

《统计软件》实践报告

|  |  |  |
| --- | --- | --- |
|  |  |  |
| **学** | **院** | 理学院 |
| **专** | **业** | 数据科学与大数据技术 |
| **班** | **级** | 大数据232 |
| **学** | **号** | 1230510006 |
| **姓** | **名** | 张方波 |

2025 年 7 月 11 日

# 一 概率模型

### 实验一 古典概率的计算

|  |
| --- |
| **一、实验目的：**  1.熟悉概率的概念和性质  2.掌握古典概率的计算方法，并通过实例加深对概率概念和性质的理解 |
| **二、实验环境：**   1. 操作系统：Win 2. 环境： R编译器及RStudio |
| **三、实验内容、要求**  见实验指导书 |
| **四、实验步骤（对实验步骤的说明应该能够保证根据该说明即可重复完整的实验内容，得到正确结果。）**  输入的指令、代码、运行结果及截图。  例题1.假设每人的生日在一年 365 天中的任一天是等可能的,即都等于 1/365 ,求 ( nn≤ 365) 个人中至少有2人生日相同的概率。  在键入函数的函数同时按shift+回车就可以不执行代码换行，直接shift+回车键入n=seq(10,110,10)，p(n)即可直接回车执行  p=function(n)  {  p=1-choose(365,n)\*factorial(n)/365^n  p  }  n=seq(10,110,10)  p(n)    例题2.从5双不同的鞋子中任取4只，问这4只鞋子中至少有两只配成一双的概率。  p=(choose(5,1)\*choose(4,2)\*choose(2,1)\*choose(2,1)+choose(5,2))/choose(10,4)  p    练习1. 某接待站在某一周曾接待过 12次来访，已知所有这 12 次接待都是在周二和周四进 行的，问是否可以推断接待时间是有规定的。  observed = c(8, 4)  expected = c(6, 6)  result = chisq.test(observed, p = c(0.5, 0.5))  print(result)  if (result$p.value < 0.05) {  cat("结论：接待时间存在明显的规律性。\n")  } else {  cat("结论：无法判断接待时间是否有规律，可能是随机的。\n")  }    复制直接运行即可，可以修改observed中的内容改变预测值。  练习2. 50 只铆钉随机地取来用在10个部件上，其中有3只铆钉强度太弱.每个部件用3只 铆钉。 若将3只强度太弱的铆钉都装在一个部件上，则这个部件强度就太弱，求发生一个 部件强度太弱的概率。  set.seed(2025)  n\_trials = 100000  simulate\_once = function() {  rivets = c(rep("weak", 3), rep("strong", 47))  sample\_rivets = sample(rivets, 30)  parts = matrix(sample\_rivets, nrow = 10, byrow = TRUE)  any(apply(parts, 1, function(part) all(part == "weak")))  }  probability = mean(replicate(n\_trials, simulate\_once()))  cat("估计至少有一个部件全弱的概率为：", round(probability, 6), "\n") |
| **五、实验结果与分析(可以是实验详细设计、数据结构定义、算法源程序、流程图、算法时空复杂度分析等任意心得体会)：**  1 做的过程中遇到什么问题？如何解决的？  （1.不知道如何将代码换行，回车之后直接运行了代码：按shift+回车就可以实现换行  2 知识总结与体会  通过本次实验，我掌握了古典概率的基本概念和常见计算方法，并学会了使用 R 语言进行组合数、排列数以及概率的编程计算。实验中通过两个实际问题的求解，我对古典概率模型有了更深的理解：   1. 在“生日问题”中，运用了排列的思想计算至少有两人生日相同的概率，体会到随着人数增加，重复的可能性迅速上升。 2. 在“配对鞋子问题”和“铆钉分配问题”中，学习了如何用组合数分析复杂情况，并通过模拟估计实际概率。   此外，我也学会了使用 choose()、factorial()、chisq.test()、sample() 等 R 语言函数进行概率计算与模拟。通过编程实现，增强了对抽象概率模型的直观理解，提高了动手能力和分析问题的能力。  本次实验让我认识到：**概率并不是凭感觉判断，而是可以通过严谨的数学与模拟分析得出的科学结论。** |

