Pre-Exam Marathon

CA Foundation Dec 23

MATH, LR & STATS

Session 1

Time Value of Money, Statistical Description of Data, Measures of Central Tendency and Dispersion, Correlation Regression

SESSION LINK:

https://www.youtube.com/live/9-nM_-VOj3Q?si=FI2oDiyqKBirnTia

JOIN TELEGRAM CHANNEL FOR ALL UPDATES AND NOTES:

https://telegram.me/learnwithpranav


ULTIMATE CA
CA MOHNISH VORA
CA DEEPIKA RATHI
CA INTERMEDIATE
NOVEMBER 24
Early Bird Batch
CA TEJAS SUCHAK
CA INDRESH GANDHI
STARTING 24TH JANUARY
CA VIVEK GABA
CA PRANAV POPAT

Time Value of Money

Past Trends

Attempt	SI & CI	Annuity and Other	Total
May 2018	3	3	6
Nov 2018	11	3	14
Jun 2019	7	3	10
Nov 2019	10	3	10
Nov 2020	7	7	14
Jan 2021	10	4	14
Jul 2021	6	7	13
Dec 2021	4	3	7
Jun 2022	2	8	10
Dec 2022	8	6	14
June 2023	7	7	14

Calculator Tricks & Basics

Power (Integer)	Base [×][=][=][=]... (square, cube)
n^th Power (Non-Integer)	Base [√][√][√]...12times..[−1][×][n][+1][×=][×=][×=]...12times
n^th Root	Base [√][√][√]...12times..[−1][÷][n][+1][×=][×=][×=]...12times
Reciprocal of any number	[÷][=]
Reasons to pay/ receive Interest	Opportunity Cost: ▪ To lend money to others, we sacrifice the return on investing that money somewhere else Inflation: ▪ Time Factor: Due to inflation a given amount of money buys fewer goods in the future than it will now Liquidity Preference: ▪ After lending, money is not available for immediate use Risk Factor: ▪ Due to inflation a given amount of money buys fewer goods in the future than it will now

Basic Terms	

Interest	Interest is the **price paid** by a borrower for the **use of a lender's** money.
Principal	Principal is initial value of **lending** (or **borrowing**).
Rate of Interest	The rate at which the interest is charged for a defined length of time for use of principal generally on a **yearly basis** is known to be the rate of interest.
Accumulated Balance	Accumulated amount is the **final value** of an investment. It is the **sum total** of principal and interest earned.

Simple Interest

Concept	▪ Simple interest is the interest computed on the principal **for the entire period** of borrowing. ▪ It is calculated on the **principal amount only** and not on interest previously earned. ▪ Value of Interest **remains constant** for each year
Formula of Simple Interest	$SI = \frac{P.r.t}{100}$ where, P = Principal Value, r = rate of interest per annum, t = time in years
Formula of Amount as per Simple Interest	$A = P + SI$ $A = P + \frac{P.r.t}{100} = P(1 + \frac{rt}{100})$

Compound Interest

Basics	▪ We can define the compound interest as the interest **that accrues** when earnings for each specified period **added to the principal.** ▪ In CI, after every conversion period we increase the **principal base** on which **subsequent interest** is computed.
Conversion Period	**Conversion Period**: Period for which interest is computed

Conversion Period	Description	Number of Conversion Period in a year
1 day	Compounded Daily	365
1 month	Compounded Monthly	12
3 months	Compounded Quarterly	4
6 months	Compounded Semi Annually	2
12 months	Compounded Annually	1

Formula for Amount as per Compound Interest	$A = P(1+i)^n$ where, P = Initial Principal, i = adjusted interest rate, n = no. of periods $i = \frac{r\%}{nocppy}$ $n = t \times noccpy$
Formula for Compound Interest	$CI = A - P$ $CI = P(1+i)^n - P$ $CI = P[(1+i)^n - 1]$ where, P = Initial Principal, i = adjusted interest rate, n = no. of periods
Trick for Amount as per Compound Interest	P + i% + i% + ...n times Suitable when value of n is small
Effective Rate of Interest	Equivalent **annual rate** of interest compounded annually if interest is compounded **more than once a year.** Effective rate is not dependent on Principal. $E = [(1+i)^n - 1]$
CI Concept in WDV Depreciation	$A = P(1-i)^n$ where, P = Historical Cost of Asset, A = Scrap Value/ Residual value of asset, n = no. of periods, i = Depreciation %

MCQs

PYQ Jun 19

The certain sum of money became ₹ 692 in 2 years and ₹800 in 5 years then the principal amount

a. 520 b. 720
c. 620 d. 820

Ans: c

PYQ Nov 18

A certain money doubles itself in 10 years when deposited on simple interest, it would triple itself in

a. 25 years b. 15 years c. 20 years d. None

Ans: c

PYQ Jun 19

In simple interest if the principal is Rs.2,000 and the rate and time are the roots of the equation $x^2 - 11x + 30 = 0$ then simple interest is:

a. 500 b. 600 c. 700 d. 800

Ans: b

PYQ May 18

₹8,000 becomes ₹10,000 in two years at SI. The amount that will become ₹6,875 in 3 years at the same rate of interest is

a. 4850 b. 5000 c. 5500 d. 5275

Ans: b

PYQ Nov 18

A certain sum of money Q was deposited for 5 years and 4 months at 4.5% simple interest and amounted to ₹248, then the value of Q is

a. ₹200 b. ₹210 c. ₹220 d. ₹240

Ans: a

PYQ Nov 18

A man deposited Rs.8000 in a bank for 3 year at 5% per annum compound interest, after 3 years he will get:

a. 8800 b. 9200 c. 9261 d. 1261

Ans: c

PYQ Nov 18

If Rs. 10,000 is invested at 8% per year compounded quarterly, then the value of investment after 2 years is

a. 11716.59 b. 10716.59 c. 117.1659 d. None

Ans: a

Exercise

The population of a town increases every year by 2% of the population at the beginning of that year. The number of years by which the total increase of population be 40% is

a. 7 years b. 10 years c. 17 years d. None

Ans: c

PYQ Nov 18

A bank pays 10% rate of interest compounded annually. A sum of Rs.400 is deposited in the bank. The amount at the end of 1 year will be

a. 440 b. 441 c. 439 d. 442

Ans: a

PYQ May 18

If an amount is kept at SI it earns an interest of ₹600 in first two years but when kept at CI it earns an interest of ₹660 for the same period, then the rate of interest and principal is

a. 20%, 1200 b. 20%, 1500 c. 10%, 1200 d. 10%, 1500

Ans: b

PYQ Nov 18

If compounded interest on a sum for 2 years at 4% per annum is Rs.102, then the simple interest on the same sum for the same period at the same rate will be

a. 99 b. 100 c. 101 d. 95

Ans: b

PYQ Nov 19

The difference between CI and SI for 2 years, is 21. If rate of interest is 5% find principal

a. 8400 b. 8000 c. 4800 d. 8200

Ans: a

PYQ May 18

If Rs. 1000 be invested at interest rate of 5% and the interest is added to the principal every 10 years, then the number of years in which it will amount to Rs. 2000 is

a. 16.66 years b. 6.25 years c. 16 years d. 6.66 years

Ans: a

Exercise

The useful life of a machine is estimated to be 10 years and cost ₹ 10,000. Rate of depreciation is 10% p.a. The scrap value at the end of its life is

a. 3,486.78 b. 4,383 c. 3,400 d. None

Ans: a

PYQ Jun 19

A sum was invested for 3 years as per CI and the rate of interest for three years are respectively 9%, 6%, 3%. Find the sum if the amount in three years is ₹550?

a. 250 b. 300 c. 462.16 d. 350

Ans: c

PYQ Nov 19

What will be the population after 3 years when present population is 25,000 and population increases at the rate of 3% in 1 year at 4% in II year and at 5% III year

a. 28119 b. 27000 c. 29118 d. 30000

Ans: a

Exercise

The effective rate of interest corresponding a nominal rate of 7% p.a convertible quarterly is

a. 7% b. 7.5% c. 5% d. 7.18%

Ans: d

PYQ Jun 19

The effective rate of interest does not depend upon

a. Principal b. Conversion Period c. Rate of Interest d. None

Ans: a

Exercise

The annual birth and death rates per 1,000 are 39.4 and 19.4 respectively. The number of years in which the population will be doubled assuming there is no immigration or emigration is

a. 35 years b. 30 years c. 25 years d. None

Ans: a

PYQ Nov 20

An amount is lent at a nominal rate of 4.5% per annum compounded quarterly. What would be the gain in rupees over when compounded annually?

