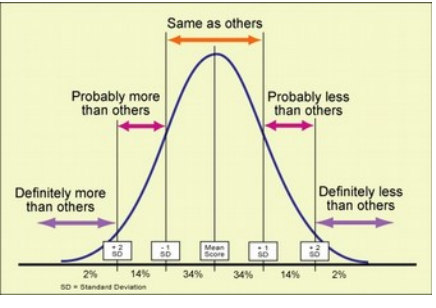
Statistics

Statistics attempts to describe a **typical representative** of a given set of individuals.

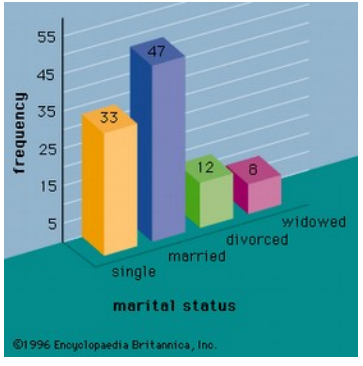
This is typical in scenario, if there is a big number of samples with different properties, or more measurement of the same phenomenon, however each measurement burdened with some error.

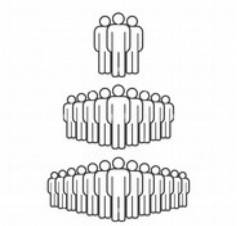
The term **average** (in some sense) is usually used in order to describe ‘typical representative’ of the given set (=data or dataset).
How much can the chosen average describe the characteristics of the set depends on the character and **variability** of the data.

Examples: Economic indicators (GDP, Per capita income,..), sociology (opinions, election polls,..) measurements (meteorology, health, experiments,..), etc.

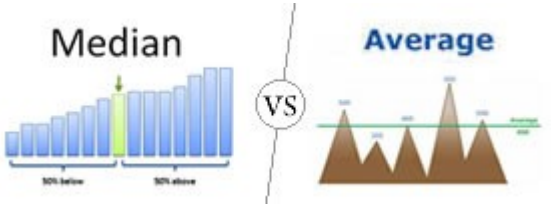


z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949





Average



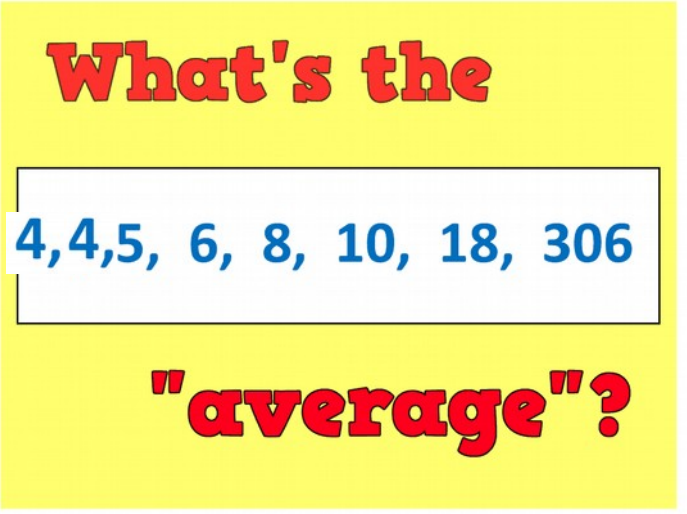
There are 3 ways how to describe ‘typical’ or ‘average’ representative of the set:

Mean
= arithmetic average; it helps to understand the overall value magnitude

$$\bar{x} = \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right) = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

Median
= the value of the middle element in the set, which is sorted

Mode
= the most frequent element of the set



There are 15 boys.

Height (cm)	160–164	165–169	170–174	175–179	180–184
Number	2	5	4	3	1

Calculate the mean and median height of boys. Determine also mode.

In the test, 15 students received a grade 1, 35 students received a grade 2, 30 students received a grade 3, 15 students received a grade 4 and the remaining 5 students received a grade 5.
Calculate the average test grade, modus, median.
Represent the test results graphically.

