

複数のテストの点数の相関

樋口さぶろお

龍谷大学理工学部数理情報学科

生活の中の統計技術 L03(2018-10-08 Mon)

最終更新: Time-stamp: "2018-10-08 Mon 14:01 JST hig"

今日の目標

- 偏差値の意味を説明できる
- 相関係数を計算できる
- 相関係数の意味を説明できる



ここまで来たよ

2 受験者データの代表値と散布度

- 偏差値の意味

3 複数のテストの点数の相関

- 2変量データと散布図
- 2変量データの相関

偏差値

偏差値

あるクラスの点数のデータの中の、67点(の成績の人)の偏差値

$$50 + 10 \times \frac{67 - \text{平均値 } m}{\text{標準偏差 } s}$$

$$\text{標準偏差} = \sqrt{\text{分散}}$$

$$\text{分散} \stackrel{\text{例 } b}{=} \frac{1}{8} [(75 - m)^2 + \dots + (85 - m)^2].$$

偏差値は代表値ではない

受験者1人1人の成績が、平均値から上、または下に離れている程度を見られる。

- 異なるテスト(難度, 満点)でも比べられる
- 同じクラスでないと比べられない
- 偏差値の平均値は , 偏差値の標準偏差は
- 偏差値はまあ'無次元の数'(1000点満点と100点満点を比較可能)

L03-Q1

Quiz(標準得点と偏差値を用いた比較)

数学と英語の学年共通のテストがあった。数学の点数は 平均値は $\bar{x} = 90$, 分散 $S_x^2 = 16$ だった。英語の点数は 平均値は $\bar{y} = 60$, 分散 $S_y^2 = 4$ だった。

- ① 数学で 86 点を取った人がいる。この 86 点に対応する標準得点と偏差値を求めよう。
- ② 数学での 86 点と英語での 57 点をくらべると、学年内での相対評価として、どちらのほうがよい成績か。式を使って理由を延べ、完結した日本語の文で答えよう。

L03-Q2

Quiz(偏差値)

学力テストの偏差値について、次のうち正しいものを1つ選ぼう。

- ① 偏差値の最低値は0である
- ② 自分の点が平均点より下のとき、分散が大きいほうが自分の偏差値はより高い
- ③ 自分の点が平均点より上のとき、標準偏差が大きいほうが自分の偏差値はより高い
- ④ 100点のテストで満点を取った場合の偏差値は75である
- ⑤ 偏差値50の人の順位は上から1/2である

L03-Q3

Quiz(偏差値の性質)

次を, 正しい, 誤り, もっともらしいが正しいとは断定できない, に分類しよう.

- ① 別の塾に転校した後, 塾内テストの偏差値が上がったことから, 成績が上がったと言える.
- ② 同じ学級内の偏差値が, 中間試験より期末試験で下がったので, 学級内の順位が下がったと言える.
- ③ 教員が全受験者に5点を加点したので, 偏差値は実際より高めに出ているはずである.
- ④ 同じ学級内での偏差値が, 数学より理科のほうが高いので, 理科のほうがより上位にいると言える.

ここまで来たよ

2 受験者データの代表値と散布度

- 偏差値の意味

3 複数のテストの点数の相関

- 2 変量データと散布図
- 2 変量データの相関

多変量データ

これまでやってたのはぜんぶ1変量データ.

3変量データはこんな例. 1人のデータ (x, y, z) . x, y, z は同じ単位でなくてよい. (身長, 体重, 年齢) でもいい.

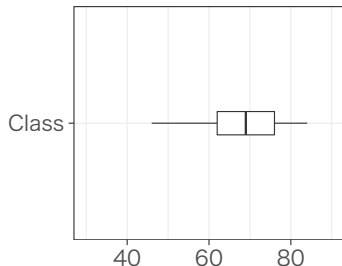
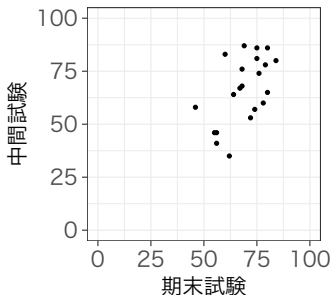
- x 期末試験の点数
- y 中間テストの点数
- z 提出課題数

(学生名)	x	y	z
学生 1	90	34	0
学生 2	68	51	9
⋮	⋮	⋮	⋮
学生 N	28	51	9
中央値	62	70	10
平均値	70	68	12
標準偏差	5	7	3

散布図=相関図

2 変量なら散布図で様子を描ける
(Excel なら 挿入 > グラフ).

横軸に x , 縦軸に y で, データ 1 個
(学生 1 人) について点を 1 個うつ.
もし 2 人以上が重なったら...



