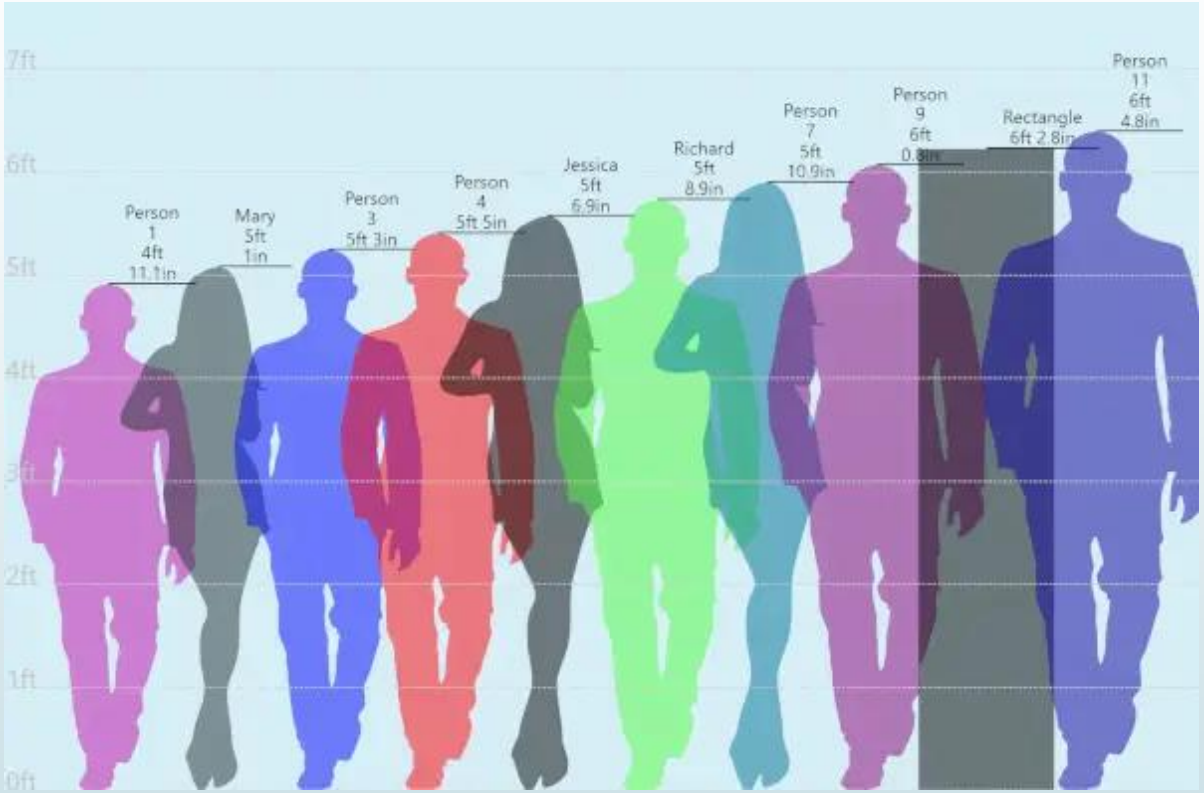
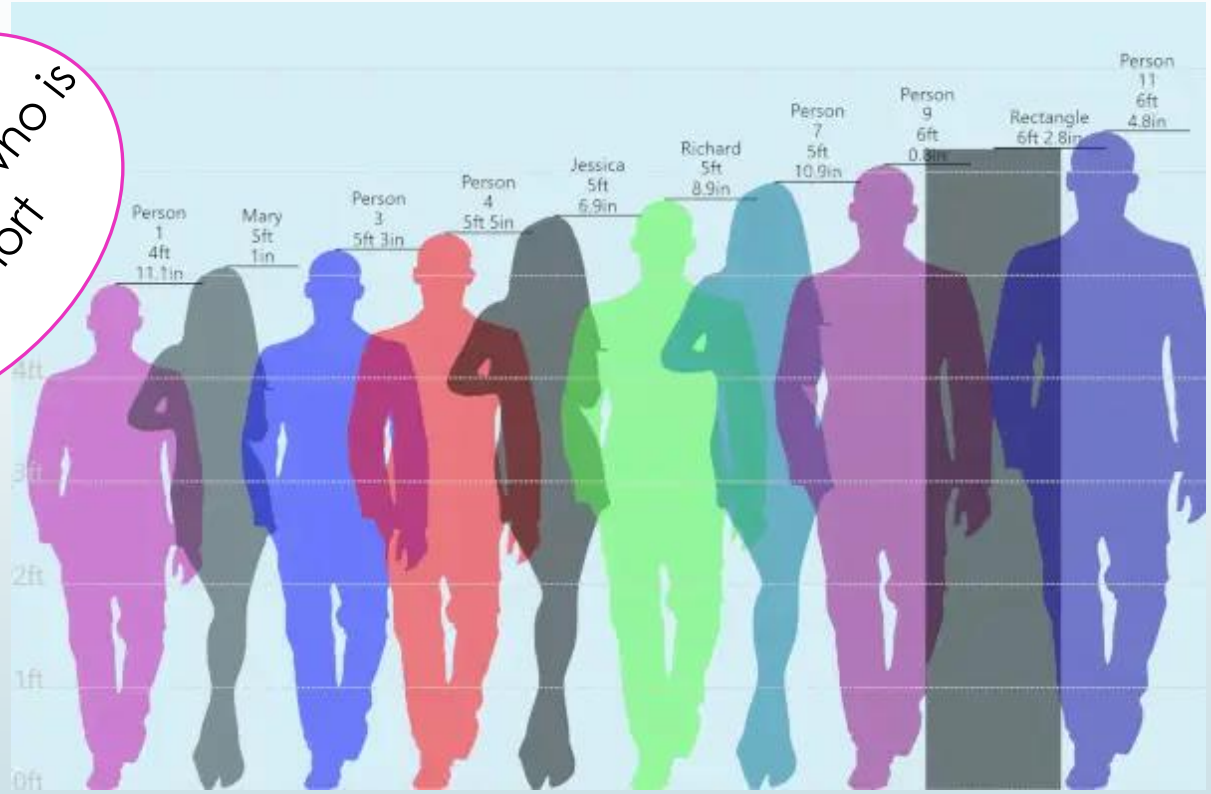
Normal Distribution

Height Comparison

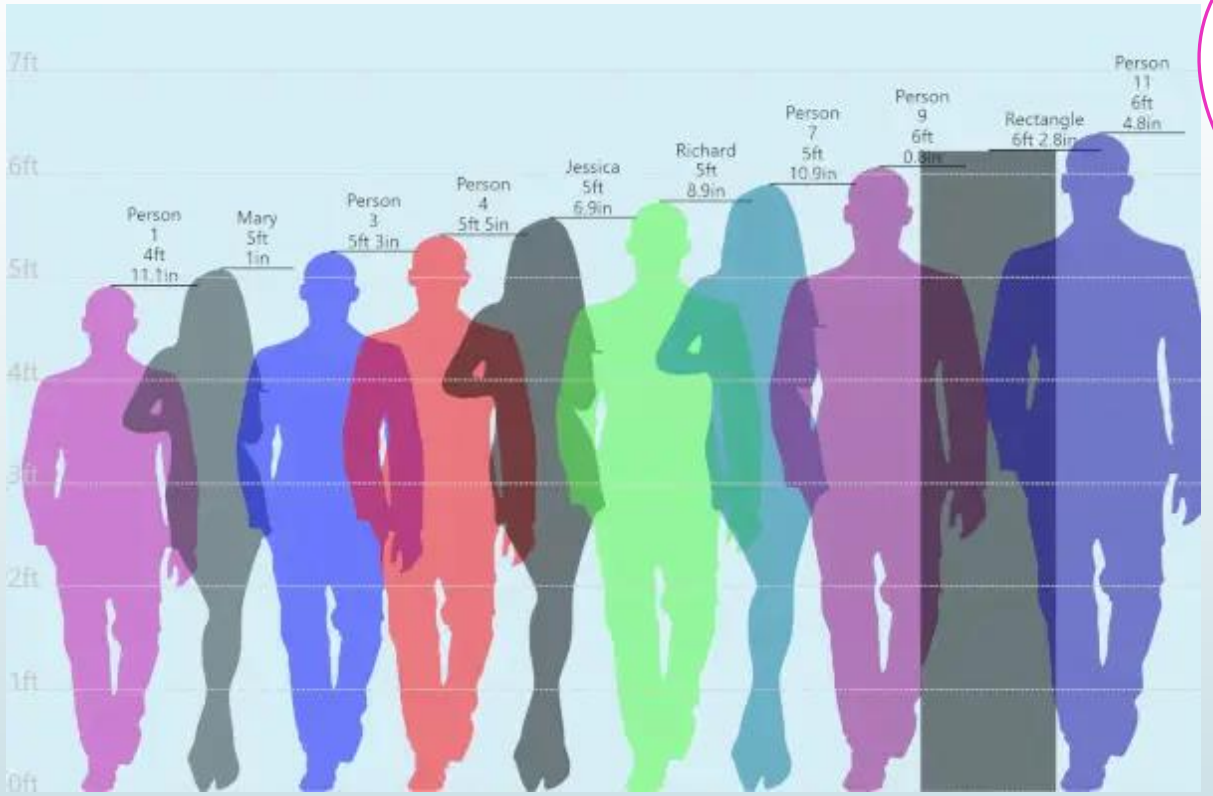


Height Comparison

It is relatively rare to see someone who is super short

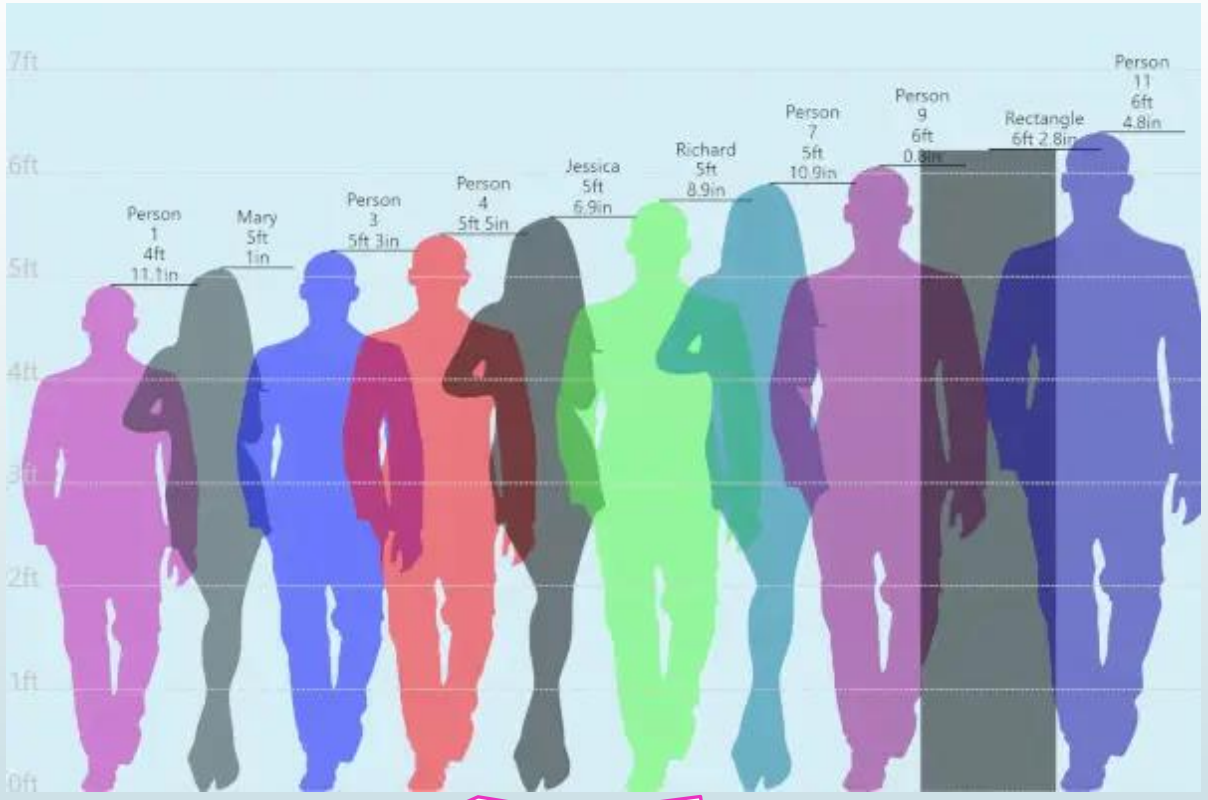


Height Comparison



And it is relatively rare to see someone who is super tall also..

