6.5 Exercise: Bootstrap re-sampling

This exercise will enable you to use Visual Inferential Tools (VIT) to construct a Bootstrap Confidence Interval for a single numeric and a single categorical variable.

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

The skills addressed are:

- Use VIT to generate a bootstrap distribution and confidence interval.
- Use VIT to generate a bootstrap confidence interval for a proportion.

[*VITonline version linked here*]

INSTRUCTIONS

Follow the instructions below to generate the graphs. Or you may prefer to print the instructions. If you have a problem doing the exercise, scroll down to **Common questions**.

Start VIT as you have done previously but this time select the **Bootstrap confidence interval construction** option in the VIT section and then click **Run selected VIT module**.
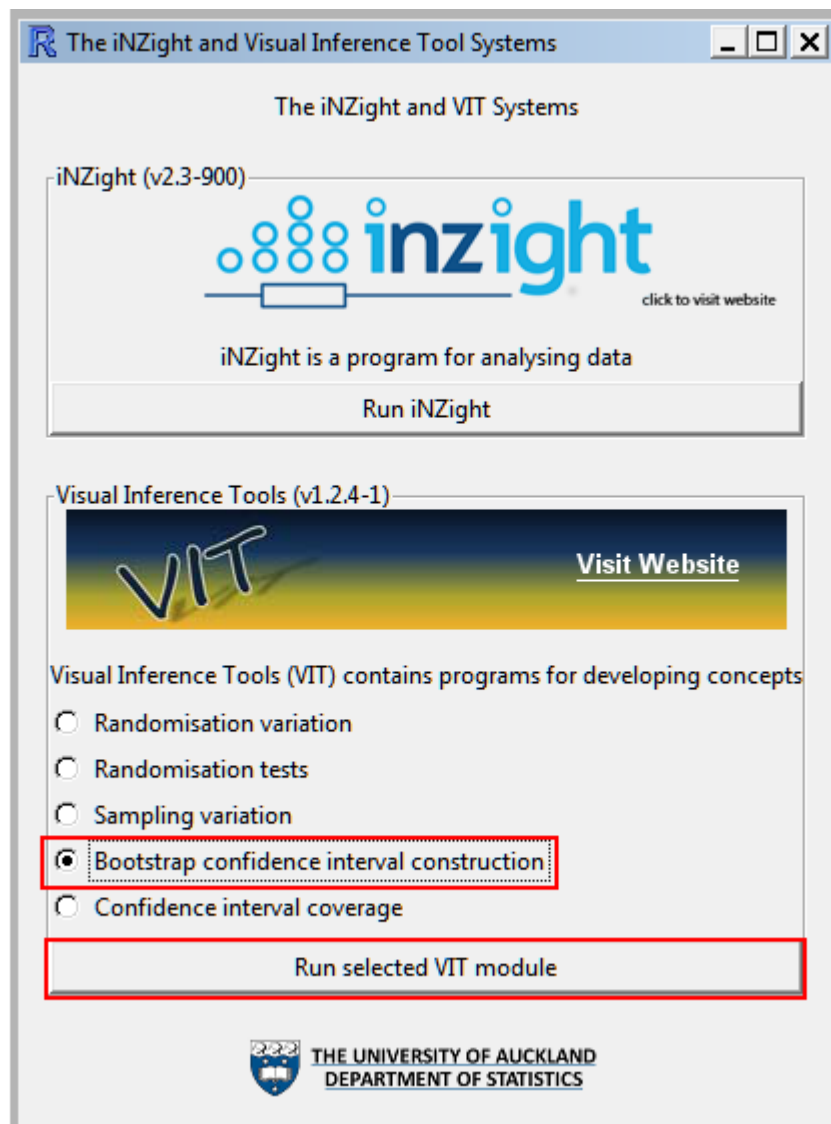
Download the **NHANES-1000-50.csv dataset here** or **NHANES-1000-50.txt dataset here** and put it on your Desktop.

Use **Import Data** and the browser window it opens up to import the **NHANES-1000-50** dataset into VIT from your Desktop. [*Your data set should have only 50 rows.*]

Once your dataset has appeared in the window, *check that your data set only has 50 rows.* [If it has more you probably imported NHANES-1000 by mistake.]

Click **View Variables** to see a list of the NHANES-1000-50 variable names.

Using VIT to generate a sample of a numeric variable

We will use our sample of 50 people to form a Bootstrap Confidence interval for the mean **BMI** of the NHANES-1000 data which we looked at in Week 5 exercises.

