ASSIGNMENT

BUSINESS MODELING AND APPLIED ANALYTICS USING-R

CSDA304

SUBMITTED TO:

Ms. SUNU MARY ABRAHAM

DEPT. OF COMPUTER SCIENCE

SUBMITTED ON:

12-07-2023

SUBMITTED BY:

NIMSHA V S

MSC CS(DA)

Roll No:20

3RD SEMESTER

Write a note on uniform and normal distribution

- **Uniform Distribution**

Suppose the probability density function or probability distribution of a uniform distribution with a continuous random variable X is f(b)=1/y-x. In that case, It is denoted by U(x,y), where x and y are constants such that x<a<y. It is written as x~U(a,b).There are two types of uniform distribution:

I. Discrete Uniform Distribution

The discrete uniform distribution is a symmetric probability distribution in probability theory and statistics in which a limited number of values are equally likely. This implies that no new outcomes are possible.

II. Continuous Uniform Distribution

Another kind of distribution that can be uniform is one that is continuous. Any result is possible if it fits within the range, and there are endless alternative outcomes.

Characteristics

The following are the key characteristics of the uniform distribution:

- The density function integrates to unity
- Each of the inputs that go in to form the function have equal weighting
- Mean of the unifrom function is given by:

$$\mu = \frac{(a+b)}{2}$$

- The variance is given by the equation:

$$V(x) = \frac{(b-a)^2}{12}$$

- **Normal Distribution**

For a finite population the mean (m) and standard deviation (s) provide a measure of average value and degree of variation from the average value. If random samples of size n are drawn from the population, then it can be shown (the Central Limit Theorem) that the distribution of the sample means approximates that of a distribution with

mean: $\boldsymbol{\mu = m}$

standard deviation: $\boldsymbol{\sigma = \frac{s}{\sqrt{n}}}$

pdf: $\boldsymbol{f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}}$

which is called the Normal Distribution. The pdf is characterized by its "bellshaped" curve, typical of phenomena that distribute symmetrically around the mean value in decreasing numbers as one moves away from the mean. The "empirical rule" is that

- approximately 68% are in the interval [μ-σ,μ+σ]
- approximately 95% are in the interval [μ-2σ,μ+2σ]
- almost all are in the interval [μ-3σ,μ+3σ]

This says that if n is large enough, then a sample mean for the population is accurate with a high degree of confidence, since σ decreases with n.

Characteristics

- The mean, median, and mode are all equal.
- The curve is known to be symmetric at the center, which is around the mean.
- Exactly 1/2 of all the values are known to be to the left of center whereas exactly half of all the values are to the right of the center.
- The total area under the curve is 1.

Generation of Random Numbers with uniform and normal distribution in R.

i. Illustration of **sample, runif and rnorm** functions with examples

Code

```
sample(1:20,size = 2)
runif(20,min = 1,max = 20)
rnorm(10,mean = 0,sd=1)
```

Output

```
> sample(1:20,size = 2)
[1]  5 10
> runif(20,min = 1,max = 20)
 [1] 14.308676 15.453668  4.149955 12.318287  6.911597  8.396801 14.131169  4.214194  4.943534 17.132458
[11]  3.098855 17.707888  4.726092 16.072576 13.162720  7.957501  1.401159  7.140794  9.956365 13.873788
> rnorm(10,mean = 0,sd=1)
 [1] -0.17645208  1.13242927  0.30835757  0.07854171  0.21664433 -0.21449715  1.15693919  1.67387861
 [9] -0.16825966  3.09357152
```

ii. Write a R program to create a list of random numbers in normal distribution and count occurrences of each value.

Code

```
n<-rnorm(20)
n
table(n)
```

Output

```
> n<-rnorm(20)
> n
 [1] -0.08466007 -0.04289822  0.81077173  0.65187392  0.08887630 -0.90658677 -1.58608223 -0.79932406
 [9] -0.57725710 -0.43015905 -0.85102290  0.29164096 -1.80289454 -0.28448266  0.48118236 -1.41899777
[17]  0.54141040 -0.71765005 -0.40951830 -0.11547567
> table(n)
n
  -1.80289454158906   -1.58608223384325   -1.41899777379039  -0.906586770262542  -0.851022897186843
                  1                   1                   1                   1                   1
 -0.799324062126308  -0.717650053092638  -0.577257101888266  -0.430159048299786  -0.409518296752904
                  1                   1                   1                   1                   1
 -0.284482656438713  -0.115475666398708 -0.0846600664533823 -0.0428982233731585  0.0888763002235373
                  1                   1                   1                   1                   1
  0.291640959386399   0.481182363013819   0.541410401086711   0.651873922271664   0.810771729031362
                  1                   1                   1                   1                   1
```

iii. Write a R program to create a vector which contains 10 random integer values between -50 and +50

