Exercise: Using VIT to run a randomisation test

This exercise will enable you to use Visual Inferential Tools (VIT) to run a randomisation test for an experiment.

VIT has a Randomisation test module that is used for testing how "significant" our difference is. We can use randomisation tests to see if our experimental results could be produced just by the "luck of the draw".

The skills addressed are:

- Use VIT to re-randomise data and create a re-randomisation distribution.
- Interpret the re-randomisation distribution and tail proportion.

This exercise cannot be done in VITonline yet because the randomisation tests capability has not been built yet. If you don't have VIT, reading over the exercise should still be useful, though, for some of the teaching points reinforced within it.

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 **Randomisation tests** option.

- Click **Run selected VIT module**.
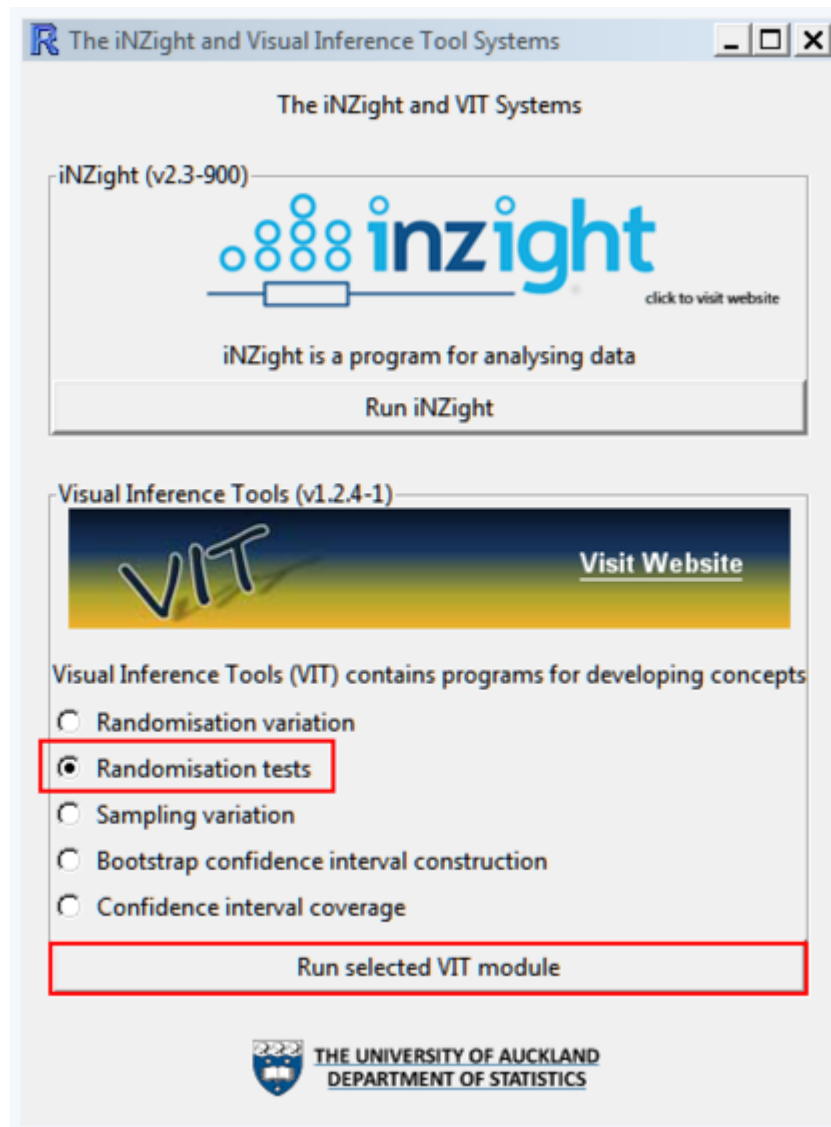
Download the [CanadaPop.csv dataset here](#) or [CanadaPop.txt dataset here](#) and put it on your Desktop.

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

This data is based on an experiment where the researcher asked 19 students to estimate (in millions) the population of Canada (see Utts, J. M. & Heckard, R. F. (2014). *Mind on Statistics*, (5th ed.). USA: Cengage Learning.) The researcher

was interested in whether the answers to a question may be influenced by information given with, or in, the question. This is called an "anchoring effect".

