**成都信息工程大学计算机学院**

**数据统计与分析实验报告**

“正态总体的参数估计和假设检验”实验

实验二 “正态总体的参数估计和假设检验”实验

**一、实验目的**

1. 掌握R语言进行正态分布均值的区间估计方法。

2. 掌握R语言进行正态分布方差的区间估计方法。

3. 掌握R语言进行单个总体的正态分布均值的假设检验方法。

4. 掌握R语言进行单个总体的正态总体方差的假设检验方法。

**二、实验原理**

1. 正态分布均值的区间估计

现实中，总体标准差未知，故可使用统计量，其中为样本的标准差，用它来代替总体标准差。注意：

(1) 统计量服从正态分布且较小，则统计量服从于自由度为的分布；

(2) 如果足够大，便可应用中心极限定理，则统计量将是渐进正态的。

当数据服从正态分布（或渐进正态）时，可以运用t分布来构造置信区间。

当未知时，由于，有。用表示自由度为的*t*分布的上分位点。于是均值的置信区间为：。

2. 正态分布方差的区间估计

取统计量，对于给定的分位点，由分布统计量查的两个分位点和，使得，于是方差的置信区间为。

3. 假设检验

假设检验的具体步骤如下：

(1) 建立假设，包括原假设与备择假设；

(2) 寻找检验统计量*T*，确定拒绝域的形式；

(3) 给出显著性水平；

(4) 给出临界值，确定拒绝域；

(5) 根据样本观察值计算检验统计量的值，根据观察值是否落在拒绝域中做出判断？

3.1 单个总体的正态分布均值的假设检验方法

单样本均值的统计量服从t分布，故采用t检验。

(1) 检验假设：，；()

(2) 给定检验水平；

(3) 计算检验统计量 ；

(4) 计算t值对应的p值；

(5) 若，拒绝，接受；若，接受，拒绝。通常取 。

在R中，单个总体的正态总体均值的假设检验可以通过t.test()函数实现。

3.2 单个总体的正态分布方差的假设检验方法

单样本均值的统计量服从卡方分布，故采用卡方检验。

(1) 检验假设：，；()

(2) 给定检验水平；

(3) 计算检验统计量；

(4) 计算值对应的p值；

(5) 若，拒绝，接受；若，接受，拒绝。通常取。

在R中，正态总体方差的假设检验可以通过var.test()函数实现。

**三、实验内容**

1. 正态分布均值的区间估计

用R语言实现教材129页习题2的置信区间计算。

解答：

|  |  |
| --- | --- |
| 统计量 | 对均值进行区间估计，但总体方差未知，故选择作为统计量。 |
| 代码 | x = c(7.2, 3.5, 4.3, 6.2, 10.1, 5.4, 6.8, 4.5, 5.1, 6.6, 3.8, 8.2)  c(mean(x)-qt(0.975,df=11)\*sd(x)/sqrt(12), mean(x)+qt(0.975,df=11)\*sd(x)/sqrt(12)) |
| 运行结果 | [1] 4.740555 7.209445 |
| 置信区间 | 均值μ的95%置信区间为 [4.7406，7.2094]。 |

2. 正态分布方差的区间估计

某一预制厂在生产过程中定期检查搅拌罐的搅拌质量，现对12块混凝土试块进行强度检验，结果如下：(单位kgf/cm2)

212 204 193 189 198 217

208 211 190 199 220 207

试求总体均值及总体标准差的置信水平为0.95的置信区间?

解答：求总体均值的置信水平为0.95的置信区间。

|  |  |
| --- | --- |
| 统计量 | 对均值进行区间估计，但总体方差未知，故选择作为统计量。 |
| 代码 | x = c(212,204,193,189,198,217,208,211,190,199,220,207)  c(mean(x)-qt(0.975,df=11)\*sd(x)/sqrt(12), mean(x)+qt(0.975,df=11)\*sd(x)/sqrt(12)) |
| 运行结果 | [1] 197.4585 210.5415 |
| 置信区间 | 均值μ的95%置信区间为 [197.4585，210.5415]。 |

解答：求总体标准差的置信水平为0.95的置信区间。

|  |  |
| --- | --- |
| 统计量 | 对方差进行区间估计，由于总体均值未知，故选择作为统计量。 |
| 代码 | x = c(212,204,193,189,198,217,208,211,190,199,220,207)  *c(sqrt(11\*sd(x)^2/21.920), sqrt(11\*sd(x)^2/3.816))* |
| 运行结果 | [1] 7.293383 17.480147 |
| 置信区间 | 总体**标准差**的95%置信区间为 [7.2933，17.4801]。 |

3. 正态分布均值的假设检验

教材129页习题4。

答案：新工艺是否提高了平均断裂强度？

|  |  |
| --- | --- |
| 统计量 | 由于是对总体均值进行假设检验，且总体方差未知，故选择为统计量。 |
| 原假设 | (学生补充) |
| 备择假设 | (学生补充) |
| 置信度 | 是\_\_右\_\_侧检验，置信度为0.05。 |
| 代码 | x = c(502,496,510,508,506,498,512,497,515,503,510,506)  t.test(x, mu=500,alternative = 'greater') |
| 运行结果 | One Sample t-test  data: x  t = 2.9564, df = 11, p-value =  0.006529  alternative hypothesis: true mean is greater than 500  95 percent confidence interval:  502.0609 Inf  sample estimates:  mean of x  505.25 |
| 是否接收H0 | 否 |

