

回帰分析

回帰モデルの考え方と推定

村田 昇

講義の内容

- 第1回：回帰モデルの考え方と推定
- 第2回：モデルの評価
- 第3回：モデルによる予測と発展的なモデル

回帰分析の例

ボルドーワインの価格と気候の関係

| VINT | LPRICE2 | WRAIN | DEGREES | HRAIN | TIME_SV |
|------|----------|-------|---------|-------|---------|
| 1952 | -0.99868 | 600 | 17.1167 | 160 | 31 |
| 1953 | -0.45440 | 690 | 16.7333 | 80 | 30 |
| 1954 | NA | 430 | 15.3833 | 180 | 29 |
| 1955 | -0.80796 | 502 | 17.1500 | 130 | 28 |
| 1956 | NA | 440 | 15.6500 | 140 | 27 |
| 1957 | -1.50926 | 420 | 16.1333 | 110 | 26 |
| 1958 | -1.71655 | 582 | 16.4167 | 187 | 25 |
| 1959 | -0.41800 | 485 | 17.4833 | 187 | 24 |
| 1960 | -1.97491 | 763 | 16.4167 | 290 | 23 |
| 1961 | 0.00000 | 830 | 17.3333 | 38 | 22 |
| 1962 | -1.10572 | 697 | 16.3000 | 52 | 21 |
| 1963 | -1.78098 | 608 | 15.7167 | 155 | 20 |
| 1964 | -1.18435 | 402 | 17.2667 | 96 | 19 |
| 1965 | -2.24194 | 602 | 15.3667 | 267 | 18 |
| 1966 | -0.74943 | 819 | 16.5333 | 86 | 17 |
| 1967 | -1.65388 | 714 | 16.2333 | 118 | 16 |
| 1968 | -2.25018 | 610 | 16.2000 | 292 | 15 |
| 1969 | -2.14784 | 575 | 16.5500 | 244 | 14 |
| 1970 | -0.90544 | 622 | 16.6667 | 89 | 13 |
| 1971 | -1.30031 | 551 | 16.7667 | 112 | 12 |
| 1972 | -2.28879 | 536 | 14.9833 | 158 | 11 |
| 1973 | -1.85700 | 376 | 17.0667 | 123 | 10 |
| 1974 | -2.19958 | 574 | 16.3000 | 184 | 9 |
| 1975 | -1.20168 | 572 | 16.9500 | 171 | 8 |
| 1976 | -1.37264 | 418 | 17.6500 | 247 | 7 |

| | | | | | |
|------|----------|-----|---------|-----|----|
| 1977 | -2.23503 | 821 | 15.5833 | 87 | 6 |
| 1978 | -1.30769 | 763 | 15.8167 | 51 | 5 |
| 1979 | -1.53960 | 717 | 16.1667 | 122 | 4 |
| 1980 | -1.99582 | 578 | 16.0000 | 74 | 3 |
| 1981 | NA | 535 | 16.9667 | 111 | 2 |
| 1982 | NA | 712 | 17.4000 | 162 | 1 |
| 1983 | NA | 845 | 17.3833 | 119 | 0 |
| 1984 | NA | 591 | 16.5000 | 119 | -1 |
| 1985 | NA | 744 | 16.8000 | 38 | -2 |
| 1986 | NA | 563 | 16.2833 | 171 | -3 |
| 1987 | NA | 452 | 16.9833 | 115 | -4 |
| 1988 | NA | 808 | 17.1000 | 59 | -5 |
| 1989 | NA | 443 | NA | 82 | -6 |

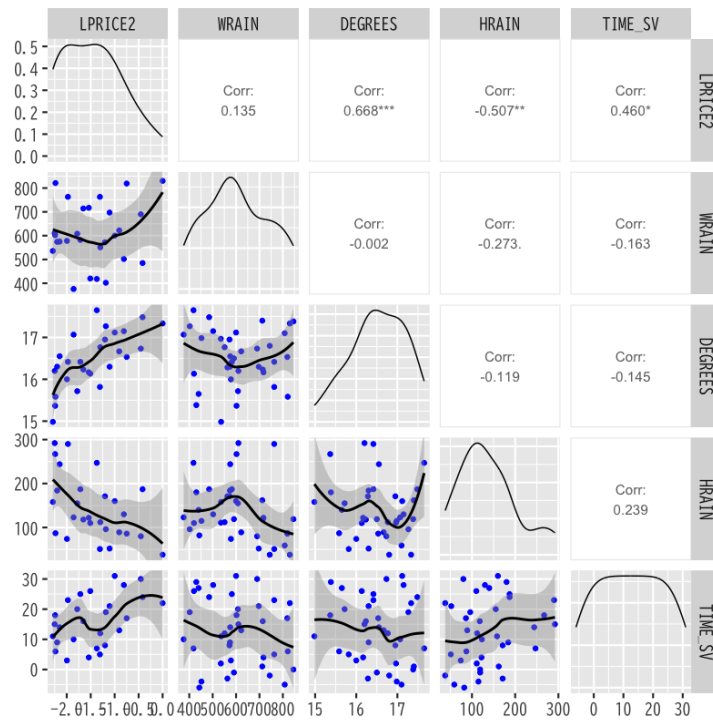


Figure 1: 価格と気候の散布図

- 回帰式

$$LPRICE2 = \beta_0 + \beta_1 \times WRAIN + \beta_2 \times DEGREES + \beta_3 \times HRain + \beta_4 \times TIME_SV$$

