



0101011001010100100100110101011001100110010011010100110
00110010101100101010010010010011001010110011001100100110101
1001001001100101011001010100100100100110011001010110011001

ANALISI STATISTICA DEI DATI

Fabio Scotti

Universita' degli studi
Dipartimento di tecnologie dell'informazione

UNITA' 1:

Strumenti e ambienti informatici per la valutazione di parametri statistici

- ◆ Excel
- ◆ Matlab
- ◆ R



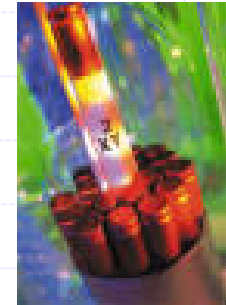
EDITING DATI

- ◆ Esaminiamo la sequenza temporale di temperature
- ◆ Dati in un file TESTO (temperature1.txt)

- ◆ Visualizzazione/Editing
 - Notepad
 - Word (...no)
 - Editor testuali
 - Excel

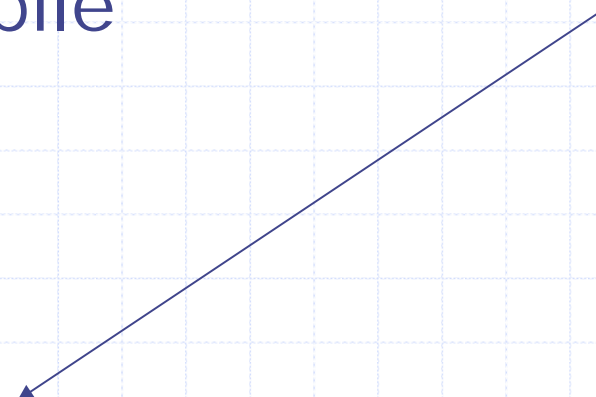


EDITING DATI



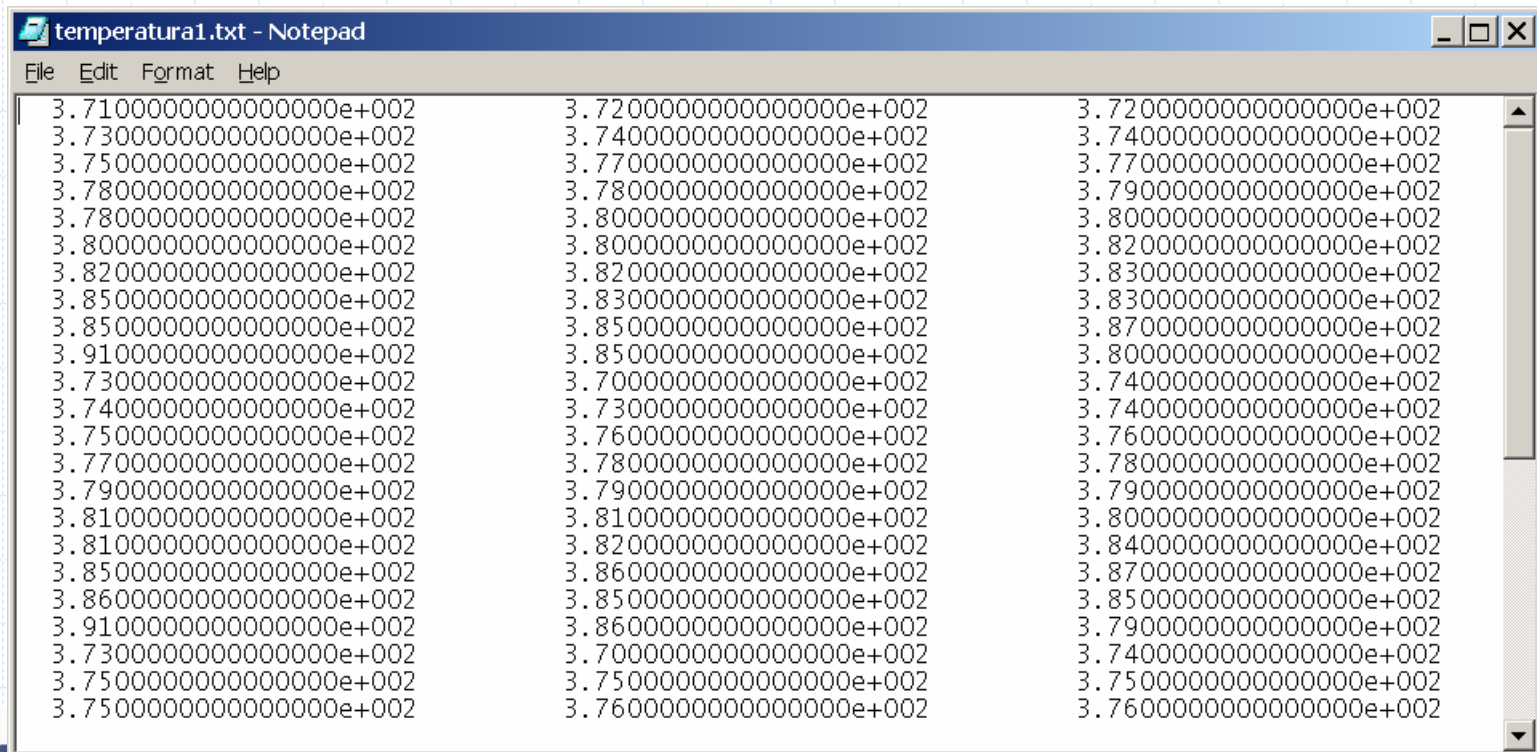
◆ Esempio ad una variabile

```
temperatura1a.txt - Notepad
File Edit Format Help
371
372
372
373
374
374
375
377
377
378
378
379
378
380
380
380
380
382
382
382
382
383
```



