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

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	☐	Student 1 EXCELLENCE	☐
Assessment guidelines	☐	Student 2 MERIT	☐
Teaching notes	☐	**Student 3 ACHIEVED**	✓
		Student 4 ACHIEVED	☐
		Student 5 NOT ACHIEVED	☐

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

Date version published:	**July 2006**
Ministry of Education Quality assurance status	**For use in internal assessment from 2006**

EXAMPLE OF ASSESSED STUDENT WORK

ASSESSMENT COVER SHEET FOR STUDENT 3 ACHIEVED

	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 the 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

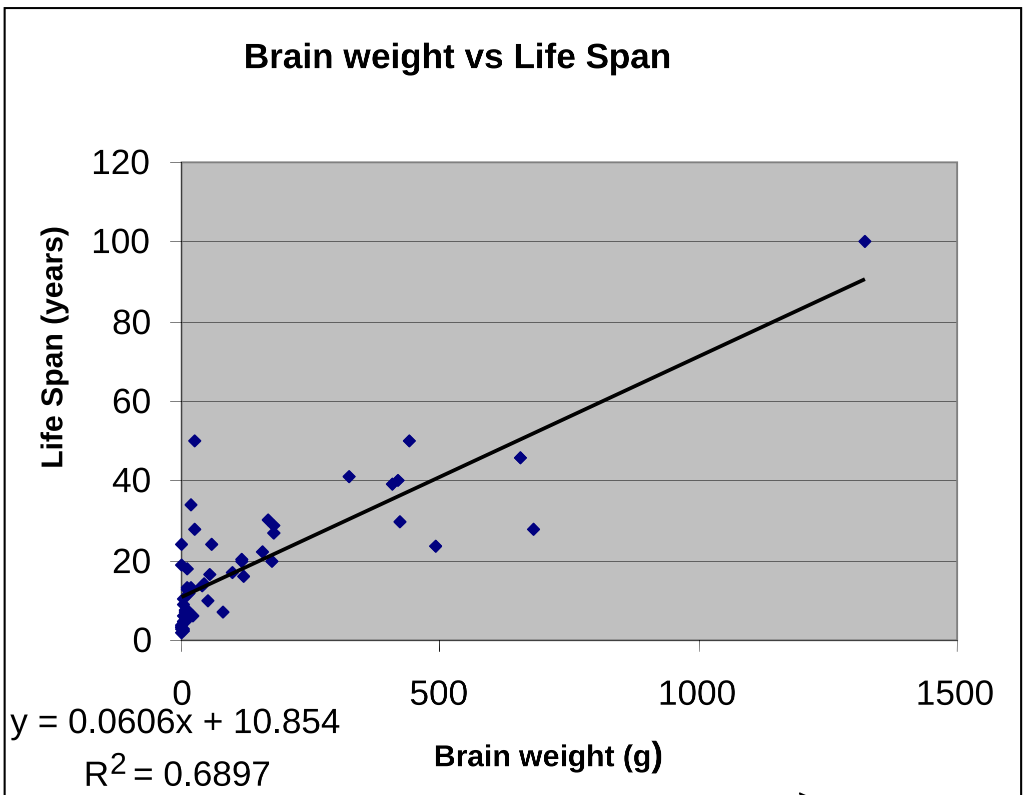
STUDENT 3 ACHIEVED

Bivariate data investigation

Data was provided for the time spent sleeping and the amount of sleep in dream sleep in animals. Data was also provided for the brain weight, total sleep, the lifespan and the gestation period of a variety of species. The data about predation, exposure and danger is discrete and cannot be used.
I am going to investigate the following:

Is there a relationship between the weight of an animal's brain (predictor variable) and the total life span of the animal (response variable)? I will compare this relationship with the relationship between the weight of an animal's brain and the gestation period.

