**Q-Q Plot:**  
The quantile-quantile plot is a graphical method for determining whether two samples of data came from the same population or not. A q-q plot is a plot of the quantiles of the first data set against the quantiles of the second data set. By a quantile, we mean the fraction (or percent) of points below the given value.

For the reference purpose, a 45% line is also plotted, if the samples are from the same population then the points are along this line. In Statistics, Q-Q(quantile-quantile) plots play a very vital role to graphically analyze and compare two probability distributions by plotting their quantiles against each other. If the two distributions which we are comparing are exactly equal then the points on the Q-Q plot will perfectly lie on a straight line y = x.

#### Usage:

The Quantile-Quantile plot is used for the following purpose:

* Determine whether two samples are from the same population.
* Whether two samples have the same tail
* Whether two samples have the same distribution shape.
* Whether two samples have common location behaviour.

#### How to Draw Q-Q plot

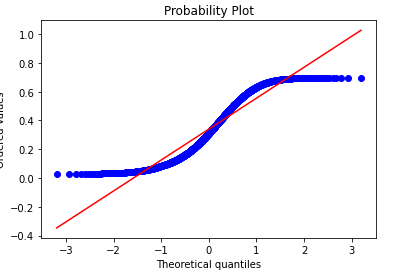
* Collect the data for plotting the quantile-quantile plot.
* Sort the data in ascending or descending order.
* Draw a normal distribution curve.
* Find the z-value (cut-off point) for each segment.
* Plot the dataset values against the normalizing cut-off points.

#### Advantages of Q-Q plot

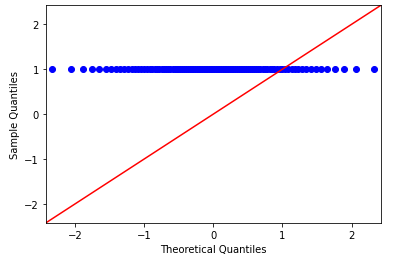
* Since Q-Q plot is like probability plot. So, while comparing two datasets the sample size need not to be equal.
* Since we need to normalize the dataset, so we don’t need to care about the dimensions of values.

#### Types of Q-Q plots

* For Left-tailed distribution: Below is the plot.



For the uniform distribution: Below is the q-q plot distribution for uniform distribution:



**# python code**

import scipy.stats as stats

import numpy as np

import matplotlib.pyplot as plt

%matplotlib inline

n = 2000

observation = np.random.binomial(n, 0.53, size=1000)/n

stats.probplot(observation, dist="norm", plot=plt)

plt.title("binomial Observation plot")

plt.show()

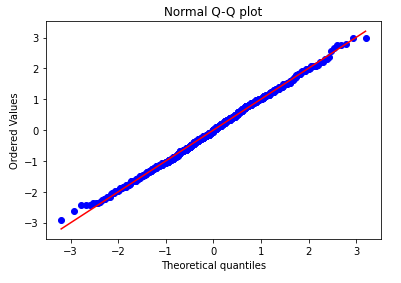
z = (observation-np.mean(observation))/np.std(observation)

stats.probplot(z, dist="norm", plot=plt)

plt.title("Normal Q-Q plot")

plt.show()

**Output**



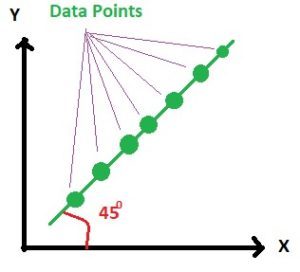
The probability density for the binomial distribution is

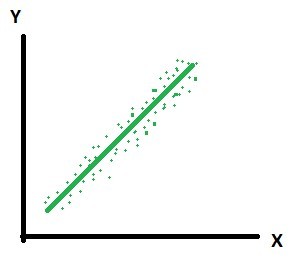
P(N) = (N C n) \* p ^ N \* (1 - p) ^ (n - N)

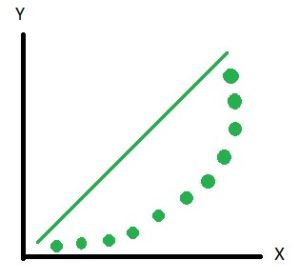
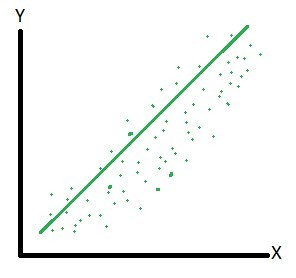
where n is the number of trials, p is the probability of success, and N is the number of successes.When estimating the standard error of a proportion in a population by using a random sample, the normal distribution works well unless the product p\*n <=5, where p = population proportion estimate, and n = number of samples, in which case the binomial distribution is used instead. For example, a sample of 15 people shows 4 who are left handed, and 11 who are right handed. Then p = 4/15 = 27%. 0.27\*15 = 4, so the binomial distribution should be used in this case.

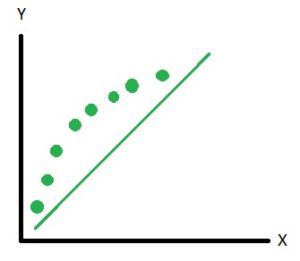
**More Details:** When the [quantiles](https://en.wikipedia.org/wiki/Quantile) of two variables are plotted against each other, then the plot obtained is known as quantile – quantile plot or qqplot. This plot provides a summary of whether the distributions of two variables are similar or not with respect to the locations.

**Interpretations**

All point of quantiles lie on or close to straight line at an angle of 45 degree from x – axis. It indicates that two samples have similar distributions.  
And in practice it is always not possible to get such a 100 percent clear straight line but the plot looks like below.

Here the points are lying nearly on the straight line.  


**The y – quantiles are lower than the x – quantiles(Y is a Left Tailed Distribution)**. It indicates y values have a tendency to be lower than x values.  
  
And in practice it is not always possible to get 100 percent as shown above but the plot looks as shown below. Here you can see that most of the points are lying below the line and few points are above the line. Hence we can say that the distributions are not the same.  


**The x – quantiles are lower than the y – quantiles(Right Tailed Distribution)**. It indicates x values have a tendency to be lower than the y values.  


Indicates that there is a breakpoint up to which the y **– quantiles are lower than the x – quantiles** and after that point the **y – quantiles are higher than the x – quantiles.** **(Tails on both sides of distribution of Y)**