### 实验二 频率稳定性

|  |
| --- |
| **一、实验目的：**  1.理解频率和概率的概念  2.理解频率和概率的关系 |
| **二、实验环境：**   1. 操作系统：Win 2. 环境： R编译器及RStudio |
| **三、实验内容、要求**  见实验指导书 |
| **四、实验步骤（对实验步骤的说明应该能够保证根据该说明即可重复完整的实验内容，得到正确结果。）**  输入的指令、代码、运行结果及截图。  例题1. 模拟10000 次投币试验的结果，统计其中随机数1（表示出现正面）和0（表示出现 反面）出现的次数，并对试验结果进行分析。  a=table(rbinom(1000,1,0.5))  a  a/1000    例题2. 向桌面上任意投掷一颗骰子，由于骰子的构造是均匀的，可知出现 1,2,...,6 这六个数 （朝上的点数）中任一个数的可能性是相同的。试产生离散均匀分布随机数对其进行模拟， 并对试验结果进行分析。  x=1:6  a=table(sample(x,1000,1/6))  a/1000    练习1. 利用随机数发生器产生10000个均匀分布随机数，分别记录其中小于0.5（表示出现正 面）和不小于0.5（表示出现反面）的随机数的个数，并对试验结果进行分析。  rand = runif(10000)  count\_head = sum(rand < 0.5)  count\_tail = sum(rand >= 0.5)  cat("正面次数：", count\_head, "\n")  cat("反面次数：", count\_tail, "\n")  cat("正面频率：", count\_head / 10000, "\n")  cat("反面频率：", count\_tail / 10000, "\n") |
| **五、实验结果与分析(可以是实验详细设计、数据结构定义、算法源程序、流程图、算法时空复杂度分析等任意心得体会)：**  1 做的过程中遇到什么问题？如何解决的？  2 知识总结与体会  通过本次实验，我深入理解了频率和概率的关系，尤其是频率的稳定性原理。通过使用 R 语言模拟投币试验和掷骰子，我看到了在大量重复试验中，事件的频率确实趋近于其理论概率，这不仅加深了我对概率本质的认识，也锻炼了我使用 R 语言进行随机模拟、计数和抽样的能力。  特别是在第三题中，我利用 runif() 函数从均匀分布中模拟了“伪硬币”，这让我更加理解了如何将数学概念转化为实际编程操作。本次实验不仅巩固了概率基础理论，也提升了数据模拟与分析的能力。 |

### 实验三 圆周率π的近似计算——蒲丰投针问题

|  |
| --- |
| **一、实验目的：**  1.理解几何概型的概率的概念和计算方法  2.频率稳定性的应用  3.掌握无理数π的近似计算方法 |
| **二、实验环境：**   1. 操作系统：Win   环境： R编译器及RStudio |
| **三、实验内容、要求**  见实验指导书 |
| **四、实验步骤（对实验步骤的说明应该能够保证根据该说明即可重复完整的实验内容，得到正确结果。）**  输入的指令、代码、运行结果及截图。  例子1.  buffon=function(n)  {  l=3 #这是小写的字母L  d=4  m=0  for(k in 1:n) #数字1  {  y=runif(1,0,d/2) #数字1  x=runif(1,0,pi) #数字1  if(y<0.5\*l\*sin(x)) #字母L  m=m+1 #数字1  else  m=m  }  PI=2\*n\*l/(m\*d) #字母L  PI  }  buffon(10000)    练习一.  # 定义函数进行圆内投点模拟  circle\_pi = function(n) {  x = runif(n, -1, 1) # 随机生成 x 坐标  y = runif(n, -1, 1) # 随机生成 y 坐标  inside = (x^2 + y^2) <= 1 # 判断是否落在单位圆内  m = sum(inside) # 圆内点个数  PI = 4 \* m / n # 根据概率公式估算 π  return(PI)  }  # 调用函数进行模拟  circle\_pi(10000) |
| **五、实验结果与分析(可以是实验详细设计、数据结构定义、算法源程序、流程图、算法时空复杂度分析等任意心得体会)：**  1 做的过程中遇到什么问题？如何解决的？  2 知识总结与体会  本次实验通过蒲丰投针实验与投点估算圆周率两种方式，全面体验了几何概型在概率中的应用与“频率稳定性”的原理。以下是我的几点体会：   1. **几何概型的特点**：本质是将“概率”转化为“面积之比”或“体积之比”，非常适合借助模拟和图形化方式进行理解。 2. **频率逼近概率**：不论是投针还是投点，随着试验次数的增大，频率逐渐逼近理论概率，这种“数值收敛”的过程让我更直观地感受到概率的定义。 3. **蒙特卡洛方法的强大**：通过大量的随机模拟，我们能够逼近 π 这样的数学常数，也启发我在以后的问题求解中可以考虑模拟而不是强行推导。 4. **R语言能力提升**：掌握了 runif()、sum()、sin()、逻辑判断、函数定义等语法，在实践中提升了数据模拟建模能力。   总之，本实验不仅加深了我对概率概型的理解，还提高了我用编程工具解决数学问题的能力，为后续的概率论学习和数据科学研究打下了坚实基础。 |

