Overview (/study/app/math-aa-hl/sid-134-cid-761926/o

Teacher view

0 (https://intercom.help/kognity)

Index

Table of contents

Notebook

Glossary

Reading assistance

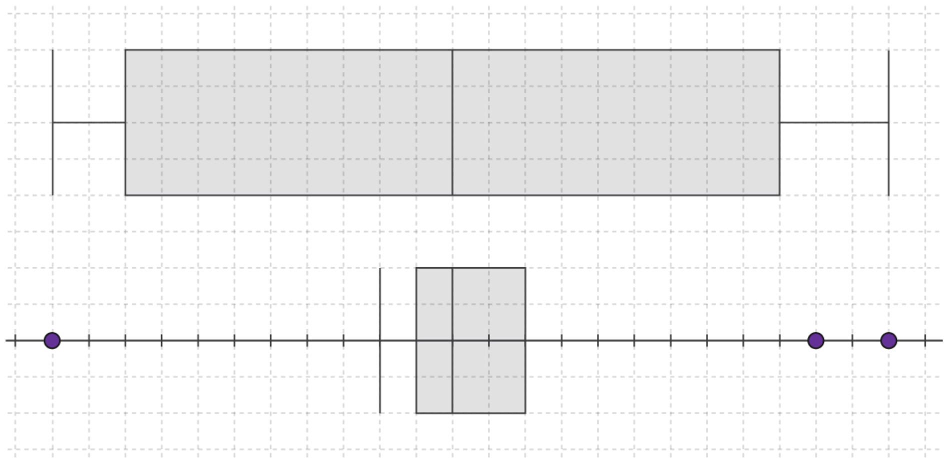
4. Probability and statistics / 4.3 Measures of central tendency and dispersion

The big picture

Consider the following two sets of data and their box-and-whisker plots:

Set A: $\{3, 4, 5, 13, 13, 14, 18, 21, 23, 25, 26\}$

Set B: $\{3, 12, 13, 13, 13, 14, 15, 16, 16, 24, 26\}$



Box-and-whisker plots for Set A (upper) and Set B (lower).

More information

The image shows two box-and-whisker plots against a grid background. The upper plot represents Set A, illustrating a wide spread with a central box that describes the interquartile range. The median is shown as a line inside the box. Whiskers extend from the box to indicate the full range of the data, with outliers marked as points beyond the whiskers. The lower plot represents Set B, showing a narrower interquartile range compared to Set A. Its median is similarly marked, with smaller whiskers indicative of data spread. Both plots allow visual comparison of data distribution and range between the two datasets.

[Generated by AI]

Student view

Overview
(/study/ap
aa-
hl/sid-
134-
cid-
761926/o

When you look at the two sets, and especially their respective box-and-whisker plots, it appears that they have little in common. However, when you calculate various statistics for the sets, you see a very different story:

Statistics for Set A and Set B.

	Set A	Set B
Mean	15	15
Median	14	14
Mode	13	13
Range	23	23

How can this happen? How could two sets with such different data values have the same statistics? What makes the difference? The answer is that the data in Set A is more widely **dispersed** than the data in Set B. The range of each set is 23, but the range considers only the maximum and minimum values. How is the rest of the data spread within that range? A box-and-whisker plot shows us the quartiles, range and interquartile range (IQR), but in this subtopic we will delve deeper and explore other statistics to help us interpret the data.

First, we will review the various measures of central tendency– the mean, median and mode – and explore how to estimate the mean for data grouped into classes. Then we will introduce two new measures of dispersion– variance and standard deviation. We will also consider how these statistics are affected when the data in a set change uniformly.

Concept

In this subtopic, we will take the opportunity to explore **relationships** between different statistical measures, such as the mean and standard deviation, as well as the impact that various **changes** have on those measures. We will use other strategies to **approximate** these measures when dealing with data that is grouped into ranges (called classes). How do these statistical measures help us understand the data better?

Student
view

Measures of central tendency

If you had to represent a data set with just one value, which one would you use? It should be a value around which the entire data set is well distributed. This first measure with which we summarise the data is a value for the **central tendency**, the one number which best represents the entire data set. In everyday language, this is referred to as the **average**. There are three common measures for the central tendency : mode, median and mean.

Mode

In section 4.2.1 (/study/app/math-aa-hl/sid-134-cid-761926/book/grouped-data-and-quartiles-id-25512/)you learned that for data grouped into intervals or classes, the class with the greatest frequency is called the **modal class**. The mode of a data set is the *value* with the greatest frequency (the word 'mode' is used for what is fashionable, what is 'popular'). For data grouped into intervals, the mode is estimated as the mid-interval value of the modal class. It is possible for a data set to have more than one mode – we call a data set like this bimodal, trimodal, and so on. If no values are repeated, or all of the values in a set are repeated the same number of times, then there is *no mode* .

Exam tip

Class intervals do not have to be equal, but for your work in this course, they will be.

Median, m

In section 4.2.2 (/study/app/math-aa-hl/sid-134-cid-761926/book/box-and-whisker-plots-id-25513/)you used **quartiles** to present data as a box-and-whisker plot and learned that the second quartile is also called the median . The median of a data set is the middle value when the data is arranged in order. It divides the data set in two, with half the data below and half the data above its value. If the data set has an even number of data points, say n, the median is estimated as the average of the data values of the $\frac{n}{2}\text{th}$ and $\left(\frac{n}{2}+1\right)$ th. If the data is grouped into classes, the median is estimated as the mid-interval of the class containing the middle value.

Student view

Overview
(/study/app/math-aa-hl/sid-134-cid-761926/o

Mean, μ

The measure of mean refers to the **arithmetic mean** of the data: the sum of the numerical data divided by the number of data points. The mean is often what is meant when the word 'average' is used. Thus, for a data set given by n data values x_i, the mean is given by

$$\mu = \frac{\sum_{i=1}^{n} x_i}{n}.$$

If the data is given by a frequency distribution table, then the mean is given by

$$\mu = \frac{\sum_{i=1}^{k} f_i x_i}{n},$$

where $f_i$is the frequency of the data value x_i and $n = \sum_{i=1}^{k} f_i$is the number of data points in the data set.

If the data is grouped into $k$classes, we can adapt this formula to *estimate* the mean:

$$\mu = \frac{\sum_{i=1}^{k} f_i x_i}{n},$$

where $f_i$is the frequency of the ith class and x_i is the mid-interval value of the ith class .

Making connections

Consider the difference that extreme values could have on these measures of centre. The mode would probably not change at all, and the median might only change a little, but what about the mean?

Section Student... (0/0) Feedback Print (/study/app/math-aa-hl/sid-134-cid-761926/book/the-big-picture-id-25517/print/) Assign

People working in fields where financial data are used must often consider which of these measures should be used in order to communicate information accurately and ethically. For example, median income is often used because it is not affected very much by one very rich person living in the area, whereas the mean income could drastically increase. Likewise, median home values are often used because one unkempt house in a neighbourhood could bring the mean value down, affecting the value of other homes.

How do you think people could use data like this unethically in order to mislead others?

Example 1

★★☆

Student view

Overview
(/study/ap
aa-
hl/sid-
134-
cid-
761926/o

The table below shows the data found by asking 20 students how many different countries they have travelled to in their lifetime, including the one in which they live. Find the mean, median and mode of the number of countries visited by the group of students.

Number of countries	Frequency
1	4
2	6
4	3
5	4
7	2
10	1

How many countries have you travelled to?

Kat72 Getty Images

More information

The image is a world map showing continents and countries in various colors. There are numerous colored map pins placed across different countries. The map spans from North and South America on the left to Europe, Africa, and Asia on the right. The Atlantic and Pacific Oceans are labeled in blue, and countries are marked with black borders. The map pins are placed in clusters over North America, Europe, and parts of the Middle East, indicating locations that might represent traveled destinations or points of interest. The pins vary in color, potentially symbolizing different categories or frequencies.

