UNIVERSITÀ DEGLI STUDI DI BERGAMO
Dipartimento di Ingegneria Gestionale, dell'Informazione e della Produzione

22059 – APPLIED TOPICS IN MANAGEMENT ENGINEERING

Excel, Access and Matlab

Prof. Giuseppe Pellegrini

Prof. Renato Redondi

Content

The course consists of the following subjects:

1- Introduction to the use of **Excel, Access and Matlab**. Solving typical problems such as data conversion between different formats and the generation of random numbers for simulation purposes.

2- Practical applications in the field of statistical quality control (sampling, quality control charts, etc.);

3- Practical applications in the field of business economics and organization, management control and information management (investment analysis, clustering and database management)

A simple case

Let us take into account this basic example.

D.C. Montgomery, Introduction to Statistical Quality Control, 6th Ed., p. 83

EXAMPLE 3.7 Tensile Strength of Paper

The tensile strength of paper used to make grocery bags is an important quality characteristic. It is known that the strength—say, x—is normally distributed with mean $\mu = 40$ lb/in^2 and standard deviation $\sigma = 2$ lb/in^2, denoted $x \sim N(40, 2^2)$. The purchaser of the bags requires them to have a strength of at least 35 lb/in^2. Calculate the probability that bags produced from this paper will meet or exceed the specification.

SOLUTION

The probability that a bag produced from this paper will meet or exceed the specification is $P\{x \geq 35\}$. Note that

$$P\{x \geq 35\} = 1 - P\{x \leq 35\}$$

To evaluate this probability from the standard normal tables, we standardize the point 35 and find

$$P\{x \leq 35\} = P\left\{z \leq \frac{35-40}{2}\right\} =$$

$$P\{z \leq -2.5\} = \Phi(-2.5) = 0.0062$$

Consequently, the desired probability is

$$P\{x \geq 35\} = 1 - P\{x \leq 35\} = 1 - 0.0062 = 0.9938$$

Figure 3.18 shows the tabulated probability for both the $N(40, 2^2)$ distribution and the standard normal distribution. Note that the shaded area to the left of 35 lb/in^2 in Fig. 3.18 represents the fraction nonconforming or "fallout" produced by the bag manufacturing process.

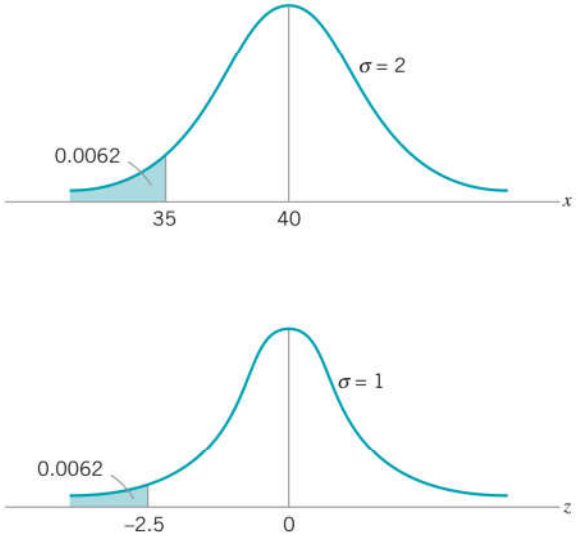


FIGURE 3.18 Calculation of $P\{x \leq 35\}$ in Example 3.7.

Solution(s)

using the tables:

Evaluate the standard variable z : $z = \frac{x - \mu}{\sigma} = \frac{35 - 40}{2} = -2.5$

using the tables, we get F(-2.5) = 0.00621 = 0.6 %

with EXCEL:

`=DISTRIB.NORM.N(35;40;2;VERO)` `=NORM.DIST(35;40;2;VERO)`

else: `=DISTRIB.NORM.ST.N(-2.5)` `=NORM.S.DIST(-2.5)`

with MATLAB:

`normcdf(35,40,2)`

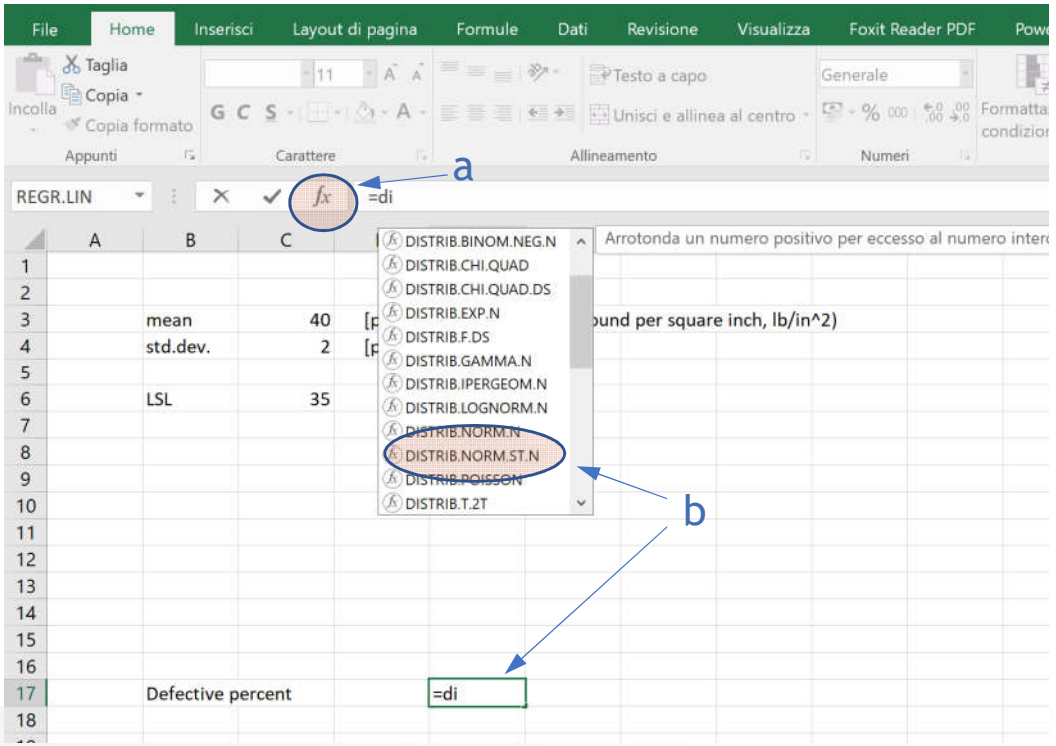
else: `normcdf(-2.5)`

Do it with EXCEL

a) you can also click on the fx icon and select a function from a list. You will get an i/o window with help. It is useful when you have no clear idea on what to do…

b) on a target cell, type "=" and a few letters. A drop list of all available functions will open;

c) you can always type the function name and input parameters (no mistakes!).

