

データ解析 課題 第1回

概要： 回帰分析のプログラムを実装して，データ解析をおこなう．作成したプログラムおよびデータ解析の結果をレポート（紙）で提出．回帰分析については，講義資料の重回帰分析の節を参照すること．

提出方法： 西 8 W のレポートボックス「データ解析」

締め切り： 5月2日（金）の 15:00

1.1 最小 2 乗法

「応用線形代数」で学習した最小 2 乗法を復習する．まずは単回帰分析． $2n$ 個の実数 $x_1, \dots, x_n, y_1, \dots, y_n$ がデータとして与えられている． x と y を「変数」と考え，その t 番目の要素 x_t と y_t に，次の関係があると考えよう．

$$y_t = \beta_0 + \beta_1 x_t + \epsilon_t, \quad t = 1, \dots, n$$

ここで， β_0, β_1 は回帰係数， ϵ_t は誤差である． x は説明変数， y は目的変数と呼ばれる．このとき，誤差の 2 乗和 $\sum_{t=1}^n \epsilon_t^2$ を最小にするように回帰係数を定める．これは，データから回帰係数を「推定」したことになる．このとき，回帰係数の推定値 $\hat{\beta}_0, \hat{\beta}_1$ を表す数式をデータの関数として示せ．なお， $n \geq 2$ などのような

条件を適当に仮定して，ほとんどの場合に通用しそうな解を示せばよいが，その条件を明記すること．

1.2 複数の変数がある場合の最小 2 乗法

説明変数 x が m 次元の行ベクトルとする．これを n 回観測したとき， t 番目の観測値を $x_t = (x_{t1}, x_{t2}, \dots, x_{tm})$, $t = 1, \dots, n$ と書く．このとき誤差の 2 乗和

$$\sum_{t=1}^n [y_t - (\beta_0 + \beta_1 x_{t1} + \dots + \beta_m x_{tm})]^2$$

を最小にする回帰係数（すなわち最小 2 乗法の解）を示せ．適当な条件を仮定してよい．

1.3 R による実装

上記の最小 2 乗法を R の関数として実装せよ．関数の入力，および，出力の意味や，各行で行っている演算の説明をきちんとコメントにしておくこと．この関数定義を含めて，つぎのデータ解析を実行するための「スクリプト」を作成し，それを実行したときの「コンソール出力」を示すこと．

「R の使い方ヒント」

```
> A <- matrix(1:6,2,3) # 2 x 3 行列
```

```
> A
```

```
      [,1] [,2] [,3]
[1,]    1    3    5
```

```

[2,]  2  4  6
> t(A) # 転置行列
      [,1] [,2]
[1,]  1  2
[2,]  3  4
[3,]  5  6
> B <- matrix(1:6,3,2) # 3 x 2 行列
> B
      [,1] [,2]
[1,]  1  4
[2,]  2  5
[3,]  3  6
> C <- A %*% B # 行列の積
> C
      [,1] [,2]
[1,]  22  49
[2,]  28  64
> D <- solve(C) # 逆行列
> D
      [,1]      [,2]
[1,]  1.7777778 -1.3611111
[2,] -0.7777778  0.6111111
> C %*% D # 確認
      [,1]      [,2]
[1,]  1.000000e-00  2.997602e-15
[2,] -1.776357e-15  1.000000e+00
> a <- 1:3 # ベクトル (行ベクトルまたは列ベクトルの区別はない)
> A %*% a # a は列ベクトルとして解釈されている
      [,1]

```

```

[1,] 22
[2,] 28
> a %*% B # aは行ベクトルとして解釈されている
      [,1] [,2]
[1,]  14  32
> as.matrix(a) # 行列形式に変換するときは，列ベクトルとして解釈される．
      [,1]
[1,]    1
[2,]    2
[3,]    3
> cbind(a,a,B) # 行列やベクトルを，列方向につなげる
      a a
[1,] 1 1 1 4
[2,] 2 2 2 5
[3,] 3 3 3 6
> cbind(1,B) # 長さが足りないときは，要素が繰り返し利用される
      [,1] [,2] [,3]
[1,]    1    1    4
[2,]    1    2    5
[3,]    1    3    6

```

1.4 データ解析

X2000data.txt を読み込み，適当な変数を選ぶ．目的変数は1個，説明変数は2から10個程度を選び，上記のプログラムを使って回帰分析を実行する．

- 関数定義を含め，データの読み込みから結果の出力までをおこなう「スクリプト」を作成する．
- スクリプトを実行したときの，コンソール出力を示せ．スクリプトの定義，および，実行結果が含まれるようにしておくこと．推定した回帰係数をわかりやすく示すこと．
- 選んだ変数には，E09504 のようなコード番号ではなく，Gakureki のようなわかりやすい名前をつけること．さらに，選んだ変数の説明を X2000name.txt からコピーして示すこと．
- 友達とレポートの相談をするのはかまわないが，全く同じ変数の組み合わせにならないように配慮すること．gakureki-rikon-12.txt にしめした 1 2 変数からは変数を選ばないこと．
- 分析結果にたいして，なんらかの解釈を与えること．

1.5 確認

自分で書いたプログラムが正しいかどうか確認する．R に組み込みの `lm` をつかって，上記と同じデータ解析を行い，その回帰係数の推定値が自分のプログラムの推定値と等価であることを確認すること．

- `lm` を実行する行，および，推定した回帰係数の値が含まれるように，コンソール出力を示すこと．