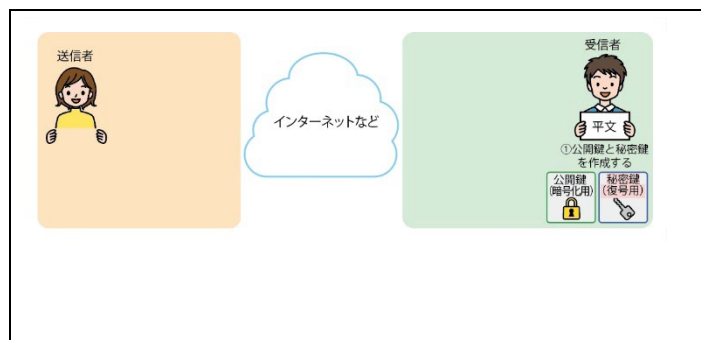
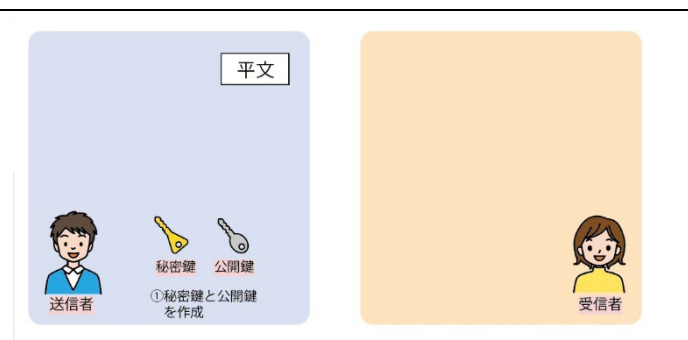


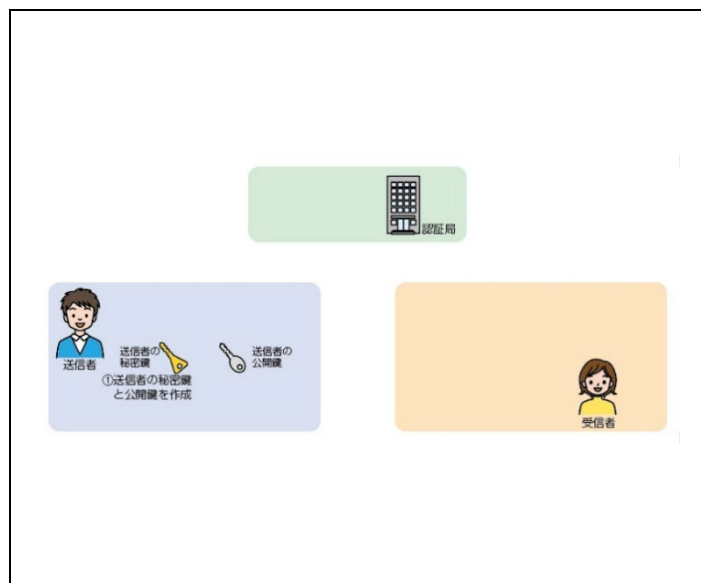
1



2



3



4

Screenshot of the e-Stat website showing the 'ネットワークと情報システム' (Network and Information System) section. The page includes a navigation bar with 'e-Stat' and '統計で見る日本' (Japan seen through statistics). The main content area lists several options: '最初から全部解く' (Solve everything from the beginning), '続きから解く' (Continue from where you left off), '不正解の問題だけ解く' (Solve only the questions you got wrong), and '登録した問題を解く' (Solve the questions you registered). A toggle switch for 'シャッフル出題: OFF' is visible. The footer includes a home button and a '使い方' (How to use) link.

5

Screenshot of a textbook page titled '5章: 章末問題 解説' (Chapter 5: End-of-chapter problems Explanation). The page contains a section titled '7 <解説>' (7 <Explanation>). The text describes the 'トレーサビリティ' (Traceability) system, which tracks the production flow from raw materials to finished products. It mentions that the system is based on cloud-based streaming services and is used for real-time monitoring across the country.

6

Screenshot of the e-Stat website showing the '統計で見る日本' (Japan seen through statistics) section. The page features a navigation bar with 'e-Stat' and '統計で見る日本'. The main content area includes a search bar and several filters for '統計データを探す' (Find statistics data), such as 'すべて' (All), '分野' (Field), '組織' (Organization), '地域' (Region), and '時期' (Period). There are also links to '統計データを活用する' (Use statistics data), '統計データの利用' (Use of statistics data), and '統計データの活用' (Use of statistics data). The footer includes a home button and a '使い方' (How to use) link.

