

R workflow

Corso introduttivo a R

Inizializzare un R project

New Repository in Github

- 1 creare un nuovo repository in Github (github.com/NAME/)
- 2 copiare l'indirizzo url del repository

Create a new repository

A repository contains all project files, including the revision history. Already have a project repository elsewhere?
[Import a repository.](#)

Required fields are marked with an asterisk ().*

Repository template

No template ▾

Start your repository with a template repository's contents.

Owner *



TMax66 ▾

Repository name *

progetto

✔ progetto is available.

Great repository names are short and memorable. Need inspiration? How about [scaling-guacamole](#) ?

Description (optional)



Public

Anyone on the internet can see this repository. You choose who can commit.



Private

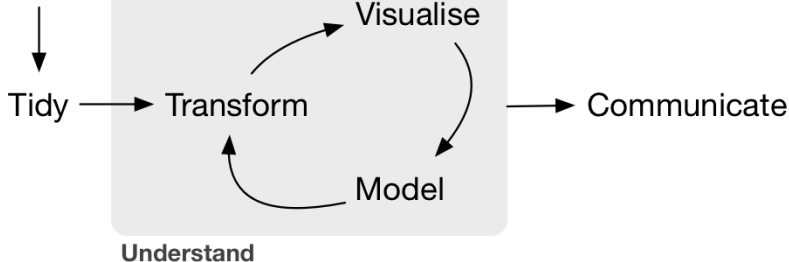
You choose who can see and commit to this repository.

tidyverse

Manipolare, analizzare e comunicare con il paradigma

Tidyverse

Import



i dati devono essere tidy

“**TIDY DATA** is a standard way of mapping the meaning of a dataset to its structure.”

—HADLEY WICKHAM

In tidy data:

each column a variable









Statistica con R

statistical significant: don't say it and don't use it

Common statistical tests are linear models

Last updated: 02 April 2019

See worked examples and more details at the accompanying notebook: <https://lindeleov.github.io/tests-as-linear>

	Common name	Built-in function in R	Equivalent linear model in R	Exact?	The linear model in words	Icon
Simple regression: $lm(y \sim 1 + x)$	y is independent of x P: One-sample t-test N: Wilcoxon signed-rank	t.test(y) wilcox.test(y)	$lm(y \sim 1)$ $lm(\text{signed_rank}(y) \sim 1)$	✓ for $N \geq 14$	One number (intercept, i.e., the mean) predicts y. - (Same, but it predicts the signed rank of y.)	
	P: Paired-sample t-test N: Wilcoxon matched pairs	t.test(y1, y2, paired=TRUE) wilcox.test(y1, y2, paired=TRUE)	$lm(y2 - y1 \sim 1)$ $lm(\text{signed_rank}(y2 - y1) \sim 1)$	✓ for $N \geq 14$	One intercept predicts the pairwise $y_2 - y_1$ differences. - (Same, but it predicts the signed rank of $y_2 - y_1$.)	
	y ~ continuous x P: Pearson correlation N: Spearman correlation	cor.test(x, y, method='Pearson') cor.test(x, y, method='Spearman')	$lm(y \sim 1 + x)$ $lm(\text{rank}(y) \sim 1 + \text{rank}(x))$	✓ for $N \geq 10$	One intercept plus x multiplied by a number (slope) predicts y. - (Same, but with ranked x and y)	
	y ~ discrete x P: Two-sample t-test P: Welch's t-test N: Mann-Whitney U	t.test(y1, y2, var.equal=TRUE) t.test(y1, y2, var.equal=FALSE) wilcox.test(y1, y2)	$lm(y \sim 1 + G_2)^4$ $glm(y \sim 1 + G_2, \text{weights} = \dots^4)$ $lm(\text{signed_rank}(y) \sim 1 + G_2)^4$	✓ for $N \geq 11$	An intercept for group 1 (plus a difference if group 2) predicts y. - (Same, but with one variance per group instead of one common.) - (Same, but it predicts the signed rank of y.)	
Multiple regression: $lm(y \sim 1 + x_1 + x_2 + \dots)$	P: One-way ANOVA N: Kruskal-Wallis	aov(y ~ group) kruskal.test(y ~ group)	$lm(y \sim 1 + G_2 + G_3 + \dots + G_n)^4$ $lm(\text{rank}(y) \sim 1 + G_2 + G_3 + \dots + G_n)^4$	✓ for $N \geq 11$	An intercept for group 1 (plus a difference if group $\neq 1$) predicts y. - (Same, but it predicts the rank of y.)	
	P: One-way ANCOVA	aov(y ~ group + x)	$lm(y \sim 1 + G_2 + G_3 + \dots + G_n + x)^4$	✓	- (Same, but plus a slope on x.) <i>Note: this is discrete AND continuous. ANCOVAs are ANOVAs with a continuous x.</i>	
	P: Two-way ANOVA	aov(y ~ group * sex)	$lm(y \sim 1 + G_2 + G_3 + \dots + G_n + S_2 + S_3 + \dots + S_k + G_2 * S_2 + G_2 * S_3 + \dots + G_n * S_k)^4$	✓	Interaction term: changing sex changes the y ~ group parameters. <i>Note: $G_{i,j,k}$ is an indicator (0 or 1) for each non-intercept levels of the group variable. Similarly for $S_{i,j,k}$ for sex. The first line (with G_j) is main effect of group, the second (with S_j) for sex and the third is the group x sex interaction. For two levels (e.g. male/female), line 2 would just be "S_2" and line 3 would be S_2 multiplied with each G_i.</i>	[Coming]
	Counts ~ discrete x N: Chi-square test	chisq.test(groupXsex_table)	Equivalent log-linear model $glm(y \sim 1 + G_2 + G_3 + \dots + G_n + S_2 + S_3 + \dots + S_k + G_2 * S_2 + G_2 * S_3 + \dots + G_n * S_k, \text{family} = \dots)^4$	✓	Interaction term: (Same as Two-way ANOVA.) <i>Note: Run glm using the following arguments: glm(model, family=poisson()) As linear-model, the Chi-square test is $\log(y) = \log(N) + \log(\alpha) + \log(\beta) + \log(\alpha\beta)$ where α and β are proportions. See more info in the accompanying notebook.</i>	Same as Two-way ANOVA
	N: Goodness of fit	chisq.test(y)	$glm(y \sim 1 + G_2 + G_3 + \dots + G_n, \text{family} = \dots)^4$	✓	(Same as One-way ANOVA and see Chi-Square note.)	1W-ANOVA

