4. Chp13 Statistical Description of Data

01 May 2024 10:57

SDOD One Shot June...

One Shot Revisions

- Live Revision on YouTube
- Coverage – Concepts, All IMP MCQs
- One Video for One Chapter
- Max Coverage in Min. Time
- Free PDF of Revision NOTES

YT Revisions – Phase I

Date	Day	Chapter Name	Category	Marks	Time
25-Apr-24	Thu	One Shot - Blood Relations	A	6	11.30 AM
27-Apr-24	Sat	One Shot - Maths for Finance	A	12	11.30 AM
28-Apr-24	Sun	One Shot - Seating Arrangements	A	4	11.30 AM
30-Apr-24*	Tue	One Shot - Statistical Description of Data	A	6	11.30 AM
02-May-24	Thu	One Shot - Direction Test	A	5	11.30 AM
04-May-24	Sat	One Shot - Central Tendency & Dispersion	A	12	11.30 AM
05-May-24	Sun	One Shot - Number Series Coding Decoding	A	5	11.30 AM
07-May-24	Tue	One Shot - Correlation Regression	B	5	11.30 AM
09-May-24	Thu	One Shot - Index Numbers	B	6	11.30 AM

*rescheduled to 1-May-24

YT Revisions – Phase II

Date	Day	Chapter Name	Category	Marks	Time
21-May-24	Tue	One Shot - Equation	B	4	TBD
23-May-24	Thu	One Shot - Linear Inequalities	B	1	TBD
25-May-24	Sat	One Shot - Ratio, Proportion, Indices, Logarithm	B	6	TBD
26-May-24	Sun	One Shot - Sequence and Series	B	4	TBD
28-May-24	Tue	Theory hai Zaroori	Special		TBD
30-May-24	Thu	Theory hai Zaroori	Special		TBD
01-Jun-24	Sat	OTM Permutations and Combinations	C	5	TBD
02-Jun-24	Sun	OTM Set Relation Functions	C	4	TBD
04-Jun-24	Tue	OTM Probability	C	6	TBD
06-Jun-24	Thu	OTM Theoretical Distribution	C	5	TBD
08-Jun-24	Sat	OTM Limit Continuity	D		TBD
09-Jun-24	Sun	OTM Calculus	D	4	TBD
11-Jun-24	Tue	Formular Marathon Maths	Special		TBD
13-Jun-24	Thu	Formular Marathon Stats	Special		TBD

ab mushkil nahi kuch bhi,
nahi kuch bhi

Stay Connected

let's get started.

Chapter 13 – Statistical Description of Data

Past Trends

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

<table>
<tr><td>Definition of Statistics</td><td>

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

</td></tr>
<tr><td>Origin of Word</td><td>
<table>
<tr><th>Language</th><th>Actual Word</th><th>Memorize by</th></tr>
<tr><td>Latin</td><td>Status</td><td>Latus</td></tr>
<tr><td>Italian</td><td>Statista</td><td>Pasta</td></tr>
<tr><td>German</td><td>Statistic</td><td>Breadstick</td></tr>
<tr><td>French</td><td>Statistique</td><td>Barbeque</td></tr>
</table>
</td></tr>
<tr><td>Publication</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

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

	• Industry and Commerce: Profit maximization using business data – sales, purchase, market etc. by consulting experts
Limitation of Statistics	• 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.
Data	• Quantitative information shown as number • Primary: first time collected by agency/ investigator • Secondary: collected data used by different person/ agency
Variable	• Measurable Data – Value can vary **Discrete Variable**: • When a variable assumes a finite or countably infinite isolated values. • Example: no. of petals in a flower, no. of road accident in locality **Continuous Variable**: • When a variable assumes any value from the given interval (can also be in decimals, fractions). • Example: height, weight, sale, money

Special point about Money	• In general, Money is a discrete variable • However, from commercial point of view, money is a continuous variable.
Attribute	• Qualitative Characteristics. Example: gender of a baby, the nationality of a person, the colour of a flower etc.
Collection of Primary Data – Interview Method	**Personal Interview**: • Where data is collected directly from respondents. • Highly Accurate – Low Coverage • Example: Natural Calamity, Door to Door Survey **Indirect Interview**: • When reaching respondent is difficult, data is collected by contacting associated persons. • Highly Accurate – Low Coverage • Example: Rail accident **Telephone Interview**: • Data is collected over phone • Quick and non-expensive method • Low Accuracy – High Coverage
Collection of Primary Data – Mailed Questionnaire Method	• 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

Collection of Primary Data – Observation Method	• 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.
Collection of Primary Data – Questionnaire Filled and sent by Enumerators	• 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.
Sources of Secondary Data	**International Sources**: World Health Organization (WHO), International Monetary Fund (IMF), International Labor Organization (ILO), World Bank **Government Sources**: 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 **Private or Quasi-govt. sources**: Indian Statistical Institute (ISI), Indian Council of Agriculture, NCERT

Scrutiny of Data	• 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.
Presentation of Data – Classification / Organization of Data	**Classification or Organisation**: putting data in a neat, precise, and condensed form, making it comparable, suitable for analysis, more understandable.