a. 0.56 b. 0.45 c. 0.076 d. 0.85

Ans: c

PYQ Nov 20

The ratio of principal and the compound interest value for three years (compounded annually) is 216 : 127. The rate of interest is:

a. 0.1777 b. 0.1567 c. 0.1666 d. 0.1587

Ans: c

PYQ Jul 21

What is the difference (in ₹) between the simple interest and the compound interest on a sum of ₹ 8,000 for $2\frac{2}{5}$ years at the rate of 10% p.a. when the interest is compounded yearly?

a. 136.12 b. 129.50 c. 151.75 d. 147.20

Ans: a

PYQ Dec 22

A sum of money invested of compound interest double itself in four years. In how many years it become 32 times of itself at the same rate of compound interest?

a. 12 years b. 16 years c. 20 years d. 24 years

Ans: c

Types of Cashflows

Single Cashflow	If single amount is paid or received initially and then direct finally at the end
Annuity	Annuity can be defined as a sequence of **constant periodic** payments (or receipts) regularly over a specified period.
Types of Annuities	Annuity Regular – First payment/receipt at the end of the period Annuity Due – First payment/receipt at the beginning of the period If Question is silent – annuity is to be taken as regular

Future Value

Future Value – Single Cashflow	▪ Future value is the **cash value** of an **investment** at some time in the future. ▪ It is **tomorrow's value** of today's money **compounded** at the rate of interest.
Formula for FV of Single Cashflow	$FV = CF(1+i)^n$ where, CF = Single Cashflow for which FV is to be calculated, i = adjusted interest rate, n = no. of periods
FV of Annuity Regular	▪ To calculate **final maturity value** of an investment like RD where sum is invested in the annuity pattern starting at the end of each period. ▪ To calculate the final value of Sinking Fund or Savings amount to achieve the **target maturity value.**
Formula for Future Value - Annuity Regular	$FVAR = A_i \times FVAF(n,i)$ $FVAR = A_i \times \left\{ \frac{[(1+i)^n - 1]}{i} \right\}$ where, **FVAR** = Future Value of Annuity Regular, $\mathbf{A_i}$ = Annuity Value (Installment), **FVAF** = Future Value Annuity Factor, **i** = adjusted interest rate, **n** = no. of periods
FV of Annuity Due	▪ To calculate final maturity value of an investment like RD where sum is invested in the annuity pattern **at the beginning of each period**

	▪ To calculate final maturity value of an investment like RD where sum is invested in the annuity pattern at the beginning of each period
Formula for Future Value - Annuity Due	$$FVAD = A_i \times FVAF(n,i) \times (1+i)$$ $$FVAD = A_i \times \left\{ \frac{[(1+i)^n - 1]}{i} \right\} \times (1+i)$$ **where, FVAD**= Future Value of Annuity Due, **A_i** = Annuity Value (Installment), **FVAF** = Future Value Annuity Factor, **i** = adjusted interest rate, **n** = no. of periods
Sinking Fund	▪ It is the **fund credited** for a specified purpose by way of **sequence of periodic payments** over a **time-period** at a specified interest rate. ▪ Interest is compounded at the **end of every period**. ▪ Size of the sinking fund deposit is same as Future Value of Annuity
Compounding and Discounting	Compounding (Adding the interest) → $\times(1+i)^n$ Discounting (Removing the interest) → $\times \frac{1}{(1+i)^n}$

Present Value

Present Value of Single Cashflow	▪ Present value is **today's value** of tomorrow's money **discounted** at the interest rate
Formula for PV of Single Cashflow	$$PV = \frac{CF}{(1+i)^n}$$ where, CF = Single Cashflow for which PV is to be calculated, i = adjusted interest rate, n = no. of periods
Present Value – Annuity Regular	**Use:** To calculate **loan amount** when periodic installments value are given and vice-versa. **Application:** Leasing, Capital Expenditure etc.
Formula for PV of Annuity Regular	$$PVAR = A_i \times PVAF(n,i)$$ $$PVAR = A_i \times \left[\frac{1}{i} \times \left\{ 1 - \frac{1}{(1+i)^n} \right\} \right]$$ **where, PVAR** = Present Value of Annuity Regular, **A_i** = Annuity Value (Installment), **PVAF** = Present Value Annuity Factor, **i** = adjusted interest rate, **n** = no. of periods
Calculator Trick for PVAF	$\boxed{1+i}\boxed{\div}\boxed{=}\boxed{=}...n-\text{times}\boxed{GT}$
Formula for Present Value of Annuity Due	$$PVAD = \left[A_i \times PVAF\{(n-1),i\} \right] + A_i$$

Applications of TVOM & Other Concepts

<table>
<tr><td>Leasing</td><td>

Lessor: Owner of Asset, who gives asset on rent. Lease Rentals are income for Lessor
Lessee: User of the asset who has taken asset on rent. Lease Rentals are expense for Lessee
Use of TVOM: Present Value of Annuity (Lease Rentals) are compared with asset cash down price to decide if leasing is preferable or not.

</td></tr>
<tr><td>Capital Expenditure Decisions</td><td>

Present value of future benefits due to new asset are compared with purchase value of asset, to decide whether asset to purchase or not.

</td></tr>
<tr><td>Valuation of Bond</td><td>

Present value of interest income and maturity value is compared with the issue price of bond
Terms

<table>
<tr><td>Bond</td><td>It is a debt security. Type of loan taken by company from public. Like debentures</td></tr>
<tr><td>Face Value/ Par Value</td><td>Value written on the document of bond. This value is used to calculate Interest Amount</td></tr>
<tr><td>Issue Price</td><td>Actual payment made to purchase the bond</td></tr>
<tr><td>Maturity Value</td><td>Amount to be received on redemption or maturity of bond</td></tr>
</table>
</td></tr>
<tr><td>PV of Perpetuity</td><td>Perpetuity: An annuity that continues till infinite period of time is called as Perpetuity.
$$PVP = \frac{A_i}{i}$$
where,
PVP = Present Value of Perpetuity, A_i = Annuity Value (Installment), i = adjusted interest rate</td></tr>
<tr><td>PV Growing Perpetuity</td><td>A stream of cashflows that grows at constant rate forever is known as growing perpetuity.
$$PVGP = \frac{A_i}{i-g}$$
where,
PVGP = Present Value of Growing Perpetuity
A_i = Annuity Value (Installment)
i = adjusted interest rate
g = growth rate</td></tr>
<tr><td>Net Present Value</td><td>
<table>
<tr><td>Formula</td><td>NPV = Present Value of Cash Inflows – Present Value of Cash Outflows</td></tr>
<tr><td>Decision Base</td><td>If NPV ≥ 0, accept the proposal,
If NPV < 0, reject the proposal</td></tr>
</table>
</td></tr>
<tr><td>Real Rate of Return</td><td>Real Rate of Return = Nominal Rate of Return – Rate of Inflation</td></tr>
<tr><td>CAGR</td><td>Compounded Annual Growth rate used to show annual growth as per CI</td></tr>
</table>

MCQs

Exercise

A person invests ₹ 500 at the end of each year with a bank which pays interest at 10% p.a. C.I. annually. The amount standing to his credit one year after he has made his yearly investment for the 12th time is.

e. 11,761.36 f. 10,692.34
g. 12,000 h. None

Ans: a

Exercise

A person bought a house paying ₹ 20,000 cash down and ₹ 4,000 at the end of each year for 25 yrs. at 5% p.a. C.I. The cash down price is

e. 75,000 f. 76,000 g. 76,375.5 h. None

Ans: c

Exercise

Johnson left ₹ 1,00,000 with the direction that it should be divided in such a way that his minor sons Tom, Dick and Harry aged 9, 12 and 15 years should each receive equally after attaining the age 25 years. The rate of interest being 3.5%, how much each son receives after getting 25 years old?

e. 50,000 f. 51,947 g. 52,000 h. None

Ans: b

Exercise

A sinking fund is created for redeeming debentures worth ₹ 5 lakhs at the end of 25 years. How much provision needs to be made from profits each year provided sinking fund investments can earn interest at 4% p.a.?

a. 11206 b. 12006 c. 13264 d. None

Ans: b

Exercise

The amount of an annuity certain of ₹ 150 for 12 years at 3.5% p.a C.I is

a. 2190.28 b. 1290.28 c. 2180.28 d. None

Ans: a

Exercise

A loan of ₹ 10,000 is to be paid back in 30 equal instalments. The amount of each installment to cover the principal and at 4% p.a CI is

a. 587.87 b. 587 c. 578.30 d. None

Ans: c

Exercise

Appu retires at 60 years receiving a pension of 14,400 a year paid in half-yearly installments for rest of his life after reckoning his life expectation to be 13 years and that interest at 4% p.a. is payable half-yearly. What single sum is equivalent to his pension?

a. 1,45,000 b. 1,44,900 c. 1,44,800 d. 1,44,700

Ans: b

PYQ June 19

A person wants to lease out a machine costing ₹5,00,000 for a 10-year period. It has fixed rental of ₹51,272 per annum payable annually starting from the end of the first year. Suppose rate of interest is 10% p.a. compounded annually. To whom this agreement is favorable?

