

## 5.13 Exercise: Sampling variation - numeric data (VITonline version)

In this exercise you will begin using [VITonline](https://www.stat.auckland.ac.nz/~wild/VITonline/) the online version of the Visual Inferential Tools (VIT) software. VITonline will enable you to see the variation in estimates from sample to sample and illustrate how the variation changes with different sample sizes.

[VITonline: <https://www.stat.auckland.ac.nz/~wild/VITonline/> ]

VIT is a developing collection of software modules for use in exploring the core concepts involved in drawing conclusions from data.

The skill addressed is:

Learn how to use the VITonline's sampling variation module to produce sampling-variation animations like those used in the Random Error video.

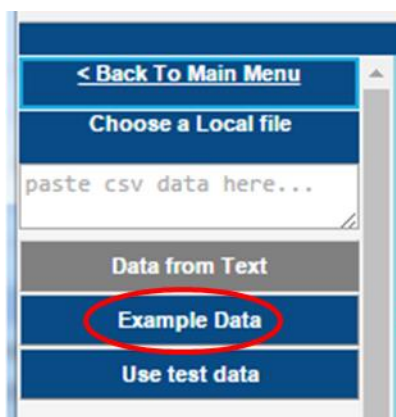
### INSTRUCTIONS

*Do not use Internet Explorer.* Use Chrome, Firefox or Safari.

Go to <https://www.stat.auckland.ac.nz/~wild/VITonline/> to start up [VITonline](https://www.stat.auckland.ac.nz/~wild/VITonline/)

To investigate sampling variation we need a population to sample from and a variable to look at. We will use the **Blood Pressures** of people in NHANES-1000.

- Select the **Sampling variation** option.



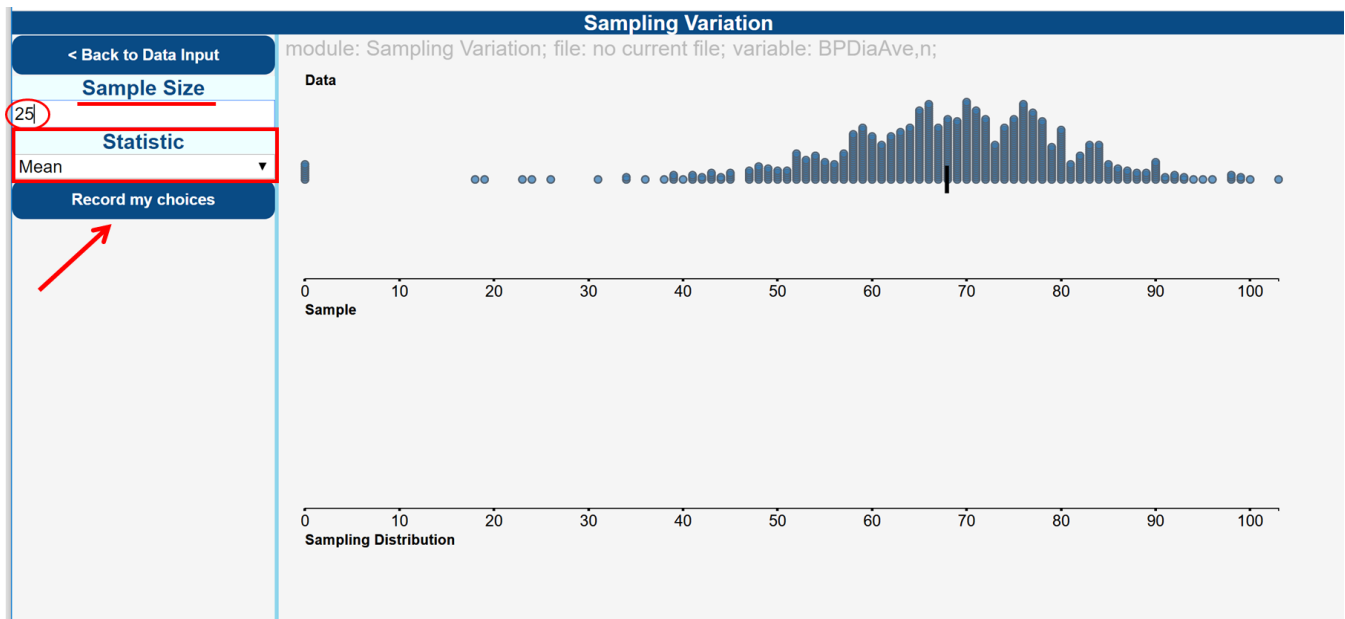
- Inside the Sampling Variation module, click on **Example Data**

Then a list of data sets will appear ...

