

Statistics

Complete Notes, Chapter Wise

FULL Question Answers

Original Study / Revision Notes Only

6 Chapters | 30 Short Questions | 12 Long Questions | 30 MCQs

Chapters 1 and 2 follow the student's contents page in detail. Chapters 3 to 6 follow the standard syllabus in general terms. All worked numbers are illustrative examples only. No board is claimed. Not official textbook notes, not board papers, not leaked material, not a guarantee of any question or marks.

Chapter 1: Introduction

Explanation

Statistics is the study of collecting, organising, presenting, analysing and interpreting numerical data so that sensible conclusions can be drawn from it. In early times, rulers counted people, land, crops and soldiers so that they could plan taxes, food supply and defence. Over time, counting grew into a proper method for comparing groups, finding averages, noticing patterns and making decisions. Today statistics is used in almost every field, from business, health and education to government and science, because decisions are safer when they rest on carefully collected facts rather than on guesses.

The word statistics is used in two senses. In the plural sense, it means numerical facts themselves, such as the marks of students or the prices of goods. In the singular sense, it means the science and the methods used to deal with those facts, including collection, classification, presentation, analysis and interpretation. Descriptive statistics describes data that we already have, using tables, graphs, averages and percentages. Inferential statistics goes one step further and uses a sample to draw conclusions about a larger population, so description comes first and inference follows from it.

Good statistical data should be collected for a clear purpose, stated in numbers, affected by many causes, reasonably accurate, comparable with other data and collected in a systematic way. Statistics also has clear limits. It deals mainly with numbers, so it cannot directly measure qualities such as honesty or kindness. It studies groups rather than single persons in detail, and its results are often averages that can hide differences inside a group. Wrong data, a wrong method or dishonest presentation can therefore give a wrong result, which is why checking the source and the method matters as much as doing the calculation correctly.

Data may be primary or secondary. Primary data is collected for the first time, directly by the investigator, for example by interview, questionnaire or observation. Secondary data already exists, having been collected by someone else, for example from books, reports or official records, and it must be checked for reliability, date and suitability before it is used again. A variable is a quantity that changes from person to person or from time to time, such as height or marks, while a constant stays fixed in a study. A discrete variable takes whole-number values only, such as the number of children, while a continuous variable can take any value in a range, such as height or weight. Qualitative data describes qualities, such as colour or opinion, while quantitative data is measured in numbers, such as age or income.

Accuracy means how close a measurement is to the true value, and it improves with careful instruments, trained observers and repeated measurements. Errors of measurement happen when a measured value differs from the true value. Compensatory errors partly or fully cancel each other out, for example when some over-estimates and some under-estimates in a large set balance each other so the total error becomes small. Biased errors push results in one direction only, for example a faulty scale that always adds weight, and they do not cancel out, so they must be corrected at the source. Significant digits are the meaningful digits in a number, scientific notation writes large or

small numbers as a number between 1 and 10 multiplied by a power of 10, rounding reduces digits to a required place, and summation, shown by the symbol Sigma, means adding all the values together.

Important Points

- Statistics began with counting for administration and grew into a method for decision-making.
- Plural sense: numerical facts. Singular sense: methods for studying those facts.
- Descriptive statistics summarises known data; inferential statistics generalises from a sample to a population.
- Primary data is collected first-hand; secondary data already exists and must be checked before reuse.
- Discrete variables are counted in whole numbers; continuous variables are measured and can take decimal values.
- Compensatory errors tend to cancel out; biased errors push results one way and must be corrected.
- Significant digits, scientific notation, rounding and summation keep numbers clear and comparable.

Short Questions with Full Answers

Q1. What is statistics in simple words?

Statistics is the study of collecting, organising, presenting, analysing and interpreting numerical data. Its purpose is to turn a long list of figures into a clear picture that helps us understand a situation. From that picture, sensible conclusions and decisions can be drawn.

Q2. Why did early rulers need statistics?

Early rulers counted people, land, crops and soldiers so that they could plan taxes, food supply and defence. Without those counts, a ruler could not know how much food was needed, how many soldiers could be raised, or how much revenue could be expected. Counting therefore gave the state the basic facts it needed to govern.

Q3. How are descriptive and inferential statistics different?

Descriptive statistics describes data that we already have, for example by finding the average marks of one class. Inferential statistics uses a sample to draw a conclusion about a larger population, for example by testing some students and estimating the average for all students in a city. Description summarises what is known, while inference goes beyond the data in hand.

Q4. What is the difference between primary and secondary data?

Primary data is collected for the first time, directly by the investigator, for the present purpose, for example by interview or questionnaire. Secondary data already exists, having been collected by someone else, for example from a book, a report or an official record. Primary data fits the purpose better but costs more time and money, while secondary data is cheaper and quicker but may be old or unsuitable.

