UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

Dipartimento di Economia Marco Biagi

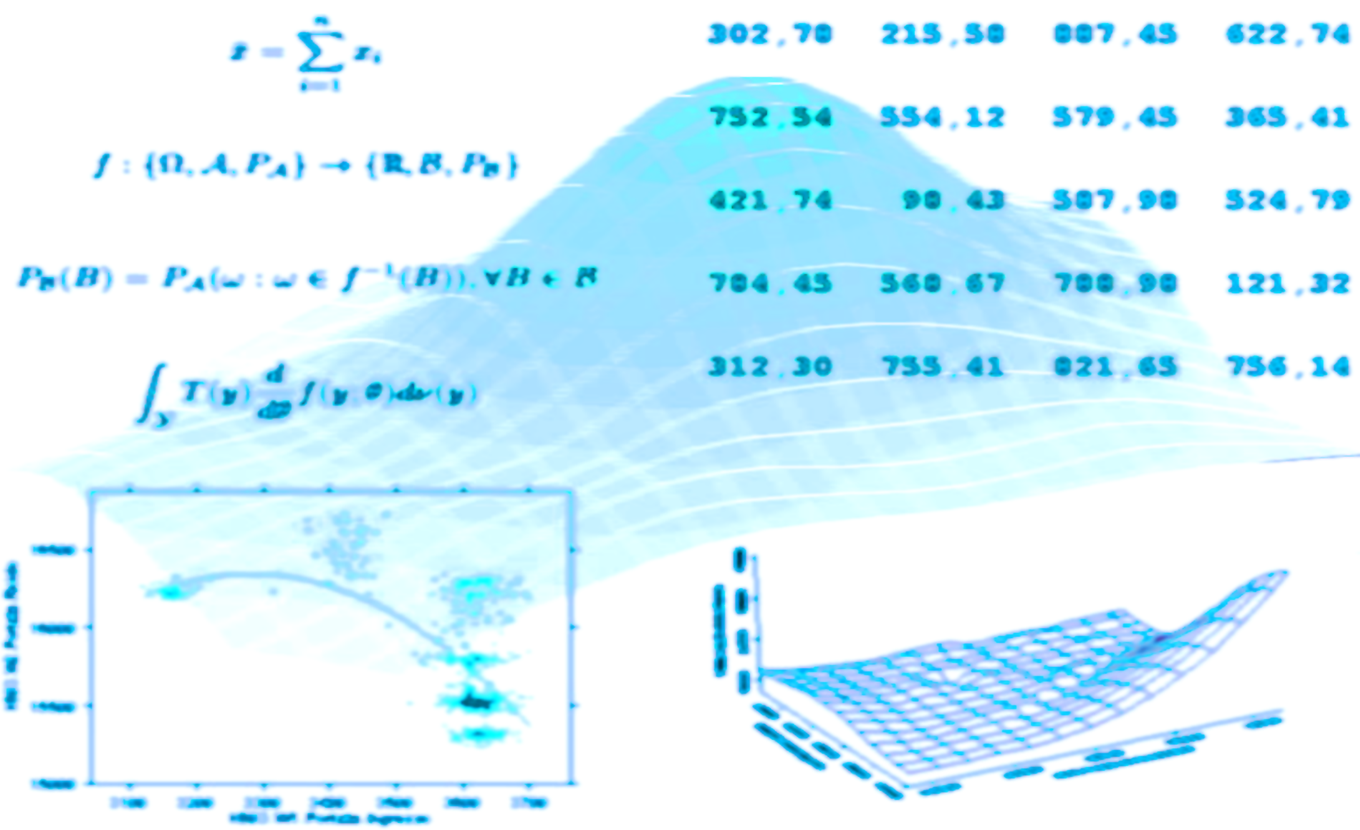
www.economia.unimore.it

Appunti di Statistica

CLEAM AA 24/25

Patrizio Frederic

Aggiornato al 07-04-2025



Indice

Avvertenza

Questo lavoro è un work in progress, questa non è la versione definitiva, sconsiglio di stampare tutto.

Introduzione

Durante il lock down della prima ondata, nel marzo del 2020, ho iniziato a trascrivere i miei appunti, raccolti negli ultimi 17 anni, in diapositive. Questa è una rielaborazione di quelle diapositive a cui è stato aggiunto parte dell'eserciziario realizzato con Michele Lalla. L'esposizione degli argomenti ricalca il corso di Statistica che con Michele Lalla abbiamo progettato per i corsi di CLEF e CLEA nel periodo che va dal 2008 al 2019, al quale ho portato alcune modifiche di forma e qualche aggiunta.

Questo manoscritto è diviso in tre parti, la prima dedicata alla statistica descrittiva nella quale vengono date le nozioni di base di descrizione di una popolazione, le sue rappresentazioni grafiche, la media e la variabilità. Nella seconda parte verranno date le nozioni di base di teoria e calcolo delle probabilità. Dal concetto di evento fino alle principali variabili casuali, il teorema del limite centrale e la distribuzione delle principali statistiche campionarie. Nella terza e ultima parte verranno sviluppati i principi di base di inferenza classica, dalla stima puntuale alla stima intervallare, la teoria dei test, il modello di regressione lineare e il test del chi-quadro.

Questo materiale, che non è ancora un libro, è da considerarsi il canovaccio d'appunti sul quale costruisco le mie lezioni. Non è da ritenersi alternativo alle lezioni in aula o ad un libro di testo. È ancora incompleto e disseminato di errori e imprecisioni.

Un ringraziamento speciale va alle mie donne e a Michele Lalla, amico, collega e maestro.

Bologna, il 07-04-2025.

Parte I
Statistica Descrittiva

I Fenomeni Collettivi

1

Scrivono Agresti, e Franklin, (2007), nel loro celebre libro:

> "Statistics is the art and science of designing studies and analyzing the data that those studies produce. Its ultimate goal is translating data into knowledge and understanding of the world around us. In short, **statistics is the art and science of learning from data**".
>
> La statistica è l'arte e la scienza di pianificare la raccolta e l'analisi dei dati che tali ricerche producono. Il suo fine è di trasformare i dati in conoscenza e comprensione del mondo circostante. In sintesi: **La statistica è l'arte e la scienza di imparare dai dati**.
>
> *Agresti, A., and Franklin, C. (2007), Statistics: the Art and Science of Learning from Data, Upper Saddle River, Pearson Prentice Hall.*

La **Statistica Descrittiva** si occupa di analizzare in modo meramente descrittivo, senza tentare di spiegare quale meccanismo li abbia generati. Di quello si occupa la Statistica Inferenziale nella terza parte del libro.

Definiremo dapprima cosa sono i dati, quindi 1.1, come si classificano 1.2 e come si raccolgono 1.4.

1.1 I Dati

In latino *datum* significa un fatto, quindi i *dati* sono collezioni di fatti. La Statistica ha come oggetto di studio i dati dalla loro definizione fino alla sintesi della conclusioni. Ovvero la statistica si occupa di definire i dati di interesse, di organizzarne la raccolta, di elaborarli e di sintetizzare le conclusioni.

I dati vengono raccolti su individui (**unità statistiche**), che sono accomunate per alcuni aspetti e presentano *variabilità* su altri, creando così un *fenomeno collettivo*. La Statistica è dunque la scienza che studia i dati che definiscono i fenomeni collettivi e che presentano forme di variabilità. La Statistica individua: concetti, metodi, e strumenti per la loro analisi.

Esempi di fenomeni collettivi che presentano variabilità:

- Reddito,
- Titolo di studio,
- Durata ricerca di occupazione,
- Preferenze di un consumatore,
- Durata di una malattia,
- Durata di una apparecchiatura

Il fenomeno è definito da concetti: caratteri o variabili (sinonimi). l'unita statistica è l'elemento su cui si osservano i caratteri oggetto di studio. una popolazione statistica o collettivo statistico e un insieme di unita statistiche omogenee rispetto a una o più caratteristiche o caratteri.

Esempio. Variabile = genere (M, F)

- Unità statistica = il singolo individuo in quest'aula
- Popolazione = gli studenti di quest'aula.

Esempio. Variabile = stato (difettoso, non difettoso)

- Unità = pezzo prodotto
- Popolazione = tutti i pezzi prodotti da settembre 2019

Esempio. Variabile = numero giorni di degenza

- Unità = individuo ricoverato
- Popolazione = tutti i ricoverati dell'ospedale XXX dal 2012 al 2020

Esempio. Variabile = kg di produzione

- Unità = ettaro coltivato con la varietà X
- Popolazione = tutti i possibili ettari coltivati con la varietà X

1.2 Variabili Statistiche

Ogni variabile statistica è caratterizzata dal numero e dal tipo di **modalità** che questa può assumere.

Esempi.

- Variabile genere: due sole possibili modalità {M, F}
- Variabile colore dei capelli: {Biondo, Castano, Rosso}
- Variabile titolo di studio: {Elementari, Medie, Superiori, Laurea, Post Laurea}
- Variabile numero di interventi: {0, 1, 2, 3, …}

Le variabili possono essere **Qualitative** o **Quantitative**.

- Le variabili **qualitative**:
  - possono essere *sconnesse* (o con scala nominale): le modalità possono essere solo uguali o diverse (né ordinate e né ordinabili). Ad esempio il genere, il tipo di titolo di studio, lo stato civile, la regione di provenienza, ecc-
  - Oppure *ordinate* (o con scala ordinale): le modalità sono ordinabili. Come per esempio: il titolo di studio, il livello di qualifica, le preferenze, ecc.
- Le variabili **quantitative** possono essere:
  - **Discrete**: le modalità possono avere una corrispondenza biunivoca con un sottoinsieme dei numeri interi. *Esempio*: numero di figli, eta , voto di laurea.
  - **Continue**: le modalità possono avere una corrispondenza biunivoca con un sottoinsieme dei numeri reali. *Esempio*: reddito, consumo, risparmio, altezza.
  - **Trasferibili**: se l'unità statistica può cedere tutto o parte del carattere posseduto a un'altra in modo sensato.

1.3 Popolazioni Statistiche

La **Popolazione Statistica** è l'insieme di tutte le unità che rispondono alla definizione. Una popolazione si dice

- **Finita** se si conosce il numero esatto delle sue unità
- **Infinita** se non si conosce il numero delle sue unità

Esempi di popolazione finita:

- Gli aventi diritto al voto
- Le imprese iscritte alla camera di commercio di Modena

Esempi di popolazione infinita:

- I consumatori della marca X
- Le aziende che hanno un gestionale con più di 5 anni

1.4 Le rilevazioni Statistiche

Si distinguono in due tipologie:

- **Sperimentali** utilizzate, in contesti scientifici, come Fisica, Medicina, Chimica e prevedono
  - Ipotesi di lavoro
  - Possibilità di controllo
- **Osservazionali** come le indagini di mercato, i sondaggi, ecc. nei quali non si ha possibilità di controllo.

L'obiettivo principale di un'indagine statistica è la **conoscenza di una popolazione obiettivo** o di riferimento ($\mathscr{P}$) intesa come insieme di unità elementari su cui si manifesta il fenomeno oggetto di studio. L'indagine è svolta quasi sempre su un campione, che è un sottoinsieme di $\mathscr{P}$, diversamente si avrebbe il **censimento**.

Per estrarre un campione occorre la **lista** di $\mathscr{P}$. La lista è l'elenco degli elementi appartenenti a $\mathscr{P}$ e rappresenta lo strumento principale per la scelta delle unità statistiche campionarie.

- **Censimento**: è una indagine completa perché esamina tutte le unità statistiche che compongono la popolazione oggetto di studio, $\mathscr{P}$.
- **Campionamento**: è una indagine parziale perché esamina solo un sottoinsieme, detto "campione", della popolazione oggetto di studio, $\mathscr{P}$.

1.4.1 Fasi dell'indagine

- definizione degli obiettivi,
- definizione delle unità e delle variabili da rilevare,
- scelta del periodo di riferimento.
- individuazione della popolazione e della lista delle unità statistiche.
- piano di campionamento
- raccolta dei dati,

- scelta della tecnica di rilevazione,
- formulazione del questionario e pretest,
- rilevazione sul campo.
- registrazione dei dati:
- controllo e correzione.
- elaborazione e analisi dei dati.

	Cens	Camp
Accuratezza delle Stime	Pro	Contro
su livelli territoriali piccoli	perfetta	alto rischio
		di non campionare
Esaustività	Pro	Contro
	sì	no
Costi	Contro	Pro
	Alti	Contenuti
Tempi di elaborazione	Contro	Pro
	Alti	Contenuti
Qualità dei dati	Contro	Pro
	Bassa	Alta
Quantità dei variabili	Contro	Pro
	Bassa	Alta

1.5 La matrice dei dati

La matrice dei dati è una tabella che consente di raccogliere in modo efficiente molti tipi diversi di dati.

i	Età	Sesso	Stato Civile	Titolo di Studio	Reddito x 1000€	Num. di Filgi
1	41	M	Non sposato	Laurea	10.23	2
2	20	F	Non sposato	Superiori	10.47	0
3	54	F	Sposato	Elementari	10.12	1
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	27	F	Non sposato	Laurea	10.07	0

Sulle RIGHE le UNITÀ STATISTICHE: si leggono le determinazioni dei caratteri oggetto di studio associati a una specifica unità statistica. Sulle COLONNE i CARATTERI: si leggono le modalità delle unità statistiche associate a uno specifico carattere.

1.6 Riepilogo sulle Variabili

- **Qualitativa**, la variabile è espressa attraverso etichette qualitative
  - *Qualitative sconnesse*: le caratteristiche che la VS può assumere hanno un ordinamento soggettivo;
    - genere,
    - stato civile,
    - settore di occupazione,
    - generi musicali.
  - *Qualitative ordinate*: le caratteristiche che la VS può assumere hanno un ordinamento oggettivo
    - titolo di studio,
    - preferenze,
    - giudizi.
- **Quantitativa**, la variabile è espressa attraverso una scala numerica.
  - *Quantitative Discrete*: le caratteristiche che la VS può assumere sono in numero finito al più numerabile $\rightarrow$ corrispondenza con i numeri interi;
    - numero di incidenti,
    - voto di laurea.
  - *Quantitative Continue*: le caratteristiche che la VS può assumere sono in numero infinito non numerabile.
    - misure di lunghezza, capienza e peso,
    - temperature,
    - reddito.

Variabili Statistiche e Distribuzioni di Frequenza

2

La statistica si fonda sulla necessità di comprendere e descrivere fenomeni complessi in modo sistematico e universale. Per farlo, è indispensabile adottare un linguaggio rigoroso che consenta di astrarre dai dettagli specifici e lavorare direttamente sulle relazioni tra i dati. La formalizzazione non è un vezzo, ma una necessità: permette di trattare insiemi di dati di qualunque dimensione o complessità senza cambiare le regole del ragionamento.

Il linguaggio formale che utilizzeremo in questo libro è pensato per garantire una sintesi efficace e una generalizzazione immediata. Una manciata di osservazioni e un archivio di miliardi di righe possono essere analizzati con gli stessi strumenti e le stesse regole. Questa uniformità concettuale è ciò che rende il pensiero statistico potente: non ci si perde nei dettagli numerici, ma si lavora direttamente sulle proprietà e sulle trasformazioni delle strutture che i dati rappresentano.

Adottare una notazione rigorosa consente inoltre di comunicare con chiarezza e precisione. Ogni simbolo, ogni operatore ha un significato ben definito che elimina ambiguità e rende possibile il ragionamento collaborativo, indipendentemente dal contesto applicativo. La formalizzazione non è solo un linguaggio per il calcolo, ma uno strumento per pensare, per astrarre e per collegare il particolare all'universale.

Lavorare in questo modo significa concentrarsi sui concetti fondamentali, svincolandosi dalla necessità di riferirsi continuamente alla numerosità dei dati o alla loro specifica natura. È così che il linguaggio statistico diventa non solo un mezzo per analizzare, ma anche per costruire modelli e sviluppare inferenze che si applicano a fenomeni ben oltre i dati osservati.

In statistica, la scelta dei simboli gode di una flessibilità maggiore rispetto ad altre discipline come la fisica, dove molte notazioni sono rigidamente standardizzate. Questa libertà riflette la natura più applicativa e interdisciplinare della statistica, che si adatta a contesti diversi e a una grande varietà di fenomeni. Tuttavia, nel tempo, si è consolidata una notazione comune, grazie alla diffusione di libri di testo e alla pratica accademica, che offre un riferimento condiviso pur lasciando margini per piccole personalizzazioni.

In fisica, al contrario, cambiare i simboli può portare a confusione o addirittura compromettere la comprensione. Ad esempio, la costante gravitazionale universale G e la velocità della luce nel vuoto c sono simboli universalmente riconosciuti. Rinominare G con un altro simbolo, come k, o c con v, rischierebbe di generare fraintendimenti, poiché quei simboli sono già associati ad altri concetti fondamentali, come la costante elastica e la velocità generica di un corpo.

In statistica, invece, sebbene esistano convenzioni consolidate – come indicare una variabile casuale con una lettera maiuscola (X) e un'osservazione specifica con una minuscola (x_i) – la comunità accetta varianti ragionevoli, purché siano chiaramente definite. In questo libro, useremo una notazione che rispetta gli standard più diffusi, integrandola con piccole personalizzazioni pensate per migliorare la chiarezza e la leggibilità, senza mai perdere il rigore. Questo approccio garantisce che il linguaggio sia al tempo stesso accessibile e conforme alle convenzioni accademiche, facilitando il collegamento con altri testi e contesti di studio.

2.1 Variabili Statistiche

Una *Variabile Statistica* (VS) è una qualunque caratteristica osservabile sugli individui (unità statistiche) della popolazione di riferimento, che *varia* da individuo ad individuo.

2.1.1 Notazione di Base

- $\mathbf{x} = (x_1, x_2, ..., x_i, ..., x_n)$, etichette simboliche per i dati, il primo dato osservato, il secondo ecc.
- i, indice che conta le osservazioni nell'ordine in cui sono state osservate
  - $i \in \{1, 2, ..., n\}$
  - il primo, il secondo, … l'i-esimo, … l'n-esimo (l'ultimo)
- n, numerosità assoluta: il numero totale di individui osservati.
- $S_X = \{\mathrm{x}_1, ..., \mathrm{x}_j, ..., \mathrm{x}_K\}$, l'insieme di tutte le modalità possibili che la variabile statistica è suscettibile di assumere.
- j, indice che conta le modalità: prima, seconda, …, j-esima,…, la K-esima.
- K, numero di modalità.

Esempio 2.1.1 (Variabile: genere)**.**

- $\mathbf{x} = (x_1 = M, x_2 = F, x_3 = M, x_4 = F, x_5 = F, x_6 = F)$
- $n = 6$
- $S_X = \{\mathrm{x}_1 = F, \mathrm{x}_2 = M\}$
- $K = 2$

Esempio 2.1.2 (Variabile: titolo di studio)**.**

- $\mathbf{x} = (x_1 = E, x_2 = M, x_3 = L, x_4 = S, x_5 = S, x_6 = S, x_7 = L, x_8 = M, x_9 = L, x_{10} = S)$
- $n = 10$
- $S_X = \{\mathrm{x}_1 = E, \mathrm{x}_2 = M, \mathrm{x}_3 = S, \mathrm{x}_4 = L\}$
- $K = 4$

Esempio 2.1.3 (Variabile: Numero di interventi di manutenzione giornalieri)**.**

- $\mathbf{x} = (x_1 = 0, x_2 = 1, x_3 = 0, x_4 = 2, x_5 = 1, x_6 = 1, x_7 = 0, x_8 = 1, x_9 = 3, x_{10} = 1)$
- $n = 10$
- $S_X = \{\mathrm{x}_1 = 0, \mathrm{x}_2 = 1, \mathrm{x}_3 = 2, \mathrm{x}_4 = 3, ...\}$
- $K = +\infty$

2.1.2 Ordinamento e conteggio

Se l'ordine di osservazione non è influente ai fini della conoscenza del fenomeno i dati possono essere permutati (mescolati) a piacimento, la sequenza:

$$(x_{(1)}, x_{(2)}, ..., x_{(i)}, ..., x_{(n)}),$$

indica i dati riordinati, dal più piccolo, al più grande. Se i dati sono numerici l'ordinamento è univoco, se i dati sono categoriali l'ordinamento è arbitrario.

Esempio 2.1.4 (Continua)**.** Continuiamo l'esempio della variabile genere discussa nell'esempio 2.1.1

- $\mathbf{x} = (x_1 = M, x_2 = F, x_3 = M, x_4 = F, x_5 = F, x_6 = F)$
- $S_X = \{\mathrm{x}_1 = F, \mathrm{x}_2 = M\}$
- $x_{(1)} = F, x_{(2)} = F, x_{(3)} = F, x_{(4)} = F, x_{(5)} = M, x_{(6)} = M$

Esempio 2.1.5 (Continua: codifica 0, 1)**.**

- Variabile: genere {M -> 0,F-> 1}
- $\mathbf{x} = (x_1 = 0, x_2 = 1, x_3 = 0, x_4 = 1, x_5 = 1, x_6 = 1)$
- $S_X = \{\mathrm{x}_1 = 0, \mathrm{x}_2 = 1\}$
- $x_{(1)} = 0, x_{(2)} = 0, x_{(3)} = 1, x_{(4)} = 1, x_{(5)} = 1, x_{(6)} = 1$
- **nota**: nella codifica 0, 1 ha senso sommare i dati:

$$x_1 + x_2 + x_3 + x_4 + x_6 = 4, \quad \text{Numero di femmine}$$

Esempio 2.1.6 (Continua: Variabile titolo di studio)**.**

- $\mathbf{x} = (x_1 = E, x_2 = M, x_3 = L, x_4 = S, x_5 = S, x_6 = S, x_7 = L, x_8 = M, x_9 = L, x_{10} = S)$
- $S_X = \{\mathrm{x}_1 = E, \mathrm{x}_2 = M, x_3 = S, x_4 = L\}$
- $x_{(1)} = E, x_{(2)} = M, x_{(3)} = M, x_{(4)} = S, x_{(5)} = S, x_{(6)} = S, x_{(7)} = S, x_{(8)} = L, x_{(9)} = L, x_{(10)} = L$

Esempio 2.1.7 (Continua: Codifica Numerica)**.**

- Variabile: titolo di studio {E -> 1, M -> 2, S -> 3, L -> 4}
- $\mathbf{x} = (x_1 = 1, x_2 = 2, x_3 = 4, x_4 = 3, x_5 = 3, x_6 = 3, x_7 = 4, x_8 = 2, x_9 = 4, x_{10} = 3)$
- $S_X = \{\mathrm{x}_1 = 1, \mathrm{x}_2 = 2, x_3 = 3, x_4 = 4\}$
- $x_{(1)} = 1, x_{(2)} = 2, x_{(3)} = 2, x_{(4)} = 3, x_{(5)} = 3, x_{(6)} = 3, x_{(7)} = 3, x_{(8)} = 4, x_{(9)} = 4, x_{(10)} = 4$
- ha senso sommare i dati?

La codifica numerica corretta sarebbe più complessa

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	Tot
E	1	0	0	0	0	0	0	0	0	0	1
M	0	1	0	0	0	0	0	1	0	0	2
S	0	0	0	1	1	1	0	0	0	1	4
L	0	0	1	0	0	0	1	0	1	0	3

I totali di colonna hanno senso e indicano il numero di individui che ha un determinato titolo di studio.

Questa codifica è sovra abbondante infatti come per maschio e femmina possiamo contare solo un colonna di presenza 1 (è femmina) e assenza 0 (non è femmina e quindi maschio), per una variabile a 4 modalità possiamo contare solo 3, ad esempio

- 0,0,0 elementari
- 1,0,0 medie
- 0,1,0 superiori
- 0,0,1 università

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	Tot
M	0	1	0	0	0	0	0	1	0	0	2
S	0	0	0	1	1	1	0	0	0	1	4
L	0	0	1	0	0	0	1	0	1	0	3

Per sapere il numero persone che ha al massimo le elementari basta fare 10 (numero totale di individui) meno 2 (medie) più 4 (superiori) più 3 (laureati):

$$10 - (2 + 4 + 3) = 1 \quad \text{con le elemntari}$$

Esempio 2.1.8 (Continua: Numero di interventi di manutenzione giornalieri)**.**

- $\mathbf{x} = (x_1 = 0, x_2 = 1, x_3 = 0, x_4 = 2, x_5 = 1, x_6 = 1, x_7 = 0, x_8 = 1, x_9 = 3, x_{10} = 1)$
- $S_X = \{\mathrm{x}_1 = 0, \mathrm{x}_2 = 1, \mathrm{x}_3 = 2, \mathrm{x}_4 = 3, ...\}$
- $x_{(1)} = 0, x_{(2)} = 0, x_{(3)} = 0, x_{(4)} = 1, x_{(5)} = 1, x_{(6)} = 1, x_{(7)} = 1, x_{(8)} = 1, x_{(9)} = 2, x_{(10)} = 3$
- ha senso sommare i dati?
- cosa rappresenta la somma dei dati?

Esempio 2.1.9 (Ore uomo dedicate a interventi di manutenzione)**.**

Supponiamo di aver collezionato il numero di ore uomo (e frazioni di ora) dedicate ad ogni intervento di manutenzione L'unità statistica sarà l'intervento, $i = 1$ il primo, $i = 2$, il secondo, ecc. e assumerà un valore decimale $x_3 = 3.5$ significa che la terza manutenzione ha impiegato un addetto per 3 ore e mezza.

- $\mathbf{x} = (x_1 = 0.4, x_2 = 2.7, x_3 = 3.5, x_4 = 1.4, x_5 = 4.3, x_6 = 4.6, x_7 = 0.2, x_8 = 1.9, x_9 = 3.4, x_{10} = 0.1)$
- $S_X = \{\mathrm{x}_1 = 0.0, \mathrm{x}_2 = 0.1, \mathrm{x}_3 = 0.2, \mathrm{x}_4 = 0.3, ...\}$
- $x_{(1)} = 0.1, x_{(2)} = 0.2, x_{(3)} = 0.4, x_{(4)} = 1.4, x_{(5)} = 1.9, x_{(6)} = 2.7, x_{(7)} = 3.4, x_{(8)} = 3.5, x_{(9)} = 4.3, x_{(10)} = 4.6$
- ha senso sommare i dati?
- cosa rappresenta la somma dei dati?

2.1.3 Le unità di misura

In statistica, ogni dato è sempre espresso in una specifica unità di misura, che conferisce significato al valore numerico registrato. Le unità di misura variano in base alla natura dei dati e al fenomeno osservato, e possono includere:

1. **Conteggi**: I dati che rappresentano quantità possono essere espressi in unità di conteggio, come unità singole, decine, centinaia, migliaia, ecc. Ad esempio, se registriamo il numero di visitatori in un parco in un anno, può essere più comodo esprimere 12.000 visitatori come "12 migliaia" (12 x 1000), facilitando la lettura. Analogamente, 1.200 prodotti venduti potrebbero essere espressi come "1,2 migliaia" per rendere i confronti immediati tra dati di grandezza diversa.

2. **Misure metriche**: Le grandezze fisiche come la lunghezza, la massa, la capacità e la temperatura sono espresse in unità metriche, come metri, chilogrammi, e litri. Ad esempio, per confrontare il peso di vari articoli, un dato in grammi potrebbe risultare scomodo se molto elevato; per articoli pesanti, come 12 000 grammi, potrebbe essere più leggibile esprimere il dato in chilogrammi come 12 kg, dove 1 kg = 1 000 grammi. Questo cambio di unità mantiene invariato il rapporto tra le osservazioni, facilitando la comprensione e l'analisi.

3. **Misure derivate o di rapporto**: Alcune unità di misura rappresentano rapporti tra grandezze, come chilometri all'ora (km/h) per la velocità, o litri per 100 chilometri (L/100 km) per il consumo di carburante. Ad esempio, se un'auto consuma 7 litri ogni 100 km, esprimere il consumo in questa unità (7 L/100 km) è più immediato e rappresentativo di un valore medio, rispetto all'uso di litri per singolo chilometro (0.07 L/km), che potrebbe risultare poco intuitivo.

4. **Misure di risultati e punteggi**: In statistica, si usano spesso trasformazioni che combinano più misure per produrre un indicatore con una scala di misura propria, utile per confrontare sinteticamente informazioni complesse. Consideriamo la seguente formula generale:

   $$x_i = f(a_{i1}, a_{i2}, a_{i3}, \dots)$$

   dove a_{ij} rappresenta le varie misure o statistiche che contribuiscono al calcolo del valore x_i, cioè l'indicatore della i-esima osservazione. Per esempio, immaginiamo di voler calcolare un **punteggio di idoneità** per i candidati a un lavoro, basato su tre misure diverse:

   - a_{i1}: esperienza in anni,
   - a_{i2}: punteggio di un test tecnico (da 0 a 100),
   - a_{i3}: valutazione del colloquio (da 0 a 5).

Per produrre un indicatore che misuri l'idoneità complessiva su una scala uniforme, potremmo usare una formula come:

$$x_i = 0.5 \cdot a_{i1} + 0.3 \cdot a_{i2} + 2 \cdot a_{i3}$$

Questa formula pondera ciascuna misura per dare maggiore peso all'esperienza (0.5) e al colloquio (2) rispetto al test tecnico (0.3). Per il quinto candidato ($i = 5$) con 5 anni di esperienza, un punteggio di 80 al test e una valutazione di 4 al colloquio, il punteggio di idoneità risulterebbe:

$$x_5 = 0.5 \cdot 5 + 0.3 \cdot 80 + 2 \cdot 4 = 2.5 + 24 + 8 = 34.5$$

In questo caso, il punteggio di idoneità x_i è su una scala propria, definita dalle scelte di peso e dalle unità combinate, e consente di confrontare i candidati in modo sintetico, anche se ogni componente ha un'unità di misura diversa.

2.1.4 Trasformazioni lineari

In matematica e statistica, una trasformazione lineare è un'operazione che modifica i dati mantenendo invariati i rapporti tra le osservazioni. La forma generale di una trasformazione lineare è:

$$y_i = a + bx_i$$

dove:

- x_i rappresenta la **misura i-esima** del dato originale.
- a rappresenta uno **spostamento dell'origine** (traslazione), che posiziona i dati su un nuovo punto di partenza, spostando l'intero grafico verso l'alto o verso il basso.
- b è il **fattore di scala**, che ridimensiona i valori, influenzando la pendenza della retta. Un fattore maggiore di 1 espande i valori, mentre uno tra 0 e 1 li contrae.

Geometricamente, il termine a trasla il grafico lungo l'asse verticale senza modificare le distanze relative tra i punti, mentre b varia la pendenza, allargando o comprimendo i dati rispetto all'origine.

2.1.4.1 Cambiamento di scala come trasformazione lineare

Un cambiamento di scala è una trasformazione lineare applicata a ogni misura i-esima del dato, che permette di adattare l'unità di misura senza alterare i rapporti tra le osservazioni. Vediamo alcuni esempi pratici.

Esempio 2.1.10 (Dati spese). Supponiamo di avere i dati su diverse spese in euro e di volerli esprimere in migliaia di euro per facilitare la lettura. La trasformazione è:

$$y_i = \frac{1}{1000} \cdot x_i$$

Dove x_i è il valore in euro della misura i-esima e y_i è il valore corrispondente in migliaia di euro. Per chiarire questo concetto, prendiamo $i = 4$ con $x_4 = 12\ 000$:

$$y_4 = \frac{1}{1000} \cdot 12\ 000 = 12$$

Quindi, invece di esprimere la quarta misura come 12 000 euro, la rappresentiamo come 12 *migliaia* di euro, semplificando la gestione e la comparabilità dei dati.

Esempio 2.1.11 (Conversione temperature). Anche la conversione da gradi Celsius (C) a gradi Fahrenheit (F) segue una trasformazione lineare, espressa come:

$$F_i = 32 + \frac{9}{5} \cdot C_i$$

Qui, il termine 32 rappresenta uno **spostamento dell'origine**, poiché 0 °C corrisponde a 32 °F, mentre il fattore $\frac{9}{5}$ espande i valori per adattarsi alla scala Fahrenheit. Ad esempio, se $i = 2$ e $C_2 = 25$:

$$F_2 = 32 + \frac{9}{5} \cdot 25 = 32 + 45 = 77$$

Questa trasformazione mantiene la proporzionalità tra le temperature, adattando la scala e le unità di misura.

Le trasformazioni lineari $y = a + bx$ mantengono le distanze relative tra i dati poiché ogni differenza tra due valori x_i e x_j viene amplificata o ridotta di un fattore costante b. In altre parole, per ogni coppia di dati x_i, x_j vale la relazione:

$$y_i - y_j = b(x_i - x_j)$$

Quindi, la proporzionalità tra le differenze dei valori originali si conserva, preservando l'ordine e la struttura relativa dei dati.

In termini più intuitivi immaginate di avere 5 punti distribuiti su una retta, rappresentati da valori numerici come coordinate lungo un asse x. Se si cambia l'unità di misura (ad esempio passando da metri a centimetri) o se si decide di nascondere i numeri, la loro disposizione relativa sulla retta rimane la stessa agli occhi di chi osserva: le distanze tra i punti e l'ordine dei punti non cambiano.

Questo accade perché una trasformazione lineare del tipo $y = a + bx$ modifica solo la scala e la posizione dell'intero sistema, ma non altera le proporzioni tra le distanze dei punti. È come osservare un disegno su un foglio: se ingrandisci o rimpicciolisci il foglio, le figure disegnate mantengono le stesse relazioni interne tra di loro.

All'occhio, quindi, i punti "restano uguali" perché la trasformazione lineare agisce uniformemente su tutti i punti, preservando le loro distanze relative e l'allineamento lungo la retta.

2.1.4.2 Rappresentazione grafica delle trasformazioni lineari

Come abbiamo visto una trasformazione lineare è data

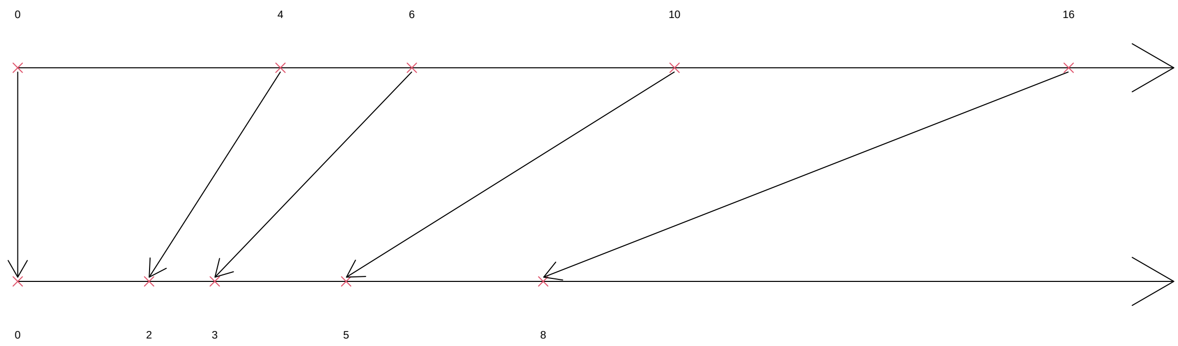
$$y_i = a + b \cdot x_i, \forall i = 1, ..., n$$

caso 1: $a = 0$**,** $b < 1$

I dati si contraggono, lo zero resta invariato Per esempio: supponiamo di avere osservato ($x_1 = 0, x_2 = 4, x_3 = 6, x_4 = 10, x_5 = 16$), e fissato $b = 1/2$

$$y_i = \frac{1}{2} \cdot x_i, \forall i = 1, ..., 5$$

allora $(x_1 = 0, x_2 = 2, x_3 = 3, x_4 = 5, x_5 = 8)$. Geometricamente



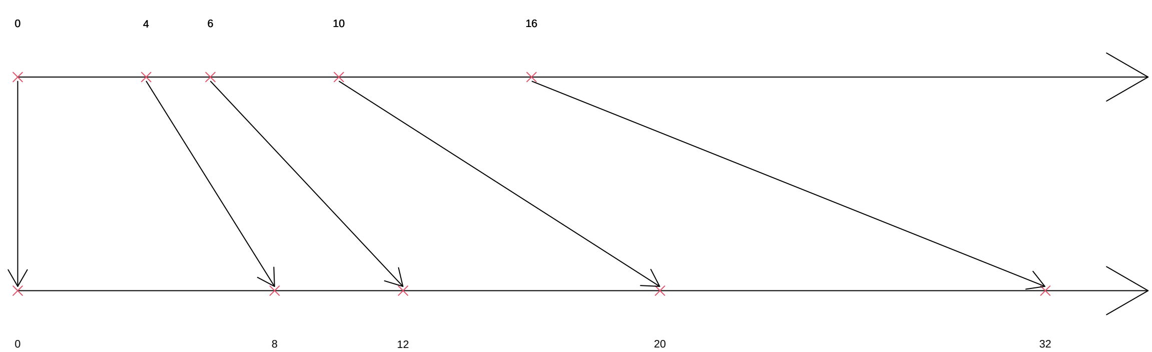
Osserviamo che i dati si restringono, i dati originari vanno da 0 a 16, quelli trasformati da 0 a 8. Notiamo che lo zero resta invariato.

caso 2: $a = 0$, $b > 1$

I dati si espandono, lo zero resta invariato. Per esempio: supponiamo di avere osservato $(x_1 = 0, x_2 = 4, x_3 = 6, x_4 = 10, x_5 = 16)$, e fissato $b = 2$

$$y_i = 2 \cdot x_i, \forall i = 1, ..., 5$$

allora $(x_1 = 0, x_2 = 8, x_3 = 12, x_4 = 20, x_5 = 32)$. Geometricamente



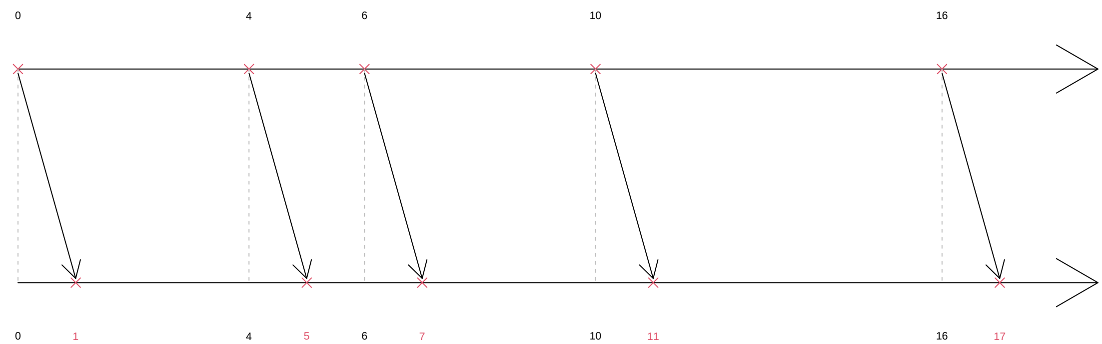
Osserviamo che i dati si espandono, i dati originari vanno da 0 a 16, quelli trasformati da 0 a 32. Notiamo che lo zero resta invariato.

caso 3: $a \neq 0$, $b = 1$

Le distanze tra i dati restano invariate, lo zero cambia. Per esempio: supponiamo di avere osservato $(x_1 = 0, x_2 = 4, x_3 = 6, x_4 = 10, x_5 = 16)$, e fissato $a = +1$

$$y_i = 1 + x_i, \forall i = 1, ..., 5$$

allora $(x_1 = 1, x_2 = 5, x_3 = 7, x_4 = 11, x_5 = 17)$. Geometricamente



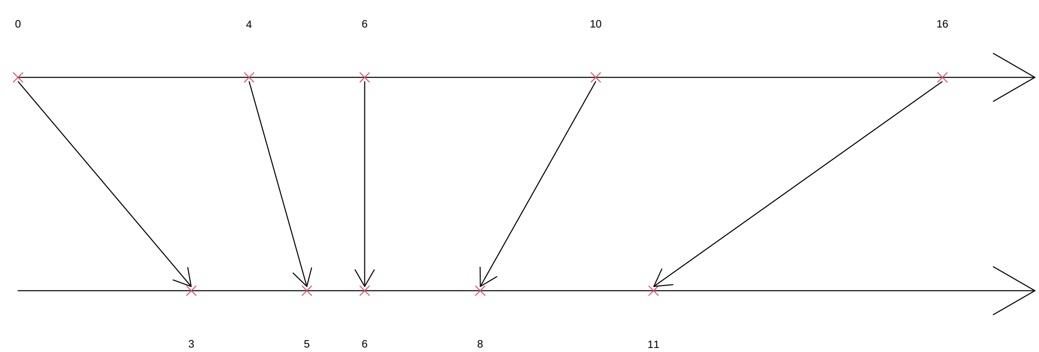
Osserviamo che i dati sono stati traslati a destra di uno, i dati originari vanno da 0 a 16, quelli trasformati da 1 a 17. Notiamo che lo zero è cambiato nella nuova scala.

caso 4: $a \neq 0$**,** $b \neq 1$

Le distanze tra i dati restano invariate, lo zero cambia. Per esempio: supponiamo di avere osservato $(x_1 = 0, x_2 = 4, x_3 = 6, x_4 = 10, x_5 = 16)$, e fissato $a = +1$ e $b = 1/2$

$$y_i = 3 + \frac{1}{2} \cdot x_i, \forall i = 1, ..., 5$$

allora $(x_1 = 3, x_2 = 5, x_3 = 6, x_4 = 8, x_5 = 11)$. Geometricamente



Osserviamo che i dati sono stati traslati a destra di uno, i dati originari vanno da 0 a 16, quelli trasformati da 3 a 11. Notiamo che lo zero è cambiato nella nuova scala.

2.1.4.3 Scelta dei coefficienti

A volte i coefficienti a e b vengono usati per modificare a piacimento il massimo e il minimo dei dati, per trasformare, ad esempio, punteggi in centesimi in punteggi in trentesimi.

Se i dati originari esistono nell'intervallo $[x_{\min}, x_{\max}]$, ovvero tutti i dati $x_{\min} \leq x_i \leq x_{\max}$ e vogliamo trasformarli in dati y_i tali che $y_{\min} \leq y_i \leq y_{\max}$, dove $y_{\min}$ e $y_{\max}$ sono fissati arbitrariamente. Se imponiamo

$$\begin{cases} y_{\min} = a + b \cdot x_{\min} \\ y_{\max} = a + b \cdot x_{\max} \end{cases}$$

Avremo:

$$b = \frac{y_{\max} - y_{\min}}{x_{\max} - x_{\min}}$$

e quindi:

$$a = y_{\min} - \frac{y_{\max} - y_{\min}}{x_{\max} - x_{\min}} \cdot x_{\min}$$

in definitiva

$$y = \left(y_{\min} - \frac{y_{\max} - y_{\min}}{x_{\max} - x_{\min}} \cdot x_{\min} \right) + \frac{y_{\max} - y_{\min}}{x_{\max} - x_{\min}} \cdot x$$

Esempio 2.1.12 (Conversione punteggi tra 0 e 50)**.** Supponiamo di aver somministrato un questionario con 25 domande, che assegnano punteggio 0 se la risposta è sbagliata, +1 se la risposta è parzialmente corretta e +2 se la risposta è corretta. Sia x_i la somma dei punteggi dello studente etichettato con i, che potrà:

1. sbagliarle tutte $x_i = 0 \cdot 25 = 0$
2. rispondere correttamente a tutte $x_i = 2 \cdot 25 = 50$.

Il punteggio originario varia in una scala tra 0 e 50.

Per trasformare il punteggio in trentesimi faremo

$$y_i = \left(0 - \left(\frac{30-0}{50-0} \right) \cdot 0 \right) + \left(\frac{30-0}{50-0} \right) x_i = \frac{3}{5} x_i$$

Il nuovo punteggio varrà quindi

1. se tutte le domande sono sbagliate $x_i = 0 \to y_i = 0$
2. se tutte le domande sono corrette $x_i = 50 \to y_i = 30$

Il punteggio trasformato varia tra 0 e 30. Se per esempio lo studente 3 ha ottenuto il punteggio $x_3 = 43$ avrà un voto in trentesimi pari a

$$y_3 = \frac{30}{50} \cdot 43 = 25.8$$

Esempio 2.1.13 (Conversione punteggi tra -25 e 25). Supponiamo di aver somministrato un questionario con 15 domande, che assegnano punteggio -1 se la risposta è sbagliata, 0 se la risposta non è data e $+1$ se la risposta è corretta

Sia x_i la somma dei punteggi dello studente etichettato con i, che potrà:

1. sbagliarle tutte $x_i = -1 \cdot 25 = -25$
2. rispondere correttamente a tutte $x_i = 1 \cdot 25 = 25$

Per trasformare il punteggio in trentesimi faremo

$$y_i = \left(0 - \left(\frac{30-0}{50-0}\right) \cdot (-25)\right) + \left(\frac{30-0}{50-0}\right) x_i = 15 + \frac{3}{5} x_i$$

Il nuovo punteggio varrà

1. se tutte le domande sono sbagliate $x_i = -25 \rightarrow y_i = 15 + 3/5 \cdot (-25) = 0$
2. se tutte le domande sono corrette $x_i = +25 \rightarrow y_i = 15 + 3/5 \cdot (+25) = 30$

Il punteggio trasformato varia tra 0 e 30. Se per esempio lo studente 3 ha ottenuto il punteggio $x_3 = 21$ avrà un voto in trentesimi pari a

$$y_3 = 15 + \frac{3}{5} \cdot 43 = 27.6$$

2.2 Distribuzione di Frequenza

La frequenza indica quanto una modalità insiste sul collettivo. Le frequenze si dividono in:

Definizione 2.2.1 (Frequenze Assolute). Si definiscono le n_j le **frequenze assolute**: il numero di individui che presentano la modalità j.

Definizione 2.2.2 (Frequenze Relative). Si definiscono le $f_j = n_j/n$ le **frequenze relative**: la proporzione di individui che presentano la modalità j.

Definizione 2.2.3 (Frequenze Percentuali). Si definiscono le $f_{\%j} = f_j \times 100$ le **frequenze percentuali**: la percentuale di individui che presentano la modalità j.

Proprietà 2.2.1. *Le proprietà della frequenze assolute(n_j) sono:*

- $0 \leq n_j \leq n, \forall j = 1, ..., K,$
- $\sum_{j=1}^{K} n_j = n.$

Proprietà 2.2.2. *Le proprietà della frequenze relative (f_j) sono:*

- $0 \leq f_j \leq 1, \forall j = 1, ..., K,$
- $\sum_{j=1}^{K} f_j = 1.$

Proprietà 2.2.3. *Le proprietà della frequenze percentuali ($f_{\%j}$) sono:*

- $0 \leq f_{\%,j} \leq 100, \forall j = 1, ..., K,$
- $\sum_{j=1}^{K} f_{\%,j} = 100.$

Definizione 2.2.4 (Distribuzione di Frequenza). Una **distribuzione di frequenza** è una tabella a cui vengono associate le modalità e le frequenze

Esempio 2.2.1 (Continua: Variabile: genere).

$$x_{(1)} = F, x_{(2)} = F, x_{(3)} = F, x_{(4)} = F, x_{(5)} = M, x_{(6)} = M$$

X	n_j	f_j	$f_{\%j}$
F	4	$4/6 = 0.67$	67%
M	2	$2/6 = 0.33$	33%
Tot	6	1.00	100%

Esempio 2.2.2 (Continua: Variabile titolo di studio). Consideriamo i 10 dati ordinati: $x_{(1)} = E, x_{(2)} = M, x_{(3)} = M, x_{(4)} = S, x_{(5)} = S\ , x_{(6)} = S, x_{(7)} = S, x_{(8)} = L, x_{(9)} = L, x_{(10)} = L$ la loro tabella di frequenza è

X	n_j	f_j	$f_{\%j}$
E	1	$1/10 = 0.1$	10%
M	2	$2/10 = 0.2$	20%
S	4	$4/10 = 0.4$	40%
L	3	$3/10 = 0.3$	30%
Tot	10	1.0	100%

2.2.1 Dati quantitativi continui

Se i dati sono quantitativi continui il numero delle modalità è spesso di gran lunga superiore al numero dei dati e non sempre è possibile fissare un limite superiore in anticipo all'osservazione dei dati. Se per esempio volessi misurare il reddito di una persona in centesimi, otterrei:

Esempio 2.2.3 (Variabile: Reddito mensile lordo in migliaia di euro).

- unità di rilevazione: famiglie del comune A a febbraio 2021

Tabella 2.1: A sinistra abbiamo troppe poche classi, si perde troppa variabilità. Al centro sono state scelte troppe classi, non si coglie la distribuzone. A destra infine le classi sono state scelte ad hoc per rappresentare al meglio i dati cercando un compromesso tra sintesi e ricchezza dei dati.

