１１．２変数の関係の分析
（相関分析）
（連関分析）

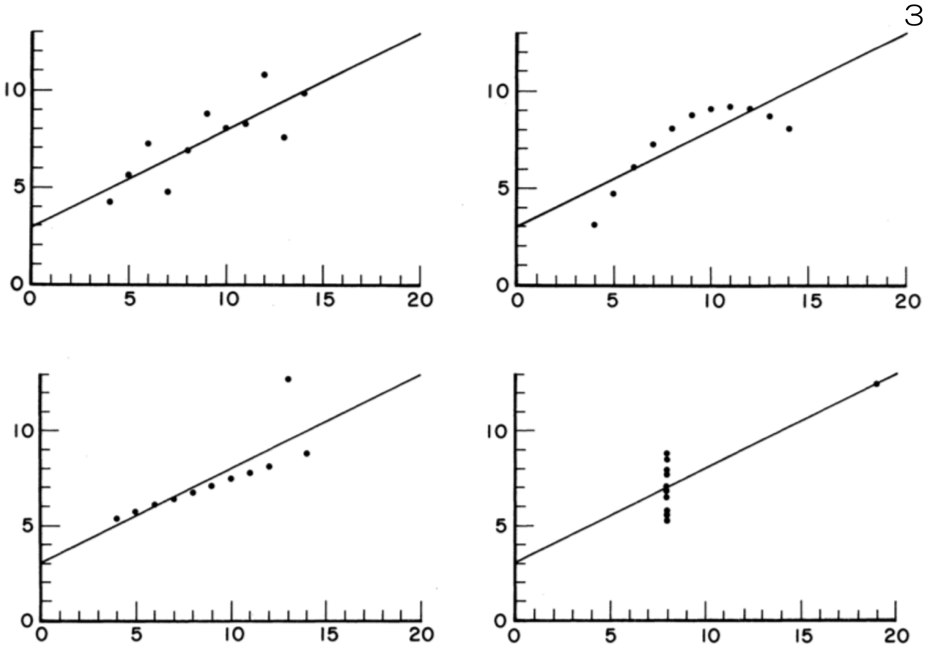
内　容

１．相関係数

２．連関係数

３．順位相関係数

3



(Anscombe 1973)

