IB SL Maths Transition Data Task

In this task you will use real data to investigate if there is an association between a country's Gross Domestic Product (GDP) per capita and the percentage of their population that has access to the internet.

This will prepare you for completing your Maths Internal Assessment (IA) in a few months' time, where you will look for an association between two variables of your choosing.

To complete this task, you will also need the Excel spreadsheet titled 'IB SL Maths Advisory Task - Data', which contains the data for this investigation, and access to the Geogebra website to analyse the data (https://www.geogebra.org/classic#spreadsheet).

Research Question:

Is there an association between a county's GDP per capita and the percentage of its population that has access to the internet?

A: Make a prediction

Here is some information about the two variables being studied to help you make a prediction of the outcome of your investigation:

- A country's Gross Domestic Product (GDP) is a measure of the total income generated by the production of goods and services in a country or region each year. GDP per capita is the GDP divided by the number of people in that country's population; it is a measure of the average economic output per person. The GDP per capita in this task is measured in US$ and is from 2021.[1]
- The percentage of a country's population with access to the internet was measured by surveying approximately 1000 people aged 15 years and over in each country. The data is from 2021.[2]

Use the information provided above as well as your general knowledge to make a prediction about whether there will be an association between a country's GDP per capita and the percentage of its population that has access to the internet:

I predict that there will / will not (delete as appropriate) be an association between a country's GDP per capita and the percentage of its population that has access to the internet because .....

.....

.....

.....

.....

.....

If you predicted that there will be an association, then also predict the nature of the association:

I predict that the association will be positive / negative (delete as appropriate), i.e. as the GDP per capita increases, the percentage of the population with access to the internet will increase / decrease (delete as appropriate) because .....

.....

.....

.....

.....

B: Take a sample of the data

The 'IB SL Maths Advisory Task - Data' spreadsheet contains information about 120 countries. To make it easier to work with the data, take a random sample of 30 of the countries by using the random number generator on the spreadsheet (follow the instructions on the spreadsheet for help with this).

Include a copy of your sampled data here, showing the Country Name, GDP per capita and Percentage of the population with internet access:

Looking at the data, do you notice anything interesting/surprising/relevant? Give details.

.....

.....

.....

.....

.....

C: Analyse each variable in turn (univariate analysis)

It is easier to analyse data using the Geogebra website than Excel, so copy the data from columns J, K and L (Country Names, GDP per capita and Percentage of the population with internet access) into a spreadsheet in Geogebra (https://www.geogebra.org/classic#spreadsheet).

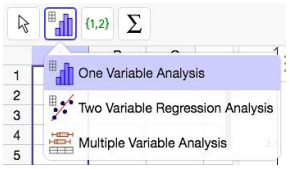
To help you get to know the data better, you first need to look at each variable separately.

Complete each of the following steps for the GDP per capita in your random sample of 30 countries and then again for the percentage of the population with internet access.

1) Use the data in Geogebra to draw a box plot.
2) Use the box plot to identify if there are any outliers and note which countries these relate to.
3) Use the data in Geogebra to identify the median and interquartile range.
4) Describe how the spread compares to the average.

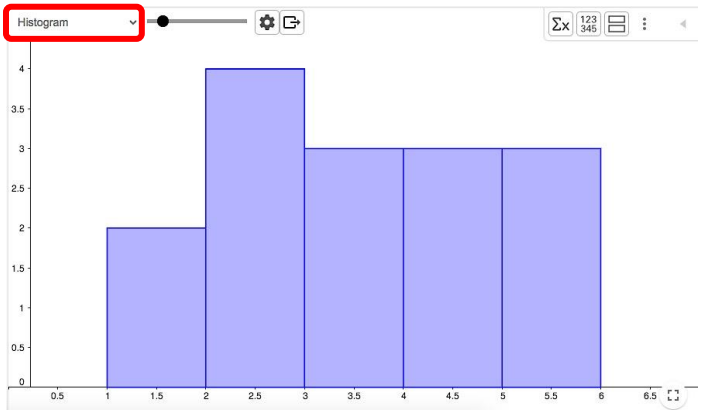
How to use Geogebra to do the above:

1) Use the data in Geogebra to draw a box plot.
    1) Highlight all of the data in the required variable by clicking on the column heading letter.
    2) Click on the blue bar chat in the toolbar then click on 'One Variable Analysis'.



