

平成22年度 統計工学

平成22年度  
統計工学 講義資料(13)

～変数間の関係性:回帰分析～

(C) Sadami SUZUKI

平成22年度 統計工学

2変数に関する情報:相関係数

以下のデータは、ある店舗における製品出荷量と、その日の予想最高気温ならびに予想最低気温の関係を示したものである。

| 番号 | 出荷量  | 予想最高気温 | 予想最低気温 |
|----|------|--------|--------|
| 1  | 240  | 29     | 24     |
| 2  | 223  | 27     | 22     |
| 3  | 254  | 29     | 25     |
| 4  | 246  | 30     | 28     |
| 5  | 271  | 31     | 24     |
| 6  | 274  | 35     | 29     |
| 7  | 266  | 33     | 25     |
| 8  | 239  | 31     | 26     |
| 9  | 235  | 32     | 26     |
| 10 | 230  | 28     | 22     |
| 11 | 217  | 25     | 24     |
| 合計 | 2695 | 330    | 275    |
| 平均 | 245  | 30     | 25     |

最高気温  $y$

最低気温  $z$

出荷量  $x$

$r_{xy} = 0.809$   
( $p = 0.003$ )

$r_{yz} = 0.694$   
( $p = 0.018$ )

$r_{xz} = 0.525$   
( $p = 0.097$ )

製品出荷量も最低気温も最高気温の影響を受ける  
□ 最高気温の影響を取り除いたときの最低気温と出荷量の関係は？  
製品出荷量も最高気温も最低気温の影響を受ける  
□ 最高気温の影響を取り除いたときの最低気温と出荷量の関係は？

…偏相関

(C) Sadami SUZUKI

平成22年度 統計工学

偏相関係数

【相関】  
 $r_{xy} = 0.809$   
( $p = 0.003$ )  
 $r_{yz} = 0.694$   
( $p = 0.018$ )  
 $r_{xz} = 0.525$   
( $p = 0.097$ )

【偏相関】  
これらの影響を取り除く  
 $y$ の影響を取り除いた  $x$  と  $z$  の偏相関

偏相関係数  
 $y$ の影響を取り除いた  $x$  と  $z$  の偏相関係数  
$$r'_{xz} = \frac{r_{xz} - (r_{xy} \cdot r_{yz})}{\sqrt{1 - r_{xy}^2} \cdot \sqrt{1 - r_{yz}^2}}$$

(C) Sadami SUZUKI

平成22年度 統計工学

回帰:変数間の関係性のモデル化

予想最高気温と出荷量の間の関係性を明示的に示すことができれば、  
出荷量の予測がたてられる…  
- 予想最高気温の関数として出荷量を捉える

目的変数 ……予測の対象となる変数(結果系)  
説明変数 ……目的変数の予測に用いる変数(原因系)

回帰  
- 説明変数の関数として目的変数を捉えること  
- 線形回帰…説明変数の一次式によるモデル化

出荷量  $y$   
200 220 240 260 280 300  
 $y = a + bx$   
20 25 30 35 40  
予想最高気温  $x$

(C) Sadami SUZUKI

平成22年度 統計工学

回帰分析 (Regression Analysis)

単回帰モデルにみる回帰式導出メカニズム

回帰式  
$$y = a + b x$$
  
目的変数 切片 回帰係数 説明変数

$y_i = a + bx_i + \epsilon_i$   
観測値  
 $\hat{y}_i = a + bx_i$   
予測値  
残差:観測値と予測値とのスレ

全サンプルデータにおける残差をなるべく小さく抑えたい…  
残差平方和に着目

(C) Sadami SUZUKI

平成22年度 統計工学

残差平方和の最小化による回帰式の決定

残差平方和  
$$S = \sum \epsilon_i^2 = \sum (y_i - \hat{y}_i)^2 = \epsilon_1^2 + \epsilon_2^2 + \dots + \epsilon_n^2$$
$$= \{y_1 - (a + bx_1)\}^2 + \{y_2 - (a + bx_2)\}^2 + \dots + \{y_n - (a + bx_n)\}^2$$
$$= \sum \{y_i - (a + bx_i)\}^2$$

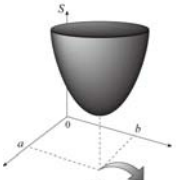
全サンプルデータにおける残差をなるべく小さく抑えたい…  
残差平方和  $S$  の最小値を与える  $(a, b)$  を求めよう！

(C) Sadami SUZUKI

平成22年度 統計工学

残差平方和の最小化による回帰式の決定

- <最小二乗法>
  - 残差平方和  $S$  の最小値を与える  $(a,b)$  を求める方法



$$(1) \frac{\partial S}{\partial a} = -2 \sum (y_i - a - bx_i) = 0$$
$$(2) \frac{\partial S}{\partial b} = -2 \sum x_i (y_i - a - bx_i) = 0$$

