# 东北大学2019年复试真题

1. **简答题**
2. **写出第一类错误和第二类错误的定义，并说明在实际中应如何控制两类错误。**

**答：在假设检验中，我们把原假设为真却被拒绝而犯的错误称为第一类错误，犯这类错误的概率为，故也称为错误或拒真错误；把原假设不真却被接受而犯的错误称为第二类错误，犯这种错误的概率为，故也称为错误或取伪错误。**

**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。**

**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。**

1. **写出值的定义并说明值与简答拒绝域与接受域的关系。**

**答：对于给定的显著性水平，当选择作为检验统计量时，若的分布函数为，则接受域与拒绝域可被定义如下：**

**拒绝域：**

**接受域：**

**显然，，其中为全集；，为空集。其意义为，当我们对某一假设进行假设检验时，其结果要么为接受原假设，要么为拒绝原假设。**

**值被定义为，当原假设为真时，样本观察值或更极端结果出现的概率。值越小，拒绝原假设的理由越充分。**

**对于给定的显著性水平，当时，拒绝原假设，此时检验统计量落在拒绝域中；当时，接受原假设，此时检验统计量落在接受域中。**

1. **写出统计量的定义和构造统计量的目的，说明充分统计量的统计意义。**

**答：设为总体的样本，为样本空间中中点的实值函数，做样本的函数，*T*的取值记为。若也为一随机变量，且不含任何未知参数，则称或为统计量。**

**由于样本来自总体，样本中包含了总体的信息，我们可以通过构造不同的统计量对总体的参数进行估计，从而对总体作出推断。**

**设总体的分布函数为，为未知参数，，为样本，为不含有未知参数的统计量，若在统计量时，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. **简答区间估计的定义、步骤，说明区间估计与点估计的关系。**

**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，分别为置信上、下限。**

**区间估计的步骤如下：**

* 1. **确定统计量的分布**
  2. **确定分布的分位数**
  3. **由关系式确定区间上、下限**
  4. **写出参数的估计区间**

**区间估计包括三个要素：点估计值，估计误差，置信水平。即区间估计是建立在点估计的基础上的，估计区间一定包含点估计值。**

1. **简答最大似然估计的思想和步骤**

**答：设总体的密度函数为，其中为未知参数，参数空间是维的。为样本，它的联合密度函数为，称：**

**为的似然函数，若有使得下式成立：**

**则称为的极大似然法估计量。**

**极大似然法的原理就是选取使得样本落在观察值的邻域里的概率达到最大的数值作为参数的估计值。**

**其步骤如下：**

* + 1. **求样本的极大似然函数。**
    2. **求对数似然函数。**
    3. **对求偏导，得到似然方程。**
    4. **解似然方程，得极大似然估计。**

1. **计算题**
2. **是来自下列总体的样本，计算各分布参数的充分统计量。**

**关于如何证明某个统计量是充分统计量：**

**定义法：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。**

**因子分解定理：若总体为连续型随机变量，密度函数为，为样本，则统计量为参数的充分统计量的充要条件是：样本的联合密度函数可表示为**

**其中是仅通过而依赖于样本且与有关的非负函数，是样本的非负函数但与参数无关。特别地，可为统计量的密度函数。**

**若总体为离散型随机变量，则密度函数代之以概率函数，有类似于上式的分解式。**

**注：极大似然估计常为充分统计量，这为计算充分统计量提供了一个思路。**

**正态分布的参数的充分统计量.**

**解：已知分别为的估计量，其中**

**样本的联合密度函数为**

**取**

**由因子分解定理知为参数向量的充分统计量.**

**泊松分布的参数的充分统计量.**

**解：已知为参数的估计量**

**样本的联合密度函数为**

**取**

**由因子分解定理知为参数的充分统计量.**

**指数分布的参数的充分统计量.**

**解：已知参数的一个估计为，其中**

**样本的联合分布函数为**

**取**

**由因子分解定理知为参数的充分统计量.**

1. **样本来自两点分布总体，试给出参数的矩估计、渐进置信区间估计，并对两种估计分别给出一个实际案例。**

**关于矩估计，不再多做介绍，下面主要说明渐进区间估计。**

**在某些情况下，根据样本，我们可以很容易的求出统计量的精确分布，进而得出相应的置信区间。而精确分布的要求较高，大多数时候，我们无法求得统计量的精确分布，只能退而求其次，利用统计量的渐进分布来进行相应的统计推断，渐进置信区间就是建立在统计量的渐进分布的基础之上的置信区间的估计。**

