

SS 理数探究 I 統計処理④

1. 前時を振り返る

・分散と標準偏差

※ 偏差値って？

平均点を 50 として得点等を換算して得られる値

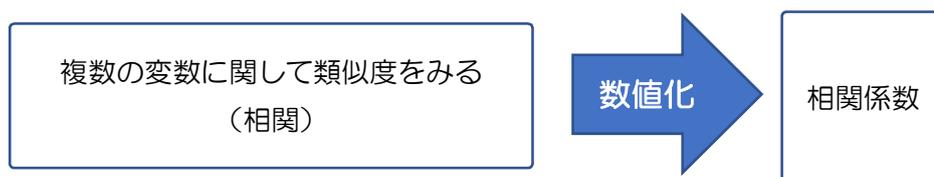
「偏差値 = (得点 - 平均点) ÷ 標準偏差 × 10 + 50」

偏差値	80	70	60	50	40	30
上位からの割合(%)	0.13	2.28	15.87	50.0	84.13	97.72

2. 相関とは

2つ以上のもつがあるときにそれらが「どれくらい類似しているか」という「類似度」を意味する。

相関係数とは、2つの（確率）変数の間の相関（類似性の度合い）を示す統計学的指標・



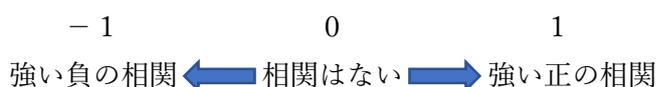
相関係数（ピアソン）の計算方法

数学 I の教科書では

$$\begin{aligned} \text{相関係数 } r &= \frac{s_{xy}}{s_x s_y} \frac{(x \text{ と } y \text{ の共分散})}{(x \text{ の標準偏差}) \times (y \text{ の標準偏差})} \\ &= \frac{((x \text{ の偏差})(y \text{ の偏差}) \text{ の和}) \div (\text{データの個数})}{(x \text{ の標準偏差}) \times (y \text{ の標準偏差})} \end{aligned}$$

と載っている。

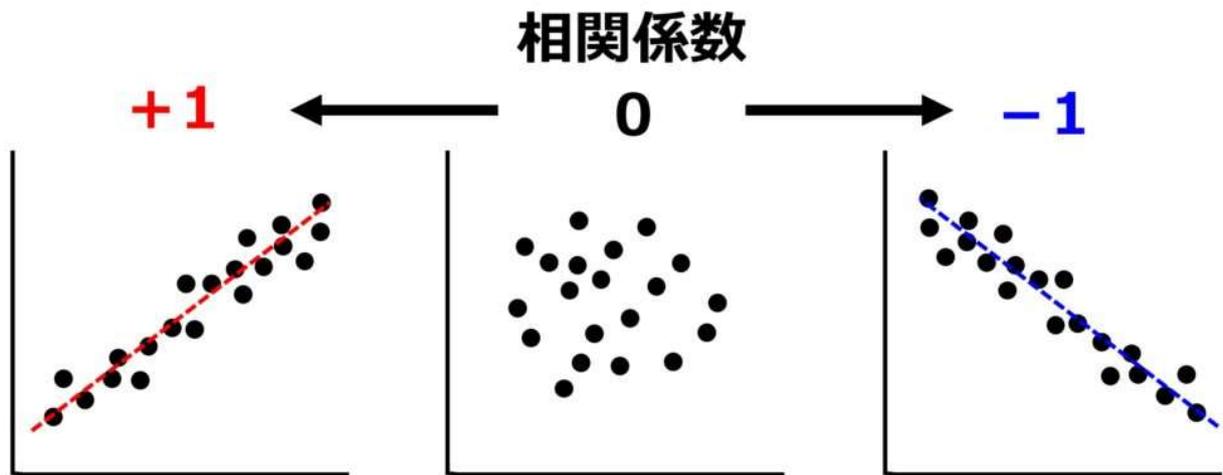
実際に計算すると $-1 \leq r \leq 1$ の値をとり、以下のような相関があるという。



