

# 盛岡三高数学科通信

## How do you solve?

## How do you teach?

### 第18号

発行責任者  
盛岡第三高等学校  
下町壽男

#### データの相関 ①

今、1年生がデータの相関に入っているようですので、今回から2回にわたって、2変量の相関についての話をしていきたいと思います。

#### ● 散布図とデータの相関関係

統計データの特徴を調べるために、中学校ではヒストグラム、数学Iでは箱ひげ図などを学んできました。また、代表値として平均値、中央値、最頻値、そして、散布度としては、分散と標準偏差の話ができました。

それらを踏まえてまず「相関」についての概観について述べていきたいと思います。

データを分析する際、ある集団のテストの点数とか、身長とか、1つの変量について考える場合と、複数の変量がどういう関係があるかということに注目する場合があります。例えば、次の表を見て下さい。

	身長	体重
大崎ナナ	162	43
小松奈々	158	46
本城蓮	182	64
高木泰士	185	72
寺島伸夫	167	52
岡崎真一	164	50
芹澤レイラ	164	48
一ノ瀬巧	183	67
藤枝直樹	178	65
上原美里	153	42
早乙女淳子	168	53
高倉京助	180	70
遠藤章司	175	60
川村幸子	146	35

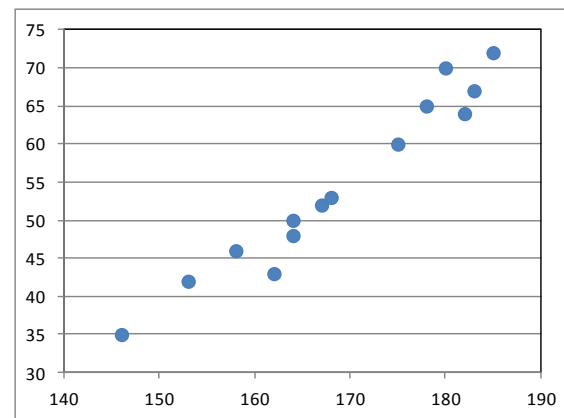
(NANA/矢沢あい より引用しました)

これは、14人の身長と体重を組にして表したものです。身長を変量  $x$ 、体重を変量  $y$  としたとき、 $x$  と  $y$  の大小に関係性があるかどうかを調べるための一つの

モノサシ(指標)が「相関係数」といわれるものです。では、まず相関係数を説明する前に、**散布図**(相関図)について説明しましょう。

今、この2つのデータの関係性を見易くするために、横軸に身長、縦軸に体重をとって、一人一人の場所をプロットしてみましょう。

この図のことを「散布図」といいます。この散布図を見てどんなことがわかるでしょうか。



まず、14人を点で表すと、直線的に並んでいることが視覚的にイメージできます。それは、身長が高いと体重も重い傾向があるということです。

このように、右上がりの直線に点が並ぶとき2つの変量には「(強い)正の相関がある」といい、右下がりのときは「(強い)負の相関がある」といいます。

相関関係とは、簡単にいうと、2つの変量の関係のことですが、もう少し具体的に言うと、次のようにまとめることができます。

#### 【相関関係】

2つの変量について、一方の大小と他方の大小が関係しているどうかを示すもの。

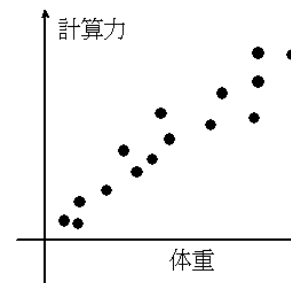
気をつけなければいけないのは、「身長と体重に強い正の相関がある」を、「身長が高ければ体重も重くなる」とは必ずしも捉えられないということです。

確かに、体重は身長の3乗に比例しているというなどの説を信じれば、一方が変化するのにもなってもう一方が変化するという関数関係と見ることができるかもしれませんが、しかし、普通、相関関係を考えるということは、あくまで2つの**変量どうしが似ているかどうか**をみるものにすぎないということです。

なので、相関関係を因果関係(一方が他方の原因になっている)と勘違いしないように気をつけることを授業では注意すべきであると思います。

一つ例をあげます。例えば、下のデータは、「算数の計算力」と「体重」について散布度に表したものです。

これを見ると、強い正の相関がみられます。



この結果から、「体重が重い方が頭がいい」とは結論づけられません。体重が重いということは、それだけ年齢が高い可能性もあります。体重が10kgだったら、赤ん坊だから当然計算なんかできないわけですね。

ですから、「体重」と「計算力」は、相関はあるけれど、因果関係にはなっていないのです。どちらの変量にも「年齢」という因子が関わっているので相関が高くなっているということですね。

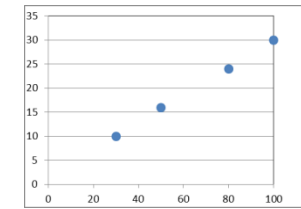
#### ● 相関係数

では、相関係数の話に進みましょう。

今度は次のようなデータを考えてみましょう。

	テスト1	テスト2
学者うさぎ	100	30
マドンナうさぎ	80	24
元気ネコ	50	16
うかれぎつね	30	10
平均	65	20

<テスト1・テスト2 散布図>



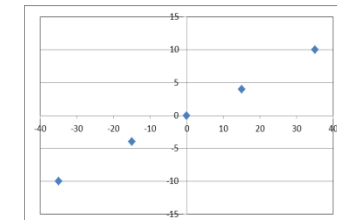
テスト1の平均は65で、テスト2の平均は20です。平均点は違いますが、正の相関が強いことは、散布図からわかります。

では次に、テスト1のデータと、テスト2のデータを比較するかわり、変量から平均を引いたものどうしを比較してみましょう。

	テスト1	テスト2	テスト3 テスト1-65	テスト4 テスト2-20
学者うさぎ	100	30	35	10
マドンナうさぎ	80	24	15	4
元気ネコ	50	16	-15	-4
うかれぎつね	30	10	-35	-10
平均	65	20	0	0

つまり(100,80,50,30)と、(30,24,16,10)の相関を調べるかわりに(35,15,-15,-35)と(10,4,-4,-10)の相関を調べるということです。

<テスト3・テスト4 散布図>



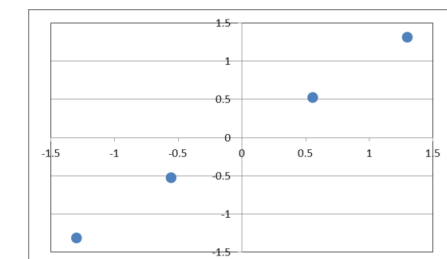
図のように、前のグラフを単に平行移動したグラフになるので、相関の強さには変化はありません。

ところで、テスト3とテスト4の標準偏差を計算すると、26.9と7.6となります。ここで、テスト3とテ

テスト5	テスト6
1.30	1.31
0.56	0.53
-0.56	-0.53
-1.30	-1.31

スト4のデータを、26.9と7.6で割ったデータの散布図を考えてみましょう。

<テスト5・テスト6 散布図>



やはり相関に変化はなさそうです。

ではこの続きは次号で。