A Level Statistics
Practice Test 4: 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: Combining and Comparing Datasets [25 marks]

1. [**12 marks**] Define and calculate combined statistics for merged datasets:

   (a) Explain the formula for combining means from two separate datasets.

   (b) Two groups have the following statistics: Group A (n = 25, mean = 78, SD = 12) and Group B (n = 35, mean = 82, SD = 15). Calculate the combined mean.

   (c) Derive and use the formula for combined variance when merging datasets.

   (d) Calculate the combined standard deviation for the groups in part (b).

   (e) A third group C (n = 20, mean = 75, SD = 10) is added. Calculate the overall mean and standard deviation for all three groups combined.

   (f) Explain why simply averaging the individual standard deviations doesn't give the correct combined standard deviation.

2. [**8 marks**] Analyze proportional changes and scaling effects:

   (a) A dataset has mean = 45 and SD = 8. If all values are increased by 15

   (b) Original dataset: mean = 120, SD = 24. After transformation y = 0.8x - 10, find the new mean and standard deviation.

   (c) Explain the general rules for how linear transformations affect measures of location and spread.

   (d) Two datasets are proportionally related. If Dataset 1 has CV = 25

3. [**5 marks**] Compare datasets using standardized measures:

   (a) Define the coefficient of variation and explain when it's more useful than standard deviation.

   (b) Calculate and compare CVs for: Dataset X (mean = 500, SD = 75) and Dataset Y (mean = 50, SD = 12).

   (c) Interpret which dataset has more relative variability and explain the practical implications.

Section B: Time Series Analysis and Moving Averages [30 marks]

4. [**15 marks**] Calculate and interpret moving averages:

(a) Define moving averages and explain their purpose in smoothing data trends.

(b) Calculate 3-point moving averages for the dataset: 15, 18, 22, 19, 25, 28, 24, 30, 27, 32

(c) Calculate 5-point moving averages for the same dataset where possible.

(d) Compare the smoothing effect of 3-point versus 5-point moving averages.

(e) Explain the trade-off between smoothing and data loss in moving averages.

(f) Describe when moving averages are particularly useful in statistical analysis.

5. [**15 marks**] Analyze monthly sales data (£000s) for a retail company over 18 months:
Jan-23: 45, Feb-23: 42, Mar-23: 48, Apr-23: 52, May-23: 55, Jun-23: 58 Jul-23: 62, Aug-23: 59, Sep-23: 54, Oct-23: 50, Nov-23: 65, Dec-23: 78 Jan-24: 48, Feb-24: 46, Mar-24: 52, Apr-24: 58, May-24: 61, Jun-24: 64

(a) Calculate the overall mean and standard deviation for all 18 months.

(b) Calculate 4-point moving averages to identify the underlying trend.

(c) Identify seasonal patterns by comparing the same months across years.

(d) Calculate the coefficient of variation to assess the relative variability of sales.

(e) Determine which quarters show the highest and lowest average sales.

(f) Use the moving averages to predict sales for July 2024.

(g) Calculate the mean absolute deviation from the trend (using moving averages).

(h) Comment on the business implications of the seasonal patterns observed.

Section C: Quality Control and Statistical Process Control [35 marks]

6. [**18 marks**] A manufacturing process produces components with target weight 250g. Quality control measurements over 25 samples show:
248, 251, 249, 252, 247, 253, 250, 248, 254, 249, 251, 246, 252, 250, 255, 249, 248, 251, 247, 253, 250, 249, 252, 248, 251

(a) Calculate the process mean and standard deviation.

(b) Determine if the process is centered on the target weight.

(c) Calculate control limits using the 3-sigma rule (mean $\pm$ 3$\times$SD).

(d) Identify any measurements that fall outside the control limits.

(e) Calculate the process capability index: $C_p = \frac{\text{USL}-\text{LSL}}{6\sigma}$ where USL = 260g, LSL = 240g.

(f) Calculate the process capability index: $C_{pk} = \min\left(\frac{\text{USL}-\mu}{3\sigma}, \frac{\mu-\text{LSL}}{3\sigma}\right)$

(g) Interpret the capability indices and assess process performance.

(h) Suggest improvements if the process is not meeting specifications.

(i) Calculate what percentage of components are expected to be within specification limits.

7. **[17 marks]** Compare the performance of three production shifts using the following summary statistics:

Shift A: n = 40, mean = 99.2, SD = 2.1, target = 100 **Shift B:** n = 35, mean = 100.8, SD = 1.8, target = 100 **Shift C:** n = 45, mean = 99.8, SD = 2.5, target = 100

Specification limits: 95 x 105

(a) Calculate the bias (difference from target) for each shift.

(b) Determine which shift has the best accuracy (closest to target).

(c) Calculate the coefficient of variation for each shift to assess precision.

(d) Calculate Cp and Cpk for each shift.

(e) Estimate the percentage of output within specifications for each shift (assume normality).

(f) Rank the shifts by overall quality and justify your ranking.

(g) Calculate the combined statistics if all three shifts are merged.

(h) Recommend which shift should be used as the benchmark for improvement.

(i) Suggest specific improvements for the worst-performing shift.

Answer Space

Use this space for your working and answers.

Formulae and Key Concepts

Combined Statistics:

Combined mean: $\bar{x}_{combined} = \frac{n_1\bar{x}_1+n_2\bar{x}_2}{n_1+n_2}$

Combined variance: $s^2_{combined} = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2+n_1(\bar{x}_1-\bar{x}_{combined})^2+n_2(\bar{x}_2-\bar{x}_{combined})^2}{n_1+n_2-1}$

Linear Transformations:

If $y = ax + b$:

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

$s_y = |a| \times s_x$

$CV_y = CV_x$ (coefficient of variation unchanged by scaling)

Moving Averages:

k-point moving average: $MA_t = \frac{x_{t-k+1}+x_{t-k+2}+...+x_t}{k}$

Centered moving average: Places average at center of time period

Number of moving averages: n - k + 1 for k-point MA

Coefficient of Variation:

$CV = \frac{\text{Standard Deviation}}{\text{Mean}} \times 100\%$
Used to compare relative variability across different scales
Lower CV indicates more consistent/precise process

Process Capability Indices:
$C_p = \frac{\text{USL}-\text{LSL}}{6\sigma}$ (measures process spread relative to specification width)
$C_{pk} = \min\left(\frac{\text{USL}-\mu}{3\sigma}, \frac{\mu-\text{LSL}}{3\sigma}\right)$ (accounts for process centering)
Interpretation: Cp, Cpk 1.33 (capable), 1.67 (highly capable)

Control Limits:
Upper Control Limit (UCL): $\mu + 3\sigma$
Lower Control Limit (LCL): $\mu - 3\sigma$
Process in control if all points within these limits

Quality Metrics:
Accuracy: Closeness to target (measured by bias = $|\bar{x} - \text{target}|$)
Precision: Consistency (measured by standard deviation or CV)
Bias: $\bar{x} - \text{target}$ (positive = above target, negative = below)

Normal Distribution Properties:
68% within ±
95% within ± 2
99.7% within ± 3

Mean Absolute Deviation:
$MAD = \frac{\sum|x_i-\bar{x}|}{n}$ (average absolute distance from mean)
Alternative measure of spread, less sensitive to outliers than SD

Seasonal Analysis:
Seasonal indices: Compare period averages to overall average
Deseasonalized data: Remove seasonal patterns to see trend
Seasonal amplitude: Range of seasonal effects

END OF TEST

Total marks: 90

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