統計モデル

統計モデル

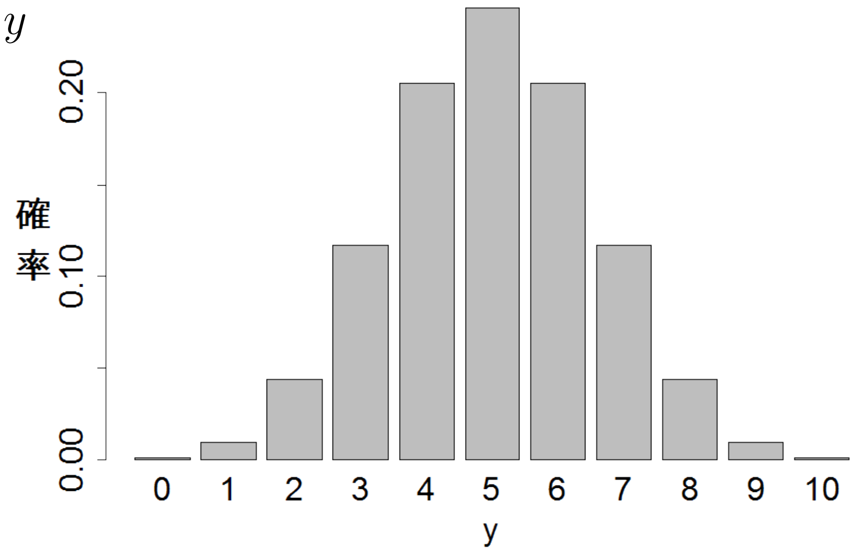
- ばらつきや誤差を含むデータの背後にある規則性，データを発生させる仕組みを数式で表したもの
- 観測誤差を含む観測データから背後にある現象を分析・予測

統計モデル

- コインを投げて表が出る確率をθとして，
  N回コインを投げたときにy回表が出る確率

$$P(y|N,\theta) = {}_N C_y \theta^y (1-\theta)^{N-y}$$

- 二項分布モデル
- $\theta = 0.5$
- $N = 10$

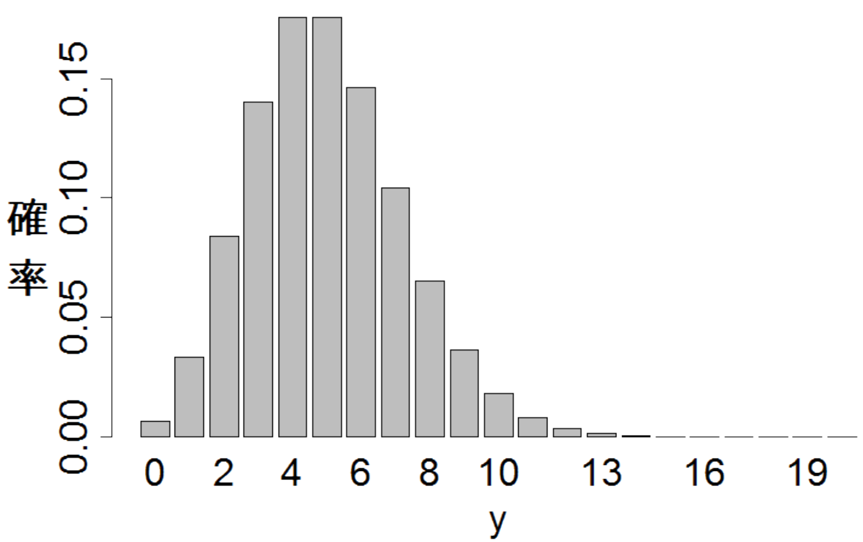


統計モデル

- 1時間に平均λ人の客が来る窓口に，
  1時間にy人の客が来る確率

$$P(y) = \frac{\lambda^y e^{-\lambda}}{y!}$$

- ポアソン分布モデル
- $\lambda = 5$



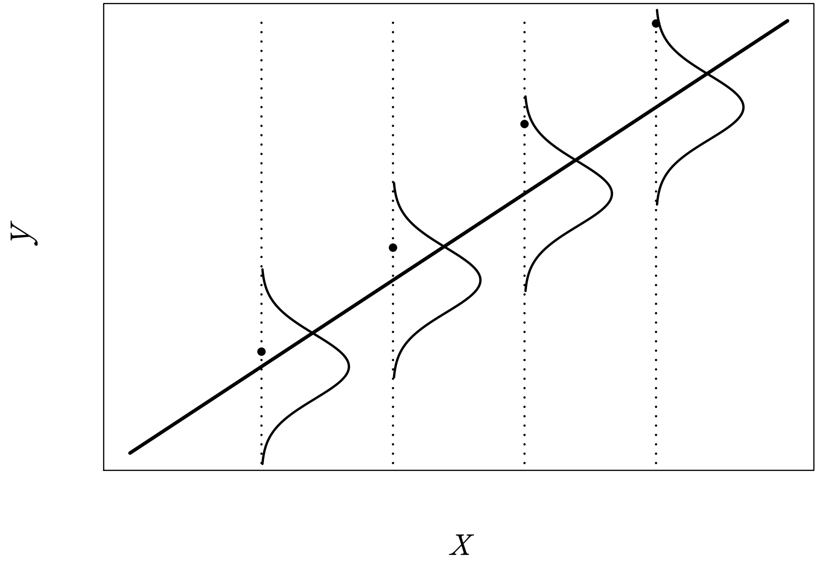
統計モデル

- 単位面積当たりの給水量(x)に対して，植物の収穫量(y)が，平均$\beta_0 + \beta_1 x$，分散σ^2 (標準偏差σ)の正規分布にしたがう
- 給水量xのときに収穫量yの確率密度

$$f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{\{y-(\beta_0+\beta_1 x)\}^2}{2\sigma^2} \right]$$

