

Analisi Multivariata dei Dati Sperimentali



Massimo Borelli

DM

- perché un corso integrato?
- perché R?
- perché in laboratorio di informatica?

perché R?

The screenshot shows the Bioconductor website with the following content:

- Header:** BIOCONDUCTOR open source software for bioinformatics. Bioconductor is an open source and open development software project for the analysis and comprehension of genomic data.
- Navigation:** home, getting started, overview, downloads, documentation, publications, workshops, cabig.
- Left Sidebar:** Getting Started, Overview, Core Developers, Screenshots, Related Projects, Advisory Board, Downloads, Documentation, Publications, Workshops, Developers, News.
- Main Content: Screenshots and Graphics**
 - affy:** Image of Affymetrix probe-level intensities, *image* function.
 - affy:** MA-plots for three chips, *mva.pairs* function.
 - eddi:** stratified matrix boxplots, *matbox.strat* function.
 - geneploater:** Comparing mean expression levels between two groups.
 - geneploater:** color image of samples x genes expression matrix, separately for each strand, *alongChrom* function.
 - geneploater:** cumulative plot of expression measures along a chromosome, *alongChrom* function.
 - marrayPlots:** image of Cy3 background intensities, *malmage* function.
- Right Sidebar:** Search bar (nel sito, cerca), News (2009-01-07: R, the open source platform used by Bioconductor, featured in a series of articles in the New York Times).

perché un corso integrato

1	Probability Theory (i): One Random Variable	1
3	Statistics (i): An Introduction to Statistical Inference	111
4	Stochastic Processes (i): Poisson Processes and Markov Chains	155
5	The Analysis of One DNA Sequence	174
6	The Analysis of Multiple DNA or Protein Sequences	220
7	Stochastic Processes (ii): Random Walks	257
9	Statistics (iii): Classical Hypothesis Testing Theory	304
	10 BLAST	
11	Stochastic Processes (iii): Markov Chains	384
	12 Hidden Markov Models	
13	Gene Expression, Microarrays, and Multiple Testing	430
	15 Phylogenetic Tree Estimation	

