



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 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 checked="" 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 4

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 4 ACHIEVED

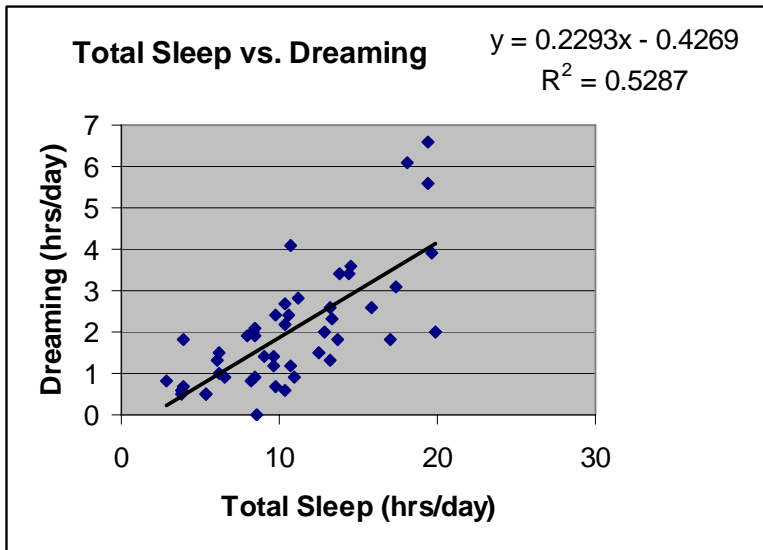
Bivariate Data Investigation

Data was provided about 62 animals. The variables I choose must be continuous therefore I can not use discrete data such as gestation, predation, exposure and danger.

I had to delete 15 rows (resulting in 48 rows of data) out of the original data because, in one or more of the columns I needed to produce results from, the data was not available or obtained for that species therefore irrelevant to my investigation.

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

The purpose of this investigation is to use regression to investigate the relationship between the total sleep of animals (predictor variable) and the time spent dreaming (response variable) and compare this relationship with the relationship between total sleep of animals (predictor variable) and the time spent non-dreaming (response variable).



Scatter graph drawn. **A**
Regression line obtained. **A**

This is a scatter graph of total sleep vs. dreaming.

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

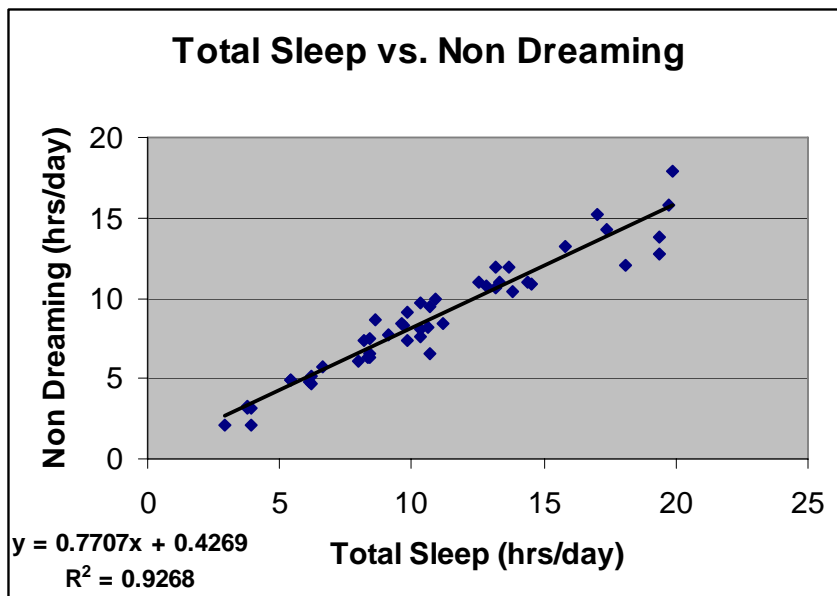
The equation is $y = 0.2293x - 0.4269$

This means that an increase of total sleep (in one hour per day) produces an increase of the dreaming rate by 0.2293hrs/day.

