

Kiwi Kapers 3

Follows Kiwi Kapers 2. This activity is split into two parts. Collectively they aim to develop the method for finding informal confidence intervals for medians.

In part one students develop the concept of dividing the sample IQR by \sqrt{n} , recognising the effect of sample on the variability of an estimate.

In part two students investigate widths of informal confidence intervals such that the informal confidence intervals capture the population median about 90% of the time.

Problem



Level 6 and below problem: **What are typical weights of kiwis?**

At this point we can improve the questions we have been working with, as we are about to be able to answer a question about a parameter, such as the median, rather than the more general typical which we have used until now. We are moving towards finding an informal confidence interval for the **median weight** of the population of kiwis.

NOTE: You could pick this point up right at the start if you want to in the future.

Level 7 problem: **What is the median weight of kiwis?**

From our first activity (Kiwi Kapers 1) we decided that the median weight of kiwis was somewhere between ____ and ____ (*this will be from the activity kiwi kapers 1 and based on the range of medians from the class data, using sample size 30 because it was more reliable than samples of size 15*).

From the second activity (Kiwi Kapers 2) where we looked at different sample sizes, we decided that samples of size ____ (*base this on decision from kiwi kapers 2 activity*) were reliable enough.

From looking at the distribution of sample medians we could say that the median weight of kiwis was somewhere between ____ and ____ (*this will be based on the distribution of sample medians for that particular sample size and probably includes about 90% of the sample medians*).

HOWEVER, in real life we don't get to take multiple samples so this process won't work for us in the long run. We are going to look at finding an informal confidence interval for the population median based on a single sample.

Make the link to the work from level 6, that when we were exploring comparison situations we looked at both the spread and the sample size when deciding to make the call, and we will use both the spread and sample size when making an informal interval estimate.

PART ONE

Plan



- We want to compare the IQR for sample medians from different sized samples.
- We want to make connections between the IQR for sample medians from different sized samples and the population IQR
- We will collate results from from kiwi kapers 1 and 2 for samples of size 30.

Data



Give students **kiwikapers3studentworksheet.pdf** (available as a .doc if you need to edit, be aware that some of the labels drop off when printing from .docs.)

Students complete the worksheet for the sample medians for $n=30$ based on the sample medians found in kiwi kapers 1 & 2 activities.

Explain that the population distribution is shown in the first dot plot and box plot.

The second dot plot and box plot is the distribution of **sample medians** for samples of size 30 from the kiwi population. Ideally we would have used their sample medians based on the samples that they did as a class. However, we didn't record this information accurately enough for the purpose of this activity, so a set of 35 samples have been generated using technology to mimic what they did in kiwi kapers 1 (make the link to their graphs).

The third dot plot and box plot show the distribution of **sample medians** for samples of size 400 from the kiwi population. Again we have used technology to do this. There are 35 samples of $n=400$ which is close to the number of samples we took as a class for $n=30$.

Analysis



The two sample sizes $n=30$ and $n=400$ have been deliberately chosen as the difference in width of the IQR for sample medians is dramatic when compared with the population IQR and with one another. The pictures to the right show the number of the sample median IQR that equal the population IQR if needed to show this.

Q3. Samples of size 30 generate sample medians with an IQR of about $1/5^{\text{th}}$ the width of the population IQR.

Q4. Samples of size 400 generate sample medians with an IQR of about $1/20^{\text{th}}$ the width of the population IQR.

Students should be able to see these differences from the summary picture.

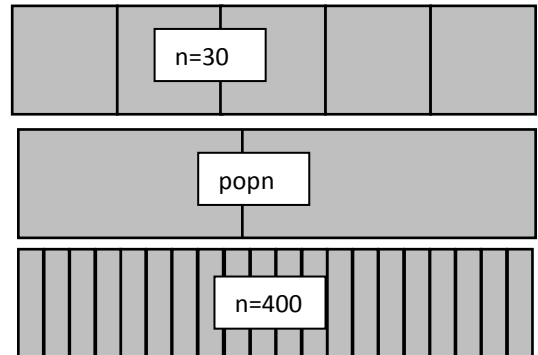
Q5. A bit more challenging is getting the $\sqrt[n]{n}$ link:

It might be useful to add in a further piece of information. For samples of size 100, the sample medians have an IQR of about $1/10^{\text{th}}$ the width of the population IQR.

