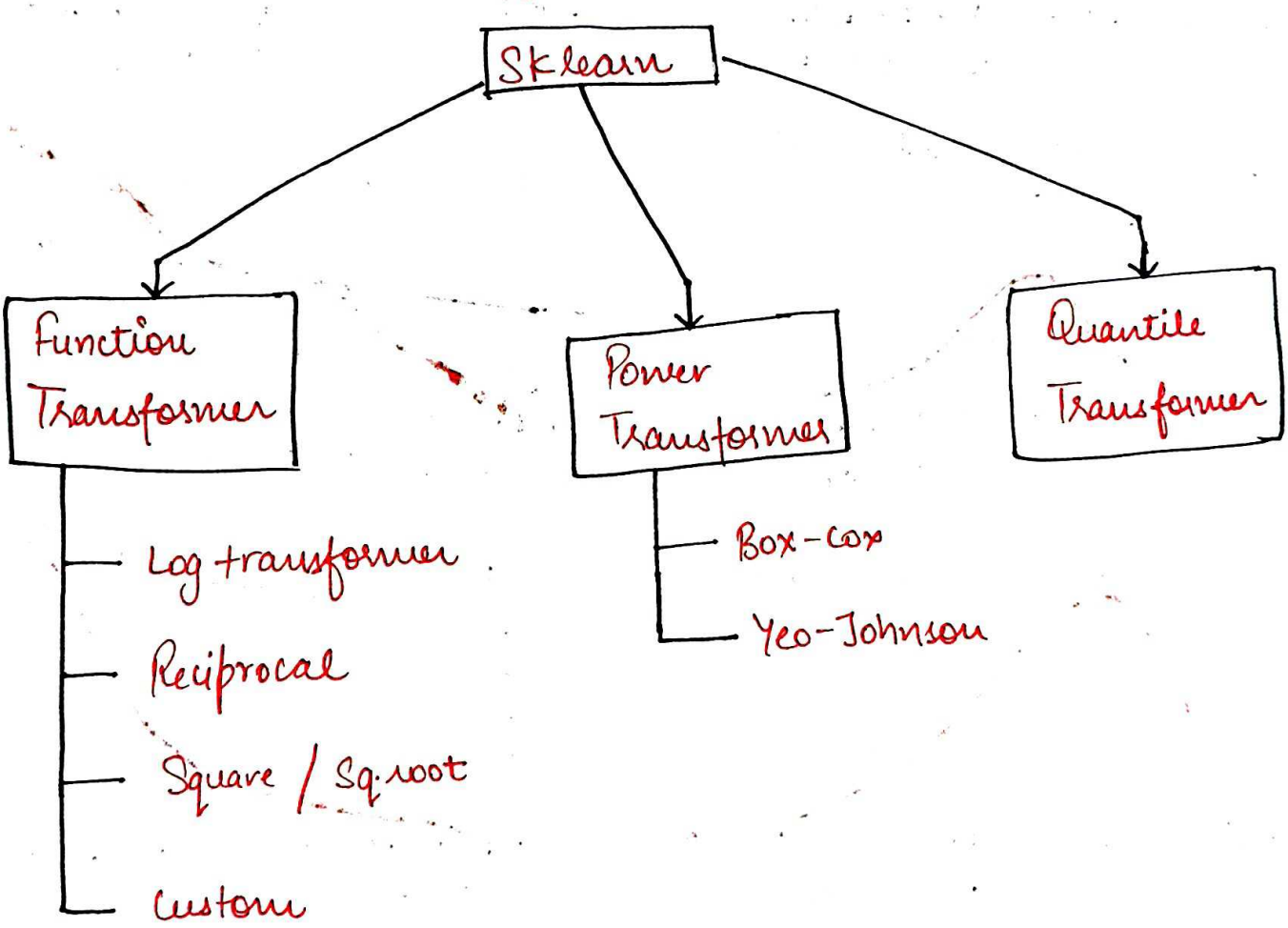
Transformation

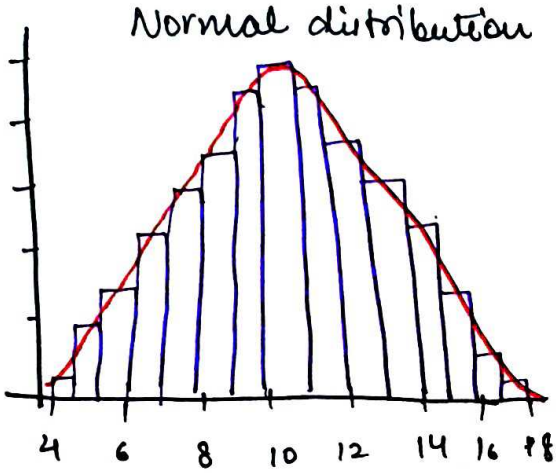
Function Transformer



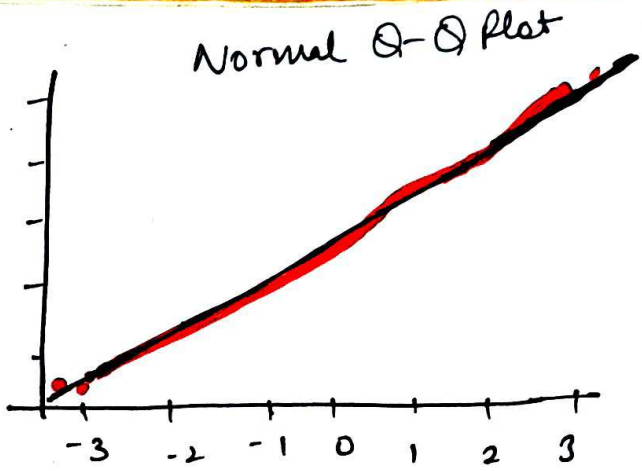
How to find if data is normal?

