K変数回帰分析、最小二乗法†

担当:　長倉　大輔
（ながくらだいすけ）

† この資料は私の講義およびゼミにおいて使用するために作成した資料です。WEBページ上で公開しており、自由に参照して頂いて構いません。ただし、内容について、一応検証してありますが、もし誤植、間違い、等があった場合でもそれによって生じるいかなる損害、不利益について責任を負いかねますのでご了承ください。誤植、間違いは発見次第、継続的に直していますが、常に存在する可能性があります。

線形回帰分析

- 説明変数が任意の数の場合

被説明変数は１つの説明変数だけではよく説明できない場合がほとんどです。例えば体重は、身長だけではなく、年齢や、カロリー摂取量など他の様々な要因にも依存しても決まってくるでしょう。

先週の回帰モデルを説明変数がたくさんある場合 (**多変量分析**という)に拡張する事を考えます。

線形回帰分析

- 多変量回帰モデル

被説明変数 Y_i $(i=1,\ldots,n)$ が k 個の説明変数 $1, X_{2i}\ldots,$ X_{ki} $(i=1,\ldots,n)$によって次のように表す事ができると仮定する (定数項の説明変数として 1 を含んでいる事に注意)

$$Y_i = \beta_1 + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + \varepsilon_i \quad (i=1,\ldots,n)$$

ここで ε_i は被説明変数の説明変数だけでは説明できない部分をまとめたものである。

線形回帰分析

- 多変量回帰モデルの行列表現

ここで

$$\mathbf{Y}=\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix},\quad \mathbf{X}=\begin{bmatrix} 1 & X_{21} & \dots & X_{k1} \\ 1 & X_{22} & \dots & X_{k2} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & X_{2n} & \dots & X_{kn} \end{bmatrix},\quad \mathbf{e}=\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix},\quad \boldsymbol{\beta}=\begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix},$$

とすると

$$\begin{aligned} Y_1 &= \beta_1 + \beta_2 X_{21} + \dots + \beta_k X_{k1} + \varepsilon_1 \\ Y_2 &= \beta_1 + \beta_2 X_{22} + \dots + \beta_k X_{k2} + \varepsilon_2 \\ &\vdots \\ Y_n &= \beta_1 + \beta_2 X_{2n} + \dots + \beta_k X_{k2} + \varepsilon_n \end{aligned} \quad \Rightarrow \quad \mathbf{Y}=\mathbf{X}\boldsymbol{\beta}+\mathbf{e}$$

のように簡潔に表現できる。

線形回帰分析

- 単回帰モデルの行列表現

単回帰モデルは

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix}, \quad \mathbf{e} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \alpha \\ \beta \end{bmatrix},$$

とすると

$$\begin{aligned} Y_1 &= \alpha + \beta X_1 + \varepsilon_1 \\ Y_2 &= \alpha + \beta X_2 + \varepsilon_2 \\ &\vdots \\ Y_n &= \alpha + \beta X_n + \varepsilon_n \end{aligned} \qquad \Rightarrow \qquad \mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$$

のように表現できる。

線形回帰分析

- 残差平方和の行列表現

$\mathbf{b} = (a, b)^{\mathrm{T}}$ とすると、残和平方和の行列表現は

$$SSR(a,b) = \sum_{i=1}^{n} (Y_i - a - bX_i)^2 = (\mathbf{Y} - \mathbf{Xb})^{\mathrm{T}} (\mathbf{Y} - \mathbf{Xb})$$

となる。行列に関する微分を知っていれば、このSSRの $\mathbf{b}$ に関する微分は

$$\frac{\partial SSR(a,b)}{\partial \mathbf{b}} = \begin{bmatrix} \dfrac{\partial SSR(a,b)}{\partial a} \\ \dfrac{\partial SSR(a,b)}{\partial b} \end{bmatrix} = -2\mathbf{X}^{\mathrm{T}}\mathbf{Y} + 2\mathbf{X}^{\mathrm{T}}\mathbf{Xb}$$

