Level 3 – AS91582

Standard 3.10

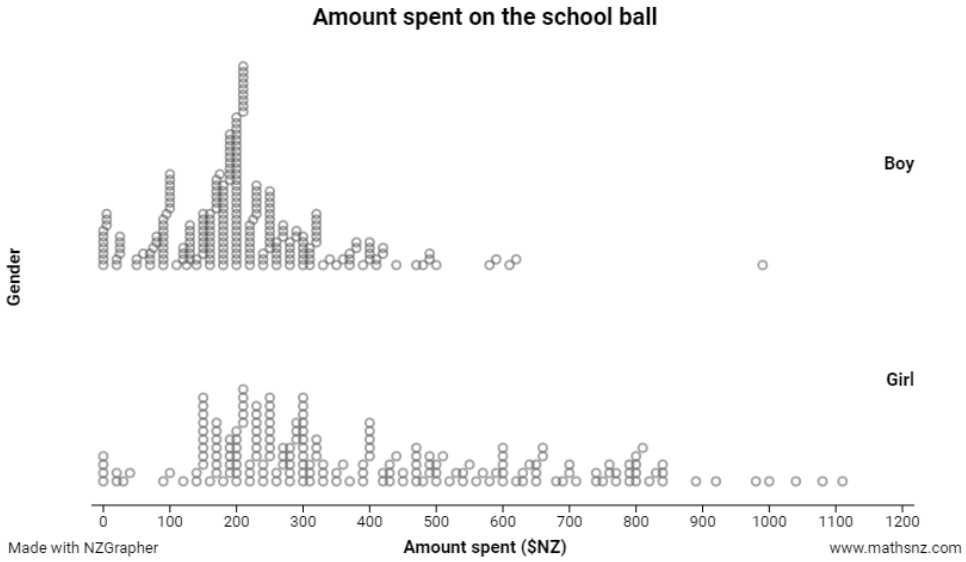
4 credits – Internal

Use statistical methods to make a formal inference

Activity: Comparing groups

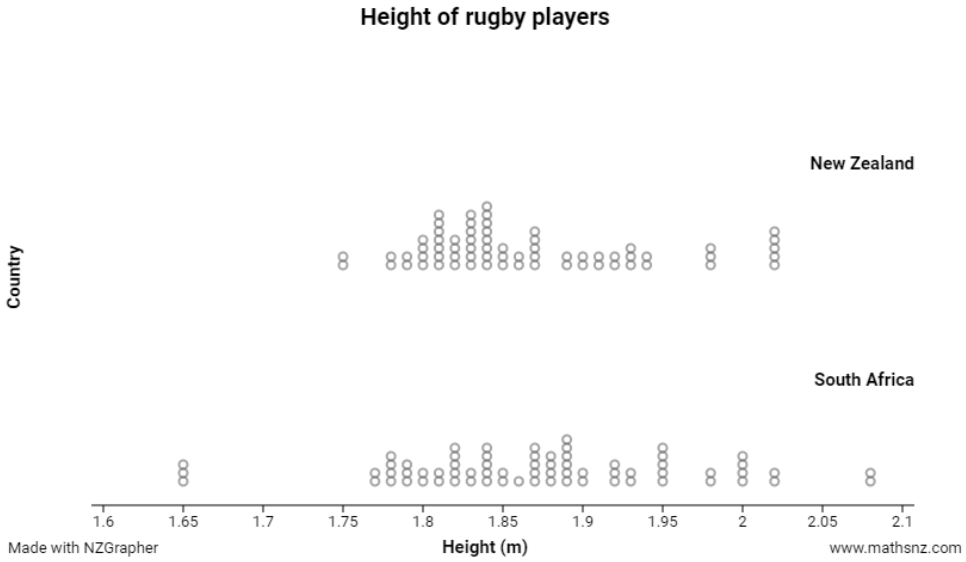
Ball spending

Here is data from 420 year 12 and 13 students going to a school ball in New Zealand in 2012 and how much they spent on their clothing and accessories. Which group, if either, spent the most on the school ball and how do you know?



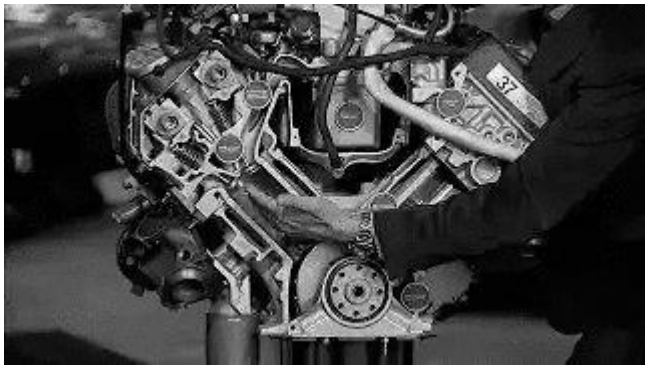
Rugby

Here is data of the heights of 135 rugby players from New Zealand and South Africa. Which country, if either, has the tallest rugby players and how do you know?

