

Statistics for Engineers (MAT2001)- Lab

Experiment-V: Normal distribution

1 Normal Distribution

R has four in built functions to generate normal distribution. They are described below.

- `dnorm(x, mean, sd)`
- `pnorm(x, mean, sd)`
- `qnorm(p, mean, sd)`
- `rnorm(n, mean, sd)`

Following is the description of the parameters used in above functions: `x` is a vector of numbers.

`p` is a vector of probabilities.

`n` is number of observations(sample size).

`mean` is the mean value of the sample data. It's default value is zero.

`sd` is the standard deviation. It's default value is 1

1.1 Normal distribution computations and graphs

`dnorm()` : This function gives height of the probability distribution at each point for a given mean and standard deviation.

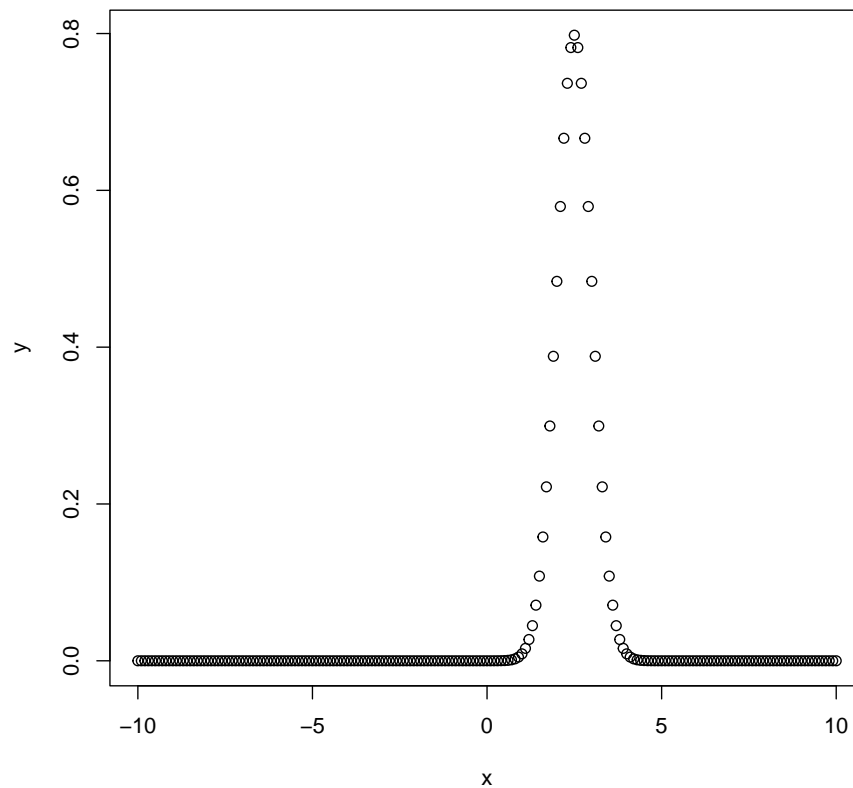
Problem:

Create a sequence of numbers between -10 and 10 incrementing by 0.1 then find the probability distribution for mean as 2.5 and standard deviation as 0.5.

```
x <- seq(-10,10,by=.1)
x
##      [1] -10.0 -9.9 -9.8 -9.7 -9.6 -9.5 -9.4 -9.3 -9.2 -9.1 -9.0 -8.9
##     [13] -8.8 -8.7 -8.6 -8.5 -8.4 -8.3 -8.2 -8.1 -8.0 -7.9 -7.8 -7.7
##     [25] -7.6 -7.5 -7.4 -7.3 -7.2 -7.1 -7.0 -6.9 -6.8 -6.7 -6.6 -6.5
##     [37] -6.4 -6.3 -6.2 -6.1 -6.0 -5.9 -5.8 -5.7 -5.6 -5.5 -5.4 -5.3
```

```
## [49] -5.2 -5.1 -5.0 -4.9 -4.8 -4.7 -4.6 -4.5 -4.4 -4.3 -4.2 -4.1
## [61] -4.0 -3.9 -3.8 -3.7 -3.6 -3.5 -3.4 -3.3 -3.2 -3.1 -3.0 -2.9
## [73] -2.8 -2.7 -2.6 -2.5 -2.4 -2.3 -2.2 -2.1 -2.0 -1.9 -1.8 -1.7
## [85] -1.6 -1.5 -1.4 -1.3 -1.2 -1.1 -1.0 -0.9 -0.8 -0.7 -0.6 -0.5
## [97] -0.4 -0.3 -0.2 -0.1  0.0  0.1  0.2  0.3  0.4  0.5  0.6  0.7
## [109]  0.8  0.9  1.0  1.1  1.2  1.3  1.4  1.5  1.6  1.7  1.8  1.9
## [121]  2.0  2.1  2.2  2.3  2.4  2.5  2.6  2.7  2.8  2.9  3.0  3.1
## [133]  3.2  3.3  3.4  3.5  3.6  3.7  3.8  3.9  4.0  4.1  4.2  4.3
## [145]  4.4  4.5  4.6  4.7  4.8  4.9  5.0  5.1  5.2  5.3  5.4  5.5
## [157]  5.6  5.7  5.8  5.9  6.0  6.1  6.2  6.3  6.4  6.5  6.6  6.7
## [169]  6.8  6.9  7.0  7.1  7.2  7.3  7.4  7.5  7.6  7.7  7.8  7.9
## [181]  8.0  8.1  8.2  8.3  8.4  8.5  8.6  8.7  8.8  8.9  9.0  9.1
## [193]  9.2  9.3  9.4  9.5  9.6  9.7  9.8  9.9 10.0

y <- dnorm(x, mean= 2.5, sd = 0.5)
plot(x,y)
```

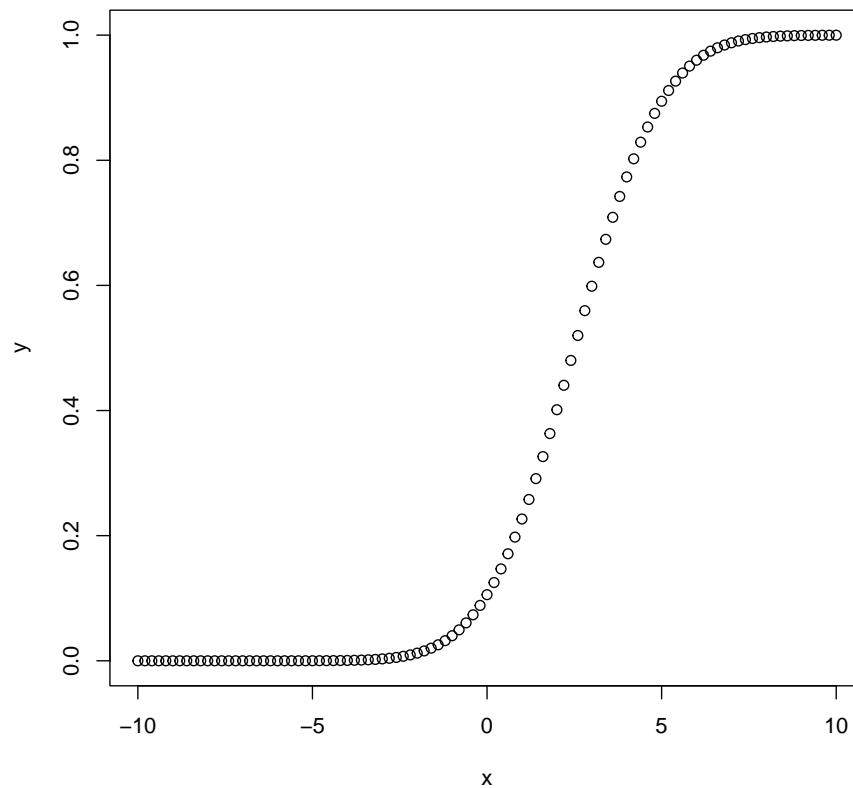


`pnorm()`: This function gives the probability of a normally distributed random number to be less than the value of a given number. It is also called "Cumulative Distribution Function".

