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

Exemplar for Internal Achievement Standard

Mathematics and Statistics Level 3

This exemplar supports assessment against:

Achievement Standard AS91582

Use statistical methods to make a formal inference

An annotated exemplar is an extract of student evidence, with a commentary, to explain key aspects of the standard. It assists teachers to make assessment judgements at the grade boundaries.

New Zealand Qualifications Authority

To support internal assessment

	Grade Boundary: Low Excellence
1.	For Excellence, the student needs 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 may include reflecting about the process; considering other relevant explanations. The student has demonstrated insight with their integration of statistical and contextual knowledge when posing a comparison question (1) and discussing the centres of the sample distributions (2). There is evidence of reflection about the process (3). This extract is from a student response which also included the student integrating statistical and contextual knowledge to an appropriate level for the award of Excellence when discussing other aspects seen in the displays of the sample distributions. For example, spread, shape, middle 50%. For a more secure Excellence, the student needs to provide more depth in the discussion in the inference/conclusion by integrating more fully the contextual knowledge they acquired both while researching the variable prior to posing the question and when explaining their findings. They could also provide further evidence considering other relevant explanations by investigating other variables, for example the number of bedrooms.

Student 1: Low Excellence

NZQA Intended for teacher use only

Context

House prices vary between New Zealand's cities and towns[1]. If a teacher is moving to New Zealand from overseas and having to choose where to take a position out of the four main centres of Wellington, Auckland, Dunedin and Christchurch from overseas house prices may influence their final choice of destination, especially if they have limited funds available to purchase a house. Having the right information available will help them to choose a centre where they can afford a house, and this will make it more possible for them to feel at home in New Zealand and not feel financial pressure. This report investigates whether there is a difference in North Island house prices compared to South Island house prices for median house sales price (NZD) in Auckland, Wellington, Christchurch and Dunedin in 2016 to help teachers choose where in New Zealand they may want to look at purchasing a house. Thus, the variables for this investigation are location by island (categorical) and 2016 house sale price (NZD). This data has come from the Real Estate Institute of New Zealand.

Question

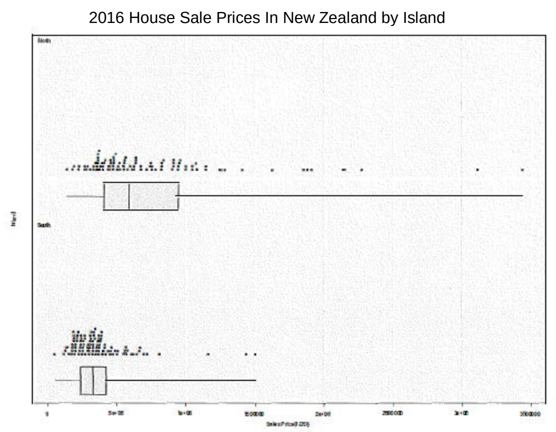
What is the difference between the median house sales price (NZD) in Auckland, Wellington, Christchurch and Dunedin in 2016 for North Island median sales price compared to South Island median sales price? 1

Hypothesis

Given that, North Island prices are higher than South Island prices, and Auckland prices are higher than prices in other cities[2] and the Auckland housing market is so 'hot'[3], it is expected this report will show the median house sales price (NZD) in Auckland, Wellington, Christchurch and Dunedin in 2016 will have a higher median sale price for North Island properties than South Island properties.

Analysis

For this data set it is appropriate to compare median rather than mean house sale price for houses sold in Auckland, Wellington, Christchurch, and Dunedin in 2016 . This is because a sample could have a small number of houses that sell for a very high price. This causes the mean to increase but would usually cause the median to change to a lesser degree thus the median is a better representation of the 'mid-point' that most of the houses are sold for. 3 REINZ also uses median figures for their reports on property sales explaining, "There are technical reasons why the median provides a more accurate picture of what is happening to the prices of houses rather than the 'average.' As an example, assume there are 11 houses sold in a month with a price range of $200,000 to $300,000 and an average price of $250,000. Now replace one of those houses with a house that sold for $1 million. The average is now $318,182, even though 10 of the 11 houses for the month sold for less than this value. The median would be the price of the middle house sold in the range (in this case the sixth house), which more accurately reflects what the majority of the houses sold for.[4]" Later in this report, Bootstrapping is used to determine the difference in median house sale prices. Median is compared here again, instead of mean, to similarly represent the sale prices of most houses in the market.



Summary Statistics

	Min	25%	Median	75%	Max	Mean	σ	IQR
North	140000	407500	585000	947000	3435000	775626	579618	539500
South	55000	239750	330000	421250	1500000	376944	229090	181500

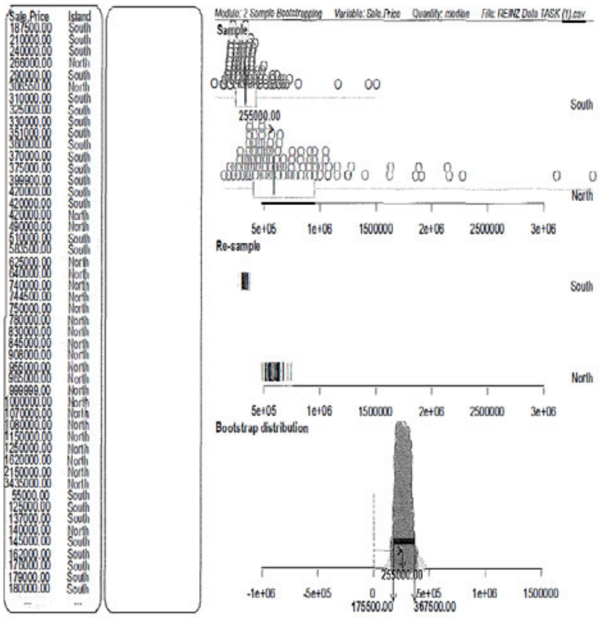
Centres

This sample shows that the median house sales price (NZD) in Auckland, Wellington, Christchurch and Dunedin in 2016 for the North Island is $585,000 NZD which is $255,000 NZD than the median sale price for the South Island at $330,000 NZD. This difference in median house sales prices is significant because a difference of $255,000 is more than 5 times the 2016 median annual income in New Zealand (the 2016 median income in New 2 Zealand is $924[5] per week; equating to an annual median income of $48048 NZD). This demonstrates that it is likely to be much harder for someone to purchase a property in the North Island. A higher median for North Island house sales prices than South Island house sales prices makes sense because of the particularly high prices in the Auckland property market. The Real Estate Institute of New Zealand (REINZ) reports that the median house sale price in Auckland is $854,500 and $537,000 in Wellington compared to $447,000 in Canterbury and $381,000 in Otago[6].