Median vs Average

MEDIAN	AVERAGE
exact middle of a list of values	sum of values divided by number of values
1, 3, 3, **6**, 7, 8, 9 median = 6	1, 3, 3, 6, 7, 8, 9 average = (1 + 3 + 3 + 6 + 7 + 8 + 9) / 7 average = 5.3
1, 3, **4**, **5**, 6, 9 median = (4 + 5) ÷ 2 median = 4.5	1, 3, 4, 5, 6, 9 average = (1 + 3 + 4 + 5 + 6 + 9) / 6 average = 4.7

Name & Meaning	Formula / Example	Used for
Arithmetic Mean [average]	$\frac{sum}{size} = \frac{a+b+c}{3}$	Most situations ("average item")
Median [middle value]	Middle of sorted list (2 middles? Average 'em)	Wildly varying samples (houses, incomes)
Mode [most popular]	Most popular value	No compromises (winner takes all)
Geometric Mean [average factor]	$\sqrt[3]{abc}$	Investments, growth, area, volume
Harmonic Mean [average rate]	$\frac{3}{\frac{1}{a}+\frac{1}{b}+\frac{1}{c}}$	Speed, production, cost

Measures of average

Mean = sum of all data ÷ number of pieces of data

Median = middle value when data is arranged in size order

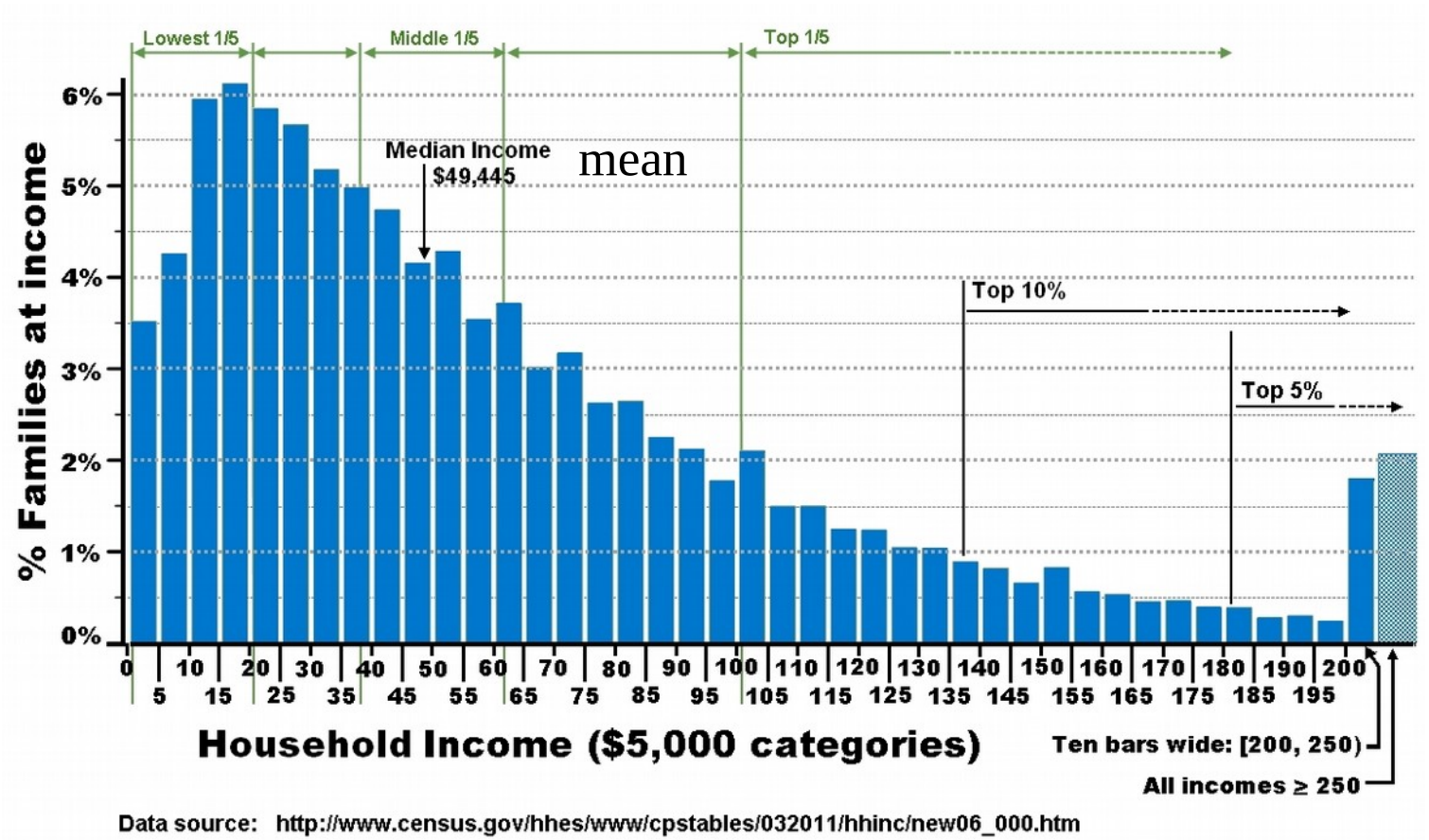
Mode = the most common data value

Eg find the mean, median and mode of these numbers:
2, 1, 4, 7, 1

Mean	Median	Mode
Sum of data = 2+1+4+7+1 = 15 15 ÷ 5 = 3 **mean = 3 goals**	Data in size order: ~~1~~, ~~1~~, (2), ~~4~~, ~~7~~ **median = 2 goals**	Most common value is 1 **mode = 1 goal**

