**相似真题补充**

**读者须知**

**本篇总计参阅近40所高校五年的432统计学考研真题，从中挑选出符合东北大学真题风格和类型的题目。**

**由于在专业课复习过程中，习题量较少，除却课后题与真题之外，大概在十月份就会进入题目真空期，而考试题的考点虽然大致相同，但试题形式层出不穷。整理此篇的目的之一是强化对考点的理解，使其能够运用相同的知识点应对不同的试题形式，另一目的是缓解在专业课题目空窗期的焦虑症状。**

**题目特点是，基本涵盖了往年计算题证明的所有类型，对于简答题和论述题也进行了补充，并适当进行了考点扩充。针对于往年曾考察过的冷僻考点，本篇也适当地给出了题目进行练习。**

**需要说明的是，这些题目虽然风格相似，但难度不一或相差较大，仅可作为练习使用，不可用于模拟。**

**计算证明题**

1、设是来自总体的样本，的概率密度函数为。求的极大似然估计。

解：参数的似然函数为

随数似然函数为

令

解得

2、已知正态总体，为来自总体的样本。

（1）分别求出的极大似然估计量。

（2）是否为的无偏估计。

解：（1）

（2）是的无偏估计，不是的无偏估计。

3、设是来自总体的样本，的概率密度函数为，试求的矩估计和极大似然估计。

解：矩估计；

极大似然估计。

4、总体X的概率分布如下：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| X | 0 | 1 | 2 | 3 |
| p |  |  |  |  |

其中：是未知参数，利用样本观测值3，1，3，0，3，1，2，3

求：的矩估计和极大似然估计。

解：矩估计；

极大似然估计。

5、设随机变量为来自X的样本，试求参数p的极大似然估计。

解：

6、根据世界卫生组织数据，我国居民肺癌患病率为38.46人/10万人。另外根据我国《居民营养与健康状况调查》结果，居民吸烟率为31%，而根据医学研究发现，吸烟者患肺癌的概率是不吸烟者得10.8倍。

（1）不吸烟者与吸烟者患肺癌的概率各是多少？

（2）随机抽取一名居民做检查后发现其患有肺癌，求这个居民是吸烟者的概率。

解：（1）不吸烟者患肺癌的概率；

吸烟者患肺癌的概率

（2）0.829

7、线性回归模型，，独立同分布于，记

求的极大似然估计。

解：由于，因此，于是似然函数为

对似然函数取对数得到

当

时，上式的解即为的极大似然估计。即

8、随机抽取10个家庭，调查了他们的家庭月收入X（单位：千元）和月支出Y（单位：千元），记录数据如下：

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| X | 15 | 10 | 15 | 20 | 12 | 15 | 13 | 14 | 17 | 11 |
| Y | 13 | 9 | 12 | 15 | 9 | 14 | 12 | 13 | 15 | 8 |

