岐阜数学教育研究
2018, Vol 17, 70-108

高校数学におけるデータ分析に関する授業実践

太田成美[1]，菱川洋介[2]，山田雅博[2]

これからの社会で活躍する生徒に求められる力の一つとして，自分が直面する問題に関するデータを正しく収集し，適切な方法によってデータを分析し，結果を考察することで問題解決する力が必要であると考える。そこで著者は，実際のデータや身近なデータを用いて分析・考察する機会を通して，データ分析の実用性を感じられる授業づくりが必要であると考え，中学校 1 年生から高校 3 年生を対象とした教材の開発と実践を行った。本論文では，開発した教材の内容を説明し，実践した結果およびその考察について報告する。

＜キーワード＞データの散らばり，標準偏差，相関，散布図，相関係数

1.　はじめに

本研究では，これからの情報化社会を生きる高校生に，様々なデータを扱う場面において，適切な方法でデータを分析し，結果を考察する力を身につけさせるための教材開発を行った。近年，我々の生活の殆ど至るところで情報文化が根付いてきたことから，社会の情報化が急速に発展してきたと言える。それゆえに，これから社会で活躍する生徒に求められる力の一つとして，自分が直面する問題に関するデータを正しく収集し，適切な方法によってデータを分析し，結果を考察することで問題解決する力が必要であると考える。実際，中央教育審議会の平成 28 年 12 月 21 日の第 109 回総会における「幼稚園，小学校，中学校，高等学校及び特別支援学校の学習指導要領等の改善及び必要な方策等について（答申）」では，「数学の学びを社会生活で活用する場面として，統計に関する学習を充実させていくことが重要である。」と記述されている。統計分野の学習を社会生活で活用できるような授業づくりの工夫が求められていることが分かる。しかし，現在の教科書の内容だけでは，統計分野の用語の定義を確認し計算することは出来るが，作業的であり，果たして実用性を感じられるのだろうかと疑問を抱いた。

そこで著者は，実際のデータや身近なデータを用いて分析・考察する機会を通し，データ分析の実用性を感じられる授業づくりが必要であると考え，本研究を進めるに至った。

本論文では，データを様々な方法で分析する活動を通して，生徒たちが身近なデータに興味・関心をもち，自分でデータを分析し考察することを目指した教材開発と実践を報告する。

2.　教材について

2.1　教材の概要

本教材について説明する。本教材では下記の 3 つをねらいとする。

(A)　データの分析に有効な手法について正しく理解し活用することができる。

(B)　身近なデータや，データを分析することに興味・関心をもち統計の有用性を感じることができる。

[1] 岐阜大学大学院教育研究科
[2] 岐阜大学教育学部

(C) 必要なデータや手法などを自分で選択し，様々な面から分析・考察することができる。

本教材は，身近であり興味の持ちやすいデータとして，スマートフォンの利用時間や視力などを取り扱う。その目的は，ねらい(B)に記す通り，生徒に興味・関心をもたせるためである。

この教材では大きく，1 変量のデータを分析する内容と，2 変量のデータの間の関係を分析する内容を取り扱う。1 変量のデータを分析する内容では，中学校の既習内容である平均値，範囲，四分位数，箱ひげ図に加え，偏差，分散，標準偏差を取り扱う。2 変量のデータの間の関係を分析する内容では，視覚的に分析することのできる散布図や，数値的に分析することのできる共分散，相関係数を取り扱う。さらに，社会生活で活用する場面が想像しやすいようなテーマを設定し，2 変量のデータの関係を相関係数や散布図を用いてグループで分析・考察する活動を最後に設ける。この活動を通して，意見の交流や議論の場面から生徒が自分の考えだけでなく他者の考えにも触れ，生徒にとってより深みのある学習となるようにしたいと考えている。

2.2 本教材における用語

本教材で用いる用語を以下に記す。この用語は，参考文献[1],[2],[3]から抜粋した。

①1 変量のデータの分析

以下，変量xについてのn個のデータをx_1，x_2，…，x_nとする。

用語 1　平均値

x_1，x_2，…，x_nの総和をnで割った値をこのデータの**平均値**といい，$\bar{x}$で表す。すなわち，n個のデータの平均$\bar{x}$は，

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

で求めることができる。

用語 2　範囲

データの最大の値から最小の値をひいた差をそのデータの**範囲**という。

用語 3　分布

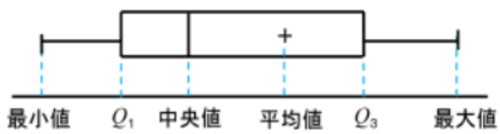
データの散らばりの様子を**分布**という。

用語 4　四分位数

データを値の大きさの順に並べ，4 等分する位置の値を**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。すなわち，Q_2より小さいデータ（下位のデータ）の中でさらに中央値をとったものがQ_1，大きいデータ（上位のデータ）の中でさらに中央値をとったものがQ_3である。

用語 5　箱ひげ図

データの分布の特徴を，5 つの値（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）を用いて表した下記のような図を**箱ひげ図**という。



箱ひげ図を用いることで，視覚的に分布を捉えたり，箱ひげ図を並べて複数のデータを比較したりすることができる。

用語 6　偏差

x_1，x_2，⋯，x_nの平均値を$\bar{x}$とするとき，各値と平均値との差を$x_i-\bar{x}$で表し，それぞれ平均値からの**偏差**という。

用語 7　平均偏差

x_1，x_2，⋯，x_nの平均値を$\bar{x}$とするとき，各値と平均値との差$x_i-\bar{x}$のそれぞれの絶対値の平均値を**平均偏差**という。変量xのデータの平均偏差は

$$\frac{1}{n}\sum_{i=1}^{n}|x_i-\bar{x}|$$

で求めることが出来る。

用語 8　分散

x_1，x_2，⋯，x_nの平均値を$\bar{x}$とするとき，各値と平均値との差$x_i-\bar{x}$のそれぞれの 2 乗の平均値を**分散**という。これをVとすると，

$$V=\frac{1}{n}\sum_{i=1}^{n}(x_i\ \ -\bar{x})^2$$

である。

このとき，分散と平均値に対して以下の公式が成り立つ。

公式 1

データx_1，x_2，⋯，x_nによって与えられる変量x^2のn個のデータx_1^2，x_2^2，⋯，x_n^2を考える。このとき，次の式を満たす。

$$V=\frac{1}{n}\sum_{i=1}^{n}x_i^2-\ (\bar{x})^2$$

つまり，
(データxの分散)

$=(x^2$ の平均値$)-(x$の平均値$)^2$

である。

用語 9　標準偏差

分散の正の平方根$\sqrt{V}$を**標準偏差**といい，sで表す。

②2 変量のデータの間の関係の分析

以下，変量xについてのn個のデータをx_1，x_2，⋯，x_n，変量yについてのn個のデータをy_1，y_2，⋯，y_nとする。

用語 10　相関

2 変量$(x,\ y)$のデータの間に，一方が増加すればそれにしたがって他方も増加する，または，他方が減少するという傾向がみられるとき，2 変量の間に**相関**がある，または**相関関係**があるという。 2 変量$(x,\ y)$のデータの間に，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 変量の間には**正の相関**があるという。また，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 変量の間には**負の相関**があるという。どちらの傾向もみられないときは，**相関がない**または**相関関係がない**という。

用語 11　散布図

2 変量からなるデータの点$(x,\ y)$を平面上に図示したものを，**散布図**という。

散布図をかくことで 2 変量の相関関係を視覚的に捉えることができる。

用語 12　共分散

n個の点の偏差の積$(x_i-\bar{x})(y_i-\bar{y})$の平均値を**共分散**といい，$s_{xy}$と表す。すなわち，

$$s_{xy}=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})$$

用語 13　相関係数

s_xをxの標準偏差，s_yをyの標準偏差，s_{xy}を共分散とする。共分散s_{xy}を各標準偏差s_x，s_yの積s_xs_yで割った値を相関係数といい，rで表す。

$$r=\frac{s_{xy}}{s_xs_y}$$

相関係数rの値は常に$-1\leq r\leq 1$であり，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。相関がないとき，rは 0 に近い値をとる。

2.3 教材の詳細

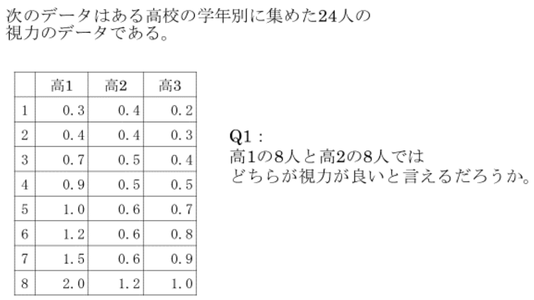
本教材を大まかに 3 つの場面に分けて説明する。

① 1 変量のデータの分析について
② 2 変量のデータの間の関係の分析について
③ グループでデータの分析・考察

以下，各場面について詳しく説明する。

①1 変量のデータの分析について

まず，1 変量のデータとして図 1 の視力のデータを提示し，分析する活動を行う。視力のデータを用いる理由は，生徒たちにとって身近なデータであると考えたからである。


次のデータはある高校の学年別に集めた24人の視力のデータである。

	高1	高2	高3
1	0.3	0.4	0.2
2	0.4	0.4	0.3
3	0.7	0.5	0.4
4	0.9	0.5	0.5
5	1.0	0.6	0.7
6	1.2	0.6	0.8
7	1.5	0.6	0.9
8	2.0	1.2	1.0

Q1：
高1の8人と高2の8人ではどちらが視力が良いと言えるだろうか。

図 1　視力のデータ

図 1 のように，高 1～高 3 それぞれ 8 人，計 24 人の視力のデータを提示し，どの学年の視力が良いといえるか生徒に問う。分析の方法として，生徒は既習である平均値や範囲を用いることが予想される。それゆえに，中学校で既習である平均値と範囲について，この場面で復習する。図 1 の高 1 と高 2 の問題は平均値を用いて比較すると高 1 の方が大きくなり，「高 1 は高 2 に比べて視力が良いと言える」と結論付ける生徒の姿が予想される。一方，高 2 と高 3 のデータを比べると，平均値と範囲がともに等しくなってしまい，比べることが難しい。この場面を扱うことで，平均値や範囲だけで全てのデータを比較できるとは限らないことを生徒に感じさせたい。

次に，平均値や範囲では比較できない問題を解決するために，図 2 のように各学年のデータを数直線にドットプロットし，データの散らばりに着目させる活動を行う。この活動は，平均値や範囲が等しくても，散らばり具合が違うことに気づかせることが目的である。