Problem:

Create a sequence of numbers between -10 and 10 incrementing by 0.1 then find the cumulative distribution for mean as 2.5 and standard deviation as 2.

```
x <- seq(-10,10,by=.2)
y <- pnorm(x,mean=2.5,sd = 2)
plot(x,y)
```

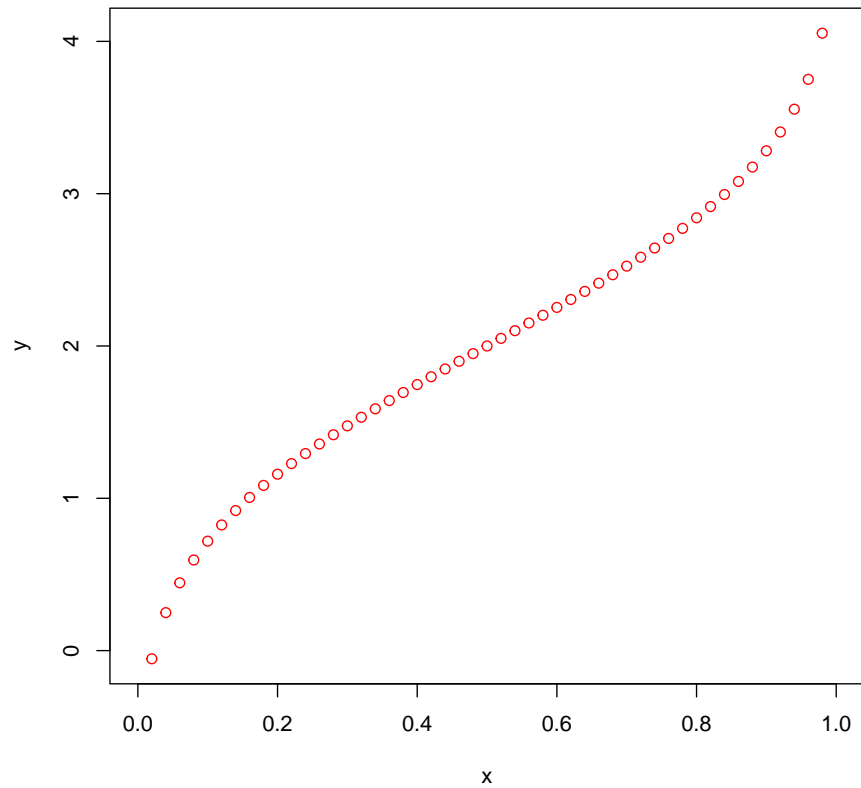


qnorm():- This function takes the probability value and gives a number whose cumulative value matches the probability value.

Problem:

Create a sequence of probability values incrementing by 0.02. Give the cumulative value matches for mean as 2 and standard deviation as 3.

```
x <- seq(0,1,by=0.02)
y <- qnorm(x,mean=2,sd=1)
plot(x,y, col="red")
```

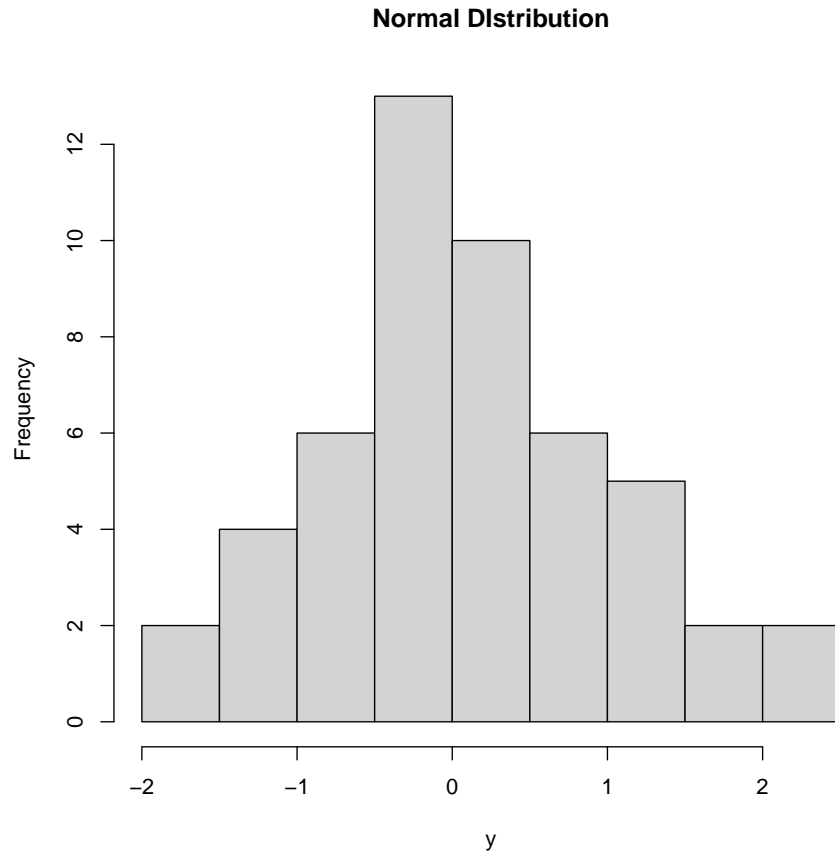


`rnorm()` This function is used to generate random numbers whose distribution is normal. It takes the sample size as input and generates that many random numbers. We draw a histogram to show the distribution of the generated numbers.

Problem:

Create a sample of 50 numbers which are normally distributed.

```
y <- rnorm(50)
hist(y, main = "Normal DIstribution")
```

