

Chapter 3: Organisation of Data

NCERT · Class 11 · Statistics

Textbook exercises

Question 1

Which of the following alternatives is true?

(i) The class midpoint is equal to: (a) The average of the upper class limit and the lower class limit. (b) The product of upper class limit and the lower class limit. (c) The ratio of the upper class limit and the lower class limit. (d) None of the above.

(ii) The frequency distribution of two variables is known as (a) Univariate Distribution (b) Bivariate Distribution (c) Multivariate Distribution (d) None of the above

(iii) Statistical calculations in classified data are based on (a) the actual values of observations (b) the upper class limits (c) the lower class limits (d) the class midpoints

(iv) Range is the (a) difference between the largest and the smallest observations (b) difference between the smallest and the largest observations (c) average of the largest and the smallest observations (d) ratio of the largest to the smallest observation

Solution

ANSWER (i) (a) The average of the upper class limit and the lower class limit. (Class mark = $(\text{Upper} + \text{Lower}) \div 2$.) (ii) (b) Bivariate Distribution — the frequency distribution of two variables together. (iii) (d) The class midpoints — once data are grouped, the class mark represents the class and is used in further calculations. (iv) (a) The difference between the largest and the smallest observations.

Question 2

Can there be any advantage in classifying things? Explain with an example from your daily life.

Solution

ANSWER Yes. Classification arranges things into groups based on common characteristics, which saves time and effort, brings order, and makes it easy to locate, compare and draw inferences. It removes the confusion of dealing with a large, mixed mass of items. **Example:** When I arrange my school books subject-wise — History, Geography, Mathematics, Science — I can instantly pick up the

book I need instead of searching through the whole pile. Similarly, a grocery shop keeps pulses, spices and toiletries on separate shelves so any item can be found quickly. In the same way, the raw data of a census becomes meaningful only when classified by gender, age, education and occupation.

Question 3

What is a variable? Distinguish between a discrete and a continuous variable.

Solution

ANSWER A **variable** is a quantity or characteristic that can take different values — for example marks, height, income or the number of students. Its value “varies” from one observation to another. Variables are of two types: discrete and continuous. **Discrete variable:** takes only certain values; its value changes by finite “jumps” and it does not take intermediate values. For example, the number of students in a class can be 25 or 26 but not 25.5, because “half a student” is absurd. **Continuous variable:** can take *any* value within a range — whole numbers, fractions, and even irrational values. For example, a student’s height growing from 90 cm to 150 cm passes through every value in between (90.85 cm, 102.34 cm, etc.). Other examples are weight, time and distance.

Question 4

Explain the ‘exclusive’ and ‘inclusive’ methods used in classification of data.

Solution

ANSWER **Exclusive method:** here an item equal to a class limit (usually the *upper limit*) is excluded from that class and placed in the next class. The classes are written in a continuous form such as 0–10, 10–20, 20–30. A value of 10 goes into the class 10–20, not 0–10. This method is used very often for continuous variables because continuity is automatically maintained. **Inclusive method:** here both the lower and upper limits of a class are *included* in that class. The classes are written with a gap, such as 0–10, 11–20, 21–30. A value of 10 is included in the class 0–10 and 11 in the class 11–20. This method is commonly used for discrete variables. **Difference:** the inclusive method shows a “gap” between the upper limit of one class and the lower limit of the next (e.g. between 10 and 11). To make a continuous distribution from it, an *adjustment* is made — subtract 0.5 from each lower limit and add 0.5 to each upper limit (so 0–10 becomes –0.5–10.5, etc.), restoring continuity.

Question 5

Use the data in Table 3.2 that relate to monthly household expenditure (in Rs) on food of 50 households and

(i) Obtain the range of monthly household expenditure on food. (ii) Divide the range into appropriate number of class intervals and obtain the frequency distribution of expenditure. (iii) Find the number of households whose monthly expenditure on food is (a) less than Rs 2000 (b) more than Rs 3000 (c) between Rs 1500 and Rs 2500.

Solution

ANSWER (i) **Range.** From Table 3.2, the highest monthly expenditure is Rs 5090 and the lowest is Rs 1007.

Range = Largest – Smallest = $5090 - 1007 = \text{Rs } 4083$. (ii) **Frequency distribution.** Taking a convenient class interval of Rs 500 (exclusive method), the 50 observations are tallied as below.

(Number of classes = $4083 \div 500 \approx 9$ classes.)

Expenditure on food (Rs)	Tally	No. of households (Frequency)
1000–1500	//// //// //// ////	20
1500–2000	//// //// ///	13
2000–2500	//// /	6
2500–3000	////	5
3000–3500	//	2
3500–4000	/	1
4000–4500	//	2
4500–5000	—	0
5000–5500	/	1
Total		50

(iii) **Number of households (counted directly from the raw data of Table 3.2):** (a) **Less than Rs 2000:** $20 + 13 = 33$ households. (b) **More than Rs 3000:** households with expenditure above 3000 are 3473, 3676, 3222, 4439, 4248 and 5090 = 6 households. (c) **Between Rs 1500 and Rs 2500** (i.e. 1500 to under 2500): $13 + 6 = 19$ households.

