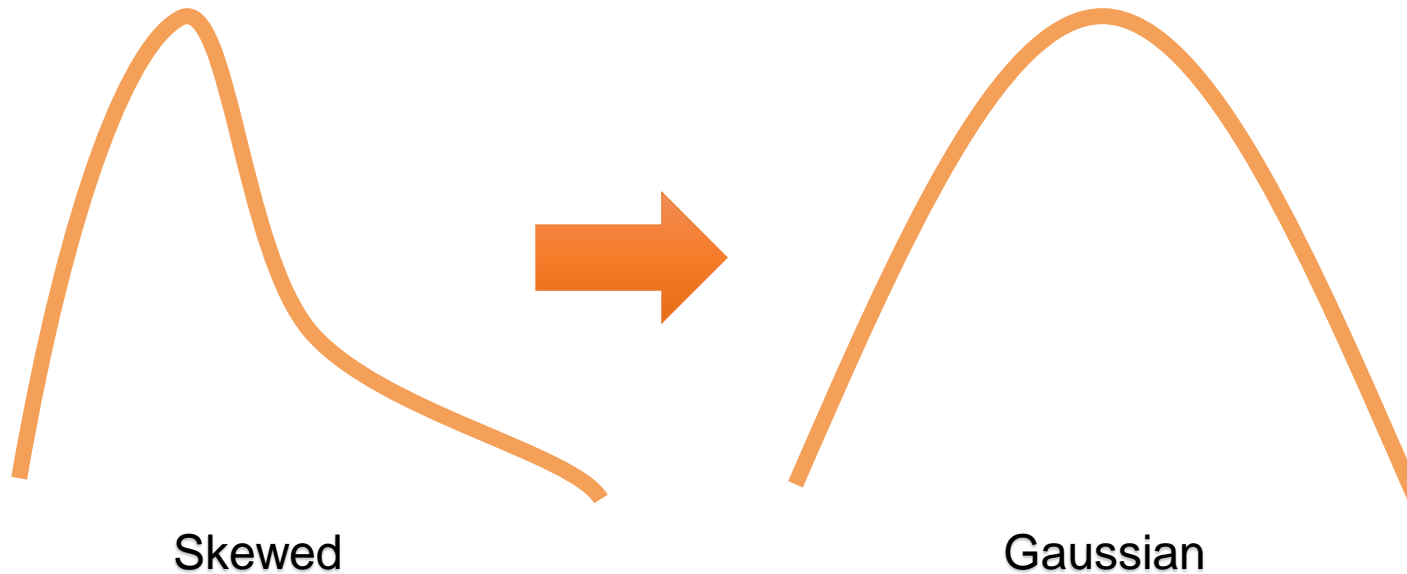




Transformation

Wrap up

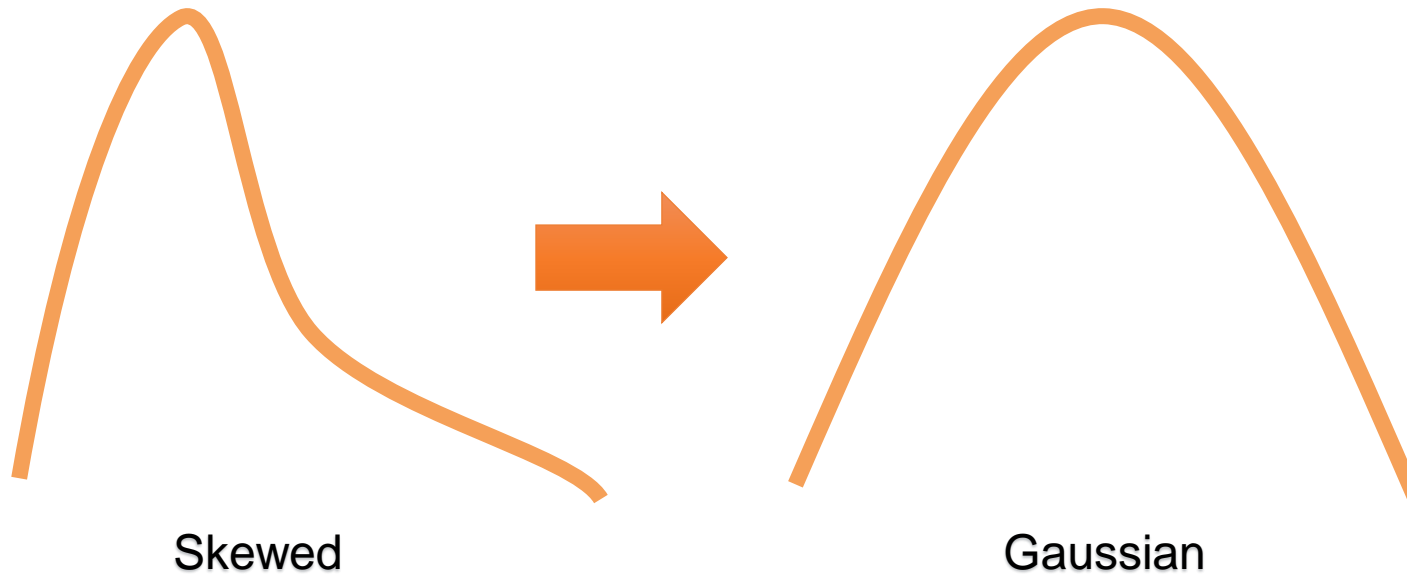
Variable transformation



Variable transformation

- Logarithmic
- Reciprocal
- Square-root
- Arcsin
- Power
- Box-Cox
- Yeo-Johnson

Variable transformation



Variable transformation

- **Logarithmic**
- Reciprocal
- Square-root
- Arcsin
- Power
- **Box-Cox**
- Yeo-Johnson

When should we transform variables?

