5.13 Exercise: Sampling variation - numeric data

In this exercise you will begin using the Visual Inferential Tools (VIT) software. VIT will enable you to see the variation in estimates from sample to sample and illustrate how the variation changes with different sample sizes.

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 VIT sampling variation to produce sampling-variation animations like those used in the Random Error video.

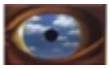
[*VITonline version linked here*]

INSTRUCTIONS

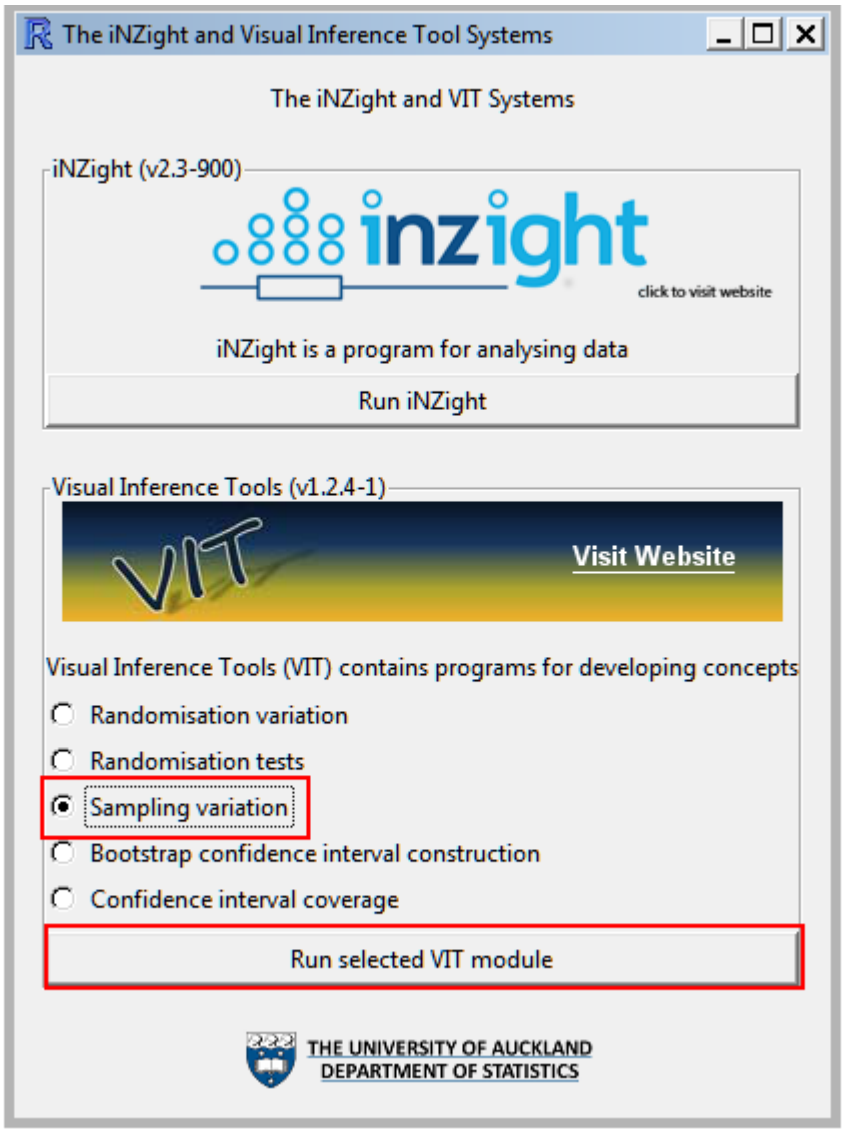
Follow the instructions below to generate the graphs. Or you may prefer to print these instructions.

On Windows: A normal installation of iNZightVIT on Windows leaves a shortcut icon to VIT like that above on your desktop. Double click to open VIT. Otherwise find this shortcut in your iNZightVIT folder wherever you installed it (by default in **Documents**).

On Mac: Use **Go > Applications > iNZightVIT**. The VIT shortcut looks like this.



When VIT opens up you should see:



Select the **Sampling variation** option and click **Run selected VIT module**.

