データ分析基礎
回帰分析

京都大学 国際高等教育院 附属データ科学イノベーション教育研究センター

せきど ひろと

關戸 啓人

sekido.hiroto.7a@kyoto-u.ac.jp

回帰分析と最小二乗法

回帰分析と回帰曲線

- ★ **多変量解析**
  - ★ 複数の確率変数の関係を調べる
    - ★「身長」と「体重」の関係
    - ★「気温」と「ビールの売上」の関係
    - ★「朝食を食べる割合」と「テストの点数」の関係

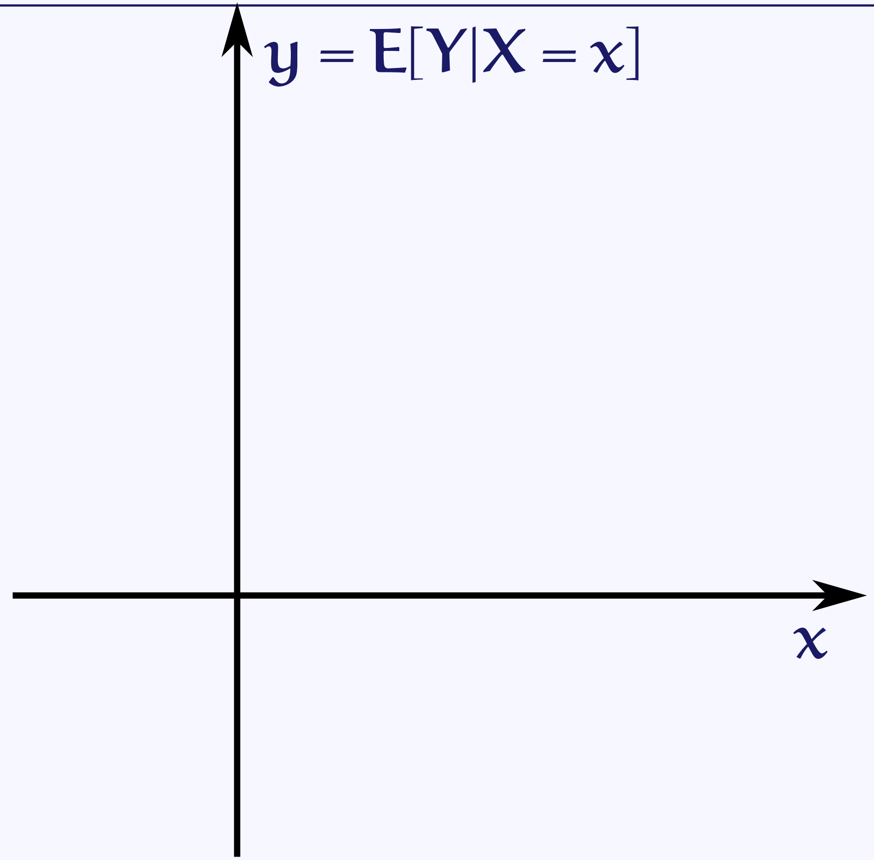
- ★ **回帰分析**
  - ★ **回帰曲線（回帰曲面）**を推定することで複数の確率変数の関係を調べる

- ★ 回帰曲線
  - ★ 2つの確率変数 X と Y を考える
  - ★ $X = x$ という条件下での Y の平均 $\mathrm{E}[Y|X = x]$ を x の関数と思ったとき，それを回帰曲線という