$[\mathrm{x}_j,$	$\mathrm{x}_{j+1})$	$f_{j\%}$	$[\mathrm{x}_j,$	$\mathrm{x}_{j+1})$	$f_{j\%}$	$[\mathrm{x}_j,$	$\mathrm{x}_{j+1})$	$f_{j\%}$
0	5	55.56	0.0	2.5	15.56	0	3	20.00
5	20	44.44	2.5	5.0	40.00	3	5	35.56
		100.00	5.0	7.5	11.11	5	10	22.22
			7.5	10.0	11.11	10	20	22.22
			10.0	12.5	8.89			100.00
			12.5	15.0	2.22			
			15.0	17.5	6.67			
			17.5	20.0	4.44			
					100.00			

- $n = 45$
- $S_X = \{0.00, 0.01, 0.02, ..., 100.00, 100.01, ..., 2000.00, ..., 10\,000, ...\}$

Qui di seguito i dati nell'ordine in cui sono stati raccolti è sono mostrati sopra, mentre i dati riordinati sono mostrati sotto:

$x_1 =$	2.13	$x_{10} =$	3.08	$x_{19} =$	4.73	$x_{28} =$	7.31	$x_{37} =$	14.54
$x_2 =$	0.74	$x_{11} =$	4.32	$x_{20} =$	4.36	$x_{29} =$	6.65	$x_{38} =$	10.52
$x_3 =$	1.17	$x_{12} =$	4.76	$x_{21} =$	3.27	$x_{30} =$	8.17	$x_{39} =$	17.59
$x_4 =$	0.27	$x_{13} =$	4.78	$x_{22} =$	4.09	$x_{31} =$	7.12	$x_{40} =$	10.84
$x_5 =$	2.89	$x_{14} =$	4.13	$x_{23} =$	4.36	$x_{32} =$	7.03	$x_{41} =$	16.04
$x_6 =$	0.03	$x_{15} =$	4.19	$x_{24} =$	4.06	$x_{33} =$	9.59	$x_{42} =$	12.30
$x_7 =$	1.72	$x_{16} =$	3.73	$x_{25} =$	3.17	$x_{34} =$	9.04	$x_{43} =$	19.67
$x_8 =$	2.29	$x_{17} =$	3.71	$x_{26} =$	8.10	$x_{35} =$	7.70	$x_{44} =$	16.05
$x_9 =$	2.62	$x_{18} =$	4.18	$x_{27} =$	5.16	$x_{36} =$	11.07	$x_{45} =$	16.40
$x_{(1)} =$	0.03	$x_{(10)} =$	3.08	$x_{(19)} =$	4.19	$x_{(28)} =$	7.03	$x_{(37)} =$	10.84
$x_{(2)} =$	0.27	$x_{(11)} =$	3.17	$x_{(20)} =$	4.32	$x_{(29)} =$	7.12	$x_{(38)} =$	11.07
$x_{(3)} =$	0.74	$x_{(12)} =$	3.27	$x_{(21)} =$	4.36	$x_{(30)} =$	7.31	$x_{(39)} =$	12.30
$x_{(4)} =$	1.17	$x_{(13)} =$	3.71	$x_{(22)} =$	4.36	$x_{(31)} =$	7.70	$x_{(40)} =$	14.54
$x_{(5)} =$	1.72	$x_{(14)} =$	3.73	$x_{(23)} =$	4.73	$x_{(32)} =$	8.10	$x_{(41)} =$	16.04
$x_{(6)} =$	2.13	$x_{(15)} =$	4.06	$x_{(24)} =$	4.76	$x_{(33)} =$	8.17	$x_{(42)} =$	16.05
$x_{(7)} =$	2.29	$x_{(16)} =$	4.09	$x_{(25)} =$	4.78	$x_{(34)} =$	9.04	$x_{(43)} =$	16.40
$x_{(8)} =$	2.62	$x_{(17)} =$	4.13	$x_{(26)} =$	5.16	$x_{(35)} =$	9.59	$x_{(44)} =$	17.59
$x_{(9)} =$	2.89	$x_{(18)} =$	4.18	$x_{(27)} =$	6.65	$x_{(36)} =$	10.52	$x_{(45)} =$	19.67

Come si osserva aver rimesso in ordine i dati non ci aiuta a capire la distribuzione del fenomeno.

2.2.2 Raggruppamenti in Classi

L'idea è quella di raggruppare i dati in **intervalli contigui** e procedere alla rappresentazione in distribuzione di frequenza. In tabella 2.1 vediamo a sinistra troppe poche classi, al centro troppe, mentre a destra vediamo che il numero delle classi e la loro ampiezza variabile rende più leggibile la distribuzione dei dati.

2.2.3 Frequenze Cumulate

Si definisce frequenza cumulata F la seguente quantità:

- $F_1 = f_1$
- $F_2 = f_1 + f_2 = F_1 + f_2$
- $F_3 = f_1 + f_2 + f_3 = F_2 + f_3$
- ...
- $F_j = f_1 + f_2 + ... + f_j = F_{j-1} + f_j$
- ...
- $F_K = f_1 + f_2 + ... + f_K = 1$

ovvero $F_j = f_1 + ... + f_j$ *cumula* tutte le frequenze dalla 1 alla j.

Esempio 2.2.4 (Continua: Variabile titolo di studio)**.**

	n_j	f_j	F_j
E	1	0.1	0.1
M	2	0.2	0.3
S	4	0.4	0.7
U	3	0.3	1.0

E si legge: $F_1 = 0.1$ ci dice che il 10% del collettivo in esame ha come massimo titolo ha non più delle elementari. $F_1 = 0.3$ ci dice che il 30% del collettivo ha come massimo titolo ha non più delle medie. $F_4 = 0.7$ ci dice che il 30% del collettivo ha come massimo titolo ha non più delle superiori e $F_5 = 1$ che il 100% del collettivo ha, al massimo, la laurea.

Esempio 2.2.5 (Continua: Reddito)**.**

$[x_j,$	$x_{j+1})$	f_j	F_j
0	3	0.20	0.20
3	5	0.36	0.56
5	10	0.22	0.78
10	20	0.22	1.00
		1.00	

E si legge: $F_1 = 0.21$ ci dice che il 20% del collettivo in esame guadagna al massimo 3 (mila euro); alternativamente leggiamo che il 20% del collettivo **non** guadagna più di 3 (mila euro). $F_2 = 0.56$ ci dice che il 56% del collettivo guadagna al massimo 5. $F_2 = 0.56$ ci dice che il 56% del collettivo guadagna al massimo 5 (mila euro); il 56% non guadagna più di 5 (mila euro). $F_3 = 0.78$ ci dice che il 78% del collettivo guadagna 10 (mila euro); il 78% non guadagna più di 10 (mila euro). E infine il 100% del collettivo guadagna al massimo 20 (mila euro).

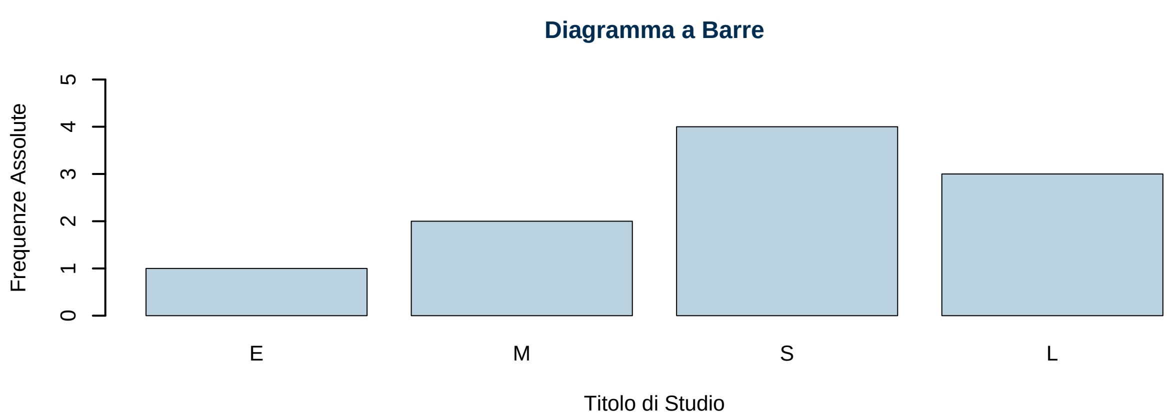
2.3 Rappresentazioni Grafiche

Diversi grafici statistici consentono di leggere in modo visivo una tabella di frequenza. Le frequenze vengono rappresentate come barre verticali -orizzontali- (diagramma a barre), o come angoli di una circonferenza (diagramma a torta).

Esempio 2.3.1 (Continua: Variabile titolo di studio). Consideriamo i 10 dati ordinati: $x_{(1)} = E, x_{(2)} = M, x_{(3)} = M, x_{(4)} = S, x_{(5)} = S\ , x_{(6)} = S, x_{(7)} = S, x_{(8)} = L, x_{(9)} = L, x_{(10)} = L$ e la loro tabella di frequenza

X	n_j	f_j	$f_{\%j}$
E	1	$1/10 = 0.1$	10%
M	2	$2/10 = 0.2$	20%
S	4	$4/10 = 0.4$	40%
L	3	$3/10 = 0.3$	30%
Tot	10	1.0	100%

Un diagramma a barre verticali trasforma le frequenze (assolute o relative) nelle altezze dei rettangoli



Mentre un diagramma a torta rappresenta le frequenze come spicchi della circonferenza e ovvero in angoli con una proporzione. Sia α_j l'angolo dello spicchio j in corrispondenza della modalità j, allora

$$\alpha_j : 360° = n_j : n$$

da cui si ricava facilmente che l'angolo di ogni spicchio è

$$\alpha_j = f_j \cdot 360°$$

o, in radianti

$$\theta_j = f_j \cdot 2\pi$$

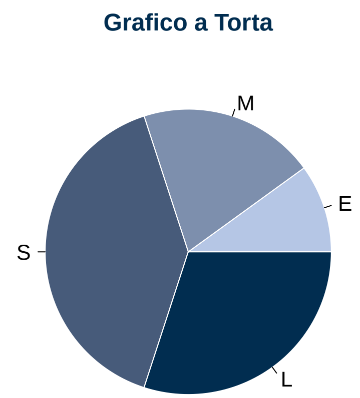


Figura 2.1: Grafico a torta

Attenzione

Attenzione la terza dimensione è inutile, non ha interpretazione fenomenica e rischi di non fare cogliere l'angolatura giusta.

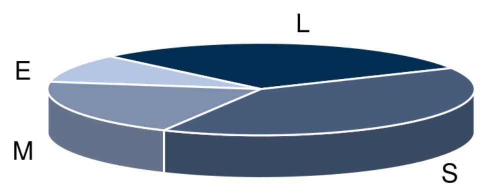


Figura 2.2: Attenzione alla terza dimensione

2.4 Istogramma di Densità

È grafico che rappresenta rettangoli contigui la cui area è la frequenza e la base è l'intervallo di raggruppamento. Usiamo il simbolo b_j per denotare l'ampiezza della base del rettangolo, l'altezza dei rettangoli viene chiamata *densità*

$$h_j = Const. \times \frac{f_j}{b_j}$$

Se $Const. = 1$ si ottiene *l'istogramma di densità relativa*, la somma delle aree dei rettangoli è 1. Se $Const. = n$ si ottiene *l'istogramma di densità assoluta*, la somma delle aree dei rettangoli è n. Se $Const. = 100$ si ottiene *l'istogramma di densità percentuale*, la somma delle aree dei rettangoli è 100. Per comodità tutti gli esempi si riferiscono all'istogramma di densità percentuale.

Tabella 2.2: Come ricavare le quantità necessarie per calcolare l'istogramma di densità percentuale

$[\mathrm{x}_j,$	$\mathrm{x}_{j+1})$	n_j	$f_j = \frac{n_j}{n}$	$b_j = \mathrm{x}_{j+1} - \mathrm{x}_j$	$h_j = 100 \times \frac{f_j}{b_j}$
$[\mathrm{x}_1 = 0,$	$\mathrm{x}_2 = 3)$	$n_1 = 9$	$f_1 = \frac{n_1}{n} = \frac{9}{45} = 0.20$	$b_1 = 3 - 0 = 3$	$h_1 = 100 \times \frac{0.20}{3} = 6.67$
$[\mathrm{x}_2 = 3,$	$\mathrm{x}_3 = 5)$	$n_2 = 16$	$f_2 = \frac{n_2}{n} = \frac{16}{45} = 0.36$	$b_2 = 5 - 3 = 2$	$h_2 = 100 \times \frac{0.36}{2} = 17.78$
$[\mathrm{x}_3 = 5,$	$\mathrm{x}_4 = 10)$	$n_3 = 10$	$f_3 = \frac{n_3}{n} = \frac{10}{45} = 0.22$	$b_3 = 10 - 5 = 5$	$h_3 = 100 \times \frac{0.22}{5} = 4.44$
$[\mathrm{x}_4 = 10$	$\mathrm{x}_5 = 20)$	$n_4 = 10$	$f_4 = \frac{n_4}{n} = \frac{10}{45} = 0.22$	$b_4 = 20 - 10 = 10$	$h_4 = 100 \times \frac{0.22}{10} = 2.22$

Esempio 2.4.1. La tabella 2.2 mostra passo, passo lo sviluppo del calcolo. La figura 2.3 la corrispondente rappresentazione grafica.

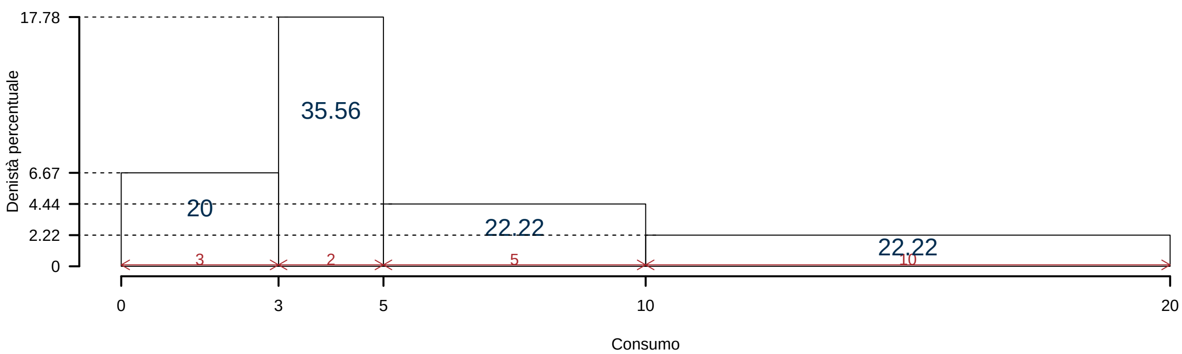


Figura 2.3: Rappresentazione grafica dell'istogramma di densità percentuale, l'area di ogni rettangolo corrisponde alla frequenza percentuale della classe, rappresentata sull'asse delle ascisse

2.5 La Funzione di Ripartizione

Se i dati sono quantitativi continui raggruppati in classi, la Funzione di Ripartizione della VS X è la funzione che misura l'area dell'istogramma di densità (le aree sommano ad 1) dal più piccolo dei dati $x_{(0)}$ fino ad un x qualunque. Se nel caso dell'esempio precedente scegliessimo $x = 7.2$, graficamente vedremmo la figura 2.4. Notiamo innanzitutto che:

$$\begin{aligned} F(\mathrm{x}_1) &= 0 \\ F(\mathrm{x}_2) &= F_1 \\ F(\mathrm{x}_3) &= F_2 \end{aligned}$$

$$\begin{array}{rcl} \vdots & & \vdots \\ F(\mathrm{x}_j) & = & F_{j-1} \\ \vdots & & \vdots \\ F(\mathrm{x}_{K+1}) & = & 1 \end{array}$$

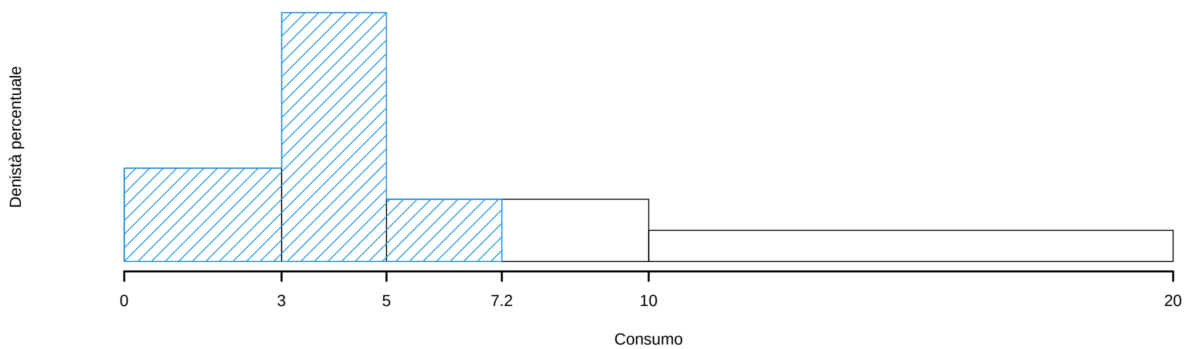


Figura 2.4: Rappresentazione grafica della Funzione di Ripartizione di X valutata nel punto 7.2, $F(7.2)$ è l'area da 0 a 7.2 dell'istogramma.

Nel nostro caso $F(7.2)$ è la comma delle frequenze fino a 5 più l'area del rettangolo di base $(7.2-5)$ e altezza $h_3 = 4.4444/100$, ovvero

$$\begin{aligned} F(7.2) &= f_1 + f_2 + \frac{(7.2-5)}{100} \times 4.4444 \\ &= F_2 + 2.5 \times 0.0444 \\ &= 0.6533 \end{aligned}$$

Se per esempio ci interessasse sapere, in modo approssimato, che percentuale e quanti individui che guadagno meno di 7.2 (mila euro) al mese, basta moltiplicare $F(7.2)$ per 100 e per n, rispettivamente.

$$\begin{aligned} \%(X < 7.2) &= 100 \times F(7.2) \\ &= 65.3333\% \\ \#(X < 7.2) &= 45 \times F(7.2) \\ &= 29.4 \end{aligned}$$

Dove $\%(X < 7.2)$ significa la *percentuale* approssimata di dati minori di 7.2 e dove $\#(X < 7.2)$ significa il *numero* approssimato di dati minori di 7.2.

Se per esempio sono interessato alla percentuale (o al numero) di dati compresi tra 2.4 e 7.2 osservo che

$$\%(2.4 < X < 7.2) = 100 \times (F(7.2) - F(2.4))$$

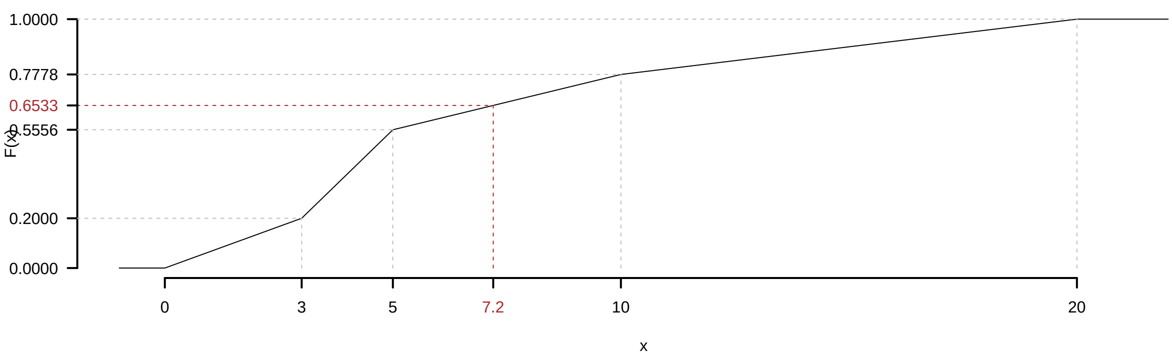
$$\begin{aligned} F(2.4) &= \frac{2.4-0}{100} \times 6.6667 \\ &= 0.16 \\ \%(2.4 < X < 7.2) &= 100 \times (0.6533 - 0.16) \\ &= 49.3333. \end{aligned}$$

Infatti calcolare l'area tra 2.4 e 7.2 equivale a calcolare l'area fino a 7.2, l'area fino a 2.4 e sottrarle.

Più in generale la funzione di ripartizione cumula l'area dal più piccolo degli x fino al più grande.

$$\begin{aligned} F(x) &= 0 \quad \text{per ogni } x \leq \mathrm{x}_1 \\ F(x) &= F(\mathrm{x}_{j^*-1}) + \frac{x - \mathrm{x}_{j^*}}{100} h_{j^*} \\ F(x) &= 1 \quad \text{per ogni } x \geq \mathrm{x}_{K+1} \end{aligned}$$

dove j^* è la classe che contiene x. Se la volessimo rappresentare graficamente, nel nostro esempio sarebbe così:



2.6 L'inversa della Funzione di Ripartizione

La funzione di ripartizione è una funziona che crescente che vale zero quando x è il più piccolo dei dati e vale uno quando x è il più grande dei dati.

$$F : S_X \to [0,1]$$

Definiamo $Q = F^{-1}$ la funzione inversa:

$$Q : [0,1] \to S_X$$

ed è tale che

$$Q(p) = x_p : F(x_p) = p, 0 \leq p \leq 1$$

2.7 Indicatori Sintetici di Centralità e di Variabilità

Un indicatore è un numero che sintetizza una caratteristica del fenomeno collettivo. Esempi di indicatori sono: il massimo del fenomeno, il minimo del fenomeno, la media del fenomeno, la modalità più ricorrente, ecc.

Gli **indicatori di centralità** sintetizzano l'intero fenomeno in un numero. Indicatori di che osserveremo centralità sono:

- La **media aritmetica** (variabili quantitative) nella sezione 3.1
- La **mediana** (variabili quantitative e variabili qualitative ordinate) nella sezione 4.1
- La **moda** (ogni tipo di variabile) nella sezione 4.4

La media aritmetica è una *media analitica* perché dipende dal valore che la variabile assume sulle unità. Mediana e Moda sono invece *medie lasche* perché dipendono dall'ordinamento dei dati.

Gli **indicatori di variabilità** misurano lo scostamento del fenomeno oggetto di studio dall'indicatore di centralità. Vedremo:

- La **varianza** 3.2 e la **standard deviation** nella sezione 3.3
- Lo **scarto interquartile** nella sezione 4.3

2.8 Riepilogo

Estremo inf	Estremo sup	freq. ass.	freq. relativa	freq. cum.	ampiezza	densità
$[\mathrm{x}_1,$	$\mathrm{x}_2)$	n_1	$f_1 = \frac{n_1}{n}$	$F_1 = f_1$	$b_1 = \mathrm{x}_2 - \mathrm{x}_1$	$h_1 = 100 \times \frac{h_1}{b_1}$
$[\mathrm{x}_2,$	$\mathrm{x}_3)$	n_2	$f_2 = \frac{n_2}{n}$	$F_2 = F_1 + f_2$	$b_2 = \mathrm{x}_3 - \mathrm{x}_2$	$h_2 = 100 \times \frac{f_2}{b_2}$
...	...	...	...	...	...	...
$[\mathrm{x}_j,$	$\mathrm{x}_{j+1})$	n_j	$f_j = \frac{n_j}{n}$	$F_j = F_{j-1} + f_j$	$b_j = \mathrm{x}_{j+1} - \mathrm{x}_j$	$h_j = 100 \times \frac{f_j}{b_j}$
...	...	...	...	...	...	...
$[\mathrm{x}_K,$	$\mathrm{x}_{K+1})$	n_K	$f_K = \frac{n_K}{n}$	$F_K = F_{K-1} + f_K$	$b_K = \mathrm{x}_{K+1} - \mathrm{x}_K$	$f_K = 100 \times \frac{f_K}{b_K}$

Media Aritmetica, Varianza e Standard Deviation

3

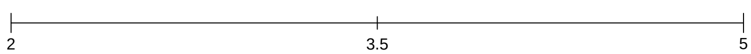
3.1 Media Aritmetica

La media tra due numeri x_1 e x_2 il punto centrale

$$\bar{x} = \frac{x_1 + x_2}{2}$$

Esempio. Posto $x_1 = 2$, $x_2 = 5$, allora

$$\bar{x} = \frac{2+5}{2} = 3.5$$



Definizione 3.1.1 (Media Aritmetica)**.** Consideriamo la serie dei dati $\mathbf{x} = (x_1, ..., x_i, ..., x_n)$, si definisce la media aritmetica:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Ovvero la media tra n numeri $x_1, x_2, ..., x_n$ è definita da

$$\bar{x} = \frac{x_1 + x_2 + ... + x_n}{n}$$

siccome la somma dei dati, rappresenta il totale (Tot) del fenomeno nel collettivo

$$Tot = x_1 + x_2 + ... + x_n = \sum_{i=1}^{n} x_i$$

allora la media aritmetica

$$\bar{x} = \frac{Tot}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

rappresenta la quantità *ipotetica* che ogni individuo possiederebbe se il totale fosse equi-ripartito.

3.1.1 La Media Aritmetica come Baricentro dell'Istogramma

La media aritmetica tiene in equilibrio l'istogramma di densità come se si trattase di un sistema fisico. Se per esempio consideriamo 3 diverse serie di dati

$$\begin{aligned} \mathbf{x}_1 &= (x_1 = 1, x_2 = 2, x_3 = 2, x_4 = 3) \\ \mathbf{x}_2 &= (x_1 = 1, x_2 = 2, x_3 = 2, x_4 = 5) \\ \mathbf{x}_3 &= (x_1 = 1, x_2 = 2, x_3 = 2, x_4 = 7) \end{aligned}$$

E osserviamo che

$$\begin{aligned} \bar{x}_1 &= \frac{1+2+2+3}{4} = 2 \\ \bar{x}_2 &= \frac{1+2+2+5}{4} = 2.5 \\ \bar{x}_3 &= \frac{1+2+2+7}{4} = 3 \end{aligned}$$

Ovvero spostando l'ultimo dato verso valori maggiori spingiamo la media su valori maggiori. Graficamente osserviamo come la medie tenga in equilibrio l'istogramma nella figura 3.1.

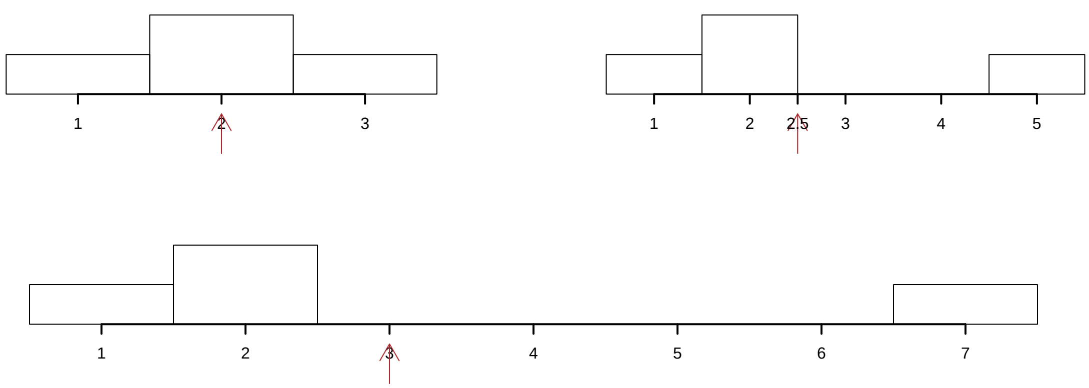


Figura 3.1: La media artimetica tiene in equiibrio l'istogramma di densità, più ci sono dati estremi molto grandi più la media sale per mantenere l'equilibrio col totale.

Allo stesso modo se osserviamo:

$$\mathbf{x}_4 = (x_1 = -1, x_2 = 2, x_3 = 2, x_4 = 3)$$

$$\mathbf{x}_5 = (x_1 = -3, x_2 = 2, x_3 = 2, x_4 = 3),$$

allora

$$\begin{aligned}
\bar{x}_1 &= \frac{1+2+2+3}{4} = 2 \\
\bar{x}_4 &= \frac{-1+2+2+3}{4} = 1.5 \\
\bar{x}_5 &= \frac{-3+2+2+3}{4} = 1
\end{aligned}$$

Ovvero spostando il primo dato verso valori minori spingiamo la media su valori minori. Graficamente osserviamo come la media tenga in equilibrio l'istogramma nella figura 3.2.

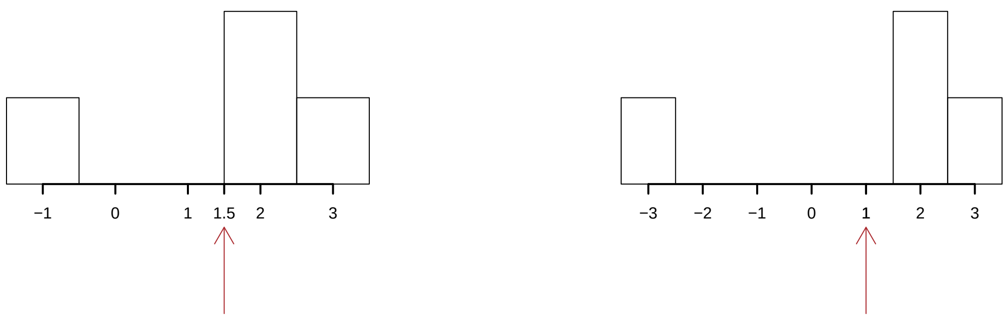


Figura 3.2: Analogamente se spostiamo un dato verso sinistra la media si sposta a sinistra anch'essa

3.1.2 Calcolo per Distribuzioni di Frequenza

Se i dati sono raccolti in distribuzione di frequenza

Modalità	x_1	x_2	...	x_j	...	x_K	
Frequenze	n_1	n_2	...	n_j	...	n_K	n

Definizione 3.1.2 (Media Artimetica per Dati Raccolti in Classi)**.**

$$\bar{x} = \frac{1}{n}\sum_{j=1}^{K} \mathrm{x}_j n_j$$

Esempio 3.1.1. Osserviamo i seguenti dati: $x_1 = 3.4$; $x_2 = 3.4$; $x_3 = 2.7$; $x_4 = 3.4$; $x_5 = 2.7$; $x_6 = 3.4$; $x_7 = 2.7$; $x_8 = 5.1$; $x_9 = 5.1$; $x_{10} = 2.7$;

La media

$$
\begin{aligned}
\bar{x} &= \frac{1}{n}\sum_{i=1}^{n} x_i \\
&= \frac{x_1+x_2+x_3+x_4+x_5+x_6+x_7+x_8+x_9+x_{10}}{10} \\
&= \frac{3.4+3.4+2.7+3.4+2.7+3.4+2.7+5.1+5.1+2.7}{10} \\
&= \frac{34.6}{10} \\
&= 3.46
\end{aligned}
$$

Riordiniamo i dati: $x_{(1)} = 2.7$; $x_{(2)} = 2.7$; $x_{(3)} = 2.7$; $x_{(4)} = 2.7$; $x_{(5)} = 3.4$; $x_{(6)} = 3.4$; $x_{(7)} = 3.4$; $x_{(8)} = 3.4$; $x_{(9)} = 5.1$; $x_{(10)} = 5.1$;

E raccogliamo in distribuzione di frequenza:

modalita	$\mathrm{x}_1 = 2.7$	$\mathrm{x}_2 = 3.4$	$\mathrm{x}_3 = 5.1$	
frequenze	4	4	2	10

la media:

$$
\begin{aligned}
\bar{x} &= \frac{1}{n}\sum_{j=1}^{K} \mathrm{x}_j n_j \\
&= \frac{\mathrm{x}_1 n_1 + \mathrm{x}_2 n_2 + \mathrm{x}_3 n_3}{n} \\
&= \frac{2.7 \times 4 + 3.4 \times 4 + 5.1 \times 2}{10} \\
&= \frac{34.6}{10} \\
&= 3.46
\end{aligned}
$$

3.1.3 Proprietà della Media Aritmetica

Proprietà 3.1.1 (della media aritmetica). *Le principale proprietà della media aritmetica sono:*

0. Internalità: $x_{\min} = x_{(1)} \leq \bar{x} \leq x_{(n)} = x_{\max}$

1. Invarianza della somma:

$$n\bar{x} = \sum_{i=1}^{n} x_i$$

2. Somma degli scarti dalla media nulla: $\sum_{i=1}^{n}(x_i - \bar{x}) = 0$

3. Minimizza la somma degli scarti al quadrato:

$$\sum_{i=1}^{n}(x_i-\bar{x})^2<\sum_{i=1}^{n}(x_i-d)^2 \quad \forall d\neq\bar{x}$$

4. Invarianza per trasformazioni lineari: se $y_i=a+bx_i$ allora $\bar{y}=a+b\bar{x}$

5. Associatività. Sia una popolazione, $\mathscr{P}$, formata da K gruppi con medie e numerosità: $(\bar{x}_1;\ n_1)$, $(\bar{x}_2;\ n_2)$, ..., $(\bar{x}_K;\ n_K)$. Allora, la media totale $\bar{x}_T$ di $\mathscr{P}=$ è data da

$$\bar{x}_T=\frac{Tot\{\mathscr{P}_1\}+\cdots+Tot\{\mathscr{P}_K\}}{n_1+\cdots+n_K}=\frac{n_1\ \bar{x}_1+\cdots+n_K\ \bar{x}_K}{n_1+\cdots+n_K}$$

Dimostrazione. Qui di seguito le dimostrazioni

0. La proprietà di internalità deriva dal fatto che la somma dei dati è maggiore della somma di n volte del più piccolo dei dati $\sum_{i=1}^{n}x_i>\sum_{i=1}^{n}x_{(1)}$. Mentre $\sum_{i=1}^{n}x_i<\sum_{i=1}^{n}x_{(n)}$ la somma dei dati è maggiore della somma di n volte del più grande dei dati.

1. La proprietà di invarianza della somma la otteniamo direttamente dalla definizione di media aritmetica.

$$\begin{aligned}\bar{x}&=\frac{1}{n}\sum_{i=1}^{n}x_i\\ n\bar{x}&=\sum_{i=1}^{n}x_i.\end{aligned}$$

2. Somma degli scarti dalla media nulla. Osserviamo che

$$\begin{aligned}\sum_{i=1}^{n}(x_i-\bar{x})&=\sum_{i=1}^{n}x_i-\sum_{i=1}^{n}\bar{x}\\ &=n\bar{x}-n\bar{x}\\ &=0.\end{aligned}$$

3. Minimizza la somma degli scarti al quadrato. Se poniamo $g(x)=\sum_{i=1}^{n}(x_i-x)^2$ osserviamo che

$$\begin{aligned}g(x)&=\sum_{i=1}^{n}(x_i-x)^2\\ &=(x_1-x)^2+\ldots+(x_n-x)^2 && \text{La funzione } g \text{ è una somma di parabole}\\ g'(x)&=-2(x_1-x)-\ldots-2(x_n-x) && \text{Dove } g' \text{ indica la derivata prima di } g\\ &=-2\sum_{i=1}^{n}x_i-2nx\end{aligned}$$

$$\begin{aligned} g'(x) &= 0 && \text{Eguagliamo } g' \text{ a zero per avere il minimo} \\ -2\sum_{i=1}^{n} x_i - 2nx &= 0 \\ x &= \frac{1}{n}\sum_{i=1}^{n} x_i. \end{aligned}$$

4. Invarianza per trasformazioni lineari: se $y_i = a + bx_i$ allora

$$\begin{aligned} \bar{y} &= \frac{1}{n}\sum_{i=1}^{n} y_i \\ &= \frac{1}{n}\sum_{i=1}^{n}(a + bx_i) \\ &= \frac{1}{n}\sum_{i=1}^{n} a + \frac{1}{n}\sum_{i=1}^{n} bx \\ &= a + b\bar{x}. \end{aligned}$$

5. Associatività. Sia una popolazione, $\mathscr{P}$, formata da K gruppi con medie e numerosità: $(\bar{x}_1;\ n_1)$, $(\bar{x}_2;\ n_2)$, ..., $(\bar{x}_K;\ n_K)$. Allora, il totale di tutte le popolazioni è $Tot = n_1\bar{x}_1 + ... + n_1\bar{x}_K$, mentre il numero totale di individui di tutte e K le popolazioni è $n_T = n_1 + ... + n_K$ E quindi la media

$$\bar{x}_T = \frac{\mathrm{Tot}\{\mathscr{P}_1\} + \cdots + \mathrm{Tot}\{\mathscr{P}_K\}}{n_1 + \cdots + n_K} = \frac{n_1\ \bar{x}_1 + \cdots + n_K\ \bar{x}_K}{n_1 + \cdots + n_K}$$

$\square$

3.2 La varianza

La media riduce un complesso di n dati in uno solo. A parità di media i dati possono essere molto diversi tra di loro. Per esempio le due serie di dati

$$\begin{aligned} \mathbf{x}_1 &= (x_1 = 2, x_2 = 2, x_3 = 2, x_4 = 2) \\ \mathbf{x}_2 &= (x_1 = 1, x_2 = 2, x_3 = 2, x_4 = 3) \\ \mathbf{x}_3 &= (x_1 = 0, x_2 = 0, x_3 = 0, x_4 = 8) \end{aligned}$$

hanno tutte la stessa media $\bar{x}_1 = \bar{x}_2 = \bar{x}_3 = 2$, ma nel primo caso tutti possiedono la media, nel secondo chi poco e chi tanto, nel terzo caso uno possiede il totale e gli altri 3 nulla.

La varianza misura la distanza dei dati dalla media.

Definizione 3.2.1 (Varianza). Si definisce la varianza la quantità:

$$\sigma^2 = \frac{1}{n}\sum_{1=1}^{n}(x_i - \bar{x})^2$$

La varianza misura lo scostamento medio quadratico dei dati dalla media aritmetica; ovvero è la media del quadrato degli scarti.

Con un po' di algebra si dimostra che

Proprietà 3.2.1 (Formula Calcolatoria della Varianza).

$$\sigma^2 = \frac{1}{n}\sum_{1=1}^{n} x_i^2 - \bar{x}^2$$

Dimostrazione.

$$\begin{aligned}
\sigma^2 &= \frac{1}{n}\sum_{1=1}^{n}(x_i - \bar{x})^2 \\
&= \frac{1}{n}\sum_{1=1}^{n}(x_i^2 + \bar{x}^2 - 2 \cdot x_i \cdot \bar{x}) \\
&= \frac{1}{n}\sum_{1=1}^{n} x_i^2 + \frac{1}{n}\sum_{1=1}^{n}\bar{x}^2 - \frac{1}{n}\sum_{1=1}^{n} 2 \cdot x_i \cdot \bar{x} \\
&= \frac{1}{n}\sum_{1=1}^{n} x_i^2 + \frac{n}{n}\bar{x}^2 - \frac{2}{n}\bar{x} \cdot n \cdot \bar{x} \\
&= \frac{1}{n}\sum_{1=1}^{n} x_i^2 - \bar{x}^2
\end{aligned}$$

□

Quindi la varianza si può calcolare o come media del quadrato degli scarti dalla media o come media dei quadrati meno il quadrato della media.

Esempio 3.2.1. Posto $x_1 = 2$, $x_2 = 2$, $x_3 = 2$, $x_4 = 2$ allora

$$\bar{x} = \frac{2+2+2+2}{4} = 2$$

$$\sigma^2 = \frac{(2-2)^2 + (2-2)^2 + (2-2)^2 + (2-2)^2}{4} = 0$$

Tutti gli individui hanno la stessa quantità che è pari alla media, non c'è variabilità, la varianza vale zero.

Esempio 3.2.2. Posto $x_1 = 1$, $x_2 = 2$, $x_3 = 2$, $x_4 = 3$ allora

$$\bar{x} = \frac{2+2+2+2}{4} = 2$$

$$\sigma^2 = \frac{(1-2)^2 + (2-2)^2 + (2-2)^2 + (3-2)^2}{4} = 0.5$$

Non tutti gli individui hanno la stessa quantità, c'è variabilità, la varianza è diversa da zero.

Esempio 3.2.3. Posto $x_1 = 0$, $x_2 = 0$, $x_3 = 0$, $x_4 = 8$ allora

$$\bar{x} = \frac{0+0+0+8}{4} = 2$$

$$\sigma^2 = \frac{(0-2)^2 + (0-2)^2 + (0-2)^2 + (8-2)^2}{4} = 12$$

Tutto il totale è posseduto da un solo individuo, c'è massima variabilità.

3.2.1 Calcolo per Distribuzioni di Frequenza

Se i dati sono raccolti in distribuzione di frequenza

Modalità	x_1	x_2	...	x_j	...	x_K	
Frequenze	n_1	n_2	...	n_j	...	n_K	n

la varianza si può calcolare

Proprietà 3.2.2 (Varianza per Dati in Distribuzione di Frequenza).

$$\sigma^2 = \frac{1}{n}\sum_{j=1}^{k}(\mathrm{x}_j - \bar{x})^2 n_j$$

e di conseguenza, con un po' di algebra otteniamo:

Proprietà 3.2.3 (Formula Calcolatoria per la Varianza per Dati in Distribuzione di Frequenza).

$$\sigma^2 = \frac{1}{n}\sum_{j=1}^{k}\mathrm{x}_j^2 n_j - \bar{x}^2$$

Esempio 3.2.4. $x_1 = 3.4$; $x_2 = 3.4$; $x_3 = 2.7$; $x_4 = 3.4$; $x_5 = 2.7$; $x_6 = 3.4$; $x_7 = 2.7$; $x_8 = 5.1$; $x_9 = 5.1$; $x_{10} = 2.7$;

La media

$$
\begin{aligned}
\bar{x} &= \frac{1}{n}\sum_{i=1}^{n} x_i \\
&= \frac{x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7 + x_8 + x_9 + x_{10}}{10} \\
&= \frac{3.4 + 3.4 + 2.7 + 3.4 + 2.7 + 3.4 + 2.7 + 5.1 + 5.1 + 2.7}{10} \\
&= \frac{34.6}{10} \\
&= 3.46
\end{aligned}
$$

La varianza

$$
\begin{aligned}
\sigma^2 &= \frac{1}{n}\sum_{1=1}^{n} (x_i - \bar{x})^2 \\
&= \frac{(x_1-\bar{x})^2 + (x_2-\bar{x})^2 + (x_3-\bar{x})^2 + (x_4-\bar{x})^2 + (x_5-\bar{x})^2 + (x_6-\bar{x})^2 + (x_7-\bar{x})^2 + (x_8-\bar{x})^2 + (x_9-\bar{x})^2}{10} \\
&= \frac{(3.4-3.46)^2 + (3.4-3.46)^2 + (2.7-3.46)^2 + (3.4-3.46)^2 + (2.7-3.46)^2 + (3.4-3.46)^2 + (2.7-3.46)^2}{10} \\
&= \frac{0.0036 + 0.0036 + 0.5776 + 0.0036 + 0.5776 + 0.0036 + 0.5776 + 2.6896 + 2.6896 + 0.5776}{10} \\
&= \frac{7.704}{10} \\
&= 0.7704
\end{aligned}
$$

Osserviamo che

$$
\begin{aligned}
\sigma^2 &= \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2 \\
&= \frac{1}{10}(x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_5^2 + x_6^2 + x_7^2 + x_8^2 + x_9^2 + x_{10}^2) - \bar{x}^2 \\
&= \frac{1}{10}(3.4^2 + 3.4^2 + 2.7^2 + 3.4^2 + 2.7^2 + 3.4^2 + 2.7^2 + 5.1^2 + 5.1^2 + 2.7^2) - 3.46^2 \\
&= \frac{1}{10}(11.56 + 11.56 + 7.29 + 11.56 + 7.29 + 11.56 + 7.29 + 26.01 + 26.01 + 7.29) - 11.9716 \\
&= 12.742 - 11.9716 \\
&= 0.7704
\end{aligned}
$$

Riordiniamo i dati: $x_{(1)} = 2.7$; $x_{(2)} = 2.7$; $x_{(3)} = 2.7$; $x_{(4)} = 2.7$; $x_{(5)} = 3.4$; $x_{(6)} = 3.4$; $x_{(7)} = 3.4$; $x_{(8)} = 3.4$; $x_{(9)} = 5.1$; $x_{(10)} = 5.1$;

E raccogliamo in distribuzione di frequenza:

modalita	$x_1 = 2.7$	$x_2 = 3.4$	$x_3 = 5.1$	
frequenze	4	4	2	10

la media:

$$\begin{aligned}
\bar{x} &= \frac{1}{n}\sum_{j=1}^{K} x_j n_j \\
&= \frac{x_1 n_1 + x_2 n_2 + x_3 n_3}{n} \\
&= \frac{2.7 \times 4 + 3.4 \times 4 + 5.1 \times 2}{10} \\
&= \frac{34.6}{10} \\
&= 3.46
\end{aligned}$$

la varianza:

$$\begin{aligned}
\sigma^2 &= \frac{1}{n}\sum_{j=1}^{k} (x_j - \bar{x})^2 n_j \\
&= \frac{1}{10}\left((x_1 - \bar{x})^2 n_1 + (x_2 - \bar{x})^2 n_2 + (x_3 - \bar{x})^2 n_3\right) \\
&= \frac{1}{10}\left((2.7 - 3.46)^2 \times 4 + (3.4 - 3.46)^2 \times 4 + (5.1 - 3.46)^2 \times 2\right) \\
&= \frac{7.704}{10} \\
&= 0.7704
\end{aligned}$$

o alternativamente

$$\begin{aligned}
\sigma^2 &= \frac{1}{n}\sum_{j=1}^{k} x_j^2 n_j - \bar{x}^2 \\
&= \frac{1}{10}(x_1^2 n_1 + x_2^2 n_2 + x_3^2 n_3) - \bar{x}^2 \\
&= \frac{1}{10}(2.7^2 \times 4 + 3.4^2 \times 4 + 5.1^2 \times 2) - 3.46^2 \\
&= \frac{1}{10} \times 127.42 - 11.9716 \\
&= 0.7704
\end{aligned}$$

Esempio 3.2.5. $x_1 = 3.61$; $x_2 = 3.32$; $x_3 = 3.16$; $x_4 = 3.74$; $x_5 = 3.61$; $x_6 = 3.61$; $x_7 = 3.61$; $x_8 = 3.46$; $x_9 = 3.61$; $x_{10} = 3.61$; $x_{11} = 3.74$; $x_{12} = 3.32$; $x_{13} = 3.74$; $x_{14} = 3.74$; $x_{15} = 3.74$; $x_{16} = 3.46$; $x_{17} = 3.46$; $x_{18} = 3.46$; $x_{19} = 3.87$; $x_{20} = 3.61$; $x_{21} = 3.61$;

La media:

$$\begin{aligned} \bar{x} &= \frac{1}{n}\sum_{i=1}^{n} x_i \\ &= \frac{1}{21}75.09 \\ &= 3.5757 \end{aligned}$$

La varianza

$$\begin{aligned} \sigma^2 &= \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2 \\ &= \frac{1}{21}269.0885 - 3.5757^2 \\ &= 9.238 \end{aligned}$$

riordiniamo i dati

$x_{(1)} = 3.16$; $x_{(2)} = 3.32$; $x_{(3)} = 3.32$; $x_{(4)} = 3.46$; $x_{(5)} = 3.46$; $x_{(6)} = 3.46$; $x_{(7)} = 3.46$; $x_{(8)} = 3.61$; $x_{(9)} = 3.61$; $x_{(10)} = 3.61$; $x_{(11)} = 3.61$; $x_{(12)} = 3.61$; $x_{(13)} = 3.61$; $x_{(14)} = 3.61$; $x_{(15)} = 3.61$; $x_{(16)} = 3.74$; $x_{(17)} = 3.74$; $x_{(18)} = 3.74$; $x_{(19)} = 3.74$; $x_{(20)} = 3.74$; $x_{(21)} = 3.87$;

E raccogliamo in distribuzione di frequenza:

x_j	$\mathrm{x}_1 = 3.16$	$\mathrm{x}_2 = 3.32$	$\mathrm{x}_3 = 3.46$	$\mathrm{x}_4 = 3.61$	$\mathrm{x}_5 = 3.74$	$\mathrm{x}_6 = 3.87$	Tot
n_j	1	2	4	8	5	1	21
$\mathrm{x}_j n_j$	10	22	48	104	70	15	269
$\mathrm{x}_j^2 n_j$	100	242	576	1352	980	225	3475

E osserviamo che

$$\begin{aligned} \bar{x} &= \frac{1}{n}\sum_{j=1}^{k} \mathrm{x}_j n_j \\ &= \frac{1}{21}75.09 \\ &= 3.5757 \end{aligned}$$

e che

$$\begin{aligned}\sigma^2 &= \frac{1}{n}\sum_{j=1}^{n} \mathrm{x}_j^2 n_j - \bar{x}^2 \\ &= \frac{1}{21}269.0885 - 3.5757^2 \\ &= 9.238\end{aligned}$$

3.2.2 Proprietà della Varianza

Proprietà 3.2.4 (della varianza σ^2). *Le principale proprietà della varianza sono:*

1. $\sigma^2 \geq 0$.
2. $\sigma^2 = 0$, *se e solo se* X *è costante.*
3. *Se* $y_i = a + bx_i$ *allora* $\sigma_Y^2 = b^2\sigma_X^2$.

Dimostrazione. Le dimostrazioni qui di seguito.

1. $\sigma^2 \geq 0$ deriva direttamente dalla definizione, essendo σ^2 la media di scarti al quadrato e quindi di quantità positive, non potrà *mai* essere negativa.

2. $\sigma^2 = 0$ solo se ogni scarto dalla media è zero e questo può avvenire solo se tutti i dati sono uguali alla media, ovvero se i dati sono tutti uguali tra di loro e quindi non variano.

3. Se $y_i = a + bx_i$ allora

$$\begin{aligned}\sigma_Y^2 &= \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2 \\ &= \frac{1}{n}\sum_{i=1}^{n}(a + bx_i - (a + b\bar{x}))^2 \\ &= b^2\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \\ &= b^2\sigma_X^2\end{aligned}$$

$\square$

3.3 La Standard Deviation

La varianza non ha un'unità di misura leggibile, è una media di quadrati degli scarti. E quindi anche l'unità di misura è elevata al quadrato

Si definisce la *standard deviation* (deviazione standard, scarto quadratico medio), la radice della varianza

$$\sigma = \sqrt{\sigma^2}$$

3.3.1 Proprietà della Standard Deviation

Proprietà 3.3.1 (della deviazione standard σ)**.** *Le principale proprietà della deviazione standard sono:*

1. $\sigma \geq 0$.
2. $\sigma = 0$, *se e solo se* X *è costante.*
3. *Se* $y_i = a + bx_i$ *allora allora* $\sigma_Y = |b|\sigma_X$

Se la distribuzione della X è abbastanza simmetrica e di forma campanulare, allora

$$\%(\bar{x} - 2\sigma \leq X \leq \bar{x} + 2\sigma) \approx 95\%$$

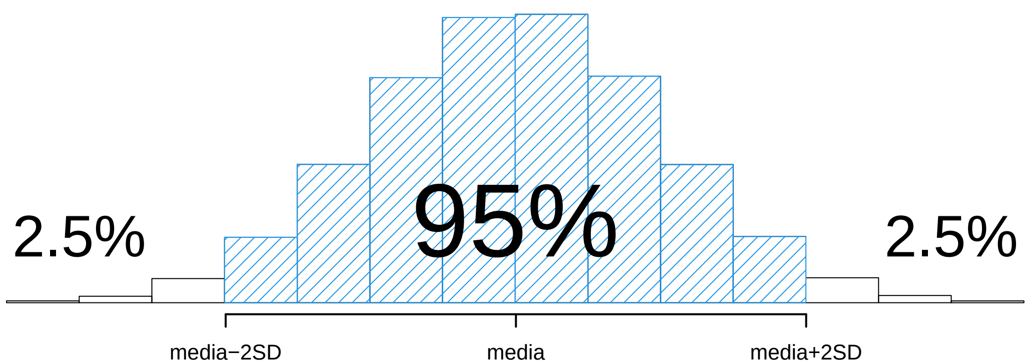


Figura 3.3: Interpretazione della SD

la figura La figura 3.3 la corrispondente rappresentazione grafica.

