

(今日の準備)  
回しているUSBメモリから  
「data\_4」というフォルダを  
デスクトップにコピーしてく  
ださい

## Rコマンダーを用いた統計解析 (4)

担当 繁榎 算男 教授

アシスタント 森 一将

### 前回の補足(1)

- USBに入っていなかったパッケージ類について
  - Rコマンダー起動時に一部パッケージのインストールを要求されました。
  - これらのパッケージは実習上必要のないものですが、別途ご用意しましたのでインストールしてください。



### 前回の補足(2)

- パッケージの読み込み手順
  - 配布したフォルダをデスクトップ上にコピー
  - Rを起動し以下の作業を実施
    - 「パッケージ」 「ローカルにあるZipファイルからのパッケージのインストール」
    - 「追加パッケージ」フォルダにあるすべてのZipファイルの選択
  - その後、Rcmdrを起動してください

## 今日の演習事項

2009年7月20日

- 2変数のデータ解析
  - 散布図を描く
  - 相関係数を求める
- 3変数(以上)のデータ解析
  - Excelファイルからのデータ読み込み
  - 散布図を描く
  - 相関係数を求める

5

心理学演習資料

## Rコマンドを用いたデータ読み込み - 直接読み込み -

2009年7月20日

- クリップボードからの直接読み込み
- 「ボール投げデータ」を用いる

距離	身長
22	146
30	169
28	160
25	156
27	161
29	168
28	154
25	153
28	160
26	152

6

心理学演習資料

## Rコマンドの起動

2009年7月20日

- Rコマンドの起動
  - USBにあるRのアイコンをダブルクリック
  - 上の「パッケージ」メニュー 「パッケージの読み込み」
  - 「Rcmdr」を選択し、下の「OK」ボタンを押す



7

## 2変数のデータ解析

2009年7月20日

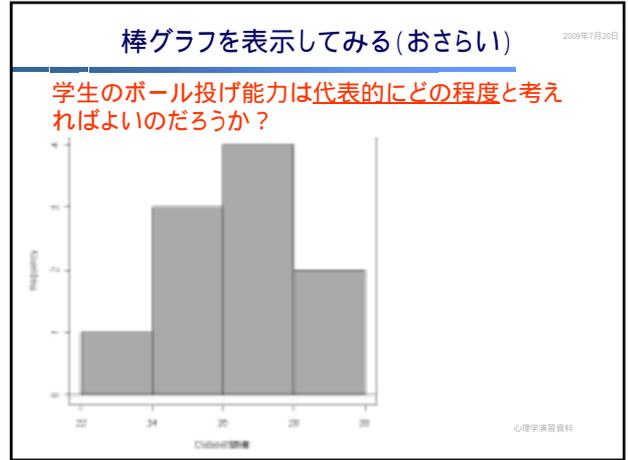
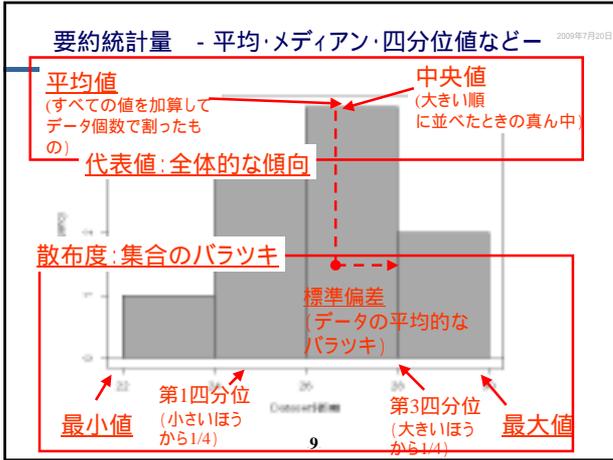
【ある学校におけるボール投げの距離と記録】

距離(m)	身長(cm)
22	146
30	169
28	160
25	156
27	161
29	168
28	154
25	153
28	160
26	152

