

# Psicometria

Corrado Caudek

Questo documento è stato realizzato con:

- $\text{\LaTeX}$  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

<b>Indice</b>	<b>iii</b>
<b>Prefazione</b>	<b>vii</b>
La psicologia e la Data Science . . . . .	vii
Come studiare . . . . .	viii
Sviluppare un metodo di studio efficace . . . . .	viii
<b>1 Valore atteso e varianza</b>	<b>1</b>
1.1 Valore atteso . . . . .	1
Interpretazione . . . . .	2
Proprietà del valore atteso . . . . .	2
Variabili casuali continue . . . . .	4
1.2 Varianza . . . . .	4
Formula alternativa per la varianza . . . . .	5
Variabili casuali continue . . . . .	5
1.3 Deviazione standard . . . . .	5
1.4 Standardizzazione . . . . .	6
1.5 Momenti di variabili casuali . . . . .	6
1.6 Funzione di ripartizione . . . . .	6
<b>Bibliografia</b>	<b>9</b>
<b>Elenco delle figure</b>	<b>11</b>



Copyright © 2022.

Data della versione presente: Dicembre 18, 2021.



## 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 psico-

logico 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. Istintivamente 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.<sup>1</sup> 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

---

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

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

Febbraio 2022

## Valore atteso e varianza

Spesso risulta utile fornire una rappresentazione sintetica della distribuzione di una variabile casuale attraverso degli indicatori caratteristici piuttosto che fare riferimento ad una sua rappresentazione completa mediante la funzione di ripartizione, o la funzione di massa o di densità di probabilità. Una descrizione più sintetica di una variabile casuale, tramite pochi valori, ci consente di cogliere le caratteristiche essenziali della distribuzione, quali: la posizione, cioè il baricentro della distribuzione di probabilità; la variabilità, cioè la dispersione della distribuzione di probabilità attorno ad un centro; la forma della distribuzione di probabilità, considerando la simmetria e la curtosi (pesantezza delle code). In questo Capitolo introdurremo quegli indici sintetici che descrivono il centro di una distribuzione di probabilità e la sua variabilità.

### 1.1 Valore atteso

Quando vogliamo conoscere il comportamento tipico di una variabile casuale spesso vogliamo sapere qual è il suo “valore tipico”. La nozione di “valore tipico”, tuttavia, è ambigua. Infatti, essa può essere definita in almeno tre modi diversi:

- la *media* (somma dei valori divisa per il numero dei valori),
- la *mediana* (il valore centrale della distribuzione, quando la variabile è ordinata in senso crescente o decrescente),
- la *moda* (il valore che ricorre più spesso).

Per esempio, la media di  $\{3, 1, 4, 1, 5\}$  è  $\frac{3+1+4+1+5}{5} = 2.8$ , la mediana è 3 e la moda è 1. Tuttavia, la teoria delle probabilità si occupa di variabili casuali piuttosto che di sequenze di numeri. Diventa dunque necessario precisare che cosa intendiamo per “valore tipico” quando facciamo riferimento alle variabili casuali. Giungiamo così alla seguente definizione.

**Definizione 1.1.** Sia  $Y$  è una variabile casuale discreta che assume i valori  $y_1, \dots, y_n$  con distribuzione  $p(y)$ , ossia

$$P(Y = y_i) = p(y_i),$$

per definizione il *valore atteso* di  $Y$ ,  $\mathbb{E}(Y)$ , è

$$\mathbb{E}(Y) = \sum_{i=1}^n y_i \cdot p(y_i). \quad (1.1)$$

A parole: il valore atteso (o speranza matematica, o aspettazione, o valor medio) di una variabile casuale è definito come la somma di tutti i valori che la variabile casuale può prendere, ciascuno pesato dalla probabilità con cui il valore è preso.