3.4 Esempi

Esempio 3.4.1. Si si è chiesta l'età a 37500 uomini e 38100 donne di un determinato collettivo, ed è risultato che gli uomini di quel collettivo hanno un'età media di 45 anni e le donne un'età media di 49 anni. La sintesi dei dati qui di seguito:

	n_j	$\bar{x}_j$
Uomini	37500	45
Donne	38100	49

Calcolare l'età media dell'intero collettivo.

Soluzione. L'età media per l'intera popolazione è

$$\bar{x}_a = \frac{37500 \times 45 + 38100 \times 49}{37500 + 38100} = 47.02.$$

Esempio 3.4.2. Uno studente iscritto al secondo anno di un CdL, ha superato 7 esami con un voto medio pari a 26/30. Sostiene un nuovo esame ottenendo un voto pari a 28/30. Qual è il voto medio dopo l'ottavo esame?

Soluzione. Sia $\bar{x}_7 = 26$ il voto medio dopo i primi 7 esami. Sia $\bar{x}_8$ il voto medio dopo l'8^o esame.

$$\begin{aligned}
\bar{x}_7 &= \frac{1}{7}\sum_{i=1}^{7} x_i \\
\bar{x}_8 &= \frac{1}{8}\sum_{i=1}^{8} x_i = \frac{1}{8}\left(\sum_{i=1}^{7} x_i + x_8\right) \\
&= \frac{1}{8}\left(7\bar{x}_7 + x_8\right) = \frac{7 \times \bar{x}_7 + 1 \times x_8}{8} \\
&= \frac{7 \times 26 + 1 \times 28}{8} = \frac{182 + 28}{8} = 26.25.
\end{aligned}$$

Esempio 3.4.3. In una contrattazione sindacale, il rappresentante del governo propone di alzare di un ammontare fisso di 100 euro lo stipendio degli impiegati statali.

a. Come cambierebbero lo stipendio medio e la varianza se questa misura fosse intrapresa?

b. Se il governo aumentasse lo stipendio di ciascun impiegato statale del 5%, come cambierebbe lo stipendio medio e la varianza?

Soluzione.

a. Sia x lo stipendio degli statali.

$$\begin{aligned}
y_i &= x_i + 100 \\
\bar{y} &= \bar{x} + 100.
\end{aligned}$$

$$\sigma_Y^2 = \sigma_X^2$$

Stipendio medio aumenta esattamente di 100€, la varianza non cambia.

b. Aumento percentuale pari al 5%.

$$\begin{aligned}
y_i &= x_i + \frac{5}{100}\, x_i = 1.05\, x_i \\
\bar{y} &= 1.05\bar{x}
\end{aligned}$$

$$\sigma_Y^2 = (1.05)^2\sigma_X^2$$

La MEDIA aumenta del 5%, la varianza aumenta in modo quadratico.

Esempio 3.4.4. La spesa per le vacanze estive (in migliaia di euro) sostenuta da 12 famiglie di un condominio è stata di:

0	0	2	2.5	4	5.1	5.8	6	7	12	15	21

(a) Determinare la spesa media e la varianza per famiglia.

$$\begin{aligned}\bar{x} &= \frac{0+0+2+2.5+4+5.1+5.8+6}{12}+\ldots \\ &+ \frac{7+12+15+21}{12} = \frac{80.4}{12} = 6.7€. \\ \sigma_X^2 &= \frac{0^2+0^2+2^2+\ldots+21^2}{12} - (6.7)^2 \\ &= 36.8517\end{aligned}$$

(b) Determinare la spesa media per famiglie con spesa $\neq 0$.

$$\begin{aligned}\bar{x} &= \frac{2+2.5+4+5.1+5.8+6+7+12+15+21}{10} \\ &= \frac{80.4}{10} = 8.04€. \\ \sigma_X^2 &= \frac{2^2+\ldots+21^2}{10} - (8.04)^2 \\ &= 33.4484\end{aligned}$$

Esempio 3.4.5. Numero di impiegati per anni di servizio di una industria

$[\mathrm{x}_j,$	$\mathrm{x}_{j+1})$	n_j	f_j	F_j	$\bar{\mathrm{x}}_j$	$\bar{\mathrm{x}}_j^2$	$\bar{\mathrm{x}}_j n_j$	$\bar{\mathrm{x}}_j^2 n_j$
0	1	7	0.0609	0.0609	0.5	0.25	3.5	1.75
1	5	18	0.1565	0.2174	3.0	9.00	54.0	162.00
5	10	45	0.3913	0.6087	7.5	56.25	337.5	2531.25
10	20	25	0.2174	0.8261	15.0	225.00	375.0	5625.00
20	30	20	0.1739	1.0000	25.0	625.00	500.0	12500.00
Totale		115	1.0000				1270.0	20820.00

Determinare media, mediana e varianza dell'età di servizio dell'industria.

$$\bar{x} = \frac{0.5\times 7+3.0\times 18+\cdots+20\times 25}{7+18+\cdots+20} = \frac{1270}{115} = 11.04$$

$$
\begin{aligned}
x_{0.5} &= x_{m;\ inf} + \frac{0.5 - F_{m-1}}{F_m - F_{m-1}}(x_{m;\ sup} - x_{m;\ inf}) \\
&= 5 + \frac{0.5 - 0.22}{0.61 - 0.22}\ (10 - 5) = 8.59 \\
\sigma_X^2 &= \frac{1}{115}20820 - (11.04)^2 \\
&= 170
\end{aligned}
$$

Mediana, Percentili e Moda

4

4.1 La Mediana

La *Mediana*, *Me* o $x_{0.5}$, è il valore centrale della serie dei dati riordinati in due: metà dei dati sono minori della mediana e metà dei dati sono maggiori della mediana. In simboli: sia $x_1, ..., x_n$ la serie dei dati e $x_{(1)}, ..., x_{(n)}$ i dati riordinati in modo crescente, allora:

1. se n è dispari
$$x_{0.5} = x_{\left(\frac{n+1}{2}\right)}$$
2. se n è pari
$$x_{0.5} = \frac{1}{2}\left(x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}\right)$$

Esempio 4.1.1. Sia
$$(x_1 = 2.9, x_2 = 3.5, x_3 = 1.2, x_4 = 2.7, x_5 = 4.2)$$
la serie dei dati, la serie dei dati riordinati sarà
$$(x_{(1)} = 1.2, x_{(2)} = 2.7, x_{(3)} = 2.9, x_{(4)} = 3.5, x_{(5)} = 4.2),$$
$n = 5$ è dispari e dunque
$$x_{0.5} = x_{\left(\frac{n+1}{2}\right)} = x_{\left(\frac{5+1}{2}\right)} = x_{(3)} = 2.9$$

Esempio 4.1.2. Sia
$$(x_1 = 2.9, x_2 = 3.5, x_3 = 1.2, x_4 = 2.7, x_5 = 4.2, x_6 = 4.2)$$
la serie dei dati, la serie dei dati riordinati sarà
$$(x_{(1)} = 1.2, x_{(2)} = 2.7, x_{(3)} = 2.9, x_{(4)} = 3.5, x_{(5)} = 4.2, x_{(6)} = 4.2)$$
$n = 6$ è pari e dunque
$$x_{0.5} = \frac{1}{2}\left(x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}\right) = \frac{1}{2}\left(x_{\left(\frac{6}{2}\right)} + x_{\left(\frac{6}{2}+1\right)}\right) = \frac{1}{2}\left(x_{(3)} + x_{(4)}\right) = \frac{2.9 + 3.5}{2} = 3.2$$

4.1.1 Dati espressi in distribuzione di frequenza

Se il fenomeno è espresso in una tabella di distribuzione di frequenza, allora la *modalità mediana* è la prima modalità tale per cui la frequenza cumulata è maggiore di 0.5

Esempio 4.1.3. Fenomeno *Titolo di Studio*, $n = 350$, numero di modalità $k = 5$.

j	x_j	n_j	f_j	F_j
1	Elementare	35	0.10	0.10
2	Media inferiore	105	0.30	0.40
3	Media Superiore	147	0.42	0.82
4	Laurea	35	0.10	0.92
5	Post Laurea	28	0.08	1.00

La modalità mediana è la terza $j = 3$, dunque *Media Superiore*, infatti $F_3 = 0.82 > 0.50$.

4.1.2 Dati espressi in classi

Se il fenomeno è espresso in classi, allora l'*intervallo mediano* è la primo intervallo tale per cui la frequenza cumulata è maggiore di 0.5.

Esempio 4.1.4.

Il reddito di $n = 4700$ famiglie è rappresentato nella seguente tabella di frequenza

j	[x$_j$,	x$_{j+1}$)	n_j	f_j	b_j	h_j	F_j
1	0	10	517	0.11	10	1.1	0.11
2	10	15	1269	0.27	5	5.4	0.38
3	15	20	1551	0.33	5	6.6	0.71
4	20	25	987	0.21	5	4.2	0.92
5	25	35	376	0.08	10	0.8	1.00
			4700	1.00	35		

La classe mediana è la terza classe $j = 3$, ovvero la classe [15,20), in quanto $F_3 = 0.71 > 0.50$.

Il *valore approssimato della mediana* è un valore che si trova all'interno dell'intervallo mediano e si ottiene dalla formula

$$x_{0.5} = x_{\text{inf};m} + \frac{0.5 - F_{m-1}}{f_m} \cdot \left(x_{\text{sup};m} - x_{\text{inf};m}\right),$$

dove m è l'indicatore della classe mediana, $x_{\text{inf};m}$ e $x_{\text{sup};m}$ sono, rispettivamente l'estremo inferiore e quello superiore dell'intervallo che contiene la mediana.

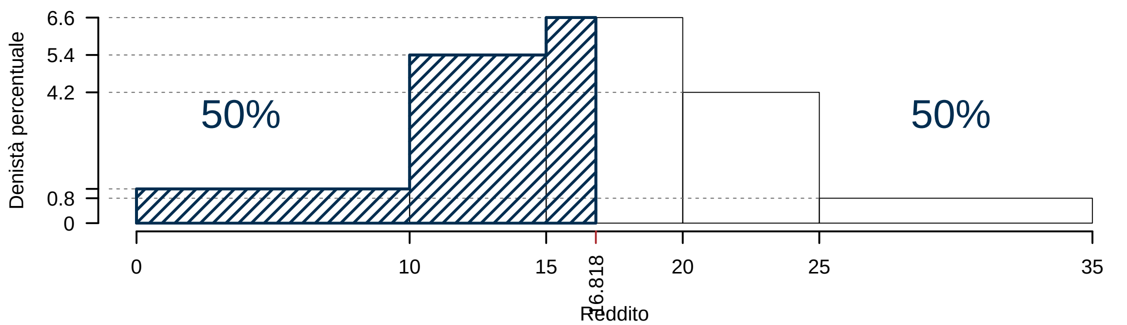
Esempio 4.1.5. Nell'esempio precedente l'intervallo mediano è [15,20) otterremo:

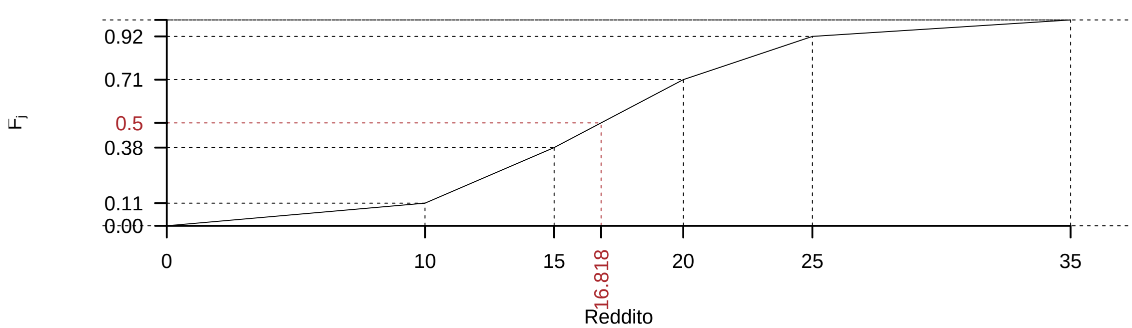
$$p = 0.5, \text{essendo } F_3 = 0.71 > 0.5 \Rightarrow j_{0.5} = 3$$

$$
\begin{aligned}
x_{0.5} &= x_{\inf;3} + \frac{0.5 - F_2}{f_3} \cdot b_3 \\
&= 15 + \frac{0.5 - 0.38}{0.33} \cdot 5 \\
&= 16.82
\end{aligned}
$$

Nota

La mediana è quel valore che taglia l'istogramma in due parti, entrambe di area pari al 50% dell'area totale





L'area tratteggiata è il 50% dell'area totale.

4.1.3 Proprietà della Mediana

Proprietà 4.1.1 (della Mediana). *La mediana di una distribuzione, $x_{0.5}$, è quel valore della per X il quale si ha $F(x_{0.5}) = 0.5$. Le proprietà della mediana ($x_{0.5}$) sono:*

1. $x_{\min} \le x_{0.5} \le x_{\max}$,
2. $\sum_{j=1}^{n} |x_j - x_{0.5}|$ *è un minimo.*
3. *Relazione Media-Mediana:*

- *Distribuzione simmetrica* $\rightarrow x_{0.5} = \bar{x}$
- *Distribuzione con coda lunga a destra* $\rightarrow x_{0.5} < \bar{x}$
- *Distribuzione con coda lunga a sinistra* $\rightarrow x_{0.5} > \bar{x}$

4.2 I Percentili

Il p-esimo percentile x_p, $0 \leq p \leq 1$, è qual valore che divide la serie dei dati riordinati in due: il $p \times 100\%$ dei dati sono minori di x_p e $(1-p) \times 100\%$ dei dati sono maggiori di x_p. Se per esempio $p = 0.30$ allora il trentesimo percentile è quel valore che ha il 30% dei dati inferiore il 70% dei dati superiore. Il p-esimo percentile di una serie di dati è il valore che occupa la posizione $\lfloor p \times n \rfloor + 1$, dove $\lfloor x \rfloor$ è l'operatore che estrae la parte intera di un numero decimale, ad esempio $\lfloor 3.001 \rfloor = \lfloor 3.21 \rfloor = \lfloor 3.94 \rfloor = 3$.

Esempio 4.2.1. Si considerino $n = 21$ osservazioni di una variabile categoriale ordinata che assume 7 valori: -2, -1, 0, 1, 2 (ad esempio una scala del tipo -2 =*in totale disaccordo*, -1 =*più in disaccordo che in accordo*, 0 =*né d'accordo, né in disaccordo*, 1 =*più d'accordo che in disaccordo*, 2 =*totalmente d'accordo*). Qui di seguito i dati riordinati:

(i)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
x_i	-2	-1	-1	-1	0	0	0	0	0	0	0	0	1	1	1	2	2	2	2	2	2

Il 15-esimo percentile è il dato che occupa la $\lfloor n \times p \rfloor + 1 = \lfloor 21 \times 0.15 \rfloor + 1 = 3 + 1 = 4$, e dunque il 15-esimo percentile è il quarto dato $x_{(4)} = -1$. È chiaro che la mediana è il 50-esimo percentile. In questo caso $x_{0.5} = x_{(\lfloor 21 \times 0.5 \rfloor + 1)} = x_{(11)} = 0$.

4.2.1 Dati espressi in distribuzione di frequenza

Se il fenomeno è espresso in una tabella di distribuzione di frequenza, allora il p-esimo percentile è la prima modalità tale per cui la frequenza cumulata è maggiore di p.

Esempio 4.2.2. Fenomeno *Titolo di Studio*, $n = 350$, numero di modalità $k = 5$.

j	x_j	n_j	f_j	F_j
1	Elementare	35	0.10	0.10
2	Media inferiore	105	0.30	0.40
3	Media Superiore	147	0.42	0.82
4	Laurea	35	0.10	0.92
5	Post Laurea	28	0.08	1.00

Il 90-esimo percentile $x_{0.90}$ è la quarta modalità, x_4 =Laurea.

Esempio 4.2.3. Fenomeno: Numero di volte che si è cercato lavoro negli ultimi 3 mesi, $n = 322$

j	x_j	n_j	f_j	F_j
1	0	20	0.0621	0.0621
2	1	47	0.1460	0.2081
3	2	80	0.2484	0.4565
4	3	64	0.1988	0.6553
5	4	46	0.1429	0.7981
6	5	36	0.1118	0.9099
7	6	16	0.0497	0.9596
8	7	9	0.0280	0.9876
9	8	3	0.0093	0.9969
10	10	1	0.0031	1.0000

Il 25-esimo percentile è la terza modalità in quanto $F_3 = 0.45 > 0.25$, $x_{0.25} = x_3 = 2$. Il 50-esimo percentile, la mediana, è $x_{0.5} = x_4 = 3$ e il 75-esimo percentile è $x_{0.75} = x_5 = 4$.

4.2.2 Dati espressi in classi

Se il fenomeno è espresso in classi, allora l'*intervallo che contiene il p-esimo percentile* è il primo intervallo tale per cui la frequenza cumulata è maggiore di p. Il *valore approssimato del percentile* è un valore che si trova all'interno dell'intervallo e si ottiene dalla formula

$$x_p = x_{\text{inf};j_p} + \frac{p - F_{j_p-1}}{f_{j_p}} \cdot \left(x_{\text{sup};j_p} - x_{\text{inf};j_p}\right)$$

dove j_p è l'indicatore della classe che contiene il p-esimo percentile, $x_{\text{inf};j_p}$ e $x_{\text{sup};j_p}$ sono, rispettivamente, l'estremo inferiore e quello superiore.

Esempio 4.2.4.

Il reddito di $n = 4700$ famiglie è rappresentato nella seguente tabella di frequenza

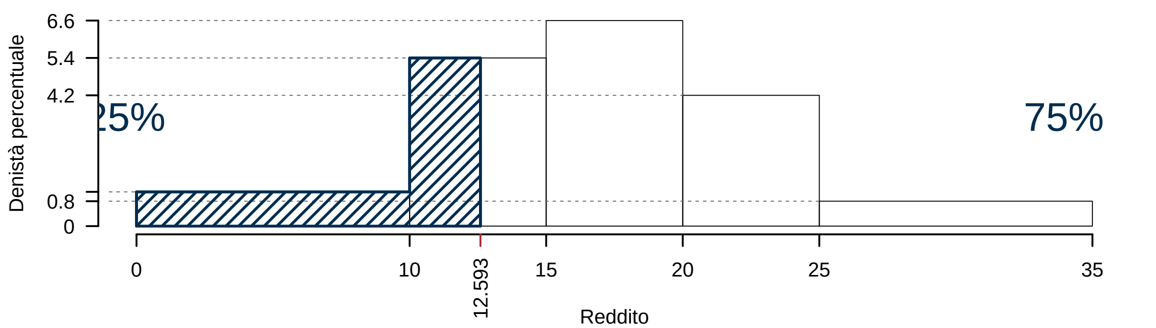
j	$[\mathrm{x}_j,$	$\mathrm{x}_{j+1})$	n_j	f_j	b_j	h_j	F_j
1	0	10	517	0.11	10	1.1	0.11
2	10	15	1269	0.27	5	5.4	0.38
3	15	20	1551	0.33	5	6.6	0.71
4	20	25	987	0.21	5	4.2	0.92
5	25	35	376	0.08	10	0.8	1.00
			4700	1.00	35		

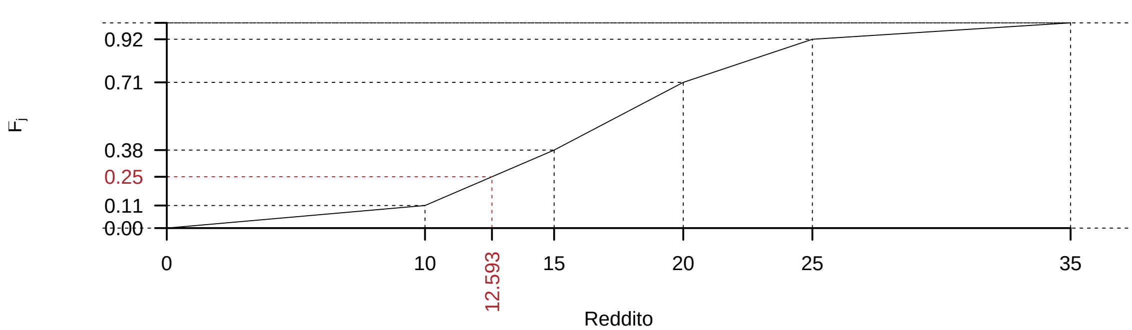
La classe che contiene il 25-esimo percentile è la seconda classe $j_{0.25} = 2$, ovvero la classe (10,15], in quanto $F_2 = 0.2081 > 0.25$.

$$
\begin{aligned}
p &= 0.25, \text{essendo } F_2 = 0.38 > 0.25 \Rightarrow j_{0.25} = 2 \\
x_{0.25} &= x_{\text{inf};2} + \frac{0.25 - F_1}{f_2} \cdot b_2 \\
&= 10 + \frac{0.25 - 0.11}{0.27} \cdot 5 \\
&= 12.59
\end{aligned}
$$

Nota

Il p-esimo percentile x_p è quel valore che taglia l'istogramma in due parti, l'area dell'istogramma alla sinistra di x_p è pari a $p \times 100\%$, mentre l'area la sua destra è $(1-p) \times 100\%$





L'area in blue è il 25% dell'area totale, quella in grigio il 75%.

4.2.3 I Quartili

Si definiscono **i quartili** della VS X, il 25-esimo, il 50.esimo e il 75-esimo percentile di X:

$$(x_{0.25}, x_{0.5}, x_{0.75})$$

4.2.4 Percentili e Funzione di Ripartizione

Se i dati sono quantitativi continui raccolti in classi e F è la funzione di ripartizione di X allora il percentile è quel valore tale che

$$F(x_p) = p$$

ovvero a sinistra di x_p c'è il $p \times 100\%$ dei dati e a destra di x_p il rimanente $(1-p) \times 100\%$. Per esempio sappiamo nel caso studiato sopra che $x_{0.25} = 12.593$ e quindi $F(12.593) = 0.25$.

Ogni valore di X dal suo minimo al suo massimo è un percentile, per esempio il valore 15 è il 38-esimo percentile di X ($x_{0.38} = 15$), infatti il 38% dei dati è inferiore a 15:

$$F(15) = F_2 = 0.38$$

Mentre la funzione inversa $Q = F^{-1}$ è la funzione che ci permette di calcolare il percentile di ordine p:

$$Q(p) = x_p.$$

Per esempio

$$Q(0.25) = x_{0.25} = 12.593$$

Questa applicazione interattiva aiuta a comprendere meglio la relazione tra istogramma e funzione di ripartizione: La Funzione di Ripartizione

4.3 Lo Scarto Interquartile

Una misura di variabilità è lo scarto interquartile

$$SI = x_{0.75} - x_{0.25}$$

4.4 La Moda

Si definisce la **moda**, x_{Mo} la modalità cui compete frequenza maggiore.

Esempio 4.4.1. Consideriamo la distribuzione del colore dei capelli

x$_j$	Cast.	Biondi	Rossi	Tot
n_j	245	68	13	326

La modalità modale (la moda) è x_{Mo}=Castano.

Esempio 4.4.2. Titolo di studio:

x$_j$	Prim.	M. inf.	M. sup.	Univ.	Post univ.	Tot
n_j	10	18	158	62	12	260

La modalità modale (la moda) è x_{Mo} =M. sup.

Esempio 4.4.3.

0	1	2	3	4
1	11	6	8	5

La modalità modale è x = 1 e osserviamo che la media è 2.1613 e la mediana è 2

4.4.1 La Moda per dati raccolti in classi

Se i dati sono sono raccolti in classi, non c'è un valore modale ma una classe modale ed è la **classe cui compete densità maggiore**.

Esempio 4.4.4.

j	$[x_j,$	$x_{j+1})$	n_j	f_j	b_j	h_j	F_j
1	0	10	517	0.11	10	1.1	0.11
2	10	15	1269	0.27	5	5.4	0.38
3	15	20	1551	0.33	5	6.6	0.71
4	20	25	987	0.21	5	4.2	0.92
5	25	35	376	0.08	10	0.8	1.00
			4700	1.00	35		

la classe modale è la terza classe, la classe $[15, 20)$

4.5 Relazione tra Media, Moda e Mediana

Se la VS X ha una sola classe modale, allora valgono le seguenti relazioni:

- Se la distribuzione presenta un'asimmetria negativa (coda lunga a sx) allora

$$\bar{x} \leq x_{0.5} \leq x_{mo}$$

- Se la distribuzione è simmetrica allora

$$x_{mo} \approx x_{0.5} \approx \bar{x}$$

- Se la distribuzione presenta un'asimmetria positiva (coda lunga a dx) allora

$$x_{mo} \leq x_{0.5} \leq \bar{x}$$

La figura 4.1 ne offre una rappresentazione grafica.

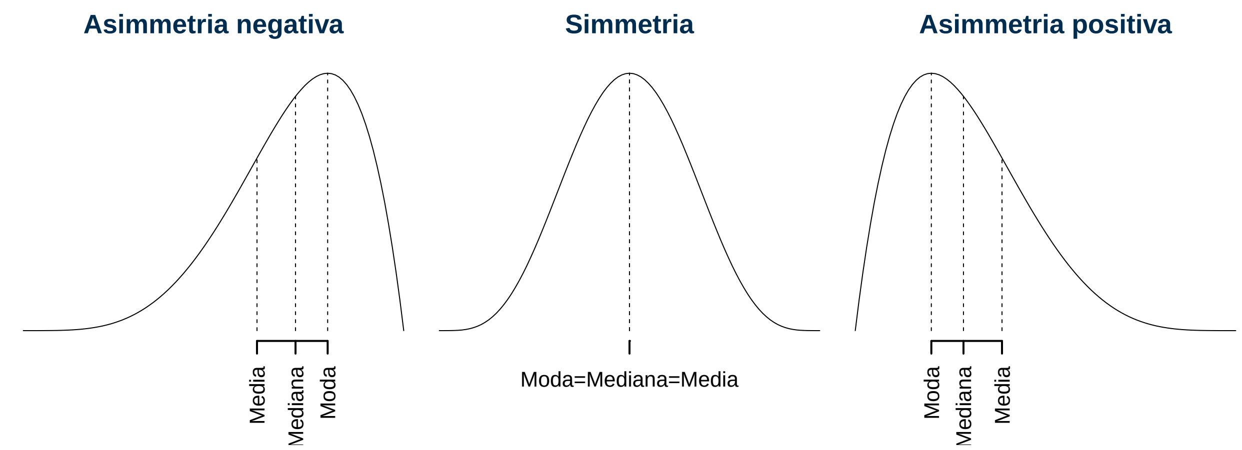


Figura 4.1: Relazione tra media mediana e moda

4.6 Istogramma e Percentili

La relazione tra istogramma di densità e percentili è evidente, i percentile di ordine p indica per quale valore di x_p l'area dell'istogramma fino misurano l'area dell'istogramma fino ad x_p vale p.

Attraverso il seguente esempio osserveremo meglio il legame tra i due concetti.

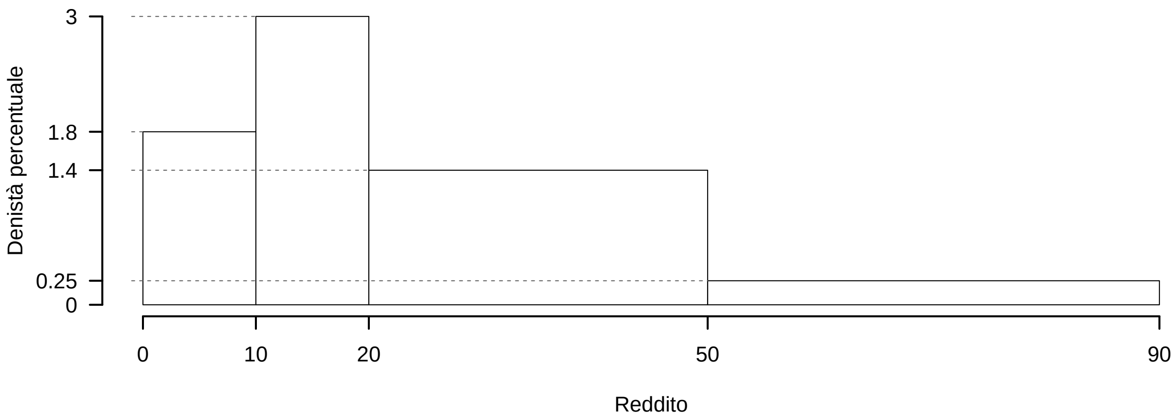
Esempio 4.6.1. Il reddito di $n = 4700$ famiglie (dati inventati) è rappresentato nella tabella di frequenza qui di seguito

$[\mathrm{x}_j,$	$\mathrm{x}_{j+1})$	f_j
0	10	0.18
10	20	0.30
20	50	0.42
50	90	0.10
		1.00

Per prima cosa calcoliamo tutta la tabella: la colonna delle frequenza assolute e cumulate, delle densità, ecc.

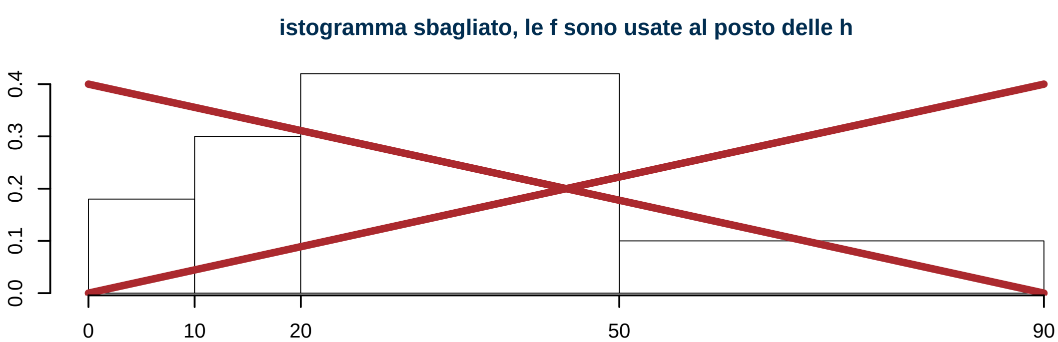
$[\mathrm{x}_j,$	$\mathrm{x}_{j+1})$	n_j	f_j	b_j	h_j	F_j	$\bar{\mathrm{x}}_j$	$\bar{\mathrm{x}}_j^2$	$\bar{\mathrm{x}}_j n_j$	$\bar{\mathrm{x}}_j^2 n_j$	$f_{j\%}$
0	10	846	0.18	10	1.80	0.18	5	25	4230	21150	18
10	20	1410	0.30	10	3.00	0.48	15	225	21150	317250	30
20	50	1974	0.42	30	1.40	0.90	35	1225	69090	2418150	42
50	90	470	0.10	40	0.25	1.00	70	4900	32900	2303000	10
		4700	1.00	90					127370	5059550	100

Quindi disegniamo l'istogramma di densità percentuale, dove le h sono usate come altezze e le $f_{\%}$ sono le aree dei rettangoli.



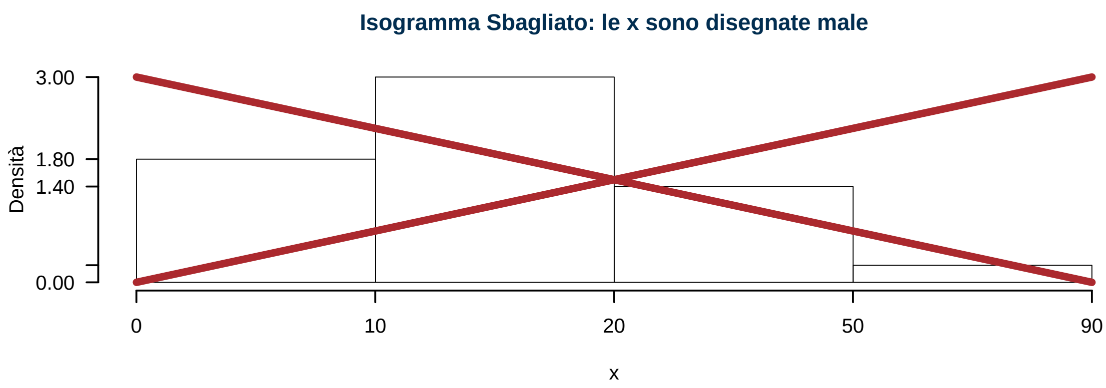
Attenzione

Se al posto delle h usiamo le f disegniamo l'istogramma *sbagliato*, . Notiamo che la classe $[20, 50)$ viene sovra rappresentata. Infatti è vero che il 45% delle famiglie si trova in quella classe, ma è anche vero che l'ampiezza della classe molto grande e l'istogramma non rappresenta i dati.



Attenzione

Per disegnare un istogramma di densità in modo corretto si devono rispettare le proporzioni tra le basi



Una volta che abbiamo ricostruito tutta la tabella e realizzato il grafico possiamo rispondere a tante domande sulla distribuzione dei dati, quali, per esempio:

- Individuare la classe modale

La classe modale è la classe [10,20) non la classe 20-50. Infatti la classe modale è la classe con densità maggiore, non la classe con frequenza maggiore.

- Individuare la classe che contiene la mediana

La classe mediana è la terza classe [20,50), infatti $F_3 = 0.90$ è il primo degli F_j che super 0.50.

- Calcolare $x_0, x_{0.10}, x_{0.20}, ..., x_{0.90}, x_1$

$$x_p = x_{\inf;j_p} + \frac{p - F_{j_p-1}}{f_{j_p}} b_{j_p}$$

$$x_0 = 0, \quad \text{il più piccol dei dati}$$

$$\begin{aligned}
p &= 0.1, \text{essendo } F_1 = 0.18 > 0.1 \Rightarrow j_{0.1} = 1 \\
x_{0.1} &= x_{\inf;1} + \frac{0.1 - F_0}{f_1} \cdot b_1 \\
&= 0 + \frac{0.1 - 0}{0.18} \cdot 10 \\
&= 5.556
\end{aligned}$$

$$\begin{aligned}
p &= 0.2, \text{essendo } F_2 = 0.48 > 0.2 \Rightarrow j_{0.2} = 2 \\
x_{0.2} &= x_{\inf;2} + \frac{0.2 - F_1}{f_2} \cdot b_2 \\
&= 10 + \frac{0.2 - 0.18}{0.3} \cdot 10 \\
&= 10.67
\end{aligned}$$

$$\vdots$$

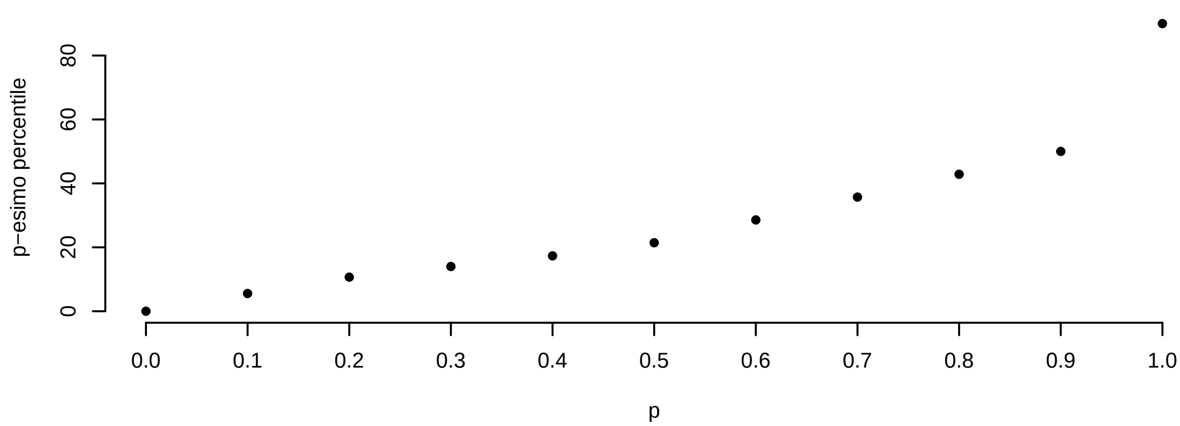
$$\begin{aligned}
p &= 0.9, \text{essendo } F_4 = 1 > 0.9 \Rightarrow j_{0.9} = 4 \\
x_{0.9} &= x_{\inf;4} + \frac{0.9 - F_3}{f_4} \cdot b_4 \\
&= 50 + \frac{0.9 - 0.9}{0.1} \cdot 40 \\
&= 50
\end{aligned}$$

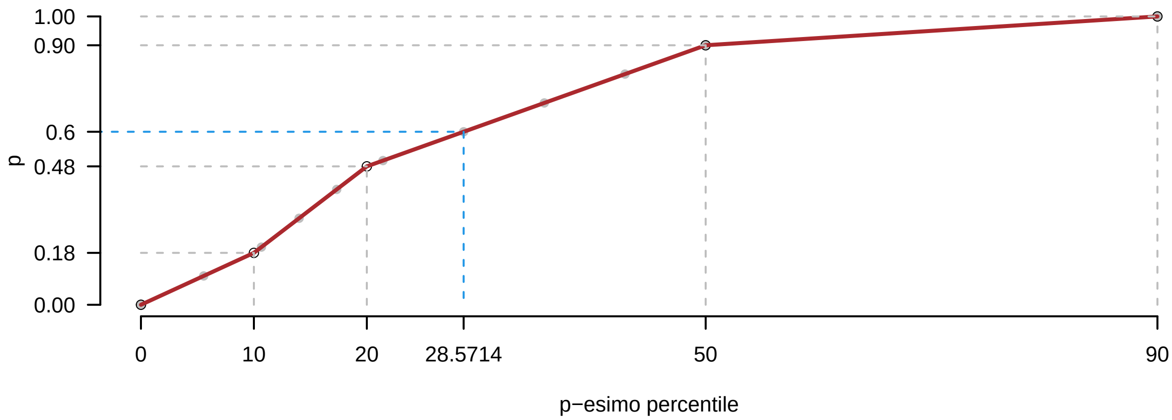
$$x_{1.0} = 90, \quad \text{il più grande dei dati}$$

Riassumendo in tabella otteniamo:

	x_0	$x_{0.1}$	$x_{0.2}$	$x_{0.3}$	$x_{0.4}$	$x_{0.5}$	$x_{0.6}$	$x_{0.7}$	$x_{0.8}$	$x_{0.9}$	x_1
Percentili	0	5.556	10.67	14	17.33	21.43	28.57	35.71	42.86	50	90

- mettere a grafico i punti $(0,x_0)$, $(0.10,x_{0.10})$,...,$(1,x_1)$) e $(x_0,0)$, $(x_{0.10},0.10)$,...,$(x_1,1)$)





- calcolare la percentuale di individui con reddito inferiore a 50

90% è la percentuale di individui con reddito minore di 50

- calcolare la percentuale di individui con reddito superiore a 20
  - 48% è la percentuale di individui con reddito minore di 20
  - 52% è la percentuale di individui con reddito maggiore di 20

- calcolare la percentuale approssimata di individui con reddito inferiore a 14

Per calcolare $\%(X < 14)$ abbiamo diversi modi, anzi tutto notiamo che la percentuale di $dat2$ minori di 14 è, approssimativamente, l'area dell'istogramma da zero a 14. E quindi, direttamente:

$$\begin{aligned} \%(X < 14) &= f_1 \times 100 + (14 - 10) \times h_2 \\ &= 0.18 \times 100 + 4 \times 3 \\ &= 30\% \end{aligned}$$

In modo del tutto analogo con la funzione di ripartizione F

$$\begin{aligned} \%(X < 14) &= 100 \times F(14) \\ &= 100 \times (f_1 + (14 - 10) \times h_2/100) \\ &= 100 \times (0.18 + 4 \times 3/100) \\ &= 30\% \end{aligned}$$

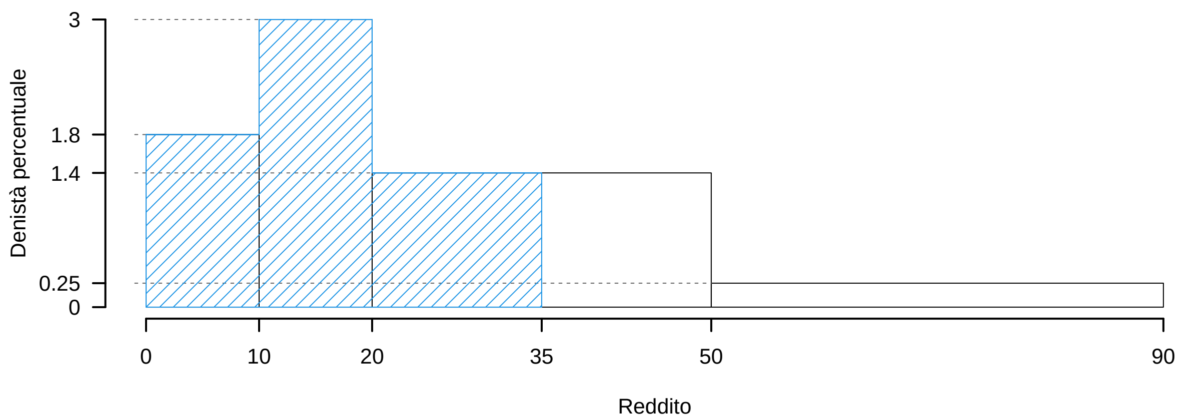
Ma anche notando che $x_{0.30} = 14$, significa che $F(x_{0.30}) = 0.30$ e $\%(X < 14) = 100 \times F(x_{0.30}) = 30\%$.

- Calcolare la percentuale approssimata di individui con reddito superiore a 28.57

Notiamo che $x_{0.60} = 28.57$ e quindi $F(x_{0.60}) = 0.60$, quindi $\%(X < 28.57) = 0.60 \times 100 = 60\%$ e quindi $\%(X > 28.57) = 40\%$

- Calcolare la percentuale approssimata di individui con reddito inferiore a 35

Si tratta di calcolare l'area in blu:

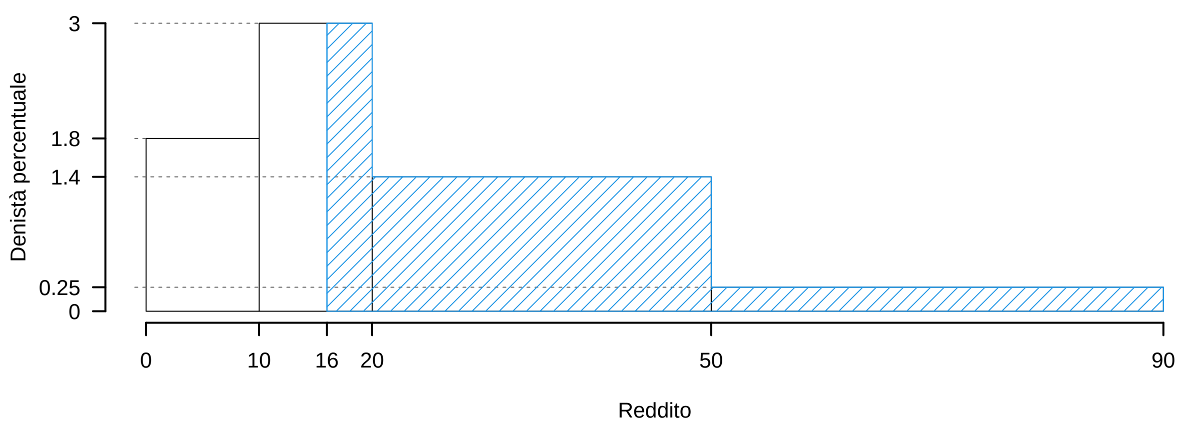


e osservare che la sua area misura 69. Infatti l'area è data dalla somma di

$$\begin{aligned}\%(X \leq 35) &= f_1 \times 100 + f_2 \times 100 + (35 - 20) \times h_3 \\ &= 18 + 30 + 15 \times 1.4 \\ &= 69\end{aligned}$$

- calcolare la percentuale approssimata di individui con reddito superiore a 16

Si tratta di calcolare l'area in blu:



e osservare che la sua area misura 64. Infatti l'area si può vedere o direttamente calcolandola da 16 a 100 oppure si può valutarla come complemento:

$$\%(X > 16) = 100\% - \%(X \leq 16)$$

e quindi:

$$\%(X \leq 16) = f_1 \times 100 + (16 - 10) \times h_2 = 36$$

Quindi $\%(X > 16) = 100 - 36 = 64$

- Individuare la media approssimata, la varianza approssimata, e i quartili approssimati

La media:

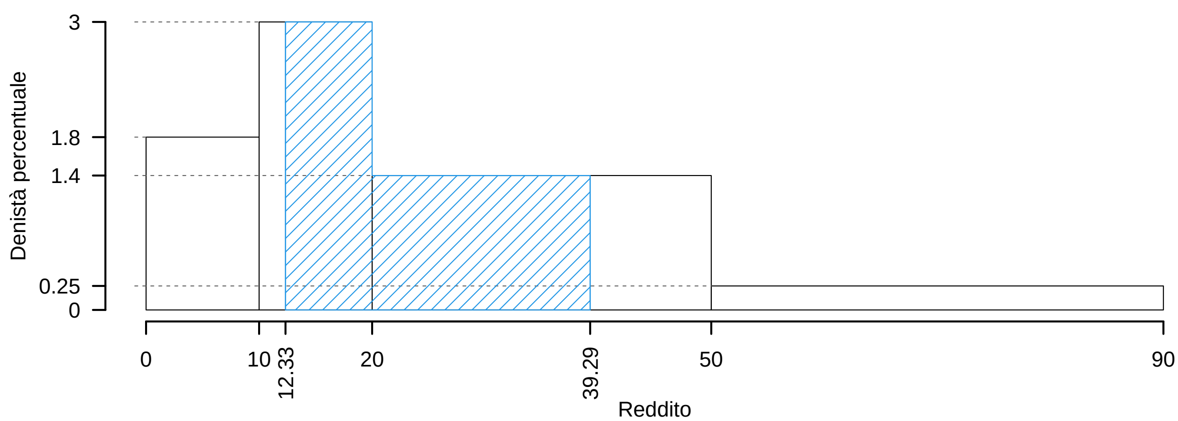
$$\bar{x} = \frac{1}{n} \sum_j x_{jc} n_j = \frac{1}{4700} 127370 = 27.1$$

La varianza:

$$Var(X) = \frac{\sum_j x_{jc}^2 n_j}{n} - \bar{x}^2 = \frac{1}{4700} 5059550 - (27.1)^2 = 342.09$$

$\bar{x} = 27.1$, $Var = 342.09$, $(x_{0.25}, x_{0.50}, x_{0.75}) = (12.3333,\ 21.4286,\ 39.2857)$

- Calcolare la percentuale di dati compresi tra il 25-esimo e il 75-esimo percentile



- calcolare $x_{0.025}$ e $x_{0.975}$

i valori sono $(x_{0.025}, x_{0.975})$ =(1.3889, 80)

- Rappresentare graficamente e calcolare la percentuale di famiglie con reddito:
  - compreso tra 1.3889 e 80
  - compreso tra 0 e 1.3889
  - compreso tra 1.3889 e 21.4286
  - compreso tra 21.4286 e 80
  - compreso tra 80 e 100
  - minore di 1.3889 o maggiore di 80

Parte II
Probabilità

Cenni di Teoria della probabilità 5

5.1 Concetti di base

La definizione più moderna che possiamo dare è che

La probabilità è una misura dell'incertezza

che un osservatore *razionale* esprime sull'accadibilità di un evento.

Questa definizione consente di ampliare i campi di applicazione della teoria della probabilità oltre il mero calcolo dei giochi d'azzardo. Ma per arrivare a questa definizione molto ampia, matematici, scienziati e filosofi discutono di caso, caos, disordine, frequenza, probabilità, possibilità, verosimiglianza, ecc. da diversi secoli. La storia della filosofia e del calcolo della probabilità è affascinate e complessa ed esula dagli obiettivi di questi appunti. Rimando ai più curiosi il libro di Costantini e Geymonat:

D. Costantini e L. Geymonat, Filosofia della probabilità, Feltrinelli, Milano 1982.

5.1.1 Eventi

Prima di definire cos'è la misura di *probabilità* dobbiamo definire a cosa si applichi. La probabilità, nell'accezione che daremo, è un numero compreso tra 0 ed 1 che gradua quanto è credibile l'accadibilità di un fatto (un *evento*).

Un evento E è un *fatto* (potenzialmente verificabile), espresso nel linguaggio comune, che non sappiamo se è vero o se è falso:

- uscirà 6 dal lancio di un dado (perfetto).
- uscirà testa dal lancio di una moneta non regolare.
- Domani pioverà.
- l'indice Dow Jones tra un'ora quoterà 35 000.

Dicendo che il *fatto* deve essere *potenzialmente verificabile* escludiamo tutti quegli eventi che sono fuori dalla portata dei nostri sensi. Ovvero speculare sull'esistenza di Dio o della vita dopo la morte o di scenari ipotetici nel passato non è compito della probabilità.