1. sns.distplot
2. pd.skew() if output is 0 then it is normal
3. QQ plot

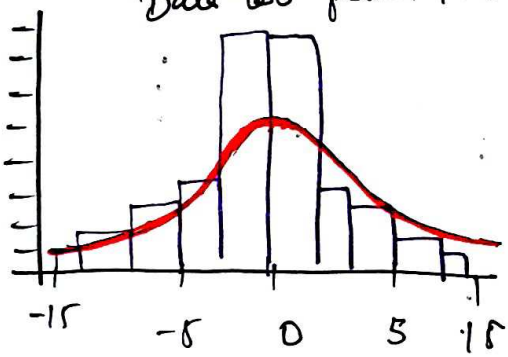
Normal distribution



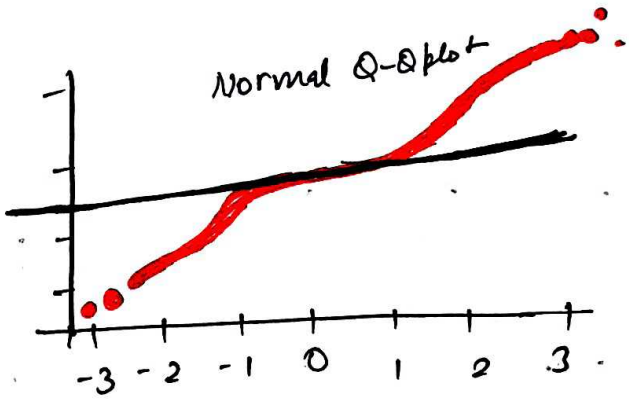
Normal Q-Q Plot



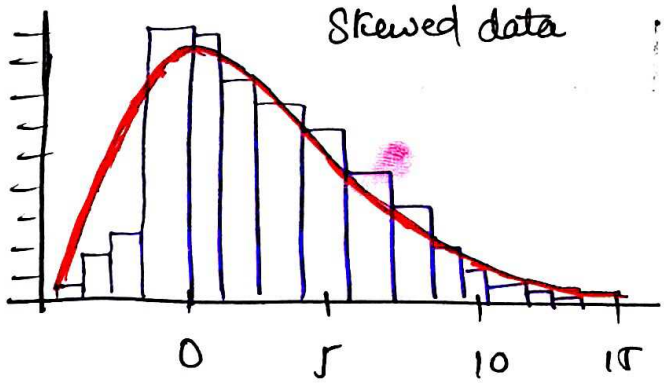
Data too peaked in middle



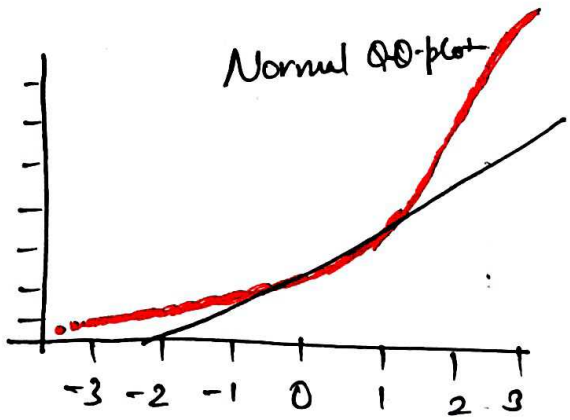
Normal Q-Q plot



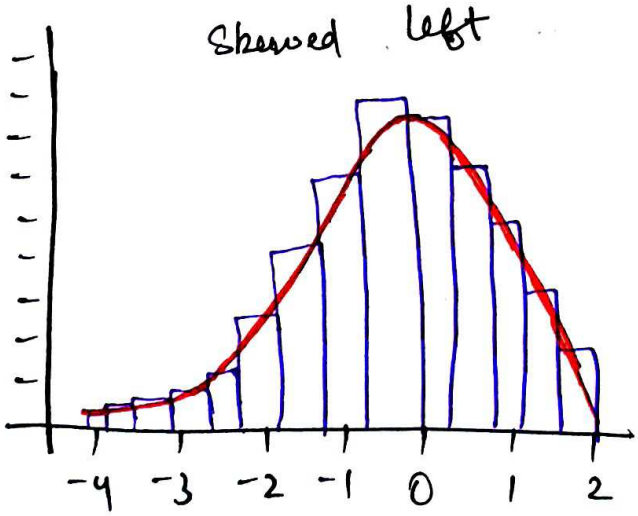
Skewed data



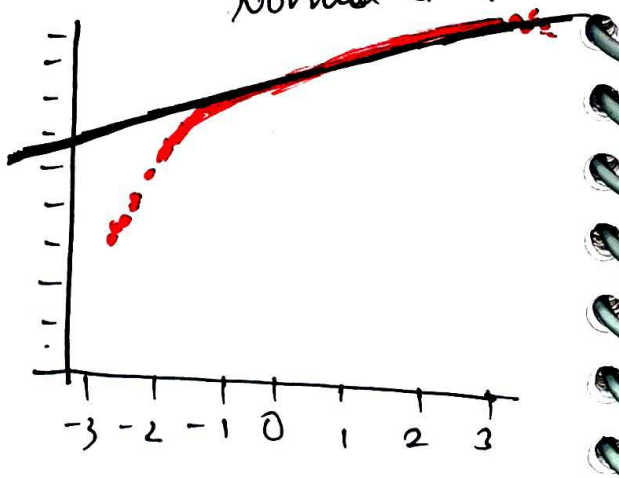
Normal QQ-plot



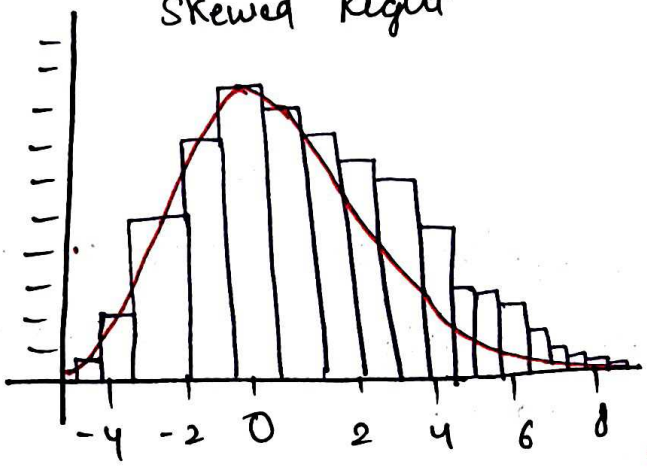
Skewed left



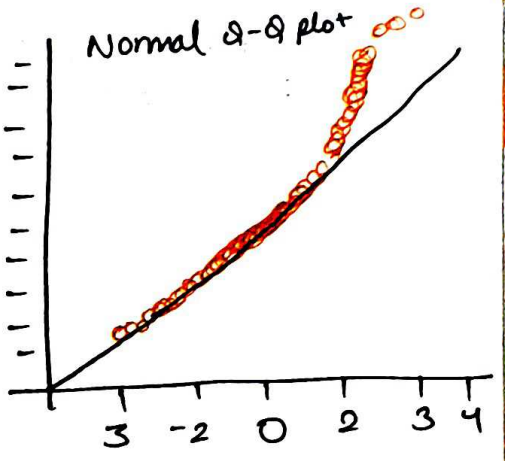