| Characteristic | Beta | 95% CI ^l | p-value |
|----------------|---------|---------------------|---------|
| (Intercept) | -12.15 | -15.65, -8.644 | <0.001 |
| WRAIN | 0.0012 | 0.0002, 0.0022 | 0.024 |
| DEGREES | 0.6164 | 0.4190, 0.8138 | <0.001 |
| HRain | -0.0039 | -0.0055, -0.0022 | <0.001 |

TIME_SV 0.0238 0.0090, 0.0387 0.003

¹CI = Confidence Interval

R² = 0.828; Adjusted R² = 0.796; Statistic = 26.4; p-value = <0.001

回帰分析の考え方

回帰分析

- ある変量を別の変量で説明する関係式を構成する
- 関係式：回帰式 (regression equation)
 - 説明される側：目的変数, 被説明変数, 従属変数, 応答変数
 - 説明する側：説明変数, 独立変数, 共変量
- 説明変数の数による分類
 - 一つの場合：単回帰 (simple regression)
 - 複数の場合：重回帰 (multiple regression)

一般の回帰の枠組

- 説明変数： x_1, \dots, x_p (p 次元)
- 目的変数： y (1 次元)
- 回帰式： y を x_1, \dots, x_p で説明するための関係式

$$y = f(x_1, \dots, x_p)$$

- 観測データ：n 個の (y, x_1, \dots, x_p) の組

$$\{(y_i, x_{i1}, \dots, x_{ip})\}_{i=1}^n$$

線形回帰

- 任意の f では一般的すぎて分析に不向き
- f として 1 次関数 を考える
 - ある定数 $\beta_0, \beta_1, \dots, \beta_p$ を用いた式：
$$f(x_1, \dots, x_p) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p$$
 - 1 次関数の場合：線形回帰 (linear regression)
 - 一般の場合：非線形回帰 (nonlinear regression)
- 非線形関係は新たな説明変数の導入で対応可能
 - 適切な多項式： $x_j^2, x_j x_k, x_j x_k x_l, \dots$
 - その他の非線形変換： $\log x_j, x_j^\alpha, \dots$
 - 全ての非線形関係ではないことに注意

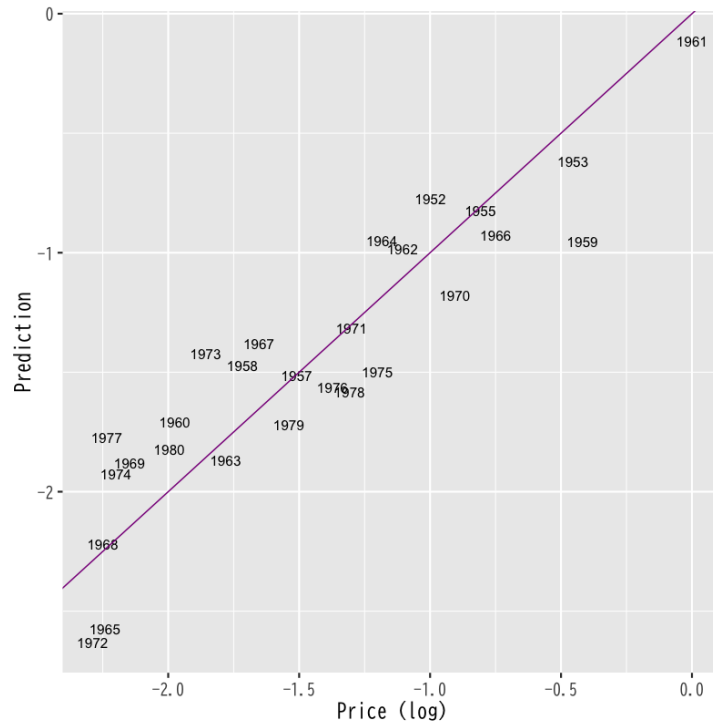


Figure 2: 重回帰による予測値と実際の価格

回帰係数

- 線形回帰式

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

- $\beta_0, \beta_1, \dots, \beta_p$: 回帰係数 (regression coefficients)
- β_0 : 定数項 / 切片 (constant term / intersection)

- 線形回帰分析 (linear regression analysis)
 - 未知の回帰係数をデータから決定する分析方法
 - 決定された回帰係数の統計的な性質を診断

回帰の確率モデル

- 回帰式の不確定性
 - データは一般に観測誤差などランダムな変動を含む
 - 回帰式がそのまま成立することは期待できない
- 確率モデル: データのばらつきを表す項 ϵ_i を追加

