CASIO

Mean and Standard Deviation

Ma. Louise De Las Penas, Phd
Ateneo de Manila University
Philippines

LEVEL
High School and University

OBJECTIVES
To study and interpret the mean and standard deviation of given data sets with the aid of the graphics calculator.

Corresponding eActivity
mean.g1e

OVERVIEW
The **mean** is a measure of central tendency calculated by dividing the sum of all values by the number of values in the data set. The **standard deviation** is a measure of the spread that is given by the positive square root of the variance.

In this paper, we will illustrate how the numerical and graphical analyses of the mean and standard deviation can be carried out using the graphics calculator.

ACTIVITIES
Example 1. In a genetic study, regular food was placed in each of 20 vials and the number of flies of a particular genotype feeding on each vial recorded. The counts of flies were also recorded for another set of 20 vials that contained grape juice. The following data sets were obtained:

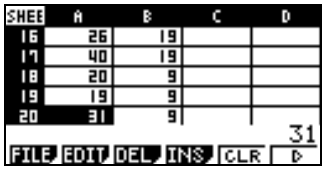
NUMBER OF FLIES (REGULAR FOOD)									
15	20	31	16	22	22	23	33	38	28
25	20	21	23	29	26	40	20	19	31

NUMBER OF FLIES (GRAPE JUICE)									
6	19	0	2	11	12	13	12	5	16
2	7	13	20	18	19	19	9	9	9

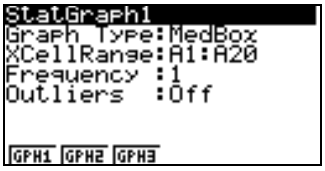
a. Make a visual comparison of the two distributions with respect to their centers and spreads.
b. Calculate the mean and standard deviation for each data set.

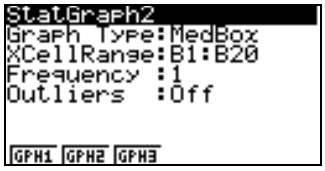
Solution:
We access the Spreadsheet Editor.

The first data list is entered in column *A* and the second data list is entered in column *B* as follows:

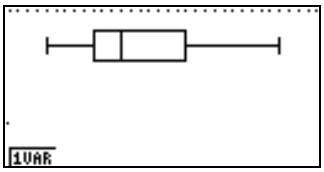


a. To display the data graphically, we access the GRPH menu. We assign the first graph, GPH1 to column *A* and the second graph, GPH2 to column *B* and select " boxplot"[MedBox] as the graph type.

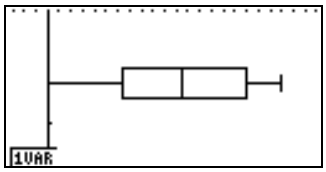




Box plots for both data lists will be generated as follows:



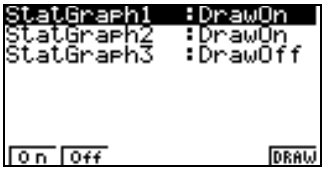
regular food

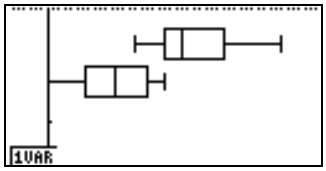


grape juice

A graphical comparison can be obtained by drawing both graphs on the screen at once. We choose SEL to select graphs to be drawn as follows:

Mean and Standard Deviation



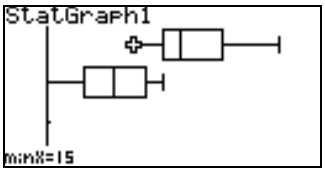


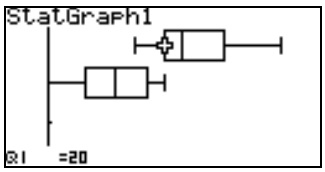
Note that the centers or the medians of both distributions vary (as represented by the vertical lines in the given boxes). The data for flies on grape juice (bottom box plot) represents a more symmetric data set, the line representing the median being in the middle of the box. Both sets of data appear to be "spread out" in the same manner.

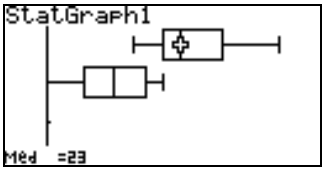
As we trace [Trace][Shift F1] through the given box plots, we see the differences in the numbers used to create them.

