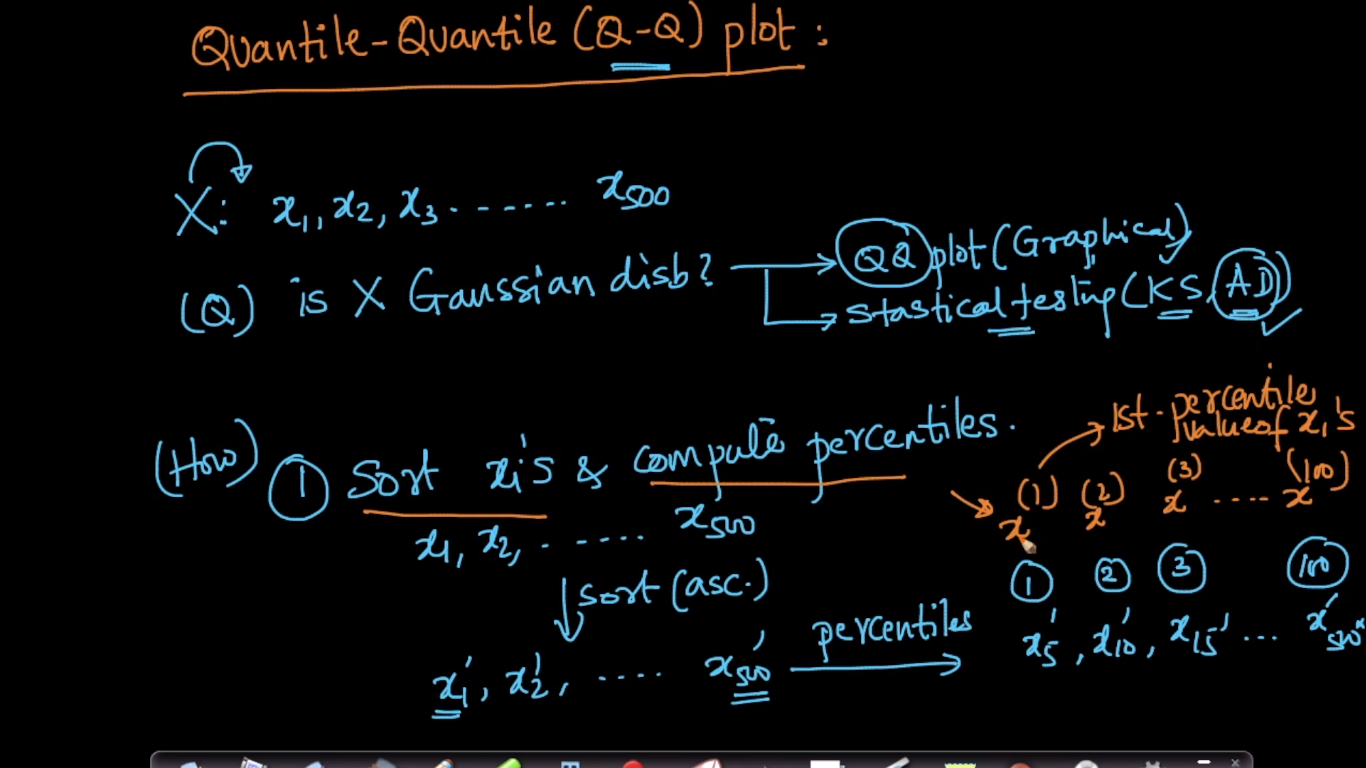
Q Q Plots (Quantile-Quantile plots) are plots of two [quantiles](https://www.statisticshowto.datasciencecentral.com/quantile-definition-find-easy-steps/)against each other. A quantile is a fraction where certain values fall below that quantile. For example, the [median](https://www.statisticshowto.datasciencecentral.com/probability-and-statistics/statistics-definitions/median-formula/)is a quantile where 50% of the data fall below that point and 50% lie above it. The purpose of Q Q plots is to find out if two sets of data come from the same distribution. A 45 degree angle is plotted on the Q Q plot; if the two data sets come from a common distribution, the points will fall on that reference line.



Given a random variable **X** which contains 500 (x1, x2, ……., x500) observations, Now using Q-Q plot we can find whether X is Gaussian distributed, To plot Q-Q plot perform following steps.

1. Sort all observations (xi’s) in ascending order.

So sorted observations are: x1’, x2’, …………, x500’.

