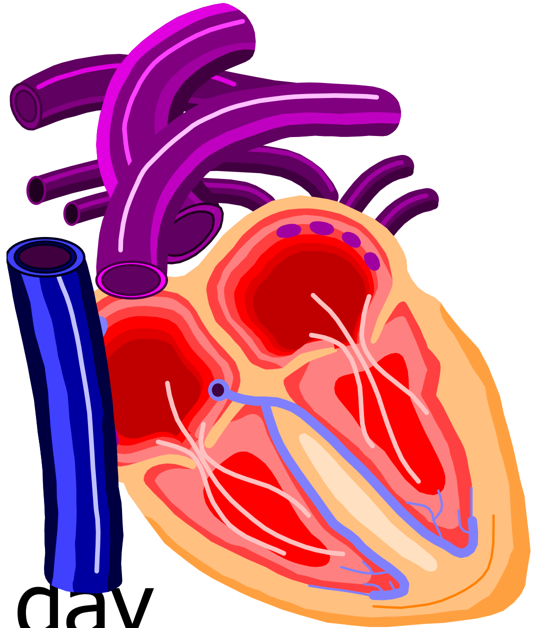
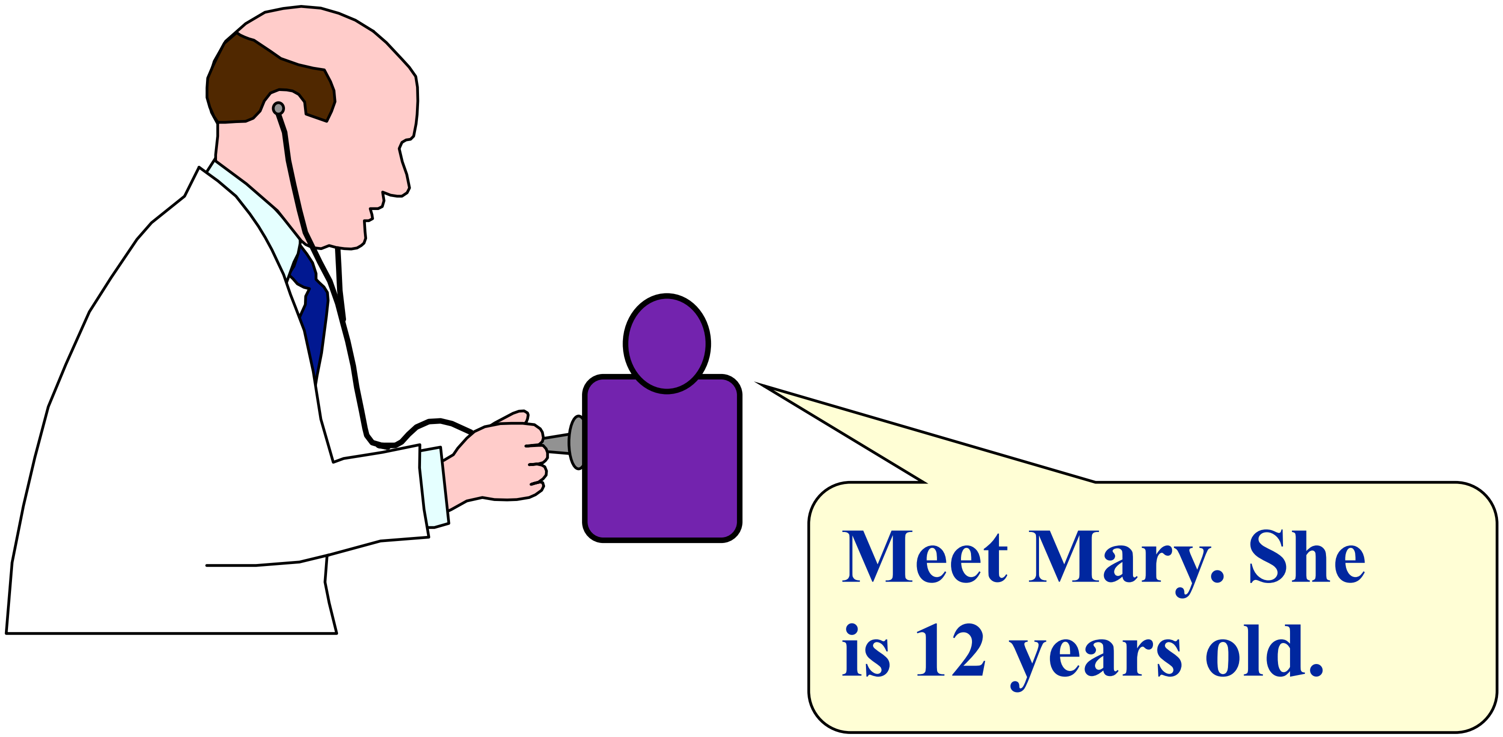
Statistics for beginners