統計モデル

- yが，平均$\beta_0 + \beta_1 x$，分散σ^2 (標準偏差σ)の正規分布にしたがう



$$y \simeq \beta_0 + \beta_1 x$$

と近似

統計モデル

モデルの構築

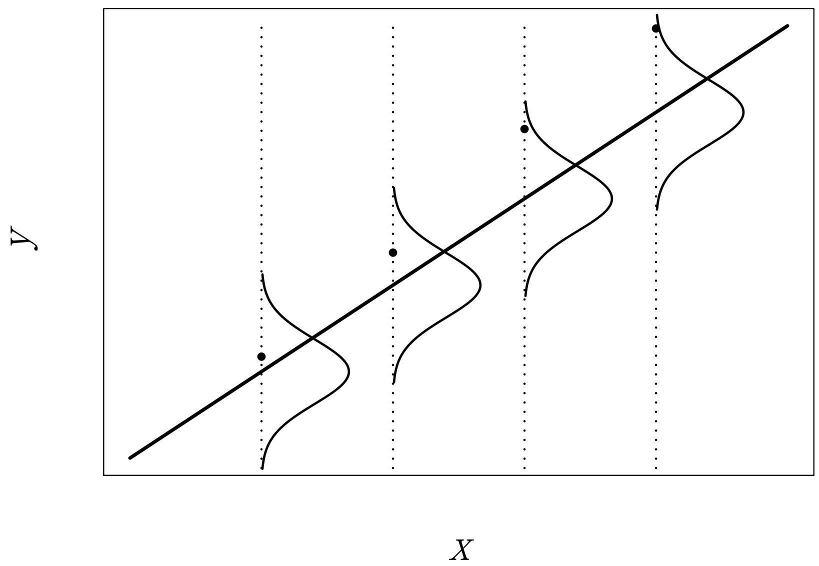
1. データの背後にある規則性，データを発生させる仕組みを関数で表す
   - 関数の型式は決まる／パラメタの値はわからない
2. 観測データと照らし合わせて最も適切なパラメタの決定 (推定)
   - 非線形の最適化
3. モデルの評価，修正

