irish.r

cules

Wed Aug 23 12:48:53 2017

if((any(grepl("readxl",installed.packages())))==FALSE)  
 install.packages("readxl")  
library(readxl)  
IRIS <- read\_excel("~/DataAnalytics/assignment1/IRIS.xlsx")  
View(IRIS)  
length(IRIS)

## [1] 6

SepalLengthmean<-apply(IRIS["SepalLengthCm"],2,mean)  
SepalWidthmean<-apply(IRIS["SepalWidthCm"],2,mean)  
PetalLengthmean<-apply(IRIS["PetalLengthCm"],2,mean)  
PetalWidthmean<-apply(IRIS["PetalWidthCm"],2,mean)  
cat("Mean Sepal Length=",SepalLengthmean,". Mean Sepal Width=",SepalWidthmean,". Mean Petal Length=",PetalLengthmean,". Mean Petal Width=",PetalWidthmean)

## Mean Sepal Length= 5.843333 . Mean Sepal Width= 3.054 . Mean Petal Length= 3.758667 . Mean Petal Width= 1.198667

#The population variance is should be quite close the true variance as the sample data is very close to population data  
SepalLengthVar<-apply(IRIS["SepalLengthCm"],2,var)  
SepalWidthVar<-apply(IRIS["SepalWidthCm"],2,var)  
PetalLengthVar<-apply(IRIS["PetalLengthCm"],2,var)  
PetalWidthVar<-apply(IRIS["PetalWidthCm"],2,var)  
cat("MSepal Length Variance=",SepalLengthVar,". Sepal Width Variance=",SepalWidthVar,". Petal Length Variance=",PetalLengthVar,". Petal Width Variance=",PetalWidthVar)

## MSepal Length Variance= 0.6856935 . Sepal Width Variance= 0.188004 . Petal Length Variance= 3.113179 . Petal Width Variance= 0.5824143

sample\_data <- IRIS[sample(1:nrow(IRIS), 50,replace=FALSE),]   
SampleSepalLengthVar<-apply(sample\_data["SepalLengthCm"],2,var)  
SampleSepalWidthVar<-apply(sample\_data["SepalWidthCm"],2,var)  
SamplePetalLengthVar<-apply(sample\_data["PetalLengthCm"],2,var)  
SamplePetalWidthVar<-apply(sample\_data["PetalWidthCm"],2,var)  
cat("MSepal Length Variance=",SampleSepalLengthVar,". Sepal Width Variance=",SampleSepalWidthVar,". Petal Length Variance=",SamplePetalLengthVar,". Petal Width Variance=",SamplePetalWidthVar)

## MSepal Length Variance= 0.6476898 . Sepal Width Variance= 0.2465469 . Petal Length Variance= 2.940751 . Petal Width Variance= 0.5846367

var.interval = function(data, conf.level = 0.95) {  
 df = nrow(data) - 1  
 chilower = qchisq((1 - conf.level)/2, df)  
 chiupper = qchisq((1 - conf.level)/2, df, lower.tail = FALSE)  
 v = var(data)  
 c(df \* v/chiupper, df \* v/chilower)  
 }  
var.interval(sample\_data["SepalLengthCm"],0.95)

## [1] 0.4519469 1.0057640