### 实验四 蒙特霍尔问题

|  |
| --- |
| **一、实验目的：**  **1.加深频率和概率概念的理解**  **2.掌握条件概率的实际含义**  **3.掌握全概率公式和贝叶斯公式的应用** |
| **二、实验环境：**   1. **操作系统：Win**   **环境： R编译器及RStudio** |
| **三、实验内容、要求**  **见实验指导书** |
| **四、实验步骤（对实验步骤的说明应该能够保证根据该说明即可重复完整的实验内容，得到正确结果。）**  **输入的指令、代码、运行结果及截图。**  **例题一.**  montyhall=function(n)  {  m=0  l=0  x=c(1,2,3)  for(i in 1:n)  {  k=sample(c(1,2,3),1)  if(x[k]==2)  {  m=m+1  l=1  }  else  {  m=m  l=l+1  }  }  nochange=m/n  change=l/n  paste("nochange=",nochange,"change=",change)  }  montyhall(1000)  montyhall(10000)  montyhall(100000)  montyhall(1000000)    **练习一.**  sensitive\_survey = function(n, a = 3, b = 7, p = 0.6, q = 0.4) {  yes\_count = 0 # 回答“是”的次数    for (i in 1:n) {  ball = sample(c("red", "white"), 1, prob = c(a / (a + b), b / (a + b)))    if (ball == "red") {  # 回答敏感问题（作弊）  if (runif(1) < p) {  yes\_count = yes\_count + 1  }  } else {  # 回答普通问题  if (runif(1) < q) {  yes\_count = yes\_count + 1  }  }  }    # 回答“是”的总体频率  prob\_yes = yes\_count / n  return(prob\_yes)  }  sensitive\_survey(10000) |
| **五、实验结果与分析(可以是实验详细设计、数据结构定义、算法源程序、流程图、算法时空复杂度分析等任意心得体会)：**  **1 做的过程中遇到什么问题？如何解决的？**  **2 知识总结与体会**  通过本次实验，我深入理解了条件概率、全概率公式与贝叶斯定理在实际中的应用，特别是：   1. **直觉 VS 理性**： 蒙特霍尔问题初看令人困惑，模拟验证了“换门胜率更高”的事实，纠正了错误直觉，增强了用数学思维分析问题的能力。 2. **频率逼近概率**： 多次模拟显示频率稳定性，换门成功概率趋于2/3，是对概率定义的又一次形象体现。 3. **贝叶斯思想初步应用**： 在敏感性问题模拟中，通过设计随机机制保护隐私的同时，依然可以估计总体行为特征，体会到了“可追踪但不可识别”的调查方法，这在统计学和隐私保护中具有重要意义。 4. **R语言提升**： 学习了 sample()、runif()、逻辑判断和计数等方法，掌握了构造概率模型与模拟估算的技能，为后续数据分析打下基础。 |

# 二. 随机变量及其分布

### 实验一 随机数的生成

|  |
| --- |
| **一、实验目的：**  1.熟悉常见分布的随机数的产生  2.掌握利用随机数进行随机模拟的方法 |
| **二、实验环境：**   1. 操作系统：Win   环境： R编译器及RStudio |
| **三、实验内容、要求**  见实验指导书 |
| **四、实验步骤（对实验步骤的说明应该能够保证根据该说明即可重复完整的实验内容，得到正确结果。）**  输入的指令、代码、运行结果及截图。  例题一.  rbinom(1,20,0.25)  rbinom(10,20,0.25)  matrix(rbinom(21,20,0.25),3,7)    例题二.  matrix(runif(81,1,3),9,9)    例题三.  rexp(9,1/3)    例题四.  rnorm(7,4,2)    练习一.  x = runif(9, min = 0, max = 1)  x    练习二.  y = rnorm(8, mean = 0, sd = 1)  y |
| **五、实验结果与分析(可以是实验详细设计、数据结构定义、算法源程序、流程图、算法时空复杂度分析等任意心得体会)：**  1 做的过程中遇到什么问题？如何解决的？  2 知识总结与体会  这些函数是R中模拟连续型随机变量分布的基本工具，在之后涉及中心极限定理、置信区间估计等实验中会频繁使用。 |

