

【手がかり用スペース】

先週の復習

代表値が同じ分布でも同じものとは限らない

ばらつき具合の数値化

ばらつき具合を数値 1 つで表現する

散布度

偏差二乗和

分散

標準偏差

標準化

標準偏差は散布度の中でも重要

個々のデータの偏差を標準偏差で割る

z スコア

$$\text{偏差値} = z \text{ スコア} \times 10 + 50$$

正規分布するデータは「標準正規分布」に変換して相互に比較できる

正規分布するデータなら「標準正規分布」のどの位置にあるかを把握できる

第 8 回以降で詳しく説明します

練習問題の解答

問題 1：平均値を求めなさい

47.0

問題 2：偏差平方和を求めなさい

2250.00

問題 3：分散を求めなさい

225.00

問題 4：標準偏差を求めなさい

15.00

問題 5：f と i の z スコアを求めなさい

f：-0.60

i：1.80

問題 6：g の偏差値を求めなさい

54

有効数字の桁数に注意してください

【要点書き出し用スペース】

【手がかり用スペース】

本日の内容

- 2 変数統計
- 共変関係の表現
 - 連関係数
 - 相関係数

共変関係

一方が変わると、もう一方も値が変わる関係

共変関係の視覚的な把握

- まずは視覚的な把握を行う
 - 散布度の時と同じ
 - 「散布図（スキャタグラム）の描画」
 - 一方の変数を x 軸に取り、もう一方の変数を y 軸に取る
 - 組み合わせたところにプロット（点を描く）

表 1 サンプルデータ

名前	国語	英語	国語 Z	英語 Z
A	16	25		
B	28	25		
C	38	30		
D	46	30		
E	46	46		
F	50	61		
G	58	63		
H	60	72		
I	66	72		
J	92	76		
平均値			0.0	0.0
標準偏差			1.00	1.00