Q5. Is the number of books discrete or continuous, and why?

The number of books is a discrete variable because books are counted in whole numbers. A person can own 3 books or 4 books, but not three and a half books in this sense. Weight, by contrast, is continuous, because it is measured and can take decimal values.

Long Questions with Full Answers

Q1. Explain the history and development of statistics, and why it matters today.

Statistics began as simple counting for practical needs. Ancient rulers needed to know how many people lived in their lands, how much land was cultivated and how many soldiers could be called upon, because taxes, food and defence all depended on those facts. As trade, industry and science grew, counting alone was no longer enough, and people needed better ways to compare groups, find averages, notice patterns and judge whether a difference was real or accidental. Statisticians therefore developed methods for sampling, averages, dispersion and probability, and those methods turned statistics from a record-keeping exercise into a tool for thinking. Today the subject matters even more, because businesses study sales and demand, doctors study disease rates and treatment results, governments plan services from population data, and researchers test ideas with experiments and surveys. In every case the same lesson applies: a decision made from carefully collected and honestly analysed data is far more reliable than a decision made from a guess, a single example or a feeling.

Q2. Explain errors of measurement, and distinguish compensatory errors from biased errors.

An error of measurement happens whenever a measured value differs from the true value, and such errors can come from the instrument, the observer or the method itself. A compensatory error is one that tends to be balanced by another error in the opposite direction. For example, if one observer over-estimates a little and another under-estimates a little, the two mistakes partly cancel each other in a large set of measurements, so the total error becomes small and the average stays close to the truth. A biased error is very different, because it pushes results in one direction only. A faulty scale that always adds two kilograms, a clock that always runs fast, or a questionnaire that leads people towards one answer will distort every reading in the same way, so the errors do not cancel out no matter how many measurements are taken. The practical conclusion is clear: compensatory errors can be reduced by taking more measurements and averaging them, but a biased error must be found and corrected at its source, by repairing the instrument, retraining the observer or rewriting the question, before any further calculation is trusted.

MCQs with Answer Key

1. Statistics in the plural sense means:

- (a) Methods only
- (b) Numerical facts
- (c) Graphs only
- (d) Probability only

Answer: (b)

2. Finding the average of a class is an example of:

- (a) Inferential statistics

- (b) Descriptive statistics
- (c) Sampling only
- (d) Estimation only

Answer: (b)

3. Drawing a conclusion about all students from 50 students is:

- (a) Descriptive statistics
- (b) Inferential statistics
- (c) Tabulation
- (d) Classification

Answer: (b)

4. A faulty scale that always adds weight causes:

- (a) Compensatory error
- (b) Biased error
- (c) No error
- (d) Sampling error

Answer: (b)

5. The number of students in a class is a:

- (a) Continuous variable
- (b) Discrete variable
- (c) Constant in every study
- (d) Qualitative variable

Answer: (b)

Chapter 2: Presentation of Data

Explanation

Raw data is usually long, unordered and hard to read, so it must be presented in a clear form before it can be understood. Classification means arranging raw data into groups or classes according to a common feature, so that it becomes easy to compare and to see patterns. Classification may be geographical, by place such as city or province; chronological, by time such as year or month; qualitative, by a quality such as gender or subject; or quantitative, by numbers such as marks or age groups. Without classification, a long list of numbers tells the reader very little, because totals, differences and patterns stay hidden inside the list.

Tabulation means presenting classified data in rows and columns in a table. A good table has a clear title and number, column headings, units, carefully entered figures, totals where they are needed, and footnotes and a source note where they are required. The steps in building a table are simple: give it a clear title, decide the rows and columns, write the headings and units, enter the figures carefully, add totals, and add notes. A table should be simple, complete and easy to read at a glance, because a table that confuses the reader has failed in its purpose even if its figures are correct.

A frequency distribution shows how many times each value or class occurs in the data. For grouped data, values are put into classes, for example marks 0-9, 10-19 and so on, and the number in each class is its frequency. Class limits are the lowest and highest values stated for a class, for example 10-19 has lower limit 10 and upper limit 19. Class boundaries are the true limits between classes, found by taking the midpoint between the upper limit of one class and the lower limit of the next. For classes 10-19 and 20-29, the boundary between them is 19.5, and histograms use boundaries so that their bars touch each other without gaps. A bar chart uses separate bars to compare categories, a histogram uses touching bars to show a frequency distribution for continuous data, and a frequency polygon joins the midpoints of histogram bars. These graphs help the reader see at once where most values lie.

