CASIO

Sampling With Graphics Calculators

Mun Chou, Fong
QED Education Scientific
Malaysia

LEVEL
Senior high schools and first year of university after students have been introduced to Statistics.

OBJECTIVES
To discuss data collection and random sampling with a graphics calculator.

CORRESPONDING eActivity
RANDOM.g1e

OVERVIEW
Application of statistics in some scientific study or quality control of production at a factory normally begins after the important process of data collection, or sampling. We shall look at using graphics calculator to help us perform random sampling, random selection and simple surveying activities.

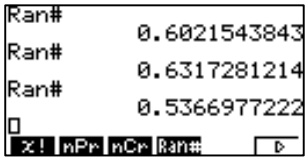
EXPLORATORY ACTIVITIES
[Note]
- We shall use small letter *x* instead of capital X as shown on the calculator throughout the paper.
- The Ran# of the calculator generates pseudo-random numbers.

Random sampling helps avoid bias-ness during data collection. A good way to choose a random sample is by drawing lots but this is not really feasible when the population size is very large. Here we look at some ways of generating "random number" in the graphics calculator, where we can use them to select a random sample.

Exploration 1: Suppose a chocolate factory produces 5,000 pieces of chocolate daily. The quality control division decided to randomly select 50 pieces (or 1%) of chocolates produced for its daily quality inspection (before analyzing the data statistically.)

As usual, today the quality inspector numbered the pieces of chocolates according to their production order, so the first piece of chocolate produced for the day is numbered as 1, the second piece is numbered as 2 and so on. Using his graphics calculator, he can randomly select 50 pieces out of these 5000 using the [Ran#] for inspection.

Open the Run strip "**Exp-1**". Now tap [OPTN] [F6] [F3] [F4] to enter [Ran#] then tap [EXE]. Follow this by tapping [EXE] a few more times.



Note that each tap of [Ran#] produced different ten-digit decimals and [Ran#] output actually observes the following inequality:

$$0 < \text{Ran\#} < 1$$

We can help the inspector to adapt this to suit his random sampling needs. If we multiply [Ran#] with 5000 we ensure that each output falls between 0 and 5000 as follow.

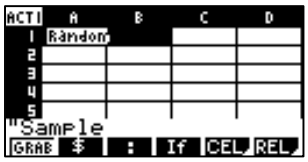
$$0 < 5000 \times \text{Ran\#} < 5000$$

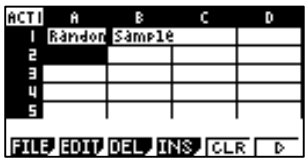
We could further adjust the [Ran#] to help him selects his random sample. While still in "**Exp-1**" strip tap [OPTN] [F6] [F4] [F2] to call out the function of [Int], followed by the expression 5000×Ran#, then add 1.

While the inspector can now generate 50 different numbers randomly out of 1 to 5000 at the [Run] window by pressing [EXE] 50 times, it is much more advantageous and efficient for him to use a spreadsheet to handle the sampling, recording the quality test results of the chocolate and even for statistical analysis.

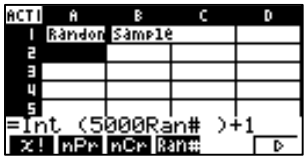
Activity 1: Use the spreadsheet mode of the graphics calculator to help the inspector randomly selects 50 chocolates to be tested today.

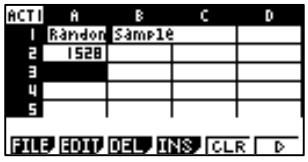
Solution:
(A) Open the Spreadsheet strip "**Act1A**". We shall use the A column for random number generation and named it as "Random". We also record the sample in the B column and named B column as "Sample". Then, scroll to the first empty cell of the "Random" column.



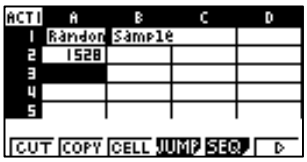


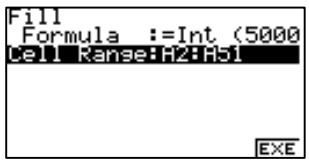
Enter the formula of "Int(5000Ran#)+1" to the selected cell of A2. It is similar to how we entered it before. As we tap [EXE] we will have generated a random number.