Chronological/ Temporal/ Time Series Data	• Data arranged based on Time • Example: Revenues YoY i.e year on year
Geographical or Spatial Series Data	• Arrangement based on regions • Example: Country wise Revenue of a global company
Qualitative or Ordinal Data	• Based on some attribute • Nationality Wise Medal Winners in Olympics
Quantitative or Cardinal Data	• Based on some variable • Example: Frequency Distribution of a Data

Mode of Presentation of Data – Textual	• 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.

Mode of Presentation of Data – Tabular Form	• 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
Components of Table	

Description	Name of Component of Table
Entire Upper Part	Box Head
Upper Part describing columns and sub-columns	Caption
Left part of the table describing rows	Stub
Main Data of Table	Body
Source of Data at the bottom of Table	Footnote

Mode of Presentation of Data – Diagrams	• Can be used by educated and uneducated section of society • Hidden trend can be traced • If priority is accuracy, then tabulation is better

Line Diagram	• 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
Bar Diagram	• 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 • 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 =** $\frac{(\text{segment value} \times 360°)}{(\text{total value})}$

ICAI SM

(1) *Which of the following statements is false?*

a. *Statistics is derived from the Latin word 'Status'*
b. *Statistics is derived from the Italian word 'Statista'*
c. *Statistics is derived from the French word 'Statistik'*
d. *None of these.*

ICAI SM

(2) *Statistics is defined in terms of numerical data in the*

a. *Singular sense* b. *Plural sense*
c. *Either (a) or (b)* d. *Both (a) & (b)*

ICAI SM

(3) *Statistics is applied in*

a. *Economics*
b. *Business management*
c. *Commerce and industry*
d. *All these.*

ICAI SM

(4) *Statistics is concerned with*

a. *Qualitative information*
b. *Quantitative information*
c. *(a) or (b)*
d. *Both (a) and (b)*

Stats deals with quant.

ICAI SM

(5) *An attribute is*

a. *A qualitative characteristic*
b. *A quantitative characteristic*
c. *A measurable characteristic*
d. *All these.*

ICAI SM

(6) *Annual income of a person is*

a. *An attribute*
b. *A discrete variable*
c. *A continuous variable*
d. *(b) or (c).*

ICAI SM

(7) *Marks of a student is an example of*

a. *An attribute*
b. *A discrete variable*
c. *A continuous variable*
d. *None of these.*

ICAI SM

(8) *Nationality of a student is*

a. *An attribute*
b. *A continuous variable*
c. *A discrete variable*
d. *(a) or (c).*

ICAI SM

(9) *Drinking habit of a person is*

a. *An attribute*
b. *A variable*
c. *A discrete variable*
d. *A continuous variable.*

ICAI SM

(10) *Age of a person is*

a. *An attribute*
b. *A discrete variable*
c. *A continuous variable*
d. *A variable.*

ICAI SM

(11) *Data collected on religion from the census reports are*

a. *Primary data*
b. *Secondary data*
c. *Sample data*
d. *(a) or (b).*

ICAI SM

(12) The data collected on the height of a group of students after recording their heights with a measuring tape are

a. Primary data
b. Secondary data
c. Discrete data
d. Continuous data

ICAI SM

(13) The primary data are collected by

a. Interview method
b. Observation method
c. Questionnaire method
d. All these.

ICAI SM

(14) The quickest method to collect primary data is

a. Personal interview
b. Indirect interview
c. Telephone interview
d. By observation.

ICAI SM

(15) The best method to collect data, in case of a natural calamity, is

a. Personal interview
b. Indirect interview
c. Questionnaire method
d. Direct observation method.

ICAI SM

(16) In case of a rail accident, the appropriate method of data collection is by

a. Personal interview
b. Direct interview
c. Indirect interview
d. All these.

ICAI SM

(17) Which method of data collection covers the widest area?

a. Telephone interview method
b. Mailed questionnaire method
c. Direct interview method
d. All these.

ICAI SM

(18) The amount of non-responses is maximum in

a. Mailed questionnaire method
b. Interview method
c. Observation method
d. All these.

ICAI SM

(19) Some important sources of secondary data are

a. International and Government sources
b. International and primary sources
c. Private and primary sources
d. Government sources.

ICAI SM

(20) Internal consistency of the collected data can be checked when

a. Internal data are given
b. External data are given
c. Two or more series are given
d. A number of related series are given.

ICAI SM

(21) The accuracy and consistency of data can be verified by

a. Internal checking
b. External checking
c. Scrutiny
d. Both (a) and (b).

