

# Laboratorio Bio-Demografico Introduzione

---

**Nicola Barban**

16 febbraio 2021

Alma Mater Studiorum- University of Bologna

---



ALMA MATER STUDIORUM  
UNIVERSITÀ DI BOLOGNA

# Outline

- 1 Struttura del corso
- 2 Rstudio
- 3 Datasets
- 4 Rstudio
- 5 Aula virtuale

# Course schedule

- 1 Martedì ore 9:30–11:00 (online)
- 2 Mercoledì 11:00-12:30 LAB AULA VIII
- 3 Giovedì 11:00-12:30 LAB AULA VIII

# Modalità di verifica dell'apprendimento

La prova finale è divisa in due parti:

- 1 **Lavoro di gruppo** (max 4 studenti) da presentare al resto della classe
  - Visualizzazione di dati
  - **Presentazione e discussione 16 e 23 Marzo**
- 2 **Report individuale o di gruppo** (max 4 studenti)
  - Report breve analisi dati
  - Modalità del report verranno distribuite entro il 19 Marzo
  - **Consegna entro 19 Aprile**
  - Discussione/prova orale con il docente

# Modalità di verifica dell'apprendimento (non frequentanti)

## 1 Report individuale

- Report Analisi dati
- Modalità del report verranno definite con il docente
- Discussione/prova orale con il docente

# Modalità del corso

- 1 Martedì → slides (in html)
- 2 Mercoledì e Giovedì. Laboratorio, R syntax
- 3 Esercizi da svolgere in laboratorio. Soluzioni disponibili dopo la lezione.

# Ricevimento Studenti

- 1 online
- 2 Da fissare con il docente (n.barban@unibo.it)

# Riferimenti bibliografici

- **Quantitative Social Science: An Introduction**

<https://press.princeton.edu/books/hardcover/9780691167039/quantitative-social-science> *Kosuke Imai, Princeton University Press*

- **Data Visualization: A Practical Introduction**

<https://socviz.co> *Kieran Healy, Princeton University Press*

## *Altre risorse*

- verranno distribuite durante le lezioni

# Docente

- Professore di demografia UniBo
- precedentemente at University of Essex, and Oxford UK
- MSc and PhD in Statistica e Demografia, Università di Padova
- interessi di ricerca: life course research, genetics and social sciences and genetics
- Principal Investigator GENPOP: Genes, genealogies and the evolution of demographic change and social inequality ([www.genpop.org](http://www.genpop.org))

# Prerequisiti

- Sample mean, sample variance and standard error
- t-test, Normal distribution, confidence interval, p-value
- logarithm, exponential function
- Covariance, correlation
- Linear Regression (OLS); Logistic Regression
- **R**. Moltissime risorse online, ecco alcuni esempi:
  - [swirlstats.com](http://swirlstats.com)
  - <https://cloud.r-project.org/doc/contrib/Torfs+Brauer-Short-R-Intro.pdf>
  - <https://rstudio.com/resources/cheatsheets/>

# Obiettivi del corso

- Diventare degli ottimi statistici

# Obiettivi del corso

- Diventare degli ottimi statistici:
  - 1 **Descrivere** fenomeni bio-demografici con diversi strumenti quantitativi
  - 2 Preparare i dati in maniera professionale
  - 3 **Comunicare** i risultati in maniera efficace e professionale (sia con presentazione orale che in un rapporto scritto)