Because the median house sale price for the two North Island cities are higher than the South Island cities it makes sense that the North Island median house sale price is higher than the South Island median house sale price.

Inference and Conclusion

From this sample, we can conclude that the median house sales prices (NZD) in Auckland, Wellington, Christchurch and Dunedin in 2016 is greater for houses sold in the North Island than houses sold in the South Island. From the bootstrapping confidence interval we can be pretty sure that the difference in medians between houses sold in the North Island compared to the South Island will be between $175,500 and $367,500. Thus, we can be pretty sure that the median sale price of North Island houses will be between $175,500 and $367,500 higher than the median sale price of South Island houses. Since this confidence interval is all positive, we can make a fairly certain call that the median price of houses are in the North Island is higher than the median price of houses in the South Island for house sales prices (N) in Auckland, Wellington, Christchurch and Dunedin in 2016; at no point in the 1000 re-samples is the median house sale price higher in the South Island than the North Island.



When using a bootstrap confidence interval our statistics modelling programme resamples the data set, with replacement, 1000 times. This means that even a small sample of data can be used to create a relatively good representation of the population as long as this sample is representative of the population. For someone coming to NZ this is an efficient, low cost, effective way for her to understand the difference in property prices in selected North and South Island cities without having to gather multiple samples.

3

Due to sampling variation, if this study was repeated with a similar sized sample we would expect to have similar (but not the same) values and a confidence interval of a similar width. We would expect to still be able to make a call about the difference in medians between the two groups because our confidence interval is significantly enough above zero that we wouldn't expect the new confidence interval to shift into such a region (ie. Have both positive and negative values) where a call could not be made. Should the analysis be repeated with a larger sample we would expect to see a narrower confidence interval thus a more precise call about the difference in medians can be made.

These results are consistent with research quoted from a variety of sources throughout this report and provide compelling evidence that a teacher should look to choose a job and move to and purchase a home in Christchurch or Dunedin (South Island) rather than Auckland or Wellington (North Island); especially if they have limited resources and no other ties to any city. In reality, choosing a job and a city to live in requires many more considerations beyond the expense of houses in that city and their decision-making process will include factors such as family, climate, recreation etc.

While Sale Method may change the price that a house sells for, so too will other factors like the number of bedrooms and bathrooms, suburb, time of year, proximity to power lines etc[7]. It might be useful to investigate these other further before deciding which island to buy a house. Also, other research might be useful such as finding out how the number of migrants may influence house prices in the North Island compared to the South Island cities. This may be worthwhile looking into further as it could influence where teachers could afford to live. If there are too many migrants the house prices might be too high because of increased demand.

[1] http://www.enz.org/new-zealand-real-estate-guide.html
[2] http://www.enz.org/new-zealand-real-estate-guide.html
[3] https://www.nzherald.co.nz/business/news/article.cfm?c_id=3&objectid=11675211
[4] https://www.reinz.eo.nz/statistics
[5] http://www.stats. govt.nz/browse_for_stats/income-and-work/Income/LabourMarketStatistics
[6] https://www.reinz.eo.nz/residential-property-data-gallery
[7] http://www.rochesterrealestateblog.com/what-factors-influence-the-sale-price-of-a-home/

	Grade Boundary: High Merit
2.	For Merit, the student needs to use statistical methods to make a formal inference, with justification. This involves linking components of the statistical enquiry cycle to the context and/or to the populations, and referring to evidence such as sample statistics, data values, or features of visual displays in support of statements made. The student has researched the context and used this to develop a comparison question (1). When discussing sample distributions, the student has used evidence in support of the statements made (2) and there is evidence of linking their discussion to the context and/or population (3). The student has also made a formal inference which has been supported with evidence (4). In their conclusion, the student has linked the comments made back to findings and to the question posed (5) and commented on sampling variability including the variability of the estimates (6). To reach Excellence, the student could integrate statistical and contextual knowledge in greater depth throughout the response, for example by reflecting on whether the discussion of the sample distributions is consistent with contextual knowledge acquired through researching the context. They could also consider other relevant explanations by including a more detailed discussion of the possible reasons for survival other than the cost of the fare.