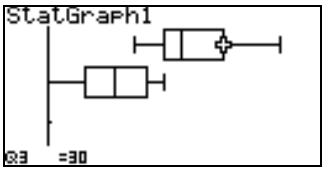
A **box plot**, gives a graphic representation of data using five measures, the median, the first and third quartiles, and the smallest and largest values in the data set between the upper and lower inner fences. The quartiles are the summary measures that divide a ranked data set into four equal parts. The second quartile is the same as the median of a data set. A box plot helps us visualize the center, the spread and the skewness of a data set. In this example, constructing box plots is helpful to allow us to compare both distributions relative to their centers and spreads.

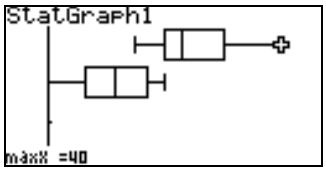
Note that for the data on the flies on the regular food, the minimum number is 15 and the maximum number is 40. The median is 23 flies.



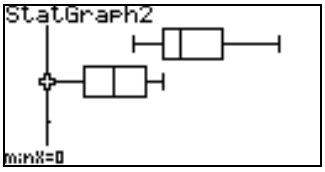


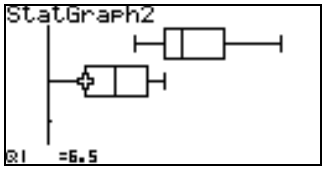




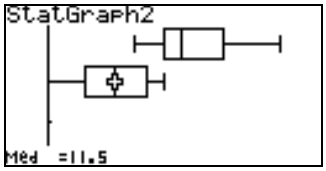


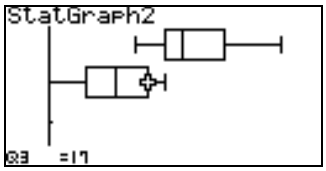
On the other hand, for the data on the flies on the grape juice, the minimum number is 0 and maximum number is 20. The median is 11.5 flies. Q1 and Q3 signify the quartiles.

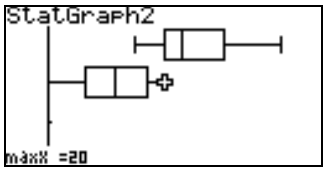




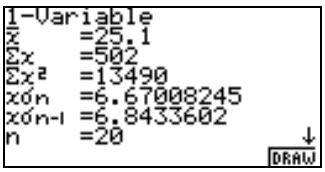
Mean and Standard Deviation



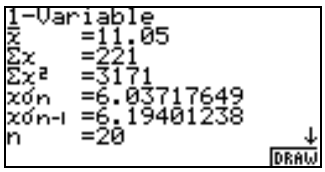




b. Pressing 1Var after a box plot is drawn corresponding to each set of data will display the univariate statistics associated with the given data. We have the following statistics corresponding to both sets of data:



regular food



grape juice

Note that pertaining to the data on the number of flies on regular food, the mean $\bar{x}$ is 25.1 and the standard deviation $x\sigma_{n-1}$ is 6.8433602. On the other hand for the data on the number of flies on grape juice, the mean $\bar{x}$ is 11.05 and the standard deviation $x\sigma_{n-1}$ is 6.19401238.

An observation we can make is that the standard deviations of both sets of data differ by only about 0.65. A graphical perspective can also help us make this observation. As we compare both box plots, we note that the "spread" of the data is almost the same. The measure of the dispersion of both sets of data is not so much different.

Note also that the mean number of flies for the regular food is higher than that for the flies for the grape juice. This observation is also consistent with the box plots we have obtained.

Example 2. Robynne surveyed the prices for a quart of a certain brand of motor oil. The sample data, in dollars per quart, are summarized below:

PRICE PER QUART	0.99	1.09	1.19	1.29	1.39	1.49
NUMBER OF STORES	2	3	7	10	14	4

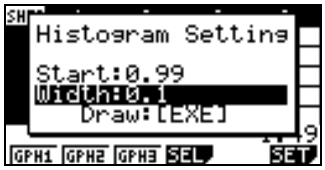
a. Represent the data graphically.
b. What is the mean and standard deviation of the prices?

Solution:

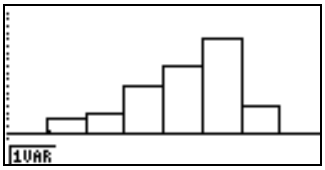
a. In this example, our data include information on price and frequency pertaining to the number of stores carrying the given price. A histogram is a graph that can provide a visual summary for data that include frequency. The prices per quart will be marked on the horizontal axis and the frequencies on the vertical axis.