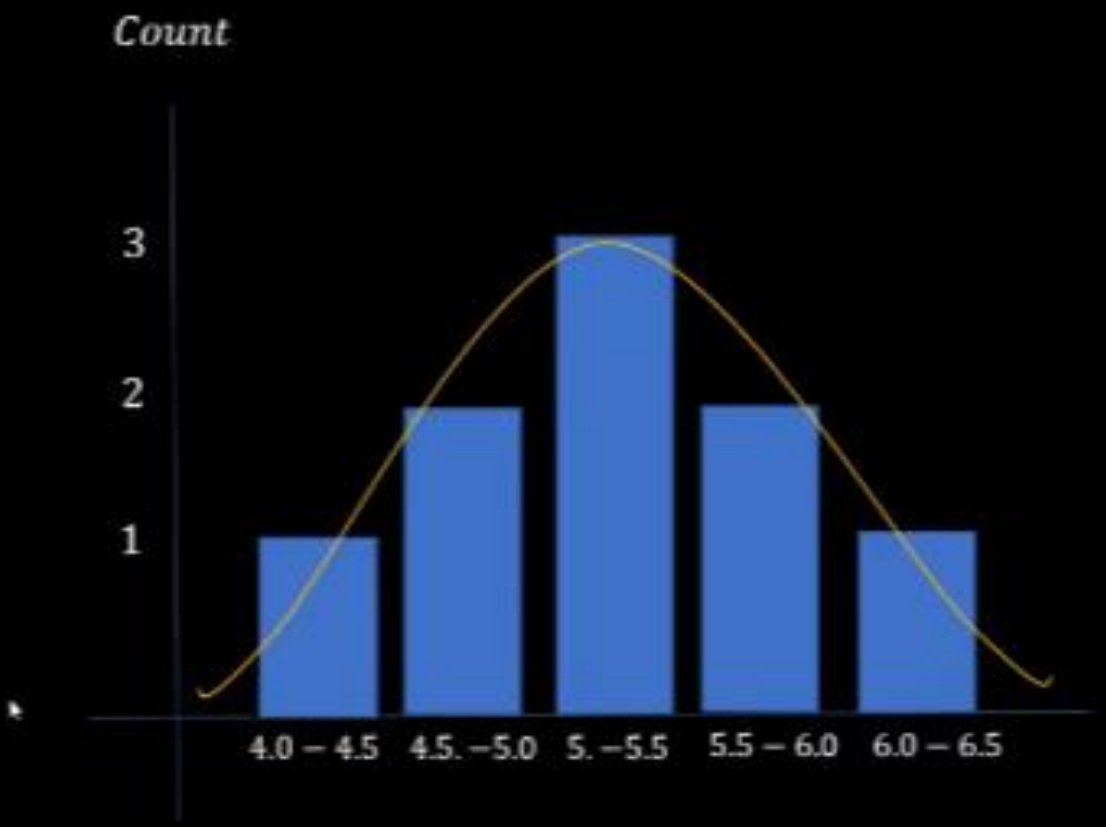
Height Comparison



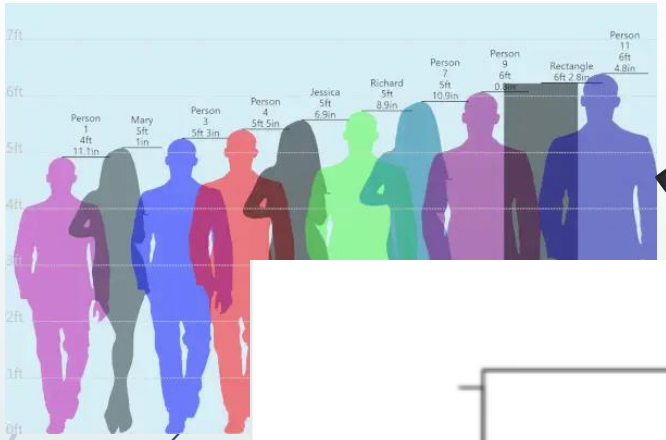
But, it is common to see someone with average heights…

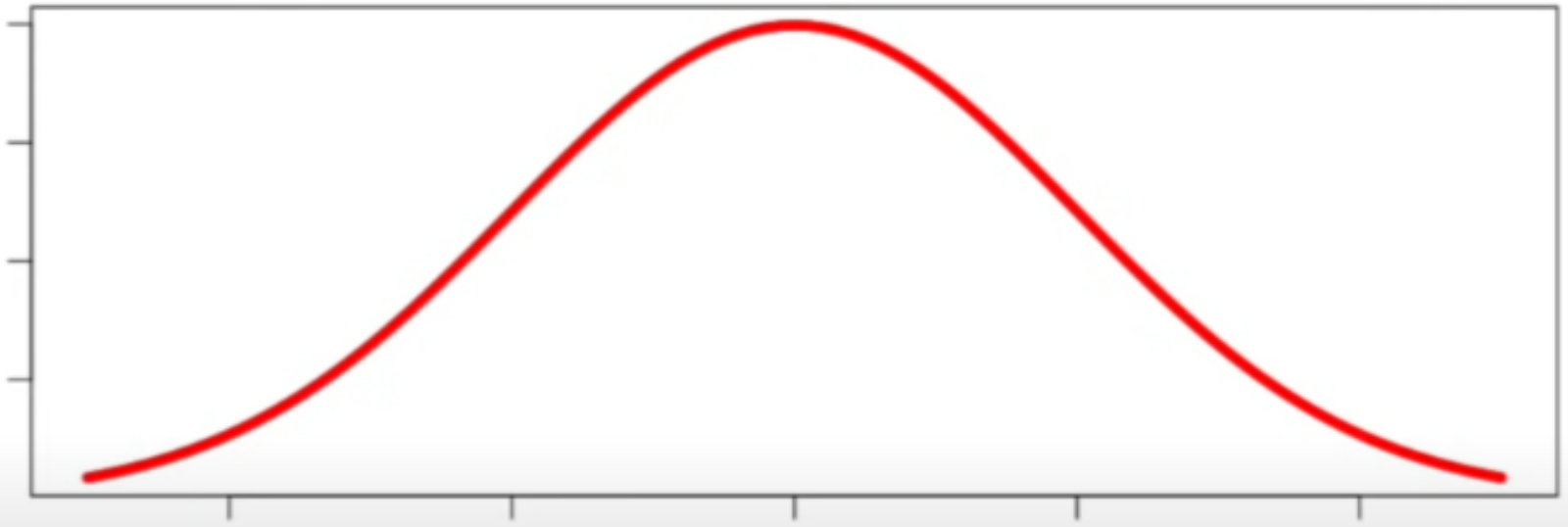
Name	Height (ft)
Rob	6.2
Thomas	5.7
Nina	4.6
Mittal	5.4
Sofia	5.9
Mohan	4.3
Tao	5.1
Deepika	5.2
Rafiq	4.9

Name	Height (ft)
Rob	6.2
Thomas	5.7
Nina	4.6
Mittal	5.4
Sofia	5.9
Mohan	4.3
Tao	5.1
Deepika	5.2
Rafiq	4.9

