

1 2. 多変量解析 1  
(重回帰分析)

主な多変量解析

2

			説明変数	
			量的	質的
目的変数	あり	量的	重回帰分析	数量化Ⅰ類
		質的	判別分析	数量化Ⅱ類
	なし		主成分分析 因子分析	数量化Ⅲ類 コレスポンデント分析
			多次元尺度構成法	数量化Ⅳ類

その他 クラスタ分析  
共分散構造分析

3

説明変数：独立変数、予測変数

目的変数：従属変数、基準変数

4

## 1. 単回帰分析

各データの構造

$$y_{\alpha} = b + ax_{\alpha} + \varepsilon_{\alpha}$$

$\alpha$  : 1, 2, ..., n (nはサンプル数)

a : 回帰係数

b : 切片

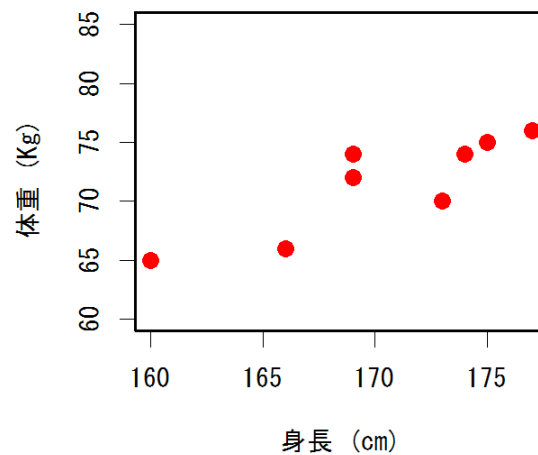
$\varepsilon$  : 誤差

回帰式

$$\hat{y} = b + ax$$

身長	166	173	177	160	174	175	169	169	<sup>5</sup>
体重	66	70	76	65	74	75	74	72	

身長から体重を予測する



身長	166	173	177	160	174	175	169	169	<sup>6</sup>
体重	66	70	76	65	74	75	74	72	

```

height <- c(166, 173, 177, 160, 174, 175, 169, 169)
weight <- c(66, 70, 76, 65, 74, 75, 74, 72)
taikei <- data.frame(height, weight)
colnames(taikei) <- c("身長", "体重")
rownames(taikei) <- 1:8
taikei.lm <- lm(体重~身長, taikei)
summary(taikei.lm)

```

7

```

Call:
lm(formula = 体重 ~ 身長, data = taikei)

Residuals:
    Min       1Q   Median       3Q      Max
-3.1841 -0.5560  0.2119  0.7451  3.3822

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -37.8083    26.4304  -1.430  0.20253
身長          0.6416     0.1551   4.138  0.00609 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
                 ' ' 1

Residual standard error: 2.278 on 6 degrees of freedom
Multiple R-squared:  0.7405,    Adjusted R-squared:  0.6972
F-statistic: 17.12 on 1 and 6 DF,  p-value: 0.006095

```

8

モデル式



```

Call:
lm(formula = 体重 ~ 身長, data = taikei)

Residuals: 残差の基本統計量
    Min       1Q   Median       3Q      Max
-3.1841 -0.5560  0.2119  0.7451  3.3822

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -37.8083    26.4304  -1.430  0.20253
身長          0.6416     0.1551   4.138  0.00609 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05
                 '.' 0.1 ' ' 1

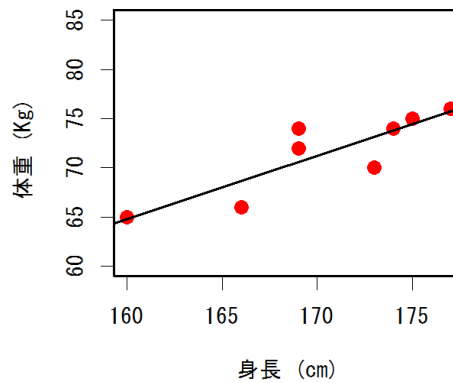
Residual standard error: 2.278 on 6 degrees of freedom
Multiple R-squared:  0.7405,    Adjusted R-squared:

```

```
par(cex=1.8, pch=16, lwd=3)
plot(taikei, ylim=c(60, 85), cex=1.5, col="red",
     xlab="身長 (cm)", ylab="体重 (Kg) ")
abline(taikei.lm)
```

9

残差  $\varepsilon_a = y_a - \hat{y}_a$



個々の残差の表示

10

taikei.lm\$residuals でも同じ



```
> (残差 <- residuals(taikei.lm))
```

```
      1          2          3          4          5
-2.6931094 -3.1841343  0.2495657  0.1563405  0.1742907
```

CCoefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-37.8083	26.4304	-1.430	0.20253
身長	0.6416	0.1551	4.138	0.00609 **

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.278 on 6 degrees of freedom  
 Multiple R-squared: 0.7405, Adjusted R-squared: 0.6972  
 F-statistic: 17.12 on 1 and 6 DF, p-value: 0.006095

標準誤差 t 値 p 値  
 切片  
 回帰係数

回帰係数

残差の2乗和を最小にする a と b

$$a = \frac{S_{xy}}{S_{xx}} = \frac{\sum (x_{\alpha} - \bar{x})(y_{\alpha} - \bar{y})}{\sum (x_{\alpha} - \bar{x})^2}$$

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

回帰式

$$Weight = -37.8083 + 0.6416 \times Height$$

回帰係数の標準誤差、 $t$  値

13

$$\text{残差平方和} \quad S_e = \sum (y_\alpha - \hat{y}_\alpha)^2$$

$$\text{残差不偏分散} \quad s_e^2 = \frac{S_e}{n - k - 1}$$

$n$ : 標本数、 $k$ : 説明変数の数

回帰係数の標準誤差

$$SE(a) = \sqrt{s_e^2 \left[ \frac{1}{n} + \frac{\bar{x}^2}{\sum (x_\alpha - \bar{x})^2} \right]}$$

