線形重回帰

July 24, 2021

1 線形重回帰

1.1 線形重回帰の行列表現

線形重回帰モデルは応答変数 Y_i $(i = 1, 2, \cdots, n)$ と説明変数 x_{ij} $(i = 1, 2, \cdots, n,\ j = 1, 2, \cdots, k)$ の間に次の関係を仮定している。

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \epsilon_i \\ \epsilon_i &\sim N(0, \sigma^2) \end{aligned}$$

ここで、β_j $(j = 0, 1, 2, \cdots, k)$ は重回帰係数、ϵ_i は平均が 0、分散 σ^2 が共通である独立なランダム変数である。

この重回帰モデルを行列で表現すると、次のように表すことができる。

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

ここで、

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} \quad , \quad \mathbf{X} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{pmatrix}$$

$$\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} \quad , \quad \epsilon = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

である。行列 $\mathbf{X}$ は $n \times (k+1)$ の行列でデザイン行列 (*design matrix*) と呼ばれる。回帰モデルをフィットしたときの値 $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \cdots \hat{\beta}_k x_{ik}$ と残差ベクトル $\mathbf{e} = \mathbf{Y} - \hat{\mathbf{Y}} = \mathbf{Y} - \mathbf{X}\hat{\beta}$ も次のように行列で表現することができる。

$$\begin{aligned} \hat{\mathbf{Y}} &= \begin{pmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \vdots \\ \hat{Y}_n \end{pmatrix} = \begin{pmatrix} \hat{\beta}_0 + \hat{\beta}_1 x_{11} + \hat{\beta}_2 x_{12} + \cdots + \hat{\beta}_k x_{1k} \\ \hat{\beta}_0 + \hat{\beta}_1 x_{21} + \hat{\beta}_2 x_{22} + \cdots + \hat{\beta}_k x_{2k} \\ \vdots \\ \hat{\beta}_0 + \hat{\beta}_1 x_{n1} + \hat{\beta}_2 x_{n2} + \cdots + \hat{\beta}_k x_{nk} \end{pmatrix} = \mathbf{X}\hat{\beta} \\ \mathbf{e} &= \begin{pmatrix} Y_1 - \hat{Y}_1 \\ Y_2 - \hat{Y}_2 \\ \vdots \\ Y_n - \hat{Y}_n \end{pmatrix} = \mathbf{Y} - \mathbf{X}\hat{\beta} \end{aligned}$$

ここで、

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_k \end{pmatrix}$$

は回帰パラメータの推定値である。

残差ベクトルはデザイン行列と直交しているので、次のように書くことができる。

$$\mathbf{X}^T(\mathbf{Y} - \mathbf{X}\hat{\beta}) = 0$$

これから、パラメータの推定値 $\hat{\beta}$ は次の正規方程式から求めることができる。

$$\mathbf{X}^T\mathbf{X}\hat{\beta} = \mathbf{X}^T\mathbf{Y}$$

逆行列 $(\mathbf{X}^T\mathbf{X})^{-1}$ が存在すれば、正規方程式の解 $\hat{\beta}$、フィットされた値のベクトル、残差ベクトルは、次のように求めることができる。

$$\begin{aligned}
\hat{\beta} &= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} && (1)\\
\hat{\mathbf{Y}} &= \mathbf{X}\hat{\beta} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} = \mathbf{H}\mathbf{Y} && (2)\\
\mathbf{e} &= \mathbf{Y} - \hat{\mathbf{Y}} = \mathbf{Y} - \mathbf{H}\mathbf{Y} = (\mathbf{I} - \mathbf{H})\mathbf{Y} && (3)
\end{aligned}$$

ここで、

$$\mathbf{H} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T \quad (4)$$

である。行列 $\mathbf{H}$ はハット行列 (*hat matrix*) と呼ばれる。

1.2 標準誤差の推定値と回帰パラメータの推定値

ランダムベクトル $\mathbf{W}$ がランダムベクトル $\mathbf{Y}$ と行列 $\mathbf{A}$ の積で表わされるとき、すなわち $\mathbf{W} = \mathbf{A}\mathbf{Y}$ であるとき、

$$\mathrm{Cov}(\mathbf{W}) = \mathbf{A}\mathrm{Cov}(\mathbf{Y})\mathbf{A}^T \quad (5)$$

が成立する。式 (5) から

$$\begin{aligned}
\mathbf{A} &= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\\
\mathbf{A}^T &= \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\\
\mathrm{Cov}(\mathbf{Y}) &= \sigma^2\mathbf{I}
\end{aligned}$$

