Outliers and box plots

118 marks

Hints online, answers free with an account: mathswariz.uk/worksheets/a-level-statistics-outliers-and-box-plots/

Date

A Do Now

Quick questions to start. Some use earlier lessons; some set a slip that costs marks.

A1 Needed today [3 marks]

Find the lower quartile, the median and the upper quartile of the eleven values 3, 7, 8, 10, 12, 13, 15, 17, 18, 22, 35.

Answer .....

A2 Needed today [3 marks]

For eight values, $\sum x = 96$ and $\sum x^2 = 1240$. Find the mean and the standard deviation.

Answer .....

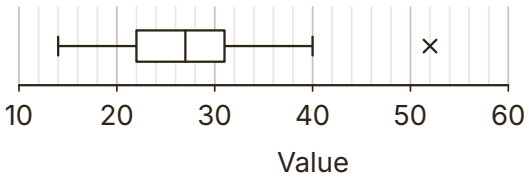
A3 Common slip [2 marks]

A set of data has mean 20 and variance 6.25. An outlier is a value more than 2 standard deviations from the mean. Find the two boundaries.

Answer .....

A4 Common slip [2 marks]

The box plot shows a set of data, with one outlier. Find the range of the data.



Answer .....

© 2026 Mathswariz. All rights reserved. Free for personal and classroom use. Not to be sold or republished.

B Example, then your turn

Your teacher works through each example with you. Then try the one beside it.

EXAMPLE 1

For a set of data, the lower quartile is 34 and the upper quartile is 46. An outlier is a value less than $Q_1 - 1.5(Q_3 - Q_1)$ or greater than $Q_3 + 1.5(Q_3 - Q_1)$. The smallest value is 12 and the largest is 70. Decide whether each of these is an outlier.

Hint Find the interquartile range first, then multiply by 1.5. Take it off Q_1 and add it to Q_3.

B1 [4 marks]

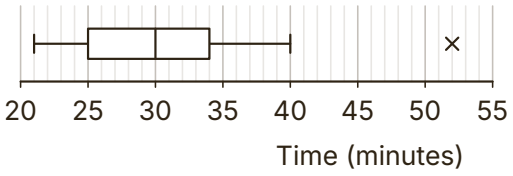
For the masses, in grams, of a sample of apples, $Q_1 = 140$ and $Q_3 = 164$. An outlier is a value more than $1 \times (Q_3 - Q_1)$ below Q_1 or above Q_3. The three smallest masses are 110, 116 and 121; the largest is 190. Find the boundaries and state which masses are outliers.

Answer .....

© 2026 Mathswariz. All rights reserved. Free for personal and classroom use. Not to be sold or republished.

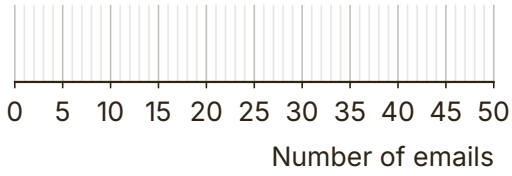
EXAMPLE 2

Fifteen students recorded the time, in minutes, they spent on homework one evening: 21, 24, 25, 25, 27, 28, 29, 30, 31, 31, 32, 34, 35, 40, 52. An outlier is a value more than $1.5 \times \text{IQR}$ below Q_1 or above Q_3. Find the quartiles and any outliers, and draw a box plot. The finished box plot is shown.



Hint $n = 15$: $\frac{15}{4} = 3.75$, so round up to the 4th value. The upper whisker stops at the largest value that is not an outlier.

B2 Nineteen people recorded the number of emails they received in a day: 2, 18, 20, 22, 23, 24, 25, 26, 26, 27, 28, 29, 30, 30, 31, 32, 33, 35, 47. An outlier is a value more than $1.5 \times \text{IQR}$ below Q_1 or above Q_3. Find the quartiles and any outliers, and draw a box plot on the grid. [5 marks]



Answer .....

© 2026 Mathswariz. All rights reserved. Free for personal and classroom use. Not to be sold or republished.

EXAMPLE 3

The heights, x cm, of ten plants give $\sum x = 1520$ and $\sum x^2 = 233\,290$. An outlier is a value more than 2 standard deviations from the mean. The shortest plant is 118 cm and the tallest is 190 cm. Are these outliers?

