

## 統計基礎 第10回

### テーマ：相関（その2）

#### 1 はじめに

- ・ 次回（12/18）に、冬休みの宿題を出します。冬休み明けの最初の授業（1/15）に提出して下さい。台風で休講になった回の授業の代わりです。もし、次回休んだ場合は、教材提示フォルダーからDLして、提出して下さい。

Q excel で文字を入力したが、はみ出した部分が印刷できない場合はどうすればよい？

A ホーム →折り返して全体を表示する

#### 前回の演習の解答例

- ・ 毎月の平均気温とカゼ薬への支出額には負の相関がある

## 2 相関 (つづき)

**演習 1**<sup>1</sup> 正の相関があると思うものを1つ、負の相関があると思うものを1つ考え、

<http://htanaka.jimdo.com/>

にアクセスし、統計基礎第10回の解答先の下リンクをクリックして、

解答して下さい。

(注) できるだけ原因となるものを前に、結果となるものを後ろにして下さい。

➤ (例) 身長と体重 (正の相関)

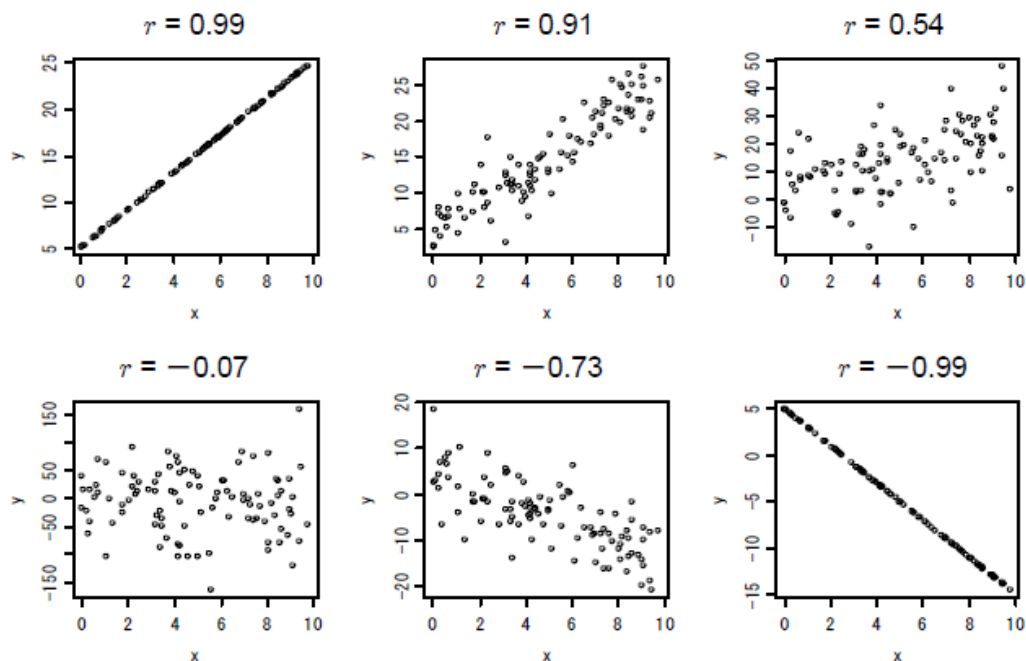
(ヒント) できるだけ数量で表現します。

- △寒くなるとコートを着る人が増える (負の相関)
  - 気温 (が低くなる) とコートを着ている人の数 (が増える) (負の相関)
- △離婚者数 (が増える) と少子化 (になる) (負の相関)
  - 離婚者数 (が増える) と出生数 (が減る) (負の相関)

