



National Certificate of Educational Achievement
TAUMATA MĀTAURANGA Ā-MOTU KUA TĀEA



MINISTRY OF EDUCATION
Te Tāhuhu o te Mātauranga

2006

Internal Assessment Resource

Subject Reference: **Statistics and Modelling 3.5**

Resource Title: “**Animal Antics**”

Achievement standard: **90645 version 2**

Standard title: *Select and analyse continuous bi-variate data*

Credit: 3

This resource has been trialled in a school and includes annotated examples of assessed student work. There are eight documents in this resource:

Task and schedule	<input type="checkbox"/>	Student 1 EXCELLENCE	<input checked="" type="checkbox"/>
Assessment guidelines	<input type="checkbox"/>	Student 2 MERIT	<input type="checkbox"/>
Teaching notes	<input type="checkbox"/>	Student 3 ACHIEVED	<input type="checkbox"/>
		Student 4 ACHIEVED	<input type="checkbox"/>
		Student 5 NOT ACHIEVED	<input type="checkbox"/>

Internal assessment resource reference number:
Statistics and Modelling/3/5 X version 1

Date version published:
Ministry of Education
Quality assurance status

July 2006
For use in internal assessment
from 2006

EXAMPLE OF ASSESSED STUDENT WORK

ASSESSMENT COVER SHEET FOR STUDENT 1

EXCELLENCE

	Achievement Criteria	Code	Evidence	Judgement (refer to Assessment Schedule for judgement statements)	Sufficiency
Achievement	Select and analyse bi-variate continuous data	A	Purpose stated.	✓	All four of code A
		A	Scatterplot drawn.	✓	
		A	Regression line obtained.	✓	
		A	Relationship explained in context.	✓	
Achievement with Merit	Carry out an in-depth analysis of bi-variate data	M	Relationship between two pairs compared with explanation.	✓	Achievement plus THREE of code M
		M	Regression equations used to obtain predictions.	✓	
		M	Appropriateness of regression model(s) discussed.	✓	
		M	R ² values interpreted correctly.		
		M	Difference between correlation and causality explained.		
Achievement with Excellence	Report on the validity of the analysis	E	Assumptions about the data stated.	✓	Merit plus THREE of code E
		E	Limitations of model given.		
		E	Piecewise or other models proposed and justified, and/or outliers identified and an approach to dealing with them suggested.	✓	
		E	Relevance and usefulness of the evidence explained.		
		E	Applicability of findings stated.	✓	

