Train In Data

Standardisation

Standardisation

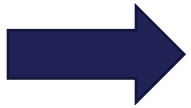
- Centres the variable at zero and sets the variance to 1.

$$\text{Z-score} = \frac{X - mean(X)}{Std(X)}$$

Standardisation: example

Price
100
90
50
40
20
100
50
60
120
40
200

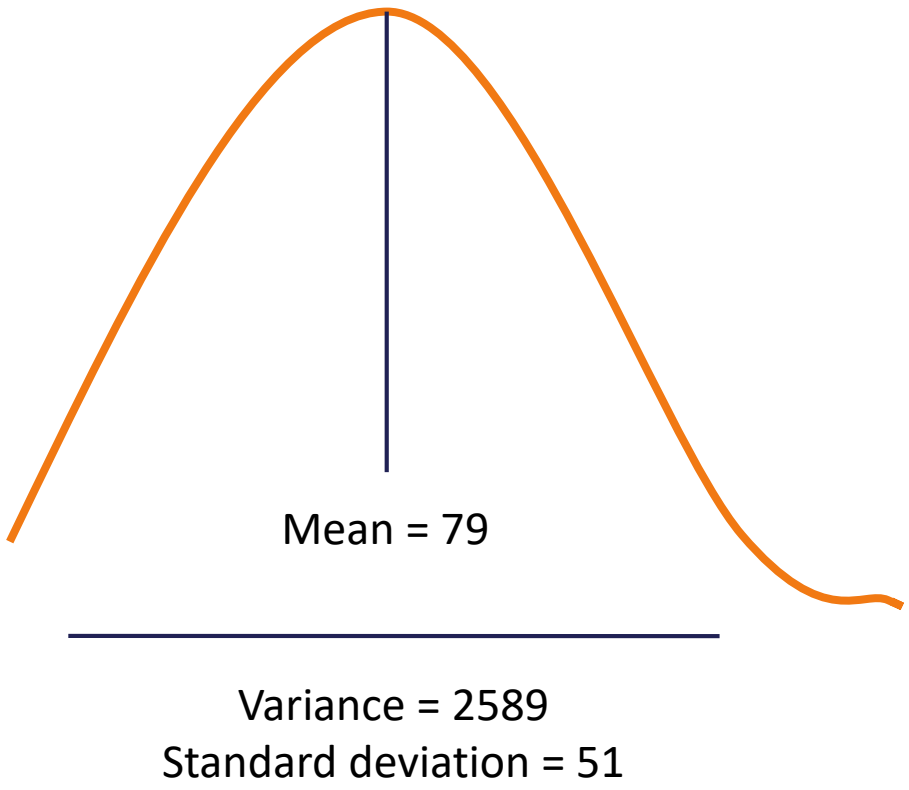
Mean = 79
Standard dev = 51



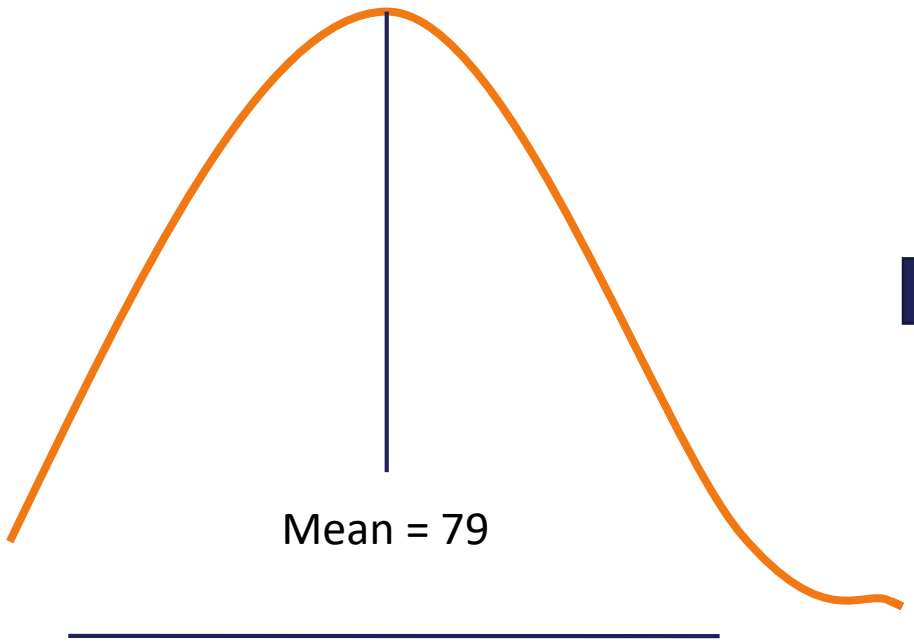
$$\frac{\text{Obs. - Mean}}{\text{Standard dev}}$$

