TOPIC 4
PROBABILITY AND STATISTICS

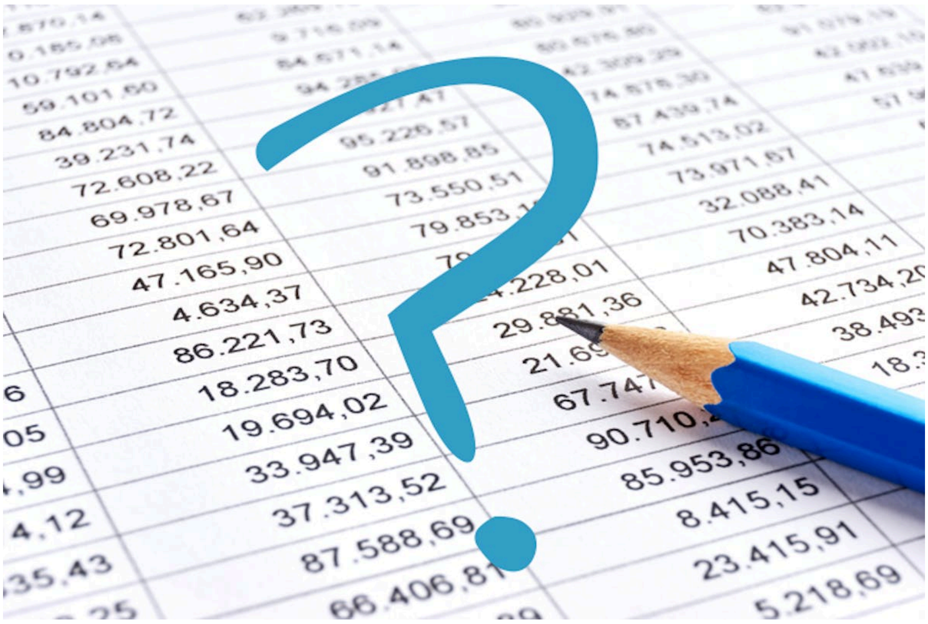
SUBTOPIC 4.2
PRESENTATION OF DATA

Overview
(/study/ap
122-
cid-
754029/

Teacher view

4. Probability and statistics / 4.2 Presentation of data

The big picture



A sheet of raw data

Credit: repinanatoly Getty Images

More information

The image shows a spreadsheet containing rows and columns of numerical data. A blue pencil is placed on the spreadsheet, and a large blue question mark is superimposed over the image. The numbers appear to be financial or statistical data, with various figures arranged in a grid format. The question mark suggests uncertainty or inquiry related to the data presented in the spreadsheet.

[Generated by AI]

'A picture is worth a thousand words.'

Student
view

Overview
(/study/ap
122-
cid-
754029/

This saying rings true in almost all areas of life, but in few areas is it more applicable than the area of statistics. With spreadsheets filled with row upon row of numbers, what we call **raw data**, even the most observant and brilliant mathematicians struggle to find patterns, make comparisons or recognise trends. When those data are compiled into a meaningful model, it is easier to discover relationships and make inferences about what the data represent. In this subtopic, you will discover how to use two models in particular: the cumulative frequencycurve and the box-and-whisker plot. You will learn how to identify important facts about the data from each model and how to construct models to communicate facts about data you might have collected.

Concept

Statisticians use **modelling** to represent data so that patterns, trends and other relationships between variables are easy to recognise. Models are used to communicate significant results and enable readers to make inferences about the population the data represents. What types of models do journalists use to tell news stories?

4. Probability and statistics / 4.2 Presentation of data

Grouped data and quartiles

Grouped data

Grapes in a vineyard.

Credit: repinanatoly Getty Images

Student
view

Imagine we are interested in how many grapes there are in the bunches in a vineyard. You can choose some bunches and count the number of grapes in each to obtain, say, the following numbers of grapes:

$$9, 12, 15, 18, 9, 11, 14, 13, 17, 8, 8, 12, 6, 10, 15, 17, 6, 7, 11, 14, 12, 5, 15, 4, 1, 14, 16,$$

$$5, 10, 7, 10, 12, 21, 6, 2, 8, 11, 13, 17, 13, 12, 14, 10.$$

It is obvious that representing your findings as this string of numbers does not aid clarity. The idea of keeping a tally leads us to consider the frequency of occurrence of a certain datum (singular of data, thus a piece of information). This is shown in the table below.

Tally and frequency table for the grape data.

Grapes per bunch	Tally	Frequency
1	√	1
2	√	1
4	√	1
5	√√	2
6	√√√	3
7	√√	2
8	√√√	3
9	√√	2
10	√√√√	4
11	√√√	3
12	√√√√√	5
13	√√√	3
14	√√√√	4
15	√√√	3
16	√	1

Overview
(/study/ap
122-
cid-
754029/

Grapes per bunch	Tally	Frequency
17	√√√	3
18	√	1
21	√	1

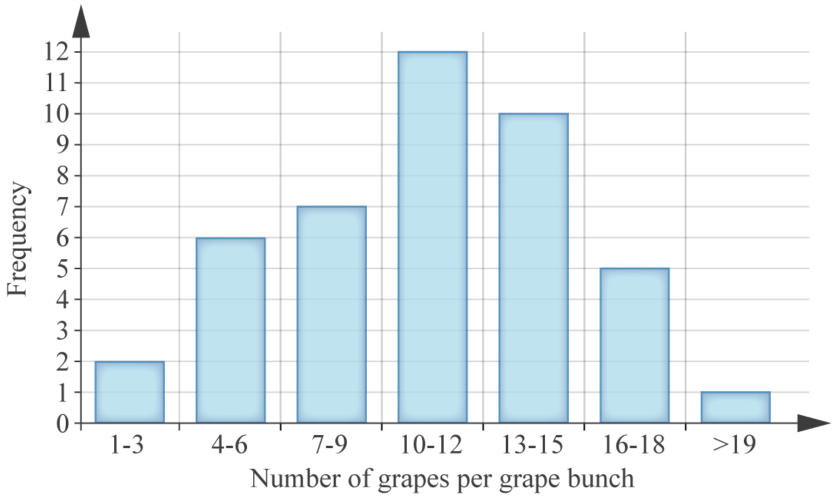
However, this does not necessarily make a pattern clear. For it to be clear, we can create bins in which we put data that share characteristics. In terms of quantitative data, this means that they fall in a range of numbers. For our grapes per bunch data, we can create bins, called **intervals** or **classes**, of, say, grapes per bunch of 1–3, 4–6, 7–9, 10–12, 13–15, 16–18 and more than 19. This leads to a **frequency distribution table**, see the table below for the number of grapes per bunch.

Tally and frequency of number for the grouped grape data.

Grapes per bunch	Tally	Frequency
1—3	√ √	2
4—6	√√√√√√	6
7—9	√√√√√√√	7
10—12	√√√√√√√√√√√√	12
13—15	√√√√√√√√√√	10
16—18	√√√√√	5
>19	√	1

Thus, the table above is a representation of the raw grape data that gives a clearer overview. A frequency table in turn leads to a **frequency diagram** as its graphical visualisation (see below).

Student view



The frequency diagram of the data shown in the table above. Note that the columns do not touch, since the data is discrete.

More information

The image is a bar chart displaying a frequency diagram of grape bunch data. The X-axis is labeled "Number of grapes per grape bunch," and it is divided into intervals: 1-3, 4-6, 7-9, 10-12, 13-15, 16-18, and >19. The Y-axis represents frequency, ranging from 0 to 12. The diagram visually represents how frequently grape bunches fall within these categories. The bars for each category do not touch, reflecting the discrete nature of the data. Noticeably, the highest frequency is at the 10-12 grouping, while the 1-3 and >19 groupings have the lowest frequencies.

[Generated by AI]

The number of grapes per bunch is *discrete data* , as we can *count it* . As a consequence, the columns in the diagram do not touch, there is no continuity between the columns, and a bunch cannot contain more than 3 but less than 4 grapes .

We can do the same for *continuous data* . For example, consider the height of students in the IB Diploma Programme at a particular school. The table below shows this data in the form of a frequency distribution table. The raw data has been grouped into classes.

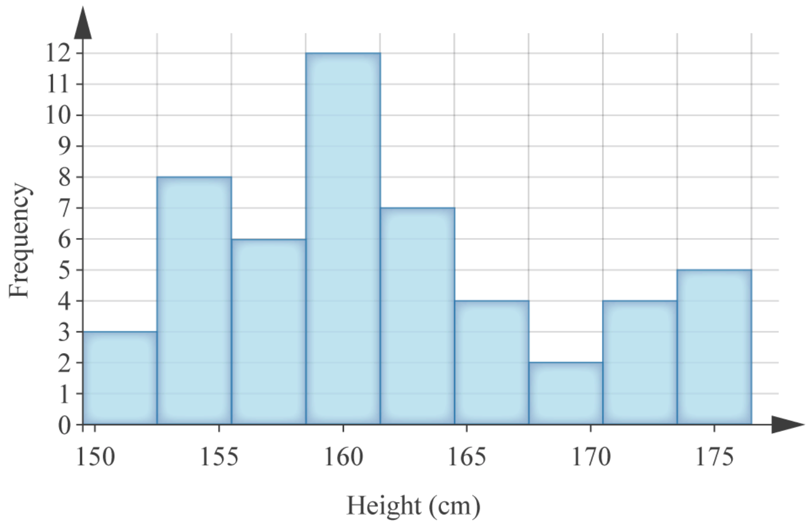
Frequency distribution table for heights of students.

Height h (cm)	**Tally**	**Frequency**
$149.5 \leq h < 152.5$	√√√	3
$152.5 \leq h < 155.5$	√√√√√√√√	8

Overview (/study/app/m/sid-122-cid-754029/...)

Height h (cm)	Tally	Frequency
$155.5 \leq h < 158.5$	√√√√√√	6
$158.5 \leq h < 161.5$	√√√√√√√√√√√√	12
$161.5 \leq h < 164.5$	√√√√√√√	7
$164.5 \leq h < 167.5$	√√√√	4
$167.5 \leq h < 170.5$	√√	2
$170.5 \leq h < 173.5$	√√√√	4
$173.5 \leq h < 176.5$	√√√√√	5

You can again present this information on a diagram. This type of diagram is called a **histogram**. In this course you will only meet histograms where the grouping intervals have the same width. In this case the height of the rectangles is proportional to the frequency.



The histogram for the distribution of heights given in the frequency distribution table. Note that the columns touch, since the data is continuous.

More information

The histogram represents the distribution of heights. The X-axis is labeled "Height (cm)" and ranges from 150 to 180 cm in intervals of 5 cm. The Y-axis is labeled "Frequency," ranging from 0 to 12. Each bar's height represents the frequency of heights in each interval. The bars are touching, indicating that the data is continuous. The first bar indicates a frequency of 3 for the height range 150-155 cm. The highest bar appears in the 160-165 cm range, reaching a frequency of 10. There are visible decreases and increases in the subsequent bars, showing variation in frequencies across different height ranges.

[Generated by AI]

Student view

Overview (/study/app/m/sid-122-cid-754029/

The height data is *continuous data* as it is *measured* . As a consequence, the columns in the histogram touch, there is a continuity between the set $149.5 \leq h < 152.5$ cm and $152.5 \leq h < 155.5$ cm, i.e. a student with a measured height of 152.4 cm falls into the first interval (or bin), whereas a student with a measured height of 152.5 cm falls into the second interval.

In either discrete or continuous data, another interesting representation is the **cumulative frequency table** (thus, distribution) where now the tally in each bin is the sum of all tallies up to and including that bin. By adding a cumulative frequency column to the table for the data of the 51 students above to include the cumulative frequency, we get the following results.

