BYE, BYE DESCRIPTIVE STATISTICS

Probability theory

Distributions

INFERENTIAL STATISTICS

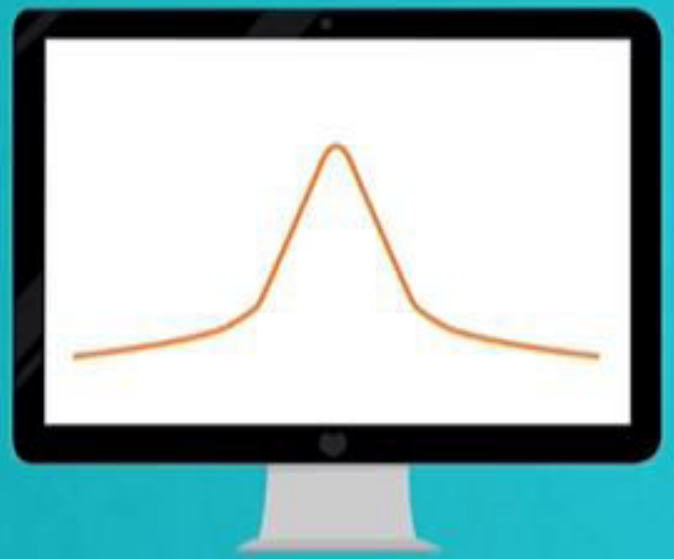
WHAT IS A DISTRIBUTION

IN STATISTICS

DISTRIBUTION

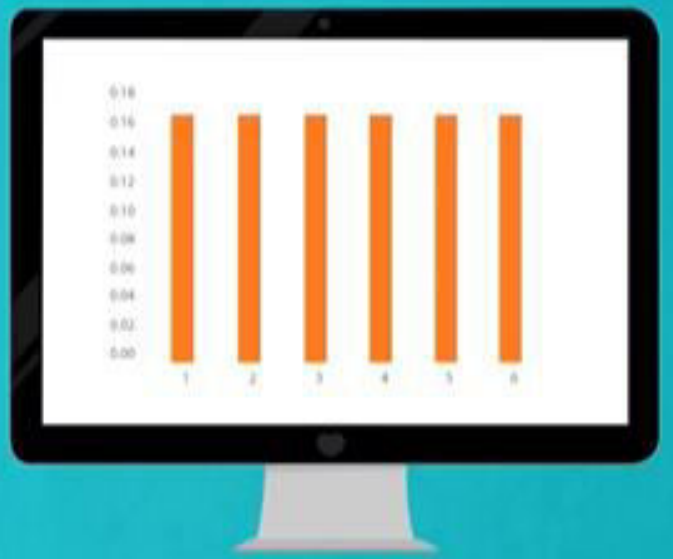
PROBABILITY DISTRIBUTION

DISTRIBUTION = PROBABILITY DISTRIBUTION



NORMAL

BINOMIAL



UNIFORM

DEFINITION

A distribution is a function that shows the possible values for a variable and how often they occur.

ROLLING A DIE

Probability of getting

$\frac{1}{6}$

ROLLING A DIE

Probability of getting ● ●

$\frac{1}{6}$

ROLLING A DIE

OUTCOME	PROBABILITY
1	1/6
2	
3	
4	
5	
6	

WHAT IS THE PROBABILITY OF GETTING A

7

ROLLING A DIE

✕ IMPOSSIBLE

ROLLING A DIE

OUTCOME	PROBABILITY
1; 2; 3; 4; 5; 6	1/6
7	0

ROLLING A DIE

OUTCOME	PROBABILITY
1; 2; 3; 4; 5; 6	0.17
All else	0

ROLLING A DIE

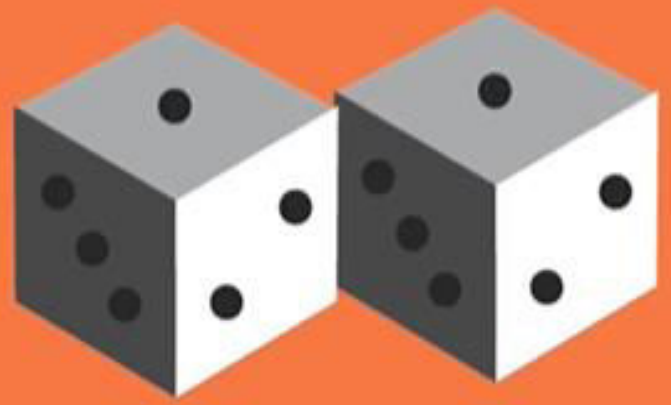
DISCRETE UNIFORM DISTRIBUTION

OUTCOME	PROBABILITY
1; 2; 3; 4; 5; 6	0.17
All else	0

VISUAL REPRESENTATION

ROLLING TWO DICE

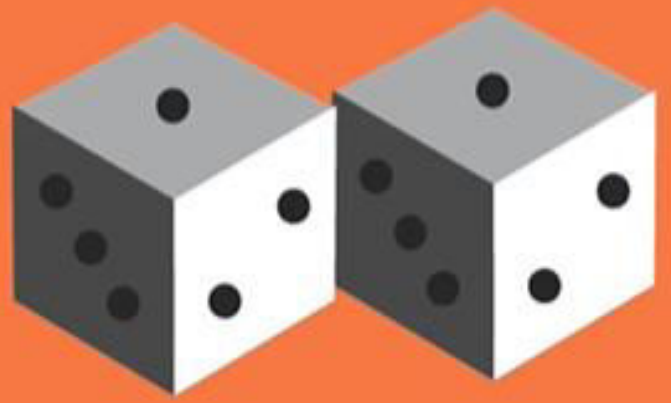
Possibilities?



ROLLING TWO DICE

Possibilities? Total: 36

(1,1) (2,1) (3,1) (4,1) (5,1) (6,1)
(1,2) (2,2) (3,2) (4,2) (5,2) (6,2)
(1,3) (2,3) (3,3) (4,3) (5,3) (6,3)
(1,4) (2,4) (3,4) (4,4) (5,4) (6,4)
(1,5) (2,5) (3,5) (4,5) (5,5) (6,5)
(1,6) (2,6) (3,6) (4,6) (5,6) (6,6)



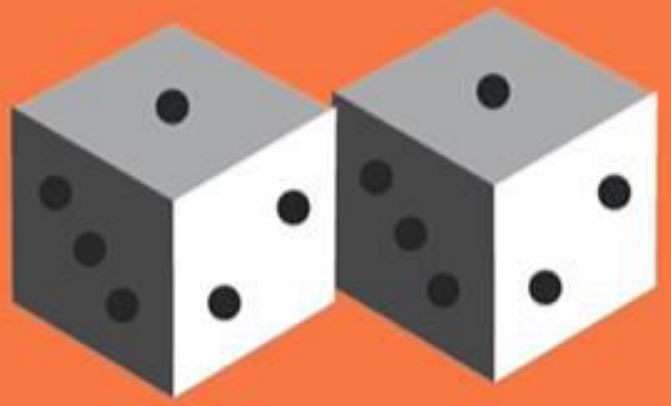
ROLLING TWO DICE

Probability of getting

ROLLING TWO DICE

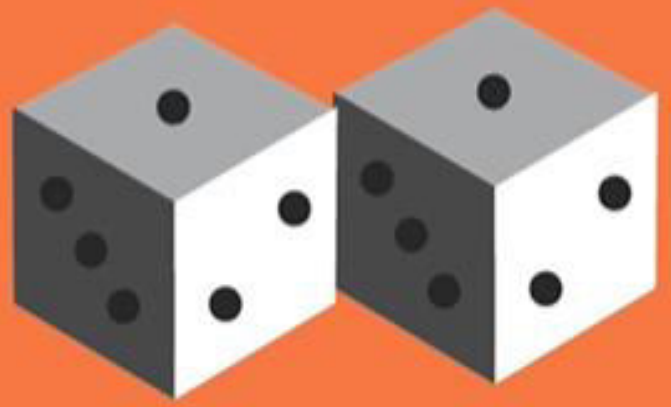
Probability of getting

0



ROLLING TWO DICE

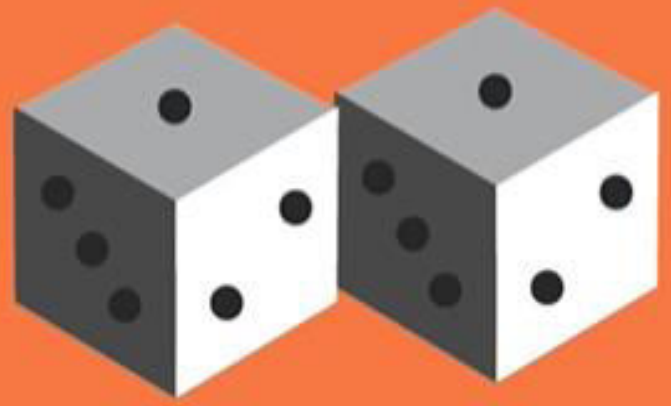
Probability of getting ●●



ROLLING TWO DICE

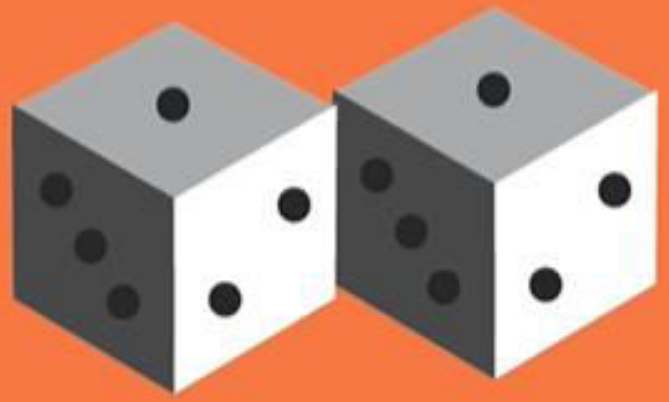
Probability of getting

$$\frac{1}{36} \text{ or } 0.03$$



ROLLING TWO DICE

Probability of getting ● ● ●



ROLLING TWO DICE

Probability of getting ● ● ●

$\frac{2}{36}$ (1,2) and (2,1)

ROLLING TWO DICE

OUTCOME	PROBABILITY
2	0.03
3	0.06
4	0.08
5	0.11
6	0.14
7	0.17
8	0.14
9	0.11
10	0.08
11	0.06
12	0.03
All else	0


0.18
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
2
3
4
5
6
7
8
9
10
11
12