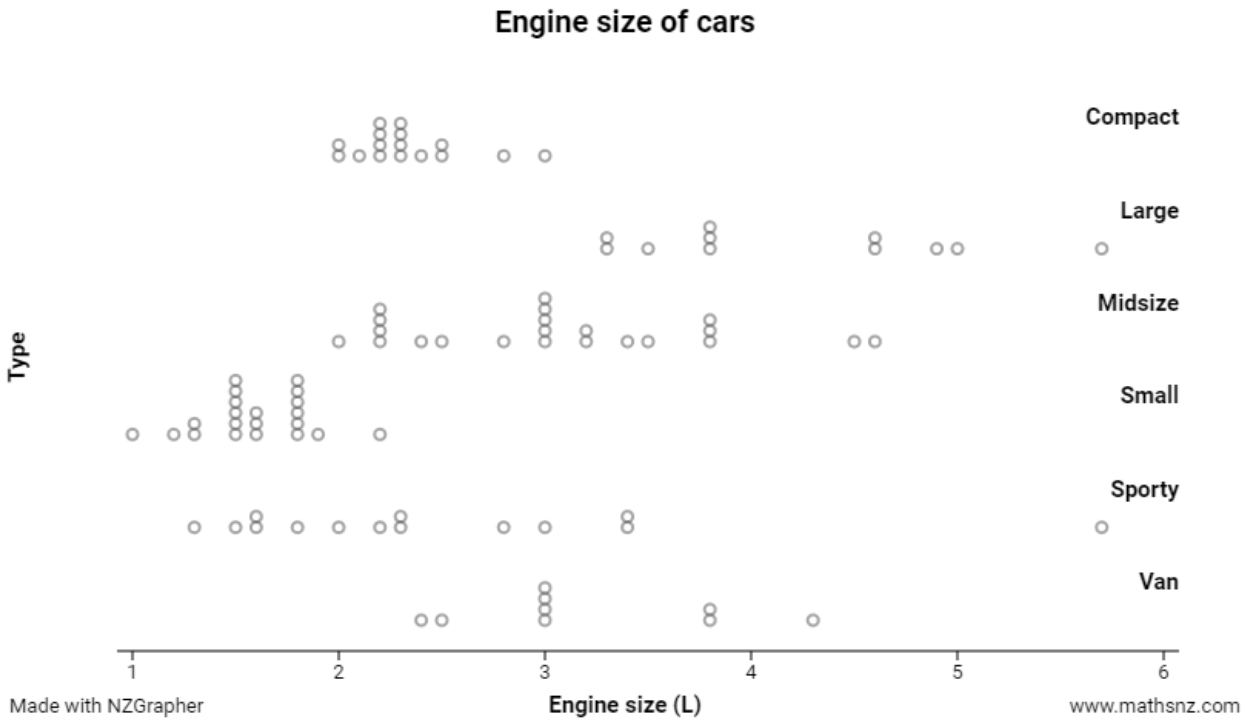


Here are two videos to introduce the idea of engine size in cars as well as data of engine sizes from 93 new car models sold in the USA in 1993. Rank the five types of cars in order from largest engines to smallest engines and explain your ordering.

Car Tech 101: Understanding engine configurations



Mazda RX8 car review - Top Gear - BBC autos



After discussing these three comparisons with your class, name the averages that you could use to compare groups and explain why averages are used to compare groups.

Activity: Introducing the dragonistics data cards

Introduction to the data cards

There is a population of dragons inhabiting an island. Each dragon has a number of attributes shown below:

Use nine data cards to fill in the table on the following page.

Then write four observations you make about the entire dragon population on the island, e.g. "red dragons tend to be stronger than green dragons on this island."

