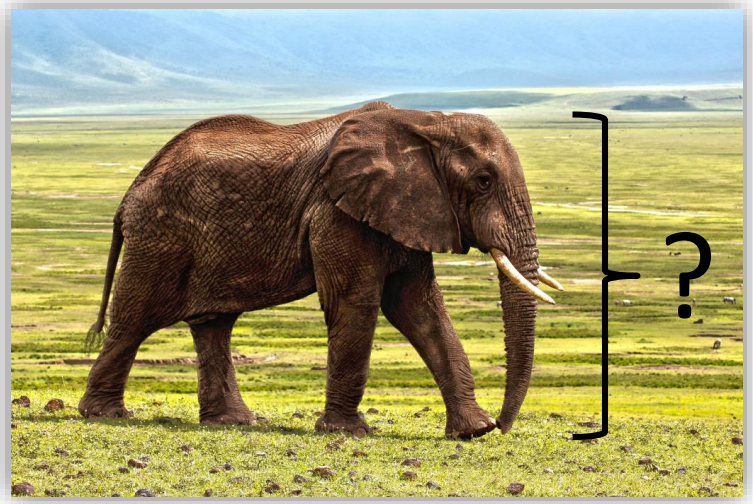
Statistics

Overview, Terms & Vocabulary

Statistics – Random Variables

- a random variable is an uncertain quantity/number.
- an outcome is an observed value of a random variable
- an event is a single outcome or a set of outcomes

3



3m

Discrete vs Continuous random variables

1, 2, 3, 4, 5, 6

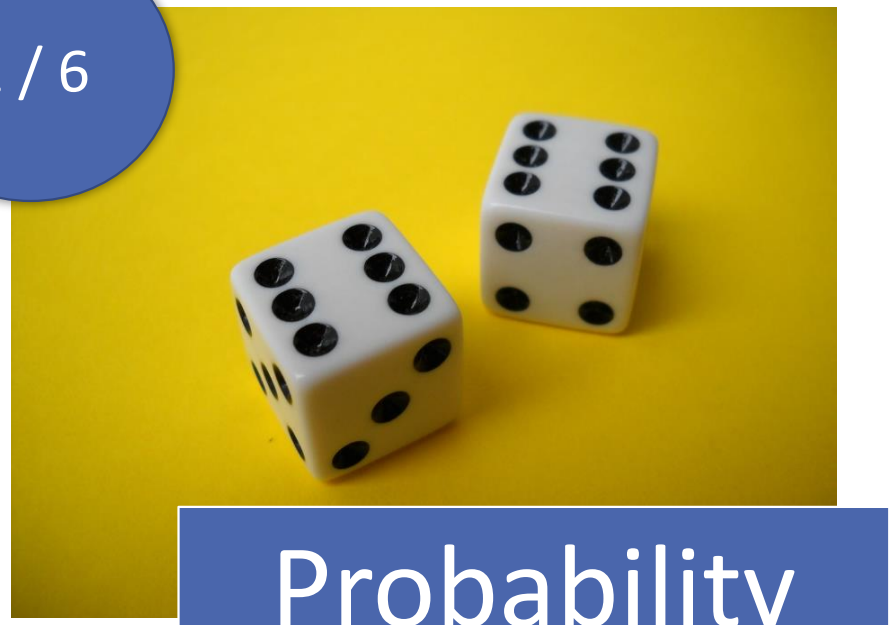
Number of possible outcomes can be counted (6).

3.128m, 2.854m, 2.945m, …

Number of possible outcomes is infinite.

Probability vs. Frequency

1 / 6



Probability

Likelihood that an event occurs.

