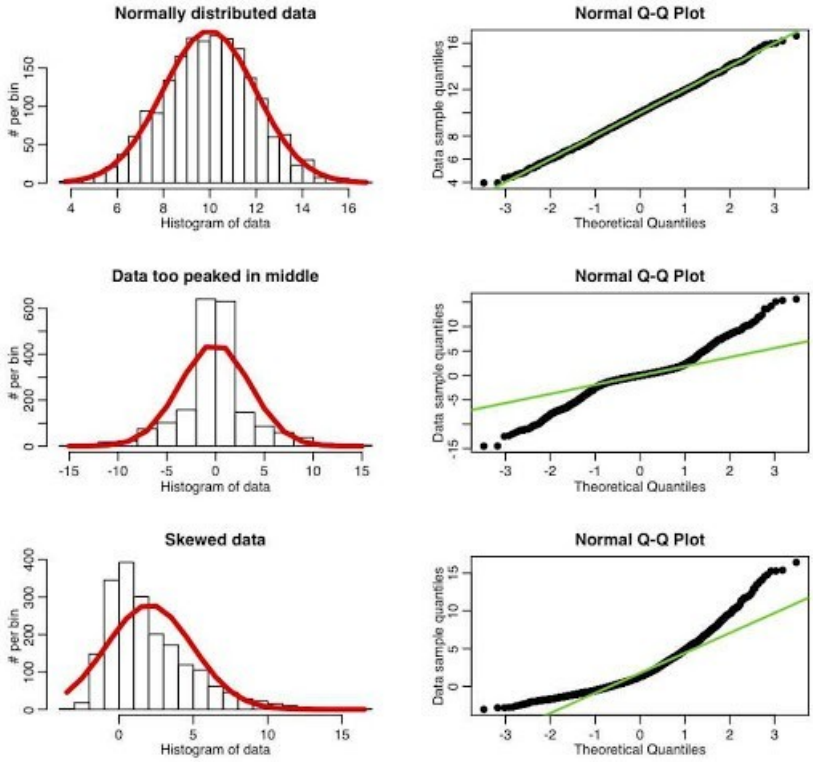
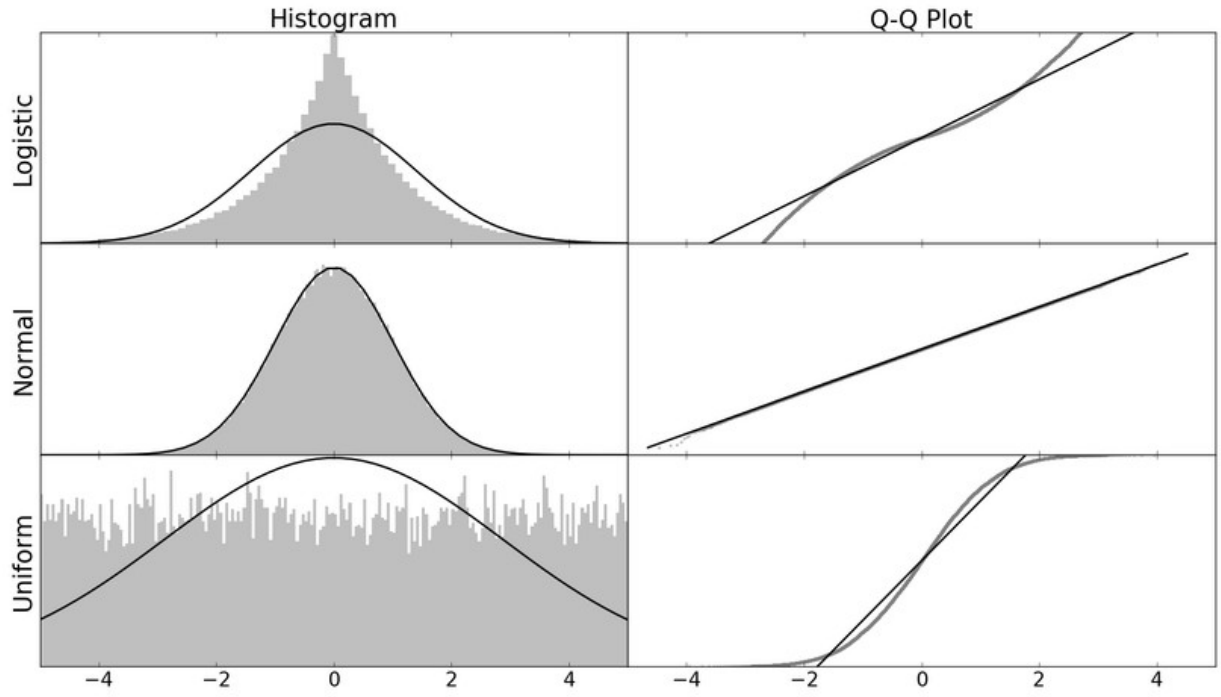
6. What is a Q-Q plot? Explain the use and importance of a Q-Q plot in linear regression?

Answer-->

Quantile-Quantile plot or Q-Q plot is a scatter point distribution between the data points and the quantile values.

