

統計演習 (13)

回帰分析

高谷 智英

Jan 8, 2020





回帰分析 (regression analysis), p211

回帰分析：一方の変数の値から、もう一方の値を予測する。

予測に用いる変数：説明変数 (または独立変数)

予測される(したい)変数：目的変数 (または従属変数)

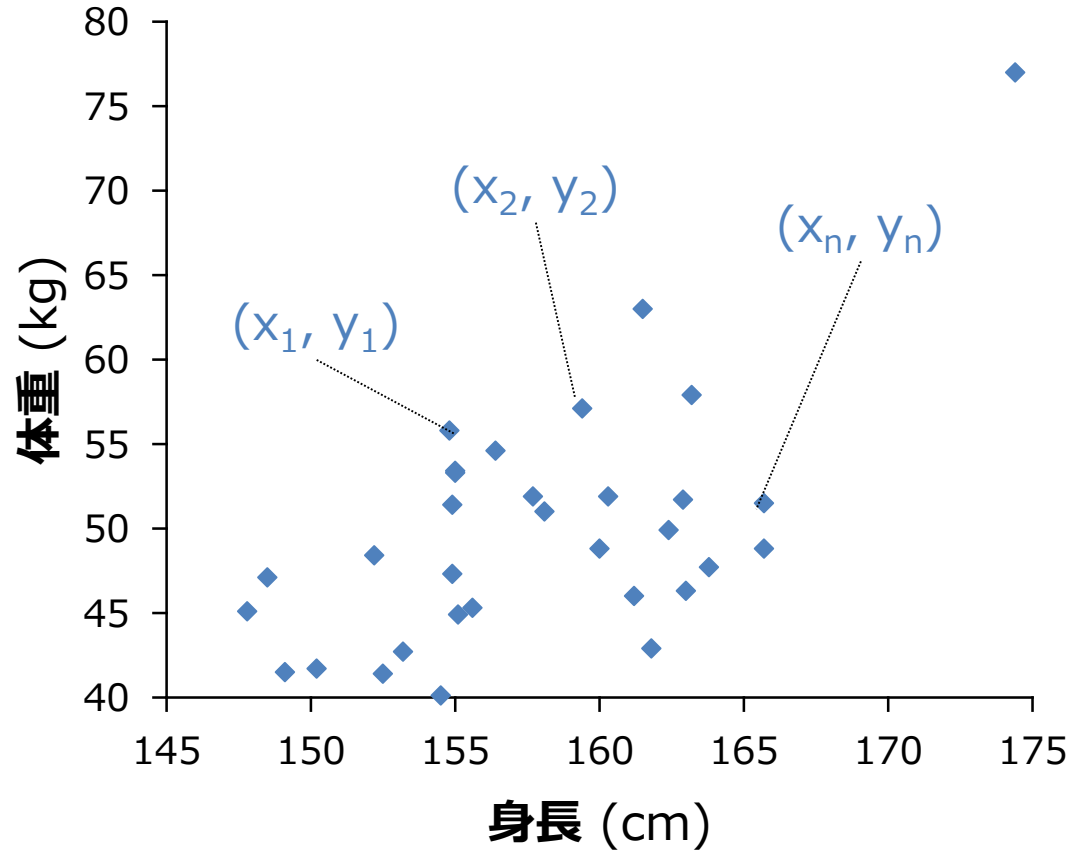
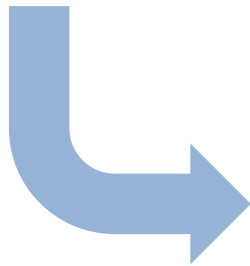
説明変数が1つの場合：単回帰分析

説明変数が2つ以上の場合：重回帰分析



単回帰分析 (説明変数が1つ), p212

| 説明変数 x | 目的変数 y |
|----------|----------|
| x_1 | y_1 |
| x_2 | y_2 |
| \vdots | \vdots |
| x_n | y_n |





回帰直線を引く, p212

最小二乗法:

各点 (x_i, y_i) から回帰直線までの垂直距離の2乗和が最小になる

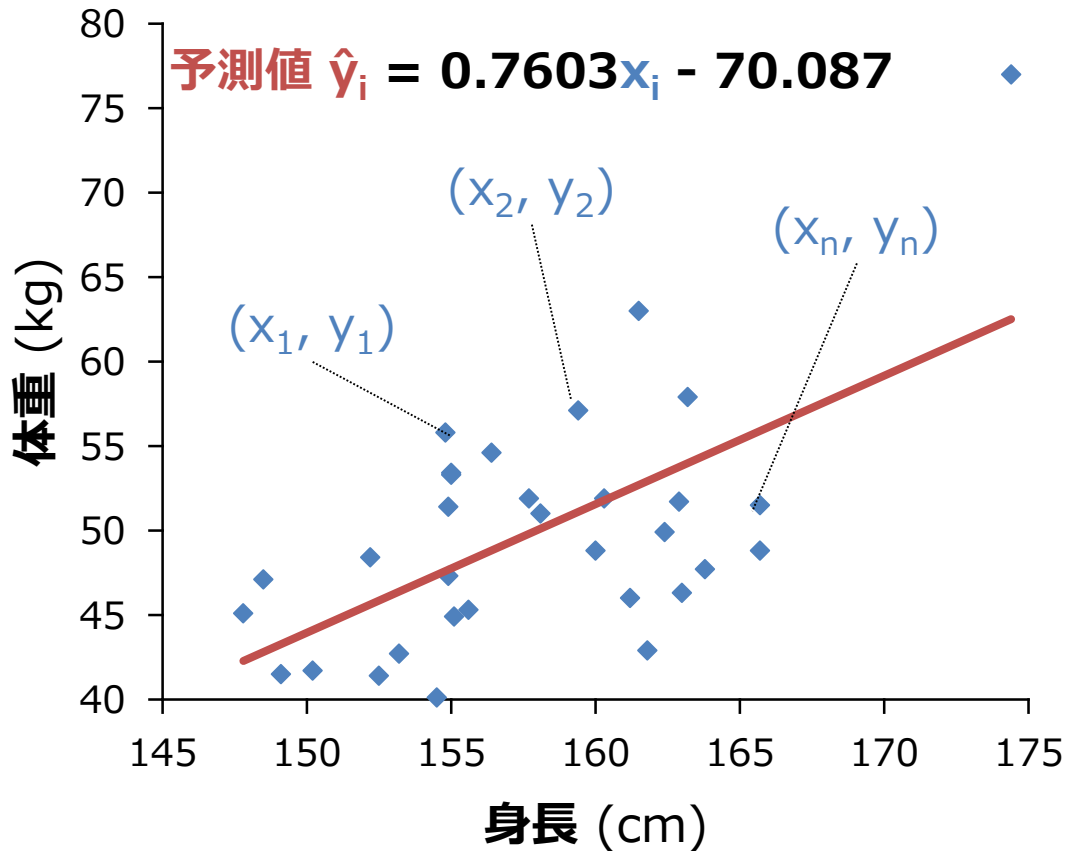
回帰直線の式を求める手法

求められる一次式:

$$\text{予測値 } \hat{y}_i = b_0 + b_1 x_i$$

b_0 : 定数項

b_1 : 回帰係数





決定係数 (寄与率) R^2 , p213

予測値 $\hat{y}_i = b_0 + b_1x_i$

と、

実測値 y_i

の差(残差)が小さいほど良い回帰である。

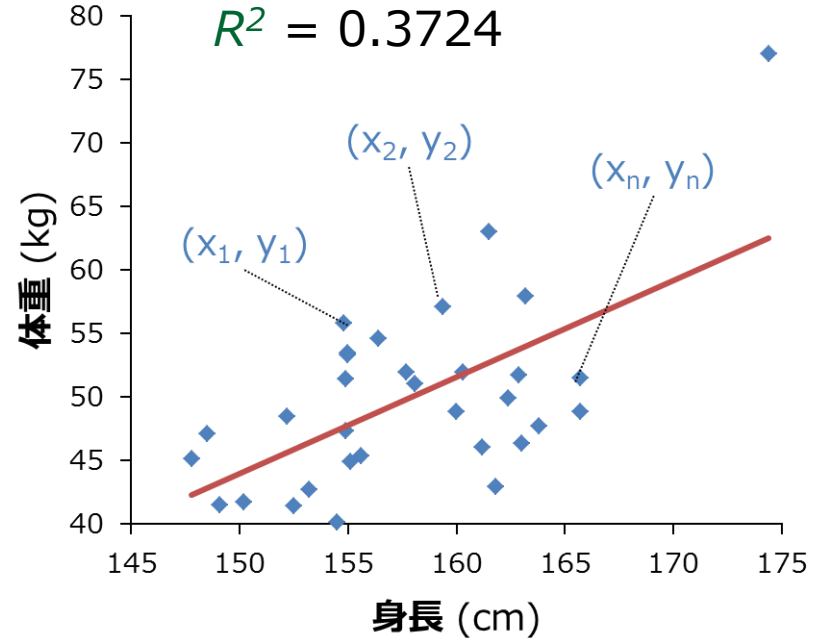
平均値 \bar{y}

とすると、

$$\frac{\sum(y_i - \bar{y})^2}{\text{全変動}} = \frac{\sum(\hat{y}_i - \bar{y})^2}{\text{回帰変動}} + \frac{\sum(y_i - \hat{y}_i)^2}{\text{残差変動}}$$