IQR for sample medians (sample size= n) is approximately $1/\sqrt[n]{n}$ of the population IQR.

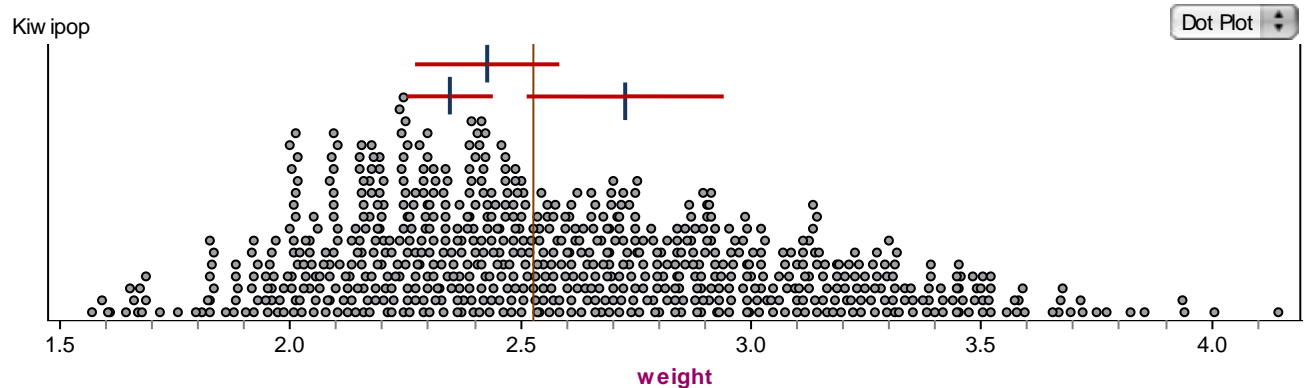
When $n=400$ the IQR of the sample medians is approximately $1/\sqrt[400]{400}$ or $1/20^{\text{th}}$ of population IQR.

When $n=30$ the IQR of the sample medians is approximately $1/\sqrt[30]{30}$ or $\sim 1/5^{\text{th}}$ of population IQR. *Note we chose not to use $n=25$ as all the previous work had been with samples of size 30.*



PART TWO

We want to use the sample median and the *sample IQR*/ \sqrt{n} to find an informal confidence interval for the population median. So we need to see how big we need to make the informal confidence interval so that we are pretty sure the interval includes the population median. Our method won't work for every interval, but we want it to work about 90% of the time. (Teachers probably want to show this picture to their students. It is on a separate page at the end to use for this purpose.)



- Informal confidence intervals are calculated using the sample median $\pm k \times \text{sample IQR}/\sqrt{n}$.
- In this picture we can see three informal confidence intervals that have been calculated using three different samples of size 30.
- Three different samples, three different medians, three different IQRs.
- We are trying to work out what would be the ideal number (k) of sample IQR/\sqrt{n} to use all the time to be pretty sure the interval includes the population median (marked by the vertical line). Population median is 2.529 kg.
- It is about us all using the same method.

Plan



- Construct an informal confidence interval (interval estimate) for the population median based on median of one sample (point estimate).
- The sample median, sample IQR and sample size are used to construct this informal confidence interval.
- We want 90% of the informal confidence intervals generated from samples to capture the population median.