出典: www.qmss.jp

心理学演習資料

8

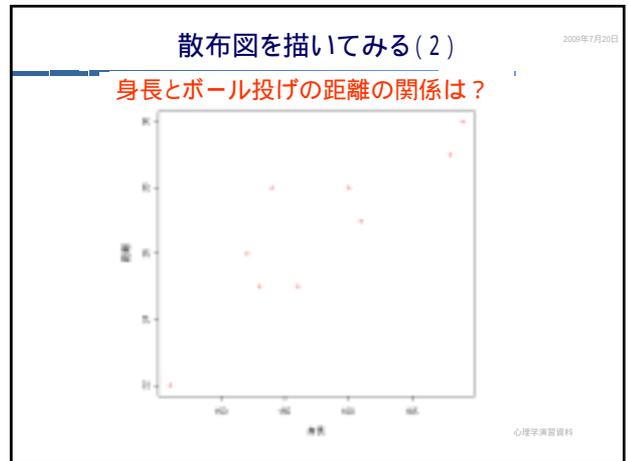


### 散布図を描いてみる(1)

2009年7月20日

- Rコマンドの「グラフ」「散布図」
- Xに「身長」Yに「距離」
- チェックをはずしOK

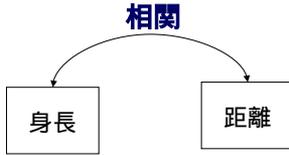
心理学演習資料



要約統計量 - 相関係数-(1)

2009年7月20日

- 相関変数 = 2つの変数の関係の強さを示したもの
- 1から1の数値で表示



13

心理学演習資料

要約統計量 - 相関係数-(2)

2009年7月20日

X:身長、Y(ボール投げの)距離 とすると  
 - 共分散: 「平均からの偏差の積」の平均

$$s_{xy} = \frac{(x_1 - \bar{x})(y_1 - \bar{y}) + (x_2 - \bar{x})(y_2 - \bar{y}) + \dots + (x_n - \bar{x})(y_n - \bar{y})}{n}$$

- 標準偏差: 分散の平均

$$s_x = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n}}$$

$$s_y = \sqrt{\frac{(y_1 - \bar{y})^2 + (y_2 - \bar{y})^2 + \dots + (y_n - \bar{y})^2}{n}}$$

14

心理学演習資料

要約統計量 - 相関係数-(3)

2009年7月20日

- 相関係数

$$r_{xy} = \frac{s_{xy}}{s_x s_y}$$

← 単位を揃えるための乗数。

- 相関変数 = 2つの変数の「関係」の強さを示したもの
- 1から1の数値で表示

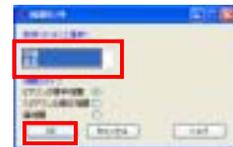
15

心理学演習資料

要約統計量 - 相関係数-(4)

2009年7月20日

- 相関変数 = 2つの変数の「関係」の強さを示したもの
- 1から1の数値で表示
- Rコマンダーより「統計量」「要約」「相関行列」
- 「距離」と「身長」を選択し、OK



心理学演習資料

## 要約統計量 - 相関係数 - (5)

2009年7月20日

### 相関行列の表示

```
> cor(Dataset[,c("距離", "身長")], use="complete.obs")
      距離      身長
距離  1.0000000  0.8764228
身長  0.8764228  1.0000000
```

身長と距離の間の相関係数

17

心理学演習資料

## 要約統計量 - 相関係数 - (6)

2009年7月20日

(参考) 相関係数の値の評価(山田、村井、2004等)

- $0 < r < 0.2$  または  $-0.2 < r < 0$  **ほとんど相関なし**
- $0.2 < r < 0.4$  または  $-0.4 < r < -0.2$  **弱い相関**
- $0.4 < r < 0.7$  または  $-0.7 < r < -0.4$  **中程度の相関**
- $0.7 < r < 1$  または  $-1 < r < -0.7$  **強い相関**