- Click on NHANES-1000.csv
  - A variable list will appear
- Select BPDiaAve
- Click Analyse (under the variable list)

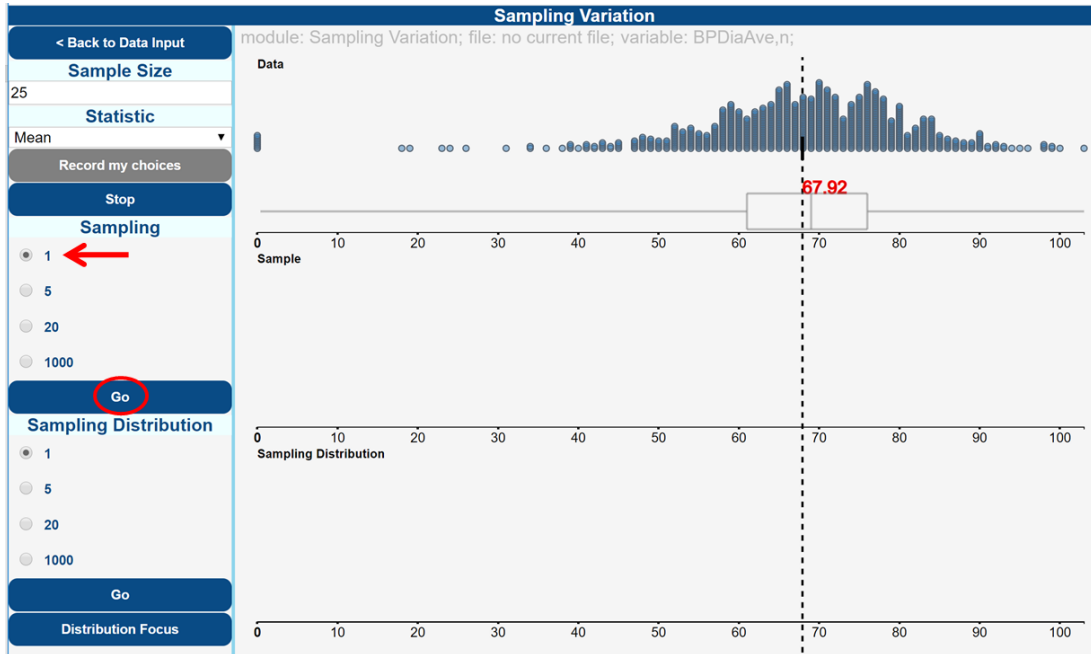
The left-hand command panel will change ...

- For **Sample Size**, use 25 for this example
- For **Statistic**, select **Mean** (the default)
- Now click **Record my choices**.

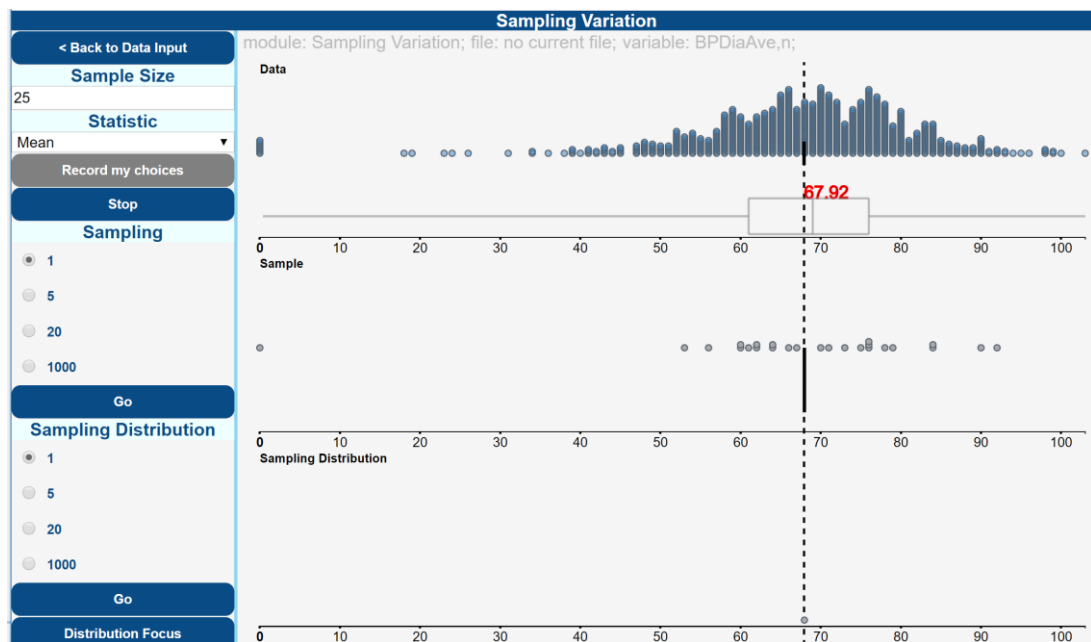


You should now see the below. Make your selections in the middle part of the window under **Sampling**.

