



CBSE Class X Data Science Set 4 (106)

Time Allowed :2 Hours	Maximum Marks :50	Total Questions :21
-----------------------	-------------------	---------------------

General Instructions

1. Please read the instructions carefully.
2. This question paper consists of 21 questions in two sections: **Section A & Section B.**
3. Section A has objective type questions whereas Section B contains subjective type questions.
4. Out of the given ($5 + 16 = 21$) questions, a candidate has to answer ($5 + 10 = 15$) questions in the allotted (maximum) time of 2 hours.
5. All questions of a particular section must be attempted in the correct order.
6. **Section A – Objective type questions (24 marks):**
 - (a) This section has 5 questions.
 - (b) There is no negative marking.
 - (c) Do as per the instructions given.
 - (d) Marks allotted are mentioned against each question/part.
7. **Section B – Subjective type questions (26 marks):**
 - (a) This section has 16 questions.
 - (b) A candidate has to do 10 questions.
 - (c) Do as per the instructions given.
 - (d) marks allotted are mentioned against each question/ part.

Section A

1. Answer any 4 out of the given 6 questions:

(i) What is a common misconception about starting a business?

- (A) You need a large amount of money to get started.
- (B) Starting a business requires no financial investment.
- (C) You can start a business with whatever money you have.
- (D) Every business idea has to be unique.

Correct Answer: (A) You need a large amount of money to get started.

Explanation: While financial planning is essential, many businesses can start with minimal investment using creative approaches or digital platforms. The misconception that large capital is mandatory often discourages potential entrepreneurs.

Quick Tip

Start small and scale as you grow. Focus on creativity and leveraging low-cost resources initially.

(ii) Which of the following is *not* true with respect to self-awareness abilities?

- (A) Being self-aware means that you can identify your strengths and weaknesses.
- (B) Self-awareness helps to convert your strengths to weaknesses.
- (C) Self-awareness helps to convert your strengths to exceptional talent.
- (D) Analyzing your strengths and weaknesses helps you to attain success in life.

Correct Answer: (B) Self-awareness helps to convert your strengths to weaknesses.

Explanation: Self-awareness focuses on identifying and enhancing strengths while managing weaknesses. It does not aim to diminish strengths.

Quick Tip

Practice regular reflection to improve self-awareness and align your goals with your abilities.



(iii) Managing ----- is about making a plan to be able to cope effectively with daily pressures and to strike a balance between life, work, relationships, relaxation, and fun.

- (A) Weight
- (B) Stress
- (C) Beliefs
- (D) Interests

Correct Answer: (B) Stress

Explanation: Stress management involves techniques to maintain a healthy balance in life by organizing tasks, practicing relaxation, and focusing on positive habits.

Quick Tip
Manage stress by setting priorities, taking breaks, and practicing mindfulness techniques.

(iv) Which of the following is a type of verbal communication?

- (A) Gestures
- (B) Body language
- (C) Facial expressions
- (D) Interpersonal communication

Correct Answer: (D) Interpersonal communication

Explanation: Verbal communication includes spoken words and conversations, while gestures and body language are forms of non-verbal communication.

Quick Tip
Effective interpersonal communication involves clarity, active listening, and constructive feedback.



(v) What is the keyboard shortcut for copying selected text or files in most operating systems?

- (A) Ctrl + X
- (B) Ctrl + Z
- (C) Ctrl + V
- (D) Ctrl + C

Correct Answer: (D) Ctrl + C

Explanation: Ctrl + C is the universal keyboard shortcut for copying text or files across most operating systems. It is commonly paired with Ctrl + V for pasting.

Quick Tip

Keyboard shortcuts like Ctrl + C and Ctrl + V save time and increase efficiency in digital tasks.

(vi) How many Sustainable Development Goals (SDGs) were adopted by the United Nations in 2015?

- (A) 5
- (B) 10
- (C) 15
- (D) 17

Correct Answer: (D) 17

Explanation: The United Nations adopted 17 Sustainable Development Goals in 2015, focusing on areas such as poverty reduction, quality education, and climate action to achieve a better and sustainable future.

