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

Exemplar for Internal Assessment Resource

Mathematics Level 3

Resource title: Sport Science

This exemplar supports assessment against:

Achievement Standard 91581

Investigate bivariate measurement data

Student and grade boundary specific exemplar The material has been gathered from student material specific to an A or B assessment resource.

Date version published by Ministry of Education — December 2012
To support internal assessment from 2013

	Grade Boundary: Low Excellence
1.	For Excellence the student is required to investigate bivariate measurement data, with statistical insight. This involves integrating statistical and contextual knowledge throughout the investigation process, and may include reflecting about the process; considering other relevant variables; evaluating the adequacy of any models, or showing a deeper understanding of the models. An appropriate relationship question has been implied. The choice of different pairs of variables should have been considered (1). Appropriate displays have been selected and used (2). Features of the data have been identified, and the nature and strength of the relationship with regard to visual aspects of the scatter about the regression line have been described in context. Comments indicate that the context of high performance sport has been considered (3). An appropriate model has been found and justified. The model has been used to make predictions. The appropriateness of a linear model has been considered and there are some contextual comments about the predictions (4). An additional model has been considered. The appropriateness of the new model and accuracy of the predictions using this model have been compared with the linear model (5). Findings have been communicated in a conclusion and comments demonstrate an understanding of the context. Issues related to specific sports and the age of the competitors could have been considered (6). There is evidence of investigating bivariate measurement data with statistical insight in the integration of contextual knowledge. For a more secure Excellence the student could have explored the relationship for male athletes, or could have considered investigating further particular sports.

Body image is a concern of modern women and just as much for female athletes as it is for other 'normal' women. http://thesportjournal.org/article/body-image-disturbances-ncaa-division-i-and-iii-female-athletes

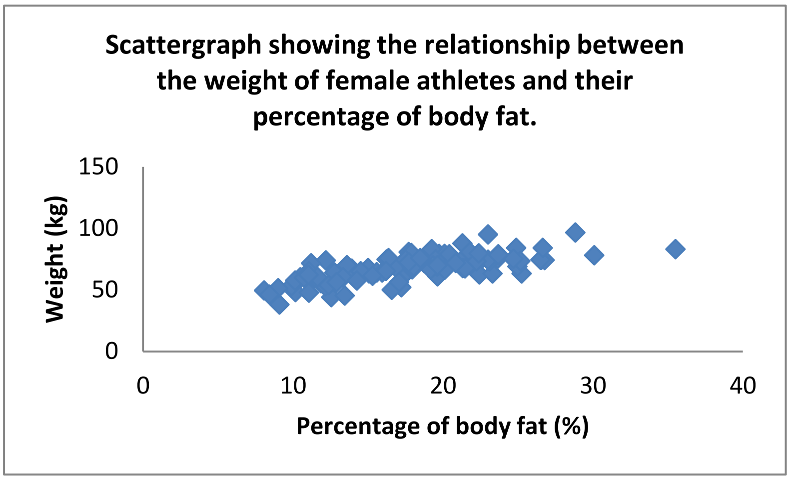
Weight charts can be unreliable as a source for determining the healthy weight for athletes and body fat percentages are sometimes used instead. Women tend to weigh in the top range for their heights due to muscle mass being heavier than fat mass.

During the 2012 London Olympics the media reportedly criticised some female athletes suggesting that they were fat rather than fit. A list of top female athletes hit back at critics in an article 'Fat? We are fit. Get over it' by Belinda Goldsmith http://sports.yahoo.com/news/fat-fit-over-women-athletes-193539328--spt.html

I am going to investigate if there is a relationship between the weight of a female athlete and their percentage of body fat. I have taken the percentage of body fat as the explanatory variable and the weight as the response variable. (1)

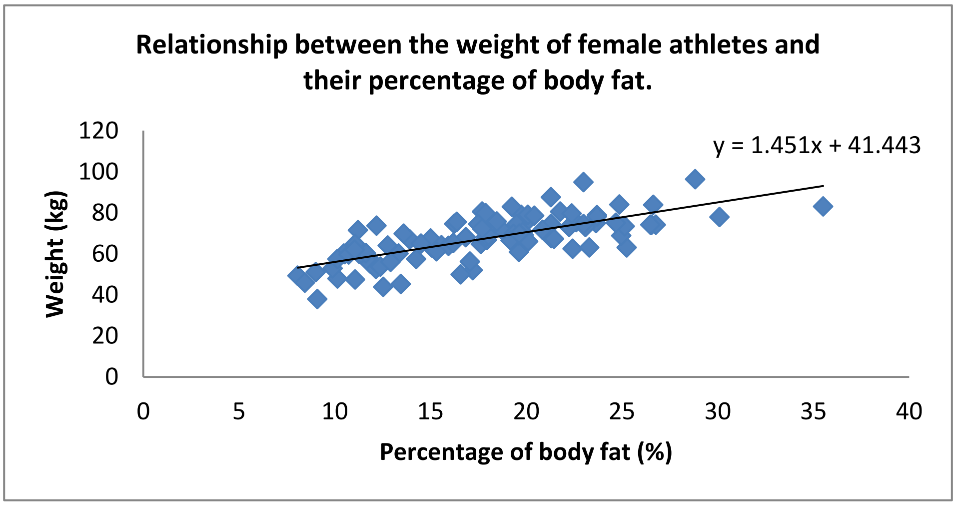
Initial Scatter graph.

This scatter graph shows us that there is a positive relationship between the weight of female athletes and their body fat percentage, i.e. people with a higher percentage of body fat tend to be heavier. This is consistent with what we would expect. It also appears to have a linear relationship and there is nothing to suggest that a different model would be better fit to the data.



The American Council on Exercise (ACE) divides body fat percentage into five different categories: essential, athletes, fitness, acceptable and obese. Essential body fat ranges from 10 to 14%, athletes 14 to 20% and fitness 21 to 24%. The acceptable range of body fat for women is 25 – 31% and a woman with a body fat percentage of over 32% is considered obese. The percentage of body fat in my scatter graph is mostly between 10 and 26% so this sort of fits what the ACE are saying. There is not much data with a percentage of body fat over 30% but I don't think this means that those in this range won't be athletes more that most athletes in most sports will be under this. There is one value of 35% body fat and looking at the original data this value is for a Netballer. I have looked at the other athletes in the data from the Australian Institute of Sport that have a body fat percentage of more than 25% and 7 of them are netballers and there is one basketball player – this suggests that maybe netballers will be in the higher range of percentage body fat.

