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

Exemplar for Internal Assessment Resource

Mathematics Level 3

Resource title: Elite athletes

This exemplar supports assessment against:

Achievement Standard 91582

Use statistical methods to make a formal inference

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 use statistical methods to make a formal inference, with statistical insight. This involves integrating statistical and contextual knowledge throughout the statistical enquiry cycle and reflecting on the process. There is evidence that statistical and contextual knowledge have been integrated. The student has used their research to develop a purpose and pose an appropriate question (1). Analysis comments are clearly linked to the context and there is evidence of considering other explanations in the investigation into the impact of gender (highlighted). For a more secure Excellence more in-depth research would be needed, and the graphical evidence from the investigation into gender should have been included.

Student 1: Low Excellence

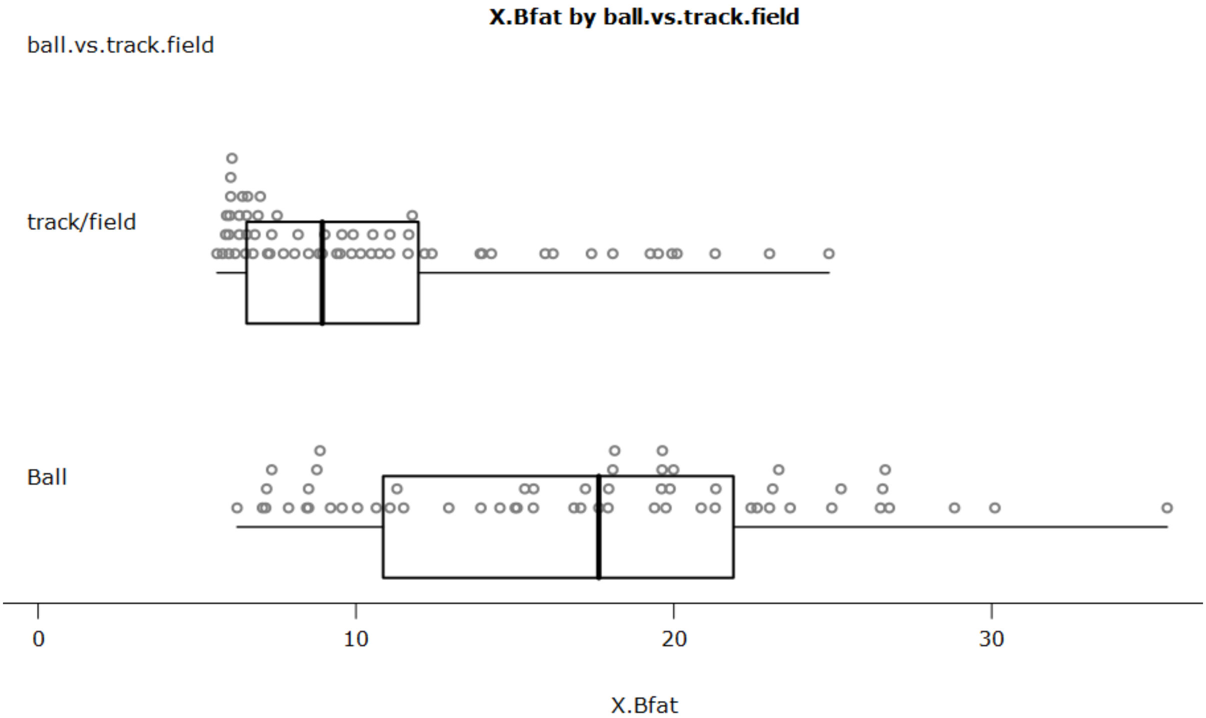
Purpose of investigation:
I am required to analyse some data about the athletes of the Australian Institute of Sport and write a concluding report about it for Sports and Recreation NZ. I will be looking to see if there is a difference between 2 particular groups looking at one variable.
In my investigation I will be determining the difference of %Bfat between different types of sport/event. Wikipedia says that essential fat for women is between 10–13% and 2–5% for men; however this is the minimum percentage a person can have while still being healthy. It also said that for athletes the average %Bfat was 14-20% for women and 6–13% for men, this however will also vary on what sport/event the athlete is competing in.

I suspect that the median %Bfat of ball sport players will be higher than the median %Bfat of track/ field athletes. I do also expect some variety in each group due to the different types of game/event each athlete participates in.

Identify variables to investigate:
I will investigate the percentage of body fat (%Bfat) in athletes of the Australian Institute of Sport (AIS). I will categorise my sports variable into two classes. I will combine all sports involving a ball (netball, basketball and tennis) into one class, called "ball" as I suspect this group will have similar %Bfat. I will combine all track and field into 1 class, called "track/field". For the purposes of this study I am only interested in athletes who fit these categories and I will ignore all the others. This is to ensure the results of this investigation are valid for the context and allow me to make conclusions.

Comparative investigative question:
What is the difference in the median body fat percentage (%Bfat) of track/field athletes and the median body fat percentage of ball sport athletes in the Australian Institute of Sport (AIS)?

1




```
Summary of X.Bfat by ball.vs.track.field
            Min. 1st Qu. Median  Mean 3rd Qu.  Max. Std.dev Sample.Size
Ball        6.26   10.75  17.43 17.05   22.15 35.52   6.959          58
track/field 5.63    6.56   8.94 10.43   11.96 24.88   4.921          63
```

Discribing and compering sample dstributions:

Centres:

The median %Bfat of ball sport players is 17.43% and the median %Bfat of track/field athletes is 8.94%. The difference between these two medians is 8.49% meaning that the median %Bfat of ball sport players is almost double the median of track/field athletes. Though I thought the median %Bfat of ball sport players would be higher than that of track/field athletes I did not expect it to be that much higher.

The median %Bfat of ball sport players is quite high – I expected it to be lower. Also the maximum %Bfat for a ball sport is 35.52% which actually puts this player in the 'obese' category, this was very surprising. This leads me to wonder what is causing such high %Bfat in these ball sport players and whether this will also be the case back in the entire population of athletes involved in AIS. Further investigation involving other countries and their athletes could be interesting.

Shape:

The distribution of the %Bfat of track/field athletes in AIS is skewed to the right. The bulk of track/field athletes have a %Bfat between 5.63 – 12%, the rest of the data is spread out to a maximum of 24.88%.

The distribution of the %Bfat of ball sport players in AIS is much more spread than that of track/field athletes. The middle 50% of the ball sport players have their % body fat between 10.75% and 22.15% which is range of 11.4% but for track and field the range of the middle 50% is from 6.56% to 11.96% body fat which is a range of 5.4%. I am not comfortable to make a call on the distribution of this graph; more data would be needed to make this judgement. There are two bunches in the data; one at the very bottom of the scale (form 6.26% - around 9%), the other bunch is situated around the median (from 17% - 22%). Between these two groups the data spreads out slightly, and after the middle bunch the data becomes quite spread up to the maximum of 35.52%. I think that the distribution of %Bfat for ball sport players could be bimodal, I will investigate this further.

After looking at the data more closely, I found that the two bunches in the %Bfat of ball sport players is caused by the difference of %Bfat of the different genders, which supports what I found in Wikipedia. The lower bunch of values is formed by the males with only one female in the labelled bunch (form 6.26% - around 9%). The bunch around the median value is formed by the females with no males found in this labelled bunch (from 17% - 22%). I went further and re-graphed the %Bfat of ball sport players, splitting it into males and females. The middle 50% of males are between 7.6% - 9.8% which is very similar to the range of the first bunch (form 6.26% - around 9%). The middle 50% of females is between 17% - 23.2% which again is very similar to the range of the second bunch (from 17% - 22%).

From further investigation on the %Bfat of ball sport players I have found that it is very likely that back in the population the distribution is bimodal with the two peaks produced by the difference between males and females.

I was initially very surprised to see such high percentages of body fat in both groups as we are talking about national elite athletes and the higher %s were higher than the figures from Wikipedia but they were for athletes which could include several sports. However I do know that in some sports/events it can actually be an advantage to have a higher %Bfat. Findings from Wilmore in 1994 gave figures for a variety of sports and showed females basketball players from 20 – 28% body fat and female track throwers 20 – 27%.

The highest %Bfat in track/field athletes was 24.88%. This value is from a female field athlete. Some field events such as shot put do not require large amounts of running so a large %Bfat will not hinder the performance of the athlete, in fact in some cases it can actually be an advantage to be bigger.