Quick Tip

The 17 SDGs aim to address global challenges by 2030, including environmental, social, and economic goals.



2. Answer any 5 out of the given 6 questions:

(i) In data analysis, why should you create subsets of your data?

- (A) To increase the size of your dataset.
- (B) To make the data easier to analyze.
- (C) To add extra information in your dataset.
- (D) To add noise to the data.

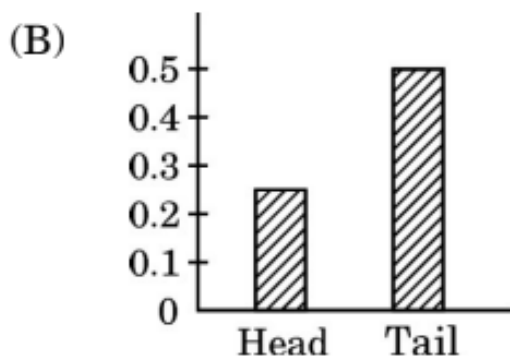
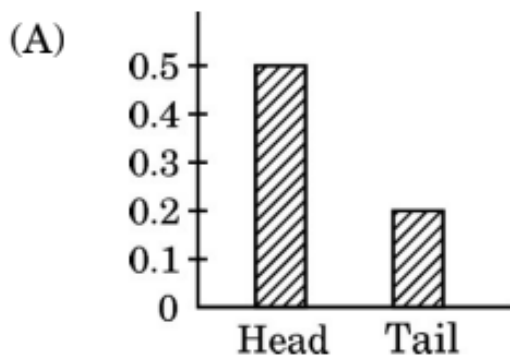
Correct Answer: (B) To make the data easier to analyze.

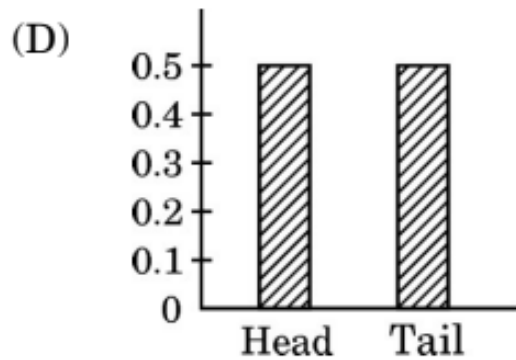
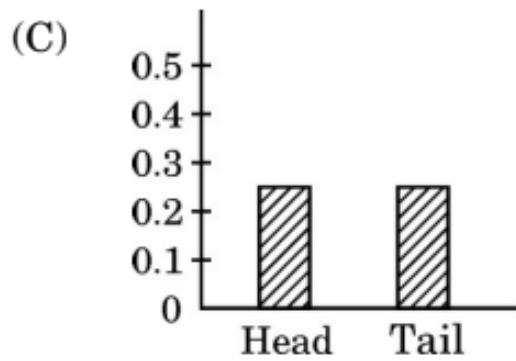
Explanation: Creating subsets of data makes it easier to focus on specific parts of the dataset, allowing for targeted analysis and better insights without being overwhelmed by the full dataset.

Quick Tip

Break large datasets into smaller subsets to simplify analysis and improve processing efficiency.

(ii) Which of the following graphs shows the probability distribution for tossing a coin?





Correct Answer: (C)

The graph in option (C) shows equal probabilities (0.5) for both heads and tails, which is the correct probability distribution for a fair coin toss.

Quick Tip

A fair coin toss results in equal probabilities (50

(iii) In Data Science, _____ is a deviation from the expected outcome in the data.

