

線形回帰分析



線形回帰分析

この資料について

- この資料は、「JIN'S PAGE」の「Rと回帰分析」の箇所をまとめて、少し補ったものになっています。
- より細かく引数を設定することによって、様々な分析を行うことができます。
参考URLを見ながら試してみてください。
- 参考URL: JIN'S PAGE
<http://www1.doshisha.ac.jp/~mjcin/R/>

データ

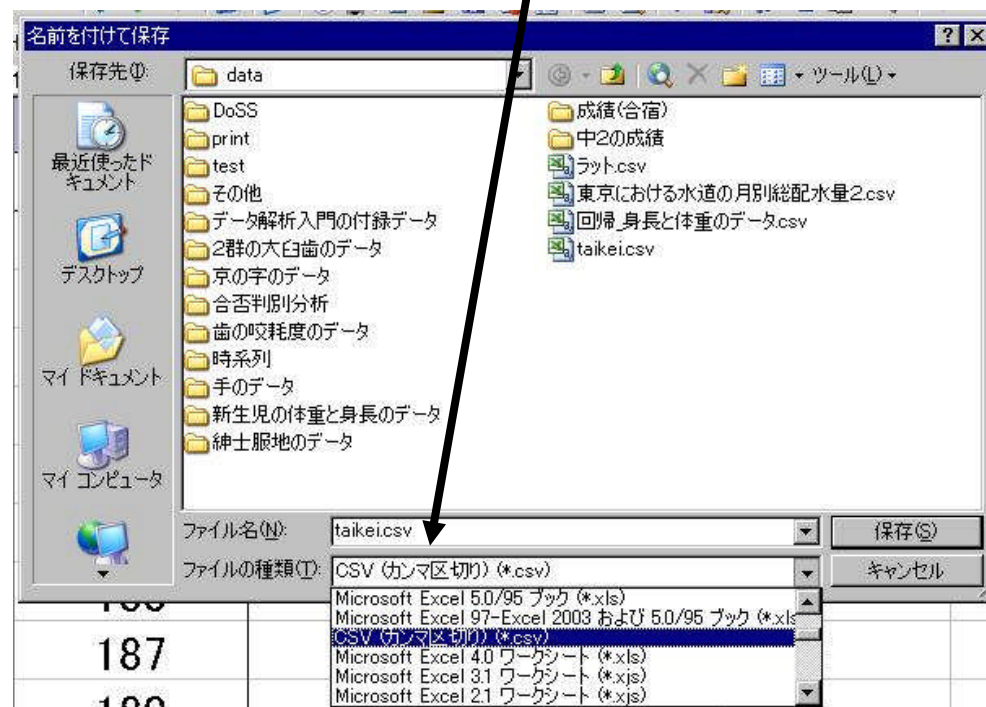
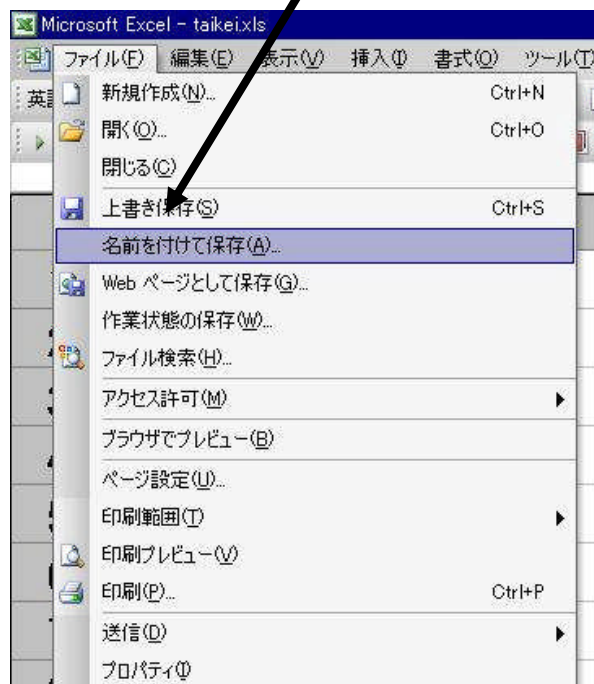
- 今回は「身長」と「体重」のデータを扱います。
- Excelで右のように入力し、
CSV形式で保存してください。
- 「R」では、データを扱う際、
行に個人を、列に変数を
対応させると良いと思います。

