

# Average and Standard Deviation

## \$1. Introduction

Is it enough for you to calculate the <mean> and <standard deviation> of a data set in statistical data analysis?

It seems to me these are not sufficient to understand the characteristics of a data set.

For example, suppose Company <A> has 50 employees and Company <B> has 20 employees.

The bank deposits of the employees are as follows;

### Company A:

Size of Bank Account	No. of employees
\$10000	2 persons
\$20000	4 persons
\$30000	6 persons
\$40000	8 persons
\$50000	10 persons
\$60000	8 persons
\$70000	6 persons
\$80000	4 persons
\$90000	2 persons

### Company B:

Size of Bank Account	No. of employees
\$50000	16 persons
\$94721	2 persons
\$5279	2 persons

Both data sets have \$50 k and \$400 k for the mean and the variance (denominator = N), respectively.

Do you think that the employees of both companies have the same characteristics in the pattern of their deposits including the <Confidence Interval>?

## \$2. Usage

### 1. Execute the Microsoft Excel.

(This application software requires the Microsoft Excel.)

### 2. Input data in the cells of the Microsoft Excel.

Even if data are randomly inputted in the cells as follows, it is OK.

	A	B	C	D	E
1	1	2			
2	2.5		3		
3		3.3	3.6	3.9	
4	4.1	4.3	4.5		
5		4.7		4.9	
6	5	5.1	5.2	5.3	
7	5.5		5.9	6	
8		6.5	6.9		
9	7.2	7.7		8.1	
10	9				
11					

### 3. Press the left mouse button at the upper-left cell and drag the mouse to the lower-right cell, by holding the mouse button down.

An example is as follows;

	A	B	C	D
1	1	2		
2	2.5		3	
3		3.3	3.6	3.9
4	4.1	4.3	4.5	
5		4.7		4.9
6	5	5.1	5.2	5.3
7	5.5		5.9	6
8		6.5	6.9	
9	7.2	7.7		8.1
10	9			

4. Select the Excel menu <Edit> → <Copy>, for copying the data to the Clipboard.

Note: In the case of saving the inputted data as an Excel file, first save it at #2 step and then copy it to the Clipboard after #3 step.

5. Close the Microsoft Excel.

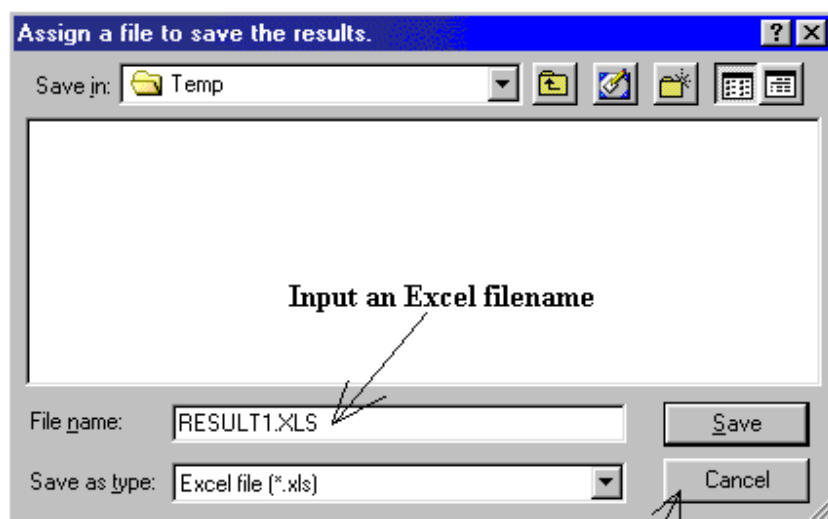
6. Then, execute this application software.

