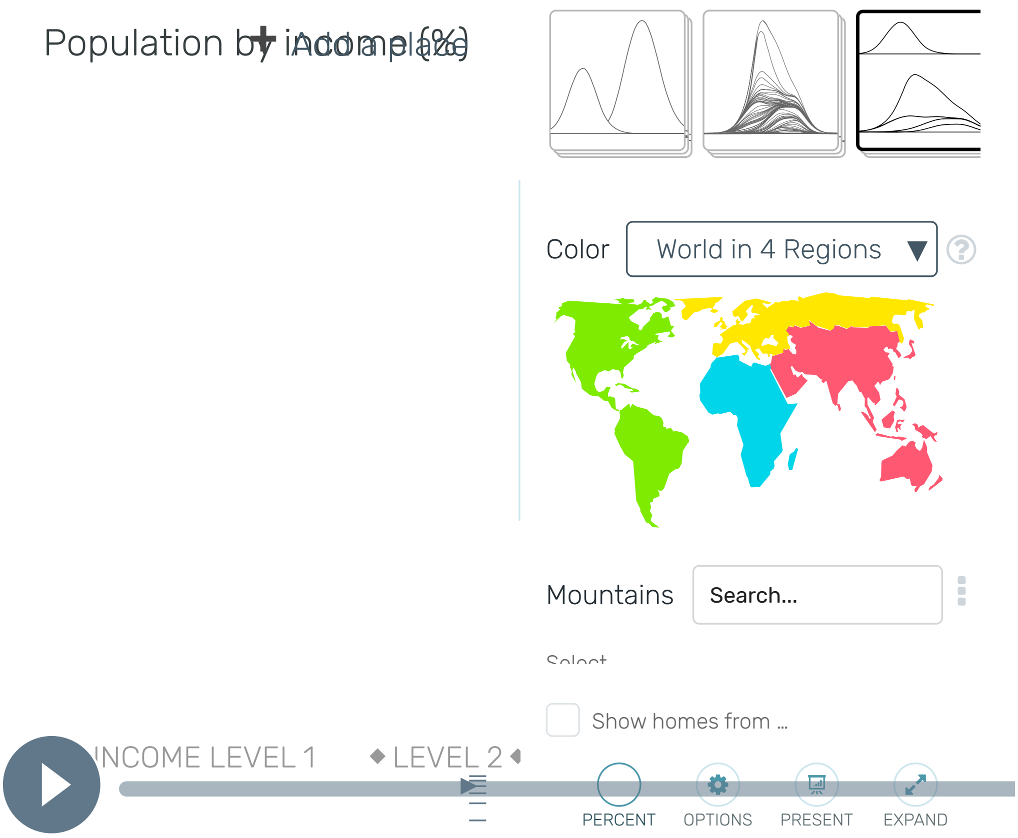
TOPIC 2
FUNCTIONS

SUBTOPIC 2.10
SCALING USING LOGARITHMS AND LINEARISING DATA

Index

The big picture

Graphs do not necessarily show a single curve between two set of axes. There are many complicated data sets that require different graphing techniques. Some data sets, such as population or gross domestic product (GDP), have a wide range of values and it can be impractical to plot them on a single graph. If you did, you may not be able to pick out the underlying pattern. What mathematical tool has been used to fit the data to a scale for the interactive graph below?



Interactive 1. What Mathematical Tool Has Been Used to Fit the Data to a Scale for the Interactive Graph?

Overview
(/study/app/math-ai-hl/sid-132-cid-761618/ov

More information for interactive 1

This interactive visualization explores global income distribution patterns across different regions and time periods, using scaled axes to effectively represent wide-ranging economic data. The tool employs logarithmic scaling to compress the vast income spectrum into a manageable visual format while preserving meaningful comparisons between populations at all economic levels.

The display presents a color-coded graph with income levels (in dollars per day) on a logarithmic x-axis and population percentages on the y-axis. There are four regional divisions: Africa in blue, Asia in red, Europe in yellow, and the Americas in green, and a vertical dashed line around 2 dollars per day labeled as "extreme poverty". Below the graph, interactive controls allow the selection of specific countries and adjustment of the period from 1800 to 2023 through a play button animation on the bottom left corner.

When users select "Asia" and animate the timeline, they observe the region's income distribution shifting rightward, showing economic growth. Choosing "Africa" reveals a different pattern, with more population concentration at lower income levels. The logarithmic scale ensures visibility of both very low incomes (under $\$1/\text{day}$) and high incomes ($\$500/\text{day}$) on the same graph, demonstrating how mathematical scaling enables comprehensive data representation.