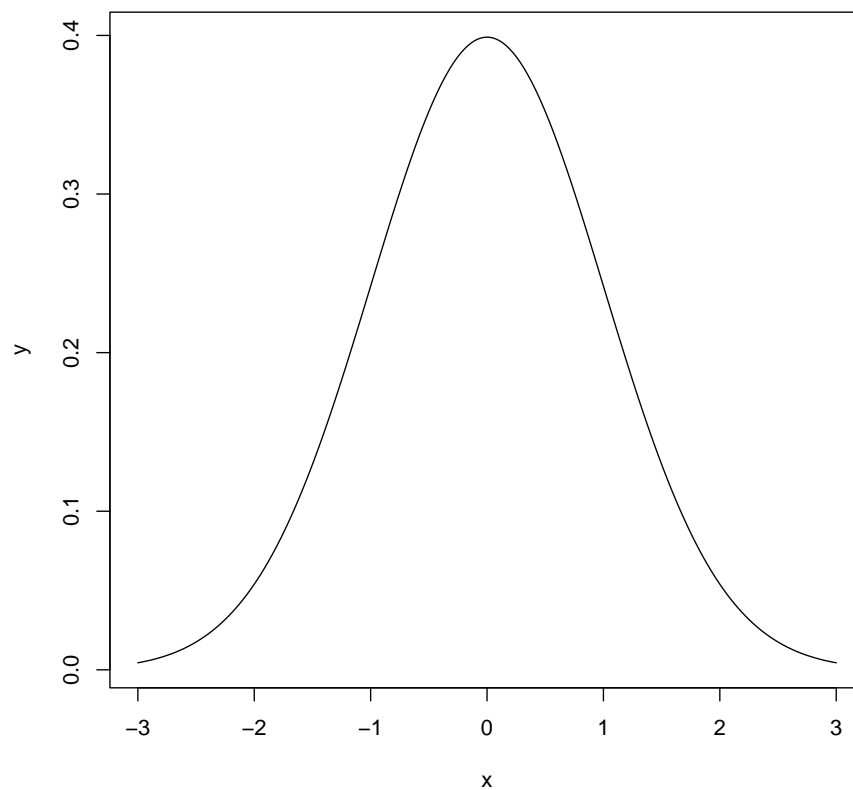


1.2 Standard Normal Probability Distribution Plotting and Finding the Area

Problem:

create a standard normal distribution sequence of 200 numbers, beginning at $x=-3$ and ending at $x=3$

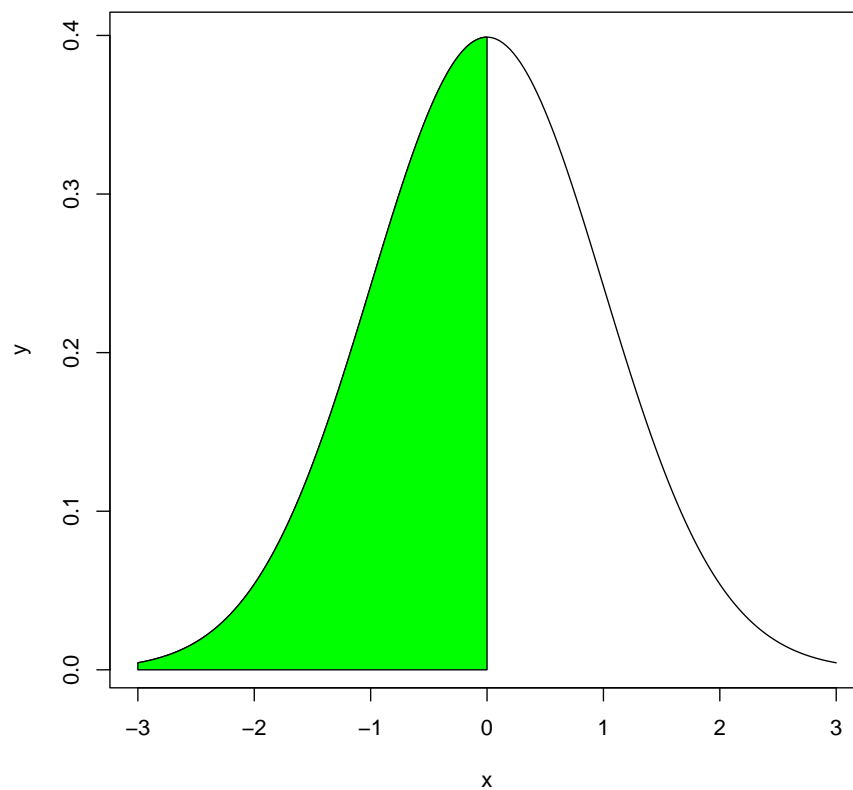
```
x=seq(-3,3,length=200)
y=dnorm(x,mean=0,sd=1)
plot(x,y,type="l")
```



Problem:

Find the area under the curve to left of the mean then find the area

```
x=seq(-3,3,length=200)
y=dnorm(x,mean=0,sd=1)
plot(x,y,type="l")
x=seq(-3,0,length=100)
y=dnorm(x,mean=0,sd=1)
polygon(c(-3,x,0),c(0,y,0),col="green")
```



```
pnorm(0,mean=0,sd=1)
```

```
## [1] 0.5
```

```
#
```

Problem:

Find the area to the left of 1. First, draw an image, then compute the area.

```
x=seq(-3,3,length=200)
```

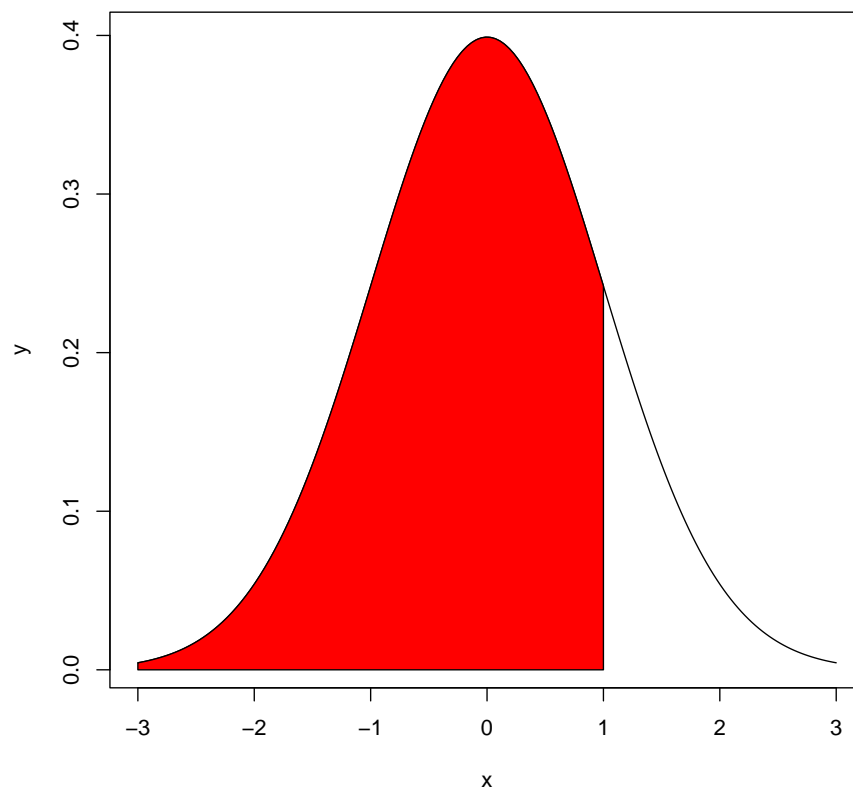
```
y=dnorm(x,mean=0,sd=1)
```

```
plot(x,y,type="l")
```

```
x=seq(-3,1,length=100)
```

```
y=dnorm(x,mean=0,sd=1)
```

```
polygon(c(-3,x,1),c(0,y,0),col="red")
```