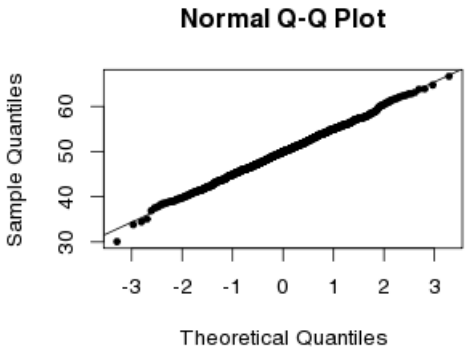


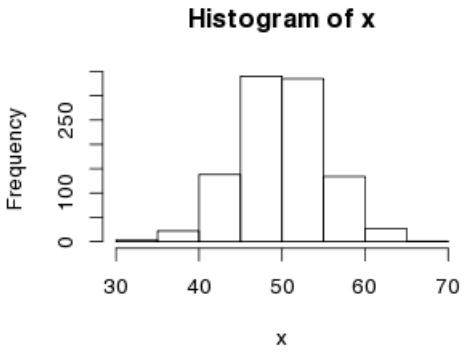
If the scatter points fits in a single line, we may assume our distribution assumption normal/uniform/skewed is correct.

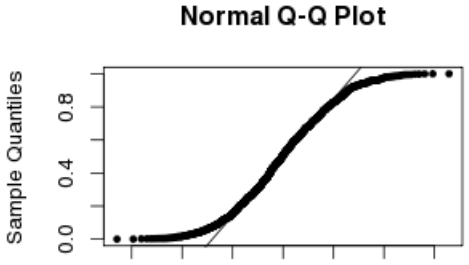


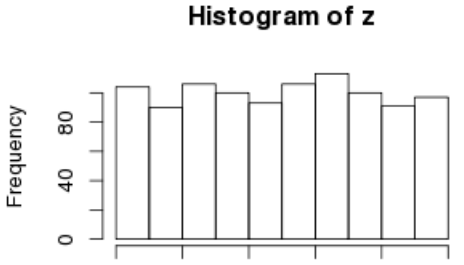
Steps:

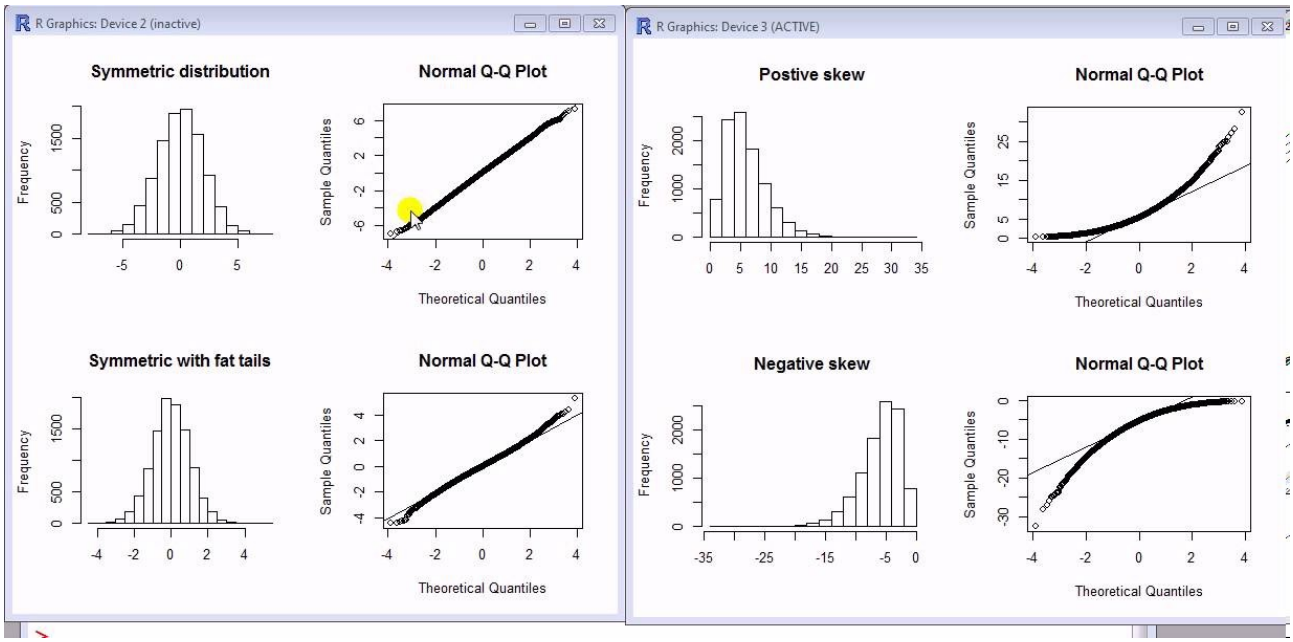
1. plot the data on the chart.
2. assign quantile values to each data point.
3. assume a uniform/normal/skewed distribution.
    a. uniform distribution – distance between each quantile value is constant.
    b. normal distribution – points at the median position are closer to each other whereas points faraway fro mean/median are faraway from each other.
    c. skewed distribution – points at one of the end are closer to each other whereas points at the other end are faraway from each other.
4. plot a scatter graph with data points and their quantile values. If all of these points lie on a line, the distribution assumed is correct, if wrong assume different distribution.


Normal Q-Q Plot
Sample Quantiles
30 40 50 60
-3 -2 -1 0 1 2 3
Theoretical Quantiles


Histogram of x
Frequency
0 100 250
30 40 50 60 70
x


Normal Q-Q Plot
Sample Quantiles
0.0 0.4 0.8


Histogram of z
Frequency
0 40 80


R Graphics: Device 2 (inactive)
Symmetric distribution
Normal Q-Q Plot
Frequency
Sample Quantiles
Theoretical Quantiles
Symmetric with fat tails
Normal Q-Q Plot
R Graphics: Device 3 (ACTIVE)
Postive skew
Normal Q-Q Plot
Negative skew
Normal Q-Q Plot