1. _____

_____

2. _____

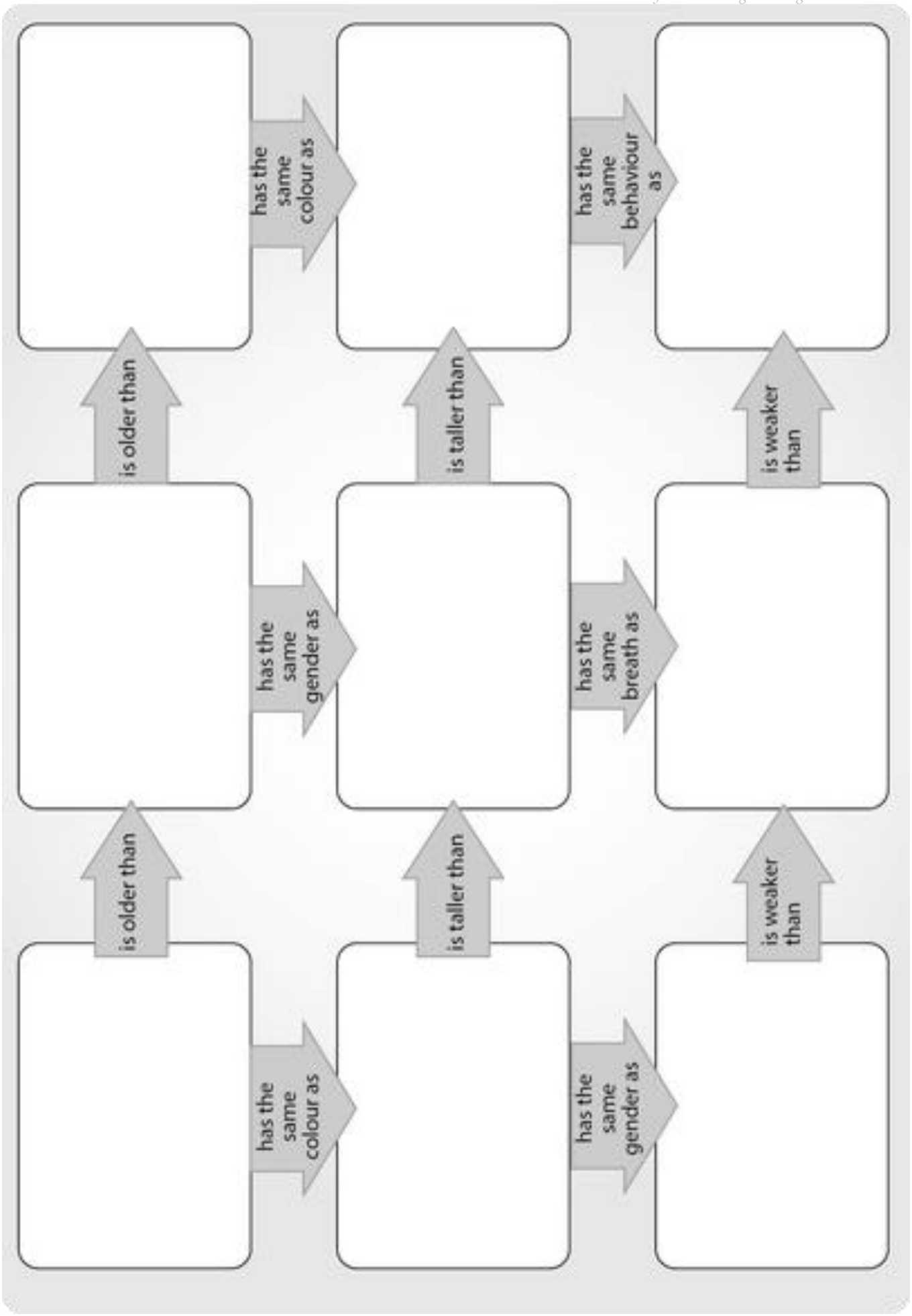
_____

3. _____

_____

4. _____

_____


is older than
has the same colour as
has the same behaviour as
is older than
has the same gender as
is taller than
has the same breath as
is weaker than
is taller than
has the same colour as
is weaker than
has the same gender as

Activity: Review of the statistical enquiry cycle (PPDAC)

Review of the statistical enquiry cycle

To review the statistical enquiry cycle, consider the following question:

On this particular island, are male dragons stronger than female dragons?

The statistical enquiry cycle is also known as PPDAC: Problem, Plan, Data, Analysis, Conclusion.

Rewrite the question so that it compares the difference between the average strength of each group.

It's often too costly or impossible to measure each element in the population so a sample is taken instead.

Write a plan to take a suitable sample. Make sure to state the sample size and sample method.

Data displays are useful to 'see' features in the data.

Enter your data into NZGrapher and generate a dot plot and box and whisker plot of your data. Print it out and stick it in the space below. Make sure to include a suitable title, suitable labels, and the summary statistics for each group. Calculate the difference between the median strength of male dragons and the median strength of female dragons *in your sample*.

The sample statistics (such as the median or mean) are just estimates of the population statistics. Therefore the difference between the sample medians is an estimate for the difference between the population median which is the original question. *"What we see is not quite the way it really is."*

To reflect this 'fuzziness', an interval (or range) of estimates is used rather than a single estimate. This is called the confidence interval.

One way to calculate an interval is to take lots of different samples, calculate their sample medians (and the difference between the sample medians of the two groups), and use this as the interval.

Write the difference between the medians for each students' sample of the dragon population.

Complete the following sentence:

The difference between the median strength of _____ dragons and _____ dragons is probably between _____ shields and _____ shields.

Taking many samples can often be costly and impractical. An alternative is to calculate an informal confidence interval (ICI) from a single sample.

$$ICI = \text{sample median} \pm 1.5 \times \frac{IQR}{\sqrt{n}}$$

For both groups in one of your samples, calculate the sample median and the interquartile range (IQR). Use these and the sample size, n, to calculate an informal confidence interval for the population median for both groups.

