

線形回帰分析

回帰分析

= 説明変数を用いて、目的変数を説明する統計モデル

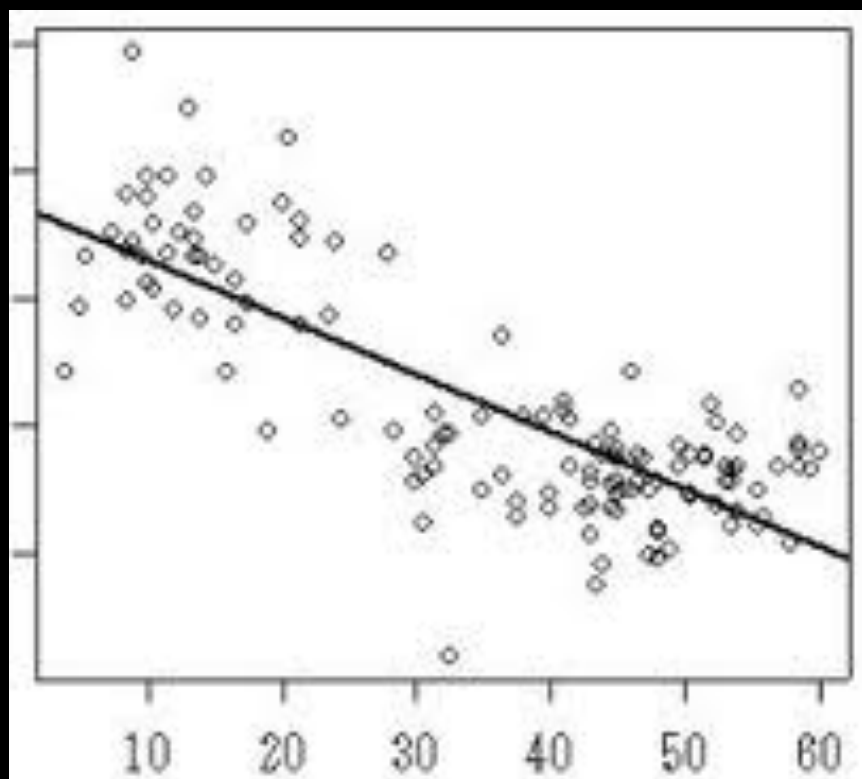
- 単回帰分析

= 説明変数が一つ

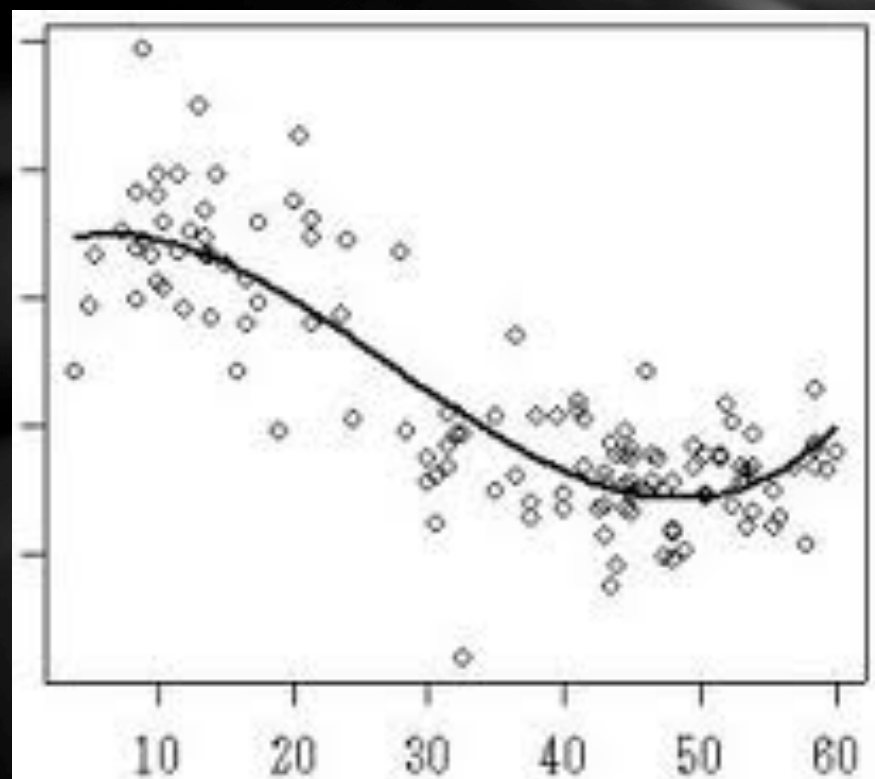
- 重回帰分析

= 説明変数が二つ

線形回帰分析と非線形回帰分析



▲線形回帰分析



▲非線形回帰分析

関数lm

`lm(formula, data, weights, subset, na.action)`

回帰分析の関連関数

コマンド	内 容	使 用 例
print	要約より簡単な結果	<code>print(taikei.lm)</code>
summary	回帰分析結果の要約	<code>summary(taikei.lm)</code>
coef	回帰係数	<code>cofe(taikei.lm)</code> 、 <code>taikei.lm\$coef</code>
fitted	用いたデータの予測値	<code>fitted(taikei.lm)</code> 、 <code>taikei.lm\$fitted</code>
deviance	残差の平方和	<code>deviance(taikei.lm)</code>
anova	回帰係数の分散分析	<code>anova(taikei.lm)</code>
predict	新たなデータに対する 予測値	<code>predict(taikei.lm)</code>
plot	回帰診断プロット	<code>plot(taikei.lm)</code>

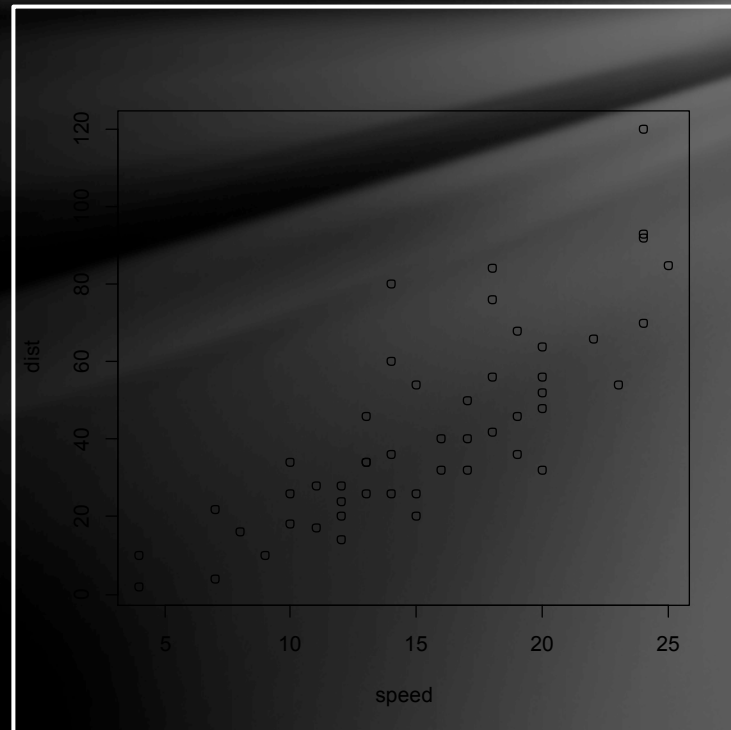
ケーススタディー

用いるデータ

(データセット **cars**)

```
> plot(cars);cor(cars$speed,cars$dist)  
[1] 0.8068949
```

	speed	dist
1	4	2
2	4	10
<中略>		
49	24	120
50	25	85



相関関係、線形関係がある

散布図

解析

```
> cars.lm<-lm(dist~speed,data=cars)
> cars.lm<-lm(dist~speed,data=cars)
> summary(cars.lm)
```

解析の結果

Call:

```
lm(formula = dist ~ speed, data = cars)
```

Residuals:

Min	1Q	Median	3Q	Max
-29.069	-9.525	-2.272	9.215	43.201

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-17.5791	6.7584	-2.601	0.0123 *
speed	3.9324	0.4155	9.464	1.49e-12 ***

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

Residual standard error: 15.38 on 48 degrees of freedom

Multiple R-squared: 0.6511, Adjusted R-squared: 0.6438

F-statistic: 89.57 on 1 and 48 DF, p-value: 1.49e-12

残差

実データ

$$y_i = a + b x_i + e_i$$

回帰直線 (推測モデル)