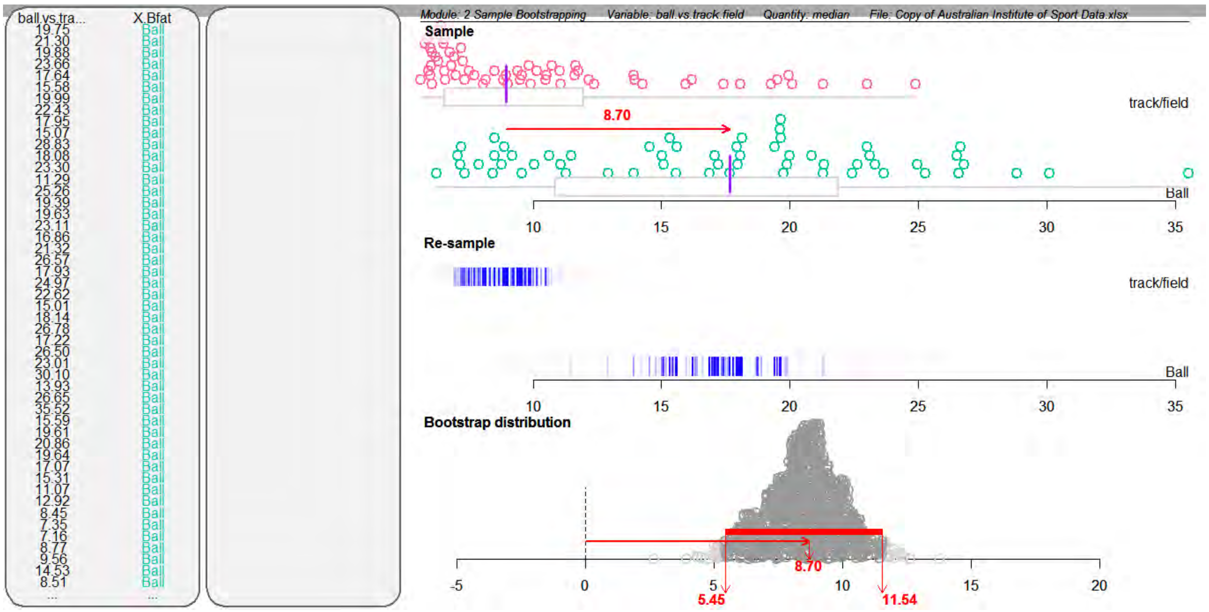
The highest %Bfat for ball sport players was 35.52%. This value is from a female netball player; from this it is likely that this particular player plays either as goal shoot or goal keep. Because these positions do not require as much running/physical activity, the player would be able to play unhindered by this larger %Bfat. That said, from my research I found that obesity for a female is said to be from 32% and above. That puts this netball player in the obese category, which to me is a worrying factor considering she is a national elite athlete.

These higher %Bfat values should be expected in athletes because of the different requirements of different sports/events. However, I am still surprised that the values are as high as they are. It is also likely that back in the population there will be more athletes will similarly high %Bfat and these values at the top end of the scale are not just outliers.

Spread:
For the %Bfat of track/field athletes, the skew mentioned above contributes to the spread of the middle 50% of the data but it is still relatively compact with an IQR of 5.4%. In comparison, the middle 50% of the %Bfat of ball sport players is much more spread with an IQR of 11.4%. For the ball sport players the LQ (10.75%) has been shifted further down because of the bunch at the lower end of the scale (mentioned above). This has caused the lower half of the middle 50% to be more spread than the upper half of the middle 50%. Because the spread of the sample of %Bfat of ball sport players is so spread, it suggests that if I was to take other samples, the median %Bfat of these samples could vary far more than medians %Bfat of other samples for the track/field athletes group. There would be a larger band of possible values of true medians for %Bfat of the ball sport players than the band of possible values for true medians of the %Bfat of track/field athletes. This larger spread of data could be caused by the different sports and positions in the group. Each sport and position in each sport can require a different build and because of this there is a lot of variation in the group of ball sport players.
Due to the range of these samples, I would expect that the data in these samples would be representative of the population. However I do think that the IQR of the %Bfat of both ball sport players and track/field athletes will decrease as the bunches effecting the position of the IQR will have more values causing the IQR to shrink.

Shift/Overlap:
There is a small overlap between the UQ of the %Bfat of track/field athletes (11.96%) and the LQ of the %Bfat of ball sport players (10.75%). The overlap is 1.21%.

Because there is such a small overlap and the medians are well separated from this overlap, I would be confident to say that the median %Bfat of ball sport players is higher than the median %Bfat of track/field athletes.
Just to be sure I will calculate a bootstrapping confidence interval to confirm this.
A bootstrap confidence interval is a range of believable values for the difference in medians of the population parameter (the %Bfat of athletes in AIS)



Statistical inference:
It is a very safe bet that for the population of athletes in AIS, the median %Bfat of ball sport players will be between 5.45% and 11.54% higher than the median %Bfat of track/field athletes.
Therefore I can make the call that the median %Bfat of ball sport players is higher than the median %Bfat of track/field athletes.

Conclusion:
I can make the call as my confidence interval did not contain zero. This means that there is a very high chance that in every sample from the AIS population the median %Bfat of ball sport players is likely to be higher than the median %Bfat of track/field athletes.
I decided to use medians as my sample statistics as the means would be influenced by the few large values. The results of my investigation were the same as my hypothesis. I was surprised by the large %Bfat of athletes in both groups however I concluded that these high values may be due to the nature of position or event. The other reason for this is that these athletes with large %Bfat may be new to the program and therefore may not be at their best competing state.
One thing I didn't consider in my study was gender. As my research showed males have a lower percentage body fat than women. I discovered this gender difference in %Bfat is connected with body structure to allow child birth and being able to support the baby in the womb. Perhaps my study would have been better if I first compared %Bfat of male track/field athletes vs %Bfat of male ball sport players and then separately %Bfat of female track/field athletes vs %Bfat of female ball sport players. I could have then seen if the differences were similar.

How could Sports and Recreation NZ use this?
Because this investigation suggests that athletes from certain types of sports have different %Bfat, Sports and Recreation NZ could use this information to alter the training programs of their athletes by looking at what %Bfat is zone specific to their sport. I suggest that further investigation into other areas of sports and even specific sports and their 'ideal' %Bfat. In doing this they will be able to ensure maximum performance of each athlete in their specific sport/event.

	Grade Boundary: High Merit
2.	For Merit the student is required to make a formal inference, with justification. This involves linking components of the statistical enquiry cycle to the context and making supporting statements which refer to evidence. The student has developed a purpose for the investigation, but has not given an explanation for the choice of team and individual (1). The discussion has been linked to the context throughout and the conclusion has been clearly justified (2). The student has attempted to seek explanations for some of the findings in the comments on young and older athletes (highlighted), but has overlooked any consideration of the difference between male and female athletes. To reach Excellence more evidence of integrating statistical and contextual knowledge is needed. This would include further research and investigation into the bimodal distribution.

Student 2: High Merit

Australian Institute of Sport Investigation

Purpose of the Investigation:

'Olympic athletes have natural talent, dedication, and drive. They devote their lives to their sport in hopes of being the best in the world. No matter how talented or driven an athlete, however, they must train hours a day to perfect their skills and maintain their phenomenal level of physical shape. An aspiring Olympic athlete spends an average of eight hours a day, seven days a week training their body and mind—more time than a full-time job' – Elizabeth Olsen, Infoplease

There are many differences between team and individual sports, and as shown in the above quote. Individual athletes put hundreds of hours into their training, particularly in the build-up to the Olympics. I am investigating the percentages of body fat in athletes who play individual sports compared to athletes who play team sports, from data from the Australian Institute of Sport. This could provide evidence for nutrition and training schedules in elite athletes, particularly for those who want to alter their training.

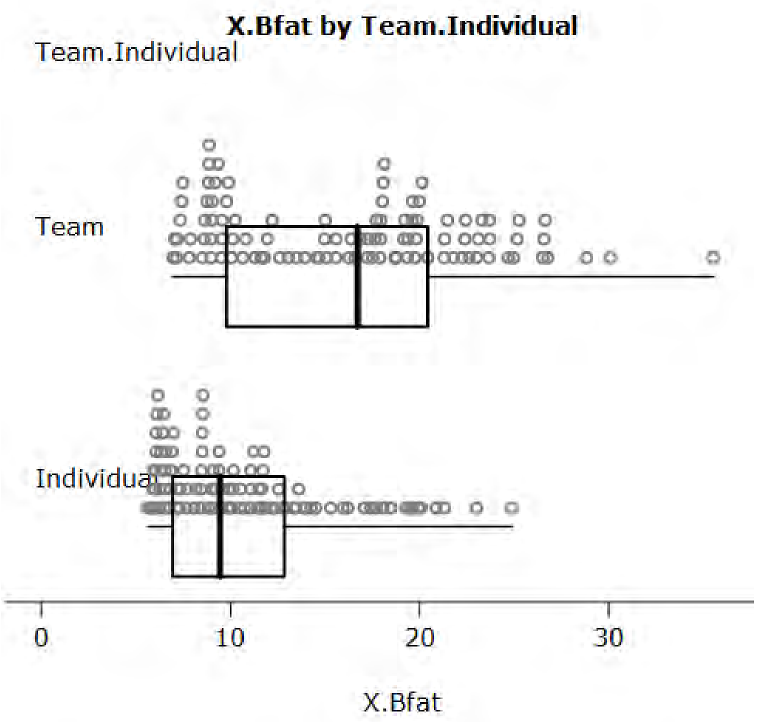
I think that individual athletes will have a lower percentage of body fat than athletes who pay team sports because of their intense training programs, and the fact that they can rely solely upon themselves in order to win a competition. Team players can rely on the fitness of other team members in order to win a game. Different sports also need different types of fitness and the length of peak performance in team sports is often longer. A sprint athlete is preparing for full intensity over a short period of their race whereas in a ball sport the full intensity is over a longer period.

Variables to be Investigated:

I will investigate the sport(team or individual) of those represented in the Australian Institute of Sport data and the 1 percentage of body fat the athletes have. I have re-categorised my data (Basketball, Field, Gym, Netball, Rowing, Swimming, Track (400m and Sprint), Tennis, Water Polo) to provide groups for Team or Individual Sports (Tennis was counted as a team sport).

Comparative Investigative Question:

What is the difference between the median percentages of body fat of athletes who compete in individual sports and the median percentages of body fat of athletes who compete in team sports?