Students were randomly allocated to two groups and the question asked was phrased in two different ways.

Group A were asked: "*The population of the U.S. is about 270 million. To the nearest million, what do you think is the population of Canada?*"
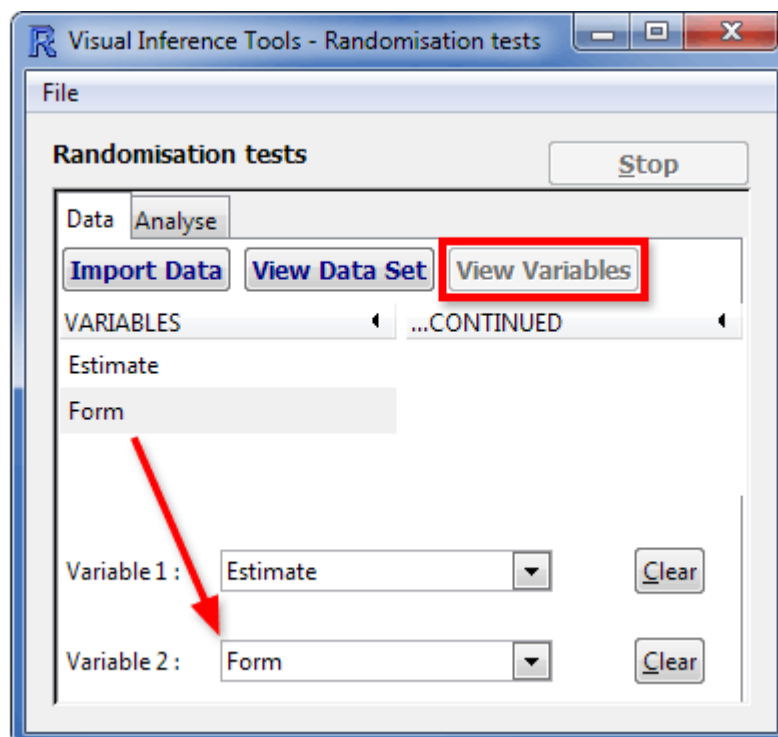
Group B were asked: "*The population of Australia is about 18 million. To the nearest million, what do you think is the population of Canada?*"

The variable **Estimate** contains their guesses/estimates of Canada's population size.

### Creating a re-randomisation distribution

Drag the variable **Estimate** into the **Variable1** slot.
Drag the variable **Form** into the **Variable2** slot.

The graphics window will show the estimates and form used on the left hand side and a dot plot of the estimates on the right.

