Overview

Outliers

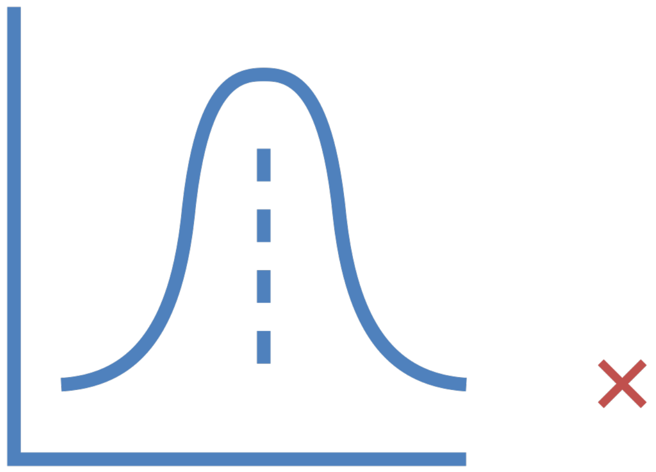
Contents

WHAT IS AN OUTLIER?

METHODS TO IDENTIFY OUTLIERS

What is an outlier?

- Data that is very different from other observations
- Suspicion that data generating process is different for these points
- Cause of the outlier determines how we handle it



Example causes

RECORDING ERRORS
(E.G., MANUAL PROCESS)

EXTERNAL EVENTS
(E.G., MARKETING)

RARE EVENT
(E.G., HEAVY TAIL DISTRIBUTION)

Approach to outliers in this course

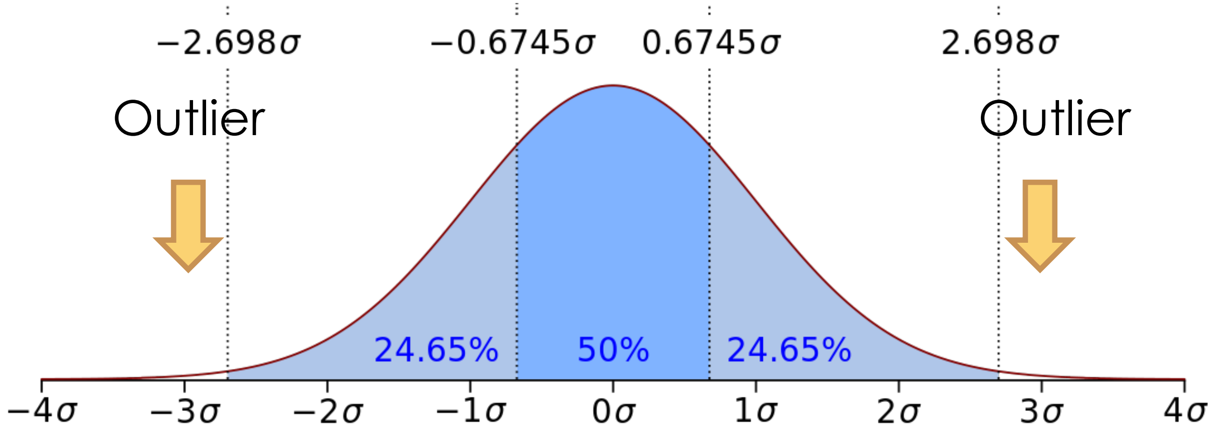
- Handle outliers in cases where they may affect model performance

- The course is tailored to improve forecasting model performance

- We will demonstrate some easy to apply outlier detection methods to achieve the above

- Outlier detection is a large field which is mostly out of scope for this course