となる。

線形回帰分析

- 最小二乗推定量の行列表現

この SSR の $\mathbf{b}$ に関する微分を $\mathbf{0}$ とおいたもの、すなわち

$$-2\mathbf{X}^{\mathrm{T}}\mathbf{Y} + 2\mathbf{X}^{\mathrm{T}}\mathbf{X}\mathbf{b} = \mathbf{0}$$

は 1 階の条件の行列表現に他ならない。

これを $\mathbf{b}$ について解くと

$$\mathbf{b} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Y}$$

という最小二乗推定量の行列表現が得られる。

線形回帰分析

- 最小二乗推定量の行列表現

この行列表現を実際に計算してみると

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Y}$$

$$= \left(\begin{bmatrix} 1 & \cdots & 1 \\ X_1 & \cdots & X_n \end{bmatrix} \begin{bmatrix} 1 & X_1 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & \cdots & 1 \\ X_1 & \cdots & X_n \end{bmatrix} \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}$$

$$= \frac{1}{n\sum_{i=1}^{n} X_i^2 - (\sum_{i=1}^{n} X_i)^2} \begin{bmatrix} (\sum_{i=1}^{n} X_i^2)(\sum_{i=1}^{n} Y_i) - (\sum_{i=1}^{n} X_i)(\sum_{i=1}^{n} X_i Y_i) \\ -(\sum_{i=1}^{n} X_i)\sum_{i=1}^{n} Y_i + n\sum_{i=1}^{n} X_i Y_i \end{bmatrix}$$

となる。

線形回帰分析

- 多変量回帰モデルの最小二乗推定量

行列表現:

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

において、$\mathbf{b} = (b_1, \ldots, b_k)^{\mathrm{T}}$ とすると

$$SSR(b_1, b_2, \ldots, b_k) = \sum_{i=1}^{n} (Y_i - b_1 - b_2 X_{2i} - \ldots - b_k X_{ki})^2$$

$$= (\mathbf{Y} - \mathbf{Xb})^{\mathrm{T}} (\mathbf{Y} - \mathbf{Xb})$$

を最小にする $\mathbf{b}$ は、まったく同じ議論により

$$\mathbf{b} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Y}$$

となる。これが $\boldsymbol{\beta}$ の最小二乗推定量、$\hat{\boldsymbol{\beta}}$ である。

線形回帰分析

- 誤差項 に関する仮定

線形回帰モデルにおいて誤差項 ε_i は確率変数として扱われる。以下のような仮定をおく。

(A1) $E(\varepsilon_i) = 0, \ \ i = 1, \ldots, n.$

(A2) (**無相関の仮定**) $\mathrm{Cov}(\varepsilon_i, \varepsilon_j) = 0, \ \ i \neq j, \ i, j = 1, \ldots, n.$

(A3) (**均一分散の仮定**) $\mathrm{Var}\,(\varepsilon_i) = \sigma^2, \quad i, j = 1, \ldots, n.$

(A4) $X_{ji}, \ j = 2, \ldots, k, \ i = 1, \ldots, n$ は非確率変数で

$\mathrm{rank}(\mathbf{X}) = k < n, \ \ \lim_{n\to\infty} \dfrac{\mathbf{X}^{\mathrm{T}}\mathbf{X}}{n} = \mathbf{Q}$ となる正則行列 $\mathbf{Q}$

が存在する。

線形回帰分析

- 誤差項に関する仮定

仮定(A1)は

$$E(\boldsymbol{\varepsilon}) = \mathbf{0}$$

仮定(A2), (A3) は、まとめて

$$\begin{aligned}\mathrm{var}(\boldsymbol{\varepsilon}) &= E([\boldsymbol{\varepsilon} - E(\boldsymbol{\varepsilon})]\,[\boldsymbol{\varepsilon} - E(\boldsymbol{\varepsilon})]^{\mathrm{T}}) \\ &= E(\boldsymbol{\varepsilon}\,\boldsymbol{\varepsilon}^{\mathrm{T}}) \\ &= \sigma^2\mathbf{I}_n \qquad (\mathbf{I}_n \text{ は } n \times n \text{ の単位行列})\end{aligned}$$

