# single samples

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

attach(data)

names(data)

summary(y)

boxplot(y)

hist(y)

# rug plot

length(table(y))

plot(range(y),c(0,10),type="n",xlab="y values",ylab="")

for (i in 1:100) lines(c(y[i],y[i]),c(0,1),col="blue")

# designing a histogram

(max(y)-min(y))/10

diff(range(y))/11

# the game of craps

score <- 2:12

ways <- c(1,2,3,4,5,6,5,4,3,2,1)

( game <- rep(score,ways) )

sample(game,1)

outcome <- numeric(10000)

for (i in 1:10000) outcome[i] <- sample(game,1)

hist(outcome,breaks=(1.5:12.5))

mean.score <- numeric(10000)

for (i in 1:10000) mean.score[i] <- mean(sample(game,3))

hist(mean.score,breaks=(1.5:12.5))

mean(mean.score)

sd(mean.score)

xv <- seq(2,12,0.1)

yv <- 10000\*dnorm(xv,mean(mean.score),sd(mean.score))

hist(mean.score,breaks=(1.5:12.5),ylim=c(0,3000),col="yellow", main="")

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

# standard normal disribution

standard.deviations <- seq(-3,3,0.01)

pd <- dnorm(standard.deviations)

plot(standard.deviations,pd,type="l",col="blue")

pnorm(-2)

pnorm(-1)

1-pnorm(3)

qnorm(c(0.025,0.975))

# shading the tails of the standard normal distribution

xv<-seq(-3,3,0.01)

yv<-dnorm(xv)

plot(c(-3,3),c(0,0.3),xlim=c(-3,3),ylim=c(0,0.4),type="n",ylab="pd",xlab="standard deviations")

polygon(c(1.96,1.96,-1.96,-1.96,xv[105:496]),c(yv[496],0,0,yv[105],yv[105:496]),col="green")

polygon(c(-1.96,-1.96,xv[1],xv[1:104]),c(yv[104],0,0,yv[1:104]),col="red")

polygon(c(xv[601],xv[601],1.96,1.96,xv[497:601]),c(yv[601],0,0,yv[496:601]),col="red")

text(0,0.2,"95%",cex=2)

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

# calculations with the sandard normal distribution

ht <- seq(150,190,0.01)

plot(ht,dnorm(ht,170,8),type="l",col="brown",

ylab="Probability density",xlab="Height")

pnorm(-1.25)

pnorm(1.875)

1 - pnorm(1.875)

pnorm(1.25) - pnorm(-0.625)

# drawing a panel of four normal distributions

par(mfrow=c(2,2))

ht <- seq(150,190,0.01)

pd <- dnorm(ht,170,8)

plot(ht,dnorm(ht,170,8),type="l",col="brown",

ylab="Probability density",xlab="Height")

plot(ht,dnorm(ht,170,8),type="l",col="brown",

ylab="Probability density",xlab="Height")

yv <- pd[ht<=160]

xv <- ht[ht<=160]

xv <- c(xv,160,150)

yv <- c(yv,yv[1],yv[1])

polygon(xv,yv,col="orange")

plot(ht,dnorm(ht,170,8),type="l",col="brown",

ylab="Probability density",xlab="Height")

xv <- ht[ht>=185]

yv <- pd[ht>=185]

xv <- c(xv,190,185)

yv <- c(yv,yv[501],yv[501])

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

plot(ht,dnorm(ht,170,8),type="l",col="brown",

ylab="Probability density",xlab="Height")

xv <- ht[ht>=160 & ht <= 180]

yv <- pd[ht>=160 & ht <= 180]

xv <- c(xv,180,160)

yv <- c(yv,pd[1],pd[1])

polygon(xv,yv,col="green")

# plots for skewness

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

attach(data)

qqnorm(values)

qqline(values,lty=2)

# speed of light data

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

attach(light)

names(light)

hist(speed)

summary(speed)

wilcox.test(speed,mu=990)

a <- numeric(10000)

for(i in 1:10000) a[i] <- mean(sample(speed,replace=T))

hist(a)

# student’s t distribution

plot(c(0,30),c(0,10),type="n",

xlab="Degrees of freedom",ylab="Students t value")

lines(1:30,qt(0.975,df=1:30),col="red")

abline(h=1.96,lty=2,col="green")

xvs <- seq(-4,4,0.01)

plot(xvs,dnorm(xvs),type="l",

ylab="Probability density",xlab="Deviates")

lines(xvs,dt(xvs,df=5),col="red")

qt(0.975,5)

# skewness

skew <- function(x){

m3 <- sum((x-mean(x))^3)/length(x)

s3 <- sqrt(var(x))^3

m3/s3 }

hist(values,main="",col="green")

skew(values)

skew(values)/sqrt(6/length(values))

1 - pt(2.949,28)

skew(sqrt(values))/sqrt(6/length(values))

skew(log(values))/sqrt(6/length(values))

# kurtosis

kurtosis <- function(x) {

m4 <- sum((x-mean(x))^4)/length(x)

s4 <- var(x)^2

m4/s4 - 3 }

kurtosis(values)

kurtosis(values)/sqrt(24/length(values))