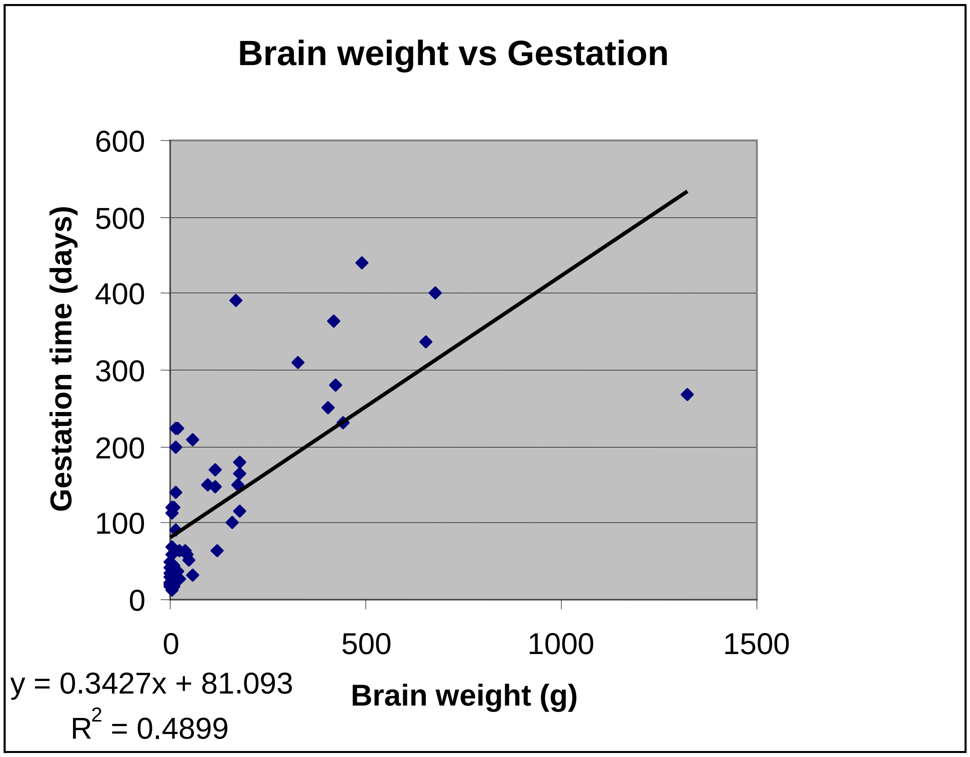
Purpose is stated. Aappropriate predictor variables are selected and stated. **A**



Scatter graph drawn and regression line found. **A**

This graph is a scatter graph of Brain weight v Life Span.
The line is the best fit straight line between the variables.
The equation is Y=0.0606X +10.854.
This means that an increase of 1 gram in brain weight results in an increase of 0.0606 years in a life span.

Regression line interpreted in context. **A**



It would be appropriate to have gestation as the predictor variable.

This is a scatter graph of brain weight v period of gestation.
The line is the best fit straight line relationship between the two variables.
The equation is Y=0.3427X + 81.093.
This means that an increase of 1 gram in brain weight results in an increase of 0.3427 days in gestation.