のように表すこともできる。

線形回帰分析

- 誤差項に関する仮定

最後の等式は

$$E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}}) = E\left(\begin{bmatrix}\varepsilon_1\\ \varepsilon_2\\ \vdots\\ \varepsilon_n\end{bmatrix}\begin{bmatrix}\varepsilon_1 & \varepsilon_2 & \dots & \varepsilon_n\end{bmatrix}\right) = E\left(\begin{bmatrix}\varepsilon_1^2 & \varepsilon_1\varepsilon_2 & \dots & \varepsilon_1\varepsilon_n\\ \varepsilon_1\varepsilon_2 & \varepsilon_2^2 & \dots & \varepsilon_2\varepsilon_n\\ \vdots & \vdots & \dots & \vdots\\ \varepsilon_1\varepsilon_n & \varepsilon_2\varepsilon_n & \dots & \varepsilon_n^2\end{bmatrix}\right)$$

$$= \begin{bmatrix}E(\varepsilon_1^2) & E(\varepsilon_1\varepsilon_2) & \dots & E(\varepsilon_1\varepsilon_n)\\ E(\varepsilon_2\varepsilon_1) & E(\varepsilon_2^2) & \dots & E(\varepsilon_2\varepsilon_n)\\ \vdots & \vdots & \dots & \vdots\\ E(\varepsilon_n\varepsilon_1) & E(\varepsilon_n\varepsilon_2) & \dots & E(\varepsilon_n^2)\end{bmatrix} = \begin{bmatrix}\sigma^2 & 0 & \dots & 0\\ 0 & \sigma^2 & \dots & 0\\ \vdots & \vdots & \dots & \vdots\\ 0 & 0 & \dots & \sigma^2\end{bmatrix}$$

$$= \sigma^2\mathbf{I}_n$$

より求まる。このように成分が確率変数であるベクトルのそれぞれの成分に対する分散と共分散を並べた行列を**分散共分散行列**とよぶ。

線形回帰分析

- 最小二乗推定量の性質

以上の仮定の下で、$\boldsymbol{\beta}$ の推定量である $\hat{\boldsymbol{\beta}}$ は以下のような性質を持っている。

(P1) $\hat{\boldsymbol{\beta}}$ は $\boldsymbol{\beta}$ の**不偏推定量**である。

(P2) $\hat{\boldsymbol{\beta}}$ は $\boldsymbol{\beta}$ の**一致推定量**である。

(P3) $\hat{\boldsymbol{\beta}}$ は $\boldsymbol{\beta}$ の**最良線形不偏推定量**である。

順に見て行く。

線形回帰分析

- 最小二乗推定量の不偏性

まず $\hat{\boldsymbol{\beta}}$ を以下のように書き直す。

$$\begin{aligned}\hat{\boldsymbol{\beta}} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Y} &= (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}(\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon}) \\ &= (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{X}\boldsymbol{\beta} + (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\boldsymbol{\varepsilon} \\ &= \boldsymbol{\beta} + (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\boldsymbol{\varepsilon}\end{aligned}$$

仮定(4)より $\mathbf{X}$ は確率変数ではないので 、

$$E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta} + (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\,E(\boldsymbol{\varepsilon})$$

よって、仮定(1)より ($E(\boldsymbol{\varepsilon}) = \mathbf{0}$)

$$E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$$

となり、$\hat{\boldsymbol{\beta}}$ は $\boldsymbol{\beta}$ の**不偏推定量である**事がわかる。

線形回帰分析

- 最小二乗推定量の一致性

一致性とは推定量が確率収束する事であった。

$\hat{\beta}$ は不偏推定量なので、あとはその分散がゼロに収束する事を示せばよい。

よってまず $\hat{\beta}$ の分散共分散行列を計算する。

線形回帰分析

- $\hat{\boldsymbol{\beta}}$ の分散共分散行列

$\hat{\boldsymbol{\beta}}$ の分散共分散行列は

$$
\begin{aligned}
\mathrm{var}(\hat{\boldsymbol{\beta}}) &= E[(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})^{\mathrm{T}}] \\
&= E[(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}}\mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}] \\
&= (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}})\mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1} \\
&= \sigma^2(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1} \\
&= \sigma^2(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}
\end{aligned}
$$