You will notice that the user interface is different. Two windows will open on your screen – the command window is no longer attached to the graphics window. VIT has a separate button next to **View Data** for **Import Data** and above that, there are separate tabs for **Data** and **Analyse**. If you have any problems during this exercise, see **Common questions** at the bottom of this page.

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

Download the NHANES-1000.csv dataset here or NHANES-1000.txt dataset here and put it on your Desktop.

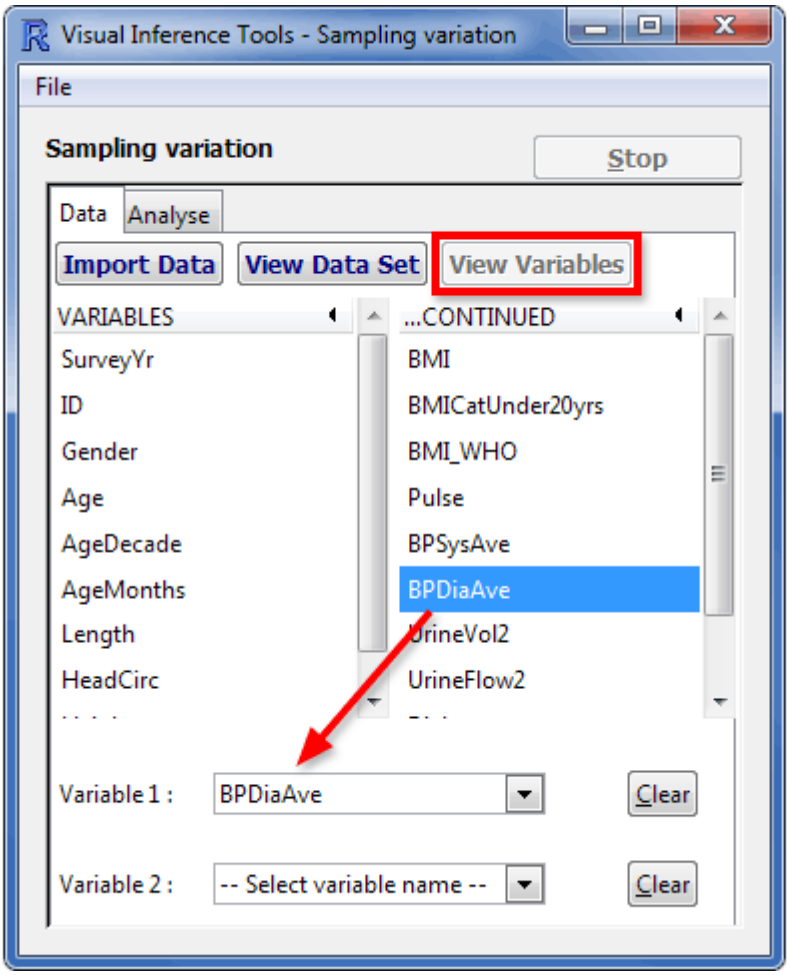
Use **Import Data** and the browser window it opens up to import the **NHANES-1000** dataset into VIT from your Desktop.

Once your dataset has appeared in the window you can click **View Variables** to see a list of the NHANES-1000 variable names.

Using VIT to generate a sample of a numeric variable

We will explore the mean of **BPDiaAve**, a numeric variable.

Drag **BPDiaAve** down into the **Variable 1** slot. You will see the **BPDiaAve** values for the entire NHANES-1000 population plotted on the graphics window.