1

2

3

164

2 代表値

データ全体の特徴を、1つの値で代表して表すことがある。このような値を、データの **代表値** という。ここでは代表値について確認していこう。

1 平均値・最頻値・中央値

一般に、データの値の個数を、そのデータの **大きさ** という。変数 x について、大きさ n のデータの値を $x_1, x_2, \dots, x_n$ のように表すことがある。このとき、値の総和をデータの大きさ n で割った値を、そのデータの **平均値** といい、 $\bar{x}$ で表す。すなわち

$$\bar{x} = \frac{1}{n}(x_1 + x_2 + \dots + x_n)$$

大きさが4のデータ3, 4, 6, 8の平均値 $\bar{x}$ を求めると

$$\bar{x} = \frac{1}{4}(3 + 4 + 6 + 8) = \frac{1}{4} \times 20 = 5$$

大きさが5のデータ12, 48, 31, 15, 28の平均値を求めよ。

データにおいて最も個数の多い値を、そのデータの **最頻値** または **モード** という。データが度数分布表に整理されているときは、度数の最も多い階級の階級値を最頻値とする。

次の表は、ある高校のクラス40人の靴のサイズ(cm)のデータである。

サイズ	22.5	23.0	23.5	24.0	24.5	25.0	25.5	26.0	26.5	27.0
人数	3	5	4	4	3	7	9	3	2	1

このデータの最頻値は 25.5 (cm)

162ページの度数分布表において、最頻値を求めよ。

4

5

180

3 データの外れ値

データに、他の値と比べて極端に大きい値や小さい値が含まれる場合がある。ここでは、そのような値について考えてみよう。

1 外れ値

右の表は、ある高校のバスケットボール部員7人が、それぞれ30本のフリースローを行ったときの成功本数の記録である。

値を小さい順に並べると

9, 11, 11, 12, 13, 14, 28

ここで、平均値は14本であるから、部員③を除く全員の記録が平均値以下になっている。

上の例において、部員③の記録は他の値と比べて極端に大きい。このように、データの値のうち、他の値と比べて極端に大きい値や小さい値を **外れ値** という。

一般に、四分位範囲を用いて、外れ値を次のように定める。

2 外れ値

データの第1四分位数を Q_1 、第3四分位数を Q_3 とするとき、 $Q_1 - 1.5(Q_3 - Q_1)$ 以下 または $Q_3 + 1.5(Q_3 - Q_1)$ 以上の値を外れ値とする。

6

166

3 四分位数と四分位範囲

データの順がばりの度合いを数値化することについて学んでいこう。

1 四分位数

データの値を小さい順に並べたとき、データを4分の1ずつに分ける位置にある3つの数を **四分位数** という。四分位数は小さい方から順に **第1四分位数**、**第2四分位数**、**第3四分位数** といい、それぞれ Q_1 、 Q_2 、 Q_3 で表す。四分位数は次のように定められる。

① データの値を小さい順に並べたとき、中央値が Q_2 である。

② Q_1 を境目として、データを前半と後半に分けたとき、前半の中央値が Q_1 、後半の中央値が Q_3 である。なお、データの大きさが奇数のとき、中央値 Q_2 は前半、後半のどちらにも含まない。

次の小さい順に並べられたデータについて、四分位数を求めよう。

4, 5, 5, 6, 7, 8, 9, 10, 11

中央値が Q_2 であるから $Q_2 = 7$

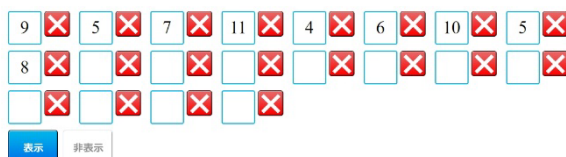
Q_1 を境目として、データを前半と後半に分ける。 Q_1 は前半の中央値であるから $Q_1 = \frac{5+5}{2} = 5$

Q_3 は後半の中央値であるから $Q_3 = \frac{9+10}{2} = 9.5$

次の小さい順に並べられたデータについて、四分位数を求めよ。

(1) 5, 6, 6, 8, 9, 10, 12 (2) 1, 3, 4, 8, 10, 12, 13, 15, 16, 17

1



- ☒ 第1四分位数
- ☒ 第2四分位数
- ☒ 第3四分位数
- ☒ 前半, 後半
- ☒ 箱ひげ図

2

174

2 データの相関

2つの位置で表されるデータについて、変量間の関係を探る方法を学んでいこう。

1 散布図

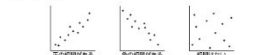
右の図は、生徒 10 人の 10 点満点の小テストのデータについて、英語の得点 x を横軸、数学の得点 y を縦軸にとり、①～④の生徒に対応する点を表示したものである。