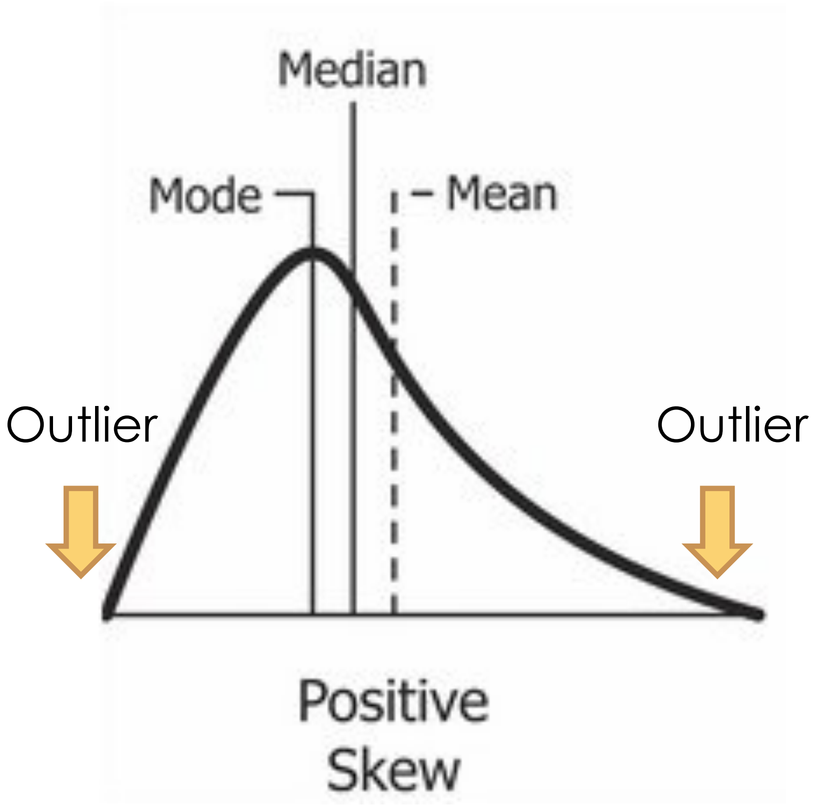
Extreme value analysis: Normal distribution



- ~99% of the observations of a normally distributed variable lie within the mean ± 3 × standard deviations.
- Values outside mean ± 3 × standard deviations are considered outliers

https://commons.wikimedia.org/wiki/File:Boxplot_vs_PDF.svg

Extreme value analysis: Skewed distributions



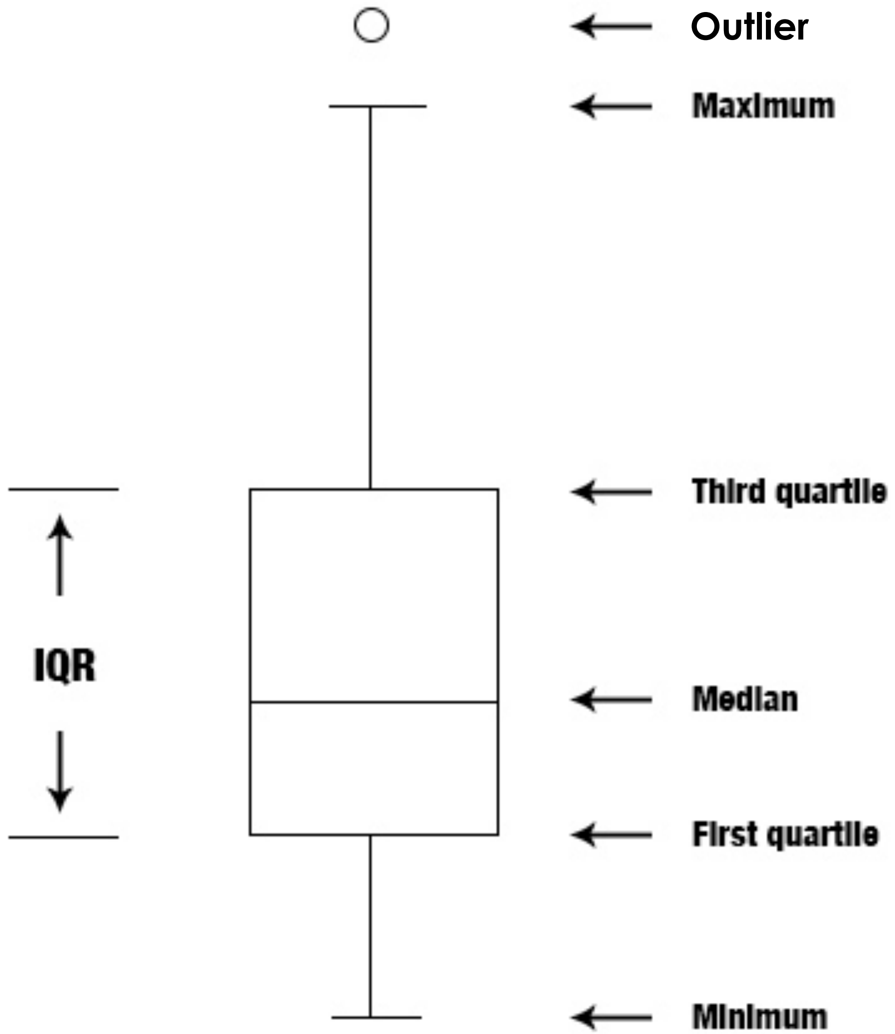
- The general approach is to calculate the quantiles, and then the inter-quantile range (IQR), as follows:
- IQR = 3rd Quartile – 1st Quartile
- Upper limit = 3rd Quartile + IQR × 1.5
- Lower limit = 1st Quartile - IQR × 1.5