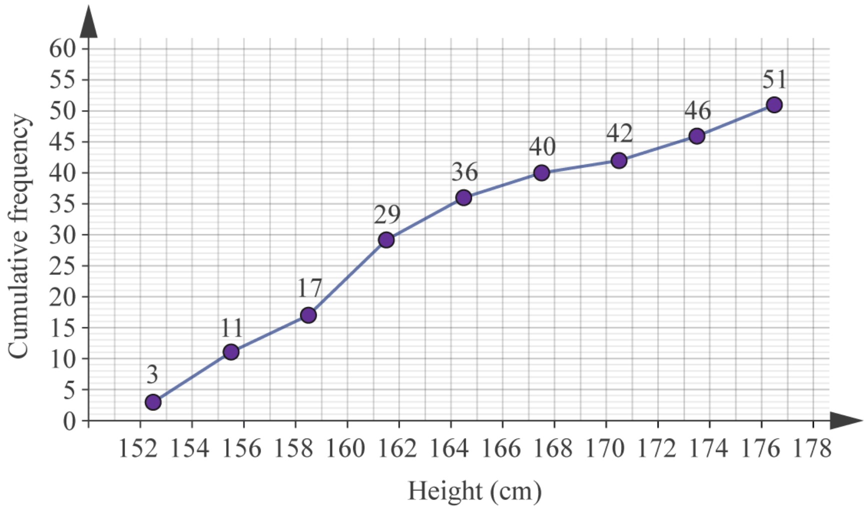
Cumulative frequency distribution table for heights.

Height h (cm)	Frequency	Cumulative frequency
$149.5 \leq h < 152.5$	3	3
$152.5 \leq h < 155.5$	8	$8 + 3 = 11$
$155.5 \leq h < 158.5$	6	$8 + 3 + 6 = 17$
$158.5 \leq h < 161.5$	12	$8 + 3 + 6 + 12 = 29$
$161.5 \leq h < 164.5$	7	$8 + 3 + 6 + 12 + 7 = 36$
$164.5 \leq h < 167.5$	4	$8 + 3 + 6 + 12 + 7 + 4 = 40$
$167.5 \leq h < 170.5$	2	$8 + 3 + 6 + 12 + 7 + 4 + 2 = 42$
$170.5 \leq h < 173.5$	4	$8 + 3 + 6 + 12 + 7+$ $+4 + 2 + 4 = 46$
$173.5 \leq h < 176.5$	5	$8 + 3 + 6 + 12 + 7+$ $+4 + 2 + 4 + 5 = 51$

The **cumulative frequency graph** is shown below.

Student view

Overview
(/study/ap
122-
cid-
754029/



The cumulative frequency graph for the cumulative distribution of heights given in the table above.

More information

This is a cumulative frequency graph representing the distribution of heights as given in the accompanying table (Table 4). The X-axis is labeled 'Height (cm)' and ranges from 150 to 180 centimeters with intervals of 2 centimeters. The Y-axis is labeled 'Cumulative frequency,' ranging from 0 to 60 with intervals of 5.

Data points are marked with purple dots and connected by a line, showing cumulative frequencies at different height intervals. The curve starts at 152.5 cm with a cumulative frequency of 3, then progressively increases through the following points: (155.5 cm, 11), (158.5 cm, 17), (161.5 cm, 29), (164.5 cm, 36), (167.5 cm, 40), (170.5 cm, 42), (173.5 cm, 46), and finally (176.5 cm, 51).

The graph illustrates the increasing trend of cumulative frequency as height increases, indicating a continuous accumulation of individuals within the specified height ranges.

[Generated by AI]

Note that the value of each interval or class or bin has been assigned to the upper bound of this interval. For example, the value 11 of the second interval $152.5 \leq h < 155.5$ is associated with the upper bound 155.5 in the graph.

We will see the cumulative frequency distribution in more detail in section 4.2.3 (/study/app/m/sid-122-cid-754029/book/cumulative-frequency-curves-id-26230/).

For data grouped into intervals or classes, we may identify the following:

1. mid-interval values
2. interval width (though it is not common to have a varying interval width)
3. lower interval boundary

Student
view

Overview (/study/ap 122- cid- 754029/

4. upper interval boundary
5. modal class (the class with the highest frequency or the tallest class in the diagram; be aware, use the tallest class in the frequency diagram, **not** in the cumulative frequency diagram).

Thus, for the number of grapes per bunch data, we have:

1. mid-interval values: $2, 5, 8, 11, 14, 17$
2. interval width: $3 - 1 = 2$
3. lower interval boundaries: $1, 4, 7, 10, 13, 16, 19$
4. upper interval boundaries: $3, 6, 9, 12, 15, 18, \infty$
5. modal class of grapes per bunch is 10–12 grapes.

Similarly, for the heights data, we have:

1. mid-interval values: $151, 154, 157, 160, 163, 166, 169, 172, 175$ (cm)
2. interval width: 3 cm
3. lower interval boundary: $149.5, 152.5, 155.5, 158.5, 161.5, 164.5, 167.5, 170.5, 173.5$ (cm)
4. upper interval boundaries: $152.5, 155.5, 158.5, 161.5, 164.5, 167.5, 170.5, 173.5, 176.5$ (cm)
5. modal class of heights is 159–161, i.e. heights between 158.5 and 161.5 (cm), including 158.5 but not including 161.5.

Be aware

Once your (raw) data has been grouped, it doesn't include the actual values of the data points. Thus, if there are 7 people with heights between 161.5 and 164.5 cm, we do not know where they lie in that interval. If we must assign a value to all 7, we must select a value. Normally we choose the mid-interval value, i.e. these 7 data points will be given the value 163 cm.

The total number of grapes actually counted using the raw data was 470. From the grouped data, we do not get the same total. Using 20 as an estimate of the highest interval:

$$\sum_{i=1}^{7} f_i \times x_i = 2 \times 2 + 6 \times 5 + 7 \times 8 + 12 \times 11 + 10 \times 14 + 5 \times 17 + 1 \times 20 = 467,$$

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

Similarly, the total height of the students in the modal class is $12 \times 160 = 1920$, where again we have taken the mid-interval value of the class as representative of the value of all the members of that class.

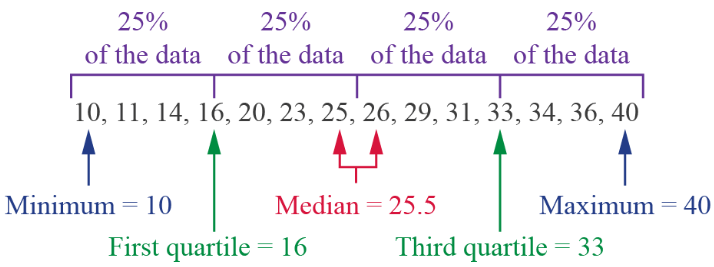
Student view

Overview (/study/ap 122-cid-754029/

Quartiles

It is often helpful to examine the values within a data set that fall above a certain percentage of the rest of the data. When you do this, it is known as finding the percentile of that value. Often, particular attention is paid to the highest and lowest values and the quartiles, the values at the 25th, 50th and 75th percentiles. Each of these values has a certain name:

- The lowest value of a data set is the minimum .
- The 25th percentile is known as the first quartile (Q_1).
- The 50th percentile is known as the second quartile, but is better known as the median .
- The 75th percentile is known as the third quartile (Q_3).
- The highest value of the data set is the maximum .



More information

This diagram visually represents the five-number summary of a data set used to make box-and-whisker plots. The data is divided into four segments, each comprising 25% of the total dataset.

- The minimum value is labeled as 10.
- The first quartile corresponds to a value of 16.
- The median is marked at 25.5.
- The third quartile indicates a value of 33.
- The maximum value is 40.

Data points leading up to these values are also included: 10, 11, 14, 16, 20, 23, 25, 26, 29, 31, 33, 34, 36, and 40. Each section is labeled to indicate that it represents 25% of the data. Arrows indicate the progression from the minimum to maximum value along the number line.

[Generated by AI]

Student view

Overview (/study/app/m/sid-122-cid-754029/...)

These five values together make up the five-number summary of the data set, which is used to make box-and-whisker plots, which we will explore in detail in section 4.2.3 (/study/app/m/sid-122-cid-754029/book/cumulative-frequency-curves-id-26230/).

To find Q_1, recognise that it is the middle of the lower half of the data, meaning it is the median of all of the data that is below the median of the data set.

To find Q_3, recognise that it is the middle of the upper half of the data, meaning it is the median of all of the data that is above the median of the data set.

✓ **Important**

When there is an odd number of data points in the set, the median is one of those points. Do not use that value when finding Q_1 or Q_3. Only use the data *below* or *above* the median.

Another measure you can use is what is known as the interquartile range (IQR). This is the range between Q_1 and Q_3. To find the IQR:

$$\text{IQR} = Q_3 - Q_1$$

Example 1

★★☆

Use the data in the table to find the five-number summary and the IQR for the data.

Value	Frequency	Cumulative frequency
2	4	4
4	7	11
6	8	19
7	5	24
10	2	26

To find the five-number summary, you need the minimum, maximum, Q_1, Q_3 and the median.

- The minimum is 2, and the maximum is 10.

Student view

Overview (/study/ap 122- cid- 754029/

- There are 26 values in the data set, so the median is the mean of the 13th and 14th values. Since both of those values equal 6, the median is 6.
- Q_1 is the median of the 13 values that are below the median, making it the 7th value in the set, which is 4.
- Q_3 is the median of values 14 through 26. There are 13 values in this group, so the median is the 7th value in the group. This makes Q_3 the 20th value in the data set, which is 7.
- Therefore, the five-number summary is:
  - minimum =*2*
  - *Q 1 =4*
  - median =*6*
  - *Q 3 =7*
  - maximum =*10* .
- Finally, $\text{IQR} = Q_3 - Q_1 = 7 - 4 = 3$.

2 section questions

4. Probability and statistics / 4.2 Presentation of data

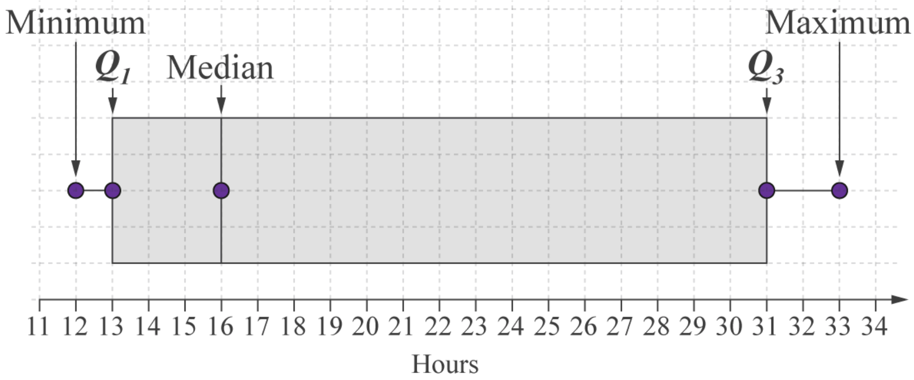
Box and whisker plots

Box-and-whisker plots

Constructing a box-and-whisker plot

In the previous section, you learned how to find a five-number summary for a set of data. One model that this summary is used to construct is the box-and-whisker plot(also known as the boxplot). To illustrate this, see the box-and-whisker plot below, which models the data set $\{12, 13, 13, 14, 15, 16, 18, 19, 31, 31, 33\}$.

Student view



More information

This image is a box-and-whisker plot constructed for the data set (\left { 12,13,13,14,15,16,18,19,31,31,33 \right }). The plot is displayed on a horizontal axis labeled as "Hours," ranging from 11 to 34. Key components of the plot include:

- **Minimum Value**: Located at 12.
- **First Quartile (Q1)**: Positioned at 13, marking the lower edge of the box.
- **Median (Q2)**: Positioned at 16, indicated by a line within the box.
- **Third Quartile (Q3)**: Positioned at 31, marking the upper edge of the box.
- **Maximum Value**: Located at 33.

The box itself spans from Q1 (13) to Q3 (31), with the whiskers extending from the minimum (12) to maximum values (33). This representation illustrates the distribution and spread of the data set based on the five-number summary.

[Generated by AI]

As you compare the box-and-whisker plot above to the set of data, you can see that the box-and-whisker plot is constructed using five primary values – the five-number summary. This box-and-whisker plot illustrates the following:

- $\text{minimum} = 12$
- $Q_1 = 13$
- $\text{median} = 16$
- $Q_3 = 31$
- $\text{maximum} = 33$

Notice that Q_1 and Q_3 make up the ends of the box, and you draw lines from Q_1 to the minimum and from Q_3 to the maximum. These lines are often called the whiskers, which is why this model is known as a box-and-whisker plot.

Student view

But that is not all that the box-and-whisker plot shows us.