### 实验二.概率作图

|  |
| --- |
| **一、实验目的：**  1.掌握常用离散型分布的分布律和分布函数图  2.掌握常用连续型分布的概率密度和分布函数图 |
| **二、实验环境：**   1. 操作系统：Win   环境： R编译器及RStudio |
| **三、实验内容、要求**  见实验指导书 |
| **四、实验步骤（对实验步骤的说明应该能够保证根据该说明即可重复完整的实验内容，得到正确结果。）**  输入的指令、代码、运行结果及截图。  例题一.  x=0:10  y=dbinom(x,10,0.3)  plot(x,y,main = "二项分布的分布律",pch=21,bg="blue")    例题二.  x=seq(-1,11,0.1)  y=pbinom(x,10,0.3)  plot(x,y,main="二项分布的分布函数",type="l",xlab="",ylab="")    例题三.  x=seq(-2,14,0.1)  y=dnorm(x,6,2)  plot(x,y,main="正态分布的概率密度图形",type="l",xlab="",ylab="")    x=seq(-2,14,0.1)  y=pnorm(x,6,2)  plot(x,y,main="正态分布的分布函数图形",type="l",xlab="",ylab="")    例题四.  x=0:25  bx=dbinom(x,25,0.15)  px=dpois(x,3.75)  data=rbind(bx,px)  barplot(data,col=c("blue","red"),beside = TRUE)  legend("topright",pch=15,c("二项概率","泊松概率"),col=c("blue","red"))    练习一（1）.  x = seq(-15, 15, 0.1)  y1 = dnorm(x, mean = -3, sd = 2)  y2 = dnorm(x, mean = 3, sd = 2)  y3 = dnorm(x, mean = 5, sd = 2)  plot(x, y1, type = "l", col = "blue", ylim = c(0, 0.25), main = "不同均值下的正态密度函数", ylab = "密度", xlab = "x")  lines(x, y2, col = "red")  lines(x, y3, col = "green")  legend("topright", legend = c("均值 = -3", "均值 = 3", "均值 = 5"), col = c("blue", "red", "green"), lty = 1)    x = seq(-15, 15, 0.1)  y1 = pnorm(x, mean = -3, sd = 2)  y2 = pnorm(x, mean = 3, sd = 2)  y3 = pnorm(x, mean = 5, sd = 2)  plot(x, y1, type = "l", col = "blue", ylim = c(0, 1), main = "不同均值下的正态分布函数", ylab = "分布函数", xlab = "x")  lines(x, y2, col = "red")  lines(x, y3, col = "green")  legend("bottomright", legend = c("mean = -3", "mean = 3", "mean = 5"), col = c("blue", "red", "green"), lty = 1)    练习一（2）.  x = seq(-5, 17, 0.1)  y1 = dnorm(x, mean = 6, sd = 1)  y2 = dnorm(x, mean = 6, sd = 2)  y3 = dnorm(x, mean = 6, sd = 3)  plot(x, y1, type = "l", col = "blue", ylim = c(0, 0.45), main = "不同标准差下的正态密度函数", ylab = "密度", xlab = "x")  lines(x, y2, col = "red")  lines(x, y3, col = "green")  legend("topright", legend = c("sd = 1", "sd = 2", "sd = 3"), col = c("blue", "red", "green"), lty = 1)    x = seq(-5, 17, 0.1)  y1 = pnorm(x, mean = 6, sd = 1)  y2 = pnorm(x, mean = 6, sd = 2)  y3 = pnorm(x, mean = 6, sd = 3)  plot(x, y1, type = "l", col = "blue", ylim = c(0, 1), main = "不同标准差下的正态分布函数", ylab = "分布函数", xlab = "x")  lines(x, y2, col = "red")  lines(x, y3, col = "green")  legend("bottomright", legend = c("sd = 1", "sd = 2", "sd = 3"), col = c("blue", "red", "green"), lty = 1)    练习二（1）.  x = seq(0, 100, 1)  y = dexp(x, rate = 1/16)  plot(x, y, type = "l", col = "blue", main = "指数分布概率密度函数 (θ = 16)", ylab = "密度", xlab = "x")    练习二（2）.  x = seq(0, 100, 1)  y = pexp(x, rate = 1/16)  plot(x, y, type = "l", col = "red", main = "指数分布分布函数 (θ = 16)", ylab = "分布函数", xlab = "x")    练习三（1）.  x = seq(0, 8, 0.1)  y = dunif(x, min = 2, max = 6)  plot(x, y, type = "l", col = "blue", main = "均匀分布密度函数 [2,6]", ylab = "密度", xlab = "x")    练习三（2）.  x = seq(0, 8, 0.1)  y = punif(x, min = 2, max = 6)  plot(x, y, type = "l", col = "green", main = "均匀分布分布函数 [2,6]", ylab = "分布函数", xlab = "x") |
| **五、实验结果与分析(可以是实验详细设计、数据结构定义、算法源程序、流程图、算法时空复杂度分析等任意心得体会)：**  1 做的过程中遇到什么问题？如何解决的？  2 知识总结与体会  本实验通过 R 语言绘制常见概率分布图，直观展示了**离散分布**（如二项分布、泊松分布）的分布律与分布函数，以及**连续分布**（如正态、指数、均匀分布）的概率密度与分布函数。  通过图形观察，我们掌握了参数变化对分布形状的影响，例如标准差影响正态分布曲线宽度，泊松分布能近似拟合大样本低概率的二项分布等规律。  实验中使用 dbinom()、dpois()、dnorm()、pexp()、punif() 等函数快速实现绘图，增强了对抽象概率概念的理解。R语言强大的绘图功能提升了我们统计分析的效率与可视化表达能力。  本实验帮助我们建立了概率分布与图形之间的直观联系，为后续数据分析与建模打下了基础。 |

### 实验三. 服务窗口设置问题

|  |
| --- |
| **一、实验目的：**  1.加深对二项分布的理解  2.掌握二项分布在实际问题中的应用 |
| **二、实验环境：**   1. 操作系统：Win   环境： R编译器及RStudio |
| **三、实验内容、要求**  见实验指导书 |
| **四、实验步骤（对实验步骤的说明应该能够保证根据该说明即可重复完整的实验内容，得到正确结果。）**  输入的指令、代码、运行结果及截图。  例题一.  n=200  s=3  p=0.1  a=0.95  for(m in 1:15)  {  l=0  c=0  for(k in 0:(m\*s))  {  l=l+choose(n,k)\*p^k\*(1-p)^(n-k)  }  cat("l(",m,")=",l,"\n")  }    p=0.1  n=200  s=3  zcs=1000  alpha=0.95  for(m in 1:15)  {  b=0  for(j in 1:zcs)  {  a=0  for(i in 1:n)  {  x=rbinom(1,1,p)  if(x==1)  a=a+1  }  if(a<=m\*s)  b=b+1  }  cat("m(",m,")=",b/zcs,"\n")  } |
| **五、实验结果与分析(可以是实验详细设计、数据结构定义、算法源程序、流程图、算法时空复杂度分析等任意心得体会)：**  1 做的过程中遇到什么问题？如何解决的？  2 知识总结与体会 |