EDITING DATI

◆ Esempio ad una variabile (tab separated)



```
temperatura1.txt - Notepad
File Edit Format Help
3.7100000000000000e+002 3.7200000000000000e+002 3.7200000000000000e+002
3.7300000000000000e+002 3.7400000000000000e+002 3.7400000000000000e+002
3.7500000000000000e+002 3.7700000000000000e+002 3.7700000000000000e+002
3.7800000000000000e+002 3.7800000000000000e+002 3.7900000000000000e+002
3.7800000000000000e+002 3.8000000000000000e+002 3.8000000000000000e+002
3.8000000000000000e+002 3.8000000000000000e+002 3.8200000000000000e+002
3.8200000000000000e+002 3.8200000000000000e+002 3.8300000000000000e+002
3.8500000000000000e+002 3.8300000000000000e+002 3.8300000000000000e+002
3.8500000000000000e+002 3.8500000000000000e+002 3.8700000000000000e+002
3.9100000000000000e+002 3.8500000000000000e+002 3.8000000000000000e+002
3.7300000000000000e+002 3.7000000000000000e+002 3.7400000000000000e+002
3.7400000000000000e+002 3.7300000000000000e+002 3.7400000000000000e+002
3.7500000000000000e+002 3.7600000000000000e+002 3.7600000000000000e+002
3.7700000000000000e+002 3.7800000000000000e+002 3.7800000000000000e+002
3.7900000000000000e+002 3.7900000000000000e+002 3.7900000000000000e+002
3.8100000000000000e+002 3.8100000000000000e+002 3.8000000000000000e+002
3.8100000000000000e+002 3.8200000000000000e+002 3.8400000000000000e+002
3.8500000000000000e+002 3.8600000000000000e+002 3.8700000000000000e+002
3.8600000000000000e+002 3.8500000000000000e+002 3.8500000000000000e+002
3.9100000000000000e+002 3.8600000000000000e+002 3.7900000000000000e+002
3.7300000000000000e+002 3.7000000000000000e+002 3.7400000000000000e+002
3.7500000000000000e+002 3.7500000000000000e+002 3.7500000000000000e+002
3.7500000000000000e+002 3.7600000000000000e+002 3.7600000000000000e+002
```



EDITING DATI

- ◆ Esempio ad una variabile (tab separated)

```
0 10 20 30 40 50 60 70 80 90 100
1 3.7100000000000000e+002 3.7200000000000000e+002 3.7200000000000000e+002 3.7300000000
2
```

For Help, press F1 Ln 1, Col. 2, CW UNIX Mod: 28/02/2004 11.15 File Size: 3122 INS



EDITING DATI

- ◆ Esempio ad una variabile (tab separated)

```
UltraEdit-32 - [E:\DATIFabio\MATLABcode\codiceCorsi\temperatura1.txt]
File Edit Search Project View Format Column Macro Advanced Window Help
temperatura1.txt
0 10 20 30 40 50 60 70 80 90 100
1 3.7100000000000000e+002 3.7200000000000000e+002 3.7200000000000000e+002 3.73000000
2
For Help, press F1 Ln 1, Col. 2, CW UNIX Mod: 28/02/2004 11.15 File Size: 3122 INS
```



GESTIONE DATI EXCEL

◆ CARICARE IL FILE (open, all files)

Text Import Wizard - Step 1 of 3 [?] [X]

The Text Wizard has determined that your data is Delimited.
If this is correct, choose Next, or choose the data type that best describes your data.

Original data type

Choose the file type that best describes your data:

Delimited - Characters such as commas or tabs separate each field.

Fixed width - Fields are aligned in columns with spaces between each field.

Start import at row: File origin:

Preview of file E:\DATIFabio\MATLABcode\codiceCorsi\temperatura1a.txt

1	371
2	372
3	372
4	373
5	374
6	374

Cancel < Back Next > Finish



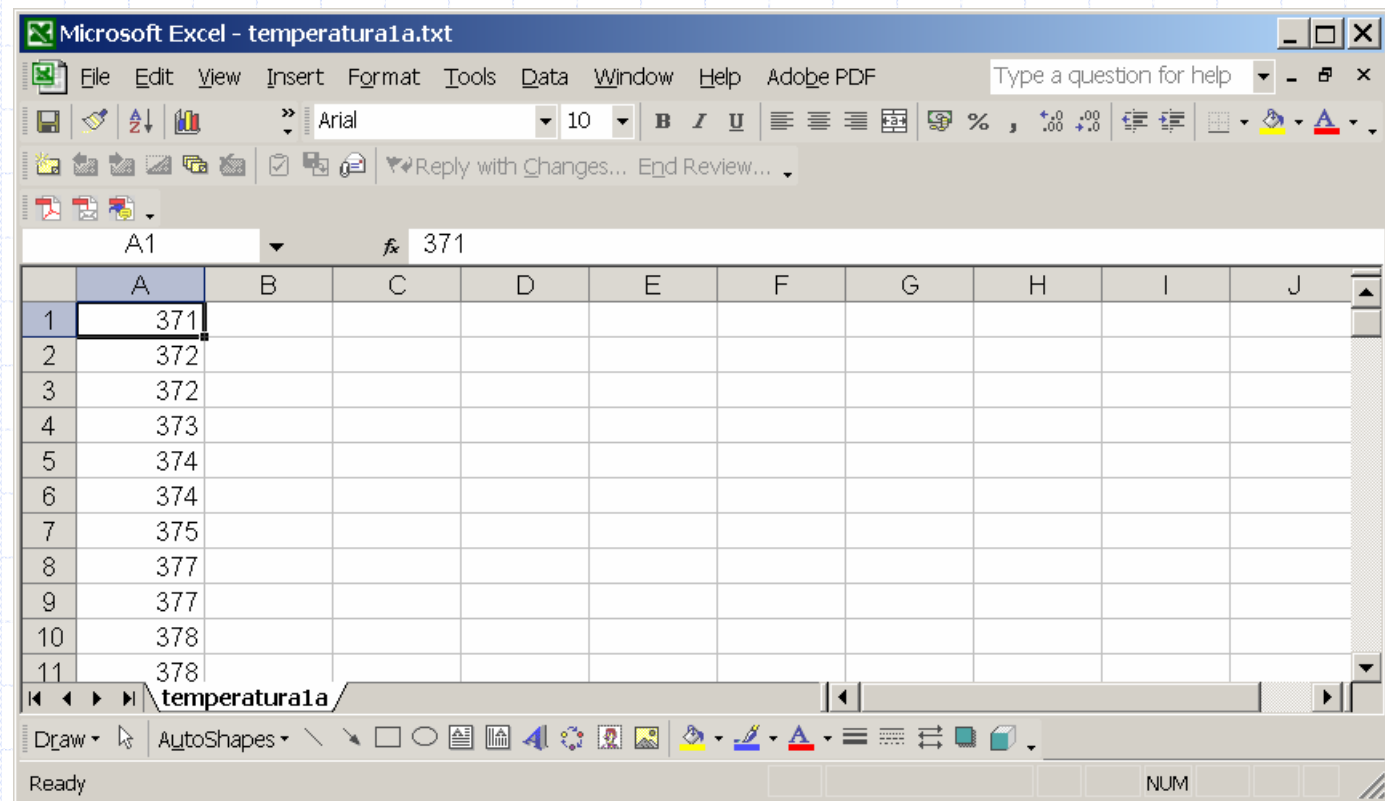


0101011001010100100100110101011001100110010011010100110
0011001010110010101001001001100101010110011001100100110101
1001001001100101011001010100100100110011001010110011001

EXCEL

GESTIONE DATI EXCEL

◆ Abbiamo i dati.



