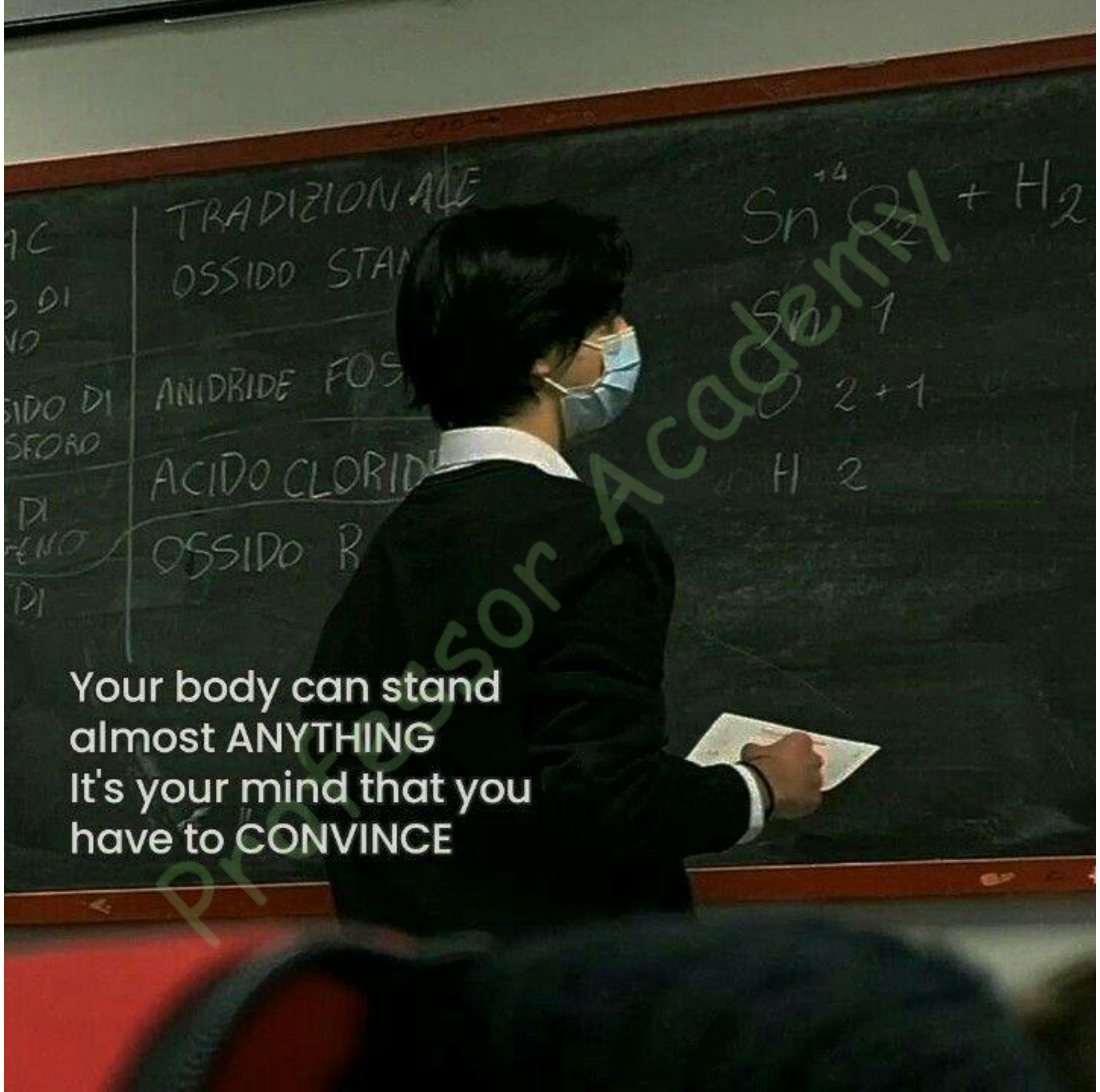

Your body can stand
almost ANYTHING
It's your mind that you
have to CONVINCE
Professor Academy