Complete the following sentence:

The difference between the median strength of _____ dragons and _____ dragons is probably between _____ shields and _____ shields.

Unfortunately, informal confidence intervals assume the population distribution is a certain shape. A superior method, bootstrapping, will be introduced later.

Lesson One: Investigative question (Problem and Plan)

In the "comparing groups" activity you saw that it was difficult to compare groups without using averages. In this standard, there are two averages that you could use to compare two groups.

What are those two averages?

_____

The investigative question

The investigative question is the core of your investigation. For 3.10 formal inference, you will investigate what the difference is between the average [insert variable here] of two groups.

For example:

- What is the difference between the median strength of green and red dragons on the island?
- What is the difference between the mean height of male and female dragons on the island?
- What is the difference between the median time spent on homework by year 12 and 13 students in New Zealand?

Notice that each investigative question specifies that the population, i.e. all dragons on the island or all year 12 and 13 students in New Zealand (as opposed to all year 12 and 13 students in the world).

Any investigative question needs to include:

- The numerical variable that you're investigating, e.g. strength.
- The population parameter you'll be inferring about, i.e. the median or mean. In lesson three: bootstrapping, you'll see the difference between the two.
- The groups being compared, e.g. green and red dragons.
- The population you're inferring about and from which you are sampling, e.g. all dragons on the island.

Write two more suitable investigative questions using the dragonistics data cards:

1. _____

   _____

2. _____

   _____

   _____

Additional information

<u>You also need to include at least a brief statement explaining the purpose of your investigation.</u>

There is other information that is useful to include in your introduction as well as the investigative question.

- Where the sample is from. Who collected it, for what purpose, how big is the sample, etc.
- Your hypothesis
- Define the variables.

Defining the variables

You need to clarify what each variable means. For example, if you are comparing the strength of green and red dragons, you would need to clarify that strength is measured as a score between 1 and 8. Or if you were comparing the weight of rugby players by their position, you should state: *"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."*

More additional information

The information already mentioned is sufficient to make a formal inference [achieved] but to make a formal inference with justification and/or insight [merit and excellence] you should also include the following:

- A justification for the investigation. How could people use your conclusion for a useful purpose?
- Contextual background **based on research**.
- Your hypothesis. What do you think it will be? Why do you think that?

Research

In this standard, research will be crucial to develop your contextual understanding. The context for the assessment won't be something you already know a lot about such as heights of students or school bag weights. It will be something more specialised like climate data, sports science, natural disasters, etc. Your research needs to be relevant to the context. Make sure you refer to it throughout your report to compare it to your analysis findings and your hypothesis.

Introduction exemplars

Below are exemplars of what an introduction could look like at each level of achievement.

Not achieved

The purpose of my investigation is to see if high risk groups tend to have a higher blood alcohol level than low risk groups involved in their car crashes.

Achieved

In 2011, there were 1,409 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. 317 samples were 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 years old 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 from a 2011 survey of 317 drivers in New Zealand?

Merit / Excellence

As year 13 students, many of us want to be more independent and plan to sit the first part of our driver's license as well as start our 'learning to drive campaign'. Despite being confident to begin driving amongst other vehicles, individuals between the ages of 15 – 24 are also known to often drink alcohol to excess. Unfortunately, one of the side-effects of alcohol is an impairment of rational thought, which combined with vehicle and/or peer pressure can have lethal consequences. There are many studies on the effects of drinking and driving that conclude that "car crashes are the leading cause of death among people aged 15 to 20." In New Zealand, approximately "1,900 people under the age of 21 die every year from car crashes involving underage drinking." Although a horrific statistic, my research has shown that as the age of the 'risk group' increases, the severity of the crash decreases. This could be due to the amount of alcohol in blood levels between the two age groups and is most likely a result of learned behaviour, i.e. the drinkers are much more experienced and have knowledge of the effects of alcohol on their bodies so drink relatively less than they may have done when in the higher risk age group. Alcohol is also a confidence booster and lowers peoples' perception of hazards and risk. Therefore, I suspect that the level of blood alcohol has the greatest influence on crash severity. If the blood alcohol level is higher, then the severity of the crash is likely to be greater than if the blood alcohol level was lower. I want to investigate if, in New Zealand, there is a difference between the median alcohol level per 100mL of blood present in the blood of drivers who have serious crashes and the median alcohol level present in drivers who have minor crashes.

