

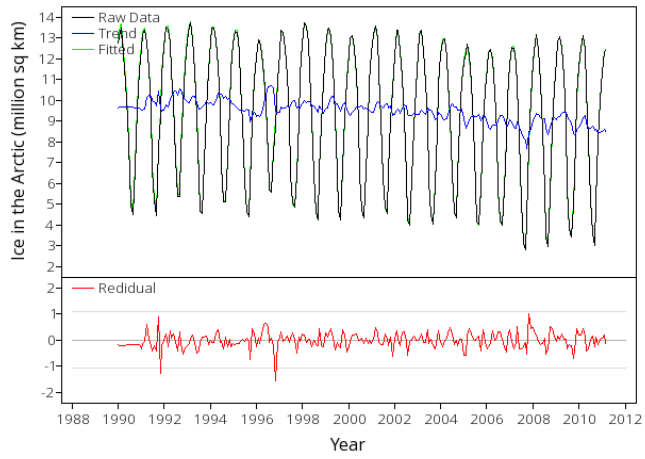
Investigate Time Series Data

Written by J Wills – MathsNZ – jwills@mathsnz.com

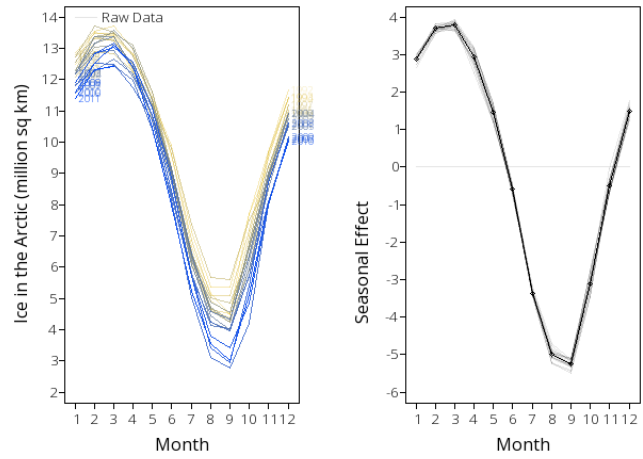
Achievement	Achievement with Merit	Achievement with Excellence
Investigate time series data.	Investigate time series data, with justification.	Investigate time series data, with statistical insight.

Problem.....	4
Purpose	4
What I’m Predicting.....	5
Research and Referencing	6
Plan	8
Explain the Variables	8
Naming the Source.....	9
Data – Using NZGrapher	10
Analysis	12
Long Term Trend	12
Seasonal Pattern	14
Outliers	16
Appropriateness of the Model	17
Predictions	18
Reasons for Confidence	20
Conclusion.....	22
Writing Your Own Internal 1	24
Writing Your Own Internal 2.....	26
Sample Internal (at Achieved Level)	28
Data Set Information.....	30
Assessment Guidelines – 91580 – Investigate Time Series Data.....	32