【要点書き出し用スペース】

【手がかり用スペース】

サンプルデータから基本統計量（代表値と散布度）を算出
表 1 に記入
データをもとにプロットを行う

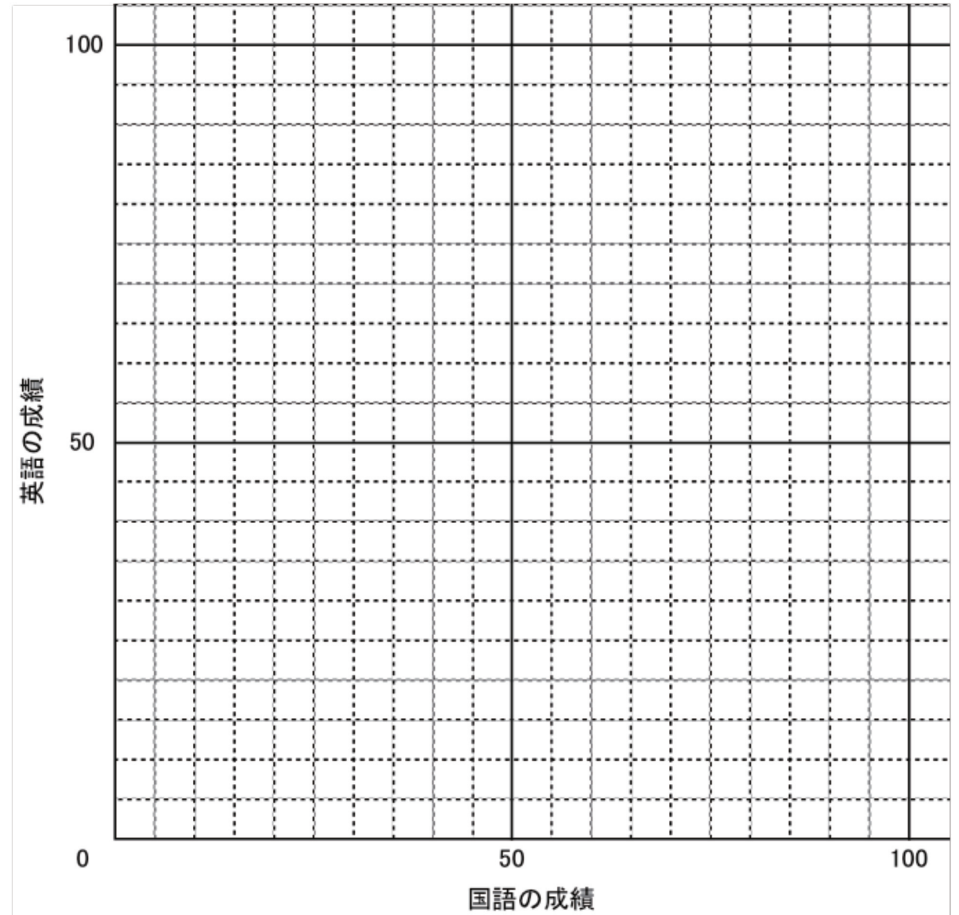


図 1 国語と英語の成績による散布図

共変関係の指標

一方の値が大きくなると、もう一方の値はどう変化するか？

大きくなる or 小さくなる

線を書き入れることで把握する

SD 直線

【要点書き出し用スペース】

【手がかり用スペース】

SD 直線

共変関係の傾向を把握する

SD 直線は 3 点を通る直線

X と Y でそれぞれ平均よりも標準偏差で 1 ずつ小さい点

X と Y でそれぞれ平均よりも標準偏差で 1 ずつ大きい点

X と Y のそれぞれの平均値

(国語の平均値 - 国語 1SD, 英語の平均値 - 英語 1SD)

(国語の平均値 + 国語 1SD, 英語の平均値 + 英語 1SD)

(30,30) と (70,70) と (50,50) を通る直線

右上がりの SD 直線の描画

もう一つの SD 直線

以下の 3 点を通る直線

(国語の平均値 - 国語 1SD, 英語の平均値 + 英語 1SD)

(国語の平均値 + 国語 1SD, 英語の平均値 - 英語 1SD)

