数理統計 補助資料
～線形回帰モデル～

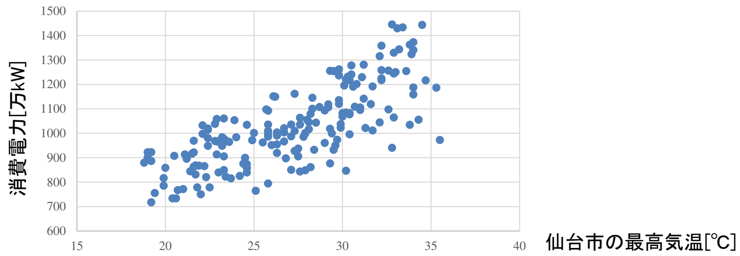
2024年度2学期:　月曜1限, 水曜3限

担当教員:　石垣　司

復習: 回帰分析の前に

基本は「散布図」
- 変数 x（最高気温）と y（消費電力）の相関関係の可視化

  この関係性を利用した予測や実証の手段が回帰分析



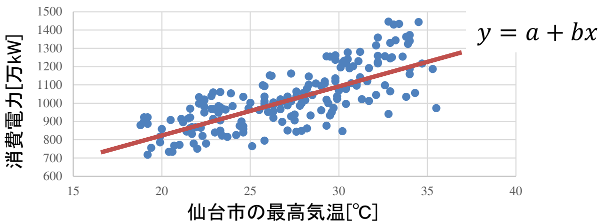
「回帰」とは, 目的変数 y の動きを, 別の説明変数 x と関数 f で予測したり説明したりすること

東北電力ネットワーク：東北6県・新潟エリアの2020&21年7月1日～9月30日の各日の12時から13時の電力使用量[万kW]
https://setsuden.nw.tohoku-epco.co.jp/download.html
気象庁：2020&21年7月1日～9月30日の各日の仙台市の最高気温
https://www.data.jma.go.jp/obd/stats/etrn/

復習: 線形単回帰モデル

関数 f に直線を仮定した説明変数が1つだけの回帰分析のためのモデル

$$y = f(x) = a + bx$$



- **因果関係がある場合は, x が原因, y が結果の表現**
- **データ$\{(x_1, y_1), \cdots, (x_N, y_N)\}$を用いて, 散布図の傾向に適合する直線の切片 a と傾き b を推定**
- **切片 a と傾き b が決まれば, 目的変数 y を予測できる**

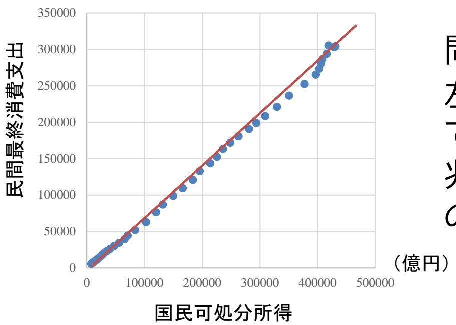
復習: 回帰分析と予測

回帰式 $y = \hat{a} + \hat{b}x$ を利用した予測

$\hat{a}$ と $\hat{b}$ はデータ $\{(x_1, y_1), \cdots, (x_N, y_N)\}$ から推定された係数

- **$\hat{y}$: 説明変数 x に対する目的変数 y の予測値**

$$\hat{y} = \hat{a} + \hat{b}x$$



1955年度～1998年度(1968SNA)
（内閣府　国民経済計算年次推計）

問題
左図の回帰係数は$\hat{a} \cong 0$, $\hat{b} \cong 0.7$である。国民可処分所得が500兆円のとき, 民間最終消費支出の予測値は？

復習：回帰係数の推定

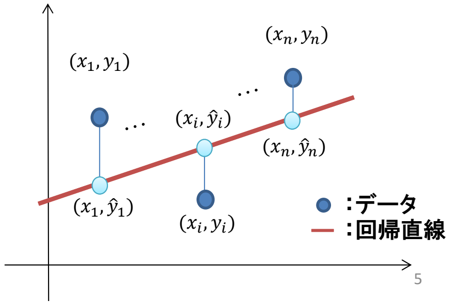
データから回帰係数 b と切片 a を決定する
- 合理的な基準と手続きに基づいた推定が必要

基準：残差平方和(RSS: Residual sum of squares)**の最小化**
- 残差 e_i　　$e_i = y_i - \hat{y}_i = y_i - (a + bx_i)$
- 残差平方和　$RSS = \sum_{i=n}^{n} e_i^2$