Student 2: High Merit
NZQA Intended for teacher use only

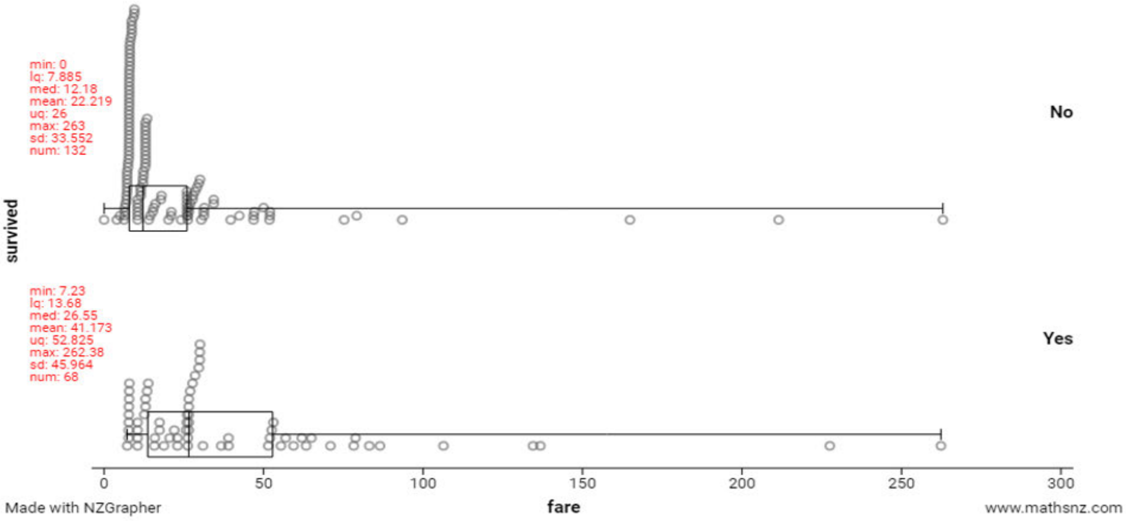
Titanic

My research shows that the median fare price that was paid by the survivors was higher than the median fare price that was paid by non-survivors. For example, my research states that a first class ticket today costs around £83200 today and the third class tickets cost around £298-£793. Of the first class riders, 60% of them had all survived. I also discovered that first-class facilities and accommodation was located on the top decks of the Titanic to avoid the vibrations and noise of the engines which were at the bottom of the ship. This also meant that first class passengers were closer to the lifeboats. Only 25% of the third class ticket buyers survived, as they were closer to the bottom of the boat and because of this they were further from the lifeboats. Also, the Titanic itself was built to gold 32 lifeboats, however, only 20 were on board at the time. One of the lifeboats had the capacity to hold 65 people, however, only 27 carried on the lifeboat, so I feel that there should have been more lifeboats on board for caution. In total there was 2228 people on board the Titanic, but only 705 survived and 1523 unfortunately perished. Using this research and data, I will be able to conclude whether or not it is statistically true that those who paid a higher price were more likely to survive.

Problem 1

I wonder what is the difference Is between the median fare of the Titanic passengers of those who survived and those who unfortunately didn't survive the horrific Titanic accident in 1912. My prediction is that the median price a person who survived paid for a ticket will be higher. The fare is the price that passengers of the Titanic paid to board the ship in British pounds.

Analysis

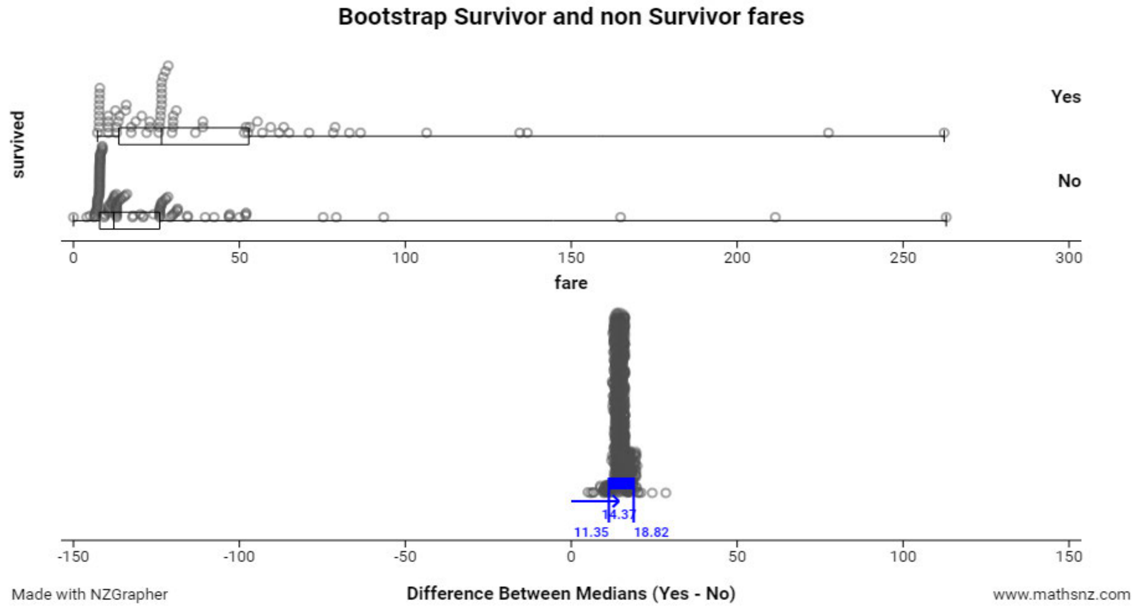


The median of the survivors is £14.37 higher than the median of the non survivors, as it is £26.55 compared to the non-survivors £12.18. I am looking at the median rather than the mean because the 2 extreme values affect what the center looks like, therefore the mean isn't used. So far this is lining up with my expectations as I expect that the people who paid a higher fare rate more likely survived. This supports my research because it showed that those who paid more were able to allow them to stay on 3 one of the higher decks of the ship, therefore increasing their chances of survival as they were further away from the danger. For example, of the third class survivors, only 25% of the people that stayed on that level survived. Whereas the percentage of people that survived that stayed in first class was 60%.

The middle 50% of the non-survivors fares are between £8 and £26, whereas the middle 50% of the 2 survivors fares are between £13.70 and £52.80 Saying this, there is quite a bit of overlap between the upper quartile of the non-survivors and the lower quartile of the survivors. So this suggests that there was quite a difference in the fares paid and that survivors on the Titanic in 1912 paid more for their 3 ticket than the non-survivors. This once again can be linked to my research as this result lined up with my expectations, as I had predicted that those who survived paid a higher median fare price. This is because if they paid more, they would have been staying in one of the rooms on the upper deck.