Code

```
n<-sample(-50:50,size = 10)
n
```

Output

```
> n<-sample(-50:50,size = 10)
> n
 [1] -49  10  25  45   4  23  -3  -1  19 -14
```

iv. Use the sample function to obtain a random sample of 10 realisations in a biased coin experiment

Code

sample(0:1,10,rep=T)

Output

```
> sample(0:1,10,rep=T)    #Tail=6 & Head=4
 [1] 0 0 0 0 1 1 1 1 0 0
```

v. Create a *fair dice* (with possible outcomes from 1 to 6) and determine the arithmetic mean and standard deviation of throwing it 10,000 times.

Code

t<-sample(1:6,10000,replace = TRUE)
mean(t)
sd(t)
t

Output

```
> t<-sample(1:6,10000,replace = TRUE)
> mean(t)
[1] 3.4896
> sd(t)
[1] 1.715397
```

vi. The most popular German lottery is known as 6 aus 49, in which a total of 7 numbers are randomly drawn: First, 6 unique numbers are randomly drawn out of the numbers from 1 to 49. Second, a single-digit "Superzahl" between 0 and 9. Simulate this lottery and run it once.

Code

sample(1:49,6)
sample(0:9,1)

Output

```
> sample(1:49,6)
[1] 11 19 46  7 49 34
> sample(0:9,1)
[1] 9
```

vii. Suppose we select a SRS of n = 3 balls from an urn containing a population of N = 6 balls (painted with the numbers 1, 2, 3, 4, 5, and 6). List the sample space S of all possible outcomes.

Code

```
library(combinat)
sample<-combinat::combn(1:6,3)
sample
```

Output

```
> sample<-combinat::combn(1:6,3)
> sample
     [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9]
[1,]    1    1    1    1    1    1    1    1    1
[2,]    2    2    2    2    3    3    3    4    4
[3,]    3    4    5    6    4    5    6    5    6
     [,10] [,11] [,12] [,13] [,14] [,15] [,16]
[1,]     1     2     2     2     2     2     2
[2,]     5     3     3     3     4     4     5
[3,]     6     4     5     6     5     6     6
     [,17] [,18] [,19] [,20]
[1,]     3     3     3     4
[2,]     4     4     5     5
[3,]     5     6     6     6
```

Write a note on Probability Distributions and its types.

- **Probability Distribution**

A Probability distribution is a mathematical function that estimates the likelihood that several possible outcomes and likelihoods that a random variable can take within a given range. This range will be bounded between the minimum and maximum possible values, but precisely where the possible value is likely to be plotted on the probability distribution depends on a

number of factors. These factors include the distribution's mean (average), standard deviation, skewness, and kurtosis. In terms of its sample space and event probability, it is a mathematical description of random phenomena (subsets of the sample space).

- **Types of Probability Distributions**

There are two Probability distribution types:

- Discrete Probability distribution
- Continuous Probability distribution

- **Discrete Probability distribution**

Discrete Probability distribution determines the probabilities of outcomes for discrete random variables. In other words, it aids in figuring out the probability that a random variable will take on a specific value within a predetermined range. A Discrete Probability distribution gives each discrete result of a random variable a probability. The Binomial distribution, which simulates events with two alternative outcomes, success and failure, is the most prevalent kind of Discrete Probability distribution. Bernoulli, Poisson, Geometric, and Negative Binomial distributions are some further instances of Discrete Probability distribution. A probability mass function can depict Discrete Probability distribution (PMF).

a. Binomial Distribution

It has a binary nature or has two alternative outcomes. It represents the probability distribution of the number of successful trials out of n trials with p success probabilities. Criteria of Binomial Distribution include Fixed and Independent trials, Fixed probability of success, and Two mutually exclusive outcomes.

b. Bernoulli's Distribution

The binomial distribution can be referred to as the Bernoulli distribution for n = 1 (one experiment). When n = 1, the Bernoulli distribution is frequently referred to as a particular case of the binomial distribution.