PROBABILITY OF GETTING A 7 IS THE HIGHEST

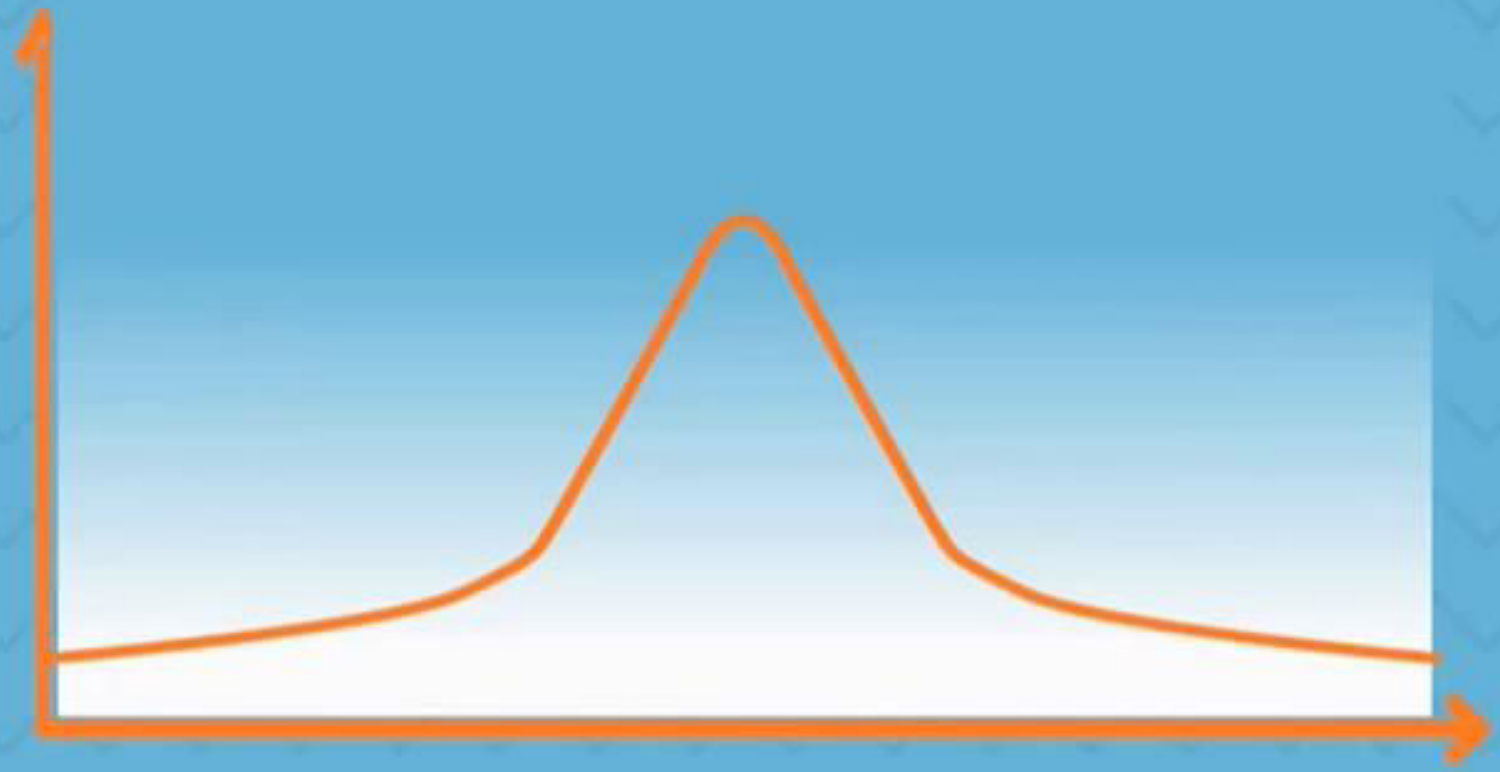
What is a distribution?

- ◯ A graph representing the probability of occurrence of a variable.
- ◯ A graph representing the possible values of a variable and the probability of their occurrence.
- ◯ A distribution is a function that shows the possible values for a variable and the probability of their occurrence.
- ◯ A distribution is a function that shows the probability of a variable.

THE DISTRIBUTION OF A DATASET

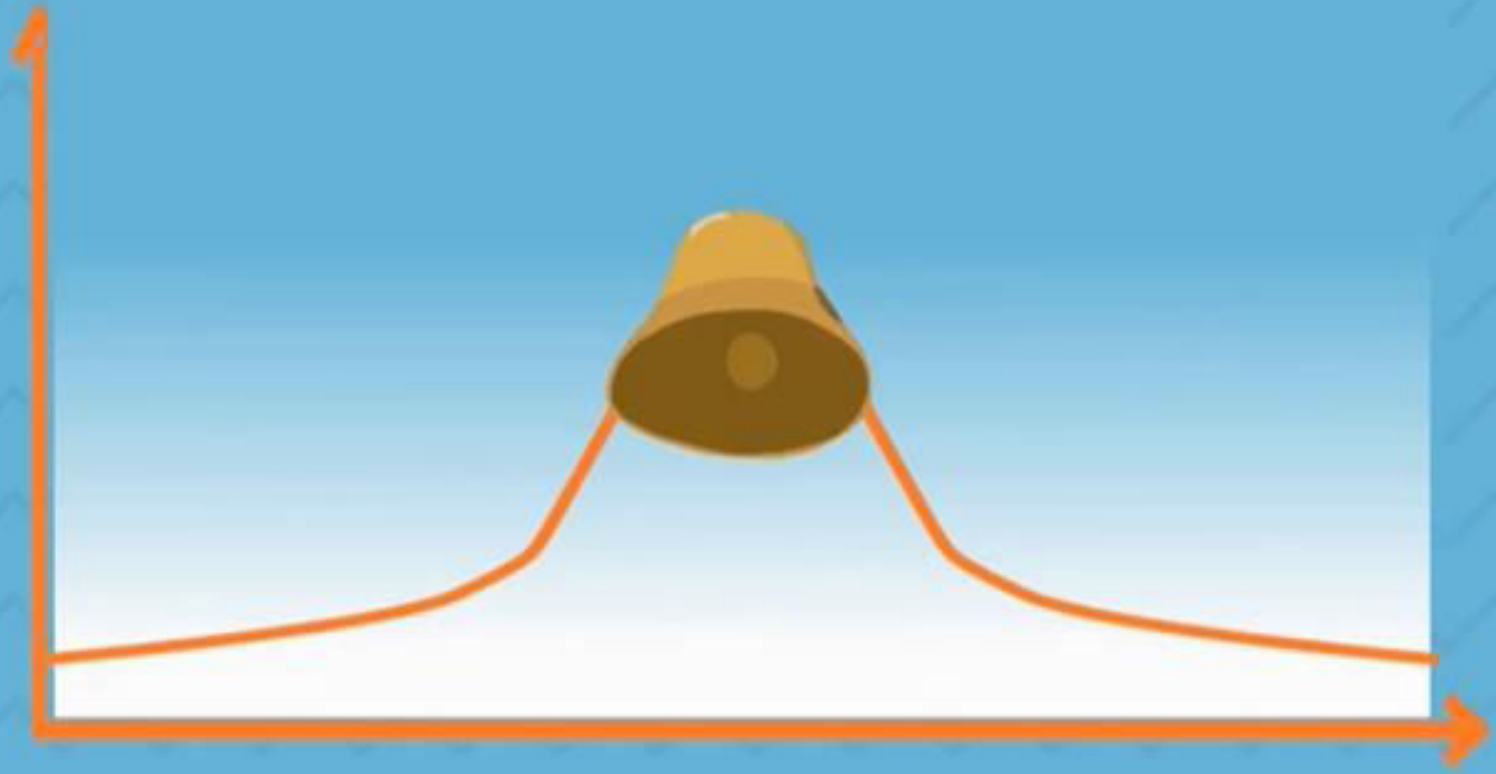
SHOWS US THE FREQUENCY AT WHICH
POSSIBLE VALUES OCCUR

