

**《数学实验》实验报告(二)**

**——统计分析和数据建模**

姓 名

学 号

专 业 信息与计算科学（嵌入式）

班 级 18（1）

**二Ｏ 二Ｏ 年 月 日**

1. **表8.4给出了1930年各国人均年消耗的烟支数以及1950年男子死于肺癌的死亡率.（注：研究男子的肺癌死亡率是因为在1930年左右几乎极少的妇女吸烟，记录1950年的肺癌死亡率是因为考虑到吸烟的效应要有一段时间才能显现.）**

|  |  |  |
| --- | --- | --- |
| **表8.4 各国烟消耗量与死于肺癌人数** | | |
| 国家 | 1930年人均烟消耗量 | 1950年每百万男子死于肺癌人数 |
| 澳大利亚 | 480 | 180 |
| 加拿大 | 500 | 150 |
| 丹麦 | 380 | 170 |
| 芬兰 | 1100 | 350 |
| 英国 | 1100 | 460 |
| 荷兰 | 490 | 240 |
| 冰岛 | 230 | 60 |
| 挪威 | 250 | 90 |
| 瑞典 | 300 | 110 |
| 瑞士 | 510 | 250 |
| 美国 | 1300 | 200 |

1. **画出该数据散点图；**
2. **该散点图是否表明在吸烟多的人中间肺癌死亡率较高？**
3. **计算两列数据的相关系数.**

**代码：**

x=[480,500,380,1100,1100,490,230,250,300,510,1300];

y=[180,150,170,350,460,240,60,90,110,250,200];

Label={'澳大利亚','加拿大','丹麦','芬兰','英国','荷兰', ...

'冰岛','挪威','瑞典','瑞士','美国'};

scatter(x,y,'filled')

xlabel('1930年人均烟消耗量');

ylabel('1950年每百万男子死于肺癌人数');

for k=1:11

text(x(k)+10,y(k),Label(k));%每个点加上具体信息

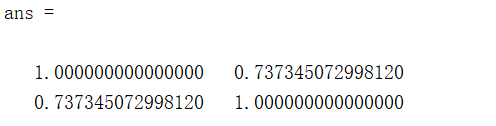
end

corrcoef(x,y)

**结果：**

（1）

1. 是



由此知两列数据的相关系数为0.7373

1. **用Monte Carlo法计算积分（任选一个）**

**，，**

**选择：**

****

**代码：**

x=unifrnd(0,1,1,10000);

y=unifrnd(0,1,1,10000);

count=0;

for k=1:10000

if(exp(-x(1,k)^2/2)/sqrt(2\*pi)>=y(1,k))

count=count+1;

end

end

result=count/10000

**结果：**



1. **第九章ppt中的第29页问题：**

**二次多项式拟合数据**

****

**且要求和,怎么办？**

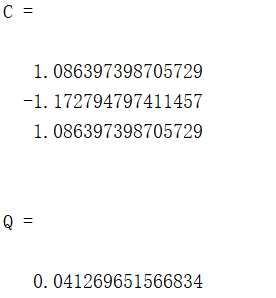
**代码：**

x=[0.1,0.2,0.15,0,-0.2,0.3];

y=[0.95,0.84,0.86,1.06,1.50,0.72];

[C,Q]=lsqlin([(x.^2)',x',ones(6,1)],y,[1,0,-1],0,[1,1,1],1,[],[])

**结果：**



1. **用二次模拟产生下列数据表：**

|  |  |  |  |
| --- | --- | --- | --- |
| 序号 | x | y | z |
| 1 | 1.61 | 9.57 | 100.85 |
| 2 | 5.50 | 4.85 | 50.43 |
| 3 | -6.78 | 8.00 | 142.48 |
| 4 | 2.59 | 1.42 | 10.03 |
| 5 | 0.96 | 4.22 | 26.72 |
| 6 | -3.92 | 9.16 | 118.65 |
| 7 | -1.30 | 7.92 | 81.02 |
| 8 | 1.03 | 9.59 | 96.59 |
| 9 | 10.74 | 6.56 | 132.83 |
| 10 | 8.31 | 0.36 | 51.80 |
| 11 | -4.05 | 8.49 | 101.82 |
| 12 | 9.10 | 9.34 | 157.13 |