$$y'_i = a + b x_i$$

両値の差

$$e_i = y_i - y'_i$$

```
> residuals(cars.lm)
```



```
1      2      3      4      5      6      7      8      9
3.849460 11.849460 -5.947766 12.052234 2.119825 -7.812584 -3.744993 4.255007 12.255007
10     11     12     13     14     15     16     17     18
-8.677401 2.322599 -15.609810 -9.609810 -5.609810 -1.609810 -7.542219 0.457781 0.457781
19     20     21     22     23     24     25     26     27
12.457781 -11.474628 -1.474628 22.525372 42.525372 -21.407036 -15.407036 12.592964
-13.339445
28     29     30     31     32     33     34     35     36
-5.339445 -17.271854 -9.271854 0.728146 -11.204263 2.795737 22.795737 30.795737 -21.136672
37     38     39     40     41     42     43     44     45
-11.136672 10.863328 -29.069080 -13.069080 -9.069080 -5.069080 2.930920 -2.933898 -18.866307
46     47     48     49     50
-6.798715 15.201285 16.201285 43.201285 4.268876
```

回帰式と回帰係数

回帰直線の係数 a と b は残差の二乗の和を最小化する連立方程式の解

```
> round(coefficients(cars.lm),2)
```

(Intercept)	speed
-17.58	3.93

$$\underline{\text{dist} = -17.58 + 3.93 \times \text{speed}}$$

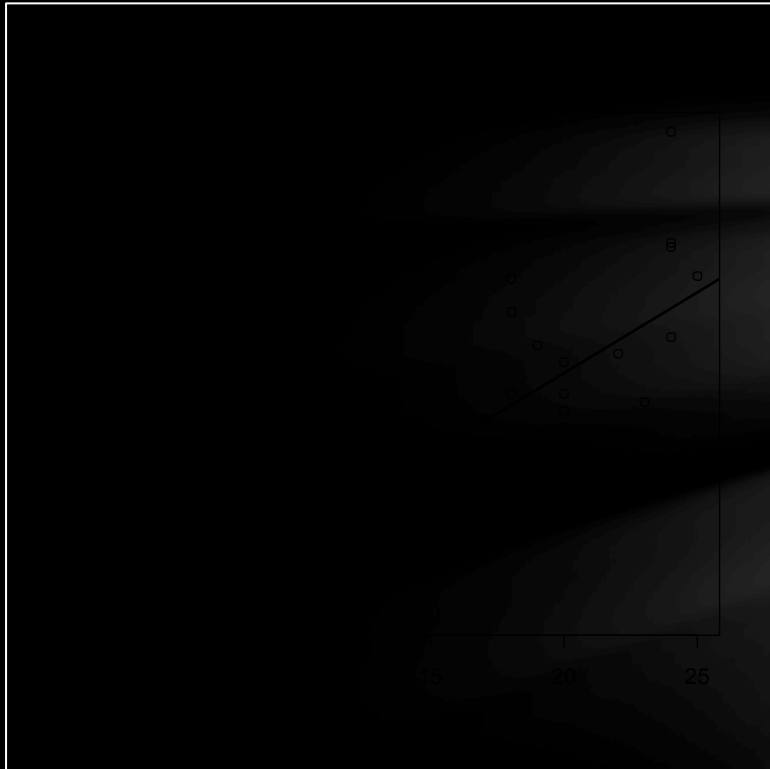
予測値

```
> 予測値<-predict(cars.lm)
> 残差<-residuals(cars.lm)
> data.frame(cars,予測値,残差)
```