The tool teaches how logarithmic transformations make extreme-value data visually interpretable, revealing patterns obscured on linear scales. Users learn to analyze global inequality trends and recognize how scaling techniques facilitate the comparison of diverse economies. This approach is fundamental for economists, policymakers, and researchers working with datasets spanning multiple orders of magnitude, from poverty analysis to macroeconomic studies.

Watch this video (http://www.gapminder.org/videos/crisis-narrows-china-uk-gap/) to understand how such data sets can be scaled. The video discusses two graphs. For each graph, note the scale on the axis that represents per capita income and how this scale helps in giving a better picture of the data .

Concept

Representing a data set with a wide range on a graph is always challenging. If the **rate of change** is high compared to the **absolute values**, then representing the whole data set on a single graph can be often achieved by changing the scale of one or both of the axes using logarithms. This method can also be used to **linearise** a curve .

Student view

Overview
(/study/app/math-ai-hl/sid-132-cid-761618/ov

Theory of Knowledge

Solving equations can sometimes seem formulaic (pun intended!) and void of creativity. Does mathematics lack creativity?

More importantly, a key knowledge question is: ‘Is the degree of creative freedom within an area of knowledge positively correlated with the expansion of knowledge within that AOK?’

2. Functions / 2.10 Scaling using logarithms and linearising data

Logarithmic scaling, linearisation, and semi-log

Logarithmic scaling

Different objects in the universe have different sizes. These values could range from as small as 10^{-15} m (size of a proton) to 10^{21} m (size of a galaxy). The sizes of some objects and their logarithmic values are given in the following table:

Object	Approximate size of object, a (m)	ln (a)
Proton	10^{-15}	-34.5388
Atom	10^{-10}	-23.0259
Blood cell	10^{-6}	-13.8155
Diameter of human hair	10^{-4}	-9.21034
Human	2	0.693147
Giraffe	5	1.609438
Blue whale	30	3.401197

Student view

Overview
(/study/app/math-ai-hl/sid-132-cid-761618/ov

Object	Approximate size of object, a (m)	ln (a)
Moon	1.731×10^6	14.36421
Earth	6.371×10^6	15.66727
Sun	6.957×10^8	20.36043
Solar system	7.48×10^{12}	29.64325
Milky Way	1.892×10^{21}	48.99192

Making connections

Logarithms, as you saw in subtopic 2.9 (/study/app/math-ai-hl/sid-132-cid-761618/book/the-big-picture-id-27495/), are used to simplify calculations such as multiplication and division involving large numbers.

Observe the magnitudes of the numbers in the two end columns in the table above. What do you observe? Which set of numbers would you prefer to work with?

This of course depends on the task you are working on. Sometimes working with logarithms is more convenient. Taking the logarithm is called logarithmic scaling.

Activity

Logarithmic scaling is used in several contexts. Here are some examples.

- The decibel scale expresses sound intensity. When you adjust the volume on an ampliphier, you see a number changing. What does that number mean?
- The Richter scale expresses the magnitude of an earthquake. What is the difference between a 5.9 and a 9.5 earthquake?
- Scientists use the pH scale to express the acidity or alkalinity of solutions. What does the pH value 7 mean? One way of measuring the pH value is using the so-called universal indicator where colours correspond to different pH-values. What pH-value is indicated by a red, yellow, green, blue and purple colour?

Section Student... (0/0) Feedback Print (/study/app/math-ai-hl/sid-132-cid-761618/book/the-big-picture-id-27500/print/) Assign

Student view

Overview (/study/app/math-ai-hl/sid-132-cid-761618/ov

As another example, consider the distances of the planets from the sun.

Planet	Distance from the sun (km)
Mercury	57 910 006
Venus	108 199 995
Earth	149 599 951
Mars	227 939 920
Jupiter	778 330 257
Saturn	1 429 400 028
Uranus	2 870 989 228
Neptune	4 504 299 579

The distances range from over 57 million km to over 4 billion km. Sometimes it can be helpful to use the logarithm of the numbers, which range from 15 to 25.

Planet	Distance to the sun, d (km)	ln(d)
Mercury	57 910 006	17.87
Venus	108 199 995	18.50
Earth	149 599 951	18.82
Mars	227 939 920	19.24
Jupiter	778 330 257	20.47
Saturn	1 429 400 028	21.08

Student view

Overview
(/study/app/math-ai-hl/sid-132-cid-761618/ov

Planet	Distance to the sun, d (km)	ln(d)
Uranus	$2\,870\,989\,228$	21.78
Neptune	$4\,504\,299\,579$	22.23

A logarithmic scale can also use other bases, such as 10.

