

# AS91264 (version 3) Mathematics and Statistics

**Use statistical methods  
to make an inference**

Te whai i ngā tikanga o te tūhuratanga  
tauanga hei whakaputa hīkaro

Te whai hua - kia ora!

**sorted**  
*in Schools*

This resource provides answers to  
the questions in the student AS91264  
Practice booklet.

LEVEL

2

CREDITS

4

SORTED THEMES

KiwiSaver  
Retirement  
Managing my Money

# Module Enquiry Context

## Module Enquiry Context Activity

---

- 1) a.** KiwiSaver is a voluntary scheme set up by the government to help people get ready for retirement.
- b.** You can access your funds when you turn 65. If you have invested in KiwiSaver for three or more years, you can also use your funds to help pay for your first home. If you end up in extreme financial difficulty, you can also apply to get access to your KiwiSaver funds.
- c.** Reasons that females often have less money than males when they reach retirement age include:
- Females tend to earn less per hour than males, so they have less money to save or invest
  - Females are more likely than males to take a career break when they have children. This means that there are periods of their life when they are not earning money.
  - In terms of finances, females are more likely to be affected negatively by relationship break-ups than males are, so they have less money to save or invest.
- 2) a.** Females in Aotearoa New Zealand do twice the amount of unpaid labour that males do.
- b.** Examples of factors include:
- race
  - class
  - sexuality
  - language and culture
  - access to education
  - age
- 

## Topic One:

### Establishing a purpose and an investigative question

---

#### Topic 1, Activity 1

Students watch the [Sorted in Schools Inequality video](#). They should discuss how the gender pay gap and financial wellbeing in retirement might be connected.

#### Topic 1, Activity 2

Identify whether the following variables are categorical variables **or** numerical variables:

Variable	Categorical or numerical
Gender	Categorical
KiwiSaver balance	Numerical
Personal income	Numerical
Employment status	Categorical
Age group	Categorical. Although we use numbers to describe how old people are, when we define people by “age group” we’re putting them into a category.

#### Topic 1, Activity 3

Other factors mentioned in the video that impact on inequality include race, class, sexuality, language, culture, age, and access to education.

#### Topic 1, Activity 4

A **population** is any large collection of objects or individuals that we want to investigate.

A **parameter** is any summary number, such as an average or percentage, that describes the entire population. This value is **unknown** because the **population** size is too large to find out what it is.

A **sample** is a group drawn from the population.

A **statistic** is any summary number, such as an average or a percentage, that describes the sample.

Because samples are manageable in size, we can work out the value of any **statistic**.

We can then use this to make an informed guess about the value of a population **parameter**.

## Topic 1, Activity 5

- a. The median is \$17,727
- b. The mean is \$30,503.
- c. The median is the most useful average for this set of data because it is not affected by the three very large values. Two thirds of the values are below the mean, so it's not a very "typical" value for this data set.

## Topic 1, Activity 6

Natasha has developed the following investigative question for her statistical enquiry:

*In 2018, was the median KiwiSaver balance of full-time workers in Aotearoa New Zealand (measured in NZ dollars) higher than the median KiwiSaver balance of part-time workers in Aotearoa New Zealand (measured in NZ dollars)?*

- a. Fill in the table below using components of Natasha's investigative question to show that she has met all of the requirements of an investigative question.

<b>The variable of interest (including units)</b>	KiwiSaver balance in New Zealand dollars
<b>The two groups being compared</b>	full-time and part-time workers



<b>The parameter being compared</b>	the median KiwiSaver balance
<b>The direction of the comparison</b>	full-time median higher than part-time median
<b>The population</b>	workers in Aotearoa New Zealand in 2018

**b.** Answers will vary. Here are two examples:

- *In 2018, was the median personal income of 18 to 34 year olds in Aotearoa New Zealand (measured before tax in NZ dollars) lower than the median personal income of 35 to 54 year olds in Aotearoa New Zealand (measured before tax in NZ dollars)?*
- *In 2018, was the median KiwiSaver balance of full-time workers in Aotearoa New Zealand (measured in NZ dollars) higher than the median KiwiSaver balance of part-time workers in Aotearoa New Zealand (measured in NZ dollars)?*

