class X 2026

Series : 2LKNM

SET ~ 4

प्रश्न-पत्र कोड **106**
Q.P. Code

रोल नं.
Roll No.

परीक्षार्थी प्रश्न-पत्र कोड को उत्तर-पुस्तिका के मुख-पृष्ठ पर अवश्य लिखें।
Candidates must write the Q.P. Code on the title page of the answer-book.

नोट / NOTE :

(I) कृपया जाँच कर लें कि इस प्रश्न-पत्र में मुद्रित पृष्ठ **19** हैं।
Please check that this question paper contains **19** printed pages.

(II) प्रश्न-पत्र में दाहिने हाथ की ओर दिए गए प्रश्न-पत्र कोड को परीक्षार्थी उत्तर-पुस्तिका के मुख-पृष्ठ पर लिखें।
Q.P. Code given on the right hand side of the question paper should be written on the title page of the answer-book by the candidate.

(iii) कृपया जाँच कर लें कि इस प्रश्न-पत्र में **21** प्रश्न हैं।
Please check that this question paper contains **21** questions.

(IV) **कृपया प्रश्न का उत्तर लिखना शुरू करने से पहले, उत्तर-पुस्तिका में यथा स्थान पर प्रश्न का क्रमांक अवश्य लिखें।**
Please write down the Serial Number of the question in the answer-book at the given place before attempting it.

(V) इस प्रश्न-पत्र को पढ़ने के लिए 15 मिनट का समय दिया गया है। प्रश्न-पत्र का वितरण पूर्वाह्न में 10.15 बजे किया जाएगा। 10.15 बजे से 10.30 बजे तक परीक्षार्थी केवल प्रश्न-पत्र को पढ़ेंगे और इस अवधि के दौरान वे उत्तर-पुस्तिका पर कोई उत्तर नहीं लिखेंगे।
15 minute time has been allotted to read this question paper. The question paper will be distributed at 10.15 a.m. From 10.15 a.m. to 10.30 a.m., the candidates will read the question paper only and will not write any answer on the answer-book during this period. []

डेटा विज्ञान

DATA SCIENCE

*निर्धारित समय : **2** घण्टे* | *अधिकतम अंक : **50***

*Time allowed : **2** hours* | *Maximum Marks : **50***

सामान्य निर्देश :

(i) कृपया निर्देशों को ध्यान से पढ़ें।

(ii) इस प्रश्न-पत्र में **दो** खण्डों में **21** प्रश्न हैं : **खण्ड क** और **खण्ड ख**।

(iii) **खण्ड क** में वस्तुनिष्ठ प्रकार के प्रश्न हैं जबकि **खण्ड ख** में विषयपरक प्रकार के प्रश्न हैं।

(iv) दिए गए (5 + 16) = 21 प्रश्नों में से, उम्मीदवार को 2 घंटे के आबंटित (अधिकतम) समय में (5 + 10) = 15 प्रश्नों के उत्तर देने हैं।

(v) किसी विशेष खण्ड के सभी प्रश्नों को सही क्रम में करने का प्रयास किया जाना चाहिए।

(vi) **खण्ड क :** वस्तुनिष्ठ प्रकार के प्रश्न (24 अंक) :

(a) इस खण्ड में **5** प्रश्न हैं।

(b) कोई नकारात्मक अंकन नहीं है।

(c) दिए गए निर्देशों के अनुसार कीजिए।

(d) प्रत्येक प्रश्न/भाग के सामने आबंटित अंकों का उल्लेख किया गया है।

(vii) **खण्ड ख** : विषयपरक प्रकार के प्रश्न (26 अंक) :

(a) इस खण्ड में **16** प्रश्न हैं।

(b) उम्मीदवार को **10** प्रश्न करने हैं।

(c) दिए गए निर्देशों के अनुसार कीजिए।

(d) प्रत्येक प्रश्न/भाग के सामने आबंटित अंकों का उल्लेख किया गया है।

खण्ड क

(वस्तुनिष्ठ प्रकार के प्रश्न) **(24 अंक)**

1. रोज़गार कौशल पर दिए गए *6* प्रश्नों में से किन्हीं *4* के उत्तर दीजिए। *4×1=4*

(i) संचार में संदेश के प्रति प्राप्तकर्ता की स्वीकृति और प्रतिक्रिया को ________ कहा जाता है।

(A) फीडबैक (B) चैनल

(C) सूचना (D) स्थानांतरण

(ii) कार्यस्थल में स्वतंत्र रूप से कार्य करने की क्षमता क्यों महत्त्वपूर्ण है ?