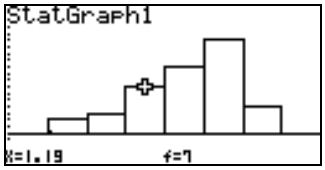
Mean and Standard Deviation

In the Spreadsheet Editor, we enter the prices in column *A* and the corresponding frequencies in column *B*. To graph the data, we select "histogram"[Hist]. One also specifies where to start the histogram and how wide to make each interval for the data:

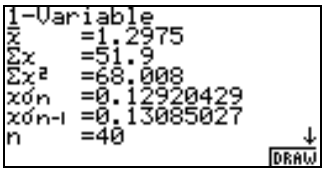


As we trace through the histogram from left to right, for each interval, the calculator shows the left endpoint as x and the interval frequency as f at the bottom of the screen:





b. The mean of the prices $\bar{x}$ is approximately \$1.2975 and the standard deviation of the prices $x\sigma_{n-1}$ is approximately 0.13085027. The low standard deviation tells us that the pricing is not so spread out and is cluttered close to the mean. We have the following screen dump:



Example 3. In order to study the composition of families in Winslow, Arizona, 40 randomly selected married couples were surveyed to determine the number of children per family. The following results were obtained:

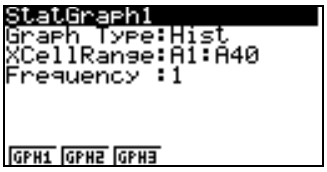
3	1	0	4	1	3	2	2	0	2	0	2	2	1
4	3	1	1	3	4	2	1	3	0	1	0	2	5
1	2	3	0	0	1	2	3	1	2	0	2		

a. Construct a histogram to represent the data.
b. Find the mean number of children per family.
c. Find the standard deviation of the number of children per family.

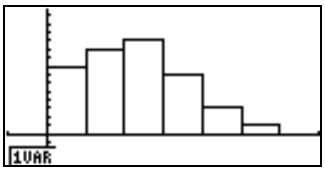
Solution:

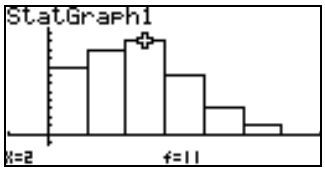
a. Once the data is entered, the calculator can provide a summary of the given data and

tabulate the frequency a number appears. The 1Var frequency is set to 1, as each data represents a single family's number of children. We adjust the histogram settings so that the width of each bar is 1.



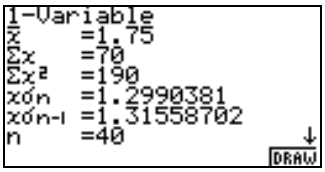
The histogram is given as follows:





b. The mean number of children per family is approximately 1.75 or about 2 per family.
c. The standard deviation of the number of children per family is about 1.32.

We have the following screen dump:



Remarks: The answers to exercises 1 to 3 above can also be obtained by entering the given data in the STAT menu. The graphs and calculations are obtained using similar commands.

EXERCISES:

1. Given the following data:

NORMAL MONTHLY RAINFALL IN SEATTLE, WASHINGTON											
JAN	FEB	MAR	APR	MAY	JUNE	JULY	AUG	SEPT	OCT	NOV	DEC
5.4	4.0	3.8	2.5	1.8	1.6	0.9	1.2	1.9	3.3	5.7	6.0

NORMAL MONTHLY RAINFALL IN PHEONIX, ARIZONA											
JAN	FEB	MAR	APR	MAY	JUNE	JULY	AUG	SEPT	OCT	NOV	DEC
0.7	0.7	0.9	0.2	0.1	0.1	0.8	1.0	0.9	0.7	0.7	1.0

a. Make a visual comparison of the two distributions with respect to their centers and spreads.
b. Calculate the mean and standard deviation for each data set.

2. The following data are the results of an exam given in a statistics class:

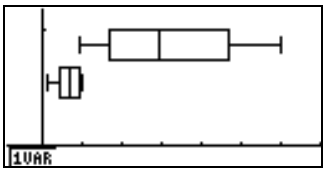
TEST SCORES(IN %)	99	74	85	93	80	70
NUMBER OF STUDENTS	2	3	7	10	14	4

What is the mean and standard deviation of the scores?

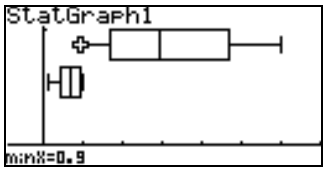
SOLUTIONS:

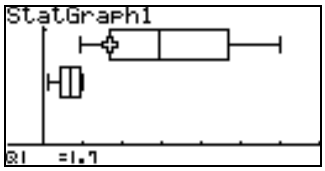
Exercise 1.

