Statistical Distribution in R

R functions produce 4 important values for commonly statistical distributions. The four important functions are:

1. The density function (d)
   Note:
   There is a difference between function (d) in discrete and continues distribution such that:

discrete distribution	continues distribution
function (d) means density at the point	function (d) means height at the point

2. The probability function (p) – or cumulative density function $P(X \leq x)=F(x)$
3. The quantile function (q) – inverse of the probability function $p(q(x))=x$ & $q(p(x))=x$.
4. The random number generation function (r) – generate random numbers from specified distribution.

For a normal distribution

To compute density at x

dnorm(x, mean=0, sd=1) , x is a vector of quantiles.

To compute the cumulative

pnorm (q, mean=0, sd=1) , q vector of quantiles.

To compute the pth quantile – inverse of the prob. function

qnorm(p, mean=0, sd=1) , p vector of probabilities.

To generate a random sample of size n from normal distribution

rnorm (n, mean=0, sd=1) , n sample size

A powerful feature of the R language is the ability to generate random numbers from a specified distribution. For example, rnorm (10) generates 10 random numbers from the standard normal distribution; rnorm (10, 12, 3) generates them from N (12, 9)

The quantile x is the value such that
p (random variable $\leq$ x) = F(x) = p(probability).
pnorm(1) ; pnorm(1.96) ; pnorm(1.64) ; pnorm(-.5 : .5)
pnorm(seq(-2, 2, 1);pnorm (2, 0, 2); pnorm (1:5, 3,1.5)

Geometrically pnorm is the area of pdf to the left of the value x under the curve in standard normal distribution.
qnorm (p) is the inverse function of (p) which gives you the quantile x.

Geometrically qnorm is the value associated with the area size p from standard normal distribution.

Examples:

rnorm(10) # generate 10 numbers from normal(0,1)
rnorm(10,12,3) # generate 10 numbers from normal(12,9)
pnorm(1);pnorm(1.96);pnorm(-.5:.5); pnorm(seq(-2,2,1))
pnorm(2,0,2); pnorm(1:5,3,1.5)
qnorm(0.8);qnorm(0.975);qnorm(0.95); qnorm(0.05);qnorm(0.025)

Note that

f (f (p)) = p
check this by
pnorm(qnorm(c(0.9,0.95,0.975,0.99))) # give prob back.

The same way we can use the four functions (d,p,q,r) for the other distributions in R. A command are invoked by placing one of four symbols (p, q, d, r) on the front of R root name associated with a specific distribution.

Distribution	R root name	Parameters	
Normal	norm	Mean = 0	sd =1
Student's t	t	df	
Chi-square	chisq	df	
F	F	df1	df2
Gamma	gamma	shape	
Beta	beta	shape1	shape2
Uniform	unif	min=0	max=1
Lognormal	lnorm	meanlog=0	sdlog=1
Logistic	logis	location=0	scale=1
Cauchy	cauchy	location=0	scale=1
Exponential	exp	rate=1	
Binomial	binom	size	probability
Poisson	pois	lambda	
Weibull	weibull	shape	

- A special distributions the uniform distribution.

Examples:

runif (12) #generates 12 random numbers from U(0, 1)

- To generate 12 random numbers equally likely between 0 and 10, then
  10*runif (12) # or runif (12, 0, 10)

- To generate random integers between 1, 10
  ceiling (runif (12, 0, 10))

- To fix the random seed to regenerate same sequence of random numbers, you can use
  set.seed (111)

More examples:

```
qt(p=0.975,df=9)  #the 5% critical value for a two sided t-test on 9 d.f.
dpois(x=3,lambda=5) # the value of density of poisson dist. With rate =5
dnorm(-2:2,2,2) # find density at ……. From Normal(   ,     )
dnorm(qnorm(c(0.05,0.1,0.9,0.95))  )  #to check symmetry of distribution
```

#Generate a specified number of random numbers from a given distribution

```
my.ran<-function(n,distribution,shape){
if (distribution=="gamma") rgamma(n,shape) else
if (distribution=="exp") rexp(n) else
if (distribution=="norm") rnorm(n) else
print("unknown distribution")
}
distribution="norm"
my.ran(20,distribution)
```