ICAI SM

(22) *The mode of presentation of data are*

a. *Textual, tabulation and diagrammatic*
b. *Tabular, internal and external*
c. *Textual, tabular and internal*
d. *Tabular, textual and external.*

ICAI SM

(23) *The best method of presentation of data is*

a. *Textual* b. *Tabular*
c. *Diagrammatic* d. *(b) and (c).*

ICAI SM

(24) *The most attractive method of data presentation is*

a. *Tabular* b. *Textual*
c. *Diagrammatic* d. *(a) or (b).*

ICAI SM

(25) *For tabulation, 'caption' is*

a. *The upper part of the table*
b. *The lower part of the table*
c. *The main part of the table*
d. *The upper part of a table that describes the column and sub-column.*

ICAI SM

(26) *'Stub' of a table is the*

a. *Left part of the table describing the columns*
b. *Right part of the table describing the columns*
c. *Right part of the table describing the rows*
d. *Left part of the table describing the rows.*

ICAI SM

(27) *The entire upper part of a table is known as*

a. *Caption* b. *Stub*
c. *Box head* d. *Body.*

ICAI SM

(28) *The unit of measurement in tabulation is shown in*

a. *Box head* b. *Body*
c. *Caption* d. *Stub.*

ICAI SM

(29) *In tabulation source of the data, if any, is shown in the*

a. *Footnote* b. *Body*
c. *Stub* d. *Caption.*

ICAI SM

(30) *Which of the following statements is untrue for tabulation?*

a. *Statistical analysis of data requires tabulation*
b. *It facilitates comparison between rows and not columns*
c. *Complicated data can be presented*
d. *Diagrammatic representation of data requires tabulation.*

ICAI SM

(31) *Hidden trend, if any, in the data can be noticed in*

a. *Textual presentation*
b. *Tabulation*
c. *Diagrammatic representation*
d. *All these.*

ICAI SM

(32) *Diagrammatic representation of data is done by*

a. *Diagrams* b. *Charts*
c. *Pictures* d. *All these.*

ICAI SM

(33) *The most accurate mode of data presentation is*

a. *Diagrammatic method*
b. *Tabulation*
c. *Textual presentation*
d. *None of these.*

ICAI SM

(34) *The chart that uses logarithm of the variable is known as*

a. *Line chart*
b. *Ratio chart*
c. *Multiple line chart*
d. *Component line chart.*

ICAI SM

(35) Multiple line chart is applied for
- a. Showing multiple charts
- b. Two or more related time series when the variables are expressed in the same unit
- c. Two or more related time series when the variables are expressed in different unit
- d. Multiple variations in the time series.

ICAI SM

(36) Multiple axis line chart is considered when
- a. There is more than one time series
- b. The units of the variables are different
- c. (a) or (b)
- d. (a) and (b).

ICAI SM

(37) Horizontal bar diagram is used for
- a. Qualitative data
- b. Data varying over time
- c. Data varying over space
- d. (a) or (c).

ICAI SM

(38) Vertical bar diagram is applicable when
- a. The data are qualitative
- b. The data are quantitative
- c. When the data vary over time
- d. (a) or (c).

Note: Error in options, option a and b should be swapped.

ICAI SM

(39) Divided bar chart is considered for
- a. Comparing different components of a variable
- b. The relation of different components to the table
- c. (a) or (b)
- d. (a) and (b).

ICAI SM

(40) In order to compare two or more related series, we consider
- a. Multiple bar chart
- b. Grouped bar chart
- c. (a) or (b)
- d. (a) and (b).

ICAI SM

(41) Pie-diagram is used for
- a. Comparing different components and their relation to the total
- b. Representing qualitative data in a circle
- c. Representing quantitative data in circle
- d. (b) or (c).

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

Classification of Grouped of Frequency Distribution

Mutually Exclusive / Overlapping Classification

Class	LCL	UCL
10-20	10	20
20-30	20	30
30-40	30	40

- 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.

Mutually Inclusive / Non-Overlapping Classification

Class	LCL	UCL
10-19	10	20
20-29	20	30
30-39	30	40

- 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.

Class Boundary

In case of Exclusive / Overlapping Classification

Class Boundary = Class Limit

Class	LCL	UCL	LCB	UCB
10-20	10	20	10	20
20-30	20	30	20	30
30-40	30	40	30	40

In case of Inclusive / Overlapping Classification

Lower Class Boundary
LCB = LCL – 0.5
UCB = UCL + 0.5

Class	LCL	UCL	LCB	UCB
10-19	10	19	9.5	19.5
20-29	20	29	19.5	29.5
30-39	30	39	29.5	39.5

Mid-Point / Class Mark / Mid Value of Class Interval

$$\frac{LCL+UCL}{2} \qquad \frac{LCB+UCB}{2}$$

- Useful in calculation of AM, GM, HM, SD in case of grouped frequency distribution