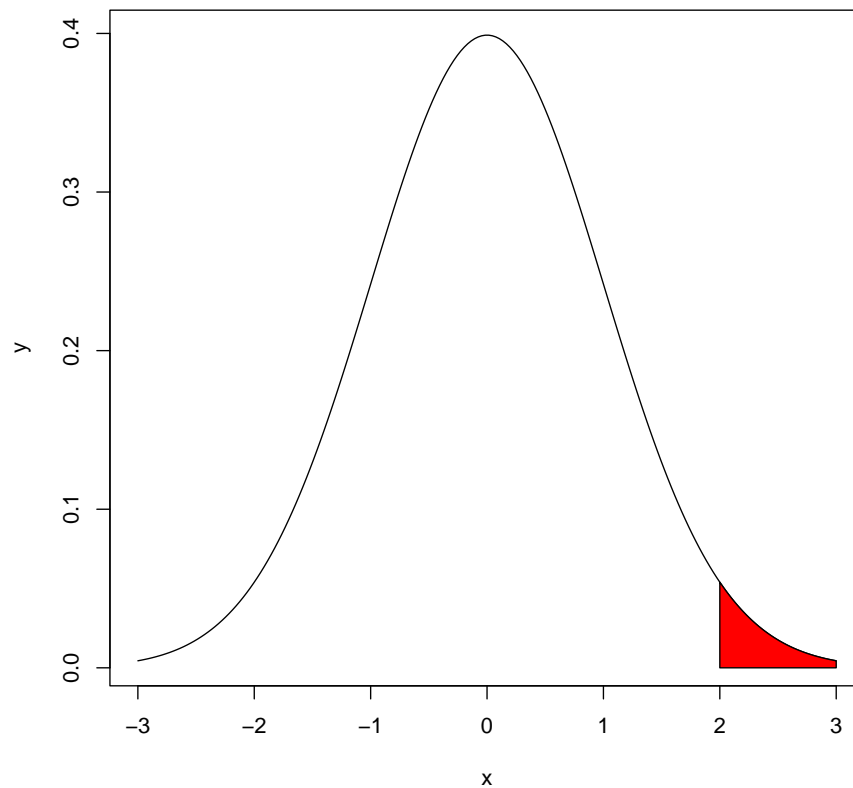
```
pnorm(1,mean=0,sd=1)
```

```
## [1] 0.8413447
```

Problem:

Get the area to the right of 2. First, draw an image, then compute the area.

```
x=seq(-3,3,length=200)
y=dnorm(x,mean=0,sd=1)
plot(x,y,type="l")
x=seq(2,3,length=100)
y=dnorm(x,mean=0,sd=1)
polygon(c(2,x,3),c(0,y,0),col="red")
```



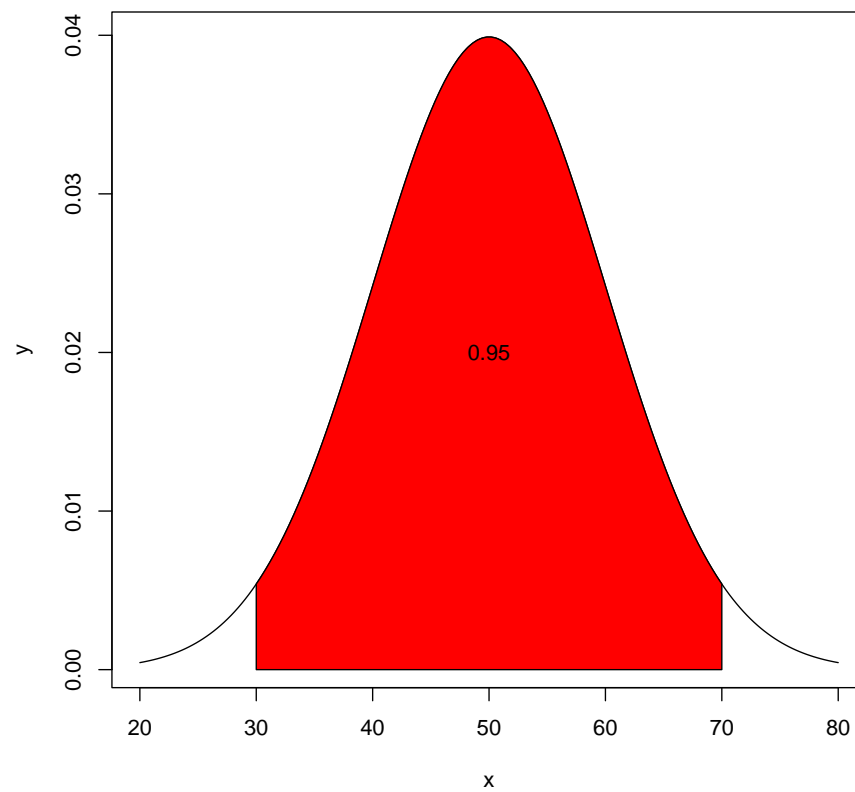
```
1-pnorm(2,mean=0,sd=1)
## [1] 0.02275013
```

Problem:

Use the pnorm command to find areas under the normal density curve, regardless of the mean and standard deviation values. Example, mean=50 and standard deviation=10.

```
x=seq(20,80,length=200)
y=dnorm(x,mean=50,sd=10)
plot(x,y,type="l")
x=seq(30,70,length=100)
y=dnorm(x,mean=50,sd=10)
polygon(c(30,x,70),c(0,y,0),col="red")
pnm <- pnorm(70,mean=50,sd=10)-pnorm(30,mean=50,sd=10)
```

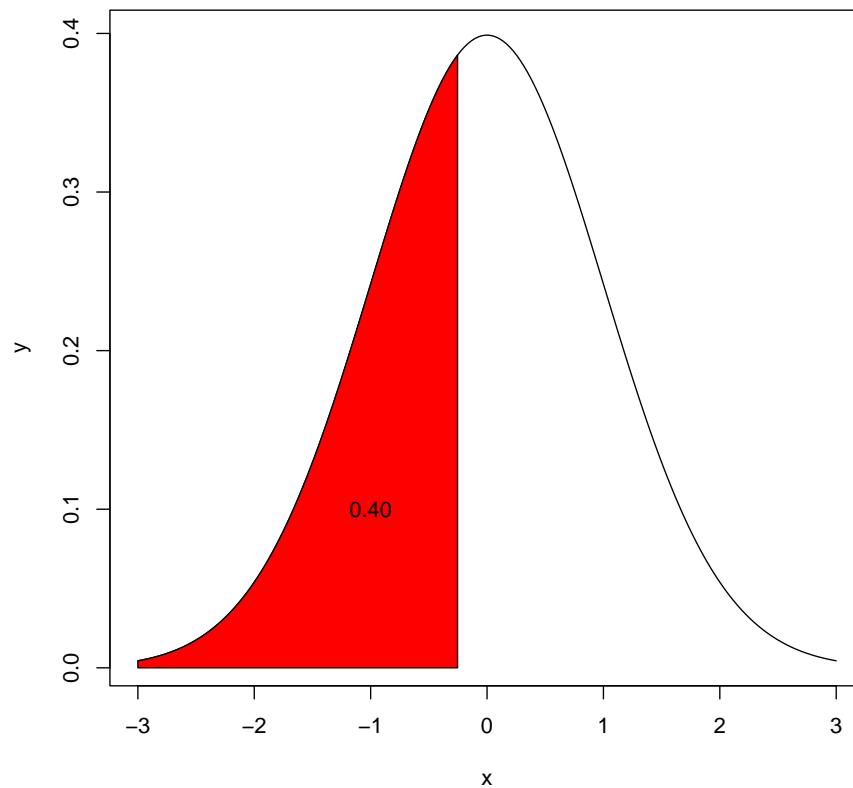
```
text(50, 0.02, round(pnm, 2))
```



Finding the Quantile (Percentile)

```
x=seq(-3,3,length=200)
y=dnorm(x,mean=0,sd=1)
plot(x,y,type="l")
x=seq(-3,-0.2533,length=100)

y=dnorm(x,mean=0,sd=1)
polygon(c(-3,x,-0.2533),c(0,y,0),col="red")
text(-1,0.1,"0.40")
```



```
qnorm(0.40,mean=0,sd=1)
```

```
## [1] -0.2533471
```

