

# Normality Check

# Normality Check

## Why?

Assumption of Machine Learning Model

Statistical Tests

Outlier Detection

# Normality Check : Histogram

**Visuality Check:**

# Normality Check : Histogram

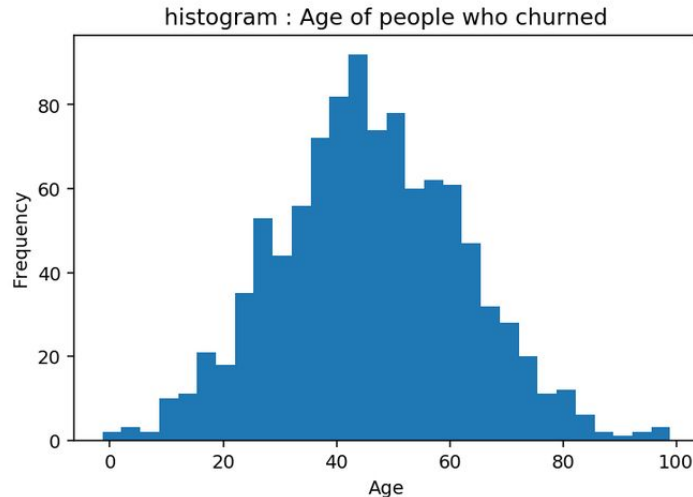
## Visuality Check:

Frequency against bins to get the Histogram.

# Normality Check : Histogram

## Visuality Check:

Frequency against bins to get the Histogram.



# Normality Check : KDE

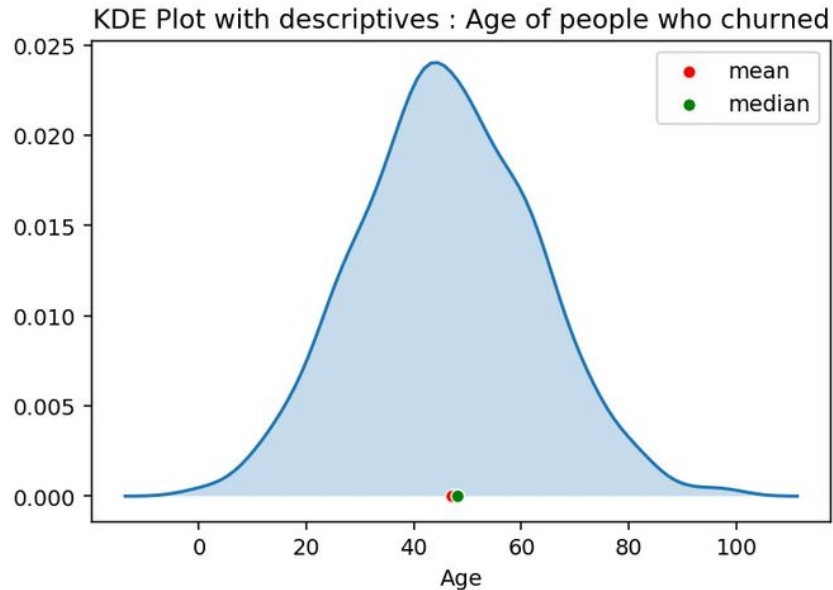
## Visuality Check:

KDE is a method to estimate pdf of data generating distribution

# Normality Check : KDE

## Visuality Check:

KDE is a method to estimate pdf of data generating distribution



# Normality Check : Q\_Q Plot



# Normality Check : Q\_Q Plot

Quantile, divide the probability distribution or observations in equal intervals.

# Normality Check : Q\_Q Plot

Quantile, divide the probability distribution or observations in equal intervals.

- 2 quantile - Median

# Normality Check : Q\_Q Plot

Quantile, divide the probability distribution or observations in equal intervals.

- 2 quantile - Median
- 4 quantile - Quartile

# Normality Check : Q\_Q Plot

Quantile, divide the probability distribution or observations in equal intervals.

- 2 quantile - Median
- 4 quantile - Quartile
- 10 quantile - Decile

# Normality Check : Q\_Q Plot

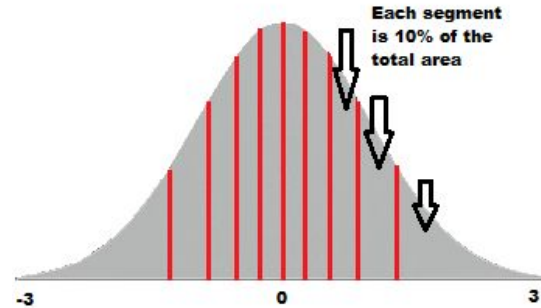
Quantile, divide the probability distribution or observations in equal intervals.

- 2 quantile - Median
- 4 quantile - Quartile
- 10 quantile - Decile
- 100 quantile - Percentile.

# Normality Check : Q\_Q Plot

Quantile, divide the probability distribution or observations in equal intervals.

- 2 quantile - Median
- 4 quantile - Quartile
- 10 quantile - Decile
- 100 quantile - Percentile.