Hint Find the mean and the variance from the sums, then take the square root to get the standard deviation before you use it.

B3 The ages, x years, of 20 members of a club give $\sum x = 1100$ and $\sum x^2 = 61\,220$. An outlier is a value more than 2 standard deviations from the mean. The youngest member is 41 and the oldest is 66. Find the boundaries and decide whether each age is an outlier. [4 marks]

Answer .....

© 2026 Mathswariz. All rights reserved. Free for personal and classroom use. Not to be sold or republished.

EXAMPLE 4

A midwife records the masses, in kg, of twelve newborn babies: 3.1, 3.4, 2.9, 3.6, 3.2, 3.8, 3.0, 3.5, 2.8, 3.3, 3.6, 34.0. Explain why one value should be removed before the data are used, and find the mean and the standard deviation before and after it is removed.

Hint A newborn baby of 34 kg is not possible: this is a recording error. Use $\sum x$ and $\sum x^2$ for each set.

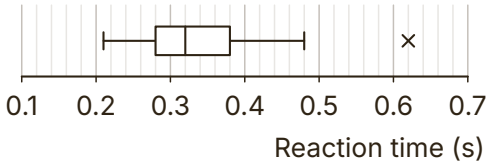
B4 [5 marks]

A student records the daily maximum temperature, in $^\circ\text{C}$, on ten days in July: 18.2, 19.5, 19.9, 20.1, 21.4, 22.0, 22.8, 23.5, 24.1, 251. Explain why 251 is an error, and not a real outlier. It should have been 25.1. Find the mean and the standard deviation of the ten corrected temperatures.

© 2026 Mathswariz. All rights reserved. Free for personal and classroom use. Not to be sold or republished.

EXAMPLE 5

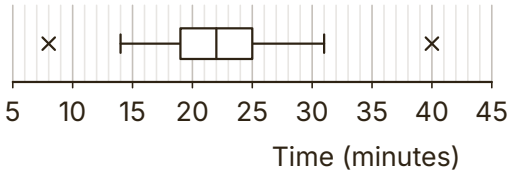
The box plot shows the reaction times, in seconds, of 40 drivers. An outlier is a value more than $1.5 \times \mathrm{IQR}$ above Q_3 or below Q_1. (a) Find the range and the interquartile range. (b) Find the upper boundary for outliers, and explain what the end of the upper whisker tells you. (c) Estimate how many drivers had a reaction time greater than 0.38 s.



Hint The range uses the outlier. The whisker stops at the largest value that is not an outlier. A quarter of the data lie above Q_3.

B5 [5 marks]

The box plot shows the times, in minutes, taken by 60 runners to finish a course. An outlier is a value more than $1.5 \times \mathrm{IQR}$ above Q_3 or below Q_1. Find the range and the interquartile range, show that both crosses are outliers, and estimate the number of runners who took more than 25 minutes.



Answer .....

© 2026 Mathswariz. All rights reserved. Free for personal and classroom use. Not to be sold or republished.

C1 A set of data has mean 32 and variance 16. An outlier is a value more than 2 standard deviations from the mean. Ali writes: "The boundaries are $32 - 2(16) = 0$ and $32 + 2(16) = 64$, so the value 21 is not an outlier." Explain Ali's mistake, and decide correctly whether 21 is an outlier. [4 marks]

© 2026 Mathswariz. All rights reserved. Free for personal and classroom use. Not to be sold or republished.

C2 Jake has the eleven values 4, 15, 17, 18, 20, 22, 23, 25, 26, 32, 41. An outlier is a value more than $1.5 \times \mathrm{IQR}$ below Q_1 or above Q_3. Jake says: "4 and 41 are both a long way from the rest, so they are both outliers." Find the boundaries and state which values are outliers. [5 marks]

© 2026 Mathswariz. All rights reserved. Free for personal and classroom use. Not to be sold or republished.

C3 For a set of data the outlier boundaries $Q_1 - 1.5(Q_3 - Q_1)$ and $Q_3 + 1.5(Q_3 - Q_1)$ are 6 and 54. Find the interquartile range and the two quartiles. [4 marks]

Answer .....

