Lab6

Fitting and Plotting of Binomial distribution & Poisson distribution

BASICS IN PROBABILITY:-

1. If you want to pick five numbers at random from the set 1:50, then you can

```
> sample(1:40,5)
[1] 15 36 10 40 24

> sample(1:40,5)
[1]  3 27  5  8 38

> sample(1:40,5)
[1] 19 17 37 20  6
```

2. Sampling with replacement is suitable for modelling coin tosses or throws of a die.

Roll a die(it gives different results)

```
> sample(1:6,10,replace=TRUE)

[1] 4 4 4 1 4 6 3 5 4 2

> sample(1:6,10,replace=TRUE)

[1] 3 3 4 3 2 1 4 2 6 3
```

3. ## roll 2 die. Even fancier   # replace when rolling dice

```
> dice = as.vector(outer(1:6,1:6,paste))
> sample(dice,5,replace=TRUE)
[1] "1 2" "5 2" "4 4" "6 2" "6 4"

> sample(dice,5,replace=TRUE)
[1] "4 3" "3 6" "1 6" "4 2" "4 6"
```

4. #Toss a coin

```
> sample(c("H","T"),10,replace=TRUE)

[1] "H" "H" "H" "H" "T" "T" "T" "H" "H" "H"

> sample(c("H","T"),10,replace=TRUE)

[1] "T" "T" "H" "H" "T" "H" "H" "H" "T" "T"
```

5. # Combination

$\binom{10}{3}$ *OR* $10c_3$

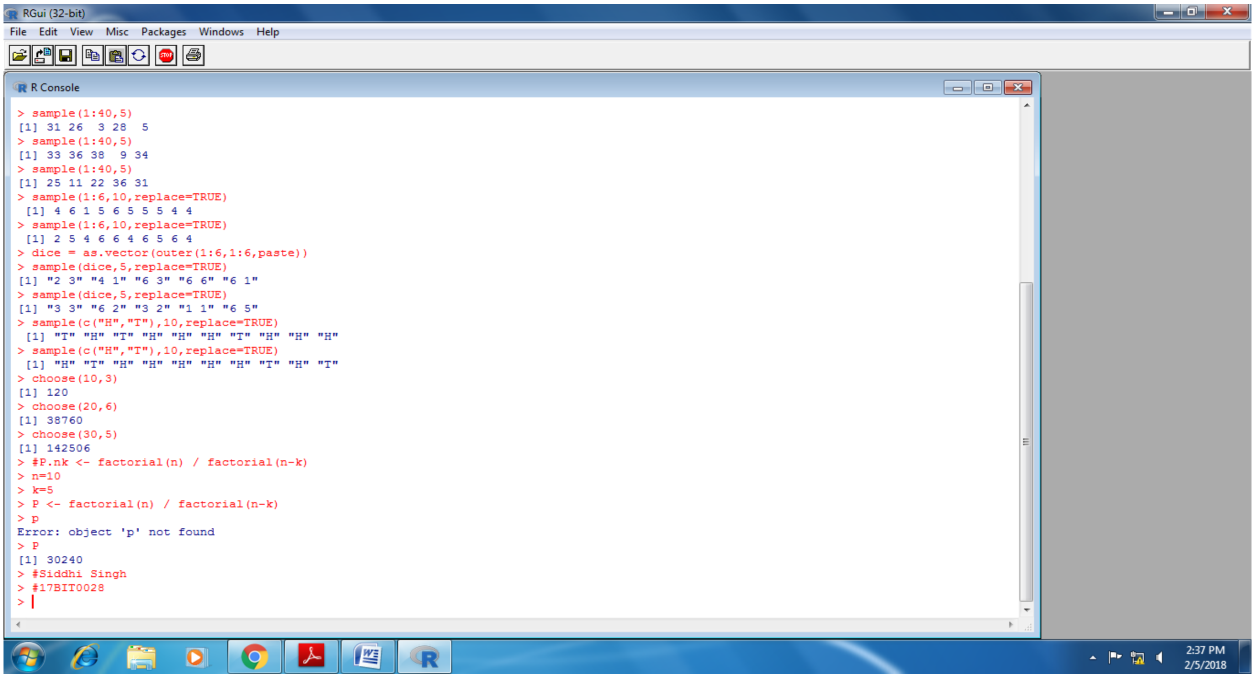
```
> choose(10,3)
[1] 120
```

$\binom{20}{6}$

```
> choose(20,6)
[1] 38760
```

$\binom{30}{5}$

```
> choose(30,5)
[1] 142506
```



6. #permutation (there is no separate permutation function in R)

```
> #P.nk <- factorial(n) / factorial(n-k)
> n=10
> k=5
> P <- factorial(n) / factorial(n-k)
> P
[1] 30240
```

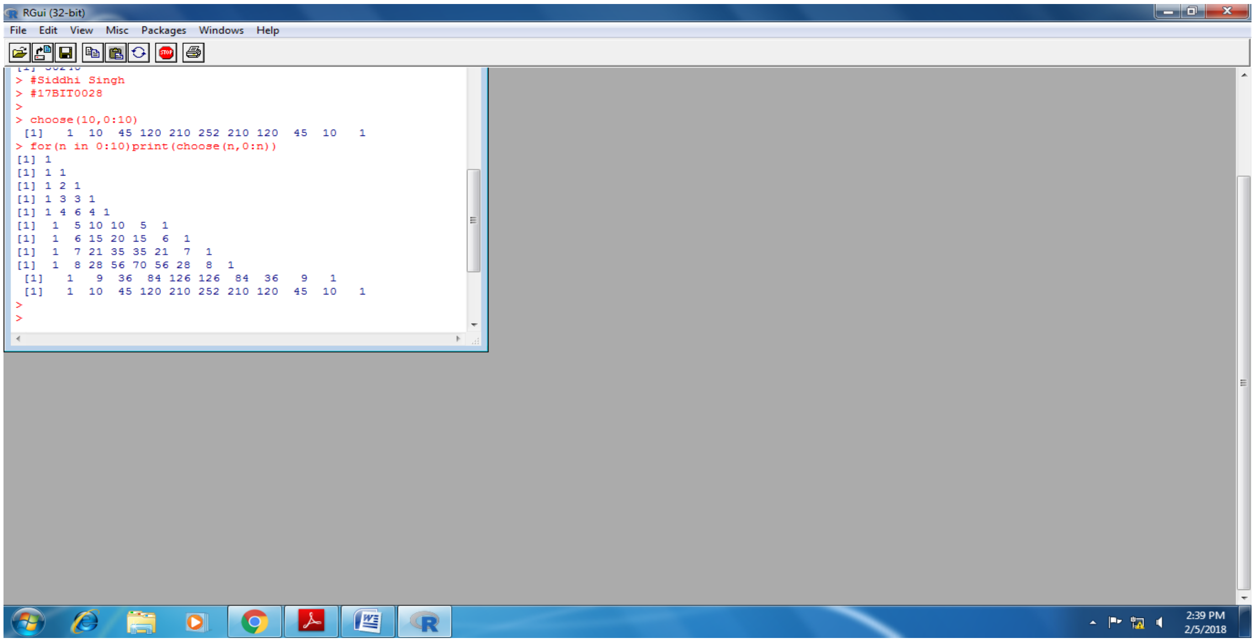
7. Give all binomial coefficients for $\binom{10}{x}$

```
>  choose(10,0:10)

 [1]  1  10  45 120 210 252 210 120  45  10   1
```

8. Use a loop to print the first several rows of pasacal's triangle.

```
> for(n in 0:10)print(choose(n,0:n))
[1] 1
[1] 1 1
[1] 1 2 1
[1] 1 3 3 1
[1] 1 4 6 4 1
[1]  1  5 10 10  5  1
[1]  1  6 15 20 15  6  1
[1]  1  7 21 35 35 21  7  1
[1]  1  8 28 56 70 56 28  8  1
 [1]   1   9  36  84 126 126  84  36   9   1
 [1]   1  10  45 120 210 252 210 120  45  10   1
```



Binomial Distribution

The **binomial distribution** is a discrete probability distribution. It describes the outcome of n independent trials in an experiment. Each trial is assumed to have only two outcomes, either success or failure. If the probability of a successful trial is p, then the probability of having x successful outcomes in an experiment of n independent trials is as follows.

$$P[X=x]=\binom{n}{x}p^x q^{n-x}, x=0,1,\ldots\ldots n$$

Mean $\mu_1^{'}=np$ **Variance:** $\mu_2=npq$

Syntax:-

```
For a binomial(n,p) random variable X, the R functions involve the abbreviation "binom":

        dbinom(k,n,p)  # binomial(n,p) density at k: Pr(X = k)

        pbinom(k,n,p)  # binomial(n,p) CDF at k: Pr(X <= k)

        qbinom(P,n,p)  # binomial(n,p) P-th quantile

        rbinom(N,n,p)  # N binomial(n,p) random variables

        help(Binomial) # documentation on the functions related
              # to the Binomial distribution
```

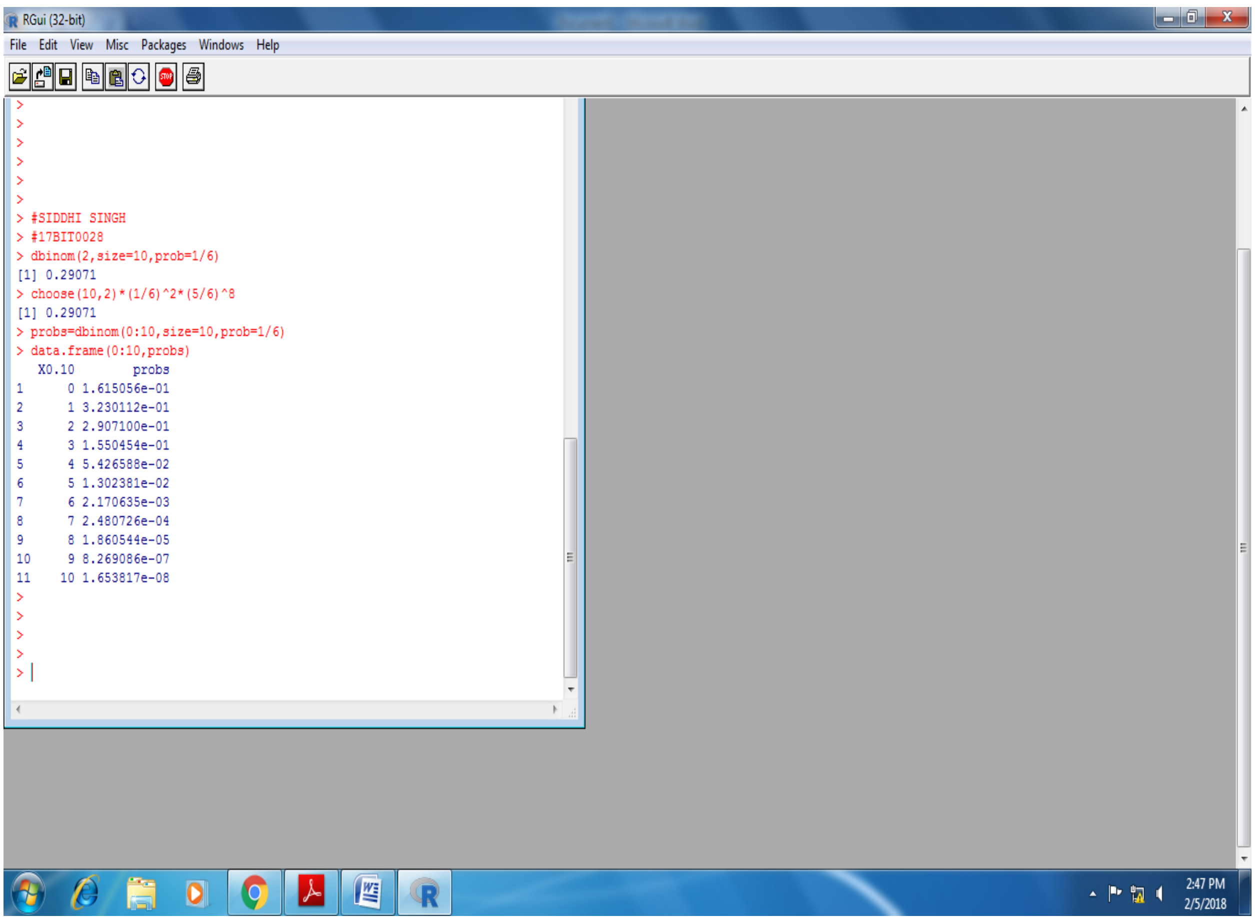
Problem1. **Find the Probability of getting two ‘4’ among ten dice**

```
>dbinom(2,size=10,prob=1/6)
[1] 0.29071
```

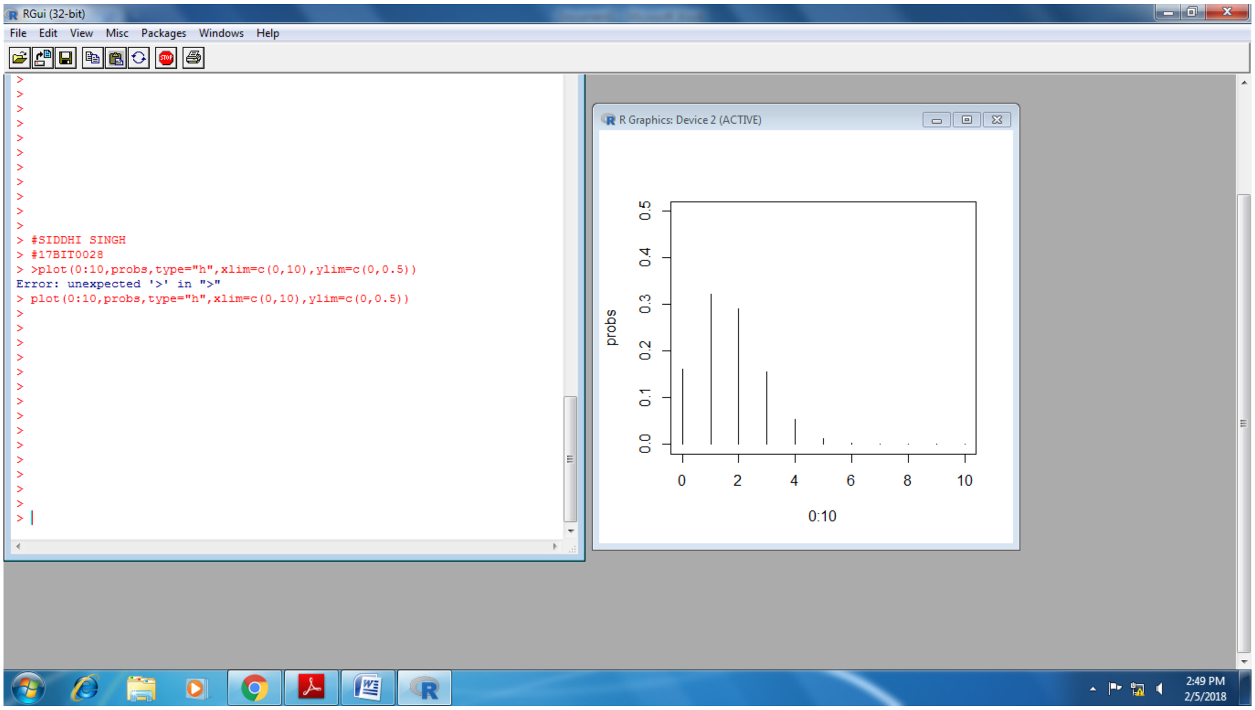
Problem 2: **Find the P(2) by using binomial probability formula**