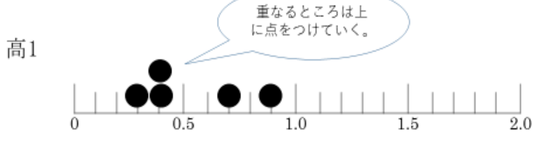


図 2　ドットプロット

この活動を通して，平均値のような代表値に加え，データを比較するには散らばり具合に着目する必要があることに気付かせたい。

その後，散らばり具合を表す新たな手法として，四分位数と箱ひげ図を定義し，高2のデータと高3のデータを比較させる。まず図3を用いて四分位数を定義する。

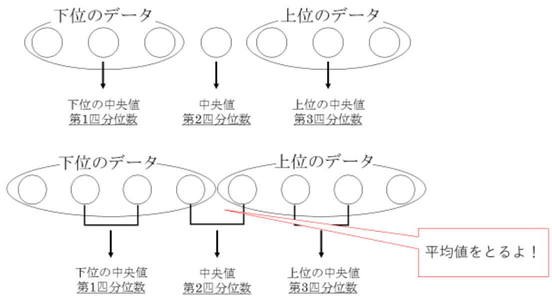


図3　四分位数

四分位数の求め方は，データの個数が偶数の場合と奇数の場合によって異なるため，図3のような1つ1つのデータを〇で表したモデルを見せ，正しい理解を図った。次に，四分位数を用いて，箱ひげ図を作成する。

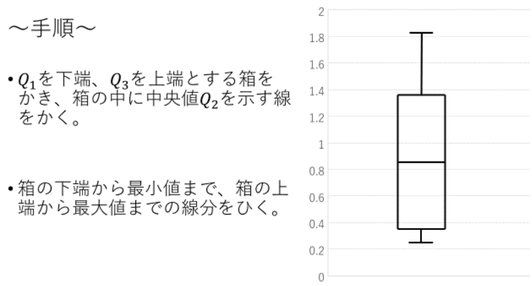


図4　箱ひげ図の作成

図4のような箱ひげ図をスライドで提示し，アニメーションをつけることで，箱ひげ図を描く手順が分かりやすくなるよう工夫をした。箱ひげ図を用いるメリットは，データの範囲やデータの広がり具合を視覚的に捉えることができるということや，複数のデータがあった場合に並べて見ることで比較しやすいことがあげられる。このことは作図を通して生徒に伝えられると考える。一方，四分位数や箱ひげ図では反映されていないデータが存在するデメリットがあることにも気付かせたい。このことから，四分位数と箱ひげ図を用いることで，データの集団同士に関する一定の比較はできるものの，個々のデータを分析・考察に反映させることには不向きであることを結論付ける。

このデメリットにより，偏差の導入がよりスムーズに行えると考えている。まず，「すべての値を反映させることができるものはないか」という考えから，データの各値と平均値の差に着目させる。偏差の平均をとることで，もとのデータの平均値からデータの各値が平均してどれくらい離れているか，すなわち，データが散らばっているのか平均値の周りに集まっているのかが，分かるのではないかと生徒に投げかけ，偏差の必要性を伝える。実際に図1の視力のデータで偏差を求め，偏差の平均値を計算させる。しかし，偏差とは各値から平均値をひいたものなので，当然負の数となることもある。また，平均値とは各値を均したものであるため，偏差の平均値を求めると0になる。そのため，散らばり具合を数値として比較することが出来ない。そのことを実感させたうえで，どうしたら0でなくなるか意見を求め，絶対値をとる，もしくは，2乗することで負の数を正の数に改められることを確認する。この場面を扱うことで，次に扱う，各偏差の絶対値の平均値である平均偏差，各偏差の2乗の平均値である分散，その正の平方根である標

準偏差の定義について理解を深めることができると考えている。

　この活動を踏まえた上で，分散と標準偏差を定義する。分散の値の計算を定義通りに進めると，計算が複雑になることから，計算がより簡潔になるように，公式１について紹介し，実際に活用させる。ここでは，公式１について，分散の定義より証明させる活動を取り入れる。

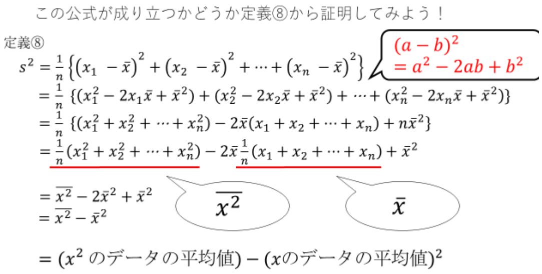


図５　分散の定理の証明

この証明では，$(a-b)^2$の展開を利用するが，まだ習っていない中学生がいることを想定して，図５のように展開の仕方をスライドで紹介し，細かく証明することにした。

　ここで，全員に電卓を配布し利用させる。以降の計算をより素早く行えるようにするための配慮である。電卓の便利な活用法として，メモリー機能について紹介する。メモリーという仮想の箱をスライドで用い，そこに数字を入れたり抜いたりするアニメーションをつけ，M+やM－などをイメージしやすくなるよう工夫をした。

②2 変量のデータの間の関係の分析について

　ここでは２変量$(x,\ y)$のデータの間の関係を取り扱う。①で扱っていた視力のデータに加え，スマートフォンの利用時間のデータを提示し，その２つのデータの間にはどのようなことが言えそうか分析するための手法について触れていく。まず，相関について定義し，思いつく限りの相関のありそうな例を挙げてもらう。この活動は，新しく定義した相関という関係について慣れてもらうことと相関について興味をもってもらうことが目的である。生徒が思いついた相関の例を全体で共有することでさらに興味・関心を高めたい。

　次に散布図について定義し，データを表す点$(x_i,\ y_i)$の散らばりに着目させる。「一方が増加すると他方も増加する」「一方が増加すると他方が減少する」という相関の定義より，データを表す点$(x_i,\ y_i)$の散らばりが直線に近いほど相関が強いことを確認する。

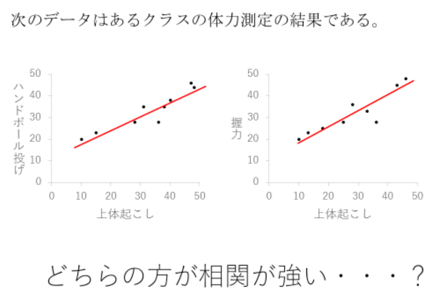


図６　散布図の比較

　次に，図６のような２つの散布図を提示し，「どちらの方が，相関が強いと思うか」という質問を生徒に投げかける。この質問の意図は，散布図が複数ある場合，相関の強さは，散布図を見るだけでは比較しにくいということに気付かせるためである。この活動を通して，相関の強さを数値化する必要性を感じさせたい。数値化する準備として，$(\bar{x},\ \bar{y})$を中心に散布図を４領域にわけ，各領域で共通点をまとめる。

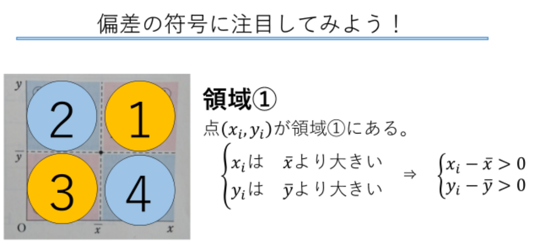


図７　４領域に分けた散布図

点の散らばりが領域①と領域③に集まった場合の相関は正の相関であり，領域②と領域④に集まった場合の相関は負の相関であることを確認し，分かりやすくなるよう図７の散布図のモデルのように，色をつける工夫をした。次に，それぞれの領域において，$(\bar{x}, \bar{y})$を中心としていることから，点(x_i, y_i)の偏差に着目させる。「それぞれの偏差にどのような特徴があるのか」という質問を投げかけたうえで，図７のようにその領域にある点の偏差の符号についてまとめる。偏差の符号に着目させることで，領域①と領域③，領域②と領域④の共通点に気付き，共分散への理解が深まると考えている。さらに，下に示した図８のスライドを使った説明の中にアニメーションを活用し，それぞれの領域について，偏差の符号によってその領域の偏差の積はどうなるのか分かりやすくなるよう工夫した。

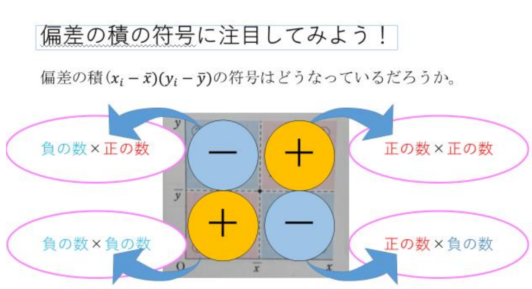


図８　各領域の偏差の積の符号

次に，相関の強さは等しそうであるが共分散の値が違うデータを提示する。それぞれのデータの分布を確認するために散布図をかき，視覚的に相関の強さが等しいことを確認させ，「なぜ共分散の値に違いがでてしまったのか」と質問を投げかける。この質問の投げかけによって，相関係数の必要性を感じながら，相関係数の定義の場面を導入できると考えている。具体的には，共分散では規模の違うデータは比べられないことを確認し，データの種類や単位に影響されない値が必要であると感じさせることである。

そのような値として相関係数を紹介する。相関係数は，共分散を各標準偏差の積で割ることで，値が必ず$-1 \leq r \leq 1$になる。すなわち，データの種類や単位に関係なく数値の範囲を統一することができる。また，相関係数が－1や1に近いほど相関が強いことを確認する。以上のことから，相関係数を用いることで，相関の強さは数値化でき，明確に比較することができる。実際に視力のデータとスマートフォンの利用時間のデータの相関係数を求め，比較させる。

③グループで分析・考察

この場面では，生徒を４～６人程度のグループに分け，グループごとでデータの分析及び考察を行わせる。あらかじめ用意しておいた都道府県ごとの様々なデータから，相関の気になる２つのデータを選ばせる。そのようにした理由は，データを授業者が選択し疑問を投げかけるのではなく，生徒たちが気になるようなデータを自ら選択させることで，より積極的に取り組めるのではないかと考えたからである。次にそ

のデータを選んだ理由や，相関はありそうかどうかという予想を立てさせ，実際のデータをもとに本時で学習した内容を自由に活用し，相関関係について調べる活動を行う。そして，ただ相関関係を調べて完結するのではなく，問題解決や改善に繋げるため，どうしてそのような結果が出たと考えられるか，何が原因となっているかなど様々な視点から分析・考察してもらう活動を設ける。グループごとで分析・考察した内容を模造紙にまとめさせ，全体交流の場で発表させる。自分のグループの気になったデータだけでなく，他のグループの発表を聞くことで，データ分析への興味がさらに深まることを期待している。