1. Now calculate all the percentiles, therefore:

1st percentile ie x(1)  = x5’

2nd percentile ie x(2) = x10’

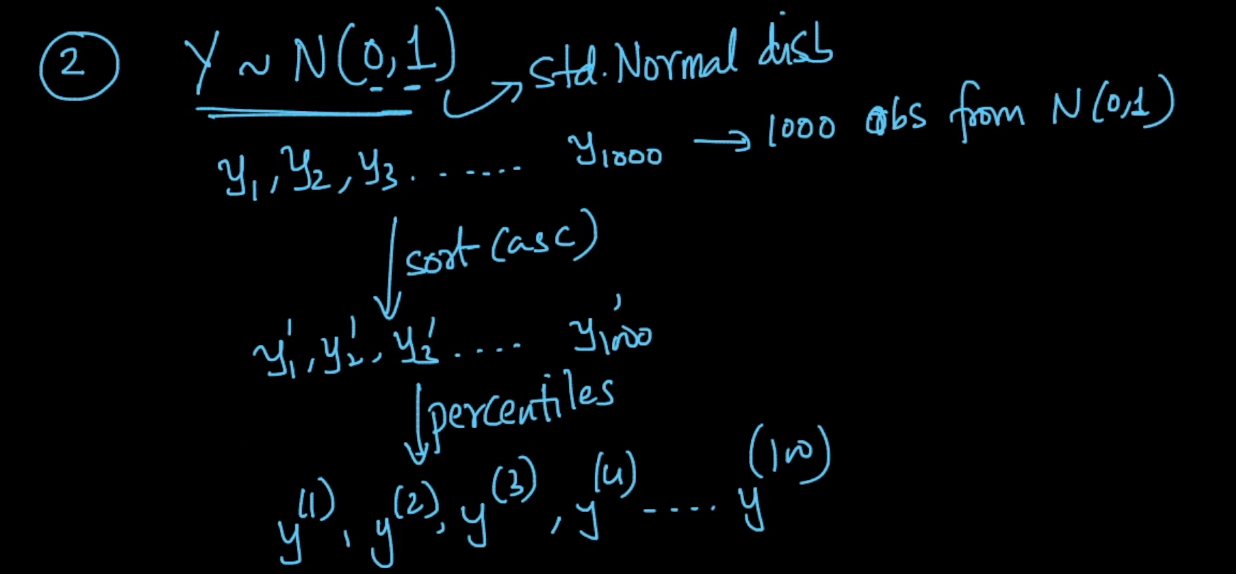
.

.

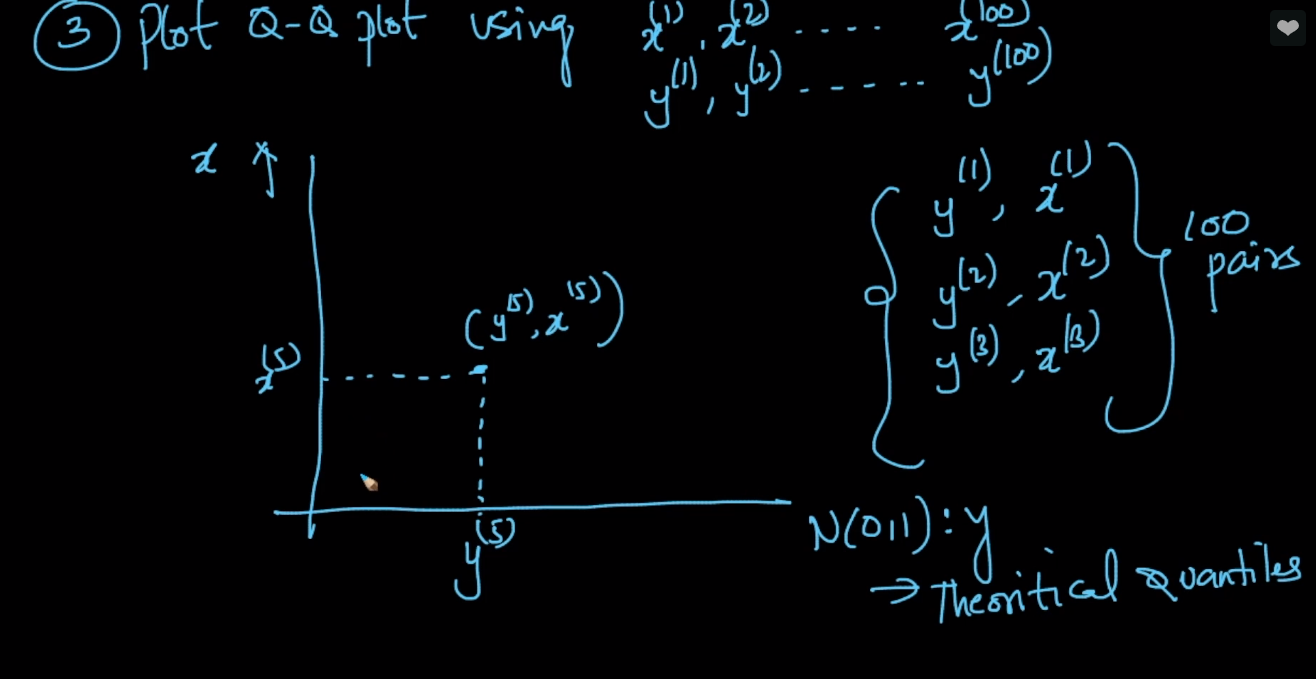
100th percentile ie x(100) =x500’

