WELLCOME

STATISTIC WITH PYTHON

EMBARKING ON A JOURNEY INTO DATA SCIENCE

YA MANON

THE CENTRAL LIMIT THEOREM

THE CENTRAL LIMIT THEOREM

In this section we'll cover **the central limit theorem** (CLT), which will allow us to apply the concepts we learned on the normal distribution to populations that follow any distribution

TOPICS WE'LL COVER:

CLT Basics

Standard Error

GOALS FOR THIS SECTION:

- *Understand the concept of a sampling distribution, and its relationship with the central limit theorem*

- *Identify the impact of the sample size on the normality & variability of the sampling distribution*

- *Calculate the standard error of a sampling distribution*

SAMPLING DISTRIBUTION OF THE MEAN

The **sampling distribution of the mean** is obtained by taking many samples from a population, calculating the mean for each, and plotting their frequencies
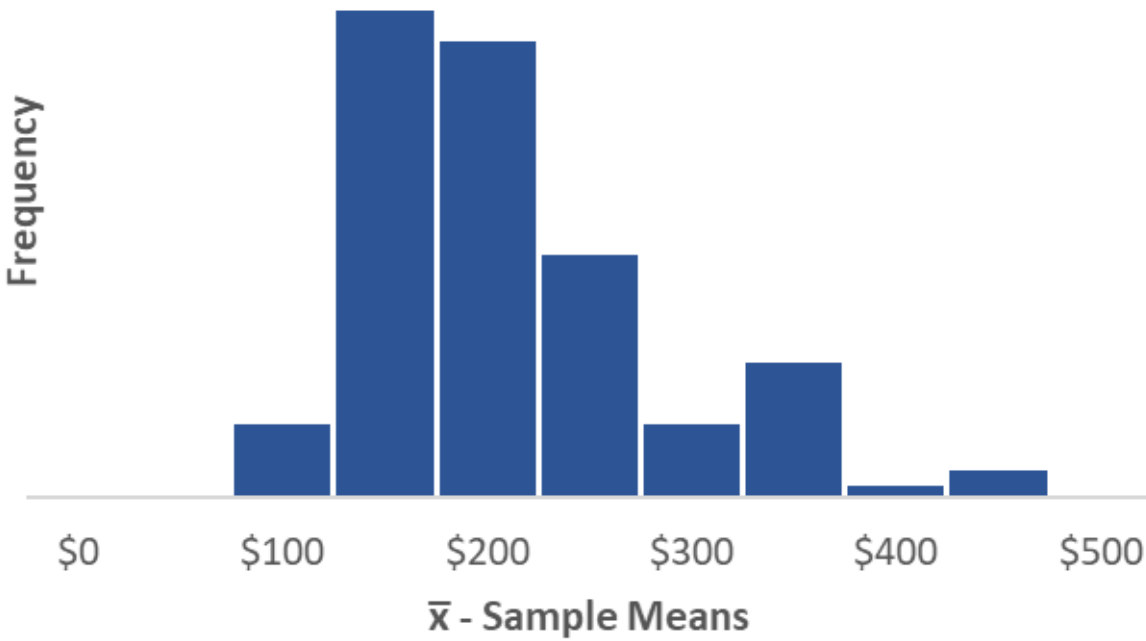
CLT Basics

Standard Error

EXAMPLE | *Daily Airbnb Rates in New York City*

Population

Price
$150
$95
$156
$259
$250
$59
$330
$85
$125
$285
$95
$171
$60

n>39,000

Sample (n=5) | *Mean*

1	$60	$305	$45	$272	$201	**$177**
2	$400	$322	$160	$233	$120	**$247**
3	$200	$174	$159	$356	$70	**$192**
4	$297	$50	$60	$55	$91	**$111**
5	$71	$112	$43	$41	$80	**$69**
6	$229	$345	$195	$774	$225	**$354**
7	$150	$90	$314	$80	$190	**$165**
8	$97	$170	$128	$225	$71	**$138**
...						
100	$312	$50	$99	$165	$213	**$168**

Sampling Distribution of the Mean (n=5)