	A	B
1	体重	身長
2	49.1	165.4
3	49.9	162.0
4	54.3	168.6
5	54.3	163.5
6	63.4	172.5
7	71.9	169.3
8	73.8	174.6
9	75.0	175.4
10	76.0	173.5
11	77.4	183.7
12	86.3	183.4

CSV形式での保存1

① 名前を付けて保存

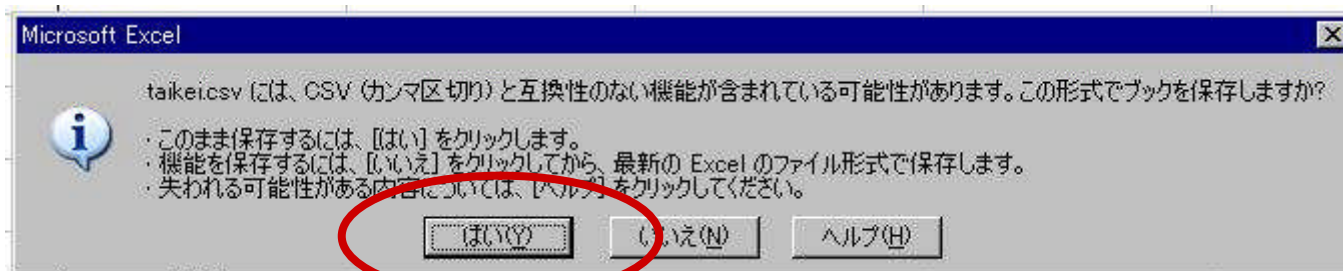
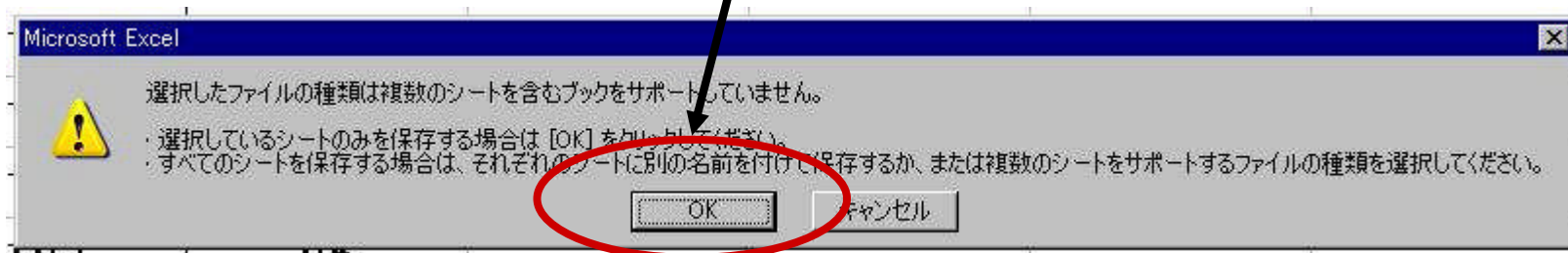
② 「ファイルの種類」⇒「CSV」



③ ファイル名を「taikei.csv」として保存

CSV形式での保存2

「OK」



「はい」

身長と体重のデータ・データ入力

- 「ファイル」⇒「ディレクトリの変更」の後、次のようにしてデータを読み込んで下さい。

```
R Console
> #データの読み込み
> taikei <- read.csv("taikei.csv", header=T)
> taikei
  体重 身長 ウェスト
1  49.1 165.4   59.4
2  49.9 162.0   62.3
3  54.3 168.6   64.4
4  54.3 163.5   67.8
5  63.4 172.5   70.0
6  71.9 169.3   85.1
7  73.8 174.6   79.5
8  75.0 175.4   78.3
9  76.0 173.5   84.3
10 77.4 183.7   76.0
11 86.3 183.4   84.4
> |
```