    You will now see a histogram of your data on the right-hand side (although this is not useful to us).

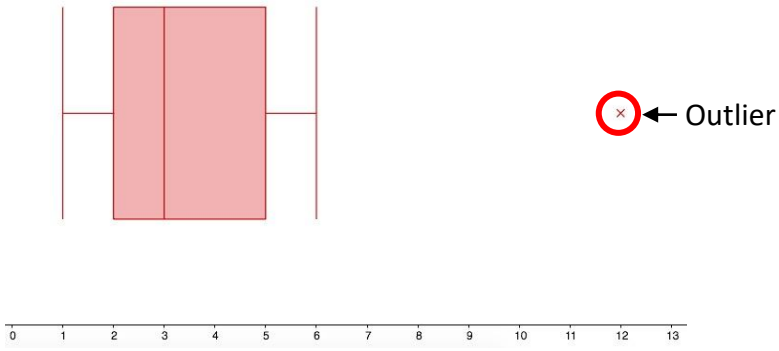
    3) To change the histogram into a box plot, click on the word 'Histogram' and then click on 'Boxplot'.



2) Use the box plot to identify if there are any outliers and note which countries these relate to.

If your data contains outliers they will be shown as crosses on the box plot,

e.g.

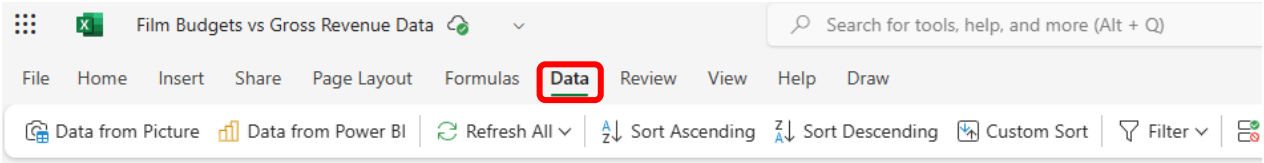


To identify which data values are the outliers,

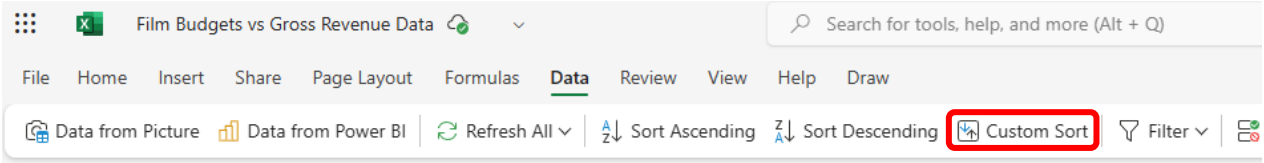
1) Go back to the Excel spreadsheet of your sampled data

2) Highlight the sampled data (including the headings) from columns **I, J, K and L**

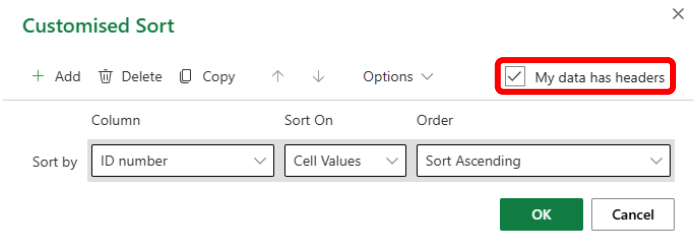
3) Click on the *Data* tab.



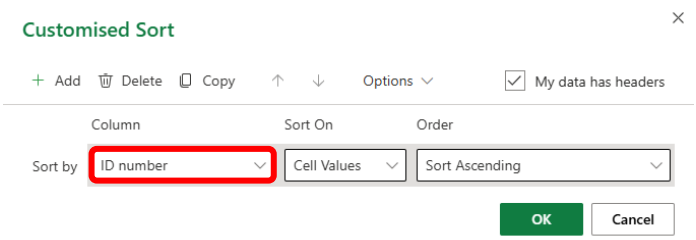
4) Then click on *Custom Sort*.



