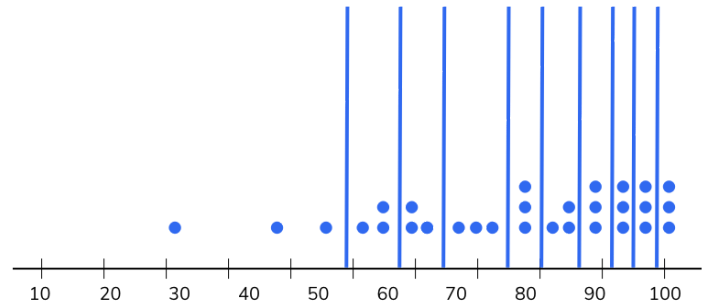


Quartiles, Quantiles, and Interquartile Range

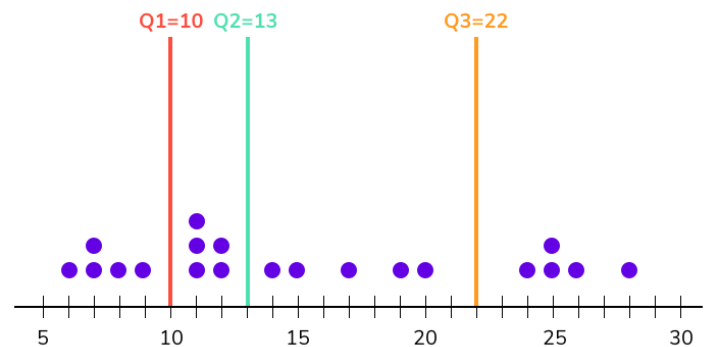
Quantiles

Quantiles are the set of values/points that divides the dataset into groups of equal size. For example, in the figure, there are nine values that splits the dataset. Those nine values are quantiles.



Quartiles

The three dividing points (or quantiles) that split data into four equally sized groups are called quartiles. For example, in the figure, the three dividing points Q1, Q2, Q3 are quartiles.



Numpy's Quantile() Function

In Python, the `numpy.quantile()` function takes an array and a number say `q` between 0 and 1. It returns the value at the `q` th quantile. For example,

`numpy.quantile(data, 0.25)` returns the value at the first quartile of the dataset `data`.

```
import numpy as np
data = [1,2,3,4,5]
first_quartile = np.quantile(data, 0.25)
```

Quantiles and Groups

If the number of quantiles is n , then the number of equally sized groups in a dataset is $n+1$.

Median in Quantiles

The median is the divider between the upper and lower halves of a dataset. It is the 50%, 0.5 quantile, also known as the 2-quantile.

```
# The value 5 is both the median and the 2-quantile
```

```
data = [1, 3, 5, 9, 20]
Second_quantile = 5
```

Interquartile Range Definition

The interquartile range is the difference between the first(Q1) and third quartiles(Q3). It can be mathematically represented as $IQR = Q3 - Q1$.

Interquartile Range and Outliers

The interquartile range is considered to be a robust statistic because it is not distorted by outliers like the average (or mean).

```
# Eventhough d_2 has an outlier, the IQR is identical for the 2 datasets
```

```
d_1 = [1,2,3,4,5,6,7,8,9]
d_2 = [-100,2,3,4,5,6,7,8,9]
```

Box Plot Values

The box in the box plot displays the dataset's median, first and third quartile, and the interquartile range. The line in the center of the box shows the median, the edges shows the first and third quartiles, and the interquartile range is visualized by the width of the box.

Usage of Side-by-side Box plots

The two datasets can be analyzed visually by placing two box plots side by side. This allows easy comparison of median, first and third quartiles and the IQR of the datasets.

Side-by-side Boxplots

In Python's Matplotlib library, if multiple datasets are specified in function `pyplot.boxplot()`, then those datasets will be visualized as side by side box plots.

```
mul_datasets = [[3, 5, 7, 2], [2, 4, 10, 43]]
pyplot.boxplot(mul_datasets)
```

Box Plot Outliers

In a box plot, the data points that fall beyond the whiskers are called outliers. They are usually labeled with a dot or an asterisk.

Box Plot Whiskers

A box plot's whiskers are the lines that extends from the 1st or 3rd quartile to points farthest from the median. The upper whisker of the box plot is the largest dataset number smaller than $1.5IQR$ above the third quartile and the lower whisker is the smallest dataset number larger than $1.5IQR$ below the first quartile.

Boxplot in Matplotlib

In Python's Matplotlib library, the `pyplot.boxplot()` function takes a dataset as input and returns a box plot.

```
# dataset= list of numbers  
pyplot.boxplot(dataset)
```