Question 6

In a city 45 families were surveyed for the number of Cell phones they used. Prepare a frequency array based on their replies as recorded below.

1 3 2 2 2 1 2 1 2 2 3 3 3 3 3 3 2 2 6 1 6 2 1 5 1 5 3 2 4 2 7 4 2 4 3 4 2 0 3 1 4 3

Solution

ANSWER “Number of cell phones” is a **discrete variable**, so its data is classified as a **frequency array** — each integral value gets its own frequency.

Number of cell phones	Tally	Number of families (Frequency)
0	/	1
1	//// //	7
2	//// //// ////	15
3	//// //// //	12
4	////	5
5	//	2
6	//	2
7	/	1
Total		45

The frequencies add up to $1 + 7 + 15 + 12 + 5 + 2 + 2 + 1 = 45$, confirming all families are accounted for. Most families (15) used 2 cell phones.

Question 7

What is ‘loss of information’ in classified data?

Solution

ANSWER ‘Loss of information’ is the inherent shortcoming of classifying raw data into a frequency distribution. When individual observations are grouped into classes, their actual values are no longer used — every value in a class is assumed to be equal to the **class mark** (mid-value), and all further statistical calculations are based on the class mark, not on the original observations. For example, in the class 20–30 the six observations 25, 25, 20, 22, 25 and 28 are all treated as 25. The distribution then tells us only that the class has 6 values, not what those values actually were. This is the loss of information. However, the gain in making the raw data concise, ordered and comprehensible more than makes up for it.

Question 8

Do you agree that classified data is better than raw data? Why?

Solution

ANSWER Yes, classified data is generally better than raw data for the purpose of analysis. Raw data are disorganised, large and cumbersome; it is a tedious task to draw any meaningful conclusion from them because they do not yield easily to statistical methods. Classification places facts of similar characteristics in the same class, which: (i) brings order to the data; (ii) makes it concise and easy to understand; (iii) enables easy comparison; (iv) helps locate any item quickly; and (v) makes it ready for further statistical analysis (averages, dispersion, etc.). Although some 'loss of information' occurs because individual values are replaced by class marks, the benefit of comprehensibility and analysability far outweighs this drawback. Hence, classified data is more useful.

Question 9

Distinguish between univariate and bivariate frequency distribution.

Solution

ANSWER **Univariate frequency distribution:** the frequency distribution of a *single* variable. It shows how the values of one variable are distributed across classes with their frequencies — for example, the marks of 100 students in mathematics classified into the classes 0–10, 10–20, ..., 90–100. **Bivariate frequency distribution:** the frequency distribution of *two* variables together, shown in a two-way table where one variable is classed in rows and the other in columns. Each cell shows the joint frequency — for example, the sales and advertisement expenditure of 20 firms, where a cell shows that 3 firms had sales of Rs 135–145 lakh and advertisement expenditure of Rs 64–66 thousand.

Question 10

Prepare a frequency distribution by inclusive method taking class interval of 7 from the following data.

28 17 15 22 29 21 23 27 18 12 7 2 9 4 1 8 3 10 5 20 16 12 8 4 33 27 21 15 3 36 27 18 9 2 4 6 32 31 29 18 14 13 15 11 9 7 1 5 37 32 28 26 24 20 19 25 19 20 6 9

Solution

ANSWER There are 60 observations. The smallest value is 1 and the largest is 37, so Range = $37 - 1 = 36$. With class interval 7 by the **inclusive method**, we start from 1: classes 1–7, 8–14, 15–21, 22–28, 29–35, 36–42 (both limits included in each class).

Class interval	Tally	Frequency
1–7	//// //// ////	15
8–14	//// //// //	12
15–21	//// //// ////	15
22–28	//// ////	10
29–35	//// /	6
36–42	//	2
Total		60

The frequencies total $15 + 12 + 15 + 10 + 6 + 2 = 60$, confirming every observation is included.

Question 11

“The quick brown fox jumps over the lazy dog”

Examine the above sentence carefully and note the numbers of letters in each word. Treating the number of letters as a variable, prepare a frequency array for this data.

Solution

ANSWER Counting the letters in each of the 9 words: The (3), quick (5), brown (5), fox (3), jumps (5), over (4), the (3), lazy (4), dog (3). The number of letters takes only whole values, so it is a **discrete variable** and is classified as a frequency array.

Number of letters in a word	Words	Frequency (number of words)
3	The, fox, the, dog	4
4	over, lazy	2
5	quick, brown, jumps	3
Total		9

The frequencies add up to $4 + 2 + 3 = 9$, which equals the number of words in the sentence.