Note, for extreme outliers, multiply the IQR by 3 instead of 1.5

Notes on quantiles

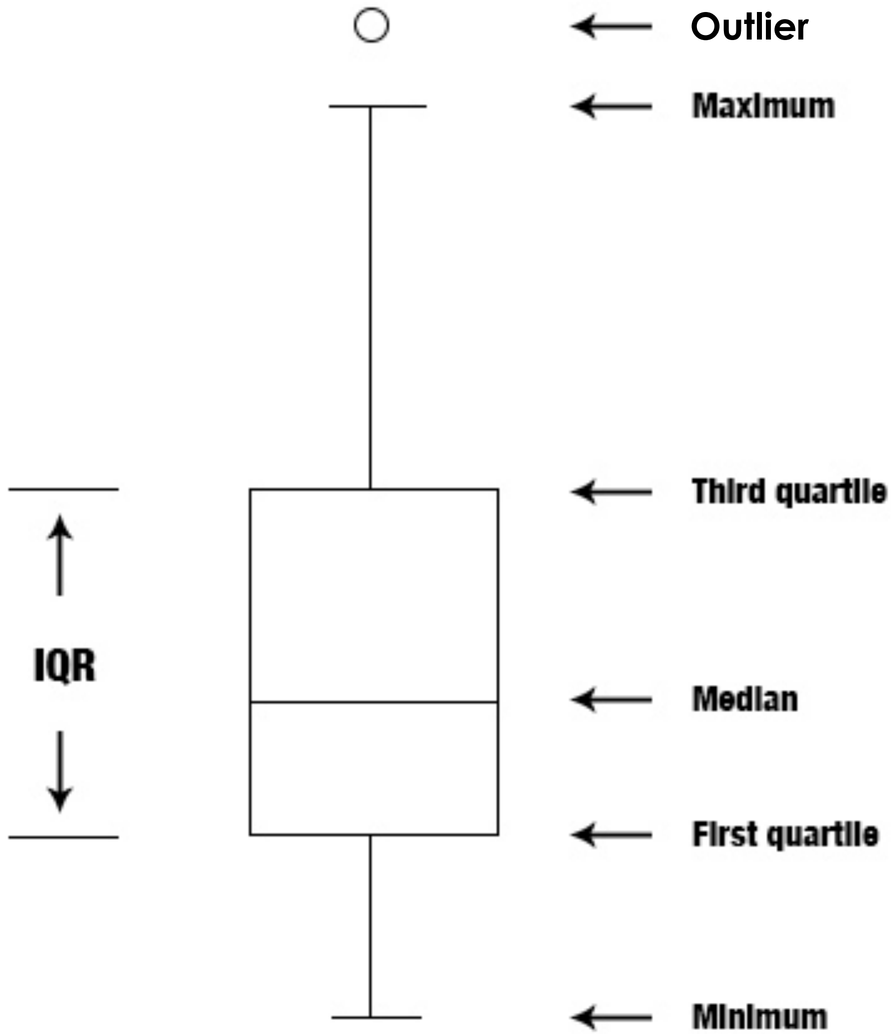
- Quartiles = dividing the distribution in 4
- Quantiles = dividing the distribution into 100
- 1st Quartile = 25th Quantile
- 3rd Quartile = 75th Quantile
- 2nd Quartile = 50th Quantile = Median
- IQR = 75th Quantile – 25th Quantile = 3rd Quartile – 1st Quartile