UNIT 10

STATISTICS AND PROBABILITY

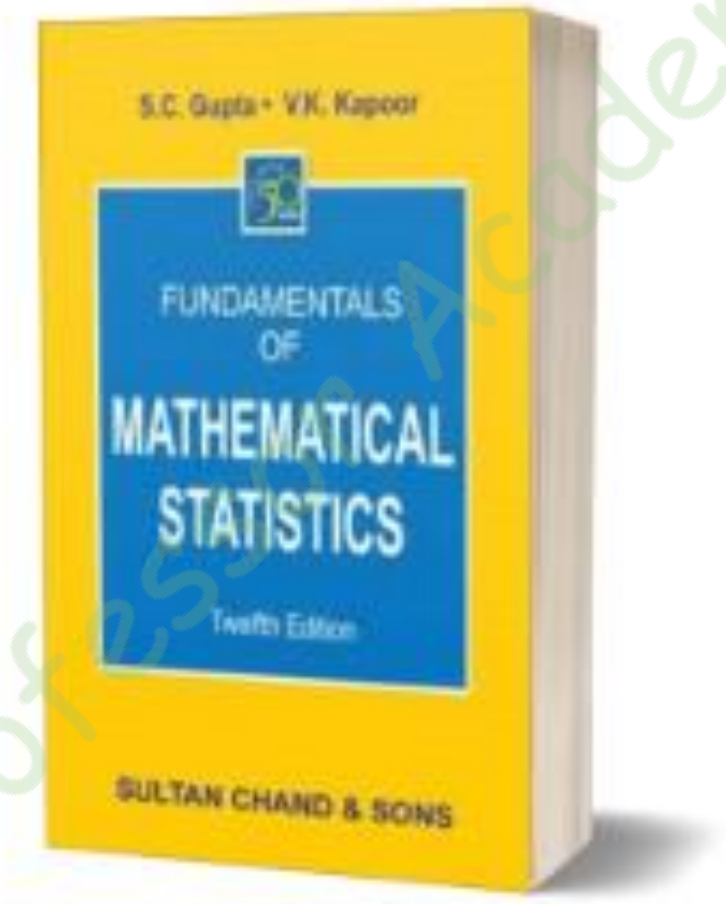
SYLLABUS

UNIT—X STATISTICS/PROBABILITY

Measures of central tendency – Measures of Dispersion – Moments – Skewness and Kurtosis – Correlation – Rank Correlation – Regression – Regression line of *x* on *y* and *y* on *x* – Index Numbers – Consumer Price Index numbers – Conversion of chain base Index Number into fixed base index numbers – Curve Fitting – Principle of Least Squares – Fitting a straight line – Fitting a second degree parabola – Fitting of power curves – Theory of Attributes – Attributes – Consistency of Data – Independence and Associate of data.

Theory of Probability – Sample Space – Axioms of Probability – Probability function – Laws of Addition – Conditional Probability – Law of multiplication – Independent – Boole's Inequality – Bayes' Theorem – Random Variables – Distribution function – Discrete and continuous random variables – Probability density functions – Mathematical Expectation – Moment Generating Functions – Cumulates – Characteristic functions – Theoretical distributions – Binomial, Poisson, Normal distributions – Properties and conditions of a normal curve – Test of significance of sample and large samples – Z-test – Student's t-test – F-test – Chi square and contingency coefficient.