（1）试求X与Y的一元线性回归方程。（保留两位小数）

（2）对所得回归方程做显著性检验。（α=0.05）

（已知）

解：（1），

（2）拒绝原假设，回归方程显著。

9、假设随机变量，是取自总体的样本，在的显著性水平下，检验，及拒绝域为，求：

（1）求常数c。（）

（2）若已知，是否可依据此样本推断？（）

（3）若以作为检验的拒绝域，试求检验的显著性水平。

解：（1）

（2）不能

（3）。（）

10、某饮料制品公司生产一种瓶装含乳饮料，设这种饮料每瓶脂肪含量X（单位：g）服从正态分布，均未知。先随机抽取7瓶这种饮料化验其脂肪含量，数据如下：

如果要求这种饮料每瓶的平均 脂肪含量是5.5g，标准差不得超过0.3g。取显著性水平α=0.05，做一下假设检验：

（1）能否认为这种饮料每瓶的平均脂肪含量符合要求？

（2）能否认为这种饮料每瓶脂肪含量的标准差符合要求？

解：（1）可以认为符合要求。

（2）不符合要求。

11、糖厂用自动打包机打包，每包标准重量是100克。每天开工后需要检验一次打包机是否正常工作。某日开工后测得9包重量如下：

已知每包的重量服从正态分布，试以显著性水平α=0.05来检验打包机是否正常工作。（）

解：不拒绝原假设，打包机正常工作。

12、设总体的密度函数为：;，

为其子样。

（1）求参数的极大似然估计。

（2）证明样本均值以及

都是的无偏估计量，并说明那个更有效。

解：（1）

（2）时，

更有效；

时，二者有效性相同；

2时，样本均值更有效。

13、在一元线性回归模型下，假设，是根据次独立抽样得到的样本数据。

（1）求的最小二乘估计。

（2）如何在显著性水平下，进行假设检验。

解：详见参考书

14、设正态分布随机变量与相互独立。

（1）分别求与的分布，并说明是否独立。

（2）求概率。（用标准正态分布函数）

解：（1），，与不独立。

（2）。

15、记随机变量服从均匀分布，为来自总体的样本。假设

现有检验：“拒绝，若”。

求上述检验的第一类错误与第二类错误的概率。

解：第一类错误的概率：

第二类错误的概率：。

16、在一元线性回归模型下，假设，是根据次独立抽样得到的样本数据。

（1）求的极大似然估计。

（2）写出的95%置信区间（）。

解：（1）

（2

17、设为来自伯努利分布的样本。

求的极大似然估计，并说明该估计不是的无偏估计。

解：

18、设总体服从如下分布。为其简单随机样本。

|  |  |  |  |
| --- | --- | --- | --- |
| X | -1 | 0 | 1 |
| P |  |  |  |

令

其中为示性函数。针对假设，构建拒绝域

求次检验问题的第一类错误和第二类错误的概率。

解：

19、设总体，为来自总体的简单随机样本。

（1）求参数的极大似然估计。

（2）证明不是的无偏估计。

解：（1）。

（2），故不是的无偏估计。

20、随机变量，为来自总体的样本，试证明与均为的无偏估计。

解：略

21、随机变量，其中均未知，为来自总体的样本，求的极大似然估计。

解：

22、有一项考试，每个人优秀者得3分，合格者得2分，不合格者得1分。且根据以往经验，优秀、合格、不合格的比例分别为0.2,0.7,0.1。现有100人参加考试，用中心极限定理估计总分在180-200之间的概率。

解：

23、假设总体，是来自总体的简单随机样本，，是参数的两个点估计量。试问：，是参数的无偏估计吗？如果是，哪一个估计更有效？为什么？

解：二者均是的无偏估计，更有效，因为。

**简答论述题**

练习1：

1、简述假设检验中显著性水平的含义。

2、简述指数分布无记忆性的特点。

3、简述大数定律及其对统计推断应用的意义。

4、简述变量的集中趋势与离散趋势的测度指标。

5、简述多元回归分析中的多重共线性现象，说明其危害性和解决方法。

6、简述假设检验中的两类错误，并说明两类错误之间的关系。

7、从连续型二维随机变量的角度，论述如何判断两个随机变量是否独立，并说明独立性与相关性之间有什么样的关系。

8、论述为什么在一元线性回归分析中方程的显著性检验和回归系数的显著性检验是等价的，而在多元回归分析中二者不等价。

9、方差分析中对数据有哪些基本假定。

答案：

1：显著性水平取意味着，当原假设成立时，如果事件发生的概率（或得到当前样本观测的概率）小于，那么我们认为原假设是不成立的，从而做出拒绝原假设的决策。在实际决策中，通常取0.1，0.05，0.01这几个值。

另外，是当原假设成立时，得出拒绝原假设结论的概率，即α是犯第一类错误的概率。这个概率值是由我们事先设定的，是可控制的，并且我们把优先控制第一类错误的假设检验称为显著性检验。

2：指数分布的无记忆性可简单描述为：若某一随机变量服从指数分布，则在已知其取值大于的情况下，其取值大于的概率，等于在未知先验信息时，其取值大于的概率，即若随机变量，则对任意的，成立

3：大数定律是说，对于大量的随机变量的序列，其前个随机变量的期望的算术平均值收敛于前个随机变量的算术平均值。常用的大数定律包括辛钦大数定律、伯努利大数定律、切比雪夫大数定律。

