

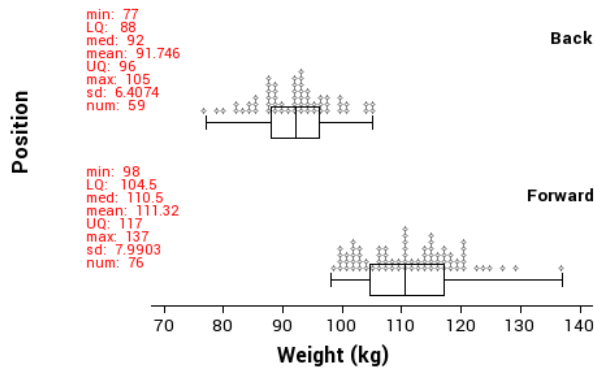
Use Statistical Methods to Make a Formal Inference

Written by J Wills – MathsNZ – jwills@mathsnz.com

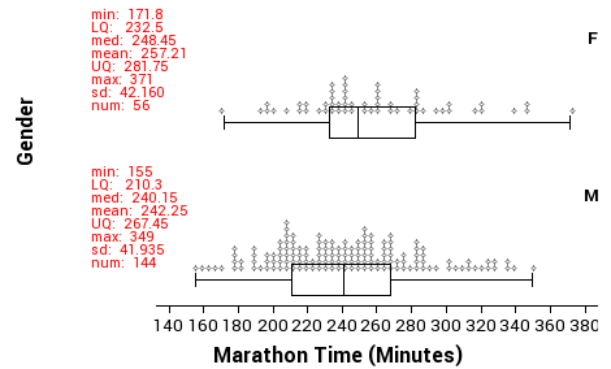
Achievement	Achievement with Merit	Achievement with Excellence
Use statistical methods to make a formal inference.	Use statistical methods to make a formal inference, with justification.	Use statistical methods to make a formal inference, with statistical insight.

Problem and Plan.....	3
Writing a Good Question.....	3
Defining the Variables	3
Sampling Variability	4
The Effect of Sample Size.....	5
Data – Using NZGrapher	6
Analysis	7
Centre – The Difference Between Medians	7
Centre – Middle 50%.....	7
Shift – Comparing the Medians and Quartiles	7
Shift – Overall Visual Spread Calculation.....	8
Spread	8
Shape.....	9
Special Features	9
Bootstrapping Activity	10
Using NZGrapher to Create a Bootstrap Confidence Interval	13
Making a Formal Inference.....	13
Writing a Conclusion.....	14
Writing Your Own Internal 1	15
Writing Your Own Internal 2	17
Sample Internal (at Achieved level)	19
Data Set Information.....	21
Assessment Guidelines – 91582 – Use Statistical Methods to Make a Formal Inference.....	24

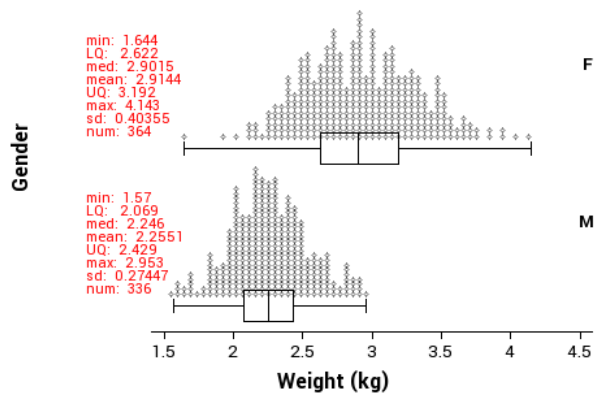
1. Rugby Players Weight by Position



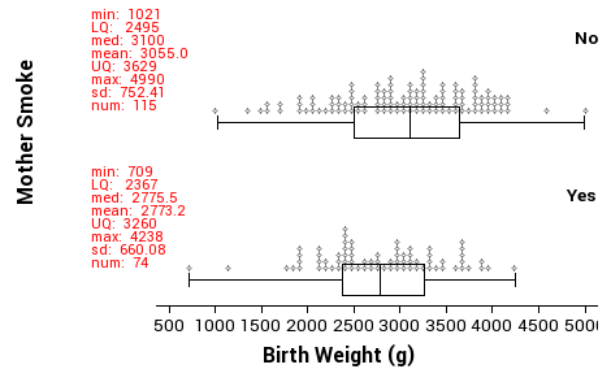
4. Marathon Times by Gender



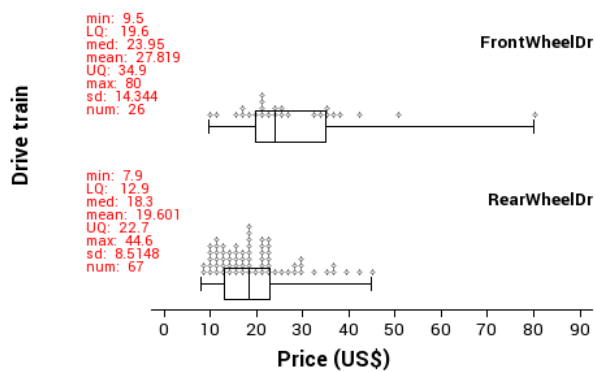
2. Weight of Kiwi Birds by Gender



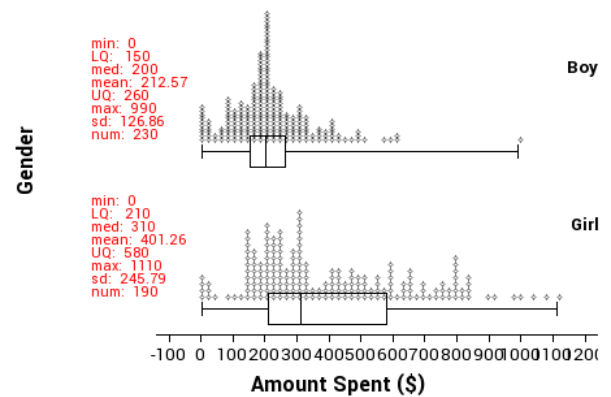
5. Birth Weight by Smoking Mother



3. Car Prices by Drive Train



6. Amount Spent on Ball by Gender



Problem and Plan

Writing a Good Question.

