Chapter 3: Organisation of Data

Exercises — Organisation of Data

Q1 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

(i) Answer: (a)

The class midpoint (mid-value) = (Upper class limit + Lower class limit) / 2, which is their average.

(ii) Answer: (b)

A frequency distribution of two variables simultaneously is called a Bivariate Distribution.

(iii) Answer: (d)

In classified (grouped) data, the actual values of individual observations are lost. Statistical calculations use the class midpoint as the representative value of each class interval.

(iv) Answer: (a)

Range = Largest observation - Smallest observation. Option (b) subtracts in the wrong order (smallest minus largest would give a negative value).

Answer: (i) a, (ii) b, (iii) d, (iv) a

Q2 Can there be any advantage in classifying things? Explain with an example from your daily life.

SOLUTION

Yes, classifying things makes them easier to locate, compare, and analyse.

Example: A library arranges books by subject (Fiction, Science, History, etc.). Without classification, finding a specific book among thousands would be tedious. With classification, a reader goes directly to the relevant section.

Advantages of classification:
1. Simplifies complex data — reduces a large mass of figures into a compact, understandable form.
2. Facilitates comparison — similar items are grouped together, making differences and similarities visible.
3. Highlights key features — patterns and characteristics that are not apparent in raw data become evident.
4. Saves time — locating information is faster when data is organised into categories.

Answer: Classification simplifies data, enables comparison, and highlights patterns. Example: books in a library arranged by subject allow readers to locate any book quickly.

Q3 What is a variable? Distinguish between a discrete and a continuous variable.

SOLUTION

A variable is a characteristic or quantity that can take different values for different individuals or at different times. Examples: marks scored, height, number of children.

Discrete Variable:
- Takes only specific, countable values (usually whole numbers).
- No intermediate value is possible between two consecutive values.
- Example: Number of students in a class (can be 30 or 31, not 30.5); number of children in a family.

Continuous Variable:
- Can take any value within a given range, including fractions and decimals.
- Intermediate values are always possible.
- Example: Height (a person can be 165.3 cm, 165.35 cm, etc.); temperature; weight.

Answer: A variable is a measurable characteristic that varies. Discrete variables take only countable (integer) values (e.g., number of siblings); continuous variables can take any value in a range (e.g., height, weight).

Q4 Explain the 'exclusive' and 'inclusive' methods used in classification of data.

SOLUTION

Exclusive Method:
- The upper class limit of one class equals the lower class limit of the next class.
- An observation equal to the upper limit of a class is excluded from that class and included in the next class.
- Example: Classes 10-20, 20-30, 30-40. An observation of 20 goes into the 20-30 class, not 10-20.
- Suitable for continuous variables; class boundaries coincide with class limits.

Inclusive Method:
- Both lower and upper class limits are included within the same class interval.
- An observation equal to the upper limit of a class is included in that class itself.
- Example: Classes 10-19, 20-29, 30-39. An observation of 19 belongs to 10-19; 20 belongs to 20-29.
- Suitable for discrete variables where gaps between classes are meaningful.

Answer: Exclusive method: upper limit of a class = lower limit of next class; the upper limit value belongs to the higher class. Inclusive method: both limits are included in the same class; a gap exists between consecutive classes.

Q5 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

(i) Range:
Minimum expenditure = Rs 1007
Maximum expenditure = Rs 5090
Range = 5090 - 1007 = Rs 4083

(ii) Frequency Distribution (Exclusive Method, class width = Rs 1000):
Number of classes = 5 (covers range 1000-6000)

Class Interval (Rs)	Frequency
1000-2000	33
2000-3000	11

Class Interval (Rs)	Frequency
3000-4000	3
4000-5000	2
5000-6000	1
Total	50

(iii) Answers:
(a) Households with expenditure less than Rs 2000 = 33 (all in the 1000-2000 class)
(b) Households with expenditure more than Rs 3000 = 3 + 2 + 1 = 6
(c) Households with expenditure between Rs 1500 and Rs 2500 = 19

Answer: (i) Range = Rs 4083. (ii) 5 classes of width Rs 1000 (1000-2000: 33, 2000-3000: 11, 3000-4000: 3, 4000-5000: 2, 5000-6000: 1). (iii) (a) 33 households, (b) 6 households, (c) 19 households.

Q6 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 3 3 4 2 2 2 2 3 7 2 2 4 2 2 2 1 6 4 2 1 3 1 6 4 2 2 2 2 1 0 3 5 3 3 1 1 3 5 4 3 3 3

SOLUTION

Arranging the data as a frequency array (discrete variable — number of cell phones):

Number of Cell Phones	Tally	Frequency
0	|	1
1	||||| ||	7
2	||||| ||||| |||||	15
3	||||| ||||| ||	12
4	|||||	5
5	||	2
6	||	2
7	|	1
Total		45

Verification: 1+7+15+12+5+2+2+1 = 45 ✓

Answer: Frequency array: 0→1, 1→7, 2→15, 3→12, 4→5, 5→2, 6→2, 7→1. Total = 45 families. Most families (15) use 2 cell phones.