Final Grade Awarded

N		A		M		E	✓
----------	--	----------	--	----------	--	----------	---

EXAMPLE OF ASSESSED STUDENT WORK

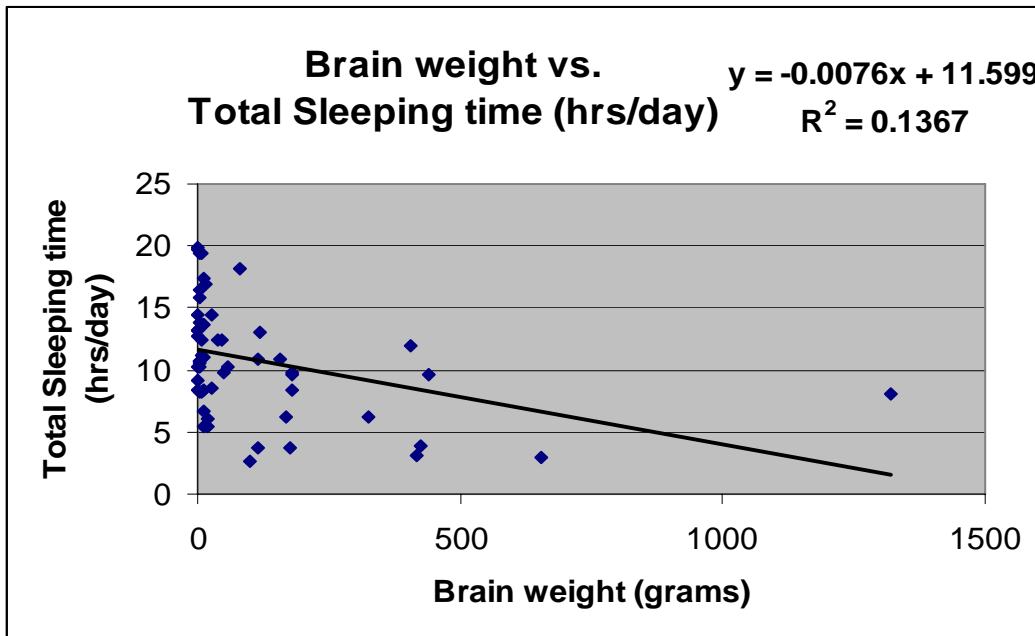
STUDENT 1 EXCELLENCE

Bivariate Data Investigation

Data was provided about 62 species of animals and their brain weight, sleeping patterns, life span, gestation, predation, danger and exposure. The data about gestation, predation, danger and exposure are discrete and cannot be used.

Is there a relationship between Brain Weight (predictor variable) and Total Sleep (response variable) and to compare this with the relationship between Brain Weight (predictor variable) and Life Span (response variable).

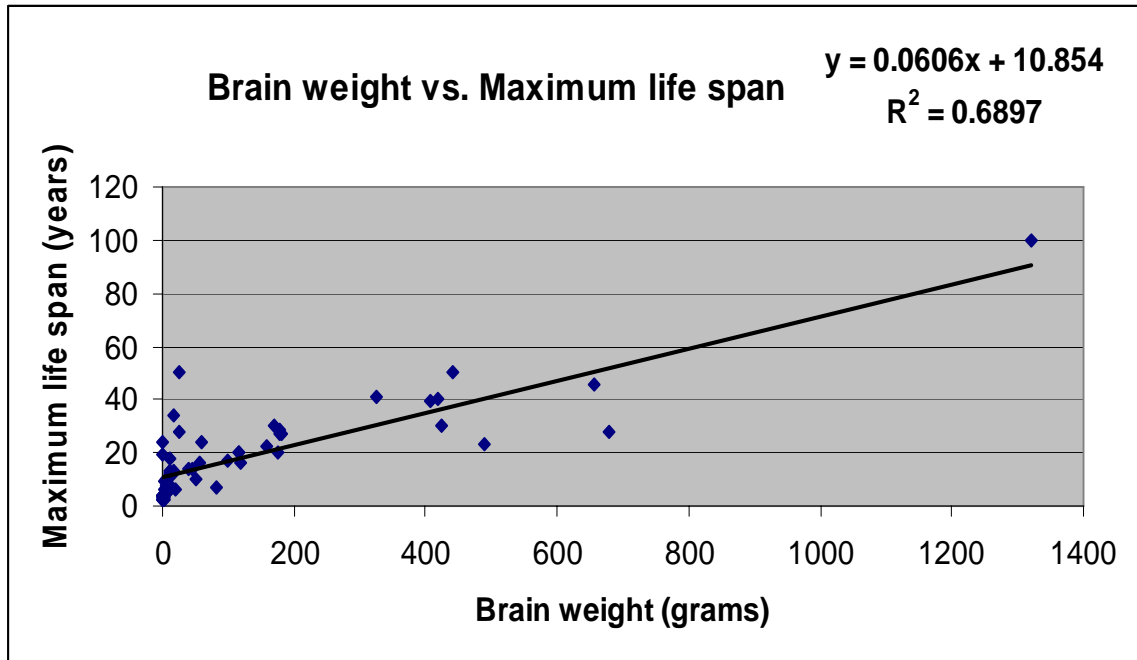
Purpose is stated.
Predictor variables selected and stated are appropriate.
A



This graph is a scatter graph of Brain weight vs. Total Sleeping time (hrs/day). The line is the best fit straight line relationship between the two variables. The equation is $y = -0.0076x + 11.599$. The correlation coefficient is $r = -0.370$. This value of r indicates that total sleep decreases as brain weight increases and that there is a weak linear relationship between the two variables.

This means that an increase in one gram of Brain weight produces a decrease of 0.0076(hrs/day) Total Sleeping time.