4

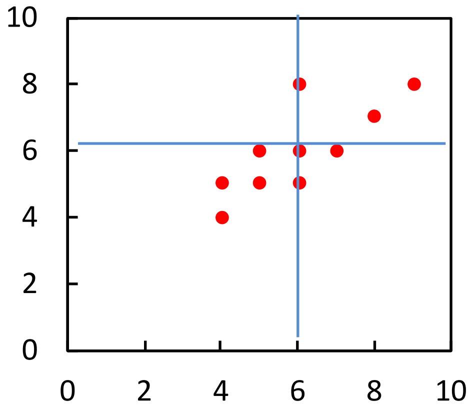
1．相関係数

5

散布図、共分散

$$(x_i - \bar{x})(y_i - \bar{y})$$

平均を原点として、
右上と左下は正
右下と左上は負



共分散

$$s_{xy} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})$$

6

（Pearsonの積率）相関係数 r

完全な正の相関があるような場合

$$s_{xy} = s_x s_y$$

完全な負の相関があるような場合

$$s_{xy} = -s_x s_y$$

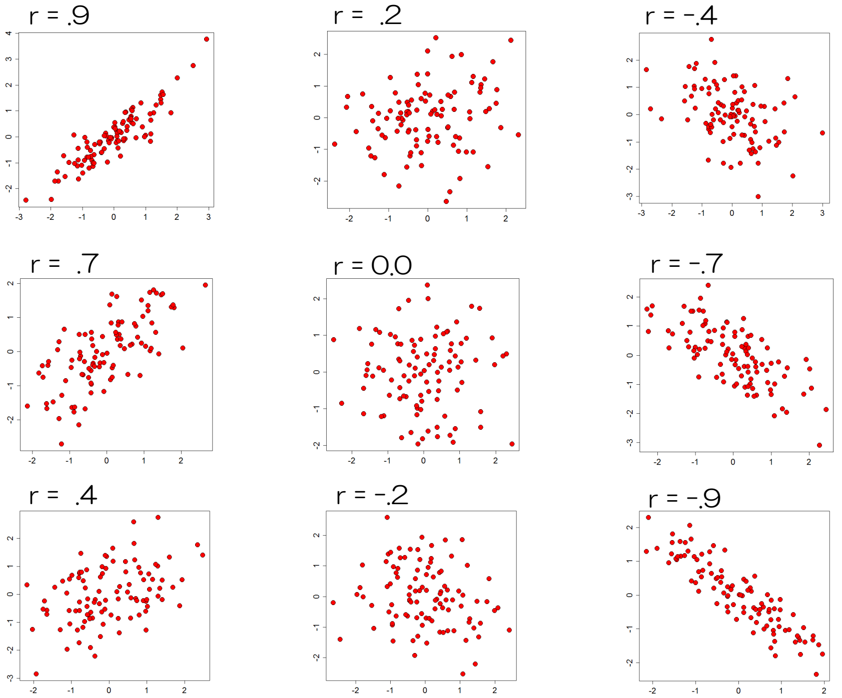
共分散の取りうる範囲

$$-s_x s_y \leq s_{xy} \leq s_x s_y$$

$s_x s_y$ で割ると

$$-1 \leq \frac{s_{xy}}{s_x s_y} \leq 1$$

7



8

関連の強さに関する、おおよその目安

$.7 < \lvert r \rvert \le 1.$	強い相関あり
$.4 < \lvert r \rvert \le .7$	比較的強い相関あり
$.2 < \lvert r \rvert \le .4$	弱い相関あり
$0 \le \lvert r \rvert \le .2$	ほとんど相関なし

