**云南大学数学与统计学院**

**上机实验报告 （ 2）**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **实验课程名** | | | **抽样技术与应用实验** | | | | | | | | | **成绩** | |  | |
| **学号** |  | | | | **姓名** | | **枫叶** | **专业** | | **统计学** | **年级** | | **2021级** | | |
| **实验项目名称** | | | | 分层随机抽样在R软件中的实现 | | | | **日期** | | **2023/4/4** | **实验时间** | | | |  |
| **指导教师** | | **潘东东老师** | | | | **实验地址（室）** | | | **格物楼3507** | | | | | | |
| **教师评语** | |  | | | | | | | | | | | | | |

1. **实验目的：**

掌握分层随机抽样的方法在R软件的实现过程，并会对样本调查数据进行处理，掌握对总体均值方差进行估计的过程，将简单估计的结果与比估计和回归估计进行比较

1. **实验内容:**

1、以iris数据集的最后一列Species为分层标志，从前四列中任选一列为总体单元指标，从每一层中各抽10个样本(备注：需把自己学号最后两位或三位数字设置成随机数种子)。

2、基于所抽得的样本计算总体均值的分层估计，并计算其方差的估计。

3、若要求总体均值的分层估计的95%的绝对误差限r=5%，求比例分配、奈曼分配下所需的总样本量及各层的样本量。

4、利用前面抽取的分层样本数据，以Sepal.length作为主指标Y，以Sepal.width作为辅助指标X，计算Sepal.length总体均值的的分别比估计、联合比估计、分别回归估计、联合回归估计的值，并比较它们的标准差估计值。

**三、使用环境:**

**R**

**四、实验步骤:**

**1、数据准备：**

**Iris为内置数据**

**2、程序实验：**

**（1）**

library(sampling)

set.seed(172)

L <- length(unique(iris$Species))

n\_h <- rep(10,L)

st <- strata(iris,"Species",n\_h,"srswor")

data <- getdata(iris,st)

**（2）**

library(sampling)

library(survey)

set.seed(172)

N <- nrow(iris)

N\_h <- table(iris$Species)

W\_h <- N\_h/N

L <- length(unique(iris$Species))

n\_h <- rep(10,L)

st <- sampling::strata(iris,"Species",n\_h,"srswor")#survey包中也有strata函数

data <- getdata(iris,st)

pw <- 1/st$Prob #样本权重

fpc <- rep(1-n\_h/N\_h,c(n\_h))#有限总体修正系数

sampledata <- cbind(data,pw,fpc)

dstrat <- svydesign(ids = ~1,strata = ~Species,weights = ~pw,data = sampledata,fpc = ~fpc)#定义抽样设计

svymean(~Sepal.Length, dstrat)

**（3）**

**比例分配**

library(sampling)

set.seed(172)

st <- sampling::strata(iris,"Species",rep(10,3),"srswor")

data <- getdata(iris,st)

v1 <- var(data[data$Stratum %in% c(1),]$Sepal.Length)

v2 <- var(data[data$Stratum %in% c(2),]$Sepal.Length)

v3 <- var(data[data$Stratum %in% c(3),]$Sepal.Length)

V=(0.05/1.96)^2

n0=(1/3\*v1+1/3\*v2+1/3\*v3)/V

n=n0/(1+(n0/150))

n

nh=n/3

nh

**奈曼分配**

library(sampling)

set.seed(172)

st <- sampling::strata(iris,"Species",rep(10,3),"srswor")

data <- getdata(iris,st)

v1 <- var(data[data$Stratum %in% c(1),]$Sepal.Length)

v2 <- var(data[data$Stratum %in% c(2),]$Sepal.Length)

v3 <- var(data[data$Stratum %in% c(3),]$Sepal.Length)

S<-1/3\*sqrt(v1)+1/3\*sqrt(v2)+1/3\*sqrt(v3)

nn=S^2/(V+(1/3\*v1+1/3\*v2+1/3\*v3)/150)

nn

nn1=1/3\*nn\*sqrt(v1)/S

nn2=1/3\*nn\*sqrt(v2)/S

nn3=1/3\*nn\*sqrt(v3)/S

nn1

nn2

nn3

**（4）**

**分别比估计**

library(sampling)

library(survey)

set.seed(172)

N <- nrow(iris)

N\_h <- table(iris$Species)

