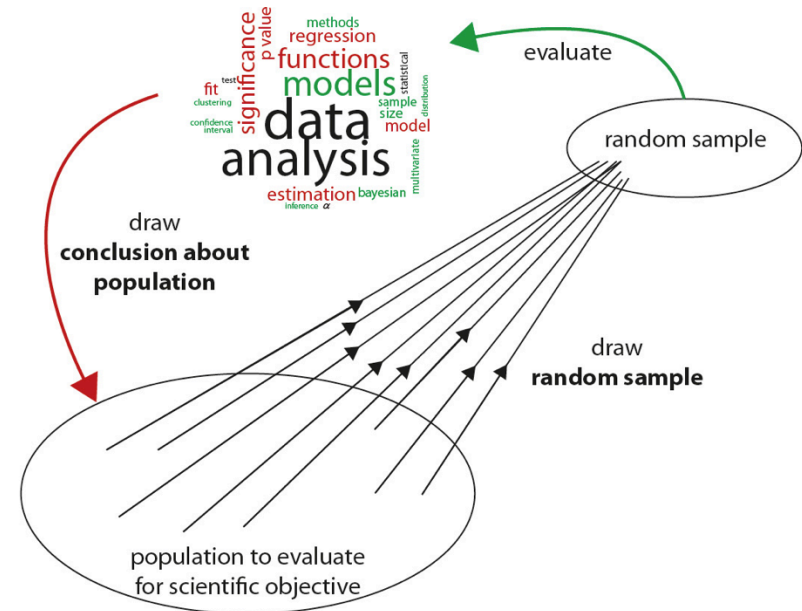


Inference Pain Points

Statistical Inference: the practice of forming judgments about the parameters of a population on the basis of random sampling.

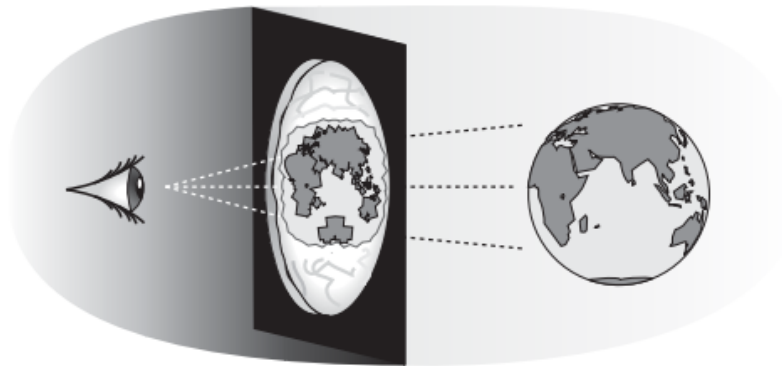
1. Strengthen your conceptual understanding of the 3 inference standards by exploring their pain points
2. Can the question be answered with the data?
3. Does the activity provide data that allows for generalization of results to a population
4. Does the analysis provide useful information for addressing the question?
5. Does the interpretation provide an appropriate answer to the statistical question?



Robyn Headifen
Team Solutions
**EDUCATION AND
SOCIAL WORK**

The overriding idea in these 3 inference standards is that we are investigating a population, a population that we cannot see.
To do this we take a random sample from the population.

Looking at the World Using data



Is like looking through a window with ripples "*Chris Wild 2008*"

Notes:

Level 1 91035	Level 2 91264	Level 3 91582
Investigate a given multivariate data set using the statistical enquiry cycle	Use statistical methods to make an inference	Use statistical methods to make a formal inference
<p>Students need to be familiar with the statistical enquiry cycle to investigate a given multivariate data set, which involves:</p> <ul style="list-style-type: none"> investigating data that has been collected from a survey situation posing an appropriate comparison question using a given multivariate data set selecting and using appropriate display(s) giving summary statistics such as the five summary values (minimum, maximum, median, quartiles) discussing features of distributions comparatively, such as shape, middle 50%, shift, overlap, spread, unusual or interesting features communicating findings, such as informal inference and supporting evidence, in a conclusion. 	<p>Using the statistical enquiry cycle to make an inference involves:</p> <ul style="list-style-type: none"> posing an appropriate investigative comparison question from a given set of population data selecting random samples selecting and using appropriate displays and measures discussing sample distributions discussing sampling variability, including the variability of estimates making an inference communicating findings in a conclusion. 	<p>Using the statistical enquiry cycle to make a formal inference involves:</p> <ul style="list-style-type: none"> 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 communicating findings in a conclusion.
What is the same across all 3 levels		
What is the difference?		