3.　授業実践について

場所：岐阜大学教育学部 A 棟 426 教室
日程：平成 28 年 10 月 4 日(木)，90 分
　　　　　　　　11 日(木)，90 分
対象：岐阜大学教育学部数学教育講座
　　　1 年生　27 人

本教材は高校生を対象として開発したが，実践は岐阜大学の教育学部数学教育講座の 1 年生を対象として行った。対象は本実践に関わる統計の内容を既習している。しかし，授業内で用語や定義，またどのような場合に用いるかを説明できるか質問したところ，2 名の学生しか答えられなかったことから，本実践の内容に関する深い知識の定着はしていないと感じた。ゆえに，本教材の実践によって，統計に関する理解が深まる良い機会になったと捉えている。

3.1 授業のねらい

授業のねらいについて改めて述べる。

(A) データの分析に有効な手法について正しく理解し活用することができる。
(B) 身近なデータや，データを分析することに興味・関心をもち統計の有用性を感じることができる。
(C) 必要なデータや手法などを自分で選択し，様々な面から分析・考察することができる。

また，活動を通して，自分の力でデータを分析したり予想と結果を比較したりすることで，達成感や統計学の楽しさを感じてほしいと考えている。

3.2 授業における生徒の姿

時間の都合上，説明を割愛した箇所や作業を減らした箇所があったが，ほとんどの生徒が本時取り扱った内容について理解することができていた。また積極的に授業に取り組む姿が多くみられ，声をかけなくても周りと確認する姿や教えあう姿がみられた。本来，活動の最後にグループごとで模造紙にまとめ発表することを予定していたが，グループ活動での計算に時間がかかってしまい実施できなかったため，グループごとで A4 の紙にまとめてもらいレポート形式で提出してもらった。以下，2.3 で示した 3 つの場面にわけて詳細を記述する。また，本実践の指導案及び学習プリントは，本論文の末尾に掲載することとする。

①について

学習プリント①で多くの生徒はデータを比較する際に平均値を思い浮かべ実際に計算し求めていた。平均値以外にも，範囲や中央値，箱ひげ図をかいている生徒もいたが数名であった。

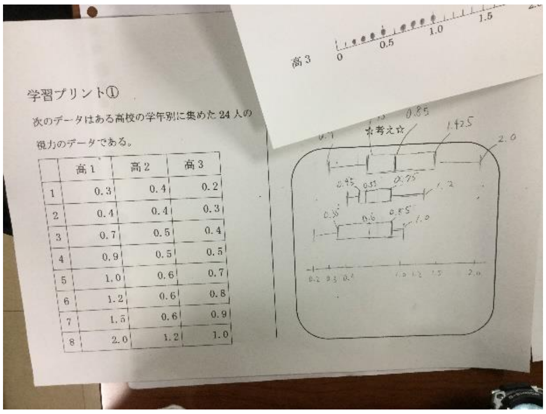


図 9　箱ひげ図を用いて考える生徒

平均値や範囲などはよく理解できていたが，四分位数については高校で既習であるとはいえ，説明できると自信をもって回答したのは全体の中で 2 名だけであった。しかし，図 3 のモデルで説明し，机間指導も行うことでほとんどの生徒が正しく四分位数を求めることが出来ていた。各偏差の平均をとると 0 になってしまうが，0 にならないためにはどうすればよいだろうかという問いでは，絶対値をとることや 2 乗することが意見として出ることを期待したが，意見が出るまでには至らなかった。

分散や標準偏差の定義について質問したところ，定義があいまいな生徒がほとんどだった。この姿から，高校在学時に学んでいるものの，分散や標準偏差の知識について正しく理解できていないと感じた。分散や標準偏差を導入する必要性を確認したうえで改めて定義すると，うなずきながら聞く生徒の姿が見られた。定義をただ教え込むのではなく，「散らばり具合を数値化するために各偏差の平均をとる必要があり，ただ各偏差の平均をとるだけでは値が 0 になってしまうため，各偏差を 2 乗して平均をとる。それが分散である」というように，定義をする動機の流れを詳細に説明することで理解が深まったのではないかと感じた。分散の値を簡潔に求める公式 1 を証明する場面では，数学教育講座の学生であることから，すらすらと証明していた。つまっている生徒にも机間指導で少し声をかけると，すんなりと進めることができていた。実際に分散の公式 1 の証明を高校生へ取り扱う場合には，対象の習熟度などを考慮したうえで取り扱うべきである。

電卓の便利な活用法では，生徒は今まで使ったことのない機能だったため非常に興味を示していた。機能をひとつひとつ説明していくと，生徒の中からその場で実際に使ってみて感心する声が聞こえていた。また，慣れるまでには少し時間がかかり，練習問題では苦戦する姿や教えあう姿が見られた。

②について

相関という用語に生徒があまり慣れていないと考え，相関がありそうな例をあげてもらった。たくさんの例はあがらないだろうと思っていたが，授業に積極的な生徒が多く，時間いっぱいまで考える姿が見られた。実際に生徒は，『年齢と髪の量』，『足の長さと足の速さ』など数多くの例を挙げた。特に，『バイト時間と GPA』のような大学生ならではの例もあり，大変興味深かった。この姿から，変量の扱い方や相関という用語の理解ができていると感じた。

一方，共分散の定義では講義形式になってしまい，授業への意欲が落ちてしまった生徒もいた。作業などを授業の間に挟むことは，集中力を切らすことなく授業をする上で有効だと感じた。

相関係数の定義の理解を確かめる練習問

題を解く場面では，計算が複雑であったことから，生徒の半数近くが混乱していた。提示した問題の数値の設定をもう少し簡単にするなどの工夫をするべきであったと考えている。

③について

座っていた席の近くで 1 グループ 4 人～6 人となるようなグループを作った。普段の大学生活でお互いを知っているということもあり，スムーズに活動に入ることが出来ていた。役割分担もしっかりしており，効率的な活動ができていたように感じる。

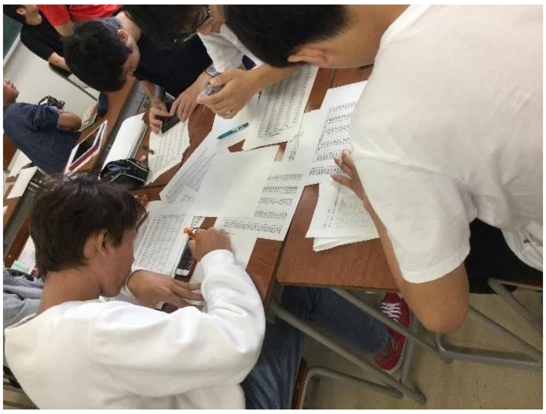


図 10　グループ活動の様子

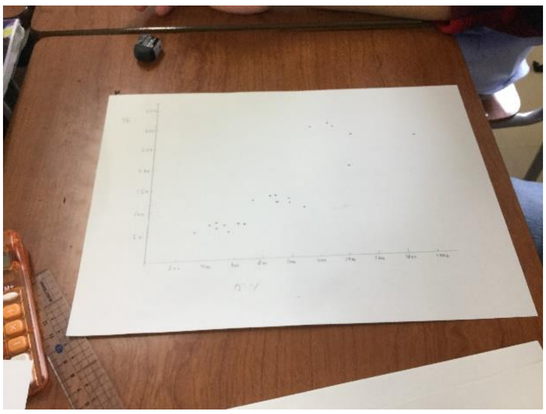


図 11　散布図を使った分析

また図 11 のように，相関係数だけでなく散布図も描き，相関係数と散布図から分析するグループもあった。分析結果や考察，またどうしてそのデータを調べてみたいと思ったかなどをレポートにして提出させた。5 グループあるうち，3 グループが自動車保有台数とガソリンスタンドの軒数の相関を調べていた。提出されたレポートを見ると，車の台数が多いとガソリンスタンドもたくさん必要だろうと予想していた。実際に相関係数を計算すると約 0.9 となり，47 都道府県では強い相関がみられると結論付けることができていた。他にも，夫婦は自分たちの家を買いたくなるのか気になり，夫婦数と持ち家数の相関を調べるグループもあった。また，ただ相関を調べるだけでなく，どうして夫婦の数と持ち家の数に相関があったのか，さらには持ち家を持つことでローンが発生することから，持ち家のある夫婦はお金を稼ぐ目的で仕事にしっかり打ち込めるのではないかという推測までたてることができていた。このように，どのグループも相関を調べるだけではなく，どうして相関があったのか，もしくは，なかったのかという理由の推測まで考察することが出来ていた。

3.3 アンケートの結果とその考察

生徒には，2 日間の授業後にアンケートを実施した。以下にアンケートの結果と考察を述べる。

Q1：資料・データなどを扱うような数学は好きですか。

1 好き・・・4 人

2 どちらかといえば好き・・・17 人

3 どちらかといえば嫌い・・・4 人

4 嫌い・・・2 人

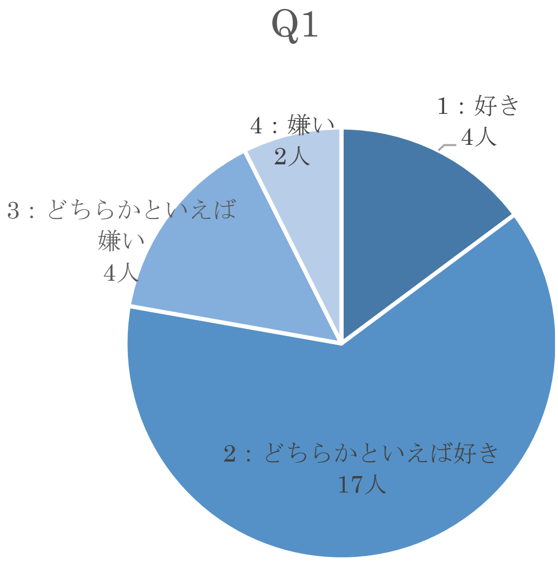


図 12　Q1 の円グラフ

円グラフで表すと図 12 のようになった。約 73%の生徒が，データを扱うような数学に対して肯定的な回答をしていた。しかし 6 人はどちらかといえば嫌い，もしくは，嫌いと回答している。こちらにも着目してアンケートを読み進めた。半数以上が統計について好感を抱いていたためか，今回の実践では積極的に授業に取り組む姿が多くみられた。

