



<http://topdrawer.aamt.edu.au/Statistics/Misunderstandings/Misunderstanding-samples-and-sampling/Increasing-sample-size>

Transcript of video Increasing Sample Size

This demonstration emphasises the importance of sample size and the ability of a sample to represent a population. At the top, we have a population of 813 Australian students and their heights. I've sketched the approximate shape of the distribution over the graph in order for us to keep an idea of the general shape.

(Pointing to the approximately normal distribution of the population plot)

At the right, we have a slider and the slider allows us to take a random sample from this population and display it on the plot at the bottom.

(Pointing to a slider labelled sample size)

I've have made the scales of the two graphs the same so that we can see how the sample represents the population as it increases in size.

(Indicating a single data value on the sample plot)

I am now taking the sample from size 1 up to size 5 and we can see the values are spread out along the axis and we don't get very much of an idea of what the distribution is like.

(Five data points spread out along the sample plot)

I will go up to 10 and we get more values. At 15, it seems that we have more, perhaps more, smaller values than we do larger values in our sample.

(Data points spread out along the sample plot but gathered toward the lower end of the scale)

By a sample size of 20, they're be getting to become more spread out. Watch what happens, 25 and up to 30.



(More values appearing at the middle and upper end of the scale)

Right?

There seem to be more values in the middle now as our sample size increases. Move up. There is 35, similar things happening. And by the time we get to a sample size of 40 I think we have a feeling that the distribution of the values are in fact fairly similar to those of the population.

(Sample plot loosely approximating a normal distribution)

At this point, we could actually look at the mean value of the population and the sample so the mean value is 159.14 for the population. For the sample, it's 156.44. So it's not exact, but it's fairly close and we have an idea of what the distribution would be like.

(Finding the means by selecting the appropriate tool from the software)

So let's watch as I increase the sample size slowly to 100 and we'll do this automatically and you can watch the values increase.

(Values added automatically until a sample size of 100 is reached)

Right, there we are at 100 and we can see that the shape is very similar and the mean is now 159.02, which is very close to the population mean by the time we get to this sample size. What we have demonstrated is that the larger the sample size, the closer the sample represents the population.

(Sample plot with a similar shape to the population plot)