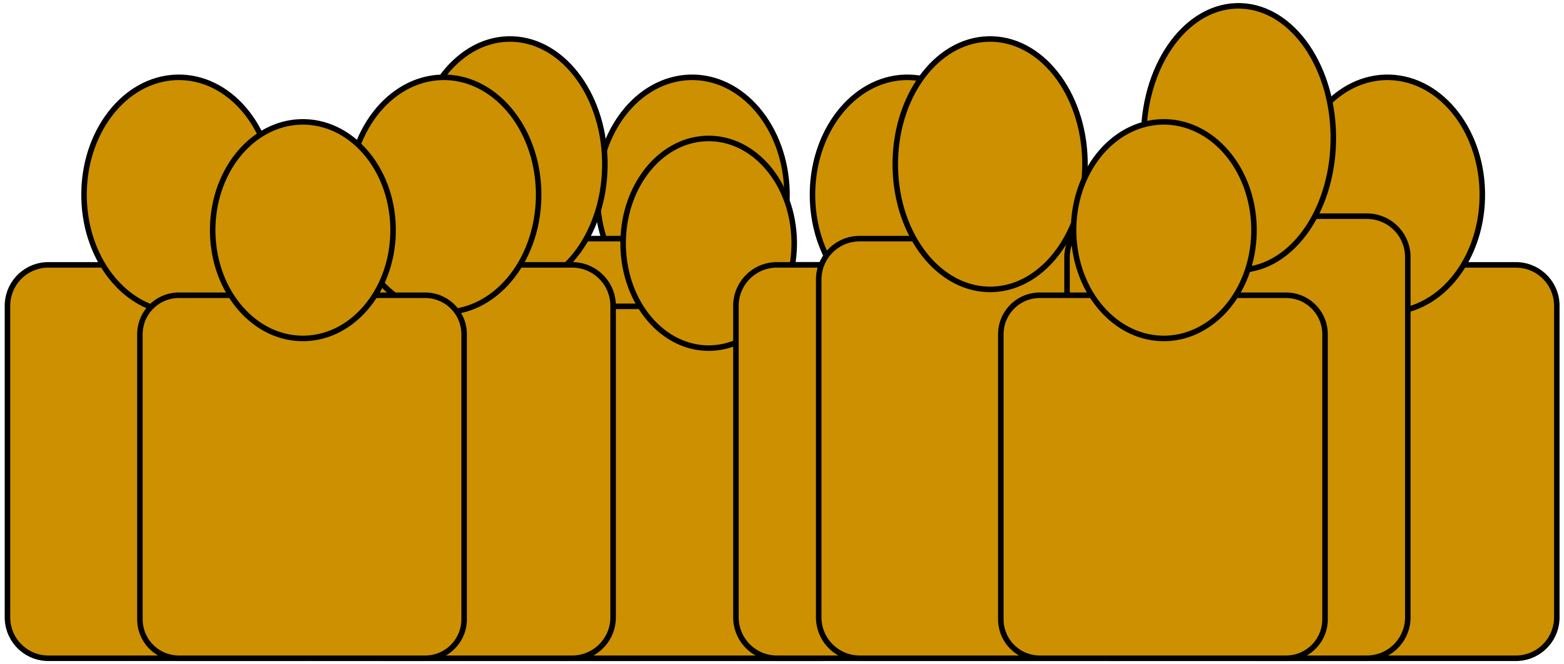
- It is fun
- You can understand it
- You use it intuitively every day

A DAY AT THE OFFICE



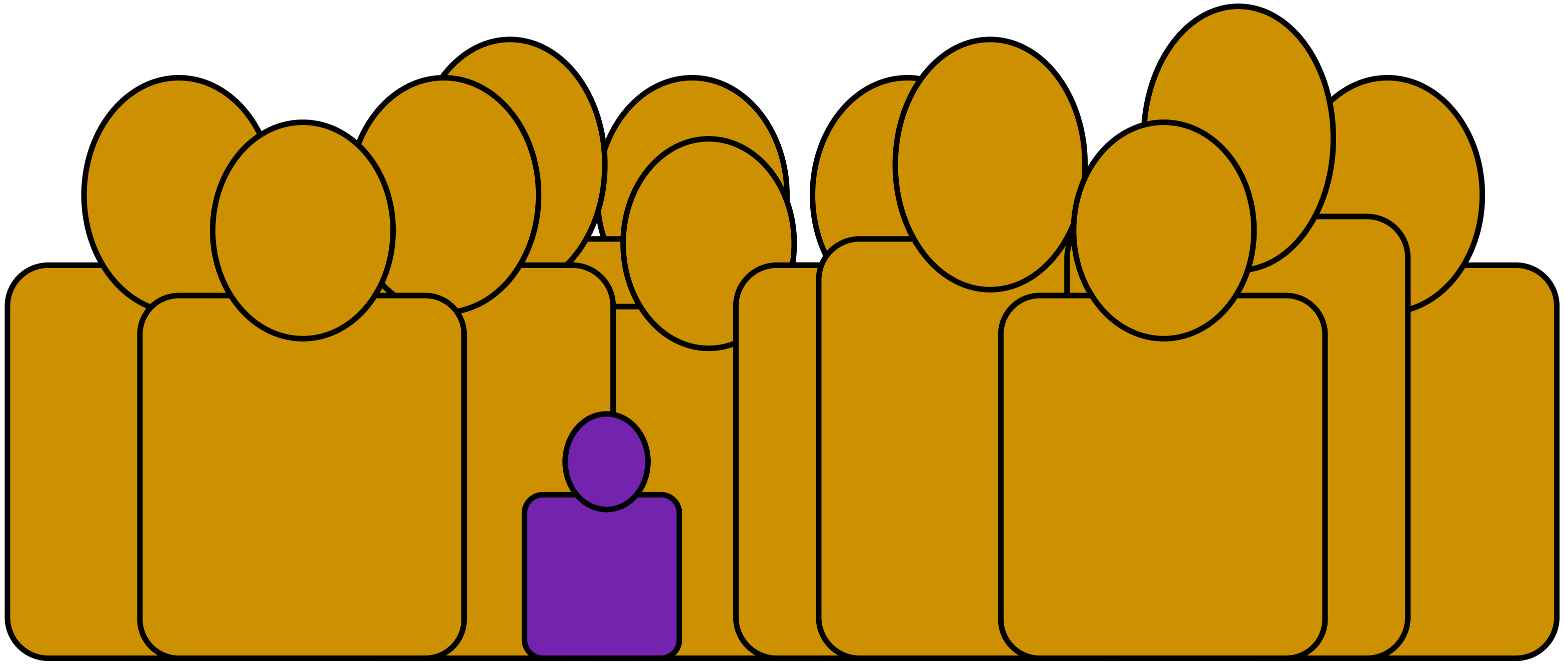
- Mary is very short
- What makes you say that she is short?

