A Level Statistics
Practice Test 1: Data Collection

Instructions:
Answer all questions. Show your working clearly.
Calculators may be used unless stated otherwise.
Draw diagrams where appropriate to illustrate your solutions.
Time allowed: 3 hours

Section A: Types of Data [25 marks]

1. [**8 marks**] Define and distinguish between different types of data:

   (a) Define qualitative data and give three examples.

   (b) Define quantitative data and give three examples.

   (c) Explain the difference between discrete and continuous data.

   (d) Give two examples each of discrete and continuous quantitative data.

   2. [**10 marks**] Classify the following variables and explain your reasoning:

   (a) Height of students in a class

   (b) Shoe sizes available in a shop

   (c) Blood type (A, B, AB, O)

   (d) Time taken to run 100 metres

   (e) Number of goals scored in football matches

   (f) Temperature measured to the nearest degree

   (g) Car colors in a car park

   (h) Annual income of employees

   (i) Star ratings (1 to 5) for restaurants

   (j) Weight of apples

   3. [**7 marks**] A researcher is studying customer satisfaction at a restaurant. They collect the following data:

   - Customer age

   - Satisfaction rating (Poor, Fair, Good, Excellent)

   - Amount spent on meal

- Number of previous visits

- Preferred cuisine type

(a) Classify each variable as qualitative or quantitative.

(b) For quantitative variables, identify whether they are discrete or continuous.

(c) Suggest appropriate methods for collecting each type of data.

Section B: Sampling Methods Fundamentals [30 marks]

4. [**12 marks**] Define and explain the four main sampling methods:

(a) Random sampling: Define and explain how to implement it.

(b) Systematic sampling: Describe the method and when it's appropriate.

(c) Stratified sampling: Explain the process and its advantages.

(d) Cluster sampling: Describe when and why this method is used.

5. [**18 marks**] A school has 1200 students: 400 in Year 12 (200 male, 200 female) and 800 in Year 13 (350 male, 450 female). A sample of 120 students is required for a survey.

(a) Describe how to select a simple random sample of 120 students.

(b) For systematic sampling, calculate the sampling interval and describe the selection process.

(c) Design a stratified sampling scheme using year group as the stratifying variable. Calculate the sample sizes for each stratum.

(d) Design a stratified sampling scheme using both year group and gender. Calculate all sample sizes.

(e) For cluster sampling, if classes are used as clusters with 30 students per class, how many clusters should be selected?

(f) Compare the advantages and disadvantages of each method for this scenario.

Section C: Bias in Sampling [35 marks]

6. [**15 marks**] Explain different types of bias and their effects on data quality:

(a) Define sampling bias and explain how it occurs.

(b) Describe selection bias and give an example.

(c) Explain response bias and suggest ways to minimize it.

(d) Define non-response bias and its potential impact on results.

(e) Describe convenience sampling bias and why it should be avoided.

7. [**20 marks**] Analyze the following scenarios for potential bias:
Scenario 1: A survey about internet usage is conducted by calling landline telephone numbers between 9 AM and 5 PM on weekdays.
Scenario 2: A political opinion poll is conducted by standing outside a particular political party's headquarters.
Scenario 3: A health survey uses a mailed questionnaire with a return rate of 25%.

Scenario 4: A study of student study habits selects participants from the library during exam period.

Scenario 5: An online survey about technology use is shared only on social media platforms.
For each scenario:

(a) Identify the type(s) of bias present.

(b) Explain how this bias might affect the results.

(c) Suggest improvements to reduce bias.

(d) Describe the likely direction of any bias in the results.

Section D: Frequency Tables and Data Organization [25 marks]

8. [**12 marks**] The following data shows the weights (in kg) of 40 adults:

$$68, 75, 82, 91, 63, 78, 85, 72, 69, 87, 94, 76, 81, 67, 73, 89, 66, 79, 83, 95,$$
$$71, 77, 86, 92, 64, 74, 88, 70, 84, 90, 65, 80, 93, 75, 82, 68, 76, 87, 71, 96$$

(a) Construct a frequency table using class intervals of width 5 kg, starting from 60 kg.