統計モデル

- yが，平均$\beta_0+\beta_1 x$，分散σ^2 (標準偏差σ)の正規分布にしたがう

$$y \simeq \beta_0 + \beta_1 x$$

と近似



統計モデル

モデルの構築

1. データの背後にある規則性，データを発生させる仕組みを関数で表す
- 関数の型式は決まる／パラメタの値はわからない

2. 観測データと照らし合わせて最も適切なパラメタの決定 (推定)
- 非線形の最適化

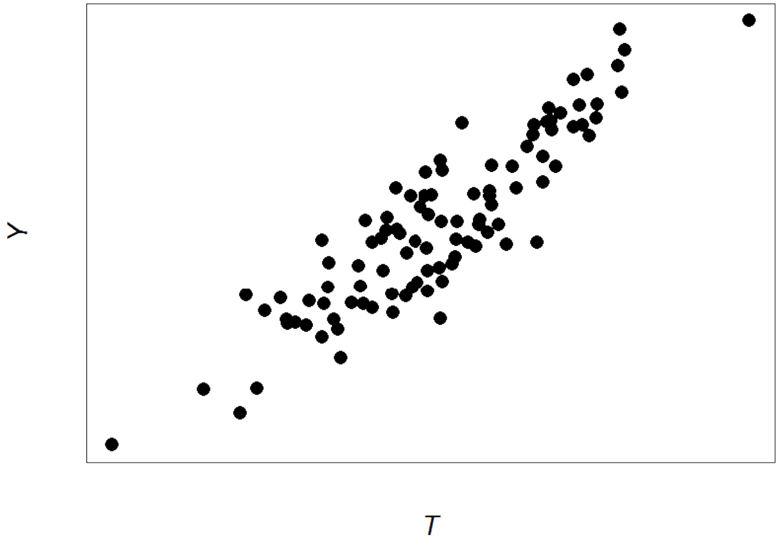
3. モデルの評価，修正

最小二乗法によるパラメタ推定

最小二乗法によるパラメタ推定

線形単回帰モデル

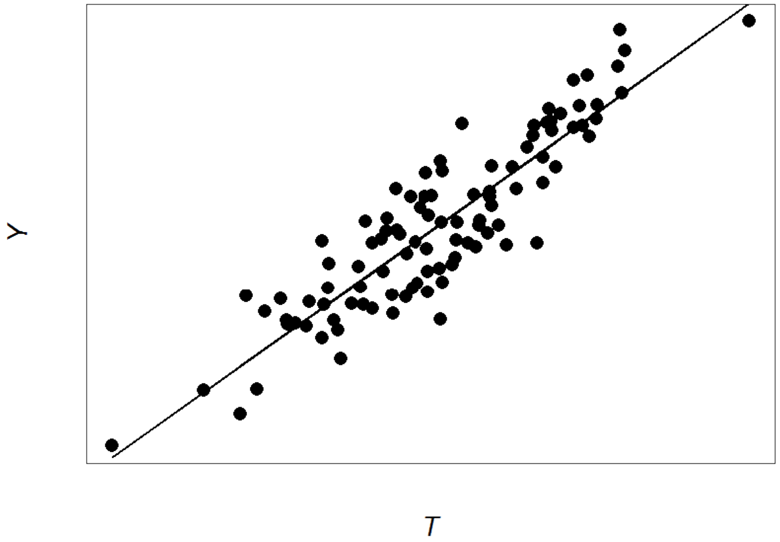
- 1日の最高気温Tとビールの販売量Yの関係



最小二乗法によるパラメタ推定

線形単回帰モデル

- 1日の最高気温Tとビールの販売量Yの関係



最小二乗法によるパラメタ推定

線形単回帰モデル

- 1日の最高気温Tとビールの販売量Yの間に近似的に直線関係が成り立ちそう

$$Y = \beta_0 + \beta_T T$$

- 原因となる変数 ⋯ **説明変数**
- 結果となる変数 ⋯ **目的変数**
- 説明変数と目的変数を結びつけるモデル
  ⋯ **回帰モデル**

最小二乗法によるパラメタ推定

線形単回帰モデル

- 説明変数Tと目的変数Yの間に近似的に

$$Y = \beta_0 + \beta_T T$$

が成り立つと仮定

- n個のYの測定値を$Y_1, Y_2, \cdots, Y_n$,
対応するTの値を各々$T_1, T_2, \cdots, T_n$

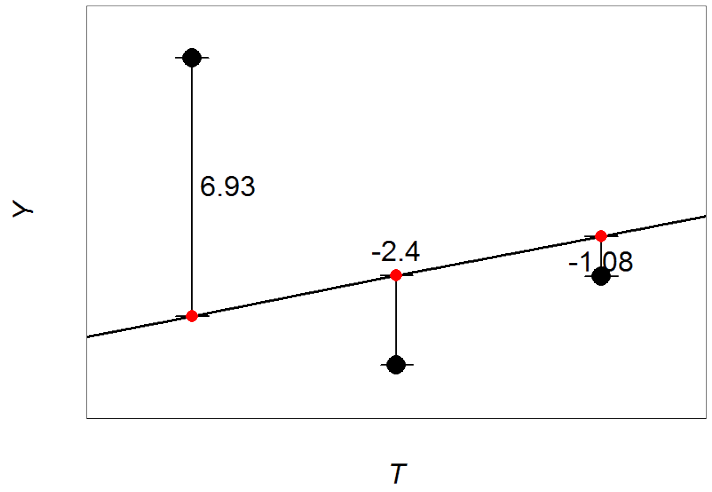
$$Y_i = \beta_0 + \beta_T T_i + \varepsilon_i$$

- ε_iは直線近似した時の誤差（残差）

最小二乗法によるパラメタ推定

線形単回帰モデル

- 各ε_iが0に近いほど直線がよく当てはまっている



最小二乗法によるパラメタ推定

線形単回帰モデル

- 各 ε_i が 0 に近いほど直線がよく当てはまっている

$$J(\beta_0, \beta_T) = \sum_{i=1}^{n} \varepsilon^2 = \sum_{i=1}^{n} \{Y_i - (\beta_0 + \beta_T T_i)\}^2$$

が最小になるようにパラメタ β_0, β_T を決定

⋯ **最小二乗法**

最小二乗法によるパラメタ推定

線形単回帰モデル

$$J(\beta_0, \beta_T) = \sum_{i=1}^{n} \varepsilon^2 = \sum_{i=1}^{n} \{Y_i - (\beta_0 + \beta_T T_i)\}^2$$

を最小とするβ_0, β_Tを決定

- Jが最小になるためには，

$$\nabla J(\beta_0, \beta_T) = \mathbf{0}$$

最小二乗法によるパラメタ推定

線形単回帰モデル

$$\frac{\partial J}{\partial \beta_0} = 2n\beta_0 - 2\sum_{i=1}^{n} Y_i + 2\beta_T \sum_{i=1}^{n} T_i = 0$$

$$\frac{\partial J}{\partial \beta_T} = 2\beta_T \sum_{i=1}^{n} T_i^2 - 2\sum_{i=1}^{n} T_i Y_i + 2\beta_0 \sum_{i=1}^{n} T_i = 0$$

$$\beta_0 = \frac{\sum_{i=1}^{n} T_i^2 \sum_{i=1}^{n} Y_i - \sum_{i=1}^{n} T_i Y_i \sum_{i=1}^{n} T_i}{n\sum_{i=1}^{n} T_i^2 - (\sum_{i=1}^{n} T_i)^2}, \quad \beta_T = \frac{n\sum_{i=1}^{n} T_i Y_i - \sum_{i=1}^{n} T_i \sum_{i=1}^{n} Y_i}{n\sum_{i=1}^{n} T_i^2 - (\sum_{i=1}^{n} T_i)^2}$$

最小二乗法によるパラメタ推定

線形単回帰モデル

$$J(\beta_0, \beta_T) = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} \{Y_i - (\beta_0 + \beta_T T_i)\}^2$$