(A) सहकर्मियों पर निर्भरता बढ़ाने के लिए

(B) नौकरी की जिम्मेदारियों को कम करने के लिए

(C) टीम वर्क और सहयोग से बचने के लिए

(D) व्यक्तिगत विकास और निर्णय लेने को बढ़ावा देने के लिए

General Instructions :

(i) *Please read the instructions carefully.*

(ii) *This question paper consists of **21** questions in **two** sections : **Section A** and **Section B**.*

(iii) ***Section A** has Objective Type Questions whereas **Section B** contains Subjective Type Questions.*

(iv) *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.*

(v) *All questions of a particular section must be attempted in the correct order.*

(vi) ***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.*

(vii) ***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

(Objective Type Questions) **(24 Marks)**

1. Answer any **4** out of the given **6** questions on Employability skills. *4×1=4*

(i) The receiver's acknowledgement and response to the message is called __________ in communication.

(A) feedback (B) channel

(C) information (D) transfer

(ii) Why is the ability to work independently important in the workplace ?

(A) To increase dependence on colleagues

(B) To reduce job responsibilities

(C) To avoid team work and collaboration

(D) To promote personal growth and decision-making

(iii) निम्नलिखित में से कौन-सी तनाव प्रबंधन तकनीक ***नहीं*** है ?

(A) शारीरिक व्यायाम

(B) ध्यान

(C) तनावों के बारे में चिंता करना

(D) योग

(iv) कंप्यूटर प्रदर्शन को अनुकूलित करने के लिए नियमित सफाई के दौरान किस प्रकार की फ़ाइलें आमतौर पर हटा दी जाती हैं ?

(A) सिस्टम फ़ाइलें

(B) सॉफ्टवेयर एप्लिकेशन

(C) अस्थायी फ़ाइलें

(D) महत्त्वपूर्ण दस्तावेज़

(v) एक उद्यमी के कौन-से कार्य में एक नई विधि, विचार या उत्पाद बनाना शामिल है ?

(A) नवाचार

(B) आय का विभाजन

(C) जोखिम प्रबंधन

(D) निर्णय लेना

(vi) निम्नलिखित में से कौन-सा पारिस्थितिक तंत्र और जैव-विविधता को नुकसान पहुँचाने वाली अस्थिर प्रथाओं का परिणाम है ?

(A) आर्थिक विकास

(B) पर्यावरणीय गिरावट

(C) अल्पकालिक मुनाफ़ा

(D) जनसंख्या वृद्धि

(iii) Which of the following is ***not*** a stress management technique ?

(A) Physical exercise

(B) Meditation

(C) Worrying about stressors

(D) Yoga

(iv) Which types of files are typically removed during a routine clean-up to optimize computer performance ?

(A) System files

(B) Software applications

(C) Temporary files

(D) Important documents

(v) Which function of an entrepreneur involves creating of a new method, idea or product ?

(A) Innovation

(B) Division of income

(C) Managing risk

(D) Making decisions

(vi) Which of the following is a consequence of unsustainable practices that harm ecosystem and biodiversity ?

(A) Economic growth

(B) Environmental degradation

(C) Short-term profits

(D) Population growth

2. दिए गए ***6*** प्रश्नों में से किन्हीं ***5*** के उत्तर दीजिए । *5×1=5*

(i) डेटा-आधारित उपसमायोजन (subsetting) का उपयोग तब किया जाता है :

(A) जब सभी पंक्तियों (Rows) को चुन लिया जाए

(B) उपसमायोजन (Subsetting) विशिष्ट डेटा दशाओं पर आधारित है

(C) जब संपूर्ण डेटा सेट का सत्यापन हो जाए

(D) जब कॉलमों (Columns) और पंक्तियों (Rows) को यादृच्छिक (randomly) रूप से चुन लिया जाए

(ii) सांख्यिकी समस्या-समाधान (Statistical Problem-solving) प्रक्रिया का मुख्य प्रयोजन क्या है ?

(A) विस्तृत रिपोर्ट तैयार करना

(B) ऑनलाइन डेटा एकत्र करना

(C) डेटा का प्रयोग करते हुए खोजी प्रश्नों के उत्तर देना

(D) चार्ट्स देखना

(iii) डेटा विज्ञान में पूर्वाग्रह (bias) क्या है ?