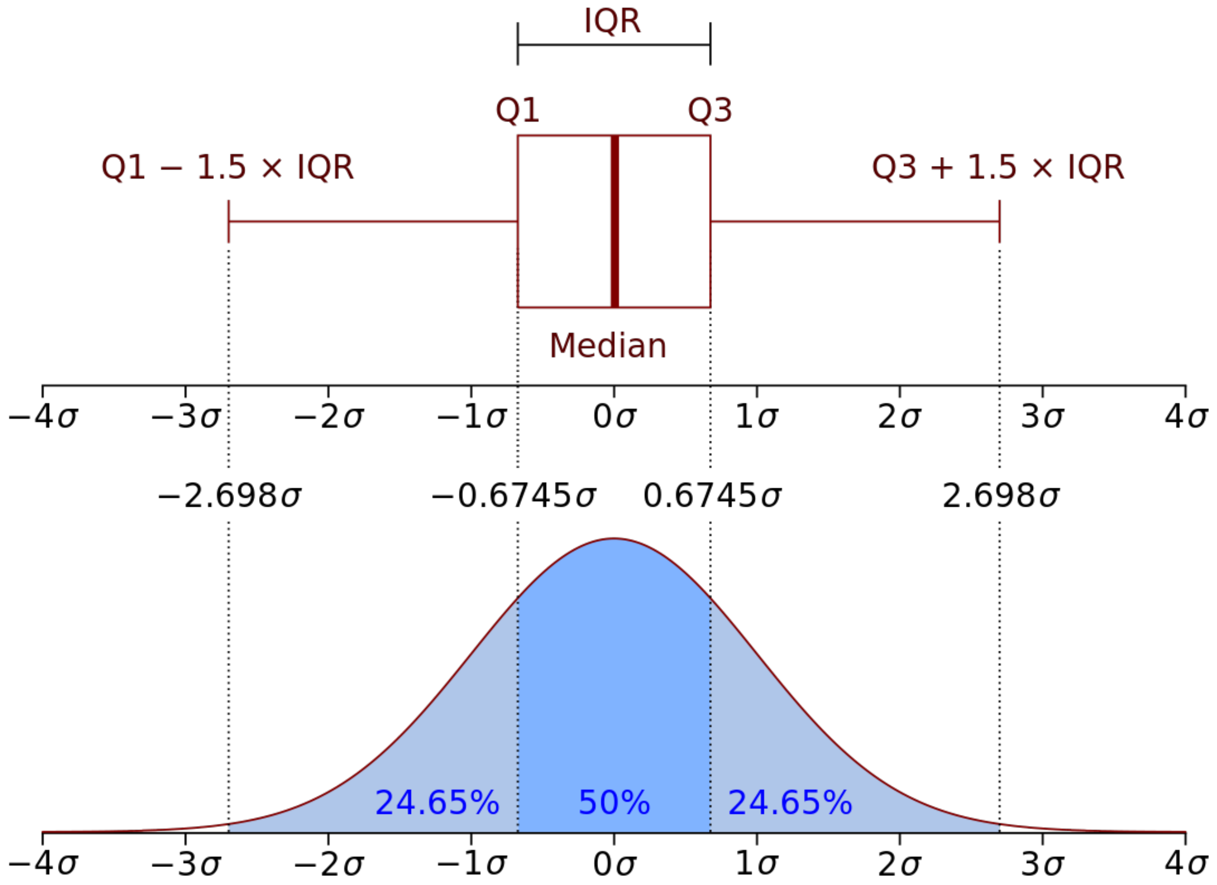
Visualising outliers - boxplots



https://commons.wikimedia.org/wiki/File:Box_plot_description.jpg

Visualising outliers - boxplots

