A LEVEL MATHEMATICS

STATISTICS YEAR 1/2 – Measures of Location and Spread

NAME _____

Standard deviation

Standard deviation = √(Variance)

Interquartile range = $\mathrm{IQR} = \mathrm{Q}_3 - \mathrm{Q}_1$

For a set of n values $x_1, x_2, \ldots x_i, \ldots x_n$

$$\mathrm{S}_{xx} = \Sigma(x_i - \bar{x})^2 = \Sigma x_i^2 - \frac{(\Sigma x_i)^2}{n}$$

Standard deviation = $\sqrt{\frac{\mathrm{S}_{xx}}{n}}$ or $\sqrt{\frac{\sum x^2}{n} - \bar{x}^2}$

Instructions

- Use **black** ink or ball-point pen.
- If pencil is used for diagrams/sketches/graphs it must be dark (HB or B).
- **Fill in the boxes** at the top of this page with your name, centre number and candidate number.
- Answer **all** questions and ensure that your answers to parts of questions are clearly labelled.
- Answer the questions in the spaces provided – *there may be more space than you need*.
- You should show sufficient working to make your methods clear. Answers without working may not gain full credit.
- Values from statistical tables should be quoted in full. If a calculator is used instead of tables the value should be given to an equivalent degree of accuracy.
- Inexact answers should be given to three significant figures unless otherwise stated.

Question 1

Chris is studying the Daily Total Rainfall, x, for Jacksonville from the large data set for 1987

(a) Using your knowledge of the large data set, write down

(i) the units of x

(ii) the modal value of x

(2)

Chris found the following summary statistics

$$n = 184 \qquad \sum x = 506.8 \qquad \sum x^2 = 10\,750$$

(b) Calculate, giving your answers to 3 significant figures,

(i) the mean of x

(1)

(ii) the standard deviation of x

(2)

Chris believes that the mean Daily Total Rainfall for Jacksonville has increased since 1987

When using the large data set, Chris

- treated the observations from 1987 and 2015 as independent random samples
- used a null hypothesis that there was **no change** in the mean
- carried out a suitable test that had a p-value of 0.0015

(c) Use this information to comment on Chris's belief.
Give a reason for your answer.

(2)

Question 2

Runners in an athletics club can train with either coach A or coach B for the 400 m race.

Coach A trains 120 runners for the 400 m and records the best time, x seconds, for each runner.

The results are summarised by the following statistics

$$\sum x = 6612 \qquad \sum x^2 = 364902$$

(a) Calculate the mean of the best times for the runners trained by coach A

(1)

(b) Calculate the standard deviation of the best times for the runners trained by coach A

(2)

The mean and standard deviation for the best times of the 100 runners trained for the 400 m by coach B are 55.1 seconds and 3.6 seconds respectively.

A 400 m race consists of equal numbers of the fastest runners trained by coach A and by coach B

(c) State, giving a reason, which coach is more likely to have trained the winner.

(2)

Question 3

Ming is studying the large data set for Perth in 2015

He intended to use all the data available to find summary statistics for the Daily Mean Air Temperature, x °C.
Unfortunately, Ming selected an incorrect variable on the spreadsheet.
This incorrect variable gave a mean of 5.3 and a standard deviation of 12.4

(a) Using your knowledge of the large data set, suggest which variable Ming selected.

(1)

The correct values for the Daily Mean Air Temperature are summarised as

$$n = 184 \qquad \sum x = 2801.2 \qquad \sum x^2 = 44\,695.4$$

(b) Calculate the mean and standard deviation for these data.

(3)

One of the months from the large data set for Perth in 2015 has

- mean $\bar{x} = 19.4$
- standard deviation $\sigma_x = 2.83$

for Daily Mean Air Temperature.

(c) Suggest, giving a reason, a month these data may have come from.

(2)

Question 4

Dian uses the large data set to investigate the Daily Total Rainfall, r mm, for Camborne.

(a) Write down how a value of $0 < r \leqslant 0.05$ is recorded in the large data set.

(1)

Dian uses the data for the 31 days of August 2015 for Camborne and calculates the following statistics

$$n = 31 \qquad \sum r = 174.9 \qquad \sum r^2 = 3523.283$$

(b) Use these statistics to calculate

(i) the mean of the Daily Total Rainfall in Camborne for August 2015,

(ii) the standard deviation of the Daily Total Rainfall in Camborne for August 2015.

(3)

Dian believes that the mean Daily Total Rainfall in August is less in the South of the UK than in the North of the UK.
The mean Daily Total Rainfall in Leuchars for August 2015 is 1.72 mm to 2 decimal places.