The screenshot shows a Microsoft Excel window titled "Microsoft Excel - temperatura1a.txt". The spreadsheet contains a single column of data in column A, with rows 1 through 11. The values are: 371, 372, 372, 373, 374, 374, 375, 377, 377, 378, 378. The status bar at the bottom indicates "Ready" and "NUM".

	A	B	C	D	E	F	G	H	I	J
1	371									
2	372									
3	372									
4	373									
5	374									
6	374									
7	375									
8	377									
9	377									
10	378									
11	378									



Media e varianza

- ◆ Sequenza temporale di dati
- ◆ Dati da superfici e volumi

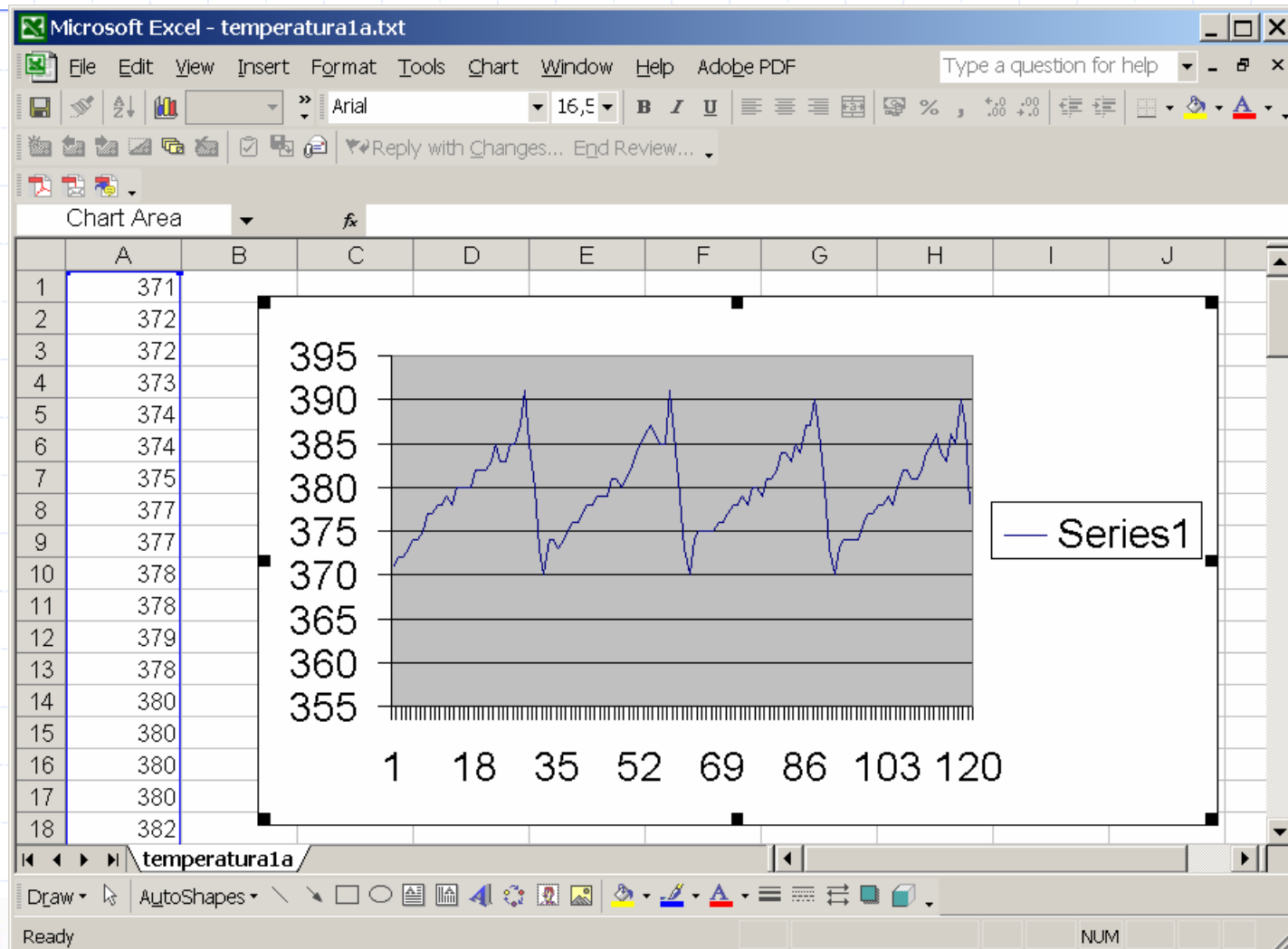
- ◆ ES: Un fermentatore ed un sensore di temperatura.

- ◆ ES: Dati degli esami ematici di un gruppo di laboratori.

