Affiliated to CBSE, New Delhi (Affiliation No. 1030642)

GNPS

GOMTI NANDAN PUBLIC SCHOOL

Gandhi Ward, Bina (M.P.) 470113. | 9165 609 333. | info@mygnps.com | www.mygnps.com

CLASS X · ARTIFICIAL INTELLIGENCE (417) · SKILL SUBJECT · CBSE BOARD EXAMINATION 2027

PART B · UNIT 4

Statistical Data

Unit Booklet · Short Notes · Board-Pattern Questions · Model Answers

Practical unit · no theory marks | 28 questions with model answers | Sonakshi Ma'am | Taught: Thu 26 Nov (lab)

INSIDE THIS BOOKLET

1	**Unit Notes** 8 note blocks, key terms, exam tips	2
A	**Objective Questions** 16 questions · 1 mark	6
B	**Short Answer Questions** 6 questions · 2 marks	7
C	**Long Answer Questions** 4 questions · 4 marks	7
D	**Case-Based Questions** 2 cases · 4 marks	8
E	**Answers** Key + model answers, marking-scheme style	9

STUDENT DETAILS

NAME _____

CLASS & SECTION _____

ROLL NO. _____

UNIT FINISHED ON _____

OBJECTIVE SELF-CHECK _____ / 16

1

READ THIS FIRST · SHORT NOTES

Unit Notes

What data science is, no-code and low-code AI compared with coding, the statistics you need (mean, median, mode, range, standard deviation) with Excel functions, the Orange Data Mining tool, and the Palmer penguins project worked through the AI project cycle.

THIS UNIT IN THE BOARD PAPER

0 / 40

Unit 4 has **no theory marks**. It is tested in the **practical exam** (Units 4–7 together carry 15 marks), the **practical file** and the **viva**, and the penguin project is one of the suggested projects. Use these questions to prepare for the viva and the lab work.

SYLLABUS 2026–27 · WHAT CBSE LISTS FOR THIS UNIT

- Data science and its applications
- No-code and low-code AI; how they differ from code-based AI; some no-code tools
- Important statistics concepts
- Orange Data Mining tool
- AI project cycle with the Palmer penguins use case
- Activity: statistical analysis in MS Excel

HOW THE CBSE SKILL-SUBJECT PAPER IS SET (THEORY 50 MARKS · 2 HOURS)

1 mark
Objective: MCQ, assertion–reason, fill in the blank, true/false, match. Q1 is Part A; Q2–Q5 (answer 5 of 6 in each) are Part B.

2 marks
Short answer in **20–30 words**: two clear points. Part B: answer any 4 of 6 (Q11–Q16).

4 marks
Long answer in **50–80 words**: four points, or 2 + 2. Answer any 3 of 5 (Q17–Q21); often case-based.

Part B = 40
Objective 20 + short 8 + long 12. Part A (Employability Skills) = 10. Practical = 50.

Write in points, start each point with the key word, and stay inside the word limit. Objective answers: write the option letter **and** the answer.

PRACTICAL WORK FOR THIS UNIT

Find mean, median, mode, range and standard deviation of a data set in MS Excel · load the penguins data in Orange and explore it with Data Table, Distributions and Scatter Plot · build a Tree or kNN model and check it with Test and Score and Confusion Matrix · write each activity in the practical file.

1 Data science

Data science: the use of statistics, computer tools and AI to collect, clean, study and draw useful conclusions from data.

Field	Use of data science
Health	Predicting disease from test reports and scans
Shopping and OTT	Recommending products, songs and films
Banking	Spotting fraud in card payments
Weather and farming	Forecasting rain; advising farmers on crops
Sports	Studying player performance and match strategy
Transport	Best routes and traffic prediction in map apps

2 No-code, low-code and code-based AI

	No-code AI	Low-code AI	Code-based AI
How	Drag-and-drop blocks or widgets, no programming	Mostly visual, with a little code	Programs written in Python etc.
Who	Beginners, teachers, domain experts	Users with some coding	Programmers, data scientists
Speed	Very quick to build	Quick	Slower; needs testing
Control	Limited to the tool's features	Some customisation	Full control

No-code tools in the syllabus: Orange Data Mining (data analysis and models), **Teachable Machine** (image, sound and pose models in a browser) and **Lobe** (image classification).