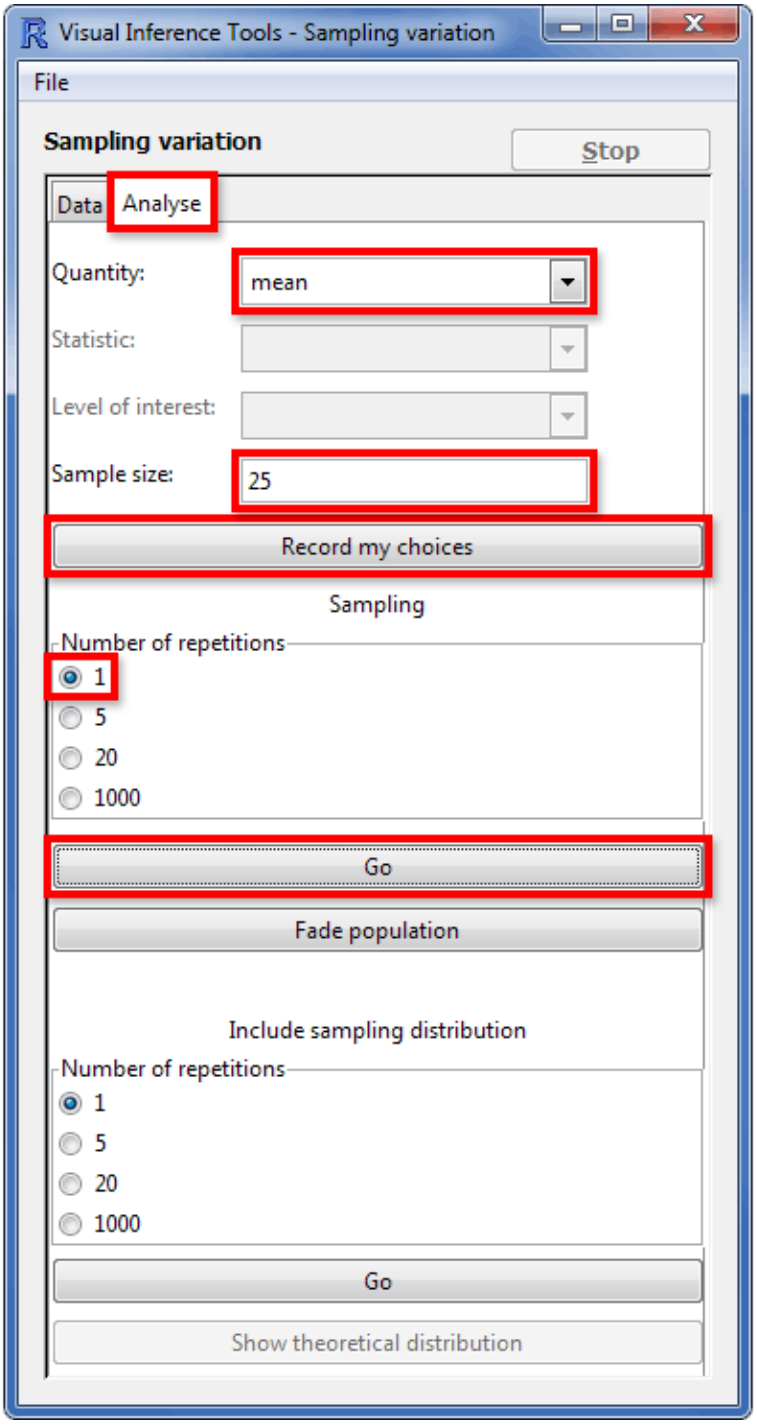


Now click the **Analyse** tab.