## 2

## Topic Two: Selecting a sample

---

### Topic 2, Activity 1

**1) a.** The population is students at your school.

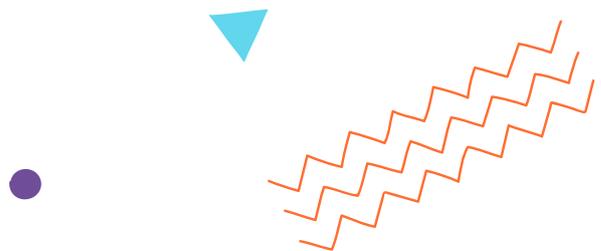
A survey is emailed to 100 students randomly selected from your school asking how much money they had spent at the school cafeteria the week before.

**b.** The variable of interest is the amount of money a student spent at the school cafeteria the week before the survey was sent out.

- c. The sample has been selected randomly. As a result, groups in the student population (such as year level or gender) may not be represented in the same proportions as they exist in the population.
- 2) A stratified sample is more likely to produce a sample that is representative of the population.
  - 3) Answers will vary. Examples of strata include: gender, ethnic group, age, region, level of education, income.

## Topic 2, Activity 2

- 1) This is a practical activity. Check that students understand how to use software to select a sample.
- 2) Using just age makes the sample too specific. In real life, there's not much difference between a 32 and a 33 year old in terms of age or income. By using an age group instead we can see whether there are differences that relate to different life stages.