(A) एक प्रकार के डेटा का अवलोकन (visualization)

(B) डेटा प्रदर्शन (representation) की प्रक्रिया

(C) डेटा विश्लेषण (data analysis) की एक पद्धति

(D) डेटा में प्रत्याशित परिणाम से परिवर्तन (विचलन)

(iv) अभिकथन (A) : विभिन्न परीक्षणों के प्राप्तांकों (scores) की तुलना करने में शतमक (percentiles) मदद करते हैं।

कारण (R) : शतमक (percentiles) एक निर्धारित डेटा प्वॉइंट पर या उससे नीचे वैल्यूज (values) के अनुपात को प्रदर्शित करते हैं।

(A) अभिकथन (A) और कारण (R) दोनों सही हैं और कारण (R), अभिकथन (A) का सही स्पष्टीकरण है।

(B) अभिकथन (A) और कारण (R) दोनों सही हैं, परंतु कारण (R), अभिकथन (A) का सही स्पष्टीकरण ***नहीं*** है।

(C) अभिकथन (A) सही है, परंतु कारण (R) ग़लत है।

(D) अभिकथन (A) ग़लत है, परंतु कारण (R) सही है।

(v) बताइए कि सही है या ग़लत :

विश्लेषण के लिए संग्रहीत डेटा को कभी भी मानव इच्छा में हस्तक्षेप (interfere) नहीं करना चाहिए।

2. Answer any **5** out of the given **6** questions. *5×1=5*

(i) Data-based subsetting is used when :

(A) All the rows are selected

(B) Subsetting is based on specific data conditions

(C) The entire dataset is verified

(D) Columns and rows are randomly selected

(ii) What is the main purpose of the statistical problem-solving process ?

(A) To prepare detailed report

(B) To collect online data

(C) To answer investigative questions using data

(D) To visualize charts

(iii) What is bias in data science ?

(A) A type of data visualization

(B) The process of representing data

(C) A method of data analysis

(D) A deviation from the expected outcome in data

(iv) Assertion (A) : Percentiles help compare scores across different tests.

Reason (R) : Percentiles represent the proportion of values at or below a certain data point.

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

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

(C) Assertion (A) is true, but Reason (R) is false.

(D) Assertion (A) is false, but Reason (R) is true.

(v) State True or False :

Data collected for analysis should never interfere with human will.

(vi) संग्रहण उपकरण (storage device) में डेटा का संग्रहण करते समय डेटा का/की __________ एक अच्छी पद्धति है, ताकि हैकर्स आपके डेटा को न पढ़ सकें।

(A) प्रतिलिपि (Copy) करना (B) फोटो लेना

(C) कोडीकरण (Encrypt) करना (D) फॉर्मेट (Format) करना

3. दिए गए ***6*** प्रश्नों में से किन्हीं ***5*** के उत्तर दीजिए। *5×1=5*

(i) अभिकथन (A) : जहाँ डेटा सेट में भिन्नता (outlier) हों वहाँ माध्यक (median) केन्द्रीय प्रवृत्ति (central tendency) का अधिक प्रभावशाली उपाय है।

कारण (R) : किसी डेटा सेट में माध्यक सभी मानों (values) का औसत (average) है।

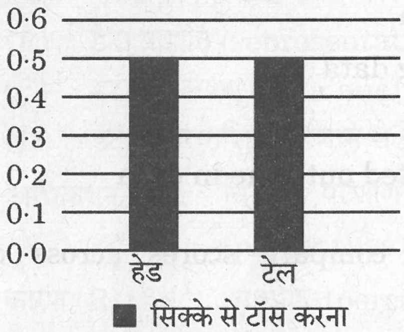
(A) अभिकथन (A) और कारण (R) दोनों सही हैं और कारण (R), अभिकथन (A) का सही स्पष्टीकरण है।

(B) अभिकथन (A) और कारण (R) दोनों सही हैं, परंतु कारण (R), अभिकथन (A) का सही स्पष्टीकरण ***नहीं*** है।

(C) अभिकथन (A) सही है, परंतु कारण (R) ग़लत है।

(D) अभिकथन (A) ग़लत है, परंतु कारण (R) सही है।

(ii) निम्नलिखित ग्राफ द्वारा प्रदर्शित वितरण (distribution) के किस्म का नाम बताइए।