(c) State, giving a reason, whether this provides evidence to support Dian's belief.

(2)

Dian uses the large data set to estimate the proportion of days with no rain in Camborne for 1987 to be 0.27 to 2 decimal places.

(d) Explain why the distribution B(14, 0.27) might **not** be a reasonable model for the number of days without rain for a 14-day summer event.

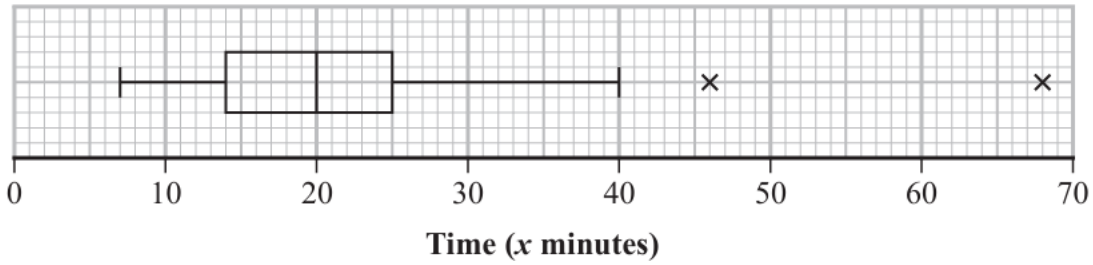
(1)

Question 5

Each member of a group of 27 people was timed when completing a puzzle.

The time taken, x minutes, for each member of the group was recorded.

These times are summarised in the following box and whisker plot.



Time (x minutes)

(a) Find the range of the times.

(1)

(b) Find the interquartile range of the times.

(1)

For these 27 people $\sum x = 607.5$ and $\sum x^2 = 17\,623.25$

(c) calculate the mean time taken to complete the puzzle,

(1)

(d) calculate the standard deviation of the times taken to complete the puzzle.

(2)

Taruni defines an outlier as a value more than 3 standard deviations above the mean.

(e) State how many outliers Taruni would say there are in these data, giving a reason for your answer.

(1)

Adam and Beth also completed the puzzle in a minutes and b minutes respectively, where $a > b$.
When their times are included with the data of the other 27 people

- the median time increases
- the mean time does not change

(f) Suggest a possible value for a and a possible value for b, explaining how your values satisfy the above conditions.

(3)

(g) Without carrying out any further calculations, explain why the standard deviation of all 29 times will be lower than your answer to part (d).

(1)

Question 6

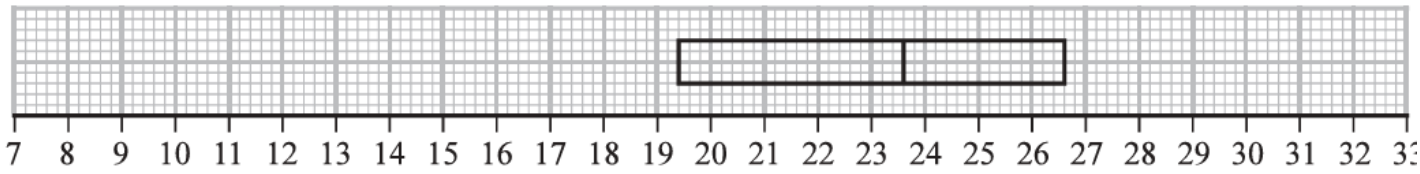


Figure 1

The partially completed box plot in Figure 1 shows the distribution of daily mean air temperatures using the data from the large data set for Beijing in 2015

An outlier is defined as a value
more than 1.5 × IQR below Q_1 or
more than 1.5 × IQR above Q_3

The three lowest air temperatures in the data set are 7.6 °C, 8.1 °C and 9.1 °C
The highest air temperature in the data set is 32.5 °C

(a) Complete the box plot in Figure 1 showing clearly any outliers.

(4)

(b) Using your knowledge of the large data set, suggest from which month the two outliers are likely to have come.

(1)

Using the data from the large data set, Simon produced the following summary statistics for the daily mean air temperature, x °C, for Beijing in 2015

$$n = 184 \qquad \sum x = 4153.6 \qquad S_{xx} = 4952.906$$

(c) Show that, to 3 significant figures, the standard deviation is 5.19 °C

(1)

Simon decides to model the air temperatures with the random variable

$$T \sim N(22.6, 5.19^2)$$

(d) Using Simon's model, calculate the 10th to 90th interpercentile range.

(3)

Simon wants to model another variable from the large data set for Beijing using a normal distribution.

(e) State two variables from the large data set for Beijing that are **not** suitable to be modelled by a normal distribution. Give a reason for each answer.

(2)