$S$  を  $a, b$  それぞれについて偏微分することによって得られる連立方程式の解として...

$$\therefore b = \frac{\hat{\sigma}_{xy}}{\hat{\sigma}_x^2}, \quad a = \bar{y} - b\bar{x}$$
$$\hat{\sigma}_{xy} = \frac{1}{n-1} \sum (x_i - \bar{x})(y_i - \bar{y}), \quad \hat{\sigma}_x^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2$$

<  $x$  と  $y$  の共分散 > <  $x$  の(不偏)分散 >

(C) Sadami SUZUKI

平成22年度 統計工学

回帰係数(共分散・不偏分散)計算の実際

- 実際に回帰係数を求める際には、次のようにすると便利(桁落ち防止)

1. 偏差平方和の計算

$$S_x = \sum (x_i - \bar{x})^2 = \sum x_i^2 - n\bar{x}^2$$

2. 偏差積和の計算

$$S_{xy} = \sum (x_i - \bar{x})(y_i - \bar{y}) = \sum x_i y_i - n\bar{x}\bar{y}$$

3. 回帰係数・切片の計算

$$b = \frac{S_{xy}}{S_x}$$
$$a = \bar{y} - b\bar{x}$$

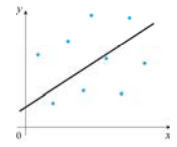
(C) Sadami SUZUKI

平成22年度 統計工学

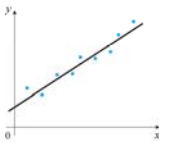
回帰の精度・当てはまり具合

- 回帰式はデータの当てはまり具合に係らず機械的に求めることが可能
  - 下に示す2つの図を比較すると...

【回帰の精度は低い】



【回帰の精度は高い】



回帰の精度を表す指標が必要！

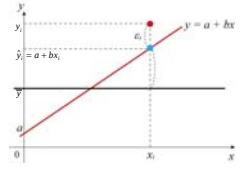
(C) Sadami SUZUKI

平成22年度 統計工学

バラツキの分解

バラツキの分解

- (観測値のバラツキ) = (予測値のバラツキ) + (残差のバラツキ)
  - 残差のバラツキが小さいほど観測値のバラツキは予測値のバラツキを用いて説明できることが分かる
- (総平方和) = (回帰平方和) + (残差平方和)  
= (回帰による変動) + (回帰からの変動)


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

平方和分解

$$S_T = S_R + S_E$$

【総平方和】 【回帰による変動】 【回帰からの変動】

(C) Sadami SUZUKI

平成22年度 統計工学

決定係数(寄与率)

- 回帰式精度・当てはまり具合(説明力の尺度)は...
  - 回帰からの変動(誤差部分のバラツキ)が回帰による変動(説明変数で説明できる部分)に比べて小さいことが望ましい

$$S_T = S_R + S_E$$

【回帰による変動】 【回帰からの変動】

$$R^2 = \frac{\text{回帰による変動 } S_R}{\text{回帰による変動 } S_R + \text{回帰からの変動 } S_E}$$
$$= \frac{\text{予測値の分散 } \hat{\sigma}_y^2}{\text{予測値の分散 } \hat{\sigma}_y^2 + \text{残差の分散 } \sigma_e^2}$$

決定係数(寄与率)

$$R^2 = \frac{S_R}{S_T} = 1 - \frac{S_E}{S_T}$$

$R$ : 変相関係数  
(観測値  $y_i$  と予測値  $\hat{y}_i$  との相関係数)

(C) Sadami SUZUKI

平成22年度 統計工学

回帰の有意性検定

- 回帰することに意味があるかどうかを統計的に検討する
  - 分散分析表の活用

| 要因          | 平方和   | 自由度                  | 平均平方(不偏分散)           | F値                |
|-------------|-------|----------------------|----------------------|-------------------|
| 回帰による変動(回帰) | $S_R$ | $\phi_R = p$         | $V_R = S_R / \phi_R$ | $F_0 = V_R / V_E$ |
| 回帰からの変動(残差) | $S_E$ | $\phi_E = n - p - 1$ | $V_E = S_E / \phi_E$ |                   |
| 計           | $S_T$ | $\phi_T = n - 1$     |                      |                   |

- 回帰係数の検定
  - 帰無仮説  $H_0: b_j = 0$  (対立仮説  $H_1: b_j \neq 0$ )
$$t_0 = \frac{b_j}{b_j \text{ の標準誤差 }}$$

自由度  $n-p-1$  の  $t$  分布に従う

(C) Sadami SUZUKI