Pain Point 1: The Question



Notes:

Good questions lead to appropriate inferences
The question we ask is about the population

91035

We are investigating whether a variable for one group tends to be bigger than that variable for another group in the population using visual analysis.

*I wonder if green kiwifruit **tend to be** heavier than gold kiwi fruit from the 2015 harvest in the Te Puke region?*

91264

We are investigating whether one group in the population's median is bigger than the other group's median.

*I wonder if the **median** weight of green kiwifruit is higher than the **median** weight of gold kiwi fruit from the 2015 harvest in the Te Puke region?*

91582

We are investigating the size of the difference between the medians/means of two groups in a population.

*I wonder **what the difference is between** the median weights of green kiwifruit and gold kiwifruit from the 2015 Harvest in the Te Puke region?*





Pain Point 2: The Inference

What is the answer to the question, using the data and the analysis?

91035

The inference is made after recording a visual analysis of the data plots. A decision can be made from the position of the medians relative to the “boxes”. More complicated rules are unnecessary.

From my analysis I can make the call that green kiwifruit do not tend to be heavier than gold kiwifruit from the 2015 Harvest in the Te Puke Region, but I can make the call that gold kiwifruit tend to be heavier than green kiwifruit

I can make this call because...

91264

The inference, conclusion, is made about the population medians, based on the random sample taken from the entire data set provided.

Informal confidence intervals are used in making the decision.

The informal confidence intervals should be interpreted for each group, then, an inference made based on whether the confidence intervals overlap or not.

My confidence intervals show green kiwi fruit are very likely to have a median weight between 108 grams & 120 grams and gold kiwifruit are very likely to have a median weight between 102 grams and 116 grams.

These two confidence intervals overlap so there is a chance the median weight for green kiwifruit could be larger or smaller, or the same as, the median weight for gold kiwifruit from the 2015 harvest.

Therefore I cannot say whether the median weight of green kiwifruit from the 2015 harvest in Te Puke is likely to be more than the median weight of the gold kiwifruit.

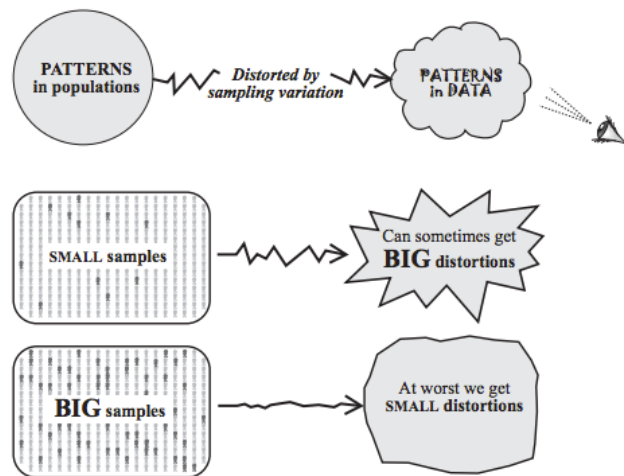
91582

At this level the inference is made based on a [bootstrap confidence interval](#) for the difference between the medians/means.

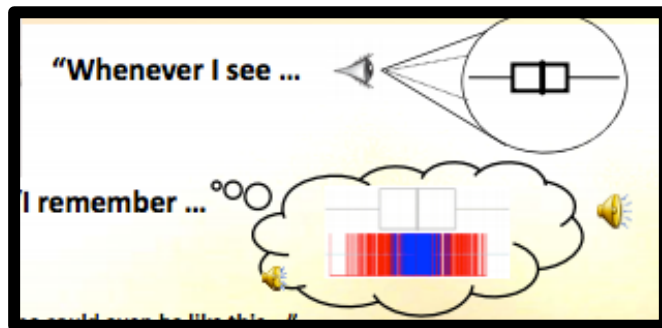
If the bootstrap interval shows -6 to 4 then we could say the median weight for the green kiwifruit from the 2015 harvest in Te Puke is likely to be from up to 6 grams lighter to 4 grams heavier than the median for the gold kiwifruit.