Normal Q-Q plot



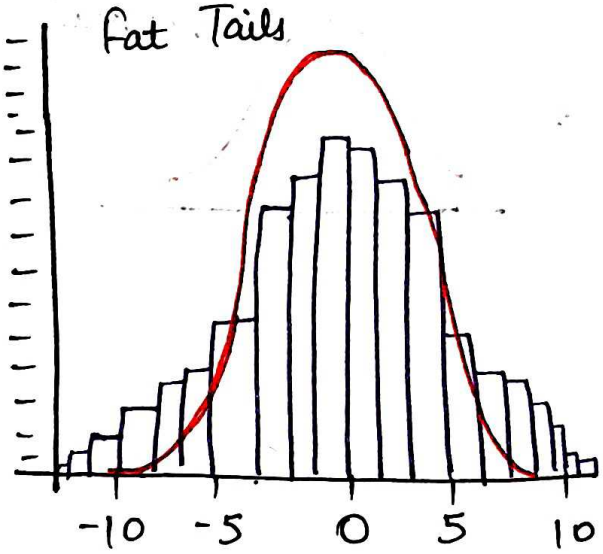
Skewed Right



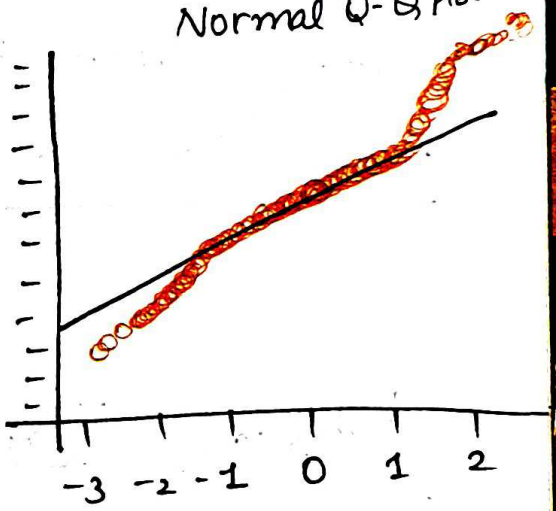
Normal Q-Q plot



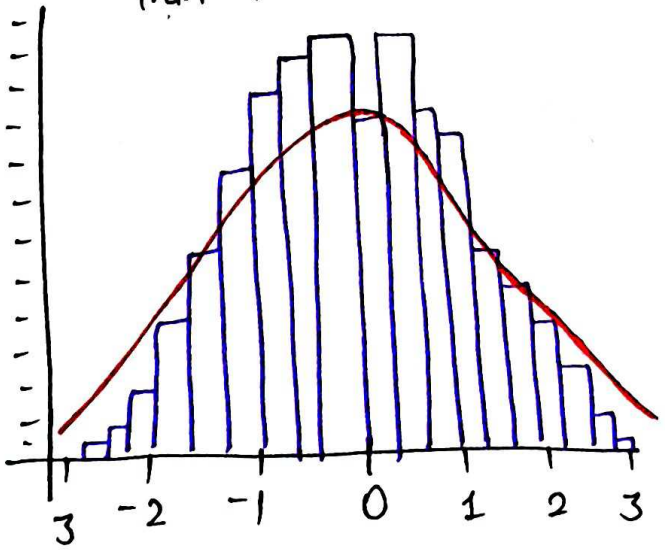
Fat Tails



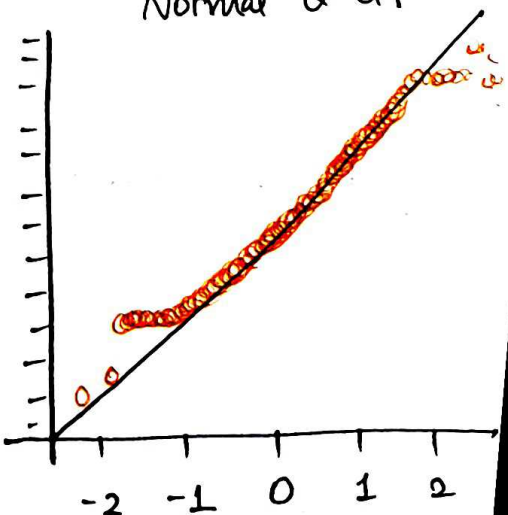
Normal Q-Q plot



Thin Tails



Normal Q-Q plot



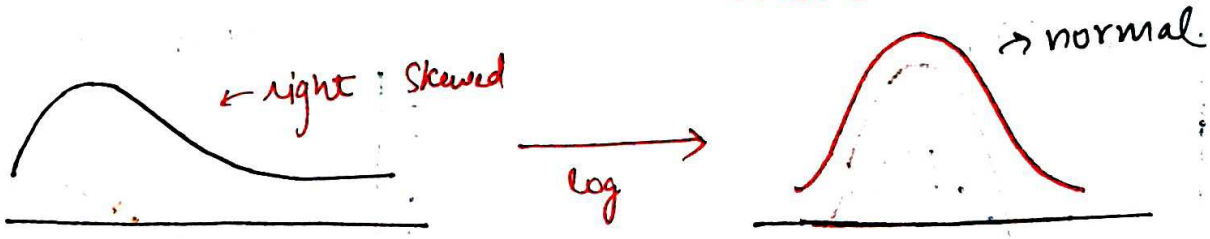
Log Transformation

* not use for -ve value.

Age
21
35
45

→ (log)

* np.log create error on 0.
* np.log1p add 1 at every value.
  ↳ in code



Resiprocal Transform

$\frac{1}{x}$ ⇒ large value convert into small value and small value convert into large value.