(iii) एक अनुसंधानकर्ता केवल उन परिणामों को देखता है जो वह अपने अध्ययन में देखना चाहता है। यह __________ का एक उदाहरण है।

(A) पुष्टिकरण पूर्वाग्रह (Confirmation bias) (B) रेखिक पूर्वाग्रह (Linearity bias)

(C) स्मरण पूर्वाग्रह (Recall bias) (D) चयन पूर्वाग्रह (Selection bias)

(iv) ज़ेड-स्कोर (Z-Score) प्रदर्शित करता है :

(A) डेटा का सर्वाधिक सामान्य (common) वेल्यू

(B) एक स्कोर से नीचे के वेल्यूज़ (values) का प्रतिशत

(C) मानक विचलनों (standard deviations) की वह संख्या जो औसत (mean) से डेटा प्वॉइंट में हो

(D) डेटा प्वॉइंट्स का कुल योग

(vi) While storing data in a storage device, it is a good practice to ________ the data, so that hackers cannot read your data.

(A) copy (B) take photographs of

(C) encrypt (D) format

3. Answer any **5** out of the given **6** questions. *5×1=5*

(i) Assertion (A) : Median is a more effective measure of central tendency where there are outliers in the data set.

Reason (R) : Median is the average of all the values in a dataset.

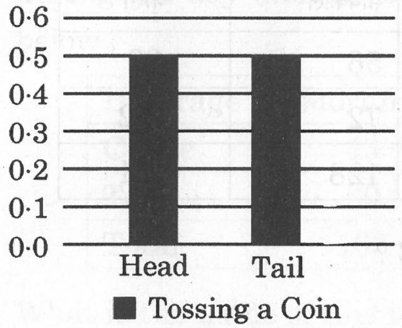
(A) Both Assertion (A) and Reason (R) are true and Reason (R) is the correct explanation of Assertion (A).

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

(C) Assertion (A) is true, but Reason (R) is false.

(D) Assertion (A) is false, but Reason (R) is true.

(ii) Name the type of distribution represented by the following graph.



(iii) A researcher only sees the results they want to see in their study. This is an example of :

(A) Confirmation Bias (B) Linearity Bias

(C) Recall Bias (D) Selection Bias

(iv) The Z-score represents :

(A) The most common value of data

(B) The percentage of values below a score

(C) The number of standard deviations a data point is from the mean

(D) The total sum of the data points

(v) एक डेटा सेट में Q1 = 20, Q3 = 50. IQR क्या है ?

(A) 30 (B) 20

(C) 50 (D) 70

(vi) निम्नलिखित में से कौन-सा डेटा विश्लेषण के लिए नैतिक दिशानिर्देश (ethical guideline) ***नहीं*** है ?

(A) डेटा के सही और विश्वसनीय स्रोत सुनिश्चित करना

(B) निजता और गोपनीयता बनाए रखना

(C) विश्लेषण के लिए केवल संगत डेटा (relevant data) का उपयोग करना

(D) डेटा विश्लेषक अपनी व्यक्तिगत पसंदों के आधार पर डेटा चुन सकते हैं

4. दिए गए ***6*** प्रश्नों में से किन्हीं ***5*** के उत्तर दीजिए । *5×1=5*

(i) दिन के समय चाय या कॉफी पीने वाले अनेक लोगों पर एक सर्वेक्षण किया गया और नीचे दिए गए फॉर्मेट में एक साथ आँकड़े (data) रखे गए :

पेय (beverage)	प्रात:	सायंकाल	कुल
कॉफी	43	56	99
चाय	60	72	132
योग	103	128	231

इस प्रकार की सारणी (table) का क्या नाम है ?

(A) सब्सेट टेबल (Subset table)

(B) टू-वे फ्रीक्वेंसी टेबल (Two-way frequency table)

(C) टू-वे रिलेटिव फ्रीक्वेंसी टेबल (Two-way relative frequency table)

(D) स्टैंडर्ड डेविएशन टेबल (Standard deviation table)

(ii) निम्नलिखित में से कौन-सा सांख्यिकी खोजी प्रश्न (Statistical investigative question) ***नहीं*** है ?

(A) एक दस वर्ष का बालक कितना तेज दौड़ सकता है ?

(B) क्या जो बच्चे उचित नाश्ता करते हैं अधिक तेज दौड़ सकते हैं ?

(C) बच्चे के कार्यकलापों को नींद किस तरह प्रभावित करती है ?