切片の標準誤差

$$SE(b) = \sqrt{\frac{s_e^2}{\sum (x_\alpha - \bar{x})^2}}$$

14

回帰係数の  $t$  値

$$t_a = \frac{a}{SE(a)}$$

切片の  $t$  値

$$t_b = \frac{b}{SE(b)}$$

推定値の標準誤差＝残差の標準偏差

Residual standard error: 2.278 on 6 degrees of freedom  
 Multiple R-squared: 0.7405, Adjusted R-squared:  
 0.6972  
 F-statistic: 17.12 on 1 and 6 DF, p-value: 0.006095

モデル全体のF値

決定係数  $R^2$  (寄与率) 調整済み決定係数  $\tilde{R}^2$

$$\frac{\text{回帰の変動}}{\text{目的変数の変動}}$$

0.8 <  $R^2$  かなりよい  
 0.5 <  $R^2$  < 0.8 まあよい

決定係数  $R^2$   $R^2 = \frac{S_R}{S_T}$

調整済み決定係数  $\tilde{R}^2$

$$\tilde{R}^2 = 1 - \frac{S_e / n - k - 1}{S_T / (n - 1)}$$

$n$ : 標本数、 $k$ : 説明変数の数

$$S_T = \sum (y_\alpha - \bar{y})^2 \quad S_R = \sum (\hat{y}_\alpha - \bar{y})^2$$

$$S_e = \sum (y_\alpha - \hat{y})^2 \quad S_T = S_R + S_e$$



17

 $F$  値

$$F = \frac{R^2}{1-R^2} \times \frac{n-k-1}{k}$$

18

予測値、残差

```
予測 <- predict(taikei.lm)
data.frame(taikei, 予測, 残差)
```

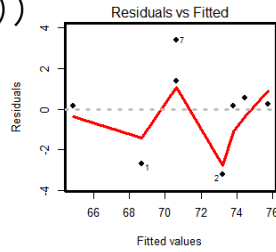
	身長	体重	予測	残差
1	166	66	68.69311	-2.6931094
2	173	70	73.18413	-3.1841343
3	177	76	75.75043	0.2495657
4	160	65	64.84366	0.1563405
5	174	74	73.82571	0.1742907
6	175	75	74.46728	0.5327157
7	169	74	70.61783	3.3821656
8	169	72	70.61783	1.3821656

回帰診断図

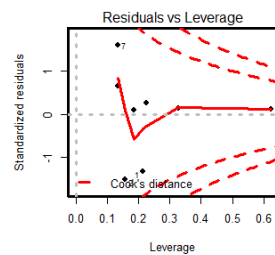
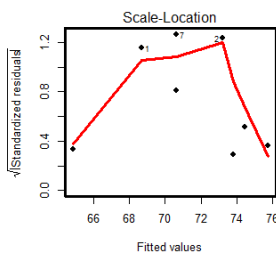
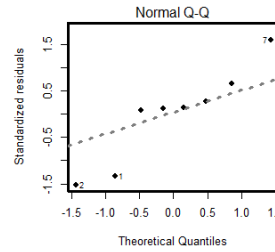
19

```
par(mfrow=c(2, 2))
plot(taikei.lm)
```

残差と予測値



残差の正規Q-Q



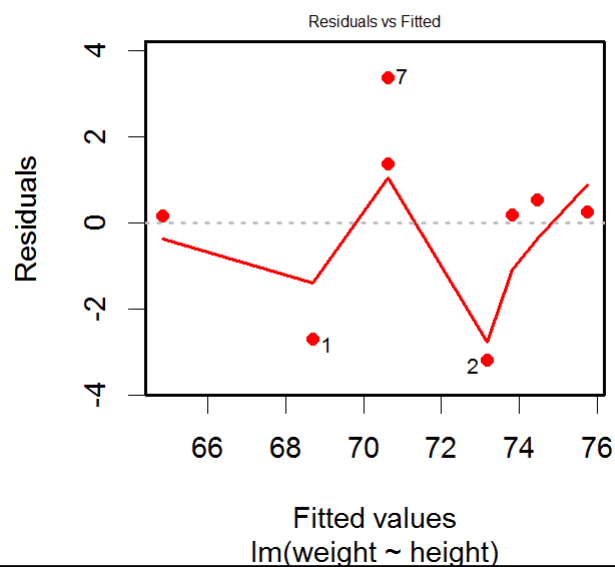
残差の平方根

てこ比・標準化残差

残差と予測値 (フィット値)

20

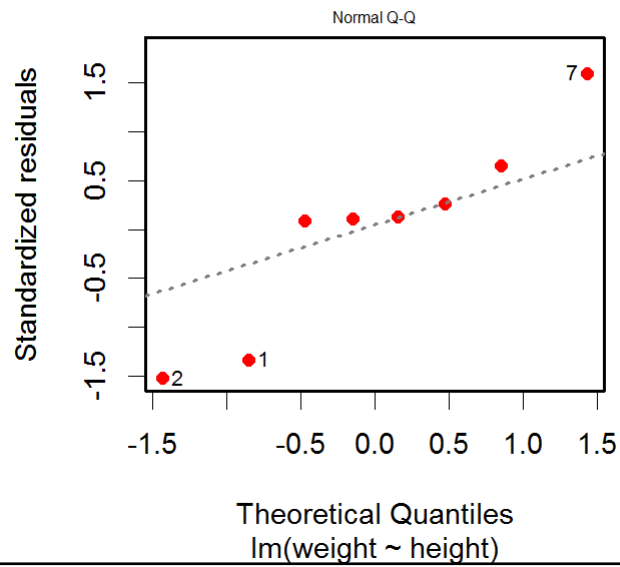
```
plot(taikei.lm, col="red", which=1)
```