NOTE for Teachers: *We are working towards finding the informal confidence interval which is **sample median $\pm 1.5 \times \text{sample IQR}/\sqrt{n}$** . We are using the sample IQR in place of the population IQR because we don't know the population IQR and sample IQR is a reasonable estimate of the population IQR. Students cannot be expected to come up with this interval by themselves; we need to scaffold their discovery of it.*

Data



Give students **kiwikapers3studentworksheetPART2.pdf** (available as a .doc if you need to edit, be aware that some of the labels drop off when printing from .docs.)

Note: this document has 3 pages.

Page 1: information sheet - for deciding number of IQR to divide by \sqrt{n} .

Page 2: write on sheet - example 1 for student books – this is for drawing on the informal confidence interval once they have found it (see activity four).

Page 3: write on sheet - Student write on sheet for testing informal confidence intervals.

Activity One: Deciding how many *sample IQR*/ \sqrt{n} are needed for the informal confidence interval

NEED: Page 1 and 3.

Explore how many times the *sample IQR*/ \sqrt{n} the three sample medians (examples 1, 2 & 3) are

away from the population median to establish what might be a sensible rule to use for most of the informal confidence intervals to include the population median.

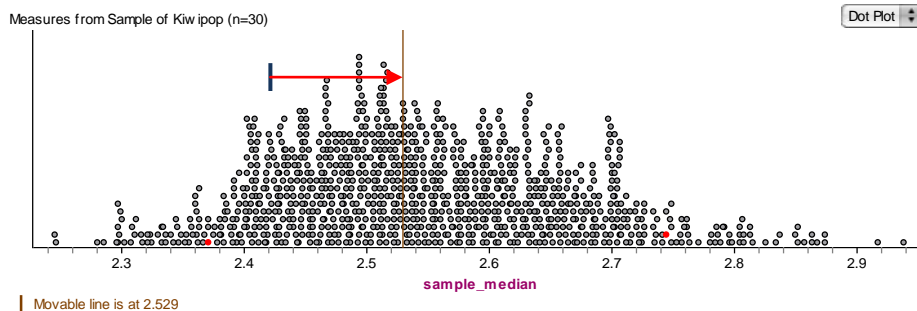
For each example:

- mark the sample median on the graph on page 3 and draw a line to the population median
- find the distance the sample median is from the population median (2.529 kg)
- divide by sample IQR/\sqrt{n}
- this gives the number of sample IQR/\sqrt{n} that the sample median is away from the population median
- this is the number we are interested in to use to try to decide what **k** is

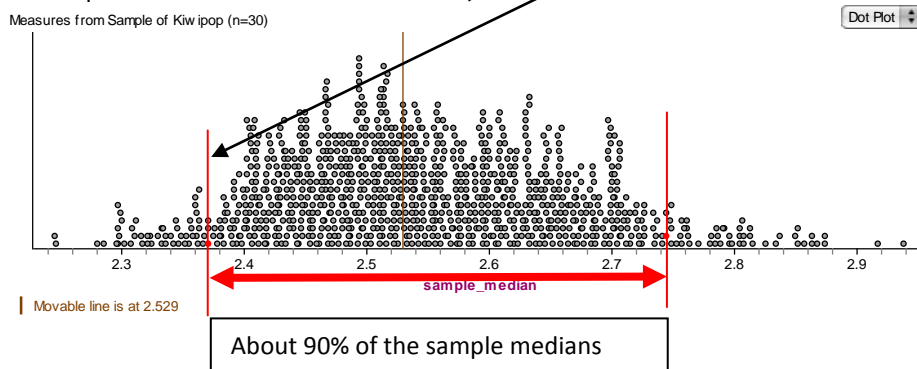
Example 1.

Distance the sample median is from the population median: $2.529 - 2.416 = 0.113$ kg
 Sample IQR/\sqrt{n} : $0.695/\sqrt{30} = 0.127$
 Distance/Sample IQR/\sqrt{n} : $0.113/0.127 = 0.82$

Interpretation: The sample median (2.416 kg) is 0.82 x **sample IQR/\sqrt{n}** below the population median. (Write above the line drawn at the start 0.82 IQR/\sqrt{n}).