a. Favor of Lessee b. Favor of Lessor
c. Not for both d. Can't be determined

Ans: a

PYQ June19

Determine the present value of perpetuity of ₹50,000 per month @ rate of interest 12% p.a. is _____

a. 45,00,000 b. 50,00,000 c. 55,00,000 d. 60,00,000

Ans: b

PYQ Jul 21

If discount rate is 14% p.a., then how much a company has to pay to receive ₹280 growing at 9% annually forever?

a. 5,600 b. 2,800 c. 1,400 d. 4,200

Ans: a

PYQ Jul 21

If the nominal rate of growth is 17% and inflation is 9% for the five years. Let P be the GDP amount at the present year then the projected real GDP after 6 years is

a. 1.587P b. 1.921P c. 1.403P d. 2.51P

Ans: a

PYQ Jul 21

Let the operating profit of a manufacturer for five years is given as:

Years	1	2	3	4	5	6
Operating Profit	90	100	106.4	107.14	120.24	157.34

Find CAGR

a. 9% b. 12% c. 11% d. 13%

Ans: b

PYQ Jul 21

If the cost of capital is 12% p.a., then the NPV from the given cashflow is

Years	0	1	2	3
Cashflow	(100)	60	40	50

a. 31048 b. 34185 c. 21048 d. 24187

Ans: c

Example

An investor intends purchasing a three-year Rs. 1000 par value bond having nominal interest rate of 10%. At what price the bond may be purchased now if it matures at par and the investor requires a rate of return of 14%?

a. 907.125 b. 900.36 c. 916.66 d. 569.22

Ans: a

PYQ Jun 19

Let a person invest a fixed sum at the end of each month in an account paying interest 12% per year compounded monthly. If the future value of this annuity after the 12^{th} payment is ₹ 55,000 then the amount invested every month is?

a. ₹ 4,837 b. ₹ 4,637 c. ₹ 4,337 d. ₹ 3,337

Ans: c

PYQ Nov 20

A stock pays annually an amount of Rs. 10 from 6^{th} year onwards. What is the present value of the perpetuity if the rate of return is 20%?

a. 20.1 b. 19.1 c. 21.1 d. 22.1

Ans: a

PYQ Jan 21

The present value of an annuity immediate is the same as

a. Annuity regular for (n-1) year plus the initial receipt in the beginning
b. Annuity regular for (n-1) years
c. Annuity regular for (n+1) years
d. Annuity regular for (n+1) year plus the initial receipt in the beginning

Ans: a

PYQ Dec 22

10 years ago the earning per share (EPS) of ABC Ltd. was ₹ 5 share. Its EPS for this year is ₹ 22. Compute at what rat, EPS of the company grow annually?

a. 15.97% b. 16.77% c. 18.64% d. 14.79%

Ans: a

PYQ Dec 22

Sinking fund factor is the reciprocal of:

a. Present value interest factor of a single cash flow
b. Present value interest factor of an annuity
c. Future value interest factor of an annuity
d. Future value interest factor of a single cash flow

Ans: c

Statistical Description of Data (Chp13)

Past Trends

Attempt	Theory	Practical	Total
May 2018	2	0	2
Nov 2018	6	1	7
Jun 2019	5	0	5
Nov 2019	1	1	1
Nov 2020	8	0	8
Jan 2021	10	0	10
Jul 2021	5	2	7
Dec 2021	3	4	7
Jun 2022	9	0	9
Dec 2022	0	4	4
Jun 2023	2	2	4

Statistical Description of Data – Basics of Statistics

Definition of Statistics

- Plural Sense: Any data – quantitative or qualitative used for statistical analysis.
- Singular Sense: Scientific method of collecting, analyzing, and presenting data to draw statistical inferences. It is also called as Science of Averages or Science of Counting

Origin of Word

Language	Actual Word	Memorize by
Latin	Status	Latus
Italian	Statista	Pasta
German	Statistic	Breadstick
French	Statistique	Barbeque

<table>
<tr><td>Publication</td><td></td><td>
<table>
<tr><td>Koutilya's Arthashastra</td><td>• Record of Birth and Deaths
• Chandragupta's reign
• 4th Century B.C</td></tr>
<tr><td>Abu Fezal's Ain-i-Akbari</td><td>• Record on Agriculture
• Akbar Reign
• 16th Century A.D.</td></tr>
<tr><td>First Census</td><td>• Egypt 300 BC to 2000 BC
• By Pharaoh</td></tr>
</table>
</td></tr>
<tr><td>Application of Statistics</td><td>• Economics: Demand Analysis, Future Projection etc.
• Business Management: Decision making using quantitative techniques not intuition
• Industry and Commerce: Profit maximization using business data – sales, purchase, market etc. by consulting experts</td></tr>
<tr><td>Limitation of Statistics</td><td>• It deals with aggregate data and not individual data
• Quantitative data can only be used, however for qualitative – it needs to be converted into quantitative
• Projections are based on conditions/ assumptions and any change in that will change the projection. Example: Future projections of sales
• Sampling based conclusions are used, improper sampling leads to improper results. Random Sampling is must.</td></tr>
<tr><td>Data</td><td>• Quantitative Information shown as number
• Primary: first time collected by agency/ investigator
• Secondary: collected data used by different person/ agency</td></tr>
<tr><td>Variable</td><td>• Measurable Data – Value can vary
<table>
<tr><td>Discrete Variable</td><td>• When a variable assumes a finite or count ably infinite isolated values.
• Example: no. of petals in a flower, no. of road accident in locality</td></tr>
<tr><td>Continuous Variable</td><td>• When a variable assumes any value from the given interval (can also be in decimals, fractions).
• Example: height, weight, sale, money</td></tr>
</table>
</td></tr>
<tr><td>Attribute</td><td>• Qualitative Characteristics. Example: gender of a baby, the nationality of a person, the colour of a flower etc.</td></tr>
</table>

<table>
<tr><td>Collection of Primary Data – Interview Method</td><td>
<table>
<tr><th>Method</th><th>Details</th></tr>
<tr><td>Personal Interview</td><td>• Where data is collected directly from respondents.
• Highly Accurate – Low Coverage
• Example: Natural Calamity, Door to Door Survey</td></tr>
<tr><td>Indirect Interview</td><td>• When reaching respondent is difficult, data is collected by contacting associated persons.
• Highly Accurate – Low Coverage
• Example: Rail accident</td></tr>
<tr><td>Telephone Interview</td><td>• Data is collected over phone
• Quick and non-expensive method
• Low Accuracy – High Coverage</td></tr>
</table>
</td></tr>
<tr><td>Collection of Primary Data – Mailed Questionnaire Method</td><td>• In this method well drafted and soundly sequenced questionnaire,
• covering all the important aspects of the data requirement is sent to respondent for filling.
• Here coverage is wide but amount of non-responses will be maximum</td></tr>
<tr><td>Collection of Primary Data – Observation Method</td><td>• In this method data is collected by direct observation or using instrument.
• For example: data on height and weight for a group of students.
• Although more accurate but it is time consuming, low coverage and laborious method.</td></tr>
<tr><td>Collection of Primary Data – Questionnaire Filled and sent by Enumerators</td><td>• Mix of Interview and Mailed Questionnaire
• Enumerator means a Person who directly interacts with respondent and fills the questionnaire.
• It is generally used in case of Surveys and Census.</td></tr>
<tr><td>Sources of Secondary Data</td><td>
<table>
<tr><td>International Sources</td><td>World Health Organization (WHO), International Monetary Fund (IMF), International Labor Organization (ILO), World Bank</td></tr>
<tr><td>Government Sources</td><td>In India – Central Statistics Office (CSO), Indian Agricultural Statistics by the Ministry of Food and Agri, National Sample Survey Office- NSSO, Regulators – RBI, SEBI, RERA, IRDA</td></tr>
<tr><td>Private or Quasi-govt. sources</td><td>Indian Statistical Institute (ISI), Indian Council of Agriculture, NCERT</td></tr>
</table>
</td></tr>
<tr><td>Scrutiny of Data</td><td>• checking accuracy and consistency of data
• There is no rule for it, one must apply his intelligence, patience and experience while scrutinizing the given information.
• Internal Consistency: When two or more series of related data are given, we should check consistency among them.</td></tr>
</table>