Overview
(/study/ap
122-
cid-
754029/

The applet below is a helpful tool that you can use to construct box-and-whisker plots for various sets of data. Give it a try.

Interactive 1. Constructing Box and Whisker Plot.

Credit: GeoGebra (https://www.geogebra.org/m/DQcs8CYw) Steve Phelps

More information for interactive 1

This interactive allows users to create box-and-whisker plots by entering their data values. Users can input a set of numbers separated by commas between curly braces { } . The interactive then generates a box-and-whisker plot based on the five-number summary (minimum, first quartile, median, third quartile, and maximum) calculated from the data.

For example: For the dataset $\{4, 5, 9, -10, 4, -11, 15\}$, the tool calculates key statistics and generates a box-and-whisker plot showing:

Minimum: −11 (left whisker)

- First quartile (Q1): The value separating the lowest 25% of data
- Median: The middle value of a dataset that has been ordered from smallest to largest.
- Third quartile (Q3): The value separating the highest 25% of data
- Maximum: 15 (right whisker)

The plot visually represents the data spread, with the box showing the interquartile range (Q1 to Q3) and whiskers extending to the minimum and maximum values. This helps users quickly analyze the distribution and identify outliers. Users can reset and enter new data at any time.

Users can visualize the distribution of their data, observe the spread, and identify key statistical measures. This tool provides a hands-on way to explore and understand how data is represented in box-and-whisker plots.

Activity

Student
view

Overview (/study/ap 122- cid- 754029/

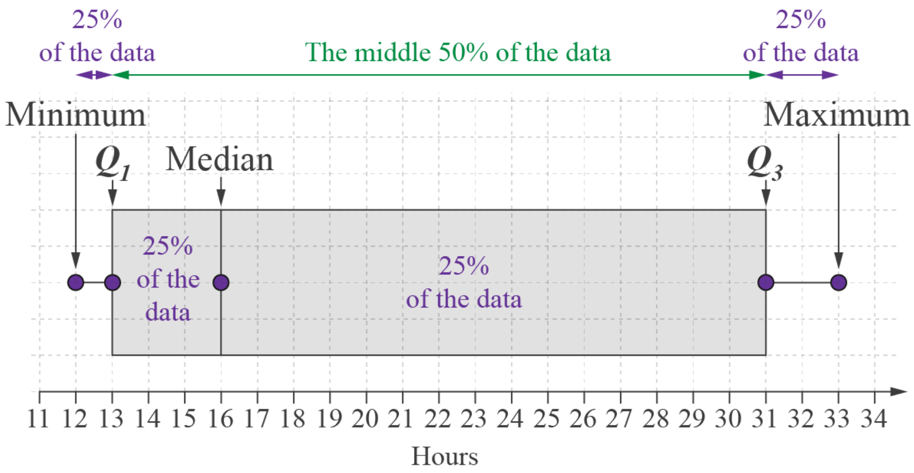
Measure the height of all the students in your class and create a box-and-whisker plot to illustrate the spread of values. What is the five-number summary?

International Mindedness

What might a box-and-whisker plot look like for the average rainfall in your country? How do you think this might compare to the box-and-whisker plots of average rainfall in other countries around the world?

Interpreting box-and-whisker plots

Remember what the quartiles stand for: Q_1, the median, and Q_3 represent the 25th, 50th and 75th percentiles, respectively. Considering this, we can tell that the lower whisker (on the left) represents 25% of the data, the part of the box between Q_1 and the median represents another 25% of the data, the part of the box between the median and Q_3 represents another 25% of the data, and the upper whisker (on the right) represents the last 25% of the data. With this in mind, we can see the data like this:



More information

The image shows a box-and-whisker plot, which provides a summary of a dataset's distribution. The X-axis is labeled 'Hours' and ranges from 11 to 34, with labels at intervals of 1 hour. The graph is divided into sections that represent different quartiles: the minimum value is marked at Hour 12, Q1 is at Hour 15, the median is at Hour 20, Q3 is at Hour 31, and the maximum value is at Hour 34. Each section of the plot represents 25% of the data distribution: the left whisker from the minimum to Q1, the first section of the box from Q1 to the median, the second section of the box from the median to Q3, and the right whisker from Q3 to the maximum. Over the box, it mentions 'The middle 50% of the data,' indicating the range between Q1 and Q3. Additionally, annotations highlight that each whisker and box section represents 25% of the dataset, illustrating the distribution spread.

[Generated by AI]

Student view

Overview
(/study/app
122-
cid-
754029/

The box-and-whisker plot is also very useful to explore the **spread** of the data over the range, because it allows us to compare various portions of the data visually. For example, the left-hand whisker is very short, meaning the values in that 25% of the data are all close together. The portion of the box between the median and Q_3 is very wide, meaning that 25% of the data contains values that are more spread out from one another. Suppose the box-and-whisker plot above represents the times recorded in a long-distance race. The short left-hand whisker shows that the fastest 25% of the competitors were bunched together – this could be the group of favourites who were trying to win the race. On the other hand, the distance from the median to the maximum is very large, indicating a long tail of slower runners coming in after the main competitors.

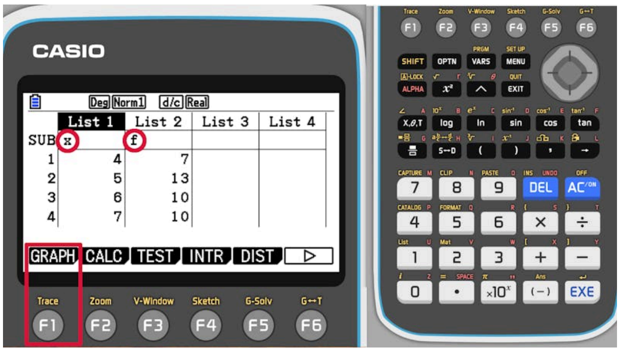
Making connections

The middle 50% of the data is often used to find a range of typical results within a given scenario. Some colleges and universities publish the middle 50% ranges for different test scores to show applicants the typical range for those who are accepted. For example, if a university states that the middle 50% range of IB Diploma points for applicants who were admitted is 30—35, then that means half of the students they accepted had a points score falling within that range. An applicant with fewer points is less likely to be accepted by that university.

When you are dealing with a large amount of data, especially if it is not already in numerical order, using a calculator can be quite helpful.

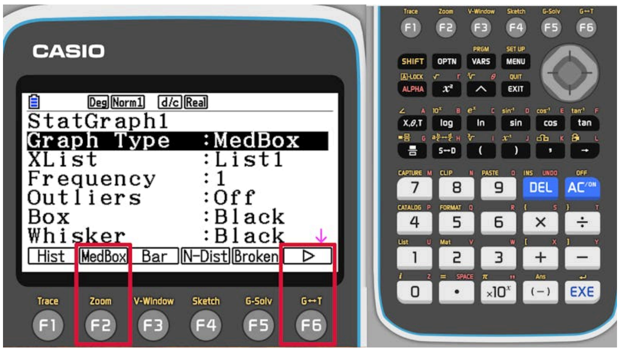
Student
view

Overview
(/study/ap
122-
cid-
754029/

Step	Explanation
Press 2 to open the statistics mode.	
Enter the data. Note, that you can also give a name to your data. In this example we use a data given using frequencies. Remember that you used List1 for storing the values and List2 for storing the frequencies, you will need this later. Once done entering the data, press F1.	

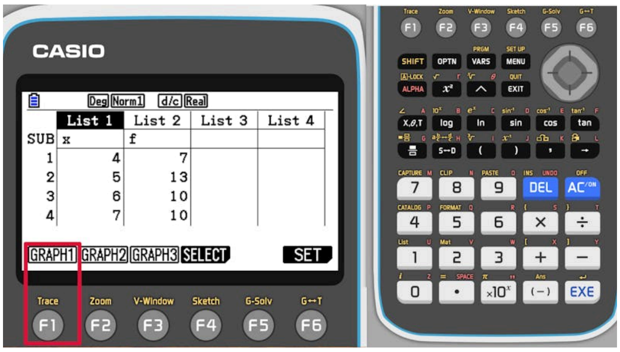
Student
view

Overview (/study/app/m/sid-122-cid-754029/

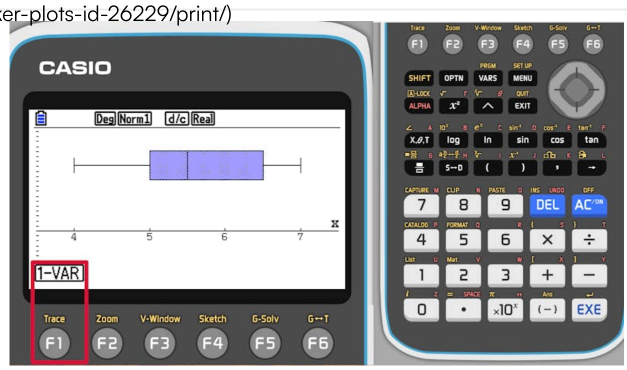
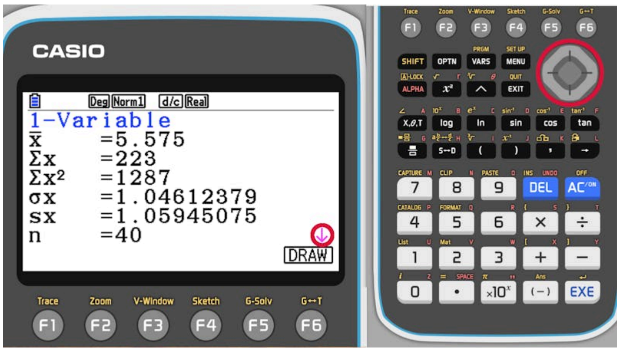
Step	Explanation
First you need to set some parameters, so press F6.	
The box and whisker plot is called MedBox on this calculator. To find this option you need to press F6 first to scroll to the right.	

Student view

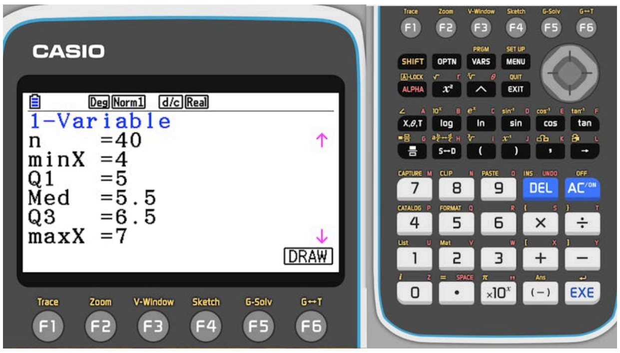
Overview
(/study/ap
122-
cid-
754029/

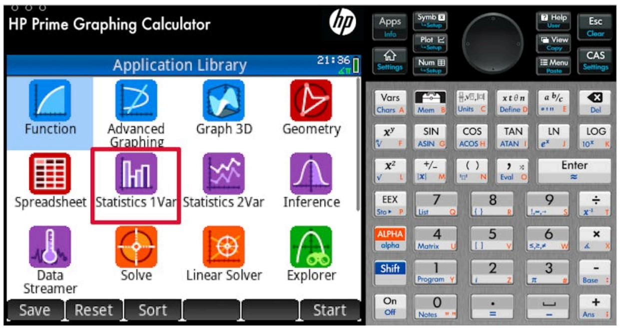
Step	Explanation
Remember, you stored the x-values in List1 and the frequencies in List2. Use F2 and the number keys to tell this to the calculator. The rest of the options is about the appearance of the box and whisker plot. Modify those too if you want. Once done, press EXE.	
The calculator is now ready to display the box and whisker plot of the data you entered.	

Student
view

Step	Explanation
You can read the five number summary from the plot, but the calculator also gives you an option to view a full statistical analysis of the data. Press F1 to see this summary.	
The arrow in the corner of the screen indicates that there is more information available. To access the five number summary you need to scroll down.	