18

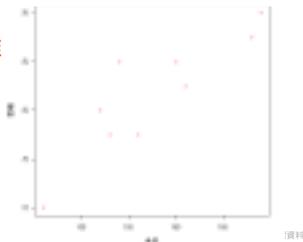
心理学演習資料

## データ分析による考察(1)

2009年7月20日

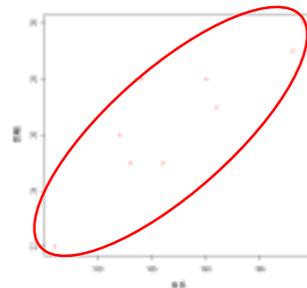
```
> cor(Dataset[,c("距離", "身長")], use="complete.obs")
      距離      身長
距離  1.0000000  0.8764228
身長  0.8764228  1.0000000
```

身長とボール投げの距離の関係は？



## データ分析による考察(2)

2009年7月20日



• 横軸(身長)が右にいくと縦軸(距離)が上に行く傾向

--

## データ分析による考察(2)

2009年7月20日

	距離	身長
距離	1.000000	0.8764228
身長	0.8764228	1.000000

- 正の相関
- 1に近い 強い正の相関

21

心理学演習資料

## 3変数のデータ解析

2009年7月20日

- 年齢・血圧・コレステロールの関係

年齢	血圧	コレステロール
25	108	187
22	108	179
24	100	229
27	128	164
33	102	170
36	130	211
31	112	151
38	108	148
41	134	270
47	134	183
48	176	185
46	126	190
53	108	213
50	140	193
55	156	174
54	162	164
66	142	264
68	164	282
64	152	250
66	178	175

出典: [www.qmss.jp](http://www.qmss.jp)

22

心理学演習資料

## ファイルからのデータ読み込み

2009年7月20日

- Rコマンドでは外部データファイルからの読み込みも可能
- 「データ」「データのインポート」「Excel またはAccess dBaseのデータセットから」
- データセット名を入力
- デスクトップの「data\_4」フォルダにある「blooddata.xls」を選択

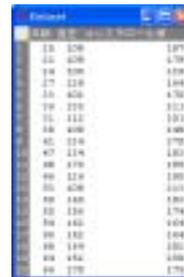
23

心理学演習資料

## 読み込めたかを確認

2009年7月20日

- データセットを表示させてみてください。



年齢	血圧	コレステロール
25	108	187
22	108	179
24	100	229
27	128	164
33	102	170
36	130	211
31	112	151
38	108	148
41	134	270
47	134	183
48	176	185
46	126	190
53	108	213
50	140	193
55	156	174
54	162	164
66	142	264
68	164	282
64	152	250
66	178	175

