

# 統計学入門 補助資料 ～単回帰分析～

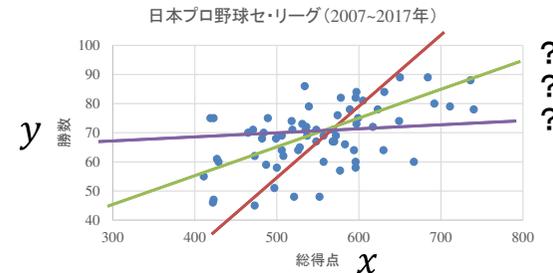
2022年度1学期: 月曜2限  
担当教員: 石垣 司

## 単回帰分析

- **統計的予測・計量経済学の基礎**
- **2変数  $(x, y)$  の関係に1次式を仮定**

$$y = a + bx$$

- $y$  を  $x$  の関数とみなす
  - 散布図のデータの散らばりに適合する直線(切片  $a$ , 傾き  $b$ )を推定
  - $x$  と  $y$  に因果関係がある場合は,  $x$  が原因,  $y$  が結果の表現



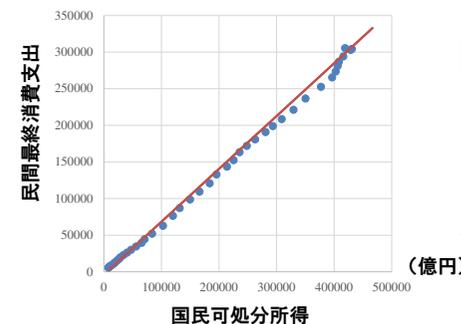
どの直線が“良い”直線?

## 単回帰分析の用語の整理

- **変数に関する用語**
  - 説明変数 変数  $x$
  - 従属変数 変数  $y$  (目的変数、被説明変数)
  - 回帰係数 係数  $b$  (パラメータ)
- **回帰に関する用語**
  - 回帰 ある変数を他の変数の関数で表現すること
  - 回帰式  $y = a + bx$  (単回帰式)
  - 回帰直線 回帰式が表現する直線
  - 単回帰分析 1つの説明変数を用いた回帰による分析
  - 重回帰分析 複数の説明変数を用いた回帰による分析

## 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  $\{(x_1, y_1), \dots, (x_n, y_n)\}$  から推定された係数
    - $\hat{y}_{n+1}$ :  $x_{n+1}$  に対する 変数  $y$  の予測値
- $$\hat{y}_{n+1} = \hat{a} + \hat{b}x_{n+1}$$



**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500兆円するとき、民間最終消費支出の予測値は?

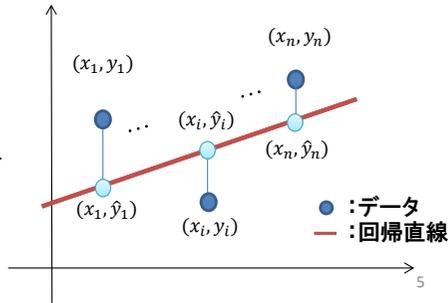
1955年度~1998年度(1968SNA)  
(内閣府 国民経済計算年次推計)

# 回帰式の推定

- 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  $RSS = \sum_{i=1}^n e_i^2$
- 手続き 最小2乗法

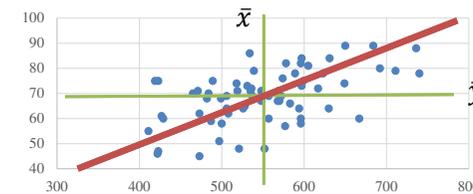
## 最小2乗推定量 check!

