A Level Statistics
Practice Test 3: 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: Skewness and Distribution Shape Analysis [25 marks]

1. **[12 marks]** Define and analyze skewness in data distributions:

(a) Define positive (right) skewness and negative (left) skewness.

(b) Explain the relationship between mean, median, and mode in skewed distributions.

(c) Calculate Pearson's coefficient of skewness using the formula: Skewness $= \frac{3(\text{mean}-\text{median})}{\text{standard deviation}}$

(d) For a dataset with mean = 85, median = 78, standard deviation = 12, calculate and interpret the skewness.

(e) Describe what the quartile-based skewness measure $\frac{(Q3-Q2)-(Q2-Q1)}{Q3-Q1}$ tells us about distribution shape.

(f) Explain how outliers affect measures of skewness and central tendency.

2. **[8 marks]** Analyze the following three datasets for distribution shape:
Dataset A: Mean = 50, Median = 52, Mode = 55, SD = 8 **Dataset B:** Mean = 72, Median = 72, Mode = 72, SD = 10 **Dataset C:** Mean = 65, Median = 70, Mode = 75, SD = 15

(a) Calculate Pearson's coefficient of skewness for each dataset.

(b) Classify each distribution as symmetric, positively skewed, or negatively skewed.

(c) Sketch the approximate shape of each distribution.

(d) For each dataset, predict which measure of central tendency would be most representative and justify your choice.

3. **[5 marks]** Analyze kurtosis and distribution characteristics:

(a) Define kurtosis and explain what it measures about distribution shape.

(b) Distinguish between platykurtic, mesokurtic, and leptokurtic distributions.

(c) Explain how kurtosis affects the interpretation of standard deviation and outliers.

Section B: Robust Statistics and Resistant Measures [30 marks]

4. **[15 marks]** Define and calculate robust statistical measures:

(a) Define robust statistics and explain why they're important in data analysis.

(b) Calculate the trimmed mean (10

(c) Compare the trimmed mean with the regular mean and explain the difference.

(d) Define the median absolute deviation (MAD) and calculate it for the dataset above.

(e) Compare MAD with standard deviation as measures of spread.

(f) Explain when robust measures should be preferred over traditional measures.

5. **[15 marks]** Analyze the impact of outliers on different statistical measures:
Consider the dataset: 85, 88, 90, 92, 94, 96, 98, 100, 102, 105, 150

(a) Identify the outlier(s) using the IQR method.

(b) Calculate: mean, median, standard deviation, and IQR with the outlier included.

(c) Recalculate the same measures after removing the outlier(s).

(d) Calculate the percentage change in each measure when the outlier is removed.

(e) Determine which measures are most resistant to outliers.

(f) Create box plots with and without outliers to visualize the effect.

(g) Calculate the 5

(h) Discuss scenarios where removing outliers would or wouldn't be appropriate.

Section C: Advanced Grouped Data Analysis and Interpolation [35 marks]

6. **[18 marks]** A survey of household incomes (£000s) in a region produces the following data:

Income (£000s)	Frequency	Cumulative Frequency
10-20	18	18
20-30	32	50
30-40	45	95
40-50	38	133
50-60	25	158
60-80	22	180
80-120	12	192
120-200	8	200

(a) Calculate the estimated mean income using appropriate midpoints.

(b) Use linear interpolation to estimate the median income.

(c) Calculate the first and third quartiles using interpolation.

(d) Estimate the 90th percentile income.

(e) Calculate the estimated standard deviation.

(f) Determine the modal class and estimate the mode using interpolation.

(g) Calculate Pearson's coefficient of skewness and interpret the result.

(h) Estimate what percentage of households earn between £25,000 and £75,000.

(i) Comment on the distribution of income and identify any unusual features.

7. **[17 marks]** Compare two different frequency distributions representing test scores:
Class A:

Score	Frequency
0-20	2
20-40	8
40-60	15
60-80	18
80-100	7

Class B:

Score	Frequency
0-25	3
25-50	12
50-75	20
75-100	15

(a) Calculate the estimated mean for each class.

(b) Estimate the median for each class using interpolation.

(c) Calculate the estimated standard deviation for each class.

(d) Determine which class has more consistent performance.

(e) Calculate the coefficient of variation for each class.

(f) Estimate the percentage of students in each class who achieved above 70

(g) Use Pearson's coefficient to assess the skewness of each distribution.

(h) Which class performed better overall? Justify your answer with statistical evidence.

(i) Discuss the challenges of comparing these distributions given their different class intervals.

Answer Space

Use this space for your working and answers.

Formulae and Key Concepts

Skewness Measures:
Pearson's coefficient: Skewness = $\frac{3(\text{mean}-\text{median})}{\text{standard deviation}}$
Quartile skewness: $\frac{(Q3-Q2)-(Q2-Q1)}{Q3-Q1}$
Interpretation: Positive = right skewed, Negative = left skewed, 0 = symmetric

Robust Measures:
Trimmed mean: Remove extreme values, calculate mean of remaining data
Median Absolute Deviation: $MAD = \text{median}(|x_i - \text{median}|)$
Interquartile range: $IQR = Q3 - Q1$ (resistant to outliers)

Interpolation for Grouped Data:
Linear interpolation: $x = L + \frac{(n \times p - CF_{before})}{f} \times h$
where L = lower boundary, p = proportion (0.5 for median, 0.25 for Q1, etc.)
CF = cumulative frequency, f = class frequency, h = class width

Modal Class Estimation:
Mode $L + \frac{f_1 - f_0}{(f_1 - f_0) + (f_1 - f_2)} \times h$
where f_1 = modal class frequency, f_0 = frequency before, f_2 = frequency after

Distribution Shape Relationships:
Symmetric: Mean Median Mode
Right skewed: Mean ¿ Median ¿ Mode
Left skewed: Mean ¡ Median ¡ Mode

Outlier Impact:
Mean: Highly affected by outliers
Median: Resistant to outliers
Standard deviation: Highly affected
IQR: Resistant to outliers

Kurtosis:
Leptokurtic: Heavy tails, sharp peak (kurtosis ¿ 3)
Mesokurtic: Normal-like (kurtosis 3)
Platykurtic: Light tails, flat peak (kurtosis ¡ 3)

Coefficient of Variation:
$CV = \frac{\text{Standard Deviation}}{\text{Mean}} \times 100\%$
Used for comparing variability between different scales

Percentile Calculations:
Position of kth percentile: $P_k = \frac{k(n+1)}{100}$
For grouped data: Use interpolation with cumulative frequencies

Resistance to Outliers (Most to Least):
1. Median (most resistant)
2. IQR
3. Trimmed mean
4. Mode
5. Mean (least resistant)
6. Standard deviation (least resistant)

END OF TEST

Total marks: 90

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