- ◆ Es: la temperatura di un corpo misurata con immagini all'infrarosso

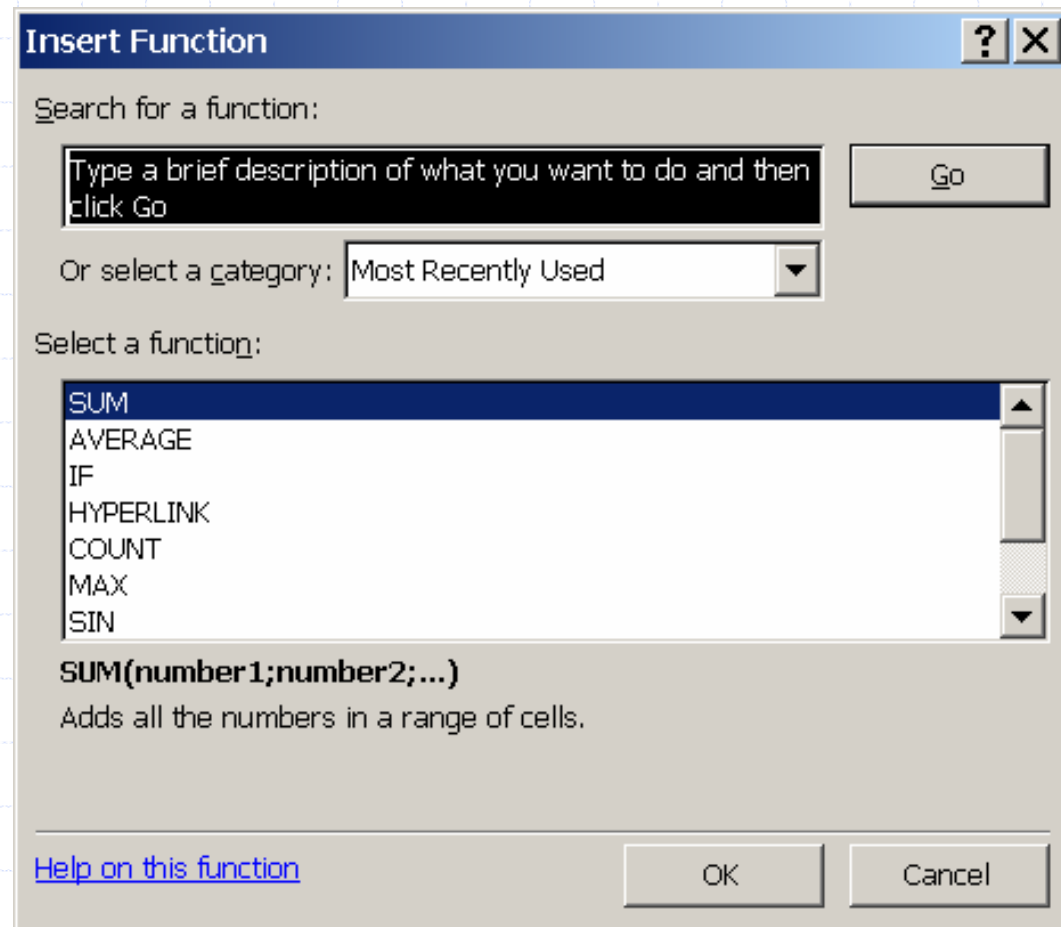


Media e varianza EXCEL



Media e varianza EXCEL

◆ Inseriamo
la funzione



Media e varianza EXCEL

◆ Uppure inseriamo il comando a mano

=AVERAGE(A1:A120)

The screenshot shows a Microsoft Excel window titled "Microsoft Excel - temperatura1a.txt". The spreadsheet contains a column of temperature values from row 1 to 12. Cell C7 contains the formula "=AVERAGE(A1:A120)" and the result "379,725". The word "MEDIA" is written in cell B7. The status bar at the bottom indicates "Ready" and "NUM".

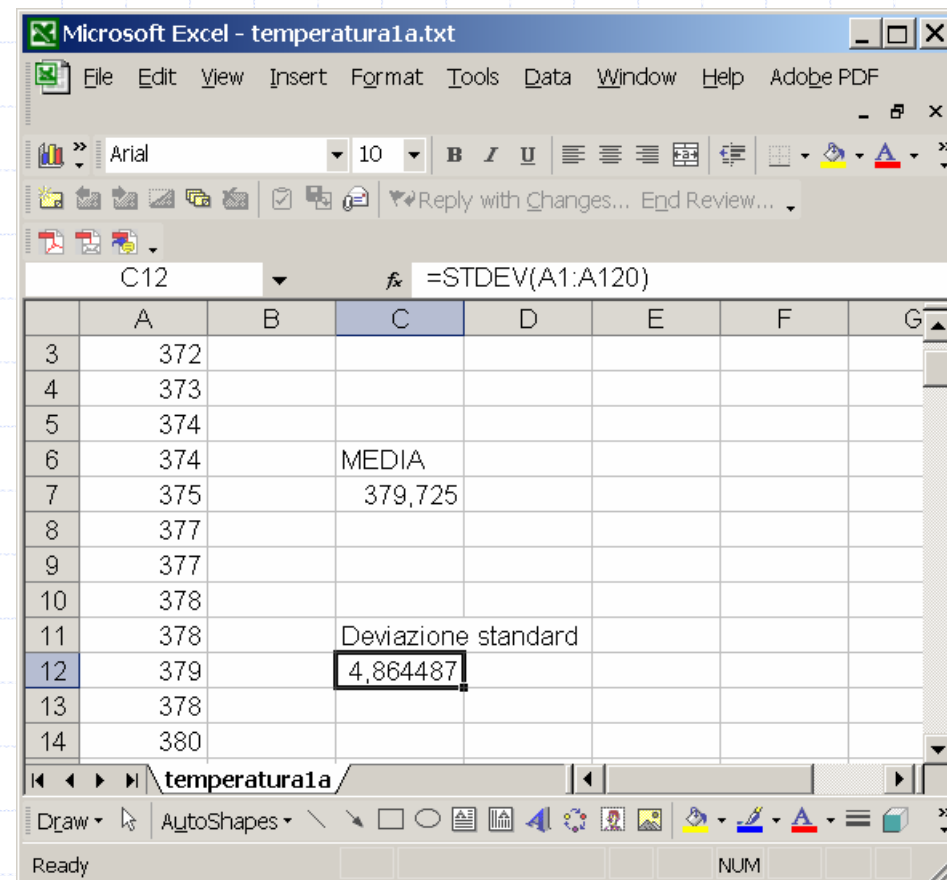
	A	B	C	D	E	F	G	H
1	371							
2	372							
3	372							
4	373							
5	374							
6	374							
7	375	MEDIA	379,725					
8	377							
9	377							
10	378							
11	378							
12	379							



Media e varianza EXCEL

◆ Deviazione standard

=STDEV(A1:A120)