を最小とする β_0, β_T を決定

最小二乗法によるパラメタ推定

線形重回帰モデル

- 1日の最高気温T，降水量Rとビールの販売量Yの間に近似的に

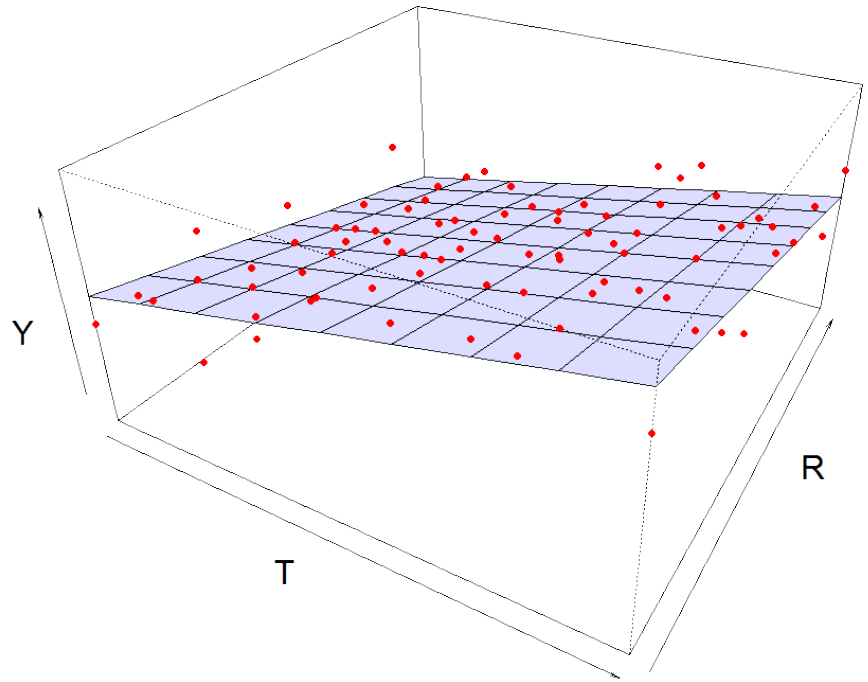
$$Y = \beta_0 + \beta_T T + \beta_R R$$

が成り立つ

最小二乗法によるパラメタ推定

線形重回帰モデル

- 1日の最高気温T, 降水量Rとビールの販売量Yとの関係



最小二乗法によるパラメタ推定

線形重回帰モデル

- p個の説明変数$x_1, x_2, \cdots, x_p$を用いて，近似的に

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p$$

　が成り立つと仮定

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i$$

- 複数の説明変数からなる線形回帰モデル
  $\cdots$ (線形)重回帰モデル

最小二乗法によるパラメタ推定

線形重回帰モデル

- 重回帰モデルのパラメタ $\beta_0, \beta_1, \cdots, \beta_p$ は

$$J(\boldsymbol{\beta}) = \sum_{i=1}^{n} \{y_i - (\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip})\}^2$$

  を最小化する最小二乗法で求められる
- $\nabla J(\boldsymbol{\beta}) = \mathbf{0}$ を解くと，$\beta_0, \beta_1, \cdots, \beta_p$ が求められる

最小二乗法によるパラメタ推定

線形重回帰モデル

$$\boldsymbol{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ & & & \ddots & \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}$$

$$\boldsymbol{y} = [y_1, y_2, \cdots, y_n]^T, \quad \boldsymbol{\beta} = [\beta_0, \beta_1, \cdots, \beta_p]^T$$

$$\boldsymbol{\varepsilon} = [\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n]^T$$

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

最小二乗法によるパラメタ推定

線形重回帰モデル

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

$$J(\boldsymbol{\beta}) = \sum_{i=1}^{n} \varepsilon_i^2 = \boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon} = (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^T(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})$$

$$\nabla J(\boldsymbol{\beta}) = -2\boldsymbol{X}^T(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) = \boldsymbol{0}$$

を解くと

$$\boldsymbol{\beta} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$$

最小二乗法によるパラメタ推定

モデルの評価

$$\boldsymbol{\beta} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$$

- $\boldsymbol{X}$ を構成するベクトルが線形従属
  ＝1つの説明変数が他の説明変数の線形和
  $\cdots$ $\boldsymbol{X}^T\boldsymbol{X}$ が逆行列を持たない
- 線形従属に近い状態 $\cdots$ パラメタが不安定
- 説明変数が冗長
- 分散拡大要因 (variance inflation factor; VIF),
  トレランス (tolerance)

