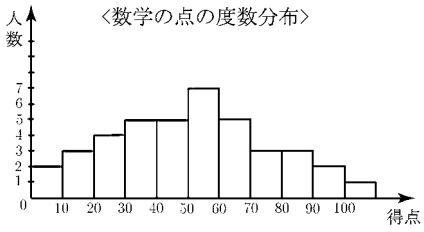
問 1 以下の数値データの平均と分散および標準偏差を求めよ。

(1) $\{3,\ 4,\ 5,\ 6,\ 7,\ 8,\ 9\}$ (2) $\left\{-\frac{3}{2},\ -1,\ -\frac{1}{2},\ 0,\ \frac{1}{2},\ 1,\ \frac{3}{2}\right\}$

分散 (および標準偏差) は「データの広がりぐあい」を示す数値である。分散が小さいほど平均の近くにデータが集まっている。実際のデータの分布を表すには以下のようなグラフを使う。

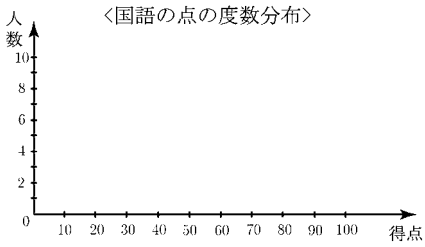
例 2 以下のデータは 40 人の数学の試験結果を低い点から順に並べたものである。

{0, 6, 12, 15, 18, 20, 22, 25, 28, 30, 33, 35, 35, 38,
40, 40, 42, 45, 48, 50, 50, 52, 54, 55, 57, 58, 60,
60, 63, 65, 67, 70, 75, 77, 80, 84, 86, 90, 94, 100}



このテストの平均は 49.475，分散は 616.75，標準偏差は 24.83 である。上図に 0 点〜9 点が 2 人，10 点台が 3 人，$\cdots$，90 点台 2 人，100 点 1 人としたグラフを書いた。このようなグラフを**度数分布**または**ヒストグラム**という。

問 2 以下のデータは 40 人の国語の試験結果である。右図にヒストグラムを作り，平均・分散・標準偏差を求めよ。

{25, 28, 30, 32, 36, 38, 40, 42, 44, 46, 46, 48, 48, 50, 50 50,
52, 52, 54, 54, 56, 58, 60, 60, 62, 64, 66, 68, 68, 70, 72 74,
76, 78, 80, 80, 82, 84, 86, 90}

この問題にかぎり EXCEL や Mathematica 等を使ってよい。