Q2：今回の授業について理解できましたか。

1 理解できた・・・19 人

2 まあまあ理解できた・・・8 人

3 あまり理解できなかった・・・0 人

4 理解できなかった・・・0 人

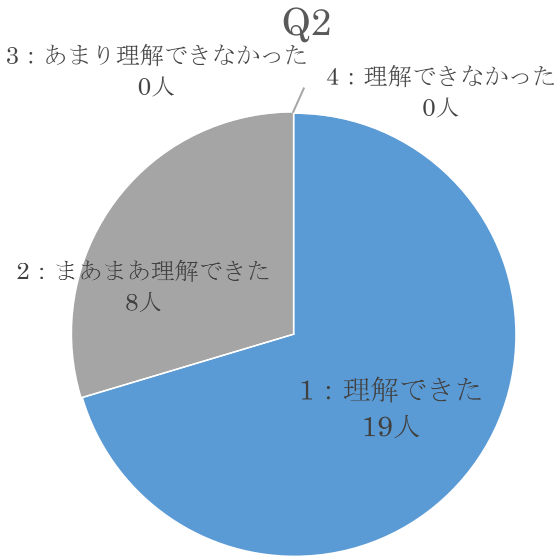


図 13　Q2 の円グラフ

円グラフで表すと図 13 のようになった。あまり理解できなかった，もしくは，理解できなかったと回答した生徒はともに 0 人であった。しかし，対象が岐阜大学教育学部数学教育講座の学生であり数学に対して好意的であるということ，また，すでに数Ⅰで既習であるということが大きく影響していると考えられる。授業で確認してみると，覚えていないところや正確に理解していないところなどが見受けられたので，今回の授業で理解がより深まったのではないかと感じた。しかしこの結果に満足せず，より分かりやすい教材となるよう習熟度に合わせてペースを配分したり，確認問題を用意したり改善していきたいと考える。

Q3：今回の授業で難しかったところを教えてください。

この問いは記述形式にした。主に目立った回答である電卓についてと相関係数の内容について述べていく。

一番多かった回答は電卓についてである。電卓の初めて知る機能をうまく使いこなせなかったという意見や，打ち間違えずに電卓で計算することが難しかったという意見を 8 人が回答していた。電卓とは限らず計算が難しかったという回答もあった。また，少数や負の数があり，計算が難しかったという回答や，データの数が多すぎたという意見もあった。練習問題では，もう少し簡単な数字を扱った方がよさそうだと感じた。

他には，相関係数rが常に$-1 \leq r \leq 1$となる理由が分からなかったという回答があった。それについては，混乱を危惧したことや時間の都合を踏まえて割愛したのだが，やはり気になる生徒もいるため，対象の学力レベルによっては解説しても良いと感じた。

Q4：2 変量のデータの間の関係について調べることに興味・関心は深まりましたか。
1 深まった・・・10 人
2 まあまあ深まった・・・15 人
3 あまり深まらなかった・・・1 人
4 深まらなかった・・・1 人

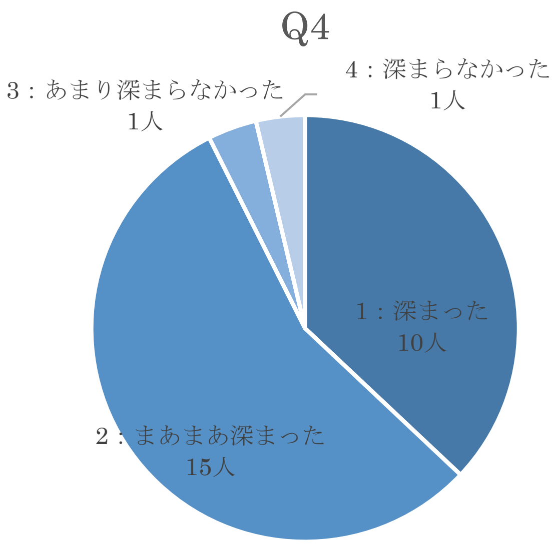


図 14　Q4 の円グラフ

円グラフで表すと図 14 のようになった。約 92%の生徒が深まった，もしくは，まあまあ深まったと肯定的な回答をしている。また，注目してほしいのは Q1 では否定的だった生徒も少なからず興味・関心は深まったと回答していることだ。ここから，今回の実践は有効であったのではないかと考えられる。しかし，深まらなかったと回答している生徒もいるので，他にもアプローチ方法を考え実践できたらよいと感じた。

Q5：日常生活で今回学んだことを使ってみようと思いましたか。
1 思った・・・5 人
2 まあまあ思った・・・17 人
3 あまり思わなかった・・・3 人
4 思わなかった・・・2 人

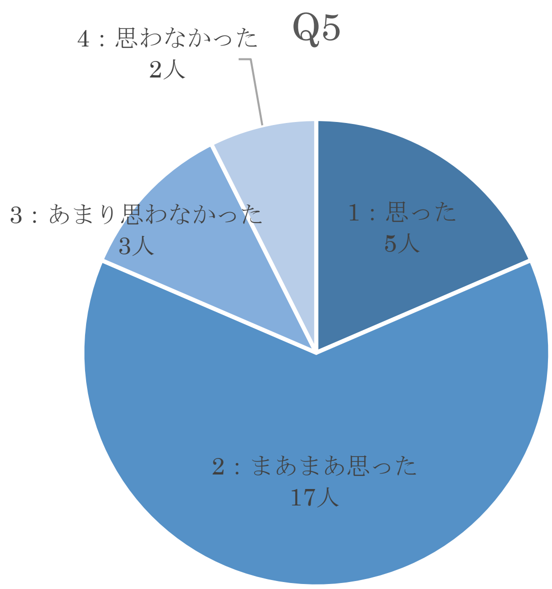


図 15　Q5 の円グラフ

円グラフで表すと図 15 のようになった。約 81%の生徒が肯定的な回答をしているが，日常生活で使ってみようとは思わなかった生徒が 5 人いた。Q2 で，理解できた，もしくはまあまあ理解できたと回答していた生徒のなかにも，日常生活で使ってみようとは思わなかった生徒が 5 人もいたことがわかる。また，興味・関心は深まったが実際に使ってみようとは思わなかった生徒もいることが読み取れる。要因として考えられるのはグループでの活動に 47 都道府県のデータを用いたことでデータの個数が多く計算が大変になってしまったことが挙げられる。Q3 の難しかった点について，計算が大変だったと回答していたことからもそのような要因が考えられる。また，より日常生活で使ってみたいと思うためにも，やはりグループ活動で調べた内容を発表する時間をとるべきだったと感じた。そうすることで，自分が調べていないことでも他のグループの考えを聞き，より一層興味・関心が深まったのではないかと考えるからである。実際に，著者は各グループのレポートをみて，それぞれが導き出した結果と考察にとても関心を抱き，純粋に面白いと感じた。生徒も自分たちがデータを調べて導き出したことに達成感や関心を抱いただろう。同様に，自分たちが調べたことだけでなく他のグループの考察を聞くと盛り上がっただろうと感じた。

また，Q5 では今回学んだことを，具体的にどのようなことに使ってみようと思ったかも記述してもらった。いくつか原文のまま抜粋して紹介する。

・CD の売り上げとライブチケットの値段
・牛乳を飲んだ量と身長
・喧嘩の回数とデートの回数
・読書時間と成績
・塾の自習室に来る時間とテストの点数
・相関がありそうなものが実はなかったという様な事象を探してみたい
・人気とモテ度

著者も相関を調べてみたいと思うような具体例がたくさん挙げられていた。著者が思い浮かばなかったものばかりであり，非常に興味深く今後の授業づくりの参考になった。やはり，自分の調べたいデータの方が，関心が強くもっと利用してみたいと思うきっかけになり，“数学のよさを認識し積極的に数学を活用しようとする態度”や“粘り強く考え数学的論拠に基づいて判断しようとする態度”に繋げていくことができるのではないかと感じた。そのために，日常生活での活用をより身近に感じられるような，生活環境にさらに近いデータを準備するか，自分たちでデータを集められるような状況を作りたい。

Q6：今回学んだこと，または，感想を教えてください。

Q6 では第一著者の今後の実践や教材開発，授業方法などの参考にするために記述形式で意見や感想を募った。いくつか抜粋する。

・電卓の便利な使い方を知ることができてよかった。
・2 つのデータの関係を読み解く方法を学んだ。
・分散の公式についてしっかりと理解できたのでよかった。
・スライドもわかりやすかったし，説明もわかりやすかったです。
・平均値ではわかりにくいデータも相関係数などで捉えることができると分かった。
・2 つのデータの間の関係を見つける方法を学んだ。公式や定理の導入が唐突だと感じるところがあった。
・ほとんど高校で習ったことだったが，電卓は昔からなんだろうと気になっていたので初めて知れて便利だと思った。
・共分散をそれぞれの標準偏差で割る理由が知りたかった。便利だからなんだろうけど数学的意味はあるのかな。
・高校で学んだデータの分析の本質的な意味についてよくわかった

やはり目立ったのは電卓のメモリー機能等のことだった。約 3 分の 1 の生徒が電卓の便利な機能について感想をよせていた。昔から電卓のメモリーボタンについて気になっていたなど，知らなかった便利な機能を新しく知れたことから，このような意見がたくさん出てきたと考えられる。生徒には今後もぜひ活用していってほしい。ほかにも，高校ですでに習った内容ではあったが，定義などが曖昧であったため，正しく定義を学び直せて良かったという意見がいくつかあった。その一方で，公式の導入が唐突だったという意見や，どうして共分散をそれぞれの標準偏差で割るのか理由が知りたかったという意見もあり，強引に定義にもっていってしまった点があると感じた。今回の実践では，できるだけ躓きを減らし，何がメリットで何がデメリットなのか，デメリットを補うために何をしたのかを伝えることを意識していたのだが，ここはまだ課題点だと捉えている。

4.　終わりに

今回は，1 変量のデータについて分析する方法や 2 変量のデータについて相関を求める方法などについて取り扱った。身近なデータを取り扱ったり，作業をすることでできた達成感や習得できた手ごたえを感じさせたり，生徒が楽しいと思えるような授業開発を意識した。その結果，データについて興味をもつ姿や，積極的に授業に取り組む姿をみることができ授業のねらいの(A)(B)は達成できていたと考える。アンケートの結果からも，今回の実践は，理解を深めさらに興味・関心をもつことに有効であったといえる。(C)については回収したレポートより，一人一人が様々な考察を述べることができていたが，グループ活動で行ったため，発言しない生徒や，他者に委ねる生徒もでてきてしまうと感じられた。グループの作り方を工夫したり机間指導で促したり，また，評価方法についても考えなければならないと感じた。今後，高校数

学では以前に増して統計学を重視し，社会にでても求められる機会が多くなることが予想される。少しでも統計学に対して，面白い，楽しい，興味深いと思ってもらえるような教材開発を今回の課題点を踏まえ行っていきたい。