- (A) Outlier
- (B) Standard Deviation
- (C) Bias
- (D) Variance

Correct Answer: (A) Outlier

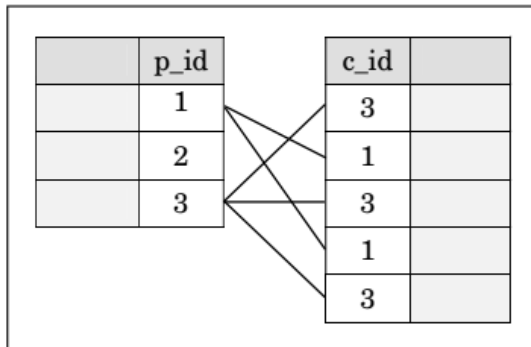
Explanation: An outlier is a data point significantly different from the other observations in the dataset, representing unexpected deviations.



Quick Tip

Identify outliers to improve model accuracy and ensure better data consistency.

(iv) Identify the type of join from the following diagram.



- (A) Many to Many
- (B) One to Many
- (C) One to One
- (D) Many to One

Correct Answer: (B) One to Many

Explanation: In the diagram, a single `p_id` maps to multiple `c_id` values, demonstrating a One-to-Many relationship.

Quick Tip

One-to-Many joins are used when a single key from one table corresponds to multiple keys in another.

(v) State True or False.

If you soft delete a particular file, it gets deleted permanently and cannot be restored.

Correct Answer: False

Explanation: Soft delete marks the file as deleted without permanently removing it, allowing for recovery later. Permanent deletion occurs with hard delete.

Quick Tip

Soft deletes retain data for recovery, while hard deletes remove it permanently.

(vi) **Statement I: Shredding and cutting are two ways to discard physical data.**

Statement II: The shredded documents can be read easily but the documents that are cut down into small pieces cannot be read.

- (A) Both Statement I and Statement II are true.
- (B) Both Statement I and Statement II are false.
- (C) Statement I is true, but Statement II is false.
- (D) Statement II is true, but Statement I is false.

Correct Answer: (C) Statement I is true, but Statement II is false.

Explanation: Shredding and cutting are effective ways to discard physical data securely. However, shredded documents are harder to read, while cut pieces are equally unreadable.

Quick Tip

For secure physical data disposal, always prefer shredding over manual cutting to ensure confidentiality.

3. Answer any 5 out of the given 6 questions:

(i) **A _____ table represents the percentages of data points that fit in each category.**

- (A) Dummy
- (B) Two-way frequency
- (C) Two-way relative frequency
- (D) Two-way direct frequency

Correct Answer: (C) Two-way relative frequency

Explanation: A two-way relative frequency table shows the proportions or percentages of data points in each category. It helps in analyzing relationships between two categorical variables.



Quick Tip

Relative frequency tables are useful for comparing data proportions and identifying trends between categories.

(ii) The value of z-score is always _____ if the value of z-score lies above the mean.

- (A) Positive
- (B) Negative
- (C) Zero (0)
- (D) Infinite

Correct Answer: (A) Positive

Explanation: A positive z-score indicates that the data point lies above the mean.

Conversely, a negative z-score means the data point is below the mean, while zero implies it is at the mean.

Quick Tip

Use z-scores to standardize data and understand how far a data point is from the mean in terms of standard deviations.

(iii) What does a distribution represent in Data Science?

- (A) The exact values for a variable and how often they occur.
- (B) The probable values for a variable and how often they occur.
- (C) The average values for a variable and how often they occur.
- (D) The maximum values for a variable and how often they occur.

Correct Answer: (B) The probable values for a variable and how often they occur.

Explanation: A distribution describes the range of possible values of a variable and the frequency of their occurrence. Common distributions include normal and binomial distributions.



Quick Tip

Study distributions to analyze patterns, probabilities, and variability in data for better decision-making.

(iv) In which bias do we assume that change in one quantity produces an equal and proportional change in another?

- (A) Linearity Bias
- (B) Confirmation Bias
- (C) Recall Bias
- (D) Selection Bias

Correct Answer: (A) Linearity Bias

Explanation: Linearity bias occurs when we mistakenly assume a direct proportional relationship between two variables without considering other influences or nonlinear relationships.

Quick Tip

Avoid linearity bias by exploring data relationships using scatterplots and statistical tests.

(v) Which of the following is an appropriate way of discarding physical copies of confidential data?

