Confidence Interval for Population Mean

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 interval estimation of population mean with the aid of a graphics calculator.

CORRESPONDING eActivity
INTERVAL.g1e

OVERVIEW
Calculating the confidence interval is very informative as it provides a range of plausible values for the population parameter being studied. In this paper our discussion shall be on simple random sampling and finding the confidence intervals of the population mean using the spreadsheet and statistics functions of the graphics calculator.

EXPLORATORY ACTIVITIES
[Note]

- We shall use small letter *x, s* instead of capital X, S as shown on the calculator throughout the paper.
- The Ran# of the calculator generates pseudo-random numbers.

The confidence interval can be efficiently found with the graphics calculator where this leaves us with more time to focus on the interpretation and analysis of the data.

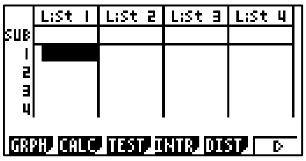
Activity 1: Suppose a random sample of 64 sweets is selected and the mean mass of these sweets is found to be 0.932 gram and the value of standard deviation *s* is 0.1 gram. Find the 95% confidence interval for the population mean mass.

<u>Solution</u>:
The sample mean is $\bar{x}$ =0.932, the sample standard deviation is s=0.1 and we also have n=64.

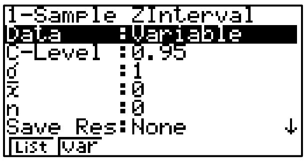
As the population variance is unknown and the sample size is > 30, we can replace the population variance σ by s. Therefore the 95% confidence interval can be calculated as follow, where $z_{0.025}$ denotes the value such that $P(|Z| < z_{0.025})$ =0.95 where $Z \sim N(0,1)$.

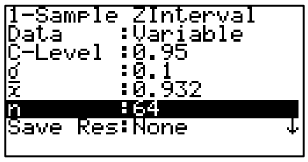
$$\left(\bar{x} - z_{0.025}\frac{s}{\sqrt{n}},\ \bar{x} + z_{0.025}\frac{s}{\sqrt{n}} \right).$$

We can find the same confidence interval at the graphics calculator. Open the eActivity INTERVAL.g1e, and then scroll down to open the <u>LIST</u> strip "**Act1A**".

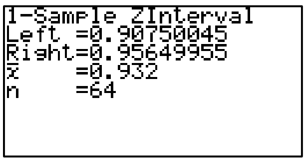


Now tap F4 F1 F1 (1-S) to enter the 1-sample Z interval menu. Key in the statistics provided to the corresponding parameter and set data source to [Variable] as we are not calculating the confidence interval using data from a sample (see below right.)





Pressing EXE would give us the lower (left) and upper (right) confidence limit as 0.9075 and 0.9565 correct to 4 decimal places.

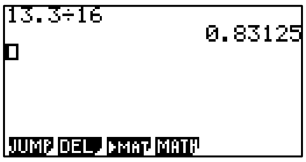


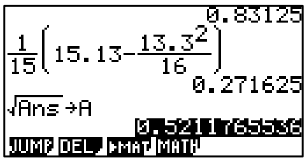
In short the approximate 95% confidence interval for the mean mass of the sweets is $(0.9075,\ 0.9565)$, or $0.9075 < \mu < 0.9565$. □

In actual practice, often the sample size is constrained by many factors and therefore sample size could be less than 30. When such small sample is taken from a normal distribution with unknown variance, it is more appropriate to base the confidence interval on a t-distribution.

Activity 2: Suppose that due to constraint of time, a random sample of 16 sweets is chosen instead of 64 in Activity 1 above. The mass of each sweet, x gram is determined and the measurements are summarized by $\sum x = 13.3$ and $\sum x^2 = 15.13$. Assuming the mean mass have a normal distribution, find the 95% confidence interval for the population mean. State the confidence interval correct to 3 decimal places.

Solution:
Although the sample mean and sample variance are not provided here, they can be calculated easily which we can do so at the Run strip "**Act2A**".





From the calculation we should find the sample mean $\bar{x} = 0.83125$ and sample variance $s^2 = 0.271625$ which in turn gives us s=0.52118, correct to 5 decimal places.