The interquartile range for the survivors is £39.10 whereas the interquartile range for the non-survivors is £18, indicating that the survivors have more variation in the fare price that they had paid than the non-survivors. The standard deviation is also higher for the survivors. Overall, visually the survivors seem to be slightly more spread out than the non-survivors. 2

The shapes of both graphs are different when compared to each other. The non-survivors fare paid seems to be a bimodal graph with a peak close to around £7-8 and another peak near £25. On the other hand the survivors fare paid seems to be a unimodal graph with a mode around the median of £26.55. Although it is very skewed to the right it has a shorter tail stretching from the upper quartile to the highest value than the non- survivors graph. This supports my prediction because on the survivors data there is more people that paid more for their fare, which is why they were more likely to survive. 2 3



It is a fairly safe bet using the bootstrapping graph, that the median fare that was paid by the survivors is likely to be between £11.35 and £18.82 more than the non-survivors on the Titanic in 1912. 4

Conclusion

From this information and data, and based on looking at the sample that I was given, I am reasonably confident that the median price paid by the survivors of the horrific Titanic accident in 1912 was higher than what was paid by the non-survivors. This is backed up by the bootstrapped confidence this as my confidence interval said that passengers that did survive the Titanic, median fares is likely likely to be between £11.35 and £18.82 more than the passengers, that did not survive the Titanic, median fares. This interval along with my analysis matches what I thought would happen. If I analysed another set of data, some of the points and the statistics may have differed and maybe people that paid a lower price could have survived more in that sample. However, I feel I will be able to claim this statement even with a different sample as the bootstrap confidence interval is positive. There are many other possible variables that could be considered during this investigation, such as the age and gender of the passenger. This would require a more detailed analysis. 5 6

References

http://www.dummies.com/education/history/suites-and-cabins-for-passengers-on-the-titanic/
https://en.wikipedia.org/wiki/First_class_facilities_of_the_RMS_Titani
http://Wl/llw.keyf1ux.com/titanic/facts.htm
http :/[/www](//www). eszl i nger. com/titan i c/titanfacts. html

	Grade Boundary: Low Merit
3.	For Merit, the student needs to use statistical methods to make a formal inference, with justification. This involves linking components of the statistical enquiry cycle to the context and/or to the populations, and referring to evidence such as sample statistics, data values, or features of visual displays in support of statements made. This student has researched the context and posed a comparison question (1). When discussing sample distributions, the student has used evidence to support the statements made (2) and there is some evidence of linking their discussion to the context and/or population (3). The student has also made a formal inference which has been supported with evidence (4). In their conclusion, the student has linked the comments made back to findings and to the question posed (5) and commented on sampling variability including the variability of the estimates (6). For a more secure Merit, this student could link the investigative question more clearly to research about the difference in fare prices and whether a passenger survives. In the discussion of sample distributions and in the conclusion, the student could further develop links to the context/population and provide research to support the comments made.

Student 3: Low Merit

NZQA Intended for teacher use only

The Titanic

RMS Titanic was a British passenger liner that sank in the North Atlantic Ocean in the early morning hours of 15 April 1912, after it collided with an iceberg during its maiden voyage from Southampton to New York City. From my research I know that there were an estimated 2,224 passengers and crew aboard the ship, and more than 1,500 died, making it one of the deadliest commercial peacetime maritime disasters in modern history. The RMS Titanic was the largest ship afloat at the time it entered service and was the second of three Olympic-class ocean liners operated by the White Star Line. If there were differences in survival rates because of the median fair being paid by passengers it could give us rough indications as to whether or not passengers had more or less access to emergency exits. I have read that some passengers who paid more has better access to emergency survival avenues i.e. emergency exit boats in relation to the median fair they paid to board the ship.

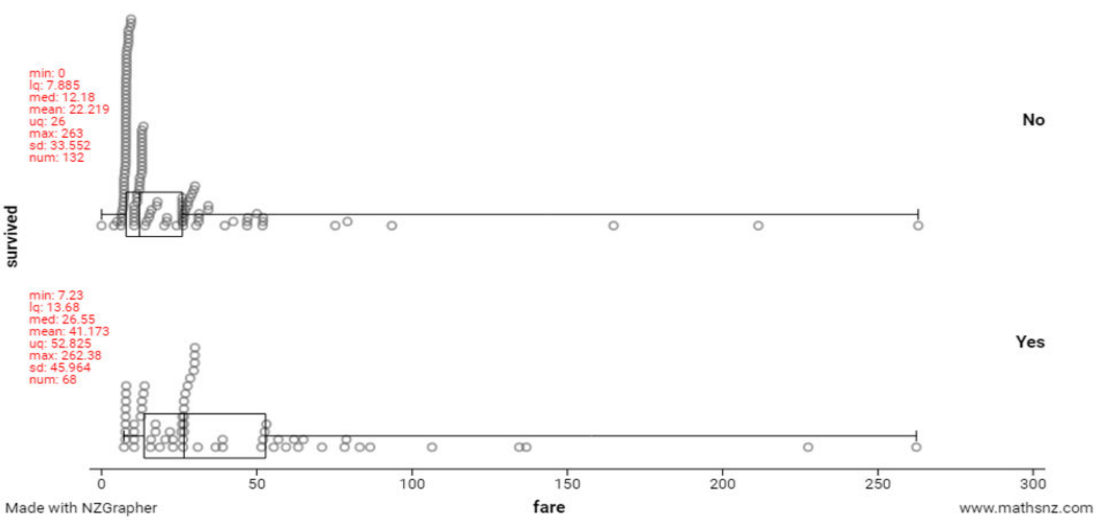
1

Question

Therefore I wonder what the difference is between the median fare of passengers, who survived and those who did not survive the Titanic disaster in 1912. I think that those who paid more would have been more likely to survive. I will use a random sample of 200 people that were on the titanic when it sank in 1912. The Numerical Variable is the fare measured in British Pounds. The Categorical Variable is the survivors and non-survivors that were on the Titanic when it sank in 1912.

Analysis

Survivor and Non Survivors fares



2

The survivors median fare is 14.37(BP) higher than the non-survivors. This is to be expected as those passengers paying more could be paying extra for easier access to emergency survival avenues i.e. emergency exit boats as stated above therefore to increased survival probability.The middle 50% of the fares for survivors are between 13.7 and 52.9 (BP) whereas the middle 50% of the fares for non-survivors is from 7.9 to 26 (BP).

