

## 5.14 Exercise: Explore sampling variation - proportions (VITonline version)

*(The graphics in this exercise were corrected 8 Nov after VITonline was changed to handle missing values correctly)*

In this exercise you will begin using [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.

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 VITonline's Sampling Variation module to produce a sampling distribution for proportions like those used in the Random Error video.

### INSTRUCTIONS

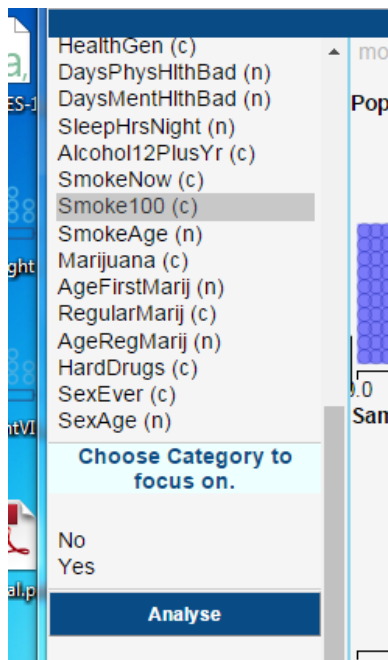
*Do not use Internet Explorer. Use Chrome, Firefox or Safari.*

Start [VITonline](#) select the **Sampling variation** option and load the NHANES-1000 data set from **Example Data** as instructed in the previous Exercise (Step 5.13).

See that the numeric variables have "(n)" after their names and the categorical variables have "(c)" after their names. In this exercise, we are looking at categorical variables.

If you have any problems during this exercise, see **Common questions** at the bottom of this page.

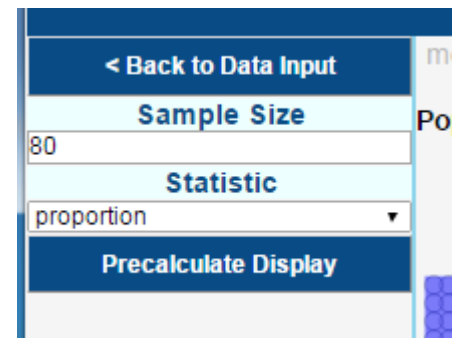
We will explore the variable **Smoke100** which is the proportion of people in the NHANES-1000 dataset who have smoked more than 100 cigarettes in their lifetime.



- Click on **Smoke100**
- At the bottom of the variable list under **Category** to focus on choose **Yes**.
- Then click the **Analyse** button

The left command panel will change

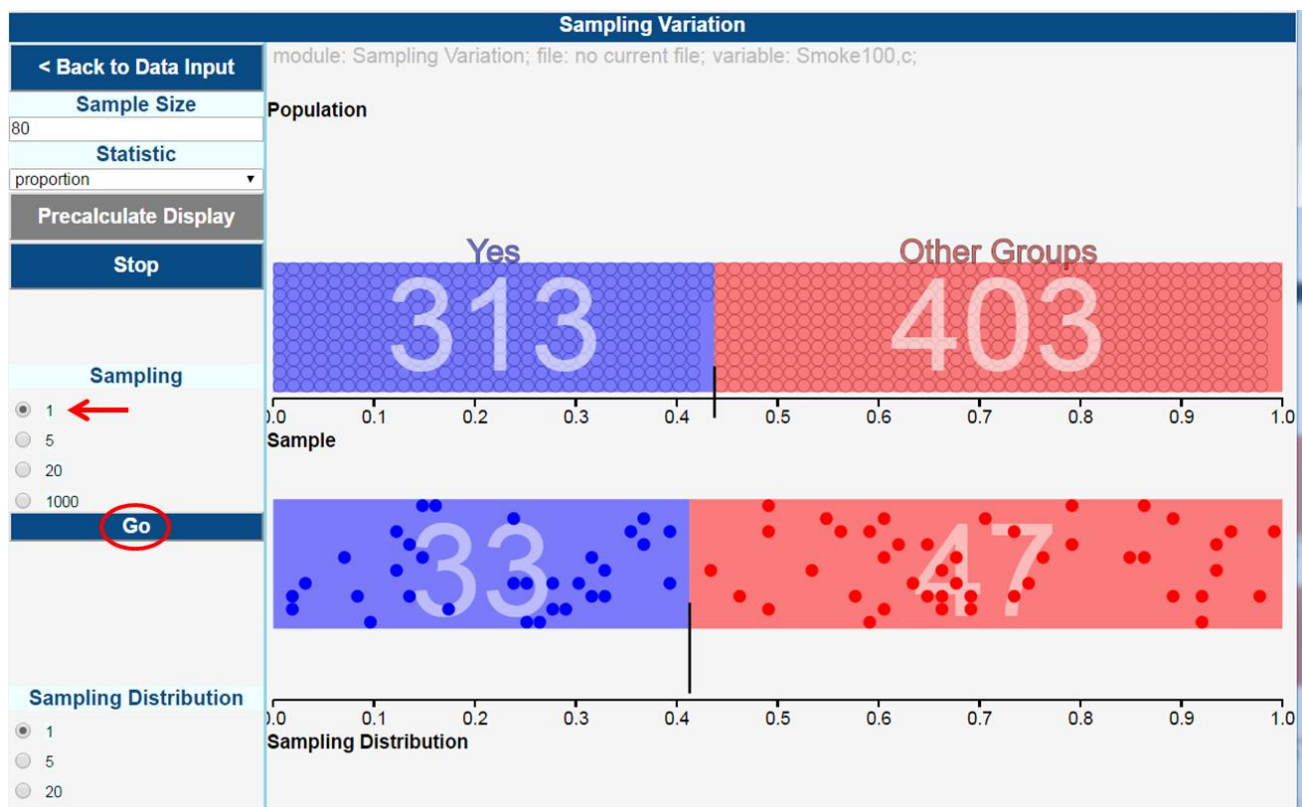
- For **Sample Size**, use **80**
- For **Statistic**, the only option is **Proportion**.
- Now click **Precalculate Display**.