Absolute Frequency

1: 1
2: 2
3: 1
4: 0
5: 1
6: 0

Relative Frequency

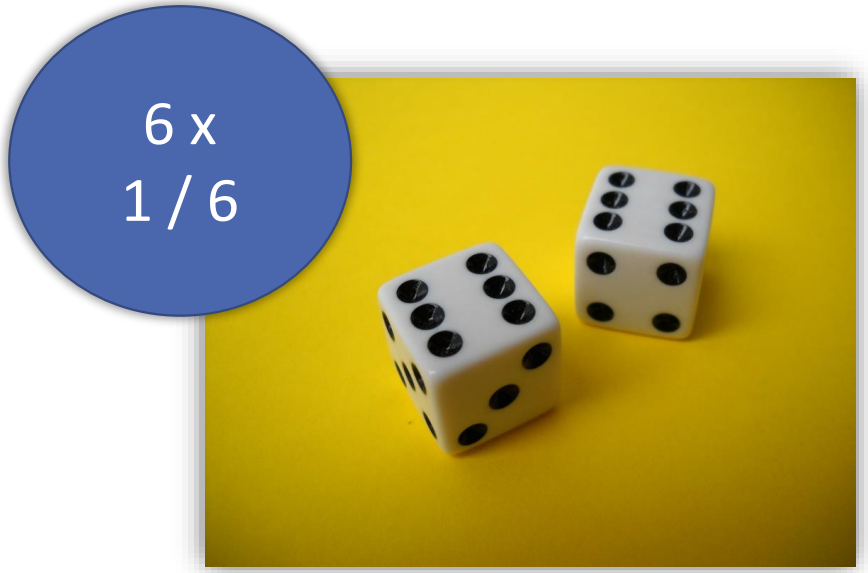
1: 20%
2: 40%
3: 20%
4: 0%
5: 20%
6: 0%

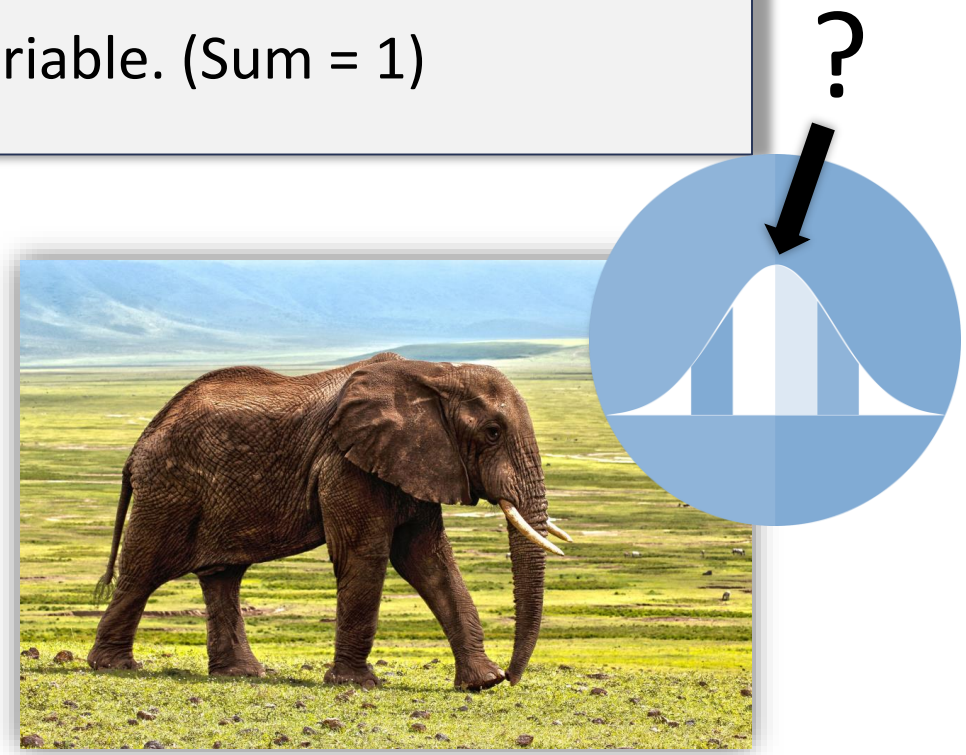
Frequency

Number of occurrences of an outcome / event
Rolling the dice 5x: (3, 5, 2, 2, 1)