図において、単一の点が右上がり分布していることから、英語の得点 x が低い生徒は、数学の得点 y も低い傾向にあることがわかる。

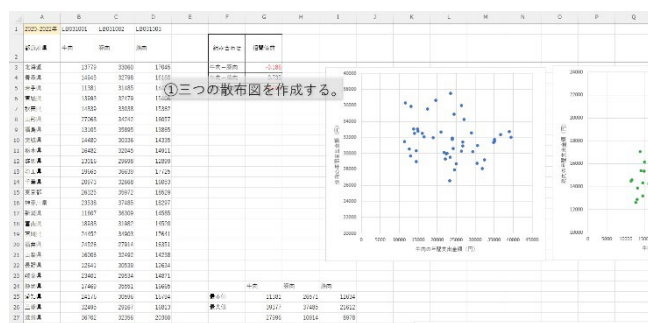
上の例の図のように、2つの変量で表されるデータを、平面上の点で表した図を **散布図** という。

一般に、2つの変量の間に、一方が大きくなり他方も大きい傾向があるとき、**正の相関** があるという。反対に、一方が大きくなり他方が小さい傾向があるとき、**負の相関** があるという。どちらの傾向もないとき、**相関はない** という。



なお、2つの変量の間に正または負の相関があるとき、散布図において、点の分布が1つの直線に近づくほど **相関が強い** といひ、散らばっているほど **相関が弱い** という。

3



4

170

2 データの分析

1 分散と標準偏差

四分位範囲は、中央値を基準としたデータの散らばりの程度を表す値であった。ここでは、平均値を基準としたデータの散らばりの程度を表す方法を学んでいこう。

分散と標準偏差

変量 x について、大きさ n のデータ $x_1, x_2, \dots, x_n$ の各値から平均値を引いた値の総和は 0 になる。

$$\begin{aligned} \text{すなわち、平均値を } \bar{x} \text{ と表すと} \\ (x_1 - \bar{x}) + (x_2 - \bar{x}) + \dots + (x_n - \bar{x}) \\ = (x_1 + x_2 + \dots + x_n) - n\bar{x} \\ = (x_1 + x_2 + \dots + x_n) - n \times \frac{1}{n}(x_1 + x_2 + \dots + x_n) = 0 \end{aligned}$$

変量 x について、大きさ n のデータ $x_1, x_2, \dots, x_n$ の平均値を $\bar{x}$ とするとき、各値から平均値を引いた値 $x_1 - \bar{x}, x_2 - \bar{x}, \dots, x_n - \bar{x}$ を、**偏差** といふ。上で示したように、偏差の総和は 0 であるから、散らばりの度合いを表すことができる。

そこで、偏差を 2 乗した値

$$(x_1 - \bar{x})^2, (x_2 - \bar{x})^2, \dots, (x_n - \bar{x})^2$$

の平均値を考える。この値を **分散** といひ、 s^2 で表す。

また、分散の正の平方根を **標準偏差** といひ、 s で表す。

$$\begin{aligned} \text{分散} \quad s^2 &= \frac{1}{n}[(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2] \\ \text{標準偏差} \quad s &= \sqrt{\frac{1}{n}[(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2]} \end{aligned}$$

5

- ★あるクラスで 100 点満点の小テストを行ったところ、得点 x の平均値は $\bar{x} = 64$ 、標準偏差は $s = 15$ であった。このとき、 $u = 10\left(\frac{x - \bar{x}}{s}\right) + 50$ によって得られる変量 u について、次の問いに答えよ。
- 得点が 70 点であるとき、 u の値を求めよ。
 - u の平均値を、標準偏差 s_u を求めよ。
 - 次の①～③のうち、正しいといえるものをすべて選べ。
 - A の得点 x_A と B の得点 x_B について $x_A < x_B$ であるとき、対応する u の値 u_A, u_B について $u_A > u_B$ となることがある。
 - $\bar{x}$ の値が大きくなると、 u の値は小さくなる。
 - s の値が変わっても、 s_u の値は変わらない。