Pain Point 3: Sampling variability

Students should be introduced to the idea of variability between samples in the junior school by taking several samples from that same “population” and comparing to develop these ideas.



We want students to instinctively understand that different samples will almost certainly give different sample statistics.



Chris Wild

Data sets on Census at School

1. [Kiwi Data](#)
2. [Koala & Trout Data](#)
3. [Census at School database](#)

Note: at level 2 time should be spent developing the CI formula rather than just given to the students. (see [Kiwi Kapers](#) on CAS)

[Animations of sampling variation](#)

Senior Secondary Guides: Statistical Investigations

AO 6-1: [Plan & Conduct investigation using the statistical inquiry cycle](#)

AO 7-1: [Carry out investigation of phenomena using the statistical inquiry cycle](#)

AO 8-1: [Carry out investigation of phenomena using the statistical inquiry cycle](#)

[Link](#) to Statistical Investigation notes at Level 7

At level 7 the new concepts are formalising sampling methods for taking a random sample, starting to formalise inference, and the reliability of an inference.

Relevant contextual Knowledge

From TKI:

The progression for contextual knowledge from level 6 to level 8 is

1. At level 6 students should be working with contexts that are familiar to them.
2. At level 7 students should be provided with relevant contextual knowledge about the situation under investigation.
3. At level 8 students should be sourcing relevant contextual knowledge about the situation under investigation from places such as the internet, the school or local library, newspapers and magazines.

Source: <http://seniorsecondary.tki.org.nz/Mathematics-and-statistics/Achievement-objectives/AOs-by-level/AO-S7-1/Statistical-graphs>

Notes:

CONTEXT --- CONTEXT --- CONTEXT

Contextual knowledge should be driving the investigation.

Context should be incorporated throughout a report and integrated with statistical knowledge.

Providing Structure

It is important to give students structure but equally important is to remove the structure over time.

Centre

Spread & shape

Interesting

What do I see in the displays?

Where do I see it?

What does it make me think about the population?

Obvious

Specific

Evidence

Meaning

PEER & PEEL

SEXY

Being Left Handed makes you taller

Researchers took a large sample of New Zealand students from the 2013 census at school database. The results of the survey are summarised below. Your task is to investigate this.


Variables : Hand: students were recorded as being left handed or right handed
 Height: students height was measured in centimetres

left

right

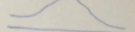
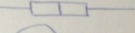
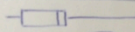


height

	min	Q1	Median	Q3	Max	N
Right	122	153	161	169	198	156
Left	141	159	170	179	210	22

Features and Keywords	Starting sentences
<p>Question</p> <p></p> <ul style="list-style-type: none">VariableGroupsPopulationCompareDirection	<p>I wonder</p>
<p>Feature 1</p> <p>Centre</p> <ul style="list-style-type: none">MedianUnitsMore/lessbigger/smaller	<p>I notice....</p> <p>The median height for right handed students is ____ cm</p> <p>The median height for left handed students is ____ cm</p> <p>The median height for left handed students is ____ cm bigger than for female students</p> <p>This makes me think ...</p>

<p>Feature 2</p> <p>Spread</p> <p>Small → large</p> <p>LQ → UQ</p> <p>IQR (middle 50%)</p> <p>Interquartile range</p>	<p>I notice ...</p> <p>The middle 50% of heights for left handed students are between ____ cm and ____ cm</p> <p>The middle 50% of heights for right handed students are between ____ cm and ____ cm</p> <p>The middle 50% of heights for right handed students are <u>more / less</u> spread out than middle 50% of heights for left handed students</p> <p>This makes me think ...</p>
<p>Feature 3</p> <p>Shape</p> <ul style="list-style-type: none">SymmetricalNot symmetrical <p>Skew to right / left</p>	<p>I notice</p>
<p>Interesting</p> <ul style="list-style-type: none">BiggestSmallestGroupGap	<p>One thing I found interesting is...</p>
<p>Making the call</p> <p>Overlap of boxes</p> <p>Where is the median ?</p> <ul style="list-style-type: none">Within the boxAbove the boxBelow the box <p>The box = middle 50%</p>	
<p>Conclusion</p> <p>Answer the question</p> <p>Use ... tend to ...</p>	<p>In conclusion</p> <p>I can / cannot make the call that</p>

