レポート提出票

科目名: 情報工学実験2

実験テーマ: 実験テーマ4 統計的推測と単回帰分析

実施日: 2020年 9月 28日

学籍番号: 4619055

氏名: 辰川力駆

共同実験者:

1 はじめに

本実験では、単回帰分析の考え方と手順を理解することを目標とする。

2 目的

1. 単回帰分析の考え方と手順

   単回帰分析の目的、考え方、手順を理解する

2. 単回帰分析における行列表現

   単回帰分析における行列表現 (線形回帰モデル、正規方程式、最小二乗推定量など) を理解する

3. 実際のデータ解析

   実際のデータに回帰分析を適用することで、解析法を実践的に利用・応用できるようにする

3 実験方法

3.1 実験 1 単回帰分析の考え方と手順

6 つの市町村の人口と行政職員数の仮想データを表 1 に示す。また、各市町村の人口を x_i, 職員数を $y_i(i = 1, ..., n(= 6))$ と表記する。

表 1: 市町村の人口と行政職員数

市町村	人口 x(千人)	職員数 y(人)
A	1	10
B	2	20
C	3	20
D	3	40
E	5	40
F	1	5
合計	15	135

1. 次の統計量を計算する。

$$\sum_{i=1}^{n} x_i, \quad \sum_{i=1}^{n} y_i, \quad \sum_{i=1}^{n} {x_i}^2, \quad \sum_{i=1}^{n} {y_i}^2, \quad \sum_{i=1}^{n} x_i y_i$$

2. 次式が整理することを証明する。

$$\sum_{i=1}^{n}(x_i-\bar{x})^2 = \sum_{i=1}^{n}{x_i}^2-n\bar{x}^2 \tag{1}$$

$$\sum_{i=1}^{n}(y_i-\bar{y})^2 = \sum_{i=1}^{n}{y_i}^2-n\bar{y}^2 \tag{2}$$

$$\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y}) = \sum_{i=1}^{n}x_iy_i-n\bar{x}\bar{y} \tag{3}$$

3. 人口 x と職員数 y の基本統計量 (データ数、平均、標準偏差、最小値、最大値) を計算する。

$$\text{人口}\ x\ \text{の平均} = \bar{x} = \frac{\sum_{i=1}^{n}x_i}{n} \tag{4}$$

$$\text{人口}\ x\ \text{の標準偏差} = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}} = \sqrt{\frac{\sum_{i=1}^{n}{x_i}^2-n\bar{x}^2}{n-1}} \tag{5}$$

$$\text{職員数}\ y\ \text{の平均} = \bar{y} = \frac{\sum_{i=1}^{n}y_i}{n} \tag{6}$$

$$\text{職員数}\ y\ \text{の標準偏差} = \sqrt{\frac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{n-1}} = \sqrt{\frac{\sum_{i=1}^{n}{y_i}^2-n\bar{y}^2}{n-1}} \tag{7}$$

4. 人口 x と職員数 y の Pearson の積率相関係数 r を計算する。

$$r = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}} = \frac{\sum_{i=1}^{n}x_iy_i-n\bar{x}\bar{y}}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}} \tag{8}$$

5. 人口 x を横軸, 職員数 y を縦軸にした散布図を作成して、両者の関係を調べる。

6. 単回帰モデル $y_i=\beta_0+\beta_1x_i+\varepsilon_i(i=1,...,n)$ をあてはめる。β_0 と β_1 の推定量を $\hat{\beta}_0$ と $\hat{\beta}_1$ とすると、目的変数 (応答変数) である職員数の予測値は $\hat{y}_i=\hat{\beta}_0+\hat{\beta}_1x_i$ で与えられる。次式の残差平方和 S_e を $\hat{\beta}_0$ と $\hat{\beta}_1$ でそれぞれ偏微分する。

$$S_e = \sum_{i=1}^{n}(y_i-\hat{y}_i)^2 = \sum_{i=1}^{n}(y_i-(\hat{\beta}_0+\hat{\beta}_1x_i))^2 \tag{9}$$

7. 正規方程式を作成する。

8. 正規方程式を解き、β_0 と β_1 の最小二乗推定量を数式で表現する。

9. 最小二乗推定量 $\hat{\beta}_0$,$\hat{\beta}_1$ の値を求める。

10. 得られた回帰直線 ($\hat{y}_i=\hat{\beta}_0+\hat{\beta}_1x_i$) を手順 5 で作成した散布図に図示して、結果を考察する。

11. データ分析 [回帰分析] を用いて、これまでに得られた結果と同様の結果が得られることを確認する。

3.2 実験 2 単回帰分析における行列表現

データ数を n とする。目的変数ベクトル $\boldsymbol{Y}$ と説明変数を含む定数行列 $\boldsymbol{X}$ を次式で定義する。

$$\boldsymbol{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

$$\boldsymbol{X} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}$$

このとき、実験 1 の単回帰モデルは

$$\boldsymbol{y} = X\beta + \varepsilon \tag{10}$$

$$\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$$

1. あ

ぱ あ

ぱあ

ソースコード 1: read_2_1byte.c

```
#include <stdio.h>
```

4 結果・考察・課題

5 まとめ

1. 単回帰分析の考え方と手順を学んだ
   - 手計算やエクセルで分析を行った
2. 単回帰分析における行列表現を学んだ
   - 実験 1 の手順を行列表現した
   - R を使い、単回帰分析を行った

6 感想

参考文献

[1] 東京理科大学工学部情報工学科 情報工学実験 2 2020 年度東京理科大学工学部情報工学科出版