



《数学建模与数学实验》第5版 统计分析 习题9.7

原创小趴菜_已于 2023-09-22 18:58:41 修改阅读量1.1w收藏191点赞数37

分类专栏：《数学建模与数学实验》文章标签：matlab回归聚类

版权

 开放原子开发者工作坊 文章已被社区收录

 《数学建模与数学实... 专栏收录该内容

186 订阅7 篇文章

加入社区

订阅专栏

文章目录

1. 回归分析
- 1.1 考察温度x对产量y的影响，测得下列10组数据：
- 1.2 某零件上有一段曲线，为了在程序控制机床上加工这一零件，需要这段曲线的解析表达式，在曲线横坐标 x_i 处测得纵坐标 y_i 共11对数据如下：
- 1.3 混凝土的抗压强度随养护时间的延长而增加，现将一批混凝土作成12个试块，记录了养护日期x（日）及抗压强度y（kg/cm²）的数据：
2. 聚类分析
- 为了得出 2007 年江苏省的13个地市的国民经济分布规律，在众多衡量经济水平的指标中我们将采用下列指标：
3. 判别分析
- 银行的贷款部门需要判别每个客户的信用好坏（是否未履行还贷责任），以决定是否给予贷款。

参考教材：《数学建模与数学实验》第5版

提示：以下是本篇文章正文内容，来自参考教材课后习题。

1. 回归分析

1.1 考察温度x对产量y的影响，测得下列10组数据：

温度	20	25	30	35	40	45	50	55	60	65
产量	13.2	15.1	16.4	17.1	17.9	18.7	19.6	21.2	22.5	24.3

求y关于x的 线性回归 方程，检验回归效果是否显著，并预测x=42℃时产量的估值及预测区间（置信度95%）.

matlab求解：

```
1 clear;clc
2 x=[20 25 30 35 40 45 50 55 60 65]';
3 X = [ones(10,1) x];
4 Y=[13.2 15.1 16.4 17.1 17.9 18.7 19.6 21.2 22.5 24.3]';
5 % 线性回归regress
6 [b,bint,r,rint,stats] = regress(Y,X,alpha)
7 % alpha为显著性水平（缺失时为0.05）；
8 % bint: 回归系数区间估计
9 % r, rint: 残差及置信区间
10 % stats: 检验回归模型的统计量
11 [b,bint,r,rint,stats] = regress(Y,X);
12 b,bint,stats
13 % 一元线性回归
14 polytool(x,Y,1)
```

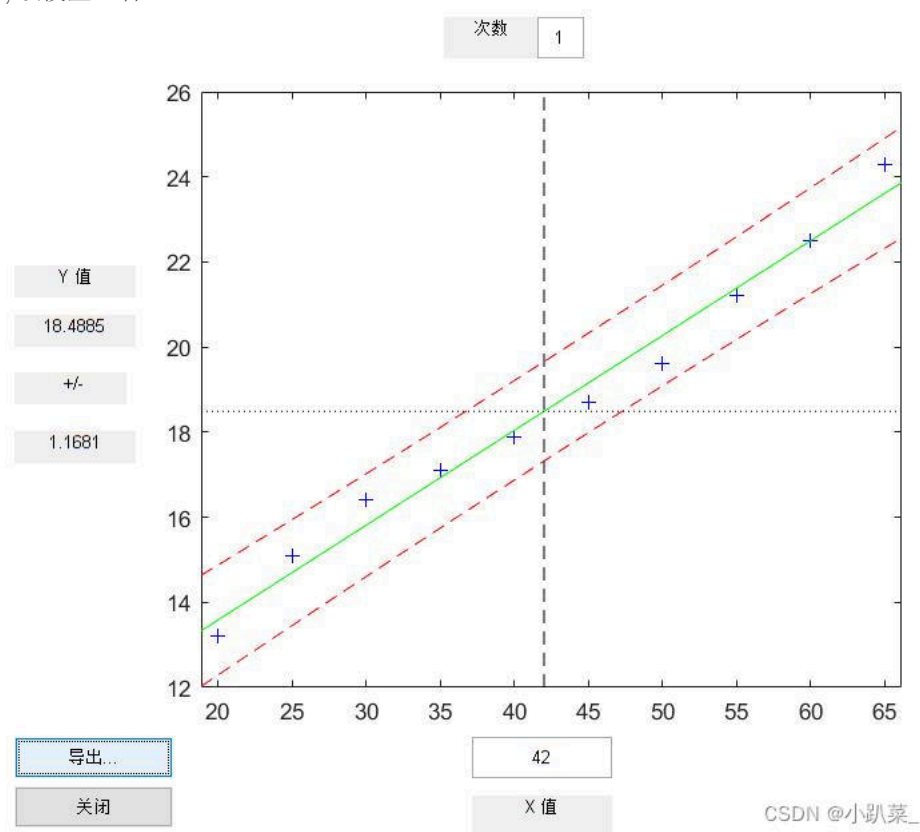
```
b =
    9.1212
    0.2230

bint =
    8.0211    10.2214
    0.1985    0.2476

stats =
    0.9821   439.8311    0.0000    0.2333
```

