

Individualised Statistics Coursework Using Spreadsheets Random Data Generator - Demonstration

by Neville Hunt, Coventry University



Introduction

ISCUS (Individualised Statistics Coursework Using Spreadsheets) is a tool that enables you to allocate a unique subset of a large data set to each of your students, typically as part of an assignment. The data subset is determined by a five-digit PIN number, e.g. the last five digits of the student's University ID number.

The motivation is twofold:

- to deter plagiarism (primary motivation),
- to make the assignment more interesting and challenging (secondary motivation).

Different levels of sophistication are possible in setting data-based assignments:

Level	Data allocation	Potential for plagiarism
0	Students all given identical data	Nothing to stop two students submitting identical assignments
1	Students each given a random sample of fixed size n from the same variables	Two students could use the same spreadsheet to analyse their data
2	Students each given a random sample of random size n from the same variables	One student need only make minor adjustments to another student's spreadsheet, but can probably still copy their report with only minor editing
3	Students each given a random sample of random size n from the same variables from a randomly selected stratum (e.g. each student gets data from a different country)	One student need only make minor adjustments to another student's spreadsheet, can copy their introduction and methodology, but conclusions will probably be very different
4	Students each given a random sample of random size n from a fixed number k of randomly selected variables (unless the dataset is huge, <u>some</u> students will still receive the same variables)	One student need only make minor adjustments to another student's spreadsheet, can use the same methodology, but introduction and conclusions are likely to be very different
5	Students given an entirely different dataset	Students can only share methodology, but one student can still ask another to do the whole assignment for them

ISCUS is not needed at Level 0, and cannot be used at Level 5. Its facilities are designed with levels 1-4 in mind.

The problem for the teacher is that as the level of sophistication increases so does the time taken to mark the assignments. At Level 0 the teacher can use a single answer sheet to check all assignments. At Level 5 the time taken to check and mark each assignment would be prohibitive.

ISCUS provides two facilities to assist the checking and marking of students' work:

- selected key outputs for each student's dataset – e.g. counts, means, correlations,
- a printed analysis of each student's dataset, based on a pre-prepared template.

Getting started

To operate ISCUS requires the following:

1. Access to the spreadsheet *RDG.xls*.
2. Access to a source dataset, either stored in an accessible Excel spreadsheet, or copied to the clipboard from some other package (e.g. Minitab) but in a format that can be pasted directly into an Excel spreadsheet.
3. Either a decision to allow any (positive) five digit PIN to be used, or access to a spreadsheet containing PINs (first column) and associated student names (second column).

Warning: Enabling Macros

It is possible that some users will be unable to run the data generator owing to their macro security setting in Excel being set to **High**. To change the security settings:

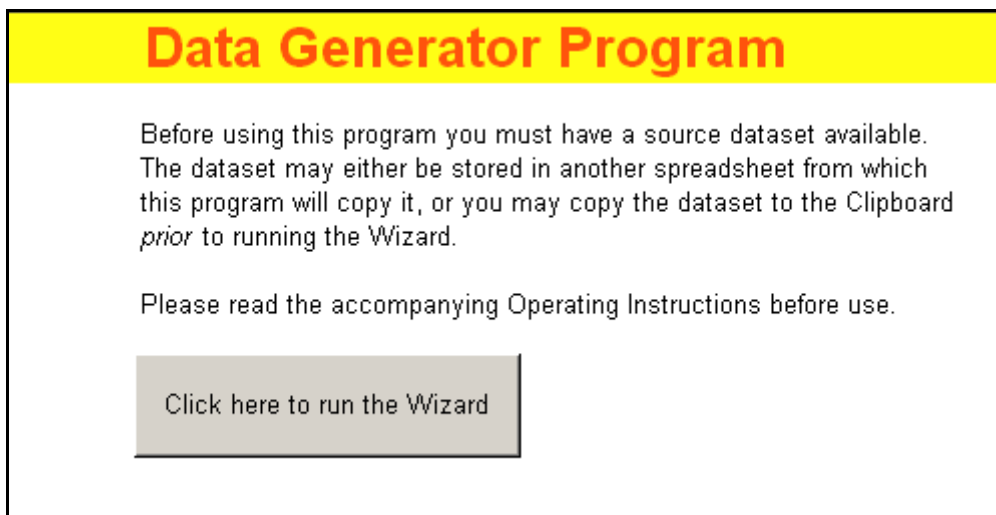
- Open Excel and from the main menu select **Tools > Macro > Security**.
- Select the **Medium** security setting, then **OK**.