List of common parametric (P) non-parametric (N) tests and equivalent linear models. The notation $y \sim 1 + x$ is R shorthand for $y = 1 + x$ which most of us learned in school. Models in similar colors are highly similar, but really, notice how similar they all are across colors! For non-parametric models, the linear models are reasonable approximations for non-small sample sizes (see "Exact" column and click links to see simulations). Other less accurate approximations exist, e.g., Wilcoxon for the sign test and Goodness-of-fit for the binomial test. The signed rank function is $\text{signed_rank} = \text{function}(x) \text{ sign}(x) * \text{rank}(\text{abs}(x))$. The variables G and S are "dummy coded" indicator variables (either 0 or 1) exploiting the fact that when $\Delta x = 1$ between categories the difference equals the slope. Subscripts (e.g., G_i or y_i) indicate different columns in data. lm requires long-format data for all non-continuous models. All of this is exposed in greater detail and worked examples at <https://lindeleov.github.io/tests-as-linear>.

⁴ See the note to the two-way ANOVA for explanation of the notation.

⁵ Same model, but with one variance per group: $glm(\text{value} \sim 1 + G_2, \text{weights} = \text{varident}(\text{form} = \sim 1 | \text{group}), \text{method} = "ML")$.



Comunicazione (rmarkdown / quarto)

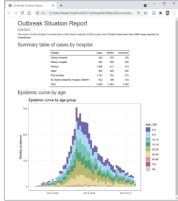
R Markdown script

[illegible]

Word



HTML



Powerpoint

Outbreak Situation Report

This report is for the Incident Command team.
 (The bottom 20 words of this page are redacted.)
 April the

Summary table of cases by hospital

Epidemic curve by age



PDR



Comunicazione (rmarkdown / quarto)

The screenshot displays the RStudio interface with an R Markdown document open. The left pane shows the source code, and the right pane shows the rendered HTML output.

Source Code (Left Pane):

```
1 ---
2 title: "Markdown Demo"
3 output: html_document
4 bibliography: rmarkdown.bib
5 ---
6
7 Markdown provides an easy way to make standard types of formatted
8 text, like
9
10 - *italics*
11 - **bold**
12 - `code`
13 - \[links\]\(rmarkdown.rstudio.com\)
14 - etc.
15
16 But did you know that you can also use R Markdown's markdown to
17 make
18
19 - Latex equations,  $E = mc^2$ 
20 - And bibliographies [@rmarkdown15].
21
22 # References
```

Rendered Output (Right Pane):

Markdown Demo

Markdown provides an easy way to make standard types of formatted text, like

- italics*
- bold**
- `code`
- [links](#)
- etc.