`read.csv("ファイル名.csv", header=T)`

線形回帰分析で用いる主な関数

```
taikei <- read.csv("taikei.csv", header=T)  
result <- lm(体重~身長,data=taikei)
```

```
summary(result)
```

```
予測値 <- predict(result)  
残差 <- residuals(result)  
data.frame(taikei,予測値,残差)
```

```
plot(taikei)  
abline(result)
```

線形回帰分析を行う関数: lm()

```
taikei <- read.csv("taikei.csv", header=T)
result <- lm(体重~身長,data=taikei)
```

- $\text{lm}(\text{目的変数} \sim \text{説明変数}, \text{データ})$

$$\text{目的変数} = \beta_0 + \beta_1 \times \text{説明変数} + \epsilon$$

定数項

- $\text{lm}(\text{体重} \sim \text{身長}, \text{data}=\text{taikei})$

$$\text{体重} = \beta_0 + \beta_1 \times \text{身長} + \epsilon$$

分析結果の要約を返す: summary()

summary(result)

- summary(obj) : 分析結果の要約を返す関数
 - “obj”はlm関数などで得られた分析結果
 - 分析で用いるモデル
 - 残差の情報
 - 回帰係数の推定量と、その有意性
 - 決定係数、調整済み決定係数等が出力されます。

関数”summary”の出力結果

```
R Console
> summary(result)

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

Residuals:
    Min     1Q  Median     3Q    Max
-7.591 -5.364  1.225  3.155  9.661

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -205.2511    46.1209  -4.450  0.00160 **
身長          1.5800     0.2679   5.897  0.00023 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.124 on 9 degrees of freedom
Multiple R-squared:  0.7944,    Adjusted R-squared:  0.7715
F-statistic: 34.77 on 1 and 9 DF,  p-value: 0.00023
```

出力結果の説明1

設定したモデル

```
R Console
> summary(result)
```

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

Residuals:
Min 1Q Median 3Q Max
-7.591 -5.364 1.225 3.100 7.591

回帰係数の推定量

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-205.2511	46.1209	-4.450	0.00160 **
身長	1.5800	0.2679	5.897	0.00011 **

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

決定係数 自由度調整済み決定係数

Residual standard error: 0.124 on 9 degrees of freedom
Multiple R-squared: 0.7944, Adjusted R-squared: 0.7715
F-statistic: 34.77 on 1 and 9 DF, p-value: 0.00023

出力結果の説明2

```
R Console
> summary(result)

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

Residuals:
    Min     1Q   Median     3Q     Max
-7.591 -5.364  1.225  3.155  9.661

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -295.2514    46.1209  -4.450  0.00160 **
身長          1.5800     0.2679   5.897  0.00023 ***
---
**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Multiple R-squared: 0.7044, Adjusted R-squared: 0.7715
F-statistic: 34.77 on 1 and 9 DF, p-value: 0.00023
```

残差の情報

決定係数の分布

回帰係数の標準偏差・有意性検定

Estimate	Std. Error	t value	Pr(> t)
(Intercept) -295.2514	46.1209	-4.450	0.00160 **
身長 1.5800	0.2679	5.897	0.00023 ***

予測値・残差 : predict() ・ residuals()

```
予測値 <- predict(result)
残差 <- residuals(result)
data.frame(taikei, 予測値, 残差)
```

- predict(obj) : 予測値を返す関数
- residuals(obj) : 残差を返す関数
- data.frame(A,B) : **データフレーム形式**で表示
- “data.frame(taikei, 予測値, 残差)”で
用いたデータと予測値・残差を並べて表示します。