Mary's peers



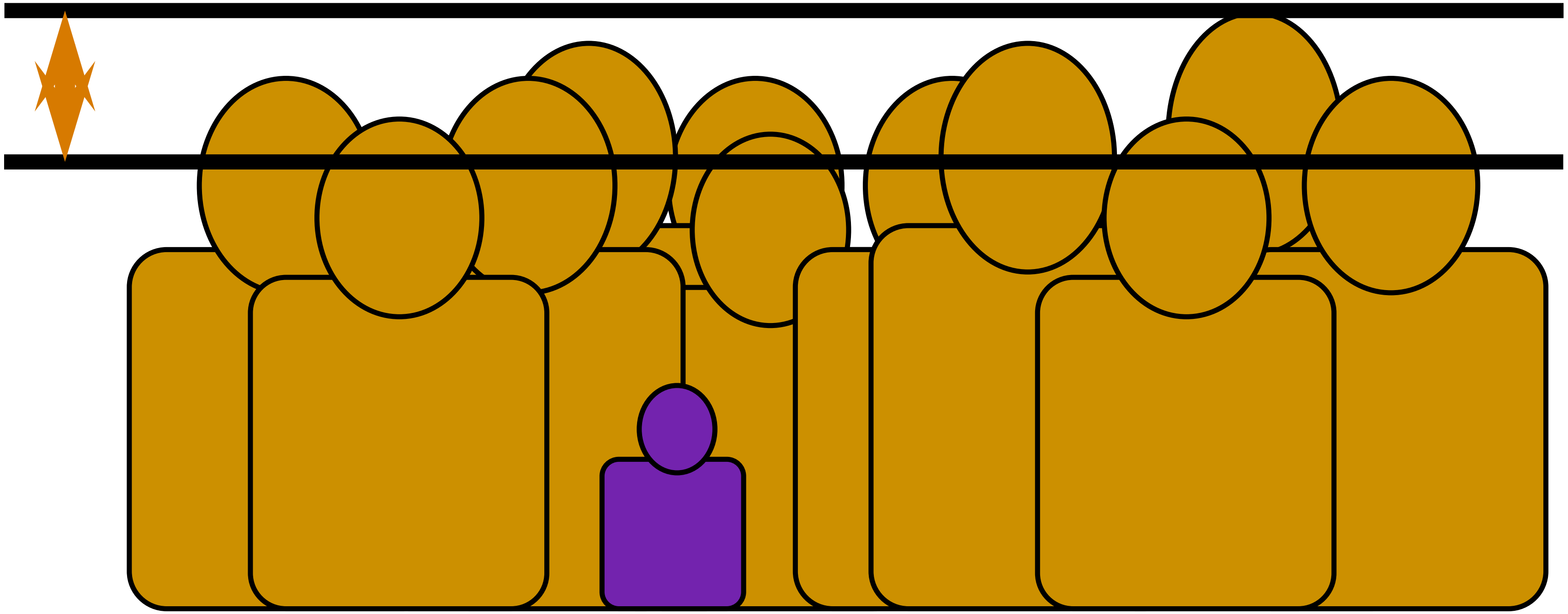
- You think Mary should be like everybody else.

Mary’s peers



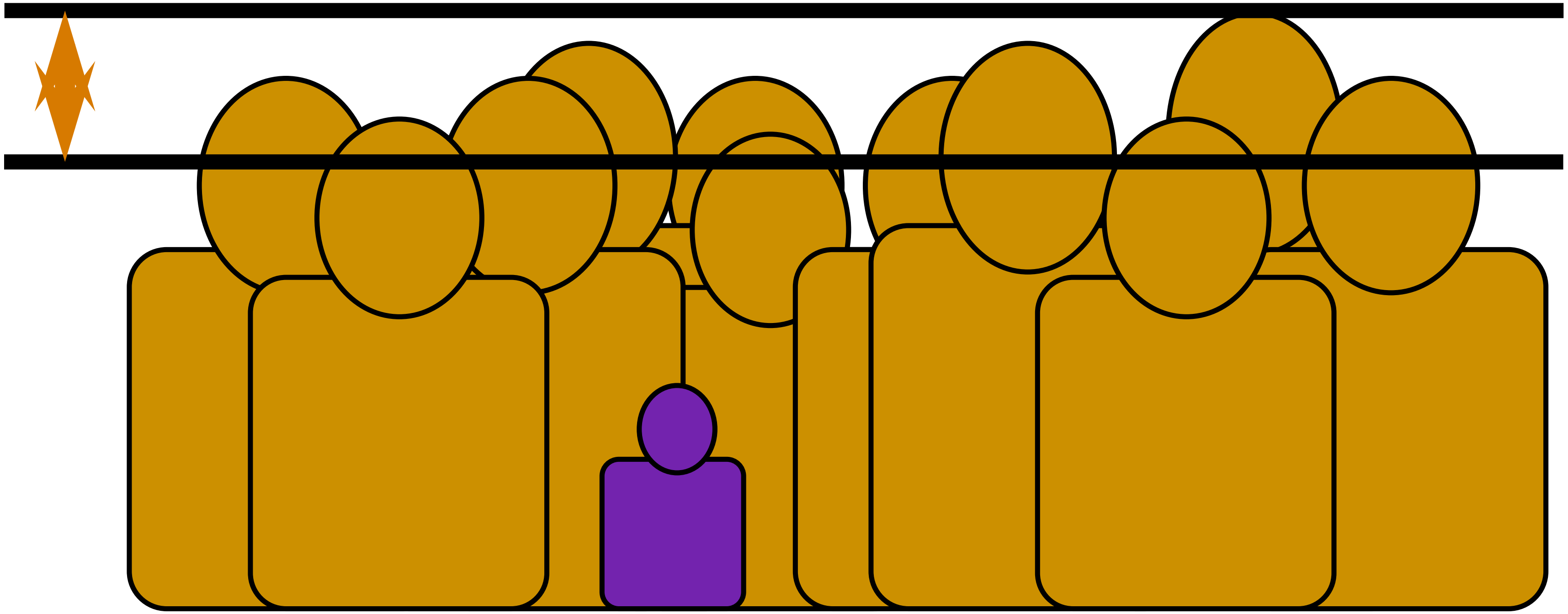
- You compare Mary with everybody else
- Is Mary’s height and ”everybody else’s” height all you know in order to make a conclusion?