[Generated by AI]

Student
view

Overview (/study/app/math-aa-hl/sid-134-cid-761926/o

To find the mean μ given a frequency table, use the formula

$$\mu = \frac{\sum_{i=1}^{n} f_i x_i}{n}.$$

In the numerator, multiply each value by its frequency and add the products. In the denominator, add the frequencies (or notice from the problem that there are 20 students total).

$$\mu = \frac{1 \times 4 + 2 \times 6 + 4 \times 3 + 5 \times 4 + 7 \times 2 + 10 \times 1}{4 + 6 + 3 + 4 + 2 + 1}$$

Simplify the expression to find the value of the mean.

$$\mu = \frac{72}{20} = 3.6$$

To find the median m from 20 students, find the mean of the 10th and 11th values.

$$m = \frac{2 + 4}{2} = 3.$$

To find the mode, note that the highest frequency is 6, for the value 2.

$$\text{mode} = 2.$$

Which measure of central tendency do you think best represents the data in this example?

Example 2

★★☆

The table below shows grouped data for the heights of 51 people. Estimate the mean, median and mode of the data.

Grouped height data.

Student view

Overview
(/study/app/math-aa-hl/sid-134-cid-761926/o

Height (cm)	Frequency f_i
150—152	3
153—155	8
156—158	6
159—161	12
162—164	7
165—167	4
168—170	2
171—173	4
174—176	5
Total	51

The table below shows the data along with the mid-interval, x_i, and the product of the frequency with the mid-interval values, $f_i x_i$.

Calculations to determine the mean of grouped data.

Height (cm)	Mid-interval (cm) x_i	Frequency f_i	$f_i x_i$
150—152	151	3	453
153—155	154	8	1232
156—158	157	6	942
159—161	160	12	1920
162—164	163	7	1141
165—167	166	4	664

Student view

Overview (/study/app/math-aa-hl/sid-134-cid-761926/o

Height (cm)	Mid-interval (cm) x_i	Frequency f_i	$f_i x_i$
168—170	169	2	338
171—173	172	4	688
174—176	175	5	875
	Total	51	8253

To estimate the mean, sum the frequencies and the products:

$$\mu = \frac{\sum_{i=1}^{9} f_i x_i}{n} = \frac{8253}{51} \approx 161.82\,\text{cm}.$$

To estimate the median height of 51 people, look for the 26th value. In this case, the 26th value is in the interval 159–161, so the estimated median is the mid-interval value 160.

To estimate the mode, take the mid-interval value of the modal class. The interval 159–161 is the modal class, so the estimated mode is 160.

Important

Calculating the mean from grouped data will, in general, not give the same result as calculating from ungrouped (raw) data. Why do you think using mid-interval values provides only an estimate of the mean? How does that differ from the mean we found in Example 1? Why are the calculations different for grouped and ungrouped data?

Making connections

While something as mundane as measuring heights can be sufficiently accurate with the ranges of data in Example 2, some disciplines require more precision than others. For example, a medical researcher testing the results of an experimental drug may use minute intervals when compiling data related to different dosages. On the other hand, an economist might look at financial data with ranges spanning several millions. Think about a field of study that you might want to study someday. Would research in that field require data in minute intervals, or would less precise data be sufficient?

Student view

Overview
(/study/app/math-aa-hl/sid-134-cid-761926/o

Theory of Knowledge

Consider the way language is used in mathematics. When finding measures of central tendency, we define **centre** in three different ways. The mean is central in relation to its distance from the other data. The median is central in its position among the other data. The mode could be considered the centre of attention, or popularity, given that it is the value that more of the data are equal to. Can you think of any other mathematical concepts that are interpreted differently based on how we define certain terms?

3 section questions

Question 1

Difficulty: ★★☆

The table below gives the daily high temperatures (in $^{\circ}C$) in Paris, France during the month of March.

Temperature (°C)	Frequency (days)
6	2
7	1
8	2
9	2
10	4
11	3
12	6
14	2
15	1
16	3
17	1

Student view

Overview
(/study/ap
aa-
hl/sid-
134-
cid-
761926/o

Temperature (°C)	Frequency (days)
18	1
19	3

Find the mean of the temperatures, rounded to the nearest tenth of a degree.

1 $12.3\,^\circ\mathrm{C}$ ✓

2 $29.3\,^\circ\mathrm{C}$

3 $12.0\,^\circ\mathrm{C}$

4 $2.4\,^\circ\mathrm{C}$

Explanation

To find the mean μ given a frequency table, use the formula

$$\mu = \frac{\sum_{i=1}^{n} f_i x_i}{n}.$$

In the numerator, multiply each value by its frequency and add the products. In the denominator, add the frequencies (or notice from the problem that there are 31 days in March).

$$\mu = \frac{6\times 2 + 7\times 1 + 8\times 2 + 9\times 2 + 10\times 4 + 11\times 3 + 12\times 6 + 14\times 2 + 15\times 1 + 16\times 3 + 17\times 1 + 18\times 1 + 19\times 3}{2+1+2+2+4+3+6+2+1+3+1+1+3}$$

Simplify the expression to find the value of the mean.

$$\mu = \frac{381}{31} = 12.3$$

Question 2

Difficulty: ★★☆

Crab fishermen count the number of crabs they catch in each trap in order to approximate the weight of the crabs they have caught towards their legal quota. The table below shows the data for their first 135 traps.

Crabs per trap	Frequency
71—80	6

Student view

Overview (/study/app/math-aa-hl/sid-134-cid-761926/o

Crabs per trap	Frequency
$81—90$	18
$91—100$	32
$101—110$	48
$111—120$	26
$121—130$	5

Estimate, to 1 decimal place, the mean (μ), mode and median (m) of the number of crabs per trap.

1 $\mu = 101.8, \text{mode} = 105.5, m = 105.5$ ✓

2 $\mu = 101.3, \text{mode} = 105, m = 105$

3 $\mu = 102.3, \text{mode} = 106, m = 106$

4 $\mu = 101.5, \text{mode} = 105.5, m = 105.5$

Explanation

For the grouped data,

$$\sum_{i=1}^{6} f_x \times x_i = 6 \times 75.5 + 18 \times 85.5 + \cdots + 5 \times 125.5 = 13742.5$$

and $\sum_{i=1}^{6} f_i = 135$ such that $\mu = \frac{13742.5}{135} \approx 101.8$.

The modal class is the interval $101—110$, which has a mid-interval value of 105.5.

The middle data point of the ordered data set is $\frac{135+1}{2} = 68$, so we need to look for the value of the 68th data point. In the first three intervals, there are $6 + 18 + 32 = 56$ data points, but the next interval consists of 48 points, so we see that the 68th data point falls in the interval corresponding to $101—110$. Hence, the median, m, is 105.5.

Question 3

Difficulty: ★★★

The following table of values is incomplete, but you know that the mean of the data is $\mu = 15$. Find the value of x. Give your answer as an integer.

Student view

Value	Frequency
4	2
5	4
13	x
14	6
21	5
23	3
26	2

4

Accepted answers

4, x = 4

Explanation

To find the mean, use the formula:

$$\mu = \frac{\sum_{i=1}^{n} f_i x_i}{n}$$

Enter the data from the table:

$$\mu = \frac{4 \times 2 + 5 \times 4 + 13 \times x + 14 \times 6 + 21 \times 5 + 23 \times 3 + 26 \times 2}{2 + 4 + x + 6 + 5 + 3 + 2}$$

Finally, set the mean equal to 15 and solve the equation for x:

$$\begin{aligned} 15 &= \frac{13x + 338}{x + 22} \\ 15(x + 22) &= 13x + 338 \\ 15x + 330 &= 13x + 338 \\ 2x &= 8 \\ x &= 4 \end{aligned}$$

Measures of dispersion

Range and interquartile range

You have already encountered the **range** and **interquartile range** (IQR), which are two basic examples of measures of dispersion. Dispersion has to do with how much data is spread out or clustered together. In this section, we will examine how measures of dispersion tell us more about the data. To explore this, we will revisit the two sets of data from section 4.3.0 (/study/app/math-aa-hl/sid-134-cid-761926/book/the-big-picture-id-25517/):

Set A: $\{3, 4, 5, 13, 13, 14, 18, 21, 23, 25, 26\}$

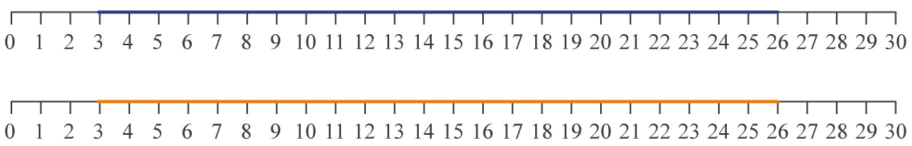
Set B: $\{3, 12, 13, 13, 13, 14, 15, 16, 16, 24, 26\}$

When we examined this data before, we recognised that the mean, median, mode and range were all identical, but even a brief look at the data indicates that something is different.