Relationship described in context. **A**

The straight line is appropriate as it shows the positive relationship, the gradient and how close the points are to the line of best fit to determine the strength of the correlation.

Dreaming and non dreaming sleep are related (sum total sleep) and this may limit the opportunity to demonstrate Excellence.



This is a scatter graph of total sleep vs. non-dreaming.
The line is the best fit straight line relationship between the two variables.
The equation is $y = 0.7707x + 0.4269$
This means that an increase of total sleep (in one hour per day) produces an increase of the non-dreaming rate by 0.7707hrs/day.
The straight line is appropriate as it shows the positive relationship, the gradient and how close the points are to the line of best fit to determine the strength of the correlation.

COMPARING THE RELATIONSHIPS

The gradient of these scatter graphs are increasing therefore they have positive relationships.

Scatter graphs

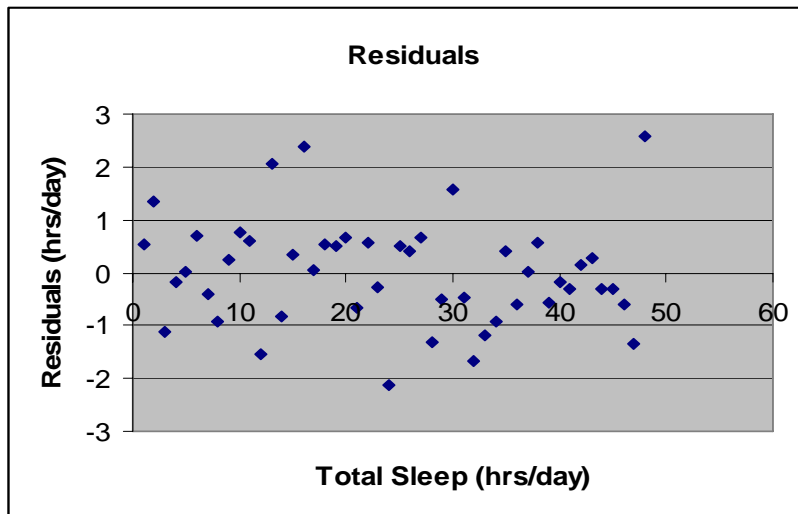
The relationship between total sleep and non-dreaming is stronger than that between total sleep and dreaming because the points on the scatter graph are closer to the line of best fit.

R²

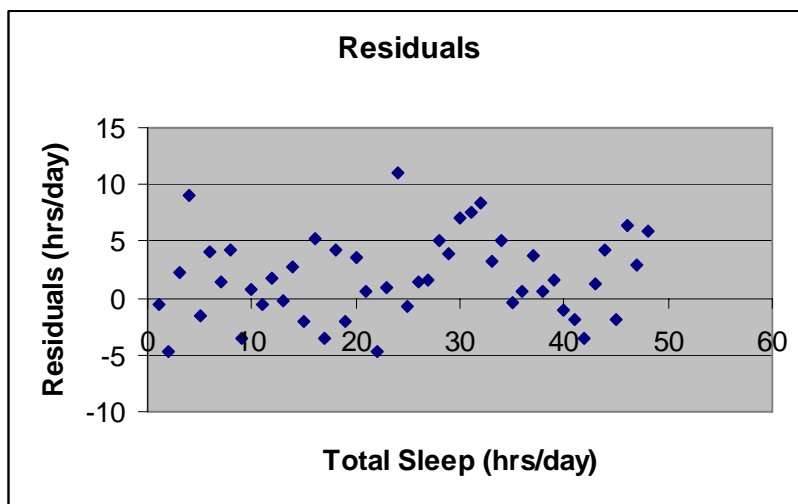
Also the R² value for between total sleep and non-dreaming is 0.9268 (this is a very strong linear relationship) and for between total sleep and dreaming, the R² value is 0.5287 (this is only a moderate relationship). This confirms the relationship is stronger.