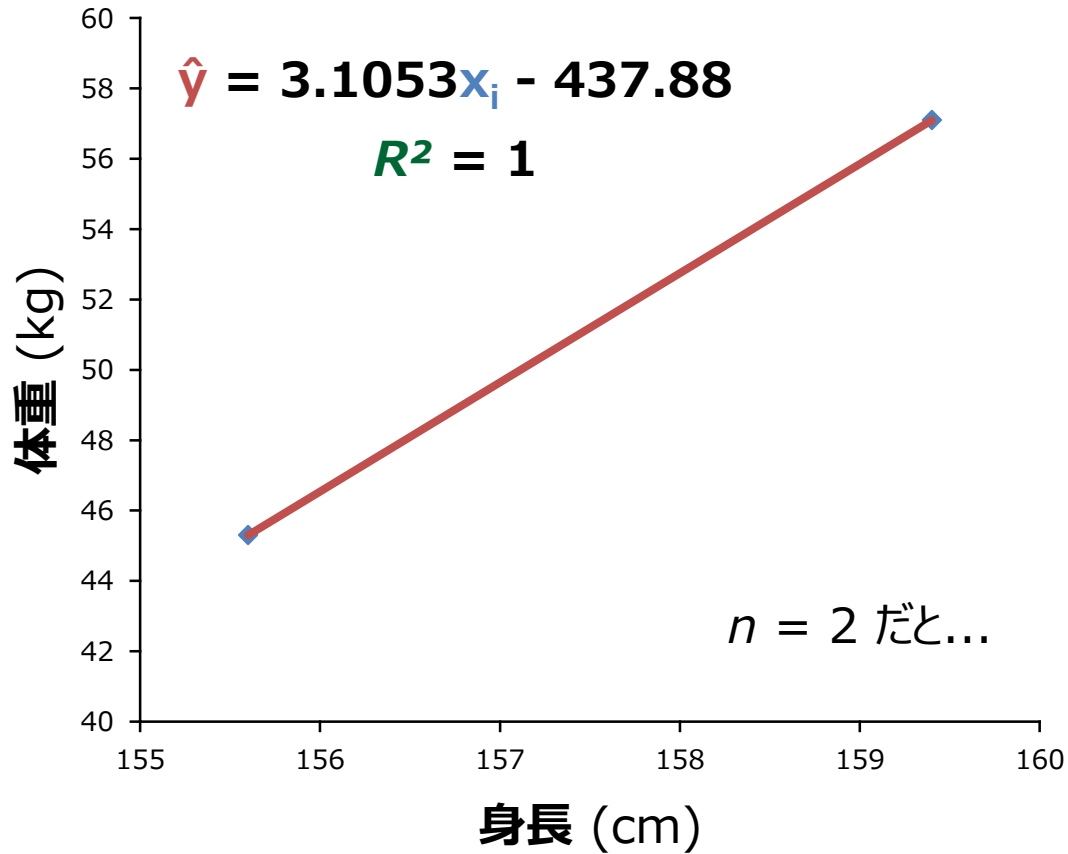
決定係数 (寄与率) $R^2 = 1 - (\text{残差変動}/\text{全変動})$

R^2 の値が大きいほど、回帰直線とデータの相関が高い。





R^2 が高けりゃ良いってもんじゃない, p214



分散分析によって、回帰式の有意性も検定する, p301



重回帰分析 (説明変数が2つ以上), p219

| 説明変数 | | | | 目的変数 |
|----------|----------|-----|----------|-------|
| x_1 | x_2 | ... | x_q | y |
| x_{11} | x_{21} | ... | x_{q1} | y_1 |
| x_{12} | x_{22} | ... | x_{q2} | y_2 |
| : | : | | : | : |
| x_{1n} | x_{2n} | ... | x_{qn} | y_n |