For each of the graphs on the previous page write a good comparative question. A question should have:

- What you are comparing (including the parameter)
- The characteristic you are grouping by
- What the population is
- Where your data is sourced from

The first one has been done for you.

To be going for Merit you need to say why you are wanting to look at these variables.

To be going for Excellence you need to use research to develop the problem.

1. I wonder what the difference is between the median weight of forward and back rugby players in New Zealand and South Africa according to a sample from <http://www.rugby-sidestep-central.com/>

Defining the Variables

The next thing that we need to do is define our variables.

Define the variables for each of the graphs on page 2.

The first one has been done as an example for you.

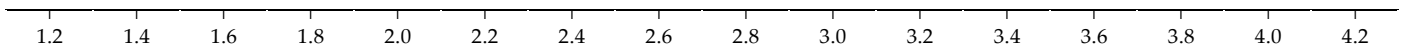
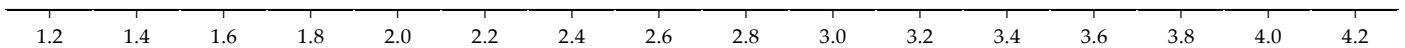
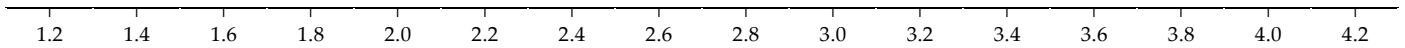
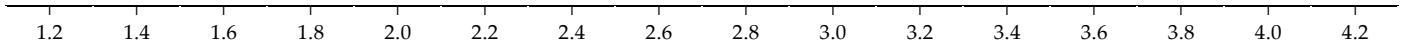
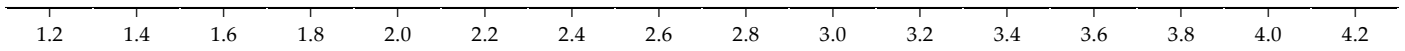
To be going for Merit or Excellence you should be using research to define the variables.

1. The weight is the weight of the rugby players in kilograms, and the position is the player's normal position on the rugby field, either forward or back.

Sampling Variability

When we take a sample there are always variations in what we choose. The more varied the data is the more varied our samples will be.

Take 5 samples of the weight of ___ kiwis using the 'Kiwi Kapers' cards, and produce a dot plot for each one on the axis below.



What do you notice as the sample size increases?

Teachers: please see the note on the dataset information on page for more information on this dataset.

Kiwi Kapers cards available from: <http://new.censusatschool.org.nz/resource/kiwi-kapers-1/>

The Effect of Sample Size

The size of our sample can also affect how reliable our sample is for predicting the population parameters. We will use the Kiwi Kapers dataset again to investigate this. Using the Sampling module of NZGrapher do 5 sampling repetitions for each sample size and record the lower quartile, median and upper quartile.

Instructions on sampling: <http://students.mathsnz.com/nzgrapher/by-graph-type/part-3-sampling-a-dataset>

Use the same sampling method (either simple random or stratified each time).

Sample Size 15

Sample	Lower Quartile	Median	Upper Quartile
1			
2			
3			
4			
5			

Sample Size 30

Sample	Lower Quartile	Median	Upper Quartile
1			
2			
3			
4			
5			

Sample Size 60

Sample	Lower Quartile	Median	Upper Quartile
1			
2			
3			
4			
5			

What do you notice as the sample size increases?

Data – Using NZGrapher

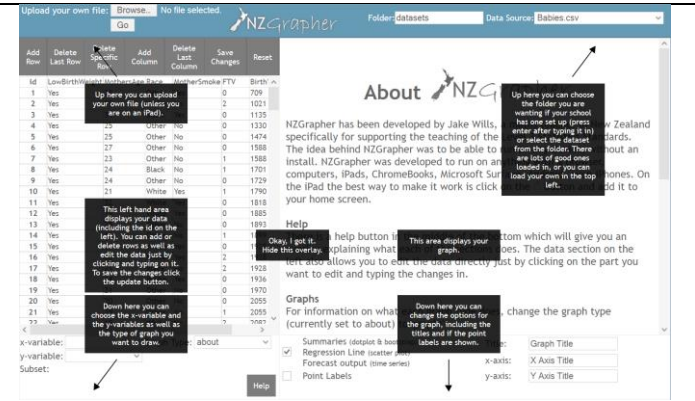
The next section that we need to do is the data section. This is reproducing the graphs on Page 2 using NZGrapher. The example below will go through using the Rugby dataset for weight by position. NZGrapher runs on anything with a browser... Macs, PCs, iPad, Android, ChromeBooks and more.

You can see a video version of this at students.mathsnz.com

First up we need to start NZGrapher by going to the link in the box to the right.

www.jake4maths.com/grapher

The first time you load NZGrapher it will display an overlay with descriptions as to what all the different areas do as shown to the right. To load your data in either select it from the dropdown in the top right, or upload it in the top left corner and press go.



To draw a dot plot there are just three things you need to do.

1. Select the x-variable... this is your numerical variable that will be on the x-axis, in this case it's weight.
2. Select the y-variable... this is your categorical variable and should have two categories, in this case it's position.
3. Select the graph type... for this we want the dotplot.