Class Length/ Width or Size

UCB – LCB only

Cumulative Frequency

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

Class Interval	Freq.	UCB	Less than type CF	More than type CF	Total of both CF
44-48	3	48.5	3	33	36
49-53	4	53.5	7	29	36
54-58	5	58.5	12	24	36
59-63	7	63.5	19	17	36
64-68	9	68.5	28	8	36
69-73	8	73.5	36	0	36
Total	36				

Frequency Density

$$\frac{\text{Class Frequency}}{\text{Class Length of class}}$$

Relative Frequency

$$\frac{\text{Class frequency}}{\text{Total Frequency}}$$

It can take values between 0 and 1 exclusive

Percentage Frequency

$$\frac{\text{Class frequency}}{\text{Total Frequency}} \times 100$$

Frequency Dist. Diagram – Histogram

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

Frequency Polygon	• *Usually preferable for ungrouped frequency distribution* • *Can be used for grouped also but only if class lengths are even*
Ogives/ Cumulative Frequency	• *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*
Frequency Curve	• *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*

CA FOUNDATION QUANTITATIVE APTITUDE | CHAPTER 13

Theory Concepts not in ICAI Study Material

Descriptive Statistics vs Inferential Statistics	*Descriptive Statistics*	• *When a statistic is used only to describe scores in a sample (and not used to make inferences about populations), that is a descriptive use.* • *For example, a teacher may compute the mean test score for her class. She simply uses this number to think about how well her students did.*
	Inferential Statistics	• *Inferential use of statistics occurs when a data analyst uses information from a sample (such as a mean or correlation) to make inferences or guesses about values of the corresponding mean or correlation in a* ***population*** • *Statistical inference only works well when sample sizes are reasonably large* • *For example, in a political poll, a polling organization obtains the percent of people who state an intention to vote for Candidate X for a sample of votes, perhaps selected by random digit telephone dialing from the population of all registered voters.*

CA FOUNDATION QUANTITATIVE APTITUDE | CHAPTER 13

Simple Classification vs Manifold Classification	*Simple Classification*	• *When based on only one attribute, the given data is classified into two classes, which is known as Simple Classification.* • *For example, when the population is divided into literate and illiterate, it is a simple classification.*
	Manifold Classification	• *When based on more than one attribute, the given data is classified into different classes, and then sub-divided into more sub-classes, which is known as Manifold Classification.* • *For example, when the population is divided into literate and illiterate, then sub-divided into male and female, and further sub-divided into married and unmarried, it is a manifold classification.*
Ideographs/ Pictograms	• *A symbol that represents an idea or a thing, rather than the sounds of a word, is called an ideograph.* • *A smiley face emoji is an ideograph that represents happiness.* • *Many street signs are ideographs, meant to convey a specific meaning without using any words.*	

Qualitative Data Types	*Nominal Data*	• *Nominal data are used to **label** variables without any quantitative value.* • *Common examples include hair color, nationalities, names of people, and so on.* • *Nominal means Name means Labels*
	Ordinal Data	• *Ordinal word comes from order* • *Ordinal scales are often used for measures of satisfaction, happiness, and so on.* • *Example: high-low-medium, strong-weak, etc.*
Research Data	***Research data can be placed into two broad categories:***	
	Quantitative data	• *Quantitative data are used when a researcher is trying to quantify a problem, or address the "what" or "how many" aspects of a research question.* • *It is data that can either be counted or compared on a numeric scale. For example, it could be the number of first year students at Macalester, or the ratings on a scale of 1-4 of the quality of food served at Cafe Mac. This data are usually gathered using instruments, such as a questionnaire which includes a ratings scale or a thermometer to collect weather data. Statistical analysis software, such as SPSS, is often used to analyze quantitative data.*

	Qualitative data	• *Qualitative data describes qualities or characteristics.* • *It is collected using questionnaires, interviews, or observation, and frequently appears in narrative form.* • *For example, it could be notes taken during a focus group on the quality of the food at Cafe Mac, or responses from an open-ended questionnaire.* • *Qualitative data may be difficult to precisely measure and analyze. The data may be in the form of descriptive words that can be examined for patterns or meaning, sometimes through the use of coding.*
Sturges' Rule	*In the early 20th century, German statistician Herbert Sturges formulated a method (now called Sturges' Rule) of choosing the optimum number of bins in a histogram that minimize the potential for these pitfalls. His formula is simple:* $k = 1 + 3.322 \log n$ *where:* *k = the number of bins* *n = the number of observations in the data set.*	

ICAI SM

(42) *A frequency distribution*

a. Arranges observations in an increasing order
b. Arranges observation in terms of a number of groups
c. Relates to a measurable characteristic
d. All these.

Note: Given answer (b) in SM is incorrect

ICAI SM

(43) *The frequency distribution of a continuous variable is known as*

a. Grouped frequency distribution
b. Simple frequency distribution
c. (a) or (b)
d. (a) and (b).