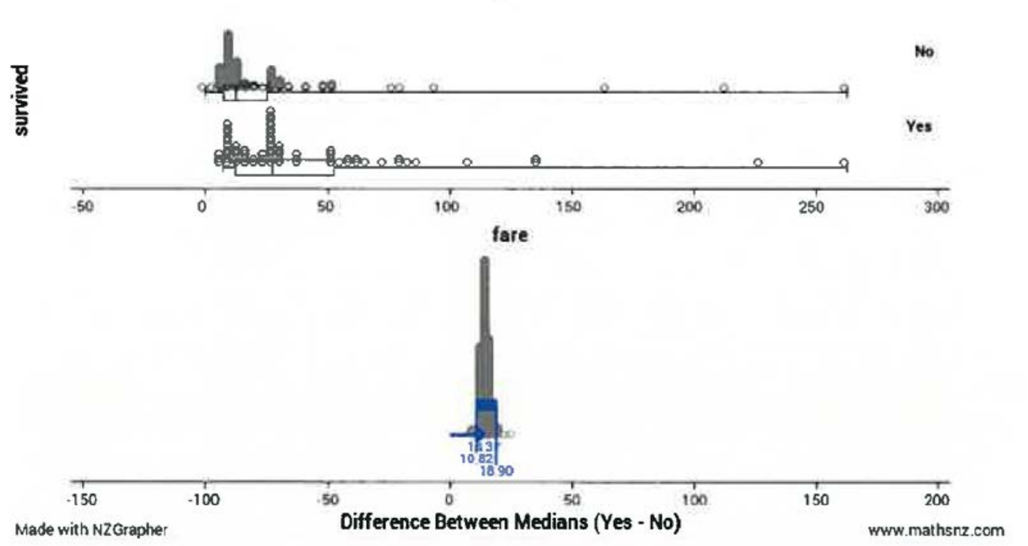
Shift I Overlap

The median fare of the survivors is higher than the Upper quartile for the non-survivors. There is an overlap where the Upper quartile, median and lower quartile fares are all higher for the survivors compared with the non-survivors. The difference between the medians is 14.37(BP) which is 0.645 (BP) of the overall spread which is a significant difference. 2

Shape The fares for both survivors and non-survivors appear to be skewed to the right
however non survivors fares seem to have a bimodal shape having two peaks whereas 2
the survivors fares are clearly unimodal (one peak)

Spread
The interquartile range for the survivors is 39.14 (BP)(whereas the interquartile range
for the non-survivors is 18.11 (BP) indicating that the survivors have more variation in 2
their fares than the non-survivors. The standard deviation is also higher for the survivors.
Overall visually the middle 50% of data for survivors seem to be more spread out than
the non-survivors. The above statements are to be expected as evidence within them
indicates that larger amounts of money paid per fair means increased probability of 3
survival to those individuals. This suggests survivors paid higher fares.

Special Features
Looking at the graphs I can see many unusual findings but the most interesting is that
the top fare paying passenger was a non survivor paying 270(BP). A close second was 2
survivor paying a fare of 262.38(BP). This is interesting yet unusual as these readings
disagree with the previous assumptions i have made about fairs and survival rates. 3



Inference
From the bootstrapping confidence interval it is fairly safe bet that survivors median fare 4
will be between 10.82(BP) and 18.9 BP) more than the non-survivors median fare on the
Titanic in 1912.

Conclusion
I am fairly sure that back in the population of all passengers boarding the Titanic in 1912 5
that survivors median fare will be more than non-survivors median fare. I can make this
call as the confidence interval says that survivors median fare is likely to between 10:82
and 18.9(BP) more than non-survivors median fare and the entire confidence interval is
Positive. I can also confidently say that there is a difference between the fare price and
the survival rates of those passengers who boarded the Titanic. I am basing this
conclusion on the bootstrap confidence interval that I calculated and via careful
analyzation of my findings. This involves resampling from my original sample of 200
people who were on board. If I were to take another sample, the results may be different 6
as that sample will contain a different makeup of passengers on the ship, but I would
expect that there would be a difference between the median fares where the median
would be higher for the survivors

1. **References -** https://www._g_oo_gle.eo.nz/search?q=titanic+wikipedia&rlz=1 C1 GGRV enN2751NZ751 &og=titanic+wikip&ags=chrome.1.69i57i015.12023j0i7 &sourceid=chrome &ie=UTF-8

	Grade Boundary: High Achieved
4.	For Achieved, the student needs to use statistical methods to make a formal inference. This involves showing evidence of using each component of the statistical enquiry cycle: posing a comparison investigative question using a given multivariate data set, selecting and using appropriate displays and summary statistics, discussing sample distributions, discussing sampling variability, including the variability of estimates, making an appropriate formal statistical inference and communicating findings in a conclusion. The student has researched the context and posed a comparison question (1). They have selected and used appropriate displays and summary statistics (2), provided evidence of discussing sample distributions (3) and linked some of their comments to statistics and visual displays (4). The student has also made an appropriate formal statistical inference (5), communicated findings in a conclusion (6) and commented on sampling variability (7). To reach Merit, the student needs to link the comments made in the analysis/inference and conclusion more clearly to the context and the actual population. They also need to include a discussion on the variability of estimates when discussing sampling variability, and further explanation about the choice of variables when developing their question.