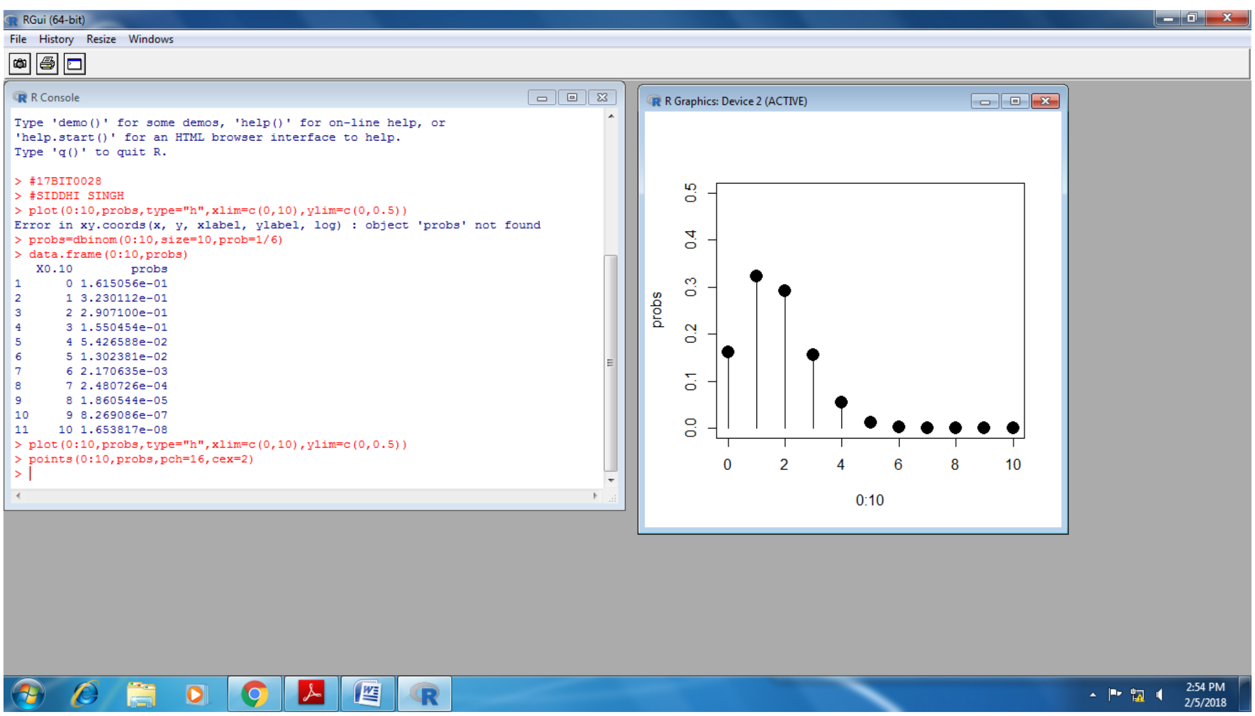
```
> choose(10,2)*(1/6)^2*(5/6)^8
[1] 0.29071
```

Problem 3: **Find the table for BIN(n=10,P=1/6)**

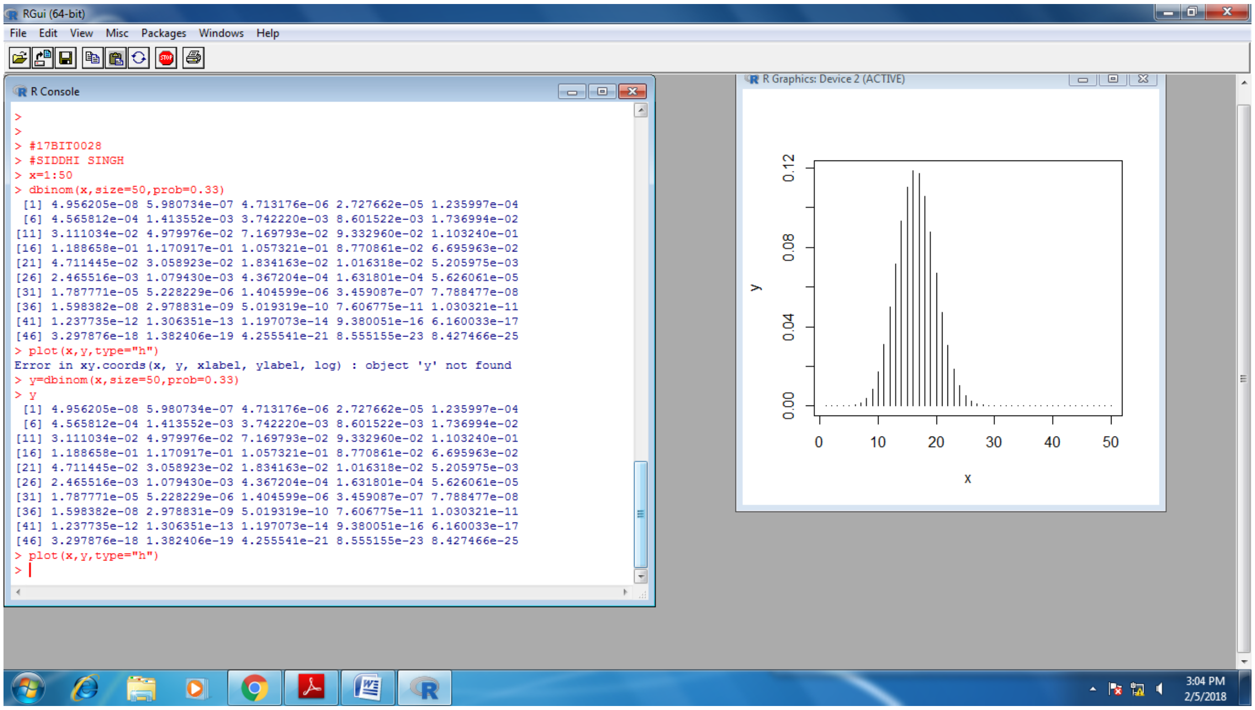


Problem4: BINOMIAL PROBABILITY PLOTS :Draw a Plot for the Binomial distribution Bin(n=10,p=1/6)



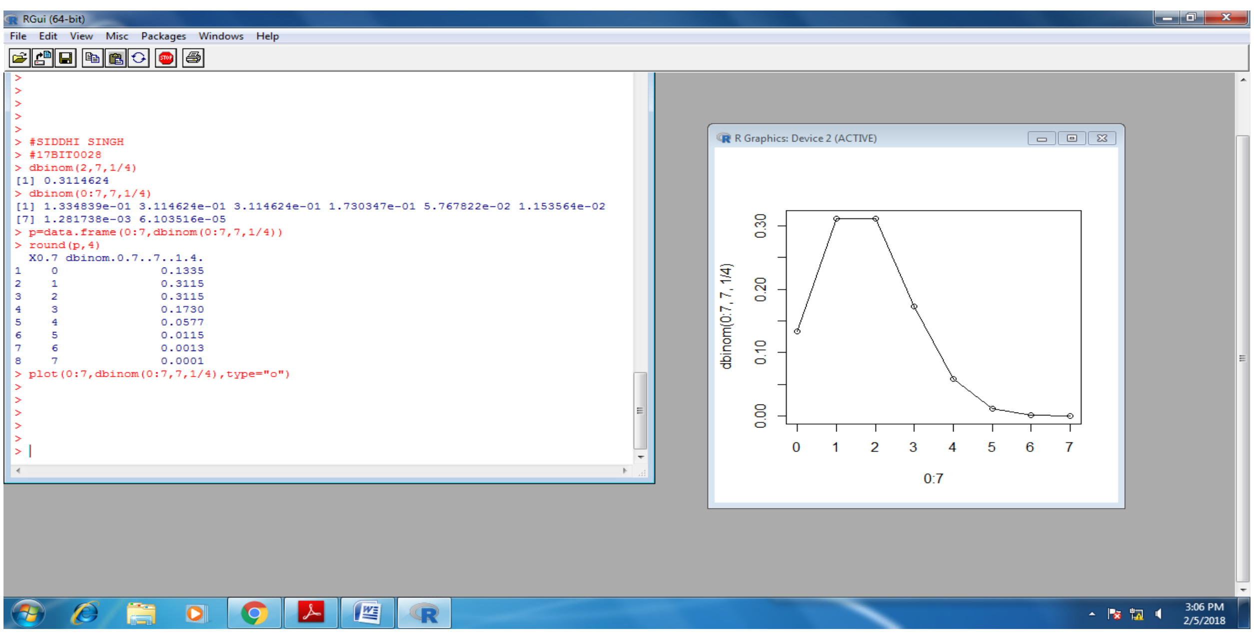


Problem 5: Plot Binomial distribution with n=50 and P=0.33

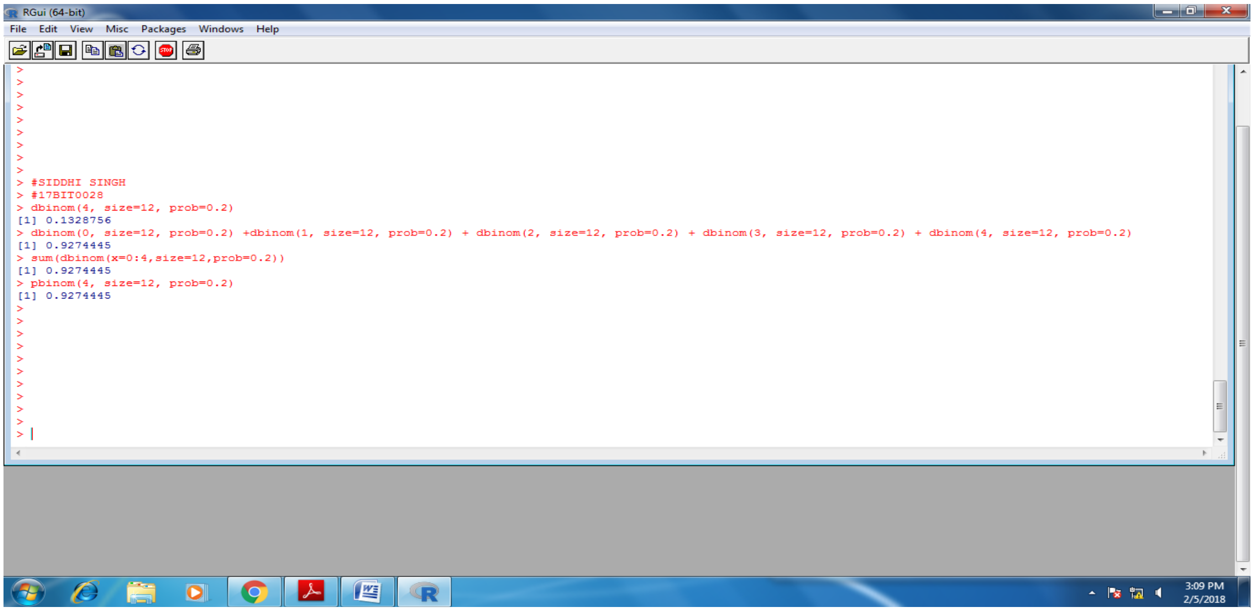


Problem 6 : For a Binomial(7,1/4) random variable named X,

i. Compute the probability of two success
ii. Compute the Probablities for whole space
iii. Display those probabilities in a table
iv. Show the shape of this binomial Distribution

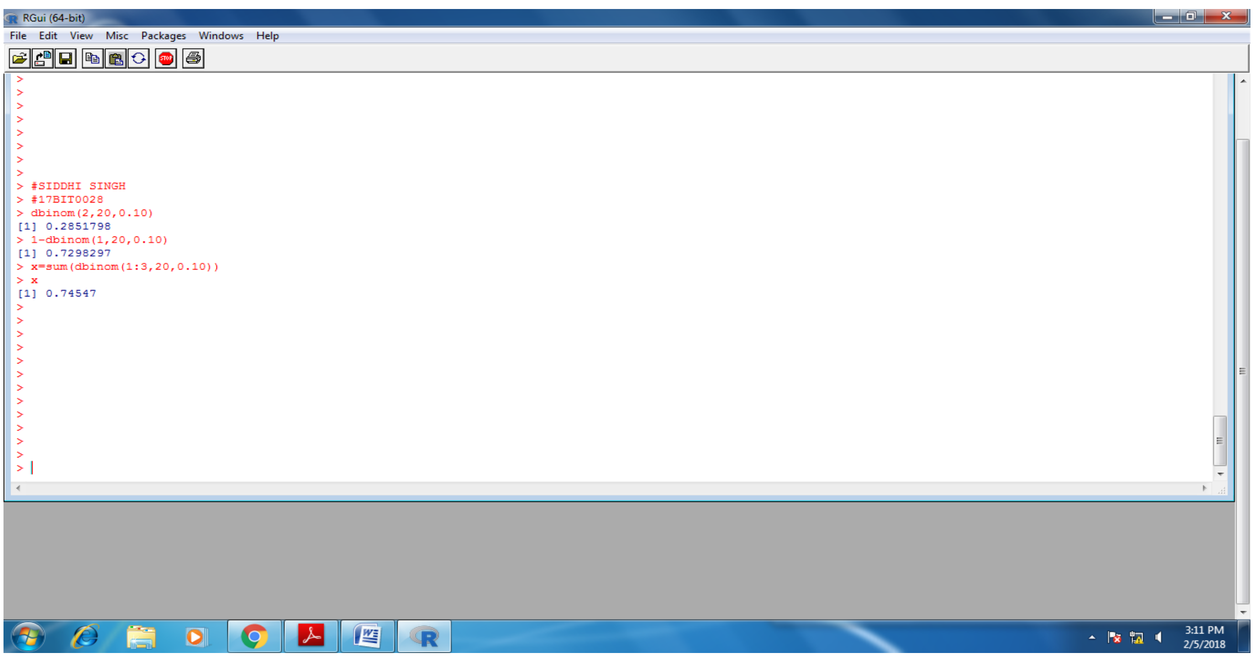


Problem 7: Suppose there are twelve multiple choice questions in an English class quiz. Each question has five possible answers, and only one of them is correct. Find the probability of having four or less correct answers if a student attempts to answer every question at random.



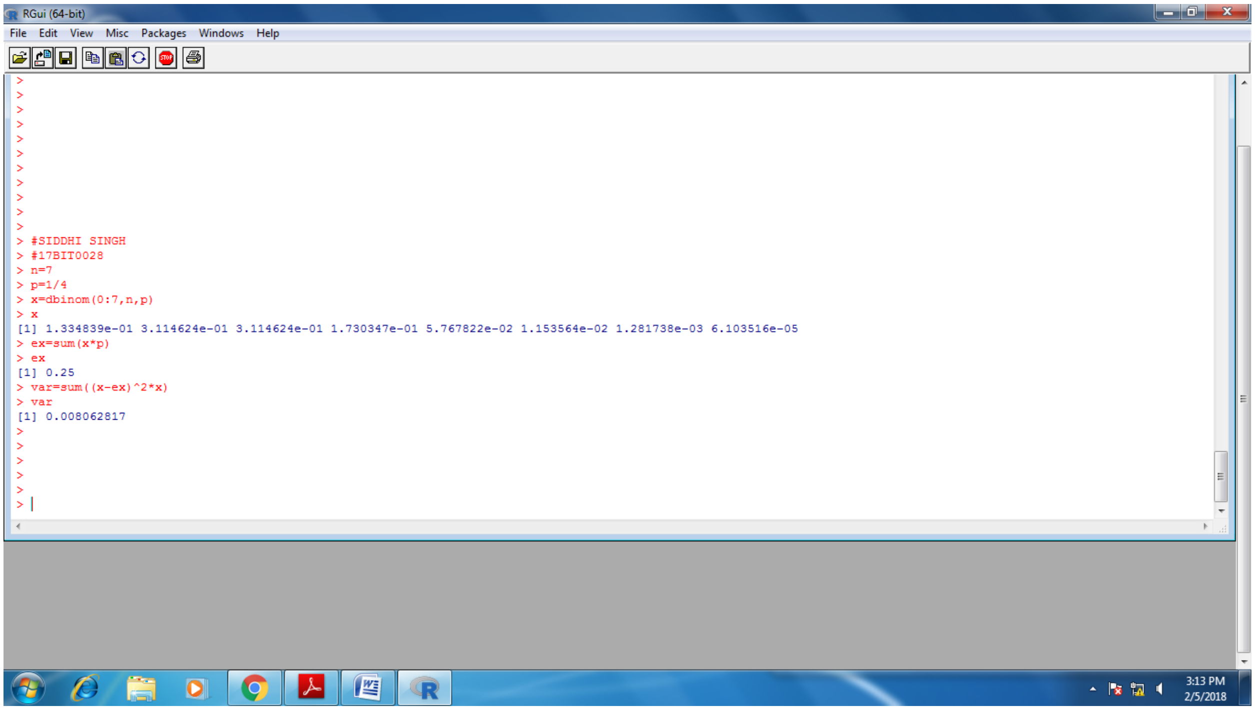
Problem 8: If 10% of the Screws produced by an automatic machine are defective, find the probability that out of 20 screws selected at random, there are

(i) Exactly 2 defective
(ii) At least 2 defectives
(iii) Between 1 and 3 defectives (inclusive)



Relationship between mean and variance :-

Problem 9: Show that Binomial distribution variance is less than mean with Binomial variable follows (7,1/4)



THE POISON DISTRIBUTION:

If the number of Bernoulli trials of a random experiment is fairly large and the probability of success is small it becomes increasingly difficult to compute the binomial probabilities. For values of n and p such that $n \geq 150$ and $p \leq 0.05$, the poisson distribution serves as an excellent approximation to the binominal distribution.