### 3 相関係数

- 相関の強さをあらわす指標です。

散布図と相関係数 (r) の関係は？



相関係数 (r) をどう解釈しますか？

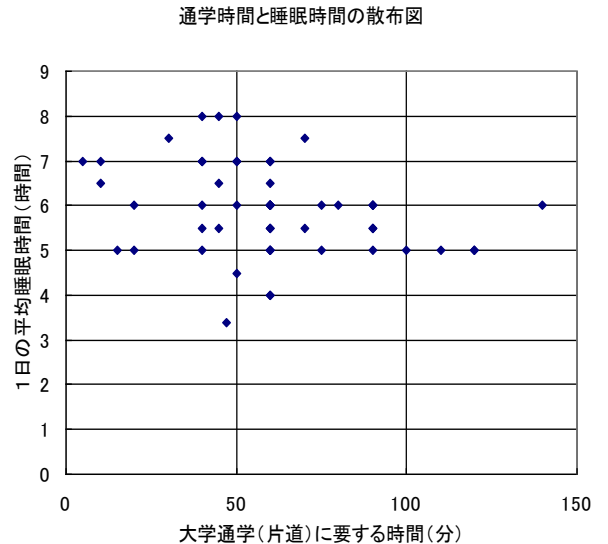
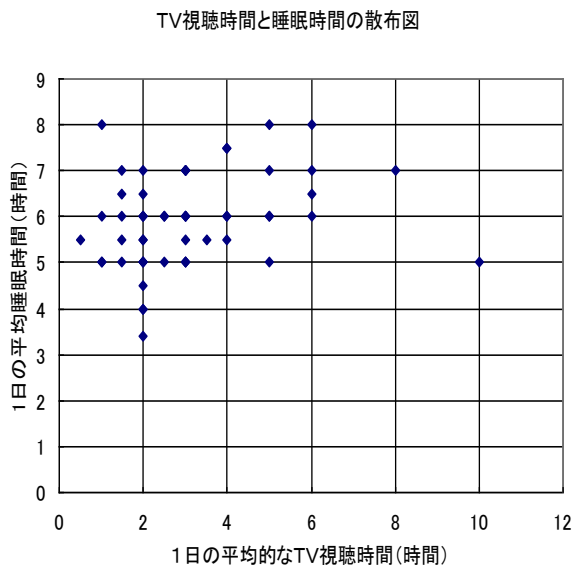
-1.0 ~ -0.4	-0.4 ~ -0.2	-0.2 ~ 0.2	0.2 ~ 0.4	0.4 ~ 1.0
負の相関	弱い負の相関	無相関	弱い正の相関	正の相関

※ 学問分野、分析者、データの種類により異なり、絶対的な基準ではありません。

**演習 2** どのように予想しますか？ ア イ ウのいずれかを選んで下さい。

- (1) テレビの視聴時間と睡眠時間は、  
 ア 正の相関がある イ 無相関 ウ 負の相関がある
- (2) 通学時間と睡眠時間は、  
 ア 正の相関がある、 イ 無相関、 ウ 負の相関がある

**演習 3** 次は、2008年度の統計基礎の受講者のデータを散布図にしたものです。2つの量の関係を述べて下さい。相関係数は、どの程度だと予想しますか？



- (1) テレビの視聴時間と睡眠時間は、  
 ア 弱い正の相関がある イ 無相関 ウ 弱い負の相関がある  
 相関係数は、  
 カ 0.27 キ 0 ク -0.27
- (2) 通学時間と睡眠時間は、  
 ア 弱い正の相関がある、 イ 無相関、 ウ 弱い負の相関がある  
 相関係数は、  
 カ 0.28 キ 0 ク -0.28

(解答例)