Mary's peers



- NO, you also have an idea in your mind that Mary's height is farther away from everybody else than random variation

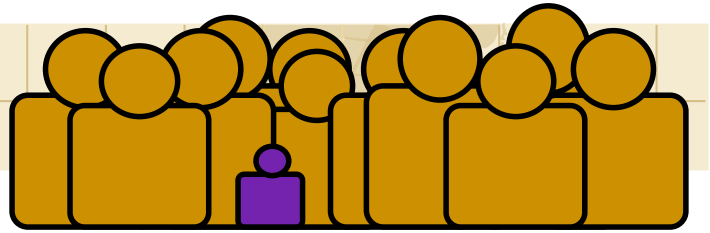
Mary's peers



- You make an estimate that Mary is more different than random variation should account for

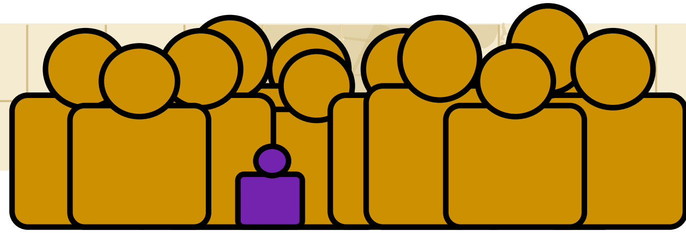
THIS IS THE IMPORTANT EYE-BALL TEST

Formula:



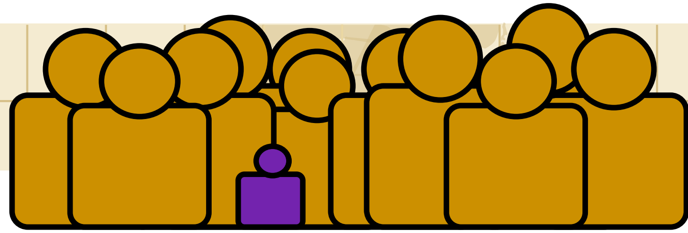
Mary's height

Formula:



Mary's height - Everybody else's height

Formula:



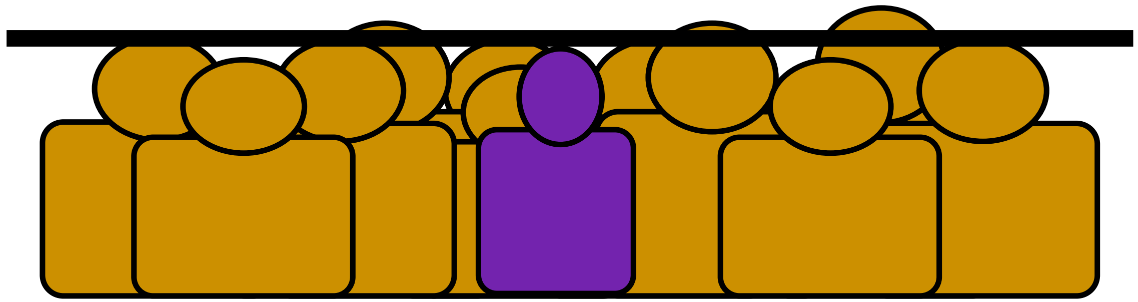
$$\frac{\text{Mary's height} - \text{Everybody else's height}}{\text{Random variation}}$$

Formula:

$$\frac{\text{Mary's height - Everybody else's height}}{\text{Random variation}}$$

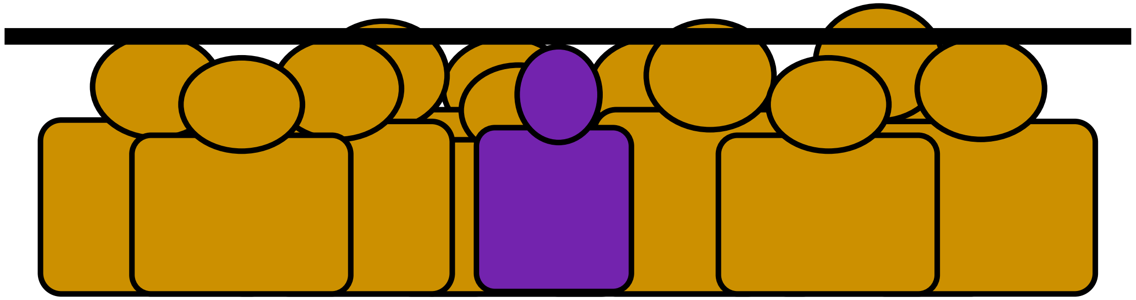
- A high figure leads you to think that Mary can not be like everybody else
- You start looking for a specific «reason» for her low height, maybe a medical condition.

Girls with Turner's syndrome



- Mary is not like everybody else. but she is like everybody else if you only compare with other girls with Turner's syndrome

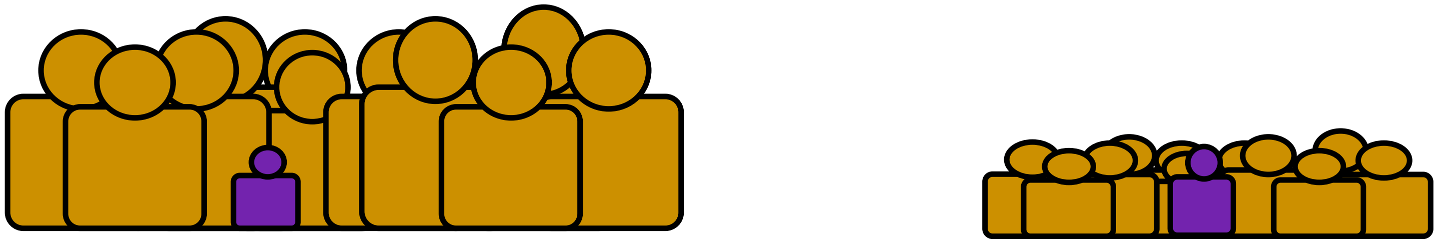
Girls with Turner's syndrome



- Mary is not like everybody else, but she is like everybody else if you only compare with other girls with Turner's syndrome

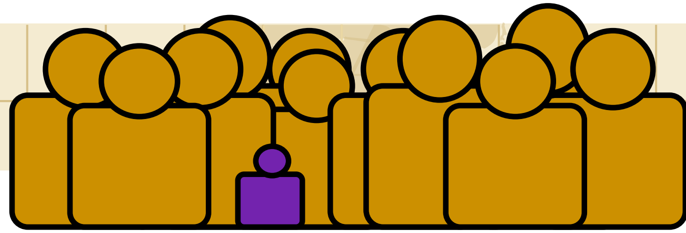
$$\frac{\text{Mary's height- Everybody else's height}_{\text{for girls with Turner's syndrome}}}{\text{Random variation}_{\text{for girls with Turner's syndrome}}}$$

Important!!!