- (A) Crumpling the papers which contain confidential data and throwing them in the dustbin.
- (B) Burning the documents.
- (C) Leaving the confidential data in an unlocked drawer.
- (D) Using the papers containing confidential data for rough work.

Correct Answer: (B) Burning the documents.

Explanation: Burning ensures that confidential data is destroyed permanently, making it unrecoverable. Other methods, like shredding, are also effective for secure disposal.



Quick Tip

Always use shredding or burning methods to ensure the secure disposal of sensitive documents.

(vi) Quartiles of a dataset partition the data in _____ equal parts.

- (A) Three
- (B) Four
- (C) Eight
- (D) Ten

Correct Answer: (B) Four

Explanation: Quartiles divide a dataset into four equal parts, each containing 25

Quick Tip

Quartiles are useful for detecting outliers and summarizing the distribution of datasets.

4. Answer any 5 out of the given 6 questions:

(i) What security measure should you employ to ensure that even in the case of a data leak, the hackers are not able to read your data?

- (A) Backing up data to an external drive.
- (B) Using any password.
- (C) Encrypting the data.
- (D) Deleting the data.

Correct Answer: (C) Encrypting the data.

Explanation: Encrypting data ensures that it is unreadable to unauthorized users, even if it is intercepted. Encryption adds an extra layer of security.



Quick Tip

Always encrypt sensitive data to secure it from unauthorized access, especially during storage or transmission.

(ii) In a dataset of income levels for a small town's residents, the income data includes a few extremely high values (outliers) due to the presence of some wealthy individuals. In this scenario, which form of central tendency would be more appropriate to describe the typical income of the town's residents?

- (A) Mean
- (B) Median
- (C) Mode
- (D) Range

Correct Answer: (B) Median

Explanation: Median is less affected by outliers compared to the mean. It represents the middle value, providing a better measure of central tendency in datasets with extreme values.

Quick Tip

Use the median to summarize data when outliers might skew the mean.

(iii) What is the correct order of steps in the Statistical Problem Solving Process?

- (A) Collect/Consider the data, Formulate statistical investigative questions, Analyse the data, Interpret the results.
- (B) Analyse the data, Interpret the results, Collect/Consider the data, Formulate statistical investigative questions.
- (C) Formulate statistical investigative questions, Collect/Consider the data, Analyse the data, Interpret the results.
- (D) Formulate statistical investigative questions, Interpret the results, Analyse the data, Collect/Consider the data.



Correct Answer: (C) Formulate statistical investigative questions, Collect/Consider the data, Analyse the data, Interpret the results.

Explanation: The process starts with defining the problem through investigative questions, then collecting data, analyzing it, and interpreting the results for decision-making.

Quick Tip

Follow the four-step approach: Ask, Collect, Analyze, and Interpret for effective problem solving in statistics.

(iv) If n is the total number of data in a sample, which of the following correctly represents the first decile?

- (A) $D_1 = 1 + (n + 1)/10$ Data.
- (B) $D_1 = 1 - (n + 1)/10$ Data.
- (C) $D_1 = n/(n + 1)$ Data.
- (D) $D_1 = (n + 1)/10$ Data.

Correct Answer: (D) $D_1 = (n + 1)/10$ Data.

Explanation: The first decile (D_1) divides the dataset into 10 equal parts, and its position is given by $(n + 1)/10$ when data is arranged in ascending order.

Quick Tip

For deciles, use $(n + 1)/10$, $(n + 1)/5$, and so on, to divide the dataset into equal parts.

(v) If you have a dataset with 100 data points, what data point defines the first quartile (Q_1)?

- (A) The 25th percentile.
- (B) The 30th percentile.
- (C) The 75th percentile.
- (D) The 100th percentile.



Correct Answer: (A) The 25th percentile.

Explanation: The first quartile (Q1) corresponds to the 25th percentile, dividing the dataset such that 25

Quick Tip

Use quartiles (Q1, Q2, Q3) to divide data into four equal sections, with Q1 at the 25th percentile.