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \epsilon_i \quad (i = 1, \dots, n)$$

- $\epsilon_1, \dots, \epsilon_n$: 誤差項 / 攪乱項 (error / disturbance term)
 - * 誤差項は独立な確率変数と仮定
 - * 多くの場合, 平均 0, 分散 σ^2 の正規分布を仮定
- 推定 (estimation): 観測データから回帰係数を決定

回帰係数の推定

残差

- 残差 (residual) : 回帰式で説明できない変動
- 回帰係数 $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)^\top$ を持つ回帰式の残差

$$e_i(\boldsymbol{\beta}) = y_i - (\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}) \quad (i = 1, \dots, n)$$

- 残差 $e_i(\boldsymbol{\beta})$ の絶対値が小さいほど当てはまりがよい

最小二乗法

- 残差平方和 (residual sum of squares)

$$S(\boldsymbol{\beta}) = \sum_{i=1}^n e_i(\boldsymbol{\beta})^2$$

- 最小二乗推定量 (least squares estimator)

残差平方和 $S(\boldsymbol{\beta})$ を最小にする $\boldsymbol{\beta}$

$$\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p)^\top = \arg \min_{\boldsymbol{\beta}} S(\boldsymbol{\beta})$$

行列の定義

- デザイン行列 (design matrix)

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

- $n \times (p+1)$ 行列

ベクトルの定義

- 目的変数, 誤差, 回帰係数のベクトル

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad \boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}$$

- $\mathbf{y}, \boldsymbol{\epsilon}$ は n 次元ベクトル

- $\boldsymbol{\beta}$ は $p+1$ 次元ベクトル

行列・ベクトルによる表現

- 確率モデル

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

- 残差平方和

$$S(\boldsymbol{\beta}) = (\mathbf{y} - X\boldsymbol{\beta})^\top (\mathbf{y} - X\boldsymbol{\beta})$$

解の条件

- 解 $\boldsymbol{\beta}$ では残差平方和の勾配は零ベクトル

$$\frac{\partial S}{\partial \boldsymbol{\beta}}(\boldsymbol{\beta}) = \left(\frac{\partial S}{\partial \beta_0}(\boldsymbol{\beta}), \frac{\partial S}{\partial \beta_1}(\boldsymbol{\beta}), \dots, \frac{\partial S}{\partial \beta_p}(\boldsymbol{\beta}) \right)^\top = \mathbf{0}$$

演習

問題

- 残差平方和 $S(\boldsymbol{\beta})$ をベクトル $\boldsymbol{\beta}$ で微分して解の条件を求めなさい

解答例

- 残差平方和を展開しておく

$$\begin{aligned} S(\boldsymbol{\beta}) &= (\mathbf{y} - X\boldsymbol{\beta})^\top (\mathbf{y} - X\boldsymbol{\beta}) \\ &= \mathbf{y}^\top \mathbf{y} - \mathbf{y}^\top X\boldsymbol{\beta} - (X\boldsymbol{\beta})^\top \mathbf{y} + (X\boldsymbol{\beta})^\top X\boldsymbol{\beta} \\ &= \mathbf{y}^\top \mathbf{y} - \mathbf{y}^\top X\boldsymbol{\beta} - \boldsymbol{\beta}^\top X^\top \mathbf{y} + \boldsymbol{\beta}^\top X^\top X\boldsymbol{\beta} \end{aligned}$$

- ベクトルによる微分を行うと以下のようなになる

$$\begin{aligned} \frac{\partial S}{\partial \boldsymbol{\beta}}(\boldsymbol{\beta}) &= -(\mathbf{y}^\top X)^\top - X^\top \mathbf{y} + (X^\top X + (X^\top X)^\top)\boldsymbol{\beta} \\ &= -2X^\top \mathbf{y} + 2X^\top X\boldsymbol{\beta} \end{aligned}$$

- したがって $\boldsymbol{\beta}$ の満たす条件は以下となる

$$\begin{aligned} -2X^\top \mathbf{y} + 2X^\top X\boldsymbol{\beta} &= 0 \quad \text{より} \\ X^\top X\boldsymbol{\beta} &= X^\top \mathbf{y} \end{aligned}$$

補足

- 成分ごとの計算は以下のようになる

$$\frac{\partial S}{\partial \beta_j}(\boldsymbol{\beta}) = -2 \sum_{i=1}^n \left(y_i - \sum_{k=0}^p \beta_k x_{ik} \right) x_{ij} = 0$$

ただし, $x_{i0} = 1$ ($i = 1, \dots, n$), $j = 0, 1, \dots, p$

$$\sum_{i=1}^n x_{ij} \left(\sum_{k=0}^p x_{ik} \beta_k \right) = \sum_{i=1}^n x_{ij} y_i \quad (j = 0, 1, \dots, p)$$

x_{ij} は行列 X の (i, j) 成分であることに注意

正規方程式

正規方程式

- 正規方程式 (normal equation)

$$X^T X \boldsymbol{\beta} = X^T \mathbf{y}$$

- Gram 行列 (Gram matrix)

$$X^T X$$

- $(p+1) \times (p+1)$ 行列 (正方行列)
- 正定対称行列 (固有値が非負)