年 組 番 氏名

統計処理④

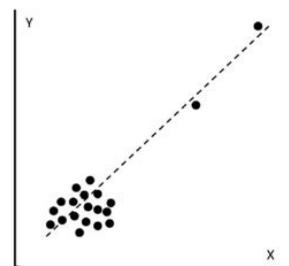
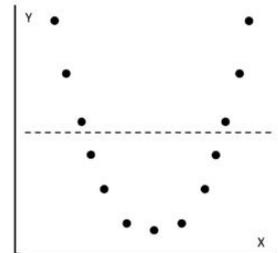
散布図：2つの変数の値を縦軸・横軸でとって、各標本の変数を図内に点を取って表したものの
相関係数との関係は以下の通り



<相関を扱う際の注意>

- ※ 相関係数は直線的な関係性（線形性）を表すので、相関係数が0で何らかの相関があることもある。また、外れ値（特異値）があると相関係数が大きく異なることもある。

例



- ※ 相関関係があっても因果関係までを説明できない。
2つの変数の関係性はいえるがどちらかの原因・要因が他方であるとはいえない

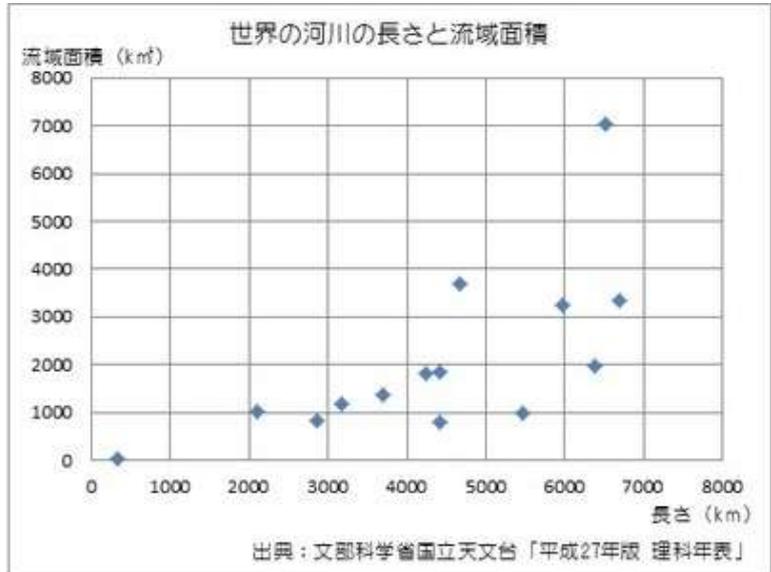
統計処理④

ワークシート

課題：① 以下の表と図は世界の主な河川の長さ（km）と流域面積（1000km²）の一覧である。河川の長さ（km）と流域面積（1000km²）にはどのような相関があるか。

河川名	長さ (km)	流域面積 (1000km ²)
アマゾン	6516	7050
コンゴ（ザイール）	4667	3700
ナイル	6695	3349
ミシシッピー-ミズーリ	5969	3250
アムール	4416	1855
マッケンジー	4241	1805
ボルガ	3688	1380
長江（揚子江）	6380	1959
オレンジ	2100	1020
黄河	5464	980
インダス	3180	1166
ドナウ	2850	815
メコン	4425	810
利根川	322	17

出典：文部科学省国立天文台「平成27年版 理科年表」



② 皆さんの身の周りで正・負いずれかの強い相関があると予想される2つの変数を各グループで一つずつ上げてください（なぜ強い相関がありそうかも考えてください）

年 組 番 氏名

統計処理④

参考

- Data Science by R and Python 統計学を、広く、深く、わかりやすく。
<http://tomoshige-n.hatenablog.com/entry/2014/08/06/035001>
- Lactivator すべてのビジネスマンのためのマーケティングリサーチの学び場
<https://lactivator.net/2020/06/05/correlation/>
- TECHPLAY
<https://techplay.jp/column/495>
- いちばんやさしい、医療統計
https://best-biostatistics.com/correlation_regression/correlation.html
- BigDataMagazine
<https://bdm.change-jp.com/?p=1254>
- なるほど統計学園高等部
https://www.stat.go.jp/koukou/howto/process/p4_3_2_5.html