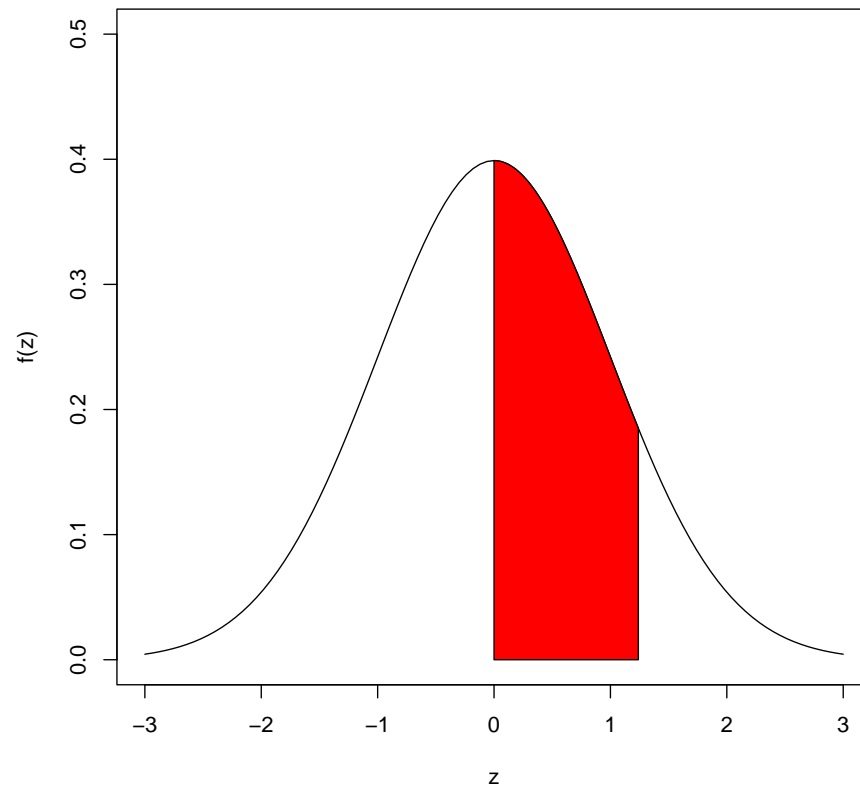
```
pnorm(-0.2533, 0, 1)
```

```
## [1] 0.4000182
```

Find $P(0 < Z < 1.24)$

```
plot.new()
curve(dnorm, xlim = c(-3, 3), ylim = c(0, 0.5), xlab = "z",
      ylab="f(z)")
zleft = 0
zright = 1.24
x = c(zleft, seq(zleft, zright, by=.001), zright)
```

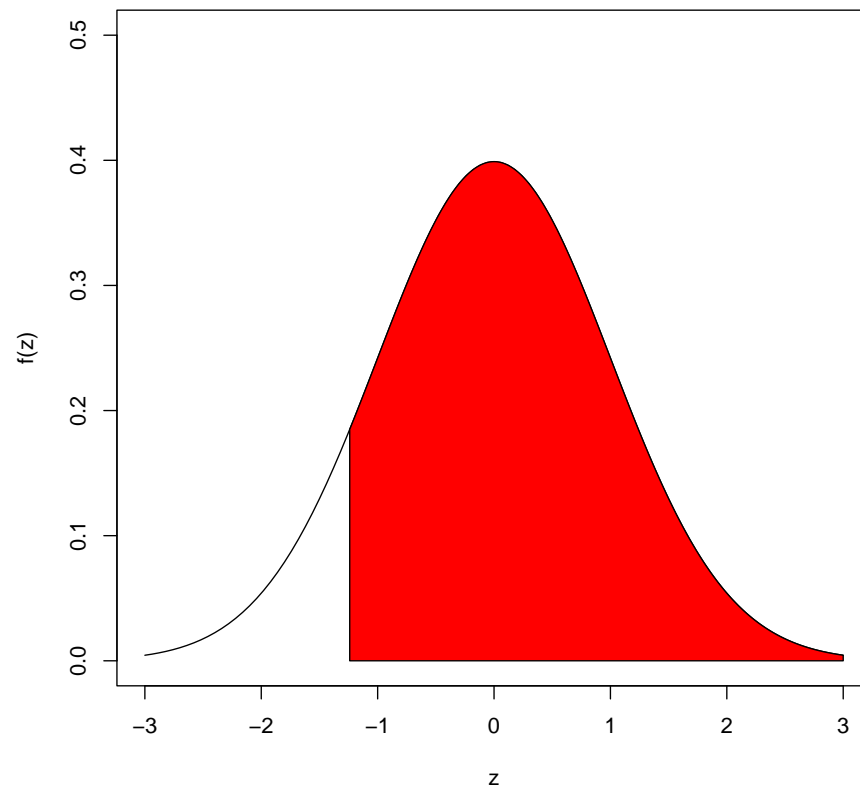
```
y = c(0, dnorm(seq(zleft, zright, by=.001)), 0)
polygon(x, y, col="red")
```



Find $P(Z > -1.24)$

```
1 - pnorm(-1.24)
## [1] 0.8925123

plot.new()
curve(dnorm, xlim = c(-3, 3), ylim = c(0, 0.5), xlab = "z",
      ylab="f(z)")
z = -1.24
x = c(z, seq(z, 3, by=.001), 3)
y = c(0, dnorm(seq(z, 3, by=.001)), 0)
polygon(x, y, col="red")
```