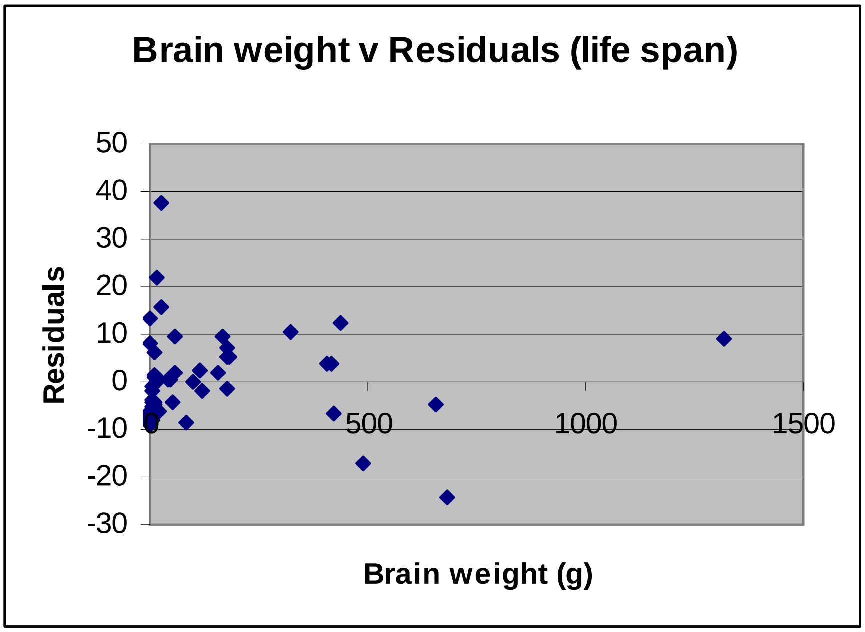
Comparing the relationships

Scatter Graph. The linear relationship between the brain weight and life span is stronger then that between brain size and gestation. This is because the majority of dots on the brain size v life span are clumped closer to the trend line.

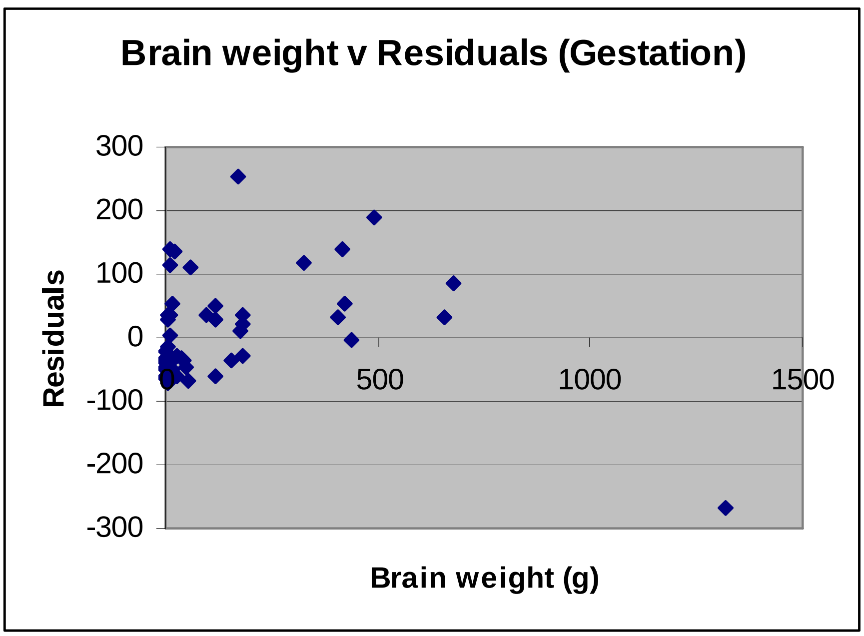
R^2 value. The R^2 value for brain weight v life span is 0.6897 and the R^2 value for the brain weight v gestation is 0.4899. This confirms that the linear relationship is stronger for the brain weight v the life span as they have a higher R^2 value.

A visual comparison of the relationships ("clumped" closer to the trend line" are key words) supported by R^2 . **M**

Residuals



For brains with a small weight the majority of the residuals are negative. There is a pattern to the residual graph and this suggests that the linear model is not appropriate.



The same pattern is observable in this graph and therefore the linear model is not appropriate.

The pattern must be explained.

Predictions

Brain weight and life span
The linear model predicts that an animal with a brain weight of 6.6 g will have a life span of 11.25 years.
The linear model predicts that an animal with a brain size of 680 g will have a life span of 52 years.

Brain weight and gestation
The linear model predicts that an animal with a brain weight of 5 g will have a gestation period of 82 days.
The linear model predicts that an animal with a brain weight of 1320 g will have a gestation period of 533 days.

> Predictions are made in context. **M**
> (This would be stronger if references were to species rather than animals.)