Identical statistics do not necessarily indicate identical data.

	Set A	Set B
Mean	15	15
Median	14	14
Mode	13	13
Range	23	23

The range is the distance between the highest value and the lowest; that is, $\text{maximum} - \text{minimum}$. While this does tell you something, you don't learn very much about the set of data.



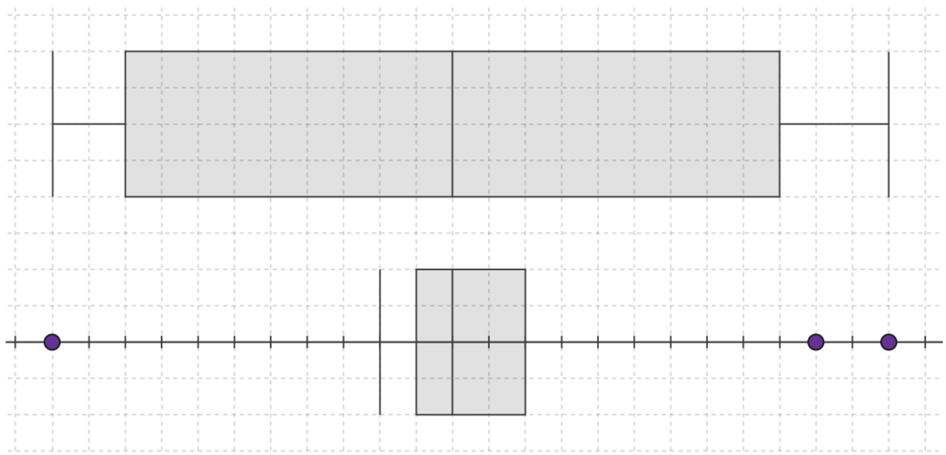
Number lines for Set A (blue) and Set B (orange)

More information

The image contains two number lines. The first number line on top is labeled for Set A and is colored blue. It spans horizontally from 0 to 30, with a solid blue line stretching from 5 to 23. The second number line below it is labeled for Set B and is colored orange. It also spans from 0 to 30, with a solid orange line extending from 13 to 16. These number lines visually represent the ranges of two different data sets, Set A and Set B, showing their respective minimum and maximum values on the number line.

[Generated by AI]

The IQR is somewhat more useful, in that it tells us the distance between the first and third quartiles, that is, $Q_3 - Q_1$, which is the range of the middle 50% of the data. This is $23 - 5 = 18$ for Set A and $16 - 13 = 3$ for Set B. The middle 50% of Set A is spread out six times as much as the middle 50% of Set B. This gives more information, but it could have been estimated simply by looking at the box-and-whisker plot. How could you describe the dispersion of the data even more specifically?



Box-and-whisker plots for Set A and Set B.

More information

The image shows box-and-whisker plots for two different data sets, Set A and Set B, displayed over a grid. Both plots illustrate the distribution of data, with each plot featuring a box that represents the interquartile range (IQR).

Overview
(/study/app/math-aa-hl/sid-134-cid-761926/o

- **Set A:** The box for Set A is significantly larger, indicating a wider IQR. The lower and upper quartiles are marked by the edges of the box which stretches horizontally from the first quartile (Q1) to the third quartile (Q3). The line inside the box represents the median of Set A. Whiskers extend from the box to depict the smallest and largest values within a specified range.
- **Set B:** Comparatively, the box for Set B is much smaller, indicating a narrow IQR. The edges of the box mark the lower and upper quartiles. A line within this box shows the median value for Set B. The whiskers for Set B also extend outward but cover a shorter range than those of Set A.

Overall, Set A demonstrates more variance in data distribution compared to Set B, highlighting the different spreads within the two data sets.

[Generated by AI]

Variance, σ^2, and standard deviation σ

Two more advanced measures of dispersion are variance and standard deviation. These measure how much the data set *varies* or *deviates* from the mean. The higher the value, the more spread out the data is.

Section Student... (0/0) Feedback Print (/study/app/math-aa-hl/sid-134-cid-761926/book/measures-of-central-tendency-id-25518/print/) Assign

For example, first calculate the difference between each data point and the mean. Then add up all these differences and divide by the number of data points. Is this a good measure of dispersion? Can you think of any problems with this method? Think about what happens when you sum positive and negative values.

To make sure that positive and negative differences don't cancel each other out, we **square** the differences when calculating the variance. The variance, written Var or often σ^2, is therefore a measure of the spread of the data around the mean. The standard deviation, σ, is the **square root** of the variance. Thus, variance and standard deviation measure the spread in the same manner, but one is the square of the other.

In this course, you will not need to calculate the variance or standard deviation except with your GDC, which you will see how to do later in this section . Thus, while **you will not be asked to apply the following formulae** , it is worth seeing how it is done.

We define the variance as

Student view

$$\sigma^2 = \frac{\sum_{i=1}^{k} f_i(x_i - \mu)^2}{n},$$

Overview (/study/app/math-aa-hl/sid-134-cid-761926/o

where k is the number of distinct values or classes. The standard deviation is therefore

$$\sigma = \sqrt{\frac{\sum_{i=1}^{k} f_i (x_i - \mu)^2}{n}}.$$

Now we can determine the standard deviation of each set of data to see the difference in the dispersion as it relates to distance from the mean.

The variance can also be expressed as

$$\sigma^2 = \frac{\sum_{i=1}^{k} f_i x_i^2}{n} - \mu^2$$

because

$$\begin{aligned}\frac{\sum_{i=1}^{k} f_i (x_i - \mu)^2}{n} &= \frac{\sum_{i=1}^{k} f_i x_i^2}{n} - 2\mu \frac{\sum_{i=1}^{k} f_i x_i}{n} + \mu^2 \frac{\sum_{i=1}^{k} f_i}{n} \\ &= \frac{\sum_{i=1}^{k} f_i x_i^2}{n} - 2\mu^2 + \mu^2 \\ &= \frac{\sum_{i=1}^{k} f_i x_i^2}{n} - \mu^2.\end{aligned}$$

This formula can be easier to remember as it is ‘the mean of the squares minus the square of the mean’.

Making connections

If the interquartile range goes together with the median, the standard deviation, which is associated with the variance of the data, goes together with the mean. Why do you think this is?

Example 1

★★☆

Find the variance and standard deviation of the data from Set A ($\mu = 15$). Since only one value is repeated, you do not need to use a frequency table.

Student view

To find the variance, start with the formula, replacing f_i with 1 and k with n:

Overview (/study/app/math-aa-hl/sid-134-cid-761926/o

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{n}$$

Since there are several calculations for each element of data, it is helpful to use a table to organise your calculations as we do in the table below.

Calculations to determine the variance of the data in Set A.

x_i	$x_i - \mu$	$(x_i-\mu)^2$
3	$3-15$	144
4	$4-15$	121
5	$5-15$	100
13	$13-15$	4
13	$13-15$	4
14	$14-15$	1
18	$18-15$	9
21	$21-15$	36
23	$23-15$	64
25	$25-15$	100
26	$26-15$	121
		$\text{Total} = 704$

Evaluating the formula with these values gives you

$$\sigma^2 = \frac{704}{11} = 64.$$

Since standard deviation is the square root of variance,

$$\sigma = \sqrt{64} = 8.$$

Student view

Overview
(/study/ap
aa-
hl/sid-
134-
cid-
761926/o

Example 2

★★☆

Find the variance and standard deviation of the data from Set B ($\mu = 15$). Since several values are repeated, use a frequency table.