(D) बच्चे ने 200 मीटर दौड़ने में कितना समय लगाया ?

(v) In a dataset, Q1 = 20, Q3 = 50. What is the IQR ?

(A) 30 (B) 20

(C) 50 (D) 70

(vi) Which of the following in ***not*** an ethical guideline for data analysis ?

(A) Ensure accurate and reliable sources of data

(B) Maintain privacy and confidentiality

(C) Use only relevant data for analysis

(D) Data analysts may select data based on their personal preferences

4. Answer any **5** out of the given **6** questions. *5×1=5*

(i) A survey was done on the number of people consuming tea or coffee during the day and the data was put together in the format given below :

Beverage	Morning	Evening	Total
Coffee	43	56	99
Tea	60	72	132
Total	103	128	231

What is the name of this type of table ?

(A) Subset table

(B) Two-way frequency table

(C) Two-way relative frequency table

(D) Standard deviation table

(ii) Which of the following is ***not*** a statistical investigative question ?

(A) How fast can a ten-year-old child run ?

(B) Do children who have proper breakfast run faster ?

(C) How does sleep affect the performance of a child ?

(D) What time did the child take to run 200 m ?

(iii) एक शहर में छात्रों की औसत लंबाई का पता लगाने के लिए एक सर्वेक्षण किया जाता है। सभी छात्रों की लम्बाई नापने के बजाय अनुसंधानकर्ता विभिन्न स्कूलों से साइज 40 के नमूने लेता है और प्रत्येक नमूने के व्यक्तिगत औसत की गणना करता है। इसके बाद इन व्यक्तिगत नमूना औसतों का औसत निकाला जाता है। यह पाया गया कि छात्रों की नमूना-औसत-लम्बाई का आयतचित्र (histogram) सामान्य वितरण (normal distribution) से समानता रखता है। यहाँ कौन-सी सांख्यिकी अवधारणा (statistical concept) प्रदर्शित की गई है ?

(A) संभाव्यता (Probability)

(B) रैखिक पूर्वाग्रह (Linear bias)

(C) केन्द्रीय सीमा प्रमेय (Central Limit Theorem)

(D) सूचना संग्रहण मूल्यांकन (Data evaluation)

(iv) निम्नलिखित में से कौन-सा चतुर्थक (Quartiles) से सही रूप में मेल खाता है ?

(A) Q1 – 25%, Q2 – 50%, Q3 – 75%, Q4 – 100%

(B) Q1 – 10%, Q2 – 20%, Q3 – 30%, Q4 – 50%

(C) Q1 – 33%, Q2 – 66%, Q3 – 100%, Q4 – 99%

(D) Q1 – 50%, Q2 – 75%, Q3 – 100%, Q4 – 25%

(v) औसत से इसकी दूरी के संदर्भ में किसी प्वॉइंट की स्थिति को बताने वाले इस सांख्यिकी शब्द की वैल्यू मानक विचलन यूनिटों (standard deviation units) में नापी जाती है। इस शब्द (term) को क्या कहते हैं ?

(vi) डेटा संचालन संरचना (data governance framework) का मुख्य उद्देश्य क्या है ?

(A) लाभ मार्जिन को बढ़ाना

(B) डेटा का मानकीकरण, एकीकरण, संरक्षण और संग्रहण

(C) केवल डुप्लीकेट रिकॉर्ड्स को हटाना

(D) डेटा प्रत्यक्षीकरण (data visualizations) को तैयार करना

5. दिए गए *6* प्रश्नों में से किन्हीं *5* के उत्तर दीजिए। *5×1=5*

(i) निम्नलिखित में से कौन-सा मानक विचलन (standard deviation) का वास्तविक-जीवन में अनुप्रयोग (real-life application) है ?

(A) Excel में डेटा ग्राफ बनाना

(B) केवल औसत वर्षा की गणना करना

(C) परीक्षा में छात्रों के परफॉर्मेन्स स्प्रेड (performance spread) का मापन करना

(D) स्प्रेडशीट में कॉलमों की संख्या का पता करना

(iii) A survey is conducted to find the average height of students in a city. Instead of measuring all students, a researcher takes samples of size 40 from different schools and calculates the individual mean of each sample. Thereafter the mean of these individual sample means is calculated. It is noticed that the histogram of sample mean heights of students resembles the normal distribution. Which statistical concept is demonstrated here ?