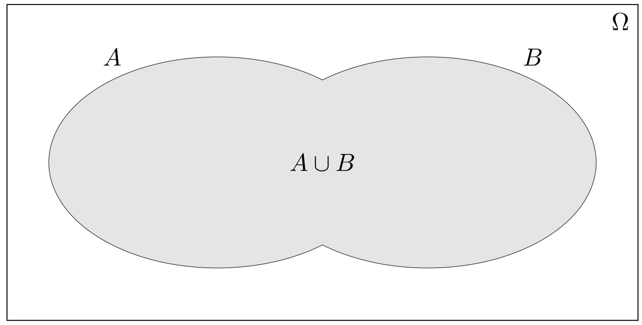
5.1.2 Algebra degli eventi

Usiamo gli operatori dell'insiemistica per combinare gli eventi tra di loro, trasformando le operazioni sintattiche del linguaggio comune in unioni ed intersezioni di eventi.

Definizione 5.1.1 (Unione tra Eventi)**.** Siano A e B due eventi, l'espressione

$$A \cup B$$

è vera se **almeno uno dei due** è vero.



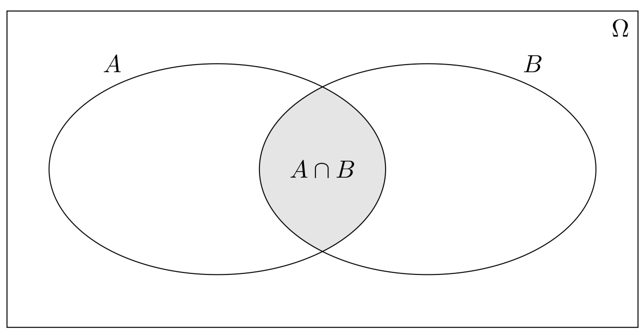
Esempio 5.1.1.

$$\begin{aligned} A &= \text{Domani pioverà} \\ B &= \text{Domani ci sarà traffico} \end{aligned}$$

$A \cup B$ sarà vera se domani pioverà **o** ci sarà traffico **o** ci sarà pioggia e traffico.

Definizione 5.1.2 (Intersezione tra Eventi)**.** Siano A e B due eventi, l'espressione

$A \cap B$ è vera se è vero A **e** è vero B.

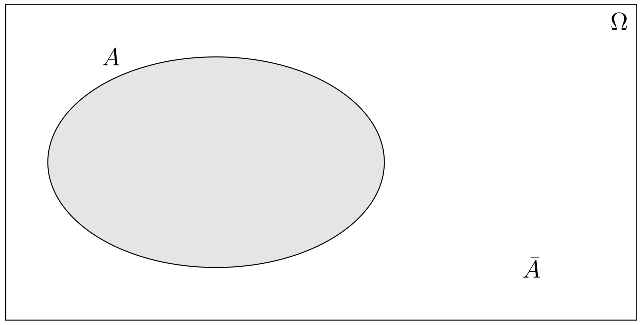


Esempio 5.1.2.

$$\begin{aligned} A &= \text{Domani pioverà} \\ B &= \text{Domani ci sarà traffico} \end{aligned}$$

$A \cap B$ sarà vera se domani pioverà **e** ci sarà traffico.

Definizione 5.1.3 (Evento Complementare)**.** Sia A un evento, si definisce $\bar{A}$ l'evento complementare di A



Esempio 5.1.3.

$$
\begin{aligned}
A &= \text{Domani piover\`a} \\
\bar{A} &= \text{Domani } \textbf{non} \text{ piover\`a}
\end{aligned}
$$

domani **non** pioverà

Definizione 5.1.4 (Evento Certo). Sia A un evento, si definisce l'evento certo Ω l'evento:

$$\Omega = A \cup \bar{A}$$

Esempio 5.1.4.

$$
\begin{aligned}
A &= \text{Domani piover\`a} \\
\bar{A} &= \text{Domani } \textbf{non} \text{ piover\`a} \\
\Omega &= \text{Domani o piover\`a o non piover\`a}
\end{aligned}
$$

Definizione 5.1.5 (Evento Impossibile). Sia A un evento, si definisce l'evento certo $\emptyset$ l'evento:

$$\emptyset = A \cap \bar{A}$$

Esempio 5.1.5.

$$
\begin{aligned}
A &= \text{Domani piover\`a} \\
\bar{A} &= \text{Domani } \textbf{non} \text{ piover\`a} \\
\emptyset &= \text{Domani piover\`a } \textbf{e} \text{ non piover\`a}
\end{aligned}
$$

5.1.3 Operazioni su insieme

Nota

Gli operatori Unione $\cup$ e Intersezione $\cap$ si comportano sugli insiemi come somme $+$ e moltiplicazione $\times$ si comportano sui numeri. In particolare

$$
\begin{aligned}
A \cup B &= B \cup A, && \text{prorietà commutativa}\\
A \cap B &= B \cap A, && \text{prorietà commutativa}\\
(A \cup B) \cup C &= A \cup (B \cup C), && \text{prorietà associativa}\\
(A \cap B) \cap C &= A \cap (B \cap C), && \text{prorietà associativa}\\
(A \cup B) \cap C &= (A \cap C) \cup (B \cap C), && \text{prorietà distributiva}\\
A \cup B \cap C &= A \cup (B \cap C), && \text{l'intersezione ha priorità sull'unione...}\\
(A \cup B) \cap C &\neq A \cup (B \cap C), && \text{a meno di opportune parentesi}
\end{aligned}
$$

5.1.4 La probabilità è una funzione

La probabilità assegna ad ogni evento il grado di credibilità

$$P(A)$$

indica la probabilità che l'evento A sia vero. In particolare

$$P(A) = P(\Omega) = 1$$

se A è un evento **certo**

$$P(A) = P(\emptyset) = 0$$

se A è un evento **impossibile** e in generale

$$0 \leq P(A) \leq 1$$

5.1.5 Definizioni di probabilità

- approccio classico (Laplace)
- approccio frequentista
- approccio soggettivista

Definizione 5.1.6 (Approccio Classico (Laplace))**.** la probabilità è il rapporto tra il numero dei casi favorevoli e il numero dei casi possibili, posto che gli eventi siano tutti equiprobabili.

$$P(A) = \frac{\#(\text{casi favorevoli ad } A)}{\#(\text{casi totali})}$$

Esempio 5.1.6. Un'urna contiene 5 sfere Rosse, 3 sfere Blu e 2 Nere. La probabilità dell'evento

$$R = \text{Estraggo una Rossa}$$

è data da

$$P(R) = \frac{5}{5+3+2} = 0.5$$

Definizione 5.1.7 (Approccio Frequentista). **Postulato empirico del caso.**
In un gruppo di prove ripetute più volte *nelle stesse condizioni*, ciascuno degli eventi possibili si presenta con una frequenza relativa che tende alla probabilità all'aumentare del numero di prove; ossia

$$P(A) = \frac{n_A}{n} + \epsilon_n \quad \text{dove} \quad \epsilon_n \to 0 \quad \text{per} \quad n \to \infty.$$

Esempio 5.1.7. Viene lanciata 200 vole una moneta (che non sappiamo se è bilanciata oppure no), Si è osservato che 136 su 200 lanci sono TESTA La probabilità dell'evento

$$T = \text{“Esce Testa”}$$

è data da

$$P(T) \approx \frac{136}{200} = 0.68$$

Viene lanciata 2000 vole una moneta (che non sappiamo se è bilanciata oppure no), Si è osservato che 1360 su 2000 lanci sono TESTA La probabilità dell'evento

$$T = \text{“Esce Testa”}$$

è data da

$$P(T) \approx \frac{1360}{2000} = 0.68$$

Ma saremo più *sicuri* che avendo lanciato solo 200 volte.

Definizione 5.1.8 (Approccio Soggettivista).

L'approccio soggettivista alla probabilità, sviluppato da Bruno de Finetti, si differenzia dagli approcci classico e frequentista. Secondo de Finetti, la probabilità non è una proprietà intrinseca degli eventi, ma rappresenta il grado di credenza o fiducia che un individuo assegna all'accadimento di un certo evento. In questo contesto, la probabilità è personale e varia da soggetto a soggetto, riflettendo l'informazione, l'esperienza e il giudizio personale.

La probabilità soggettiva si esprime attraverso le scommesse. Per esempio, dire che un evento ha probabilità del 60% equivale a dire che si è disposti a scommettere con un rapporto di 3:2 in favore dell'evento, sia come scommettitore che come allibratore. Questa visione mette in luce come le probabilità siano strettamente legate alle decisioni e alle aspettative personali, e come possano essere aggiornate in base a nuove informazioni (teoria Bayesiana).

L'approccio soggettivistico è particolarmente utile in situazioni dove i dati sono limitati o dove è difficile definire una frequenza a lungo termine, come nelle previsioni meteorologiche o nelle valutazioni di rischio in ambiti finanziari o assicurativi.

5.2 Teoria di Kolmogorov

La teoria di Kolmogorov è una teoria matematica, estremamente formalizzata, che non si preoccupa di assegnare le probabilità agli eventi ma alle regole formali per assegnarla senza cadere in contraddizioni. La teoria muove da 3 assiomi che andremo ad elencare tra poco, di estremo *buon senso* e muove per definizioni e teoremi verso strumenti di grande aiuto nella soluzione di problemi concreti.

Definizione 5.2.1. Sia Ω l'evento certo, Si definisce uno spazio probabilizzato $\{\Omega, \mathscr{A}, P\}$, dove $\mathscr{A}$ è un'algebra costruita su Ω e P è una misura di probabilità.

5.2.1 Algebra degli Eventi

L'algebra degli eventi è una particolare collezione di sottoinsiemi di Ω che siamo interessati a probabilizzare. È uno spazio astratto che etichetta tutte le i possibili eventi che possiamo costruire a partire da alcuni eventi di partenza. Un'algebra degli eventi viene solitamente indicata con una lettera calligrafica corsiva. In queste pagine useremo il carattere tipografico $\mathscr{A}$.

$\mathscr{A}$: un insieme di sotto insiemi di Ω, che contiene l'insieme vuoto, tutto Ω ed è chiuso rispetto alle unioni e alle intersezioni e al passaggio al complementare: se $A \in \mathscr{A}$ e $B \in \mathscr{A}$ allora

$$\begin{aligned} A \cup B &\in \mathscr{A} \\ A \cap B &\in \mathscr{A} \\ \bar{A} &\in \mathscr{A} \end{aligned}$$

Esempio 5.2.1. Per illustrare l'idea di un'algebra degli eventi con un esempio finito, consideriamo un insieme Ω che rappresenta lo spazio campionario di un semplice lancio di due monete. Ω è dato da tutte le possibili combinazioni dei risultati di due lanci di moneta: $\Omega = (\text{T, T}), (\text{T, C}), (\text{C, T}), (\text{C, C})$.

Certo, consideriamo ora il lancio di due monete. In questo caso, lo spazio campionario Ω è dato da tutte le possibili combinazioni dei risultati di due lanci di moneta: $\Omega = \{(\text{T, T}), (\text{T, C}), (\text{C, T}), (\text{C, C})\}$.

Costruiamo un'algebra degli eventi $\mathscr{A}$ per questo esperimento. Questa algebra potrebbe includere i seguenti sottoinsiemi di Ω:

1. L'insieme vuoto $\emptyset$, che rappresenta l'evento impossibile.
2. L'intero insieme Ω, che rappresenta l'evento certo.
3. Singoli elementi come $\{(\text{T, T})\}$, $\{(\text{T, C})\}$, $\{(\text{C, T})\}$, $\{(\text{C, C})\}$, che rappresentano gli eventi di ottenere specifiche combinazioni.

4. Combinazioni di questi eventi, come $\{(\text{T, T}), (\text{T, C})\}$, che rappresenta l'evento in cui la prima moneta mostra T, indipendentemente dal risultato della seconda moneta.

Questi sottoinsiemi rispettano le proprietà di un'algebra degli eventi:

- Contengono l'insieme vuoto e l'intero insieme Ω.
- Sono chiusi rispetto alle operazioni di unione, intersezione e passaggio al complementare. Ad esempio:
  - Unione: $\{(\text{T, T})\} \cup \{(\text{C, C})\} = \{(\text{T, T}), (\text{C, C})\} \in \mathscr{A}$.
  - Intersezione: $\{(\text{T, T}), (\text{T, C})\} \cap \{(\text{T, T}), (\text{C, T})\} = \{(\text{T, T})\} \in \mathscr{A}$.
  - Complementare: $\overline{\{(\text{T, T})\}} = \{(\text{T, C}), (\text{C, T}), (\text{C, C})\} \in \mathscr{A}$.

In questo modo, l'algebra degli eventi $\mathscr{A}$ cattura tutte le possibili combinazioni di eventi che possono verificarsi nel contesto di due lanci di moneta.

In definitiva

$$
\begin{aligned}
\mathscr{A} = \{&\emptyset, \\
&(\text{T, T}), (\text{T, C}), (\text{C, T}), (\text{C, C}) \\
&\{(\text{T, T}) \cup (\text{T, C})\}, \{(\text{T, T}) \cup (\text{C, T})\}, \{(\text{T, T}) \cup (\text{C, C})\}, \\
&\{(\text{T, C}) \cup (\text{C, T})\}, \{(\text{T, C}) \cup (\text{C, C})\}, \{(\text{C, T}) \cup (\text{C, C})\}, \\
&\{(\text{T, T}) \cup (\text{T, C}) \cup (\text{C,T})\}, \{(\text{T, T}) \cup (\text{T, C}) \cup (\text{C, C})\}, \\
&\{(\text{T, T}) \cup (\text{C, T}) \cup (\text{C,C})\}, \{(\text{T, C}) \cup (\text{C, T}) \cup (\text{C, C})\}, \\
&\Omega\}
\end{aligned}
$$

5.2.2 Assiomi di Kolmogorov

La probabilità P è una funzione che trasforma ogni evento A di $\mathscr{A}$ in un numero reale

$$P : \mathscr{A} \to \mathbb{R}, \quad \forall A \in \mathscr{A}$$

Tale che

i. $P(A) \geq 0, \ \forall A \in \mathscr{A}$

ii. $P(\Omega) = 1$

iii. $\forall A, B \in \mathscr{A} : A \cap B = \emptyset, P(A \cup B) = P(A) + P(B)$

La forza della teoria consiste nel ricavare tutti i risultati partendo da questi 3 assiomi.

5.2.3 Proprietà di P

Dagli assiomi precedenti possiamo ricavare diverse proprietà interessanti di P che non sono scritte in modo esplicito negli assiomi ma si ricavano per dimostrazione.

Proprietà 5.2.1 (Proprietà Principali di P). *Tra le tante enunciamo le più immediate ed utili:*

1. $0 \leq P(A) \leq 1, \ \forall A \in \mathscr{A}$
2. $P(\emptyset) = 0$
3. $P(A) = 1 - P(\bar{A})$
4. $P(A \cap B) = P(A) - P(A \cap \bar{B})$
5. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Dimostrazione. Nell'ordine

1. Dall'assioma *i* sappiamo che
$$P(A) > 0$$
e dall'assioma *ii* che
$$P(\Omega) = 1$$
Siccome
$$A \cup \bar{A} = \Omega$$
in virtù dell'assioma *iii* otteniamo
$$\begin{array}{rcl} P(A \cup \bar{A}) & = & P(\Omega) \\ P(A) + P(\bar{A}) & = & 1 \\ P(A) & \leq & 1 \end{array}$$

2. Siccome il complementare di Ω è l'insieme vuoto
$$\bar{\Omega} = \emptyset$$
dalla 1. sappiamo che la probabilità è compresa tra zero ed 1 e quindi
$$\begin{array}{rcl} P(\Omega \cup \emptyset) & = & P(\Omega) \\ P(\Omega) + P(\emptyset) & = & 1 \\ P(\emptyset) = 0 & & \end{array}$$

3. Siccome
$$A \cup \bar{A} = \Omega$$
in virtù dell'assioma *iii* otteniamo
$$\begin{array}{rcl} P(A \cup \bar{A}) & = & P(\Omega) \\ P(A) + P(\bar{A}) & = & 1 \\ P(A) & = & 1 - P(\bar{A}) \end{array}$$

4. Osserviamo che
$$A = (A \cap B) \cup (A \cap \bar{B})$$
e quindi
$$\begin{aligned} P(A) &= P((A \cap B) \cup (A \cap \bar{B})) \\ P(A) &= P(A \cap B) + P(A \cap \bar{B}) \\ P(A \cap B) &= P(\bar{A}) - P(A \cap \bar{B}) \\ P(A \cap \bar{B}) &= P(\bar{A}) - P(A \cap B) \end{aligned}$$

5. Notiamo che
$$A \cup B = (A \cap \bar{B}) \cup (B \cap \bar{A}) \cup (A \cap B)$$
cioè $A \cup B$ si può riscrivere come l'unione di tre eventi disgiunti $(A \cap \bar{B})$, $(B \cap \bar{A})$ e $\cup(A \cap B)$. e quindi
$$\begin{aligned} P(A \cup B) &= P((A \cap \bar{B}) \cup (B \cap \bar{A}) \cup (A \cap B)) \\ P(A \cup B) &= P(A \cap \bar{B}) + P(B \cap \bar{A}) + P(A \cap B) \\ P(A \cup B) &= P(A) - P(A \cap B) + P(B) - P(B \cap A) + P(A \cap B) \\ P(A \cup B) &= P(A) + P(B) - P(B \cap A) \end{aligned}$$
in figura 5.1 una rappresentazione grafica.

□

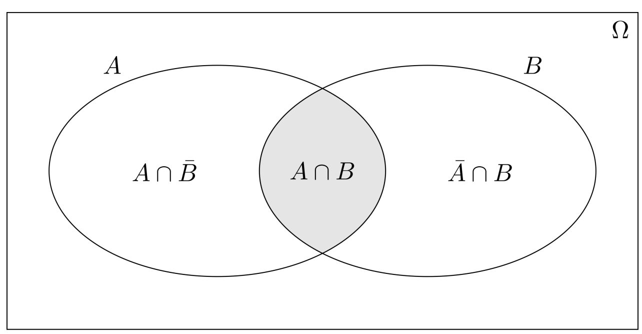


Figura 5.1: Regola di De Morgan per due insiemi: la probabilità dell'unione è la somma della probilità di tre eventi disgiunti

Definizione 5.2.2 (Eventi Incompatibili). A e B si dicono **incompatibili** se e solo se
$$A \cap B = \emptyset$$
in figura 5.2 una rappresentazione grafica.

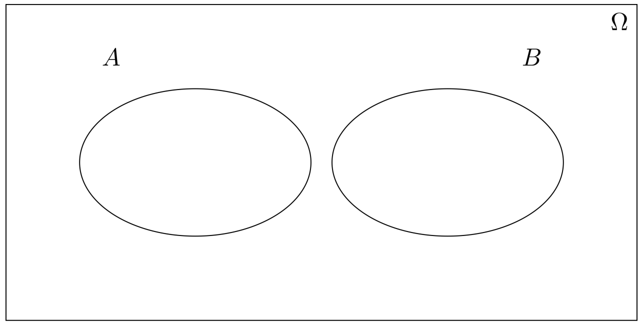


Figura 5.2: Due eventi si dicono incompatibili se sono disgiunti

Esempio 5.2.2. Un'urna ha 8 palline bianche numerate da 1 a 8 e 5 palline nere numerate da 5 a 9. L'evento

$$A = \text{esce un numero inferiore a 4},$$

e l'evento

$$B = \text{esce una pallina nera},$$

Sono chiaramente incompatibili.

5.3 Probabilità Condizionata

La probabilità di un evento A condizionata ad un evento B risponde alla domanda

Se B fosse vero, con quale probabilità sarebbe vero A?

Definizione 5.3.1 (Probabilità Condizionata). Si definisce probabilità di A condizionata a B (probabilità di A dato B) la quantità

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Esempio 5.3.1. Un'urna ha 8 palline bianche numerate da 1 a 8 e 5 palline nere numerate da 5 a 9. Si considerino l'evento

$$A = \text{esce un numero maggiore o uguale a 6},$$

e l'evento

$$B = \text{esce una pallina nera}.$$

Ovviamente

$$\begin{aligned} P(A) &= \frac{7}{13} = 0.5385 \\ P(B) &= \frac{5}{13} = 0.3846 \end{aligned}$$

$$P(A \cap B) \;=\; \frac{4}{13} = 0.3077$$

e infine

$$\begin{aligned} P(A|B) &= \frac{P(A \cap B)}{P(B)} \\ &= \frac{\frac{4}{13}}{\frac{5}{13}} = 0.8 \\ P(B|A) &= \frac{P(A \cap B)}{P(A)} \\ &= \frac{\frac{4}{13}}{\frac{7}{13}} = 0.5714 \end{aligned}$$

La probabilità condizionata ci consente di esprimere la probabilità dell'intersezione come prodotto di probabilità condizionate. Infatti in alcune circostanze è più facile ricavare una probabilità condizionata invece della probabilità dell'intersezione. La *Chain Rule* è il modo di esprimere probabilità dell'intersezione come prodotto di condizionate.

Versione a coppie

$$P(A \cap B) = P(A)P(B|A) = P(B)P(A|B)$$

la dimostrazione deriva direttamente dalla definizione 5.3.1.

Versione a triple

$$P(A \cap B \cap C) = P(A)P(B|A)P(C|A \cap B)$$

Per mostrare la versione a triple basta analizzare prima l'intersezione di due eventi A e $B \cap C$, ovvero

$$\begin{aligned} P(A \cap B \cap C) &= P(A \cap (B \cap C)) \\ &= P(A)P(B \cap C|A) \\ &= P(A)P(B|A)P(C|A \cap B) \end{aligned}$$

Versione generale

$$P(A_1 \cap A_2 \cap A_3 \cap ... \cap A_n) = P(A_1)P(A_2|A_1)P(A_3|_1 \cap A_2)...P(A_n|A_1 \cap A_2 \cap ... \cap A_{n-1})$$

Per mostrare la versione generale basta iterare il ragionamento in triplette.

5.3.1 Indipendenza tra Eventi

L'**indipendenza** è un concetto estremamente importante in probabilità. Se due eventi A e B sono *indipendenti* significa che l'accadere o meno dell'uno non altera in alcun modo la probabilità dell'altro.

Definizione 5.3.2 (Indipendenza tra Eventi). Due eventi A e B si dicono **indipendenti** se e solo se

$$\begin{aligned} P(A|B) &= P(A) \\ P(B|A) &= P(B) \end{aligned}$$

e quindi, se A e B sono indipendenti

$$P(A \cap B) = P(A)P(B)$$

Infatti

$$P(A \cap B) = P(A)P(B|A) = P(A)P(B)$$

5.3.2 Indipendenza e Incompatibilità

Se A e B sono **incompatibili** allora **non** sono **indipendenti**, vice versa se A e B sono **indipendenti** allora **non** sono **incompatibili**:

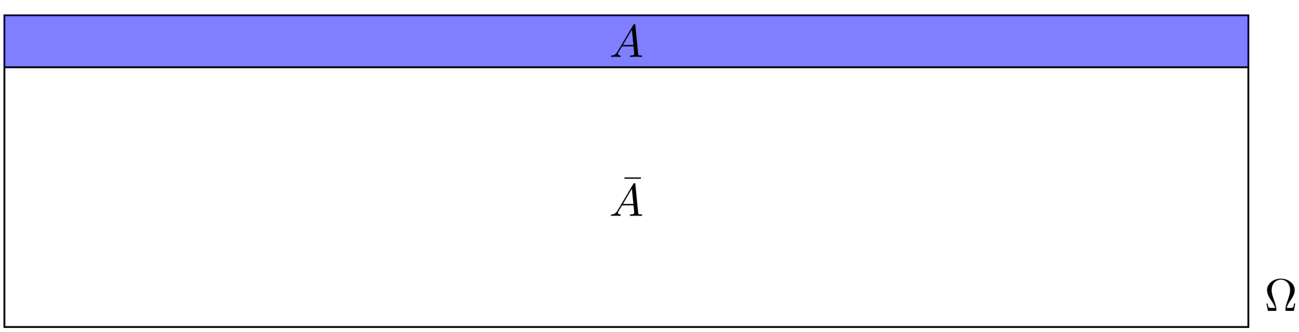
Attenzione

- Se $A \neq \emptyset$ e $B \neq \emptyset$ sono incompatibili, allora $P(A \cap B) = 0 \neq P(A)P(B)$
- Se A e B sono indipendenti, allora $P(A \cap B) = P(A)P(B) \neq 0$

5.3.3 Partizioni di Ω

Un *partizione* di Ω è una collezione di due o più eventi disgiunti che uniti insieme restituiscono Ω. Se A è un evento e $\bar{A}$ è il suo complementare è immediato che la coppia $\{A, \bar{A}\}$ è una *partizione* in quanto $A \cup \bar{A} = \Omega$. In questo caso parleremo di *bipartizione*.

Definizione 5.3.3 (Bipartizione). Sia A è un evento e $\bar{A}$ è il suo complementare, allora la coppia $\{A, \bar{A}\}$ è una *bipartizione* di Ω.



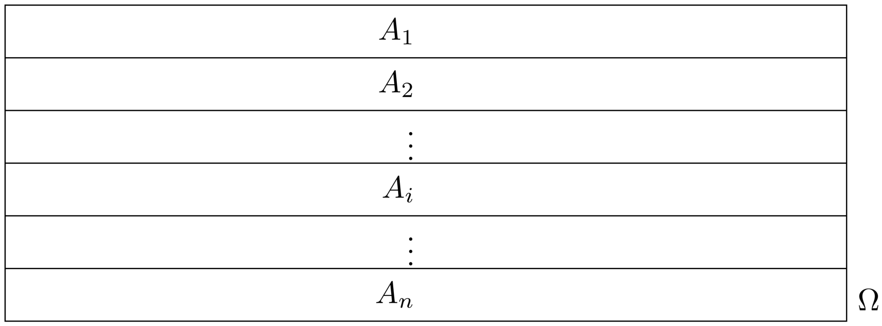
Più in generale definiamo una *partizione finita*:

Definizione 5.3.4 (Partizione finita). Sia $\{A_1, ..., A_n\}$, $n < +\infty$ una collezione di eventi di Ω tali che

1. $A_i \cap A_j = \emptyset, \ \forall i \neq j$

2. $\bigcup_{i=1}^{n} A_i = \Omega$

Allora $\{A_1, ..., A_n\}$ è detta una *partizione* (finita) di Ω

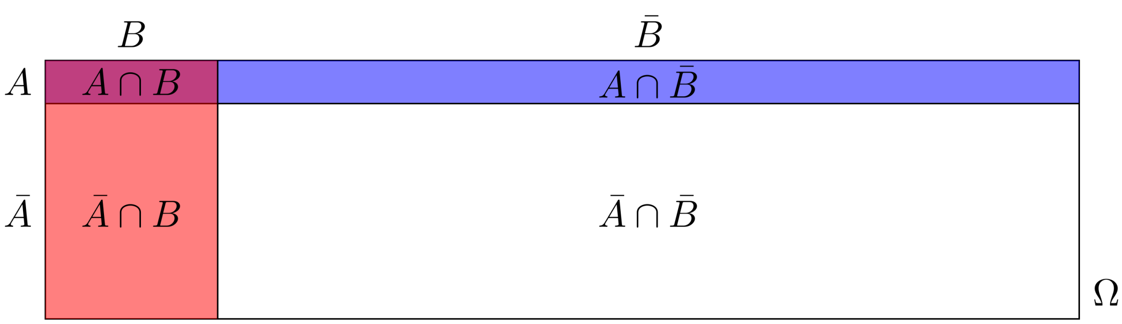


5.3.4 Teorema delle probabilità totali

Il teorema dell probabilità totali permette di esprime la probabilità un evento come somma della probabilità delle intersezioni che lo compongono.

Teorema 5.3.1 (Probabilità Totali versione a coppie). *Siano A e B due eventi diversi dal vuoto, allora*

$$P(B) = P(A)P(B|A) + P(\bar{A})P(B|\bar{A})$$



Dimostrazione. Si parte considerando l'identità insiemistica

$$B = (A \cap B) \cup (\bar{A} \cap B)$$

e quindi

$$\begin{aligned} P(B) &= P(A \cap B) + P(\bar{A} \cap B) \\ &= P(A)P(B|A) + P(\bar{A})P(B|\bar{A}), \qquad \text{per la chain rule} \end{aligned}$$

□

Esempio 5.3.2. Quando piove Giulio arriva in ritardo con probabilità 0.18, mentre se non piove arriva in ritardo con probabilità 0.01. Le previsioni del tempo dicono che domani pioverà con probabilità 0.85. Calcolare la probabilità che Giulio, domani, arrivi in ritardo.

Soluzione. Sia B = "Domani Giulio arriverà in ritardo" e sia A = "Domani pioverà". Sappiamo dalle previsioni che $P(A) = 0.85$, mentre $P(\bar{A}) = 1 - 0.85 = 0.15$. Inoltre sappiamo che $P(B|A) = 0.18$ mentre $P(B|\bar{A}) = 0.01$ Inoltre osserviamo che B si può dividere nell'intersezione in due insiemi disgiunti: piove *ed* arriva in ritardo *oppure* non piove *ed* arriva in ritardo, in simboli

$$B = (A \cap B) \cup (\bar{A} \cap B)$$

Dal teorema dell probabilità totali

$$\begin{aligned} P(B) &= P(A)P(B|A) + P(\bar{A})P(B|\bar{A}) \\ &= 0.85 \times 0.18 + 0.15 \times 0.01 \\ &= 0.1545 \end{aligned}$$

Il teorema può essere esteoso alla versione a triple: Siano $\{A_1, A_2, A_3\}$ e $\{B_1, B_2, B_3\}$ due partizioni di Ω: $A_1 \cap A_2 = \emptyset$, $A_1 \cap A_3 = \emptyset$, $A_2 \cap A_3 = \emptyset$ e $\Omega = B_1 \cup B_2 \cup B_3$, e $B_1 \cap B_2 = \emptyset$, $B_1 \cap B_3 = \emptyset$, $B_2 \cap B_3 = \emptyset$ e $\Omega = B_1 \cup B_2 \cup B_3$. Allora

$$\begin{aligned} P(B_1) &= P(A_1)P(B_1|A_1) + P(A_2)P(B_1|A_2) + (A_3)P(B_1|A_3) \\ P(B_2) &= P(A_1)P(B_2|A_1) + P(A_2)P(B_2|A_2) + (A_3)P(B_2|A_3) \\ P(B_3) &= P(A_1)P(B_3|A_1) + P(A_2)P(B_3|A_2) + (A_3)P(B_3|A_3) \end{aligned}$$

Teorema 5.3.2 (Probabilità Totali versione Generale)**.** *Siano $\{A_1, ..., A_n\}$ e $\{B_1, ..., B_m\}$ due partizioni di Ω, ovvero $A_i \cap A_j = \emptyset$, $\forall i \neq j$ e $\Omega = \bigcup_{i=1}^n A_i$ e $B_i \cap B_j = \emptyset$, $\forall i \neq j$ e $\Omega = \bigcup_{i=1}^n B_i$ Allora*

$$P(B_j) = \sum_{i=1}^{n} P(A_i)P(B_j|A_i), \qquad j = 1, ..., m$$

	B_1	B_2	...	B_j	...	B_m	
A_1	$A_1 \cap B_1$	$A_1 \cap B_2$		$A_1 \cap B_j$		$A_1 \cap B_m$	
A_2	$A_2 \cap B_1$	$A_2 \cap B_2$		$A_2 \cap B_j$		$A_2 \cap B_m$	
$\vdots$							
A_i	$A_i \cap B_1$	$A_i \cap B_2$		$A_i \cap B_j$		$A_i \cap B_m$	
$\vdots$							
A_n	$A_n \cap B_1$	$A_n \cap B_2$		$A_n \cap B_j$		$A_n \cap B_m$	Ω

Dimostrazione. Si parte considerando l'identità insiemistica

$$B_j = (A_1 \cap B_j) \cup (A_2 \cap B_j) \cup ... \cup (A_n \cap B_j)$$

e quindi

$$\begin{aligned} P(B_j) &= P(A_1 \cap B_j) + P(A_2 \cap B_j) + ... + P(A_n \cap B_j) \\ &= P(A_1)P(B_j|A_1) + P(A_2)P(B_j|A_2) + ... + P(A_n)P(B_j|A_n), \qquad \text{per la chian rule} \\ &= \sum_{i=1}^{n} P(A_i)P(B_j|A_i) \end{aligned}$$

□

Esempio 5.3.3. Un'urna contiene 10 palline: due etichettate con A, tre etichettate con B e le rimanenti cinque etichettate con C. Se si estrae la pallina A si estrae da l'urna $\mathcal{A}$ che contiene 3 palline vincenti e una perdente, se si estrae la pallina B si estrae da l'urna $\mathcal{B}$ che contiene due palline vincenti e una perdente e se si estrae la pallina C si estrae da l'urna $\mathcal{C}$ che contiene 1 pallina vincente e due perdenti. Qual è la probabilità di vincere?

Soluzione. Anzitutto notiamo che

$$\begin{aligned} P(A) &= \frac{2}{10} \\ P(B) &= \frac{3}{10} \\ P(C) &= \frac{5}{10} \end{aligned}$$

Sia $V =$ "Vincere", dai dati abbiamo che

$$\begin{aligned} P(V) &= P(A)P(V|A) + P(B)P(V|B) + (C)P(V|C) \\ &= \frac{2}{10} \times \frac{3}{4} + \frac{3}{10} \times \frac{2}{3} + \frac{5}{10} \times \frac{1}{3} \\ &= 0.5167 \end{aligned}$$

5.3.5 Il Teorema di Bayes

Il teorema di Bayes è un risultato probabilitistico che consente di esprimere le probabilità condizionate ed è utilizzato come base per quella che è nota come teoria statistica bayesiana, di cui non entreremo nel dettaglio.

Teorema 5.3.3 (Teorema di Bayes versione a coppie.). *Si considerino due eventi A e B di cui sono note $P(A)$, $P(B|A)$ e $P(B|\bar{A})$, allora*

$$P(A|B) = \frac{P(A)P(B|A)}{P(A)P(B|A) + P(\bar{A})P(B|\bar{A})}$$

Dimostrazione.

$$\begin{aligned} P(A|B) &= \frac{P(A \cap B)}{P(B)}, && \text{per definizione} \\ &= \frac{P(A)P(B|A)}{P(B)}, && \text{per la chain rule} \\ &= \frac{P(A)P(B|A)}{P(A)P(B|A) + P(\bar{A})P(B|\bar{A})}, && \text{per il teorema delle prob. tot.} \end{aligned}$$

□

Esempio 5.3.4. Quando piove Giulio arriva in ritardo con probabilità 0.18, mentre se non piove arriva in ritardo con probabilità 0.01. Le previsioni del tempo dicono che domani pioverà con probabilità 0.85. Se il giorno dopo Giulio entrasse in ritardo con qual probabilità avrebbe piovuto?

Soluzione. Sia B = "Domani Giulio arriverà in ritardo" e sia A = "Domani pioverà". Sappiamo dalle previsioni che $P(A) = 0.85$, mentre $P(\bar{A}) = 1 - 0.85 = 0.15$. Inoltre sappiamo che $P(B|A) = 0.18$ mentre $P(B|\bar{A}) = 0.01$ Inoltre osserviamo che B si può dividere nell'intersezione in due insiemi disgiunti: piove *ed* arriva in ritardo *oppure* non piove *ed* arriva in ritardo, in simboli

$$B = (A \cap B) \cup (\bar{A} \cap B)$$

Dal teorema dell probabilità totali

$$\begin{aligned} P(B) &= P(A)P(B|A) + P(\bar{A})P(B|\bar{A}) \\ &= 0.85 \times 0.18 + 0.15 \times 0.01 \\ &= 0.1545 \end{aligned}$$

Dal teorema di Bayes

$$\begin{aligned} P(B|A) &= \frac{P(A)P(B|A)}{P(A)P(B|A) + P(\bar{A})P(B|\bar{A})} \\ &= \frac{0.85 \times 0.18}{0.85 \times 0.18 + 0.15 \times 0.01} \\ &= 0.9903 \end{aligned}$$

Teorema 5.3.4 (Teorema di Bayes versione Generale). *Siano $\{A_1, ..., A_n\}$ e $\{B_1, ..., B_n\}$ due partizioni di Ω, di cui sono note $P(A_i), \forall i$ e $P(B_j|A_i), \forall i, j$, allora*

$$P(A_i|B_j) = \frac{P(A_i)P(B_j|A_i)}{\sum_{i=1}^{n} P(A_i)P(B|A_i)}$$

Esempio 5.3.5. Un'urna contiene 10 palline: due etichettate con A, tre etichettate con B e le rimanenti cinque etichettate con C. Se si estrae la pallina A si estrae da l'urna $\mathcal{A}$ che contiene 3 palline vincenti e una perdente, se si estrae la pallina B si estrae da l'urna $\mathcal{B}$ che contiene due palline vincenti e una perdente e se si estrae la pallina C si estrae da l'urna $\mathcal{C}$ che contiene 1 pallina vincente e due perdenti. Giulio ha appena giocato e ha vinto, qual è la probabilità che sia uscita una pallina etichettata con C?

Soluzione. Anzitutto notiamo che

$$\begin{aligned} P(A) &= \frac{2}{10} \\ P(B) &= \frac{3}{10} \\ P(C) &= \frac{5}{10} \end{aligned}$$

Sia $V =$ "Vincere", dai dati abbiamo che

$$\begin{aligned} P(V) &= P(A)P(V|A) + P(B)P(V|B) + (C)P(V|C) \\ &= \frac{2}{10} \times \frac{3}{4} + \frac{3}{10} \times \frac{2}{3} + \frac{5}{10} \times \frac{1}{3} \\ &= 0.5167 \end{aligned}$$

In virtù del teorema di Bayes

$$\begin{aligned} P(C|V) &= \frac{P(C)P(V|C)}{P(A)P(V|A) + P(B)P(V|B) + P(C)P(V|C)} \\ &= \frac{\frac{5}{10} \times \frac{1}{3}}{\frac{2}{10} \times \frac{3}{4} + \frac{3}{10} \times \frac{2}{3} + \frac{5}{10} \times \frac{1}{3}} \\ &= 0.3226 \end{aligned}$$

5.4 Specchietto finale utile per gli esercizi elementari

$0 \leq P(A) \leq 1, \ \forall A \in \mathscr{A}$	la probabilità è compresa tra 0 e 1.
$P(\Omega) = 1$	la prob. dell'evento certo è 1,
$P(\emptyset) = 0$	la prob. dell'insieme vuoto è zero.
$P(A) = 1 - P(\bar{A})$	regola del complementare
$P(A \cup B) = P(A) + P(B) - P(A \cap B)$	regola della somma (de Morgan)
$P(A \cup B) = P(A) + P(B)$	**se e solo se A e B sono incompatibili:**
	terzo assima di Kolmogorov
$P(A \cap B) = P(A)P(B\|A) = P(B)P(A\|B)$	regola del prodotto (chain rule)
$P(A \cap B) = P(A)P(B)$	se e solo se A e B sono indipendenti
$P(B) = P(A)P(B\|A) + P(\bar{A})P(B\|\bar{A})$	Teorema delle probabilità totali

Variabili Casuali

6

6.1 Definizione formale di una VC discreta

Una variabile casuale, o numero casuale, è un numero che non sappiamo in anticipo quale valore varrà: il numero di Teste nei prossimi 10 lanci, i millimetri di pioggia che cadranno in Emilia Romagna nelle prossime 24 ore, il numero di ricoverati nell'ospedale Maggiore di Bologna nel 2025.

Dal punto di vista formale una variabile casuale è una numerizzazione dell'evento certo. Si consideri un partizione finita o al più numerabile dell'evento certo:

$$\{E_1, ..., E_k\},\ E_i \cap E_j,\ \forall i \neq j,\ E_1 \cup ... \cup E_k = \Omega$$

Una VC X è una funzione che mappa la partizione sulla retta reale

$$X : \Omega \to \mathbb{R}$$

.

$$X(E_j) = x_j,\ \text{ad ogni evento } E_j \text{ viene assegnato un numero}$$

Si definisce S_X il **supporto** della VC X, ed è l'insieme dei valori che la VC è suscettibile di assumere.

Resta definita la **funzione di probabilità** di X che è data da:

$$f(x) = P(X = x) = P(E_j : X(E_j) = x)$$

Una variabile casuale X è un numero che ancora non sappiamo quanto varrà, potrà assumere uno qualunque dei valori x del supporto S_X e lo assumerà con una data probabilità $f(x) = P(X = x)$.

Esempio 6.1.1. Consideriamo il lancio di due monete identiche, la partizione associata è

- $E_1 = \{T, T\}$, la prima e la seconda moneta mostrano Testa.
- $E_2 = \{T, C\}$, la prima moneta mostra Testa e la seconda Croce.
- $E_3 = \{C, T\}$, la prima moneta mostra Croce e la seconda Testa.
- $E_4 = \{C, C\}$, la prima e la seconda moneta mostrano Croce.

Si considerino le tre variabili nella tabella qui sotto

j	E_j	$X(E_j)$	$Y(E_j)$	$W(E_j)$
1	$E_1 = \{T, T\}$	$X(E_1) = 2$	$Y(E_1) = 0$	$W(E_1) = 1$
2	$E_2 = \{T, C\}$	$X(E_2) = 1$	$Y(E_2) = 1$	$W(E_2) = 0$
3	$E_3 = \{C, T\}$	$X(E_3) = 1$	$Y(E_3) = 1$	$W(E_3) = 0$
4	$E_4 = \{C, C\}$	$X(E_4) = 0$	$Y(E_4) = 2$	$W(E_4) = 1$

X conta il numero di Teste (0, 1 o 2), Y conta il numero di Croci (0, 1 o 2), W vale uno se i due lanci sono identici e vale zero altrimenti.

Se la moneta è perfetta ($P(T) = P(C) = 0.5$) allora:

$$\begin{aligned}
P(E_1) &= P(\{T, T\}) \\
&= P(T)P(T), \qquad \text{in virtù dell'indipendenza} \\
&= \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} \\
P(E_2) &= P(\{T, C\}) \\
&= P(T)P(C), \qquad \text{in virtù dell'indipendenza} \\
&= \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} \\
P(E_3) &= P(\{C, T\}) \\
&= P(C)P(T), \qquad \text{in virtù dell'indipendenza} \\
&= \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} \\
P(E_4) &= P(\{C, C\}) \\
&= P(C)P(C), \qquad \text{in virtù dell'indipendenza} \\
&= \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}
\end{aligned}$$

e quindi:

$$\begin{aligned}
P(X = 0) &= P(E_1) \\
&= \frac{1}{4} \\
P(X = 1) &= P(E_2 \cup E_3) \\
&= P(E_2) + P(E_3) \\
&= \frac{1}{4} + \frac{1}{4} = \frac{2}{4} \\
P(X = 2) &= P(E_4) \\
&= \frac{1}{4}.
\end{aligned}$$

6.1.1 Descrizione di una VC

Un X VC ho molte analogie con una VS, X ha un *supporto* ovvero l'insieme dei valori che X può assumere, una funzione d probabilità che produce numeri compresi tra zero e uno che sommano ad uno (come le frequenze relative), la Funzione di Ripartizione che, anziché cumulare le frequenze, cumula le probabilità, il Valore Atteso che è l'analogo della media e la Varianza che è l'analogo della varianza descrittiva. In sintesi

Nota

Una VC discreta è **identificata** in maniera univoca da

- Il suo supporto S_X.
- La sua funzione di probabilità $f(x)$ o la su funzione di ripartizione F.
- Il suo valore atteso $E(X)$.
- La sua varianza $V(X)$.

Definizione 6.1.1 (Supporto). Sia X una VC, si definisce S_X il supporto di X, l'insieme di tutti i possibili valori che X è suscettibile di assumere.

Esempio 6.1.2. Lancio una moneta due volte e definisco X la VC che *conta il numero di volte che osservo Testa in due lanci*. La VC X potrà assumete solo 3 valori: 0 (zero volte Testa in due lanci), 1 (una volta Testa in due lanci), e 2 (2 volte Testa in due lanci). E quindi

$$S_X = \{0, 1, 2\}$$

Definizione 6.1.2 (Funzione di Probabilità). Sia X una VC con supporto S_X, si definisce f la funzione di probabilità: è la probabilità che la VC X assuma esattamente il valore x

$$f(x) = P(X = x), \;\; x \in S_X$$

Esempio 6.1.3. Si consideri un'urna che contiene 4 palline:

$$E_1 = \boxed{1}, E_2 = \boxed{2}, E_3 = \boxed{2}, E_4 = \boxed{3}$$

Sia X la VC che rappresenta il numero sulla pallina:

$$\begin{aligned} X(E_1) &= 1 \\ X(E_2) &= 2 \\ X(E_3) &= 2 \\ X(E_4) &= 3 \end{aligned}$$

ha come supporto:

$$S_X = \{1, 2, 3\}$$

e come funzione di probabilità:

$$\begin{aligned} f(1) &= P(X = 1) = P(E_1) = \frac{1}{4} \\ f(2) &= P(X = 2) = P(E_2 \cup E_3) = \frac{1}{4} + \frac{1}{4} = \frac{2}{4} \\ f(3) &= P(X = 3) = P(E_3) = \frac{1}{4} \end{aligned}$$

La funzione di probabilità f si comporta come le frequenze relative, la f è compresa tra zero e uno, per ogni x e la somma della f calcolata su ogni x dà uno.

Proprietà 6.1.1 (Funzione di probabilità). *Sia X una VC con supporto S_X e funzione di probabilità f, allora*

- $0 \leq f(x) \leq 1, \forall x \in S_X$
- $\sum_{x \in S_X} f(x) = 1.$

La Funzione di Ripartizione F di una VC X cumula tutta la probabilità fino ad x:

Definizione 6.1.3 (Funzione di Ripartizione di una VC).

$$F(x) = P(X \leq x) = \sum_{x^* \leq x} f(x^*)$$

Esempio 6.1.4 (Continua). Continuando l'esempio precedente:

$$\begin{aligned} F(x) &= P(X \leq x) \\ &= \begin{cases} 0, & x < 1 \\ \frac{1}{4}, & 1 \leq x < 2 \\ \frac{3}{4} & 2 \leq x < 3 \\ \frac{4}{4} & x \geq 3 \end{cases} \end{aligned}$$

La funzione di ripartizione gode di diverse proprietà. Una, in particolare sarà utilissima per il calcolo delle probabilità della VC continua che vedremo più avanti.

Proprietà 6.1.2 (Funzione di Ripartizione). *La funzione di ripartizione F di una VC X è, per definizione:*

$$F(x) = P(X \leq x)$$

F gode delle seguenti proprietà:

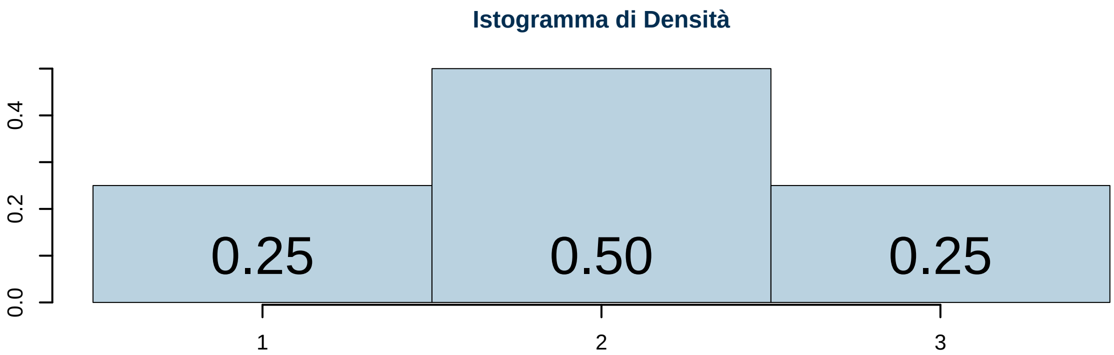
1. *Non decrescente, ossia $x_1 < x_2 \Rightarrow F(x_1) \leq F(x_2)$*
2. $\lim_{x \to -\infty} F(x) = 0, \qquad \lim_{x \to \infty} F(x) = 1.$
3. *Continua a destra, ossia $\lim_{x \to x_0^+} F(x) = F(x_0)$.*
4. *$P(a < X \leq b) = F(b) - F(a)$.*

La funzione di ripartizione di una VC si comporta la funzione di ripartizione di una VS. Altre caratteristiche si possono calcolare, come i **percentili** di ordine p: *Il percentile x_p* è quel valore che divide la distribuzione di X in due: $P(X \leq x_p) = p, \quad P(X > x_p) = 1 - p$

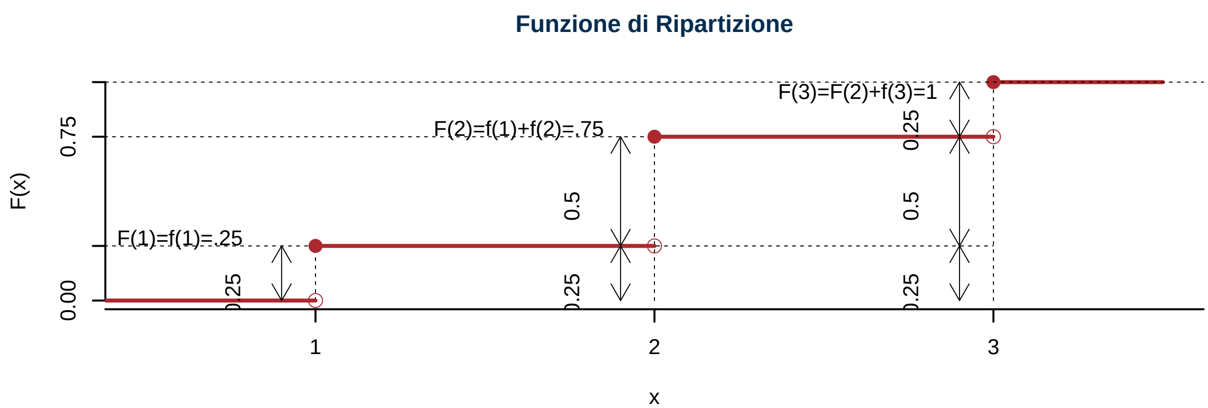
Definizione 6.1.4 (Percentie di una VC). Sia X una VC con support S_X e con Fuzione di Ripartizione F, si definisce il p-esimo percentile di X, il vaolre x_p, tale che:

$$x_p : F(x_p) = p$$

La distribuzione può essere rappresentata graficamente con un istogramma di densità. Esattamente come per la VC quantitative discrete una VC può essere rappresentata in intervalli di ampiezza uno costruiti intorno ai dati del supporto. Nel nostro esempio di prima:



Analogamente possiamo costruire la funzione di ripartizione cumulando le probabilità. Nell'esempio di sopra otterremmo:



6.1.2 Operazioni tra VC

Le VC sono numeri che non sappiamo in anticipo che valore assumeranno. Ma siccome diventeranno numeri li potremo sommare, sottrarre, moltiplicare, ecc. sia con numeri costanti che con altre VC.

