Level 2 – AS91264 – 4 Credits – Internal

**2.9 Use Statistical Methods to Make an Inference**

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

|  |  |  |
| --- | --- | --- |
| **Achievement** | **Achievement with Merit** | **Achievement with Excellence** |
| Use statistical methods to  make an inference. | Use statistical methods to make an inference, with  justification. | Use statistical methods to  make an inference, with  statistical insight. |

[Part 1: Sampling Methods 2](#_Toc447560967)

[Part 1.1: Cluster Sample 2](#_Toc447560968)

[Part 1.2: Simple Random Sample 2](#_Toc447560969)

[Part 1.3: Stratified Sample 2](#_Toc447560970)

[Part 1.4: Systematic Sample 2](#_Toc447560971)

[Part 2: Sampling Variability 3](#_Toc447560972)

[Part 3: Writing a Good Question 4](#_Toc447560973)

[Part 4: Interpreting Dot Plots 6](#_Toc447560974)

[Part 4.1: Measures of Centre 6](#_Toc447560975)

[Part 4.2: Measures of Spread 7](#_Toc447560976)

[Part 4.3: Looking at Shape 7](#_Toc447560977)

[Part 4.4: Looking at Unusual Features 7](#_Toc447560978)

[Part 5: Making an Inference 7](#_Toc447560979)

[Part 6: Pulling it All Together 7](#_Toc447560980)

[Part 6.1: Fill in the blanks 1 7](#_Toc447560981)

[Part 6.2: Fill in the blanks 2 7](#_Toc447560982)

[Part 6.3: Write your own internal 7](#_Toc447560983)

[Assessment Guidelines – 91264 – 2.9 Inferences 8](#_Toc447560984)

# Part 1: Sampling Methods

In real life it is often difficult or expensive to sample every single person (or animal or whatever) so hence a sample needs to be taken in order to estimate what we think the population is, without having to spend lots and lots of money doing a census (asking everyone)

## Part 1.1: Cluster Sample

A method of sampling in which the population is split into naturally forming groups (the clusters), with the groups having similar characteristics that are known for the whole population. A simple random sample of clusters is selected. Either the individuals in these clusters form the sample or simple random samples chosen from each selected cluster form the sample.

**Example**

Consider obtaining a sample of secondary school students from Wellington. The secondary schools in Wellington are suitable clusters. A simple random sample of these schools is selected. Either all students from the selected schools form the sample or simple random samples chosen from each selected school form the sample.

## Part 1.2: Simple Random Sample

A sample in which, at any stage of the sampling process, each object or individual (which has not been chosen) in the population has the same probability of being chosen in the sample.

In a simple random sample, an object or individual in the population can be chosen once at most. This is often called **sampling without replacement**.

## Part 1.3: Stratified Sample

A method of sampling in which the population is split into non-overlapping groups (the strata), with the groups having different characteristics that are known for the whole population. A simple random sample is taken from each stratum.

**Example**

Consider obtaining a sample of students from a secondary school with students from year 9 to year 13. The year levels are suitable strata, and the simple random samples taken from each year level form the sample.

## Part 1.4: Systematic Sample

A method of sampling from a list of the population so that the sample is made up of everykth member on the list, after randomly selecting a starting point from 1 to k.

**Example**

Consider choosing a systematic sample of 20 members from a population list numbered from 1 to 836.

To find k, divide 836 by 20 to get 41.8.

Rounding gives k = 42.

Randomly select a number from 1 to 42, say 18.

Start at the person numbered 18 and then choose every 42nd member of the list.

The sample is made up of those numbered

18, 60, 102, 144, 186, 228, 270, 312, 354, 396, 438, 480, 522, 564, 606, 648, 690, 732, 774, 816

Sometimes rounding may cause the sample size to be one more or one less than the desired size.

