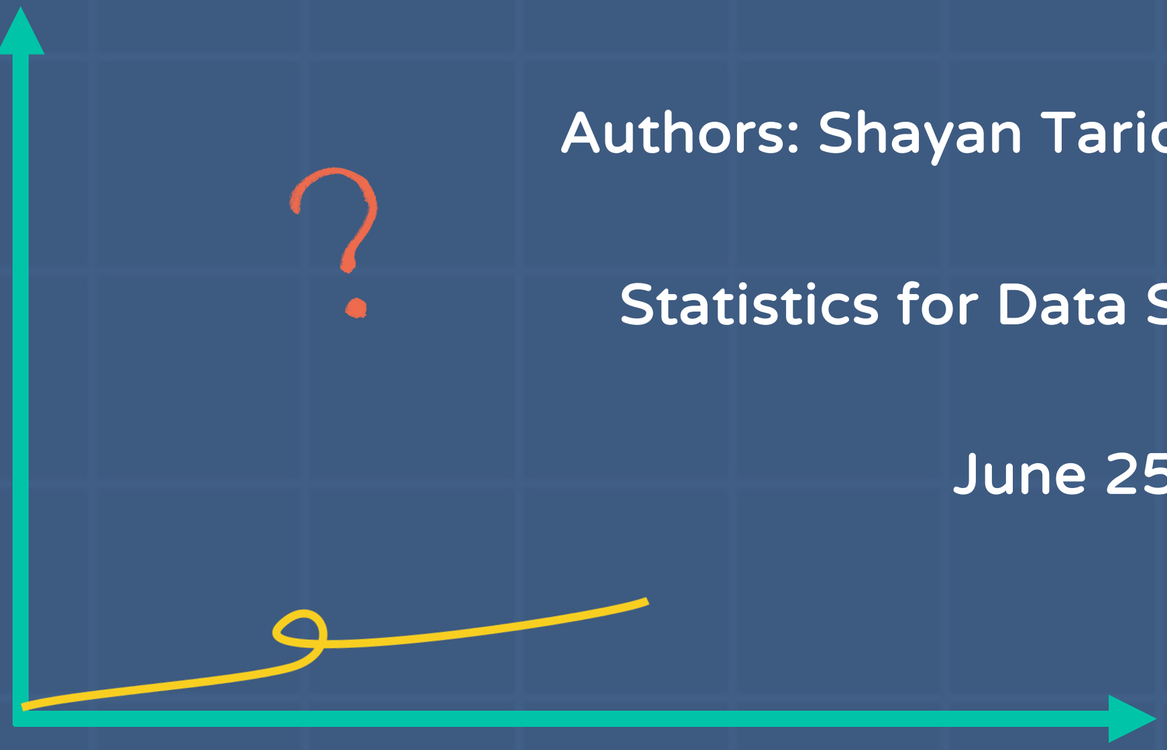
Estimating Measures of Location and Scale

Authors: Shayan Tariq, Muhammad Hussain

Statistics for Data Science Final Project

June 25, 2024



OUTLINE

- ✓ A Bootstrap Estimate of a Standard Error
- ✓ The Sample Trimmed Mean
- ✓ Estimators
- ✓ Comparisons of the Location Estimators
- ✓ Outlier Detection Methods

