《系统工程导论》

第四章作业 1

Matlab 实现一元线性回归

姓名：卢志

班级：自 43 班

学号：2014011497

2016 年 3 月 28 日

1. 题目描述

利用 Matlab，编程实现**一元线性回归**。

要求：

1）实现函数 function linear_regression1(data,alpha)；

2）**输入为 N*2 的矩阵 data**，第一列为 Y，第二列为 X;显著性水平 alpha;

3）**打印出回归直线方程**（有必要的话，也可输出重要的中间数据）;

4）用 **F 检验法进行统计检验**，显著性水平为**输入 alpha**，提示 Matlab 中得到 F 分布对于给定显著性水平和自由度的分位数函数是 finv，请大家自行用 help 工具确定其用法；输出检验结果，如果输入数据满足线性关系，那么 5）和 6），否则结束；

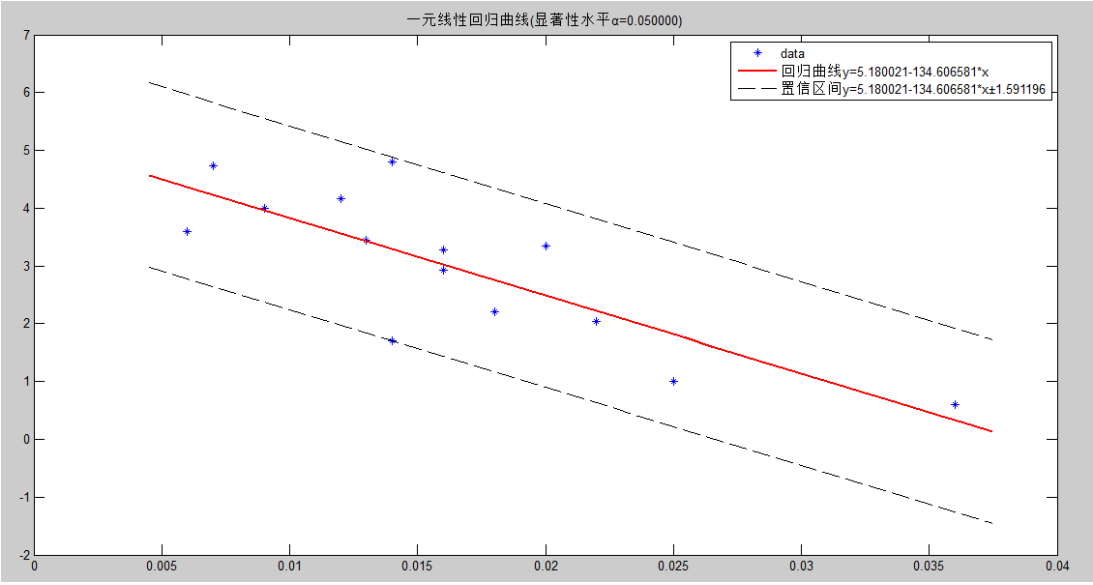
5）打印出置信区间，提示 Matlab 中得到标准正态分布相应分位数的函数是 norminv，请大家自行用 help 工具确定其用法；

6）在**一个 figure 中，画出:a 所有数据点，b 回归直线，c 置信区间相应的两条边界直线**。

7）将“第 4 章作业(1)_数据.ppt”中的数据用你编写的程序处理，将所有结果贴到作业报告中，显著性水平取 0.05。

2. 作业解答

2.1 算例结果



由上述结果图可知，得到拟合后的一元线性回归曲线方程式为

$$y=5.18-134.61x$$

且经过 F 检验后发现符合显著性水平 α=0.05 的要求。

计算得其置信区间是 y=5.18-134.61x±1.59。

2.2 操作步骤

① 输入的测试值如下：

编号	成分 A(x)	成分 B(y)	编号	成分 A(x)	成分 B(y)
1	0.009	4.0	8	0.014	1.7
2	0.013	3.44	9	0.016	2.92
3	0.006	3.6	10	0.014	4.8
4	0.025	1.0	11	0.016	3.28
5	0.022	2.04	12	0.012	4.16
6	0.007	4.74	13	0.020	3.35
7	0.036	0.6	14	0.018	2.2

② 求解一元线性回归方程

原理为：设已经得到了 x 和 y 的若干数据对 xi 和 yi，i=1,2,…,N，称为样本点。如果 x 和 y 存在某种线性关系，则 x 和 y 可用

$$y=a+bx+\varepsilon$$

表示，a 和 b 是待定系数，ε是随机变量。该模型为一元回归模型。a 和 b 为回归系数。

我们利用最小二乘原理，使目标误差平方和最小，从而得到：

$$X_i = [x_1 - \bar{x}, x_2 - \bar{x}, \ldots, x_N - \bar{x}], \bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

$$Y_i = [y_1 - \bar{y}, y_2 - \bar{y}, \ldots, y_N - \bar{y}], \bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i$$

$$\hat{b} = \frac{\sum X_i Y_i}{\sum X_i^2} = \frac{L_{xy}}{L_{xx}}, \hat{a} = \bar{y} - \hat{b}\bar{x}$$

由此可得到回归方程:

$$y = \hat{a} + \hat{b}x$$

③ 显著性检验

利用 F 检验法进行统计性检验，根据讲义，统计量

$$F = \frac{ESS/f_E}{RSS/f_R}$$

服从自由度为(n,N-n-1)的 F 分布，对于给定的显著性水平，可查 F 分布表，做假设检验。可利用 finv 函数，用法如下:

x=finv(p,v1,v2)。分子自由度为 v1，分母自由度为 v2，求置信度（显著性水平）为 1-p 的情况下得到的 F 的值即为 x。此时分子为解释平方和，自由度 1；分母为剩余平方和，自由度为 N-2（N 为数据总数)。