*Source:* [*http://seniorsecondary.tki.org.nz/Mathematics-and-statistics/Glossary/*](http://seniorsecondary.tki.org.nz/Mathematics-and-statistics/Glossary/)

# Part 2: Sampling Variability

The variation in a sample statistic from sample to sample.

Suppose a sample is taken and a sample statistic, such as a sample mean, is calculated. If a second sample of the same size is taken from the same population, it is almost certain that the sample mean calculated from this sample will be different from that calculated from the first sample. If further sample means are calculated, by repeatedly taking samples of the same size from the same population, then the differences in these sample means illustrate sampling variation.

*Source:* [*http://seniorsecondary.tki.org.nz/Mathematics-and-statistics/Glossary/*](http://seniorsecondary.tki.org.nz/Mathematics-and-statistics/Glossary/)

Let’s look at a more concrete example. Say there are 30 people in your maths class (which for us is our population). Their heights are: 152, 153, 154, 156, 158, 158, 163, 163, 165, 166, 166, 168, 168, 169, 170, 170, 171, 171, 174, 175, 177, 179, 181, 181, 182, 183, 183, 185, 185, 190.

If we take a random sample of 5 from them we might get 163, 170, 171, 182, 190 which gives us a mean height of 175.2cm … this is fine for this sample, but it obviously won’t be the same for every sample we take. The bigger the sample the less variation we should get, but there will still always be some variation due to the fact we have taken a sample and not asked everyone.

**Task:** Complete the table below by taking simple random samples of 5 from the ‘population’ of you class. The dataset can be accessed at: <http://www.jake4maths.com/grapher/?folder=students&dataset=Class%20Heights.csv>

To take a sample press the ‘Sample and More’ Button and then press ‘Sample’. To reset the dataset and get all the points back you can press the ‘Reset Dataset’ button.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Sample** | **Point 1** | **Point 2** | **Point 3** | **Point 4** | **Point 5** | **Mean** |
| **1** | 163 | 170 | 171 | 182 | 190 | 172.5 |
| **2** |  |  |  |  |  |  |
| **3** |  |  |  |  |  |  |
| **4** |  |  |  |  |  |  |
| **5** |  |  |  |  |  |  |
| **6** |  |  |  |  |  |  |
| **7** |  |  |  |  |  |  |
| **8** |  |  |  |  |  |  |
| **9** |  |  |  |  |  |  |
| **10** |  |  |  |  |  |  |

# Part 3: Writing a Good Question

The investigative question that is posed must involve a comparison and needs to include

* the variable (eg: height)
* the population groups being compared (eg: year 12 boys and year 12 girls)
* the population parameter the inference will be about (eg: median) and
* the direction of the comparison (eg: boys **greater** than girls).

A suitable question would be

“I wonder if the median height of NZ year 12 boys is greater than the median height of NZ year 12 girls?”

Write a good question for each of the graphs below, the first one has been done for you.

|  |  |
| --- | --- |
| http://www.jake4maths.com/grapher/imagetemp/2aPFp5B80u.png | I wonder if the median height of NZ year 12 boys is greater than the median height of NZ year 12 girls? |
| http://www.jake4maths.com/grapher/imagetemp/u32WKi7Szs.png | I wonder if the \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ greater than the \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_?  Data is for Year 12 students from your school. |
| http://www.jake4maths.com/grapher/imagetemp/CEVl6RBrsy.png | I wonder if the \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ greater than the \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_?  Data is for Kiwi Birds from around New Zealand. |
| http://www.jake4maths.com/grapher/imagetemp/XKOkc7LqrA.png | Data is from professional rugby players from New Zealand and South Africa. |
| http://www.jake4maths.com/grapher/imagetemp/m96XChvA2b.png | Data is from marathon runners competing in events in New Zealand. Smaller times means it was completed faster. |

# Part 4: Interpreting Dot Plots / Boxes and Whiskers

For this internal we analyse dot plots and box and whiskers… you can see some examples of dot plots and box and whisker graphs in the previous section. This involves looking at a few different things which we go into more detail below.