NORMAL DISTRIBUTION



NORMAL DISTRIBUTION

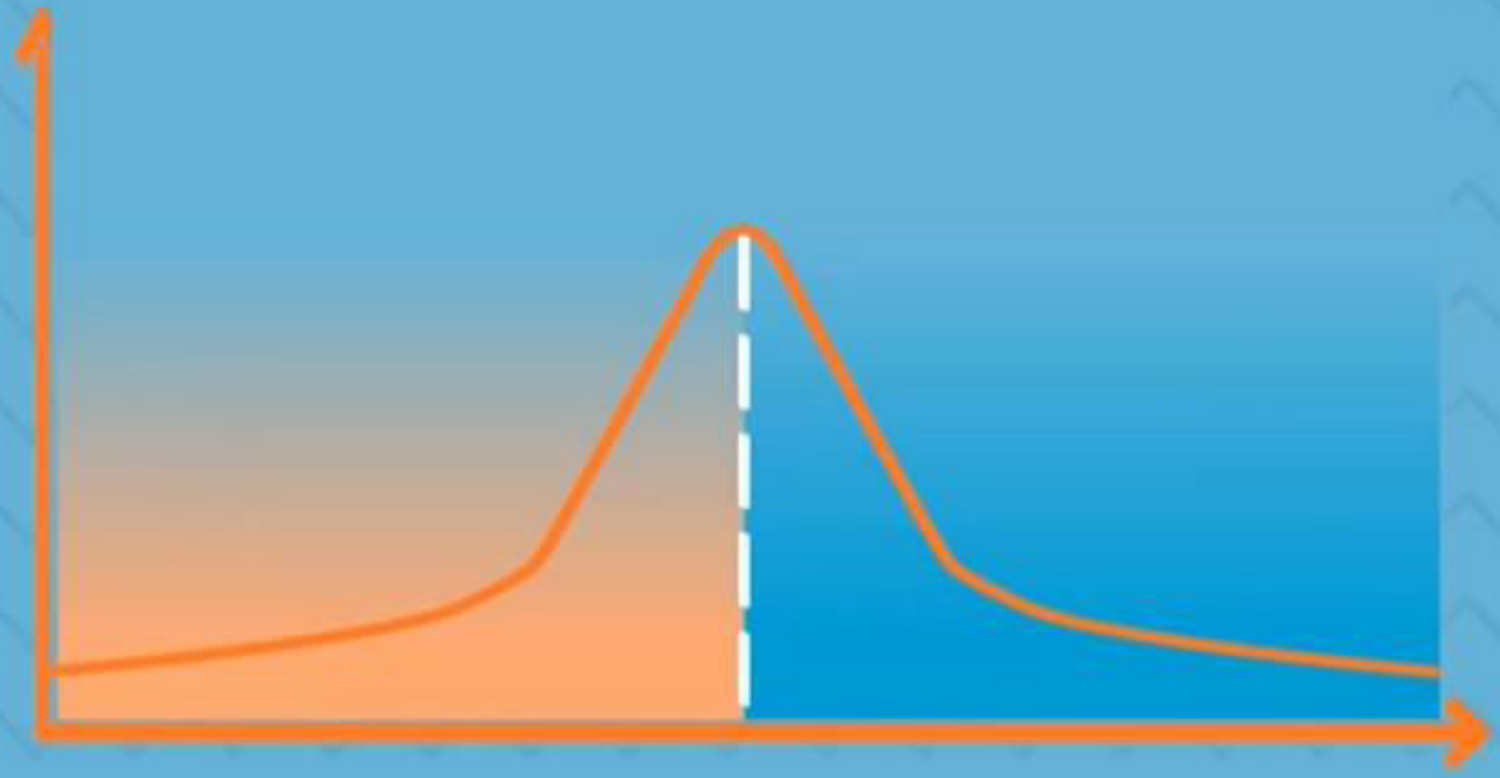
GAUSSIAN DISTRIBUTION



BELL CURVE

NORMAL DISTRIBUTION

GAUSSIAN DISTRIBUTION



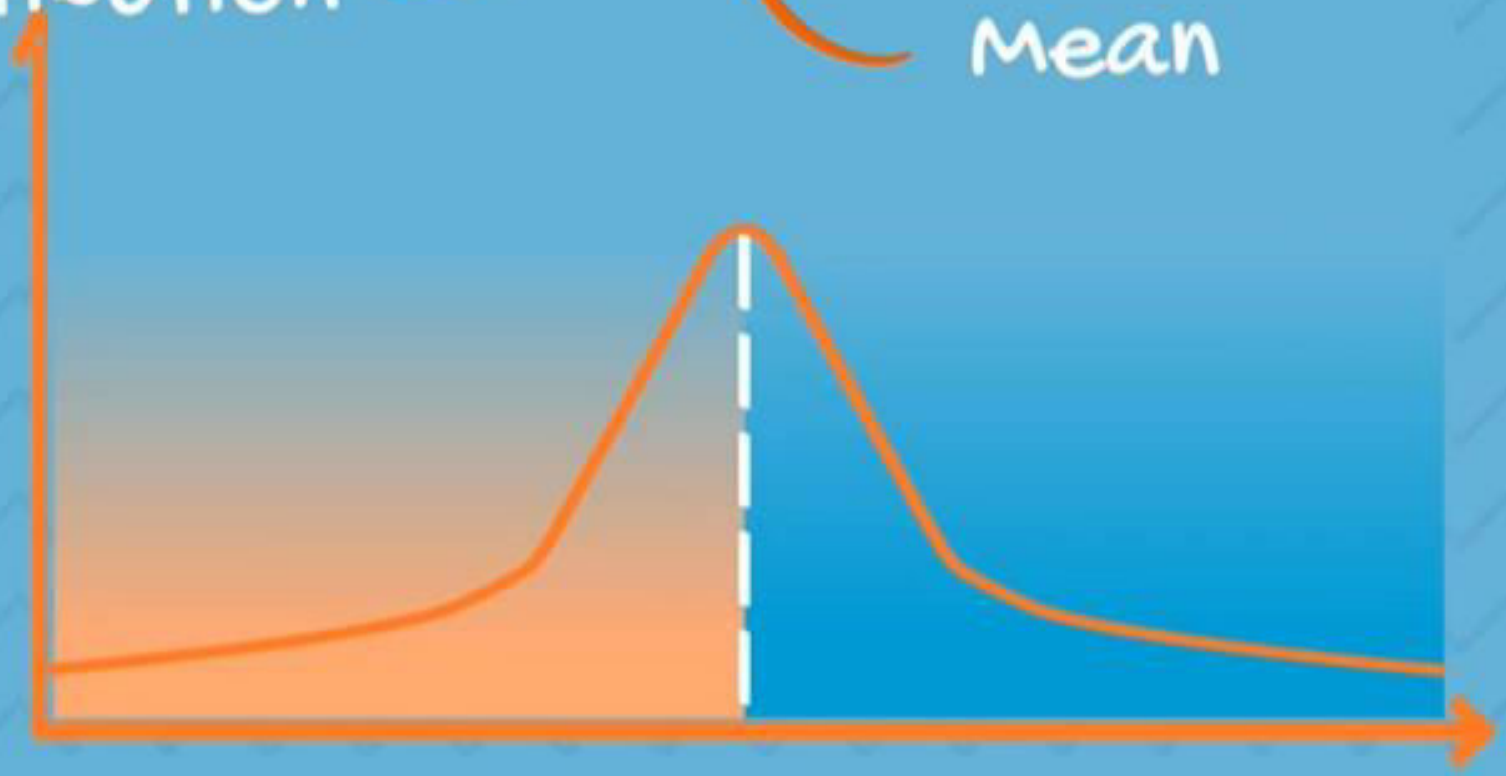
mean = median = mode

Normal → $N \sim (\mu, \sigma^2)$

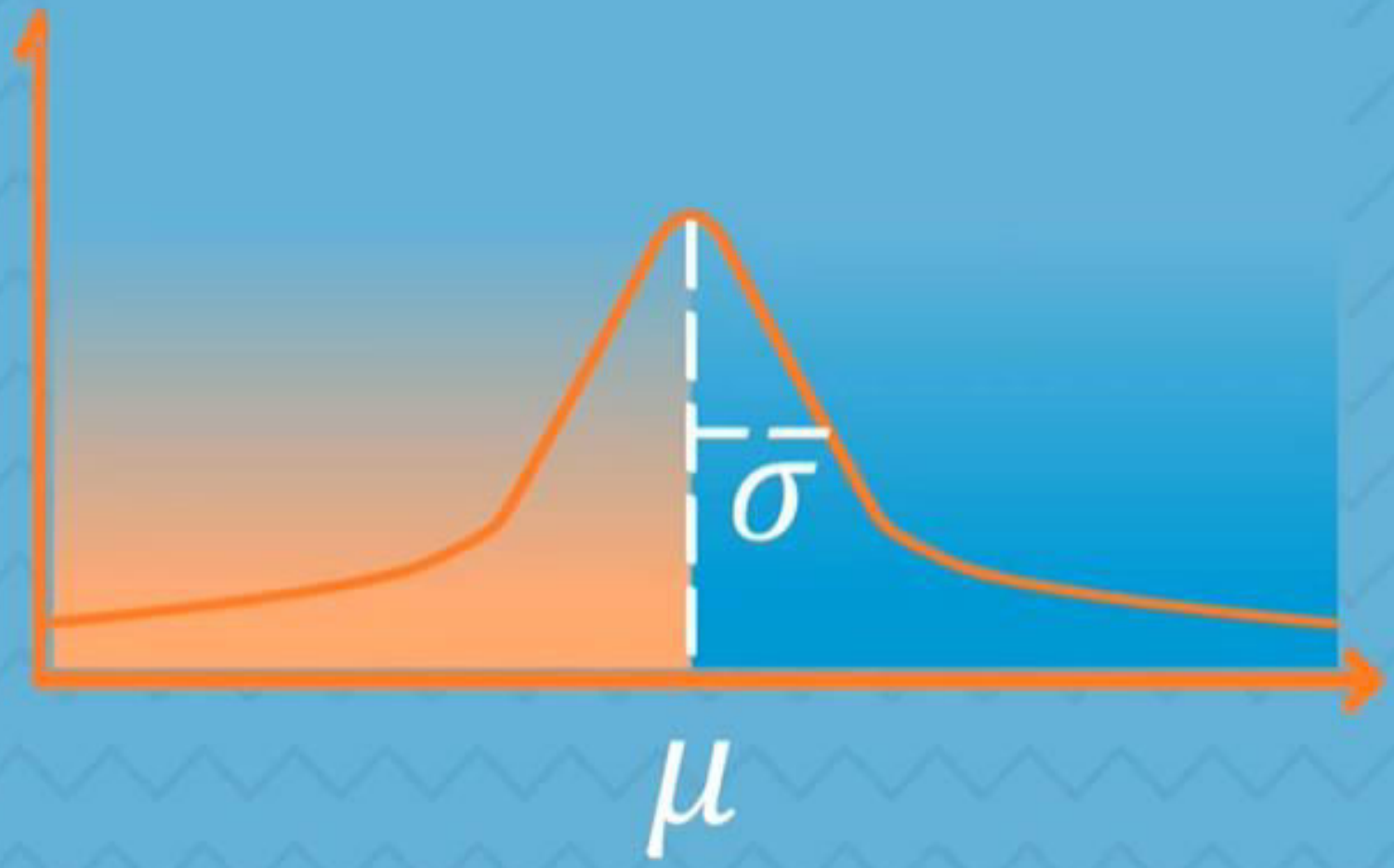
Distribution

Mean

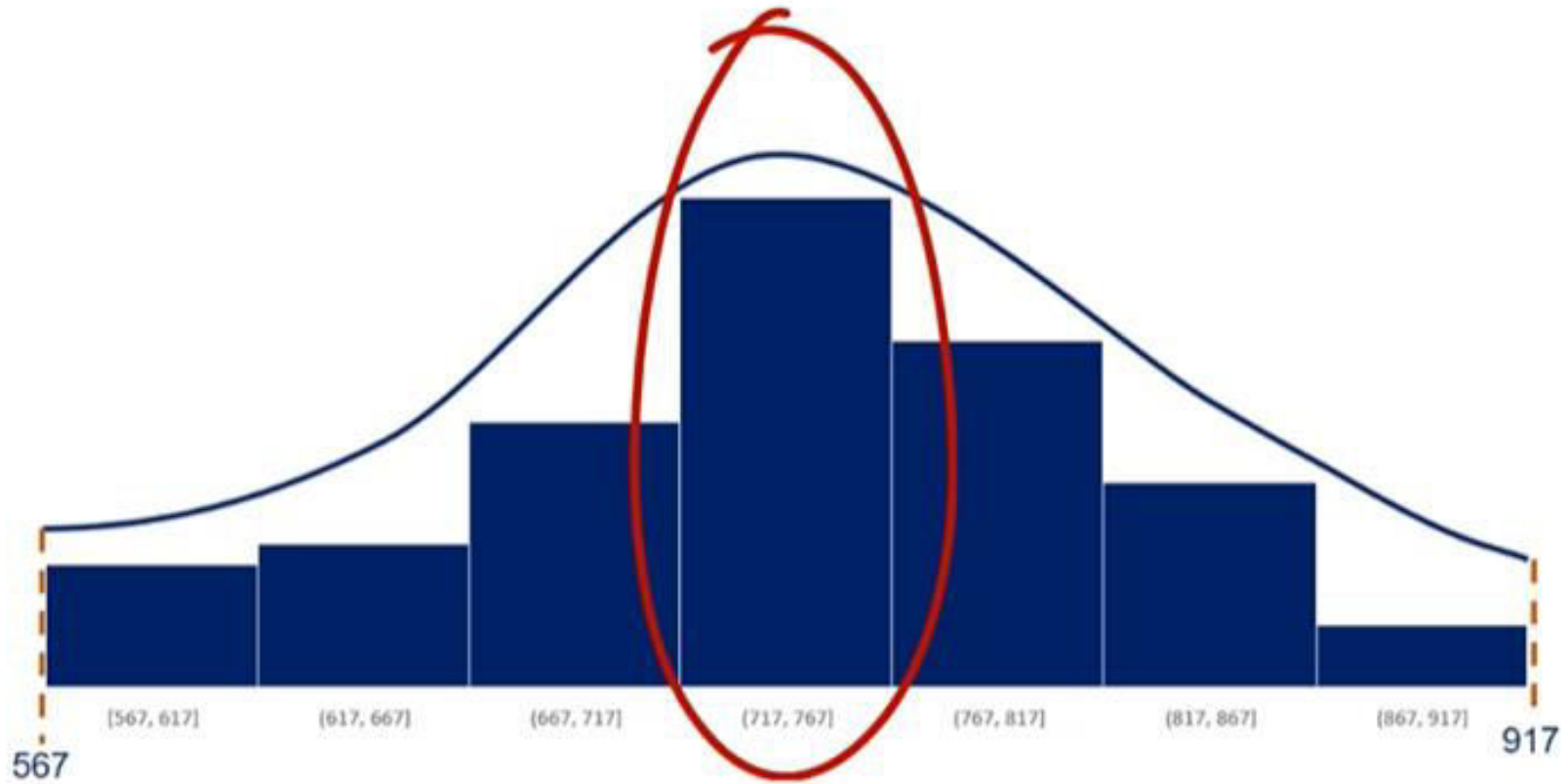
Variance



$$N \sim (\mu, \sigma^2)$$

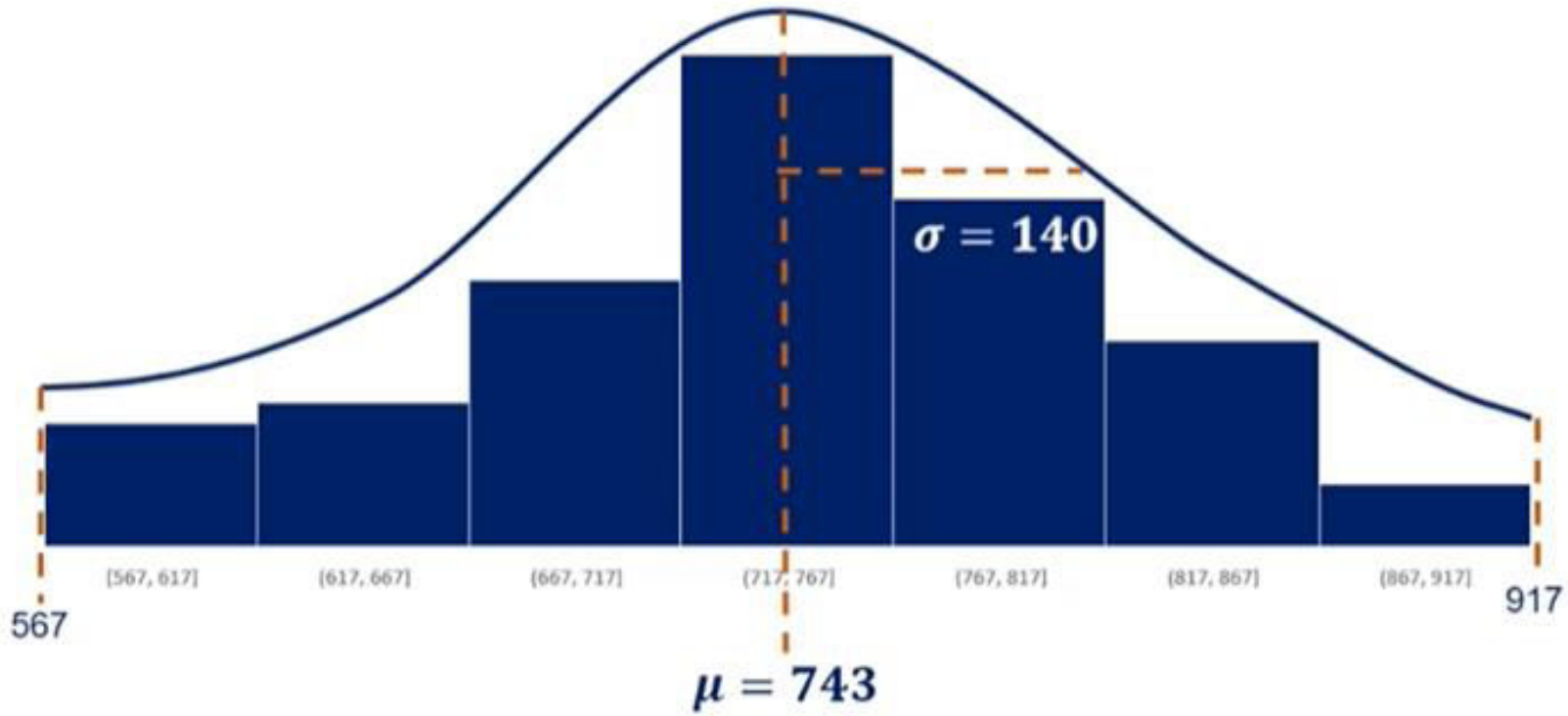


Normal distribution



mean = median = mode

Normal distribution



STANDARDIZATION

STANDARDIZATION

every distribution can be standardized