となる。

線形回帰分析

- 最小二乗推定量の一致性

この分散共分散行列が n が大きくなると $\mathbf{0}$ に収束する事を示せばよい。

仮定(A4) より

$$
\begin{aligned}
\lim_{n\to\infty} \mathrm{var}(\hat{\boldsymbol{\beta}}) &= \lim_{n\to\infty} \sigma^2 (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1} \\
&= \lim_{n\to\infty} \frac{\sigma^2}{n} \left(\frac{\mathbf{X}^{\mathrm{T}}\mathbf{X}}{n} \right)^{-1} \\
&= 0 \cdot \mathbf{Q}^{-1} \\
&= \mathbf{0}
\end{aligned}
$$

となる。よって $\hat{\boldsymbol{\beta}}$ は $\boldsymbol{\beta}$ の**一致推定量である**事がわかる。

線形回帰分析

- 最小二乗推定量の最適性

最小二乗推定量 $\hat{\boldsymbol{\beta}}$ は $\boldsymbol{\beta}$ の最良線形不偏推定量
(BLUE: Best Linear Unbiased Estimator)
であるとは全ての不偏な($\mathbf{Y}$ に関して)線形な推定量の中で最小の分散を持っているという事である。

これは**ガウス・マルコフの定理**と呼ばれる。

この結果は ε_i の分布に関わらず、仮定(A1) ~ (A3)が満たされていれば成立する。

線形回帰分析

- 最小二乗推定量の分布

以下の仮定をおく。

(A5) ε_i は正規分布に従う。

この時 $\hat{\boldsymbol{\beta}}$ は多変量(k 変量)正規分布

$$N(\boldsymbol{\beta}, \sigma^2(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1})$$

に従う。

線形回帰分析

- m変量正規分布

m 変量確率変数 $\mathbf{Z} = (Z_1,\ldots, Z_m)^{\mathrm{T}}$ が m 変量正規分布 $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ に従うとは $Z_1,\ldots, Z_m$ のそれぞれが正規分布に従いその平均が $E(\mathbf{Z}) = \boldsymbol{\mu}$、分散共分散行列が $\mathrm{var}(\mathbf{Z}) = \boldsymbol{\Sigma}$ であるような分布の事である。

線形回帰分析

- 誤差項の分散の推定

線形回帰モデルの誤差項 ε_i の分散 σ^2 の推定を考えよう

通常 σ^2 の推定には最小二乗推定量を代入して得た
残差平方和 $SSR(\hat{\boldsymbol{\beta}})$を $n-k$ で割った

$$s^2 = \frac{\sum_{i=1}^{n} \hat{\varepsilon}_i^2}{n-k}$$

が用いられる。ここで

$$\hat{\varepsilon}_i = Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \ldots - \hat{\beta}_k X_{ki}, \quad i = 1, \ldots, n$$

である。

線形回帰分析

- s^2の性質

誤差項 ε_i の分散 σ^2 の推定量として s^2 には次の性質がある。

仮定(1)～(5)のもとで

(S1) s^2 は σ^2 の**不偏推定量**である。

(S2) s^2 は σ^2 の**一致推定量**である。

(S3) s^2 は σ^2 の**最良2次不偏推定量**である。

以下では(S1)だけを確認する。

線形回帰分析

- s^2の不偏性

s^2 が σ^2 の不偏推定量である事を見て行こう。
これはそれほど難しくないが、いろいろと準備がいる。
まず $\mathbf{e} = (\hat{\varepsilon}_1, \hat{\varepsilon}_2, \ldots, \hat{\varepsilon}_n)^{\mathrm{T}}$ とするとこの $\mathbf{e}$ は

$$\begin{aligned} \mathbf{e} &= \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}} \\ &= \mathbf{Y} - \mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Y} \\ &= (\mathbf{I}_n - \mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}})\mathbf{Y} \\ &= \mathbf{M}\mathbf{Y} \end{aligned}$$