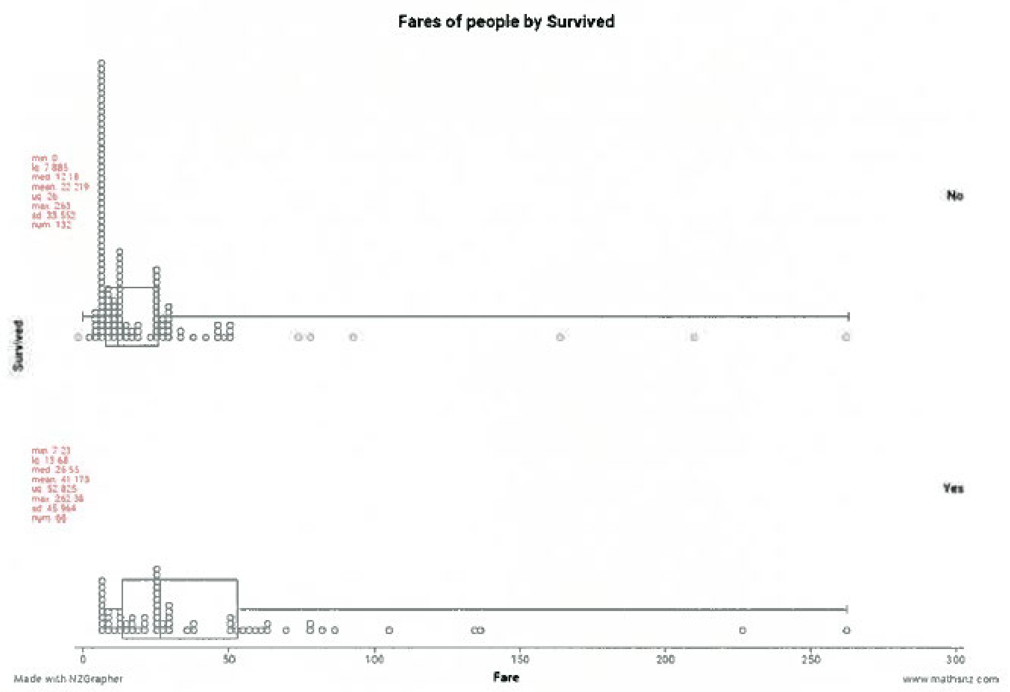
Statistical Inference Report

Student 4: High Achieved
NZQA Intended for teacher use only

According to Wikipedia when the titanic sank it only had enough lifeboats to carry about half of those on board and third-class passengers who I think would have paid way less for their fares were largely left to fend for themselves, causing many of them to become trapped below decks as the ship filled with water. In fact “54% of those in third class died”.

I wonder if there is a difference between the median fares of people that were on the Titanic in 1912, who survived and who did not, according to a sample provided from a Titanic passenger list. I am doing this investigation to discover whether the more money you payed increased a passenger's chances of survival. I think that if you were in first class on the Titanic, it would have increased your chances of survival because in Wikipedia it says that the first class passengers were closer to the lifeboats. 1

The fare is the amount a passenger on the Titanic paid to board it in British pounds (B.p), and survived is whether a passenger survived the Titanic sinking or not. The fares are different due to different class.



The passengers that survived, had a median fare that compared, with the passengers that did not survive, was 14.37 B.p greater. 3

The middle 50% of the passenger's fares, that survived the Titanic, are between 13.68 B.p and 52.83 8.p whereas the middle 50% of the passenger's fares, that didn't survive the Titanic, are between 7.89 B.p and 26.00 B.p. The passengers that survived the Titanic have a median fare that is greater than the upper quartile of the passengers that did not survive the Titanic. 4 3

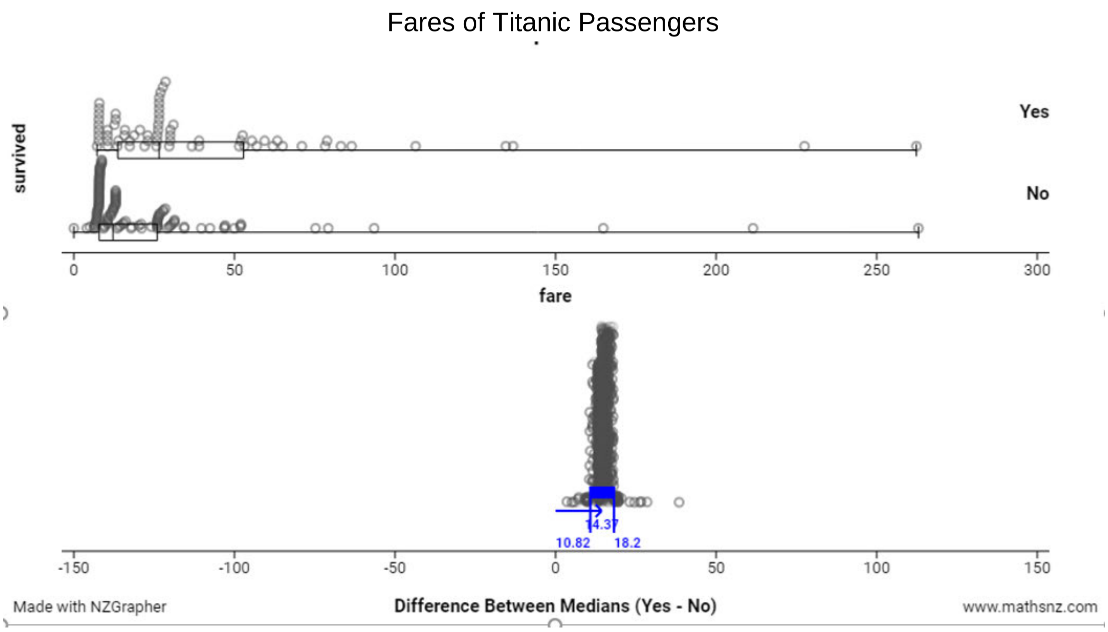
The difference between the medians is 14.37 B.p which is 0.320 of the overall visual spread, which is a significant difference.

The interquartile range for the fares of Titanic passengers that did not survive is 18.115 B.p whereas 4
the fares of Titanic passengers that did survive, have an interquartile range that is 39.145 B.p. So the interquartile range is higher for passengers who did not survive. The standard deviation for the
passengers that did survive the Titanic is also higher and overall visually, the middle 50% of fares of 3
passengers that did survive the Titanic seem to be more spread out than for the passengers that did not survive.

The fares for passengers that did not survive the Titanic appear to be skewed to the right and the
passengers that did survive also seem to be skewed to the right. The passengers that did survive the 3
Titanic appears to be slightly unimodal and the passengers that did not survive also seems to be slightly unimodal.

There is not really any special features looking at this data set.

I will now be adding a bootstrapping graph.



2

From the bootstrapping confidence interval comparing medians shows that the median fares of
passengers that did survive the Titanic disaster in 1912 was likely to be between 10.82 B.p and 18.2 5
B.p more than the median fares of passengers that did not survive the Titanic disaster.