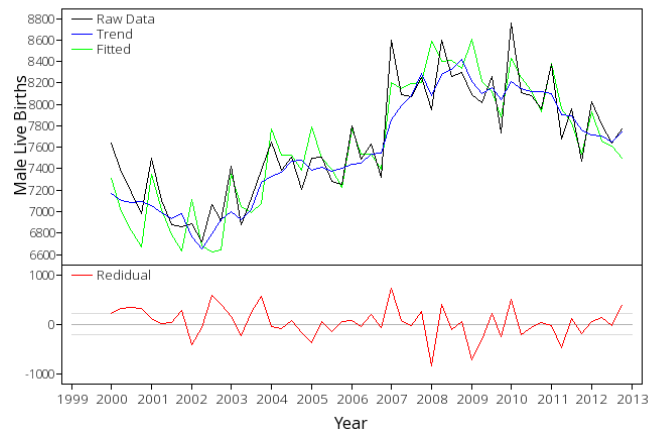
1. Sea Ice



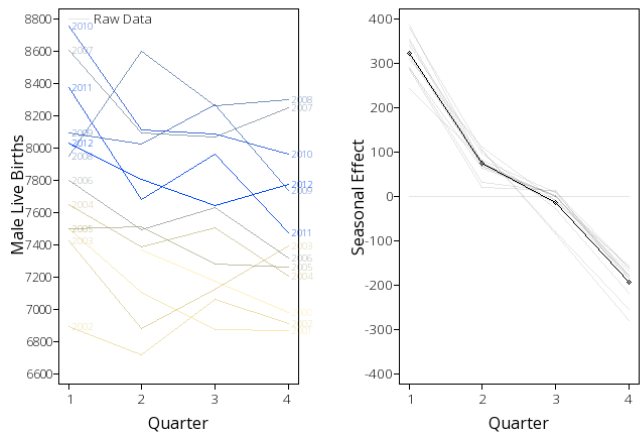
1. Sea Ice



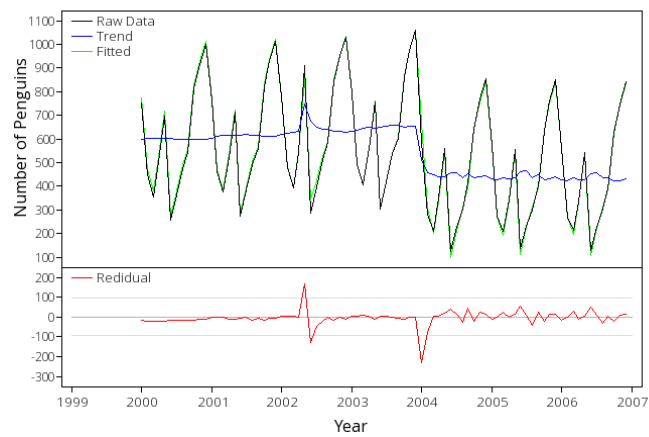
2. Births and Deaths



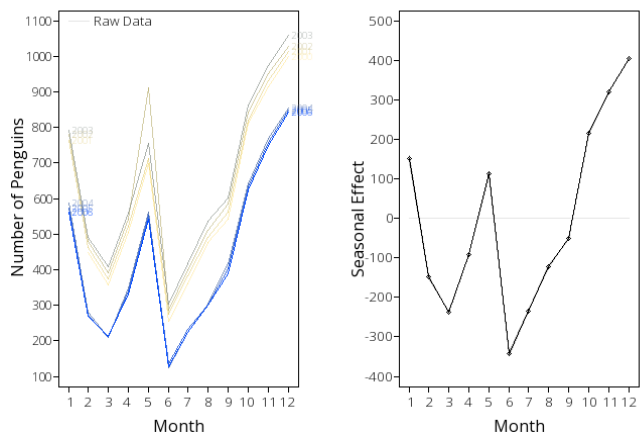
2. Births and Deaths



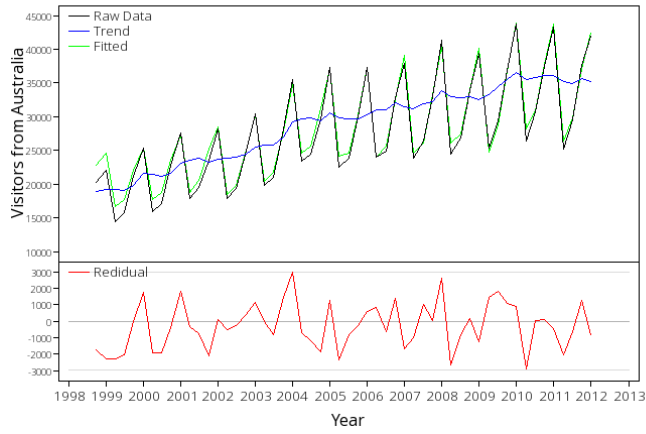
3. Penguin



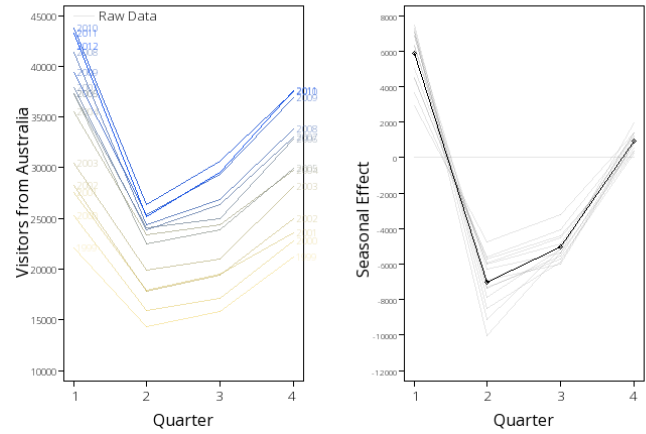
3. Penguin



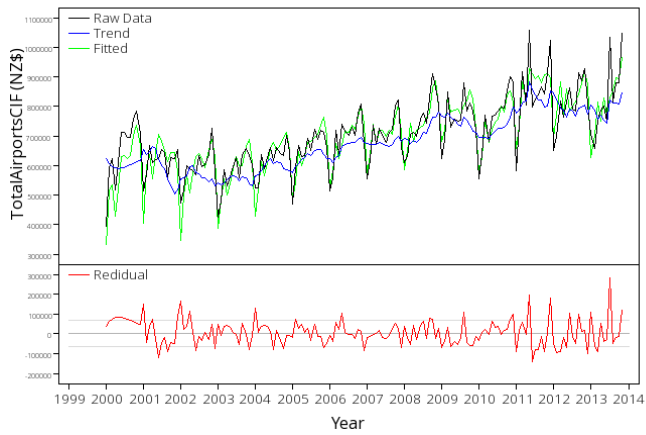
4. Visitors



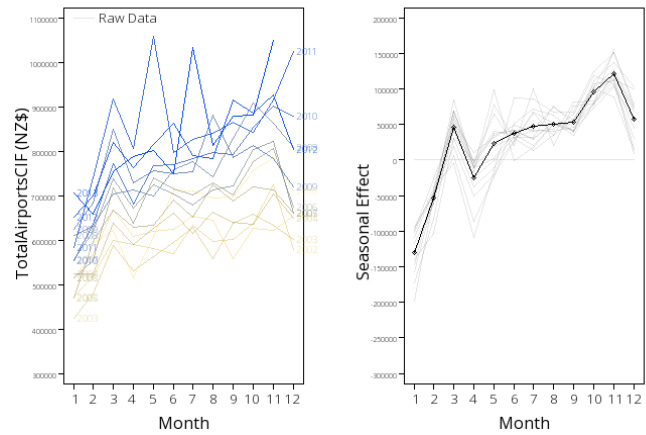
4. Visitors



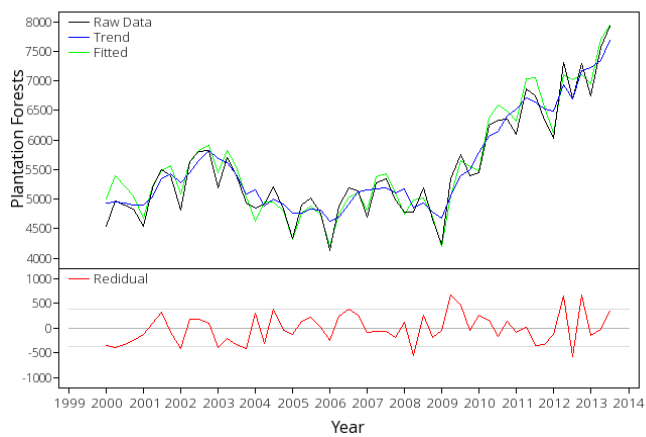
5. Imports



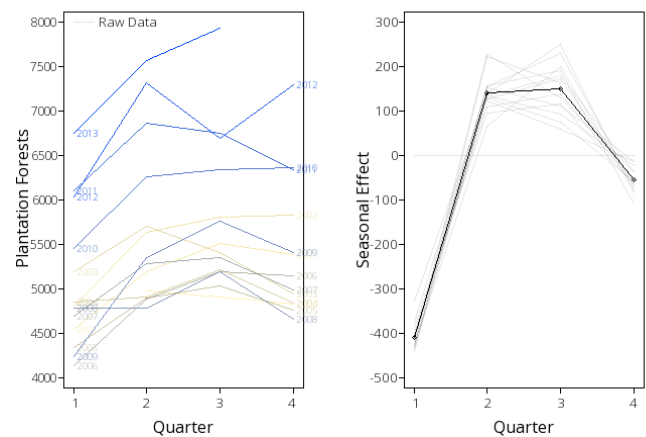
5. Imports



6. Forestry



6. Forestry



Problem

Purpose

For each of the graphs on the previous pages write a good purpose. A purpose should give a reason as to why you would want to look at this dataset. The first two have been done for you.

