CSCU9YM Practical Three
Experimenting with ABMs

This sheet contains one checkpoint.

In the last practical you saw how to explore an ABM by using sliders to control the values of parameters and exporting data to a file for later analysis. Manipulating the parameters of a model directly via the NetLogo interface is convenient for performing small-scale investigations of the behaviour of a model, and for testing it. However, this approach becomes tedious and impractical if we want to carry out large-scale studies involving a range of values for a parameter (or set of parameters).

For example, we might want to systematically investigate how changing *q* causes the final width of the butterfly corridors to change. We *could* do this by manually setting *q* to values such as 0, 0.1, 0.2, ... up to 1, running the model for each of these values, and then collating and graphing the results. But this would be very time-consuming. Even worse, because the butterfly model has *stochastic* behaviour, it is not enough to run the model just once for each value of *q*. This is because different model runs may end up with different corridor widths because of random variations in the way the butterflies move. To allow for this variation, we need to run the model several times, say 20 times, for each value of *q*, and then look at the average corridor width over all 20 runs. This is far too time-consuming to do by hand.

BehaviorSpace to the rescue! BehaviorSpace is a tool built into NetLogo which allows us to design and run experiments on models. BehaviorSpace runs the model many times, systematically varying the parameters and recording the results of each model run. You can control how the parameters are varied (this is called “parameter sweeping”) and how the results are recorded. In this practical we will see how to use BehaviorSpace to explore how changing q causes the final width of the butterfly corridors to change.

You can read more about BehaviorSpace in the NetLogo User Manual (under “Features”) and in Chapter 8 of Railsback and Grimm.

Download ButterflyCorridorsExperimentStarter.nlogo and open it in NetLogo. This file contains a slightly modified version of the butterfly corridor model you created in the last practical. [The differences are that in this version, all the butterflies start on the same, specified patch, and there is no writing of output data.] Read through the code and check that you understand it. Run the model in the usual way and check that it works.

We are now going to modify this model so that it can be used with BehaviorSpace. First we will split the `setup` procedure into two separate procedures. The first procedure, which we will call `setup`, creates the landscape, and will be run only once (because it is inefficient to re-create the landscape each time the model is run.) The second procedure, which we will call `setup-run`, will create the butterflies and perform all the actions needed at the start of each model run. The code for these procedures is shown below:

```
to setup    ;; called once, to create the landscape
  clear-all
  make-hills
end

to setup-run   ;; called before each model run
  ask patches [set  visited? false]    ;; reset patch variables
  clear-ticks       ;; clear the tick counter
  clear-turtles     ;; remove all turtles
  clear-drawing     ;; removes lines drawn by turtles
  clear-all-plots   ;; clears all plots
  clear-output      ;; clears any output text (there is none)
  make-butterflies  ;; create and initialise the butterflies
  reset-ticks       ;; restart tick counter at 0
end
```

Run `setup` to set up the landscape. Now open BehaviorSpace (from the Tools menu). Create a new experiment. Set it up as follows:

The variables are varied so that q sweeps through values from 0 to 1 in increments of 0.1, and num-butterflies is fixed as 50, as below:

```
["q" [0 0.1 1]]
["num-butterflies" 50]
```

Leave the number of repetitions as 1.