Important Points

- Classification arranges raw data into groups by a common feature.
- Four types: geographical (place), chronological (time), qualitative (quality), quantitative (numbers).
- A good table needs a title, headings, units, careful figures, totals and a source note.
- Frequency is the number of times a value or class occurs.
- Class limits are stated values; class boundaries are true dividing points, for example 19.5 between 10-19 and 20-29.
- Bar charts compare categories; histograms show continuous frequency distributions with touching bars.

Short Questions with Full Answers

Q1. What is classification, and why is it needed?

Classification means arranging raw data into groups or classes according to a common feature. It is

needed because a long, unordered list of numbers is hard to read and hard to compare. Once data is grouped, totals, differences and patterns become clear at a glance.

Q2. Give an example of chronological and geographical classification.

Arranging the sales of a shop in 2022, 2023 and 2024 by year is chronological classification, because the grouping is by time. Arranging the same sales by city, for example Lahore, Karachi and Multan, is geographical classification, because the grouping is by place. Both methods make comparison easier, but they answer different questions.

Q3. What must a good statistical table contain?

A good table must have a clear title and number, so the reader knows what it shows. It must also have column headings and units, figures entered carefully, totals where they are needed, and footnotes and a source note where they are required. A table without a title or headings may contain correct figures and still confuse the reader.

Q4. What is a frequency distribution?

A frequency distribution shows how many times each value or class occurs in a set of data. For grouped data, values are put into classes such as marks 0-9 or 10-19, and the number of values in each class is called its frequency. This arrangement turns a long list into a short summary that shows where most values fall.

Q5. What is the difference between class limits and class boundaries?

Class limits are the lowest and highest values stated for a class, for example 10 and 19 for the class 10-19. Class boundaries are the true dividing points between classes, found by taking the midpoint between the upper limit of one class and the lower limit of the next. For classes 10-19 and 20-29, the boundary is 19.5, and histograms use boundaries so their bars touch without gaps.

Long Questions with Full Answers

Q1. Why is the presentation of data important, and how do classification, tabulation and graphs help?

Raw data is long and confusing, and a reader who is given hundreds of unordered figures cannot easily see what they mean. Classification helps first, because it arranges those figures into groups by a common feature, so that a reader can compare one group with another instead of struggling with single values. Tabulation helps next, because it places the classified figures in rows and columns with a title, headings, units and totals, so that the whole picture can be read at a glance and checked for mistakes. Graphs help last, because a bar chart, a histogram or a frequency polygon turns numbers into a shape that the eye can grasp at once, showing where most values lie and where the data is thin. Together, these three steps do the same job in different ways: they make data short, clear and comparable, they expose patterns and mistakes that a raw list would hide, and they allow a reader to reach a conclusion without having to repeat the whole calculation. Presentation is therefore not decoration added after the work is done; it is part of the statistical work itself, because data that cannot be read clearly cannot be used wisely.

Q2. Explain class limits and class boundaries with an illustrative example, and say why histograms use boundaries.

Class limits are the values actually stated for a class, while class boundaries are the true dividing points between one class and the next. Take the classes 20-29 and 30-39 as an illustrative example. The stated limits are 20 and 29 for the first class and 30 and 39 for the second, but there is a gap between 29 and 30 if the limits alone are used. The boundary between the two classes is found by taking the midpoint of that gap, which is 29 plus 30 divided by 2, giving 29.5. In the same way, the lower boundary of the class 20-29 is 19.5 and its upper boundary is 29.5, so the class really covers the continuous range from 19.5 up to but not including 29.5. Histograms use boundaries rather than limits because a histogram is meant to show a continuous distribution, and its bars must touch each other without gaps. If limits were used, the bars would stand apart and suggest that no values could fall in the gaps, which would be false for continuous data such as height or weight. Boundaries therefore give the honest picture: the data flows continuously from one class into the next, and the height of each bar shows how many values fall in that range.

MCQs with Answer Key

1. Arranging data by year is:

- (a) Geographical classification
- (b) Chronological classification
- (c) Qualitative classification
- (d) Quantitative classification

Answer: (b)

2. A good table must have:

- (a) Only numbers
- (b) A title and headings
- (c) No totals
- (d) No source note

Answer: (b)

3. For classes 10-19 and 20-29, the class boundary between them is:

- (a) 19
- (b) 20
- (c) 19.5
- (d) 29.5

Answer: (c)

4. A histogram is used for:

- (a) Qualitative data only
- (b) A frequency distribution of continuous data
- (c) Pie charts only
- (d) A raw list of numbers

Answer: (b)

5. Frequency means:

- (a) The total of the values
- (b) The number of occurrences

(c) The average

(d) The range

Answer: (b)

Chapter 3: Measures of Central Tendency

Explanation

Note: this chapter follows the standard Statistics syllabus in general terms, as stated in the introduction. It is not a word-for-word copy of any textbook chapter.