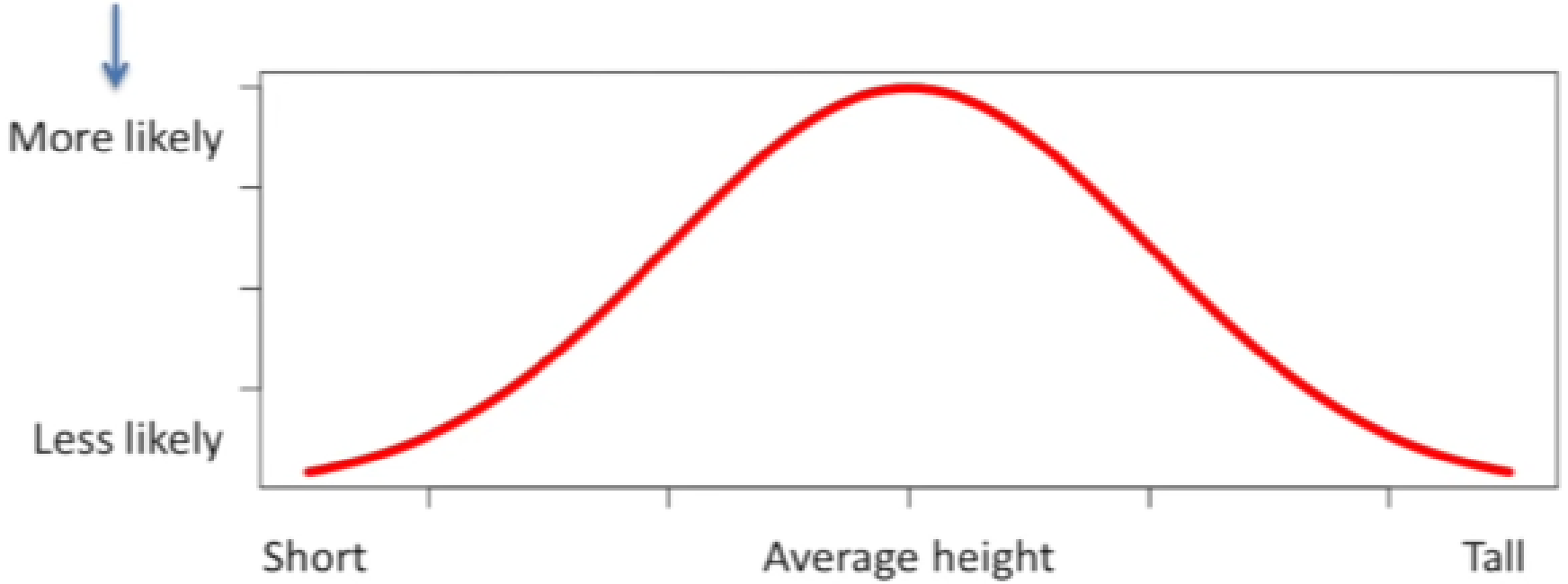


Height Comparison





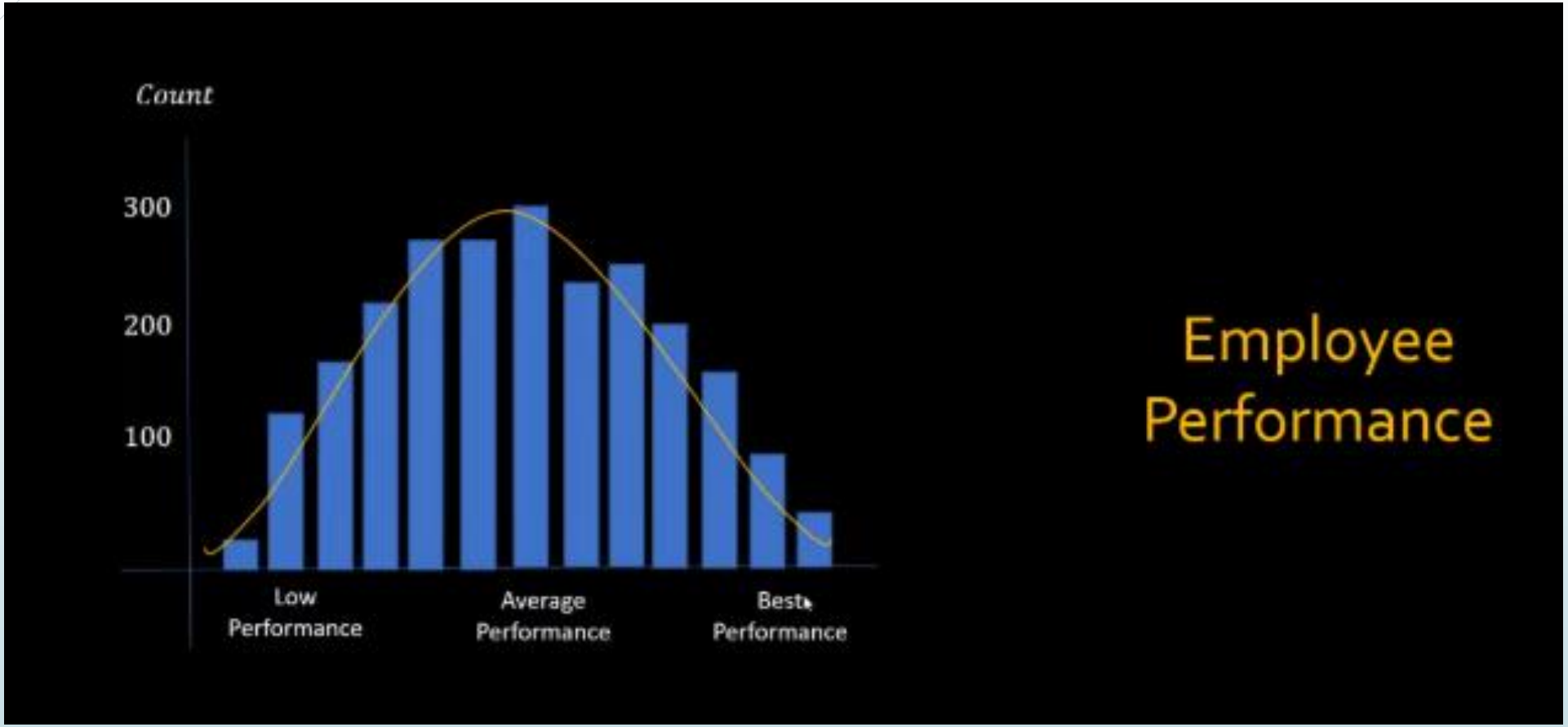
The y-axis represents the relative probability of observing someone who is really short, or really tall, or who has an average height.

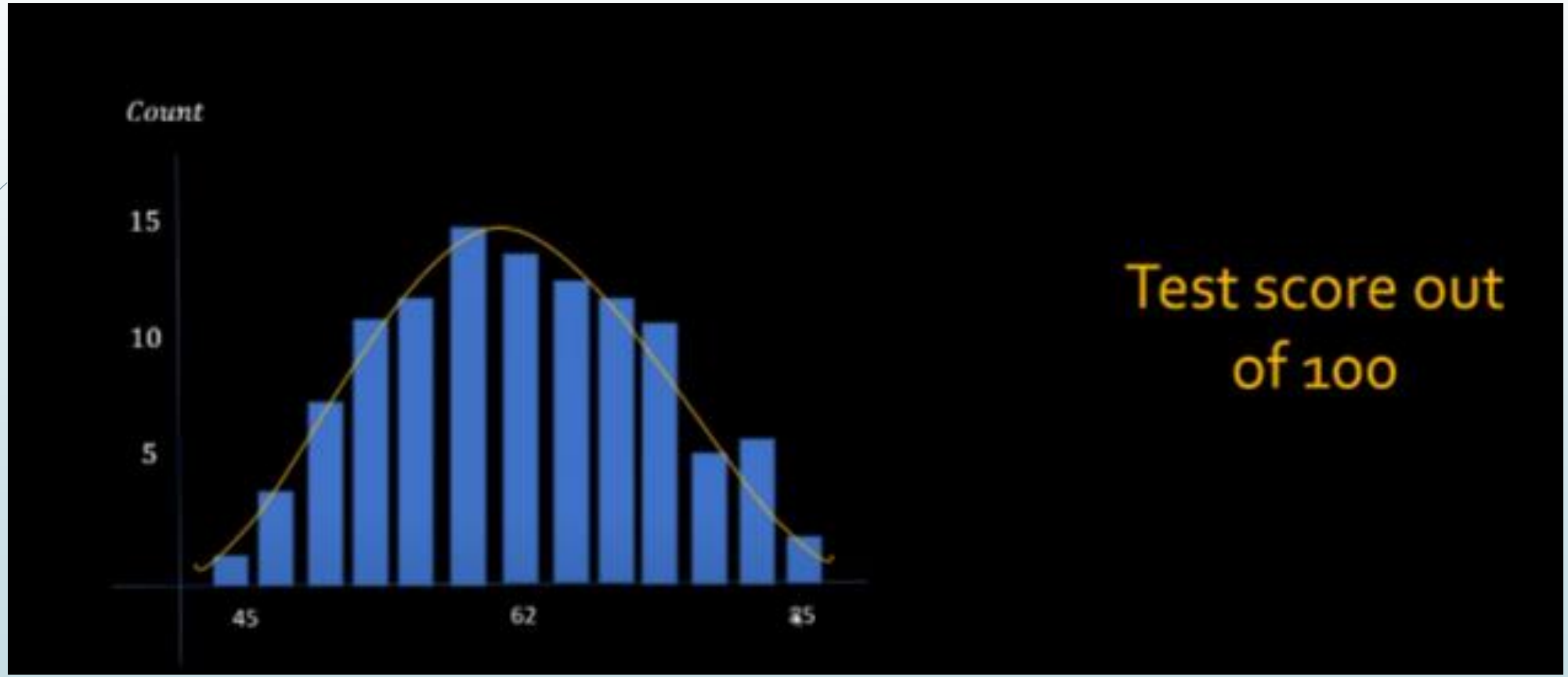


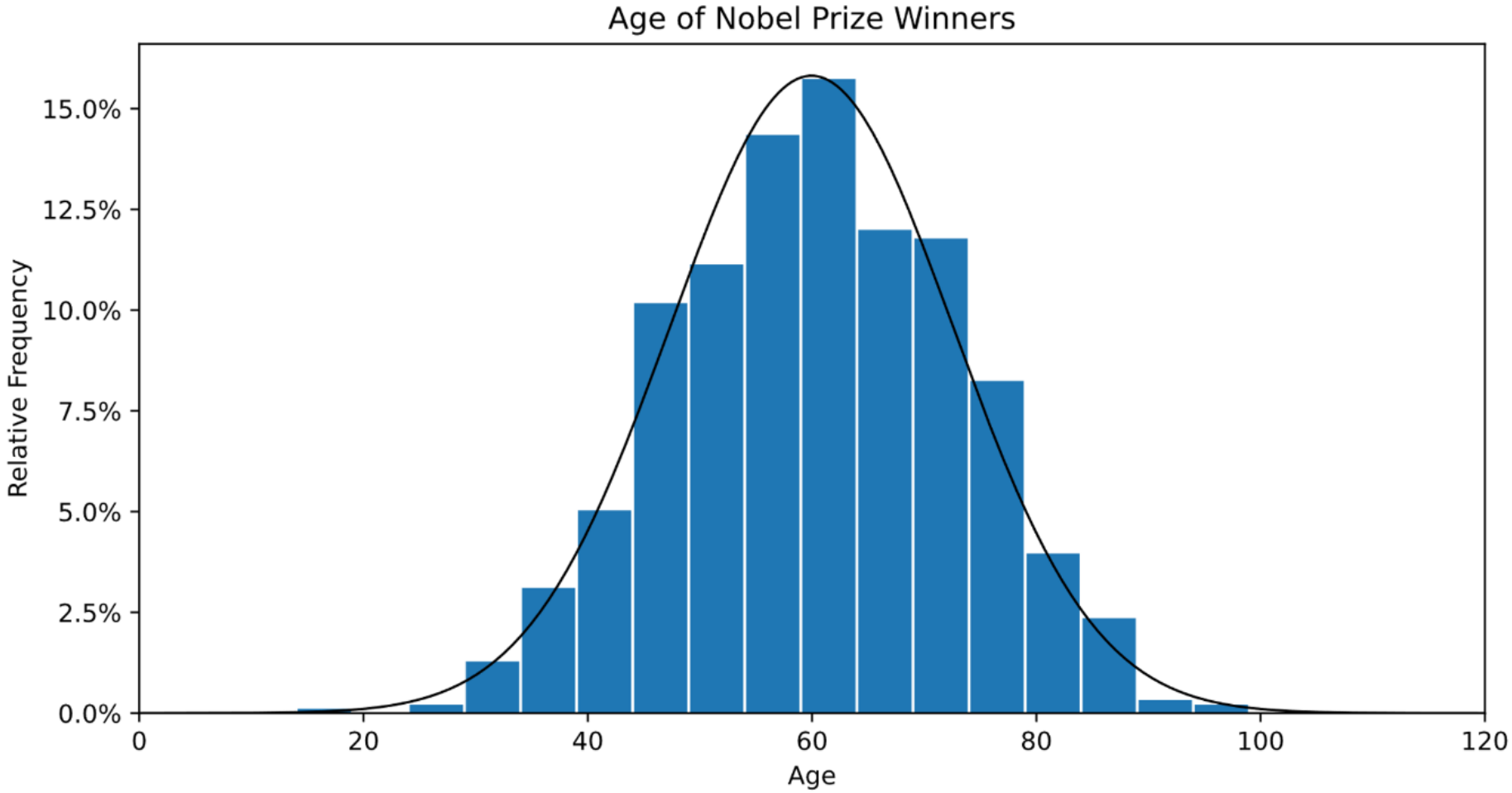
Normal distribution

A normal distribution is a type of continuous probability distribution in which most data points cluster toward the middle of the range, while the rest taper off symmetrically toward either extreme.

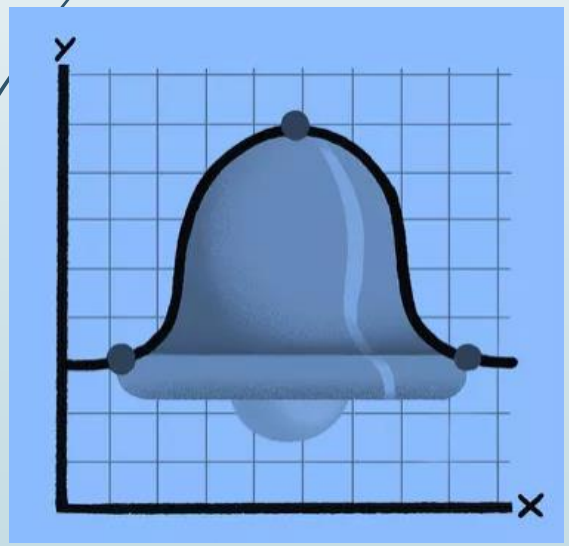
Examples:

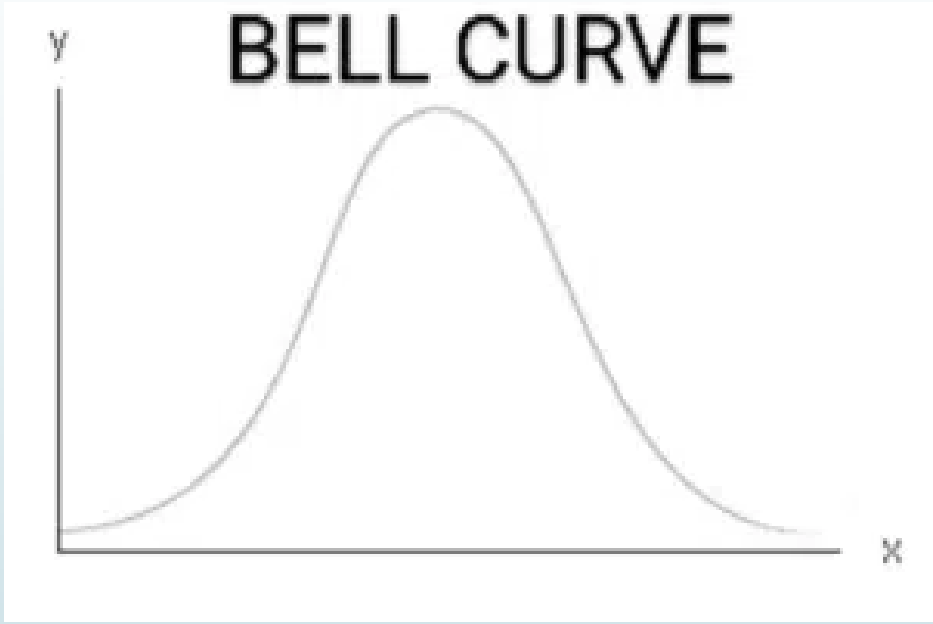

Count
300
200
100
Low Performance
Average Performance
Best Performance
Employee Performance