Linear model



(2)

We see by the trend line that this is indeed a positive relationship - as the percentage of body fat of female athletes increases their weight in kg tends to increase. The gradient of the trend line shows us that for every 5% that your body fat increases you can expect to add approximately 8kg to your total weight. Most of the data points are close to the regression line indicating that the relationship between percentage body fat and weight of females athletes is strong. (3)

The y-intercept of a weight of around 41kg seems unrealistic as it is impossible to live with 0% body fat as it insulates the body and is your body's energy source. The influence of the y intercept (a body fat % of zero) on the model needs to be considered as while it is possible to have a very low percentage of body fat the essential body fat percentage quoted by ACE is between 10 and 13%. The model that I have found fits the data well and I don't think the body fat percentages in the lower ranges are of interest in my investigation so I will not change the y intercept and look for a new linear model. (It is not possible to have a body fat percentage of 0 and so this part of the model is of little use to us). (3)

Prediction

Using my linear model, which fits the data reasonably well, I can make predictions on how much a female athlete will weigh (kg) given her percentage of body fat, because of the fit I expect my predictions will be fairly accurate. For example:

If female athlete X had a body fat percentage of 20% we would expect her to weigh

$y = 1.451*20+41.443 = 70.463 = 70.5$kg or

if female athlete X had a body fat percentage of 30% we would expect her to weigh

$y = 1.451*30+41.443 = 84.973 = 85.0$kg. (4)

When I look at my model I can see that some data points are above and some below (weights) near the body fat percentage of 20% and a weight of 70.5kg is somewhere in the middle of the values, so the model appears reliable. The other value that I chose to predict the weight for is a percentage body fat where there weren't many data points (30%). The prediction of 85kg seems reasonable but it is closer to the weights under the linear model rather than those above for body fat percentages near 30%. (4)

I noticed that the some of the athletes with higher percentages of body fat were netballers which suggest that the type of sport that an athlete participates in could change the weight and body fat percentage relationship. For example athletes who complete in sports such as shot put could have a higher weight due to a higher percentage of muscle mass but not necessarily have a higher percentage of body fat. In the data set provided all the track and field athletes are put together as 'field' so it is not possible to verify this.

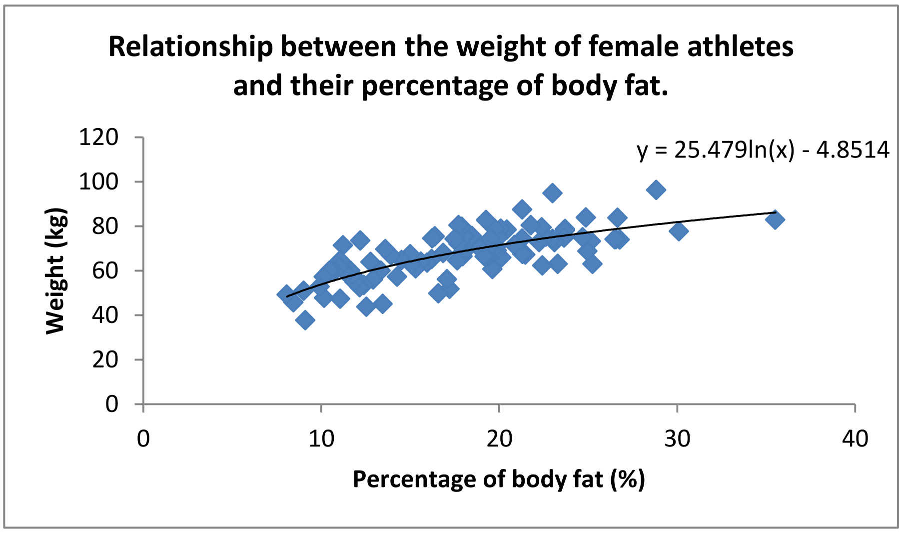
If we also consider the differences that athletes such as swimmers who would have a lower percentage of muscle mass compared to the shot putters we might observe a difference in the relationship between the weight and body fat percentage and the sport that the female athlete competes in. (http://www.livestrong.com/article/510269-what-is-the-normal-body-fat-of-an-athlete/)

The amount of body fat seems to depend on the type of sport in the data supplied from the Australian Institute of Sport for instance shot putters may have more but other sports that involve running and jumping etc maybe benefit from less. Coaches and trainers have the task of balancing a programme of exercise, training and diet to enable the athletes to perform at their best in their chosen sport and this is likely to involve making decisions about the ideal weight or body fat percentage.

Improvements to model

I noticed that the data points beyond a body fat percentage of approximately 26 % seem further away from the linear model and do not appear to be a good fit to the linear model. Because of this we might consider another model, looking at the scatter plot it looks like a logarithmic model might be better.

In the logarithmic model we see visually that the trend line seems to fit the data better.



I would consider the relationship moderately strong to strong. The data seems to be evenly scattered close to the log model from the smaller values up to the largest body fat % of 35%. There is a positive relationship - as the percentage of body fat increases the weight (kg) tends to increase and unlike the previous linear model this log model levels off meaning that as the percentage body fat reaches the higher levels the weight tends to increase at a slower rate. While this might not be the case for 'the average women' it seems reasonable in the case of athletes. (5)

Using this model to make a prediction for the same body fat percentage that I found before:
For a female athlete with a body fat percentage of 20% we would expect her to weigh
y= 25.478*ln(20)-4.8514 = 71.48 = 71.5kg or for a female athlete with a body fat percentage of 30% we would expect her to weigh
y= 25.478*ln(30)-4.8514 = 81.81 = 81.8kg. (5)

Comparing with these with my earlier predictions using the linear model we see that for 20% body fat that this new model provides a similar prediction of the weight of the female athlete but looking at the graph the data points around 30% body fat these are closer to this logarithmic model so I think that this prediction is probably a more reliable prediction than the linear one used earlier. (5)

Conclusion
The relationship between the weight of a female athlete in kg and the percentage of body fat they have is a strong to moderately strong, positive logarithmic relationship.

The lower data values between 10 and 20% of body fat fit both the linear and logarithmic equally well but for the upper values greater than 26% of body fat the logarithmic model is a better fit.

I believe that there are possibly other factors that should be considered. I could have investigated what the effect of the sport the female athlete was completing in was on the relationship between the weight (kg) and the body fat percentage. This would likely be useful for the coaches and trainers of the athletes in the different sports. Knowing the ideal weight and associated body fat % to enable top performance could be incorporated into the planning of training and diet plans for athletes. This is likely to be different in different sports.

I expect that many athletes typically have a lower body fat than the average person because of the training that they do and the sports that they take part in. Although there is no perfect shape for an athlete, low body fat can help improve performance for sports such as gymnastics, basketball, triathlons and running.