年齢

血液

コレステロール

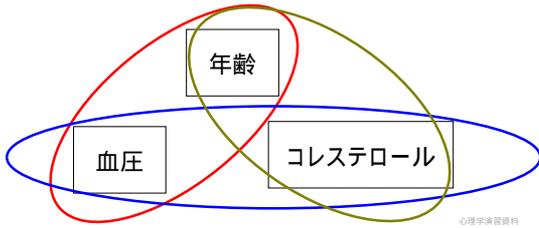
24

心理学演習資料

### 3変数の散布図

2009年7月20日

- 散布図
  - (通常は)2変数間の関係を示したもの
  - 3変数の場合は3つの散布図が必要



25

心理学演習資料

### Rコマンドによる3変数の散布図(1)

2009年7月20日

- 「グラフ」「散布図行列」
- 「最小2乗直線」「平滑線」のチェックをはずす
- 「コレステロール値」「血圧」「年齢」を選択
- 「ヒストグラム」を選択

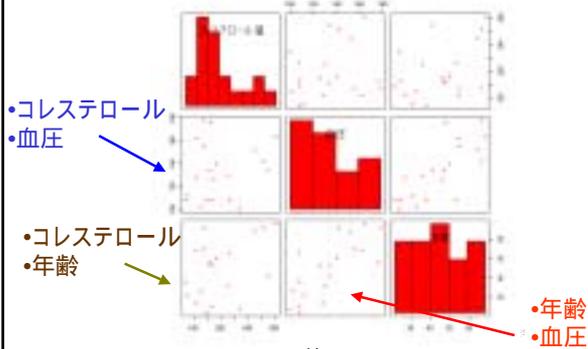


26

心理学演習資料

### Rコマンドによる3変数の散布図(2)

2009年7月20日



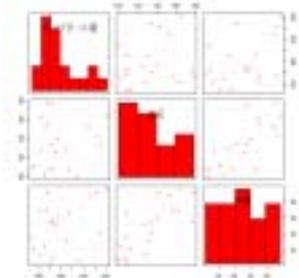
27

心理学演習資料

### Rコマンドによる3変数の散布図(3)

2009年7月20日

- コレステロール値が高い人はどのような人だろうか？



心理学演習資料

## 要約統計量 - 3変数の相関係数-(1)

2009年7月20日

- 相関係数を計算してみる



```
PS C:\Users\user> .\ps\ps13\ps13_01.ps -c "コレステロール値", "血圧", "年齢" | .\ps\ps13\ps13_02.ps  
コレステロール値    血圧    年齢  
平均値 1.09000000 0.18500000 0.42749822  
標準偏差 0.18500000 0.18000000 0.74247258  
相関係数 0.42749822 0.74247258 1.00000000
```

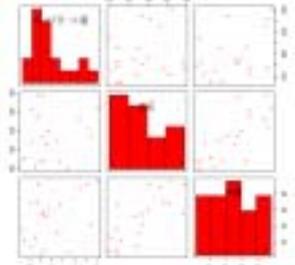
心理学演習資料

29

## データ分析による考察(1)

2009年7月20日

- コレステロール値が高い人はどのような人だろうか？
- 血圧の高い人は？

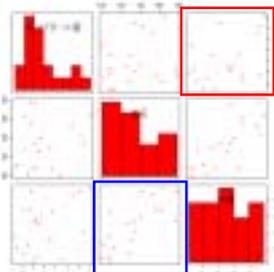


```
PS C:\Users\user> .\ps\ps13\ps13_01.ps -c "コレステロール値", "血圧", "年齢" | .\ps\ps13\ps13_02.ps  
コレステロール値    血圧    年齢  
平均値 1.09000000 0.18500000 0.42749822  
標準偏差 0.18500000 0.18000000 0.74247258  
相関係数 0.42749822 0.74247258 1.00000000
```

## データ分析による考察(2)

2009年7月20日

- コレステロール値と年齢の相関が高い
- 血圧と年齢の関係が高い



```
PS C:\Users\user> .\ps\ps13\ps13_01.ps -c "コレステロール値", "血圧", "年齢" | .\ps\ps13\ps13_02.ps  
コレステロール値    血圧    年齢  
平均値 1.09000000 0.18500000 0.42749822  
標準偏差 0.18500000 0.18000000 0.74247258  
相関係数 0.42749822 0.74247258 1.00000000
```

## (注意)相関係数と因果関係(1)

2009年7月20日

- 因果関係  
- 原因 結果の「ヤジルシの向き」が存在
- 相関係数  
- 「ヤジルシの向き」は分からない

32

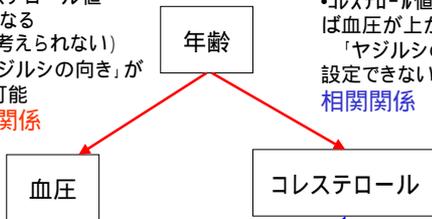
心理学演習資料

## (注意) 相関関係と因果関係 (2)

2009年7月20日

年齢が高くなれば  
• 血圧  
• コレステロール値  
が高くなる  
(逆は考えられない)  
「ヤジルシの向き」が  
設定可能  
**因果関係**

• 血圧が高くなればコレステロール値が高くなる？  
• コレステロール値が高くなれば血圧が上がる？  
「ヤジルシの向き」が  
設定できない  
**相関関係**



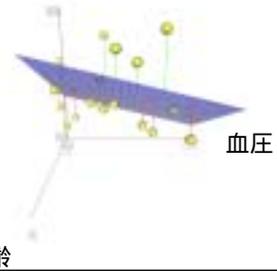
一般的にデータ解析からのみでは因果関係は判明しないことが多い  
因果関係を考えなければいけないときは注意すること

## おまけ

2009年7月20日

- 3次元の散布図

コレステロール値



心理学演習資料