Count
15
10
5
45
62
85
Test score out of 100

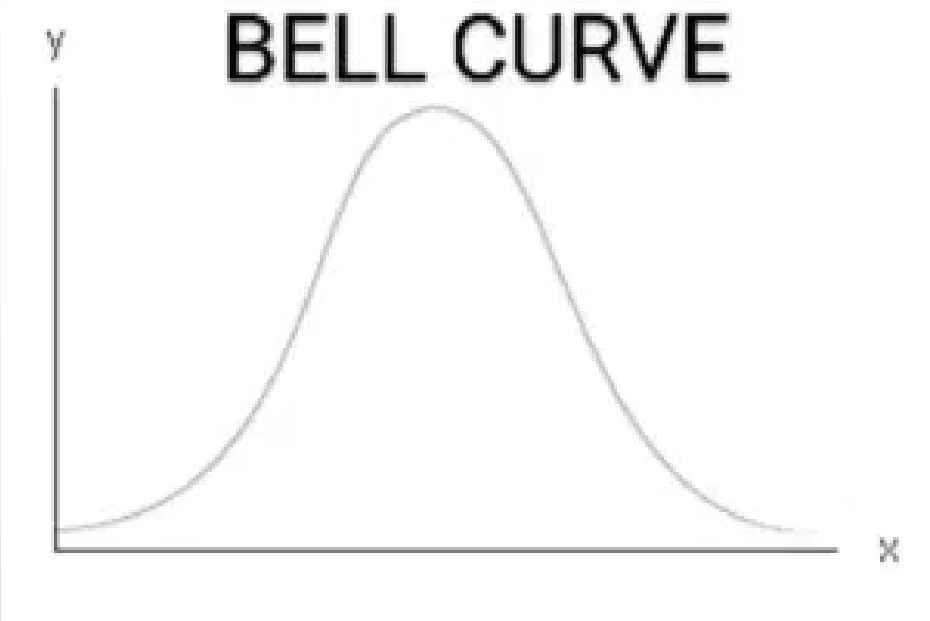

Age of Nobel Prize Winners
Relative Frequency
15.0%
12.5%
10.0%
7.5%
5.0%
2.5%
0.0%
0
20
40
60
80
100
120
Age

What Is a Bell Curve?



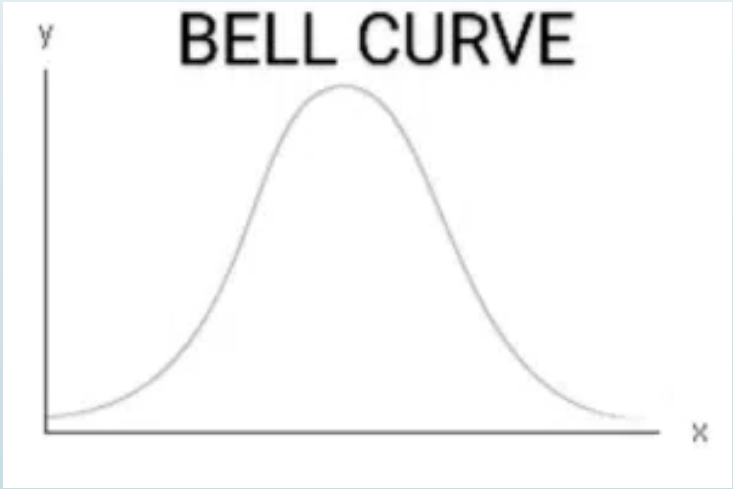


What Is a Bell Curve?



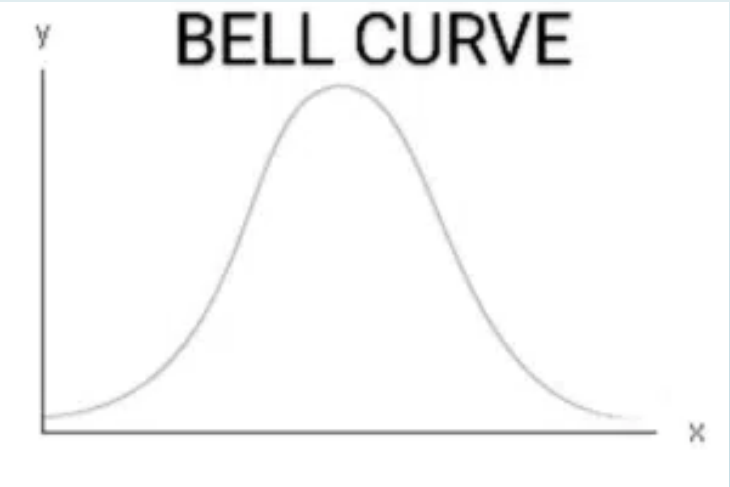
The normal distribution is often referred to as a '**bell curve**' because of it's shape.

A bell curve is the graph that depicts a normal probability distribution.



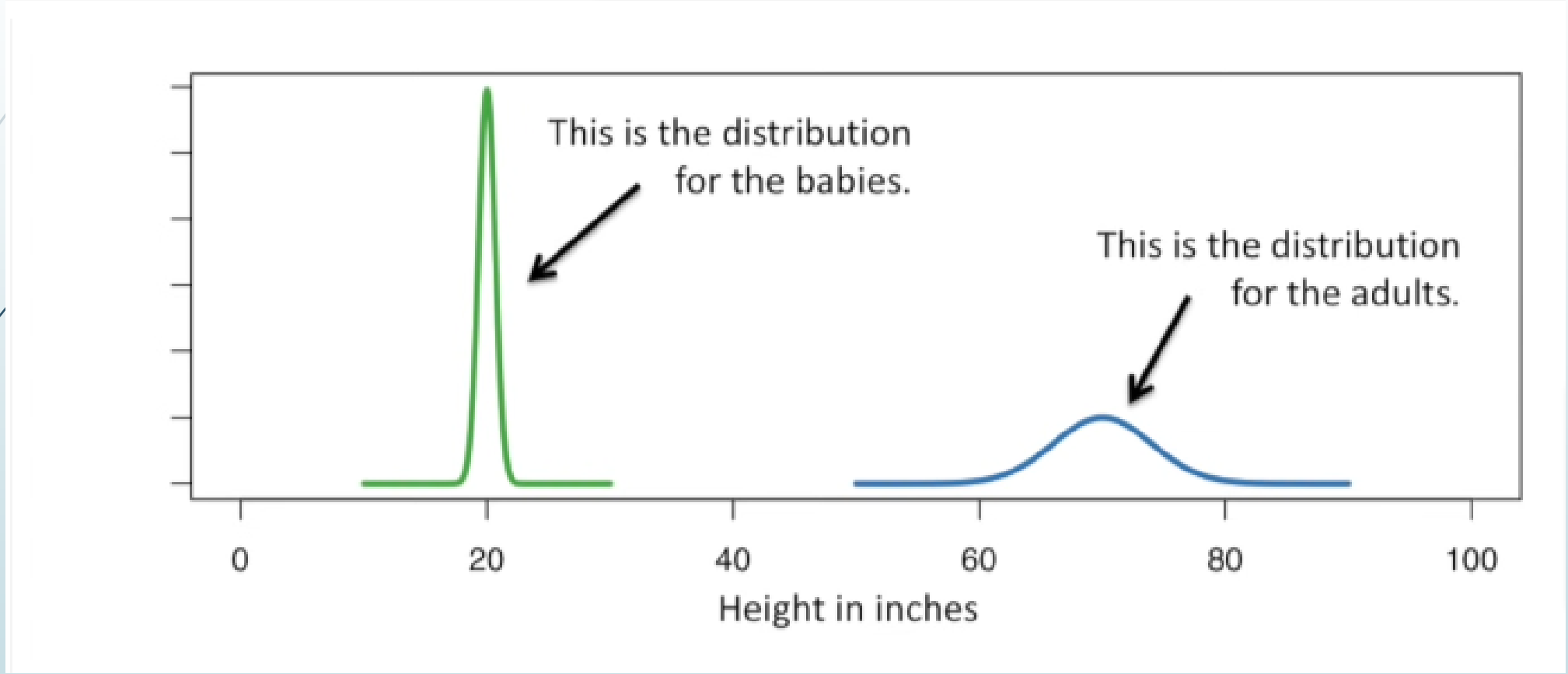
A bell curve is the graph that depicts a normal probability distribution.

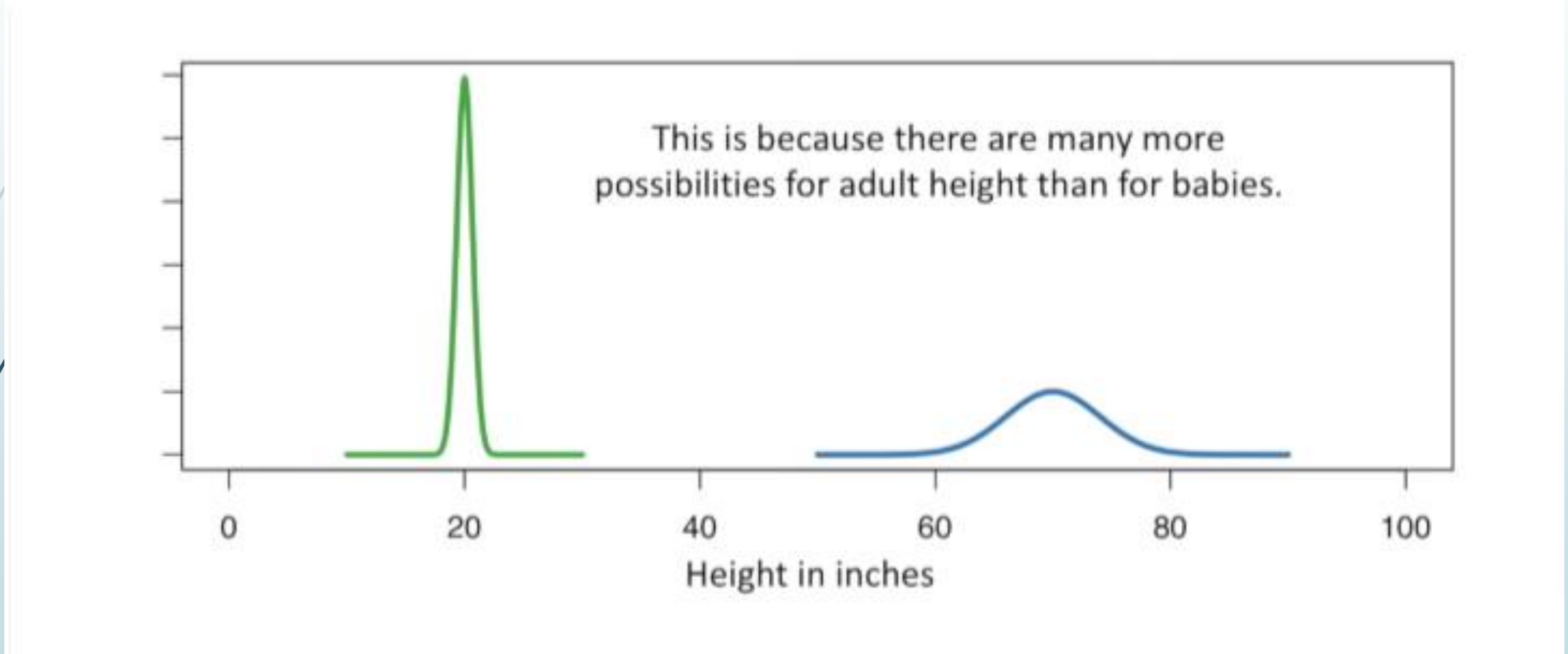
- The bell curve is perfectly symmetrical.
- It is concentrated around the peak and decreases on either side.
- In a bell curve, the peak represents the most probable event in the dataset while the other events are equally distributed around the peak.