Scatter graph drawn, regression line found and interpreted (both gradient and r) in context. **A**



This graph is a scatter graph of Brain weight vs. Maximum life span.

The line is the best fit straight line relationship between the two variables.

The equation is $y = 0.0606x + 10.854$

$r = 0.83$

This value of r indicates that maximum life span increases with brain weight and that there is a strong linear relationship between the two variables.

This means that an increase in one gram of Brain weight produces an increase of 0.0606(years) life span.

Comparing the relationships

Scatter graphs

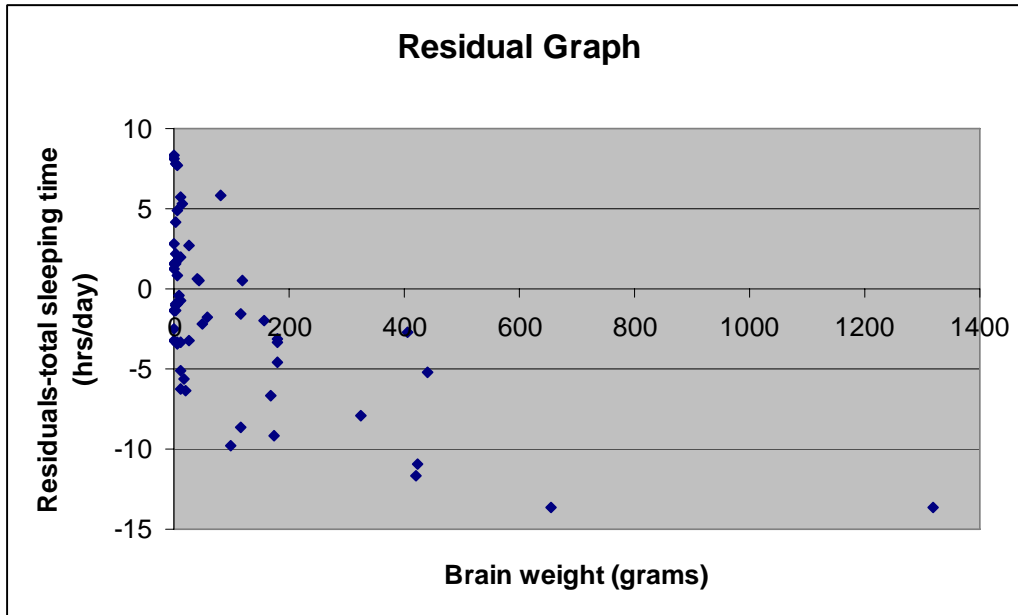
The relationship between Brain weight and Maximum life span is stronger than that between Brain weight and Total Sleeping time because the plots on the graph are closer to the trendline.

R^2 value

The R^2 value for Brain weight and Maximum life span is 0.6897 and for Brain weight and Total sleeping time the R^2 value is 0.1367. This confirms that the Brain weight and Maximum life span relationship is stronger as the R^2 value is larger.