1. People who live in low lying areas are concerned about the amount of ice in the sea as when it melts they are worried about their homes flooding.
2. A doctor at the local birthing unit has asked you to look at the number of births so that he can work out if he should be employing more staff or not.

3. _____

4. _____

5. _____

6. _____

What I'm Predicting

You then need to state what you are going to be predicting, and link this to your purpose.

NB: iNZight will always give predictions for the next two years.

1. One of the main areas of ice in the world is at the North Pole, or the Arctic, so I am going to predict the surface area of the ice at the North Pole for the next two years.
2. The number of males and females born are approximately the same, so I am going to predict the number of live male births in New Zealand for the next two years.

3. _____

4. _____

5. _____

6. _____

Research and Referencing

For this internal we will again be using the APA style of referencing... you should refer to your last booklet (3.9 Bivariate Data) if you have forgotten how to do it.

For each of the data sets find a relevant source and quote or paraphrase a quote. On the next page write up the bibliography entry for each of the sources that you have used.

1. "Even a modest rise in sea levels could cause flooding problems for low-lying coastal areas."
(Strickland & Grabianowski, 2014)
2. The department of internal affairs records all births, deaths and marriages on a large register. (DIA, 2014)

3. _____

4. _____

5. _____

6. _____

Bibliography entry for each source.

You can use the free online tool: <http://www.bibme.org/> to make the bibliography really easily. Just make sure you use the APA formatting.

1. Strickland, J., & Grabianowski, E. (2005, April 21). How Global Warming Works. *HowStuffWorks*. Retrieved February 8, 2014, from <http://science.howstuffworks.com/environmental/green-science/global-warming4.htm>
2. The Department of Internal Affairs. (n.d.). *Births, deaths and marriages*. Retrieved February 10, 2014, from <http://www.dia.govt.nz/Births-deaths-and-marriages>

3. _____

4. _____

5. _____

6. _____

Plan

Explain the Variables

The next thing that we need to do is explain our variable(s) and say what units are being used.

Identify the variables for each of the graphs on pages 2 and 3.

The first two have been done as examples for you.

1. The sea ice is the surface area of sea ice in the Arctic Circle which is measured in millions of square kilometres.

2. Male live births is the number of males born in each quarter in New Zealand.

3. _____

4. _____

5. _____

6. _____

Naming the Source

In order for our report to have validity, we need to state where the data has come from.

Name the source for each of the graphs on page 2 and 3. The first two has been done as examples for you.

Hint: Information of all of the data sources in in the appendix of this booklet.

1. The data used in this investigation is from the National Snow and Ice Data Center from 1990 - 2011.

2. The data used in this investigation is from Statistics New Zealand from 2000 - 2012.

3. _____

4. _____

5. _____

6. _____

Data – Using NZGrapher

The next section that we need to do is the data section. This is reproducing the graphs on Page 2 using NZGrapher. The example below will go through using the Rugby dataset for weight by position. NZGrapher runs on anything with a browser... Macs, PCs, iPad, Android, ChromeBooks and more.

The example below uses the sea ice dataset.

First up we need to start NZGrapher by going to the link in the box to the right.

www.jake4maths.com/grapher

The first time you load NZGrapher it will display an overlay with descriptions as to what all the different areas do as shown to the right. To load your data in either select it from the dropdown in the top right, or upload it in the top left corner and press go.

The screenshot shows the NZGrapher interface. On the left, there's a table with columns: Id, LowBirth, Birth, and Smoke. The table contains data for various birth records. On the right, there's a 'Data Source' dropdown menu set to 'Babies.csv'. Below the table, there are buttons for 'Add Row', 'Delete Row', 'Add Column', 'Delete Column', 'Save Changes', and 'Reset'. There are also checkboxes for 'Summaries (desktop & laptop)', 'Regression Line (scatter plot)', 'Forecast output (time series)', and 'Point Labels'. The 'Graph Type' is set to 'Time series re-composition'.

To draw a dot plot there are just three things you need to do.

1. Select the x-variable... this is your time series variable that will be on the x-axis, in this case it's 'Time'.
2. Select the y-variable... this is your actual data, in this case it's 'Arctic'.
3. Select the graph type... for this we want the 'time series re-composition'.

You then just need to check the graph title and axis labels to make sure they are appropriate (include units where necessary) and add press update graph to save the titles.

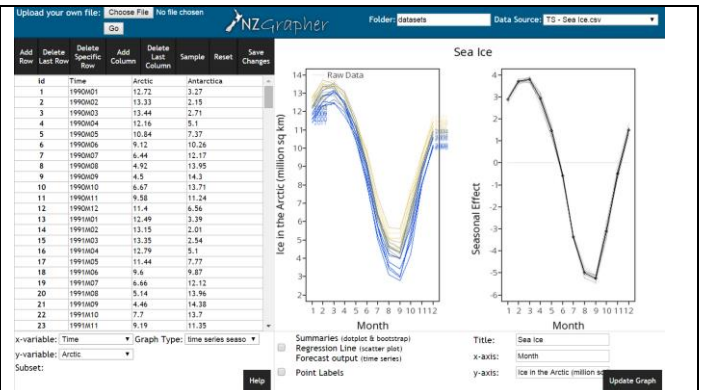
To save the graph just right click on it and press 'Save Image As' or whatever your device says that is similar.

Note: If you want to identify the extreme points, if you click the 'Point Labels' checkbox this will add little numbers next to the points that correspond with the point id.