REFERENCE BOOK



STATISTICS

Measures of central tendency
Measures of Dispersion
Moments
Skewness and Kurtosis

Correlation
Rank Correlation
Regression
Regression line of *x* on *y* and *y* on *x*

Index Numbers
Consumer Price Index numbers
Conversion of chain base Index Number into fixed base index numbers

Theory of Attributes
Attributes
Consistency of Data
Independence and Associate of data

Curve Fitting
Principle of Least Squares
Fitting a straight line
Fitting a second-degree parabola
Fitting of power curves

PROBABILITY

1

Theory of Probability
Sample Space
Axioms of Probability
Probability function
Laws of Addition
Conditional Probability
Law of multiplication
Independent
Boole's Inequality
Bayes' Theorem

Theoretical distributions
Binomial, Poisson, Normal distributions
Properties and conditions of a normal curve

Test of significance of sample and large samples
Z-test
Student's t-test
F-test
Chi square and contingency coefficient

Basics

Random Variables
Distribution function
Discrete and continuous random variables
Probability density functions
Mathematical Expectation
Moment Generating Functions
Cumulates
Characteristic functions

Measures of central tendency

Measures of Dispersion

Moments

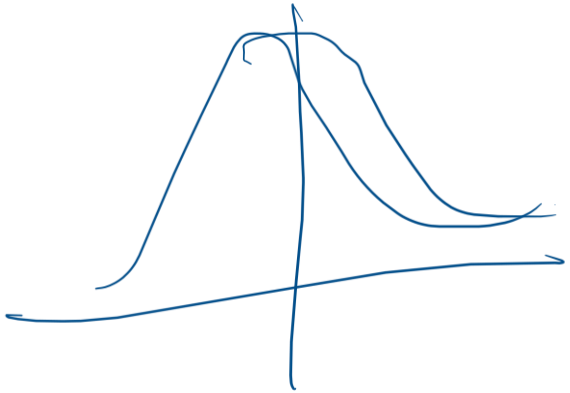
Skewness and Kurtosis

MEASURES OF CENTRAL TENDANCY OR AVERAGES

They show a tendency to concentrate at ascertain values, usually somewhere in the center of the distribution

MEASURE OF VARIATION OR DISPERSION

"deviate"



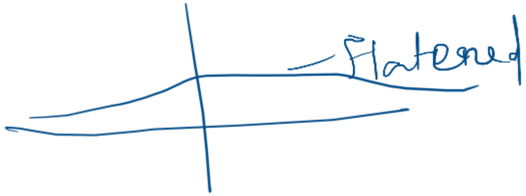
They show how much the data vary about a measure of central tendency.

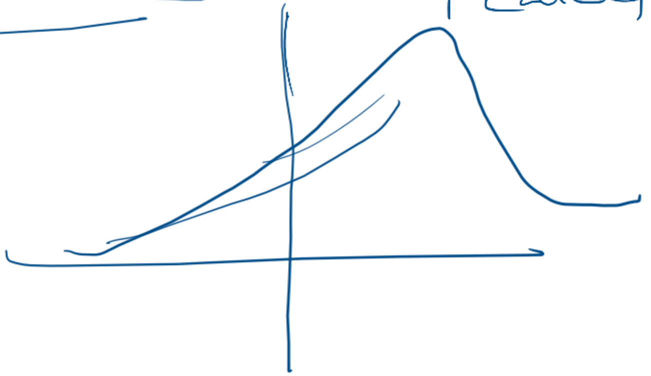
MEASURE OF SKEWNESS

Symmetric

This measures the direction and degree of asymmetry of a data in the frequency distribution.

MEASURE OF KURTOSIS





The measure of flatness or peakedness of the frequency curves.

5

MEASURES OF CENTRAL TENDENCY

1. Arithmetic Mean or Mean
2. Median
3. Mode – ill defined
4. Geometric Mean
5. Harmonic Mean ✓

Non-zero Data.

Population → 1000 → Samples

1, 2, 2, 5, 5 → Mode ?