つながる数学



6

92

3 仮説検定

さいころを続けて投げるとき、偶数の目ばかり、または奇数の目ばかりが出た場合、さいころが正しくつくりられているかどうかをどのように判断すればよいだろうか。

このような判断をするには、確率を用いて次のように考えればよい。

- 「さいころは正しくつくりられている」と仮説を立てる。
- 仮説を否定するかどうかの基準となる確率を定める。たとえば、確率 0.05 とする。
- さいころを 10 回投げるとき、偶数の目が出る回数を X とする。正しくつくりたいさいころであれば、1 回ごとの偶数の目が出る確率は $\frac{1}{2}$ であるから、 X は二項分布 $B(10, \frac{1}{2})$ に従い、確率分布は右の表のようになる。

X	P
0	0.001
1	0.010
2	0.054
3	0.117
4	0.205
5	0.246
6	0.205
7	0.117
8	0.054
9	0.010
10	0.001
計	1

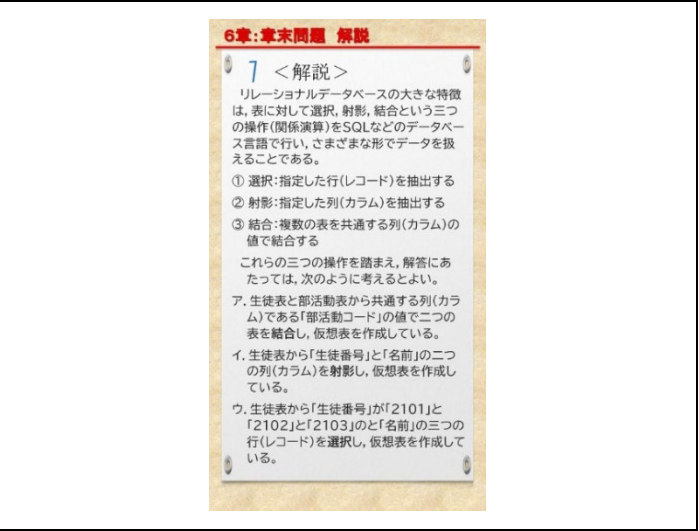
となり、基準の確率 0.05 以下である。よって、仮説「さいころは正しくつくりられている」は誤りと判断する。

上の例のように、母集団についての仮説が誤りかどうか確率を用いて判断する方法を **仮説検定** といふ。仮説検定においては、仮説を **帰無仮説** とよび、帰無仮説と対立する仮説を **対立仮説** とよび、

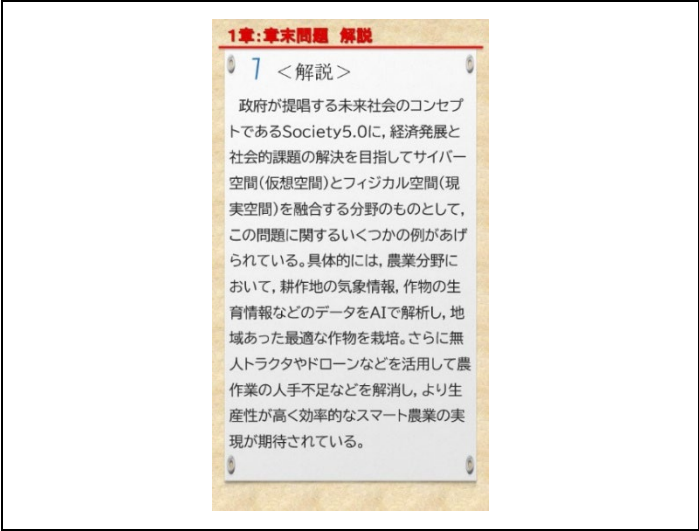
1



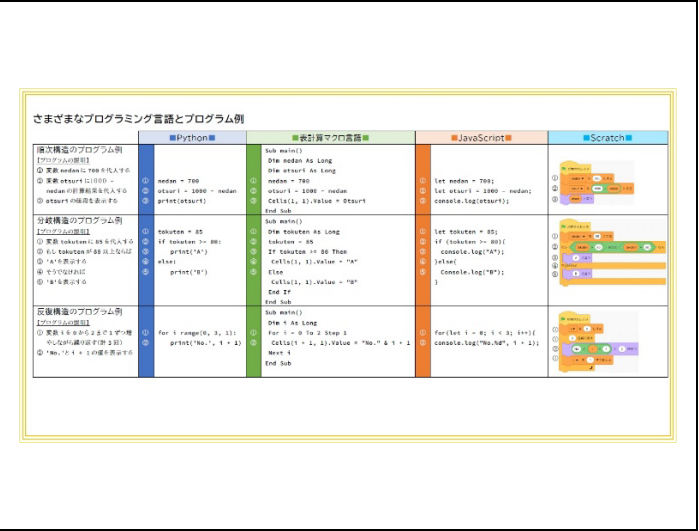
2



3



4



5