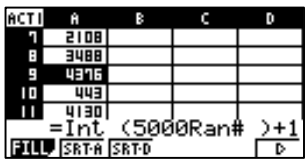




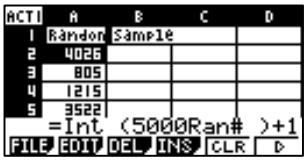
(B) For this activity we want to generate 49 more random numbers, and there is a quick way to do so. While still in the spreadsheet and cell A3 is highlighted, tap [F2] (EDIT) followed by [F6] [F1] (FILL). Enter the formula of "=Int(5000Ran#)+1" (enter only between the quotation marks) to [Formula] and set [Cell Range] to A3:A51 to generate the other 49 random numbers.

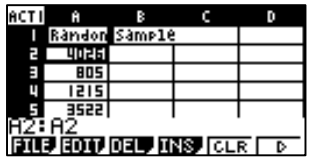


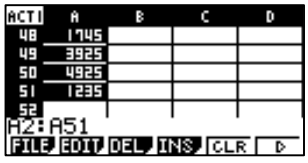




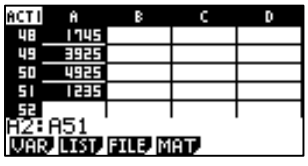
(C) Next we save the numbers in "Random" column to a list before recalling the same set of numbers to the "Sample" column. Scroll to the first number cell, or A2, of "Random" and tap [SHIFT] [8] (CLIP). With A2 highlighted, scroll down till the last of the number cell of A51. The cell reference displayed at the bottom left of the screen means we have highlighted the cell A2 till A51.



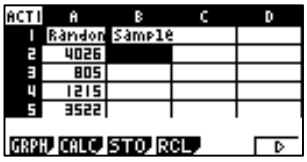


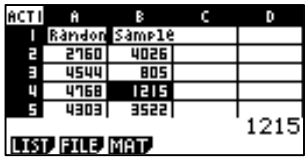


To store these numbers into a list, tap [F6] [F3] (STO) and followed by [F2] and set to store these numbers into [List1].

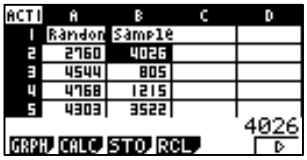


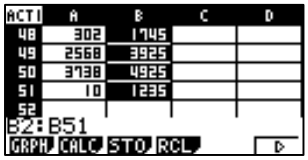
Go to the first empty cell of the "Sample" column and tap [EXIT] until we see the screenshot below left. Tap [F4] [F1] to recall entries of [List1] and put them in "Sample" column beginning at B2.

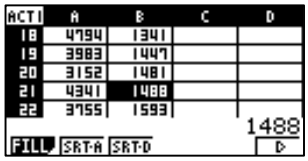




(D) Finally we sort the numbers at "Sample" column in ascending order and check for possible repeating numbers. Go to the first cell of "Sample" column, tap [SHIFT] [8] and scroll to the last non-empty cell of B51. To sort the column tap [EXIT] [F6] [F2] [F6] [F2] (SRT.A), then explore the column and check to see if there is any repeating number.







The random numbers we have in this discussion contains no repeating number (the random numbers you generate should be different from the above,) but if we do see repeating numbers, we just begin once again at section (C).

Therefore according to the random numbers we helped generate, the quality inspector will be inspecting the chocolates produced according to the order displayed in the "Sample" column in the Spreadsheet strip of "**Act1A**" today. □

The quality inspector could always use this same spreadsheet to select a random sample for his work everyday and even to record the quality check. Furthermore, the spreadsheet could be improvised to do more and be more flexible.

Activity 2: The factory sometimes produces between 4000 and 4500 pieces of chocolate daily, apart from the usual 5000. The sample size required for quality testing is always 1% of the daily production amount. In the test, each piece of chocolate either passes (given 1 point), or fails (given 0 point.) Not more than 8% of the sample should fail the daily quality test or further control action will be taken. The following is the planned daily production for the next 5 days.