Whether the relationship investigated here for female athletes is the same for the average female has not been investigated. For instance the acceptable norm range of % body fat for women is between 21 – 24%. This would be a further investigation ie looking at the relationship of weight and the percentage of body fat in the average female to see how that differs to the relationship in a female athlete. (6)

	Grade Boundary: High Merit
2.	For Merit the student is required to investigate bivariate measurement data, with justification. This involves linking components of the statistical enquiry to the context, and referring to evidence such as statistics, data values, trends, or features of visual displays in support of statements made. An appropriate relationship question has been posed. There is some contextual reflection on the nature of the data. The choice of a different pair of variables has been considered (1). Appropriate displays have been selected and used (2). Features of the data have been identified and the nature and strength of the relationship have been described in context. There is some contextual discussion on features of the relationship (3). An appropriate model has been found and justified. The model has been used to make predictions. There is contextual discussion about the predictions (4). An additional model has been considered with discussion about suitability (5). The student has acknowledged that they appear to have found a linear relationship between the variables, but that this does not imply causation (6). Findings have been communicated, but there is no clear link to a purpose for the investigation. There has been some contextual consideration with provision of some background information (7). There is evidence of investigating bivariate measurement data with justification in the provision of contextual evidence to support statements for components of the statistical enquiry cycle. To reach Excellence all aspects of the investigation would need to be supported with reflective comments which demonstrate a deeper contextual understanding, supporting a clearer purpose of the investigation.

Student 2: High Merit

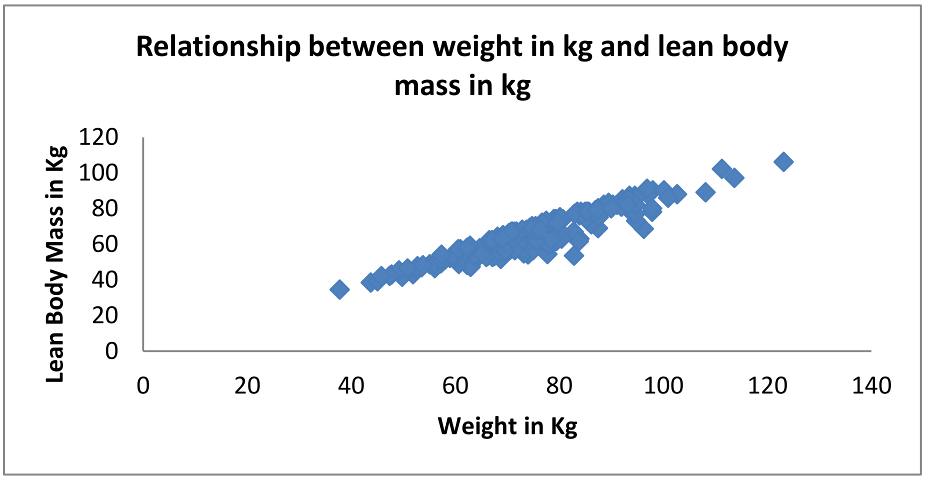
Using the data set provided from the Australian Institute of Sport I will look at different pairs of variables to quickly see the kinds of relationships that there might be. I have decided to investigate the relationship between weight and lean body mass.

Investigative question:

I wonder if there is a relationship between weight in kg and lean body mass in kg in males and females that play sports?

According to the Journal of Romanian Sports Medicine Society[1] body composition is a factor contributing to sport performance and the assessment of body composition is an important component of the on-going monitoring of athletes interested in improving their performance.

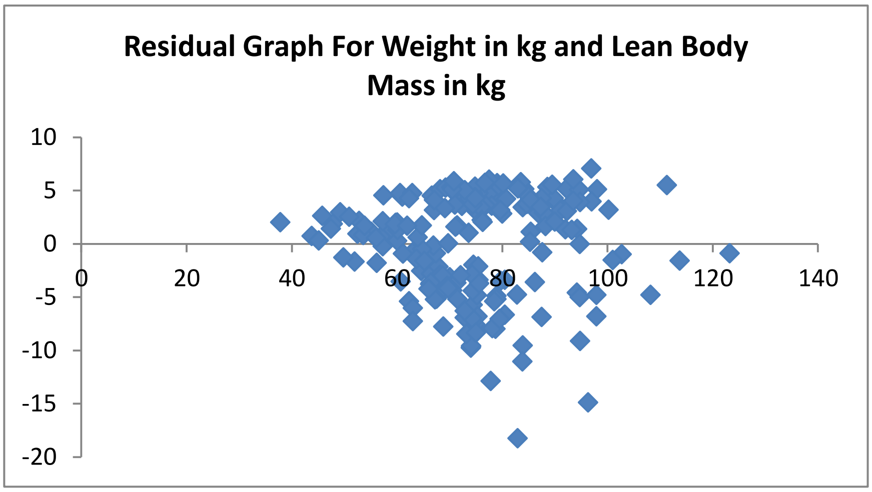
Lean body mass (LBM) is how much you weigh without your body fat so I expect that there will be some relationship. Without the body fat you should be able to get an idea of how much muscle you are gaining or losing as a result of training and diet. LBM is fairly easy to calculate once you have weighed yourself and figured out your body fat percentage. You just calculate your body fat in kg and subtract that from your body weight. Weight in kg is the explanatory or independent variable and LBM in kg is the response or dependent variable. (1)



From the graph it seems that there is a positive linear relationship between the weight and LBM of an athlete and there is no other evidence to suggest that a different model would be better. This is seen by how the data points on the graph are very close together and are increasing in a positive direction and are close to an imaginary linear trend line. This suggests that there is a strong relationship between the two variables. The relationship is also seen as positive due to the fact that the data seems to go upwards and increases. It seems that when weight in kg increases the lean body mass tends to increase for males and females that play a sport. There seems to be outliers in the data. There are some values that are not so close to the main group of data these can be found in the range between the weights of 111kg to 124kg and the weight 38kg.

(2)

When a trend line has been added to the graph it is visible from the positive gradient that there is a positive relationship and the points are scattered along the regression line quite closely, therefore there is a strong relationship between weight and LBM of athletes. The correlation coefficient (r) demonstrates the strength of the relationship between the two variables. This has a value of 0.93090405 and confirms that there is a very strong, positive relationship because the value is very close to 1. The correlation coefficient expresses how close the points are to the trend line and in this case they are very close. The gradient of the linear trend line 0.8737 expresses that as the weight in kg for males and females that play a sport increase by 1kg, the lean body mass in kg increases by 0.8737kg. (3)



From this graph of residuals for weight in kg and lean body mass in kg it is visible that there is quite an even spread of the points. About half of the points are positive and the other half is negative. This reinforces that a linear model is appropriate.

Predictions

I am going to use my linear model for weight in kg and lean body mass in kg to make predictions. The correlation coefficient for model is 0.93090405 and it is likely that my predictions will be fairly reliable. I used the linear equation, which is y = 0.8737x - 0.6627 and substituted my x value, which is the weight for an athlete. I will make several predictions.