手続き：最小2乗法

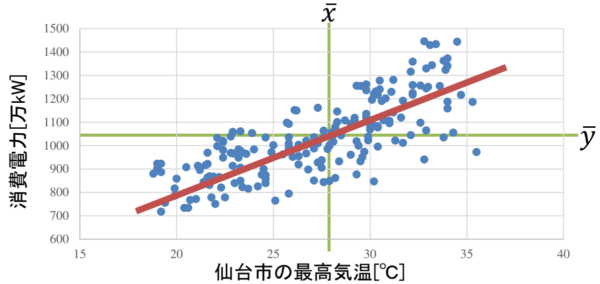
最小2乗推定量　check!

$$\hat{b} = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2},\ \hat{a} = \bar{y} - \hat{b}\bar{x}$$



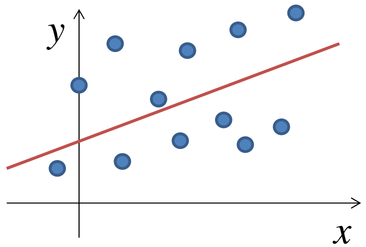
復習：推定された回帰直線の性質

最小2乗法で推定された回帰直線が満たす性質
1. 推定された回帰直線は$(\bar{x}, \bar{y})$を通る
2. $\sum_{i=1}^{n} e_i = 0$（残差の和は0）
3. $\sum_{i=1}^{n} e_i x_i = 0$（残差と説明変数 x の積和は0）
   残差と説明変数のベクトルは直交する

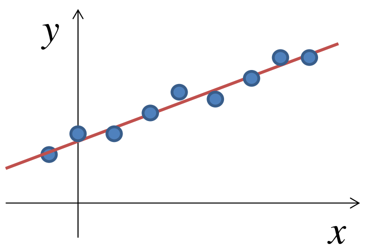


復習：決定係数 R^2

単回帰分析の回帰式の適合度(goodness of fit)**の指標**
- 同じ回帰式でもデータの説明力が異なる



説明力が低い回帰直線 ($R^2 \approx 0$)



説明力が高い回帰直線 ($R^2 \approx 1$)

異なるデータから推定された同じ回帰直線

決定係数 R^2　$(R^2 \leq 1)$
- 適合度が高いと1に近く, 低いとゼロに近い

復習：決定係数 R^2 の定義と意味

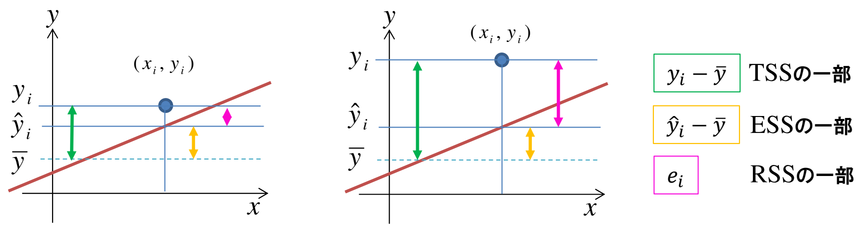
決定係数 R^2 の定義

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$$

全変動(TSS: total sum of squares)
回帰変動(ESS: explained SS)
残差平方和(RSS: residual SS)

- 標本分散の分解　check!

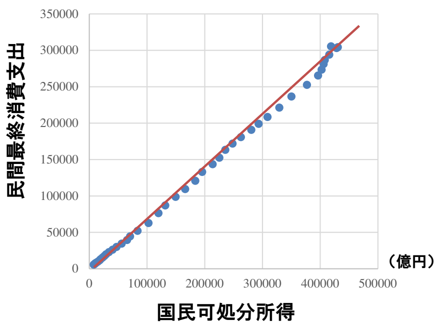
$$\underbrace{\sum_{i=1}^{n}(y_i - \bar{y})^2}_{\text{TSS}} = \underbrace{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}_{\text{ESS}} + \underbrace{\sum_{i=1}^{n} e_i^2}_{\text{RSS}}$$



適合度が高いデータの例　　適合度が低いデータの例

復習：決定係数の例～所得と消費

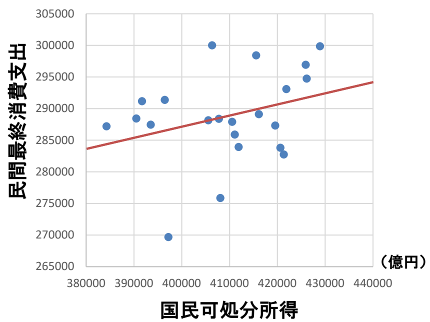
国民可処分所得と民間最終消費支出（内閣府　国民経済計算年次推計）



1955年度～1998年度(1968SNA)