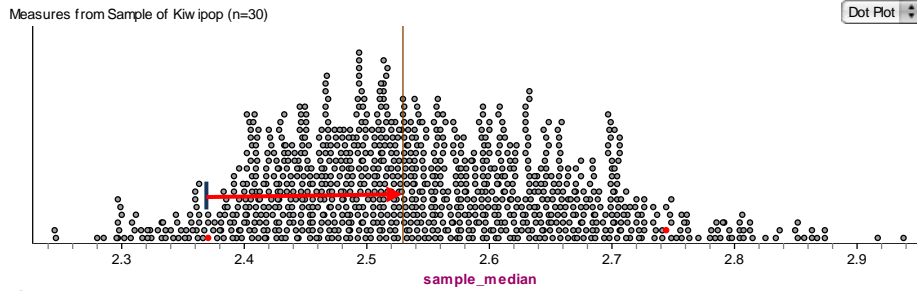


Example 2 (this sample median is at about the 5th percentile, so this is probably the lower bound of sample medians we want to include).



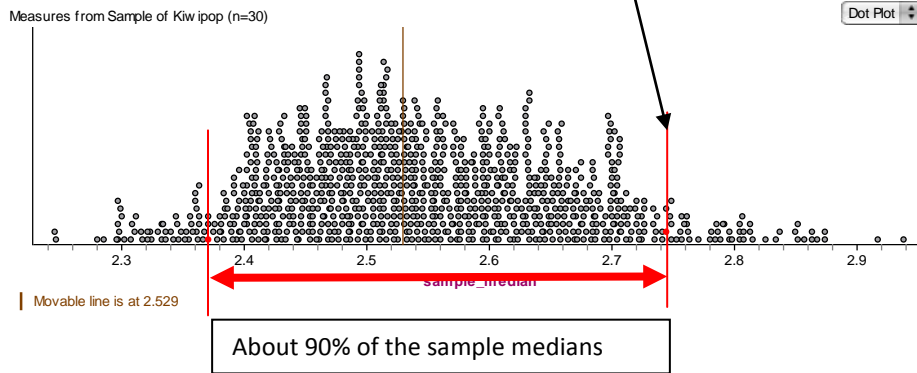
Distance the sample median is from the population median: $2.529 - 2.37 = 0.159$ kg
 Sample IQR/\sqrt{n} : $0.589/\sqrt{30} = 0.108$
 Distance/Sample IQR/\sqrt{n} : $0.159/0.108 = 1.47$

Interpretation: The sample median (2.37 kg) is 1.47 x **sample IQR/\sqrt{n}** below the population median. (Write above the line drawn at the start 1.47 IQR/\sqrt{n}).



Movable line is at 2.529

Example 3 (this sample median is at about the 95th percentile, so this is probably the lower bound of sample medians we want to include).

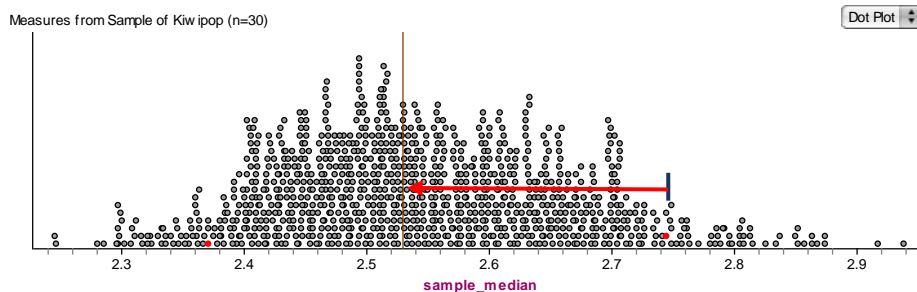


Movable line is at 2.529

About 90% of the sample medians

Distance the sample median is from the population median: $2.741 - 2.529 = 0.212$ kg
 Sample IQR/ \sqrt{n} : $0.81/\sqrt{30} = 0.147$
 Distance/Sample IQR/ \sqrt{n} : $0.212/0.147 = 1.44$

Interpretation: The sample median (2.741 kg) is 1.44 x **sample IQR/ \sqrt{n}** above the population median. (Write above the line drawn at the start 1.44 IQR/ \sqrt{n}).