An average, or a measure of central tendency, gives a single value that represents a whole set of data. The mean is the ordinary average: add all the values and divide by their number. The median is the middle value when the data is arranged in order, and if there are two middle values, their average is taken as the median. The mode is the value that occurs most often in the data. As an illustrative example, the mean of 4, 6 and 10 is 20 divided by 3, which is about 6.67; the median of 3, 5, 7, 9 and 11 is 7, because it sits in the middle; and the mode of 2, 3, 3 and 5 is 3, because it occurs more often than any other value.

Each measure has its own strength. The mean uses every value, so it is sensitive to the whole set, but that sensitivity is also its weakness, because one very large or very small value can pull the mean away from the centre. The median is less affected by extreme values, so it is often better for incomes or house prices, where a few very high values would distort the mean. The mode is the only measure that works well for qualitative data, such as the most popular colour or the most common answer in a survey, and it is useful when we want to know what is typical in the sense of most frequent. A data set can have no mode if no value repeats, and it can have two modes if two values tie for the highest frequency.

Important Points

- Mean: add all values and divide by their number; it uses every value.
- Median: the middle value in ordered data; it resists extreme values.
- Mode: the most frequent value; it works for qualitative data too.
- A set can have no mode, one mode, or two modes if two values tie.
- Extreme values pull the mean but have little effect on the median.
- Choose the measure that fits the data type and the question being asked.

Short Questions with Full Answers

Q1. How is the mean calculated? Give an illustrative example.

The mean is calculated by adding all the values in a set and dividing the total by the number of values. For example, the mean of 5, 10 and 15 is 30 divided by 3, which is 10. The mean therefore uses every value in the set, which is both its strength and its weakness.

Q2. How is the median found?

The median is found by arranging the data in order, from smallest to largest, and taking the middle value. For example, the median of 2, 4, 6, 8 and 10 is 6, because two values lie below it and two lie above it. If there are two middle values, their average is taken as the median.

Q3. What is the mode, and can a set have no mode?

The mode is the value that occurs most often in a set of data. Yes, a set can have no mode, and

this happens when no value repeats, for example in the set 1, 2, 3, 4, where every value occurs exactly once. A set can also have two modes if two values tie for the highest frequency.

Q4. When is the median better than the mean?

The median is better than the mean when the data contains very large or very small extreme values. Such values pull the mean towards themselves and away from the centre of the data, while the median stays in the middle. For incomes, where a few very high incomes would distort the average, the median therefore gives a fairer picture of a typical person.

Q5. When is the mode the most useful measure?

The mode is most useful when we want to know what occurs most often, especially for qualitative data that cannot be averaged. For example, a shopkeeper wants to know the most popular shoe size, and a survey researcher wants to know the most common answer to a question. In both cases the mean and the median would be meaningless, while the mode answers the question directly.

Long Questions with Full Answers

Q1. Compare the mean, the median and the mode, with an illustrative example of an extreme value.

The mean, the median and the mode all try to describe the centre of a set of data, but they do so in different ways and they behave differently when the data contains an extreme value. The mean adds all the values and divides by their number, so it uses every value and gives a precise balance point, but it is pulled by extremes. The median is simply the middle value in ordered data, so it ignores how far the extremes lie from the centre and stays stable. The mode is the most frequent value, so it tells us what is common, but it may ignore most of the data altogether. Consider, as an illustrative example, the incomes 10, 12, 13, 14 and 101, measured in thousands. The mean is 150 divided by 5, which is 30, a figure higher than four of the five incomes and therefore misleading as a picture of a typical person. The median is 13, which sits in the middle and describes the typical income far better. The conclusion is practical: use the mean when the data is fairly even and every value should count, use the median when extremes are present, and use the mode when the question is about what occurs most often.

Q2. How should a student choose between the mean, the median and the mode in a real problem?

Choosing a measure of central tendency begins with the type of data and the question being asked, not with habit. If the data is numerical and fairly even, with no extreme values, the mean is a good choice, because it uses every value and is the basis for further statistical work such as variance and standard deviation. If the data contains extremes, or if it is open-ended, the median is safer, because it stays in the middle and is not dragged by a few very large or very small values, as with incomes or house prices. If the data is qualitative, or if the question is simply about popularity, the mode is the only sensible choice, because you cannot average colours or answers, but you can count which one occurs most often. A careful student therefore states the measure, explains why it was chosen, and checks the result against common sense. An average that is higher than almost every value in the set, or a mode that occurs only twice in a hundred values, should both be

questioned, because a measure is only useful when it honestly represents the data behind it.