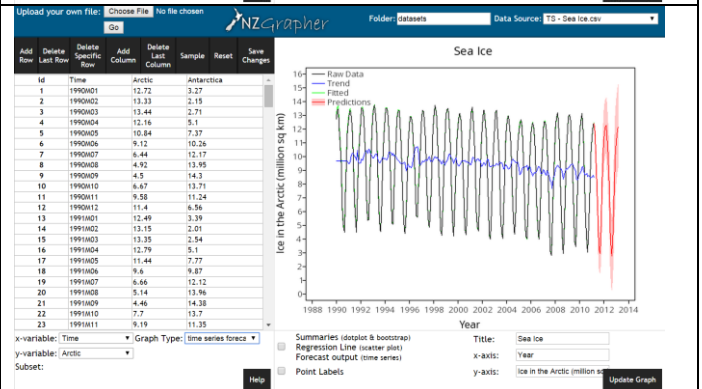
To get the seasonal effect graph just change the graph type to 'time series seasonal effects'

Note: you will need to change the x-axis title to reflect if it is monthly or quarterly (or something else).



The final graph we need is the forecast graph.
To get the visual output you just change the graph type to 'time series forecasts'

You will also need the text output of the predictions which you can get by clicking on the tick box below the graph.



Now it's your turn. For each dataset produce all the outputs shown above.

Teachers: for tips and tricks on how to format data for time series click on the time series link on:

<http://www.mathsnz.com/inzight-tips/>

Analysis

We now start on the analysis section of our report. In all of these sections you should be attempting to explain the cause of what you can see, ideally with references to back up your statements.

Long Term Trend

For the long term trend we want to be discussing the overall trend. This is what is happening to the blue line on our recomposed data graphs. Is it increasing, decreasing or doing something else? It is good to give numbers for different dates to back up your comments. For higher grades you should be giving possible reasons and link to your purpose.

For each of the graphs on pages 2 and 3 write a comment about the long term trend.

The first two have been done for you.

1. By looking at the recomposed data I can see that overall the amount of sea ice in the artic area appears to be generally decreasing from about 10 million square kilometres on average in 1990 to approximately 8 million square kilometres on average in 2011.
2. By looking at the recomposed data I can see that overall the amount of male live births has increased since 2000. In 2000 there were approximately 7400 births per quarter on average, this dropped to approximately 6800 per quarter on average in 2002, climbed to a peak of 8200 per quarter on average in 2009, possibly due to the financial boom, and has dropped off slightly to approximately 7800 per quarter on average in 2013.

3. _____

[illegible][illegible][illegible]

Seasonal Pattern

With the seasonal pattern we are looking at what is happening in each month or each quarter. You want to be commenting on any highs or lows and also the change between them. For higher grades you should be giving possible reasons and link to your purpose.

For each of the graphs on pages 2 and 3 write a comment about the long term trend.

The first two have been done for you.

1. The largest amount of sea ice at the North Pole normally occurs during February and March when the amount of sea ice is approximately 4 million square kilometres above the long term trend. Between April and August there is a consistent drop from month to month as temperatures are rising. The smallest amount of sea ice regularly occurs in September where the amount of sea ice is approximately 5 million square kilometres below the long term trend. After this the temperatures start cooling down again increasing the amount of ice on a month to month basis until it returns to its peak in February and March.
2. The seasonal pattern for male live births is not very consistent. On average the highest number of births occurs in the first quarter (January to March) where there are approximately 220 more births than the long term trend. The lowest on average occurs in the fourth quarter (October to December) where there are approximately 180 births less than the long term trend, however there are many years that it does not follow this pattern.

[illegible]

This image shows a blank sheet of white paper with horizontal ruling lines. The lines are evenly spaced and extend across the width of the page. There are no margins, text, or other markings on the paper.

This image shows a single page of white paper with horizontal blue or grey ruling lines, typical of notebook paper. The lines are evenly spaced and run across the width of the page. There is no handwriting or other markings on the paper.

[illegible]

Outliers

The next step involves a little bit of mathematical calculations. To work out if a point is an outlier tend to use the rule of thumb that the residual is more than 10% of the overall spread.

To calculate this we use the following calculation:

$$\frac{\text{Absolute Highest Value} - \text{Absolute Lowest Value}}{10}$$

If any of the residuals are either larger or smaller than this value we need to comment on them and what might be causing them.

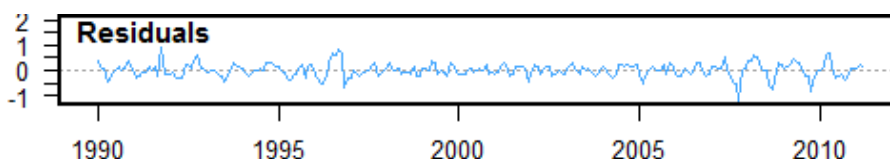
For each of the graphs on pages 2 and 3 calculate if there are any residuals and if there are comment on them.

The first two have been done for you.

1. *Absolute Highest Value: 14,*
Absolute Lowest Value: 3.

$$\frac{14-3}{10} = 1.1$$

Looking at the residuals



graph there is only one point that is more than 1.1 million square kilometres away from the trend. This occurred in September 2007 and may have been due to an unusually hot summer.

2. *Absolute Highest Value: 8700, Absolute Lowest Value: 6700, $\frac{8700-6700}{10} = 200$*

Looking at the residuals graph due to the inconsistency of the data there are a large number of residuals between 2007 and 2010 that are outside the acceptable range. This is during the financial boom so it may be due to people being more willing to have children and therefore not worrying about the timing so much, therefore the normal patterns do not happen.

3. _____

4. _____

5. _____

6. _____

Appropriateness of the Model

For this you need to make a comment around how well the model fits the data by looking at how the recomposed (trend + seasonal) data fits with the raw data, particularly commenting on any time periods where it does not fit particularly well. Try and link these to how accurate you think the predictions will be.

For each of the graphs on pages 2 and 3 comment on how well the model fits the data.

The first two have been done for you.