$\hat{a} = -2950, \hat{b} = 0.698, R^2 = 0.998$

消費額の全変動の99.8%は国民可処分所得で説明可能



1994年度～2015年度(2008SNA)

$\hat{a} = 215600, \hat{b} = 0.178, R^2 = 0.099$

消費額の全変動の約10%は国民可処分所得で説明可能

$\hat{a}$	基礎消費、
$\hat{b}$	限界消費性向

線形重回帰分析

複数の説明変数による回帰分析

- 目的変数：　変数 y
- 説明変数：　変数 $x_1, x_2, \cdots, x_P$
- データ：　$\{y_i, x_{i1}, x_{i2}, \cdots, x_{iP}\}(i = 1, \cdots, N)$
- 偏回帰係数：　係数 $b_1, b_2, \cdots, b_P$（パラメータ）
- 切片：　係数 b_0（パラメータ）

重回帰式

$$y = b_0 + b_1x_1 +, \cdots, + b_Px_P$$

重回帰モデル（データによる記述）

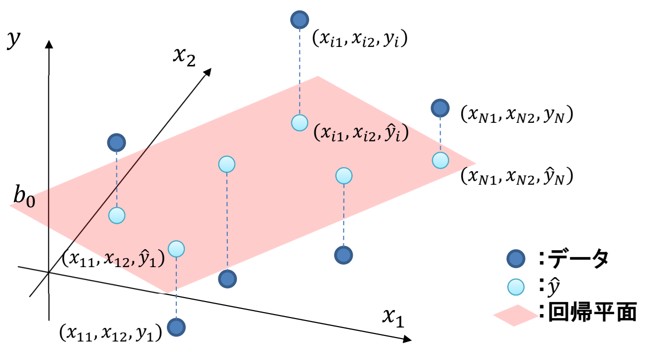
$$y_i = b_0 + b_1x_{i1} +, \cdots, + b_Px_{iP} + e_i \ \ (i = 1, \cdots N)$$

重回帰モデルのイメージ

2つの説明変数による重回帰モデルのイメージ

- 推定された重回帰モデルによる予測値

$$\hat{y}_i = \hat{b}_0 + \hat{b}_1x_{i1} +, \cdots, + \hat{b}_Px_{iP} + e_i \ \ (i = 1, \cdots N)$$



#メモ　ここではまだ誤差項e_iの分布に正規分布を仮定する必要がないことに注意

線形回帰モデルと行列

線形重回帰モデル

$$\begin{cases} y_1 = b_0 + b_1x_{11} +, \cdots, + b_Px_{1P} + e_1 \\ \quad\vdots \\ y_i = b_0 + b_1x_{i1} +, \cdots, + b_Px_{iP} + e_i \\ \quad\vdots \\ y_N = b_0 + b_1x_{N1} +, \cdots, + b_Px_{NP} + e_N \end{cases}$$

$$\boldsymbol{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_i \\ \vdots \\ y_N \end{bmatrix} \quad X = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1P} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{i1} & \ddots & x_{iP} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{N1} & \cdots & x_{NP} \end{bmatrix} \quad \boldsymbol{b} = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_p \\ \vdots \\ b_P \end{bmatrix} \quad \boldsymbol{e} = \begin{bmatrix} e_1 \\ \vdots \\ e_i \\ \vdots \\ e_N \end{bmatrix}$$

線形回帰モデルの行列表現

$$\boldsymbol{y} = X\boldsymbol{b} + \boldsymbol{e}$$

偏回帰係数の推定

最小2乗法による残差平方和の最小化

– 残差ベクトル $\boldsymbol{e} = \boldsymbol{y} - \hat{\boldsymbol{y}} = \boldsymbol{y} - X\boldsymbol{b}$

– 残差平方和 $\mathrm{RSS} = \sum_{i=n}^{n} e_i^2 = \boldsymbol{e}^T\boldsymbol{e} = (\boldsymbol{y} - X\boldsymbol{b})^T(\boldsymbol{y} - X\boldsymbol{b})$

