Chapter 2: Collection of Data

Exercises — Collection of Data

Q1 Frame at least four appropriate multiple-choice options for the following questions: (i) Which of the following is the most important when you buy a new dress? (ii) How often do you use computers? (iii) Which of the newspapers do you read regularly? (iv) Rise in the price of petrol is justified. (v) What is the monthly income of your family?

SOLUTION

(i) Which of the following is the most important when you buy a new dress?
(a) Colour (b) Price (c) Brand (d) Fabric/Material

(ii) How often do you use computers?
(a) Daily (b) 2–3 times a week (c) Once a week (d) Rarely or never

(iii) Which of the newspapers do you read regularly?
(a) The Hindu (b) Hindustan Times (c) The Times of India (d) Any regional newspaper

(iv) Rise in the price of petrol is justified.
(a) Strongly agree (b) Agree (c) Disagree (d) Strongly disagree

(v) What is the monthly income of your family?
(a) Less than Rs 10,000 (b) Rs 10,000–25,000 (c) Rs 25,000–50,000 (d) More than Rs 50,000

Multiple-choice (closed-ended) questions like these give respondents a fixed set of options, making them easy to answer, score, and codify for analysis. The 'Any Other' option can be added where an individual's true response may not be among the listed alternatives.

Answer: Sample multiple-choice options framed for each of the five questions above, covering relevant categories and providing at least four options per question.

Q2 Frame five two-way questions (with 'Yes' or 'No').

SOLUTION

Two-way questions are closed-ended questions with only two possible answers: 'Yes' or 'No'. Five sample two-way questions are:

(i) Do you own a computer at home? Yes / No
(ii) Do you read a newspaper daily? Yes / No
(iii) Do you use a mobile phone? Yes / No
(iv) Do you think the price of essential commodities has risen in the last year? Yes / No
(v) Do you use public transport to travel to school or work? Yes / No

Two-way questions are best used when the answer can clearly be either yes or no. They are simple to use, easy to score, and quick for respondents to answer. However, they may oversimplify issues where respondents would prefer a more nuanced answer.

Answer: Five sample two-way (Yes/No) questions on everyday topics such as computer ownership, newspaper reading, mobile phone use, price rise perception, and use of public transport.

Q3 State whether the following statements are True or False. (i) There are many sources of data. (ii) Telephone survey is the most suitable method of collecting data, when the population is literate and spread over a large area. (iii) Data collected by investigator is called the secondary data. (iv) There is a certain bias involved in the non-random selection of samples. (v) Non-sampling errors can be minimised by taking large samples.

SOLUTION

(i) True. Statistical data can be obtained from two broad sources — primary sources (collected by the researcher directly through surveys) and secondary sources (published government reports, documents, newspapers, books, websites, etc.). Within these, many individual sources exist at national and state levels.

(ii) False. Mailing (questionnaire) survey is more suitable when the population is literate and spread over a large area, because it can reach remote areas at low cost. Telephone survey is cheaper than personal interview but is limited by the fact that many people may not own telephones, and access to the entire population cannot be guaranteed.

(iii) False. Data collected by the investigator through direct enquiry or surveys is called primary data. Secondary data is data that has already been collected and processed by some other agency and is later used by another researcher.

(iv) True. In non-random sampling, units are selected on the basis of the investigator's judgement, convenience, purpose, or quota. This means not every member of the population has an equal chance of being selected, introducing bias into the sample.

(v) False. Non-sampling errors are more serious than sampling errors precisely because they cannot be minimised by taking larger samples. Non-sampling errors — such as errors in data acquisition, non-response errors, and sampling bias — can occur even in a Census. Sampling errors, on the other hand, can be reduced by increasing sample size.

Answer: (i) True (ii) False (iii) False (iv) True (v) False

Q4 What do you think about the following questions? Do you find any problem with these questions? Describe. (i) How far do you live from the closest market? (ii) If plastic bags are only 5 per cent of our garbage, should it be banned? (iii) Wouldn't you be opposed to increase in price of petrol? (iv) Do you agree with the use of chemical fertilisers? (v) Do you use fertilisers in your fields? (vi) What is the yield per hectare in your field?

SOLUTION

(i) 'How far do you live from the closest market?' — This question is ambiguous. Different respondents may interpret 'how far' differently: some may answer in kilometres, others in minutes of travel, and others by number of stops. The question should specify the unit of measurement, e.g., 'How far (in kilometres) do you live from the closest market?'