IDEAL MEASURES OF CENTRAL TENDENCY – Prof. Yule

1. Well defined
2. Easy to calculate
3. Should be based on all the observations
4. Should be suitable for further mathematical treatment
5. Should be affected as little as possible by fluctuations of sampling
6. Should not be affected much by extreme values

2 – 2 times
5 – 2 times

mode = 2, 5 → Bimodal

change

least – greatest

ARITHMETIC MEAN OR AVERAGES

Sum of observations divided by the number of observations.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

MEDIAN

The value of the variable which divides it into two equal parts.

MODE

Most repeated value in the data.

I. 4, 5, 10, 12, 15, 10 → mode
5

Mean = $\frac{4 + 5 + 10 + 12 + 15}{5} =$

$\boxed{500}$ → Height.

List?

$P_1 \rightarrow 5$ ft
$P_2 \rightarrow 6$ f
$P_3 \rightarrow 5.5$ ft
$P_4 \rightarrow 5$ ft

II. Frequency - Repeat.

x (ft): height	4	4.5	5
f:	20	10	25

frequency.

$\boxed{4.001}$

continuous data ← { 4.6, 4.9, 4.62 }

4 ... 5

III.

x:	0-1	1-2	2-3	4-5
f:	10	15	20	10

→ Grouped data

ARITHMETIC MEAN OR AVERAGES

100 → 10, 30, 10 ↓ 50

Sum of observations divided by the number of observations.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

→ discrete [finite, only data]

In case of frequency distribution

Total frequency $= \Sigma f_i = N$

No. of observation

$$\bar{x} = \frac{\sum f_i x_i}{\sum f_i} = \frac{1}{N}\Sigma f_i x_i$$

→ finite, frequency

When considering deviation 'd' $= x - A$

$$\bar{x} = A + \frac{1}{N}\Sigma f_i d_i$$

$$d = \frac{x-A}{h}$$

$$\bar{x} = A + \frac{h}{N}\Sigma f_i d_i$$

x: $\underline{1}^{x_1}$ 2^{x_2} 3^{x_3} 4^{x_4} 5^{x_5} 6^{x_6} 7^{x_7}

f: $5 + 9 + 12 + 17 + 14 + 10 + 6 = 73 \rightarrow \Sigma f_i$

$$\text{Mean} = ? = \frac{\Sigma x_i f_i}{\Sigma f_i}$$

mean of TN

$$= \frac{5 + 18 + 36 + 68 + 70 + 60 + 42}{73}$$

$$= \frac{299}{73} = 4.09$$

$$\sum_{i=1}^{7} (x_i - \bar{x}) = 0 \;?$$

Midpoint of class

Class	f	x	$x-A = x-28$	$\frac{x-A}{h} = d$	fd
0-8	8	4	−24	−3	−24
8-16	7	12	−16	−2	−14
16-24	16	20	−8	−1	−16
24-32	24	28	0	0	0
32-40	15	36	8	1	15
40-48	7	44	16	2	14

$N = 77$

$A = 28$, $h = 8$

mean

$\boxed{\Sigma f_i d_i = -25}$

$$\bar{x} = A + \frac{\Sigma f_i d_i}{N} \times h$$

$$= 28 + \frac{(-25)}{77} \times 8$$

$$= 28 - \frac{200}{77} = \frac{2156-200}{77} = \frac{1956}{77} = 25.404$$

PROPERTIES OF ARITHMETIC MEAN OR AVERAGES

$$\Sigma (x_i - \bar{x})^2$$

DIY 1. Algebraic sum of the deviations of a set of values (x) from their arithmetic mean is zero.

