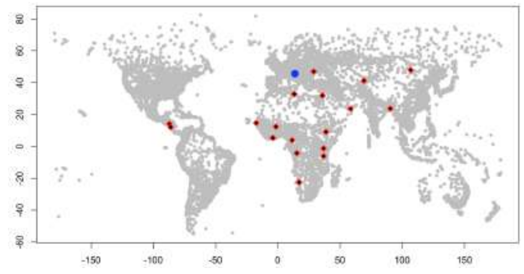
P5.2 Statistics for Medicine

Massimo Borelli

Master of Advanced Studies in Medical Physics

1. Welcome, introduction and objectives
   - Information about our Course

welcome, introduction and objectives



welcome, introduction and objectives

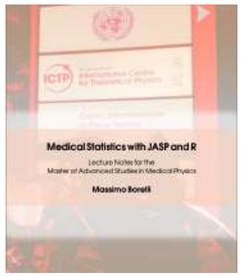
`github.com/MassimoBorelli/ictpmmp`

- free copy of the Lecture Notes
- all the slides
- homeworks (for the final exam)

welcome, introduction and objectives

brief Syllabus

1. Descriptive Statistics
2. Probability and Medicine
3. Sampling and Inference
4. the linear model
5. the generalized linear model



welcome, introduction and objectives

goal to achieve at the end of the course

- to be able to summarize a biomedical dataset, by means of properly choosen statistical indicators
- to be able to provide basic statistical inference, choosing the proper statistical test or regression model
- to be able to properly interpret the frequentist and bayesian reporting

welcome, introduction and objectives

About the exam

1. 'homework' assignments
2. final exam
   1. median vote of the homeworks


Homework #1

welcome, introduction and objectives

Safety, first

my advice: to study and to work together

P5.2 Statistics for Medicine

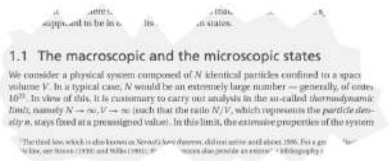
Massimo Borelli

Master of Advanced Studies in Medical Physics

UNIVERSITÀ DEGLI STUDI DI TRIESTE
ICTP International Centre for Theoretical Physics

1. What are we talking about
   - shifting Statistics from Physics to Medicine
   - frequently used softwares

shifting Statistics from Physics to Medicine /1 of 2



1.1 The macroscopic and the microscopic states

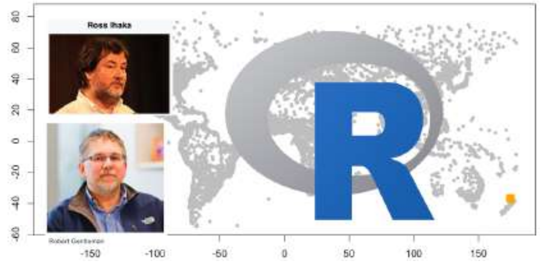
We consider a physical system composed of N identical particles confined to a space volume V. In a typical case, N would be an extremely large number — generally, of order 10^{23}. In view of this, it is customary to carry out analysis in the so-called *thermodynamic limit*, namely $N \to \infty$, $V \to \infty$ (such that the ratio N/V, which represents the *particle density* n, stays fixed at a preassigned value). In this limit, the *extensive* properties of the system

- $N \longrightarrow \infty$?
- $j \in \{1, ..., N\}$!

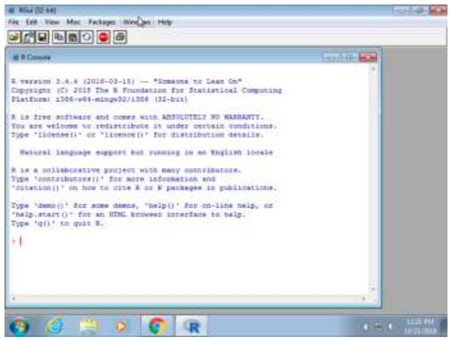
shifting Statistics from Physics to Medicine /2 of 2

Timestamp	Name	Surname	Daybirth	Monthbirth	Yearbirth	Id
28/11/2021 10.55.31	James	Wang	29	12	1966	
28/11/2021 10.56.53	Mary	Chen	19	10	1978	
28/11/2021 10.56.59	Robert	Singh	9	7	1957	
28/11/2021 10.58.00	Patricia	Kumar	12	8	1980	
28/11/2021 11.01.35	John	Ali	11	11	1976	
28/11/2021 11.03.07	Jennifer	Nguyen	7	12	1968	
28/11/2021 11.04.33	Michael	Khan	11	9	1977	
28/11/2021 11.05.04	Linda	Ahmed	26	1	1982	
28/11/2021 11.05.55	William	Khatun	22	1	1960	
28/11/2021 11.06.14	Elizabeth	Silva	18	3	1980	
28/11/2021 11.07.27	David	Tang	13	9	1983	
28/11/2021 11.07.47	Barbara	Mohamed	2	5	1962	
28/11/2021 11.07.47	Richard	Xie	23	8	1966	
28/11/2021 11.08.19	Susan	Han	20	4	1972	