### 实验四. 巴拿赫火柴盒问题

|  |
| --- |
| **一、实验目的：**  1.加深对二项分布的理解  2.掌握二项分布在实际问题中的应用 |
| **二、实验环境：**   1. 操作系统：Win   环境： R编译器及RStudio |
| **三、实验内容、要求**  见实验指导书 |
| **四、实验步骤（对实验步骤的说明应该能够保证根据该说明即可重复完整的实验内容，得到正确结果。）**  输入的指令、代码、运行结果及截图。  练习一.  n=20  for(k in 0:20)  {  p=choose(2\*n-k,n)\*0.5^(2\*n-k)  cat("p(",k,")=",p,"\n")  }    k=2  for(n in 10:20)  {  p=choose(2\*n-k,n)\*0.5^(2\*n-k)  cat("p(",n,")=",p,"\n")  }    n=20  p=0.5  a=1000  for(k in 1:n)  {  m=0  for(j in 1:a)  {  n1=n  nr=n  for(i in 1:(2\*n))  {  x=rbinom(1,1,p)  if(x==0)  n1=n1-1  else  nr=nr-1  if(n1==0|nr==0)  break    }  if((n1==0&nr==k)|(nr==0&n1==k))  m=m+1  }  cat("P(",n,",",k,")=",m/a,"\n")  } |
| **五、实验结果与分析(可以是实验详细设计、数据结构定义、算法源程序、流程图、算法时空复杂度分析等任意心得体会)：**  1 做的过程中遇到什么问题？如何解决的？  2 知识总结与体会  掌握了如何将实际问题抽象为概率模型；理解了二项分布在非对称情境下的应用；学会使用 R语言编写逻辑与计数结构，实现复杂事件的概率模拟；体会到理论计算与蒙特卡洛模拟的互补作用。本实验不仅提高了我们对随机事件建模的能力，也培养了我们用计算手段验证概率结论的实践技能 |

### 实验五. 公交汽车车门高度设计

|  |
| --- |
| **一、实验目的：**  1.加深对正态分布的理解  2.掌握正态分布在实际问题中的应用 |
| **二、实验环境：**   1. 操作系统：Win   环境： R编译器及RStudio |
| **三、实验内容、要求**  见实验指导书 |
| **四、实验步骤（对实验步骤的说明应该能够保证根据该说明即可重复完整的实验内容，得到正确结果。）**  输入的指令、代码、运行结果及截图。 |
| **五、实验结果与分析(可以是实验详细设计、数据结构定义、算法源程序、流程图、算法时空复杂度分析等任意心得体会)：**  1 做的过程中遇到什么问题？如何解决的？  2 知识总结与体会  通过此次实验，我们不仅学到了正态分布的实际应用，更提高了将数学模型与现实问题结合的能力。 |

### 实验六. 考试录取问题

|  |
| --- |
| **一、实验目的：**  1.加深对正态分布的理解  2.掌握正态分布在实际问题中的应用 |
| **二、实验环境：**   1. 操作系统：Win   环境： R编译器及RStudio |
| **三、实验内容、要求**  见实验指导书 |
| **四、实验步骤（对实验步骤的说明应该能够保证根据该说明即可重复完整的实验内容，得到正确结果。）**  输入的指令、代码、运行结果及截图。  31/1821  qnorm(1-0.017,0,1)  194/2.1201  pnorm(90/91.5,0,1)  1-pnorm(90/91.5,0,1)  0.1627\*1821 |
| **五、实验结果与分析(可以是实验详细设计、数据结构定义、算法源程序、流程图、算法时空复杂度分析等任意心得体会)：**  1 做的过程中遇到什么问题？如何解决的？  2 知识总结与体会  本实验强化了正态分布的实战运用能力，对实际人事选拔、分级录取等问题具有重要参考价值 |

# 三 大数定律和中心极限定理

### 实验一 定积分的计算

|  |
| --- |
| **一、实验目的：**  1.加深对大数定理的认识，对其背景和应用有直观的理解  2.了解R软件在模拟仿真中的应用 |
| **二、实验环境：**   1. 操作系统：Win   环境： R编译器及RStudio |
| **三、实验内容、要求**  见实验指导书 |
| **四、实验步骤（对实验步骤的说明应该能够保证根据该说明即可重复完整的实验内容，得到正确结果。）**  输入的指令、代码、运行结果及截图。  times = 1000  x = runif(times)  y = x^2  I = mean(y)  I  y = function(x) {  y = x^2  }  integrate(y, 0, 1) |
| **五、实验结果与分析(可以是实验详细设计、数据结构定义、算法源程序、流程图、算法时空复杂度分析等任意心得体会)：**  1 做的过程中遇到什么问题？如何解决的？  2 知识总结与体会  本实验是大数定律在数值计算中应用的直观体现，体现了概率思想在工程计算和建模中的实际价值 |