<table>
<tr><td>Presentation of Data – Classification / Organization of Data</td><td>Classification or Organisation: putting data in a neat, precise, and condensed form, making it comparable, suitable for analysis, more understandable.
<table>
<tr><td>Chronological/ Temporal/ Time Series Data</td><td>• Data arranged based on Time
• Example: Revenues YoY i.e year on year</td></tr>
<tr><td>Geographical or Spatial Series Data</td><td>• Arrangement based on regions
• Example: Country wise Revenue of a global company</td></tr>
<tr><td>Qualitative or Ordinal Data</td><td>• Based on some attribute
• Nationality Wise Medal Winners in Olympics</td></tr>
<tr><td>Quantitative or Cardinal Data</td><td>• Based on some variable
• Example: Frequency Distribution of a Data</td></tr>
</table></td></tr>
<tr><td>Mode of Presentation of Data – Textual</td><td>• This method comprises presenting data with the help of a paragraph or several paragraphs.
• This is not a suitable mode of presentation as it is dull, monotonous and non-comparable.</td></tr>
<tr><td>Mode of Presentation of Data – Tabular Form</td><td>• When data is shown in the form of Table.
• Useful in easy comparison
• Complicated data can be presented
• Table is must to create a diagram
• No analysis possible without table
• Components of Table</td></tr>
<tr><td>Components of Table</td><td>
<table>
<tr><th>Description</th><th>Name of Component of Table</th></tr>
<tr><td>Entire Upper Part</td><td>Box Head</td></tr>
<tr><td>Upper Part describing columns and sub-columns</td><td>Caption</td></tr>
<tr><td>Left part of the table describing rows</td><td>Stub</td></tr>
<tr><td>Main Data of Table</td><td>Body</td></tr>
<tr><td>Source of Data at the bottom of Table</td><td>Footnote</td></tr>
</table></td></tr>
<tr><td>Mode of Presentation of Data – Diagrams</td><td>• Can be used by educated and uneducated section of society
• Hidden trend can be traced
• If priority is accuracy, then tabulation is better</td></tr>
<tr><td>Line Diagram</td><td>• Time Series is generally in x axis
• For wide fluctuation – log chart or ratio chart is used
• Two or more series of same unit – Multiple Line Chart
• Two or more series of different unit – Multiple Axis Chart</td></tr>
<tr><td>Bar Diagram</td><td>• Bar means rectangle of same width and of varying length drawn horizontally or vertically
• For comparable series – multiple or grouped bar diagrams can be used
• For data divided into multiple components – subdivided or component bar diagrams</td></tr>
</table>

	• For relative comparison to whole, percentage bar diagrams or divided bar diagrams • Vertical Bar Diagram: Useful for Data varying over Time and Quantitative Data • Horizontal Bar Diagram: Useful for Data varying over Space and Qualitative Data
Pie Chart	• Used for circular presentation of relative data (% of whole) • Summation of values of all components/segments are equated to 360 Degree (total angle of circle) • **Segment angle = $\frac{\text{(segment value x 360°)}}{\text{(total value)}}$**

Statistical Description of Data – Frequency Distribution

Frequency and Distribution	• Frequency means number of times a particular observation is repeated. • Frequency Distribution is table which contains observation or class intervals in one column and corresponding frequency in the other. • Definition: A frequency distribution may be defined as a – tabular representation of statistical data, usually in an ascending order, – relating to a measurable characteristic – according to individual value or a group of values of the characteristic under study.	
Types of Frequency Distribution	**Ungrouped/ Simple Frequency Distribution**	• When there are limited number of distinct observations, frequency can be assigned to each one of them. • This distribution is simple
	Grouped Frequency Distribution	• When there are large no. of observations, grouping is done among them (generally in ascending order). • Each group is called as class interval and frequency is assigned to group and not individual values, • this is called Grouped Frequency Distribution
Class Limit	• For a class interval CL is the minimum and maximum value the class interval may contain • Minimum Value – Lower Class Limit • Maximum Value – Upper Class Limit	

Class Interval	Frequency	LCL	UCL
10-19	10	10	19
20-29	5	20	29
30-39	8	30	39

<table>
<tr><td rowspan="2">Classification of Grouped of Frequency Distribution</td><td>Mutually Exclusive / Overlapping Classification
<table><tr><td>Class</td><td>LCL</td><td>UCL</td></tr><tr><td>10-20</td><td>10</td><td>20</td></tr><tr><td>20-30</td><td>20</td><td>30</td></tr><tr><td>30-40</td><td>30</td><td>40</td></tr></table></td><td>• Here UCL an interval and LCL of next interval are same
• This is usually applicable for continuous variable.
• An observation which is equivalent to common class limit is excluded from the class interval where it is UCL and taken in the class where it is LCL.</td></tr>
<tr><td>Mutually Inclusive / Non-Overlapping Classification
<table><tr><td>Class</td><td>LCL</td><td>UCL</td></tr><tr><td>10-19</td><td>10</td><td>20</td></tr><tr><td>20-19</td><td>20</td><td>30</td></tr><tr><td>30-39</td><td>30</td><td>40</td></tr></table></td><td>• There is no common class limits between two intervals.
• This is usually applicable to discrete variable.
• All observation including UCL and LCL will be taken in the same class interval as there is no confusion.</td></tr>
<tr><td rowspan="2">Class Boundary</td><td>In case of Exclusive / Overlapping Classification</td><td>Class Boundary = Class Limit
<table><tr><td>Class</td><td>LCL</td><td>UCL</td><td>LCB</td><td>UCB</td></tr><tr><td>10-20</td><td>10</td><td>20</td><td>10</td><td>20</td></tr><tr><td>20-30</td><td>20</td><td>30</td><td>20</td><td>30</td></tr><tr><td>30-40</td><td>30</td><td>40</td><td>30</td><td>40</td></tr></table></td></tr>
<tr><td>In case of Inclusive / Overlapping Classification</td><td>Lower Class Boundary
LCB = LCL – 0.5
UCB = UCL + 0.5
<table><tr><td>Class</td><td>LCL</td><td>UCL</td><td>LCB</td><td>UCB</td></tr><tr><td>10-19</td><td>10</td><td>19</td><td>9.5</td><td>19.5</td></tr><tr><td>20-29</td><td>20</td><td>29</td><td>19.5</td><td>29.5</td></tr><tr><td>30-39</td><td>30</td><td>39</td><td>29.5</td><td>39.5</td></tr></table></td></tr>
<tr><td>Mid-Point / Class Mark / Mid Value of Class Interval</td><td colspan="2">$\frac{LCL+UCL}{2}$ $\frac{LCB+UCB}{2}$
• Useful in calculation of AM, GM, HM, SD in case of grouped frequency distribution</td></tr>
<tr><td>Class Length/ Width or Size</td><td colspan="2">UCB – LCB only</td></tr>
</table>

<table>
<tr><td>Cumulative Frequency</td><td>

Less than type: It shows no. of observations less than UCB
More than type: It shows no. of observations more than UCB

<table>
<tr><th>Class Interval</th><th>Freq.</th><th>UCB</th><th>Less than type CF</th><th>More than type CF</th><th>Total of both CF</th></tr>
<tr><td>44-48</td><td>3</td><td>48.5</td><td>3</td><td>33</td><td>36</td></tr>
<tr><td>49-53</td><td>4</td><td>53.5</td><td>7</td><td>29</td><td>36</td></tr>
<tr><td>54-58</td><td>5</td><td>58.5</td><td>12</td><td>24</td><td>36</td></tr>
<tr><td>59-63</td><td>7</td><td>63.5</td><td>19</td><td>17</td><td>36</td></tr>
<tr><td>64-68</td><td>9</td><td>68.5</td><td>28</td><td>8</td><td>36</td></tr>
<tr><td>69-73</td><td>8</td><td>73.5</td><td>36</td><td>0</td><td>36</td></tr>
<tr><td>Total</td><td>36</td><td></td><td></td><td></td><td></td></tr>
</table>
</td></tr>
<tr><td>Frequency Density</td><td>$\frac{\text{Class Frequency}}{\text{Class Length of class}}$</td></tr>
<tr><td>Relative Frequency</td><td>$\frac{\text{Class frequency}}{\text{Total Frequency}}$
Its can have values between 0 and 1</td></tr>
<tr><td>Percentage Frequency</td><td>$\frac{\text{Class frequency}}{\text{Total Frequency}} \times 100$</td></tr>
<tr><td>Frequency Dist. Diagram – Histogram</td><td>

It is a convenient way to represent FD
Comparison between frequency of two different classes possible
It is useful to calculate mode also

</td></tr>
<tr><td>Frequency Polygon</td><td>

Usually preferable for ungrouped frequency distribution
Can be used for grouped also but only if class lengths are even

</td></tr>
<tr><td>Ogives/ Cumulative Frequency</td><td>

This graph can be made by both type of Cumulative Frequency and called as Less than Ogive or More than Ogive
It can be used for calculating quartiles, median

</td></tr>
<tr><td>Frequency Curve</td><td>

It is a limiting form of Area Diagram (Histogram) or Frequency Polygon
It is obtained by drawing smooth and free hand curve though the mid points
Most used curve is Bell Shaped

</td></tr>
</table>

PYQ Jul 21 There are 200 employees in an office in which 150 were married. Total male employees were 160 out of which 120 were married. What was the number of female unmarried employees?

