

STATISTICS

SCIENCE OF DATA

{ COLLECTION
ANALYSIS. ←
PRESENTATION
INTERPRETATION

ANALYSIS { INFORMATIVE
NON-INFORMATIVE

MODEL - LESS COMPLICATED REPRESENTATION
OF SOMETHING
THAT RETAINS ITS KEY FEATURES

SAMPLING : MOST DATA WE MEET
WILL BE INTERPRETABLE

AS A SAMPLE FROM SOME
POPULATION

We are usually interested in ESTIMATING
certain numerical characteristic(s) of the POPULATION,
but to determine this directly is usually "difficult".

We thus hope that we can (a) obtain a REPRESENTATIVE
sample and (b) use it to ESTIMATE the PARAMETER(S)
OF INTEREST.

STATISTICAL MODEL

A probabilistic/mathematical model for a data generating process.

- SAMPLE SPACE : Set of all possible datasets one could obtain in a given context.
- EACH SUBSET (of interest / TYPE OF SAMPLE)
i.e. EACH EVENT is assigned a PROBABILITY (simply a number in $[0, 1]$) in such a way that the collection of all such numbers "behave like proportions".

- Such models allow us to

- describe what has (or might have) happened
- predict or OR (with the aid of a computer)
- Simulate what might happen in the future.

CLASSICAL PROBABILITY

We shall use the term "classical probability" to describe a scenario where we have a sample space with

- a finite number of
- EQUALLY LIKELY outcomes.

there used to be some debate over what "equally likely" means, but we assume this is not an issue.

↙ "fair/balanced"

Example: Flip a coin 3 times in sequence. We can represent the sample space as

$\{ HHH, HHT, HTH, THH, HTT, THT, TTH, TTT \}$

there are $8 = 2^3$ outcomes.