Relationships compared
(visual comparison supported
by R²). **M**

RESIDUALS



This is a graph showing the residuals of the total sleep vs. dreaming. There are no patterns so the residual graph suggests the linear model is appropriate.



There are errors in this graph. The total sleep per day cannot be more than 24, and the graph does not match the corresponding scatter graph. The values of the residuals are wrong as well. They are not as one would expect from the scatter graph.

This is a graph showing the residuals of the total sleep vs. non-dreaming. There are no patterns (apart from around the 30hrs per day mark where all the points are above the line) so the residual graph suggests the linear model is appropriate.

PREDICTIONS

I calculated my predictions using Microsoft Excel and used the formula modelled as $y=mx+c$ and substituted the value of total sleep into x .

Total sleep and dreaming

The linear model predicts that when a Cow sleeps for of 3.9 hours its average dreaming time would be 0.46737hrs/day.

The linear model predicts that when a Little Brown Bat sleeps for of 19.9 hours its average dreaming time would be 4.13617hrs/day.

Total sleep and non-dreaming

The linear model predicts that when an Echidna sleeps for of 8.6 hours its average non-dreaming time would be 7.05492hrs/day.

The linear model predicts that when a Water Opossum sleeps for of 19.9 hours its average non-dreaming time would be 15.37848hrs/day.

There is one error: the value of 15.378 should be 15.763. This one error in 4 calculations should not preclude this part contributing to Merit. In general these predictions are given to too many s.f. **M**

CAUSALITY OR CORRELATION

The relationship between total sleep and non-dreaming is causality as there is a definite relationship which is very strong. Whereas total sleep versus dreaming is mostly a correlation because the relationship is not as strong and the points are more grouped than in a straight line and are more of a coincidence that they are moderately close to the line of best fit.

Causality is clearly misunderstood here.

DISCUSSION

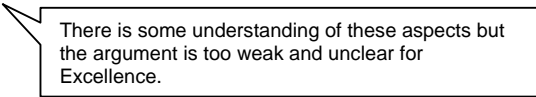
The limitations of my analysis are that there might actually be a causality relationship, between total sleep and dreaming with animals, and there might also be a scientific reason behind it, but my results showed more of a correlation so therefore my conclusion was that the relationship was more of a coincidence than a certain relationship. Also that I had less information to work with as I had to delete some rows of data as some were not available and that also contributes to my results being less accurate.

Muddled thinking about causality and correlation.

The relevance and usefulness of the evidence is quite limited as all the species of animals in the world were not included and again, some were deleted as the data was not provided. Therefore these relationships do not apply to all animal species, even if the data did have all the relevant and up to date data, these results could not be applied or used to predict dreaming and non-dreaming patterns (of a singular species) as all species would have their own relationships to do with what kind of activities they do, how much they eat etc. This would contribute to determine how long they sleep per day and therefore predicting their dreaming and non-dreaming time. For example, take the Cat species and produce a set of results, similar to these, to create assumptions that are more accurate and useful. It would also depend on the country in which the data was collected as animals in different countries may vary in size and health because of the food availability to them and also they may sleep more in warmer climate etc.

This paragraph starts well but becomes muddled and unclear. Not sufficient for Excellence.

Consequently, my findings can not be applied (let alone widely) as they are not accurate and do not prove much – apart from finding that there is a very strong relationship between total sleep time and non-dreaming – mostly because non-dreaming time takes up most of the total sleeping time so therefore are similar amounts of time and giving a near-direct association. Also the animals researched are all mammals, leaving out reptiles, amphibians etc. So you can not base my findings on all animals. This data is also out of date therefore the results and technology used to produce this data would be as well, as it was collected in 1976 (30 years ago), it would definitely be updated by now so there will be more relevant and useful findings that can be more widely applied.



There is some understanding of these aspects but the argument is too weak and unclear for Excellence.