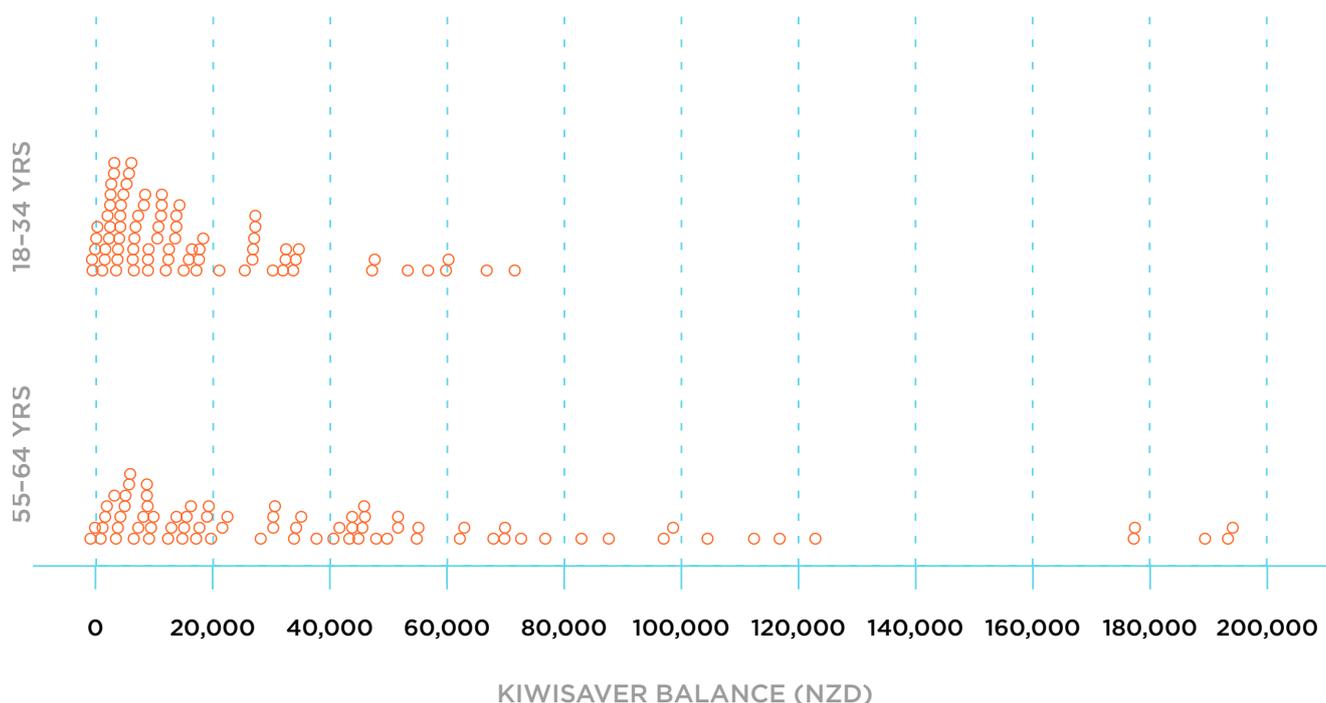


## Topic Three:

### Displaying data and calculating statistics

#### Topic 3, Activity 1

KiwiSaver balances by age group



Features:

#### Centre

The KiwiSaver balances of the sample from the older age group (55 to 64 year olds) have a higher median (middle value) than the sample from the younger group (18 to 34 year olds). I can tell this because the old age group is spread out reasonably evenly from 0 to 120,000 while the highest KiwiSaver balance of the younger sample group has a highest value of around \$80,000.

#### Symmetry

Neither sample is symmetric, they both tend to have more values clustered more towards the lower end of the KiwiSaver balances. Both samples are skewed to the right.

## Spread

This KiwiSaver balances of the younger group sample are more clustered towards the lower end of the scale than the KiwiSaver balances of the older group sample. The range of the KiwiSaver balances for the older age group is much larger than the range of the younger age group. This is likely to be because older investors have had a longer time to invest money in KiwiSaver. They are also more likely to have large incomes and spare money that they can invest than younger people.

## Outliers

The sample of KiwiSaver balances of the older group has a few potential outliers (values over \$170,000). However, because there are 5 values in this range, they may not be outliers – this just suggests that some people in the population have a lot more money than other people. It's not surprising that all of the largest values are in the older age group as their earning and investing power is likely to be far greater than that of young people in the population.

## Topic 3, Activity 2

- 1) This is a practical activity. Check that students understand how to use software to create a dot plot and box plot and how to display summary statistics on their graph.
- 2) a. Graphs and summary statistics will vary depending on the sample.
  - b. It's likely that the sample includes some very large values. Outliers affect the mean but not the median.