1. Overall the model seems to fit the data very well. Most of the fitted values match up with the raw data, however there does appear to be an increase in the magnitude of the residuals since 2008 which may be due to an increase in climate change altering the pattern.
2. Due to the highly variable nature of the seasonal effects the model doesn't fit the data particularly well. It is not too bad before 2006 and appears to have settled down after 2011. Provided the pattern remains settled the predictions should be reasonably accurate.

3. _____

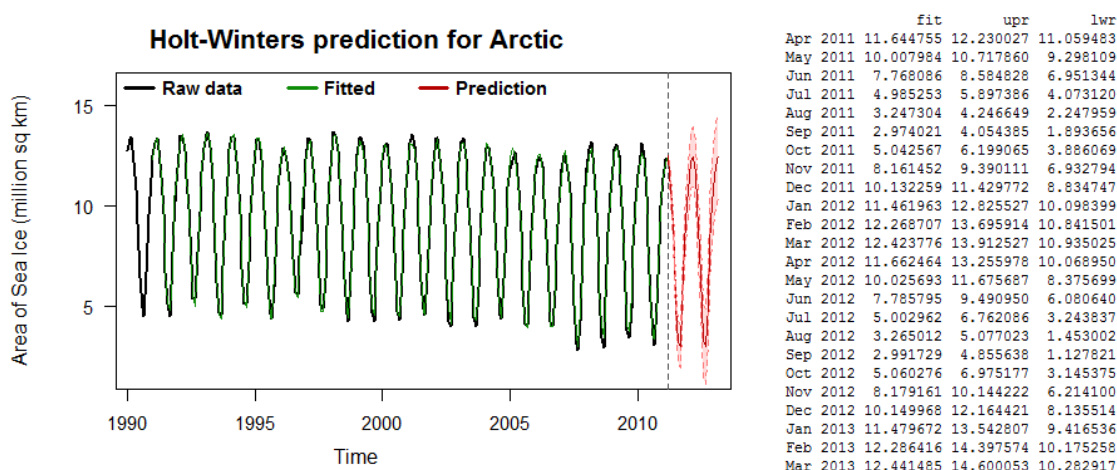
4. _____

5. _____

6. _____

Predictions

At this point in your report you should include the forecast graph and the prediction output we produced earlier eg:



We then need to write up **two** of these predictions in sentences. In these sentences it is VITAL you include the units, and round the numbers appropriately, and not only give the exact number as well as a range that it is likely to be between.

For each of the graphs on pages 2 and 3 make two predictions, the first two have been done for you.

1. In April 2011 I would expect the amount of sea ice for the Arctic to be 11.64 million square kilometres. However I cannot be completely certain about my prediction, but I would expect the area of ice will between 11.06 million square kilometres and 12.23 million square kilometres.
In March 2013 I would expect the amount of sea ice for the Arctic to be 12.44 million square kilometres. However I cannot be completely certain about my prediction, but I would expect the area of ice will between 10.28 million square kilometres and 14.60 million square kilometres.
2. In January to March 2013 I would expect there to be approximately 8060 male live births. However I cannot be completely confident in my prediction, but I would expect the number of male live births for this quarter to be somewhere between 7480 and 8640.
In October to December 2014 I would expect there to be approximately 7440 male live births. However I cannot be completely confident in my prediction, but I would expect the number of male live births for this quarter to be somewhere between 6210 and 8660.

3. _____

4.

5.

6.

Reasons for Confidence

With our predictions we normally cannot be completely confident in our predictions. It is useful to discuss other factors that may be influencing the predictions and how much effect these will have on your predictions.

This is a great time to include some extra research.

For each of the graphs on pages 2 and 3 make comments about your predictions.

The first two have been done for you.

1. We can't be completely certain about our predictions as there are a number of factors that could be affecting the amount of sea ice. One of these factors is possibly global temperatures which is affected by sun spots (NASA, 2014).
2. We can't be completely certain about our predictions as there are a number of factors that could be affecting the number of male live births. One of these factors is the economic conditions as when times are tougher people think more before having children, so it depends on what happens to New Zealand's economy as to how many live births there are likely to be.

3. _____

4. _____

5. _____

This image shows a single sheet of white paper with horizontal ruling lines. The lines are evenly spaced and run across the width of the page. There are no margins, text, or other markings on the paper.

Conclusion

In the conclusion we need to summarise and link back to the purpose. It is important to:

- Give a brief summary
- Link back to the research... do our findings match up?
- What is the impact of these findings?

For each of the graphs on pages 2 and 3 write up a conclusion, the first two have been done for you.

1. In conclusion, there appears to be a definite decrease in the amount of sea ice at the North Pole. This should be of concern to people living in low lying areas as this means the sea ice levels are likely to be rising, potentially putting their homes at risk of flooding.
2. In conclusion, the amount of male births fluctuates widely depending on the economic climate, therefore if the doctor feels like economic times are improving he might want to employ more staff, but if he thinks economic times are likely to get tougher then he should not employ more staff.

3. _____

4. _____

This image shows a single sheet of white paper with horizontal blue or grey ruling lines. The lines are evenly spaced and run across the width of the page, typical of notebook paper. There are no margins, text, or other markings on the page.

[illegible]

able to write up your own internal. Don't forget to give your report a title.

Writing Your Own Internal 1

Using the framework below write a report on the diamonds data. You can use the sample internal at the end of the booklet to help you if you need it.

