**云南大学数学与统计学院**

**实验报告**

**实验课名称： 应用多元统计分析实验**

**指导教师： 李会琼**

**专业（年级）： 统计学2021级**

**学生姓名： 枫叶 学号:**

**实验名称： 多元正态分布均值向量的检验**

**实验成绩：**

**《应用多元统计分析实验》实验报告 4**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 实验名称 | 多元正态分布均值向量的检验 | | 实验成绩 |  | |
| 学号 |  | | 姓名 | 枫叶 | |
| 实验时间 | 2024年4月19日 | 实验地点 | 格物楼3508 | 指导教师 | 李会琼 |
| 1. **实验目的**   学习使用R软件对多元正态分布均值向量进行检验。   1. **实验要求**   1. 对所使用的方法与所得到的结果进行适当的文字描述。  2. 在实验结果的相应部分附上完整的代码与适当的注释。  3. 采用一定的可视化方法体现出对应计算结果。   1. **实验内容** 2. 某小麦良种的四个主要经济性状的理论值为。现在从外地引入一新品种，在21个小区种植，取得如表所示数据。设新品种的四个性状，在显著性水平下，试检验假设  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | | 小区号  性状 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | |  | 22.88 | 22.74 | 22.60 | 22.93 | 22.74 | 22.53 | 22.67 | |  | 32.81 | 32.56 | 32.76 | 32.95 | 32.74 | 32.53 | 32.58 | |  | 51.51 | 51.49 | 51.50 | 51.17 | 51.45 | 51.36 | 51.44 | |  | 61.53 | 61.39 | 61.22 | 60.91 | 61.56 | 61.22 | 61.30 |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | | 小区号 性状 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | |  | 22.74 | 22.62 | 22.67 | 22.82 | 22.67 | 22.81 | 22.67 | |  | 32.67 | 32.57 | 32.67 | 32.80 | 32.67 | 32.67 | 32.67 | |  | 51.44 | 51.23 | 51.64 | 51.32 | 51.21 | 51.43 | 51.43 | |  | 60.30 | 61.39 | 61.50 | 60.97 | 61.49 | 61.15 | 61.15 | | 小区号 性状 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | |  | 22.81 | 23.02 | 23.02 | 23.15 | 22.88 | 23.16 | 23.13 | |  | 33.02 | 33.05 | 32.95 | 33.15 | 33.06 | 32.78 | 32.95 | |  | 51.70 | 51.48 | 51.55 | 51.58 | 51.45 | 51.48 | 31.38 | |  | 61.49 | 61.44 | 61.62 | 61.65 | 61.54 | 61.41 | 61.58 |  1. 鸢尾花(iris)是R软件自带的一个常用数据集，鸢尾花分为三个品种（Species: setosa, versicolour, virginica)，从每个品种选取50株作为样本，共记录了150株鸢尾花的信息。花的特征用4种属性描述：萼片长度(Sepal.Length)、萼片宽度(Sepal.Width)、花瓣长度(Petal.Length)、花瓣宽度(Petal.Width)。请用R语言编程，通过实验结果说明：setosa和versicolour两个品种的鸢尾花在形态上的差异是否显著（显著性水平）？ 2. **实验软件**   R语言   1. **实验结果**  加载包 library(readxl) library(dplyr) 第一题 #单个总体均值向量检验(方差未知)  mu.test.unknown=function(data, mu0, alpha = 0.05) {   data=as.matrix(data) #将数据框转化为矩阵  n=nrow(data) #n 行  p=ncol(data) #p 列  X.bar=apply(data, 2, mean) #按列求均值  A=(n-1)\*var(data)#求离差阵  T0=(n-1)\*n\*t(X.bar-mu0)%\*%solve(A)%\*%(X.bar-mu0)#构造统计量  F\_stat=(n-p)/((n-1)\*p)\*T0#构造统计量  a2 = qf(1-alpha,p,n-p)#求下侧分位点  p.two=1-pf(F\_stat, p, n-p)#求 p 值  return(list(统计量=as.vector(F\_stat), 临界值=a2, p.value=as.vector(p.two)))  } data <- read\_xlsx("D:/预删除文件夹/大三下/多元统计/实验4数据.xlsx") mu\_0 <- c(22.75,32.75,51.50,61.50) mu.test.unknown(data,mu\_0)  ## $统计量 ## [1] 3.029811 ##  ## $临界值 ## [1] 2.964708 ##  ## $p.value ## [1] 0.04678598  由p值可知在0.05的显著性水平下应该拒绝原假设，即四个形状的均值向量与理论值不同 第二题 #两总体协方差阵的检验 multiple.var.test=function(data1,data2,alpha=0.05){  nn=matrix(NA,nrow=1,ncol=2)#储存两个总体的行数  nn[1,1]=nrow(data1)#第一个总体有n1行  nn[1,2]=nrow(data2)#第二个总体有n2行  n=sum(nn)#两个总体的总行数  p=ncol(data1)#p列  AA=array(NA,c(p,p,2))#储存两个总体的离差阵  AA[,,1]=(nn[1,1]-1)\*var(data1)#第一个总体的离差阵  AA[,,2]=(nn[1,2]-1)\*var(data2)#第二个总体的离差阵  A=AA[,,1]+AA[,,2]#两个总体的离差阵之和  S=1#lambda的分母  for(i in 1:2){S=S\*(det(AA[,,i]/(nn[1,i]-1))^(-(nn[1,i]-1)/2))}  lambda=((det(A/(n-2)))^(-(n-2)/2))/S#统计量  M=-2\*log(lambda)#统计量  b=apply(1/(nn-1),1,sum)  if(nn[1,1]==nn[1,2]){d=((2\*p^2+3\*p-1)\*3)/(6\*(p+1)\*(n-2))}  else{d=((2\*p^2+3\*p-1)/6\*(p+1)\*(1))\*(b-(1/(n-2)))}  T0=(1-d)\*M#检验统计量  f=(1/2)\*p\*(p+1)\*(1)#自由度  a2=qchisq(1-alpha,f)#求下侧分位点  p.value=1-pchisq(T0,f)#求p值  return(list(统计量=as.vector(T0),临界值=a2,p.value=as.vector(p.value))) } data(iris) setosa <- filter(iris,Species=="setosa") %>%  select(-5) versicolor <- filter(iris,Species=="versicolor") %>%  select(-5) multiple.var.test(setosa,versicolor)  ## $统计量 ## [1] 66.81048 ##  ## $临界值 ## [1] 18.30704 ##  ## $p.value ## [1] 1.823304e-10  #视作成对样本检验 mu.test.unknown(setosa-versicolor,mu0 = 0)  ## $统计量 ## [1] 538.2213 ##  ## $临界值 ## [1] 2.574035 ##  ## $p.value ## [1] 0  先做两正态总体协方差阵是否相等的检验，结果显示在0.05的显著性水平下拒绝原假设，即两者协方差阵不相等，进一步基于协方差阵不等的假设来检验形态均值向量是否相等，作为著名的Behren-Fisher问题，其尚无较好的检验方法，这里参考教材上的做法，将其视为成对样本做检验，结果显示二者差异非常显著 | | | | | |