CSDN @小趴菜_

回归方程为： $y = 9.1212 + 0.2230x$
 $R^2 = 0.9821, P = 0$, 该模型显著。



输入需要预测的值42，导出结果：

变量已在基础工作区中创建。

```
>> yci, yhat
```

```
yci =
    17.3203    19.6566
```

```
yhat =
    18.4885
```

CSDN @小趴菜_

以上结果可得：

当温度为42的预测值为18.4885，预测区间为[17.3203，19.6566]

1.2 某零件上有一段曲线，为了在程序控制机床上加工这一零件，要求这段曲线的解析表达式，在曲线横坐标 x_i 处测得纵坐标

共14组数据如下：

x_i	0	2	4	6	8	10	12	14	16	18	20
y_i	0.6	2.0	4.4	7.5	11.8	17.1	23.3	31.2	39.6	49.7	61.7

求这段曲线的纵坐标y关于横坐标x的二次多项式回归方程.

matlab求解:

```

1 | x=0:2:20;
2 | y=[0.6,2.0,4.4,7.5,11.8,17.1,23.3,31.2,39.6,49.7,61.7];
3 | % 确定多项式系数polyfit
4 | [p,s]=polyfit(x,y,2);
5 | % 预测值
6 | yy=@(x)p(1).*x.^2+p(2).*x+p(3);
7 |
8 | plot(x,y,'+',x,yy(x),'b');

```

p =

0.1403
0.1971
1.0105

s =

包含以下字段的 [struct](#):

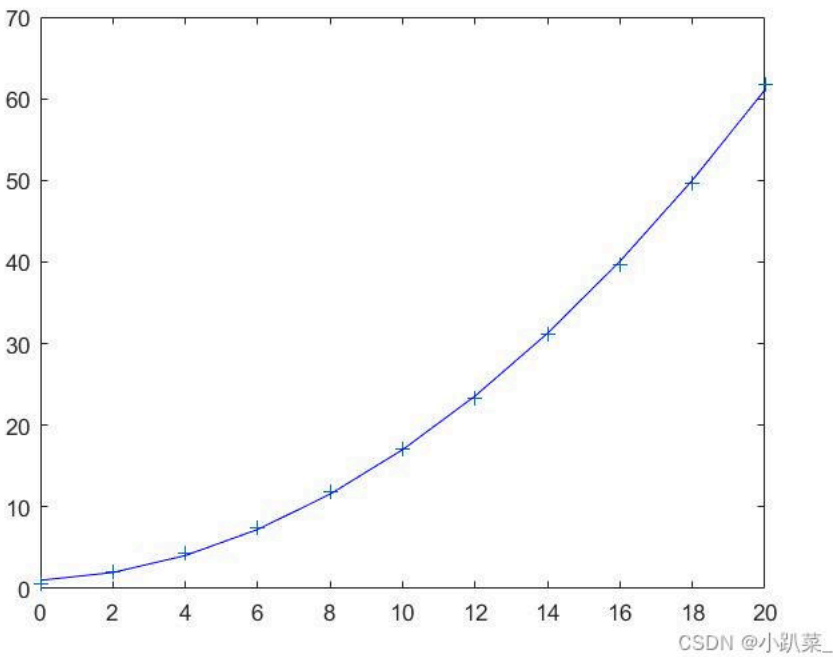
R: [3×3 double]