最小二乗法によるパラメタ推定

モデルの評価

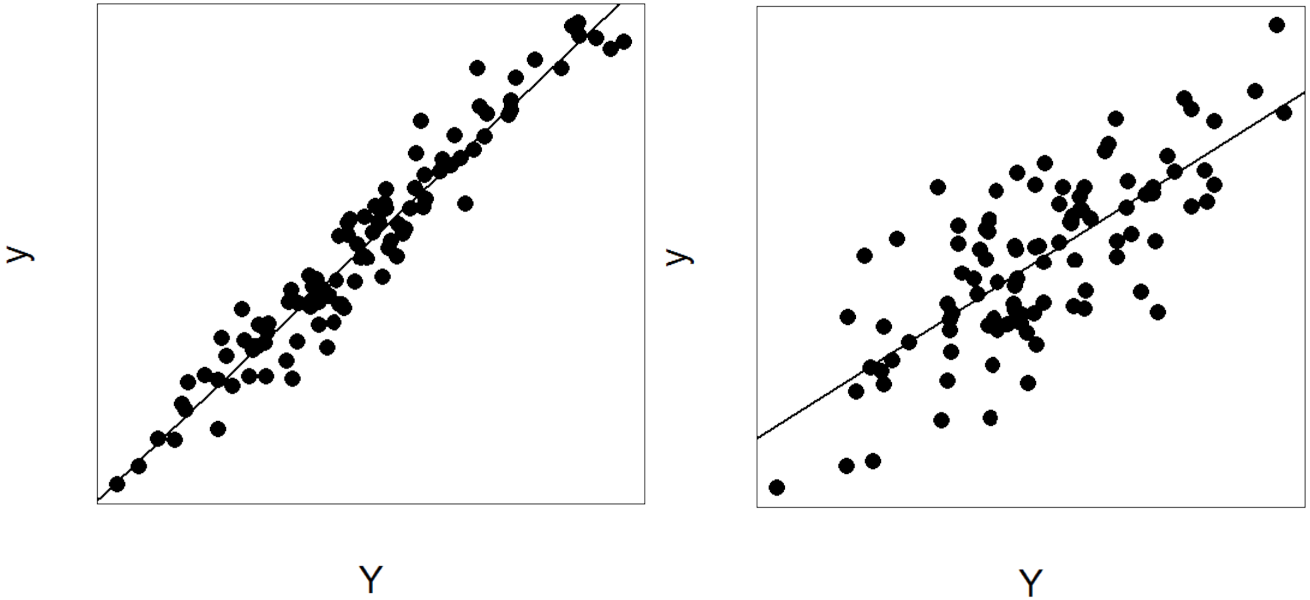
- 測定値$\boldsymbol{y}$と推定値$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta}$の相関係数(重相関係数)

$$R = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}}$$

- $-1 \leq R \leq 1 \cdots$ 増減をともにする程度
- 決定係数$R^2 \cdots y$の変動をモデルで説明できる割合

最小二乗法によるパラメタ推定

モデルの評価



$R^2 = 0.94$ $R^2 = 0.64$

最小二乗法によるパラメタ推定

モデルの評価

- $-1 \leq R \leq 1 \cdots$ 増減をともにする程度
- 決定係数 $R^2 \cdots y$ の変動をモデルで説明できる割合
- 説明変数の数が多いほど R^2 は大きい
  $\cdots$ 説明変数が多過ぎると予測精度低下
- 自由度調整済み決定係数

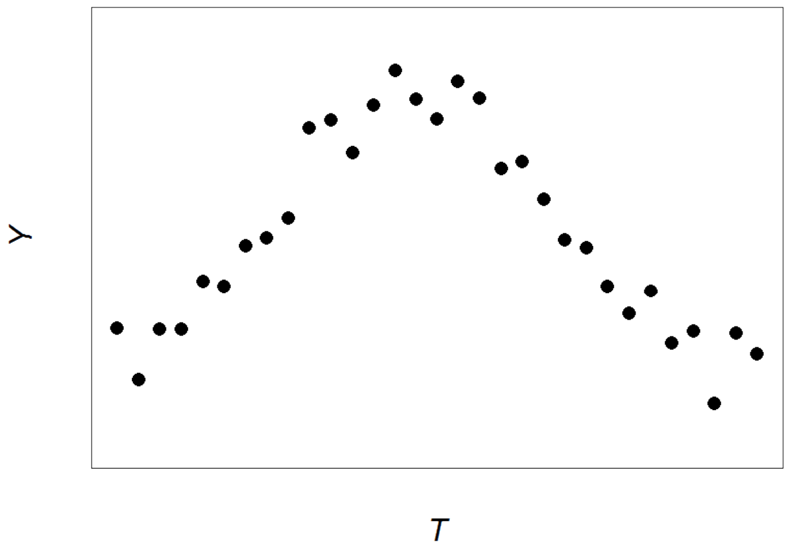
$$\bar{R}^2 = 1 - \frac{n-1}{n-p-1}(1-R^2)$$