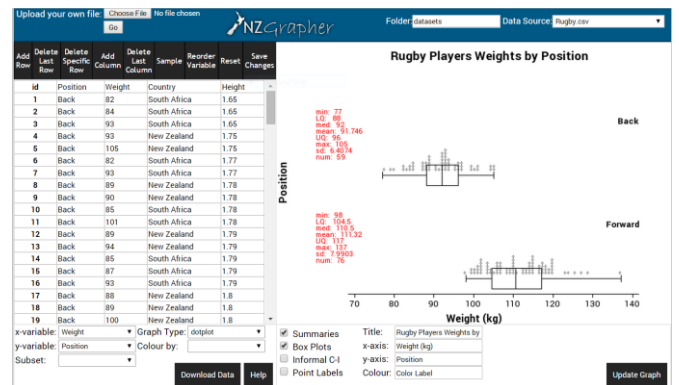
You then just need to check the graph title and axis labels to make sure they are appropriate (include units where necessary) and add press update graph to save the titles.

To save the graph just right click on it and press 'Save Image As' or whatever your device says that is similar.

Note 1: The summary statistics are automatically overlaid in red, if you want to remove them just un-tick the summary statistics box.

Note 2: sometimes you may want to only use some of the dataset... you can either delete each row you don't want in the data viewer, or open it in excel and delete the parts you do not want.

Note 3: If you want to identify the extreme points, if you click the 'Point Labels' checkbox this will add little numbers next to the points that correspond with the point id.



Now it is your turn. For each dataset you need to produce the box and whisker plot with the summary statistics overlaid.

Analysis

We now start on the Analysis section of our report. This section can be abbreviated to CSI.

The C stands for Centre, then there are 4 S's, Shift, Spread, Shape and Special Features. I stands for Inference.

Centre – The Difference Between Medians

We now need to state what the difference between the medians is. This is calculated by subtracting one median from the other.

Again the first one has been done for you.

To be going for Merit you need to say why you are looking at the median rather than the mean and state if this lines up with expectations. To be going for Excellence you need to link this to research.

1. The forwards' median weight is 18.50 kg higher than the backs' median weight.

Centre – Middle 50%

The centre is looking at what is happening with the middle 50% of the data, which is between the lower quartile (1st Qu.) and the upper quartile (3rd Qu.).

Discuss the centre for each of the sets of data, the first one has been done for you.

To be going for Merit you need to say if there is an overlap and state if this lines up with expectations.

To be going for Excellence you need to link this to research.

1. The middle 50% of the forward's weights are between 104.8 kg and 117.0 kg whereas the middle 50% of the back's weights are between 88.0 kg and 95.5 kg.

Shift – Comparing the Medians and Quartiles

With the shift we need to look at what parts of the box and whisker graphs overlap, and which parts are shifted along. You need to consider where the median and upper / lower quartiles are for the two groups of data.

Compare the medians and quartiles for each of the sets of data, the first one has been done for you.

1. The lower quartile for the forwards weight is higher than the upper quartile of the weight of the backs.

Shift – Overall Visual Spread Calculation

You also need to consider the difference in the medians (which we calculated earlier) in relation to the overall visual spread (the highest upper quartile minus the lowest lower quartile).

The calculation that you need to do is $\frac{\text{difference between medians}}{\text{overall visual spread}}$ to tell you how significant the difference is. In the example we have been working through this would be $\frac{18.5}{117.0 - 88.0} = 0.638$. The closer this number is to one the more significant the difference is.

You can use the table below to work out if the number you have calculated indicates there is a significant difference or not.

Sample Size	Difference NOT Significant	Difference IS Significant
30 – 99	0 – 0.33	0.33 – 1
100 – 999	0 – 0.20	0.20 – 1
1000 +	0 – 0.10	0.10 – 1

Discuss the shift for each of the sets of data, the first one has been done for you.

1. The difference between the medians is 18.5 kg which is 0.638 of the overall visual spread which is a significant difference.

Spread

To calculate the spread we normally look at the inter-quartile range (IQR) for the two data sets. The IQR is calculated by subtracting the lower quartile off the upper quartile. You can also look at the standard deviation for each of the two data sets. You should also comment on what you see visually.

Discuss the spread for each of the sets of data, the first one has been done for you.

1. The inter quartile range for the forwards is 12.2 kg whereas the interquartile range for the backs is 7.5 kg indicating that the forwards have more variation in their weights than the backs. The standard deviation is also higher for the forwards. Overall visually the forwards seem to be slightly more spread out than the backs.

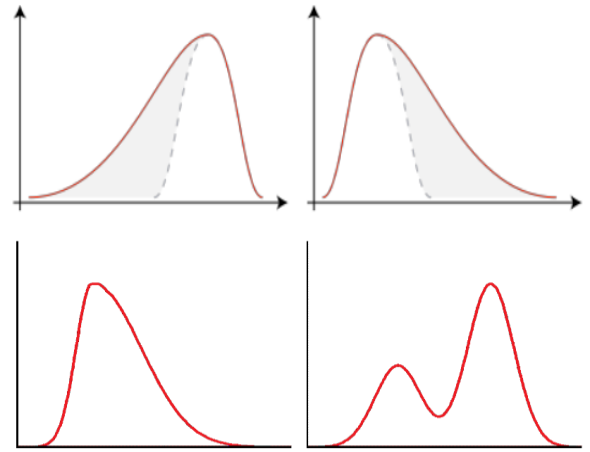
Shape

In the shape we need to look at two things... the skew and the modality.

If the distribution has a long tail to the left, it is skewed to the left (like left diagram).

If it has a long tail to the right it is skewed to the right (like right diagram).

We also need to say if there is one mode (unimodal, left diagram) or two modes (bimodal, right diagram).



Discuss the shape for each of the sets of data, the first one has been done for you.

1. The forwards weights appear to be skewed to the right whereas the backs weights seem reasonably symmetrical. The backs appear to be unimodal whereas the forwards are potentially bimodal.

