

回帰直線

Code ▼

regression line

回帰直線

- data の裏にある隠された法則を見つける
 - 二次元の data に当てはめ → 線形回帰直線 → 予測 data の取得

営業所	広告費 (万円)	売上 (万円)
A	12.5	141
B	20.2	188
C	11.1	111
D	18.2	150
E	19.9	156
F	14.3	154

$y = bx + a$

↓

手持ちのdataで 傾き・切片 を求める

↓

予測 data 取得

最小二乗法

- 誤差が二乗の総和 が最小になる様な直線を見つける

値	公式
実際の値	y_i
直線上の値	$\hat{y}_i = bx_i + a$
誤差	$e_i = y_i - (bx_i + a)$
誤差の二乗の総和	$\sum e_i^2$

公式

$$b = \frac{\text{共分散}}{x\text{の分散}}$$
$$a = y\text{の}\mu - b \times x\text{の}\mu$$

回帰直線の特徴

- 決定係数 r^2 : 求めた回帰直線の 当てはまりの良さ を表す値
 - 相関係数 r の二乗 $0 \leq r \leq 1 \rightarrow 1$ に近いほど当てはまりが良い

y の変動数

||

回帰直線 + 残差



回帰直線の説明力

因果関係

- 回帰分析でも因果関係の有無に注意
 - 原因
 - 説明変数, 独立変数
 - 結果
 - 被説明変数, 従属変数

原因(x) = 結果(y)

結果(y) \neq 原因(x)