The following activity helps you to understand how base 10 can be used in the above example.

Activity

The table below shows the values when you take the base 10 logarithm of the distances.

Planet	Distance to the sun, d (km)	$\log_{10}$
Mercury	$57\,910\,006$	7.7627[illegible]
Venus	$108\,199\,995$	8.0342[illegible]
Earth	$149\,599\,951$	8.1749[illegible]
Mars	$227\,939\,920$	8.3578[illegible]
Jupiter	$778\,330\,257$	8.8911[illegible]
Saturn	$1\,429\,400\,028$	9.1551[illegible]
Uranus	$2\,870\,989\,228$	9.4580[illegible]
Neptune	$4\,504\,299\,579$	9.6536[illegible]

- Is there an advantage using base 10 over the natural logarithm?
- In some areas of science base 2 logarithm is often used. Find the base 2 logarithm of the distances. What is the range of these values? Is there an advantage of this transformation over the other two logarithms mentioned previously?

Student view

Overview
(/study/app/math-ai-hl/sid-132-cid-761618/ov

Linearisation

Sometimes using logarithmic scaling on one or both axis of a coordinate system can help in finding models. Let's look at some examples.

Example 1

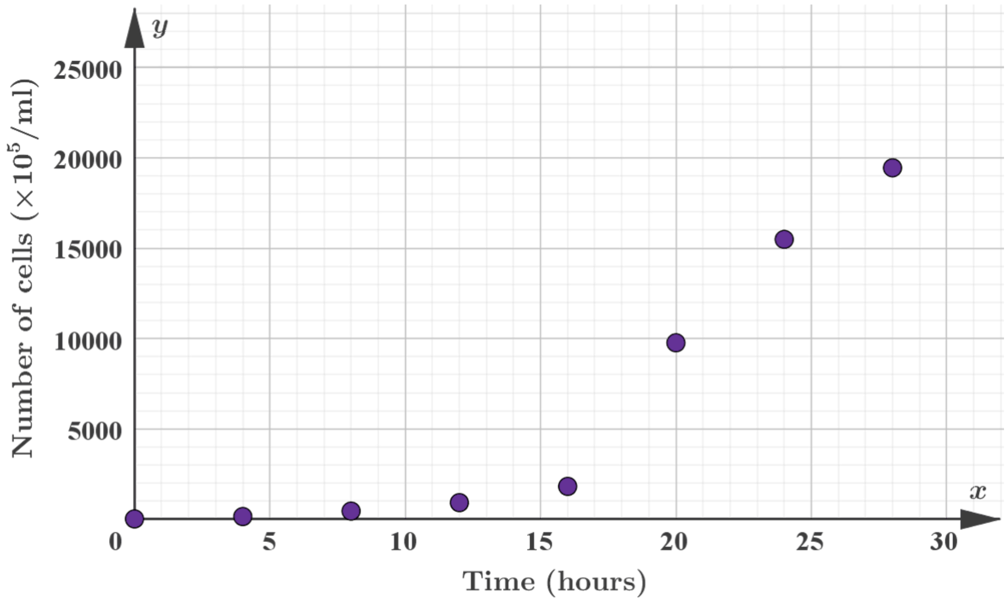
The table below gives the growth in the number of bacteria cells ($\times 10^5$ per ml) in an infected person's blood taken every 4 hours.

Time (hours)	**Number of bacteria cells ($\times 10^5$ per ml)**
0	20
4	150
8	453
12	920
16	1820
20	9765
24	15 487
28	19 450

Plotting the data gives the graph below:

Student view

Overview
(/study/app/math-ai-hl/sid-132-cid-761618/ov



More information

The graph depicts the growth of bacteria cells over time. The X-axis represents time in hours, ranging from 0 to 30 hours. The Y-axis shows the number of bacterial cells, multiplied by 10^6 per milliliter, with intervals of 5000, ranging from 0 to 25000. Multiple data points are plotted, showing a trend where the number of cells starts low and increases with time. Initially, the curve is less steep, indicating slower growth, but it becomes steeper, suggesting an acceleration in bacterial growth over time. This pattern reflects the explanation provided that fitting a good trendline may be challenging due to the wide range of cell numbers.

