A Level · Edexcel · Maths

🕐 3 hours   ❓ 36 questions

2.1 Statistical Measures

2.1.1 Basic Statistical Measures / 2.1.2 Frequency Tables / 2.1.3 Standard Deviation & Variance / 2.1.4 Coding

Easy (9 questions)	/46
Medium (9 questions)	/43
Hard (9 questions)	/54
Very Hard (9 questions)	/57
Total Marks	**/200**

Scan here to return to the course
or visit savemyexams.com

Easy Questions

1 (a) Students' marks, given as a percentage, on their recent statistics test were:

38   41   19   33   22   0   27   19   10   99

Find the mode, range, mean and median of the students' marks.

(4 marks)

(b) Give a reason why the median is an appropriate measure of location for these data.

(1 mark)

2 (a) Two sets of data are given below:

set 1	1	2	3	4	5	6	7	8	9
set 2	1	5	5	5	5	5	5	5	9

For set 1,

(i) Calculate the mean, $\bar{x}$, of the data.

(ii) Calculate the variance, σ^2, of the data using the formula:

$$\sigma^2 = \frac{\Sigma(x - \bar{x})^2}{n}$$

(3 marks)

(b) For set 2,

(i) Calculate the mean, $\bar{x}$, of the data.

(ii) Calculate the variance, σ^2, of the data using the formula:

$$\sigma^2 = \frac{\Sigma x^2}{n} - \bar{x}^2$$

(3 marks)

3 (a) Seven friends decide to see how long they can hold their breath underwater. Their times, in seconds, are shown below.

59   72   69   105   77   81   92

Write down the

(i)   lower quartile, Q_1,

(ii)   median, Q_2,

(iii)   and the upper quartile, Q_3,

of the data.

(3 marks)

(b) One more friend comes along and decides to join in. He holds his breath for 85 seconds. Comment on how this 8th value will affect your answers in part (a).

(2 marks)

4 (a) Lucy is working with some grouped, continuous data. For a set of 100 items of data, she has calculated that $\Sigma xf = 357$ and $\Sigma(x - \bar{x})^2 f = 42$, where f is the frequency for each group.

(i) Give a reason why Lucy has decided to group her data.

(ii) Briefly explain what is meant by Σxf and $\Sigma(x - \bar{x})^2 f$.

(3 marks)

(b) To calculate the standard deviation, Lucy could choose to use either of the following two formulae.

$$\text{standard deviation} = \sqrt{\frac{\Sigma(x - \bar{x})^2 f}{\Sigma f}} \text{ or } \sqrt{\frac{\Sigma x^2 f}{\Sigma f} - \bar{x}^2} \text{ where } \bar{x} = \frac{\Sigma xf}{\Sigma f}$$

(i) Calculate the mean of Lucy's data.

(ii) Using the appropriate formula from above, calculate the standard deviation of Lucy's data.

(3 marks)

5 As part of her veterinary course, Harriet measured the weight, x grams, of 50 new-born kittens and summarised their data as $\Sigma x = 6342$ and $\Sigma x^2 = 879013$.

Calculate the mean and standard deviation of the weights of the kittens.

6 (a) Katie is collecting information on Jupiter's moons for a research project. She collects data on the diameters of 78 of Jupiter's known moons and organises the information into the table below.

Diameter d (km)	Number of moons f
$0 < d \leq 1$	6
$1 < d \leq 2$	20
$2 < d \leq 5$	23
$5 < d \leq 50$	17
$50 < d \leq 1000$	8
$1000 < d \leq 6000$	4

(i) Write down the modal class interval.

(ii) Write down the class interval that contains the median.

(iii) Katie discovers another moon, Valetudo, which has a diameter of 1 km. Write down the class interval which should include the diameter of Valetudo.

(3 marks)

(b) Katie calculates the mean diameter of Jupiter's moons to be 6500 km. Explain how you know Katie is incorrect.

(1 mark)

7 (a) The number of goals scored by the 24 teams that played in the first 44 games of the UEFA Euro cup 2020 can be summarised in the table below.

Goals scored	0 - 1	2 - 3	4 - 5	6 - 7	8 - 9	10 - 11
Frequency f	3	5	5	6	4	1

Estimate the mean number of goals scored by each team.

(2 marks)

(b) Find the standard deviation of the number of goals scored by each team.

(2 marks)

8 (a) The Mythical Creatures Research Centre measures the heights, h, of nine unicorns to the nearest centimetre. The heights are shown below:

$$276 \quad 219 \quad 198 \quad 154 \quad 213 \quad 243 \quad 192 \quad 161 \quad 218$$

Use your calculator to find the mean and standard deviation of the nine heights.