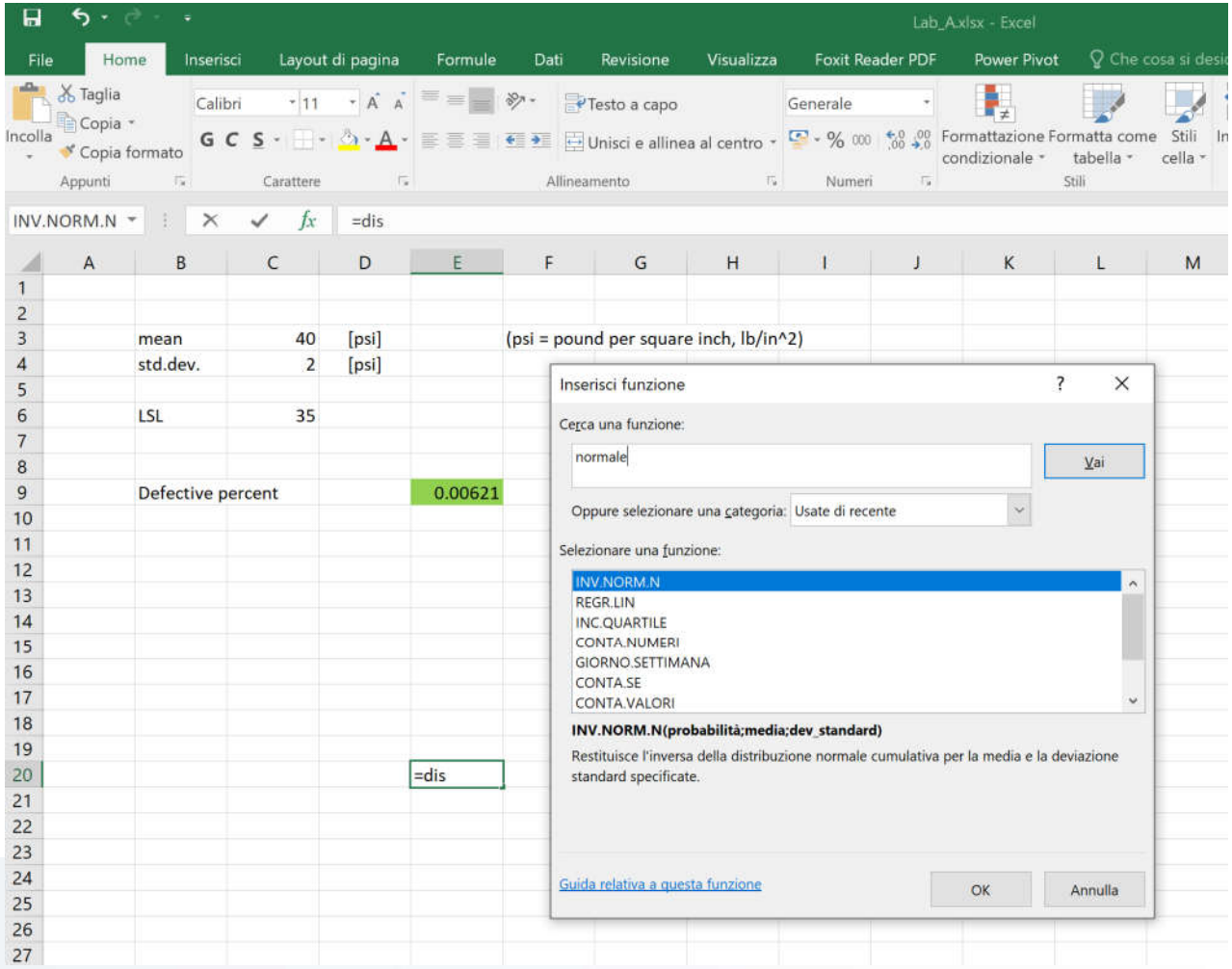


Using the DISTRIB.NORM.N function

Access the function via the fx icon.

Type «`normale`» to look for a function on this subject, or select «`Statistical`» from the function list.

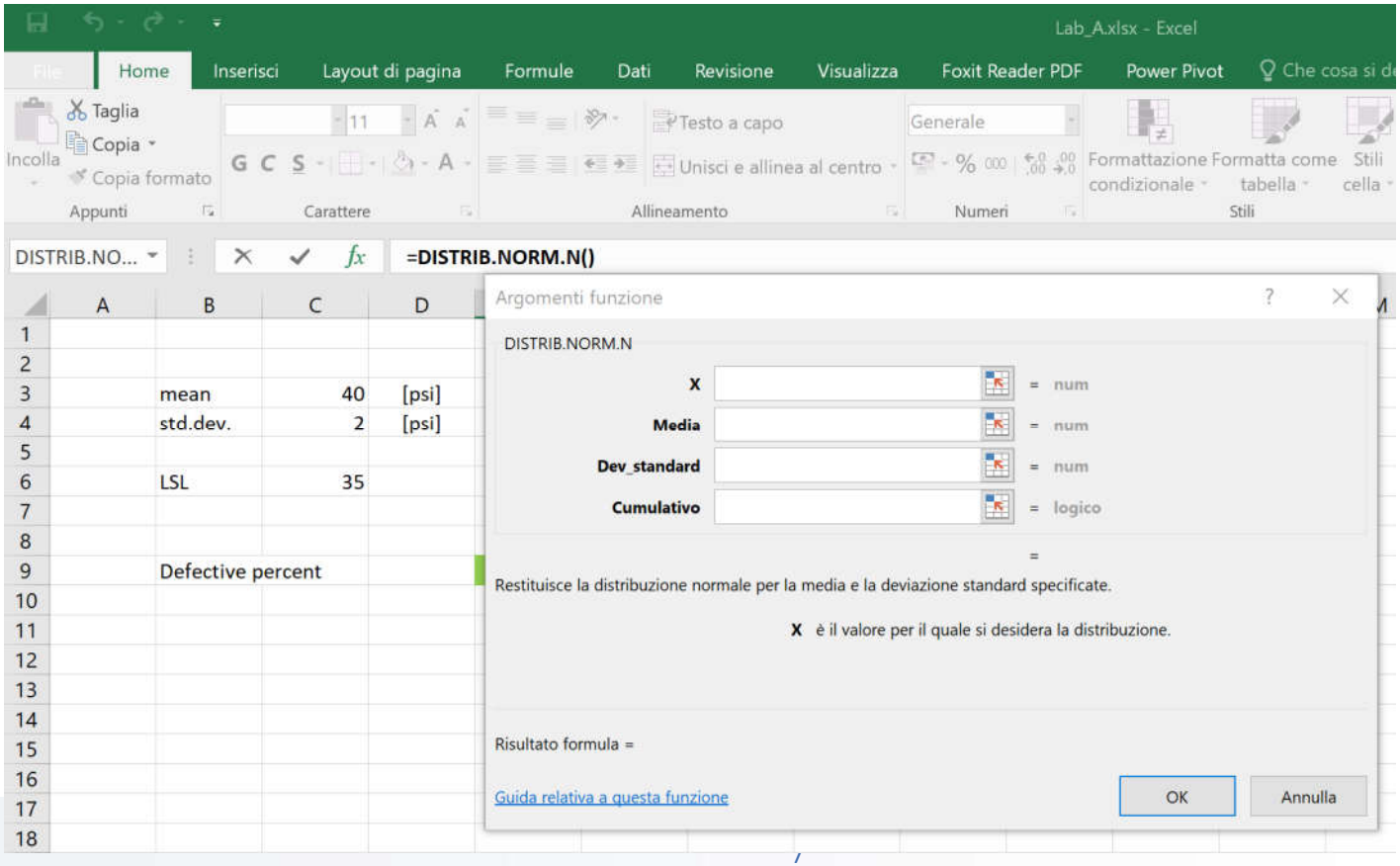
Supply input parameters through the window (with a little help).