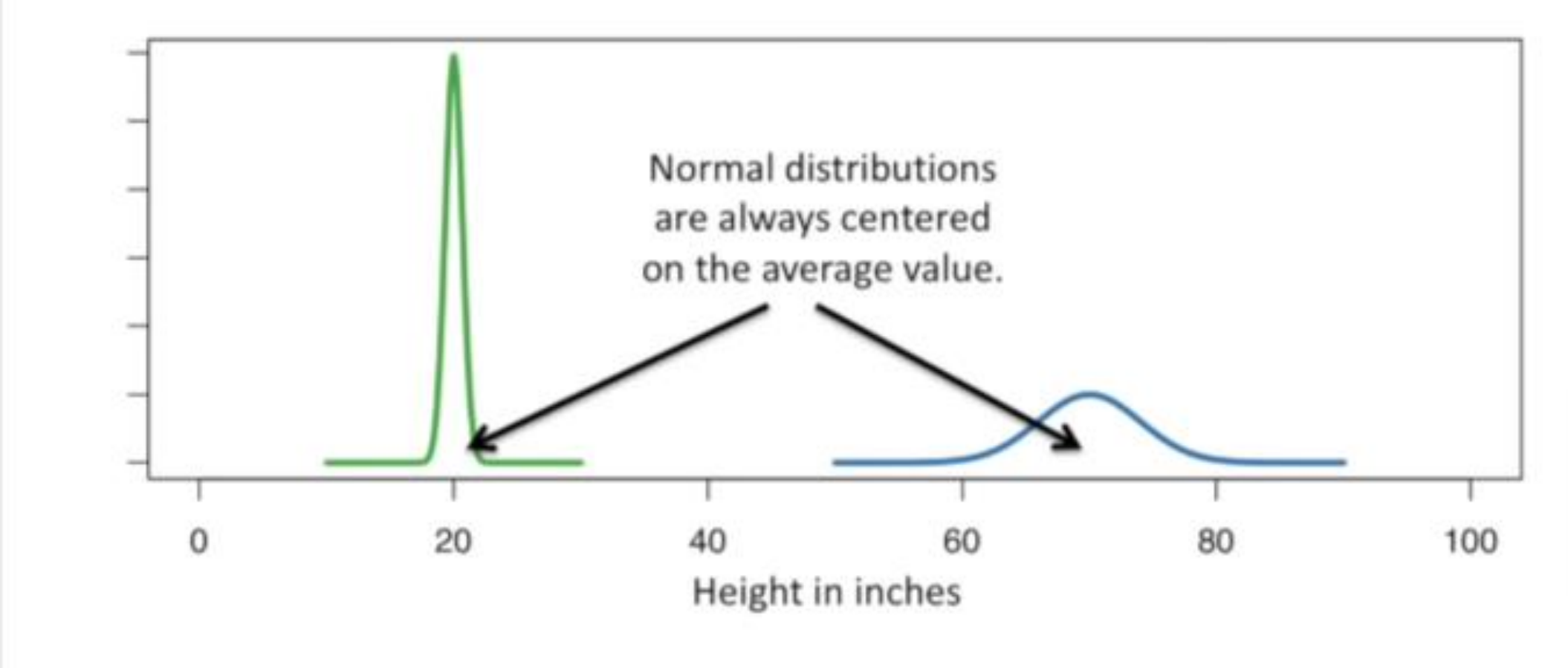


Central Tendency of Data

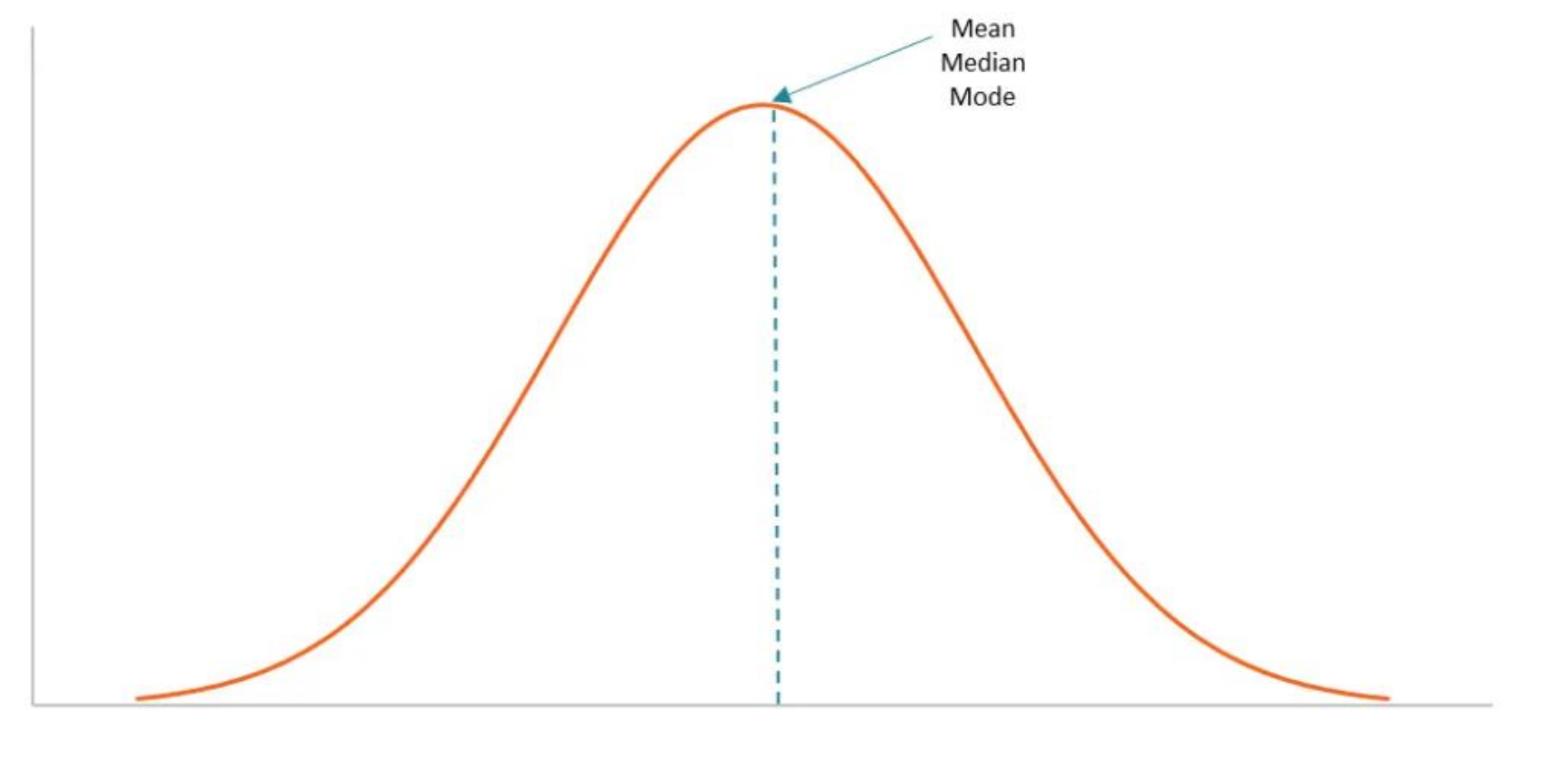
Comparing two Distributions:




This is because there are many more possibilities for adult height than for babies.
0
20
40
60
80
100
Height in inches

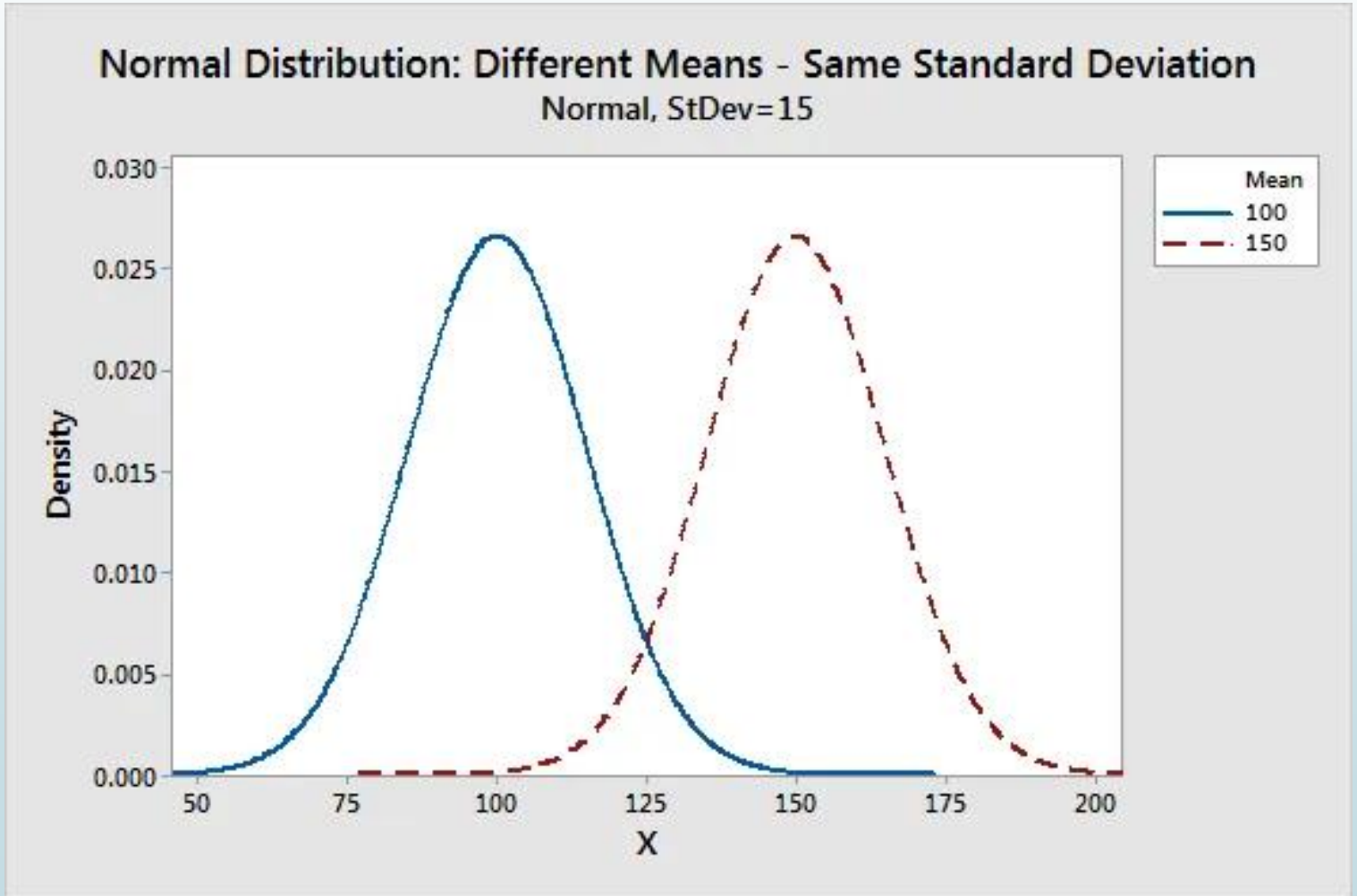

Normal distributions
are always centered
on the average value.
0
20
40
60
80
100
Height in inches

Mean, Median, and Mode: In a normal distribution, the mean, median, and mode are all equal and located at the centre of the distribution.