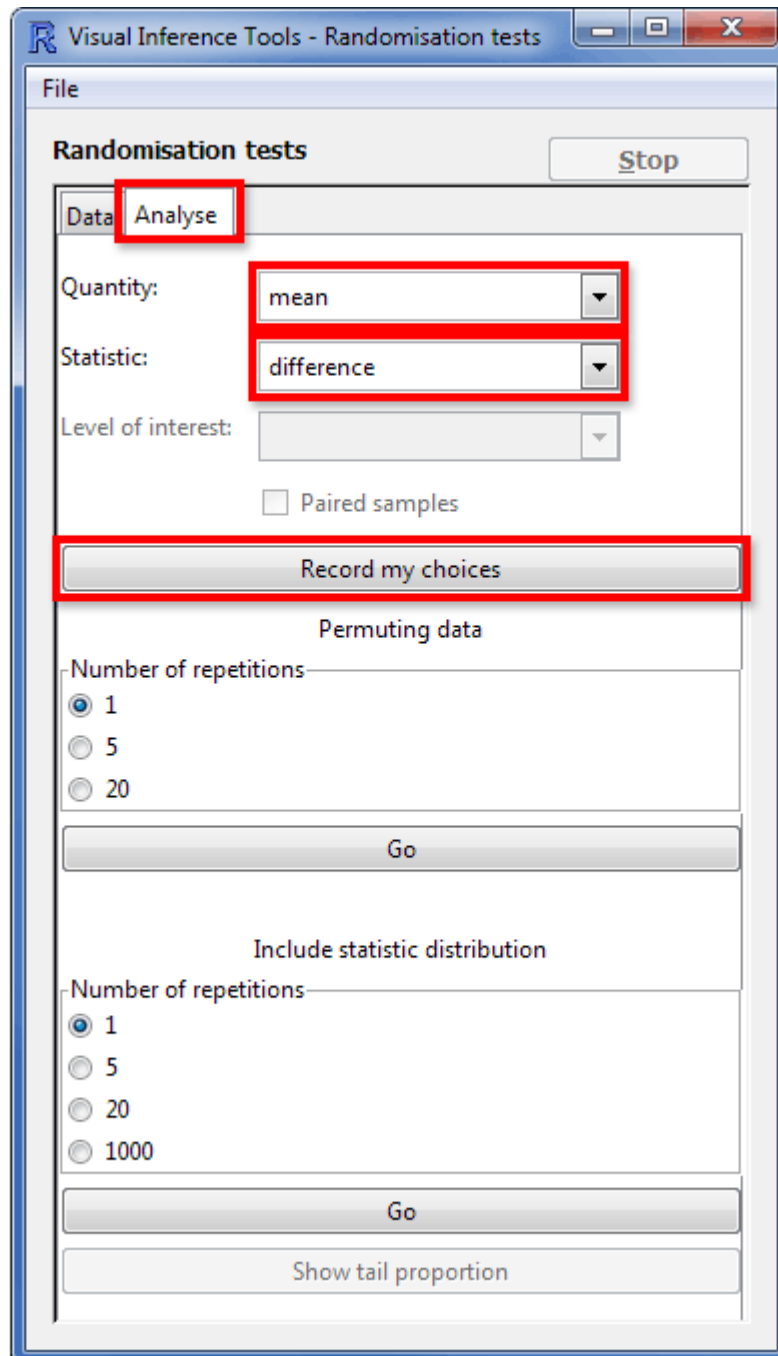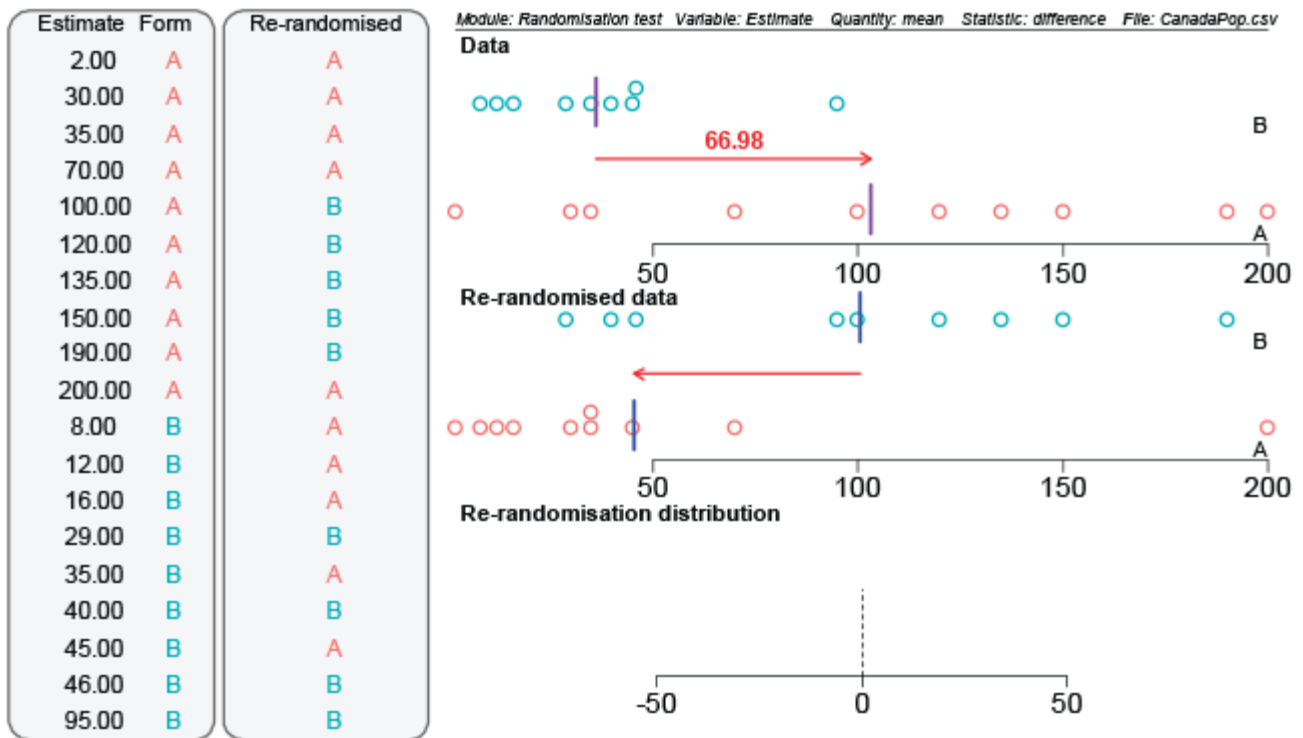| Estimate	Form
2.00	A
30.00	A
35.00	A
70.00	A
100.00	A
120.00	A
135.00	A