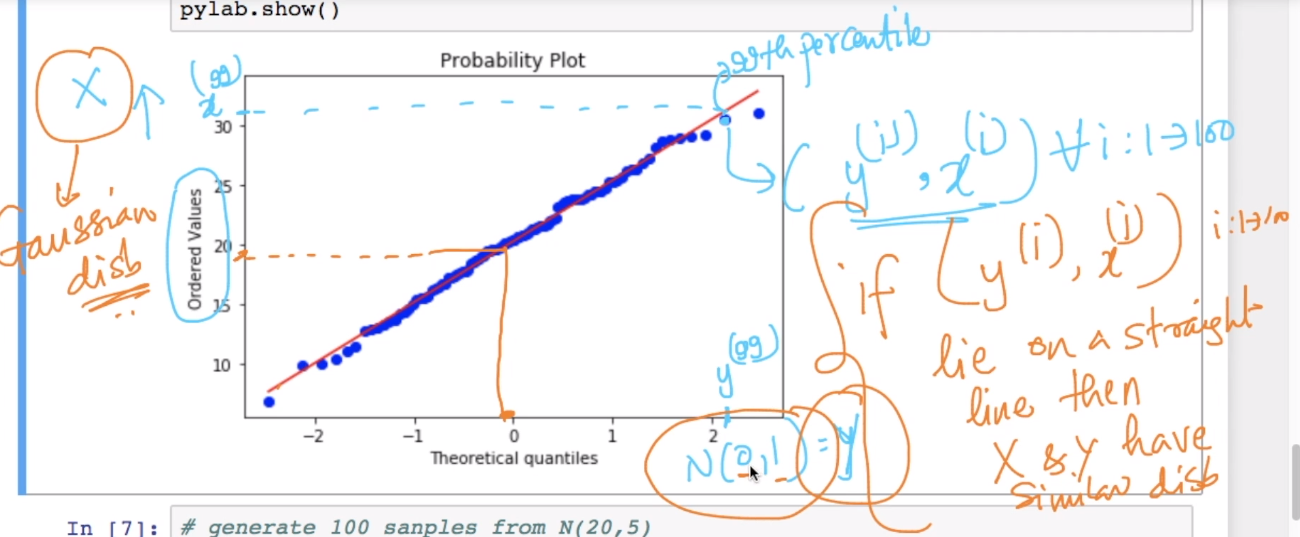
1. Since we are checking whether X is Gaussian distributed, so now we find the percentile of standard gaussian distributed random variable(whose mean is 0 and variance is 1).

In below figure we follow the same steps to calculate all the percentile of standard gaussian distributed **Y.**



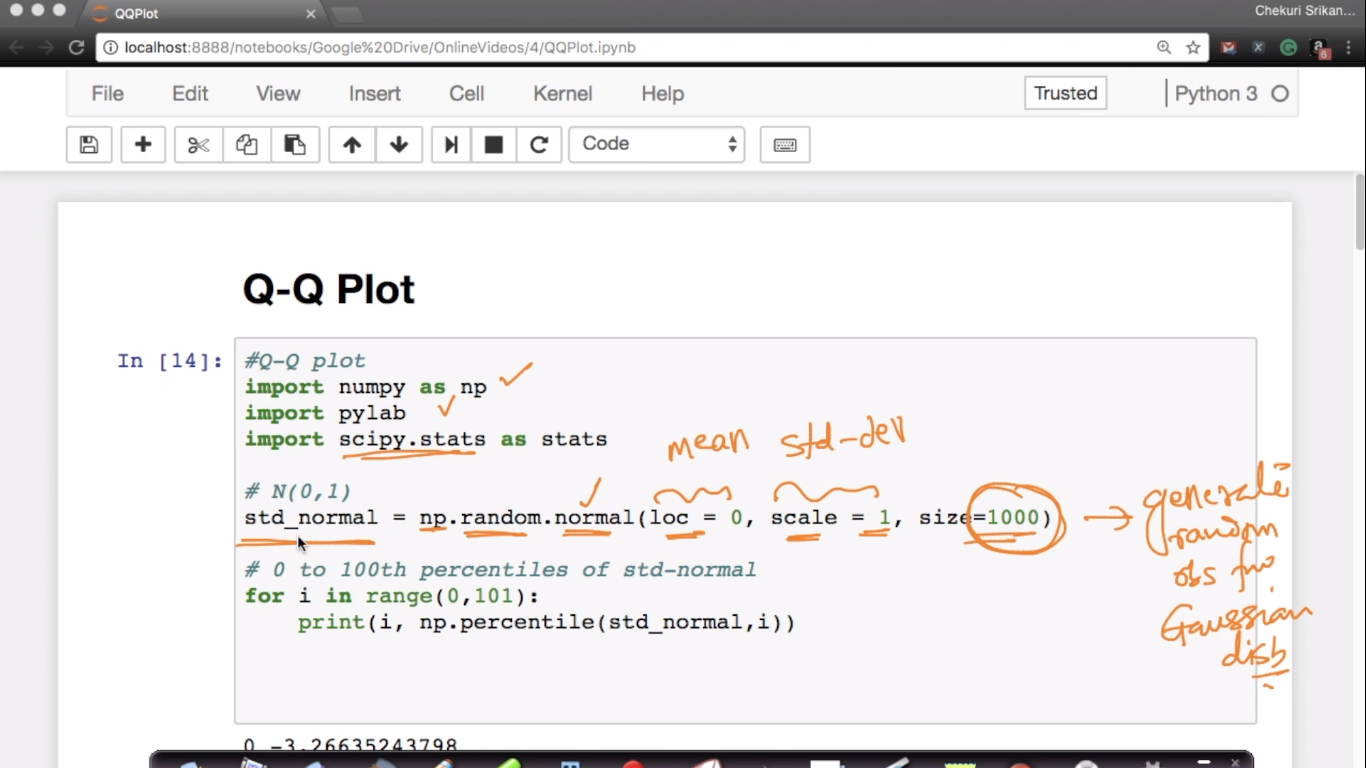
1. Now we’ll plot the given random variable’s (**X**) percentile on y-axis and standard normal distributed variable’s(**Y**) percentile on x-axis.