df: 8

normr: 1.1097

CSDN @小趴菜_

拟合图：



回归方程： $y = 0.1403x^2 + 0.1971x + 1.0105$

1.3 混凝土的抗压强度随养护时间的延长而增加，现将一批混凝土作成12个试块，记录了养护日期x（日）及抗压强度y（kg/cm2）的数据：

x	2	3	4	5	7	9	12	14	17	21	28	56
y	35	42	47	53	59	65	68	73	76	82	86	99

试求 $y = a + blnx$ 型回归方程。

matlab求解:

```
4 x=[2,3,4,5,7,9,12,14,17,21,28,56]';
5 y=[35,42,47,53,59,65,68,73,76,82,86,99]';
6 % 回归系数初值
7 beta0 = [5,1]';
8 % 确定回归系数nlinfit
9 [beta,r,J] = nlinfit(x,y,volum,beta0);
10 beta
11 a=beta(1),b=beta(2)
12 %拟合图
13 yy=a+b*log(x);
14 plot(x,y,'+',x,yy,'r')
```

```
beta =

    21.0058
    19.5285

a =

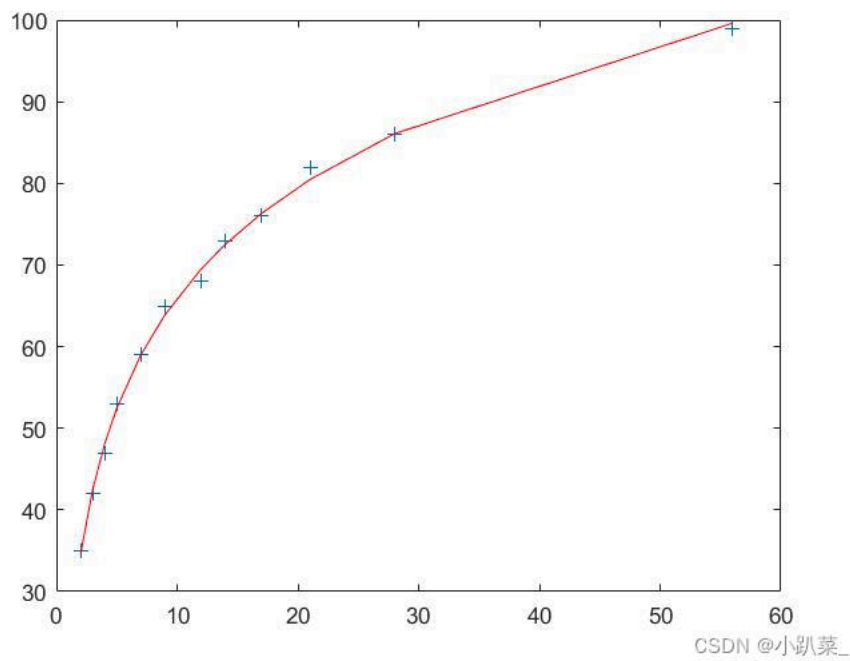
    21.0058

b =

    19.5285
```

fx >> CSDN @小趴菜_

拟合图：



回归方程：y=21.0058+19.5285lnx

2. 聚类分析

为了得出 2007 年江苏省的13个地市的国民经济分布规律，在众多衡量经济水平的指标中我们将采用下列指标：

x1：年末户籍人口（万人）x2：城镇化率(%) x3：地区生产总值 GDP（亿元）x4：第三产业占 GDP 的比重(%)x5：城镇固定资产投资额（亿元）x6：社会消费品零售总额（亿元）x7：城市居民人均可支配收入（元）x8：恩格尔系数城市(%) x9：农村居民人均纯收入（元）x10恩格尔系数农村(%)

