1.60

Tie-ins to Statistics

Statistics VS Characteristics

Sample

60% of 1000 people

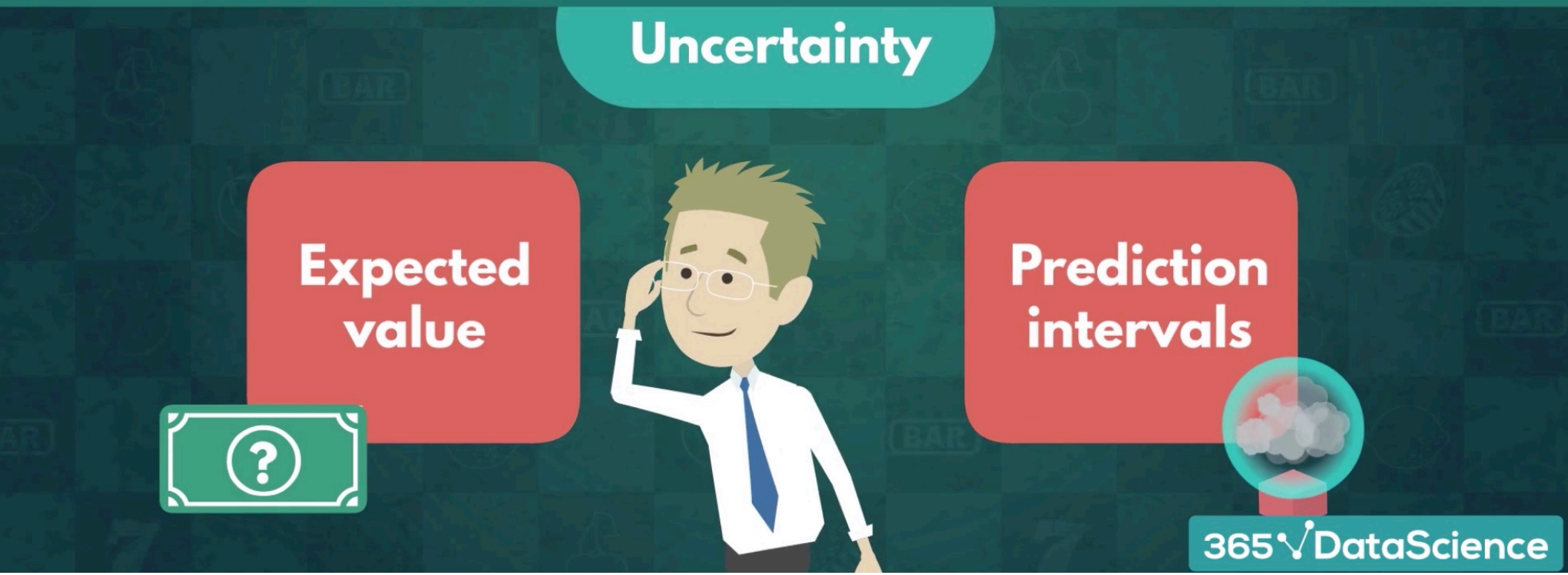
Population

365 DataScience

Statistics

Statistics focuses predominantly on samples and incomplete data

Fullscreen



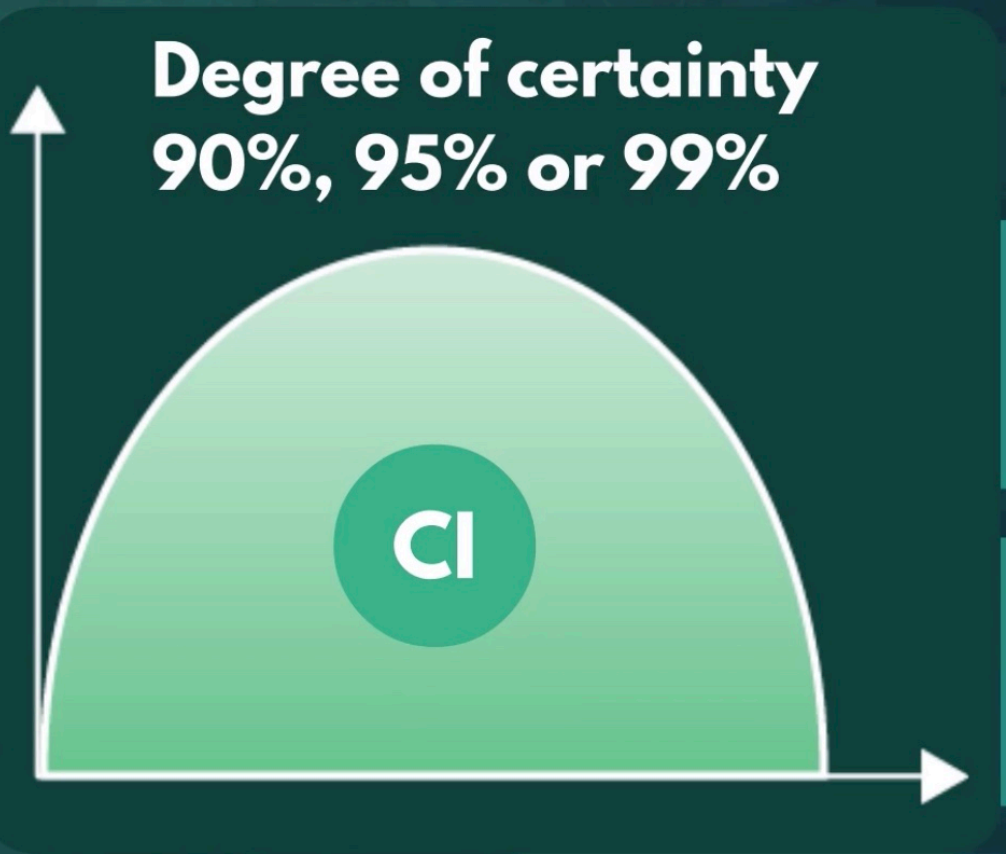
1.60

Statistics

Fullscreen

Experimental data

- Many useful concepts based on probability theory
- Express the likelihood of the population mean being within that interval