Based on looking at my data sample of Titanic survivors and non-survivors fares, I am confident that the passengers that did survive the Titanic median fares will tend to be greater than the median fares
of passengers on the Titanic in that did not survive. I can state this as my bootstrap confidence interval 6
for the difference in the price of fares for passengers that did survive and passengers that did think it is safe to assume this as the entire confidence interval is positive. I am basing this conclusion on the confidence interval that i calculated. This involves resampling from my original sample of 200
passengers on the Titanic. I am assuming my original sample was a representative of the population 7
of all passengers on the Titanic. If I were to make another sample, the results may differ but i would still expect the same outcome of my confidence interval being positive.

Reference - https://en.wikipedia.org/wiki/RMS_Titanic

	Grade Boundary: Low Achieved
5.	For Achieved, the student needs to use statistical methods to make a formal inference. This involves showing evidence of using each component of the statistical enquiry cycle: posing a comparison investigative question using a given multivariate data set, selecting and using appropriate displays and summary statistics, discussing sample distributions, discussing sampling variability, including the variability of estimates, making an appropriate formal statistical inference and communicating findings in a conclusion. This evidence is from a student's response to the TKI task 'New Zealand crash statistics'. The student has researched the context and posed a comparison question (1). They have selected and used appropriate displays and summary statistics (2) and provided evidence of discussing sample distributions (3). The student has also commented on sampling variability, including the variability of estimates (4), made an appropriate formal statistical inference (5), and communicated findings in a conclusion (6). For a more secure Achieved, this student needs to ensure that the conclusion refers to the actual population. The student also needs to make the link between the research and the context clearer when developing their question.

Student 5: Low Achieved

NZQA Intended for teacher use only

Assessment

New Zealand Crash Data

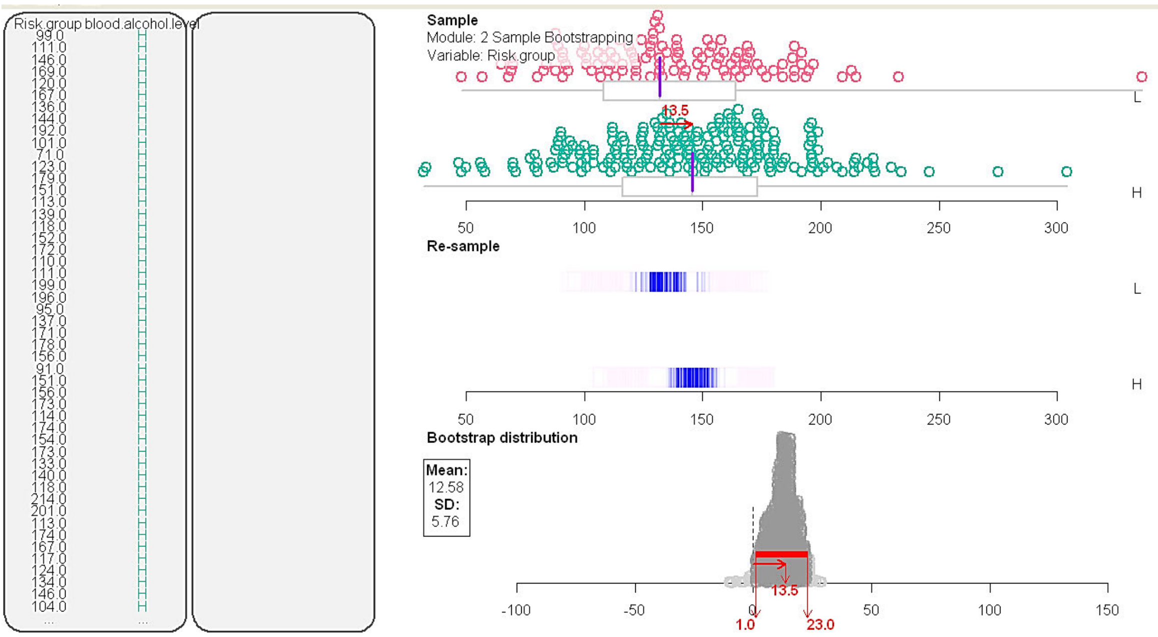
Problem:

In 2011 there were 1409 serious or minor crashes where alcohol or drugs were recorded as a factor. A random sample was taken from these drivers and they were interviewed in person by researchers. A sample of 317 drivers was taken to analyse. The variables that I will be using are different risk groups and the blood alcohol level. The Two risk groups are a high risk group which is an age group between 15-24 and a low risk group which is between the ages of 50-59.

What is the difference between the median blood alcohol levels for drivers aged 50-59 and drivers aged 15-24 in New Zealand during 2011? 1

Data:

A random sample was taken from these drivers and they were interviewed in person by researchers. A sample of 317 drivers was taken to analyse.



2

Analysis:

Overall Visual Comparisons:

My initial impression of this sample is that people aged 15-24 (high risk group) have their blood alcohol levels more spread out than people aged 50-59 (low risk group). Also there are more people in the high risk group than there are in the low risk group. You can see this clearly in the sample. 3

Centres:

The sample median blood alcohol level for people aged 50-59 (low risk group) is 13.5 milligrams of alcohol per 100 millilitres of blood lower than people aged 15-24 (high risk group). 3

I was surprised how many high risk drivers there were involved in serious and minor crashes in 2011. Out of the 317 people involved in this sample, 206 were high risk drivers and only 111 were low risk drivers.

Shape:

The distributions of both high and low risk groups are slightly right skewed, with most of the data between 0 milligrams of alcohol per 100 millilitres of blood and 150 milligrams of alcohol per 100 millilitres of blood. 3

Unusual:

In the sample the person who has the highest blood alcohol level is 336 milligrams of alcohol per 100 millilitres of blood yet they are in the low risk group. There must have been some extreme circumstances which is why they are an outlier.

Shift/Overlap:

There is large overlap between the middle 50% of the blood alcohol level for both risk groups with the median not too far away from each other. Still, this overlap is not extreme and with the medians being different I can make a clear call on what risk group has the higher blood alcohol level. 3

I will calculate a bootstrapping confidence interval anyway to confirm this.