Esempio 6.1.5. Sia X la VC con supporto $S_X = \{-1, 0, +1\}$ e con funzione di probabilità

$$
\begin{aligned}
f(-1) &= \frac{1}{5} \\
f(0) &= \frac{3}{5} \\
f(+1) &= \frac{1}{5}
\end{aligned}
$$

e sia Y la VC con supporto $S_Y = \{0, +1\}$ e con funzione di probabilità

$$\begin{aligned} f(0) &= \frac{1}{2} \\ f(+1) &= \frac{1}{2} \end{aligned}$$

Se poniamo

$$S = X + Y$$

S è la VC che rappresenta la somma di due VC: casuale è X, casuale è Y, casuale sarà la somma tra X ed Y.

	-1;	$1/5$	0;	$3/5$	1;	$1/5$
0; $1/2$	-1;	$1/10$	0;	$3/10$	1;	$1/10$
1; $1/2$	0;	$1/10$	1;	$3/10$	2;	$1/10$

E ricaviamo la distribuzione di, S

S	-1	0	1	2
$P(S)$	$1/10$	$4/10$	$4/10$	$1/10$

6.2 Valore Atteso, e Varianza di una VC

Il **valore atteso** di una VC X è l'analogo della media aritmetica di una VC, le modalità di X vengono pesate con le probabilità invece che con le frequenze.

Definizione 6.2.1 (Valore Atteso di una VC discreta). Si definisce $E(X)$ il valore atteso della VC X con supporto S_X e funzione di probabilita f:

$$E(X) = \sum_{x \in S_x} x f(x)$$

Attenzione

Il valore atteso di una VC è un numero.

La **varianza** di una VC è del tutto analoga alla varianza di una VS.

Definizione 6.2.2 (Varianza di una VC discreta). Si definisce $V(X)$ la varianza della VC X con supporto S_X e funzione di probabilita f:

$$\begin{aligned} V(X) &= E\left((X - E(X))^2\right) \\ &= \sum_{x \in S_x} (x - E(X))^2 f(x), \qquad \text{oppure equivalentemente} \end{aligned}$$

$$
\begin{aligned}
&= E(X^2) - E^2(X) \\
&= \sum_{x \in S_X} x^2 f(x) - E^2(X).
\end{aligned}
$$

così come la sua standard deviation.

Definizione 6.2.3 (Standard Deviation di una VC discreta). Si definisce $SD(X)$ la Standard Deviation della VC X con supporto S_X e funzione di probabilita f, la radice della sua varianza

$$SD(X) \quad = \quad \sqrt{V(X)}$$

Esempio 6.2.1 (Continua). Continuiamo l'esempio precedente e otteniamo, il valore atteso:

$$E(X) = 1 \cdot \frac{1}{4} + 2 \cdot \frac{2}{4} + 3 \cdot \frac{1}{4} = 2$$

e la varianza di X

$$V(X) = 1^2 \cdot \frac{1}{4} + 2^2 \cdot \frac{2}{4} + 3^2 \cdot \frac{1}{4} - 2^2 = 0.5$$

Le proprietà di valore atteso e varianza di una VC sono del tutto analoghe alle rispettive di una VS.

Proprietà 6.2.1 (Proprietà del Valore Atteso di una VC). *Le proprietà del valore atteso, $E(X)$ sono:*

1. $x_{\min} \leq E(X) \leq x_{\max}, \quad x_{\min},\ x_{\max} \in S_X$,
2. $E\Big(X - E(X)\Big) = 0$,
3. $E\Big(X - E(X)\Big)^2 < E(X-d)^2 \quad \forall\ d \neq E(X)$,
4. $E(a + bX) = a + b\ E(X)$.
5. $E(aX + bY) = aE(X) + bE(Y)$

Esiste anche l'analogo della proprietà di associatività per il valore atteso, ma richiede alcuni risultati che esulano dallo scopo di questi appunti e non verrà riportata.

Anche per la varianza di una VC valgono le stesse proprietà della varianza di una VS. In particolare

Proprietà 6.2.2 (Proprietà della Varianza di una VC). *Le proprietà della Varianza, $V(X)$ sono:*

1. $V(X) \geq 0$,
2. $V(X) = 0$ *se e solo se* $P(X = x) = 1$
3. $$V(a + bX) = b^2 V(X)$$
4. *Se X e Y sono* ***indipendenti****, allora*

$$V(aX + bY) = V(aX - bY) = a^2 V(X) + b^2 V(Y), \forall\ a, b \in \mathbb{R}$$

Notiamo che nel caso di una VC X la sua varianza vale zero se e solo se la VC assume un solo valore con probabilità uno. Quindi una VC che non varia, ovvero una costante.

Attenzione

Se $a = 1$ e $b = 1$ allora

$$V(X+Y) = V(X-Y) = V(X) + V(Y)$$

le varianze non si sottraggono mai.

Proprietà 6.2.3 (Proprietà della SD di una VC). *Le proprietà della Standard Deviation di X, $SD(X)$ sono:*

1. $SD(X) \geq 0$,
2. $SD(X) = 0$ *se e solo se* $P(X = x) = 1$
3. $$SD(a + bX) = |b|V(X)$$
4. *Se X e Y sono **indipendenti**, allora*
$$SD(aX + bY) = SD(aX - bY) = \sqrt{a^2V(X) + b^2V(Y)}, \forall\ a, b \in \mathbb{R}$$

Attenzione

Se $a = 1$ e $b = 1$ allora

$$SD(X+Y) = SD(X-Y) = \sqrt{V(X) + V(Y)}$$

la SD di una somma non si esprime come la comma delle SD degli addendi.

6.3 Indipendenza tra VC

In generale due VC X e Y si dicono indipendenti se e solo se:

$$P(X \in A \cap Y \in B) = P(X \in A) \cdot P(Y \in B), \quad \forall A \subset S_X, \forall B \subset S_Y$$

Per le VC discrete la relazione di indipendenza si può scrivere:

$$P(X = x \cap Y = y) = P(X = x) \cdot P(Y = y), \quad \forall x \in S_X, \forall y \in S_Y$$

6.4 VC condizionate (complementi)

La probabilità condizionata di X dato Y si scrive

$$P(X \in A | Y \in B) = \frac{P(X \in A \cap Y \in B)}{P(Y \in B)}$$

Se X e Y sono discrete si può scrivere

$$\begin{aligned} f(x|y) &= P(X = x|Y = y) \\ &= \frac{P(X = x \cap Y = y)}{P(Y = y)} \end{aligned}$$

e si legge: *la probabilità che* X assuma il valore x dato che Y ha assunto il valore y è $f(x|y)$.

6.4.1 Valore atteso e varianza condizionata (complementi)

Sia X una VC discreta con supporto S_X. Si definisce il valore atteso di X condizionato ad $Y = y$, la quantità

$$E(X|Y = y) = \sum_{x \in S_X} x f(x|y)$$

Si definisce varianza di X condizionato ad $Y = y$, la quantità

$$V(X|Y = y) = \sum_{x \in S_X} (x - E(X|Y = y))^2 f(x|y)$$

6.4.2 Esempio di indipendenza tra VC

Sia Y una VC con supporto $S_Y = \{-1, +1\}$ e con funzione di probabilità

$$\begin{aligned} f_Y(-1) &= \frac{1}{2} \\ f_Y(+1) &= \frac{1}{2} \end{aligned}$$

la Y al piede della f serve a non confondere la funzione di probabilità della Y con quella della X. Se X ed Y sono **indipendenti** allora:

$$\begin{aligned} P(X = x \cap Y = y) &= f(x, y) \\ &= f_X(x) f_Y(y) \end{aligned}$$

	$f_X(1) = \frac{1}{4}$	$f_X(2) = \frac{2}{4}$	$f_X(3) = \frac{1}{4}$
$f_Y(-1) = \frac{1}{2}$	$f(1, -1) = 0.125$	$f(2, -1) = 0.25$	$f(3, -1) = 0.125$
$f_Y(+1) = \frac{1}{2}$	$f(1, +1) = 0.125$	$f(2, +1) = 0.25$	$f(3, +1) = 0.125$

6.5 Specchietto finale per le VC discrete

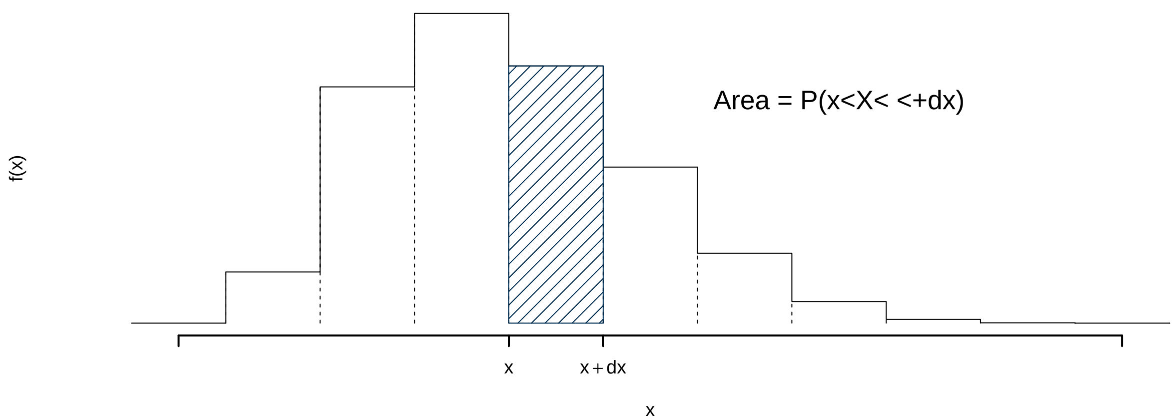
S_X	il supporto della VC X:
	l'insieme di tutti i possibili valori che la VC può assumere.
	Se X è una VD discreta il suo supporto ha:
$S_X = \{x_1, \ldots, x_k\}$	un numero finito,
$S_X = \{x_1, x_2, x_3, \ldots\}$	o al più numerabile di elementi.
$f(x) = P(X = x),\ x \in S_X$	f è la funzione di probabilità,
	indica la probabilità che la VC X assuma esattamente il valore x.
$E(a + bX) = a + bE(X)$	linearità
$E(aX + bY) = aE(X) + bE(Y)$	
$V(X) = E\left(\left(X - E(X)\right)^2\right)$	Varianza della VC X
$= \sum_{x \in S_X} x^2 f(x) - E^2(X)$	
$V(a + bX) = b^2 V(X)$	
$SD(X) = \sqrt{V(X)}$	Standard Deviation della VC X
$SD(a + bX) = \lvert b\rvert SD(X)$	
Indipendenza tra VC	
$P(X \in A \cap Y \in B) = P(X \in A) \cdot P(Y \in B)$	$\forall A \subset S_X, \forall B \subset S_Y$
$P(X = x \cap Y = y) = P(X = x) \cdot P(Y = y)$	$\forall x \in S_X, \forall y \in S_Y$
$V(aX + bY) = a^2 V(X) + b^2 V(Y)$	se e solo se X e Y sono indipendenti
$SD(aX + bY) = \sqrt{a^2 V(X) + b^2 V(Y)}$	se e solo se X e Y sono indipendenti.
	n.b. la SD di una somma non
	si può esprimere con la somma delle SD.

6.6 Le VC continue

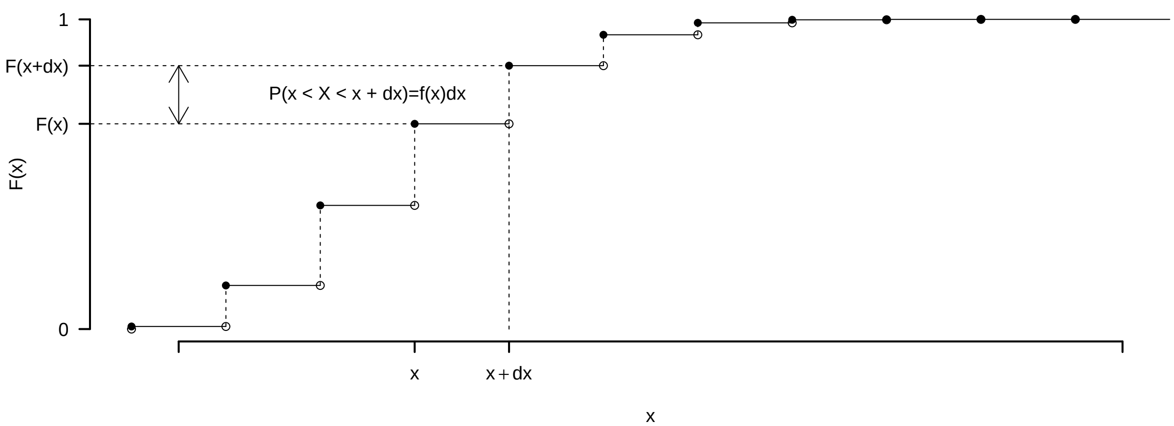
Le VC che hanno un supporto più che numerabile, $S_X \subseteq \mathbb{R}$, cioè un sottoinsieme della retta reale o la retta reale stessa vengono chiamate VC continue. Siccome un intervallo di numeri reali contiene una quantità più che numerabile di punti è impossibile probabilizzare tutti i punti dell'intervallo. Invece di probabilizzare i singoli numeri vengono probabilizzati gli intervalli.

Concettualmente si definiscono immaginando di mandare la precisione di una misura all'infinito. Ogni misura è un conteggio, una lunghezza si può misurare in *quanti* metri, o in *quanti centimetri* centimetri, o in *quanti millimetri*, ecc.

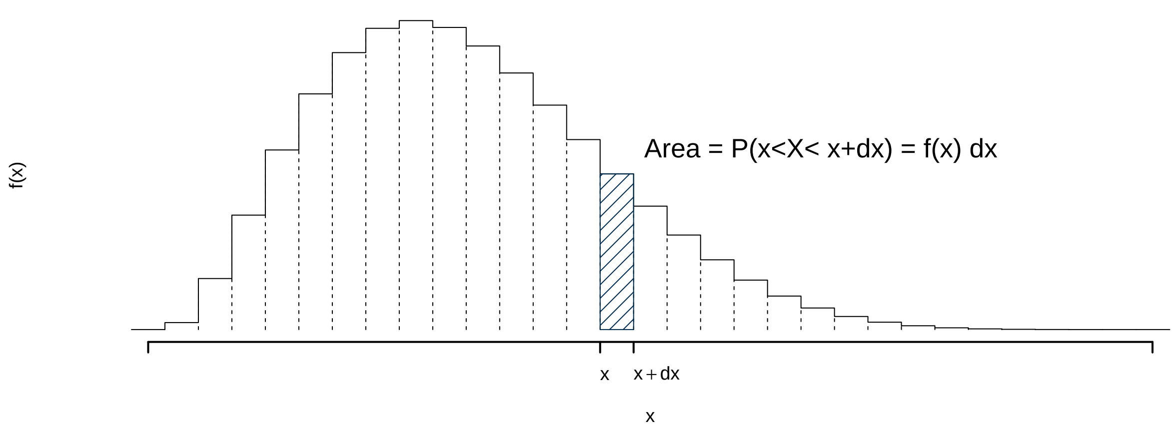
L'idea è di dividere il supporto in classi e costruire un istogramma di densità tale che l'area sottesa ad una classe si la probabilità della classe stessa. Se per esempio dividiamo l'intervallo in 11 intervalli, otteniamo, graficamente



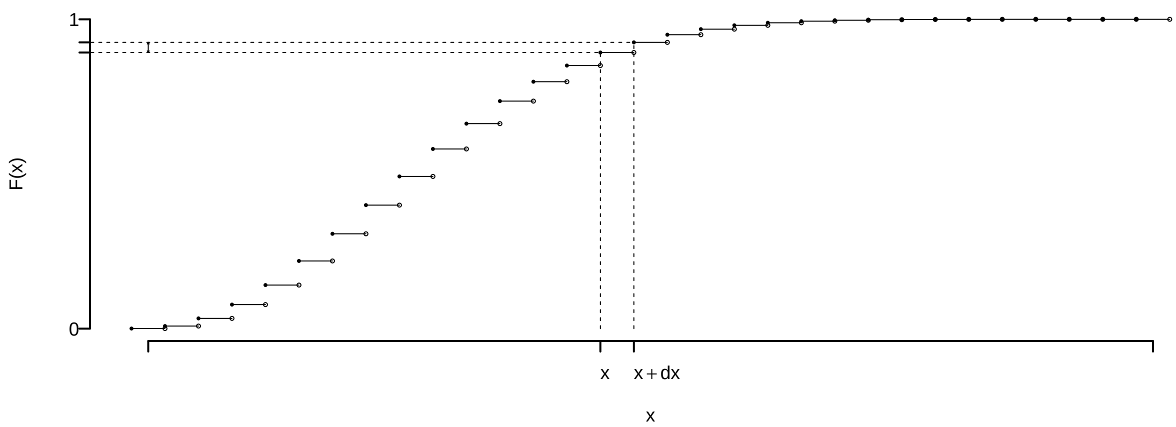
La relativa Funzione di Ripartizione F sarà



Mandare dx a zero significa farlo diventare progressivamente sempre più piccolo

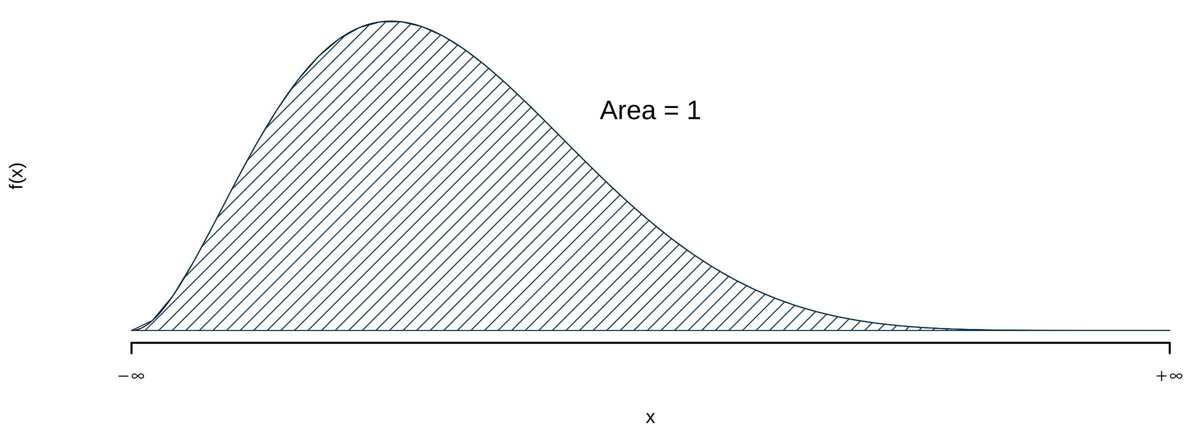


La relativa Funzione di Ripartizione F sarà

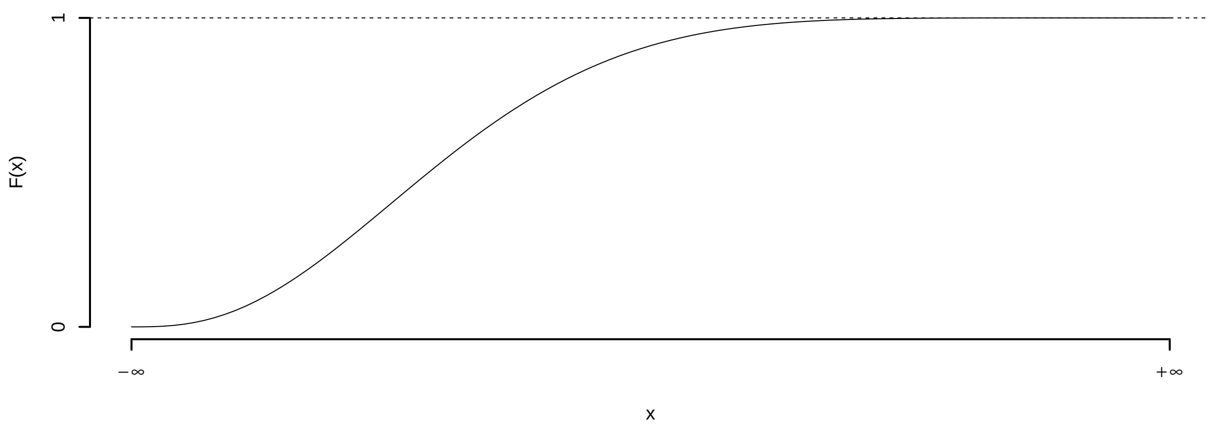


Una VC continua X è caratterizzata dal supporto $S_X \subseteq \mathbb{R}$ e dalla funzione di densità f la cui area sottostante a S_X è uguale ad 1.

$$\int_{-\infty}^{+\infty} f(x)dx = 1$$

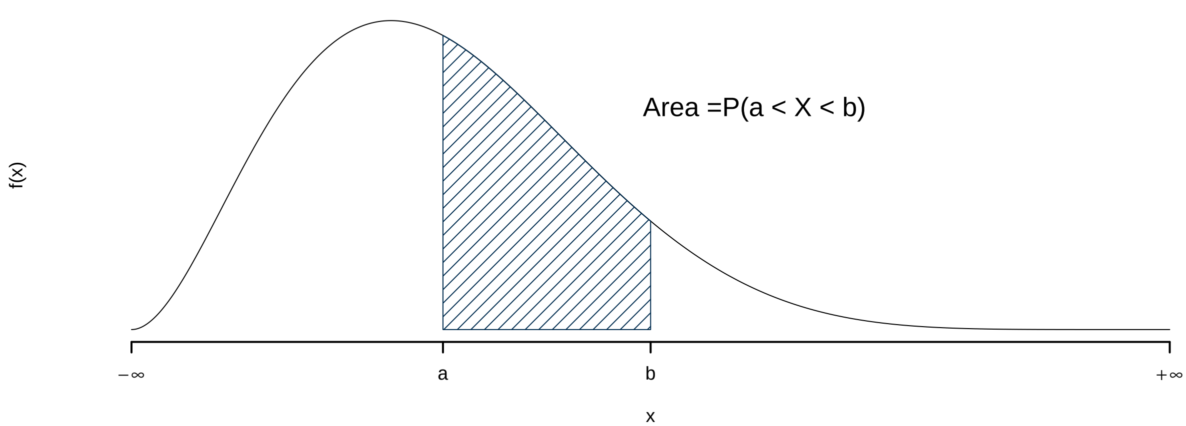


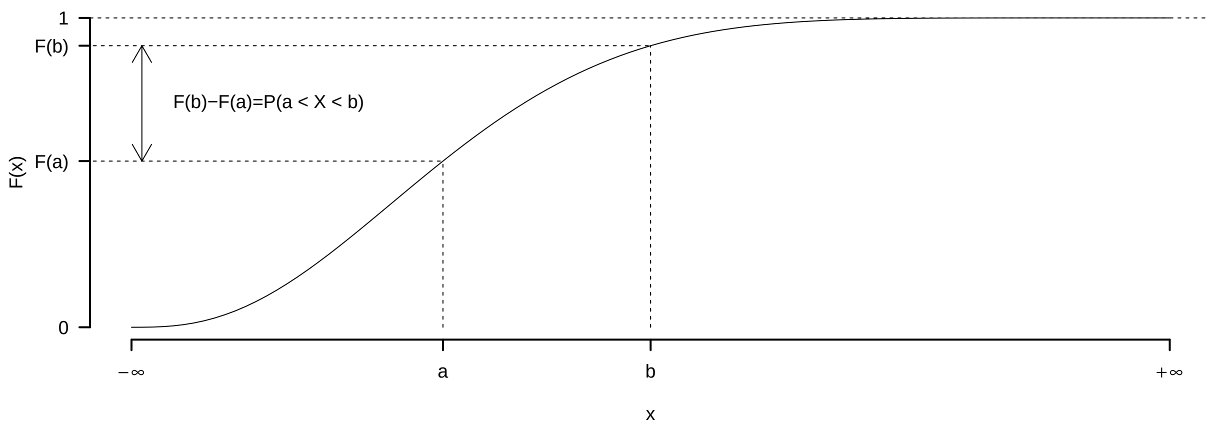
La relativa Funzione di Ripartizione F sarà



La probabilità di un intervallo qualunque (a, b) è l'area sottesa ad f

$$P(a < X < b) = \int_a^b f(x)dx = F(b) - F(a)$$





Questa applicazione esemplifica il passaggio dal discreto al continuo e il concetto di *modello*: VC continue.

6.6.1 Valore Atesso e Varianza di una VC continua

Se X è una VC con supporto $S_X = (x_{\min}, x_{\max})$, con $-\infty \leq x_{\min} < x_{\max} \leq +\infty$, allora si definisce

$$E(X) = \int_{x_{\min}}^{x_{\max}} x f(x) dx$$

e

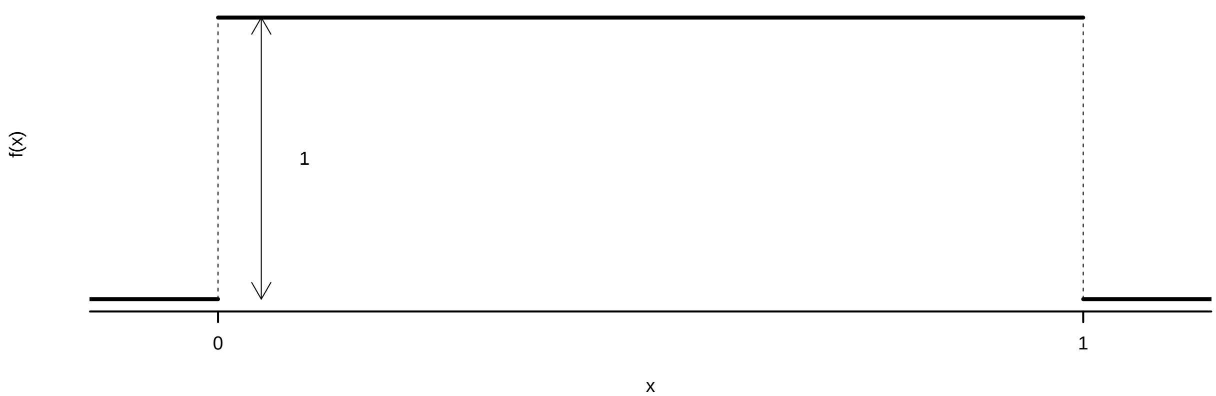
$$V(X) = \int_{x_{\min}}^{x_{\max}} (x - E(X))^2 f(x) dx$$

6.6.2 La VC uniforme

La VC uniforme è utile per prendere confidenza con le VC continue ed è definita ne seguente modo: $S_X = [0, 1]$ e

$$f(x) = \begin{cases} 1, & \text{se } 0 \leq x \leq 1 \\ 0, & \text{altrimenti} \end{cases}$$

Si tratta di una funziona a gradino che vale uno nell'intervallo [0,1] e zero altrove.

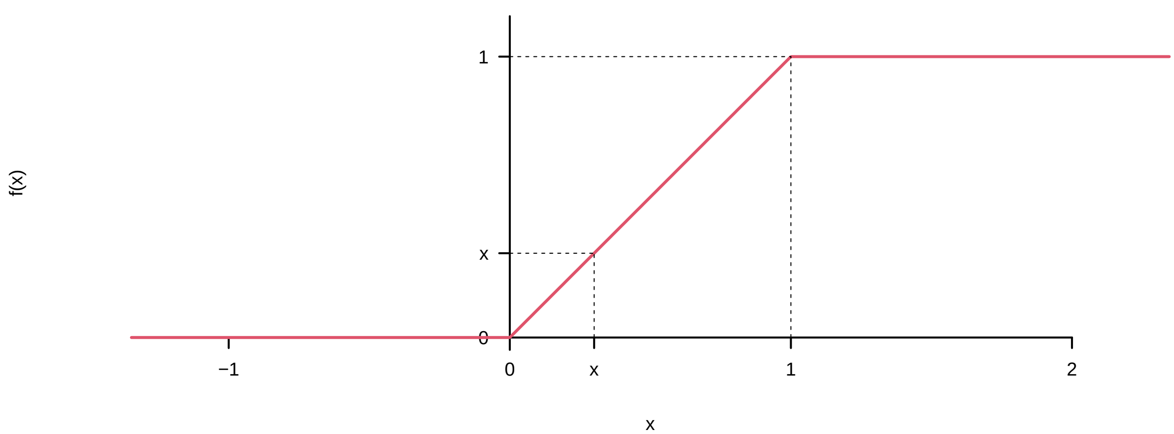


La sua funzione di ripartizione è

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt$$

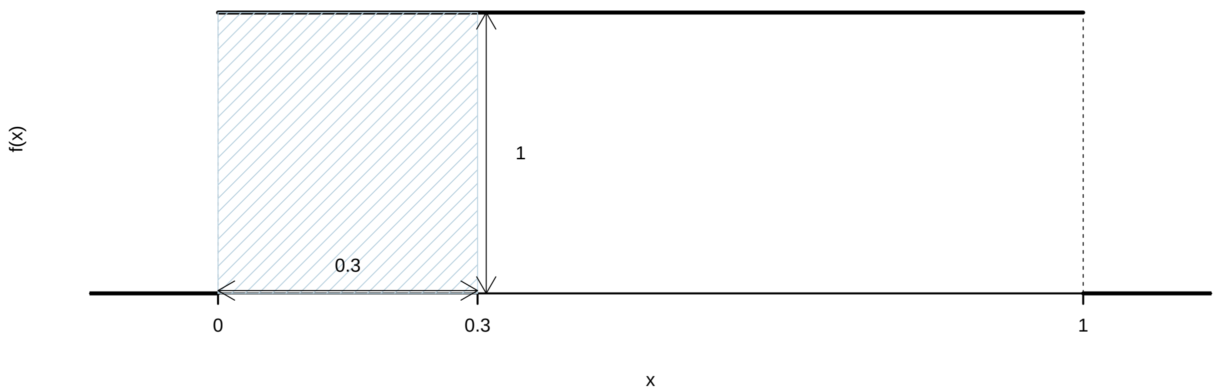
Usiamo la lettera t invece della x perché la x è usata per definire F essendo $f(x) = 0$ per ogni $x < 0$ allora $\int_{-\infty}^{x} f(t)dt = \int_{0}^{x} f(t)dt$ e dunque

$$F(x) = \begin{cases} 0, & \text{se } x < 0 \\ x, & \text{se } 0 \leq x \leq 1 \\ 1, & \text{se } x > 1 \end{cases}$$



Ovvero la probabilità che $X \leq x$ è l'area tra zero ed x della funzione a gradino. Per esempio

$$F(0.3) = P(X \leq 0.3) = (\text{base} = 0.3) \times (\text{altezza} = 1) = 0.3$$



Il valore atteso è

$$
\begin{aligned}
E(X) &= \int_{-\infty}^{+\infty} x f(x) dx \\
&= \int_0^1 x dx \qquad \text{poiché } f \text{ vale 1 in [0,1] e 0 altrove} \\
&= \left[\frac{x^2}{2}\right]_0^1 \\
&= \frac{1^2}{2} - \frac{0^2}{2} \\
&= \frac{1}{2}
\end{aligned}
$$

La varianza

$$
\begin{aligned}
V(X) &= E(X^2) - E^2(X) \\
E(X^2) &= \int_{-\infty}^{+\infty} x^2 f(x) dx \\
&= \int_0^1 x^2 dx \qquad \text{poiché } f \text{ vale 1 in [0,1] e 0 altrove} \\
&= \left[\frac{x^3}{3}\right]_0^1 \\
&= \frac{1^3}{3} - \frac{0^2}{3} \\
&= \frac{1}{3}
\end{aligned}
$$

$$\begin{aligned} V(X) &= \frac{1}{3} - \left(\frac{1}{2}\right)^2 \\ &= \frac{1}{3} - \frac{1}{4} \\ &= \frac{4-3}{12} \\ &= \frac{1}{12} \end{aligned}$$

6.7 Operazioni sulle VC

Le VC sono numeri che non sappiamo in anticipo che valore avranno, quindi siamo autorizzati a fare operazioni. Sia X una VC con supporto S_X e funzione di probabilità f. Se $g : \mathbb{R} \to \mathbb{R}$, posto

$$Y = g(X)$$

Allora Y è una VC.

Esempio 6.7.1. Sia X una VC con supporto $\{-1, 0, 1\}$ r con funzione di probabilità

$$f(x) = \begin{cases} \frac{1}{4}, & \text{se } x = -1 \\ \frac{2}{4}, & \text{se } x = 0 \\ \frac{1}{4}, & \text{se } x = +1 \end{cases}$$

posto

$$Y = g(X) = 2 + 3X$$

allora

$$S_Y = \{-1, 2, 5\}$$

ovvero se $x = -1$, $y = 2 + 3 \times (-1) = -1$, se $x = 0$, $y = 2 + 3 \times 0 = 2$, se $x = 1$, $y = 2 + 3 \times 1 = 5$ e quindi

$$f_Y(y) = \begin{cases} \frac{1}{4}, & \text{se } y = -1 \\ \frac{2}{4}, & \text{se } x = +2 \\ \frac{1}{4}, & \text{se } x = +5 \end{cases}$$

Esempio 6.7.2. Sia X una VC con supporto $\{-1, 0, 1\}$ r con funzione di probabilità

$$f(x) = \begin{cases} \frac{1}{4}, & \text{se } x = -1 \\ \frac{2}{4}, & \text{se } x = 0 \\ \frac{1}{4}, & \text{se } x = +1 \end{cases}$$

posto

$$Y = g(X) = X^2$$

allora

$$S_Y = \{0, 1\}$$

ovvero se $x = -1$, $y = (-1)^2 = 1$, se $x = 0$, $y = 0^2 = 0$, se $x = 1$, $y = 1^2 = 1$ e quindi

$$P(Y = 0) = P(X = 0) = \frac{2}{4} = \frac{1}{2} P(Y = 1) = P(X = -1 \cup X = 1) = P(X = -1) + P(X = 1) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

$$f_Y(y) = \begin{cases} \frac{1}{2}, & \text{se } y = 0 \\ \frac{1}{2}, & \text{se } y = 1 \end{cases}$$

Esempio 6.7.3. Sia X una VC con supporto $\{-1, 0, 1\}$ r con funzione di probabilità

$$f(x) = \begin{cases} \frac{1}{4}, & \text{se } x = -1 \\ \frac{2}{4}, & \text{se } x = 0 \\ \frac{1}{4}, & \text{se } x = +1 \end{cases}$$

posto

$$Y = g(X) = X^2$$

allora

$$S_Y = \{0, 1\}$$

ovvero se $x = -1$, $y = (-1)^2 = 1$, se $x = 0$, $y = 0^2 = 0$, se $x = 1$, $y = 1^2 = 1$ e quindi

$$P(Y = 0) = P(X = 0) = \frac{2}{4} = \frac{1}{2} P(Y = 1) = P(X = -1 \cup X = 1) = P(X = -1) + P(X = 1) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

$$f_Y(y) = \begin{cases} \frac{1}{2}, & \text{se } y = 0 \\ \frac{1}{2}, & \text{se } y = 1 \end{cases}$$

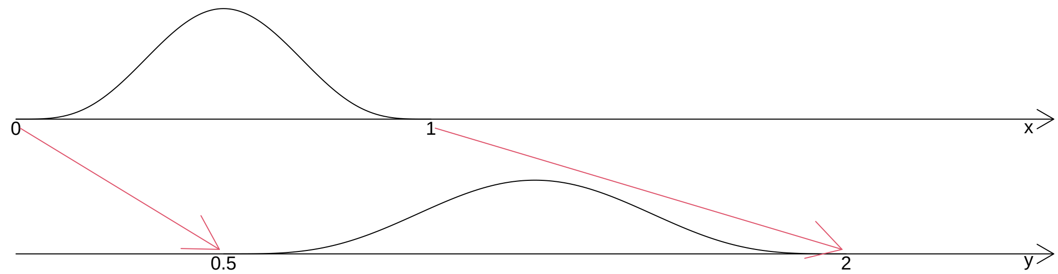
Esempio 6.7.4. Sia X un VC continua con supporto $S_X = [0, 1]$ e funzione di densità $f(x)$. Posto

$$Y = 0.5 + 1.5 \cdot X$$

allora

$$S_Y = [0.5, 2]$$

infatti se $X = 0$ allora $Y = 0.5 + 1.5 \cdot 0 = 0.5$, mentre se $X = 1$ allora $Y = 0.5 + 1.5 \cdot 1 = 2$.


0
1
x
0.5
2
y

7 Variabili Casuali di particolare interesse

Le variabili casuali rappresentano i mattoncini fondamentali con cui si costruiscono i modelli probabilistici e statistici. Alcune variabili casuali, in particolare, emergono come strumenti privilegiati per descrivere classi molto ampie di problemi. La loro importanza deriva dalla semplicità e generalità con cui riescono a rappresentare fenomeni complessi, rendendole elementi centrali in numerose applicazioni.

Queste variabili casuali, che chiamiamo **di particolare interesse**, sono caratterizzate da una distribuzione nota e convenzionalmente identificata da un nome. Esse non sono semplicemente funzioni generiche, ma rappresentano **modelli probabilistici** completamente definiti, a eccezione di un **numero finito di parametri** che devono essere specificati. Questi parametri controllano aspetti fondamentali della distribuzione, come il centro, la dispersione o la forma.

Per rappresentare formalmente questa idea, adottiamo la seguente notazione: se X è una variabile casuale con distribuzione identificata dalla sigla $\mathscr{L}$ e parametrizzata da un vettore di parametri θ, scriviamo:

$$X \sim \mathscr{L}(\theta)$$

Questa espressione si legge:

> *X è distribuita come una variabile casuale di tipo $\mathscr{L}$ con parametro θ.*

La scelta del simbolo θ è intenzionalmente astratta. In questo libro, useremo θ esclusivamente per rappresentare un **parametro generico** o un vettore di parametri. Nessun modello specifico che vedremo avrà mai un parametro chiamato θ: i parametri dei modelli saranno sempre identificati da simboli più specifici, come μ, σ^2, λ, che rifletteranno il significato del parametro nel contesto della distribuzione. Questa scelta è importante per mantenere il rigore formale e per distinguere il livello generale della teoria da quello specifico delle applicazioni.

In ogni caso, l'insieme dei valori possibili per θ, chiamato **spazio dei parametri** e indicato con Θ, definisce il dominio di validità della distribuzione. È l'insieme dei valori per cui la distribuzione ha senso probabilistico.

Questa notazione, benché astratta, è fondamentale per costruire un linguaggio che sia universale, sintetico e adattabile a contesti diversi. Essa ci permette di comunicare non solo che X segue una certa distribuzione, ma anche di precisare come i parametri controllano il comportamento della distribuzione stessa. Sarà utile sia nel contesto probabilistico, dove i parametri sono spesso fissati a priori, sia in quello statistico, dove i parametri devono essere stimati dai dati osservati.

7.1 La VC di Bernoulli

La VC di Bernoulli codifica ogni bipartizione di Ω nei numeri zero e uno. Sia A un evento qualunque la VC di Bernoulli assegna $X = 1$ se A è vero e $X = 0$ se A è falso. Se per esempio estraggo a caso uno studente da un'aula codifico con 1 il fatto che sia maschio e con 0 che sia femmina. La VC X ha come supporto

$S_X = \{0, 1\}$

Definiamo con π la probabilità che $X = 1$ e dunque

$$\begin{cases} X = 1, & \text{con probabilità } \pi \\ X = 0, & \text{con probabilità } 1 - \pi \end{cases}$$

La precedente può essere scritta in modo più compatto

$$f(x) = P(X = x) = \pi^x (1 - \pi)^{1-x}, \quad x \in \{0, 1\}$$

infatti

$$\begin{aligned} f(1) &= P(X = 1) = \pi^1 (1 - \pi)^{1-1} = \pi \\ f(0) &= P(X = 0) = \pi^0 (1 - \pi)^{1-0} = 1 - \pi \end{aligned}$$

Diremo che X è distribuita come una Bernoulli di parametro π e si scrive $X \sim \text{Ber}(\pi)$. Lo spazio dei parametri in questo caso è $\pi \in [0, 1]$, cioè π essendo un probabilità non può essere scelto più piccolo di zero o più grande di uno.

Esempio. Un lancio di una moneta perfetta è descritto da una Bernoulli di paramento $\pi = \frac{1}{2}$

Esempio. L'estrazione da un'urna che contiene 1 sfera vincente e 9 perdenti è descritto da una Bernoulli $\pi = \frac{1}{10}$

7.1.1 Valore Atteso e Varianza

Se $X \sim \text{Ber}(\pi)$, allora

$$\begin{aligned} E(X) &= \sum_{x \in S_X} x f(x) \\ &= 1 \cdot \pi + 0 \cdot (1 - \pi) \\ &= \pi \end{aligned}$$

$$V(X) = \sum_{x \in S_X} (x - E(X))^2 f(x)$$

$$
\begin{aligned}
&= (1-\pi)^2 \cdot \pi + (0-\pi)^2 \cdot (1-\pi) \\
&= (1-\pi)^2 \cdot \pi + \pi^2 \cdot (1-\pi) \\
&= \pi((1-\pi)^2 + \pi(1-\pi)) \\
&= \pi(1-\pi)(1-\pi+\pi) \\
&= \pi(1-\pi)
\end{aligned}
$$

7.1.2 In Sintesi

Se X è distribuita come una Bernoulli di parametro π, allora

Notazione	$X \sim \text{Ber}(\pi)$
Supporto	$S_X = \{0, 1\}$
Funzione di probabilità	$f(x) = \pi^x (1-\pi)^{1-x}$
Spazio dei parametri	$\pi \in [0, 1]$
Valore Atteso	$E(X) = \pi$
Varianza	$V(X) = \pi(1-\pi)$

7.2 La VC Binomiale

La VC Binomiale consente di calcolare la probabilità di avere un certo numero di successi in un certo numero di prove indipendenti le une dalle altre. Se per esempio lanciamo una moneta perfetta 15 volte potremmo chiederci con quale probabilità otterremmo 8 volte testa su 15 lanci.

7.2.1 La VC Binomiale attraverso un esempio

Si consideri un'urna che contiene 1 bussolotto rosso e 2 bianchi, estraiamo 5 volte con reintroduzione, vogliamo calcolare la probabilità di avere 3 bianchi (e quindi 2 rossi), su 5 estrazioni.

$$E = 3 \text{ bianchi su 5 estrazioni}$$

Chi sono gli eventi che realizzano E?

$$
\begin{aligned}
E = \{E_1 &= (B_1 \cap B_2 \cap B_3 \cap R_4 \cap R_5) \cup \\
E_2 &= (B_1 \cap B_2 \cap R_3 \cap B_4 \cap R_5) \cup \\
E_3 &= (B_1 \cap R_2 \cap B_3 \cap B_4 \cap R_5) \cup \\
...\}
\end{aligned}
$$

Osserviamo che ogni estrazione dall'urna è descritta da una Bernoulli $X_1, ..., X_5$, ovvero che la VC che conta 1 se esce Bianco e 0 se esce Rosso è:

$$X_i \sim \text{Ber}(\pi = 2/3), \quad i = 1, ..., 5$$

Le 5 Bernoulli che rappresentano le 5 estrazioni dall'urna sono tutte tra di loro **Indipendenti** e **Identicamente Distribuite**, sono **IID**.

Osserviamo pure che l'evento E è vero se e solo se $X = X_1, ..., X_5 = 3$. E dunque la probabilità dell'evento E_1 sarà

$$\begin{aligned}
P(E_1) &= P(X_1 = 1 \cap X_2 = 1 \cap X_3 = 1 \cap X_4 = 0 \cap X_5 = 0) \\
&= P(X_1 = 1)P(X_2 = 1)P(X_3 = 1)P(X_4 = 0)P(X_5 = 0), \quad \text{in virtù della indipendenza} \\
&= \frac{2}{3} \cdot \frac{2}{3} \cdot \frac{2}{3} \cdot \frac{1}{3} \cdot \frac{1}{3} \\
&= \left(\frac{2}{3}\right)^3 \left(\frac{1}{3}\right)^2 \\
&= 0.0329
\end{aligned}$$

Notiamo che $1 + 1 + 1 + 0 + 0 = 3$. E dunque la probabilità dell'evento E_2 sarà

$$\begin{aligned}
P(E_2) &= P(X_1 = 1 \cap X_2 = 1 \cap X_3 = 0 \cap X_4 = 1 \cap X_5 = 0) \\
&= P(X_1 = 1)P(X_2 = 1)P(X_3 = 0)P(X_4 = 1)P(X_5 = 0), \quad \text{in virtù della indipendenza} \\
&= \frac{2}{3} \cdot \frac{2}{3} \cdot \frac{1}{3} \cdot \frac{2}{3} \cdot \frac{1}{3} \\
&= \left(\frac{2}{3}\right)^3 \left(\frac{1}{3}\right)^2 \\
&= 0.0329
\end{aligned}$$

Anche stavolta notiamo che $1 + 1 + 0 + 1 + 0 = 3$. E dunque la probabilità dell'evento E_3 sarà

$$\begin{aligned}
P(E_3) &= P(X_1 = 1 \cap X_2 = 0 \cap X_3 = 1 \cap X_4 = 1 \cap X_5 = 0) \\
&= P(X_1 = 1)P(X_2 = 0)P(X_3 = 1)P(X_4 = 1)P(X_5 = 0), \quad \text{in virtù della indipendenza} \\
&= \frac{2}{3} \cdot \frac{1}{3} \cdot \frac{2}{3} \cdot \frac{2}{3} \cdot \frac{1}{3} \\
&= \left(\frac{2}{3}\right)^3 \left(\frac{1}{3}\right)^2 \\
&= 0.0329
\end{aligned}$$

Tutti gli eventi E_i hanno la stessa probabilità e dunque

$$P(E_i) = \left(\frac{2}{3}\right)^3 \left(\frac{1}{3}\right)^2 = 0.0329$$

E dunque

$$\begin{aligned}
P(E) &= P(E_1 \cup E_2 \cup E_3 \cup ...) \\
&= P(E_1) + P(E_2) + P(E_3) + ..., \quad \text{per via della incompatibilità} \\
&= \left(\frac{2}{3}\right)^3 \left(\frac{1}{3}\right)^2 + \left(\frac{2}{3}\right)^3 \left(\frac{1}{3}\right)^2 + \left(\frac{2}{3}\right)^3 \left(\frac{1}{3}\right)^2 + \ ... \\
&= \#\{\text{di modi in cui posso disporre 3B su 5}\} \cdot 0.0329
\end{aligned}$$

In quanti modi posso disporre 3 successi nelle 5 prove?

sequenze	conteggio
1,1,1,0,0	1
1,1,0,1,0	2
1,1,0,0,1	3
1,0,1,1,0	4
1,0,1,0,1	5
1,0,0,1,1	6
0,1,1,1,0	7
0,1,1,0,1	8
0,1,0,1,1	9
0,0,1,1,1	10

In 10 modi diversi, per cui:

$$P(E) = 10 \cdot 0.0329 = 0.3292$$

Il *Coefficiente Binomiale* calcola il numero di possibili combinazioni senza doverle scrivere tutte (in quanti modi posso riempire n scatole numerate con k palline indistinguibili).

Nel nostro caso:

$$\binom{5}{3} = \frac{5!}{3!(5-3)!} = \frac{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{(3 \cdot 2 \cdot 1)(2 \cdot 1)} = 10$$

7.2.2 Il modello

Siano $X_1, ..., X_n$ n VC IID, tutte di Bernoulli

$$X_i \sim \text{Ber}(\pi), \quad \pi \in [0,1]$$

Sia

$$X = X_1 + ... + X_n$$

allora X è chiamata VC **Binomiale**

$$X \sim \text{Binom}(n; \pi)$$

è la VC che conta *il numero di successi in* n *prove* indipendenti e identicamente distribuite di Bernoulli, ogni prova ha probabilità π di avere *successo*.