Special Features

We also need to discuss any unusual features that we notice with the data sets. This could be an extreme value (a point with a much higher value than the others) or anything else that you notice. It is good to give a possible explanation for anything you notice. Going back to the original data set to find out more information about the data point is often useful as well.

Discuss the unusual for each of the sets of data, the first one has been done for you.

1. Looking at the graphs I can see that the forwards have one player that weighs more than most of the other forwards. He is a New Zealander weighing 137 kg and is 1.81 m tall. This could be because he is a stockier player that is quite large with more muscles causing him to weigh more.

Bootstrapping Activity

Bootstrapping is sampling from the sample with replacement. It normally involves sampling until you have the same number as in your original sample, but for the sake of this activity when we are doing it manually we are just going to take samples of 30 in total, which means we may end up with different numbers of forwards and backs.

Record the weights of the forwards and backs below (you won't end up filling up the whole table), and then use your calculator to work out the median for the forwards and the backs from the bootstrap, and find the difference between the two.

This activity can also be done online at: <http://www.jake4maths.com/mboot.php>

Bootstrap 1

Forwards	Backs
Med:	Med:
Difference:	

Bootstrap 2

Forwards	Backs
Med:	Med:
Difference:	

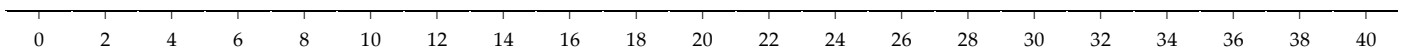
Bootstrap 3

Forwards	Backs
Med:	Med:
Difference:	

Bootstrap 4

Forwards	Backs
Med:	Med:
Difference:	

Plot the differences from both your bootstraps, as well as the bootstraps from your class as a dot plot on the axis below.



This gives us a fairly good idea of how accurate our samples are going to be, and if there is going to be a difference between the two groups (in this case the forwards' and the backs' weights). It is a very tedious process though, so we normally will use a computer to speed it up.

Bootstrapping Activity

Below is all of the rugby players from Data Set 1. You will need to cut them all out in order to do the activity on page 15. This activity can also be done online at: <http://www.jake4maths.com/mboot.php>

Back	82	Back	84	Back	88	Forward	119	Forward	102
Back	84	Back	90	Back	96	Forward	100	Forward	120
Back	93	Back	99	Back	79	Forward	109	Forward	98
Back	93	Back	83	Back	97	Forward	114	Forward	115
Back	105	Back	87	Back	101	Forward	115	Forward	99
Back	82	Back	88	Forward	116	Forward	117	Forward	100
Back	93	Back	85	Forward	120	Forward	105	Forward	103
Back	89	Back	93	Forward	102	Forward	108	Forward	110
Back	90	Back	96	Forward	110	Forward	107	Forward	115
Back	85	Back	105	Forward	137	Forward	111	Forward	103
Back	101	Back	89	Forward	102	Forward	117	Forward	115
Back	89	Back	92	Forward	112	Forward	118	Forward	124
Back	94	Back	93	Forward	103	Forward	102	Forward	110
Back	85	Back	95	Forward	123	Forward	103	Forward	116
Back	87	Back	97	Forward	114	Forward	107	Forward	99
Back	93	Back	92	Forward	115	Forward	117	Forward	101
Back	88	Back	92	Forward	116	Forward	107	Forward	110
Back	89	Back	94	Forward	118	Forward	113	Forward	110
Back	100	Back	77	Forward	125	Forward	106	Forward	106
Back	104	Back	92	Forward	102	Forward	113	Forward	106
Back	92	Back	87	Forward	120	Forward	101	Forward	112
Back	92	Back	96	Forward	101	Forward	108	Forward	114
Back	94	Back	89	Forward	104	Forward	106	Forward	114
Back	95	Back	91	Forward	107	Forward	115	Forward	117
Back	97	Back	94	Forward	109	Forward	104	Forward	120
Back	104	Back	93	Forward	118	Forward	110	Forward	119
Back	80	Back	99	Forward	127	Forward	129	Forward	120

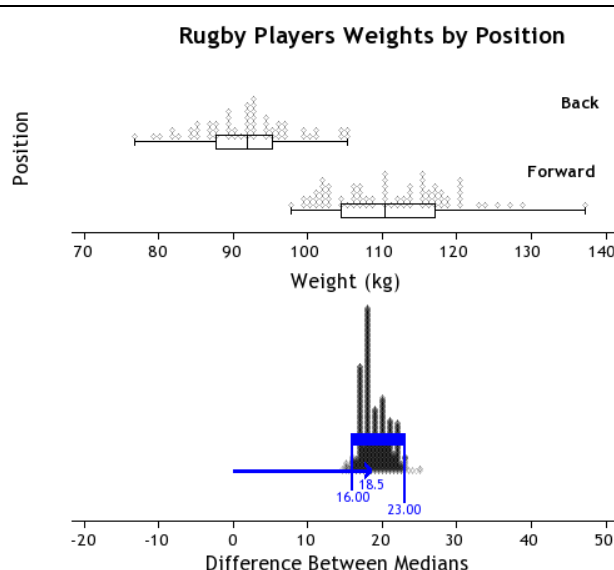
This page has been deliberately left blank (as you are cutting out the other side)

Using NZGrapher to Create a Bootstrap Confidence Interval

This part is really easy... all you need to do is change the graph type from the graph that we did earlier to a bootstrap confidence interval.

I recommend you use the median, as that is what we focus on in this booklet, but if you think the mean is better for your dataset, then use that... just justify it in your report somewhere.

This gives the output shown to the right, which tells us the difference between the medians is 18.50kg, but that we can be reasonably confident that forwards will be between 16kg and 23kgs on average heavier than the backs.