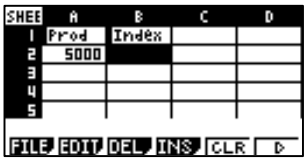
Day	1	2	3	4	5
Chocolate Production	5000	4000	4500	5000	5000

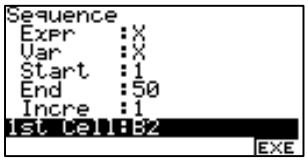
Improvise the spreadsheet to help the inspector perform the above on a daily basis.

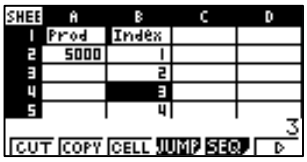
Solution:
In summary here are the daily sample size needed and the 8% of sample size.

	Day1	Day2	Day3	Day4	Day5
Chocolate Production	5000	4000	4500	5000	5000
Sample size	50	40	45	50	50
8% of sample size	4	3.2	3.6	4	4

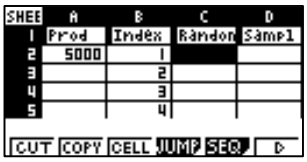
(A) Open the Spreadsheet strip "**Act2A**". Here we set the daily production as parameter in the random number generation. Cell A2 is used to record the daily production. We also use the B column as an indexing column for easy reference to the amount produced and the random numbers. We begin with the daily production of 5000 for Day1.

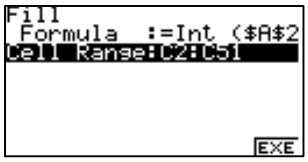


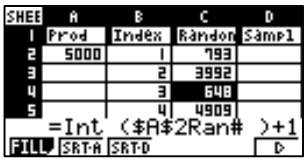




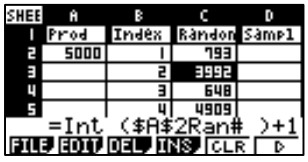
(B) The "Random" column and "Sample" column are put in C and D columns respectively. To set up the spreadsheet to generate 50 random numbers, tap [F2] (EDIT) followed by [F6] [F1] (FILL). This time we enter the formula of "=Int(A2Ran#)+1" to [Formula] and set [Cell Range] to C2:C51.



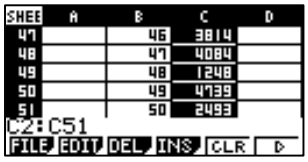




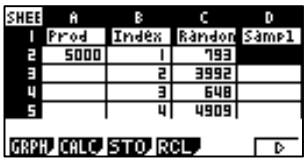
(C) Here we have generated 50 random numbers for Day1. Again we want to save the numbers in the "Random" column to [List1] before recalling the same set of numbers to the column "Sample", sort these numbers and check for repeating number. First let's return to the main spreadsheet menu by tapping [EXIT] (see below.)

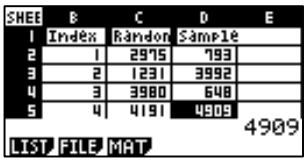


Go to cell C2 tap [SHIFT] [8] (CLIP) then scroll down till cell C51. With cell C2:C51 selected, tap [F6] [F3] (STO) and followed by [F2] and store these numbers into [List1].

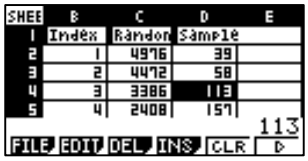


Press [EXIT] to return to the main spreadsheet menu (see screenshot below left.) Go to the first empty cell of the "Sample" column, tap [F4] [F1] to recall entries of [List1].

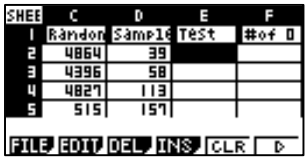


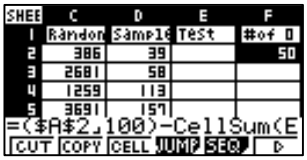


(D) Now we sort the numbers at the "Sample" column in ascending order and check for possible repeating number. Again use [SHIFT] [8] to highlight cell D2 till cell D51 and sort the column with [F6] [F2] [F6] [F2] (SRT.A), then explore and check for repeating number.