ICAI SM

(44) *The distribution of shares is an example of the frequency distribution of*

a. A discrete variable
b. A continuous variable
c. An attribute
d. (a) or (c).

ICAI SM

(45) *The distribution of profits of a blue-chip company relates to*

a. *Discrete variable*
b. *Continuous variable*
c. *Attributes*
d. *(a) or (b).*

ICAI SM

(46) *Mutually exclusive classification*

a. *Excludes both the class limits*
b. *Excludes the upper class limit but includes the lower class limit*
c. *Includes the upper class limit but excludes the upper class limit*
d. *Either (b) or (c).*

ICAI SM

(47) *Mutually inclusive classification is usually meant for*

a. *A discrete variable*
b. *A continuous variable*
c. *An attribute*
d. *All these.*

ICAI SM

(48) *Mutually exclusive classification is usually meant for*

a. *A discrete variable*
b. *A continuous variable*
c. *An attribute*
d. *Any of these.*

ICAI SM

(49) *The LCB is*

a. *An upper limit to LCL*
b. *A lower limit to LCL*
c. *(a) and (b)*
d. *(a) or (b).*

ICAI SM

(50) *The UCB is*

a. *An upper limit to UCL*
b. *A lower limit to LCL*
c. *Both (a) and (b)*
d. *(a) or (b).*

ICAI SM

(51) *length of a class is*

a. *The difference between the UCB and LCB of that class*
b. *The difference between the UCL and LCL of that class*
c. *(a) or (b)*
d. *Both (a) and (b).*

ICAI SM

(52) *For a particular class boundary, the less than cumulative frequency and more than cumulative frequency add up to*

a. *Total frequency*
b. *Fifty per cent of the total frequency*
c. *(a) or (b)*
d. *None of these.*

ICAI SM

(53) *Frequency density corresponding to a class interval is the ratio of*

a. *Class frequency to the total frequency*
b. *Class frequency to the class length*
c. *Class length to the class frequency*
d. *Class frequency to the cumulative frequency.*

ICAI SM

(54) *Relative frequency for a particular class*

a. *Lies between 0 and 1*
b. *Lies between 0 and 1, both inclusive*
c. *Lies between –1 and 0*
d. *Lies between –1 to 1.*

ICAI SM

(55) *Mode of a distribution can be obtained from*

a. *Histogram*
b. *Less than type ogives*
c. *More than type ogives*
d. *Frequency polygon*

ICAI SM

(56) *Median of a distribution can be obtained from*

a. *Frequency polygon*
b. *Histogram*
c. *Less than type ogives*
d. *None of these.*

Note: Correct option c should be ogives not less than ogives.

ICAI SM

(57) *A comparison among the class frequencies is possible only in*

a. *Frequency polygon*
b. *Histogram*
c. *Ogives*
d. *(a) or (b).*

ICAI SM

(58) *Frequency curve is a limiting form of*

a. *Frequency polygon*
b. *Histogram*
c. *(a) or (b)*
d. *(a) and (b).*

Note: Given ans d in SM is incorrect.

ICAI SM

(59) *Most of the commonly used frequency curves are*

a. *Mixed*
b. *Inverted J-shaped*
c. *U-shaped*
d. *Bell-shaped.*

ICAI SM

(60) *The distribution of profits of a company follows*

a. *J-shaped frequency curve*
b. *U-shaped frequency curve*
c. *Bell-shaped frequency curve*
d. *Any of these.*

Numerical MCQs

ICAI SM

(66) *Out of 1000 persons, 25 per cent were industrial workers and the rest were agricultural workers. 300 persons enjoyed world cup matches on TV. 30 per cent of the people who had not watched world cup matches were industrial workers. What is the number of agricultural workers who had enjoyed world cup matches on TV?*

a. 260
b. 240
c. 230
d. 250

	Agri	Indust	Total
TV	260	40	300
Non TV	490	210	700
Total	750	250	1000

30%

ICAI SM

(67) *A sample study of the people of an area revealed that total number of women were 40% and the percentage of coffee drinkers were 45 as a whole and the percentage of male coffee drinkers was 20. What was the percentage of female non-coffee drinkers?*

a. 10
b. 15
c. 18
d. 20

	Male	female	
Coffee	20	25	45
no coffee	40	15	55
	60	40	100

ICAI SM

(68) *Cost of sugar in a month under the heads raw materials, labour, direct production and others were 12, 20, 35 and 23 units respectively. What is the difference between the central angles for the largest and smallest components of the cost of sugar?*

a. 72°
b. 48°
c. 56°
d. 92°

12 → smallest
20
35 → largest
23
90

$12/90 \times 360^\circ = 48^\circ$

$35/90 \times 360^\circ = 140^\circ$

diff 92°

ICAI SM

(69) *The number of accidents for seven days in a locality are given < :*

No. of accidents	0	1	2	3	4	5	6
Frequency	15	19	22	31	9	3	2

What is the number of cases when 3 or less accidents occurred?