Il supporto di X è

$$S_X = \{0, 1, 2, ..., n\}$$

Lo spazio dei parametri è

$$0 \leq \pi \leq 1 \rightarrow \pi \in [0, 1]$$

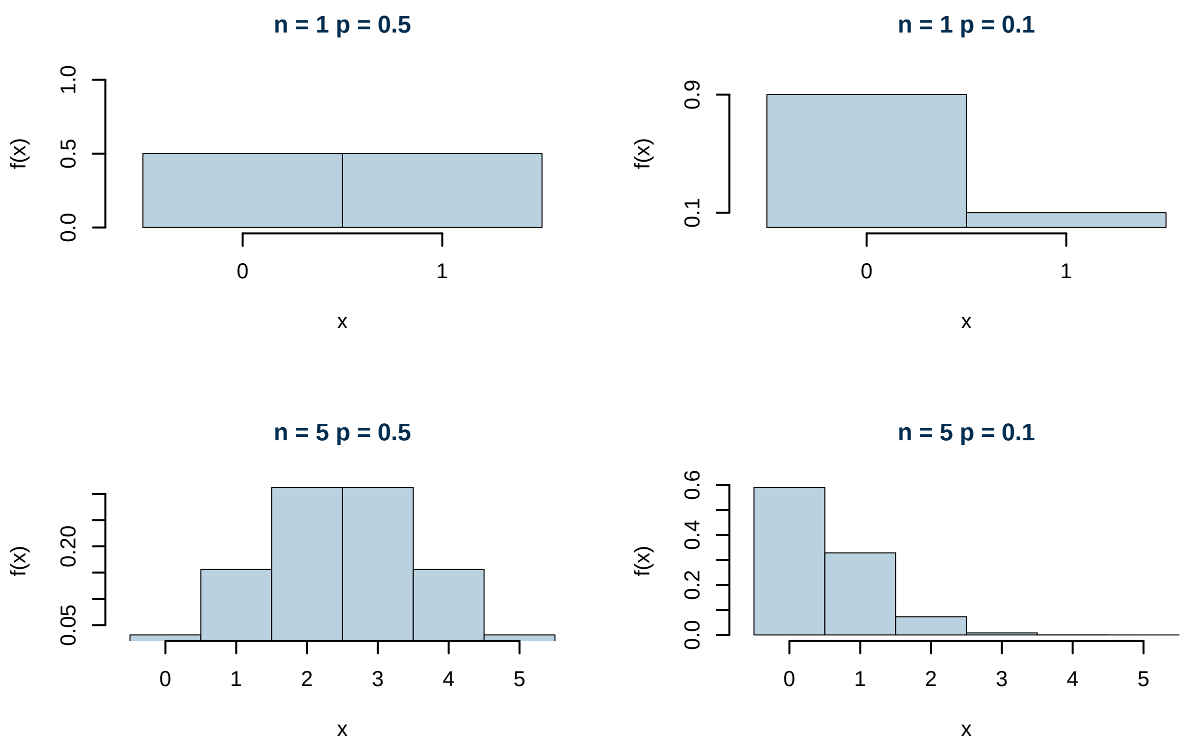
La funzione di probabilità di X è

$$f(x) = P(X = x) = \binom{n}{x} \pi^x (1 - \pi)^{n-x}$$

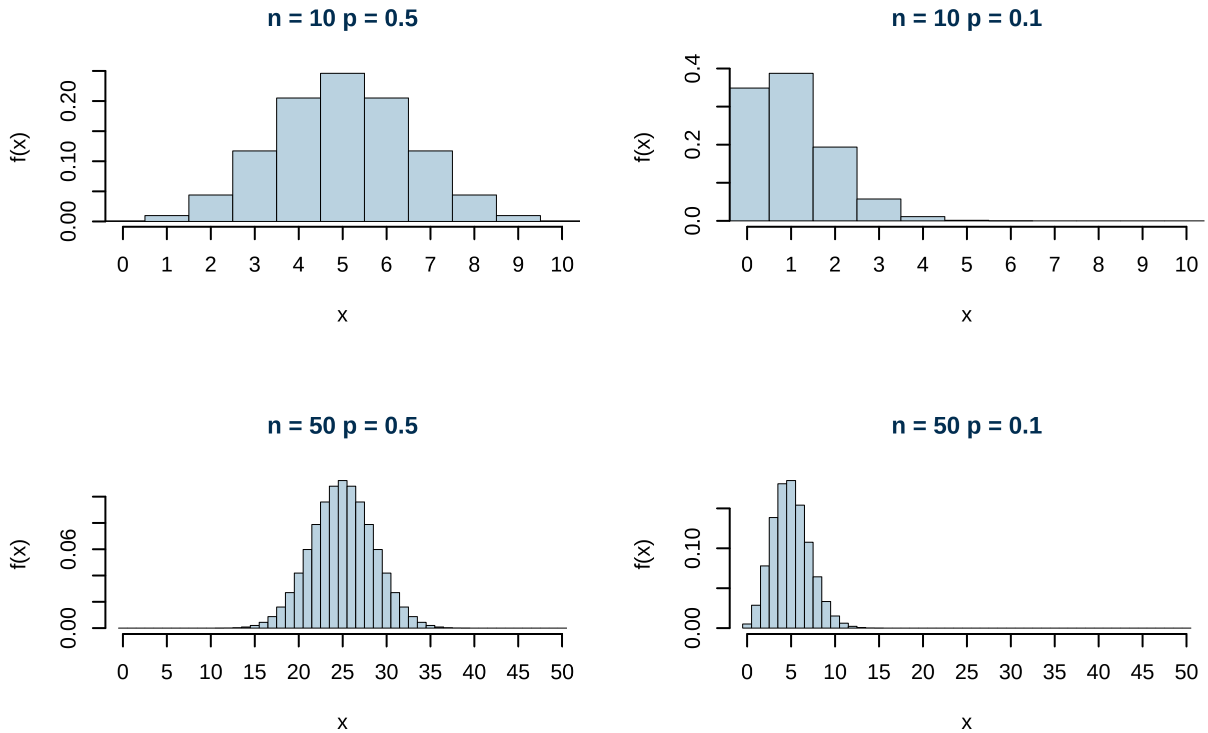
Il valore atteso e la varianza sono, rispettivamente

$$E(X) = n\pi \qquad V(X) = n\pi(1 - \pi)$$

Esempi $(n = 1$ **e** $n = 5)$



Esempi $(n = 10$ **e** $n = 50)$



7.2.3 Dimostrazione del Valore atteso e della Varianza

se $X \sim \text{Binom}(n; \pi)$ allora è la somma di n prove di Bernoulli IID $X_i \sim \text{Ber}(\pi)$:

$$X = X_1 + ... + X_n$$

e sappiamo che

$$E(X_i) = \pi, \qquad V(X_i) = \pi(1-\pi), \;\; \forall i = 1, ..., n$$

quindi

$$\begin{aligned}
E(X) &= E(X_1 + ... + X_n) \\
&= E(X_1) + ... + E(X_n) \\
&= \pi + ... + \pi \\
&= n\pi \\
V(X) &= V(X_1 + ... + X_n) \\
&= V(X_1) + ... + V(X_n), \quad \text{in virtù dell'indipendenza} \\
&= \pi(1-\pi) + ... + \pi(1-\pi) \\
&= n\pi(1-\pi)
\end{aligned}$$

7.2.4 Esempio

Si lancia un dado a sei facce non truccato $n = 5$ volte, si vince se si ottiene un punteggio strettamente maggiore di 4.

1. Calcolare la probabilità non vincere nessuna volta in $n = 5$ tentativi.

Soluzione. La variabile casuale che conta il numero di vittorie in $n = 5$ lanci è la Binomiale, la probabilità di vittoria nel singolo lancio si calcola facilmente:

$$\pi = P(\text{Dado} > 4) = P(\text{Dado} = 5 \cup \text{Dado} = 6) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$$

La probabilità di non vincere in nessuno degli n tentativi equivale a calcolare la probabilità che X sia uguale a zero

$$P(X = 0) = \binom{n}{x}\pi^x(1-\pi)^{n-x} = \binom{5}{3}\frac{1}{3}^0\left(1-\frac{1}{3}\right)^{5-0} = 1 \cdot \left(\frac{2}{3}\right)^5 = 0.0412$$

2. Calcolare la probabilità vincere *al massimo* due volte in $n = 5$ tentativi.

Soluzione. La probabilità di vincere al massimo due volte su n tentativi equivale a calcolare la probabilità che X sia minore o uguale due

$$\begin{aligned}
P(X \leq 2) &= P(X = 0 \cup X = 1 \cup X = 2) \\
&= P(X = 0) + P(X = 1) + P(X = 2) \\
&= \binom{5}{0}\left(\frac{1}{3}\right)^0\left(1-\frac{1}{3}\right)^{5-0} + \binom{5}{1}\left(\frac{1}{3}\right)^1\left(1-\frac{1}{3}\right)^{5-1} + \binom{5}{2}\left(\frac{1}{3}\right)^2\left(1-\frac{1}{3}\right)^{5-2} \\
&= 1 \cdot \left(\frac{2}{3}\right)^5 + 5 \cdot \left(\frac{1}{3}\right)^1\left(\frac{2}{3}\right)^4 + 10 \cdot \left(\frac{1}{3}\right)^2\left(\frac{2}{3}\right)^3 \\
&= 0.7901
\end{aligned}$$

3. Calcolare la probabilità vincere *almeno* due volte in $n = 5$ tentativi.

Soluzione. La probabilità di vincere almeno due volte su n tentativi equivale a calcolare la probabilità che X sia maggiore o uguale due

$$\begin{aligned}
P(X \geq 2) &= P(X = 2 \cup X = 3 \cup X = 4 \cup X = 5) \\
&= 1 - P(X < 2) \\
&= 1 - (P(X = 0) + P(X = 1)) \\
&= 1 - \left(1 \cdot \left(\frac{1}{3}\right)^0\left(1-\frac{1}{3}\right)^{5-0} + 5 \cdot \left(\frac{1}{3}\right)^1\left(1-\frac{1}{3}\right)^{5-1}\right) \\
&= 0.5391
\end{aligned}$$

7.2.5 Proprietà

Proprietà 7.2.1 (Variabile Casuale Binomiale). *Sia $X \sim Binom(n;\pi)$, allora:*

1. *$E(X)$ e $V(X)$ crescono al crescere di n*
2. *La distribuzione di X è perfettamente simmetrica se $\pi = 0.5$*
3. *La distribuzione di X è tende a diventare simmetrica se $n \to \infty$, per ogni valore di $0 < \pi < 1$*
4. ***Proprietà riproduttiva:** Siano $X_1, X_2, ..., X_m$ m VC indipendenti, ognuna $X \sim Binom(n_i, \pi), i = 1, ..., m$, allora*

$$X_1 + ... + X_m \sim Binom(n_1 + ... + n_m; \pi)$$

Si osservi che una Binomiale per $n = 1$ e π equivale ad una Bernoulli di parametro π

$$X \sim \text{Binom}(1, \pi) \sim \text{Ber}(\pi)$$

e quindi dalla proprietà riproduttiva abbiamo che se $X_1, X_2, ..., X_n$ sono n VC indipendenti, ognuna $X \sim \text{Binom}(1, \pi), i = 1, ..., m$, allora

$$X_1 + ... + X_m \sim \text{Binom}(n; \pi)$$

Lanciare 5 volte la stessa moneta e contare il numero di teste equivale, dal punto di vista probabilistico, a lanciare 5 monete identiche e contare il numero di teste.

7.2.6 In Sintesi

Se X è distribuita come una Binomiale di parametro π, allora

Notazione	$X \sim \text{Binom}(n, \pi)$
Supporto	$S_X = \{0, 1, 2, ..., n\}$
Funzione di probabilità	$f(x) = \binom{n}{x} \pi^x (1-\pi)^{n-x}$
Spazio dei parametri	$\pi \in [0, 1], n \in \{0, 1, 2, 3, ...\}$
Valore Atteso	$E(X) = n\pi$
Varianza	$V(X) = n\pi(1-\pi)$

Proprietà riproduttiva: Siano $X_1, X_2, ..., X_m$ m VC indipendenti, ognuna $X \sim \text{Binom}(n_i, \pi), i = 1, ..., m$, allora

$$X_1 + ... + X_m \sim \text{Binom}(n_1 + ... + n_m; \pi)$$

7.3 La VC di Poisson

7.3.1 Obiettivo

Trovare una variabile casuale che ci consenta di probabilizzare particolari fenomeni espressi in forma di conteggi:

- Numero di telefonate in arrivo ad un centralino di emergenza in un'ora
- Numero di interventi richiesti in assistenza in 24h
- Numero di automobili in fila ad un semaforo in orario di punta
- Numero di volte che si è cercato lavoro negli ultimi tre mesi

Sono tutti fenomeni che hanno un supporto indefinito, perché il fenomeno potrebbe essere zero, oppure uno, oppure due, oppure... e non è possibile fissare manualmente infiniti valore di probabilità.

7.3.2 Storia

La distribuzione *Poisson* fu introdotta da Siméon-Denis Poisson nel 1838 nel suo articolo "Recherches sur la probabilité des jugements en matière criminelle et en matière civile". Tra le tante applicazioni venne utilizzata per studiare la variabile casuale X *numero di morti per calcio di cavallo*. Evento raro, ma possibile nel IX secolo. Se il numero n di calci da cavallo fossero stati pochi, la distribuzione sarebbe binomiale, $X \sim \text{Binom}(n;\pi)$, dove π è la probabilità di morire avendo preso un calcio. Pur essendo π un numero molto basso, nel 1800 n invece è un numero molto elevato

7.3.3 Il modello

Poisson assume che il valore atteso di X sia $E(X) = n \cdot \pi = \lambda$, dove λ è un *numero fissato*. Si deduce facilmente che

$$\pi = \frac{\lambda}{n}$$

La distribuzione di X diventa allora

$$\begin{aligned}
P(X = x) &= \binom{n}{x}\pi^x(1-\pi)^{n-x} \\
&= \binom{n}{x}\left(\frac{\lambda}{n}\right)^x\left(1-\frac{\lambda}{n}\right)^{n-x} \\
&= \frac{n!}{x!(n-x)!}\left(\frac{\lambda}{n}\right)^x\left(1-\frac{\lambda}{n}\right)^{n-x} \\
&= \frac{\lambda^x}{x!}\frac{n!}{(n-x)!n^x}\left(1-\frac{\lambda}{n}\right)^{n-x}
\end{aligned}$$

Infine si calcola il limite per $n \to \infty$. Anzitutto analizziamo il supporto

$$S_X = \{0, 1, ..., n\}$$

Se $n \to \infty$ il supporto di X diventa

$$S_X = \{0, 1, ..., +\infty\} = \mathbb{N}$$

Osserviamo la funzione di probabilità:

$$\lim_{n\to\infty} \frac{\lambda^x}{x!} \frac{n!}{(n-x)!n^x} \left(1 - \frac{\lambda}{n}\right)^{n-x} = \frac{\lambda^x}{x!} e^{-\lambda}$$

In quanto

$$\begin{aligned}
\lim_{n\to\infty} \frac{n!}{(n-x)!n^x} &= 1 \\
\lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^{n-x} &= e^{-\lambda}, \qquad \text{si tratta di un limite notevole}
\end{aligned}$$

La variabile di Poisson $X \sim \text{Pois}(\lambda)$ è quella variabile che ha supporto tutti i numeri naturali

$$S_X = \mathbb{N}$$

con funzione di probabilità

$$f(x) = P(X = x) = \frac{\lambda^x}{x!} e^{-\lambda}$$

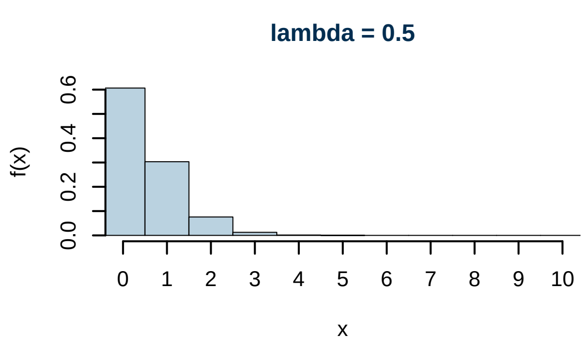
lo spazio dei parametri è

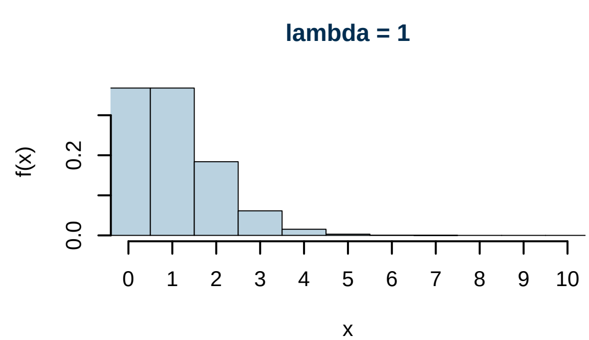
$$\lambda > 0 \to \Theta = (0, +\infty) = \mathbb{R}^+$$

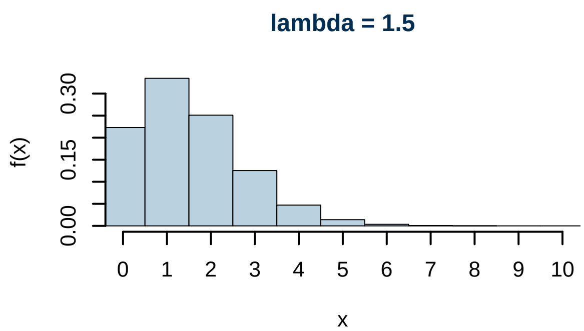
Il valore atteso e la varianza sono

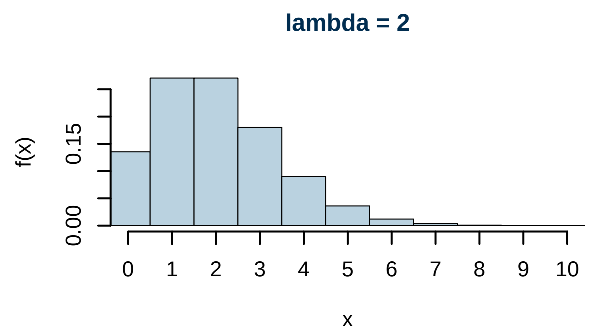
$$E(X) = \lambda, \qquad V(X) = \lambda$$

Alcuni esempi per $\lambda \leq 2$

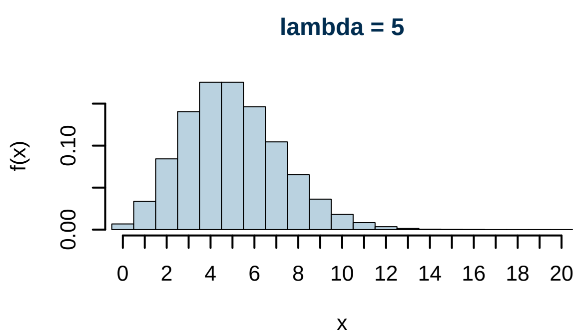


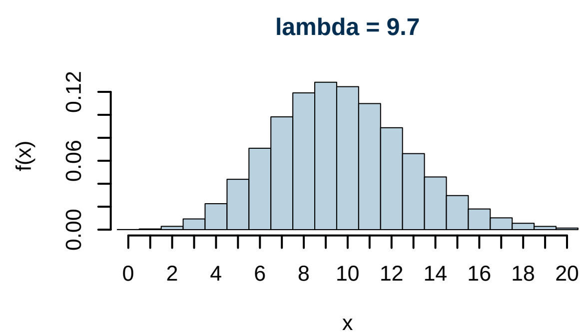


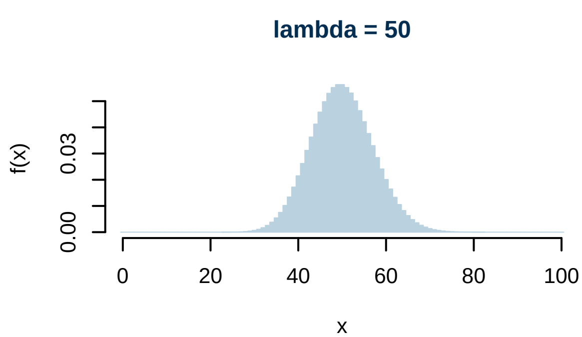


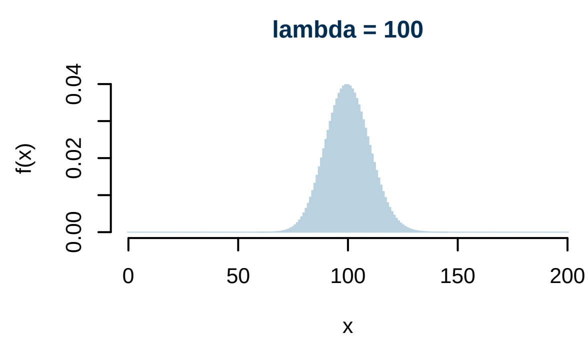


Alcuni esempi per $\lambda \geq 5$









7.3.4 Dimostrazione del Valore atteso e della Varianza della Poisson

Ricordiamo che la Poisson è il limite di una binomiale, il valore atteso della binomiale $n\pi$, al limite si ottiene

$$E(X) = \lim_{n\to\infty} n\pi = \lim_{n\to\infty} n\frac{\lambda}{n} = \lambda$$

la varianza della binomiale è $n\pi(1-\pi)$, al limite

$$V(X) = \lim_{n\to\infty} n\pi(1-\pi) = \lim_{n\to\infty} n\frac{\lambda}{n}\left(1-\frac{\lambda}{n}\right) = \lambda$$

7.3.5 Esempio

Sia X il numero di persone in fila alla cassa 1 di in un supermercato in orario di punta, si *assume* che X sia *ben modellata* da una Poisson di parametro $\lambda = 3.5$, e si scrive $X \sim \text{Pois}(3.5)$.

1. Calcolare la probabilità che *nessuno* sia in fila

Soluzione.

$$P(X = 0) = \frac{\lambda^x}{x!}e^{-\lambda} = \frac{3.5^0}{0!}e^{-3.5} = 0.0302$$

2. Calcolare la probabilità di avere *al massimo* 2 persone in fila

Soluzione.

$$\begin{aligned} P(X \le 2) &= P(X = 0 \cup X = 1 \cup X = 2) \\ &= P(X = 0) + P(X = 1) + P(X = 2) \\ &= \frac{3.5^0}{0!}e^{-3.5} + \frac{3.5^1}{1!}e^{-3.5} + \frac{3.5^2}{2!}e^{-3.5} \\ &= 0.0302 + 0.1057 + 0.185 = 0.3208 \end{aligned}$$

3. Calcolare la probabilità di avere *almeno* 2 persone in fila

Soluzione.

$$\begin{aligned} P(X \ge 2) &= P(X = 2 \cup X = 3 \cup X = 4 \cup ...) \\ &= 1 - P(X < 2) \\ &= 1 - (P(X = 0) + P(X = 1)) \\ &= 1 - (0.0302 + 0.1057) = 0.8641 \end{aligned}$$

4. Calcolare la probabilità che il numero di persone in fila sia *compreso* tra 2 e 4, estremi inclusi

Soluzione.

$$\begin{aligned} P(2 \le X \le 4) &= P(X = 2 \cup X = 3 \cup X = 4) \\ &= 0.185 + 0.2158 + 0.1888 = 0.1359 \end{aligned}$$

7.3.6 Proprietà della Poisson

Proprietà 7.3.1 (Variabile Casuale di Poisson)**.**

1. *Siano* $X_1, X_2, ..., X_n$*,* n *VC indipendenti, tali che*

$$X_i \sim Pois(\lambda_i)$$

allora

$$X_1 + X_2 + ... + X_n \sim Pois(\lambda_1 + \lambda_2 + ... + \lambda_n)$$

2. *Sia* $X \sim Binom(n; \pi)$*, se* n *cresce* $(n \uparrow)$ *e* π *decresce* $(\pi \downarrow)$ *e il loro prodotto rimane fisso in*

$$n\pi = \lambda > 0$$

allora $\lim_{n\to\infty} X \sim Pois(\lambda)$*.*

7.3.7 In Sintesi

Se X è distribuita come una Poisson di parametro λ, allora

Notazione	$X \sim \text{Pois}(\lambda)$
Supporto	$S_X = \{0, 1, 2, ...\} \equiv \mathbb{N}$
Funzione di probabilità	$f(x) = \dfrac{\lambda^x}{x!} e^{-\lambda}$
Spazio dei paramteri	$\lambda \in [0, +\infty) \equiv \mathbb{R}^+$
Valore Atteso	$E(X) = \lambda$
Varianza	$V(X) = \lambda$

Proprietà riproduttiva: Siano $X_1, X_2, ..., X_n$, n VC indipendenti, tali che

$$X_i \sim \text{Pois}(\lambda_i)$$

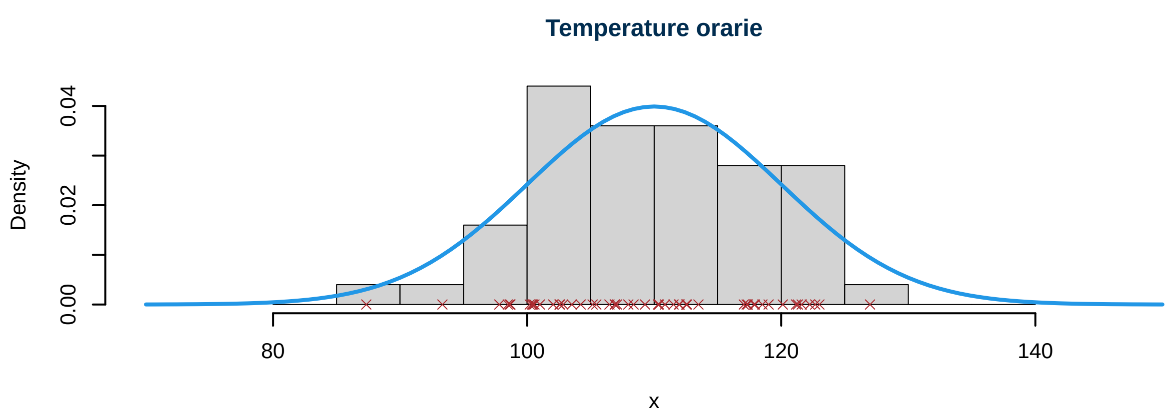
allora

$$X_1 + X_2 + ... + X_n \sim \text{Pois}(\lambda_1 + \lambda_2 + ... + \lambda_n)$$

7.4 La VC Normale

7.4.1 Obiettivo

L'obiettivo è trovare una variabile casuale che ci consenta di probabilizzare fenomeni simmetrici espressi in forma di misure: Temperatura di un locale, utile netto giornaliero di un negozio, errore commesso nel misurare una distanza, ecc. Sono tutti fenomeni che si esprimono con numeri decimali a volte con una precisione che può essere indefinitamente aumentata (ad esempio il peso, il tempo, la temperatura, ecc.), che possono assumere anche valori negativi



7.4.2 Storia

La distribuzione *Normale* fu proposta da Gauss (1809) nell'ambito della teoria degli errori, ed è stata attribuita anche a Laplace (1812), che ne definì le proprietà principali in anticipo rispetto alla trattazione più completa fatta da Gauss.

Nata in ambito fisico, la distribuzione è detta *distribuzione degli errori accidentali* in quanto la distribuzione degli errori commessi nel misurare ripetutamente una stessa grandezza, è molto bene approssimata da questa VC.

$$\text{Misura osservata} = \text{Misura vera} + \text{Errore di misura}$$

e dunque

$$\text{Errore di misura} = \text{Misura osservata} - \text{Misura vera}$$

L'errore di misura è, in tantissimo casi, simmetrico rispetto allo zero:

$$P(\text{Errore di misura} > +\epsilon) = P(\text{Errore di misura} < -\epsilon), \forall \epsilon \geq 0$$

(la probabilità di sovrastimare la misura di una quantità maggiore ϵ è uguale alla probabilità di sottostimarla di una minore quantità ϵ)

$$\lim_{\epsilon \to \infty} P(|\text{Errore di misura}| > \epsilon) = 0$$

7.4.3 Il modello

La VC *Normale*, $X \sim N(\mu, \sigma^2)$ è una distribuzione *continua* con supporto con supporto

$$S_X = [-\infty, +\infty] = \mathbb{R}$$

Dipende da due parametri μ e σ^2. Lo spazio dei parametri è

$$\Theta = \{(-\infty < \mu < +\infty) \cap (\sigma^2 > 0)\} = \mathbb{R} \times \mathbb{R}^+$$

Con densità di probabilità (non richiesta all'esame)

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \qquad \text{Attenzione, in questo caso } \pi = 3.1415... \text{ è la costante trigonometrica}$$

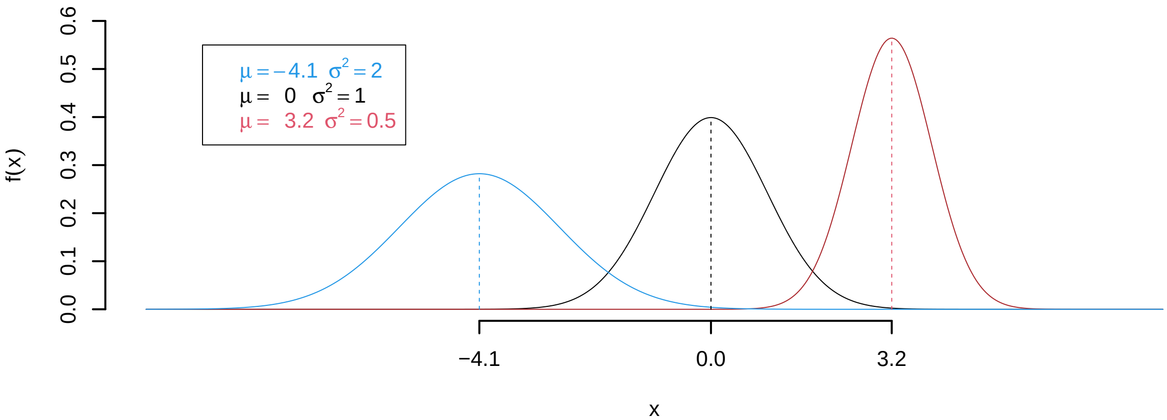
Descrizione:

- ha forma *campanulare*;
- è simmetrica rispetto al parametro μ;
- ha due punti di flesso in $\mu \pm \sigma$;

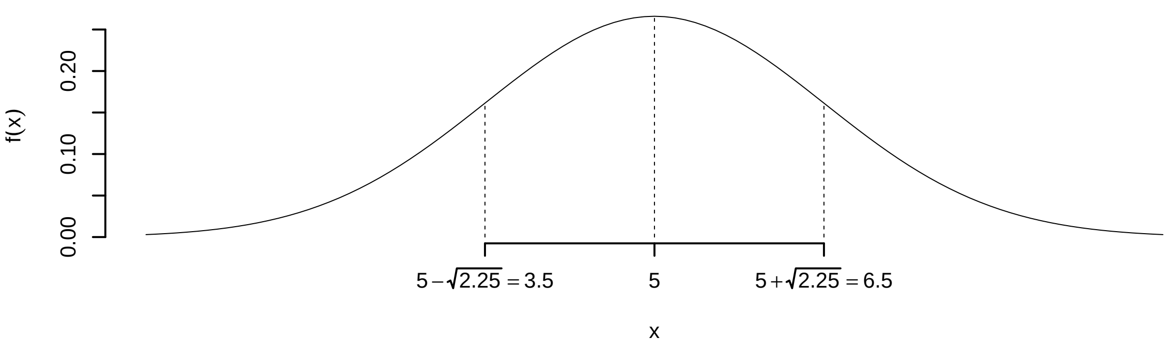
Con valore atteso e varianza:

$$E(X) = \mu, \qquad V(X) = \sigma^2$$

Esempi



Esempio $X \sim N(5, 2.25)$, posizione dei flessi



7.4.4 Proprietà della Normale

La normale gode di molte proprietà interessanti, tutte le combinazioni lineari di una normale sono normali e anche la combinazione lineare di normali indipendenti è normale. Più formalmente

Proprietà 7.4.1 (Proprietà della Normale)**.** *Le principali proprietà dell normale sono*

1. *Sia* $X \sim N(\mu, \sigma^2)$*, posto*

$$Y = a + bX, \qquad a \text{ e } b \text{ due numeri reali qualunque}$$

   allora

$$Y \sim N(a + b\mu, b^2\sigma^2)$$

2. *Siano* $X_1, X_2, ..., X_n$*,* n *VC indipendenti, tali che*

$$X_i \sim N(\mu_i, \sigma_i^2)$$

posto

$$Y = a_1 X_1 + ... + a_n X_n$$

dove $a_1, ..., a_n$ *sono* n *numeri reali qualunque, allora*

$$Y \sim N(a_1\mu_1 + ... + a_n\mu_n, a_1^2\sigma_1^2 + ...a_n^2\sigma_n^2)$$

Il punto 1. ci dice che ogni normale è legata ad un'altra, infatti se $X \sim N(\mu_X, \sigma_X^2)$ e $Y \sim N(\mu_Y, \sigma_Y^2)$, allora

$$\mu_Y + \frac{\sigma_Y}{\sigma_X} \cdot (X - \mu_X) \sim N(\mu_Y, \sigma_Y^2)$$

Il punto 2. ci dice che la combinazione lineare di normali indipendenti è normale, con media e varianza che si combinano secondo le regole di media e varianza.

Casi speciali per $n = 2$

Se $X_1 \sim N(\mu_1, \sigma_1^2)$ e $X_2 \sim N(\mu_2, \sigma_2^2)$ allora:

$$X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

Attenzione

Attenzione

$$X_1 - X_2 \sim N(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2)$$

Esempio 7.4.1. Sia $X_1 \sim N(2.1, 3.4)$ e $X_2 \sim N(5.2, 2.7)$, allora

$$\begin{aligned} X_1 + X_2 &\sim N(2.1 + 5.2 = 7.3, \quad 3.4 + 2.7 = 6.1) \\ X_1 - X_2 &\sim N(2.1 - 5.2 = -3.1, \quad 3.4 + 2.7 = 6.1) \end{aligned}$$

Spiegazione

Dalla proprietà 2. osserviamo che, se $X_1 \sim N(\mu_1, \sigma_1^2)$ e $X_2 \sim N(\mu_2, \sigma_2^2)$ posto

$$Y = X_1 - X_2$$

allora

$$Y = (+1)X_1 + (-1)X_2$$

$a_1 = +1$, $a_2 = -1$ e dunque

$$\begin{aligned} Y &\sim N((+1)\mu_1 + (-1)\mu_2, (+1)^2\sigma_1^2 + (-1)^2\sigma_2^2) \\ &\sim N(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2) \end{aligned}$$

Nella differenza di due normali le medie si **sottraggono** le varianze si **sommano**.

7.4.5 La normale standard

Si definisce la *Normale Standard* la normale con parametri $\mu = 0$ e $\sigma^2 = 1$. Viene indicata con la lettera Z, $Z \sim N(0,1)$.

Dalla proprietà 1. si osserva che se $X \sim N(\mu, \sigma^2)$, posto

$$Z = \frac{X - \mu}{\sigma} = -\frac{\mu}{\sigma} + \frac{1}{\sigma}X$$

allora

$$E(Z) = -\frac{\mu}{\sigma} + \frac{1}{\sigma}E(X) = -\frac{\mu}{\sigma} + \frac{\mu}{\sigma} = 0$$

e

$$V(Z) = \left(\frac{1}{\sigma}\right)^2 V(X) = \frac{\sigma^2}{\sigma^2} = 1$$

Sia $Z \sim N(0,1)$, siano μ e $\sigma^2 > 0$ due numeri qualunque, posto

$$X = \mu + \sigma Z$$

allora

$$X \sim N(\mu, \sigma^2)$$

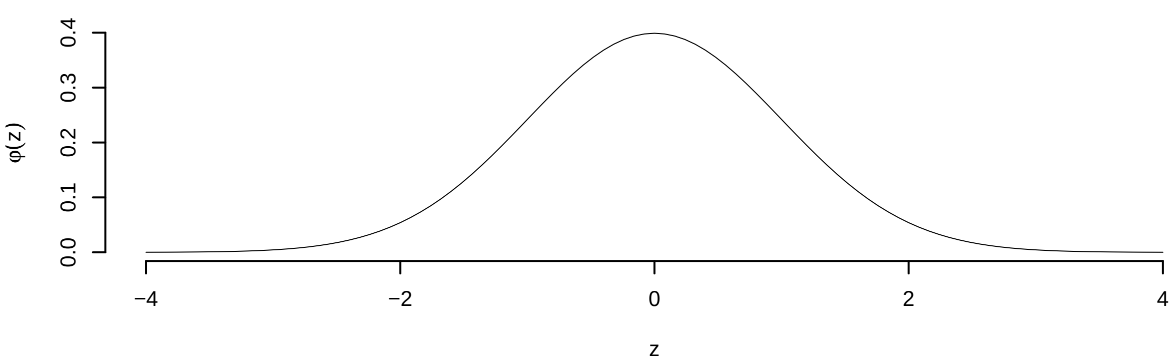
Quindi tutte le normali sono legate tra di loro, in particolare sono tutte legate alla Normale Standard.

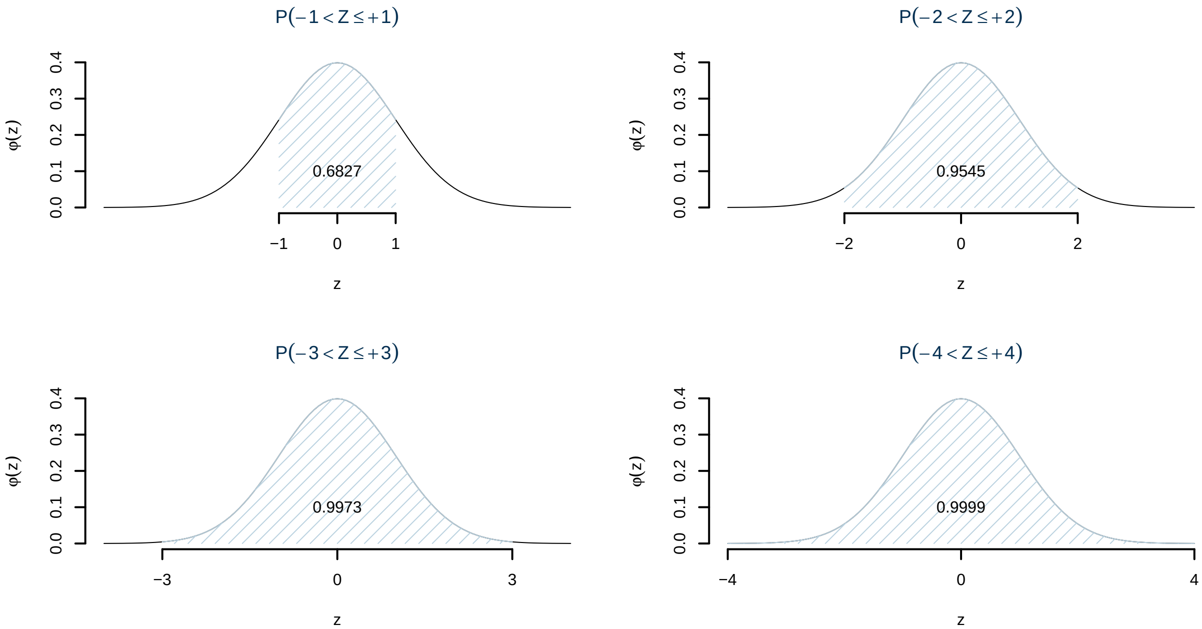
La normale standard ha densità di probabilità chiama ϕ (da non richiesta all'esame)

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$

È una funzione simmetrica intorno a zero e dunque:

$$\phi(-z) = \phi(z)$$



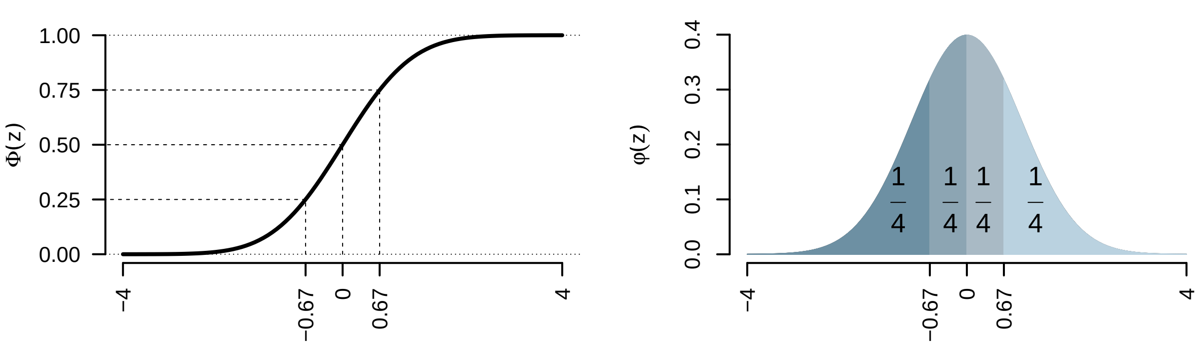


7.4.6 La Funzione di Ripartizione della Normale Standard

Si definisce Φ la funzione di ripartizione della normale standard, $Z \sim N(0,1)$

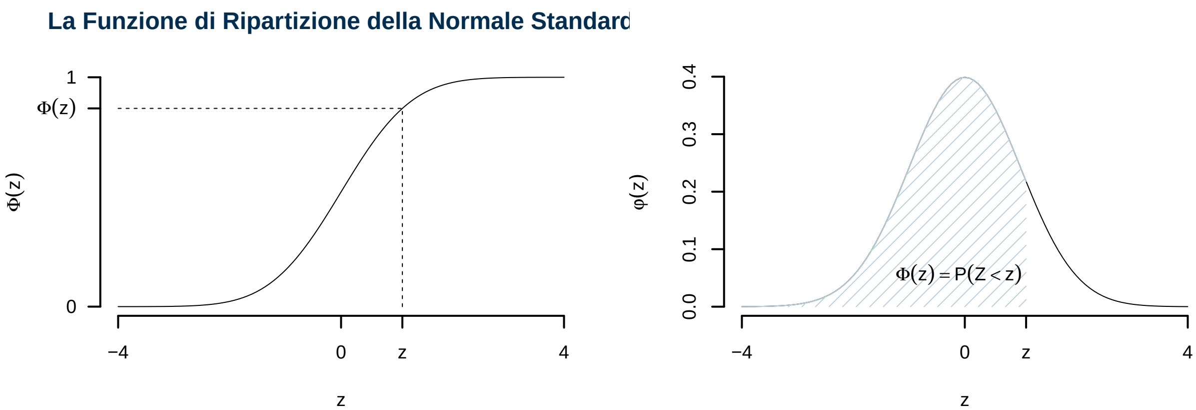
$$\Phi(z) = P(Z \leq z) = \int_{-\infty}^{z} \phi(t)dt$$

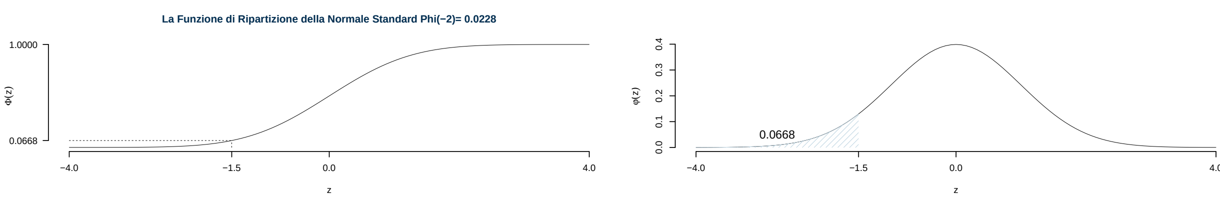
Rappresenta la probabilità che la variabile Z assuma un valore minore o uguale a z

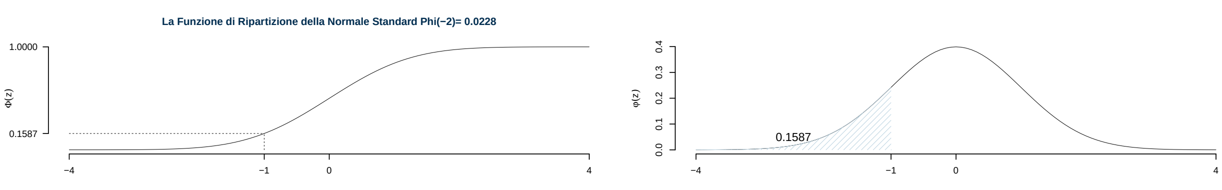


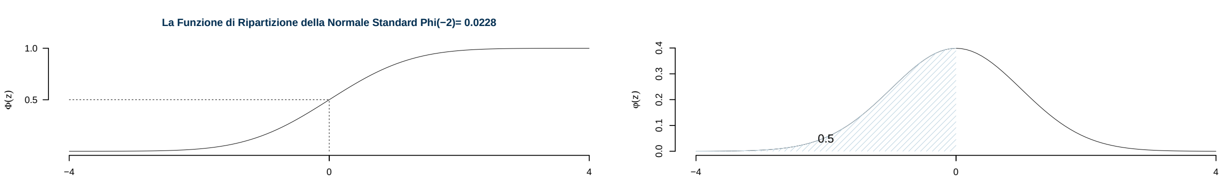
Osserviamo che

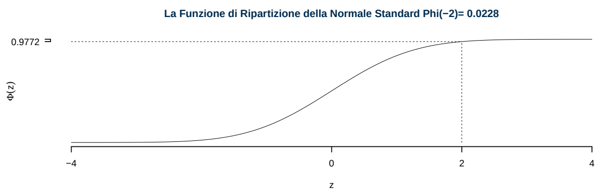
$$
\begin{aligned}
\Phi(0) &= \frac{1}{2} \\
\Phi(z) &= 1 - \Phi(-z)
\end{aligned}
$$

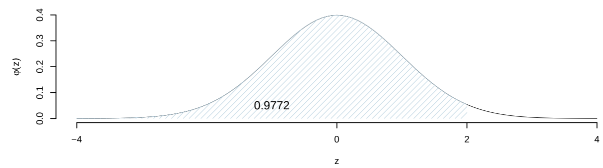


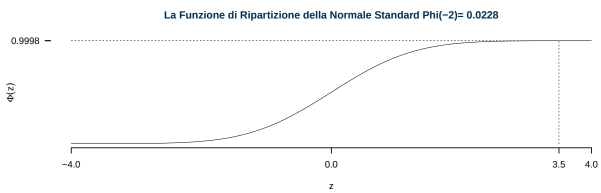


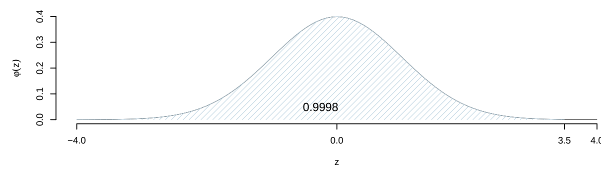


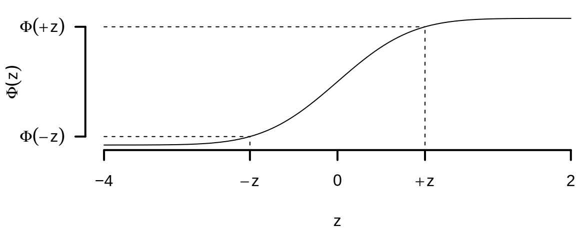


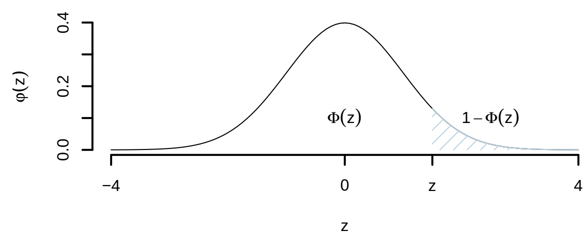


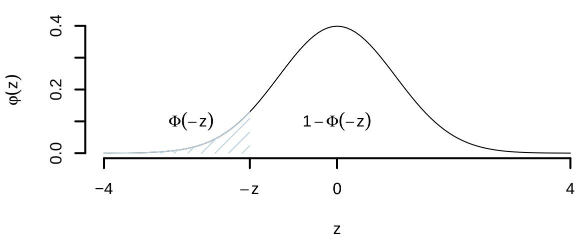




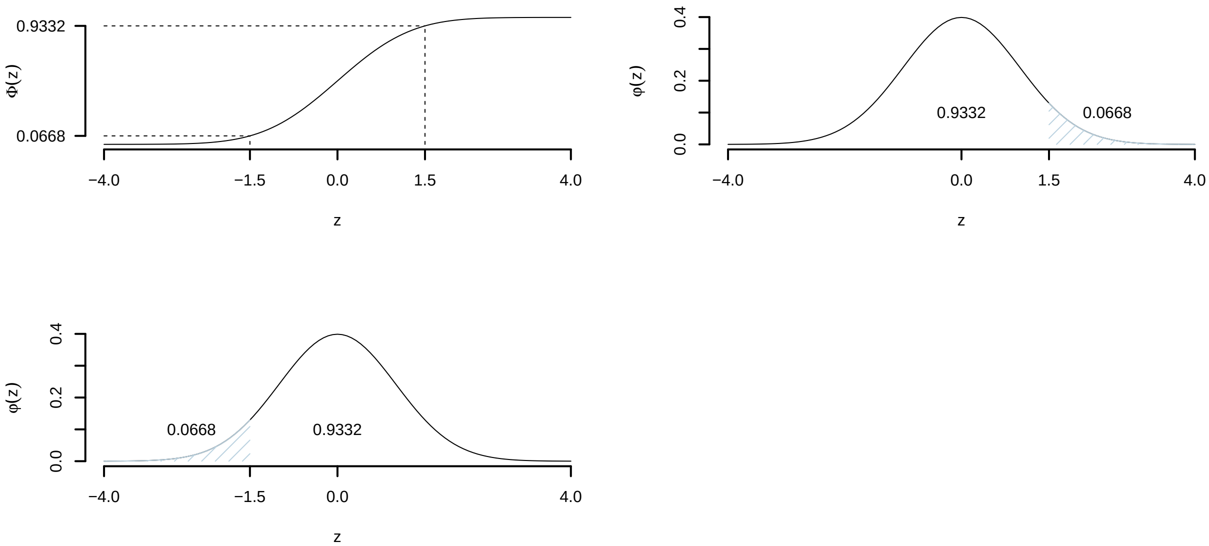




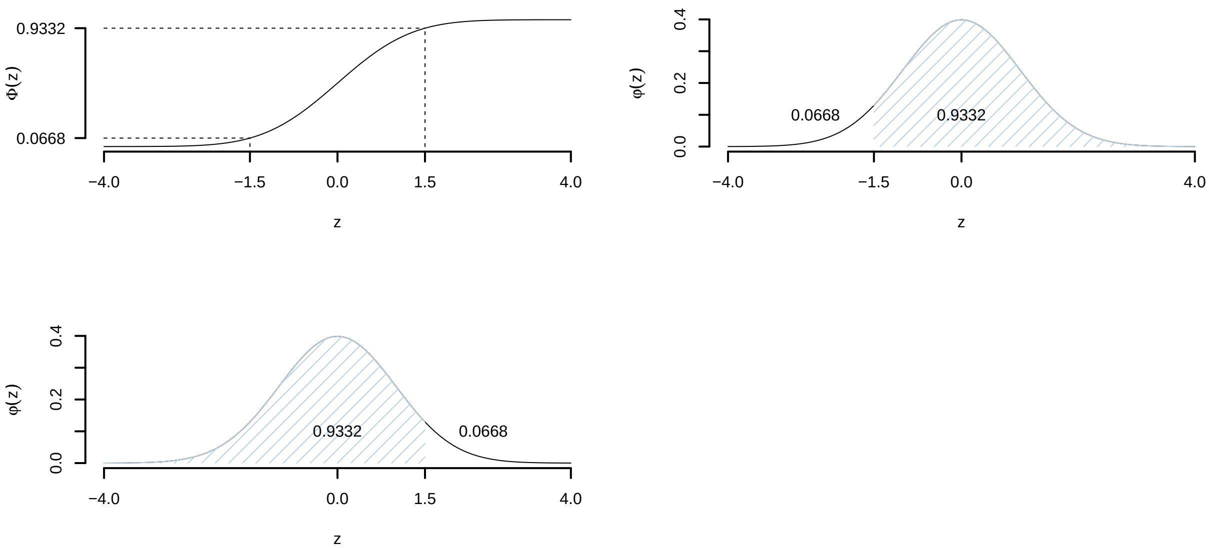




Esempio $z = -1.5$, $\Phi(-1.5) = 1 - \Phi(+1.5) = 0.0668$



Esempio $z = +1.5$, $\Phi(+1.5) = 1 - \Phi(-1.5) = 0.9332$



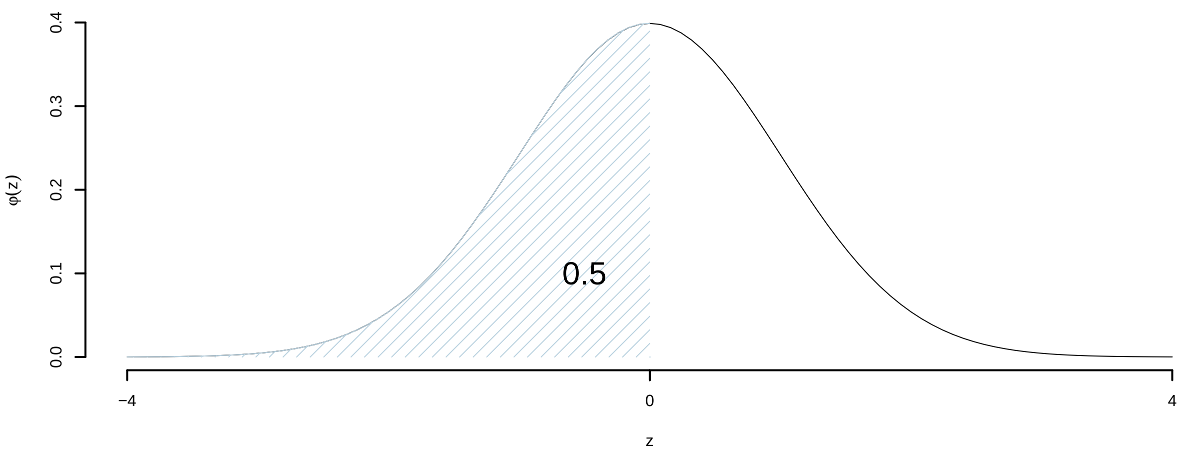
7.4.7 La tavole Statistiche della Z

La VC normale standard $Z \sim N(0, 1)$ è tabulata nelle tavole statistiche. La tabulazione calcola $\Phi(z)$ per $z = 0.00, 0.01, 0.02, ..., 4.08, 4.09$ I valori negativi non sono tabulati ma si ricavano dalle proprietà di simmetria della normale. Le tavole si presentano come una matrice dimensione 40×10 Le righe rappresentano gli interi e i decimi di z, le colonne i centesimi

Esempi:

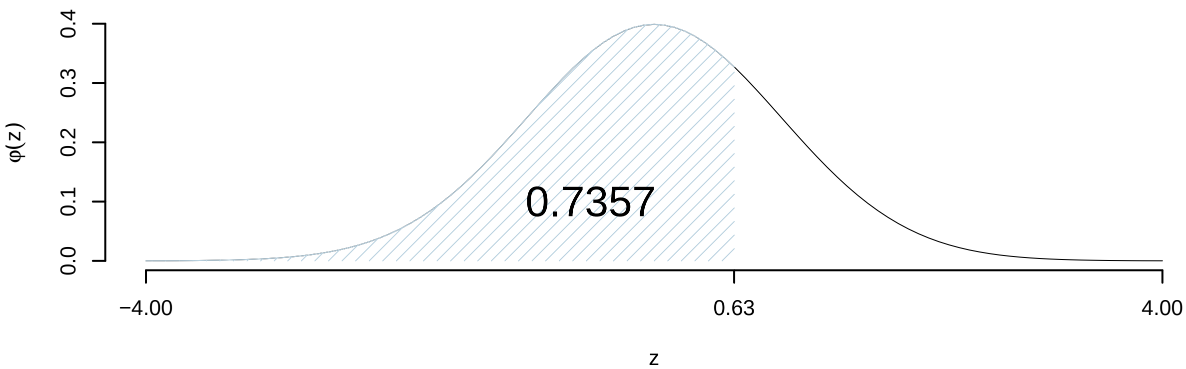
$$\Phi(0) = \frac{1}{2} = 0.5$$

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
continua	...	...	...	...	...	...	...	...	...	...