正規方程式の解

- 正規方程式の基本的な性質
 - 正規方程式は必ず解をもつ (一意に決まらない場合もある)
 - 正規方程式の解は最小二乗推定量であるための必要条件
- 解の一意性の条件
 - Gram 行列 $X^T X$ が **正則**
 - X の列ベクトルが独立 (後述)
- 正規方程式の解

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

最小二乗推定量の性質

解析の上での良い条件

- 最小二乗推定量がただ一つだけ存在する条件
 - $X^T X$ が正則
 - $X^T X$ の階数が $p+1$
 - X の階数が $p+1$
 - X の列ベクトルが **1 次独立**

これらは同値条件

解析の上での良くない条件

- 説明変数が 1 次従属：多重共線性 (multicollinearity)
- 多重共線性が強くないように説明変数を選択
 - X の列 (説明変数) の独立性を担保する
 - 説明変数が互いに異なる情報をもつように選ぶ
 - 似た性質をもつ説明変数の重複は避ける

推定の幾何学的解釈

- あてはめ値 / 予測値 (fitted values / predicted values)

$$\hat{y} = X\hat{\beta} = \hat{\beta}_0 X_{\text{第0列}} + \cdots + \hat{\beta}_p X_{\text{第p列}}$$

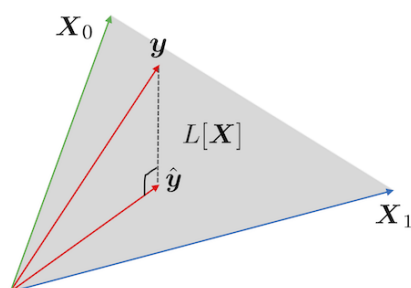


Figure 3: $n = 3, p + 1 = 2$ の場合の最小二乗法による推定

- 最小二乗推定量 \hat{y} の幾何学的性質
 - $L[X]$: X の列ベクトルが張る \mathbb{R}^n の線形部分空間
 - X の階数が $p+1$ ならば $L[X]$ の次元は $p+1$ (解の一意性)
 - \hat{y} は y の $L[X]$ への直交射影
 - 残差 (residuals) $\hat{e} = y - \hat{y}$ はあてはめ値 \hat{y} に直交

$$\hat{e} \cdot \hat{y} = 0$$

線形回帰式と標本平均

- $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^T$: i 番目の観測データの説明変数
- 説明変数および目的変数の標本平均

$$\bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i, \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i,$$

- $\hat{\beta}$ が最小二乗推定量のとき以下が成立

$$\bar{y} = (1, \bar{\mathbf{x}}^T) \hat{\beta}$$

演習

問題

- 最小二乗推定量について以下を示しなさい
 - 残差の標本平均が 0 となる
目的変数や残差のベクトルについて以下を示せばよい

$$\mathbf{1}^T(\mathbf{y} - \hat{\mathbf{y}}) = \mathbf{1}^T \hat{\boldsymbol{\epsilon}} = 0$$

ただし $\mathbf{1} = (1, \dots, 1)^T$ とする

- 回帰式が標本平均を通る

$$\bar{y} = (1, \bar{\mathbf{x}}^T) \hat{\boldsymbol{\beta}}$$

解答例

- 残差の表現を整理する

$$\begin{aligned} \hat{\boldsymbol{\epsilon}} &= \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - X\hat{\boldsymbol{\beta}} \\ &= \mathbf{y} - X(X^T X)^{-1} X^T \mathbf{y} \end{aligned}$$

- 左から X^T を乗じる

$$X^T \mathbf{y} - X^T X (X^T X)^{-1} X^T \mathbf{y} = X^T \mathbf{y} - X^T \mathbf{y} = 0$$

- 行列 X の 1 列目が $\mathbf{1}$ であることより明らか
- 説明変数の標本平均をデザイン行列で表す

$$\mathbf{1}^T X = n(1, \bar{\mathbf{x}}^T)$$

- したがって以下が成立する

$$\begin{aligned} n(1, \bar{\mathbf{x}}^T) \hat{\boldsymbol{\beta}} &= \mathbf{1}^T X \hat{\boldsymbol{\beta}} \\ &= \mathbf{1}^T \hat{\mathbf{y}} = \mathbf{1}^T \mathbf{y} \\ &= n\bar{y} \end{aligned}$$

残差の分解

最小二乗推定量の残差

- 観測値と推定値 $\hat{\boldsymbol{\beta}}$ による予測値の差

$$\hat{\epsilon}_i = y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \dots + \hat{\beta}_p x_{ip}) \quad (i = 1, \dots, n)$$

- 誤差項 $\epsilon_1, \dots, \epsilon_n$ の推定値
- 全てができるだけ小さいほど良い
- 予測値とは独立に偏りが無いほど良い

- 残差ベクトル

$$\hat{\boldsymbol{\epsilon}} = \mathbf{y} - \hat{\mathbf{y}} = (\hat{\epsilon}_1, \hat{\epsilon}_2, \dots, \hat{\epsilon}_n)^T$$