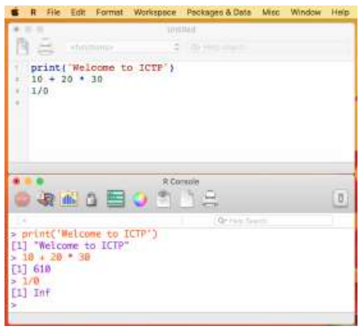
Softwares used by Statisticians /1



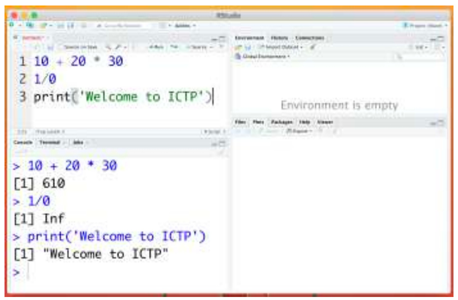
standard console of R /2



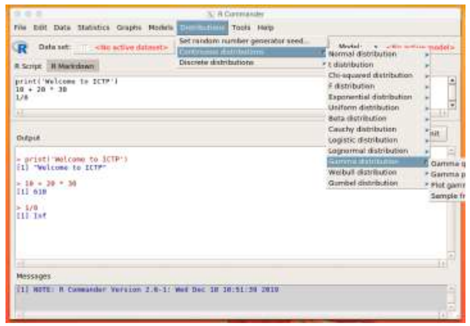
working with scripts in R /3



best interface: R Studio /4



Helping beginners: R Commander /5



Integrating spreadsheet: Jamovi /5

a promising 'new entry': JASP /6

0th homework (optional, not compulsory)

- In hospitals, spreadsheets are routinary
- Very often data not properly masked

protecting privacy in a spreadsheet

As an exercise, download on your computer the `privacy` dataset (at `https://github.com/MassimoBorelli/ictpmmp`), explore it with your favourite spreadsheet and create a new column of data by means of a text function (or joining together the outputs of different text functions) in order to provide a unique identifier for each row ('record') of the dataset.

P5.2 Statistics for Medicine

Massimo Borelli

Master of Advanced Studies in Medical Physics

UNIVERSITÀ DEGLI STUDI DI TRIESTE
ICTP International Centre for Theoretical Physics

Background

To describe the features of a quantitative dataset:

- the **location** of the data
- and their **variability**

Elise Whitley, Jonathan Ball.
Statistics review 1: Presenting and summarising data
`https://ccforum.biomedcentral.com/articles/10.1186/cc1455`

Alla Katsnelson.
Colour me better: fixing figures for colour blindness
`https://www.nature.com/articles/d41586-021-02696-z`

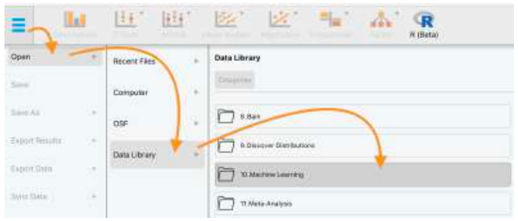
historical example: the `iris` dataset

- *setosa*
- *versicolor*
- *virginica*

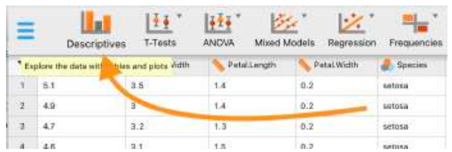
- petal length, petal width
- sepal length, sepal width

preview

`iris` is already stored in JASP



menu Descriptives



lab guided activity /1

Example (position and dispersion measures)

Are we able to understand?

- **measures of central tendency / location**
- **measures of shapes / dispersions**
- the concepts of quantiles
- a **balanced** dataset
- a **complete** dataset

Jonathan Blitzstein, Jessica Hwang.
Introduction to Probability.
`https://projects.iq.harvard.edu/stat110/home`

lab guided activity /2

'A picture is worth a thousand words'

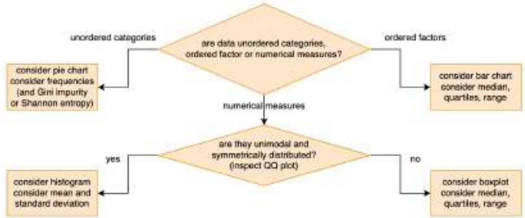
Example (graphs)

- dot plots
- distribution plots
- boxplots (quantiles and outliers?)
- scatter plots

1st Homework Activity (final exam)

Mario de Denaro and Mara Severgnini (Radiation Oncology)

suggested algorithm



parametric vs. non-parametric

Table 1. Characteristics of Infants with Very Low Birth Weight and Those Born at Term.*

Characteristic	Study Participants	Study Nonpart
Very low birth weight		
No. of subjects	166	89
Gestational age — wk	29.17±2.22	29.17±2.
Birth weight — g	1120±221	1130±20

	n	Mean (SD) g/ week	Median g/ week
1994	362	128 (147)	79

funny ;-)

Results

Reliability of automated analysis

For the analysis of the datasets, the two expert analysts manually detected, on average, 4,562 (range 4,439 to 4,686) events (EAdi or P_V events). ICCs for the NeuroSync$_{MANU}$

.. like Sherlock Holmes .. / 1 of 2

$$\mu \approx \frac{a+2m+b}{4} + \frac{a-2m+b}{4n} \approx \frac{a+2m+b}{4}$$

$$\mu \approx \frac{a+2Q_1+2m+2Q_3+b}{8}$$

$$\mu \approx \frac{Q_1+m+Q_3}{3}$$

.. like Sherlock Holmes .. / 2 of 2

$$\sigma \approx \frac{b-a}{\xi(n)}$$

$$\sigma \approx \frac{Q_3-Q_1}{\eta(n)}$$

$$\sigma \approx \frac{1}{2}\left(\frac{b-a}{\xi(n)} + \frac{Q_3-Q_1}{\eta(n)}\right)$$