線形回帰モデルの最小2乗推定量

$$\hat{\boldsymbol{b}} = (X^TX)^{-1}X^T\boldsymbol{y}$$

正規方程式とよばれる

– 最小2乗推定量 $\{\hat{b}_p\}_{p=1,\dots,P}$ は偏回帰係数と呼ばれる

偏相関係数と同様に, p 番目の説明変数以外の影響を取り除いた場合の, 目的変数 y と説明変数 x_p の単回帰係数に等しい

#メモ 推定量 $\hat{b}_p$ は X と y の各要素から影響を受けるため, 行列とベクトルを使用しないと一般的な表記が困難。早い段階（学部1,2年生）から行列とベクトルの取り扱いに慣れましょう

自由度調整済み決定係数

定義：

$$\text{adjusted } R^2 = 1 - \frac{\frac{RSS}{N-P-1}}{\frac{TSS}{N-1}}, \qquad R^2 = 1 - \frac{\frac{RSS}{N-1}}{\frac{TSS}{N-1}}$$

自由度調整済み決定係数　　決定係数

– 決定係数は説明変数の種類 (P) の増加で RSS が単調減少

説明変数をたくさん利用すると, R^2 は 1 に近い値をとってしまう

– 自由度調整済み決定係数は, P の増加による RSS の減少を補正している

– 重回帰分析では, 決定係数ではなく, 自由度調整済み決定係数で回帰の当てはまりの良さを

回帰分析の結果の幾何学的理解～内積と正射影

内積の定義：2つの N 次元ベクトル $\boldsymbol{a}, \boldsymbol{b}$ のなす角が θ のとき

$$\boldsymbol{a}\cdot\boldsymbol{b} = \|\boldsymbol{a}\|\|\boldsymbol{b}\|\cos\theta = a_1b_1 + a_2b_2 + \cdots + a_Nb_N$$

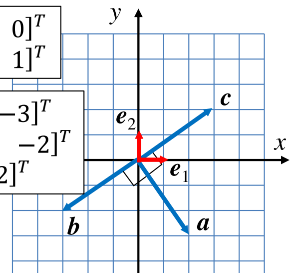
内積の性質

– $\boldsymbol{a}\cdot\boldsymbol{b} = 0 \Leftrightarrow \boldsymbol{a}, \boldsymbol{b}$ は直交する

– $\boldsymbol{a}\cdot\boldsymbol{b} = \tilde{\boldsymbol{a}}\cdot\boldsymbol{b}$（$\tilde{\boldsymbol{a}}$ は $\boldsymbol{a}$ から $\boldsymbol{b}$ 上への正射影ベクトル）

$\boldsymbol{e}_1 = [1 \quad 0]^T$
$\boldsymbol{e}_2 = [0 \quad 1]^T$

$\boldsymbol{a} = [2 \quad -3]^T$
$\boldsymbol{b} = [-3 \quad -2]^T$
$\boldsymbol{c} = [3 \quad 2]^T$

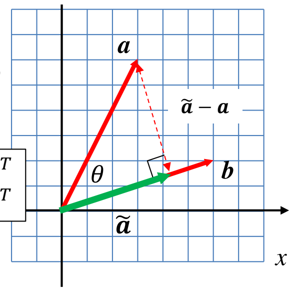


直交の例：$\boldsymbol{e}_1\cdot\boldsymbol{e}_2 = 0, \ \boldsymbol{a}\cdot\boldsymbol{b} = \boldsymbol{a}\cdot\boldsymbol{c} = 0$

ベクトル $\boldsymbol{a}$ の指す座標点とベクトル $\boldsymbol{b}$ 上の点で最も距離が小さいのは正射影ベクトル $\tilde{\boldsymbol{a}}$ の指す座標点

$$\tilde{\boldsymbol{a}} = \frac{\boldsymbol{a}\cdot\boldsymbol{b}}{\|\boldsymbol{b}\|^2}\boldsymbol{b}$$

$\boldsymbol{a} = [3 \quad 6]^T$
$\boldsymbol{b} = [6 \quad 2]^T$

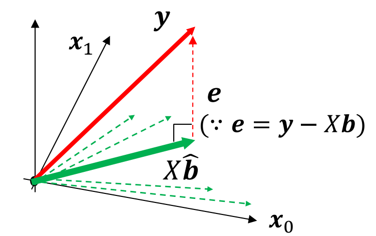


回帰分析の結果の幾何学的理解

最小2乗推定の結果は, N 次元空間内のベクトル $\boldsymbol{y}$ との距離が最も小さくなる P 次元空間 $(N > P)$ 内のベクトル $X\boldsymbol{b}$ を見つけることと同じである（X はデータとして与えられているので, ↑を満たす $\boldsymbol{b}$ を探すことと同じ）