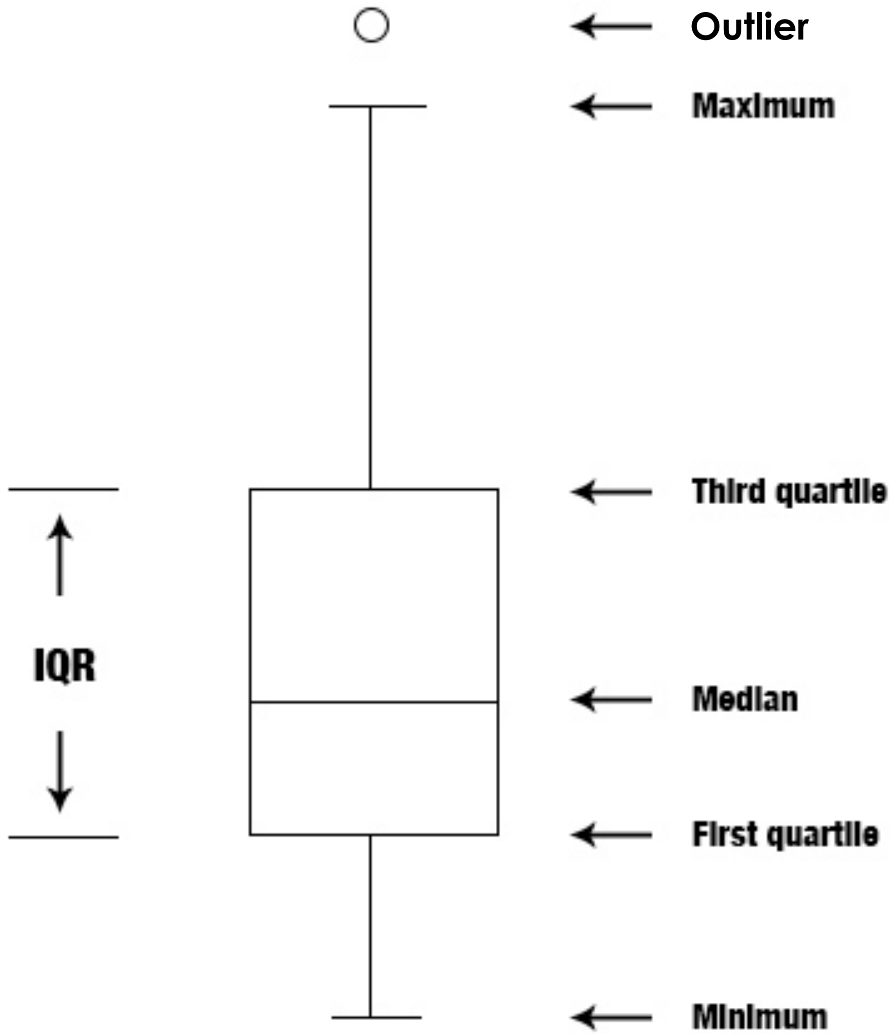


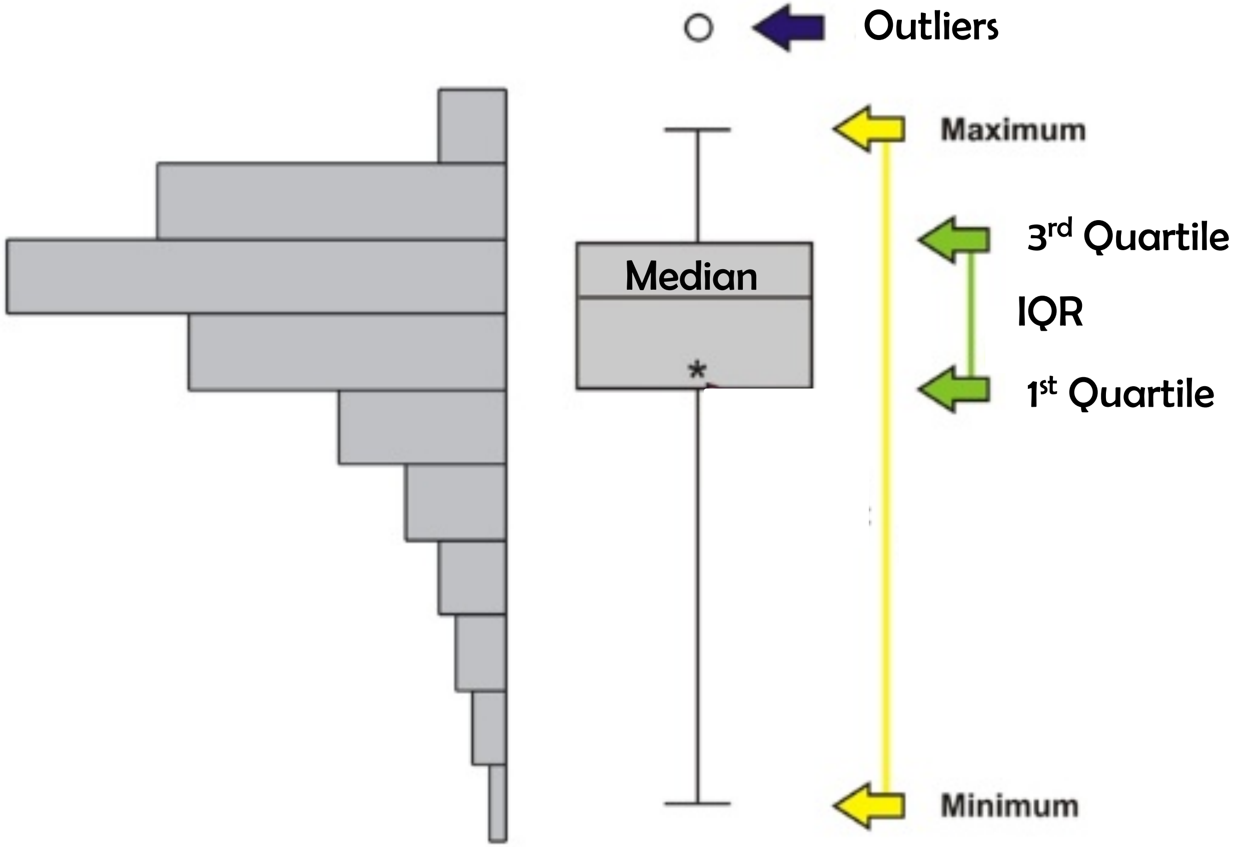


https://commons.wikimedia.org/wiki/File:Box_plot_description.jpg

https://commons.wikimedia.org/wiki/File:Boxplot_vs_PDF.svg

Visualising outliers - boxplots





https://commons.wikimedia.org/wiki/File:Box_plot_description.jpg