残差の正規Q-Q

21

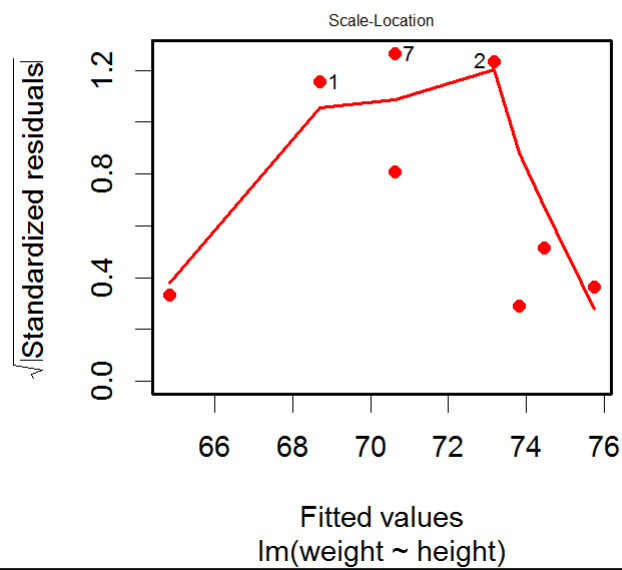
`plot(taiei.lm, col="red", which=2)`



残差の平方根

22

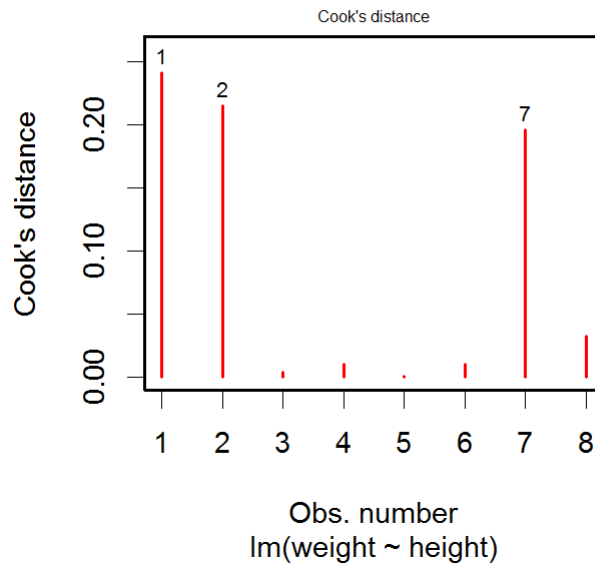
`plot(taiei.lm, col="red", which=3)`



Cookの距離

23

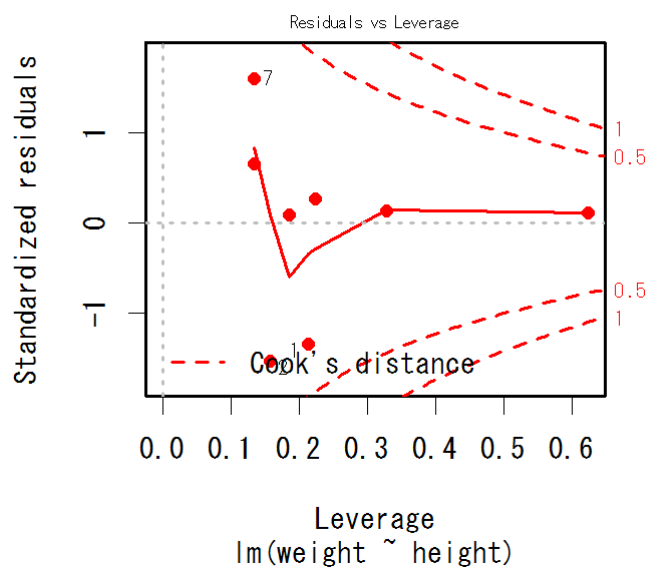
plot(taikei.lm, col="red", which=4)



てこ比・標準化残差

24

plot(taikei.lm, col="red", which=5)



例題) データセットcarsを使って、速度から  
停止距離を回帰分析せよ

> (cars)

	speed	dist
1	4	2
2	4	10
3	7	4
4	7	22
...	...	...
47	24	92
48	24	93
49	24	120
50	25	85

組み込みのデータセットは

> data()

で一覧を見ることができる

さらにヘルプで詳細が表示される

## 2. 重回帰分析

各データの構造 (説明変数が  $i$  個)

$$y_{\alpha} = b + a_1 x_{\alpha 1} + a_2 x_{\alpha 2} + \cdots + a_i x_{\alpha i} + \varepsilon_{\alpha}$$

$\alpha$  : 1, 2, ..., n (nはサンプル数)

$a_i$  : 偏回帰係数

$b$  : 切片

$\varepsilon$  : 誤差

重回帰式

$$y = b + a_1 x_1 + a_2 x_2 + \cdots + a_i x_i$$

誤差の独立性、不偏性、等分散性、正規性を仮定

身長	体重	ウエスト	胸囲	27
166	66	79	83	
173	70	78	80	
177	76	80	86	
160	65	78	81	
174	74	79	84	
175	75	81	84	
169	74	84	91	
169	72	81	85	

```

height <- c(166, 173, 177, 160, 174, 175, 169, 169)
weight <- c(66, 70, 76, 65, 74, 75, 74, 72)
waist <- c(79, 78, 80, 78, 79, 81, 84, 81)
chest <- c(83, 80, 86, 81, 84, 84, 91, 85)
taikei <- data.frame(height, weight, waist, chest)
colnames(taikei) <- c("身長", "体重", "ウエスト", "胸囲")
rownames(taikei) <- 1:8

```

身長	体重	ウエスト	胸囲	28
166	66	79	83	
173	70	78	80	
177	76	80	86	
160	65	78	81	
174	74	79	84	
175	75	81	84	
169	74	84	91	
169	72	81	85	

```

height <- c(166, 173, 177, 160, 174, 175, 169, 169)
weight <- c(66, 70, 76, 65, 74, 75, 74, 72)
waist <- c(79, 78, 80, 78, 79, 81, 84, 81)
chest <- c(83, 80, 86, 81, 84, 84, 91, 85)
taikei <- data.frame(height, weight, waist, chest)
colnames(taikei) <- c("身長", "体重", "ウエスト", "胸囲")
rownames(taikei) <- 1:8