A Bootstrap Estimate of a Standard Error

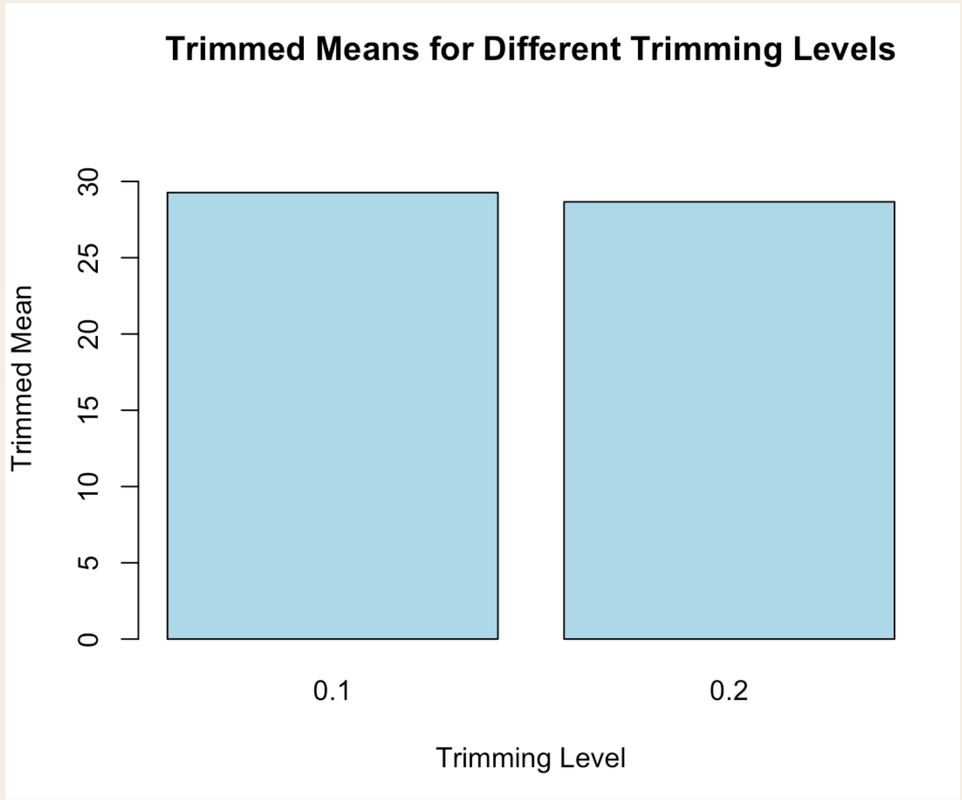
The bootstrap method, you can create many sets of data (called bootstrap samples) and calculate your estimator (like an average) for each set1. The variation among these estimates gives you an idea of the standard error.

- ✓ Original Sample Mean Age: 29.6991
- ✓ Bootstrap Standard Error of the Mean Age: 0.538734
- ✓ Original Sample Standard Error of the Mean Age: 0.5436405

The Sample Trimmed Mean

Sample Trimmed Mean: Describes how to calculate the trimmed mean by removing a certain percentage of the largest and smallest values from a data set.

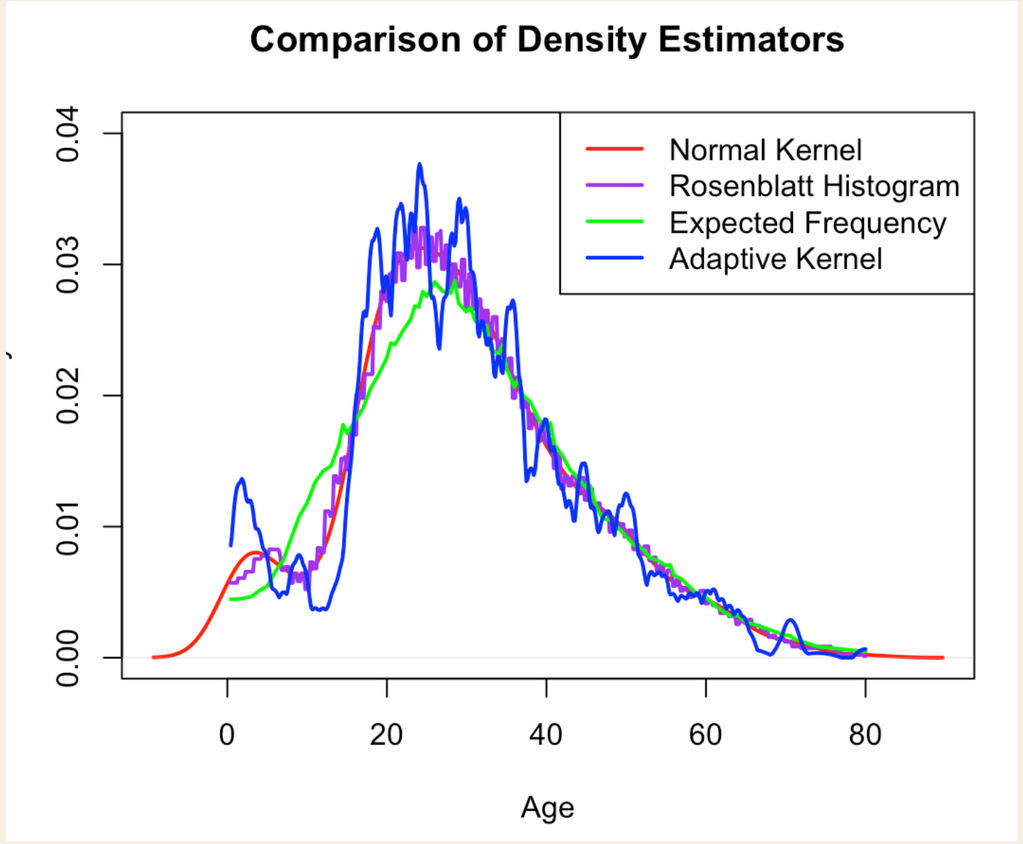
Trim_Level	Trimmed_Mean	SE
0.1	29.26923	0.5357630
0.2	28.66395	0.5276420