**例如：对于来自未知分布总体的样本，对其进行均值的区间估计。由于总体分布未知，我们无法确定其统计量的抽样分布，而由中心极限定理可知，当样本量足够大时，渐进服从正态分布，此时可以由渐进分布来获得未知总体的均值的渐进区间估计。**

**解：由矩估计原理，令**

**其中**

**得到的矩法估计量为**

**由中心极限定理知**

**即**

**对进行标准化处理，有**

**由**

**而等价于**

**记**

**这里**

**于是得到的渐进置信区间为**

**举例：某小区拟实行垃圾分类投放管理条例并制定了一系列小区奖惩措施，现对小区300户居民进行民意调查，同意实施者220户，反对者80户。**

**记每户投票结果为，则居民投票通过率的矩估计为**

**的95%的渐进置信区间估计为**

**3、 现有观测数据，拟对其建立线性回归模型进行参数估计，考虑以下两种情况：**

1. **时，回归参数的最小二乘估计。**

**解：令*,*则有**

**由于，故满秩，从而满秩，即可逆，由正则方程可得**

1. **时，应如何处理。**

**解：由(1)可知，当时，是非满秩的，这将使得不再可逆，正则方程式当然也不再适用。此时若想获得参数的最小二乘估计，有以下几种做法：**

**方法一：继续观测，获得更多的观测值，以使得，将其转化为(1)的情形即可。**

**方法二：若条件不允许继续获得观测，可退一步，以损失估计的优良性为代价，以矩阵的广义逆进行计算。这样，即使不满秩，我们也可以在新的规则下求得的逆矩阵，进一步利用正则方程求解。**

**方法三：如果既不能继续获得观测，又想要保留估计优良性，那么可以考虑删减变量，将与模型相关度较低的变量从模型中剔除。若操作之后能够使得，则可以按照(1)中方法进行求解。**

**方法四：当每个变量都与模型相关度较高或无法分析变量与模型的相关性的时候，可以考虑对变量进行线性组合，组合为少数几个综合变量，这样就把问题转化为求解少数几个综合变量的系数的估计的问题。但这样的估计同样存在问题，即丧失了一些最小二乘估计原有的优良性。**

**4、 已知随机变量，参数函数.**

1. **证明参数函数不存在无偏估计.**

**证：反证法，设是参数函数的无偏估计，则应有**

**上式左端为的次多项式，由于是有界的，故必定有界。**

**而当时，上式右端趋向于无穷大，即右端是无界函数。从而上式不可能恒成立，这与假设相矛盾，故参数函数不存在无偏估计。**

1. **求的最大似然估计，并讨论其相合性.**

**解：记,则参数的似然函数为**

**对数似然方程为**

**解得：**

**有最大似然估计的不变性，知参数函数的最大似然估计为**

**下面讨论其相合性：**

**由切比雪夫不等式知，对任意的，成立下式**

**这里，显然上式可写为**

**故是参数函数的相合估计。**

**第4题版本2**

**总体,为来自总体的样本,参数函数.**

1. **证明参数函数不存在无偏估计.**

**证明：反证法，假设存在无偏估计，记为，则有**

**由数学期望的定义**

**当且仅当上述积分收敛时，的数学期望存在，若上述积分收敛，则其值唯一，这与已知**

**相矛盾，故的无偏估计不存在.**

1. **求的最大似然估计，并讨论其相合性.**

**解：由于最大似然估计似然估计具有不变性，自然想到求的最大似然估计，然后利用不变性得到的最大似然估计。**

**的似然函数为**

**由上述函数对参数求偏导数，并令其为**

**解得的最大似然估计为**

**由最大似然估计的不变性，有**

**下面讨论的相合性，由题知**

**当时，对于任意的**

**故此时作为的最大似然估计是相合的.**

**当时**

**前面我们有**

**满足马尔科夫条件，由马尔可夫大数定律，对于任意的**

**即是的相合估计，由于，从而**

**亦即**

**故此时依然是的相合估计.**