Frequency

$\bar{x}$ - Sample Means

$0 $100 $200 $300 $400 $500

THE CENTRAL LIMIT THEOREM
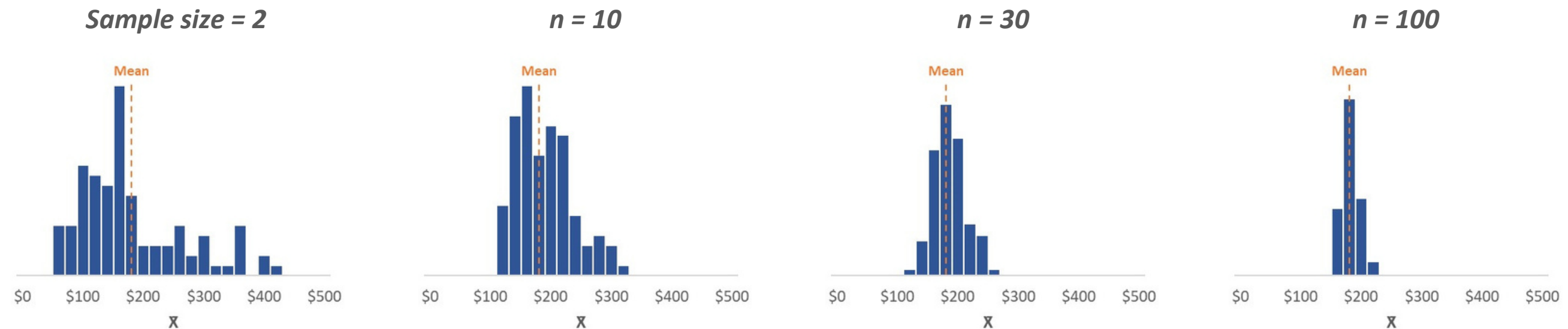
The **central limit theorem** states that the means of large enough samples of *any* population will be normally distributed around the population mean

- A sample size of **30** or more is typically required *(n > 30)*

CLT Basics

Standard Error

SAMPLING DISTRIBUTION (100 samples):

Sample size = 2   n = 10   n = 30   n = 100

HEY THIS IS IMPORTANT!

As sample size increases, the sampling distribution approximates a normal distribution

STANDARD ERROR

As you know, normal distributions are described by their mean & standard deviation

For the normal distribution of the sample means, the mean is the same as its population's mean, but the standard deviation is known as the **standard error**

- The standard error is the standard deviation of the sample means around the population mean
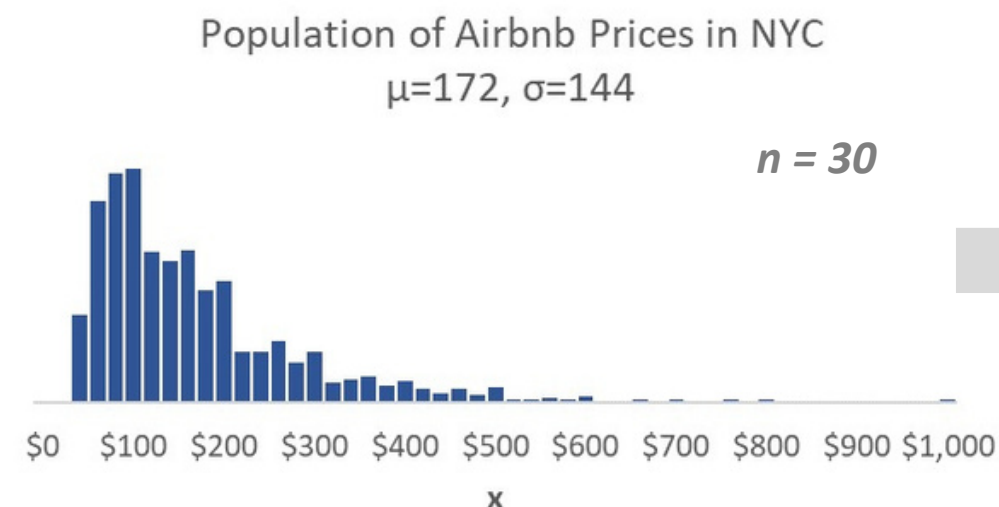
$$.SE =. \frac{\sigma}{\sqrt{n}}$$

To calculate the standard error, simply divide the standard deviation of the population by the square root of the sample size

⚠️ **HEY THIS IS IMPORTANT!**

As sample size increases, the standard error decreases

Population of Airbnb Prices in NYC
μ=172, σ=144

n = 30

$$SE= \frac{144}{\sqrt{30}}$$ **26.3**

Sampling Distribution of the Mean (n=30)
x̄ ~ N (172, 26.3)

You don't need all the actual samples!

$0 $100 $200 $300 $400 $500 $600 $700 $800 $900 $1,000
x

$50 $100 $150 $200 $250 $300
x̄