Measure runs using the reporter `corridor-width`

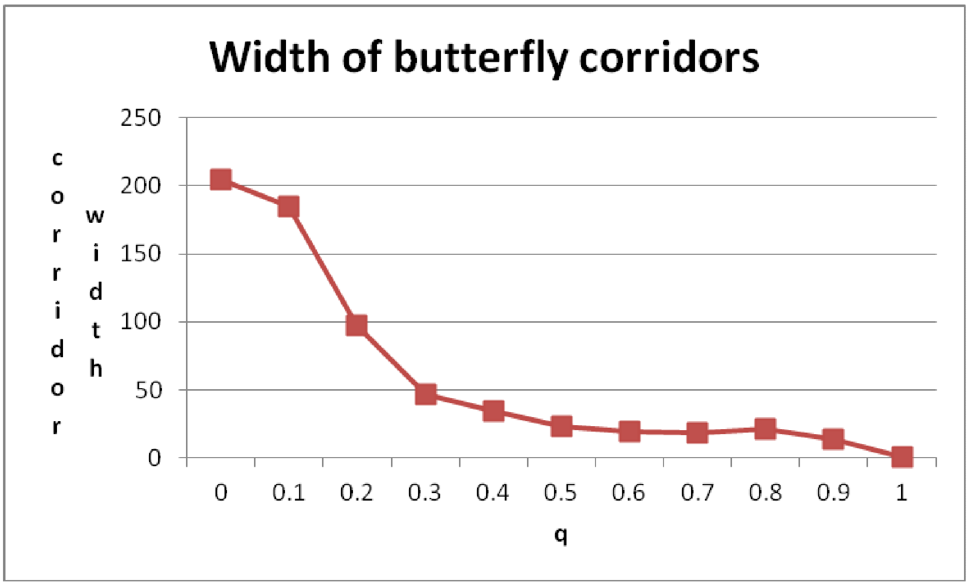
Untick the “Measure runs at every step” tick-box.

Make the setup command `setup-run` and the go command `go`.

When you have finished setting up the experiment, click Run to run it. Choose table output, and do 1 run in parallel. (Doing more than 1 run in parallel is very useful when doing large numbers of runs, but makes the result file slightly harder to manage.). Save the output file as a `.csv` file with a name of your choice.

The experiment should complete within a few minutes. Open the output file in Excel and look at the results. Use your Excel skills to make a chart containing a line graph of corridor-width on the y-axis versus q on the x-axis. The graph should look something like the one below.

(If you need a refresher in drawing graphs with Excel, have a look at the tutorials on lynda.com You can access these for free via the Student Portal.)



You may notice that your graph is not very smooth, and it probably has a slightly different shape from mine. The butterfly corridor model has random (stochastic) behaviour, and this means that model runs are not deterministic: different model runs will give different results. To get a better idea of the behaviour of the model we need to run it multiple times and graph the average results.

Set up a second BehaviorSpace experiment that is identical to your first one but with 20 repetitions. Run the experiment, storing the data in a table as before. This may take some time to complete. When it is finished, open the result file in Excel. You will see that for each of the 11 values of q, there are 20 rows, containing the results of the replicated model runs with that value of q.

Using Excel Pivot Tables for data analysis

To easily calculate the mean and standard deviation of the corridor width over the 20 runs for each value of q, you are going to use a pivot table. This is a very useful feature of Excel that you may not have used before. Follow these steps:

- Select (by clicking and dragging) the cells containing the data from your experiments, including the row of column headers. In my spreadsheet this is the data range A7:E227.
- Go to the Insert menu and choose PivotTable. In the dialog box that appears, click on OK. A new worksheet appears, in which you can build a Pivot Table report. You should see that the column headers in your data are listed on the right within a "Pivot Table Field List". There are four areas below this list, labelled Report Filter, Column Labels, Row Labels, and Values.
- Drag the field q from the field list into the Row Labels area.
- Now drag the field corridor-width from the field list into the Values area.
- The pivot table report now shows each value of q next to the sum of the corridor-widths for the runs with that q. However, we are interested in the average corridor-width, rather than the sum. To change the sum to the average, click on the field Sum of corridor-width where it appears in the Values area, and choose Value Field Settings from the menu which

appears. In the dialog box which appears, go to the Summarize by tab and select Average. Your report should changes to show the average corridor-width for each q, rather than the sum.

