

計測工学11

回帰分析と相関

1

回帰分析

- 測定値群の変数間の因果関係を数式で記述し、予測にも役立つようにする方法が**回帰法**(regression)である。
- 関数: $y=f(x_1, x_2, \dots, a, b, \dots)$ という関数関係を仮定し、 x_i に対し誤差を含む測定値 y から、未知パラメータ $a, b, \dots$ を推定し、1本の線を引くことを**回帰分析**という。

「回帰: ひとまわりしてもとにもどそうとすること」国語大辞典(新装版)

2

回帰分析(例)

何名かの体重と身長が分かっているとき、体重の値は分かっているが、身長が不明の人がいるとする。このようなとき、すでに得ているデータから身長と体重の関係を調べ、その相関を求め、**体重**から身長不明の人の**身長**を予測する。

予測に使う値: 説明変量
求めるもの: 目的変量

説明変量が1つの場合を単回帰分析という

3

回帰分析(例)

豆を多く収穫するために、肥料の量を5, 10, 15の3種類、水の量を2, 4, 6の3種類、計9種類の育て方で収穫の違いを調べる実験を行った。この実験結果から、**肥料の量**が8で**水の量**が5の場合の**豆の収穫**を予測する。

説明変量 2つ 目的変量 1つ

説明変量が2つ以上の場合を多重回帰分析という

4

線形回帰分析

- 未知パラメータが線形結合の場合
 $y=ax_1+bx_2+\dots+lx_n$
として、未知数 $a, b, \dots, l$ を最小二乗法で推定する。(未知数が x でなく、 a, b となっていることに注意)

5

線形回帰分析: $y=p+qx$ の場合

$$\text{平均値: } \bar{x} = \frac{1}{n} \sum x_i, \quad \bar{y} = \frac{1}{n} \sum y_i$$

$$x \text{ の残差の二乗和: } S_{xx} = \sum (x_i - \bar{x})^2$$

$$\text{残差の積の総和: } S_{xy} = \sum (x_i - \bar{x})(y_i - \bar{y})$$

とおくと

$$p = \bar{y} - \frac{S_{xy}}{S_{xx}} \bar{x}, \quad q = \frac{S_{xy}}{S_{xx}} \quad (5.5)$$

となり、回帰直線は

$$y = p + qx \quad \text{または} \quad y - \bar{y} = \frac{S_{xy}}{S_{xx}} (x - \bar{x}) \quad (5.6)$$

6

最小二乗法の式(3.19)との関係

(3.19)式 (p39)

$$p = \frac{1}{\Delta} \left(\sum x_i^2 \sum y_i - \sum x_i \sum x_i y_i \right)$$

$$q = \frac{1}{\Delta} \left(n \sum x_i y_i - \sum x_i \sum y_i \right)$$

$$\Delta = n \sum x_i^2 - \left(\sum x_i \right)^2$$

(5.5)式(p56)

$$p = \bar{y} - \frac{S_{xy}}{S_{xx}} \bar{x} = \frac{1}{n} \sum y_i - \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} \left(\frac{1}{n} \sum x_i \right)$$

$$q = \frac{S_{xy}}{S_{xx}} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

7

qの式変形の例(3.19式との関係)

$$q = \frac{S_{xy}}{S_{xx}} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

$$\text{分母 } \sum (x_i - \bar{x})^2 = \sum (x_i^2 - 2x_i \bar{x} + \bar{x}^2)$$

$$= \sum x_i^2 - 2\bar{x} \sum x_i + \bar{x}^2 \sum 1$$

$$= \sum x_i^2 - 2 \left(\frac{1}{n} \sum x_i \right) \sum x_i + \left(\frac{1}{n} \sum x_i \right)^2 n$$

$$= \sum x_i^2 - \frac{1}{n} \left(\sum x_i \right)^2$$

$$\text{分子 } \sum (x_i - \bar{x})(y_i - \bar{y}) = \sum (x_i y_i - x_i \bar{y} - y_i \bar{x} + \bar{x} \bar{y})$$

$$= \sum x_i y_i - \bar{y} \sum x_i - \bar{x} \sum y_i + \bar{x} \bar{y} \sum 1$$

$$= \sum x_i y_i - \left(\frac{1}{n} \sum y_i \right) \sum x_i - \left(\frac{1}{n} \sum x_i \right) \sum y_i + \left(\frac{1}{n} \sum x_i \right) \left(\frac{1}{n} \sum y_i \right) n$$

$$= \sum x_i y_i - \frac{1}{n} \sum x_i \sum y_i$$

8

分散、共分散

$$x \text{ の残差の二乗和 : } S_{xx} = \sum (x_i - \bar{x})^2$$

$$\text{残差の積の総和 : } S_{xy} = \sum (x_i - \bar{x})(y_i - \bar{y})$$

とよくと

xの不偏分散Vは

$$V = \frac{S_{xx}}{n-1}$$

共分散 c_{xy} は

$$c_{xy} = \frac{S_{xy}}{n-1}$$

9

因子、水準、欠測値

- 因子(factor): 実験結果の大きさを評価するために変動させる変数(豆の収穫予測のための実験における、肥料の量と水の量)
- 水準(level): 因子設定の段係数(豆の実験では、肥料の量3水準、水の量3水準)
- 欠測値: 何らかの事故のため得られなかった実験値(豆の実験をしたところ、肥料の量5、水の量10の豆が洪水で全て流れてしまい収穫できなければ欠測値となる)

10

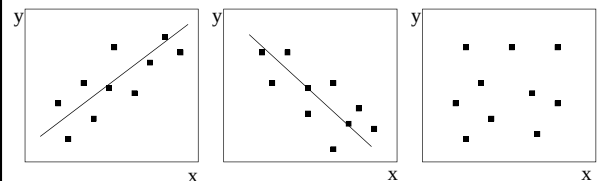
例題5. 1

- Excelで表を完成し、p, qを求める(式5.5よりp, qを求める)
- Excelの近似曲線でp, qを求め、式(5.5)の結果と比較する

11

相関関係

- 相関とは、2つの変量間の関係を表わす尺度である。



正の相関

負の相関

相関がない

12

相関係数

x, y を無次元化し、相関係数 r を

$$r = \sum \left\{ \frac{(x_i - \bar{x})}{\sqrt{\sum (x_i - \bar{x})^2}} \cdot \frac{(y_i - \bar{y})}{\sqrt{\sum (y_i - \bar{y})^2}} \right\} = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}}$$

と定義する。(x, yの共分散をxの標準偏差とyの標準偏差で割ったもの)

相関係数の特性

$$|r| \leq 1$$

$r > 0$: 正の相関 (x, yの増減傾向が一致)

$r < 0$: 負の相関 (x, yの増減傾向が逆)

$0 \leq |r| \leq 0.2$: ほとんど相関がない

$0.2 < |r| \leq 0.4$: 弱い相関がある

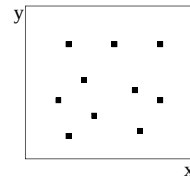
$0.4 < |r| \leq 0.7$: ある程度相関がある

$0.7 < |r| \leq 1.0$: 強い相関がある

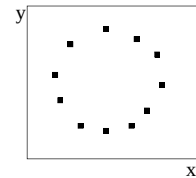
13

相関と独立

- 「2つの変数に相関がない」と「2つの変数が独立である」は同じ意味ではない



相関係数=0で独立



相関係数=0であるが独立でない

14

相関の演習

- 例題5. 2 Excelで表5.2を完成し、相関係数を求める

15