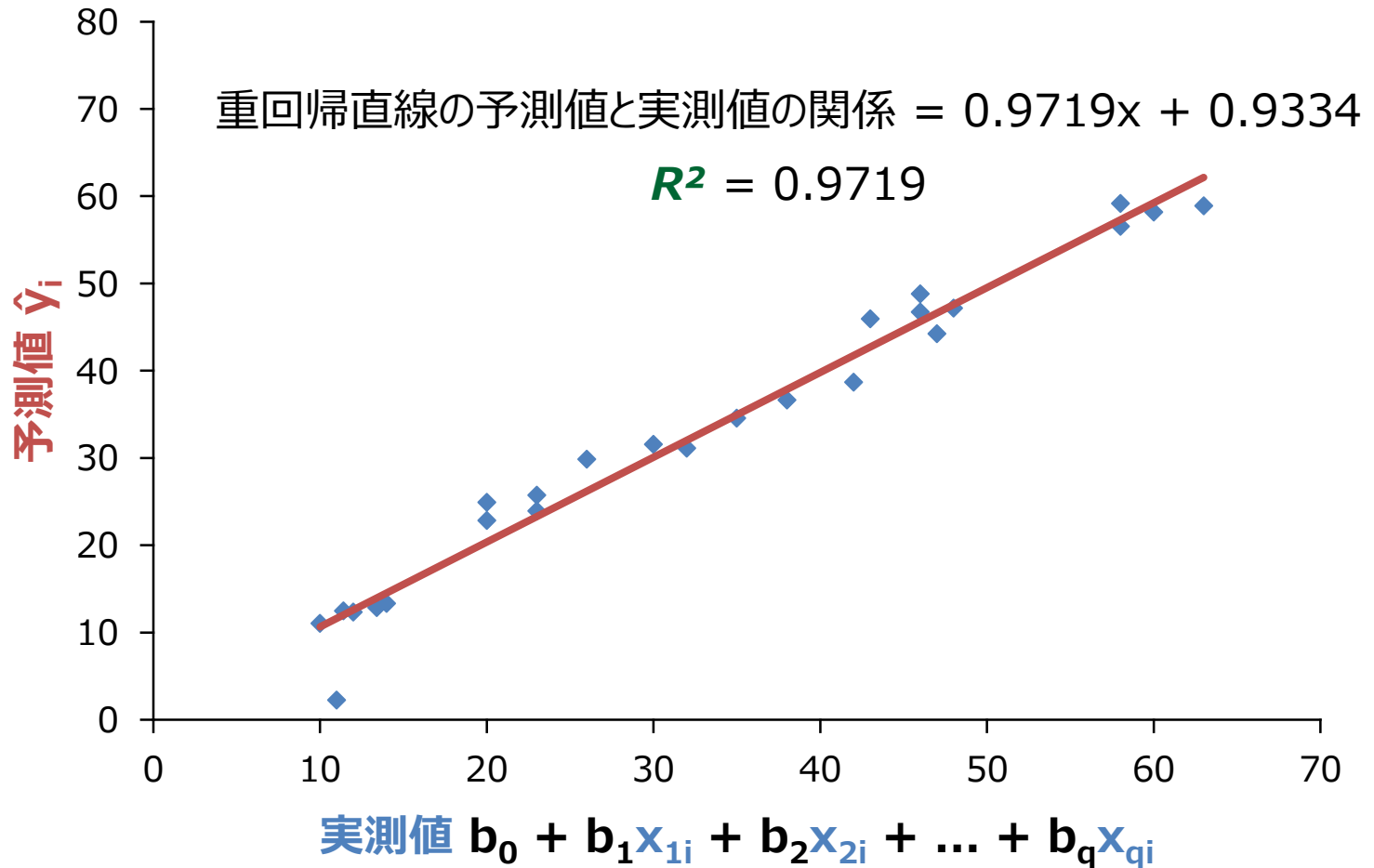
最小二乗法によって、重回帰直線の式を求める。

$$\text{予測値 } \hat{y}_i = b_0 + b_1 x_{1i} + b_2 x_{2i} + \dots + b_q x_{qi}$$



重回帰分析 (説明変数が2つ以上), p210, example 40

重回帰直線 : 予測値 $\hat{y}_i = b_0 + b_1x_{1i} + b_2x_{2i} + \dots + b_qx_{qi}$





変数選択-重回帰分析, p224, example 41

重回帰直線：予測値 $\hat{y}_i = b_0 + b_1x_{1i} + b_2x_{2i} + \dots + b_qx_{qi}$

説明変数 x が全て予測に役立つとは限らない！
有意でない説明変数 x が入ることによって、
逆に予測の精度が下がることもある。

変数選択-重回帰分析:

複数の説明変数 x から、目的変数 y に大きく影響を与えている
変数だけを用いて重回帰関数を求める方法。

1. **変数増加法**：全ての変数 x を外す → 変数 x を加えていく。
2. **変数減少報**：全ての変数 x を含む → 変数 x を取り除いていく。

整次多項式による回帰分析, p231, example 42

説明変数 x が一つの場合でも、

一次式 $y = b_0 + b_1x$

では不十分な場合がある。そのような場合は、

整次多項式 $y = b_0 + b_1x + b_2x^2 + \dots + b_qx^q$

を求める。

| 説明変数 | 目的変数 |
|----------|----------|
| x | y |
| x_1 | y_1 |
| x_2 | y_2 |
| \vdots | \vdots |
| x_n | y_n |



| 説明変数 | | | | 目的変数 |
|----------|----------|---------|----------|----------|
| x | x^2 | \dots | x^q | y |
| x_1 | x_1^2 | \dots | x_1^q | y_1 |
| x_2 | x_2^2 | \dots | x_2^q | y_2 |
| \vdots | \vdots | | \vdots | \vdots |
| x_n | x_n^2 | \dots | x_n^q | y_n |

単回帰分析

重回帰分析として考えることができる