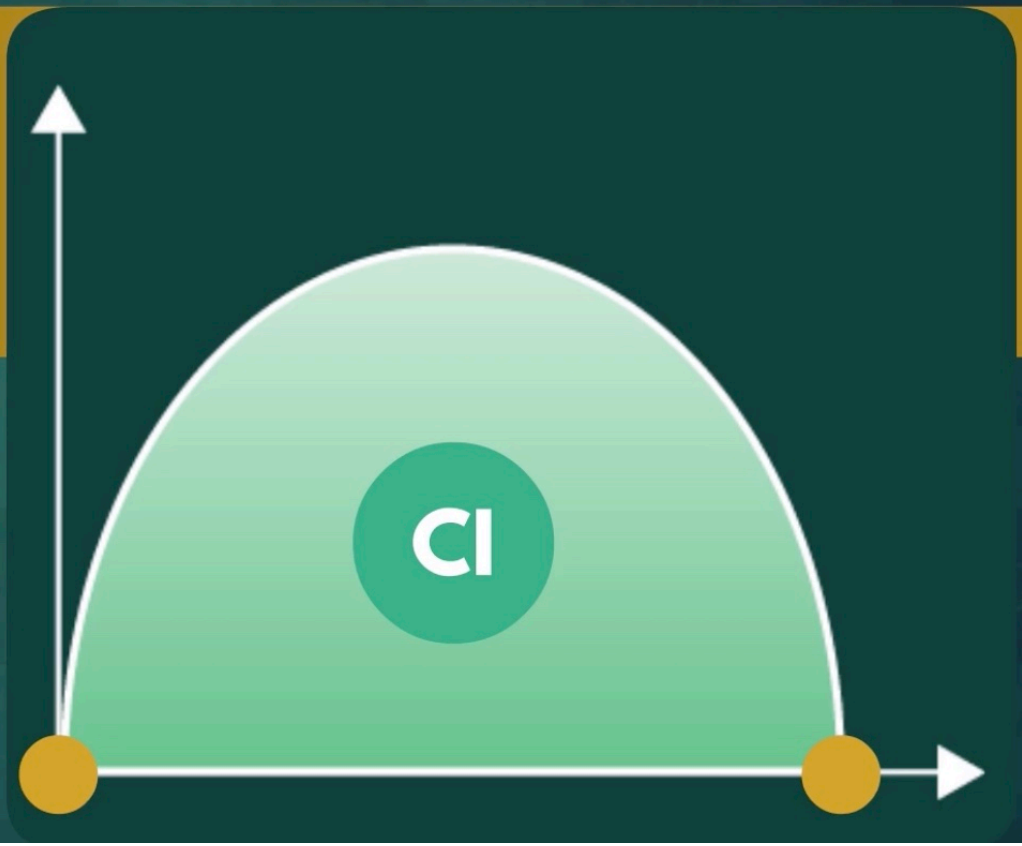


1.60

Confidence Intervals

Fullscreen

Approximate some margins for the mean of the entire population



1.60

Confidence Intervals

Fullscreen

To calculate these CIs we must know :