Density Estimators

Density estimators allows for the visualisation and analysis of data distribution patterns without assuming a specific form for the underlying distribution.

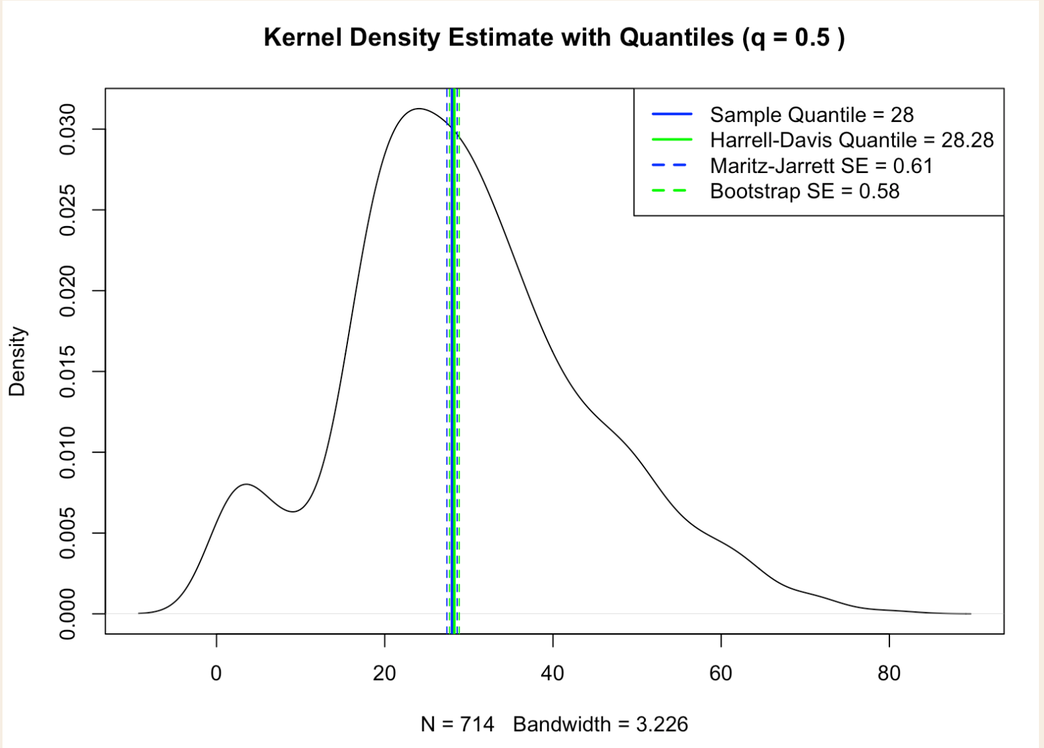
$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right),$$