1. **用二元线性模拟拟合上述数据表，并作回归分析；**

**代码：**

x=[1.61,5.50,-6.78,2.59,0.96,-3.92,-1.30,1.03,10.74,8.31,-4.05,9.10];

y=[9.57,4.85,8.00,1.42,4.22,9.16,7.92,9.59,6.56,0.36,8.49,9.34];

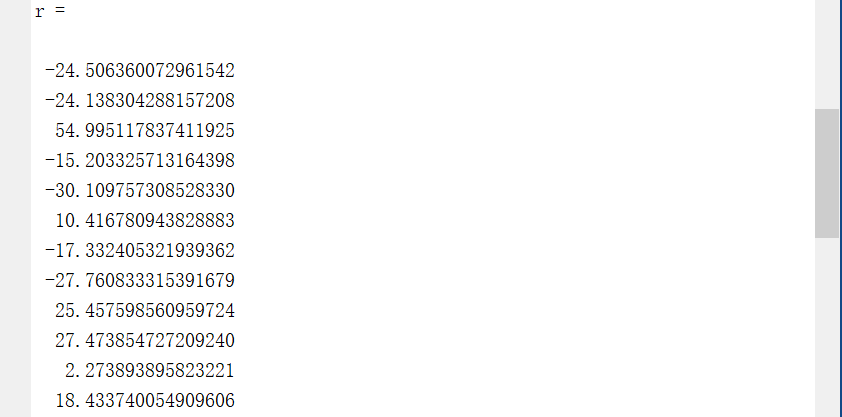
z=[100.85,50.43,142.48,10.03,26.72,118.65,81.02,96.59,132.83, ...

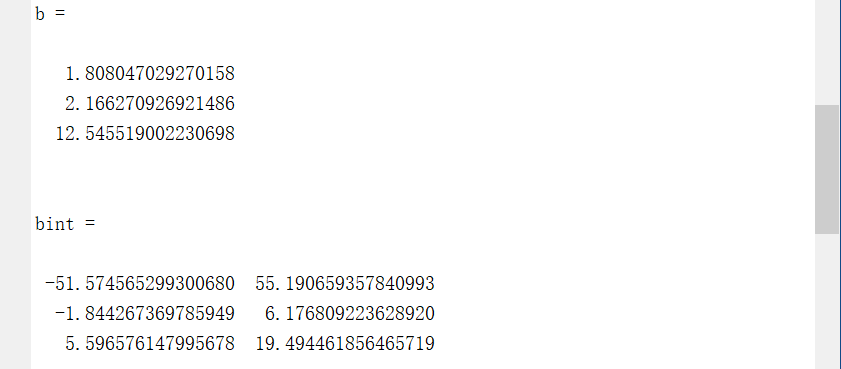
51.80,101.82,157.13];