7. Press <Set Conditions> button (  ) and set the analysis conditions, if necessary.

8. Press <Analysis> button (  ).

9. A dialog box which prompts to input a filename for saving the results appears on the screen, if no error occurred.

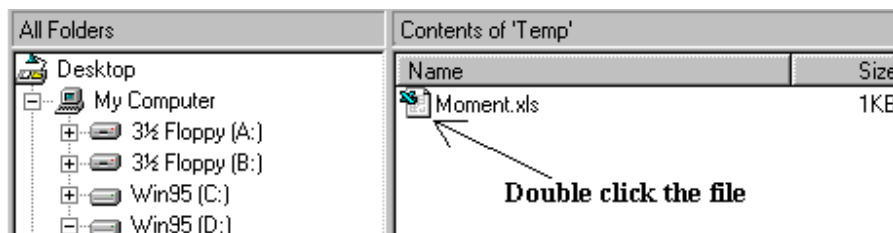
Input an Excel filename (xxxx.XLS) to save the results, and then press <Save> button.



Even pressing <Cancel>, the calculation is executed. So, you can get a histogram.

10. After saving the results to the assigned Excel file, a success message appears.

11. Look for the assigned Excel file by using a filer (e.g. Explorer), and double click the file. Then, you can read the results by the Excel.



#### Note:

Sometimes, a part of the Excel table has "#####."

In this case, please carry out the following procedures;

- (1) Drag the mouse from the upper-left corner to the lower-right corner of the whole data area so as the background color being inverted.
- (2) Select Excel menu <Format> → <Cells...>, and select <Number> tab of the dialog.
- (3) Then, select "General" in the selection box of the dialog, and press <OK> button.

### \$3. Data input to the Excel cells

1. Confirm no data in the Excel cells. Then, press the left mouse button at the upper-left cell and drag the mouse to the lower-right cell, by holding the mouse button down, corresponding to the whole cells in which data are inputted.

2. Select the Excel menu <Format> → <Cells...> → sheet <Number> of a dialog appeared.

3. Select one of the following two processes.

- (1) In the case of the input by standard numerical value format ;

Select "General" in a selection box of the dialog, and press <OK> button.

In this case, only input by standard numerical value format ( e.g. 10, -3, 14.32, -0.03333, etc. ) is allowed at the following #4 step.

Allowed characters for inputting are as follows;

0 1 2 3 4 5 6 7 8 9 . -

- (2) In the case of the input by several numerical value formats ;

Select "Text" in a selection box of the dialog, and press <OK> button.

In this case, input by several numerical value formats ( e.g. 12, -3.5, 1.3E-2, -2/3, 3/7E2, 3E-2/7, etc. (\*1) ) is allowed at the following #4 step.

(\*1) : 3/7E-2 is same as 3/(7E-2), and 3E-2/7 is same as (3E-2)/7.

Allowed characters for inputting are as follows;

0 1 2 3 4 5 6 7 8 9 . - E /

An example is as follows;

	A	B	C	D
1	1/3	2		
2	2.5		3.5E2/9.1	
3		3.3	3.6	3.9
4	4.1	4.3E-1	4.5	
5		4.7		4.9

4. Then, start data input to the Excel cells.

### \$4. Others

1. Number of bars in graph

According to statistical data analysis, reasonable number of bars in a graph is estimated by the following equation;

$$3.3 \log_{10}(\text{Number of Data}).$$

2. Population

It is assumed that the Population has a Gaussian Distribution.

3. Copyright and Warranties

Copyright holder, Kimio Kanda, does not have any responsibility for damage, caused by the use of copyright holder's software, to software of user(s), if it happens.

## \$5. Supplement

If

“m” is defined as “mean (unbiased estimator),”

“V<sup>2</sup>” is defined as “variance (unbiased estimator),” and

“s” is defined as “standard deviation (unbiased estimator),”

we can get the following equations.

$$m = \frac{\sum_{i=1}^N X_i}{N}, \quad V^2 = \frac{\sum_{i=1}^N (X_i - m)^2}{N-1}, \quad s = \sqrt{\frac{N-1}{2}} \frac{\Gamma\left(\frac{N-1}{2}\right)}{\Gamma\left(\frac{N}{2}\right)} \sqrt{V^2},$$

where,  $\Gamma(x)$  is the Gamma function.

© Kimio Kanda