Density and Cumulative Probability

The last input field «**Cumulativo**» is a logical flag.

When set to "**FALSE**", ("FALSO", the default value), returns the probability **density**; when set to "**TRUE**" ("VERO") returns the **cumulative** values.



Note

- Every 1000 items, about 6 elements are likely to be nonconforming (that is 0.6%).
- Small deviations from this value are ok (this value is very small and data are random), but large differences are unlikely.
- Check file **Lab_A_data.txt** (or **Lab_A_data.csv**) or and count the defective items (how?)

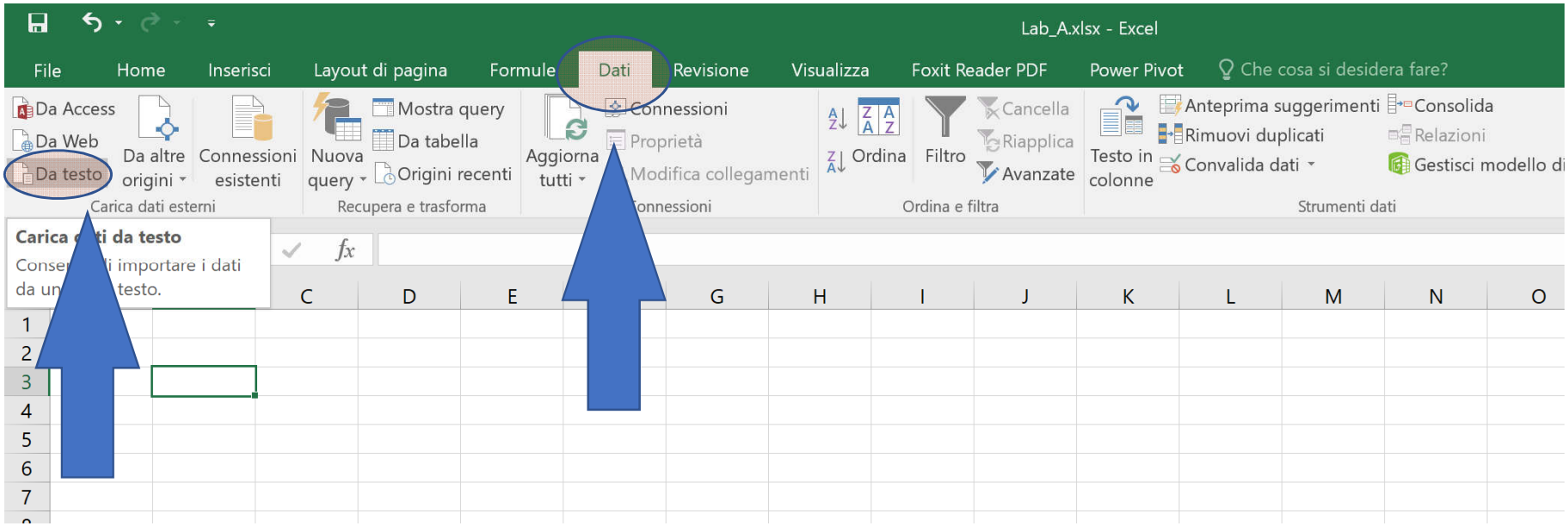
Import data

- Both **Lab_A_data.txt** and **Lab_A_data.csv** are ASCII files (try and open them with a text editor, e.g. Notepad).
- Sometimes the OS associates .csv and, more seldom, .txt with Excel. Anyway, do not rely on this.
- You may cut and paste data (using Ctrl-C and Ctrl-V) or use Excel import function («Dati», «da testo»), that will work in most cases.

Warning: sometimes you will have problems with the data format (. or , as a decimal separator), try **Lab_A_data_bis.txt**…

Importing data with Excel

This is the best way of importing large sets of data. In the modern versions of Excel there are **1,048,576 rows**. In older versions of Excel (2003 and prior) there were 65,536 rows. Handling more than 65,000 is not recommended.



Finding «defective items»

- A «defective item» is a number smaller than 35 psi, our Lower Specification Limit.
- To find such values, you may use the function COUNTIF (CONTA.SE). Use online help for its format.
- Alternatively (safer, IMHO), sort the data (“Dati”, “Ordina”); the smaller numbers will appear on top.

```
=CONTA.SE(C3:C13;"<35”)
=CONTA.SE(C3:C13;"<"&H3)
```

Look at data

With large data sets, it is impossible to understand data by just «looking at» them.

Evaluate the average and the standard deviation for each data set («AVERAGE» and «STDEV», that is «MEDIA» e «DEV.ST.C»).