[Generated by AI]

Notice that the curve is less steep at the beginning and gets steeper. Fitting a good trendline can be difficult because of the wide range in the number of cells (y-axis). By taking logarithms for the y-axis, where N is the number of bacteria cells, we get:

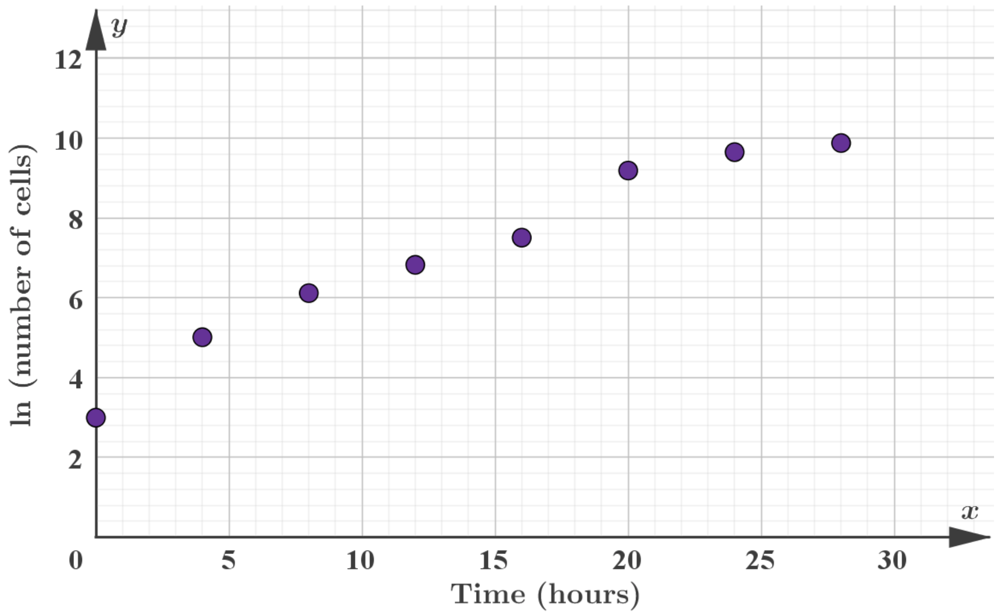
Time (hours)	$\ln N$
0	2.995732274
4	5.010635294
8	6.115892125
12	6.82437367
16	7.50659178

Student view

Overview
(/study/app/math-ai-hl/sid-132-cid-761618/ov

Time (hours)	$\ln N$
20	9.186559843
24	9.647756241
28	9.875602349

The graph of the above table is as shown below. Look at the y-values in this table and compare them with the original y-values. What do you observe?



More information

The image is a scatter plot graph displaying the natural logarithm of the number of cells on the Y-axis and time in hours on the X-axis. The Y-axis ranges from 0 to 12 and is labeled "In (number of cells)", while the X-axis ranges from 0 to 30 and is labeled "Time (hours)". Purple data points are plotted at intervals along the X-axis to indicate observed measurements. The data points show a trend of increasing cell numbers over time. Key data points include an increase from approximately 2 at hour 0 to around 11 at hour 28, indicating a positive trend. Each point represents data collected at specific time intervals: around hour 5, 10, 15, 20, and 25, demonstrating steady growth in the number of cells as time progresses.

[Generated by AI]

Which of the above two plots do you think would be easier to find a good regression model? Why?

Student view

Overview
(/study/app/math-ai-hl/sid-132-cid-761618/ov

It looks like that the second plot can be modelled with a linear function, which would fit most of the points to a reasonable approximation. Let's use the linear regression option of the calculator to find a linear model.

$$y = 0.24178249x + 3.76043807$$

Note that $y = \ln N$, so let's rewrite this equation to get a relationship between time and the number of cells.

$$\begin{aligned}
\ln N &= 0.24178249x + 3.76043807 \\
N &= \mathrm{e}^{0.24178249\mathrm{x}+3.76043807} \\
&= \mathrm{e}^{3.76043807}\mathrm{e}^{0.24178249\mathrm{x}} \\
&= 42.96724452\mathrm{e}^{0.24178249\mathrm{x}} \\
N &\approx 43.0\mathrm{e}^{0.242\mathrm{x}}
\end{aligned}$$

This model can also be written in a different form.

$$\begin{aligned}
N &= 42.96724452\mathrm{e}^{0.24178249\mathrm{x}} \\
N &= 42.96724452(\mathrm{e}^{0.24178249})^{\mathrm{x}} \\
&= 42.96724452 \times 1.27351716^{x} \\
N &\approx 43.0 \times 1.27^{x}
\end{aligned}$$

You can confirm this result by using the exponential regression option of your calculator.

⚠ Be aware

If you use a logarithmic scale for the y -axis, you need to be careful while modelling and interpreting the graph since the y -values are the logarithms of the original values and not the actual values.

This process of transforming an exponential trend into a linear trend is called linearisation .

