回帰分析

モデルの評価

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

回帰分析の復習

線形回帰モデル

- **目的変数** を **説明変数** で説明する関係式を構成
  - 説明変数 : $x_1, \dots, x_p$ (p 次元)
  - 目的変数 : y (1 次元)
- **回帰係数** $\beta_0, \beta_1, \dots, \beta_p$ を用いた一次式

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

- **誤差項** を含む確率モデルで観測データを表現

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

簡潔な表現のための行列

- デザイン行列 (説明変数)

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

簡潔な表現のためのベクトル

- ベクトル (目的変数・誤差・回帰係数)

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

問題の記述

- 確率モデル

$$\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \text{確率分布}$$

- 回帰式の推定 : **残差平方和** の最小化

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

解の表現

- 解の条件 : **正規方程式**

$$X^\top X\boldsymbol{\beta} = X^\top \boldsymbol{y}$$

- 解の一意性 : **Gram 行列** $X^\top X$ が正則

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

最小二乗推定量の性質

- **あてはめ値** $\hat{\boldsymbol{y}} = X\hat{\boldsymbol{\beta}}$ は X の列ベクトルの線形結合
- **残差** $\hat{\boldsymbol{\epsilon}} = \boldsymbol{y} - \hat{\boldsymbol{y}}$ はあてはめ値 $\hat{\boldsymbol{y}}$ と直交

$$\hat{\boldsymbol{\epsilon}}^\top \hat{\boldsymbol{y}} = 0$$

- 回帰式は説明変数と目的変数の **標本平均** を通過

$$\bar{y} = (1, \bar{\boldsymbol{x}}^\top)\hat{\boldsymbol{\beta}}, \quad \bar{\boldsymbol{x}} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i, \quad \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i,$$

寄与率

- **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

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 (再掲)
  - 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