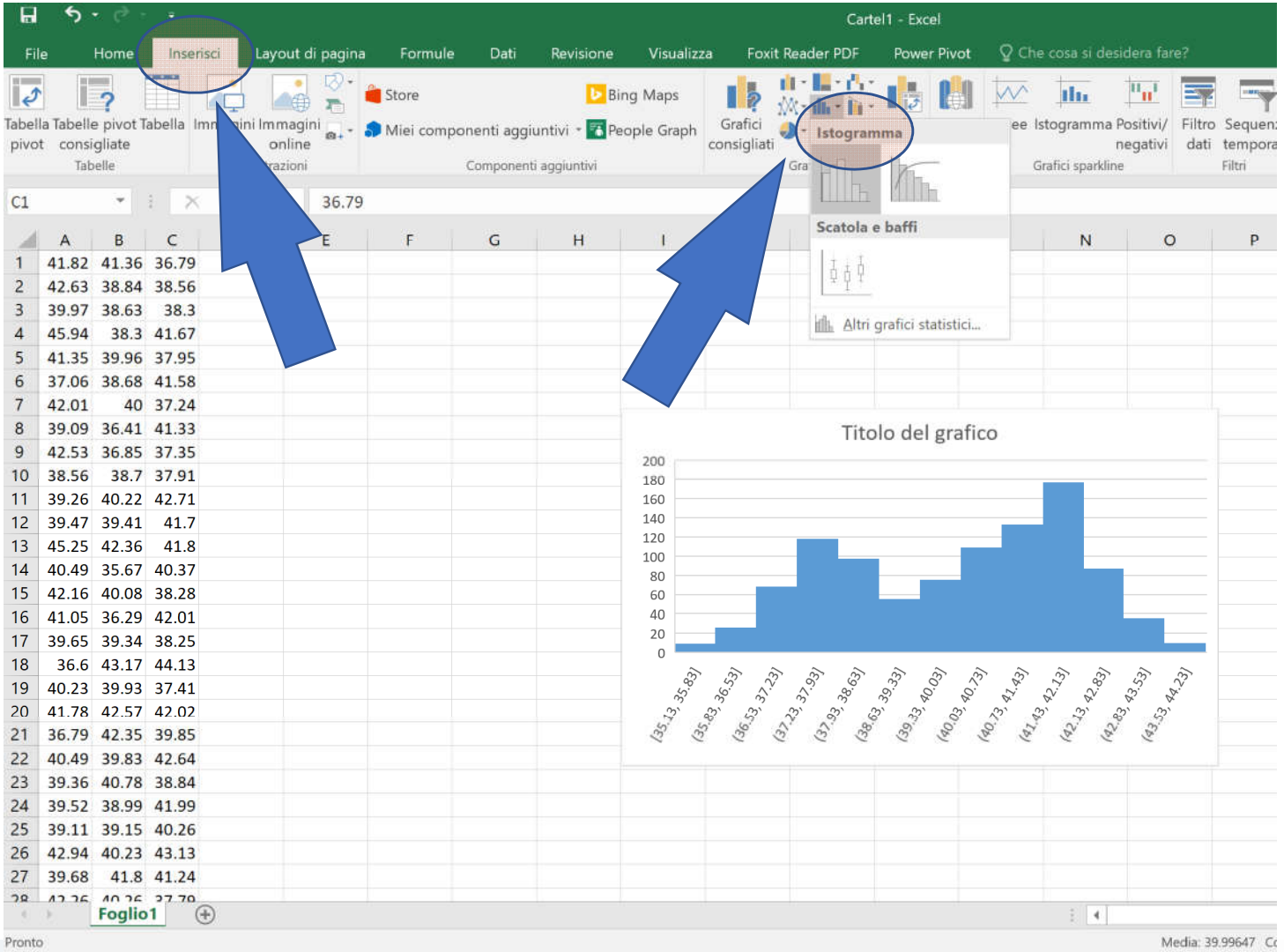
…

Then, try and make histograms to check the data distribution. Select the number of classes, the class limits and then count how many observations are in the class. Plot the frequency table.

Histograms

Excel has a built-in procedure for making histograms.

You shoud adjust the number of bins.