国語の平均点, 算数の平均点

どちらの SD 直線に近いかで関係性を把握する

右肩上がり

正の共変関係

右肩下がり

負の共変関係

完全な関係なら SD 直線上に全ての点が並ぶ

z スコアでの散布図と SD 直線

単位が違うもの同士の比較等には z スコアが必要

基本的に素点での散布図と同等

※来週までに作成しておくこと

共変関係の数値化

視覚的には把握できたが他人に伝えるには数値化する必要がある

1 つの値で示したい

【要点書き出し用スペース】

【手がかり用スペース】

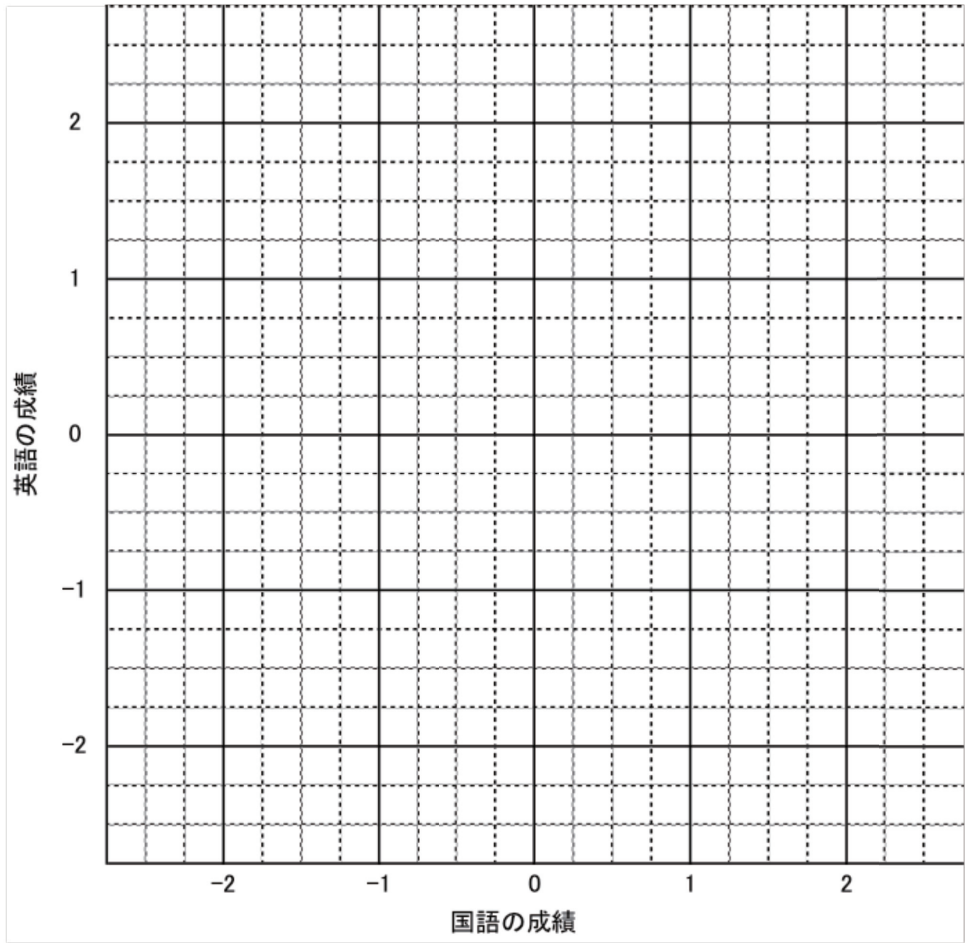


図 2 z スコア変換後の散布図

表 2 分割表

	平均より小さい	平均より大きい	合計
平均より大きい	c	a	(a+c)
平均より小さい	d	b	(b+d)
合計	c+d	a+b	(a+b+c+d)

【要点書き出し用スペース】

【手がかり用スペース】

分割表の作成

平均値を元にデータを分割し度数を数える
平均値と同値の場合は 0.5 度数として考える

共変関係の方向

正の共変関係

表の a と d が b と c より大きいはず

負の共変関係

表の b と c が a と d より大きいはず

a と d、b と c の組み合わせを比較すれば良い

連関係数

分割表を元に、数値を 1 つ算出

ϕ (ふあい) 係数 (四分点相関係数)

ad-bc を計算

周辺の度数を全部掛け合わせたものの平方根で割る

データ数によって調整している

計算結果は $-1 \leq \phi \leq 1$ になる

正の値なら右上がり

負の数なら右下がり

名義尺度以上で利用可能

質的データの分析等で利用される

順序尺度以上であれば相関係数を用いる

関係の強さを 1 つの数値で表現

ピアソンの積率相関係数

連関係数は度数しか利用していない

情報量が少ない

→より多くの情報を利用して関連性を 1 つの数値で表現する

全てのデータの値を利用する

情報量が多い

いわゆる「相関係数」が求められる

【要点書き出し用スペース】

【手がかり用スペース】

相関係数の求め方

まずデータを標準化する

z スコアを用いて算出する

異なる単位同士の関係でも利用可能

それぞれのデータで、x と y の z スコアを掛け合わせる

それぞれの値がどれだけ平均値から離れているかの算出

変数 X の z スコアと変数 Y の z スコアの積

c	a
d	b

a と d は正の値

c と b は負の値

掛け合わせた値の平均値を求める

求められた値は「相関係数 (r)」(ピアソンの積率相関係数)

$-1 \leq r \leq 1$ (ただし尺度は順序尺度)

-1 は完全な負の相関、+1 は完全な正の相関

相関の強さ

$0.0 \leq |r| < 0.2$ …ほとんど関係なし

$0.2 < |r| \leq 0.4$ …弱い相関あり

$0.4 < |r| \leq 0.7$ …中程度の相関あり

$0.7 < |r| \leq 1.0$ …強い相関あり

今回の結果はどうだったか？

【要点書き出し用スペース】

【手がかり用スペース】

練習問題

1. 空欄に用語欄から適切な語を選び、挿入しなさい

複数のデータセットを比較するためには、(A) を算出して比較する必要がある。(A) を算出するためにはそれぞれの素点から (B) を引き (C) で割る。この手続きは (D) と呼ばれる。この結果、データは (E) 上に配置されることになり、単一の物差し上で比較することができるようになる。また、いわゆる (F) は (A) を 10 倍して 50 を加算した結果である。従って 50 近辺の値であれば (B) に近いと考えられる。

語群

ア：平均値 イ：中央値 ウ：Z スコア エ：分散 オ：標準偏差 カ：偏差値
キ：標準正規分布 ク：標準化 ケ：尺度化 コ：偏差平方和)

2. 次の値の平均値、標準偏差を求めなさい

26, 27, 28, 30, 33, 34, 38, 42, 43, 51
59, 61, 61, 63, 64, 65, 67, 68, 69, 71

3. 次の二変数間のデータについて問いに答えなさい

	国語	英語
A	15	20
B	25	40
C	30	30
D	35	40
E	45	40

3-1 Cさんの成績は共に30点だが、どちらの方が成績が良かったといえるか。根拠を示しつつ明らかにしなさい

3-2 Dさんの国語の成績について、偏差値を算出しなさい

3-3. 国語と英語の相関係数を求めなさい

【要点書き出し用スペース】