# Normality Check : Q\_Q Plot

## Visuality check:

Q-Q plots take theoretical normal distribution quantiles and compare them against observed data quantiles.

# Normality Check : Q\_Q Plot

## Visuality check:

Q-Q plots take theoretical normal distribution quantiles and compare them against observed data quantiles.

The scatter plot will roughly form a straight line

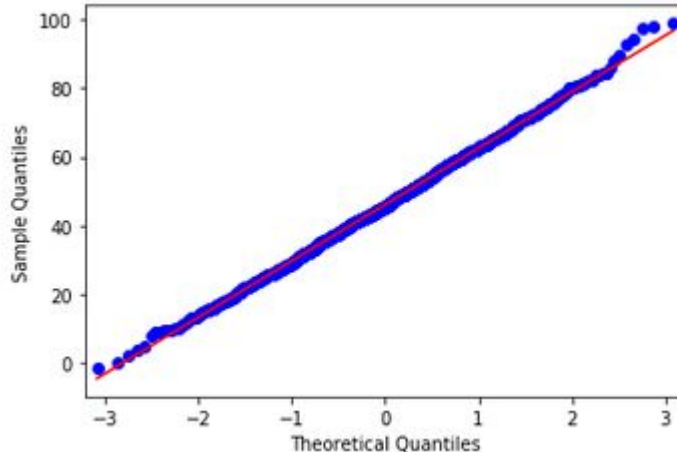


# Normality Check : Q\_Q Plot

## Visuality check:

Q-Q plots take theoretical normal distribution quantiles and compare them against observed data quantiles.

The scatter plot will roughly form a straight line.



# Normality Check : Q\_Q Plot

Example :- If the following data comes from a normal distribution ?  
106.75, 101.57, 100.78, 89.67, 120.08, 106.93, 125.61, 111.8 ,101.51

# Normality Check : Q\_Q Plot

Example :- If the following data comes from a normal distribution ?  
106.75, 101.57, 100.78, 89.67, 120.08, 106.93, 125.61, 111.8 ,101.51

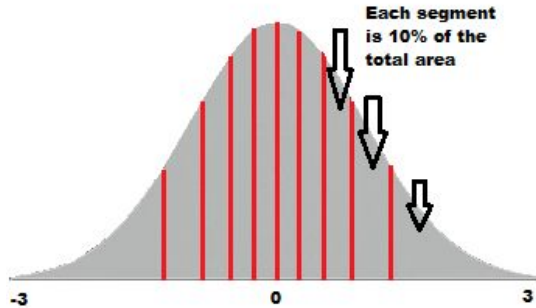
Step 1 : Sort values

89.67, 100.78, 101.51, 101.57, 106.75, 106.93, 111.8 , 120.08, 125.61

# Normality Check : Q\_Q Plot

Example :- If the following data comes from a normal distribution ?  
106.75, 101.57, 100.78, 89.67, 120.08, 106.93, 125.61, 111.8 ,101.51

Step 2: Divide a Normal curve into  $n+1$  (here 10) equally-sized areas.



# Normality Check : Q\_Q Plot

Example :- If the following data comes from a normal distribution ?  
106.75, 101.57, 100.78, 89.67, 120.08, 106.93, 125.61, 111.8 ,101.51

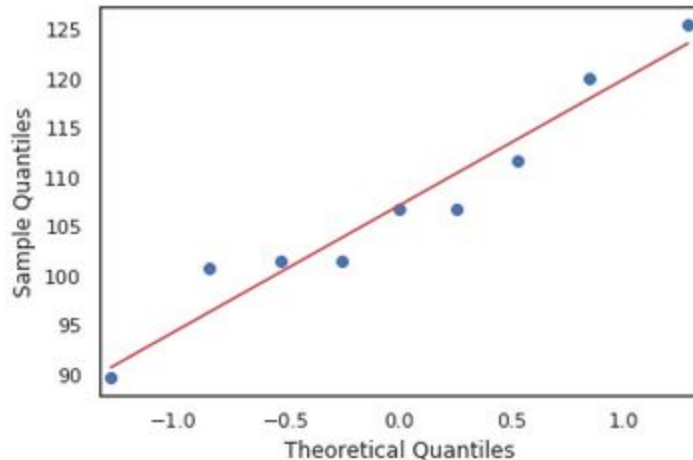
Step 3: Find the z value for each of this segment (how to use z table, covered in later section)

10% = -1.28	, 20% = -0.84,	30% = -0.52,	40% = -0.25
50% = 0,	60% = 0.25,	70% = 0.52,	80% = 0.84
90% = 1.28,	100% = 3.0		

# Normality Check : Q\_Q Plot

Example :- If the following data comes from a normal distribution ?  
106.75, 101.57, 100.78, 89.67, 120.08, 106.93, 125.61, 111.8 ,101.51

Step 4 : Plot the sorted data values (step 1) against observed z value (step 3)



Thank You!