c. Poisson Distribution

The Poisson distribution gives the probability of an event happening k number of times within a given interval of time or space.
The probability density function (pdf) for this distribution is:

$$\mathbf{f(x; \lambda) = P(X = x) = \lambda x \times e\text{-}\lambda/x!}$$

λ: Average Success Rate and x=Number of success

- **Continuous Probability distribution**

Continuous Probability distribution deals with random variables that can have any continuous value within a specific range. Contrary to Discrete Random Variables, which can

have only definite, precise values, continuous random variables can take on various values. Like height, weight, and volume, continuous random variables are frequently used in Mathematics. The radioactive decay rate or sound waves' speed are two examples of physical processes often modeled using continuous probability distributions. Continuous Probability distributions come in various forms, each with its shape. The Normal bell-shaped distribution is the most prevalent. Continuous Probability distributions can represent a wide range of real-world phenomena.

a. Normal Distribution

A continuous probability distribution for a real-valued random variable. Most of the observations are centered around the central peak of this symmetric distribution, and the probability for values that are further from the mean taper off equally in both directions.
A bell-shaped density curve with a Mean and Standard deviation represents it. The Gaussian distribution is another name for it.

b. Continuous Uniform Distribution

A probability distribution with a constant probability is known as a Uniform Distribution or a rectangle distribution.
Two factors, a and b, determine this distribution:

- The minimum is a.
- The maximum is b.

The distribution is written as U(a, b).

c. Log-Normal Distribution

A probability distribution with a normally distributed logarithm is known as a lognormal (log-normal or Galton) distribution. If a random variable's logarithm has a normal distribution, it is said to be lognormally distributed.
This kind of distribution frequently fits skewed distributions with low mean values, high variation, and only positive values. Since log(x) can only exist for positive values of x, values must be positive.
The probability density function is defined by the mean μ and standard deviation, σ:
The shape of the lognormal distribution is defined by three parameters:

- σ, the shape variable. Additionally, the log-normal Standard Deviation impacts the distribution's overall form. These parameters are often known from past data. You might occasionally be able to estimate it using recent data. The location and height are unaffected by the shape parameter.
- m, the scale parameter (this is also the median). This parameter shrinks or stretches the graph.
- Θ (or μ), the location parameter, tells you where the graph is on the x-axis.

d. Exponential Distribution

The exponential distribution is a Continuous Probability distribution used in statistics that frequently deals with how long until a particular event occurs. Events occur continually, independently, and at a steady average pace during this process. The crucial characteristic of the exponential distribution is that it has no memory.

The continuous random variable, say X, is said to have an Exponential distribution if it has the following Probability density function:

Where λ is called the distribution rate and the mean of the Exponential distribution is 1/λ, and variance is 1/λ2, and the memoryless quality of the exponential distribution is its most significant characteristic.

The density functions of exponential distributions concerning different parameters λ

Probability Distributions: Demonstration of CDF and PDF uniform and normal, binomial & Poisson distributions in R.

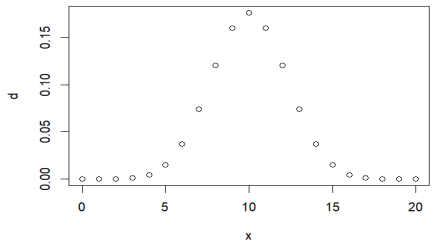
i. Illustration of PDF & CDF functions of normal, binomial & Poisson distributions.

