



Mathematical Modeling

数学建模

同济大学数学科学学院

第十二章

回归模型

目录/Contents | 第十二章 回归模型

第一节 牙膏的销售量

第二节 软件开发人员的薪金



数学建模的基本方法

机理分析

测试分析

由于客观事物内部规律的复杂及人们认识程度的限制,无法分析实际对象内在的因果关系, 建立合乎机理规律的数学模型。

通过对数据的统计分析, 找出与数据拟合最好的模型

回归模型是用统计分析方法建立的最常用的一类模型

不涉及回归分析的数学原理和方法

通过实例讨论如何选择不同类型的模型

对软件得到的结果进行分析, 对模型进行改进

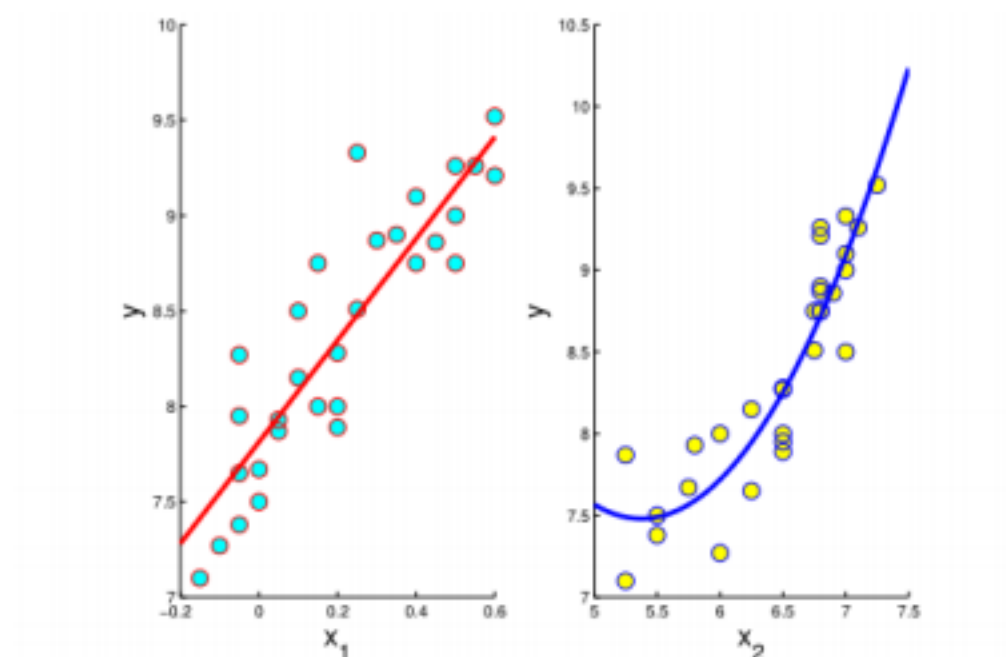
**问题**

建立牙膏销售量与价格、广告投入之间的模型

预测在不同价格和广告费用下的牙膏销售量

收集了30个销售周期本公司牙膏销售量、价格、广告费用，及同期其它厂家同类牙膏的平均售价

销售周期	本公司价格(元)	其它厂家价格(元)	广告费用(百万元)	价格差(元)	销售量(百万支)
1	3.85	3.80	5.50	-0.05	7.38
2	3.75	4.00	6.75	0.25	8.51
...
29	3.80	3.85	5.80	0.05	7.93
30	3.70	4.25	6.80	0.55	9.26



$$y = \beta_0 + \beta_1 x_1 + \varepsilon \quad y = \beta_0 + \beta_1 x_2 + \beta_2 x_2^2 + \varepsilon$$

- y 本公司牙膏销售量(被解释变量)
- x_1 差价, 解释变量, 回归变量
- x_2 广告费, 解释变量, 回归变量
- ε 均值为0的正太分布变量

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_2^2 + \varepsilon$$

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_2^2 + \varepsilon$$

Matlab的统计工具箱， 回归函数 **regress**

[b,bint,r,rint,stats] = regress(y,x,alpha)

- **y**: n 维数据
- **x**: $n \times 4$ 数据矩阵
- **alpha**: 置信水平，通常设为0.05
- **b, bint**: 参数的估计值及其区间
- **r, rint**: 残差的估计值及其区间

