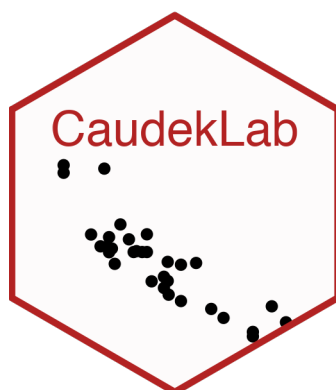


Psicometria

Corrado Caudek

Questo documento è stato realizzato con:

- \LaTeX e la classe memoir (<http://www.ctan.org/pkg/memoir>);
- R (<http://www.r-project.org/>) e RStudio (<http://www.rstudio.com/>);
- bookdown (<http://bookdown.org/>) e memoir (<https://ericmarcon.github.io/memoir/>).



Nel blog della mia pagina personale sono forniti alcuni approfondimenti degli argomenti qui trattati. <https://ccaudek.github.io/caudeklab/>

Indice

Indice	iii
Prefazione	vii
La psicologia e la Data science	vii
Come studiare	viii
Sviluppare un metodo di studio efficace	viii
A Simbologia di base	1
B Introduzione al linguaggio R	3
B.1 Prerequisiti	3
Installare R e RStudio	3
Utilizzare RStudio per semplificare il lavoro	4
Eseguire il codice	4
B.2 Installare <code>cmdstan</code>	5
Bibliografia	7
Elenco delle figure	9

Data della versione presente: Gennaio 12, 2022.

Prefazione

Data Science per psicologi contiene il materiale delle lezioni dell'insegnamento di *Psicometria B000286* (A.A. 2021/2022) rivolto agli studenti del primo anno del Corso di Laurea in Scienze e Tecniche Psicologiche dell'Università degli Studi di Firenze. *Psicometria* si propone di fornire agli studenti un'introduzione all'analisi dei dati in psicologia. Le conoscenze/competenze che verranno sviluppate in questo insegnamento sono quelle della Data science, ovvero un insieme di conoscenze/competenze che si pongono all'intersezione tra statistica (ovvero, richiedono la capacità di comprendere teoremi statistici) e informatica (ovvero, richiedono la capacità di sapere utilizzare un software).

La psicologia e la Data science

It's worth noting, before getting started, that this material is hard. If you find yourself confused at any point, you are normal. Any sense of confusion you feel is just your brain correctly calibrating to the subject matter. Over time, confusion is replaced by comprehension [...]

— Richard McElreath

Sembra sensato spendere due parole su un tema che è importante per gli studenti: quello indicato dal titolo di questo Capitolo. È ovvio che agli studenti di psicologia la statistica non piace. Se piacesse, forse studierebbero Data science e non psicologia; ma non lo fanno. Di conseguenza, gli studenti di psicologia si chiedono: “perché dobbiamo perdere tanto tempo a studiare queste cose quando in realtà quello che ci interessa è tutt'altro?” Questa è una bella domanda.

C'è una ragione molto semplice che dovrebbe farci capire perché la Data science è così importante per la psicologia. Infatti, a ben pensarci, la psicologia è una disciplina intrinsecamente statistica, se per statistica intendiamo quella disciplina che studia la variazione delle caratteristiche degli individui nella popolazione. La psicologia studia *gli individui* ed è proprio la variabilità inter- e intra-individuale ciò che vogliamo descrivere e, in certi casi, predire. In questo senso, la psicologia è molto diversa dall'ingegneria, per esempio. Le proprietà di un determinato ponte sotto certe condizioni, ad esempio, sono molto simili a quelle di un altro ponte, sotto le medesime condizioni. Quindi, per un ingegnere la statistica è poco importante: le proprietà dei materiali sono unicamente dipendenti dalla loro composizione e restano costanti. Ma lo stesso non può dirsi degli individui: ogni individuo è unico e cambia nel tempo. E le variazioni tra gli individui, e di un individuo nel tempo, sono l'oggetto di studio proprio della psicologia: è dunque chiaro che i problemi che la psicologia si pone sono molto diversi da quelli affrontati, per esempio, dagli ingegneri. Questa è la ragione per cui abbiamo tanto bisogno della Data science in psicologia: perché la Data science ci consente di descrivere la variazione e il cambiamento. E queste sono appunto le caratteristiche di base dei fenomeni psicologici.