Overview
(/study/ap
122-
cid-
754029/

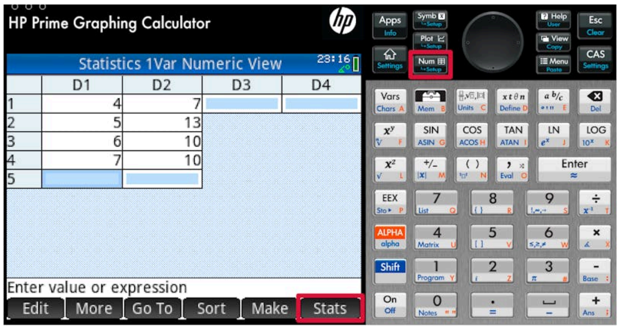
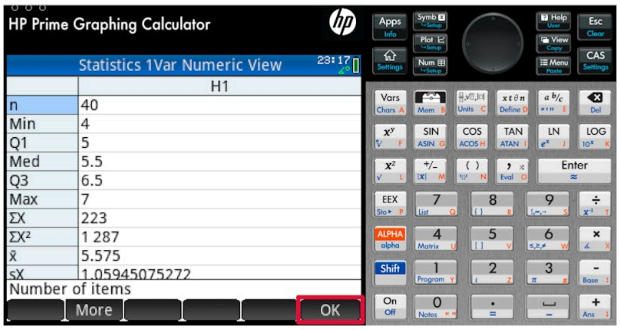
Step	Explanation
On this screen you can see the minimum, the maximum, the quartiles and the median. Notice that you also have an option to go back to see the plot.	

Step	Explanation
Choose the 1-variable statistics application.	

Section Student... (0/0) Feedback Print (/study/app/m/sid-122-cid-754029/book/grouped-data-and-quartiles-id-26228/print/) Assign

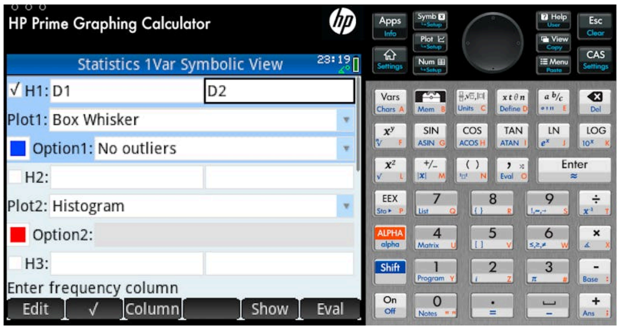
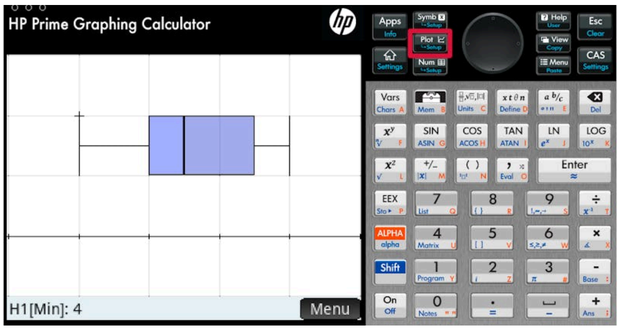
Student view

Overview
(/study/ap
122-
cid-
754029/

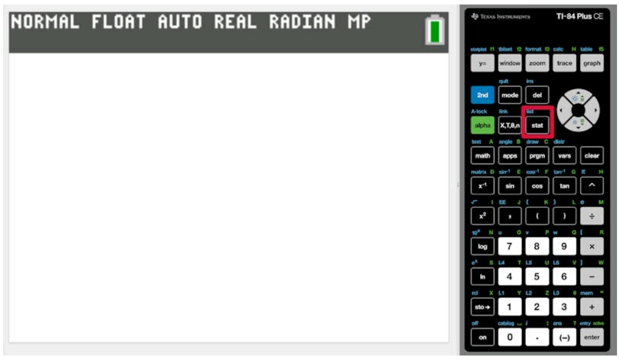
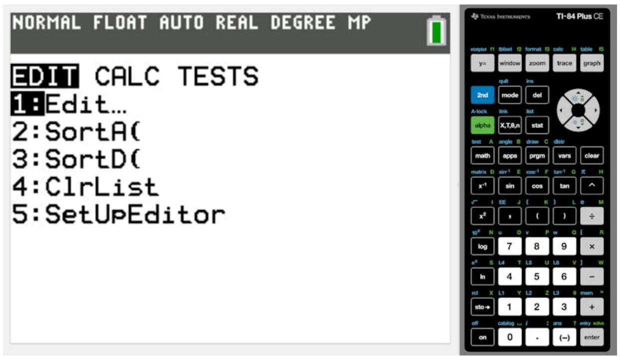
Step	Explanation
In the numeric view of the application, enter the data. In this example we use a data given using frequencies. Remember that you used D1 for storing the values and D2 for storing the frequencies, you will need this later. Once done entering the data, press Stats to view a full statistical analysis of the data.	
In the first few lines you can see the size of the data set and the five number summary (the minimum, the maximum, the quartiles and the median). There is of course more information available if you scroll down. Press OK when you finished evaluating the information.	
In the symbolic view you can choose to see a box and whisker plot. In the top row enter D1 to tell the calculator where you stored the data, ...	

Student
view

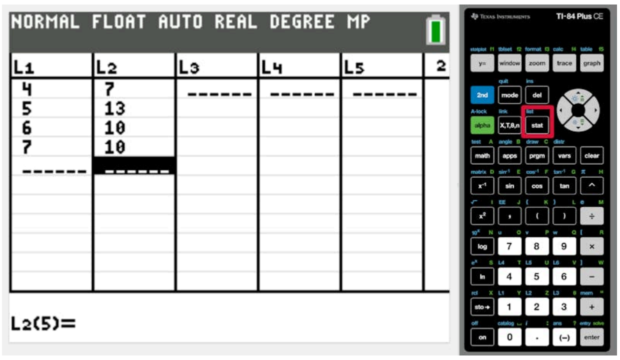
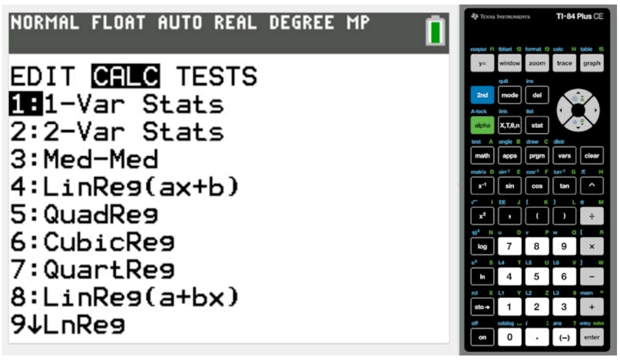
Overview
(/study/ap
122-
cid-
754029/

Step	Explanation
... and enter D2 to tell the calculator where the frequencies are.	
In the plot view you can see the box and whisker plot. You can move left and right with the cursor to see the values in the five number summary. Note, that you will probably need to use the plot setup to specify the viewing window.	

Student
view

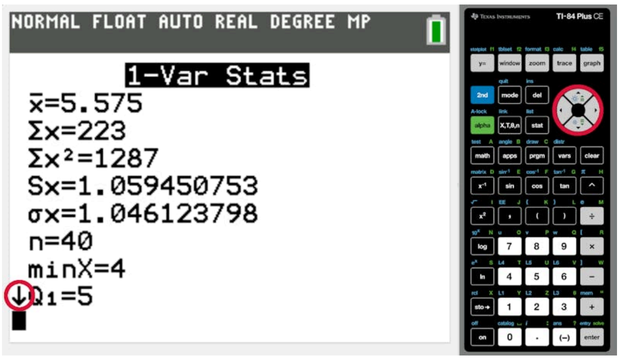
Step	Explanation
Press stat to start working with data sets.	
First you need to enter your data, so choose to bring up the editing screen.	

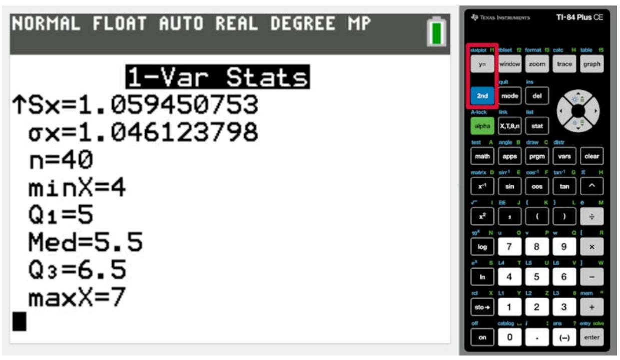
Overview
(/study/ap
122-
cid-
754029/

Step	Explanation
Enter the data. In this example we use a data given using frequencies. Remember that you used L1 for storing the values and L2 for storing the frequencies, you will need this later. Once done entering the data, press stat again.	
Choose the option to calculate the 1-variable statistics.	

Student
view

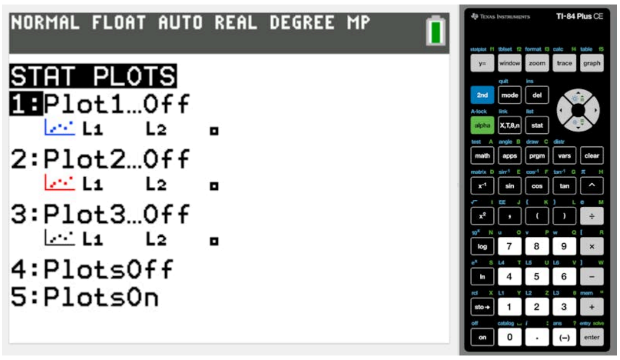
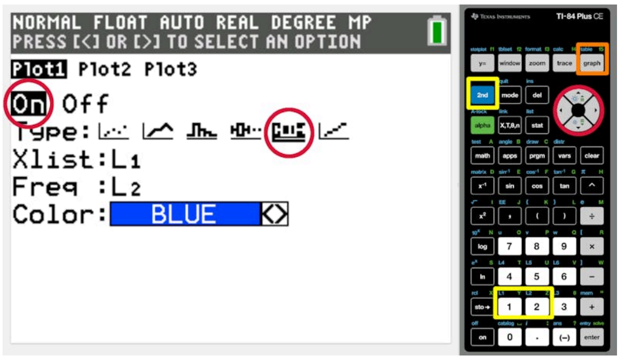
<table>
<tr><th>Step</th><th>Explanation</th></tr>
<tr><td>
Remember, you stored the x-values in L1 and the frequencies in L2. Use 2nd and the number keys to tell this to the calculator.
Once done, move down to the last row and press enter.
</td><td>


</td></tr>
<tr><td>
The arrow in the corner of the screen indicates that there is more information available. To access the five number summary you need to scroll down.
</td><td>


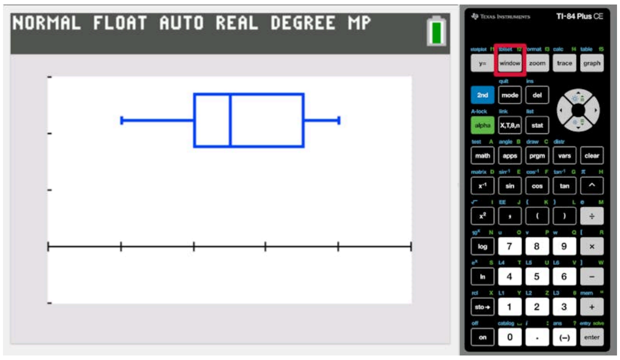
</td></tr>
<tr><td>
On this screen you can see the minimum, the maximum, the quartiles and the median.
When you finished evaluating the information, choose the stat plot option to draw a box and whisker diagram.
</td><td>


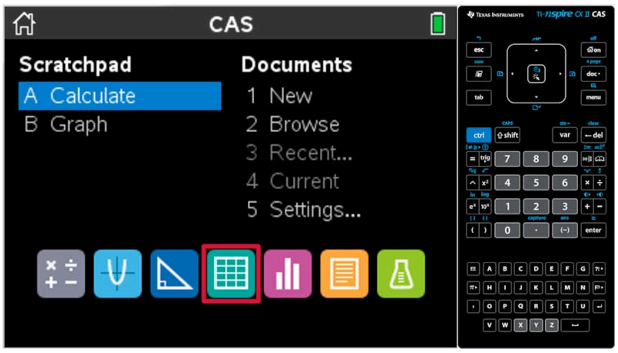
</td></tr>
</table>

Overview
(/study/ap
122-
cid-
754029/

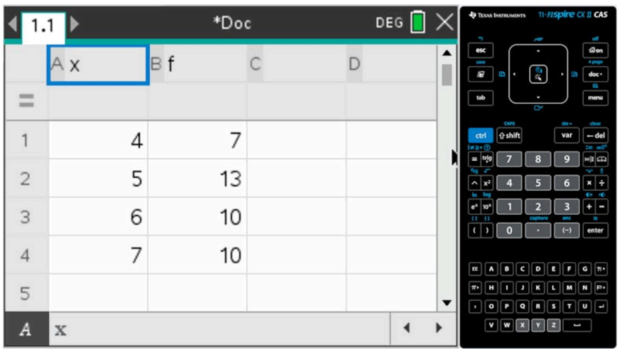
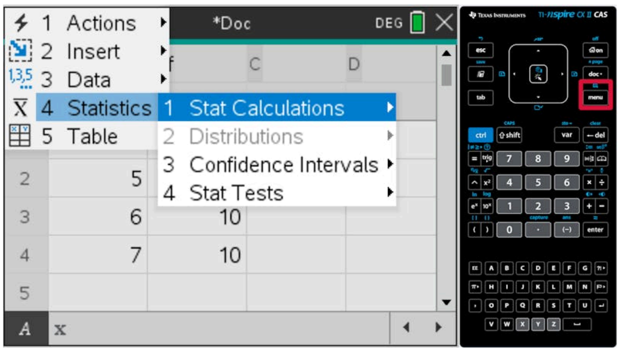
Step	Explanation
The calculator can display several plots, choose any of these.	
Navigate to the appropriate positions to turn the plot on and choose box and whisker plot as the type. Remember, you stored the x-values in L1 and the frequencies in L2. Use 2nd and the number keys to tell this to the calculator. Once you are done setting the options, press graph to see the plot.	