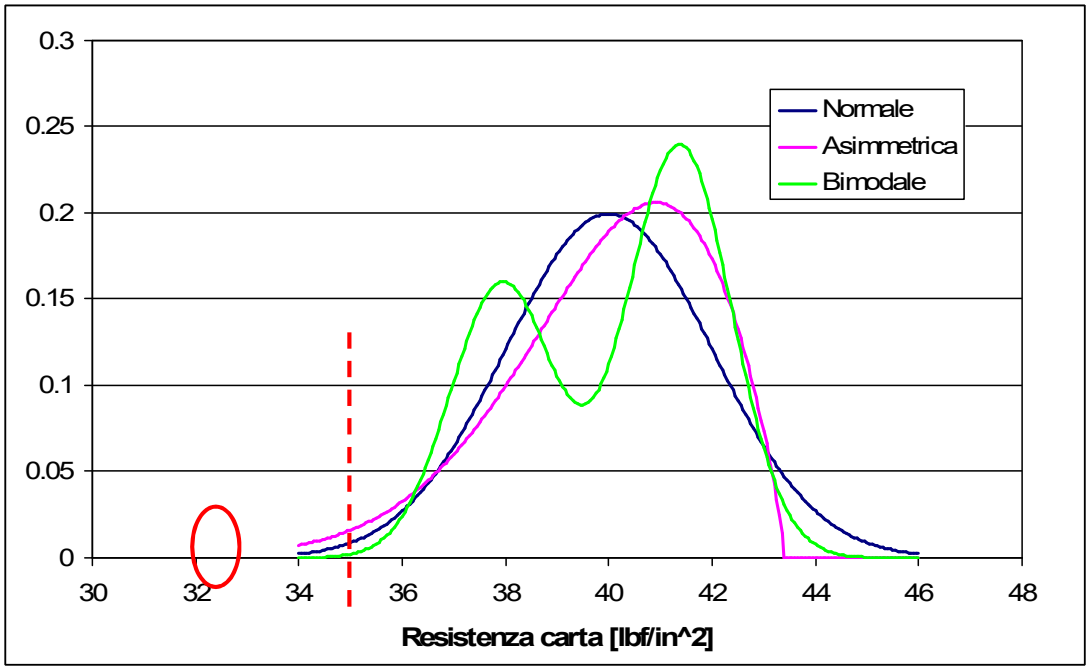


Esempio 1 (1)

Nota: la normalita' dei dati

e' l'ipotesi piu' difficile da verificare e puo' influenzare di molto il risultato

Tutte queste distribuzioni hanno media 40 e dev.st. 2



Esempio 1 (2)

