**What is sampling variation?**

[**https://www.stat.auckland.ac.nz/~wild/WPRH/**](https://www.stat.auckland.ac.nz/~wild/WPRH/)

|  |  |
| --- | --- |
| **Two samples dot and box plots** | **Two samples – Box plots with a memory** |
| [**https://www.stat.auckland.ac.nz/~wild/WPRH/Animations/CtsVars\_2samp\_dots\_30.pdf**](https://www.stat.auckland.ac.nz/~wild/WPRH/Animations/CtsVars_2samp_dots_30.pdf) | [**https://www.stat.auckland.ac.nz/~wild/WPRH/Animations/boxes\_2samp\_mem\_30.pdf**](https://www.stat.auckland.ac.nz/~wild/WPRH/Animations/boxes_2samp_mem_30.pdf) |
| [**https://www.stat.auckland.ac.nz/~wild/WPRH/Animations/CtsVars\_2samp\_dots\_100.pdf**](https://www.stat.auckland.ac.nz/~wild/WPRH/Animations/CtsVars_2samp_dots_100.pdf) | [**https://www.stat.auckland.ac.nz/~wild/WPRH/Animations/boxes\_2samp\_mem\_100.pdf**](https://www.stat.auckland.ac.nz/~wild/WPRH/Animations/boxes_2samp_mem_100.pdf) |
| [**https://www.stat.auckland.ac.nz/~wild/WPRH/Animations/CtsVars\_2samp\_dots\_300.pdf**](https://www.stat.auckland.ac.nz/~wild/WPRH/Animations/CtsVars_2samp_dots_300.pdf) | [**https://www.stat.auckland.ac.nz/~wild/WPRH/Animations/boxes\_2samp\_mem\_300.pdf**](https://www.stat.auckland.ac.nz/~wild/WPRH/Animations/boxes_2samp_mem_300.pdf) |



**Sample Variability** - how each sample varies from other samples (but is similar)

**Sampling Variation** – the variation (changes) in a sample statistic from sample to sample.

Suppose a sample is taken and a sample statistic, such as a [sample mean](http://nzmaths.co.nz/category/glossary/sample-mean), is calculated. If a second sample of the same size is taken from the same [population](http://nzmaths.co.nz/category/glossary/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.

This means if we were to take another sample from the population this is likely to result in different displays and summary statistics.

**Bootstrapping**

**Goal**: Take repeated samples to allow for the effects of sampling variation

**Problem**: Repeated sampling from a population may be impractical, expensive or not possible.

**Solution**: If we cannot resample from the population, (the true sampling distribution is unavailable) then we resample from the best approximation of the population we have - which is the sample itself (producing a bootstrap distribution)

**** So instead we take repeated re-samples from the original sample. We then use these re-samples to calculate an estimate for the population statistic (mean or median). This is called Bootstrapping. A sample size of at least 5 values is required. To calculate the bootstrapping confidence interval many (1000+) re-samples are needed. We do this using iNZight.

Assuming our sample was representative of the population then the bootstrapped confidence interval can be used as an estimate of the true population mean (or median).

**Everybody stand up…** We are a sample of students in NZ, Y13 students, Y13 statistics students?

**Bootstrapping using iNZight**

|  |  |  |  |
| --- | --- | --- | --- |
| 1. iNZite Bootstrap confidence intervals
 | **http://maths.nayland.school.nz/Year_13_Maths/3.10_Inference/Home_images/ScreenShot016.gif** | 9. Include Bootstrap Distribution | **http://maths.nayland.school.nz/Year_13_Maths/3.10_Inference/Home_images/ScreenShot024.gif** |
| 1. Import data
 | **http://maths.nayland.school.nz/Year_13_Maths/3.10_Inference/Home_images/ScreenShot017.gif** | 1. Repeat 100 times and...
 | **http://maths.nayland.school.nz/Year_13_Maths/3.10_Inference/Home_images/ScreenShot025.gif** |
| 1. Add a Variable
 | **http://maths.nayland.school.nz/Year_13_Maths/3.10_Inference/Home_images/ScreenShot018.gif** | 1. Show the CI to see the Bootstrap confidence interval.

This is the interval within which we can be reasonably sure the population median lies. | **http://maths.nayland.school.nz/Year_13_Maths/3.10_Inference/Home_images/ScreenShot026.gif** |
| 1. Analyse
 | **http://maths.nayland.school.nz/Year_13_Maths/3.10_Inference/Home_images/ScreenShot019.gif** |
| 1. Mean or median
 | **http://maths.nayland.school.nz/Year_13_Maths/3.10_Inference/Home_images/ScreenShot020.gif** |
| 1. Record your choices
 | **http://maths.nayland.school.nz/Year_13_Maths/3.10_Inference/Home_images/ScreenShot021.gif** |
| 1. Resample and animation
 | **http://maths.nayland.school.nz/Year_13_Maths/3.10_Inference/Home_images/ScreenShot022.gif** |
| 1. From this sample select a single item, record it, with replacement, repeat selection process 'n' times then calculate the median (or mean)
 | **http://maths.nayland.school.nz/Year_13_Maths/3.10_Inference/Home_images/ScreenShot023.gif** Top box plot is of the sampleBottom box plot is of the resample and the median of the resample (Bootstrap) | Now discuss these results What does the bootstrap confidence interval mean? |