结果分析

参数	参数估计值	置信区间
β_0	17.3244	[5.7282 28.9206]
β_1	1.3070	[0.6829 1.9311]
β_2	-3.6956	[-7.4989 0.1077]
β_3	0.3486	[0.0379 0.6594]
$R^2=0.9054$ $F=82.9409$ $p=0.0000$		

- ✓ y 的 90.54% 可由模型确定
- ✓ p 远小于 $\alpha=0.05$
- ✓ β_2 的置信区间包含零点(右端点距零点很近)
- ✓ F 远超过 F 检验 的临界值

销售量预测

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_2^2$$

价格差 x_1 =其它厂家价格 x_3 -本公司价格 x_4

调整 x_4 \Rightarrow 控制 x_1 \Rightarrow 通过 x_1, x_2 预测 y

控制价格差 $x_1=0.2$ 元, 投入广告费 $x_2=650$ 万元

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_2^2 = 8.2933 \text{ (百万支)}$$

销售量预测区间为 [7.8230, 8.7636] (置信度95%)

上限用作库存管理的目标值

下限用来把握公司的现金流

若估计 $x_3=3.9$, 设定 $x_4=3.7$, 则可以95%的把握知道销售
额在 $7.8320 \times 3.7 \approx 29$ (百万元) 以上

模型改进

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_2^2 + \varepsilon$$

 x_1 和 x_2 对 y 的影响独立

参数	参数估计值	置信区间
β_0	17.3244	[5.7282 28.9206]
β_1	1.3070	[0.6829 1.9311]
β_2	-3.6956	[-7.4989 0.1077]
β_3	0.3486	[0.0379 0.6594]
$R^2=0.9054$ $F=82.9409$ $p=0.0000$		

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_2^2 + \beta_4 x_1 x_2 + \varepsilon$$

 x_1 和 x_2 对 y 的影响有交互作用

参数	参数估计值	置信区间
β_0	29.1133	[13.7013 44.5252]
β_1	11.1342	[1.9778 20.2906]
β_2	-7.6080	[-12.6932 -2.5228]
β_3	0.6712	[0.2538 1.0887]
β_4	-1.4777	[-2.8518 -0.1037]
$R^2=0.9209$ $F=72.7771$ $p=0.0000$		

目录/Contents | 第十二章 回归模型

第一节 牙膏的销售量

第二节 软件开发人员的薪金



建立模型研究薪金与资历、管理责任、教育程度的关系

分析人事策略的合理性，作为新聘用人员薪金的参考

46名软件开发人员的档案资料

编号	薪金	资历	管理	教育	编号	薪金	资历	管理	教育
01	13876	1	1	1	42	27837	16	1	2
02	11608	1	0	3	43	18838	16	0	2
03	18701	1	1	3	44	17483	16	0	1
04	11283	1	0	2	45	19207	17	0	2
...	46	19346	20	0	1

资历 ~ 从事专业工作的年数；

管理 ~ 1=管理人员，0=非管理人员；

教育 ~ 1=中学，2=大学，3=更高程度

分析与假设

 y 薪金, x_1 资历 (年) $x_2 = 1$ 管理人员, $x_2 = 0$ 非管理人员

教育

1=中学
2=大学
3=更高

$$x_3 = \begin{cases} 1, & \text{中学} \\ 0, & \text{其它} \end{cases}$$

$$x_4 = \begin{cases} 1, & \text{大学} \\ 0, & \text{其它} \end{cases}$$

中学: $x_3=1, x_4=0$;大学: $x_3=0, x_4=1$;更高: $x_3=0, x_4=0$ 为什么不用
 $x_3=0,1,2$ 表示?资历每加一年薪金的增长是常数;
管理、教育、资历之间无交互作用

线性回归模型

$$y = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_3 + a_4 x_4 + \varepsilon$$

 a_0, a_1, \dots, a_4 是待估计的回归系数, ε 是随机误差

模型求解

$$y = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4 + \varepsilon$$

参数	参数估计值	置信区间
a_0	11032	[10258 11807]
a_1	546	[484 608]
a_2	6883	[6248 7517]
a_3	-2994	[-3826 -2162]
a_4	148	[-636 931]
$R^2=0.957$ $F=226$ $p=0.000$		