予測値と残差

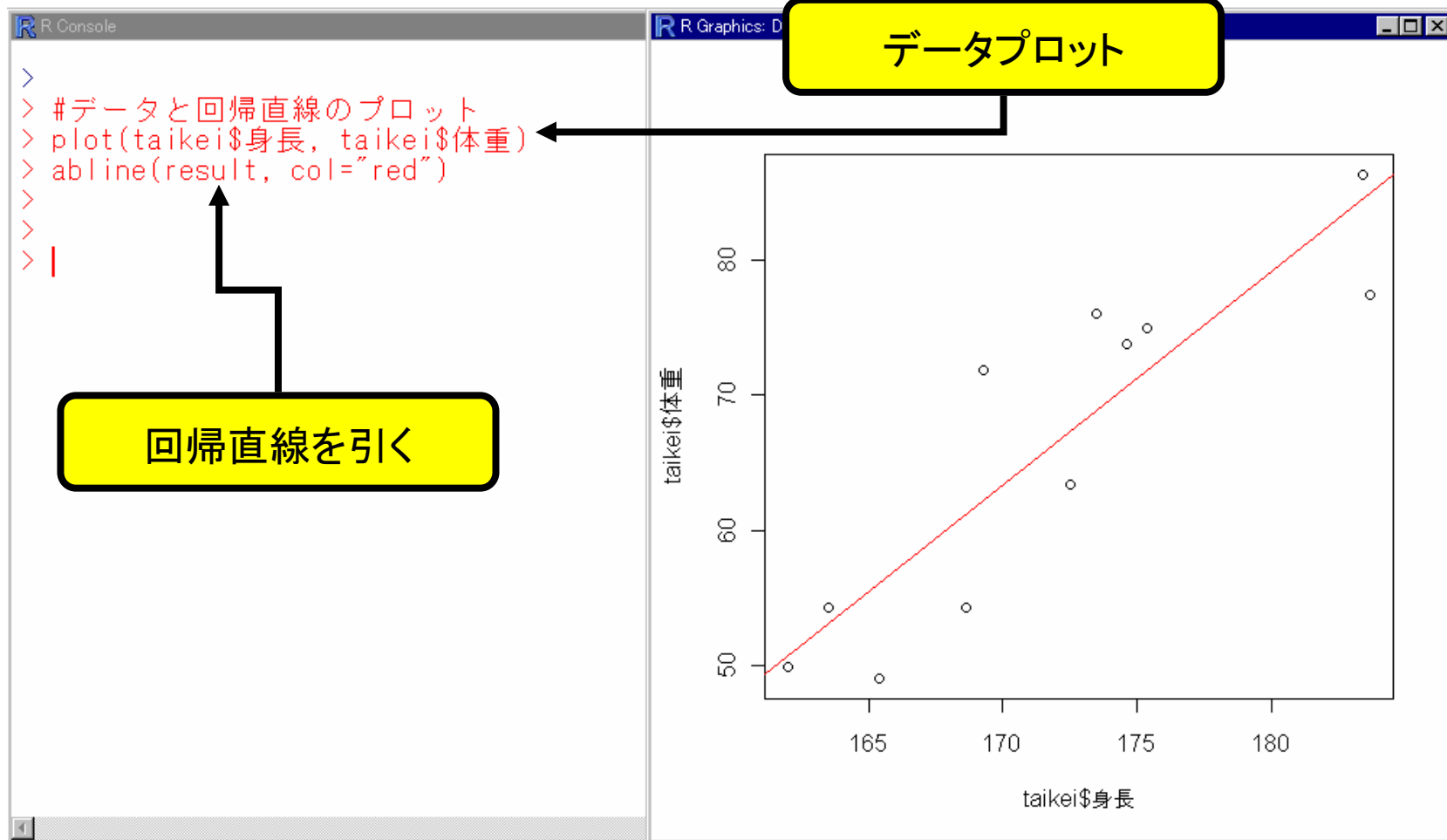
```
R Console
> #予測値を出力する関数：predict()
> 予測値 <- predict(result)
>
> #残差を出力する関数：residuals()
> 残差 <- residuals(result)
>
> #データと、当てはめによる予測値・残差
> data.frame(taikei, 予測値, 残差)
  体重 身長 予測値 残差
1  49.1 165.4 56.07741 -6.9774099
2  49.9 162.0 50.70548 -0.8054807
3  54.3 168.6 61.13334 -6.8333433
4  54.3 163.5 53.07545  1.2245505
5  63.4 172.5 67.29526 -3.8952621
6  71.9 169.3 62.23933  9.6606713
7  73.8 174.6 70.61322  3.1867816
8  75.0 175.4 71.87720  3.1227982
9  76.0 173.5 68.87524  7.1247587
10 77.4 183.7 84.99103 -7.5910290
11 86.3 183.4 84.51704  1.7829648
```