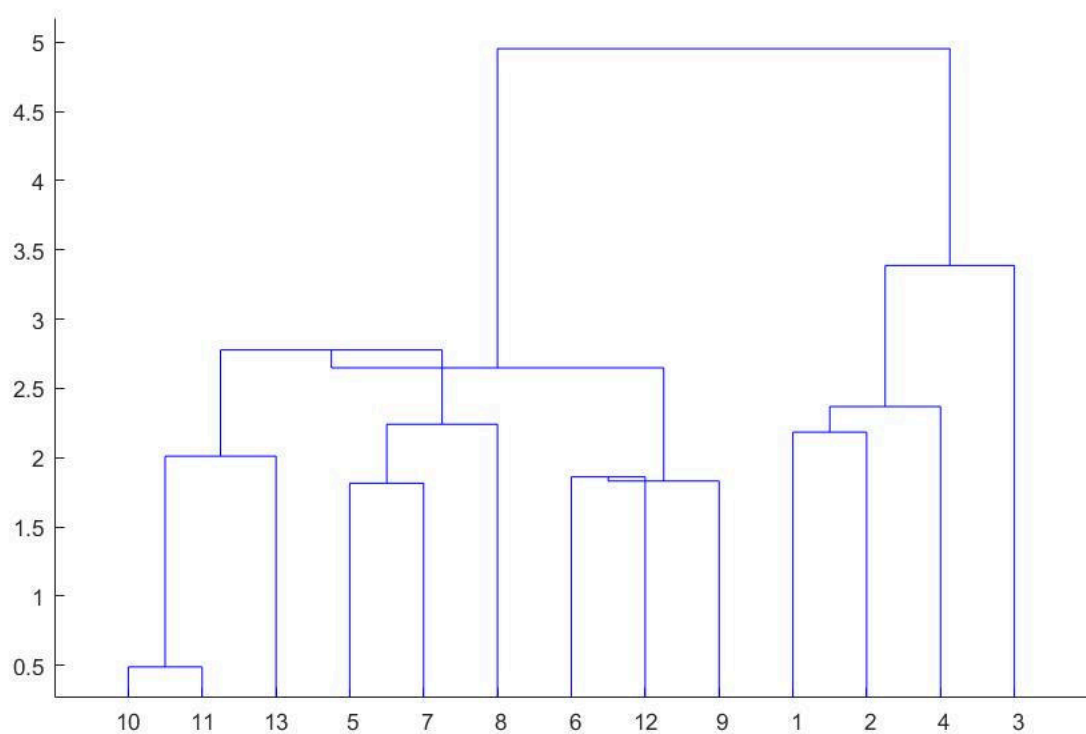
对2007年江苏省13个城市的国民经济进行聚类分析。

地市	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10

地市	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10
常州	357.38	60.9	1881.28	18.6	748.89	610.85	19089	35.0	9033	38.0
南京	617.17	76.8	3283.73	11.0	1443.40	1380.46	20317	35.3	8020	37.4
镇江	268.78	59.6	1206.69	24.5	363.73	331.36	16775	38.7	7668	39.4
南通	766.13	48.6	2111.88	35.1	633.94	736.54	16451	38.5	6905	37.9
扬州	459.25	50.2	1311.89	35.3	438.35	418.90	15057	37.9	6586	38.9
泰州	500.70	47.6	1201.82	33.2	347.73	321.07	14940	43.1	6469	38.1
徐州	940.95	45.8	1679.56	36.0	769.59	543.01	14875	34.9	5534	39.0
连云港	482.23	40.5	618.18	36.2	409.56	249.08	13254	38.9	4828	43.7
淮安	534.00	39.9	765.23	34.8	394.91	269.40	12164	38.9	5010	43.2
盐城	809.79	43.7	1371.26	34.1	470.06	433.74	13857	38.5	6092	41.7
宿迁	531.53	34.1	542.00	32.0	256.18	158.87	9468	42.4	4783	46.0