(A) Probability (B) Linear bias

(C) Central Limit Theorem (D) Data evaluation

(iv) Which of the following correctly matches quartiles ?

(A) Q1 – 25%, Q2 – 50%, Q3 – 75%, Q4 – 100%

(B) Q1 – 10%, Q2 – 20%, Q3 – 30%, Q4 – 50%

(C) Q1 – 33%, Q2 – 66%, Q3 – 100%, Q4 – 99%

(D) Q1 – 50%, Q2 – 75%, Q3 – 100%, Q4 – 25%

(v) The value of this statistical term that describes the position of a point in terms of its distance from the mean, when it is measured in standard deviation units. Name this term.

(vi) What is the main aim of a data governance framework ?

(A) To increase profit margins

(B) To standardize, integrate, protect, and store data

(C) To delete duplicate records only

(D) To design data visualizations

5. Answer any **5** out of given **6** questions. *5×1=5*

(i) Which of the following is a real-life application of standard deviation ?

(A) Plotting data graphs in Excel

(B) Calculating average rainfall only

(C) Measuring students' performance spread in a test

(D) Finding the number of columns in a spreadsheet

(ii) डेटा संग्रहण (collection) डिज़ाइन को डेटा की परिवर्तनशीलता (variability) को स्वीकार करना चाहिए। डेटा में परिवर्तनशीलता को कम करने और उसका पता लगाने के लिए कुछ पद्धतियों का प्रयोग किया जाता है। निम्नलिखित में से किस पद्धति का उपयोग किया जा सकता है ?

(A) सांख्यिकीय प्रक्रिया नियंत्रण (Statistical Process Control)

(B) संभाव्यता (Probability)

(C) वितरण (Distribution)

(D) इवेंट (Event)

(iii) अभिकथन (A) : पूर्वाग्रहपूर्ण डेटा (Biased data) से ग़लत पूर्वानुमानित मॉडल हो सकता है।

कारण (R) : पूर्वानुमानित मॉडल (Predictive models) केवल उन डेटा पर विचार करते हैं जो इसके प्रशिक्षण के लिए प्रणाली (system) में फीड किया गया हो।

(A) अभिकथन (A) और कारण (R) दोनों सही हैं और कारण (R), अभिकथन (A) का सही स्पष्टीकरण है।

(B) अभिकथन (A) और कारण (R) दोनों सही हैं, परंतु कारण (R), अभिकथन (A) का सही स्पष्टीकरण *नहीं* है।

(C) अभिकथन (A) सही है, परंतु कारण (R) ग़लत है।

(D) अभिकथन (A) ग़लत है, परंतु कारण (R) सही है।

(iv) आँकड़ों के विलय (data merging) की प्रक्रिया के बारे में निम्नलिखित में से कौन-सा सही है ?

(A) विलीन (merge) किए जा रहे सभी आँकड़ा स्रोत (data sources) हमेशा समान तरीके से वर्गीकृत (grouped) होते हैं।

(B) बहुल आँकड़ा स्रोतों (multiple data sources) के बीच काफी अंतर होता है।

(C) सभी आँकड़ा स्रोत एक ही उद्देश्य से एक ही समय में तैयार किए जाते हैं।

(D) विभिन्न आँकड़ा स्रोतों से विलय किए गए आँकड़ों को सुधारने की कोई आवश्यकता नहीं है।

(ii) Data collection designs must acknowledge variability in data. Few methods are used to reduce and detect variability in data. Which of the following method can be used ?

(A) Statistical Process Control

(B) Probability

(C) Distribution

(D) Event

(iii) Assertion (A) : Biased data can lead to inaccurate predictive models.

Reason (R) : The predictive models consider only the data that is fed into the system for training it.

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

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

(C) Assertion (A) is true, but Reason (R) is false.

(D) Assertion (A) is false, but Reason (R) is true.

(iv) Which of the following is true about the process of data merging ?

(A) All data sources being merged are always grouped in similar manner.

(B) There is a lot of difference between multiple data sources.

(C) All data sources are created with the same objective and at the same time.

(D) No correction is required on the data merged from different data sources.