This will allow users to choose whether or not to enable macros.

Example 1 Student data only

It is assumed that you have a dataset stored in a file called *MyData.xls*.

- Start Excel and select **File > Open** from the main menu. In the dialog box that appears choose to view **Files of type:** Excel (*.xls) and in the **Look in** box navigate to the folder containing *RDG.xls*. Click on this file and then **Open**.
- Select **Enable Macros** otherwise the Wizard will not work.



- Click on the grey button to run the seven-step Wizard.

Step 1

- Click on **Data stored in spreadsheet** and then on the **Open file** button.
- Navigate to the appropriate folder and open your data file *MyData.xls*.
- Click inside the box headed **Select the sheet (and range) containing the data**, then use your mouse to highlight the data range to be used as your source data set – we suggest **A1:H51** would be adequate for this exercise.
- Click on **Next**.
- A message will confirm the size of the dataset.

Step 1 of 7 - Source Data

Data can either be retrieved from the clipboard or from a file not currently open

Data copied to clipboard

Data stored in spreadsheet

Select the sheet (and range) containing the data

Sheet1!\$A\$1:\$H\$51

Step 2

- Accept the default setting of **Allow any 5 digit PIN**.
- Click on **Next**.

Step 2 of 7 - Student PINs

Allow any 5 digit PIN

Check PIN against list

List of PINs/Names stored in file:

Select the sheet (and range) containing the PINs/Names

Step 3

- Click on the **Same sample size for each student** option and enter **30**.
- Click on **Next**.

Step 3 of 7 - Sample Size

Same sample size for each student

Random sample size for each student

Between

And

Step 4

- Accept the default setting of **all columns**.
- Click on **Next**.

Step 4 of 7 - Choice of Columns

Student data is to be drawn from:

- all columns
- a random selection of columns
- a selection of columns as structured below

Structure>> Selection>>

Cancel Back **Next**

Step 5

- Accept the default setting of **None** under **Stratify using levels in column**.
- Click on **Next**.

Step 5 of 7 - Choice of Rows

Stratify using levels in column None

- Sample from all levels
- Sample from randomly chosen levels equally or proportionately
- Use weights

Level	Weight
<input type="text"/>	<input type="text"/>

Cancel Back **Next**

Step 6

- Accept the default setting of **Student data only**.
- Click on **Next**.

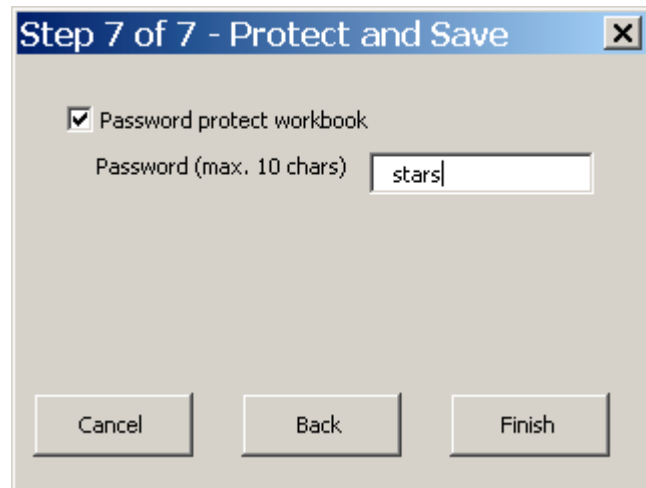
Step 6 of 7 - Mode of Operation

- Student data only
- Student data and questions in file:
 - Open file
 - Hide data
- Key answers for teacher
 - Options>>
- Teacher answers template in file:
 -
 - Print for all PINs

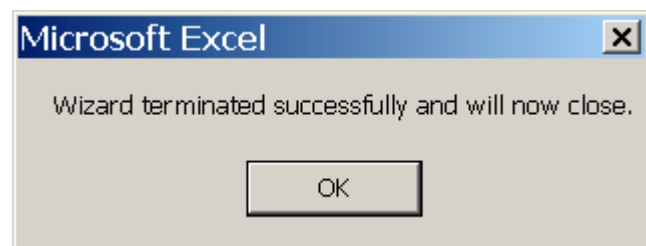
Cancel Back **Next**

Step 7

- Click in the checkbox to enable password protection of the student data generator.
- Enter the password **stars**.
- Click on **Finish**.
- You will be presented with the usual Excel **Save As** dialog. Choose to save the file under the name *STARS1.xls* in some appropriate folder.