- **Select 1** repetition
- Click **Go**.



You will see a random sample of 25 values being selected from the NHANES-1000 population. They will drop down onto the middle graph labelled **Sampling**.



## What is happening in the graphics window?

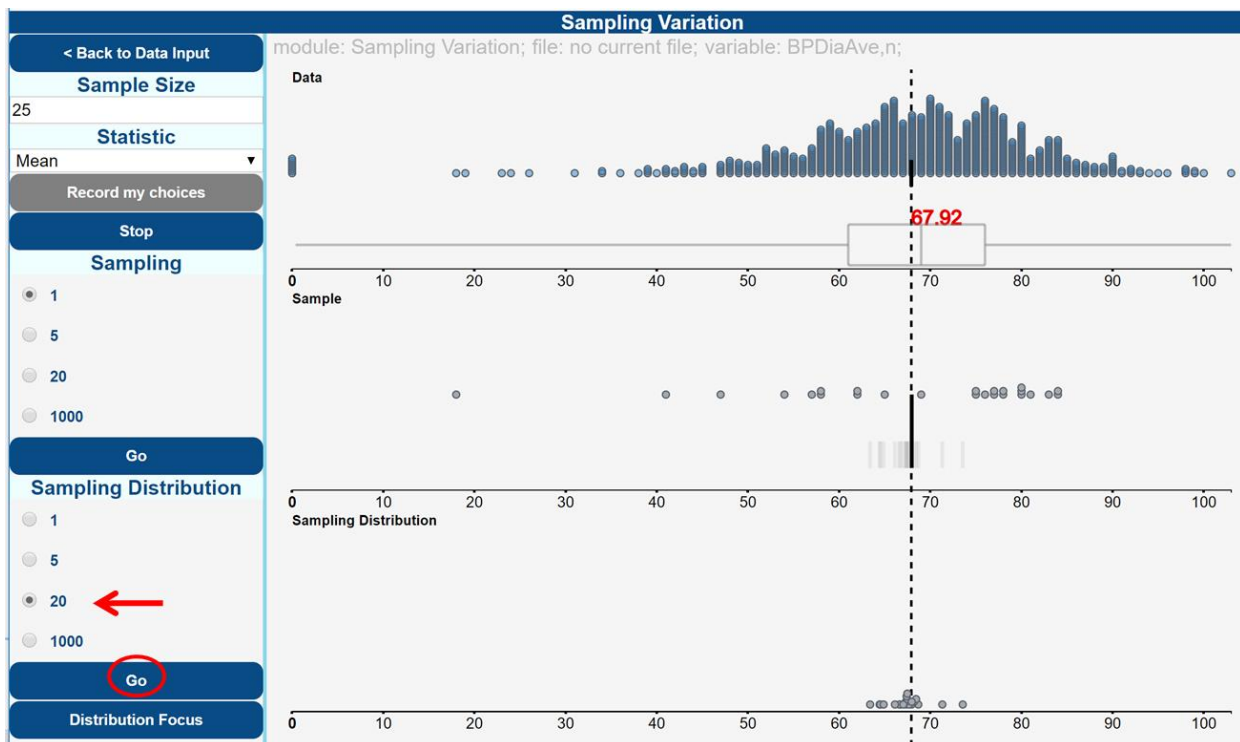
VIT has taken a sample of size 25 from the population of 1000 and calculated the mean of the sample. The sample and the position of its mean (black line) are evident on the middle plot.

- Now **select 5** repetitions
- Click **Go**.

This will generate 5 different samples of 25 and calculate the mean of each sample. A black line will appear and leave a grey "footprint" behind for the mean of each sample taken.

Now we'll use the bottom part of the **Analyse** window that says **Sampling Distribution**.

