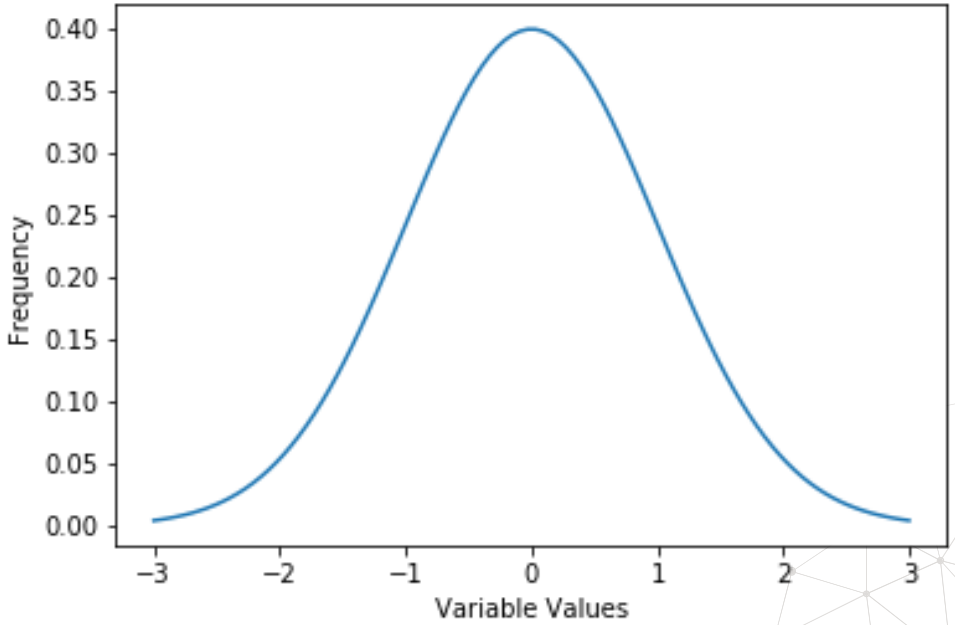
Train In Data

Variable Transformation

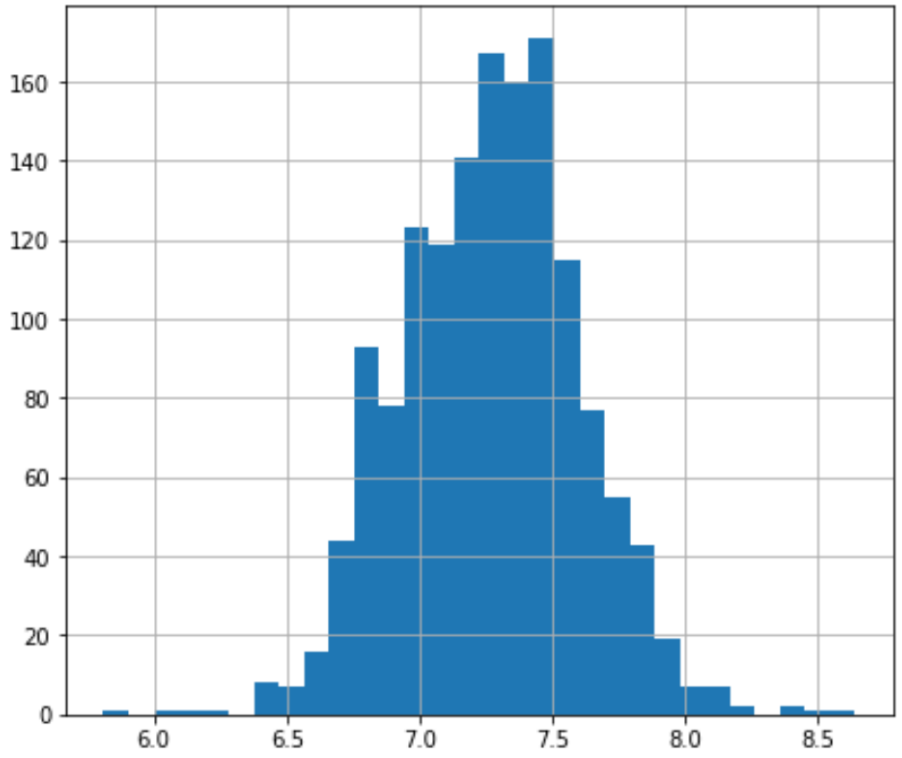
Normality in linear models

- Variables follow a Gaussian Distribution
- Normality can be assessed with histograms and Q-Q plots