(2 marks)

(b) Before calculating the mean, the researchers choose to form a new variable, y, using the formula $y=h$-200.

(i) Write down the nine values of y.

(ii) Use your calculator to find the mean and standard deviation of the nine values of y.

(3 marks)

(c) By comparing your answers to (a) and (b)(ii), describe how subtracting a value from each piece of data affects the mean and standard deviation.

(2 marks)

9 (a) A selection of data from the large data set relating to the daily mean cloud cover, measured in oktas, in Heathrow for the first 10 days in May 1987 is given below.

$$7 \quad 4 \quad 5 \quad 2 \quad 7 \quad 4 \quad 2 \quad 0 \quad 3 \quad 5$$

Using your knowledge of the large data set, explain why a value of 10 oktas would be impossible.

(1 mark)

(b) (i) Find the value of the median of the data.

(ii) Find the value of the lower quartile of the data.

(iii) Calculate the interquartile range of the data.

(4 marks)

Medium Questions

1 A pharmacy sells face masks in a variety of sizes. Their sales over a week are recorded in the table below:

Size	Kids		Adults			
	Small	Large	S	M	L	XL
Frequency f	29	4	8	24	15	4

(i)   Write down the mode for this data.

(ii) Explain why, in this case, the mode from part (i) would not be particularly helpful to the shop owner when reordering masks.

(iii) Given that the shop is open every day of the week, calculate the mean number of masks sold per day.

(4 marks)

2 Fran sits three Maths papers and six Science papers during her final A Level exams. She achieves a mean score of 62.7% across the three Maths exam papers, and needs an overall mean score of 78.5% across all nine papers to get into her chosen University. After getting the results of four out of her six Science papers, her mean score in Science is 84.2%.

Given that each of the nine papers is weighted equally when working out the mean scores, calculate the mean score she must achieve on her final two science papers in order to gain a place at University.

(4 marks)

3 (a) Coffee4Life manufactures reusable coffee cups out of coffee plant waste. Coffee cups are tested to see how many times they can be used before they begin to disintegrate. A sample of 15 cups are tested, giving the following results for numbers of uses:

$$31 \quad 36 \quad 41 \quad 43 \quad 47$$

$$49 \quad 51 \quad 56 \quad 58 \quad 62$$

$$62 \quad 63 \quad 68 \quad 69 \quad 72$$

(i) Write down the modal number of times a cup can be used.

(ii) Find the values of the lower quartile, median and upper quartile.

(4 marks)

(b) The advertising department at Coffee4Life designs an advert which says;

"If used once a day, $\dfrac{3}{4}$ of our cups last longer than 9 weeks."

Explain the mistake that the advertising department has made, and state how the advert could be reworded to make it correct.

(2 marks)

4 The lengths (l cm) of a sample of nine otters, measured to the nearest centimetre by a wildlife research team, are:

$$76 \quad 77 \quad 91 \quad 65 \quad 63 \quad 83 \quad 92 \quad 61 \quad 88$$

Calculate the mean and standard deviation of the nine recorded lengths.

(3 marks)

5 (a) A machine is set to fill sacks of potatoes to a weight of 50 kg. In a random sample the following weights in kg were recorded:

$$53.4 \quad 46.8 \quad 47.2 \quad 42.4 \quad 43.0 \quad 47.4 \quad 55.2$$

Code the weights using the coding $y = 0.5$ *(weight in kg-40)*

(2 marks)

(b) (i)   Calculate the mean of the coded data values.

(ii)   Use your answer to part (i) to show that the mean of the original data is less than 50 kg.

(4 marks)

6 (a) During initial training for the Royal Air Force new recruits must sit an aptitude test. Test scores for the latest round of recruits are shown in the table below:

Score	Frequency f
0 – 154	5
155-199	6
200-234	12
235-260	5

Recruits who score below the 25th percentile are disqualified.

Calculate an estimate for the score recruits must have achieved to avoid disqualification.

(2 marks)

(b) Those who score in the top 30% move on to the next stage of training and the rest must re-sit the test.

One of the recruits, Amelia, achieves a score of 231. Estimate whether Amelia will need to re-sit the test or will be moved on to the next stage of training.

(3 marks)

7 (a) A random sample of 50 students were asked how long they spent revising for their Maths exam in the 24 hours before the exam. The results are shown in the table below:

Time t (minutes)	Number of students f
$0 \le t < 60$	5
$60 \le t < 120$	6
$120 \le t < 180$	17
$180 \le t < 240$	14
$240 \le t < 300$	8

For this data, use linear interpolation to estimate the median.