(vi) Which type of bias is commonly associated with recommendation systems and personalized advertisements, where the sample data is not representative of the true future population of cases that the model will see?

- (A) Linearity Bias
- (B) Recall Bias
- (C) Selection Bias
- (D) Survivor Bias

Correct Answer: (C) Selection Bias.

Explanation: Selection bias occurs when the dataset used to train or evaluate a model is not representative of the entire population, leading to skewed results.

Quick Tip

Prevent selection bias by using diverse and representative datasets for training machine learning models.

5. Answer any 5 out of the given 6 questions:

(i) Statement I: In Data Science, we can perform data merging by implementing data joins on the databases in frame.

Statement II: There are five categories of data joins.

- (A) Both statement I and statement II are correct.



- (B) Both statement I and statement II are incorrect.
- (C) Statement I is correct, but statement II is incorrect.
- (D) Statement II is correct, but statement I is incorrect.

Correct Answer: (A) Both statement I and statement II are correct.

Explanation: Data Science often uses data joins such as inner join, left join, right join, full join, and cross join to merge datasets effectively.

Quick Tip

Understand the five types of joins in SQL (inner, left, right, full, and cross) for effective data merging in Data Science.

(ii) You have a dataset of ages for a group of 5 people. The ages are as follows: 35, 32, 29, 35, and 27. What is the median age for this dataset?

- (A) 27
- (B) 28
- (C) 29
- (D) 30

Correct Answer: (C) 29.

Explanation: Median is the middle value in a sorted dataset. Arranging the dataset as 27, 29, 32, 35, 35, the middle value is 29.

Quick Tip

For datasets with an odd number of elements, the median is the middle value when sorted.

(iii) You have a bag with two red marbles and two green marbles. If you randomly select one marble from the bag without looking, what is the probability that it will be a red marble?



- (A) 0.50
- (B) 0.25
- (C) 1.0
- (D) 0.75

Correct Answer: (A) 0.50.

Explanation: Total marbles = 4 (2 red and 2 green). The probability of selecting a red marble is $\frac{\text{favorable outcomes}}{\text{total outcomes}} = \frac{2}{4} = 0.50$.

Quick Tip

Probability = $\frac{\text{Number of favorable outcomes}}{\text{Total number of outcomes}}$. Always check the total outcomes first.

(iv) In statistical analysis, what minimum sample size is generally considered enough for the Central Limit Theorem to hold?

- (A) 5
- (B) 10
- (C) 20
- (D) 30

Correct Answer: (D) 30.

Explanation: The Central Limit Theorem states that for a sufficiently large sample size (typically $n \geq 30$), the sampling distribution of the mean approximates a normal distribution.

Quick Tip

For accurate results using the Central Limit Theorem, ensure the sample size is at least 30.

(v) Deciles sort the data into how many equal parts?

- (A) 5
- (B) 10



(C) 8

(D) 4

Correct Answer: (B) 10.

Explanation: Deciles divide the dataset into 10 equal parts, with each part representing 10

Quick Tip

Remember: Quartiles = 4 parts, Deciles = 10 parts, Percentiles = 100 parts.

(vi) Assertion (A): Privacy does not always mean confidentiality of data.

Reason (R): Private data may need to be audited based on the relevant requirements.

(A) Both (A) and (R) are correct and (R) is the correct explanation of (A).

(B) Both (A) and (R) are correct, but (R) is not the correct explanation of (A).

(C) (A) is correct, but (R) is incorrect.

(D) (R) is correct, but (A) is incorrect.

Correct Answer: (A) Both (A) and (R) are correct and (R) is the correct explanation of (A).

Explanation: Privacy focuses on user control over personal data, whereas confidentiality ensures data is only accessible to authorized individuals. Both can overlap but are distinct concepts.

Quick Tip

Distinguish privacy and confidentiality: Privacy = User control, Confidentiality = Controlled access.

Section B

Answer any 3 out of the given 5 questions on Employability Skills.

Answer each question in 20-30 words.



