8. 累積相対度数：ある階級以下の相対度数を合計したもの

2　代表値 (P.15)

度数分布表，ヒストグラム：統計データを整理し，母集団に関する情報を得る一つの方法。

分布の状態を数値で表したい。

代表値： データを代表する値 $\Longrightarrow$ 平均値，分散，標準偏差，中央値 (メディアン)，最頻値 (モード)，•••

2.1　平均値 (P.16)

n 個のデータ： $x_1, x_2, \cdots, x_n$

算術平均 (P.16)：

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

表 1.3 (P.7) のデータから

$$\overline{x} = \frac{1}{20}(4.3 + 5.2 + \cdots + 5.9) = 5.53$$

となる。

加重平均 (P.16)：

階級値	階級境界値 (以上)　(未満)	度数
m_1	$a_0 \sim a_1$	f_1
m_2	$a_1 \sim a_2$	f_2
$\vdots$	$\vdots$	$\vdots$
m_k	$a_{k-1} \sim a_k$	f_k
	合計	n

ただし，$m_1 = \dfrac{a_0 + a_1}{2}, m_2 = \dfrac{a_1 + a_2}{2}, \cdots, m_k = \dfrac{a_{k-1} + a_k}{2}$ とする。

上のような度数分布表が利用可能なとき，

$$\overline{x} = \frac{1}{n}(f_1 m_1 + f_2 m_2 + \cdots + f_k m_k) = \frac{1}{n}\sum_{i=1}^{k} f_i m_i$$

として，平均値を計算することが出来る。$\Longrightarrow$ 加重平均 (各階級値を度数でウエイトづけして平均したもの)

$$\overline{x} = \sum_{i=1}^{k} \frac{f_i}{n} m_i$$

$\dfrac{f_i}{n}$ は相対度数である。

上の表のデータの平均を求めると，

$$\overline{x} = \frac{1}{20}\Big(2 \times 3.45 + 3 \times 4.45$$

$$+8 \times 5.45 + 5 \times 6.45 + 2 \times 7.45\Big)$$

$$= 5.55$$

階級の幅の選び方によって，多少，値は異なる。

2.2　分散，標準偏差 (P.20)

分散，標準偏差： データの散らばり具合を表す

分散，標準偏差が大きければ，データの存在する範囲が広い

標準偏差＝分散の平方根

分散 (s^2 で表す) の定義：

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

ただし， $\overline{x} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} x_i$ **とする。**

標準偏差： s

分散の実際の計算には，

$$s^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \overline{x}^2$$

を用いる。

なぜなら，

$$
\begin{aligned}
s^2 &= \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}(x_i^2 - 2\overline{x}x_i + \overline{x}^2) \\
&= \frac{1}{n}\Big(\sum_{i=1}^{n} x_i^2 - 2\overline{x}\sum_{i=1}^{n} x_i + \sum_{i=1}^{n}\overline{x}^2\Big) \\
&= \frac{1}{n}\Big(\sum_{i=1}^{n} x_i^2 - 2n\overline{x}^2 + n\overline{x}^2\Big) \\
&= \frac{1}{n}\Big(\sum_{i=1}^{n} x_i^2 - n\overline{x}^2\Big) \\
&= \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \overline{x}^2
\end{aligned}
$$

となる。

表 1.3 (P.7) のデータの分散を求めると，

$$s^2 = \frac{1}{20}\Big((4.3-5.53)^2 + (5.2-5.53)^2 + \cdots$$
$$+(5.9-5.53)^2\Big)$$
$$= 1.1591$$

または，

$$s^2 = \frac{1}{20}(4.3^2 + 5.2^2 + \cdots + 5.9^2) - 5.53^2$$
$$= 1.1591$$

$s = 1.0766$ **===> 標準偏差**

表 2.1 (P.17) の度数分布表からの計算では，

$$s^2 = \frac{1}{n}\sum_{i=1}^{k} f_i(m_i - \overline{x})^2$$

となる。ただし，$\overline{x} = \frac{1}{n}\sum_{i=1}^{k} f_i m_i$ とする。

実際の計算には，

$$s^2 = \frac{1}{n}\sum_{i=1}^{k} f_i m_i^2 - \overline{x}^2$$

を使う。

なぜなら，

$$
\begin{aligned}
s^2 &= \frac{1}{n}\sum_{i=1}^{k} f_i(m_i - \overline{x})^2 \\
&= \frac{1}{n}\sum_{i=1}^{k} f_i(m_i^2 - 2\overline{x}m_i + \overline{x}^2) \\
&= \frac{1}{n}\Big(\sum_{i=1}^{k} f_i m_i^2 - 2\overline{x}\sum_{i=1}^{k} f_i m_i + \overline{x}^2\sum_{i=1}^{k} f_i\Big) \\
&= \frac{1}{n}\Big(\sum_{i=1}^{k} f_i m_i^2 - 2n\overline{x}^2 + n\overline{x}^2\Big) \\
&= \frac{1}{n}\Big(\sum_{i=1}^{k} f_i m_i^2 - n\overline{x}^2\Big) \\
&= \frac{1}{n}\sum_{i=1}^{k} f_i m_i^2 - \overline{x}^2
\end{aligned}
$$

となる。

上の表のデータの分散を求めると，

$$s^2 = \frac{1}{20}\Big(2(3.45-5.55)^2 + 3(4.45-5.55)^2$$

$$+8(5.45-5.55)^2 + 5(6.45-5.55)^2$$

$$+2(7.45-5.55)^2\Big)$$

$$= 1.19$$

または，

$$s^2 = \frac{1}{20}(2\times 3.45^2 + 3\times 4.45^2$$

$$+8\times 5.45^2 + 5\times 6.45^2 + 2\times 7.45^2) - 5.55^2$$

$$= 1.19$$

すなわち，$s = 1.0909$，

2.3　範囲，四分位点，メディアン，モード (P.18)

- 範囲：最大値－最小値
- 四分位点：

  25 %点 (第 1 四分位点)，50 %点 (第 2 四分位点)，75 %点 (第 3 四分位点) のこと
- 四分位範囲：第 3 四分位点－第 1 四分位点
- メディアン（中央値）：

  大きい順に並べて，真ん中の値 (第 2 四分位点) $\longrightarrow$ 表 1.3 (P.7) のデータでは，大きい順に並べて 10 番目と 11 番目のデータの平均で，$(5.6 + 5.6)/2 = 5.6$

- **モード（最頻値）：**

  最も多い度数の階級値 $\longrightarrow$ 表 1.3 (P.7) のデータでは 5.45，階級の幅によって変わる

2.4　相関係数 (P.23)

2 変数データの組に関する代表値 $\Longrightarrow$ 共分散，相関係数

例： 100 人の家計からの消費と所得，身長と体重

n **組のデータ** (x_1, y_1), (x_2, y_2), $\cdots$, (x_n, y_n)

共分散 s_{xy}

$$s_{xy} = \frac{1}{n}\Big((x_1 - \overline{x})(y_1 - \overline{y}) + (x_2 - \overline{x})(y_2 - \overline{y})$$