



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 <u>linked here</u>]

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**.

限 The iNZight and Visual Inference Tool Systems					
The iNZight and VIT Systems					
iNZight (v2.3-900)					
Run iNZight					
Visual Inference Tools (v1.2.4-1) Visit Website					
Visual Inference Tools (VIT) contains programs for developing concepts					
O Randomisation variation					
C Randomisation tests					
C Sampling variation					
Bootstrap confidence interval construction					
O Confidence interval coverage					
Run selected VIT module					
THE UNIVERSITY OF AUCKLAND DEPARTMENT OF STATISTICS					

Download the <u>NHANES-1000-50.csv dataset here</u> or <u>NHANES-1000-50.txt</u> <u>dataset here</u> 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).

R	Visual Inference	Tools - Boot	tstra	p confidence	_		x	
Fi	ile							
	Bootstrap confidence intervals <u>Stop</u>							
	Data Analyse							
	Import Data	View Data	a Se	t View Varia	ables			
	VARIABLES	•	÷.	CONTINUED		•	*	
	SurveyYr			BMI				
	ID			BMCatUnder20)yrs			
	Gender			вмі_who			-	
	Age			Pulse			=	
	AgeDecade			PSysAve				
	AgeMonths			8PDiaAve				
	Race1			BPSys1				
	Race3			BPDia1				
	Education			BPSys2				
	MaritalStatus		1	BPDia2				
	HHIncome			BPSys3				
	HHIncomeMid			BPDia3				
	Poverty			Testosterone				
	HomeRooms			DirectChol				
	HomeOwn			TotChol				
	Work			UrineVol1				
	Weight			UrineFlow1				
	Length			UrineVol2				
	HeadCirc		-	UrineFlow2			-	
	112.15			D' L A				
	Variable 1 : B	MI		•		Clear		
	Variable 2 :	Select varia	ble r	name 💌		<u>C</u> lear		

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.

💦 Visual Inference T	ools - Bootstrap confidence.						
File	File						
Bootstrap confidence intervals Stop							
Data Analyse							
Quantity:	mean	•					
Level of interest:		T					
	Additional options						
	Record my choices						
	Re-sampling						
 ● 1 ● 5 ● 20 ● 1000 ■ Animate points and track sample Go 							
Include bootstrap distribution -Number of repetitions							
	Go						
Show CI Fade on/off Bootstrap distribution summaries							

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 Cl.



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 $\frac{1}{2}$ % 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.

Ŗ Visual Inference Tools - Bootstrap confidence 💶 💷 🗾							
File							
Bootstrap confidence intervals							
Data Analyse							
Quantity:	proportion 💌						
Level of interest:	30.0_plus						
Additional options							
	Record my choices						
	Re-sampling						
 1 5 20 1000 Animate points and track sample 							
	Go						
Include bootstrap distribution Number of repetitions 1 5 20 1000							
	Go						
Show C	TI Fade on/off						
Bootstrap distribution summaries							
,							



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 Cl.

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.

Why, when I repeat the 1,000 bootstrap distribution simulation, do I keep getting the same confidence-interval values?

With just pressing Go again it simply replays the same set of values. In VIT to get a fresh simulation set, press Record My Choices again.

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

See <u>iNZight FAQ entry</u>.