平方和の分解

- $\bar{\mathbf{y}} = \bar{y}\mathbf{1} = (\bar{y}, \bar{y}, \dots, \bar{y})^T$: 標本平均のベクトル
- いろいろなばらつき
 - $S_y = (\mathbf{y} - \bar{\mathbf{y}})^T(\mathbf{y} - \bar{\mathbf{y}})$: 目的変数のばらつき
 - $S = (\mathbf{y} - \hat{\mathbf{y}})^T(\mathbf{y} - \hat{\mathbf{y}})$: 残差のばらつき ($\hat{\boldsymbol{\epsilon}}^T \hat{\boldsymbol{\epsilon}}$)
 - $S_r = (\hat{\mathbf{y}} - \bar{\mathbf{y}})^T(\hat{\mathbf{y}} - \bar{\mathbf{y}})$: あてはめ値(回帰)のばらつき
- 3つのばらつき(平方和)の関係

$$(\mathbf{y} - \bar{\mathbf{y}})^T(\mathbf{y} - \bar{\mathbf{y}}) = (\mathbf{y} - \hat{\mathbf{y}})^T(\mathbf{y} - \hat{\mathbf{y}}) + (\hat{\mathbf{y}} - \bar{\mathbf{y}})^T(\hat{\mathbf{y}} - \bar{\mathbf{y}})$$

$$S_y = S + S_r$$

演習

問題

- 以下の関係式を示しなさい
 - あてはめ値と残差のベクトルが直交する

$$\hat{\mathbf{y}}^T(\mathbf{y} - \hat{\mathbf{y}}) = \hat{\mathbf{y}}^T \hat{\boldsymbol{\epsilon}} = 0$$

- 残差平方和の分解が成り立つ

$$S_y = S + S_r$$

解答例

- 残差の表現を整理する

$$\begin{aligned}\hat{\boldsymbol{\epsilon}} &= \mathbf{y} - X(X^T X)^{-1} X^T \mathbf{y} \\ &= (I - X(X^T X)^{-1} X^T) \mathbf{y}\end{aligned}$$

- 左から $\hat{\mathbf{y}}$ を乗じる

$$\begin{aligned}\hat{\mathbf{y}}^T \hat{\boldsymbol{\epsilon}} &= \hat{\boldsymbol{\beta}}^T X^T (I - X(X^T X)^{-1} X^T) \mathbf{y} \\ &= \hat{\boldsymbol{\beta}}^T (X^T - X^T X (X^T X)^{-1} X^T) \mathbf{y} \\ &= \hat{\boldsymbol{\beta}}^T (X^T - X^T) \mathbf{y} = 0\end{aligned}$$

- 以下の関係を用いて展開すればよい

$$\mathbf{y} - \bar{\mathbf{y}} = \mathbf{y} - \hat{\mathbf{y}} + \hat{\mathbf{y}} - \bar{\mathbf{y}}$$

$$\text{ただし } \bar{\mathbf{y}} = \bar{y}\mathbf{1}$$

- このとき以下の項は0になる

$$(\hat{\mathbf{y}} - \bar{\mathbf{y}})^T(\mathbf{y} - \hat{\mathbf{y}}) = \hat{\mathbf{y}}^T(\mathbf{y} - \hat{\mathbf{y}}) - \bar{y}\mathbf{1}^T(\mathbf{y} - \hat{\mathbf{y}}) = 0$$

決定係数

回帰式の寄与

- ばらつきの分解

$$S_y \text{ (目的変数)} = S \text{ (残差)} + S_r \text{ (あてはめ値)}$$

- 回帰式で説明できるばらつきの比率

$$\text{(回帰式の寄与率)} = \frac{S_r}{S_y} = 1 - \frac{S}{S_y}$$

- 回帰式のあてはまり具合を評価する代表的な指標

決定係数 (R^2 値)

- 決定係数 (R-squared)

$$R^2 = 1 - \frac{\sum_{i=1}^n \hat{\epsilon}_i^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

- 自由度調整済み決定係数 (adjusted R-squared)

$$\bar{R}^2 = 1 - \frac{\frac{1}{n-p-1} \sum_{i=1}^n \hat{\epsilon}_i^2}{\frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2}$$

- 不偏分散で補正している

解析の事例

実データによる例

- 気象庁より取得した東京の気候データ
 - 気象庁 <https://www.data.jma.go.jp/gmd/risk/obsdl/index.php>
 - データ https://noboru-murata.github.io/multivariate-analysis/data/tokyo_weather.csv