Price
0.41
0.22
-0.57
-0.76
-1.16
0.41
-0.57
-0.37
0.80
-0.76
2.37

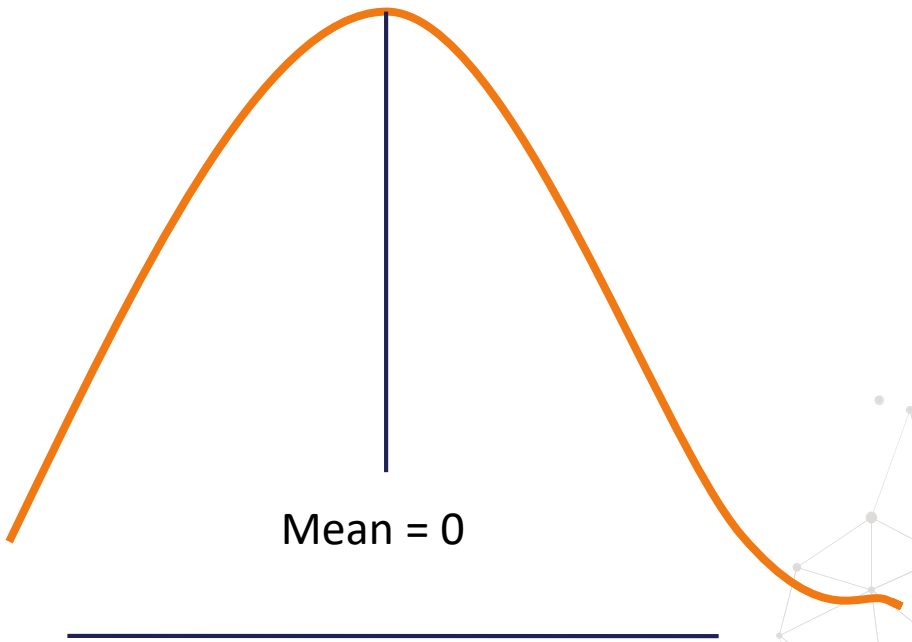
Standardisation: effect



Standardisation: effect



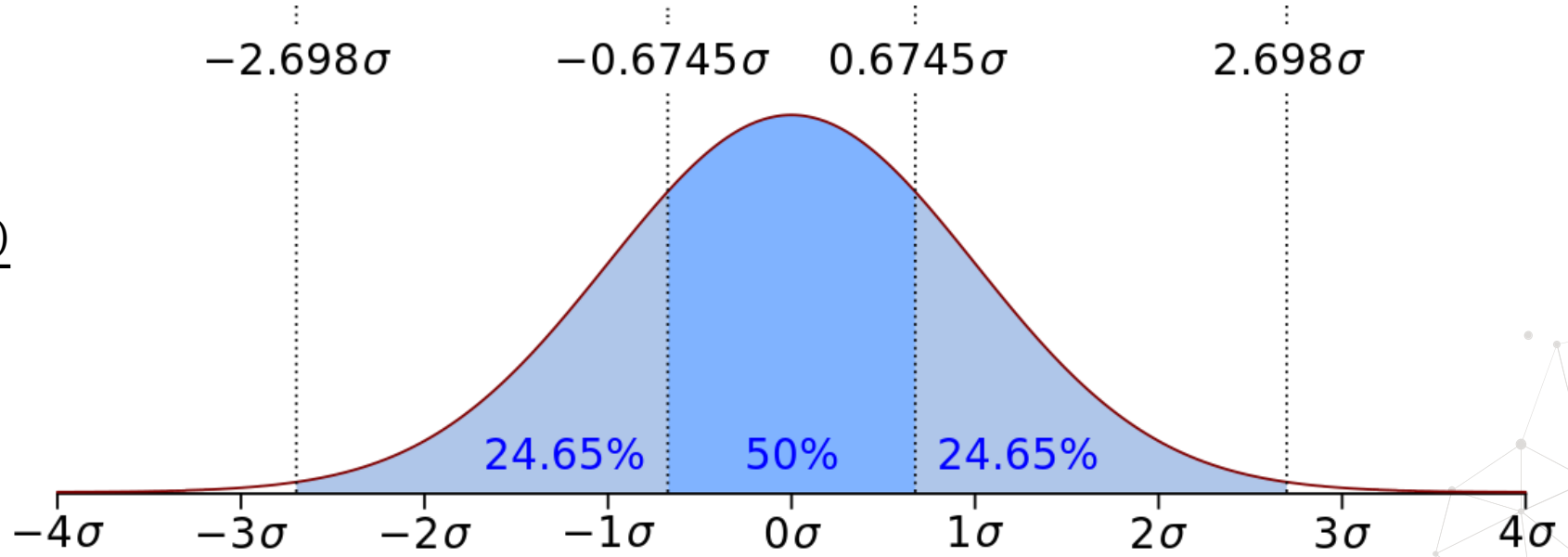
Variance = 2589
Standard deviation = 51



Variance = 1
Standard deviation = 1

Standardised variable: meaning

$$\text{Z-score} = \frac{X - mean(X)}{Std(X)}$$



Standardisation: summary

- Centres the mean at 0
- Scales the variance at 1
- Preserves the shape of the original distribution
- Minimum and maximum values vary
- Preserves outliers

Accompanying Jupyter Notebook

- Read the accompanying Jupyter Notebook
- Standardisation with Scikit-learn

Train In Data

THANK YOU

www.trainindata.com