# 东北大学2018年复试真题

1. **简答题**
2. **什么是充分统计量？解释其概念及意义。**

**答：设总体的分布函数为为未知参数，,为样本，为不带有未知参数的统计量，若在给定统计量时，的条件分布与无关，则称为的充分统计量。**

**充分统计量已经从样本中充分的提取了所包含的有关总体参数的全部信息。**

1. **简要概述什么是主成分分析法，其大致步骤是什么？**

**答：主成分分析的思想:主成分分析是利用降维的思想，在损失很少信息的前提下，把多个指标转化为几个综合指标的多元统计方法。其思想是在保留原始变量尽可能多的信息的前提下达到降维的目的，从而简化问题的复杂性并抓住问题的主要矛盾。**

**主成分分析的基本步骤如下:**

**① 根据研究问题选取初始分析变量；**

**② 根据初始变量特性判断由协方差阵求主成分还是由相关阵求主成分；**

**③ 求协方差阵或相关阵的特征根与相应标准特征向量；**

**④ 判断是否存在明显的多重共线性，若存在，则回到第①步；**

**⑤ 得到主成分的表达式，并确定主成分个数，选取主成分；**

**⑥ 结合主成分对研究问题进行分析并深入研究。**

**3、简述区间估计的含义。**

**答：设为一给定的常数，满足，如果关系式:**

**成立，并用这个随机区间作为参数的估计，则称［］是参数的置信水平为的区间估计，称为显著性水平，分别为上、下置信限。**

**其95%置信区间的含义为：如果用某种方法构造的所有区间中有95％的区间包含总体参数的真值，5％的区间不包含总体参数的真值，那么，用该方法构造的区间称为置信水平为95％的置信区间。**

**区间估计的一些特点：**

* 1. **总体参数的真值是固定的、未知的，而用样本构造的区间则是不固定的。**

**②由一个(特定)样本所构造的区间是一个特定的区间，不再是随机区间，即一个特定的区间“总是包含”或“绝对不包含”参数的真值，不存在“以多大概率包含总体参数”的问题。**

**4、描述方差分析的相关内容。**

**答：通过检验各总体的均值是否相等来判断分类型自变量对数值型因变量是否有显著影响。它是通过对数据误差来源的分析来判断不同总体的均值是否相等，进而分析自变量对因变量是否有显著性影响。**

**方差分析中的基本假定:**

**①（正态性）：每个总体(水平)都应服从正态分布；**

**②（方差齐性）：各个总体的方差必须相同；**

**③（独立性）：各观测值是相互独立的；**

**④各总体误差的期望值为零。**

**单因素方差分析的步骤:**

**①计算各样本的样本均值；**

1. **计算全部观测值的总均值；**
2. **计算各误差平方和；**
3. **计算统计量的值；**
4. **做出决策。**

**误差分解：**

**①是对随机误差和系统误差大小的度量，它反映了自变量对因变量的影响，也称自变量效应或因子效应；**

**②是对随机误差大小的度量，它反映了除自变量对因变量的影响之外的其他因素对因变量的总影响，因此，也称为残差变量，它所引起的误差称为残差效应；**

**③是对全部数据总误差程度的度量，它反映了自变量和残差变量的共同影响，因此它等于自变量效应加残差效应。**

1. **证明题**

**1、多元回归分析模型， (I为单位矩阵)，，已有样本观测.**

**(1) 求的最小二乘估计.**

**解：记为的最小二乘估计，令**

**由极值条件知，可由对求偏导，并令其为零，即**

**由于总是假设满秩，故可逆，从而得到为**

**(2) 证明：.**

**证：由上面所求由，又，则**

1. **总体,是来自总体的样本，显著性水平为，试给出*、*的置信区间的推导过程.**

**证：由于未知，需要对其进行估计，已知**

**其中**

**由抽样分布定理知，**

**再由**

**可以解得的置信区间为**

**又知**

**则有**

**令**

**可得的置信区间为**

1. **计算综合题**
2. **将下面方差分析表补充完整，并判断因子对因变量是否有显著性影响，给定显著性水平*.***

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **误差来源** | **平方和** | **自由度** | **均方** | **比** |
| **组间误差** |  |  |  |  |
| **组内误差** |  |  |  |  |
| **总计** |  |  |  |  |

1. **给定协方差阵，从协方差阵出发，求主成分和贡献率.**

**解：根据主成分分析的步骤，先求的特征值和标准正交特征向量，令**

**解得**

**对应的标准正交特征向量分别为**

**记其主成分分别为，原始变量为，则其主成分为**

**记各主成分的贡献率为，则**

1. **给出五个样本，将每个样品归为一个类，记为，由欧氏距离计算距离矩阵如下，利用最短距离法进行聚类，并画出聚类系谱图.**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

**解：用表示类与类之间的距离，其中距离最小的为故将合并为一个新类，记为，进一步计算各类与新类之间的距离矩阵**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

**这里为最小距离，将进行并类，记新类为，重新计算各类与之间的距离矩阵**

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

**这一步的最小距离为，将进行并类，记为。**

**计算之间的距离为**

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |
|  |  |  |

**将进行并类，记为，聚类完成。**

**聚类过程如下：表示将样品聚为一类**

**聚类系谱图如下：**