[**Note**] The sample variance is found by calculating

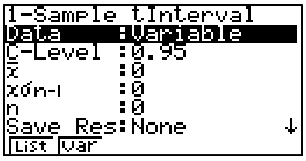
$$s^2 = \frac{n\sum x^2 - (\sum x)^2}{n(n-1)} = \frac{1}{n-1}\left(\sum x^2 - \frac{(\sum x)^2}{n}\right).$$

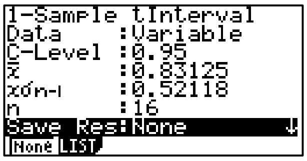
This confidence interval will be based on the t_{15}-distribution, and the 95% confidence interval can be calculated using

$$\left(\bar{x} - c\frac{s}{\sqrt{n}},\ \bar{x} + c\frac{s}{\sqrt{n}}\right).$$

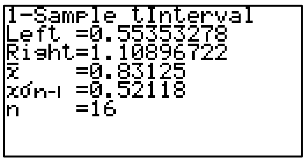
Here c is the relevant percentage point for probability of p=0.975 and degrees of freedom 15. The probability is 0.975 because the combined probability of 1-0.95=0.05 in the two tails of the symmetric distribution implies 0.025 in the upper tail and hence p=0.975.

We can find this same confidence interval at the graphics calculator. Open the LIST strip "**Act2B**". Now tap F4 F2 F1 (1-S) to enter the 1-sample t interval menu. Set data source to [Variable] and set [C-Level] to 0.95. Key in $\bar{x} = 0.83125$ we have calculated just now, and as this confidence interval is based on a t-distribution, use s=0.52118 to replace the unknown σ (shown as $x\sigma_{n-1}$ in the calculator.)





Pressing EXE should give us the lower and upper confidence limit as 0.554 and 1.109 respectively, correct to 3 decimal places.



In short the approximate 95% confidence interval for the mean mass of the population is $0.554 < \mu < 1.109$. □

In the following activity we look at using the spreadsheet to provide immediate feedback.

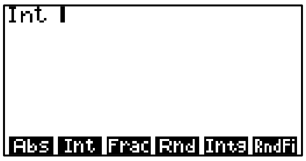
Suppose a company which produces sweets wants to find out the mean mass of the sweets they produces such that that they can print this information on the wrapper. As the factory produces 10,000 pieces of sweets per day, choosing a random sample is the more feasible approach. You as the marketing manager decided to choose a random sample of 40 sweets for this task and since nothing of this nature has happened before, there is no information about the products (population.)

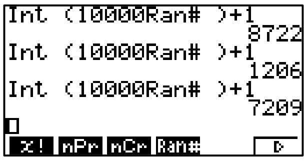
Activity 3: Use the graphics calculator to,
(a) Select a random sample of 40 pieces of sweets.
(b) Record the mass of each sweet and calculate the summary statistics of the data.
(c) Find the 99% confidence interval for the population mean.
Then, discuss about the mean mass of the population.

<u>Solution</u>:
(a) We begin by randomly select 40 numbers using the [Ran#] feature. Note that [Ran#] outputs different ten-digit decimals at every tap which observes this inequality:

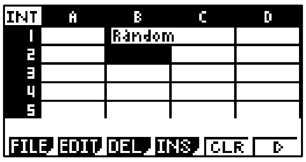
$$0 < \text{Ran\#} < 1$$

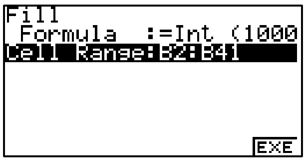
When we multiply [Ran#] by 10000, we would have 0<10000×Ran#<10000. If we call out the [Int] function (tap OPTN F6 F4 F2 in the <u>Run</u> strip "**Act3A**"), followed by the expression of 10000×Ran#, and add 1 to it, we should have an expression which produces integers between 1 and 10,000 inclusive.