matlab求解：

```
1 | x = [624.43 65.6 5700.85 7.4 1704.27 1250.05 21260 37.9 10475 35.7
2 | 461.74 67.4 3858.54 9.1 1180.74 1134.75 20898 39.8 10026 37.6
3 | 357.38 60.9 1881.28 18.6 748.89 610.85 19089 35.0 9033 38.0
4 | 617.17 76.8 3283.73 11.0 1443.40 1380.46 20317 35.3 8020 37.4
5 | 268.78 59.6 1206.69 24.5 363.73 331.36 16775 38.7 7668 39.4
6 | 766.13 48.6 2111.88 35.1 633.94 736.54 16451 38.5 6905 37.9
7 | 459.25 50.2 1311.89 35.3 438.35 418.90 15057 37.9 6586 38.9
8 | 500.70 47.6 1201.82 33.2 347.73 321.07 14940 43.1 6469 38.1
9 | 940.95 45.8 1679.56 36.0 769.59 543.01 14875 34.9 5534 39.0
10 | 482.23 40.5 618.18 36.2 409.56 249.08 13254 38.9 4828 43.7
11 | 534.00 39.9 765.23 34.8 394.91 269.40 12164 38.9 5010 43.2
12 | 809.79 43.7 1371.26 34.1 470.06 433.74 13857 38.5 6092 41.7
13 | 531.53 34.1 542.00 32.0 256.18 158.87 9468 42.4 4783 46.0];
14 | % 数据标准化
15 | x1 = zscore(x);
16 | % 采用欧氏距离计算对象之间距离
17 | y1 = pdist(x1,"euclidean");
18 | % 采用质心距离法计算系统聚类数
19 | z1 = linkage(y1,"centroid")
20 | % 利用生成的y1和z1计算cophenet相关系数
21 | c1 = cophenet(z1,y1)
22 | % 利用生成的z1创建分类
23 | t = cluster(z1,6)
24 | % 生成谱系图
25 | h = dendrogram(z1)
26 |
```



CSDN @小趴菜_

由以上结果可知：结合实际情况分析采用质心距离法把13个地市分为两类：

第一类：连云港、淮安、宿迁、镇江、扬州、泰州、南通、盐城、徐州

第二类：苏州、无锡、南京、常州

3. 判别分析

银行的贷款部门需要判别每个客户的信用好坏（是否未履行还贷责任），以决定是否给予贷款。

可以根据贷款申请人的年龄（ X_1 ）、受教育程度（ X_2 ）、现在所从事工作的年数（ X_3 ）、未变更住址的年数（ X_4 ）、收入（ X_5 ）、负债收入比例（ X_6 ）、信用卡债务（ X_7 ）、其它债务（ X_8 ）等来判断其信用情况。下表是从某银行的客户资料中抽取的部分数据，和某客户的如上情况资料为（53，1，9，18，50，11.20，2.02，3.58），根据样本资料分别用马氏距离判别法、线性判别法、二次判别法对其进行信用好坏的判别。

目前信用 好坏	客户 序号	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
已履行还 贷责任	1	23	1	7	2	31	6.60	0.34	1.71
	2	34	1	17	3	59	8.00	1.81	2.91
	3	42	2	7	23	41	4.60	0.94	.94
	4	39	1	19	5	48	13.10	1.93	4.36
	5	35	1	9	1	34	5.00	0.40	1.30
未履行还 贷责任	6	37	1	1	3	24	15.10	1.80	1.82
	7	29	1	13	1	42	7.40	1.46	1.65
	8	32	2	11	6	75	23.30	7.76	9.72
	9	28	2	2	3	23	6.40	0.19	1.29
	10	26	1	4	3	27	10.50	2.47	3.6

CSDN @小趴菜_

matlab求解：

```

1 clc,clear;
2 %用于构造判别函数的训练样本数据矩阵
3 training=[23 1 7 2 31 6.60 0.34 1.71
4           34 1 17 3 59 8.00 1.81 2.91
5           42 2 7 23 41 4.60 0.94 0.94
6           39 1 19 5 48 13.10 1.93 4.36

```

```

10      32 2 11 6 75 23.30 7.76 9.72
11      28 2 2 3 23 6.40 0.19 1.29
12      26 1 4 3 27 10.50 2.47 0.36];
13 %参数group是与training相应的分组变量
14 group=[1;1;1;1;1;2;2;2;2;2];
15 % 待判别的样本数据矩阵
16 sample=[53,1,9,18,50,11.20,2.02,3.58];
17 %[class,err]=classify(sample,training,group,'mahalanobis') %使用马氏距离判别法分类,需要降维
18 [class,err]=classify(sample,training,group,'linear') %使用线性判别法分类
19 %[class,err] = classify(sample,training,group,'diagQuadratic') %使用二次判别法分类
20 %[class,err] = classify(sample,training,group,'quadratic') %使用二次判别法分类,需要降维

```

```

class =

    1

err =

    0

```

该客户属于“已履行还贷责任”类客户，出错概率为0.

显示推荐内容