気温に影響を与える要因の分析

| 日付 | 気温 | 降雨 | 日射 | 降雪 | 風向 | 風速 | 気圧 | 湿度 | 雲量 |
|------------|------|------|-------|----|-----|-----|--------|----|------|
| 2023-09-01 | 29.2 | 0.0 | 24.01 | 0 | SSE | 4.3 | 1012.1 | 71 | 2.0 |
| 2023-09-02 | 29.6 | 0.0 | 22.07 | 0 | SSE | 3.1 | 1010.3 | 72 | 8.0 |
| 2023-09-03 | 29.1 | 3.5 | 18.64 | 0 | ENE | 2.8 | 1010.6 | 74 | 9.3 |
| 2023-09-04 | 26.1 | 34.0 | 7.48 | 0 | N | 2.6 | 1007.5 | 96 | 10.0 |
| 2023-09-05 | 29.3 | 0.0 | 22.58 | 0 | S | 3.5 | 1005.2 | 77 | 3.5 |
| 2023-09-06 | 27.5 | 0.5 | 13.17 | 0 | SSW | 2.6 | 1003.6 | 79 | 10.0 |
| 2023-09-07 | 27.0 | 0.5 | 11.01 | 0 | ENE | 2.5 | 1007.9 | 72 | 10.0 |

| | | | | | | | | | |
|------------|------|-------|-------|---|-----|-----|--------|----|------|
| 2023-09-08 | 21.9 | 107.5 | 2.10 | 0 | NW | 3.4 | 1007.8 | 98 | 10.0 |
| 2023-09-09 | 24.8 | 1.0 | 8.81 | 0 | S | 2.2 | 1006.8 | 93 | 7.5 |
| 2023-09-10 | 27.8 | 0.0 | 17.57 | 0 | S | 3.1 | 1009.1 | 83 | 6.3 |
| 2023-09-11 | 28.1 | 0.0 | 17.19 | 0 | SSE | 3.1 | 1010.1 | 79 | 9.0 |
| 2023-09-12 | 27.7 | 0.0 | 20.02 | 0 | SSE | 2.8 | 1010.0 | 76 | 4.8 |
| 2023-09-13 | 28.0 | 0.0 | 22.00 | 0 | SE | 2.4 | 1010.9 | 74 | 4.5 |
| 2023-09-14 | 28.2 | 0.0 | 14.54 | 0 | SSE | 2.8 | 1009.9 | 80 | 7.0 |
| 2023-09-15 | 27.4 | 10.5 | 9.21 | 0 | NE | 2.0 | 1010.9 | 88 | 8.5 |
| 2023-09-16 | 27.9 | 0.0 | 11.78 | 0 | SSE | 2.0 | 1011.5 | 86 | 10.0 |
| 2023-09-17 | 28.7 | 0.0 | 14.84 | 0 | S | 3.2 | 1011.5 | 80 | 4.0 |
| 2023-09-18 | 28.9 | 0.0 | 19.59 | 0 | S | 4.2 | 1011.6 | 74 | 1.8 |
| 2023-09-19 | 29.0 | 0.0 | 19.93 | 0 | S | 3.3 | 1010.1 | 72 | 2.3 |
| 2023-09-20 | 27.2 | 6.0 | 10.65 | 0 | N | 1.9 | 1009.3 | 82 | 8.3 |
| 2023-09-21 | 26.7 | 2.0 | 6.65 | 0 | S | 4.1 | 1006.7 | 87 | 9.5 |
| 2023-09-22 | 24.8 | 59.5 | 6.83 | 0 | ENE | 2.5 | 1008.1 | 93 | 10.0 |
| 2023-09-23 | 22.1 | 4.0 | 4.48 | 0 | NE | 2.6 | 1012.5 | 89 | 10.0 |
| 2023-09-24 | 22.2 | 0.0 | 15.81 | 0 | N | 3.0 | 1017.2 | 67 | 7.0 |
| 2023-09-25 | 22.4 | 0.0 | 15.49 | 0 | N | 2.5 | 1017.1 | 69 | 6.5 |
| 2023-09-26 | 24.6 | 0.0 | 16.08 | 0 | NNW | 2.0 | 1012.7 | 71 | 6.0 |
| 2023-09-27 | 25.3 | 0.0 | 11.59 | 0 | SSE | 1.9 | 1008.1 | 81 | 9.0 |
| 2023-09-28 | 27.4 | 0.0 | 14.03 | 0 | ESE | 1.9 | 1004.7 | 79 | 5.8 |
| 2023-09-29 | 26.3 | 0.0 | 10.11 | 0 | SSE | 3.0 | 1009.0 | 75 | 8.5 |
| 2023-09-30 | 25.6 | 0.0 | 7.98 | 0 | S | 2.5 | 1007.5 | 77 | 7.0 |

- 気温を説明する 5 種類の線形回帰モデルを検討
 - モデル 1: 気温 = F(気圧)
 - モデル 2: 気温 = F(日射)
 - モデル 3: 気温 = F(気圧, 日射)
 - モデル 4: 気温 = F(気圧, 日射, 湿度)
 - モデル 5: 気温 = F(気圧, 日射, 雲量)

