A Level Statistics
Practice Test 5: 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: Advanced Statistical Measures [25 marks]

1. **[8 marks]** Define and explain advanced measures of central tendency and spread:

(a) Define the geometric mean and explain when it's more appropriate than arithmetic mean.

(b) Explain weighted averages and provide a practical example.

(c) Define variance and standard deviation, explaining their relationship.

(d) Describe standardized scores (z-scores) and their interpretation.

2. **[10 marks]** A company has three departments with the following salary data:

Department	Number of Employees	Average Salary (£)
Sales	25	32,000
Technical	15	45,000
Management	10	65,000

Additional data for Technical department salaries: Standard deviation = £8,000

(a) Calculate the overall weighted mean salary for the company.

(b) If a Technical department employee earns £53,000, calculate their z-score.

(c) Interpret the meaning of this z-score.

(d) Calculate the total salary bill for each department.

(e) Determine what percentage of the total wage bill goes to Management.

3. **[7 marks]** Investment returns over 5 years are: 8

(a) Calculate the arithmetic mean return.

(b) Calculate the geometric mean return.

(c) Explain why the geometric mean is more appropriate for investment returns.

(d) If you invested £1000 initially, calculate the final value using the geometric mean.

Section B: Data Transformation and Standardization [30 marks]

4. **[12 marks]** Explain data transformation techniques:

(a) Define linear transformation and explain its effect on mean and standard deviation.

(b) Describe logarithmic transformation and when it's used.

(c) Explain standardization (z-score transformation) and its benefits.

(d) Describe rank transformation and its applications.

5. **[18 marks]** A dataset of house prices (£000s) has been collected: 180, 195, 210, 225, 240, 255, 270, 285, 300, 320, 340, 380, 450, 520, 680

(a) Calculate the mean and standard deviation of the original data.

(b) Apply a linear transformation: New price = (Original price - 200) ÷ 50.

(c) Calculate the mean and standard deviation of the transformed data.

(d) Verify the relationship between original and transformed statistics.

(e) Identify any outliers using the 2-standard deviation rule on original data.

(f) Transform the data to z-scores and identify which houses are more than 1.5 standard deviations from the mean.

(g) Calculate the 5-number summary for the original data.

(h) Create a box plot and identify outliers using the IQR method.

(i) Compare outlier detection methods and comment on their effectiveness.

Section C: Research Ethics and Data Protection [35 marks]

6. **[15 marks]** Explain ethical considerations in data collection and research:

(a) Define informed consent and explain its importance in research.

(b) Describe confidentiality and anonymity in data collection.

(c) Explain the principle of "do no harm" in statistical research.

(d) Describe data protection requirements for personal information.

(e) Explain voluntary participation and the right to withdraw from studies.

7. **[20 marks]** Evaluate the ethical and practical considerations in these research scenarios:

Scenario 1: A social media company analyzes user behavior patterns from millions of accounts to improve their recommendation algorithm, using posts, likes, and browsing history.

Scenario 2: A medical researcher studies the effectiveness of a new treatment by tracking patient recovery times, requiring access to detailed medical records.

Scenario 3: A school district evaluates teacher performance using student test scores, classroom observations, and parent feedback surveys.

Scenario 4: A marketing company conducts surveys about shopping habits by offering gift vouchers to participants, collecting income and spending data.

For each scenario, address:

(a) Key ethical issues and potential risks to participants.

(b) Data protection and privacy concerns.

(c) Informed consent requirements and challenges.

(d) Potential benefits versus risks of the research.

(e) Recommendations for ethical data collection practices.

Answer Space

Use this space for your working and answers.

Formulae and Key Concepts

Advanced Means:

Arithmetic Mean: $\bar{x} = \frac{\sum x}{n}$

Weighted Mean: $\bar{x}_w = \frac{\sum w_i x_i}{\sum w_i}$

Geometric Mean: $\bar{x}_g = \sqrt[n]{x_1 \times x_2 \times \ldots \times x_n}$

For percentages: $\bar{x}_g = \sqrt[n]{(1+r_1)(1+r_2)\ldots(1+r_n)} - 1$

Measures of Spread:

Variance: $s^2 = \frac{\sum(x-\bar{x})^2}{n-1}$

Standard Deviation: $s = \sqrt{s^2}$

Range: Maximum - Minimum

IQR: Q3 - Q1

Standardization:

Z-score: $z = \frac{x-\bar{x}}{s}$

Interpretation: Number of standard deviations from mean

$|z| > 2$: Potential outlier, $|z| > 3$: Likely outlier

Linear Transformation:

If $Y = aX + b$, then:

$\bar{y} = a\bar{x} + b$

$s_y = |a| \times s_x$

Outlier Detection:

Z-score method: $|z| > 2$ or $|z| > 3$

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

Modified z-score using median and MAD

Box Plot Components:

Five-number summary: Min, Q1, Median, Q3, Max

Whiskers extend to furthest non-outlier
Outliers plotted separately

Research Ethics Principles:
Autonomy: Respect for persons and their decisions
Beneficence: Maximizing benefits and minimizing harm
Justice: Fair distribution of benefits and burdens
Non-maleficence: "Do no harm"

Data Protection Requirements:
Lawful basis for processing
Data minimization principle
Accuracy and timeliness
Storage limitation
Security and confidentiality
Accountability and transparency

Informed Consent Elements:
Purpose of research clearly explained
Procedures and time commitments
Risks and benefits disclosed
Voluntary participation emphasized
Right to withdraw without penalty
Contact information provided

END OF TEST

Total marks: 90

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