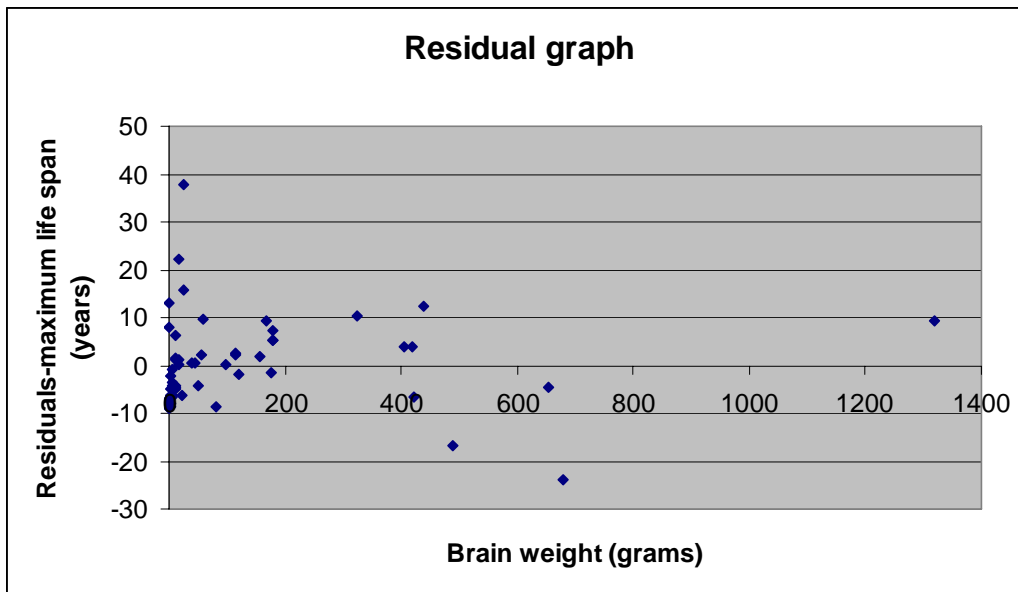
Relationships are compared and appropriateness of the models discussed. A visual comparison is supported by R^2 . **M**

Residuals



This residual graph shows a pattern in that the lower values are both equally positive and negative but as the values get higher the residuals are all negative. Therefore you cannot put a straight trendline on this graph and a linear model would not be suitable.

The final residual is negative but the scatter graph indicates it should be positive.



This residual graph doesn't show much of a pattern as the residuals are spread unevenly. Therefore you could put a straight trendline on this graph and a linear model would be suitable in this case.

The relationships are compared further (through the residuals), with an explanation. M

PREDICTIONS

Brain weight and Total Sleeping time

The linear model predicts that a Brain weight of 179.5 grams will produce a total sleeping time of 10.2348 (hrs/day).

The linear model predicts that a Brain weight of 655 grams will produce a total sleeping time of 6.621 (hrs/day).

Brain weight and Maximum Life span

The linear model predicts that a Brain weight of 5 grams will produce a Maximum life span of 11.157 years.

The linear model predicts that a Brain weight of 1320 grams will produce a Maximum life span of 90.846 years.

Predictions are made in context. **M**

DISCUSSION

Correlation and Causality

Correlation is a relationship that exists between the 2 variables. This was not found in the relationship between Brain weight and Total Sleeping time as there was no pattern in the scatter graph and the trend line did not represent the data appropriately. However correlation was seen in the 2 variables, Brain weight and Life span, as there was a pattern in the scatter graph and the plots were fairly close to the trend line, showing appropriateness of a linear model. Causality is a relationship where one variable causes or affects the other variable. This was found in both the relationships. Brain weight has an effect on the Sleeping time of animals as the larger the animal (*and therefore the larger the Brain weight*), the less time it needs for sleep. E.g. according to the data, a mouse needs 13 hours of sleep a day, compared to a Roe Deer that only needs 3 hours. Smaller animals with a smaller brain have a limited capacity for energy storage. They lose heat energy at a greater rate than larger species because their surface area is proportionately larger to their mass. In addition, larger animals with larger brain sizes generally have more fat insulation and thicker fur than smaller animals which means they are better protected against the loss of heat. Brain weight also has an effect on the life span of an animal as the larger the brain, the longer the life span. This can be seen in elephants and also humans.

There is some discussion on causality, although it is not totally convincing. Causality is defined but not discussed clearly in this context. This is marginal for Merit. However if the weight of evidence for Excellence is sufficient, this could be used to make a holistic judgement and award Excellence.

Interpolation and Extrapolation

The regression model for the first relationship (Brain weight and Total sleeping time) would not be very accurate for interpolation and extrapolation as the linear model is not appropriate and did not represent the data accurately. This was seen as the plots were not close to the line, the R^2 value was very small and the residual graph showed a pattern. The outlier also affects the trendline, further causing inaccuracy when trying to extrapolate and interpolate. The second relationship (Brain weight and Life span) would be more accurate when extrapolating

and interpolating. This is because the linear model and regression represented the data more accurately. The plots were closer to the line, the R^2 value was much higher and the residual graph didn't show a pattern (meaning a linear model is appropriate). And although there was an outlier it did not affect the trendline as much as the last graph. All these factors would make it easier for someone to extrapolate and interpolate on the 2nd graph as it is more accurate.