Drag **BMI** down into the **Variable1** slot. You will see the BMI values for 49 people in our sample plotted on a graph with a box plot below it. (One person had a missing value for BMI).
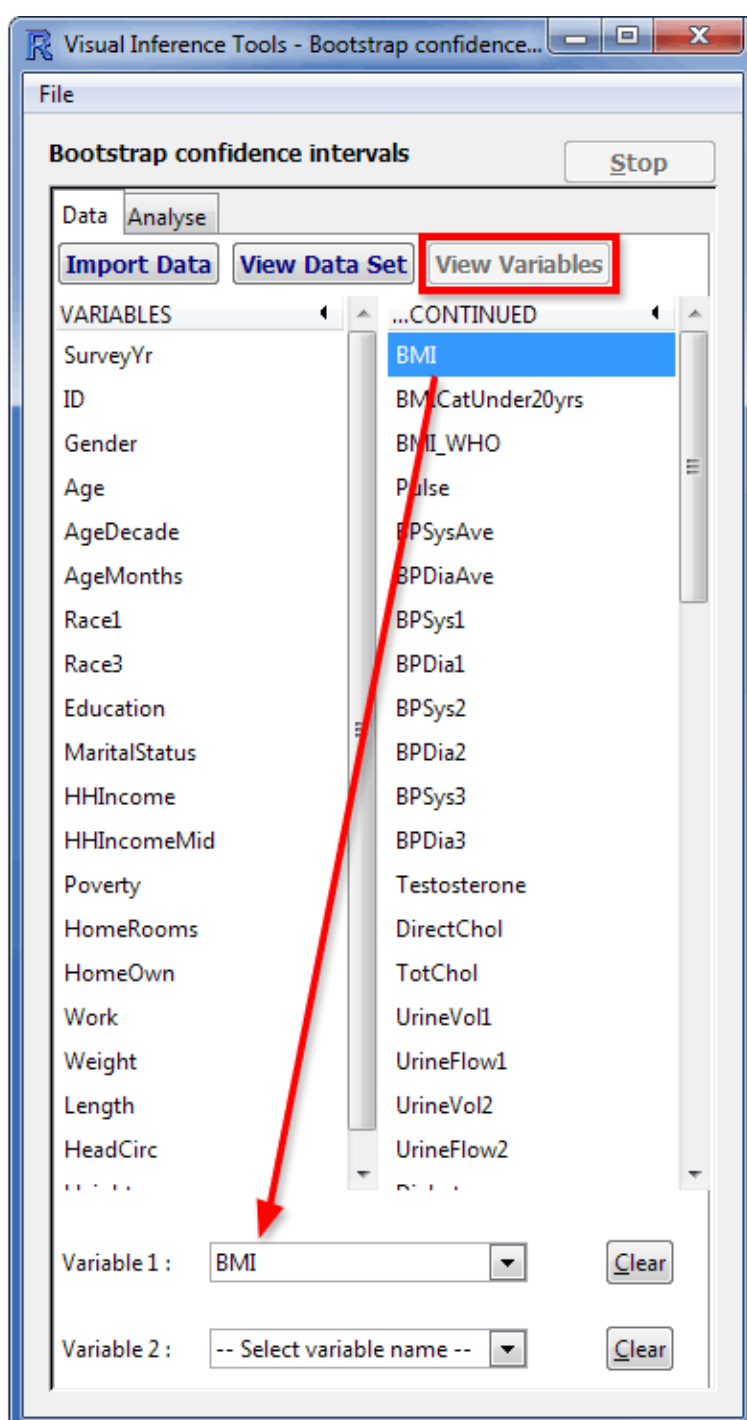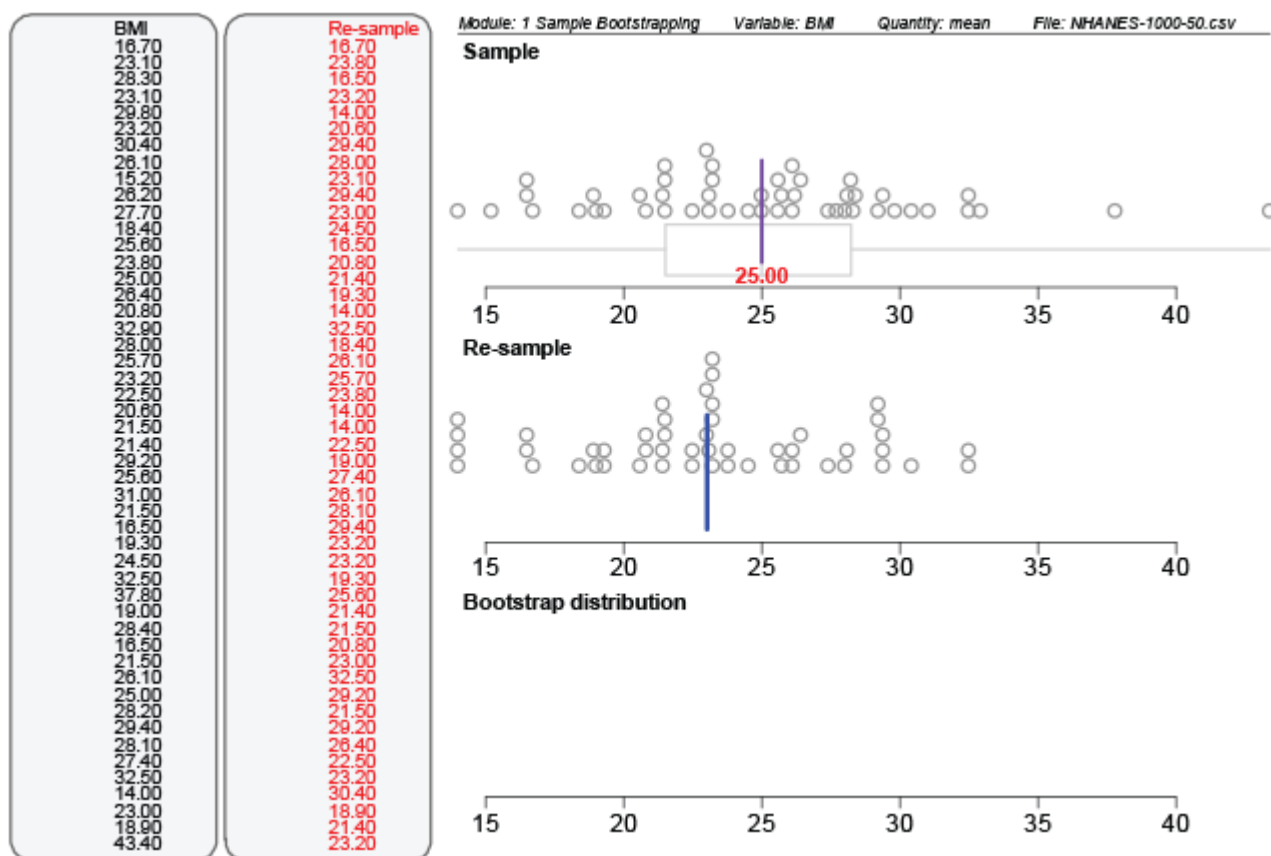
Click the **Analyse** tab.

