

03/2025

作业:

1. 下表是北京 1951-1980 年的一月平均气温(y, 单位: 摄氏度)和 3 个环流要素 x1, x2, x3, 请建立这三个预报因子和预报量 y 的多元回归方程。

x1	x2	x3	y
32.300	10.100	32.600	-2.700
31.100	9.700	31.500	-5.900
32.300	9.400	31.100	-3.400
31.400	8.100	32.100	-4.700
32.500	9.900	32.500	-3.800
31.300	9.100	30.800	-5.300
30.600	10.800	32.300	-5.000
32.300	9.800	31.600	-4.300
31.600	7.100	30.000	-5.700
32.800	10.100	32.800	-3.600
32.900	10.000	31.900	-3.100
32.400	9.500	31.000	-3.900
32.300	10.500	31.300	-3.000
32.000	9.700	31.900	-4.900
30.800	9.000	30.700	-5.700
31.500	10.600	32.700	-4.800
32.000	8.600	29.900	-5.600
31.000	9.800	29.900	-6.400
31.800	9.400	29.400	-5.600
31.900	10.000	33.700	-4.200
31.200	7.700	31.100	-4.900
32.400	10.300	32.500	-4.100
33.600	9.600	31.800	-4.200
31.900	10.300	31.400	-3.300
31.900	8.100	32.400	-3.700
31.900	8.100	28.800	-7.600
32.800	10.900	31.500	-3.500
33.300	10.400	31.000	-4.200
32.300	9.600	32.000	-4.500
30.700	9.900	31.300	-4.800

问题:

1. 回归方程是: $y = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + e$, e 是误差。请使用最小二乘法

$$\mathbf{b} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$$

数值计算回归系数向量 $\mathbf{b} = [b_0 \ b_1 \ b_2 \ b_3]^T$.

2. 解释回归系数的物理意义
3. 计算离差平方和 S_{yy} , 回归平方和 U , 残差平方和 Q , 验证 3 者之间的关系
4. 计算回归解释方差 R^2 , 比较复相关系数 R 和 $r_{y\hat{y}}$
5. 计算调整相关系数 R_a^2 , 比较 R^2 和 R_a^2
6. 对回归方程计算 F 统计量, 进行显著性检验
7. 对回归系数 b_1, b_2, b_3 计算 t 统计量, 进行统计性检验, 是否每个回归因子是否通过显著性检验?
8. 这里有一个回归因子不显著, 将 y 和另外 2 个因子再次做多元线性回归
 - 8.1 计算新的回归系数
 - 8.2 对这个 2 个因子的多元线性回归进行显著性检验
 - 8.3 计算 R^2 和 R_a^2 , 并和第 4 个问题 (3 个因子回归) 得到的 R^2 和 R_a^2 分别进行比较

2. 进行多元线性回归

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + \cdots + b_px_p$$

后, 得到新的量 \hat{y} , 比较 \hat{y} 和 y , 下面哪些统计量发生了变化? 为什么?

- 算术平均值
- 方差(标准差)
- 相关系数