4：集中趋势是指一组数据向某一中心值靠拢的趋势，他反映了一组数据中心点的位置所在。集中趋势的度量指标包括平均值、中位数、众数、四分位数等。

离散趋势是指数据之间的差异程度或频数分布的分散程度，它反映的是各变量值远离数据中心的趋势。离散趋势的测度指标包括极差、标准差、四分位差等。

5：多重共线性现象是指，在多元线性回归模型中，两个或两个以上自变量之间存在相关关系。

多重共线性产生的危害：

①变量之间高度相关时，可能会使回归结果混乱，甚至把分析结果引入歧途。

②可能会导致估计的参数的正负号相反。

③会使显著性检验失去意义。

④会使回归系数的失去实际意义。

解决方法：

①将一个或多个相关的自变量从模型中剔除，使保留的自变量尽可能的不相关。

②使用有偏的参数估计方法，如岭估计等。

6：第一类错误是指原假设为真，但却作出了拒绝原假设所犯的错误；

第二类错误是指原假设为不真，却作出了接受原假设所犯的错误。

对于固定的样本量，要减小第一类错误的概率，就会增大第二类错误的概率，要减小第二类错误的概率，就会增大犯第一类错的概率，增大样本量可以同时减小两类错误的概率。但样本量不可能无限增加，一般会对假设检验中的两类错误进行控制。在假设检验中，大家都在执行这样一个原则，即首先控制犯第一类错误的概率。因为原假设是什么通常是已知的、明确的，而备择假设一般是未知的、模糊的。

7：设及，分别是二维随机变量的分布函数及边缘分布函数，若对于所有的和有

，

即

，

则称随机变量是相互独立的。

对于而言，由的分布可以确定关于、关于的边缘分布。反之，由关于和关于的边缘分布一般是不能确定的联合分布的。只有当相互独立时，由两边缘分布能确定的分布。

相关系数有时候也称为线性相关系数，它是一个可以用来描述随机变量的两个分量之间的线性关系紧密程度的数字特征。当较小时，的线性相关程度较差；当时，称不相关。不相关是指之间不存在线性关系，不相关，他们还可能存在除线性关系以外的关系。而相互独立是针对的一般关系而言的，即若相互独立，则之间不存在任何关系。

相互独立，则一定不相关；

相关，则一定不独立；

不独立，则有可能线性相关，也可能线性无关；

不相关，则有可能相互独立，也可能不相互独立。

特别地，对于二维正态随机变量不相关与相互独立是等价的。

8：在一元线性回归分析中，方程的显著性检验与回归系数的显著性检验是等价的。这是因为在方程的显著性检验中，目的是检验自变量与因变量的线性关系是否显著，其原假设与备择假设的形式为：

在回归系数的显著性检验中，目的是为了检验该自变量对因变量的影响是否显著，在一元回归分析中，只存在一个自变量，故其假设的形式也为：

可以发现二者的假设形式是一样的，若其中一个检验显著，那么另一个检验也一定是显著的，若其中一个检验不显著，则另一个检验也一定不显著。因此，其二者的检验时等价的。

在多元回归分析的线性关系的显著性检验中，只要有一个自变量是显著的，那么检验就可以通过，只有当所有变量都不显著时，检验才不通过，其假设的形式为：

在其回归系数的显著性检验中，若一个变量的检验结果是该回归系数不显著，那么就可以剔除该变量，若其检验结果是显著的，就保留该变量，其对某个系数的假设的形式为：

故而在多元回归分析中，线性关系的显著性检验与回归系数的显著性检验是不等价的。

9：方差分析是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。

练习2：

1、说明参数估计中矩估计法与极大似然估计法的基本思想。

2、简述中心极限定理的含义。

3、简述平均数、中位数、众数之间的关系。

4、简述评价估计量好坏的标准。

5、简述一元回归分析中的基本假定。

6：简述时间序列的构成因素及常用预测方法。

7、简述值的含义，并说明如何使用值作出决策。

8、简述充分统计量的定义。

9、简述顺序统计量的定义。

答案：

1：矩估计法的基本思想：一般来说，样本矩依概率收敛于相应的总体矩，样本矩的连续函数依概率收敛于总体矩的连续函数，我们用样本矩作为相应的总体矩的估计量，而以样本矩的连续函数作为相应的总体矩的连续函数的估计量，这种估计方法称为矩估计法，其理论基础是大数定理。