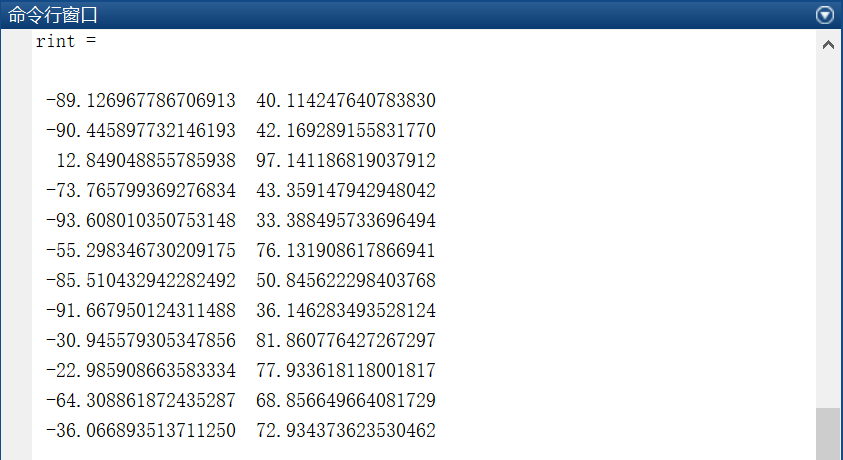
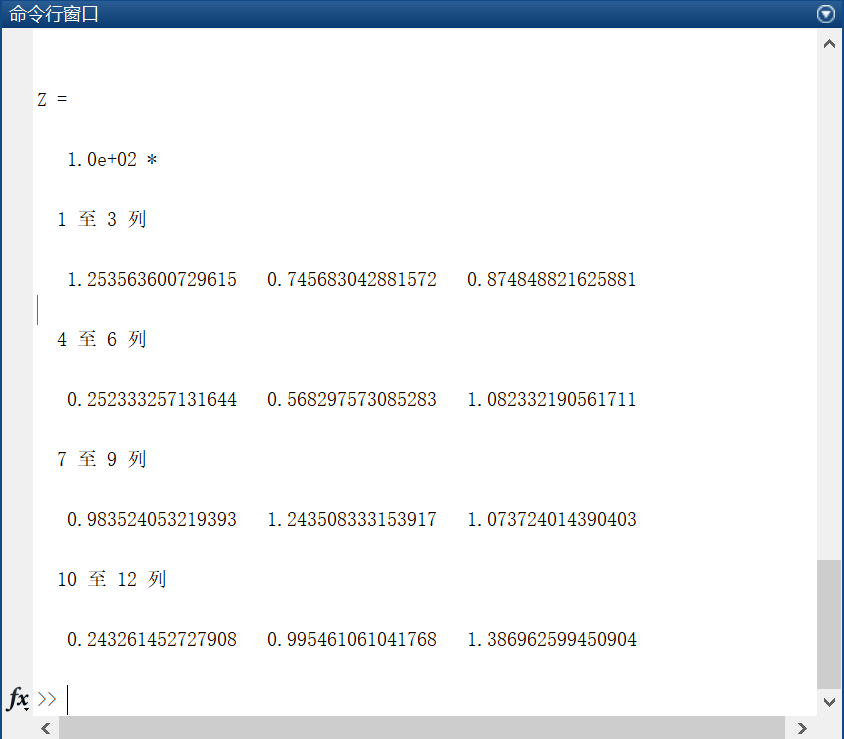
XY=[ones(12,1),x',y'];

[b,bint,r,rint,stats]=regress(z',XY)

Z=b(1)+b(2)\*x+b(3)\*y

**结果：**





1. **用完全二次模拟拟合上述数据表，并作回归分析；**

**代码：**

x=[1.61,5.50,-6.78,2.59,0.96,-3.92,-1.30,1.03,10.74,8.31,-4.05,9.10];

y=[9.57,4.85,8.00,1.42,4.22,9.16,7.92,9.59,6.56,0.36,8.49,9.34];