9

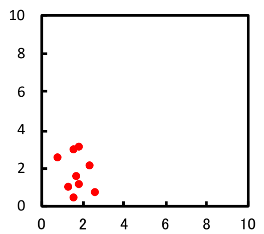
相関係数の留意点

- 外れ値の影響
- 切断効果
- 分割相関（層別相関）
- 曲線相関
- 疑似相関

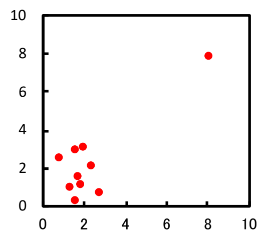
10

外れ値の影響

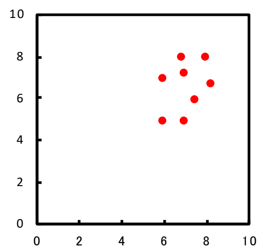
r = -.232



r = .844



切断効果
選抜効果



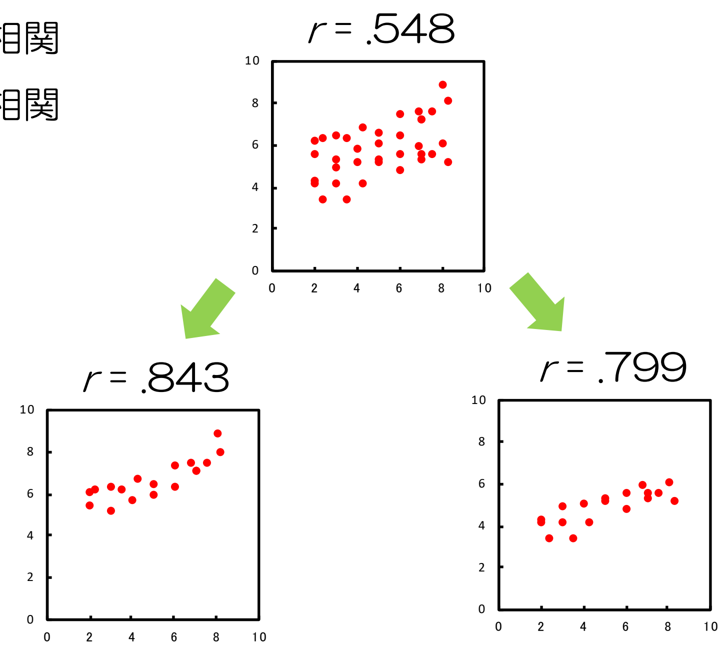
r = .765

r = .329　（横軸6以上）

11

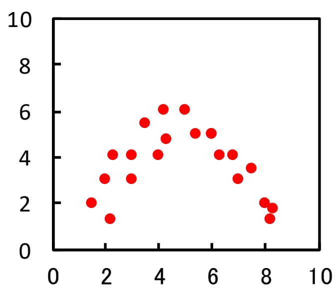
分割相関

層別相関



12

曲線相関



r = .123

疑似相関

身長

相関？

小学校の学年

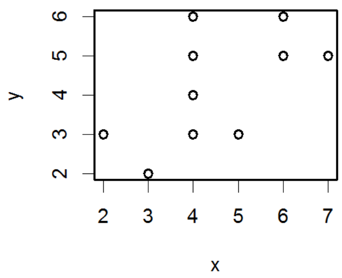
学力

13

x	6	4	4	2	6	5	4	7	4	3
y	5	4	3	3	6	3	5	5	6	2

```
x <- c(6, 4, 4, 2, 6, 5, 4, 7, 4, 3)
y <- c(5, 4, 3, 3, 6, 3, 5, 5, 6, 2)
cor(x,  y)
```

```
[1] 0.5791078
```



14

無相関検定

H_0 ：母相関係数＝0

H_1 ：母相関係数≠0

$$t_0 = \frac{|r|\sqrt{n-2}}{\sqrt{1-r^2}} \sim t_{n-2}\text{分布}$$

15

x	6	4	4	2	6	5	4	7	4	3
y	5	4	3	3	6	3	5	5	6	2

```
x <- c(6, 4, 4, 2, 6, 5, 4, 7, 4, 3)
y <- c(5, 4, 3, 3, 6, 3, 5, 5, 6, 2)
cor.test(x, y)
```

ピアソンの積率相関係数

データ: x と y
t値 = 2.0092, 自由度 = 8, P値 = 0.07938
対立仮説: 母相関は, 0ではない
95 パーセント信頼区間: −0.07950929
0.88576510
標本推定値:
相関係数
0.5791078

16

相関の差の検定

H_0 ： 母相関係数の間に差がない
H_1 ： いずれかの母相関係数の間に差がある

互いに独立な k 個の標本の相関係数の有意差検定

$$z' = \frac{1}{2}\log_e \frac{1+r}{1-r}$$

フィッシャーの z' 変換

$$\chi^2 = \sum_i^k z_i'^2 (N_i - 3) - \left\{ \sum_i^k z'_i (N_i - 3) \right\}^2 / \sum_i^k (N_i - 3)$$

～ χ^2（k-1）分布

17

group	1	2	3	4
n	25	20	30	25
r	0.543	0.458	0.632	0.521

```
n <- c(25,20,30,25)
r <- c(0.543,0.458,0.632,0.521)
eq.cor(n, r)
```

標本相関係数の同等性の検定

```
data:  n and r
chi sq. = 0.729, df = 3, p-value = 0.8664
sample estimates:
Estimated rho
    0.5515522
```

18

2．連関係数

カテゴリカルなデータの相関

２×２分割表

- 四分点相関係数
- ϕ係数
- Yuleの連関係数（Q）

$l \times m$分割表

- Craméｒの連関係数（V）
- Goodman-Kruskalの順序連関係数（γ）

四分点相関係数

各カテゴリーを0と1に置き換えて、相関係数を求めたもの

	はい	いいえ
男子	2	3
女子	1	2

ケース	1	2	3	4	5	6	7	8
性別	1	1	0	1	1	1	0	0
意見	1	1	1	0	0	0	0	0

φ係数

$$\phi = \sqrt{\frac{\chi^2}{n}}$$

$$0 \le \phi \le 1$$

	はい	いいえ
男子	23	26
女子	12	19

```
dat <- matrix(c(23,26,12,19),ncol=2, byrow=T)
x <- chisq.test(dat,correct=F)
(a <- sqrt(x$statistic/sum(dat)))
```

```
[1] 0.08081479
```

21

Cramér の連関係数（V）

$l \times m$ 分割表（クロス集計表）

	a_1	a_2	a_3	Σ
b_1				
b_2				
Σ				

$$V = \sqrt{\frac{\chi^2}{(\min(l,m)-1)\times n}}$$

$$0 \le V \le 1$$

22

名義尺度の例題）

	優	良	可	不可
A	7	12	18	13
B	11	15	15	9
C	20	12	13	5

```
dat <-
matrix(c(7,12,18,13,11,15,15,9,20,12,13,5),ncol=4,byrow=T)
x <- chisq.test(dat,correct=F)
(v <- sqrt(x$statistic/((min(ncol(dat),nrow(dat))-
1)*sum(dat))))
```

```
[1] 0.1986889
```