Sono sicuro che, leggendo queste righe, a molti studenti sarà venuta in mente la seguente domanda: perché non chiediamo a qualche esperto di fare il “lavoro sporco” (ovvero le analisi statistiche) per noi, mentre noi (gli psicologi) ci occupiamo solo di ciò che ci interessa, ovvero dei problemi psicologici slegati dai dettagli “tecnici” della Data

science? La risposta a questa domanda è che non è possibile progettare uno studio psicologico sensato senza avere almeno una comprensione rudimentale della Data science. Le tematiche della Data science non possono essere ignorate né dai ricercatori in psicologia né da coloro che svolgono la professione di psicologo al di fuori dell'Università. Infatti, anche i professionisti al di fuori dall'università non possono fare a meno di leggere la letteratura psicologica più recente: il continuo aggiornamento delle conoscenze è infatti richiesto dalla deontologia della professione. Ma per potere fare questo è necessario conoscere un bel po' di Data science! Basta aprire a caso una rivista specialistica di psicologia per rendersi conto di quanto ciò sia vero: gli articoli che riportano i risultati delle ricerche psicologiche sono zeppi di analisi statistiche e di modelli formali. E la comprensione della letteratura psicologica rappresenta un requisito minimo nel bagaglio professionale dello psicologo.

Le considerazioni precedenti cercano di chiarire il seguente punto: la Data science non è qualcosa da studiare a malincuore, in un singolo insegnamento universitario, per poi poterla tranquillamente dimenticare. Nel bene e nel male, gli psicologi usano gli strumenti della Data science in tantissimi ambiti della loro attività professionale: in particolare quando costruiscono, somministrano e interpretano i test psicometrici. È dunque chiaro che possedere delle solide basi di Data science è un tassello imprescindibile del bagaglio professionale dello psicologo. In questo insegnamento verranno trattati i temi base della Data science e verrà adottato un punto di vista bayesiano, che corrisponde all'approccio più recente e sempre più diffuso in psicologia.

Come studiare

I know quite certainly that I myself have no special talent. Curiosity, obsession and dogged endurance, combined with self-criticism, have brought me to my ideas.

— Albert Einstein

Il giusto metodo di studio per prepararsi all'esame di Psicometria è quello di seguire attivamente le lezioni, assimilare i concetti via via che essi vengono presentati e verificare in autonomia le procedure presentate a lezione. Incoraggio gli studenti a farmi domande per chiarire ciò che non è stato capito appieno. Incoraggio gli studenti a utilizzare i forum attivi su Moodle e, soprattutto, a svolgere gli esercizi proposti su Moodle. I problemi forniti su Moodle rappresentano il livello di difficoltà richiesto per superare l'esame e consentono allo studente di comprendere se le competenze sviluppate fino a quel punto sono sufficienti rispetto alle richieste dell'esame.

La prima fase dello studio, che è sicuramente individuale, è quella in cui è necessario acquisire le conoscenze teoriche relative ai problemi che saranno presentati all'esame. La seconda fase di studio, che può essere facilitata da scambi con altri e da incontri di gruppo, porta ad acquisire la capacità di applicare le conoscenze: è necessario capire come usare un software (R) per applicare i concetti statistici alla specifica situazione del problema che si vuole risolvere. Le due fasi non sono però separate: il saper fare molto spesso ci aiuta a capire meglio.

Sviluppare un metodo di studio efficace

Memorization is not learning.

— Richard Phillips Feynman

Avendo insegnato molte volte in passato un corso introduttivo di analisi dei dati ho notato nel corso degli anni che gli studenti con l'atteggiamento mentale che descriverò qui sotto generalmente ottengono ottimi risultati. Alcuni studenti sviluppano naturalmente questo approccio allo studio, ma altri hanno bisogno di fare uno sforzo per maturarlo.

Fornisco qui sotto una breve descrizione del “metodo di studio” che, nella mia esperienza, è il più efficace per affrontare le richieste di questo insegnamento (Burger & Starbird, 2012).