(2 marks)

(b) Using x to represent the mid-point of each class, Σfx = 8340 and Σfx^2 = 1 636 200.

Estimate the mean and the standard deviation of the amount of time students spent revising.

(3 marks)

8 A college needs to standardise the test scores of all students before submitting them to the Exam Board. The scores are standardised by using the coding $y = \dfrac{x + 53}{200}$, where x represents the raw test score and y represents the standardised score. The college calculates the mean standardised test score to be 0.74. Find the mean of the raw test scores.

(2 marks)

Daily Mean Temp. °C Beijing October 1987	20.6	19.1	21.1	20.4	19.8	19.3	17.1	16.5	18	18.9
Daily Mean Temp. °C Beijing October 2015	16.1	19.4	18.6	18.4	18.9	20.3	20.5	14.5	14.7	14

9 (a)

A selection of data from the large data set relating to the mean daily air temperature in Beijing for the first 10 days in October in both 1987 and 2015 is given above. Climate activists use temperature data to track changes over time.

Using the data given above, find the mean of the daily mean air temperature for both 1987 and 2015.

(4 marks)

(b) Climate activists claim that temperature changes of more than 0.5°C each decade would indicate significant evidence of irreparable climate change damage.

(i)   Use your answers to part (a) to comment on the activists' claim.

(ii)   Give one reason why the sample used above should not be used to draw wider conclusions and suggest how it could be improved.

(4 marks)

Hard Questions

1 a, b, c and d are 4 integers written in order of size, starting with the smallest.

The sum of a, b and c is 70
The mean of a, b, c and d is 25
The range of the 4 integers is 14.

Work out the median of a, b, c and d

(4 marks)

2 (a) The speeds (*s*), to the nearest mile per hour, of 80 vehicles passing a speed camera were recorded and are grouped in the table below.

Speed, *s* (mph)	$20 \leq s < 25$	$25 \leq s < 30$	$30 \leq s < 35$	$s \geq 35$
Number of vehicles	23	48	7	2

(i) Write down the modal class for this data.

(ii) Write down the class group that contains the median.

(2 marks)

(b) (i) Assuming that ≥35 means 'at least 35 mph but less than 40 mph', calculate an estimate for the mean speed of the 80 vehicles.

(ii) It is now discovered that ≥35 means 'at least 35 mph but less than 60 mph'. Without further calculation, state with a reason whether this would cause an increase, a decrease or no change to the value of the estimated mean.

(4 marks)

3 (a) A veterinary nurse records the weight of puppies (in kg) at birth and again at their eight week check-up. The table below summarises the weight gain of 50 small breed puppies over their first eight weeks.

Weight gain w (kg)	Number of puppies f
$0.0 \leq w < 0.5$	1
$0.5 \leq w < 1.0$	8
$1.0 \leq w < 1.5$	19
$1.5 \leq w < 2.0$	18
$2.0 \leq w < 2.5$	4

Use linear interpolation to estimate the median and interquartile range of the weight gain of the 50 puppies.

(4 marks)

(b) Give a reason why it is not possible to determine the exact median for this data.

(1 mark)

(c) The veterinary nurse decides to monitor any puppies whose weight gain during their first 8 weeks was less than 0.8 kg.

Estimate the number of puppies whose weight gain could be below 0.8 kg, and explain why the vet would need more information before determining for certain how many puppies would need to be monitored.

(2 marks)

4 (a) Workers at an elephant sanctuary measure the health of their elephants by weighing the amount of dung (d kg) each one produces. The data for the mass of dung produced in one day by 18 elephants can be summarised as $\Sigma d = 895$ and $\Sigma d^2 = 45\,810$.

Calculate the mean and variance of the amount of dung produced that day.

(3 marks)

(b) To calculate how much food (f kg) to give each elephant, staff use the formula $f = 3d - 25$.

Calculate the mean and variance of the amount of food the workers should give to the elephants each day.

(2 marks)

5 (a) A machine is set to fill sacks of potatoes to a target weight of 50 kg, although the actual weight of the sacks (w kg) can vary from that target.

To test the accuracy of the machine, a random sample of 20 sacks is taken and the values of $y=(w$-50$)$ are recorded.

The mean and standard deviation of y are found to be -1.8 and 3.1 respectively.

Write down the mean and standard deviation of w.

(2 marks)

(b) Calculate the value of

(i) Σw

(ii) Σw^2

(3 marks)

(c) Another 10 sacks of potatoes are sampled and the mean of these is found to be 51.2 kg. Calculate the mean of all 30 sacks of potatoes and comment on the accuracy of the machine.

