

Summary (descriptive) statistics

EXAMPLE 3.2. Length (mm) of two-spot ladybirds (*Adalia bipunctata*)

BOX 3.1. How to calculate a standard deviation and variance for normally distributed (parametric) data

In this section we explain how you can use Excel to work out terms such as the **mean**, **median**, **mode**, and **variance**. Our explanation is short, so unlike other chapters on this web site we have not followed the section headings that are in the book. Read through the complete explanation and you will quickly locate the critical step to calculate any one (or more) of these summary statistics. The example we use is Example 3.2. We reproduce this here.

EXAMPLE 3.2. Length (mm) of two-spot ladybirds (*Adalia bipunctata*)

An investigator was interested in the length of two-spot ladybirds (*Adalia bipunctata*). In an observational investigation she measured the length (mm) of 50 ladybirds collected at random from a garden (Table 3.8).

Table 3.8. The length (mm) of 50 *Adalia bipunctata* sampled in a garden

Length of <i>Adalia bipunctata</i> (x)									
1	5	2	5	7	8	3	6	7	4
4	5	6	4	5	5	7	5	3	5
4	5	1	7	9	2	6	5	6	3
3	6	8	6	4	6	6	8	5	6
7	4	8	9	5	4	3	4	2	5

Step 1. Put the data into a worksheet. The data must be in a rectangular array, and this can most easily be achieved by placing it all in a single column. (If you have a prime number of measurements (e.g. 19), you cannot get a rectangular array except by using a single row or a single column.)

The screenshot shows an Excel window titled "Microsoft Excel - Descriptive Statistics". The menu bar includes File, Edit, View, Insert, Format, Tools, Data, Window, and Help. The toolbar shows various icons for file operations and editing. The active cell is A1, containing the text "Lengths of 2-spot ladybirds (mm)". The data is organized in a 10-column grid (A-J) and 9 rows (1-9). Row 1 is the header, and rows 3-7 contain numerical data.

	A	B	C	D	E	F	G	H	I	J
1	Lengths of 2-spot ladybirds (mm)									
2										
3	1	5	2	5	7	8	3	6	7	4
4	4	5	6	4	5	5	7	5	3	5
5	4	5	1	7	9	2	6	5	6	3
6	3	6	8	6	4	6	6	8	5	6
7	7	4	8	9	5	4	3	4	2	5
8										
9										

Step 2. You can either calculate what you need individually (step 2), use the fx paste function (step 3) or use an Excel function that gives you the whole lot in one go (step 4).

To calculate each statistic individually, you need to allocate a cell in which the statistic is to be displayed, and it is good practice to label it so that you don't forget what is in it (and anyone else looking at your work will see immediately what you have been doing). For example for the mean, type '=average(a3:j7)' into the cell where you want the mean to appear. (In this example, the data is in the rectangular array with top left cell a3, and bottom right cell j7: replace these with cell locations to correspond to the location of your data.)

Kurtosis =kurt(cellrange)

Skewness =skew(cellrange)

Smallest value =min(cellrange)

Highest value =max(cellrange)

Range =max(cellrange) – min(cellrange)

Sum of all the measurements (Σx)
=sum(cellrange)

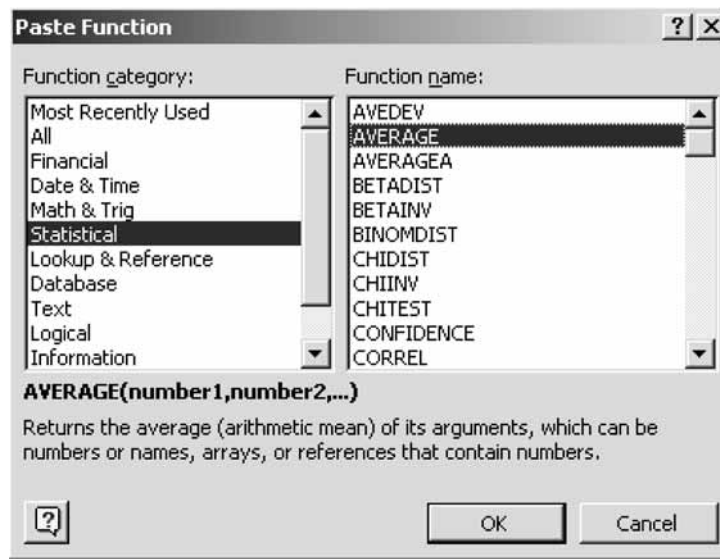
Number of observations (n) =**count(cellrange)**

[Note: If you want population parameters rather than sample statistics, use ‘=varp(cellrange)’ and ‘=stdevp(cellrange)’, where cellrange is the range of cells whose contents you wish to use.]

