**华中师范大学数学与统计学学院**

**实验报告书**

实验7 共线性

课程名称 统计模型实验

专业 统计学

年级 2020

学生姓名 陈启源

学号 2020211946

指导教师 左国新

华中师范大学数学与统计学学院

2022年11月

**实验课程: 统计分析与软件**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 实验项目编号 | | | 实验7 | 实验项目名称 | | 共线性 | | |
| 实 验 时 间 | | | 11-3 | 实验计算机号 | |  | | |
| 实验目的: 利用Matlab判断共线性 | | | | | | | | |
| 实  验  内  容 | **问题重述：**    **问题求解：**      Y=randn(100,2);       X1=Y(:,1);  X2=Y(:,2);  U=(X1-mean(X1))/sqrt(var(X1));      beta=(X2'\*X2)\X2'\*U;  e=U-X2\*beta;  W=(e-mean(e))/sqrt(var(e));      r=0.3;  V=r\*U+sqrt(1-r\*r)\*W;      mean(U)  var(U)  mean(V)  var(V)  corrcoef(U,V)  ans =  8.8818e-18  ans =  1.0000  ans =  -8.8818e-18  ans =  1.5721  ans =  1.0000 0.9997  0.9997 1.0000  其中，U的均值为0，方差为1，满足理论假设。但是V的均值为0，而方差不为1，而是1.5712。相关系数也不是设定的0.3，而是0.9997。并不精确。      E=randn(100,1);  Y=U+V+E;      X=[U,V];  beta=(X'\*X)\X'\*Y  beta =  -1.1182  2.6351  所以回归方程可以写作：y = -1.1182U + 2.6351V      E=randn(100,1000);  Y=zeros(100,1000);  beta=zeros(1000,2);  for i=1:1000  Y(:,i)=U+V+E(:,i);  beta(i,:)=(X'\*X)\X'\*Y(:,i);  end;      a的直方图    b的直方图    a+b的直方图     1. b的直方图         问题七要求对问题六中分析的四个参数代数分析其相关信息。首先从问题六所画的直方图中可以看到，a和b的均值都为1，那么可以大致判断a和b的真实值应为1，从理论上进行分析，a和b是通过设计矩阵X和响应变量Y直接得到的，是最优化的估计系数，而又由问题三知，Y的值是通过U和V以系数均为1的方程计算出的，也就是说，其实问题四所拟合的回归方程：中所有的数据都是来自问题三中的方程，那么所拟合的回归系数也就自然是问题三中U和V的系数，即都为1，实验的事实也证明如此。  在这些参数：a，b，a+b，a-b中参数a-b应是最容易估计的，而参数a+b是最难估计的，同样的根据最小二乘原理可以知道，在设计矩阵X固定的情况下，a和b的估计依赖与Y的取值，而Y的取值不仅和U、V有关，还与随机误差项有关，也就是说，参数a和b都是依赖与随机误差的，对于同一组参数的估计值a和b，它们的取值中随机误差项所占的比重相同，那么通过a-b就可以消掉随机误差项对其的影响，而a+b则会加倍随机误差项对其的影响，所以说参数a-b是最容易估计的，而参数a+b是最难估计的。  **附加题：**  同样的对问题3，参数U和V的数据都是标准化的数据，所以当将它们向Y回归时也不需要截距。对于第4步中直接利用问题2中得到的U和V的数据来进行多次回归和重新产生U和V的数据进行回归是一样的，从问题2中我们可以看到，U和V是通过Matlab软件按照标准正态分布产生的随机数，即使每次都重复产生，它们本质上都是服从正态分布的，因此均值和方差都是在一个特定值附件摆动，而随机误差项也是有系统随机产生的，那么当U和V固定时它们的摆动情况可以由误差项来描述。当U和V不固定是，它们的摆动情况也可以附加到误差项上，由此可见，不管U和V是不是固定的对于本次模拟实验来说都是没有关系的。 | | | | | | | |
| 小结 | 1. 利用matlab对随机数进行模拟和计算 2. 利用matlab来求解回归方程 | | | | | | |
| 实验成绩 | |  | | | 指导教师 | |  |