W\_h <- N\_h/N

L <- length(unique(iris$Species))

n\_h <- rep(10,L)

st <- sampling::strata(iris,"Species",n\_h,"srswor")#survey包中也有strata函数

data <- getdata(iris,st)

pw <- 1/st$Prob #样本权重

fpc <- rep(1-n\_h/N\_h,c(n\_h))#有限总体修正系数

sampledata <- cbind(data,pw,fpc)

dstrat <- svydesign(ids = ~1,strata = ~Species,weights = ~pw,data = sampledata,fpc = ~fpc)#定义抽样设计

ratio <- svyratio(~Sepal.Length, ~Sepal.Width, dstrat, separate=TRUE)#分别比估计

popm <- data.frame(Width=tapply(iris$Sepal.Width, INDEX=iris$Species, FUN=mean))#辅助变量总体均值

predict(ratio, popm$Width\*W\_h)

**联合比估计**

library(sampling)

library(survey)

set.seed(172)

N <- nrow(iris)

N\_h <- table(iris$Species)

W\_h <- N\_h/N

L <- length(unique(iris$Species))

n\_h <- rep(10,L)

st <- sampling::strata(iris,"Species",n\_h,"srswor")#survey包中也有strata函数

data <- getdata(iris,st)

pw <- 1/st$Prob #样本权重

fpc <- rep(1-n\_h/N\_h,c(n\_h))#有限总体修正系数

sampledata <- cbind(data,pw,fpc)

dstrat <- svydesign(ids = ~1,strata = ~Species,weights = ~pw,data = sampledata,fpc = ~fpc)#定义抽样设计

ratio <- svyratio(~Sepal.Length, ~Sepal.Width, dstrat, separate=FALSE)#联合比估计

popm <- data.frame(Width=mean(iris$Sepal.Width))#辅助变量总体均值

predict(ratio, popm$Width)

**分别回归估计**

library(sampling)

library(survey)

set.seed(172)

N <- nrow(iris)

N\_h <- table(iris$Species)

W\_h <- N\_h/N

L <- length(unique(iris$Species))

n\_h <- rep(10,L)

st <- sampling::strata(iris,"Species",n\_h,"srswor")#survey包中也有strata函数

data <- getdata(iris,st)

pw <- 1/st$Prob #样本权重

fpc <- rep(1-n\_h/N\_h,c(n\_h))#有限总体修正系数

sampledata <- cbind(data,pw,fpc)

#分别比估计

sampledata1 <- sampledata[sampledata$Stratum %in% c(1),]

sampledata2 <- sampledata[sampledata$Stratum %in% c(2),]

sampledata3 <- sampledata[sampledata$Stratum %in% c(3),]

#定义抽样设计

dstrat1 <- svydesign(ids = ~1,strata = ~Species,weights = ~pw,data = sampledata1,fpc = ~fpc)

dstrat2 <- svydesign(ids = ~1,strata = ~Species,weights = ~pw,data = sampledata2,fpc = ~fpc)

dstrat3 <- svydesign(ids = ~1,strata = ~Species,weights = ~pw,data = sampledata3,fpc = ~fpc)

reg1 <- svyglm(Sepal.Length~Sepal.Width,dstrat1)

reg2 <- svyglm(Sepal.Length~Sepal.Width,dstrat2)

reg3 <- svyglm(Sepal.Length~Sepal.Width,dstrat3)

popm <- data.frame(Sepal.Width=tapply(iris$Sepal.Width, INDEX=iris$Species, FUN=mean))#辅助变量总体均值

popm1 <- data.frame(Sepal.Width=popm[1,])

popm2 <- data.frame(Sepal.Width=popm[2,])

popm3 <- data.frame(Sepal.Width=popm[3,])

a<-predict(reg1,popm1)

b<-predict(reg2,popm2)

c<-predict(reg3,popm3)

sum(c(a,b,c)\*W\_h)

**联合回归估计**

library(sampling)

library(survey)

set.seed(172)

N <- nrow(iris)

N\_h <- table(iris$Species)

W\_h <- N\_h/N

L <- length(unique(iris$Species))

n\_h <- rep(10,L)

st <- sampling::strata(iris,"Species",n\_h,"srswor")#survey包中也有strata函数

data <- getdata(iris,st)

pw <- 1/st$Prob #样本权重

fpc <- rep(1-n\_h/N\_h,c(n\_h))#有限总体修正系数

sampledata <- cbind(data,pw,fpc)

dstrat <- svydesign(ids = ~1,strata = ~Species,weights = ~pw,data = sampledata,fpc = ~fpc)#定义抽样设计