- Beside Quantity, select **mean**
- Click **Record my choices**

Next, make your selections in the middle part of the window under **Re-sampling**.

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

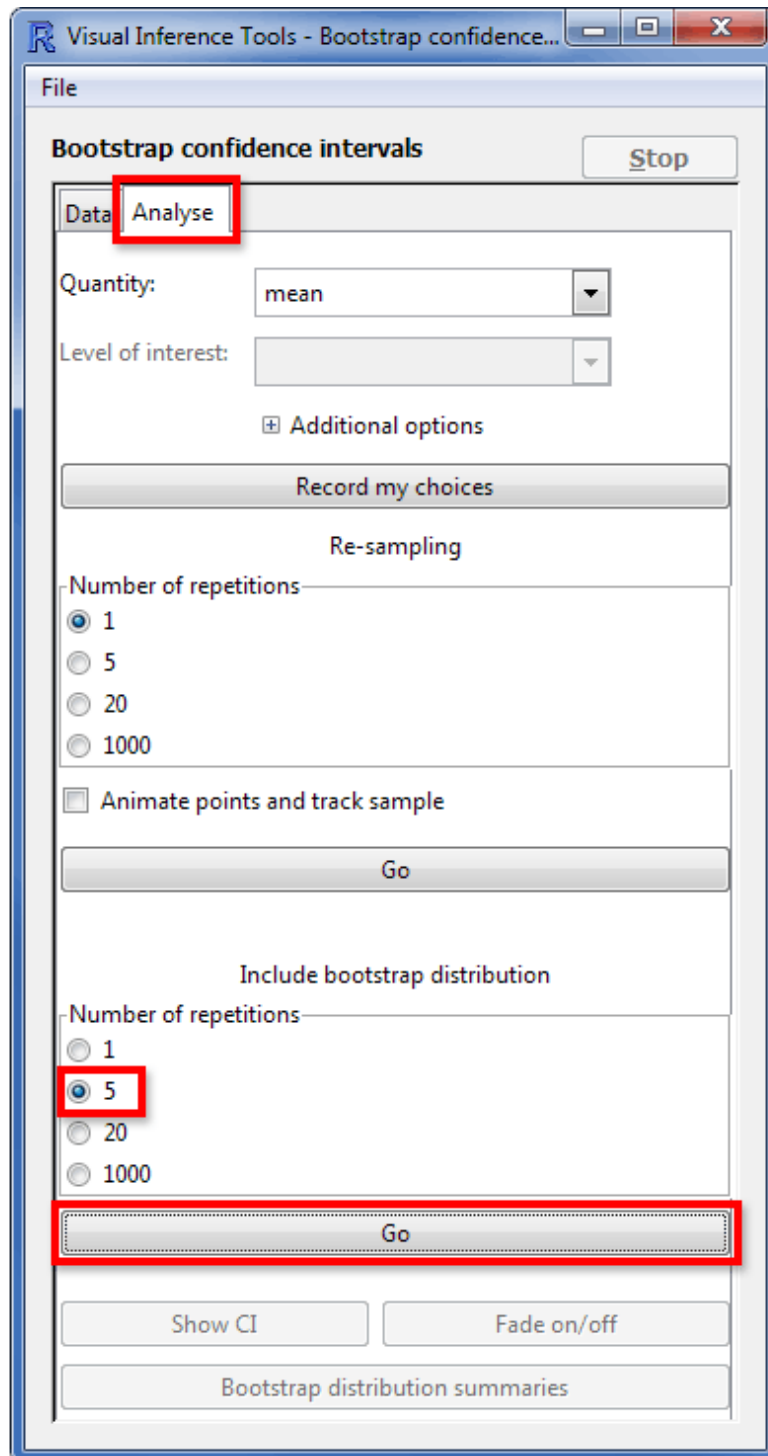You will see a re-sample of size 49 from the original sample being selected in red.

[Enlarge image](#)

What is happening in the graphics window?

A plot of the **Re-sample** will appear in the middle part of the graphics window.

