

信息与计算科学专业 实验任务书指导书

学生姓名: 李金哲 学号: 201707010119 实验日期: 2020.5.22 成绩:

- (1) 熟悉数据的回归分析方法;
- (2) 熟悉撰写数据分析报告的方法;
- (3) 熟悉数据分析软件 SPSS, 并运用其完成数据回归分析。

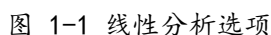
根据题目的具体要求, 完成实验报告。

1、某医院为了解病人对医院工作的满意程度 Y 和病人的年龄 X_1 、病情的严重程度 X_2 和病人的忧虑程度 X_3 之间的关系，随机调查了该医院的 30 位病人，得数据如表（见附近 data.txt）所示。

- 型及有关误差分布正态性假定的合理性;
- (2) 用逐步回归法选择最优回归方程;
- (3) 对选择的最优回归方程做残差分析, 与 (1) 中的相应结果比较有何变化。

—

1) 分析 \rightarrow 回归 \rightarrow 线性 (图 1-1)



2) 将 y 放入因变量，x1、x2、x3 放入块，方法选择输入（图 1-2）

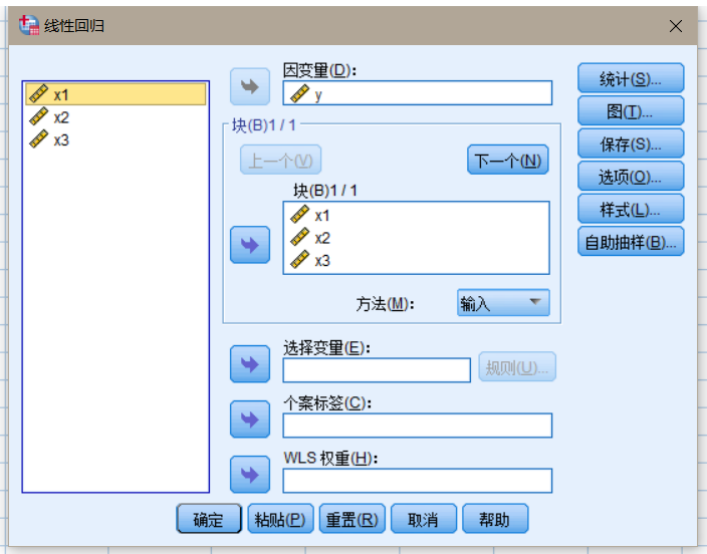


图 1-2 具体选项 1

3) 选择统计，将置信区间勾选；然后点击图将 ZPRED 和 ZRESID 拖入，勾选直方图和正态概率

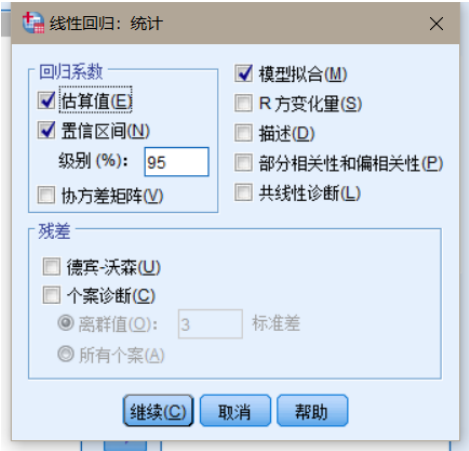


图 1-3 详细信息 (a)



图 1-3 详细信息 (b)

结果分析

表 1-1 检验结果

ANOVA ^a						
模型		平方和	自由度	均方	F	显著性
1	回归	4559.322	3	1519.774	14.810	.000 ^b
	残差	2668.044	26	102.617		
	总计	7227.367	29			
a. 因变量: y						
b. 预测变量: (常量), x3, x1, x2						

表1-2 系数表
系数^a

模型		未标准化系数		标准化系数	t	显著性	B 的 95.0% 置信区间	
		B	标准误差	Beta			下限	上限
1	(常量)	154.034	23.751		6.485	.000	105.213	202.855
	x1	-1.231	.249	-.652	-4.942	.000	-1.743	-.719
	x2	-.436	.729	-.114	-.598	.555	-1.935	1.063
	x3	-8.933	10.362	-.159	-.862	.396	-30.232	12.365

a. 因变量: y

由表 1-1 可知，回归方程显著性检验的 F 统计量的观测值为 14.810，其对应的概率 P-值近似为 0。若显著性水平 α 为 0.05，因概率 P 值小于 α ，拒绝回归方程显著性检验的原假设，即回归系数不同时为 0，解释变量与被解释变量间存在显著的线性关系，选择线性模型具有合理性。