150.00	A
190.00	A
200.00	A
8.00	B
12.00	B
16.00	B
29.00	B
35.00	B
40.00	B
45.00	B
46.00	B
95.00	B
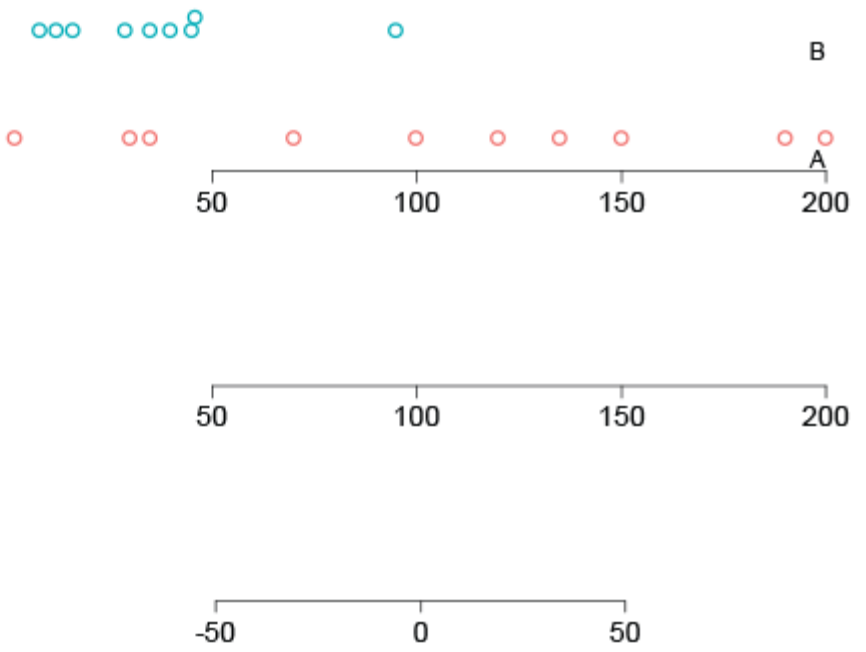
Enlarge image