- **Benefits:** easy, fast, visual; anyone can try AI; good for learning.
- **Limits:** less flexible; hard for very large or unusual projects; depends on the tool.

3 Statistics – key ideas

- **Statistics:** collecting, organising, analysing and interpreting data.
- **Numerical data:** numbers – discrete (counted: 3 goals) or continuous (measured: 1.62 m).
- **Categorical data:** groups or labels – species, colour, island.
- **Mean** = sum of values ÷ number of values (the average).
- **Median** = middle value after arranging in order; for an even count, the average of the two middle values.
- **Mode** = the value that occurs most often.
- **Range** = largest – smallest.
- **Variance** = average of the squared distances from the mean; **standard deviation (SD)** = √variance. A small SD means values are close to the mean; a large SD means they are spread out.

4 Worked example

Data: **2, 4, 4, 4, 5, 5, 7, 9** (8 values)

Measure	Working	Answer
Mean	40 ÷ 8	**5**
Median	(4 + 5) ÷ 2 (4th and 5th values)	**4.5**
Mode	4 occurs three times	**4**
Range	9 – 2	**7**
Variance	(9+1+1+1+0+0+4+16) ÷ 8 = 32 ÷ 8	**4**
SD	√4	**2**

Here variance divides by the number of values (n), as Excel's VAR.P and STDEV.P do.

5 Statistics in MS Excel

Function	Gives	For B2:B9 holding the data above
=AVERAGE(B2:B9)	Mean	5
=MEDIAN(B2:B9)	Median	4.5
=MODE.SNGL(B2:B9)	Mode	4
=MAX(B2:B9)−MIN(B2:B9)	Range	7
=VAR.P(B2:B9)	Variance	4
=STDEV.P(B2:B9)	Standard deviation	2
=COUNT(B2:B9)	Number of values	8

Excel can also draw **charts** (column, line, scatter) to show patterns.

6 Orange Data Mining

- Free, **open-source** tool for data analysis and machine learning, made at the University of Ljubljana, Slovenia.
- Uses **visual programming**: you place **widgets** on a canvas and join them with lines to make a **workflow**; data flows along the lines.

Widget	Job
File	Loads a data set
Data Table	Shows the data in rows and columns
Distributions · Box Plot · Scatter Plot	Charts to explore the data
Feature Statistics	Mean, median, spread and missing values of each column
Tree · kNN · Logistic Regression	Learners that build classification models
Test and Score	Checks the model: accuracy (CA), precision, recall, F1
Confusion Matrix · Predictions	Shows right and wrong predictions

7 Palmer penguins – through the AI project cycle

Stage	In the penguins project
Problem scoping	Identify a penguin's **species** from body measurements. Who: researchers; Where: Palmer Archipelago, Antarctica; Why: quick, correct identification without experts.
Data acquisition	Data set of 344 penguins: 3 species (**Adelie, Chinstrap, Gentoo**), 3 islands, bill length, bill depth, flipper length, body mass, sex.
Data exploration	Scatter plot of flipper length vs bill length shows three groups; Gentoo are heaviest with the longest flippers. A few values are missing.
Modelling	Classification: a learner such as Tree or kNN is trained on labelled data (features → species).
Evaluation	Test and Score and Confusion Matrix show how many penguins are classified correctly.

8 No-code workflow in Orange

1. **File** – load the penguins data
2. **Data Table** – look at the rows, columns and target (species)
3. **Scatter Plot / Distributions** – explore patterns
4. **Learner** (Tree or kNN) – build the model
5. **Test and Score** – note accuracy (CA) and F1
6. **Confusion Matrix** and **Predictions** – see right and wrong answers

Each Orange step matches a stage of the AI project cycle: File and Data Table = data acquisition; plots = exploration; learner = modelling; Test and Score = evaluation.