1. If the points lies on a straight line then **X**  is gaussian distributed, or we can say if points lie on straight line the X an Y belongs to similar distribution family.

**How to plot q-q plot**



**np.random.normal(loc, scale, size)**

normal is function present in random’s package of numpy.

Loc is the mean we want in the resultant data.

Scale is the variance we specified for the resultant data.

Size is the counts of how many observations we want.

Fortunately we don’t need to do all this thing manually there is a package called **stats** present in **scipy**, which will all these things itself, you only need to specify that random variable which you need to check for it’s family.

Here **stats.probplot(rv, dist, plot=pylab)**

prbplot is a function to print probability plot/QQ plot.

rv is the Random variable given to find it’s distribution family

dist is the parmeter which specify to which distribution family we want to check, in our example we are checking measurements against normal distribution.

plot is the plot we use to plot QQ plot

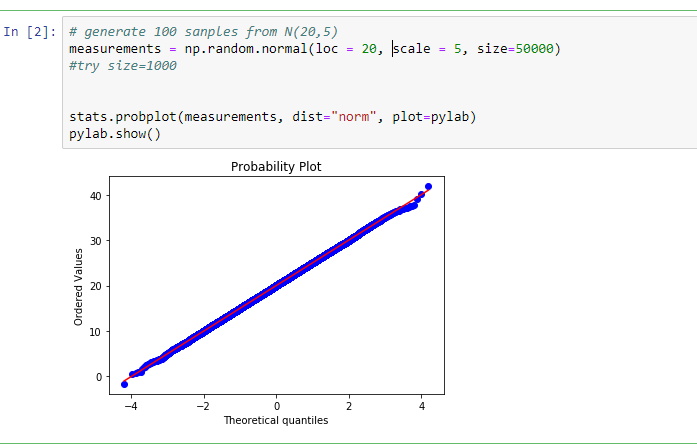
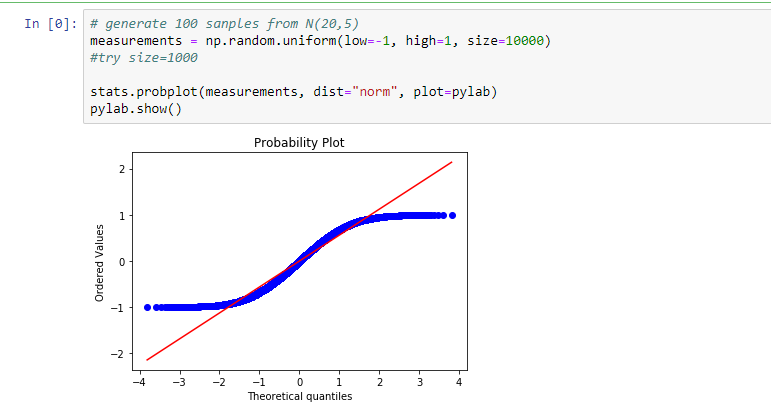


Figure on next page shows the comparison of uniform distribution with normal distribution, since both are different distribution, hence their points do not lie on straight line.



**Note:** Both the rv’s (X and Y) size’s need not to be equal.