### 实验二 参加家长会人数问题

|  |
| --- |
| **一、实验目的：**  1.加深对中心极限定理的认识，对其背景和应用有直观的理解  2.了解R软件在模拟仿真中的应用 |
| **二、实验环境：**   1. 操作系统：Win   环境： R编译器及RStudio |
| **三、实验内容、要求**  见实验指导书 |
| **四、实验步骤（对实验步骤的说明应该能够保证根据该说明即可重复完整的实验内容，得到正确结果。）**  输入的指令、代码、运行结果及截图。  data = matrix(sample(c(0,1,2), 400000, c(0.05, 0.8, 0.15), replace = TRUE), 400, 1000)  s1 = apply(data, 2, sum)  length(which(s1 > 450))  p1 = length(which(s1 > 450)) / 1000  paste("参加会议的家长人数X超过450的概率:", p1)  s2 = 0  for(i in 1:1000) {  s2[i] = length(which(data[, i] == 1))  }  length(which(s2 <= 340))  p2 = length(which(s2 <= 340)) / 1000  paste("有1名家长来参加会议的学生人数不多于340的概率:", p2) |
| **五、实验结果与分析(可以是实验详细设计、数据结构定义、算法源程序、流程图、算法时空复杂度分析等任意心得体会)：**  1 做的过程中遇到什么问题？如何解决的？  2 知识总结与体会  本实验是中心极限定理在实际建模中的典型应用，通过模拟学生家长参会情况，展示了大量独立随机变量之和近似服从正态分布的规律。我们利用 R 语言进行蒙特卡罗模拟，不仅验证了理论概率的计算结果，也加深了对中心极限定理的直观理解。该实验体现了概率统计在工程决策与资源调配中的重要作用，为解决实际中的不确定问题提供了有效的建模工具和计算手段。 |

### 实验三 保险公司盈利问题

|  |
| --- |
| **一、实验目的：**  1.加深对中心极限定理的认识，对其背景和应用有直观的理解  2.了解R软件在模拟仿真中的应用 |
| **二、实验环境：**   1. 操作系统：Win   环境： R编译器及RStudio |
| **三、实验内容、要求**  见实验指导书 |
| **四、实验步骤（对实验步骤的说明应该能够保证根据该说明即可重复完整的实验内容，得到正确结果。）**  输入的指令、代码、运行结果及截图。  low = 5.9  up = 6.1  n = 10000  fee = 12  p = 0.006  fp = 1000  Ex = fp \* p  Dx = fp \* p \* (1 - p)  Exx = Ex  Dxx = Dx / n  P1 = pnorm((up - Exx) / sqrt(Dxx)) - pnorm((low - Exx) / sqrt(Dxx))  P1  yn = n \* fee / fp  m = n \* p  v = n \* p \* (1 - p)  P2 = 1 - pnorm(yn, m, sqrt(v))  P2  P2 = 1 - pnorm(90, m, sqrt(v))  P2  P2 = pnorm(80, m, sqrt(v))  P2 |
| **五、实验结果与分析(可以是实验详细设计、数据结构定义、算法源程序、流程图、算法时空复杂度分析等任意心得体会)：**  1 做的过程中遇到什么问题？如何解决的？  2 知识总结与体会  本实验基于中心极限定理，从理论与模拟两个角度分析了保险公司在大样本投保人群中的盈利风险。我们通过 R 语言计算了每户的期望赔付与方差，进一步估算了平均赔付金额落在特定区间的概率、公司整体亏损的概率及高利润概率。  结果表明，在投保人数达到 10,000 时，赔付金额的波动极小，平均每户赔付近似为常数，说明中心极限定理对大数样本下的总赔付具有良好的近似能力。特别地，保险公司亏本的概率几乎为零，而每年利润超过 4 万元的概率高达 99.5%，显示出保险业务的高稳定性与低风险特征。  本实验不仅加深了我们对中心极限定理的理解，也展现了统计理论在金融风险评估中的实际应用价值，培养了我们运用概率模型解决现实问题的能力。 |

# 四 抽样分布

### 实验一 模拟抽样

|  |
| --- |
| **一、实验目的：**  1.理解样本的随机性  2.掌握抽样过程的具体步骤  3.了解R软件在模拟抽样中的应用 |
| **二、实验环境：**   1. 操作系统：Win   环境： R编译器及RStudio |
| **三、实验内容、要求**  见实验指导书 |
| **四、实验步骤（对实验步骤的说明应该能够保证根据该说明即可重复完整的实验内容，得到正确结果。）**  输入的指令、代码、运行结果及截图。  sample(c(76,56,55,85,23,74,63,46,81,57), 5, replace = TRUE) |
| **五、实验结果与分析(可以是实验详细设计、数据结构定义、算法源程序、流程图、算法时空复杂度分析等任意心得体会)：**  1 做的过程中遇到什么问题？如何解决的？  2 知识总结与体会  本实验通过模拟从一个固定总体中抽取样本的过程，加深了我们对抽样随机性的认识。通过使用 sample() 函数进行随机抽样，可以直观地理解每次抽样结果的变动性，体现了样本的代表性并非绝对，而是具有一定的随机性和不确定性。此外，也认识到合理设计抽样过程对于保证推断的准确性至关重要，为后续统计推断奠定基础。 |