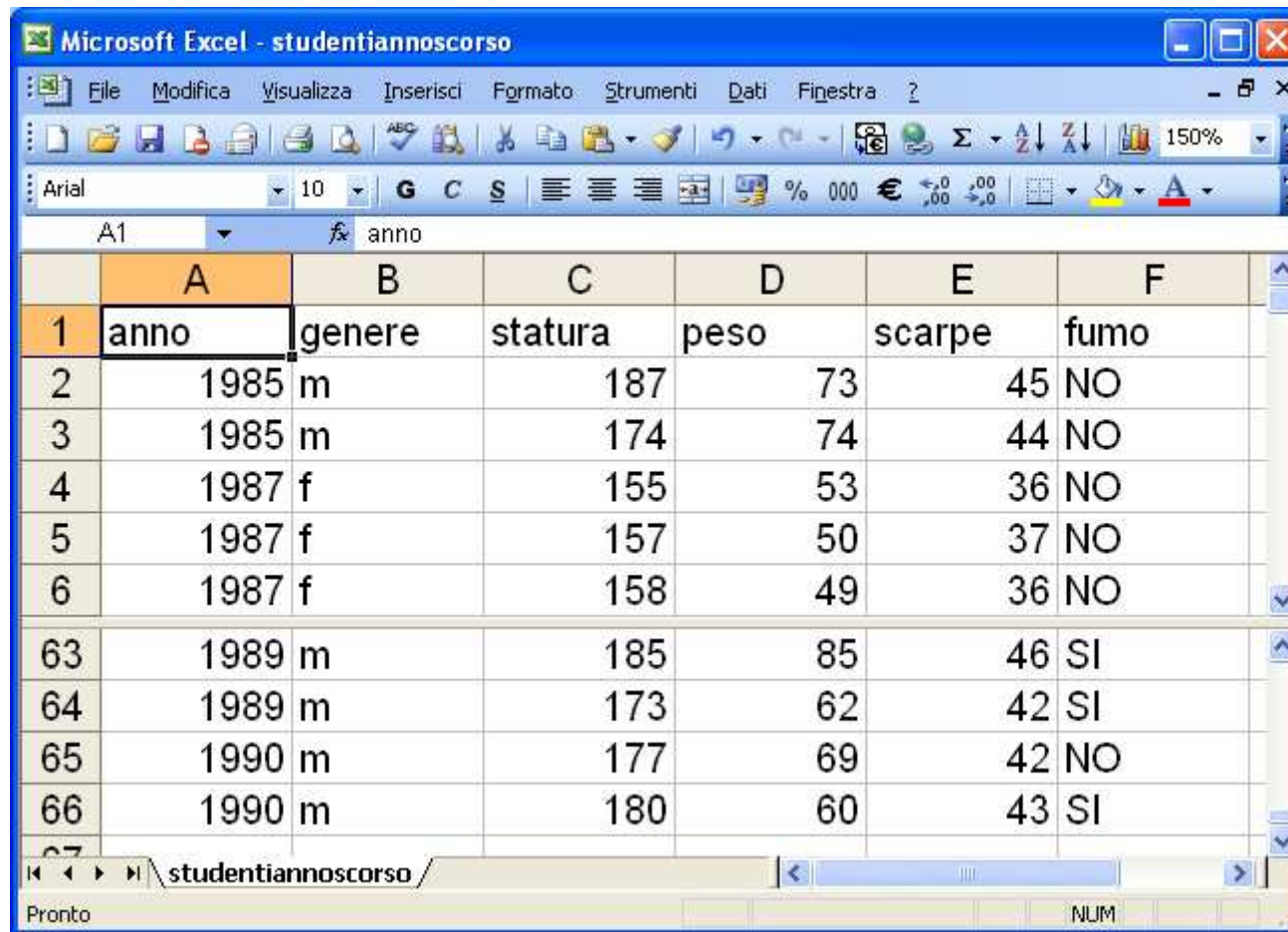
Statistical Methods in Bioinformatics:

An Introduction

perché in laboratorio?