The screenshot shows a Microsoft Excel window titled 'Microsoft Excel - Descriptive Statistics'. The spreadsheet contains data for 'Lengths of 2-spot ladybirds (mm)' in columns A through J, rows 3 through 7. Below the data, a summary table provides statistical measures:

	A	B	C	D	E	F	G	H	I	J
1	Lengths of 2-spot ladybirds (mm)									
2										
3	1	5	2	5	7	8	3	6	7	4
4	4	5	6	4	5	5	7	5	3	5
5	4	5	1	7	9	2	6	5	6	3
6	3	6	8	6	4	6	6	8	5	6
7	7	4	8	9	5	4	3	4	2	5
8										
9	mean		5.08							
10	median		5							
11	mode		5							
12	variance		3.748571							
13	standard deviation		1.936123							

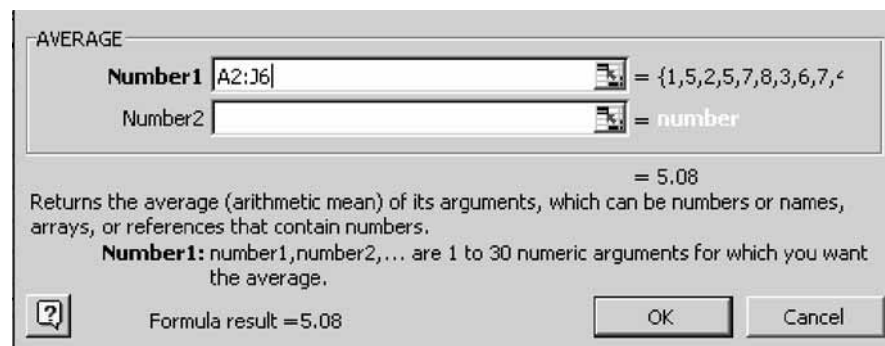
Step 3. By using fx paste function, the above procedure can be speeded up. This feature is found on the top tool bar. Select a cell where you want the result to be returned. Click on fx paste function and a box will open. From ‘Function Category’, select ‘Statistical’ and from ‘Function Name’, select ‘Average’. Note that a description of the function is given below the selection area. Click on OK.



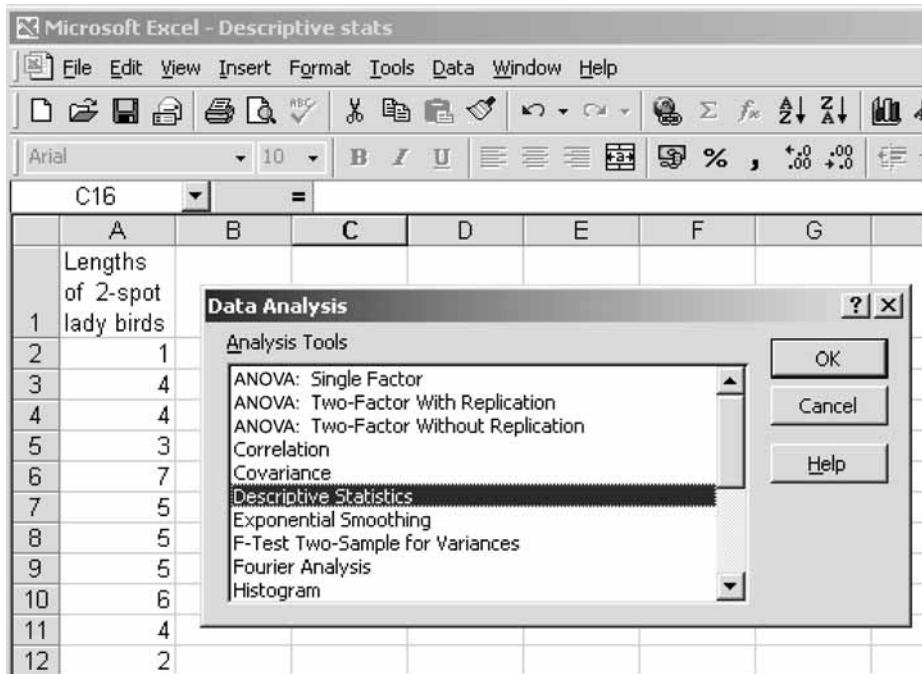
A new box will open.