To find the variance, start with the formula:

$$\sigma^2 = \frac{\sum_{i=1}^{k} f_i(x_i - \mu)^2}{n}$$

The table below shows the data from Set B arranged with frequencies and the calculations to find the variance .

Calculations to determine the variance of the data in Set B.

x_i	f_i	$x_i - \mu$	$f_i(x_i - \mu)^2$
3	1	$3 - 15$	144
12	1	$12 - 15$	9
13	3	$13 - 15$	12
14	1	$14 - 15$	1
15	1	$15 - 15$	0
16	2	$16 - 15$	2
24	1	$24 - 15$	81
26	1	$26 - 15$	121
Total	11		370

Evaluating the formula with these values gives you

$$\sigma^2 = \frac{370}{11} \approx 33.64.$$

Student
view

Since standard deviation is the square root of variance,

Overview (/study/app/math-aa-hl/sid-134-cid-761926/o

$$\sigma \approx \sqrt{33.64} \approx 5.80.$$

Based on our measures of the mean and standard deviation, we can summarise the data from each set by quoting its mean and standard deviation in the form:

Set A: $\mu \pm \sigma = 15 \pm 8$

Set B: $\mu \pm \sigma = 15 \pm 5.8$

The smaller standard deviation of Set B means that, on average, the data in Set B is closer to the mean than the data in Set A.

Theory of Knowledge

You will notice above that there are actually two formulae for finding variance.

$$\sigma^2 = \frac{\sum_{i=1}^{n} f_i(x_i - \mu)^2}{n} \text{ and}$$

$$\sigma^2 = \frac{\sum_{i=1}^{k} f_i x_i^2}{n} - \mu^2.$$

In this case we saw how the second was derived from the first, but in other cases it may not be evident how two formulae that find the same value are related. If two different formulae can consistently give you the same value, what does this tell us about the concept of truth in mathematics?

Calculating the mean, standard deviation and variance on a calculator

Data may consist of lots of numbers. Statistical analysis consists of applying various formulae to all these numbers. Using technology, we can compute statistics for large sets of data with a speed and accuracy that would be impossible for us to achieve on our own.

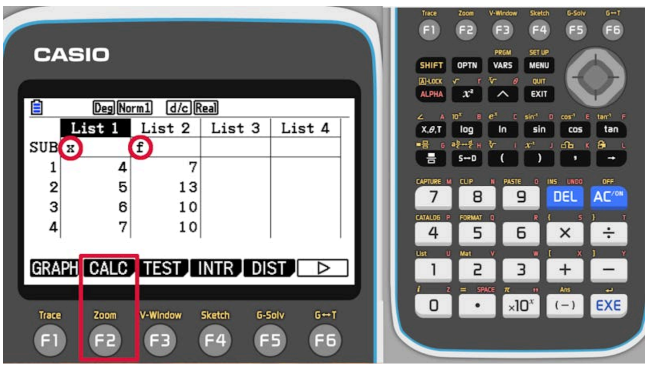
Exam tip

Knowing how to use your GDC is essential to success on the final exam. Make sure you learn how to use your calculator efficiently so you do not waste time during the exam trying to figure out where functions are and how they work.

Student view

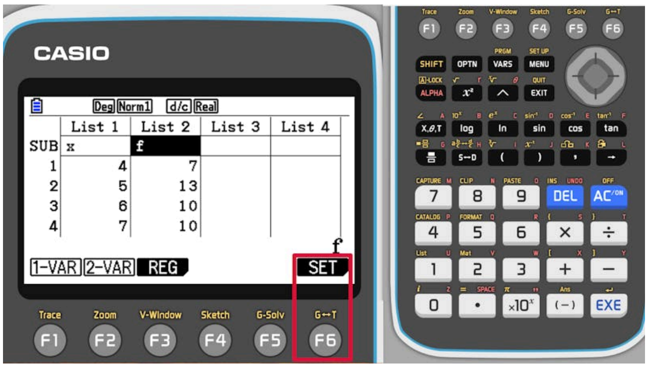
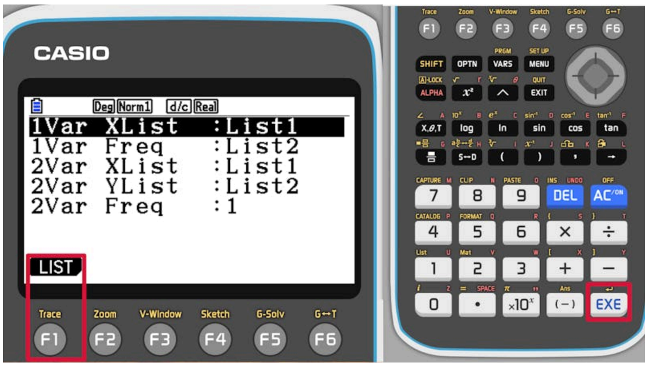
Overview
(/study/app/math-aa-hl/sid-134-cid-761926/o

In subtopic 4.2 (/study/app/math-aa-hl/sid-134-cid-761926/book/the-big-picture-id-25511/)you learned how to find the five-number summary of a data set using your GDC. The instructions below show how to use your calculator to find the mean, variance and standard deviation of a data set.

Step	Explanation
Press 2 to open the statistics mode.	
Enter the data. Note, that you can also give a name to your data. Remember that you used List1 and List2 for storing the data, you will need this later. Once done entering the data, press F2.	

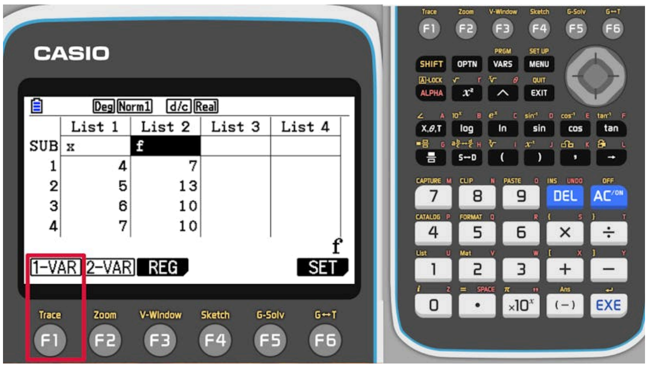
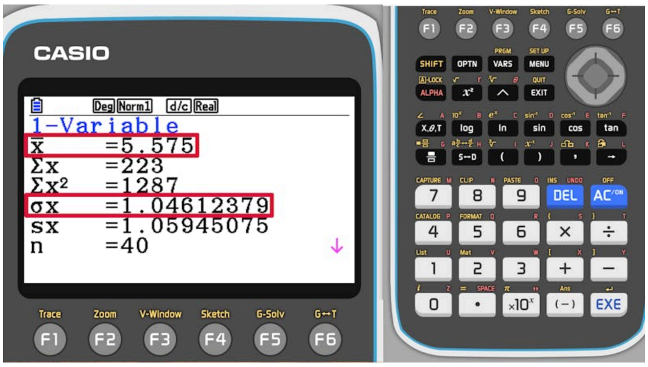
Student view

Overview (/study/ap aa-hl/sid-134-cid-761926/o

Step	Explanation
First you need to set some parameters, so press F6.	
Remember, you stored the x-values in List1 and the frequencies in List2. Use F1 and the number keys to tell this to the calculator. You are interested in the settings for 1-variable statistics now. Once done, press EXE.	

