Standard Deviation

A measure of how spread out data are.

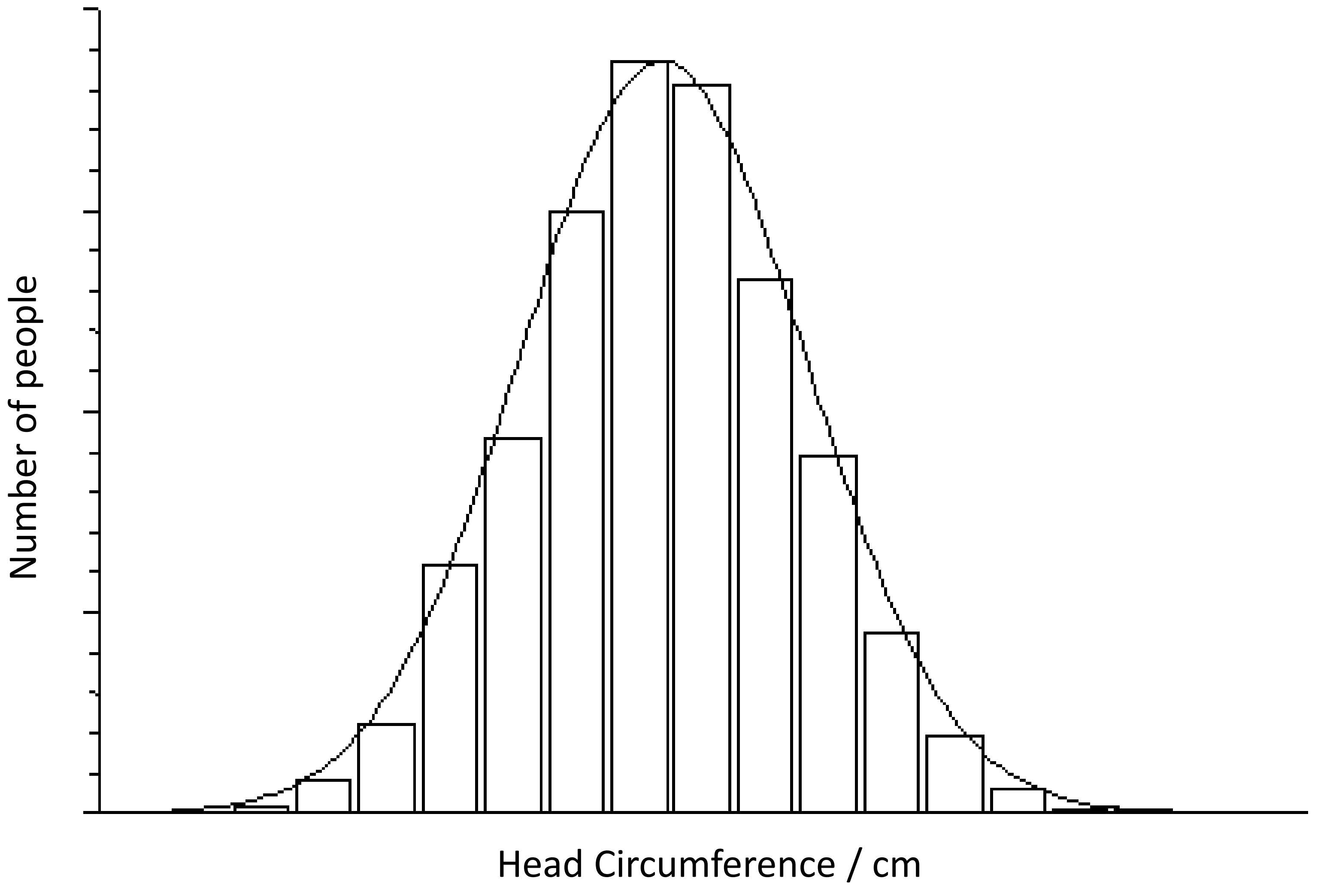
So, what is standard deviation?