Probability Distributions

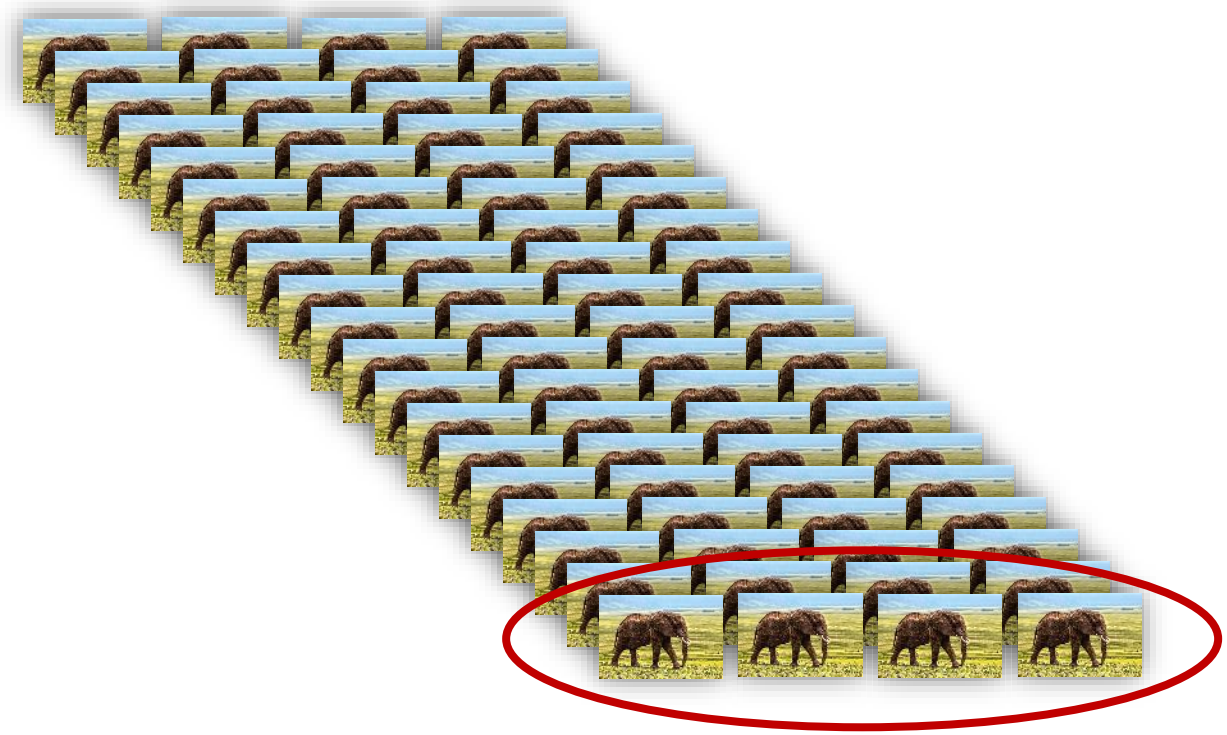
- a probability distribution describes the probabilities of all the possible outcomes for a random variable. (Sum = 1)

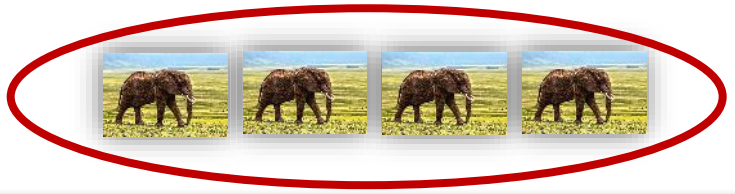




Population vs. Sample

Population (set of all possible members of a group)