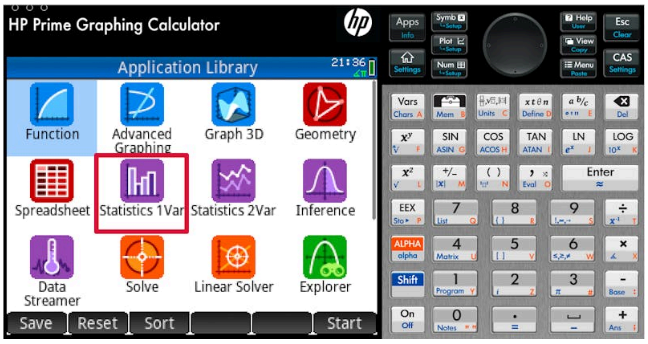
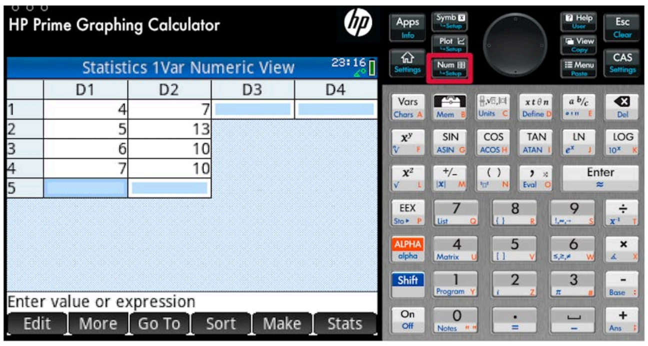
Student view

Overview
(/study/app/math-aa-hl/sid-134-cid-761926/o

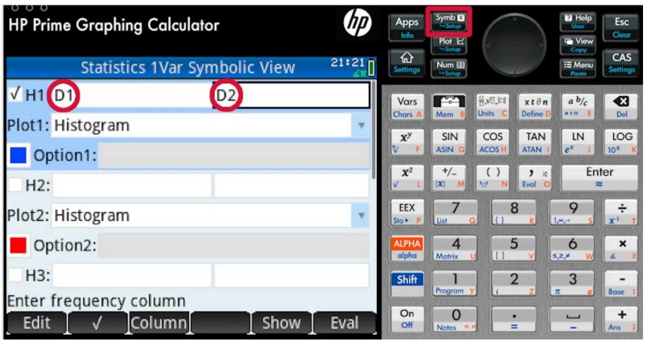
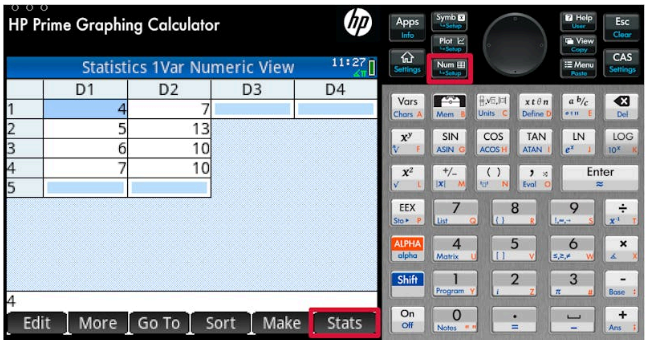
Step	Explanation
Press now F1 to view the 1-variable ststistics summary.	
In the summary, the mean is denoted by $\overline{x}$ and the standard deviation is σx. The variance is the square of the standard deviation, it is not given in the summary directly.	

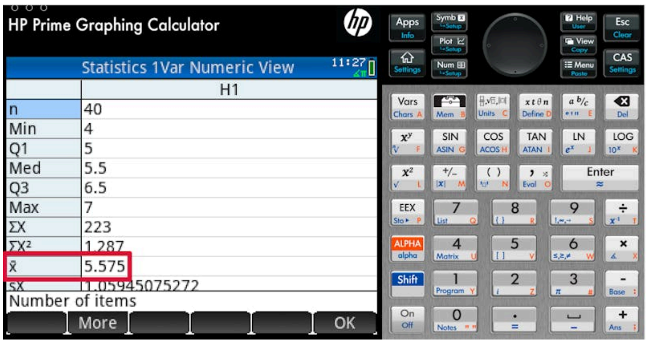
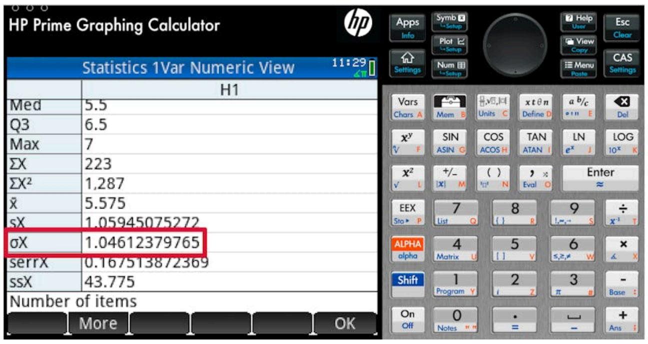
Student view

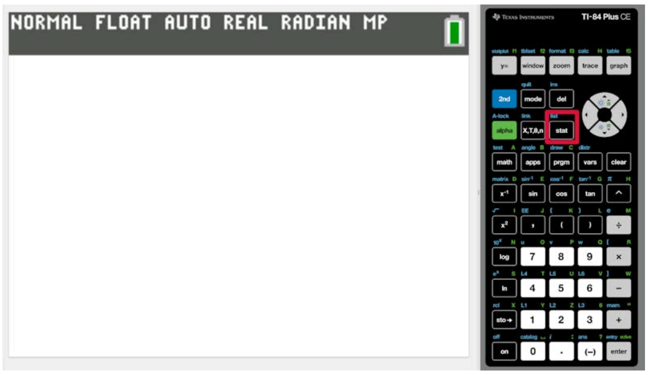
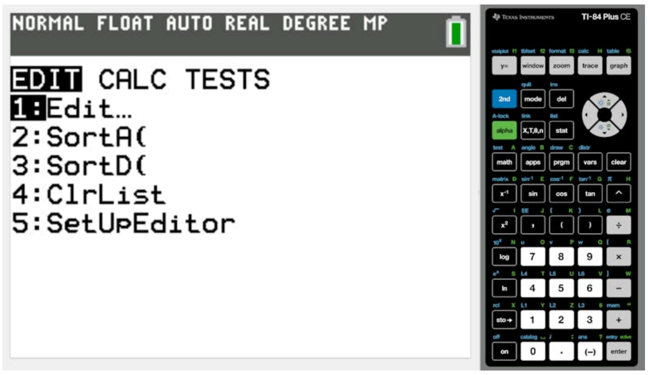
Overview
(/study/ap
aa-
hl/sid-
134-
cid-
761926/o

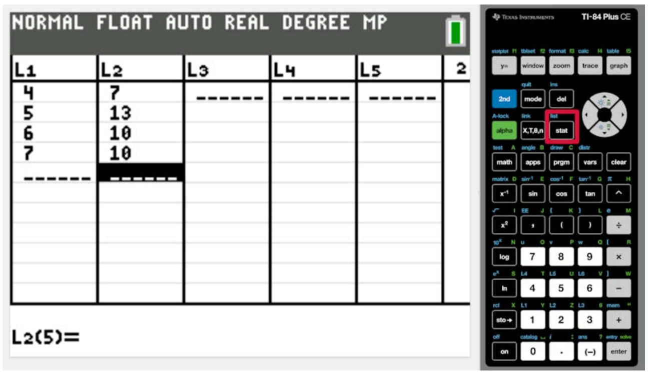
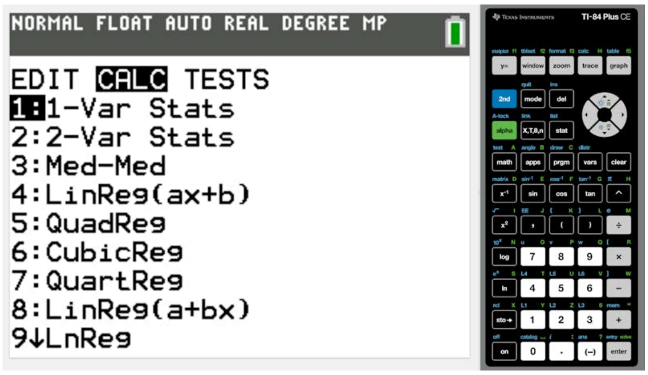
Step	Explanation
Choose the 1-variable statistics application.	
In the numeric view of the application, enter the data. Remember that you used D1 and D2 for storing the data, you will need this in the next step.	