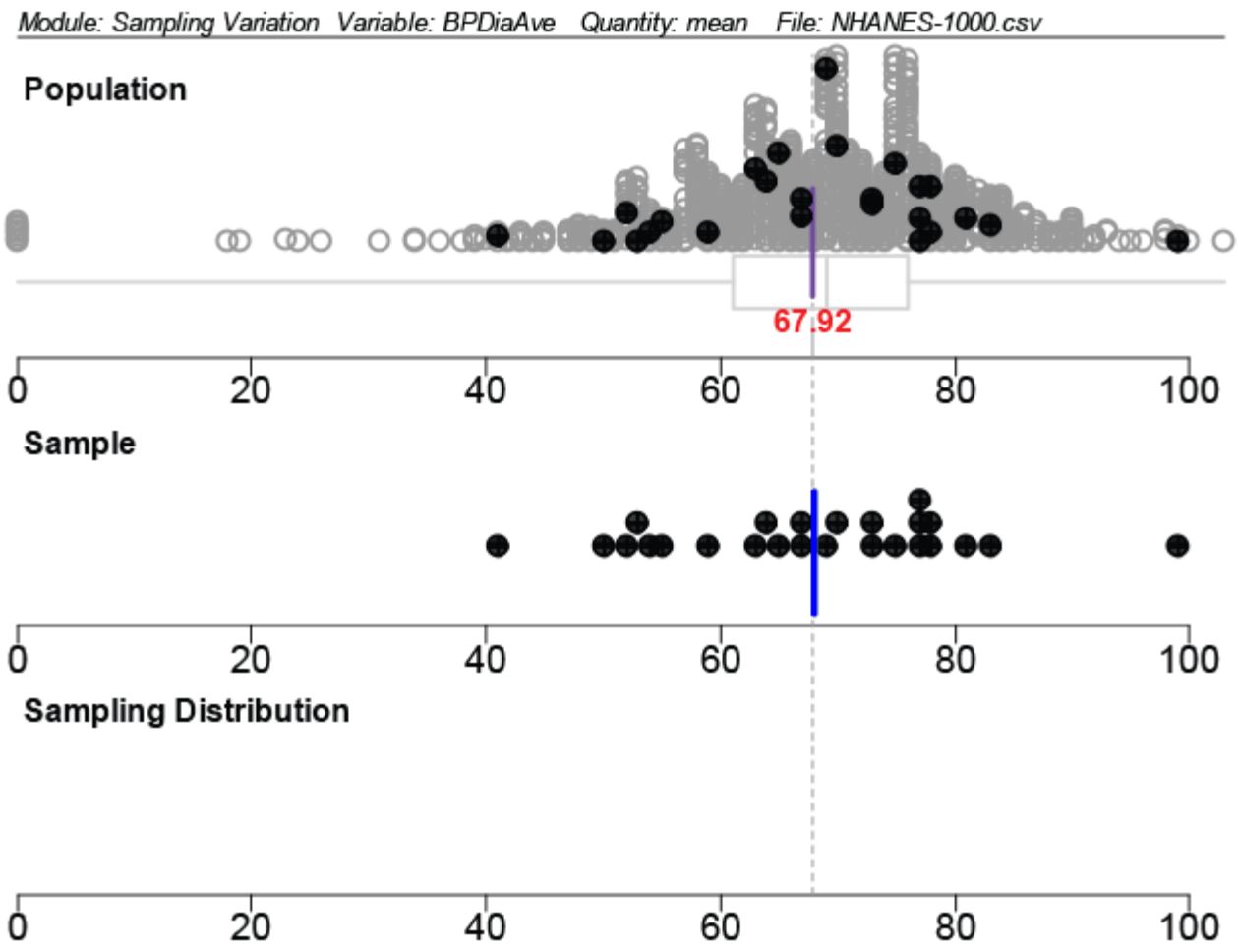
- Beside **Quantity**, select **mean**
- Type in your sample size, use **25** for this example
- Click **Record my choices**.

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 **Sample**.



What is happening in the graphics window?

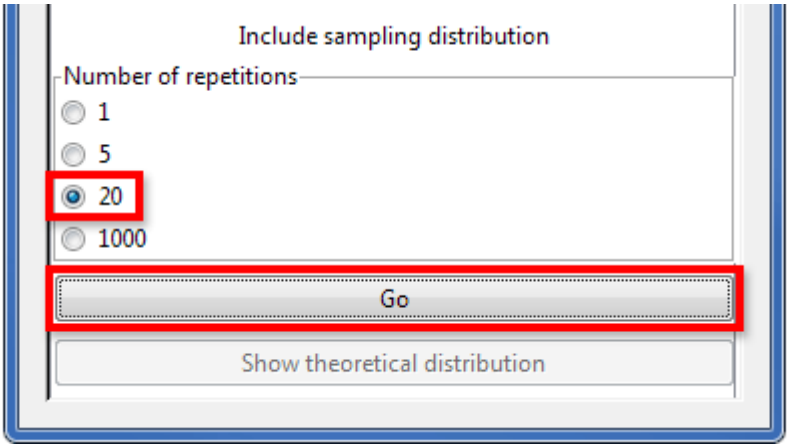
VIT has taken a sample of size 25 from the population of 1000 and calculated the mean of the sample. The sample and the position of its mean (blue line) are evident on the middle plot.

