# Variance

y <- c(13,7,5,12,9,15,6,11,9,7,12)

plot(y,ylim=c(0,20))

# range

range(y)

plot(1:11,y,ylim=c(0,20),pch=16,col="blue")

lines(c(4.5,4.5),c(5,15),col="brown")

lines(c(4.5,3.5),c(5,5),col="brown",lty=2)

lines(c(4.5,5.5),c(15,15),col="brown",lty=2)

# residuals

plot(1:11,y,ylim=c(0,20),pch=16,col="blue")

abline(h=mean(y),col="green")

for (i in 1:11) lines(c(i,i),c(mean(y),y[i]),col="red")

# sum of squares

y - mean(y)

y - mean(y))^2

sum((y - mean(y))^2)

# variance

variance <- function (x) sum((x-mean(x))^2)/(length(x)-1)

variance(y)

var(y)

# ozone example

ozone <- read.csv("c:\\temp\\gardens.csv")

attach(ozone)

ozone

mean(gardenA)

gardenA - mean(gardenA)

(gardenA - mean(gardenA))^2

sum((gardenA - mean(gardenA))^2)

sum((gardenA - mean(gardenA))^2)/9

mean(gardenB)

gardenB - mean(gardenB)

(gardenB - mean(gardenB))^2

sum((gardenB - mean(gardenB))^2)

sum((gardenB - mean(gardenB))^2)/9

mean(gardenC)

gardenC - mean(gardenC)

(gardenC - mean(gardenC))^2

sum((gardenC - mean(gardenC))^2)

sum((gardenC - mean(gardenC))^2)/9

var(gardenC)/var(gardenB)

# critical value of Fisher’s F

2\*(1 - pf(10.667,9,9))

var.test(gardenB,gardenC)

# sample size

plot(c(0,32),c(0,15),type="n",xlab="Sample size",ylab="Variance")

for (df in seq(3,31,2)) {

for( i in 1:30){

x <- rnorm(df,mean=10,sd=2)

points(df,var(x)) }}

# standard error of a mean

sqrt(var(gardenA)/10)

sqrt(var(gardenB)/10)

sqrt(var(gardenC)/10)

# quantiles of the t distribution

qt(.025,9)

qt(.975,9)

qt(.995,9)

qt(.9975,9)

qt(.975,9)\*sqrt(1.33333/10)

# bootstrap intervals

data <- read.csv("c:\\temp\\skewdata.csv")

attach(data)

names(data)

plot(c(0,30),c(0,60),type="n",xlab="Sample size",

ylab="Confidence interval")

for (k in seq(5,30,3)){

a <- numeric(10000)

for (i in 1:10000){

a[i] <- mean(sample(values,k,replace=T))

}

points(c(k,k),quantile(a,c(.025,.975)),type="b",pch=21,bg="red")

}

quantile(a,c(.025,.975))

xv <- seq(5,30,0.1)

yv <- mean(values)+1.96\*sqrt(var(values)/xv)

lines(xv,yv,col="blue")

yv <- mean(values)-1.96\*sqrt(var(values)/xv)

lines(xv,yv,col="blue")

yv <- mean(values)-qt(.975,xv)\*sqrt(var(values)/xv)

lines(xv,yv,lty=2,col="green")

yv <- mean(values)+qt(.975,xv)\*sqrt(var(values)/xv)

lines(xv,yv,lty=2,col="green")