回帰分析と回帰曲線

★ $X = x$ という条件下での Y の平均 $\mathrm{E}[Y|X = x]$ を x の関数と思ったとき，それを回帰曲線という

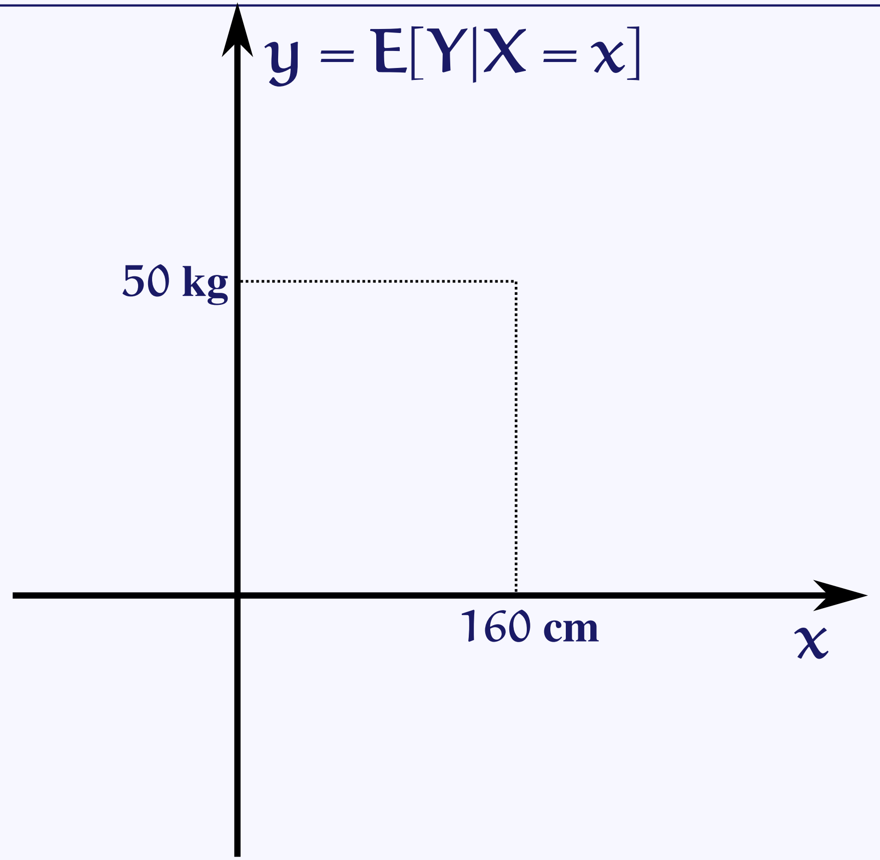
★ 例）X は身長，Y は体重を表すとする



回帰分析と回帰曲線

★ $X = x$ という条件下での Y の平均 $\mathrm{E}[Y|X = x]$ を x の関数と思ったとき，それを回帰曲線という

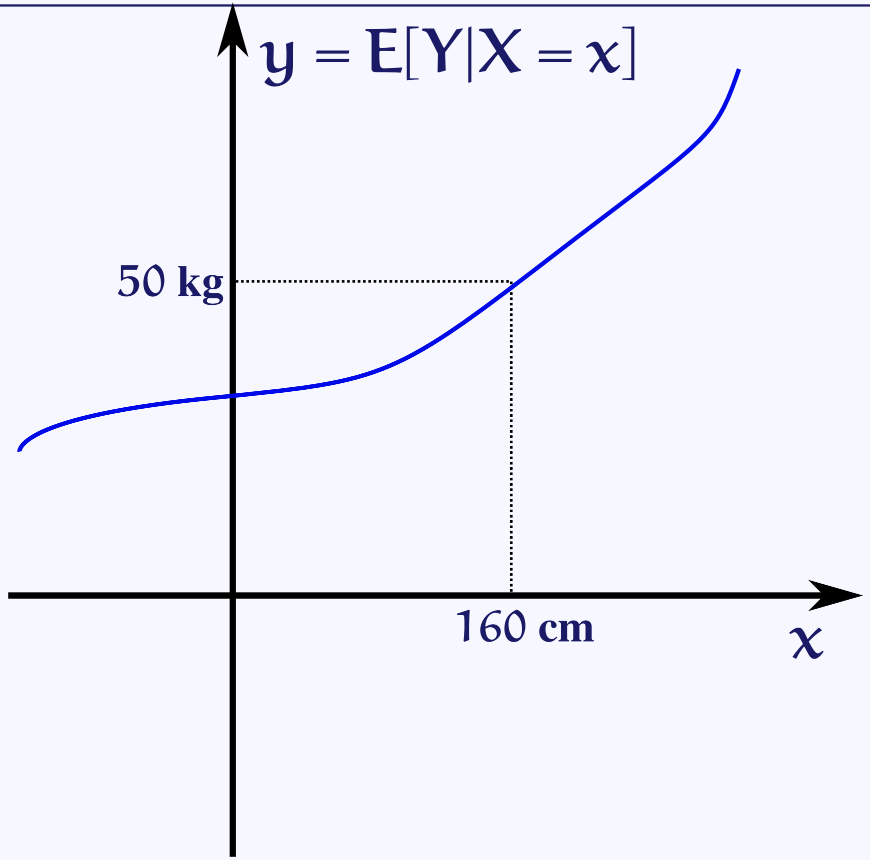
★ 例）X は身長，Y は体重を表すとする



回帰分析と回帰曲線

★ $X = x$ という条件下での Y の平均 $\mathrm{E}[Y|X = x]$ を x の関数と思ったとき，それを回帰曲線という

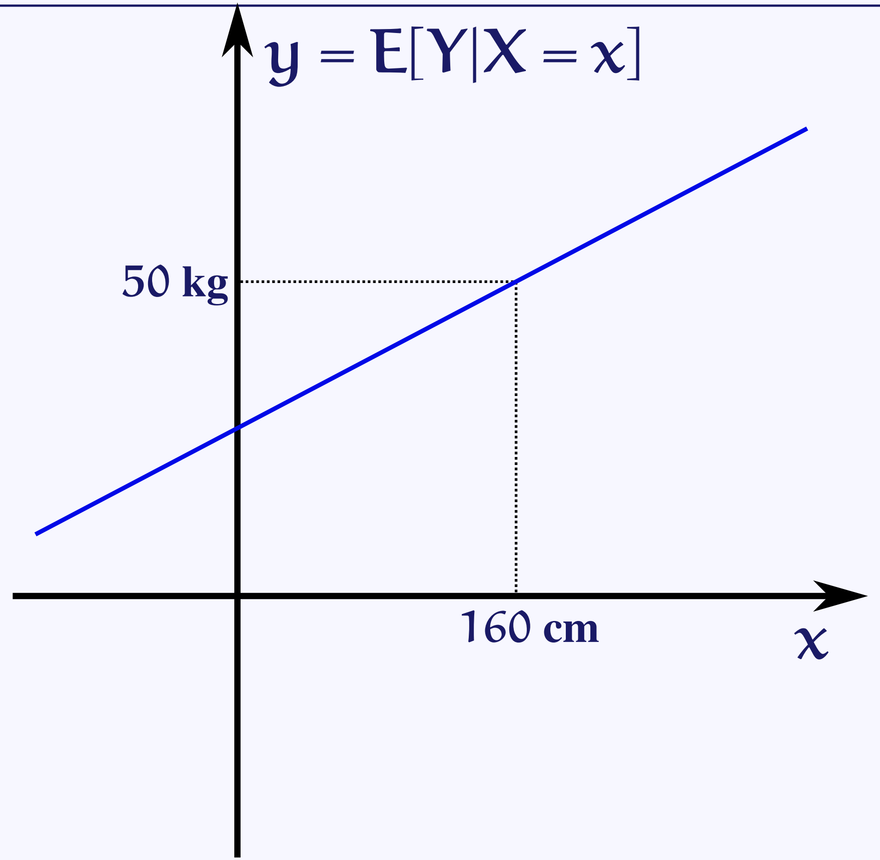
★ 例）X は身長，Y は体重を表すとする



回帰分析と回帰曲線

★ $X = x$ という条件下での Y の平均 $\mathrm{E}[Y|X = x]$ を x の関数と思ったとき，それを回帰曲線という