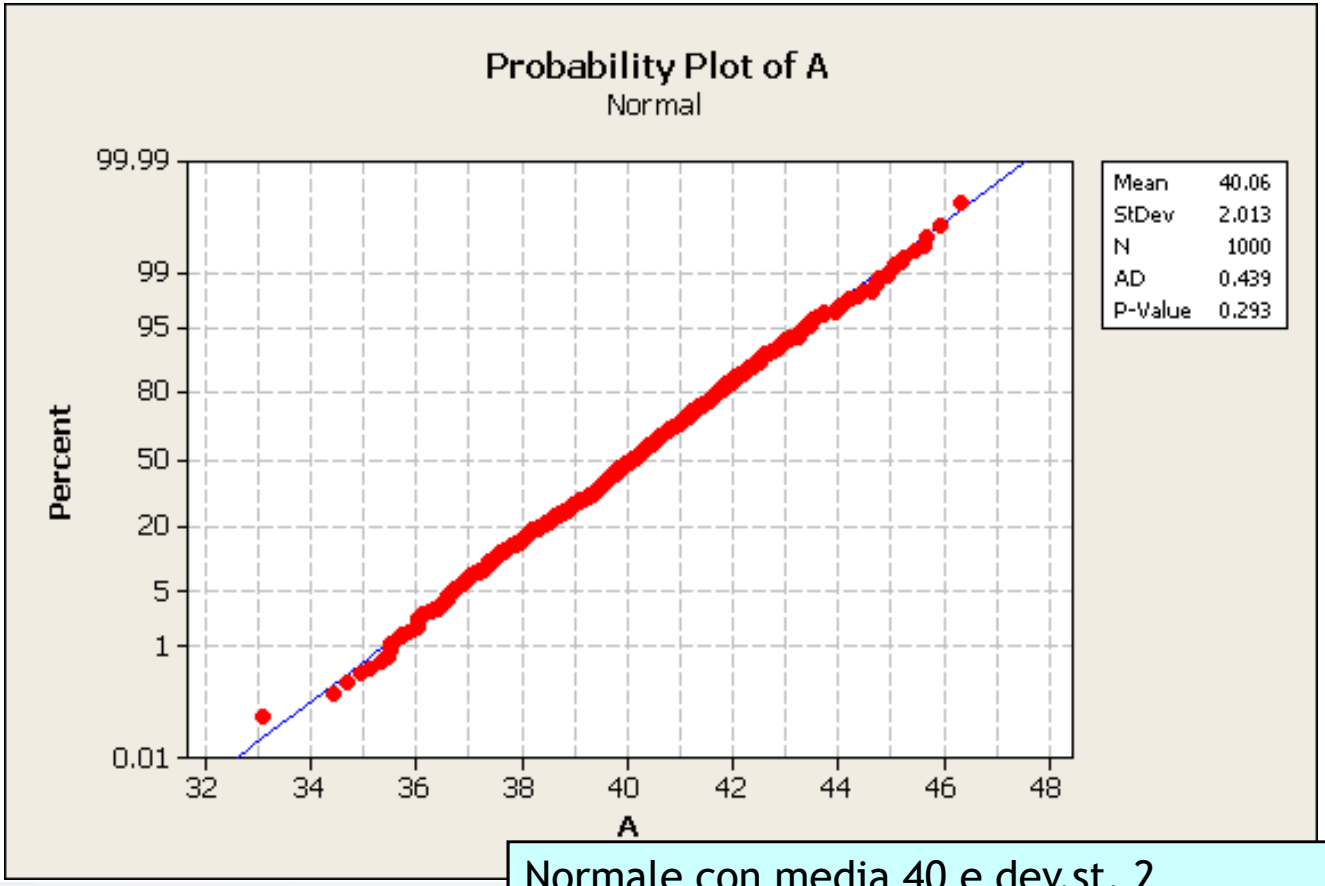
Interpretazione dei casi

- Normale: e’ molto frequente, nessun problema. La deviazione standard misura la ripetibilita’ intrinseca alla lavorazione.

- Asimmetrica: e’ comunque un caso possibile. In questo caso gli scarti sono circa 8 volte piu’ numerosi…

- Bimodale: puo’ essere causata da differenze nel processo (due macchine con regolazione diversa, due lotti di materia prima ecc.). In questo caso ogni macchina e’ piu’ “precisa” (dev. st. pari a 1) ma il risultato finale puo’ essere peggiore.

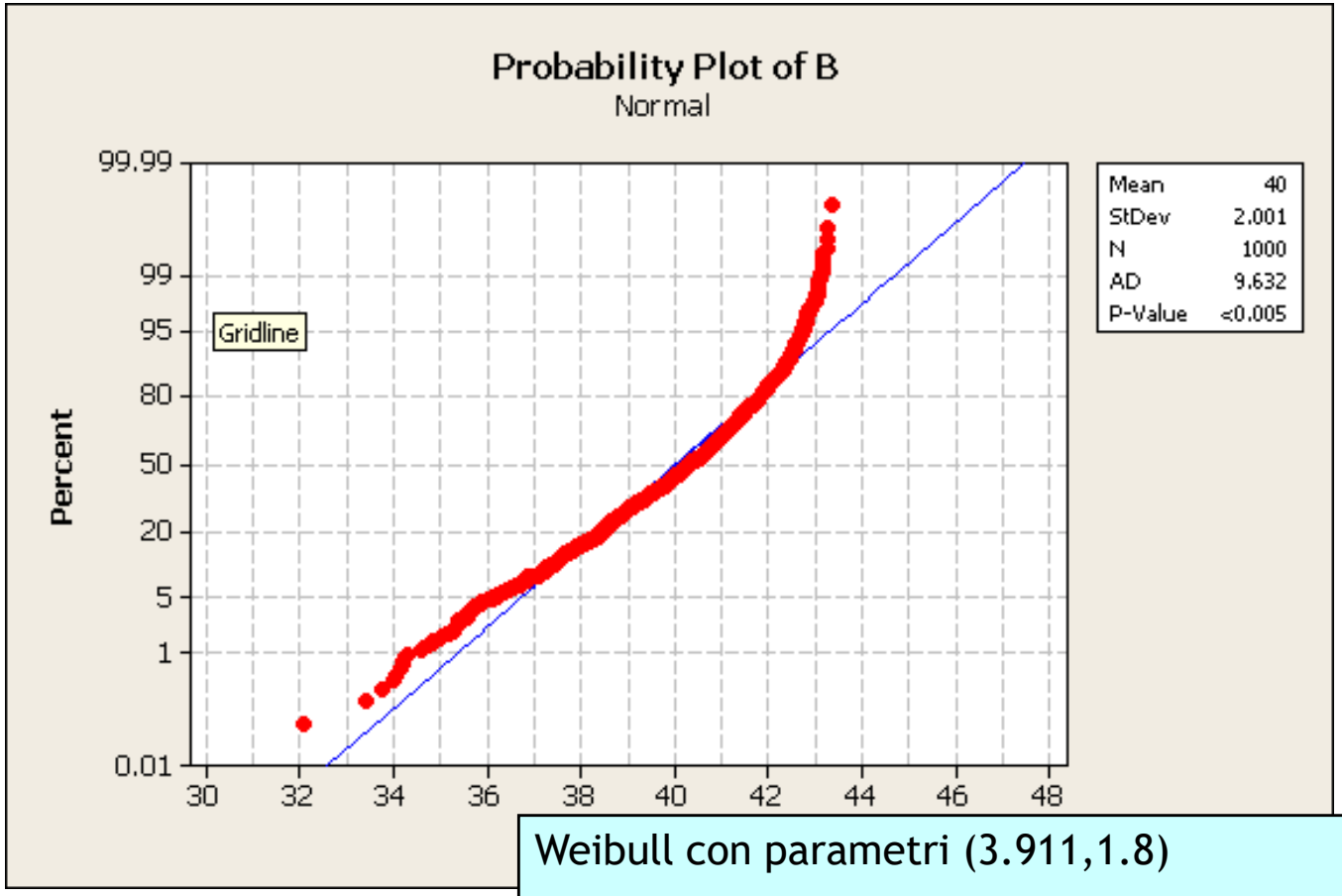
Test di normalita' - normale



Normale con media 40 e dev.st. 2

=DISTRIB.NORM(B5;40;2;FALSO)