Summary of X.Bfat by Team.Individual

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Std.dev	Sample.Size
Individual	5.63	6.972	9.45	10.77	12.82	24.88	4.548	100
Team	6.96	9.815	16.72	16.20	20.35	35.52	6.422	102

Describing and comparing Sample Distributions:

Initial Visual Interpretation:

My initial impression is that the body fat percentage of those who play team sports is on average higher than those who compete in individual sports. This confirms my initial hypothesis. Athletes who compete in team sports may need to alter their training to have elements that are more similar to individual training, as well as the relevant training for their sport if they want to decrease the amount of body fat they have.

Centre:

The median % body fat for the team athletes is 7.27% higher than the median for the individual athletes. I think this a significant number in terms of body fat because that means the percentage body fat for some team athletes would be double some individual ones, even just by adding 7.27% to the fat.

Shape:

The distribution of % body fat for team athletes is clearly bimodal. The majority of the data is found in two regions, (roughly 7-10%) and (roughly 17-26%). Interestingly, the first region is fairly similar to the group for the individual athletes, while the second group is significantly higher. This makes me think that back in the population, there could actually be two groups of team athletes – those who perhaps are young and extremely fit, and maybe those who are older and have the experience so can make the team but are perhaps not as fit as they used to be. These might be the players who rely on other members' fitness. They might want to increase training plans or change nutrition in order to lose body fat if they do actually want to.

The distribution of % body fat for individual sports is clearly right skewed. The left part of the graph where the individual athletes with lower % body fat are has a much smaller spread than the right side of the graph. Looking at the box and whisker the lowest 25% of the % body fat for the individual sports has a very small spread from 5.63% to 6.97% with 75% of the % body fat below the upper quartile of 12.82 and then the 25% of individual athletes with the highest % body fat have measurements from 12.82% to 24.88% which is a spread of almost 12% body fat. The spread of the % body fat for these athletes (12.82% to 24.88%) is more than the spread of the remaining individual athletes (5.63% to 12.82%). It is interesting that the curve is not symmetrical and then right skewed, but actually around the minimum % body fat there is the most number of athletes. This makes me quite confident in my initial hypothesis, that individual athletes do have a lower percentage of body fat back in the population.

Spread:

Both the full range and the interquartile range of the % body fat is a lot greater for the group of team athletes than the individual athletes. This didn't surprise me and I would expect the same to be true in the population. I would also expect quite a bit of variation in the individual team members especially in some teams where different positions demand different physiques, for example in rugby there are quite different physiques in the forwards (tall and lean for locks but heavier and often shorter front row). For the team athletes, the range is 28.56% and the interquartile range is 10.535%. For individual athletes the range is 16.25(significantly less than the range of the team players), and the interquartile range is 5.848(roughly half of the IQR of the team athletes!). These ranges show me that back in the population, the percentages of body fat for team athletes will include much higher percentages than those of the individual athletes.

Shift/Overlap:

The % body fat for team athletes is shifted higher up the scale than the % body fat for individuals. For team athletes, the top 75% of the data has a greater % of body fat than the bottom 50% of the data for individual athletes. There is a small overlap in the middle 50% of both groups, with the rest of the team middle 50% extending to the right of the graph. This is a clear difference. It shows me that back in the population, the median % body fat of team athletes would most likely be higher than those of individual athletes.

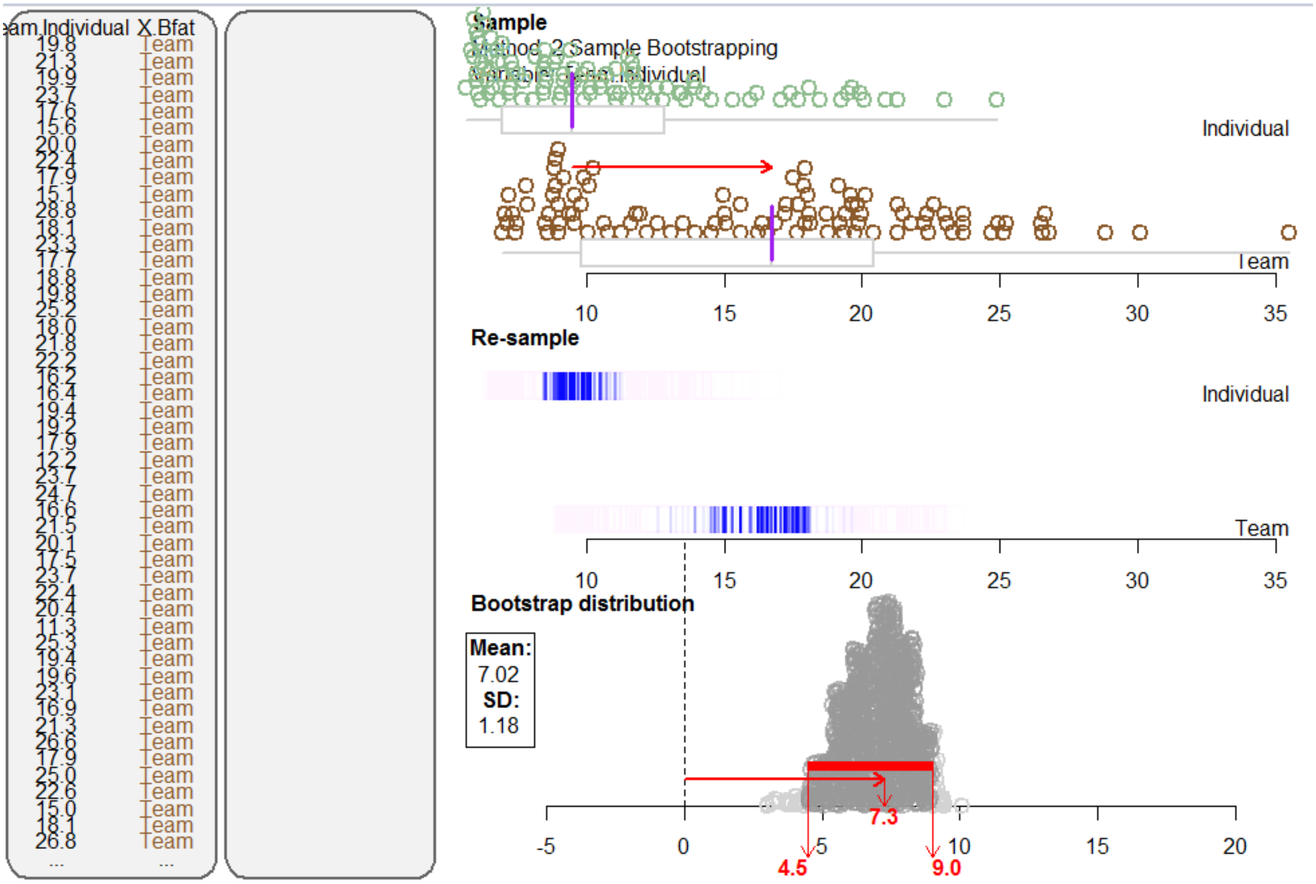
Unusual Features:

I do not see any unusual features.

Statistical Inference:

This data shows me that it is quite likely that back in the population there is in fact a difference between the medians for the percentage of body fat in individual athletes and team athletes. My findings agree with my initial hypothesis - that is that the athletes who compete in individual sports have a lower median percentage of body fat than that of those who compete in team sports'

I created a bootstrap confidence interval for the difference between these two medians so that I get a clearer picture of the difference in the median percentages of body fat back in the population. I resampled a thousand times, and I am assuming that this data is representative of all elite athletes with the Australian Institute of Sport.



My bootstrap confidence interval is from 4.5 to 9%. This does not include 0, and therefore I am confident that there is a difference between the percentages of body fat in the athlete population.

2

I think it is fairly safe to say based on this interval, that the median percentage of body fat in team athletes is 4.5-9% higher than the percentage of body fat in individual athletes from the Australian Institute of Sport.

Conclusion:

Initially I thought that athletes who compete in Individual sports would have a lower percentage of body fat than those who compete in team sports because of the fact that some team players rely on others for their fitness levels, while individual athletes can only rely on themselves. My research and confidence interval confirmed my initial thoughts, with a confidence interval that showed that the median percentage of body fat of team athletes was between 4.5-9% higher than individuals. This interval works based on the assumption that this sample is reflective of the entire population of athletes in the Australian Institute of Sport, and because of the significant range of the interval, I'm fairly sure that another sample would provide very similar results.

I think that one piece of data that should have been taken into account is the age of the athletes, as a younger athlete is more likely to have lower body fat than an older athlete. This may have shown me that in my bimodal curve for the team athletes, those in lower region were all younger than 30, while those in the higher region were those older than 30. This may have altered my final report. However, I am still confident that this data does give a good idea of what athletes look like back in the population.

	Grade Boundary: Low Merit
3.	For Merit the student is required to make a formal inference, with justification. This involves linking components of the statistical enquiry cycle to the context and making supporting statements which refer to evidence. An explanation for the choice of variables has been given and an appropriate question that identifies the population and the difference between the medians of the two groups has been posed (1). Some of the discussion (highlighted) on the distributions has been linked to the context or investigative question. The student has made a correct call that has been justified (2). For a more secure Merit the linking of the analysis discussion to the population needs to be developed and more depth of understanding needs to be evident in the conclusion.