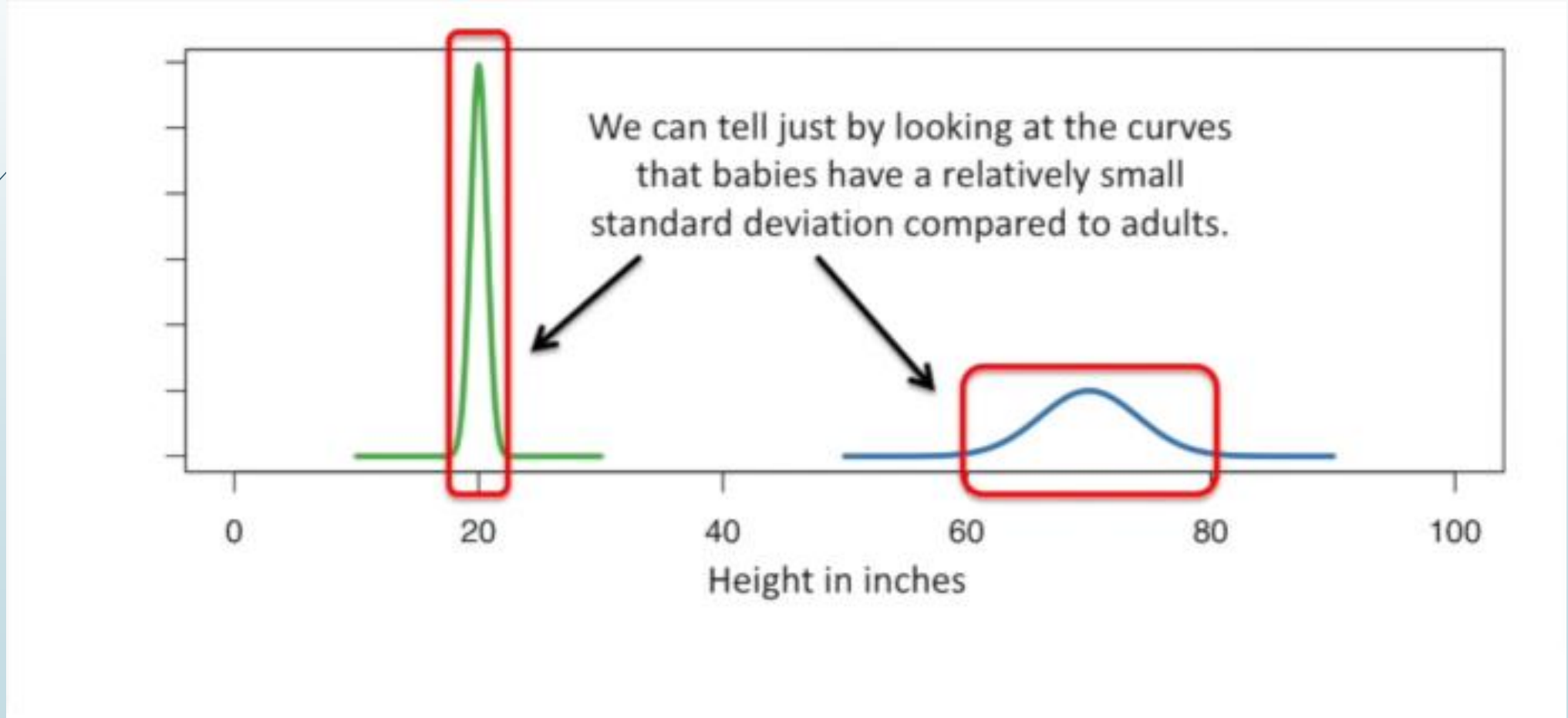
- **Symmetry:** The normal distribution is symmetric around its mean. This means the left side of the distribution mirrors the right side.

If we change the value of the mean, then the curve of normal distribution moves either to the left or right along the X-axis.

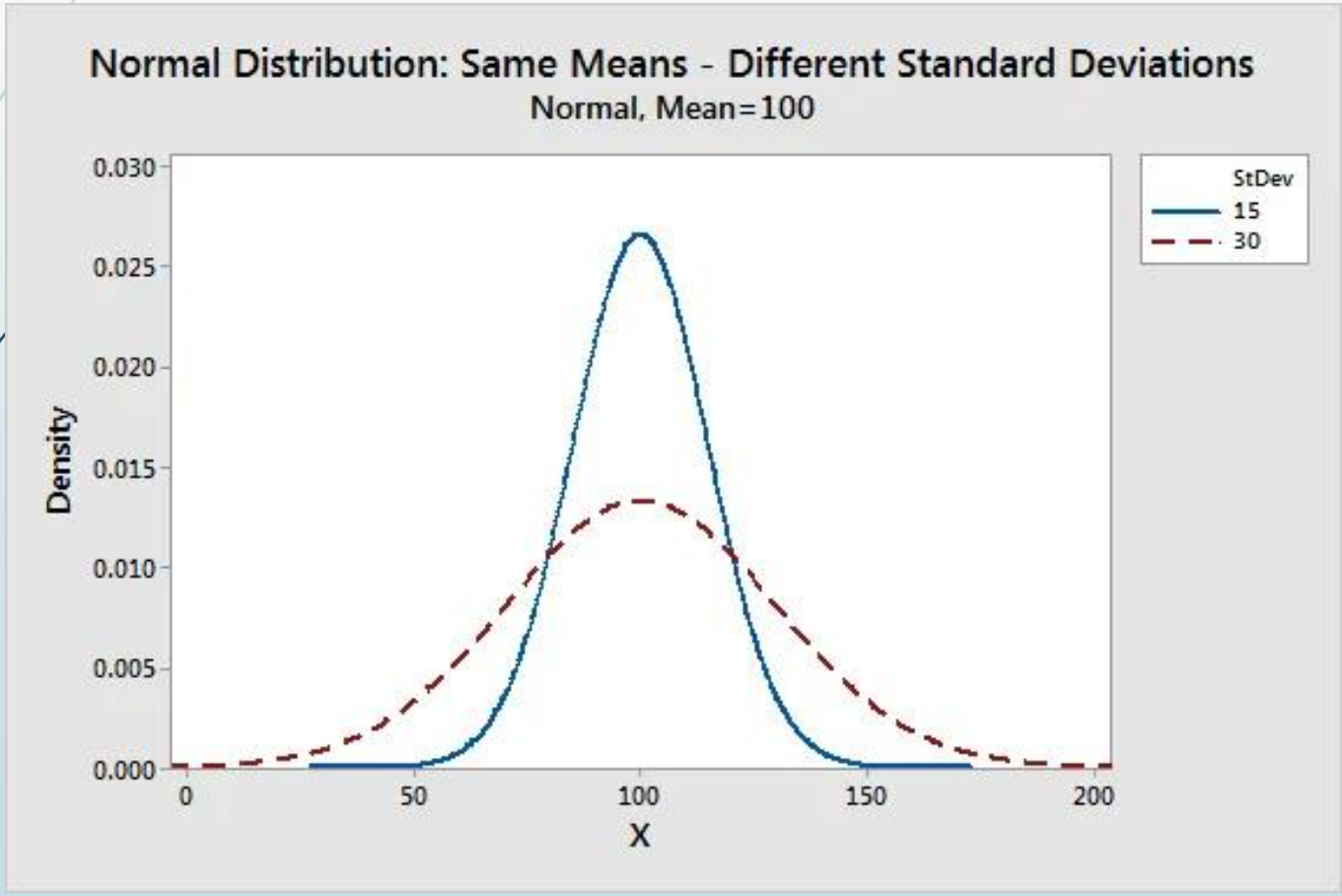


Spread of Data

Standard deviation (σ) measures how far a 'typical' observation is from the average of the data (μ).

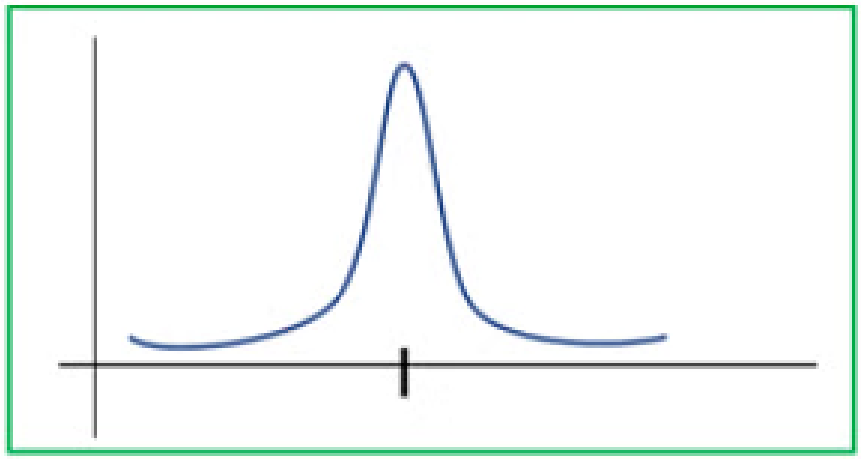


The standard deviation defines the width of the graph.

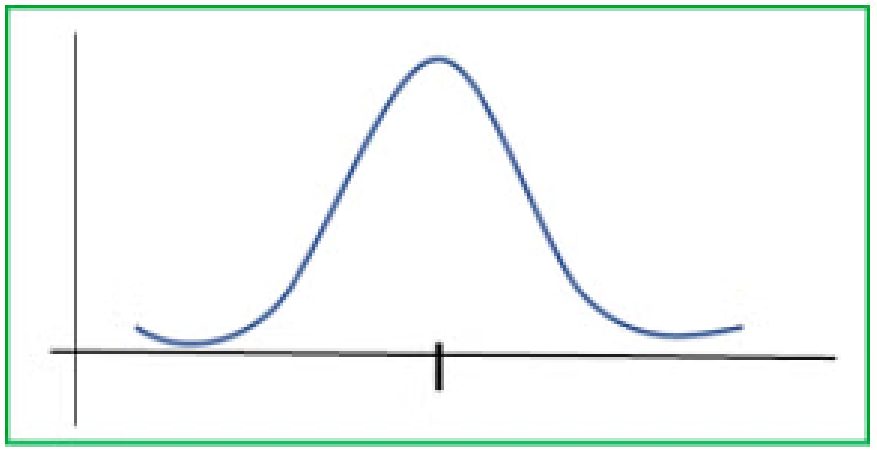


Normal distributions with the same mean but different standard deviations.

Income in rural and urban areas:



The individual income in rural areas does not vary by much, hence we see less deviation from the average