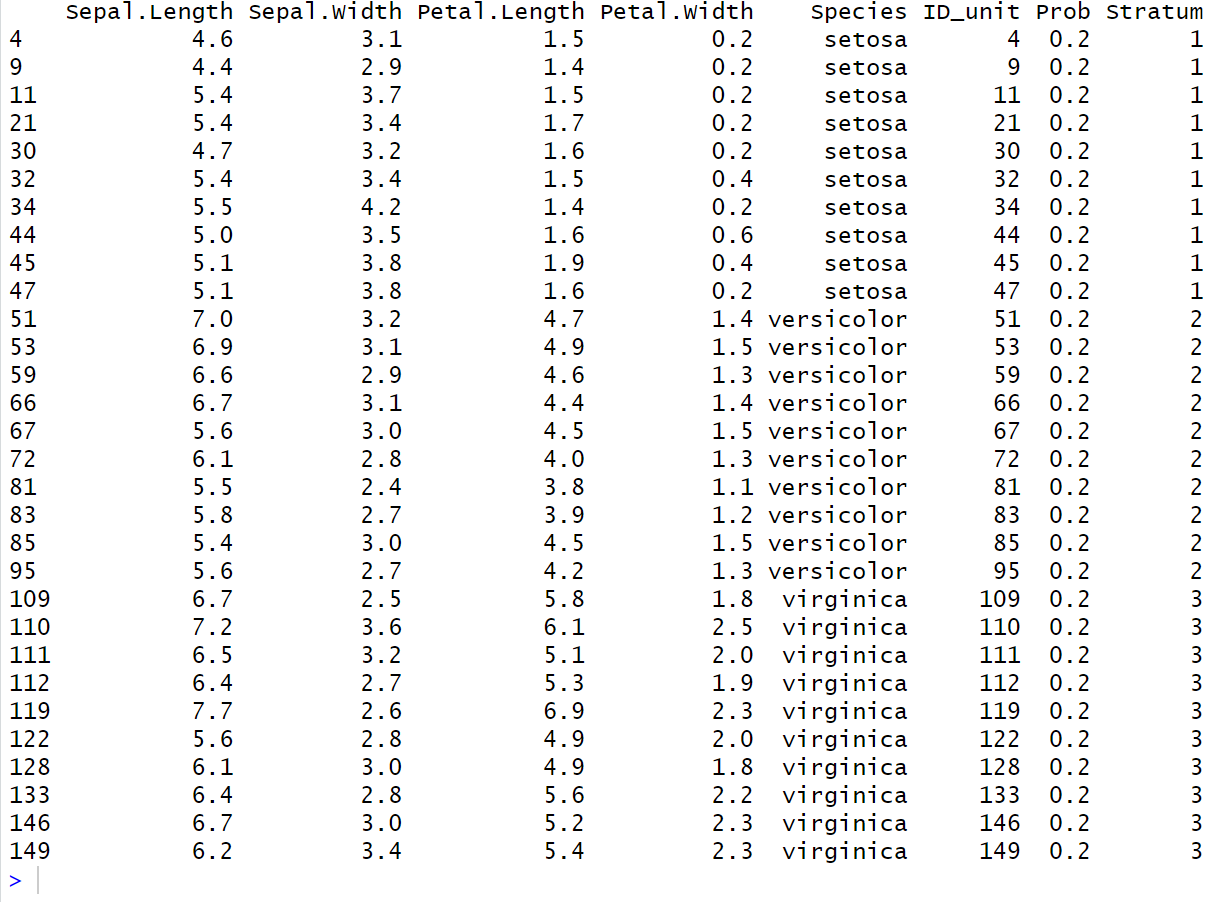
reg <- svyglm(Sepal.Length~Sepal.Width,dstrat)#联合比估计

popm <- data.frame(Sepal.Width=mean(iris$Sepal.Width))#辅助变量总体均值

predict(reg,popm)

**3、程序结果：**

**（1）**



**（2）**

****

**（3）**

**比例分配**

总样本量

****

各层样本量

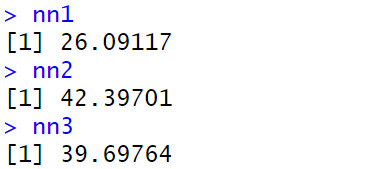
****

**奈曼分配**

总样本量

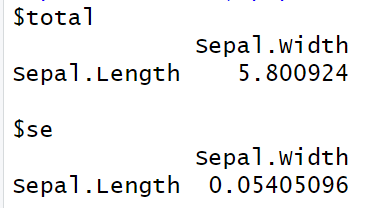
****

各层样本量

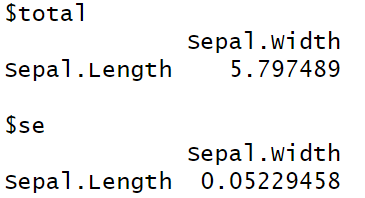
****

**（4）**

**分别比估计**

****

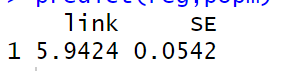
**联合比估计**

****

**分别回归估计**

****

**联合回归估计**

****

**五、实验结果及分析：**

在本次实验中，联合比估计的标准误最小，但各种估计之间的标准误差别不大