- When we analyze data through linear statistical tests like ANOVA and when training linear regression models.
- Not necessary when training non-linear models.

Variable transformation summary

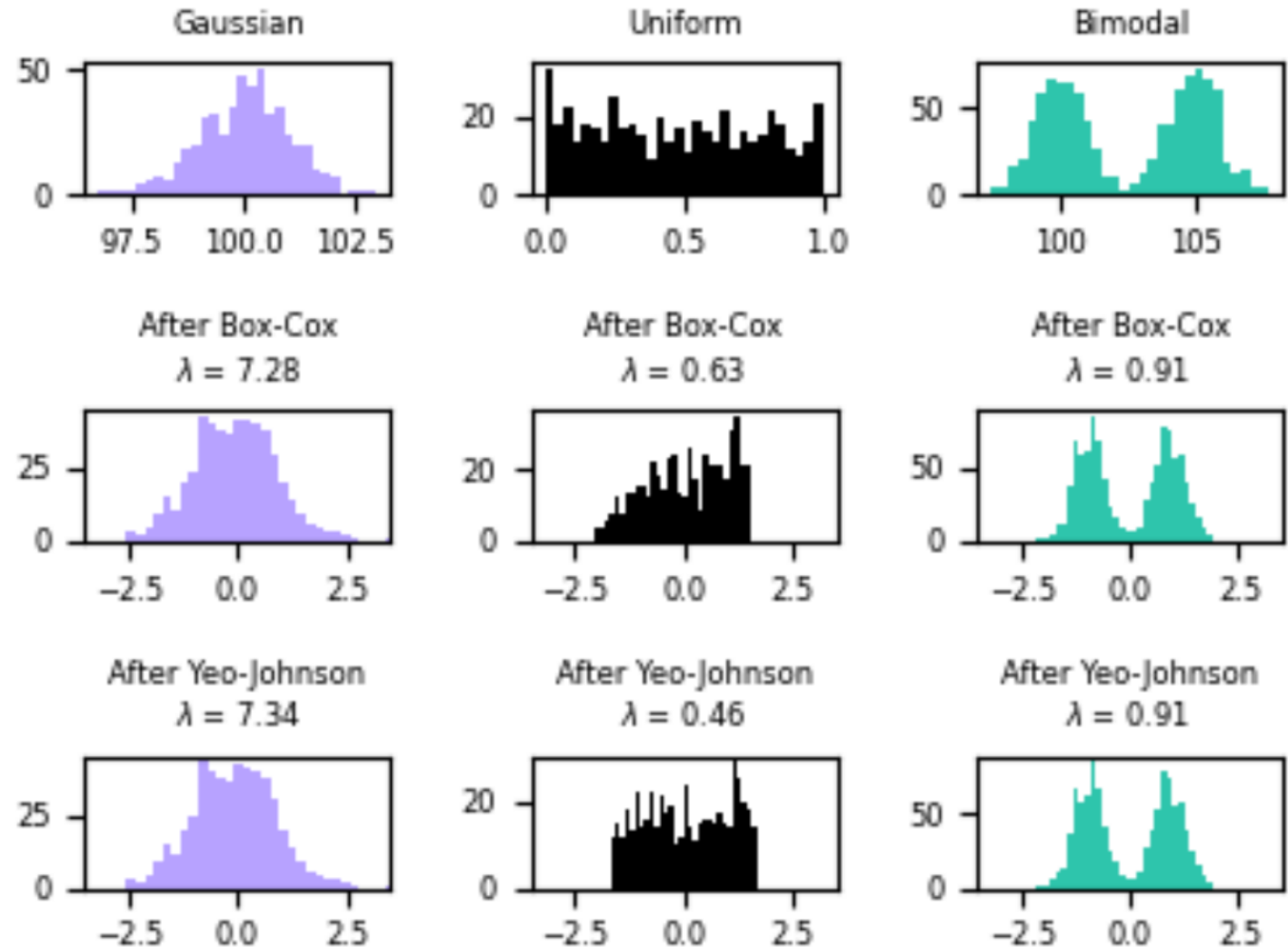
- The logarithm deals with **positive data** with a **right-skewed distribution**.
- The reciprocal transformation is useful when we have **ratios**.
- The square root is suitable for variables with **counts**.
- The arcsin helps in dealing with probabilities, percentages, and proportions.
- Box-Cox automatically finds the best transformation (includes the above except arcsin).
- Yeo-Johnson extends Box-Cox to zero and negative variables.

Analyse the data after the transformation

After transforming variables, we should take a look and make sure we obtained the expected result.

Analyse the data after the transformation

Taken from Scikit-learn documentation.



Variable transformation in Python



Feature-Engine

Content



For each lecture:

- Presentation and video
- Accompanying Jupyter notebook
 - Implementation in **numpy-scipy**
 - Implementation in **sklearn**
 - Implementation in **Feature-engine**

THANK YOU

www.trainindata.com