a. 56
b. 6
c. 68
d. 87

ICAI SM

(70) *The following data relate to the incomes of 86 persons :*

Income in Rs	500–999	1000-1499	1500-1999	2000-2499
No. of persons	15	28	36	7

What is the percentage of persons earning more than Rs. 1500?

a. 50
b. 45
c. 40
d. 60

$\frac{43}{86} \times 100 = 50\%$

ICAI SM

(71) *The following data relate to the marks of a group of students:*

ICAI SM

(69) The number of accidents for seven days in a locality are given < :

No. of accidents	0	1	2	3	4	5	6
Frequency	15	19	22	31	9	3	2

What is the number of cases when 3 or less accidents occurred?

a. 56 b. 6

c. 68 d. 87

ICAI SM

(70) The following data relate to the incomes of 86 persons :

Income in Rs	500-999	1000-1499	1500-1999	2000-2499
No. of persons	15	28	36	7

What is the percentage of persons earning more than Rs. 1500?

a. 50 b. 45

c. 40 d. 60

ICAI SM

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

Marks	<10	<20	<30	<40	<50
No. of students	15	38	65	84	100

How many students got marks more than 30?

a. 65 b. 50

c. 35 d. 43

Sampling (added in New Syllabus)

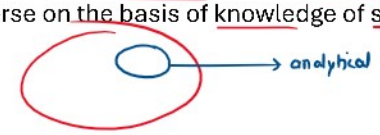
Disclaimer

- This topic is uses the concept of probability, theoretical distribution & combinations also
- Getting detailed understanding is not required
- If you get overview and able to solve Theory MCQs – that is also fine

Sampling (added in New Syllabus)

Need of Sampling

- Due to time and cost, it is impossible to do the complete enumeration of all units of population/ universe
- Hence we use sample to draw conclusion/ inference about unknown universe on the basis of knowledge of sample observations



ICAI SM

(1) *Sampling can be described as a statistical procedure*

a. *To infer about the unknown universe from a knowledge of any sample*

b. *To infer about the known universe from a knowledge of a sample drawn from it*

c. *To infer about the unknown universe from a knowledge of a random sample drawn from it*

d. *Both (a) and (b).*

ICAI SM

(5) *Statistical decision about an unknown universe is taken on the basis of*

a. *Sample observations* b. *A Sampling frame*

c. *Sample survey* d. *Complete Enumeration*

Sampling (added in New Syllabus)

Population

- Aggregate of all the units under consideration
- Number of units of population is called as population size
- Types – Finite, Infinite, Existent, Hypothetical

ICAI SM

(4) *The population of roses in Salt Lake City is an example of*

a. *A finite population* b. *An infinite population*

c. *A hypothetical population* d. *An imaginary population.*

Sampling (added in New Syllabus)

Sample

- part of a population selected with a view to represent the population and all its characteristics.
- Number of units of in a sample is called as sample size
- Unitsforming the sample are known as “Sampling Units”
- A detailed and complete list of all the sampling units is known as a “Sampling Frame”

Sampling (added in New Syllabus)

Parameter

- characteristic of a population based on all the units of the population.
- Example: Population Mean, Population Variance etc.

Statistic

- statistical measure of sample observation
- function of sample observations
- It is used to estimate parameter hence it is also called as estimator
- Example of Statistic – Sample Mean, Sample Variance etc.

Population 100,000 units

Sample randomly 500 persons

/estimator
Statistic

1-Apr-2024.	Sample I	sample $\bar{x}$ = 54 kg
15-Apr-2024	Sample II	sample $\bar{x}$ = 55 kg
30-Apr-2024	Sample III	sample $\bar{x}$ = 56 kg

Population mean = 55 kg

CA FOUNDATION QUANTITATIVE APTITUDE | CHAPTER 13

ICAI SM

(7) *A parameter is a characteristic of*

a. *Population* b. *Sample*

c. *Both (a) and (b)* d. *(a) or (b)*

ICAI SM

(8) *A statistic is*

a. *A function of sample observations* b. *A function of population units*

c. *A characteristic of a population* d. *A part of a population.*

Sampling (added in New Syllabus)

Sampling Fluctuations
- value of Statistic varies from sample to sample
- this variation is called as sampling fluctuations

Sampling Distribution
- distribution of value of statistic for each sample along with its corresponding probability
- It is similar to Random Variable Probability Distribution
- Mean of Sampling Distribution is called as Expectation
- SD of Sampling Distribution is called as Standard Error

sampling dist

Statistic	Probability
54 kg	25%
55 kg	40%
56 kg	5%
57 kg	10%

Mean → Expectation

SD → Standard Error

ICAI SM

(9) *Sampling Fluctuations may be described as*
- a. *The variation in the values of a statistic*
- b. *The variation in the values of a sample*
- c. *The differences in the values of a parameter*
- d. *The variation in the values of observations.*