- After a short delay the Wizard terminates.
- Click on **OK** to close the program.



What the program has done is to create a new program called *STARS1.xls* which is used by the students to generate their personalised datasets.

To try it out:

- Open the file *STARS1.xls*.
- Click on **Enable Macros**, otherwise nothing will work.
- Type any five digit PIN number into the blue box and press the Enter key.
- Click on the **Create Data** button.

Data Generator Program

1. Please type your five digit PIN number into the blue box and press Return.
 PIN (e.g. 05678): 12345
2. Now click on the **Create Data** button below.

Create Data

PS. This may take a minute or two

- After a short pause, a new spreadsheet of data with a default name (e.g. **Book1**) should appear and the STARS1 spreadsheet should have disappeared.

- Run *STARS1.xls* again with a different PIN and another new spreadsheet (e.g. **Book2**) will be created. Compare **Book1** and **Book2** to satisfy yourself that a different sample has been created.
- Run *STARS1.xls* again with your first choice of PIN to check that you get the same data as you did the first time you used that PIN.
- Open *STARS1.xls* again and check that the workbook is protected. You should not be able to unhide the hidden sheets using **Format > Sheet > Unhide**.
- If, for whatever reason, you needed to access any of the hidden sheets then (without entering a PIN) you could use **Tools > Protection > Unprotect Workbook** and enter the password **stars** to unlock the workbook. Using **Format > Sheet > Unhide** you should now be able to access the hidden sheets Teacher, Settings and Pins. In this example the Pins sheet is blank, but had you provided a list of PINs then you might have wanted to edit the list at some later date without having to re-create the data generator program from scratch.
- The data generator can be re-protected using **Tools > Protection > Protect Workbook**. The general advice is always to protect the data generator, although it is not clear what advantage a student might gain from accessing the hidden worksheets.

Example 2 Student data and questions

In this example, each student will be allocated two randomly selected columns from the source dataset in order to carry out a regression analysis. To further personalise the task, the teacher has supplied a randomised question stored in a spreadsheet file.

Preliminaries

- Start Excel and open the file *Questions.xls* to inspect it. You will see that the workbook contains two sheets, one to receive the data from ISCUS and the other containing a randomised question that contains data-dependent elements.
- Select the **Data** sheet. This shows a typical student's data, with a column of x-values in A, y-values in B and summary information in columns D and E, including the student's PIN in cell E1.
- Select the **Questions** sheet and click on cell G2. In the formula bar you will see that the cell formula is =**Data!B1**. This cell is picking up the name of the y-variable in the student's dataset.
- Click now on cell C5 and examine its formula. This cell is specifying the value of the student's x-variable at which a prediction is to be made. It is taking the student's average x-value, multiplying it by a factor between 0 and 2, then taking the integer part. The factor is determined by the number formed by the last two digits of the student's PIN, divided by 50. The thinking behind this is that some students' prediction will involve interpolation and others' extrapolation.
- Close *Questions.xls* without saving changes.

Wizard

- Start Excel and select **File > Open** from the main menu. In the dialog box that appears choose to view **Files of type: Excel (*.xls)** and in the **Look in** box navigate to the folder containing *RDG.xls*. Click on this file and then **Open**.
- Select **Enable Macros** otherwise the Wizard will not work.
- Click on the grey button to run the seven-step Wizard.

Step 1

- Click on **Data stored in spreadsheet** and then on the **Open file** button.
- Navigate to the appropriate folder and open your data file *MyData.xls*.
- Click inside the box headed **Select the range containing the data**, then use your mouse to highlight the data range to be used as your source data set – we suggest **A1:H51** would be adequate for this exercise.
- Click on **Next**.