4. 正态分布均值的假设检验（P129第3题）

根据过去大量资料，某厂生产的灯泡的使用寿命服从正态分布N~(1020，1002)。现从最近生产的一批产品中随机抽取16只，测得样本平均寿命为1080小时。试在0.05的显著性水平下判断这批产品的使用寿命是否有显著提高？

解答：在0.05的显著性水平下判断这批产品的使用寿命是否有显著提高？

|  |  |
| --- | --- |
| 统计量 | 对总体均值进行假设检验，且总体方差已知，故选择统计量。 |
| 原假设 | (学生补充) |
| 备择假设 | (学生补充) |
| 置信度 | 是\_\_\_\_左\_\_\_\_侧检验，置信度为0.05。 |
| 代码 | (1080-1020)/(100/sqrt(16)) |
| 运行结果 | [1] 2.4 |
| 是否接收H0 | 否 |

5. 正态分布方差的假设检验

某厂商生产出一种新型的饮料装瓶机器，按设计要求，该机器装一瓶一升(1000cm3)的饮料误差为1cm3。如果达到设计要求，表明机器的稳定性非常好。现从该机器装完的产品中随机抽取25瓶，分别进行测定(用样本减1000cm3)，得到如下结果。检验该机器的性能是否达到设计要求 (α=0.05)？

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 0.3 | -0.4 | -0.7 | 1.4 | -0.6 |
| -0.3 | -1.5 | 0.6 | -0.9 | 1.3 |
| -1.3 | 0.7 | 1 | -0.5 | 0 |
| -0.6 | 0.7 | -1.5 | -0.2 | -1.9 |
| -0.5 | 1 | -0.2 | -0.6 | 1.1 |

解答：检验该机器的性能是否达到设计要求 (α=0.05)？

|  |  |
| --- | --- |
| 统计量 | 对总体方差进行假设检验，选择作为统计量。 |
| 原假设 |  |
| 备择假设 |  |
| 置信度 | 显著性水平为0.05，为双边检验。 |
| 代码 | x=c(0.3,-0.4,-0.7,1.4,-0.6,-0.3,-1.5,0.6,-0.9,1.3,-1.3,0.7,1,-0.5,0,-0.6,0.7,-1.5,-0.2,-1.9,-0.5,1,-0.2,-0.6,1.1)  (25-1)\*sd(x)^2/1 |
| 运行结果 | [1] 20.7816 |
| 是否接收H0 | 接收 |

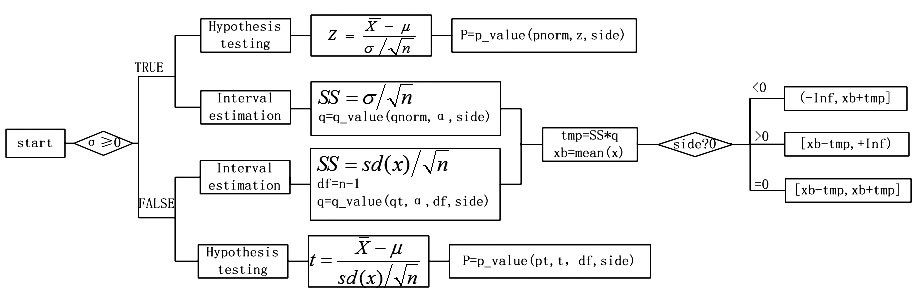
**四、实验组织运行要求**

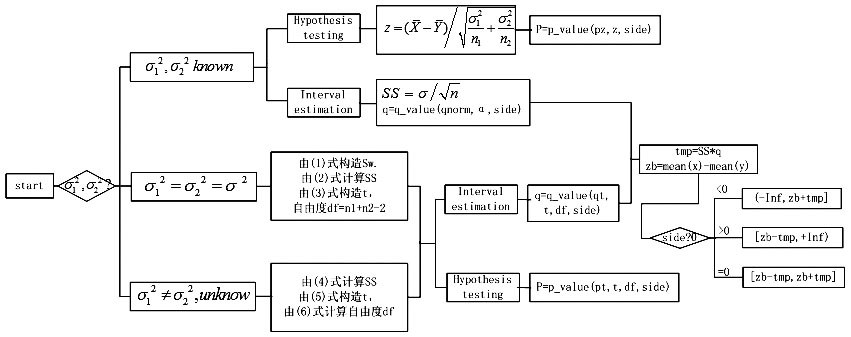
以学生自主训练为主的开放模式组织教学。

**五、思考题**

如何进行正态总体均值和方差的单侧假设检验，并用R语言实现？

**答：**





**六、实验总结**

R语言是一个极其全面的统计研究和可进行交互式数据分析和探索的强大平台，拥有顶尖水准的制图工具。这一点后起之秀的Python显得比较单薄，特别是对于可视化这一部分，虽然Python有一些不错的可视化库，如Seaborn，Bokeh和Pygal。但相比于R，呈现的结果并不总是那么顺眼。所以靠着统计学吃饭的同学，R是必须掌握语言，Python作为爱好比较合适。类似MALTAB的工具箱，Python的库，R也有自己的R package。这些包可以轻松地从各种类型的数据源导入数据，包括文本文件、数据库管理系统、统计软件，乃至专门的数据库。在后面会有较为介绍。在R语言中，R package起到了至关重要的作用：R package就是R函数、数据、预编译代码以一种定义完善的格式组成的集合。可以说这个是最让人开心的一部分了，因为正是由于包的存在，使得编译人员可以通过可选模块的下载和安装来实现很多功能。