STANDARDIZATION

$$\sim (\mu, \sigma^2) \Rightarrow \sim (0,1)$$

STANDARDIZATION

$$\sim (\mu, \sigma^2) \rightarrow \sim (0,1)$$

$$\frac{x - \mu}{\sigma}$$

STANDARDIZATION

of a Normal distribution

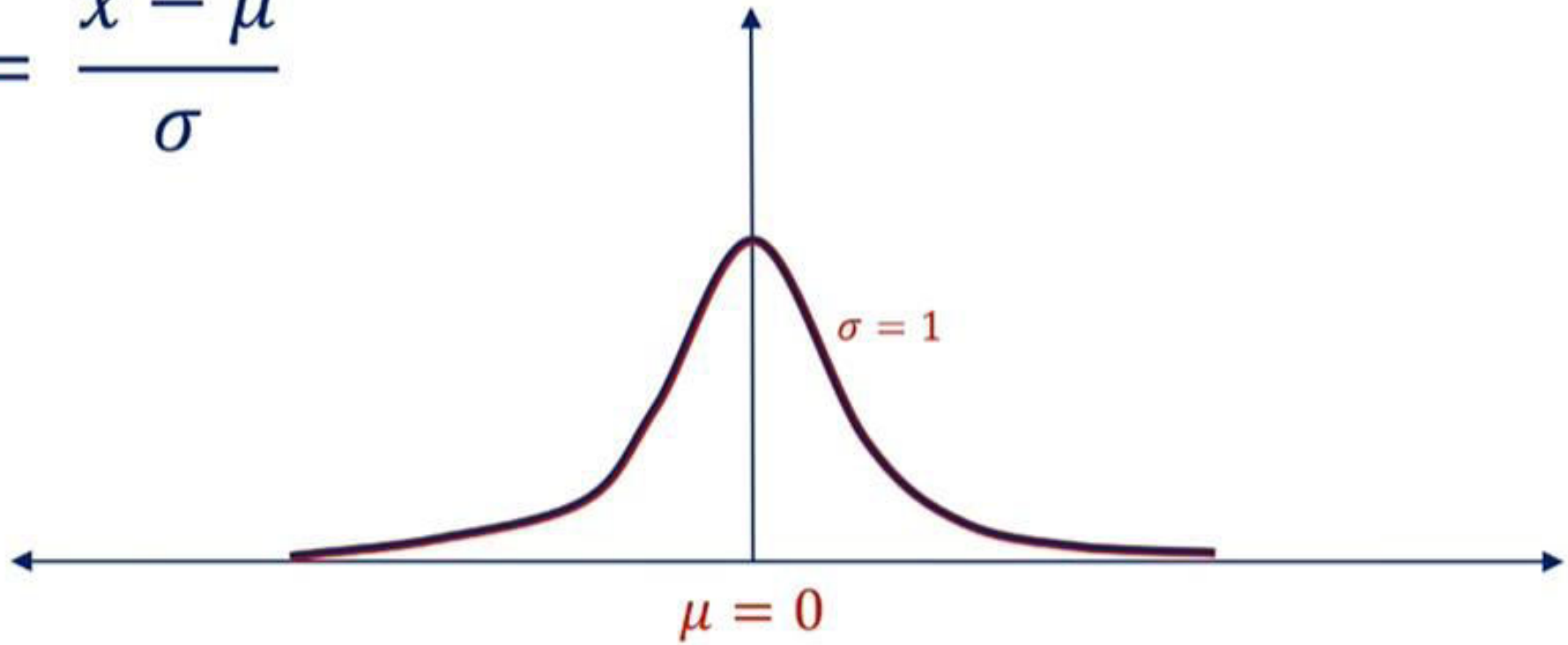
$$\sim N(\mu, \sigma^2) \rightarrow \sim N(0,1)$$

$$Z = \frac{x - \mu}{\sigma}$$

When we standardize a Normal distribution, the result is a Standard Normal distribution

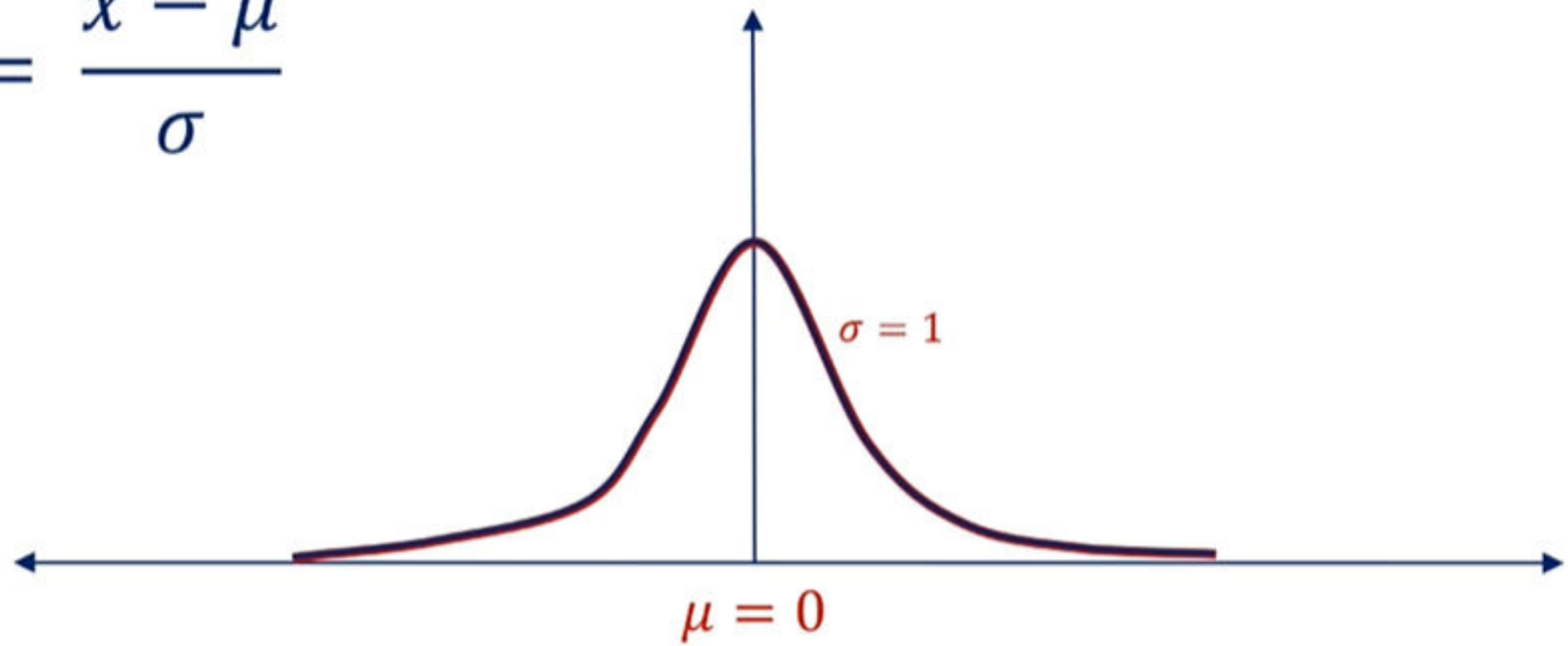
STANDARDIZATION

$$z = \frac{x - \mu}{\sigma}$$



STANDARDIZATION

$$z = \frac{x - \mu}{\sigma}$$



$$z \sim N(0,1)$$

Question 1:

Imagine we have a variable X, which follows a Normal Distribution with a mean of 4 and a variance of 9. We want to standardize Z, so what formula do we use for the transformation?

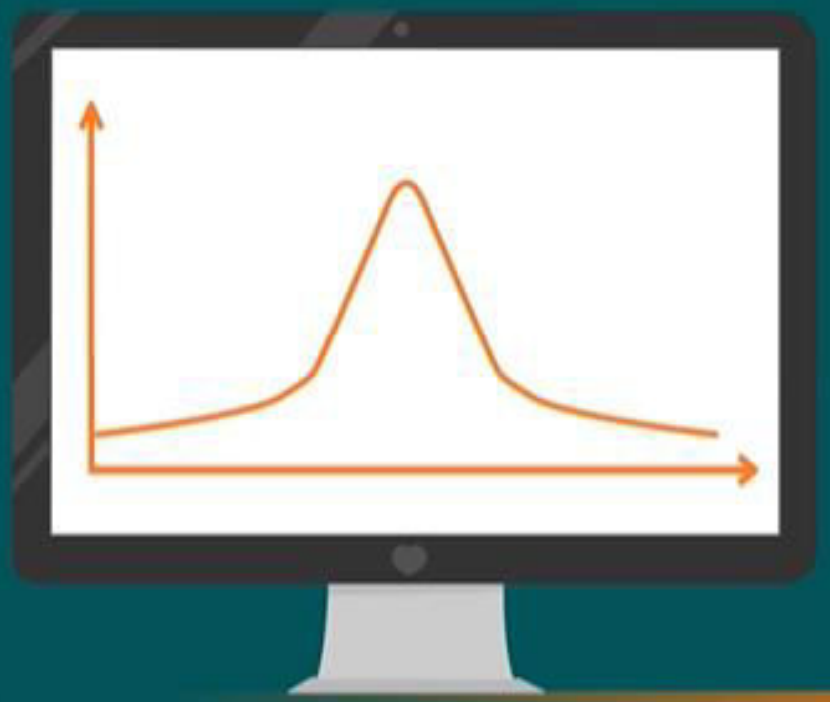
a) $z = \frac{x-4}{3}$

b) $z = \frac{4-x}{3}$

c) $z = \frac{4-x}{9}$

d) $z = \frac{4-x}{-3}$

- ○ a)
- ○ b)
- ○ c)
- ○ d)