Merit / Excellence

I wonder what the difference in median blood alcohol levels is between minor and severe crashes involving alcohol in New Zealand for the most (15 – 24 years old) and least (50 – 59 years old) likely people to crash groups. My analysis relates to recorded accidents in 2011. I am doing this investigation as I have, as a teenager, recently started driving. By the end of this year, I will be old enough to drink. It is known that these two areas do not tend to mix very well, with it generally believed that alcohol is causing the crash, but to what extent does the amount of blood alcohol impact on the severity of the crash you have? Contrary to what one would expect, a study from Cedars-Sinai Medical Center, in Los Angeles, suggests that if someone suffered a brain injury in a motor vehicle crash (MVC), the had a better chance of surviving if they had been drinking before the MVC. This is refuted by others, though, such as the Injury Research Center at the University of Michigan, who have found that alcohol can exacerbate the effects of an MVC, making things worse. While my analysis will not delve into such deep issues and as detailed conclusions, I do expect to find a higher blood alcohol level in MVCs that are more severe, and a lower blood alcohol level in MVCs that are less severe.

Activity: Sampling the dragonistics data cards

Take a sample
Take a sample of dragons from the population. It doesn't have to be a perfectly random sample but make an effort because the sample needs to be representative of the whole population. Choose a sample size between roughly 10 and 30.

In the last activity, you wrote four observations. Choose one on them and rewrite it as an appropriate investigative question.

Explain why you can't simply compare the two population medians (or means) straight away.

Estimating the population median
The best estimate of the population median is the sample median. For example, the best estimate of the median strength of all green dragons on the island is the median strength of green dragons in your sample.

Calculate the sample medians using your sample of dragons and calculate the difference between the medians.

Why can't you use the difference between the sample medians to answer your investigative question?

Construct informal confidence intervals for the population medians using the formula from 2.9 inference and interpret those intervals.

$$\text{ICI} = \text{sample median} \pm 1.5 \times \frac{IQR}{\sqrt{n}}$$

Use the informal confidence intervals to complete the following sentence and answer your investigative question.

For example:
I am pretty sure that the median height of red dragons is between 2.1m and 2.8m taller than the median height of green dragons on the island.

I am pretty sure that the median _____ of _____ dragons

is between _____ and _____ _____ than the median

_____ of _____ dragons on the island.

Note: This is similar to the process you will eventually use to answer your investigative question but not quite the same.

Activity: Using NZGrapher (Data)

Making a box and whisker plot

Step One: Upload the data
Click "Choose file" in the top left corner and select the .csv file containing your data.

Step Two: Select graph type
Change the graph type to "Dot plot(and box and whisker)" in the bottom left corner.

Step Three: Select variables
Set variable one as your numerical variable, e.g. strength and variable two as your categorical variable (the groups you're comparing), e.g. dragon colour.

Step Four: Title and labels
Create an appropriate title for your graph. A good default is: "(Numerical variable) by (categorical variable)", e.g. "Strength by dragon colour" or "Marathon time by gender". Have suitable axis labels including units.

Step Five: Annotate graph
In the bottom centre section, tick the boxes to show:

- summaries
- box plots [or high box plots or box (no outlier)]

You may also want to show the mean dot as well.

Step Six: Remove outliers (if applicable)
You may wish to remove some points as you consider them to be outliers. To do so, identify which row of the spreadsheet contains that point. You can do this by turning on point labels in the bottom centre section and selecting "Row +/-" in the top left corner and deleting the specific row that matches that data point. Make sure to turn off point labels before copying your graph.

Alternatively you can sort the spreadsheet by that variable by clicking "sample and more" in the top left corner. Sort by your numerical variable. Now you can remove the outlier by deleting the last (or first) row.

Making a bootstrap distribution
Later in your report, you'll need to refer to a bootstrap distribution.

Step One: Select graph type
Change the graph type to "bootstrap confidence interval" and select either median or mean depending on which parameter you want to infer about.

Step Two: Title and labels
Create an appropriate title for your graph. A suitable title could be "(Parameter) (numerical variable) of both groups in the population", e.g. "Median strength of both groups in the population "

Lesson Two: Bootstrapping

Different approaches to investigating a population
Let's answer the investigative question: "I wonder what the difference is between the median strength of green dragons and the median strength of red dragons on this island?"

Suggestion: Take a census of the dragons on the island.

What are some issues with this approach?

Suggestion: Take a sample of the dragons on the island and assume that the population medians are the same as the sample medians.

What are some issues with this approach?

Suggestion: Take many samples of the dragons on the island and average the various sample medians to estimate the population medians.

What are some issues with this approach?

Suggestion: Take a single sample of the dragons on the island and calculate an informal confidence interval to estimate the population medians.

An issue with this approach is that is assumes the population distribution is normally distributed, in other words it is unimodal and symmetrical. This is the normal distribution:

Obviously, data is not always going to be symmetrical and unimodal. Bootstrapping is a similar process that doesn't assume the population distribution is normally distributed. Instead, it assumes the population distribution is similar to the sample distribution.

This is another reason why it's important that the sample is representative of the population

Take a large enough representative sample of the dragons on the island and input it into the statistical software iNZight.

What is the difference between the sample medians?

The bootstrapping process
The bootstrapping process samples from the sample itself as if the sample were the population. This is why it's important that the sample distribution is representative of the population distribution. The sample taken from the sample is called the resample.