5) In the box that appears ensure the 'My data has headers' box is ticked.



6) Select the required variable name in the 'Column' section, then click *OK*.



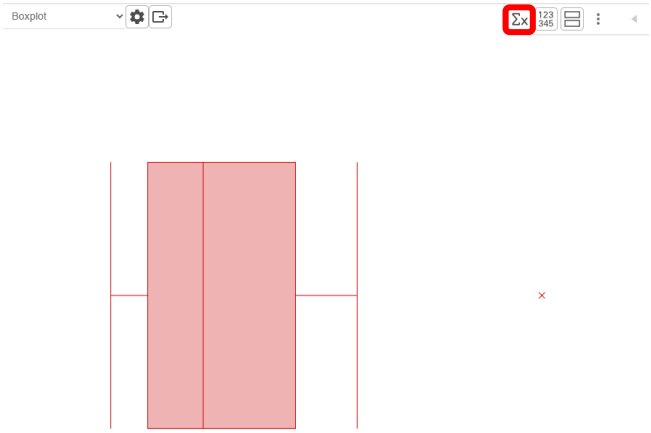
7) The data will now be sorted in ascending order for the variable you selected, i.e. the lowest values are at the top of the list, and the highest values are at the bottom of the list. Use the list to help you identify which countries were the outliers identified on the box plot.

Please note:

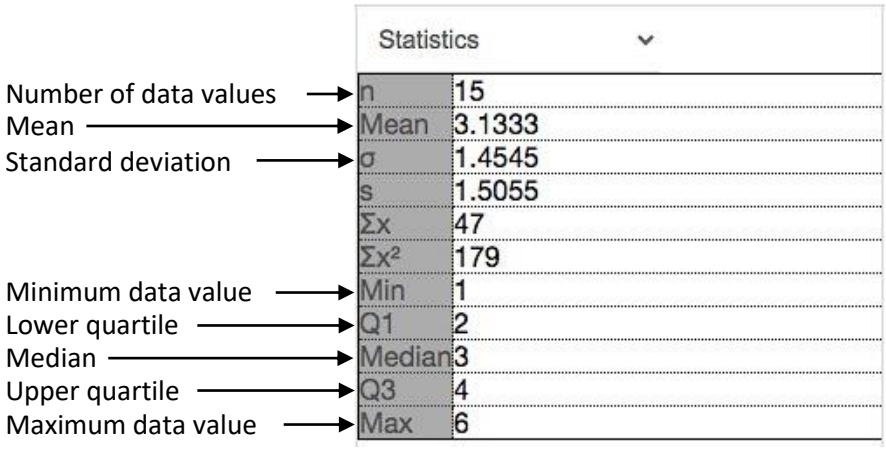
It is not mathematically correct to remove outliers from your analysis unless they are found to be inaccurate (e.g. a person's age recorded as 200 years old or a person's height recorded as 3.5 m). Therefore, if outliers are found, you should do some additional research into that particular data value to assess its accuracy. If it is found to be accurate, it must remain in your analysis moving forwards. If it is found to be inaccurate, remove it from your analysis moving forwards. Ensure you discuss your additional research into the outlier as well as your decision whether to continue to include it in your analysis or not in your report.

3) Use the data in Geogebra to identify the median and interquartile range.

To find statistics such as mean, median, quartiles, standard deviation etc, click on the Σx icon.



The statistics window will then show the following useful statistics:



GDP per capita analysis:

1) Insert a copy of your box plot here (make sure you include the numerical scale and add a label to show the variable name):

2) If there are outliers, which countries are these for?

   .....

   Are you surprised by the finding of outliers / no outliers? Give reasons.

   .....

   .....

   .....

   .....

   .....

3) Median: $.....

   Interquartile range: $.....

4) The spread of the GDP per capita is less than / similar to / greater than (delete as appropriate) the average, meaning there is comparatively small / moderate / large (delete as appropriate) variation.

   Note: Comparing the spread to the average is a useful way to assess whether the variation within the data is large or small while also taking into account the general size of the numbers in the data set.