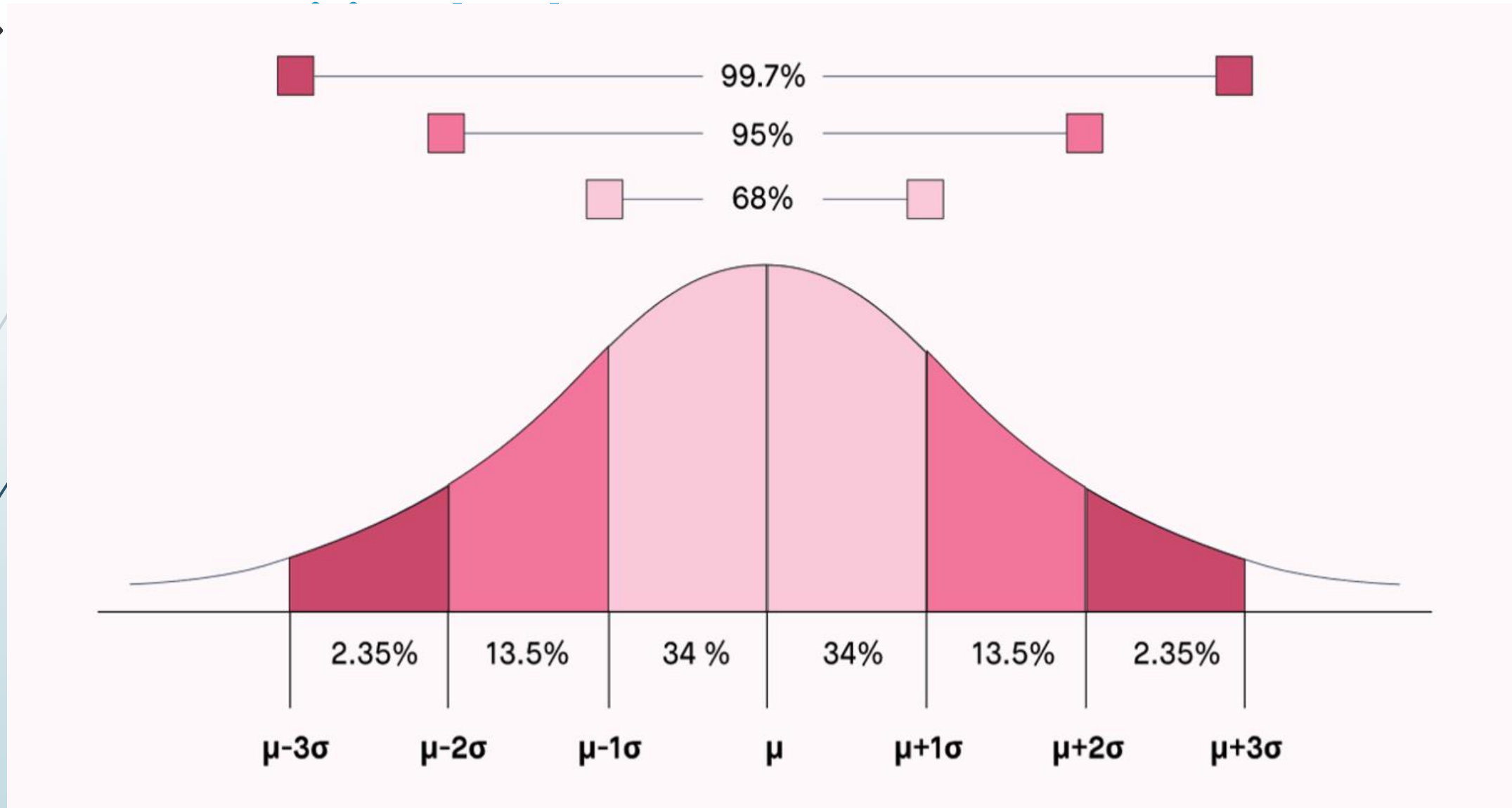


The individual income in urban areas varies due to uneven wealth distribution, hence deviation is more spread out

Empirical rule

Empirical rule

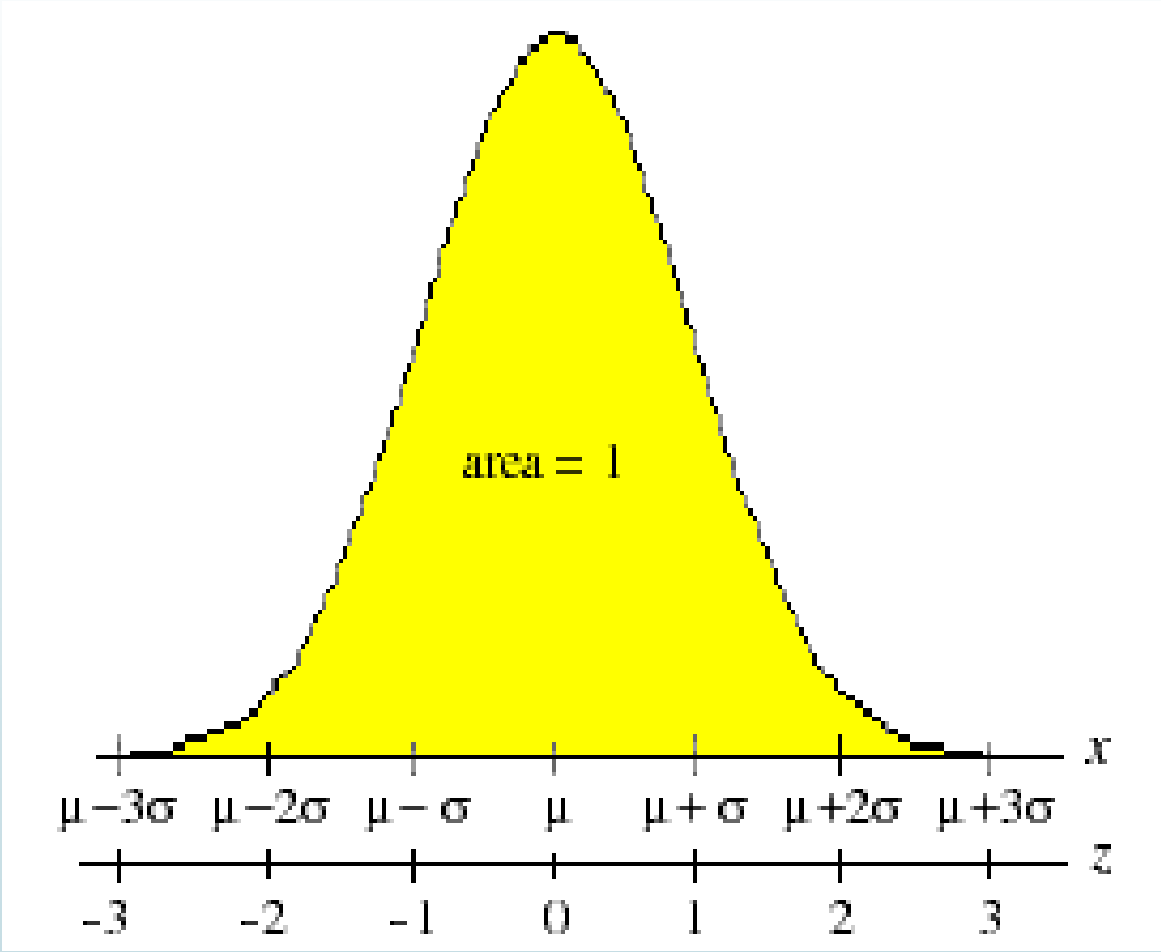
An empirical rule in statistics states almost all of the observations in a normal distribution lie within three standard deviations from the mean.


99.7%
95%
68%
2.35%
13.5%
34 %
34%
13.5%
2.35%
μ-3σ
μ-2σ
μ-1σ
μ
μ+1σ
μ+2σ
μ+3σ

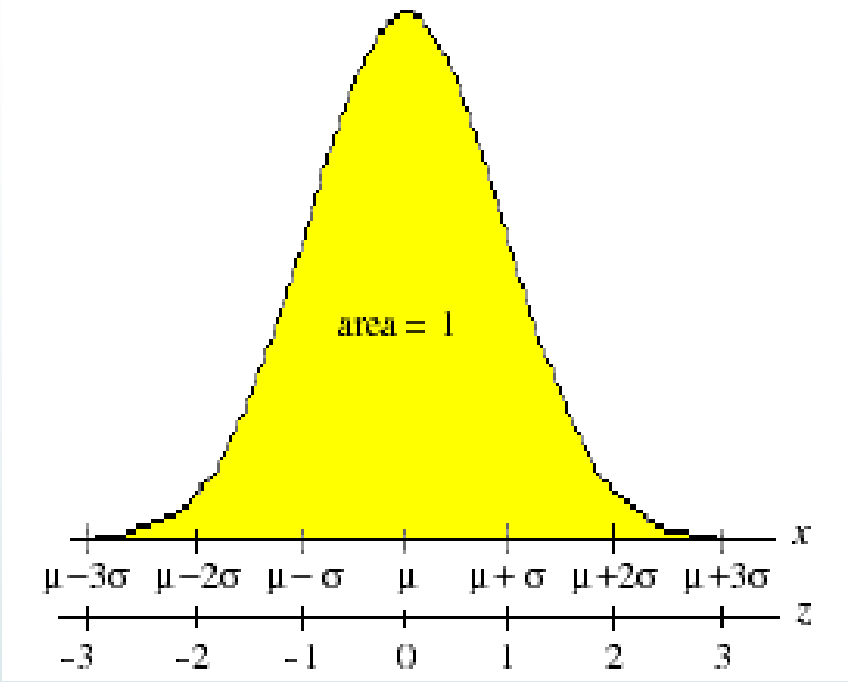
Empirical rule

- 68% of the observations will lie within (+-) 1 standard deviation from the mean
- 95% of the observations will lie within (+-) 2 standard deviations from the mean
- 99.7% of the observations will lie within (+-) 3 standard deviations from the mean