Next, make your selections in the middle part of the window under **Sampling**.

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

A sample of 80 from the NHANES-1000 population will be selected and drop down onto the middle graph labelled **Sample**.



## What is happening in the graphics window?

VIT has taken a sample of size 80 from the population of 1000 and calculated the proportion of people in the sample who have smoked 100 cigarettes or more. In the sample above, the proportion of smokers is about 0.41 or 41% but yours will probably be different. The blue segment is the proportion of smokers in the sample. Remember that our proportion of 'smokers' in the population is in the top plot.

- Select 5 repetitions and observe 5 more samples of size 80.

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

- Under Number of repetitions **Select 1000** and click **Go**.



## What is happening in the graphics window?

You should see 1000 different samples of size 80 taken from the NHANES-1000 population. 1000 different sample proportions of 'smokers' are calculated from the 1000 different samples. Each sample proportion is shown as a blue line in the middle plot and circles in the **Sampling Distribution** plot.

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

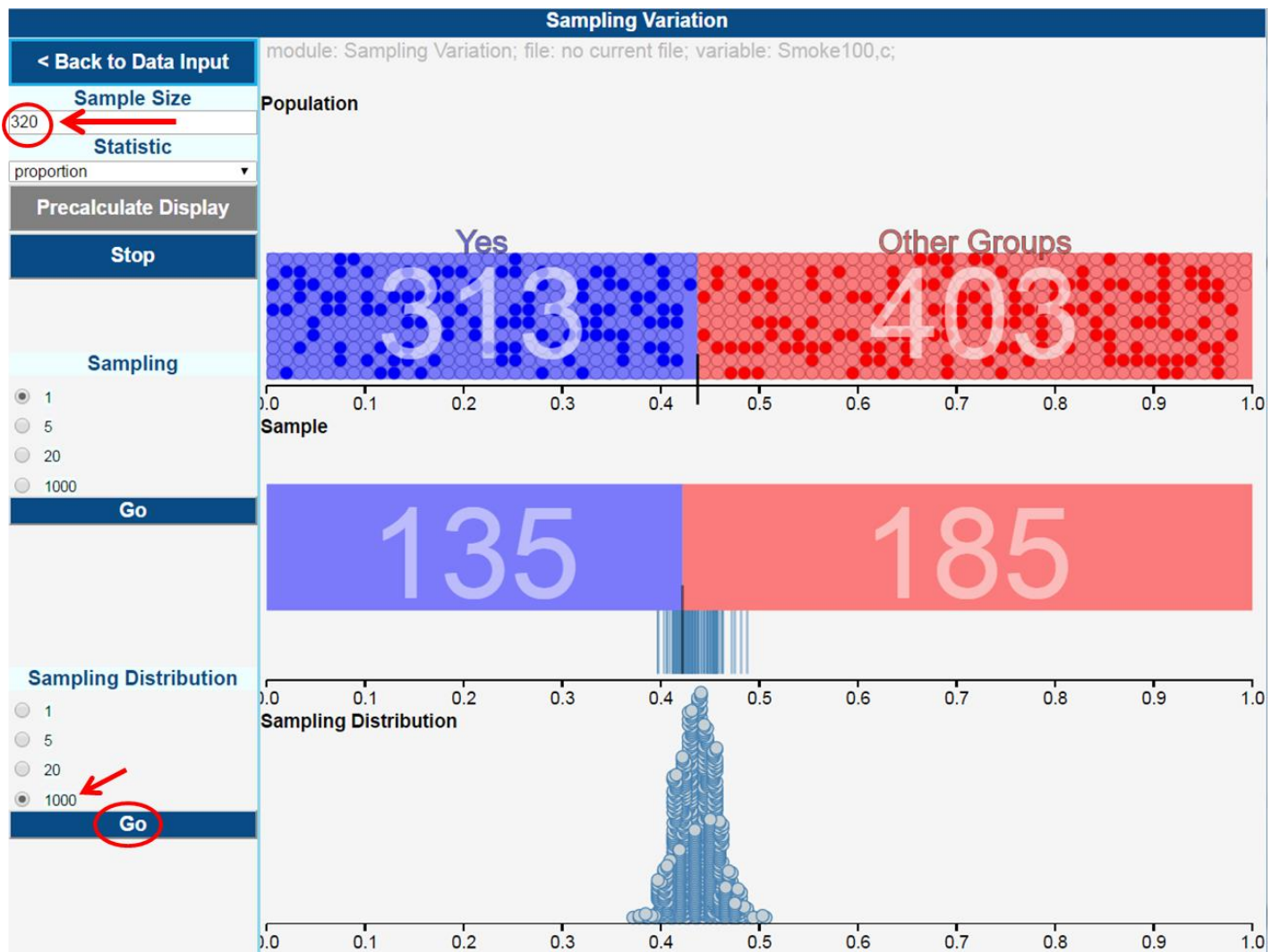
This exercise demonstrates that if we take a sample of 80 from a population of size 1000, our estimate of the proportion of people who have smoked more than 100 cigarettes is likely to be somewhere between about 0.27 and 0.6.