ここまで来たよ

2 受験者データの代表値と散布度

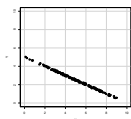
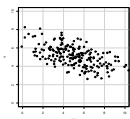
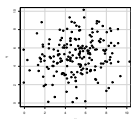
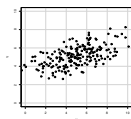
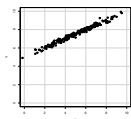
- 偏差値の意味

3 複数のテストの点数の相関

- 2 変量データと散布図
- 2 変量データの相関

正の相関・負の相関・無相関

相関係数 r で x, y の傾向の一部がとらえられる。計算方法は以下。



強い正の相関

$$r = 0.99$$

弱い正の相関

$$r = 0.55$$

無相関

$$r = 0$$

弱い負の相関

$$r = -0.55$$

強い負の相関

$$r = -0.99$$

相関

‘正の/負の相関がある’: x が大きい $\Leftrightarrow y$ が大きい/小さい傾向がある

‘相関が強い/弱い’: 傾向がはっきりしている/していない

‘相関がない’ x が大きい $\Leftrightarrow y$ が大きいとも小さいともいえない。相関が極端に弱い場合。

共分散 高校 数学 I 発展

相関の強さを数で表したい
復習:

$$x \text{ の平均値 } \bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$$

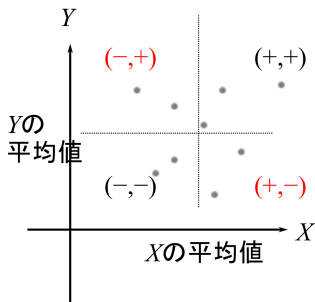
$$x \text{ の分散 } S_x^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})(x_i - \bar{x})$$

\bar{y}, S_y^2 も同様.

共分散 (covariance)

$$x, y \text{ の共分散 } C_{xy} = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x}) \times (y_i - \bar{y})$$

共分散の意味



$(+, -) = (x_i - \bar{x}$ の符号, $y_i - \bar{y}$ の符号).

共分散が正に/負に大きい \Leftrightarrow 正の/負の相関が強い (?)

なぜなら

しか～し (次のスライド)

相関係数 高校 数学 I

共分散は

- テストの点数 x を 20 点満点から 100 点満点に変更すると 5 倍になる.

相関係数は、これらの影響を受けずに、相関の強さをそのまま表す.

相関係数 (correlation coefficient)

$$x, y \text{ の相関係数 } r = \frac{C_{xy}}{S_x \times S_y}$$

L03-Q1

Quiz(共分散)

- ① x, y の共分散を求めよう
- ② x, y の相関係数を求めよう. ただし, y の標準偏差 $= \sqrt{\frac{122}{5}} = 4.94$ は使っちゃっていい.

x	y
1	5
3	15
4	14
5	11
7	20

L03-Q2

Quiz(共分散と相関係数)

次は、ある材料と製法で作った棒の長さ x cm と質量 y g のデータである.

$x(\text{cm})$	$y(\text{g})$
9	40
9	45
9	50
10	50
10	50
10	50
10	50
10	55
11	55
12	55

平均値はそれぞれ、 $\bar{x} = 10\text{cm}$ 、 $\bar{y} = 50\text{g}$ である. 次の量を、(単位があるものには) 単位をつけて答えよう.

- ① x の分散 s_x^2 .
- ② x と y の共分散 s_{xy}
- ③ x と y の相関係数 r

相関係数の性質

- $-1 \leq r \leq +1$
- 正/負の相関 $\Leftrightarrow r$ が正負
- 相関が強い/弱い $\Leftrightarrow |r|$ が 1 / 0 に近い.
- $r = \pm 1 \Leftrightarrow$ 散布図の点がすべて傾き正/負の一直線上 $\Leftrightarrow y$ は x の 1 次関数.
- $r = 0 \Leftrightarrow$ '相関がない' しかし…
- 相関係数は
- r は x, y の 1 次関数による変換のもとで不変

L03-Q3

Quiz(相関係数の性質)

2 変量データ (x, y) の相関係数を考える.

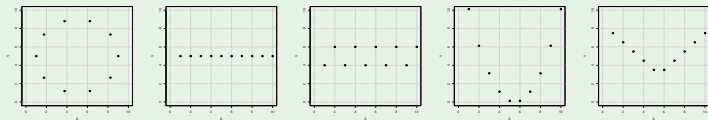
- ① x に一斉に 5 を加えたとき, 相関係数はどうなる?
- ② x を一斉に 2 倍したとき, 相関係数はどうなる?
- ③ y を一斉に -2 倍したとき, 相関係数はどうなる?
- ④ x, y をともに一斉に -2 倍したとき, 相関係数はどうなる?

相関係数=0 と 「関係がない」 は別の話

L03-Q4

Quiz(相関係数)

次のうち、相関係数 r がもっとも大きいものはどれ?



Anscombe(1973)

みかけの相関: 相関係数 $\neq 0$ と「AがBの原因」は別の話

期末試験の成績と、「授業中に鉛筆を回す回数」には正の相関があった…



受験者集団をあわせたとき，相関係数は和にも平均にもならない

A 組では，平常点と期末試験の点数に正の相関があった．

B 組では，平常点と期末試験の点数に正の相関があった．

では，A 組と B 組をあわせた学年全体では？

統計検定 3 級 (2016-11) から

2016 年 11 月統計検定 3 級問 5

お知らせ

- 2018-11-25 日 統計検定の一般会場団体受験 (10%ディスカウント受験) (大阪会場) の 10 月初めまでに申込み. 申込
<https://goo.gl/forms/XUGUXr2TJw33wR3j1>