(3 marks)

6 Whilst in lockdown, 100 people were asked to record the length of time, rounded to the nearest minute, that they spent exercising on a particular day.

The results are summarised in the table below:

Time *mins*	Frequency f
$0 \leq t \leq 10$	1
$10 < t \leq 20$	12
$20 < t \leq 30$	25
$30 < t \leq 40$	a
$40 < t \leq 50$	b
$50 < t \leq 60$	14

Given that the estimate of the mean time spent exercising based on this table is 35.4 minutes, find the values of a and b.

(4 marks)

7 (a) The ages, x years, of 200 people attending a vaccination clinic in one day are summarised by the following: $\Sigma x = 7211$ and $\Sigma x^2 = 275\,360$.

Calculate the mean and standard deviation of the ages of the people attending the clinic that day.

(3 marks)

(b) One person chooses not to get the vaccine, so their data is discounted. The new mean is exactly 36. Calculate the age of the person who left and the standard deviation of the remaining people.

(4 marks)

8 (a) In January of 2021, the UK government announced a nationwide lockdown to control the spread of the coronavirus. The table below shows the means and standard deviations of the average amounts of time spent indoors per day by some people in London, UK and in Wellington, New Zealand, in January of 2021.

	Number of people	Mean (hr)	Standard deviation
London	25	20.9	1.51
Wellington	15	15.1	2.87

Suggest a reason, in the context of the question, for why

(i)   the mean in London is higher than the mean in Wellington

(ii)   the standard deviation in London is lower than the standard deviation in Wellington.

(2 marks)

(b) Based on the data in the table, do you think the government in New Zealand had imposed the same restrictions as those in the UK? Give a reason for your answer.

(1 mark)

(c) Calculate the overall mean for the average amounts of time spent indoors by all 40 people.

(2 marks)

9 (a) Data on the daily maximum gust in Camborne is gathered from the large data set. For the third week of October 1987, and the same week in 2015, the daily maximum gust in knots (kn) is shown in the table below:

Daily maximum gust, kn Camborne 15th – 21st October 1987	53	59	55	54	36	14	29
Daily maximum gust, kn Camborne 15th – 21st October 2015	15	16	17	13	14	14	24

Using the data given above, find the mean and standard deviation of the daily maximum gust for Camborne for the third week of October 1987.

(4 marks)

(b) Without carrying out any calculations, state whether you think the mean and standard deviation for the 2015 data would be greater or less than your answers for part (a). Use your knowledge of the large data set to suggest a reason for each of your answers.

(4 marks)

Very Hard Questions

1 Whilst in lockdown, a group of people were asked to record the length of time, t hours, they spent browsing the internet on a particular day.

The results are summarised in the table below.

Time, t (hours)	Frequency, f
t ≤ 2	3
2 < t ≤ 4	5
4 < t ≤ 6	a
6 < t ≤ 8	10
8 < t ≤10	2

From this data an A Level Statistics student calculated that the estimated mean time spent browsing the internet is 5 hours and 15 minutes. Calculate the value of a and find the estimated variance of the length of time spent browsing the internet.

(4 marks)

2 (a) Zisien measures the speeds, x miles per hour, of a number of cars passing her house one day. She knows that the speed limit is 30 miles per hour so she decides to use the coding $y = x - 30$ when she records the data.

She finds that $\Sigma y = 13.4$ and $\Sigma y^2 = 1470$.

Is it more likely that more of the cars in Zisien's sample were going over or under the speed limit that day? Give a reason for your answer.

(2 marks)

(b) Given that the mean speed of the cars in Zisien's sample, $\bar{x}$, is 30.67 miles per hour, calculate the standard deviation of x.

(3 marks)

(c) Zisien's sister, Ying, used the code $z = x - 20$ to record the data for the same cars.

    (i)   Calculate the value of Σz.

    (ii)   Ying decides to work out the median of her coded data and discovers that it is 9.4. Use this information to comment on your answer to part (a).

(2 marks)

3 Two friends, Anna and Connor, are playing a gaming app on their phones. As they play, they can choose from three different booster options. They are unaware that each of the three options are charging them automatically from their mobile accounts. The number of in-app purchases they each make are shown in the table below.

	Super-charge	Re-energise	Level-up
Anna	4	0	2
Connor	3	6	1

(i) The mean and standard deviation of the cost of Anna's in-app purchases are £0.50 and £0 respectively. Write down the cost of a single in-app purchase to 'Level-up'.