```

29

相関行列

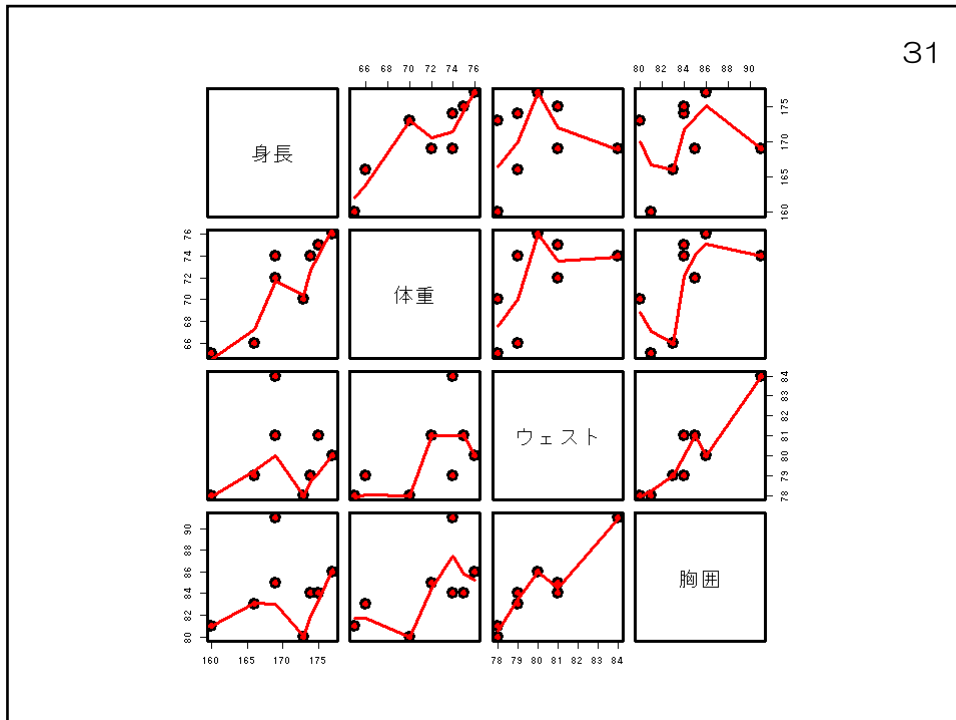
```
round(cor(taikei), 3)
```

	身長	体重	ウエスト	胸囲
身長	1.000	0.861	0.180	0.208
体重	0.861	1.000	0.569	0.594
ウエスト	0.180	0.569	1.000	0.933
胸囲	0.208	0.594	0.933	1.000

30

対散布図

```
pairs(taikei, pch=21, bg="red", cex=2,  
panel=panel.smooth)
```



32

重回帰分析

```
(taikei.lm <- lm(体重~., data=taikei))
```

Call:  
lm(formula = 体重 ~ ., data = taikei)

Coefficients:  
(Intercept)      身長      ウェスト      胸囲  
-85.4741      0.5765      0.4175      0.3009



## summary(taikei.lm)

33

```
Call:
lm(formula = 体重 ~ ., data = taikei)

Residuals:
    1     2     3     4     5     6     7     8
-2.1841 -0.8996  0.1540  1.2942  0.9029  0.4913 -0.4082  0.6495

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -85.4741    33.5930  -2.544  0.06368 .
身長          0.5765     0.1040   5.543  0.00518 **
ウエスト      0.4175     0.7826   0.533  0.62197
胸囲          0.3009     0.4671   0.644  0.55457
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
                 ' ' 1

Residual standard error: 1.494 on 4 degrees of freedom
Multiple R-squared:  0.9256,    Adjusted R-squared:  0.8699
F-statistic: 16.6 on 3 and 4 DF,  p-value: 0.01011
```

## Call:

34

```
lm(formula = 体重 ~ ., data = taikei)
```

```
Residuals: 個々のデータの残差
```

```
    1     2     3     4     5     6     7
-2.1841 -0.8996  0.1540  1.2942  0.9029  0.4913 -0.4082  0
```

```
Coefficients:
```

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -85.4741    33.5930  -2.544  0.06368 .
身長          0.5765     0.1040   5.543  0.00518 **
ウエスト      0.4175     0.7826   0.533  0.62197
胸囲          0.3009     0.4671   0.644  0.55457
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.'
                 ' ' 1
```

```
Residual standard error: 1.494 on 4 degrees of freedom
```

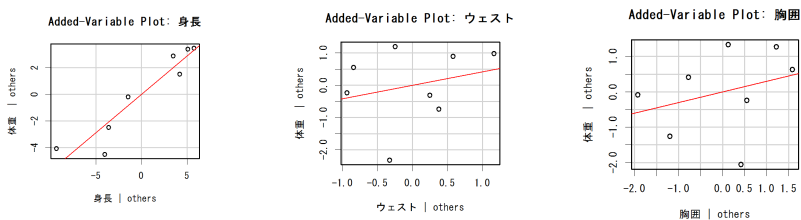
Coefficients: 偏回帰係数 標準誤差 *t* 値 *p* 値  
 Estimate Std. Error *t* value Pr(>|*t*|)  
 (Intercept) -85.4741 33.5930 -2.544 0.06368 .  
 身長 0.5765 0.1040 5.543 0.00518 \*\*  
 ウェスト 0.4175 0.7826 0.533 0.62197  
 胸囲 0.3009 0.4671 0.644 0.55457  
 ---  
 Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05  
 '.' 0.1 ' ' 1  
 Residual standard error: 1.494 on 4 degrees of freedom  
 Multiple R-squared: 0.9256, Adjusted R-squared:  
 0.8699  
 F-statistic: 16.6 on 3 and 4 DF, p-value: 0.01011

推定値の標準誤差 = 残差の標準偏差

モデル全体のF値 決定係数  $R^2$  調整済み決定係数  $R^2$

偏回帰プロット

