**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： 多元正态分布协方差阵的检验**

**实验成绩：**

**《应用多元统计分析实验》实验报告 5**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 实验名称 | 多元正态分布协方差阵的检验 | | 实验成绩 |  | |
| 学号 |  | | 姓名 | 枫叶 | |
| 实验时间 | 2024年4月26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协方差阵进行检验。   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**   为了研究某种疾病，对一批人同时测量了4个指标：脂蛋白()，甘油三酯()，脂蛋白()，前脂蛋白()。按不同年龄、不同性别分为三组(20至35岁的女性、20至25岁的男性和35至50岁的男性)，数据见下表。   1. 这三个组的4项指标间有无显著性差异()? 2. 这三个组的协方差矩阵是否相等()？  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | 身体指标化验数据 | | | | | | | | | | | | | | | | |  |  |  |  | 组 |  |  |  |  | 组 |  |  |  |  | 组 | | 260 | 75 | 40 | 18 | 1 | 310 | 122 | 30 | 21 | 2 | 320 | 64 | 39 | 17 | 3 | | 200 | 72 | 34 | 17 | 1 | 310 | 60 | 35 | 18 | 2 | 260 | 59 | 37 | 11 | 3 | | 240 | 87 | 45 | 18 | 1 | 190 | 40 | 27 | 15 | 2 | 360 | 88 | 28 | 26 | 3 | | 170 | 65 | 39 | 17 | 1 | 225 | 65 | 34 | 16 | 2 | 295 | 100 | 36 | 12 | 3 | | 270 | 110 | 39 | 24 | 1 | 170 | 65 | 37 | 16 | 2 | 270 | 65 | 32 | 21 | 3 | | 205 | 130 | 34 | 23 | 1 | 210 | 82 | 31 | 17 | 2 | 380 | 114 | 36 | 21 | 3 | | 190 | 69 | 27 | 15 | 1 | 280 | 67 | 37 | 18 | 2 | 240 | 55 | 42 | 10 | 3 | | 200 | 46 | 45 | 15 | 1 | 210 | 38 | 36 | 17 | 2 | 260 | 55 | 34 | 20 | 3 | | 250 | 117 | 21 | 20 | 1 | 280 | 65 | 30 | 23 | 2 | 260 | 110 | 29 | 20 | 3 | | 200 | 107 | 28 | 20 | 1 | 200 | 76 | 40 | 17 | 2 | 295 | 73 | 33 | 21 | 3 | | 225 | 130 | 36 | 11 | 1 | 200 | 76 | 39 | 20 | 2 | 240 | 114 | 38 | 18 | 3 | | 210 | 125 | 26 | 17 | 1 | 280 | 94 | 26 | 11 | 2 | 310 | 103 | 32 | 18 | 3 | | 170 | 64 | 31 | 14 | 1 | 190 | 60 | 33 | 17 | 2 | 330 | 112 | 21 | 11 | 3 | | 270 | 76 | 33 | 13 | 1 | 295 | 55 | 30 | 16 | 2 | 345 | 127 | 24 | 20 | 3 | | 190 | 60 | 34 | 16 | 1 | 270 | 125 | 24 | 21 | 2 | 250 | 62 | 22 | 16 | 3 | | 280 | 81 | 20 | 18 | 1 | 280 | 120 | 32 | 18 | 2 | 260 | 59 | 21 | 19 | 3 | | 310 | 119 | 25 | 15 | 1 | 240 | 62 | 32 | 20 | 2 | 225 | 100 | 34 | 30 | 3 | | 270 | 57 | 31 | 8 | 1 | 280 | 69 | 29 | 20 | 2 | 345 | 120 | 36 | 18 | 3 | | 250 | 67 | 31 | 14 | 1 | 370 | 70 | 30 | 20 | 2 | 360 | 107 | 25 | 23 | 3 | | 260 | 135 | 39 | 29 | 1 | 280 | 40 | 37 | 17 | 2 | 250 | 117 | 36 | 16 | 3 |  1. **实验软件**   R语言   1. **实验结果**  加载包 library(readxl) library(dplyr) library(purrr) 第一题 multple.var.test=function(data1,data2,data3,k=3,alpha = 0.05){   nn=matrix(NA,nrow=1,ncol=k)#储存 k 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  nn[1,3]=nrow(data3)#第三个总体有 n3 行  n=apply(nn,1,sum)#三个总体的总行数  p=ncol(data1)#p 列  AA=array(NA,c(p,p,k))#储存 k 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  AA[,,3]=(nn[1,3]-1)\*var(data3)#第三个总体的离差阵  A=AA[,,1]+AA[,,2]+AA[,,3]#三个总体的离差阵之和  S=1#lambda 的分母  for(i in 1:k){   S=S\*(det(AA[,,i]/(nn[1,i]-1))^(-(nn[1,i]-1)/2))   }   lambda=((det(A/(n-k)))^(-(n-k)/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]||nn[1,2]==nn[1,3]){   d=((2\*p^2+3\*p-1)\*(k+1))/(6\*(p+1)\*(n-k))   }else{   d=((2\*p^2+3\*p-1)/6\*(p+1)\*(k-1))\*(b-(1/(n-k)))   }   T0=(1-d)\*M#检验统计量  f=(1/2)\*p\*(p+1)\*(k-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 <- read\_xlsx("D:/预删除文件夹/大三下/多元统计/实验5数据.xlsx") data1 <- data %>%   group\_split(组别) %>%  map(\(x) x[,1:4]) multple.var.test(data1[[1]],data1[[2]],data1[[3]],alpha = 0.01)  ## $统计量 ## [1] 20.33162 ##  ## $临界值 ## [1] 37.56623 ##  ## $p.value ## [1] 0.4373646  先对数据作协方差阵检验，以便判断使用何种均值向量检验的方式，检验结果显示在0.01的显著性水平下不拒绝原假设，即认为协方差阵相等  X <- as.matrix(data[,1:4]) Y <- factor(data$组别) model <- manova(X~Y) summary(model,test = "Wilks")  ## Df Wilks approx F num Df den Df Pr(>F)  ## Y 2 0.66212 3.0907 8 108 0.003538 \*\* ## Residuals 57  ## --- ## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  基于协方差阵相等的假定进一步作均值向量的检验，检验结果显示在0.01的显著性水平下拒绝原假设，即认为均值向量不全相等   1. **教师评语** | | | | | |