(ii) 'If plastic bags are only 5 per cent of our garbage, should it be banned?' — This is a leading question. By stating the fact that plastic bags are 'only 5 per cent of garbage', it hints to the respondent that banning them may not be necessary. A good question should not give a clue about the expected answer. The question should be reframed without the embedded premise, e.g., 'Do you think plastic bags should be banned?'

(iii) 'Wouldn't you be opposed to increase in price of petrol?' — This is a question with a double negative. Questions starting with 'Wouldn't you' or 'Don't you' are likely to lead to biased responses because they suggest the expected answer. It should be reframed as a direct positive question: 'Are you opposed to an increase in the price of petrol?'

(iv) 'Do you agree with the use of chemical fertilisers?' — This question is vague and ambiguous. It does not specify the context (crop type, scale of farming, region) and the word 'agree' is subjective. Responses would be difficult to compare or analyse. A better version would separate the issue into more specific sub-questions.

(v) 'Do you use fertilisers in your fields?' — This is a good, clear two-way question with a simple Yes/No answer. However, it could be improved by specifying the type of fertiliser (chemical or organic) to get more precise data.

(vi) 'What is the yield per hectare in your field?' — This question may be difficult for many farmers who may not know the exact yield per hectare and may not be familiar with the unit 'per hectare'. It could lead to inaccurate responses. A better question might ask for the total yield from the field and the size of the field separately, then calculate yield per hectare from the data.

Answer: Questions (ii), (iii) are leading/biased questions; (i) and (vi) are ambiguous; (iv) is vague; (v) is broadly acceptable but could be more specific.

Q5 You want to do a research on the popularity of Vegetable Atta Noodles among children. Design a suitable questionnaire for collecting this information.

SOLUTION

A suitable questionnaire for collecting information on the popularity of Vegetable Atta Noodles among children could include the following questions. The questionnaire should be easy to understand, with clear and precise questions, and should not be too long.

Name: _____ Age: _____ Class: _____

Q1. Have you ever eaten Vegetable Atta Noodles? (a) Yes (b) No

Q2. How often do you eat Vegetable Atta Noodles?
(a) Daily (b) 2–3 times a week (c) Once a week (d) Rarely (e) Never

Q3. Why do you prefer Vegetable Atta Noodles? (You may choose more than one)
(a) Taste (b) Health benefits (c) Easy to prepare (d) Affordable price (e) Any other: _____

Q4. How do you rate the taste of Vegetable Atta Noodles?
(a) Excellent (b) Good (c) Average (d) Poor

Q5. Would you recommend Vegetable Atta Noodles to your friends? (a) Yes (b) No

The questionnaire moves from general to specific questions, uses both two-way and multiple-choice formats, avoids leading or ambiguous questions, and keeps the total number of questions short and child-friendly.

Answer: A sample questionnaire with five questions covering consumption frequency, reasons for preference, taste rating, and recommendation likelihood.

Q6 In a village of 200 farms, a study was conducted to find the cropping pattern. Out of the 50 farms surveyed, 50% grew only wheat. What is the population and the sample size?

SOLUTION

Population refers to the totality of all items under study. In this case, the study is about the cropping pattern of farms in a particular village.

Population: All 200 farms in the village.

Sample: The 50 farms that were actually surveyed. The sample is a smaller representative group selected from the population to obtain information.

The sample size is 50, which is 25 per cent of the total population of 200 farms. From this sample, we observe that 50% of the 50 surveyed farms (i.e., 25 farms) grew only wheat. This sample finding is used to draw conclusions about the cropping pattern of all 200 farms in the village.

Answer: Population = 200 farms (all farms in the village); Sample size = 50 farms (the farms actually surveyed).

Q7 Give two examples each of sample, population and variable.

SOLUTION

Population (or Universe) refers to the totality of all items under study. A sample is a smaller representative group selected from the population. A variable is a characteristic that varies from one item to another and is represented by values.

Examples of Population:
(i) All students studying in Class XI in India — if we want to study the study habits of Class XI students across the country.
(ii) All agricultural labourers in a district — if we want to study the economic condition of agricultural labourers in that district.

Examples of Sample:
(i) 100 students selected from all Class XI students in India to study their study habits.
(ii) 10 per cent of agricultural labourers selected from a district to estimate the economic condition of all labourers in the district.

