# 2019年东北大学432统计学

**一、简答题(共9题，每题5分，共45分)**

**1、简述分布与分布之间的关系。**

**答：分布定义：若随机变量，，且、相互独立，则称随机变量**

**是服从分布的。**

**分布定义：若随机变量，，且、相互独立，则称随机变量**

**是服从分布的。**

**显然，由上述定义可知若一个随机变量，则有。**

**2、简述两个随机变量独立与不相关的定义，并指出在什么样的条件下两个随机变量的独立与不相关是等价的。**

**答：独立：记随机变量的概率密度函数分别为，分布函数分别*为。*称随机变量是独立的，若**

**或**

**不相关：记为两个随机变量，若的协方差**

**则称不相关。**

**当均服从正态分布时，独立与不相关是等价的。**

**3、对正态总体的参数进行区间估计和假设检验，二者的结论中有什么共同之处？**

**答：二者所选取的统计量相同，均为统计量。**

**二者使用同一样本，同一显著性水平。**

**的双侧置信区间即为其双边假设检验的接受域。**

**的单侧置信区间即为其单边假设检验的接受域。**

**4、请说明在假设检验中原假设和备择假设不能随意互换位置的原因。**

**答：①一般来说，哪一类错误所带来的后果越严重，危害越大，在假设检验中就应当把哪一类错误作为首要的控制目标。在假设检验中，一般首先控制犯错误。主要的原因在于，原假设是什么通常是明确的，而备择假设是什么通常是模糊的。**

**②在进行显著性检验时，犯第Ⅰ类错误的概率是由我们控制的，取得小，则概率｛当为真拒绝｝就越小，这保证了当原假设为真时错误地拒绝原假设的可能性很小。这意味着原假设是受到保护的，也表明原假设和备择假设的地位不是对等的。**

**因此，在假设检验中原假设和备择假设不能随意交换位置。**

**5、简述最小二乘法的思想。**

**答：选择使得误差平方和达到最小的参数值作为总体参数的取值，这样的方法称为最小二乘法。**

**6、在多元回归分析中，出现多重共线性会带来什么问题？**

**答：①变量之间高度相关时，可能会使回归的结果混乱，甚至会把分析引入歧途。**

**②多重共线性可能对参数估计的正负号产生影响，特别是正负号有可能与预期的正负号相反。**

**7、在多元回归分析中，为什么要对拟合优度进行调整？**

**答：由于增加自变量的个数总是使得因变量的变差中被回归方程所解释的比例增大，为避免增加自变量而高估，统计学家提出用样本量和自变量个数去调整。**

**8、简述最优无偏估计量的定义。**

**答：设为总体参数的一个无偏估计，即，为任意一个的无偏估计，若**

**总成立，则称为的最优无偏估计量。**

**9、简述两类错误的定义，并指出其关系。**

**答：第Ⅰ类错误是原假设为真却被我们拒绝了，犯这种错误的概率用表示，所以也称错误或弃真错误；第Ⅱ类错误是原假设为伪我们却没有拒绝，犯这种错误的概率用表示，所以也称错误或取伪错误。**

**对于一定的样本量，如果减少错误，就会增大犯错误的机会；若减少错误就会增大犯错误的机会。增大样本量可以同时减少和。**

**二、计算题(共5题，每题15分，共75分)**

**1、设相互独立，且分别服从参数为的指数分布，概率密度函数分别为。回答下列问题**

**(1)**

**解：由于变量间相互独立，则有**

**(2)**

**解：**

**(3) 若，求。**

**解：**

**2、设总体，今从中抽取容量分别为10和15的两个独立样本，试求这两个样本均值之差的绝对值大于0.3的概率是多少？(注：标准正态分布)。**

**解：记两样本的均值分别为，则**

**由于两样本独立，故也独立，从而有**

**于是**

**3、设随机变量的概率密度函数为：，是来自总体的一个样本，求：**

**(1) 的矩估计量**

**解：随机变量的一阶矩为**

**样本一阶矩为**

**由矩估计原理，令**

**得的矩估计为**

**(2) 的极大似然估计量**

**解：参数的似然函数为**

**显然是关于的单调递减函数，又若使得，则有**

**故的极大似然估计为**

**4、某生物科技公司为生产一种新产品，需要对产品重复进行伯努利实验，已知每次实验成功的概率为6%，实验进行200次，求200次实验中成功次数不少于6次的概率。(注：结果保留小数点两位；可能用到的数据：标准正态分布，，)。**

**解：由题意知**

**其中，。则**

**有中心极限定理知**

**即**

**故**

**5、在某校的一个班级中随机抄录了25名男生的身高数据，测得平均值，标准差，假定男生身高服从正态分布，试求：(所有结果保留小数点后两位)**

**(1) 的置信水平为0.95的置信区间(上分位数：)**

**解：样本量，因此选择统计量，有**

**由**

**得到的置信水平为0.95的置信区间为**

**(2) 的置信水平为0.95的置信区间(上分位数：，)**

**解： 由抽样分布定理知**

**再由**

**得到的置信水平为0.95的置信区间为**

**三、论述题(共2题，每题15分，共30分)**

**1、单因素方差分析模型为：，其中相互独立同分布于，未知，，，调查得到的样品数据如下：**

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | **…** |  |
|  |  | **…** |  |
|  |  | **…** |  |
| **…** | **…** | **…** | **…** |
|  |  | **…** |  |

**样本总数为，记的均值为，样本总均值为。**

**定义总平方和：**

**组间平方和：**

**组内平方和：**

**证明：**

**(1)**

**证：**

**其中**

**故有**

**(2)**

**证：由于各总体都是正态的，由抽样分布定理知**

**又各总体间方差相等且相互独立，由卡方分布的独立可加性可得**

**即**

**2、一元线性回归分析中，，其中，做次独立观测，得到一个容量为的样本，记，使用最小二乘法求得的估计量为。**

**(1) 求**

**解：由，可得**

**又由假设知，故**

**(2) 求的无偏估计量**

**解：记**

**其中**

**由于**

**故**

**即为的无偏估计。**

**(3) 请构造出的原假设的拒绝域。**

**解：记**

**则**

**由**

**得到拒绝域为**

**或**