引用・参考文献

[1] 大島利雄　他 12 名，高等学校数学 I (平成 23 年 3 月 9 日検定済)　数研出版株式会社

[2] 小寺平治，新統計入門(2015)，株式会社裳華房

[3] 猪野富秋，伊藤正義，数理統計入門(1981)，森北出版株式会社

[4] 御園生善尚　他 4 名，統計学大要(1986)，株式会社養賢堂

[5] 文部科学省，高等学校学習指導要領（平成 30 年 3 月公示）

[6] 文部科学省，高等学校学習指導要領解説(平成 21 年 11 月）

[7] 文部科学省，中学校学習指導要領(平成 29 年 3 月）

[8] 都道府県別統計とランキングで見る県民性[とどラン] https://todo-ran.com/

[9] 内田学，兼子良久，仕事が 10 倍速くなる！統計学の活かし方，PHP ビジネス新書

学習指導案（略案）

1．本時のねらい

本時のねらいは以下の 3 つである。

(A) データの分析に有効な手法について正しく理解し活用することができる。

(B) 身近なデータや，データを分析することに興味・関心をもち統計の有用性を感じることができる。

(C) 必要なデータや手法などを自分で選択し，様々な面から分析・考察することができる。

2．本時の展開

<table>
<tr><th></th><th>学習内容</th><th>指導・援助</th></tr>
<tr><td>導入</td><td>
相関関係のありそうなデータを紹介する。
・おしゃれ努力度と自由時間
・スマートフォン利用時間と視力
<table>
<tr><th></th><th>高 1</th><th>高 2</th><th>高 3</th></tr>
<tr><td>1</td><td>0.3</td><td>0.4</td><td>0.2</td></tr>
<tr><td>2</td><td>0.4</td><td>0.4</td><td>0.3</td></tr>
<tr><td>3</td><td>0.7</td><td>0.5</td><td>0.4</td></tr>
<tr><td>4</td><td>0.9</td><td>0.5</td><td>0.5</td></tr>
<tr><td>5</td><td>1.0</td><td>0.6</td><td>0.7</td></tr>
<tr><td>6</td><td>1.2</td><td>0.6</td><td>0.8</td></tr>
<tr><td>7</td><td>1.5</td><td>0.6</td><td>0.9</td></tr>
<tr><td>8</td><td>2.0</td><td>1.2</td><td>1.0</td></tr>
</table>
既習である平均値・範囲に触れる。
○高 1 と高 2 では，どちらのほうが視力が良いと言えるだろうか。
・平均値で比べれば分かるのではないか。
・高 1 の視力の平均値は 1.0 である。
・高 2 の視力の平均値は 0.6 である。
・高 1 と高 2 では高 1 のほうが視力が良いのではないか。

定義①　平均

○高 2 と高 3 では，どちらのほうが視力が良いと言えるだろうか。
・高 3 の視力の平均値は 0.6 である。
・高 2 と高 3 では平均値が同じで比べられない。

定義②　範囲
</td><td>
・興味のありそうなデータを持ってきて関心を高める。

・既習である平均値を活用できているか確認する。

・範囲が大きければ大きいほど，データ
</td></tr>
</table>

	・範囲も同じで比べられない。 ○数直線を用いて自分たちで点をとる。 ・数直線をみると，散らばり具合に差がある。 ・高 2 のほうが，データに散らばりがある。 定義③　分布 ○範囲以外にデータの分布を表す数値や図にはどんなものがあるだろうか。 定義④　四分位数 定義⑤　箱ひげ図 ○高 1 のデータを箱ひげ図に表してみる。 ○高 2 と高 3 の視力のデータを箱ひげ図に表してみる。	の散らばりが大きいことを確認する。 ・平均値も範囲も同じ場合，何が違うのか気付かせる。 ・範囲も分布を表していることに注意する。 ・範囲は等しかったが，箱ひげ図にしてみると，箱の位置のかたよりに差があることに気付かせる。
展開	○四分位数では，すべてのデータを反映できているだろうか。 データの値すべてを用いて散らばりの具合を表すことができる値はないか考えてみよう。 定義⑥　偏差 ○偏差の平均値を求めてみよう。 ・0 になってしまう。 ○0 にならないためにはどうしたらよいか考えてみよう。 ・絶対値をとる。 ・2 乗してから平均値をとる。 定義⑦　平均偏差 定義⑧　分散 定義⑨　標準偏差 ○電卓を使って，高 1 の視力のデータの平均偏差・分散・標準偏差をそれぞれ求める。 発表	・四分位数は，第 1 四分位数，第 2 四分位数，第 3 四分位数を使っており，すべてのデータを反映できていないことを確認する。 ・第 2 四分位数は中央値と同じであることに触れておく。 ・偏差は平均との差なので，正の値と負の値がでてしまい，足すと常に 0 になってしまうことを確認する。 ・電卓を配っておく。

展開	公式①　分散の公式 ○どうしてこの式で求めることができるのか定義⑧から求めてみよう。 電卓の便利な使い方を紹介する。 **「M＋」（メモリープラス）**：電卓に表示されている数字を，メモリーに足す。 **「M-」（メモリーマイナス）**：電卓に表示されている数字を，メモリーから引く。 **「MRC」（メモリーリコール/メモリークリア）**：1 度押すとメモリー内容を呼び出し，もう 1 度押すとクリア（0 に）する。 **「×＝」（２乗）**：その数字の 2 乗の値を出す。 ○練習問題を各自やってみる。 (1)　$2\times3+4\times5=$ (2)　$2\times3+4\times5+6\times7=$ (3)　$4\times5-2\times3=$ (4)　$2\times3+4\times5+8\times9-6\times7=$ (5)　$27\times3-54\times15-12\times62=$ (6)　$(12\times12-9\times9)\div3=$ (7)　$24\times24+19\times19-13\times13=$ (8)　$124\times124+92\times29-13\times17=$ (9)　$41\times34-15\times11+113\times113+55\times4=$ (10)　$20\times18+7\times28-201\times8+72\times8=$ ○全員で高 1 の視力のデータの分散と標準偏差を，電卓の便利な使い方を用いて求め，一致することを確かめる。 ○高 2，高 3 の視力のデータの分散と標準偏差を，電卓をうまく活用し求めてみよう。 ここからは2 変量のデータの相関関係に注目していく。 視力とスマートフォンの使用時間に関係があるかどうかを調べるため	・証明をパワーポイントで紹介する。 ・機能をまとめた学習プリント⑨を配る。 ・パワーポイントを使って 2 問ほどやってみせる。 ・プリント⑩を配る。 ・慣れるまで時間がかかると思うので，補助員は計算の仕方をマスターしておく。 ・答え合わせの際に，押すキーの順番が分かりやすいよう工夫してパワーポイントに載せる。

	にはどうすればよいだろうか。 定義⑪　相関とは 定義⑫　散布図 ○高2と高3のデータをそれぞれ散布図に表してみよう。 ○散布図からどんなことが読み取れるだろうか。 散布図だけでなく，データの値から相関関係を読み取ることを考える。 定義⑬　共分散 定義⑭　相関係数 ○高1，高2，高3ののデータの相関係数を求め，比較してみよう。	・例として散布図をいくつかだし，どれが一番相関が強そうか考えさせる。 ・数値化する必要性を感じさせる。
グループ活動	4 人〜6 人のグループを作り，都道府県別の様々なデータの中から相関を調べたい 2 変量を選び，相関について分析・考察する。 ○どうしてその 2 変量を選んだのか，また，どうしてそのような結論に至ったのかを明確にして模造紙にまとめよう。 発表	・こちらであらかじめ用意していた都道府県別のデータを配る。 ・どうして相関があったのか，または，どうして相関がなかったのかまで考察させるよう意識する。

学習プリント①

次のデータはある高校の学年別に集めた 24 人の視力のデータである。

	高 1	高 2	高 3
1	0.3	0.4	0.2
2	0.4	0.4	0.3
3	0.7	0.5	0.4
4	0.9	0.5	0.5
5	1.0	0.6	0.7
6	1.2	0.6	0.8
7	1.5	0.6	0.9
8	2.0	1.2	1.0

Q1：高 1 の 8 人と高 2 の 8 人ではどちらが視力が良いと言えるだろうか。

☆考え☆

学習プリント②

次のデータはある高校の学年別に集めた 24 人の視力のデータである。

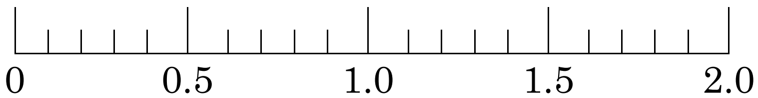
	高 1	高 2	高 3
1	0.3	0.4	0.2
2	0.4	0.4	0.3
3	0.7	0.5	0.4
4	0.9	0.5	0.5
5	1.0	0.6	0.7
6	1.2	0.6	0.8
7	1.5	0.6	0.9
8	2.0	1.2	1.0

Q2：高 2 の 8 人と高 3 の 8 人ではどちらが視力が良いと言えるだろうか。

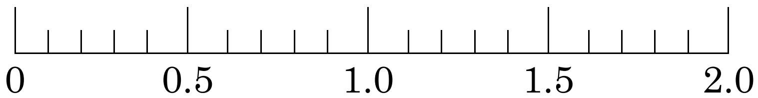
☆考え☆

学習プリント③

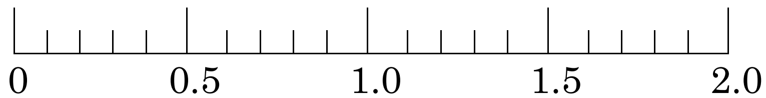
高 1



高 2



高 3



☆分かったこと☆

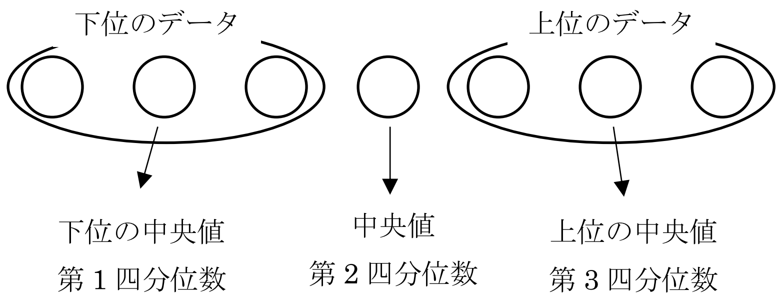
定義③　分布

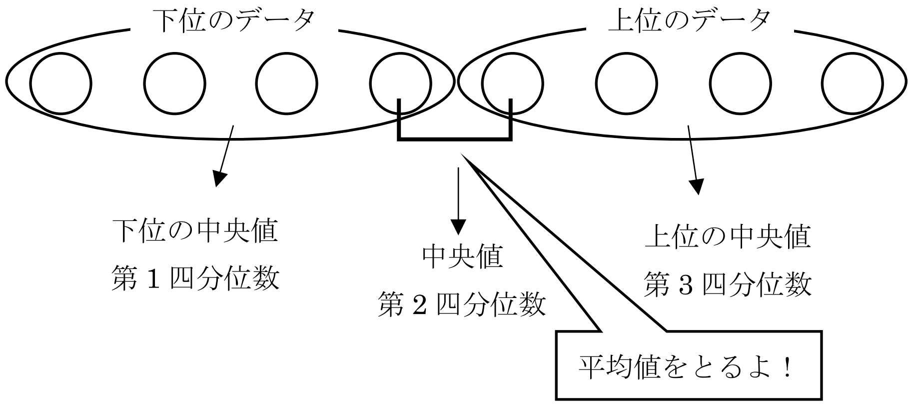
データの散らばりの様子を**分布**という。

学習プリント④

定義④　四分位数

データを値の大きさの順に並べ、4 等分する位置の値を**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Q_1 , Q_2 , Q_3で表す。Q_2より小さいデータ（下位のデータ）の中でさらに中央値をとったものがQ_1、大きいデータ（上位のデータ）の中でさらに中央値をとったものがQ_3である。