- データの概要

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

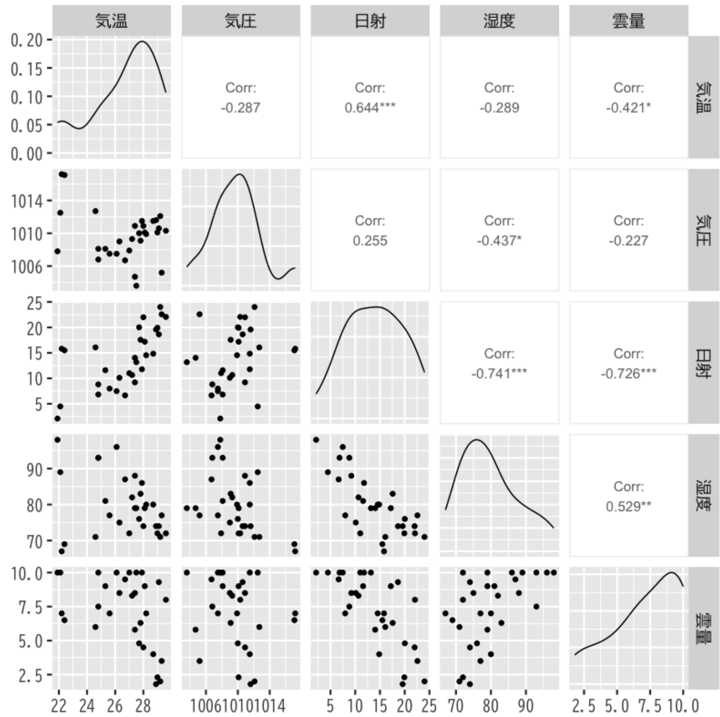

Figure 1: 散布図

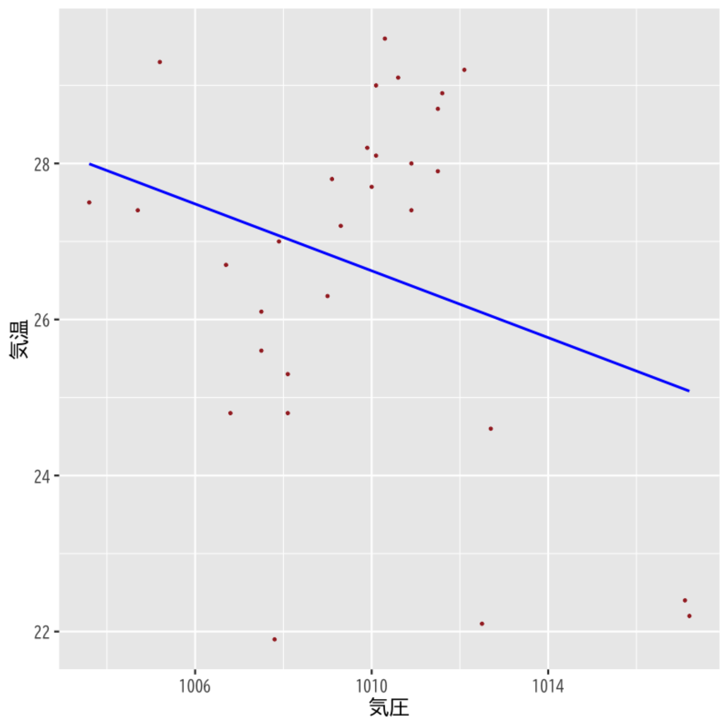

Figure 2: モデル 1

- モデル 2 の推定結果

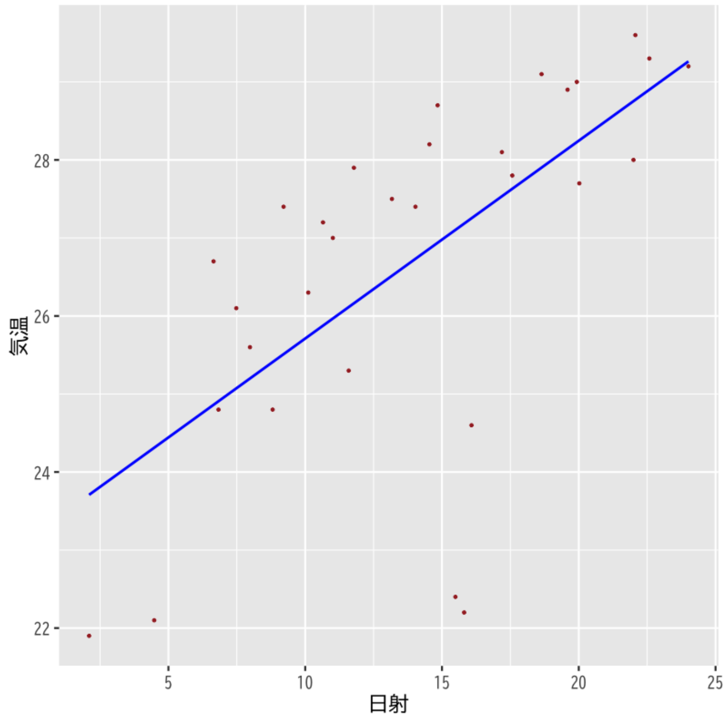

Figure 3: モデル 2

- 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)

	モデル 1		モデル 2		モデル 3		モデル 4		モデル 5	
Characteristic	**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	