- The Standard Deviation is a measure of how spread out numbers are.
- Its symbol is **σ** (the greek letter sigma)
- The formula is easy: it is the **square root** of the **Variance.** So now you ask, "What is the Variance?"

Variance

- Defined as:
  - The average of the squared differences from the mean.
- To calculate the variance follow these steps:
  - Work out the Mean (the simple average of the numbers)
  - Then for each number: subtract the Mean and square the result (the *squared difference*).
  - Then work out the average of those squared differences.

Example

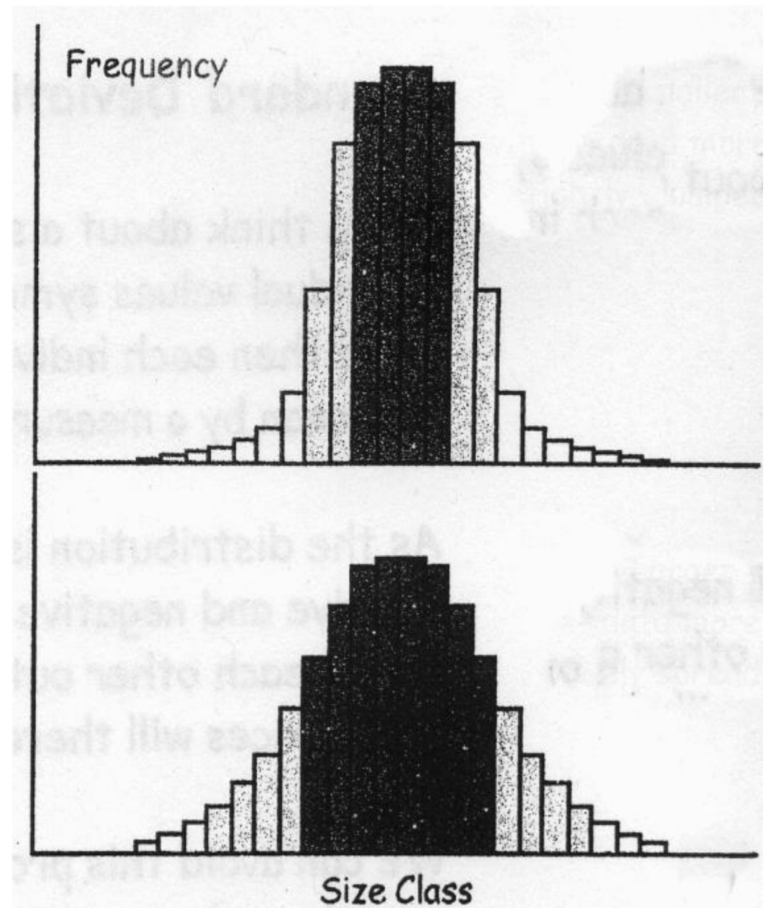


- Normal distributions can vary in how big the spread of values is

- All normal distributions can be described if we know:

- The central point

- The degree of spread

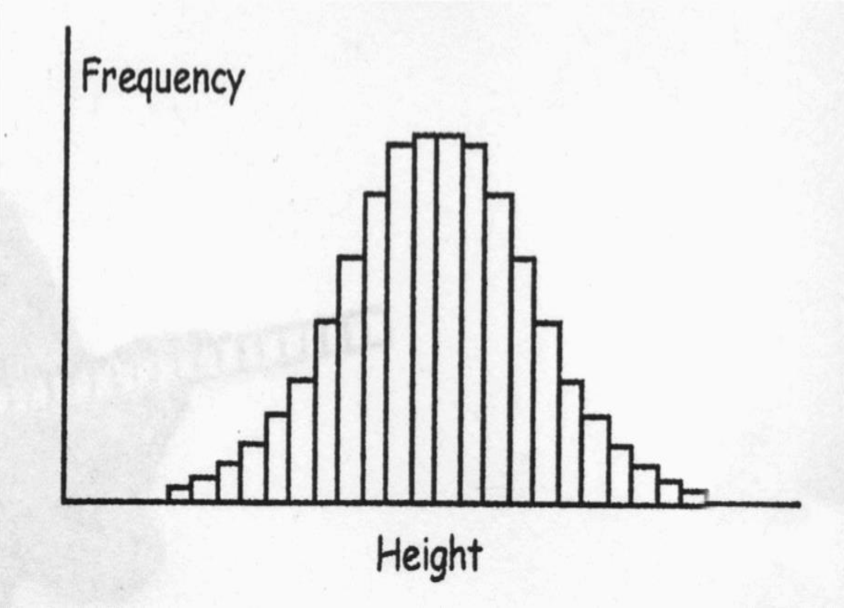


Central Tendency

- **Central tendency** is indicated by the average

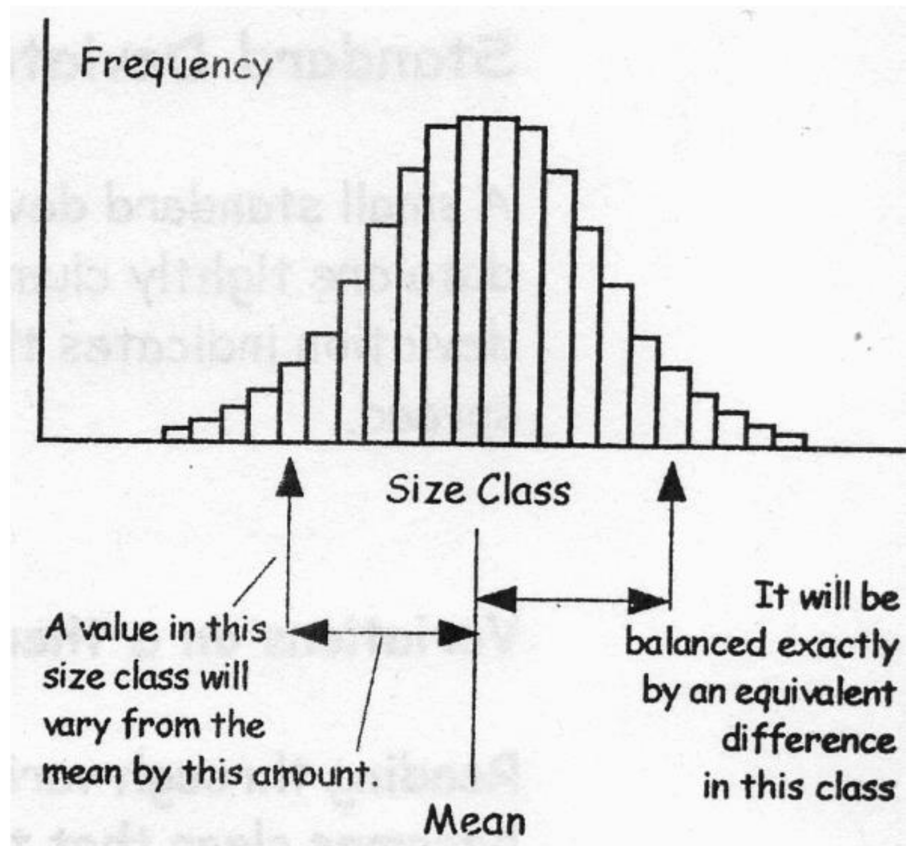
- In normal data the mean, median and mode are the same