C4 A sample from the large data set gives the daily mean temperature at Beijing on each day from May to October. A box plot of the sample has three low outliers. Sam says: "These must be errors, so remove them." Give the month in which the low outliers were most likely recorded, with a reason, and explain whether Sam is right. [4 marks]

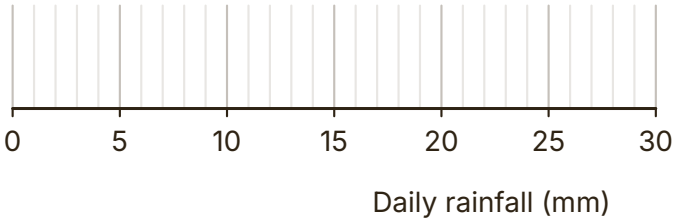
© 2026 Mathswariz. All rights reserved. Free for personal and classroom use. Not to be sold or republished.

C5 For a set of data $Q_1 = 18$, $Q_2 = 23$ and $Q_3 = 27$. The smallest value is 12 and the two largest values are 36 and 49. An outlier is a value more than $1.5 \times \mathrm{IQR}$ above Q_3 or below Q_1. Mia draws the upper whisker of the box plot to 40.5. Explain her mistake, and say where the upper whisker should end and how 49 is shown. [4 marks]

© 2026 Mathswariz. All rights reserved. Free for personal and classroom use. Not to be sold or republished.

D1 A sample of 23 values from the large data set gives the daily rainfall, in mm, at one weather station, in order: tr, tr, tr, 0, 0.2, 0.4, 0.6, 0.8, 1.0, 1.4, 1.8, 2.2, 2.6, 3.0, 3.6, 4.2, 4.8, 5.2, 6.4, 7.6, 8.8, 19.6, 27.2. An outlier is a value more than $1.5 \times \mathrm{IQR}$ below Q_1 or above Q_3.

(a) State what "tr" means in the large data set, and the value used for it in calculations. [2 marks]



(b) Find the median and the quartiles. [3 marks]

Answer .....

© 2026 Mathswariz. All rights reserved. Free for personal and classroom use. Not to be sold or republished.

(c) Find the outliers. [3 marks]

Answer .....

(d) On the grid, draw a box plot for these data. [2 marks]

Answer .....

(e) A student says the two outliers should be removed before the data are used. Comment on this. [2 marks]

Total for D1: 12 marks

© 2026 Mathswariz. All rights reserved. Free for personal and classroom use. Not to be sold or republished.

D2

For 25 values, $\sum x = 1250$ and $\sum x^2 = 63\,400$. Tom writes: "Mean $= 50$. Variance $= \dfrac{63\,400}{25} - 50 = 2486$, so the standard deviation is 49.9, and no value can be an outlier by the mean ± 2 standard deviations rule."

Tom is wrong.

(a) Explain Tom's mistake and find the standard deviation. [3 marks]

(b) An outlier is a value more than 2 standard deviations from the mean. The smallest value is 35. Find the boundaries and decide whether 35 is an outlier. [3 marks]

Answer .....

© 2026 Mathswariz. All rights reserved. Free for personal and classroom use. Not to be sold or republished.

(c) It is found that 35 was recorded in error, and the value should be 53. Find the mean and the standard deviation of the corrected data. [4 marks]

Answer .....

Total for D2: 10 marks

© 2026 Mathswariz. All rights reserved. Free for personal and classroom use. Not to be sold or republished.

D3 The times, x minutes, for 20 journeys give $\sum x = 858$ and $\sum x^2 = 40\,479$. One time, 98 minutes, is found to be an error and is removed. Find the mean and the standard deviation of the remaining 19 times. An outlier is a value more than 2 standard deviations from the mean of these 19 times. Decide whether a further journey of 54 minutes would be an outlier. [7 marks]

Answer .....

© 2026 Mathswariz. All rights reserved. Free for personal and classroom use. Not to be sold or republished.

D4 Eleven values are 12, 14, 15, 17, 19, 20, 22, 24, 28, 30, 41. An outlier is a value more than $1.5 \times \mathrm{IQR}$ below Q_1 or above Q_3.