☆実際に四分位数を求めてみよう！☆

高 1

第 1 四分位数 ________________

第 2 四分位数 ________________

第 3 四分位数 ________________

高 2

第 1 四分位数 ________________

第 2 四分位数 ________________

第 3 四分位数 ________________

高 3

第 1 四分位数 ________________

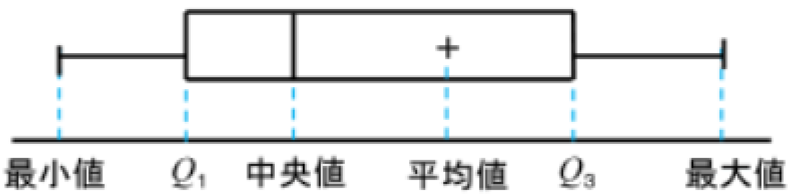
第 2 四分位数 ________________

第 3 四分位数 ________________

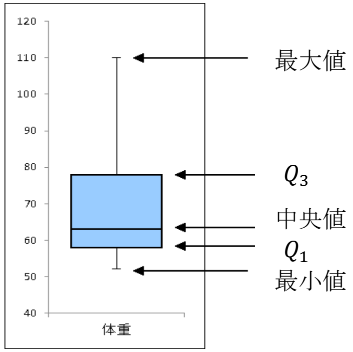
学習プリント⑤

定義⑤　箱ひげ図

データの分布を、次のような図で表すことができる。



これを**箱ひげ図**という。データの分布の特徴を、5 つの値（最小値、第 1 四分位数Q_1、中央値、第 3 四分位数Q_3、最大値）で簡明に表している。



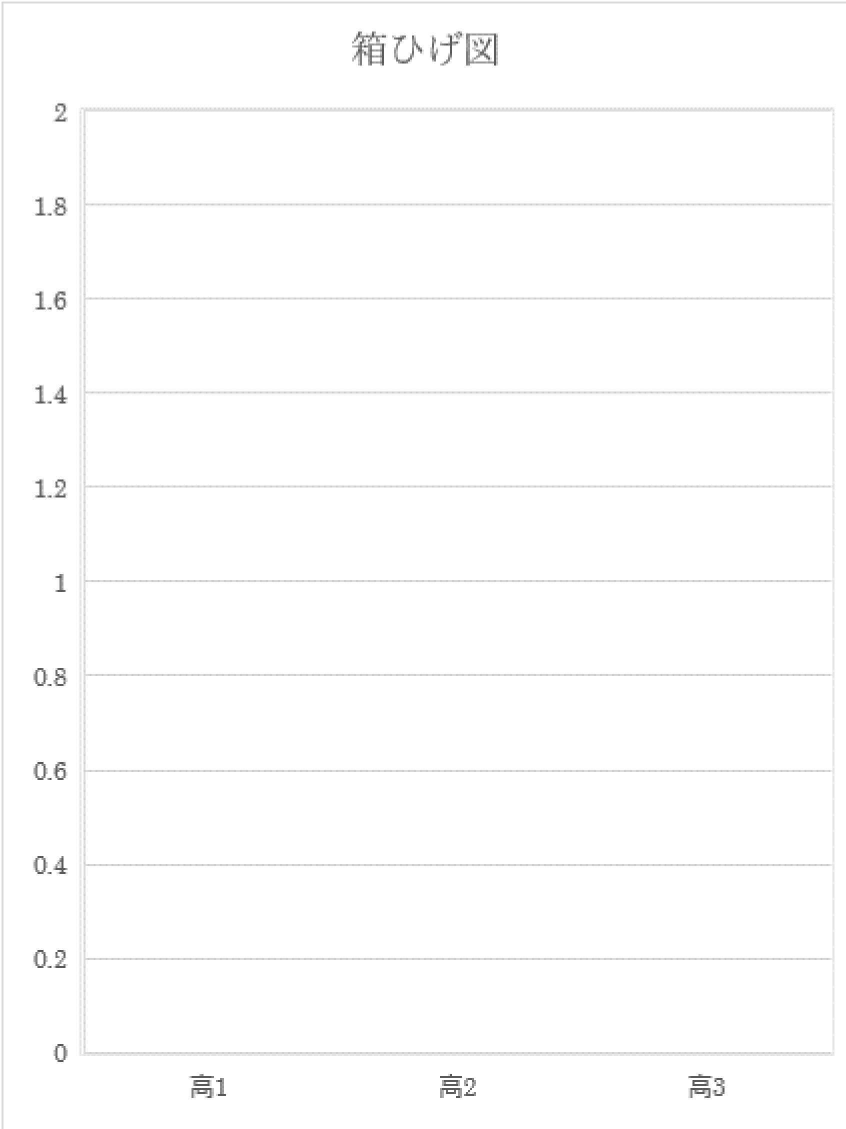
箱ひげ図には
縦向きのものも
あります。

～手順～

① Q_1を下端、Q_3を上端とする箱をかき、箱の中に中央値Q_2を示す線をかく。

② 箱の下端から最小値まで、箱の上端から最大値までの線分をひく。

☆実際に箱ひげ図を描いてみよう！☆



学習プリント⑥

定義⑥　偏差

　データの平均値のまわりに、データの各値がどのように分布しているのかを示す値として、各値と平均値の差を**偏差**という。

　データ$(x_1\,,\ x_2\,,\cdots,x_n)$の平均値を$\bar{x}$とするとき、各値と平均値との差$(x_1-\bar{x},\ x_2-\bar{x},\cdots,x_n-\bar{x})$を、それぞれ平均値からの偏差といい、$x\quad-\bar{x}$で表す。

☆偏差を求めてみよう！☆

$\bar{x}=1.0$

	高１	偏差
1	0.3	
2	0.4	
3	0.7	
4	0.9	
5	1.0	
6	1.2	
7	1.5	
8	2.0	

☆偏差の平均値を求めてみよう！☆

定義⑦　平均偏差

　データ$(x_1\,,\ x_2\,,\ldots,x_n)$の平均値を$\bar{x}$とするとき、

　各値と平均値との差$(x_1-\bar{x},\ x_2-\bar{x},\ldots,x_n-\bar{x})$のそれぞれの<u>絶対値</u>の平均値を**平均偏差**という。

☆平均偏差を求めてみよう！☆

学習プリント⑦

定義⑧　分散

データ$(x_1\ ,\ x_2\ ,\cdots,x_n\ )$の平均値を$\bar{x}$とするとき、各値と平均値との差$(x_1\ -\bar{x},$ $x_2\ -\bar{x},\cdots,x_n\ -\bar{x})$のそれぞれの２乗の平均値を**分散**という。

（例）

データ	偏差	$(偏差)^2$
20		
40		
60		
80		

このデータの
平均値は５０

$$(分散)=\frac{1}{4}(900+100+100+900)=\frac{1}{4}\times 2000=500$$

定義⑧　標準偏差

分散の正の平方根を**標準偏差**といい、s で表す。

$$\mathrm{s}\ =\sqrt{分散}=\sqrt{500}=22.36\cdots$$

分散	$s^2=\frac{1}{n}\ \{(x_1\ -\bar{x})^2+(x_2\ -\bar{x})^2+\cdots+(x_n\ -\bar{x})^2\}$
標準偏差	$\mathrm{s}\ =\sqrt{分散}$

☆電卓を使って分散と標準偏差を求めてみよう！☆

	高１	偏差	$(偏差)^2$
1	0.3		
2	0.4		
3	0.7		
4	0.9		
5	1.0		
6	1.2		
7	1.5		
8	2.0		

分散　＿＿＿＿＿＿＿＿

標準偏差　＿＿＿＿＿＿＿＿

学習プリント⑧

定理①　分散公式

変量x の n 個のデータが$(x_1, x_2, \cdots, x_n)$のとき、$(x_1^2, x_2^2, \cdots, x_n^2)$を変量$x^2$ の n 個のデータと考えることにする。

このとき、

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

である。

(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$)^2$

☆この公式が成り立つかどうか定義⑧から証明してみよう！☆

定義⑧

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

学習プリント⑨

☆電卓の便利な機能☆

<u>M（メモリー）機能を使いこなそう！</u>

・**[M＋]（メモリープラス）**：電卓に表示されている数字を、メモリーに足す。

例）$100 \times 2 + 20 \times 3$を計算するには…

[100][×][2][M +][20][×][3][M +]

の順にボタンを押す。

・**[MRC]（メモリーリコール/メモリークリア）**：1 度押すとメモリー内容を呼び出し、もう 1 度押すとクリア（0 に）する。

例）$100 \times 2 + 20 \times 3$の計算結果を見るには…

[100][×][2][M +][20][×][3][M+][MRC]

の順にボタンを押す。

・**[M－]（メモリーマイナス）**：電卓に表示されている数字を、メモリーから引く。

例）$100 \times 2 - 20 \times 3$の計算結果を見るには…

[100][×][2][M +][20][×][3][M −][MRC]

の順にボタンを押す。

・**[×][=]（2 乗）**：その数字の 2 乗の値を出す。

例）215×215を計算するには…

[215][×][=]

例）$215 \times 215 + 40 \times 40$の計算結果を見るには…

[215][×][=][M +][40][×][=][M +][MRC]

[M]（メモリー）という箱に

[M+]で数字を貯めて

[M－]で数字を引き出して

[MRC]で箱の中身を見るイメージ！

例）$15 \times 15 + 100 \times 2 - 20 \times 3 =$

[15][×][=][M+][100][×][2][M +][20][×][3][M −][MR]