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

- **Select** 5 repetitions and click **Go**.

This will generate 5 re-samples of size 49 and the mean of each re-sample will be calculated and be shown by a blue line in the **Re-sampling** plot. A disc will be generated from the mean of each re-sample and will drop down into the **Bootstrap distribution** plot.
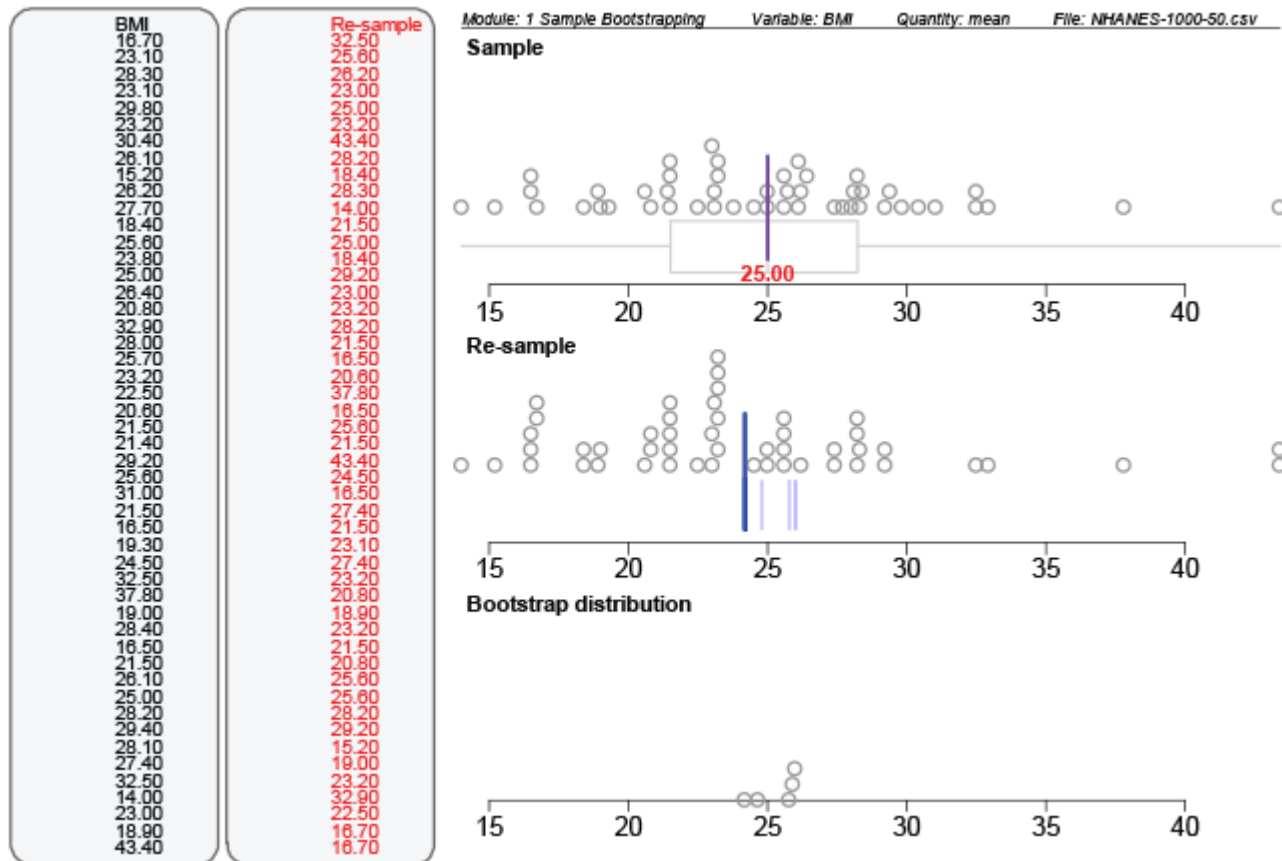
[Enlarge image](#)