$R^2, F, p \rightarrow$ 模型整体上可用

解释

资历增加1年薪金增长546

管理人员薪金多6883

中学程度薪金比更高的少2994

大学程度薪金比更高的多148

a_4 置信区间包含零点，解释不可靠！

结果分析 残差分析方法

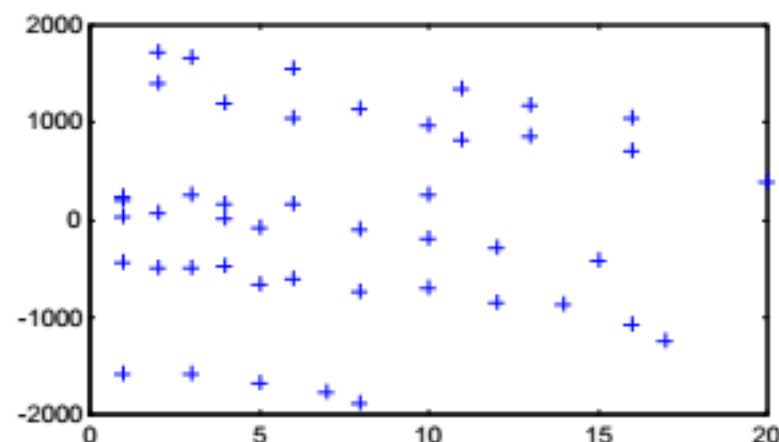
$$\hat{y} = \hat{a}_0 + \hat{a}_1 x_1 + \hat{a}_2 x_2 + \hat{a}_3 x_3 + \hat{a}_4 x_4$$

残差 $e = y - \hat{y}$

管理与教育的组合

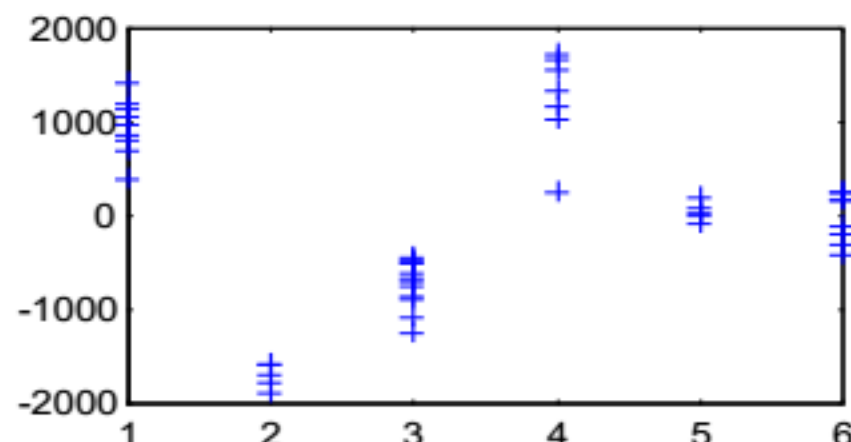
组合	1	2	3	4	5	6
管理	0	1	0	1	0	1
教育	1	1	2	2	3	3