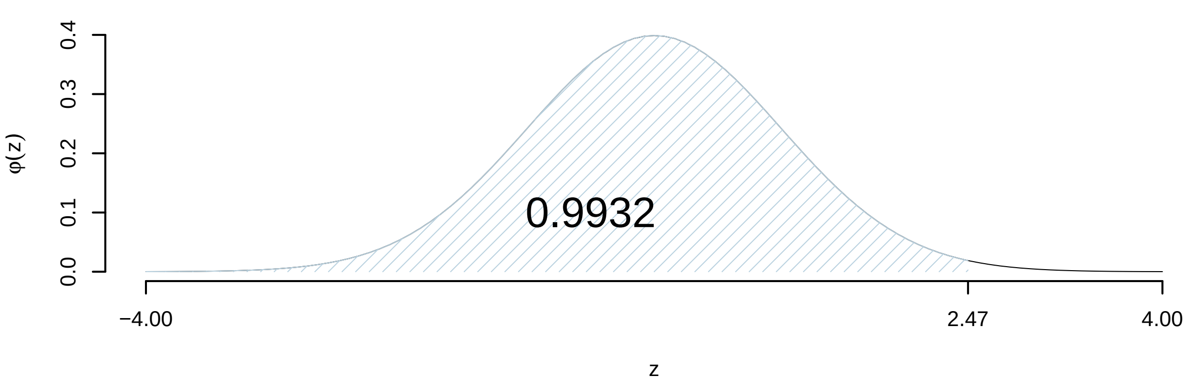
Esempi $\Phi(0.63) = 0.7357$

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
continua	...	...	...	...	...	...	...	...	...	...



Esempi $\Phi(2.47) = 0.9932$

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
continua	...	...	...	...	...	...	...	...	...	...
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
continua	...	...	...	...	...	...	...	...	...	...



7.4.7.1 Uso delle tavole se $z < 0$

Siccome

$$\Phi(-z) = 1 - \Phi(z)$$

allora

$$\Phi(z) = \begin{cases} \text{se } z \geq 0, & \Phi(z), \text{ da leggere direttamente dalle tavole} \\ \text{se } z < 0, & 1 - \Phi(+|z|) \end{cases}$$

Esempio 7.4.2. Calcolare $P(Z \leq 1.12)$: $1.12 \geq 0$ e quindi dalle tavole

$$P(Z \leq 1.12) = \Phi(1.12) = 0.8686$$

Esempio 7.4.3. Calcolare $P(Z \leq -1.12)$: $-1.12 < 0$ e quindi

$$P(Z \leq -1.12) = \Phi(-1.12) = 1 - \Phi(+1.12) = 0.1314$$

7.4.7.2 Calcolo di $P(Z > z)$

Ricordiamo che per definizione

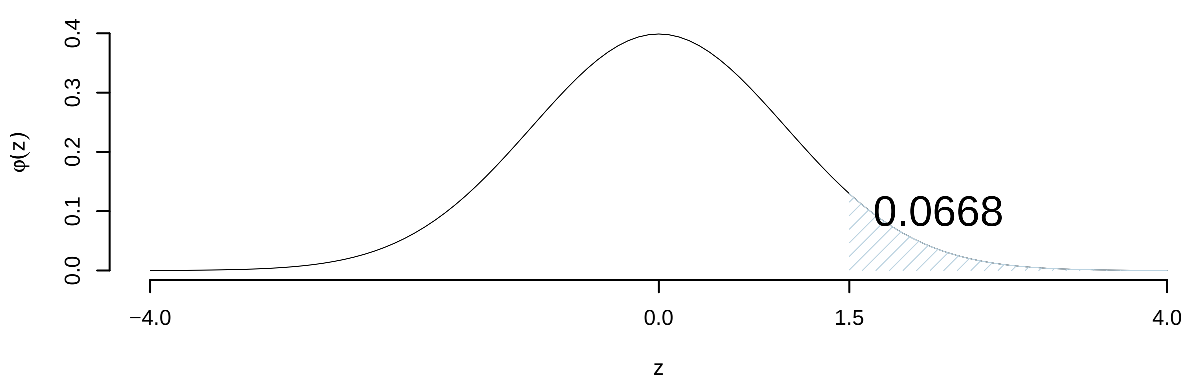
$$\Phi(z) = P(Z \leq z)$$

Siccome il complementare di $\{Z \leq z\}$ è $\{Z > z\}$, dalle proprietà della probabilità abbiamo

$$P(Z > z) = 1 - P(Z \leq z) = 1 - \Phi(z)$$

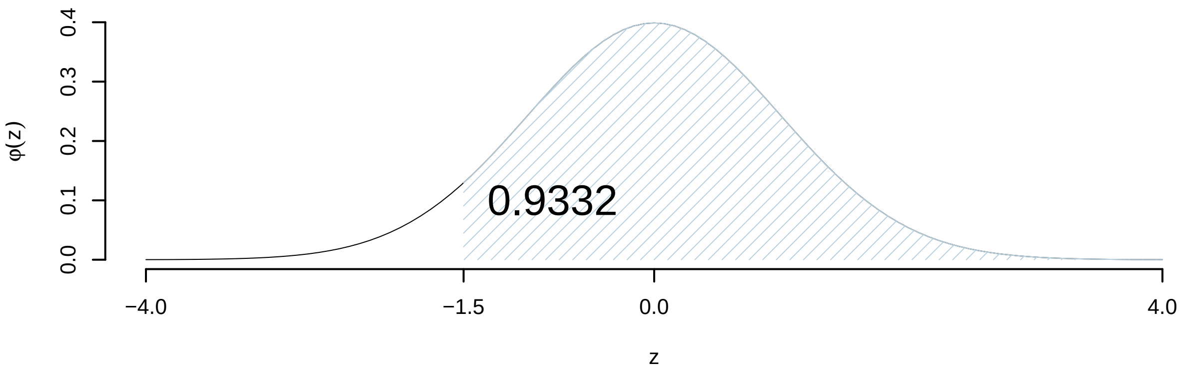
Esempio 7.4.4. calcolare la probabilità che Z si maggiore di 1.5

$$P(Z > 1.5) = 1 - \Phi(1.5) = 1 - 0.9332 = 0.0668$$



Esempio 7.4.5. Calcolare la probabilità che Z si maggiore di -1.5

$$P(Z > -1.5) = 1 - \Phi(-1.5) = 1 - (1 - \Phi(+1.5)) = \Phi(+1.5) = 0.9332$$



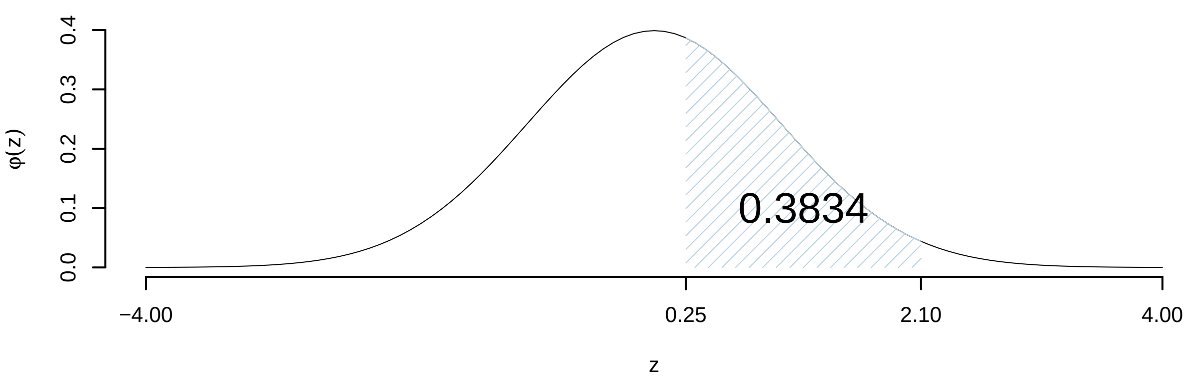
7.4.7.3 Calcolo della probabilità di intervalli

Obiettivo: calcolare la probabilità che Z si avveri nell'intervallo $(z_1, z_2]$. Dalle proprietà della funzione di ripartizione abbiamo

$$P(z_1 < Z \leq z_2) = \Phi(z_2) - \Phi(z_1)$$

Esempio 7.4.6. Calcolare la probabilità che Z si avveri tra 0.25 e 2.1

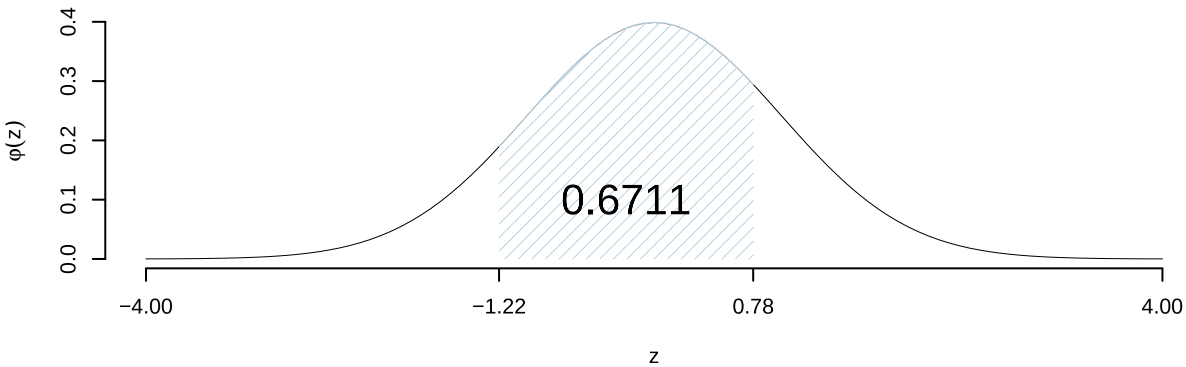
$$P(0.25 < Z \leq 2.1) = \Phi(2.1) - \Phi(0.25) = 0.9821 - 0.5987 = 0.3834$$



Esempio 7.4.7. Calcolare la probabilità che Z si avveri tra -1.22 e 0.78

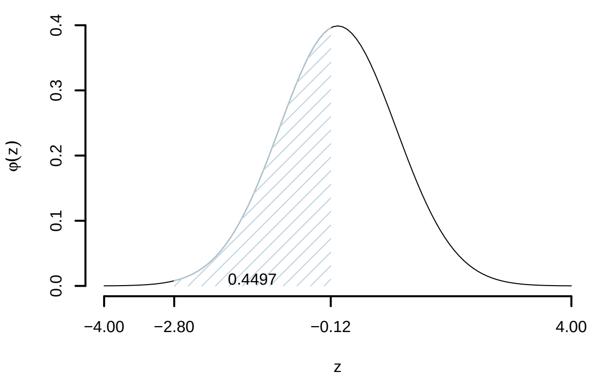
$$P(-1.22 < Z \leq 0.78) \quad = \quad \Phi(0.78) - \Phi(-1.22)$$

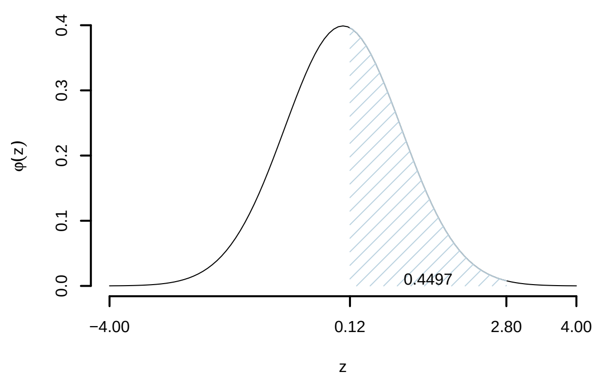
$$
\begin{aligned}
&= \Phi(0.78) - (1 - \Phi(+1.22)) \\
&= 0.7823 - (1 - 0.8888) \\
&= 0.6711
\end{aligned}
$$



Esempio 7.4.8. Calcolare la probabilità che Z si avveri tra -2.8 e -0.12

$$
\begin{aligned}
P(-2.8 < Z \leq -0.12) &= \Phi(-0.12) - \Phi(-2.8) \\
&= (1 - \Phi(+0.12)) - (1 - \Phi(+2.8)) \\
&= \Phi(+2.8) - \Phi(+0.12) \\
&= 0.4497
\end{aligned}
$$





7.4.7.4 Calcolo su una normale qualunque $X \sim N(\mu, \sigma^2)$

Sia data $X \sim N(\mu, \sigma^2)$, obiettivo calcolare

$$P(x_1 < X \leq x_2)$$

dove x_1 e x_2 sono due numeri reali qualunque.

Ricordiamo che

$$Z = \frac{X-\mu}{\sigma}, \qquad Z \sim N(0,1)$$

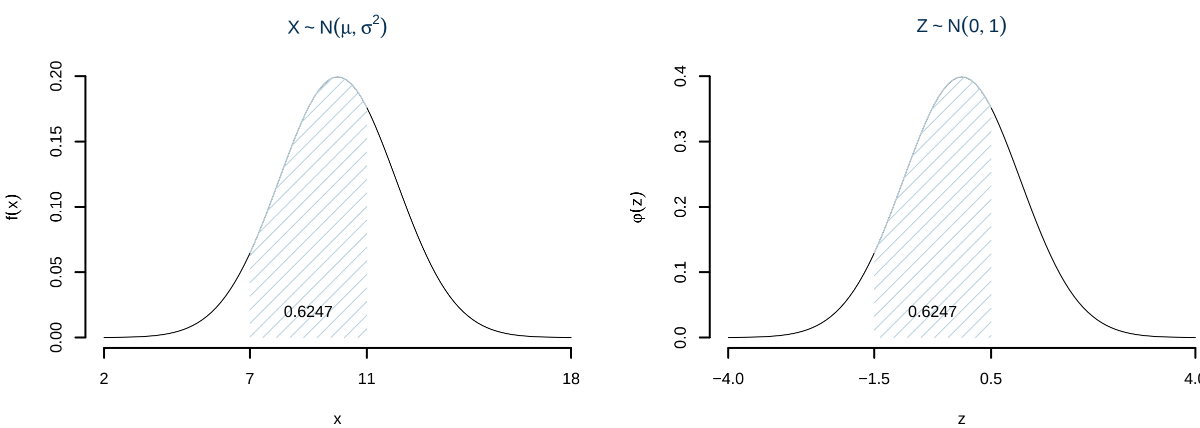
E dunque

$$\begin{aligned} P(x_1 < X \le x_2) &= P\left(\frac{x_1-\mu}{\sigma} < \frac{X-\mu}{\sigma} \le \frac{x_2-\mu}{\sigma}\right) \\ &= P(z_1 < Z \le z_2) \end{aligned}$$

dove $z_1 = \frac{x_1-\mu}{\sigma}$ e $z_2 = \frac{x_2-\mu}{\sigma}$

Esempio 7.4.9. Sia $X \sim N(10,4)$, calcolare la probabilità che $7 < X \le 11$

$$\begin{aligned} P(7 < X \le 11) &= P\left(\frac{7-10}{\sqrt{4}} < \frac{X-\mu}{\sigma} \le \frac{11-10}{\sqrt{4}}\right) \\ &= P\left(-1.5 < Z \le 0.5\right) \\ &= \Phi(0.5) - \Phi(-1.5) \\ &= \Phi(0.5) - (1-\Phi(1.5)) \\ &= 0.6915 - (1-0.9332) \\ &= 0.6247 \end{aligned}$$



Questa applicazione esemplifica l'uso delle tavole e risolve ogni tipo intervallo: Normale

Osservazione

Abbiamo detto che per definizione la Funzione di Ripartizione di una VC X è:

$$F(x) = P(X \le x)$$

Nelle funzioni di ripartizione di VC *continue*, e quindi anche per la normale, accade che

$$P(X \le x) \ge 0$$

ma

$$P(X = x) = 0$$

e quindi, se X è VC continua

$$P(X \leq x) = P(X < x)$$

e analogamente

$$P(X \geq x) = P(X > x)$$

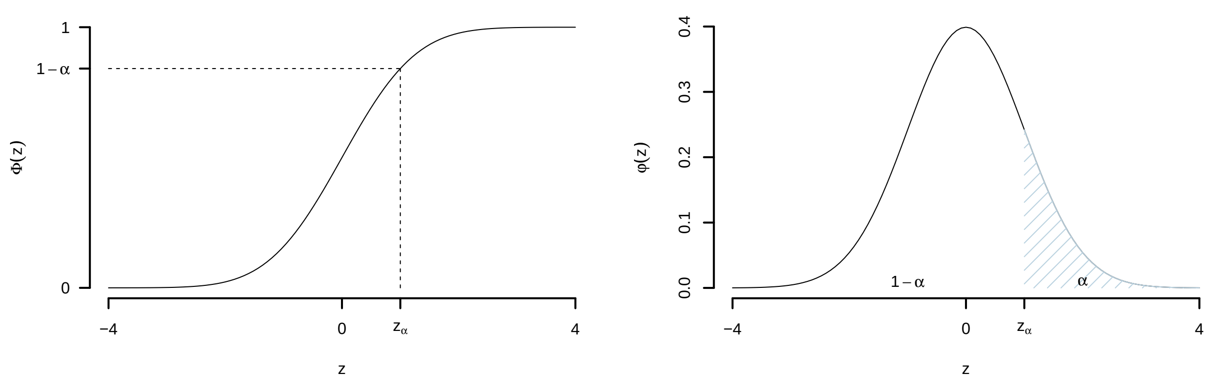
Se X è VC continua

$$P(x_1 < X \leq x_2) = P(x_1 \leq X \leq x_2) = P(x_1 \leq X < x_2) = P(x_1 < X < x_2)$$

7.4.7.5 Il problema inverso

Obbiettivo: data $Z \sim N(0,1)$, sia $0 < \alpha < 1$ un numero fissato, si cerca qual valore z_α tale che:

$$P(Z > z_\alpha) = \alpha$$



Esempio 7.4.10. Per α qualunque. Sia $\alpha = 0.15$ cerchiamo quel valore $z_{0.15}$ tale che

$$P(Z > z_{0.15}) = 0.15$$

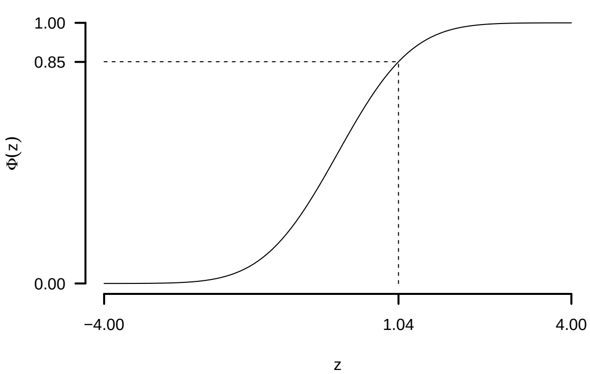
Dalla regola del complemento otteniamo

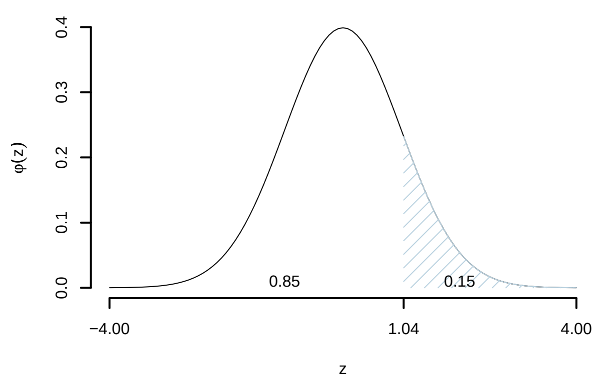
$$P(Z \leq z_{0.15}) = 1 - 0.15 = 0.85$$

Cerchiamo nelle tavole della normale quale valore di z si avvicina di più a 0.85

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
continua	...	...	...	...	...	...	...	...	...	...
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
continua	...	...	...	...	...	...	...	...	...	...

e dunque $z_{0.15} \approx 1.04$





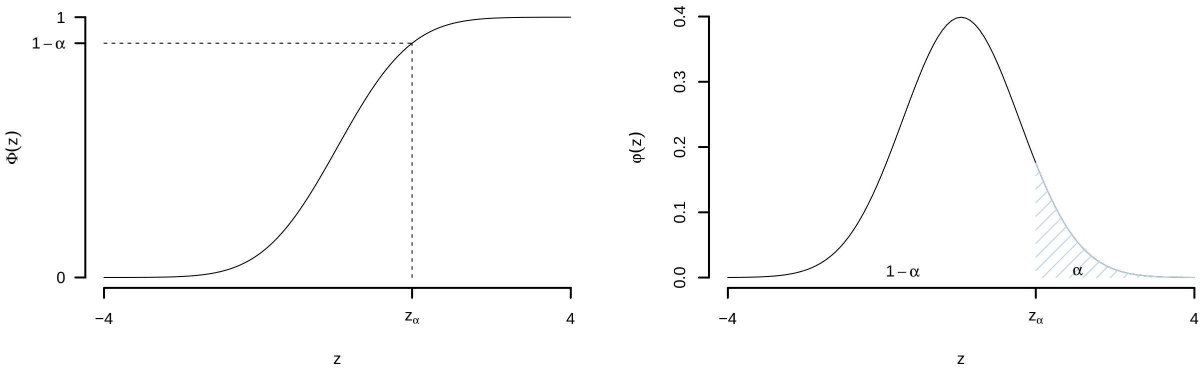
7.4.7.6 Alcuni valori speciali di α

Alcuni valori di α sono di particolare interesse per le future applicazioni questi valori sono

$$0.5, 0.25, 0.1, 0.05, 0.025, 0.01, 0.005, 0.001, 0.0005$$

Per tali valori le tavole della normale vengono arricchite con una tabella supplementare

α	0.5	0.1	0.05	0.025	0.01	0.005	0.001	0.0005
z_α	0.0000	1.2816	1.6449	1.9600	2.3263	2.5758	3.0902	3.2905
$\Phi(z_\alpha) = 1 - \alpha$	0.5	0.9	0.95	0.975	0.99	0.995	0.999	0.9995



Esempio 7.4.11. Se $\alpha = 0.1$ allora $z_{0.1} = 1.2816$

$$P(Z > 1.2816) = 0.1, \qquad P(Z \leq 1.2816) = 0.9$$

Se $\alpha = 0.025$ allora $z_{0.025} = 1.96$

$$P(Z > 1.96) = 0.025, \qquad P(Z \leq 1.96) = 0.975$$

7.4.8 In Sintesi

Se X è distribuita come una Normale di parametri μ e σ^2, allora

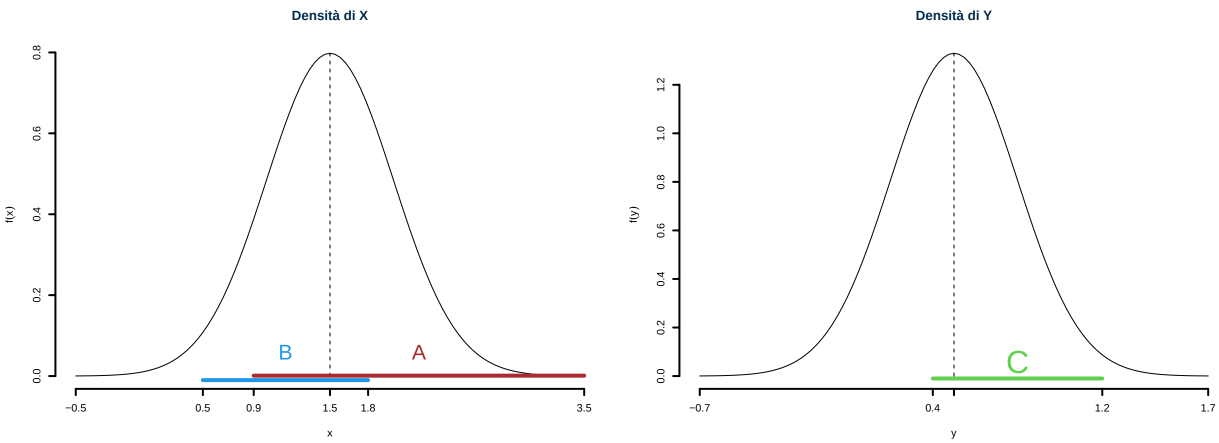
Notazione	$X \sim N(\mu, \sigma^2)$
Supporto	$S_X = (\infty, +\infty) \equiv \mathbb{R}$
Funzione di probabilità	$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$
Spazio dei paramteri	$\mu \in (-\infty, +\infty) \equiv \mathbb{R}$
	$\sigma^2 \in (0, +\infty) \equiv \mathbb{R}^+$
	$(\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}^+$
Valore Atteso	$E(X) = \mu$
Varianza	$V(X) = \sigma^2$

Proprietà:

1. Sia $X \sim N(\mu, \sigma^2)$, posto
$$Y = a + bX, \qquad a \text{ e } b \text{ due numeri reali qualunque}$$
allora
$$Y \sim N(a + b\mu, b^2\sigma^2)$$
2. Siano $X_1, X_2, ..., X_n$, n VC indipendenti, tali che
$$X_i \sim N(\mu_i, \sigma_i^2)$$
posto
$$Y = a_1X_1 + ... + a_nX_n$$
dove $a_1, ..., a_n$ sono n numeri reali qualunque, allora
$$Y \sim N(a_1\mu_1 + ... + a_n\mu_n, a_1^2\sigma_1^2 + ... + a_n^2\sigma_n^2)$$

7.4.9 Esempio

Siano $X \sim N(1.5, 0.25)$ e sia $Y \sim N(0.5, 0.09)$, X e Y **indipendenti**. Si considerino i seguenti insiemi: $A = \{X > 0.9\}$, $B = \{0.5 < X < 1.8\}$, $C = \{0.4 < Y < 1.2\}$



1. Calcolare $P(B)$.

$$
\begin{aligned}
P(0.5 < X \leq 1.8) &= P\left(\frac{0.5-1.5}{\sqrt{0.25}} < \frac{X-\mu_X}{\sigma_X} \leq \frac{1.8-1.5}{\sqrt{0.25}}\right) \\
&= P\left(-2 < Z \leq 0.6\right) \\
&= \Phi(0.6) - \Phi(-2) \\
&= \Phi(0.6) - (1-\Phi(2)) \\
&= 0.7257 - (1-0.9772) \\
&= 0.7029
\end{aligned}
$$

2. Calcolare di $P(A)$.

$$
\begin{aligned}
P(X > 0.9) &= P\left(\frac{X-\mu_X}{\sigma_X} > \frac{0.9-1.5}{\sqrt{0.25}}\right) \\
&= P\left(Z > -1.2\right) \\
&= 1 - P(Z < -1.2) \\
&= 1 - (1-\Phi(1.2)) \\
&= 0.8849
\end{aligned}
$$

alternativamente possiamo osservare che

$$
A = \{X > 0.9\} = \{0.9 < X \leq +\infty\}
$$

$$
\begin{aligned}
P(0.9 < X \leq +\infty) &= P\left(\frac{0.9-1.5}{\sqrt{0.25}} < \frac{X-\mu_X}{\sigma_X} \leq \frac{+\infty-1.5}{\sqrt{0.25}}\right) \\
&= P\left(-1.2 < Z \leq +\infty\right) \\
&= \Phi(+\infty) - \Phi(-1.2) \\
&= \Phi(+\infty) - (1-\Phi(1.2)) \\
&= 1-(1-0.8849) \\
&= 0.8849
\end{aligned}
$$

3. Calcolare $P(C)$.

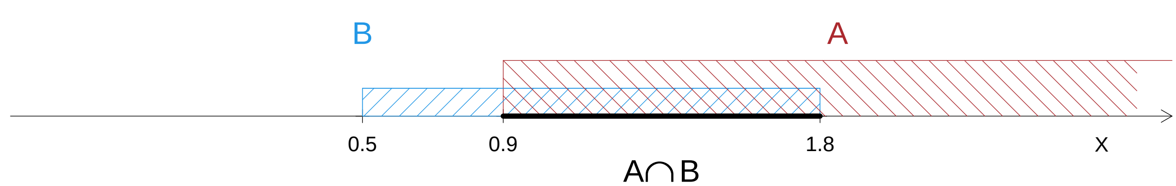
$$
\begin{aligned}
P(0.4 < Y \leq 1.2) &= P\left(\frac{0.4-0.5}{\sqrt{0.09}} < \frac{Y-\mu_Y}{\sigma_Y} \leq \frac{1.2-0.5}{\sqrt{0.09}}\right) \\
&= P\left(-0.33 < Z \leq 2.33\right) \\
&= \Phi(2.33) - \Phi(-0.33) \\
&= \Phi(2.33) - (1-\Phi(0.33)) \\
&= 0.9901-(1-0.6293) \\
&= 0.6194
\end{aligned}
$$

4. Individuare l'insieme $A \cap B$.

Osserviamo che

$$A \cap B = \{X > 0.9\} \cap \{0.5 < X < 1.8\} = \{0.9 < X < 1.8\}$$

e quindi $A \cap B$ è un **intervallo** sulla retta di supporto di X



5. Calcolare $P(A \cap B)$.

$$
\begin{aligned}
P(0.9 < X \leq 1.8) &= P\left(\frac{0.9-1.5}{\sqrt{0.25}} < \frac{X-\mu_X}{\sigma_X} \leq \frac{1.8-1.5}{\sqrt{0.25}}\right) \\
&= P\left(-1.2 < Z \leq 0.6\right)
\end{aligned}
$$

$$
\begin{aligned}
&= \Phi(0.6) - \Phi(-1.2) \\
&= \Phi(0.6) - (1 - \Phi(1.2)) \\
&= 0.7257 - (1 - 0.8849) \\
&= 0.6106
\end{aligned}
$$

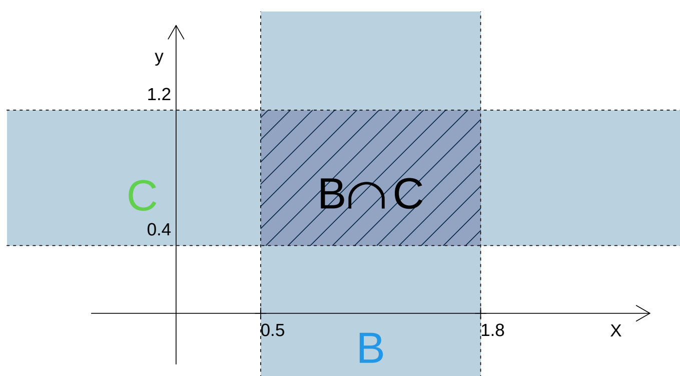
Osservare che $0.6107 = P(A \cap B) \neq P(A)P(B) = 0.6221$, il motivo è che A e B **non sono eventi indipendenti**, perché sono due insiemi della stessa variabile X.

6. Individuare l'insieme $B \cap C$.

Quando invece osserviamo $B \cap C$ vediamo che

$$B \cap C = \{0.5 < X < 1.8\} \cap \{0.4 < Y < 1.2\} = \{0.5 < X < 1.8\} \times \{0.4 < Y < 1.2\}$$

e dunque $B \cap C$ è un **rettangolo** sul piano (X, Y)



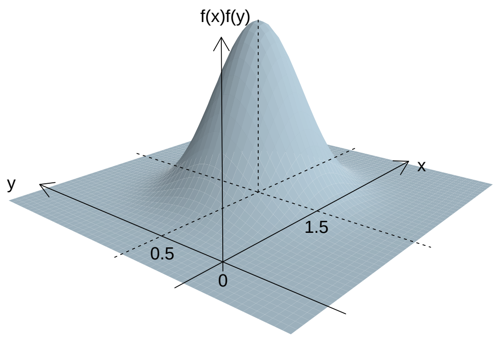
7. Calcolare $P(B \cap C)$.

Anzitutto osserviamo che la coppia di VC (X, Y) vive sul piano e la densità di probabilità è una superficie

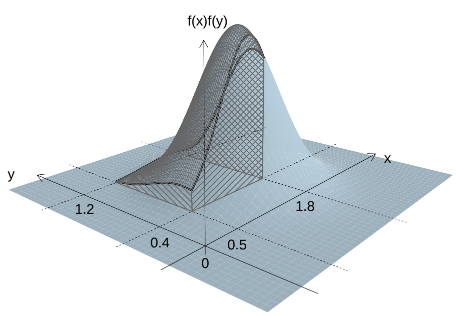
$$f(x, y) = f(x)f(y)$$

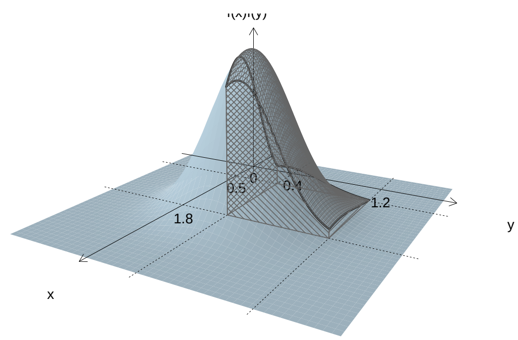
dove $f(x, y)$ e la densità della normale doppia in (x, y) $f(x)$ è la densità di X e $f(y)$ è la densità di Y.

Dal punto di vista grafico:



Calcolare $P(B \cap C)$ equivale a calcolare il **volume** sotteso della densità doppia al rettangolo $B \cap C$:





Operativamente basta moltiplicare le due probabilità:

$P(\{x_1 < X < x_2\} \cap \{y_1 < Y < y_2\}) = P(\{x_1 < X < x_2\})P(\{y_1 < Y < y_2\}),$ in virtù dell'indipendenza

E dunque:

$$P(B \cap C) = P(B)P(C) = 0.703 \times 0.6207 = 0.4364$$

8. Calcolare di $P(A \cup B)$.

Iniziamo osservando che

$$A \cup B = \{X > 0.9\} \cup \{0.5 < X < 1.8\} = \{X > 0.5\}$$

e dunque

$$P(X > 0.5) \quad = \quad P\left(\frac{X - \mu_X}{\sigma_X} > \frac{0.5 - 1.5}{\sqrt{0.25}}\right)$$

$$\begin{aligned} &= P(Z > -2) \\ &= 1 - P(Z < -2) \\ &= 1 - (1 - \Phi(2)) \\ &= 0.9772 \end{aligned}$$

9. Calcolare $P(B \cup C)$.

Iniziamo osservando che

$$B \cup C = \{0.5 < X < 1.8\} \cup \{0.4 < Y < 1.2\}$$

e dunque, dalla regola della somma otteniamo:

$$P(B \cup C) = P(B) + P(C) - P(B \cap C) = 0.703 + 0.6207 - 0.4364 = 0.8874$$

10. Calcolare $P(A|B)$.

Dalla definizione di probabilità condizionata otteniamo:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

e dunque:

$$P(A|B) = \frac{0.6107}{0.703} = 0.8687$$

11. Calcolare $P(B|C)$.

$$P(B|C) = \frac{P(B \cap C)}{P(C)} = \frac{P(B)P(C)}{P(C)} = P(B) = 0.703$$

Osservazione: X e Y sono **indipendenti** e quindi B e C sono a loro volta eventi indipendenti: la probabilità che B sia *vero* non cambia se C è *vero* o se C è *falso*.

12. Sia $W = X + Y$, calcolare la probabilità che $W < 1$.

Dalle proprietà della normale otteniamo che $W \sim N(1.5 + 0.5, 0.25 + 0.09)$ e dunque

$$\begin{aligned} P(W < 1) &= P\left(\frac{W - \mu_W}{\sigma_W} < \frac{1-2}{\sqrt{0.34}}\right) \\ &= P(Z < -1.71) \\ &= 1 - \Phi(1.71) \\ &= 0.0436 \end{aligned}$$

13. Sia $V = X - Y$, calcolare la probabilità che $V < 1.2$.

Dalle proprietà della normale otteniamo che $V \sim N(1.5 - 0.5, 0.25 + 0.09)$ e dunque

$$\begin{aligned} P(V < 1.2) &= P\left(\frac{V-\mu_V}{\sigma_V} < \frac{1.2-1}{\sqrt{0.34}}\right) \\ &= P\left(Z < 0.34\right) \\ &= \Phi(0.34) \\ &= 0.6331 \end{aligned}$$

14. Sia $V = X - Y$, posto $A = \{V > 0\}$ e $B = \{V < 1.2\}$. Calcolare $P(A|B)$

L'evento A è vero se $V > 0$ e l'evento B è vero se $V < 1.2$. Chiedersi quanto vale $P(A|B)$ equivale a chiedersi "è stato estratto un numero casuale V, non so quanto vale ma so che è minore di 1.2, con quale probabilità è maggiore di zero?"

Dalla definizione di probabilità condizionata

$$\begin{aligned} P(A|B) &= \frac{P(A \cap B)}{P(B)} \\ &= \frac{P(\{V > 0\} \cap \{V < 1.2\})}{P(V < 1.2)} \\ &= \frac{P(0 < V < 1.2)}{P(V < 1.2)} \end{aligned}$$

$$\begin{aligned} P(0 < V \leq 1.2) &= P\left(\frac{0-1}{\sqrt{0.34}} < \frac{V-\mu_V}{\sigma_V^2} \leq \frac{1.2-1}{\sqrt{0.34}}\right) \\ &= P\left(-1.71 < Z \leq 0.34\right) \\ &= \Phi(0.34) - \Phi(-1.71) \\ &= \Phi(0.34) - (1 - \Phi(1.71)) \\ &= 0.6331 - (1 - 0.9564) \\ &= 0.5895 \end{aligned}$$

E quindi

$$\begin{aligned} P(A|B) &= \frac{P(A \cap B)}{P(B)} \\ &= \frac{P(0 < V < 1.2)}{P(V < 1.2)} \\ &= \frac{0.5894}{0.6331} \\ &= 0.9311 \end{aligned}$$

14. Sia $V = X - Y$, posto $A = \{V > 0\}$ e $B = \{V < 1.2\}$. Calcolare $P(B|A)$

$$\begin{aligned} P(A|B) &= \frac{P(A \cap B)}{P(A)} \\ &= \frac{P(\{V > 0\} \cap \{V < 1.2\})}{P(V > 0)} \\ &= \frac{P(0 < V < 1.2)}{P(V > 0)} \end{aligned}$$

dobbiamo solo calcolare $P(V > 0)$

$$\begin{aligned} P(V > 0) &= P\left(\frac{V - \mu_V}{\sigma^2} > \frac{0 - 1}{\sqrt{0.34}}\right) \\ &= P\left(Z > -1.71\right) \\ &= 1 - P(Z < -1.71) \\ &= 1 - (1 - \Phi(1.71)) \\ &= 0.9564 \end{aligned}$$

E quindi

$$\begin{aligned} P(B|A) &= \frac{P(A \cap B)}{P(A)} \\ &= \frac{0.5894}{0.9564} \\ &= 0.6163 \end{aligned}$$

Il Teorema del Limite Centrale 8

8.1 Successioni di VC

Una successione di variabili casuali è espressa da un numero infinito di variabili casuali $\{X_1, X_2, ..., X_n, ...\}$.

Esempio 8.1.1. Sia

$$\begin{array}{ccl} X_1 & \sim & N(0,1) \\ X_2 & \sim & N\left(0, \frac{1}{2}\right) \\ X_3 & \sim & N\left(0, \frac{1}{3}\right) \\ ... & & \\ X_n & \sim & N\left(0, \frac{1}{n}\right) \\ ... & & \end{array}$$

Esempio 8.1.2. Sia

$$\begin{array}{ccl} X_1 & \sim & \text{Binom}(1, \pi) \\ X_2 & \sim & \text{Binom}(2, \pi) \\ ... & & \\ X_n & \sim & \text{Binom}(n, \pi) \\ ... & & \end{array}$$

Siamo interessati a sapere se la successione converge ad una VC X. Ma essendo VC e non numeri il concetto di convergenza è più complesso. Esistono diversi tipi di convergenza e non entreremo nella trattazione sistematica del tema. Mostreremo solo le convergenze che ci interessano per sviluppare il resto della teoria.

Definizione 8.1.1 (Convergenza in Distribuzione)**.** Si dice che la successione $\{X_1, X_2, ..., X_n, ...\}$ converge in distribuzione alla VC X se e solo se

$$\lim_{n\to\infty} F_n(x) = F(x), \ \forall x \in S_X$$

dove F_n e F rappresentano la funzione di ripartizione di X_n e di X, rispettivamente. E si scrive

$$X_n \underset{d}{\to} X$$

Definizione 8.1.2 (Convergenza in Probabilità)**.** Si dice che la successione $\{X_1, X_2, ..., X_n, ...\}$ converge in probabilità alla VC X se e solo se

$$\lim_{n\to\infty} P(|X_n - X| < \varepsilon) = 1$$

e si scrive

$$X_n \underset{P}{\rightarrow} X$$

Definizione 8.1.3 (Convergenza in Media Quadratica)**.** Si dice che la successione $\{X_1, X_2, ..., X_n, ...\}$ converge in media quadratica alla VC X se e solo se

$$\lim_{n \to \infty} E\big((X_n - X)^2\big) = 0$$

e si scrive

$$X_n \underset{L^2}{\rightarrow} X$$

8.2 Somme e Medie di VC

Siano $X_1, ..., X_n$ n VC IID, tali che $E(X_i) = \mu$ e $V(X_i) = \sigma^2$. Chiamiamo S_n la somma delle X_i

$$S_n = X_1 + ... + X_n = \sum_{i=1}^{n} X_i$$

Dalle proprietà del valore atteso e della varianza otteniamo

$$\begin{aligned}
E(S_n) &= E(X_1 + ... + X_n) \\
&= E(X_1) + ... + E(X_n) \\
&= \mu + ... + \mu \\
&= n\mu \\
V(S_n) &= V(X_1 + ... + X_n) \\
&= V(X_1) + ... + V(X_n) \\
&= \sigma^2 + ... + \sigma^2 \\
&= n\sigma^2
\end{aligned}$$

Chiamiamo $\bar{X}$ la media delle X_i

$$\bar{X} = \frac{S_n}{n} = \frac{X_1 + ... + X_n}{n} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

Dalle proprietà del valore atteso e della varianza otteniamo

$$\begin{aligned}
E(\bar{X}) &= E\left(\frac{S_n}{n}\right) \\
&= \frac{1}{n} E(S_n) \\
&= \mu \\
V\left(\bar{X}\right) &= V\left(\frac{S_n}{n}\right)
\end{aligned}$$

$$= \frac{1}{n^2}V(S_n)$$
$$= \frac{\sigma^2}{n}$$

Attenzione

Le VC $X_1, ..., X_n$ sono VC qualunque non necessariamente normali, il fatto che chiamiamo $E(X_i) = \mu$ e $V(X_i) = \sigma^2$ è solo una convenzione e non deve fare pensare ai parametri della normale.

Teorema 8.2.1 (Legge dei Grandi Numeri). *Siano $X_1, ..., X_n$ n VC IID, tali che $E(X_i) = \mu$ e $V(X_i) = \sigma^2$. Posto*

$$\bar{X} = \frac{S_n}{n} = \frac{X_1 + ... + X_n}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

allora $\bar{X}$ converge in media quadratica alla VC X che assume il valore μ con probabilità 1.

Dimostrazione. Dalla definizione di convergenza in media quadratica, dobbiamo studiare il limite

$$\lim_{n\to\infty} E((\bar{X} - X)^2)$$

Siccome X è tale che $P(X = \mu) = 1$ allora

$$\lim_{n\to\infty} E((\bar{X} - \mu)^2)$$

ma essendo $E((\bar{X} - \mu)^2) = V(\bar{X}) = \sigma^2/n$ otteniamo

$$\lim_{n\to\infty} \frac{\sigma^2}{n} = 0$$

□

Quindi la media di VC converge alla media dell'urna se il numero di VC aumenta all'infinito. Ma mentre converge al punto della media, cosa succede? A questa domanda rispondono i teoremi centrali del limite.

8.3 Teoremi del Limite Centrale

I *Teoremi del Limite Centrale* (TLC), central limit theorems, sono una famiglia di teoremi sul limite delle somme di VC. Occupano un posto centrale nella teoria della probabilità e dell'inferenza statistica. Esistono molti enunciati a seconda delle ipotesi di partenza. In questo corso mostriamo il teorema più noto e lo decliniamo in tre casi particolari.