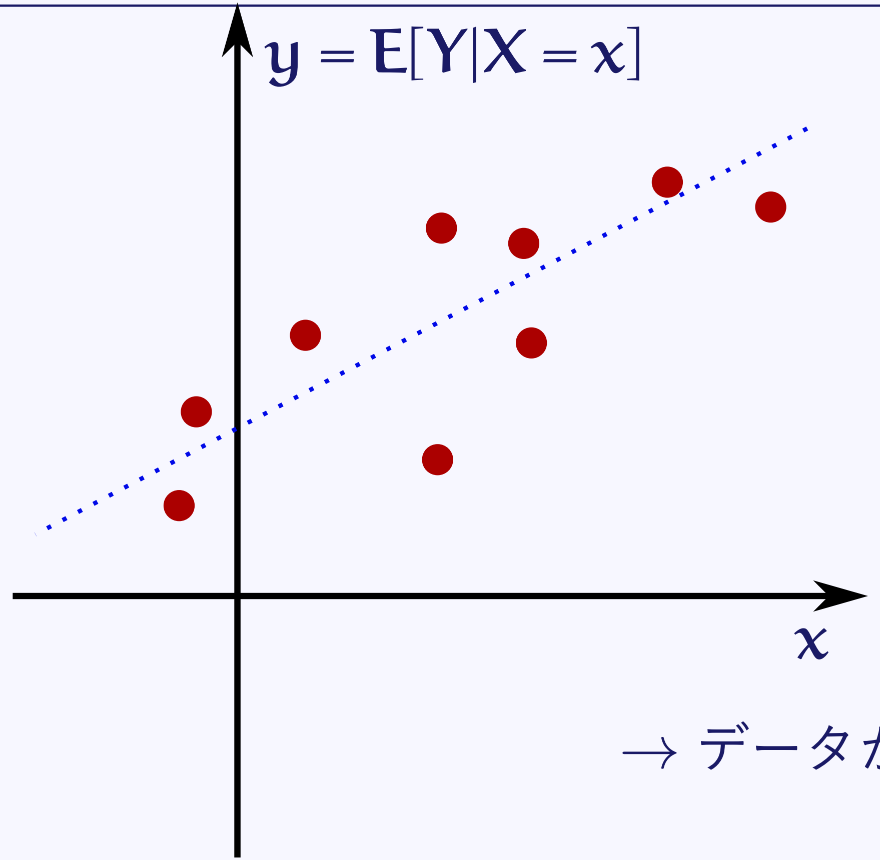
★ 簡単のため直線 $y = ax + b \ (a, b \in \mathbb{R})$ であると仮定することが多い



回帰分析と回帰曲線

★ $X = x$ という条件下での Y の平均 $\mathrm{E}[Y|X = x]$ を x の関数と思ったとき，それを回帰曲線という

★ 簡単のため直線 $y = ax + b\ (a, b \in \mathbb{R})$ であると仮定することが多い



$\to$ データから a, b を推定する

単回帰分析と重回帰分析

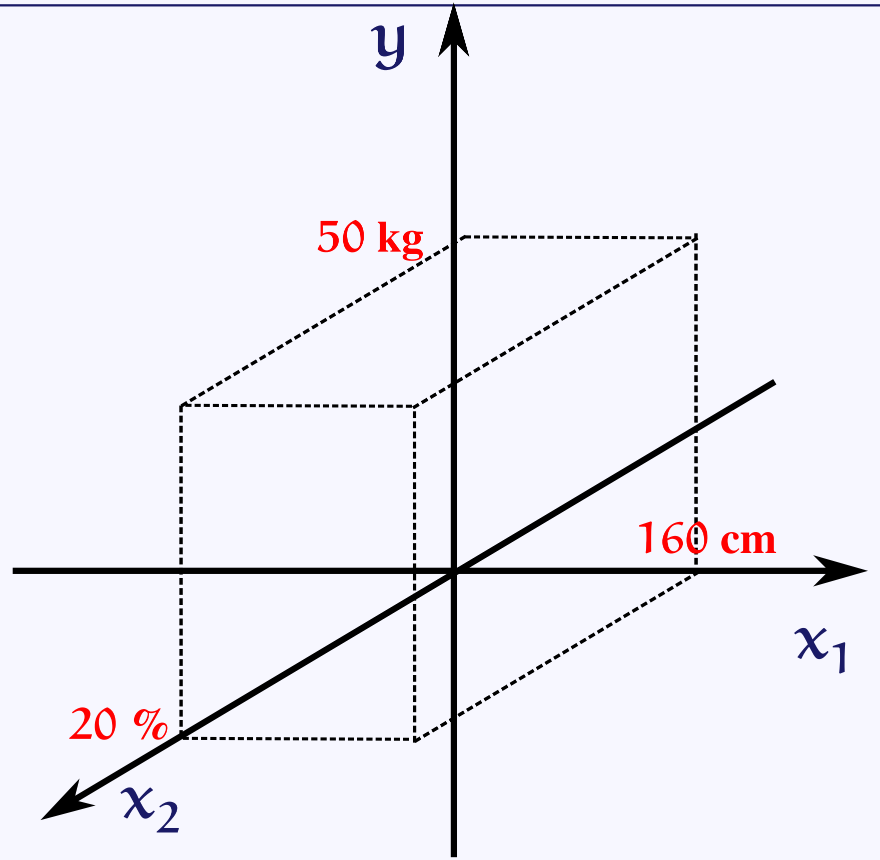
- ★ $\mathrm{E}[Y|X=x]$ を推定するときは，X **は説明変数，**Y **は被説明変数（目的変数）**と呼ばれる
- ★ つまり，Y がどのような値を取るかは X によって決まる，と考えている
  - ★ Y: ビールの売上，X: 気温
  - ★ Y: テストの点数，X: 朝食を食べる割合