Square Transformation (x^2)

↳ left skewed

Square Root ($\sqrt{x}$)

Box - Cox Transform

$$x_i^{(\lambda)} = \begin{cases} \dfrac{x_i^{\lambda} - 1}{\lambda} & \text{if } \lambda \neq 0, \\ \ln(x_i) & \text{if } \lambda = 0, \end{cases}$$

The exponent here is a variable called lambda (λ) that varies over the range of -5 to 5, and in the process of searching, we examine all values of λ. Finally, we choose the optimal value (resulting in the best approximation to a normal distribution) for your variable.

* Box-Cox use where number is greater than 0 and and non-negative number.

Yeo - Johnson Transform

$$x_i^{(\lambda)} = \begin{cases} [(x_i+1)^{\lambda} - 1]/\lambda & \text{if } \lambda \neq 0,\ x_i \geq 0, \\ \ln(x_i)+1 & \text{if } \lambda = 0,\ x_i \geq 0, \\ -[(-x_i+1)^{2-\lambda} - 1]/(2-\lambda) & \text{if } \lambda \neq 2,\ x_i < 0, \\ -\ln(-x_i+1) & \text{if } \lambda = 2,\ x_i < 0 \end{cases}$$

This transformation is somewhat of an adjustment to the Box-Cox transformation, by which we can apply it to negative numbers.