- Dedicate un tempo sufficiente al materiale di base, apparentemente facile; assicuratevi di averlo capito bene. Cercate le lacune nella vostra comprensione. Leggere presentazioni diverse dello stesso materiale (in libri o articoli diversi) può fornire nuove intuizioni.
- Gli errori che facciamo sono i nostri migliori maestri. Istitivamente cerchiamo di dimenticare subito i nostri errori. Ma il miglior modo di imparare è apprendere dagli errori che commettiamo. In questo senso, una soluzione corretta è meno utile di una soluzione sbagliata. Quando commettiamo un errore questo ci fornisce un’informazione importante: ci fa capire qual è il materiale di studio sul quale dobbiamo ritornare e che dobbiamo capire meglio.
- C’è ovviamente un aspetto “psicologico” nello studio. Quando un esercizio o problema ci sembra incomprensibile, la cosa migliore da fare è dire: “mi arrendo”, “non ho idea di cosa fare!”. Questo ci rilassa: ci siamo già arresi, quindi non abbiamo niente da perdere, non dobbiamo più preoccuparci. Ma non dobbiamo fermarci qui. Le cose “migliori” che faccio (se ci sono) le faccio quando non ho voglia di lavorare. Alle volte, quando c’è qualcosa che non so fare e non ho idea di come affrontare, mi dico: “oggi non ho proprio voglia di fare fatica”, non ho voglia di mettermi nello stato mentale per cui “in 10 minuti devo risolvere il problema perché dopo devo fare altre cose”. Però ho voglia di *divertirmi* con quel problema e allora mi dedico a qualche aspetto “marginale” del problema, che so come affrontare, oppure considero l’aspetto più difficile del problema, quello che non so come risolvere, ma invece di cercare di risolverlo, guardo come altre persone hanno affrontato problemi simili, oppure lo stesso problema in un altro contesto. Non mi pongo l’obiettivo “risolvi il problema in 10 minuti”, ma invece quello di farmi un’idea “generale” del problema, o quello di capire un caso più specifico e più semplice del problema. Senza nessuna pressione. Infatti, in quel momento ho deciso di non lavorare (ovvero, di non fare fatica). Va benissimo se “parto per la tangente”, ovvero se mi metto a leggere del materiale che sembra avere poco a che fare con il problema centrale (le nostre intuizioni e la nostra curiosità solitamente ci indirizzano sulla strada giusta). Quando faccio così, molto spesso trovo la soluzione del problema che mi ero posto e, paradossalmente, la trovo in un tempo minore di quello che, in precedenza, avevo dedicato a “lavorare” al problema. Allora perché non faccio sempre così? C’è ovviamente l’aspetto dei “10 minuti” che non è sempre facile da dimenticare. Sotto pressione, possiamo solo agire in maniera automatica, ovvero possiamo solo applicare qualcosa che già sappiamo fare. Ma se dobbiamo imparare qualcosa di nuovo, la pressione è un impedimento.
- È utile farsi da soli delle domande sugli argomenti trattati, senza limitarsi a cercare di risolvere gli esercizi che vengono assegnati. Quando studio qualcosa mi viene in mente: “se questo è vero, allora deve succedere quest’altra cosa”. Allora verifico se questo è vero, di solito con una simulazione. Se i risultati della simulazione sono quelli che mi aspetto, allora vuol dire che ho capito. Se i risultati sono diversi da quelli che mi aspettavo, allora mi rendo conto di non avere capito e ritorno indietro a studiare con più attenzione la teoria che pensavo di avere capito – e ovviamente mi rendo conto che c’era un aspetto che avevo frainteso. Questo tipo di verifica è qualcosa che dobbiamo fare da soli, in prima persona: nessun altro può fare questo al posto nostro.
- Non aspettatevi di capire tutto la prima volta che incontrate un argomento nuovo.¹ È utile farsi una nota mentalmente delle lacune nella vostra comprensione e tornare su di esse in seguito per cercare di colmarle. L’atteggiamento naturale, quando

¹Ricordatevi inoltre che gli individui tendono a sottostimare la propria capacità di apprendere (Horn & Loewenstein, 2021).

non capiamo i dettagli di qualcosa, è quello di pensare: “non importa, ho capito in maniera approssimativa questo punto, non devo preoccuparmi del resto”. Ma in realtà non è vero: se la nostra comprensione è superficiale, quando il problema verrà presentato in una nuova forma, non riusciremo a risolverlo. Per cui i dubbi che ci vengono quando studiamo qualcosa sono il nostro alleato più prezioso: ci dicono esattamente quali sono gli aspetti che dobbiamo approfondire per potere migliorare la nostra preparazione.