Student 3: Low Merit

Investigation for Sport NZ

Purpose of the investigation: An article written by Krista Sheehan, written on the 16th of April 2011, posted her article on the livestrong.com website[1]. Her thoughts concluded that while you exercise for a long period of time (high endurance sports) your white blood count goes down as the blood goes down as the blood is being pumped around the body at a faster rate. This can result in bad circulation. Some sports "Over a long period of time can cause improper blood circulation" Sheehan, says. Bad circulation means that the vital oxygen and white blood cells cannot be pumped around the body to meet demand. Bad circulation can lead to Diabetes, Arthritis, high cholesterol, high blood pressure, angina and heart disease[2]. The result of my investigation will be helpful to Sport New Zealand ((Sport NZ) is the new name for the government organisation responsible for sport and recreation (formerly SPARC"[3]).This investigation could help those, help athletes perform better as they can take steps to increase WCC and better circulation. The data I will be using to conclude my investigation is from The Australian Institute of Sport (AIS.) The White blood Cell Count is counted in thousands per cubic millilitre of blood[4].

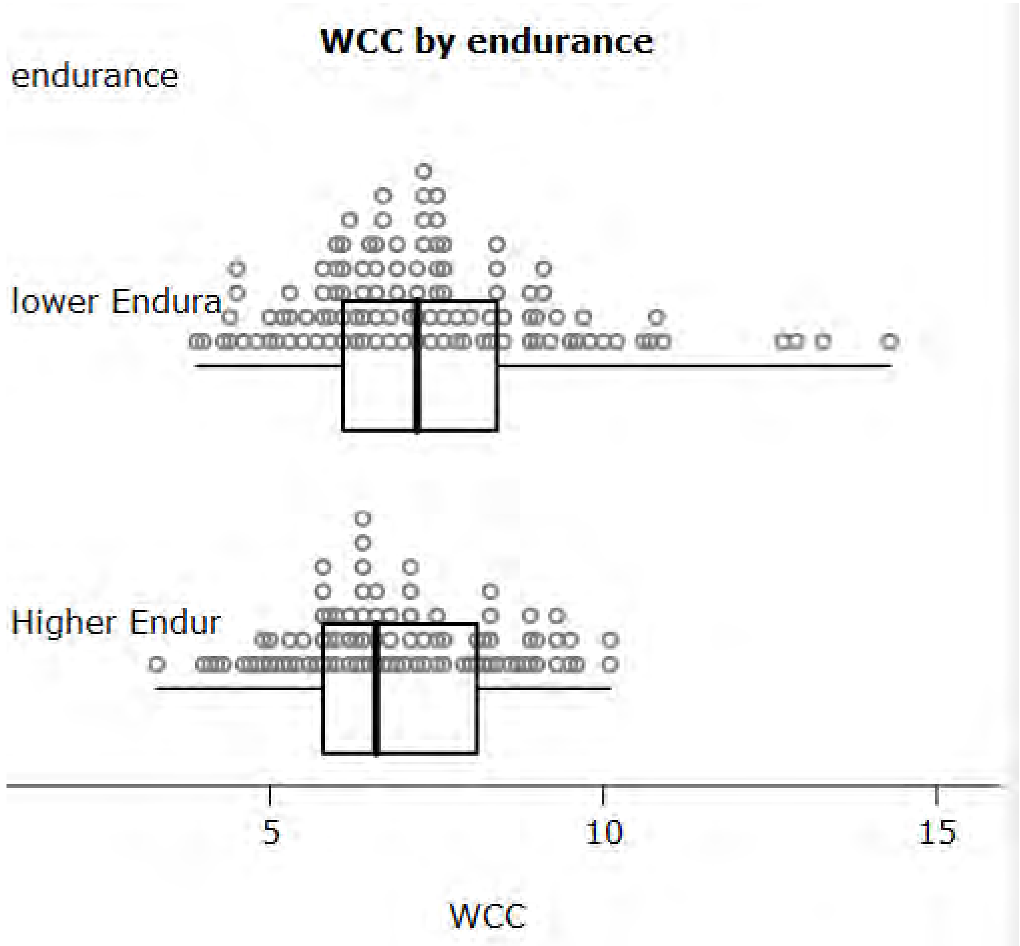
I want to investigate White blood cell count of elite athletes in different sports at the AIS. If I separated the Lower and Higher endurance sports then we can see if the blood counts are similar or different. Sheehan suggests that people who participate in higher endurance sports are more likely to have a lower white blood cell count during exercise and eventually result in poor circulation. By targeting the individuals who participate in higher endurance sports then we can help them by to get better WCC and better circulation preventing unwanted conditions and diseases.

I think that the people who participate in higher endurance sports will have lower white blood cell count as they need a higher constant supply of oxygen and white blood cell count at a higher rate, over a longer period of time. In comparison to the lower endurance sport who need more oxygen and white blood cells over a shorter period of time.

Variables to be investigate: I will investigate the White blood cell count of those surveyed from the Australian Institute of Sport (AIS.) I have added a new Column to show lower and higher endurance sports. The higher endurance sports include: Gym, Rowing, Swimming, Tennis and field. The lower endurance sports include: Basketball, water polo, netball, Tsprint and T400m.

Comparative Question: What is the difference between the median white blood count of the lower endurance sports and the median white blood cell count of the individuals who participate than higher endurance sports from the Australian institute of sport (AIS)? 1

[1] http://www.livestrong.com/article/422049-the-wbc-drops-with-exercise/
[2] http://godswaynutrition.com/disorders/poorcirculation.html
[3] http://www.sportnz.org.nz/
[4] http://www.medterms.com/script/main/art.asp?articlekey=9983



Summary of WCC by endurance

	Min.	1st Qu	Median	Mean	3rd Qu.	Max.	Std.dev	Sample. Size
Higher Endurance	3.3	5.8	6.6	6.849	8.1	10.1	1.538	93
Lower Endurance	3.9	6.1	7.2	7.330	8.4	14.3	1.978	109

Initial visual Interpretation:

My initial impression is that the median of the white blood cell count for the lower endurance sports are higher on average than the higher endurance sports. This is what I predicted. But I don't think that I could safely say that individuals that participate in higher endurance sports have lower WCC as the medians have a difference of 0.6 (about 600 white blood cells.)

Centre: The median WCC (white blood count) for the lower endurance sports group sample has a higher median of WCC by 0.6. This shows me that back in the population there may be a difference between lower and higher endurance sports. Although a median 0.6 higher does not convince me that the lower endurance sports have higher WCC.

Shape: The distribution of the lower endurance group has a unimodal distribution. It is slightly skewed to the right with 4 higher values creating a high maximum. 92.7% of the data is below 10.

In comparison the Higher endurance sports have a lower median and has a fairly unimodal shape and is slightly skewed distribution to the left. A large chunk of the data is between 6-8 WCC.

Spread: The Lower endurance sports are more evenly spread as the median cuts through roughly the middle. In comparison the higher endurance sports are more spread as the upper Quartile is stretched. This could mean that the individuals participating in lower impact sports have an even concentration of WCC generally compared to the high endurance sports where there is a range between WCC.

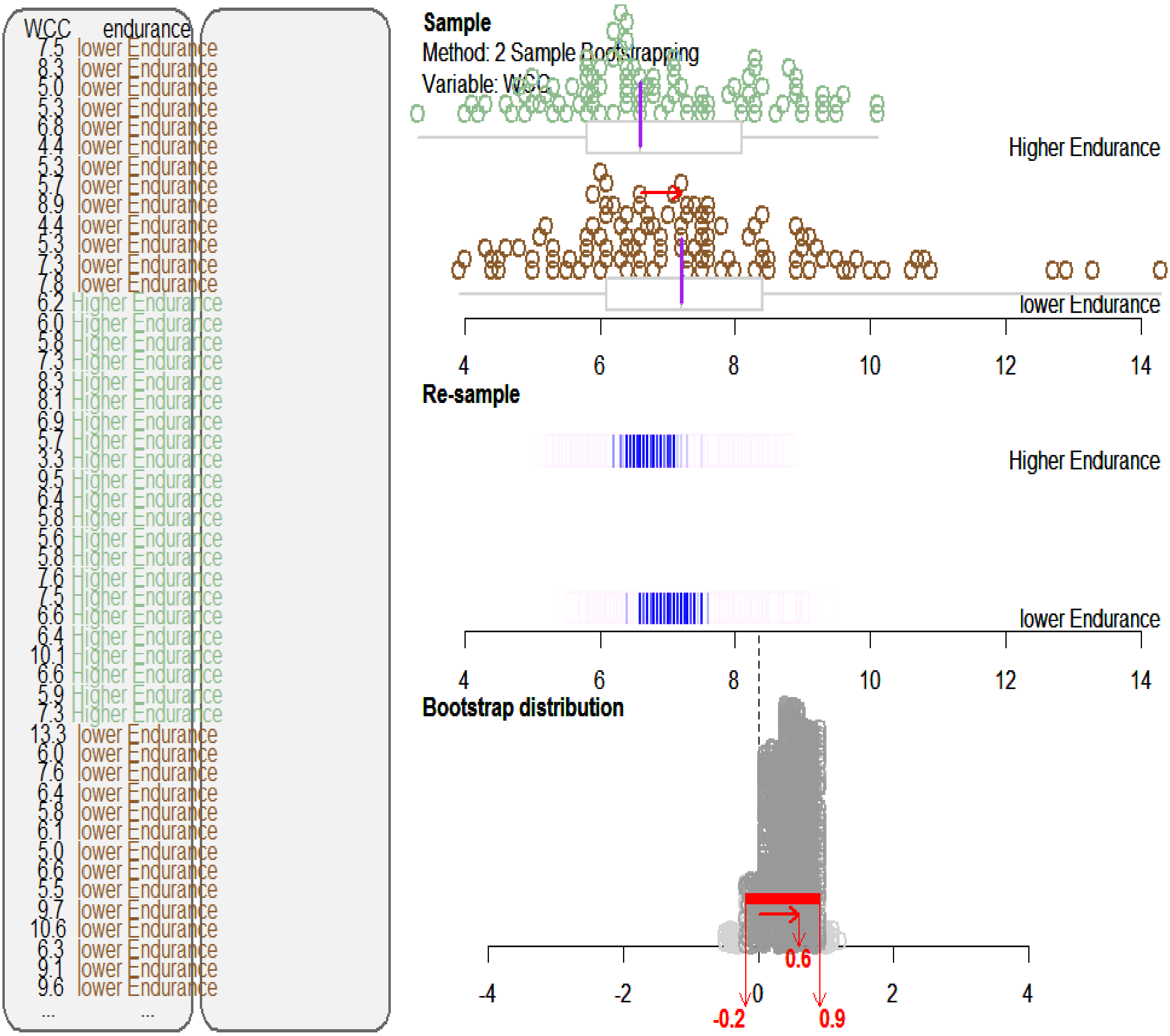
Shift/overlap: The median of the higher endurance sports WCC has shifted to the left of the lower endurance sports. There is a lot of overlap between the boxes suggesting that there may not be much difference between the WCC of high and low endurance sports back in the population. The Lower Endurance sports WCC is more even while the Higher endurance sports is clumped resulting in the upper quartile being larger than the lower quartile.