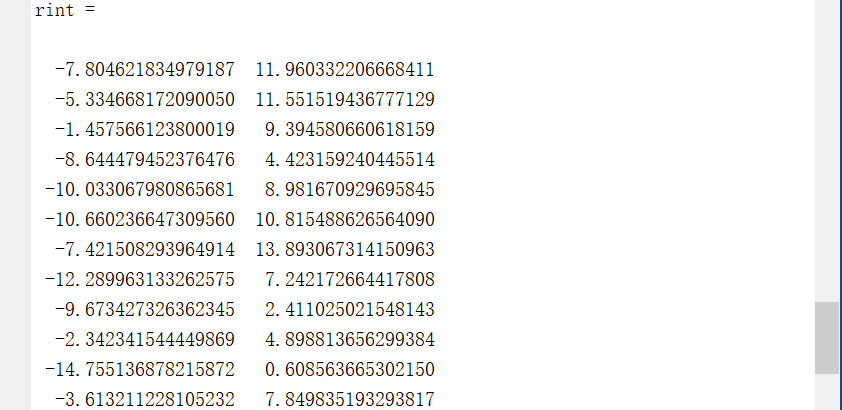
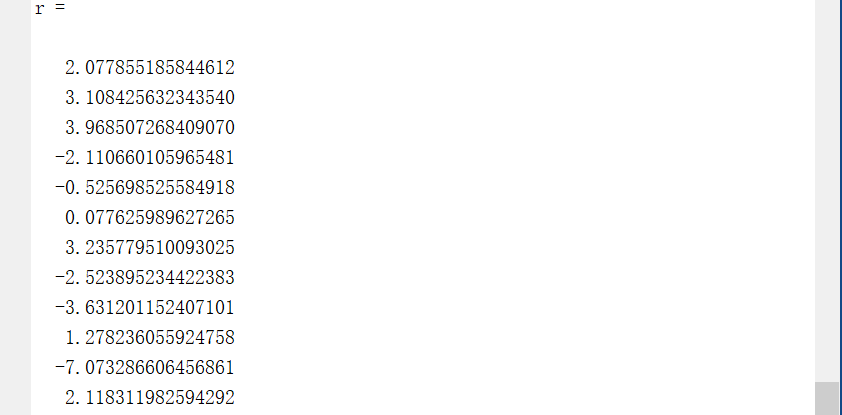
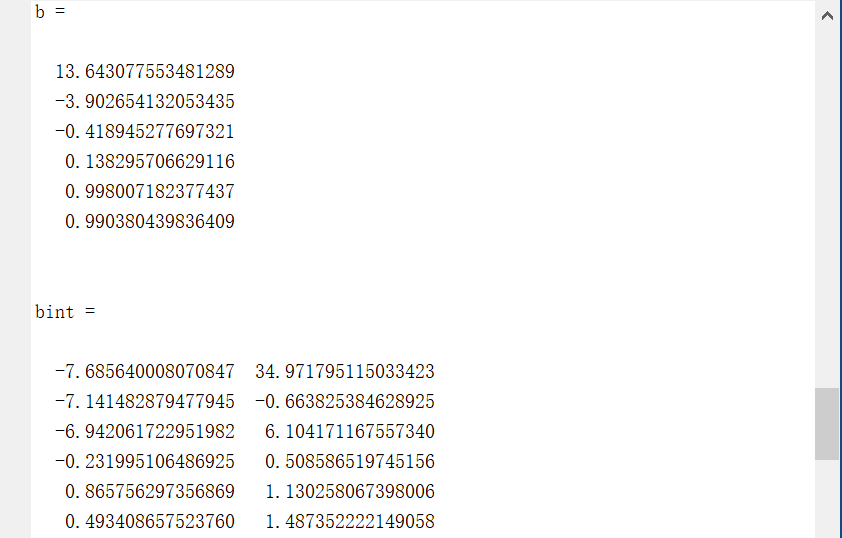
z=[100.85,50.43,142.48,10.03,26.72,118.65,81.02,96.59,132.83, ...

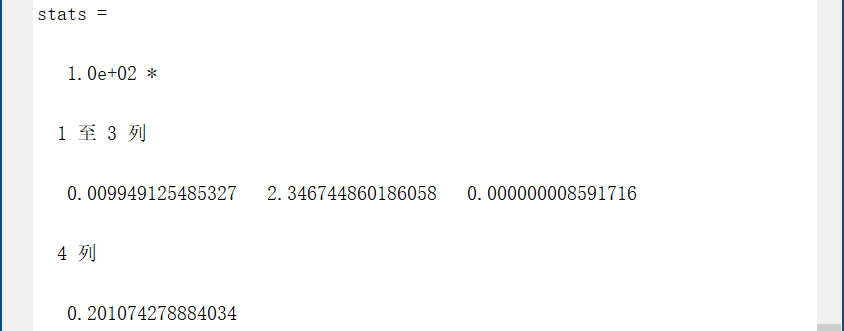
51.80,101.82,157.13];