Cont ...

Now click on the **Analyse** tab. Accept the default settings, i.e.

- Next to Quantity, select **mean**
- Beside Statistic, select **difference**
- Click **Record my choices**.

Next, make your selections in the middle part of the window under Permuting data.

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

Estimate	Form	Re-randomised
2.00	A	A
30.00	A	A
35.00	A	A
70.00	A	A
100.00	A	B
120.00	A	B
135.00	A	B
150.00	A	B
190.00	A	B
200.00	A	A
8.00	B	A
12.00	B	A
16.00	B	A
29.00	B	B
35.00	B	A
40.00	B	B
45.00	B	A
46.00	B	B
95.00	B	B

Module: Randomisation test   Variable: Estimate   Quantity: mean   Statistic: difference   File: CanadaPop.csv
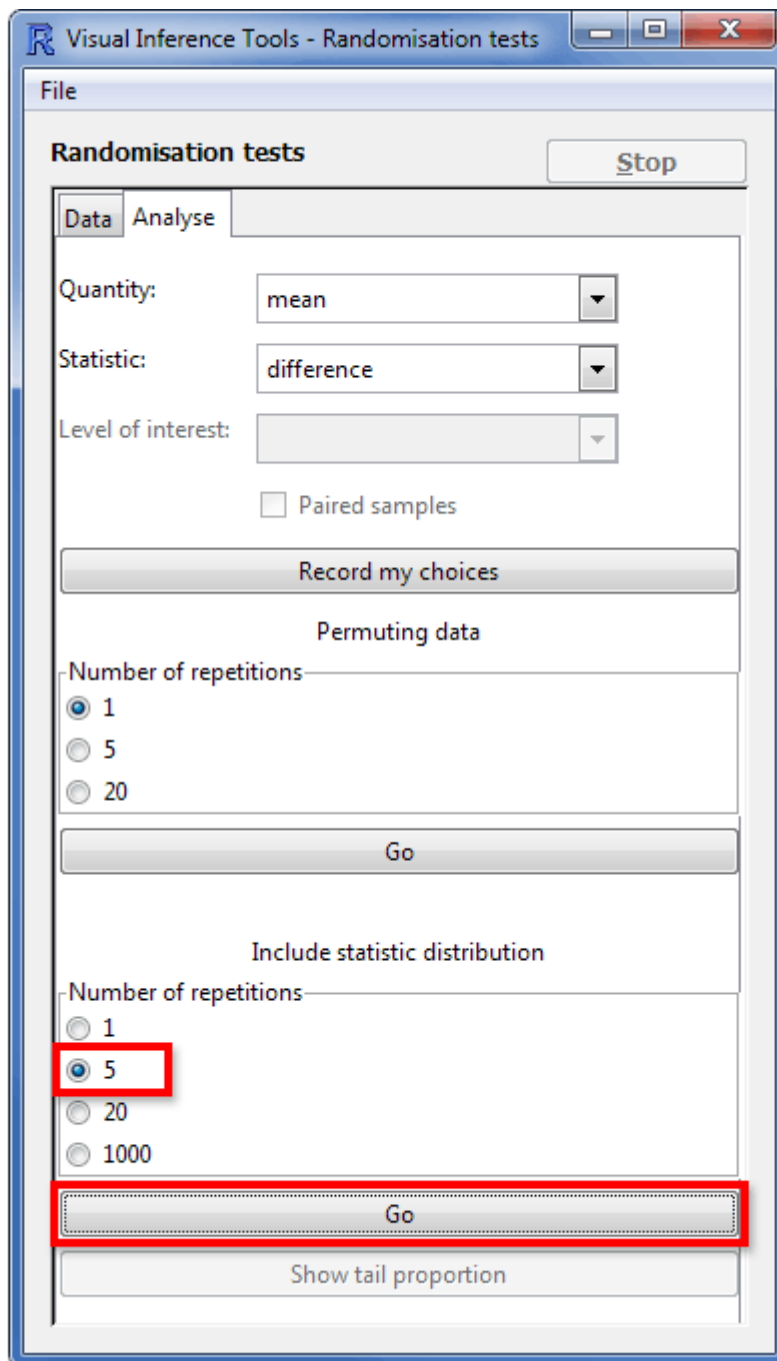
[Enlarge image](#)

What is happening in the graphics window?