Estimating Quantiles

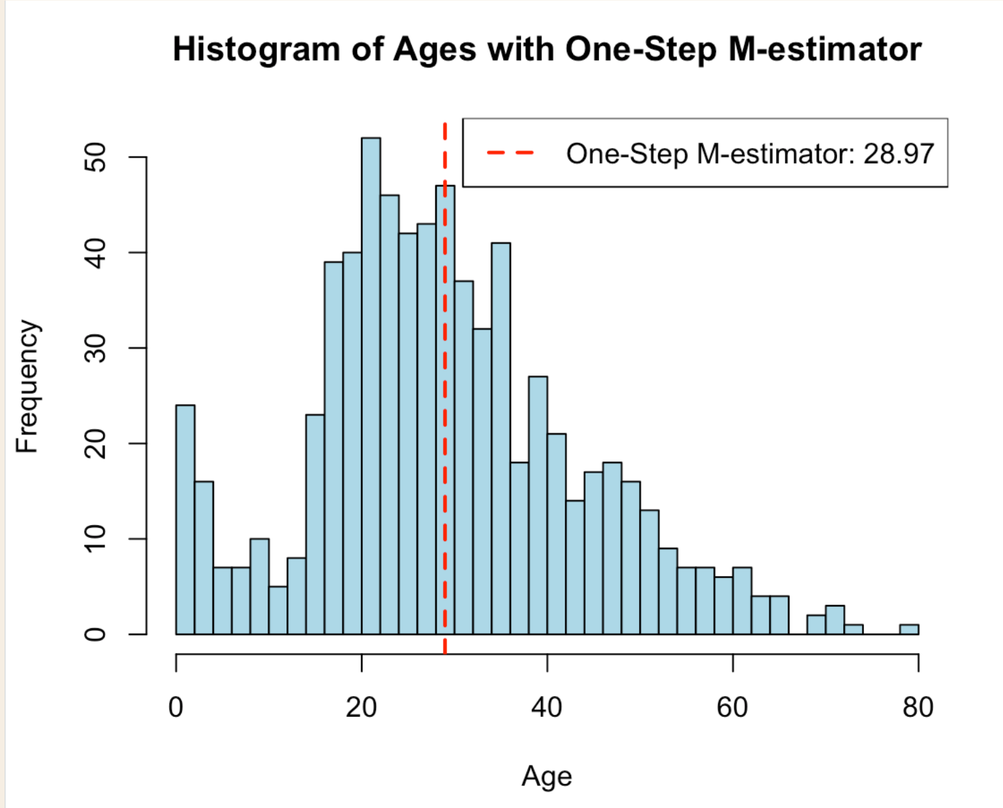
Quantiles are points in your data that divide the distribution into intervals with equal probabilities. Estimating quantiles involves finding these points accurately and understanding their variability, which is important for statistical analysis.

$$IF_q(x) = \begin{cases} \frac{q-1}{f(x_q)}, & \text{if } x < x_q \\ 0, & \text{if } x = x_q \\ \frac{q}{f(x_q)}, & \text{if } x > x_q, \end{cases}$$



