

相関の指標

高木 廣文
東邦大学医学部看護学科
国際保健看護学研究室

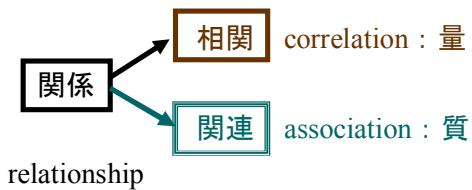
関係を調べるための方法

- ・身長が大きいと体重も重い。
- ・塩分の摂りすぎは高血圧症のもと。
- ・タバコの吸いすぎは肺ガンの原因。



物事には何らかの関係があるものが多い。
「関係」を統計学的に調べるには？

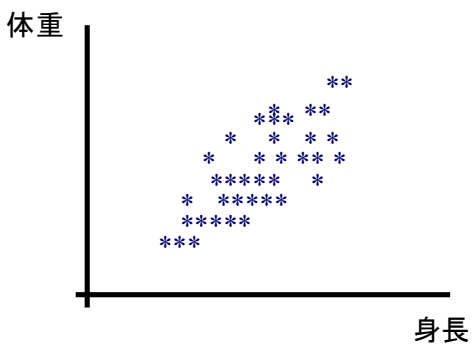
関係の分類



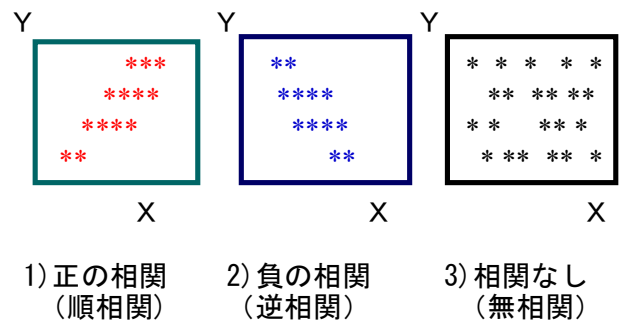
相関関係について

- 2つの量的な変数間の関係の程度を示す。
- 相関図 (散布図)
correlation diagram, scatter plots
2変数間の相関関係を図示。
- 相関係数 correlation coefficient
2変数間の相関関係の大きさを数値化。

相関図の例



いろいろな相関関係



回帰 regression

回帰直線 regression line:

Yの予測値 = [係数] × [Xの値] + [定数]

$$\hat{y} = b x + a$$

最小2乗法によるパラメータ推定

最小2乗法 least squares method

$$L = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - b x_i - a)^2$$

Lを最小にするように、b, aを定める。

回帰直線のパラメータ推定

回帰係数の推定:

$$b = \frac{XとYの共分散}{Xの分散}$$

定数項の推定:

$$a = \bar{y} - b \times \bar{x}$$

共分散 covariance

2変数の変動関係の指標

・ 2変数 X, Yの偏差の積の平均値:

$$\text{共分散 } S_{xy} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

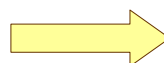
相関係数 correlation coefficient

量的な2変数間の関係の強さを示す指標。
通常は、2変数間の直線的な関係の程度を表すために用いられる。

Pearsonの積率相関係数, 単相関係数
Pearson's product moment cor. coef.
simple correlation coefficient

単相関係数の定義

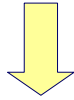
$$r_{xy} = \frac{S_{xy}}{S_x \times S_y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$



$$-1 \leq r_{xy} \leq 1$$

決定係数 coefficient of determination

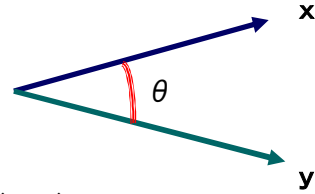
- 相関係数の2乗
一方の変動により，他方の変動を
どれだけ説明することができるか。



$$0 \leq r_{xy}^2 \leq 1$$

相関係数の幾何学的意味

2つのベクトルの余弦と相関係数

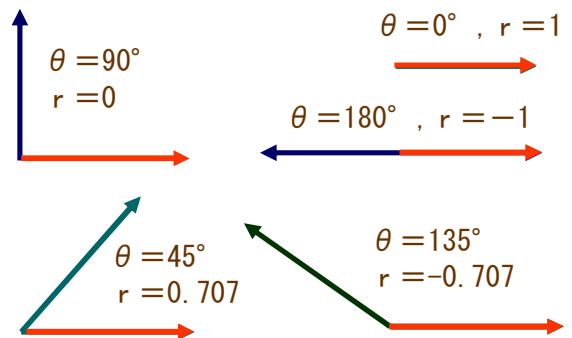


$$\cos \theta = \frac{(\mathbf{x}, \mathbf{y})}{\|\mathbf{x}\| \cdot \|\mathbf{y}\|} = \text{XとYの相関係数 } r_{XY}$$

(参考)ベクトルの知識

- ベクトルの掛け算：
 $\mathbf{x} = (x_1, x_2, \dots, x_N)$, $\mathbf{y} = (y_1, y_2, \dots, y_N)$
 $\mathbf{xy}' = x_1y_1 + x_2y_2 + \dots + x_Ny_N$
 対応する要素の積の和。
 「'」はベクトルの縦，横を転置する意味
- ベクトルのノルム (要素の2乗の和)：
 $\mathbf{xx}' = x_1x_1 + x_2x_2 + \dots + x_Nx_N = \|\mathbf{x}\|^2$