Key terms

15 terms

Data science: drawing useful conclusions from data

No-code AI: building AI by drag and drop, no programming

Low-code AI: mostly visual with a little code

Mean: sum ÷ number of values

Median: middle value in order

Mode: most frequent value

Range: largest – smallest

Variance: average squared distance from mean

Standard deviation: square root of variance; spread

Categorical data: labels or groups

Widget: a block in Orange that does one job

Workflow: widgets joined to pass data

Classification: predicting a category, e.g. species

CA: classification accuracy in Orange

Palmer penguins: data set of 344 penguins of 3 species

Common mistakes and exam tips

AVOID THESE

- Median: **arrange in order first**. With an even count, average the two middle values.
- Mean is affected by extreme values; median is not.
- SD tells the **spread**, not the average.
- In Orange, data flows from **left to right** along the lines between widgets.
- Penguin species are the **target (label)**; measurements are the **features**.

A

PART A · 1 MARK EACH

Objective Questions

16 questions: MCQs, assertion–reason, fill in the blanks and true/false, like Q2–Q5 of the board paper. Answers in Part E.

ASSERTION–REASON: HOW TO ANSWER

(a) Both A and R are true, and R is the correct explanation of A.
(b) Both A and R are true, but R is not the correct explanation of A.
(c) A is true, but R is false.
(d) A is false, but R is true.

A1. Which of these is a no-code AI tool? [1]

(a) Python **(b)** Orange Data Mining
(c) C++ **(d)** Java

A2. The median of 11, 3, 8, 5, 9 is: [1]

(a) 8 **(b)** 5
(c) 9 **(d)** 7.2

A3. The mode of 3, 7, 7, 2, 9, 7, 4 is: [1]

(a) 2 **(b)** 7
(c) 9 **(d)** 5

A4. The mean of 6, 8, 10, 12, 14 is: [1]

(a) 8 **(b)** 10
(c) 12 **(d)** 50

A5. Which measure tells how spread out the data is? [1]

(a) Mean **(b)** Mode
(c) Standard deviation **(d)** Median

A6. Which Excel function gives the average? [1]

(a) =MEDIAN() **(b)** =AVERAGE()
(c) =MODE.SNGL() **(d)** =COUNT()

A7. In Orange, a block that does one job, such as loading a file, is called a: [1]

(a) Cell **(b)** Widget
(c) Layer **(d)** Pixel

A8. In the penguins project, predicting the species is a: [1]

(a) Regression task **(b)** Clustering task
(c) Classification task **(d)** Sorting task

A9. Which Orange widget reports the accuracy (CA) of a model? [1]

(a) File **(b)** Scatter Plot
(c) Test and Score **(d)** Data Table

A10. Species, colour and island are examples of: [1]

(a) Continuous data (b) Categorical data
(c) Discrete numbers (d) Standard deviations

A11. ASSERTION–REASON [1]

Assertion (A): No-code tools let beginners build AI models quickly.
Reason (R): They use drag-and-drop blocks instead of written programs.

(a) Both A and R are true, and R is the correct explanation of A.
(b) Both A and R are true, but R is not the correct explanation of A.
(c) A is true, but R is false.
(d) A is false, but R is true.

A12. ASSERTION–REASON [1]

Assertion (A): Code-based AI gives less control than no-code AI.
Reason (R): No-code tools are limited to the features they offer.

(a) Both A and R are true, and R is the correct explanation of A.
(b) Both A and R are true, but R is not the correct explanation of A.
(c) A is true, but R is false.
(d) A is false, but R is true.

A13. FILL IN THE BLANK [1]

Range = largest value – _____ value.

A14. FILL IN THE BLANK [1]

The three penguin species in the data set are Adelie, Chinstrap and _____.

A15. TRUE OR FALSE [1]

Standard deviation is the square root of variance.

A16. TRUE OR FALSE [1]

To find the median you do not need to arrange the data in order.

B — Short Answer Questions

PART B · 2 MARKS EACH

Answer in 20–30 words: two clear points. Model answers in Part E.

