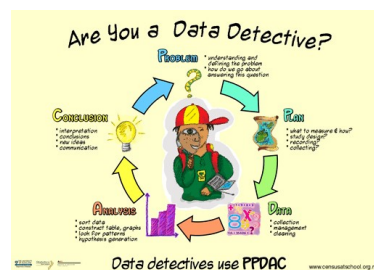


Overview of Bivariate Measurement Data Report Writing

Key components of the statistical enquiry cycle for investigating bivariate measurement data:

- posing an appropriate relationship question using a given multivariate data set
- selecting and using appropriate displays
- identifying features in data
- finding an appropriate model
- describing the nature and strength of the relationship and relating this to the context
- using the model to make a prediction
- communicating findings in a conclusion.



Aiming for Excellence

Achievement	Merit	Excellence
Investigate bivariate measurement data involves showing evidence of using each component of the statistical enquiry cycle.	Investigate bivariate measurement data, <i>with justification</i> involves linking components of the statistical enquiry cycle to the context, and referring to evidence such as statistics, data values, trends, or features of visual displays in support of statements made.	Investigate bivariate measurement data, <i>with statistical insight</i> involves integrating statistical and contextual knowledge throughout the statistical enquiry cycle, and may include reflecting about the process; considering other relevant variables; evaluating the adequacy of any models; or showing a deeper understanding of models.

Bivariate Data	☹	☺	☺
<ul style="list-style-type: none"> • Bivariate data compares two variables that are potentially connected e.g. ice cream sales and temperature on that day 			
<ul style="list-style-type: none"> • In this assessment you will be given some raw data that you will be required to analyse by drawing a scatter plot (using iNZight) and writing a report. 			

Overview of Time Series Report (Use headings 1-6 to organise your report)	☹	☺	☺
I notice / I wonder (do not include these notes in your final report) – use Scatter Plot Matrix in iNZight			
1. Introduction / Background			
2. Identify features in the data (association, features, trend)			
3. Select and justify an appropriate model (linear / non-linear)			
4. Make a prediction in context with units and sensible rounding.			
5. Examine other Models			
6. Conclusion			

1. Writing Introductions for Bivariate Measurement Data	Exemplar statements <i>Italicised statements give alternative statements for describing the data</i>	☹	☺	☺
Description and Investigative Question <ul style="list-style-type: none"> • Description of topic / variables from topic (one sentence). • Relationship Question (one only). • Aim / Interest (Why worth investigating? Questions?) 	<ul style="list-style-type: none"> • Measurements such as height, weight and lean body mass are useful ways of comparing athletes' health and performance. • Variables that might be likely to have a relationship are... because... • This report will investigate if there a relationship between Haematocrit levels and Red Blood Cell Count for athletes from the Australian Institute of Sport. • <i>This report will investigate the nature of the relationship between Haematocrit levels and Red Blood Cell Count for athletes from the AIS.</i> • <i>This report will investigate if an athlete's Haematocrit levels can be used to predict their Red Blood Cell Count.</i> • An understanding of the relationship between these variables might be useful to... because... 			
Data / Survey <ul style="list-style-type: none"> • Source • Definition and description of explanatory and response variables. • Important aspects of data collection details / validity. 	<ul style="list-style-type: none"> • The source of this data is 120 athletes from the Australian Institute of Sport. • The explanatory variable being investigated is... • The response variable being investigated is... • This data is likely to be valid as it is collected by doctors at the AIS. • <i>This data was collected by xxx and therefore may not be a valid measure of ...</i> 			
Research <ul style="list-style-type: none"> • Link findings to what you know / have researched about the variables. • Research suggests these two variables may be related because... 	<ul style="list-style-type: none"> • In this context it seems likely that there would be a relationship between these two variables as the Haematocrit level measures the percentage of red blood cells and the red blood cell gives a count of these same cells. • One important aspect of this data is that all of the data points are for athletes and so any relationships may not be applicable to non-athletes. 			