The above plot shows 5 re-samples have been taken so far because there are 5 circles in the **Bootstrap distribution** plot.
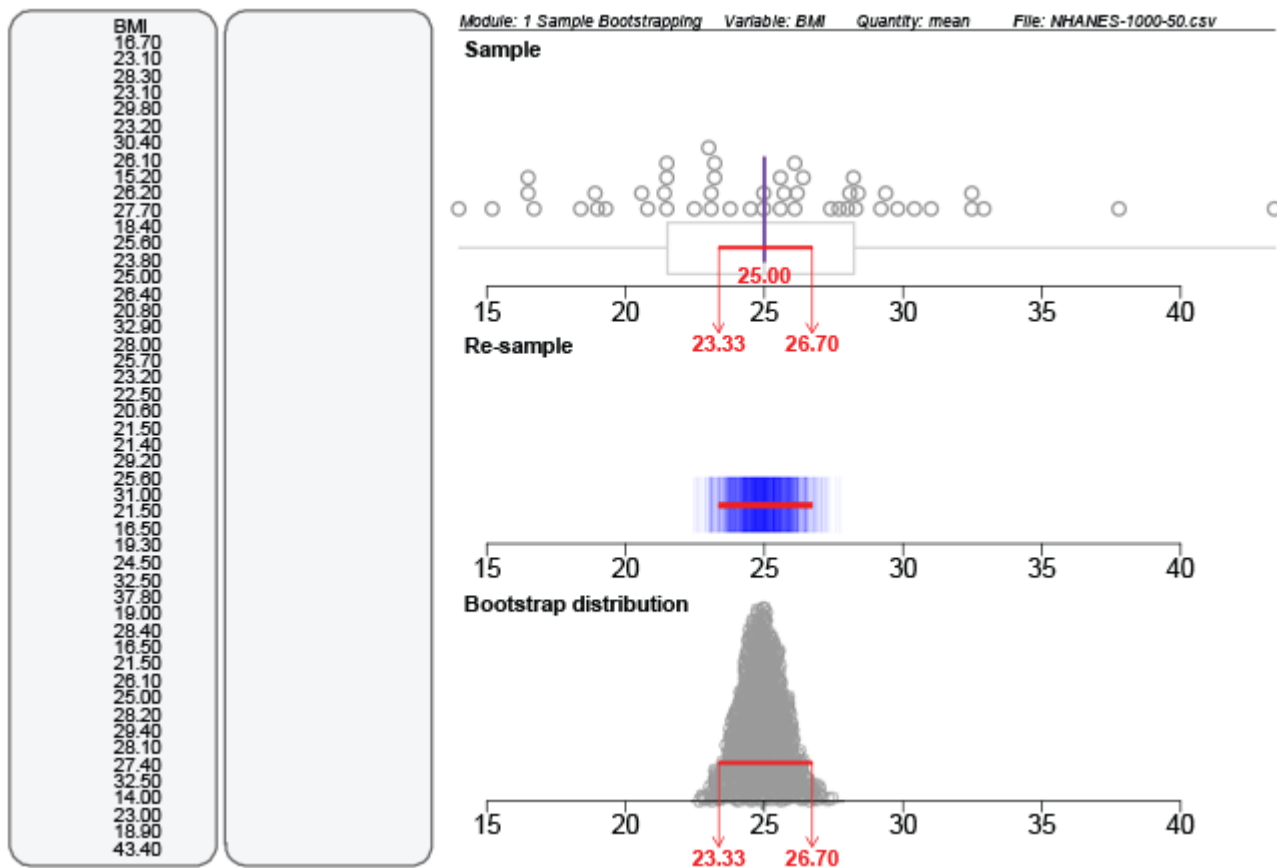
In the bottom part of the **Analyse** window (**Include bootstrap distribution**)

- **Select** 1000 repetitions click **Go**.