```
install.packages("car")
library(car)
avPlot(taikei.lm, "身長")
avPlot(taikei.lm, "ウェスト")
avPlot(taikei.lm, "胸囲")
```



## 説明変数の選択

37

予測精度を低めずに、  
必要最小限の説明変数を選択する

$$AIC = n \log\left(\frac{SS_e}{n}\right) + 2(q+1)$$

AICが小さいほど良いモデル

```
(taikei2.lm <- step(taikei.lm))
```

38

```
Start: AIC=8.87
体重 ~ 身長 + ウエスト + 胸囲

      Df Sum of Sq  RSS   AIC
- ウエスト 1    0.635  9.559  7.4239
- 胸囲     1    0.926  9.849  7.6636
<none>                    8.924  8.8741
- 身長     1   68.540 77.464 24.1630

Step: AIC=7.42
体重 ~ 身長 + 胸囲

      Df Sum of Sq  RSS   AIC
<none>                    9.559  7.4239
- 胸囲  1   21.583 31.142 14.8729
- 身長  1   68.127 77.686 22.1858

Call:
lm(formula = 体重 ~ 身長 + 胸囲, data = taikei)

Coefficients:
(Intercept)      身長      胸囲
   -71.2316      0.5743      0.5327
```

39

Start: **AIC=8.87**

体重 ~ 身長 + ウェスト + 胸囲

	Df	Sum of Sq	RSS	AIC
- ウェスト	1	0.635	9.559	7.4239
- 胸囲	1	0.926	9.849	7.6636
<none>			8.924	8.8741
- 身長	1	68.540	77.464	24.1630

Step: **AIC=7.42**

体重 ~ 身長 + 胸囲

	Df	Sum of Sq	RSS	AIC
<none>			9.559	7.4239
- 胸囲	1	21.583	31.142	14.8729
- 身長	1	68.127	77.686	22.1858

40

Call:

lm(formula = 体重 ~ 身長 + 胸囲, data = taikei)

Coefficients:

(Intercept)	身長	胸囲
-71.2316	0.5743	0.5327

回帰式

$$\text{体重} = -71.2316 + 0.5743 \times \text{身長} + 0.5327 \times \text{胸囲}$$

## summary(taikei2.lm)

41

```
Call:
lm(formula = 体重 ~ 身長 + 胸囲, data = taikei)

Residuals:
    1     2     3     4     5     6     7     8 
-2.3214 -0.7437 -0.2372  1.1900  0.5512  0.9769 -0.3060  0.8902

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -71.23163   18.87440  -3.774  0.01297 *
身長          0.57433    0.09621   5.970  0.00189 **
胸囲          0.53270    0.15854   3.360  0.02011 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
                 ' ' 1

Residual standard error: 1.383 on 5 degrees of freedom
Multiple R-squared:  0.9203,    Adjusted R-squared:  0.8885 
F-statistic: 28.89 on 2 and 5 DF,  p-value: 0.001791
```

## 予測値、残差

42

```
残差 <- residuals(taikei2.lm)
```

```
予測 <- predict(taikei2.lm)
```

```
data.frame(taikei, 予測, 残差)
```

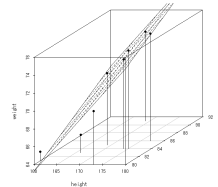
	身長	体重	ウエスト	胸囲	予測	残差
1	166	66	79	83	68.32142	-2.3214242
2	173	70	78	80	70.74366	-0.7436585
3	177	76	80	86	76.23717	-0.2371723
4	160	65	78	81	63.81004	1.1899647
5	174	74	79	84	73.44878	0.5512196
6	175	75	81	84	74.02311	0.9768872
7	169	74	84	91	74.30600	-0.3060005
8	169	72	81	85	71.10982	0.8901839

## 予測平面

43

```
install.packages("scatterplot3d")
library(scatterplot3d)

s3d <-
scatterplot3d(cbind(height, chest, weight),
type="h", pch=16)
s3d$plane3d(taikei2.lm, lty.box = "solid")
```



VIF : Variance Inflation factor (分散拡大係数)<sup>44</sup>

V I F が 4 ~ 10 程度以上は多重共線性が  
疑われる

トレランス (許容度) =  $1 / V I F$

多重共線性 (multicollinearity)

説明変数間に関係が強い場合

解が求まらないあるいは安定しない

45

VIF

```
install.packages("car")
```

```
library(car)
```

```
vif(taikei.lm)
```

	身長	ウェスト	胸囲
	1.046850	7.687803	7.774920

```
vif(taikei2.lm)
```

	身長	胸囲
	1.045224	1.045224

46

相関行列

```
round(cor(taikei), 3)
```

	身長	体重	ウェスト	胸囲
身長	1.000	0.861	0.180	0.208
体重	0.861	1.000	0.569	0.594
ウェスト	0.180	0.569	1.000	0.933
胸囲	0.208	0.594	0.933	1.000

47

## 標準回帰係数

測定値を標準得点化

```
staikei <- data.frame(scale(taikei))
staikei
```

	身長	体重	ウエスト	胸囲
1	-0.7878180	-1.3283762	-0.5	-0.37091592
2	0.4726908	-0.3622844	-1.0	-1.26111411
3	1.1929816	1.0868533	0.0	0.51928228
4	-1.8682541	-1.5698991	-1.0	-0.96438138
5	0.6527635	0.6038074	-0.5	-0.07418318
6	0.8328362	0.8453303	0.5	-0.07418318
7	-0.2475999	0.6038074	2.0	2.00294594
8	-0.2475999	0.1207615	0.5	0.22254955

48