(1) テレビ視聴時間と睡眠時間は弱い正の相関がある。

相関係数は 0.27。

自由になる時間がたくさんある人は、テレビを長時間視聴し、睡眠時間も多くとっているようである。すなわち、睡眠時間を削ってテレビを見ている訳ではないようである。

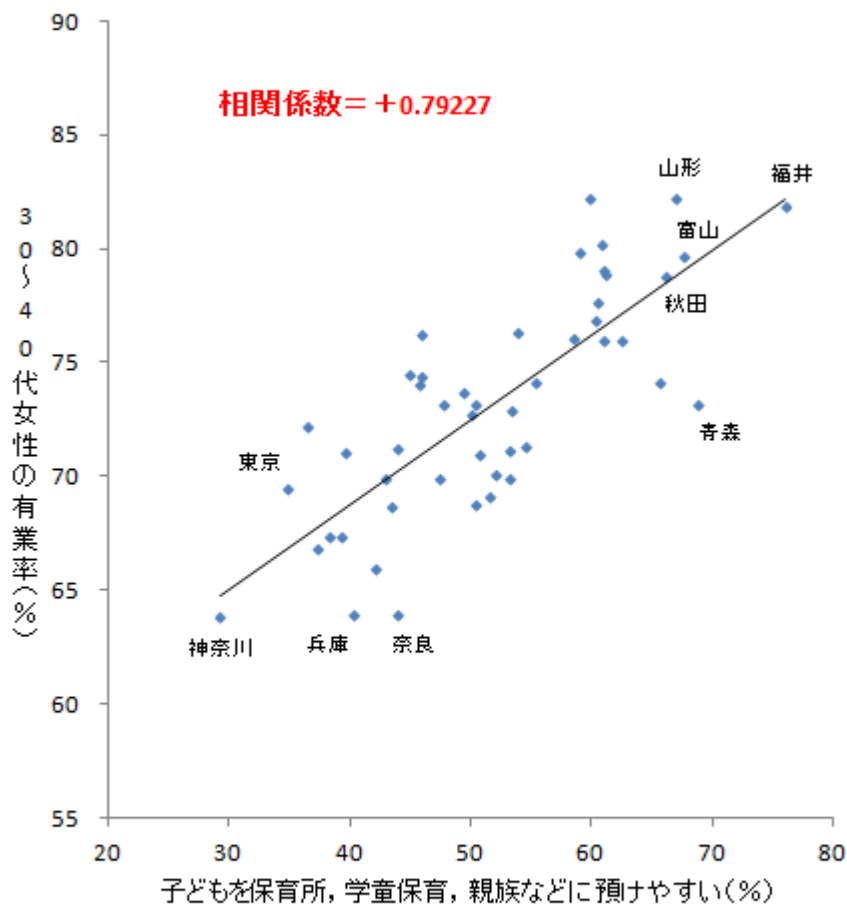
(2) 通学時間と睡眠時間は弱い負の相関がある。

相関係数は -0.28。

通学時間が長いと、睡眠時間は短くなる傾向にある。

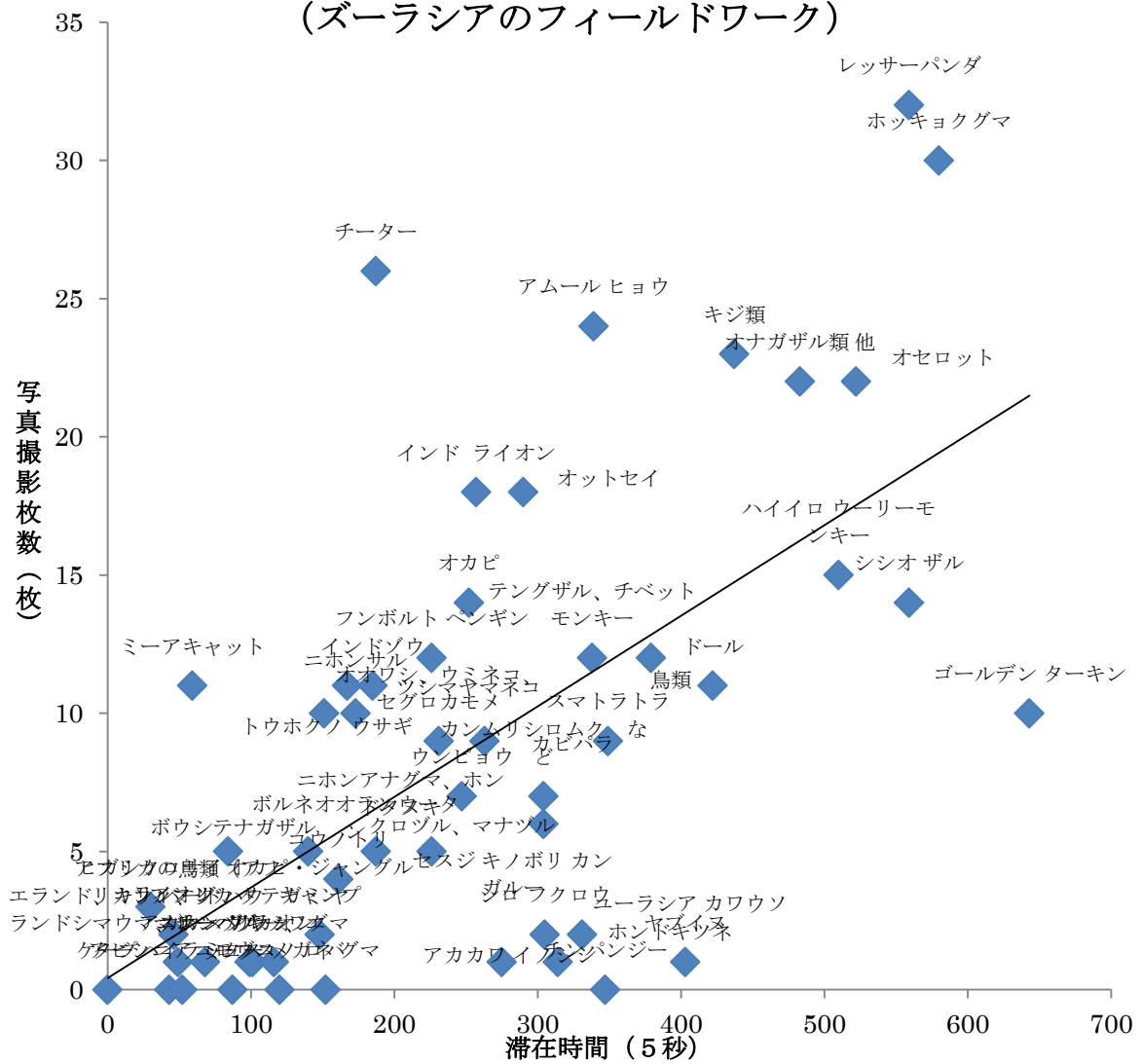
参考

子どもの預けやすさと、子育て期の女性の有業率の相関



※横軸は、高校生以下の子がいる成人男女の肯定率。  
資料:内閣府「地域における女性の活躍に関する意識調査」(2015年),  
総務省「就業構造基本調査」(2012年)  
作成者:舞田敏彦(@tmaita77)

