北京建筑大学 2019/2020 学年 第 2 学期 考试

数据分析 课程期末 考试试卷 B

题号	_	1	11	四	五	六	七	八	九	+	总分
题分	30	20	25	25							100
得分											
阅卷人											
复核人											

一、附件 1 (data1. sav) 是 2018 年北京市西城区某社区医院对 110 位老人的健康监测体 检数据,包括 YW、TC、TG、HbA1C、HDLC、LDLC 六项指标。请按照题目的要求,选择适当的方法完成下列各题:

- (1) 计算 HbA1C、HDLC、LDLC 三项观测数据的均值:
- (2) 对 HbA1C 指标计算方差、标准差、偏度、峰度;
- (3) 对 HbA1C 计算中位数、上下四分位数;
- (4) 对 HbA1C 作出直方图;
- (5) 计算观测数据 HbA1C、HDLC、LDLC 的 Spearman 相关矩阵 Q,并做相关性的显著性检验。

1、考生必须将姓 名、学号等项填 写在装订线左 侧内,不得超 出,否则按违纪 处理。

姓名:

学号:

班级:

- 2、请将证件放在 桌角处备查。
- 3、 遵守考场规则。
- 4、凡违反考纪者 按规定给予相 应处分。
- 二、为研制一种治疗枯草热病的药物,将两种成分 (A和B) 各按三种不同剂量 (低、中、高)混合,将36位自愿受试患者随机分为9组,每组4人服用各种剂量混合下的药物,记录其病情缓解的时间(单位:小时)(数据见附件(data2.sav)),假设所给数据服从方差分析模型。
- (1)建立方差分析表,分别判断 A 与 B 两因素在显著性水平 α =0.05 下是否显著?
- (2)分别计算因素 A 的各水平 Ai 上的均值 μ i 和因素 B 的各水平 Bj 上的均值 μ j,及其 95%的置信区间 (i, i=1,2,3)。

三、某企业为了研究其生产的商品在全国二线城市的年销售量 Y(单位:万台)与该城市人口数 X1(单位:万人)以及该市人均月收入 X2(单位:元)之间的关系,2018 年该企业在 20 个城市做了调查,得上述各量的观测值(数据见附件 data3. sav)。

(1) 对年销售量 Y 与 X1、X2 之间做线性回归分析,请你根据概率 p 值,对 α =0.05,检验 Y 与 X1、X2 之间满足下列哪种线性关系

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon \quad Y = \beta_0 + \beta_1 X_1 + \varepsilon \quad Y = \beta_0 + \beta_2 X_2 + \varepsilon$$

并求出相应回归方程的回归系数,写出回归方程并对回归系数作出解释;

- (2) 分别求出由(1) 得到的线性方程系数的置信度为 95%的置信区间:
- (3)该家电企业欲根据(1)得到的线性关系,对另外一城市预测下一年的家电销售量,该市 X1=180 (万人),人均月收入 X2=3240 的新的城市中销售其产品,求其销量的预测值及置信度为 95%的置信区间(根据(1)得到的线性方程预测)。

四、数据(见附件 data4. sav): 给出了 1991 年我国 30 个省市、城镇居民的月平均消费数据,所考察的八个指标如下:(单位均为元/人)

X1: 人均粮食支出;

X2: 人均副食支出;

X3: 人均烟酒茶支出;

X4: 人均其他副食支出;

X5:人均衣着商品支出:

X6:人均日用品支出;

X7:人均燃料支出;

X8: 人均非商品支出。

- (1) 求样本相关系数矩阵 R。
- (2) 从 R 出发做主成分分析,求出各主成分的贡献率及前两个主成分的累积贡献率;
- (3) 根据(2)中得到的前两个主成分,按最短距离法(最近邻元素),对样本进行谱系聚类分析;其中距离采用欧式平方距离,并画出谱系聚类图:给出分为3类的聚类结果。