The random variable X is said to follow the Poisson distribution if and only if

$$p[X = x] = \frac{e^{-\lambda}\lambda^{x}}{\lfloor x}, x = 0,1,2,.....$$

Assumptions:-

1. Number of Bernoulli trials (n) is indefinitely large, ($n \to \infty$)
2. The trials are independent.
3. Probability of success (p) is very small, ($p \to 0$)

   λ = np is constant, $\lambda = np \Rightarrow p = \frac{\lambda}{n}$

4. Mean and variance in poison distribution are equal

Syntax:-

```
dpois(x, lambda, log = FALSE)
ppois(q, lambda, lower.tail = TRUE, log.p = FALSE)
qpois(p, lambda, lower.tail = TRUE, log.p = FALSE)
rpois(n, lambda)
```

Problem 1:

a. #P(x=5) with parameter 7

```
> dpois(x=5,lambda=7)
[1] 0.1277167
```

b. #P(x=0)+P(x=1)+……….+P(x=5)

```
> dpois(x=0:5,lambda=7)

[1] 0.000911882 0.006383174 0.022341108 0.052129252 0.091226192 0.127716668
```

c.

```
> #P(x<=5)
> sum(dpois(0:5,lambda=7))
[1] 0.3007083
```

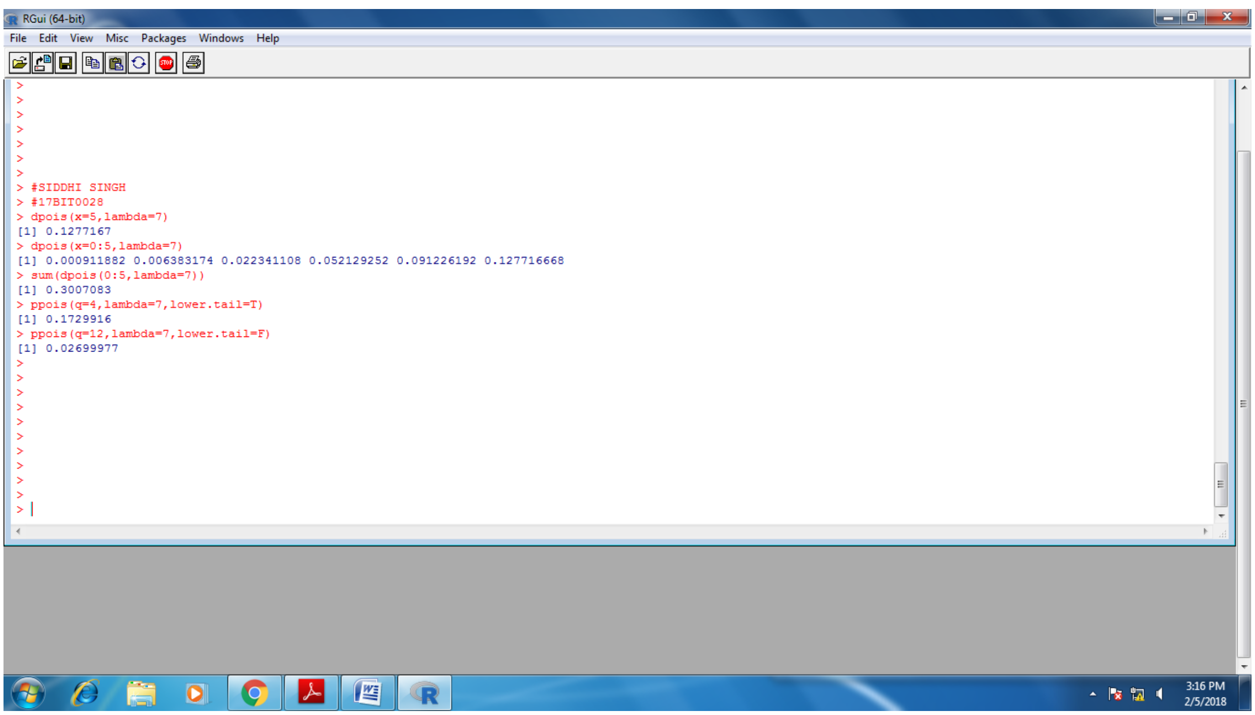
Or

```
> ppois(q=4,lambda=7,lower.tail=T)
[1] 0.1729916
```

d.

```
> ppois(q=12,lambda=7,lower.tail=F)

[1] 0.02699977
```



Problem 2 : Check the relationship between mean and variance in Poisson distribution(4) with n=100

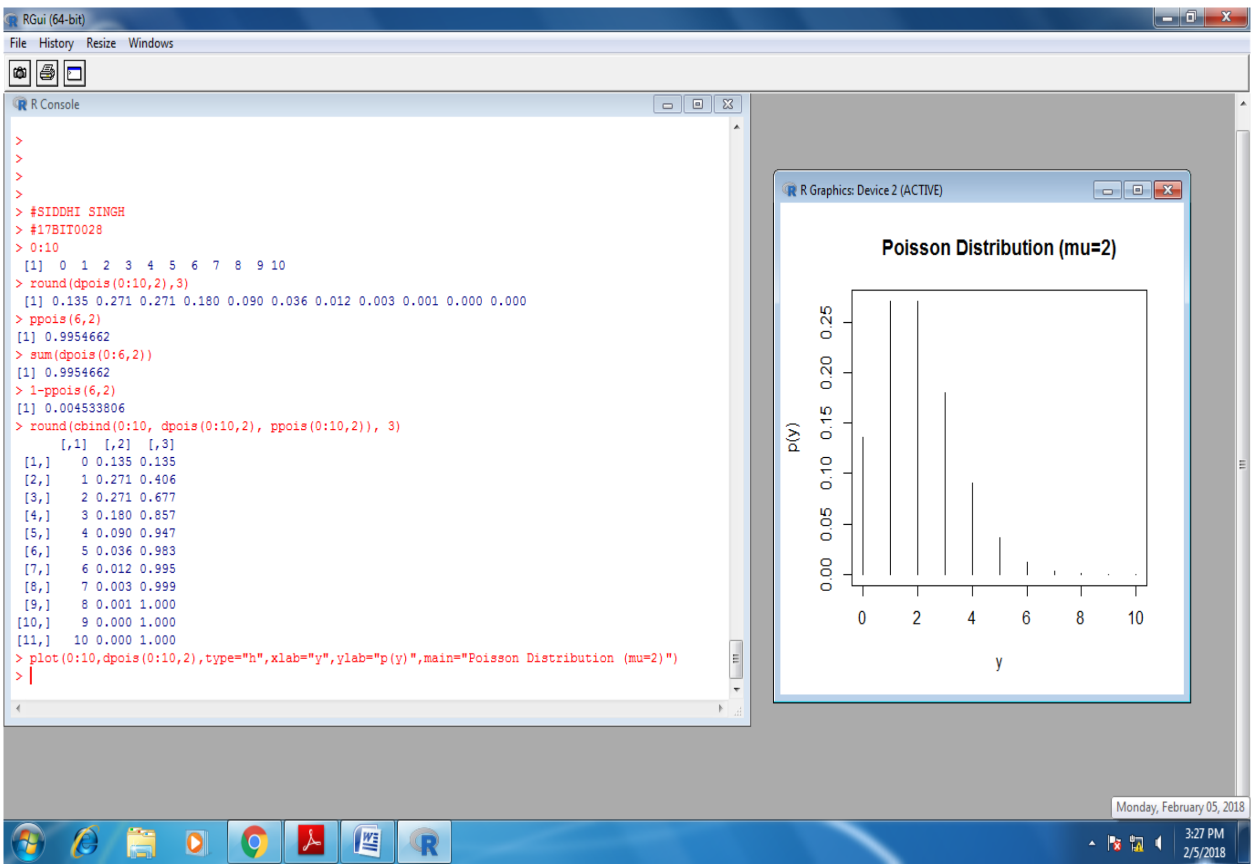
```
>
>
>
>
>
>
>
>
>
>
> #SIDDHI SINGH
> #17BIT0028
>
> X.val=0:100
> P.val=dpois(X.val,4)
> EX=sum(X.val*P.val)
> EX
[1] 4
> sum((X.val-EX)^2*P.val)
[1] 4
>
>
>
>
>
>
>
>
>
> |
```

Problem 3 : Compute Probabilities and cumulative probabilities of the values between 0 and 10 for the parameter 2 in poisson distribution.

```
>
>
> dpois(0:10,2)
 [1] 1.353353e-01 2.706706e-01 2.706706e-01 1.804470e-01 9.022352e-02 3.608941e-02 1.202980e-02 3.437087e-03 8.592716e-04 1.909493e-04 3.818985e-05
> P=data.frame (0:10,dpois(0:10,2))
> round (P,4)
   X0.10 dpois.0.10..2.
1      0         0.1353
2      1         0.2707
3      2         0.2707
4      3         0.1804
5      4         0.0902
6      5         0.0361
7      6         0.0120
8      7         0.0034
9      8         0.0009
10     9         0.0002
11    10         0.0000
> ppois(0:10,2)
 [1] 0.1353353 0.4060058 0.6766764 0.8571235 0.9473470 0.9834364 0.9954662 0.9989033 0.9997626 0.9999535 0.9999917
> p=data.frame(0:10,ppois(0:10,2))
> round(p,4)
   X0.10 ppois.0.10..2.
1      0         0.1353
2      1         0.4060
3      2         0.6767
4      3         0.8571
5      4         0.9473
6      5         0.9834
7      6         0.9955
8      7         0.9989
9      8         0.9998
10     9         1.0000
11    10         1.0000
> |
```

Problem 3: Poisson distribution with parameter '2'

1. How to obtain a sequence from 0 to 10
2. Calculate P(0),P(1),...,P(10) when lambda =2 and Make the output prettier
3. Find P(x <= 6)
4. Sum all probabilities
5. Find P(Y>6)
6. Make a table of the first 11 Poisson probs and cumulative probs when # mu=2 and make the output prettier
7. Plot the probabilities Put some labels onthe axes and give the plot a title:



AIM:* *Computing/plotting and visualising the following probability distributions

About Normal Distribution:-

THE NORMAL DISTRIBUTION :

A random variable X is said to posses normal distribution with mean μ and variance σ^2 , if its probability density function can be expressed of the form,

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} , \quad -\infty < x < +\infty$$

The standard notation used to denote a random variable to follow normal distribution with appropriate mean and variance is, $X \sim N(\mu , \sigma^2)$

STANDARD NORMAL DISTRIBUTION :

If a random variable X follows normal distribution with mean μ and variance σ^2 , its transformation $Z = \frac{X-\mu}{\sigma}$ follows standard normal distribution (mean 0 and unit variance)

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} , \quad -\infty < z < +\infty$$

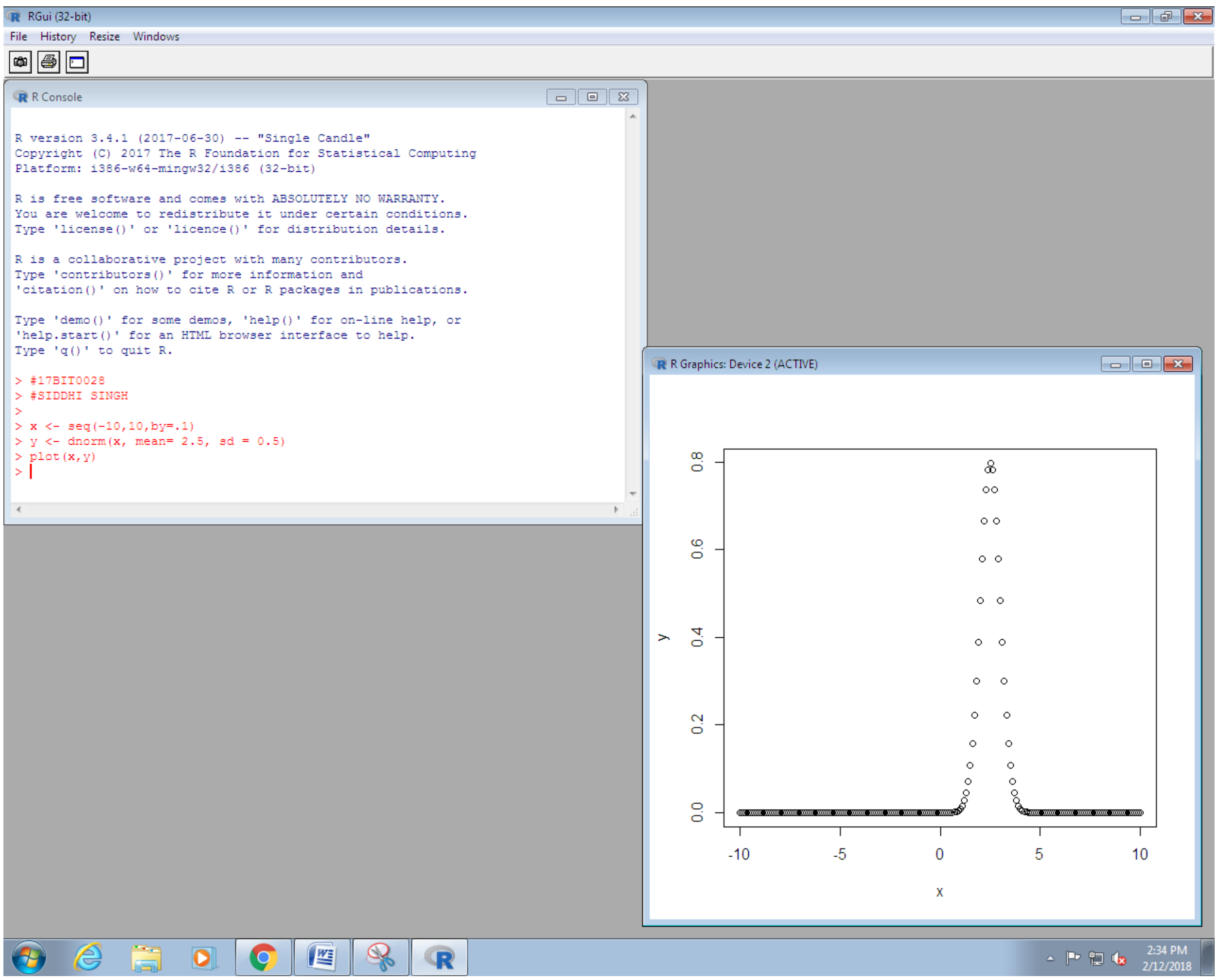
The distribution function of the standard normal distribution

$$F(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \, dx$$

(I) Normal distribution computations and graphs

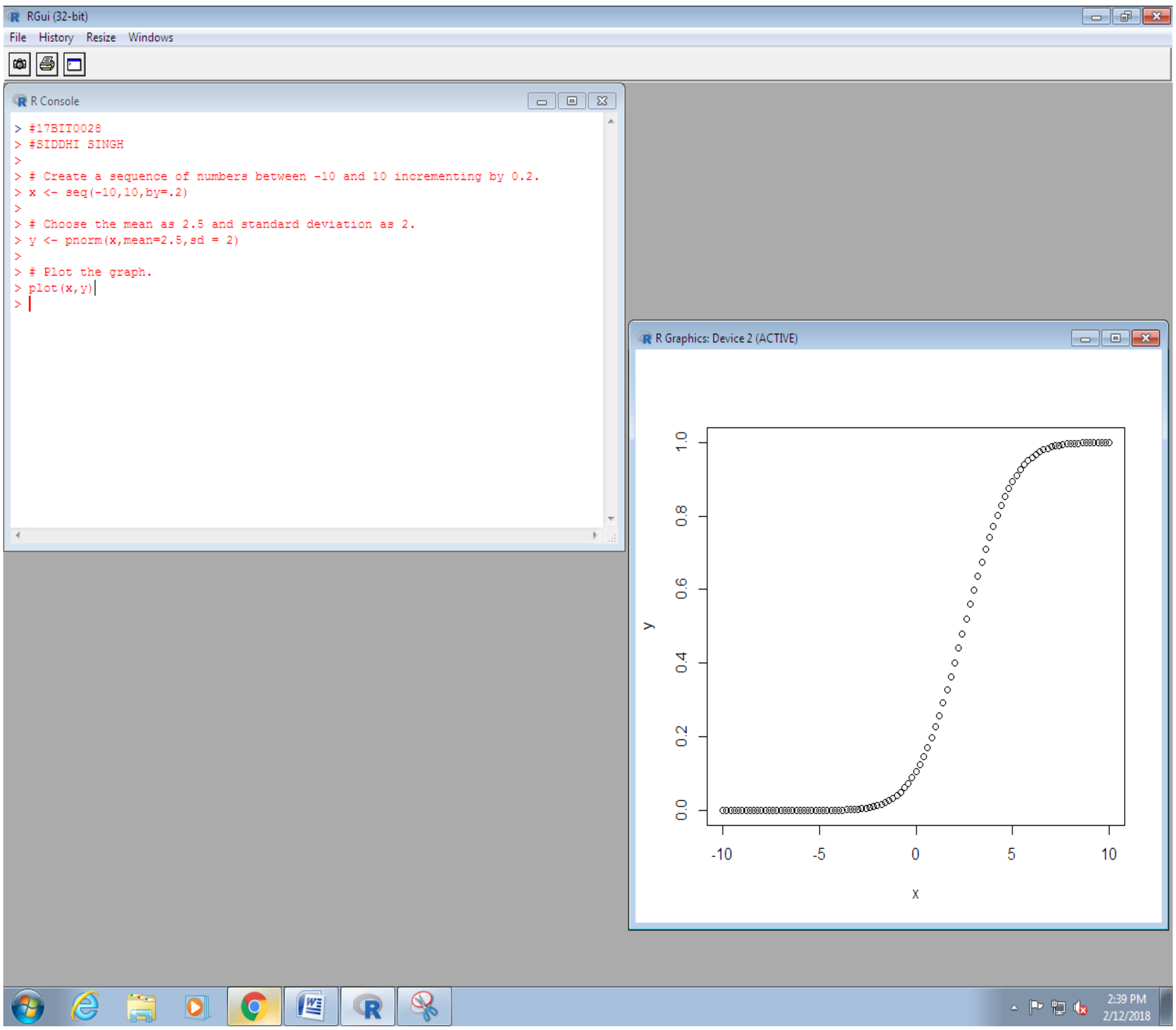
dnorm() :

This function gives height of the probability distribution at each point for a given mean and standard deviation.