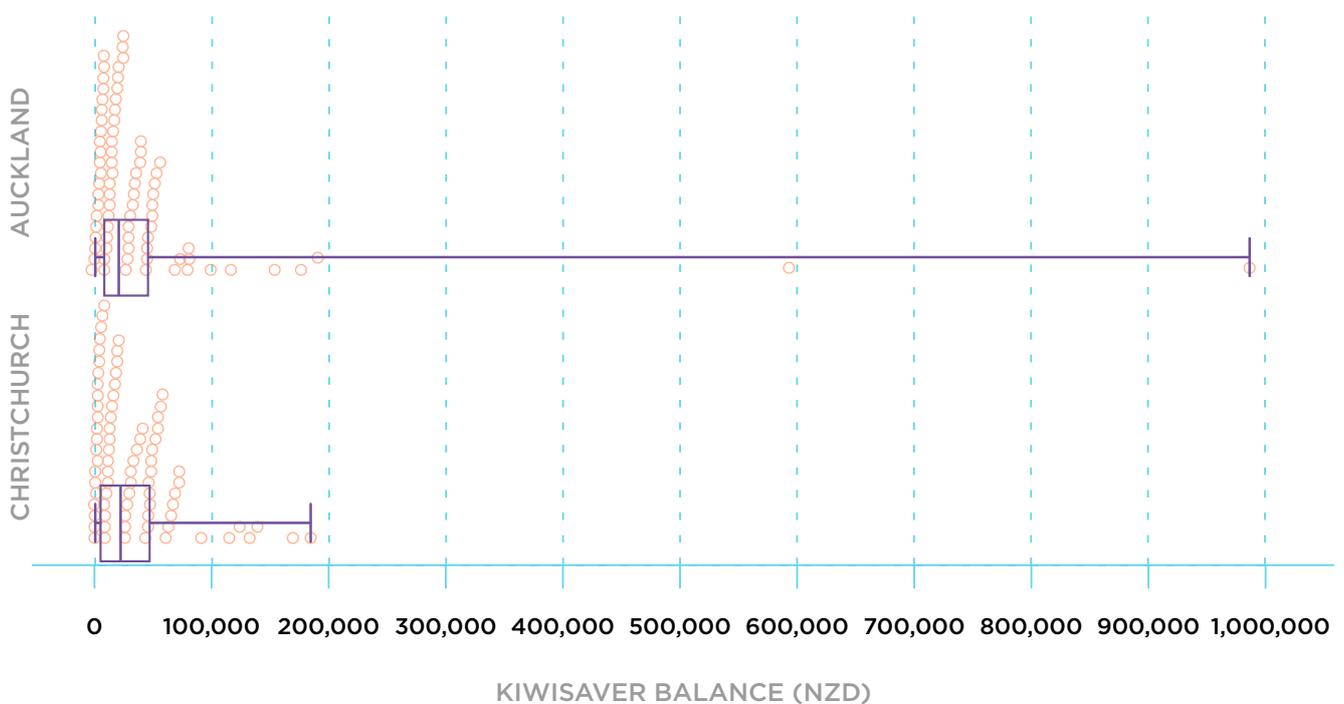
## Topic Four:

### Comparing box plots and summary statistics

#### Topic 4, Activity 1

- 1) The graph below shows the KiwiSaver balances of a sample of 80 people from Auckland and 80 people from Christchurch.

**KiwiSaver balances of people from Auckland and Christchurch**



Summary Statistics		
	Auckland sample	Christchurch sample
Minimum	\$157	\$528
Lower Quartile	\$7,154	\$4,892
Median	\$18,780	\$19,384
Mean	\$50,748	\$34,301
Upper quartile	\$44,726	\$47,998
Maximum	\$989,010	\$184,149

## Shape

<b>Point</b>	Both samples are heavily skewed to the right.
<b>Explanation</b>	This means that the data values are more closely clustered together in the lower half of the data.
<b>Evidence</b>	As a result, the mean of each sample is much higher than the median. For example, the sample median of the Auckland KiwiSaver balances is \$18,780 while the sample mean is \$50,798
<b>Link</b>	This reflects the income inequality that is part of Aotearoa New Zealand society. Some people enter retirement with far more money invested than others.

## Symmetry

<b>Point</b>	For both samples, the middle 50 percent of the data is also not symmetric.
<b>Explanation</b>	This is evident in the box plots of the sample KiwiSaver balances, because the median is closer to the lower quartile than to the upper quartile.
<b>Evidence</b>	For example, for the Christchurch sample, the lower quartile is around \$14,500 less than the median, while the upper quartile is over \$28,600 more than the median.
<b>Link</b>	It would be interesting to make comparisons with the incomes of these two sample groups. It's likely that they would also be unsymmetrical because our income levels influence how much money we have left over for investing.

## Spread

<b>Point</b>	The Auckland sample has smaller interquartile range than the Christchurch sample.
<b>Explanation</b>	This means that in the sample, the middle 50 percent of KiwiSaver balances of people from Auckland is less spread out than the middle 50 percent of KiwiSaver balances of people from Christchurch.

<b>Evidence</b>	The interquartile range for the Auckland sample of KiwiSaver balances is \$37,572 (\$44,726 - \$ 7,154) while the interquartile range of the Christchurch sample of KiwiSaver balances is \$43,106 (\$47,998 - \$ 4,892)
<b>Link</b>	I was surprised by this difference, and I would need to learn more about the populations of Auckland and Christchurch to understand whether this is significant. In the scale of things, this difference might not mean much. It could also just be these particular samples.