であることを使うと、

$$\mathrm{Cov}(\hat{\beta}) = \mathbf{A}\sigma^2\mathbf{I}\mathbf{A}^T = \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1} = \sigma^2(\mathbf{X}^T\mathbf{X})^{-1} \quad (6)$$

が導かれる。

σ^2 の不偏推定量 s^2 は

$$s^2 = \frac{\mathbf{e}^T\mathbf{e}}{n-k-1}$$

で表されるので、これを式 (6) に代入すると、分散共分散行列の推定値 (*estimated variance-covariance matrix*) を得ることができる。

$$s^2(\hat{\beta}) = s^2(\mathbf{X}^T\mathbf{X})^{-1}$$

これから、パラメータ $\hat{\beta}$ の推定値の標準誤差の推定値 $s(\hat{\beta})$ を求めることができる。

1.3 重決定係数

全変動 (sum of total squares(SST))、回帰変動 (sum of the squared regression(SSR))、残差変動 (sum of the squared errors(SSE)) はそれぞれ次のように定義される。

$$\begin{aligned}
\mathrm{SST} &= \sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \mathbf{Y}^T\mathbf{Y} - n\bar{Y}^2\\
\mathrm{SSR} &= \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 = \hat{\mathbf{Y}}^T\hat{\mathbf{Y}} - n\bar{Y}^2\\
\mathrm{SSE} &= \sum_{i=1}^{n}e_i^2 = \mathbf{e}^T\mathbf{e}
\end{aligned}$$

これらは次の関係も満す。

$$\mathrm{SST} = \mathrm{SSR} + \mathrm{SSE}$$

重決定係数 R^2 は次のように定義される。

$$R^2 = \frac{\mathrm{SSR}}{\mathrm{SST}}$$

自由度調整済み決定係数 R_a^2 は次のように定義される。

$$R_a^2 = 1 - \left(\frac{n-1}{n-k-1}\right)(1-R^2)$$

1.4 回帰パラメータの信頼区間

回帰パラメータの信頼区間は、

$$\frac{\hat{\beta}_i - \beta_i}{s(\hat{\beta}_i)} \sim t_{n-k-1}$$

という関係から導くことができる。ここで、$s(\hat{\beta}_i)$ はパラメータの推定値 $\hat{\beta}_i$ の標準誤差の推定値である。また、t_{n-k-1} は自由度 $(n-k-1)$ のステューデントの t 分布を示す。例えば、信頼係数 100(1-α)%での回帰パラメータの推定値 $\hat{\beta}_i$ の信頼区間は

$$\hat{\beta}_i \pm s(\hat{\beta}_i) t_{n-k-1}\left(\frac{\alpha}{2}\right)$$

で与えられる。

1.5 Studentized residual

残差ベクトル $\mathbf{e}$ はハット行列、式 (4) を使うと、次のように書くことができる。

$$\mathbf{e} = \mathbf{Y} - \hat{\mathbf{Y}} = \mathbf{Y} - \mathbf{H}\mathbf{Y} = (\mathbf{I} - \mathbf{H})\mathbf{Y}$$

式 (5) から

$$\begin{aligned} \mathbf{H}^T &= \mathbf{H}, \\ (\mathbf{I}-\mathbf{H})^2 &= (\mathbf{I}-\mathbf{H}) \end{aligned}$$

であるので、

$$\mathrm{Cov}(\mathbf{e}) = (\mathbf{I}-\mathbf{H})\sigma^2\mathbf{I}(\mathbf{I}-\mathbf{H})^T = \sigma^2(\mathbf{I}-\mathbf{H})^2 = \sigma^2(\mathbf{I}-\mathbf{H})$$

が導かれる。

したがって、i 番目の残差 e_i の分散は次のように示される。

$$V(e_i) = \sigma^2(1-h_{ii})$$

ここで、h_{ii} はハット行列の i 行目の対角要素である。

分散 σ^2 をその不偏推定量 s^2 で置き換えると、i 番目の残差の標準誤差の推定値 (*estimated standard error for the ith residual*) を $s(e_i)$ で表すと、次のように与えられる。