CENTRAL LIMIT THEOREM

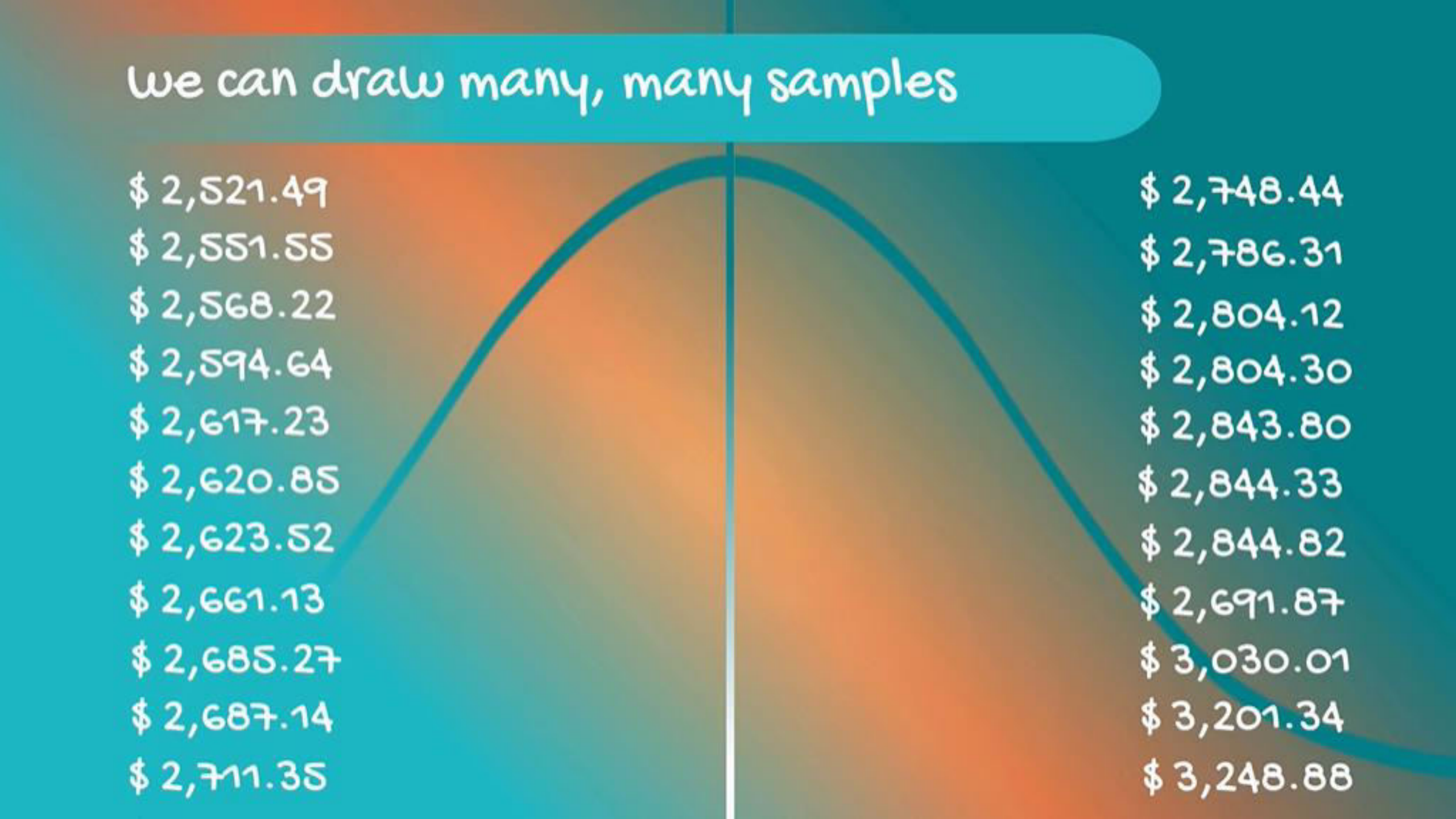
POPULATION OF USED CARS IN A CAR SHOP


POPULATION OF USED CARS IN A CAR SHOP
$\mu, \sigma_x, \sigma_{xy}$

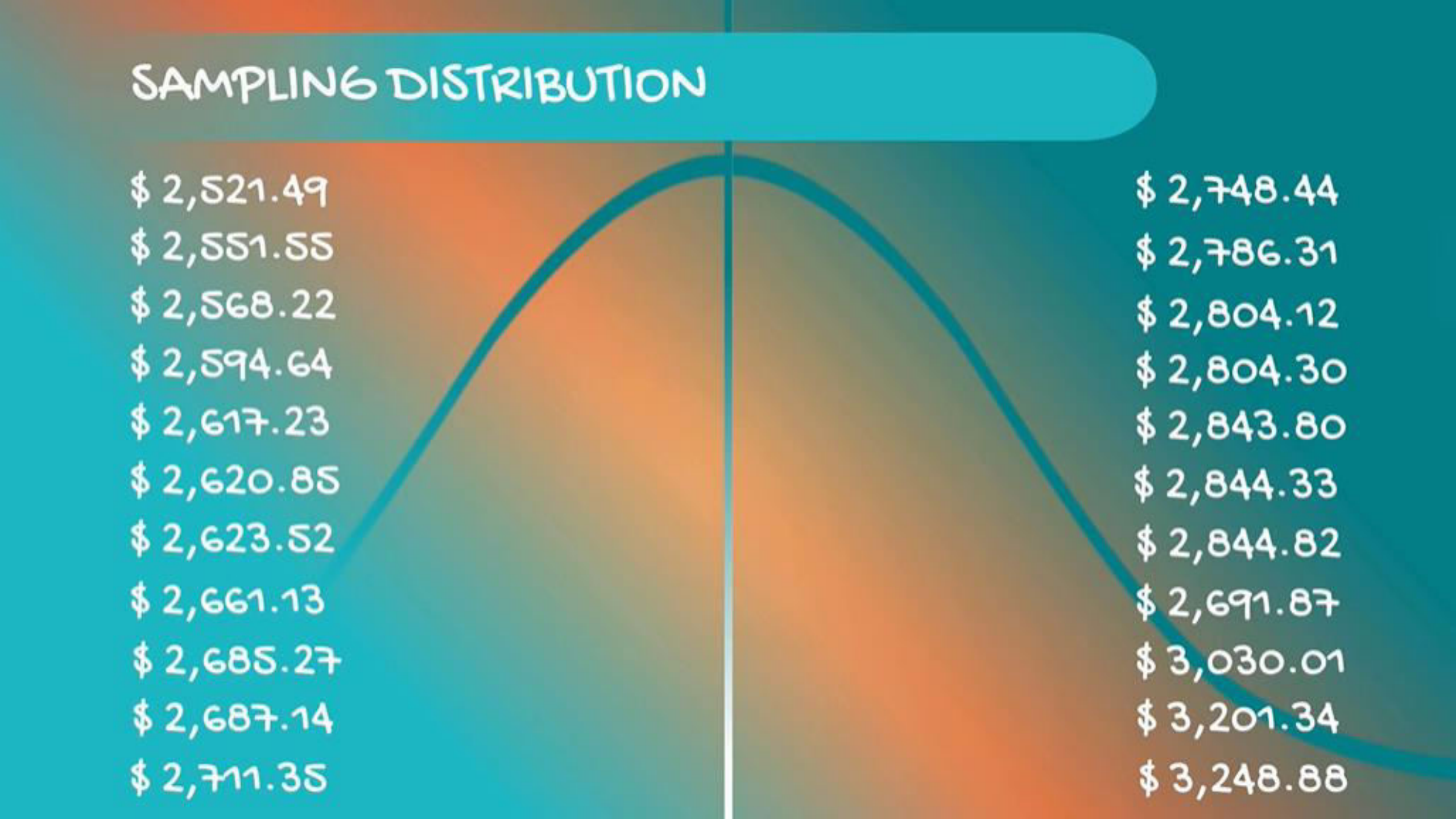
THE MEAN $\bar{x}$

Sample #1:	$2,617.23
Sample #2:	$3,201.34
Sample #3:	$2,844.33

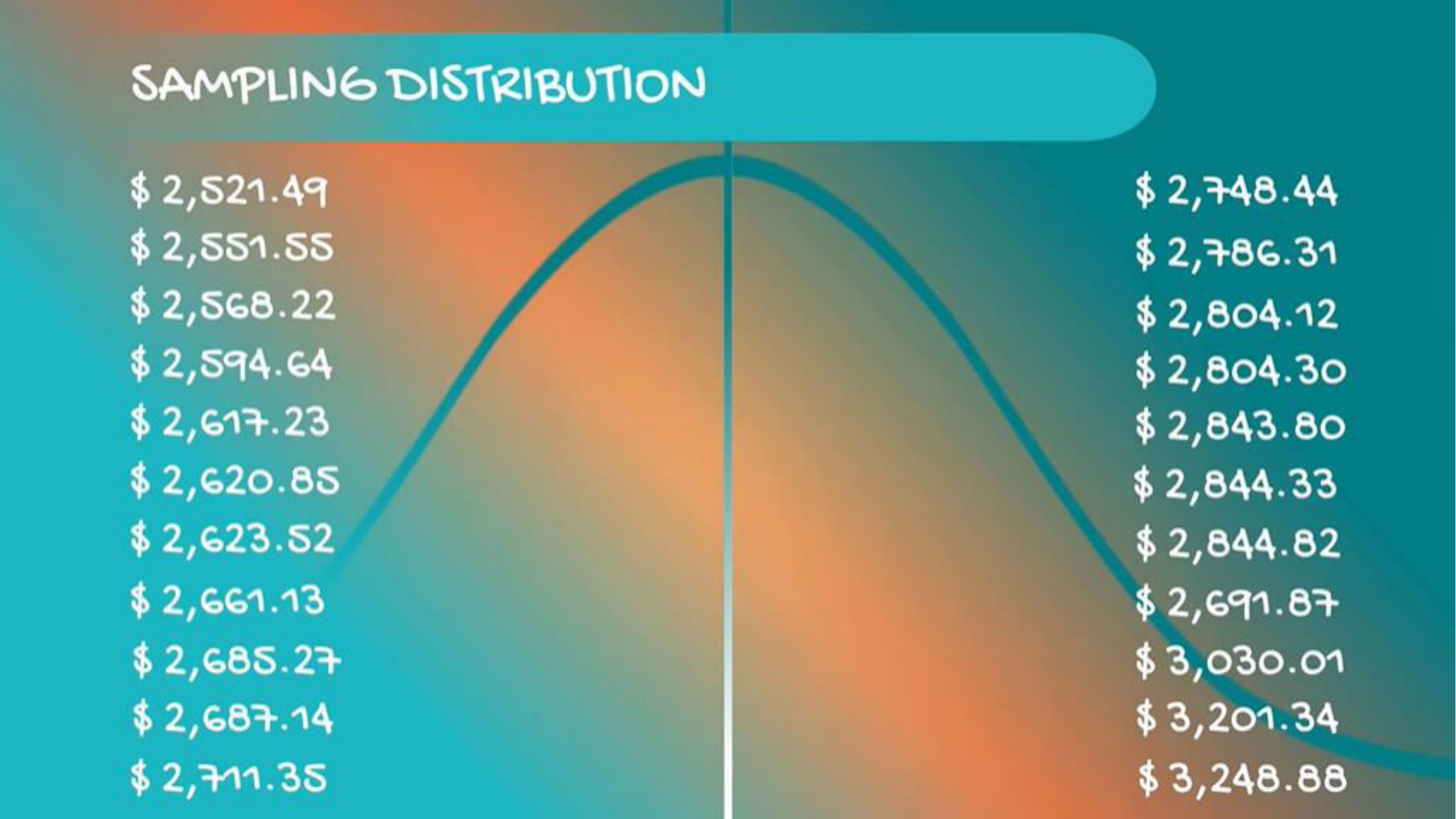
we can draw many, many samples



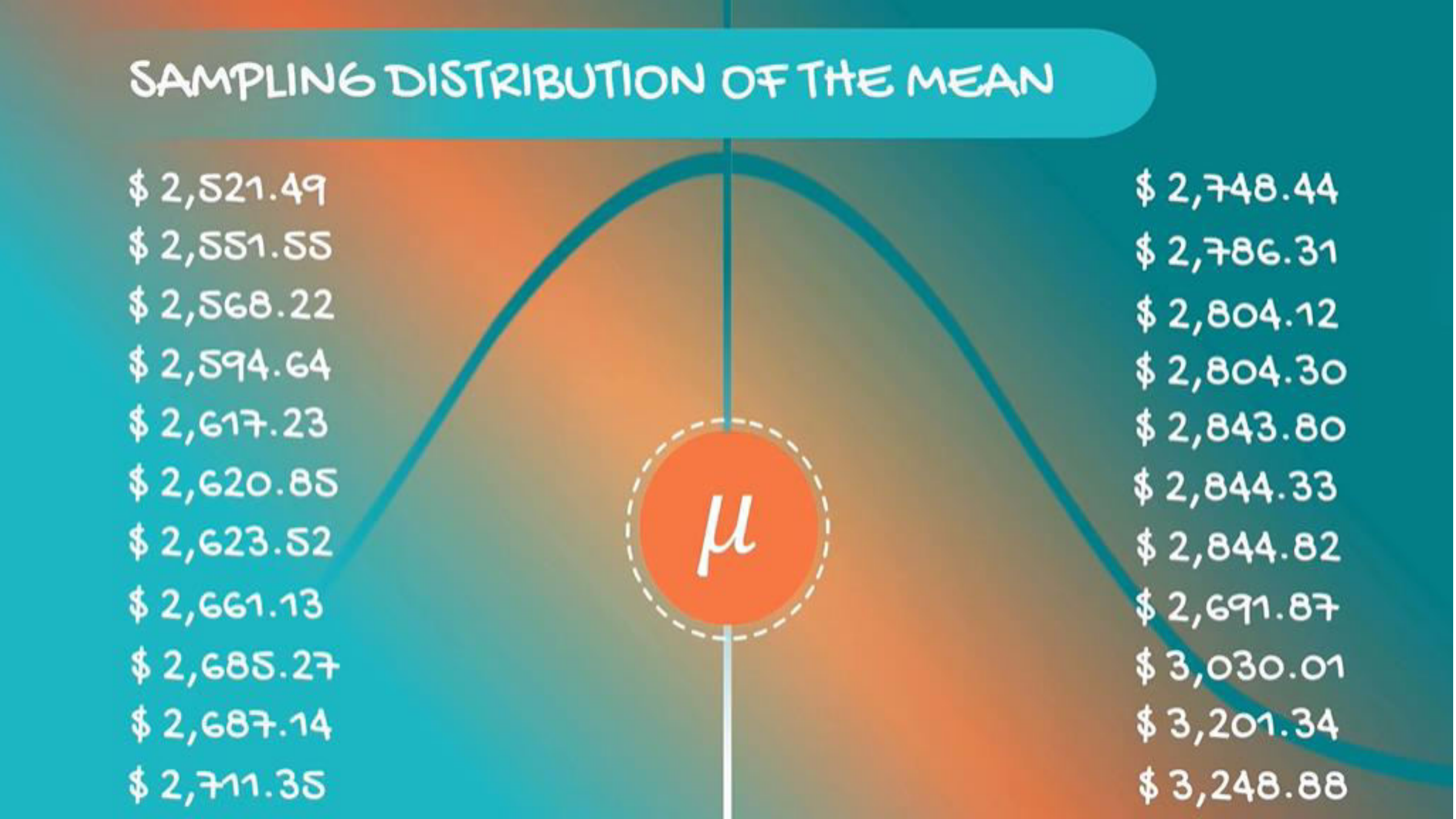
SAMPLING DISTRIBUTION