# 5

## Topic Five: Confidence intervals

### Topic 5, Activity 1

- 1) No two samples taken from a population will ever be identical.

### Topic 5, Activity 2

1)

	Lower limit	Upper limit
Female sample	\$10,443.75	\$18,771.25
Male sample	\$12,611.35	\$24,002.65

### Topic 5, Activity 3

- 1) a. We can be reasonably sure that the median KiwiSaver balance of females in the population lies between \$10,444 and \$18,771.

We can be reasonably sure that the median KiwiSaver balance of males in the population lies between \$12,611 and \$24,002.

- b. The two confidence intervals overlap. Therefore, based on this sample, there is no evidence to suggest that the median KiwiSaver balance of females in Aotearoa New Zealand is lower than the median KiwiSaver balance of males in Aotearoa New Zealand.

# 6

## Topic Six:

### Writing your report

---

#### Topic 6, Activity 1

Responses will vary. A sample answer has been provided below.

This extended activity provides an opportunity for you to work your way through the process of using statistical methods to make an inference.

#### Problem

I want to investigate whether males in Aotearoa New Zealand tend to have higher savings than females in Aotearoa New Zealand.

It would be interesting to explore this because having savings is one way to buffer yourself from the financial ups and downs of everyday life.

Investigative question: In 2018, was the median amount of money that males in Aotearoa New Zealand have in savings higher than the median amount that females in Aotearoa New Zealand?

#### Plan and data

The data set I am using for my investigation is the 2018 ANZ financial wellbeing data set.

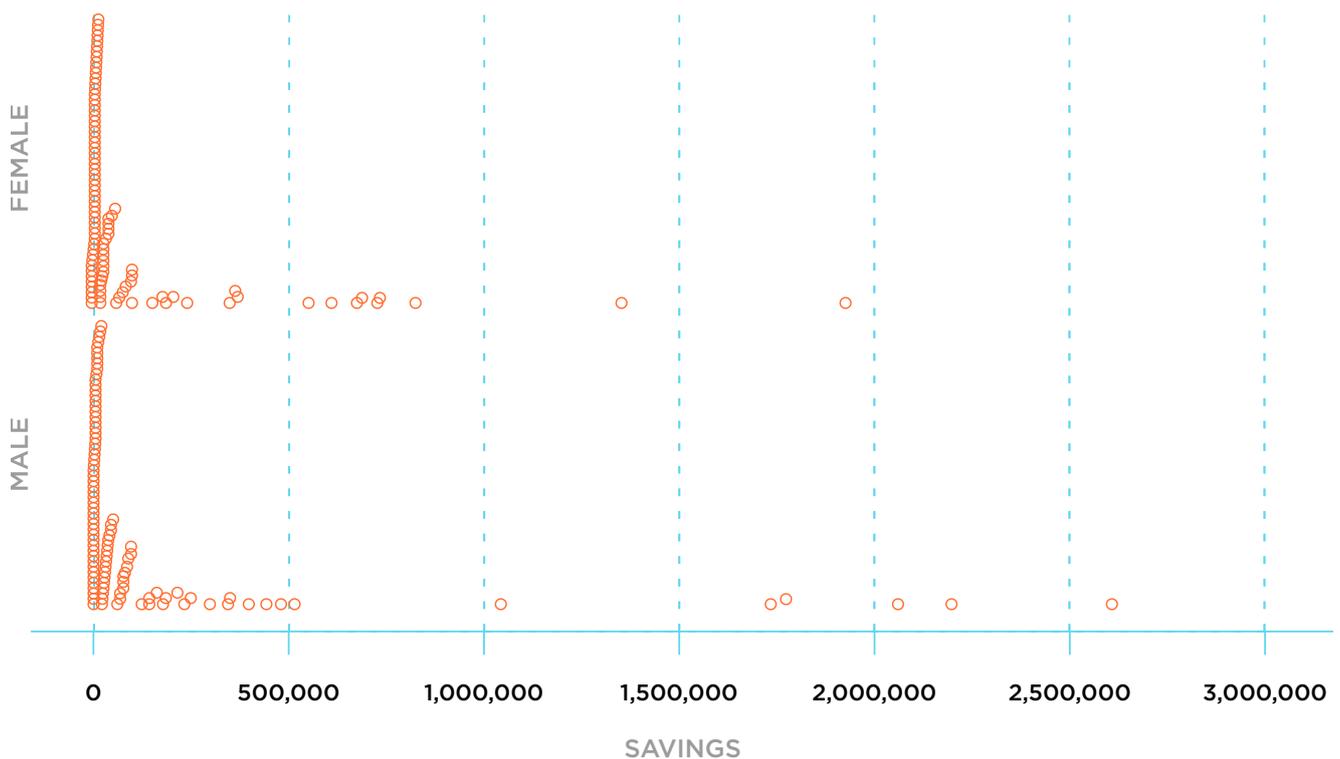
I will use simple random sampling to select a sample of 80 males and 80 females from the data set. I think that samples of size 80 will be big enough to reveal patterns in the data.

