Stat474W/574 Activity 2

Dr. Islam

Winter, Jan 20, 2022

# Objectives

* Review definitions of population and sample variances
* Compute population variance
* Compute sample variance
* Compute pooled sample variance for two populations with common variance
* Make graphical diaplays of data via boxplot and histogram

# 1. Useful definitions

**Definition 1.1**

Given a finite population ,

* the population mean, denoted by , is defined by
* the population variance, denoted by is defined by

**Definition 1.2**

Given a sample from a population with mean and variance ,

* the sample mean, denoted by , is defined by
* the sample variance, denoted by is defined by

**Definition 1.3**

* Given two samples and from two populations with unknown but common variance , an estimate of the common variance is given by

where

is called the pooled sample variance of the unknown but common variance of two populations.

# 2 Functions to be used

* sum()
* mean()
* var()
* sd()
* length()
* par()
* hist()
* boxplot()

# Example 1

The data below refers to a population of circumference of 7 orange trees, measured in mm, at breast height, a standard measurement in forestry.

209 58 203 145 201 51 75

1. Assign this population data values to a vector
2. Find the population mean of circumference of trees ().
3. Find the population variance of circumference of trees ().

Solution

There is no function in R to compute population variance . We can easily compute it via self-coding using known functions:

#(a)  
X<-scan(text="209 58 203 145 201 51 75")  
N=7  
#(b)  
mu=mean(X)  
#(c)  
pop.var<-sum((X-mu)^2)/7

# Example 2

Below is sample of creative writing score of students of a writing class, after receiving intrinsic motivation:

19.3 21.3 22.6 20.3 17.5 22.2 26.7 29.7 20.5 20.6 22.1 18.2 12.0 19.1 23.1

1. Compute sample variance by self-coding in R.
2. The var() function in R computes sample variance. Compute sample variance using var() function.
3. verify that the results in (a) and (b) agree.

x<-scan(text="19.3 21.3 22.6 20.3 17.5 22.2 26.7 29.7 20.5 20.6 22.1 18.2 12.0 19.1 23.1")  
n=length(x)  
xbar<-mean(x)  
#(a)  
samp.var<-sum((x-xbar)^2)/(n-1)  
# the function var() also compute sample variance  
#(b)  
samp.var2=var(x)  
#(c)  
c(samp.var, samp.var2)

[1] 16.16981 16.16981

# Example 3

Given two samples x and y below, x from intrinsic score and y from extrinsic score, as follows:

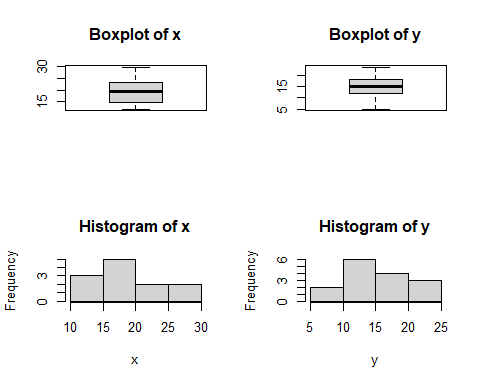
x<-scan(text="16.6 26.7 17.5 24.0 22.2 29.7 19.3 19.1 12.9 12.0 19.8 12.0")  
y<-scan(text="17.2 15.0 6.1 11.8 19.2 16.8 14.8 12.3 12.0 24.0 20.7 17.2 10.9 22.1 5.0")

1. Compute , variance of the sample x.
2. Compute , variance of the sample y.
3. Assume that the two populations of intrinsic and extrinsic scores have unknown but equal variances . Estimate pooled sample variance of the common variance .

x<-scan(text="16.6 26.7 17.5 24.0 22.2 29.7 19.3 19.1 12.9 12.0 19.8 12.0")  
y<-scan(text="17.2 15.0 6.1 11.8 19.2 16.8 14.8 12.3 12.0 24.0 20.7 17.2 10.9 22.1 5.0")  
n1<-length(x)  
n2<-length(y)  
s1sq<-var(x)  
s2sq<-var(y)  
s2pooled<-((n1-1)\*s1sq+(n2-1)\*s2sq)/(n1+n2-2)  
c(s1sq,s2sq,s2pooled)

[1] 31.97970 29.78924 30.75304

par(mfrow=c(2,2))#row-wise fill  
boxplot(x, main="Boxplot of x")  
boxplot(y, main="Boxplot of y")  
hist(x)  
hist(y)



par(mfcol=c(2,2))#column-wise fill  
boxplot(x, horizontal = T, main="Boxplot of x")  
boxplot(y, horizontal = T, main="Boxplot of y")  
hist(x)  
hist(y)