MCQs with Answer Key

1. The mean of 5, 10 and 15 is:

- (a) 5
- (b) 10
- (c) 15
- (d) 30

Answer: (b)

2. The median of 2, 4, 6, 8 and 10 is:

- (a) 4
- (b) 6
- (c) 8
- (d) 10

Answer: (b)

3. The mode is:

- (a) The middle value
- (b) The most frequent value
- (c) The average
- (d) The range

Answer: (b)

4. Which measure is least affected by an extreme value?

- (a) Mean
- (b) Median
- (c) Range
- (d) Total

Answer: (b)

5. A set in which no value repeats has:

- (a) Two modes
- (b) One mode
- (c) No mode
- (d) A mode of zero

Answer: (c)

Chapter 4: Measures of Dispersion

Explanation

Note: this chapter follows the standard Statistics syllabus in general terms, as stated in the introduction. It is not a word-for-word copy of any textbook chapter.

An average alone does not tell the whole story, because two sets of data can have the same average and yet be very different. Measures of dispersion describe how spread out the values are around their average. The range is the simplest measure: it is the difference between the largest and the smallest values, so for the illustrative values 2, 4 and 6, the range is 6 minus 2, which is 4. The range is easy to calculate, but it uses only two values and is therefore strongly affected by extremes. Variance measures the average squared distance of the values from their mean, and standard deviation is the square root of variance. Standard deviation is used more often in practice, because it is measured in the same units as the original data and is therefore easier to interpret.

A small standard deviation means that the values are close to the mean and the data is consistent, while a large standard deviation means that the values are widely spread and the average is a less reliable summary. For example, two classes may both average 70 marks, but if one class has marks close to 70 and the other has marks ranging from 30 to 100, the two classes have performed very differently, and only a measure of dispersion reveals that difference. Dispersion is therefore used in quality control, in comparing the consistency of machines, workers, students or treatments, and in judging whether an average can be trusted.

Important Points

- Dispersion shows spread around the average; an average alone can mislead.
- Range: largest value minus smallest value; simple but sensitive to extremes.
- Variance: average squared distance from the mean.
- Standard deviation: square root of variance, in the same units as the data.
- Small standard deviation means consistent data; large means widely spread data.
- Two sets with the same mean can differ greatly in dispersion.

Short Questions with Full Answers

Q1. Why is an average alone not enough?

An average alone is not enough because two sets of data can have the same average and yet be very different. One class may have all marks close to 70, while another has marks from 30 to 100, and both may average 70. Only a measure of dispersion shows that the first class is consistent and the second is widely spread.

Q2. What is the range, and what is its weakness?

The range is the difference between the largest and the smallest values in a set of data. For example, the range of 3, 7 and 12 is 12 minus 3, which is 9. Its weakness is that it uses only two values, so a single extreme value can make the range very large even when all the other values are close together.

Q3. How are variance and standard deviation related?

Variance measures the average squared distance of the values from their mean. Standard deviation is the square root of variance. Standard deviation is used more often in practice because it is measured in the same units as the original data, while variance is measured in squared units and is harder to interpret directly.

Q4. What does a small standard deviation mean?

A small standard deviation means that the values are close to the mean and the data is consistent. For example, if a machine fills packets with a small standard deviation in weight, most packets are close to the stated weight. A large standard deviation, by contrast, means that the values vary widely and the average is a less reliable summary.

Q5. Why is standard deviation easier to interpret than variance?

Standard deviation is easier to interpret because it is expressed in the same units as the original data, such as marks, kilograms or centimetres. Variance is expressed in squared units, such as squared marks, which have no direct everyday meaning. Taking the square root of variance therefore brings the measure back to a scale the reader can understand.

Long Questions with Full Answers

Q1. Explain the range, variance and standard deviation step by step, with an illustrative example.

Dispersion is best understood by working through its measures in order, from the simplest to the most useful. The range comes first: subtract the smallest value from the largest, so for the illustrative values 2, 4, 6 and 8, the range is 8 minus 2, which is 6. The range is quick, but it ignores every value except the two extremes, so one unusual value can distort it completely. Variance comes next: find the mean, subtract the mean from each value, square each difference so that negative and positive differences do not cancel, add those squares and divide by the number of values. Squaring is necessary, because without it the differences would add to zero, but it leaves the result in squared units. Standard deviation is the final step: take the square root of variance, which returns the measure to the original units. The result is a single figure that tells us, in ordinary units, how far values typically lie from the mean, and that is why standard deviation, rather than range or variance alone, is the standard measure of spread in most statistical work.

Q2. Two classes have the same average marks but very different results. Explain how dispersion resolves this puzzle.