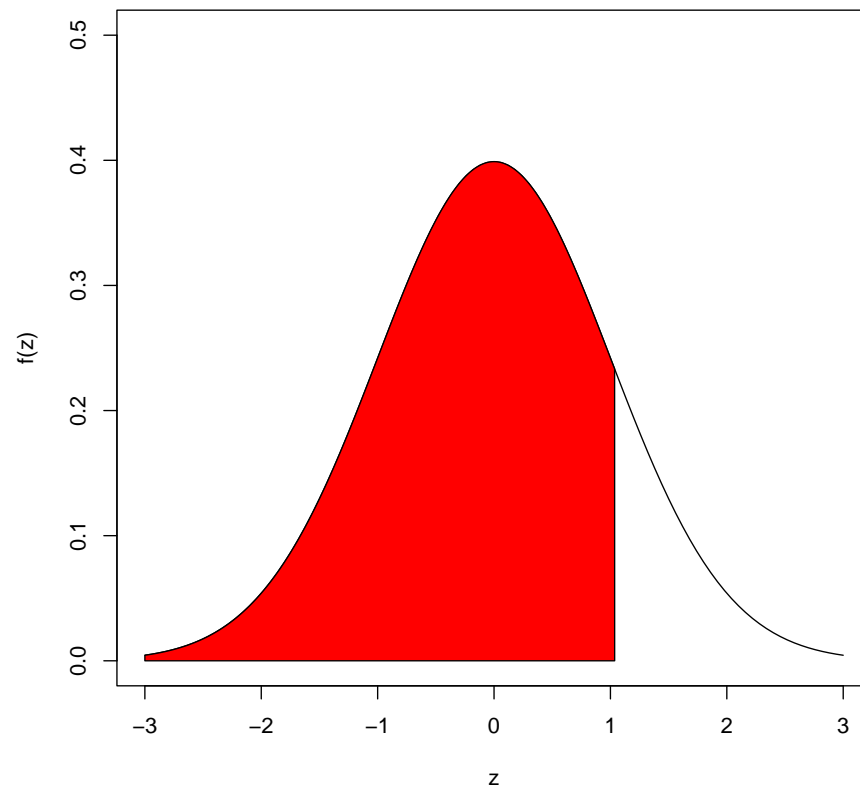


Find the 85th percentile of the standard normal “z” distribution.

```
qnorm(0.85)

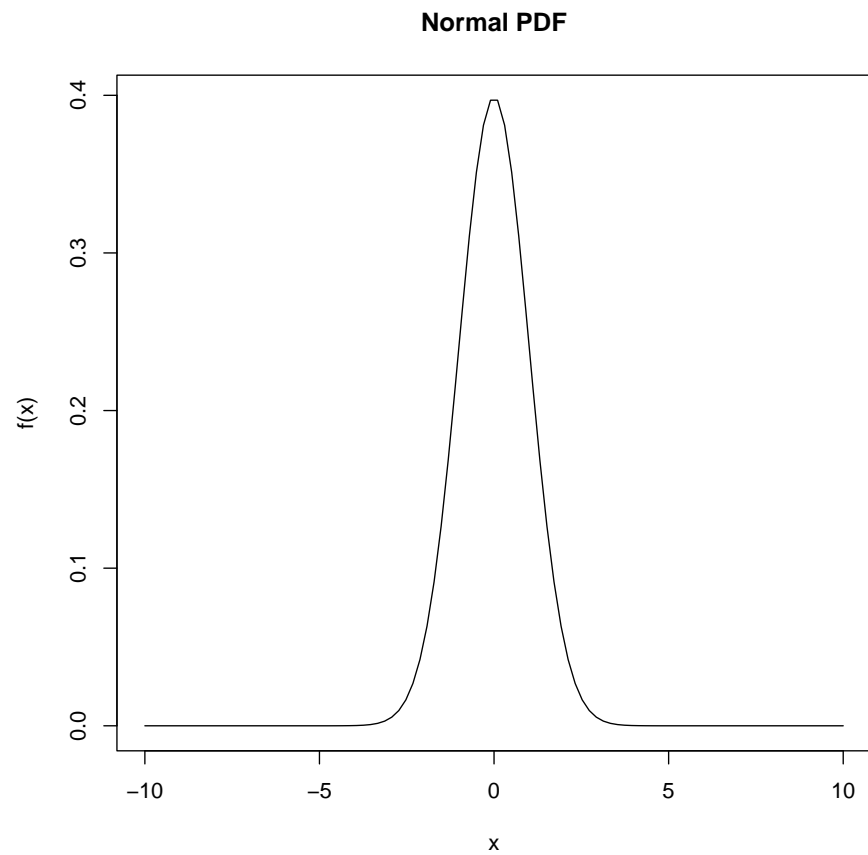
## [1] 1.036433

plot.new()
curve(dnorm, xlim = c(-3, 3), ylim = c(0, 0.5), xlab = "z",
      ylab="f(z)")
prob = 0.85
x = c(-3, seq(-3, qnorm(prob), by=.001), qnorm(prob))
y = c(0, dnorm(seq(-3, qnorm(prob), by=.001)), 0)
polygon(x, y, col="red")
```

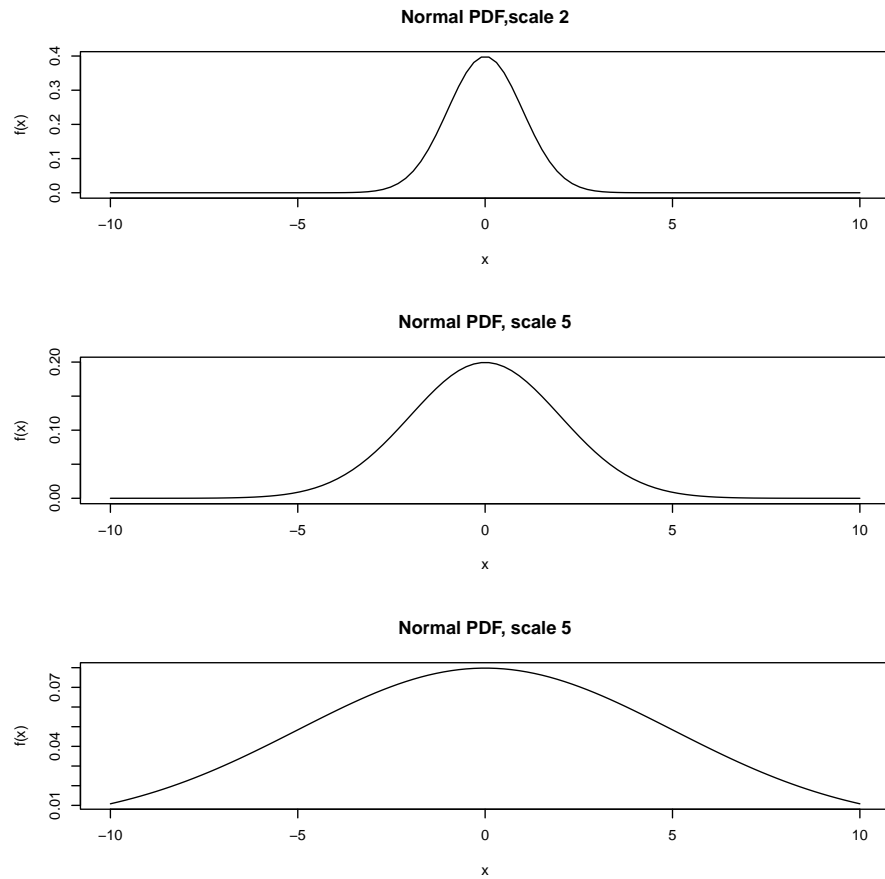


Plot a normal density for a range of x from -10 to 10 with mean 0 and standard deviation 1 :(This Problem Explains types of kurtosis by changing standard deviation)

```
x<-seq(-10,10,length=100)
plot(x,dnorm(x,0,1),xlab="x", ylab="f(x)", type='l', main="Normal PDF")
```