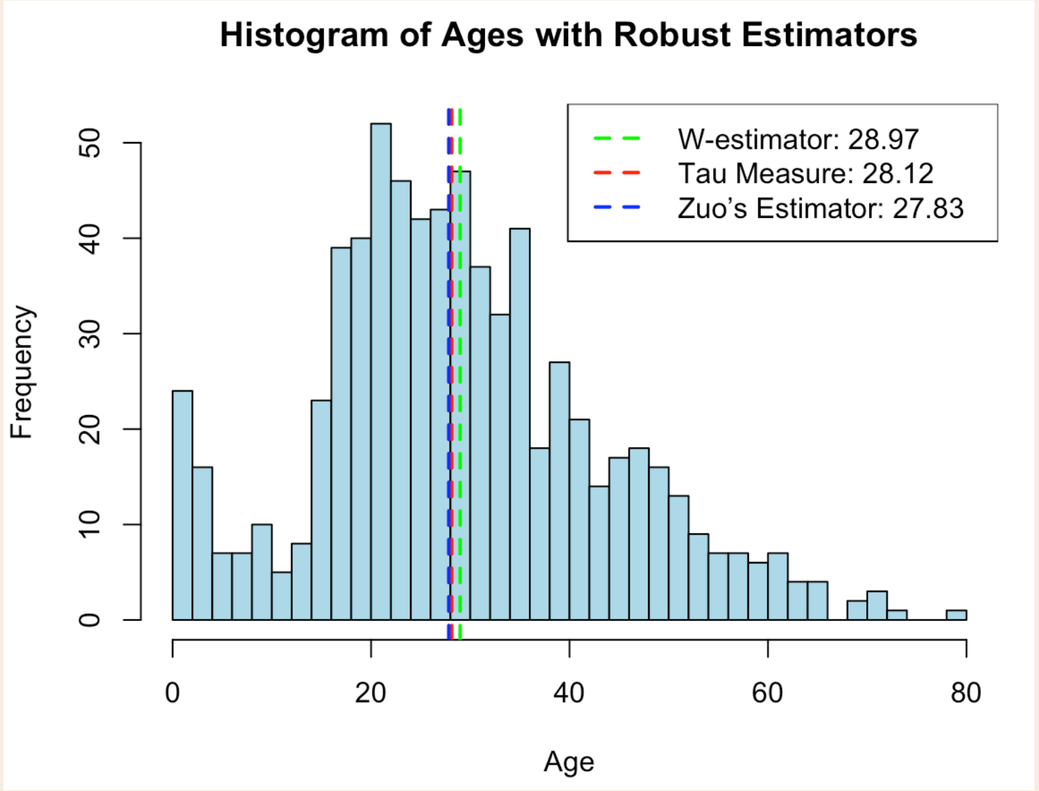
One-Step M-estimator

A One-Step M-estimator is a simplified robust estimator that provides an approximation to a full M-estimator with less computational effort. starts with an initial robust estimate, such as the median or trimmed mean, and refines it through one iteration of reweighting and adjustment.