- Usual to use **mean**



Spread

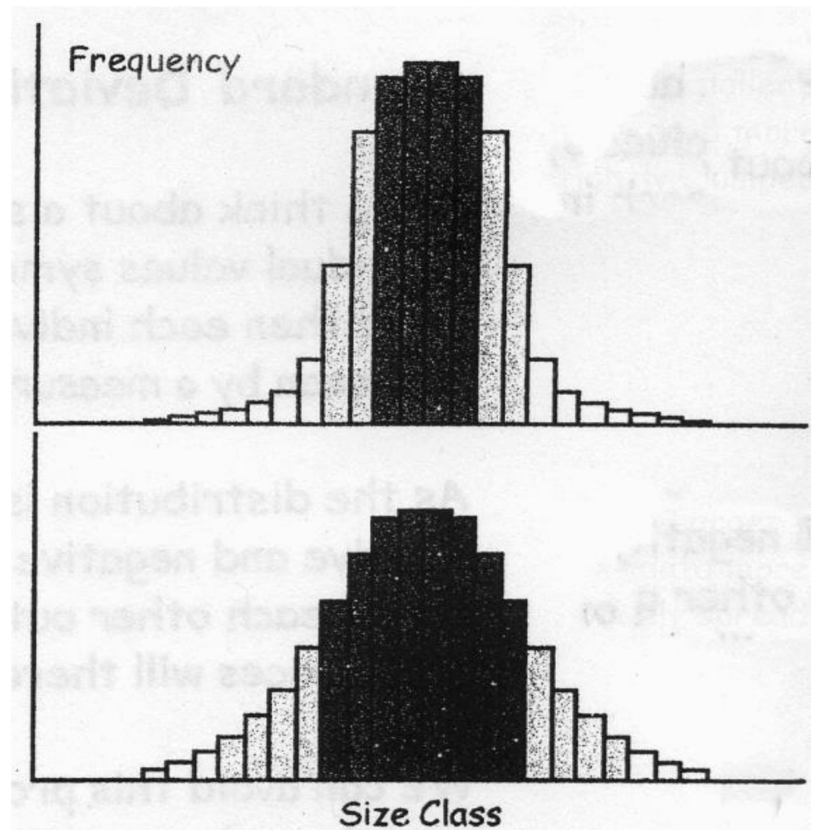
- Each data point varies from the mean by a measurable amount
- We use these values to calculate a **standard deviation**



Standard Deviation 1

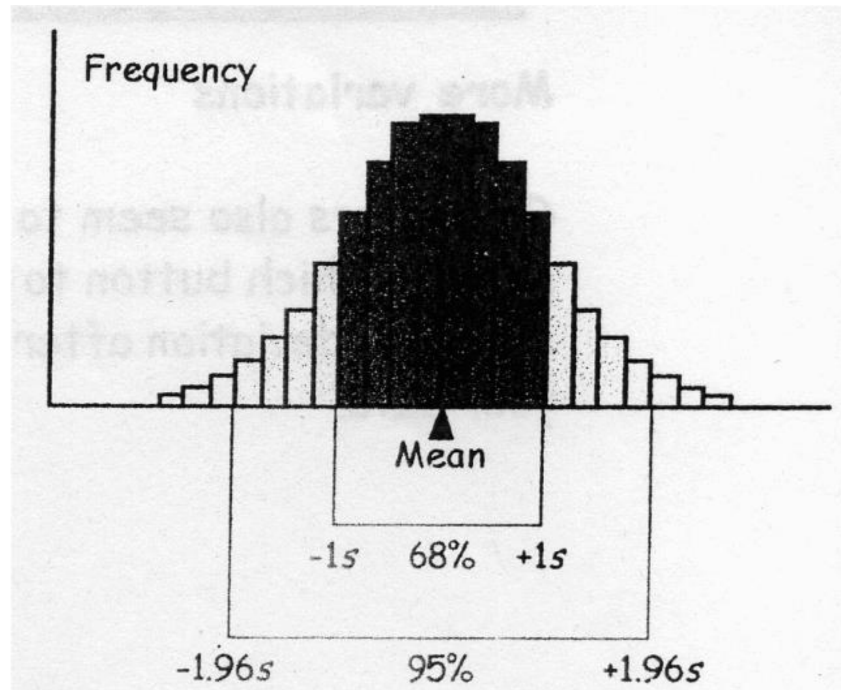
- Data tightly **clumped** -
- **Small** standard deviation