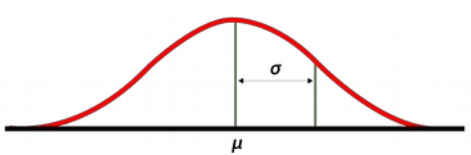
Histogram

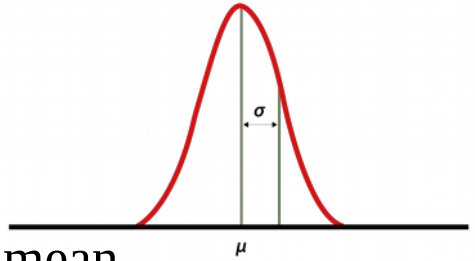
A histogram is a visual representation of the distribution of numeric data.



Variability

There are 3 ways how to describe ‘typical departure’ from ‘typical value’ of the set.
(This can be also understand with help of histogram: how much is distribution stretched.)



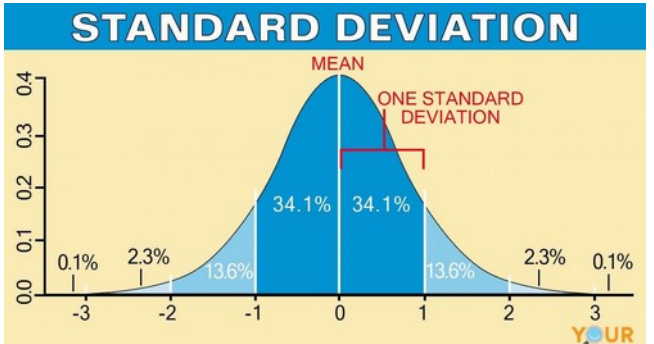


Variance
= is a measure of how far a set of numbers is spread out from mean
. denoted as Var(x) or σ^2

Var(x) = σ^2 and see definition below

Standard deviation
= is a different measure of the amount of variation of elements about its mean
= typically denoted as σ

$$\sigma = \sqrt{\frac{1}{N}\left[(x_1-\mu)^2+(x_2-\mu)^2+\cdots+(x_N-\mu)^2\right]}, \text{ where } \mu = \frac{1}{N}(x_1+\cdots+x_N),$$

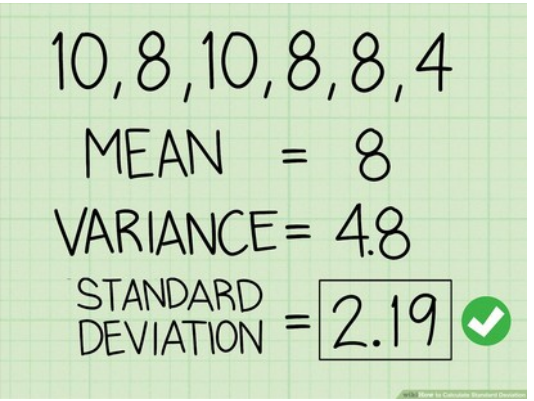


Standard deviation

Let’s have a set of numbers:

Then we need to compute **mean.**

Based on it **variance.**

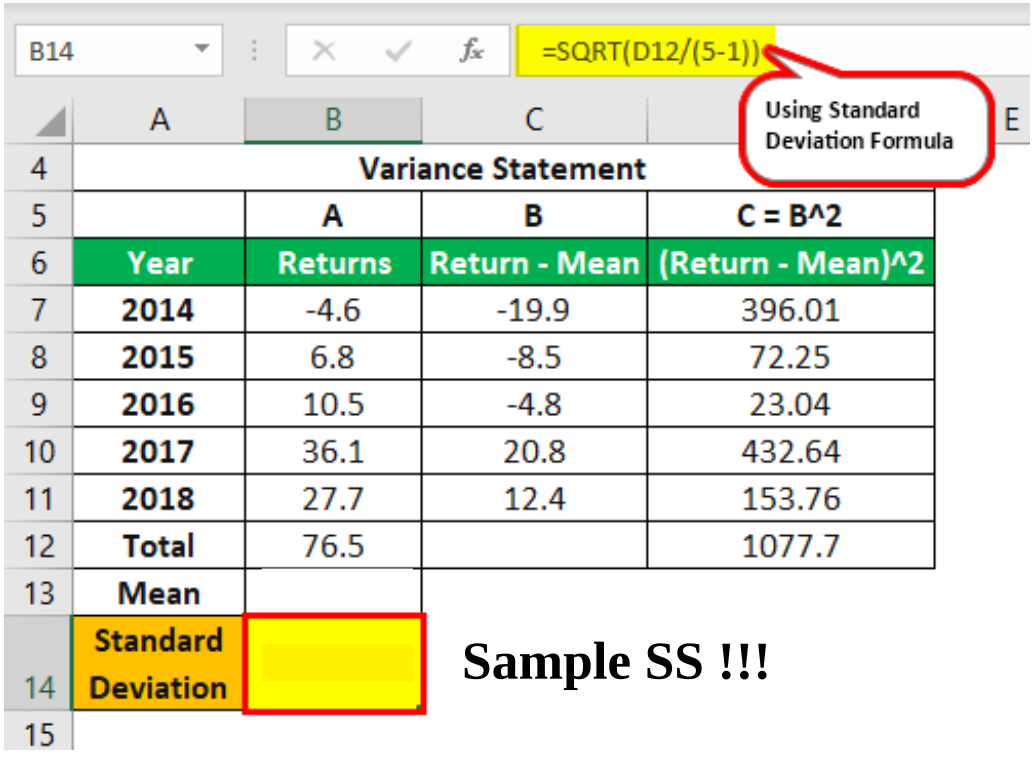


And standard deviation is already **square root of it.**

* There is also a sample standard deviation s, for explanation of a tiny difference, see web. Very nice page, which I suggest to read.

** For explanation of standard error, see web.
*** Covariance vs. correlation vs. variance, see web2.

Population	Sample
$\sigma = \sqrt{\frac{\Sigma(x_i-\mu)^2}{n}}$	$S = \sqrt{\frac{\Sigma(x_i-\bar{x})^2}{n-1}}$
μ - Population Average xi - Individual Population Value n - Total Number of Population	X - Sample Average x_i - Individual Population Value n - Total Number of Sample



	A	B	C	E
4		Variance Statement		
5		A	B	C = B^2
6	Year	Returns	Return - Mean	(Return - Mean)^2
7	2014	-4.6	-19.9	396.01
8	2015	6.8	-8.5	72.25
9	2016	10.5	-4.8	23.04
10	2017	36.1	20.8	432.64
11	2018	27.7	12.4	153.76
12	Total	76.5		1077.7
13	Mean			
14	Standard Deviation			
15				

Sample SS !!!

Range (in statistics)

= the range of a set of data is the difference between the largest and smallest values

How to Calculate the Range:

Example Values

12, 7, 17, 9, 16, 29, 8, 4, 15, 7, 29, 6, 17, 2, 5, 12, 8, 17, 6, 6, 17, 17

Step 1:
Sort all the values from low to high.

2, 4, 5, 6, 6, 6, 7, 7, 8, 9, 12, 12, 15, 16, 17, 17, 17, 17, 17, 29

Step 2:
Subtract the lowest value from the highest value to determine the range.

2, 4, 5, 6, 6, 6, 7, 7, 8, 9, 12, 12, 15, 16, 17, 17, 17, 17, 17, 29

Range = 29 - 2 = 27

What's the

4, 5, 6, 8, 10, 18, 306

range?

In checking the weight of the biscuits, 10 boxes of biscuits were checked and the following values were found:
250g, 247g, 251g, 249g, 252g, 248g, 251g, 250g, 251g, 248g.

Calculate the average weight of a box of biscuits, variance and the standard deviation.

Tba!

There are 4 classes in school – A, B, C and D.

Class	Students	Average math mark
A	27	2,08
B	25	2,18
C	26	2,70
D	22	2,37

What is the average math mark in the whole school?

The average mark of a pupil (for whole school) is determined
when the sum of the marks of all pupils divided by the total number of pupils in the year.
Total number of pupils in the year. . . . . [???]
Sum of marks in class A: Number of pupils in A multiplied by the average grade.
Sum of all grades ……... . . . . [233]
Response. The average grade of a second year pupil is . . . . . [???]