Movable line is at 2.529

An informal confidence interval is calculated using the sample median $\pm k \times \text{sample IQR}/\sqrt{n}$. These results suggest that $\pm 1.5 \times \text{sample IQR}/\sqrt{n}$ might be about right.

Analysis



Activity Two:

Let's explore $\pm 1 \times \text{sample IQR}/\sqrt{n}$, $\pm 1.5 \times \text{sample IQR}/\sqrt{n}$ and $\pm 2 \times \text{sample IQR}/\sqrt{n}$ for a few samples to see if what happens in terms of the informal confidence interval for the different values of k .

For these five samples

1. sample median 2.784; sample IQR 0.758
2. sample median 2.645; sample IQR 0.468
3. sample median 2.594; sample IQR 0.806
4. sample median 2.532; sample IQR 0.691
5. sample median 2.515; sample IQR 0.845

calculate the three informal confidence intervals, using the sample median and IQR, and draw these informal confidence intervals onto page 3. Draw all the ± 1 , ± 1.5 and ± 2 in the same place (and the same colour?). The page has been divided into three to allow for this.

The intervals are given in the table below. Highlighted intervals include the population median.

Sample median	LInt1	UInt1	LInt1_5	UInt1_5	LInt2	UInt2	IQR
1) 2.784	2.65	2.92	2.58	2.99	2.51	3.06	0.758
2) 2.645	2.56	2.73	2.52	2.77	2.47	2.82	0.468
3) 2.594	2.45	2.74	2.37	2.81	2.30	2.89	0.806
4) 2.532	2.41	2.66	2.34	2.72	2.28	2.78	0.691
5) 2.515	2.36	2.67	2.28	2.75	2.21	2.82	0.845

Activity Three:

Exploring more intervals, a large number of generated intervals are available in the excel spreadsheet **intervaltest1000samplesize30.xlsx**. This has 20 pages of 50 intervals for ± 1 , ± 1.5 and ± 2 . Each line is one sample of 30, with the ± 1 , ± 1.5 and ± 2 interval boundaries calculated. Students need to see if the population median (2.529 kg) is within the interval or not.

Get the students to highlight the intervals that capture the population median. See the snapshot below to get an idea of what this will look like (note the first five samples in the excel spreadsheet are the samples used in activity two). The snapshot below shows the first 23 samples from the first sheet.

sample_median	LInt1	UInt1	LInt1_5	UInt1_5	LInt2	UInt2
2.784	2.64561	2.92239	2.57641	2.99159	2.50722	3.06078
2.645	2.55956	2.73044	2.51683	2.77317	2.47411	2.81589
2.594	2.44685	2.74115	2.37327	2.81473	2.29969	2.88831
2.532	2.40584	2.65816	2.34276	2.72124	2.27968	2.78432
2.515	2.36072	2.66928	2.28359	2.74641	2.20645	2.82355
2.486	2.37299	2.59901	2.31648	2.65552	2.25997	2.71203
2.426	2.28962	2.56238	2.22143	2.63057	2.15323	2.69877
2.453	2.30767	2.59833	2.23501	2.67099	2.16234	2.74366
2.5595	2.41143	2.70757	2.3374	2.7816	2.26336	2.85564
2.605	2.46752	2.74248	2.39878	2.81122	2.33004	2.87996
2.7035	2.56986	2.83714	2.50303	2.90397	2.43621	2.97079
2.461	2.35985	2.56215	2.30928	2.61272	2.25871	2.66329
2.5975	2.50037	2.69463	2.45181	2.74319	2.40324	2.79176
2.607	2.49946	2.71454	2.4457	2.7683	2.39193	2.82207
2.499	2.38799	2.61001	2.33249	2.66551	2.27699	2.72101
2.5	2.40981	2.59019	2.36471	2.63529	2.31962	2.68038
2.706	2.54607	2.86593	2.4661	2.9459	2.38613	3.02587
2.573	2.47258	2.67342	2.42238	2.72362	2.37217	2.77383
2.702	2.59319	2.81081	2.53878	2.86522	2.48437	2.91963
2.683	2.58094	2.78506	2.52991	2.83609	2.47888	2.88712
2.512	2.31811	2.70589	2.22116	2.80284	2.12421	2.89979
2.393	2.24931	2.53669	2.17747	2.60853	2.10563	2.68037
2.5955	2.47884	2.71216	2.4205	2.7705	2.36217	2.82883