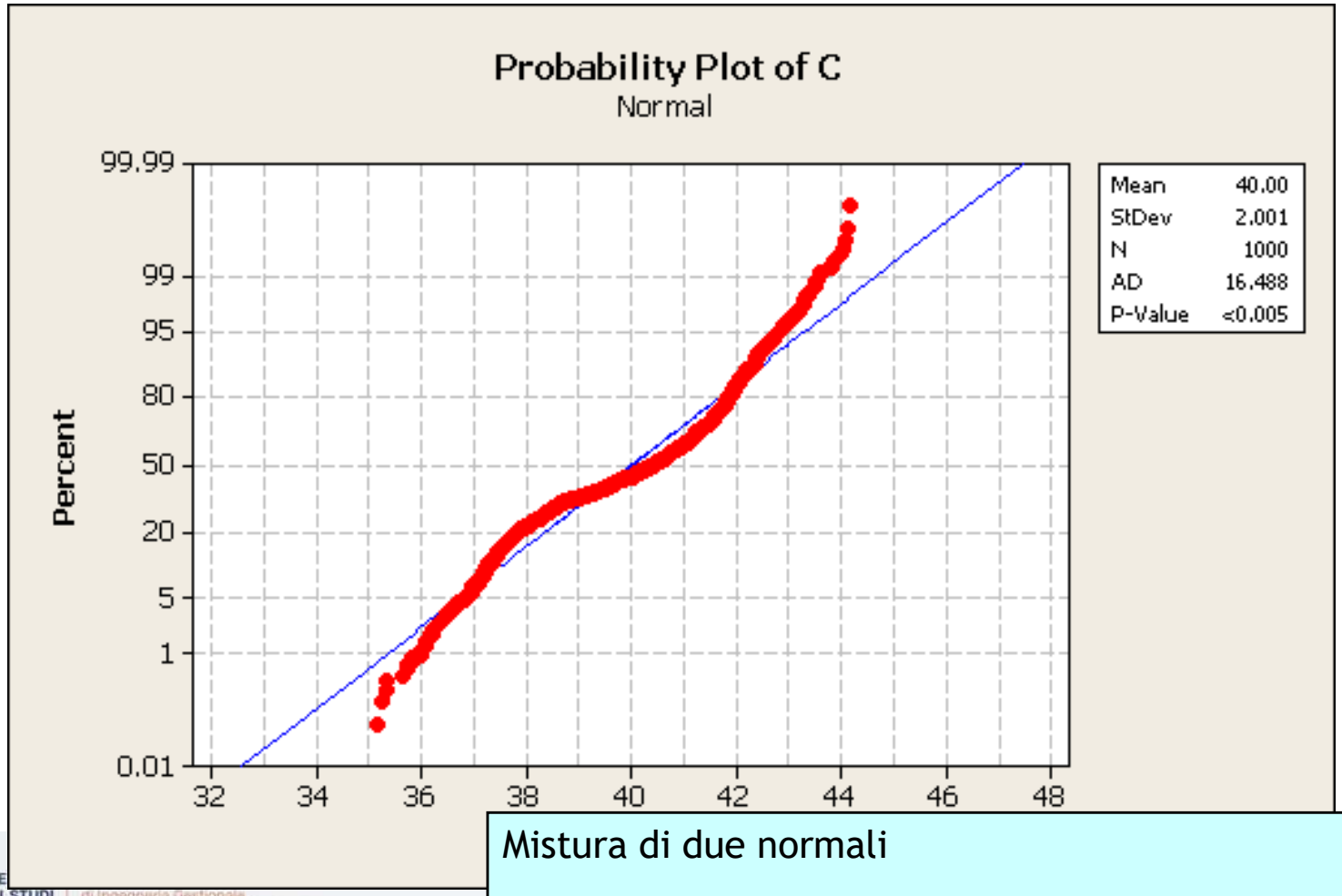
Test di normalita' - asimmetrica



Weibull con parametri (3.911,1.8)

=43.4-weibull(3.911,1.8)

Test di normalita' - bimodale



Mistura di due normali

60%*normale(41.37,1)+40%*normale(37.94,1)