Binomial

Code : PDF

```
x<-seq(0,20,by=1)
d<-dbinom(x,20,0.5)
plot(x,d)
```

Output

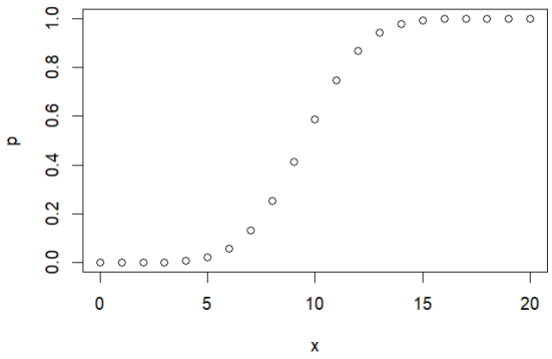


Code : CDF

```
x<-seq(0,20,by=1)
p<-pbinom(x,20,0.5)
```

```
plot(x,p)
```

Output

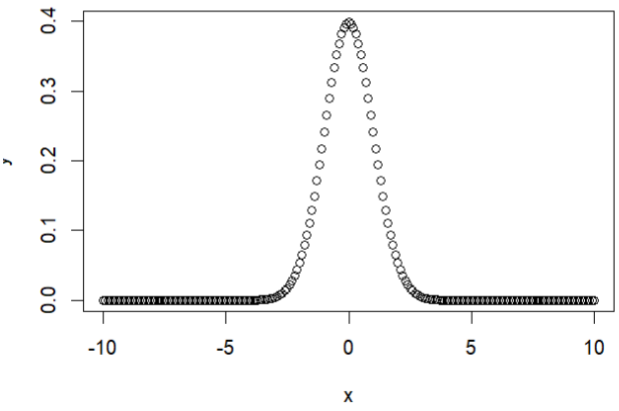


Normal

Code : PDF

```
x<-seq(-10,10,by=0.1)
y<-dnorm(x,mean = 0,sd=1)
plot(x,y)
```

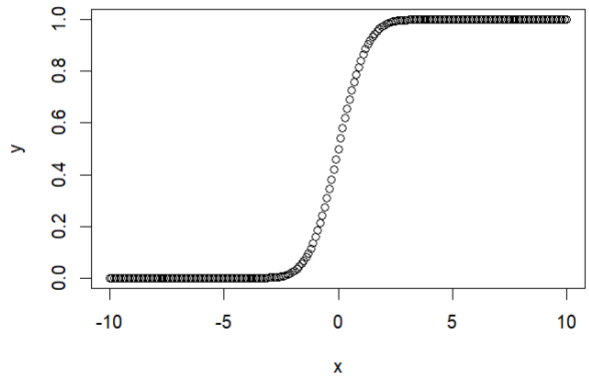
Output



Code : CDF

```
x<-seq(-10,10,by=0.1)
y<-pnorm(x,mean = 0,sd=1)
plot(x,y)
```

Output

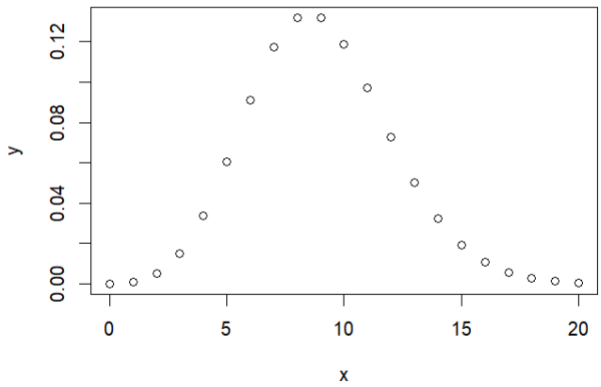


Poisson

Code : PDF

```
x<-seq(0,20)
y<-dpois(x,lambda = 9)
plot(x,y)
```

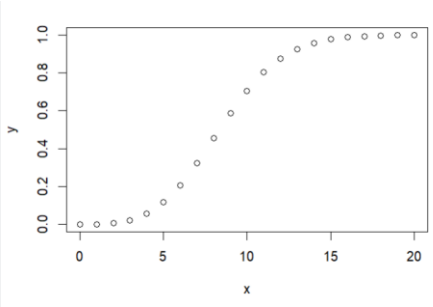
Output



Code : CDF

```
x<-seq(0,20)
y<-ppois(x,lambda = 9)
plot(x,y)
```

Output



ii. Calculate the following probabilities:

Probability that a normal random variable with mean 22 and variance 25
(i)lies between 16.2 and 27.5

Code

```
s<-sqrt(25)
pnorm(27.5,mean = 22,sd=s) - pnorm(16.2,mean = 22,sd=s)
```

Output

```
> pnorm(27.5,mean = 22,sd=s) - pnorm(16.2,mean = 22,sd=s)
[1] 0.7413095
```

(ii)is greater than 29

Code

```
1-pnorm(29,mean = 22,sd=s)
```

Output

```
> 1-pnorm(29,mean = 22,sd=s)
[1] 0.08075666
```

(iii)is less than 17

Code

```
pnorm(17,mean = 22,sd=s)
```

Output

```
> pnorm(17,mean = 22,sd=s)
[1] 0.1586553
```

(iv)is less than 15 or greater than 25

Code

(1-pnorm(25,mean = 22,sd=s)) + pnorm(15,mean = 22,sd=s)

Output

```
> (1-pnorm(25,mean = 22,sd=s)) + pnorm(15,mean = 22,sd=s)
[1] 0.3550098
```