(参考) 動物の前の滞在時間と撮影写真枚数の関係  
(ズーラシアのフィールドワーク)



相関係数 = 0.63

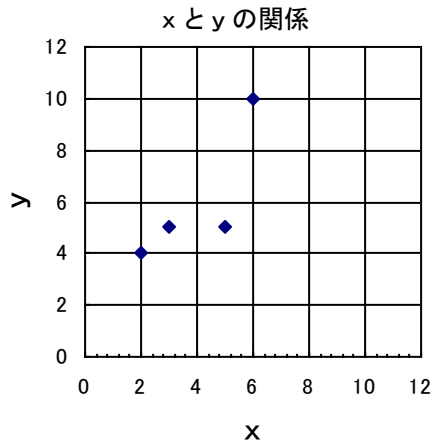
### 3 相関係数の求め方

- ・  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$  というデータがあり、 $x$  の平均を  $\bar{x}$ 、 $y$  の平均を  $\bar{y}$  とすると、次のとおりです。偏差とは、平均からのずれです。

$$\text{相関係数} = \frac{x \text{ と } y \text{ の偏差積和}}{\sqrt{x \text{ の偏差 2 乗和} \times y \text{ の偏差 2 乗和}}} \quad \text{①}$$

$$= \frac{(x_1 - \bar{x})(y_1 - \bar{y}) + (x_2 - \bar{x})(y_2 - \bar{y}) + \dots + (x_n - \bar{x})(y_n - \bar{y})}{\sqrt{\{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2\} \cdot \{(y_1 - \bar{y})^2 + (y_2 - \bar{y})^2 + \dots + (y_n - \bar{y})^2\}}} \quad \text{②}$$

演習 4  $(x, y) = \{(2, 4), (3, 5), (5, 5), (6, 10)\}$  の散布図を描くとともに、相関係数を計算して下さい。



「xが増えるとyも増えているか？」の指標

$x$	$y$	$x - \bar{x}$ xの偏差	$y - \bar{y}$ yの偏差	$(x - \bar{x})^2$ (xの偏差)の2乗	$(y - \bar{y})^2$ (yの偏差)の2乗	$(x - \bar{x})(y - \bar{y})$ (xの偏差) × (yの偏差)
2	4	-2	-2	4	4	4
3	5	-1	-1	1	1	1
5	5	1	-1	1	1	-1
6	10	2	4	4	16	8
合計→	16			10	22	12
平均→	4					
	$\uparrow \bar{x}$	$\uparrow \bar{y}$		$\uparrow S_{xx}$ 変数 x の 偏差 2 乗和	$\uparrow S_{yy}$ 変数 y の 偏差 2 乗和	$\uparrow S_{xy}$ 変数 x, y の 偏差積和

$$\text{相関係数} = \frac{S_{xy}}{\sqrt{S_{xx} \times S_{yy}}} = \frac{12}{\sqrt{10 \times 22}} = 0.809$$

相関係数を直接計算する Excel 関数

=correl(範囲 1, 範囲 2)



演習5 以下の表は、毎月の平均気温とカゼ薬への支出額を示しています。平均気温とカゼ薬への支出額に関する相関係数を求めて下さい。

平均気温とカゼ薬への支出額の関係

月	平均気温 (℃)①	カゼ薬への 支出額 (円)②	①の偏差	②の偏差	(①の偏 差)の2乗	(②の偏 差)の2乗	(①の偏 差)×(② の偏差)
1	4.4	241					
2	4.8	263					
3	7.6	235					
4	14.3	160					
5	17.3	155					
6	22.7	118					
7	25	111					
8	27.3	100					
9	23.8	113					
10	18	204					
11	11	268					
12	4.4	314					
合計							
平均値③					↑④	↑⑤	↑⑥

(ヒント)

平均値=average(範囲)

偏差: データの値-平均値

偏差の2乗: 偏差<sup>2</sup>

相関係数 = ⑥/sqrt(④\*⑤)

相関係数 = correl(範囲1, 範囲2)

相関係数(上の方法)

相関係数(Excel関数)