   The interquartile range is a measure of spread and the median is a type of average. Both of these measures are less affected by outliers than their counterparts the range and the mean, which makes them useful for this assessment.

Are you surprised by the size of the average and/or variation? Give reasons.

.....

.....

.....

.....

.....

Percentage of the population with internet access analysis:

1) Insert a copy of your box plot here (make sure you include the numerical scale and add a label to show the variable name):

2) If there are outliers, which countries are these for?

.....

Are you surprised by the finding of outliers / no outliers? Give reasons.

.....

.....

.....

.....

.....

3) Median: ……%

Interquartile range: ……%

4) The spread of the percentage of the population with internet access is less than / similar to / greater than (delete as appropriate) the average, meaning there is comparatively small / moderate / large (delete as appropriate) variation.

Note: Comparing the spread to the average is a useful way to assess whether the variation within the data is large or small while also taking into account the general size of the numbers in the data set.

The interquartile range is a measure of spread and the median is a type of average. Both of these measures are less affected by outliers than their counterparts the range and the mean, which makes them useful for this assessment. You could also compare the standard deviation to the mean.

Are you surprised by the size of the average and/or variation? Give reasons.

……

……

……

……

……

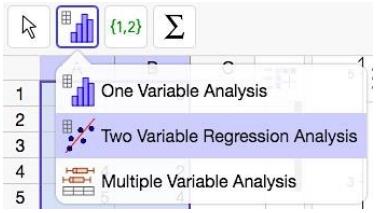
D: Analyse the association between the variables (bivariate analysis)

To help you analyse the association between a county's GDP per capita and the percentage of their population with internet access:

1) Use the data on Geogebra to draw a scatter graph.
2) Use the scatter graph to identify any linear correlation or non-linear association.

How to use Geogebra to do the above:

1) Use the data on Geogebra to draw a scatter graph.
   1) Highlight all of the data in both of the required variables by clicking on the column heading letter of the first variable, then holding the *Shift* key while clicking on the column heading letter of the second variable.
   2) Click on the blue bar chat in the toolbar then click on 'Two Variable Regression Analysis'.



You will now see a scatter graph of your data on the right-hand side of the screen.

1) Insert a copy of the scatter graph here (make sure you include the numerical scales and add labels to show the variable names):

2) The scatter graph shows no association / linear correlation / non-linear association (delete as appropriate).

Are you surprised by the association shown? Give reasons linked to your earlier predictions.

.....

.....

.....

.....

.....

E: Form a conclusion

Summarise your findings from sections B, C and D and form a conclusion relating to the original question: is there an association between a county's GDP per capita and the percentage of its population that has access to the internet?

……

……

……

……

……

……

……

……

……

……

Refer back to your earlier predictions and describe whether or not they were confirmed by the conclusion. If they were not confirmed, suggest possible reasons for this.

……

……

……

……

……

……

……

……

……

F: Reflection

What might your conclusion mean for an internet-based company that is looking to expand into other countries around the world?

……

……

……

……

……

……

……

……

Suggest some ways that this project could be improved.

……

……

……

……

……

……

……

……

Suggest some ways this project could be extended.

……

……

……

……

……

……

……

……

References:

[1]: World Bank Group, 2025. *GDP per capita (current US$)* https://data360.worldbank.org/en/indicator/WB_WDI_NY_GDP_PCAP_CD. Accessed 6th June 2025.

[2]: Word Bank Group, 2025. *Has access to the internet* https://data360.worldbank.org/en/indicator/WB_FINDEX_INTERNET. Accessed 6th June 2025.

IB SL Maths Transition Data Task – EXAMPLE ANSWERS

Note: The answers provided below are purely examples. You may have different predictions and/or reflections based on your own thoughts and the graphs/values produced will vary slightly depending on the data values selected in your random sample. However, you should find *similar* results to those given here.

A: Make a prediction

Use the information provided above as well as your general knowledge to make a prediction about whether there will be an association between a country's GDP per capita and the percentage of its population that has access to the internet:
I predict that there will be an association between a country's GDP per capita and the percentage of its population that has access to the internet because a country's GDP per capita is related to the development status of the country and the wealth of its population, both of which will impact the demand for and availability of internet access.