Sunglasses

Title Is
Given

Problem

Purpose

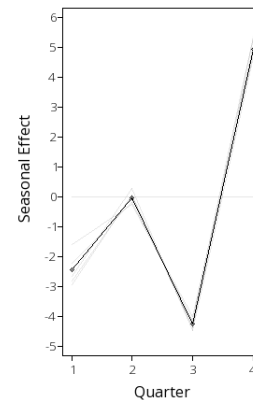
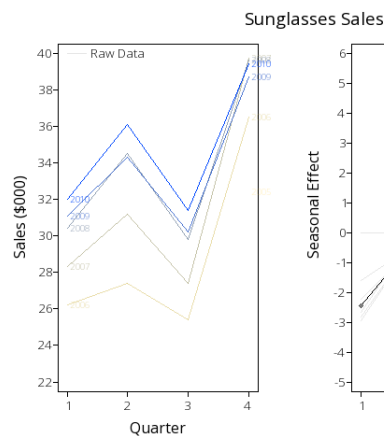
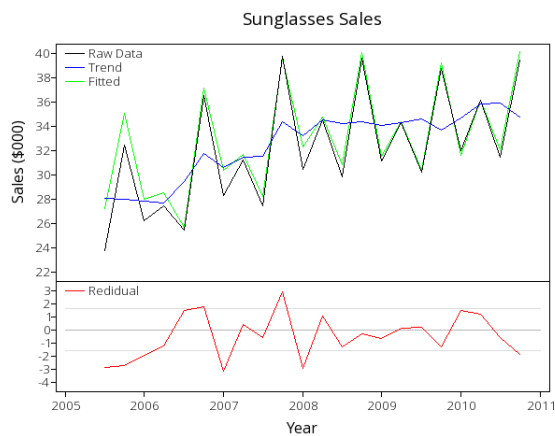
What I'm
Predicting

Plan

Explain the
Variables

Naming the
Source

Data



The Graph
including
adding the
axis label.

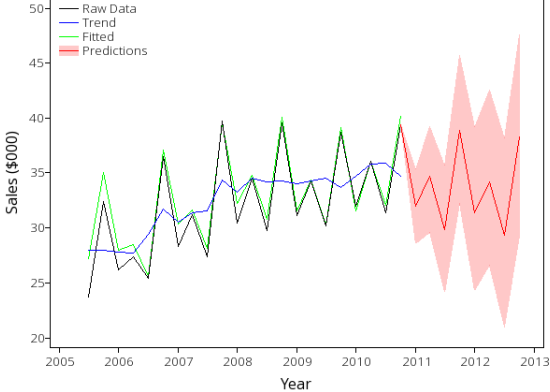
Analysis

Long Term
Trend

Seasonal
Pattern

[illegible]

Predictions



Time	Min	Prediction	Max
2011Q1	28.793	31.942	35.293
2011Q2	30.055	34.735	39.331
2011Q3	23.999	29.863	35.549
2011Q4	32.264	38.859	45.712
2012Q1	23.875	31.394	39.520
2012Q2	25.322	34.187	42.582
2012Q3	20.061	29.314	38.583
2012Q4	29.006	38.311	48.124

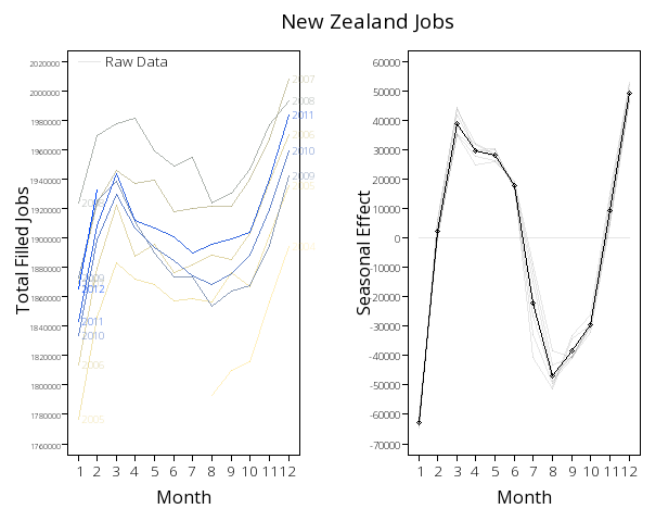
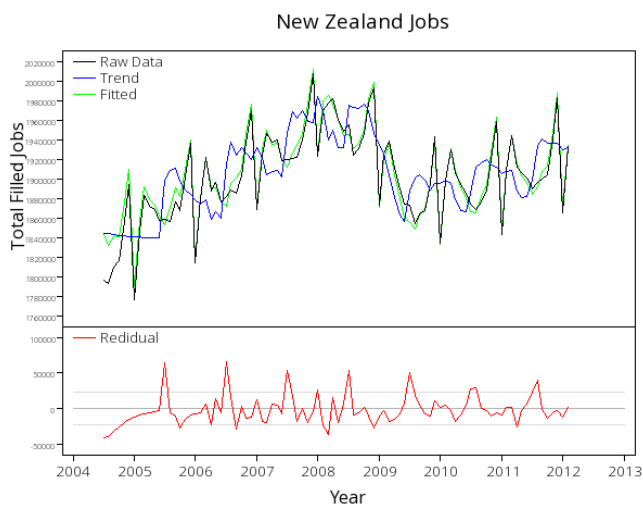
Quarter	2011Q1	2011Q2	2011Q3
2011Q1	32.264	38.859	45.712
2012Q1	23.875	31.394	39.520
2012Q2	25.322	34.187	42.582
2012Q3	20.061	29.314	38.583

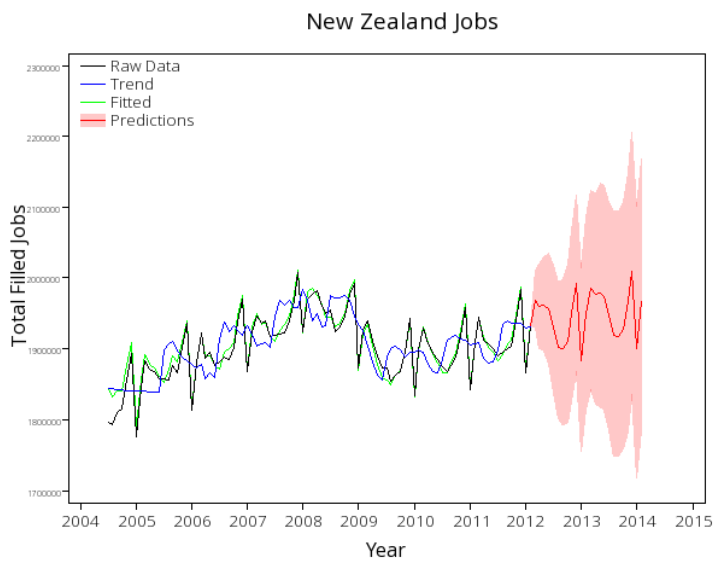
This image shows a blank sheet of white paper with horizontal ruling lines. The lines are evenly spaced and run across the width of the page. There are no margins, text, or other markings on the paper.