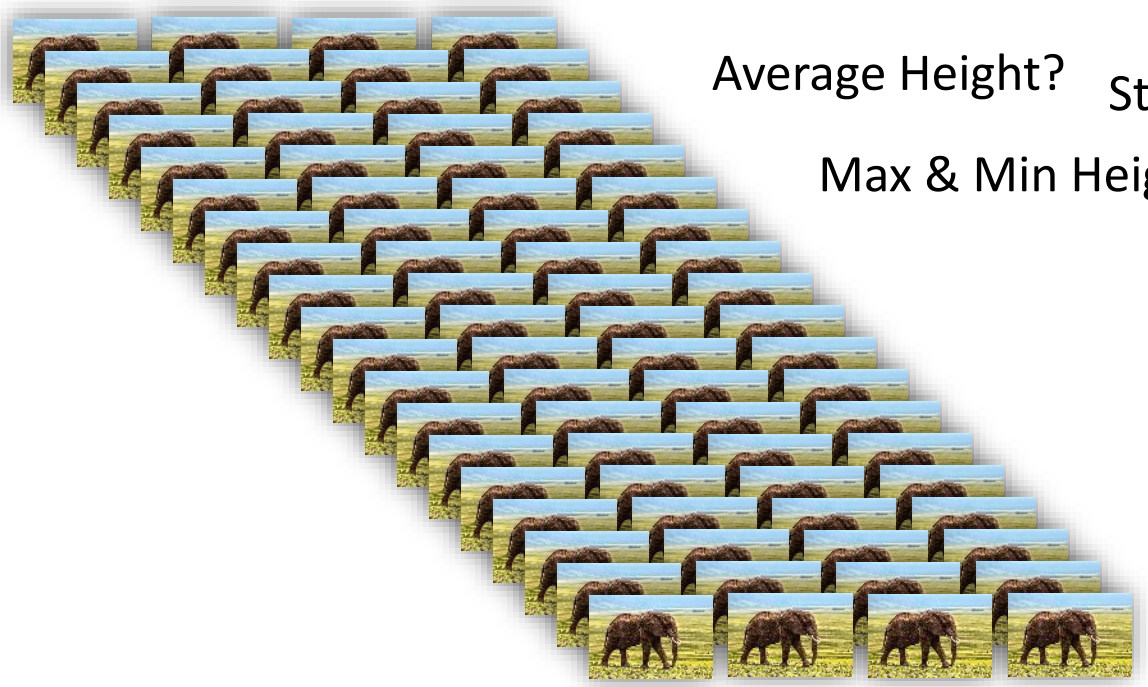




Sample (subset of a population)

Descriptive Statistics vs …

Population (set of all possible members of a group)



Average Height?

Standard Deviation of Height?

Max & Min Height?

Max & Min Height?

Standard Deviation of Height?

Average Height?

Sample (subset of a population)

… Inferential Statistics

Hypothesis Test

Population (set of all possible members of a group)

not available / observable

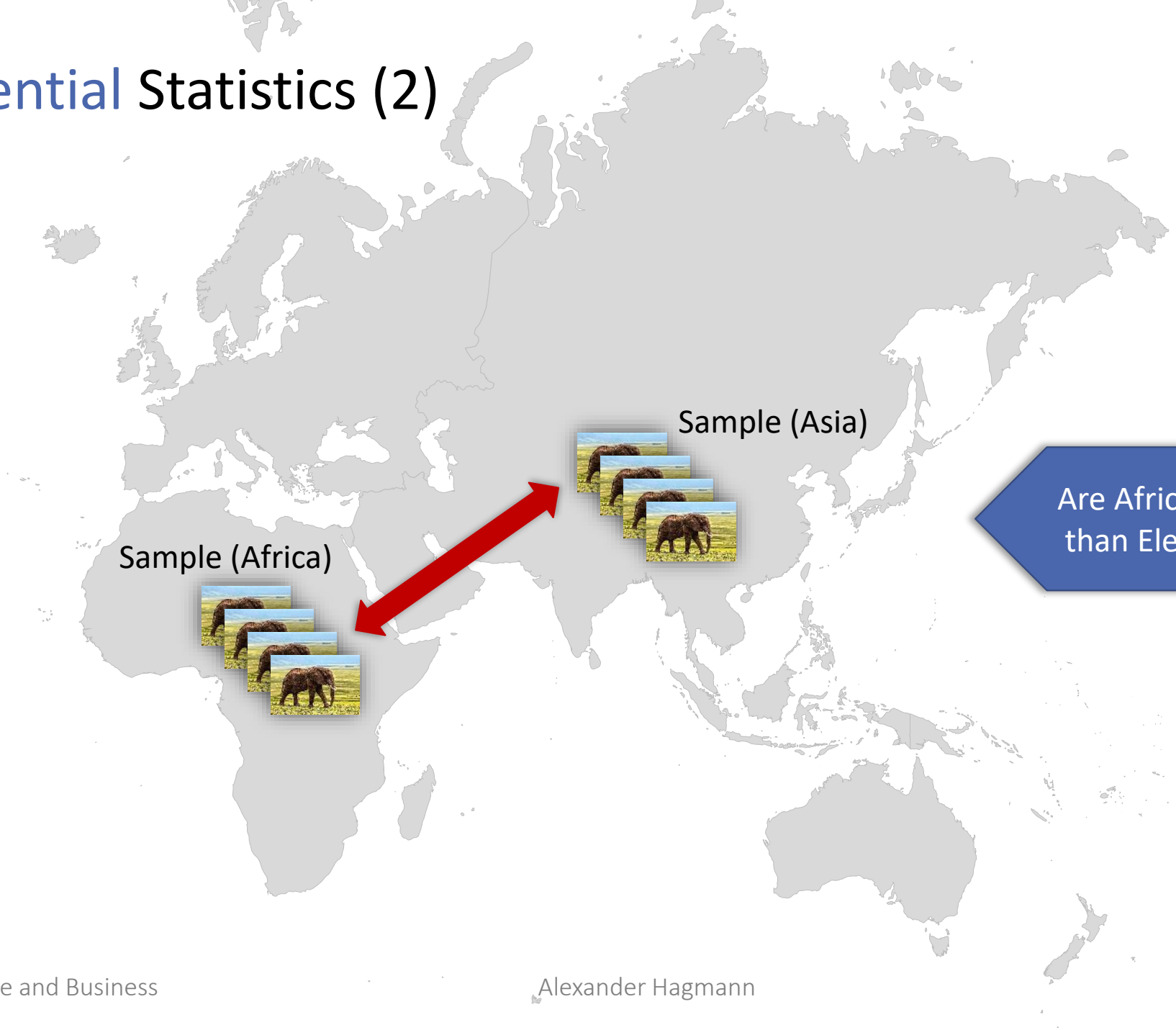
Estimate

What is the average Height of an Elephant (Pop)?