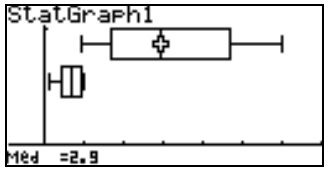
a. The box plots are given in one screen below. Choose SEL to allow graphs to be shown simultaneously on the screen:

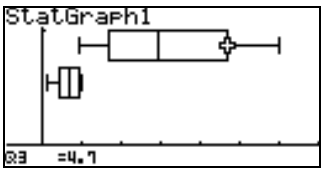


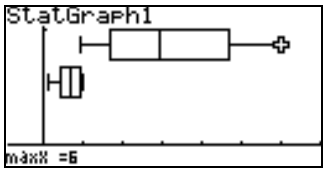
The box plot at the top corresponds to the rainfall in Seattle. The minimum and maximum values are 0.9 and 6 respectively, with median 2.9.



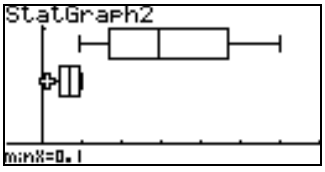


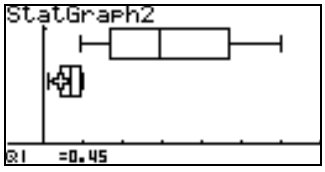




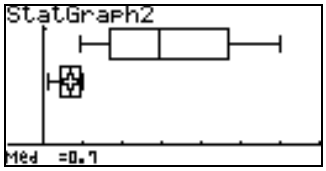


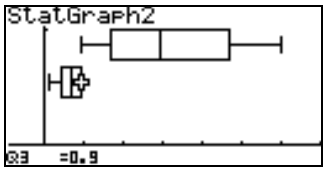
On the other hand, the box plot at the bottom corresponds to the rainfall in Arizona. Note that the minimum and maximum values are smaller in comparison, 0.1 and 1 respectively, with median 0.7.

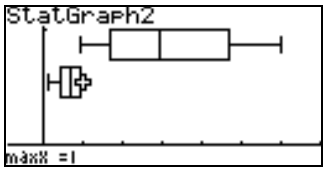




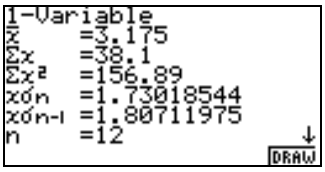
Mean and Standard Deviation



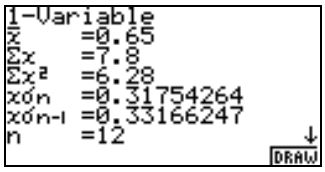




b. Note that pertaining to the data on rainfall in Seattle, the mean $\bar{x}$ is 3.175 and the standard deviation $x\sigma_{n-1}$ is 1.80711975. On the other hand, for the data on rainfall in Arizona, the mean $\bar{x}$ is 0.65 and the standard deviation $x\sigma_{n-1}$ is 0.33. The rainfall in Seattle is more spread out, it has a higher standard deviation. These results are confirmed by the appearances of their box plots.



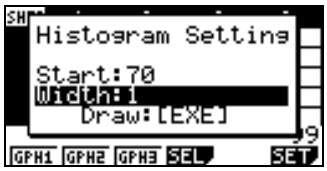
Seattle



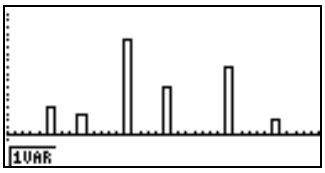
Arizona

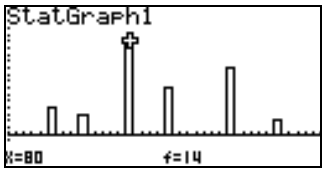
Exercise 2.

a. We have the following settings to draw the histogram:

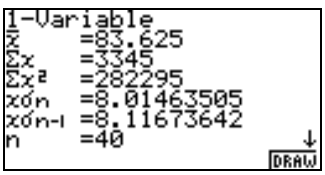


The histogram is given indicating the frequencies of the test scores:





b. The mean of the test scores is approximately 83.625, while the standard deviation is approximately 8.11673642.



REFERENCE

[1] Johnson, et al. *Statistics*, *Principles and Methods*, 3rd Edition, John Wiley and Sons, 1996.