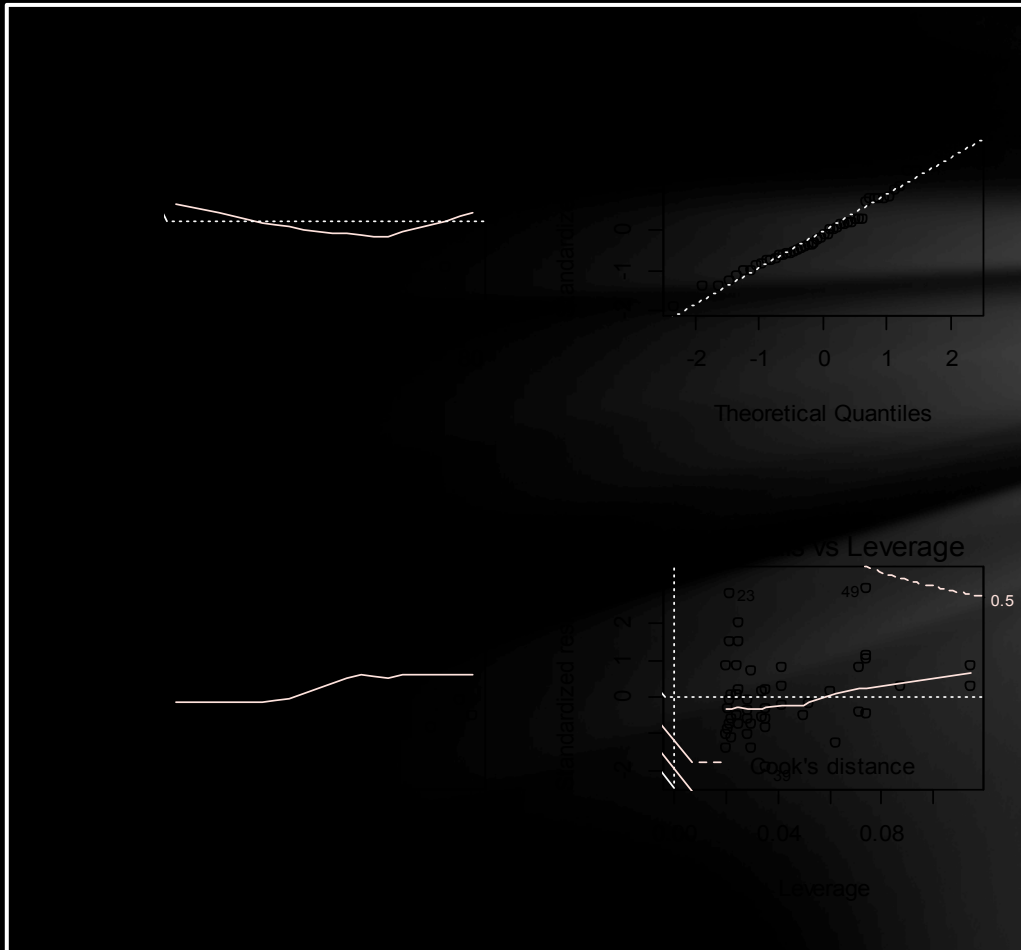


speed	dist	予測値	残差	
1	4	2	-1.849460	3.849460
2	4	10	-1.849460	11.849460
<中略>				
49	24	120	76.798715	43.201285
50	25	85	80.731124	4.268876

グラフによる分析



回帰診断図



線形重回帰分析

```
> pairs(airquality[,1:4])  
> air.lm<-lm(Ozone~Solar.R+Wind  
+Temp,data=airquality)  
> summary(air.lm)
```



結果省略

解析と結果

```
> round(coefficients(air.lm),2)
```



(Intercept)	Solar.R	Wind	Temp
-64.34	0.06	-3.33	1.65

回帰式： $-64.34 + 0.06 \times \text{Solar.R} - 3.33 \times \text{Wind} + 1.65 \times \text{Temp}$

相互作用モデル

```
> air.lm2<-lm(Ozone~(Solar.R+Wind+Temp)^2,data=airquality)
> summary(air.lm2)
```



```
Call:
lm(formula = Ozone ~ (Solar.R + Wind + Temp)^2, data = airquality)
<中略>
Multiple R-squared: 0.6863, Adjusted R-squared: 0.6682
F-statistic: 37.93 on 6 and 104 DF, p-value: < 2.2e-16
```

変数・モデルの選択

AICを用いる

```
> air.lm3<-step(air.lm2)
```



```
Start: AIC=662.37
```

```
Ozone ~ (Solar.R + Wind + Temp)^2
```

	Df	Sum of Sq	RSS	AIC
- Solar.R:Wind	1	429.42	38635	661.61
<none>		38205	662.37	
- Solar.R:Temp	1	1574.75	39780	664.86
- Wind:Temp	1	2748.20	40954	668.08

```
Step: AIC=661.61
```

```
Ozone ~ Solar.R + Wind + Temp + Solar.R:Temp + Wind:Temp
```

	Df	Sum of Sq	RSS	AIC
<none>		38635	661.61	
- Solar.R:Temp	1	2141.1	40776	665.60
- Wind:Temp	1	4339.8	42975	671.43

```
> round(coefficients(air.lm3),2)
```



(Intercept)	Solar.R	Wind	Temp	Solar.R:Temp
-136.81	-0.35	11.15	2.45	0.01
Wind:Temp				
-0.19				


$$\text{Ozone} = -136.81 - 0.35 \times \text{Solar.R} + 11.15 \times \text{Wind} + 2.45 \times \text{Temp} + 0.01 \times \text{Solar.R} \times \text{Temp} - 0.19 \times \text{Wind} \times \text{Temp}$$