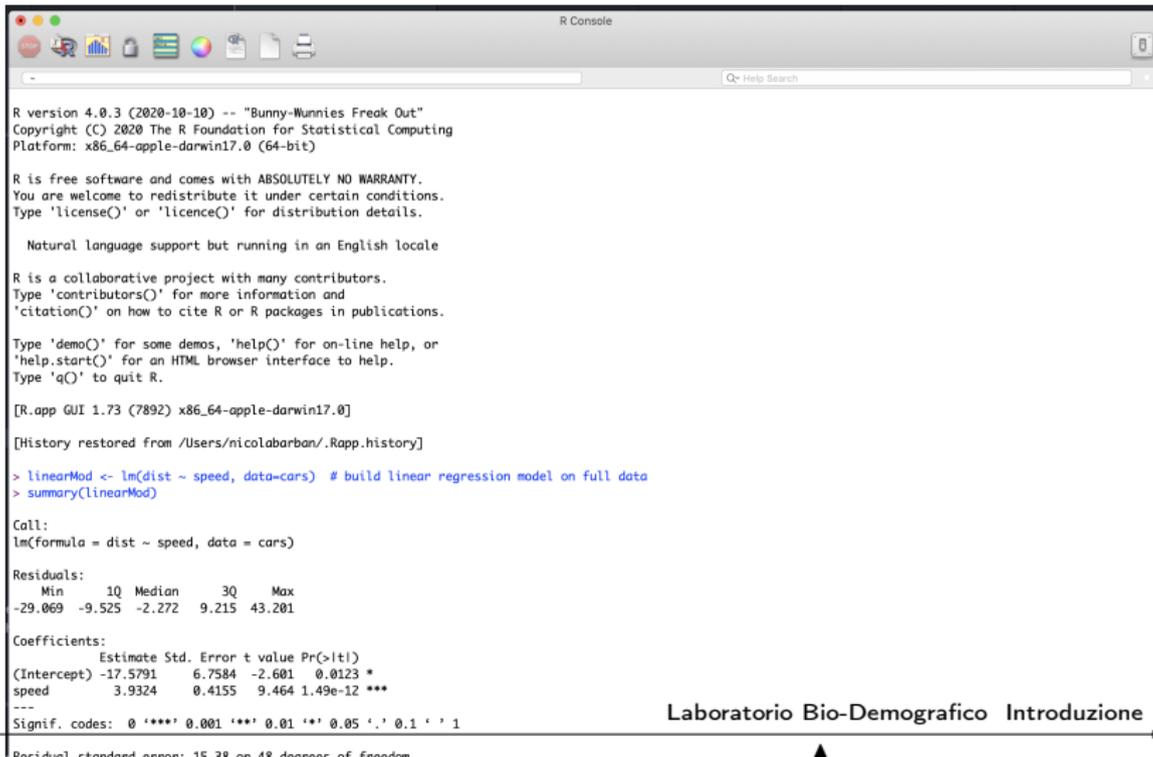
# Rstudio

# R and Rstudio

- **R** works through a command line, or console. At its most basic, you launch it from your Terminal application (on a Mac) or Command Prompt (on Windows) by typing **R**.
- **Rstudio** is

# Example R code

```
1 linearMod <- lm(dist ~ speed, data=cars) # build linear regression model on
  full data
2 summary(linearMod)
```



R Console

```
R version 4.0.3 (2020-10-10) -- "Bunny-Wunnies Freak Out"
Copyright (C) 2020 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin17.0 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[R.app GUI 1.73 (7892) x86_64-apple-darwin17.0]

[History restored from /Users/nicolabarban/.Rapp.history]

> linearMod <- lm(dist ~ speed, data=cars) # build linear regression model on full data
> summary(linearMod)

Call:
lm(formula = dist ~ speed, data = cars)

Residuals:
    Min     1Q   Median     3Q     Max
-29.069  -9.525  -2.272   9.215  43.201

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -17.5791     6.7584  -2.601  0.0123 *
speed         3.9324     0.4155   9.464 1.49e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 15.28 on 48 degrees of freedom
```

Laboratorio Bio-Demografico Introduzione

# Dove trovare i dati

Durante le classi darò esempi dove trovare dataset per i lavori di gruppo

HMD Main Menu	
Registration	
New User	
Change Password	
User's Agreement	
Project	
FAQ	
Overview	
History	
HMD Events	
People	
Acknowledgements	
Research Teams	
HMD Publications	
Methods	
Brief Summary	
Full Protocol	
Special Methods	
Data	
What's New	
Explanatory Notes	
Data Availability	
Zipped Data Files	
Citation Guidelines	
Links	
Max Planck Institute	
UC Berkeley	
UC Berkeley Demography	
INED	
Human Life Table	
Database	
Canadian HMD	
French Human Mortality	
Database	
US Mortality Database	
General	
Contact us	

## The Human Mortality Database

---

Vladimir Shkolnikov, <i>Director</i>	Max Planck Institute for Demographic Research
Magali Barbieri, <i>Associate Director</i>	University of California, Berkeley and INED, Paris
John Wilmoth, <i>Founding Director</i>	United Nations and formerly University of California, Berkeley