Sample (subset of a population)

… Inferential Statistics (2)

Hypothesis Test



Are African Elephants (Pop) taller than Elephants from Asia (Pop)?

Inferential Statistics & Regression

Hypothesis Test

Independent Variables

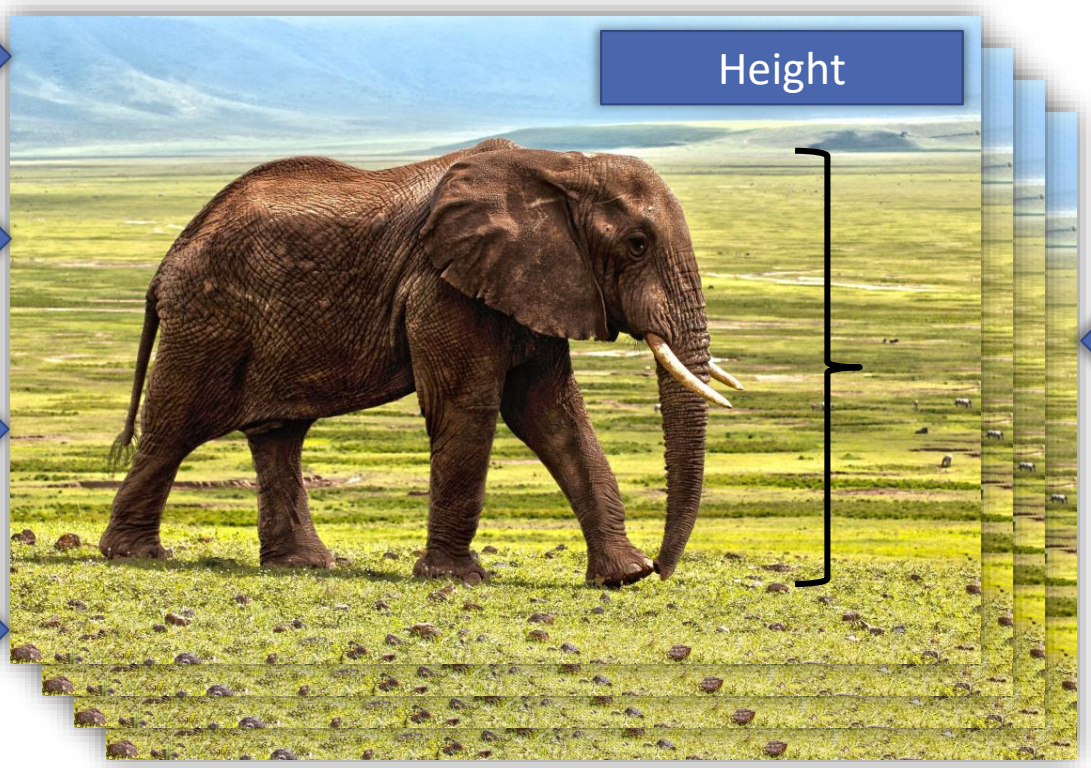
Dependent Variable

Age

Gender

Parents