根据表1-2，可以得到回归方程为： $y = 154.034 - 1.231 * X_1 - 0.436 * X_2 - 8.933 * X_3$

但是若显著性水平 α 为0.05，x2的概率p值0.555，大于 α ；x1，x3的概率p值小于 α 。故可以认为x2接受原假设，即x2和y没有显著的线性关系；而x1、x3拒绝原假设，即x1、x3和y存在显著的线性关系。能够判断该回归方程不是最优解

表 1-3 共线性诊断

共线性诊断 ^a							
模型	维	特征值	条件指标	方差比例			
				(常量)	x1	x2	x3
1	1	3.964	1.000	.00	.00	.00	.00
	2	.028	11.971	.03	.95	.01	.01

	3	.007	23.812	.41	.02	.00	.44
	4	.002	48.491	.56	.03	.99	.55
a. 因变量: y							

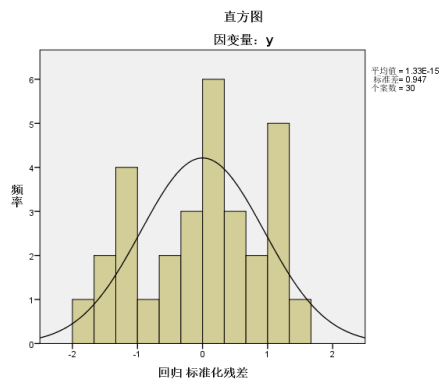


图 1-4 直方图

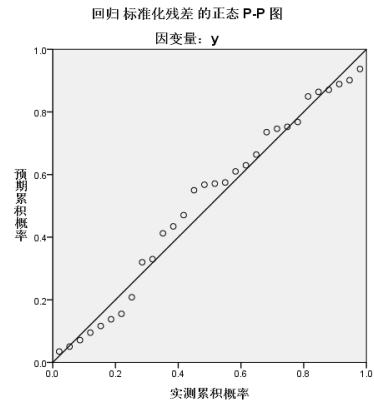


图1-5正正态p-p图

根据表1-3共轭线性诊断，能够得出最大特征值为3.964，其余依次快速减小，可以认为多重共线性较弱。

由图1-5的正态性图像结果图。可以看出，参数围绕准基线存在一定规律性。

二

实验步骤

仅需要将图1-2中的输入替换成步进即可（图2-1），其他的还如图1-1到图1-3



图 2-1 更新后的图1-2

结果分析

表2-1 线性回归分析结果

系数 ^a										
模型		未标准化系数		标准化系数	t	显著性	B 的 95.0% 置信区间		共线性统计	
		B	标准误差	Beta			下限	上限	容差	VIF
1	(常量)	118. 835	9. 179		12. 946	. 000	100. 032	137. 638		
	x1	-1. 433	. 232	-. 759	-6. 170	. 000	-1. 908	-. 957	1. 000	1. 000
a. 因变量: y										

根据表2-1可以得出最优方程为

$$Y = 118.835 - 1.433 * X_1$$

三

结果分析

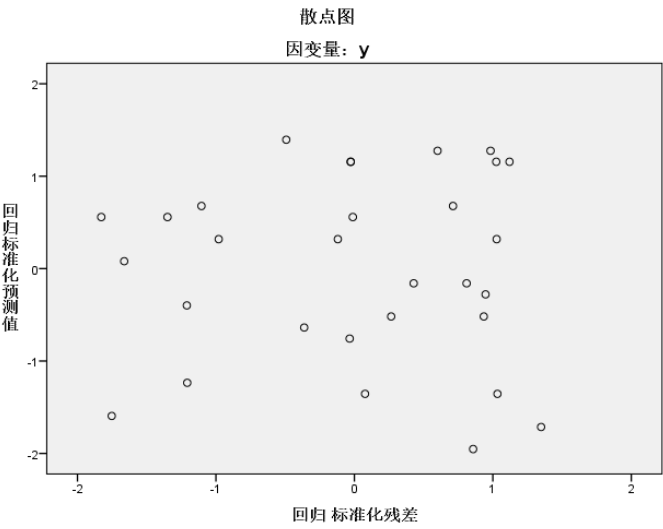


图 3 -1 标准化回归期望值

根据图3-1图像，表明在舍弃X₂和X₃后，不存在明显的异方情况，即问题二得出的方程 $Y = 118.835 - 1.433 * X_1$ 是最终的回归方程。而且当X₁提高1时，y评价改变1.433个变化量