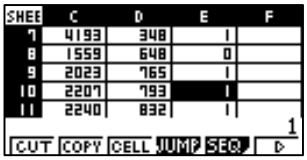
(E) We add two other columns called "Test" to record the result of the quality test, and another called "#of 0" to calculate the number of 0 (fail quality test.) The "#of 0" column actually consists of just the cell F2 which tells the number of fail cases with the formula "$\frac{\$A\$2}{100}$-CellSum(E2:E51)". Now the inspector can use this spreadsheet for his daily work.

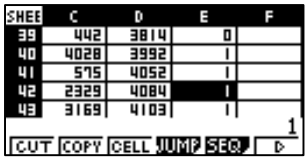


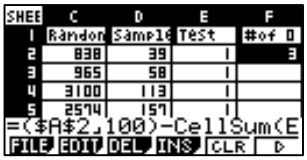


Day1

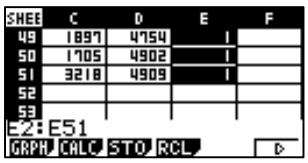
For the discussion of this activity, we use the random numbers generated above for Day1 quality test and the test results, shown partially, is supposed to be as follow.







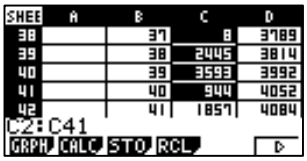
The results mean that the chocolates produced on Day1 passed the quality inspection and no further control action is needed. We can save the random numbers and test results in [List1] and [List2] as record using [SHIFT] [8] (CLIP) and [STO].



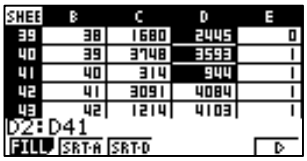
So the Day1 random sample and test result are stored in [List1] and [List2].

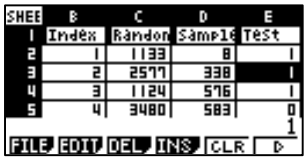
Day2

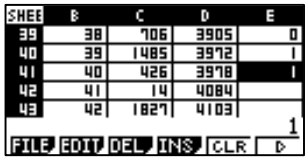
Change the daily production at cell A2 to 4000. Repeat process in section (C) and (D), but consider only the first 40 numbers in the "Random" column. The random sampling is not affected in this case although we generate 50 random numbers at the "Random number". Also, use [List3] as the temporary storing list instead.



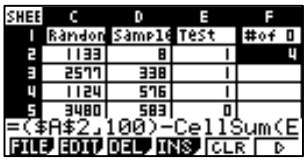
When pasting entries of [List3] to the "Sample" column, contents of cell D2 till D41 are replaced with the new set of random numbers. We can use the "Index" column as guide in selecting the newly generated 40 random numbers for sorting, and avoid selecting cell D42 till D51 (see below left.)





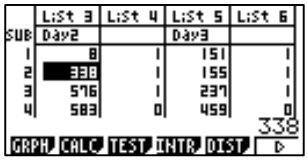


We assume that the test results, shown partially above, as the actual results. Contents in the cell of E42 till E51 must be deleted to ensure the correct number of 0 is displayed.



The number of 0 recorded is 4, which means that the chocolates produced on Day2 do not pass the quality inspection and appropriate control action is needed. We now save the random numbers and test record in [List3] and [List4] as record.

With similar approach the quality inspector selects the appropriate random sample for his Day3, Day4 and Day5 tests, then stores the corresponding random numbers and their corresponding results to [List5] and [List6] for Day3, [List7] and [List8] for Day4, and [List9] and [List10] for Day5.



You can view the full records in the [List] mode by tapping SHIFT) in the "**Act2A**" spreadsheet strip and select to view [List Editor]. The numbers you will be working with should be different from the numbers recorded there. □

The sampling method we have used above is called simple random sampling. One other method of random sampling is stratified sampling which based on identifiable strata. Useful strata might be 'males' and 'female', or 'obese' and 'non-obese'. A stratified sampling is made up of separate simple random samples for each of the strata. If we know the proportions of the population falling into these different categories, we should ensure that the proportions are reproduced by the sample.