$x_{4,5,6...}$

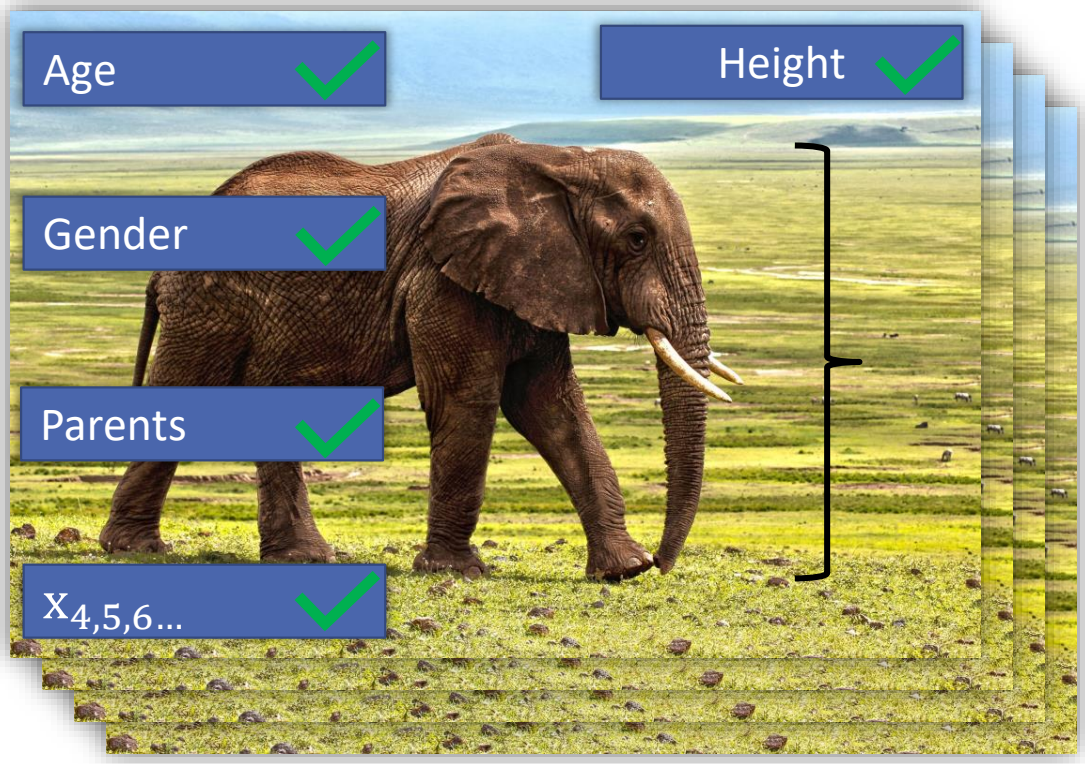


Which Factors significantly influence the height of elephants?

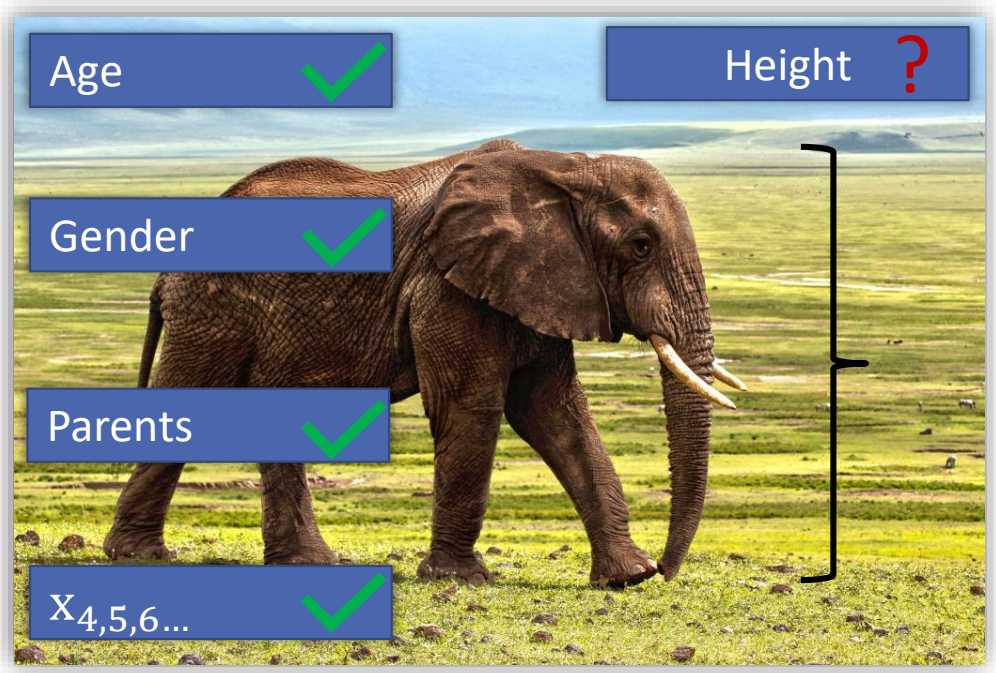
Machine Learning

Training & Updating ML Models

Training Set



Estimate / Forecast



New Observation

Best Estimate/Forecast for the unknown variable?