Questions:

1. find probability that x=6 where x is poisson(lambda=4)
2. find probability that x less than or equal to 6 where x is Normal(15,16)
3. find probability that P(X≤x)=0.5 from uniform(2,8)
4. find probability that P(X≤x)=0.5,0.95,0.65 from t(2,8)
5. Explore the following:

```
n=1000
xnorm=rnorm(1000)
stem(xnorm)
hist(xnorm)
plot(sort(xnorm),(1:n)/n,type="s",ylim=c(0,2))
boxplot(xnorm)
mean(xnorm)
median(xnorm)
range(xnorm)
quantile(xnorm)
sqrt(var(xnorm))
```

Graphics In R

One of the strongest capabilities and most attractive features.

Graphical data exploration function:

R function	Description
boxplot **Single group** **Grouped data**	**Boxplot chart** **Boxplot(y~x) y described using x** **Example: boxplot(mpg~cyl)** **Boxplot(x,y)** **Example: boxplot(mpg,qsec)**
stem	**Leaf and Stem plot**
dotchart	**Dot chart**
hist (S&G)	**Histogram**
Pie	**Pie chart**
qqnorm	**Quantile-quantile plot for Normal distribution**
qqplot(G)	**Quantile-quantile plot**
Barplot(table) **Catogrical data**	**Creates a bar graph** **table(gear,cyl)** **barplot(gc,col=c(1,2,3))**

Examples

```
x=rnorm(100)
qqnorm(x)
```

#A sample from 30 men and 50 women. Create a pie and bar charts for this sample

#pie(frequency,name of categories)
#barplot(frequency,name of categories)

```
pie(c(30,50), labels=c("men","women"))
pie(c(30,50), labels =c("men","women"),col=3:4)

barplot(c(30,50),name=c("men","women"))
```

The plot command:

It is one of the most important command to create graphics.

Syntax : plot(vector x, . . .),

Where . . . stands for options, some of which include the following

Argument	Description
type=	plot type. " p " for points, "l" for lined, "b" for both, "o" for overlaid, "n" for nothing, "s" for stairstep, and "h" for height bars.
pch=	plot characters at the points. Square(0); octagon(1); triangle(2); cross(3); x(4); diamond (5) and inverted triangle(6) or "character"
lty=	line type. 1 for solid, 2 for dotted, 3 for small breaks, etc
lwd=	line width, 1=default, 2=twice as thick, etc
xlab, ylab,	x-axis and y-axis labels
xlim, ylim	x-axis and y-axis limits (min, max)
box=T / F	draw / or not a box around the plot
Axes=T / F	with. Without axes
main= " " sub=" "	add " main title " , " subtitle " to the plot

Examples:

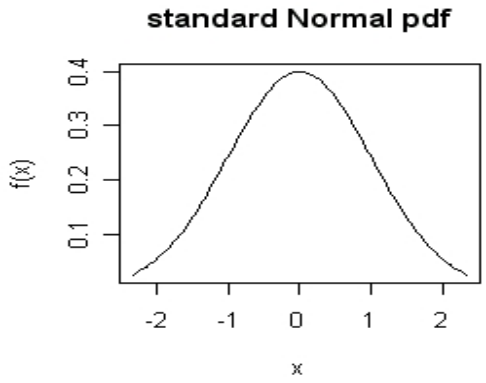
Draw the curve of f(x)-x^2

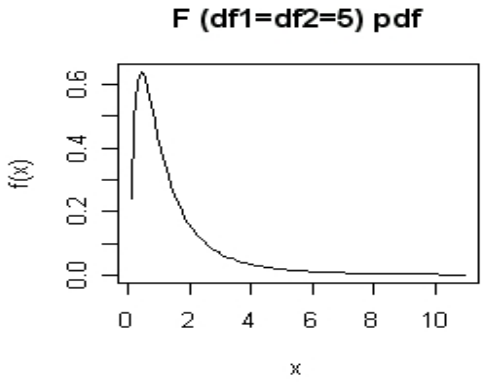
```
x=-5:5
y=x^2
par(mfrow=c(3,2))
plot(x,y)
plot(x,y,type="l")
plot(x,y,type="b")
plot(x,y,type="h")
plot(x,y,type="o")
plot(x,y,type="n")
plot(x,y,type="l",lty=2,lwd=2,xlab="x",ylab="y",ylim=c(0,28),axes=F)
```