Statistical Inference:

From the bootstrapping confidence interval you can see that the median blood alcohol level for high risk drivers in New Zealand in 2011 is somewhere between 1 milligram of alcohol per 100 millilitres of blood and 23 milligrams of alcohol per 100 millilitres of blood higher than the blood alcohol level for those in the low risk group. 5

Conclusion:

From these samples, I can make the call that there is a difference in the blood alcohol levels for people in the low risk group (people aged 50-59) and the people in the high risk group (people aged 15-24). That is, I can make the call that people in the high risk group in New Zealand have a higher blood alcohol level. 6

I am pretty sure that the median blood alcohol level for high risk drivers is somewhere between 1 milligram of alcohol per 100 millilitres of blood and 23 milligrams of alcohol per 100 millilitres of blood more than the median blood alcohol level for low risk drivers. I can make the call from these samples that there is a difference in the blood alcohol level for low and high risk drivers involved in crashes in New Zealand.

My sample did highlight that the middle section of the high risk group had more drivers involved to that of driver in the low risk group.

I am fairly confident that, if I took another sample even though the sample might be different and the initial graphs and statistics might be different, I would still make the same call as I think the difference between the medians is quite high and positive. 4

	Grade Boundary: High Not Achieved
6.	For Achieved, the student needs to use statistical methods to make a formal inference. This involves showing evidence of using each component of the statistical enquiry cycle: posing a comparison investigative question using a given multivariate data set, selecting and using appropriate displays and summary statistics, discussing sample distributions, discussing sampling variability, including the variability of estimates, making an appropriate formal statistical inference and communicating findings in a conclusion. This evidence is from a student's response to the TKI task 'Elite athletes'. The student has researched the context and posed a comparison question (1). They have selected and used appropriate displays and summary statistics (2) and provided evidence of discussing sample distributions (3). The student has also discussed sampling variability, including the variability of estimates (4). To reach Achieved, the student needs to make an appropriate formal statistical inference by using the values from the bootstrap distribution to give an interval estimate for the difference in the population medians. The student then needs to make a conclusion that goes on to answer the investigative question. This needs to include the direction of the difference.

Australian Institute of Sport

Student 6: High Not Achieved
NZQA Intended for teacher use only

Purpose

Sports anemia is common in athletes, but depending on the sports they play, their levels of haemoglobin can differ. 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).

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.

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 at the Australian Institute of Sport? 1

Analysis

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

2

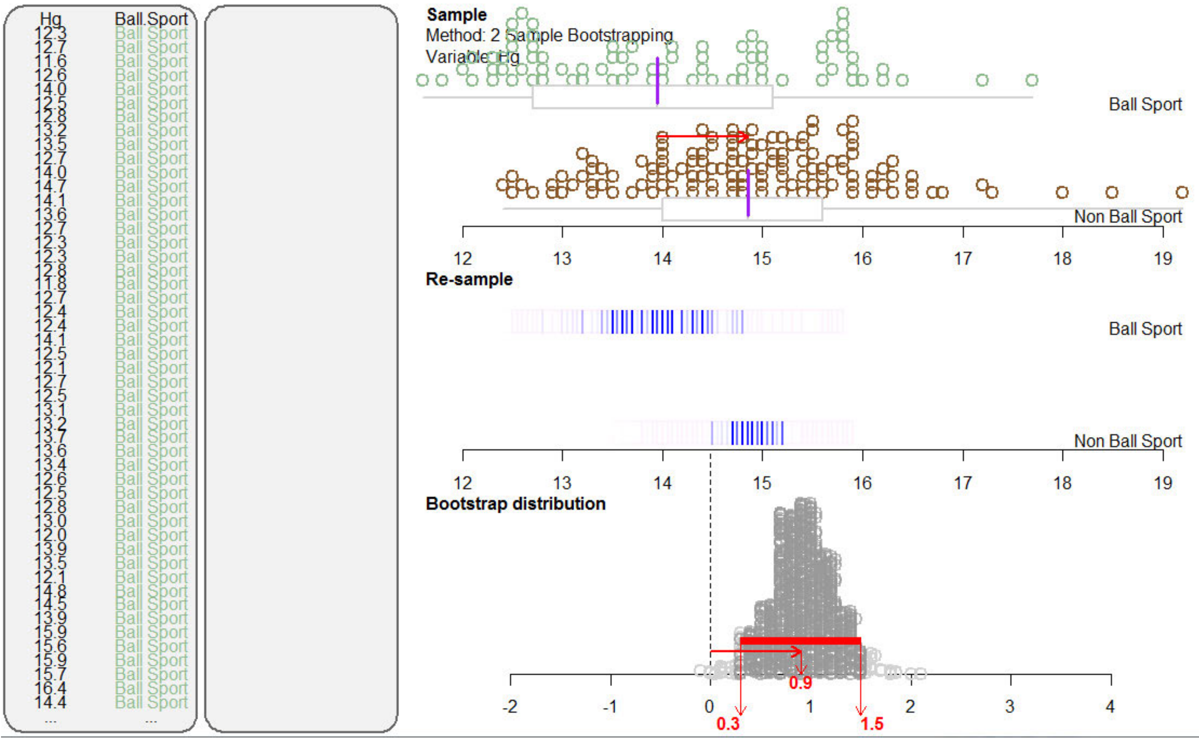
The median haemoglobin level of those who play non-ball sports is 0.93g/dL higher than those that do play ball sports. The ball sport athlete's levels range between 11.6g/dL and 17.7g/dL, so a difference of 0.93g/dL is not a huge difference but is still significant. 3

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 and 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. 3

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%. 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 he non-ball sports data has higher levels of haemoglobin. 3

From the data we can see 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 and that back in the population athletes who play ball sports have lower levels of haemoglobin, which is the opposite of what I expected the results to be.

I am going to resample 1000 times from the data and construct a bootstrap confidence interval to see the difference between the population medians.



2

Conclusion

My bootstrap has shown that those who play non-ball sports actually have a higher level of haemoglobin by between 0.3 and 1.5 g/dL. It is not a huge difference but this is obviously enough to make ball sport athletes haemoglobin levels lower than what is healthy. 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.

From my data results, I can confidently say that there is a difference in haemoglobin levels between athletes who play ball sports and 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 and graphs is very likely but because there is a big enough difference I think that it will still to show that ball sport athletes will have different levels of haemoglobin. 4