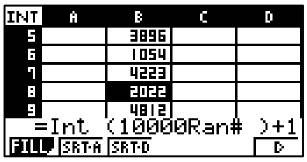




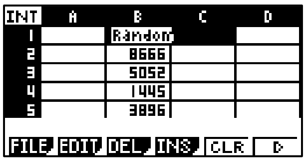
This expression can be duplicated in the spreadsheet mode. Open the <u>Spreadsheet</u> strip "**Act3B**". We leave the A column for later use and named the B column 'Random'. To set up 'Random' to generate 40 integers out of 1 to 10,000, tap F2 (EDIT) followed by F6 F1 (FILL). Then, enter the formula of "=Int(10000Ran#)+1" to [Formula] and set [Cell Range] to B2:B41.



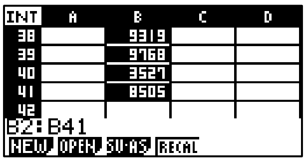


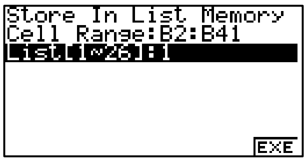


Having randomly generated the 40 numbers, we save these numbers to [List1] before recalling the same set of numbers to the column of C. At the column we shall sort the numbers in ascending order and check for any repeating number. First, tap EXIT once to return to the main spreadsheet menu.

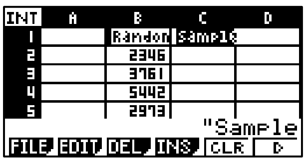


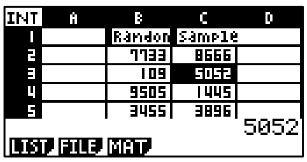
Go to cell B2, tap SHIFT 8 (CLIP) then scroll down till cell B41. With cell B2:B41 selected, tap F6 F3 (STO) and followed by F2 and to store these numbers into [List1].



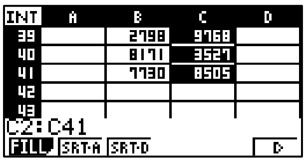


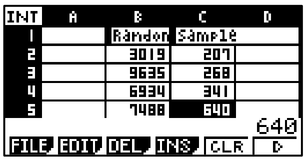
Press EXIT to return to the main spreadsheet menu and named the C column as 'Sample' then go to the first empty cell of the "Sample" column, tap F4 F1 to paste the entries of [List1].





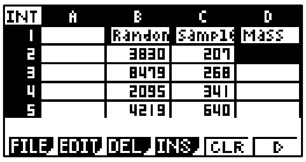
To sort 'Sample" column, again use [SHIFT] [8] to highlight cell C2 till cell C41 and sort the column with [F6] [F2] [F6] [F2] (SRT.A), then explore and check for repeating number.

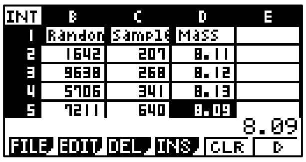


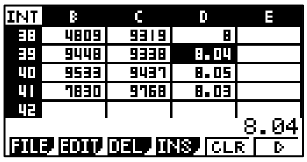


The list has no repeating number and therefore we have selected a random sample of 40 sweets. So the first sweet to be weighed is the 207th sweet out of production, the second sweet to be weighed is the 268th and so on (view the 'Sample' column in "**Act3B**".)

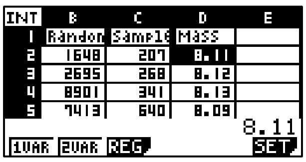
(b) To begin recording the mass of the sample, we create a column called 'Mass' to store the data of the sweets mass. When the production begins we collect the mass of the 207th, 268th, 341st …9437th and 9768th sweets produced and key in the masses accordingly. Assume the partial data shown here, in gram, are actual data recorded (explore the 'Mass' column in "**Act3B**".)

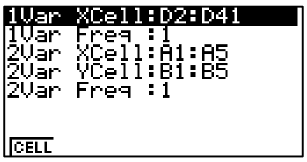




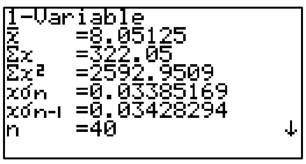


To have a sense of the sample, we can calculate the summary statistics. While still in the menu as displayed in the screenshots above, tap [F6] [F2] (CALC) and tap [F6] to set the calculator to find summary statistics of the data recorded in cell D2 to D41.