For a weight of 40kg, I predict a LBM of 0.8737*40 - 0.6627 = 34.3kg, for a weight of 63kg I predict an LBM of 0.8737*54.3- 0.6627 = 54.3kg

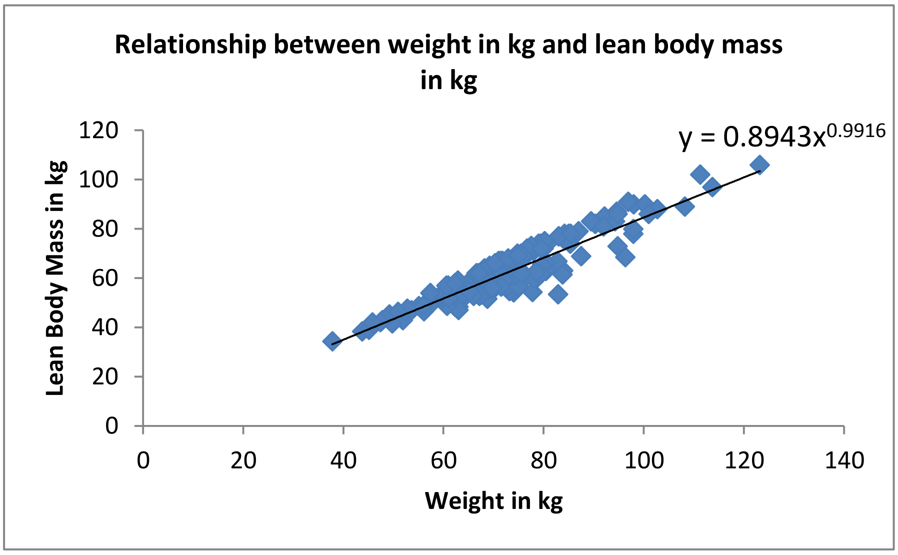
I think that the linear model that I fitted to the data is a good fit as the data points are all close to it, and this is true for this for all the values so I think that these predictions will be reasonably accurate. There are not many data points around the weight of 40kg in my scatter graph but as I said the points are very close to the line fitted so I believe that this prediction will still be ok. (4)

For a weight of 110kg I predict an LBM of 0.8737*110 - 0.6627 = 95.4kg, even though there are fewer data points for weights beyond 100kg they are very close to the regression line so I think that the model will still give accurate predictions of LBM for athletes for these heavier weights. (4)

It seems generally the heavier the lean body mass the heavier the weight of a female or male that plays a sport.

Other Models – Even though there is no immediate evidence for it I will investigate whether a power model would represent the relationship between weight in kg and LBM in kg.

Power Model



Visually the data does not seem to fit this model any better than the linear model - the data is still scattered in a similar way in this polynomial model. I don't think that predictions made with this model are likely to be any more reliable that those given by the linear model so I haven't made any. (5)

From the linear graph (my first model), it appears that there is a strong relationship between the weight in kg and the lean body mass in kg for males and females that play a sport. The correlation coefficient value is 0.93090405. This value expresses that there is a very strong and positive relationship between the two variables. Even though there appears to be a strong relationship between weight in kg and LBM in kg for athletes we cannot be sure that an increase in weight is completely responsible for an increase in LBM as there may be other factors involved that were not controlled in this investigation. (6)

This data is for athletes at the Australian Institute of Sport and so may not represent data for typical Australian men and women or actually people outside of Australia. It might not represent athletes from other countries either.

In conclusion there is a relationship between the weight in kg and lean body mass in kg for males and females that play a sport and a linear model seems to suit this data best. I would expect this information to be useful to the coaches and trainers in their on-going monitoring of these athletes - as it could be used to get an idea of how much muscle the athletes are gaining or losing as a result of training and diet. It also might be useful to do an analysis of the weight and LBM of male and female athletes separately or to look at relationships for athletes from other countries. (7)

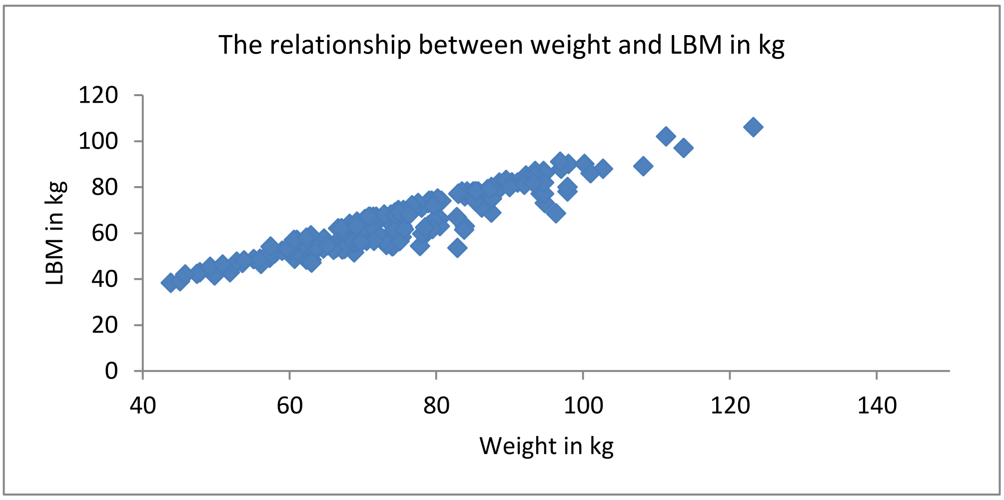
1 http://www.medicinasportiva.ro/SRoMS/english/Journal/No.6/The%20importance%20of%20body%20composition%20measurement%20at%20athletes%20and%20non%20athletes%20full.html

	Grade Boundary: Low Merit
3.	For Merit the student is required to investigate bivariate measurement data, with justification. This involves linking components of the statistical enquiry to the context, and referring to evidence such as statistics, data values, trends, or features of visual displays in support of statements made. A relationship question has been implied. The variable selection has been justified and there is contextual consideration of the data source (1). Appropriate displays have been selected and used (2). Features of the data have been identified and the nature and strength of the relationship have been described in context, including considering unusual features. Some contextual reasons have been used to justify comments (3). An appropriate model has been found and justified. The model has been used to make predictions. There is a contextual comment about the predictions (4). The student has considered that there may be differences in the relationship being investigated for male and female athletes. Two further models have been found and used to make predictions which are compared with the predictions made using the model for the entire group (5). Findings have been communicated (6). There is evidence of investigating bivariate measurement data with justification in the provision of contextual evidence to support statements for components of the statistical enquiry cycle. For a more secure Merit there would be more contextual discussion on key aspects such as how lean body mass affects an athlete's performance in difference sports, and a clear indication of the purpose of the investigation.

I am investigating if there is a relationship between weight in kg and lean body mass (LBM) for the athletes in the data set supplied from the Australian Institute of Sport.
The weight of a person is made up of body fat and lean body mass and according to Bodybuilding.com the LBM is the amount of weight you carry on your body that isn't fat. Erin Coleman, R.D., L.D. in an article on Livestrong.com says that "Lean body mass includes more than two-thirds water, according to Medline Plus, and the remainder other lean tissues such as muscle, organs and bone"
(http://www.livestrong.com/article/175858-the-average-lean-body-mass/#ixzz26y4L97ck).