例：$\boldsymbol{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_i \\ \vdots \\ y_N \end{bmatrix}, X = [\boldsymbol{x}_0 \quad \boldsymbol{x}_1] = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_i \\ \vdots & \vdots \\ 1 & x_N \end{bmatrix}, \boldsymbol{b} = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix}, \boldsymbol{e} = \begin{bmatrix} e_1 \\ \vdots \\ e_i \\ \vdots \\ e_N \end{bmatrix}$

#メモ より正確に理解するためには線形ベクトル空間、基底、部分空間を張るベクトル、直交などの概念を復習しよう。少し難しいが、より一般的にはヒルベルト空間での直交射影定理により、距離が最小の点はただ一つに定まることが分かる



$X\boldsymbol{b}$（- - -→）は $\boldsymbol{x}_0 - \boldsymbol{x}_1$ 平面上を動く

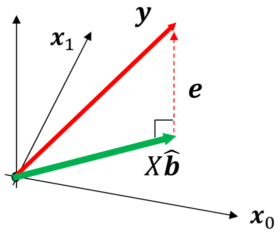
【図の意味の説明】

1. $\boldsymbol{y}, \boldsymbol{x}_0, \boldsymbol{x}_1$ は N 次元空間を動くベクトル
2. $X\boldsymbol{b} = b_0\boldsymbol{x}_0 + b_1\boldsymbol{x}_1$ なので, ベクトル $X\boldsymbol{b}$ は $\boldsymbol{x}_0$ と $\boldsymbol{x}_1$ で張られる平面内のみを動く
3. 最小2乗法は $\boldsymbol{e}^2$ を最小化。これは $\boldsymbol{e}$ の大きさ, つまり, $\boldsymbol{y}$ と $X\boldsymbol{b}$ の距離の最小化と等しい
4. $\boldsymbol{y}$ と $X\boldsymbol{b}$ の距離が最小のとき, ベクトル $\boldsymbol{e}$ と $\boldsymbol{x}_0 - \boldsymbol{x}_1$ 平面は直交する。言い換えると, ベクトル $\boldsymbol{y}$ の正射影ベクトルは $X\hat{\boldsymbol{b}}$ となる

推定された回帰直線の性質

最小2乗法で推定された回帰直線が満たす性質

1. **推定された回帰直線は**$(\bar{x}, \bar{y})$**を通る**
2. $\sum_{i=1}^{n} e_i = 0$ **（残差の和は**0**）**
3. $\sum_{i=1}^{n} e_i x_i = 0$ **（残差と説明変数** x **の積和は**0**）**
   残差と説明変数のベクトルは直交する



- **ベクトル** $\boldsymbol{e}$ **と ベクトル** $\boldsymbol{x}_0 = [1,1,\dots,1]^T$ **が直交するため**,「2. $\sum_{i=1}^{n} e_i = 0$」**が成立**
- **ベクトル** $\boldsymbol{e}$ **と ベクトル** $\boldsymbol{x}_1 = [x_1,\dots,x_N]^T$ **が直交するため**,「3. $\sum_{i=1}^{n} e_i x_i = 0$」**が成立**

演習問題

線形回帰モデルの最小2乗推定量が

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

となることを示しなさい。

ヒント：

- **残差平方和** RSS $= (\boldsymbol{y} - X\boldsymbol{b})^T(\boldsymbol{y} - X\boldsymbol{b})$ **は**2**次式。**
  ベクトル $\boldsymbol{b}$ **に対する** RSS **の最小化のためには・・・**
- **転置行列の公式：** $(A+B)^T = A^T + B^T$, $(AB)^T = B^T A^T$
- **ベクトルの微分の公式**1**：** $f(\boldsymbol{x}) = \boldsymbol{a}^T\boldsymbol{x}$ **のとき** $\frac{df(\boldsymbol{x})}{d\boldsymbol{x}} = \boldsymbol{a}$
- **ベクトルの微分の公式**2**：** A **が対称行列のとき**
  $f(\boldsymbol{x}) = \boldsymbol{x}^T A \boldsymbol{x}$ **のとき** $\frac{df(\boldsymbol{x})}{d\boldsymbol{x}} = 2A\boldsymbol{x}$