Probability that in 60 tosses of a fair coin the head comes up

(i) 20,25 or 30 times

Code

dbinom(20,size = 60,prob = 0.5) + dbinom(25,size = 60,prob = 0.5) + dbinom(30,size = 60,prob = 0.5)

Output

```
[1] 0.1512435
```

(ii) less than 20 times

Code

dbinom(20,size = 60,prob = 0.5)

Output

```
[1] 0.003635846
```

(iii)between 20 and 30 times

Code

```
dbinom(30,size = 60,prob = 0.5) - dbinom(20,size = 60,prob = 0.5)
```

Output

```
[1] 0.09894233
```

A random variable X has Poisson distribution with mean 7. Find the probability that

(i)X is less than 5

Code

```
ppois(5,lambda = 7)
```

Output

```
[1] 0.3007083
```

(ii)X is greater than 10 (strictly)

Code

```
1-ppois(10,lambda = 7)
```

Output

```
[1] 0.09852079
```

(iii)X is between 4 and 16

Code

```
ppois(16,lambda = 7) - ppois(4,lambda = 7)
```

Output

```
[1] 0.8260502
```

iii. Simulate normal distribution values. Imagine a population in which the average height is 1.70 m with a standard deviation of 0.1. Use rnorm to simulate the height of 1000 people and save it in an object called heights.

Code

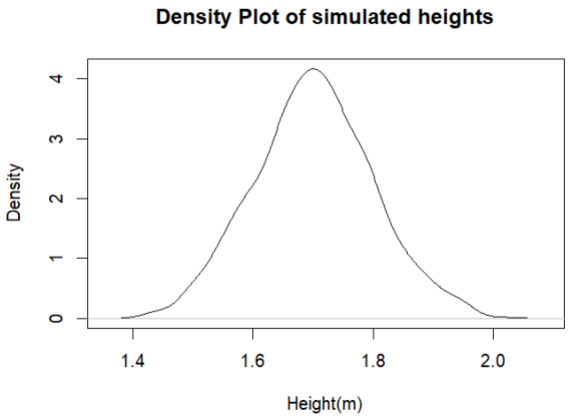
```
set.seed(123)
height<-rnorm(1000,mean = 1.70,sd=0.1)
```

a) Plot the density of the simulated values.

Code

```
plot(density(height),main = "Density Plot of simulated heights",xlab =
"Height(m)")
```

Output

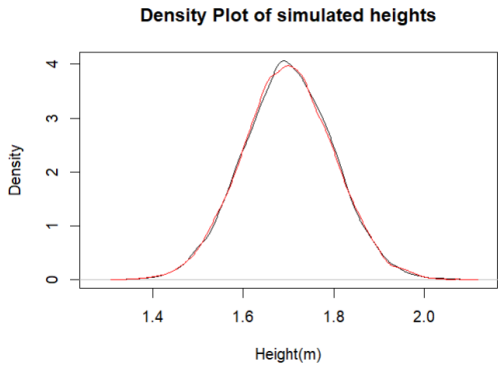


b) Generate 10000 values with the same parameters and plot the respective density function on top of the previous plot in red to differentiate it.

Code

```
height<-rnorm(10000,mean = 1.70,sd=0.1)
lines(density(height), col = "red")
```

Output



iv. You roll a die 100 times and get just 10 sixes?

- What is the probability of getting just 10 sixes?

Code

```
prob<-dbinom(10,size = 100,prob = 1/6)
prob
```

Output

```
[1] 0.02140327
```

- What is the probability of getting 10 or fewer sixes?

Code

```
prob_1<-pbinom(10,size = 100,prob = 1/6)
prob_1
```

Output

```
[1] 0.04269568
```

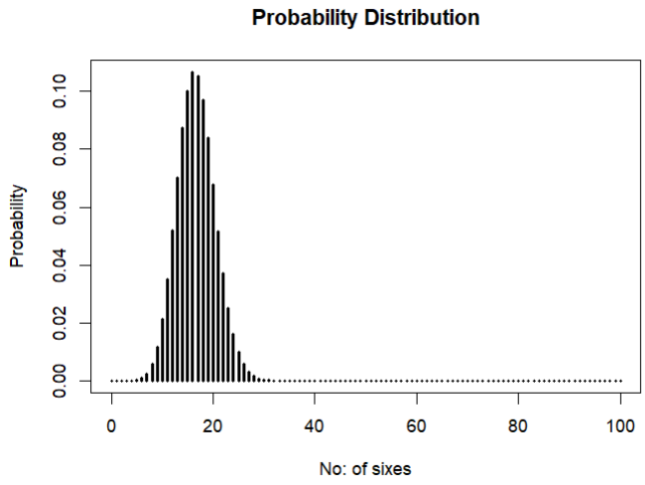
- Draw the probability distribution.