You will see a **Bootstrap distribution** of the means of 1000 re-samples taken from the initial sample of size 49.

Get the Bootstrap Confidence Interval

- Click **Show CI**.

[Enlarge image](#)

What is happening in the graphics window?

The **Bootstrap distribution** is a build-up of 1000 re-sample means and shows the variation between re-sample means.

The Bootstrap Confidence Interval has trimmed off the bottom and top 2 ½% of the re-sample means and appears in red (23.33, 26.70). This interval will appear on the two plots above as well.

The resulting interval is the values within which we are fairly sure that the mean BMI value for the NHANES-1000 population lies.

Note that your values will probably be slightly different.

Construct a Bootstrap Confidence Interval for a categorical outcome (proportion)

Now, use the sample of size 49 to generate a bootstrap distribution for estimating the proportion of people in the NHANES-1000 population who are classified as obese (BMI over 30) using the variable BMI_WHO (levels of BMI as decided by the World Health Organisation).
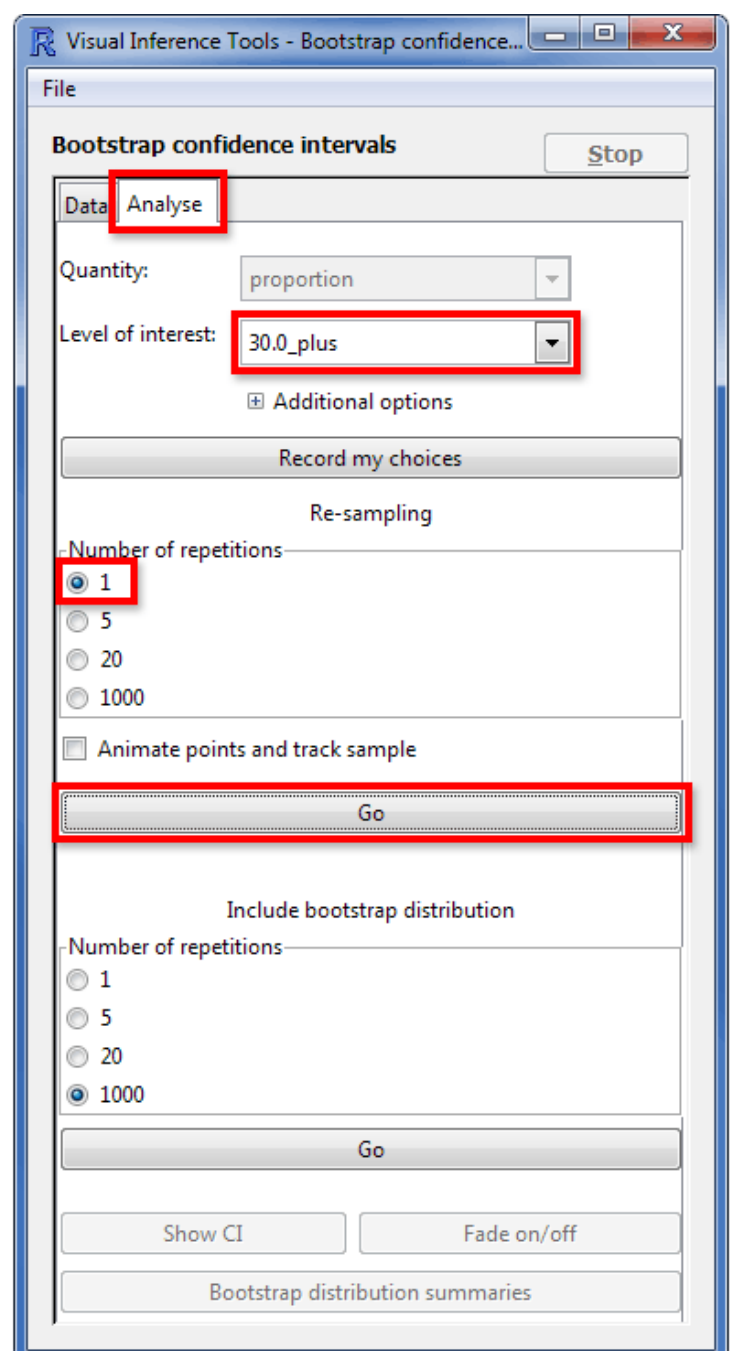
Under the **Data** tab