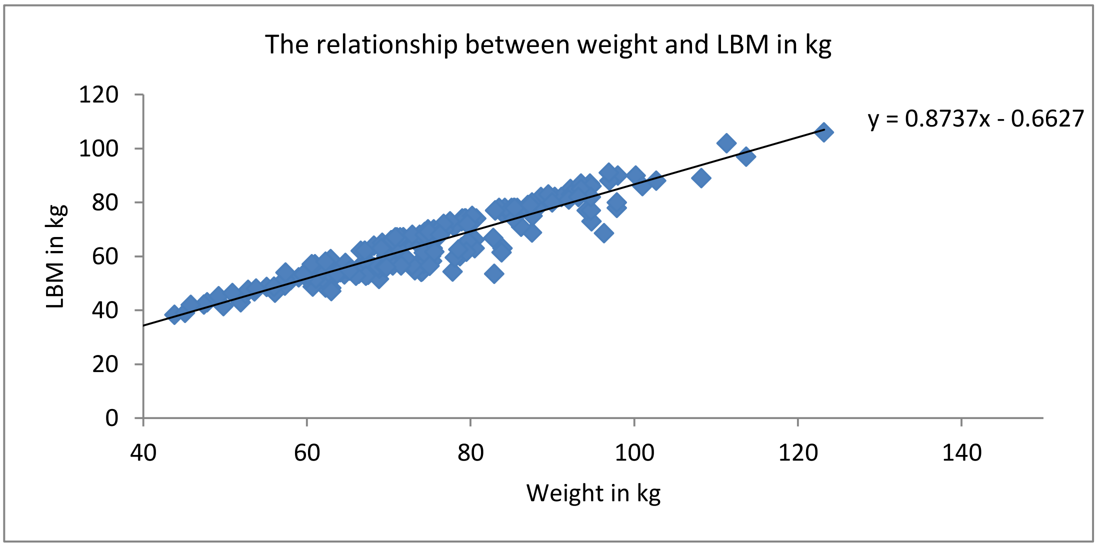
LBM affects athletic performance, appearance and weight. In some sports the amount of LBM can affect your performance and this information is likely to be useful to the coaches and trainers.
http://www.humankinetics.com/excerpts/excerpts/normal-ranges-of-body-weight-and-body-fat (1)

I have drawn an initial scatter plot using weight in kg as an explanatory variable and LBM as the response variable.



(2)

From the graph the weight in kg and LBM have a positive relationship which means that as weight in kg increases the LBM tends to increase. This may be because heavier athletes have heavy organs or muscles. The points are grouped closely together although the data for weights over 102kg are more spaced out they appear to be following the same line. I think that a linear model would be the best one and there isn't anything to suggest that a different one would be better. There does not appear to be any outliers or unusual features and the points are clustered in the middle of the graph. (3)

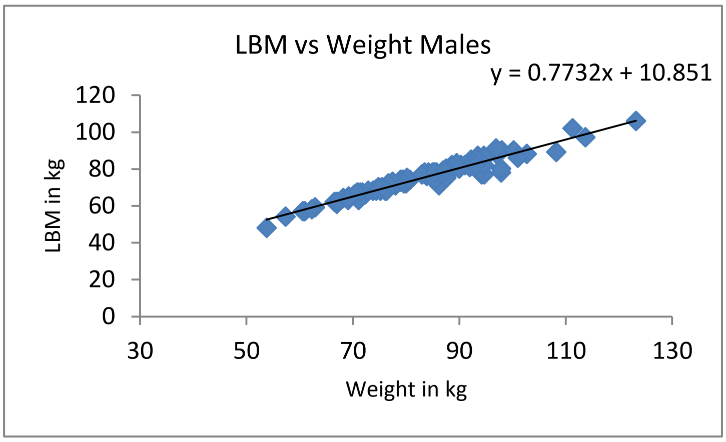


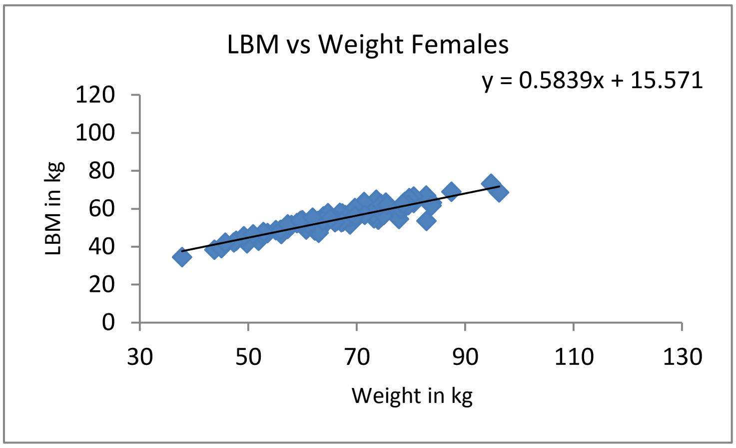
The trend line equation is y = 0.8737x – 0.6627. The relationship is strong because the points are close to the regression line. The model proposed suggests that as the weight increases by 1kg, the LBM increases by 0.8737kg. (3)

Prediction.
According to the Australian Bureau of Statistics the average weight of a man of average height is 85.2kg and a woman of average height is 70.1kg.
(http://www.abs.gov.au/ausstats/abs@.nsf/Lookup/4841.0Chapter22011).
While these will not be the same necessarily for the average weights of high performance athletes I will use a weight within this weight range to predict the LBM of athletes.

I will use the weights of 75kg and 100kg to predict what the LBM will be. Since the linear model appears to fit the data well I will use this to make my predictions. For the prediction, I substituted the weights into y = 0.8737x – 0.6627.
y = 0.8737*75 – 0.6627 = 64.8648 = 64.9kg which is above the data points (LBM values) near this weight,
y = 0.8737*100 – 0.6627 = 86.70739 = 86.7kg which is in the middle of the data points (LBM values) for this weight on my graph. (4)

I wonder if there are differences in the relationship between male and female athletes and also between different sports in the data set. I will now look at the relationship between for LBM and weight in kg for men and women separately.





The linear model for males is y = 0.7732x + 10.851 and for females the linear model is y = 0.5839x + 15.571. For both of these relationships the points are close to the regression line. Looking at the model proposed for the men the data points appear to be scattered more closely to the line than that for the men and women together and I think it is likely if I use this model to predict the LBM for a male athlete the results will be more reliable. Visually the scatter on the graph for the female athletes' is similar to the one with all the men and women together although there are no data points beyond a weight of 97kg.