Now it is your turn. For each dataset you need to produce the bootstrap confidence interval... don't forget to press the show CI button and write down the confidence intervals so you can refer back to them later.

1. 16 kg to 23 kg
2. _____ to _____
3. _____ to _____
4. _____ to _____
5. _____ to _____
6. _____ to _____

Making a Formal Inference

We now come to the most important part of the internal, where we have been leading up to the whole time, making a formal inference. This is about linking it back to the population that we care about. To get the interval we look at the bootstrap distributions that we produced earlier.

Make a formal inference for each of the sets of data, the first one has been done for you.

1. From the bootstrapping confidence interval it is a fairly safe bet that forwards median weight will be between 16.0 kg and 23.0 kg more than backs median weight.

Writing a Conclusion

We also need to make a conclusion that summarises what we have found so far. We need to say what the call is that we are making and why we can make the call (or if we can't make the call). We can only make the call if the entire interval is positive or the entire interval is negative, as if zero is in the interval then there might be a difference of zero or the difference might be the other way round. You also need to discuss the sampling variability. For Merit and Excellence it would be good to talk about what you think the population looks like (and why) and how this will affect the sample.

Make a conclusion for each of the sets of data, the first one has been done for you. You will need to use some extra paper for the last few.

1. Based on looking at my sample I am reasonably confident that back in the population of all rugby players in New Zealand and South Africa that forwards median weight will be more than backs median weight. I can make this call as the confidence interval says that forwards median weight is likely to be between 16.0 kg and 23.0 kg more than backs median weight. I can make the call as the entire confidence interval is positive.

I am basing this conclusion on the bootstrap confidence interval that I calculated. This involves re-sampling from my original sample of 145 rugby players. I am assuming my original sample was representative of the population of all rugby players. If I were to take another sample, the results may have differed as that sample will contain a different makeup of rugby players.

Congratulations, you now have written up a report for 5 different sets of data.

Writing Your Own Internal 1

Using the framework below write a report on the diamonds data. You can use the sample internal at the end of the booklet to help you if you need it.