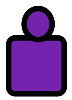
- THIS IS BASICALLY ALL THERE IS TO STATISTICAL TESTING
- REMEMBER, you do it every day just by «eye-balling»

Formula:

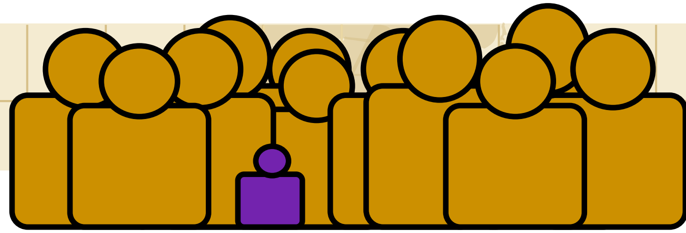


$$\frac{\boxed{\text{Mary's height}} - \text{Everybody else's height}}{\text{Random variation}}$$

- This measurement carries few difficulties

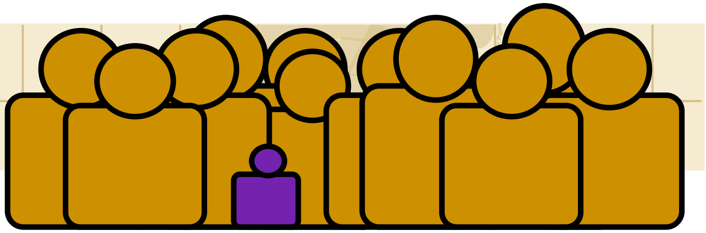


Formula:



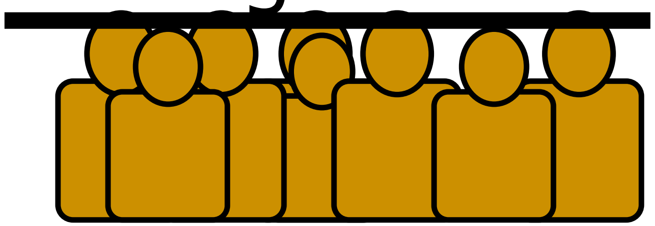
$$\frac{\boxed{135\text{ cm}}\ -\ \text{Everybody else's height}}{\text{Random variation}}$$

Formula:

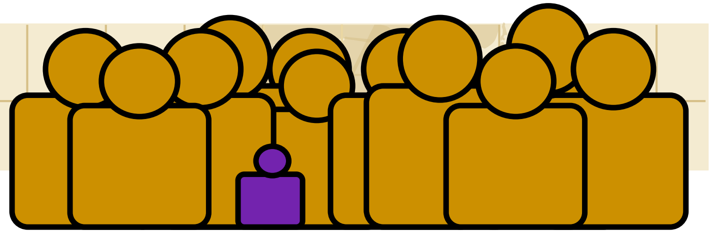


$$\frac{135\text{ cm} \quad - \boxed{\text{Everybody else's height}}}{\text{Random variation}}$$

- What is everybody else's height?

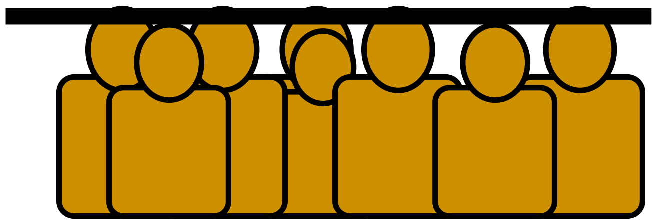


Formula:

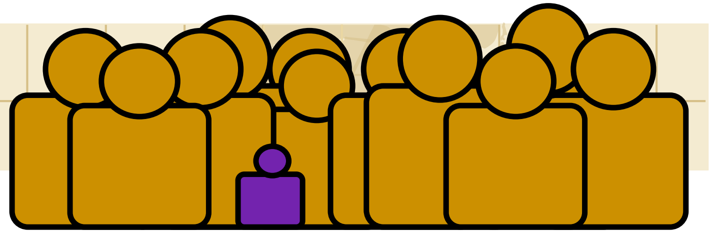


Measures of centrality:

- Mode
  - The value occurring most often
- Median
  - The middle value when all values are ranked
- Mean
  - Sum of all values divided by number of values
  - $\Sigma x/N = \mu$



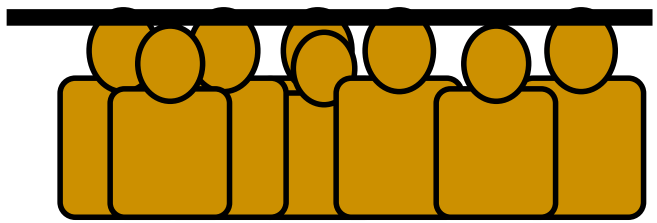
Formula:



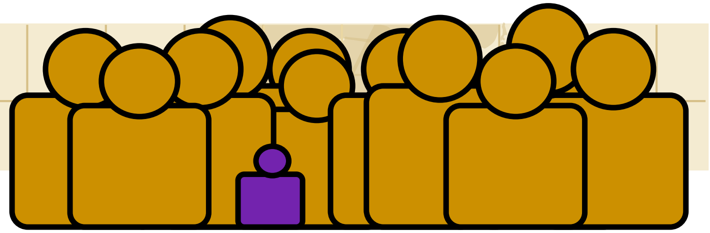
Measures of centrality:

- If Mary's peers are a statistical "population" of 10 twelve-year olds :

$$\mu = \frac{147 + 152 + 155 + 156 + 151 + 153 + 151 + 159 + 162 + 154}{10} = \mathbf{154\ cm}$$



Formula:



$$\frac{135\ \text{cm} - \boxed{154\ \text{cm}}}{\text{Random variation}}$$

Formula:

$$\frac{135\text{ cm} - 154\text{ cm}}{\boxed{\text{Random variation}}}$$

- What is random variation?

