A Level Statistics
Practice Test 5: Measures of Location and Spread

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 Percentiles and Quantile Analysis [25 marks]

1. **[12 marks]** Define and calculate advanced quantile measures:

(a) Define quantiles and explain how they divide a dataset into equal parts.

(b) Calculate the 5th, 15th, 85th, and 95th percentiles for the dataset: 12, 15, 18, 22, 25, 28, 31, 35, 38, 42, 45, 48, 51, 55, 58, 62, 65, 68, 72, 75

(c) Define and calculate the interdecile range (P90 - P10).

(d) Explain how the interdecile range compares to the interquartile range as a measure of spread.

(e) Calculate the coefficient of quartile deviation: $\frac{Q3-Q1}{Q3+Q1}$ and interpret its meaning.

(f) Create a percentile rank table showing what percentage of data falls below various values.

2. **[8 marks]** Analyze quantile-based distribution characteristics:

(a) Explain how to use quantiles to assess distribution symmetry.

(b) For a dataset where P25 = 40, P50 = 55, P75 = 65, assess the distribution shape using quantile spacing.

(c) Calculate and interpret the quantile skewness measure: $\frac{(P75-P50)-(P50-P25)}{P75-P25}$.

(d) Compare this quantile-based skewness with Pearson's coefficient method in terms of robustness.

3. **[5 marks]** Apply quantile analysis to risk assessment:

(a) Define Value at Risk (VaR) in terms of percentiles.

(b) Investment returns show P5 = -8

(c) Explain how conditional Value at Risk (CVaR) extends beyond simple VaR calculations.

Section B: Multimodal Distributions and Complex Patterns [30 marks]

4. **[15 marks]** Analyze datasets with multiple modes and complex patterns:

(a) Define unimodal, bimodal, and multimodal distributions.

(b) For the dataset: 15, 16, 17, 17, 18, 18, 18, 25, 26, 27, 27, 28, 28, 28, identify all modes and classify the distribution.

(c) Calculate the mean and median for this bimodal dataset.

(d) Explain why traditional measures of central tendency may be misleading for multimodal data.

(e) Describe methods for identifying the number of modes in a continuous distribution.

(f) Calculate the separation between modes and assess if they represent distinct populations.

5. **[15 marks]** Examine employee salary data showing potential bimodal structure:
Salaries (£000s): 28, 30, 32, 34, 35, 36, 38, 42, 44, 46, 58, 62, 65, 68, 70, 72, 75, 78, 82, 85, 88, 92, 95, 98

(a) Create a histogram and identify potential modes in the distribution.

(b) Calculate the overall mean, median, and standard deviation.

(c) Identify the two apparent groups and calculate separate statistics for each.

(d) Group 1 (£50k): Calculate mean, median, and standard deviation.

(e) Group 2 (¿£50k): Calculate mean, median, and standard deviation.

(f) Compare the coefficient of variation for each group versus the combined dataset.

(g) Explain how the bimodal structure affects the interpretation of overall statistics.

(h) Calculate the proportion of employees in each salary group.

(i) Discuss the business implications of this salary distribution pattern.

Section C: Statistical Modeling and Prediction [35 marks]

6. **[18 marks]** Use statistical measures for performance prediction and modeling:
A sports analytics team tracks player performance scores over a season. Here are the final 20 game scores for two players:
Player A: 78, 82, 75, 88, 79, 91, 84, 77, 85, 89, 83, 76, 90, 81, 87, 74, 92, 86, 80, 88 **Player B:** 85, 72, 94, 68, 89, 76, 91, 83, 69, 95, 74, 88, 82, 71, 93, 87, 70, 96, 79, 90

(a) Calculate comprehensive summary statistics for both players (mean, median, SD, CV, quartiles).

(b) Determine which player has more consistent performance using appropriate measures.

(c) Calculate the probability that each player scores above 85 (assume normal distribution).

(d) Create box plots for both players and identify any outliers.

(e) Calculate the range that contains 68

(f) If team strategy requires consistent scoring above 80, which player should be selected?

(g) Predict the likely range of scores for each player's next game (using mean ± 1SD).

(h) Calculate confidence intervals for the true mean performance of each player.

(i) Assess which player has better clutch performance (top quartile scores).

7. **[17 marks]** Apply advanced statistical analysis to market research data:
A market research study examines customer spending patterns across different age groups:
Young (18-30): n=45, mean=£142, SD=£38, median=£135, Q1=£118, Q3=£165 **Middle (31-50):** n=60, mean=£284, SD=£67, median=£275, Q1=£238, Q3=£320 **Senior (51+):** n=35, mean=£198, SD=£52, median=£205, Q1=£165, Q3=£235

(a) Calculate the coefficient of variation for each age group and interpret the results.

(b) Determine which age group has the most predictable spending patterns.

(c) Calculate the combined mean and standard deviation for all customers.

(d) Identify which age groups show skewed spending distributions and the direction of skew.

(e) Calculate the percentage of customers in each group spending above £300.

(f) Determine the optimal price point that captures 75

(g) Compare the spread of the middle 50

(h) Create a comprehensive market segmentation strategy based on these statistics.

(i) Calculate weighted percentiles for the combined customer base.

Answer Space

Use this space for your working and answers.

Formulae and Key Concepts

Advanced Percentiles:

Percentile position: $P_k = \frac{k(n+1)}{100}$ for kth percentile
Interdecile range: $P_{90} - P_{10}$ (80% of data)
Coefficient of quartile deviation: $\frac{Q_3-Q_1}{Q_3+Q_1}$
Percentile rank of value x: $\frac{\text{Number below x}}{\text{Total}} \times 100$

Quantile-Based Skewness:

Bowley's skewness: $\frac{(Q_3-Q_2)-(Q_2-Q_1)}{Q_3-Q_1}$
Interpretation: Positive = right skewed, Negative = left skewed
More robust than moment-based skewness measures

Risk Measures:

Value at Risk (VaR): Percentile corresponding to confidence level
95% VaR = 5th percentile of loss distribution

Conditional VaR: Expected loss beyond VaR threshold

Distribution Classification:
Unimodal: Single peak/mode
Bimodal: Two distinct peaks
Multimodal: Three or more peaks
Mode separation: Distance between modal values

Normal Distribution Properties:
68% within $\pm$
95% within $\pm$ 2
99.7% within $\pm$ 3
Standard score: $z = \frac{x-\mu}{\sigma}$

Confidence Intervals:
Mean: $\bar{x} \pm t_{\alpha/2}\frac{s}{\sqrt{n}}$ (for small samples)
Mean: $\bar{x} \pm z_{\alpha/2}\frac{s}{\sqrt{n}}$ (for large samples)

Combined Statistics:
Weighted mean: $\bar{x}_w = \frac{\sum w_i x_i}{\sum w_i}$
Weighted variance: More complex calculation involving group means
Weighted percentiles: Account for group sizes in calculations

Coefficient of Variation:
$CV = \frac{\text{Standard Deviation}}{\text{Mean}} \times 100\%$
Lower CV indicates more consistent/predictable performance
Allows comparison across different scales

Performance Analysis:
Consistency: Measured by CV or standard deviation
Reliability: Probability of exceeding threshold
Range prediction: Mean $\pm$ k$\times$SD for various confidence levels
Outlier identification: Values beyond Q1-1.5$\times$IQR or Q3+1.5$\times$IQR

Market Segmentation Metrics:
Market coverage: Percentage within price ranges
Segment variability: CV within each segment
Optimal pricing: Based on percentile analysis
Cross-segment comparison: Standardized measures

END OF TEST

Total marks: 90

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