**Esempio 1.1.** Calcoliamo il valore atteso della variabile casuale  $Y$  corrispondente al lancio di una moneta equilibrata (testa:  $Y = 1$ ; croce:  $Y = 0$ ).

$$\mathbb{E}(Y) = \sum_{i=1}^2 y_i \cdot P(y_i) = 0 \cdot \frac{1}{5} + 1 \cdot \frac{1}{5} = 0.5.$$

**Esempio 1.2.** Supponiamo ora che  $Y$  sia il risultato del lancio di un dado equilibrato. Il valore atteso di  $Y$  diventa:

$$\mathbb{E}(Y) = \sum_{i=1}^6 y_i \cdot P(y_i) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + \dots + 6 \cdot \frac{1}{6} = \frac{21}{6} = 3.5.$$

### Interpretazione

Che interpretazione può essere assegnata alla nozione di valore atteso? Bruno de Finetti adottò lo stesso termine di *previsione* (e lo stesso simbolo) tanto per la probabilità che per la speranza matematica. Si può pertanto dire che, dal punto di vista bayesiano, la speranza matematica è l'estensione naturale della nozione di probabilità soggettiva.<sup>1</sup>

### Proprietà del valore atteso

La proprietà più importante del valore atteso è la linearità: il valore atteso di una somma di variabili casuali è uguale alla somma dei loro rispettivi valori attesi:

$$\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y). \quad (1.2)$$

La (1.2) sembra ragionevole quando  $X$  e  $Y$  sono indipendenti, ma è anche vera quando  $X$  e  $Y$  sono associati. Abbiamo anche che

$$\mathbb{E}(cY) = c\mathbb{E}(Y). \quad (1.3)$$

La (1.3) ci dice che possiamo estrarre una costante dall'operatore di valore atteso. Tale proprietà si estende a qualunque numero di variabili casuali. Infine, se due variabili casuali  $X$  e  $Y$  sono indipendenti, abbiamo che

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y). \quad (1.4)$$

**Esercizio 1.1.** Si considerino le seguenti variabili casuali:  $Y$ , ovvero il numero che si ottiene dal lancio di un dado equilibrato, e  $Z$ , il numero di teste prodotto dal lancio di una moneta equilibrata. Poniamoci il problema di trovare il valore atteso di  $X + Y$ .

---

<sup>1</sup>Per completezza – che potrebbe anche essere evitata – aggiungo qui l'interpretazione frequentista. I frequentisti pensano al valore atteso della variabile casuale  $Y$  come alla media di un enorme numero di realizzazioni della  $Y$ . Se si potesse esaminare un numero infinito di realizzazioni  $Y$ , allora la media di tali infiniti valori sarebbe esattamente uguale al valore atteso. Allora perché abbiamo bisogno di introdurre un concetto diverso da quello di “media”? La risposta è che la media aritmetica è una somma divisa per  $n$ :  $\bar{Y} = \frac{\sum_{i=1}^n y_i}{n}$ . Le variabili casuali possono generare un numero infinito di valori possibili. Dato che dividere per infinito non è possibile, è necessario procedere in un altro modo. Possiamo dire, in termini frequentisti, che il valore atteso di una variabile casuale è una *media ponderata* in cui il valore assegnato a ciascun evento elementare dello spazio campionario viene “pesato” per la sua probabilità di verificarsi, nel caso di variabili casuali discrete, o per la sua densità di probabilità, nel caso di variabili casuali continue. Il termine valore atteso è però un po' fuorviante. Infatti, esso potrebbe non corrispondere a nessuno dei valori che possono essere generati dalla variabile casuale. Quindi il valore atteso non è “atteso” nel senso che ci aspettiamo di vederlo comparire spesso. Ci aspettiamo invece che sia simile alla media di qualsiasi campione sufficientemente grande di realizzazioni della variabile casuale:  $\mathbb{E}(Y) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n y_i$ .

Per risolvere il problema iniziamo a costruire lo spazio campionario dell'esperimento casuale consistente nel lancio di un dado e di una moneta.