If you predicted that there will be an association, then also predict the nature of the association:
I predict that the association will be positive, i.e. as the GDP per capita increases, the percentage of the population with access to the internet will increase because people in countries with a higher GDP will typically have more finances available to buy the infrastructure and devices required to allow internet access, and conversely, people in countries with a lower GDP will typically not have as many available funds for this purpose. Additionally, there will be more demand for internet access from businesses in countries with a higher GDP as more businesses here tend to provide services and require digital technologies than those in countries with lower GDPs.

B: Take a sample of the data

Include a copy of your sampled data here, showing the Country Name, GDP per capita and Percentage of the population with internet access:

Country Name	GDP per capita (US$)	% of population with internet access
Italy	36852.54254	90.87201357
Paraguay	5976.931798	72.44016528
Australia	60607.77886	93.78851652
Namibia	4412.835569	50.30946136
Bolivia	3384.844865	81.34078979
Latvia	20262.84339	94.91681457
United States	71318.30736	94.80496049
Spain	30817.68286	91.49932265
Armenia	4685.179971	81.74697757
New Zealand	49624.18167	94.51241493
Russian Federation	12521.52246	92.76232123
Thailand	7058.0695	80.05141616
Iran, Islamic Rep.	4334.792686	68.58661175
Kyrgyz Republic	1349.997307	84.81687307
Nepal	1252.750767	39.92656469
Congo, Rep.	2516.162551	30.15149236
Belgium	51655.78833	93.54858994
Mongolia	4517.61578	82.54463673
Mali	862.4674001	25.74126124
Poland	18635.51049	93.55544448
Malaysia	10903.11164	87.47013211
El Salvador	4642.607431	62.90085316
Myanmar	1242.721344	73.77069592
China	12617.5051	82.9610765
Iraq	4868.494311	72.58049846
Norway	93072.89251	98.88783693
Serbia	9680.527981	83.78220201
Sweden	61174.96802	95.27791739
Burkina Faso	895.5352884	23.75941724
Egypt, Arab Rep.	3827.354154	37.66866028

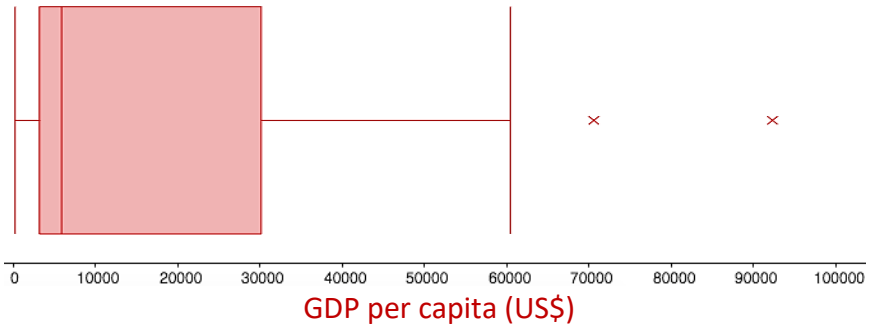
Looking at the data, do you notice anything interesting/surprising/relevant? Give details.
There seems to be a large variation in both variables; GDP per capita ranges from approximately $862 (Mali) to $93100 (Norway) while the percentage of the population with internet access ranges from approximately 25% (Burkina Faso) to 99% (Norway).

It appears that countries with a lower GDP per capita have a lower percentage of the population with internet access, e.g. Mali has GDP per capita approximately $860 and approximately 26% of the population has access to the internet. Additionally, it appears that countries with a higher GDP per capita have a higher percentage of the population with internet access, e.g. Norway has GDP per capita approximately $93000 and approximately 99% of the population has access to the internet. However, there are also countries which do not fit this trend as clearly, e.g. Myanmar has a relatively low GDP per capita of approximately $1200 but a relatively high percentage of the population with internet access of approximately 74%.