a. 30 b. 40 c. 50 d. 10

Ans: d

PYQ Dec 21 A student scored marks in five subject S1, S2, S3, S4 and S5 are 86, 79, 90, 88 and 89. If we need to draw a Pie chart to represent these markers, then what will be the Central angle for S3 is ___ degree

a. 103.2 b. 75 c. 105.6 d. 94.8

Ans: b

PYQ Dec 21

The following data relate to the marks of a group of students:

Marks	<10	<20	<30	<40	<50
F	15	38	65	84	100

How many students got marks more than 30?

a. 65 b. 50 c. 35 d. 43

Ans: d

MTP Nov 19

Let L be the lower class boundary of a class in a frequency distribution and m be the mid point of the class. Which one of the following is the higher class boundary of the class?

a. $m+\frac{m+2}{2}$ b. $L+\frac{m+L}{2}$ c. 2m-L d. None

Ans: c

Central Tendency and Dispersion

Past Trends

Attempt	Practical	Theory	Total
May 2018	3	4	7
Nov 2018	10	2	12
Jun 2019	9	3	12
Nov 2019	10	7	17
Nov 2020	4	5	9
Jan 2021	4	5	9
Jul 2021	11	1	12
Dec 2021	6	5	11
Jun 2022	6	3	9
Dec 2022	13	3	16
Jun 2023	0	15	15

Central Tendency - Basics

Meaning	• Central Tendency is the **tendency** of a given set of observations to **cluster** around a single **central or middle value.** • The **single value** that **represents** the given set of observations is described as a measure of central tendency.
Different Measures of Central Tendency	• Arithmetic Mean (AM) • Median (Me) • Mode (Mo) • Geometric Mean (GM) • Harmonic Mean (HM)
Types of Formula based Questions	• Discrete Observations • Simple Frequency Distribution • Grouped Frequency Distribution

Arithmetic Mean

Discrete Observations	$\bar{x} = \frac{x_1 + x_2 + x_3 + ... + x_n}{n}$ $\quad$ $\bar{x} = \frac{\sum x}{n}$
Frequency Distribution	$\bar{x} = \frac{\sum fx}{N}$ In case of simple frequency distribution: x = individual values In case of grouped frequency distribution: x = mid-point of class intervals N = total number of observations: $N = \sum f$
Assumed Mean / Step-Deviation Method	AM using assumed mean / step deviation method $\bar{x} = A + \frac{\sum fd}{N} \times C$ where $d = \frac{x - A}{C}$, A is assumed mean, C is class length
Property 1	If all the observations are constant, AM is also constant
Property 2	the algebraic sum of deviations of a set of observations from their AM is zero
Property 3	AM is affected both due to change of origin and scale If $y = a + bx$ then $\bar{y} = a + b\bar{x}$
Property 4	Combined AM $\bar{x}_c = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}$
General Review	• AM is best measure of central tendency • AM is based on all observations • AM is affected by sampling fluctuations • AM is amenable to mathematical property • AM cannot be used in case of open end classification

PYQ May 18 If the variables x and z are related by z = ax + b where a and b are constant, then $\bar{z} = a + b\bar{x}$

a. True b. False c. Both d. None

Ans: a

PYQ May 18 If each item is reduced by 15 then AM is

a. Reduced by 15 b. Increased by 15
c. Reduced by 10 d. None

Ans: a

PYQ Nov 18

The mean of 20 items of a data is 5 and if each item is multiplied by 3, then the new mean will be

a. 5 b. 10 c. 15 d. 20

Ans: c

PYQ Nov 18

The algebraic sum of the deviations of a set of values from their AM is

a. >0 b. =0 c. <0 d. None

Ans: b

PYQ Nov 18

If the frequencies of three series are 50, 60 and 90 and their means are 12, 15 and 20 respectively, then the mean of their composite series is

a. 16 b. 15.5 c. 16.5 d. 14.5

Ans: c

PYQ Jun 19

The AM of 15 observations is 9 and AM of first 9 observations is 11 then the AM of remaining observations is

a. 11 b. 6 c. 5 d. 9

Ans: b

Extra

If assumed mean is 419.5 and sum of product of frequency and deviation from assumed mean is -43. Total observations 308. Find the AM (given class length of distribution is 20)

a. 397.51 b. 410.66 c. 416.71 d. 432.55

Ans: c

PYQ Jul 21

The mean of n observations is x. If k is added to each observation, then the new mean is

a. k b. xk c. x-k d. x+k

Ans: d

PYQ Dec 21

If there are 3 observations 15, 20 and 25 then the sum of deviation of the observations from their AM is

a. 0 b. 5 c. -5 d. 10

Ans: a

PYQ Dec 21

If average mark for a group of 30 girls is 80 and for group of boys is 70 and combined average is 76, then how many are in the boy's group?

a. 21 b. 20 c. 22 d. 19

Ans: b

PYQ Jun 22

The mean of 20 observations is 38. If two observations are taken as 84 and 36 instead of 48 and 63 find new means.

e. 38.45 f. 41.15 g. 37.55 h. 40.05

Ans: c

Median

Discrete Observations	• If n = odd, then middle term • If n = even, average of two middle terms

<table>
<tr><td>Simple Frequency Distribution</td><td>• First make column of less than cumulative frequency
• Apply same formula as discrete</td></tr>
<tr><td>Grouped Frequency Distribution</td><td>Median in case of grouped frequency distribution
<table>
<tr><td>Step 1</td><td>Prepare a less than type cumulative frequency distribution</td></tr>
<tr><td>Step 2</td><td>Calculate $\frac{N}{2}$ and check between which class boundaries it falls and call it as Median Class</td></tr>
<tr><td>Step 3</td><td><table><tr><td>l_1</td><td>N_u</td><td>N_l</td><td>C</td></tr><tr><td>LCB of Median Class</td><td>Cum Freq. of Median Class</td><td>Cum. Freq. of Pre-Median Class</td><td>Class length of Median Class</td></tr></table></td></tr>
<tr><td>Step 4</td><td>Apply Formula
$Me = l_1 + \left(\frac{\frac{N}{2} - N_l}{N_u - N_l}\right) \times C$</td></tr>
</table></td></tr>
<tr><td>Property 1</td><td>For a set of observations, the sum of absolute deviations is minimum, when the deviations are taken from the median.
$\sum|x - Me|$ is minimum</td></tr>
<tr><td>Property 2</td><td>Median is also affected by both change of origin and scale.</td></tr>
<tr><td>General Review</td><td>• Median is also called as positional average
• Median is not based on all observations
• Median is not affected by sampling fluctuations
• Median is best measure of central tendency in case of open-end classification</td></tr>
</table>

PYQ Nov 18
PYQ Dec 21

The median of the data 5, 6, 7, 7, 8, 9, 10, 11, 11, 12, 15, 18, 18 and 19 is

a. 10.5 b. 10 c. 11 d. 11.5

Ans: a

PYQ Jun 19

Find median for the below distribution:

X	1	2	3	4	5	6
F	6	9	10	14	12	8

a. 3.5 b. 3 c. 4 d. 5

Ans: c

PYQ Nov 19

The deviations are minimum when taken from

a. Mean b. Median c. Mode d. None

Ans: b

PYQ Nov 19

Find the median of the following:

Class	0-10	10-20	20-30	30-40	40-50
Frequency	2	3	4	5	6

a. 35 b. 32 c. 36 d. 37.5

Ans: b

Partition Values

Meaning

- These may be defined as **values dividing** a given **set of observations** into number of **equal parts**
- When we want to divide the given set of observations into two equal parts, we consider median, similarly there are quartiles, deciles, percentiles

Name of PV	No. of equal parts	No. of PVs	Symbol
Median	2	1	Me
Quartile	4	3	Q_1, Q_2, Q_3
Decile	10	9	$D_1, D_2, \ldots, D_9$
Percentile	100	99	$P_1, P_2, \ldots, P_{99}$

Formula – Discrete Observations

- Rank Calculation $(n+1)p^{th}$ term
- Value of p depends on partition value

#	Median	Quartile	Decile	Percentile
First	1/2	1/4	1/10	1/100
Second		2/4	2/10	2/100
…				
Last		3/4	9/10	99/100

Quartiles Grouped FD

Quartiles in case of Grouped Frequency Distribution: Steps are like median with few modifications.

1st Quartile	3rd Quartile
Find Q_1 class using $\frac{N}{4}$	Find Q_3 class using $\frac{3N}{4}$
$Q_1 = l_1 + \left(\frac{\frac{N}{4} - N_l}{N_u - N_l}\right) \times C$	$Q_3 = l_1 + \left(\frac{\frac{3N}{4} - N_l}{N_u - N_l}\right) \times C$

Deciles Grouped FD

Deciles in case of Grouped Frequency Distribution: Steps are like median with few modifications.