- ★ 説明変数は複数あっても良い
- ★ 説明変数が $X_1, X_2, \ldots, X_n$ で，$\mathrm{E}[Y|X_1=x_1, X_2=x_2, \ldots, X_n=x_n]$ を考えても良い
  - ★ 説明変数が1個の場合を**単回帰分析**，複数の場合を**重回帰分析**という

重回帰分析の例

★ 説明変数が2個の場合：$y = \mathrm{E}[Y | X_1 = x_1, X_2 = x_2]$

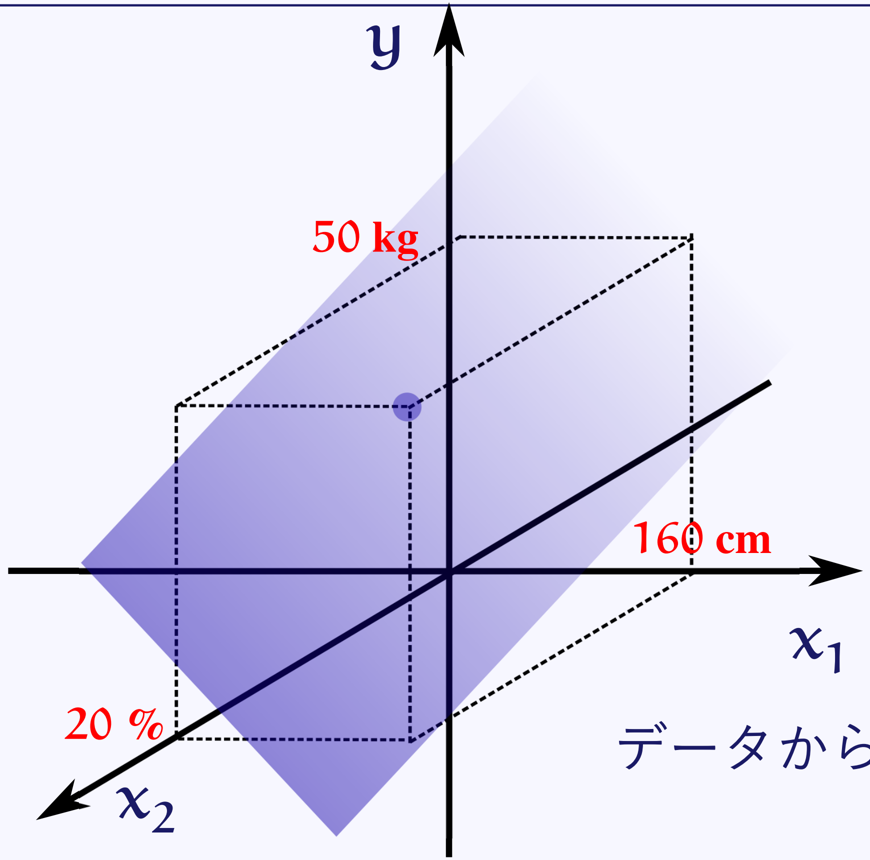
★ 例）X_1 は身長，X_2 は体脂肪率，Y は体重を表すとする



重回帰分析の例

★ 説明変数が2個の場合：$y = \mathrm{E}[Y|X_1 = x_1, X_2 = x_2]$

★ 回帰曲面は平面 $y = a_1x_1 + a_2x_2 + b$ $(a_1, a_2, b \in \mathbb{R})$ であると仮定することが多い

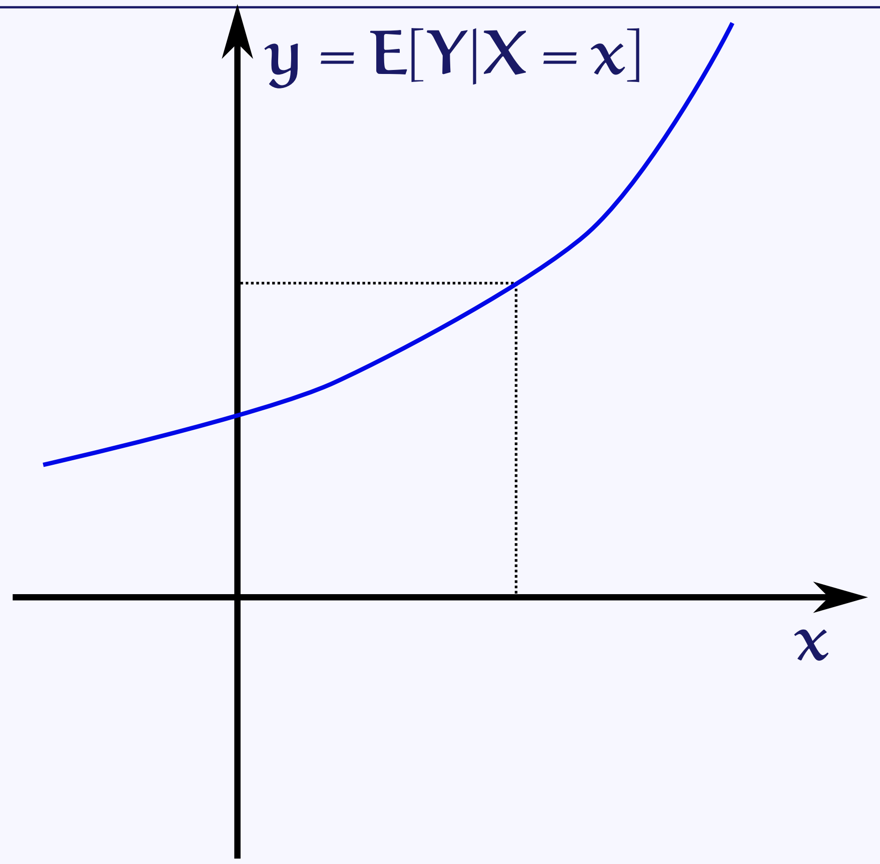


データから a_1, a_2, b を推定する

重回帰分析の例: 二次関数

★「身長 (X)」と「体重 (Y)」の関係は直線なのか？

- ★ BMI などを考慮すると二次関数 $y = ax^2 + bx + c$ と仮定したほうが良いのでは？

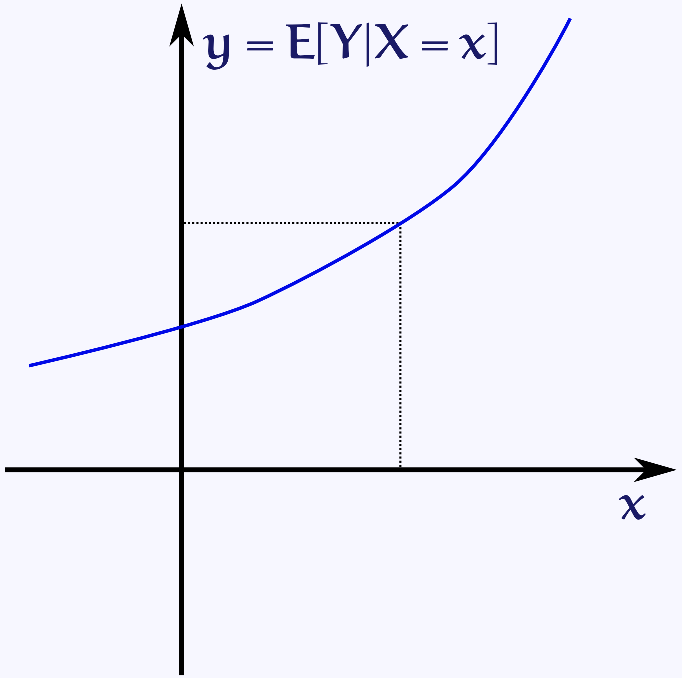


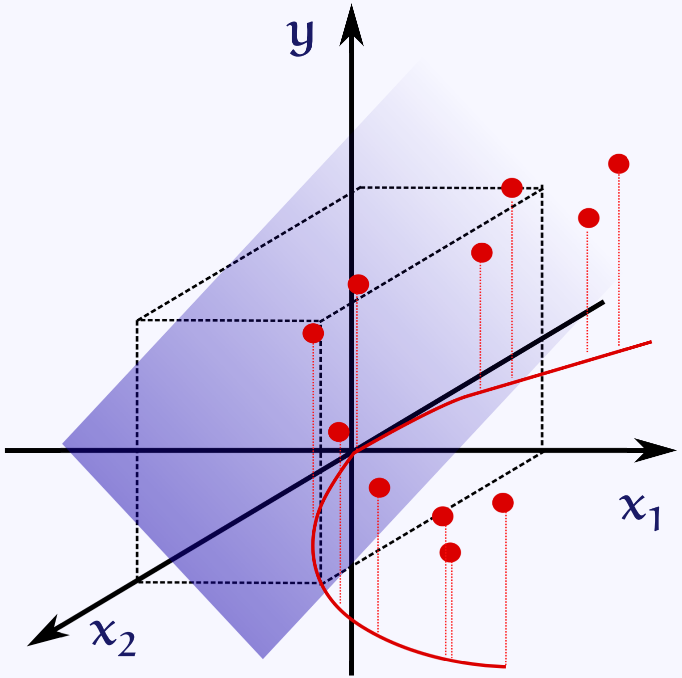
重回帰分析の例: 二次関数

★「身長 (X)」と「体重 (Y)」の関係は直線なのか？

★ BMIなどを考慮すると二次関数 $y = ax^2 + bx + c$ と仮定したほうが良いのでは？
→ 重回帰分析

★ $y = a_1x_1 + a_2x_2 + b$ において $(a_1, a_2, b, x_1, x_2) \to (a, b, c, x^2, x)$ と読み替えれば良い





問題

- ★ 確率変数 X は血圧を表すとし，Y は年収を表すとする
- ★「血圧」と「年収」の関係を回帰分析で調べた場合どうなるか？

- ★ 回帰直線は右肩上がりになる
  - ★ $y = ax + b$ とすると $a > 0$

- ★ 年収を上げるには血圧を上げれば良い！
  - ★ **と考えるのは危険**

解説

★「年収」と「血圧」には確かに正の相関があるが因果関係などは何も言っていない

★ 年収が多い人は，ストレスが掛かる仕事をしており，血圧が高いかもしれない
- ★ 実はこの場合はこれもほぼ正しくない

★「年収」も「血圧」も「年齢」と正の相関がある
- ★ 確率変数 X_1 は血圧を，X_2 は年齢を，Y は年収を表すとする
- ★ 重回帰分析をすると $y = a_1x_1 + a_2x_2 + b$ において $a_2 > 0$ だが $a_1 > 0$ とは限らない

★ 仮定が良くなかった

朝食を食べる割合の例について検証

- ★「朝食を食べる割合 (X)」と「テストの点数 (Y)」の関係を回帰分析で調べた場合はどうなるか？

- ★ 回帰直線は右肩上がりになる
  - ★ $y = ax + b$とすると$a > 0$

- ★ テストの点数を上げるには朝食を食べれば良い！
  - ★ 栄養がある状態のほうが頭が働いて勉強できる
  - ★ 朝食を食べる割合が多い家庭はしつけができてるだけなのでは…，無理やり朝食を食べてもテストの点数は変わらないよ
  - ★ 例え，朝食を食べることとテストの点数に直接的な因果関係がなくても，無理やり朝食を食べたら生活環境とかの影響でテストの点数上がるかも
    - ★ よくわからない

