Series ωZWYX

Set-4

Q.P. Code 106

Roll No.

Candidates must write the Q.P. Code on the title page of the answer-book.

DATA SCIENCE

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

- Please check that this question paper contains **11** printed pages.
- 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.
- Please check that this question paper contains **21** questions.
- **Please write down the serial number of the question in the answer-book before attempting it.**
- 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 students will read the question paper only and will not write any answer on the answer-book during this period.

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, the 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) Stress management can help you to

(a) have a joyful life.

(b) have a stressful life.

(c) be less energetic.

(d) have no time for friends and family.

(ii) Which type of motivation makes us do things because they make us happy ?

(a) Internal motivation

(b) External motivation

(c) Both (a) and (b)

(d) Neither (a) nor (b)

(iii) Which of the following is a mobile operating system ?

(a) Ubuntu (b) Microsoft Windows

(c) Google Android (d) macOS

(iv) Installing the _____ helps in increasing the performance of the computer.

(a) spam files

(b) disk cleaner software

(c) temporary files

(d) extra files

(v) Making decisions is an important _____ of an entrepreneur.

(a) misconception (b) hindrance

(c) function (d) myth

(vi) A/An _____ is a line of work that a person takes for life.

(a) impression (b) commission

(c) quality (d) career

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

(i) State whether the following statement is *true* or *false* :

Subsetting helps to focus first on the required set of data.

(ii) Which of the following statement is *true* about median ?

(a) It is the smallest element of a sorted data set.

(b) It is the middle point of a sorted data set.

(c) It is the middle point of an unsorted data set.

(d) It is the smallest element of an unsorted data set.

(iii) Standard Deviation represents how much the data is spread out around the _____ or an average.

(a) Standard deviation

(b) Mean deviation

(c) Mean

(d) Median

(iv) Two-way relative frequency tables represent the percentage of _____ that fit in each category.

(a) preference (b) data points

(c) mean values (d) frequency

(v) _____ is an example of discrete data.

(a) Pass or Fail in a Math Test

(b) Length of a road

(c) Weather forecast

(d) Number of stars

(vi) Which of the following is ***not*** *true* for discarding the data ?

(a) This helps us to prevent unauthorized access to the data.

(b) It is important for all of us to make sure that we discard the digital data in a proper way.

(c) In most of the devices, if you do a soft delete of a particular file, this file deletes from the original space and cannot be restored.

(d) With the increased amount and intensity of cyber attacks, it is important for all of us to make sure that we discard the digital data in a proper way.

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

(i) Data Merging is the process of combining two or more data sets into a single _____ .

(a) Data Frame

(b) Data Designing

(c) Data Graphing

(d) Data Handling

(ii) _____ bias is an outcome of seeing what you want to see in the data.

(a) Linearity

(b) Selection

(c) Survivor

(d) Confirmation

(iii) _____ is all about counting randomness.

(a) Mean

(b) Median

(c) Probability

(d) Graphs

(iv) _____ is an example of real-life implementation of standard deviation.

(a) Grading tests

(b) Sum of values

(c) Sorting of data

(d) Spellcheck

(v) The private information that is shared should _____ be handled with confidentiality.

(a) never

(b) sometimes

(c) always

(d) often

(vi) There are two ways in which you can store the data — in the _____ format or as a physical copy.

(a) analog

(b) printed

(c) hardcopy

(d) digital

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

(i) A percentile can be defined as the percentage of the total ordered observations at or _____ it.

(a) above

(b) equal

(c) below

(d) not at

(ii) The value of a z-score always tells us the number of standard deviations we are away from _____ .

(a) median

(b) mean

(c) standard deviation

(d) percentile

(iii) One-to-one join works by _____ the data tables using the Primary key.

(a) merging

(b) deleting

(c) inserting

(d) concatenating

(iv) Which of the following is *true* about distribution of an event ?

(a) It shows no values.

(b) It consists of only input values that can be seen.

(c) It consists of all possible values but not input values.

(d) It consists of both input values that can be seen and also all possible values.

(v) Which of the following is the last step of the Statistical Problem Solving Process ?

(a) Analyse the data

(b) Formulate statistical investigative questions

(c) Interpret the data

(d) Collect the data

(vi) Which of the following is used to find out the interquartile range ?

(a) Values of quartiles

(b) Values of percentiles

(c) Values of deciles

(d) Values of z-score

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

(i) The weight of students in a class has a mean of 22 with standard deviation of 4. What is the z-score for a student who weighs 30 kg ?

(a) 0 (b) 1

(c) 2 (d) 3

(ii) The median of the given dataset values 7, 12, 15, 16, 20 is

(a) 15 (b) 10

(c) 11 (d) 12

(iii) The mean absolute deviation for the data 3, 5, 9, 10, 15, 16 is

(a) 5 (b) 3

(c) 4 (d) 6

(iv) If two coins are tossed at the same time, what is the probability of getting Head-Tail ?

(a) 0·50

(b) 0·45

(c) 0·25

(d) 0·35

(v) Data collection designs must acknowledge _____ in data.

(a) Statistical Process Control

(b) Test Data

(c) Variability

(d) Experimental Designs

(vi) If the data that is fed into the system is _____, model accuracy and fidelity are compromised.

(a) biased

(b) predictive

(c) accurate

(d) continuous

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.* *3×2=6*

6. What is self-motivation ? Give an example.

7. Explain any two steps for Effective Time Management.

8. Define any ***two*** of the following :

(a) Virus

(b) Identity Theft

(c) Software Piracy

9. What are the two ways of employment in which one can earn a living ? Give an example for each.

10. What are the qualities of an entrepreneur ?

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

11. What is Data-based subsetting ? Give an example.

12. What is Continuous data ? Give an example.

13. Why is Central Limit Theorem important ?

14. What is Many-to-Many Join ? Give an example to support your answer.

15. Name the two ways in which data can be stored. Give an example.

16. Define the term percentile. Give an example.

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

17. Give the steps to calculate Standard Deviation.

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

(a) Formulate statistical investigative questions

(b) Collect/Consider the data

19. Differentiate between Recall Bias and Survivor Bias. Give an example of each.

20. What is the Interquartile Range ? Give an example to calculate an interquartile range.

21. Mention any two ways to discard the data stored in physical copy and any two ways to discard digital data.