cars data

- 車の speed(速さ): x軸 と dist(制動距離): y軸 の関係
 - 相関係数 : 0.8068949

Hide

cars

speed	dist
<dbl>	<dbl>
4	2
4	10
7	4
7	22
8	16
9	10
10	18
10	26
10	34
11	17

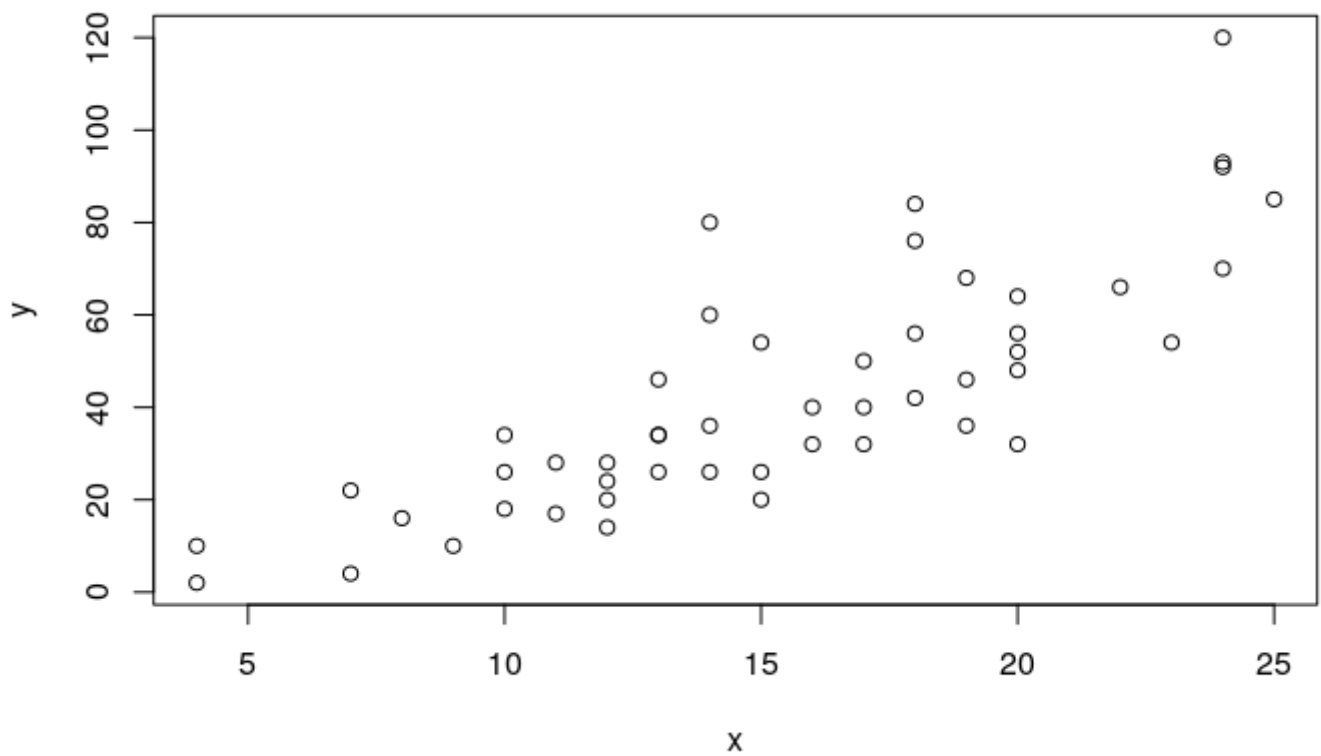
1-10 of 50 rows

Previous12345Next

Hide

```
x <- cars$speed
y <- cars$dist

plot(x, y)
```



Hide

```
cor(x, y)
```

```
[1] 0.8068949
```

線形回帰直線を求めて plot

- 誤差が最小になる値の直線を引く
 - 傾き : 3.9324088
 - 切片 : -17.5790949

Hide

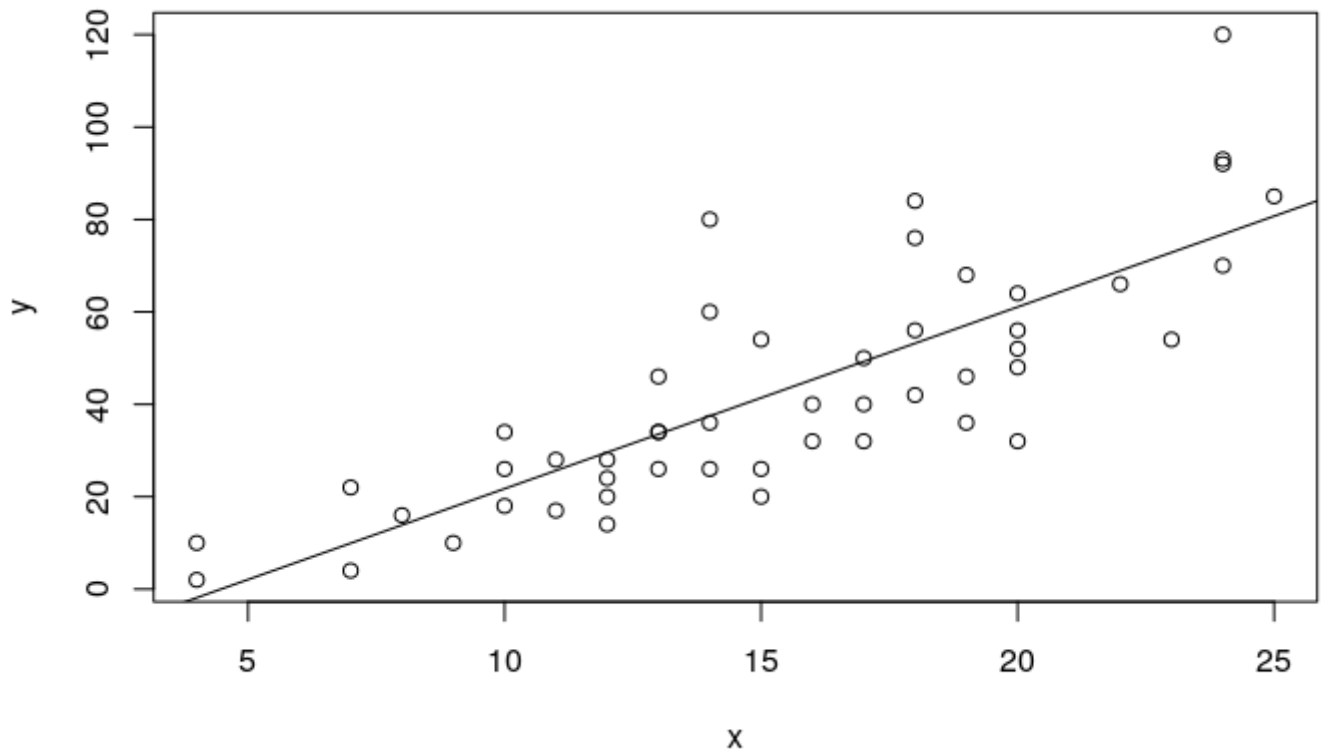
```
b <- cov(x, y) / var(x)
a <- mean(y) - b*mean(x)
c(b, a)
```

```
[1]  3.932409 -17.579095
```

Hide

```
plot(x, y)
abline(a, b, colors = "red")
```

```
"colors" is not a graphical parameter
```