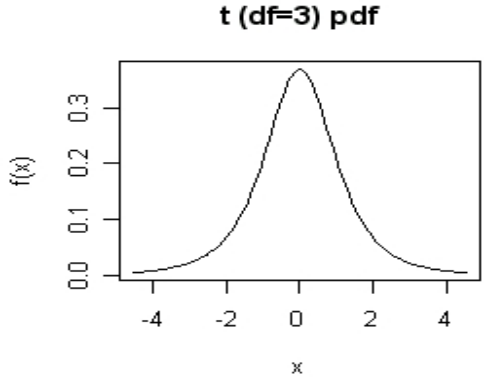
Draw the probability density function for the following distribution:

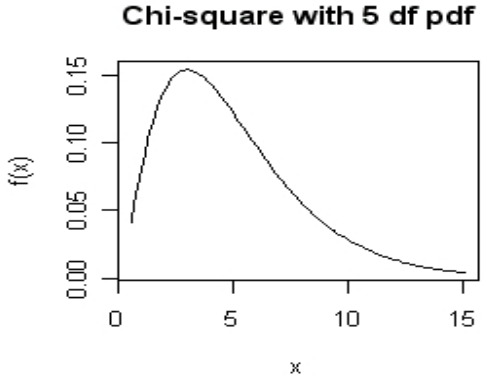
N(0,1), F(5,5), t with 3 degrees of freedom, chi-square with 5 d.f.

```
x=seq(qnorm(0.01),qnorm(0.99),length=100)
fx=dnorm(x)
plot(x,fx,type="l", xlab="x",ylab="f(x)",main="standard Normal pdf")

x=seq(qf(0.01,5,5),qf(0.99,5,5),length=100)
fx=df(x,5,5)
plot(x,fx,type="l", xlab="x",ylab="f(x)",main="F (df1=df2=5) pdf")

x=seq(qt(0.01,3),qt(0.99,3),length=100)
fx=dt(x,3)
plot(x,fx,type="l", xlab="x",ylab="f(x)",main="t (df=3) pdf")

x=seq(qchisq(0.01,5),qchisq(0.99,5),length=100)
fx=dchisq(x,5)
plot(x,fx,type="l", xlab="x",ylab="f(x)",main="Chi-square with 5 df pdf")
```









Functions to add to existing graphs:

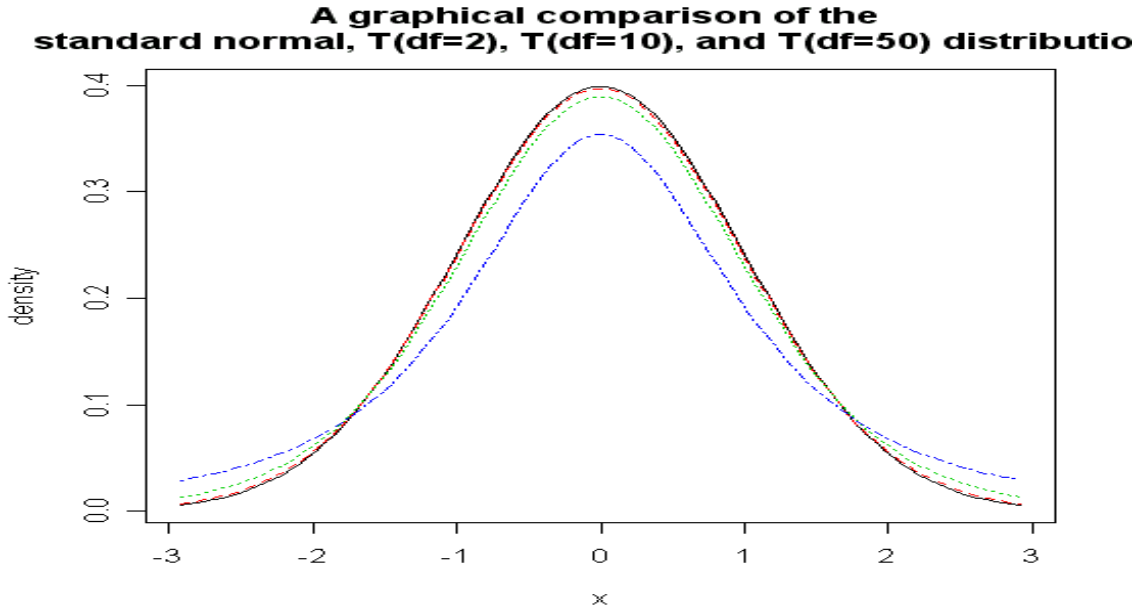
Function	Description
par (mfrow=c(2, 3))	create 2× 3 layout of figures
lines( ),	add lines to existing graph
points( )	add points to existing graph
axis(n)	add an axis to side n, n=1 for x-axis, 2 for y-axis
text( )	add text at a specified location
title( )	add title
abline (v=pos)	add vertical line at a specified position
abline (h=pos)	add horizontal line at a specified position
abline(a, b)	add line with intercept a, slope b
mtext( )	add text on the margins

