STAT 231

Roadmap

- Introduction to Statistical Inference:
  - Some definitions
  - Types of inference problems
  - STATISTICAL MODEL
  - Examples
  - ~~Lik~~ Estimation using the method of maximum likelihood

DESCRIPTIVE STATISTICS	STATISTICAL INFERENCE
describe the main properties of the data, its shape, centre, variability, etc.	analyzing the properties of our population of interest using a sample

INDUCTIVE SCIENCE : small to the big / special to the general.

vs

DEDUCTIVE SCIENCE : General to the specific: AXIOMS $\longrightarrow$ THEOREMS

THREE MAJOR TYPES OF INFERENCE PROBLEMS

- <u>ESTIMATION</u>: We want to "guess" the attributes of a population from our collected sample.

- HYPOTHESIS TESTING

- PREDICTION PROBLEMS

<u>Example</u> : ~~To~~ Average income of all recent immigrants in the K-W area : μ

We use a sample of $\{y_1, \ldots y_n\}$ and try to estimate μ.

We are trying to estimate the unknown attributes of the population

- Proportion of US College Educated voters who plan to vote for Trump

HYPOTHESIS TESTING PROBLEM

A hypothesis is a claim made about some attribute (unknown) of the population

Draw a sample from the population and "test" whether the claim is "reasonable".

Example: (i) Paul the Octopus.

Test whether Paul is Psychic.

(ii) Test whether Canadians are better in Jeopardy than Americans

---

PREDICTION PROBLEMS

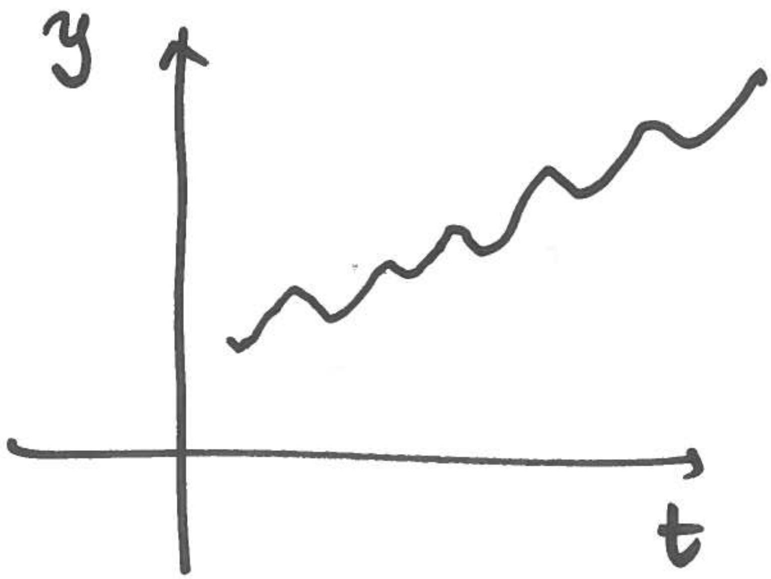
- We want to predict a future value of an population attribute based on past observations

FORECASTING ⟶

TIME SERIES ANALYSIS

$y_1, \ldots y_n$ = Stock prices in year t

$\hat{y}_{n+1} = ?$



Time series $\left\{ \begin{array}{l} \text{Trend} \\ \text{Seasonality} \end{array} \right\}$

STATISTICAL MODELLING

Definition: A statistical model is the identification of the distribution from which your data set, and the unknown attribute of interest is a parameter of the distribution.

NOTATIONS

$\{y_1, \ldots y_n\} \longrightarrow$

$y \longrightarrow$ Data point, a known #

$\underset{\sim}{Y} \longrightarrow$ random variable

$\mu, \sigma, \longrightarrow$ unknown constants.

Data points y_i, can be thought of as outcomes of a random experiment from a r.v. Y_i

<u>Example</u> Toss a coin with probability of Head $= \pi$

$$y = \left.\begin{matrix} 1 \text{ if } H \\ 0 \text{ if } T \end{matrix}\right\}$$

STAT MODEL

$$Y = \left\{\begin{matrix} 1 & \text{wp } \pi \\ 0 & \text{wp } 1-\pi \end{matrix}\right\}$$

π = Trudeau's approval rating

n y.

200 people 6. 120

\# of successes

$y = 120$.

y's a~~re~~ are drawn from a

$$\boxed{Y \sim Bin(200, \pi)}$$

Income problem

To Interested in → μ = av. salary of an immigrant in K-W

$\{y_1, \ldots y_n\}$ independently

unknown

$$\boxed{Y_i \sim N(\mu, \sigma^2)}$$

MODEL.

μ, σ are not random variables but unknown Constants

Jeopardy problem

$\{y_1, \dots y_n\}$

y_i = # of shows the i^{th} contestant appeared in

$\{1, 2, 1, 3, 1\}$ — # of trials before the 1^{st} failure.

π = Prob that a Canadian will win Jeopardy

Test whether $\pi > 1/3$ or not

$Y_i \sim \text{Geom}(\pi)$