$x/y$	1	2	3	4	5	6
0	(0, 1)	(0, 2)	(0, 3)	(0, 4)	(0, 5)	(0, 6)
1	(1, 1)	(1, 2)	(1, 3)	(1, 4)	(1, 5)	(1, 6)

ovvero

$x/y$	1	2	3	4	5	6
0	1	2	3	4	5	6
1	2	3	4	5	6	7

Il risultato del lancio del dado è indipendente dal risultato del lancio della moneta. Pertanto, ciascun evento elementare dello spazio campionario avrà la stessa probabilità di verificarsi, ovvero  $Pr(\omega) = \frac{1}{12}$ . Il valore atteso di  $X + Y$  è dunque uguale a:

$$\mathbb{E}(X + Y) = 1 \cdot \frac{1}{12} + 2 \cdot \frac{1}{12} + \dots + 7 \cdot \frac{1}{12} = 4.0.$$

Lo stesso risultato si ottiene nel modo seguente:

$$\mathbb{E}(X + Y) = \mathbb{E}(X) + E(Y) = 3.5 + 0.5 = 4.0.$$

**Esercizio 1.2.** Si considerino le variabili casuali  $X$  e  $Y$  definite nel caso del lancio di tre monete equilibrate, dove  $X$  conta il numero delle teste nei tre lanci e  $Y$  conta il numero delle teste al primo lancio. Si calcoli il valore atteso del prodotto delle variabili casuali  $X$  e  $Y$ .

La distribuzione di probabilità congiunta  $P(X, Y)$  è fornita nella tabella seguente.

$x/y$	0	1	$p(Y)$
0	1/8	0	1/8
1	2/8	1/8	3/8
2	1/8	2/8	3/8
3	0	1/8	1/8
$p(y)$	4/8	4/8	1.0

Il calcolo del valore atteso di  $XY$  si riduce a

$$\mathbb{E}(XY) = 1 \cdot \frac{1}{8} + 2 \cdot \frac{2}{8} + 3 \cdot \frac{1}{8} = 1.0.$$

Si noti che le variabili casuali  $X$  e  $Y$  non sono indipendenti. Dunque non possiamo usare la proprietà ???. Infatti, il valore atteso di  $X$  è

$$\mathbb{E}(X) = 1 \cdot \frac{3}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8} = 1.5$$

e il valore atteso di  $Y$  è

$$\mathbb{E}(Y) = 0 \cdot \frac{4}{8} + 1 \cdot \frac{4}{8} = 0.5.$$

Dunque

$$1.5 \cdot 0.5 \neq 1.0.$$

## Variabili casuali continue

Nel caso di una variabile casuale continua  $Y$  il valore atteso diventa:

$$\mathbb{E}(Y) = \int_{-\infty}^{+\infty} yp(y)dy \quad (1.5)$$

Anche in questo caso il valore atteso è una media ponderata della  $y$ , nella quale ciascun possibile valore  $y$  è ponderato per il corrispondente valore della densità  $p(y)$ . Possiamo leggere l'integrale pensando che  $y$  rappresenti l'ampiezza delle barre infinitamente strette di un istogramma, con la densità  $p(y)$  che corrisponde all'altezza di tali barre e la notazione  $\int_{-\infty}^{+\infty}$  che corrisponde ad una somma.

Un'altra misura di tendenza centrale delle variabili casuali continue è la moda. La moda della  $Y$  individua il valore  $y$  più plausibile, ovvero il valore  $y$  che massimizza la funzione di densità  $p(y)$ :

$$\text{Mo}(Y) = \arg \max_y p(y). \quad (1.6)$$

## 1.2 Varianza

La seconda più importante proprietà di una variabile casuale, dopo che conosciamo il suo valore atteso, è la *varianza*.

**Definizione 1.2.** Se  $Y$  è una variabile casuale discreta con distribuzione  $p(y)$ , per definizione la varianza di  $Y$ ,  $\text{Var}(Y)$ , è

$$\text{Var}(Y) = \mathbb{E}[(Y - \mathbb{E}(Y))^2]. \quad (1.7)$$

A parole: la varianza è la deviazione media quadratica della variabile dalla sua media.<sup>2</sup> Se denotiamo  $\mathbb{E}(Y) = \mu$ , la varianza  $\text{Var}(Y)$  diventa il valore atteso di  $(Y - \mu)^2$ .

**Esempio 1.3.** Posta  $S$  uguale alla somma dei punti ottenuti nel lancio di due dadi equilibrati, poniamoci il problema di calcolare la varianza di  $S$ .