- ★ 多角的に分析し，良さそうなら実際に試してみる

最小二乗法の概要

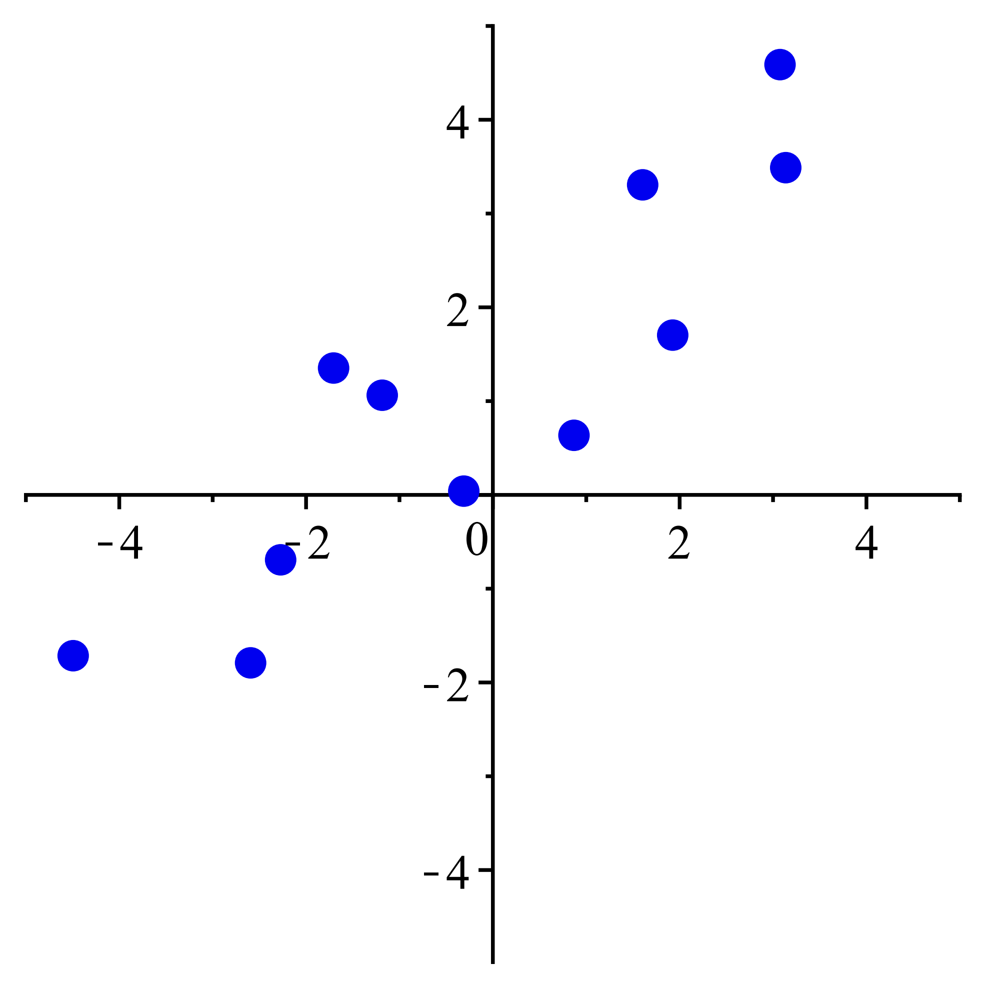
★ **未知な関数**を得られた**データから推定**したい

★ 未知関数 $f(x)$ の形はわかっていて，未知パラメータを含む形で書かれる

★ データ (x_j, y_j) は $f(x_j)$ での値が y_j であることを「示唆」する

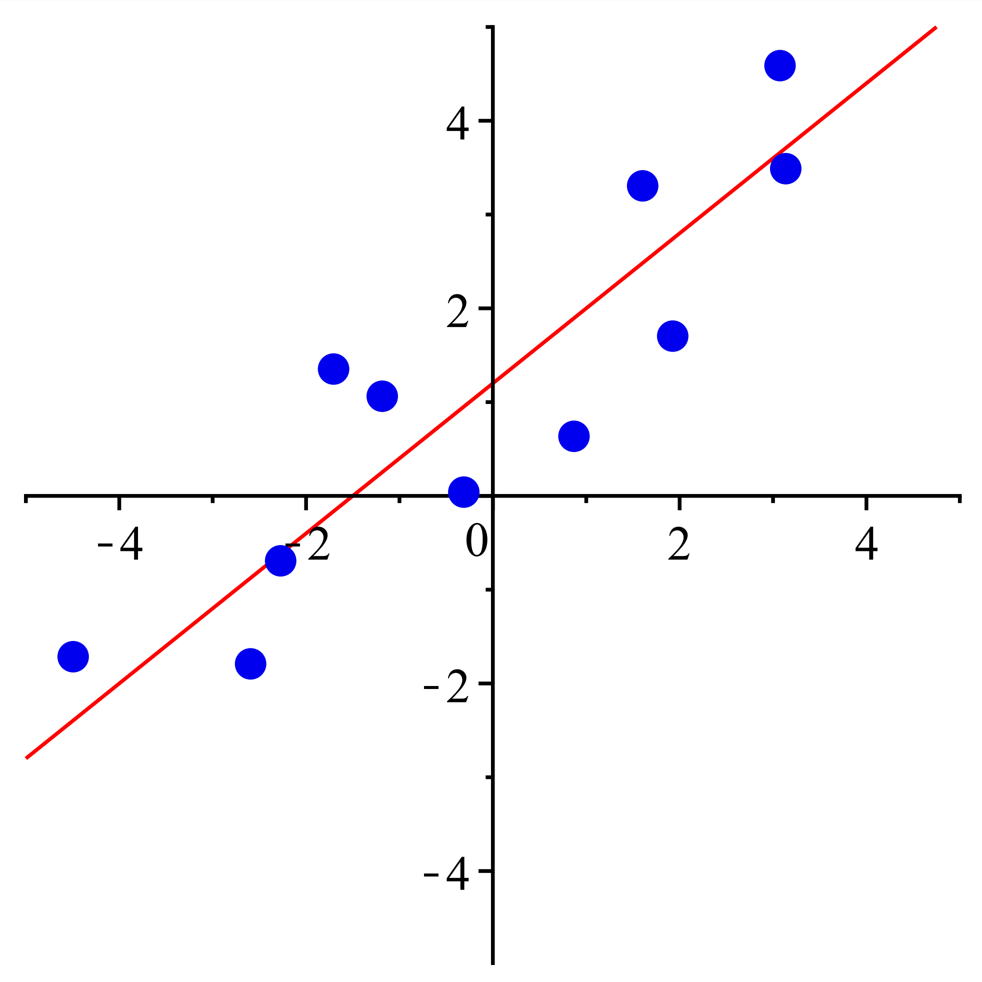
★ データは厳密に「正しい」訳ではない．つまり厳密に $f(x_j) = y_j$ とは限らない（測定誤差などが含まれている）

★ 2 変数以上の場合は x はベクトルだと思えば良い



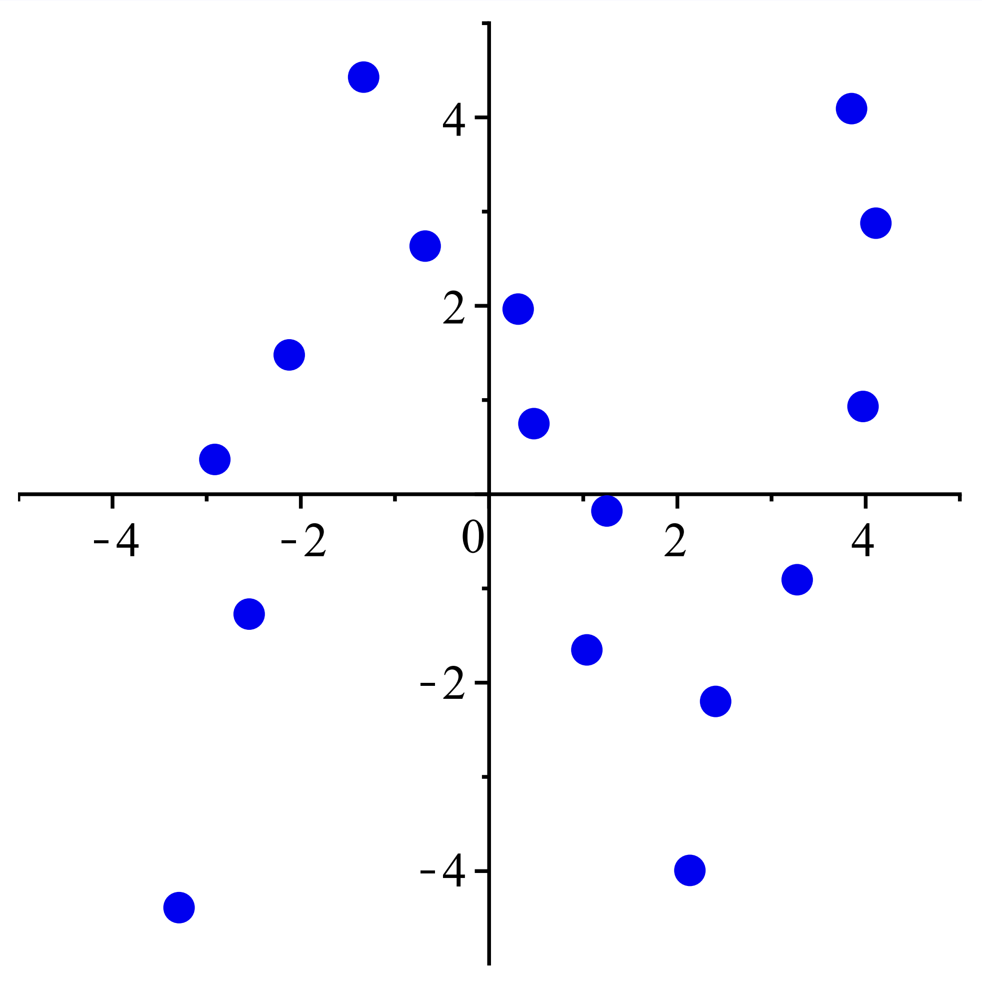
$$f(x) = \theta_1 x + \theta_0$$

最小二乗法の例 (その1)