- Data more widely **spread** -
- **Large** standard deviation



Standard Deviation 2

- Whether large or small
- One standard deviation each way encloses 68% of all the values
- 1.96 *S* enclose 95% of all the values



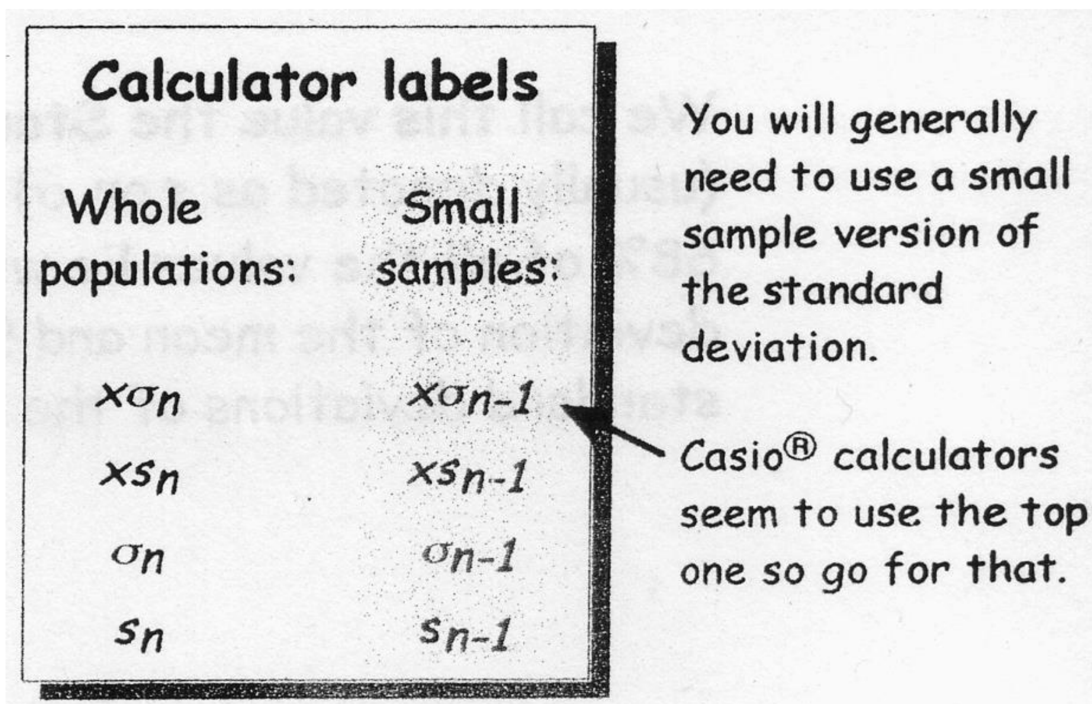
Standard Deviation 3

- There are several versions of the formula

Time consuming as need individual deviations to be calculated and can introduce rounding errors.	Quicker methods, which are less prone to rounding errors.	**Which formula?**
$s = \sqrt{\dfrac{\sum(\overline{x} - x)^2}{n}}$	$s = \sqrt{\dfrac{\sum x^2 - \dfrac{(\sum x)^2}{n}}{n}}$	If we could measure the entire population.
$s = \sqrt{\dfrac{\sum(\overline{x} - x)^2}{n - 1}}$	$s = \sqrt{\dfrac{\sum x^2 - \dfrac{(\sum x)^2}{n}}{n - 1}}$	Small samples from an entire population tend to underestimate the actual standard deviation so we use $n - 1$ as the divisor.

Standard Deviation 4

- Fortunately, calculators can give you the standard deviation at the touch of a button…..but which one?



Calculator labels

Whole populations:	Small samples:
$x\sigma_n$	$x\sigma_{n-1}$
xs_n	xs_{n-1}
σ_n	σ_{n-1}
s_n	s_{n-1}

You will generally need to use a small sample version of the standard deviation.