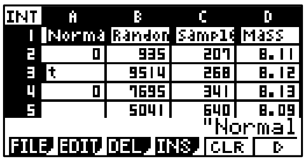


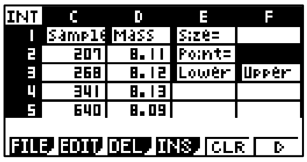
Then by tapping [F1] to select [1 VAR] we get the summary statistics of the sample.

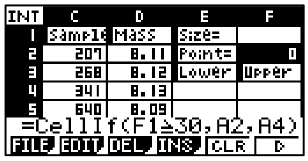


Note that we have $\bar{x}$ =8.05125 and $x\sigma_{n-1}$ =0.03428294.

(c) We can improvise the spreadsheet concerned to calculate the confidence limit and thus the confidence interval of the sweets mean mass. For this activity we set up the A column for the normal and *t* distribution percentage points as in the screenshot below left. We also arrange for the spreadsheet to use the appropriate percentage point (see cell F2 below) based on the sample size entered. The general rule of thumb is to regard a sample of 30 data as large, and the percentage point is based on standard normal distribution.



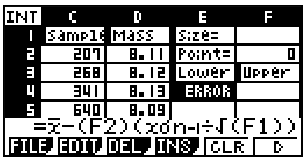


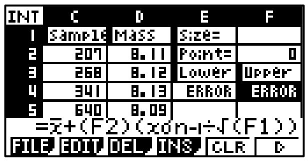


In a real life situation as with this activity, the variance σ is seldom known. When this happens we use the sample standard deviation s to replace σ, irrespective of the sample size. Then the confidence interval is calculated as follow, where *p* is the appropriate percentage point.

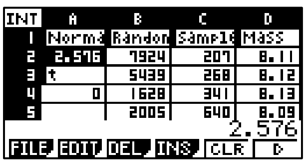
$$\left(\bar{x} - p\frac{s}{\sqrt{n}},\ \bar{x} + p\frac{s}{\sqrt{n}}\right)$$

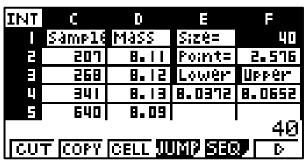
With this in mind the expressions to find the lower and upper confidence limits are entered, in this case, into cells E4 and F4 as follow.

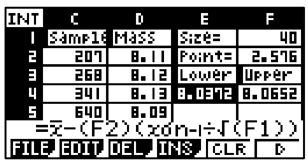




Note that $\bar{x}$ at present stores the value of 8.05125, and $x\sigma n-1$ is 0.03428294 when we calculate the summary statistics for the mean mass just now. Also, we entered both $\bar{x}$ and $x\sigma n-1$ of cells E4 and F4 with the catalog (tap SHIFT 4.)
For this activity the sample size (40) is considered large and hence the sample mean is taken to have a normal distribution. From a standard normal distribution table we have the percentage point which corresponds to 99% confidence level as 2.576.



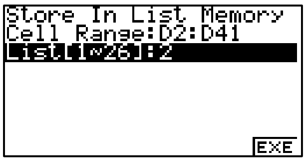


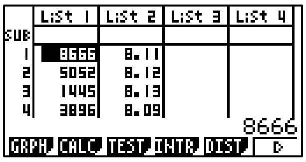


Since the sample mean is considered to have normal distribution, we entered the value of 2.576 to the cell A2, followed by the sample size of 40. As observed the lower and upper confidence limits are 8.0372 and 8.0652 respectively, which means that the 99% confidence interval for the population mean mass is $(8.0372,\ 8.0652)$.