		1st Decile: Find D_1 class using $\frac{N}{10}$; $D_1 = l_1 + \left(\frac{\frac{N}{10} - N_l}{N_u - N_l}\right) \times C$	9th Decile: Find D_9 class using $\frac{9N}{10}$; $D_9 = l_1 + \left(\frac{\frac{9N}{10} - N_l}{N_u - N_l}\right) \times C$
Percentiles Grouped FD	Percentiles in case of Grouped Frequency Distribution: Steps are like median with few modifications.	1st Percentile: Find P_1 class using $\frac{N}{100}$; $P_1 = l_1 + \left(\frac{\frac{N}{100} - N_l}{N_u - N_l}\right) \times C$	99th Percentile: Find P_{99} class using $\frac{99N}{10}$; $P_{99} = l_1 + \left(\frac{\frac{99N}{10} - N_l}{N_u - N_l}\right) \times C$

PYQ Nov 19

For 899, 999, 391, 384, 390, 480, 485, 760, 111, 240
Rank of Median is

a. 2.75 b. 5.5 c. 8.25 d. None

Ans: b

PYQ Nov 20

50th Percentile is equal to

a. Median b. Mode c. Mean d. None

Ans: a

PYQ Jun 22

The 3rd Decile for the numbers
15, 10, 20, 25, 18, 11, 9, 12 is

a. 13 b. 10.7 c. 11 d. 11.5

Ans: b

PYQ Jun 19

The QD of 6 numbers 15, 8, 36, 40, 38, 41 is equal to

a. 12.5 b. 25 c. 13.5 d. 37

Ans: c

Mode

Meaning	Mode is the **value** that **occurs the maximum** number of times
Special Thing about Mode	• If two or more observations are having maximum frequency then there are **multiple modes** [multimodal distribution] • If there are **exactly two** modes then distribution is called as **Bimodal** Distribution • If all observations are having same frequency then distribution has **no mode** • We can say that Mode is **not rigidly defined**

<table>
<tr><td>Grouped Frequency Distribution</td><td>

Find Modal Class: Class with highest frequency and obtain below values
<table>
<tr><td>f_{-1}</td><td>f_0</td><td>f_1</td></tr>
<tr><td>frequency of pre modal class</td><td>frequency of the modal class</td><td>frequency of the post modal class</td></tr>
</table>
Apply Formula
$$Mo = l_1 + \left(\frac{f_0 - f_{-1}}{2f_0 - f_{-1} - f_1} \right) \times C$$

</td></tr>
<tr><td>Property 1</td><td>If all the observations are constant, mode is also constant</td></tr>
<tr><td>Property 2</td><td>Mode is also affected both due to change of origin and scale</td></tr>
<tr><td>General Review</td><td>

Mode is not based on all observations
Mode is not rigidly defined
Mode is not amenable to Mathematical Property

</td></tr>
</table>

PYQ Nov 19

Find the mode of the following data:

Class	3-6	6-9	9-12	12-15	15-18	18-21
Frequency	2	5	10	23	21	12

a. 25 b. 4.6 c. 14.6 d. 13.5

Ans: c

PYQ Jan 21
PYQ Jul 21

If y = 3 + 4.5x and mode for x is 20, then the mode for y is

a. 3.225 b. 12 c. 24.5 d. 93

Ans: d

Relationship between Mean, Median and Mode

In case of Symmetric Distribution	**Mean = Median = Mode**
In case of Moderately Skewed Distribution (Empirical relationship)	**Mean – Mode = 3 (Mean – Median)**

PYQ May 18

Relation between Mean, Median and Mode is

a. Mean – Mode = 2 (Mean – Median)
b. Mean – Median = 3 (Mean – Mode)
c. Mean – Median = 2 (Mean – Mode)
d. Mean – Mode = 3 (Mean – Median)

Ans: d

PYQ Nov 18
PYQ Jun 19
PYQ Dec 21

If in a moderately skewed distribution, the values of mode and mean are 32.1 and 35.4 respectively then the value of median is

a. 34.3 b. 33.3 c. 34 d. 33

Ans: a

PYQ Jun 19

For a symmetric distribution

a. Mean = Median = Mode
b. Mode = 3 Median – 2 Mean
c. Mode = 1/3 Median = 1/2 Mean

d. None

Ans: a

PYQ Dec 21

For a moderately skewed distribution, the median is twice the mean, then the mode is _____ times the median.

a. 3 b. 2 c. 2/3 d. 3/2

Ans: b

Geometric Mean

Definition	For a given set of n **positive observations**, the geometric mean is defined as the n^{th} root of the product of the observations
Formula – Discrete	$G = \left(x_1 \times x_2 \times \ldots \times x_n\right)^{1/n}$
Formula – Frequency Distribution	$G = \left(x_1^{f_1} \times x_2^{f_2} \times \ldots \times x_n^{f_n}\right)^{1/N}$
Property 1	Logarithm of G for a set of observations is the AM of the logarithm of the observations $\log G = \frac{1}{n}\sum \log x$
Property 2	If all the observations are constant, GM is also constant
Property 3	If z = xy, then GM of z = GM of $x \times$ GM of y
Property 4	If z = x/y, then GM of $z = \frac{\text{GM of } x}{\text{GM of } y}$

PYQ Nov 18

The GM of 3, 6, 24, 48 is

a. 8 b. 12 c. 24 d. 6

Ans: b

PYQ Dec 21

If two variables are related by c = ab then GM of c is equal to

a. GM of a + GM of b
b. GM of a × GM of b
c. GM of a – GM of b
d. GM of a/ GM of b

Ans: b

Harmonic Mean

Definition	For a given set of **non-zero** observations, harmonic mean is defined as the **reciprocal of the AM of the reciprocals of the observation**

Formula – Discrete	$H = \frac{n}{\Sigma(\frac{1}{x})}$
Formula – Frequency Distribution	$H = \frac{N}{\Sigma(\frac{f}{x})}$
Property 1	If all observations are constant HM is also constant
Property 2	Combined $HM = \frac{n_1 + n_2}{\frac{n_1}{H_1} + \frac{n_2}{H_2}}$

PYQ Nov 20

Given the weights for the numbers 1, 2, 3, …, n is respectively $1^2, 2^2, 3^2, \ldots, n^2$ then weighted HM is

a. $\frac{2n+1}{4}$ b. $\frac{2n+1}{6}$ c. $\frac{2n+1}{3}$ d. $\frac{2n+1}{2}$

Ans: c

PYQ Nov 20

The HM of A and B is 1/3 and HM of C and D is 1/5. The HM of A, B, C and D is

a. 8/15 b. 1/4 c. 1/15 d. 5/3

Ans: b

PYQ Jan 21

If there are two groups with n_1 and n_2 observations and H_1 and H_2 are respective HMs, then HM of combined observations is

a. $\frac{n_1H_1 + n_2H_2}{n_1 + n_2}$ b. $\frac{n_1H_1 + n_2H_2}{H_1 + H_2}$

c. $\frac{n_1 + n_2}{n_1H_1 + n_2H_2}$ d. $\frac{(n_1 + n_2)H_1H_2}{n_1H_2 + n_2H_1}$

Ans: d

Use of GM and HM

Both	Both are used for calculating average rates
GM	Appropriate for rates having percentages
HM	Appropriate for rates other than percentages

PYQ Nov 20

A fire engine rushes to a place of fire accident with a speed of 110kmph and after the completion of operation returned to the base at a speed of 35kmph. The average speed per hour in per direction is obtained as _____ of speeds

a. AM b. GM c. HM d. None

Ans: c

Relationship between AM, GM, and HM

Relation	**Scenario** / **Relation** When all the observations are same — AM = GM = HM When observations are distinct — AM > GM > HM In question is silent — $AM \geq GM \geq HM$
Special Relation	If there are only two observations: $AM \times HM = (GM)^2$

PYQ Nov 20 If the AM and HM of two numbers are 6 and 9 respectively, then GM is

a. 7.35 b. 8.5 c. 6.75 d. None

Ans: a

Weighted Average

When to use	If the observations are not of equal importance and we need to treat observations according to their hierarchical importance, then we use Weighted Average	
Formulas	Weighted AM	$\frac{\sum wx}{\sum w}$
	Weighted GM	$\left(x_1^{w_1} \times x_2^{w_2} \times x_3^{w_3} \times \ldots \times x_n^{w_n}\right)^{\frac{1}{\sum w}}$
	Weighted HM	$\frac{\sum w}{\sum\left(\frac{w}{x}\right)}$

Measures of Dispersion

Meaning of Measure of Dispersion	• Dispersion for a given set of observations may be defined as • the **amount of deviation** of the observations, • usually, from an **appropriate** measure of **central tendency**	
Types of Measure of Dispersion	Absolute Measures of Dispersion	• These are with units • These are not useful for comparison of two variables with different units. • Example: Range, Mean Deviation, Standard Deviation, Quartile Deviation
	Relative Measures of Dispersion	• These are unit free measures • These are useful for comparison of two variables with different units.