Step 2

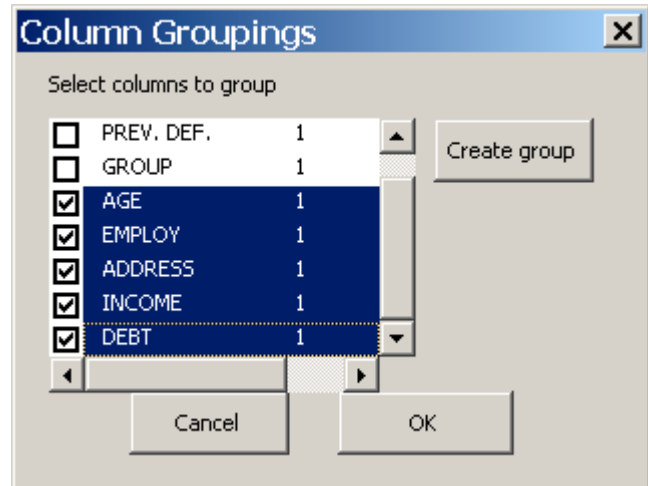
- Accept the default setting of **Allow any 5 digit PIN**.
- Click on **Next**.

Step 3

- Click on the **Random sample size for each student** option and enter Between **30** And **50**.
- Click on **Next**.

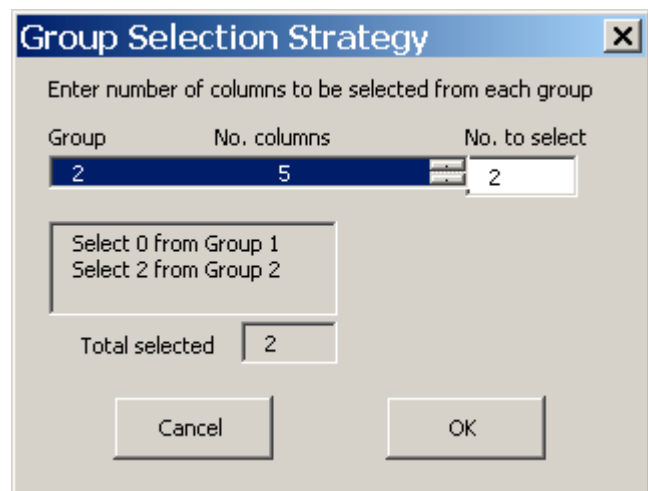
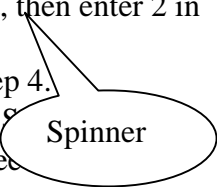
Step 4

- Click on a selection of columns as structured below, then click on the **Structure>>** button.
- Click in the checkboxes to select the continuous variables AGE through to DEBT.
- Click on **Create group**, to classify these variables as Group 2, then **OK**.
- Now click on the **Selection>>** button.



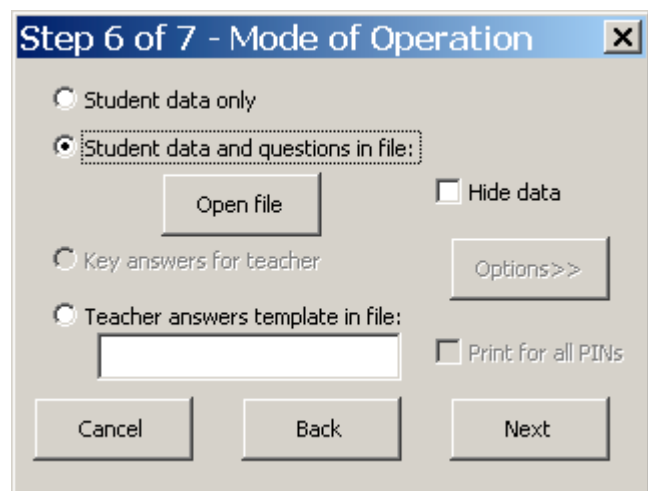
Step 4 (continued)

- Accept the default of selecting 0 columns from Group 1. Use the “spinner” to scroll to Group 2.
- Click on Group 2 to select it, then enter 2 in the **No. to select** box.
- Click on **OK** to return to Step 4.
- Click on **Next** to proceed to Step 5.
- Click on **Next** again to proceed to Step 6.



Step 6

- Click on the **Student data and questions in file** option and then on the **Open file** button.
- Navigate to the appropriate folder and open the *Questions.xls* file.
- Leave the **Hide data** option unselected.
- Click on **Next**.
- At this point the *Questions.xls* file will be copied into ISCUS and then closed. If you later return to this step to change a setting then you will need to re-open the questions file.