I understand...	I need to work on...
-----------------	----------------------

2. Identify features		Exemplar statements	☹	☺	☺
Graph	<p><i>Inzight Graph: Scatter Graph (check explanatory and response variables)</i></p> <ul style="list-style-type: none"> Re-label graph axes if needed with full title and units. 	<p>Scatter Plot showing relationship between haematocrit level and Red blood Cell Count for 120 Athletes at AIS.</p>			
Association	<ul style="list-style-type: none"> What is nature of relationship? Inc / Inc or Inc / Dec Justify by reference to visual aspects. 	<ul style="list-style-type: none"> The scatter plot shows that as the Haematocrit level increases the Red Blood Cell Count also increases. <i>The scatter plot shows that as x increases y decreases.</i> 			
Context	<ul style="list-style-type: none"> This is to be expected because... (wider population – not just this group). Name other variables that might impact on the response variable and suggest how they might impact. e.g. gender age. 	<ul style="list-style-type: none"> This is to be expected as it is likely that the percentage of red blood cells may well impact on the number of such blood cells. Other factors that may affect a person's red blood count are... because... 			
Trend	<ul style="list-style-type: none"> Linear / non-linear? 	<ul style="list-style-type: none"> From the scatter plot it appears that there is a linear relationship between Haematocrit levels and Red Blood Cell Count. <i>From the scatter plot it appears that there is a non-linear relationship between x and y.</i> 			

I understand...	I need to work on...
-----------------	----------------------

3. Find a model		Exemplar statements	☹	☺	☺
Graph	<p><i>Inzight Graph: Add to plot, select linear / non-linear trend</i></p>	<p>Scatter Plot showing relationship between haematocrit level and Red blood Cell Count for 120 Athletes at AIS.</p>			
R	Reason for linear / non-linear model	<ul style="list-style-type: none"> For the reasons given above a linear regression model has been fitted to the data. <i>For the reasons given above a non-linear regression model has been fitted to the data.</i> 			
D	Description of model <ul style="list-style-type: none"> Gradient statement if linear 	<ul style="list-style-type: none"> The linear model shows that Red Blood Cell Counts increase by 0.1 for each increase of 1 in Haematocrit value. 			
Appropriateness of model	<ul style="list-style-type: none"> Discussion of fit throughout the range of x values. Look at how well the points align with the trend line for the range of x values. 	<ul style="list-style-type: none"> This model appears to be a good fit of the data throughout the range of Haematocrit levels with all points aligning with the linear trend. The number of points above the trend line is also similar to the number of points below. However, there are no athletes with Haematocrit levels from 53 to 59 and so we are unable to describe the fit for this data range. This means the model may not be as appropriate for assessing the relationship between these variables when the Haematocrit levels are over 52. 			
	<ul style="list-style-type: none"> Consider number of data points 	<ul style="list-style-type: none"> This is a relatively high number of data points (120) which enhances the reliability of the model. <i>The relatively low number of data points means this model may not be particularly reliable.</i> 			

	<ul style="list-style-type: none"> Correlation / causation 	<ul style="list-style-type: none"> This relationship is only statistical and does not imply that an increase in Haematocrit level <i>causes</i> an increase in Red Blood Cell Count. 			
Strength of relationship	<ul style="list-style-type: none"> Look at scatter Strong / moderate-to-strong / moderate / weak-to-moderate / weak Look at amount of scatter about the regression line If linear then r / correlation coefficient variation in scatter – constant / non-constant / fanning out 	<ul style="list-style-type: none"> This relationship appears to be moderate-to-strong as there is some scatter along the trend line but it is not a large amount. The correlation coefficient is also relatively high at 0.93 indicating there is evidence of a fairly strong linear relationship between Haematocrit values and Red Blood Cell Counts. <i>The scatter along the trend line is non-constant, with more scatter after x. This suggests a stronger relationship for x = and a potentially weaker relationship for x =.</i> <i>There is an increase in scatter after x = . This suggests the relationship may not be as strong after this point.</i> 			
Unusual values	<ul style="list-style-type: none"> Visual description, numerical description, discuss possible effect on model. 	<ul style="list-style-type: none"> One unusual value is present with a Haematocrit level of 60 and a Red Blood Cell Count over 6.5. This value is along the same trend line as the rest of the data and so may inappropriately increase the strength of the relationship. 			
Groupings / Clusters	<ul style="list-style-type: none"> Visual description, numerical description, discuss possible reasons for differences. 	<ul style="list-style-type: none"> No groupings are apparent from the scatter plot. <i>Two groups are suggested in the scatter plot – the first with x < ... and the second with x > ... One possible reason for these differences may be...</i> 			

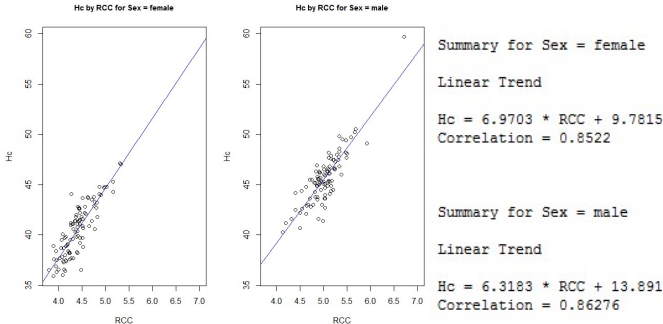
I understand...	I need to work on...
-----------------	----------------------