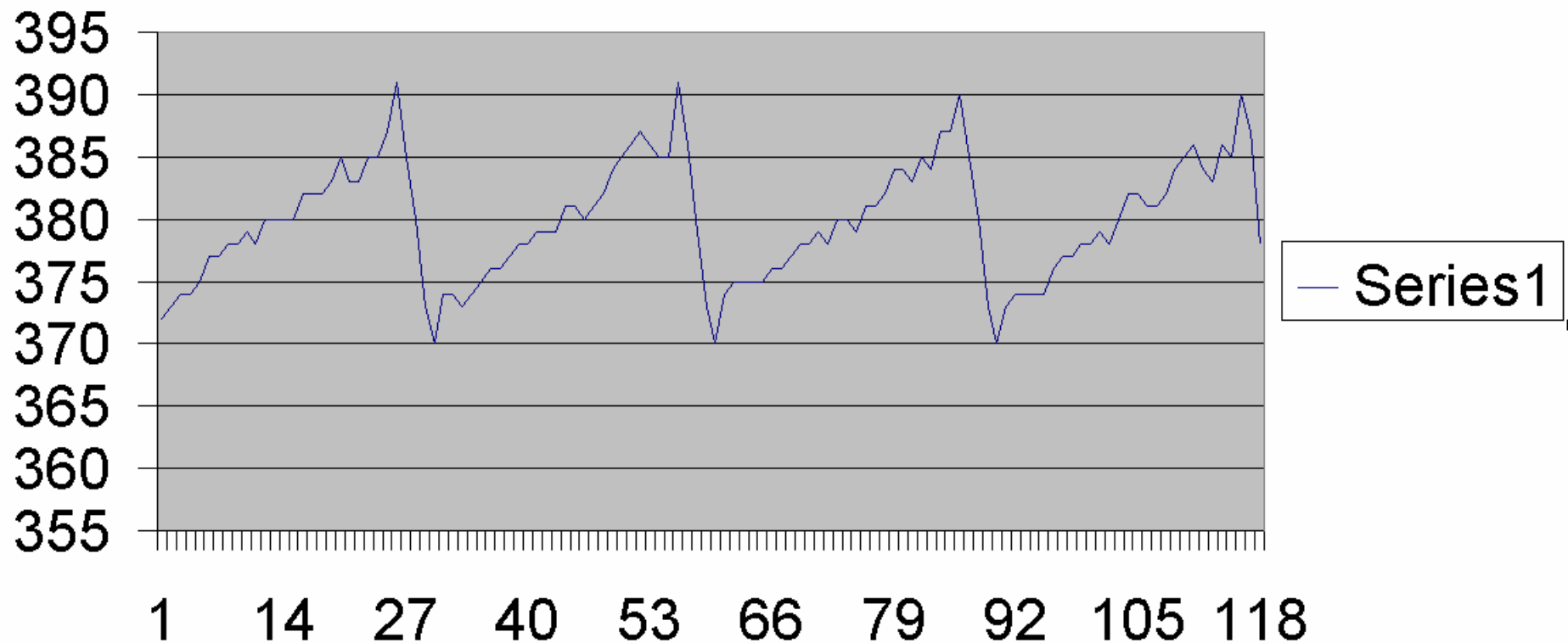
The screenshot shows a Microsoft Excel window titled "Microsoft Excel - temperatura1a.txt". The spreadsheet contains a list of temperatures in column A (rows 3-14) and statistical results in column C (rows 6-12). The formula bar shows the formula =STDEV(A1:A120) in cell C12. The status bar at the bottom indicates "Ready" and "NUM".

	A	B	C	D	E	F	G
3	372						
4	373						
5	374						
6	374		MEDIA				
7	375		379,725				
8	377						
9	377						
10	378						
11	378		Deviazione standard				
12	379		4,864487				
13	378						
14	380						