In our example, the linear function is $\ln N \approx 0.242x + 3.76$, where only one variable (N) was converted to a logarithm. This type of model is called a semi-log model.

Student view

Overview
(/study/ap
ai-
hl/sid-
132-
cid-
761618/ov

Example 2

Let's revisit now the example from the previous section about the box office data of the Frozen II movie in the weeks after it was released in November 2019.

Week (after release)	Total gross (USD)
1	202 867 358
2	302 924 901
3	347 360 072
4	374 233 961
5	404 790 889
6	438 585 364
7	453 623 042
8	461 151 690
9	467 263 568
10	470 637 370

You saw already that the scatter plot shows a pattern that is increasing, but as opposed to the previous example, this time the increase is slowing down. In these cases taking the logarithm of the numbers in the left coloumn might give a linear pattern. Let's try this.

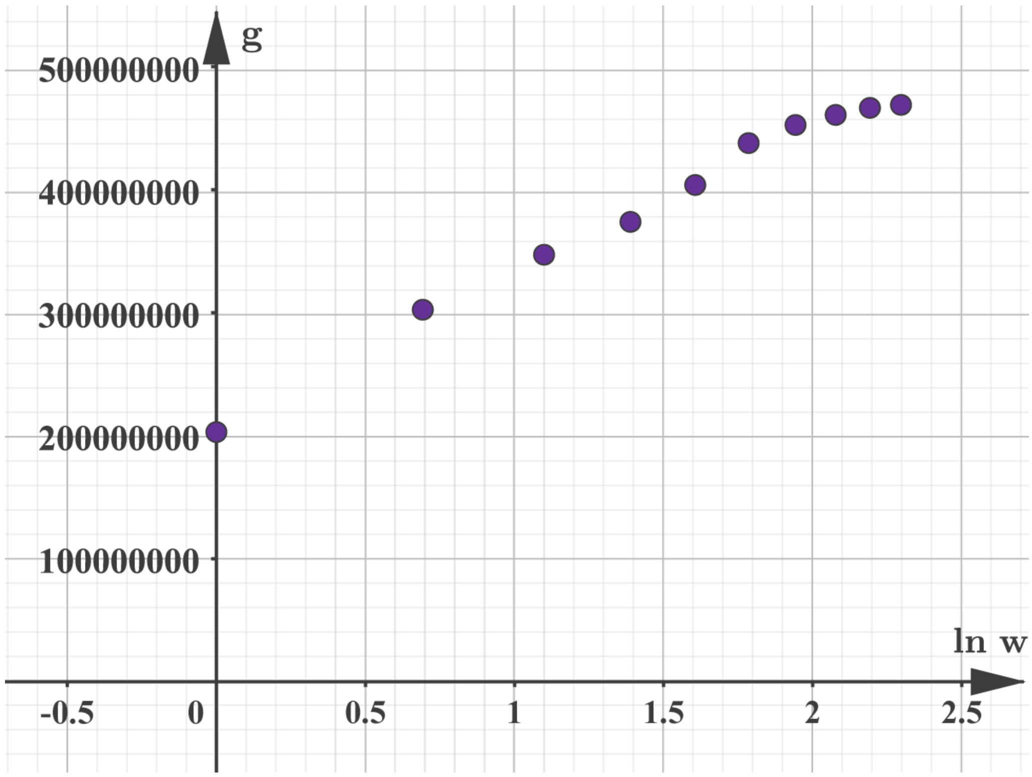
$\ln w$	g
$\ln 1 = 0$	202 867 358
$\ln 2 \approx 0.693147$	302 924 901
$\ln 3 \approx 1.098612$	347 360 072
$\ln 4 \approx 1.386294$	374 233 961

Student
view

Overview
(/study/ap
ai-
hl/sid-
132-
cid-
761618/ov

$\ln w$	g
$\ln 5 \approx 1.609438$	404 790 889
$\ln 6 \approx 791759$	438 585 364
$\ln 7 \approx 1.945910$	453 623 042
$\ln 8 \approx 2.079442$	461 151 690
$\ln 9 \approx 2.197225$	467 263 568
$\ln 10 \approx 2.302585$	470 637 370

Let's see the corresponding scatter plot.



More information

The image is a scatter plot graph. The X-axis is labeled "ln w" and ranges from -0.5 to 2.5. The Y-axis is labeled "g" with values ranging from 0 to 50,000,000 in intervals of 10,000,000. A series of data points is plotted, showing an upward trend with a slight curve downward towards the end. The data points are marked with purple dots and the pattern suggests a linear relationship with slight curvature.

