**一、简答题**

1. 主成分分析的基本思想及其作用？

2. 请阐述主成分分析法的基本步骤。

3. 用主成分分析法进行综合评价时，如何构建综合评价函数？

4. 简述主成分分析的应用。

5. 简述提取样本主成分的原则。

6. 简述主成分分析的适用范围。

7. 简述量纲对主成分分析的影响及消除方法。

**二、计算题**

1. 为研究某一树种的叶片形态，选取50片叶测量其长度和宽度，按数据（略）求得平均值和协方差矩阵为



（1）求出S的两个特征值及其对应的单位特征向量；

（2）写出主成分表达式；

（3）求两个主成分的方差贡献率；

（4）解释两个主成分的实际意义；

（5）如果一片叶的长、宽数据是，试计算它的两个主成分得分，据此你对这片叶子的形态作如何判断？

（6）写出主成分的协方差矩阵；

（7）求主成分与每一个原变量的相关系数，主成分对每一个原变量的方差贡献。

2. 某中学十二名女生的身高x1，体重x2的数据如下：

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** |
| **身高** | 155 | 153 | 157 | 154 | 158 | 152 | 160 | 156 | 158 | 157 | 159 | 161 |
| **体重** | 48 | 46 | 48 | 45 | 49 | 46 | 48 | 45 | 51 | 47 | 48 | 50 |

根据数据计算两个变量协方差矩阵的特征值及其对应的单位特征向量，写出主成分表达式，并解释主成分的实际意义。

**三、上机分析题**

1. 用主成分分析探讨城市工业主体结构，EXE5\_1是某市工业部门13个行业8个指标的数据。

X1：年末固定资产净值（万元） X2：职工人数（人）

X3：工业总产值（万元） X4：全员劳动生产率（元/人年）

X5：百元固定资产原值实现产值（元） X6：资金利税率（％）

X7：标准燃料消费量（吨） X8：能源利用效果（万元/吨）

（1）试利用主成分分析确定8个指标的几个主成分（综合变量），并解释主成分的含义；

（2）利用主成分得分对13个行业进行排序和分类。

2．EXEC5\_2是美国50个州每10万人中7种犯罪形式的比率数据，这7种犯罪分别是：

X1：杀人罪， X2：强奸罪， X3：抢劫罪， X4：斗殴罪

X5：夜盗罪， X6：偷盗罪， X7：汽车犯罪

试对数据进行主成分分析。

3. EXE5\_3是反映我国31个大陆地区高新技术产业竞争力的13项指标数据。请用主成分分析对地区高新技术产业竞争力状况进行分析比较。

X1 人均生产总值（元）

X2 企业数（个）

X3 资产总计（亿元）

X4 主营业务收入（亿元）

X5 利润总额（亿元）

X6 R&D经费内部支出（万元）

X7 新产品开发项目数（项）

X8 新产品开发经费支出（万元）

X9 新产品销售收入（万元）

X10专利申请数（件）

X11拥有发明专利数（件）

X12新增固定资产（亿元）

X13 从业人员年平均人数（人）

4. 用主成分分析对各省级区域的区域创新能力状况进行分析，EXE5\_4是反映各区域创新能力的指标数据。

X1：有研发机构的企业数（个）

X2：R&D经费内部支出占主营业务收入比重（%）

X3：R&D经费外部支出占主营业务收入比重（%）

X4：R&D人员占从业人员比重（%）

X5：每万从业人员有效发明专利数（项）

X6：新产品销售收入占主营业务收入比重（%）

X7：技术市场成交金额（万元）

X8：大专学历以上人数（人）

X9：人均GDP（元）

X10：高新技术产业企业占比（%）

5. EXE5\_5中的数据包含以下4个指标。

Y 国内旅游人数（百万人）

X1 高速公路里程数（万公里）

X2 城镇居民人均可支配收入（元）

X3 农村居民人均纯收入（元）

（1）计算4个变量的相关系数矩阵；

（2）用普通最小二乘估计建立Y与X1, X2, X3的回归方程；

（3）利用主成分方法建立Y与X1, X2, X3的回归方程。