Lecture 6: Introduction to Statistics

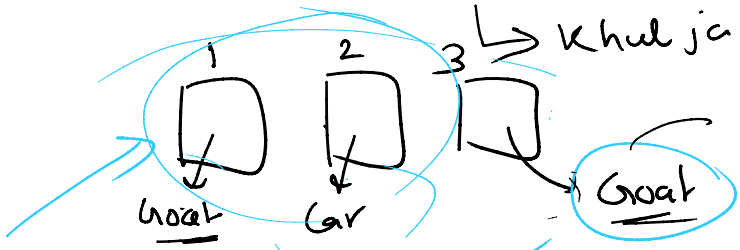
25 July 2022 17:40

{ SK-learn, Pandas, Mat plotlib }.

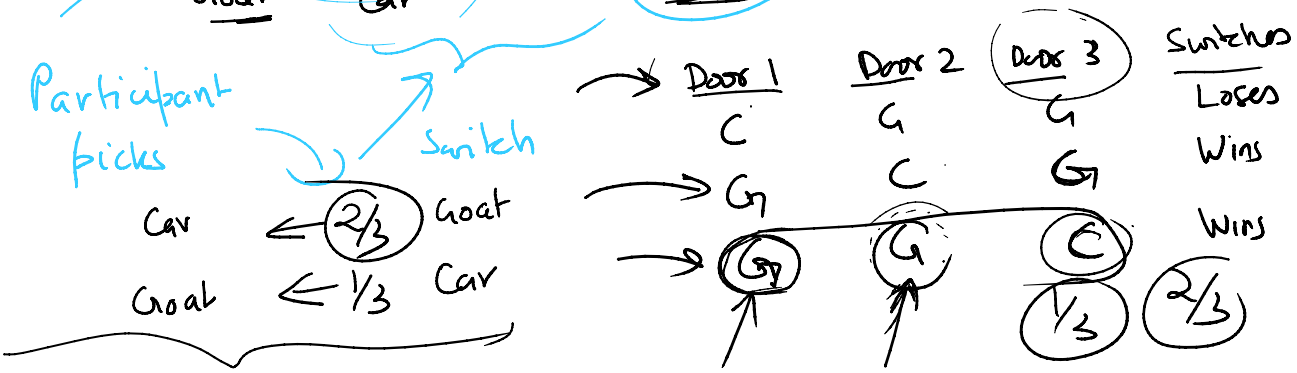
Linear Algebra Optimization Python Statistics

* Let's make a deal → Monty Hall Problem

↳ Khul ja sim sim



Optimal strategy is to always Switch



* Fair Coin → H, T 1/2

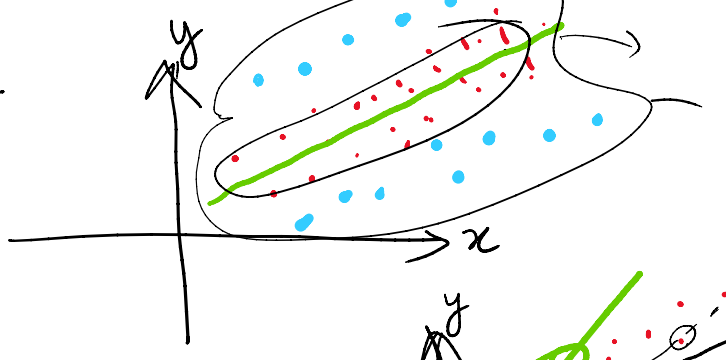
HHH - - - H 1/2^n

1/2^10 HTHT - - T 1/2^10

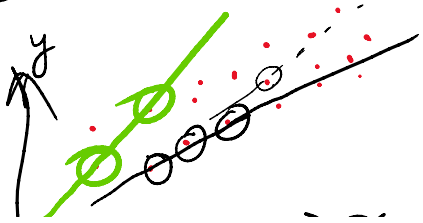
T T T - - - T 1/2^10

HHTTTHTTH

* Descriptive Statistics



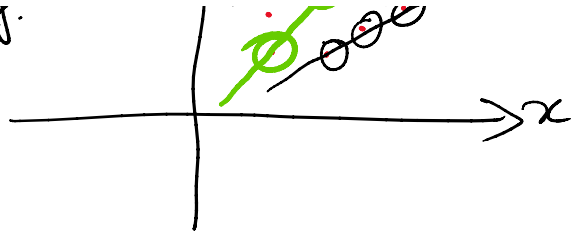
* Sampling:



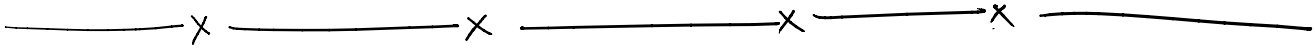
→ How to sample wisely?

