### 2017年东北大学研究生考试432统计学

**一、简答题（每题5分，共45分）**

**1.简答指数分布的无记忆性。**

**答：指数分布的定义：称随机变量服从参数为的指数分布，如果其概率密度函数具有形式**

**指数分布的无记忆性是指，，因为**

**2.简答一维随机变量分布函数的性质。**

**答：设一维连续型随机变量的分布函数为，则有如下性质，**

**①非负性：即对任意给定的实数，有。**

**②有界性：，且**

**③单调不减：对任意的实数，若，则成立**

**④右连续性：**

**3.简答正态分布的“”法则。**

**答：设随机变量，则**

**约有68％的数据在平均数个标准差的范围内；**

**约有95％的数据在平均数个标准差的范围内；**

**约有99％的数据在平均数个标准差的范围内。**

**在平均数个标准差之外的数据，统计上称为离群点。**

**4.简答箱线图（盒子图）包含的五个数字及箱线图的绘制过程。**

**答：箱线图是由最大值、最小值、上四分位数、下四分位数、中位数构成的，主要用于反映原始数据的分布特征，也可用于比较不同数据分布特征的比较研究。**

**绘制过程：首先画出一条直线，在直线上标出各个数字特征的位置，连接两个四分位数画出箱子，将最大值与上四分位数连接，最小值与下四分位数连接，中位数在中间。**

**5.简答无偏估计的意义。**

**答：无偏估计的定义：设为总体参数，为的一个估计量，若满足**

**则称为的无偏估计。**

**无偏估计的意义是，的某个估计量在具体取值时有时大于参数真值，有时小于参数真值，但在期望的意义下是相等的**

**6.简答假设检验的两类错。**

**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。**

**7.简答样本均值与中位数的区别。**

**答：样本均值与中位数都是描述数据集中趋势的统计量，但样本均值容易受到极端值的影响，稳健性较差；中位数顺序统计量，它不受样本极端值的影响，相对于样本均值具有稳健性。另外，样本均值用于度量数值型数据，而中位数可用于顺序型数据和数值型数据。**

**8.简答方差分析与回归分析的区别。**

**答：①方差分析是研究的是变量间的相关关系，而回归分析研究的是变量间的线性关系。**

**②方差分析中两个变量都是随机的，在回归分析中，因变量是随机的，而自变量是非随机的。**

**③方差分析中两变量的地位是相同的，因此可交换；而在回归分析中，因变量与自变量通常存在依赖关系，因此地位不可交换。**

**④方差分析的检验统计量是F统计量，回归分析中对回归系数进行显著性检验时用的是t统计量，对方程进行显著性检验时用的是F统计量。**

**9.矩估计的原理和理论基础。**

**答：矩估计原理是以样本矩或样本矩的连续函数来代替总体矩或总体矩的连续函数。**

**理论基础是大数定律，即前个随机变量的期望的均值收敛于他们均值的期望。**

**二、计算题（每题15分，共75分）**

**1.为来自于总体的简单随机样本，试求：**

**注:**

**（1）**

**解：由题意知**

**则**

**（2）**

**解：由题，若记的分布函数为，则的分布函数为:**

**故**

**其中**

**从而**

**（3）**

**解：由(1)知**

**故**

**于是**

**2.设来自正态总体，样本容量为18。**

**（1）确定常数，使得统计量服从分布。**

**解：由题可得**

**已知相互独立，从而和也相互独立，由t分布的定义可知**

**即当时，统计量U服从t分布。**

**（2）确定常数，使得统计量服从分布。**

**解：由(1)中讨论知**

**同样地，和相互独立，由F分布的定义可知**

**即当时，统计量U服从F分布。**

**3.设样本来自总体，总体的概率密度如下，是未知参数，请给出的矩估计量和最大似然估计量。**

**解：由题意**

**样本均值为**

**由矩估计原理，令**

**得**

**的极大似然函数为**

**对数似然函数为**

**上式对求偏导得**

**即函数单调递增，为使得，需成立**

**因此，的极大似然估计为**

**4.设是来自于总体服从泊松分别的样本，是未知参数，设估计量**

**（1）试说明是的无偏估计。**

**解：由题意知**

**从而**

**故均为的无偏估计。**

**（2）上述无偏估计中哪个估计是更有效。**

**解：由题**

**因为**

**故最有效。**

**5.为了称某大型设备的重量，将设备100个零部件分别称重，设每个零部件称重误差（单位kg）服从均匀分布，请利用中心极限定理估计整个设备称重的点误差绝对值超过30kg的概率（注：）**

**解：每个零件的重量用一个随机变量来表示，则**

**其期望和方差分别为**

**由中心极限定理知**

**即**

**则**

**三、证明题（每题15分，共30分）**

**1.从某品牌某规格的所有产品中，随机抽取5袋，测量其重量如下：（单位：g）502，497，508，501，502并假定每袋产品的重量服从正态分布，参数均未知，请问在显著性水平时，是否可以接受假设:这批产品每袋的平均重量为500g（注：，）**

**解：每个样本的重量用来表示，则样本的均值和方差分别为**

**选择t检验统计量，当原假设成立时，计算统计量的值为**

**由于，落在接受域内，故接受原假设，认为这批产品每袋的平均重量为500g。**

**2.对于一元线性回归模型，为已知常数，并记为某容量为的样本。**

**（1）请用最小二乘法给出参数的估计值。**

**解：记**

**其中，为的最小二乘估计。**

**上式分别对求偏导并令其为0，得**

**解得**

**（2）说明是的无偏估计。**

**证：的期望为**

**由于在回归分析中总是假定自变量是非随机的，故**

**故是的无偏估计。**