Suppose two classes both average 70 marks, and a teacher who looks only at the average concludes that the classes performed equally well. That conclusion may be completely wrong, because the average hides how the marks are distributed. In the first class, most marks may lie between 65 and 75, which shows steady, consistent teaching and learning, with no student far behind and none far ahead. In the second class, marks may range from 30 to 100, which shows a very different situation: some students have mastered the work, others have understood very little, and the average of 70 describes almost nobody in the class. A measure of dispersion makes this difference visible at once, because the first class will have a small standard deviation and the

second a large one. The practical lesson is that an average should never be reported without some measure of spread, whether in a school, a factory or a hospital. Consistency matters as much as the average itself, and only dispersion tells us whether a typical value is truly typical or merely the middle point of a widely scattered set.

MCQs with Answer Key

1. The range of 3, 7 and 12 is:

- (a) 3
- (b) 9
- (c) 12
- (d) 15

Answer: (b)

2. Standard deviation is the square root of:

- (a) Mean
- (b) Median
- (c) Variance
- (d) Range

Answer: (c)

3. A small standard deviation means the data is:

- (a) Widely spread
- (b) Close to the mean
- (c) Always wrong
- (d) Without an average

Answer: (b)

4. The range uses:

- (a) All values equally
- (b) Only the largest and smallest values
- (c) Only the middle value
- (d) Only the mode

Answer: (b)

5. Standard deviation is measured in:

- (a) Squared units only
- (b) The same units as the data
- (c) Percentages only
- (d) No units at all

Answer: (b)

Chapter 5: Probability Basics

Explanation

Note: this chapter follows the standard Statistics syllabus in general terms, as stated in the introduction. It is not a word-for-word copy of any textbook chapter.

Probability is a number between 0 and 1 that measures how likely an event is. A probability of 0 means the event is impossible, a probability of 1 means it is certain, and a probability of 0.5 means it is equally likely to happen or not to happen. An experiment is an action whose outcome is not known in advance, such as tossing a coin or rolling a die. The sample space is the set of all possible outcomes of the experiment, and an event is a particular outcome or set of outcomes that we are interested in. For a fair coin, the sample space is heads and tails, and the probability of getting a head, as an illustrative example, is 1 divided by 2, which is 0.5. For a fair die, the sample space is 1, 2, 3, 4, 5 and 6, and the probability of rolling a 4 is 1 divided by 6.

Probability is calculated, in the simplest case, by dividing the number of favourable outcomes by the total number of equally likely outcomes. This rule only works when the outcomes are equally likely, which is why a fair coin and a fair die are the usual examples. Two events are mutually exclusive if they cannot happen at the same time, such as rolling a 2 and rolling a 5 on a single throw. Two events are independent if the outcome of one does not affect the outcome of the other, such as two separate coin tosses. The complement of an event is the event not happening, and its probability is 1 minus the probability of the event itself, so if the probability of rain is 0.3, the probability of no rain is 0.7.

Important Points

- Probability lies between 0 (impossible) and 1 (certain).
- Experiment: an action with an uncertain outcome. Sample space: all possible outcomes. Event: an outcome of interest.
- Simple probability: favourable outcomes divided by total equally likely outcomes.
- Mutually exclusive events cannot happen together; independent events do not affect each other.
- Complement rule: probability of not happening equals 1 minus the probability of happening.

Short Questions with Full Answers

Q1. What is probability?

Probability is a number between 0 and 1 that measures how likely an event is to happen. A probability of 0 means the event is impossible, a probability of 1 means it is certain, and values in between show degrees of likelihood. Probability therefore turns a vague word such as likely into a precise number.

Q2. What is a sample space, and what is an event?

A sample space is the set of all possible outcomes of an experiment. For rolling a die, the sample space is 1, 2, 3, 4, 5 and 6. An event is a particular outcome or set of outcomes that we are interested in, for example rolling an even number, which is the set 2, 4 and 6.

Q3. What is the probability of getting a head on a fair coin?

The probability of getting a head on a fair coin is 1 divided by 2, which is 0.5. This is because there is one favourable outcome, heads, out of two equally likely outcomes, heads and tails. The same reasoning gives a probability of 0.5 for getting a tail.

Q4. What do probabilities of 0 and 1 mean?

A probability of 0 means that the event is impossible and cannot happen, for example rolling a 7 on an ordinary die. A probability of 1 means that the event is certain and must happen, for example rolling a number between 1 and 6 on that die. Every other event has a probability strictly between 0 and 1.

Q5. Why can a probability never be greater than 1 or less than 0?

A probability can never be greater than 1 because the favourable outcomes cannot exceed the total number of possible outcomes. It can never be less than 0 because the number of favourable outcomes cannot be negative. The probability scale is therefore fixed between 0 and 1, which makes probabilities from different problems directly comparable.

