線型推測論

第02回 一般線型モデル

千葉大学大学院医学研究院
長島 健悟

一般線型モデル

- 観測値がパラメータの線型式と誤差の和で表わされるモデル

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

  - $\mathbf{Y}$: $n \times 1$次元の観測値確率変数ベクトル
  - $\mathbf{X}$: $n \times p$次元の定数行列
  - $\boldsymbol{\beta}$: $p \times 1$次元の未知パラメータベクトル
  - $\boldsymbol{\epsilon}$: $n \times 1$次元の平均が0の誤差確率変数ベクトル

一般線型モデル

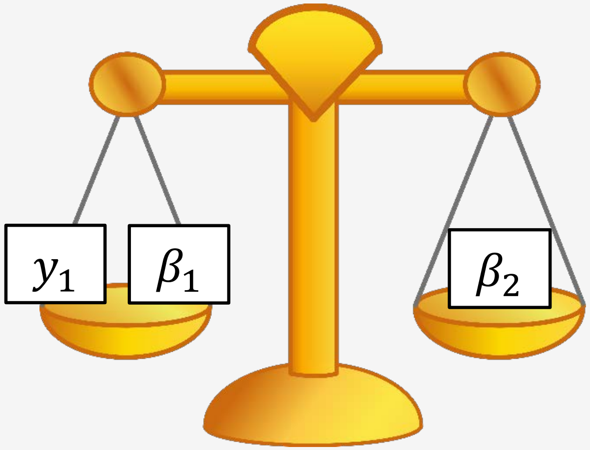
- 基本はこの形

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{11} & \cdots & x_{1p} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

- 最初は要素ごとの表記に慣れておき
- 最終的には行列ベクトルでの式変形を考えられる様になることを目標に

秤量問題

- 未知の重さの物体 (β_1, β_2) がある
- 既知の重さy_1の分銅とβ_1を秤の左側に, β_2を右側において釣り合ったとする
- 秤には計測誤差があるとする
- これを数式であらわすと？



$$y_1 = -\beta_1 + \beta_2 + \epsilon_1$$

秤量問題

- 左にy_2, 右にβ_1とβ_2

$$y_2 = \beta_1 + \beta_2 + \epsilon_2$$

- 左にy_3とβ_2, 右にβ_1

$$y_3 = \beta_1 - \beta_2 + \epsilon_3$$

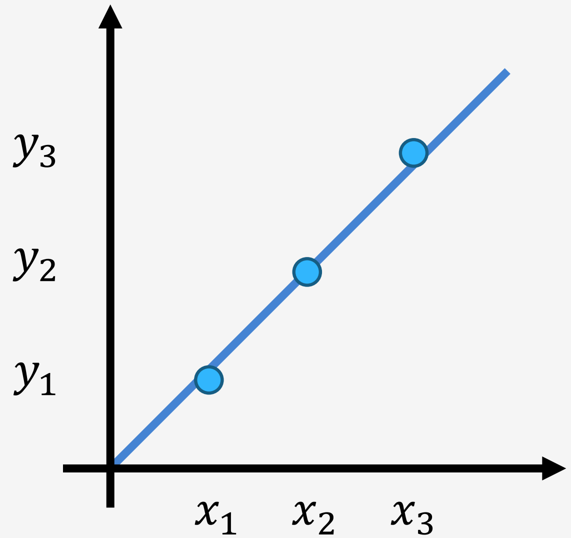
- この問題は線型モデルで定式化できそう

練習問題

- このモデルを行列で表現すると？
  - $y_1 = -\beta_1 + \beta_2 + \epsilon_1$
  - $y_2 = \beta_1 + \beta_2 + \epsilon_2$
  - $y_3 = \beta_1 - \beta_2 + \epsilon_3$
- $\begin{pmatrix} \\ \\ \\ \end{pmatrix} = \begin{pmatrix} & & \\ & & \\ & & \end{pmatrix} \begin{pmatrix} \\ \\ \end{pmatrix} + \begin{pmatrix} \\ \\ \\ \end{pmatrix}$

検量線の問題

- 液体の濃度が変わると吸光度(光の吸収量)が変化するという性質を用いて, 濃度と吸光度の関係式を作る
- この関係式を使えば未知試料の濃度を調べることができる



検量線の問題

- 濃度と測定値の組を(x_i, y_i)としたとき, どんなモデルが考えられるでしょうか
  - 濃度0の吸光度は0、$i = 1, \ldots, 5$とします
  - $y_1 =$
  - $y_2 =$
  - $y_3 =$
  - $y_4 =$
  - $y_5 =$

練習問題

- 検量線のモデルを行列で表現すると？
- $$\begin{pmatrix} \\ \\ \\ \end{pmatrix} = \begin{pmatrix} \\ \\ \\ \end{pmatrix} \left(\quad \right) + \begin{pmatrix} \\ \\ \\ \end{pmatrix}$$