Male predictions:
75kg, LBM = 0.7732*75 + 10.851= 68.841 = 68.8kg.
100 kg, LBM = 0.7732*100 + 10.851 = 88.171 = 88.2kg.
As this model looks to be a better fit than the combined model I think that these predictions are fairly accurate. (5)

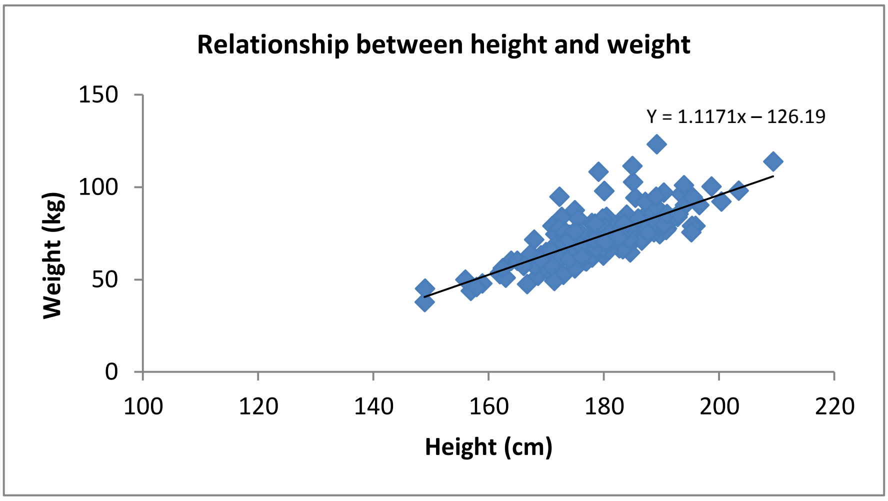
Female predictions:
75kg, LBM = 0.5839*75 + 15.571 = 59.36 = 59.4kg. Looking at the other data points (LBM) around this weight, I think that this is a reliable prediction for the LBM.
100 kg, LBM = 0.5839*100 + 15.571 = 73.961 = 74.0kg – there aren't any female athletes of this weight in the graph so this prediction may be more unreliable than those for weights of less than 100kg.

I think that the combined linear regression model fits the data investigated well but the separate models are likely to be more reliable and appear to be a better fit. This information could be useful to coaches and trainers. I didn't look at how well these models might fit the 'average population' or how the different sports the athletes took part in may influence the relationships. (6)

	Grade Boundary: High Achieved
4.	For Achieved the student is required to investigate bivariate measurement data. This involves showing evidence of using each component of the statistical enquiry cycle. An appropriate relationship question has been posed. The explanatory and response variables have been identified (1). Appropriate displays have been selected and used (2). Features of the data have been identified and the nature and strength of the relationship have been described in context (3). Appropriate models have been found and have been used to make predictions (4). The student has considered another pair of variables and compared the strength of this relationship with the original investigation (5). Findings have been communicated in a conclusion. There is some contextual reasoning in the discussion, but this is not supported (6). There is evidence of investigating bivariate measurement data in the use of each component of the statistical enquiry cycle. To reach Merit the student would need to provide more in the way of contextual references to support statements made. There would need to be some evidence that other background research material had been considered.

Student 4: High Achieved

Is there a relationship between the height in cm and the weight in kg for athletes in the data supplied from the Australian Institute of Sport? I think it is likely. There is an app from the BBC where you can put in your height and weight and see which Olympian is your body twin http://www.bbc.co.uk/news/uk-19050139 - this is based on the relationship between the height and weight of athletes.
I have made height the explanatory variable and weight the response variable. I think this may be useful to the institute when they are identifying future athletes for training. (1)



(2)

The scatter graph has a strong positive relationship between the height and weight of sports athletes at the Australian Institute of Sport. The data in the graph tells us that as an athlete gets taller (cm) their weight (kg) will generally increase as well. The relationship is strong as the data points are close to the fitted trend line and 'r' is 0.78090629 which also confirms this.

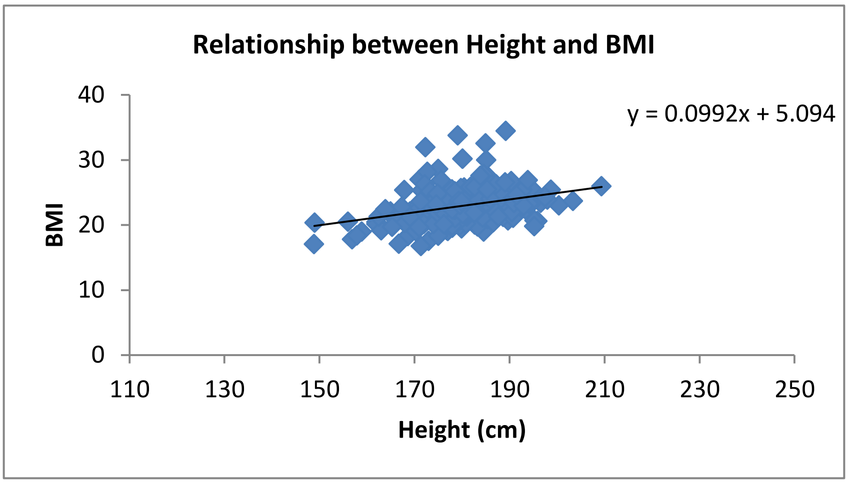
The linear model fits this data quite well as it is positive and shows generally as height increases so does weight this confirms my statement from above, this is because the taller the athlete is the more muscle/fat they will have making them weigh more. There are some values more scattered away from this linear model - such as the one that weighs around 120kg and is around 180cm tall, this is probably due to different sports requiring different physiques, in this instance it is a male 'field' athlete. This graph also shows me that in the data supplied there are only a few athletes taller than 200cm and these are all basketball players and the linear model stops at around 200cm. (3)

To make my prediction I used a male TSPRNT individual with a height of 189.1cm to predict his weight in kg. I used the equation that I got from my linear model to work this out.

Prediction = 1.1171*189.1 – 126.19 = 85.05361 = 85.1kg

I did another prediction with data outside of my graph I used a height of 210cm to predict the weight. I also used the equation from my linear model to work out the prediction.
Prediction = 1.1171*210 – 126.19 = 108.401 = 108.4kg (4)
This confirms my thinking that with a shorter person their weight was less than a person much taller.

I decided to also investigate if there is a relationship between Height (cm) and BMI (weight/height^2).
The scatter graph shows a moderate positive relationship between the height and BMI (body mass index) in athletes at the Australian Institute of Sport. The graph shows that as the height of an athlete increases their BMI will increase. As BMI is worked out from (weight/height^2) we might expect this.