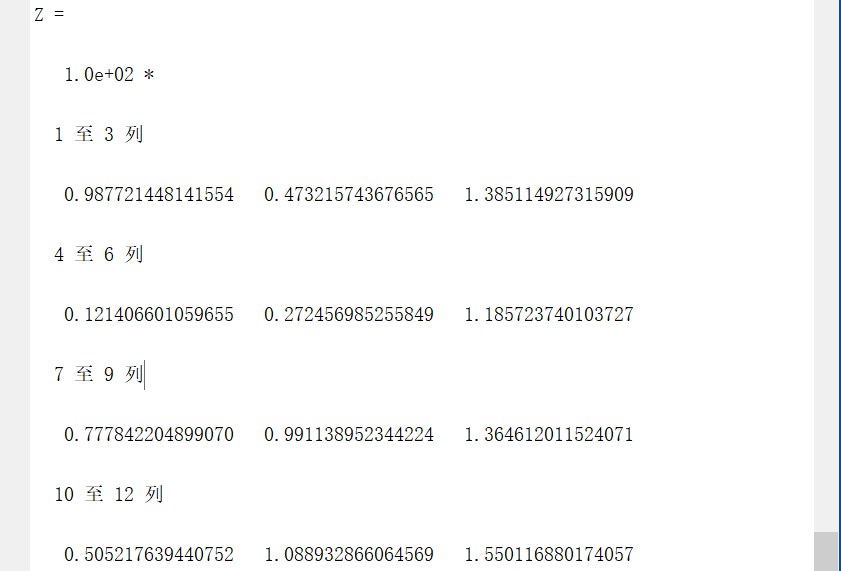
XY=[ones(12,1),x',y',(x.\*y)',(x.^2)',(y.^2)'];

[b,bint,r,rint,stats]=regress(z',XY)

Z=b(1)+b(2)\*x+b(3)\*y+b(4).\*x.\*y+b(5)\*(x.^2)+b(6)\*(y.^2)

**结果：**





1. **用逐步回归法精简上述二次模型；**

**代码：**

stepwise(XY(:,2:6),z)

**结果：**

****

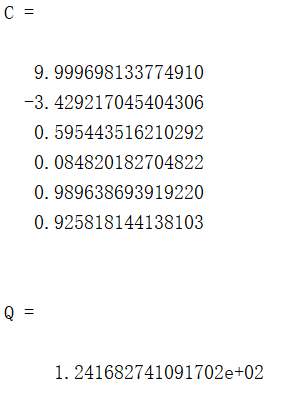
逐步回归推荐的精简模型为：

1. **若限定不超过10，用二次模型拟合上述数据表，并作误差分析；**

**代码：**

[C,Q]=lsqlin(XY,z,[],[],[],[],[],[10,inf,inf,inf,inf,inf])

**结果：**



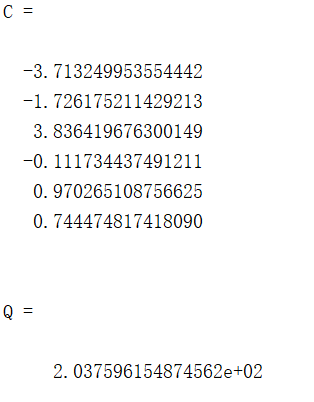
**分析：**由Q知结果误差较大，C知系数拟合较好

1. **若限定不超过10，且，用二次模型拟合上述数据表，并作误差分析.**

**代码：**

[C,Q]=lsqlin(XY,z,[],[],[1,1,1,1,1,1],0,[],[10,inf,inf,inf,inf,inf])

**结果：**



**分析：**由Q知结果误差较大，C知系数拟合也较差