Activity 3: Suppose your school student population can be described as follow:

Male students	Female students
45%	55%

Use stratified sampling to choose a random sample of 40 students.

Solution:
In this case, we would choose a simple random sample of 40×45%=18 male students and another simple random sample of 40×55%=22 female students. □

EXERCISE

Exercise 1

Suppose the student population of a school is 1750 and the male-female composition is as described below.

Male students	Female students
38%	62%

Use the spreadsheet to select a stratified sample of 50 students for a survey on students' preference between almond and milk chocolate. Design the spreadsheet to record the responses to the survey if possible.

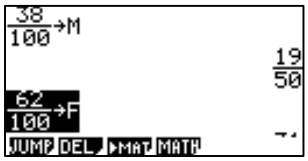
SOLUTIONS to EXERCISE

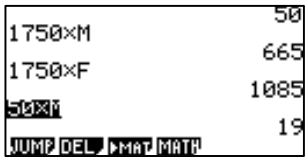
Exercise 1

First we select the stratified sample then use the technique of Activity 2 on the sample. In summary here is the information about the two categories.

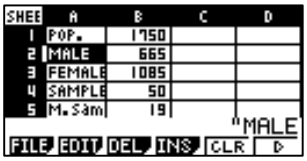
	Male students	Female students
% of population	38%	62%
Of 1750 population (%×1750)	665	1085
Stratified sample size (%×50)	19	31

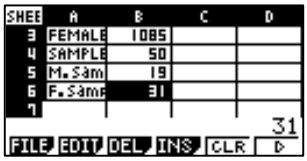
You can use Run strip "**Ex1A**" to calculate the relevant information shown above.



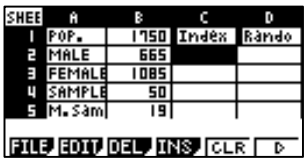


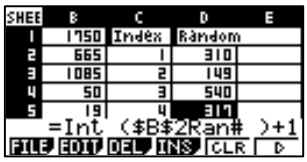
One possible ways to design the spreadsheet is as follow:
- Open the Spreadsheet strip "**Ex1A**". Use columns A and B to record the population and the sample size of each stratum, so you would have more control of the parameters.



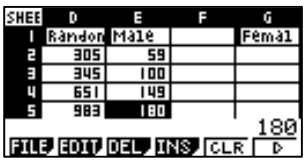


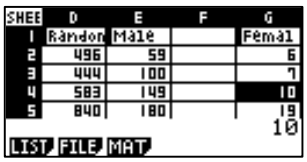
- Set up the 3rd column for indexing and 4th column for random number generation. Firstly select the random sample of male students with formula of "=Int(B2Ran#)+1" for D2 till D20 (sample size of 19.)



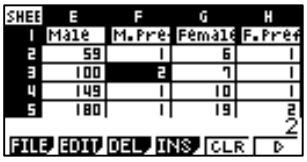


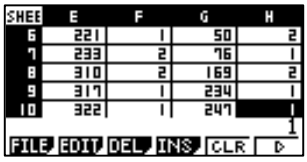
- Create two more columns called 'Male' and 'Female'. Copy, paste and sort the numbers at the 'Male' column with methods used in Activity 2. Then select the random sample of female students with the formula "=Int(B3Ran#)+1" for D2 till D32 (sample size of 31) at the "Random" column and copy-sort the numbers to the 'Female' column.





- Use the F column to record the responses of male students for the survey and the H column for female students' responses. Below are screenshots of possible responses where 1 is for students who prefer almond chocolate and 2 for students who milk chocolate. □





REFERENCE

[1] G. Upton and I. Cook, *Introducing Statistics 2nd Edition*, Oxford University Press, 2001. ISBN: 0 19 914 801 5.
[2] S. Dobbs and J. Miller, *Statistics 2*, Cambridge University Press, 2001. ISBN: 0 521 78604 5.