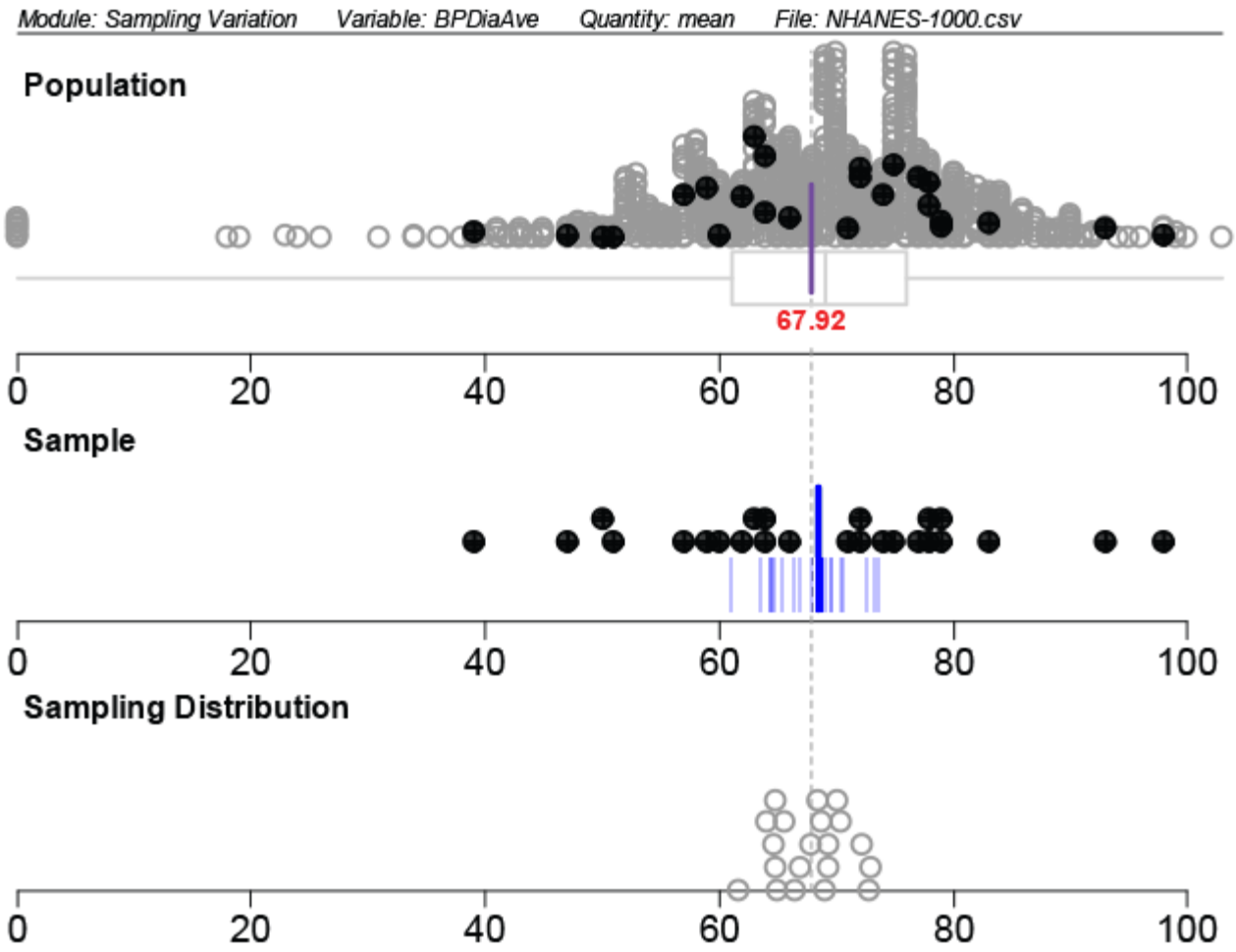
- Now **select 5** repetitions
- Click **Go**.

This will generate 5 different samples of 25 and calculate the mean of each sample. A blue line will appear and remain for the mean of each sample taken.

Now we'll use the bottom part of the **Analyse** window that says **Include sampling distribution**.

- Under Number of repetitions, **select 20** and click **Go**.





What is happening in the graphics window?

