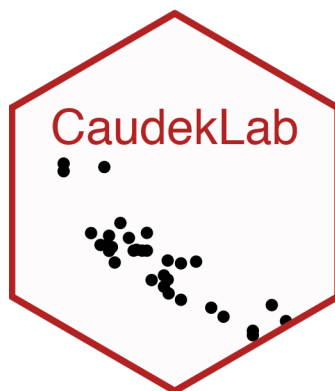


Psicometria

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 memoirR (<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	v
La psicologia e la Data Science	v
Come studiare	vi
Sviluppare un metodo di studio efficace	vii
0.1 Input/Output	1
La funzione <code>read.table()</code>	1
File di dati forniti da R	1
Esportazione di un file	1
Pacchetto rio	2
Dove sono i miei file?	2
Bibliografia	5
Elenco delle figure	7

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.

L'insegnamento di 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 le 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. Istantaneamente 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 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.

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

-
- 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

Febbraio 2022

0.1 Input/Output

I dati raccolti dallo psicologo sono contenuti in file aventi formati diversi: solo testo, CSV, Excel, eccetera. R prevede diverse funzioni di importazione dei dati. Esamineremo qui la funzione `read.table()` per l'importazione di dati in formato solo testo, ma funzioni analoghe possono essere usate per molti altri formati possibili.

La funzione `read.table()`

Ci sono tanti modi per importare un file dal nostro computer. R permette di utilizzare delle funzioni che sono già nella libreria di base, oppure possiamo utilizzare delle funzioni specifiche, a seconda del tipo di file da importare, che sono contenute in pacchetti aggiuntivi. Per leggere i dati da file in R è conveniente preliminarmente generare un file di dati in formato ASCII, disponendoli come si farebbe in una matrice di dati, e mettere questo file nella cartella di lavoro corrente. Fatto questo, si può utilizzare la funzione `read.table()` presente nella libreria di base per leggere l'intero dataset. Se la prima riga del file contiene l'intestazione delle variabili, allora `read.table("my_file.txt", header = TRUE)` interpreterà la prima riga del file come una riga dove sono contenuti i nomi delle variabili, assegnando ciascun nome alle variabili del data frame:

```
mydata <- read.table("my_file.txt", header = TRUE)
```

In alternativa, si può impiegare la funzione `read.csv()`, che è adatta a leggere dati salvati in `.csv`. Utilizzando altre funzioni, si possono leggere in R i dati contenuti in file aventi formati diversi da quelli considerati qui, quali Excel, SPSS, ecc.

File di dati forniti da R

In R esistono comunque oltre 50 insiemi di dati contenuti nel package `base` e altri sono disponibili in altri packages. Per vedere l'elenco degli insiemi di dati disponibili nel package `base` basta usare l'istruzione `data()`; per caricare un particolare insieme di dati, ad esempio `cars`, basta utilizzare l'istruzione

```
data(cars)
```

Nella maggior parte dei casi questo corrisponde a caricare un oggetto, solitamente un `data.frame` dello stesso nome: per l'esempio considerato si avrebbe un data frame di nome `cars`.

Esportazione di un file

Per esportare un `data.frame` in formato `.csv` possiamo scrivere il seguente codice

```
write.csv(df_esempio, file = "esempio.csv", row.names = FALSE)
```

dove `df_esempio` è il `data.frame` da salvare e `esempio.csv` è il file che verrà salvato all'interno della nostra cartella di lavoro.

Pacchetto rio

Un'alternativa più semplice è fornita dalle funzioni fornite dal pacchetto `rio`. Per importare i dati da un file in qualsiasi formato si usa

```
my_data_frame <- rio::import("my_file.csv")
```

Per esportare i dati in un file avente qualsiasi formato si usa invece

```
rio::export(my_data_frame, "my_file.csv")
```

Dove sono i miei file?

Quello che abbiamo detto finora, a proposito dell'importazione ed esportazione dei file, si riferisce a file che si trovano nella cartella di lavoro (*working directory*). Ma non sempre ci troviamo in questa situazione, il che è una buona cosa, perché se dobbiamo gestire un progetto anche leggermente complesso è sempre una buona idea salvare i file che usiamo in cartelle diverse. Per esempio, possiamo usare una cartella chiamata *psicometria* dove salviamo tutto il materiale di questo insegnamento. Nella cartella *psicometria* ci potrà essere una cartella chiamata *scripts* dove salveremo gli script con il codice R utilizzato per i vari esercizi, e una cartella chiamata *data* dove possiamo salvare i dati. Questa organizzazione minimale ci pone, però, di fronte ad un problema: i dati che vogliamo caricare in R non si trovano nella cartella dove sono contenuti gli script. Quando importiamo un file di dati dobbiamo dunque specificare il percorso che identifica la posizione del file sul nostro computer.

Questo problema può essere risolto in due modi: specificando l'indirizzo assoluto del file, o l'indirizzo relativo. Specificare l'indirizzo assoluto di un file comporta una serie di svantaggi. Il più grande è che non sarà possibile utilizzare quell'istruzione su una macchina diversa. Dunque, è molto più conveniente specificare l'indirizzo dei file in modo relativo. Ma relativo rispetto a cosa? Rispetto alla *working directory* che definirà l'origine del nostro percorso.

È ovvio che la *working directory* cambia da progetto a progetto. Infatti, per ciascun progetto dobbiamo specificare una diversa *working directory*. Per esempio, potremmo avere un progetto relativo all'insegnamento di *Psicometria* e un progetto relativo alla prova finale.

Per organizzare il lavoro in questo modo, si procede come segue. Supponiamo di creare una cartella chiamata *psicometria* che contiene, al suo interno, le cartelle *scripts* e *data*:

```
psicometria/
├── data
└── scripts
```

Supponiamo che queste cartelle contengano i file che ho specificato sopra. Chiudiamo RStudio, se è aperto, e lo riapriamo di nuovo. Dal menu selezioniamo `File -> New Project...` In questo modo si aprirà un menu che ci chiederà, tra le altre cose, se vogliamo creare un nuovo progetto (`New project`). Selezioniamo quell'opzione e navighiamo fino alla cartella `psicometria` e selezioniamo `open`. Questo creerà un file chiamato `psicometria.Rproj` nella cartella `psicometria`.

Chiudiamo ora RStudio. Se vogliamo accedere al progetto “psicometria”, che abbiamo appena creato, dobbiamo semplicemente cliccare sul file `psicometria.Rproj`. Questo aprirà RStudio e farà in modo che la *working directory* coincida con la cartella `psicometria`. Ogni volta che vogliamo lavorare sui dati del progetto “psicometria” chiudiamo dunque RStudio (se è già aperto) e lo riapriamo cliccando sul file `psicometria.Rproj`.

A questo punto possiamo definire l'indirizzo dei file in modo relativo – ovvero, relativo alla cartella `psicometria`. Per fare questo usiamo le funzionalità del pacchetto `here`. Supponiamo di volere caricare un file di dati che si chiama `dati_depressione.txt` e si trova nella cartella `psicometria/data`. Per importare i dati (dopo avere caricato i pacchetti `rio` e `here`) useremo l'istruzione seguente:

```
rio::import(here("data", "dati_depressione.txt"))
```

In altre parole, così facendo specifichiamo il percorso relativo del file `dati_depressione.txt` (in quanto l'origine corrisponde alla cartella `psicometria`). L'istruzione precedente significa che, partendo dalla cartella che coincide con la *working directory* (ovvero, `psicometria`) ci spostiamo nella cartella `data` e lì dentro troviamo il file chiamato `dati_depressione.txt`.

Bibliografia

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

Elenco delle figure

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.