Interpolation and Extrapolation

These linear models of both brain weight v life span and brain weight v gestation are not appropriate for extrapolation because the majority of data plotted is under 500 g. This means that the line of best fit is not as relevant, as the line of best fit has derived from brains with a small weight. Therefore we are unable to estimate other values outside the range as the estimated values do not follow logically.

The linear model of brain weight v lifespan is appropriate for interpolation, as much of the data is plotted relatively close together and close to the line of best fit. This means that we are able to estimate a value between two known values. This graph has a stronger relationship as it has a higher R2 value and thus the line of best fit is more accurate.

The linear model of brain weight v gestation is not appropriate for interpolation as a section of the data is plotted randomly and therefore it is not accurate enough to be able to estimate any values.

> This contradicts previous statements and the meaning is not clear.

Limitation

The species selected did not include fish or birds. This limited the statistical investigation because we were not able to relate these results to those species. It would have been helpful to include these animals as there would have been more comparison and the results could have been used more widely.

We only had a few of the variables to work with. They did not include all the other relevant variables that could have affected brain weight, life span etc.

> Demonstrates a poor understanding of causality, which is not directly related to the purpose of the investigation.

Variables such as temperature/humidity, or the amount of resources available such as food or water etc. Because we didn't have these other variables we could come to the conclusion that one of the variables listed was the cause of another variable listed, when in actual fact it could be caused by a variable not listed. To make the data more accurate many more of the abiotic and biotic factors would need to be taken into account.

> The limitations need to refer to the model.

The data was collected in 1976 and therefore may not be as reliable, as some animals may have had slight evolutionary changes in the last 30 years.

> Earlier comment about fish and birds demonstrates a misunderstanding of the nature of the data set.

Some of the variable columns were not completely full as they had NA (not available) in them. This means that some of the animals were not included in the investigation. It also meant that the sample number decreased and was therefore not as accurate.

Both sets of variables had weak relationships and therefore we are limited to how accurate our results are and how widely we can apply them.

Improving the regression

In the linear model brain weight v gestation I would place another line of best fit where the group of dots lie between 250 g and 700 g. This is because the original line of best fit does not include this section of the data and therefore alter the reliability of the gradient equation.

> The first sentence is appropriate. The second sentence is confused and shows a lack of understanding.

Correlation and causality

Both linear models (brain weight v lifespan and brain weight v gestation) have a small correlation. They both have small R2 values which prove this. There is a minimal connection between both sets of variables.

> This is incorrect as r is quite high.

There is no causality between brain weight and life span. They are both independent of each other. This is shown on the scatter graph, as an animal whose brain weight is close to that of 10g and an animal whose brain weight is close to 475 g have roughly the same life span.

There is no causality between brain weight and gestation. They are also both independent of each other and have a very weak relationship. This is shown on the scatter graph, as some of the data is shown to have the same brain weight but their gestation period can be different by a couple of hundred days.

> No supporting evidence is offered for this statement.

The relevance and usefulness of the evidence

This evidence may be found useful in relation to biological evolution. This data may be compared with data recorded many years ago, and may be useful in the future as a comparison. This gives us a way of recording biological changes in numerous animals and comparing the different sets of data. Using this data we are able to see changes in a range of different animals, we are then able to categorize these animals in many ways, e.g. if we were to categorize the animals in their habitat, (tropical, snow, desert etc) from here we are able to see which animals in which habitat are experiencing the largest evolutionary changes. This can be done to a number of scenarios eg the animals could be categorized in size. This data is therefore useful to biologists etc.

This data is also very relevant because it shows us such things as evolution in mankind. The data shows us that although most of the larger animals have a larger brain weight in comparison to smaller animals, humans have the largest brain weight by far and yet there are many animals larger eg a horse. Mankind is considered the most intelligent mammal, and because of this known fact the data then suggests that those animals with heavier brains are more intelligent. It would be both useful and relevant to compare the weight of a human's brain now and in 50 years and so on.

These statements are too speculative, but could have been developed and justified to provide evidence for Excellence.