Tutorial for Probability Distributions in R

Types of probability distributions:

1) Normal

   Function for probability density = dnorm
   Function for cumulative distribution = pnorm
   Function for quantile = qnorm

   The function **dnorm** returns the probability distribution for a mean and standard deviation of a normal distribution. The syntax is **dnorm(x, mean = 0, sd = 1, log = FALSE)**. x is the vector of qualities. The log can be specified as false or true, or you can omit this part. If true, then probabilities are returned in the form log(p). If you do not specify the mean and standard deviation, then the defaults are 0 and 1.

   Examples:

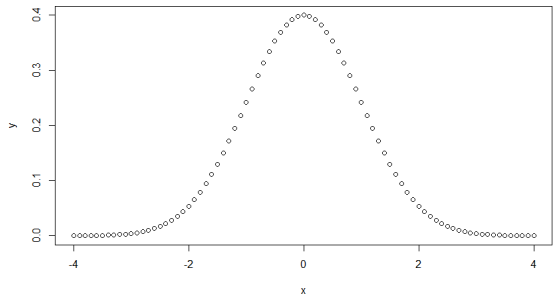
   ```
   dnorm(1, mean = 0, sd = 1)
   output: 0.2419707

   dnorm(10, mean = 7.8, sd = 0.6)
   output: 0.0008004511
   ```

   To create a plot with the dnorm() function, you need to create a sequence with the **seq** function. The syntax is **seq(from, to, by)**. The first number shows where the plot will start, the second number shows where the plot will end, and the last number shows the distance between each point in the plot. To turn it into a plot, you will make x the sequence, and you will make y the probability distribution. Then, you will use the **plot** function. The basic syntax is **plot(x,y)**.

   Example:

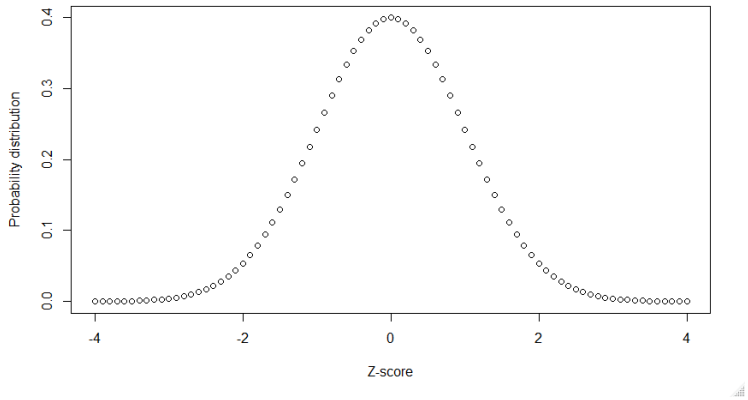
   ```
   x <- seq(-4,4,by=0.1)
   y <- dnorm(x, mean=0, sd=1)
   plot(x,y)
   ```



You can add labels by adding more information to the **plot** function. The syntax is **plot(x, y ,xlab = "Title for x-axis", ylab = "Title for y-axis")**.

Example:

```
x <- seq(-4,4,by=0.1)
y <- dnorm(x, mean=0, sd=1)
plot(x, y, xlab="Z-score", ylab="Probability distribution")
```



The function **pnorm** returns the cumulative distribution function. The syntax is the same as dnorm, but the output will be different since it is cumulative.

Examples:

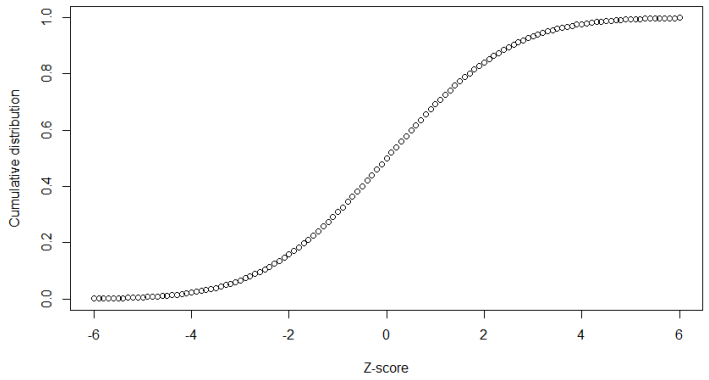
```
pnorm(1, mean = 0, sd = 1)
output: 0.8413447

pnorm(1, mean=1, sd=0.34)
output: 0.5
```

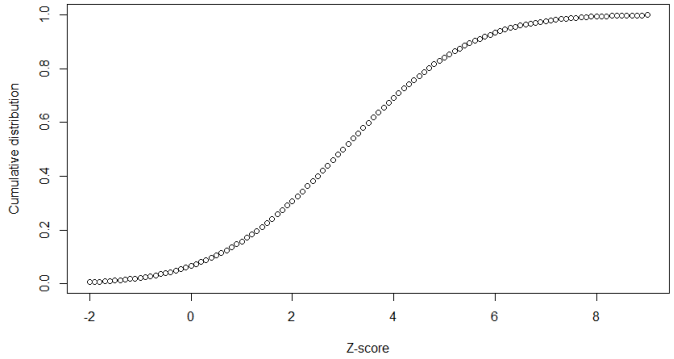
Creating a plot is like the function dnorm. However, there are a few different ways to make a plot if you want to look into other options.