B1. What is data science? Give one application. MUST DO [2]

B2. Differentiate between no-code and code-based AI. MUST DO [2]

B3. Find the mean and median of 4, 10, 6, 8, 2. [2]

B4. Name two widgets of Orange and their use. [2]

B5. Give one advantage and one limitation of no-code AI tools. [2]

B6. What do mean and standard deviation tell us about data? [2]

C — Long Answer Questions

PART C · 4 MARKS EACH

Answer in 50–80 words: four points, or two points with an example each. Model answers in Part E.

C1. Explain the Palmer penguins project through the stages of the AI project cycle. MUST DO [4]

C2. For the data 2, 4, 4, 4, 5, 5, 7, 9 find the mean, median, mode and range. MUST DO [4]

C3. Describe the steps to build a classification model in Orange. [4]

C4. Write four Excel functions used for statistical analysis and what each gives. [4]

PART D · CASE AND COMPETENCY BASED

D Case-Based Questions

Read the situation, then answer from it and from the unit. These are the 4-mark board questions that test application.

D1. Case-Based · Weekly temperatures in Excel [4]

Ishaan records the afternoon temperature (°C) for five days in cells B2 to B6 of a spreadsheet: **30, 32, 31, 35, 32**. His teacher asks him to find some statistics using Excel functions.

(i) Write the formula for the mean and find its value. [1]

(ii) Find the median and the mode. [1]

(iii) Find the range. Which Excel function shows how spread out the temperatures are? [2]

D2. Competency-Based · Which penguin is it? [4]

Students load the penguins data into Orange. A scatter plot of flipper length against bill length shows three clear groups. They train a Tree model, and Test and Score shows a classification accuracy (CA) of 0.97.

(i) What is the target (label) and what are the features? [1]

(ii) What does CA = 0.97 mean? [1]

(iii) Name the AI project cycle stage for (a) the scatter plot and (b) Test and Score. Name one more widget to see wrong predictions. [2]

E

PART E · ANSWERS IN CBSE MARKING-SCHEME STYLE

Answers

Objective key with a reason, then model answers: numbered value points, key word first, the mark for each point in orange, and the word count under each answer. Any other correct, relevant point also earns the mark.

Objective questions – answer key

Part A

Write the option letter **and** the answer, as the marking scheme does.

Q	Answer	Why (one line)
A1	**(b) Orange Data Mining**	Drag-and-drop widgets.
A2	**(a) 8**	In order: 3, 5, 8, 9, 11 – middle is 8.
A3	**(b) 7**	7 occurs three times.
A4	**(b) 10**	50 ÷ 5 = 10.
A5	**(c) Standard deviation**	Small SD = values close together.
A6	**(b) =AVERAGE()**	Mean.
A7	**(b) Widget**	Widgets are joined into workflows.
A8	**(c) Classification task**	Output is a category.
A9	**(c) Test and Score**	Evaluation stage.
A10	**(b) Categorical data**	Labels or groups.
A11	**(a) Both true; R explains A**	R explains A.
A12	**(d) A false, R true**	A false (code gives full control); R true.
A13	**smallest**	Measures spread simply.
A14	**Gentoo**	Gentoo are the heaviest.
A15	**True**	SD = √variance.
A16	**False**	Correct: you must arrange it in order first; then take the middle.

Short answers

Part B · 2 marks

B1. What is data science? Give one application. [2]

1. **Meaning:** using statistics, computers and AI to draw conclusions from data. 1
2. **Application:** recommending films on OTT or spotting bank fraud. 1

20 words · CBSE asks 20–30 words | Marking: 1 + 1 = 2 | 1 + 1 | MUST DO

B2. Differentiate between no-code and code-based AI. [2]

1. **No-code:** drag-and-drop, no programming, quick but limited. 1
2. **Code-based:** written programs, slower, full control. 1

13 words · CBSE asks 20–30 words | Marking: 1 + 1 = 2 | 1 + 1 | MUST DO

B3. Find the mean and median of 4, 10, 6, 8, 2. [2]

1. **Mean:** 30 ÷ 5 = 6. 1
2. **Median:** in order 2, 4, 6, 8, 10 – median 6. 1

16 words · CBSE asks 20–30 words | Marking: 1 + 1 = 2 | 1 + 1

B4. Name two widgets of Orange and their use. [2]