一般線型モデル

- 観測値$\mathbf{Y}$から未知パラメータ$\boldsymbol{\beta}$に対する推測を行いたい際に用いる事ができそう

- $\mathbf{X}$: は適用したい問題によって異なる定数行列であることが分かる
- 解きたい問題に応じて, 適切に$\mathbf{X}$を選択・設定できなければならない

一般線型モデルの例

- 一般線型モデルの例
  - t 検定
  - 単回帰モデル
  - 重回帰モデル
  - 分散分析モデル
  - 共分散分析モデル
  - ・・・

線型モデルの統一表現

- 何がうれしいか？
  - 仮説検定の方法
  - 推定の方法
  - 計算アルゴリズム
  - 実際のデータに対するモデルの構成方法の考え方

これらを全て同じように扱うことができる

- さらに, かなり多くの統計モデルは線型モデルの一般化になっている

回帰モデル

- 単回帰

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

- 重回帰

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \epsilon_i$$

一般線型モデルの境目

- パラメータに対して線型であれば良いので以下も一般線型モデル

$$y_i = \beta_0 - \beta_1 \sin x_{1i} + \beta_2 x_{2i}^2 + \epsilon_i$$

- パラメータに対して線型でない関数がかかるものは一般線型モデルに含まれない

$$y_i = \beta_0 - \exp(\beta_1)\, x_{1i} + \log \beta_2\, x_{2i}^2 + \epsilon_i$$

$$\log(y_i) = \beta_0 + \beta_1 x_i + \epsilon_i \Leftrightarrow$$
$$y_i = \exp(\beta_0 + \beta_1 x_i + \epsilon_i)$$

行列表現

- 何度も登場済みだが, 一般線型モデルであれば全部この形

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{11} & \cdots & x_{1p} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

- 具体例を通じて慣れていこう

練習問題：回帰モデル

- $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \epsilon_i$
  - ※$\beta_0, \dots, \beta_p$まで, β_0に注意

$$\begin{pmatrix} \\ \\ \vdots \\ \\ \end{pmatrix} = \begin{pmatrix} & & & \cdots & \\ & & & \cdots & \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ & & & \cdots & \end{pmatrix} \begin{pmatrix} \\ \\ \\ \vdots \\ \\ \end{pmatrix} + \begin{pmatrix} \\ \\ \vdots \\ \\ \end{pmatrix}$$

一元配置分散分析モデル

- 幾つかの集団 or 条件下での平均値の比較のモデル
  - $y_{ij} = \mu_i + \epsilon_{ij}$ または
  - $y_{ij} = \mu + \alpha_i + \epsilon_{ij}$

一元配置分散分析モデル

- $y_{ij} = \mu + \alpha_i + \epsilon_{ij}$
  - ※$i = 1, \dots, 3, j = 1, \dots, n$ の場合を考える

$$\begin{pmatrix} y_{11} \\ \vdots \\ y_{1n} \\ y_{21} \\ \vdots \\ y_{2n} \\ y_{31} \\ \vdots \\ y_{3n} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} + \begin{pmatrix} \epsilon_{11} \\ \vdots \\ \epsilon_{1n} \\ \epsilon_{21} \\ \vdots \\ \epsilon_{2n} \\ \epsilon_{31} \\ \vdots \\ \epsilon_{3n} \end{pmatrix}$$

練習問題：一元配置分散分析

- $y_{ij} = \mu_i + \epsilon_{ij}$ではどうなるでしょう？
  - ※$i = 1, \dots, 3, j = 1, \dots, n$ の場合を考える

$$\begin{pmatrix} \\ \\ \\ \\ \\ \\ \end{pmatrix} = \begin{pmatrix} & & \\ & & \\ & & \\ & & \\ & & \\ & & \end{pmatrix} \begin{pmatrix} \\ \\ \end{pmatrix} + \begin{pmatrix} \\ \\ \\ \\ \\ \\ \end{pmatrix}$$

練習問題：二元配置分散分析

- $y_{ijk} = \alpha_i + \beta_j + \epsilon_{ijk}$ではどうなるでしょう？
  - ※$i = 1,2, j = 1,2, k = 1,2$ の場合を考える

$$\begin{pmatrix} \\ \\ \\ \\ \\ \\ \\ \end{pmatrix} = \begin{pmatrix} & & & & & & \\ \\ \\ \\ \\ \\ \\ \end{pmatrix} \begin{pmatrix} \\ \\ \\ \\ \end{pmatrix} + \begin{pmatrix} \\ \\ \\ \\ \\ \\ \\ \end{pmatrix}$$

線型モデルの統一表現

- 上述で紹介した全てのモデルは$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$の形に表すことができた
- 観測値$\mathbf{Y}$から未知パラメータ$\boldsymbol{\beta}$に対する推測を行いたい際に利用できるのであった
- $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$の形のままで$\boldsymbol{\beta}$を推測する手順があれば、全てのモデルを同じように取り扱うことができる