VIT is randomly labelling values as from Group A or Group B regardless of which group the measurement initially came from. This is called re-randomised data. It then calculates the new group means for Groups A and B and calculates the difference between them.
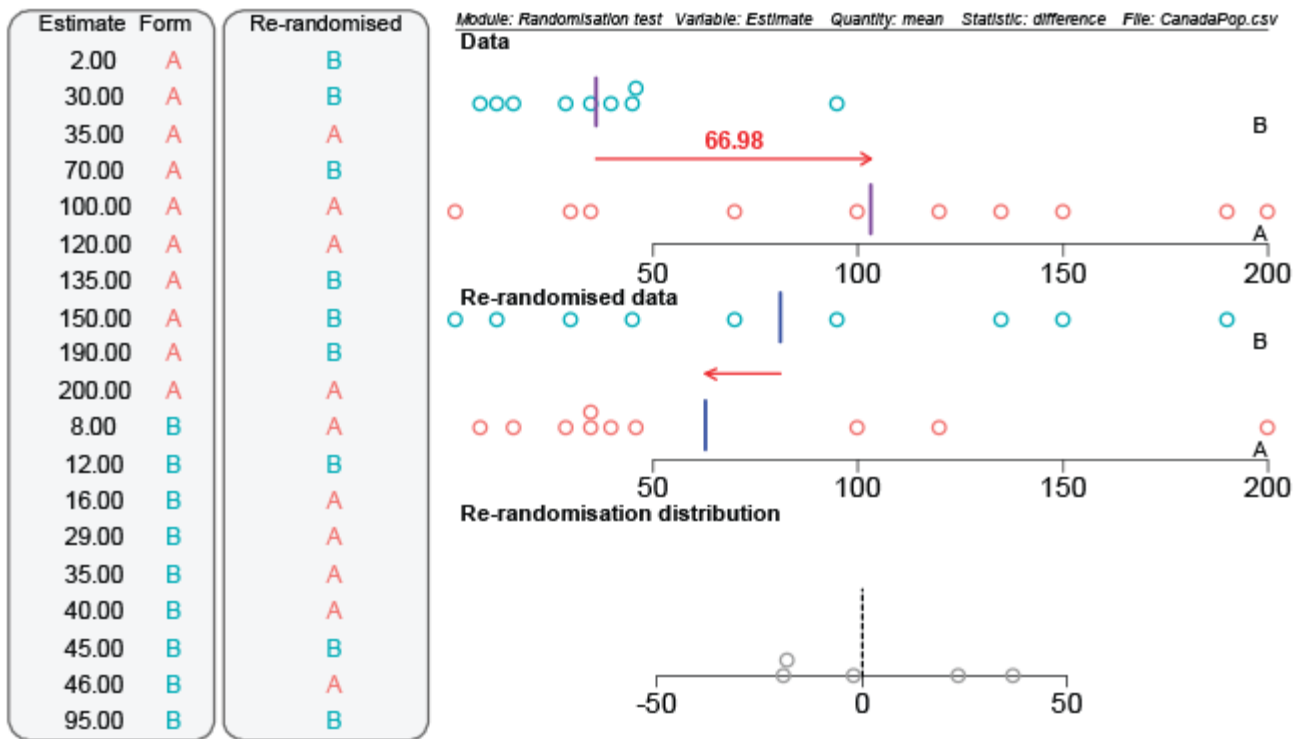
Now we'll use the bottom part of the Analyse window that says **Include statistic distribution**.

Under Number of repetitions:

- **Select 5** repetitions
- Click **Go**.

The above process of re-randomising will happen 5 times and each difference between the new randomly assigned groups A and B will now be recorded in the bottom plot called the **Re-randomisation distribution**.
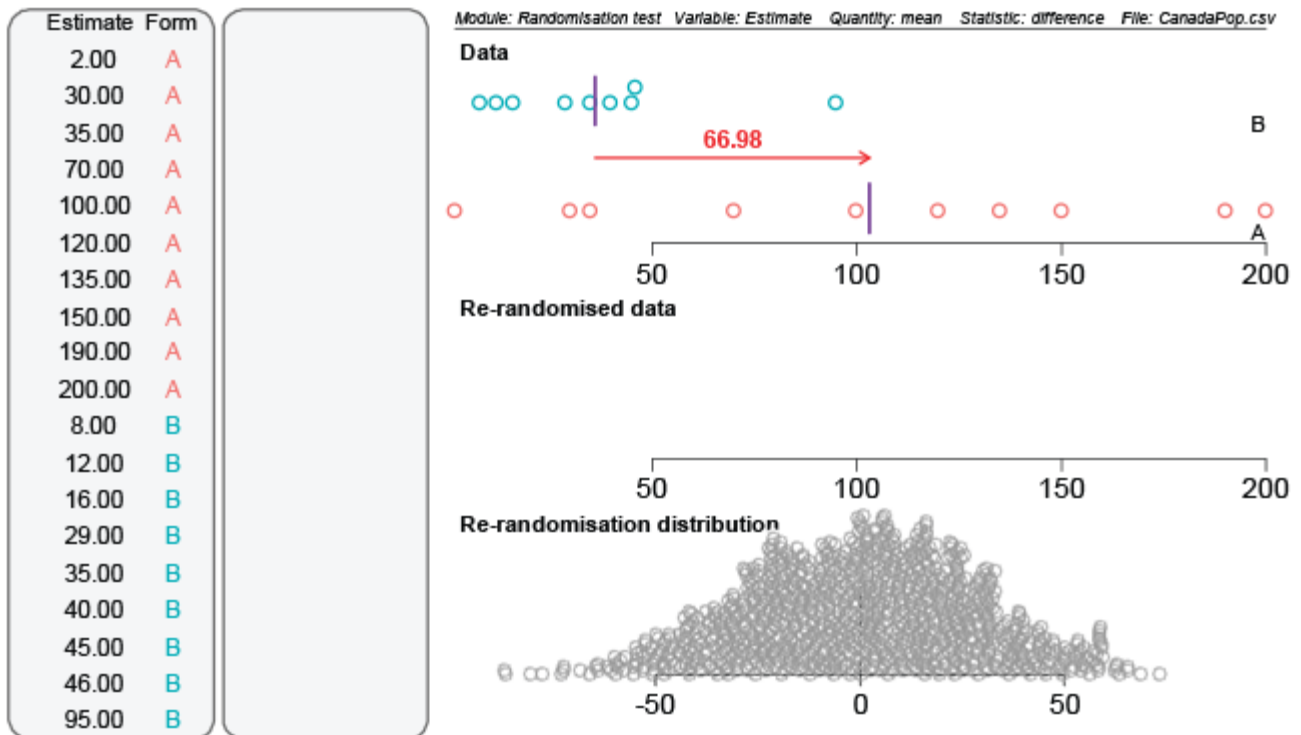