Long Questions with Full Answers

Q1. Explain the basic ideas of probability, using a coin and a die as illustrative examples.

Probability begins with an experiment, which is any action whose outcome is not known in advance, such as tossing a coin or rolling a die. The sample space is the complete set of possible outcomes: heads and tails for a coin, and 1, 2, 3, 4, 5 and 6 for a die. An event is the particular outcome we are interested in, such as getting a head, or rolling an even number. When all outcomes are equally likely, the probability of an event is found by dividing the number of favourable outcomes by the total number of outcomes in the sample space. For a fair coin, there is one favourable outcome out of two, so the probability of a head is 0.5. For a fair die, rolling a 4 has one favourable outcome out of six, so its probability is 1 divided by 6, while rolling an even number has three favourable outcomes, 2, 4 and 6, out of six, so its probability is 3 divided by 6, which is 0.5. These examples show the whole method in miniature: list the sample space, count the favourable outcomes, and divide. The same method, with more careful counting, is used for larger problems, from drawing cards to sampling students, and it works only when the outcomes are genuinely equally likely, which is why fairness of the coin or die must be assumed or checked.

Q2. Why is probability useful in real life, even though it cannot predict a single outcome with certainty?

Probability is useful precisely because real life is uncertain, and it gives us a disciplined way to deal with that uncertainty instead of guessing. A doctor cannot say with certainty whether one patient will recover, but from large numbers of similar cases, probability can say how likely recovery is, and that figure helps the patient and the doctor choose a treatment. A business cannot know next month's sales exactly, but probability based on past sales helps it decide how much stock to hold and how much risk to accept. An insurance company cannot predict which house will catch fire, but it can estimate how many houses in a thousand will do so, and set its prices accordingly. In each case, probability does not promise what will happen in a single trial, and a probability of 0.9 does not

mean that the event must happen. What it does is turn experience into a number, so that decisions can be compared, risks can be weighed, and plans can be made on evidence rather than on hope or fear. That is why probability is the bridge between descriptive statistics, which summarises what has happened, and inferential statistics, which draws conclusions about what is likely to happen next.

MCQs with Answer Key

1. The probability of an impossible event is:

- (a) 1
- (b) 0.5
- (c) 0
- (d) 2

Answer: (c)

2. The probability of a certain event is:

- (a) 0
- (b) 0.5
- (c) 1
- (d) 1.5

Answer: (c)

3. The probability of a head on a fair coin is:

- (a) 0
- (b) 0.25
- (c) 0.5
- (d) 1

Answer: (c)

4. The sample space for rolling a die contains:

- (a) 2 outcomes
- (b) 4 outcomes
- (c) 6 outcomes
- (d) 12 outcomes

Answer: (c)

5. If the probability of rain is 0.3, the probability of no rain is:

- (a) 0.3
- (b) 0.6
- (c) 0.7
- (d) 1.3

Answer: (c)

Chapter 6: Index Numbers and Sampling

Explanation

Note: this chapter follows the standard Statistics syllabus in general terms, as stated in the introduction. It is not a word-for-word copy of any textbook chapter.

Index numbers compare prices or quantities over time, so that change can be seen clearly even when the actual figures are large or complicated. A base year is chosen as the reference year, and its value is set to 100. The index for any other year is then found by dividing that year's value by the base-year value and multiplying by 100. As an illustrative example, if a basket of goods cost 200 in the base year and 250 in a later year, the later index is 250 divided by 200 multiplied by 100, which is 125, meaning prices have risen by 25 percent since the base year. Index numbers are used for the cost of living, for wages, for production and for trade, because they turn a long series of prices into a simple series of comparisons.

Sampling means studying a part, called a sample, instead of the whole population, in order to save time, money and effort, while still getting useful estimates. A population is the complete set of items or people under study, and a sample is a part of it chosen for study. Random sampling gives every member of the population a fair chance of selection, which helps the sample to be representative, meaning that it resembles the population in its important features. Other methods include stratified sampling, where the population is divided into groups and a sample is taken from each group, and systematic sampling, where every n th member is chosen after a random start. Sampling always carries some sampling error, because a sample is not the whole population, but a well-chosen sample gives a good estimate at a fraction of the cost of a full census, which is a study of every member.

Important Points

- Index numbers compare values over time; the base year is set to 100.
- Index formula: current value divided by base-year value, multiplied by 100.
- An index of 125 means a 25 percent rise since the base year.
- Population: the whole set under study. Sample: a part chosen for study.
- Random sampling gives every member a fair chance and helps representativeness.
- Sampling saves time and cost, but always carries some sampling error.

Short Questions with Full Answers

Q1. What is an index number?