[1]SE = Standard Error

あてはめ値の性質

あてはめ値

- さまざまな表現

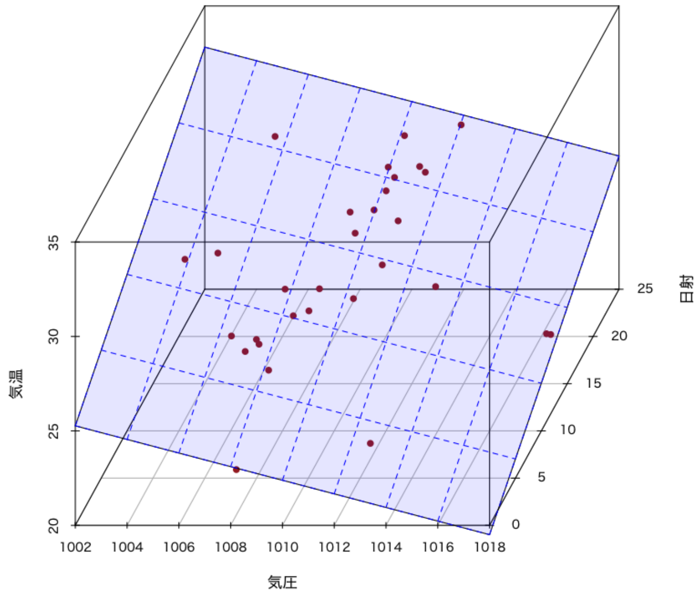

Figure 4: モデル 3

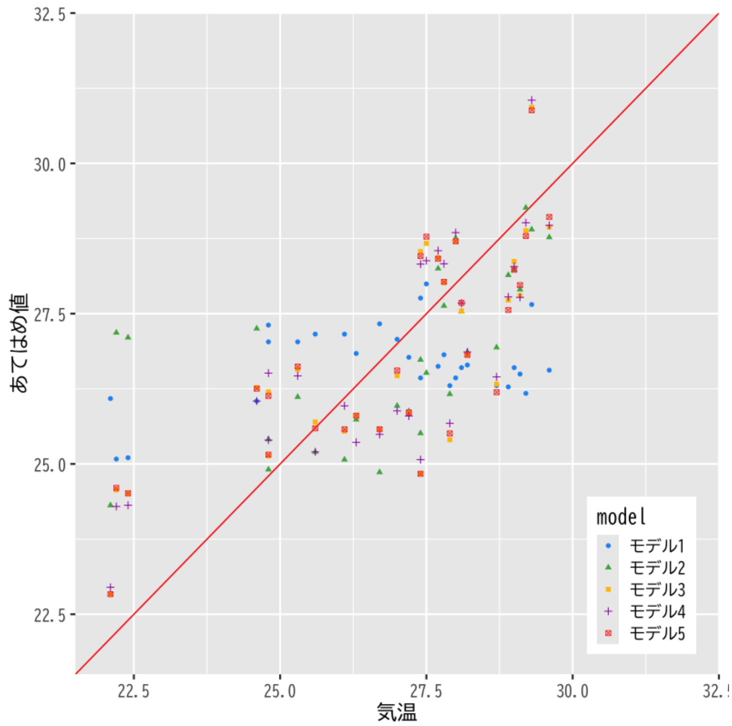

Figure 5: モデルの比較

$$
\begin{aligned}
\hat{\boldsymbol{y}} &= X\hat{\boldsymbol{\beta}} \\
&\qquad (\hat{\boldsymbol{\beta}} = (X^\top X)^{-1} X^\top \boldsymbol{y}\text{を代入}) \\
&= X(X^\top X)^{-1} X^\top \boldsymbol{y} && (A) \\
&\qquad (\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}\text{を代入}) \\
&= X(X^\top X)^{-1} X^\top X\boldsymbol{\beta} + X(X^\top X)^{-1} X^\top \boldsymbol{\epsilon} \\
&= X\boldsymbol{\beta} + X(X^\top X)^{-1} X^\top \boldsymbol{\epsilon} && (B)
\end{aligned}
$$

  - (A) あてはめ値は **観測値の重み付けの和** で表される
  - (B) あてはめ値と観測値は **誤差項** の寄与のみ異なる

あてはめ値と誤差

- 残差と誤差の関係

$$
\begin{aligned}
\hat{\boldsymbol{\epsilon}} &= \boldsymbol{y} - \hat{\boldsymbol{y}} \\
&= \boldsymbol{\epsilon} - X(X^\top X)^{-1} X^\top \boldsymbol{\epsilon} \\
&= \left(I - X(X^\top X)^{-1} X^\top\right)\boldsymbol{\epsilon} && (C)
\end{aligned}
$$

  - (C) 残差は **誤差の重み付けの和** で表される

ハット行列

- 定義

$$
H = X(X^\top X)^{-1} X^\top
$$

- ハット行列 H による表現

$$
\begin{aligned}
\hat{\boldsymbol{y}} &= H\boldsymbol{y} \\
\hat{\boldsymbol{\epsilon}} &= (I - H)\boldsymbol{\epsilon}
\end{aligned}
$$

  - あてはめ値や残差は H を用いて簡潔に表現される

ハット行列の性質

- 観測データ (デザイン行列) のみで計算される
- 観測データと説明変数の関係を表す
- 対角成分 (**テコ比**; leverage) は観測データが自身の予測に及ぼす影響の度合を表す

$$
\hat{y}_j = (H)_{jj} y_j + (\text{それ以外のデータの寄与})
$$

  - $(A)_{ij}$ は行列 A の (i, j) 成分
  - テコ比が小さい : 他のデータでも予測が可能
  - テコ比が大きい : 他のデータでは予測が困難