Using simple random sampling means that my samples will be unbiased. In other words, everyone in the population of males and females has an equal chance of being included in my sample.

## Analysis

I used NZGrapher to display the two samples.

### Savings of males and females in Aotearoa New Zealand



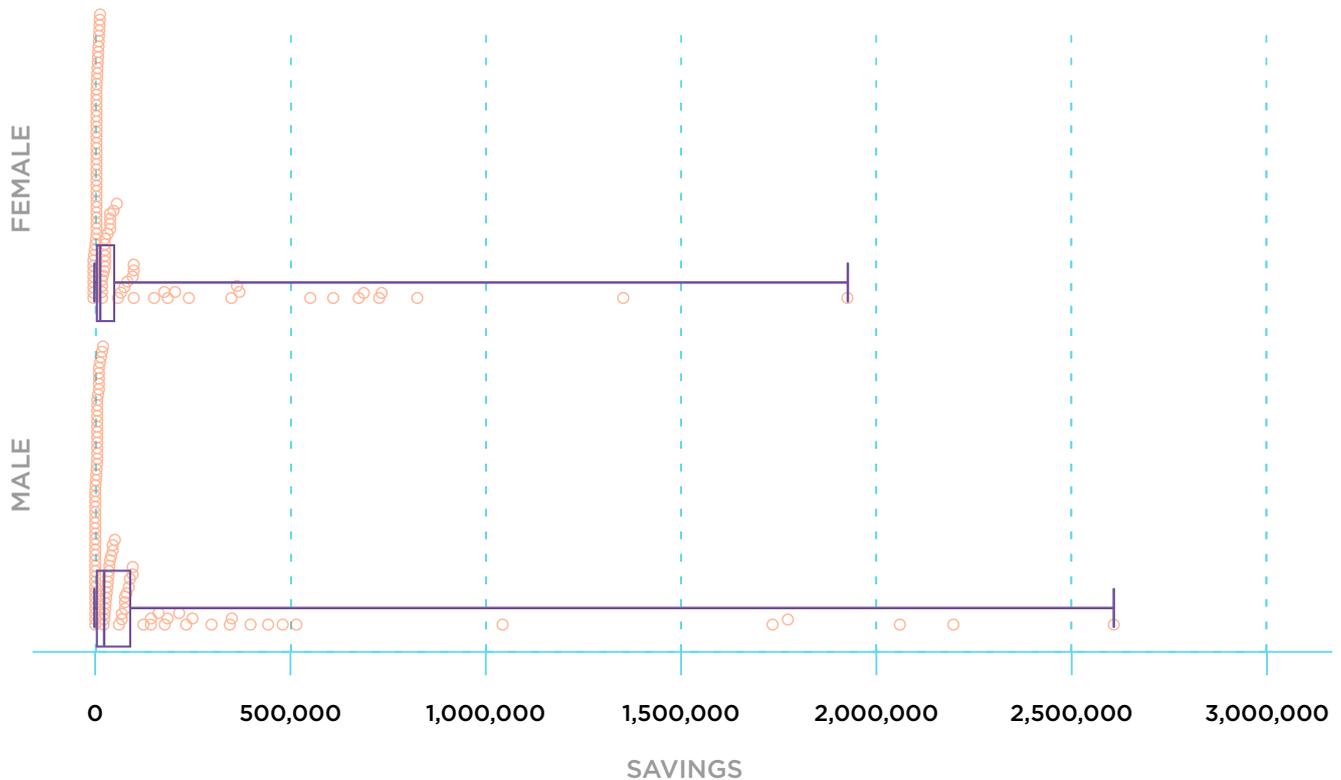
Looking at the two samples, it is clear that the savings of people in the samples are both heavily skewed to the right. I can see this because the data values are more closely clustered towards the lower end of the scale.