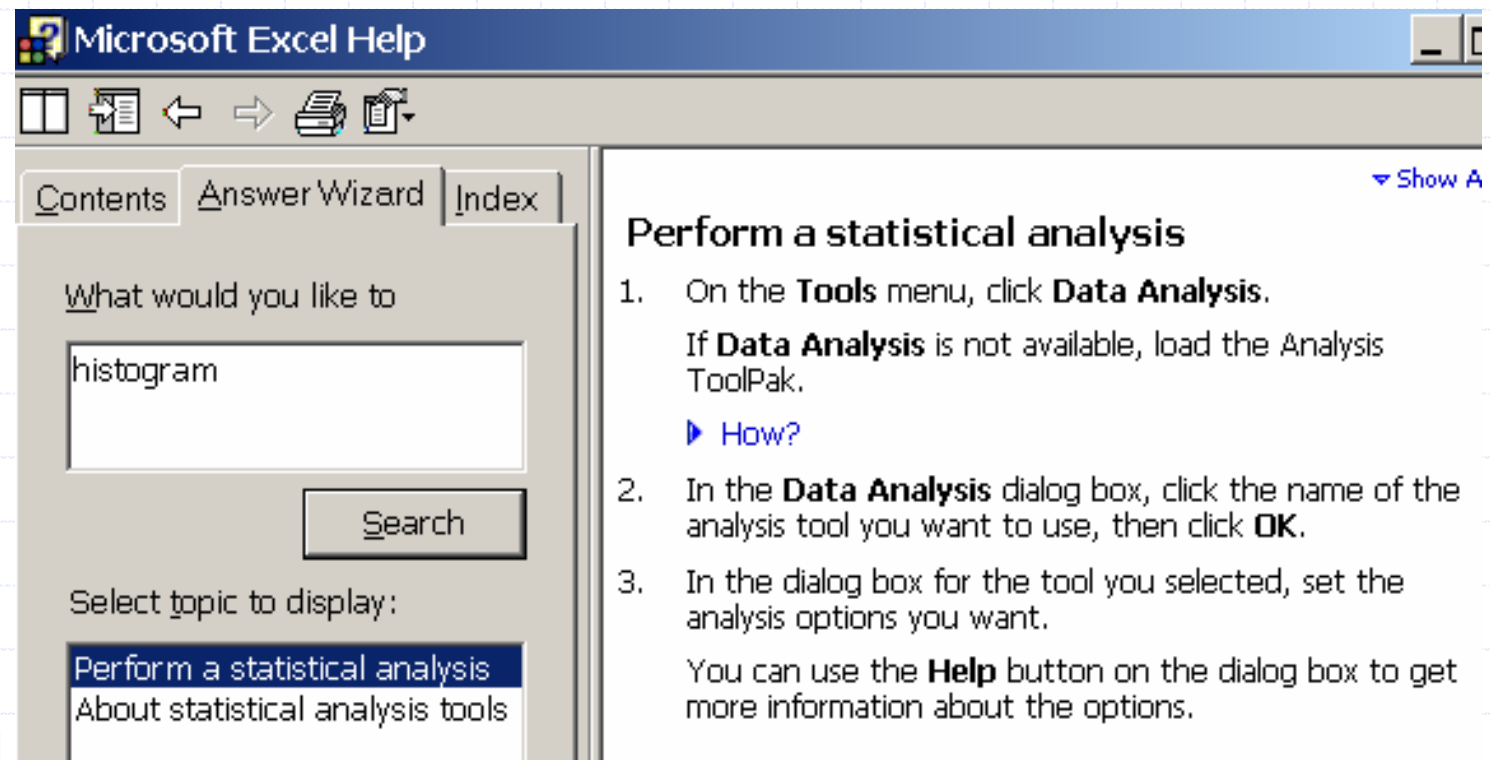
Media e varianza EXCEL

- ◆ Il sistema e' controllato adeguatamente in temperatura?



Media e varianza EXCEL

- ◆ Per rispondere meglio possiamo avvalerci del toolbox statistico di Excel



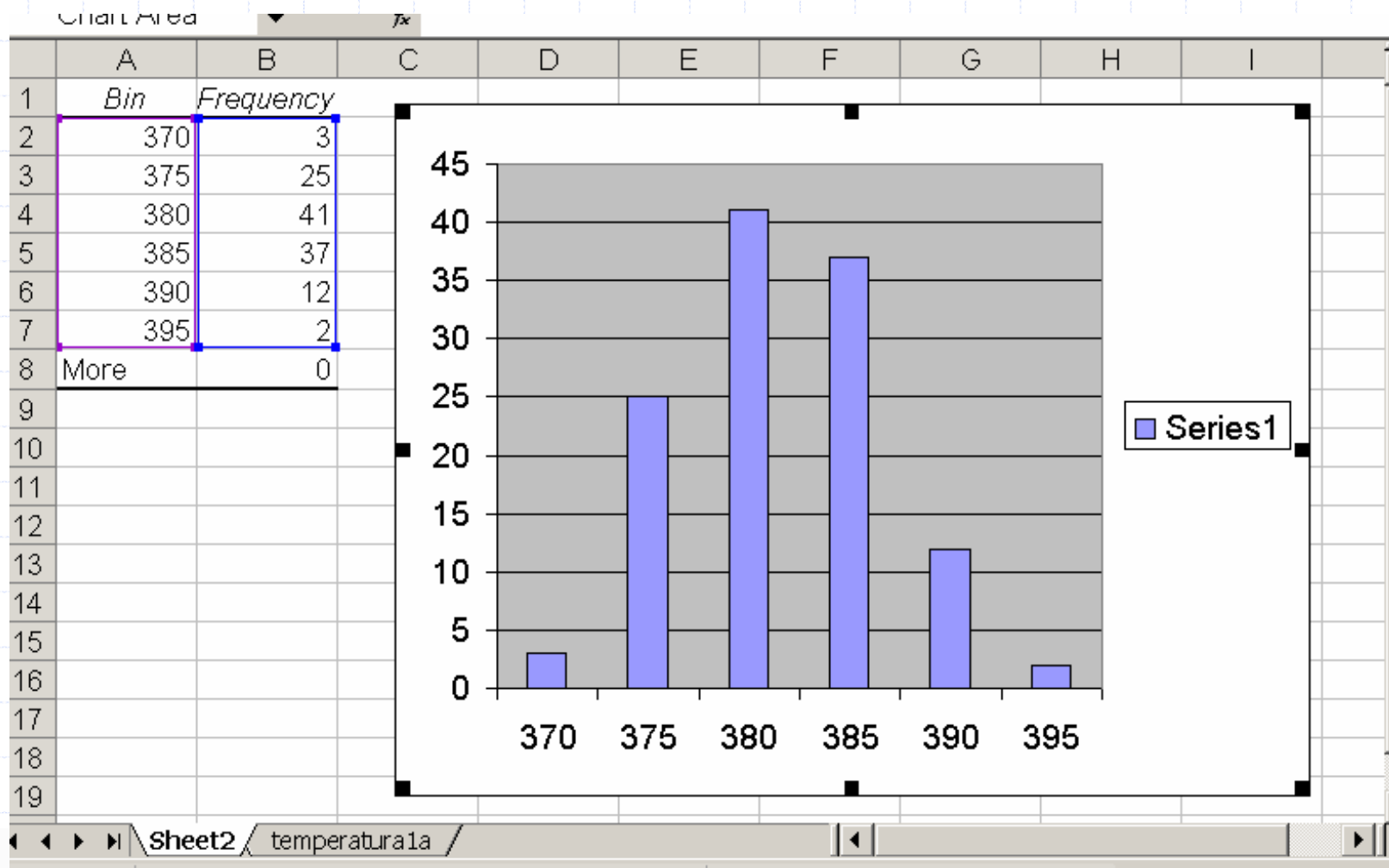
DISTRIBUZIONE con EXCEL

- ◆ Per rispondere meglio possiamo avvalerci del toolbox statistico di Excel
(Tools → Addins → Selezionare Analysis tool)
- ◆ Inserite il toolPAK
- ◆ Preparate i "bin" dell'istogramma
- ◆ Selezionate istogramma

78		
78		Deviazione standard
79		4,864487
78		
30		
30		BIN
30		370
30		375
32		380
32		385
32		390
33		395
35		