データと回帰直線のプロット

```
plot(taikei$身長, taikei$体重)  
abline(result)
```

- plot() : データの散布図を描く
- abline() : 本来は、2点を結ぶ直線を描く関数
lm関数で得られたオブジェクトに
適用すると、回帰直線を描く

データと回帰直線のプロット



重回帰分析へ...

- 身長のみを説明変数として分析を行ったが、**体型に関する他の情報をモデルに組み込むこと**で、より良い分析ができると考えられる。
- そこで、新たに「**ウエスト**」を説明変数に加えて解析を行う。
- 右のデータを入力し、ファイル名を「**taikei2.csv**」として保存して下さい。

	A	B	C
1	体重	身長	ウエスト
2	49.1	165.4	59.4
3	49.9	162.0	62.3
4	54.3	168.6	64.4
5	54.3	163.5	67.8
6	63.4	172.5	70.0
7	71.9	169.3	85.1
8	73.8	174.6	79.5
9	75.0	175.4	78.3
10	76.0	173.5	84.3
11	77.4	183.7	76.0
12	86.3	183.4	84.4

説明変数を加えた分析

```
taikei2 <- read.csv("taikei2.csv", header=T)
result2 <- lm(体重~身長+ウエスト,data=taikei2)

summary(result2)

予測値2 <- predict(result2)
残差2 <- residuals(result2)
data.frame(taikei2,予測値2,残差2)
```

説明変数を加えたモデル

```
taikei2 <- read.csv("taikei2.csv", header=T)
result2 <- lm(体重~身長+ウエスト,data=taikei2)
```

- $\text{lm}(\text{目的変数} \sim \text{説明変数1} + \text{説明変数2}, \text{データ})$

$$\text{目的変数} = \beta_0 + \beta_1 \times \text{説明変数1} + \beta_2 \times \text{説明変数2} + \varepsilon$$

- $\text{lm}(\text{体重} \sim \text{身長} + \text{ウエスト}, \text{データ})$

$$\text{体重} = \beta_0 + \beta_1 \times \text{身長} + \beta_2 \times \text{ウエスト} + \varepsilon$$

2つのモデルの比較

- 残差平方和・決定係数による比較
 - 「ウエスト」を変数に取り込んだモデルは、どの指標においても体重のみのモデルより良い結果を示している。

説明変数	残差平方和	決定係数	調整済決定係数
体重のみ	337.5	0.7944	0.7715
体重とウエスト	13.5	0.9918	0.9897

追記

□ 適切なモデルの構成

- 残差分析：早川 毅著「回帰分析の基礎」
- 交互作用の検討
- 変数選択

□ 参考URL

- R-Tips：第71節「回帰分析と重回帰分析」
- JIN'S PAGE：「Rと回帰分析」, 「Rと重回帰分析」
<http://www1.doshisha.ac.jp/~mjin/R/>

参考URL

- R-Tips

<http://cse.naro.affrc.go.jp/takezawa/r-tips/r2.html>

- JIN'S PAGE

<http://www1.doshisha.ac.jp/~mjn/R/>