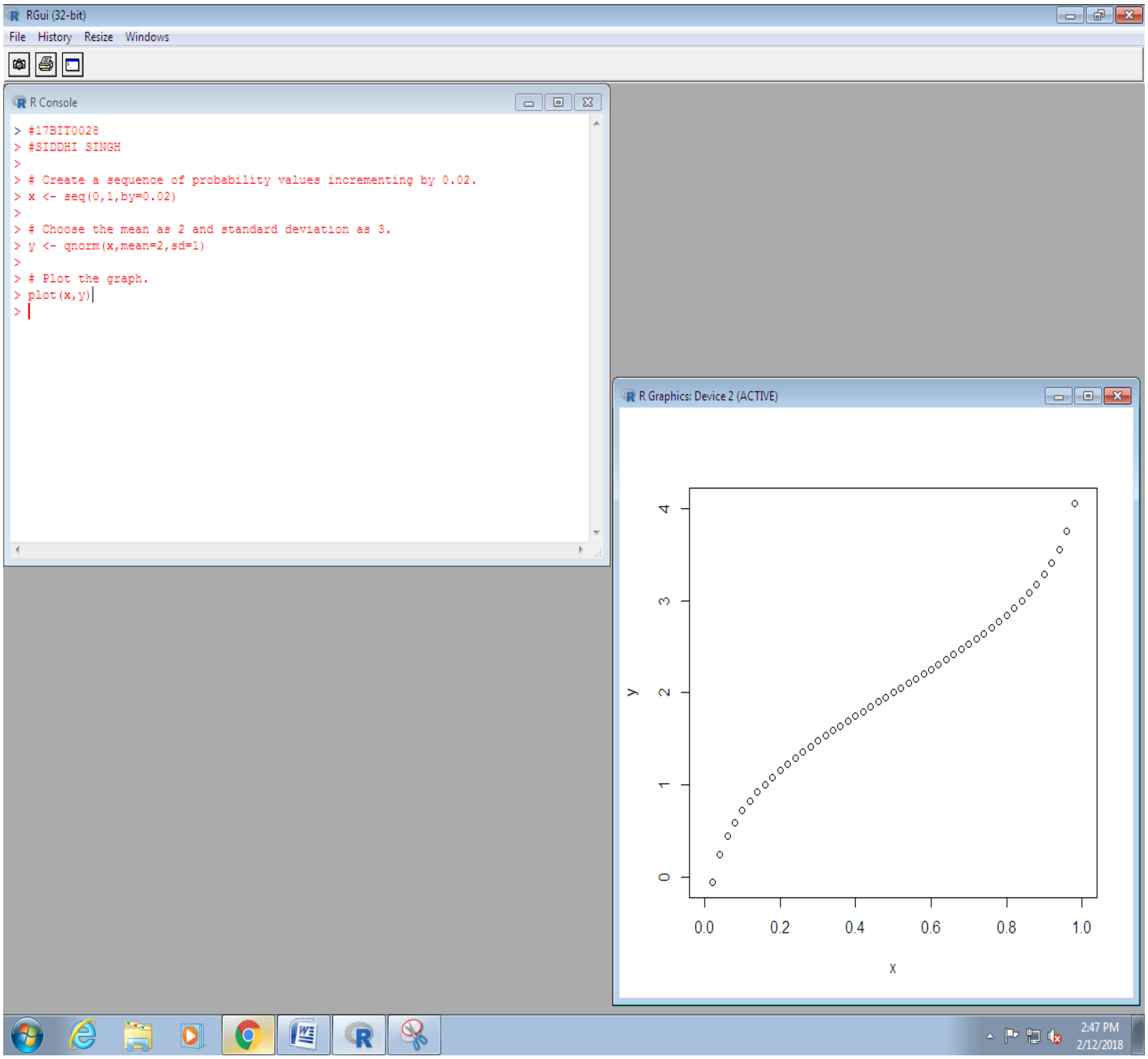
pnorm():

This function gives the probability of a normally distributed random number to be less that the value of a given number. It is also called "Cumulative Distribution Function".



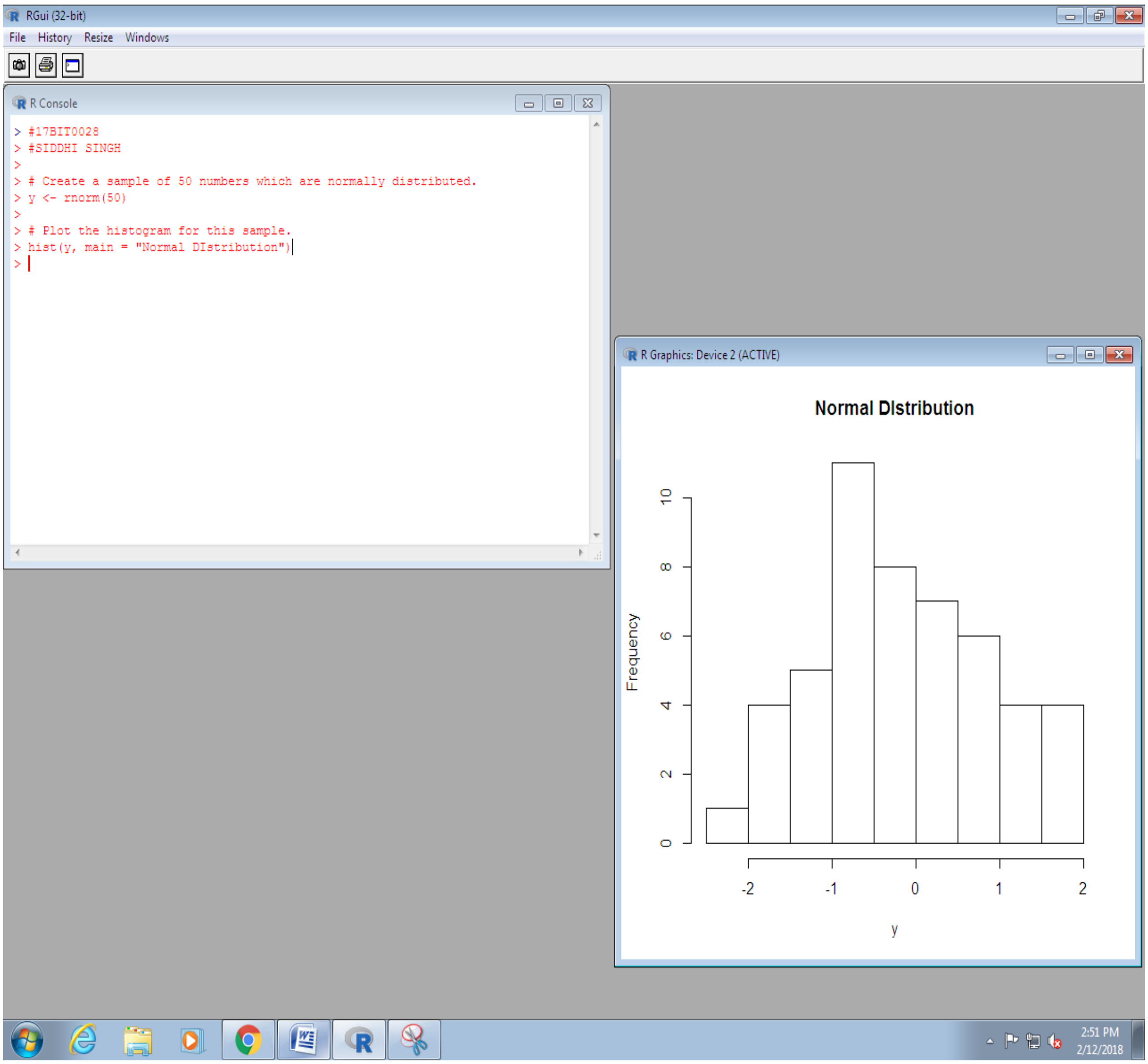
qnorm()

This function takes the probability value and gives a number whose cumulative value matches the probability value.



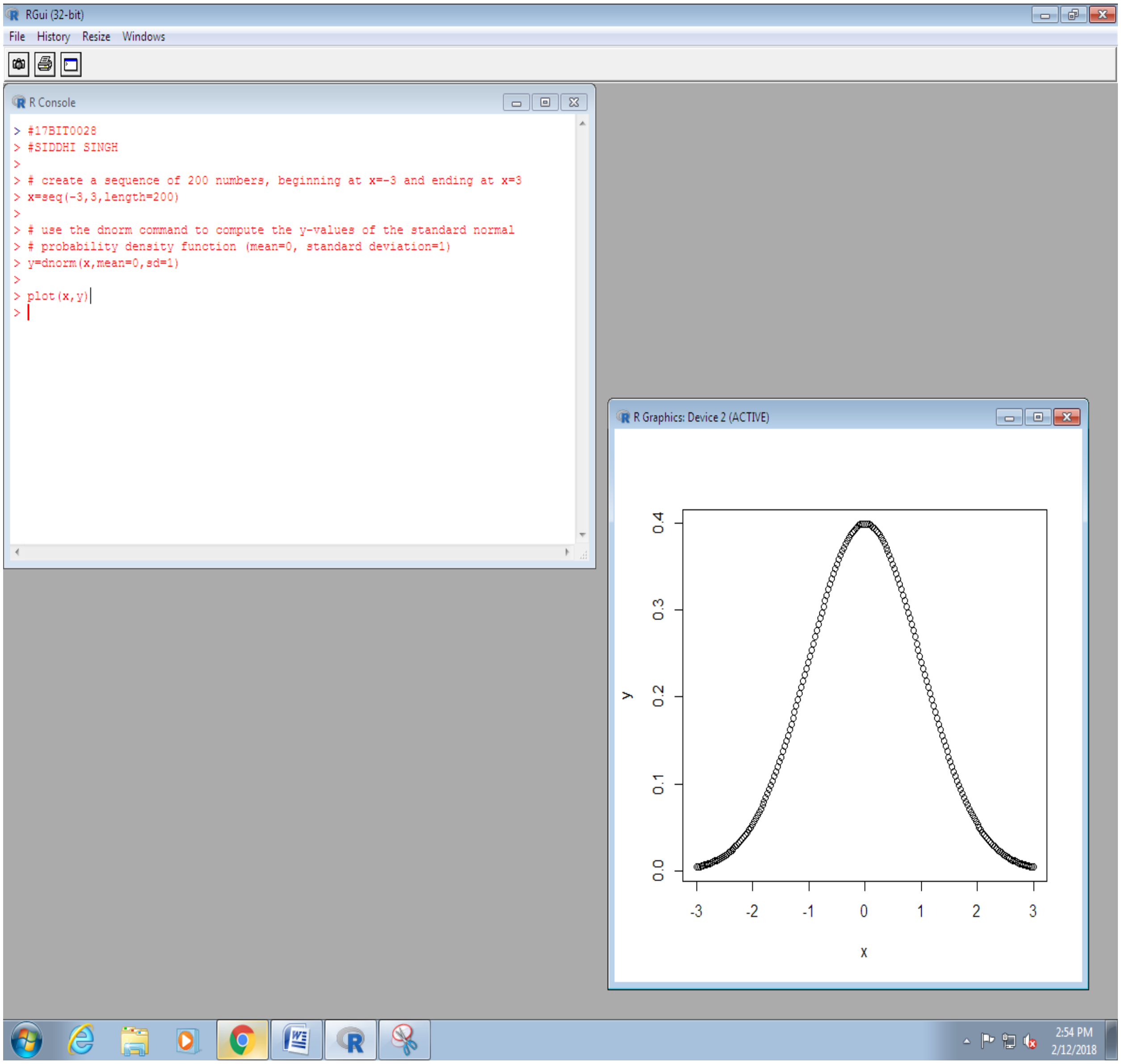
rnorm()

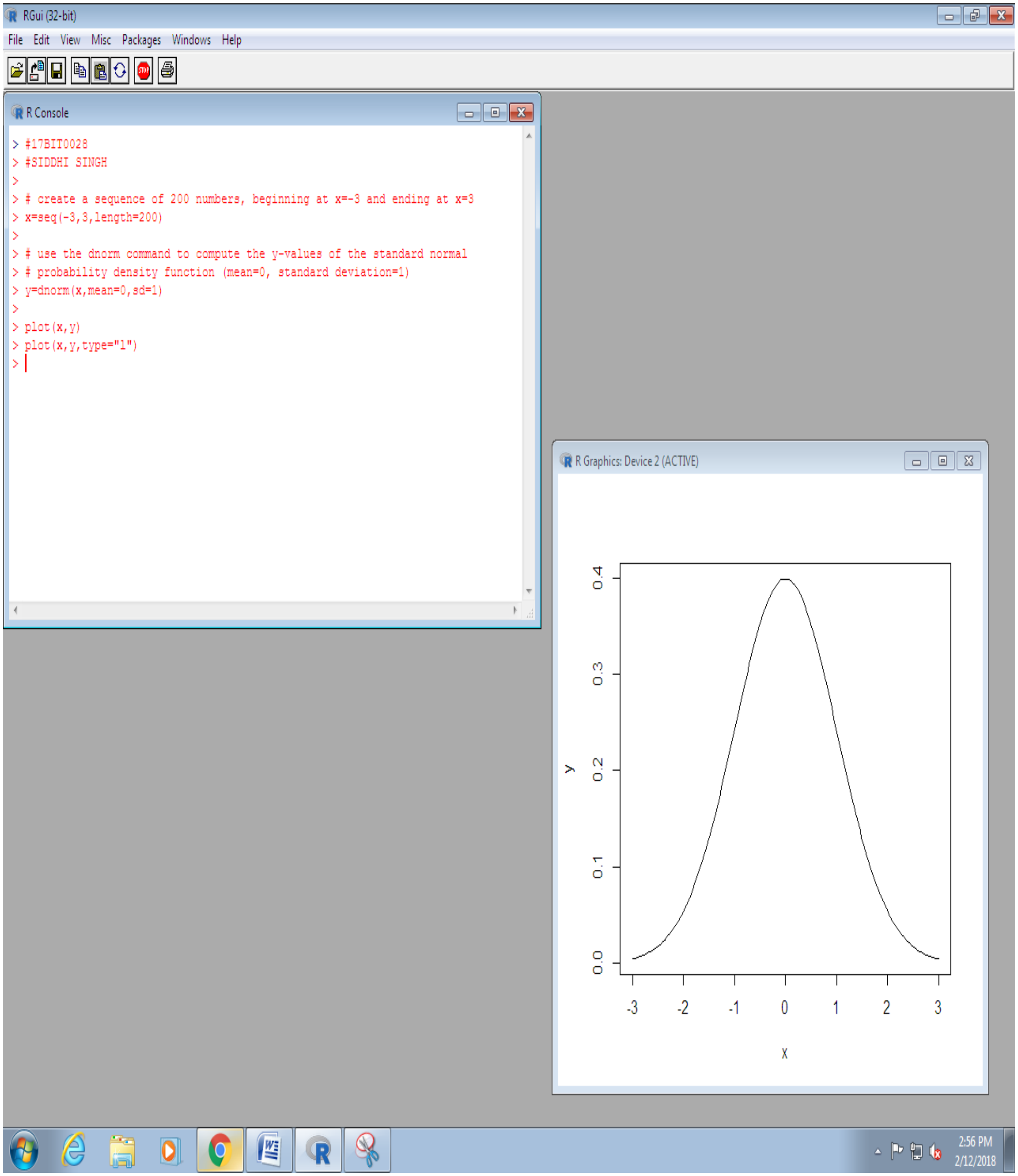
This function is used to generate random numbers whose distribution is normal. It takes the sample size as input and generates that many random numbers. We draw a histogram to show the distribution of the generated numbers.