演習

問題

- ハット行列 H について以下を示しなさい
  - H は対称行列であること
  - H は冪等であること

$$H^2 = H, \quad (I - H)^2 = I - H$$

  - 以下の等式が成り立つこと

$$HX = X, \quad X^\top H = X^\top$$

ヒント

- いずれも H の定義にもとづいて計算すればよい

$$\begin{aligned}
H^\top &= (X(X^\top X)^{-1} X^\top)^\top \\
H^2 &= (X(X^\top X)^{-1} X^\top)(X(X^\top X)^{-1} X^\top) \\
(I - H)^2 &= I - 2H + H^2 \\
HX &= (X(X^\top X)^{-1} X^\top)X \\
X^\top H &= (HX)^\top
\end{aligned}$$

推定量の統計的性質

最小二乗推定量の性質

- 推定量と誤差の関係

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

- 正規分布の重要な性質 (**再生性**)

  正規分布に従う独立な確率変数の和は正規分布に従う

推定量の分布

- **誤差の仮定** : 独立，平均 0 分散 σ^2 の **正規分布**
- 推定量は以下の多変量正規分布に従う

$$\begin{aligned}
\mathbb{E}[\hat{\boldsymbol{\beta}}] &= \boldsymbol{\beta} \\
\mathrm{Cov}(\hat{\boldsymbol{\beta}}) &= \sigma^2 (X^\top X)^{-1}
\end{aligned}$$

$$\hat{\boldsymbol{\beta}} \sim \mathcal{N}(\boldsymbol{\beta}, \sigma^2 (X^\top X)^{-1})$$

演習

問題

- 誤差が独立で，平均 0 分散 σ^2 の正規分布に従うとき，最小二乗推定量 $\hat{\boldsymbol{\beta}}$ について以下を示しなさい
  - 平均は $\boldsymbol{\beta}$ (真の母数) となること
  - 共分散行列は $\sigma^2(X^\top X)^{-1}$ となること

解答例

- 定義にもとづいて計算する

$$
\begin{aligned}
\mathbb{E}[\hat{\boldsymbol{\beta}}] &= \mathbb{E}[\boldsymbol{\beta} + (X^\top X)^{-1} X^\top \boldsymbol{\epsilon}] \\
&= \boldsymbol{\beta} + (X^\top X)^{-1} X^\top \mathbb{E}[\boldsymbol{\epsilon}] \\
&= \boldsymbol{\beta}
\end{aligned}
$$

  - 定義にもとづいて計算する

$$
\begin{aligned}
\mathrm{Cov}(\hat{\boldsymbol{\beta}}) &= \mathbb{E}[(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^\top] \\
&= \mathbb{E}[(X^\top X)^{-1} X^\top \boldsymbol{\epsilon}\boldsymbol{\epsilon}^\top X (X^\top X)^{-1}] \\
&= (X^\top X)^{-1} X^\top \mathbb{E}[\boldsymbol{\epsilon}\boldsymbol{\epsilon}^\top] X (X^\top X)^{-1} \\
&= (X^\top X)^{-1} X^\top (\sigma^2 I) X (X^\top X)^{-1} \\
&= \sigma^2 (X^\top X)^{-1}
\end{aligned}
$$

誤差の評価

寄与率 (再掲)

- **決定係数 (R-squared)**
  - 回帰式で説明できるばらつきの比率

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

- **自由度調整済み決定係数 (adjusted R-squared)**
  - 決定係数を不偏分散で補正

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

各係数の推定量の分布

- 推定された回帰係数の精度を評価
  - 誤差 ϵ の分布は平均 0 分散 σ^2 の正規分布
  - $\hat{\boldsymbol{\beta}}$ の分布 : p+1 変量正規分布

$$
\hat{\boldsymbol{\beta}} \sim \mathcal{N}(\boldsymbol{\beta}, \sigma^2 (X^\top X)^{-1})
$$

- $\hat{\beta}_j$ の分布 : 1 変量正規分布

$$\hat{\beta}_j \sim \mathcal{N}(\beta_j, \sigma^2((X^\top X)^{-1})_{jj}) = \mathcal{N}(\beta_j, \sigma^2\zeta_j^2)$$

