# problem set 1.1 -- playing with probability distributions.

# the symbols formerly known as the pound sign tell R that what follows are comments, so R will ignore these.

# Please ALWAYS read ALL the commented text

# 0.

# open R : find programs, departmental, R

# then click on “file – new” to create an R document – we’ll do this together in lab

# Copy and paste **ALL of this** file into an R document window.

# 1. run the following chunk (to #stop here – it’s quite long) by selecting all of it and then pressing <ctrl> R (if using a pc) or

<ctrl> <enter> (if using a mac)

# this is a function that we will use

# R is based on the use of functions you will get used to the syntax very quickly

#####################################################################

# gnorm plots normal curves and computes and displays probabilities #

# two parameters are necessary: the mean mu and standard #

# deviation sigma. Others options are available but optional only. #

#####################################################################

gnorm = function(mu, sigma,a=NA,b=NA,calcProb=!all(is.na(c(a,b))),quantile=NA,calcQuant=!is.na(quantile))

{

values = seq(-1,1,.005) \* 4 \* sigma + mu

probs = dnorm(values, mu, sigma)

plot(values, probs, axes = F, type = "n", xlab = "Possible Values",

ylab = "Probability Density",

main = substitute(paste("Normal Distribution with ",mu == m,", ",sigma == s),list(m=mu,s=sigma)))

axis(1, pos = 0)

abline(0,0,col=1)

lines(values, probs, col = 2)

lo = mu - 4 \* sigma

hi = mu + 4 \* sigma

h = dnorm(mu,mu,sigma)

cex=0.8

if(calcProb==T)

{

if(!is.na(a) && !is.na(b) && a > b){

d = a; a = b; b = d

}

if(is.na(a) || a <= lo){ ulo = lo }

else if(a <= hi){ ulo = a }

else { ulo = hi }

if(is.na(b) || b >= hi){ uhi = hi }

else if(b >= lo){ uhi = b }

else { uhi = lo }

u = seq(ulo,uhi,length=601)

lines(u,dnorm(u,mu,sigma),type="h",col=2)

if(!is.na(a) && !is.na(b)){

text(mu - 3.9 \* sigma, 0.8 \* h,

paste("P( ",a," < X < ",b," ) = ",

round(pnorm(b,mu,sigma)-pnorm(a,mu,sigma),digits=4),sep=""),

adj=0,col=4,cex=cex)

text(mu - 3.9 \* sigma, 0.6 \* h,

paste("P( X < ",a," ) = ",

round(pnorm(a,mu,sigma),digits=4),sep=""),adj=0,col=4,cex=cex)

text(mu + 3.9 \* sigma, 0.5 \* h,

paste("P( X > ",b," ) = ",

round(1-pnorm(b,mu,sigma),digits=4),sep=""),adj=1,col=4,cex=cex)

}

else if(!is.na(a) && is.na(b)){

text(mu - 3.9 \* sigma, 0.6 \* h,

paste("P( X < ",a," ) = ",

round(pnorm(a,mu,sigma),digits=4),sep=""),adj=0,col=4,cex=cex)

text(mu + 3.9 \* sigma, 0.5 \* h,

paste("P( X > ",a," ) = ",

round(1-pnorm(a,mu,sigma),digits=4),sep=""),adj=1,col=4,cex=cex)

}

else if(is.na(a) && !is.na(b)){

text(mu - 3.9 \* sigma, 0.6 \* h,

paste("P( X < ",b," ) = ",

round(pnorm(b,mu,sigma),digits=4),sep=""),adj=0,col=4,cex=cex)

text(mu + 3.9 \* sigma, 0.5 \* h,

paste("P( X > ",b," ) = ",

round(1-pnorm(b,mu,sigma),digits=4),sep=""),adj=1,col=4,cex=cex)

}

}

else if(calcQuant==T)

{

zoffset = -0.02

if( quantile <= 0 || quantile >= 1) quantile = 0.5

x = qnorm(quantile,mu,sigma)

if( x > lo && x < hi)

{

u = seq(lo,x,length=601)

lines(u,dnorm(u,mu,sigma),type="h",col=2)

text(x, zoffset \* h,

paste("z = ",round(qnorm(quantile),2),sep=""),adj=0.5,col=4,cex=cex)

}

else if(x >= hi)

{

u = seq(lo,hi,length=601)

lines(u,dnorm(u,mu,sigma),type="h",col=2)

text(hi, zoffset \* h,

paste("z = ",round(qnorm(quantile),2),sep=""),adj=0.5,col=4,cex=cex)

}

else if( x <= lo)

{

text(lo, zoffset \* h,

paste("z = ",round(qnorm(quantile),2),sep=""),adj=0.5,col=4,cex=cex)

}

text(mu - 3.9 \* sigma, 0.5 \* h,

paste("P( X < ",signif(x,4)," ) = ",

round(quantile,digits=4),sep=""),adj=0,col=4,cex=cex)

}

return(invisible())

}

#stop here

# 3. use the following code to make a figure, copy and paste it into this word document, then try to explain what it means. (Please note that you may have to reshape the graphic window to make the graphic easy to read.):

gnorm(mu=100, sigma=15)

**mean = 100, normal distribution**



# 4. do the same for this one:

gnorm(mu=100, sigma=15, a = 90, b=110)

**49.5% of values fall within the interval of 90-110**



# 5. do the same for this one:

gnorm(mu=100, sigma=15, a = 80, b=120)

**81.76% of values fall between 80-120**



# 6. do the same:

gnorm(mu=100, sigma=15, a = 75)

**95.22% of values are greater than 75**



# 7. do the same:

gnorm(mu=100, sigma=15, a = 130)

**97.71% of values are less than 130**



# 8. can you relate the answer to #7 to the concept of a p-value?

**There is a 97% confidence that the number will be less than 130.**

# 9. what is the probability that 1 randomly selected person will have a value greater than 100?

**50%**

# 10. what is the probability that 30 randomly selected people will all have values greater than 100? Do you remember how to combine probabilities?

**.5\*30 = 15%**