(b) Calculate the relative frequency for each class.

(c) Determine the modal class.

(d) Calculate the percentage of adults weighing between 70 kg and 85 kg.

(e) Construct a cumulative frequency table.

9. [**13 marks**] A survey of household incomes (in thousands of pounds) in a town gives the following frequency table:

Income (£'000)	Frequency	Cumulative Frequency
20-30	15	15
30-40	28	43
40-50	35	78
50-60	42	120
60-70	38	158
70-80	25	183
80-90	12	195
90-100	5	200

(a) Calculate the relative frequency for each class.

(b) Determine which income bracket contains the median.

(c) Estimate the median income using interpolation.

(d) Calculate the percentage of households earning less than £55,000.

(e) Find the interquartile range by estimating Q1 and Q3.

(f) Comment on the distribution of household incomes.

Section E: Bar Charts and Pie Charts [20 marks]

10. [**10 marks**] A survey of 240 students' favorite subjects gives the following results:

Subject	Number of Students
Mathematics	72
Science	54
English	48
History	36
Art	30

(a) Construct a bar chart for this data.

(b) Calculate the angles for a pie chart representation.

(c) Draw a pie chart for this data.

(d) Compare the effectiveness of bar charts versus pie charts for this data.

(e) If the survey were extended to include 60 more students with the same proportions, calculate the new frequencies.

    11. [**10 marks**] A company's quarterly sales data (in millions) is:

	Q1	Q2	Q3	Q4
2022	15	18	22	25
2023	17	20	24	28

(a) Create a comparative bar chart showing both years.

(b) Calculate the percentage increase in sales from Q1 to Q4 in each year.

(c) Determine which quarter showed the highest year-on-year growth.

(d) Construct a pie chart for 2023 data showing the proportion of annual sales in each quarter.

(e) Comment on the seasonal trends visible in the data.

Section F: Histograms [30 marks]

12. [**15 marks**] Explain the construction and interpretation of histograms:

(a) Define a histogram and explain how it differs from a bar chart.

(b) Explain why frequency density is used instead of frequency on the y-axis.

(c) Describe how to determine appropriate class intervals for a histogram.

(d) Explain how to interpret the shape of a histogram.

(e) Describe what equal and unequal class widths mean for histogram construction.

    13. [**15 marks**] The following data shows the time (in minutes) spent on homework by 80 students:

Time (minutes)	Frequency
0-20	8
20-30	12
30-40	18
40-60	25
60-80	12
80-120	5

(a) Calculate the frequency density for each class interval.

(b) Construct a histogram using these frequency densities.

(c) Identify the modal class.

(d) Estimate the mean time spent on homework.

(e) Describe the shape of the distribution.

(f) Estimate the median time using the histogram.

(g) Calculate what percentage of students spent more than 50 minutes on homework.

Section G: Box Plots [25 marks]

14. [**12 marks**] Explain box plots and their components:

(a) Define the five-number summary used in box plots.

(b) Explain how to identify outliers using the $1.5 \times$ IQR rule.

(c) Describe what each part of a box plot represents.

(d) Explain how box plots are useful for comparing distributions.

(e) Describe the difference between box plots with and without outliers marked.

15. [**13 marks**] The following data represents test scores for two classes:
Class A: 45, 52, 58, 61, 65, 67, 69, 71, 73, 75, 77, 79, 82, 85, 88, 91, 94
Class B: 38, 41, 55, 59, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 83, 87, 95

(a) Calculate the five-number summary for each class.

(b) Construct box plots for both classes on the same scale.

(c) Identify any outliers in either dataset.

(d) Compare the central tendency of the two classes.

(e) Compare the spread of scores between the classes.

(f) Calculate the interquartile range for each class.

(g) Which class performed better overall? Justify your answer.

Section H: Choosing Appropriate Representations [15 marks]

16. [**8 marks**] For each scenario, choose the most appropriate data representation and justify your choice:

(a) Showing the market share of different smartphone brands.

(b) Displaying the distribution of heights in a population.

(c) Comparing exam scores between three different schools.

(d) Showing the relationship between temperature and ice cream sales over time.