(a) Find the mean and the median. [2 marks]

Answer .....

(b) Find the quartiles, and show that there are no outliers. [3 marks]

Answer .....

© 2026 Mathswariz. All rights reserved. Free for personal and classroom use. Not to be sold or republished.

(c) Two more values, a and b with $a \leqslant b$, are added. The mean of the thirteen values is still 22, and the median increases. Find the set of possible values of a. [3 marks]

Answer .....

(d) Given that $a = 21$, show that 41 is now an outlier. [3 marks]

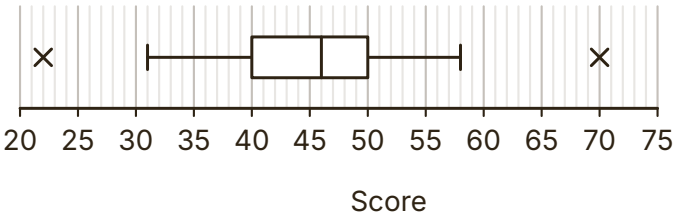
Answer .....

Total for D4: 11 marks

© 2026 Mathswariz. All rights reserved. Free for personal and classroom use. Not to be sold or republished.

D5 The box plot shows the scores of 120 students in a test. An outlier is a value more than $1.5 \times \text{IQR}$ below Q_1 or above Q_3.

(a) Show that 22 and 70 are outliers. [3 marks]



Answer .....

(b) Find the range and estimate the number of students who scored between 40 and 50. [2 marks]

Answer .....

© 2026 Mathswariz. All rights reserved. Free for personal and classroom use. Not to be sold or republished.

(c) The teacher says the score of 70 is an outlier, so it should be removed. Comment on the teacher's statement. [2 marks]

(d) It is found that the score of 22 should have been 32. Describe how the box plot changes. [2 marks]

Total for D5: 9 marks

© 2026 Mathswariz. All rights reserved. Free for personal and classroom use. Not to be sold or republished.

D6 For a set of data, $Q_1 = 40$ and $Q_3 = 60$. An outlier is a value more than $k \times \mathrm{IQR}$ below Q_1 or above Q_3. The smallest value is 12 and the largest is 96.

(a) Taking $k = 1.5$, decide which of 12 and 96 are outliers. [2 marks]

Answer .....

(b) Find the set of values of k for which exactly one of 12 and 96 is an outlier. [4 marks]

Answer .....

Total for D6: 6 marks

© 2026 Mathswariz. All rights reserved. Free for personal and classroom use. Not to be sold or republished.

E Extension

Stretch

No route is given. The hints are at the end of the sheet. Use one at a time.

E1 Nine values have mean 10 and variance 4. A tenth value, x, is added. An outlier is a value more than 2 standard deviations from the mean, where the mean and the standard deviation are those of all ten values. Find the set of values of x for which x is an outlier. Give the boundaries in surd form. [9 marks]

© 2026 Mathswariz. All rights reserved. Free for personal and classroom use. Not to be sold or republished.

Extension hints

Cover these with paper. Uncover one at a time.

E1

Hint 1 For the nine values, $\sum x = 90$ and $\sum x^2 = 9(4 + 10^2) = 936$.

Hint 2 Write the mean of the ten values, $\dfrac{90 + x}{10}$, and their variance, $\dfrac{936 + x^2}{10} - \left(\dfrac{90 + x}{10}\right)^2$, in terms of x.

Hint 3 The condition is $(x - \text{mean})^2 > 4 \times \text{variance}$. Multiply out and you get a quadratic inequality in x.

How did it go?

Colour one circle on each line: not yet, nearly, yes.

I can find the outlier boundaries from the quartiles or from the mean and standard deviation, and say which values are outliers.	○ ○ ○
I can decide whether an outlier is an error to remove or a real value to keep, and recompute the mean and standard deviation.	○ ○ ○
I can draw a box plot with its outliers, and read the range, the interquartile range and proportions from one.	○ ○ ○
I can find the quartiles, or a range of values of k, from given outlier boundaries.	○ ○ ○

© 2026 Mathswariz. All rights reserved. Free for personal and classroom use. Not to be sold or republished.