2. The sum of squares of deviations of a set of values is minimum when taken about mean.

P.T $y = \Sigma (x_i - A)^2$ is minimum when $A = \bar{x}$.

discrete

$y' = 0$

$$2\Sigma (x_i - A) = 0$$

$$\Sigma (x_i - A) = 0$$

$$\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} A = 0 \Rightarrow \Sigma x_i = nA$$

$$\frac{\Sigma x_i}{n} = A \Rightarrow \boxed{A = \bar{x}}$$

3. Mean of composite series.

$$\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2 + \cdots + n_k\bar{x}_k}{n_1 + n_2 + \cdots + n_k}$$

4. Weighted Average mean.

$$\boxed{\bar{x} = \frac{\Sigma \omega_i x_i}{\sum \omega_i}}$$

$f_i \rightarrow w_i$

MEAN

Sum of observations divided by the number of observations.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

In case of frequency distribution

$$\bar{x} = \frac{\sum f_i x_i}{\sum f_i} = \frac{1}{N}\Sigma f_i x_i$$

When considering deviation 'd' $= x - A$

$$d = \frac{x-A}{h}$$

$$\bar{x} = A + \frac{1}{N}\Sigma f_i d_i$$

$$\bar{x} = A + \frac{h \Sigma f_i d_i}{N}$$

1.Data Set: 74,000, 82,000, 75,000, 96,000, 88,000
1.Data Set: 74,000, 82,000, 75,000, 96,000, 88,000
1.Data Set: 74,000, 82,000, 75,000, 96,000, 88,000

MEDIAN

The median is the middle value in a data set that has been ordered from least to greatest.

Example: In the set {5, 2, 7, 12, 9}, the median is 7

In the ordered set {2, 5, 4, 7}, the median is 4.5 (the average of 4 and 5)

2, 5, 7, 9, 12

2, $\boxed{4, 5}$, 7

$$\frac{4+5}{2} = \frac{9}{2} = 4.5$$

Find the median of the data: 74,000, 82,000, 75,000, 96,000, 88,000

MEDIAN FOR CONTINUOUS FREQUENCY DISTRIBUTION

The class corresponding to the c.f. just greater than N/2 is called median class and the value of median is obtained by the following formula: **Median** $= l + \frac{i}{f}\left(\frac{N}{2} - c\right)$

Class Interval (X)	Frequency (f)
0-5	5
5-10	3
10-15	4
(l_1) 15-20	8 (f)
20-25	7
25-30	3
	N = ∑f = 30

Median class ←

l – lower limit = 15

f –

$\frac{30}{2} = 15$

Class Interval (X)	Frequency (f)	Cumulative Frequency (c.f.)
0-5	5	5
5-10	3	5 + 3 = 8
10-15	4	8 + 4 = 12 (c.f)
(l_1) 15-20	8 (f)	12 + 8 = 20 Median Class
20-25	7	20 + 7 = 27
25-30	3	27 + 3 = 30
	N = $\sum$f = 30	

Median class

Median(M)=Size of [N/2]th item

$= Size\ of\ [N/2]\ th\ item = Siz\ of\ 15^{th}\ item$

+

Hence, the median lies in the class 15-20.

l= 15, f = 8, i = 5, c.f. = 12

cumulative (preceding)

step size

Now apply the following formula

$$Median = l + \frac{\frac{N}{2} - c.f.}{f} \times i = 15 + \frac{\frac{30}{2} - 12}{8} \times 5 = \mathbf{16.875}$$

MODE - Most repeated value

The mode is the value that appears most frequently in a data set.

Example

Consider the following data set of hockey scores: {7, 5, 0, 7, 8, 5, 5, 4, 1, 5}.

Count the frequency of each number:

7 appears 2 times.

5 appears 4 times.

0 appears 1 time.

8 appears 1 time.

4 appears 1 time.

1 appears 1 time.

The mode of this data set is 5.

4

MODE FOR CONTINUOUS FREQUENCY DISTRIBUTION

The mode for a continuous frequency distribution is calculated using the following formula.

$$Mode = l + \frac{h(\overset{\text{Mod}}{f_1} - \overset{\text{Pre}}{f_0})}{2f_1 - f_0 - \underset{\text{Succ}}{f_2}}$$

f_m, f_p, f_s

Class interval	Frequency
0 – 10	5
10 – 20	8
20 – 30	7
30 – 40	12 (f_0)
40 – 50 ← modal class	**28** (f_1)
50 – 60	20(f_2)
60 – 70	10
70 – 80	10

$l = 40$; $h = 10$

$f_1 = 28$, $f_0 = 12$, $f_2 = 20$

GEOMETRIC MEAN

$x/y = \frac{x}{y}$ $y \neq 0$

x: 2, 5, 10, 20, 1, 0

Geometric mean of a set of n observations is the nth root of their product.