  - $(A)_{jj}$ は行列 A の (j, j) (対角) 成分

標準誤差

- **標準誤差** (standard error)
  - $\hat{\beta}_j$ の標準偏差の推定量

$$\text{s.e.}(\hat{\beta}_j) = \hat{\sigma}\zeta_j = \sqrt{\frac{1}{n-p-1}\sum_{i=1}^n \hat{\epsilon}_i^2} \cdot \sqrt{((X^\top X)^{-1})_{jj}}$$

  - 未知母数 σ^2 は不偏分散 $\hat{\sigma}^2$ で推定
  - $\hat{\beta}_j$ の精度の評価指標

演習

問題

- 以下を示しなさい
  - 不偏分散 $\hat{\sigma}^2$ が母数 σ^2 の不偏な推定量となる

    以下が成り立つことを示せばよい

$$\mathbb{E}\left[\sum_{i=1}^n \hat{\epsilon}_i^2\right] = (n-p-1)\sigma^2$$

解答例

- ハット行列 H を用いた表現を利用する

$$\begin{aligned}
\hat{\boldsymbol{\epsilon}} &= (I_n - H)\boldsymbol{\epsilon} \\
\mathbb{E}\left[\sum_{i=1}^n \hat{\epsilon}_i^2\right] &= \mathbb{E}[\hat{\boldsymbol{\epsilon}}^\top \hat{\boldsymbol{\epsilon}}] \\
&= \mathbb{E}[\mathrm{tr}(\hat{\boldsymbol{\epsilon}}\hat{\boldsymbol{\epsilon}}^\top)] \\
&= \mathbb{E}[\mathrm{tr}(I_n - H)\boldsymbol{\epsilon}\boldsymbol{\epsilon}^\top(I_n - H)] \\
&= \mathrm{tr}(I_n - H)\mathbb{E}[\boldsymbol{\epsilon}\boldsymbol{\epsilon}^\top](I_n - H) \\
&= \mathrm{tr}(I_n - H)(\sigma^2 I_n)(I_n - H) \\
&= \sigma^2\mathrm{tr}(I_n - H)
\end{aligned}$$

  - I_n は $n \times n$ 単位行列
- さらに以下が成立する

$$\begin{aligned}
\mathrm{tr}H &= \mathrm{tr}X(X^\top X)^{-1}X^\top \\
&= \mathrm{tr}(X^\top X)^{-1}X^\top X \\
&= \mathrm{tr}I_{p+1} \\
&= p + 1
\end{aligned}$$

  - 行列のサイズに注意

係数の評価

t 統計量

- **回帰係数の分布** に関する定理

  t **統計量** (t-statistic)

  $$t = \frac{\hat{\beta}_j - \beta_j}{\mathrm{s.e.}(\hat{\beta}_j)} = \frac{\hat{\beta}_j - \beta_j}{\hat{\sigma}\zeta_j}$$

  は自由度 $n{-}p{-}1$ の t 分布に従う
  - 証明には以下の性質を用いる
    - $\hat{\sigma}^2$ と $\hat{\boldsymbol{\beta}}$ は独立となる
    - $(\hat{\beta}_j - \beta_j)/(\sigma\zeta_j)$ は標準正規分布に従う
    - $(n{-}p{-}1)\hat{\sigma}^2/\sigma^2 = S(\hat{\boldsymbol{\beta}})/\sigma^2$ は自由度 $n{-}p{-}1$ の χ^2 分布に従う

t 統計量による検定

- 回帰係数 β_j が回帰式に寄与するか否かを検定
  - 帰無仮説 $H_0 : \beta_j = 0$ (t 統計量が計算できる)
  - 対立仮説 $H_1 : \beta_j \neq 0$
- p 値 : 確率変数の絶対値が $|t|$ を超える確率
  - $f(x)$ は自由度 $n{-}p{-}1$ の t 分布の確率密度関数

  $$(p\ 値) = 2\int_{|t|}^{\infty} f(x)dx \quad (両側検定)$$

  帰無仮説 H_0 が正しければ p 値は小さくならない

回帰係数の信頼区間

- 以下の確率変数は自由度 $n{-}p{-}1$ の t 分布に従う

  $$\frac{\hat{\beta}_j - \beta_j}{\mathrm{s.e.}(\hat{\beta}_j)} = \frac{(\hat{\beta}_j - \beta_j)/\sqrt{\mathrm{Var}(\hat{\beta}_j)}}{\sqrt{(n{-}p{-}1)\mathrm{s.e.}(\hat{\beta}_j)^2/(n{-}p{-}1)\mathrm{Var}(\hat{\beta}_j)}}$$