Each of the samples has some values that are much, much larger than the bulk of the values. For example, 8 people had more than a million dollars in savings. This is a stark contrast with the low amount of savings most people in each sample had.

It's not clear from the graph whether the median amount of savings for males and females in the samples are very different. The median saving amount for each sample won't be influenced by the extremely large values in the same way that the mean would be.

The graph below shows the box plots and summary statistics for each sample:

### Savings of males and females in Aotearoa New Zealand



Here are the summary statistics

Summary Statistics		
	Females	Males
Minimum	\$248	\$203
Lower Quartile	\$946	\$2,771
Median	\$10,014	\$17,472
Upper quartile	\$58,718	\$91,801
Maximum	\$1,924,092	\$2,609,210
Mean	\$115,710	\$178,270

### Shape

The summary statistics support my observation that both samples are skewed to the right. For example, for the sample of savings of females in Aotearoa, the lowest 25 percent of values lie between \$248 and \$946 (a difference of \$698). The top 25 percent of values in the same data set lie between \$58,718 and

\$1,924,092 (a difference of \$1,865,374). This shows the scale of inequality that there is in savings.

### Symmetry

The middle 50 percent of the sample data for each group is not symmetric either. For example, for the female data set, the lower quartile is around \$9,068 less than the median, while the upper quartile is around \$48,704 more than the median.

This unevenness is even more extreme for the male sample savings data. The lower quartile is around \$14,701 less than the median, while the upper quartile is around \$74,300 more than the median.

### Outliers

Both groups have extremely large values. For example, in the sample, the maximum amount of savings for females was over 2 million dollars and for the males there is a value over \$2,500,000. This reflects the income inequality that is part of Aotearoa New Zealand society.

### Confidence intervals

This graph shows the confidence intervals based on the medians, interquartile range, and sample size of each sample.



	Lower limit	Upper limit
Females	\$1,354	\$18,674
Males	\$4,117	\$30,827

Based on this sample, we can be reasonably sure that the median amount of money females in Aotearoa New Zealand have in savings is between \$1,354 and \$18,674.

We can be reasonably sure that the median amount of money males in Aotearoa New Zealand have in savings is between \$4,117 and \$30,827.

### Conclusion

Because the two confidence intervals overlap, based on this sample, there is no evidence that males in Aotearoa New Zealand have more savings than females in Aotearoa New Zealand.

One thing that has affected this investigation is the very large variation in the amount of savings that people in Aotearoa New Zealand have, particularly males. This is reflected in the width of the confidence interval for the median amount of savings males in Aotearoa New Zealand have, which is 1.5 times as wide as the confidence interval based on the female sample.

If I repeated this investigation using a different sample, it is likely that I would get a different result. However, if the variability of the sample of male savings amounts reflects a lot of variability in the population (which is likely), then repeating the investigation using another sample could still give me a confidence interval for males that is so wide that it is impossible to confidently confirm whether there is a difference between the two population medians.