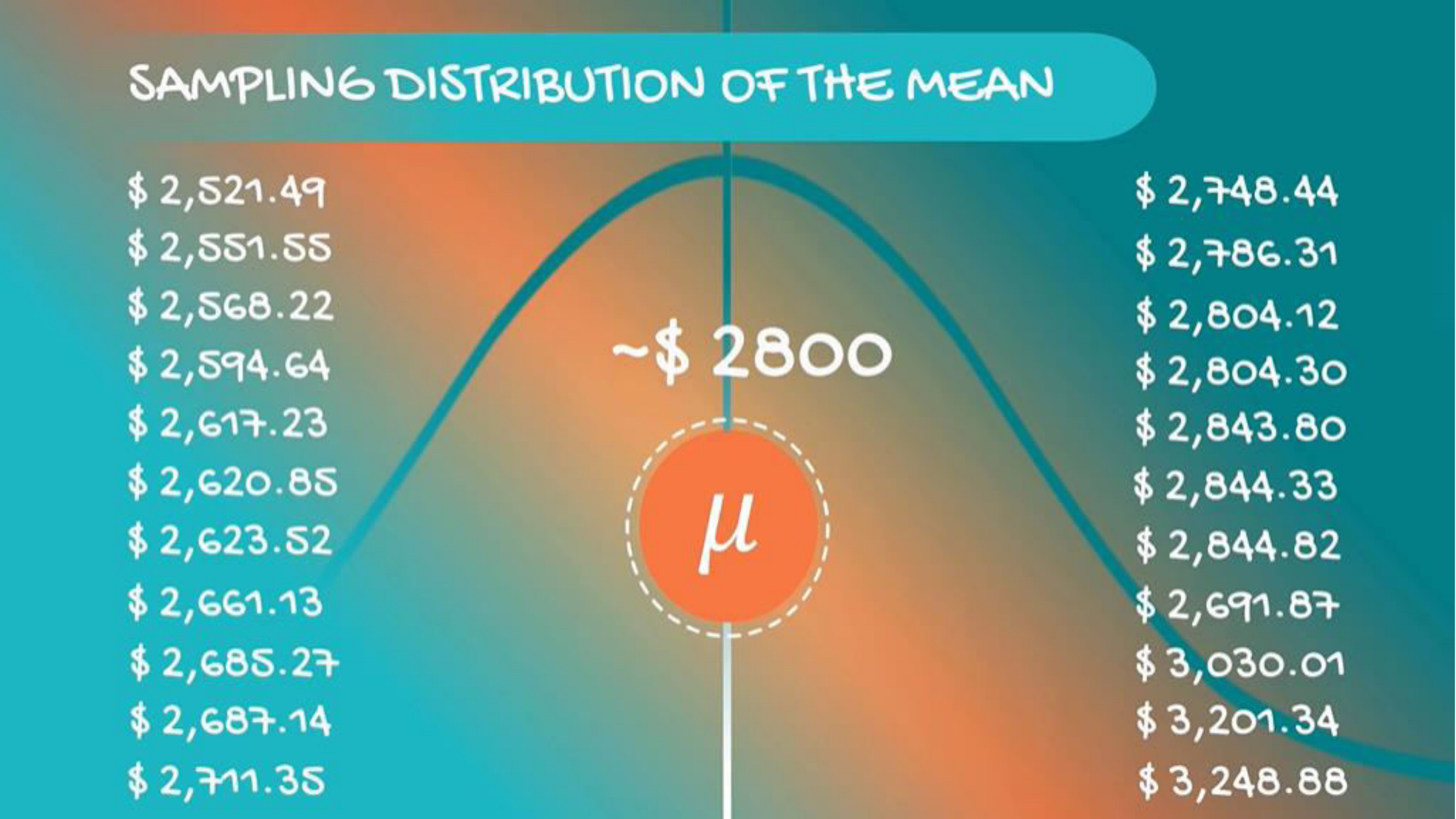
SAMPLING DISTRIBUTION



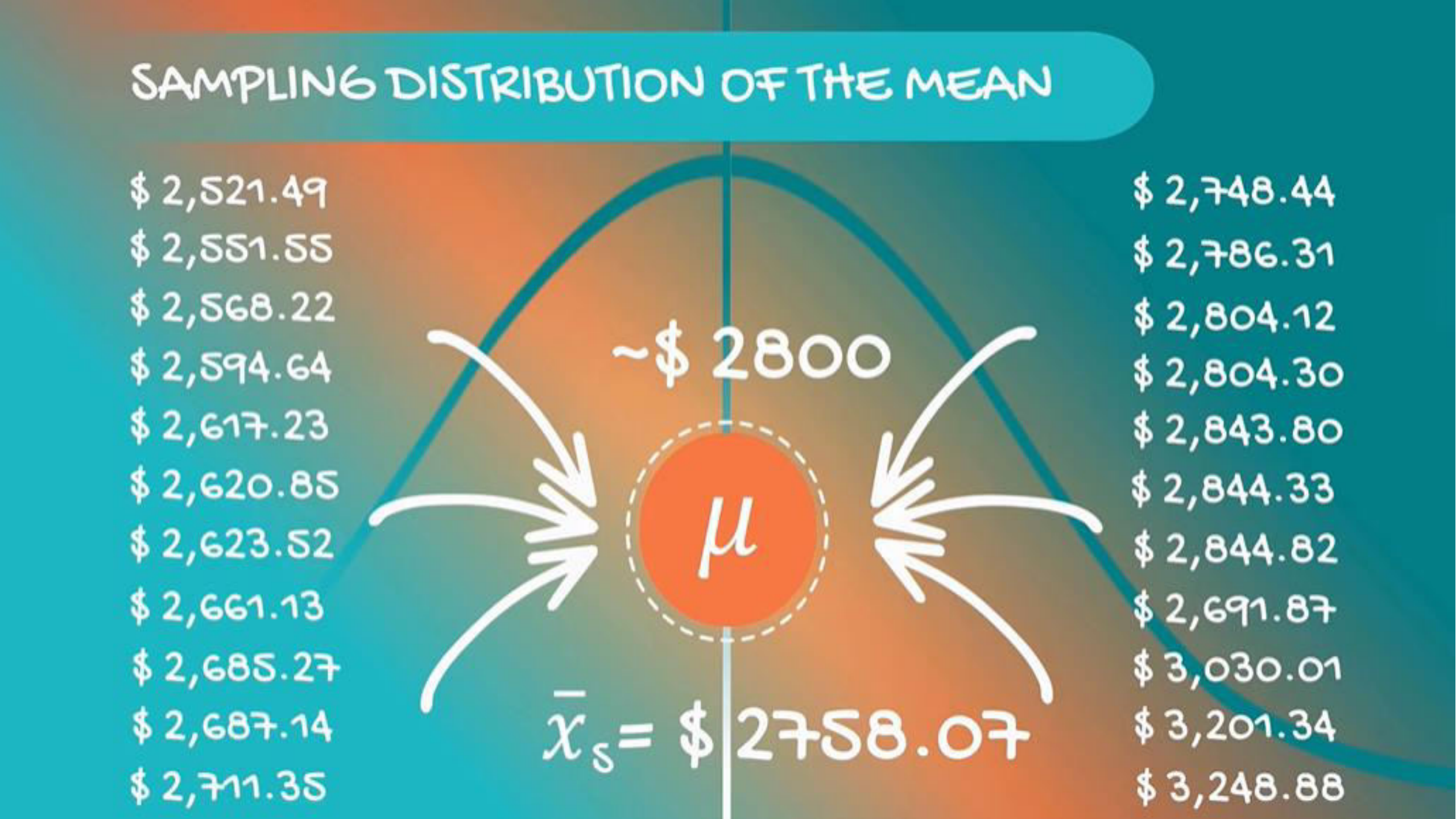
SAMPLING DISTRIBUTION OF THE MEAN



SAMPLING DISTRIBUTION OF THE MEAN



SAMPLING DISTRIBUTION OF THE MEAN



CENTRAL LIMIT THEOREM

Original distribution

$\mu \quad \sigma^2$

Sampling distribution

$N\left(\mu, \frac{\sigma^2}{n}\right)$

No matter the underlying distribution,
the sampling distribution approximates a Normal