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



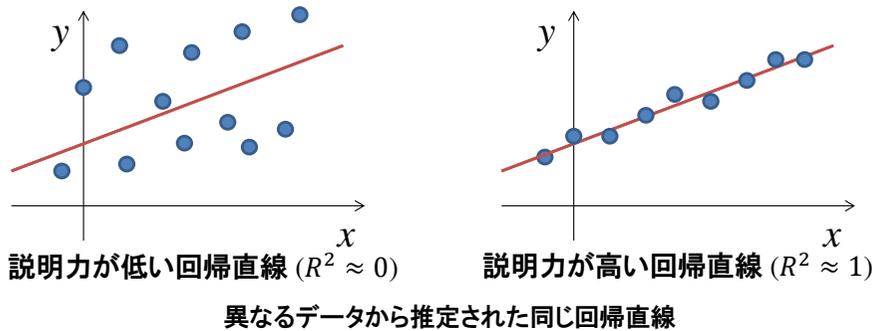
# 推定された回帰式が満たす性質

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



# 決定係数 $R^2$

- 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



- 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

# 決定係数 $R^2$ の定義と意味

- 決定係数  $R^2$  の定義

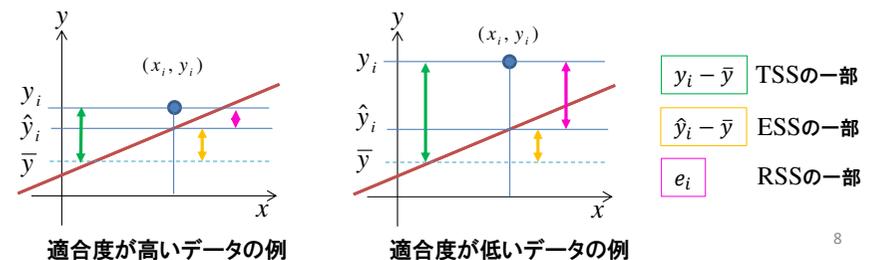
$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$$

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

- 標本分散の分解

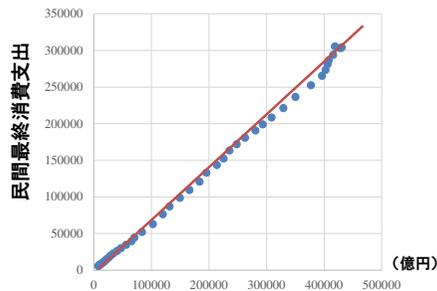
$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n e_i^2$$

TSS                      ESS                      RSS



# 決定係数の例～所得と消費

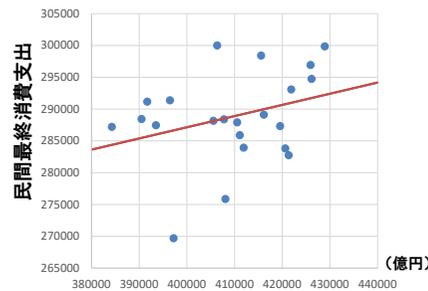
## 国民可処分所得と民間最終消費支出(内閣府 国民経済計算年次推計)



国民可処分所得  
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



国民可処分所得  
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

# 補足: 偏相関係数～計算方法

## 標本偏相関係数の定義

– 回帰により  $z$  の影響を除去した  $x$  の残差

$$u_i = x_i - (\hat{c} + \hat{d}z_i)$$

– 回帰により  $z$  の影響を除去した  $y$  の残差

$$v_i = y_i - (\hat{e} + \hat{f}z_i)$$

–  $z$  の影響を除去した  $x$  と  $y$  の標本偏相関係数

$$r_{xy,z} = \frac{\sum_{i=1}^n u_i v_i}{\sqrt{\sum_{i=1}^n u_i^2 \sum_{i=1}^n v_i^2}} = \frac{r_{xy} - r_{xz}r_{yz}}{\sqrt{(1 - r_{xz}^2)(1 - r_{yz}^2)}}$$

# 演習問題

## ある商品に関する広告費( $x$ ), 売上数( $y$ ), 価格( $z$ )の各データの代表値は次であった

- 標本平均  $\bar{x} = 80, \bar{y} = 100, \bar{z} = 50$
- 標本分散  $S_{xx} = 16, S_{yy} = 16, S_{zz} = 1$
- 標本共分散  $S_{xy} = 12, S_{xz} = -2, S_{yz} = -3$
- 標本相関係数  $r_{xz} = -0.5, r_{yz} = -0.75$

## 問題

1.  $x$  を説明変数,  $y$  を目的変数とした単回帰分析により, 広告費が 120 の時の売上数を予測しなさい
2. 相関係数  $r_{xy}$  の値を求めなさい
3.  $z$  の影響を削除した偏相関係数  $r_{xy,z}$  の値を求めなさい