- Mean, variance and standard deviation

- A good understanding of probability is crucial

1.60

London Example

YELP

10 restaurants and pubs

Fullscreen

The Restaurant ★★★★

$\bar{x} = £22.5$

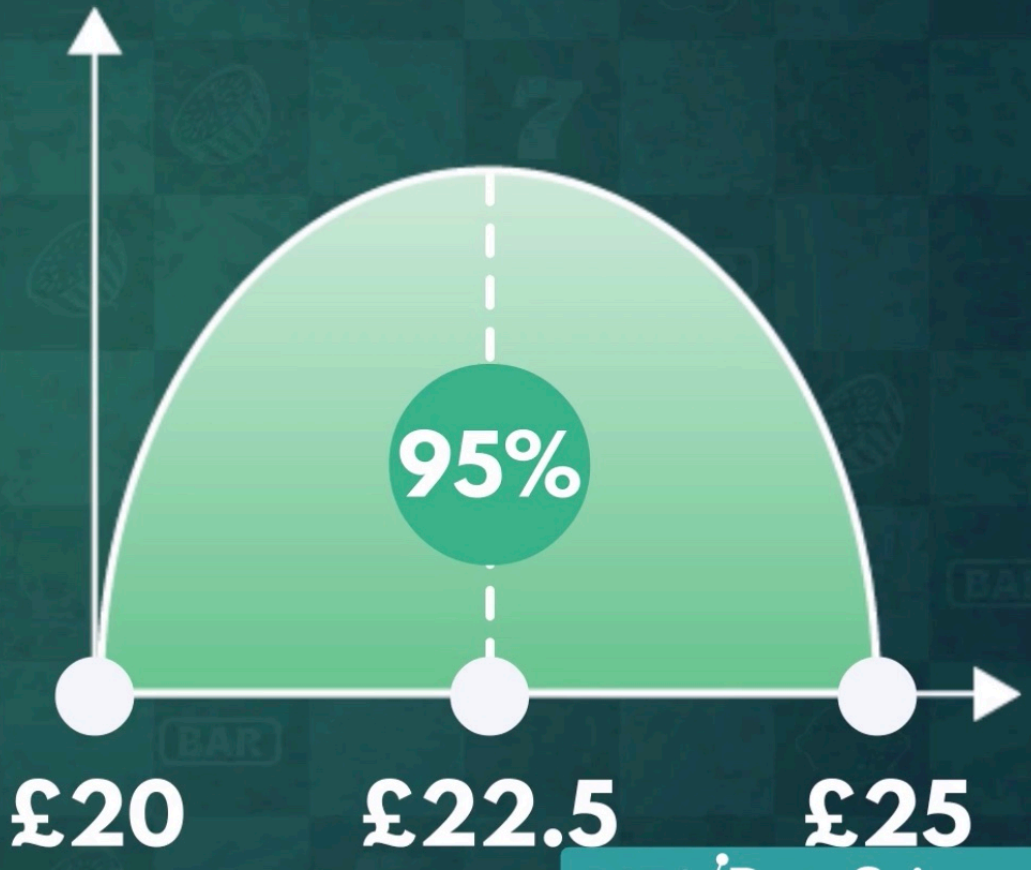
$E(x) = 22.5$

1.60

Fullscreen

London Example

CI = [20; 25]



1.60

Hypothesis Testing

Fullscreen

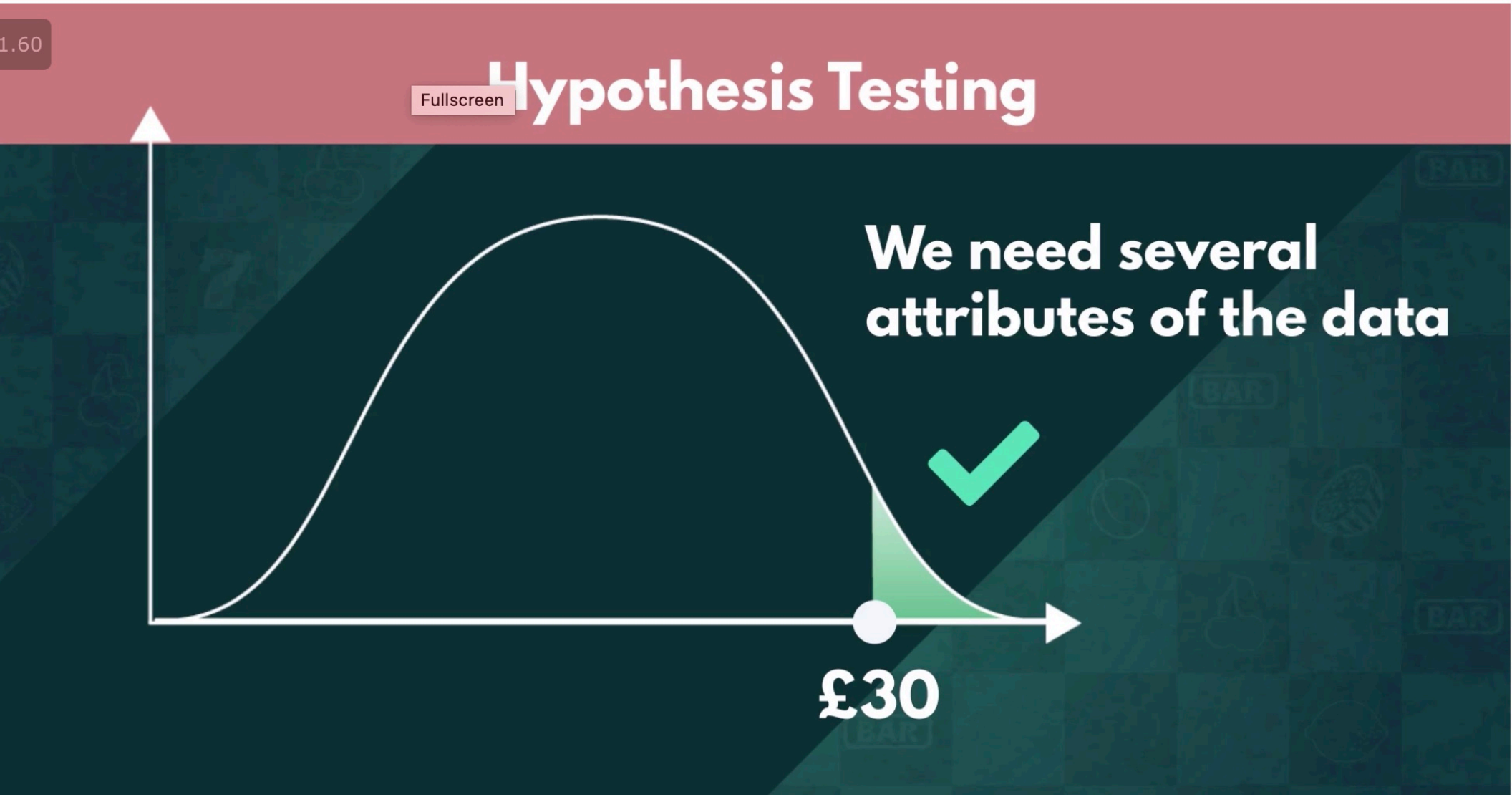
A hypothesis is an idea that can be tested