説明変数の数をペナルティ

最小二乗法によるパラメタ推定

非線形回帰モデル

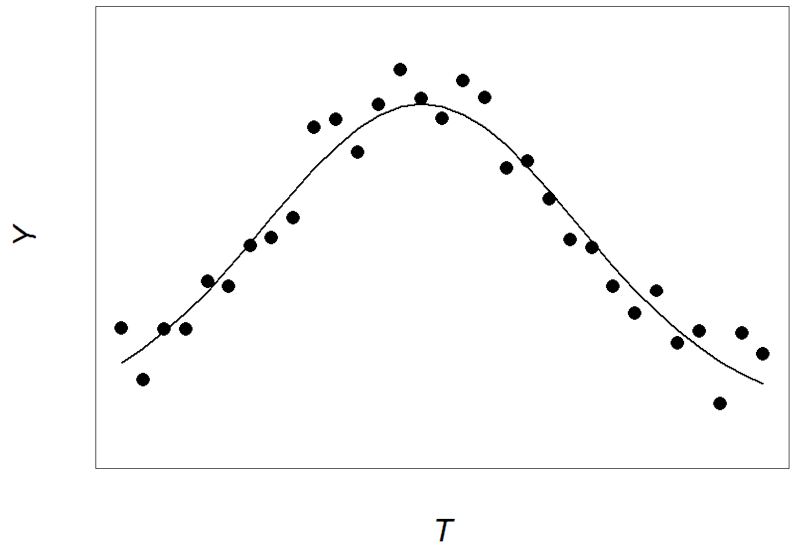
- 1日の最高気温Tとビールの販売量Yの関係



最小二乗法によるパラメタ推定

非線形回帰モデル

- 1日の最高気温Tとビールの販売量Yの関係



最小二乗法によるパラメタ推定

非線形回帰モデル

- Y と T の間に近似的に

$$Y = f(T|\boldsymbol{\beta})$$

  が成り立つと仮定
- $\boldsymbol{\beta}$ はパラメタのベクトル
- n 個の Y の測定値を $Y_1, Y_2, \cdots, Y_n$,
  対応する T の値を各々 $T_1, T_2, \cdots, T_n$

$$Y_i = f(T_i|\boldsymbol{\beta}) + \varepsilon_i$$

最小二乗法によるパラメタ推定

非線形回帰モデル

$$Y_i = f(T_i|\boldsymbol{\beta}) + \varepsilon_i$$

- パラメタ $\boldsymbol{\beta}$ は線形回帰モデルの場合と同様に

$$J(\boldsymbol{\beta}) = \sum_{i=1}^{n} \{Y_i - f(T_i|\boldsymbol{\beta})\}^2$$

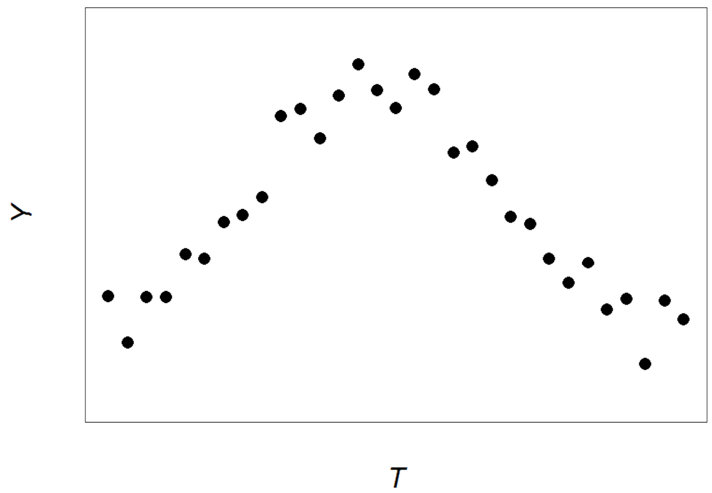
  を最小化する最小二乗法で求めることができる
- 一般にはニュートン法などの非線形最適化法を用いて数値的に解を求める

最小二乗法によるパラメタ推定

非線形回帰モデル

$$Y = f(T|\boldsymbol{\beta}) = \beta_1 \exp\left(-\frac{(T-\beta_2)^2}{\beta_3^2}\right)$$

$$Y_i = \beta_1 \exp\left(-\frac{(T_i-\beta_2)^2}{\beta_3^2}\right) + \varepsilon_i$$



最小二乗法によるパラメタ推定

$$J(\boldsymbol{\beta}) = \sum_{i=1}^{n} \varepsilon^2 = \sum_{i=1}^{n} \left\{ Y_i - \beta_1 \exp\left(-\frac{(T_i - \beta_2)^2}{\beta_3^2} \right) \right\}^2$$

$$\nabla J(\boldsymbol{\beta}) = \begin{bmatrix} \frac{\partial J}{\partial \beta_1} & \frac{\partial J}{\partial \beta_2} & \frac{\partial J}{\partial \beta_3} \end{bmatrix}^T$$

$$\nabla^2 J(\boldsymbol{\beta}) = \begin{bmatrix} \frac{\partial^2 J}{\partial \beta_1^2} & \frac{\partial^2 J}{\partial \beta_1 \partial \beta_2} & \frac{\partial^2 J}{\partial \beta_1 \partial \beta_3} \\ \frac{\partial^2 J}{\partial \beta_2 \partial \beta_1} & \frac{\partial^2 J}{\partial \beta_2^2} & \frac{\partial^2 J}{\partial \beta_2 \partial \beta_3} \\ \frac{\partial^2 J}{\partial \beta_3 \partial \beta_1} & \frac{\partial^2 J}{\partial \beta_3 \partial \beta_2} & \frac{\partial^2 J}{\partial \beta_3^2} \end{bmatrix}$$

最小二乗法によるパラメタ推定

非線形回帰モデル

- $\nabla J(\boldsymbol{\beta})$ の計算

$$\frac{\partial}{\partial \beta_1} J(\beta_1, \beta_2, \beta_3) = \lim_{\Delta \to 0} \frac{J(\beta_1 + \Delta, \beta_2, \beta_3) - J(\beta_1, \beta_2, \beta_3)}{\Delta}$$