The par ( ) command see help

Examples of Graphics in R

Compare standard normal distribution with t distributions with 2, 10, and 50 d.f., Study the effect of increasing degrees of freedom on the t distribution.

```
x=c(0.05,0.95)
nlim=qnorm(x)
t10lim=qt(x,10)
 t2lim=qt(x,2)
 t50lim=qt(x,50)
xlim=range(nlim,t10lim,t2lim,t50lim) # give min and max values for all dist.
points.x=seq(xlim[1],xlim[2],length=100)
fxnorm=dnorm(points.x)
fxt2=dt(points.x,2)
 fxt10=dt(points.x,10)
 fxt50=dt(points.x,50)
ylim=range(fxnorm, fxt2, fxt10,fxt50)
plot(points.x,fxnorm,col=1,lty=1,xlim=xlim,ylim=ylim,xlab="x",ylab="density")
lines(points.x,fxt2,col=2,lty=2)
lines(points.x,fxt10,col=3,lty=3)
lines(points.x,fxt50,col=4,lty=4)
title("A graphical comparison of the \n Normal, T(df=2), T(df=10), and T(df=50) distributions")
```



Explore these examples in page 26 (Notes on S_PLUS) and comment

3) plot the functions, sin(x), cos(x), and sin(x)+cos(x) for the interval [-10, 10] in one graph.

```
x=seq(-10,10,length=1000)
plot(x,sin(x)+cos(x),type="l",lty=1,xlab="x values",ylab="y values", main="Trigonometric Function")
lines(x,cos(x),lty=2)
lines(x,sin(x),lty=3)
```

4) Draw the function sin(x), where x $\in$ [0, 0.2π]

```
X=seq(0,2*pi,length=21)
Y=sin(x)
Plot(x,y,axes=F,type="b",pch="x",xlab="",ylab="y values")
Axis(1,at=c(0,1,2,pi,4,5,2*pi),labels=c(0,1,2,"pi",4,5,"2*pi"),pos=0)
Axis(2,at=c(-1,-0.5,0,0.5,1),labels=c(-1,-0.5,0,0.5,1))
Abline(h=c(-1,-0.5,0,0.5,1),lty=3)
Text(pi,0.1,"sin (pi)=0",adj=0)
Title("The sine function/n from 0 to Pi")
```

5) Start with N(0, 1), experiment with plots to see the effect of changing the location parameter once and the scale parameter another. Draw all plots on the same graph.

Number of Selected Variables	Common Plots
1	Box Plot, histogram, density, histogram density, QQ normal with line, dot, bar, pie.
2	Scatter, line, line/scatter, isolated points, high density, horizontal high density, step, curve fit (linear, robust linear, polynomial, nonlinear, exponential, natural log, power, log 10, spline, supersmooth, loess), QQ, Y series.
3	Bubble, color, text-as-symbol, scatter plot, matrix, grouped bar, stacked bar with error, contour (line or filled), 3D scatter, 3D line scatter, 3D dropline.
4 or more	XY pairs, scatter plot matrix, grouped bar, stacked bar, high-low-close, error bar.