Step 7

- This is the same as in Example 1.
- Click in the checkbox to enable password protection of the student data generator.
- Enter the password **stars**.
- Click on **Finish**.

- You will be presented with the usual Excel **Save As** dialog. Choose to save the file under the name *STARS2.xls* in some appropriate folder,
- When the “Wizard terminated” message appears click on **OK** to close Excel.

To try it out:

- Open *STARS2.xls*, enable macros, enter the PIN 12345, then **Create Data** to create a new workbook of student data and questions.
- Examine the **Data** and **Questions** sheets. Note that the “random” cells of the question now contain fixed values not formulae.
- The idea is that the student now saves this file either to disk or to their personal network folder.
- Try running *STARS2.xls* again with a different PIN and check that both data and question are different.

Note that *STARS2.xls* is free-standing and is not linked in any way to the *Questions.xls* spreadsheet. There is no reason why a student should not take a copy of *STARS2.xls* and run it remotely on their personal computer.

Example 3 Key answers corresponding to student data

In this example, you will produce a spreadsheet of key “outputs” or answers for each student who uses the *STARS1.xls* data generator constructed in Example 1. This mode of operation **requires a list of five-digit PINs to be provided** in a spreadsheet, with the PINs in the first column and the corresponding names in the second column.

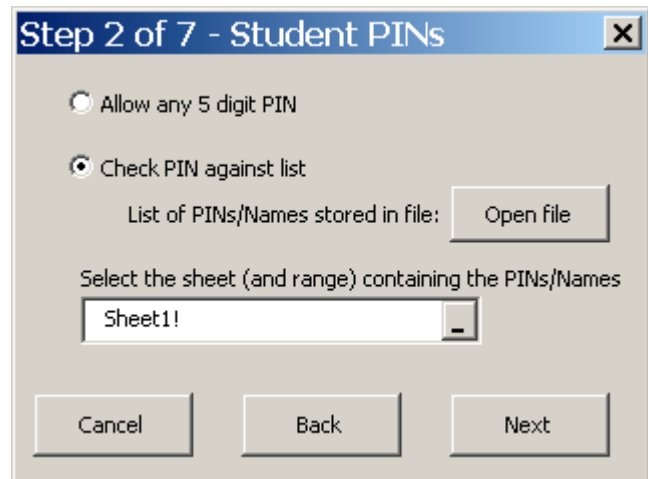
- Start Excel and select **File > Open** from the main menu. In the dialog box that appears choose to view **Files of type:** Excel (*.xls) and in the **Look in** box navigate to the folder containing *RDG.xls*. Click on this file and then **Open**.
- Select **Enable Macros** otherwise the Wizard will not work.
- Click on the grey button to run the seven-step Wizard.

Step 1

- Click on **Data stored in spreadsheet** and then on the **Open file** button.
- Navigate to the appropriate folder and open your data file *MyData.xls*.
- Click inside the box headed **Select the range containing the data**, then use your mouse to highlight the data range to be used as your source data set – we suggest **A1:H51** would be adequate for this exercise.
- Click on **Next**.

Step 2

- Click on the **Check PIN against list** option.
- Click on **Open file**, navigate to the appropriate folder and open the spreadsheet *Pins.xls*.
- Click inside the box headed **Select the range containing the PINs/Names**, then use your mouse to highlight the range A2:B5. (Alternatively, simply enter **Sheet1!** to indicate that you want to select all the PINs/Names on this worksheet.)
- Click on **Next** to continue. A message will confirm the number of PINs detected. Click **OK** to continue.



At Steps 3, 4 and 5 it is essential that the options chosen match those in Example 1. If not, the answers calculated for each PIN will not correspond to the student data generated.

Step 3

- Click on the **Same sample size for each student** option and enter **30**.
- Click on **Next**.

Step 4

- Accept the default setting of **all columns**.
- Click on **Next**.

Step 5

- Accept the default setting of **None** under **Stratify using levels in column**.