[Enlarge image](#)

In the bottom part of the analyse window (**Include statistic distribution**):

- **Select 1000** repetitions, click **Go**.
- When that is finished, click **Show tail proportion**.

You will see a **Re-randomisation distribution** of the difference in the Group means of 1000 re-randomisations of the estimates. Each dot in the re-randomisation distribution represents the difference between the group means from one re-randomisation.

Estimate	Form
2.00	A
30.00	A
35.00	A
70.00	A
100.00	A
120.00	A
135.00	A
150.00	A
190.00	A
200.00	A
8.00	B
12.00	B
16.00	B
29.00	B
35.00	B
40.00	B
45.00	B
46.00	B
95.00	B

Module: Randomisation test   Variable: Estimate   Quantity: mean   Statistic: difference   File: CanadaPop.csv

[Enlarge image](#)

What is happening in the graphics window?

The difference (of 66.98 million) between the means of Groups A and B is shown by the red arrow in the top plot. The difference is then superimposed over the re-randomisation distribution to see how many times out of 1000 re-randomisations we got a difference the same as (or larger than) our experimental difference. In the above example, the difference of 66.98 or larger happened 9 times out of 1000.

Note that your values will probably be slightly different.

EXPLORE (~5 min)

Run your randomisation test another few times and pause it so that you can understand what is happening at each part of the process. *It is important* that you always click **Record my choices** between runs. This is when all re-randomisation and the calculations are done. If you don't you'll just keep replaying the same set of results and always get the same answer.

Do you think there was an "anchoring" effect? Post a comment.

Common questions

Why do the data change their group allocation?

VIT is randomly (and artificially) relabelling the population estimates the people have made as belonging to Groups "A" and "B" to try to see how the experimental results compare to what random allocation acting alone could do. It strips the real group labels away from the estimate (all dots turn black) and randomly re-allocates group labels to the population estimates (dots re-coloured). Then VIT measures the new difference in means between the new artificially formed groups. This is one re-randomisation.

Why is the tail proportion of my plot different from the example?

The re-randomisation process is different every time VIT runs it because it randomly re-allocates the estimates to groups. You should get a tail proportion that is fairly similar. If we ran it a much larger number of times (say 100,000) we would always get the same answer. (The only reason we don't do 100,000 is because plotting would be too slow.)

What does a small tail proportion mean?

The smaller the tail proportion, the less likely it is that the "luck of the draw" would produce a difference this big. A small tail proportion tells us that our experimental results are out in the tails of the randomisation distribution, so that random reallocation almost always gives differences that are smaller than what we got in our experiment.

What does a larger tail proportion mean?

It tells us that random re-allocation often produces a difference larger than we got in our experiment. Our experimental results therefore give no evidence a true treatment difference exists.