(v) निम्नलिखित कथन (statements) एक-से-एक जुड़ाव (one-to-one join) के संबंध में हैं :

(i) एक टेबल में प्रत्येक पंक्ति (row) को दूसरी टेबल के एकल पंक्ति (row) से जोड़ा जाता है।

(ii) **'एक छात्र अनेक पाठ्यक्रमों मे पंजीकरण कर सकता है'** इस जुड़ाव (join) का वैध (मान्य) उदाहरण है।

(iii) दो टेबल्स की सिंगल पंक्ति (single row) के बीच जोड़ने का कार्य की कॉलम (Key column) के उपयोग से किया जाता है।

(iv) टेबल्स को जोड़ने के लिए प्रयुक्त की फील्ड (Key field) को अद्वितीय मूल्यों (Unique values) को अंतर्विष्ट (Contain) करने के लिए तैयार किया जाता है।

(v) **'एक छात्र के पास केवल एक पहचान पत्र (Id) हो सकता है'** इस जुड़ाव (join) का एक वैध (मान्य) उदाहरण है।

दिए गए कथनों में से कौन-सा सही हैं ?

(A) (i), (ii) और (iv)
(B) (i), (iii), (iv) और (v)
(C) (ii), (iv) और (v)
(D) (ii), (iii), (iv) और (v)

(vi) सहमति के साथ किसी व्यक्ति से एकत्रित निजी आँकड़े ऐसे होने चाहिए कि :

(A) जिसको आवश्यकता हो ऐसे किसी भी व्यक्ति के साथ इनकी बिना रोक-टोक के साझेदारी हो सके
(B) आँकड़ा विश्लेषण के लिए इन्हें उपलब्ध किया जा सके
(C) हमेशा गोपनीयता के साथ इनका संचालन किया जा सके
(D) किसी भी परिस्थिति में कभी भी इनकी लेखापरीक्षा (audit) न हो

खण्ड ख

(विषयपरक प्रकार के प्रश्न) (26 अंक)

*रोज़गार कौशल पर दिए गए **5** प्रश्नों में से किन्हीं **3** के उत्तर दीजिए। प्रत्येक प्रश्न का उत्तर 20 – 30 शब्दों में दीजिए।* *3×2=6*

6. संचार में भाषाई बाधाएँ क्या हैं ?

7. भावनात्मक बुद्धिमत्ता को प्रबंधित करने के चरण बताइए।

8. (क) स्पैम (SPAM) मेल क्या हैं ?

(ख) क्या हमें उनका जवाब देना चाहिए ?

9. उद्यमिता के बारे में कोई दो मिथक लिखिए।

10. हम संपोषणीय (sustainable) शहर कैसे बना सकते हैं ?

(v) The following statements are with respect to one-to-one join :

(i) Each row in one table is linked to a single row in another table.

(ii) **'A student can register in multiple courses'** is a valid example of this join.

(iii) The linking between single row of two tables is done using a key column.

(iv) The key field used to link the tables is designed to contain unique values.

(v) **'A student can have only one Student Id'** is a valid example of this join.

Which of the given statements are correct ?

(A) (i), (ii) and (iv)
(B) (i), (iii), (iv) and (v)
(C) (ii), (iv) and (v)
(D) (ii), (iii), (iv) and (v)

(vi) Private data collected from a person with consent should :

(A) Be freely shared with anyone who needs it
(B) Be made available for data analysis
(C) Always be handled with confidentiality
(D) Never be audited under any circumstance

SECTION B

(Subjective Type Questions) (26 Marks)

*Answer any **3** out of the given **5** questions on Employability skills. Answer each question in 20 – 30 words.* $3 \times 2 = 6$

6. What are linguistic barriers to communication ?

7. State the steps to manage emotional intelligence.

8. (a) What are SPAM mails ?
   (b) Should we respond to them ?

9. Write any two myths about entrepreneurship.

10. How can we create sustainable cities ?

दिए गए 6 प्रश्नों में से किन्हीं 4 के उत्तर 20 – 30 शब्दों (प्रत्येक) में दीजिए। 4×2=8

11. निम्नलिखित डेटासेट पर विचार कीजिए :
[5, 10, 2, 1, 20, 6, 15]
इसका माध्य (mean) और माध्यक (median) ज्ञात कीजिए।

12. डेटा पृथक (discrete) या सतत (continuous) हो सकता है। पृथक और सतत डेटा के बीच एक अंतर बताइए। प्रत्येक का एक उदाहरण भी दीजिए।

13. चयन पूर्वाग्रह (selection bias) को परिभाषित कीजिए। यह कब होता है ?

