第５章　データの分析

１　データの管理　　２　データの代表値

1　データ

人の身長，体重や運動の記録などのように，ある特性を表す数量を **変量** といい，ある変量の測定値や観測値の集まりを **データ** という。

2　度数分布

度数分布表　データのとる値をいくつかの区間に区切って階級を定め，各階級に度数を対応させた表。各階級の中央の値を **階級値** という。

ヒストグラム　度数分布表を柱状のグラフで表したもの。

3　代表値

平均値　変量 x のデータの値が x_1，x_2，……，x_n であるとき，このデータの平均値 $\overline{x}$ は　$\overline{x}=\dfrac{1}{n}(x_1+x_2+\cdots\cdots+x_n)$

最頻値（モード）　データにおいて，最も個数の多い値。データが度数分布表に整理されているときは，度数が最も大きい階級の階級値。

中央値（メジアン）　データを値の大きさの順に並べたとき，その中央の位置にくる値。データの大きさが偶数のときは，中央の２つの値の平均値。

３　データの散らばりと四分位数

1　四分位数

範囲　データの最大値から最小値を引いた差。

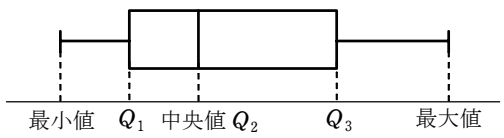
四分位数　データの値を大きさの順に並べたとき，４等分する位置の値。
小さい方から順に，**第１四分位数**，**第２四分位数**，**第３四分位数** といい，順に Q_1，Q_2，Q_3 で表す。第２四分位数はデータの中央値である。

四分位範囲　データの第３四分位数 Q_3 と第１四分位数 Q_1 の差 Q_3-Q_1

四分位偏差　四分位範囲の半分の値。

2　箱ひげ図

箱ひげ図　データの分布を，次の５つの値で表した図。平均値を記入することもある。
最小値，第１四分位数 (Q_1)，中央値 (Q_2)，第３四分位数 (Q_3)，最大値



４　分散と標準偏差

1　分散，標準偏差　　x_1，x_2，……，x_n を変量 x のデータの値とする。

偏差　データの各値と平均値 $\overline{x}$ との差 $x-\overline{x}$

分散　偏差の２乗の平均値。s^2 で表す。

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

$(x$ のデータの分散$)=(x^2$ のデータの平均値$)-(x$ のデータの平均値$)^2$

標準偏差　分散の正の平方根。s で表す。

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

$(x$ のデータの標準偏差$)=\sqrt{(x^2\text{ のデータの平均値})-(x\text{ のデータの平均値})^2}$

研究　変量の変換

a，b は定数とする。変量 x のデータから $y=ax+b$ によって新しい変量 y のデータが得られるとき，x，y のデータの平均値を $\overline{x}$，$\overline{y}$，分散を $s_x{}^2$，$s_y{}^2$，標準偏差を s_x，s_y とすると　$\overline{y}=a\overline{x}+b$，　$s_y{}^2=a^2s_x{}^2$，　$s_y=|a|s_x$

５　データの相関

1　散布図

散布図　２つの変量からなるデータを平面上に図示したもの。

正の相関　２つの変量からなるデータにおいて，一方が増加すると他方も増加する傾向がみられるとき，２つの変量には正の相関があるという。

負の相関　２つの変量からなるデータにおいて，一方が増加すると他方が減少する傾向がみられるとき，２つの変量には負の相関があるという。

散布図において，正の相関があるとき，データを表す点は全体的に右上がりとなり，負の相関があるとき，データを表す点は全体的に右下がりとなる。

2　相関係数

相関係数　相関の正負と強弱を表す値。２つの変量 x，y からなるデータにおいて，$(x-\overline{x})(y-\overline{y})$ の総和を a，$(x-\overline{x})^2$ の総和を b，$(y-\overline{y})^2$ の総和を c とするとき，次の値 r を x と y の **相関係数** という。

$$r=\frac{a}{\sqrt{bc}}\quad(-1\leqq r\leqq 1)$$

相関係数 r が１に近いほど正の相関が強く，-1 に近いほど負の相関が強い。また，相関がないとき，r は０に近い値をとる。

補足　表計算ソフトによるデータの分析

1　平均値，分散，標準偏差，相関係数

データ A1 から A5 について　**平均値**　＝AVERAGE(A1：A5)
分散　＝VARP(A1：A5)
標準偏差　＝STDEVP(A1：A5)
データ A1 から A5，B1 からB5 の**相関係数**　＝CORREL(A1：A5，B1：B5)