Casio® calculators seem to use the top one so go for that.

Real Data Ivy Leaves

- Once we have some data we need to…
- …check if they are normally distributed

n	Width / mm
1	63
2	74
3	46
4	46
5	47
6	53
7	66
8	47
9	54
10	54
11	27
12	36
13	59
14	35
15	35
16	57
17	38
18	58
19	48
20	40
21	48
22	40
23	54
24	41
25	55
26	41
27	42
28	45
29	21
30	33
Sum	1403
Mean	46.8

Checking normality…

- …draw out a **tally chart**
- Calculate difference between smallest and largest values
- Divide by 10 to get size class interval
- Calculate size classes
- Tally up measurements into their respective size classes
- Do the data appear Normal?

Tally Chart

Smallest = 21
Largest = 74
Difference = 53
Interval = 53/10 = 5.3

Size classes

20 - 25	I
26 - 31	I
32 - 37	IIII
38 - 43	IIIIII
44 - 49	IIIIIII
50 - 55	IIIII
56 - 61	III
62 - 67	II
68 - 73	
74 - 79	I

Standard Deviation 5

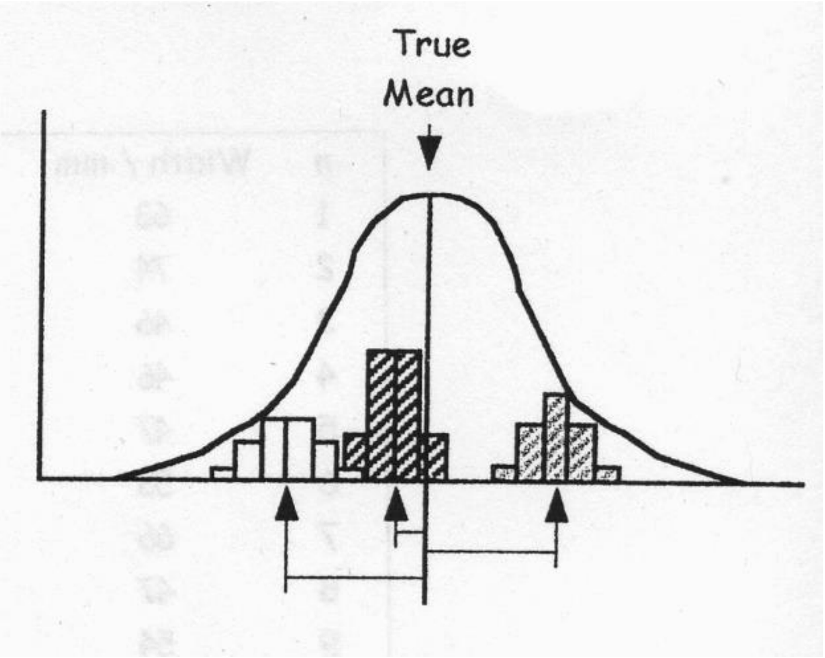
- Then we can calculate the mean and standard deviation for the data set:
- $\overline{X} = 46.8$
- $S = 11.63$

n	Width / mm
1	63
2	74
3	46
4	46
5	47
6	53
7	66
8	47
9	54
10	54
11	27
12	36
13	59
14	35
15	35

n	Width / mm
16	57
17	38
18	58
19	48
20	40
21	48
22	40
23	54
24	41
25	55
26	41
27	42
28	45
29	21
30	33

Standard Error

- The means of the samples vary from the true mean by a measurable amount
- The means form a narrow normal distribution around the population mean
- The standard deviation of these means is called the **Standard Error**



Standard Error 2

Calculate the standard error as: Standard error = $s / \sqrt{n}$

where s is the standard deviation of the sample and n the number of measurements in the sample

So for the first set of leaves: Standard error = $11.63 / \sqrt{30} = 2.12$

- The standard error is the standard deviation divided by the square root of the number of measurements, *n.*