Unusual values: In the lower endurance sports there are 3 very high values. This is thought provoking. It may be because they have a very high WCC or maybe they have a very good immune system that works super efficiently. It may be possible that on person an individual exposed themselves to viruses they could help their body to have higher immunity so that when they play sport they are less likely to become sick and therefore be unable to play especially with the Olympics coming up soon.

Statistical inference:

The data suggests that it is it is unclear at this stage to say that there is a difference between the medians of white blood cell count for lower and higher endurance sports. The previous graph did show a slight difference between the medians of 0.6 (the lower endurance had a higher blood count.) This is what I had first predicted.

I am now going to use bootstrapping to find out if there is a gap between the medians of the higher and lower endurance groups. I am going to assume that the data is representative of the whole professional sports personal from the Australian institute of sport (AIS.) I will resample 1000 times from the data set to construct a bootstrap confidence level to see if there is a difference between the medians of higher and lower endurance sports.



Looking at my bootstrap confidence level I don't think I can say safely that there is a difference between WCC between the sports players from high endurance and lower endurance sports. My confidence level includes zero so this means that there is little difference between the WCC of high and lower endurance sports.

From theses samples, I cannot fairly say that there is a difference between the medians of lower and higher endurance sports as the confidence interval showed a negative difference of 0.2 (about 200 white blood cells per cubic millimetre of blood) to 0.9 (about 900 white blood cells per cubic millilitre of blood.) This shows that there is a little overlap between the medians. I don't think I could be confident that there is a difference between medians of the lower and higher endurance sports. 2

Therefore I would advise possibly (sports and recreation NZ) to do some more research on this subject to see if there is any validity in Sheehan's idea.

Conclusion: the fact that the lower endurance sports have a higher median, which I predicted. Although there was only a difference of medians of 0.6 (about 600 cells per cubic millilitre of blood.) It is very unnerving to think higher endurance sports could be harming sports men and women with WCC leading to poor circulation resulting in conditions and or disease. It would bring a massive question over exercise as exercise is greatly encouraged by health organisations, doctor, dieticians, the government and many more organisations and is greatly loved especially by kiwis.

I then used a confidence interval to measure the difference between the medians. My results showed that there was a +0.2-0.9 difference between the medians. I cannot confidently say that there is a substantial enough gap between medians to agree with Sheehan's ideals. Boot strapping relies on the fact that the sample is represented of the whole population. I have to rely on the AIS data to show a range of WCC that are true back in the population. I am sure that if this investigation was done again with different data then the result would be different because there is so little difference between the medians.

Sheehan's idea is central to the fact that WCC drops during exercise. The data is unclear whether the samples of blood were taken before, during or after exercise. This is significant to the research as I can only use the data that I have been given by the AIS. This may mean that my investigation is invalid.

	Grade Boundary: High Achieved
4.	For Achieved the student is required show evidence of using each component of the statistical enquiry cycle to make a formal inference. The student has provided evidence of each component of the statistical enquiry cycle to make a formal inference. The investigative question identifies the comparison, the parameter and the population (1). The sample distributions have been discussed and some comments (highlighted) are linked to the context. The bootstrap confidence interval has been used to make a correct call (2). For the award of Merit the analysis comments need to be more clearly linked to the context and further explanation is needed for the purpose and choice of variables.

Student 4: High Achieved

Australian Institute of Sport Body Fat Percentage Investigation

Purpose of the Investigation:

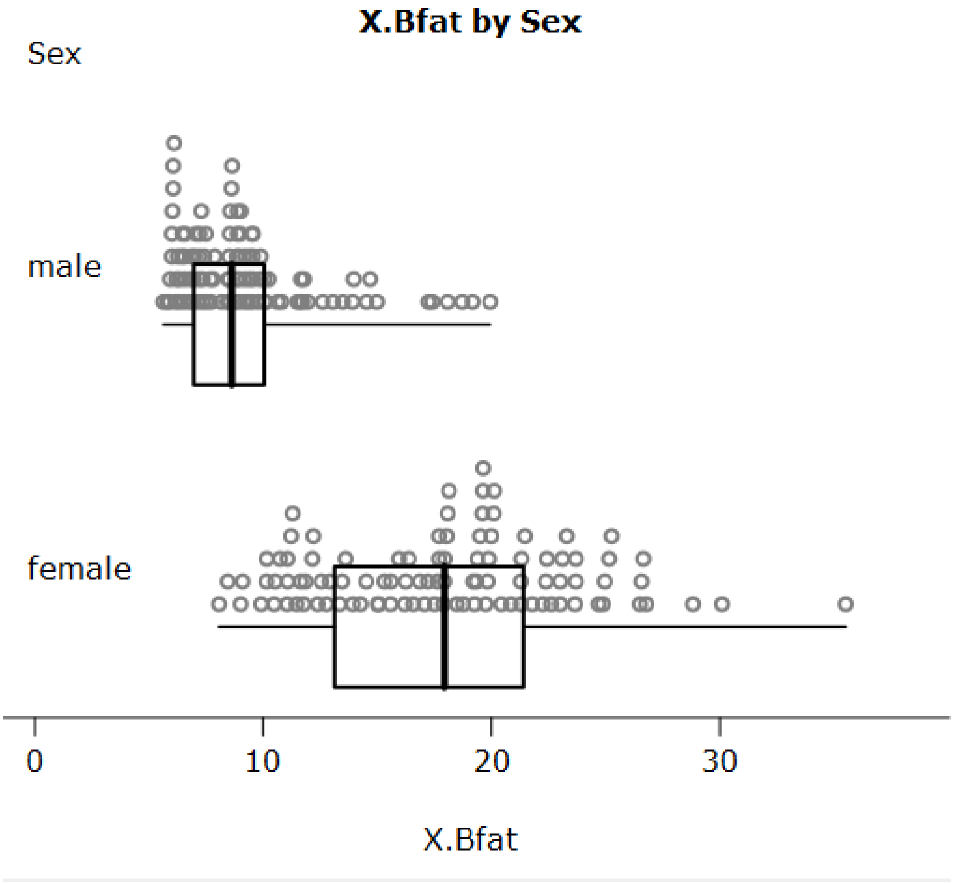
I believe that females from the 'Australian Institute of Sport' will be carrying more body fat on average in comparison to males from the same group.

Variables to be investigated:

I will investigate the percentage of body fat of those represented in the 'Australian Institute of Sport' data with both male and female individuals.

Comparative Investigative Question: 1

What is the difference between the median percentage of body fat between female and male athletes at the Australian Institute of Sport?



Summary of X.Bfat by Sex

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Std.dev	Sample.Size
female	8.07	13.240	17.940	17.850	21.36	35.52	5.453	100
male	5.63	6.967	8.625	9.251	10.01	19.94	3.185	102

Initial visual interpretation:

My initial impression is that the median for the female data in comparison to the male group is on average higher in terms of body fat percentage, which is what I would have predicted.

Centre:

The median body fat percentage for females is 9.315% higher than the median body fat percentage for males.

Shape:

The distribution for males is right-skewed bimodal with the bulk of the data ranging from 5-10% and the rest skewed off with no body fat percentage exceeding 20%. Considering the median percentage for the females is 17.940% suggests that back in the population the median body fat percentage is higher than the median percentage of males.

The distribution for the female data is different in comparison to the males. You could say that it is fairly symmetrical and would be heading towards unimodal. The data is more scattered however I would say that it occurs in chunks. The first one being between 10-13%, then we see a decrease until we get to the median of the data being 17.940%-25%. The athletes in the sample are from different sports and this could be a reason for chunks of data.

Spread:

The middle 50% of the female group is far more spread than that of the male data. The close spread of the body fat percentage for the males and the fact it skews off to the right would suggest that back in the population the female group would include a higher percentile of body fat than the male group.

Shift/Overlap

The middle 50% of the female data is to the right of the male middle 50%. We can see that there is no overlapping occurring between the two groups and it would be fair to say that the median body fat percentage for females is much higher than the male group back in the population.