$1/5$

$$G = \text{Antilog}\left[\frac{1}{n}\sum_i \log x_i\right] \Leftarrow G = (x_1 . x_2 \dots . x_n)^{1/n}$$

$$GM = (2 \times 5 \times 10 \times 20 \times 1 \times 0) = 0$$

Geometric mean of the combined group: $G = Antilog\left(\frac{n_1 \log G_1 + n_2 \log G_2 + \cdots + n_k \log G_k}{n_1 + n_2 + \cdots + n_k}\right)$

HARMONIC MEAN

"NON-ZERO VALUES"

It is the reciprocal of the arithmetic mean of the reciprocals of the given data.

1. Data → Reciprocal
2. AM
3. Reciprocal

$\{0, 2, 4, 6, 8\} \to HM$

$1/0$?

$\{\frac{1}{2}, \frac{1}{4}, \frac{1}{6}, \frac{1}{8}\} \to AM$

$\infty + 1/2 = ?$

$$H = \frac{1}{\frac{1}{N}\sum_{i=1}^{n}\frac{1}{x_i}} = \frac{106}{25}$$

$$= \frac{1}{4}\left[\frac{1}{2} + \frac{1}{4} + \frac{1}{6} + \frac{1}{8}\right] = \frac{1}{4}\left[\frac{12 + 6 + 4 + 3}{24}\right] = \frac{25}{106}$$

PARTITION VALUES

Q_1, Q_2, Q_3

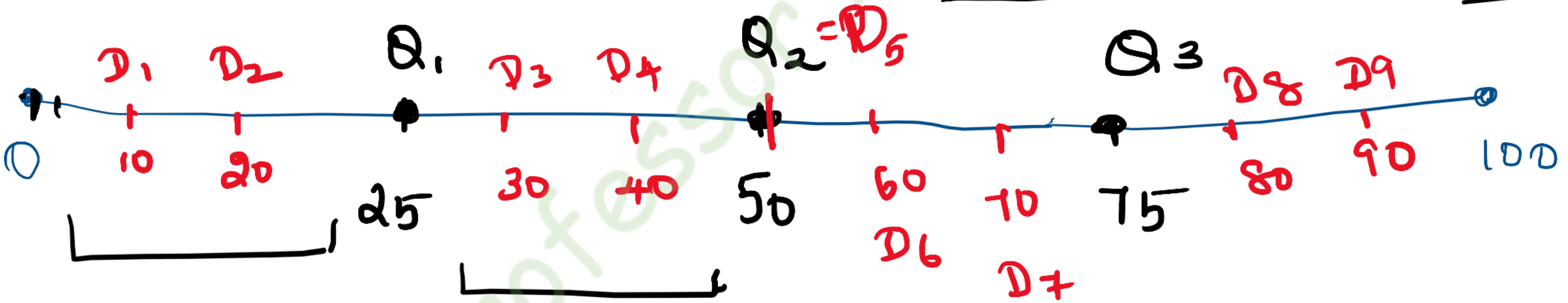
The three points which divide the series into four equal parts are called **quartiles**. 'Q'

D_1 to D_9

The nine points which divide the series into ten equal parts are called **deciles**. 'D'

The ninety-nine points which divide the series into hundred equal parts are called **percentiles**.

P_1 to P_{99}



$$Q_1 = \frac{N}{4} \qquad Q_2 = \frac{N}{2} \qquad Q_3 = \frac{3N}{4}$$