Diamond Testing

Title is given

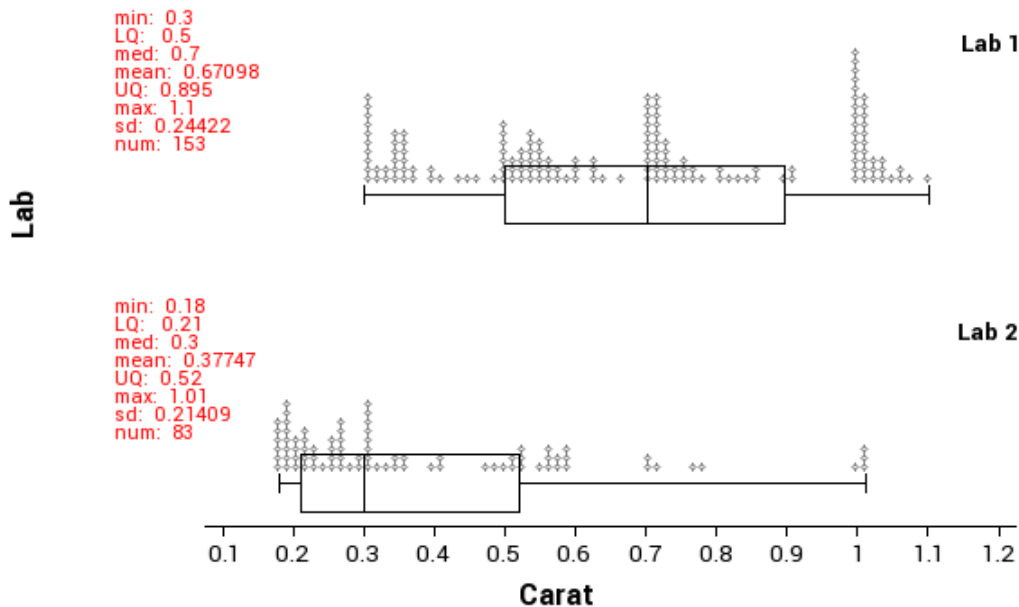
Problem

Comparative
question posed
and source
identified

Variables
Identified

Data

Diamond Carat by Lab



Dot plots and box
and whisker
plots are
produced with
summary
statistics

Analysis

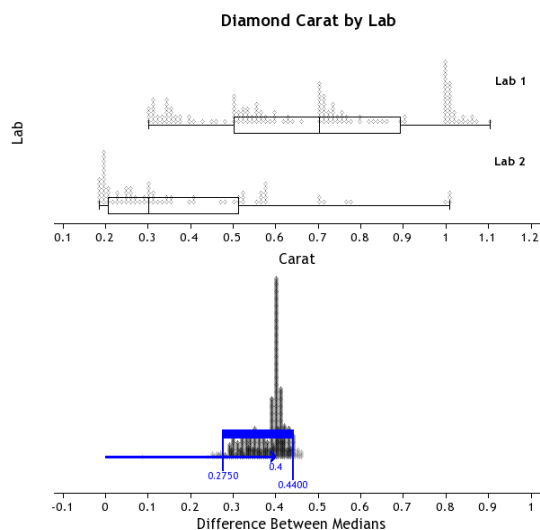
Centre

Shift / Overlap

Spread

Shape

Unusual Features
given with
possible
explanations



Conclusion

A formal
inference is made
using resampling

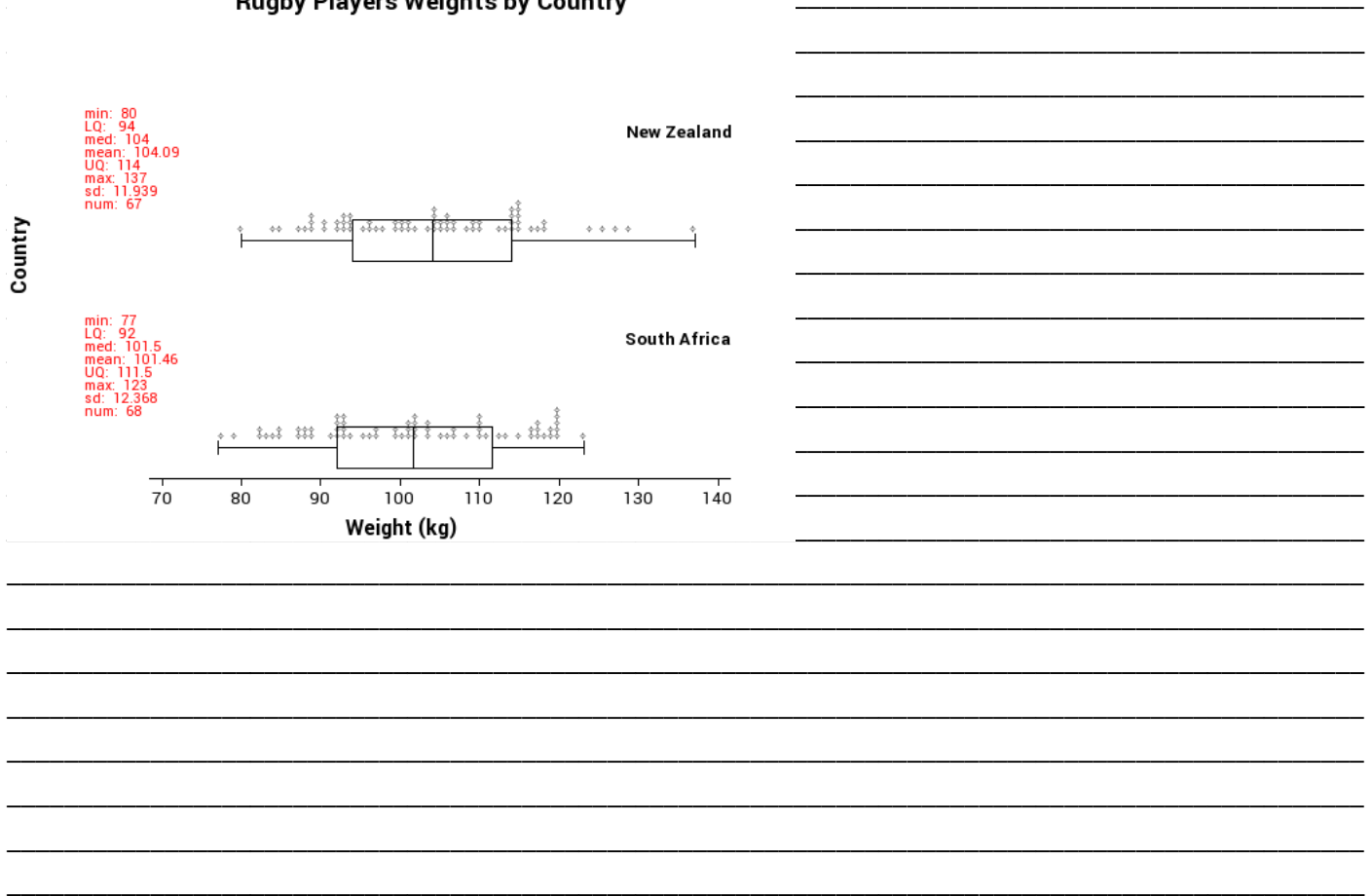
The question is
answered

and

sampling
variation is
discussed.

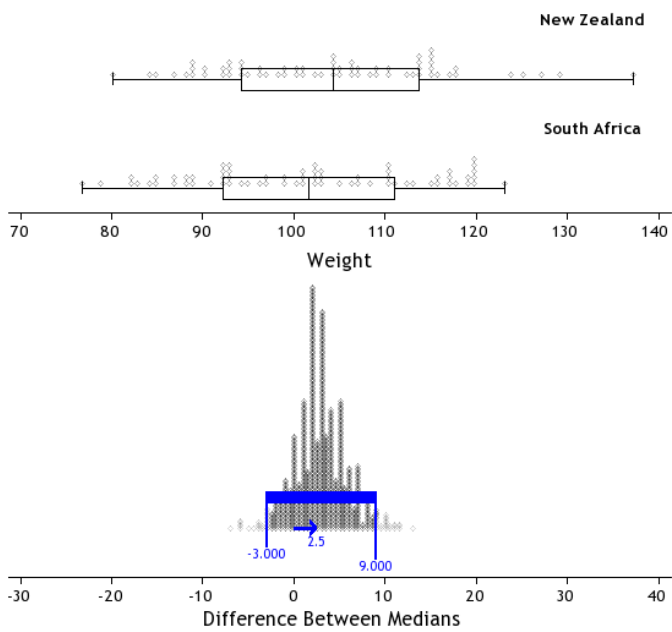
Writing Your Own Internal 2

This time you have just been provided with a title and graphs. Using these write your own internal.