Conclusion

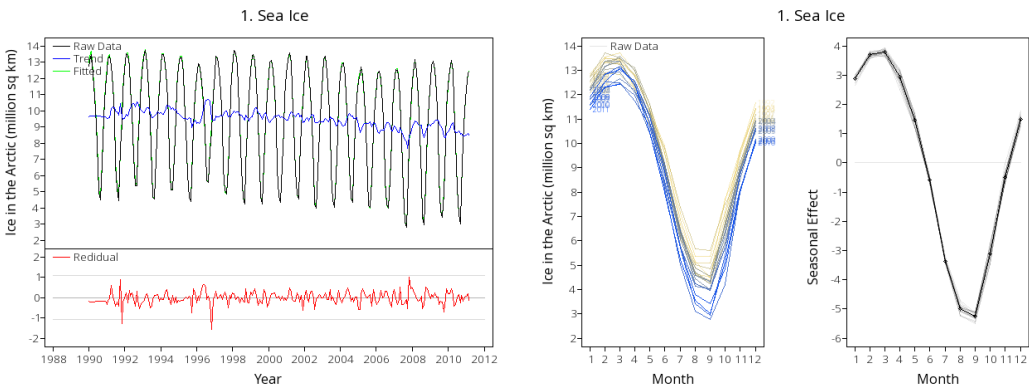
Writing Your Own Internal 2

This time you have just been provided with a title and graphs. Using these write your own internal.





Time	Min	Prediction	Max
2012M03	1929100	1968000	2009800
2012M04	1900300	1959200	2018800
2012M05	1892800	1961600	2033100
2012M06	1876600	1956300	2036700
2012M07	1836400	1929900	2020100
2012M08	1794500	1901600	2000800
2012M09	1788500	1900800	2005100
2012M10	1792700	1910900	2026600
2012M11	1831600	1949000	2068500
2012M12	1866400	1992600	2117400
2013M01	1753000	1883600	2012800
2013M02	1805300	1949700	2083300
2013M03	1837100	1985300	2125800
2013M04	1827600	1976500	2121900
2013M05	1818900	1978900	2126800
2013M06	1813500	1973700	2126400
2013M07	1779100	1947200	2110600
2013M08	1750400	1919000	2091400
2013M09	1739600	1918200	2096600
2013M10	1754300	1928200	2112700
2013M11	1788400	1966300	2152300
2013M12	1822200	2009900	2203200
2014M01	1709300	1900900	2090500
2014M02	1771100	1967000	2156300

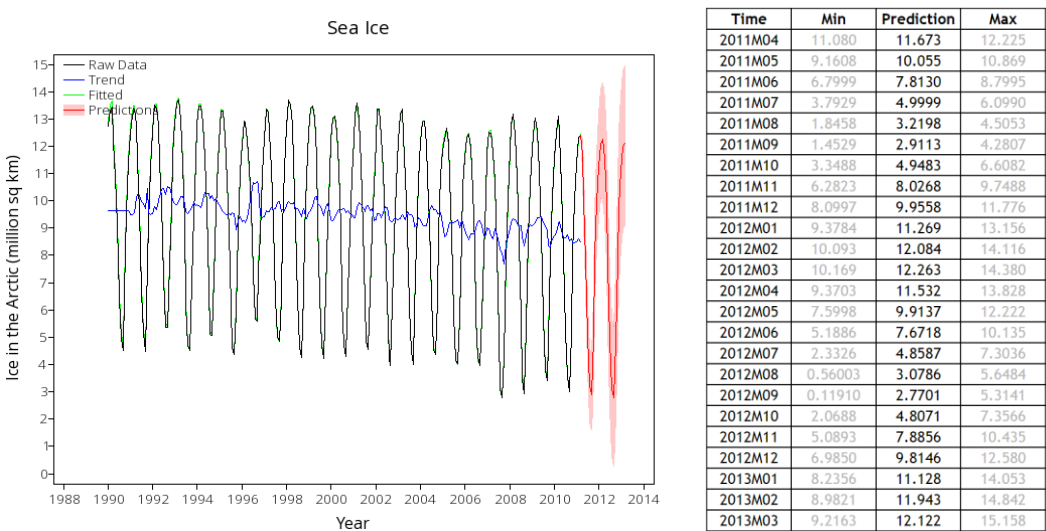
<h2>Ice at the North Pole</h2>		Title Is Given
<p>Problem</p> <p>People who live in low lying areas are concerned about the amount of ice in the sea as when it melts they are worried about their homes flooding.</p>	Purpose	
<p>One of the main areas of ice in the world is at the North Pole, or the Arctic, so I am going to predict the surface area of the ice at the North Pole for the next two years.</p>	What I'm Predicting	
<p>Plan</p> <p>The sea ice is the surface area of sea ice in the Arctic Circle which is measured in millions of square kilometres.</p>	Explain the Variables	
<p>The data used in this investigation is from the National Snow and Ice Data Center from 1990 - 2011.</p>	Naming the Source	
<p>Data</p> <div></div>	The Graph including adding the axis label.	
<p>Analysis</p> <p>By looking at the recomposed data I can see that overall the amount of sea ice in the arctic area appears to be generally decreasing from about 10 million square kilometres on average in 1990 to approximately 8 million square kilometres on average in 2011.</p>	Long Term Trend	
<p>The largest amount of sea ice at the North Pole normally occurs during February and March when the amount of sea ice is approximately 4 million square kilometres above the long term trend. Between April and August there is a consistent drop from month to month as temperatures are rising. The smallest amount of sea ice regularly occurs in September where the amount of sea ice is approximately 5 million square kilometres below the long term trend. After this the temperatures start cooling down again increasing the amount of ice on a month to month basis until it returns to its peak in February and March.</p>	Seasonal Pattern	
<p>Absolute Highest Value: 14, Absolute Lowest Value: 3. $\frac{14-3}{10} = 1.1$</p>	Outliers	

Looking at the residuals graph there is only one point that is more than 1.1 million square kilometres away from the trend. This occurred in September 2007 and may have been due to an unusually hot summer.