Code

```
x<-0:100
prob_dist<-dbinom(x,size = 100,prob = 1/6)
```

plot(x,prob_dist,type = "h",lwd=3,xlab = "No: of sixes",ylab = "Probability",main = "Probability Distribution")

Output

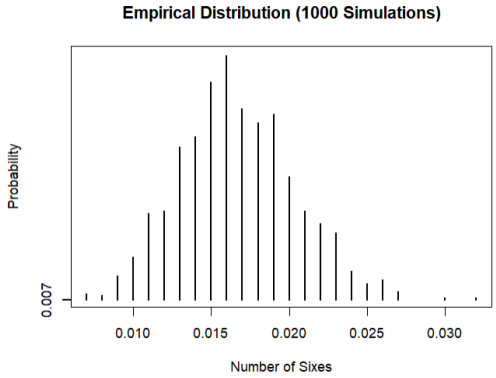


- Simulate the described experiment 1000 times and compute the empirical distribution. Compare it to the theoretical one. Then do the same with 1,000,000 simulations.

Code

s<-rbinom(1000,size = 100,prob = 1/6)
emp<-table(s/1000)
plot(as.numeric(names(emp)), emp, type = "h", lwd = 2, xlab = "Number of Sixes", ylab = "Probability", main = "Empirical Distribution (1000 Simulations)")

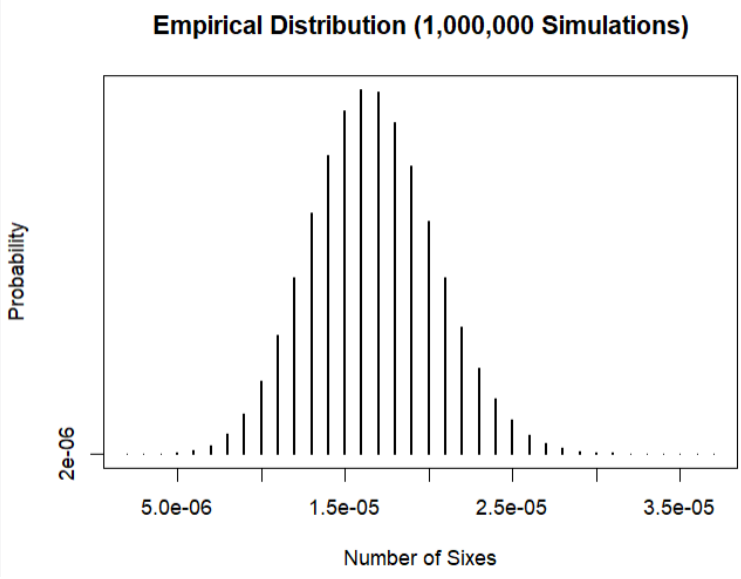
Output



Code

```
s<-rbinom(1000000,size = 100,prob = 1/6)
emp<-table(s/1000000)
plot(as.numeric(names(emp)), emp, type = "h", lwd = 2, xlab = "Number of Sixes", ylab = "Probability", main = "Empirical Distribution (1,000,000 Simulations)")
```

Output



v. Using the function rbinom to generate 10 unfair coin tosses with probability success of 0.3. Set the seed to 1.

Code

```
set.seed(1)
rbinom(10,size=1,prob = 0.3)
```

Output

```
[1] 0 0 0 1 0 1 1 0 0 0
```

vi. Simulate normal distribution values. Imagine a population in which the average height is 1.70 m with an standard deviation of 0.1, using rnorm simulate the height of 100 people and save it in an object called heights.

Code

```
set.seed(1)
height<-rnorm(100,mean = 1.70,sd=0.1)
```

a) What's the probability that a person will be smaller or equal to 1.90 m ?

Code

```
pnorm(1.90,mean = 1.70,sd=0.1)
```

Output

```
[1] 0.9772499
```

b) What's the probability that a person will be taller or equal to 1.60 m?

Code

```
1-pnorm(1.60,mean = 1.70,sd=0.1)
```

Output

```
[1] 0.8413447
```