	• Example: Coefficient of Range, Coefficient of Mean Deviation, Coefficient of variation, Coefficient of Quartile Deviation

Range

Discrete – Formula	$L - S$ where L: Largest Observation, S: Smallest Observation
Grouped Frequency Distribution – Formula	$L - S$ where Largest Observation = **UCB** of last class interval, Smallest Observation = **LCB** of first-class interval
Coefficient of Range	$\frac{L-S}{L+S} \times 100$
Property 1	• **Not affected** by change of **origin** • Affected by **change of scale (only value)** • **No impact of sign** of change of scale • Note: **Measure of Dispersion can never be negative**
General Review	• Not Based on All Observations • Easy to Compute

PYQ Nov 18 If the range of a set of values is 65 and maximum value in the set is 83, then the minimum value in the set is

a. 74 b. 9 c. 18 d. None

Ans: c

PYQ Jul 21 If the relationship between x and y is given by 2x + 3y = 10 and the range of y is 10, then what is the range of x?

a. 10 b. 18 c. 8 d. 15

Ans: d

PYQ Dec 21 The marks secured by 5 students in a subject are 82, 73, 69, 84, 66. What is the coefficient of Range

a. 0.12 b. 12 c. 120 d. 0.012

Ans: b

Mean Deviation

Meaning	• Mean deviation is defined as the • **arithmetic mean** of the • **absolute deviations** of the observations • from an **appropriate measure** of central tendency

Formula – Discrete	$MD_A = \frac{1}{n}\Sigma\lvert x - A\rvert$ where, A = Appropriate Central Tendency Measure
Formula – Frequency Distribution	$MD_A = \frac{1}{N}\Sigma f\lvert x - A\rvert$
Coefficient of Mean Deviation	Coefficient of Mean Deviation: $\frac{\text{Mean Deviation about A}}{A} \times 100$
Property 1	Mean Deviation takes its **minimum value** when deviations are taken from **Median**
Property 2	Change of Origin – **No Affect**, Change of Scale – **Affect of value not sign**
General Review	• Based on **all observations** • Improvement over Range • **Difficult to compute** • **Not amenable to Mathematical Property** because of usage of **Modulus**

PYQ Jan 21 Find the coefficient of mean deviation about mean for the data: 5, 7, 8, 10, 11, 13, 19

a. 17.28 b. 28.57 c. 32.11 d. 18.56

Ans: c

PYQ Jun 22 Mean Deviation about Mode from the data: 3, 10, 10, 4, 7, 18, 5

a. 4.39 b. 4.70 c. 4.14 d. 5.24

Ans: c

Standard Deviation

Meaning	• Improvement over Mean Deviation • It is defined as the **root mean square deviation** when the deviations are <u>taken from the AM</u> of the observations
Formula – Discrete	$\sigma_x = SD_x = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n}}$ $\sigma_x = SD_x = \sqrt{\frac{\Sigma x^2}{n} - (\bar{x})^2}$
Formula – Frequency Distribution	$\sigma_x = SD_x = \sqrt{\frac{\Sigma f(x-\bar{x})^2}{N}}$ $\sigma_x = SD_x = \sqrt{\frac{\Sigma fx^2}{N} - (\bar{x})^2}$

Coefficient of Variation	$\frac{SD_x}{\bar{x}} \times 100$
SD for any two numbers	$SD = \frac{Range}{2}$
SD for first n natural numbers	$s = \sqrt{\frac{n^2 - 1}{12}}$
Property 1	If all the observations are constant, SD is **ZERO**
Property 2	No effect of change of origin but affected by change of scale in the magnitude (ignore sign)
Property 3	$SD_c = \sqrt{\frac{n_1 s_1^2 + n_2 s_2^2 + n_1 d_1^2 + n_2 d_2^2}{n_1 + n_2}}$ $d_1 = \bar{x}_c - \bar{x}_1 \; d_2 = \bar{x}_c - \bar{x}_2$

PYQ May 18
PYQ Jun 19
If the SD of the 1st n natural number is $\sqrt{30}$
a. 19 b. 20 c. 21 d. None
Ans: a

PYQ Nov 18
If the variance of 5, 7, 9 and 11 is 4, then the coefficient of variation
a. 15 b. 25 c. 17 d. 19
Ans: b

PYQ Nov 18
Marks obtained by a student in monthly tests are 30, 35, 25, 20, 15. Find SD of marks
a. 25 b. $\sqrt{50}$ c. $\sqrt{30}$ d. 50
Ans: b

PYQ Jun 19
If variance is 100 and coefficient of variation is 20% then AM is
a. 60 b. 70 c. 80 d. 50
Ans: d

PYQ Jun 19
SD is _____ times of $\sqrt{MD \times QD}$
a. 2/3 b. 4/5 c. $\sqrt{\frac{15}{8}}$ d. $\sqrt{\frac{8}{15}}$
Ans: c

PYQ Jun 19
The sum of mean and SD of a series is a + b, if we add 2 to each observation of the series then the sum of mean and SD is
a. a + b + 2 b. 6 – a + b
c. 4 + a – b d. a + b + 4
Ans: a

PYQ Nov 19
Find SD of 1, 2, 3, 4, 5, 6, 7, 8, 9

PYQ Jan 21 a. 2.58 b. 60/9 c. 60/3 d. 3.20

Ans: a

PYQ Jul 21
PYQ Jun 22 If the numbers are 5, 1, 8, 7, 2 then the coefficient of variation is

a. 56.13% b. 59.13% c. 48.13% d. 44.13%

Ans: b

PYQ Jun 22 AM and Coefficient of variation of x is 10 and 40. What is the variance of 30-2x

a. 64 b. 56 c. 49 d. 81

Ans: a

Quartile Deviation

Formula	$QD_x = \frac{Q_3 - Q_1}{2}$
Calculation	Quartiles are calculated same as we studied in Central Tendency
Coefficient of Quartile Deviation	$\frac{Q_3 - Q_1}{Q_3 + Q_1} \times 100$
General Review	• It is the **best measure** of dispersion for **open-end** classification • It is also **less affected** due to sampling fluctuations • Like other measures of Dispersion, **QD** is also not affected by change of origin but affected by scale ignoring sign
Relationship between SD, MD and QD	$4SD = 5MD = 6QD$ or $SD : MD : QD = 15 : 12 : 10$

PYQ Jun 19 Coefficient of QD is 1/4 then Q_3 / Q_1 is

a. 5/3 b. 4/3 c. 3/4 d. 3/5

Ans: a

OTM: Correlation and Regression

Past Trends

Attempt	Practical	Theory	Total
May 2018	1	6	8
Nov 2018	3	2	5
Jun 2019	5	1	6
Nov 2019	3	1	4
Nov 2020	0	3	3
Jan 2021	1	4	5
Jul 2021	5	1	6
Dec 2021	2	2	4
Jun 2022	2	4	6
Dec 2022	4	1	5
June 2023	4	1	5

Bivariate Data

Definition	• When data are collected on two variables **simultaneously**, they are known as **bivariate data** • and the corresponding frequency distribution, derived from it, is known as **Bivariate Frequency Distribution**
Marginal Distribution	• It is the frequency distribution of **one variable** (x or y) across the other variable's **full range of values** • **Number of Marginal Distribution = 2**
Conditional Distribution	• It is the frequency distribution of **one variable** (x or y) across a **particular sub-population** of the other variable. • **No. of Conditional Distributions = m + n** m = no. of class interval of x n = no. of class interval of y • Number of Cells = m × n

MTP Nov 21 For a 4 x 7 classification of bivariate data, the maximum number of conditional distributions is:

a. 11 b. 28 c. 35 d. None

Ans: a

MTP Oct 21 For a p x q bivariate frequency table, the maximum number of marginal distributions is

a. p b. p+q c. 1 d. 2

Ans: d

Scatter Diagram

Theory about Scatter Diagram	• It helps us to find **Nature** and **Relative Strength** of Correlation • It is useful for **Non-Linear** Correlation also • It **cannot** be used to determine **value** • Diagrams are **time taking**

PYQ Nov 20

If the plotted points in a scatter diagram lie from upper left to lower right, then correlation is

a. Positive b. Negative c. Zero d. None

Ans: b

PYQ Nov 20

If the plotted points in a scatter diagram lie from upper left to lower right, then correlation is

a. Find the type of correlation
b. Identify whether variables correlated or not
c. Determine the linear or non-linear correlation
d. Find the numerical value of correlation coefficient