Student
view

Step	Explanation
Note, that you will probably need to use the window button to specify the viewing window.	

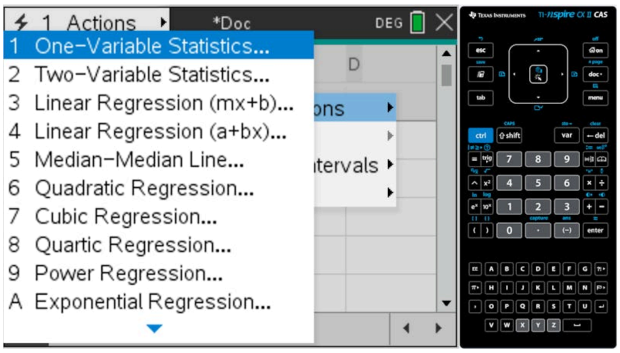
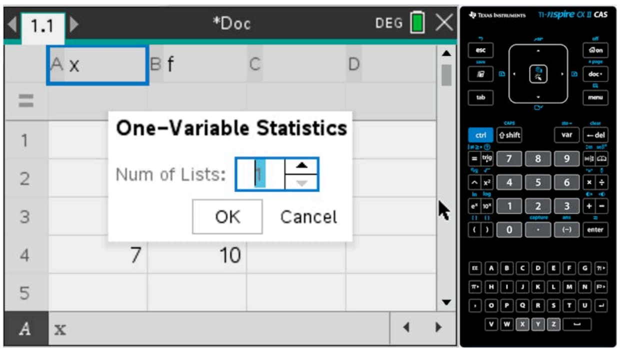
Step	Explanation
To start, open a spreadsheet document.	

Overview
(/study/ap
122-
cid-
754029/

Step	Explanation
Enter the data. Note, that you can also give a name to your data. In this example we use a data given using frequencies. Remember that you used a list named x for storing the values and a list named f for storing the frequencies, you will need this later.	
To see the five number summary, open the menu...	

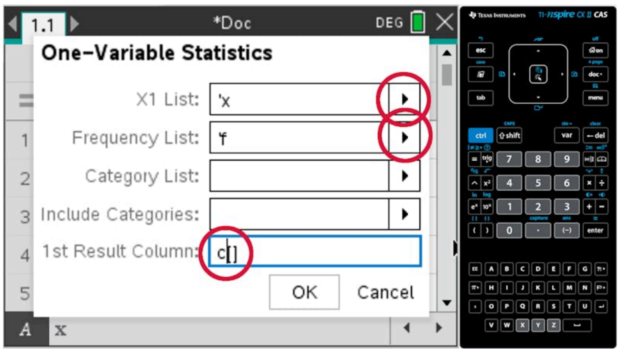
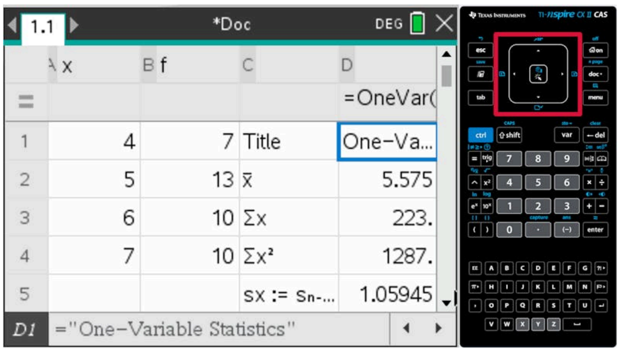
Student
view

Overview
(/study/ap
122-
cid-
754029/

Step	Explanation
... and choose one-variable statistics.	
You have one list (you will specify the frequencies later).	

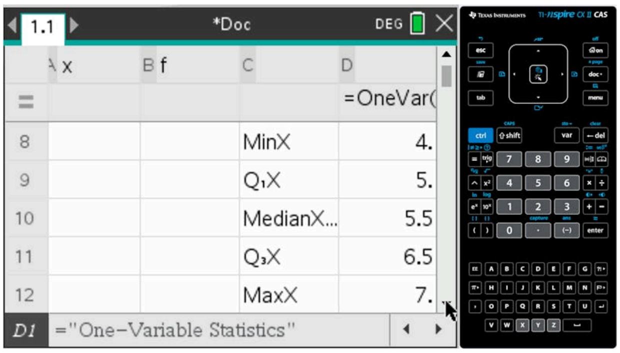
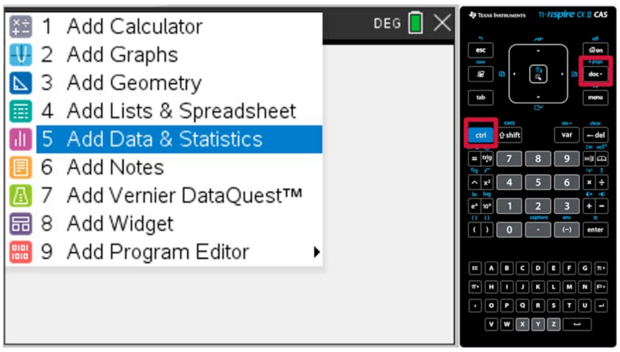
Student
view

Overview
(/study/ap
122-
cid-
754029/

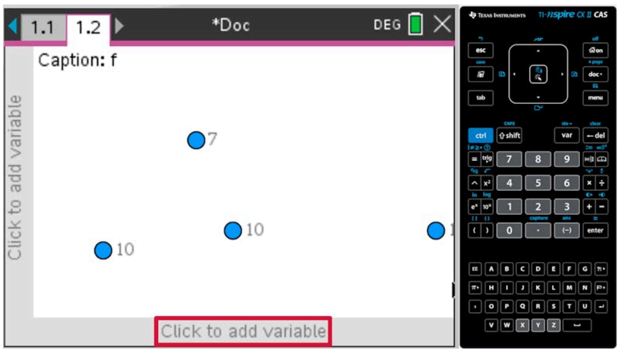
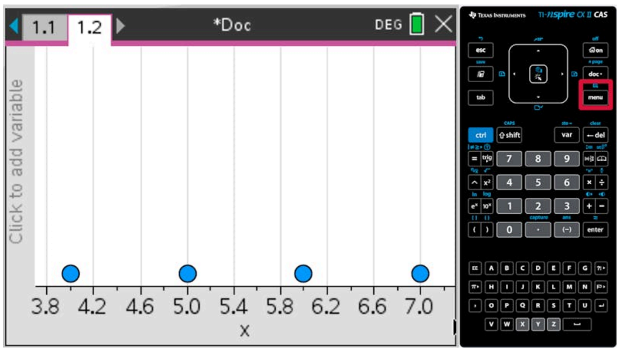
Step	Explanation
Choose these names you used to store the x-values and the frequencies from the pull-down menu. The results will be stored in the spreadsheet. Since the data is stored in the first two columns (a and b), choose a different column for the result.	
To access the five number summary you need to scroll down.	

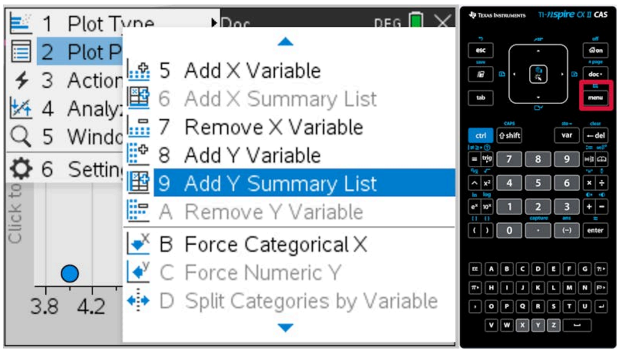
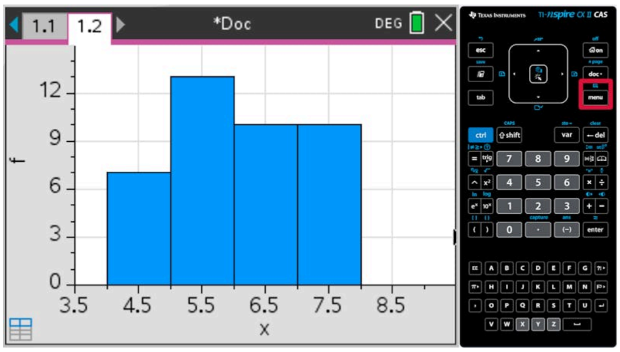
Student
view

Overview
(/study/ap
122-
cid-
754029/

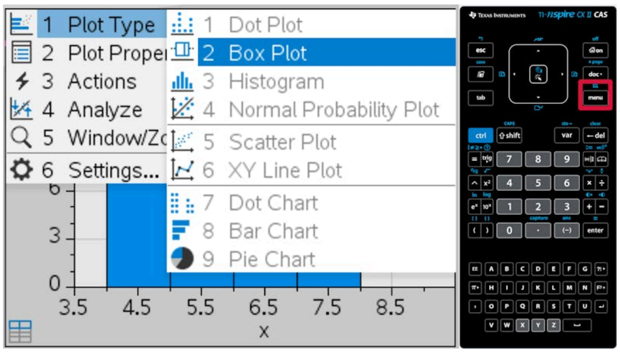
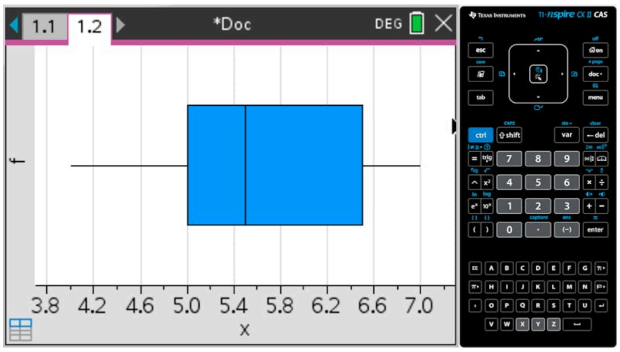
Step	Explanation
On this screen you can see the minimum, the maximum, the quartiles and the median.	
To draw the box and whisker plot, you need to add a new page.	

Student
view

Step	Explanation
What you see is of course not the box and whisker plot. You need to tell the calculator several options before it can draw the plot. Start with specifying x as the variable (since this is the name you used for the data).	
The calculator still does not know about the frequencies. Adding this information can be done through the menu.	

Step	Explanation
Choose to add a Y summary list (and use the name you gave to the frequency column in the spreadsheet).	
What you can see is still not a box and whisker plot. You can change the plot type through the menu again.	

Overview
(/study/ap
122-
cid-
754029/

Step	Explanation
Choose box plot.	
	