Examples:

```
x <- seq(-6,6,by=0.1)
y <- pnorm(x, mean=0, sd=2)
plot(x, y, xlab="Z-score", ylab="Cumulative distribution")
```



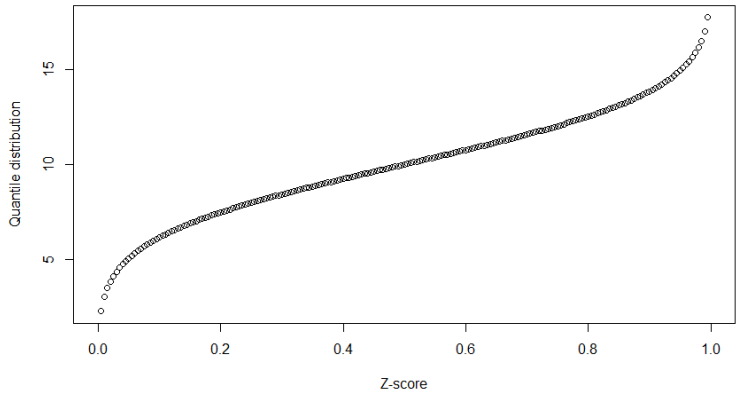

```
x <- seq(-2,9,by=0.1)
y <- pnorm(x, mean=3, sd=2)
plot(x, y, xlab="Z-score", ylab="Cumulative distribution")
```



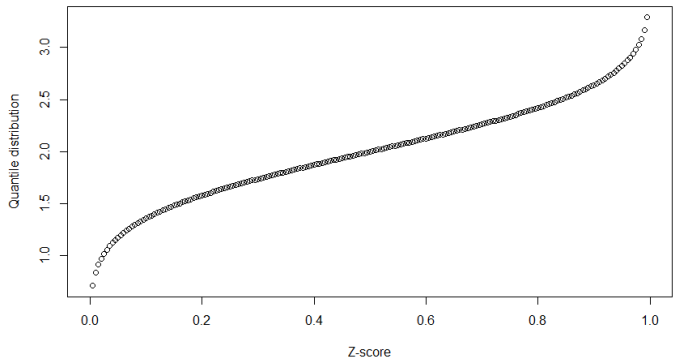
The function **qnorm** returns the quantile function, which is the inverse of the cumulative distribution function (pnorm). The syntax is the same as dnorm and pnorm, but the results are different. Also, you will need different starting and ending values. I recommend 0 and 1 by 0.005, which you will see below.

Examples:

```
x <- seq(0,1,by=0.005)
y <- qnorm(x, mean=10, sd=3)
plot(x, y, xlab="Z-score", ylab="Quantile distribution")
```




```
x <- seq(0,1,by=0.005)
y <- qnorm(x, mean=2, sd=0.5)
plot(x, y, xlab="Z-score", ylab="Quantile distribution")
```



Other ways to plot these distributions exist, but this method is one good method. If you would like more information, here are a few websites:

https://statisticsglobe.com/normal-distribution-in-r-dnorm-pnorm-qnorm-rnorm

https://www.tutorialkart.com/r-tutorial/r-plot-x-y-labels/

https://r-lang.com/dnorm-function-in-r-with-example/

If you want to find a probability with these functions…

Substitute your x-value with the number of the value you are using.

Examples:

The range is 200 to 900. The mean is 544. The standard deviation is 103.

Find the value of the density function at 550.
dnorm(540, mean=544, sd=103)
output: 0.003870306

Find the probability of a number less than 480.
pnorm(480, mean=544, sd=103)
output: 0.2671816

Find the probability of a number that is 460 or greater.
value <- pnorm(460, mean=544, sd=103)
1-value
output: 0.7926166

Find the probability of a number between 480 and 730.
pnorm(730, mean=544, sd=103)-pnorm(480, mean=544, sd=103)
output: 0.6973455

Normal Distribution probabilities Using R - VRCBuzz

2) Poisson

Function for Poisson density = dpois
Function for Poisson cumulative distribution = ppois
Function for Poisson quantile = qpois

The function **dpois** returns the poisson density. The syntax is **dpois(successes, mean per interval, log)**. The value for log is not required. Only use it if it is true, and if you put true, then the function will return the probability in the form of log.

