Confidence Intervals

Jozef Hvorecky
Vysoká škola manažmentu / City University
Bratislava, Slovakia

LEVEL
High school or university students with basic knowledge in Mathematics and first notions of Statistics.

OBJECTIVES
To use the calculator's built-in functions to make elementary statistical evaluations for real-life problems.

Corresponding eActivity
S06INTVL.g1e (for Activity1)

OVERVIEW
We will see why data collections for a statistical evaluation must be limited and still remain a base for relevant decisions.

EXPLORATORY ACTIVITIES
[Note]
We shall use small letter *x* instead of capital X as shown on the calculator throughout the paper.

Here we describe one activity. For its mathematical background refer for example to [LM], page 226-255.

Activity 1 (S06INTVL.g1e):

Statistics studies real-life situations and tries to make relevant conclusions based on collected data. All we know that real-life data are not constant. People are born, grow up, move from place to place, put on or loose their weight and so on. For that reason, a seemingly trivial question can hardly be solved completely. Say: "What is the mean weight of the population in our town?"

First, weighting thousands (possibly, hundreds of thousands or millions) persons is a time-consuming task. No one can guarantee that some citizens refuse doing so. Even if all agree, how long will the measurement last? Until the last figure is determined, the first person can be a few kilograms lighter (or heavier) – and we can start from the beginning. Even if everyone agreed to measure their weight on 17 September, 8:30, we are not sure that their balance is accurate. Moreover, we have to collect and evaluate all the inputs. It also takes time. So, the value we get was correct at the date – but it is already outdated.

For that reason, in Statistics we simplify the task. Instead of chasing for unrealistic "ideal solutions", we simplify the tasks, search for their answers and try to estimate a level of precision in which the answers can be assumed (sufficiently) correct.

Let us solve a similar problem: "What is the total number of pets owned by students of our school?"

Solution:
One – the less realistic – method is to ask all students about the number of pets they own. As we expressed above, it is time consuming and there is not yet 100% warranty that the answer is true.

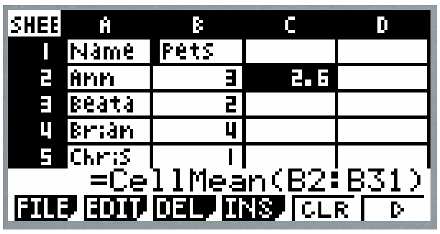
The second method is based on selecting an appropriate sample. Our class can be it. Let us explain why:

- We form a reasonably large proportion of the student body so the results will give a good picture on the real situation.
- We are not specific from the point of view. It is unlikely that we would own much more (or much less) pets that comparable students in other classes.

Notice that the second presumption may not always be true. If our class is a special (e.g. biology-oriented) one with extreme interests in animals among students, we are not "the right sample". Or, our class might consist of a group of students with an unusually high level of allergy to furs. In both cases, the evaluation would be negatively affected by the wrongly pre-selected combination of individuals. In such a situation, it is more appropriate to find another sample group – more similar to "average students".

(a) (Refer to <u>Pets</u>) The class of 30 students recorded the numbers of their pets in a spreadsheet table. Presuming that the class is an "average class", how many pets are owned by all 480 students of the school?

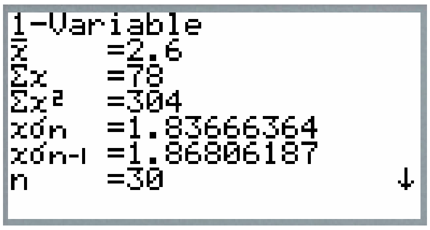
The mean is 2.6 pets per student, which is recorded in the cell C2 as shown below:



With 480 students in the school, the answer is

2.6 x 480 = 1248 animals.

Another value we will soon need is the standard deviation. It is calculated automatically after pressing F6 F2 (CALC) and then F1 (1VAR).



In this list, it is the value specified as $x\sigma_n$ i.e. 1.83666364.

(b) (Refer to Standard Error) It is difficult to trust the result ("1248 animals") without any reservations. Can we make any error estimation?

In Statistics, the estimation is derived from the so-called standard error of the mean σ_x which is defined as follows:

$$\sigma_x = \frac{\sigma}{\sqrt{n}}$$

where σ is the standard deviation and n is the size of the sample.

Calculate the standard error presuming the standard deviation is 1.836 and the sample consists of 30 students:

$$\sigma_x = \frac{1.836}{\sqrt{30}} = 0.335$$

(c) (Refer to Confidence Int) One can hardly expect that all students really own 1248 pets. At the same time, we would like to have a certain confidence in the result. We can presume that even if it is not exactly 1248, there is an interval in which the correct value belongs. It is certainly something between 0 and 3 million, but it is a too rough estimation.
Using the standard error, we can calculate two important values:
95 percent confidence interval:

$$\overline{X} \pm 1.96\frac{\sigma}{\sqrt{n}}$$

99 percent confidence interval:

$$\overline{X} \pm 2.58\frac{\sigma}{\sqrt{n}}$$

So, the intervals are calculated from the mean, standard deviation and the size of the sample.
Notice that, by a consensus among statisticians, the calculations are assumed acceptable when the sample is large enough (i.e. having at least 30 elements). This holds for our sample.

(d) (Refer to 95% Confidence) With 95 percent confidence, the mean number of pets per student is between 1.945 and 3.255.

$$\overline{X} \pm 1.96\frac{\sigma}{\sqrt{n}} = \begin{cases} 2.6 - 1.96\dfrac{1.836}{\sqrt{30}} \\ 2.6 + 1.96\dfrac{1.836}{\sqrt{30}} \end{cases} = \begin{cases} 1.945 \\ 3.255 \end{cases}$$

Multiplying these limits of the confidence interval by 480 students, we get 933 and 1562 pets as the limits of their intervals of confidence.

(e) (Refer to 99% Confidence) The results are now 1.738 and 3.462. This implies the numbers of pets between 834 and 1662. As above, the limits for the 99% confidence interval are results of multiplying the limits for the mean per student (i.e. 1.738 and 3.462) by 480.

EXERCISES

***Exercise 1*.**
Using spreadsheet calculation, calculate the mean of pets and the standard deviation for your sample group. Form a group of at least 30 persons to make your sample group big enough for further calculations.

***Exercise 2*.**
Using the data from Exercise 1, compute the standard error. Estimate the number of pets owned by students of your school within the 95 percent confidence interval.

***Exercise 3*.**
A survey of 49 smokers shows that their mean expense for cigarettes is $20 per week with the deviation $5. Using 0.95 degree of confidence, how much an individual spends for the cigarettes per week?

SOLUTIONS to EXERCISES

Exercise 1
The process should resemble the one described in the above eActivity. Do not forget to specify what “a pet” means. For example, decide whether aquarium fish is counted or not.

Exercise 2
From the spreadsheet data and results, you can determine the mean per student and the standard error. Then, you have to find out the number of students visiting your school and multiply it by the figures.

Exercise 3
Between $18.60 and $21.40.

REFERENCE

[LM] Douglas A. Lind and Robert D. Mason, *Basic Statistics for Business and Economics*, Irwin/McGraw-Hill, 1997. ISBN 0-256-19408-4