- Under Number of repetitions, **select 5** and click **Go**.
- Then, under Number of repetitions, **select 20** and click **Go** to collect 20 more samples



## What is happening in the graphics window?

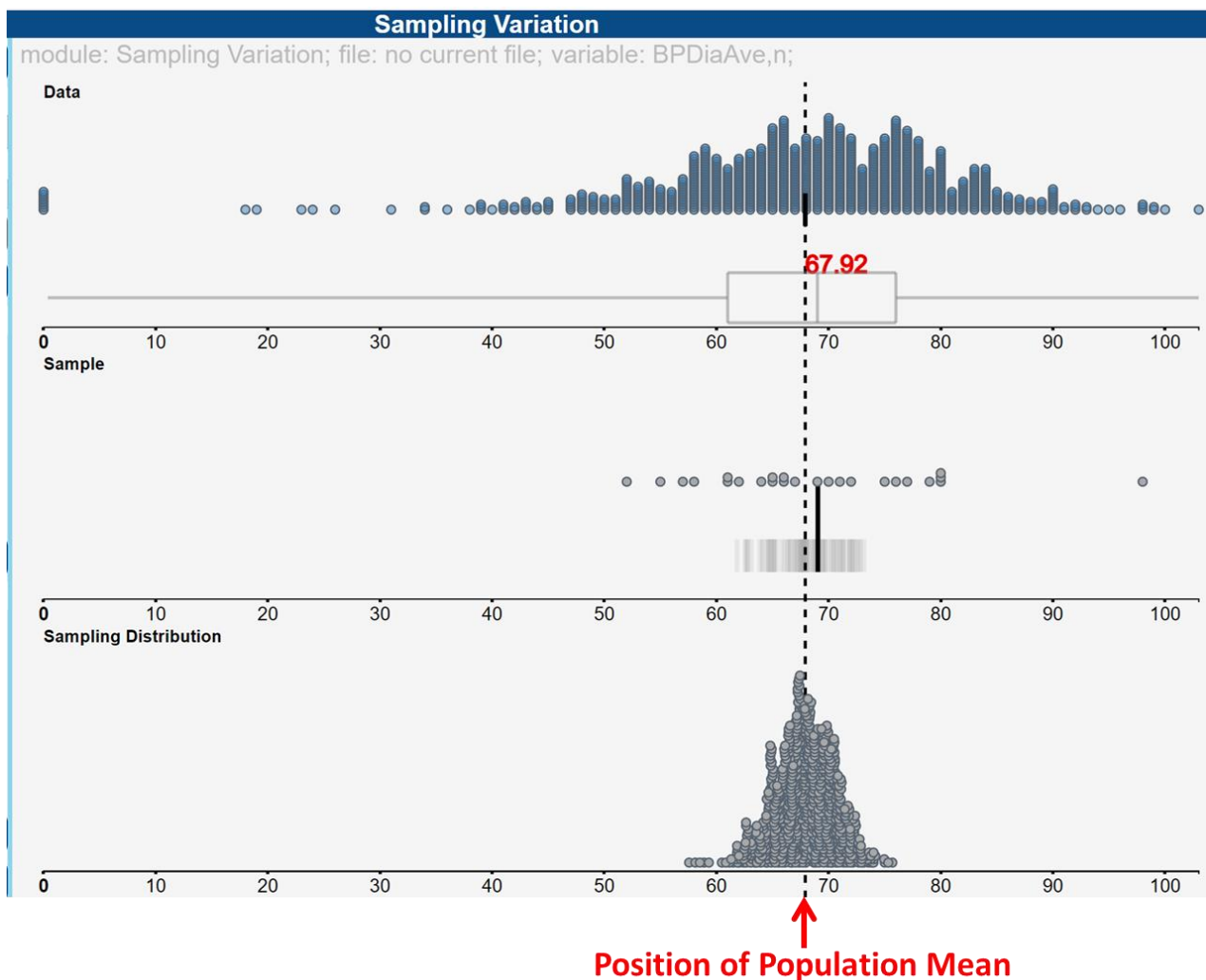
You will see many different samples of size 25 taken from the population. A mean is calculated for each of these samples and those means are shown by grey lines in the middle plot and circles in the **Sampling Distribution** plot.

The dotted line is the mean of the population 67.92 (which you would not usually know if you were sampling from a population).

Try taking 1000 different samples and recording their population means:

- Under the heading **Sampling distribution**, Select **1000**
- Click **Go**.

You will see a sampling distribution of the means of 1000 different samples taken from the population build up in the bottom plot.