```
staikei.lm <- lm(体重~., data=staikei)
staikei2.lm <- step(staikei.lm)
```

```
Start: AIC=-13.86
体重 ~ 身長 + ウエスト + 胸囲
```

	Df	Sum of Sq	RSS	AIC
- ウエスト	1	0.0370	0.5576	-15.3087
- 胸囲	1	0.0540	0.5745	-15.0691
<none>			0.5205	-13.8586
- 身長	1	3.9982	4.5187	1.4303

```
Step: AIC=-15.31
体重 ~ 身長 + 胸囲
```

	Df	Sum of Sq	RSS	AIC
<none>			0.5576	-15.3087
- 胸囲	1	1.2590	1.8166	-7.8598
- 身長	1	3.9741	4.5317	-0.5468



```
summary(staikei2.lm) 49

Call:
lm(formula = 体重 ~ 身長 + 胸囲, data = staikei)

Residuals:
    1     2     3     4     5     6
-0.56068 -0.17961 -0.05728  0.28740  0.13313  0.23594

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 2.222e-10  1.181e-01    0.00  1.00000
身長       7.703e-01  1.290e-01    5.97  0.00189 **
胸囲       4.336e-01  1.290e-01    3.36  0.02011 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3339 on 5 degrees of freedom
```

標準回帰係数

7.703e-01  
4.336e-01

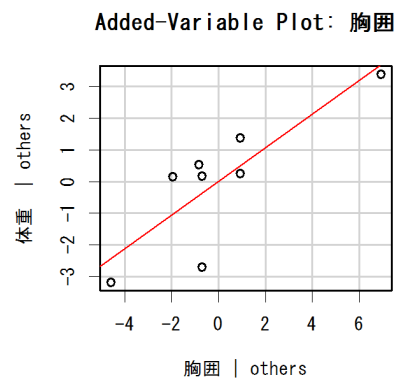
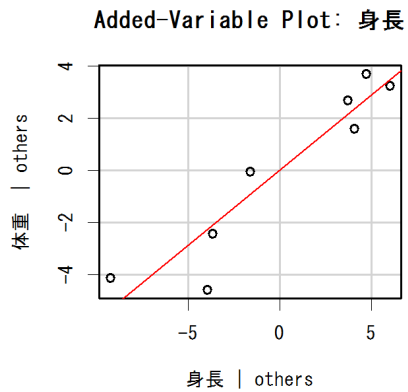
```
標準化残差 | 標準化残差 | <= 3 50

標準化残差 <- residuals(taikei2.lm)
標準化予測 <- predict(taikei2.lm)
data.frame(taikei, 標準化予測, 標準化残差)

  身長 体重 ウェスト 胸囲 標準化予測 標準化残差
1  166  66     79   83   68.32142 -2.3214242
2  173  70     78   80   70.74366 -0.7436585
3  177  76     80   86   76.23717 -0.2371723
4  160  65     78   81   63.81004  1.1899647
5  174  74     79   84   73.44878  0.5512196
6  175  75     81   84   74.02311  0.9768872
7  169  74     84   91   74.30600 -0.3060005
```

偏回帰プロット

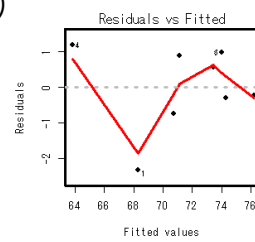
avPlot(taikei2.lm, "身長")  
avPlot(taikei2.lm, "胸囲")



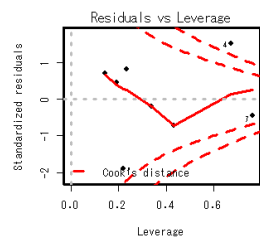
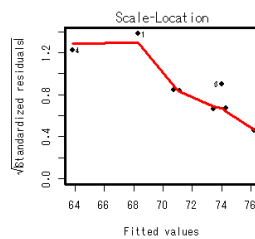
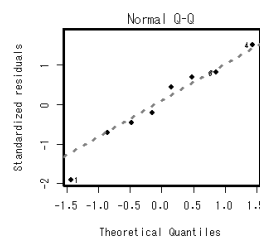
回帰診断図

par(mfrow=c(2, 2))  
plot(taikei2.lm)

残差と予測値



残差の正規Q-Q



残差の平方根

てこ比・標準化残差

55

## 青木先生のmregを使う方法

目的変数は最右列におく

```
source("../R/all.R", encoding="euc-jp")
taikei3 <- data.frame(height, waist, chest,
weight)
colnames(taikei3) <- c("身長", "ウエスト", "胸
囲", "体重")
rownames(taikei3) <- 1:8

(a <- mreg(taikei3))
```

56

	偏回帰係数	標準誤差	t 値	P 値	標準化偏回帰係数	トレイ
身長	0.57652	0.10401	5.54284	0.0051809	0.77326	0.
ウエスト	0.41752	0.78264	0.53348	0.6219695	0.20168	0.
胸囲	0.30087	0.46709	0.64413	0.5545698	0.24489	0.
定数項	-85.47413	33.59295	-2.54441	0.0636822		

## 回帰の分散分析表

	平方和	自由度	平均平方	F 値	P 値
回帰	111.0764	3	37.0255	16.597	0.010108
残差	8.9236	4	2.2309		
全体	120.0000	7	17.1429		

重相関係数 = 0.9621

重相関係数の二乗 = 0.92564

自由度調整済重相関係数の二乗 = 0.86986

対数尤度 = -11.78855

AIC = 33.57709

(sreg(taikei3))

57

```

有効ケース数: 8
従属変数: 体重
      平均値 不偏分散 標準偏差
身長  170.38 30.839 5.5533
ウエスト 80.00 4.000 2.0000
胸囲  84.25 11.357 3.3700
体重  71.50 17.143 4.1404