Student
view

<table>
<tr><th>Step</th><th>Explanation</th></tr>
<tr><td>In the symbolic view you can tell the calculator where you stored the data. The first field is for the x-values (so enter D1). Enter D2 in the second field to tell the calculator where the frequencies are.</td><td>
</td></tr>
<tr><td>Go back in the numeric view and press stats to get the 1-variable statistics summary.</td><td>
</td></tr>
</table>

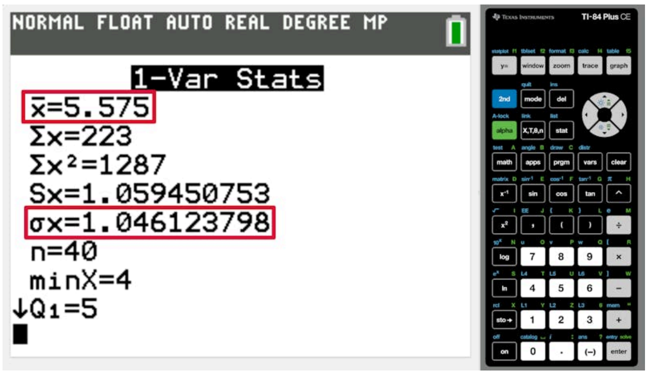
Step	Explanation
On the first screen (among other information) you can see the mean (denoted by $\overline{x}$).	
To see the standard deviation (which is denoted by σX), you need to scroll down. The variance is the square of the standard deviation, it is not given in the summary directly.	

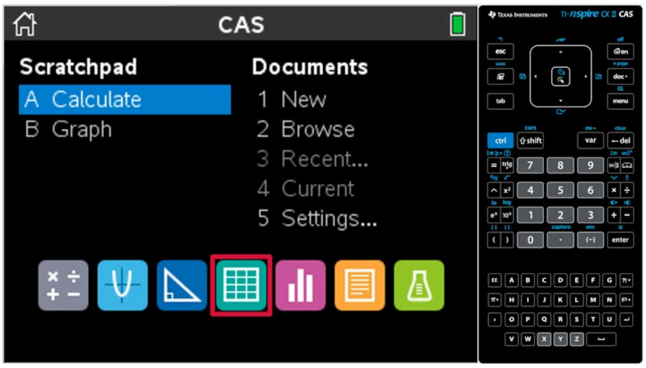
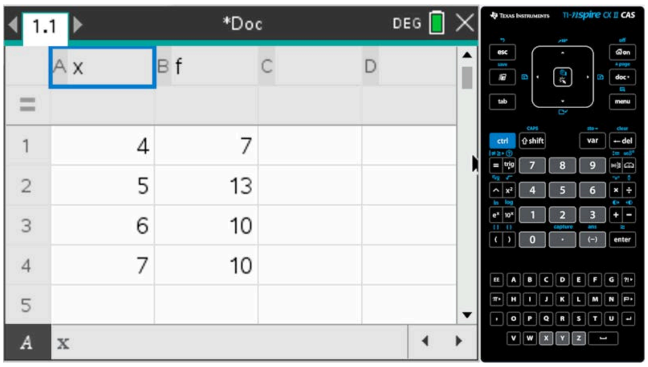
<table>
<tr><th>Step</th><th>Explanation</th></tr>
<tr><td>Press stat to start working with data sets.</td><td>
</td></tr>
<tr><td>First you need to enter your data, so choose to bring up the editing screen.</td><td>
</td></tr>
</table>

Step	Explanation
Enter the data. Remember that you used L1 and L2 for storing the data, you will need this later. Once done entering the data, press stat again.	
Choose the option to calculate the 1-variable statistics.	

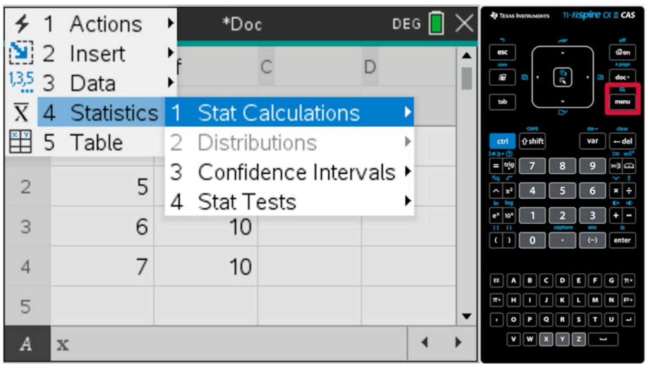
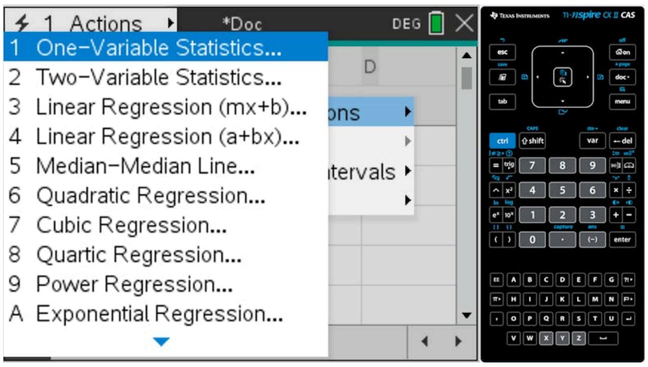
Student view

<table>
<tr><th>Step</th><th>Explanation</th></tr>
<tr><td>Remember, you stored the x-values in L1 and the frequencies in L2. Use 2nd and the number keys to tell this to the calculator.

Once done, move down to the last row and press enter.</td><td>
</td></tr>
<tr><td>In the summary, the mean is denoted by $\overline{x}$ and the standard deviation is σx. The variance is the square of the standard deviation, it is not given in the summary directly.</td><td>
</td></tr>
</table>

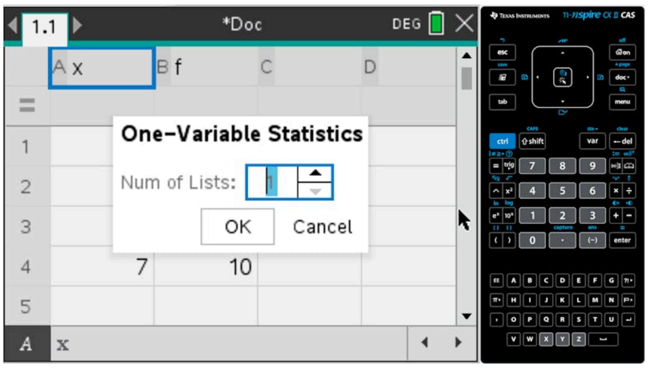
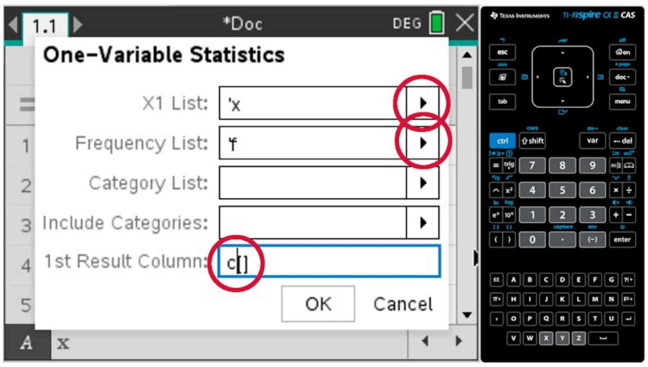
<table>
<tr><th>Step</th><th>Explanation</th></tr>
<tr><td>To start, open a spreadsheet document.</td><td>
</td></tr>
<tr><td>Enter the data. Note, that you can also give a name to your data. Remember these names, you will need this later.</td><td>
</td></tr>
</table>

Overview
(/study/app/math-aa-hl/sid-134-cid-761926/o

Step	Explanation
To see the mean and stadard deviation, open the menu...	
... and choose one-variable statistics.	