- <http://www.dmi.units.it/~borelli/fotobiotec/P9020071.JPG>

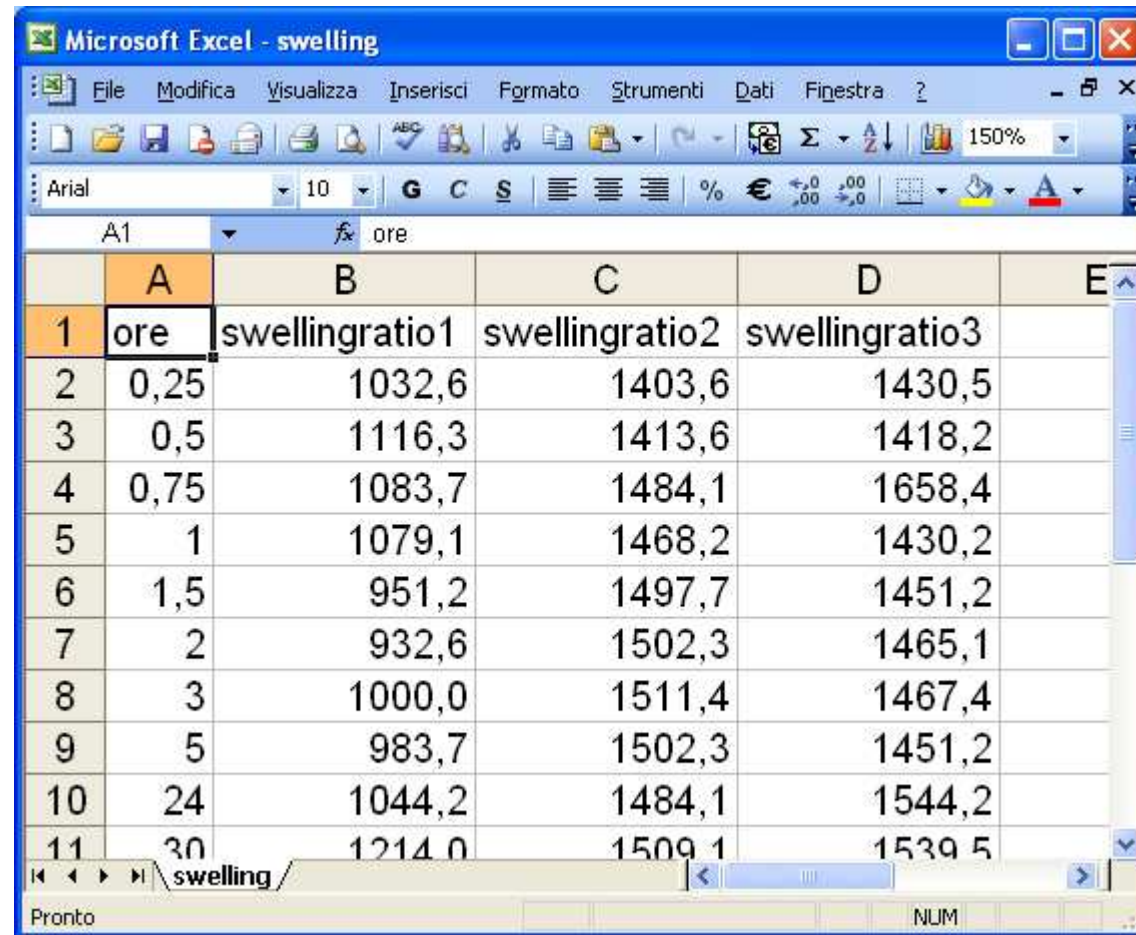
i dataset



The screenshot shows a Microsoft Excel window titled "Microsoft Excel - studentiannoscorso". The spreadsheet contains a dataset with the following columns: "anno" (year), "genere" (gender), "statura" (height), "peso" (weight), "scarpe" (shoes), and "fumo" (smoking). The data is organized into rows, with the first row serving as a header. The rows are numbered 1 through 66, with a jump from row 6 to row 63. The status bar at the bottom indicates "Pronto" and "NUM".

	A	B	C	D	E	F
1	anno	genere	statura	peso	scarpe	fumo
2	1985	m	187	73	45	NO
3	1985	m	174	74	44	NO
4	1987	f	155	53	36	NO
5	1987	f	157	50	37	NO
6	1987	f	158	49	36	NO
63	1989	m	185	85	46	SI
64	1989	m	173	62	42	SI
65	1990	m	177	69	42	NO
66	1990	m	180	60	43	SI

i dataset



Microsoft Excel - swelling

File Modifica Visualizza Inserisci Formato Strumenti Dati Finestra ?