La variabile casuale  $S$  ha la seguente distribuzione di probabilità:

$s$	2	3	4	5	6	7	8	9	10	11	12
$P(S = s)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Essendo  $\mathbb{E}(S) = 7$ , la varianza diventa

$$\begin{aligned} \text{Var}(S) &= \sum (S - \mathbb{E}(S))^2 \cdot P(S) \\ &= (2 - 7)^2 \cdot 0.0278 + (3 - 7)^2 \cdot 0.0556 + \dots + (12 - 7)^2 \cdot 0.0278 \\ &= 5.8333. \end{aligned}$$

---

<sup>2</sup>Data una variabile casuale  $Y$  con valore atteso  $\mathbb{E}(Y)$ , le “distanze” tra i valori di  $Y$  e il valore atteso  $\mathbb{E}(Y)$  definiscono la variabile casuale  $Y - \mathbb{E}(Y)$  chiamata *scarto*, oppure *deviazione* oppure *variabile casuale centrata*. La variabile  $Y - \mathbb{E}(Y)$  equivale ad una traslazione di sistema di riferimento che porta il valore atteso nell'origine degli assi. Si può dimostrare facilmente che il valore atteso della variabile scarto  $Y - \mathbb{E}(Y)$  vale zero, dunque la media di tale variabile non può essere usata per quantificare la “dispersione” dei valori di  $Y$  relativamente al suo valore medio. Occorre rendere sempre positivi i valori di  $Y - \mathbb{E}(Y)$  e tale risultato viene ottenuto considerando la variabile casuale  $(Y - \mathbb{E}(Y))^2$ .

### Formula alternativa per la varianza

C'è un modo più semplice per calcolare la varianza:

$$\begin{aligned}\mathbb{E}[(X - \mathbb{E}(Y))^2] &= \mathbb{E}(X^2 - 2X\mathbb{E}(Y) + \mathbb{E}(Y)^2) \\ &= \mathbb{E}(Y^2) - 2\mathbb{E}(Y)\mathbb{E}(Y) + \mathbb{E}(Y)^2,\end{aligned}$$

dato che  $\mathbb{E}(Y)$  è una costante; pertanto

$$\text{Var}(Y) = \mathbb{E}(Y^2) - (\mathbb{E}(Y))^2. \quad (1.8)$$

A parole: la varianza è la media dei quadrati meno il quadrato della media.

**Esempio 1.4.** Consideriamo la variabile casuale  $Y$  che corrisponde al numero di teste che si osservano nel lancio di una moneta truccata con probabilità di testa uguale a 0.8. Il valore atteso di  $Y$  è

$$\mathbb{E}(Y) = 0 \cdot 0.2 + 1 \cdot 0.8 = 0.8.$$

Usando la formula tradizionale della varianza otteniamo:

$$\text{Var}(Y) = (0 - 0.8)^2 \cdot 0.2 + (1 - 0.8)^2 \cdot 0.8 = 0.16.$$

Lo stesso risultato si trova con la formula alternativa della varianza. Il valore atteso di  $Y^2$  è

$$\mathbb{E}(Y^2) = 0^2 \cdot 0.2 + 1^2 \cdot 0.8 = 0.8.$$

e la varianza diventa

$$\text{Var}(Y) = \mathbb{E}(Y^2) - (\mathbb{E}(Y))^2 = 0.8 - 0.8^2 = 0.16.$$

### Variabili casuali continue

Nel caso di una variabile casuale continua  $Y$ , la varianza diventa:

$$\text{Var}(Y) = \int_{-\infty}^{+\infty} [y - \mathbb{E}(Y)]^2 p(y) dy \quad (1.9)$$

Come nel caso discreto, la varianza di una v.c. continua  $y$  misura approssimativamente la distanza al quadrato tipica o prevista dei possibili valori  $y$  dalla loro media.

### 1.3 Deviazione standard

Quando lavoriamo con le varianze, i termini sono innalzati al quadrato e quindi i numeri possono diventare molto grandi (o molto piccoli). Per trasformare nuovamente i valori nell'unità di misura della scala originaria si prende la radice quadrata. Il valore risultante viene chiamato *deviazione standard* e solitamente è denotato dalla lettera greca  $\sigma$ .