ICAI SM

(10) *The sampling distribution is*
- a. *The distribution of sample observations*
- b. *The distribution of random samples*
- c. *The distribution of a parameter*
- d. *The probability distribution of a statistic.*

ICAI SM

(11) *Standard error can be described as*
- a. *The error committed in sampling*
- b. *The error committed in sample survey*
- c. *The error committed in estimating a parameter*
- d. *Standard deviation of a statistic.*

ICAI SM

(12) *A measure of precision obtained by sampling is given by*
- a. *Standard Error*
- b. *Sampling fluctuation*
- c. *Sampling distribution*
- d. *Expectation*

ICAI SM

(13) *As the sample size increases, standard error*
- a. *Increases*
- b. *Decreases*
- c. *Remains constant*
- d. *Decreases Proportionately*

Sampling (added in New Syllabus)

Formulas to find number of possible samples

- If population size is n and sample size is r then number of possible samples in both cases are below
- In case of without replacement: ${}^nC_r$
- In case of with replacement: n^r

CA FOUNDATION QUANTITATIVE APTITUDE | CHAPTER 13

ICAI SM

(23) If a random sample of size 2 with replacement is taken from the population containing the units 3,6 and 1, then the samples would be

a. (3,6),(3,1),(6,1)
b. (3,3),(6,6),(1,1)
c. (3,3),(3,6),(3,1),(6,6),(6,3),(6,1),(1,1),(1,3),(1,6)
d. (1,1),(1,3),(1,6),(6,1),(6,2),(6,3),(6,6),(1,6),(1,1)

ICAI SM

(24) If a random sample of size two is taken without replacement from a population containing the units a,b,c and d then the possible samples are

a. (a, b),(a, c),(a, d)
b. (a, b),(b, c), (c, d)
c. (a, b), (b, a), (a, c),(c,a), (a, d), (d, a)
d. (a, b), (a, c), (a, d), (b, c), (b, d), (c,d)

CA FOUNDATION QUANTITATIVE APTITUDE | CHAPTER 13

ICAI SM

(14) If from a population with 25 members, a random sample without replacement of 2 members is taken, the number of all such samples is

a. 300
b. 625
c. 50
d. 600

ICAI SM

(15) A population comprises 5 members. The number of all possible samples of size 2 that can be drawn from it with replacement is

a. 100
b. 15
c. 125
d. 25

Sampling (added in New Syllabus)

Sample Survey
- study of the unknown population
- Based on proper representative sample drawn from it

Principle of Sample Survey
- Law of Statistical Regularity: for large sample size, its characterstics averagely match with populations'
- Principle of Inertia: result of large sample are more reliable and accurate
- Principle of Optimization: good sample design gives optimum efficiency at min cost
- Principle of validity: sample design is valid only if it is random

ICAI SM

(2) *The Law of Statistical Regularity says that*

a. *Sample drawn from the population under discussion possesses the characteristics of the population*

b. *A large sample drawn at random from the population would posses the characteristics of the population*

c. *A large sample drawn at random from the population would possess the characteristics of the population on an average*

d. *An optimum level of efficiency can be attained at a minimum cost.*

Sampling (added in New Syllabus)

complete 20 × 100,000 = 20L

samp 50 × 1000 = 50k

Preference to Sample Survey over complete enumeration
- Speed
- Cost: In Sampling per unit cost is high if you hire trained staff but total is less as compared to complete enumeration
- Reliability: Trained Enumerators, Better Supervision, Modern Technique
- Accuracy: Sampling Errors can be reduced by proper design and size increase
- Necessity: If pop size is very large it is necessity not a choice

If population size is small then no need of sampling

Sampling (added in New Syllabus)

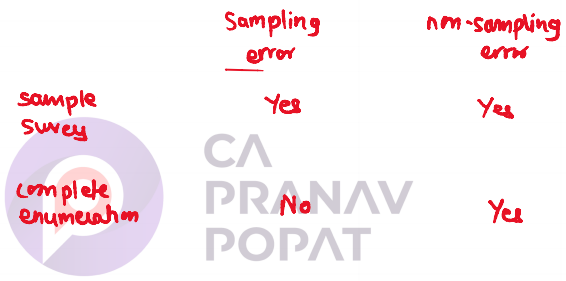
Errors

- Deviation between actual value and observed value
- Sampling Errors: defective design, substitution, wrong choice of statistic, high variability in population
- Non-Sampling Errors: lapse of memory, preference for certain digits, ignorance, non-responses, wrong measurement, communication gap

(3) *A sample survey is prone to*

a. *Sampling errors* b. *Non-sampling errors*

c. *Either (a) or (b)* d. *Both (a) & (b)*



Sampling (added in New Syllabus)

Types of Sampling

- Probability: chance of each unit to get select is same. It is used in these sampling processes - Simple random sampling, Stratified sampling, Multi Stage sampling etc.
- Non-Probability: based entirely on the judgement of the sampler, it is also called as Purposive or Judgement Sampling.
- Mixed Sampling: based partly on some probabilistic law and partly on some pre decided rule. Systematic sampling belongs to this category.