CODE 1 :-

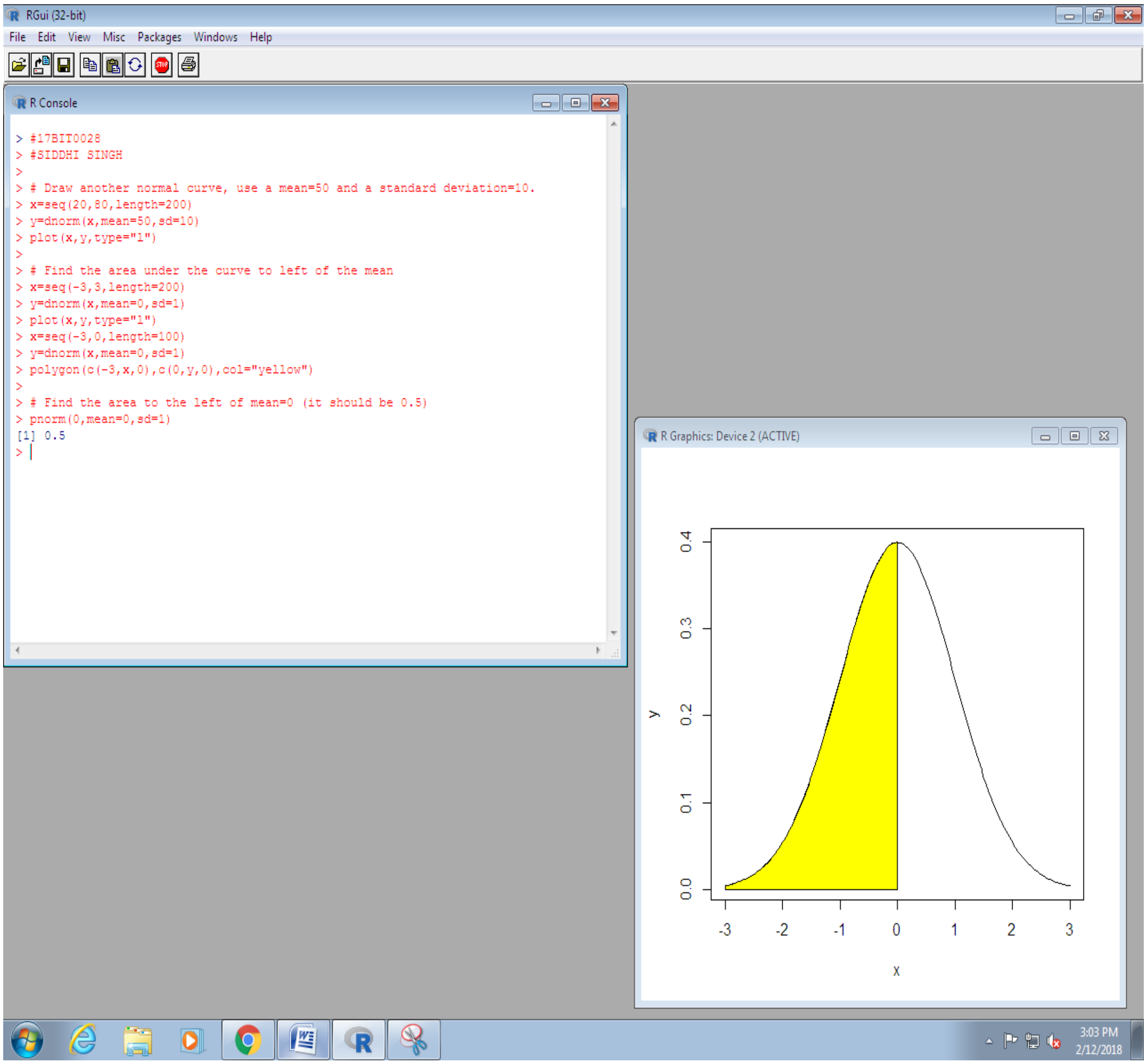
 # create a sequence of 200 numbers, beginning at x=-3 and ending at x=3




RGui (32-bit)
File Edit View Misc Packages Windows Help
R Console
> #17BIT0028
> #SIDDHI SINGH
>
> # create a sequence of 200 numbers, beginning at x=-3 and ending at x=3
> x=seq(-3,3,length=200)
>
> # use the dnorm command to compute the y-values of the standard normal
> # probability density function (mean=0, standard deviation=1)
> y=dnorm(x,mean=0,sd=1)
>
> plot(x,y)
> plot(x,y,type="l")
>
R Graphics: Device 2 (ACTIVE)
0.0
0.1
0.2
0.3
0.4
y
-3
-2
-1
0
1
2
3
x
2:56 PM
2/12/2018

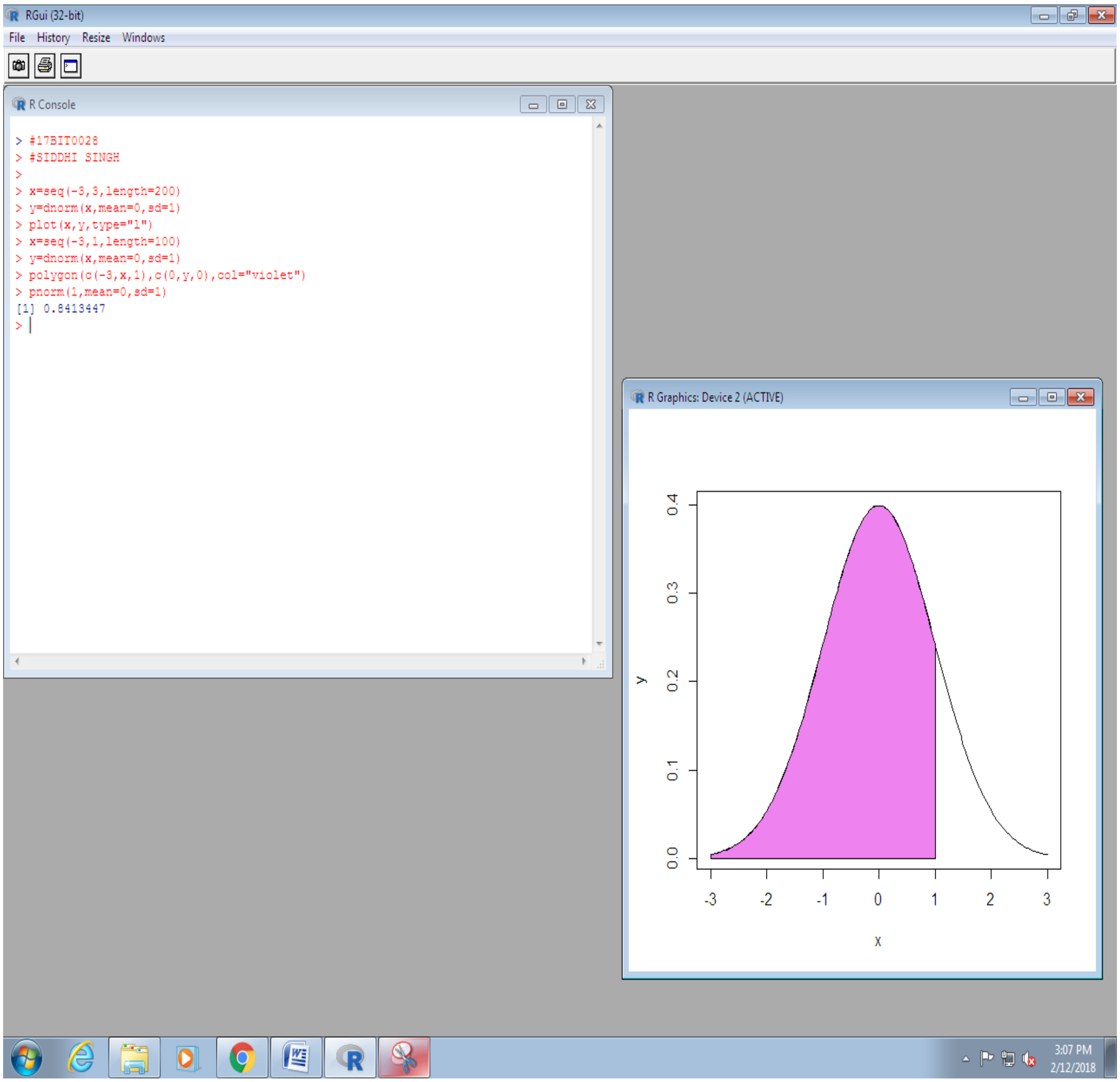
CODE 2:

 # Draw another normal curve, use a mean=50 and a standard deviation=10.



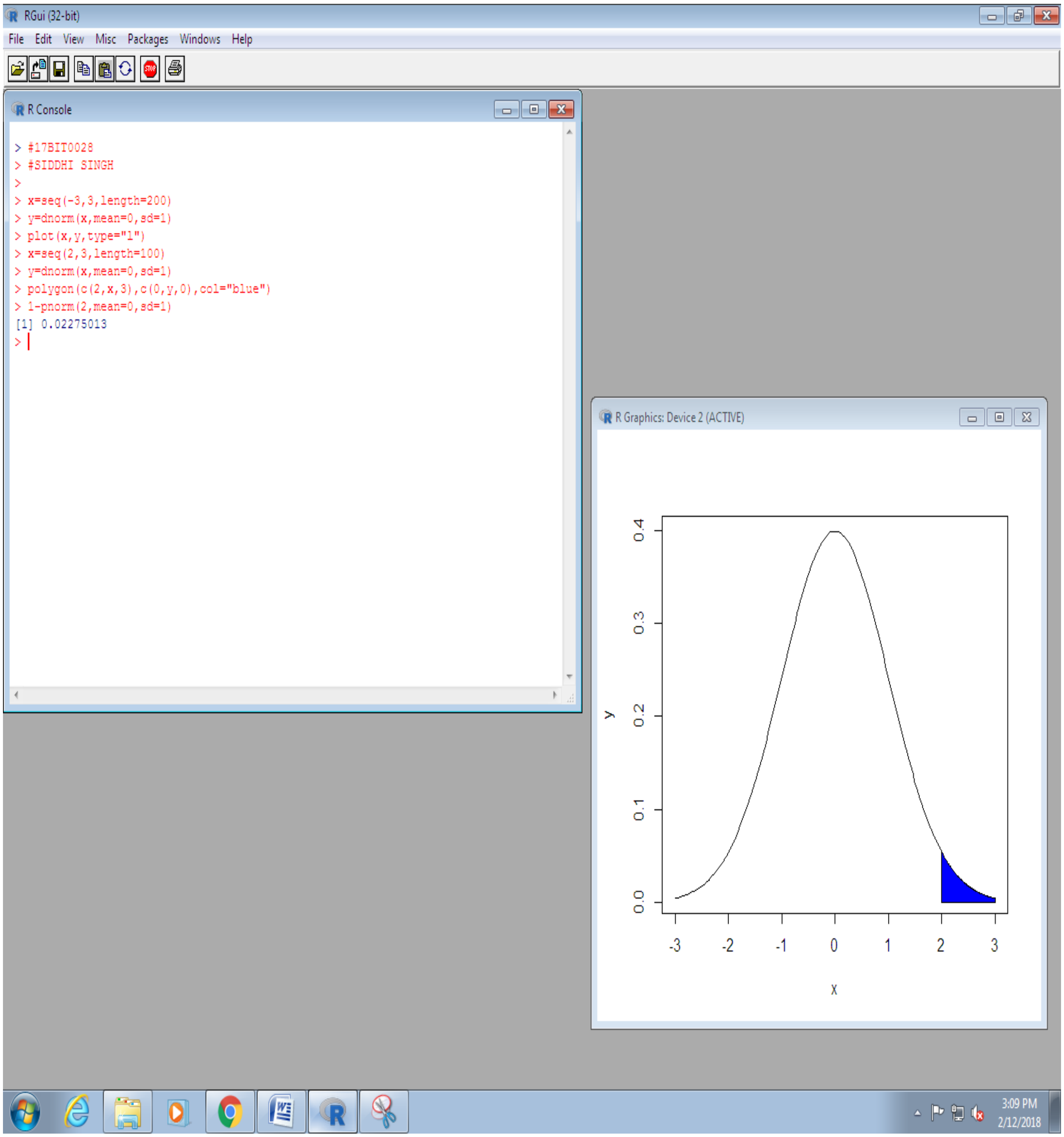
CODE 3:

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



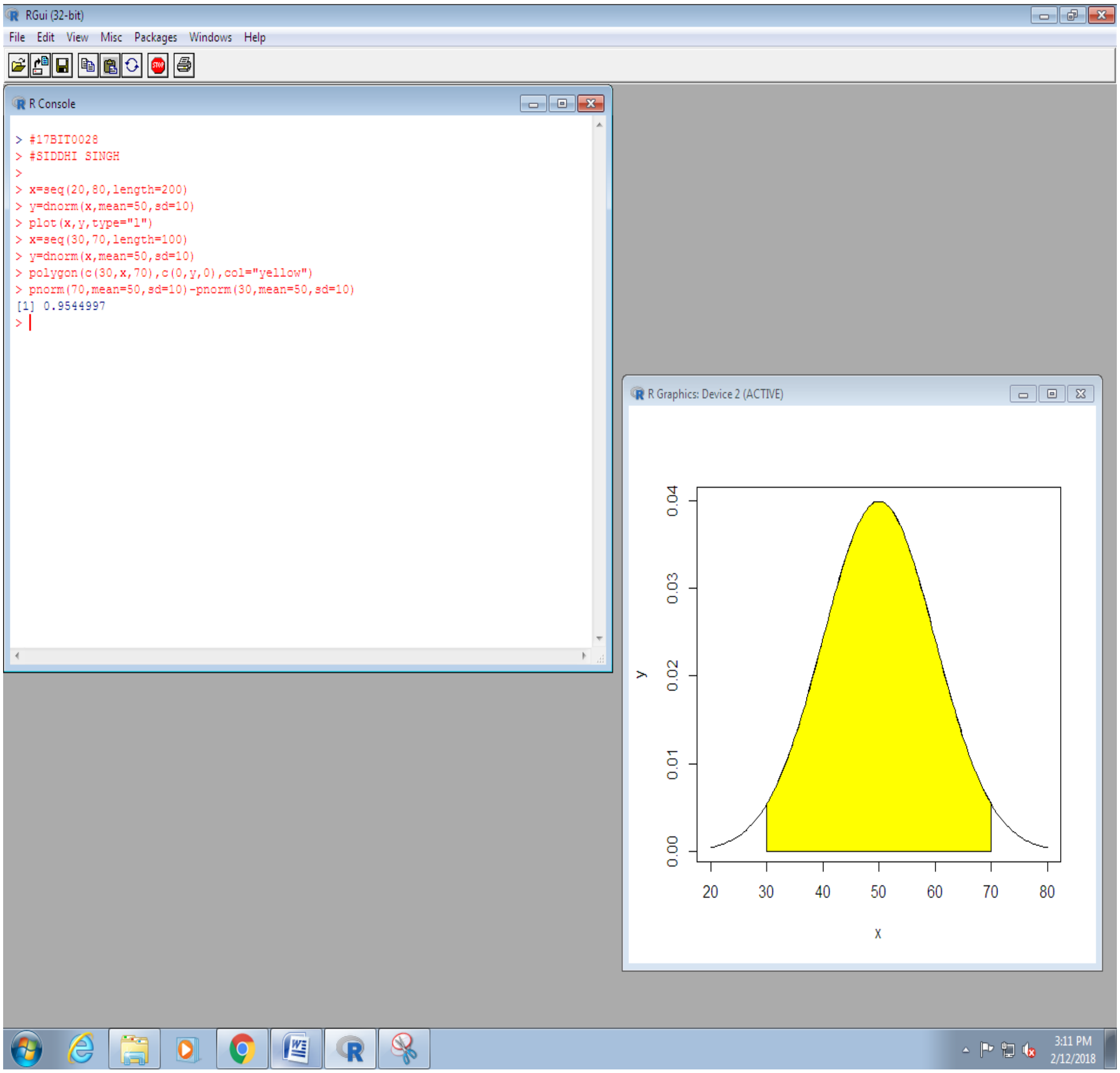
CODE 4:-

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



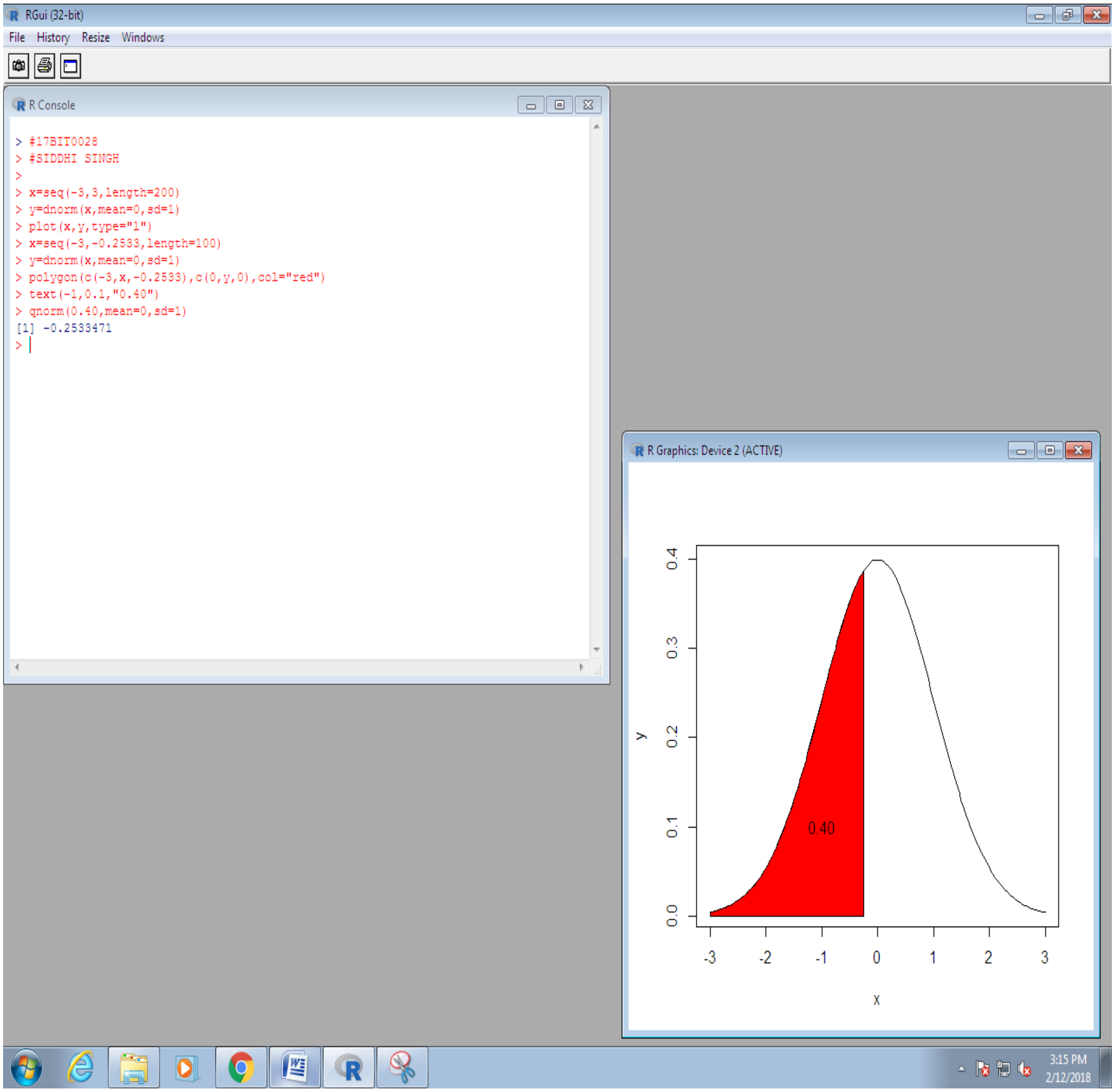
CODE 5:

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.



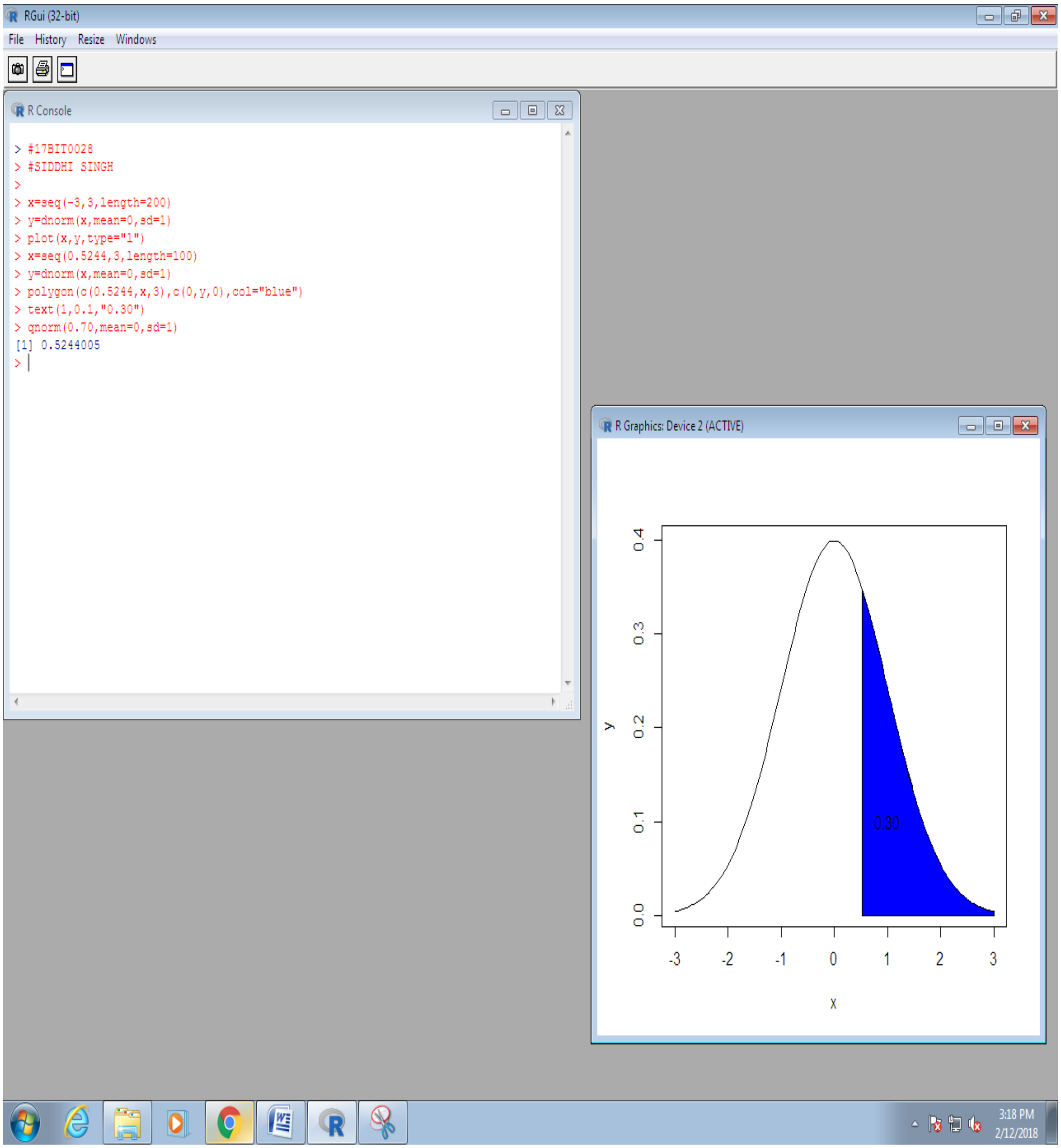
CODE 5:

Find the Quantile (Percentile) - i.e., reverse the process.
That is, given the area, find the value of x.



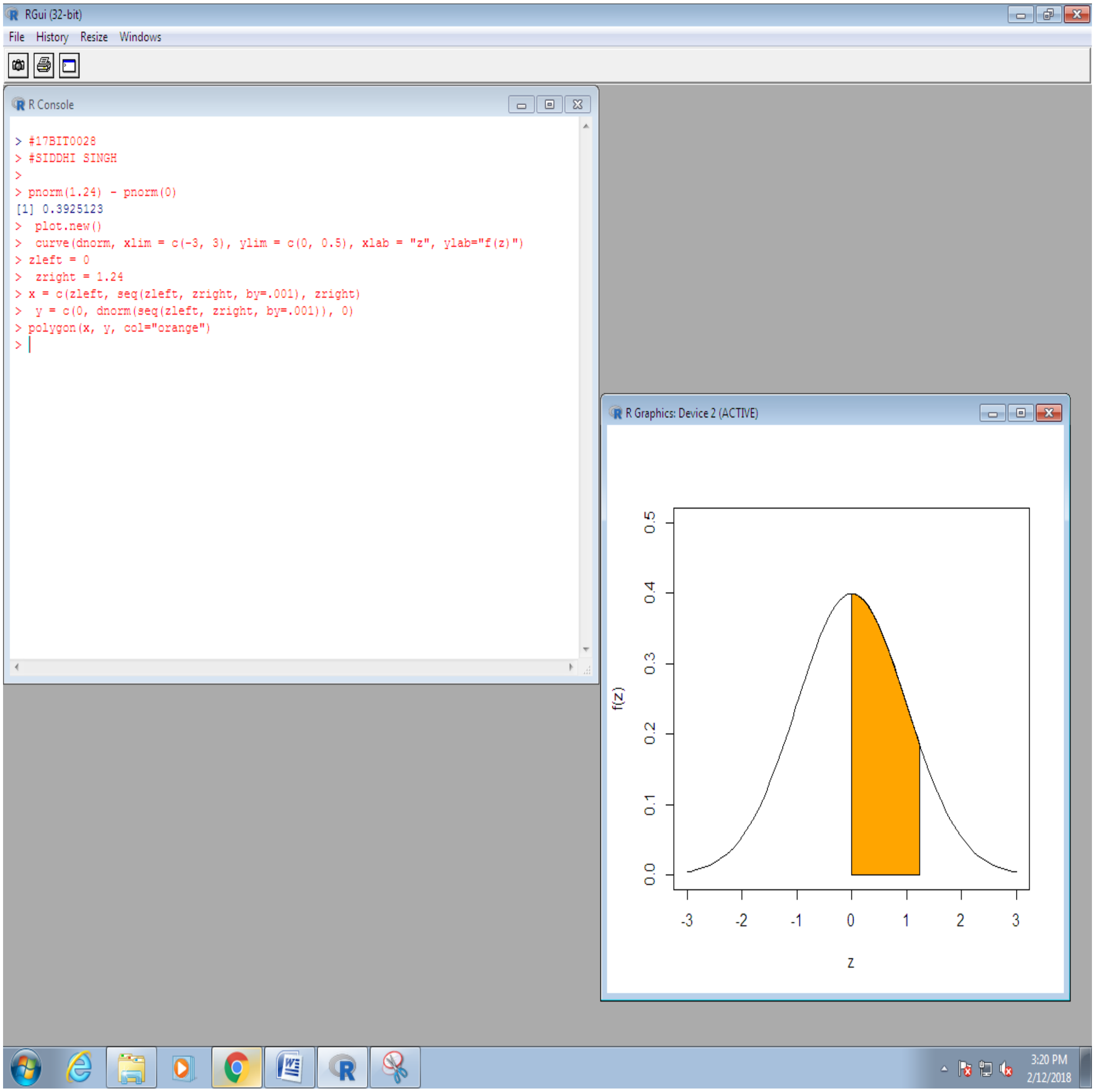
CODE 6:

density curve to the right of x is the given area.

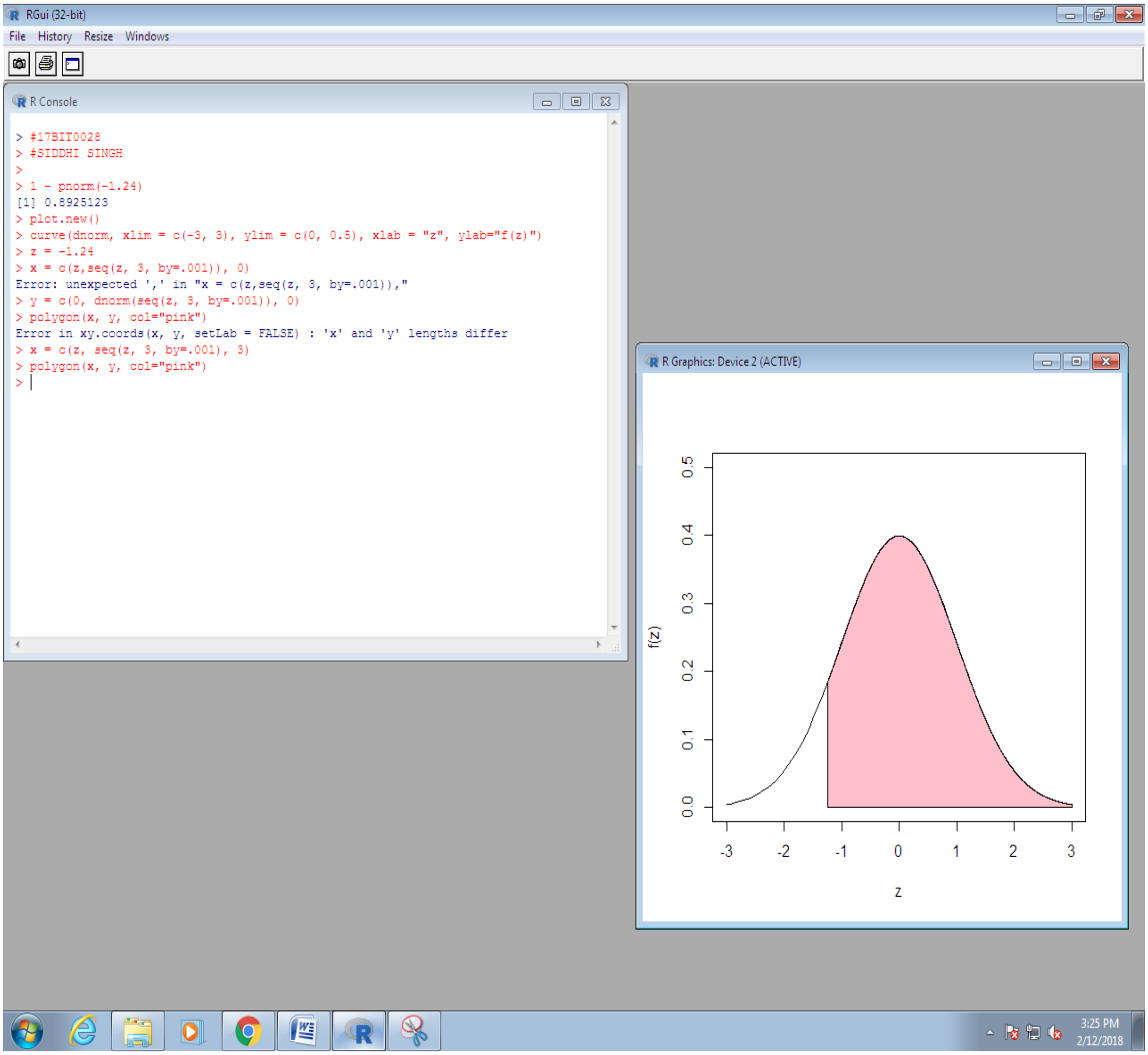


CODE 7:

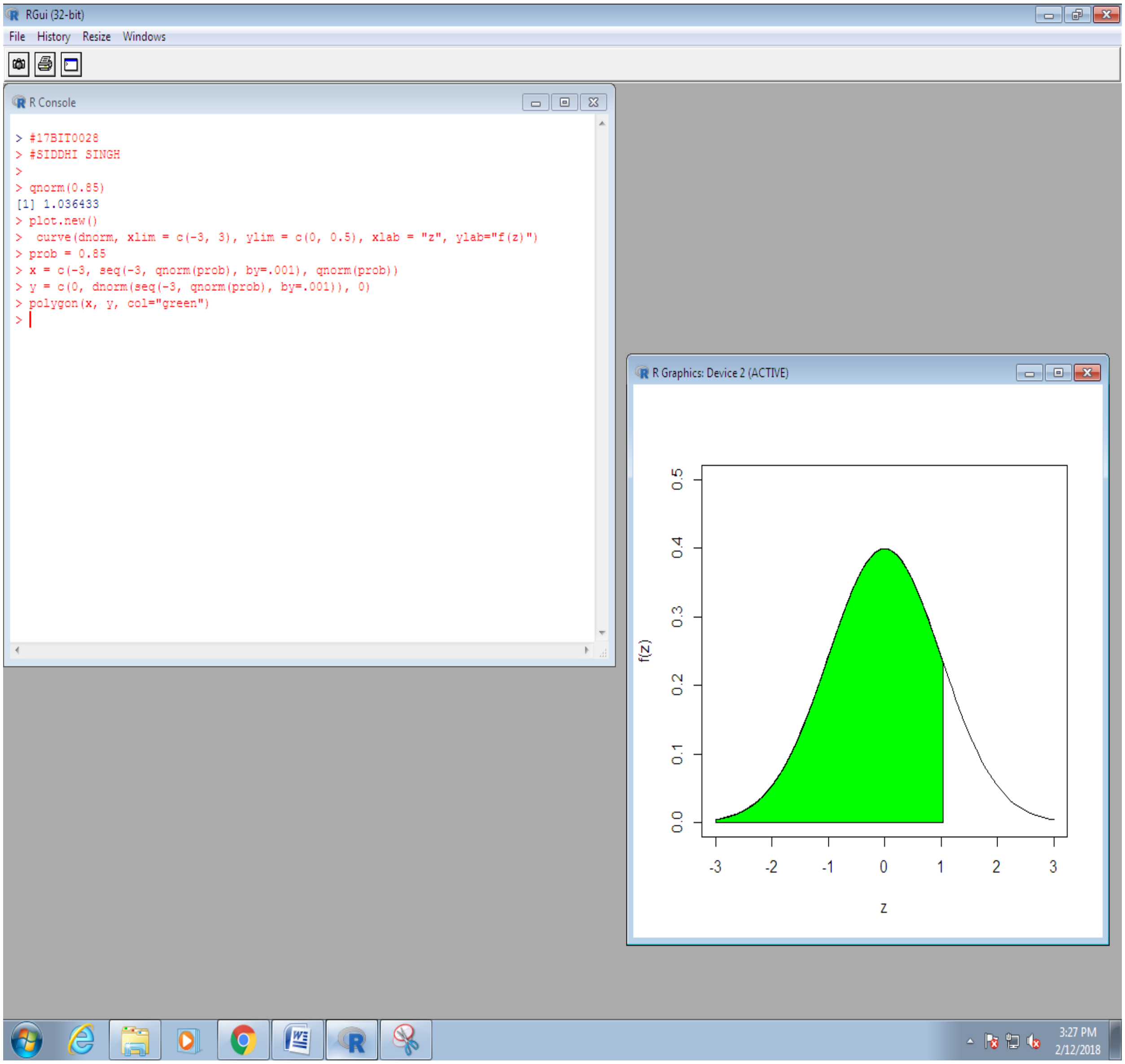
1. Find $P(0 < Z < 1.24)$



2. Find $P(Z > -1.24)$

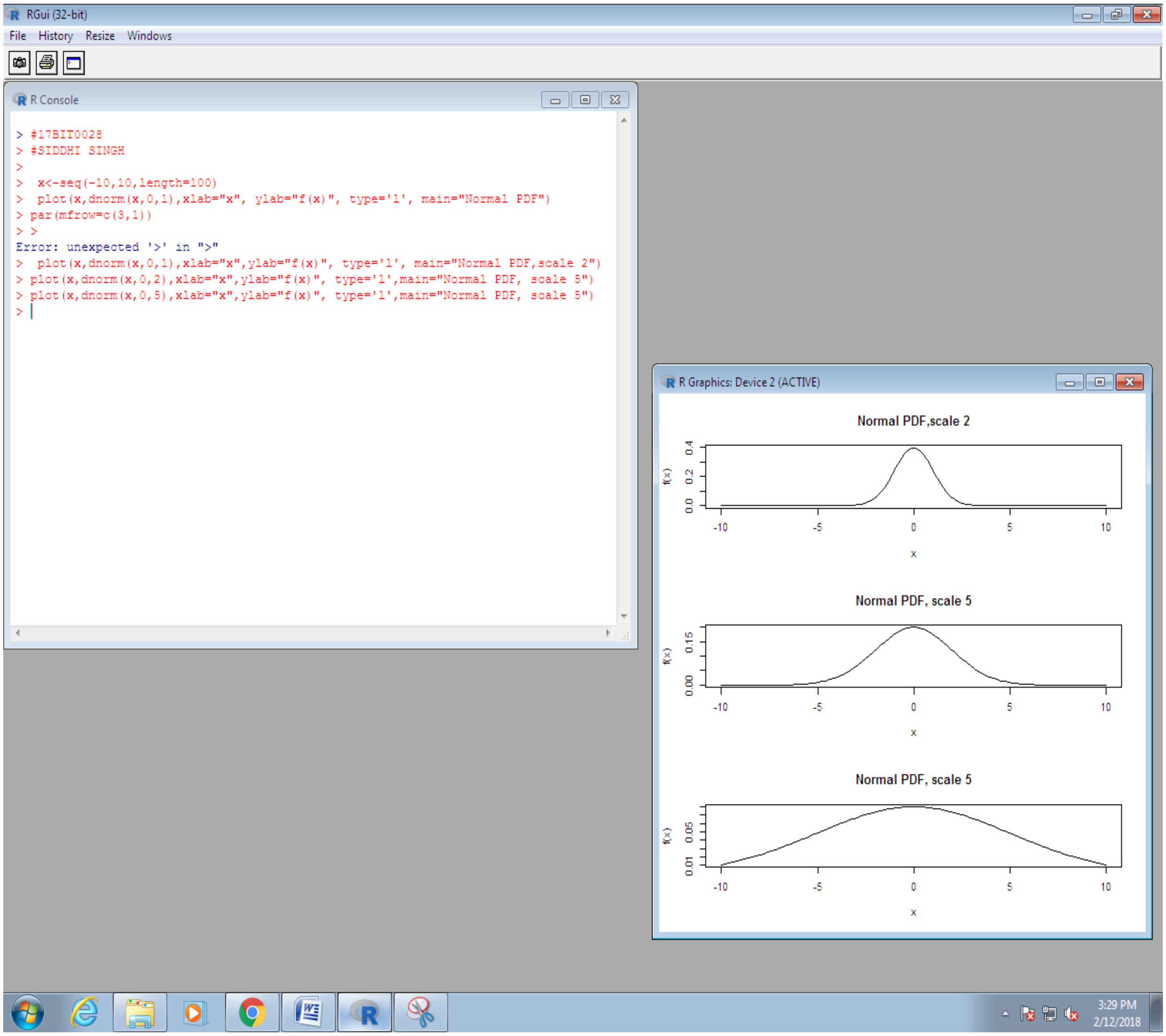

RGui (32-bit)
File History Resize Windows
R Console
> #17BIT0028
> #SIDDHI SINGH
>
> 1 - pnorm(-1.24)
[1] 0.8925123
> plot.new()
> curve(dnorm, xlim = c(-3, 3), ylim = c(0, 0.5), xlab = "z", ylab="f(z)")
> z = -1.24
> x = c(z,seq(z, 3, by=.001)), 0)
Error: unexpected ',' in "x = c(z,seq(z, 3, by=.001)),"
> y = c(0, dnorm(seq(z, 3, by=.001)), 0)
> polygon(x, y, col="pink")
Error in xy.coords(x, y, setLab = FALSE) : 'x' and 'y' lengths differ
> x = c(z, seq(z, 3, by=.001), 3)
> polygon(x, y, col="pink")
>
R Graphics: Device 2 (ACTIVE)
f(z)
0.0
0.1
0.2
0.3
0.4
0.5
-3
-2
-1
0
1
2
3
z
3:25 PM
2/12/2018

3. Find P_{85}, the 85th percentile of the standard normal "z" distribution.

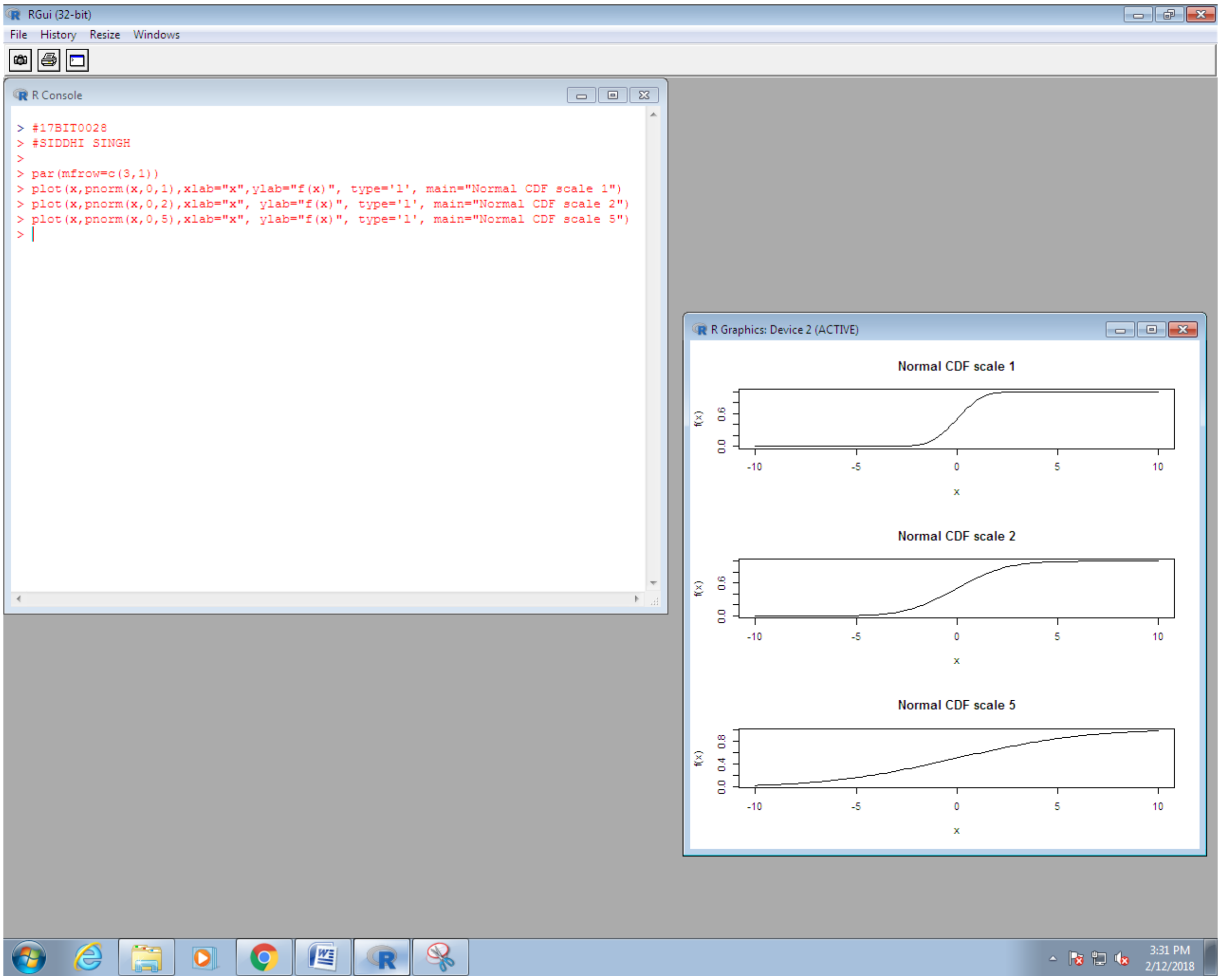


CODE 8:

(a) 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}



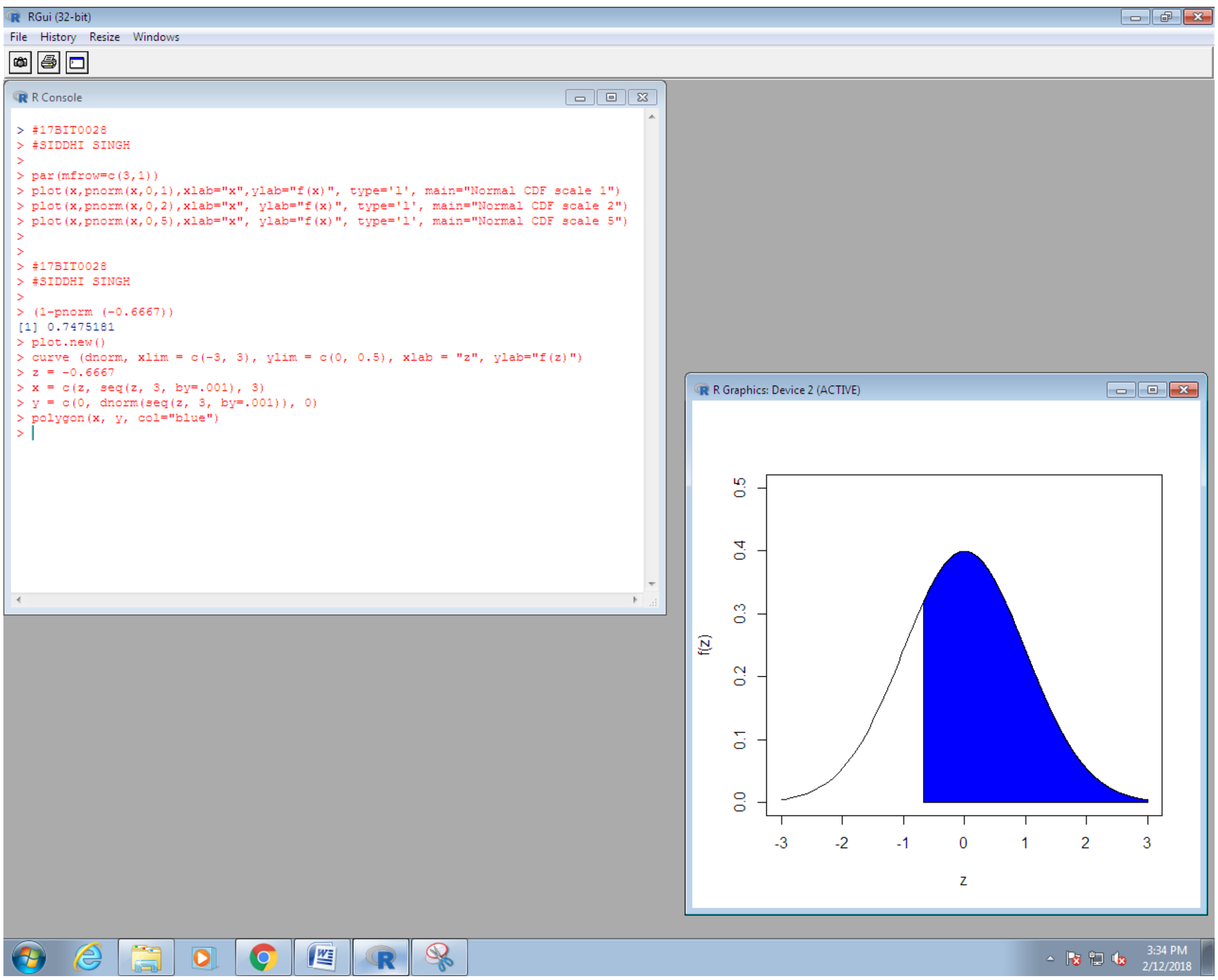
(b) Normal distribution Cumulative Distribution Function with different scale parameters