Examples of Variable:
(i) The income of farmers — it varies from farmer to farmer and from year to year, so income is a variable.
(ii) The production of food grains in India — it varies from year to year and from crop to crop, making it a variable.

Answer: Population: (i) all Class XI students in India, (ii) all agricultural labourers in a district. Sample: (i) 100 selected Class XI students, (ii) 10% of labourers in a district. Variable: (i) income of farmers, (ii) food grain production across years.

Q8 Which of the following methods give better results and why? (a) Census (b) Sample

SOLUTION

Both Census and Sample methods have their own merits, but in most practical situations, the Sample method gives better results. The reasons are as follows.

A sample can provide reasonably reliable and accurate information at a much lower cost and in a shorter time compared to a Census. Since samples are smaller than the population, more detailed and intensive enquiries can be conducted on the selected units. It is also easier to train and supervise a smaller team of enumerators, which improves the quality of data collected.

In a Census, every element of the population must be covered, which requires a large number of enumerators, substantial expenditure, and considerable time. Despite this, a Census can still contain non-sampling errors, which are difficult to minimise.

However, Census is preferred when complete and continuous demographic records are needed, as in the Census of India, which collects data on population size, density, sex ratio, literacy, migration, etc. For such purposes, Census is indispensable.

In general, for most research and statistical purposes, the Sample method is preferred because

it is cost-effective, time-saving, and can yield sufficiently accurate results when a good representative sample is chosen.

Answer: Sample method generally gives better results because it is less expensive, faster, allows intensive enquiry, and enables effective supervision of enumerators, while still providing reasonably accurate estimates of the population.

Q9 Which of the following errors is more serious and why? (a) Sampling error (b) Non-Sampling error

SOLUTION

Non-sampling error is more serious than sampling error.

Sampling error arises due to the difference between the sample estimate and the actual value of the corresponding population parameter. For example, if the true average income of farmers is Rs 600 but the sample estimate gives Rs 550, the sampling error is Rs 50. The magnitude of sampling error can be reduced by taking a larger sample. Thus, sampling error is controllable.

Non-sampling errors, on the other hand, are more serious because they cannot be minimised simply by taking a larger sample. Non-sampling errors can arise from multiple sources: errors in data acquisition (e.g., recording 13 instead of 31), non-response errors (when a person refuses to respond or cannot be contacted), and sampling bias (when some members of the population have no chance of being included in the sample).

Crucially, non-sampling errors can occur even in a Census, where every unit of the population is covered. This makes them particularly dangerous because no increase in sample size can eliminate them — only careful survey design, proper training of enumerators, and thorough verification of data can help reduce non-sampling errors.

Answer: Non-sampling error is more serious because it cannot be reduced by increasing sample size, can occur even in a Census, and arises from data acquisition mistakes, non-response, and bias — whereas sampling error can be minimised by taking a larger sample.

Q10 Suppose there are 10 students in your class. You want to select three out of them. How many samples are possible?

SOLUTION

The number of possible samples when selecting 3 students out of 10 is calculated using the combination formula:

Number of samples = 10C3 = 10! / (3! × 7!) = (10 × 9 × 8) / (3 × 2 × 1) = 720 / 6 = 120

So, 120 different samples of 3 students are possible from a class of 10 students.

In random sampling, each of these 120 possible samples has an equal chance of being selected, ensuring that every individual student also has an equal probability of being included in the chosen sample.

Answer: 120 samples are possible (10C3 = 120).

Q11 Discuss how you would use the lottery method to select 3 students out of 10 in your class.

SOLUTION

The lottery method is a simple and practical way to conduct random sampling, where every individual has an equal chance of being selected.

To select 3 students out of 10 using the lottery method, the following steps would be taken:

Step 1: Write the names (or assign numbers 1 to 10) of all 10 students on separate, identical slips of paper.

Step 2: Fold all the slips in the same way so that no slip can be identified from the outside.

Step 3: Put all the folded slips into a bowl or box and mix them thoroughly.

Step 4: Without looking, draw one slip at random from the bowl. Note the name or number on it — this is the first selected student.

Step 5: Put the slip back or set it aside (depending on whether sampling is with or without replacement). Mix again and draw a second slip — this is the second selected student.