④ 预测精度（求取置信区间）

F检验的结果为x与y存在线性关系，因此求取置信区间。S_δ 为y的剩余均方差，表示变量y偏离回归直线的误差:

$$S_\delta = \sqrt{\frac{\sum_{i=1}^{N}(y_i - \hat{y})^2}{N-2}} = \sqrt{\frac{RSS}{f_R}}$$

S_δ服从正态分布，可用norminv函数求取置信区间。norminv函数用法如下:

x=norminv(p,u,sigma)。norminv是正态分布的反函数，已知概率求百分位点

（方差前的系数）。P为概率，u为正态分布的均值，sigma为正态分布的方差。由于本次中使用的是标准正态分布，故u=0，sigma=1。语句为：

z=norminv(1-alpha/2,0,1)%即得到正数的z进行置信区间计算

2.3 原始代码

主函数 HW3.m

```
%%
%清除缓存区
clear all;
close all;

%%
%输入原始数据
N=14;
data = zeros(N,2);
%按照要求输入数据：第一列为Y，第二列为X
data(:,1) = [4.0,3.44,3.6,1.0,2.04,4.74,0.6,1.7,2.92,4.8,3.28,4.16,3.35,2.2];
data(:,2) =
[0.009,0.013,0.006,0.025,0.022,0.007,0.036,0.014,0.016,0.014,0.016,0.012,0.020,
0.018];

%%
%调用自己编写的线性回归拟合函数
%取显著性水平0.05
alpha=0.05;
linear_regression1(data,alpha);
```

函数linear_regression1(data,alpha)

```
function linear_regression1(data,alpha)
%自主编写的线性回归拟合函数

%%
%PART1：求解线性最小二乘回归的参数
%标出原始坐标点
figure;grid on;
plot(data(:,2),data(:,1),'b*');
hold on;
[N,~] = size(data);
%计算平均值
```

```
sum_x = 0;
sum_y = 0;
for i = 1:N
   sum_y = sum_y + data(i,1);
   sum_x = sum_x + data(i,2);
end
aver_y = sum_y / N;
aver_x = sum_x / N;
%计算Lxx,Lyy,Lxy
Lxy = 0;Lxx = 0;Lyy = 0;
for i = 1:N
   Lxy = Lxy + (data(i,1) - aver_y) * (data(i,2) - aver_x);
   Lxx = Lxx + (data(i,2) - aver_x) * (data(i,2) - aver_x);
   Lyy = Lyy + (data(i,1) - aver_y) * (data(i,1) - aver_y);
end
%计算参数a,b
b = Lxy / Lxx;
a = aver_y - b * aver_x;

%%
%PART2: 进行F检验
ESS = 0;RSS = 0;
for i = 1:N
    esti_y = a + b * data(i,2); %计算y的拟合值
    ESS = ESS + (esti_y - aver_y)^2; %解释平方和
    RSS = RSS + (data(i,1) - esti_y)^2; %剩余平方和
end
F = (N - 2) * ESS / RSS; %F服从F分布
%x=finv(p,v1,v2)。分子自由度为v1，分母自由度为v2
%求置信度（显著性水平）为1-p的情况下得到的F的值即为x
Fa = finv(1 - alpha,1,N-2);
%检验显著性水平，若拒绝H0，则输出图像
if F > Fa
    begin_x = min(data(:,2)) - 0.05 * (max(data(:,2)) - min(data(:,2)));%起始位
置
    end_x = max(data(:,2)) + 0.05 * (max(data(:,2)) - min(data(:,2)));%终止位置
    x = begin_x:0.001:end_x;
    %打印回归曲线
    y = a + b * x;
    plot(x,y,'r','LineWidth',2);hold on;

%%
%PART3: 预测精度（求取置信区间）
    S = sqrt(RSS / (N - 2)); %剩余均方差S
```

```
    %x=norminv(p,u,sigma)。
    %norminv是正态分布的反函数，已知概率求百分位点（方差前的系数）。
    %P为概率，u为正态分布的均值，sigma为正态分布的方差。
    %由于本次中使用的是标准正态分布，故u=0，sigma=1。
    z_half_alpha = norminv(1 - alpha/2,0,1); %半区间
    y_low = a + b * x - z_half_alpha * S; %置信区间下界
    y_up = a + b * x + z_half_alpha * S; %置信区间上界
    %画出置信区间
    plot(x,y_low,'k--');hold on;
    plot(x,y_up,'k--');hold on;
    if b >= 0
        tip1 = sprintf('回归曲线y=%f+%f*x',a,b);
        tip2 = sprintf('置信区间y=%f+%f*x±%f',a,b,z_half_alpha*S);
    else
        tip1 = sprintf('回归曲线y=%f%f*x',a,b);
        tip2 = sprintf('置信区间y=%f%f*x±%f',a,b,z_half_alpha*S);%b的负号
    end
    legend('data',tip1,tip2);
    title1 = sprintf('一元线性回归曲线(显著性水平α=%f)',alpha);
    title(title1);
%如果不符合显著性水平要求，则输出结果
else
   close all;
   warning = sprintf('一元线性回归曲线不满足α=%f)',alpha);
   warndlg(warning,'错误');
end

end
```