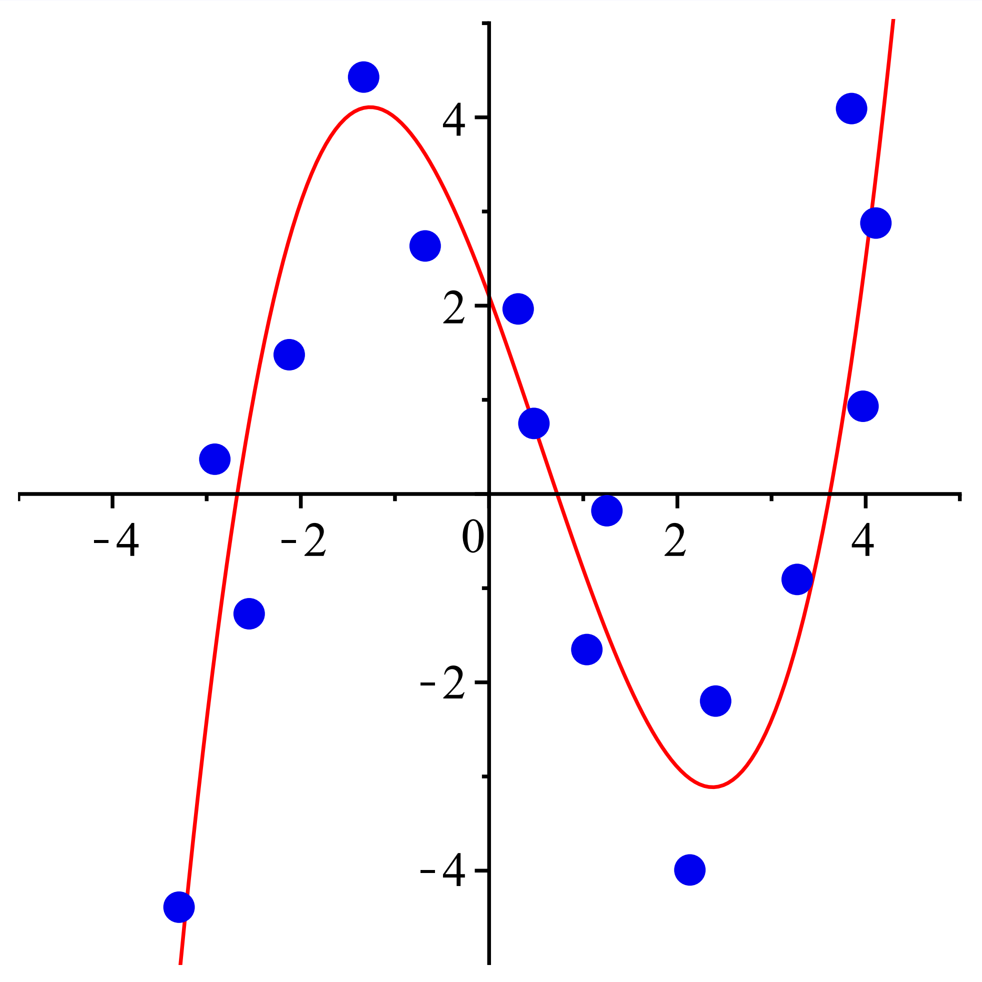
$$f(x) = 0.8x + 1.2$$

最小二乗法の例 (その2)



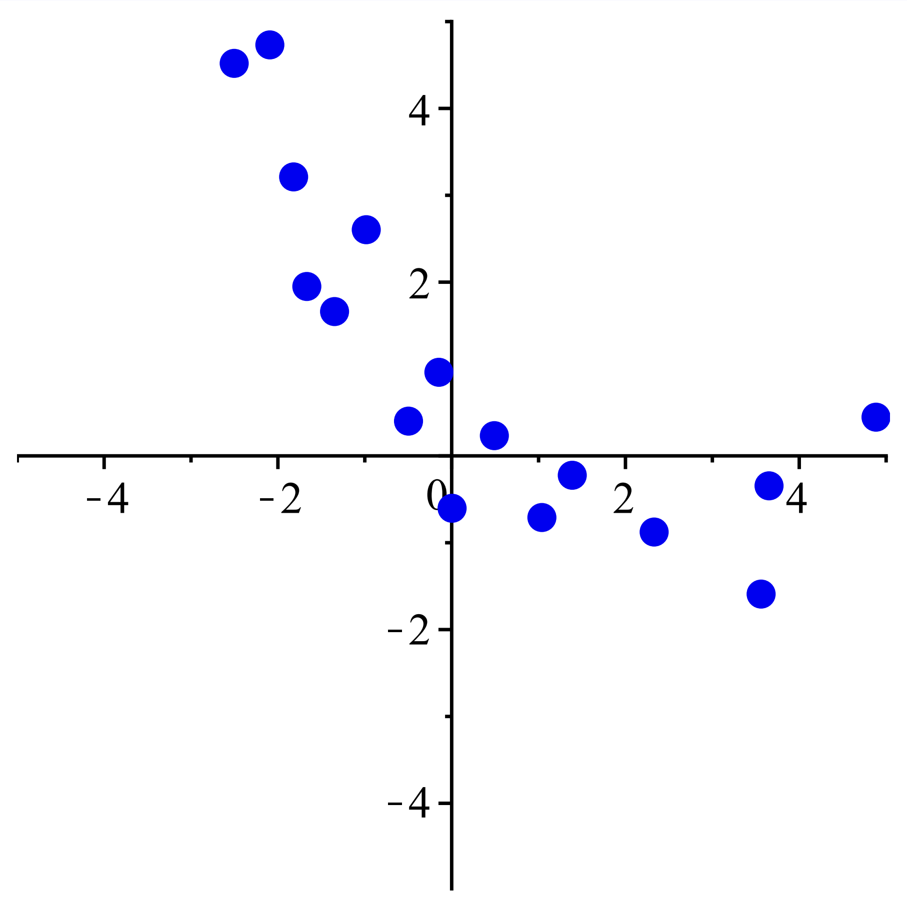
$$f(x) = \theta_3 x^3 + \theta_2 x^2 + \theta_1 x + \theta_0$$

最小二乗法の例 (その2)



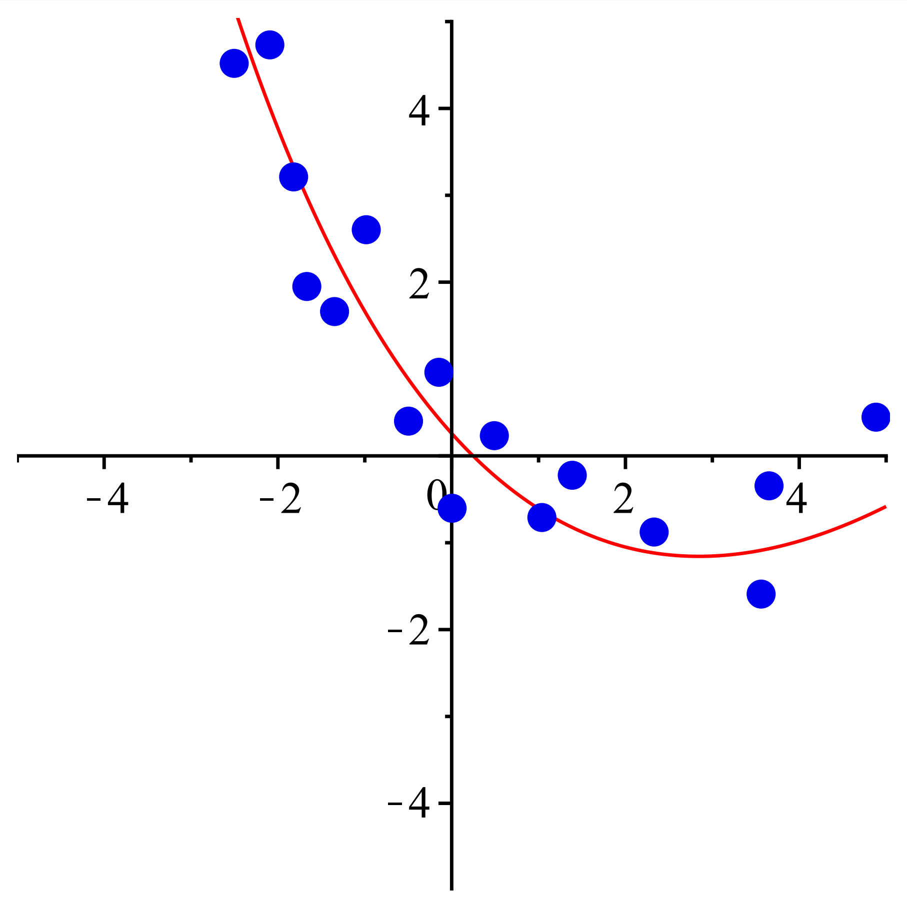
$$f(x) = 0.3x^3 - 0.5x^2 - 2.7x + 2.1$$

最小二乗法の例 (その3)



$$f(x) = \frac{\theta_2 x^2 + \theta_1 x + \theta_0}{x + \theta_3}$$

最小二乗法の例 (その3)



$$f(x) = \frac{2.1x^2 - 13.1x + 3.1}{x + 12.0}$$

最小二乗法の例

- その1: **直線で近似する場合**
  - $f(x) = \theta_0 + \theta_1 x$
  - （単純な）単回帰分析
- その2: **未知関数がパラメータについて線形（線形最小二乗法）**
  - $f(x) = \theta_0 f_0(x) + \theta_1 f_1(x) + \cdots + \theta_{m-1} f_{m-1}(x)$
  - （単純な）重回帰分析，以下では主にこれを説明する
- その3: **未知関数がパラメータについて非線形（非線形最小二乗法）**
  - $f(x) = f(x; \theta_0, \theta_1, \ldots, \theta_{m-1})$
  - 複雑な式の形を指定した場合，解く場合は最適化の理論を用いる