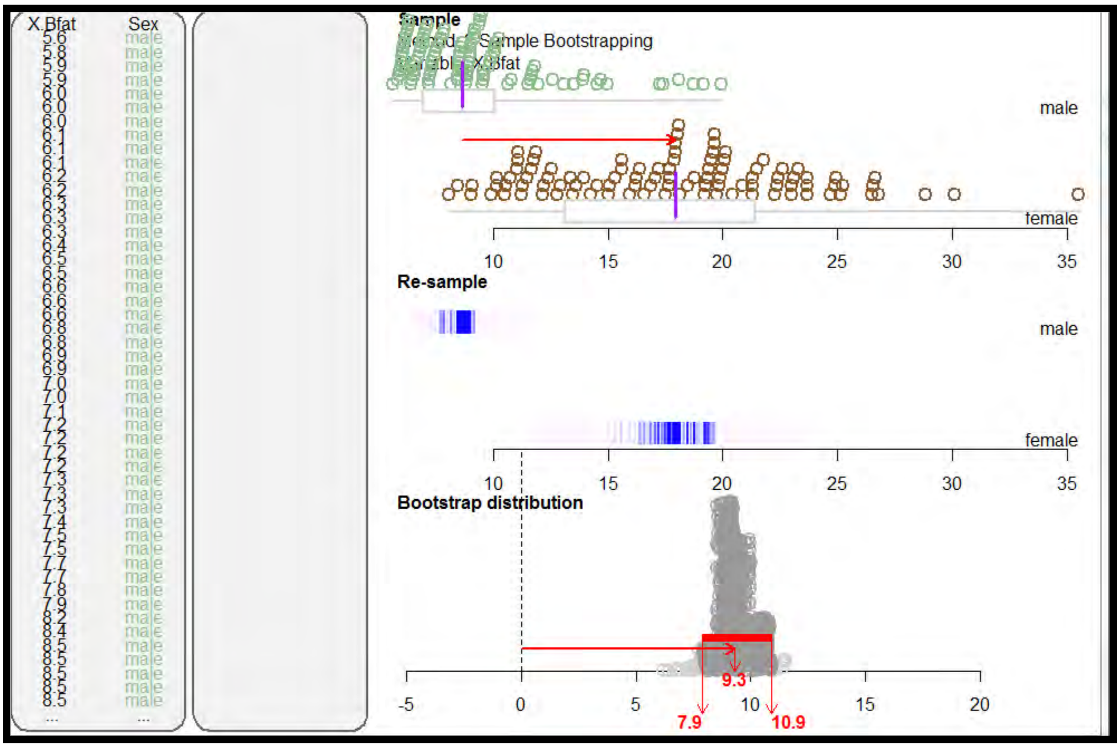
Unusual Features

There is an unusual data in the female data being the highest amount of body fat percentage being 35.52%. It makes me wonder whether a female carrying this amount of body fat would be eligible and would meet requirements for the ‘Australian Institute of Sport’ group.

Statistical Inference

From this data, it is fair to say that there is a difference in the medians for the amount of body fat percentage between the female and male individuals from the 'Australian Institute of Sport' group have. In fact, the data suggests that the median body fat percentage for the female group is higher than the median in the male group.

To get an idea of the size of the difference between the two medians of each group (female and male), I am going to assume that the data is representative of all individuals within the 'Australian Institute of Sport' and resample 1000 times from the data to construct a bootstrap confidence interval for the difference between population medians.



Looking at my bootstrap confidence interval I can fairly state that there is a difference in the amount of body fat percentage between the female and male group from the 'Australian Institute of Sport' group. This is clearly evident in the fact that my bootstrap confidence interval does not include 0. 2

From these samples, I am fairly confident that the median body fat percentage of females from the 'Australian Institute of Sport' is somewhere between 7.9% and 10.9% higher than the median body fat percentage of males from the 'Australian Institute of Sport.'

Conclusion The fact that the females carry more body fat reassures my initial prediction. I believe that this is normal seeing that the average female in general would carry more body fat than the average male.

	Grade Boundary: Low Achieved
5.	For Achieved the student is required show evidence of using each component of the statistical enquiry cycle to make a formal inference. While the variables (*% body fat, female and male*) and parameter (*median*) have been identified, the investigative question is not well expressed (1). The student has discussed the sample distributions, but there are some inaccuracies (highlighted) in the comments on shape. The bootstrap confidence interval has been used to make a formal inference about the difference in the population medians (2). For a more secure Achieved there needs to be more depth in the discussion on the distributions and the links to the context and population need to be clearer.

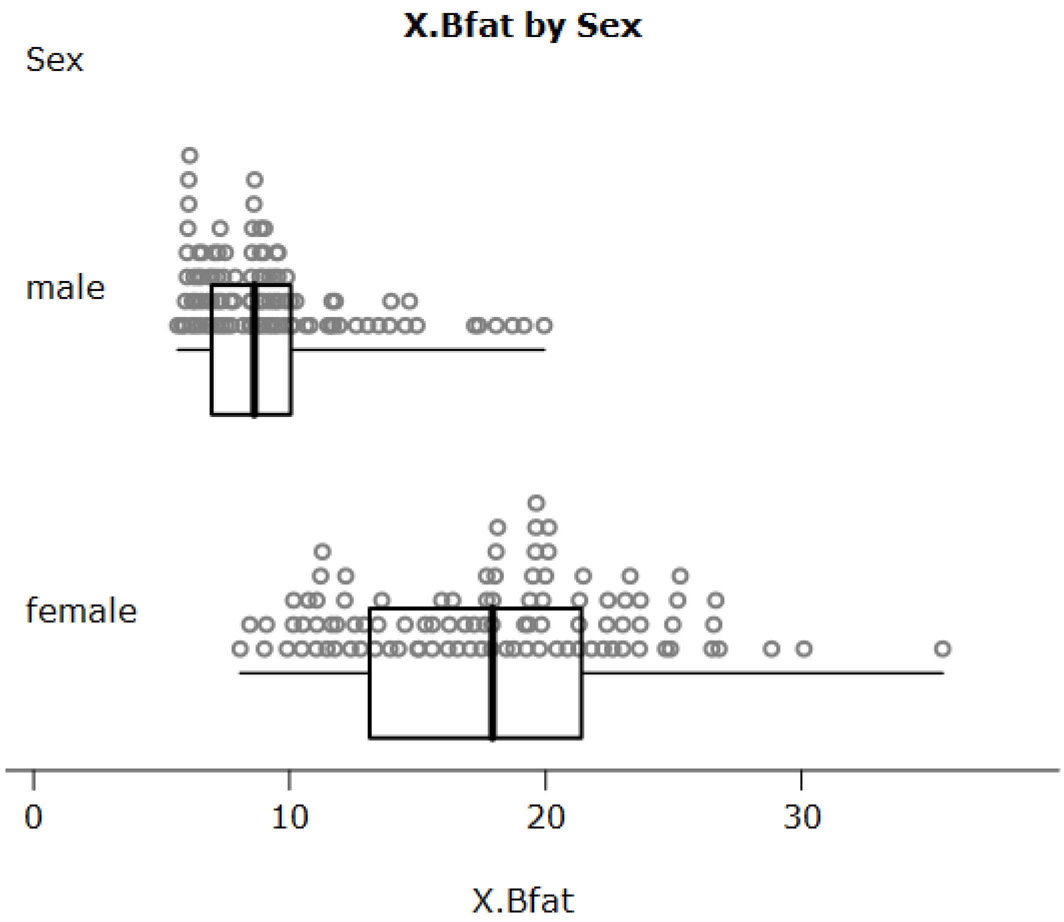
Student 5: Low Achieved

Comparative investigative question: comparing of the median % of body fat difference in female and male. 1

Variables to be investigated: in this data of the different sports from the Australian institution of sport (AIS) I will be comparing the % body fat between the female and male in the sports listed by AIS.

Summary of X.Bfat by Sex

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Std.dev	Sample.Size
female	8.07	13.240	17.940	17.850	21.36	35.52	5.453	100
male	5.63	6.967	8.625	9.251	10.01	19.94	3.185	102



Initial visual interpretation:

From this initially we can see that the males % of body fat is quite a bit lower with the box and whisker plot and the summary showing that from the median of the data is mostly lower than 10 % and concludes the most the data is clustered toward the 5-9% mark therefore we can say that men have very low % of body fat compared with the female's data. The females data is very spread out and ranging mainly from the 10-22% mark. From this initial look I can fairly say that the female % of body fat is generally higher than the male.

Centre: the median of the male data is sitting around the 8.5% mark with the middle 50% of the data ranging 7-10% and with the female data the median is sitting about 18% and the middle 50% of the data ranging from 13-22%, this is a big enough difference between the two % of body fats between female and male to make me think that there is also more % of body fat on the female than the male in these sports back in the population. The difference in the medians is 9.315% so the females % median is more than double the males.

Shape: the distribution of the % of body fat for males is showing that it is very symmetrical with most of the data being below 10% body fat considering the median of the data is 8.6%, this again suggests that it is likely that back in the population the median of the % of body fat in males is lower than the % of body fat in females. The distribution of the % of body fat for females is showing a left skew and slightly bimodal. Comparing these two, with this evidence, I state that males have a lower % of body fat and the females have a higher % of body fat so I suggest that back in the population the % of males body fat is higher than the females Because generally female are the gender that have a higher body fat % overall as males are built with more testosterone therefore more muscle is naturally in their bodies.