Ans: d

PYQ June 22

If the plotted point in a scatter diagram lie from lower left to upper right then correction is:

a. Positive b. Negative c. Zero d. None

Ans: a

Karl Pearson's Correlation Coefficient

Formula	$r_{xy} = \frac{Cov(x,y)}{(\sigma_x \times \sigma_y)}$
Formula of Covariance	$Cov(x,y) = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{n}$ or $\frac{\Sigma xy}{n} - \bar{x}.\bar{y}$
Property 1	The Coefficient of Correlation is a **unit-free measure**
Property 2	Value lies from **-1 to +1**
Property 3	Change of Origin: No impact Change of Scale: No impact of value, but if change of scale of both variables are of **different sign** then **sign of r** will also change

PYQ May 18

A relationship $r^2 = 1 - \frac{500}{300}$ is not possible. This statement is

a. True b. False c. Both d. None

Ans: a

PYQ Nov 18

If the correlation coefficient between variables X and Y is 0.5, then the correlation coefficient between the variables $2x-4$ and $3-2y$ is

a. 1 b. 0.5 c. -0.5 d. 0

Ans: c

PYQ Jun 19

Given that

X	-3	-3/2	0	3/2	3
Y	9	9/4	0	9/4	9

The Karl Pearson's Correlation Coefficient is

a. Positive b. Zero c. Negative d. None

Ans: b

PYQ Nov 19

What is the correlation coefficient from the following data:

x	1	2	3	4	5
y	5	4	3	2	6

a. 0 b. -0.75 c. -0.85 d. 0.82

Ans: a

PYQ Jan 21

For the set of observations {(1,2),(2,5),(3,7),(4,8),(5,10)} the value of Karl Pearson's coefficient of correlation is approximately given by

a. 0.755 b. 0.655 c. 0.525 d. 0.985

Ans: d

PYQ Jan 21

The coefficient of correlation between x and y is 0.5, the covariance is 16 and variance of x is 16, then SD of y is

a. 4 b. 8 c. 16 d. 64

Ans: b

Set B – Q3

If the covariance between two variables is 20 and the variance of one of the variables is 16. What would be the variance of the other variable?

a. $(\sigma_y)^2 \geq 25$ b. More than 10 c. Less than 10 d. More than 1.25

Ans: a

Spearman's Rank Correlation Coefficient

Usage	• find the level of **agreement (or disagreement)** between two judges so far as assessing a **qualitative characteristic (attribute)** is concerned • Use in case of ranks
Formula (Regular)	$r_R = 1 - \frac{6\Sigma d^2}{n(n^2-1)}$ d = difference in ranks
Formula (In case of Tie)	Spearman's Rank Correlation Coefficient (in case of tied values) $r_R = 1 - \frac{6\left(\Sigma d^2 + A\right)}{n(n^2-1)}$ here A is adjustment value

	$A = \frac{\Sigma(t^3 - t)}{12}$ where t = tie length (calculate t value for each of the ties)

PYQ Jun 19

Determine the Spearman's rank correlation coefficient from the given data $\sum d^2 = 30$ and n = 10

a. 0.82 b. 0.32 c. 0.40 d. None

Ans: a

ICAI Example

Compute rank correlation coefficient between Eco and Stats Marks

Eco	80	56	50	48	50	62	60
Stats	90	75	75	65	65	50	65

a. 0.2053 b. 0.15 c. 0.40 d. None

Ans: b

Set B Q12

While computing rank correlation coefficient between profits and investment for 10 years of a firm, the difference in rank for a year was taken as 7 instead of 5 by mistake and the value of rank correlation coefficient was computed as 0.80. What would be the correct value of rank correlation coefficient after rectifying the mistake?

a. 0.3 b. 0.945 c. 0.25 d. 0.28

Ans: b

Coefficient of Concurrent Deviations

Usage	A very **quick, simple** and **casual** method of finding correlation when we are not serious about the magnitude of the two variables
Formula	$r_c = \pm\sqrt{\pm\left(\frac{2c-m}{m}\right)}$ where c is number of concurrent deviations (same direction) m is number of pairs compared (equals to n-1)

MTP Jun 22

For 10 pairs of observations, number of concurrent deviations was found to be 4. What is the value of the coefficient of concurrent deviation?

a. $\sqrt{0.2}$ b. 1/3 c. -1/3 d. $-\sqrt{0.2}$

Ans: c

PYQ Jun 22

If concurrent coefficient is $\frac{1}{\sqrt{3}}$ and number of concurrent deviations is 6. Find the number of pairs of data?

a. 9 b. 8 c. 10 d. 11

Ans: c

Regression Basics

Meaning	Estimation of one variable for a **given value** of another variable on the basis of an **average mathematical relationship** between the two variables
Requirements	• Estimation of Y when X is given • Estimation of X when Y is given
General Points	Perfect Correlation: • When linear relationship exists between two variables, correlation is perfect. • Perfect Correlation is represented by a linear equation and this equation can be used for regression purpose directly. • Same equation can be used in both ways Imperfect Correlation: • In case of imperfect correlation there is no definite line and equation • We will use method of least square to estimate both regression lines
Formula of Regression Equations/ Lines	Estimation of Y when X is given: • Use Regression line of **Y on X** • Equation Format: $Y-\bar{Y}=b_{yx}(X-\bar{X})$ b_{yx} is regression coefficient of Y on X Estimation of X when Y is given: • Use Regression line of **X on Y** • Equation Format: $X-\bar{X}=b_{xy}(Y-\bar{Y})$ b_{xy} is regression coefficient of X on Y
Regression Coefficient	Regression Coefficient of Y on X: $b_{yx}=r.\frac{SD_y}{SD_x}$ and $b_{yx}=\frac{cov(x,y)}{(SD_x)^2}$ Regression Coefficient of X on Y: $b_{xy}=r.\frac{SD_x}{SD_y}$ and $b_{xy}=\frac{cov(x,y)}{(SD_y)^2}$
Property 1	Change of Origin/ Scale for Regression Coefficients: Origin no impact, Scale impact of both magnitude and sign. $b_{vu}=b_{yx}\times\frac{\text{change of scale of y}}{\text{change of scale of x}}$ $b_{uv}=b_{xy}\times\frac{\text{change of scale of x}}{\text{change of scale of y}}$
Property 2	Two regression lines (if not identical) will intersect at the point [means] $(\bar{x},\bar{y})$
Property 3	Correlation Coefficient is the GM of regression coefficients: $r_{xy}=\pm\sqrt{b_{xy}\times b_{yx}}$ Note: r_{xy}, b_{xy}, b_{yx} all will have same sign

PYQ Nov 18
PYQ Nov 19

If the two lines of regression are $x+2y-5=0$ and $2x+3y-8=0$ then the regression line of y on x is

a. $x+2y-5=0$ b. $2x+3y-8=0$ c. Both d. None

Ans: a

PYQ Jul 21
PYQ Dec 22

If b_{yx} = -1.6 and b_{xy} = -0.4, then r_{xy} will be

a. 0.4 b. -0.8 c. 0.64 d. 0.8

Ans: b

PYQ Nov 18

If the two regression lines are $3x=y$ and $8y=6x$, then the value of correlation coefficient is

a. 0.5 b. -0.5 c. 0.75 d. -0.80

Ans: a

PYQ Jun 19

If the regression line of y on x is given by y = x + 2 and Karl Pearson's coefficient of correlation is 0.5 then $\left(\sigma_y / \sigma_x\right)^2$ is

a. 9 b. 2 c. 4 d. 3

Ans: c

PYQ Jul 21

If the slope of regression line is calculated to be 5.5 and the intercept is 15 then the value of Y is if x = 6

a. 88 b. 48 c. 18 d. 78

Ans: b

Probable Error

Formula	Probable Error in correlation: $0.6745\times\frac{1-r^2}{\sqrt{N}}$
Use	• Correlation is calculated using sample, value for sample may differ from population, this difference is probable error • If there is significant probable error, there is no evidence of real correlation
Limits of Sample Correlation Coefficient	$r\pm PE$
How to check evidence of Correlation using PE	**Case** / **Conclusion** If r is less than PE / There is no evidence of correlation If r is greater than six times of PE / The presence of correlation is certain Since r lies from -1 to +1 / PE can never be negative

PYQ Jun 19

Find the probable error if $r=\frac{2}{\sqrt{10}}$ and n = 36

a. 0.6745 b. 0.06745 c. 0.5287 d. None

Ans: b

Coefficient of Determination and Non-Determination

Coefficient of Determination Accounted Variance/ Explained Variance	$(r_{xy})^2$
Coefficient of Non-Determination Unaccounted Variance/ Unexplained Variance	$1-(r_{xy})^2$

MTP Nov 18 If the two regression co-efficient are 4 and 0.16 the percentage of unexplained variation is:

a. 64 b. 36 c. 54 d. 46

Ans: b

MTP Nov 18 If the coefficient of correlation between two variables is 0.7 then the percentage of variation accounted for is

a. 49% b. 30% c. 51% d. 36%

Ans: a