分析の視覚化

- 関連するデータの散布図
- モデル 1 の推定結果
- モデル 2 の推定結果
- モデル 3 の推定結果
- 観測値とあてはめ値の比較

モデルの比較

- 決定係数 (R^2 , Adjusted R^2)

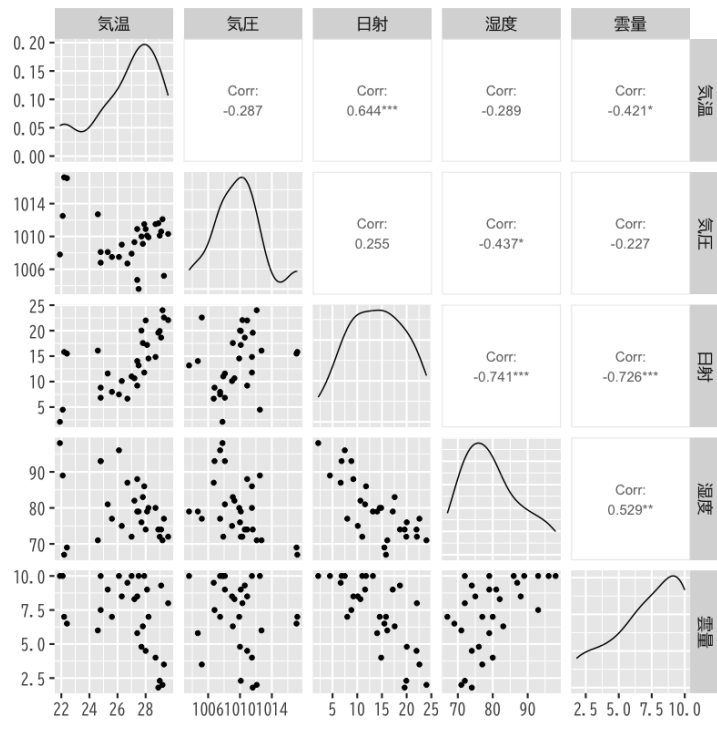


Figure 4: 散布図

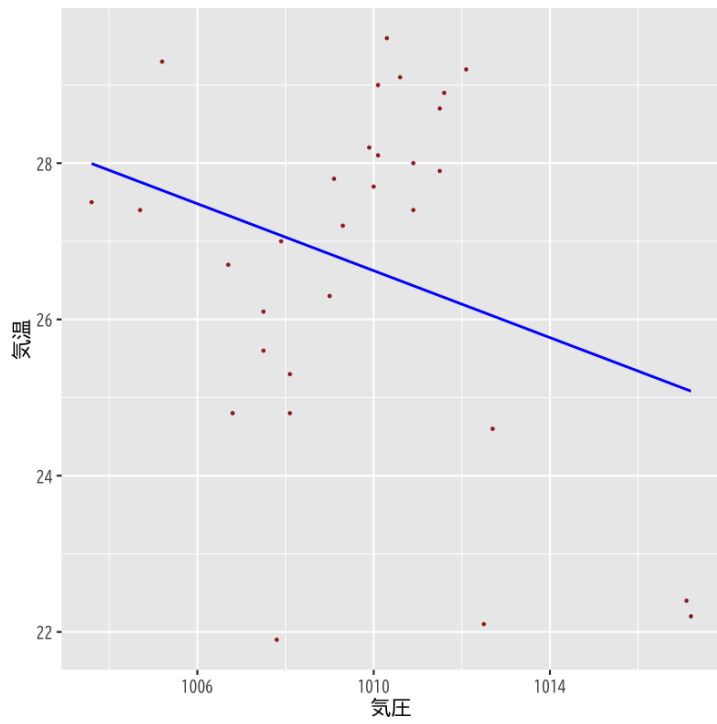


Figure 5: モデル 1

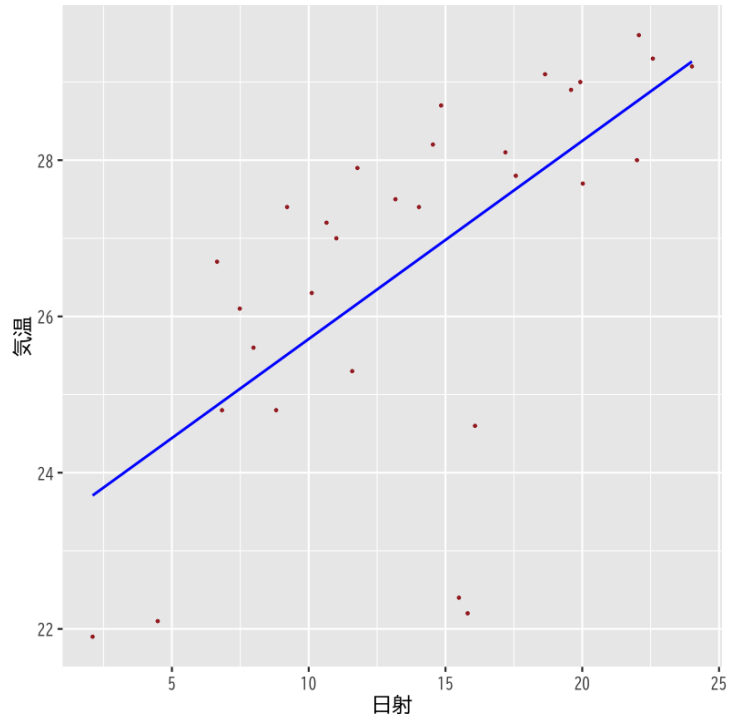


Figure 6: モデル 2

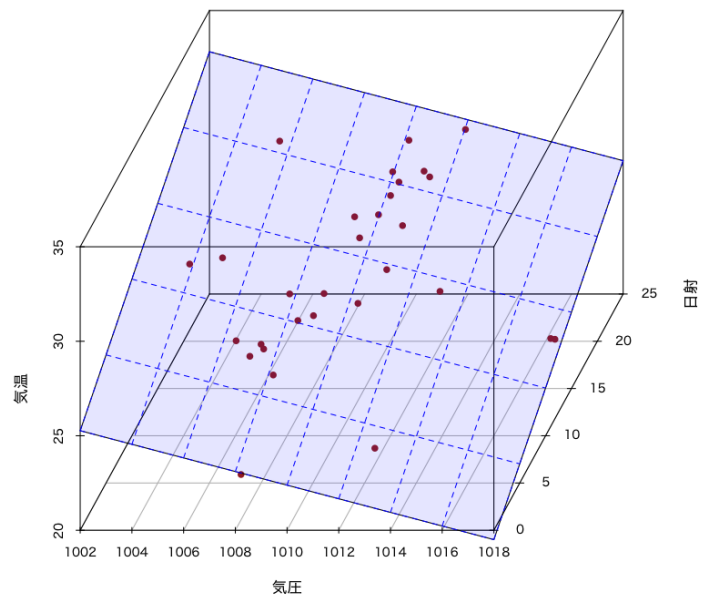


Figure 7: モデル 3

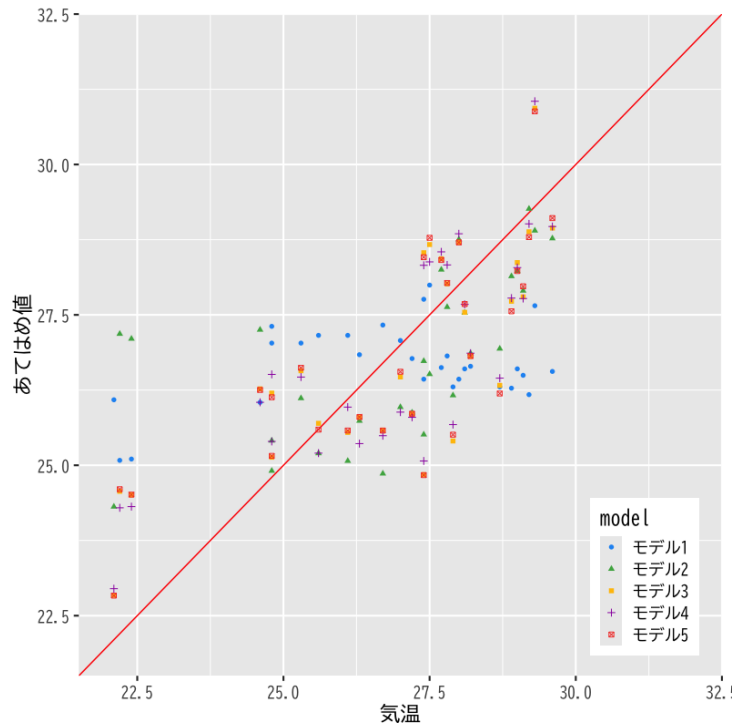


Figure 8: モデルの比較

| Characteristic | モデル 1 | | モデル 2 | | モデル 3 | | モデル 4 | | モデル 5 | |
|-------------------------|-------|-----------------|-------|-----------------|-------|-----------------|-------|-----------------|-------|-----------------|
| | Beta | SE ¹ | Beta | SE ¹ | Beta | SE ¹ | Beta | SE ¹ | Beta | SE ¹ |
| 気圧 | -0.21 | 0.135 | | | -0.36 | 0.090 | -0.32 | 0.098 | -0.36 | 0.092 |
| 日射 | | | 0.25 | 0.057 | 0.30 | 0.048 | 0.35 | 0.069 | 0.32 | 0.069 |
| 湿度 | | | | | | | 0.05 | 0.052 | | |
| 雲量 | | | | | | | | | 0.05 | 0.151 |
| R ² | 0.082 | | 0.414 | | 0.632 | | 0.644 | | 0.633 | |
| Adjusted R ² | 0.049 | | 0.393 | | 0.604 | | 0.603 | | 0.591 | |

¹SE = Standard Error

次回の予定

- 第 1 回 : 回帰モデルの考え方と推定
- 第 2 回 : モデルの評価
- 第 3 回 : モデルによる予測と発展的なモデル