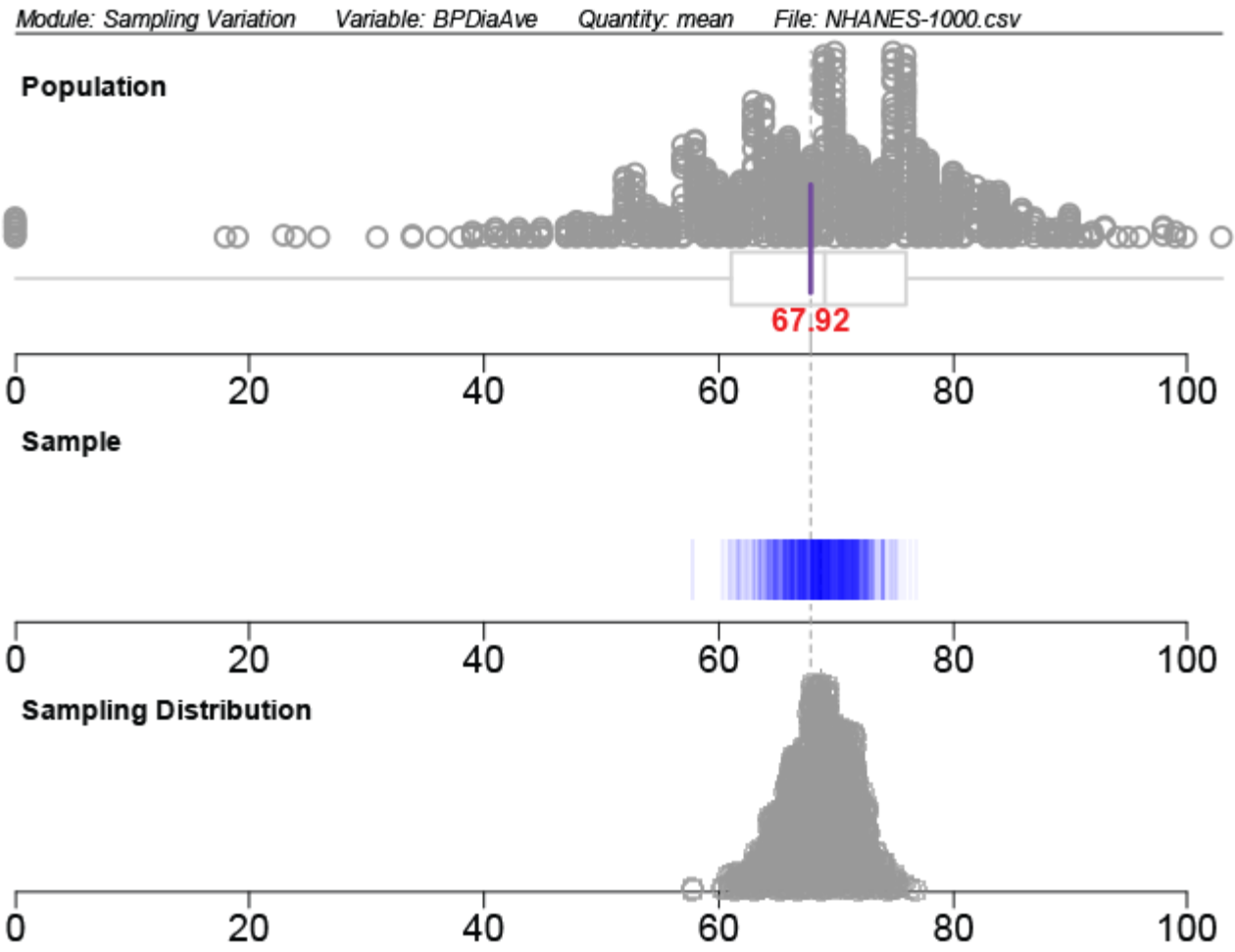
You will see 20 different samples of size 25 taken from the population. 20 different means are calculated from the 20 samples and those means are shown by blue lines in the middle plot and circles in the **Sampling Distribution** plot.

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 Include 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.