- È utile sviluppare una visione d’insieme degli argomenti trattati, capire l’obiettivo generale che si vuole raggiungere e avere chiaro il contributo che i vari pezzi di informazione forniscono al raggiungimento di tale obiettivo. Questa organizzazione mentale del materiale di studio facilita la comprensione. È estremamente utile creare degli schemi di ciò che si sta studiando. Non aspettate che sia io a fornirvi un riepilogo di ciò che dovete imparare: sviluppate da soli tali schemi e tali riassunti.
- Tutti noi dobbiamo imparare l’arte di trovare le informazioni, non solo nel caso di questo insegnamento. Quando vi trovate di fronte a qualcosa che non capite, o ottenete un oscuro messaggio di errore da un software, ricordatevi: “Google is your friend”.

Corrado Caudek

Appendice A

Simbologia di base

Per una scrittura più sintetica possono essere utilizzati alcuni simboli matematici.

- $\log(x)$: il logaritmo naturale di x .
- L'operatore logico booleano \wedge significa “e” (congiunzione forte) mentre il connettivo di disgiunzione \vee significa “o” (oppure) (congiunzione debole).
- Il quantificatore esistenziale \exists vuol dire “esiste almeno un” e indica l'esistenza di almeno una istanza del concetto/oggetto indicato. Il quantificatore esistenziale di unicità $\exists!$ (“esiste soltanto un”) indica l'esistenza di esattamente una istanza del concetto/oggetto indicato. Il quantificatore esistenziale \nexists nega l'esistenza del concetto/oggetto indicato.
- Il quantificatore universale \forall vuol dire “per ogni.”
- \mathcal{A}, \mathcal{S} : insiemi.
- $x \in A$: x è un elemento dell'insieme A .
- L'implicazione logica “ \Rightarrow ” significa “implica” (se ...allora). $P \Rightarrow Q$ vuol dire che P è condizione sufficiente per la verità di Q e che Q è condizione necessaria per la verità di P .
- L'equivalenza matematica “ \Leftrightarrow ” significa “se e solo se” e indica una condizione necessaria e sufficiente, o corrispondenza biunivoca.
- Il simbolo $|$ si legge “tale che.”
- Il simbolo \triangleq (o $:=$) si legge “uguale per definizione.”
- Il simbolo Δ indica la differenza fra due valori della variabile scritta a destra del simbolo.
- Il simbolo \propto si legge “proporzionale a.”
- Il simbolo \approx si legge “circa.”
- Il simbolo \in della teoria degli insiemi vuol dire “appartiene” e indica l'appartenenza di un elemento ad un insieme. Il simbolo \notin vuol dire “non appartiene.”
- Il simbolo \subseteq si legge “è un sottoinsieme di” (può coincidere con l'insieme stesso). Il simbolo \subset si legge “è un sottoinsieme proprio di.”
- Il simbolo $\#$ indica la cardinalità di un insieme.
- Il simbolo \cap indica l'intersezione di due insiemi. Il simbolo \cup indica l'unione di due insiemi.
- Il simbolo \emptyset indica l'insieme vuoto o evento impossibile.
- In matematica, argmax identifica l'insieme dei punti per i quali una data funzione raggiunge il suo massimo. In altre parole, $\operatorname{argmax}_x f(x)$ è l'insieme dei valori di x per i quali $f(x)$ raggiunge il valore più alto.
- a, c, α, γ : scalari.
- x, y : vettori.
- X, Y : matrici.
- $X \sim p$: la variabile casuale X si distribuisce come p .
- $p(\cdot)$: distribuzione di massa o di densità di probabilità.
- $p(y | x)$: la probabilità o densità di y dato x , ovvero $p(y = Y | x = X)$.

- $f(x)$: una funzione arbitraria di x .
- $f(X; \theta, \gamma)$: f è una funzione di X con parametri θ, γ . Questa notazione indica che X sono i dati che vengono passati ad un modello di parametri θ, γ .
- $\mathcal{N}(\mu, \sigma^2)$: distribuzione gaussiana di media μ e varianza σ^2 .
- $\text{Beta}(\alpha, \beta)$: distribuzione Beta di parametri α e β .
- $\mathcal{U}(a, b)$: distribuzione uniforme con limite inferiore a e limite superiore b .
- $\text{Cauchy}(\alpha, \beta)$: distribuzione di Cauchy di parametri α (posizione: media) e β (scala: radice quadrata della varianza).
- $\mathcal{B}(p)$: distribuzione di Bernoulli di parametro p (probabilità di successo).
- $\text{Bin}(n, p)$: distribuzione binomiale di parametri n (numero di prove) e p (probabilità di successo).
- $\mathbb{KL}(p || q)$: la divergenza di Kullback-Leibler da p a q .