C: Analyse each variable in turn (univariate analysis)

GDP per capita analysis:

1) Insert a copy of your box plot here (make sure you include the numerical scale and add a label to show the variable name):



2) If there are outliers, which countries are these for?
There are 2 outliers: Norway and the United States

Are you surprised by the finding of outliers / no outliers? Give reasons.
I am slightly surprised that Norway and the United States have been classified as outliers as their values are not too much higher than the next highest GDPs. However, this highlights the fact that most countries in the sample have GDPs that are further towards the lower end of the scale.

3) Median: $6520 (3 significant figures)
Interquartile range: $27000 (3 significant figures)

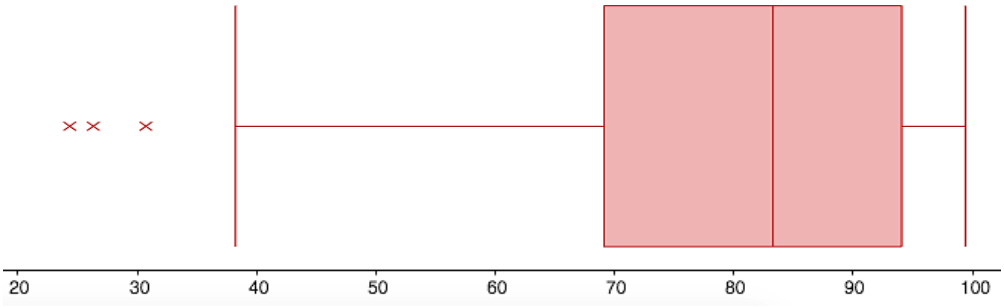
4) The spread of the GDP per capita is greater than the average, meaning there is comparatively large variation.

Are you surprised by the size of the average and/or variation? Give reasons.
The median GDP per capita is lower than I expected it to be; I expected it to be closer to the mid-point between the lowest and highest values as I expected the GDP per capita to be more evenly spread rather than more grouped towards the lower end of the scale.

I am not surprised that there is a large variation in GDP per capita as this highlights the vast differences in economies around the world.

Percentage of the population with internet access analysis:

1) Insert a copy of your box plot here (make sure you include the numerical scale and add a label to show the variable name):



Percentage of the population with internet access

2) If there are outliers, which countries are these for?
There are 3 outliers: Burkina Faso, Mali and the Republic of Congo

Are you surprised by the finding of outliers / no outliers? Give reasons.
I am not really surprised that there are outliers at the lower end of the scale as the majority of the countries in the sample have a high percentage of internet access.

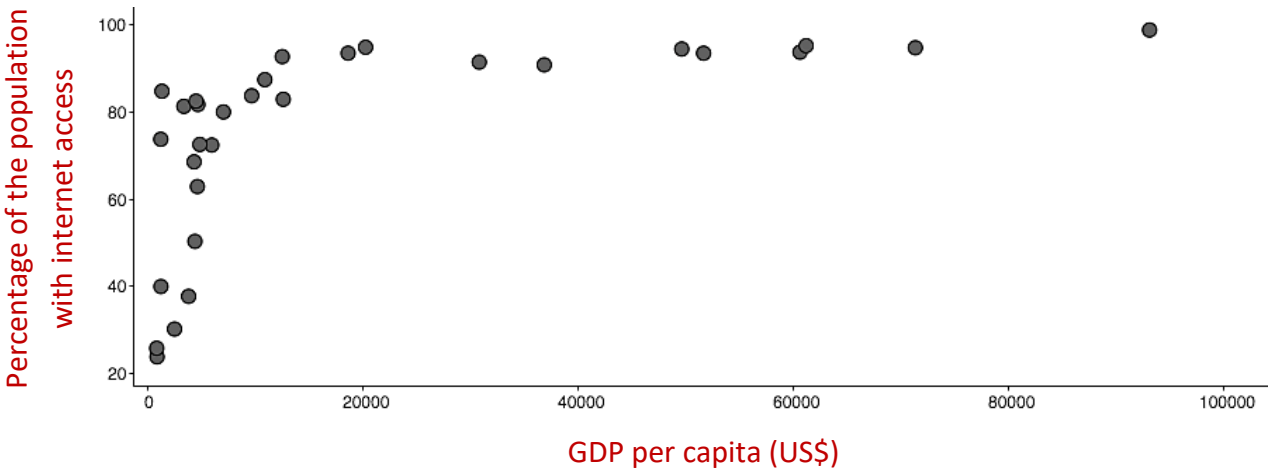
3) Median: 82.8% (3 significant figures)
Interquartile range: 25.0% (3 significant figures)

4) The spread of the percentage of the population with internet access is less than the average, meaning there is comparatively small variation.