From the class capture the number (%) of ± 1 that include the population median, then ± 1.5 and ± 2 . (In the snapshot above 18 intervals include the population median for ± 1 ; 21

intervals include the population median for ± 1.5 ; 23 intervals include the population median for ± 2).

Results should confirm that ± 1 is not wide enough, and ± 1.5 is sufficient to include the population median about 90% of the time, ± 2 is too wide and lacks precision and is no advantage over ± 1.5 because the narrower the informal confidence interval the more precise the interval estimate is.

Activity Four:

Calculate the informal confidence interval for the very first example once $\pm 1.5 \times \text{sample IQR} / \sqrt{n}$. Drawing the informal confidence interval onto the graph (give students a copy of the graph, page two of **kiwikapers3studentworksheetPART2.pdf**).

Conclusion



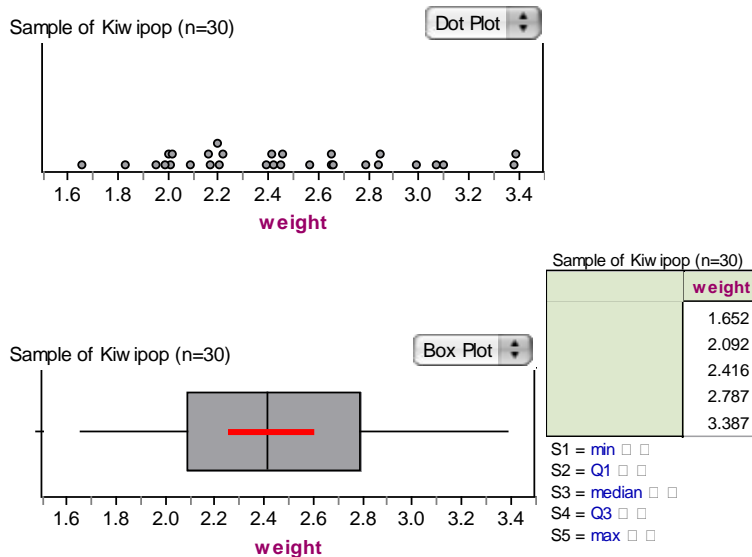
We are now in a position to answer our original investigative question.

P: **What is the median weight of kiwis?**

P: We took a sample of 30 kiwis from the population of kiwis.

D: We recorded the weight of the kiwis.

A:



The distribution of weights of these kiwis is approx symmetrical and unimodal, with a slight tail to the right. The weights range from 1.652 kg to 3.387 kg. The weights peak around 2.2 kg. the middle 50% of the weights lie between 2.1 kg and 2.8 kg. The median weight for this sample of kiwis is 2.416 kg and an informal confidence interval for the population median is from 2.23 kg to 2.61 kg. (Draw this on the box plot – see in red).

C: Based on this sample of 30 kiwis I am pretty sure that the median weight of kiwis is between 2.23 kg and 2.61 kg. I know that if I took another sample of 30 kiwis I would get a different sample median weight and a different informal confidence interval.

Reflection Extension activity

Kiwipop

Dot Plot