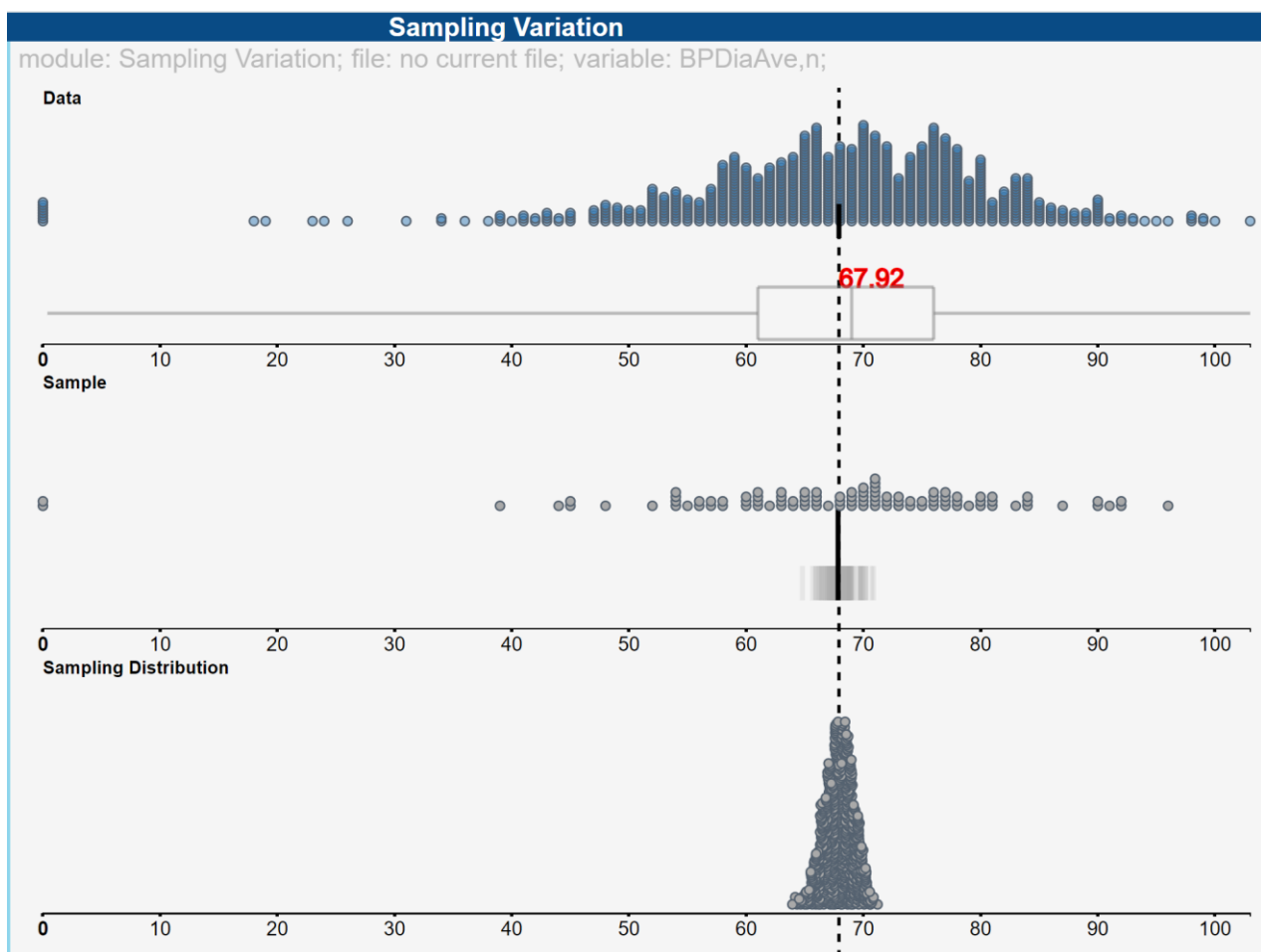
We can see that using samples of size 25, there is sampling error, i.e. the differences between the mean of each sample and the population mean. We ended up with estimates of the population mean, **BPDiaAve**, between about 60 and 75.

Your sampling distribution will not look exactly like this one but it should be quite similar.

## Exploring the effect of increasing the sample size

With VITonline we can easily explore the effects of changing the size of our samples.

- Change your **Sample size** to 100 (4 times the previous amount) and press **Enter** (on your keyboard)
- Click **Precalculate Display**.
- Under the Command panel **Sampling Distribution**, select 1000
- Click **Go**.



## What is happening in the graphics window?

1000 different samples of size 100 are being taken from the NHANES-1000 population. The mean of each of those samples is being recorded as a grey line on the **Sample** plot.

Compare the sampling distribution for sample sizes of 100 and 25 and post a comment stating anything interesting from the two sets of graphs.

## Common questions

### *Does it matter which Go button I press?*

Yes it does. The first **Go** button sets off animations that only happen in the middle graphics panel. The second **Go** button causes the sampling distribution to be formed in the bottom graphics panel.. It will use the number of repetitions selected above the button that you press.

### *What happens if I keep pressing Go?*

VIT will sample again and add the new samples to the current plot.

*(Mac retina) The text and/or points in the VIT modules are too large/unreadable*

See [iNZight FAQ entry](#).