Your sampling distribution will not look exactly like this one but it should be 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 **320** (4 times the previous amount)
- Click **Precalculate Display**.
- Under the Command panel **Sampling distribution**, select **1000**
- Click **Go**.

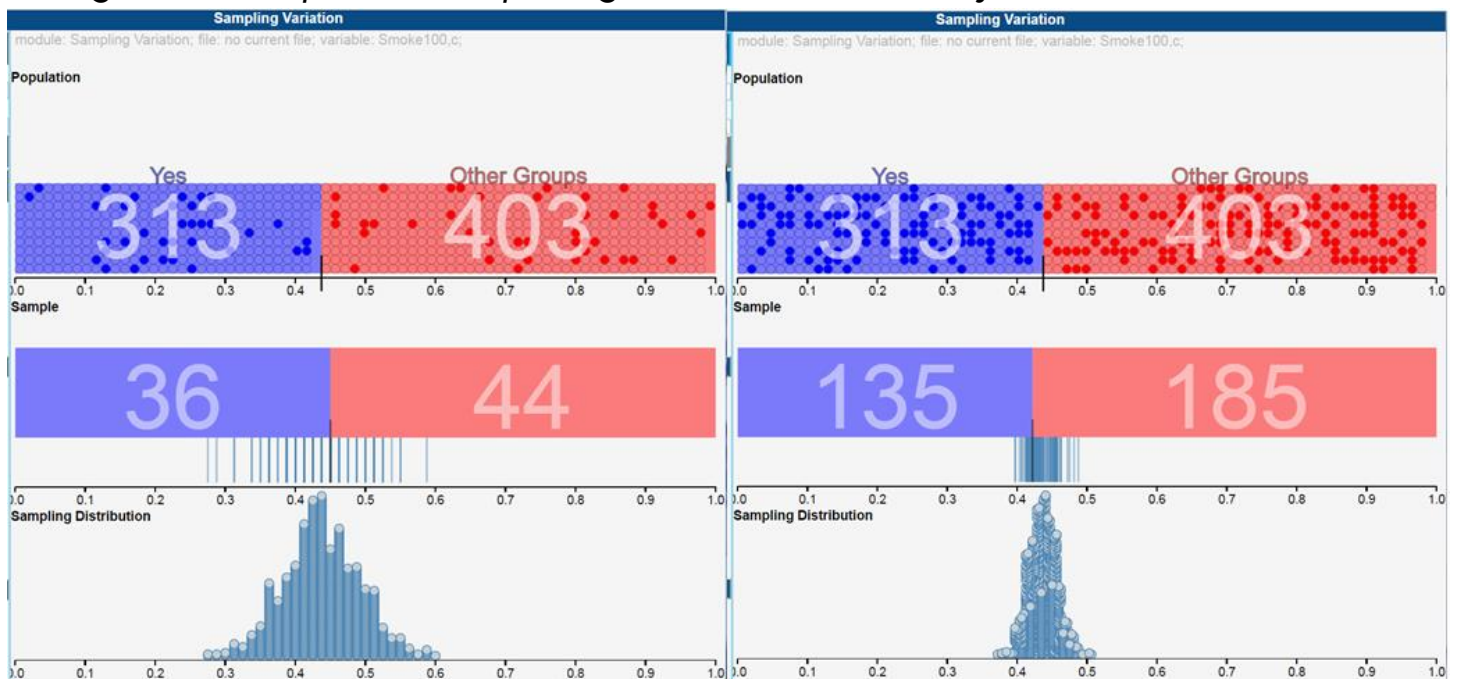


What is happening in the graphics window?

1000 different samples of size 320 are being taken from the NHANES-1000 population. The sample proportion of each of those samples is being recorded as a blue line on the **Sample plot** and a small circle in the **Sampling Distribution**.



(firing VITonline up twice and putting the 2 windows side by side ...)



### PRACTICE (~5 min)

Explore the effects of using different sample sizes and try out some other categorical variables of interest. If you notice or discover anything interesting, post a comment.

### 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.