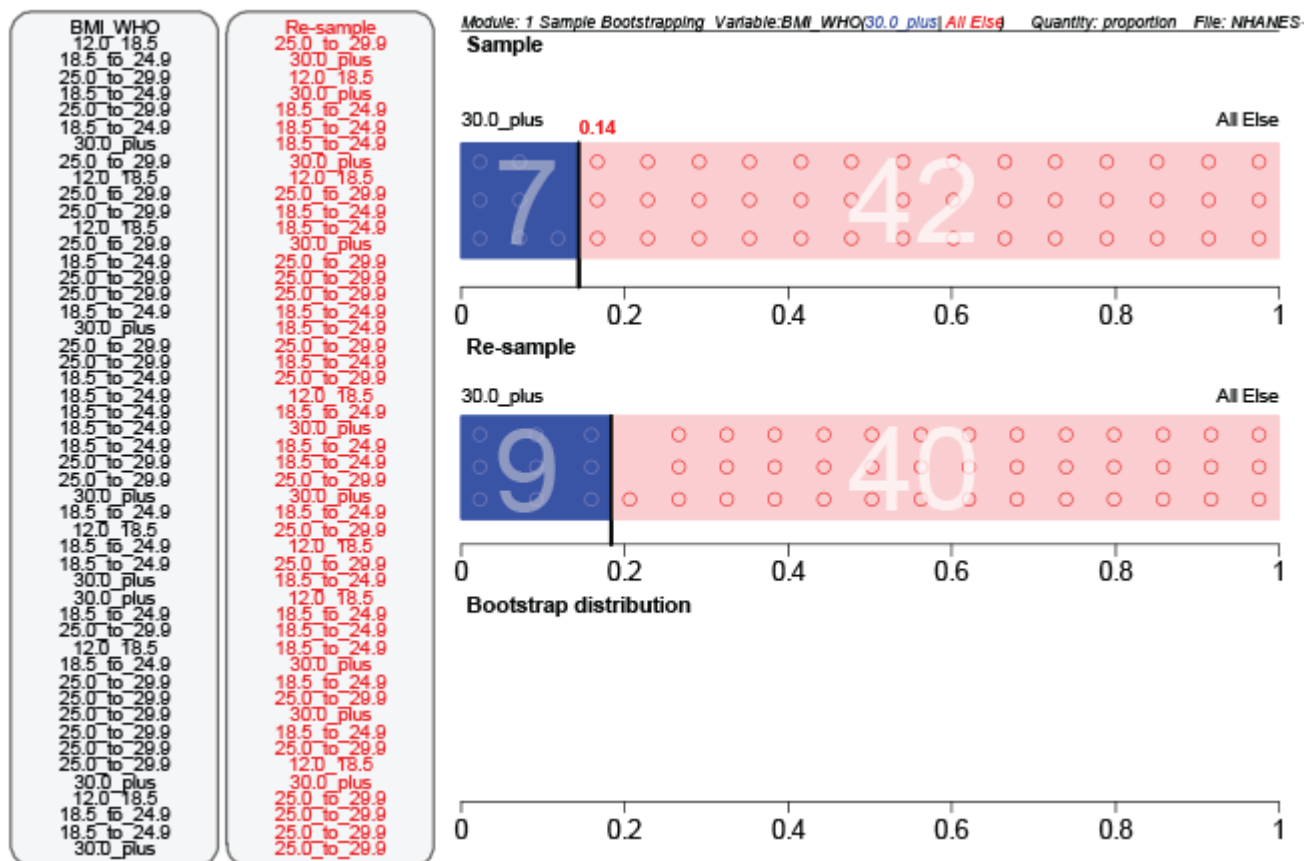- Drag variable **BMI_WHO** into the **Variable 1** slot

Under the Analyse tab

- Click **Level of interest** and select **30.0_plus** (BMI's over 30.0 which are regarded as Obese)
- Click **Record my choices**