极大似然估计法的基本思想：固定样本观察值，在取值的可能范围内挑选使似然函数达到最大的参数值，作为参数的估计值，这种估计方法称为极大似然估计法。

极大似然估计还有一个性质：设的函数，具有单值反函数，。又假设是的概率分布中参数的最大似然估计，则是的极大似然估计，这一性质称为极大似然估计的不变性。

2：中心极限定理是确定在什么条件下，大量随机变量之和的分布逼近于正态分布。

中心极限定理表明，在相当一般的条件下，当独立随机变量的个数不断增加时，其和的分布趋于正态分布。这一事实阐明了正态分布的重要性。也揭示了为什么在实际应用中会经常遇到正态分布，也就是揭示了产生正态分布变量的源泉。另一方面，它提供了独立同分布的随机变量之和的近似分布，只要和式中加项的个数充分大，就可以不必考虑和式中的随机变量服从什么分布，都可以用正态分布来近似。

3：平均数也称均值，它是一组数据相加后除以数据个数得到的结果。平均数是集中趋势的最主要测度值，它主要适用于数值型数据，而不适用于分类数据和顺序数据。平均数利用了全部数据信息。

中位数是一组数据排序后处于中间位置上的变量值，用表示。显然，中位数将全部数据等分成两部分，每部分包含50％的数据，一部分数据比中位数大，另一部分则比中位数小。中位数主要用于测度顺序数据的集中趋势，当然也适用于测度数值型数据的集中趋势，但不适用于分类数据。中位数是一个位置代表值，其特点是不受极端值影响。

众数是一组数据中出现次数最多的变量值。众数主要用于测度分类数据的集中趋势，也可以作为顺序数据以及数值型数据集中趋势的测度值。一般情况下，只有在数据量较大的情况下众数才有意义。

众数是一个位置代表值，它不受数据中极端值的影响。从分布的角度看，众数是具有明显集中趋势点的数值，一组数据分布的最高峰点所对应的数值即为众数。当然，如果数据的分布没有明显的集中趋势或最高峰点，众数可能不存在；如果有两个或多个最高峰点，则可以有两个或多个众数。

三者之间的关系：

①如果数据的分布是对称的，众数、中位数、平均数必定相等；

②如果数据是左偏分布，说明数据存在极小值，必然拉动数据向极小值一方靠，而众数和中位数由于是位置代表值，不受极值的影响，因此三者的关系表现为:平均数中位数众数；

③如果数据是右偏分布，说明数据存在极大值，必然拉动数据向极大值一方靠，因此众数中位数平均数。

4：无偏性是指估计量抽样分布的数学期望等于被估计的总体参数。总体参数为，所选择的估计量为，如果，则称为的无偏估计量。

有效性是指对同一总体参数的两个无偏估计量，有更小标准差的估计量更有效。假定分别是总体参数的两个无偏估计量，他们的抽样分布的方差用表示，如果，则称比更有效。

一致性是指随着样本量的增大，估计量的值越来越接近被估计总体的参数。一般而言，一个好的估计量应该具有一致性。

5：①因变量与自变量之间具有线性关系；

②在重复抽样中，自变量的取值是固定的，即假定是非随机的；

③误差项是一个期望值为的随机变量，即；

④对于所有的值，的方差都相同，这意味着对于一个特定的值，的方差都等于。即。

⑤误差项是一个服从正态分布的随机变量，且相互独立。即

6：时间序列的组成要素:

①趋势:时间序列在长期内呈现出来的某种持续上升或持续下降的变动，也称长期趋势；

②季节性(季节变动):时间序列在一年内重复出现的周期性波动；

③周期性(循环波动):时间序列中呈现出来的围绕长期趋势的一种波浪形或震荡式变动。循环波动无固定规律，变动周期多在一年以上，且周期长短不一。

④随机性(不规则波动):时间序列中除去趋势、周期性和季节性之后的偶然性波动。

时间序列预测方法:

①若时间序列不存在趋势且不存在季节性，则可以用平滑预测法，平滑预测法包括:简单平均法、移动平均法、指数平滑法；

②若时间序列不存在趋势但存在季节性，或时间序列存在趋势且存在季节性，则可以使用季节性预测法，季节性预测法包括:季节多元回归模型、季节自回归模型、时间序列分解；

③若时间序列存在趋势但不存在季节性，则可以使用趋势预测法，趋势预测法包括:线性趋势推测、非线性趋势推测、自回归预测模型。

7：假设检验问题的值是由检验统计量的样本观测值得出的原假设可被拒绝的最小显著性水平。

值决定因素:

①样本数据与原假设之间的差异；

②样本量；

③被假设参数的总体分布。

对于任意给定的显著性水平:

① 若值，则在显著性水平下拒绝原假设；

② 若值，则在显著性水平下接受原假设。

值表示反对原假设的依据的强度，值越小，反对原假设的依据越强、越充分。

一般若值，称推断拒绝原假设的依据很强或称检验是高度显著的；若值，称推断拒绝原假设的依据是强的或称检验是显著的；若值，称推断拒绝原假设的理由是弱的或称检验是不显著的；若值，一般来说没有理由拒绝原假设。

8：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。

9：*设*为总体的样本，现由样本建立个函数:

其中为这样的统计量，它的观察值为，为样本的观察值中由小到大排列之后的第位数值，则称为顺序统计量。

练习3：

1、简述判断一组数据中的异常值的方法。

2、简述分布的定义，并说明当其自由度趋向于无穷大时的极限分布。

3、在显著性检验中，原假设与备择假设的地位有何不同，在设置假设时应该准守哪些原则。

4、简述箱线图的绘制方法及其主要用途。

5、简述单因素方差分析的基本假设和一般步骤。

6、简述小概率事件原理，并说明其在统计学中的应用。

7、简述概率分布函数的特质。

8、简述什么是区间估计？如何理解置信区间与参数真值的关系？

9、试论述两个事件的独立性与两个随机变量独立性的关系。

答案：

1：①对于低维数据，我们可以通过绘制数据分布的盒子图或散点图。如果一个数据点的分布在散点图中是离群的，那么这个数据点就是异常点；而在盒子图中，记中位数为，四分位差为，则分布在区间之外的点就是异常点.

②对于小批量的一维数据，可以对数据进行排序，通过观察极大值（极小值）与数据中心的距离，若这个距离明显异常，那么极大值（极小值）就是异常点。

③“法则”：记一组数据的均值为标准差为，则若一个数据分布在区间

之外，那么我们认为这个数据是异常点。

2：若随机变量，随机变量，且相互独立，则

随着的增大，分布的图像逐渐与正态分布相重合，即t分布在自由区趋向于无穷时是以正态分布为极限分布的。

3：假设检验的目的在于搜集证据拒绝原假设，从而支持备择假设。因为假设检验只提供了不利于原假设的证据，因此当拒绝原假设时，可以说我们有充分的证据拒绝原假设；而当不能拒绝原假设时，只能说没有充分的证据拒绝原假设，而不能说接受原假设，因为我们没有提供证据能够证明备择假设成立。由于假设检验是围绕原假设是否成立而展开的，备择假设是作为原假设不成立时的替换，因此二者并不是非此即彼的关系。

在假设检验中，通常把我们想要证明错误的结论放在原假设上，把想要接受的结论放在备择假设上，并且等号一定在原假设中。

4：数据集的箱线图是由箱子和直线组成的图形，它是基于以下五个数的图形概括:最小值，第一四分位数，中位数，第三四分位数和最大值。它的作法如下:

① 画一水平数轴，在轴上标上，，，，。在数轴上方画一个上、下侧平行于数轴的矩形箱子，箱子的左右两侧分别位于，的上方。在点的上方画一条垂直线段。线段位于箱子内部。

② 自箱子左侧引一条水平线直至最小值；在同一水平高度自箱子右侧引一条水平线直至最大值。

它主要用于反映原始数据分布的特征，还可以进行多组数据分布特征的比较。此外，箱线图还可用于异常点的检验。

5：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。

一般步骤如下:

① 计算各水平的样本均值；

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；

⑤ 做出决策。

6：小概率事件原理是说，概率很小的事件在一次实验中几乎是不可能发生的。换句话说，如果在一次实验中，我们所认识的小概率事件发生了，那么我们有理由怀疑我们认识的正确性。

小概率事件在统计学中的通常应用于假设检验，它是假设检验的概率基础，即若在原假设成立的情况下，如果取到当前样本的概率足够小，那么原假设就应当被拒绝。

7：一维随机变量的分布函数的定义:设是一个随机变量，是任意实数，函数

称为的分布函数。

基本性质:

①是一个不减的函数；

②，且；

③ ，即是右连续的。

二维随机变量分布函数的定义:设是二维随机变量，对于任意实数，二元函数:

称为二维随机变量的分布函数，或称为随机变量和的联合分布函数。

基本性质:

①是变量和的不减函数；

②，且

对于任意固定的，

对于任意固定的，

；

③，即关于右连续，关于右连续；

④ 对于任意，下述不等式成立:

。

8：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。

二者的关系：

①的含义是，在当前显著性水平下，若重复抽取样本100次，那么由这100组样本构造的置信区间中，约有95%的置信区间包含着总体参数的真值。

②在构造置信区间时，总体参数真值是固定的，而估计区间是随机的，即估计区间是个随机区间。我们可以说估计区间以的概率包含着总体真值，但不能说总体真值有的概率存在于该估计区间中。

③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。

9：两事件相互独立：记为两随机事件，若相互独立，则有

两随机变量独立：若为两个随机变量，为其分布函数，为其概率密度，若成立

或

则称随机变量是相互独立的。

实际上随机变量之间的独立性是随机事件独立性的扩充，如果我们取事件A：“”，事件B：“”则此时随机变量的独立与随机事件的独立是等价的。

练习4：

1、试写出概率的公理化定义。

2、简述泊松分布及其性质。

3、试简述假设检验的值的含义以及用值进行假设检验相对于传统方法的优点，并简述用值做假设检验的步骤。

4、简述单侧检验和双侧检验的区别。

5、简述参数估计中常用的点估计方法。

6、简述为什么对多元线性回归分析中的多重判定系数进行调整。

7、在高斯-马尔科夫假定时，线性回归模型的最小二乘估计有哪些性质？

8、简述置信区间与假设检验的联系。

9、论述方差分析和两独立样本的检验的异同。

答案：

1：设是随机试验，是它的样本空间。对于的每一事件，赋予一个实数，记为，称为事件的概率，则集合函数满足：

①（非负性）对任一事件，有。

②（规范性）对于必然事件，有。

③（可列可加性）若是两两互不相容的事件，则

2：泊松分布是用来描述在一指定时间范围内或在指定的面积或体积之内某一事件出现的次数的分布。其定义如下

若某一随机变量的概率分布为

则称是服从参数为的泊松分布的。

性质：

①可加性：若随机变量分别服从参数为的泊松分布，且二者相互独立，则随机变量服从参数为的泊松分布。

②正态近似：当样本量足够大时，泊松分布的近似分布为正态分布。

③泊松定理：若随机变量，且足够大，非常小，记，则有

3：假设检验问题的值是由检验统计量的样本观测值得出的原假设可被拒绝的最小显著性水平。

显著性水平与值的比较:

确定以后，拒绝域的位置也就相应确定了，其好处是进行决策的界限清晰，但缺陷是进行决策面临的风险是笼统的。

值的长处是它反映了观察到的实际数据与原假设之间不一致的概率值，与传统的拒绝域范围相比，是一个具体的值，这样就提供了更多的信息。

用值做假设检验的步骤：

①确定原假设与备择假设。

②选择合适的检验统计量。

③计算该假设检验问题的值。

④根据所选择显著性水平作出决策。

4：双侧检验的原假设的形式为，他有两个拒绝域，两个临界值每个拒绝域的面积均为，故在双侧检验中，统计量落在任何一个拒绝域中，都可以拒绝原假设。

单侧检验的原假设形式为或，它只有一个拒绝域和临界值，拒绝域的面积为；在单侧检验中，左单侧检验的拒绝域在左侧，右单侧检验的拒绝域在右侧。

5：点估计方法就是用统计量的某一个值作为总体参数的真值，常用参数点估计方法包括矩法、极大似然法和最小二乘法。

矩估计法:一般来说，样本矩依概率收敛于相应的总体矩，样本矩的连续函数依概率收敛于总体矩的连续函数，我们用样本矩作为相应的总体矩的估计量，而以样本矩的连续函数作为相应的总体矩的连续函数的估计量，这种估计方法称为矩估计法，其理论基础是大数定理。

极大似然估计法:固定样本观察值，在取值的可能范围内挑选使似然函数达到最大的参数值，作为参数的估计值，这种估计方法称为极大似然估计法。

最小二乘法：使因变量的观测值与预测值之间的离差平方和达到最小来估计参数。

6：判定系数测度了回归直线对观测数据的拟合程度。的取值范围是，越接近于1，表明回归平方和占总平方和的比例越大，回归直线与各观测点越接近，用的变化来解释值变差的部分就越多，回归直线的拟合程度就越好；反之，越接近于0，回归直线的拟合程度就越差。

他的实际意义是：在因变量的变差中有可以由自变量与因变量之间的线性关系解释。

由于在回归模型中增加自变量的数量，总是会导致增大，但过多的自变量不仅会带来沉重的计算负担，还会造成过拟合现象。为了这类情况的出现，统计学家提出调整的系数，通过增添自变量数目的惩罚因子来控制的增加，从而更加合理地选择自变量的数目。

7：①无偏性：最小二乘估计具有无偏性，即对于回归系数的最小二乘估计量满足。

②由最小二乘估计量求得的估计值与总体真值的离差平方和最小。

③线性最优：最小二乘估计是回归系数的所有线性估计中方差最小的，这一点已经由马尔科夫定理保证。

④由最小二乘估计得到的估计向量是观察向量在估计空间中的垂直投影。

8、参数估计和假设检验是统计推断中两个非常重要内容，参数估计是通过样本统计量估计总体参数真值，而假设检验是通过样本统计量来检验总体参数的先验假设是否成立。

区别：①区间估计求得的是一个以点估计值为基准的对称双侧置信区间或单侧置信区间，而假设检验以假设总体参数值为基准，不仅有双侧检验也有单侧检验。

②区间估计立足于大概率事件，通常以较大的概率（置信水平）去把握。与此相反，假设检验立足于小概率事件，通常是给定很小的显著性水平去检验对总体参数的先验假设是否成立。

③区间估计中的参数是未知的、待估计的，而假设检验中的参数值是已知的、待检验的。

联系：①假设检验与区间估计都是根据样本信息对总体参数进行推断，都是以抽样分布为理论依据，都是建立在概率基础之上的推断，推断结果都有可信度并且带有一定的风险。

②对同一问题的参数进行的推断，二者使用同一样本，同一统计量，同一分布，因此二者可以相互转换。区间估计的置信区间对应于假设检验的接受域，置信区间之外的部分对应于假设检验的拒绝域。

9：方差分析：它是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。

独立样本的检验：它是通过检验在某一因素的作用前后的两个样本的均值是否相等，来判断该因素对总体是否产生了显著性影响。

相同点：

①方差分析和独立样本的检验都是研究因素对总体的影响是否显著。

②在基本假设中，方差分析和检验均要求各观测值相互独立且服从正态分布。

③当使用单因素方差分析时涉及两个水平，检验涉及两个样本，且均要求相互独立。

不同点：

①方差分析是通过研究方差来推断均值是否相等，而检验时通过研究样本差值是否为零来判断均值是否相等。

②方差分析所使用的检验统计量为统计量，检验形式为单侧检验，检验所使用的检验统计量为统计量，检验形式为双侧检验。

③方差分析可以针对两个及以上的因素进行分析，而检验只能针对两个样本进行检验。