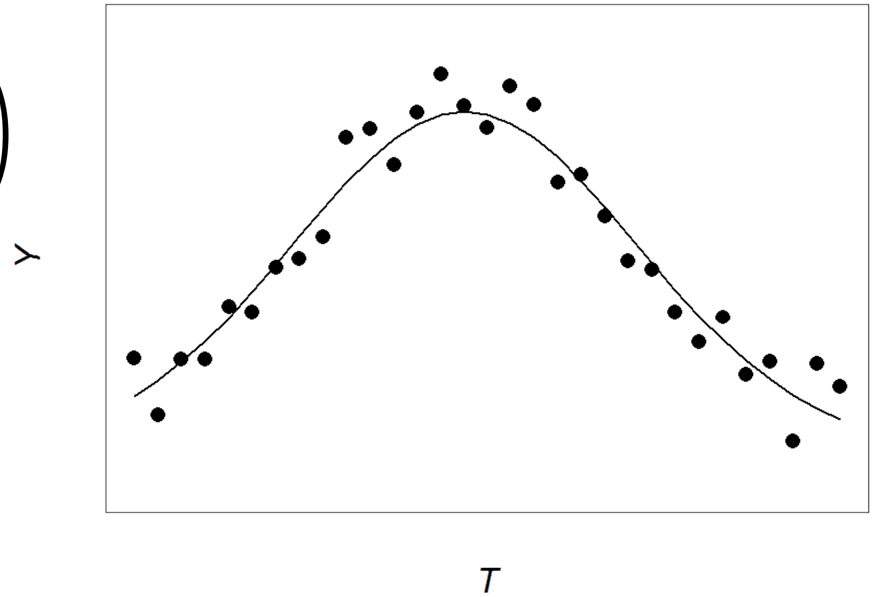
- Δ を 0 に近い数として差分近似

$$\frac{\partial}{\partial \beta_1} J(\beta_1, \beta_2, \beta_3) \simeq \frac{J(\beta_1 + \Delta, \beta_2, \beta_3) - J(\beta_1, \beta_2, \beta_3)}{\Delta}$$

最小二乗法によるパラメタ推定

非線形回帰モデル

$$y = \beta_1 \exp\left(-\frac{(T-\beta_2)^2}{\beta_3^2}\right)$$



最尤推定法によるパラメタ推定

最尤推定法によるパラメタ推定

ゆうど
尤度 (尤度関数)

- ある統計モデル $f(y|\boldsymbol{\theta})$ を仮定しているときに，特定の観測データが得られる確率 (確率密度)

$$p(y_1, y_2, \cdots, y_n|\boldsymbol{\theta}) = \prod_{i=1}^{n} f(y_i|\boldsymbol{\theta})$$

- パラメタ $\boldsymbol{\theta}$ の関数

$$L(\boldsymbol{\theta}|y_1, y_2, \cdots, y_n) = \prod_{i=1}^{n} f(y_i|\boldsymbol{\theta})$$

最尤推定法によるパラメタ推定

最尤推定法

- 尤度（尤度関数）

$$L(\boldsymbol{\theta}|y_1, y_2, \cdots, y_n) = \prod_{i=1}^{n} f(y_i|\boldsymbol{\theta})$$

- 観測データに統計モデルが「どれぐらいあてはまっているか」を表す尺度
- 最尤推定法では，観測データのもとで尤度が最大になるようにパラメタ値を決定

最尤推定法によるパラメタ推定

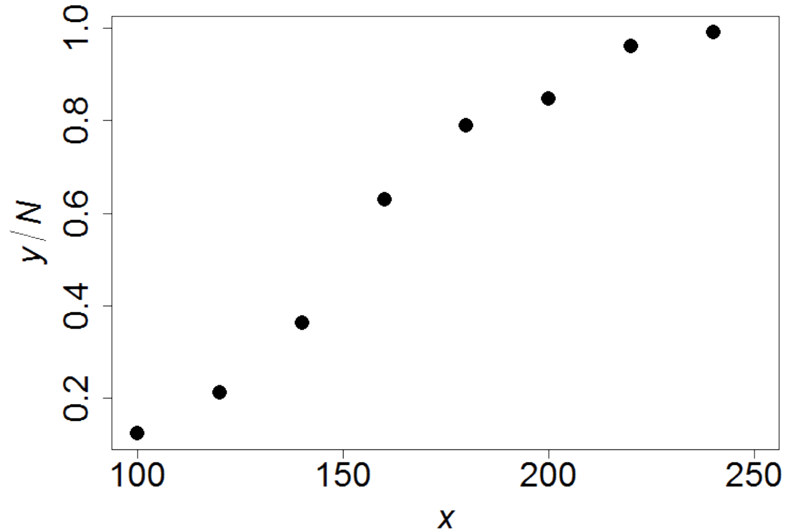
- 材料は温度を上げると化学的特性が変化
- 加熱温度 x (°C)，サンプル数 N (個)，そのうち特性が変化したサンプル数 y (個)
- 加熱温度と特性が変化する確率の関係を求めよ

i	1	2	3	4	5	6	7	8
x_i	100	120	140	160	180	200	220	240
N_i	96	99	99	97	100	98	99	100
y_i	12	21	36	61	79	83	95	99