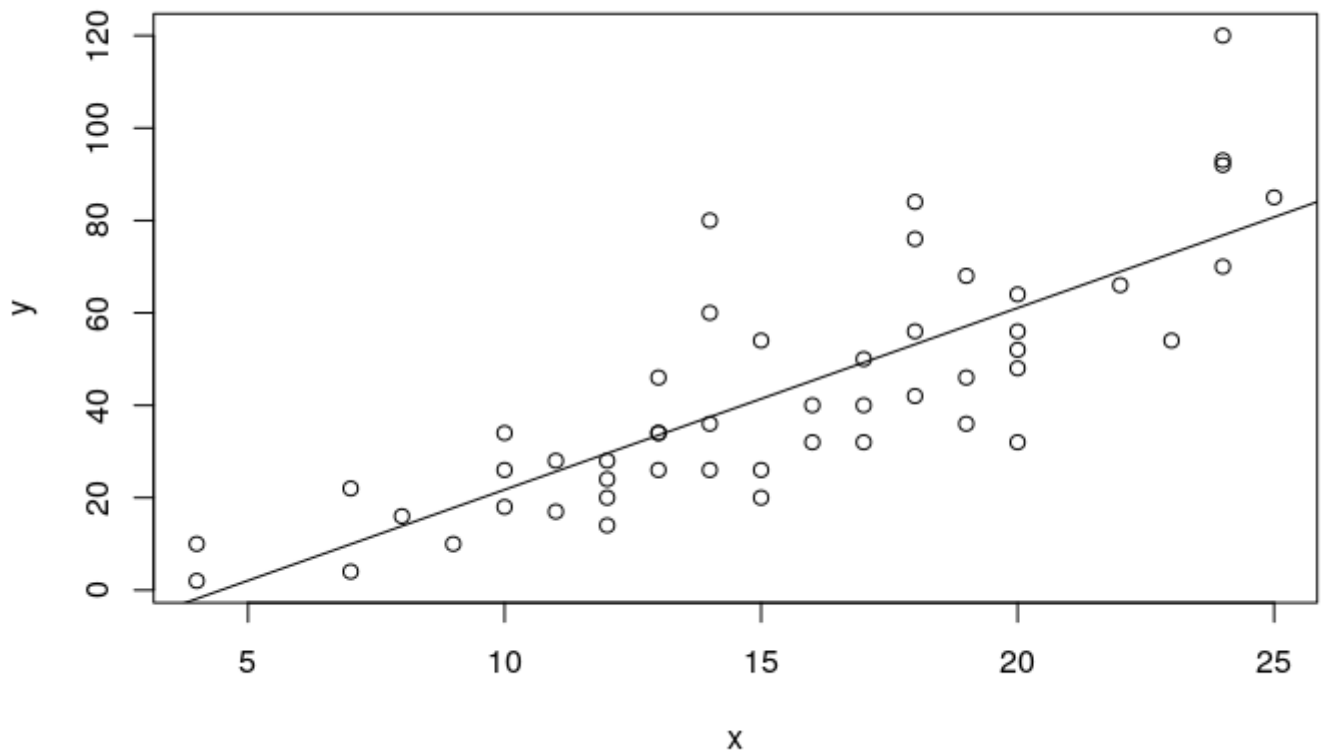
線形回帰直線

- R言語の関数

```
lm(目的変数 ~ 説明変数)
```

Hide

```
res <- lm(y ~ x)
plot(x, y)
abline(res)
```



重回帰分析

- 説明変数が(原因)が2つ以上の回帰分析
 - 変数が1つの時は **直線** → 2つの時は **平面** を当てはめる

重回帰分析

||

1つの結果に対して...

↓

複数の原因がある

$$y = a + b_1x_1 + b_2x_2$$

- a, b_1, b_2 を求めると式が決定

公式

重回帰分析 paramerter

$$\sum e_i^2 = \{y_i - (a + b_1x_1 + b_2x_2)\}^2$$

- 最小になる a, b1, b2 を求めるのだが... -> 難しいので pc に任せよう！

標準回帰係数

- あらかじめ説明変数を **標準化** することで比較が可能になる
 - **scale** を合わせる事により, category が違うものを比較可能に出来る
 - 広告費, 人員数, 販売個数 etc...

$$\text{標準化} = \frac{X - \mu}{\sigma}$$

- 各値 : X
- 平均 : μ
- 標準偏差 : σ

重回帰分析を計算

- R で計算

```
lm(目的変数 ~ 説明変数1, 説明変数2,...)
```

Hide

```
y1 <- c(82, 97, 130, 108, 144, 165)
x1 <- c(11, 10, 14, 9, 15, 20)
x2 <- c(3, 4, 5, 7, 6, 7)

res1 <- lm(y1~x1+x2)
res1
```

```
Call:
lm(formula = y1 ~ x1 + x2)

Coefficients:
(Intercept)      x1      x2
    6.876    5.387    8.100
```

標準化して重回帰分析

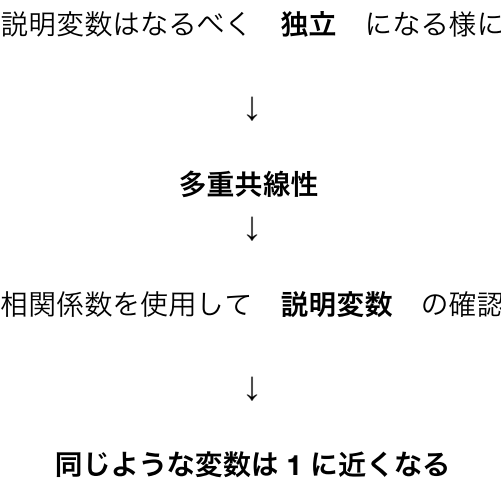
```
ys <- scale(y1)
xs1 <- scale(x1)
xs2 <- scale(x2)

res2 <- lm(ys~xs1+xs2)
res2
```

Call:
lm(formula = ys ~ xs1 + xs2)

Coefficients:
(Intercept) xs1 xs2
-1.734e-16 7.068e-01 4.263e-01

重回帰分析の注意点



説明変数は多いほど良いのか？