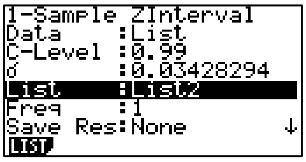
There is a 99% chance the population mean mass lies in the interval above. The company could state the sweet mass as 8.05g, correct to 2 decimal places, on the wrapper. □

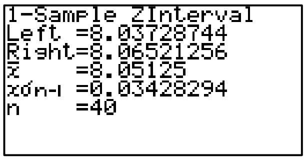
Apparently the quicker way to find the confidence interval of Activity 3 is by using the [INTR] function at Statistics mode, as seen in Activity 1, where firstly, we store the data at cell D2 to D41 to say [List2]. Then, tap [SHIFT] [,] to switch to the [List Editor] window.





Tap [F4] [F1] [F1] (1-S) to enter the 1-sample Z interval menu. Set data source to [List2], where the data is stored, then set [C-Level] to 0.99. As variance is unknown, we use s=0.03428294 as an approximation of σ (shown as $x\sigma_{n-1}$ in the calculator.) The confidence interval should be same as the one found in Activity 3.





Finding the confidence interval using the above method is very much faster but a properly created spreadsheet is more interactive, versatile and flexible. Moreover, the data and key information are available for viewing at the same window.

EXERCISES

Exercise 1

The random variable Y has a normal distribution with mean μ and unknown variance. A random sample of 200 observations of Y gives $\sum y$ =541.2, $\sum y^2$ =1831.42. Find (i) a 90% confidence interval for μ, and (ii) a 98% confidence interval for μ.

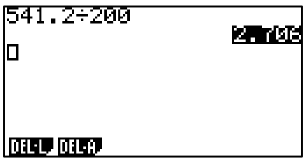
Exercise 2

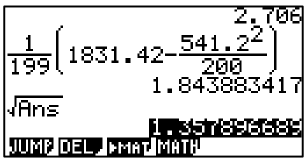
Redo Activity 3 but select a random sample of 18 only.

SOLUTIONS to EXERCISES

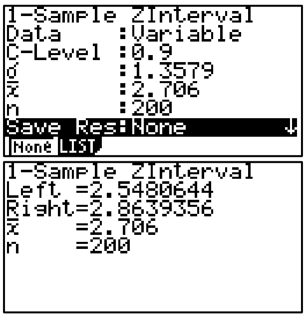
Exercise 1

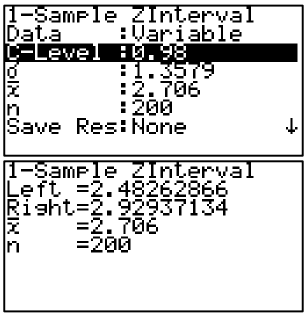
The sample mean and s can be calculated easily at the Run strip "**Ex1A**".





We should get the sample mean $\overline{x} \approx 2.7060$ and $s \approx 1.3579$.

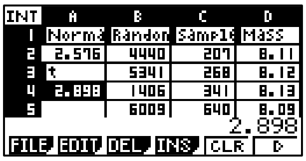


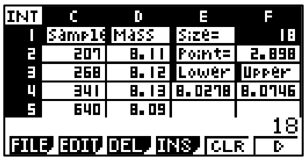


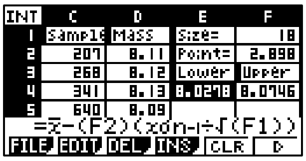
Using the <u>LIST</u> strip "**Ex1B**" we should find the 90% confidence interval for μ as (2.548, 2.864) and the 98% confidence interval for μ as (2.483, 2.929), correct to 3 decimal places. □

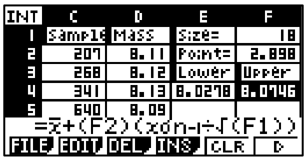
Exercise 2

With the sample size of 18, it is more appropriate to base the confidence interval on the t_{17}-distribution. From the table we should have the percentage point for degree of freedom of 17 and probability of 0.995 as 2.898. You should try creating a spreadsheet similar to the one in Activity 3 to solve the problem but for the discussion of this solution we used the spreadsheet at Activity 3 and input the new percentage point and the new sample size.









The 99% confidence interval for the population mean should be (8.0278, 8.0746). □

REFERENCE

[1] G. Upton and I. Cook, *Introducing Statistics 2nd Edition*, Oxford University Press, 2001. ISBN: 0 19 914 801 5.