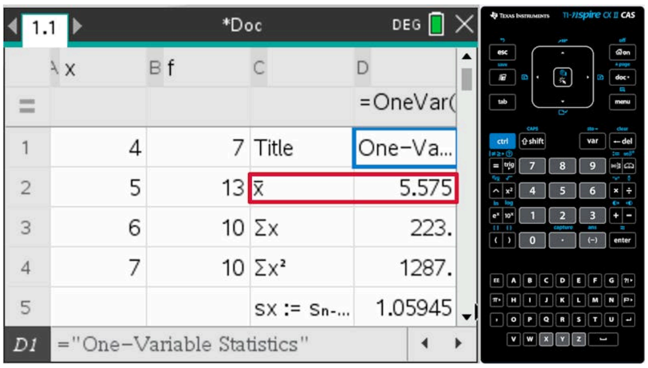
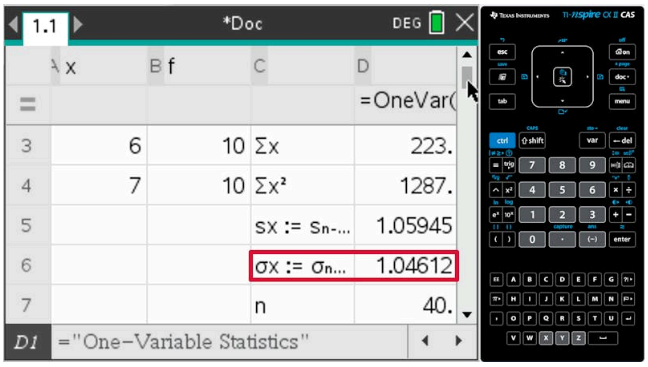
Student view

Overview
(/study/app/math-aa-hl/sid-134-cid-761926/o

Step	Explanation
You have one list (you will specify the frequencies later).	
Choose these names you used to store the x-values and the frequencies from the pull-down menu. The results will be stored in the spreadsheet. Since the data is stored in the first two columns (a and b), choose a different column for the result.	

Student view

Overview
(/study/ap
aa-
hl/sid-
134-
cid-
761926/o

Step	Explanation
On the first screen (among other information) you can see the mean (denoted by $\overline{x}$).	
To see the standard deviation (which is denoted by σX), you need to scroll down. The variance is the square of the standard deviation, it is not given in the summary directly.	

Be aware

It is easy to make a mistake in data entry when using your GDC to calculate statistical measures. Always double check your input. As well, once the analysis is done by the GDC, check that the number of data points n matches the number of data points in the

Student
view

Overview (/study/app/math-aa-hl/sid-134-cid-761926/o

question, which may be mentioned in the question or is, otherwise, easy to find.

International Mindedness

Imagine you are an economist researching market trends across the globe, or a political strategist trying to determine the impact of a trade agreement on the several countries involved. While two countries may be drastically different in population and economics, statistics can still be used to make comparisons and determine how certain factors impact the country. How do you think measures of centre and dispersion might be used to compare or contrast very large countries with smaller countries? Or how might prosperous countries be compared with less prosperous countries?

Theory of Knowledge

Statistical analysis in regard to significance as well as measures of dispersion is a key area within the human sciences where mathematics is applied. It is this use of analysis, and drawing conclusions from the mathematical analysis, that contributes to subject areas such as psychology and economics being considering a science.

Psychologists and economists have often had 'statistically significant' results of various studies, though the **application** of their findings do not always apply to all people in all contexts. How can this be? If the mathematics was accurate and valid then why would mathematical modelling lack ecological validity?

3 section questions

Question 1

Difficulty: ★★☆

In the 2017 Track and Field World Championship, 13 athletes qualified for the finals in the men's javelin event. Their distances (rounded to the nearest metre) are given in the frequency distribution below.

Distance (m)	Frequency
91	1
86	4

Student view

Overview
(/study/app/math-aa-hl/sid-134-cid-761926/o

Distance (m)	Frequency
85	1
84	6
83	1

Use your GDC to find the standard deviation. Give your answer to 3 significant figures.

1.96 ✓

Accepted answers

1.96, 1,96, 1.96 metres, 1.96m, 1,96 metres, 1,96m

Explanation

Use the 1-Variable Statistics function on the calculator with the distances as the list of values and the frequencies in the frequency list. The standard deviation on the calculator is denoted $\sigma x = 1.955117696$. Rounding to 3 significant figures gives $\sigma = 1.96$.

Question 2

Difficulty: ★★☆

A cell phone company compiled data from 200 of its customers to analyse texting activity by finding the average number of texts they sent each day, rounded to the nearest whole number. The frequency table below shows their findings.

Number of texts	Frequency
12	6
13	15
14	29
15	41
16	35
17	25
18	18

Student view

Overview
(/study/ap
aa-
hl/sid-
134-
cid-
761926/o

Number of texts	Frequency
19	19
20	12

To 1 decimal place where appropriate, what is the mean, standard deviation, first quartile, median and third quartile?

1 $\mu = 16.0,\, \sigma = 2.1,\ Q_1 = 14.5,\ m = 16,\ Q_3 = 17$ ✓

2 $\mu = 16.0,\, \sigma = 2.1,\ Q_1 = 14,\ m = 16,\ Q_3 = 17$

3 $\mu = 16.0,\, \sigma = 4.4,\ Q_1 = 14.5,\ m = 16,\ Q_3 = 17$

4 $\mu = 15.9,\, \sigma = 2.1,\ Q_1 = 15.0,\ m = 16.0,\ Q_3 = 17.0$

Explanation

Using our GDC and rounding to 1 decimal place where necessary, these are the values.

Question 3

Difficulty: ★★☆

The 20 clubs in a football league finished the 2018—19 season with the following number of goals per club:

$$89, 79, 69, 65, 63, 59, 50, 50, 49, 48, 44, 44, 44, 43, 41, 35, 33, 32, 30, 30.$$

Use the data to determine which choice best summarises the number of goals scored per club.

1 49.9 ± 15.9 goals ✓

2 49.9 ± 16.3 goals

3 15.9 ± 49.9 goals

4 16.3 ± 49.9 goals

Explanation

Use the 1-Variable Statistics function on the calculator with the numbers of goals as the list of values. A list of frequencies is not necessary, since the data are listed individually and there are not many repeated values.

Student
view

Overview
(/study/ap
aa-
hl/sid-
134-
cid-
761926/o

The mean on the calculator is denoted $\bar{x} = 49.85$, which is rounded to 49.9. Many calculators use $\bar{x}$ instead of μ, but that is just another way of denoting the mean.

The standard deviation on the calculator is denoted $\sigma x = 15.93510276$. Rounding to 3 significant figures gives $\sigma = 15.9$.

Using these two values, you can summarise the number of goals each club scored with the range $\mu \pm \sigma = 49.9 \pm 15.9$.

4. Probability and statistics / 4.3 Measures of central tendency and dispersion

Data transformation

Section Student... (0/0) Feedback Print (/study/app/math-aa-hl/sid-134-cid-761926/book/data-transformation-id-25520/print/) Assign

Growing plants

Credit: stevanovicigor Getty Images

Imagine you have been collecting agricultural data by measuring the heights of several plants at various stages throughout their growth. After completing your entire experiment, you realise that the measuring stick you used has been cut off and is missing the first 3 cm. Is all your work wasted? Do you need to plant all new plants and collect all the data again?

Thankfully, no. All you need to do is to subtract 3 cm from all your measurements. But what will that do to the statistics you've already calculated? Will the mean change? What about the standard deviation? Will it get bigger or smaller, or will it stay the same?

Student
view

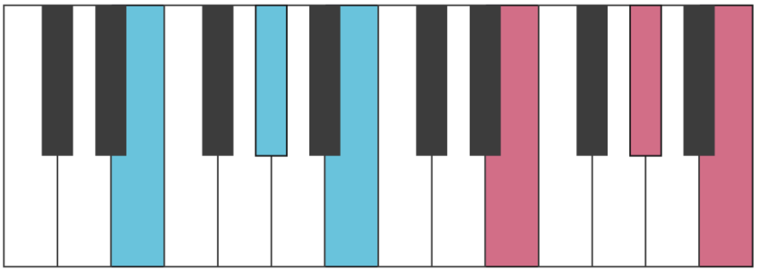
Overview
(/study/ap
aa-
hl/sid-
134-
cid-
761926/o

Important

When *adding or subtracting* the same value to all the data in a set, the mean will increase or decrease by the same amount, but the standard deviation will remain the same. This is because the distance between the data values has not changed. They have not got any further apart or closer together; they have just shifted up or down.

Making connections

The change described here is similar to what a pianist does when playing a chord in different octaves. The chord sounds the same, because the notes are all still the same distance apart. This pitch is different, because they are playing in a different octave.