最尤推定法によるパラメタ推定

ロジスティック回帰モデル

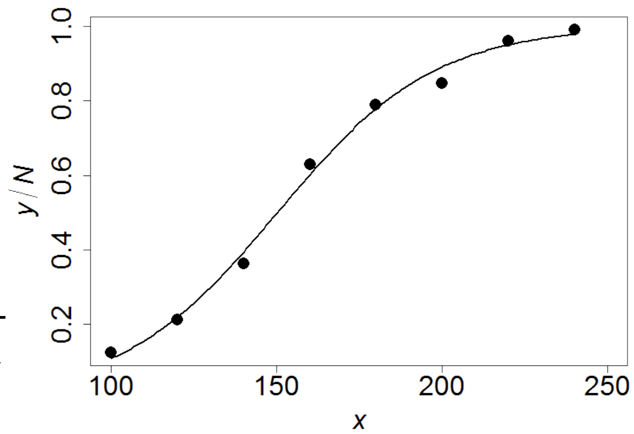
- 加熱温度 x (°C)
- サンプル数 N (個)
- 特性が変化したサンプル数 y (個)



最尤推定法によるパラメタ推定

ロジスティック回帰モデル

$$p(x|\beta_0, \beta_1) = \frac{\exp(\beta_0 + \beta_1 x)}{1 + \exp(\beta_0 + \beta_1 x)} = \frac{1}{1 + \exp\{-(\beta_0 + \beta_1 x)\}}$$



最尤推定法によるパラメタ推定

ロジスティック回帰モデル

- 加熱温度がx_i (°C)のときにN_i個のサンプル中y_i個の特性が変化する確率は二項分布にしたがう

$$P(y_i|x_i, Ni, P_i) = {}_{N_i}\mathrm{C}_{y_i} P_i^{y_i}(1-P_i)^{N_i-y}$$
$$P_i = p(x_i|\beta_0, \beta_1) = \frac{1}{1+\exp\{-(\beta_0+\beta_1 x)\}}$$

- N_i個の中からy_i個を選ぶ組み合わせの数

$$ {}_{N_i}\mathrm{C}_{y_i} = \begin{pmatrix} N_i \\ y_i \end{pmatrix} = \frac{N_i!}{(N_i-y_i)!y_i!}$$

最尤推定法によるパラメタ推定

ロジスティック回帰モデル

- $i = 1, 2, \cdots, n \quad (n = 8)$ において，
  加熱温度が x_i (°C) のときに N_i 個のサンプル中
  y_i 個の特性が変化する確率

$$f(y_1, y_2, \cdots, y_n | \beta_0, \beta_1) = \prod_{i=1}^{n} {}_{N_i}\mathrm{C}_{y_i} P_i^{y_i} (1 - P_i)^{N_i - y_i}$$

- $f(y_1, y_2, \cdots, y_n | \beta_0, \beta_1)$ をパラメタ β_0, β_1 の関数
  $L(\beta_0, \beta_1) \cdots$ 尤度

最尤推定法によるパラメタ推定

ロジスティック回帰モデル

- 対数尤度関数

$$
\begin{aligned}
\log L(\beta_0, \beta_1) = & \sum_{i=1}^{n} \log {}_{N_i}\mathrm{C}_{y_i} + \sum_{i=1}^{n} (y_i \log P_i) \\
& + \sum_{i=1}^{n} \{(N_i - y_i) \log(1 - P_i)\}
\end{aligned}
$$

最尤推定法によるパラメタ推定

正規分布モデルと最小二乗法

- 回帰モデル $y = f(\boldsymbol{x}|\boldsymbol{\beta})$ を当てはめるとき，
  y が平均 $f(\boldsymbol{x}|\boldsymbol{\beta})$，分散 σ^2 の正規分布にしたがう

$$g(y|\boldsymbol{x}, \boldsymbol{\beta}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{\{y - f(\boldsymbol{x}|\boldsymbol{\beta})\}^2}{2\sigma^2}\right]$$

- 最尤推定法によるパラメタの推定値と
  最小二乗法によるパラメタの推定値は一致する

最尤推定法によるパラメタ推定

正規分布モデルと最小二乗法

$$
\begin{aligned}
L(\boldsymbol{\beta}) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{\{y_i - f(\boldsymbol{x}_i|\boldsymbol{\beta})\}^2}{2\sigma^2}\right] \\
&= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left[-\frac{\boxed{\sum_{i=1}^{n}\{y_i - f(\boldsymbol{x}_i|\boldsymbol{\beta})\}^2}}{2\sigma^2}\right]
\end{aligned}
$$

$$
J = \sum_{i=1}^{n} \varepsilon_i^2 = \boxed{\sum_{i=1}^{n}\{y_i - f(\boldsymbol{x}_i|\boldsymbol{\beta})\}^2}
$$

最尤推定法によるパラメタ推定

モデル選択

- パラメタ数が多いほど良いモデルという訳ではない
- 予測精度の高いモデル，本質を捉えたモデル
- 様々なモデル選択基準
- 赤池の情報量基準 (AIC)

$$AIC(k) = -2\text{最大対数尤度} + 2k$$

k はパラメタ数
AICが小さいほど良いモデル

最尤推定法によるパラメタ推定

- コンピュータを内蔵した機器同士がネットワークにより接続され，自律的に連係動作することにより，人の生活を支援する技術と環境
  $\cdots$ **ユビキタスネットワーク**，
  IoT (Internet of Things)
- 多様で大量のデータ
- 役に立つ変数はほんの一部
- モデル選択，正則化