Formula:

147
151
151
152
153
154
155
156
159
162

Measures of spread:

- Difference between highest and lowest value: 162 cm - 147 cm = 15 cm

This is called RANGE

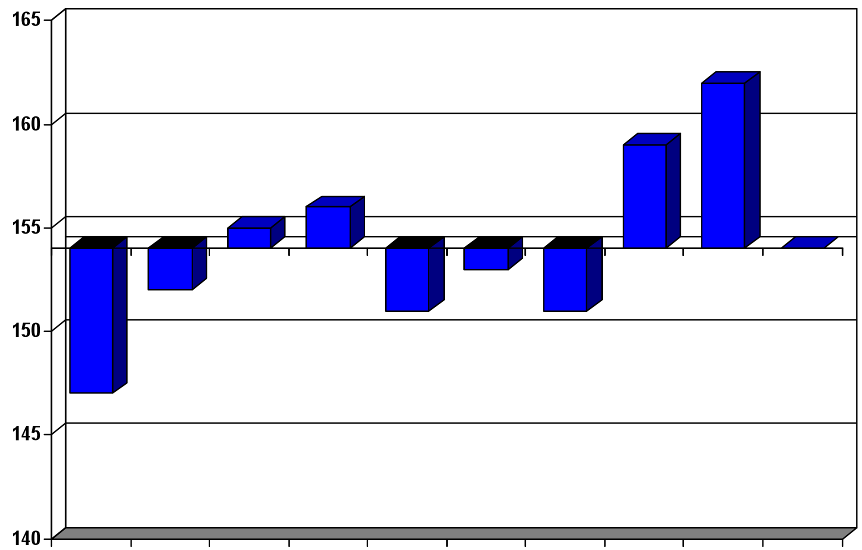
- Utilizes only a small part of the available information

Formula:

Measures of spread:

- Mean distance from the mean

$$\frac{\Sigma(x-\mu)}{N}$$



What is the value of this mathematical expression?

Formula:

Measures of spread:

- Mean distance from the mean:

$$\frac{\Sigma(x-\mu)}{N} = \frac{0}{10} = 0$$

x	x-μ
147	-7
152	-2
155	+1
156	+2
151	-3
153	-1
151	-3
159	+5
162	+8
154	0
Σ	0

Formula:

Measures of spread:

- Mean distance from the mean:

Is it possible to ignore the sign?

Formula:

Measures of spread:

- Absolute mean distance from the mean:

$$\frac{\Sigma\,|(x-\mu)|}{N} = \frac{32}{10} = 3{,}2$$

x	|x-μ|
147	7
152	2
155	1
156	2
151	3
153	1
151	3
159	5
162	8
154	0
Σ	32

Formula:

Measures of spread:

- Mean distance from the mean:

Is there another way of getting rid of the sign?

Formula:

Measures of spread:

- Mean squared distance from the mean:

$$\frac{\Sigma(x-\mu)^2}{N} = \frac{166}{10} = 16.6$$

x	$(x-\mu)$	$(x-\mu)^2$
147	-7	49
152	-2	4
155	+1	1
156	+2	4
151	-3	9
153	-1	1
151	-3	9
159	+5	25
162	+8	64
154	0	0
Σ	0	166

This is called:

VARIANCE

Formula:

Measures of spread:

- Variance is in this case expressed in cm^2
- This is not very practical when the mean height is expressed as cm

Measures of spread:

- Square root of the mean squared distance from the mean:

$$\frac{\Sigma(x-\mu)^2}{N} = 16.6$$

$$\sqrt{\frac{\Sigma(x-\mu)^2}{N}} = 4.1$$

x	(x-μ)	$(x-\mu)^2$
147	-7	49
152	-2	4
155	+1	1
156	+2	4
151	-3	9
153	-1	1
151	-3	9
159	+5	25
162	+8	64
154	0	0
Σ	0	166

This is called:

STANDARD DEVIATION

Formula:

Measures of spread:

$$\sqrt{\frac{\Sigma(x-\mu)^2}{N}}$$

The <u>STANDARD DEVIATION</u>

carries the symbol σ.

This is the measure of spread in a population

Formula:

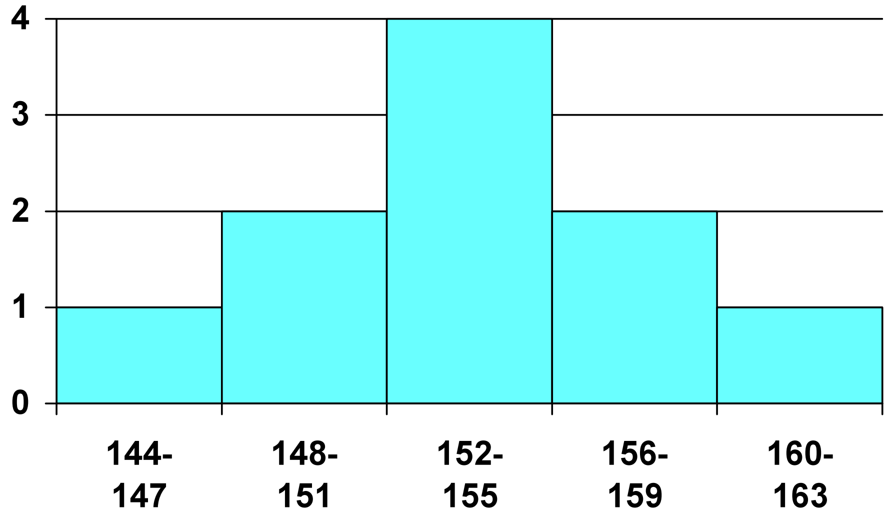
$$\frac{135\text{ cm} - 154\text{ cm}}{4.1\text{ cm}} = 4.63$$

SO WHAT??