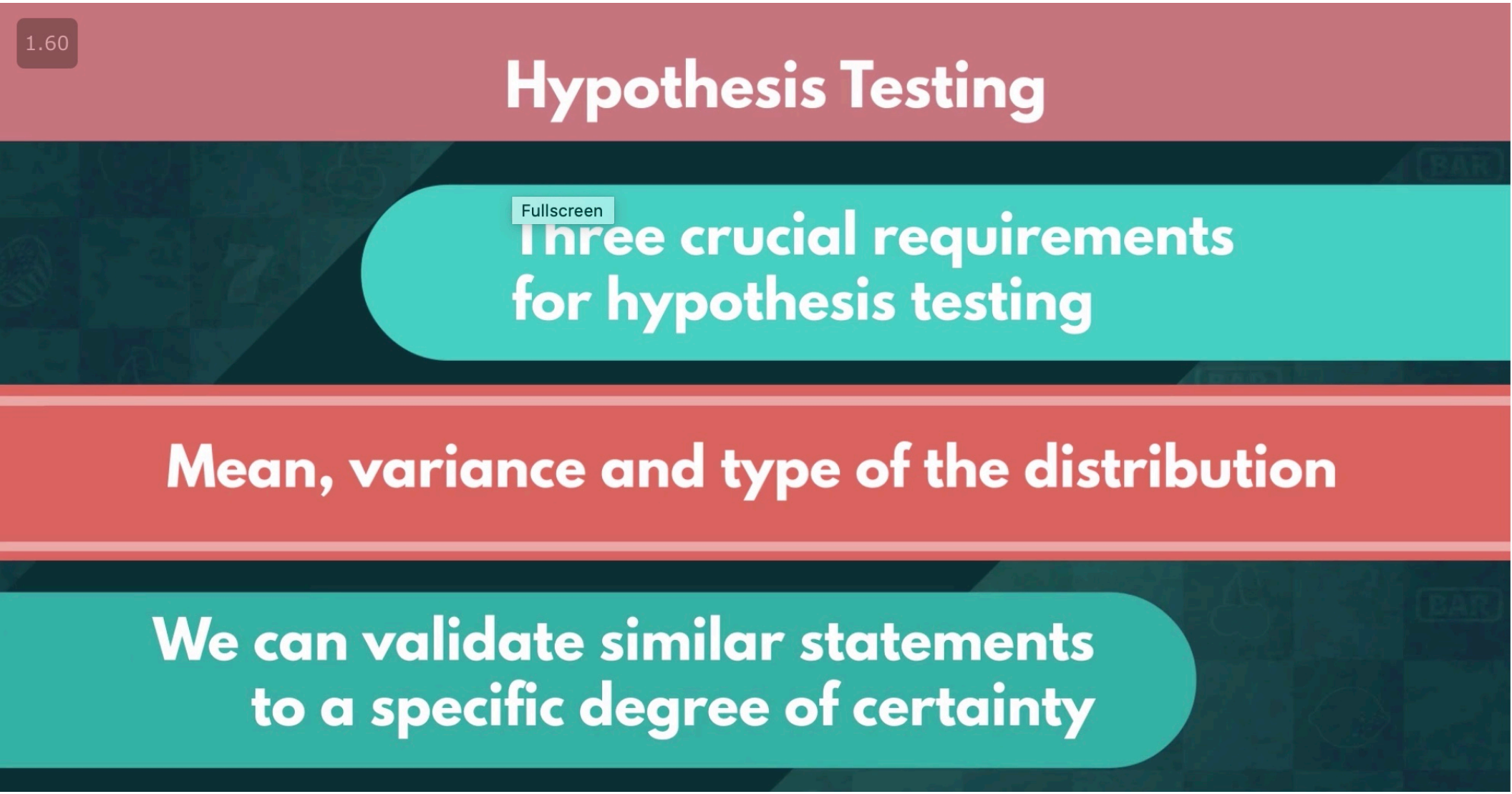
$H_0: P < £30$

1.60

Fullscreen

Hypothesis Testing



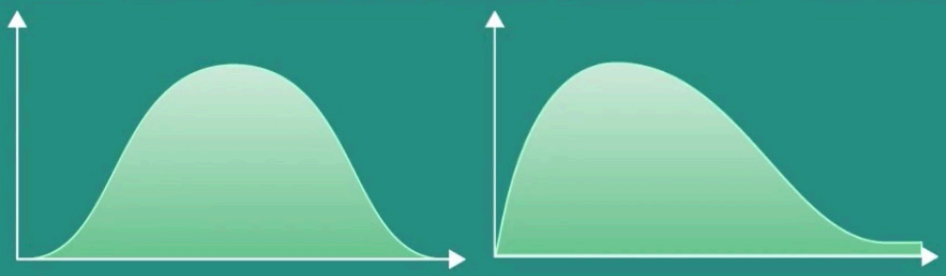

1.60
Hypothesis Testing
Fullscreen
Three crucial requirements for hypothesis testing
Mean, variance and type of the distribution
We can validate similar statements to a specific degree of certainty

1.60

Types of Distributions

- We are often provided sample data without knowing the type

- Determine type:
  - Shape



  - Characteristics μ, σ^2

1.60