The resample is taken with replacement. This means that whenever a data point from the sample is selected to be in the resample, it's replaced so that it can be selected again.

The resample is exactly the same size as the original sample and each data point in the original sample could be selected for the resample no times, one time, two times, etc.

Watch the animation of resampling on iNZight and give a description in your own words of taking a resample from the sample.

Resampling many times
Computers are able to take many resamples very quickly giving you many resample medians. **The distribution of the resample medians is the same as the distribution of sample medians if you had taken many samples from the population.**

This is an important point. Resampling many times from the original sample with replacement has pretty much the same outcome as taking many samples from the population. The obvious advantage to resampling is it can be done by computers and doesn't require collecting any new data.

This is where the name 'bootstrap' comes from. It's derived from the phrase "to pull yourself up by your bootstraps". Resampling allows you to effectively take new samples from the population without actually taking any new samples. More precisely, the patterns of variation generated by sampling from a population are approximated by the patterns of variation generated by resampling.

Use the Kiwi dataset to first take 1 000 samples from the population. Then take a single sample from the population and resample it 1 000 times. Compare the two distributions.

For each resample, the difference between the resample medians is recorded in what's called the bootstrap distribution. Doing this 1 000 times and excluding the top and bottom 2.5% give a bootstrapped confidence interval for the likely difference between the population medians.

Use iNZight to take many resamples of the dragon sample. Interpret the confidence interval and use this to answer the investigative question: *"I wonder what the difference is between the median strength of green dragons and the median strength of red dragons on this island?"* Can you make the call that one group of dragons has a higher median strength than the other group? You may find the exemplars helpful in wording this effectively.

When the confidence interval includes zero
If any of the points on the bootstrap distribution are negative, that means the resample represented by that point had the difference between the groups the other way around, e.g. if the original sample had the median strength of red dragons higher than green dragons, then this particular resample had the median strength of *green* dragons higher than red dragons.

If, after removing the extreme ends of the bootstrap distribution, the interval includes zero, then there is not enough evidence that there is a different between the population medians. Note that this doesn't mean that there is no difference, only that if there is one it's too hard to tell which way it goes.

Use iNZight and/or NZGrapher to create a bootstrap confidence interval for the difference between the median age of smokers and non-smokers. Interpret that interval and state whether you can make a call that one group's median age is older than the other's in the population?

Median vs. mean
It matters which parameter you choose to bootstrap. The median is a good default because it's unaffected by skew and outliers so can be used without much thought. However, the mean is more efficient that the median resulting in a narrower bootstrap confidence interval. Therefore if the bootstrap confidence interval for the difference between the population medians contains zero, you could create a bootstrap confidence interval for the different between the population means and see if that doesn't contain zero.

However, since the mean is affected by skew and outliers, it shouldn't be used for samples that have either skew or outliers (after removing all appropriate outliers).

Use iNZight or NZGrapher to create a bootstrap confidence interval for the difference between the **mean** strength of green and red dragons on the island. State that confidence interval. You don't need to interpret it.

Lesson Three: Analysis

The need to analyse the original sample
If the bootstrap confidence interval is what will enable you to determine the difference between the population medians (or means) and answer your investigative question, then what is the need to analyse the original sample?

There is important information in the sample which will inform your resampling and bootstrapping.

Analysing the sample
That important information is:

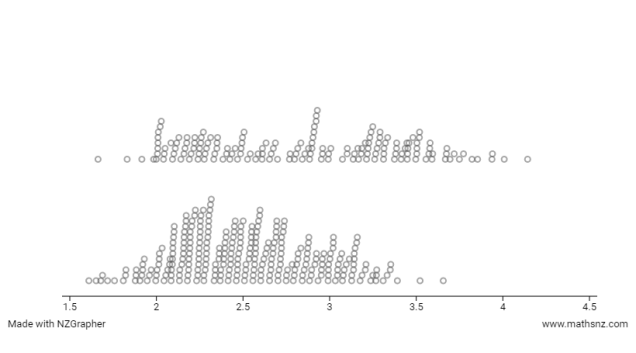
- The centre/shift
- The spread
- The shape
- Any unusual features such as outliers

Centre/shift
The centre/shift is important as it gives you an indication as to whether you can expect there to be a difference between the population medians. A large shift between the sample medians suggests a large difference between the population medians. Similarly, a small shift between the sample medians suggests a small difference between the population medians and that it will be too close to call which population median is bigger. Of course, this is only a suggestion and bootstrapping is required to calculate a confidence interval for the likely difference between the population medians.

