

17/02/2021



# MA2040: Probability, Stochastic Process & Statistics.

\*

INSTRUCTORS: Prof. S.H. Kulkarni

shk@iitpkd.ac.in

Prof. Ashok Kumar

ashokm@iitpkd.ac.in

\*

TAs : Abhinav Thakur

142002002@smail.iitpkd.ac.in

Shree Ganesh Sharma M S

142002015@smail.iitpkd.ac.in

\*

SLOTS : G1 Wed 1130-1230 hrs

T4 Thu 1400-1530 hrs

- Evaluation: 1 assignment (10 Marks)  
4 Quizzes (60 Marks)  
1 endsem exam (30 Marks)



Quizzes ~ 8 Mar, 26 Mar, 19 Apr, 7 May (0900-0945 hrs)

\*

Textbook: Introduction to Probability &  
Statistics for Engineers & Scientists



— Sheldon. M. Ross  
5<sup>th</sup> ed

- Probability, Random Variables & Stochastic Processes

— Athanasios Papoulis & S. Unnikrishna Pillai



Take own notes.



Discuss actively in the forum.

Freely ask doubts, keep your shyness  
(& ego) aside:)

## • Basic steps in Statistics

- a) Data Collection (& cleaning)  
Outliers removal, standardization, ...
- b) Representing the data  
Visualization - plot, histogram, ...
- c) Summarize the data  
Central tendency, dispersion,  
\* Skewness, \* Kurtosis
- d) Conclusion



Given mean,  $\mu = 5.4$ , } for a dataset X,  
std deviation,  $\sigma = 1.2$  }

→ what can you say about the distribution of the data (distro)

→ Significance of sample size  $n$  of X in making inference about X

$n = 2$

$n > 2$

eg:  $\leftarrow \begin{array}{ccc} \times & \vdots & \times \\ \mu - \sigma & \mu & \mu + \sigma \end{array} \rightarrow$

$\leftarrow \begin{array}{ccccccc} \times & \times & \times & \times & \times & \times & \times \\ & & & & & & \vdots \\ & & & & & & \mu \end{array} \rightarrow$

17/02/2021



18/02/2021

a)

## DATA COLLECTION

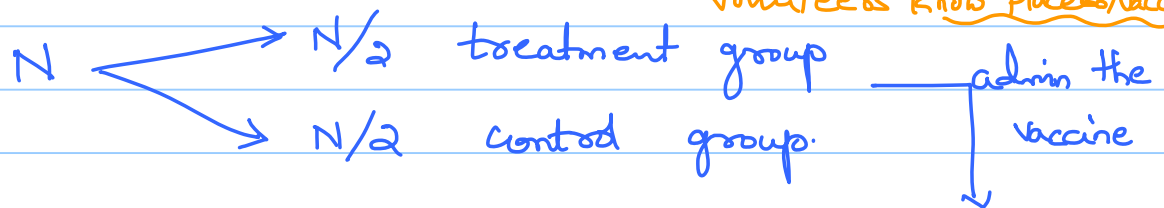
- eg: Clinical testing of efficacy of a vaccine.

Design of }  
experiments }

$N$  individuals/volunteers.  
Treatment group (expt. treatment)      Control group (Placebo)

## LECTURE 2

\* Some objectives.



Expectation:

After incubation period, check on "patients".

- More people in control group are infected.

⚠

		CONFUSION MATRIX	
Disease		present	absent
	Test		
	detected	TP	FP
	not detected	FN	TN

[https://en.wikipedia.org/wiki/False\\_positives\\_and\\_false\\_negatives](https://en.wikipedia.org/wiki/False_positives_and_false_negatives)

⚠

What's a good  $N$ ? 100 / 1000 / 10000 ...  
to  $\rightarrow$  make good decision.  
Will find out in due course.

• Minimize  $N$

F2 score etc.

• Still have good testing accuracy, efficiency, & confidence in inference.

## b) REPRESENTING / DESCRIBING THE DATA

• MATLAB (optional, useful for assignment).

++ Can use MATLAB online using institute mail-id  
@  
matlab.mathworks.com.

- Setup: 30 students choose an integer b/w 1 & 10. (frequently used). Denoted by  $X$ .

$$X = [1, 2, 2, \dots, 10]$$

a) Frequency table

# written, $x_i$	1	2	3	4	5	6	7	8	9	10	Total
# of students, $f_i$	1	3	3	2	1	4	9	3	3	1	30
frequency.											(N)

discrete data,  $x_i \in \{1, 2, \dots, 10\}$

Plot

<https://en.wikipedia.org/wiki/Histogram>



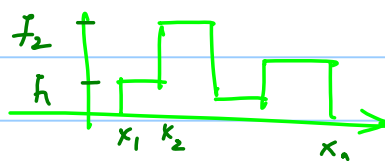
b) Bar chart

- for discrete observations



c) Histogram.

- Usually for grouped/continuous intervals.



- Normalized / relative frequency = proportion of students

$$f_i(\text{rel}) = f_i / \text{length}(X)$$

$$f_i(\text{rel}) \in [0, 1] \quad \forall i$$

- !! • area under this curve is 1  $\rightsquigarrow$  Similar to a probability mass/density function.

- relative frequency is analogous to probability.

- d) • pie-chart is a helpful tool / appropriate for this.

$$\left. \begin{array}{l} \text{area of sector} \\ \text{angle of sector} \end{array} \right\} \propto \text{frequency of observation.}$$

c) \* Stem and leaf plot.

↳ Helps keep record of the data in a concise manner.

Say all observations are in lakhs or in fractions.  
eg1. eg2.

phone #s, pincodes, ...  
9886784591, 9901425926, ...  
9886907843, 9482569380, ...

0.0011, 0.0025, 0.0019,  
0.0042, 0.0069, ...

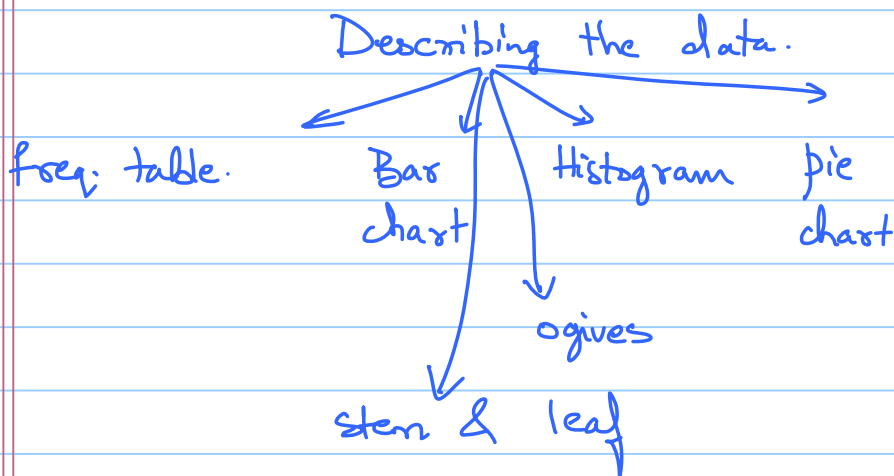
airtel  
vodafone  
BSNL

Stem	Leaf
← 990	1425926
← 988	6784591, 6907843
← 948	2569380

stem	leaf
0.001	1, 9
0.002	5
0.004	2
0.006	9

← first 3 digits of phone #  
rest of the observation.

++ Stem & leaf plot helps us group/form clusters based on the stem.



& many more.

18/02/2021  
↑

Textbook : Chapter 1  
Chapter 2 till section 2.2 page 14.