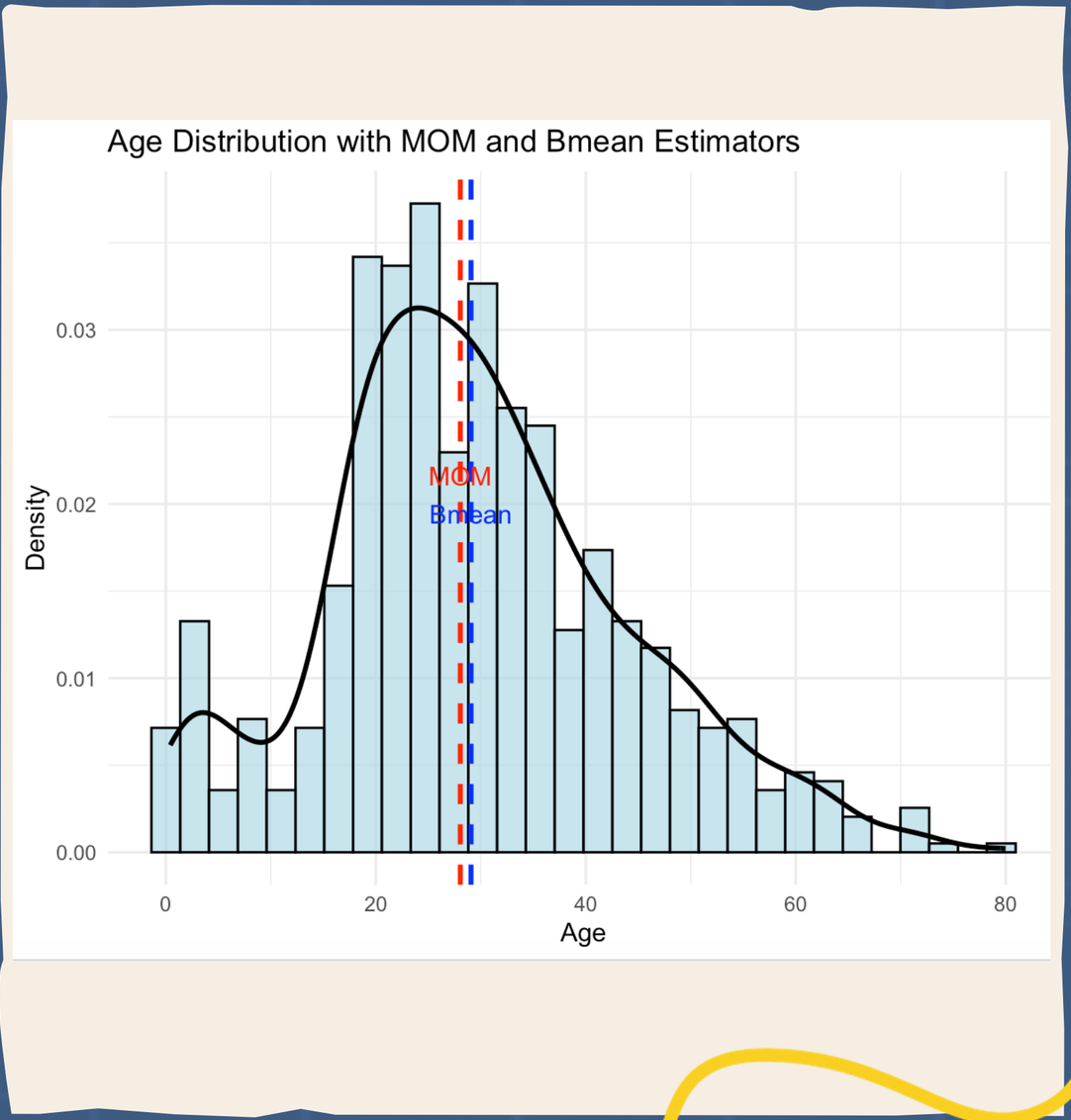
W-estimators

W-estimators are a class of robust estimators that use weights to downweight the influence of outliers, making them more robust against data points that deviate significantly from the rest of the data. W-estimators of location are closely related to M-estimators and usually they give identical results.



Skipped Estimators

Skipped estimators of location are robust statistical methods used to estimate the central tendency of a dataset while accounting for potential outliers. specific skipped estimators: the Modified One-Step M-estimator (MOM) and the Boxplot-based mean estimator.

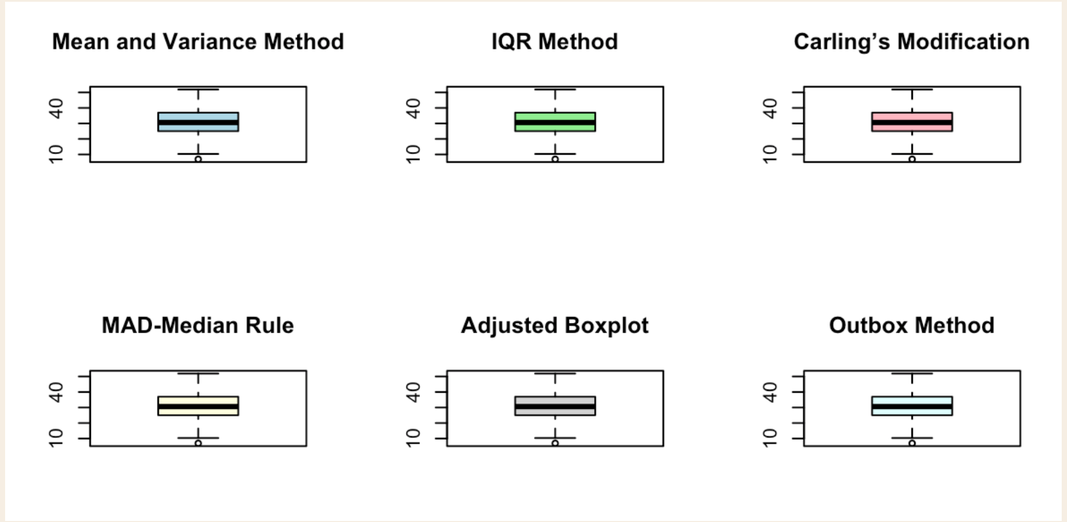


Comparisons of the Location Estimators

Comparing location estimators is crucial to understanding their performance and robustness under various conditions, such as the presence of outliers or deviations from normality. Here, we will compare several commonly used location estimators.

Outlier Detection Methods

Outlier detection is an essential part of data analysis as outliers can significantly influence the results of statistical analyses. Various methods are used to detect outliers, each with its strengths and weaknesses.



SUMMARY

- It presents both traditional and bootstrap methods for estimating standard errors, highlighting the practical importance and some counterintuitive aspects of these processes

- The chapter emphasizes the need for robust methods, including the detection and treatment of outliers and the use of graphical methods for data summarization.

Thanks