Q7 What is 'loss of information' in classified data?

SOLUTION

When raw data is classified into frequency distributions, the exact value of each individual observation is replaced by the class midpoint for statistical calculations.

For example, if the class interval is 10-20 and it has 5 observations, we do not know the actual values (say 11, 13, 15, 17, 19). We only know that 5 observations lie in 10-20. All calculations treat each observation as if it equals the midpoint (15).

This replacement of actual values by the midpoint constitutes the 'loss of information'. The original data cannot be recovered from the frequency distribution.

Consequence: Calculations (mean, median, etc.) from grouped data are approximations, not exact values.

Answer: Loss of information refers to the fact that in classified data, the exact value of each observation is unknown — only the class in which it falls is known. All calculations use the class midpoint as an approximation, making results estimates rather than exact values.

Q8 Do you agree that classified data is better than raw data? Why?

SOLUTION

Yes, classified data is generally better than raw data for the following reasons:

1. Comprehensibility: Raw data is a disorganised mass of figures. Classification condenses it into a compact, readable form that is easy to understand at a glance.

2. Comparison: Classified data allows comparison across classes, regions, or time periods. Raw data makes such comparison very difficult.

3. Pattern detection: Trends, concentrations, and distributional patterns (e.g., most households spend Rs 1000-2000) are visible only after classification.

4. Statistical analysis: Measures like mean, median, and mode can be computed from classified data efficiently, even for large datasets.

However, classified data has one drawback — loss of information (the exact values of individual observations are replaced by class midpoints). Despite this, the advantages far outweigh this limitation for large datasets.

Answer: Yes. Classified data simplifies complex raw data, enables comparison, reveals patterns, and facilitates statistical analysis. The only drawback is loss of information (exact values replaced by midpoints), but this is acceptable for large datasets.

Q9 Distinguish between univariate and bivariate frequency distribution.

SOLUTION

Univariate Frequency Distribution:
- Records the frequency of only one variable.
- Shows how often each value (or class) of a single variable occurs.
- Example: Distribution of 50 households by monthly food expenditure (only expenditure is studied).

Bivariate Frequency Distribution:
- Records the joint frequency of two variables simultaneously.
- Shows how often combinations of values of two variables occur together.
- Presented as a two-way table (rows = one variable, columns = the other).
- Example: Distribution of students by marks in Maths AND marks in Science — shows the relationship between performance in the two subjects.

Key difference: Univariate studies ONE variable; bivariate studies TWO variables and their relationship.

Answer: Univariate frequency distribution shows the frequency of one variable (e.g., height of students). Bivariate frequency distribution shows the joint frequency of two variables simultaneously (e.g., height and weight), revealing the relationship between them.

Q10 Prepare a frequency distribution by inclusive method taking class interval of 7 from the following data.
28 1 3 15 19 17 8 36 11 20 15 3 27 9 6 22 10 18 7 9 29 5 9 1 21 20 2 5 23 16 4 37 27 12 6 32 18 8 32 28 12 4 31 26 7 33 29 24 2 27 18 20 9 21 14 19 4 15 13 25

SOLUTION

Data: 60 observations. Min = 1, Max = 37. Range = 37 - 1 = 36.
Class interval = 7 (inclusive method).
Number of classes = 36 / 7 ≈ 5.14 → 6 classes.
Starting value = 1 (minimum). Classes: 1-7, 8-14, 15-21, 22-28, 29-35, 36-42.

Frequency Distribution (Inclusive Method):

Class Interval	Frequency
1-7	15
8-14	12
15-21	15
22-28	10
29-35	6

Class Interval	Frequency
36-42	2
Total	60

Verification: 15+12+15+10+6+2 = 60 ✓

Note: In the inclusive method, both the lower and upper limits are included in the same class. For example, an observation of 7 belongs to class 1-7, and an observation of 8 belongs to class 8-14.

Answer: Frequency distribution (inclusive, class width 7): 1-7→15, 8-14→12, 15-21→15, 22-28→10, 29-35→6, 36-42→2. Total = 60.

Q11 "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

Counting letters in each word:

Word	Number of Letters
The	3
quick	5
brown	5
fox	3
jumps	5
over	4
the	3
lazy	4
dog	3

Total words = 9.

Frequency Array:

Number of Letters (Variable)	Frequency
3	4

Number of Letters (Variable)	Frequency
4	2
5	3
Total	9

Verification: 4+2+3 = 9 words ✓

Answer: Frequency array: 3 letters→4 words (The, fox, the, dog), 4 letters→2 words (over, lazy), 5 letters→3 words (quick, brown, jumps). Total = 9 words.