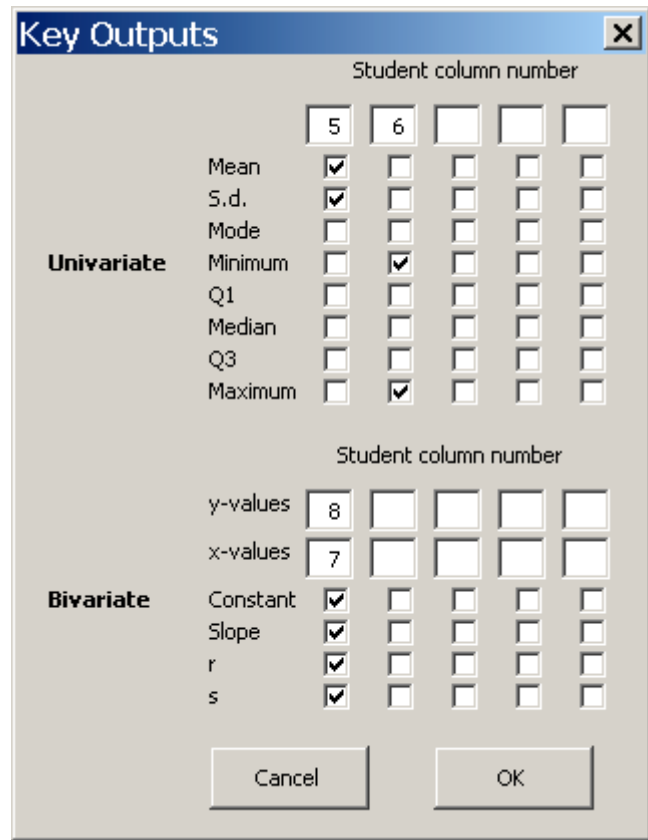
- Click on **Next**.

Step 6

- Click on the **Key answers for teacher** option, then click on the **Options>>** button.
- Recall that the student has been allocated all 8 columns of data. Click in the relevant checkboxes to calculate the answers shown, then **OK**.
- Click on **Next**.

Step 7

- This is the same as in Examples 1 and 2.
- Click in the checkbox to enable password protection of the student data generator.
- Enter the password **stars**.
- Click on **Finish**.
- You will be presented with the usual Excel **Save As** dialog. Choose to save the file under the name *STARS3.xls* in some appropriate folder.
- When the “Wizard terminated” message appears click on **OK** to close Excel.



To try it out:

- Open *STARS3.xls* and enable macros.
- Do not enter any PIN but simply click on **Create Data**.
- The answer sheet below will be generated in a new workbook.

	A	B	C	D	E	F	G	H	I	J	K
1	PIN	Name	N	Mean_5	Sd_5	Min_6	Max_6	a(8v7)	b(8v7)	r(8v7)	s(8v7)
2	12345	Brad Payne	30	3.7	2.890532	0	6	1.953779	-0.01379	-0.06472	1.120262
3	11111	Vanessa Simonite	30	3.966667	2.87058	0	6	1.266259	0.024533	0.08772	1.443631
4	54321	Penny Bidgood	30	4.266667	2.863966	0	6	1.87503	0.010476	0.032863	1.516118
5	99999	Neville Hunt	30	3.966667	3.145696	0	6	1.618088	-0.0034	-0.01746	0.999442

The abbreviated headings refer to columns in the student’s dataset. Mean_5 indicates the mean of the student’s 5th column, while a(8v7) indicates the intercept (or constant) term in the regression of the student’s column 8 (y) on column 7 (x). r is the correlation coefficient and s is the standard error about the regression line.

Note that further outputs can be added retrospectively to this spreadsheet using formulae. For example, if the teacher wanted to know the value of r-squared for each student, they could simply enter the formula =J2*J2 in cell L2 and copy down to L5. This significantly extends the range of possible answers.

Example 4 **Answers spreadsheet corresponding to student data**

In this example, a pre-prepared spreadsheet of statistical analysis will be linked to the data generator so that a comprehensive answer sheet can be prepared (and optionally printed out) for each student. Here answers will be produced for the student regression datasets generated in Example 2.