14. यदि Z-score का वैल्यू (value) धनात्मक (positive) है या ऋणात्मक (negative), तो आप इससे क्या समझते हैं ? अपने उत्तर के समर्थन में उदाहरण दीजिए।

15. अनेक के साथ अनेक जुड़ाव (Many-to-many join) को एक उदाहरण के साथ स्पष्ट कीजिए।

16 एक कंपनी, ग्राहकों के खरीददारी आँकड़े का विश्लेषण कर रही है। डेटासेट को छोटा बनाने के लिए विश्लेषक बिना औचित्य के 18 वर्ष से नीचे के ग्राहकों के सारे रिकॉर्ड छोड़ने का निर्णय लेता है।

(a) इस मामले में नैतिक मुद्दे का पता कीजिए।

(b) इसे निकालते (discard) समय क्या कंपनी को डेटा सॉफ्ट डिलीट कर देना चाहिए। क्यों या क्यों नहीं ?

दिए गए 5 प्रश्नों में से किन्हीं 3 के उत्तर 50 – 80 शब्दों (प्रत्येक) में दीजिए। 3×4=12

17. औसत निरपेक्ष विचलन (Mean Absolute Deviation) (MAD) को परिभाषित कीजिए। उदाहरण के रूप में निम्नलिखित डेटासेट को लेकर MAD संगणना के कदमों (steps) को स्पष्ट कीजिए।
[10, 12, 14, 16, 18]

18. डेटा विज्ञान में वितरण (distribution) से आप क्या समझते हैं ? यह संभाव्यता (probability) से किस प्रकार भिन्न है ? सिक्का उछालना (Tossing the coin) इवेंट की सहायता से स्पष्ट कीजिए।

19. (a) सांख्यिकी में केन्द्रीय सीमा प्रमेय (Central Limit Theorem) (CLT) को महत्त्वपूर्ण क्यों माना जाता है ?

(b) CLT के संदर्भ में, यदि डेटा का नमूना आकार बढ़ता है, तो त्रुटि बढ़ेगी या कम होगी ?

(c) CLT के किन्हीं दो वास्तविक दुनिया में अनुप्रयोगों (real world applications) का उल्लेख कीजिए।

20. (a) दशमक (deciles) क्या हैं ?

(b) 15 संख्याओं के निम्नलिखित डेटासेट पर विचार कीजिए :
[77, 60, 63, 36, 54, 57, 36, 72, 55, 51, 32, 56, 33, 42, 55]
D_1 तथा D_5 के डेटा स्थिति और decile value की गणना कीजिए।

21. डेटा का प्रयोग हो जाने के बाद उसे अच्छी तरह निकाल देना (discard) क्यों महत्त्वपूर्ण है ? गोपनीय डेटा की हार्ड कॉपी (physical copies) हटाने के तीन तरीकों को स्पष्ट कीजिए।

*Answer any **4** out of the given **6** questions in 20 – 30 words each.* *4×2=8*

11. Consider the following dataset :
[5, 10, 2, 1, 20, 6, 15]
Find the mean and median.

12. The data can be discrete or continuous. Give one difference between discrete and continuous data. Also give one example of each.

13. Define selection bias. When does it occur ?

14. What do you infer if the value of Z-score is positive or negative ? Give example to support your answer.

15. Explain Many-to-Many join with an example.

16. A company is analyzing purchase data of customers. The analyst decides to discard all records from customers under 18 without justification, to make the dataset smaller.

(a) Identify the ethical issue in this case.

(b) Should the company soft delete the data while discarding it ? Why/Why not ?

*Answer any **3** out of the given **5** questions in 50 – 80 words each.* *3×4=12*

17. Define Mean Absolute Deviation (MAD). Explain the steps to calculate MAD taking the following dataset as an example :
[10, 12, 14, 16, 18]

18. What do you mean by distribution in data science ? How is it different from probability ? Explain with the help of the event – Tossing the coin.

19. (a) Why is Central Limit Theorem (CLT) considered important in statistics ?

(b) With reference to CLT, if the sample size of data increases, will the error increase or decrease ?

(c) Mention any two real world applications of CLT.

20. (a) What are deciles ?

(b) Consider the following dataset of 15 numbers :

[77, 60, 63, 36, 54, 57, 36, 72, 55, 51, 32, 56, 33, 42, 55]

Calculate the Data Position and decile value for D_1 and D_5.

21. Why is it important to discard data properly once its use is over ? Explain the three methods to discard physical copies of confidential data.