6. Give any two ways to overcome barriers of effective communication.

Answer:

- Use active listening to ensure mutual understanding between parties.
- Simplify the message to eliminate confusion and use clear, concise language.

Quick Tip

Effective communication can be enhanced by practicing active listening and using simple, direct language.

7. Differentiate between wage-employed people and self-employed people.

Answer:

- **Wage-employed:** Individuals work for an employer and earn a fixed wage or salary.
- **Self-employed:** Individuals run their own business or practice and earn profits.

Quick Tip

Wage-employed individuals depend on employers for income, while self-employed individuals generate income independently.

8. Explain any two key strategies or initiatives that can contribute towards achieving Sustainable Development Goals (SDGs).

Answer:

- **Promoting renewable energy:** Investing in solar, wind, and other renewable energy sources reduces carbon footprints.
- **Quality education:** Providing inclusive and equitable quality education ensures opportunities for all.



Quick Tip

Achieving SDGs requires a focus on environmental sustainability and social inclusiveness.

9. Define any two of the following terms:

- (a) ICT
- (b) Trojan Horse
- (c) Firewall

Answer:

- **ICT:** Information and Communication Technology refers to digital tools that facilitate communication and the transfer of data.
- **Trojan Horse:** A type of malware disguised as legitimate software that grants unauthorized access to hackers.
- **Firewall:** A network security system that monitors and controls incoming and outgoing network traffic based on predetermined security rules.

Quick Tip

ICT is integral for digital communication, while firewalls and malware prevention tools ensure cybersecurity.

10. What does the acronym SMART stand for in the context of goal setting?

Why is goal setting an essential factor in your personal life?

Answer:

- **SMART:** Specific, Measurable, Achievable, Relevant, Time-bound.
- **Importance:** Setting SMART goals helps prioritize tasks, track progress, and maintain focus to achieve personal and professional objectives.



Quick Tip

SMART goals ensure clarity and structure in planning, helping individuals achieve milestones effectively.

11. Discuss any two ways of protecting confidential data that is stored in digital form.

Answer:

- **Encryption:** Encrypting sensitive data ensures that only authorized individuals with decryption keys can access the information.
- **Firewalls:** Installing firewalls helps prevent unauthorized access to the network and protects data from cyberattacks.

Quick Tip

Always use encryption and firewalls together for a robust digital security strategy.

12. Write steps to calculate Mean Absolute Deviation (MAD).

Answer:

1. Find the mean of the dataset.
2. Subtract the mean from each data point and take the absolute value of the differences.
3. Sum all the absolute deviations.
4. Divide the sum by the total number of data points.

Quick Tip

MAD provides an average measure of variability in the dataset, useful for identifying consistency.



13. Explain the concept of Recall Bias in Statistics.

Answer: Recall bias occurs when participants in a study are unable to accurately remember past events, leading to errors in the data collected. It is common in retrospective studies where self-reported information is required.

Quick Tip

To minimize recall bias, use multiple data sources or conduct surveys soon after the event.

14. For each of the following scenarios, classify the data as either discrete or continuous:

- (a) Result of an admission test
- (b) Weight of apples at a grocery store
- (c) Football score in a match
- (d) Depth of a river

Answer:

- (a) Discrete
- (b) Continuous
- (c) Discrete
- (d) Continuous

Quick Tip

Discrete data involves countable values, while continuous data includes measurements.

15. Explain one-to-one join with the help of an example.

Answer: A one-to-one join links two tables where each row in one table corresponds to exactly one row in another.



Example: Linking employee details in Table A (Employee ID, Name) with payroll data in Table B (Employee ID, Salary).

Quick Tip

Ensure both tables have a unique key column for one-to-one joins to maintain data integrity.

16. In a dataset representing the ages of a group of people, the lower quartile (Q1) is 30 years, and the upper quartile (Q3) is 50 years. Calculate the interquartile range (IQR) for this dataset.

Answer: The interquartile range (IQR) is calculated as:

$$\text{IQR} = Q3 - Q1 = 50 - 30 = 20 \text{ years.}$$

Quick Tip

IQR measures data spread by focusing on the middle 50% of the dataset, reducing the impact of outliers.