[illegible]

Country

Rugby Players Weight By Contry



Sports Science and BMI

Title is given

Problem

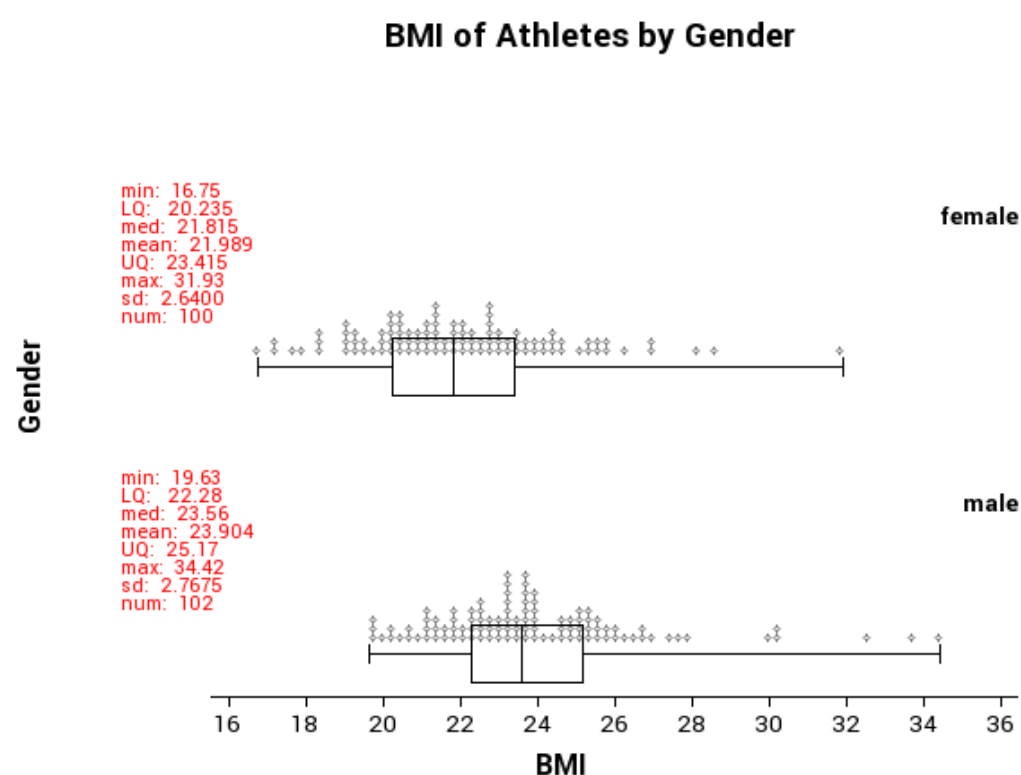
I wonder what the difference is between the median Body Mass Index (BMI) of male and female athletes in the Australian Institute of Sport (AIS) according to a sample provided from the AIS.

Comparative question posed and source identified

BMI is the Body Mass Index, and is calculated by taking the weight in kilograms and dividing by the height in meters squared. Gender is either male or female.

Variables Identified

Data



Dot plots and box and whisker plots are produced with summary statistics

Analysis

The males' median BMI is 1.74 higher than the females. The middle 50% of the data for females goes from 20.27 to 23.39 whereas the middle 50% of the data goes from 22.29 to 25.16 for the males.

Centre

The median BMI of the males is higher than the upper quartile for the females, and the median BMI for the females is lower than the lower quartile for the males. The males median is 1.74 (3sf) higher than for the females which is 0.356 of the overall visual spread indicating there might be a difference.

Shift / Overlap

The interquartile range of the BMI for the males is 2.87 whereas for the females it is 3.12. This shows that the females are more spread out than the males, however the standard deviation of the BMI for the males (2.77) is higher than for the females (2.64) indicating that the males are slightly more spread out. Overall visually there is not much difference between how spread out the males and females are.

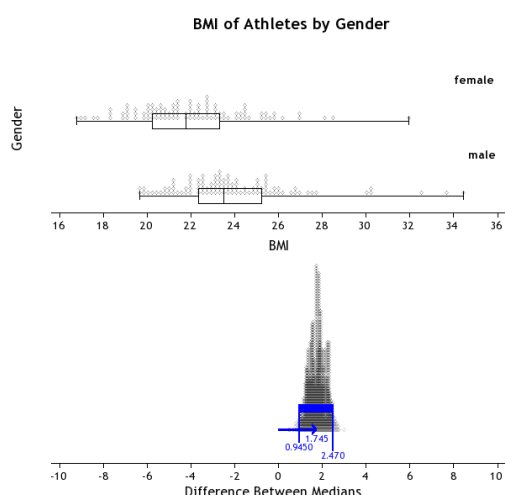
Spread

Both the males and the females BMI is skewed to the right as indicated by the longer tails to the right. Both seem to be reasonably unimodal.

Shape

The females have one data point that is significantly higher than the others. This indicates this female has a much larger BMI than the other female athletes. On looking her up in the data set I found that she is a field athlete with quite a large weight, so she quite possibly is a shot-putter.

Unusual Features given with possible explanations



From the bootstrapping confidence interval I can be reasonably confident that male athletes will have a median BMI that is between 0.97 and 2.44 higher than female athletes median BMI.

A formal inference is made using resampling

Conclusion

Based on looking at my sample it is a fairly safe bet that back in the population of all athletes at the AIS that male athletes will have a higher median BMI than female athletes. I can make this call as the confidence interval says that

The question is answered and

males are likely to have a median BMI between 0.97 and 2.44 higher than females. I can make the call as the entire confidence interval is positive.

I am basing this conclusion on the bootstrap confidence interval that I calculated. This involves re-sampling from my original sample of 202 athletes. I am assuming my original sample was representative of the population of all athletes. If I were to take another sample, the results may have differed as that sample will contain a different makeup of athletes.

sampling variation is discussed.

Data Set Information

Babies

The data on 189 births were collected at Baystate Medical Center, Springfield, Mass. during 1986.

The goal of this study was to identify risk factors associated with giving birth to a low birth weight baby (weighing less than 2500 grams). Data was collected on 189 women, 59 of which had low birth weight babies and 130 of which had normal birth weight babies.

Variable	Description
LowBirthWeight	No = Birth Weight \geq 2500g Yes = Birth Weight $<$ 2500g
MothersAge	Age of the Mother in Years
Race	Race of the mother
MotherSmoke	Smoking Status During Pregnancy
FTV	Number of Physician Visits During the First Trimester
BirthWeight	Birth Weight in Grams

BallWear

Data was recorded of students going to the school ball in 2012 as to how much they spent on their clothing and accessories.