Good discussion of the appropriateness of the model via interpolation. Extrapolation was not discussed. **M**

Limitations

There were a number of limitations in this investigation. Many pieces of data were missing or N/A (not available). This would have had a definite effect on the outcome of the scatter graphs and linear models. Also, most of the data was the result of a sample which had been averaged. The mean of this sample may not have appropriately represented the different species used, so therefore may not have been accurate. The research that was carried out was done in 1976, 30 years ago. This limits the investigation as there probably have been new developments in this area since then and new technology to further the accuracy of the research. The species that were chosen for this investigation were mostly mammals and could not have represented all the animals on earth appropriately. This means that it is harder for us to investigate correlation and causality between variables and make predictions and estimations from the data.

The only relevant comments are about the missing data points and new technology. The limitations aspect needs to refer to the model.

Assumptions

There were many assumptions made throughout this investigation. We do not know where on the Earth this research was done. Different parts of the world have different climates, seasons and weather which would definitely affect the sleeping patterns and life span of these animals. The investigation also assumes that the animals used in the research were of the same species. This affects the accuracy of the investigation as different species will have different sleeping patterns and different life spans. Once again, averages of the data will be inaccurate. Some animals may also be predators in one part of the world, so would have a longer life span and larger sleeping times (as they are not being killed as easily), but could be hunted in other parts of the world, either by other animals or even humans (this means shorter life span and smaller sleeping time as they are in danger). This investigation also does not take into account hibernation or migrations which are both factors that would change sleeping patterns. Hibernation means that they are in a permanent state of sleep so would affect the averages of the data. Migration also affects sleeping patterns as some animals will migrate for months at a time, sometimes not sleeping for days on end. This would also directly affect averages of the data and therefore making the investigation inaccurate.

These comments have synthesised knowledge from more than one subject to produce appropriate answers. **E**

Improving the regression

In the scatter graph showing the relationship between Brain weight and Sleeping time, the regression model could have been improved by dividing the graph into two. I would do this by dividing at 250 grams Brain weight, as before this there are a lot clumped plots at the beginning which would have a different trendline to the overall trendline of the entire graph. The plots shown after 250 grams Brain weight will also have a differing trendline and would take into account the outlier which is not accurately represented by the trendline in the first graph. Dividing up the graph and fitting two trendlines will definitely make the investigation more accurate and will deal with the outlier more appropriately. The scatter graph that represents the relationship between Brain weight and life span is fairly straightforward and I believe the regression does not need much improvement. However, there is still an outlier which affects the trendline so might have to be taken out or put into another graph with similar outlying values. Regression may have also been improved if I had chosen different variables to work with. I chose Brain weight as my predictor variable for both my graphs. My results and regression may have shown stronger relationships and correlation and causality if I had chosen to investigate the sleeping patterns more closely.

Piece-wise model proposed and justified. **E**

Relevance and usefulness of the Data

Since there was a large variety in the species from which the original data was obtained, the regression lines I found could be used to estimate or predict the sleeping time or maximum life span of other species of mammals from their mean brain weight. The results would only apply to adults of these other animals since the data used for the analysis related to adults only. Any predictions made would need to be for mean brain weight of adults of the species as mean brain weight of juveniles is likely to be significantly different from mean brain weight of adults. Also, the mean sleeping time of adults of a species may be significantly different from the mean sleeping times of juveniles (although life span is a characteristic of a species as a whole and not of adults or juveniles). There may be some limitations on the actual species to which the results could be extended if there are other significant factors that determine sleeping times that were not characteristic of the mammals in the dataset (for example, bats may spend more time each day sleeping than is typical of the species in the dataset). Since the original data related only to species of mammals, the results could only be extended to other species of mammals and not to other creatures such as birds.

An appropriate answer for the applicability of the findings aspect of Excellence (notwithstanding that the heading is for the relevance and usefulness of the evidence aspect). **E**