4. Make a prediction		Exemplar statements	☹	☺	😊
	<i>Inzight Graph: as above</i>				
Prediction	<ul style="list-style-type: none"> Make a prediction for the response variable using the equation of the trend line. Round answer sensibly and include units if appropriate. Don't relate to observed y-values. 	<ul style="list-style-type: none"> From this model I predict that the red blood count of a person with a Haematocrit level of 50 will be 5.5. (= 0.11565 * 50 – 0.26) 			
Justification	<ul style="list-style-type: none"> Justification regarding how accurate prediction might be – reference to stat evidence from analysis. Reflect on prediction by discussing their relevance to wider population. Justify choice of variables to use by giving reasons for using the selected one rather than others. 	<ul style="list-style-type: none"> Given the moderate-to-strong relationship found in the data it is likely that this prediction will be quite accurate. This prediction is likely to only be accurate for athletes as it is likely that they will have higher general Haematocrit and red blood cell levels than the rest of the population as they exercise more ¹ Haematocrit is likely to be the best explanatory variable because... <i>Given the weak relationship found in the data this prediction is unlikely to be particularly accurate and should only be taken as a rough indication of y at point x.</i> 			

I understand...	I need to work on...
-----------------	----------------------

5. Further Considerations		Exemplar statements	☹	☺	😊
Graph					

¹ <http://www.livestrong.com/article/299082-the-effect-of-athletic-training-on-the-rbc-count/>

<p>If unusual values:</p>	<ul style="list-style-type: none"> • Comment on the effect any unusual values might have on the model. • Justify why these values could be removed. • Extend the investigation by developing models with data with and without the unusual values. 	<ul style="list-style-type: none"> • The data point at Haematocrit value 59 does not appear to fit in with the rest of the data. This could be a valid extreme value, or it also could be an error in measurement. • For these reasons the model will be tested with and without this data point to see the effect, if any, on the prediction made. • As can be seen in the graph above... 	
<p>Graph</p>	 <p>Summary for Sex = female Linear Trend $Hc = 6.9703 * RCC + 9.7815$ Correlation = 0.8522</p> <p>Summary for Sex = male Linear Trend $Hc = 6.3183 * RCC + 13.891$ Correlation = 0.86276</p>		
<p>If subsets / groups:</p>	<ul style="list-style-type: none"> • Comment on the effect the difference subsets might have on the model. • Comment on the number of points now being investigated. • Extend the investigation by developing models with data that has been separated into relevant subsets. 	<ul style="list-style-type: none"> • One factor that may influence the relationship between Haematocrit levels and Red Blood count is the gender of the athlete. • For these reasons the data will be split into these two groups and reanalysed. • As can be seen in the graph above... 	
<p>Re-predict</p>	<ul style="list-style-type: none"> • Prediction made using alternative models • Consider accuracy of these alternative models • Compare with original prediction 	<ul style="list-style-type: none"> • Compare and contrast original prediction with updated (unusual value OR subset /group). 	

<p>I understand...</p>	<p>I need to work on...</p>
-------------------------------	------------------------------------

<p>6. Conclusion</p>		<p>Exemplar statements</p>	<p>☹ ☺ ☺</p>
<p>Summary</p>	<ul style="list-style-type: none"> • Give concise summary linked to original purpose of the investigation • Purpose of report • Brief description of model, including trend, strength and numeric. 	<ul style="list-style-type: none"> • This report investigated whether a relationship exists between the Haematocrit levels and red blood cells for 120 athletes from the AIS. • Analysis of the data showed a moderate-to-strong linear relationship between Haematocrit levels and red blood cells. • This relationship showed that Red Blood Cell Counts increase by 0.1 for each increase by 1 in Haematocrit value. 	
<p>Prediction</p>	<ul style="list-style-type: none"> • What the model predicted • Link to context • Accuracy of prediction 	<ul style="list-style-type: none"> • This model was then used to predict the Red Blood Cell Count for an athlete with a Haematocrit level of 50. The Red Blood Count was predicted to be... This means... • This prediction is likely to be accurate because... 	
<p>Extension</p>	<ul style="list-style-type: none"> • Summary of investigation into other relevant variable • Summary of what this means in context / research / future investigations • Usefulness / Limitations / Improvements / Possible uses / Future Investigations 	<ul style="list-style-type: none"> • A further variable that was investigated was... • Possible limitations of this model include... • These findings may be useful because... 	

<p>I understand...</p>	<p>I need to work on...</p>
-------------------------------	------------------------------------