- To get the standard deviations, again drag corridor-width from the field list into the Values area. Follow the instructions in the previous step, but this time, summarize the corridor-widths by standard deviation (stdev). Your report should now look something like this:

	Values	
Row Labels	**Average of corridor-width**	**StdDev of corridor-width**
0	219.0041041	13.35411745
0.1	153.2984301	13.19062157
0.2	91.84398871	7.585011408
0.3	53.9996412	3.716349615
0.4	37.28838341	3.045428185
0.5	26.37215181	2.787998163
0.6	20.72060616	2.573556619
0.7	19.75457426	1.71583359
0.8	19.91762954	2.233418231
0.9	14.26136637	2.075873525
1	1.028991511	0
Grand Total	**59.77180611**	**65.83387753**

- Now let's graph this summarized data. Drag your mouse to select all the data cells in the pivot table report. This is all the data in the white cells above. Then Insert a Line chart with markers (just as you did before).

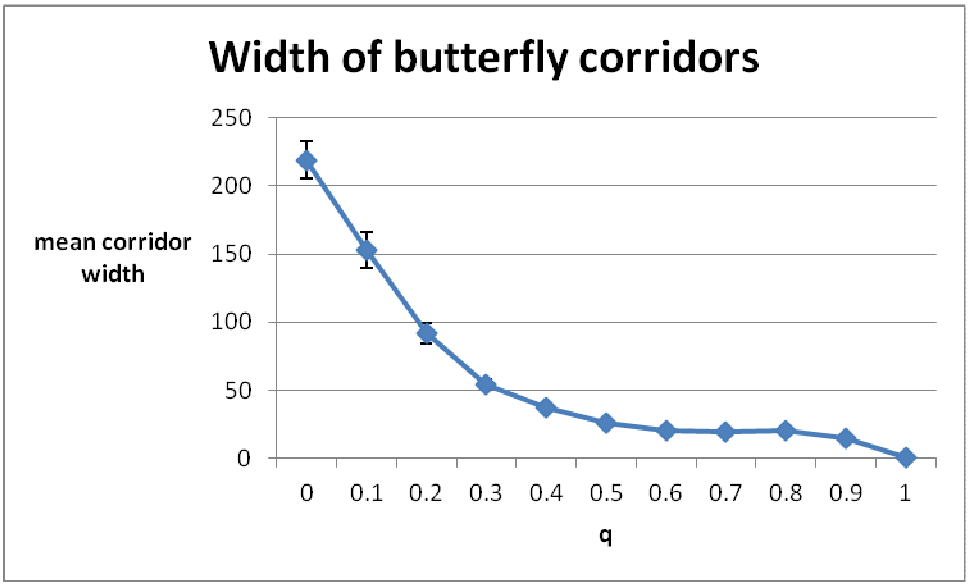
- This time Excel has correctly worked out that the data in the leftmost column should be used for the x-axis labels. However, the chart shows two series: the average corridor width (what we want) but also the standard deviation (which we don't want). Unfortunately, there seems to be no way to remove a data series from a chart based on a pivot table. However, you can click on the offending series and hide it by formatting its "Shape fill" as "No fill" and its "Shape Outline" as "No outline". In effect, the series remains on the chart but is invisible.

- Finally, follow the steps below to add error bars to your graph, showing plus and minus one standard deviation. These error bars will indicate the amount of variability in your data.

- Click within your chart to select the corridor-width series. In "PivotChart tools", select the "Layout" tab. Click on "Error Bars" and then choose "More Error Bars Options" at the bottom. A dialog box appears, called "Format Error Bars". We want a vertical error bar that goes both above and below the data point to show plus and minus one standard deviation. So choose the direction of the error bars to be "both". Make the "End Style" to be "Cap". However, do NOT choose "Standard deviation" under the "Error amount" heading, as we want to use our own standard deviation data, and not the default values provided by Excel. Instead, click on "custom", then click the "Specify Value" button. This brings up a dialog box that lets you choose the data to be used for the error bars.