(e) Displaying the number of cars sold each month by a dealership.

(f) Comparing the spread of house prices in different neighborhoods.

(g) Showing the proportion of different blood types in a population.

(h) Displaying the distribution of waiting times at a hospital.

17. [**7 marks**] A market research company needs to present data about consumer preferences to a client. They have collected both quantitative data (spending amounts, ages) and qualitative data (brand preferences, satisfaction levels).

(a) Suggest appropriate representations for the quantitative data.

(b) Recommend suitable charts for the qualitative data.

(c) Explain why different representations are needed for different types of data.

(d) Describe how to make the presentation accessible to non-technical stakeholders.

Section I: Data Collection Design [25 marks]

18. [**15 marks**] Design a data collection strategy for the following research question: "What factors influence students' choice of university?"

(a) Define the target population for this study.

(b) List the variables you would need to collect data on.

(c) Classify each variable as qualitative/quantitative and discrete/continuous where appropriate.

(d) Choose an appropriate sampling method and justify your choice.

(e) Design a sampling strategy including sample size considerations.

(f) Identify potential sources of bias and how to minimize them.

(g) Suggest appropriate data collection methods (surveys, interviews, etc.).

19. [**10 marks**] A local council wants to assess resident satisfaction with public services. They plan to survey 500 residents from a town of 50,000 people.

(a) Design a stratified sampling scheme using appropriate strata.

(b) Calculate sample sizes for each stratum.

(c) Identify potential challenges in data collection.

(d) Suggest methods to improve response rates.

(e) Describe how to ensure the sample is representative.

(f) Propose appropriate questions for measuring satisfaction.

Section J: Comprehensive Data Analysis [30 marks]

20. [**20 marks**] A fitness center collected data on 200 members' ages and weekly gym visits:

Age Group	Number of Members	Average Weekly Visits
18-25	45	4.2
26-35	68	3.8
36-45	52	3.1
46-55	25	2.6
56-65	10	2.1

(a) Construct a frequency table showing the age distribution.

(b) Create a histogram for the age distribution.

(c) Calculate the mean age of gym members (use midpoint method).

(d) Determine which age group has the median member.

(e) Create a bar chart showing average weekly visits by age group.

(f) Calculate the overall average weekly visits across all members.

(g) Construct a pie chart showing the proportion of members in each age group.

(h) Comment on the relationship between age and gym usage.

21. [**10 marks**] An online retailer wants to analyze customer data to improve their service. They have access to: - Order values - Customer locations - Product categories purchased - Customer age groups - Delivery times - Customer satisfaction ratings

(a) Classify each variable by type.

(b) Suggest appropriate sampling methods if they can't analyze all customers.

(c) Recommend suitable data representations for each variable.

(d) Identify potential sources of bias in their data.

(e) Propose ways to collect additional useful data.

(f) Design a dashboard layout to present key findings to management.

Answer Space
Use this space for your working and answers.

Formulae and Key Concepts

Frequency Density:
$$\text{Frequency Density} = \frac{\text{Frequency}}{\text{Class Width}}$$

Sampling:
$$\text{Systematic sampling interval} = \frac{\text{Population size}}{\text{Sample size}}$$
$$\text{Stratified sample size for stratum} = \frac{\text{Stratum size}}{\text{Population size}} \times \text{Total sample size}$$

Box Plot (Five-number summary):
Minimum, Q1, Median, Q3, Maximum

$$IQR = Q3 - Q1$$

Outliers: Values $<$ Q1 - 1.5 $\times$ IQR or $>$ Q3 + 1.5 $\times$ IQR

Pie Chart:

Angle for category $= \frac{\text{Category frequency}}{\text{Total frequency}} \times 360$

Relative Frequency:

Relative Frequency $= \frac{\text{Frequency}}{\text{Total number of observations}}$

Interpolation for grouped data:

Estimated value = L $+ \frac{(n \times p - CF)}{f} \times h$

where L = lower boundary, n = total frequency, p = proportion,
CF = cumulative frequency before the class, f = class frequency, h = class width

END OF TEST

Total marks: 280

For more resources and practice materials, visit:
stepupmaths.co.uk