

# Exploratory Data Analysis [ EDA ]

## Statistics - Descriptive Statistics

Descriptive statistics gives us insight into data without having to look at all of it in detail.

### **The Variation of the Data**

Measures of variation are statistics of how far away the values in the observations (data points) are from each other.

There are different measures of variation. The most commonly used are:

Range

Quartiles and Percentiles

Interquartile Range

Standard Deviation

Measures of variation combined with an average (measure of center) gives a good picture of the distribution of the data.

## Statistics - Range

The range is a measure of variation, which describes how spread out the data is.

### Calculating the Range

The range can only be calculated for numerical data.

First, find the smallest and largest values of this example:

13, 21, 21, 40, 48, 55, 72

Calculate the difference by subtracting the smallest from the largest:

$$72 - 13 = \underline{59}$$

## Statistics - Quartiles and Percentiles

Quartiles and percentiles are a measures of variation, which describes how spread out the data is.

Quartiles and percentiles are both types of **quantiles**.

### Quartiles

**Quartiles** are values that separate the data into four equal parts.

```
import numpy
```

```
values = [13,21,21,40,42,48,55,72]
```

```
x = numpy.quantile(values, [0,0.25,0.5,0.75,1])
```

```
print(x)
```

## Percentiles

**Percentiles** are values that separate the data into 100 equal parts.

```
import numpy
```

```
values = [13,21,21,40,42,48,55,72]
```

```
x = numpy.percentile(values, 65)
```

```
print(x)
```

## Statistics - Interquartile Range

Interquartile range is a measure of variation, which describes how spread out the data is.

```
from scipy import stats
```

```
values = [13,21,21,40,42,48,55,72]
```

```
x = stats.iqr(values)
```

```
print(x)
```

## Statistics - Standard Deviation

Standard deviation is the most commonly used measure of variation, which describes how spread out the data is.

```
import numpy
```

```
values = [4,11,7,14]
```

```
x = numpy.std(values)
```

```
print(x)
```