- Mary's height is four and a half times further away from the mean than the standard deviation (random variation)
- This value is called z

Normal curve:

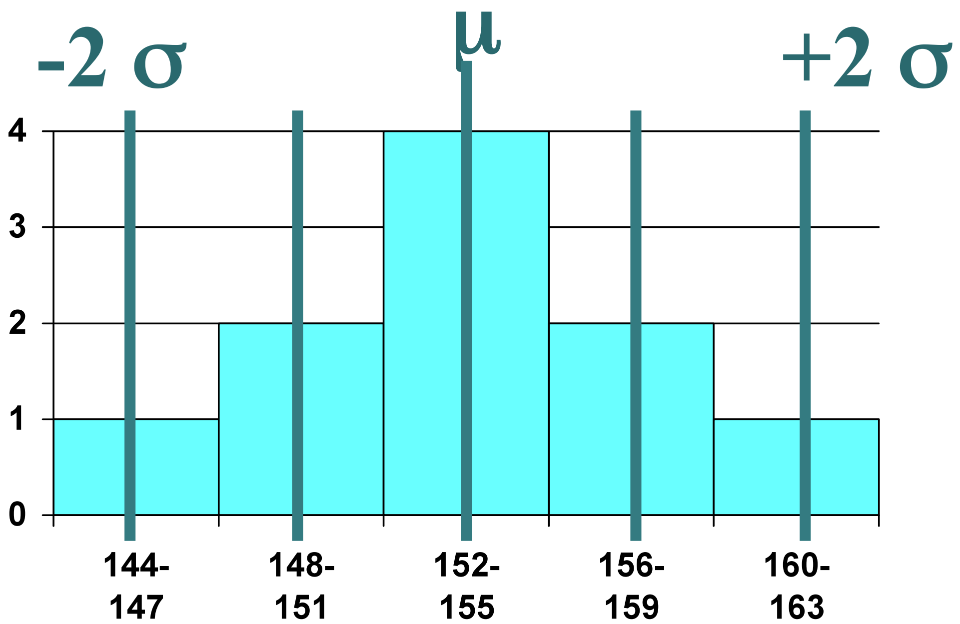
Histogram of the height of Mary's peers



Most of the girls fall in the middle, a few further out.

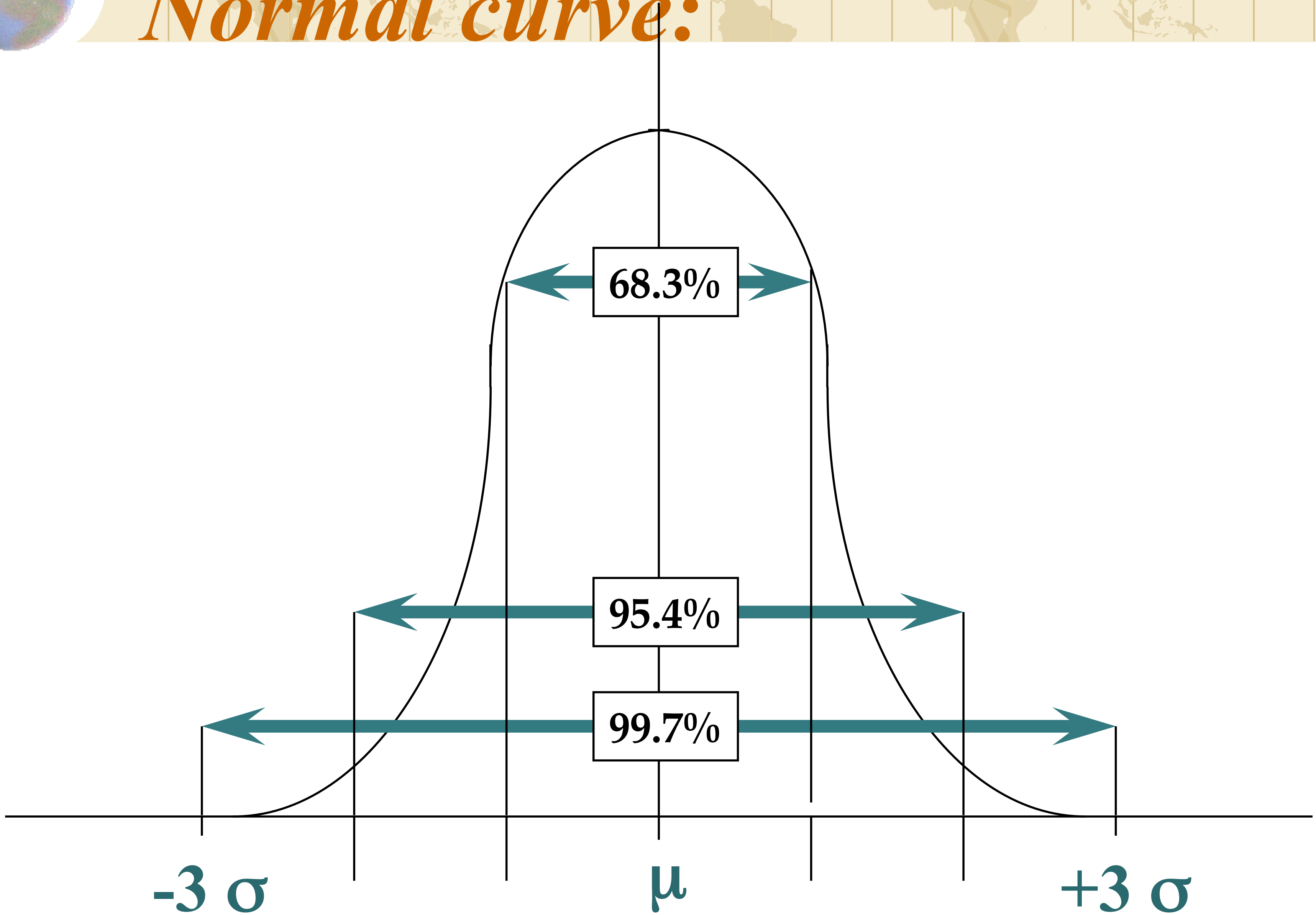
Normal curve:

Histogram of the height of Mary's peers



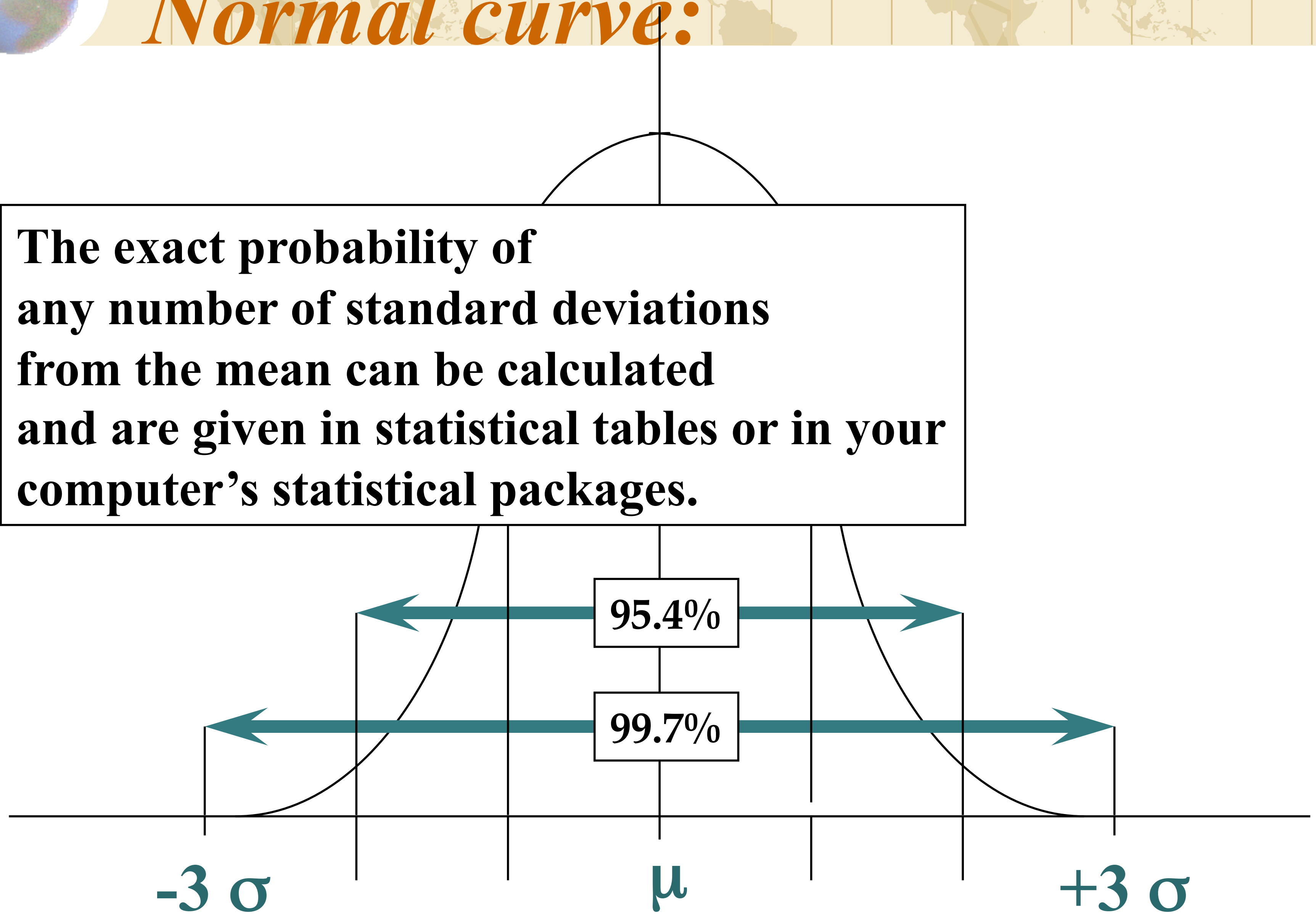
Most of the girls fall in the middle, a few further out.

Normal curve:



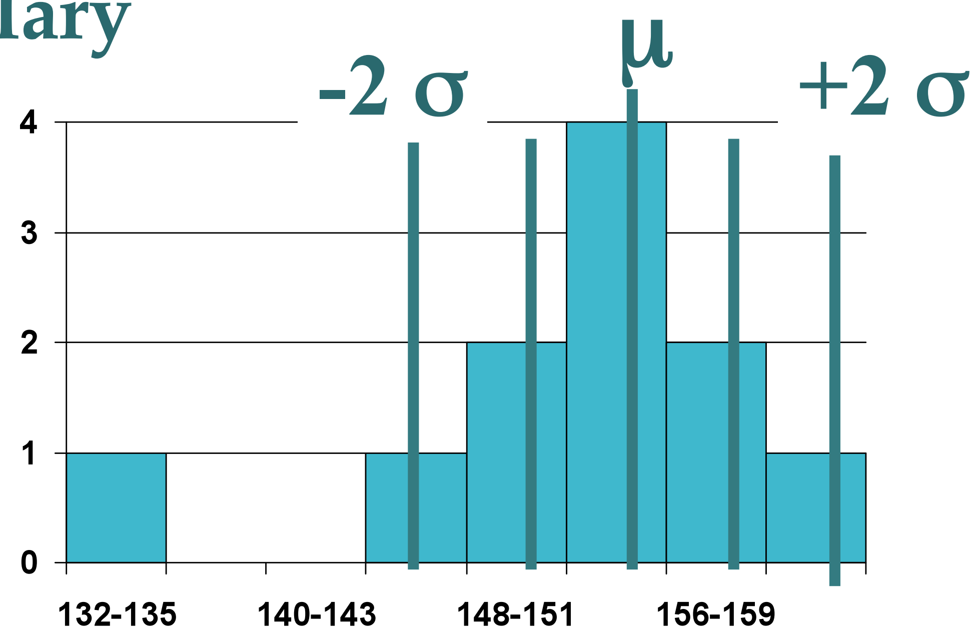
Normal curve:

The exact probability of
any number of standard deviations
from the mean can be calculated
and are given in statistical tables or in your
computer's statistical packages.



Normal curve:

Histogram of the height of all Mary's peers **and Mary**



Most of the girls fall in the middle, a few further out. Mary is way out there: **The statistics agree with our previous «eye-balling»**

Was that all?

Yes, the principles of statistical testing are not more than this.