Area under the curve

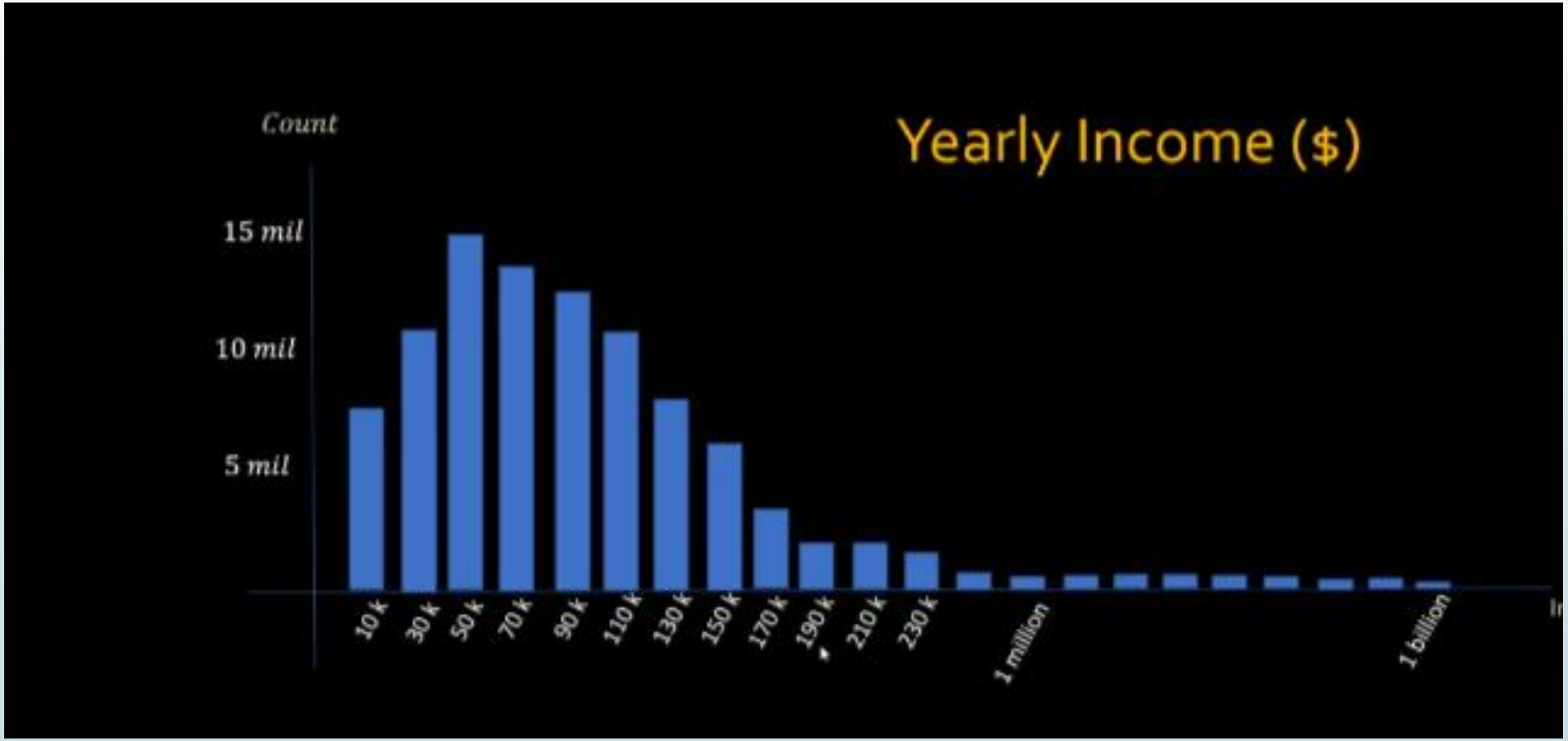


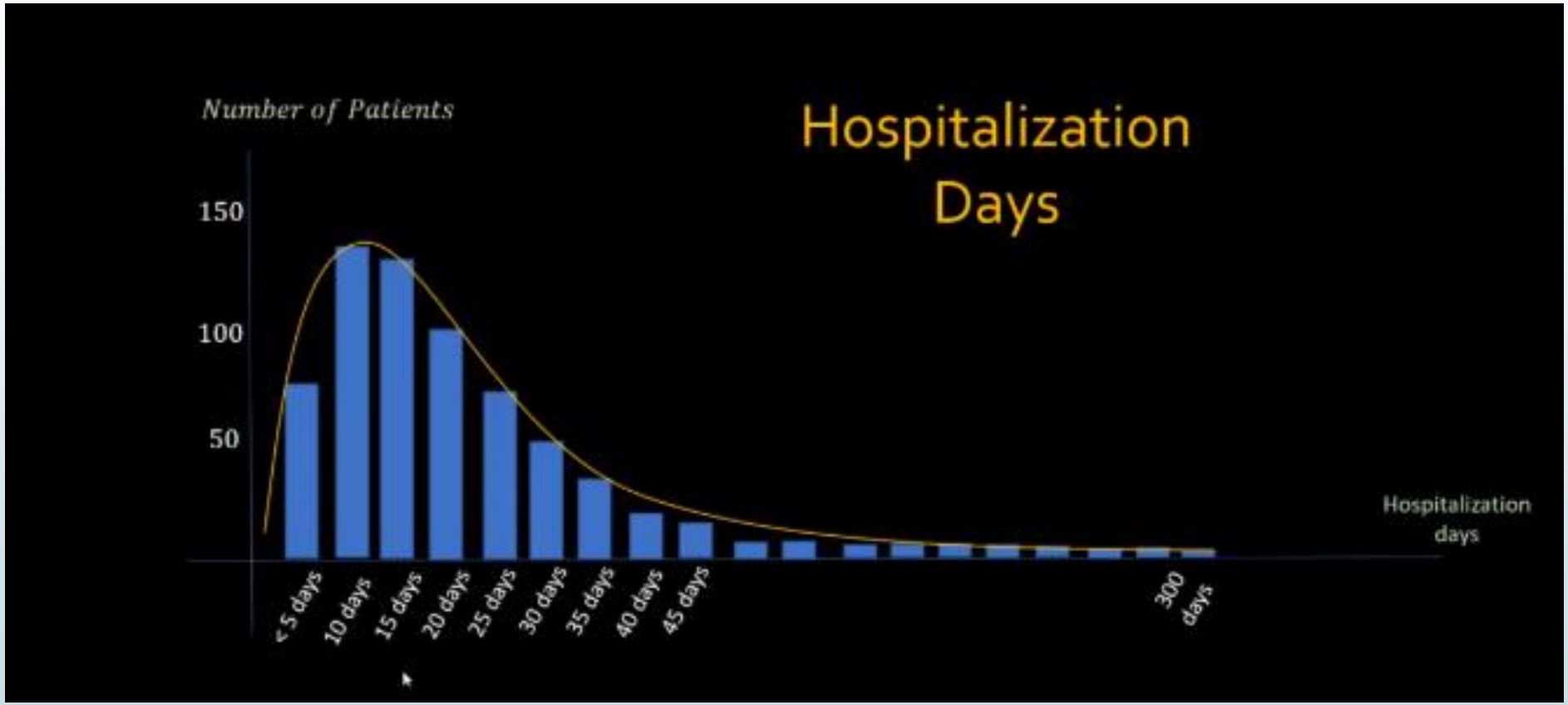
Area under the curve

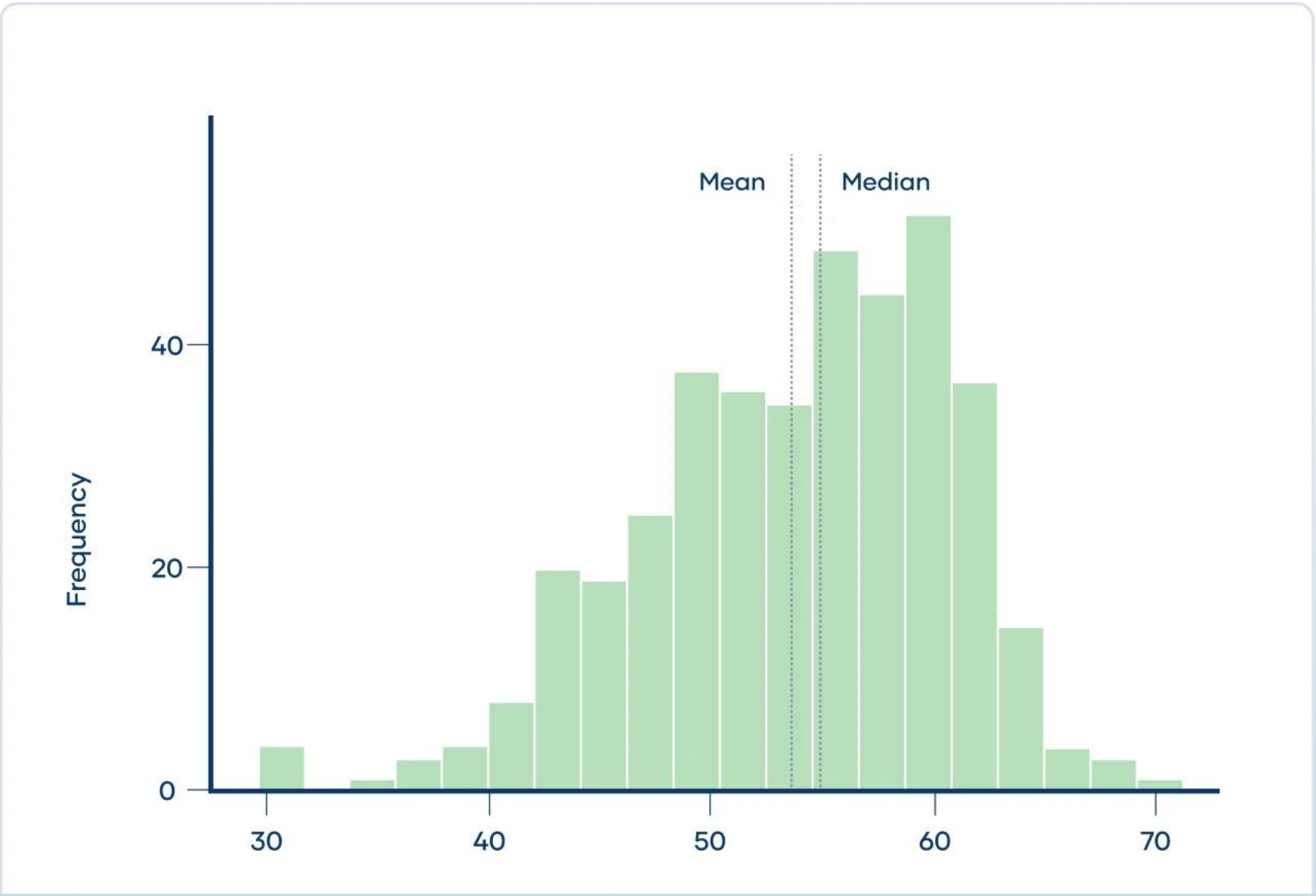


- The area under the curve of the normal distribution represents probabilities for the data.
- The area under the whole curve is equal to 1, or 100%

How different the distribution is from a normal distribution?


Count
Yearly Income ($)
15 mil
10 mil
5 mil
10 k
30 k
50 k
70 k
90 k
110 k
130 k
150 k
170 k
190 k
210 k
230 k
1 million
1 billion

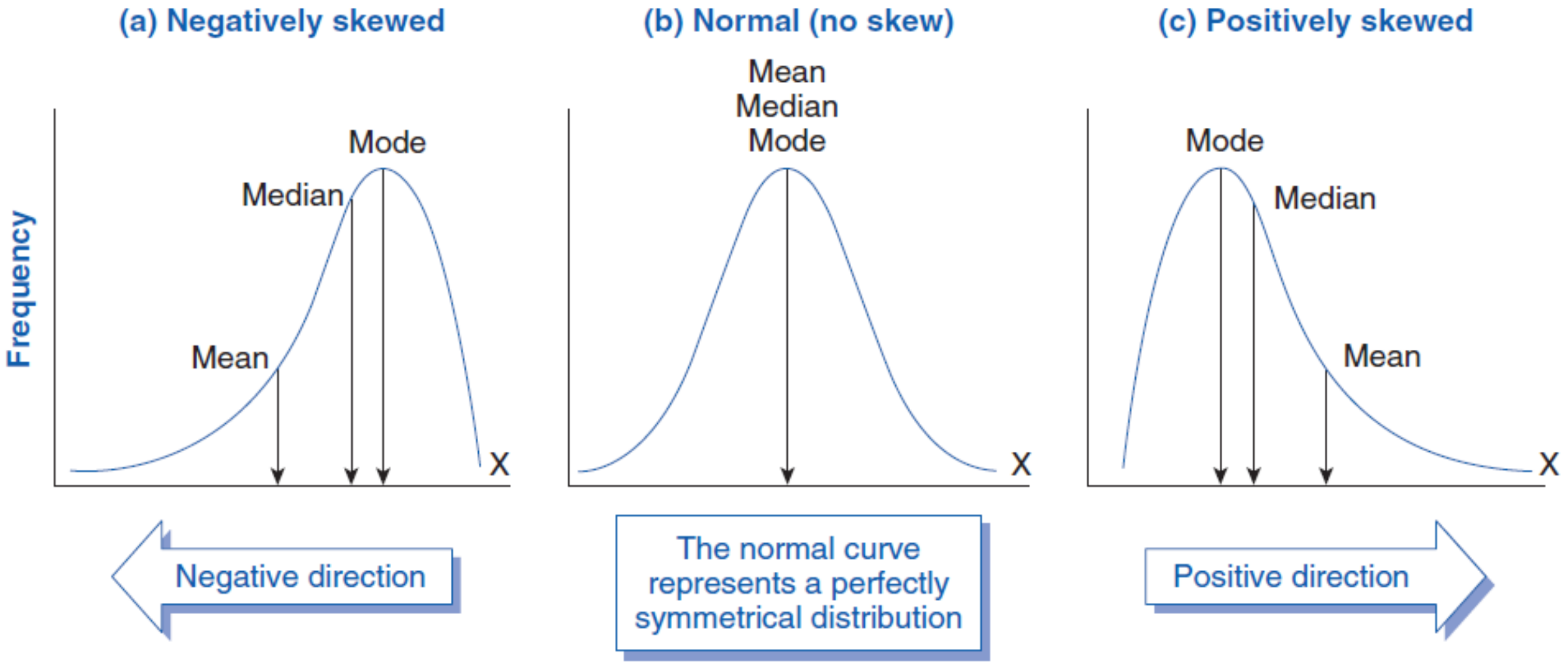

Number of Patients
Hospitalization Days
150
100
50
< 5 days
10 days
15 days
20 days
25 days
30 days
35 days
40 days
45 days
300 days
Hospitalization days


Mean
Median
Frequency
40
20
0
30
40
50
60
70

Skewness

- **Skewness** is a measure of the asymmetry of a distribution.
- A normal distribution has a skewness of zero, which means it is symmetric and balanced.
- In a distribution with zero skew, the mean and median are equal.

Skewness


(a) Negatively skewed
(b) Normal (no skew)
(c) Positively skewed
Frequency
Mode
Median
Mean
Mean
Median
Mode
Mode
Median
Mean
X
X
X
Negative direction
The normal curve represents a perfectly symmetrical distribution
Positive direction

Skewness

Right skew (positive skew)

- Longer on the right side of its peak than on its left.
- Has a long tail on its right side.
- Right skew:

  mean > median

Left skew (negative skew)

- Longer on the left side of its peak than on its right.
- Has a long tail on its left side.
- Left skew:

  mean < median

The normal Distribution is kind of magical in that wee see it a lot in nature.

That makes it super useful for statistics and data science as well.