Appendice B

Introduzione al linguaggio R

In questa sezione della dispensa saranno presentate le caratteristiche di base e la filosofia dell'ambiente R, passando poi a illustrare le strutture dati e le principali strutture di controllo. Verranno introdotte alcune funzioni utili per la gestione dei dati e verranno forniti i rudimenti per realizzare semplici funzioni. Verranno introdotti i tipi di file editabili in RStudio (script, markdown, ...). Nello specifico, dopo aver accennato alcune caratteristiche del sistema `tidyverse`, verranno illustrate le principali funzionalità dell'IDE RStudio e dei pacchetti `dplyr` e `ggplot2`. Sul web sono disponibili tantissime introduzioni all'uso di R come, ad esempio, [Hands-On Programming with R](#), [R for Data Science](#), [Data Science for Psychologists](#), e [Introduction to Data Science](#).

B.1 Prerequisiti

Al fine di utilizzare R è necessario eseguire le seguenti tre operazioni nell'ordine dato:

1. Installare R;
2. Installare RStudio;
3. Installare R-Packages (se necessario).

Di seguito viene descritto come installare R e RStudio.

Installare R e RStudio

R è disponibile gratuitamente ed è scaricabile dal sito <http://www.rproject.org/>. Dalla pagina principale del sito `r-project.org` andiamo sulla sezione `Download` e scegliamo un server a piacimento per scaricare il software d'installazione. Una volta scaricato l'installer, lo installiamo come un qualsiasi software, cliccando due volte sul file d'installazione. Esistono versioni di R per tutti i più diffusi sistemi operativi (Windows, Mac OS X e Linux).

Il R Core Development Team lavora continuamente per migliorare le prestazioni di R, per correggere errori e per consentire l'uso di nuove tecnologie. Di conseguenza, periodicamente vengono rilasciate nuove versioni di R. Informazioni a questo proposito sono fornite sulla pagina web <https://www.r-project.org/>. Per installare una nuova versione di R si segue la stessa procedura che è stata seguita per la prima installazione.

Insieme al software si possono scaricare dal sito principale sia manuali d'uso che numerose dispense per approfondire diversi aspetti di R. In particolare, nel sito <http://cran.r-project.org/other-docs.html> si possono trovare anche numerose dispense in italiano (sezione "Other languages").

Dopo avere installato R è opportuno installare anche RStudio. RStudio si può scaricare da <https://www.rstudio.com/>. Anche RStudio è disponibile per tutti i più diffusi sistemi operativi.

Utilizzare RStudio per semplificare il lavoro

Possiamo pensare ad R come al motore di un automobile e a RStudio come al cruscotto di un automobile. Più precisamente, R è un linguaggio di programmazione che esegue calcoli mentre RStudio è un ambiente di sviluppo integrato (IDE) che fornisce un'interfaccia grafica aggiungendo una serie di strumenti che facilitano la fase di sviluppo e di esecuzione del codice. Utilizzeremo dunque R mediante RStudio. In altre parole,

non aprite



aprite invece



L'ambiente di lavoro di RStudio è costituito da quattro finestre: la finestra del codice (scrivere-eseguire script), la finestra della console (riga di comando - output), la finestra degli oggetti (elenco oggetti-cronologia dei comandi) e la finestra dei pacchetti-dei grafici-dell'aiuto in linea.