Arial 10 G C S % € 150%

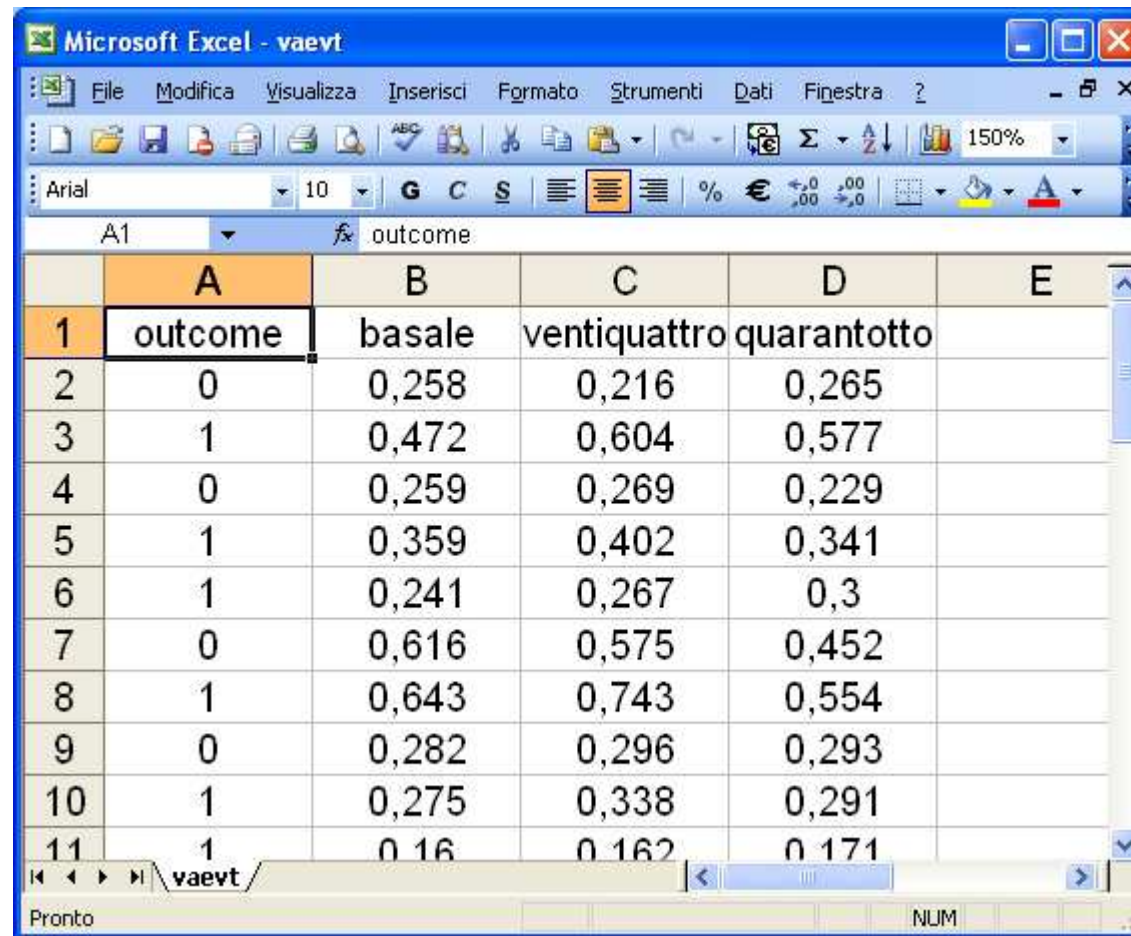
A1 ore

	A	B	C	D	E
1	ore	swellingratio1	swellingratio2	swellingratio3	
2	0,25	1032,6	1403,6	1430,5	
3	0,5	1116,3	1413,6	1418,2	
4	0,75	1083,7	1484,1	1658,4	
5	1	1079,1	1468,2	1430,2	
6	1,5	951,2	1497,7	1451,2	
7	2	932,6	1502,3	1465,1	
8	3	1000,0	1511,4	1467,4	
9	5	983,7	1502,3	1451,2	
10	24	1044,2	1484,1	1544,2	
11	30	1214,0	1509,1	1539,5	

swelling / NUM

Pronto

i dataset



Microsoft Excel - vaevt

File Modifica Visualizza Inserisci Formato Strumenti Dati Finestra ?