→ What is the minimum number of samples needed for a "good" estimate?

* sampling.



needed for a "good" estimate?
↳ Sample size



Data: Most valuable entities
↳ Types of data and scales of measurement

Data at the highest level can be broadly classified into two categories.

Qualitative (Categorical)
about qualities, information that can't be measured
{ like color of eye
type of car one drives
type of business owns
↳ Categorize them "Yes" or "No" answers

Quantitative (Numerical)
Defines information that can be counted/measured
↳ height, width, temperature, speed, etc.

House price prediction
# of rooms → Quant
Area → Quant
Type of neigh. → Qualitative

Quantitative

A bookshelf with two books and two cms tall
↳ color is red → Qualitative

Quantitative data can further be broken into two types

Discrete
# of books/pets
(cannot be divided into individual parts, fractions, decimals)
↳ Something that can be counted

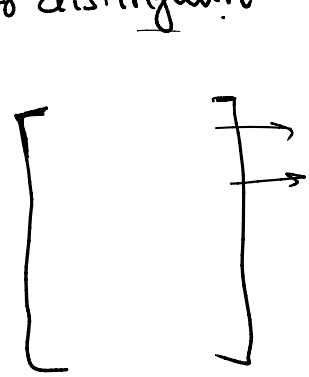
Continuous
Height of the bookshelf

that can be counted

(cannot be divided into individual parts, fractions, decimals)

Something that can be measured

Easy way to distinguish



Types of data

Quantitative (Numerical)

- Discrete
- Continuous
  - Interval
    Numbers with known differences between variables (such as time)
  - Ratio
    Numbers that have measurable intervals where differences can be determined, such as height, weight

While interval and ratio data can both be categorized, ranked, and have equal spacing b/w adjacent values, only ratio scales have a true zero

Ex: Temperature in Celcius or Fahrenheit

T_1 20°C
T_2 40°C
20°C

$T_2 = T_1 + 20$

$T_2 \neq 2T_1$

Celcius or Fahrenheit → Interval

vs

Qualitative (Categorical)

- Nominal
  Data used for naming variables, such as color of the eye
- Ordinal
  Data used to describe order. such as
  Happy = 1
  Neutral = 2
  Unhappy = 3

↙ ← → Interval $T_2 \neq 2T_1$

vs

Kelvin 0K — -273°C → T_1 200K, T_2 400K, $T_2 = 2T_1$

↙ Ratio

↓ there is absence of thermal energy

0°C → 32°F

$\frac{T_2}{T_1} = 2$ (10°C → 20°C →) $\frac{18}{5}$ $\frac{°C}{5}$

$T_1 - 0 = T_1$

Properties and Scales of Measurement

Scales of Measurement is how variables are defined and categorized.

Psychologist Stanley Stevens developed four common scales of measurement

↳ Nominal
Ordinal
Interval
Ratio

↳ Properties (associated with each scale of measurement)

- ↳ Identity → each value having a unique meaning
- ↳ Magnitude → Ordered Relationship to one another
- ↳ Equal Intervals → Equally spaced difference b/w data points along the scale
- ↳ Minimum Value of Zero → Scale has a true zero.

The Four Scales of Measurement

Nominal Scale: → defines identity property of data

No form of numerical meaning. Data can be placed in categories but cannot be multiplied, divided, added...
It's also not possible to measure differences b/w data pts.

... the order: "Cold, warm, hot, very hot"

It's also not pos...

- Nominal with order: "Cold, warm, hot, very hot"
- " without " : Male and Female
- Dichotomous : "Yes" and "NO"

Ordinal Scale of Measurement:
↳ Data is placed in a specific order
While each value is ranked, there is no information that specifics what differentiates the categories from each other
↳ Values can't be added or subtracted

Interval Scale of Measurement:
↳ Properties of nominal and ordered data, the difference b/w data points can be quantified.

✓ They can be added to or subtracted from, but cannot be multiplied or divided.

Ex. 20°C is not 10°C multiplied by 2.

↳ This scale is also characterized by the fact that number zero is an existing variable.

Ratio Scale of Measurement: It has all four properties
↓
scale has a 'true zero'.