Spread: the middle 50% of the % of body fat of the females is more spread out than the middle of the male % of body fat.

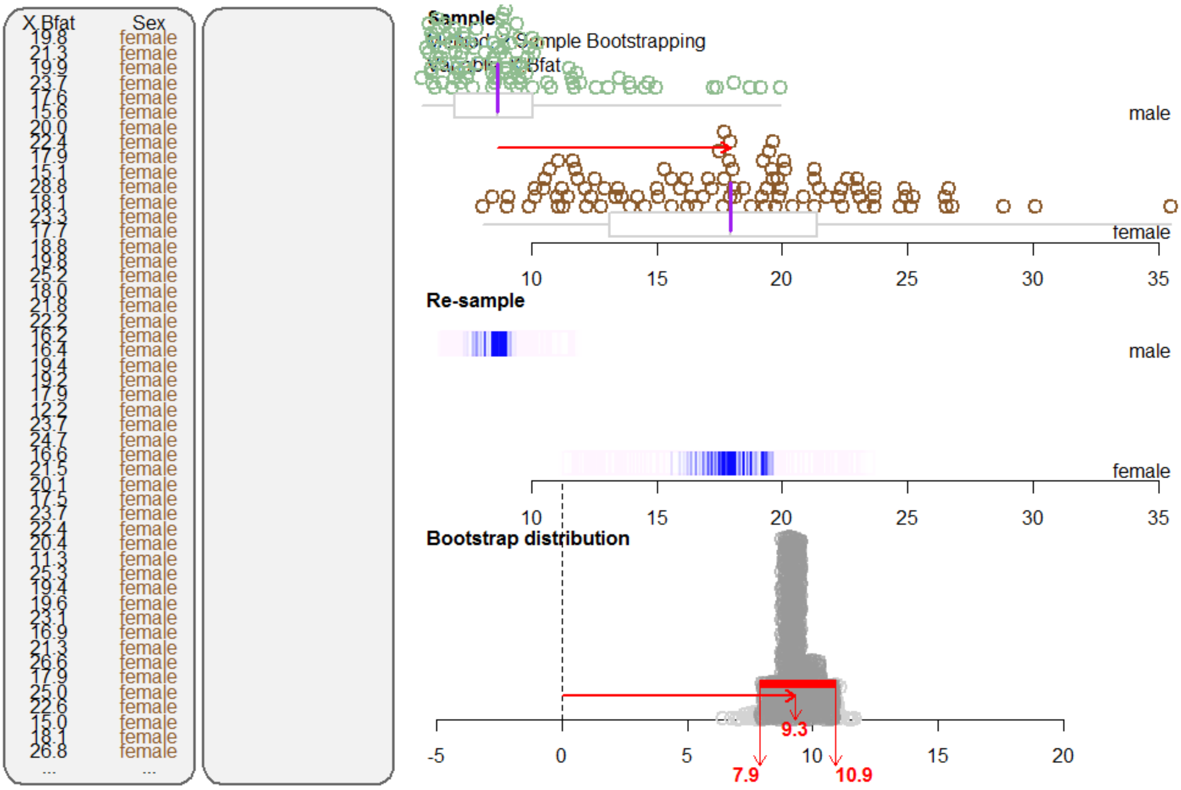
Shift/overlap: The data of the male % of body fat is shifted to the left of the female % body fat. In fact, there is no overlap with the data of the middle 50% of % body fat of males being increasingly lower than the female. With such a difference it is fair to say that the % body fat of males is greatly lower than that of the % body fat of females back in the population.

Unusual features: most of the female data is spread out but near the 19% there is more data than the rest which makes me wonder. This makes me wonder back in the population if the different sports make a difference to how much body fat is on them and also what they are putting into their bodies' nutrition wise. This would make a group worth targeting as looking at the males their big clustered group of % body fat is near the 8% mark making me notice that they have very different body builds and maybe they have different nutrition diets that could assist in the body fat difference that they are carrying in their bodies. Also makes me wonder how much the fitness levels vary in these different sports and how much it changes the outcome. With them knowing this information, it could affect the way they think to what they are giving their body and possibly to change their ways to reduce body fat on females if necessary.

Statistical reference: the data does suggest that it is fairly safe to suggest that back in the population that comparing % body fat of males and females that there is quite a large difference between them. In fact, the data suggests that the median % body fat for males is lower than the median % body fat in females, which is totally what I expected to find.

- http://www.livestrong.com/article/444398-what-is-the-normal-amount-of-body-fat-a-man-or-woman-should-have/

To get an idea of the comparing of the female and male % of body fat difference I am going to assume that it is from all over the population of % body fat of both female and male playing the variations of sport from about age 18 up to about 40 and resample the data 1000 times from the data set to construct a bootstrap confidence interval for the comparing of median % body fat between males and females in the population.



Looking at my bootstrap confidence interval I can fairly state that there is a difference in the female and male median % of body fat in the different sports from the AIS and ages 18 up to about 40. This is clearly evident that in the fact that my bootstrap confidence interval does not include 0.

2

Conclusion:

The fact that the males had a lower % of body weight definitely shows what I had predicted. But perhaps it reflects the issue of the different sports that the % of body fat varies within the two genders to how much each individual in each sport may have. If the female have more % of body fat this may just mean that over time this is how each sport pays off in how active they are etc but also in mind that generally this is how males and females' work with how they put on weight and where. Either way perhaps my initial idea of how much fitness levels they are doing and could include the nutrition they are feeding into their bodies with in my generally males eat far more than females' do. Maybe we need to educate what a good nutrition diet is for these athletes for a particular sport so that they can pick up good habits.

Bibliography:

http://www.drlenkravitz.com/Articles/genderdiffer.html

http://www.humankinetics.com/excerpts/excerpts/normal-ranges-of-body-weight-and-body-fat

	Grade Boundary: High Not Achieved
6.	For Achieved the student is required show evidence of using each component of the statistical enquiry cycle to make a formal inference. The student has posed a comparison question (highlighted), but has not clearly identified the population. There is some discussion about the sample distributions (box). The student has used resampling to construct a confidence interval for the difference in population medians (1). To reach Achieved the student needs to use the bootstrap confidence interval to make a formal inference to answer the investigative question.

Student 6: High Not Achieved

Australian Institute of Sport

Purpose

According to a sports science article, athletes can often suffer from low haemoglobin levels due to aerobic exercises, as it expands the baseline plasma volume. When this occurs, the concentration of red blood cells (which contain haemoglobin) is decreased, which can lead to sports anemia. Haemoglobin levels are considered low if they are lower than 12-13g/dL (grams per decilitre).

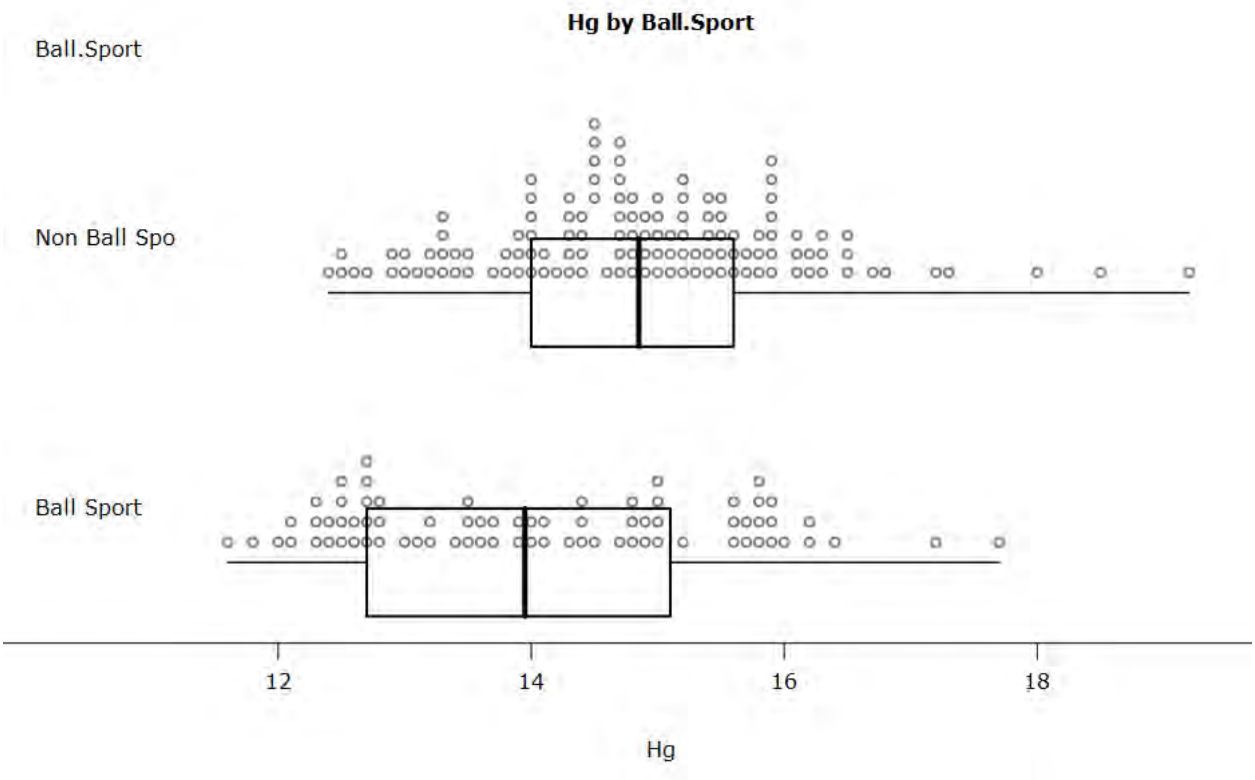
Sports anemia is common in athletes, but depending on the sports they play, their levels of haemoglobin can differ. I believe that if athletes play a ball sport their haemoglobin levels will not be as low as those who do not play a ball sport. This is because ball sports focus more on the muscles that play with the ball, while non-ball sports often focus on most of the body rather than just certain parts.