と表す事ができる。ここで $\mathbf{M} = \mathbf{I}_n - \mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}$ である。
$\mathbf{M}$ は $\mathbf{M}^2 = \mathbf{M}\mathbf{M} = \mathbf{M}$ を満たす(演習問題)。

線形回帰分析

- s^2の不偏性

さらに

$$\begin{aligned}\mathbf{MX} &= (\mathbf{X} - \mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{X}) \\ &= \mathbf{X} - \mathbf{X} \\ &= \mathbf{0}\end{aligned}$$

であるから

$$\begin{aligned}\mathbf{e} &= \mathbf{MY} \\ &= \mathbf{M}(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) \\ &= \mathbf{MX}\boldsymbol{\beta} + \mathbf{M}\boldsymbol{\varepsilon} \\ &= \mathbf{M}\boldsymbol{\varepsilon}\end{aligned}$$

である。

線形回帰分析

- s^2の不偏性

よってs^2 の分子の部分は

$$\sum_{i=1}^{n} \hat{\varepsilon}_i^2 = \mathbf{e}^{\mathrm{T}}\mathbf{e} = (\mathbf{M}\boldsymbol{\varepsilon})^{\mathrm{T}}(\mathbf{M}\boldsymbol{\varepsilon}) = \boldsymbol{\varepsilon}^{T}\mathbf{M}\mathbf{M}\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}^{T}\mathbf{M}\boldsymbol{\varepsilon}$$

となる。期待値をとると

$$\begin{aligned} E(\boldsymbol{\varepsilon}^{\mathrm{T}}\mathbf{M}\boldsymbol{\varepsilon}) &= E\left(\sum_{i=1}^{n}\sum_{j=1}^{n} m_{ij}\varepsilon_i\varepsilon_j\right) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} m_{ij}E(\varepsilon_i\varepsilon_j) \\ &= \sigma^2\sum_{i=1}^{n} m_{ii} = \sigma^2\,\mathrm{tr}(\mathbf{M}) \end{aligned}$$

となる。ここで m_{ij} は $\mathbf{M}$ の(i, j)成分であり $\mathrm{tr}\,(\mathbf{A})$ とは行列$\mathbf{A}$の対角要素の和を表す。

線形回帰分析

- s^2の不偏性

(もう少し…)

ここで $\mathrm{tr}(\mathbf{M}) = n - k$ である事が示せるので(演習問題)、

$$E(\boldsymbol{\varepsilon}^{\mathrm{T}}\mathbf{M}\boldsymbol{\varepsilon}) = (n-k)\sigma^2$$

が得られる。よって s^2 の期待値は

$$E(s^2) = \left(\frac{\sum_{i=1}^{n}\hat{\varepsilon}_i^2}{n-k}\right) = \frac{1}{n-k}E(\boldsymbol{\varepsilon}^{\mathrm{T}}\mathbf{M}\boldsymbol{\varepsilon}) = \sigma^2$$

となる。したがって　s^2 は σ^2 の**不偏推定量**である。

演習問題

p.23 に出てきた行列

$$\mathbf{M} = \mathbf{I}_n - \mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}$$

について

問題1 $\mathbf{M}$ は対称行列である事を証明しなさい

問題2 $\mathbf{M}$ は $\mathbf{M}^2 = \mathbf{MM} = \mathbf{M}$ を満たす事を証明しなさい

(このような行列を巾等行列という)

問題3 $\mathrm{tr}(\mathbf{M}) = n - k$ である事を証明しなさい。

問題3のヒント: 2 つの行列 $\mathbf{A}$, $\mathbf{B}$ に対して、もし $\mathbf{AB}$ も $\mathbf{BA}$ も定義されるならば $\mathrm{tr}(\mathbf{AB}) = \mathrm{tr}(\mathbf{BA})$である。