While it is also possible to use your GDC to construct a box-and-whisker plot, once you have the five-number summary, the box-and-whisker plot is relatively easy to construct by hand.

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, and once the analysis is done by the GDC, check that the number of data points (n) matches the number of data points in the question.

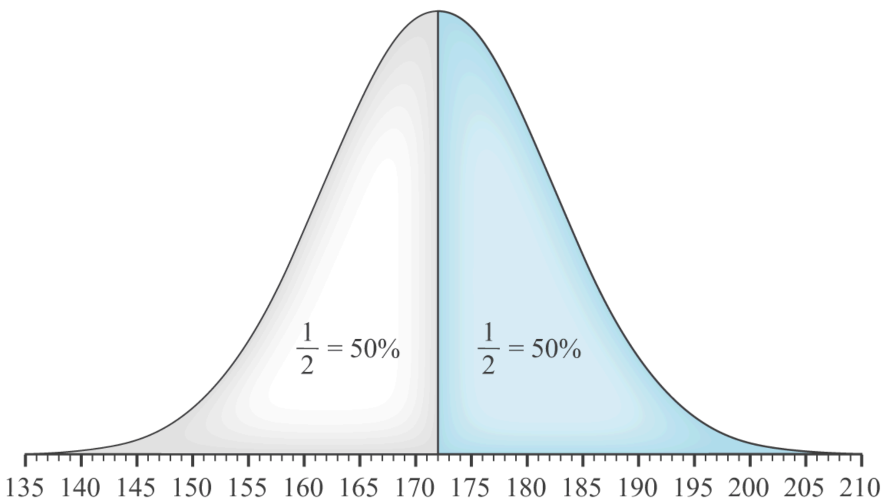
Student
view

Overview (/study/app/m/sid-122-cid-754029/

Distribution and skewing of the mean

Another aspect of spread that a box-and-whisker plot can indicate is whether the data is approximately **normally distributed**, **positively skewed** or **negatively skewed**.

Consider data found by measuring the heights of a random sample of people. In general, people tend to be close to average height, while some are a little taller, a few are much taller, some others are a little shorter than average and a few more are very short. Chances are, you have seen this other similar characteristics depicted in a bell curve like the one shown in below. This distribution of data is called a normal distribution , because as the measurements get further from the mean, the frequency of those measurements decreases.



More information

The image depicts a bell curve representing a normal distribution with a centered peak. The X-axis, shown at the bottom, is labeled with values ranging from 135 to 210. The peak of the curve is aligned with the average value around 172.5. The curve is divided into two symmetrical halves, each representing 50% of the data, with a label indicating "1/2 = 50%" on each side. The curve smoothly tapers off towards the ends, illustrating that most data points cluster around the mean, while fewer measurements are found towards the extremes.

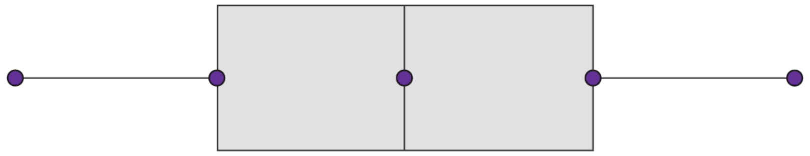
[Generated by AI]

When data is approximately in a normal distribution , the mean and median will be close to the same value and the rest of the data will be evenly dispersed on either side, as the bell curve above shows. When you have a box-and-whisker plot, you can determine if it represents an approximately normal

Student view

Overview (/study/app/m/sid-122-cid-754029/

distribution by recognising symmetry between the two halves of the box and seeing whiskers that are approximately the same length. Below you can see a box-and-whisker plot of the same data that gave us the bell curve above.



More information

The image displays a box-and-whisker plot. The plot consists of a central box divided into two parts, representing the interquartile range (IQR) of the data. The line inside the box shows the median of the data. Extending horizontally from each end of the box are whiskers, which represent the range of the dataset outside the IQR. In this plot, the whiskers are of equal length, indicating symmetry in the data distribution. Each whisker ends with a dot, which highlights the maximum and minimum values within the dataset. The symmetry of the box and equal length of the whiskers suggest an approximately normal distribution of the data.

[Generated by AI]

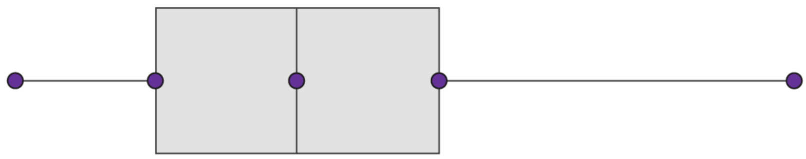
Activity

Think of other types of data that might follow a normal distribution. How could you verify your answers?

If the data contains data that is mostly similar but some is much higher than the rest, then the mean of the data will be significantly higher than the median. This data is therefore positively skewed. The box-and-whisker plot below shows what would happen if the sample of people included a few very tall people.

Student view

Overview
(/study/ap
122-
cid-
754029/

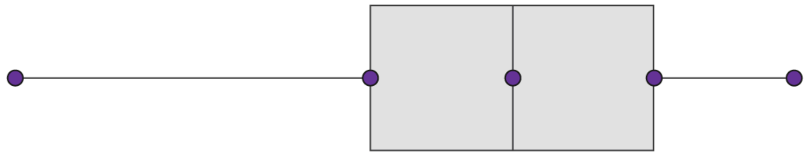


More information

The image is a box-and-whisker plot. The plot has a horizontal line representing the data range from the minimum to the maximum value. The box in the plot captures the interquartile range, with vertical lines marking the first quartile (Q1), median (Q2), and third quartile (Q3). The median is near the center of the box, indicating its distribution. Whiskers extend from the edges of the box to the smallest and largest data points, not including outliers, if any. The context provided suggests this plot represents a data set with a few very tall people, causing a positive skewness in the distribution. This implies that the median is lower than the mean, and the right whisker is longer than the left one.

[Generated by AI]

On the other hand, if the data contains data that is mostly similar but some is much lower than the rest, then the mean of the data will be significantly lower than the median. The box-and-whisker plot below represents data that is negatively skewed due to the sample including a few very short people.



More information

The image is a box-and-whisker plot which represents a dataset that is negatively skewed. The X-axis represents an undefined variable, as no specific labels or units are visible in the image. The Y-axis likely represents the frequency or occurrence, but specific intervals or scale markers are not provided.

The box plot consists of a central rectangle, a line inside the rectangle, and two whiskers extending from either end. Five data points are marked along the axis:

Student
view

Overview (/study/app/m/sid-122-cid-754029/

1. The leftmost point (start of the left whisker) marks the minimum value, representing very short individuals in the dataset.
2. The start of the box marks the first quartile (Q1).
3. A line inside the box marks the median (second quartile, Q2), which is higher than the mean due to the skew of the data.
4. The end of the box marks the third quartile (Q3).
5. The rightmost point (end of the right whisker) indicates the maximum value.

This configuration shows that most data points are clustered between Q1 and Q3, with a long tail extending to the left, suggesting a negative skew caused by a few much shorter individuals.

[Generated by AI]

Activity

Think of some contexts that might produce positively or negatively skewed data. How could you verify your answers?

Outliers

An outlier is a particularly unusual data point. But what determines whether the data is 'unusual' enough to qualify as an outlier? In one-variable statistics there is a general rule for determining whether a certain value should be considered an outlier or not by using the IQR. Any data that is more than $1.5 \times \text{IQR}$ below Q_1 is an outlier. Likewise, any data that is more than $1.5 \times \text{IQR}$ above Q_3 is also an outlier.

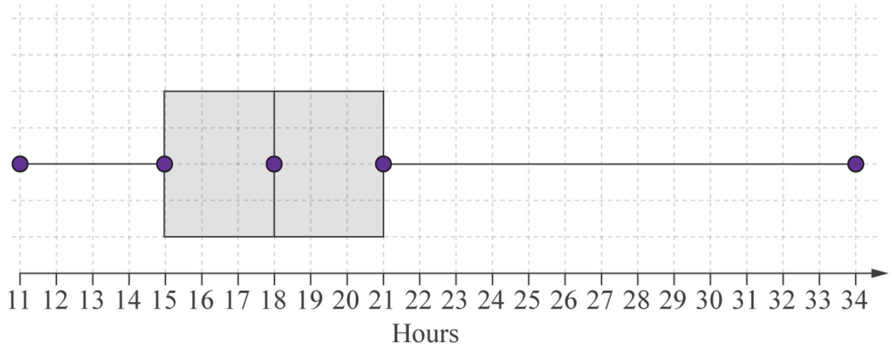
In other words:

$$\begin{gathered}\text{outlier} < Q_1 - 1.5 \times \text{IQR} \\ \text{or} \\ \text{outlier} > Q_3 + 1.5 \times \text{IQR}\end{gathered}$$

If outliers are indicated on a box-and-whisker plot, then the whiskers may no longer extend to the maximum or minimum value. Outliers can be indicated with a cross (x), although you may also see people use an asterisk (*) or even just a point. If you do not indicate when an outlier exists, you can easily draw false conclusions about the data. To illustrate this, consider the following box-and-whisker plot that does not show the outliers.

Student view

Overview
(/study/ap
122-
cid-
754029/

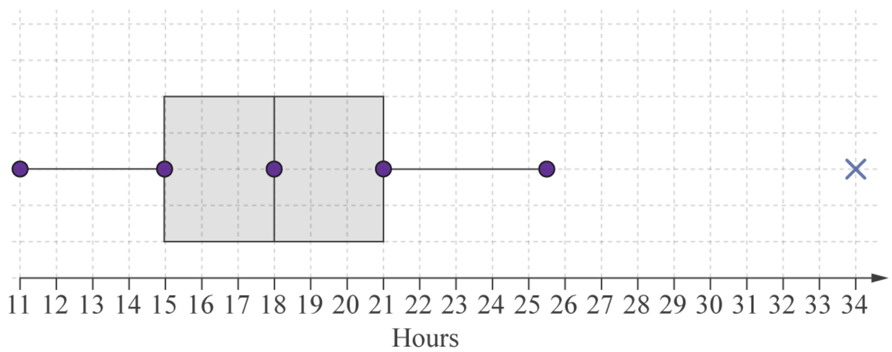


More information

The image is a box-and-whisker plot that represents a data set. The X-axis is labeled 'Hours,' ranging from 11 to 34. The box plot shows the spread of data between the quartiles, specifically between approximately 14 and 21 hours. The median is marked within the box, around 18.5 hours. Whiskers extend to the minimum non-outlier data point on the left and to an outlier on the right marked at 34 hours, shown with a purple dot. The main block of data (50%) spans from around 14 to 21 hours, with the interquartile range clearly visible. An outlier at 34 hours shifts the perception of data spread, extending beyond the general clustering of data within the central box.

[Generated by AI]

It appears as though the highest 25% of the data is very spread out, but is that really true? It may be, or there may be an outlier. The diagram below shows the box-and-whisker plot of the same data set, but reveals that there was a single outlier at 34. In this case the upper whisker extends to the largest data that is not an outlier.



More information

The image is a box-and-whisker plot illustrating the distribution of a dataset along a horizontal axis labeled 'Hours.' The scale runs from 11 to 34, which includes an identified outlier at 34. The box stretches from 14 to 22, representing the interquartile range (IQR), with median marked at 18.5. The 'whiskers' extend from the minimal value at 11 to the maximum non-outlier value at 25. Each

Student
view

Overview
(/study/ap
122-
cid-
754029/

quartile is represented by a different section of the plot, and the outlier is marked distinctly at 34, isolated from the main plot by an 'x.' The outlier signifies that the highest 25% of the data is not as spread out as initially assumed, with the largest value skewed by this single outlier.

[Generated by AI]

Now we can see that there is an outlier at 34 and that the highest 25% of the data is actually not spread out much more than the rest of the data.

Be aware

Graphing calculators can draw box-and-whisker plots with or without outliers. In either case the whiskers end on a point that is part of the data set. When outliers are indicated, whiskers are drawn to the last data point that is *not* an outlier.