学習プリント⑩

☆電卓の便利な機能を使ってみよう☆

(1)　$2 \times 3 + 4 \times 5 =$

(2)　$2 \times 3 + 4 \times 5 + 6 \times 7 =$

(3)　$4 \times 5 - 2 \times 3 =$

(4)　$2 \times 3 + 4 \times 5 + 8 \times 9 - 6 \times 7 =$

(5)　$27 \times 3 - 54 \times 15 - 12 \times 62 =$

(6)　$(12 \times 12 - 9 \times 9) \div 3 =$

(7)　$24 \times 24 + 19 \times 19 - 13 \times 13 =$

(8)　$124 \times 124 + 92 \times 29 - 13 \times 17 =$

(9)　$41 \times 34 - 15 \times 11 + 113 \times 113 + 55 \times 4 =$

(10)　$20 \times 18 + 7 \times 28 - 201 \times 8 + 72 \times 8 =$

☆定理①を使い電卓を駆使して分散と標準偏差を求めてみよう！☆

	高 1
1	0.3
2	0.4
3	0.7
4	0.9
5	1.0
6	1.2
7	1.5
8	2.0

定理①

(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

①2 乗のデータの平均値を求める。

[0.3][×][=][M +][0.4][×][=][M +] ⋯ [2.0][×][=][M +][MR][÷][8][=]

②データの平均値の 2 乗を求め、①から引く。

学習プリント⑪

☆定理①を使い電卓を駆使して分散と標準偏差を求めてみよう！☆

	高 2
1	0. 4
2	0. 4
3	0. 5
4	0. 5
5	0. 6
6	0. 6
7	0. 6
8	1. 2

$\bar{x} = 0.6$

分散 ＿＿＿＿＿＿＿＿

標準偏差 ＿＿＿＿＿＿＿＿

	高 3
1	0. 2
2	0. 3
3	0. 4
4	0. 5
5	0. 7
6	0. 8
7	0. 9
8	1. 0

$\bar{x} = 0.6$

分散 ＿＿＿＿＿＿＿＿

標準偏差 ＿＿＿＿＿＿＿＿

学習プリント⑫

次のデータは高１の視力とスマホの利用時間についてのデータである。

高１	視力	スマホの利用時間（時間）
1	0.3	2.5
2	0.4	3.0
3	0.7	1.0
4	0.9	1.5
5	1.0	1.0
6	1.2	1.5
7	1.5	1.0
8	2.0	0.5

定義⑪　相関

2 つの変量からなるデータの間に、一方が増加すればそれにしたがって他方も増加する、または他方が減少するという傾向がみられるとき、2 つの変量の間に**相関**がある、または**相関関係**があるという。

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。また、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。どちらの傾向もみられないときは、**相関がない**または**相関関係がない**という。

（例）
・平均気温とアイスの売り上げ
・小テストの点数と期末テストの点数

相関がありそうな 2 つのデータの例を考えてみよう！

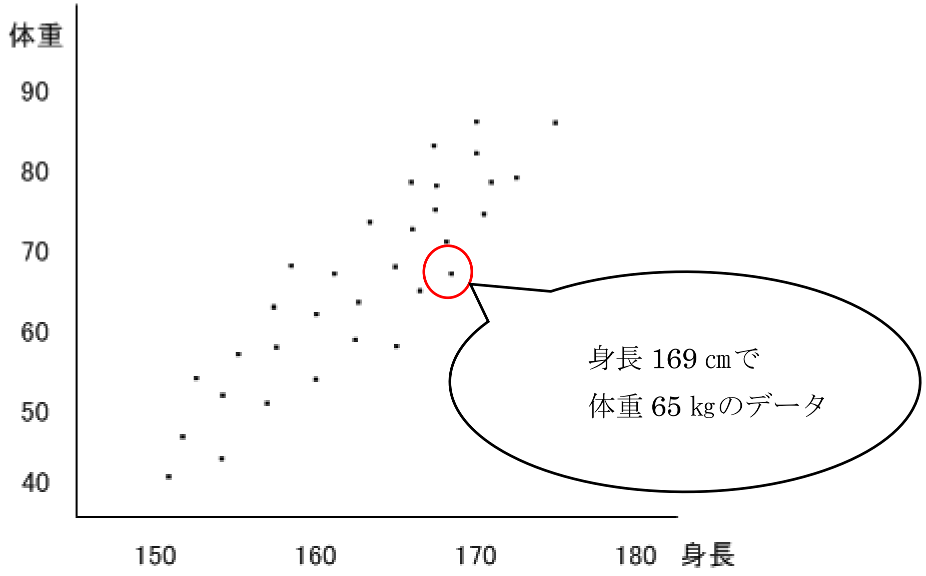
学習プリント⑬

定義⑫　散布図

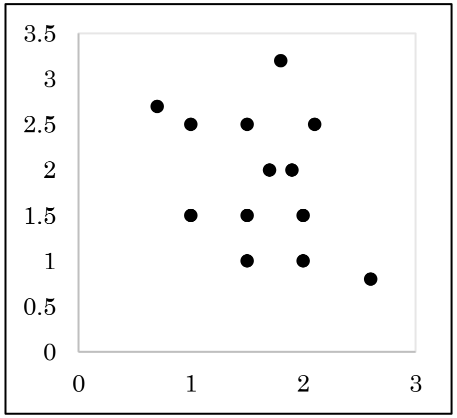
2 つの変量からなるデータを平面上に図示したものを、**散布図**という。

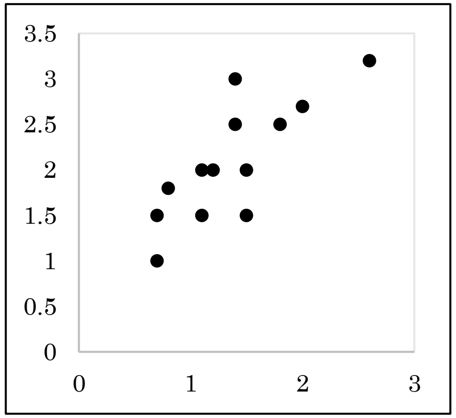
2 つの変量の間の関連性は散布図をかくことで視覚的にとらえることができる。

（例）体重と身長についての散布図



☆どちらの散布図が相関が強そうだろうか☆

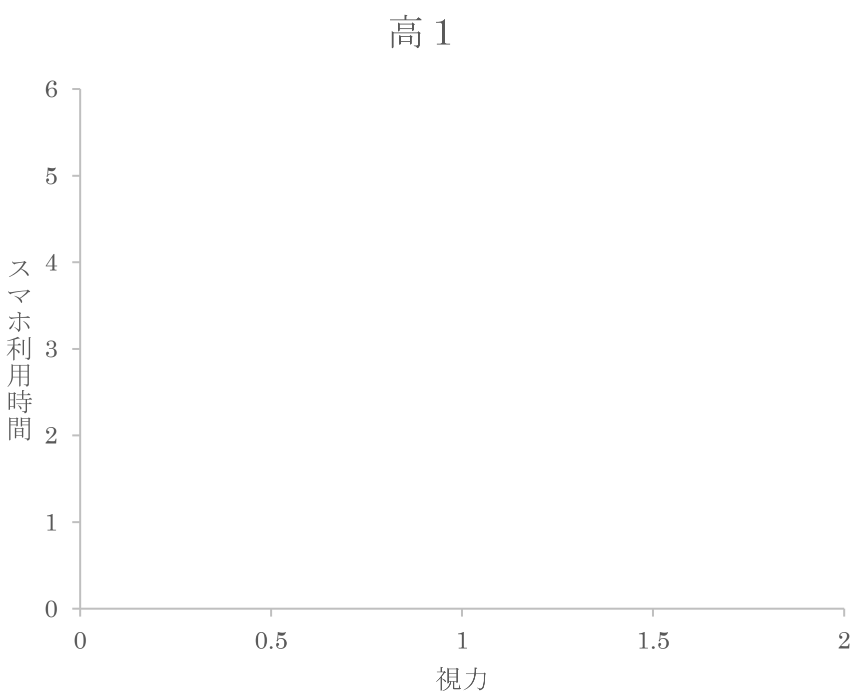




2 つの変量の間に相関があるとき、散布図における点の分布の様子が 1 つの直線に接近しているほど**相関が強い**といい、散らばっているほど**相関が弱い**という。

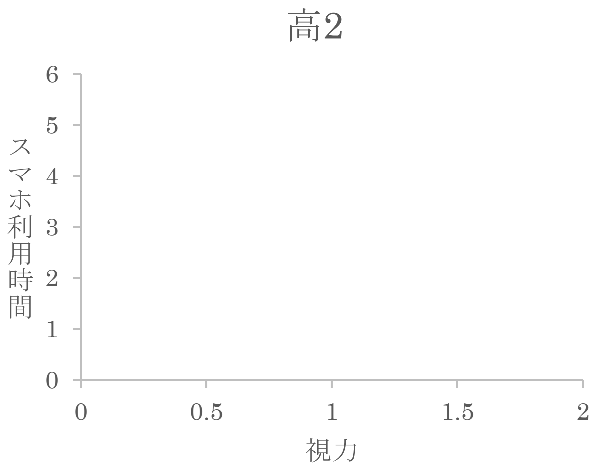
学習プリント⑭

☆実際に散布図をかいてみよう！☆



☆高２のデータで散布図をかいてみよう！☆

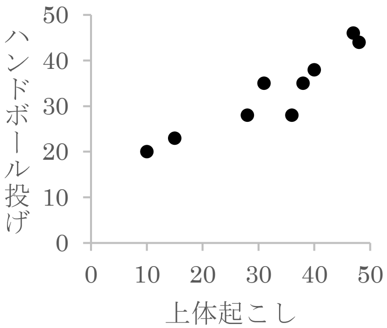
高２	視力	スマホの利用時間
1	0.4	1.0
2	0.4	5.0
3	0.5	3.0
4	0.5	2.0
5	0.6	4.0
6	0.6	3.5
7	0.6	4.5
8	1.2	1.0

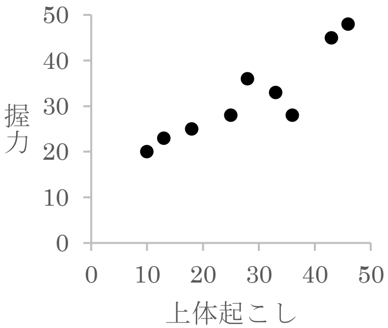


学習プリント⑮

散布図から相関があるかないかは、なんとなく検討をつけることができる。しかし具体的にどれくらい相関が強いのかは分からない。また複数の散布図を比べる場合には、見比べてもどちらが相関が強いかは判定しにくい場合がある。