Under the **Re-sampling** section

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

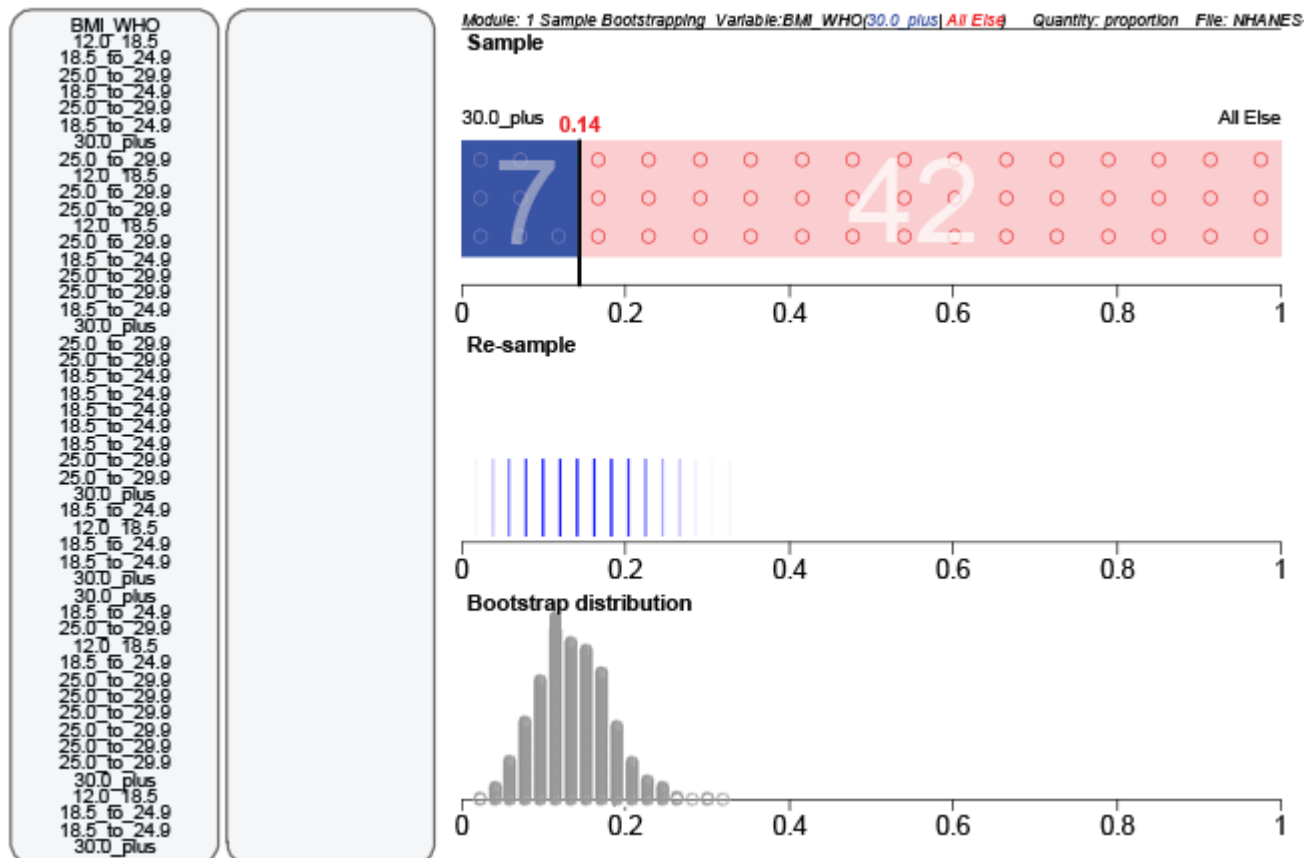[Enlarge image](#)

What is happening in the graphics window?

A re-sample of size 49 is being taken from the original sample.

Under the **Include bootstrap distribution** section

- **Select** 1000
- Click **Go**.

Once the 1000 re-samples have been selected and the bootstrap distribution completed:

- Select **Show CI**.

[Enlarge image](#)

What proportion of participants in the NHANES-1000 population do we expect to be classified as "Obese"?

EXPLORE (~5 min)

What other numeric or categorical variables can you form a bootstrap confidence interval for using your sample of 50 NHANES participants? How do your results compare with the sampling variations in the previous week?

Post a comment about anything interesting you have found.

Common questions

What does CI mean?

CI stands for Confidence Interval. If you click the **Show CI** button, a 95% bootstrap confidence interval will show on the plots.

What is Fade on/off?

You can only click **Fade on/off** after you have put a confidence interval on your plots. Then the re-sampling and bootstrap distributions will fade and you'll be left with your initial estimated proportion from your initial sample with a 95% bootstrap confidence interval superimposed on it. Fading may not work if you have resized your plotting window.

Do I have to do the animation?

No. The animation is there for you to see how each re-sample is selected from the initial sample.

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

See [iNZight FAQ entry](#).