4 section questions

4. Probability and statistics / 4.2 Presentation of data

Cumulative frequency curves

Section Student... (0/0) Feedback Print (/study/app/m/sid-122-cid-754029/book/cumulative-frequency-curves-id-26230/print/) Assign

Let us consider data collected from a survey of the members of a badminton club. Each of the 70 members was asked to record their age. The data was compiled into ranges and used to create the frequency table below, in which variable X represents a member’s age (see the table below **)**. We can use this table to explore a more precise way of obtaining the median, m, first quartile, Q_1, and third quartile, Q_3 , than simply taking the mid-interval values of the intervals in which they fall.

Frequency table for X.

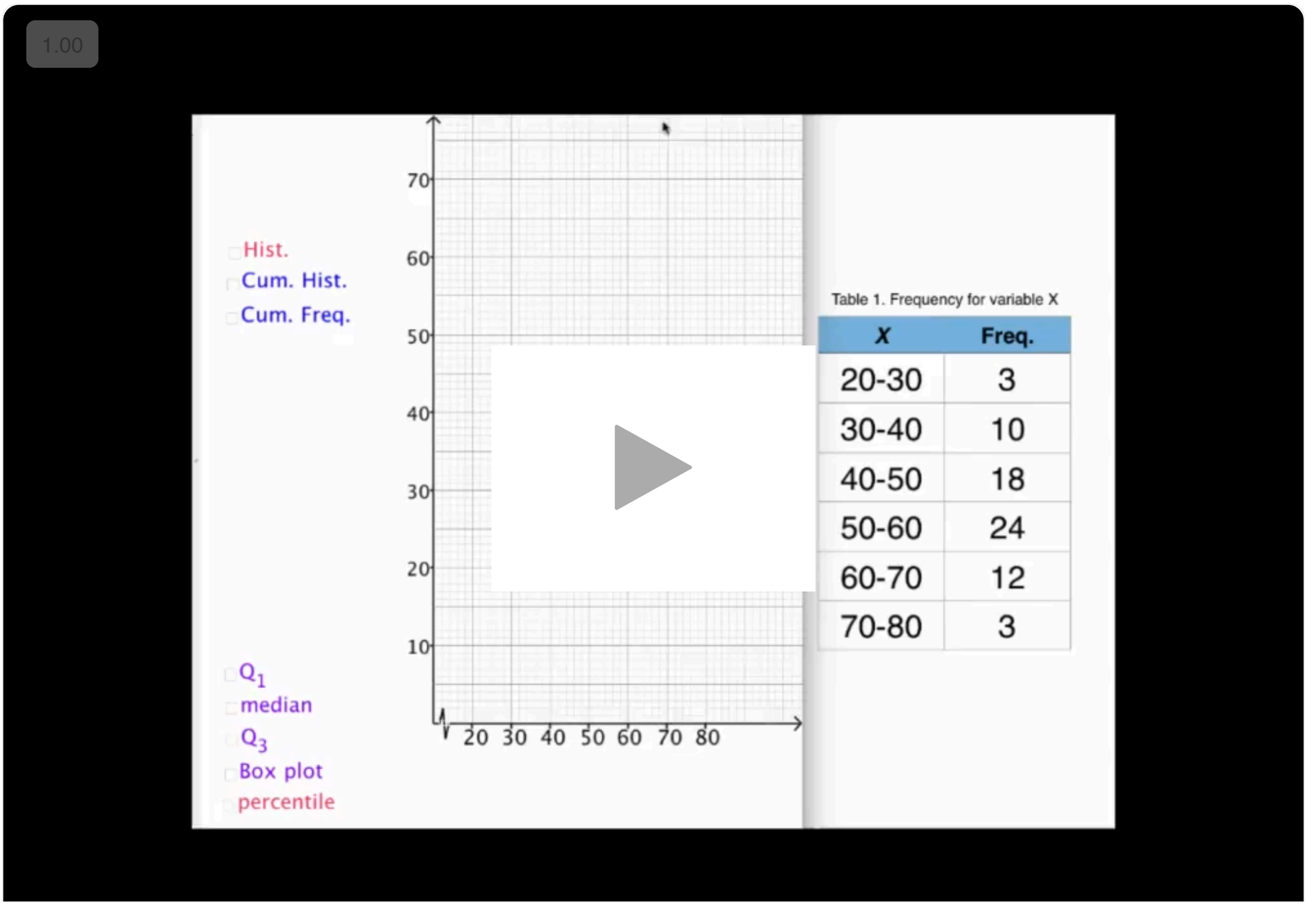
X	**Frequency**
$20—30$	3
$30—40$	10
$40—50$	18

Student view

Overview
(/study/ap
122-
cid-
754029/

X	Frequency
50—60	24
60—70	12
70—80	3

We do this in the following video.



Video 1. Understanding Cumulative Frequency Curves.

More information for video 1

1

00:00:00,967 --> 00:00:03,267

narrator: In this video we're going to

look at cumulative frequencies,

2

00:00:03,333 --> 00:00:06,067

but we're going to start

with a frequency table of a variable X

3

00:00:06,133 --> 00:00:07,533

given in certain classes

Student
view

Overview (/study/ap 122- cid- 754029/

4

00:00:07,600 --> 00:00:10,433

with the respective frequencies

given in a frequency column.

5

00:00:10,533 --> 00:00:13,167

And of course

this allows us to represent it

6

00:00:13,233 --> 00:00:16,200

with a histogram shown over here

where the height of the bars

7

00:00:16,267 --> 00:00:21,100

are given by the respective frequencies

in those class intervals.

8

00:00:22,467 --> 00:00:26,033

Now you can see that the central tendency

lies somewhere around 50,

9

00:00:26,100 --> 00:00:29,133

but looking at cumulative frequencies

allows us a much greater position.

10

00:00:29,600 --> 00:00:31,300

So to explore cumulative frequencies,

11

00:00:31,367 --> 00:00:34,533

we are going

to add another column to this table,

12

00:00:34,600 --> 00:00:38,367

and the entries in the column

are the frequencies added up

13

00:00:38,433 --> 00:00:39,633

that are before that.

14

00:00:39,733 --> 00:00:42,600

So 13, the second entry is $3 + 10$,

15

Student view

00:00:43,033 --> 00:00:46,767

31 is $3 + 10 + 18$, or of course $13 + 18$.

16

00:00:46,967 --> 00:00:51,167

55 is $3 + 10 + 18 + 24$ or $31 + 24$.

17

00:00:51,500 --> 00:00:55,900

67 is $55 + 12$ and 70 is $67 + 3$

18

00:00:55,967 --> 00:00:57,900

and 70 is a total number of data points.

19

00:00:58,033 --> 00:01:01,133

Now we can plot this to the cumulative

frequency histogram,

20

00:01:01,200 --> 00:01:03,667

and those are the blue bars

over here where you can see

21

00:01:03,733 --> 00:01:06,567

that it keeps on growing

because you keep adding the bar

22

00:01:06,633 --> 00:01:10,000

before that to the bar

that was the histogram before.

23

00:01:10,200 --> 00:01:14,567

So you can see here

how the blue one has made out of

24

00:01:14,633 --> 00:01:19,133

the previous blue bars

and the underlying red one.

25

00:01:20,400 --> 00:01:21,333

Now, why is this useful?

26

00:01:21,400 --> 00:01:23,500

Well, the median remember

is the data point

Overview
(/study/ap
122-
cid-
754029/

27

00:01:23,567 --> 00:01:26,733

is in the middle,

so the middle of 70 is 35.

28

00:01:26,933 --> 00:01:31,567

But to help us in here, we are going

to plot the cumulative frequency polygon.

29

00:01:31,900 --> 00:01:36,967

Now here, these points

are the end points of the interval.

30

00:01:37,033 --> 00:01:40,167

So that is important to realize you use

the end points of the interval,

31

00:01:40,233 --> 00:01:43,133

and that doesn't mean that we need

to create one at the very beginning,

32

00:01:43,467 --> 00:01:44,900

in this case, $(20, 0)$,

33

00:01:44,967 --> 00:01:48,133

which in a cumulative frequency polygon,

we joined by straight lines,

34

00:01:48,200 --> 00:01:49,733

you can also use a curve.

35

00:01:49,800 --> 00:01:52,600

And now of course we can use

a median which slides halfway

36

00:01:52,667 --> 00:01:56,500

to 70, which is 35,

which you can see lies around 52,

37

00:01:56,567 --> 00:02:00,500

which much greater position than we had

without a cumulative frequency polygon.

Student
view

Overview (/study/app/m/sid-122-cid-754029/

38

00:02:00,800 --> 00:02:03,967

Of course, we can also find Q1 and Q3,

39

00:02:04,033 --> 00:02:07,667

which lies at 25% and 75% up to 70.

40

00:02:07,733 --> 00:02:11,000

And here we have those values, 43, 52,

41

00:02:11,067 --> 00:02:12,967

and about 59,

42

00:02:13,700 --> 00:02:16,567

and of course, with the endpoints

as well at 20 and 80,

43

00:02:16,633 --> 00:02:18,833

we can now plot our box on whisker plots

44

00:02:18,900 --> 00:02:22,333

with the whiskers at the ends,

the box given by the values

45

00:02:22,400 --> 00:02:24,567

of Q3 and Q1

46

00:02:24,633 --> 00:02:28,000

and the median giving

them line inside the box.

47

00:02:28,400 --> 00:02:31,067

Now there's of course, nothing really

special about the quartiles.

48

00:02:31,133 --> 00:02:34,367

We can actually use arbitrary percentiles,

49

00:02:34,600 --> 00:02:35,933

which we're gonna use over years.

50

00:02:36,033 --> 00:02:39,367

Student view

Overview
(/study/ap
122-
cid-
754029/

So for example,

if I'm interested in a 10th percentile,

51

00:02:39,433 --> 00:02:45,233

then I could go 10% of 70, which is 7,

and then we can see this around 34.

52

00:02:45,300 --> 00:02:47,233

So 34 demarcates at 10th percentile.

53

00:02:47,300 --> 00:02:50,033

Here is our 50th percentile,

also called the median.

54

00:02:50,400 --> 00:02:52,167

If I'm interested in 90th percentile,

55

00:02:52,233 --> 00:02:54,933

I go to vertically to about 63,

56

00:02:55,233 --> 00:03:00,367

and you can see that

the 90th percentile lies around 67.

57

00:03:00,633 --> 00:03:05,033

So if I'm above 67,

I'm in the top 90th percentile.

58

00:03:05,300 --> 00:03:08,333

So quickly, here is our histogram,

59

00:03:08,400 --> 00:03:10,700

which allows us

to build a cumulative histogram,

60

00:03:10,800 --> 00:03:13,067

which then given the endpoints,

61

00:03:13,333 --> 00:03:17,433

allows us to create the cumulative

frequency polygon,

62

Student
view

Overview
(/study/ap
122-
cid-
754029/

00:03:17,500 --> 00:03:19,533

or if you use a smooth light in the graph,

63

00:03:19,600 --> 00:03:22,600

this easily allows us to find Q1 ,

64

00:03:22,667 --> 00:03:25,833

the median, and Q3, giving us

the box and whisker plot.

65

00:03:25,900 --> 00:03:29,967

But we can also use it

for arbitrary percentiles given our data.

66

00:03:30,033 --> 00:03:30,967

And that is that.

In summary, we create a **cumulative frequency table** (below) based on the frequency table (above). Notice how the cumulative frequency is calculated. It is the sum of all the frequencies up to and including a particular interval.

Cumulative frequency table for the ages of the 70 members of the club shown in Table 1. Notice how the cumulative frequency is calculated. It is the sum of all the frequencies up to and including a particular interval.