Arial 10 G C S % €

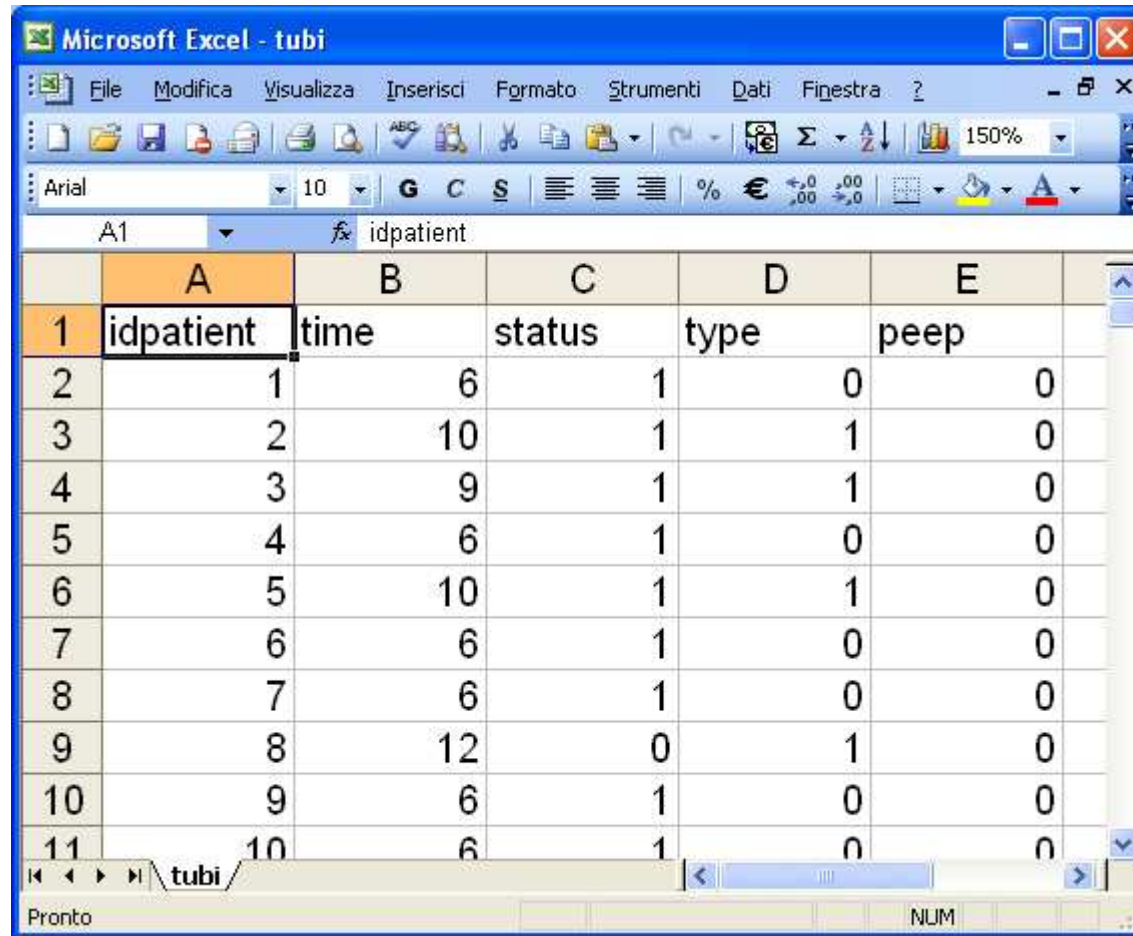
A1 outcome

	A	B	C	D	E
1	outcome	basale	ventiquattro	quarantotto	
2	0	0,258	0,216	0,265	
3	1	0,472	0,604	0,577	
4	0	0,259	0,269	0,229	
5	1	0,359	0,402	0,341	
6	1	0,241	0,267	0,3	
7	0	0,616	0,575	0,452	
8	1	0,643	0,743	0,554	
9	0	0,282	0,296	0,293	
10	1	0,275	0,338	0,291	
11	1	0,16	0,162	0,171	