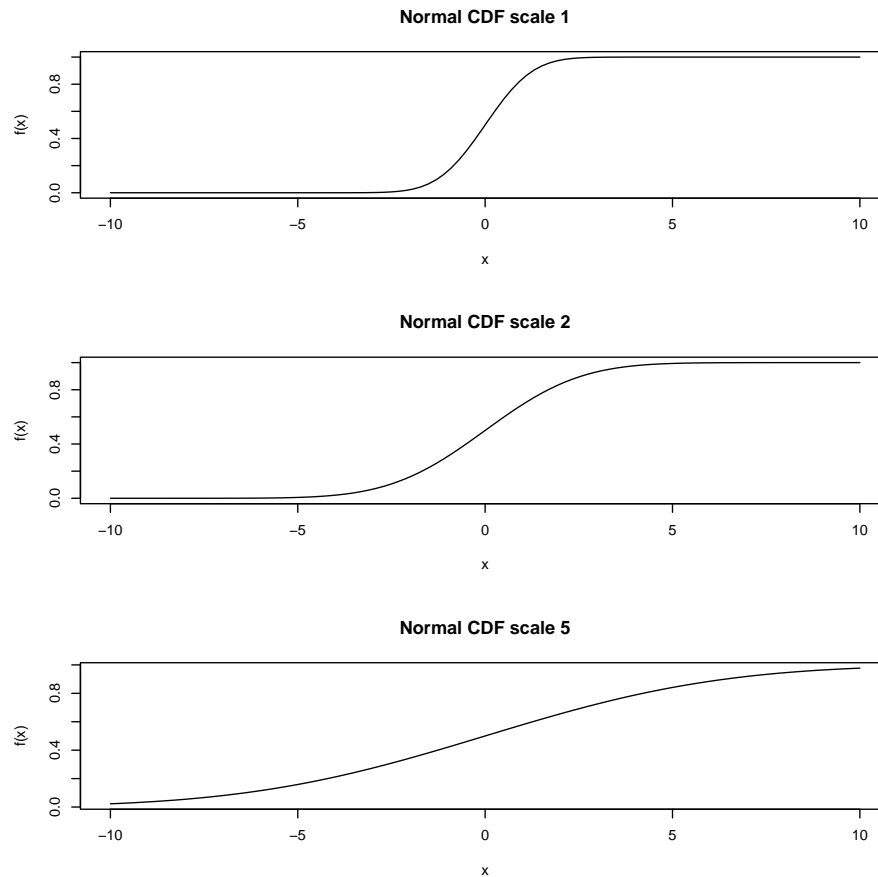


```
par(mfrow=c(3,1))
plot(x,dnorm(x,0,1),xlab="x",ylab="f(x)", type='l', main="Normal PDF,scale 2")
plot(x,dnorm(x,0,2),xlab="x",ylab="f(x)", type='l',main="Normal PDF, scale 5")
plot(x,dnorm(x,0,5),xlab="x",ylab="f(x)", type='l',main="Normal PDF, scale 5")
```

Normal distribution Cumulative Distribution Function with different scale parameters

```
par(mfrow=c(3,1))
plot(x,pnorm(x,0,1),xlab="x",ylab="f(x)", type='l', main="Normal CDF scale 1")
plot(x,pnorm(x,0,2),xlab="x", ylab="f(x)", type='l', main="Normal CDF scale 2")
plot(x,pnorm(x,0,5),xlab="x", ylab="f(x)", type='l', main="Normal CDF scale 5")
```



In a photographic process the developing times of prints may be looked upon as a random variable having the normal distribution with a mean of 16.28 seconds and a standard deviation 0.12 second. Find the probability that it will take (i) Atleast 16.20 seconds to develop one of the prints; (ii) atleast 16.35 seconds to develop one of the prints

Solution: i) $P(X \geq 16.20) = P(Z \geq -0.6667)$

ii) $P(X \leq 16.35) = P(X \leq 0.5833) = 0.5 + 0.2190 = 0.7190$

```
(1-pnorm (-0.6667))

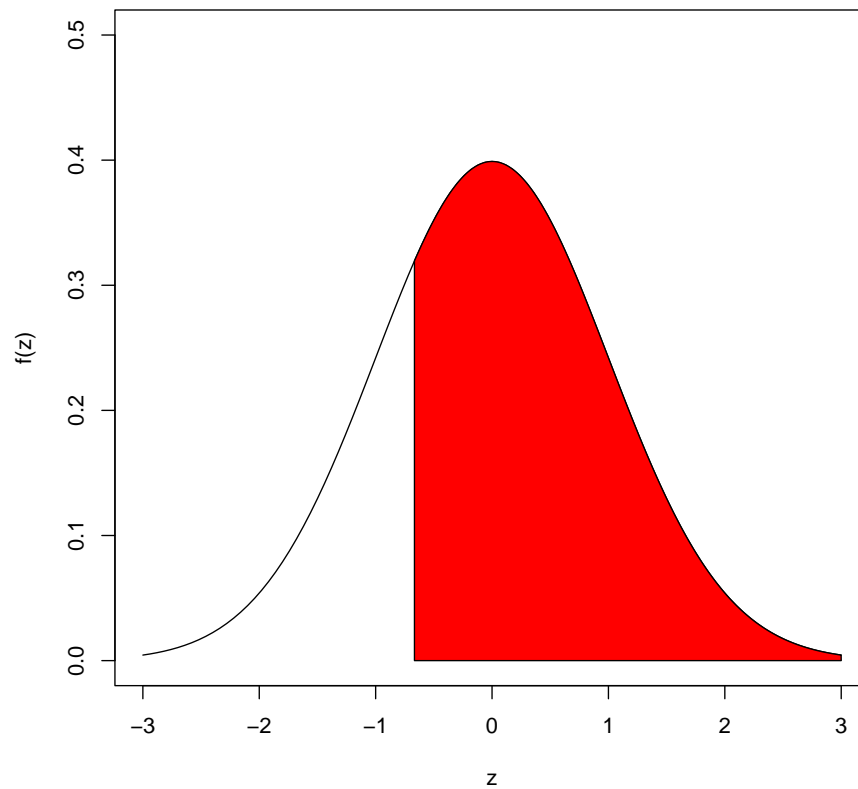
## [1] 0.7475181

plot.new()
curve (dnorm, xlim = c(-3, 3), ylim = c(0, 0.5), xlab = "z",
      ylab="f(z)")
z = -0.6667
```

```

x = c(z, seq(z, 3, by=.001), 3)
y = c(0, dnorm(seq(z, 3, by=.001)), 0)
polygon(x, y, col="red")