Answer any 3 out of the given 5 questions in 50-80 words each

17. What is a two-way frequency table? Explain its features with a suitable example.

Answer: A two-way frequency table displays the frequency of data across two categorical variables. It helps summarize relationships between the variables.

Example: A survey records gender and preferred subject:

	Math	Science	Total
Male	20	15	35
Female	25	10	35
Total	45	25	70

This table shows the relationship between gender and subject preference.



Quick Tip

Use two-way frequency tables to analyze categorical data and identify trends or correlations effectively.

18. Explain the following components of the Statistical Problem Solving Process:

1. Analyse the data
2. Interpret the results

Answer:

1. **Analyse the data:** Use statistical tools to process and summarize data, identify patterns, and test hypotheses.
2. **Interpret the results:** Derive meaningful conclusions from the analysis, link findings to the problem statement, and propose actionable insights.

Quick Tip

Always ensure the analysis aligns with the problem goals and clearly communicates the results for better decision-making.

19. Explain the Central Limit Theorem. Give any two real-world scenarios in which it is used.

Answer: The Central Limit Theorem states that the sampling distribution of the sample mean approaches a normal distribution as the sample size increases, regardless of the population distribution.

Scenarios:

- **Quality control:** Checking the average weight of packets in a factory.
- **Election surveys:** Predicting the average voter preference using sample surveys.



Quick Tip

The Central Limit Theorem enables reliable inference and hypothesis testing in statistics, even with non-normal population distributions.

20. What is z-score in data science? Write the formula to calculate z-score. Also, explain the importance of z-score in data science.

Answer: A z-score measures how many standard deviations a data point is from the mean. It is calculated using:

$$z = \frac{x - \mu}{\sigma}$$

Where:

- x : Data point
- μ : Mean
- σ : Standard deviation

Importance:

- Identifies outliers.
- Standardizes data for comparison across different datasets.
- Facilitates hypothesis testing.

Quick Tip

Use z-scores to detect anomalies and normalize data, especially when comparing datasets with varying scales.

21. You are a data analyst working for a healthcare organisation. Your team is responsible for analysing patient data to improve healthcare services and outcomes. Ethical considerations are crucial in handling this sensitive information. Recently, you received a dataset containing patient records, including medical history, personal information, and treatment details.



(i) What is the most important ethical consideration when handling patient's data?

1. Data accuracy and quality
2. Data sharing with external organisations without patients' knowledge
3. Data storage cost optimisation
4. Patient's privacy protection

Correct Answer: (D) Patient's privacy protection

Explanation: Protecting patient privacy is the most crucial ethical consideration in healthcare data analysis. It ensures compliance with legal frameworks like GDPR and builds trust between patients and healthcare providers.

Quick Tip

Always anonymize data and follow relevant laws such as HIPAA or GDPR to safeguard patient privacy.

(ii) A colleague suggests using patients' data to train an artificial intelligence model without the patients' consent. What is your ethical stance on this suggestion?

- (A) Support the idea as it can lead to technological advancements
- (B) Implement the idea without informing the patients
- (C) Reject the idea and ensure patients' consent is obtained for any data usage
- (D) Share the colleague's idea with other organisations for wider adoption

Correct Answer: (C) Reject the idea and ensure patients' consent is obtained for any data usage

Explanation: Using data without patient consent violates ethical principles and legal requirements. Informed consent ensures respect for individuals and prevents misuse of sensitive data.

Quick Tip

Obtain informed consent before using data for secondary purposes like training AI models to avoid legal and ethical violations.



(iii) List any two ethical guidelines around data analysis.

Answer:

1. Ensure transparency in data usage by informing all stakeholders about the purpose and scope of analysis.
2. Maintain data security by encrypting sensitive information and limiting access to authorized personnel only.

Quick Tip
Always follow ethical guidelines such as accountability, fairness, and privacy to maintain the integrity of data analysis.