vaevt

Pronto NUM

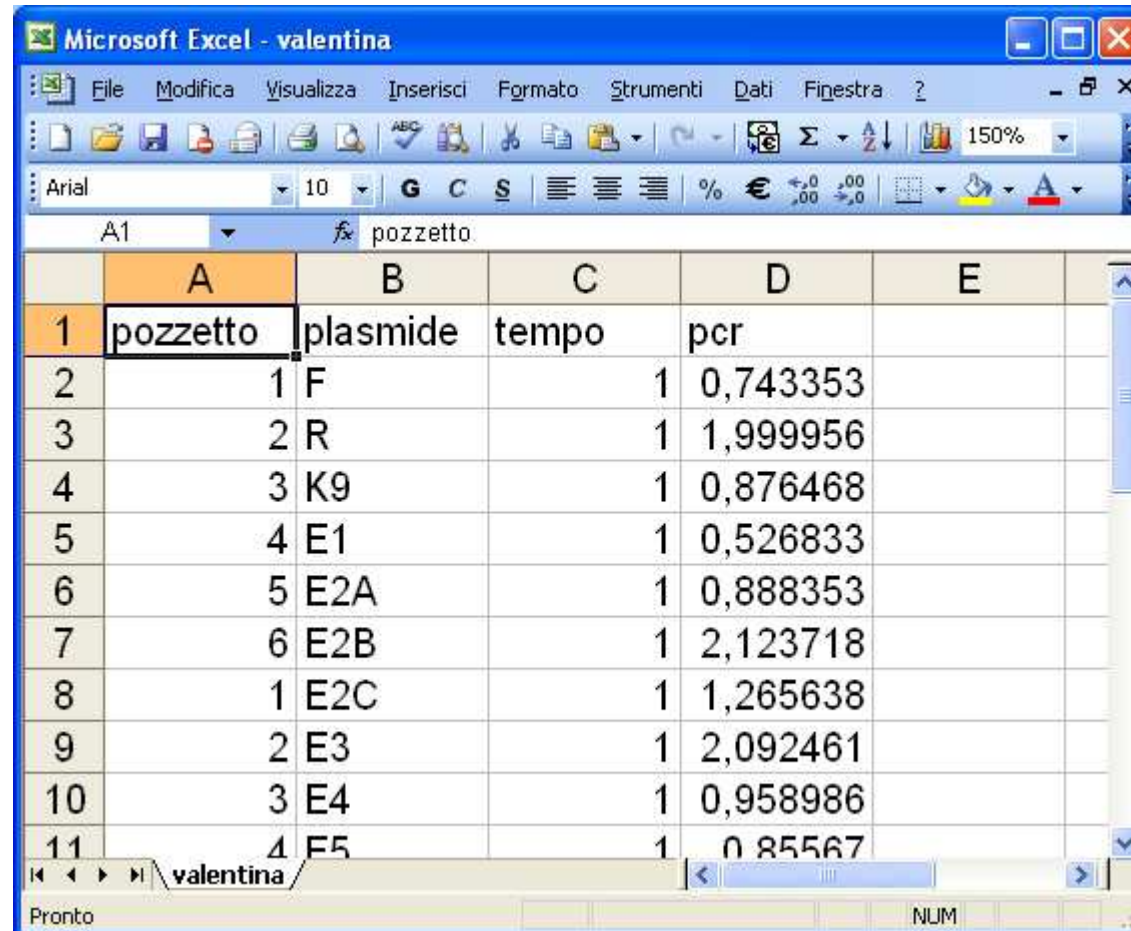
i dataset



The screenshot shows a Microsoft Excel spreadsheet with the following data:

	A	B	C	D	E
1	idpatient	time	status	type	peep
2	1	6	1	0	0
3	2	10	1	1	0
4	3	9	1	1	0
5	4	6	1	0	0
6	5	10	1	1	0
7	6	6	1	0	0
8	7	6	1	0	0
9	8	12	0	1	0
10	9	6	1	0	0
11	10	6	1	0	0

i dataset



Microsoft Excel - valentina

File Modifica Visualizza Inserisci Formato Strumenti Dati Finestra ?

Arial 10 G C S % €

A1 pozzetto.

	A	B	C	D	E
1	pozzetto	plasmide	tempo	pcr	
2	1	F		1 0,743353	
3	2	R		1 1,999956	
4	3	K9		1 0,876468	
5	4	E1		1 0,526833	
6	5	E2A		1 0,888353	
7	6	E2B		1 2,123718	
8	1	E2C		1 1,265638	
9	2	E3		1 2,092461	
10	3	E4		1 0,958986	
11	4	E5		1 0,85567	

valentina

Pronto NUM