次のデータはあるクラスの体力測定の結果である。



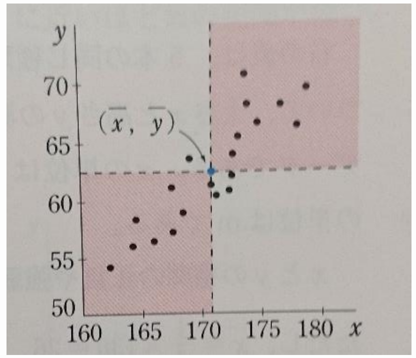


上体起こしの結果が良い方がハンドボール投げの結果も良い。
上体起こしの結果が良い方が握力の結果も良い。
ではハンドボール投げの結果と握力の結果、どちらの方が相関が強いだろうか？
具体的にどれだけ相関が強いか分かれば比較できるのでは…。

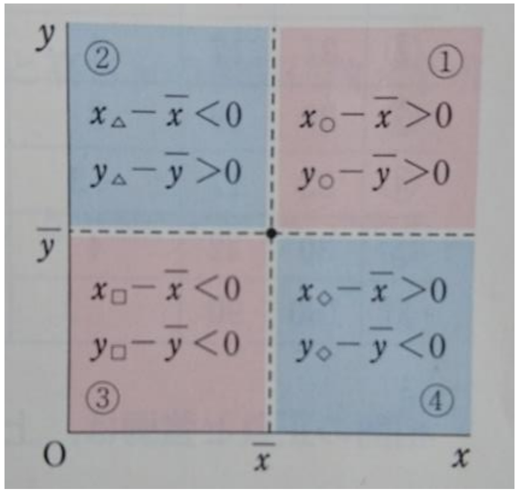
そこで…！

2 つの変量の関係を＿＿＿＿＿することを考える。

まず、変量xの平均値$\bar{x}$と,変量yの平均値$\bar{y}$を考える。
それを散布図に記すと右の図のようになる。（身長xと体重yのデータ）



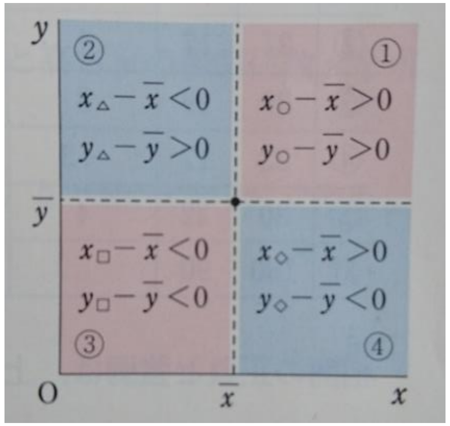
このデータは正の相関があり、$(\bar{x},\bar{y})$を基準に、点は右図のあみかけ部分に多く集まっていることがわかる。



次に、それぞれの値の平均値との差（偏差）を考える。
$(\bar{x},\bar{y})$を基準にして、座標平面を 4 分割することができる。
左図のように領域①～④を定める。
①と③に多く点が集まるような場合は**正の相関**があり、②と④に多く点が集まるような場合は**負の相関**があるといえる。

学習プリント⑯

領域①と領域③、領域②と領域④にはどんな共通点があるか？



偏差の符号に注目してみよう！

領域①と領域③

点(x_i, y_i)が領域①にある。 $\Rightarrow$ $\begin{cases} x_i\text{は}\quad \bar{x}\text{より大きい} \\ y_i\text{は}\quad \bar{y}\text{より大きい} \end{cases}$ $\Rightarrow$ $\begin{cases} x_i - \bar{x} > 0 \\ y_i - \bar{y} > 0 \end{cases}$

点(x_i, y_i)が領域③にある。 $\Rightarrow$ $\begin{cases} x_i\text{は}\quad \bar{x}\text{より小さい} \\ y_i\text{は}\quad \bar{y}\text{より小さい} \end{cases}$ $\Rightarrow$ $\begin{cases} x_i - \bar{x} < 0 \\ y_i - \bar{y} < 0 \end{cases}$

領域②と領域④

点(x_i, y_i)が領域②にある。 $\Rightarrow$ $\begin{cases} x_i\text{は}\quad \bar{x}\text{より小さい} \\ y_i\text{は}\quad \bar{y}\text{より大きい} \end{cases}$ $\Rightarrow$ $\begin{cases} x_i - \bar{x} < 0 \\ y_i - \bar{y} > 0 \end{cases}$

点(x_i, y_i)が領域④にある。 $\Rightarrow$ $\begin{cases} x_i\text{は}\quad \bar{x}\text{より大きい} \\ y_i\text{は}\quad \bar{y}\text{より小さい} \end{cases}$ $\Rightarrow$ $\begin{cases} x_i - \bar{x} > 0 \\ y_i - \bar{y} < 0 \end{cases}$

偏差の積の符号に注目してみよう！

点(x_i, y_i)の偏差の積$(x_i - \bar{x})(y_i - \bar{y})$が
正になるような点が多いならば**正**の相関

点(x_i, y_i)の偏差の積$(x_i - \bar{x})(y_i - \bar{y})$が
負になるような点が多いならば**負**の相関

ということが言える。

学習プリント⑰

定義⑬　共分散

n 個の点の偏差の積$(x_i - \bar{x})(y_i - \bar{y})$の平均値を求め、その符号を調べる。
これを**共分散**といい、s_{xy}と表す。

$$s_{xy} = \frac{1}{n}\ \{(x_1 - \bar{x})(y_1 - \bar{y}) + (x_2 - \bar{x})(y_2 - \bar{y}) + \ldots + (x_n - \bar{x})(y_n - \bar{y})\}$$

$$= \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

共分散は x と y の間に正の相関があるときは正になり、負の相関があるときは負になる。
相関がないときは、偏差の積のうち正のものと負のものが打ち消しあって０に近い値になる。

☆高１のデータの共分散を求めてみよう！

高１	視力	スマホの利用時間（時間）	$x_i - \bar{x}$	$y_i - \bar{y}$
1	0.3	2.5		
2	0.4	3.0		
3	0.7	1.0		
4	0.9	1.5		
5	1.0	1.0		
6	1.2	1.5		
7	1.5	1.0		
8	2.0	0.5		

$\bar{x} = 1.0$

$\bar{y} = 1.5$

学習プリント⑱

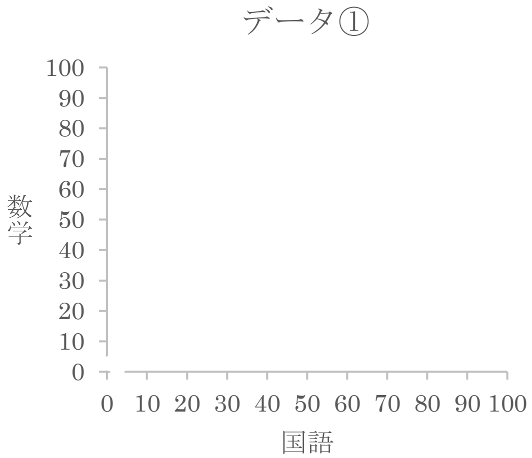
データ①

	A さん	B さん	C さん	D さん
国語	90	60	70	60
数学	50	70	60	60

国語の点数の平均　$\bar{x}$：70 点
数学の点数の平均　$\bar{y}$：60 点
共分散を求めると…

$$s_{xy} = \frac{1}{4}\{(90-70)(50-60) \quad + (60-70)(70-60) \quad + (70-70)(60-60) + (60-70)(60-60) \quad \} = \frac{1}{4}\{(-200) + (-100) \quad \} = -75$$

共分散が負の値なので、このデータの国語の点数と数学の点数には負の相関があるといえそうだ。



データ②

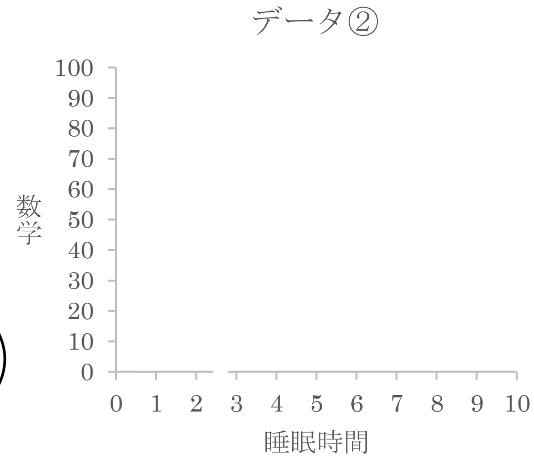
	A さん	B さん	C さん	D さん
睡眠時間	9	6	7	6
数学	50	70	60	60

睡眠時間の平均：7 時間
数学の点数の平均：60 点
共分散を求めると…

$$s_{xy} = \frac{1}{4}\{(9-7)(50-60) \quad + (6-7)(70-60) \quad + (7-7)(60-60) + (6-7)(60-60) \quad \} = \frac{1}{4}\{(-20) + (-10) \quad \} = -7.5$$

データ①の共分散：−75
データ②の共分散：−7.5

共分散の値が大きい。
↓
相関が強い？



学習プリント⑲

データ①とデータ②の散布図を見てみると、この２つのデータの分布は等しいことが分かる
しかし共分散の値には違いがでてしまった。

共分散では、数値の規模がそろっておらず比較できない…。
相関は同じはずなのに、数値は変わってしまう。

データの種類や単位に影響されず統一された数値が必要である。
そこで共分散s_{xy}を各標準偏差s_x,s_yで割った値を考える。

それが**相関係数**である。

相関係数は常に$-1 \leq \mathrm{r} \leq 1$になり、
規模が統一されるので比較することができる。

定義⑭　相関係数

$$\text{相関係数}\mathrm{r} = \frac{s_{xy}}{s_x s_y}$$

s_x：xの標準偏差

s_y：yの標準偏差

s_{xy}：共分散

相関係数rの値は常に$-1 \leq \mathrm{r} \leq 1$であり、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。
相関がないとき、rは０に近い値をとる。

学習プリント⑳

☆高１、高２、高３の相関係数を求めよう！

高３	視力	スマホ利用時間（時間）
1	0.2	5.0
2	0.3	5.5
3	0.4	4.0
4	0.5	5.0
5	0.7	3.5
6	0.8	4.0
7	0.9	2.0
8	1.0	1.0

高１

標準偏差 s_x s_y

共分散s_{xy}

相関係数r

高２

標準偏差 s_x s_y

共分散s_{xy}

相関係数r

高３

標準偏差 s_x s_y

共分散s_{xy}

相関係数r