The graph shows that the linear model fitted for this data seems ok, the data appears to be mostly clumped together – there are still some BMI values for heights between 175cm and 189cm where the data is scattered further away from the regression line. (5)

Discussion
Looking at my scatter graphs for the two relationships the relationship between height and weight is a much stronger relationship than that between height and BMI – the data points are much more closely scattered around the regression line for height/weight. This is supported by the correlation for height/weight being 0.78 and the correlation for height/BMI being 0.34. This tells us that the relationship between height and weight of an athlete is much stronger the relationship height and BMI of an athlete. The height/weight model may be better fitting to the linear graph due to it comparing the height and weight and as someone gets taller they generally have more muscles with larger bones, whereas for the height/BMI the BMI is calculated already from weight/height^2 therefore it is already including the height in the graph. (5)

Wikipedia tells us that BMI is inaccurate for athletes and people who are fit because they have higher amounts of muscle which puts them in the "overweight" category even though their body fat percentages are in the 10-15% category.

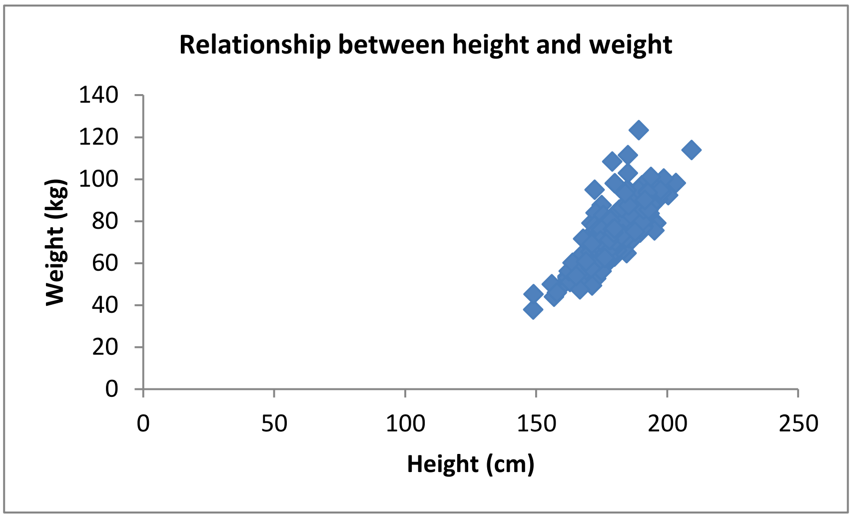
Being an institute of sport they will be focusing on particular sports and building muscles so it may also be dependent on the sport, for example in sprints you don't need the whole body at equal strength as you may need in water polo, due to sprints mainly involving timing and the power in your legs; whereas water polo you need to be able to stay afloat through your legs and be able to throw/catch the ball.

I think that the height/weight model would be useful for the sports institute to identify potential future athletes. (6)

	Grade Boundary: Low Achieved
5.	For Achieved the student is required to investigate bivariate measurement data. This involves showing evidence of using each component of the statistical enquiry cycle. An appropriate relationship question has been implied. The explanatory and response variables have been identified. A simple purpose which relates to the Australian Institute of Sport has been identified (1). Appropriate displays have been selected and used (2). Features of the data have been identified and the nature and strength of the relationship with regard to visual aspects of the scatter about the regression line have been described in context. Comments are limited in terms of contextual understanding (3). An appropriate model has been found and it has been used to make predictions. There is limited discussion about the predictions (4). Findings have been communicated in the conclusion and linked to the purpose of the investigation (5). There is evidence of investigating bivariate measurement data in the use of each component of the statistical enquiry cycle. For a more secure Achieved there would have been more depth in the description of the relationship and features of the display, particularly with respect to relating descriptions to the context of the investigation.

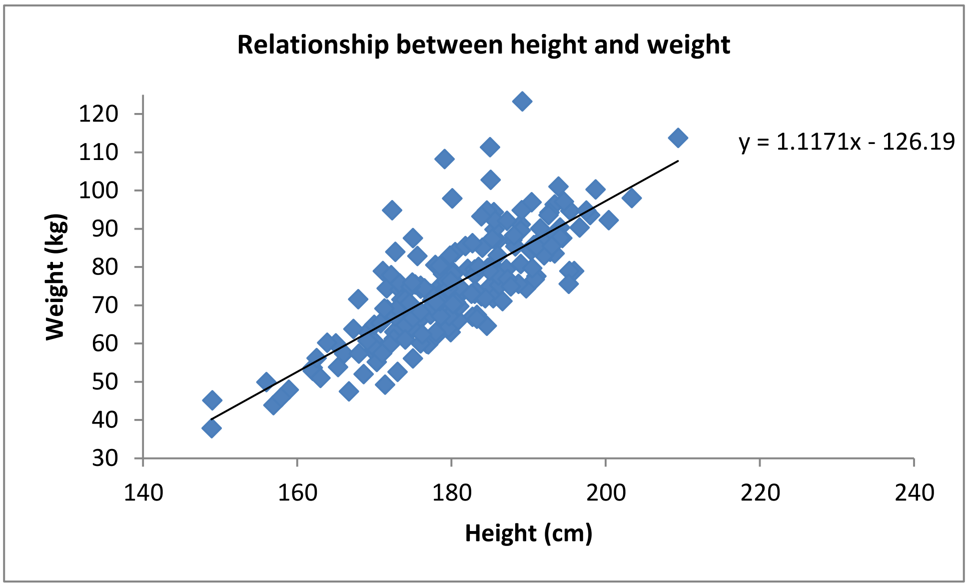
I will investigate if there is a relationship between a person's height and a person's weight for athletes from the Australian Institute of Sport as this may be useful for coaches and trainers. Height is the explanatory variable and weight is the response variable. For instance the South Australian Sport Institute's (SASI) Talent Search Program run a talent identification programme which encourages South Australians between 13 and 25 years to submit their details and some test results via a Facebook application. They don't only want people who are already good at sports – they say that a person's physical and physiological characteristics might tell them whether a person has the potential to make it in a sport that they have never have even tried before. http://www.recsport.sa.gov.au/sasi/talent-id-development.html

(1)



From looking at this graph, I can see that there is a strong positive relationship between the two variables. This means that the taller a person is in height, the heavier they are in weight.

This graph shows me that it is not very likely for an athlete to be taller than 200cm and weigh more than 190kg. The data tells me that the average height of an athlete is 180cm tall and the average weight is 75kg.