1. **File:** loads the data set. 1
2. **Scatter Plot:** shows the relation between two features. 1

13 words · CBSE asks 20–30 words | Marking: 1 + 1 = 2 | 1 + 1 (any two widgets)

B5. Give one advantage and one limitation of no-code AI tools. [2]

1. **Advantage:** quick and easy for non-programmers. 1
2. **Limitation:** less flexible; limited to the tool's features. 1

14 words · CBSE asks 20–30 words | Marking: 1 + 1 = 2 | 1 + 1

B6. What do mean and standard deviation tell us about data? [2]

1. **Mean:** the average or central value. 1
2. **Standard deviation:** how spread out the values are from the mean. 1

17 words · CBSE asks 20–30 words | Marking: 1 + 1 = 2 | 1 + 1

Long answers

Part C · 4 marks

C1. Explain the Palmer penguins project through the stages of the AI project cycle. [4]

1. **Problem scoping:** identify species from body measurements. 1
2. **Data acquisition:** 344 penguins, three species, measurements. 1
3. **Exploration:** scatter plots show three groups. 1
4. **Modelling and evaluation:** train a Tree or kNN; check with Test and Score. 1

33 words · CBSE asks 50–80 words | Marking: 1 + 1 + 1 + 1 = 4 | 1 mark each | MUST DO

C2. For the data 2, 4, 4, 4, 5, 5, 7, 9 find the mean, median, mode and range. [4]

1. **Mean:** 40 ÷ 8 = 5. 1
2. **Median:** (4 + 5) ÷ 2 = 4.5. 1
3. **Mode:** 4. 1
4. **Range:** 9 – 2 = 7. 1

22 words · CBSE asks 50–80 words | Marking: 1 + 1 + 1 + 1 = 4 | 1 mark each | MUST DO

C3. Describe the steps to build a classification model in Orange. [4]

1. **Load:** File widget loads data; Data Table shows it. 1
2. **Explore:** Scatter Plot and Distributions show patterns. 1
3. **Model:** connect a learner such as Tree or kNN. 1
4. **Evaluate:** Test and Score and Confusion Matrix check accuracy. 1

34 words · CBSE asks 50–80 words | Marking: 1 + 1 + 1 + 1 = 4 | 1 mark each

C4. Write four Excel functions used for statistical analysis and what each gives. [4]

1. **AVERAGE:** mean. 1
2. **MEDIAN:** middle value. 1
3. **MODE.SNGL:** most frequent value. 1
4. **STDEV.P:** standard deviation. 1

12 words · CBSE asks 50–80 words | Marking: 1 + 1 + 1 + 1 = 4 | 1 mark each

Case-based answers

Part D

D1. Case-Based · Weekly temperatures in Excel (passage printed in Part D) [4]

(i) Write the formula for the mean and find its value. [1]
=AVERAGE(B2:B6) = 160 ÷ 5 = **32 °C**.

(ii) Find the median and the mode. [1]
In order 30, 31, 32, 32, 35: median = **32**; mode = **32**.

(iii) Find the range. Which Excel function shows how spread out the temperatures are? [2]
Range = 35 − 30 = **5 °C**. Spread: **=STDEV.P(B2:B6)** (standard deviation).

Marking: 1 + 1 + 2 = 4 | 1 + 1 + (1 + 1)

D2. Competency-Based · Which penguin is it? (passage printed in Part D) [4]

(i) What is the target (label) and what are the features? [1]
Target: **species**. Features: bill length, bill depth, flipper length, body mass (and island, sex).

(ii) What does CA = 0.97 mean? [1]
The model classified **97%** of the penguins correctly.

(iii) Name the AI project cycle stage for (a) the scatter plot and (b) Test and Score. Name one more widget to see wrong predictions. [2]
(a) **Data exploration**; (b) **evaluation**. **Confusion Matrix** (or Predictions) shows the wrong predictions.

Marking: 1 + 1 + 2 = 4 | 1 + 1 + 2