$$s(e_i) = s\sqrt{1-h_{ii}}$$

Studentized residual を e_i^* で表すと、次のように定義される。

$$e_i^* = \frac{e_i}{s(e_i)}$$

1.6 線形重回帰における平均応答の信頼区間と予測の信頼区間

いま、ベクトル

$$\mathbf{x}_0^T = (1, x_{01}, x_{02}, \cdots, x_{0k})$$

は説明変数の値を示すものとする。すなわち、

$$X_1 = x_{01}, X_2 = x_{02}, \cdots, X_k = x_{0k}$$

であるとする。平均応答の推定値を $\hat{Y}(\mathbf{x}_0)$ で表すと、行列の積として次のように書くことができる。

$$\hat{Y}(\mathbf{x}_0) = \mathbf{x}_0^T \hat{\beta}$$

式 (2)、

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

を使うと、

$$\mathbf{x}_0^T \hat{\beta} = \mathbf{A}\mathbf{Y}$$

のように表すことができる。ここで、

$$\begin{aligned} \mathbf{A} &= \mathbf{x}_0^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T \\ \mathbf{A}^T &= \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_0 \end{aligned}$$

である。式 (5) から、$\mathrm{Cov}(\mathbf{Y}) = \sigma^2\mathbf{I}$ を使うと、

$$\mathrm{Cov}(\mathbf{x}_0^T\hat{\beta}) = \mathbf{A}\sigma^2\mathbf{I}\mathbf{A}^T = \sigma^2\mathbf{x}_0^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_0 = \sigma^2\mathbf{x}_0^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_0$$

が導かれる。$\hat{Y}(\mathbf{x}_0)$ の分散は

$$\mathrm{V}(\hat{Y}(\mathbf{x}_0)) = \sigma^2\mathbf{x}_0^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_0$$

で与えられるので、いま、予測の標準誤差の推定値 (*standard error of prediction*) を $s(\hat{Y}(\mathbf{x}_0))$ で表すと、σ^2 をその不偏推定量 $s^2 = \mathrm{SSE}/(n-k-1)$ で置き換えると、

$$s(\hat{Y}(\mathbf{x}_0)) = s\sqrt{\mathbf{x}_0^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_0}$$

が得られる。

値 $\mathbf{x}_0$ における信頼係数 $100(1-\alpha)\%$での平均応答の信頼区間は

$$\mathbf{x}_0^T\hat{\beta} \pm t_{n-k-1}\left(\frac{\alpha}{2}\right)s\sqrt{\mathbf{x}_0^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_0}$$

で与えらえる。

将来の応答は $Y(\mathbf{x}_0) = \beta_0 + \beta_1 x_{01} + \cdots + \beta_k x_{0k} + \epsilon_0$ によって与えられる。また、説明変数 $\mathbf{x}_0$ において予測される将来の応答は、$\hat{Y}(\mathbf{x}_0) = \hat{\beta}_0 + \hat{\beta}_1 x_{01} + \cdots + \hat{\beta}_k x_{0k}$ で与えられる。ここで、ϵ_0 は平均が 0、分散が σ^2 の独立な正規ランダム変数であり、$\hat{Y}(\mathbf{x}_0)$ とは独立である。

その結果、将来の応答 $Y(\mathbf{x}_0)$ と予測される将来の応答 $\hat{Y}(\mathbf{x}_0)$ との差の分散、$V(Y(\mathbf{x}_0) - \hat{Y}(\mathbf{x}_0))$ は、

$$\mathrm{V}(Y(\mathbf{x}_0) - \hat{Y}(\mathbf{x}_0)) = \mathrm{V}(Y(\mathbf{x}_0)) + \mathrm{V}(\hat{Y}(\mathbf{x}_0)) = \sigma^2(1 + \mathbf{x}_0^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_0)$$

で与えられる。

同様に、将来の応答 $Y(\mathbf{x}_0)$ と予測される将来の応答 $\hat{Y}(\mathbf{x}_0)$ との差の標準誤差の推定値は、

$$s(Y(\mathbf{x}_0) - \hat{Y}(\mathbf{x}_0)) = s\sqrt{1 + \mathbf{x}_0^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_0}$$

で与えられる。

したがって、信頼係数 $100(1-\alpha)\%$での将来の応答 $Y(\mathbf{x}_0)$ の予測信頼区間は次のように与えられる。

$$\mathbf{x}_0^T\hat{\beta} \pm t_{n-k-1}\left(\frac{\alpha}{2}\right)s\sqrt{1 + \mathbf{x}^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}}$$

References

[1] W.A.Rosenkrantz, "Introduction to Probability and Statistics for Scientists and Engineers", McGraw-Hill、1997