- $\gamma \in (0,1)$ に対する β_j の $1-\gamma$ 信頼区間

  $$\left[\hat{\beta}_j - t_{1-\gamma/2}(n{-}p{-}1)\cdot \mathrm{s.e.}(\hat{\beta}_j),\ \hat{\beta}_j + t_{1-\gamma/2}(n{-}p{-}1)\cdot \mathrm{s.e.}(\hat{\beta}_j)\right]$$

モデルの評価

F 統計量

- **ばらつきの比** に関する定理

  $\beta_1 = \cdots = \beta_p = 0$ ならば F **統計量** (F-statistic)

  $$F = \frac{\frac{1}{p}S_r}{\frac{1}{n-p-1}S} = \frac{n{-}p{-}1}{p}\frac{R^2}{1-R^2}$$

  は自由度 $p, n{-}p{-}1$ の F 分布に従う

- 証明には以下の性質を用いる
  - S_r と S は独立となる
  - S_r/σ^2 は自由度 p の χ^2 分布に従う
  - S/σ^2 は自由度 $n-p-1$ の χ^2 分布に従う

F 統計量を用いた検定

- 説明変数のうち 1 つでも役に立つか否かを検定
  - 帰無仮説 $H_0 : \beta_1 = \cdots = \beta_p = 0$ (S_r が χ^2 分布になる)
  - 対立仮説 $H_1 : \exists j\ \beta_j \neq 0$
- p 値 : 確率変数の値が F を超える確率
  - $f(x)$ は自由度 $p, n-p-1$ の F 分布の確率密度関数

$$(p\text{ 値}) = \int_F^\infty f(x)dx \quad (\text{片側検定})$$

帰無仮説 H_0 が正しければ p 値は小さくならない

解析の事例

気温に影響を与える要因の分析 (再掲)

- データの概要

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 (再掲)