Overall the model seems to fit the data very well. Most of the fitted values match up with the raw data, however there does appear to be an increase in the magnitude of the residuals since 2008 which may be due to an increase in climate change altering the pattern.

Appropriateness of the Model

Predictions



Output from NZGrapher

In April 2011 I would expect the amount of sea ice for the Artic to be 11.64 million square kilometres. However I cannot be completely certain about my prediction, but I would expect the area of ice will between 11.06 million square kilometres and 12.23 million square kilometres.

In March 2013 I would expect the amount of sea ice for the Artic to be 12.44 million square kilometres. However I cannot be completely certain about my prediction, but I would expect the area of ice will between 10.28 million square kilometres and 14.60 million square kilometres.

Putting in Context

We can't be completely certain about our predictions as there are a number of factors that could be affecting the amount of sea ice. One of these factors is possibly global temperatures which is affected by sun spots.

Confidence in Predictions

Conclusion

In conclusion, there appears to be a definite decrease in the amount of sea ice at the North Pole. This should be of concern to people living in low lying areas as this means the sea ice levels are likely to be rising, potentially putting their homes at risk of flooding.

Summarise and Link Back to the Purpose

Data Set Information

Births and Deaths

Data on the number of births and deaths in New Zealand.

The data is sourced from Statistics New Zealand.

Variable	Description
Quarter	Quarterly
Male Live Births	Number of males born during the quarter
Female Live Births	Number of females born during the quarter
Male Deaths	Number of male deaths during the quarter
Female Deaths	Number of female deaths during the quarter

Forestry

The volume of wood removed from different types of forests in New Zealand.

The data is sourced from the Ministry for Primary Industries.

Variable	Description
Quarter	Quarterly
Natural Forests	The volume of wood removed from Natural Forests in millions of m ³
Plantation Forests	The volume of wood removed from Plantation Forests in millions of m ³

Imports

Information on imports to and from New Zealand.

The data is sourced from Statistics New Zealand.

Variable	Description
Month	Monthly
TotalAirportsCIF	Cost, insurance and freight of imported goods in NZ\$(000)
TotalParcelPostCIF	Cost, insurance and freight of imported goods in NZ\$(000)
TotalSeaportsCIF	Cost, insurance and freight of imported goods in NZ\$(000)
TotalAirportsWeight	Weight of imported goods in tonnes
TotalParcelPostWeight	Weight of imported goods in tonnes
TotalSeaportsWeight	Weight of imported goods in tonnes

Jobs

The number of people in employment in New Zealand.

The data is sourced from Statistics New Zealand.

Variable	Description
Month	Monthly
Total Filled Jobs	The number of jobs that are filled

Penguin

Data on the number of penguins at the Phillip Island Penguin Parade in Australia.

Variable	Description
Month	Monthly
Number	The number of penguins in the colony

Sea Ice

The data is the surface area of sea ice in millions of square kilometres.

The data is sourced from the National Snow and Ice Data Center.

Variable	Description
Time	Monthly
Arctic	Million Square Kilometres of Ice in the Arctic
Antarctica	Million Square Kilometres of Ice in Antarctica

Sunglasses

Data on the value of sunglasses sold.

Variable	Description
Quarter	Quarterly
Sales	Amount of sales in thousands of dollars

Visitors

The visitors' dataset is the number of people entering New Zealand on a Visitor Visa from Australia, China, Japan and the UK.

The data is sourced from Statistics New Zealand.

Variable	Description
Date	Quarterly
Australia	Number of visitors in the quarter from Australia
China, People's Republic of	Number of visitors in the quarter from China
Japan	Number of visitors in the quarter from Japan
United Kingdom	Number of visitors in the quarter from the UK

Assessment Guidelines – 91580 – Investigate Time Series Data

	Achieved (all compulsory)	Merit... Achieved PLUS	Excellence... Merit PLUS
Problem	Identify a purpose for the investigation	Purpose is clear (compulsory)	Research is used to develop purpose (compulsory)
Plan	Select a variable to investigate that links to the purpose	Context is researched	
Data	<p>Graph the raw and the smoothed data</p> <p>Appropriate model is fitted and can be given as an equation, a fitted line or a graph of the smoothed data</p> <p>Must have the following:</p> <ul style="list-style-type: none"> • Recomposition graph • Seasonal Plot and estimated seasonal effects • Raw data plus predictions plus prediction intervals <p><i>Although iNZight does not include units you should write the units on the graph by hand or by using text boxes</i></p>	Graph must have a title, correctly labelled axes and any series shown on the graph(s) are clearly identified	Other relevant variables are discussed. This could include creating a new variable from the variables given
Analysis	<p>Give quantitative description of the trend and linked to context</p> <p>Seasonal pattern described and linked to context</p> <p>Must discuss:</p> <ul style="list-style-type: none"> • Long term trend • Seasonal effects • Residuals <p>Other relevant features are identified</p> <p>Forecasts are made with correct units</p>	<p>The appropriateness of the model is justified throughout the entire range of x-values</p> <p>Other relevant features must be explained in context and comments made must be supported with statistical evidence</p> <p>Forecasts are given in context and rounded correctly</p> <p>Discussion given on how precise the predictions might be based on reliability of the trend or seasonal components</p> <p>Understanding shown that forecasts are estimates</p>	<p>Provide possible explanations for the features of the graph</p> <p>A comparison of predicted values for the most recent data values of the model could be made</p> <p>Comparison between two data sets could be made</p>
Conclusion	Conclusion is consistent with the purpose of the investigation	Conclusion must be given in context (compulsory)	<p>The purpose of the investigation is addressed (compulsory)</p> <p>There is a reflection on the analysis with respect to the background research undertaken</p> <p>Impact of the findings is discussed</p>

Final grades will be decided using professional judgement based on a holistic examination of the evidence provided against the criteria in the Achievement Standard.