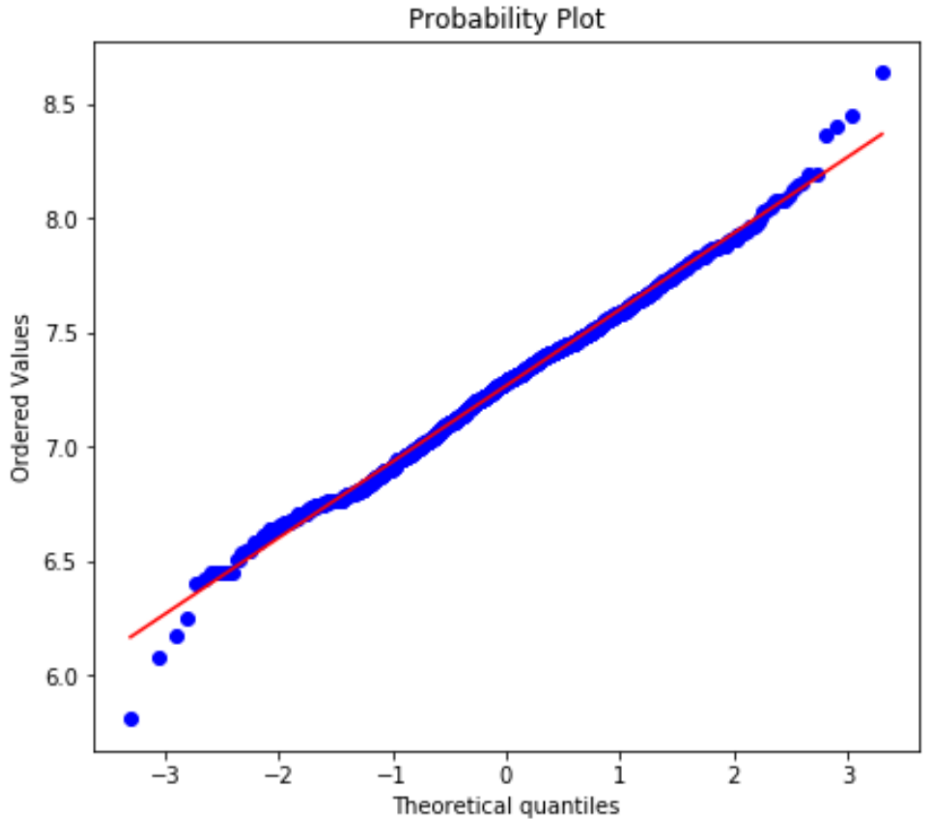


Normality assessment

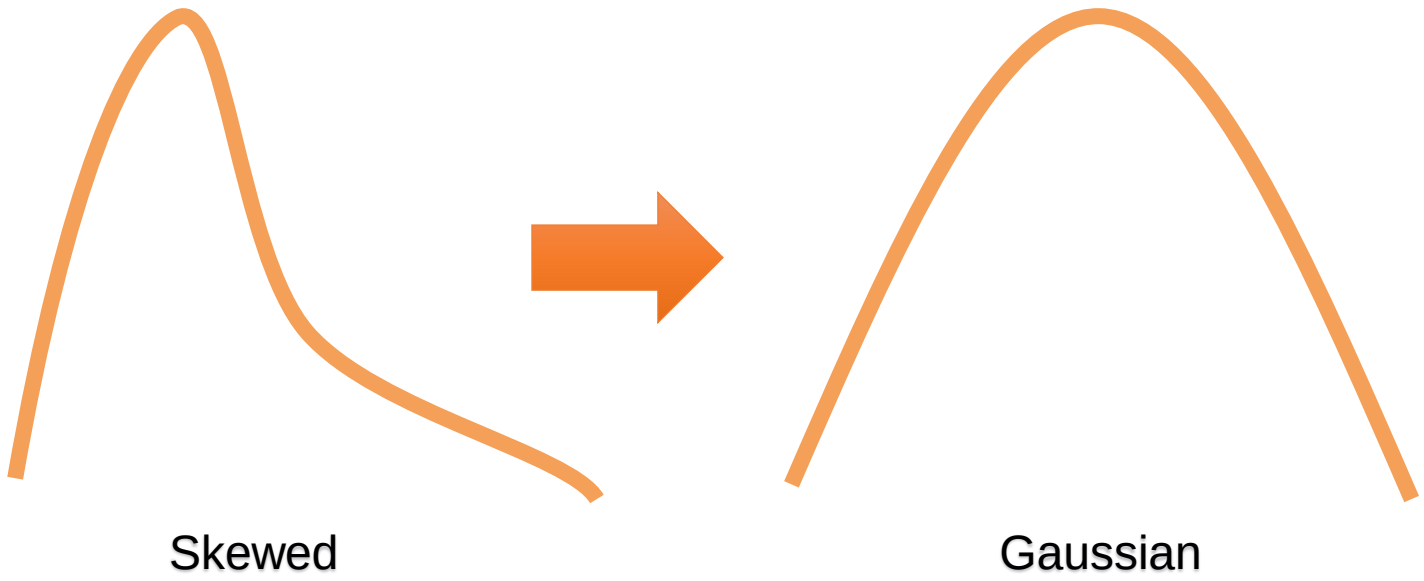
Histogram



Q-Q plot



Mathematical transformations



Variable transformation

- Logarithmic
- Exponential
- Reciprocal
- Box-Cox
- Yeo-Johnson

Mathematical transformations

Logarithmic	Reciprocal	Power / Exponential	Exponential special cases
☐ Log(X)	☐ 1 / X	☐ X exp(⋋)	☐ Box-Cox (X>0)
☐ X > 0	☐ X != 0	☐ sqr(X) / cube(X)	☐ Yeo-Johnson
		☐ Not defined for all X	

Box-Cox transformation

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

Yeo-Johnson transformation

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

Train In Data

THANK YOU

www.trainindata.com