Examples:

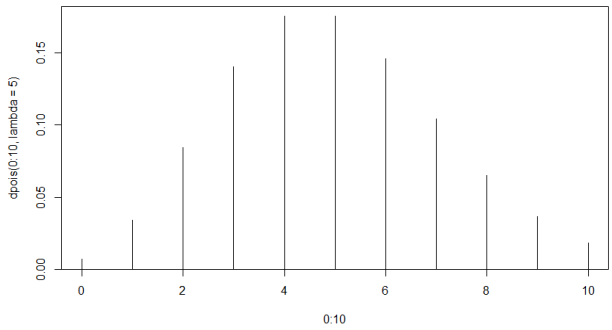
dpois(20,20)
output: 0.08883532

dpois(5,10)
output: 0.03783327

To plot the poisson density function, use **plot(success, dpois(success, lambda=5), type)**. Lambda is the mean, and the type (type='h') indicates the type of graph that we want (type='h' means we want it to be a histogram).

Example:

plot(0:10, dpois(0:10,lambda=5), type='h')



The function **ppois** returns the cumulative distribution function (cdf) of the poisson distribution. The syntax is **ppois(success, mean per interval, lower.tail, log)**. Only include lower.tail if it is true, and if you do this then the left tail is considered. If not, then the right tail is automatically considered. Only include log if it is true. If you make it true, then it will return the probability it the form of log. Otherwise, it will be false.

Examples:

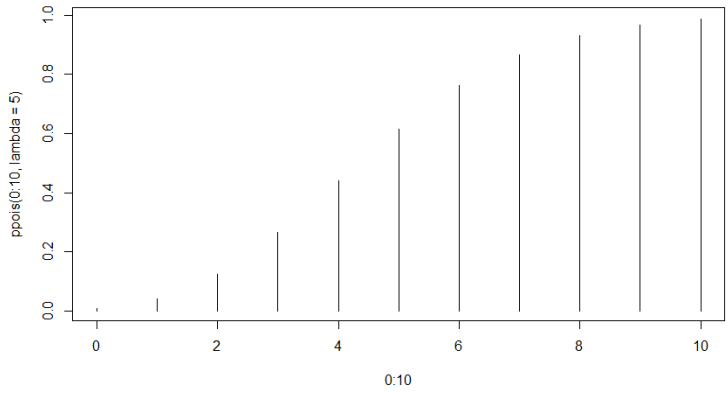
ppois(10,14)
output: 0.1756812

ppois(3,4)
output: 0.4334701

To plot the poisson cdf, the syntax is **plot(success, ppois(success, lambda), type)**.

Example:

plot(0:10, ppois(0:10, lambda=5), type='h')



The function **qpois** returns the poisson quantile function. the syntax is **qpois(percentile, lambda)**.

Examples:

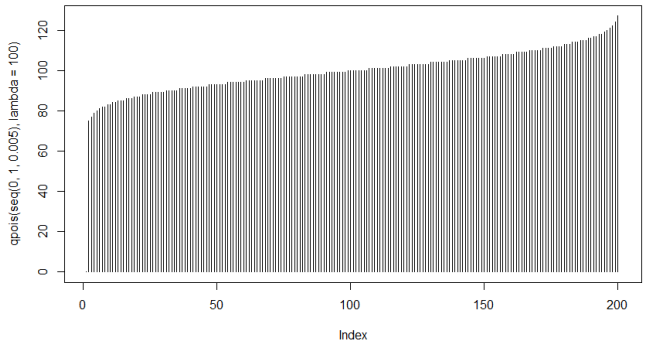
qpois(0.85, lambda=10)
output: 13

qpois(0.6, lambda=100)
output: 102

To plot the poisson inverse cdf (meaning the quantile function), the syntax is **plot(qpois(seq(0, 1, 0.005), lambda)**.

Examples:

plot(qpois(seq(0, 1, 0.005), lambda=100), type='h')



If you want to find a probability with these functions…

Substitute your value for successes with the number related to the question.

Examples:

An event occurs 3 times per month on average.

Find the probability of this event happening exactly 2 times next month.
dpois(2,3)
outcome: 0.2240418

Find the probability of this event happening at least 3 times next month.
1-ppois(2,3)
outcome: 0.5768099

Find the probability of this even happening 2, 3, or 4 times next month.
ppois(4,3)-ppois(1,3)
outcome: 0.616115

Poisson distribution probabilities using R - VRCBuzz

The normal distribution is a continuous distribution. Other continuous distributions include:

- Student's T Distribution
- Chi-square Distribution
- Exponential Distribution
- Gamma Distribution
  https://www.geeksforgeeks.org/understanding-the-t-distribution-in-r/
  https://www.geeksforgeeks.org/chi-square-distribution-in-r/
  https://mse.redwoods.edu/darnold/math15/UsingRInStatistics/ContinuousDistributions.php
  https://www.geeksforgeeks.org/gamma-distribution-in-r-programming-dgamma-pgamma-qgamma-and-rgamma-functions/

The Poisson distribution is a discrete distribution. Other discrete distributions include:

- Binomial Distribution
- Geometric Distribution
- Hypergeometric Distribution
  https://mse.redwoods.edu/darnold/math15/UsingRInStatistics/DiscreteBinom.php
  https://statisticsglobe.com/geometric-distribution-in-r-dgeom-pgeom-qgeom-rgeom
  https://statisticsglobe.com/hypergeometric-distribution-in-r-dhyper-phyper-qhyper-rhyper