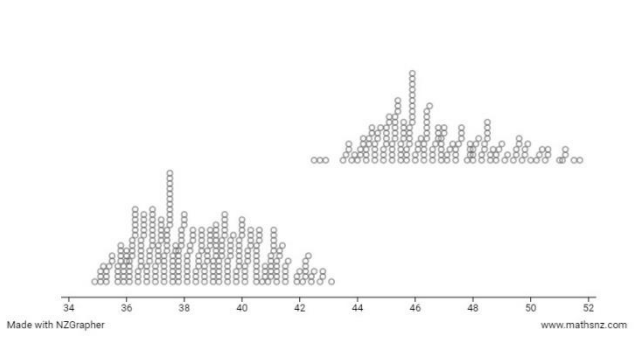
The best comparison of centre is by using your eyes; a visual comparison. You could use either the median or the mean as a measure of centre to back up your visual comparison.

The middle 50% is constructed using the lower and upper quartiles. You could use the middle 50% to discuss the extent to which the groups' data overlap each other. You could also use it to compare the spread between the two groups, i.e. the larger the middle 50%, the more variation there is.

Spread
The spread of the data is important for a similar reason. If there is more variation *within* the two groups than *between* them, that also suggests that it may not be possible to make a call with the bootstrap distribution as to which population median is bigger.



More variation within than between groups.



More variation between than within groups.

The best comparison of spread is a visual comparison. Both the interquartile range (the width of the boxes) or the standard deviation can be used as measures of spread to back up your visual comparison. Remember that the interquartile range is based on the quartiles which is based on the median so is unaffected by skew and outliers whereas the standard deviation is based on the mean so is vulnerable to skew and outliers.

Shape
The shape of the data is important, particularly if it is skewed or symmetrical. This affects your choice of median vs. mean and interquartile range vs. standard deviation. You should also state whether the distributions are unimodal or bimodal. Again, the best observation is a visual observation. You could use the length of each half of the boxes or the length of the whiskers to back up your visual observation regarding skew.

Justification and insight
Analysing the sample with justification is required (among other things) to achieve the standard with merit. Doing so with insight is required to achieve with excellence.

To analyse the sample with justification you need to justify your observations. For example:

- Using the median / mean / middle 50% to back up your visual comparison of the centre / shift.
- Using the interquartile range / width of the boxes / standard deviation to back up your visual comparison of the spread.
- Using the symmetry (or not) of the box and whisker graphs to back up your visual observation of the shape of the distribution.
- Linking back to the population. Always be thinking: "what does this suggest about the population or how does this help me answer my investigative question?"

To analyse the sample with insight you need to show you understand the connections between the different stages of the statistical enquiry cycle (Problem, Plan, Data, Analysis, Conclusion). You also need to integrate statistical and contextual knowledge throughout your analysis (and throughout your report). For example:

- Referencing knowledge from your research and integrating it into your analysis.
- Seeking explanations for features of the data, i.e. thinking more deeply about the consequences of your observations on the population. This is also called the "so what?" factor.
- Justifying why you're using the median rather than the mean or why you're using the interquartile range rather than the standard deviation (or vice versa) and considering the impact on the investigation, i.e. how would it be different if you used the other parameter?

A note about sampling variability

Sampling variability is the idea that two samples from the same population will be different because they're both only estimates. The difference between the samples doesn't reflect a difference in the populations (they're from the same population!). Whenever you analyse a sample, you must keep reminding yourself that there is some fuzziness and uncertainty.

If you want to integrate statistical insight throughout your report, you could discuss sampling variability throughout your analysis. For example, if you took another sample, you would expect the sample statistics such as the median, mean, interquartile range, and standard deviation to be different but at least similar if your sample is representative.

Lesson Four: Conclusion

Making the call

Your conclusion is a summary of your investigation. Somewhere in your conclusion you need to answer your investigative question. In other words, you need to make a call as to which group has a bigger median and by how much (or that you're unable to make the call).

For example, if your investigative question was: *"What is the difference between the median strength of green and red dragons on the island?"* then a suitable answer would be: *"I can make the call that red dragons have a stronger median strength than green dragons on this island by between 2 and 4 shields."*

In order to say how big the difference between the population medians is, you need to refer to your confidence interval, i.e. *"I am pretty sure the median strength of red dragons is between 2 and 4 shields stronger than the median strength of green dragons on this island."* You can be 'pretty sure' as in most cases, the confidence interval contains the true difference between the population medians at least 90% of the time. This figure increases the more representative the sample is of the population.

You can only make this call if the entire confidence interval is positive, i.e. doesn't include zero. If zero or negative values are in the interval it means that the difference could be the other way, e.g. green dragons could actually have a stronger median strength in the population. You need to state why you can (or can't) make the call. For example: *"Because the confidence interval doesn't include zero I can make the call that red dragons…"*. Remember that the bootstrapping process relies on the sample being representative of the population.

If you're unable to make the call, that doesn't mean that there is no difference between the population medians, only that you don't know what the difference is. There could be a difference but it's not possible to tell as the size of the difference is small compared to the sampling variation.