(2)

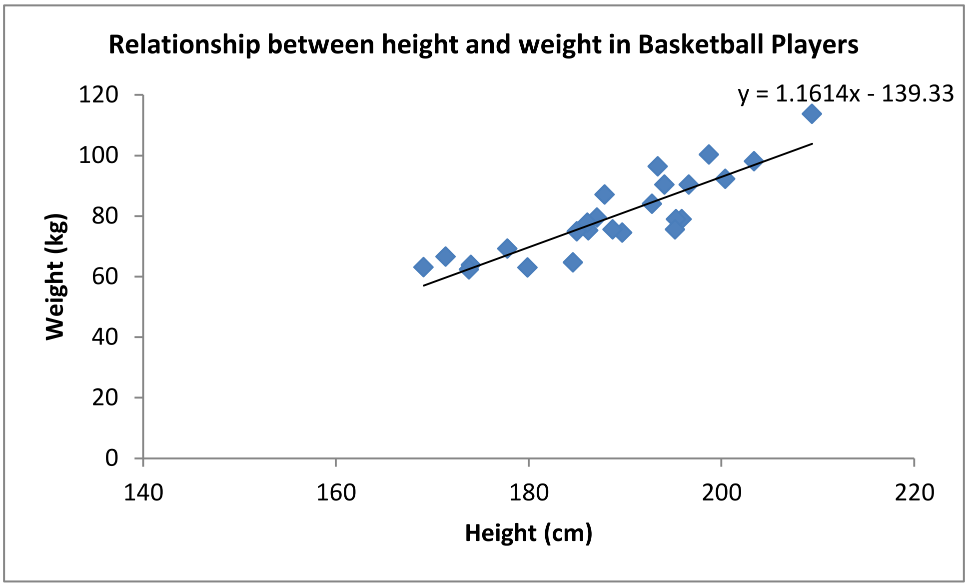
The straight line fits this data well and most of the data points are close to the line indicating a strong relationship between the height and weight of an athlete. The positive gradient means that as the height increases, so does the weight of an athlete. However, there are some data which are further away from the regression line this might mean that another model could also be used. (3)

The equation of this data is weight = 1.1171*height – 126.19

For an athlete of height 180cm my prediction for their weight would be w = 1.1171*180-126.19 = 74.888 = 74.9kg which seems reasonable.

For an athlete of height 200cm my prediction for their weight would be w = 1.1171*200 – 126.19 = 97.23 = 97.2kg which is a bit higher than we might expect from the data. (4)

Looking at the data I can see that Basketball Players are at the top end for height and weight, I have decided to look at the relationship between height and weight for only Basketball players from the data set.



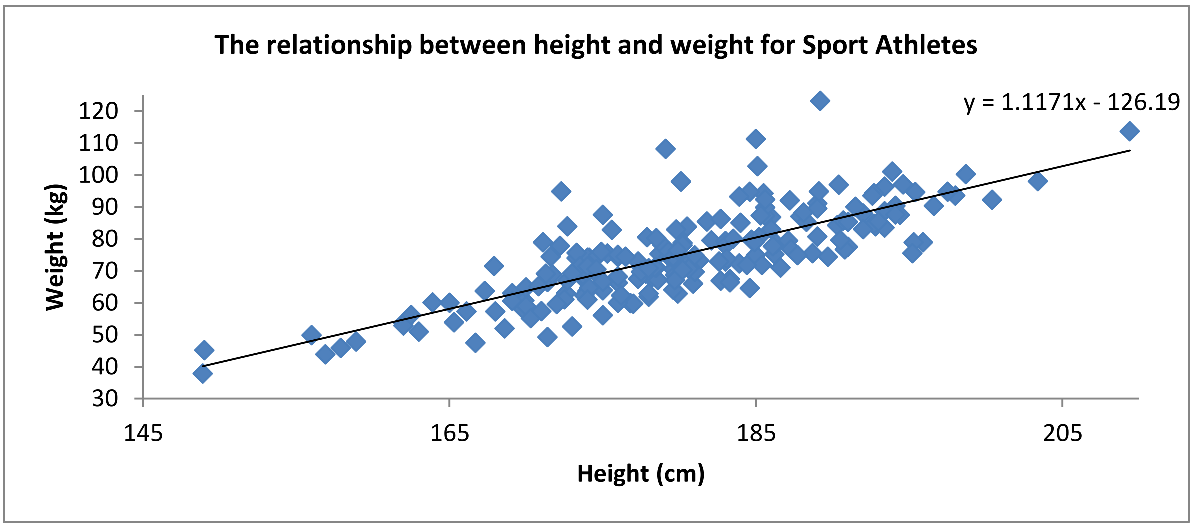
Using this model for just the basketball players my predicted weight for an athlete of height 200cm would be w = 1.1614*200 – 139.33 = 92.95 = 92.9kg which is closer to the actual values in the data set.

In conclusion I think there is a strong positive relationship between the height and weight of athletes – the taller an athlete is the more that they will weigh. (5)

My model for just the basketball players shows a relationship that could be used to identify potential players by their height and weight. I expect that other things such as hand eye coordination could be important but I think that this is a quick identifier that would be useful to coaches and trainers in their selection and training of athletes. (5)

	Grade Boundary: High Not Achieved
6.	In order to achieve the standard the student is required to investigate bivariate measurement data. This involves showing evidence of using each component of the statistical enquiry cycle. An appropriate relationship question has been implied. The explanatory and response variables have been identified (1). Appropriate displays have been selected and used (2). Some features of the data have been identified and the nature of the relationship has been described in context. The strength of the relationship relies on the value of the correlation coefficient, rather than referring to the visual aspect of scatter about the regression line (3). An appropriate model has been found and it has been used to make predictions (4). Findings have been not been communicated clearly in a conclusion. The requirements for Achieved have not been met because the description of the strength of the relationship needs to refer to visual aspects of scatter about the regression line, rather than just rely on the value of the correlation coefficient. There is no conclusion. To reach Achieved the student would need to have described the strength of the relationship by discussing visual aspects of scatter about the regression line in context. They would also have needed to provide a conclusion.

I will look at the data from the Australian Sports Institute to see whether or not there is a relationship between the height (cm) and weight (kg) for sport athletes. The explanatory variable is the height while the response variable is weight. (1)



(2)

The relationship between height and weight appears to be linear so I will fit a linear model to the data. This data shows a strong positive relationship. This means that there is a relationship between the height of an athlete from the Australian Institute of Sport and their weight - as athletes grow taller they generally weigh more. For this data the weight is more likely to be muscle as these people are athletes and depending where the most weight is, muscle weighs more than fat. (3)

The correlation coefficient, also known as r, is 0.78090629 therefore showing that r is positive and that the relationship is of reasonably strong strength.

There are a few large weights compared to their height for example 123.2 kg for someone of a height of 189.2 cm. Between around 170cm and 190cm there are a few plots higher up, this could likely be because these athletes could be very strong and the weight is muscle because of the sports they do. (3)

The data values between 170cm and 190cm in the scatter plot are more scattered above the line and it maybe another model could be used is for these values such as exponential or logarithmic as these may be better fit for the data.

A possible prediction I could make would be the weight of someone that is 160cm tall. To do this I would use the equation 1.1171*160 – 126.19 = to work out this prediction. The answer for this is 52.5 (rounded to one decimal place). Therefore someone that is 160cm tall would be approximately 52.5kg. This seems reasonable as a female athlete in the data set with a height of 162cm has a weight of 52.8kg.

If I want to see how much someone would weigh that is 211cm tall by using the same equation as previously used 1.1171 *211– 126.19, my answer would be 109.5kg (rounded to one decimal place). There are only three values over 200cm on the data plot and their weights are between 92 and 114kg. (4)