回帰モデルの例 (1) — 単回帰モデル

★ 体重を意味する確率変数を W

★ 身長を意味する確率変数を H

★ モデル：$W = \theta_1 H + \theta_0 + \varepsilon$

★ データは，例えば

	体重 (kg)	切片	身長 (cm)
A 氏	56.8	1	163.3
B 氏	52.1	1	160.2
C 氏	52.6	1	158.0
D 氏	23.4	1	129.0
E 氏	32.1	1	139.7
F 氏	40.6	1	141.4

回帰モデルの例 (2-1) — 重回帰モデル

★ 体重を意味する確率変数を W

★ 身長を意味する確率変数を H

★ モデル：$W = \theta_2 H^2 + \theta_1 H + \theta_0 + \varepsilon$

★ データは，例えば

	体重 (kg)	切片	身長 (cm)	身長^2(cm^2)
A 氏	56.8	1	163.3	26666.89
B 氏	52.1	1	160.2	25664.04
C 氏	52.6	1	158.0	24964.00
D 氏	23.4	1	129.0	16641.00
E 氏	32.1	1	139.7	19516.09
F 氏	40.6	1	141.4	19993.96

回帰モデルの例 (2-2) — 重回帰モデル

★ 体重を W，身長 H，体脂肪率を F，性別を S

★ 性別は女性を 1，男性を 0 で表す

★ モデル：$W = \theta_3 S + \theta_2 F + \theta_1 H + \theta_0 + \varepsilon$

★ データは，例えば

	体重 (kg)	切片	身長 (cm)	体脂肪率 (%)	性別
A 氏	56.8	1	163.3	14.3	0
B 氏	52.1	1	160.2	15.3	0
C 氏	52.6	1	158.0	21.2	1
D 氏	23.4	1	129.0	13.3	1
E 氏	32.1	1	139.7	16.8	0
F 氏	40.6	1	141.4	19.6	1

線形最小二乗法の定義，および，性質 1

★ 観測と応答の関係

$$Y = \sum_{k=0}^{m-1} \theta_k f_k(x) + \varepsilon = f(x, \theta) + \varepsilon$$

は**線形回帰モデル**と呼ばれる

★ $f_k(x)$ は既知の関数

★ θ_k は未知のパラメータ， $\theta = (\theta_0, \theta_1, \ldots, \theta_{m-1})^{\mathrm{T}}$

★ ε は確率変数で平均0（$\mathrm{E}[\varepsilon] = 0$）

★ 実際に n 個のデータ $(x_1, y_1), \ldots, (x_n, y_n)$ を用いて

$$y_j = f(x_j, \theta) + \varepsilon_j, \quad j = 1, 2, \ldots, n$$

とする

★ y_j, ε_j は確率変数

★ ε_j は j 回目の観測における誤差

線形最小二乗法の定義，および，性質 2

★ $y_j = f(x_j, \theta) + \varepsilon_j, \ \ j = 1, 2, \dots, n$

★ 今回は，誤差 ε_j に対して以下の仮定を置く

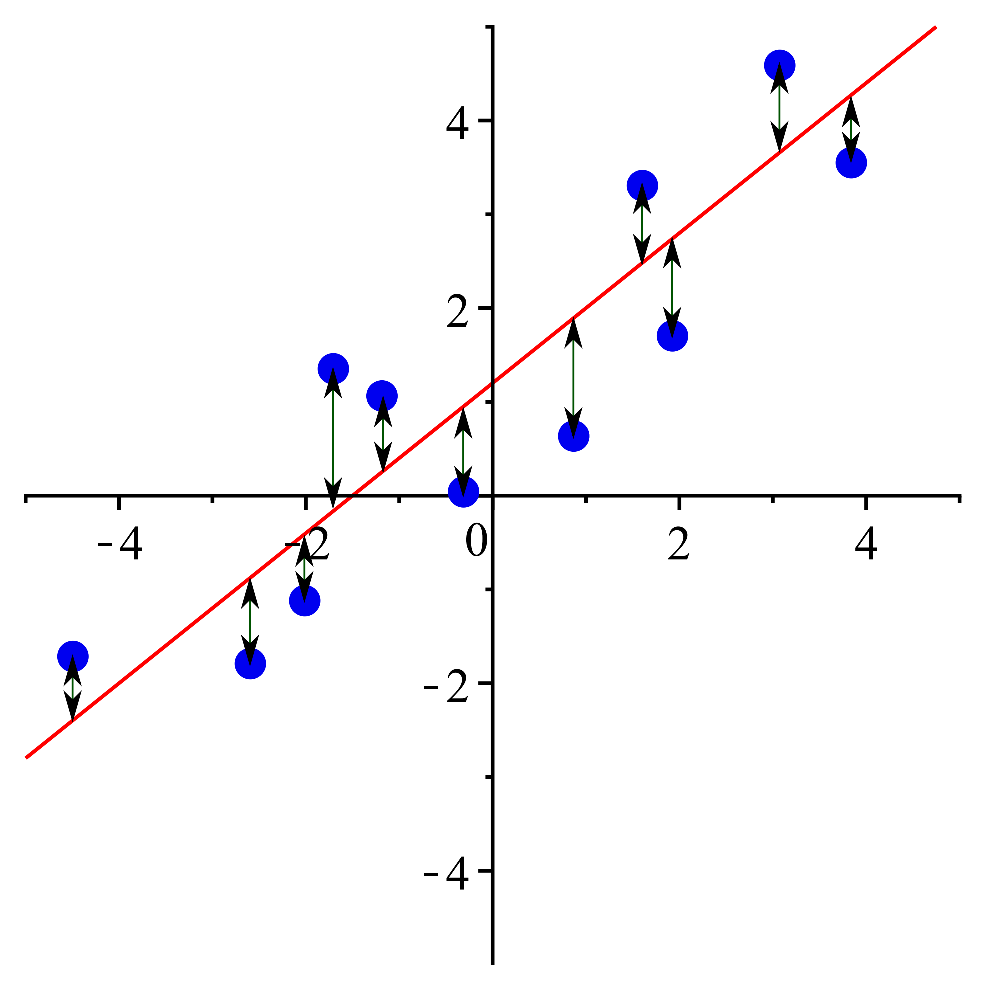
- ★ 平均は 0．つまり，$\mathrm{E}[\varepsilon_j] = 0$
- ★ 誤差の分散は等しく，正．つまり，$\mathrm{V}[\varepsilon_j] = \sigma^2 > 0$
- ★ 誤差は互いに無相関．つまり，$\mathrm{E}[\varepsilon_i \varepsilon_j] = 0, \ \ i \neq j$

★ **残差二乗和**

$$S(\beta) = \sum_{k=1}^{n} (y_k - f(x_k, \beta))^2$$

を**最小化**する未知パラメータベクトル β を**最小二乗推定量** $\hat{\theta}$ と言う

絵で見る最小二乗法



緑の線の長さの二乗和を最小化するように，未知パラメータθを推定

線形最小二乗法の定義，および，性質 3

- ★ 最小二乗推定量 $\hat{\theta}$ は，最良線形不偏推定量である
  - ★ $\mathrm{E}[\hat{\theta}] = \theta$（不偏）
  - ★ $\hat{\theta}$ は，y_j について線形の式で書ける（線形）
  - ★ その中で，分散がある意味で最小（最良）
    - ★ 任意の不偏性と線形性を満たす β に対して，$\mathrm{Cov}[\beta] - \mathrm{Cov}[\hat{\theta}]$ が非負定値

- ★ 誤差 ε が正規分布に従うとき，最小二乗推定量 $\hat{\theta}$ は，最尤推定量である
  - ★ つまり，$x_1, \ldots, x_n$ を固定して，測定結果として $y_1, \ldots, y_n$ が得られる確率を θ の関数として考えたとき，その確率の値が最大となるのが $\theta = \hat{\theta}$ のとき

最小二乗法推定量 (その1)

★ 方針

★ 残差二乗和

$$S(\beta) = \sum_{k=1}^{n} (y_k - f(x_k, \beta))^2$$

を最小化したいのだから，$\beta_0, \beta_1, \ldots, \beta_{m-1}$ で偏微分して 0 になる β を見つければ良い

最小二乗法推定量 (その1)