Preliminaries

- Start Excel and open the file *RegAnswers.xls* to inspect it. You will see that the workbook contains two sheets, one to receive the data from ISCUS and the other containing a full regression analysis of the data.
- Select the **Data** sheet. This shows a typical student's data, with a column of x-values in A, y-values in B and summary information in columns D and E.
- Now switch to the **Answers** sheet. This contains all the regression output usually provided by a statistical package. The formulae generating the output have been constructed in such a way that they update whenever the **Data** sheet changes.
- It is anticipated that the answers sheet will normally be prepared by the teacher, tailoring it to his or her own needs. However, this does assume a significant competence in Excel and is the subject of a further training session.

Wizard

- Start Excel and select **File > Open** from the main menu. In the dialog box that appears choose to view **Files of type:** Excel (*.xls) and in the **Look in** box navigate to the folder containing *RDG.xls*. Click on this file and then **Open**.
- Select **Enable Macros** otherwise the Wizard will not work.
- Click on the grey button to run the seven-step Wizard.

Step 1

- Click on **Data stored in spreadsheet** and then on the **Open file** button.
- Navigate to the appropriate folder and open your data file *MyData.xls*.
- Click inside the box headed **Select the range containing the data**, then use your mouse to highlight the data range to be used as your source data set – we suggest **A1:H51** would be adequate for this exercise.
- Click on **Next**.

Step 2

- Accept the default setting of **Allow any 5 digit PIN**.
- Click on **Next**.

Step 3

- Click on the **Random sample size for each student** option and enter Between **30** And **50**.
- Click on **Next**.

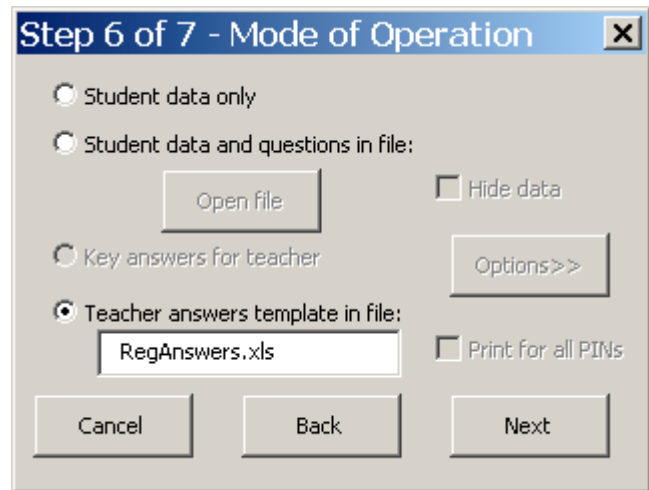
Step 4

- Click on **a selection of columns as structured below**, then click on the **Structure>>** button.
- Click in the checkboxes to select the continuous variables AGE through to DEBT.
- Click on **Create group**, to classify these variables as Group 2, then **OK**.
- Now click on the **Selection>>** button.

- Accept the default of selecting 0 columns from Group 1. Use the “spinner” to scroll to Group 2.
- Click on Group 2 to select it, then enter 2 in the **No. to select** box.
- Click on **OK** to return to Step 4.
- Click on **Next** to proceed to Step 5.
- Click on **Next** again to proceed to Step 6.

Step 6

- Click on the **Teachers answers template in file** option.
- Enter **RegAnswers.xls** in the white box, then click on **Next**.
- A warning message appears (see Note [1] below). Click **OK** to continue.
- The **Print for all PINs** option is not available here since no list of Pins has been provided. (Note too that for a big class this could take a very long time to complete!)



Step 7

- This is the same as in the preceding examples.
- Click in the checkbox to enable password protection of the student data generator.
- Enter the password **stars**.
- Click on **Finish**.
- You will be presented with the usual Excel **Save As** dialog. Choose to save the file under the name *STARS4.xls* in some appropriate folder, then close Excel.

To try it out:

- Start Excel and open *STARS4.xls*.
- Enable macros, type in the PIN 12345, press the Enter key, and click on **Create Data**.
- The answers sheet for this student should appear.
- Unlike in previous examples, the *STARS4.xls* spreadsheet is still open. This is to allow you to see the answers for another student. Use the **Window** menu to switch to *STARS4.xls* and enter the PIN 11111 instead. After a short pause, this should update the *RegAnswers.xls* workbook.

Note

- [1] *STARS4.xls* is electronically linked to *RegAnswers.xls* and will not therefore function unless the two files are stored together in the same folder.
- [2] The *RegAnswers.xls* workbook could have also contained a sheet of questions like the one used in Example 2.