DISTRIBUZIONE con EXCEL

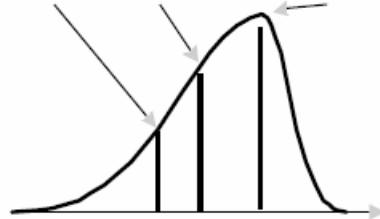


Valutazione Della forma della distribuzione con EXCEL

1. Describes How Data Are Distributed
2. Measures of Shape
 - Skew = Symmetry

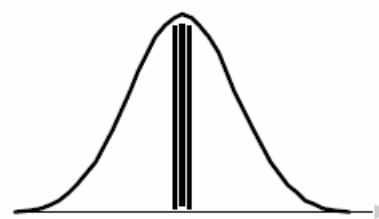
Left-Skewed

Mean Median Mode



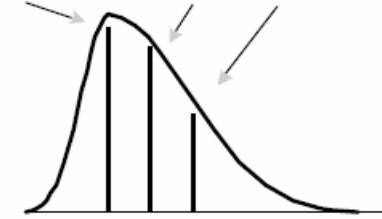
Symmetric

Mean = Median = Mode

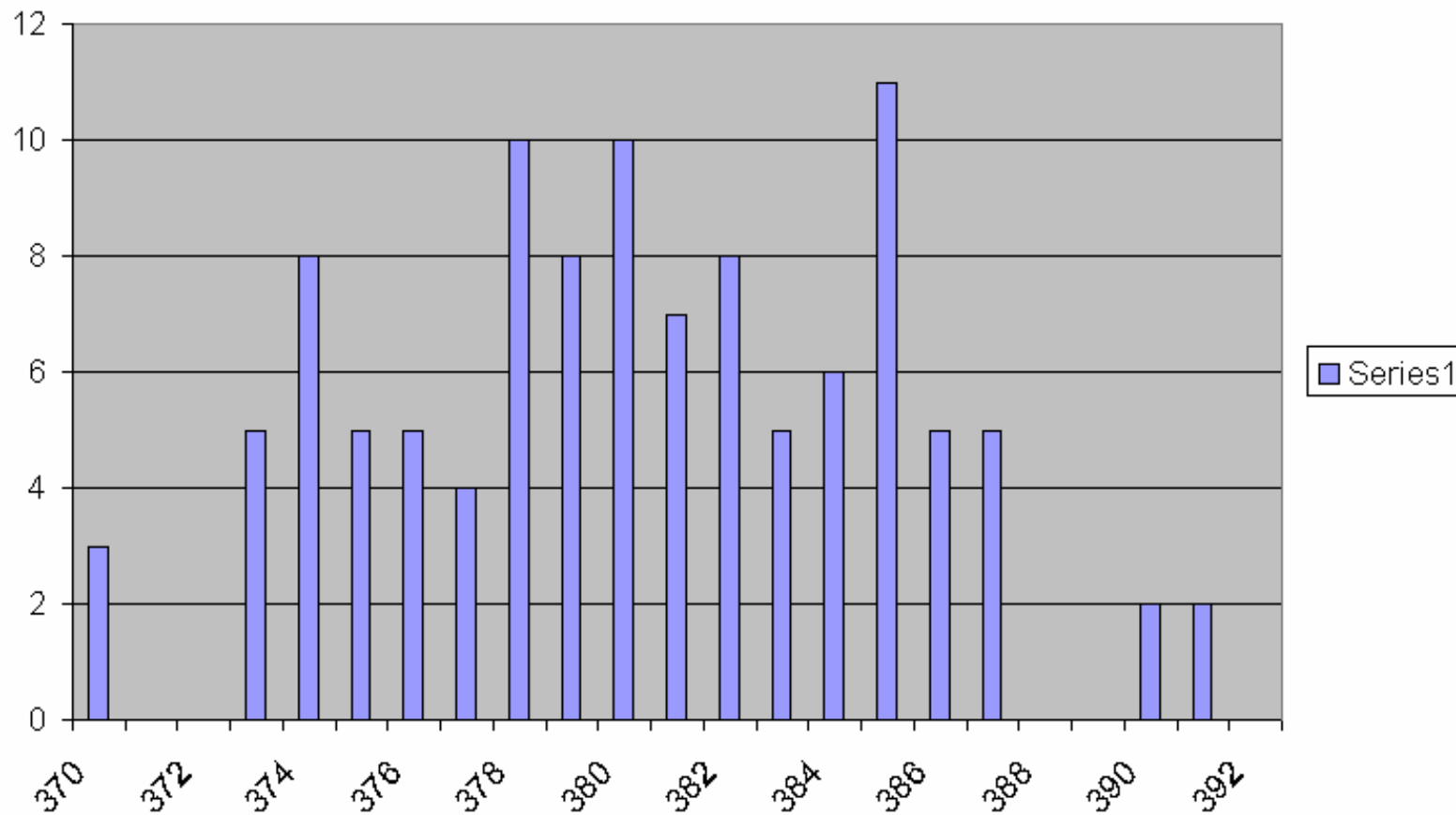


Right-Skewed

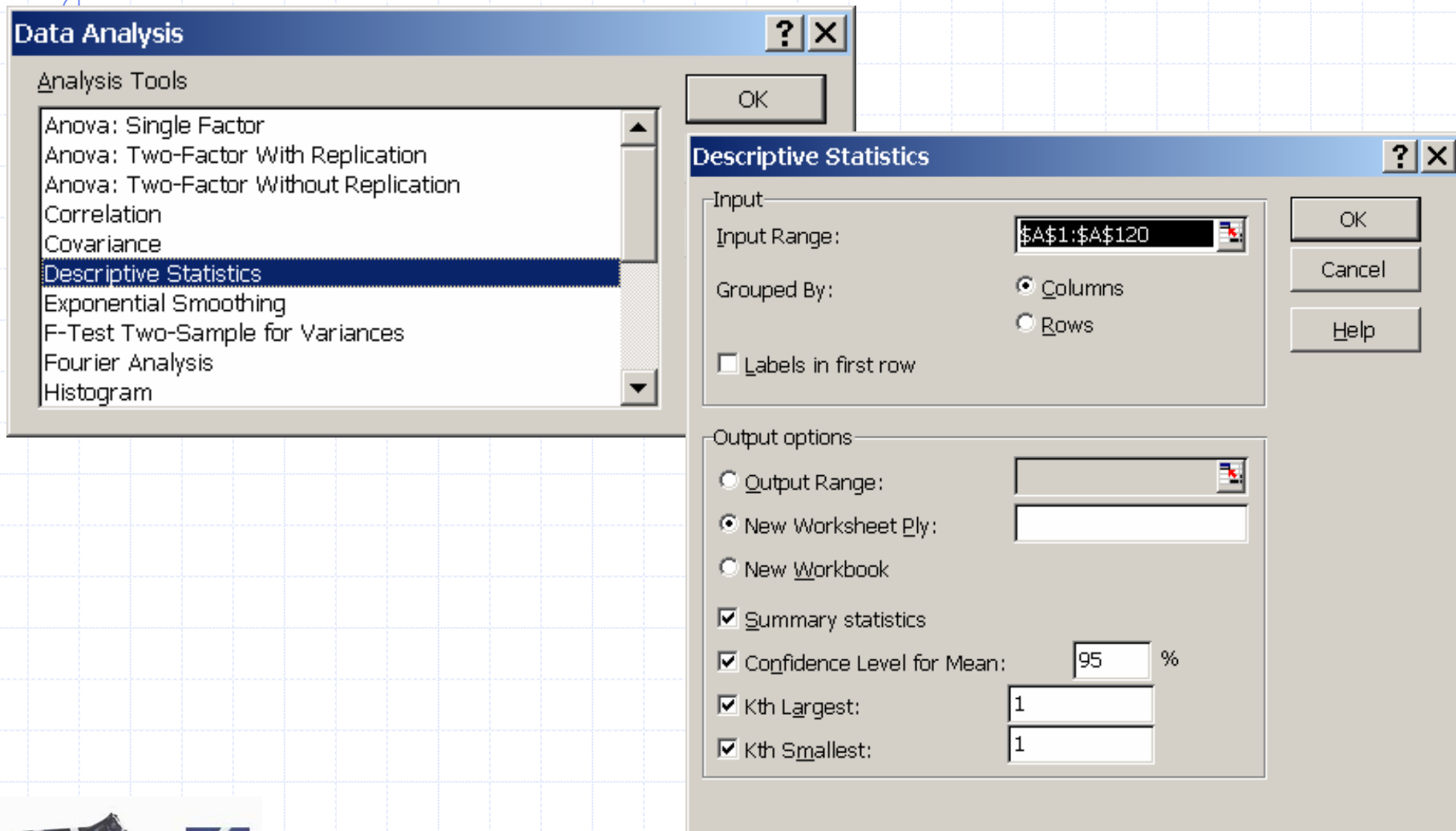
Mode Median Mean



Valutazione Della forma della distribuzione con EXCEL



RIASSUNTO STATISTICO con EXCEL



The image shows two overlapping dialog boxes from Microsoft Excel. The 'Data Analysis' dialog box is on the left, with 'Descriptive Statistics' selected in the 'Analysis Tools' list. The 'Descriptive Statistics' dialog box is on the right, showing the following settings:

- Input Range:** \$A\$1:\$A\$120
- Grouped By:** Columns (selected)
- Labels in first row
- Output options:**
 - Output Range:
 - New Worksheet Ply:
 - New Workbook
 - Summary statistics
 - Confidence Level for Mean: 95 %
 - Kth Largest: 1
 - Kth Smallest: 1



Risorse in rete

(<http://www.mathtools.net/MATLAB/Statistics/Excel/>)

Mathtools.net >

The technical computing portal for all your scientific and engineering needs

Mathtools.net > MATLAB > Statistics > Excel

- [SolverStat](#) - SolverStat is an Excel add-in (xla) for Microsoft Excel 97, 2000, XP. The built-in optimization package Solver yields minimized least squares worksheet functions makes Excel a flexible tool to solve linear and non linear regression.
- [Analyse-it, a statistics add-on](#) - Analyse-It for Microsoft Excel now available!! The parametric, and clinical research statistics. Download FREE 14-day trial of Analyse-It.
- [Blockage Analysis](#) - MaceTech Blockage Analysis is a freeware Excel Add-in. Blockage analysis is a common problem amongst options on account of unawareness, lack of knowledge, unavailability, in both positive and negative impacts of blockage, and, the separate impacts of multiple blockages.
- [CAMPUS](#) - This paper describes the development of a spreadsheet-based CAA program for quantitative disciplines. The CAMPUS program (Computer Aided Marking Program) provides a tool for marking and feedback when assessing students.
