Experiment – 6

Normal distribution, Poisson distribution

Aim: To understand Poisson distribution and Normal distribution using R functions

Introduction:

A discrete distribution is one in which the data can only take on certain values, for example integers. For a discrete distribution, probabilities can be assigned to the values in the distribution. These distributions model the probabilities of random variables that can have discrete values as outcomes. Example: Binomial distribution, Poisson distribution

Poisson distribution is a discrete probability distribution that is widely used in the field of finance. It gives the probability that a given number of events will take place within a fixed time period. The notation is written as X ~ Pois(λ), where λ>0. The pmf is given by the following formula:

$$P(X = x) = \frac{e^{-\lambda}\lambda^{x}}{x!}, \qquad x = 0,1,2,\ldots$$

Procedure:

- Import the data set
- Determine the probabilities of the random variable using Poisson distribution in R
- Visualize the probability distribution using R functions

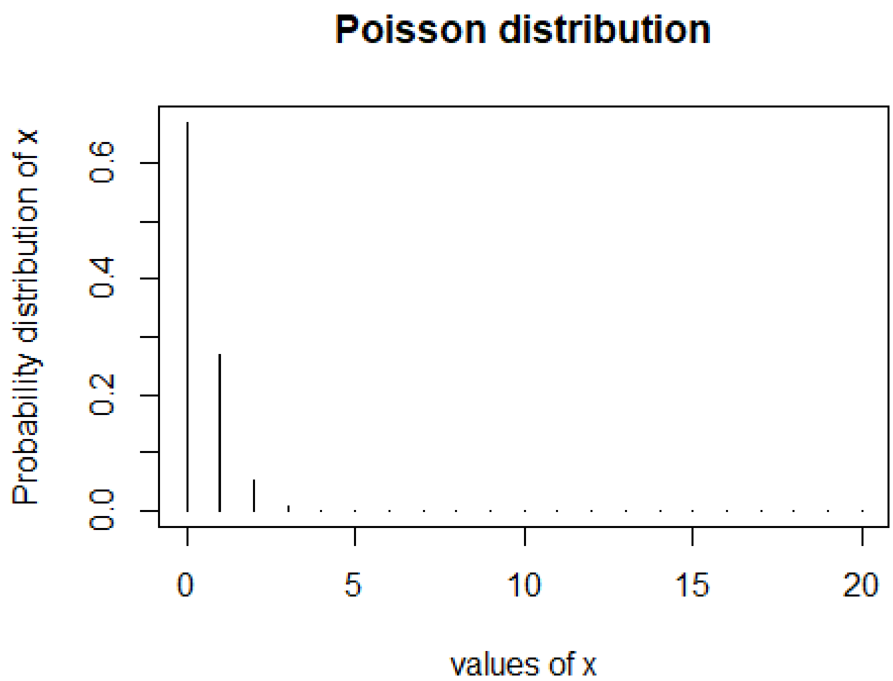
Problem:

A manufacturer of pins knows that 2% of his products are defective. If he sells pins in boxes of 20 and find the number of boxes containing (i) at least 2 defective (ii) exactly 2 defective (iii) at most 2 defective pins in a consignment of 1000 boxes (iv) plot the distribution (v) E(x) (vi) Variance of X?

Codes and Results:

```
#Poisson Distribution
# number of trails
m=20
m

## [1] 20

# probability of success
ps=0.02
# poisson parameter
lambda=m*ps
lambda
```

```
## [1] 0.4

#at least 2 defectives
p1=sum(dpois(2:m,lambda))
p1

## [1] 0.06155194

# (i) number of boxes containing at least 2 defectives
round(1000*p1)

## [1] 62

#exactly 2 defectives
p2=dpois(2,lambda)
p2

## [1] 0.0536256

# (ii) number of boxes containing exactly 2 defectives
round(1000*p2)

## [1] 54

#at most 2 defectives
p3=sum(dpois(0:2, lambda))
p3

## [1] 0.9920737

# (iii) number of boxes containing at most 2 defectives
round(1000*p3)

## [1] 992

# (iv) plot the distribution
x1=0:m
px1=dpois(x1,lambda)
plot(x1,px1,type="h",xlab="values of x",ylab="Probability distribution of
x",main="Poisson distribution")
```




```
#(v) E(x)
Ex1=weighted.mean(x1,px1)
Ex1

## [1] 0.4

# (vi) variance of x
Varx1=weighted.mean(x1*x1,px1)-(weighted.mean(x1 ,px1))^2
Varx1

## [1] 0.4
```

Normal Distribution

The Normal Distribution is defined by the probability density function for a continuous random variable in a system. Let us say, f(x) is the probability density function and X is the random variable.
$f(x) \geq 0$ for all $x \in (-\infty, \infty)$ and $\int_{-\infty}^{\infty} f(x)dx = 1$

The probability density function of normal or Gaussian distribution is given by;

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

Where,
x is the variable
μ is the mean
σ is the standard deviation

Procedure:

- Generating the data set
- Determine the probabilities of the random variable using Normal distribution in R
- Visualize the probability distribution using R functions

Problem:
A company finds that the time taken by one of its engineers to complete or repair job has a normal distribution with mean 20 minutes and S.D 5 minutes. State what proportion of jobs take:

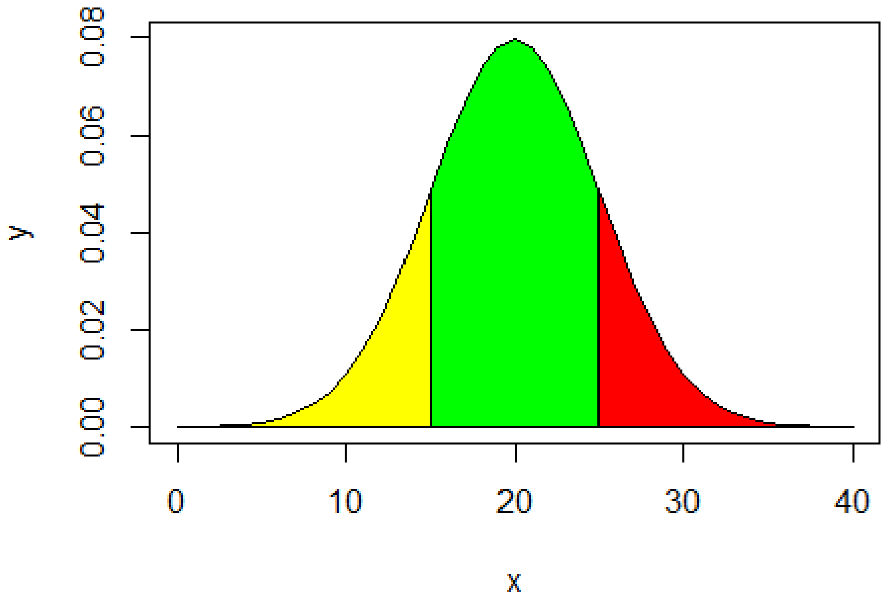
i. Less than 15 minutes
ii. More than 25 minutes
iii. Between 15 and 25 minutes
iv. Plot the distribution
v. Table the distribution

Code and Results:

```
# Generating the data x
x=seq(0,40)
x

##  [1]  0  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22
23 24
## [26] 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40

# find the density function of x
y=dnorm(x,mean=20,sd=5)
y

##  [1] 2.676605e-05 5.838939e-05 1.223804e-04 2.464438e-04 4.768176e-04
##  [6] 8.863697e-04 1.583090e-03 2.716594e-03 4.478906e-03 7.094919e-03
## [11] 1.079819e-02 1.579003e-02 2.218417e-02 2.994549e-02 3.883721e-02
## [16] 4.839414e-02 5.793831e-02 6.664492e-02 7.365403e-02 7.820854e-02
## [21] 7.978846e-02 7.820854e-02 7.365403e-02 6.664492e-02 5.793831e-02
## [26] 4.839414e-02 3.883721e-02 2.994549e-02 2.218417e-02 1.579003e-02
## [31] 1.079819e-02 7.094919e-03 4.478906e-03 2.716594e-03 1.583090e-03
## [36] 8.863697e-04 4.768176e-04 2.464438e-04 1.223804e-04 5.838939e-05
## [41] 2.676605e-05

# plot the normal distribution curve
plot(x,y,type='l')
# Proportion of jobs take less than 15 minutes
p1=pnorm(15,mean=20,sd=5)
p1
```

```
## [1] 0.1586553

x2=seq(0,15)
x2

##  [1]  0  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15

y2=dnorm(x2,mean=20,sd=5)
y2

##  [1] 2.676605e-05 5.838939e-05 1.223804e-04 2.464438e-04 4.768176e-04
##  [6] 8.863697e-04 1.583090e-03 2.716594e-03 4.478906e-03 7.094919e-03
## [11] 1.079819e-02 1.579003e-02 2.218417e-02 2.994549e-02 3.883721e-02
## [16] 4.839414e-02

polygon(c(0,x2,15),c(0,y2,0),col='yellow')

#Proportion of jobs take more than 25 minutes
p2=pnorm(40,mean=20,sd=5)-pnorm(25,mean=20,sd=5)
p2

## [1] 0.1586236

x1=seq(25,40)
x1

##  [1] 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40

y1=dnorm(x1,mean=20,sd=5)
y1

##  [1] 4.839414e-02 3.883721e-02 2.994549e-02 2.218417e-02 1.579003e-02
##  [6] 1.079819e-02 7.094919e-03 4.478906e-03 2.716594e-03 1.583090e-03
## [11] 8.863697e-04 4.768176e-04 2.464438e-04 1.223804e-04 5.838939e-05
## [16] 2.676605e-05

polygon(c(25,x1,40),c(0,y1,0),col='red')

#Proportion of jobs take between 15 and 25 minutes
p3=pnorm(25,mean=20,sd=5)-pnorm(15,mean=20,sd=5)
p3

## [1] 0.6826895

x3=seq(15,25)
x3

##  [1] 15 16 17 18 19 20 21 22 23 24 25

y3=dnorm(x3,mean=20,sd=5)
y3
```

```
##  [1] 0.04839414 0.05793831 0.06664492 0.07365403 0.07820854 0.07978846
##  [7] 0.07820854 0.07365403 0.06664492 0.05793831 0.04839414

polygon(c(15,x3,25),c(0,y3,0),col='green')
```




```
# Probability distribution
data.frame(p1,p2,p3)

##          p1        p2        p3
## 1 0.1586553 0.1586236 0.6826895
```

Conclusion: Poisson distribution and Normal distribution have been explored using R.