線形回帰

線形回帰とは？

- 目的変数Yと説明変数X_i ,$i\ =\ 1, \dots ,p$ および擾乱項ε の関係を以下のようにモデル化したものである。

$$Y = \beta_0 + \beta_1 X_1 + \ \dots + \beta_p X_p + \epsilon$$

- ϵ はXと独立である。

- 最小二乗法によって得られる回帰パラメタの推定量は

$$\hat{\beta} = \left(\mathrm{X}^{\mathrm{T}}\mathrm{X}\right)^{-1} X^T \vec{y}$$

性質

- 不変推定であることはもちろん、このβは最良線形不偏推定量となっている。
- 説明変数が一つの時は以下の式で与えられる

$$\hat{\beta} = \frac{\sum_n (x_n - \bar{x})(y_n - \bar{y})}{\sum_n (x_n - \bar{x}\ )^2}$$

$$\alpha = \bar{y} - \hat{b}\bar{x}$$

決定係数 (R^2)

- 統計学において、独立変数（説明変数）が従属変数（目的変数）のどれくらいを説明できるかを表す値である。(wiki)

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2}$$

線形回帰の場合は

$$R^2 = \frac{\sum_i (\hat{y}_i - \bar{y})^2}{\sum_i (y_i - \bar{y})^2} = \frac{\sum_i \left(\hat{y}_i - \bar{\hat{y}}\right)^2}{\sum_i (y_i - \bar{y})^2}$$

$$R^2 = \frac{SS_{\hat{y}y}^2}{SS_{\hat{y}} SS_y}$$

- 最小二乗法はこの定義を最大にするようなパラメータの選択法でもある。
- 特に、単回帰の場合は相関係数の２乗にもなる

Adjusted R^2

$$\bar{R}^2 = 1 - \frac{\frac{SS_{res}}{n-p-1}}{\frac{SS_{tot}}{n-1}}$$

擾乱項が正規分布に従うモデル

- 統計量$S^2 = \frac{(\vec{y}-X\hat{\beta})\cdot(\vec{y}-X\hat{\beta})}{n-p-1}$は分散の不変推定量になる。

有意性検定

- 擾乱項が正規分布に従うモデルでは、回帰係数の有意性を検定できる

- $T = \frac{\widehat{\beta}_i - \beta_i}{\sqrt{(X^T X)_{ii}^{-1} S^2}}$ は自由度n-p-1のt分布に従う。 Sは錯乱項が持つ分散の不変推定量詳しくは以下のurlを参照

https://starpentagon.net/analytics/multiple_linear_regression_coef_test/