- 関連するデータの散布図

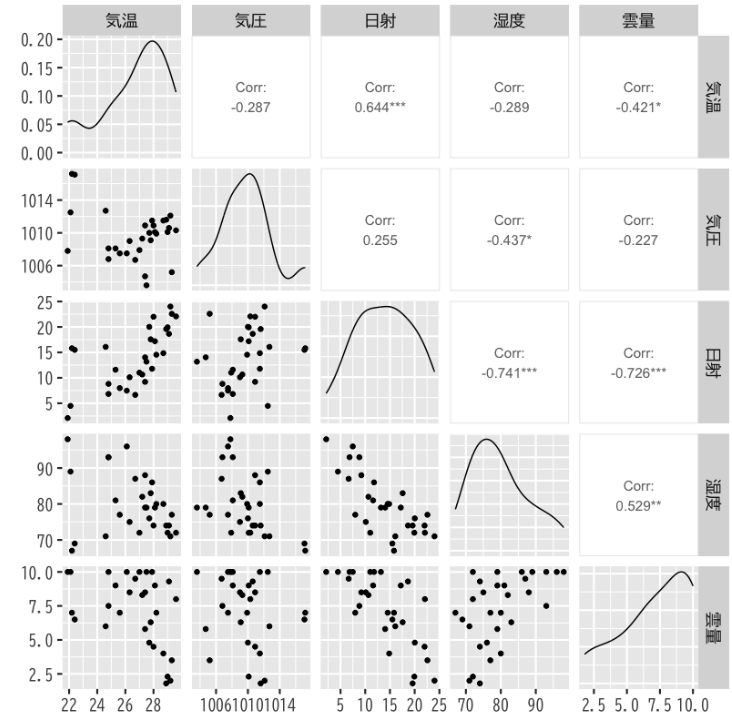

Figure 6: 散布図

- 観測値とあてはめ値の比較

モデルの評価

- t 統計量・F 統計量によるモデルの比較

	モデル 1				モデル 2				モデル 3			
Characteristic	**Beta**	**SE**[1]	**Statistic**	**p-value**	**Beta**	**SE**[1]	**Statistic**	**p-value**	**Beta**	**SE**[1]	**Statistic**	**p-value**
(Intercept)	243	137	1.78	0.086	23	0.855	27.1	<0.001	386	91.0	4.25	<0.001
気圧	-0.21	0.135	-1.58	0.12					-0.36	0.090	-3.99	<0.001
日射					0.25	0.057	4.45	<0.001	0.30	0.048	6.35	<0.001
Statistic	2.51				19.8				23.1			
p-value	0.12				<0.001				<0.001			

[1] SE = Standard Error

	モデル 4	モデル 5

Characteristic	Beta	SE[1]	Statistic	p-value	Beta	SE[1]	Statistic	p-value
(Intercept)	346	101	3.44	0.002	384	92.8	4.13	<0.001
気圧	-0.32	0.098	-3.32	0.003	-0.36	0.092	-3.90	<0.001
日射	0.35	0.069	5.05	<0.001	0.32	0.069	4.62	<0.001
湿度	0.05	0.052	0.948	0.4				
雲量					0.05	0.151	0.317	0.8
Statistic	15.7				14.9			
p-value	<0.001				<0.001			

[1]SE = Standard Error

- 様々な統計量によるモデルの比較

	モデル 1	モデル 2	モデル 3	モデル 4	モデル 5
Characteristic	Beta (95% CI)[1,2]	Beta (95% CI)[1,2]	Beta (95% CI)[1,2]	Beta (95% CI)[1,2]	Beta (95% CI)[1,2]
(Intercept)	243 (-37, 523)	23 (21, 25)***	386 (200, 573)***	346 (139, 553)**	384 (193, 575)***
気圧	-0.21 (-0.49, 0.06)		-0.36 (-0.55, -0.18)***	-0.32 (-0.53, -0.12)**	-0.36 (-0.55, -0.17)***
日射		0.25 (0.14, 0.37)***	0.30 (0.20, 0.40)***	0.35 (0.21, 0.49)***	0.32 (0.18, 0.46)***
湿度				0.05 (-0.06, 0.16)	
雲量					0.05 (-0.26, 0.36)
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
Statistic	2.51	19.8	23.1	15.7	14.9
p-value	0.12	<0.001	<0.001	<0.001	<0.001

[1]*p<0.05; **p<0.01; ***p<0.001
[2]CI = Confidence Interval

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**

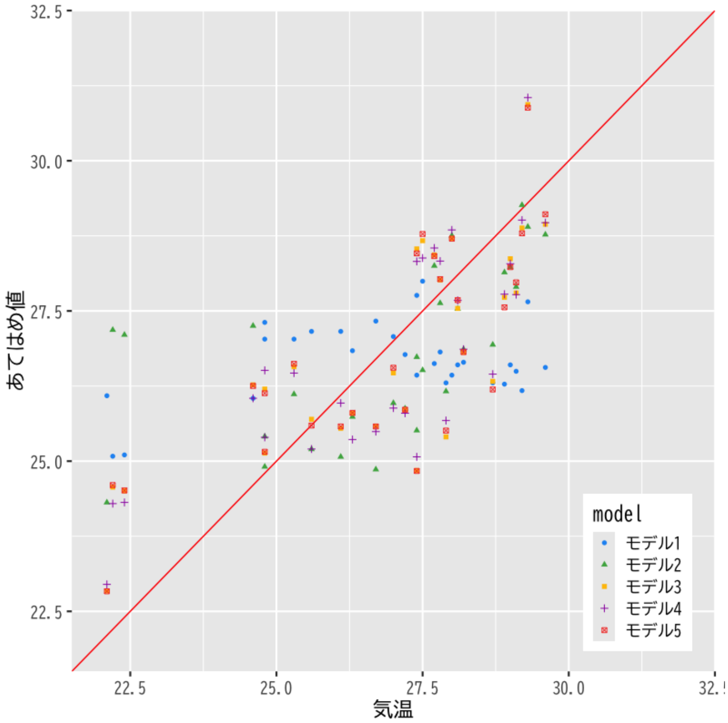

Figure 7: モデルの比較