[Generated by AI]

Student
view

Overview (/study/app/math-ai-hl/sid-132-cid-761618/ov

The pattern indeed looks linear, although a bit curving down towards the end. You can use the calculator to find the linear regression line.

$$y = 118\,612\,402x + 213\,186\,757$$

When you substitute $x = \ln w$ and $y = g$, you get the model of the previous section for the box office performance of the movie in terms of the number of weeks passed since its release.

$$g = 118\,612\,402 \ln(w) + 213\,186\,757$$

Log-log models

Before moving on to the next type of linearisation, let's summarise the methods seen in Example 1 and Example 2.

Important

- A linear relationship between x and $\ln y$ models an exponential relationship between x and y.
  $$\text{If } \ln y = mx + c, \text{ then } y = \mathrm{e^c e^{mx}} = \mathrm{e^c (e^m)^x} = \mathrm{a \times b^x}$$
- A linear relationship between $\ln x$ and y models a logarithmic relationship between x and y.
  $$y = m \ln(x) + c$$

Let's see what type of relationship we get when the correspondance between $\ln x$ and $\ln y$ is linear.

$$\begin{aligned}
\ln y &= m \ln x + c \\
y &= \mathrm{e^{m \ln x + c}} \\
y &= \mathrm{e^{m \ln x} e^c} \\
y &= \mathrm{e^c \left(e^{\ln x}\right)^m} \\
y &= \mathrm{e^c x^m} \\
y &= ax^m
\end{aligned}$$

In this case, both x and y have been converted to logarithms. This type of model is called a log-log model.

Student view

Overview
(/study/app/math-ai-hl/sid-132-cid-761618/ov

Important

- A linear relationship between $\ln x$ and $\ln y$ models a power law relationship between x and y.
  $y = ax^m$

Example 3

★☆☆

Which kind of function would we use the following models for?

1. A semi-log model would be used for ________________.
2. A log-log model would be used for ____________.

1. An exponential or logarithmic function.
2. A power law.

Activity

There are several relationships that can be expressed using a power law. Here are the names of a few people whose name is associated to such laws. Do some research and find these laws.

- Anders Knutsson Ångström (1888-1981)
- Johannes Kepler (1571-1630)
- Max Kleiber (1893-1976)
- Alfred James Lotka (1880-1949)
- Stanley Smith Stevens (1906-1973)
- Lionel Roy Taylor (1924-2007)

Example 4

★★★

Student view

Overview (/study/app/math-ai-hl/sid-132-cid-761618/ov

Look at the following table that gives the distance of the planets in the solar system from the Sun and their orbital periods.

source: NASA Planetary fact sheet (https://nssdc.gsfc.nasa.gov/planetary/factsheet/)

Planet	Distance from Sun (10^6 km)	Orbital period (days)
Mercury	57.9	88.0
Venus	108.2	224.7
Earth	149.6	365.2
Mars	227.9	687.0
Jupiter	778.6	4331
Saturn	1433.5	10 747
Uranus	2872.5	30 589
Neptune	4495.1	59 800
Pluto	5906.4	90 560

It is known that the relationship between the distance from the Sun ($d \times 10^6$) and the orbital period (p) is related according to the formula $p = ad^r$.

(a) Use the data for Earth and Uranus to estimate the value of a and r.

(b) Create a table using the natural logarithm of all numbers in the table.

(c) Draw a scatter plot for your table of values.

(d) Find the line through the first and last point on your plot and use this line to estimate the value of a and r.

(e) Use your calculator to find the regression line for your scatter plot and use this line to estimate the value of a and r.

Student view

Overview
(/study/app/math-ai-hl/sid-132-cid-761618/ov

(f) Use the power regression option of your calculator to find a model for the original data.

(g) Draw a scatter plot of the original data and include this last model on the plot.

(h) Estimate the orbital period of Ceres, a dwarf planet which is 413.6×10^6 kilometres from the Sun.

(a) The data for Earth gives $365.2 = a \times 149.6^r$

The data for Uranus gives $30\,589 = a \times 2872.5^r$

Dividing the two equations eliminates a, so we can solve the resulting equation for r.

$$\begin{aligned}\frac{a \times 2872.5^r}{a \times 149.6^r} &= \frac{30\,589}{365.2}\\ 19.2012^r &= 83.7596\\ r &= \log_{19.2012} 83.7596 \approx 1.50\end{aligned}$$

You can substitute this value in either of the equation to get the value of a.

$$\begin{aligned}365.2 &= a \times 149.6^{1.5}\\ a &= \frac{365.2}{149.6^{1.5}} \approx 0.200\end{aligned}$$

According to these calculations, an approximate model for the relationship is $p = 0.2d^{1.5}$.