### 实验二 总体与抽样分布

|  |
| --- |
| **一、实验目的：**  1.理解抽样分布的概念  2.理解总体分布与抽样分布的关系 |
| **二、实验环境：**   1. 操作系统：Win   环境： R编译器及RStudio |
| **三、实验内容、要求**  见实验指导书 |
| **四、实验步骤（对实验步骤的说明应该能够保证根据该说明即可重复完整的实验内容，得到正确结果。）**  输入的指令、代码、运行结果及截图。  一.  X = c(2,4,6)  x = matrix(c(2,4,2,6,4,6,4,2,6,2,6,4,2,2,4,4,6,6), 2, 9)  xbar = apply(x, 2, mean)  Exbar = mean(xbar)  se = sqrt(sum((xbar - Exbar)^2) / length(xbar))  u = mean(X)  SD = sqrt(sum((X - u)^2) / length(X))  SD / sqrt(nrow(x))    二.  X = c(2,4,6)  x = matrix(c(2,4,2,6,4,6,2,2,4,4,6,6), 2, 6)  xbar = apply(x, 2, mean)  p = c(2/9, 2/9, 2/9, 1/9, 1/9, 1/9)  Exbar = sum(p \* xbar)  se = sqrt(sum((xbar - Exbar)^2 \* p))  u = mean(X)  SD = sqrt(sum((X - u)^2) / length(X))  SD / sqrt(nrow(x))    三  X = c(2,4,6)  x = matrix(c(2,4,2,6,4,6,4,2,6,2,6,4), 2, 6)  xbar = apply(x, 2, mean)  Exbar = mean(xbar)  se = sqrt(sum((xbar - Exbar)^2) / length(xbar))  u = mean(X)  SD = sqrt(sum((X - u)^2) / length(X))  SD \* sqrt((length(X) - nrow(x)) / ((length(X) - 1) \* nrow(x)))    四  X = c(2,4,6)  x = matrix(c(2,4,2,6,4,6), 2, 3)  xbar = apply(x, 2, mean)  Exbar = mean(xbar)  se = sqrt(sum((xbar - Exbar)^2) / length(xbar))  u = mean(X)  SD = sqrt(sum((X - u)^2) / length(X))  SD \* sqrt((length(X) - nrow(x)) / ((length(X) - 1) \* nrow(x))) |
| **五、实验结果与分析(可以是实验详细设计、数据结构定义、算法源程序、流程图、算法时空复杂度分析等任意心得体会)：**  1 做的过程中遇到什么问题？如何解决的？  2 知识总结与体会  本实验围绕抽样分布展开，通过四种不同抽样方式的对比，进一步理解了总体分布、样本分布和抽样分布之间的关系。抽样分布体现了样本统计量的分布规律，揭示了样本均值在重复抽样中的波动程度。通过实践，我们验证了理论中关于标准误差的公式，也体会到样本容量、抽样方式（是否放回、是否考虑顺序）对抽样分布的影响。这为今后进行统计推断、区间估计和假设检验等操作提供了理论依据和实践经验。 |

# 五 参数估计

### 实验一 点估计

|  |
| --- |
| **一、实验目的：**  1.加深理解矩估计原理  2.加深理解极大似然估计原理 |
| **二、实验环境：**   1. 操作系统：Win   环境： R编译器及RStudio |
| **三、实验内容、要求**  见实验指导书 |
| **四、实验步骤（对实验步骤的说明应该能够保证根据该说明即可重复完整的实验内容，得到正确结果。）**  输入的指令、代码、运行结果及截图。  例题一.  data = c(74.001, 74.005, 74.003, 74.001, 74.000, 73.998, 74.006, 74.002)  mean(data)  sum((data - mean(data))^2 / length(data))  var(data)    例题二.  r = 0:5  s = c(44, 42, 21, 9, 4, 2)  sum(r \* s) / sum(s)  exp(-sum(r \* s) / sum(s))    练习一.  x = 0:10  f = c(0, 1, 6, 7, 23, 26, 21, 12, 3, 1, 0)  total\_samples = sum(f)  total\_successes = sum(x \* f)  p\_mle = total\_successes / (total\_samples \* 10)  p\_mle |
| **五、实验结果与分析(可以是实验详细设计、数据结构定义、算法源程序、流程图、算法时空复杂度分析等任意心得体会)：**  1 做的过程中遇到什么问题？如何解决的？  2 知识总结与体会  本实验通过对活塞环直径数据的分析、泊松分布下极大似然估计的计算以及实际地质数据的练习，系统地加深了对\*\*点估计方法（矩估计和极大似然估计）\*\*的理解。  矩估计法是通过将样本矩代入总体矩公式中推导参数估计，具有计算简单、直观的优点。  \*\*极大似然估计法（MLE）\*\*则通过最大化样本发生的可能性，得到在观测数据下最可能的参数值。MLE 拥有一致性、渐近正态性等良好性质，是现代统计推断的核心工具之一。  在 R 软件中，我们使用 mean() 和 var() 函数快速获得样本统计量，配合指数函数 exp()，即可得到泊松分布下的事件概率估计结果。 |