## Part 4.1: Measures of Centre

In this internal we look at both the means and the medians… but which ones are better and why?

Let’s look at an example to illustrate this. Let’s say you earn $4 pocket money a week… you feel like you are a little poorly done by so you go and ask some of your friends and find out that they get $1, $2, $3 and $40 each. You then work out the average amount is $10. You worked this out by going: . You go home and tell your parents this and they say, no the average amount is $3, as that is the middle number. Who is right?

It turns out you are both right. The mean amount is $10, and the median amount is $3.

To work out the **mean** you add all the numbers together and divide by the number of items.

To work out the **median** you put the numbers in order and find the middle number, so for the example below it is 1, 2, **3**, 4, 10… so $3 is the middle amount.

When you have an even number of amounts, eg: 1, 2, **3, 4**, 5, 6… the median is the average of the middle two numbers, in this case **3 and 4** which gives a median of 3.5.

**It is really important to make sure you put the numbers in order before finding the medians.**

**Task:** Find the mean / median of the data below. Give answer to at least 3 significant figures.

1. 1, 7, 10, 5, 3  
   mean =
2. 6, 8, 10, 4, 6  
   median =
3. 3, 6, 9, 9, 9, 8  
   median =
4. 8, 1, 3, 4, 4  
   mean =   
   median =
5. 78, 33, 98, 21, 100  
   mean =   
   median =
6. 13, 59, 35, 93, 99, 50, 53, 56, 71, 44, 48, 41, 74, 20, 30  
   mean =   
   median =
7. 26, 40, 12, 38, 35, 32, 34, 100, 27, 29  
   mean =   
   median =

Want more practice… [click here](http://students.mathsnz.com/qg/onscreen.html#statistics1meanmed6,statistics1meanmed6,statistics1meanmed6,statistics1meanmed6,statistics1meanmed6,).

**Answers: 1.** 5.2 **2.** 6 **3.** 8.5 **4.** mean: 4.00, median: 4.00 **5.** mean: 66.0, median: 78.0   
**6.** mean: 52.4, median: 50.0 **7.** mean: 37.3, median: 33.0

## Part 4.2: Measures of Spread

Range

IQR

## Part 4.3: Looking at Shape

Symetrical vs Skew – overall and middle 50%

## Part 4.4: Looking at Unusual Features

Extreme Values

# Part 5: Making an Inference

median ± 1.5 IQR / √n

# Part 6: Pulling it All Together

## Part 6.1: Fill in the blanks 1

## Part 6.2: Fill in the blanks 2

# Sample Internal

# Assessment Guidelines – 91264 – 2.9 Inferences

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Achieved (all compulsory)** | **Merit… Achieved PLUS** | **Excellence… Merit PLUS** |
| **Problem** | Pose a clear investigative question linked to the situation | Must have a hypothesis / prediction | Hypothesis is justified |
| **Sample** | Select an appropriate sample with evidence of method | Select an appropriate sample with evidence of method and contextual reasons have been given for deciding on the use of a sampling method **or** the sample size | Select an appropriate sample with evidence of method and contextual reasons have been given for deciding on the use of a sampling method **and** the sample size |
| **Data** | Select and use appropriate displays and measures (dot plots and box plots) | Informal confidence intervals must be shown on graph |  |
| **Analysis** | Discuss Sample Distribution with comparative statements   * Centre * Spread * Shape   State one informal confidence Interval OR include point estimate of population parameter  Answer question including making a correct inference | Discussion is linked to the investigative question and population with supporting evidence  State in **context** both informal confidence intervals  Inference is supported in context. | Discussion is supported with contextual knowledge or alternative explanations |
| **Conclusion** | Communicate findings clearly  Discuss sampling variability including variability of estimates | Has linked findings to context and population  The effect of one aspect has been considered with sampling variability | Has reflected on the process |

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