Choosing a Distribution

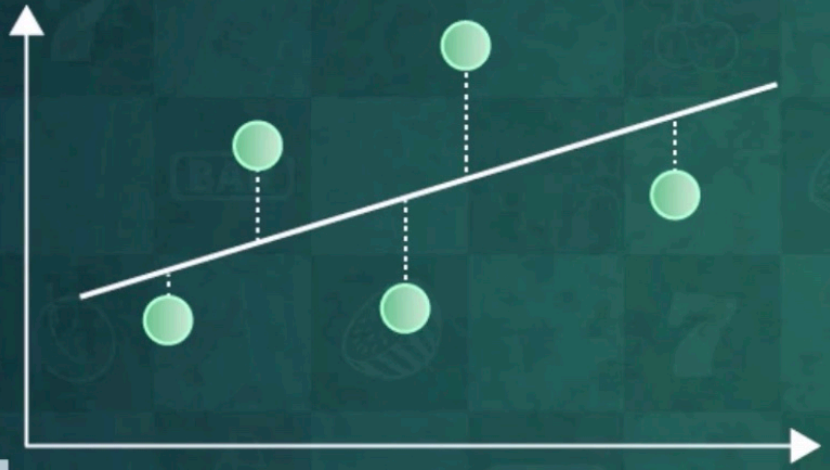
- The same logic is often applied in statistics
- Any distribution predicts a value for all points within our dataset
- The distribution anticipates the actual data point

A type of anticipated average value

1.60

Knowing the Type of a Distribution

Create different models



Computationally expensive

Computer software

MATHEMATICAL MODELING

1.60

Mathematical Modeling

An extension of statistics that data scientists deal with