ICAI SM

(6) *Random sampling implies*

a. *Haphazard sampling*
b. *Probability sampling*
c. *Systematic sampling*
d. ✓ *Sampling with the same probability for each unit*

ICAI SM

(22) *Which sampling is subjected to the discretion of the sampler?*

a. *Systematic sampling*
b. *Simple random sampling*
c. ✓ *Purposive sampling*
d. *Quota sampling*

Judgement
non-prob.

Sampling (added in New Syllabus)

Simple Random Sampling

- each unit belonging to the population has an equal chance of being a part of the sample
- Used when (i) the population is not very large (ii) the sample size is not very small and (iii) the population under consideration is not heterogeneous

ICAI SM

(16) *Simple random sampling is very effective if*

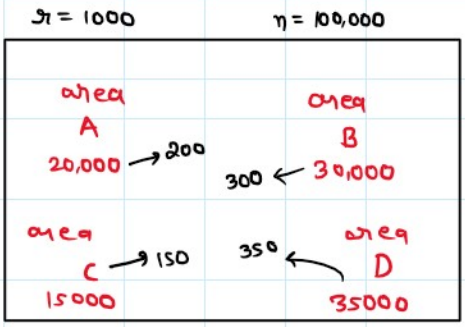
a. *The population is not very large*
b. *The population is not much heterogeneous*
c. *The population is partitioned into several sections.*
d. ✓ *Both (a) and (b)*

ICAI SM

(17) *Simple random sampling is*

a. ✓ *A probabilistic sampling*
b. *A non- probabilistic sampling*
c. *A mixed sampling*
d. *Both (b) and (c)*

Sampling (added in New Syllabus)



Stratified Sampling

- Used If the population is large and heterogeneous
- dividing the population into a number of strata or subpopulations
- Purpose is (i) to make representation of all the sub populations (ii) to provide an estimate of parameter not only for all the strata but also and overall estimate (iii) reduction of variability and thereby an increase in precision.
- We consider Proportional allocation" or "Bowley's allocation" where the sample sizes for different strata are taken as proportional to the population sizes.
- When the strata-variances differ significantly among themselves, we use "Neyman's allocation" where sample size vary jointly with population size and population standard deviation.

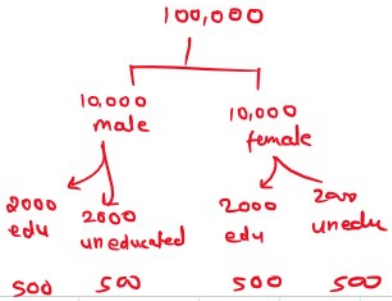
CA FOUNDATION QUANTITATIVE APTITUDE | CHAPTER 13

		ICAI SM
(18)	*According to Neyman's allocation, in stratified sampling*	
	a.	*Sample size is proportional to the population size* 4 pop SD
	b.	*Sample size is proportional to the sample SD*
	c.	*Sample size is proportional to the sample variance*
	d.	*Population size is proportional to the sample variance.*
		ICAI SM
(19)	*Which sampling provides separate estimates for population means for different segments and also an over all estimate?*	
	a. Multistage sampling	*b. Stratified sampling*
	c. Simple random sampling	*d. Systematic sampling*

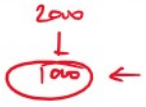
Sampling (added in New Syllabus)

Multi-stage sampling

- Stage wise samples are drawn till we reach ultimate sample size
- Coverage is high and saved labour, cost effective
- Adds flexibility to sampling process
- Less accurate then stratified



Sampling (added in New Syllabus)



Systematic Sampling

- units constituting the sample are selected at regular interval after selecting the very first unit at random i.e., with equal probability
- is partly probability sampling
- It is less time consuming, less expensive and simple
- Drawback – due to any unknown periodicity, it may give highly biased result

- It is less time consuming, less expensive and simple
- Drawback – due to any unknown periodicity, it may give highly biased result

17th 22nd
27th 32nd

ICAI SM

(20) *Which sampling adds flexibility to the sampling process?*

a. *Simple random sampling* b. *Multistage sampling*
c. *Stratified sampling* d. *Systematic sampling*

ICAI SM

(21) *Which sampling is affected most if the sampling frame contains an undetected periodicity?*

a. *Simple random sampling* b. *Stratified sampling*
c. *Multistage Sampling* d. *Systematic sampling*