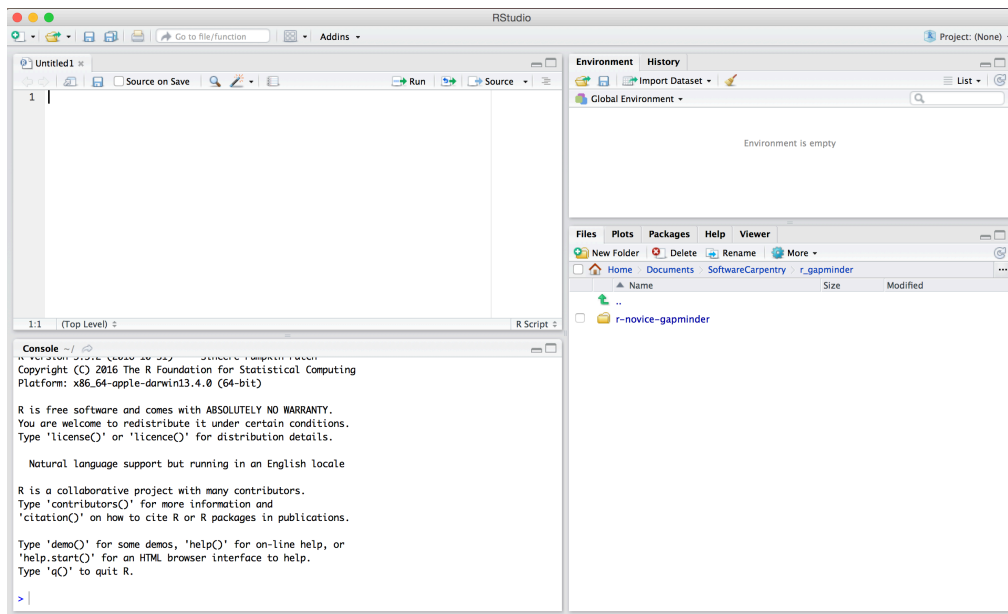


Figura B.1: La console di RStudio.

Eseguire il codice

Mediante il menu a tendina di RStudio, scegliendo il percorso

File > New File > R Notebook

oppure

File > New File > R Script

l'utente può aprire nella finestra del codice (in alto a destra) un R Notebook o un R script dove inserire le istruzioni da eseguire.

In un R script, un blocco di codice viene eseguito selezionando un insieme di righe di istruzioni e digitando la sequenza di tasti `Command + Invio` sul Mac, oppure `Control + Invio` su Windows. In un R Notebook, un blocco di codice viene eseguito schiacciando il bottone con l'icona ► (“Run current chunk”) posizionata a destra rispetto al codice.

B.2 Installare `cmdstan`

È possibile installare `cmdstan` in almeno tre modi. Per informazioni dettagliate, si vedano le istruzioni [CmdStan Installation](#).

Prima di installare `cmdstan`, tre raccomandazioni generali:

- usare la versione più recente del sistema operativo;
- usare la versione più recente di RStudio;
- usare la versione più recente di R.

Se i tre vincoli precedenti sono soddisfatti, l'installazione di `cmdstan` dovrebbe procedere senza intoppi. Altrimenti si possono creare dei problemi di non facile soluzione.

Il modo più semplice per installare `cmdstan` è quello di installare prima `cmdstanr` per poi utilizzare le funzionalità di quel pacchetto per l'installazione di `cmdstan`.

Un secondo metodo (che è quello che io uso normalmente) è quello di installare dal sorgente, seguendo le istruzioni riportate su [CmdStan Installation](#).

Un terzo metodo (che richiede una minima comprensione delle funzionalità della shell e di Python) richiede, avendo prima installato [Anaconda](#), di digitare sulla console del proprio computer (la shell) le seguenti istruzioni:

```
conda create -n stan-env -c conda-forge cmdstan
conda activate stan-env
```

Su macOS, prima di installare `cmdstan`, è necessario installare la versione più recente di Xcode. Dopo avere installato Xcode, aprire la app. Verrà chiesto all'utente se si vogliono installare delle componenti aggiuntive. Questo passaggio è cruciale, perché senza queste componenti aggiuntive `cmdstan` non funzionerà. Dopo avere installato le componenti aggiuntive, aprire Xcode e, in caso, accettare i termini della licenza. A quel punto si può chiudere Xcode. Ogni volta che Xcode viene aggiornato (deve sempre essere aggiornato quando un aggiornamento è disponibile), queste operazioni vanno ripetute.

Bibliografia

- Burger, E. B., & Starbird, M. (2012). *The 5 elements of effective thinking*. Princeton University Press. (Cit. a p. [ix](#)).
- Horn, S., & Loewenstein, G. (2021). Underestimating Learning by Doing. *Available at SSRN 3941441* (cit. a p. [ix](#)).

Elenco delle figure

B.1 La console di RStudio.	4
------------------------------------	---

Abstract This document contains the material of the lessons of Psicometria B000286 (2021/2022) aimed at students of the first year of the Degree Course in Psychological Sciences and Techniques of the University of Florence, Italy.

Keywords Data science, Bayesian statistics.