★ $f(x,\beta) = \beta_1 x + \beta_0$ の場合

★ $S(\beta) = \sum_{k=1}^{n} (y_k - \beta_1 x_k - \beta_0)^2$ であるから

★ $$\frac{\partial}{\partial \beta_1} S(\beta) = 2\sum_{k=1}^{n} (x_k^2 \beta_1 + x_k \beta_0 - x_k y_k) = 0$$

★ $$\frac{\partial}{\partial \beta_0} S(\beta) = 2\sum_{k=1}^{n} (x_k \beta_1 + \beta_0 - y_k) = 0$$

★ つまり，次の連立一次方程式を解けば良い

★ $$\begin{pmatrix} \sum x_k^2 & \sum x_k \\ \sum x_k & n \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_0 \end{pmatrix} = \begin{pmatrix} \sum x_k y_k \\ \sum y_k \end{pmatrix}$$

★ $\beta_1 = \frac{n\sum x_j y_j - \sum x_j \sum y_j}{n\sum x_j^2 - (\sum x_j)^2}$, $\beta_0 = \frac{\sum x_j^2 \sum y_j - \sum x_j y_j \sum x_j}{n\sum x_j^2 - (\sum x_j)^2}$

最小二乗法推定量 (その 2)

- ★ $f(x, \beta) = \beta_0 f_0(x) + \beta_1 f_1(x) + \cdots + \beta_{m-1} f_{m-1}(x)$ の場合
  - ★ $S(\beta) = \sum_{k=1}^{n} \left(y_k - \sum_{j=0}^{m-1} \beta_j f_j(x_k) \right)^2$ であるから
    - ★ $\frac{\partial}{\partial \beta_i} S(\beta) = 2 \sum_{k=1}^{n} f_i(x_k) \left(\left(\sum_{j=0}^{m-1} f_j(x_k) \beta_j \right) - y_k \right) = 0$
    - ★ $\sum_{j=0}^{m-1} \sum_{k=1}^{n} f_i(x_k) f_j(x_k) \beta_j = \sum_{k=1}^{n} f_i(x_k) y_k$

正規方程式

★ つまり，連立一次方程式 $B\beta = b$ を解けば良い

- ★ $B \in M_m(\mathbb{R}), \quad B_{ij} = \sum_{k=1}^{n} f_i(x_k) f_j(x_k)$
- ★ $b \in \mathbb{R}^m, \quad b_i = \sum_{k=1}^{n} f_i(x_k) y_k$

★ 行列 B がフルランクであれば，最小二乗推定量が一意に定まる

★ $B\beta = b$ は正規方程式と呼ばれる

★ **数値計算する際は，この方程式を直接解くよりも高精度な方法が存在する**

正規方程式

★ 行列 $A \in M_{n,m}(\mathbb{R})$ を以下で定義（ヤコビアン，データ行列）

- ★ $A_{ij} = f_j(x_i) = \frac{\partial}{\partial \beta_j} f(x_i)$

★ $B = A^{\mathrm{T}} A$

★ $b = A^{\mathrm{T}} y$　(ただし $y = (y_1 \ \cdots \ y_n)^{\mathrm{T}}$)

★ **正規方程式**は以下のように書き直される

- ★ $A^{\mathrm{T}} A \beta = A^{\mathrm{T}} y$

★ 行列 A が列フルランクの場合

- ★ 最小二乗推定量は $\hat{\theta} = (A^{\mathrm{T}} A)^{-1} A^{\mathrm{T}} y$

補足：そもそも最初から行列とベクトルで

★最小化したい残差二乗和は

$$S(\beta) = \sum_{k=1}^{n} (y_k - f(x_k, \beta))^2 = (A\beta - y, A\beta - y) = \|A\beta - y\|_2^2$$

★βで微分すると以下：これが0になるとおくと，正規方程式を得る

$$2A^{\mathrm{T}}A\beta - 2A^{\mathrm{T}}y$$

★補足1：

$$\begin{aligned}(A\beta - y, A\beta - y) &= (A\beta, A\beta) - 2(A\beta, y) + (y, y) \\ &= \beta^{\mathrm{T}}A^{\mathrm{T}}A\beta - 2(A^{\mathrm{T}}y)^{\mathrm{T}}\beta + y^{\mathrm{T}}y\end{aligned}$$

★補足2（ベクトルで微分する）：

$$\frac{\mathrm{d}f}{\mathrm{d}\beta} = \left(\frac{\mathrm{d}f}{\mathrm{d}\beta_0} \ \cdots \ \frac{\mathrm{d}f}{\mathrm{d}\beta_{m-1}}\right)^{\mathrm{T}}$$

★ $\frac{\mathrm{d}}{\mathrm{d}x}(a^{\mathrm{T}}x) = a, \ \frac{\mathrm{d}}{\mathrm{d}x}(x^{\mathrm{T}}Ax) = (A + A^{\mathrm{T}})x$

QR 分解を用いて解く

- **行列 A は列フルランクで QR 分解できたとする**
  - $A = QR$
    - $Q \in M_{n,m}(\mathbb{R})$ は列ベクトルが長さ 1 で互いに直交
    - $R \in M_m(\mathbb{R})$ は正則な上三角行列

- このとき，正規方程式は
  - $A^{\mathrm{T}}A\beta = A^{\mathrm{T}}y$
  - $(QR)^{\mathrm{T}}QR\beta = (QR)^{\mathrm{T}}y$
  - $R^{\mathrm{T}}Q^{\mathrm{T}}QR\beta = R^{\mathrm{T}}Q^{\mathrm{T}}y$
  - $R^{\mathrm{T}}R\beta = R^{\mathrm{T}}Q^{\mathrm{T}}y$ $\quad (Q^{\mathrm{T}}Q = I)$
  - $R\beta = Q^{\mathrm{T}}y$ $\quad$ (R^{T} は正則)

- R は上三角行列であるから，これは簡単に解ける

行列 A が列フルランクでない場合

- 行列 A が列フルランクでない場合は，最小二乗推定量は一意に定まらない（これは**そもそもナンセンスな場合が多い**）
  - 最小二乗推定量の中で，$\|\beta\|_2$ を最小とするものを求めることが多い
    - $\|\beta\|_2 = \|\beta\| = \sqrt{\beta_0^2 + \beta_1^2 + \cdots \beta_{m-1}^2} = \sqrt{\beta^{\mathrm{T}}\beta}$

- 結論を言うと，A の Moore–Penrose の一般逆行列を A^+ と書くと $A^+y = R^+Q^{\mathrm{T}}y$ が答え
  - ある程度ロバストに計算できる方法は特異値分解
  - 高速に計算するなら完全ピボット選択付き QR 分解をして直交変換

一般逆行列

★ 正則でなくても，長方行列でも良い行列 $A \in M_{mn}(\mathbb{R})$ に対して，$AXA = A$ を満たす行列 $X \in M_{nm}(\mathbb{R})$ を一般逆行列といい A^{-} で表す

- ★ A^{-} は必ず存在し，一般的には A^{-} は一意ではなく複数存在する
- ★ 連立一次方程式 $Ax = b$ の解の一つは，存在するならば $x = A^{-}b$ と書ける
- ★ 連立一次方程式 $Ax = b$ の解は，存在するならば，任意のベクトル y を用いて $x = A^{-}b + (I - A^{-}A)y$ と書ける
- ★ 連立一次方程式 $Ax = b$ は $(I - AA^{-})b = 0$ ならば解が存在する

Moore–Penrose の一般逆行列

★ 正則でなくても，長方行列でも良い行列 $A \in M_{mn}(\mathbb{R})$ に対して，$AXA = A, XAX = X, (AX)^{\mathrm{T}} = AX, (XA)^{\mathrm{T}} = XA$ を満たす行列 $X \in M_{nm}(\mathbb{R})$ を Moore–Penrose の一般逆行列といい A^{+} で表す

- ★ A^{+} は必ず存在し，一意である
- ★ 連立一次方程式 $Ax = b$ の解が存在するならば, その中で $\|x\|_2$ が最小となるものは $x = A^{+}b$ となる
- ★ 連立一次方程式 $Ax = b$ の解が存在しなければ，$\|Ax - b\|_2$ が最小とするのは $x = A^{+}b$ となる