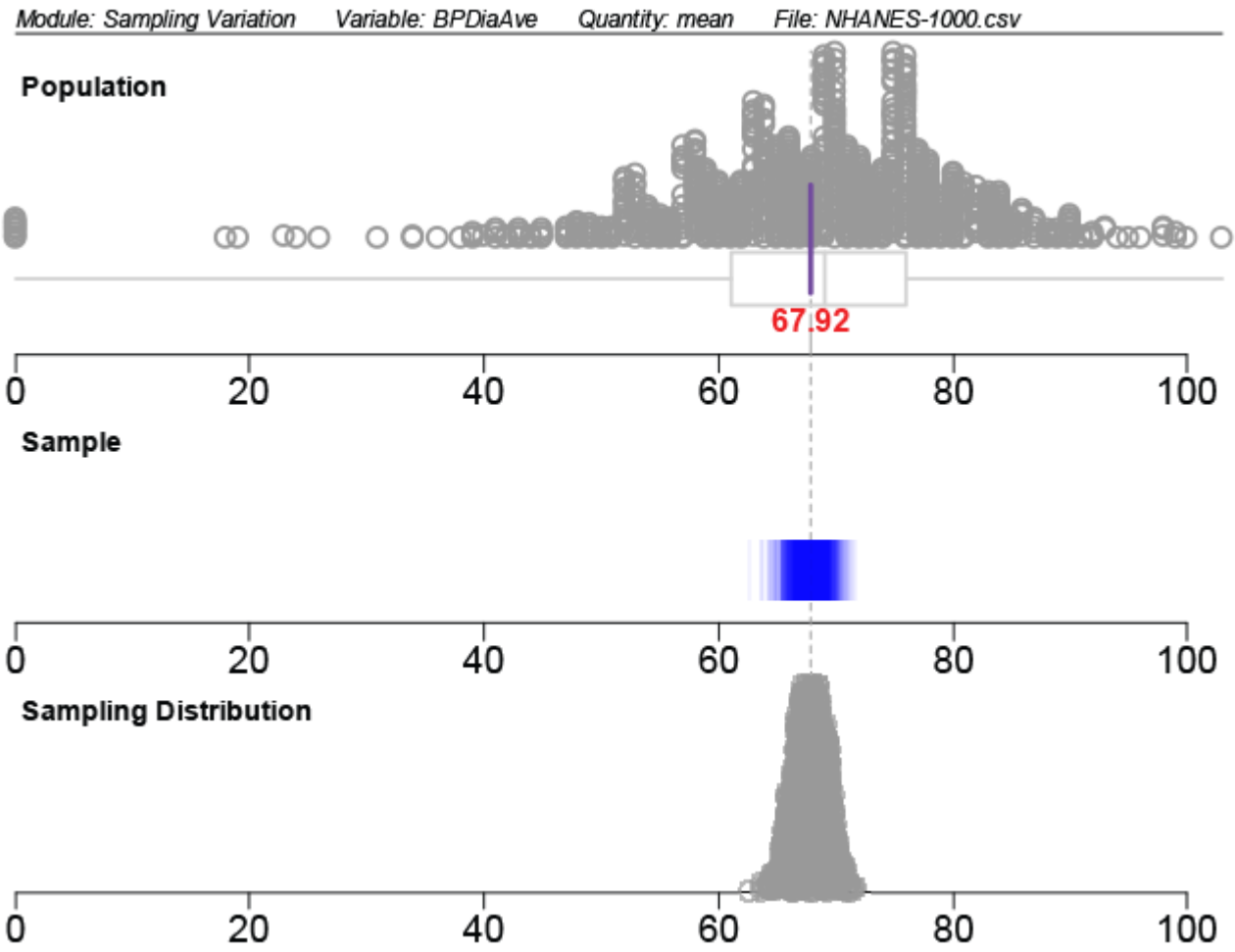
We can see that using samples of size 25, there is sampling error, i.e. variation between samples due to sampling rather than using the entire population. We ended up with estimates of the population mean, **BPDiaAve**, between about 60 and 75.

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

- Change your **Sample size** to **100** (4 times the previous amount)
- Click **Record my choices.**
- Under the heading **Include 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 blue 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.

Now:

- Change the quantity being investigated from **Mean** to **Median.**
- Leave the sample size of **100**
- Click **Record my choices**
- Click **Go**

How does the pattern generated differ? You could also repeat for the first (lower) quartile.

Common questions

When I change the sample size and click Go, the sample size changes back to the old value.
You need to click **Record my choices** to reset the plot window and save your new sample size.

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. The exception is when 1000 repetitions are selected. In that case the old display is cleared before starting.

How do I clear my plot and start again?
Click **Record my choices** to clear the plot window and save the current sample size.

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

See iNZight FAQ entry.