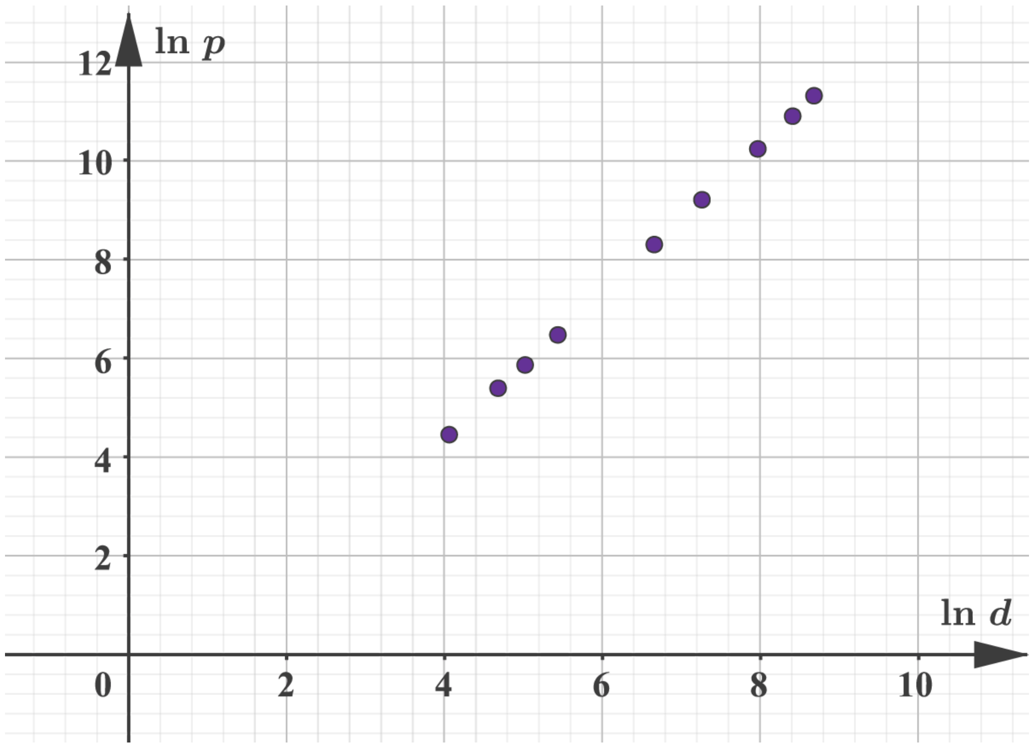
(b) To find a different model, let's create the table of logarithms.

Planet	$\ln d$	$\ln p$
Mercury	4.0587	4.4773
Venus	4.6840	5.4148
Earth	5.0080	5.9004
Mars	5.4289	6.5223

Student view

Planet	$\ln d$	$\ln p$
Jupiter	6.6575	8.3736
Saturn	7.2679	9.2824
Uranus	7.9629	10.3284
Neptune	8.4107	10.9988
Pluto	8.6838	11.4138

(c) Let's draw the scatter plot as asked.



(d) This is clearly a linear relationship. The next task is to find the equation of the line through $(4.0587, 4.4773)$ and $(8.6838, 11.4138)$,the first and last point. You can use the equation from the formula booklet.

$$y - y_1 = \frac{y_2 - y_1}{x_2 - x_1}(x - x_1)$$
$$y - 4.4773 = \frac{11.4138 - 4.4773}{8.6838 - 4.0587}(x - 4.0587)$$
$$y - 4.4773 = 1.49975(x - 4.0578)$$
$$y = 1.49975x - 1.608391$$

Overview (/study/app/math-ai-hl/sid-132-cid-761618/ov

Let's substitute $y = \ln p$ and $x = \ln d$ to find the relationship between p and d.

$$\begin{aligned}\ln p &= 1.49975 \ln d - 1.608391\\ p &= \mathrm{e}^{1.49975 \ln \mathrm{d} - 1.608391}\\ p &= \left(\mathrm{e}^{\ln \mathrm{d}}\right)^{1.49975} \mathrm{e}^{-1.608391}\\ p &= 0.200209 d^{1.49975}\\ p &\approx 0.2 d^{1.5}\end{aligned}$$