More information

This image is a diagram of a piano keyboard, featuring several keys highlighted with different colors. The layout follows the standard pattern of a piano, alternating between black and white keys. Several white keys are colored in blue and pink, indicating specific notes or keys of interest. White keys are between sets of two and three black keys, mimicking the natural arrangement. The blue keys are second and fourth in the sequence, and pink keys are sixth and eighth, correlating with certain notes or functions on the keyboard. This visual representation may be used to demonstrate a musical concept or lesson related to piano playing or music theory.

[Generated by AI]

Student
view

Now suppose the measuring stick was printed with the wrong scale, and all the measurements need to be multiplied by 2. How will this change the mean? Will it affect the variance or the standard deviation, and if so, how?

Overview (/study/app/math-aa-hl/sid-134-cid-761926/o

Important

If each data point is *multiplied or divided* by a constant factor, then the spread *does* change. This is because as values are multiplied, the distances between them get multiplied as well.

So in our example, the mean height will be doubled and the distance between each height and the mean height will also be doubled. What will be the effect of this on the variance and the standard deviation? Remember that the variance is based on the square of the distance from each value to the mean, and the standard deviation is the square root of the variance.

For a data set $\{x\}$ with mean μ_x, standard deviation σ_x and variance σ_x^2, if every value in $\{x\}$ is multiplied by a constant c, then:

- the mean of the transformed data set is $c\mu_x$
- the standard deviation of the transformed data set is $|c|\,\sigma_x$
- the variance of the transformed data set is $c^2\sigma_x^2$.

Example 1

★☆☆

A set of data has a mean of 11 and a standard deviation of 2.3. If every value in the set is decreased by 5, write down the mean and standard deviation of the new set of data.

When every value in a set of data is increased or decreased by the same amount, the mean shifts by the same amount, but all the values will be the same distance apart. Since the data were all decreased by 5, the mean will now be $11 - 5 = 6$ and the standard deviation will stay at 2.3.

Example 2

★☆☆

The same set of data, with a mean of 11 and a standard deviation of 2.3, is modified so that every value in the set is multiplied by 4. Find the mean and standard deviation of the new set of data.

Student view

Overview (/study/app/math-aa-hl/sid-134-cid-761926/o

When every value in a set of data is multiplied by the same number, the mean is also multiplied by that amount, as well as the distances between each of the data. Since the data were all multiplied by 4, the mean will now be $11 \times 4 = 44$ and the standard deviation will increase to $2.3 \times 4 = 9.2$.

You can see this relationship for yourself by examining how the mean and standard deviation change using the applet below. Can you predict the new mean and standard deviation before finding the values with the applet?

Interactive 1. Mean and Standard Deviation.

Credit: GeoGebra (https://www.geogebra.org/m/UcXWCHfW) Maths Learning Centre University of Adelaide

More information for interactive 1

This interactive allows users to explore how linear transformations affect a data set by adjusting the values of constants ‘a’ (ranging from −5 to 5) and ‘ b’ (ranging from −10 to 10). Also users can move red points on the first graph, which dynamically changes the dataset's mean and standard deviation. A second graph displays a target distribution shown as a purple point. Users can apply transformations of the form $aX + b$ to the data set, using the 'a' and 'b' , sliders to apply linear transformations, the user's goal is to match the transformed data's position to the purple point's distribution, effectively recreating the target statistical properties through proper scaling and shifting adjustments.

For example, we start with an original dataset (X) having a mean of 2 and a standard deviation of 3, as shown on the first number line. When we apply the linear transformation $aX + b$ with $a = 1.1$ (scaling factor) and $b = -0.6$ (shifting factor), the tool dynamically updates the statistics:

The new mean becomes 1.6 (calculated as:

$a \times \text{original mean} + b = 1.1 \times 2 + (-0.6) = 1.6$).

The new standard deviation becomes 3.3 (calculated as: $|a| \times \text{original sd} = 1.1 \times 3 = 3.3$).

Student view

Overview (/study/app/math-aa-hl/sid-134-cid-761926/o

The second graph displays the transformed distribution, visually demonstrating how: Scaling (a): Stretches ($a > 1$) or compresses ($0 < a < 1$) the spread of data, affecting the standard deviation and Shifting (b): Moves the entire distribution left ($b < 0$) or right ($b > 0$), altering the mean without changing the spread.

Users can further experiment by adjusting a and b to observe real-time changes in the statistical measures and their visual representation on the number line.

3 section questions

Question 1

Difficulty: ★☆☆

The maths students in a school count their money and calculate the mean to be $\mu = €15$. Everyone in the class agrees to donate $€5$ to a local charity. Find the mean amount of money the students have now.

1 $\mu = €10$ ✓

2 $\mu = €20$

3 $\mu = €75$

4 $\mu = €3$

Explanation

Since every student is donating $€5$, the money each one has is decreased by 5. Since they have subtracted 5 from all of the data, the mean will also decrease by 5.

Therefore, $\mu = €15 - 5 = €10$.

Question 2

Difficulty: ★★★

A population of fish is studied through a sample of 25 whose masses were measured. We have the following grouped data.

Student view

Overview
(/study/ap
aa-
hl/sid-
134-
cid-
761926/o

Mass (g)	Frequency
$0—1$	1
$1—2$	5
$2—3$	9
$3—4$	7
$4—5$	3

A calibration error in the weighing scale means that each mass is underestimated by 3.5 g. What is the correct estimate of the mean mass of the fish?

1 6.24 g ✓

2 2.74 g

3 6.28 g

4 6.14 g

Explanation

The mean of the original data is given by

$$\mu = \frac{\sum_{i=1}^{5} f_i \times x_i}{\sum_{i=1}^{5} f_i} = \frac{1 \times 0.5 + 5 \times 1.5 + 9 \times 2.5 + 7 \times 3.5 + 3 \times 4.5}{1+5+9+7+3} = \frac{68.5}{25} = 2.74.$$

Thus, the true mean is 3.5 larger, that is 6.24 g.

Question 3

Difficulty: ★★★

For a given data set, the variance is 12.8. Each data point in the data set is multiplied by 1.5, then increased by 3.7. What is the standard deviation of the new data set? Round your final answer to 3 significant figures.

5.37 ✓

Student
view

Accepted answers

5.37, 5,37

Overview
(/study/app/math-aa-hl/sid-134-cid-761926/o

Explanation

The first detail you need to notice is that you are given the variance of the first data set, but you are asked for the standard deviation of the new data set.

Therefore, the standard deviation of the first data set is $\sigma = \sqrt{\text{variance}} = \sqrt{12.8} = 3.577708764$

(Note: For problems that require multiple steps to solve, do not round intermediate answers.)

Next, since the data are all multiplied by 1.5, the standard deviation is also multiplied by 1.5. Hence, $\sigma = 1.5 \times 3.577708764 = 5.36656314$.

Finally, when the data are transformed with addition or subtraction, the standard deviation remains unchanged.

Therefore, the final answer rounded to 3 significant figures is $\sigma = 5.37$.

4. Probability and statistics / 4.3 Measures of central tendency and dispersion

Checklist

Section Student... (0/0) Feedback Print (/study/app/math-aa-hl/sid-134-cid-761926/book/checklist-id-25521/print/) Assign

What you should know

By the end of this subtopic you should be able to:

- calculate the mean of a frequency distribution by hand or with a calculator
- use mid-interval values to estimate the mean of data grouped in intervals
- find the variance and standard deviation of a frequency distribution with a calculator
- find the new mean, variance and standard deviation when a set of data is transformed by addition, subtraction, multiplication or division.

4. Probability and statistics / 4.3 Measures of central tendency and dispersion

Investigation

Student view

Here's a problem-solving challenge you can do with a friend.

1. Create a set of data, with or without frequencies depending on the level of challenge you want.
2. Calculate the mean and standard deviation of the set and check your values with the GDC.
3. Roll dice, use a random number generator, or have a friend choose a new mean and standard deviation for you to use as your target.
4. Transform your set of data by adding, subtracting, multiplying or dividing the set by constants until the mean and standard deviation equal the target.
5. You and a friend can make it a game to see who can work it out faster or who can come up with the most original ways to reach the target.