e 与资历 x_1 的关系



残差大概分成3个水平，6种管理—教育组合混在一起，未正确反映。

e 与管理—教育组合的关系



残差全为正，或全为负，管理—教育组合处理不当

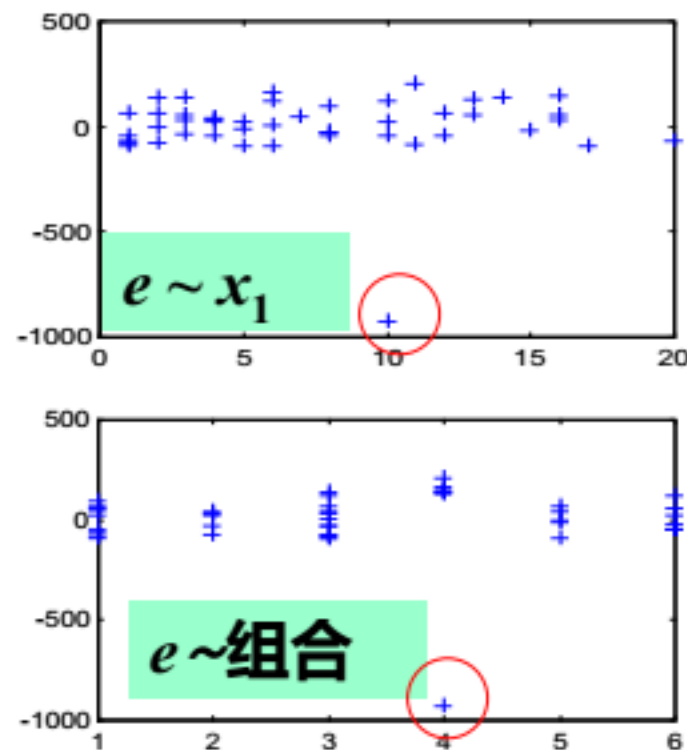
应在模型中增加管理 x_2 与教育 x_3, x_4 的交互项

进一步的模型 增加管理 x_2 与教育 x_3, x_4 的交互项

$$y = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4 + a_5x_2x_3 + a_6x_2x_4 + \varepsilon$$

参数	参数估计值	置信区间
a_0	11204	[11044 11363]
a_1	497	[486 508]
a_2	7048	[6841 7255]
a_3	-1727	[-1939 -1514]
a_4	-348	[-545 -152]
a_5	-3071	[-3372 -2769]
a_6	1836	[1571 2101]
$R^2=0.999$ $F=554$ $p=0.000$		

R^2, F 有改进, 所有回归系数置信区间都不含零点, 模型完全可用



消除了不正常现象

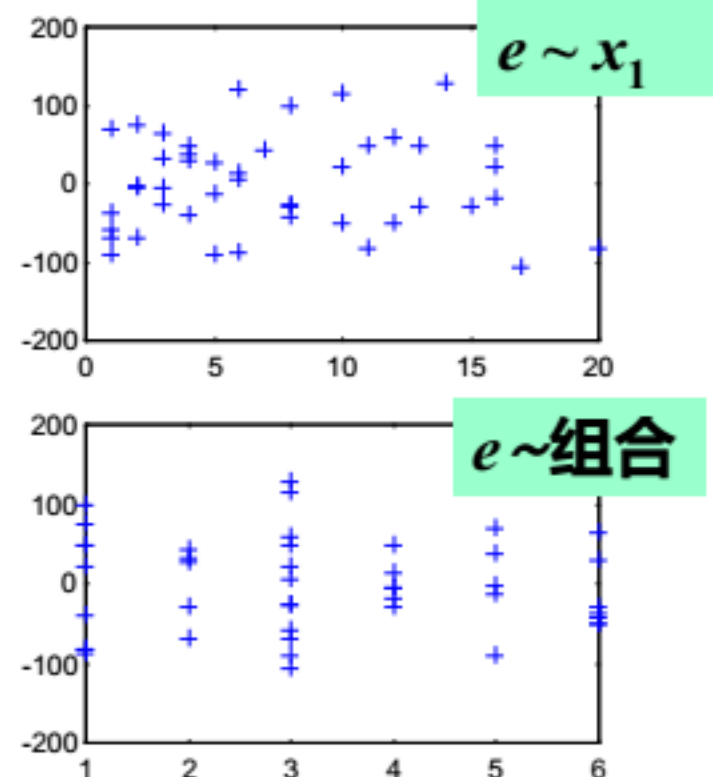
异常数据(33号)应去掉

去掉异常数据后的结果

参数	参数估计值	置信区间
a_0	11200	[11139 11261]
a_1	498	[494 503]
a_2	7041	[6962 7120]
a_3	-1737	[-1818 -1656]
a_4	-356	[-431 -281]
a_5	-3056	[-3171 -2942]
a_6	1997	[1894 2100]
$R^2=0.9998$ $F=36701$ $p=0.0000$		

 $R^2: 0.957 \rightarrow 0.999 \rightarrow 0.9998$ $F: 226 \rightarrow 554 \rightarrow 36701$

置信区间长度更短



残差图十分正常

最终模型的结果可以应用

模型应用

$$\hat{y} = \hat{a}_0 + \hat{a}_1x_1 + \hat{a}_2x_2 + \hat{a}_3x_3 + \hat{a}_4x_4 + \hat{a}_5x_2x_3 + \hat{a}_6x_2x_4$$

制订6种管理—教育组合人员的“基础”薪金(资历为0)

