

应用数理统计简述题

1、简述线性回归分析的作用、原理及计算步骤。

作用：确定两种或两种以上变量间相互依赖的定量关系

原理：在含有变量的系统中，一些变量对另一些变量之间可能存在或简单或复杂的函数关系，于是我们可以找出一个最优的函数，用这个函数来表示因变量与自变量的关系

步骤：(1)根据自变量与因变量的现有数据以及关系，初步设定回归方程

(2)求出合理的回归系数

(3)进行相关性检验，确定相关系数

(4)在符合相关性要求后，即可根据已得的回归方程与具体条件相结合，来确定事物的未来状况，并计算预测值的置信区间

2、简述主成分分析的作用、原理及计算步骤。

作用：信息降维

原理：通过构造原变量的适当的线性组合,以产生一系列互不相关的新信息,从中选出少数几个新变量并使它们含有尽可能多的原变量带有的信息,从而使使用这几个新变量代替原变量分析问题和解决问题成为可能.当研究的问题确定之后,变量中所含“信息”的大小通常用该变量的方差或样本方差来度量.

步骤：(1)对原始数据进行标准化

(2)计算相关系数矩阵

(3)计算特征值与特征向量

(4)选择 p 个主成分，计算每个主成分的贡献率

(5)得到各主成分得分，确定要使用的前 $m(m < p)$ 个主成分。

3、简述典型相关分析的作用、原理及计算步骤。

作用：研究多个变量间的相关关系

原理：为了从总体上把握两组指标之间的相关关系，分别在两组变量中提取有代表性的两个综合变量 U 和 V （分别为两个变量组 X 和 Y 中各变量的线性组合），利用这两个综合变量之间的相关关系来反映两组指标之间的整体相关性。

步骤：(1)计算协方差矩阵

(2)确定 a, b 使得 U 与 V 之间的相关系数最大，此时得到一对 U 和 V 称为典型变量。其中 a, b 分别为两个变量组 X 和 Y 的线性组合系数。

(3)如果只有一对 U, V 还不足以反映 X 和 Y 之间的相关性，可进一步构造与 U, V 互不相关的另外一对典型变量。

4、简述因子分析的作用、原理及计算步骤。

作用：因子分析是一种数据简化的技术。

原理：它通过研究众多变量之间的内部依赖关系，探索观测数据中的基本结构，并用少数几个假想变量来表示其基本的数据结构。这几个假想变量能够反映原来众多变量的主要信息。原始的变量是可观测的显在变量，而假想变量是不可观测的潜在变量，称为因子。

步骤：(1)对原始数据标准化，以消除量纲的影响

(2)建立指标间的相关系数阵 R

(3)求 R 的特征值和特征向量

(4)建立因子载荷阵

(5)对因子载荷阵实行方差最大旋转

5、简述贝叶斯判别法的作用、原理及计算步骤。

作用：解决分类问题

原理：根据总体的先验概率，使误判的平均损失达到最小而进行的判别。

步骤：(1)根据已知样本，估计出类别先验概率和类别条件概率

(2)根据贝叶斯公式，结合类别先验概率和类别条件概率得到所给样品的判别分类

5、简述距离判别法的作用、原理及计算步骤。

作用：解决分类问题

原理：根据已知分类的数据，分别计算各类的重心即分组(类)的均值，判别准则是对任给的一次观测，若它与第 i 类的重心距离最近，就认为它来自第 i 类。

步骤：(1)计算各类的均值

(2)计算所给样品到每类的均值的距离

(3)得到最短距离，最短距离所对应的类别即为样本的类别

5、简述判别法的作用、原理及计算步骤。

作用：解决高维空间不可分而在低维空间可分的分类问题

原理：把 d 维空间的样本投影到一条直线上，形成一维空间，即压缩到一维，这条直线满足：不同类别的样本点在此直线上的投影点尽可能地分开。于是可以将 d 维分类问题转化为一维分类问题，只要确定一个阈值，将所给样本的投影点与该阈值做比较，便可以做出决策。

步骤：(1)计算各变量的均值及类内协方差矩阵

(2)建立判别式

(3)求判别临界值，对所给样品判别分类

免责声明：简述题均为个人结合资料整理，请同学们自行斟酌。

祝大家考的全会，蒙的全对！