Are you surprised by the size of the average and/or variation? Give reasons.
I am not surprised that the median is quite high or that there is quite a small variation because the internet is now very widespread.

D: Analyse the association between the variables (bivariate analysis)

1) Insert a copy of the scatter graph here (make sure you include the numerical scales and add labels to show the variable names):



2) The scatter graph shows non-linear association.

Are you surprised by the association shown? Give reasons linked to your earlier predictions.
I am surprised that the association shows such a strong non-linear pattern; I was expecting to see a more linear relationship because I thought there would be a more linear relationship between the GDP and the funds available for the infrastructure and devices required for internet access.

The scatter graph seems to have two sections, those countries with a GDP per capita of at most $15000 and those with a GDP per capita of more than $15000. Both sections individually show a more linear and positive correlation, however the gradient of the section with the lower GDP is much steeper than that with the higher GDP.

E: Form a conclusion

Summarise your findings from sections B, C and D and form a conclusion relating to the original question: is there an association between a county's GDP per capita and the percentage of its population that has access to the internet?

The average GDP per capita was found to be $6520 (median to 3 sig fig) with an interquartile range of $27000 (3 sig fig), showing a comparatively large variation and the majority of the data values towards the lower end of the scale. There were 2 outliers at the top end of the scale (Norway $93100 and the United States $71300 each correct to 3 sig fig).

The average percentage of the population with internet access was found the be 82.8% (median to 3 sig fig) with an interquartile range of 25.0% (3 sig fig), showing a comparatively small variation and the majority of the data values towards the higher end of the scale. There were 3 outliers at the low end of the scale (Burkina Faso 23.8%, Mali 25.7% and Republic of Congo 30.2% each correct to 3 sig fig).

There was found to be a positive non-linear association between the two variables, meaning that countries with higher GDP tend to have a higher percentage of the population with access to the internet. However, there seemed to be two distinct sections on the scatter graph with countries with a GDP per capita of at most $15000 and those with a GDP per capita of more than $15000. Both sections individually show a more linear and positive correlation, however the gradient of the section with the lower GDP is much steeper than that with the higher GDP. This suggests that there is much more variation in the percentage of the population with access to the internet in countries with lower GDP per capita compared to those with higher GDP per capita.

Refer back to your earlier predictions and describe whether or not they were confirmed by the conclusion. If they were not confirmed, suggest possible reasons for this.

I predicted that there would be a positive association between the two variables, which was what was found. However, I expected to see a more linear correlation instead of the non-linear association that was found. The reason for the non-linear association seeming to show two distinct sections, each with a linear correlation, could be because the percentage of the population with internet access begins to approach its maximum possible value (100%) as the GDP per capita approaches approximately $20000. Therefore, if the GDP per capita continues to increase from this point, it is not possible for the percentage of the population with internet access to continue to increase at the same rate as previously, i.e. the rate of the increase must decrease.

F: Reflection

What might your conclusion mean for an internet-based company that is looking to expand into other countries around the world?

Countries with a lower GDP per capita tend to have a lower percentage of the population with internet access, so the company may struggle to expand into countries with a lower GDP per capita purely because people cannot access their website. Whereas it may be easier for them to expand into countries with a higher GDP because more of the population has internet access.

Suggest some ways that this project could be improved.

To ensure a more representative sample, a larger sample size could be used. Additionally, a stratified sample could be used instead of a simple random sample, where the strata are related to the development status of the country (lower economically developed countries vs more economically developed countries).

Suggest some ways this project could be extended.

The two distinct sections seen on the scatter graph could be individually investigated.

Additionally, you could investigate the association between GDP per capita and other measures of technology usage across the world, such as energy use per capita, mobile phone ownership, percentage of households with a TV, etc.