Variable	Description
Gender	Boy = new student is male Girl = new student is female
Amount.spent	The amount spent on clothing and accessories in New Zealand Dollars.

Cars

With rising costs of owning and running a car, and environmental awareness, buyers are becoming more conscious of the features when purchasing new cars. The data supplied is for new vehicles sold in America in 1993.

Variable	Description
Vehicle Name	
Origin	Country of manufacture <ul style="list-style-type: none">AmericaForeign
Price	US \$1000
Type	Small, midsize, large, compact, sporty, van
City	Fuel efficiency in kilometres per litre in cities and on motorways
OpenRoad	Fuel efficiency in kilometres per litre on country and open roads
Drive Train	Front Wheel Drive Rear Wheel Drive
Engine Size	Size in litres
Manual Transmission	Yes No
Weight	Weight of car in Kg

Diamonds

Every diamond is unique, and there are a variety of factors which affect the price of a diamond. Insurance companies in particular are concerned that stones are valued correctly.

Data on 308 round diamond stones was collected from a Singapore based retailer of diamond jewellery, who had the stones valued.

Variable	Description
Carat	Weight of diamond stones in carat units 1 carat = 0.2 grams
Colour	Numerical value given for quality of colour ranging from 1=colourless to 6=near colourless
Clarity	Average = score 1, 2 or 3 Above average = score 4, 5 or 6
Lab	Laboratory that tested & valued the diamond 1 = laboratory 1 2 = laboratory 2
Price	Price in US dollars

Kiwi

A sample of kiwi birds around New Zealand was collected in order to help with conservation efforts. The original data is from: <http://www.kiwisforkiwi.org/> and was sourced from the secondary school guides (<http://seniorsecondary.tki.org.nz/Mathematics-and-statistics/Achievement-objectives/AOs-by-level/AO-S7-1>)

Variable	Description
Species	GS-Great Spotted NIBr-NorthIsland Brown Tok-Southern Tokoeka
Gender	M-Male F-Female
Weight(kg)	The weight of the kiwi bird in kg
Height(cm)	The height of the kiwi bird in cm
Location	NWN-North West Nelson SF-South Fiordland CW-Central Westland N-Northland EC-Eastern Canterbury E-East North Island StI-Stewart Island W-West North Island NF-North Fiordland

Teachers note: this is a synthesised dataset based on real data. At the time of creating the data set there were around 25,000 brown, 17,000 great spotted and 34,500 southern tokoeka. These numbers formed the basis of the data set, but instead of being out of around 76,000 the data set contains around 700 birds.

The data was generated using the population parameters, including gender, location, height, weight and species in Fathom. The size of the population was so that it was too big to use all the data (when doing by hand) but not too big that it couldn't be created for students to use as a "population" to sample from.

Marathon

The data is a sample taken from marathons in NZ.

It is a simple random sample of 200 athletes.

Variable	Description
Minutes	How many minutes they completed the marathon in
Gender	Male (M) or Female (F)
AgeGroup	Younger (under 40) or older (over 40)
StridlengthCM	The persons average stride length over the marathon in cm.

Rugby

The data is real data and comes from <http://www.rugby-sidestep-central.com/>

Variable	Description
Country	New Zealand or South Africa
Position	Forward or Back
Weight	The weight of the player in kilograms (kg)
Height	The height of the player in metres (m)

Assessment Guidelines – 91582 – Use Statistical Methods to Make a Formal Inference

Text in **bold** indicated a change from the previous level of achievement.

	Achieved	Merit	Excellence
Problem	The question is a comparison investigative question that clearly identifies the comparison and the population(s).	A comparison investigative question has been posed and includes an explanation for the choice of variables for the investigation.	The research is used to develop the purpose for their investigation and the contextual knowledge is used to pose a comparison investigative question.
Data	Dot plots and box and whisker plots are produced and summary statistics, including the difference between the sample medians, have been calculated. A bootstrap interval must be constructed and displayed	Dot plots and box and whisker plots are produced and summary statistics, including the difference between the sample medians, have been calculated. A bootstrap interval must be constructed and displayed	Dot plots and box and whisker plots are produced and summary statistics, including the difference between the sample medians, have been calculated. A bootstrap interval must be constructed and displayed
Analysis	The sample distributions are discussed and compared in context. This could involve comparing the shift/centre, spread, shape, and unusual features – using features of the displays and the summary statistics. A formal statistical inference is made by using resampling (bootstrapping) to construct a confidence interval.	The sample distributions are discussed and compared in context. This will involve comparing the shift/centre, spread, shape, and unusual features, with reference to features of the displays and the summary statistics and links to the population or investigative question. A formal statistical inference is made by using resampling (bootstrapping) to construct a confidence interval.	The sample distributions are discussed and compared in context. This includes seeking explanations for features of the data, which have been identified including justifying the choice of using median and considering the impact of these on the context or investigative question. Reference to knowledge from the research needs to be included in the discussion. A formal statistical inference is made by using resampling (bootstrapping) to construct a confidence interval.
Conclusion	The formal inference is used to answer the investigative question. An understanding of sampling variability is evident.	The formal inference is used to answer the investigative question, justifying the call and making links to the context. The conclusion includes an interpretation of the confidence interval. An understanding of sampling variability is evident.	The formal inference is used to answer the investigative question, justifying the call and linking back to the purpose of the investigation. The conclusion includes an interpretation of the confidence interval and a discussion of sampling variability. Findings are clearly communicated and linked to the context and populations. There is a reflection on the process or other explanations for the findings have been considered which may involve re-examining the data from a different perspective.

Final grades will be decided using professional judgement based on a holistic examination of the evidence provided against the criteria in the Achievement Standard.