### 实验二 正态总体参数的区间估计的模拟

|  |
| --- |
| **一、实验目的：**  1.加深对置信区间的理解  2.理解上α分位点的概念 |
| **二、实验环境：**   1. 操作系统：Win   环境： R编译器及RStudio |
| **三、实验内容、要求**  见实验指导书 |
| **四、实验步骤（对实验步骤的说明应该能够保证根据该说明即可重复完整的实验内容，得到正确结果。）**  输入的指令、代码、运行结果及截图。  m = matrix(rnorm(900, 8, 1), 100, 9)  Mean = apply(m, 1, mean)  low = Mean - (1 / sqrt(9)) \* qnorm(0.975)  up = Mean + (1 / sqrt(9)) \* qnorm(0.975)  CIs = cbind(low, up)  CIlow = apply(CIs, 1, min)  CIup = apply(CIs, 1, max)  CIs = data.frame(low = CIlow, up = CIup)  plot(0, xlim = c(0, 100), ylim = c(min(CIs) - 0.2, max(CIs) + 0.2), type = "n", xlab = "", ylab = "")  for (i in 1:nrow(CIs)) {  lines(rep(i, 2), c(CIs[i, 1], CIs[i, 2]))  points(i, Mean[i], col = "green", cex = 0.5)  }  abline(h = 8) |
| **五、实验结果与分析(可以是实验详细设计、数据结构定义、算法源程序、流程图、算法时空复杂度分析等任意心得体会)：**  1 做的过程中遇到什么问题？如何解决的？  2 知识总结与体会  本实验通过R语言模拟正态总体 X∼N(8,1)X \sim N(8,1)X∼N(8,1)，每组样本容量为9，共生成100组样本。我们对每组样本计算均值，并构造95%置信区间，观察这些区间是否包含真实均值8。  实验结果表明，大约有95%的置信区间覆盖了总体均值，验证了置信区间的含义和构造方法的正确性。通过图形直观展示，我们对区间估计的频率解释有了更清晰的认识。  本实验加深了我们对置信区间、正态分布分位点和R语言模拟的理解，验证了统计推断在实际问题中的可靠性和直观性。 |

### 实验三 单个正态总体均值与方差的区间估计

|  |
| --- |
| **一、实验目的：**  1.加深对置信区间的理解  2.单个正态总体均值的置信区间  3.单个正态总体方差的置信区间 |
| **二、实验环境：**   1. 操作系统：Win   环境： R编译器及RStudio |
| **三、实验内容、要求**  见实验指导书 |
| **四、实验步骤（对实验步骤的说明应该能够保证根据该说明即可重复完整的实验内容，得到正确结果。）**  输入的指令、代码、运行结果及截图。  例题一.  data=c(506,508,499,503,504,510,497,512,  514,505,493,496,506,502,509,496)  t.test(data)    例题二.  data=c(506,508,499,503,504,510,497,512,  514,505,493,496,506,502,509,496)  var.conf.int=function(x,alpha)  {  n=length(x)  s=var(x)  df=n-1  c(sqrt(df\*s/qchisq(1-alpha/2,df)), sqrt(df\*s/qchisq(alpha/2,df)))  }  var.conf.int(data,0.05)    练习一.  tire=c(41250,40187,43175,41010,39265,41872,42654,41287,  38970,40200,42550,41095,40680,43500,39775,40400)  t.test(tire, conf.level=0.95)  var.conf.int=function(x,alpha)  {  n=length(x)  s=var(x)  df=n-1  c(sqrt(df\*s/qchisq(1-alpha/2,df)), sqrt(df\*s/qchisq(alpha/2,df)))  }  var.conf.int(tire,0.05) |
| **五、实验结果与分析(可以是实验详细设计、数据结构定义、算法源程序、流程图、算法时空复杂度分析等任意心得体会)：**  1 做的过程中遇到什么问题？如何解决的？  2 知识总结与体会  本实验通过R语言实现对正态总体均值与方差的区间估计，加深了我对置信区间概念的理解。我们探讨了两种情况：方差已知与未知时均值的估计方法，利用t分布进行均值置信区间计算，使用卡方分布构造方差的置信区间。实验通过真实样本数据模拟，强化了理论知识与实际计算的结合，提升了我们对统计推断严谨性与适用条件的认识。同时学会了如何使用R语言进行置信区间的图示与函数封装，使得复杂的公式计算变得直观高效。该实验有助于为后续的假设检验与多样本分析打下扎实基础。 |

### 实验四 两个正态总体均值差与方差比的区间估计

|  |
| --- |
| **一、实验目的：**  1.加深对置信区间的理解  2.两个正态总体均值差的置信区间  3.两个正态总体方差比的置信区间 |
| **二、实验环境：**   1. 操作系统：Win   环境： R编译器及RStudio |
| **三、实验内容、要求**  见实验指导书 |
| **四、实验步骤（对实验步骤的说明应该能够保证根据该说明即可重复完整的实验内容，得到正确结果。）**  输入的指令、代码、运行结果及截图。  例题一.  A = c(0.143, 0.142, 0.143, 0.137)  B = c(0.138, 0.14, 0.134, 0.138, 0.142)  t.test(A, B, var.equal = TRUE)    例题二.  var.test(A, B) |
| **五、实验结果与分析(可以是实验详细设计、数据结构定义、算法源程序、流程图、算法时空复杂度分析等任意心得体会)：**  1 做的过程中遇到什么问题？如何解决的？  2 知识总结与体会  本实验通过R语言对两个正态总体的均值差和方差比进行了区间估计，掌握了t检验和F检验在不同条件下的应用。利用 t.test 和 var.test 函数，能够快速得出置信区间并结合P值进行统计推断。实验结果显示，两组导线在均值和方差上均无显著差异，说明在小样本下推断存在一定不确定性。本实验强化了对置信区间含义、两总体比较方法以及分布上分位点应用的理解，为后续假设检验打下了实践基础。 |