**** 相関係数行列 ****
      身長 ウエスト 胸囲 体重
身長  1.00000 0.18007 0.20801 0.86051
ウエスト 0.18007 1.00000 0.93259 0.56930
胸囲  0.20801 0.93259 1.00000 0.59382
体重  0.86051 0.56930 0.59382 1.00000

変数編入基準 Pin: 0.05
変数除去基準 Pout: 0.05
編入候補変数: 身長      P : 0.00609 **** 編入されました

**** ステップ 1 **** 編入変数: 身長

```

偏回帰係数 標準誤差 t値 P値 標準化偏回帰係数 トランス 分散拡大要因

途中省略

58

```

===== 結果 =====
      偏回帰係数 標準誤差 t値 P値 標準化偏回帰係数 トランス 分散拡大要因
身長  0.57433 0.096209 5.9697 0.00189      0.77032 0.95673 1.0452
胸囲  0.53270 0.158537 3.3601 0.02011      0.43358 0.95673 1.0452
定数項 -71.23163 18.874396 3.7740 0.01297

      平方和 自由度 平均平方 F値 P値
回帰 110.4415  2 55.2207 28.886 0.00179
残差  9.5585  5 1.9117
全体 120.0000  7

重相関係数 0.95935
決定係数(重相関係数の二乗) 0.92035
自由度調整済み重相関係数の二乗 0.88848
対数尤度 -12.06347
AIC 32.12695

```

59

## \$分析結果

	偏回帰係数	標準誤差	t値	P値	標準化偏回帰係数	トランス	分散拡大要因
身長	0.57433	0.096209	5.9697	0.00189	0.77032	0.95673	1.0452
胸囲	0.53270	0.158537	3.3601	0.02011	0.43358	0.95673	1.0452
定数項	-71.23163	18.874396	3.7740	0.01297			

## \$分散分析表

	平方和	自由度	平均平方	F値	P値
回帰	110.4415	2	55.2207	28.886	0.00179
残差	9.5585	5	1.9117		
全体	120.0000	7			

## \$決定係数

重相関係数 0.95935  
 決定係数(重相関係数の二乗) 0.92035  
 自由度調整済み重相関係数の二乗 0.88848  
 対数尤度 -12.06347  
 AIC 32.12695

60

## \$予測

	観察値	予測値	残差	標準化残差
1	66	68.32142	-2.3214242	-1.9011451
2	70	70.74366	-0.7436585	-0.7142030
3	76	76.23717	-0.2371723	-0.2110322
4	65	63.81004	1.1899647	1.5053339
5	74	73.44878	0.5512196	0.4436400
6	75	74.02311	0.9768872	0.8068656
7	74	74.30600	-0.3060005	-0.4555635
8	72	71.10982	0.8901839	0.6962727

61

### 重回帰分析をする前の注意点

- 説明変数は10個未満とする。
- データ数は説明変数の数より20以上多い。
- 説明変数の数 < データ数 / 3
- 各説明変数は目的変数と直線相関がある。
- 説明変数間の相関はあまり大きすぎない。

62

### 重回帰分析をした後の注意点

- 疑似相関  
相関係数大きい、標準偏回帰係数が小さい
- 多重共線性  
相関係数と標準偏回帰係数が異符号で大きい  
説明変数間の相関が高い場合が多い

例題)

63

```
data(attitude)
attitude
```

rating	会社に対する総合評価
complaints	従業員の苦情の取り扱い
privileges	特別な特権は許さない
learning	学習の機会
raises	能力に基づいた昇給
critical	加重
advance	昇進

ratingを他の変数で予測

```
attitude2 <- cbind(attitude[, 2:7], rating=attitude[, 1])
```

相関行列

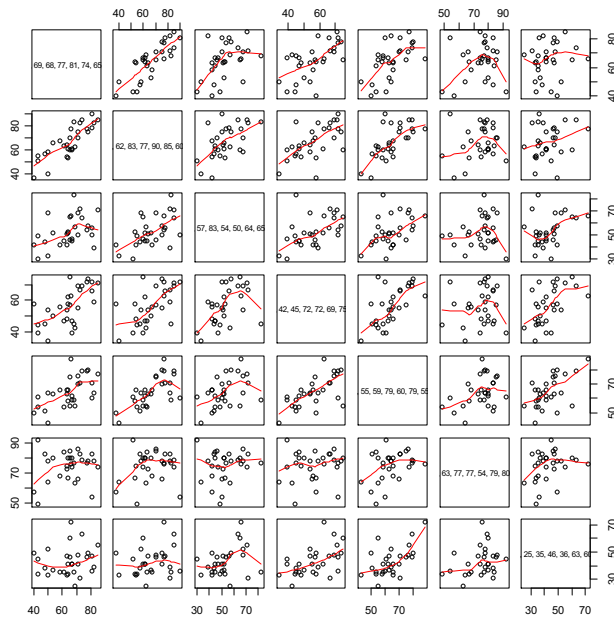
64

```
round(cor(attitude), 2)
```

	rating	complaints	privileges	learning	raises	critical	advance
rating	1.00	0.83	0.43	0.62	0.59	0.16	0.16
complaints	0.83	1.00	0.56	0.60	0.67	0.19	0.22
privileges	0.43	0.56	1.00	0.49	0.45	0.15	0.34
learning	0.62	0.60	0.49	1.00	0.64	0.12	0.53
raises	0.59	0.67	0.45	0.64	1.00	0.38	0.57
critical	0.16	0.19	0.15	0.12	0.38	1.00	0.28
advance	0.16	0.22	0.34	0.53	0.57	0.28	1.00

pairs (attitude, panel=panel. smooth, attitude)

65