Variables to be Investigated

I am going to investigate the sports athletes play, whether they are ball sports or non-ball sports, and their haemoglobin levels. The data will come from the Australian Institute of Sport, which is Australia's premier sport training organisation.

Investigative Question

What is the difference between the median haemoglobin levels of athletes who play a ball sport and the median haemoglobin levels of those who play a non-ball sport?



Summary of Hg by Ball.Sport

	Min.	*LQ.*	*Median.*	*Mean.*	*UQ.*	*Max.*	*Std.dev*	*Sample.Size*
Ball Sport	11.6	12.70	13.95	14.06	15.05	17.7	1.445	76
Non Ball Sport	12.4	14.02	14.85	14.87	15.60	19.2	1.216	126

Describing and Comparing Sample Distributions

Initial Visual Interpretation

My initial impression was that the athletes who play non-ball sports would have lower levels of haemoglobin than those who do play ball sports because non-ball sports would be more aerobic than ball sports. However the results are different than I expected which shows that it is necessary to focus more on those who play ball sports and make sure that they have a healthy diet with good amounts of iron so they do not suffer from sport anemia.

Centre

The median haemoglobin level of those who play non-ball sports is 0.93g/dL higher than those that do play ball sports. The ball sport athlete's levels range between 11.6g/dL and 17.7g/dL, so a difference of 0.93g/dL is not a huge difference but is still significant, which leads me to thinking that there may also be a difference back in the Australian population.

Shape

The distribution of haemoglobin levels of those who play a non-ball sport is unimodal, with a lot of the data between 14.5 and 15.5g/dL which suggests that back in the population, considering the median of those who play ball sport is 13.95g/dL, those who play ball sports do have lower levels of haemoglobin.

The distribution of haemoglobin levels of those who play ball sports is bimodal with most of the data between 12 and 13g/dL at the biggest clump and then at the next clump between 15.5 and 16g/dL. Back in the population, we can assume that those who play ball sports will have lower levels of haemoglobin, as the median is 13.85, which is compatible with the largest clump. However this may vary depending on the amount of sport the athletes play along with the amount of iron they consume in their diet.

Spread

In the ball sports data, the middle 50% is mostly all together and isn't really that spread out. The middle 50% is between 14 and 16g/dL. The middle 50% of the non-ball sports data is very spread out, and the clumping is on the outskirts of the 50%. Back in the population those who play ball sport will mostly have lower levels of haemoglobin, which we can see from the spread of the data, while we can see that those who do not play ball sports have higher levels of haemoglobin.

Shift/Overlap

The non-ball sports middle 50% has shifted and is overlapping the right side of the middle 50% of the ball sports. The main clumps of both data plots are not overlapping each other. The main clump of the non-ball sports data has higher levels of haemoglobin which shows that back in the population, the results will be similar and the athletes who play ball sports will have lower haemoglobin levels.

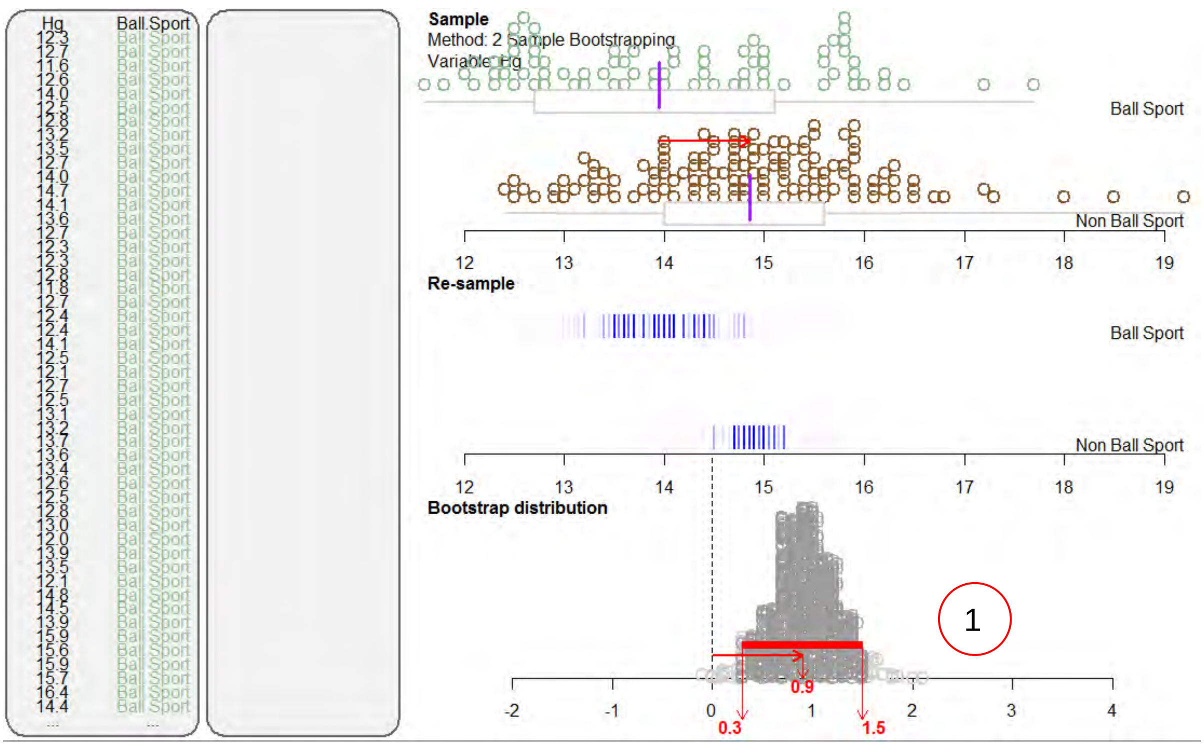
Unusual Features

Most athletes who play non-ball sports are at a healthy haemoglobin level, while athletes who play ball sports are in two main clumps. One clump is around 16g/dL which is a healthy level, but the other clump which is bigger, has most athletes around 12.5 and 13g/dL. This is still a healthy level, but they are at a minimum, which leads me to believe that this group in particular should ensure that they maintain a healthy diet that has a good amount of iron in it so that they do not let their haemoglobin levels get too low. However, some may already be eating a sufficient amount of iron, but are doing too much exercise and may need to look at cutting down the amount of sport they play.

Statistical Inference

From the data we can assume that there is a difference between the median haemoglobin levels of athletes who play ball sports and non-ball sports. There is not a huge difference between the variables, but there is still enough to show that there most likely is a difference. The data shows that athletes who play ball sports have lower levels of haemoglobin, which is the opposite of what I expected the results to be.

To show the difference of the median haemoglobin levels between ball sports and non-ball sports, I am going to assume that the data I have used represents the whole population of Australia for those who are athletes. I am going to resample 1000 times from the data and construct a bootstrap confidence interval to see the difference between the population medians.



By looking at my bootstrap confidence interval, I can see that there is a difference between the haemoglobin levels of athletes who play ball sports and those who play non-ball sports. The minimum number of those who play ball sport is 11.6g/dL, and the maximum is 17.7g/dL. While in non-ball sports the minimum is 12.4g/dL and the maximum is 19.2g/dL. It is not a huge difference, but we can see that the ball sport minimum is 0.8g/dL less than the non-ball sport minimum, and this is obviously enough to make athletes haemoglobin levels lower than what is healthy.

Conclusion

Before I did the investigation I thought that the athletes who played non-ball sports would have lower levels of haemoglobin because I thought that those sports would require exercise for the whole body. However, the data has shown that in fact those who play non-ball sports actually have a higher level of haemoglobin. If these athletes do not ensure that they maintain a healthy haemoglobin level, then they could fall beneath the ideal level and end up suffering from sport anemia. To prevent this, the Australian Institute of Sport could make sure that as they are training athletes, the athletes exercise for a healthy amount of time. They shouldn't play sport and exercise so much that they end up having unhealthy haemoglobin levels, but not so little that they become unfit and put on too much weight. The institute could also make sure that they inform and encourage their athletes to eat a diet with a healthy amount of iron, to ensure that they do not reach low levels and do not suffer from sport anemia.

From my data results, I can confidently say that athletes who play ball sports will have lower haemoglobin levels than those who do not play ball sports. Assuming that my sample reflects the entire Australian population, the bootstrapping does work. The probability that another sample of this will have different results is not very likely because there is a big enough difference to show that ball sport athletes will have lower levels of haemoglobin.

However I still need to consider my data's significance. This data is from the Australian Institute of Sport and only reflects the haemoglobin levels of athletes who train in that institute. Athletes in other organisations may have higher levels due to a different exercises that are done during warm ups etc. It is possible that the athletes who train at this particular institute do exercises that are far more aerobic than others, and that they train more. They may also have their athletes on a stricter diet which may not have as much iron in it. I have data on the sports the athletes play and on their haemoglobin levels, but I do not have data on what training they do and what their diet is. However from my data, the Australian Institute of Sport can try to ensure their athletes maintain a healthy haemoglobin level through exercise and diet, and could particularly focus on those that play a ball sport.