- Click on the box underneath "Positive Error Value". Then select the cell range that contains your standard deviation values, and press enter. Next, click on the box underneath "Negative Error Value". Again, select the cell range containing your standard deviation data (the same cell range you selected before). You need to use the same data both times because you want the errors bars to extend both up and down by one standard deviation.

- Add a chart title and axis titles. You should end up with a graph that looks similar to the one below:



CHECKPOINT [EXPERIMENT]

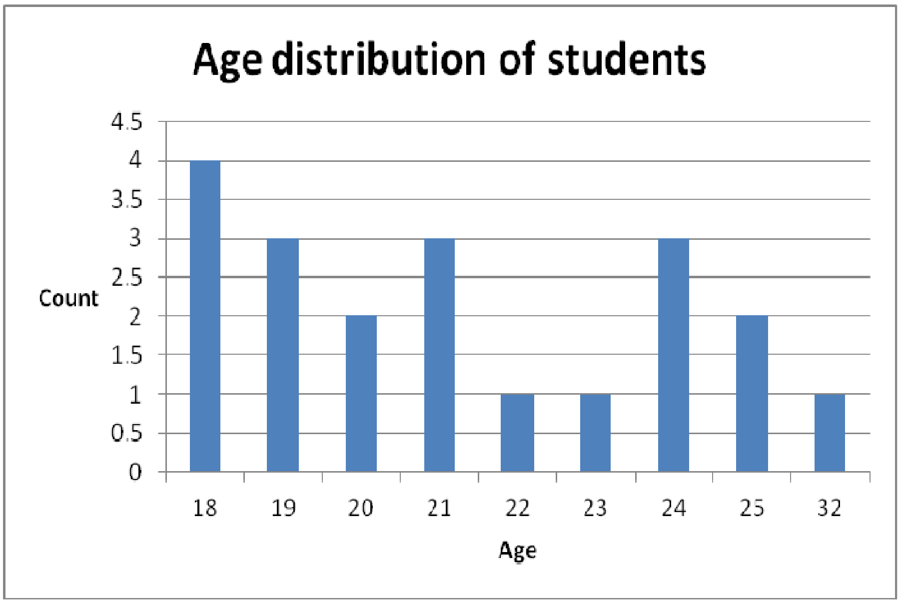
Show the demonstrator your data and graph, and discuss what you can conclude from the experiment.

If you want to learn more about Pivot Tables in Excel, I recommend this YouTube video: http://youtu.be/h_Mr4sb2BIs (Excel 2010 Statistics 06: Data Analysis: Sorting and Pivot Tables).

Creating a histogram with a pivot table

A histogram is a useful way of seeing the "shape" of the distribution of a set of data. Suppose that you have a table of data of students' ages (below left). You can use a pivot table in Excel to produce a histogram of this data (below right). Here is how.

Age
18
18
21
19
21
23
18
25
24
24
24
25
21
19
20
20
18
32
22
19



In Excel, select the data range and then choose Insert > PivotTable to create a pivot table in a new worksheet. The Pivot table field list should contain the single field age (the column header in your data). Drag this field from the field list into the Row label area. Then, drag this field *again* from the field list but this time put it into the Values area. Change the Value Field Setting from Sum to Count. Your PivotTable report now shows each value of age next to a count of the number of students with that age.

Click within the Pivot table report. From the Options Tab in the menu bar, click on PivotChart. Select a clustered column chart. Et voila, your histogram appears! You can then format the chart with titles and axis labels in the usual way.

Be aware that you will not be able to change the data used to produce the chart, as this is tied to the Pivot table report. If you change either the chart or the report the other will change automatically. If you do not want a pivot chart to show some data that is present in the report, you can hide it by formatting it so that it is invisible.

For instructions on how to create a histogram with "bins" (e.g., age ranges such as 18-25, > 25 etc) see http://datapigtechnologies.com/blog/index.php/creating-a-frequency-distribution-with-a-pivot-table/.