(ii) Given that the mean cost of Connor's in-app purchases is £0.38, find the standard deviation of the costs of Connor's purchases.

(5 marks)

4 (a) Botanists at a research centre are carrying out research on a new type of fertiliser. They collect data on the heights of one group of geraniums growing without the fertiliser (the control group) and of another group growing with the fertiliser (the experimental group). They take care to keep all other growing conditions the same for both groups.

The table below shows the heights of the control group of geraniums 15 weeks after planting.

Height (cm)	< 10	< 15	< 20	< 25	< 30	< 35	< 40
Cumulative Frequency	2	7	12	19	34	39	40

(i)   Write down the modal class for the heights of geraniums in the control group.

(ii)   Find the smallest and largest possible values for the interquartile range of the heights of the control group.

(iii) Use linear interpolation to calculate the median height.

(6 marks)

(b) The data for the group of geraniums growing in the experimental group were summarised as follows:

$$Q_1 = 23.4 \text{ cm} \qquad Q_2 = 27.1 \text{ cm} \qquad Q_3 = 28.5 \text{ cm}$$

The shortest plant in the experimental group was 15.2 cm and the tallest was 33.5 cm.

Compare the distribution of the heights of the plants in the two groups.

(2 marks)

5 (a) Wildlife researchers are studying the swimming speeds, x kmph, of two species of penguin, the emperor penguin and the gentoo penguin. The mean swimming speed of 40 gentoo penguins was found to be 31.4 kmph and the standard deviation was found to be 3.8 kmph.

Allowing x_G to represent the swimming speeds of the gentoo penguins, show that Σx_G =1256 and calculate the value of $\Sigma x_G{}^2$.

(3 marks)

(b) The swimming speeds of 20 emperor penguins (x_E) were also recorded and the mean swimming speed of all 60 penguins surveyed was found to be 24.1 kmph. Given that Σx^2=41891, calculate the mean and standard deviation of the 20 emperor penguins.

(4 marks)

6 (a) Some entomologists were studying the amount of time two different species of butterflies spent cocooned. The table shows the means and standard deviations of the time spent cocooned, measured in days, by 15 Monarch butterflies and 25 Common Blue butterflies.

Species	Mean	Standard deviation
Monarch		1.51
Common Blue	13.4	1.24

Given that the overall mean time for all 40 butterflies was 11.93 days, calculate the mean number of days the Monarch butterflies spent cocooned and complete the table.

(3 marks)

(b) Calculate the overall standard deviation of the time spent cocooned by all 40 butterflies.

(4 marks)

7 (a) Lab technicians were studying the effect of caffeine on mice. The resting heart rates, x beats per minute (bpm), of some mice were recorded and the results were summarised by $\Sigma(x-a) = 150$ and $\Sigma(x-a)^2 = 1050$, where a is a constant.

Given that the variance of the resting heart rates was found to be 10 bpm², calculate the two possible options for the number of mice in the study.

(4 marks)

(b) The mean resting heart rate is found to be 605 bpm. Using this information, write down the two possible options for the value of a.

(2 marks)

8 (a) Hattie's homeroom teacher decides to summarise the number of minutes, *t*, she has been late to school during the last year in preparation for a parents' meeting. The results are shown in the table below.

Time t (mins)	Frequency f
-10 ≤ t < -5	3
-5 ≤ t < 0	19
0 ≤ t < 5	32
5 ≤ t < 10	a
10 ≤ t < 20	53
20 ≤ t < 60	24

Write down, in the context of the question, what the time interval -10 ≤ t < -5 represents.

(1 mark)

(b) **(i)** Using x to represent the mid-point of each class, write an expression in terms of a for Σfx, giving your answer in simplified form.

(ii) Given that Σfx = 2132.5 and Σfx^2 = 53568.75, calculate the estimated mean and standard deviation of the amount of time Hattie was late for school last year.

(3 marks)

(c) Hattie happens to notice that on three of the days she was recorded as being 40 minutes late for school, she had actually arrived 40 minutes early. Calculate the corrected estimate for the mean amount of time she was late for school last year.

(2 marks)

9 Roger has been looking at some data on the daily mean air temperature, t, in two different locations, Perth and Jacksonville, taken from the large data set. All the data is taken from the month of July in 2015.

	n	Σt	Σt^2	t	σ
Location A	31	836.3	22593.0		
Location B	31			13.3	2.167

Unfortunately, some of the information has been lost and Roger does not know which data is for which location.

(i) Complete the table.

(ii) Using your knowledge of the large data set, state which of the locations is most likely to be Jacksonville, giving a reason for your answer.

(7 marks)