X	Frequency	Cumulative frequency
$20—30$	3	3
$30—40$	10	$3 + 10 = 13$
$40—50$	18	$13 + 18 = 31$
$50—60$	24	$31 + 24 = 55$
$60—70$	12	$55 + 12 = 67$
$70—80$	3	$67 + 3 = 70$

By taking as data points the **endpoints** of each interval and connecting these points, we create a **cumulative frequency graph**. There are two types of cumulative frequency graph:

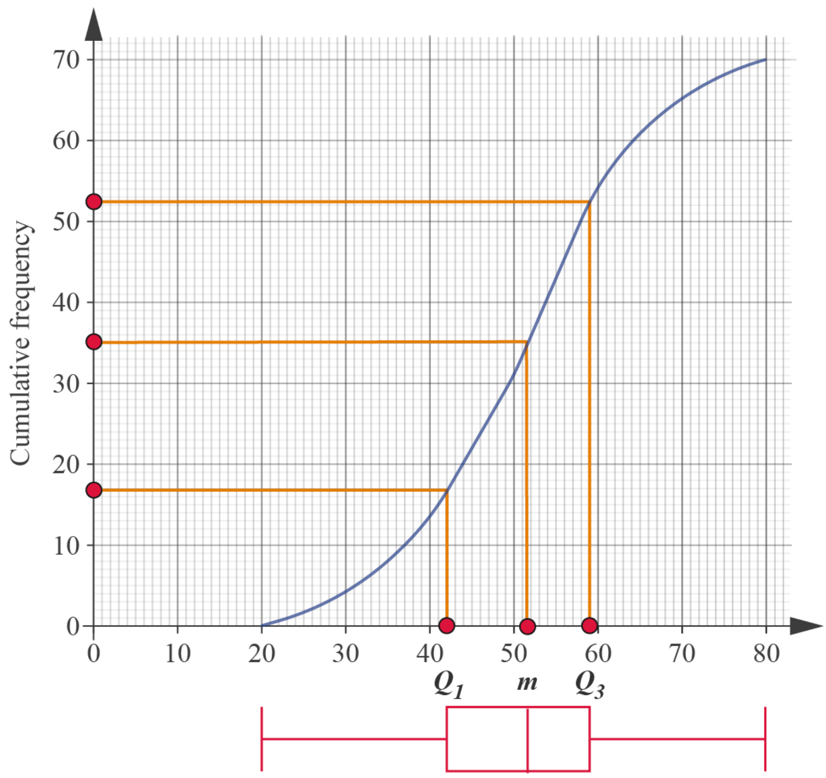
Student view

Overview
(/study/ap
122-
cid-
754029/

1. *Cumulative frequency* ***polygon*** : The data points are connected by straight lines, implying a linear distribution of the data points within an interval.
2. *Cumulative frequency* ***curve*** : All the data points are connected by a smooth curve.

Important

The vertical axis is always the cumulative frequency, while the horizontal axis will always be the independent variable being counted or measured.



More information

The image presents a cumulative frequency graph accompanied by a box-and-whisker plot directly underneath. The graph features horizontal and vertical grids. The x-axis is labeled with a range from 0 to 80, while the y-axis displays increments of 10 from 0 to 70, labeled as 'Cumulative frequency'. The graph includes marked points with lines connecting to horizontal lines on the graph, indicating cumulative frequencies for specific data points.

The cumulative frequency curve is shown in a smooth, increasing shape, starting near the lower left corner and rising to the upper right. Key data points are marked at 20, 40, 50, and 70 on the y-axis, corresponding to specific cumulative frequencies, and located at about 23, 40, 50, and 60 on the x-axis for points Q1, median (m), and Q3 respectively.

The box-and-whisker plot beneath the graph shows a horizontal line from approximately the smallest value on the x-axis to the largest, with a box indicating quartile 1 (Q1) to quartile 3 (Q3), and a central line at the median.

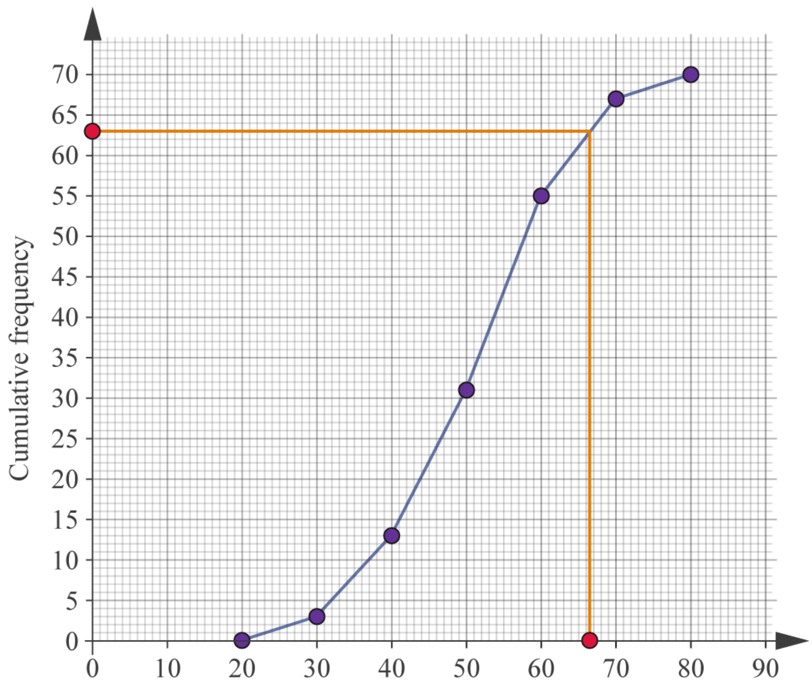
Student
view

Overview
(/study/ap
122-
cid-
754029/

[Generated by AI]

In the figure above, the median, m , and quartiles, Q_1 and Q_3, were found using the cumulative frequency graph. Together with the smallest and highest data points, this gives the five-number summary, which can be represented by the box-and-whisker plot. These are shown directly under the figure.

Naturally, apart from finding any quartile, one can use the cumulative frequency graph to find the value of any percentile of the data, as explored in the video above. Thus, the 90 th percentile of the data shown in the tables above, is the value of the 63rd data point, since $63 = 70 - 7$ where 7 is 1 0 % of 70 . As shown in the figure below, this is around 66.5. Hence, any member of the club who is over 66.5 years old is in the top 10 th percentile.



More information

This image is a cumulative frequency graph. The X-axis represents the data values, ranging from 0 to 90. The Y-axis represents cumulative frequency, ranging from 0 to 70. The graph includes a smooth curve connecting several data points marked with red and purple dots. An orange line intersects the curve at a data point corresponding to a cumulative frequency of 63, indicating the 90th percentile at approximately a data value of 66.5. The graph effectively illustrates how to determine percentiles using cumulative frequency.

Student
view

[Generated by AI]

Overview (/study/app/m/sid-122-cid-754029/

Below is an applet that replicates the process described in the video above for a frequency distribution of your own. You first give the leftmost endpoint of the classes, then the width of each class and finally the frequencies (which also implies the number of classes). Note that the frequencies must be given in the format described in the input box (separated with commas and in curly brackets). The frequency histogram is displayed by default and by ticking the relative boxes, you can also reveal the cumulative histogram, the cumulative frequency and a point on the x -axis, which you can move along and get the respective percentiles (including, of course, Q_1, the median and Q_3 for 25 % , 50 % and 75 % , respectively).

Interactive 1. Cumulative Frequency Curves.

More information for interactive 1

This interactive allows users to create and analyze frequency distributions by inputting their own data and customizing class intervals. Users can specify the leftmost endpoint and the width of each interval, then enter the frequency values for each class in a predefined format. Based on this input, the interactive automatically generates a frequency histogram, offering a visual representation of how data is distributed across the defined class ranges.

Student view

Overview
(/study/ap
122-
cid-
754029/

Users can further enhance their analysis by enabling additional options such as cumulative histogram, cumulative frequency, and percentile views.

For instance, using the frequency set $\{8, 5, 3, 6, 1, 9, 7\}$ with a starting class value of 10 and interval width of 15, the histogram displays the distribution across intervals like $(10 - 25, 25 - 40, \ldots)$. The x-axis (ranging from -20 to 140) and y-axis (from 0 to 40) provide clear scale references, and users can interactively move a point along the x-axis to identify key statistical measures and specific percentiles.

This dynamic interaction offers deeper insight into data spread, central tendency, and variability, helping users understand how class width and starting points influence the overall shape of the distribution.

This hands-on experience makes abstract statistical concepts more concrete, supporting comprehension of topics such as cumulative data, percentile ranks, and outlier detection. A reset function is available to allow users to quickly clear and input new datasets for further exploration.

Important

The cumulative frequency graph may rise with a high slope or low slope and may even appear flat at certain points, but it will never decrease. Why is this?

Exam tip

In examinations, usually the values of any percentile to be obtained via a cumulative frequency graph fall within a range, as even when using a ruler to obtain the value of a percentile, the precision is limited.

Theory of Knowledge

As you've read in this subtopic, there is a great deal of **power** in regard to choice of both data inclusion as well as data representation when working with statistics. For example as the statistician, you have the **power** to choose if you include outliers or not. You also have the power to decide on the range of your data 'bins'. All of this affects the knowledge created.

Many people believe statistics and mathematics to be **value** -free. However consider the ways in which **values** and **perspectives** are indeed integrated into mathematics, such as in the case of statistics.

3 section questions

Student
view

Overview
(/study/ap
122-
cid-
754029/

Checklist

Section Student... (0/0) Feedback Print (/study/app/m/sid-122-cid-754029/book/checklist-id-26231/print/) Assign

What you should know

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

- construct a frequency distribution diagram from a frequency distribution table
- identify the following from data grouped into intervals:
  - mid-interval values
  - interval width (though it is not common to have a varying interval width)
  - lower interval boundary
  - upper interval boundary
  - modal class (the class with the highest frequency or the tallest class in the diagram)
- find the five-number summary, range and IQR from a set of data
- construct and interpret a box-and-whisker plot, including identifying any outliers
- interpret a box-and-whisker plot to determine if data is approximately normally distributed, positively skewed or negatively skewed
- construct a cumulative frequency curve from a frequency table
- interpret a cumulative frequency curve with percentiles.

4. Probability and statistics / 4.2 Presentation of data

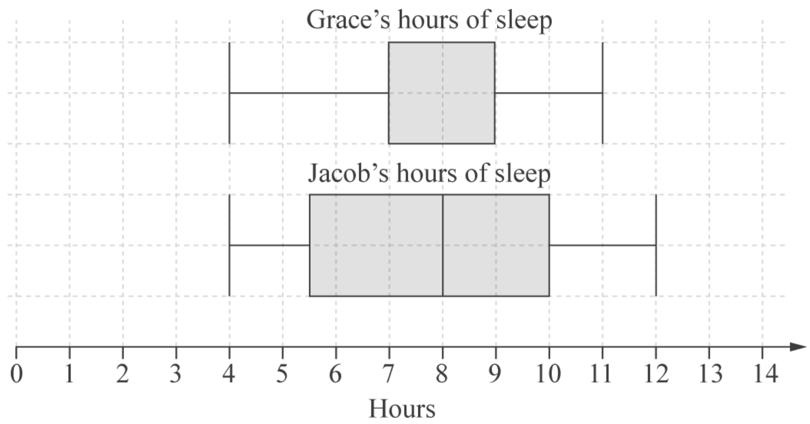
Investigation

Section Student... (0/0) Feedback Print (/study/app/m/sid-122-cid-754029/book/investigation-id-26232/print/) Assign

We have explored how to construct and interpret box-and-whisker plots, but what if you have data from different groups of people that you want to compare? You can make certain comparisons by using box-and-whisker plots that are plotted next to each other. For example, the figure below shows a comparison of the number of hours of sleep for two students, Jacob and Grace.

Student
view

Overview
(/study/ap
122-
cid-
754029/



Comparative box-and-whisker plot for Jacob and Grace.

More information

The image is a box-and-whisker plot comparing the hours of sleep for Jacob and Grace. The X-axis represents hours, ranging from 0 to 14. Grace's box plot shows a range between 8 and 14 hours, indicating her typical sleep range. Jacob's box plot displays a range from 6 to 10 hours, reflecting his sleep pattern. Both plots show the interquartile range, median, and any potential outliers. The graph provides a visual representation of the difference and spread in their sleeping hours.

[Generated by AI]

By comparing the two students in this way, you can see that while Jacob occasionally gets much more sleep than Grace, the number of hours he gets each night varies much more. Grace gets between 7 and 9 hours of sleep 50% of the time, so she seems to have a more consistent routine.

You can make comparisons like this one as well: think of some measurable characteristic that you would like to explore and two groups of students who might have a significant difference in that characteristic. Survey students in your school until you have at least $10–15$ values for each group, and construct box-and-whisker plots for each set of data. Do you see anything interesting? What comparisons can you make? Were the results what you expected? Finally, trade your plots with one of your classmates and see if you each come up with the same observations and comparisons for each other's data.

Rate subtopic 4.2 Presentation of data

Help us improve the content and user experience.

Student
view