23

関連の強さに関する、おおよその目安

$.8 < r_c$	非常に強い相関あり
$.5 \le r_c < 0.8$	やや強い相関あり
$.25 \le r_c < 0.5$	やや弱い相関あり
$r_c < 0.25$	ほとんど相関なし

24

3．順位相関係数

順位相関係数

Spearman

Kendall

一致性係数

25

Spearmanの順位相関係数 r_s

順位をデータ値としてスピアマンの積率相関係数を計算したもの

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

関連の強さの目安は積率相関係数と同じ

例題）

26

A、Bの２名で８つのサンプルに順位を付けた。２名の順位の付け方について、スピアマン順位相関係数を求め、関連性について検討せよ。

	1	2	3	4	5	6	7	8	9	10
A	8	9	2	10	1	4	5	3	7	6
B	6	10	2	9	4	5	3	1	7	8

```
dataA <- c(8,9,2,10,1,4,5,3,7,6)
dataB <- c(6,10,2,9,4,5,3,1,7,8)
cor.test(dataA,dataB,method="spearman")
```

27

```
dataA <- c(8,9,2,10,1,4,5,3,7,6)
dataB <- c(6,10,2,9,4,5,3,1,7,8)
cor.test(dataA,dataB,method="spearman")
```

スピアマンの順位相関係数

データ： dataA と dataB
S = 28, P値 = 0.005557
対立仮説：母相関（ρ）は，0ではない
標本推定値：
ρ
0.830303

28

Kendallの順位相関係数　τ

$$\tau = \frac{2P}{\frac{1}{2}n(n-1)} = \frac{4P}{n(n-1)} - 1$$

Pは大小関係が一致する組の数

例題）

A、Bの２名で８つのサンプルに順位を付けた。
２名の順位の付け方について、ケンドール順位相関係数を求め、関連性について検討せよ。

	1	2	3	4	5	6	7	8	9	10
A	8	9	2	10	1	4	5	3	7	6
B	6	10	2	9	4	5	3	1	7	8

```
dataA <- c(8,9,2,10,1,4,5,3,7,6)
dataB <- c(6,10,2,9,4,5,3,1,7,8)
cor.test(dataA,dataB,method="kendall")
```



```
dataA <- c(8,9,2,10,1,4,5,3,7,6)
dataB <- c(6,10,2,9,4,5,3,1,7,8)
cor.test(dataA,dataB,method="kendall")
```

ケンドールの順位相関係数

データ：　dataA と dataB
T = 36, P値 = 0.01667
対立仮説：母相関（τ）は，0ではない
標本推定値：
　τ
0.6

31

Kendallの一致性係数　W

k組、n個の資料

$$S = \sum_{j=1}^{n} \left\{ S_j - \frac{1}{2} k(n+1) \right\}^2$$

$$W = \frac{12S}{k^2(n^3 - n)}$$

32

例題）

A、B、Cの２名で８つのサンプルに順位を付けた。２名の順位の付け方について、ケンドールの一致性係数を求め、関連性について検討せよ。

	1	2	3	4	5	6	7	8	9	10
A	8	9	2	10	1	4	5	3	7	6
B	6	10	2	9	4	5	3	1	7	8
C	7	10	3	9	2	4	6	1	8	5

33

```
source("all.R", encoding="euc-jp")
dat <- matrix ( c (
                8,9,2,10,1,4,5,3,7,6,
                6,10,2,9,4,5,3,1,7,8,
                7,10,3,9,2,4,6,1,8,5
), byrow=FALSE, ncol=3)
kendall.w(dat)
```

ケンドールの一致度係数

データ: dat
Kendall W = 0.9111, chi sq. = 24.6000, 自由度 = 9,
P値 = 0.003447

34

ポリコリック相関係数

順序尺度と順序尺度の相関係数

```
install.packages("polycor")
library(polycor)

polychor(x,y)
```

35

ポリシアル相関係数

連続尺度と順序尺度の相関係数

install.packages(″polycor″)
library(polycor)

polyserial(x,y)

xが連続変数、yが順序変数

36

尺度が混在する場合

事前に順序尺度をfactorで指定

install.packages(″polycor″)
library(polycor)

hetcor(x,y)