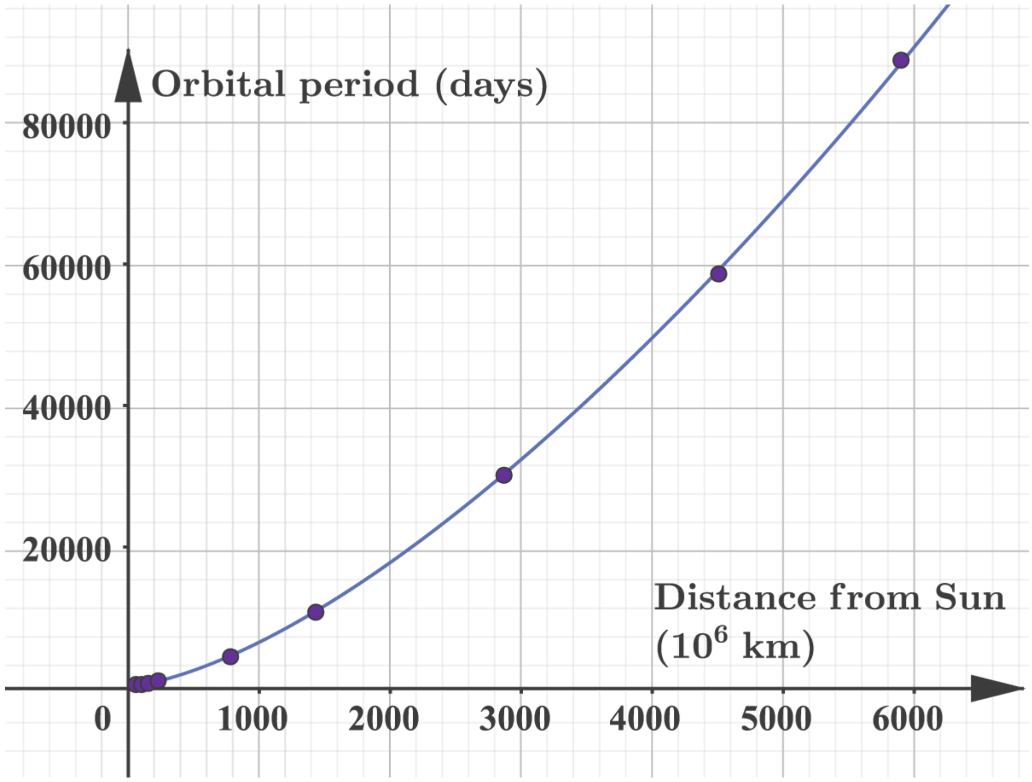
This method also gave the approximate values $a = 0.2$ and $r = 1.5$.

(e) The linear regression option of the calculator gives $y = 1.4987995x - 1.605784$, and if you convert this to the relationship between p and d as before, you get the same approximate model.

(f) The power regression option gives $p = 0.200732077 d^{1.498799523}$. With rounded coefficients you get the same model as before.

$$p \approx 0.2 d^{1.5}$$

(g) The diagram below shows the scatter plot with the model.



Student view

Overview
(/study/app/math-ai-hl/sid-132-cid-761618/ov

(h) We can substitute the distance from the Sun in the model to estimate the orbital period of Ceres.

$$p = 0.2 \times 413.6^{1.5} \approx 1682 \text{ days}$$

International Mindedness

One of the major breakthroughs in mathematics that has applications in many areas of knowledge is big data analysis. The development of regression techniques over the last 100 years has coincided with ever-increasing quantities of real-world data that need processing, since statistical analysis is relatively young compared with algebra, geometry, etc. Whilst these techniques solve many real-life problems, there are still many challenges and requirements for the further development of mathematical representations such as bubble graphs, multiple-axis graphs, etc.

How have mathematical concepts evolved in recent years with the developments in data analysis?

3 section questions

2. Functions / 2.10 Scaling using logarithms and linearising data

Checklist

Section Student... (0/0) Feedback Print (/study/app/math-ai-hl/sid-132-cid-761618/book/checklist-id-27502/print/) Assign

What you should know

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

- use a logarithmic scale to analyse a data set
- linearise a data set using logarithms
- interpret log-log and semi-log graphs.

Student view

Investigation

Monthly traffic (total number of visits in that month) to a website is shown in the following data:

Table 1. Total number of website visits per month.

Month	Number of Website
January	14
February	448
March	3402
April	14 336
May	43 750
June	108 864
July	235 298
August	458 752
September	826 686
October	1 400 000
November	2 254 714
December	3 483 648

Model the above data using a power law of the form $y = ax^b$ and an exponential function of the form $y = ae^{bx}$.

Overview (/study/app/math-ai-hl/sid-132-cid-761618/ov

Which of the models are a good fit to the data? Will using a logarithmic scale and linearising both models give you a better fit? If so, which model would you choose? Justify your answer.

Rate subtopic 2.10 Scaling using logarithms and linearising data

Help us improve the content and user experience.

Student view