- [DE Histograms](#) - This Excel Add-In provides fast and easy graphical and statistical analysis of data.





0101011001010100100100110101011001100110010011010100110
00110010101100101010010010010011001010110011001100100110101
1001001001100101011001010100100100110011001010110011001

MATLAB

Statistica descrittiva MATLAB

```
data = [1 2 3 4 50];
```

```
% The arithmetic mean of the data:  
mean(data) ans: 12
```

```
%The median of the data:  
median(data) ans: 3
```

```
% The standard deviation:  
std(data) ans: 21.27
```

```
% The smallest value in the data:  
min(data) ans: 1
```

```
% The largest value in the data:  
max(data) ans: 50
```



RIASSUNTO STATISTICO con EXCEL

	A	B
1	<i>Column1</i>	
2		
3	Mean	379,725
4	Standard Error	0,444065
5	Median	380
6	Mode	378
7	Standard Deviation	4,864487
8	Sample Variance	23,66324
9	Kurtosis	-0,63292
10	Skewness	0,106953
11	Range	21
12	Minimum	370
13	Maximum	391
14	Sum	45567
15	Count	120
16	Largest(1)	391
17	Smallest(1)	370
18	Confidence Level(95,0%)	0,879292
19		



Risorsa in rete

(<http://www.mathtools.net/MATLAB/Statistics/Excel/>)

Mathtools.net >

The technical computing portal for all your scientific and engineering needs

Mathtools.net > MATLAB > Statistics > Excel

- [SolverStat](#) - SolverStat is an Excel add-in (xla) for Microsoft Excel 97, 2000, XP. The built-in optimization package Solver yields minimized least squares worksheet functions makes Excel a flexible tool to solve linear and non linear regression.
- [Analyse-it, a statistics add-on](#) - Analyse-It for Microsoft Excel now available!! The parametric, and clinical research statistics. Download FREE 14-day trial of Analyse-It.
- [Blockage Analysis](#) - MaceTech Blockage Analysis is a freeware Excel Add-in. Blockage analysis is a common problem amongst options on account of unawareness, lack of knowledge, unavailability, in both positive and negative impacts of blockage, and, the separate impacts of multiple blockages.
- [CAMPUS](#) - This paper describes the development of a spreadsheet-based CAA program for quantitative disciplines. The CAMPUS program (Computer Aided Marking Program) provides a means of marking and feedback when assessing students.
- [DE Histograms](#) - This Excel Add-In provides fast and easy graphical and statistical analysis of data.