```



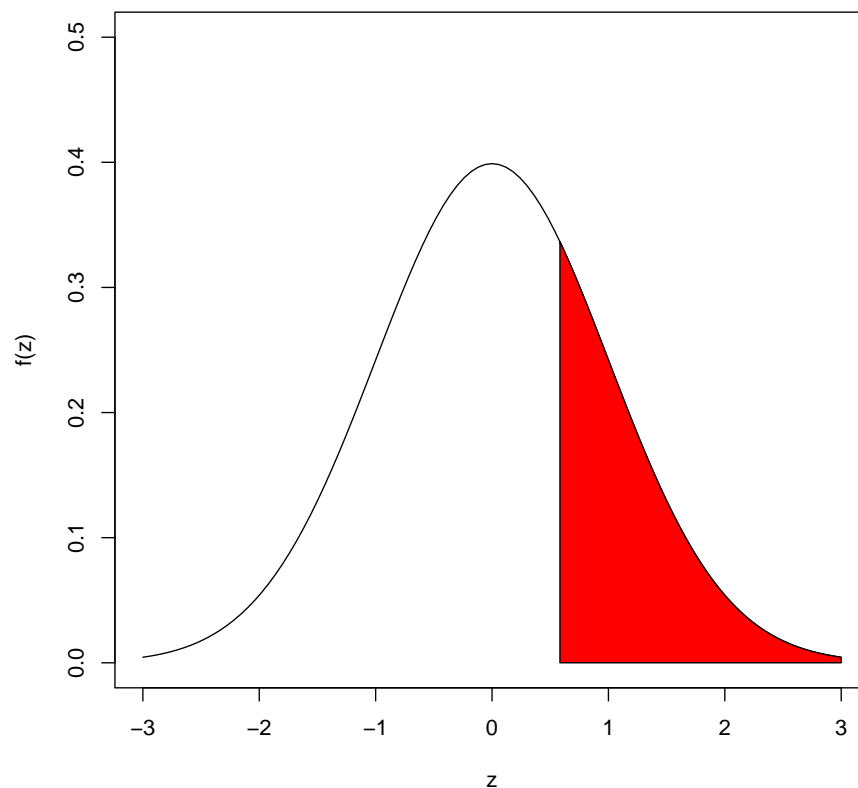
```

pnorm(0.5833)

## [1] 0.7201543

plot.new()
curve (dnorm, xlim = c(-3, 3), ylim = c(0, 0.5), xlab = "z",
       ylab="f(z)")
z = 0.5833
x = c(z, seq(z, 3, by=.001), 3)
y = c(0, dnorm(seq(z, 3, by=.001)), 0)
polygon(x, y, col="red")

```



1.3 GENERAL NORMAL PROBABILITY DISTRIBUTIONS

Problem

Suppose X is normal with mean 527 and standard deviation 105. Compute $P(X \leq 310)$.

```
pnorm(310,527,105)
```

```
## [1] 0.01938279
```

Problem

If $X \sim N(\mu = 100pts, \sigma = 15pts)$

i) Find $P(80pts < X < 95pts)$

```
pnorm(95, mean=100, sd=15) - pnorm(80, mean=100, sd=15)
## [1] 0.2782301
```

ii) Find $P(X > 125pts)$.

```
1 - pnorm(125, mean=100, sd=15)
## [1] 0.04779035
```

iii) Find P_{75} , the 75th percentile of the above distribution. This is the same as the 0.75 quantile.

Problem

The weekly wages of 1000 workmen are normally distributed around a mean of Rs. 70 with S.D of Rs 5. Estimate the number of workers whose weekly wages will be (i) Between Rs 69 and Rs 72 (ii) Less than Rs 69 (iii) More than Rs 72

```
( pnorm(72, mean=70, sd=5) - pnorm(69, mean=70, sd=5))*1000
## [1] 234.6815
```

```
(pnorm(69, mean=70, sd=5))*1000
## [1] 420.7403
```

```
(1 - pnorm(72, mean=70, sd=5))*1000
## [1] 344.5783
```

Problem

In a test on 2000 Electric bulbs ,it was found that the life of particular make, was normally distributed with an average life of 2040 hours and S.D of 60 hours. Estimate the number of bulbs likely to burn for (i) More than 2150 hours (ii) Less than 1950 hours (iii) More than 1920 hours but less than 2160 hours

```
(1 - pnorm(2150, mean=2040, sd=60))*2000
## [1] 66.75302

# The number of bulbs expected to burn for more than 2150 hours is
# 67 (approximately)
```

```
(pnorm(1950, mean=2040, sd=60))*2000

## [1] 133.6144

# The number of bulbs expected to burn for less than 1950 hours
# is 134 (approximately)
```

```
( pnorm(2160, mean=2040, sd=60) - pnorm(1920, mean=2040,
sd=60))*2000

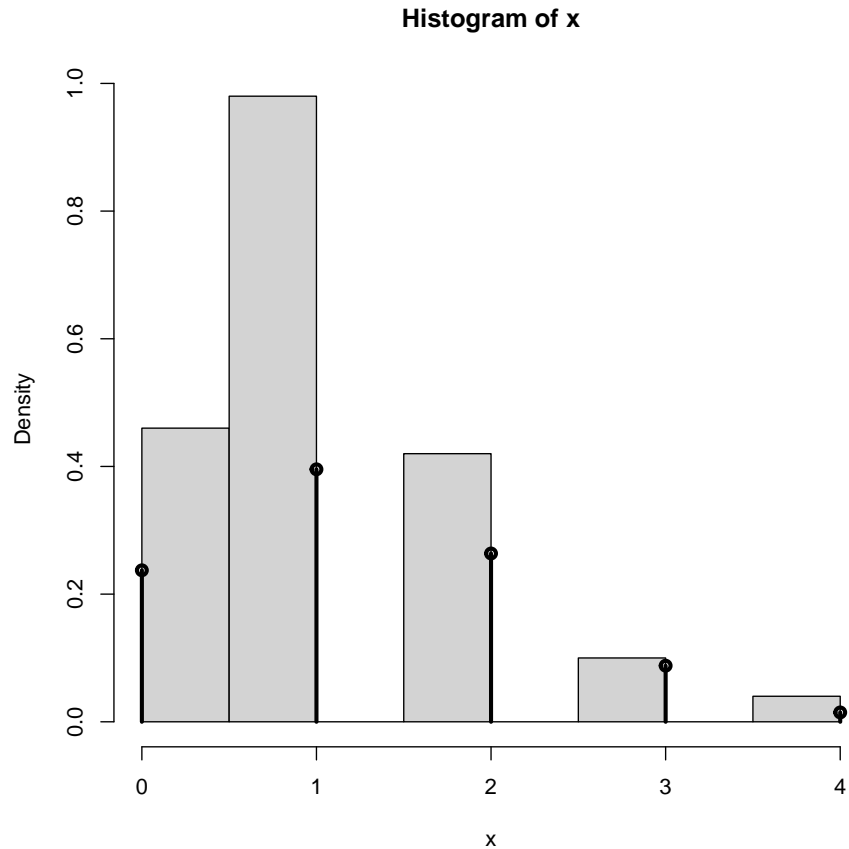
## [1] 1908.999

# The number of bulbs expected to burn more than 1920 hours but
# less than 2160 is 1909 (approximately)
```

1.4 BINOMIAL DISTRIBUTION TENDS TO NORMAL DISTRIBUTION AS 'n' TENDS TO INFINITY:

Binomial distribution tends to Normal distribution

```
n=5;p=.25
x=rbinom(100,n,p)
hist(x,probability=TRUE,)
## use points, not curve as dbinom wants integers only for x
xvals=0:n;points(xvals,dbinom(xvals,n,p),type="h",lwd=3)
points(xvals,dbinom(xvals,n,p),type="p",lwd=3)
```



Experiment

- Find i) $P(0.8 \leq Z \leq 1.5)$ ii) $P(Z \leq 2)$ iii) $P(Z \geq 1)$ Find These probability values and Plot the graph.
(Standard Normal distribution)
- If mean=70 and Standard deviation is 16 i) $P(38 \leq X \leq 46)$ ii) $P(82 \leq X \leq 94)$ iii) $P(62 \leq X \leq 86)$
(General Normal Distribution.)
- 1000 students had Written an examination the mean of test is 35 and standard deviation is 5. Assuming the to be normal find i) How many students Marks Lie between 25 and 40 ii) How many students get more than 40 iii) How many students get below 20 iv) How many students get 50
(Standard Normal distribution or General Normal Distribution)