**Definizione 1.3.** Si definisce scarto quadratico medio (o deviazione standard o scarto tipo) la radice quadrata della varianza:

$$\sigma_Y = \sqrt{\text{Var}(Y)}. \quad (1.10)$$

Interpretiamo la deviazione standard di una variabile casuale come nella statistica descrittiva: misura approssimativamente la distanza tipica o prevista dei possibili valori  $y$  dalla loro media.

**Esempio 1.5.** Per i dadi equilibrati dell'esercizio 1.3, la deviazione standard della variabile casuale  $S$  è uguale a  $\sqrt{5.8333} = 2.4152$ .

## 1.4 Standardizzazione

**Definizione 1.4.** Data una variabile casuale  $Y$ , si dice variabile standardizzata di  $Y$  l'espressione

$$Z = \frac{Y - \mathbb{E}(Y)}{\sigma_Y}. \quad (1.11)$$

Solitamente, una variabile standardizzata viene denotata con la lettera  $Z$ .

## 1.5 Momenti di variabili casuali

**Definizione 1.5.** Si chiama *momento* di ordine  $q$  di una v.c.  $X$ , dotata di densità  $p(x)$ , la quantità

$$\mathbb{E}(X^q) = \int_{-\infty}^{+\infty} x^q p(x) dx. \quad (1.12)$$

Se  $X$  è una v.c. discreta, i suoi momenti valgono:

$$\mathbb{E}(X^q) = \sum_i x_i^q p(x_i). \quad (1.13)$$

I momenti sono importanti parametri indicatori di certe proprietà di  $X$ . I più noti sono senza dubbio quelli per  $q = 1$  e  $q = 2$ . Il momento del primo ordine corrisponde al valore atteso o speranza matematica di  $X$ . Spesso i momenti di ordine superiore al primo vengono calcolati rispetto al valor medio di  $X$ , operando una traslazione  $x_0 = x - \mathbb{E}(X)$  che individua lo scarto dalla media. Ne deriva che il momento centrale di ordine 2 corrisponde alla varianza.

## 1.6 Funzione di ripartizione

Il concetto di funzione di ripartizione è molto importante nella teoria della probabilità, sia nel caso discreto, sia in quello continuo. L'insieme  $\{\omega : Y \leq y\}$  è un evento in  $\Omega$  e si può scrivere  $(Y \leq y)$ . A tale evento è possibile assegnare una probabilità  $P(Y \leq y)$  che, al variare di  $y \in \mathbb{R}$ , definisce la funzione di ripartizione della variabile casuale  $Y$ .

**Definizione 1.6.** Si chiama *funzione di ripartizione* o *funzione di distribuzione* della variabile casuale  $X$  la funzione  $F(X)$  definita da

$$F(X) \triangleq P(X \leq x), \quad x \in \mathbb{R}. \quad (1.14)$$

Detto a parole: la funzione di distribuzione cumulata, o funzione di ripartizione di  $X$ , misura la probabilità che  $X$  assuma valori minori o uguali al valore  $x$ .

La funzione di ripartizione è sempre non negativa, monotona non decrescente tra 0 e 1, tale che:

$$\lim_{x \rightarrow -\infty} F_x(X) = F_X(-\infty) = 0, \quad \lim_{x \rightarrow +\infty} F_x(X) = F_X(+\infty) = 1.$$

**Esempio 1.6.** Consideriamo l'esperimento casuale corrispondente al lancio di due monete. Sia  $X$  il numero di volte in cui esce testa. La distribuzione di probabilità di  $X$  è:

$$P(X) = \begin{cases} 0, & 0.25, \\ 1, & 0.50, \\ 2, & 0.25. \end{cases}$$



La funzione di ripartizione di  $X$  è:

$$F(X) = \begin{cases} 0, & \text{se } x < 0, \\ 1/4, & \text{se } 0 \leq x < 1, \\ 3/4, & \text{se } 1 \leq x < 2, \\ 1, & \text{se } 2 \leq x. \end{cases}$$

Il valore della funzione di ripartizione in corrispondenza di  $x = 1.5$ , ad esempio, è:

$$F(1.5) = P(X \leq 1.5) = P(X = 0) + P(X = 1) = \frac{1}{4} + \frac{2}{4} = \frac{3}{4}.$$



## Bibliografia

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



## 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.