Step 6: Repeat the process to draw the third slip — this is the third selected student.

The three students whose names were drawn form the random sample. This method is also called the lottery method. In practice, computer programmes are now used to select random samples in the same manner, but the lottery method remains valid, especially for small populations.

Answer: Assign numbers 1–10 to all students, write them on identical slips, mix in a bowl, and draw 3 slips one by one at random — the three drawn names form the random sample.

Q12 Does the lottery method always give you a random sample? Explain.

SOLUTION

The lottery method gives a random sample only if it is carried out properly and without any bias. It does not always guarantee a truly random sample if the process is not followed carefully.

For the lottery method to produce a random sample, every individual must have an equal chance of being selected. This requires that all the slips be identical in size, shape, and feel, folded in exactly the same way, mixed thoroughly before each draw, and drawn without looking. If any of these conditions are not met — for example, if some slips are larger, not properly mixed, or can be identified by touch — then the selection is no longer truly random and some individuals may have a higher or lower chance of being selected.

In practice, for large populations, the lottery method becomes difficult to implement fairly because physically mixing thousands of slips thoroughly is challenging. In such cases, random number tables or computer-generated random numbers are used instead, as they provide a

more reliable way to select a random sample.

In summary, the lottery method gives a random sample when executed correctly with identical slips and thorough mixing. Any deviation from these conditions introduces bias and the sample may no longer be truly random.

Answer: Not always — the lottery method gives a random sample only if all slips are identical, mixed thoroughly, and drawn without bias. For large populations, it is impractical and random number tables are more reliable.

Q13 Explain the procedure for selecting a random sample of 3 students out of 10 in your class by using random number tables.

SOLUTION

Random number tables are specially constructed tables containing digits (0–9) arranged in a random order, used to select a random sample without bias.

The procedure for selecting 3 students out of 10 using random number tables is as follows:

Step 1: Assign a unique number from 1 to 10 to each of the 10 students. For example, student 1 gets number 01, student 2 gets number 02, and so on up to student 10 who gets number 10.

Step 2: Open the random number table at any page and start at any random point in the table (this starting point can itself be chosen randomly, such as by closing your eyes and placing a finger on the table).

Step 3: Read two-digit numbers consecutively from the table in any direction (across a row, down a column, or diagonally — the direction must be decided before starting and kept consistent).

Step 4: Select the number if it falls within the range 01 to 10. Skip any number outside this range (11–99, 00) as it does not correspond to any student.

Step 5: Continue reading until 3 valid, distinct numbers have been selected. If a number repeats, skip it and move to the next.

Step 6: The 3 students corresponding to the selected numbers form the random sample.

This method ensures that every student has an equal chance of being included in the sample, making it a valid random sample.

Answer: Number students 01–10, open the random number table at a random point, read two-digit numbers sequentially and select those in the range 01–10, skipping repeats and out-of-range numbers, until 3 students are chosen.

Q14 Do samples provide better results than surveys? Give reasons for your answer.

SOLUTION

In most practical situations, samples (sample surveys) provide better results than complete surveys (Census), for several reasons.

First, a sample is less expensive. A Census requires covering every element of the population, which demands a large number of enumerators and substantial financial resources. A sample, by contrast, requires far fewer resources while still providing reasonably accurate information.

Second, a sample takes less time. A complete survey of the entire population is time-consuming. A sample survey can be completed quickly, which is important when timely decisions are needed.

Third, a sample allows for more detailed and intensive enquiry. Because the sample is smaller, each unit can be studied more thoroughly. It is also easier to train and supervise a smaller team of enumerators, improving the quality of data collected.

Fourth, in cases where testing is destructive — for example, testing the life of electric bulbs — a Census would destroy all items, which is clearly not feasible. A sample allows testing without destroying the entire population.

However, it must be noted that samples do not always give better results. The accuracy of sample results depends on how well the sample represents the population. A poorly chosen sample can produce highly misleading results. Also, for purposes that require complete enumeration — such as the Census of India which records demographic data for all citizens — a complete survey is necessary and a sample cannot substitute for it.

Overall, when conducted with proper sampling techniques and a representative sample, sample surveys generally provide better practical results than complete surveys for most research and statistical purposes.

Answer: Yes, in most situations samples provide better results than complete surveys because they are less expensive, faster, allow intensive enquiry, and enable better supervision — though the quality depends on how representative the sample is.