An index number is a figure that compares a price or a quantity in one period with the same item in a base period. It turns a long series of actual prices into a simple series of comparisons, so that change over time can be seen at a glance. Index numbers are widely used for the cost of living, wages, production and trade.

Q2. What is a base year, and why is it set to 100?

A base year is the reference year against which other years are compared in an index number

series. Its value is set to 100 by convention, so that later values can be read as percentages of the base year. An index of 125 therefore means a 25 percent rise since the base year, and an index of 90 means a 10 percent fall.

Q3. What is the difference between a population and a sample?

A population is the complete set of items or people under study, for example all students in a city. A sample is a part of that population chosen for study, for example 200 of those students. Studying the whole population is called a census, while studying a sample is called sampling, and it is used when a census would be too costly or too slow.

Q4. Why is random sampling considered fair?

Random sampling is considered fair because it gives every member of the population an equal chance of being selected. This fairness helps the sample to resemble the population, rather than favouring one group over another. A sample chosen by favouritism, by convenience alone, or by picking only the easiest cases to reach is likely to be biased and misleading.

Q5. What is sampling error?

Sampling error is the difference between a result obtained from a sample and the result that would have been obtained from the whole population. It arises because a sample is only a part of the population, and different samples give slightly different results. Sampling error can be reduced by taking a larger, well-chosen sample, but it cannot be removed entirely while only a part is studied.

Long Questions with Full Answers

Q1. Explain how index numbers work, with an illustrative example, and state their uses and limits.

An index number works by fixing one period as a base and expressing every other period as a percentage of that base, so that change over time becomes easy to read. The base year is chosen first, and its value is set to 100 by convention. The index for any other year is then found by dividing that year's value by the base-year value and multiplying by 100. As an illustrative example, suppose a basket of goods cost 200 in the base year and 250 in a later year. The later index is 250 divided by 200 multiplied by 100, which is 125, and the meaning is immediate: prices have risen by 25 percent since the base year, without the reader needing to remember the original prices. Index numbers are used for the cost of living, for wages, for industrial production and for trade, because they compress long price series into a single, comparable line of figures. They also have limits that must be remembered. The choice of base year affects the picture, the basket of goods may become out of date as habits change, and an index for a whole country may hide very different changes for different groups, such as the poor and the rich. An index is therefore a useful summary, but it should be read with its base year, its basket and its coverage clearly in mind.

Q2. Why do researchers use samples instead of studying the whole population, and what makes a sample good?

Researchers use samples because studying a whole population, which is called a census, is often too expensive, too slow or simply impossible. A census of a large city would need enormous staff, time and money, and by the time it was finished, the population would already have changed. A

well-chosen sample avoids those problems, because a few hundred carefully selected cases can give a good estimate of the whole population at a fraction of the cost and in a fraction of the time. What makes a sample good is representativeness: it should resemble the population in the features that matter for the study, such as age, gender, income or region. Random sampling helps achieve this, because giving every member a fair chance of selection prevents the researcher, consciously or not, from favouring one group. Stratified sampling goes further by dividing the population into groups and sampling from each group, so that small groups are not missed by chance. Even the best sample, however, carries sampling error, because it is only a part, and it can also suffer from non-sampling errors, such as wrong answers or missing cases. The practical conclusion is that a sample is a trade-off: it saves time and money, but its results must always be reported as estimates, with their limits honestly stated, rather than as exact facts about the whole population.

MCQs with Answer Key

1. In index numbers, the base-year value is usually set to:

- (a) 0
- (b) 10
- (c) 100
- (d) 1000

Answer: (c)

2. If a basket cost 200 in the base year and 250 later, the index is:

- (a) 80
- (b) 100
- (c) 125
- (d) 200

Answer: (c)

3. Random sampling means:

- (a) Choosing favourites
- (b) Every member has a fair chance
- (c) Choosing only experts
- (d) No selection at all

Answer: (b)

4. Studying the whole population is called a:

- (a) Sample
- (b) Census
- (c) Index
- (d) Estimate

Answer: (b)

5. Sampling error arises because:

- (a) The sample is the whole population
- (b) Only a part of the population is studied
- (c) No data is collected

(d) The base year is wrong

Answer: (b)

Revision Checklist

- Chapter 1: meanings, descriptive vs inferential, primary vs secondary, variables, errors, significant digits, scientific notation, rounding, summation.
- Chapter 2: classification types, tabulation rules, frequency distribution, class limits vs boundaries, bar chart vs histogram.
- Chapter 3: mean, median, mode - know when each is best.
- Chapter 4: range, variance, standard deviation - what spread means.
- Chapter 5: probability 0 to 1, sample space, simple examples.
- Chapter 6: index numbers, base year, random sampling.
- Practice: attempt each chapter's MCQs without looking, then check answers, then revise weak topics from the textbook.