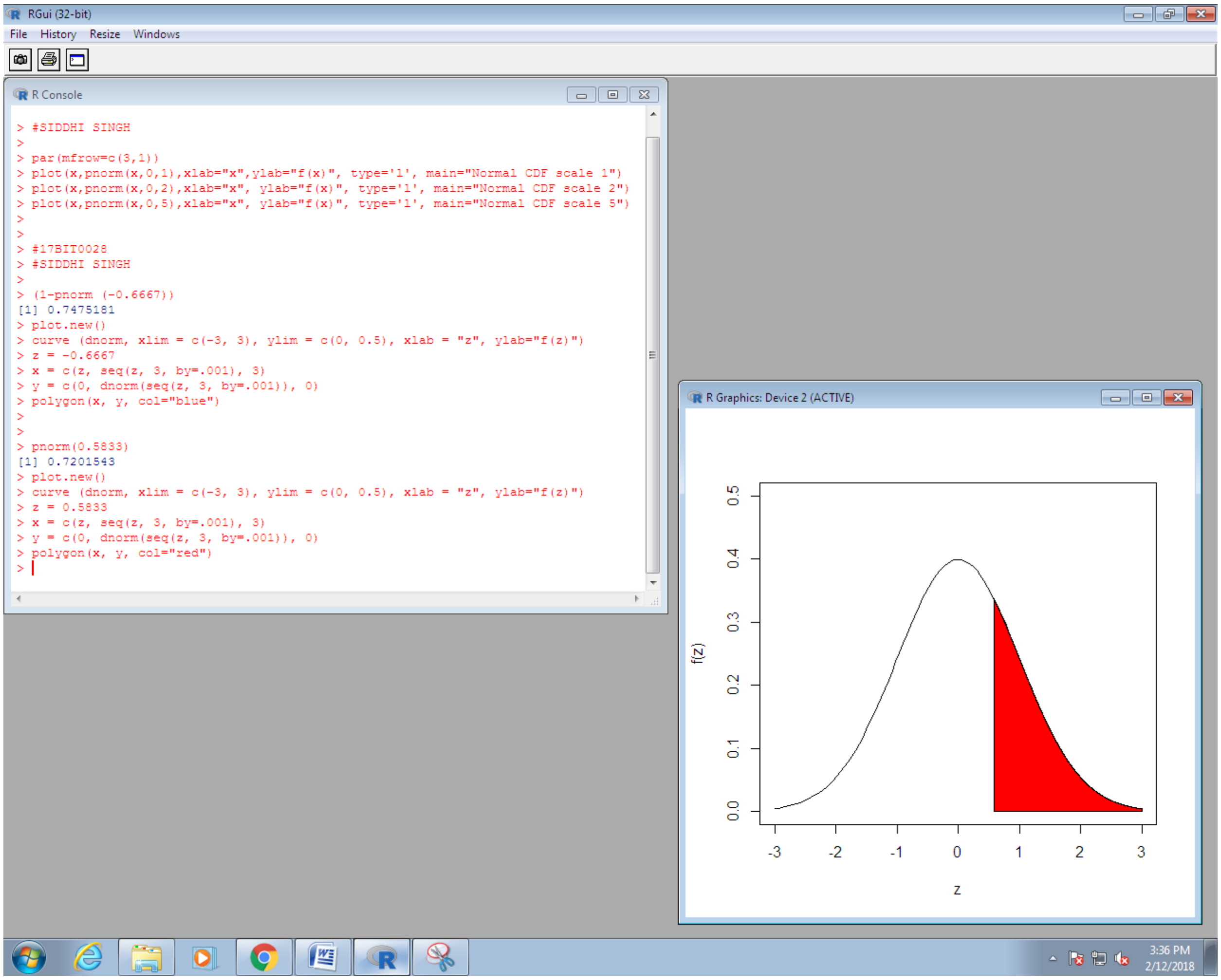


CODE 9:

Problem : 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) atmost 16.35 seconds to develop one of the prints




RGui (32-bit)
File History Resize Windows
R Console
> #SIDDHI SINGH
>
> 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")
>
>
> #17BIT0028
> #SIDDHI SINGH
>
> (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="blue")
>
>
> 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")
>
R Graphics: Device 2 (ACTIVE)
f(z)
0.0
0.1
0.2
0.3
0.4
0.5
-3
-2
-1
0
1
2
3
z
3:36 PM
2/12/2018

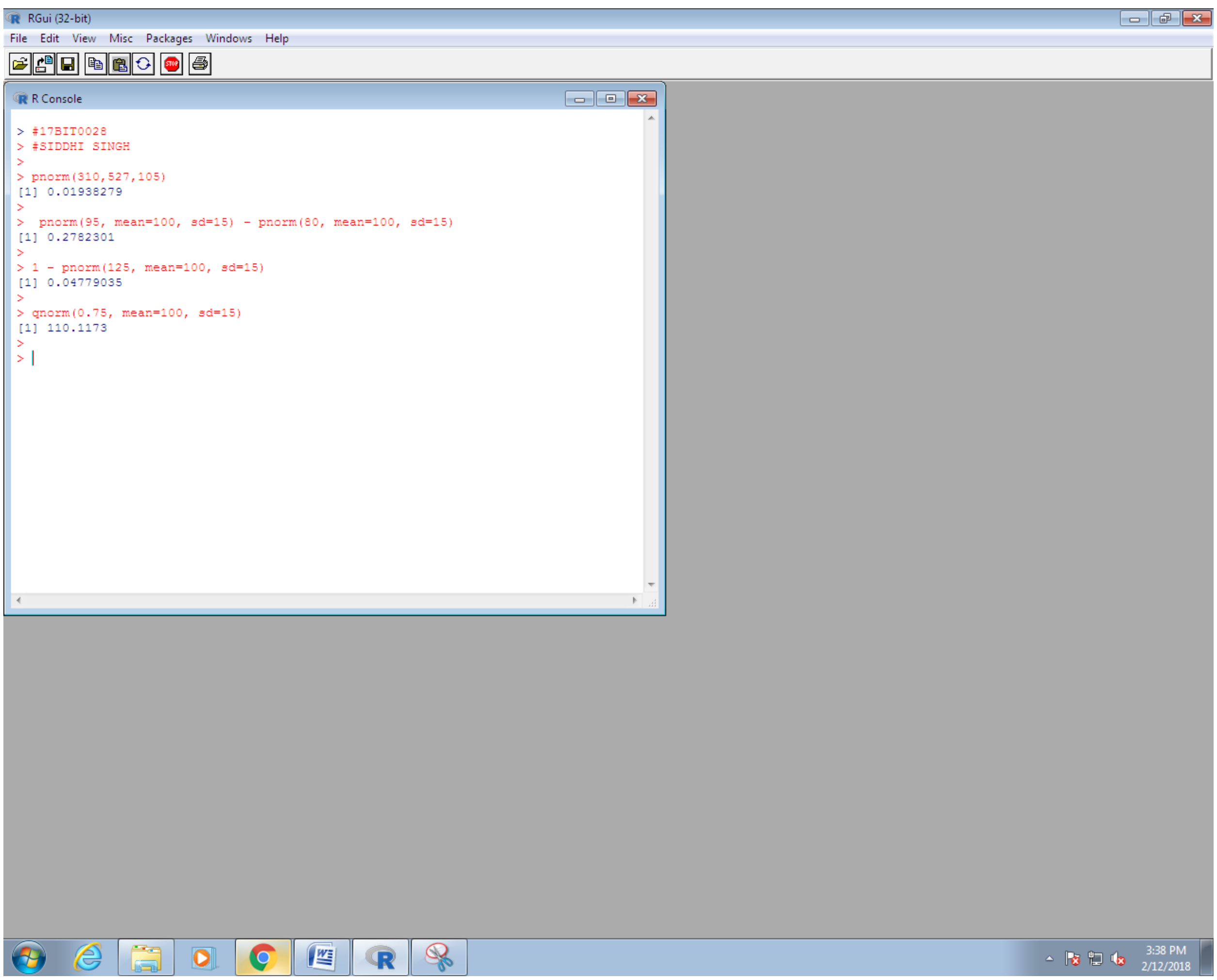
1. *Suppose X is normal with mean 527 and standard deviation 105. Compute $P(X \leq 310)$.*

2. *If* $X \sim N(\mu = 100 \text{ pts.}, \sigma = 15 \text{ pts.})$

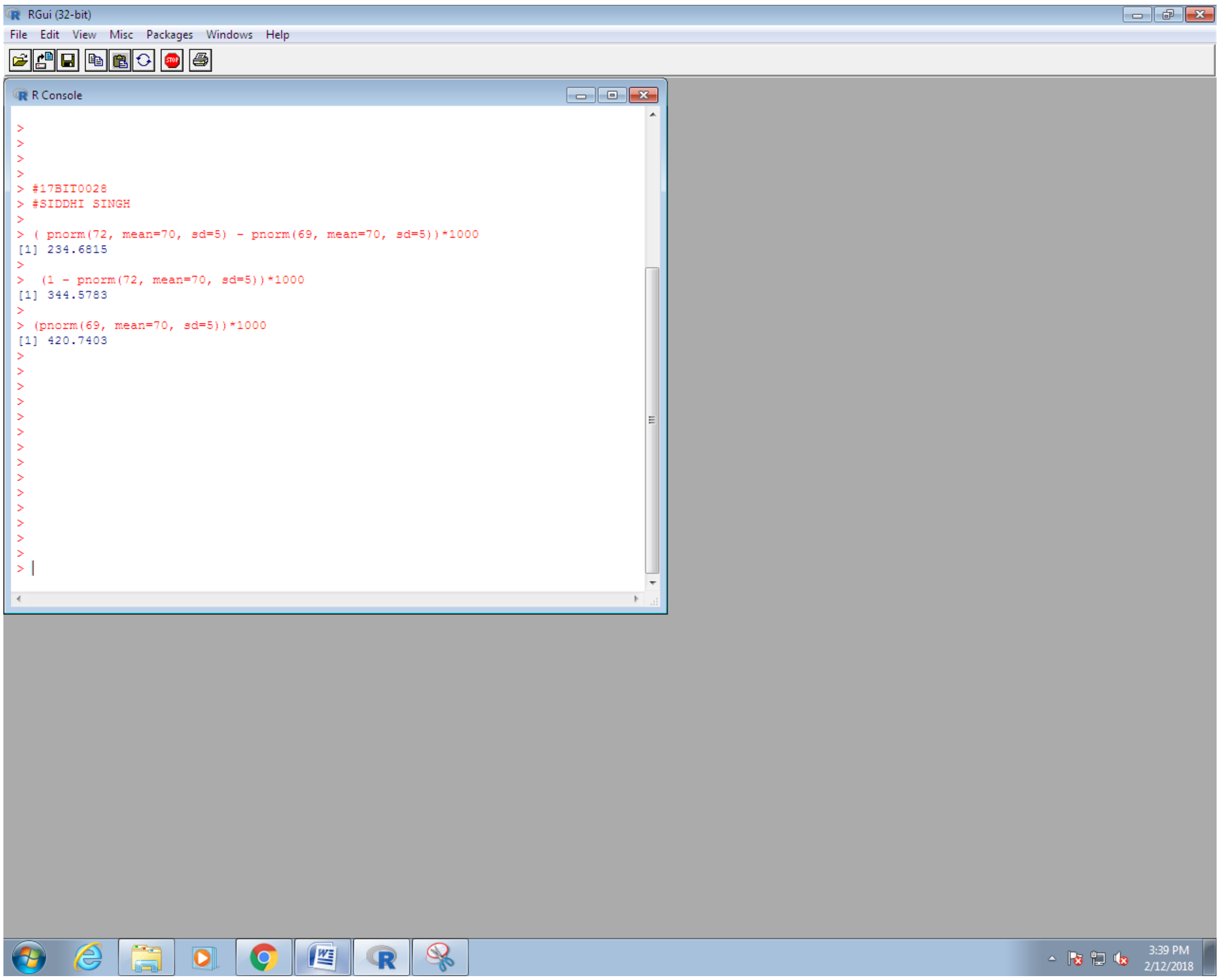
(i) Find $P(80 \text{ pts.} < X < 95 \text{ pts.})$

(ii) Find $P(X > 125 \text{ pts.})$.

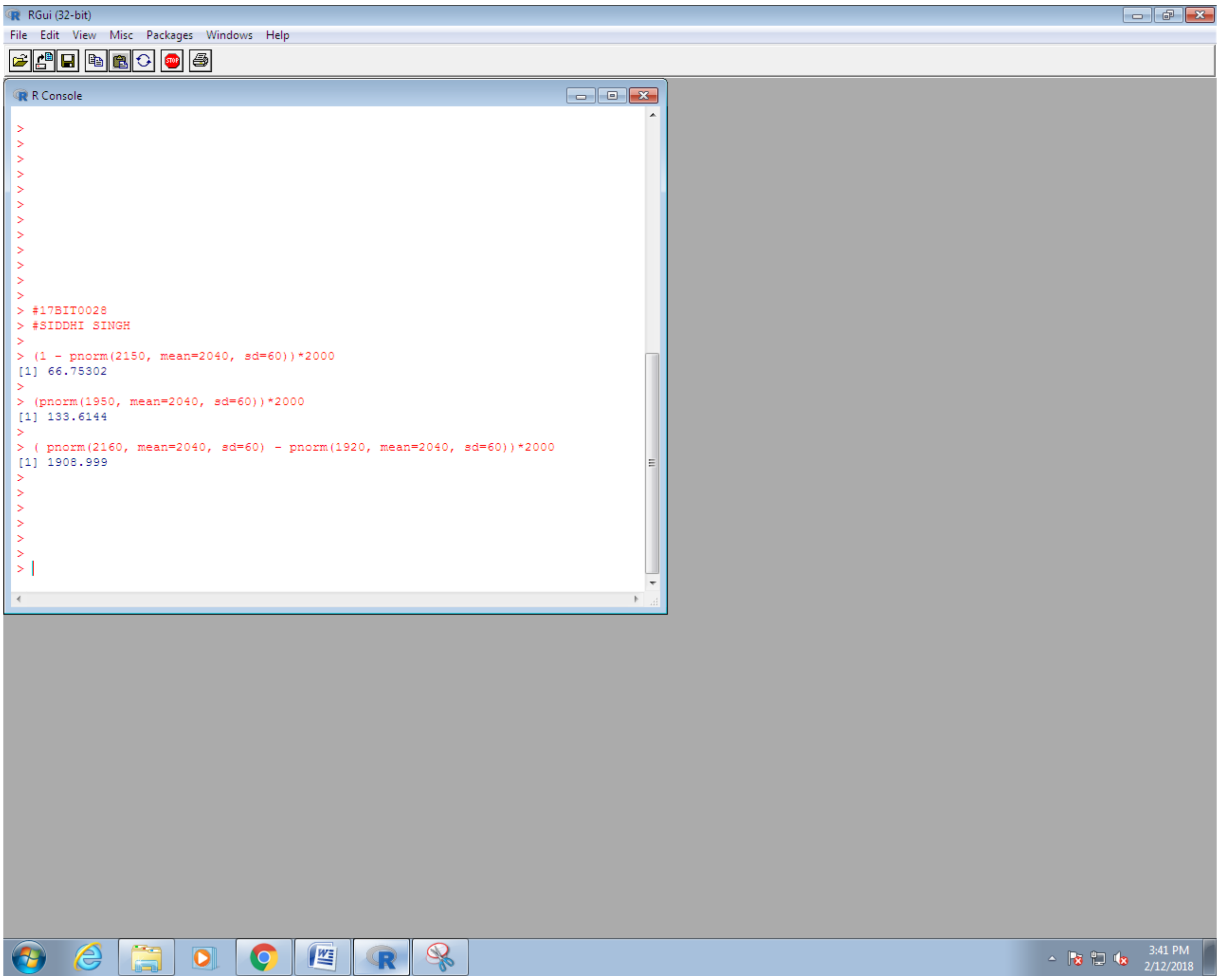
(iii) Find P_{75}, the 75_{th} percentile of the above distribution. This is the same as the 0.75 quantile.



3. *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*



4. 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
   (iv) More than 2150 hours



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