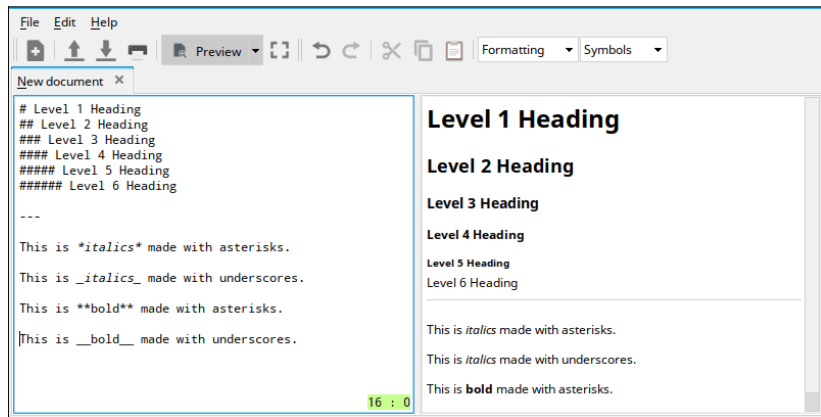
But did you know that you can also use R Markdown's markdown to make

- Latex equations, $E = mc^2$
- And bibliographies ([JJ Allaire 2015](#)).

References

JJ Allaire, et. al. 2015. *R Markdown*. <http://rmarkdown.rstudio.com>.

Comunicazione (rmarkdown / quarto)



The image shows the Quarto editor interface. The left pane contains the source R Markdown code, and the right pane shows the rendered HTML output.

Source Code (Left Pane):

```
# Level 1 Heading
## Level 2 Heading
### Level 3 Heading
#### Level 4 Heading
##### Level 5 Heading
##### Level 6 Heading

---

This is italics made with asterisks.

This is italics made with underscores.

This is bold made with asterisks.

This is bold made with underscores.
```

Rendered Output (Right Pane):

Level 1 Heading

Level 2 Heading

Level 3 Heading

Level 4 Heading

Level 5 Heading

Level 6 Heading

This is *italics* made with asterisks.

This is *italics* made with underscores.

This is **bold** made with asterisks.

The status bar at the bottom left of the editor shows "16 : 0".

Riproducibilità, archiviazione, controllo versioni

Piled Higher and Deeper by Jorge Cham

www.phdcomics.com

"FINAL".doc



FINAL.doc!



FINAL_rev.2.doc



FINAL_rev.6.COMMENTS.doc



FINAL_rev.8.comments5.
CORRECTIONS.doc



FINAL_rev.18.comments7.
corrections9.MORE.30.doc

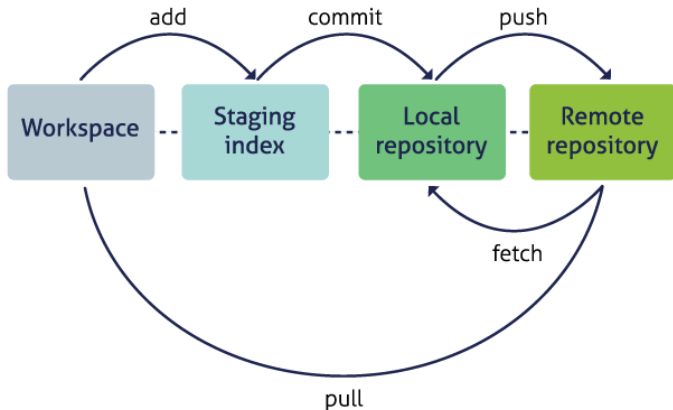


FINAL_rev.22.comments49.
corrections.10.#@\$%WHYDID
ICOMETOGRADSCHOOL?????.doc

github



il ciclo di git



- `git clone` : per copiare un intero repository remoto in locale
- `git add` (stadio) : notifica a git di tenere traccia di particolari cambiamenti
- `git commit` : memorizza quelle modifiche come versione
- `git pull` : unire le modifiche da un repository remoto al nostro repository locale
- `git push` : copia le modifiche dal nostro repository locale in un repository remoto
- `git status` : determinare lo stato di tutti i file nel repository locale
- `git log` : stampa la cronologia delle modifiche in un repository

scrittura collaborativa