---

In response to the COVID-19 pandemic, the HMD team decided to establish a new data resource: **Short-term Mortality Fluctuations (STMF) data series**. Objective and internationally comparable data are crucial to determine the effectiveness of different strategies used to address epidemics. Weekly death counts provide the most objective and comparable way of assessing the scale of short-term mortality elevations across countries and time. [Here](#) we provide weekly death counts for 38 countries: Austria, Australia (Doctor certified deaths), Belgium, Bulgaria, Chile, Canada, Croatia, Czech Republic, Denmark, England and Wales, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Israel, Italy, Latvia, Lithuania, Luxembourg, Netherlands, New Zealand, Northern Ireland, Norway, Poland, Portugal, Republic of Korea, Russia, Scotland, Slovenia, Slovakia, Spain, Sweden, Switzerland, Taiwan and the USA. The same data in the pooled CSV file are available for download [here](#). Data formats and methods are described in the [STMFNote](#). We also strongly recommend reading the [metadata text](#). Following the HMD practice, we also publish [original input data in standardized format](#). During the next few weeks data will be frequently updated and new countries will be added. The most recent STMF update is: 2020-12-31.

**New:** We invite you to explore this data with our online [STMF visualization toolkit](#).

The Human Mortality Database (HMD) was created to provide detailed mortality and population data to researchers, students, journalists, policy analysts, and others interested in the history of human longevity. The project began as an outgrowth of earlier projects in the [Department of Demography at the University of California, Berkeley, USA](#), and at the [Max Planck Institute for Demographic Research in Rostock, Germany](#) (see [history](#)). It is the work of two teams of researchers in the USA and Germany (see [research teams](#)), with the help of financial backers and scientific collaborators from around the world (see [acknowledgements](#)). The Center on the Economics and Development of Aging ([CEDA](#)) French Institute for Demographic Studies ([INED](#)) has also supported the further development of the database in recent years.

# Step 1. Download Rstudio

RStudio Desktop 1.3.1093 - [Release Notes](#)

1. Install R. RStudio requires R 3.0.1+.
2. Download RStudio Desktop. Recommended for your system:



Requires macOS 10.13+ (64-bit)



## All Installers

Linux users may need to [import RStudio's public code-signing key](#) prior to installation, depending on the operating system's security policy.

RStudio requires a 64-bit operating system. If you are on a 32 bit system, you can use an [older version of RStudio](#).

OS	Download	Size	SHA-256
Windows 10/8/7	<a href="#">↓ RStudio-1.3.1093.exe</a>	171.62 MB	62b9e60a
macOS 10.13+	<a href="#">↓ RStudio-1.3.1093.dmg</a>	148.66 MB	bdc4d3a4

# Rstudio Layout

The screenshot displays the RStudio interface with two main panels highlighted by green boxes. The left panel, labeled "Source", contains the R console output and input. The right panel, labeled "Plots", shows a scatter plot of highway mileage (hwy) versus engine displacement (displ) with a smoothed trend line and a confidence interval.

**Console Output:**

```
R version 3.4.1 (2017-06-30) -- "Single Candle"
Copyright (C) 2017 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin15.6.0 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> library(tidyverse)
Loading tidyverse: ggplot2
Loading tidyverse: tibble
Loading tidyverse: tidyr
Loading tidyverse: readr
Loading tidyverse: purrr
Loading tidyverse: dplyr
Conflicts with tidy packages -----
filter(): dplyr, stats
lag():    dplyr, stats
> ggplot(data = mpg, mapping = aes(x = displ, y = hwy)) + geom_point() + geom_smooth()
'geom_smooth()' using method = 'loess'
>
```

**Plots Panel:**

The plot shows a scatter of data points with a blue smoothed trend line and a grey shaded confidence interval. The x-axis is labeled "displ" and ranges from 2 to 7. The y-axis is labeled "hwy" and ranges from 20 to 40. The text "See results in the Plots tab" is overlaid in green above the plot.

**Annotations:**

- A green box highlights the console area with the text "Type commands at the Console" below it.
- A green box highlights the plots area with the text "See results in the Plots tab" above it.

# Aula virtuale

- Slides lezione (in html)
- Script R da compilare con esercizi da svolgere
- script R con soluzioni
- Risorse aggiuntive (link interessanti, ulteriore materiale didattico ecc. )
- Sondaggi con Veevox

www.vevox.app //

www.vevox.app

ID 119-513-549

Join at:  
**vevox.app**

ID:  
**119-513-549**