$x_1=0$; $x_2=1\sim$ 管理, $x_2=0\sim$ 非管理

中学: $x_3=1, x_4=0$; 大学: $x_3=0, x_4=1$; 更高: $x_3=0, x_4=0$

组合	管理	教育	系数	“基础” 薪金
1	0	1	a_0+a_3	9463
2	1	1	$a_0+a_2+a_3+a_5$	13448
3	0	2	a_0+a_4	10844
4	1	2	$a_0+a_2+a_4+a_6$	19882
5	0	3	a_0	11200
6	1	3	a_0+a_2	18241

大学程度管理人员比更高程度管理人员的薪金高

大学程度非管理人员比更高程度非管理人员的薪金略低

硅酸盐(Si_3N_4)制陶材料是一种强度高、耐磨、抗氧化和耐高温的材料,它广泛应用于高温结构的材料中,如切割工具、齿轮、内燃机部件及航空、航天飞行器的有关部件等.影响这种材料的强度的因素有:

- A. 加热方案, A_1 =两步, A_2 =一步, 其中“两步”包括“一步”上的预烧结阶段;
- B. 四种烧结添加剂 CaO , Y_2O_3 , MgO 和 Al_2O_3 的总量, B_1 =14%摩尔, B_2 =16%摩尔, B_3 =18%摩尔;
- C. CaO 的含量, C_1 =0.0%摩尔, C_2 =1.0%摩尔, C_3 =2.0%摩尔;
- D. Y_2O_3 的%摩尔与 MgO 的%摩尔的比率, D_1 =1:1, D_2 =1:2, D_3 =1:6;
- E. Y_2O_3 的%摩尔与 Al_2O_3 的%摩尔的比率, E_1 =2:1, E_2 =1:1, E_3 =1:4;
- F. 烧结温度, F_1 =1800°C, F_2 =1850°C, F_3 =1900°C;
- G. 烧结时间, G_1 =1h, G_2 =2h, G_3 =3h.



为了寻找使得该种材料的强度达到最高的工艺条件, 特此安排了如下试验方案, 测量数据见表1.

因素栏中数字 i 表示因素在试验中处于第 i 水平.

- ✓ 根据该表的测量数据, 试建立合理的数学模型, 并对试验结果进行分析;
- ✓ 寻找使得强度最大的最优工艺条件;
- ✓ 对你所建立的模型进行误差分析并做出评价;
- ✓ 你能否提出一种更合理的试验设计计划及试验结果的分析方法?
- ✓ 就你的研究对有关部门试写一份申报科技进步奖的报告.



试验号	因素							强度				
	A	B	C	D	E	F	G					
1	1	2	2	1	3	1	3	996.8	783.6	796.9		
2	1	2	1	2	2	3	1	843.8	816.2	714.3	824.4	
3	1	2	3	3	1	2	2	647.1	667.9	534.3	617.7	
4	1	3	2	1	2	3	2	616.3	552.3	552.6	596.0	
5	1	3	1	2	1	2	3	517.8	526.1	498.1	499.5	
6	1	3	3	3	3	1	1	1002.0	1097.0	882.9	940.1	
7	1	1	2	2	3	2	1	806.5	933.5	964.9	1046.0	
8	1	1	1	3	2	1	2	801.5	803.2	846.2	756.4	
9	1	1	3	1	1	3	3	739.2	863.3	797.0	929.6	
10	2	2	2	3	1	3	1	615.0	627.5	583.9	597.1	563.9
11	2	2	1	1	3	2	2	795.9	854.0	937.0	999.2	724.8
12	2	2	3	2	2	1	3	850.9	921.8	990.6	943.5	840.9
13	2	3	2	2	1	1	2	513.0	665.9	718.9	646.4	
14	2	3	1	3	3	3	3	831.3	981.4	912.5	950.7	987.3
15	2	3	3	1	2	2	1	806.1	908.1	627.6	855.0	
16	2	1	2	3	2	2	3	727.3	643.9	584.0	643.4	602.1
17	2	1	1	3	2	2	3	836.8	716.3	862.9	796.2	
18	2	1	3	1	1	1	1	1001.0	937.6	955.3	995.8	1009.0



Mathematical Modeling 数学建模

同济大学数学科学学院

学海无涯，祝你成功！