Shape of dot plot

- Symmetric = 
- Right skewed = 
- Left skewed = 
- clusters
- bimodal 
- unimodal 

The distributional shape of A is _____

The distributional shape of B is _____

The distributional shape may reflect _____ back in the Population for sample A/B.

This means that

This makes sense because

min

LA

med

UQ

max

Foldable
(Lucy Edmonds)

Investigate a given multivariate data set using the statistical enquiry cycle AS 91035

The bigger picture: We want to tell the stories about a population based on a set of sample data

PERFORMANCE INDICATORS:

Achieved:

Using the statistical enquiry cycle involves using each component of the statistical enquiry cycle to make comparisons.

Merit

Using the statistical enquiry cycle with justification involves linking aspects of the statistical enquiry cycle to the context and the population and making supporting statements which refer to evidence such as summary statistics, data values, trends or features of visual displays.

Excellence

Using the statistical enquiry cycle with statistical insight involves integrating statistical and contextual knowledge throughout the statistical enquiry cycle, and may involve reflecting on the process or considering other explanations for the findings.

Understanding

- We investigate the data using PPDAC
- We must be careful because: "What we see is not quite the way it really is"
- We look at samples to try and understand what is going on in the population
- We notice what we see in front of us (the kiwifruit) and wonder what we don't see in front of us (all kiwi fruit)

Knowledge

- The components of the statistical inquiry cycle (PPDAC) for this standard and what they mean
- The information I can use to justify my findings
- The general knowledge I have to reflect on the findings

Criteria

- Investigating data that has been collected from a survey situation
- Posing an appropriate comparison question using a given multivariate data set selecting and using appropriate display(s) giving summary statistics such as the five summary values (minimum, maximum, median, quartiles)
- Discussing features of sample distributions comparatively, such as shape, middle 50%, shift, overlap, spread, unusual or interesting features
- Communicating findings, such as informal inference and supporting evidence, in a conclusion.

Modes of Assessment

Written report assignment	Open ended
Google Docs	One Note
Oral presentation	Story boards
Blog	Prezzie

Essential questions

- What is obvious in the data? **What do I see?**
- What have I noticed specifically? **Where do I see it?**
- **What does this mean** when I compare it with what I already know?

Key Idea

- Students need to understand the difference between "the kiwi fruit" (the sample) and "kiwifruit" (the population).

Essential vocabulary

Compare	Variable	Distribution	Middle
Group	Less than	Shape	Spread
Sample	More than	Population	Overlap
Make a call	Tends to	Inference	Unusual

Pain Points.

The key idea in this standard is the understanding of inference. We are drawing a conclusion about a much bigger group (the population) by comparing 2 sub groups in one sample. There is a chance the features displayed in the sample are not a true reflection of the population hence we describe our findings in terms of a tendency as we are not absolutely sure.

Issues arise when the question does not reflect this. To ensure that students make an appropriate inference a question should be about the population and include the groups & variables being compared & a "fuzzy" word like tends to

Do green kiwi fruit tend to weigh more than gold kiwi fruit from the 2015 harvest in the Te Puke region?

Based on a comparison of features seen in the displays the inference could be ***Green kiwi fruit tend to weigh more than gold kiwi fruit from the 2015 harvest in the Te Puke region of New Zealand***

Skills

- Writing a comparative investigative question
- Obtaining dot and box/whisker plots
- Describing what we see in the displays
- Answering the comparative question
- Presenting a conclusion

Resources:

Wild about Statistics

<https://www.stat.auckland.ac.nz/~wild/wildaboutstatistics/index.html>



inzight for data Analysis

<https://www.stat.auckland.ac.nz/~wild/iNZight/index.php>

NZGrapher for data Analysis

<http://www.jake4maths.com/grapher/>

CensusAtSchool
NEW ZEALAND

Take part in the
census

Explore the data

Resources for
teaching statistics

<http://new.censusatschool.org.nz/resources/>

1. Add to database by having students take the survey
2. Use data by taking random samples and using the data viewer
3. Find resources for teaching

Statistics Teachers NZ

<https://www.facebook.com/groups/statsteachers/>



Notes

Worthwhile Reading

Set aside some time for reading each week

The Statistical Education of Teachers



SET. American Statistics Association

<http://www.amstat.org/education/SET/SET.pdf>

Chapter 5 & 6 are of particular interest. Preparing middle & high school teachers to teach statistics

Let's talk about literacy

Preparing students for the transition to tertiary learning

LISA EMERSON, with KEN KILPIN and ANGELA FEEKERY

http://www.nzcer.org.nz/system/files/set2015_1_003_0.pdf

This article talks about the need for students to have the opportunity to do their own thinking. Well worth reading



Telling Data Stories : Essential Dialogue for Comparative Reasoning;
Maxine Pfannkuch and Chris Wild

<http://www.amstat.org/publications/jse/v18n1/pfannkuch.pdf>

On Line MOOC's

Keep your eye out for free on line education.



<https://place.fi.ncsu.edu/>

Teaching Statistics through data Investigations

This MOOC starts again on March 9 2016.

Through this MOOC I found the newsletter called the Statistics Teachers Network. Published twice a year

<http://www.amstat.org/education/stn/>

Home Grown on Future Learn

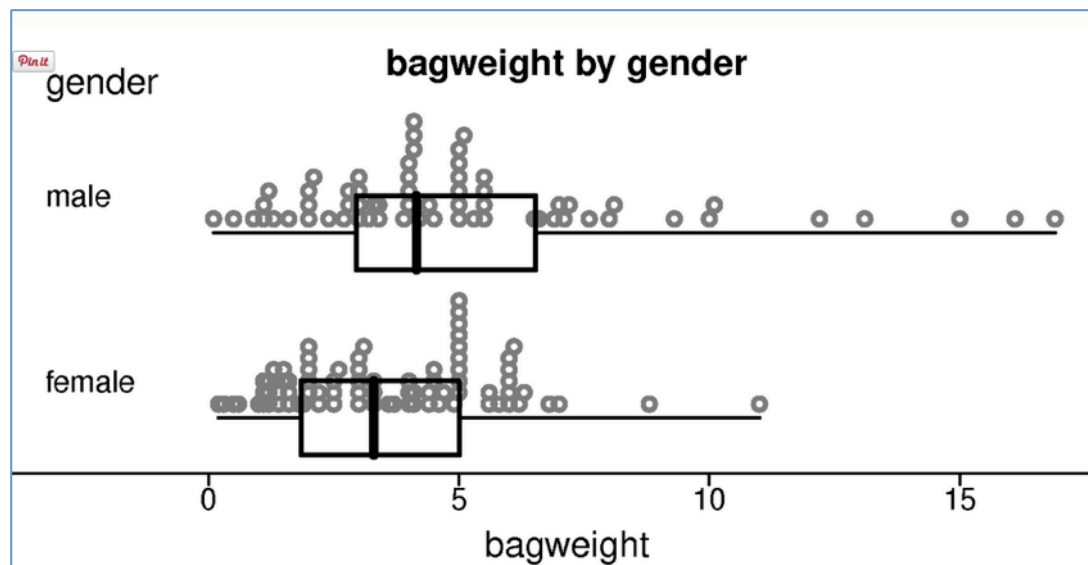
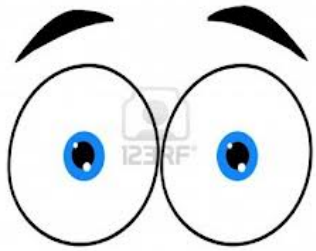


Data to Insight

an introduction to statistical data analysis for those new to the subject as well as those wanting a reminder

<https://www.futurelearn.com/courses/data-to-insight>

This course is in progress. Join now to gain access to the content



Source Data: New Zealand Census at School secondary school students

W.W.W.