ベクトル表示による相関関係



相関係数の大きさと関係の強さ

絶対値	相関関係の強さ
0.0~0.2	ほとんど相関関係がない
0.2~0.4	やや(弱い)相関関係がある
0.4~0.7	かなり(強い)相関関係がある
0.7~1.0	強い相関関係がある

単相関係数の計算例:

- 5人の身長と体重のデータ

身長	体重
170	65
165	57
180	72
168	55
162	50

平均値 : 169 59.8
分散 : 37.6 60.6

共分散と相関係数の計算

$$\begin{aligned} \text{共分散} &= [(170-169) \times (65-59.8) + (165-169) \times (57-59.8) \\ &\quad + (180-169) \times (72-59.8) + (168-169) \times (55-59.8) \\ &\quad + (162-169) \times (50-59.8)] \div 5 \\ &= 1 \times 5.2 + (-4) \times (-2.8) + 11 \times 12.2 \\ &\quad + (-1) \times (-4.8) + (-7) \times (-9.8) \\ &= 224/5 = 44.8 \end{aligned}$$

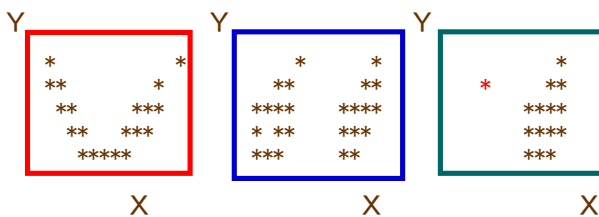
$$\text{単相関係数} = \frac{44.8}{\sqrt{37.6} \times \sqrt{60.6}} = 0.939$$

単相関係数の練習問題

【問1】5人の最高血圧(SysP)と最低血圧(DiaP)の単相関係数を求めよ。

SysP	DiaP
120	80
122	78
115	75
124	60
130	82

問題のある相関図



- 1) 曲線相関 2) 層化が必要 3) 外れ値

順位相関係数 rank cor. coef.

- データは量的でも、その数値自体の信頼性が低い場合。
- データがもともと大小関係のみが与えられた順序尺度である場合。



データの順位に基づく相関係数を求める。

「順位相関係数」

よく用いられる順位相関係数

(1) スピアマンの順位相関係数

Spearman's rank correlation coefficient

順位データにピアソンの積率相関係数の計算方法をそのまま用いたもの。

(2) ケンドールの順位相関係数

Kendall's rank correlation coefficient

データの大小関係のみから計算する。

スピアマンの順位相関係数

N標本中のケース i の2変数のデータ (x_i, y_i) が順位で与えられている場合、スピアマンの順位相関係数 r_s :

$$r_s = 1 - \frac{6 \sum_{i=1}^n (x_i - y_i)^2}{n(n^2 - 1)}$$

タイのある場合のスピアマンの順位相関係数の補正(1)

変数Xにk個の同順位のデータの組があり、各組には t_{xk} 個のデータが含まれている。
 同様に、変数Yには、m個の同順位のデータの組があり、各 t_{ym} 個のデータが含まれているとする。

修正項：

$$T_X = \sum_{i=1}^k t_{xi} (t_{xi}^2 - 1)$$

$$T_Y = \sum_{j=1}^m t_{yj} (t_{yj}^2 - 1)$$

タイのある場合のスピアマンの順位相関係数の補正(2)

修正したスピアマンの順位相関係数：

$$r_s = \frac{N(N^2 - 1) - (T_X + T_Y) / 2 - 6 \sum_i (x_i - y_i)^2}{\sqrt{N(N^2 - 1) - T_X / 2} \sqrt{N(N^2 - 1) - T_Y / 2}}$$

ケンドールの順位相関係数 τ

N組のデータから、任意に2組のデータを取り出した場合：

- ・ 2組でのXとYの順位の大小関係が一致する組の個数をP
- ・ 一致しない組の個数をQ

$$\tau = \frac{2(P - Q)}{n(n - 1)}$$

タイのある場合のケンドールの順位相関係数の修正

順位に同順位(タイ)がある場合：

T_X , T_Y を変数X, Yの各タイの組み合わせ数の総数とする。

$$\tau = \frac{P - Q}{\sqrt{N(N^2 - 1) - T_X / 2} \sqrt{N(N^2 - 1) - T_Y / 2}}$$

順位相関の計算例(1)

●5人の身長と体重のデータ

スピアマンの順位相関係数のための表

身長	体重	身長 順位	体重 順位	差	差の 2乗
170	65	4	4	0	0
165	57	2	3	-1	1
180	72	5	5	0	0
168	55	3	2	1	1
162	50	1	1	0	0

順位相関の計算例(2)

順位の差の2乗和 = $0 + 1 + 0 + 1 + 0 = 2$

スピアマンの順位相関係数

$$= 1 - 6 \times 2 / [5 \times (5 \times 5 - 1)]$$

$$= 1 - 12 / 120$$

$$= 0.9$$

順位相関の計算例(3)

ケンドールの順位相関係数の計算のためのデータの大きさの比較

番号	1	2	3	4	5
1	--				
2	大,大	--			
3	小,小	小,小	--		
4	大,大	小,大	大,大	--	
5	大,大	大,大	大,大	大,大	--
一致	4	2	2	1	--
不一致	0	1	0	0	--

順位相関の計算例(4)

- ・一致数 $P = 4 + 2 + 2 + 1 = 9$
- ・不一致数 $Q = 0 + 1 + 0 + 0 = 1$

ケンドールの順位相関

$$\begin{aligned}
 &= 2 \times (9 - 1) / [5 \times (5 - 1)] \\
 &= 16 / 20 \\
 &= 0.8
 \end{aligned}$$