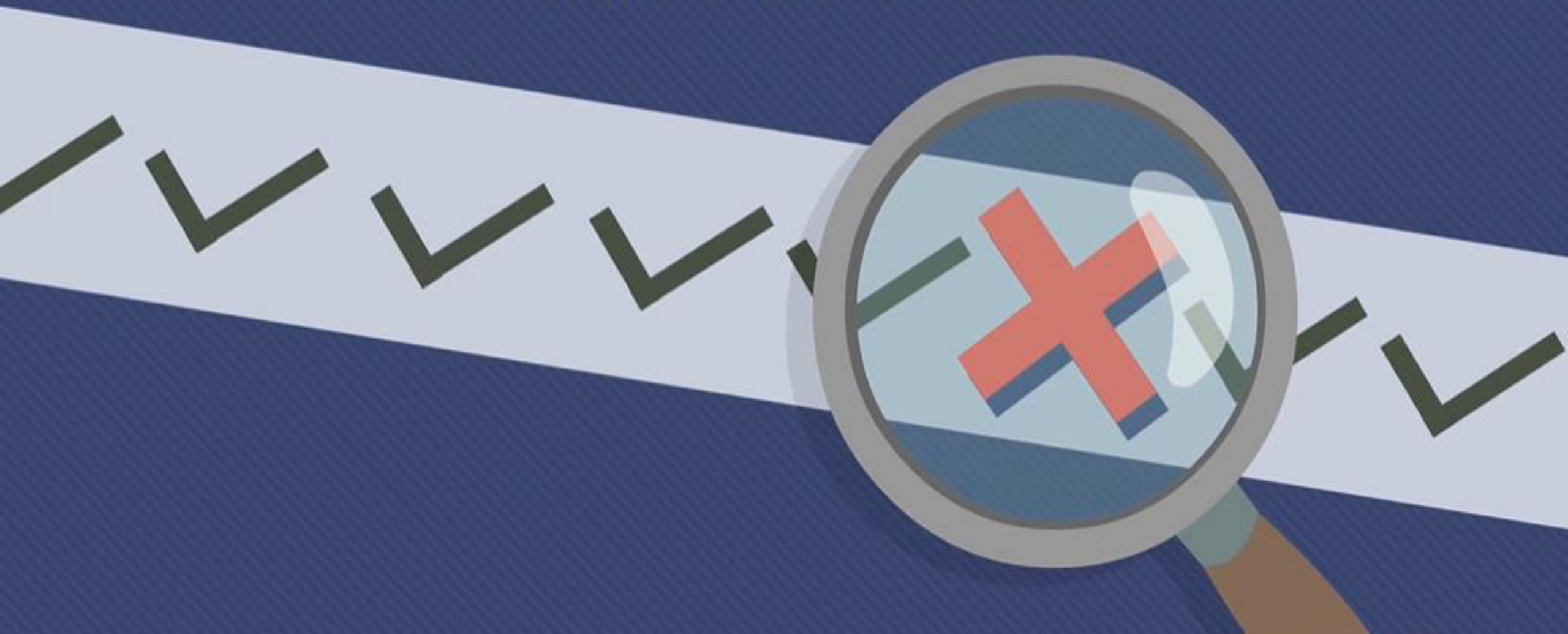
REASONS TO USE THE NORMAL DISTRIBUTION

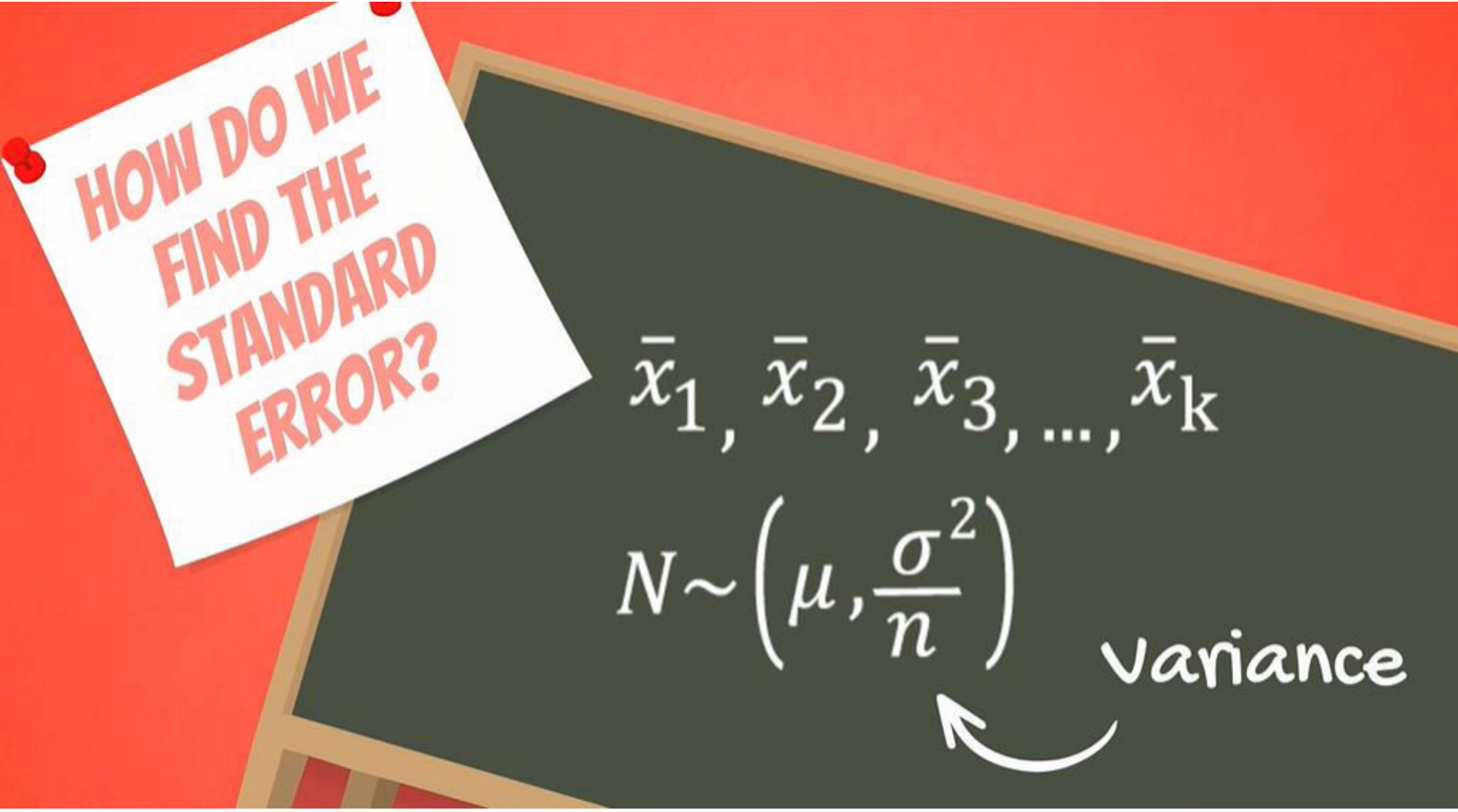
CLT allows us to perform tests, solve problems and make inferences using the Normal distribution, even when the population is not normally distributed

- They approximate a wide variety of random variables
- Distributions of sample means with large enough sample sizes could be approximated to normal
- All computable statics are elegant
- Decisions based on normal distribution insights have a good track record


STANDARD ERROR
THE STANDARD DEVIATION OF THE DISTRIBUTION
FORMED BY THE SAMPLE MEANS

SO HOW DO WE FIND THE STANDARD ERROR

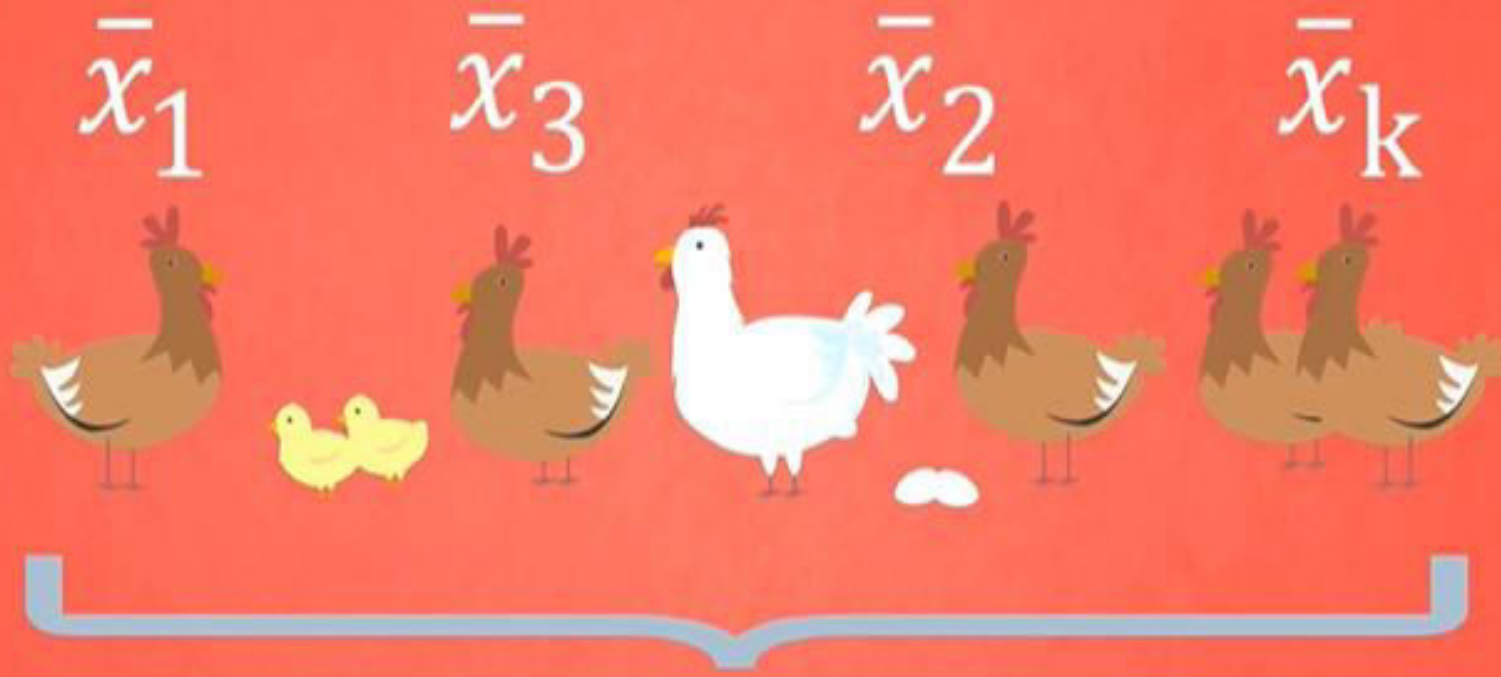



HOW DO WE FIND THE STANDARD ERROR?
$\bar{x}_1, \bar{x}_2, \bar{x}_3, \ldots, \bar{x}_k$
$N \sim \left(\mu, \frac{\sigma^2}{n}\right)$
variance

HOW DO WE FIND THE STANDARD ERROR?

standard deviation (of the sampling distribution) $= \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$

MEANING OF THE STANDARD ERROR

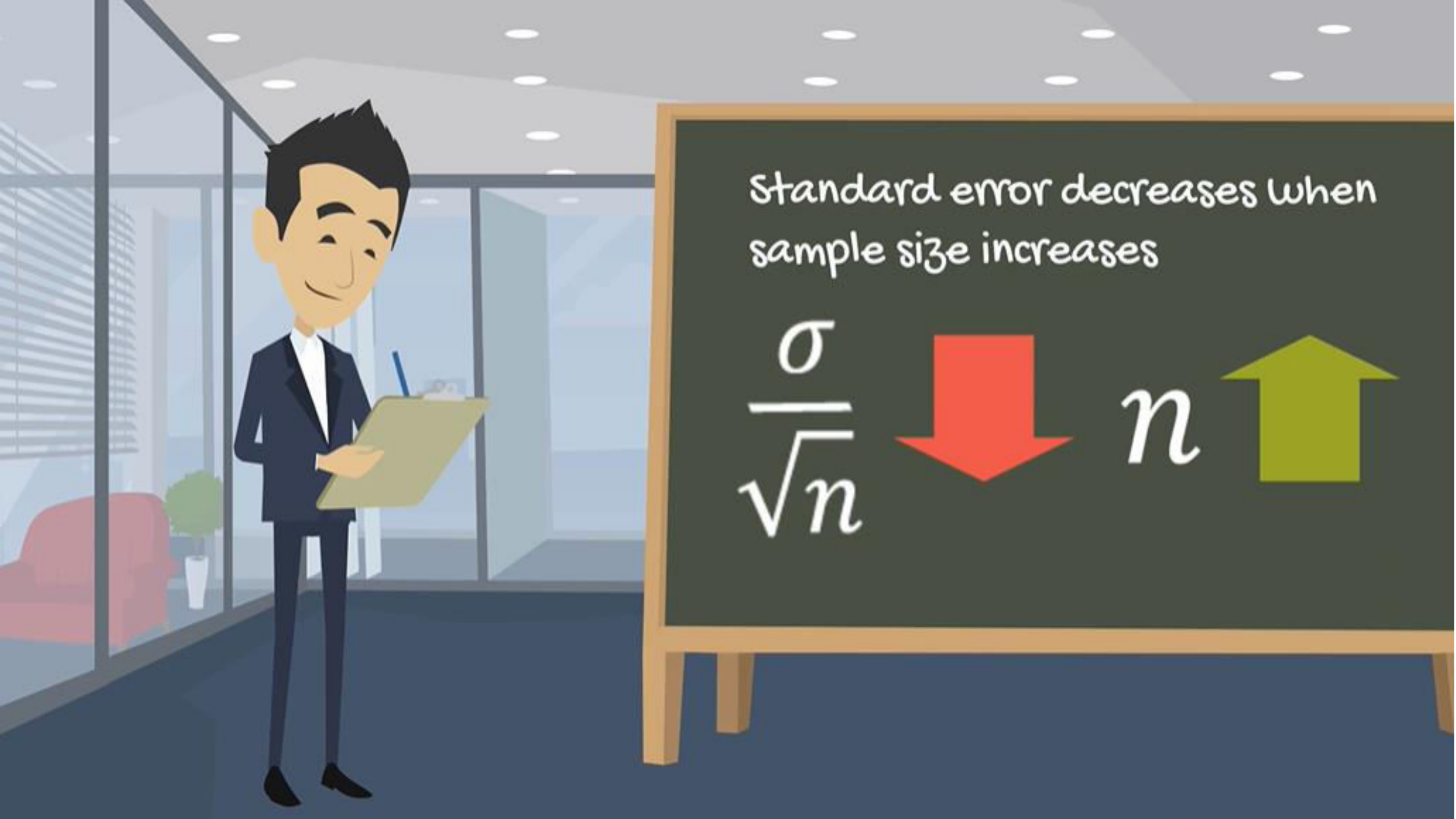


Like any standard deviation, it shows variability

WHY IS IT IMPORTANT?

Used in most statistical tests

Because it shows how well you approximated the true mean


Standard error decreases when sample size increases
$\frac{\sigma}{\sqrt{n}}$
n

Question 1:

How does sample size (*n*) affect SE (Standard Error)?

- ◯ As sample size increases, so does SE.
- ◯ As sample size decreases, so does SE
- ◯ It doesn't.
- ◯ None of the above.