```
attitude2 <-
cbind(attitude[, 2:7], rating=attitude[, 1])
(a <- mreg(attitude2))
```

66



67

	偏回帰係数	標準誤差	t 値	P 値	標準化偏回帰係数	トレランス
complaints	0.613188	0.16098	3.80902	0.00090287	0.670725	0.37494
privileges	-0.073050	0.13572	-0.53822	0.59559392	-0.073427	0.62465
learning	0.320332	0.16852	1.90085	0.06992535	0.308870	0.44033
raises	0.081732	0.22148	0.36903	0.71548009	0.069812	0.32486
critical	0.038381	0.14700	0.26111	0.79633426	0.031200	0.81426
advance	-0.217057	0.17821	-1.21799	0.23557705	-0.183464	0.51240
定数項	10.787076	11.58926	0.93078	0.36163372		

## 回帰の分散分析表

	平方和	自由度	平均平方	F 値	P 値
回帰	3148.0	6	524.661	10.502	1.2404e-05
残差	1149.0	23	49.957		
全体	4297.0	29	148.171		

重相関係数 = 0.85592

重相関係数の二乗 = 0.7326

自由度調整済重相関係数の二乗 = 0.66285

対数尤度 = -97.24991

AIC = 210.4998

68

## 回帰分析の要約情報

```
attitude.lm1 <- lm(rating ~ ., data = attitude)
summary(attitude.lm1)
```

69

```

Call:
lm(formula = rating ~ ., data = attitude)

Residuals:
    Min       1Q   Median       3Q      Max
-10.9418  -4.3555   0.3158   5.5425  11.5990

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  10.78708    11.58926   0.931 0.361634
complaints    0.61319     0.16098   3.809 0.000903 ***
privileges  -0.07305     0.13572  -0.538 0.595594
learning     0.32033     0.16852   1.901 0.069925 .
raises       0.08173     0.22148   0.369 0.715480
critical     0.03838     0.14700   0.261 0.796334
advance     -0.21706     0.17821  -1.218 0.235577

```

70

```

Residual standard error: 7.068 on 23 degrees of freedom
Multiple R-squared:  0.7326, Adjusted R-squared:  0.6628
F-statistic: 10.5 on 6 and 23 DF,  p-value: 1.240e-05
4.058e-05

```

71

AICによるモデルの変数選択

$$AIC = -2 \times (\text{モデルの最大対数尤度}) + 2 \times (\text{モデルのパラメータ数})$$

```
attitude.lm2 <- step(attitude.lm1)
```

72

```
Start: AIC=123.36
rating ~ complaints + privileges + learning + raises
+ critical + advance
```

	Df	Sum of Sq	RSS	AIC
- critical	1	3.41	1152.4	121.45
- raises	1	6.80	1155.8	121.54
- privileges	1	14.47	1163.5	121.74
- advance	1	74.11	1223.1	123.24
<none>			1149.0	123.36
- learning	1	180.50	1329.5	125.74
- complaints	1	724.80	1873.8	136.04

73

Step: AIC=121.45

rating ~ complaints + privileges + learning + raises  
+ advance

	Df	Sum of Sq	RSS	AIC
- raises	1	10.61	1163.0	119.73
- privileges	1	14.16	1166.6	119.82
- advance	1	71.27	1223.7	121.25
<none>			1152.4	121.45
- learning	1	177.74	1330.1	123.75
- complaints	1	724.70	1877.1	134.09

74

Step: AIC=119.73

rating ~ complaints + privileges + learning + advance

	Df	Sum of Sq	RSS	AIC
- privileges	1	16.10	1179.1	118.14
- advance	1	61.60	1224.6	119.28
<none>			1163.0	119.73
- learning	1	197.03	1360.0	122.42
- complaints	1	1165.94	2328.9	138.56

```
Step: AIC=118.14
rating ~ complaints + learning + advance
```

75

	Df	Sum of Sq	RSS	AIC
- advance	1	75.54	1254.7	118.00
<none>			1179.1	118.14
- learning	1	186.12	1365.2	120.54
- complaints	1	1259.91	2439.0	137.94

```
Step: AIC=118
rating ~ complaints + learning
```

	Df	Sum of Sq	RSS	AIC
<none>			1254.7	118.00
- learning	1	114.73	1369.4	118.63
- complaints	1	1370.91	2625.6	138.16

最終結果

76

```
summary(attitude.lm2)
```

77

```

Call:
lm(formula = rating ~ complaints + learning, data =
attitude)

Residuals:
    Min       1Q   Median       3Q      Max
-11.5568  -5.7331   0.6701   6.5341  10.3610

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   9.8709     7.0612   1.398   0.174
complaints    0.6435     0.1185   5.432 9.57e-06 ***
learning      0.2112     0.1344   1.571   0.128

```

78

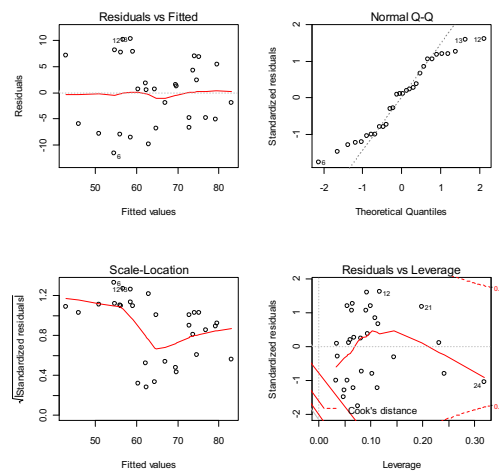
Residual standard error: 6.817 on 27 degrees of freedom  
Multiple R-squared: 0.708, Adjusted R-squared: 0.6864  
F-statistic: 32.74 on 2 and 27 DF, p-value: 6.058e-08

$$\text{rating} = 9.8709 + 0.6435 \cdot \text{complaints} + 0.2112 \cdot \text{learning}$$

## 回帰診断図

79

```
par(mfrow=c(2, 2))
plot(attitude.lm2)
```



## 例題)

80

airquality

1973年5月～9月のニューヨーク

Ozone	オゾンの量 (ppb)
Solar.R	太陽の放射の量 (lang)
Wind	風力 (mph)
Temp	温度 (華氏F)
Month	月 1-12
Day	月の日数 1-31

Ozoneを他の変数で予測