The box adjacent to 'Number 1' will be highlighted black. To enter the cell locations of the data, click on the first cell with numerical data (A2) and drag across the entire data set. The cell locations will be entered into the box. Note that the actual values are listed along side of the box and the answer is given beneath. Click on OK.

The answer will be returned in the selected cell.



Step 4. To get all these statistics together you **MUST** have your data in a single row or column. Put a label in the top cell and then list the data immediately under this. From the top tool bar, select 'Tools', 'Data Analysis' and 'Descriptive Statistics'.



Click on 'OK'.

A dialogue box will open into which you enter the input data range. To do this, ensure that the Input Range box is highlighted, then click on the first cell (cell A1) and holding down the mouse button drag down the entire data set. This will become highlighted and note that the cell locations are now entered into the box. As the first cell contains a data label you must click in the box marked 'Labels' and a tick will appear. It is helpful to include a label as the output data will then have the appropriate label. This assists with subsequent use of the results. To include a label, the text must always be written in the cell immediately above the first datum. Excel then knows to exclude the first as it is a label only (and probably not numeric).

Select 'Output Range', and click in the box along side. To enter the location where you want the output to be located, click on a cell and this will indicate the top left hand corner of the output box. Excel will determine the size of the box required. Select 'Summary Statistics'.

The screenshot shows the Microsoft Excel interface with the 'Descriptive Statistics' dialog box open. The spreadsheet data is as follows:

	A	B	C	D	E	F	G	H
1	Lengths of 2-spot lady birds							
2	1							
3	4							
4	4							
5	3							
6	7							
7	5							
8	5							
9	5							
10	6							
11	4							
12	2							
13	6							
14	1							
15	8							
16	8							
17	5							
18	4							
19	7							
20	6							
21	9							
22	7							

The 'Descriptive Statistics' dialog box is configured as follows:

- Input:**
 - Input Range:
 - Grouped By: Columns Rows
 - Labels in First Row
- Output options:**
 - Output Range:
 - New Worksheet Ply:
 - New Workbook
 - Summary statistics
 - Confidence Level for Mean: %
 - Kth Largest:
 - Kth Smallest:

Buttons: OK, Cancel, Help

Click on 'OK'.

The screenshot shows the Microsoft Excel interface with the 'Descriptive stats' window open. The data is organized in columns A through E. Column A contains the data labels, column B contains the corresponding values, and columns C and D contain the statistical results. The output table is highlighted in grey.

	A	B	C	D	E
1	Lengths of 2-spot lady birds		<i>Lengths of 2-spot lady birds</i>		
2	1				
3	4		Mean	5.08	
4	4		Standard Error	0.27380911	
5	3		Median	5	
6	7		Mode	5	
7	5		Standard Deviation	1.936122782	
8	5		Sample Variance	3.748571429	
9	5		Kurtosis	-0.293549691	
10	6		Skewness	-0.047184265	
11	4		Range	8	
12	2		Minimum	1	
13	6		Maximum	9	
14	1		Sum	254	
15	8		Count	50	
16	8		Confidence Level(95.0%)	0.550239673	
17	5				

The output box is returned, highlighted in colour. Expand the cells by clicking and dragging the cells apart on the letter row (line up the cursor with the dividing line between C and D, click and holding down the button drag the cells apart). All the values for the statistics are returned in columns Note that the output table is headed by the data label.