Sampling variability

You need to discuss sampling variability. Sampling variability causes samples to vary from one another so if anyone took another sample you'd expect that their sample would look different and have different sample statistics (i.e. different sample medians). As a consequence, you'd expect them to have a different confidence interval as well. But what wouldn't change is the call that they can make. If you were able to make the call, you'd expect that anyone with another sample would also be able to make the call and vice versa.

Justification and insight

If you wish to show justification and insight you could back up your conclusion and inference with a summary of relevant observations from your analysis, i.e. did you observe a large or small shift in your sample and does that match your inference for the population? Also, is zero just outside of your confidence interval or is it well outside of it?

You could also link your analysis and inference to research. Where do they agree and disagree?

You could also consider other possible factors that are influencing the numerical variable. For example, it may be that aggressive dragons or older dragons are stronger and that colour isn't the only factor. Use your common sense and contextual knowledge to consider whether other influences could play a part; you don't need to redo your analysis to consider them.

You could also relate your conclusion to the purpose of the investigation, i.e. what do you think should happen in the real world given your results?

Sample conclusions

There are some sample conclusions at each level of achievement on the following page.

Sample conclusions
Here are some conclusions at each level of achievement.

Not achieved
I am pretty sure that the population median of blood alcohol level for high risk groups involved in car crashes in New Zealand is somewhere between 0 milligrams and 23 milligrams higher than the population median low risk groups.

Based on these bootstrapping 1000 samples I cannot make the call that there is a difference in blood alcohol level between high risk groups in New Zealand and low risk groups, because zero is the bootstrapped confidence interval

Achieved
Based on my samples I can conclude that there is in fact a difference between median blood alcohol levels recorded for serious crashes and median blood alcohol level recorded for minor crashes. This is backed up by the bootstrapped confidence interval which shows that the population blood alcohol levels recorded for serious accidents do tend to be higher than those recorded at minor crashes by between 11.5 and 34.5mg/100ml blood.

With a bootstrapped confidence interval as large as 11.5mg/100ml to 34.5mg/100ml, I am fairly certain that should another sample be taken and processed the resulting confidence interval would still remain above zero and indicate that blood alcohol levels recorded for serious crashes are greater than blood alcohol levels recorded for minor crashes.

Merit
From these samples, I suggest that there is a difference in median alcohol level for drivers involved in serious and minor crashes in New Zealand in 2011. Furthermore, I am pretty sure that the difference in median for blood alcohol is somewhere between 11.00mg/100mls and 34.50mg/100mls from the bootstrapping process. Therefore, my investigation shows that drivers involved in serious crashes have relatively more alcohol per 100mls of blood, than those involved in minor crashes.

From the sample data of crash severity I have concluded that there is a difference the amount of alcohol absorbed into the blood stream (measured in milligrams per 100mls) depending on the severity of the crash. This conclusion is also confirmed with my bootstrap confidence interval for the difference between the population medians alcohol levels of drivers involved in serious and minor crashes in New Zealand, which does not contain zero, indicating that the median alcohol level per 100mls of blood for drivers involved is greater than the median for those involved in minor crashes. This is what I expected and is also what my research led me to believe would be my conclusion.

If I were to choose another identical age group sample, the results may have differed as that sample could have contained individuals that were relatively more responsible than individuals in this particular sample. However, despite these differences, I believe that I would still be likely to make the same call.

Excellence
I can make the call, with the sample that I am given, that the blood alcohol level is higher in serious crashes than minor crashes, for the most and least at-risk of crashing groups in New Zealand, as there is a significant gap in the median values of 135 and 160mg/100mL. This is backed up by the bootstrapped confidence interval, as a difference of 0 is well outside the confidence interval of 9mg/100ml to 35mg/100mL.

However we cannot know in these crashes whether the driver who was drunk did the cause of the accident, nor how the level of blood alcohol affected their injuries. As such, we cannot completely disprove the research saying that alcohol protects you in a crash: an identical crash with a person who was less intoxicated may have resulted in more injuries. This would require much more data, and a significantly more detailed analysis.

In this sample, the outliers will only have a small effect on the means. I would not immediately hesitate in using them to find a bootstrapping CI for the means and use this as a second source of evidence. This may only apply to this sample of crashes, though, so median would still be a better average.

Assessment guidelines - 3.10 Make a formal inference

	Achieved	Merit	Excellence
Problem / Plan	The question is a comparison investigative question that clearly identifies the comparison and the population(s). A brief explanation is given of the purpose of the investigation.	A comparison investigative question has been posed **and gives a more detailed justification for the purpose of 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.		
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.** An understanding of sampling variability is evident. **The conclusion includes an interpretation of the confidence interval.**	The formal inference is used to answer the investigative question, justifying the call and **linking back to the purpose of the investigation.** **Sampling variability and the variability of estimates have been discussed – an understanding that there will be a variation in sample statistics with a different sample is indicated.** The conclusion includes an interpretation of the confidence interval. **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. Evidence for the award of a grade may be found anywhere in the report.