I TLC riguardano la convergenza in distribuzione di una successione di somme di una VC. La potenza del teorema è che, non importa quale sia la distribuzione di partenza delle X_i, la loro somma, per n abbastanza grande, è approssimabile con una distribuzione normale. Enunciamo qui tre diversi teoremi che in realtà sono tre diverse declinazioni dello stesso. Iniziamo dal primo e più famoso.

Teorema 8.3.1 (TLC per la Somma). *Siano $X_1,...,X_n$, n Variabili Casuali (VC) Indipendenti e Identicamente Distribuite (IID), tali che $E(X_i) = \mu$, $V(X_i) = \sigma^2$, $\forall i = 1,...,n$. Posto*

$$S_n = X_1 + ... + X_n,$$

allora

$$S_n \underset{a}{\sim} N\left(n\mu, n\sigma^2\right)$$

Il TLC per la somma asserisce che la somma di VC IID, se il numero di addendi è sufficientemente grande, si può approssimare con una normale semplificando notevolmente il calcolo.

Siccome una media non è altro che una somma diviso n, valore atteso e varianza le abbiamo già ricavate, otteniamo

Teorema 8.3.2 (TLC per la Media). *Siano $X_1,...,X_n$, n Variabili Casuali (VC) Indipendenti e Identicamente Distribuite (IID), tali che $E(X_i) = \mu$, $V(X_i) = \sigma^2$, $\forall i = 1,...,n$. Posto*

$$\bar{X} = \frac{S_n}{n} = \frac{X_1 + ... + X_n}{n},$$

allora

$$\bar{X} \underset{a}{\sim} N\left(\mu, \frac{\sigma^2}{n}\right)$$

Il TLC per la media asserisce che la media di VC IID, se il numero di elementi che la contengono è sufficientemente grande, si può approssimare con una normale semplificando notevolmente il calcolo. La varianza di questa normale σ^2/n va a zero per n che diverge, e ci riporta alla legge dei grandi numeri.

Se consideriamo le X_i tutte Bernoulli di parametro π sappiamo che $E(X_i) = \pi$ e $V(X_i) = \pi(1-\pi)$. Dedichiamo alla media di Bernoulli un simbolo speciale che sarà più chiaro più avanti.

$$\hat{\pi} = \frac{X_1 + ... + X_n}{n}$$

sostituendo valore atteso e varianza nel TLC della media otteniamo:

Teorema 8.3.3 (TLC per la Proporzione). *Siano $X_1,...,X_n$, n Variabili Casuali (VC) Indipendenti e Identicamente Distribuite (IID), tali che $X_i \sim \mathrm{Ber}(\pi)$, $\forall i = 1,...,n$. Posto*

$$\hat{\pi} = \frac{S_n}{n} = \frac{X_1 + ... + X_n}{n},$$

allora

$$\hat{\pi} \underset{a}{\sim} N\left(\pi, \frac{\pi(1-\pi)}{n}\right)$$

Il TLC per la proporzione è sempre il TLC per la media ma per VC di Bernoulli. Ci dice che la proporzione osservata su un campione di n VC di Bernoulli è normale con media π e varianza che va a zero con n che diverge.

Nota

La notazione $\sim_a$ non è una notazione standard ma è diventata una prassi con i miei studenti per semplificare la notazione completa che sarebbe più elaborata. Per esempio nel caso del TLC della somma anziché scrivere

$$S_n \underset{a}{\sim} N(n\mu, n\sigma^2)$$

avremmo dovuto scrivere

$$S_n \underset{d}{\rightarrow} X, \qquad X \sim N(n\mu, n\sigma^2)$$

che complica troppo la trattazione.

8.3.1 Esempio Somma

Siano $X_1, X_2, ..., X_n$ n VC IID con supporto $S_X = \{-1, 0, +1\}$ e con funzione di probabilità

$$\begin{cases} P(X=-1) & = \frac{1}{3} \\ P(X=0) & = \frac{1}{3} \\ P(X=+1) & = \frac{1}{3} \end{cases}$$

Consideriamo

$$S_2 = X_1 + X_2$$

il supporto di S_2 sarà

$$S_{S_2} = \{-2, -1, 0, +1, +2\}$$

$S_2 = -2$ se sia $X_1 = -1$ e $X_2 = -1$, $S_2 = -1$ se sia $X_1 + X_2 = -1$ ecc. mettiamo in tabella

	-1;	$1/3$	0;	$1/3$	1;	$1/3$
-1; $1/3$	-2;	$1/9$	-1;	$1/9$	0;	$1/9$
0; $1/3$	-1;	$1/9$	0;	$1/9$	1;	$1/9$
1; $1/3$	0;	$1/9$	1;	$1/9$	2;	$1/9$

E ricaviamo la distribuzione di, S_2

S_2	-2	-1	0	1	2
$P(S_2)$	$1/9$	$2/9$	$3/9$	$2/9$	$1/9$

Se siamo interessati ad S_3

$$S_3 = X_1 + X_2 + X_3 = S_2 + X_3$$

lavoriamo come prima facendo la somma tra S_2 e X_1

	-1;	$1/3$	0;	$1/3$	1;	$1/3$
$-2; 1/9$	-3;	$1/27$	-2;	$1/27$	-1;	$1/27$
$-1; 2/9$	-2;	$2/27$	-1;	$2/27$	0;	$2/27$
$0; 3/9$	-1;	$3/27$	0;	$3/27$	1;	$3/27$
$1; 2/9$	0;	$2/27$	1;	$2/27$	2;	$2/27$
$2; 1/9$	1;	$1/27$	2;	$1/27$	3;	$1/27$

E ricaviamo la distribuzione di, S_3

S_3	-3	-2	-1	0	1	2	3
$P(S_3)$	$1/27$	$3/27$	$6/27$	$7/27$	$6/27$	$3/27$	$1/27$

Iteriamo il ragionamento per S_4

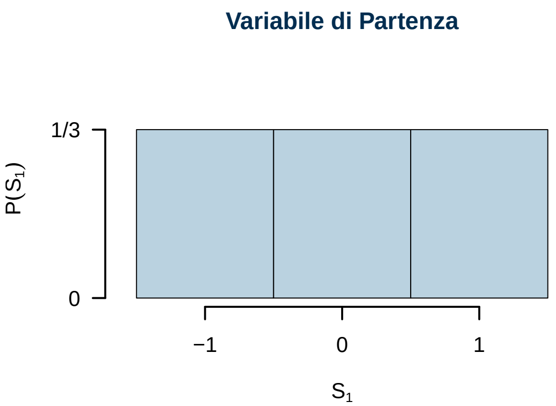
$$S_4 = S_3 + X_4$$

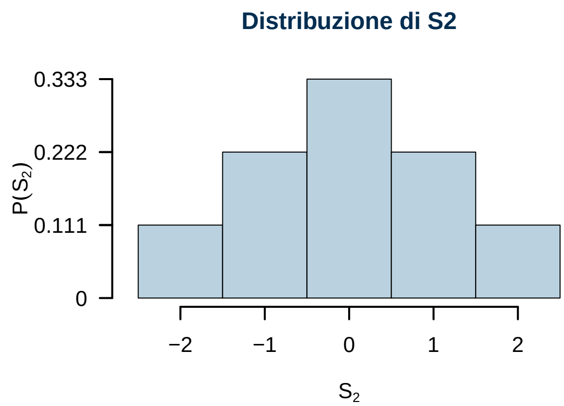
	-1;	$1/3$	0;	$1/3$	1;	$1/3$
$-3; 1/27$	-4;	$1/81$	-3;	$1/81$	-2;	$1/81$
$-2; 3/27$	-3;	$3/81$	-2;	$3/81$	-1;	$3/81$
$-1; 6/27$	-2;	$6/81$	-1;	$6/81$	0;	$6/81$
$0; 7/27$	-1;	$7/81$	0;	$7/81$	1;	$7/81$
$1; 6/27$	0;	$6/81$	1;	$6/81$	2;	$6/81$
$2; 3/27$	1;	$3/81$	2;	$3/81$	3;	$3/81$
$3; 1/27$	2;	$1/81$	3;	$1/81$	4;	$1/81$

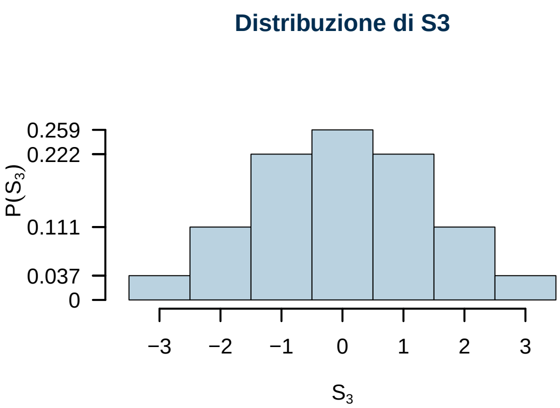
E ricaviamo la distribuzione di, S_4

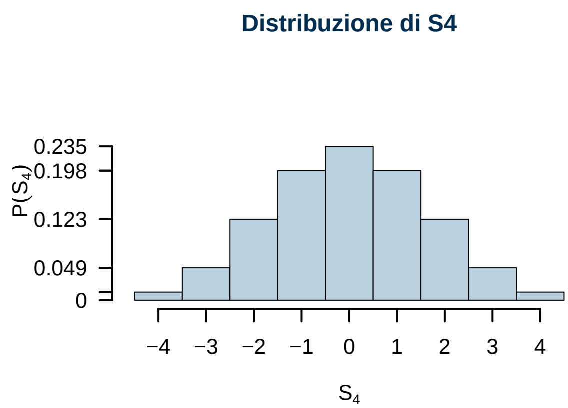
S_4	-4	-3	-2	-1	0	1	2	3	4
$P(S_4)$	$1/81$	$4/81$	$10/81$	$16/81$	$19/81$	$16/81$	$10/81$	$4/81$	$1/81$

Osserviamo i grafici per $n = 1, ..., 4$









Calcolare S_n se n è un numero elevato è difficoltoso. Il TLC ci viene incontro. Possiamo calcolare valore atteso e varianza della VC della VC che ha generato il sistema.

$$
\begin{aligned}
E(X) &= -1\frac{1}{3} + 0\frac{1}{3} + 1\frac{1}{3} = 0 \\
V(X) &= (-1)^2\frac{1}{3} + 0^2\frac{1}{3} + 1^2\frac{1}{3} - 0^2 = \frac{2}{3}
\end{aligned}
$$

e quindi in virtù del TLC per la somma

$$S_n \underset{a}{\sim} N(n \cdot 0, n \cdot 2/3)$$

Se per esempio $n = 50$

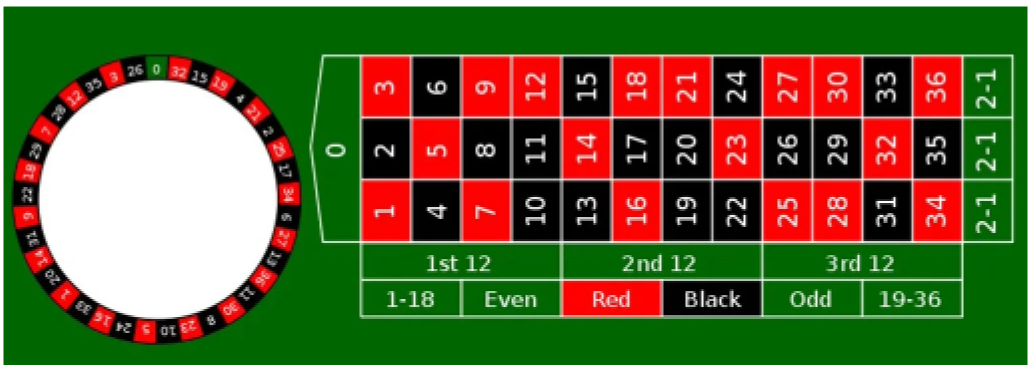
$$S_{50} \underset{a}{\sim} N(0, 50 \cdot 2/3)$$

e quindi se volessi calcolare la probabilità che $S_{50} < 3$ useremmo la distribuzione normale

$$
\begin{aligned}
P(S_n < 3) &= P\left(\frac{S_n - \mu}{V(S_n)} < \frac{3-0}{\sqrt{33.33}}\right) \\
&= P\left(Z < 0.52\right) \\
&= \Phi(0.52) \\
&= 0.6985
\end{aligned}
$$

8.3.2 Roulette

Il gioco. Una giocata dalla roulette equivale ad estrarre da un'urna che contiene 37 bussolotti numerati da 0 a 36.



Si può puntare su diverse combinazioni: pari o dispari, rosso o nero, da 1 a 18 o da 19 a 36,... e altre combinazioni. Se puntiamo 1€, per esempio, su Rosso la vincita/perdita sarà:

$$
\begin{cases} +1€, & \text{se esce Rosso} \\ -1€, & \text{se non esce Rosso} \end{cases}
$$

Giocheremo, sempre puntando un euro alla volta, per n volte. La VC R_i che descrive l'evento Rosso o non Rosso, nella giocata i, è una Bernoulli di parametro

$$\pi = \frac{18}{37} = 0.4865$$

Sia la VC X_i che descrive la vincita/perdita

$$X_i = -1 + 2R_i$$

Se $R_i = 1$ allora $X_i = -1 + 2 \times 1 = +1$, mentre se $R_i = 0$ allora $X_i = -1 + 2 \times 0 = -1$ La VC $R_i \sim \text{Ber}(\pi = 0.4865)$ e quindi

$$
\begin{aligned}
E(R_i) &= 0.4865 \\
V(R_i) &= 0.4865 \times (1 - 0.4865)
\end{aligned}
$$

E quindi

$$E(X_i) = -1 + 2 \times 0.4865$$

$$\begin{aligned} &= -0.02703 \\ V(X_i) &= 2^2 \times 0.4865 \times (1 - 0.4865) \\ &= 0.9993 \end{aligned}$$

Le X_i sono tutte tra di loro Indipendenti e Identicamente Distribuite (IID). Se quindi giochiamo n volte la VC che conta il numero di euro vinti/persi è

$$S_n = X_1 + ... + X_n$$

Riscrivendo in termini di R_i

$$\begin{aligned} S_n &= -1 + 2R_1 - 1 + 2R_2 + ... + (-1) + 2R_n \\ &= -n + 2(R_1 + ... + R_n) \\ &= -n + 2R \end{aligned}$$

$R = R_1 + ... + R_n \sim \mathrm{Bin}(n; \pi = 0.4865)$ Se per esempio gioco due ($n = 2$) volte

$$S_2 = X_1 + X_2$$

il supporto di S_2 è l'insieme $\{-2, 0, +2\}$.

$$\begin{aligned} P(S_2 = -2) &= P(R_1 + R_2 = 0) = \binom{2}{0} 0.4865^0 (1 - 0.4865)^2 = 0.2637 \\ P(S_2 = 0) &= P(R_1 + R_2 = 1) = \binom{2}{1} 0.4865^1 (1 - 0.4865)^1 = 0.4996 \\ P(S_2 = +2) &= P(R_1 + R_2 = 2) = \binom{2}{0} 0.4865^2 (1 - 0.4865)^0 = 0.2367 \end{aligned}$$

Se per esempio gioco tre ($n = 3$) volte

$$S_2 = X_1 + X_2 + X_3 = S_2 + X_3$$

il supporto di S_3 è l'insieme $\{-3, -1, +1, +3\}$.

$$\begin{aligned} P(S_3 = -3) &= P(R_1 + R_2 + R_3 = 0) = \binom{3}{0} 0.4865^0 (1 - 0.4865)^3 = 0.1354 \\ P(S_3 = -1) &= P(R_1 + R_2 + R_3 = 1) = \binom{3}{1} 0.4865^1 (1 - 0.4865)^2 = 0.3849 \\ P(S_3 = +1) &= P(R_1 + R_2 + R_3 = 2) = \binom{3}{2} 0.4865^2 (1 - 0.4865)^1 = 0.3646 \\ P(S_3 = +3) &= P(R_1 + R_2 + R_3 = 3) = \binom{3}{3} 0.4865^3 (1 - 0.4865)^0 = 0.1151 \end{aligned}$$

Se per esempio gioco tre ($n = 10$) volte

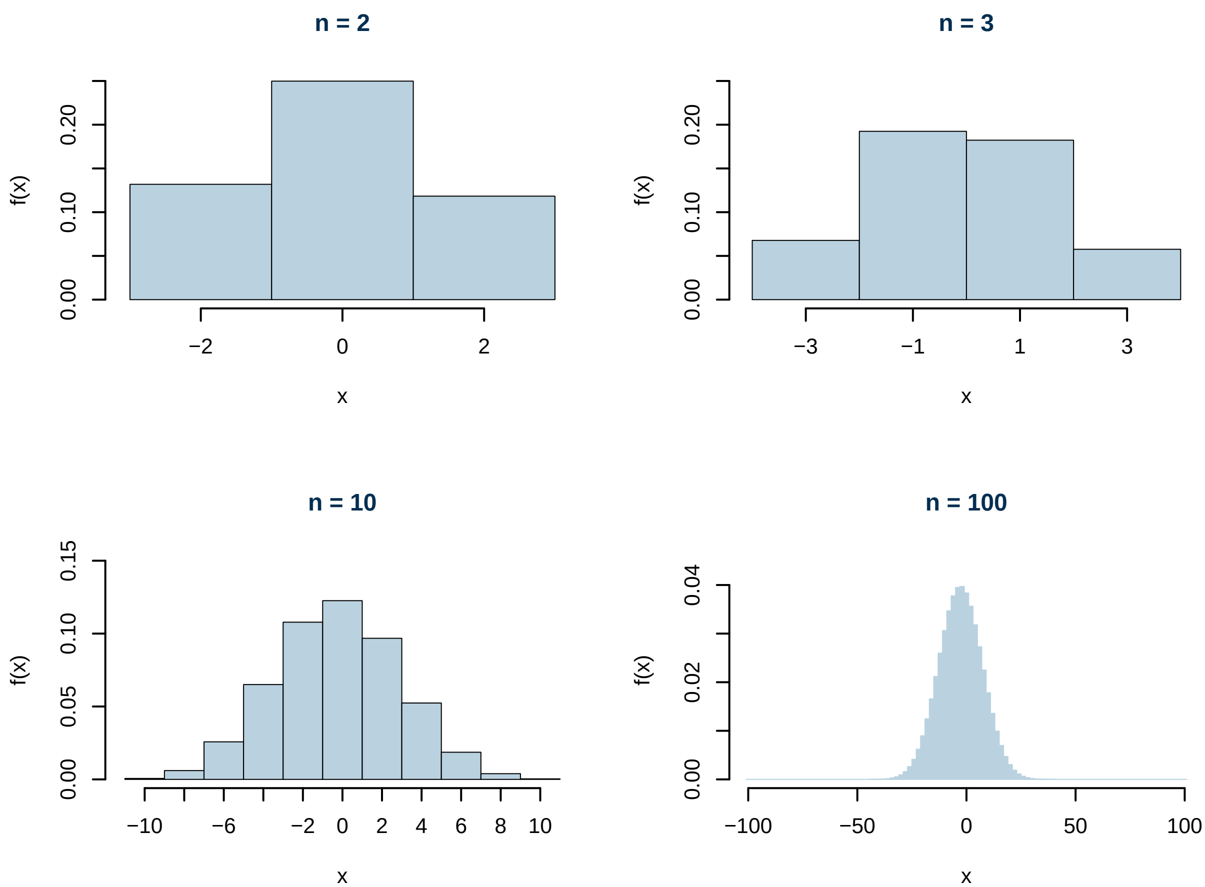
$$S_{10} = S_9 + X_{10}$$

Se per esempio gioco tre ($n = 100$) volte

$$S_{100} = S_{99} + X_{100}$$

La probabilità di non perdere è data da.

$$\begin{aligned}
P(S_{100} > 0) &= P(-100 + 2R > 0) \\
&= P(R > 100/2) \\
&= P(R > 50) \\
&= P(R = 51) + P(R = 52) + ... + P(R = 100) \\
&= \binom{100}{51} 0.4865^{51}(1 - 0.4865)^{49} + \binom{100}{52} 0.4865^{52}(1 - 0.4865)^{48} + ... \\
&+ \binom{100}{100} 0.4865^{100}(1 - 0.4865)^{0} \\
&= 0.3553
\end{aligned}$$



Approssimazione normale

$$\begin{aligned}
\mu &= E(X_i) \\
&= -1 + 2 \times 0.4865 \\
&= -0.02703 \\
\sigma^2 &= V(X_i) \\
&= 2^2 \times 0.4865 \times (1 - 0.4865) \\
&= 0.9993
\end{aligned}$$

posto

$$S_n = X_1 + ... + X_n$$

e dunque

$$\begin{aligned}
E(S_n) &= E(X_1 + ... + X_n) \\
&= E(X_1) + ... + E(X_n) \\
&= \mu + ... + \mu \\
&= n\mu \\
&= -2.7027 \\
V(S_n) &= V(X_1 + ... + X_n) \\
&= V(X_1) + ... + V(X_n) \\
&= \sigma^2 + ... + \sigma^2 \\
&= n\sigma^2 \\
&= n \times 0.9993
\end{aligned}$$

Per $n = 100$

$$\begin{aligned}
E(S_n) &= 100 \times (-0.02703) \\
&= -2.7027 \\
V(S_n) &= 100 \times 0.9993 \\
&= 99.927
\end{aligned}$$

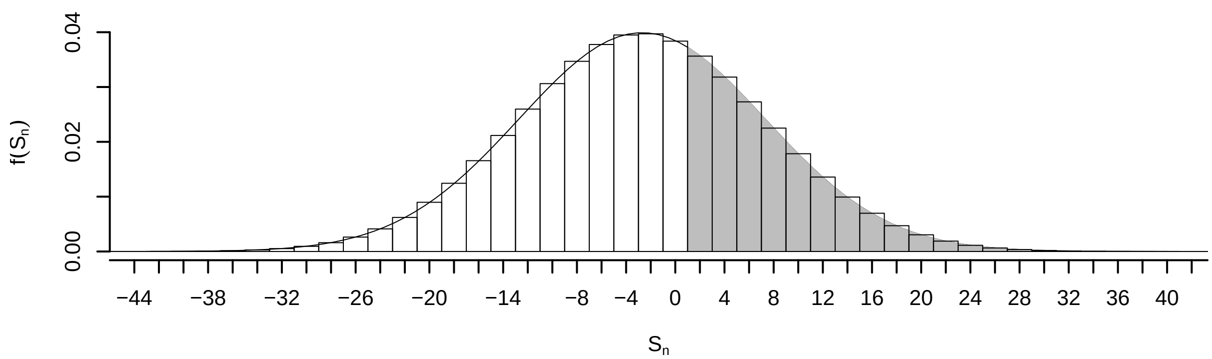
In virtù del teorema del limite centrale

$$S_{100} \quad \underset{a}{\sim} N(100 \times (-0.02703), 100 \times 0.9993)$$

$$\underset{a}{\sim} N(-2.7027, 99.927)$$

E quindi

$$\begin{aligned}
P(S_n > 0) &= P\left(\frac{S_n - E(S_n)}{\sqrt{V(S_n)}} > \frac{0-(-2.7027)}{\sqrt{99.93}}\right)\\
&= P\left(Z > 0.27\right)\\
&= 1 - P(Z < 0.27)\\
&= 1 - \Phi(0.27)\\
&= 0.3936
\end{aligned}$$



Nota

I due valori di probabilità esatta 0.3553 e quello approssimato 0.3934 sono diversi nella seconda cifra decimale. Questo è dovuto al fatto che la normale calcola anche parte dell'istogramma della binomiale in zero. Per ovviare e migliorare l'approssimazione basta spostare sulla fine del rettangolo dello zero il calcolo della normale. Senza entrare nel dettaglio si ricava che il rettangolo finisce per $S_n = 1$ e dunque

$$\begin{aligned}
P(S_n > 1) &= P\left(\frac{S_n - E(S_n)}{\sqrt{V(S_n)}} > \frac{1-(-2.7027)}{\sqrt{99.93}}\right)\\
&= P\left(Z > 0.37\right)\\
&= 1 - P(Z < 0.37)\\
&= 1 - \Phi(0.37)\\
&= 0.3557
\end{aligned}$$

Statistiche campionarie

9

9.1 Risultati preliminari

Consideriamo n variabili casuali, $X_1, ..., X_n$, IID, con valore atteso e varianza, rispettivamente, μ e σ^2. Sono dati che stiamo per osservare e quindi sono casuali.

Definizione 9.1.1. Una **statistica campionaria**, S, è una funzione dei dati $X_1, ..., X_n$

$$S(X_1, ..., X_n) = s \in \mathbb{R}$$

Come funzione di VC $S = S(X_1, ..., X_n)$ è una VC. Ad esempio la *media* dei dati è una statistica, il *25-esimo percentile* è una statistica, la *mediana* dei dati, la *varianza*, ecc. Ci possiamo porre alcune domande, per esmepio:

- come si distribuisce una media campionaria di valori casuali?

$$\hat{\mu} = \bar{X} \underset{a}{\sim} N\left(\mu, \frac{\sigma^2}{n}\right)$$

  dove, $E(\hat{\mu}) = \mu$, $V(\hat{\mu}) = \frac{\sigma^2}{n}$
- Come si distribuisce la varianza campionaria di valori casuali?

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \hat{\mu})^2 \sim ?$$

9.2 La distribuzione Chi-quadro χ^2

Definizione 9.2.1. Siano $Z_1, ..., Z_n$, n VC, IID, $Z_i \sim N(0, 1)$, posto,

$$Y = Z_1^2 + ... + Z_n^2, \qquad \text{allora} \qquad Y \sim \chi_n^2$$

La distribuzione della somma del quadrato di n normali standard è distribuita come un chi-quadro con n *gradi di libertà*

La VC ha come supporto tutta la retta reale positiva:

$$S_Y = \{y > 0\} = \mathbb{R}^+$$

Lo spazio dei parametri non ha interesse statistico

$$n \in \mathbb{Z}^+$$

Funzione di probabilista o densità:

$$f(x) = \frac{x^{\frac{n}{2}-1}e^{-\frac{x}{2}}}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}$$

dove

$$\Gamma(t) = \int_0^{+\infty} x^{t-1} e^{-x} dx$$

- per $n = 1$ ha una forma iperbolica;
- per $n > 2$ e a forma campanulare con un'asimmetria positiva (coda lunga a dx);
- in virtù del TLC se n diverge allora $Y \tilde{a} N(n, 2n)$.

$$E(Y) = n, \qquad V(Y) = 2n$$

In figura 9.1 e 9.1 la forma della densità al variare di n.

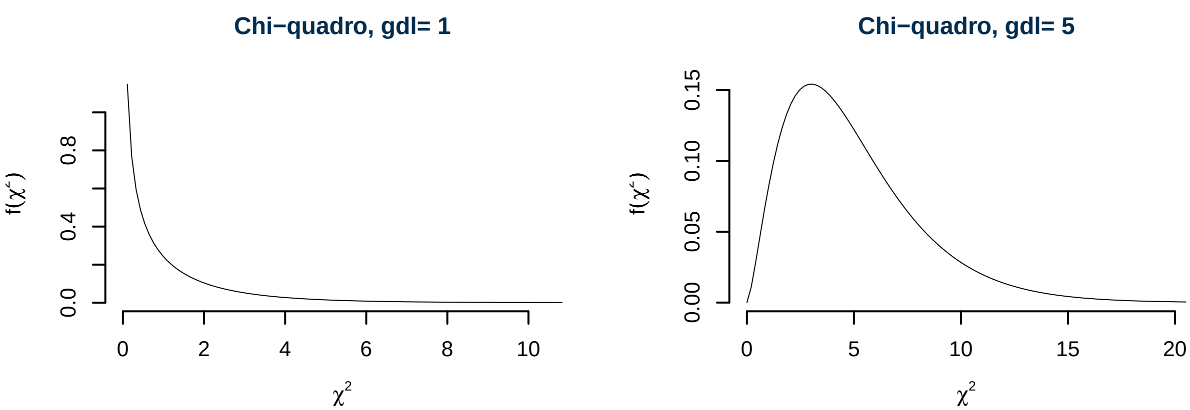


Figura 9.1: La densità della VC Chi-quadro per diversi valori di n

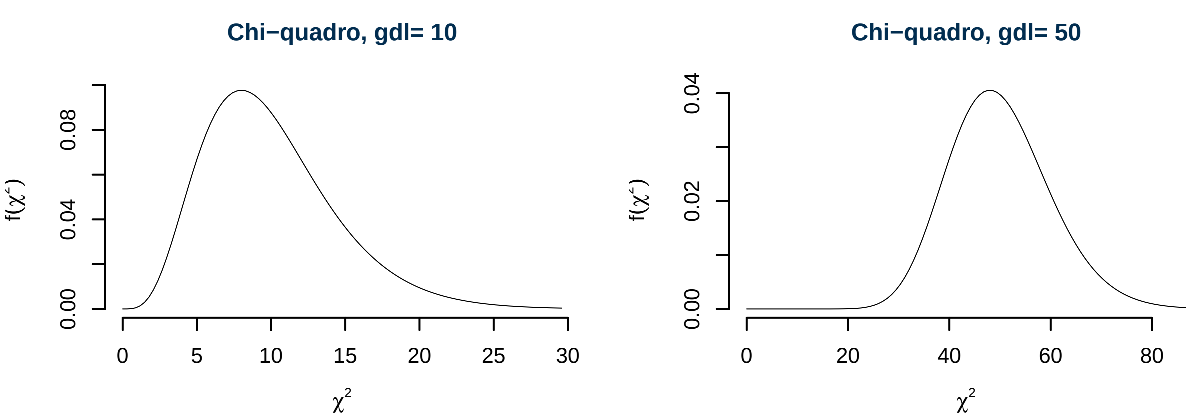


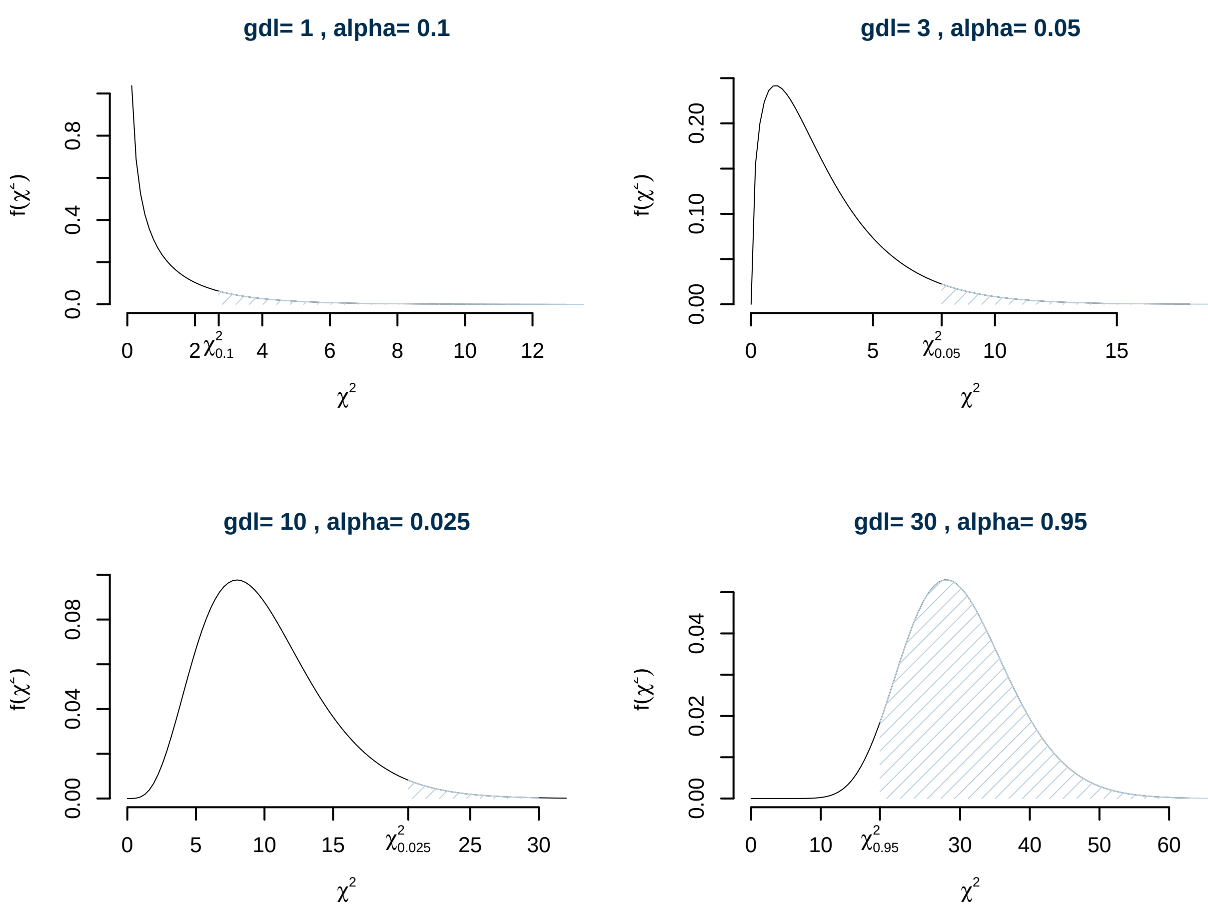
Figura 9.2: La densità della VC Chi-quadro per diversi valori di n

9.2.1 Le tavole del χ^2

Non c'è una sola distribuzione χ^2 ma tante quante sono i possibili gradi di libertà. Per comodità editoriale vengono mostrati solo alcuni valori delle code, per alcuni gradi di libertà. Offrono il percentile della χ^2 per diversi gradi di libertà e diversi valori di α, ovvero

$$P(\chi^2_n > \chi^2_{n;\alpha}) = \alpha$$

per alcuni valori di n e di α. Le tavole si presentano in forma tabellare dove ogni riga è indicizzata dal grado di libertà e ogni colonna dal valore di probabilità α.



Quindi per esempio se sono interessato a sapere quale valore del χ^2_3 lascia alla sua destra lo 0.05 dell'area dovrò cercare sulla terza riga in corrispondenza della colonna 0.05 e quindi

$$\chi^2_{3;0.05} = 7.8147$$

Tabella 9.1: Prime 8 righe delle tavole del χ^2 (1/2)

GdL	$\alpha = 0.99950$	$\alpha = 0.99900$	$\alpha = 0.99500$	$\alpha = 0.99000$	$\alpha = 0.97500$	$\alpha = 0.95000$	$\alpha = 0.90000$
1	0.00000	0.00000	0.00004	0.00016	0.00098	0.00393	0.01579
2	0.00100	0.00200	0.01003	0.02010	0.05064	0.10259	0.21072
3	0.01528	0.02430	0.07172	0.11483	0.21580	0.35185	0.58437
4	0.06392	0.09080	0.20699	0.29711	0.48442	0.71072	1.06362
5	0.15814	0.21021	0.41174	0.55430	0.83121	1.14548	1.61031
6	0.29941	0.38107	0.67573	0.87209	1.23734	1.63538	2.20413
7	0.48487	0.59849	0.98926	1.23904	1.68987	2.16735	2.83311
8	0.71038	0.85710	1.34441	1.64650	2.17973	2.73264	3.48954

Tabella 9.2: Prime 8 righe delle tavole χ^2 (2/2)

GdL	$\alpha = 0.10000$	$\alpha = 0.05000$	$\alpha = 0.02500$	$\alpha = 0.01000$	$\alpha = 0.00500$	$\alpha = 0.00100$	$\alpha = 0.00050$
1	2.70554	3.84146	5.02389	6.63490	7.87944	10.82757	12.11567
2	4.60517	5.99146	7.37776	9.21034	10.59663	13.81551	15.20180
3	6.25139	7.81473	9.34840	11.34487	12.83816	16.26624	17.73000
4	7.77944	9.48773	11.14329	13.27670	14.86026	18.46683	19.99735
5	9.23636	11.07050	12.83250	15.08627	16.74960	20.51501	22.10533
6	10.64464	12.59159	14.44938	16.81189	18.54758	22.45774	24.10280
7	12.01704	14.06714	16.01276	18.47531	20.27774	24.32189	26.01777
8	13.36157	15.50731	17.53455	20.09024	21.95495	26.12448	27.86805

9.3 La distribuzione t-di Student

Definizione 9.3.1. Siano $Z \sim N(0,1)$ e $Y \sim \chi^2_n$, Z e Y indipendenti, posto,

$$T = \frac{Z}{\sqrt{Y/n}} \qquad \text{allora} \qquad T \sim t_n$$

Il rapporto tra una normale standard e un la radice di un chi-quadro diviso per i suoi gradi di libertà è distribuito come una t-Student con n *gradi di libertà*

La VC ha come supporto tutta la retta reale:

$$S_T = \mathbb{R}$$

Lo spazio dei parametri non ha interesse statistico

$$n \in \mathbb{Z}^+$$

Funzione di probabilista o densità.

- è a forma campanulare
- è simmetrica rispetto a zero
- all'aumentare di n le code si abbassano
- Se $n \to \infty$, allora $t_n \to N(0,1)$

$$E(Y) = 0, \qquad V(Y) = \frac{n}{n-2}$$

In figura 9.3 e 9.4 il confronto tra la t-di Student e la normale standard, per diversi valori di n.

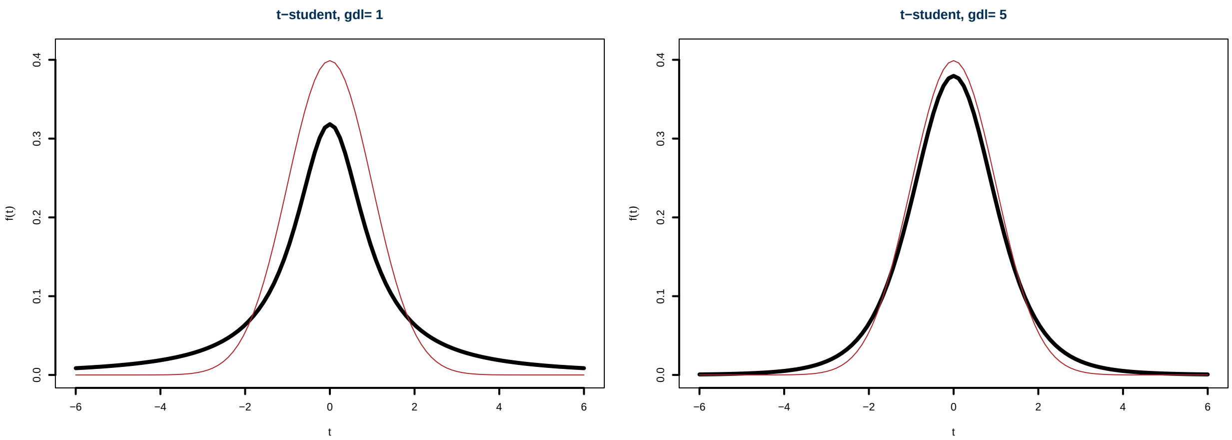


Figura 9.3: confronto tra la t-di student e la normale standard, per diversi valori di n

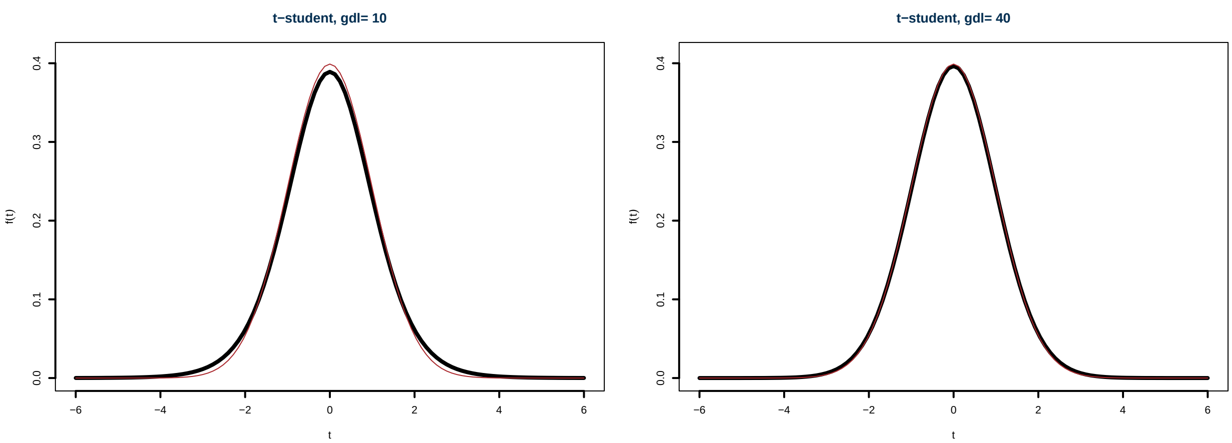


Figura 9.4: confronto tra la t-di student e la normale standard, per diversi valori di n

9.3.1 Le tavole della t

Non c'è una sola distribuzione t ma tante quante sono i possibili gradi di libertà. Per comodità editoriale vengono mostrati solo alcuni valori delle code, per alcuni gradi di libertà. Sulle tavole leggiamo:

$$P(T > t_{n;\alpha}) = \alpha$$

per alcuni valori di n e di α. Le tavole si presentano in forma tabellare dove ogni riga è indicizzata dal grado di libertà e ogni colonna dal valore di probabilità α. Per conoscere quale valore $t_{6;0.025}$ della t_6 con 6 gradi di libertà, tale che

$$P(T > t_{6;0.025}) = 0.025$$

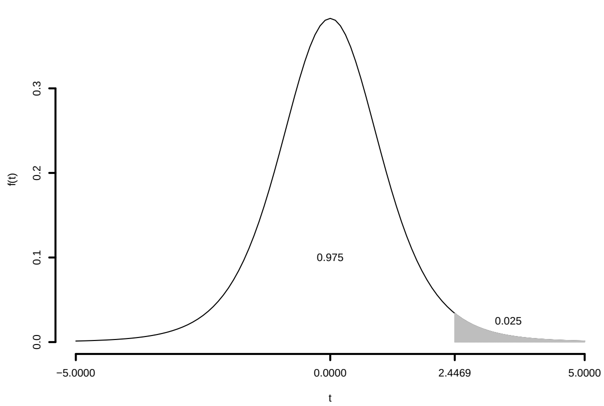
Tabella 9.3: $\alpha = 0.025$, con 6 gradi di libertà

GdL	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.025$	$\alpha = 0.01$	$\alpha = 0.005$	$\alpha = 0.001$	$\alpha = 0.0005$
1	3.0777	6.3138	12.7062	31.8205	63.6567	318.3088	636.6192
2	1.8856	2.9200	4.3027	6.9646	9.9248	22.3271	31.5991
3	1.6377	2.3534	3.1824	4.5407	5.8409	10.2145	12.9240
4	1.5332	2.1318	2.7764	3.7469	4.6041	7.1732	8.6103
5	1.4759	2.0150	2.5706	3.3649	4.0321	5.8934	6.8688
6	1.4398	1.9432	2.4469	3.1427	3.7074	5.2076	5.9588
7	1.4149	1.8946	2.3646	2.9980	3.4995	4.7853	5.4079
8	1.3968	1.8595	2.3060	2.8965	3.3554	4.5008	5.0413
9	1.3830	1.8331	2.2622	2.8214	3.2498	4.2968	4.7809
10	1.3722	1.8125	2.2281	2.7638	3.1693	4.1437	4.5869
...	...	...	...	...	...	...	...

Tabella 9.4: $\alpha = 0.005$, con 15 gradi di libertà

GdL	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.025$	$\alpha = 0.01$	$\alpha = 0.005$	$\alpha = 0.001$	$\alpha = 0.0005$
...	...	...	...	...	...	...	...
13	1.3502	1.7709	2.1604	2.6503	3.0123	3.8520	4.2208
14	1.3450	1.7613	2.1448	2.6245	2.9768	3.7874	4.1405
15	1.3406	1.7531	2.1314	2.6025	2.9467	3.7328	4.0728
16	1.3368	1.7459	2.1199	2.5835	2.9208	3.6862	4.0150
17	1.3334	1.7396	2.1098	2.5669	2.8982	3.6458	3.9651
...	...	...	...	...	...	...	...

Dunque $t_{6;0.025} = 2.44691$



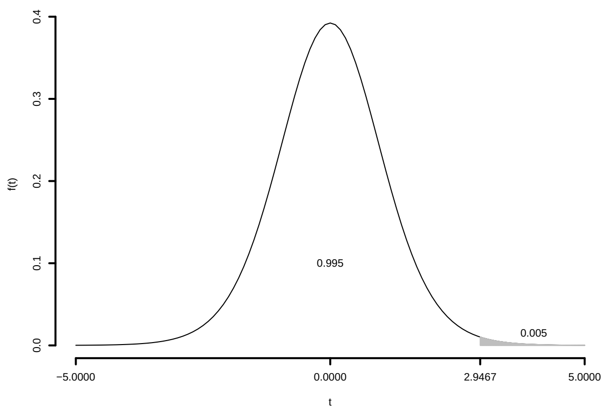
Per conoscere quale valore $t_{15;0.005}$ della t_{15} con 15 gradi di libertà, tale che

$$P(T > t_{15;0.005}) = 0.005$$

Dunque $t_{15;0.005} = 2.94671$

Tabella 9.5: $\alpha = 0.0005$, con 49 gradi di libertà

GdL	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.025$	$\alpha = 0.01$	$\alpha = 0.005$	$\alpha = 0.001$	$\alpha = 0.0005$
...	...	...	...	...	...	...	...
46	1.3002	1.6787	2.0129	2.4102	2.6870	3.2771	3.5150
47	1.2998	1.6779	2.0117	2.4083	2.6846	3.2729	3.5099
48	1.2994	1.6772	2.0106	2.4066	2.6822	3.2689	3.5051
49	1.2991	1.6766	2.0096	2.4049	2.6800	3.2651	3.5004
50	1.2987	1.6759	2.0086	2.4033	2.6778	3.2614	3.4960
...	...	...	...	...	...	...	...



Per conoscere quale valore $t_{49;0.0005}$ della t_{49} con 49 gradi di libertà, tale che

$$P(T > t_{49;0.001}) = 0.001$$

Dunque $t_{49;0.001} = 3.2651$

Nota

L'ultima riga della t, per un numero infinito di GdL, coincide con la tabella aggiuntiva della Z. Infatti

GdL	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.025$	$\alpha = 0.01$	$\alpha = 0.005$	$\alpha = 0.001$	$\alpha = 0.0005$
80	1.2922	1.6641	1.9901	2.3739	2.6387	3.1953	3.4163
90	1.2910	1.6620	1.9867	2.3685	2.6316	3.1833	3.4019
100	1.2901	1.6602	1.9840	2.3642	2.6259	3.1737	3.3905
∞	1.2816	1.6449	1.9600	2.3263	2.5758	3.0902	3.2905

La tabella dei percentili della normale è

α	0.1	0.05	0.025	0.01	0.005	0.001	0.0005
z_α	1.2816	1.6449	1.9600	2.3263	2.5758	3.0902	3.2905
$\Phi(z_\alpha) = 1-\alpha$	0.9	0.95	0.975	0.99	0.995	0.999	0.9995

9.4 La distribuzione di $\hat{\sigma}^2$

Siano $X_1, ..., X_n$, n VC IID, replicazioni della stessa $X \sim N(\mu, \sigma^2)$. La si definisce la varianza campionaria

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \hat{\mu})^2, \qquad \hat{\mu} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

allora

$$\hat{\sigma}^2 \sim \frac{\sigma^2}{n}\chi^2_{n-1}.$$

Osserviamo:

$$\begin{aligned} E(\hat{\sigma}^2) &= \frac{\sigma^2}{n}E(\chi^2_{n-1}) \\ &= \frac{n-1}{n}\sigma^2 \end{aligned}$$

9.5 La distribuzione della statistica standardizzata

Siano $X_1, ..., X_n$, n VC, IID, $X_i \sim N(\mu, \sigma^2)$. Si definisce *standardizzazione* di $\bar{X}$ dati rispetto a μ:

$$T = \frac{\hat{\mu} - \mu}{\hat{\sigma}/\sqrt{n-1}}$$

Allora

$$T \sim t_{n-1}.$$

Parte III
Inferenza

10 Inferenza: concetti introduttivi

Dalla Treccani si definisce l'atto di inferire come:

> Trarre, partendo da una determinata premessa o dalla constatazione di un fatto, una conseguenza, un giudizio, una conclusione.

Inferire è un atto fondamentale non solo del pensiero umano, ma anche della vita stessa. Gran parte delle specie evolute possiede un istinto innato per trarre conclusioni a partire dall'osservazione dell'ambiente circostante. Che si tratti di un animale che associa il rumore di un predatore al pericolo o di una pianta che risponde agli stimoli luminosi per orientare la crescita, l'inferenza è una strategia essenziale per adattarsi e sopravvivere.

L'uomo, tuttavia, si distingue per un aspetto cruciale: la capacità di prendere coscienza di questo processo, di formalizzarlo e di costruirci sopra un linguaggio. La riflessione sull'inferenza, il tentativo di comprendere e replicare consapevolmente ciò che molte specie fanno istintivamente, ha dato origine a discipline come la logica, la matematica e, infine, la statistica. Studiando l'inferenza, l'uomo è riuscito a trasformare un'abilità naturale in un processo razionale, formale e comunicabile, che non solo migliora la comprensione del mondo, ma rende possibile la condivisione e la riproduzione di quel sapere.

Esistono diverse modalità di inferenza, che riflettono modi distinti di trarre conclusioni: **deduttiva**, **induttiva diretta** e **induttiva inversa**.

L'inferenza deduttiva, tipica della logica formale, parte da premesse certe per arrivare a conclusioni necessarie. È il tipo di ragionamento che governa la matematica: dati alcuni assunti iniziali, le conclusioni sono logicamente inevitabili e già contenute nelle premesse. Questo tipo di inferenza è deterministico e privo di incertezze.

L'inferenza induttiva diretta, invece, opera nel campo della probabilità. Partendo da un modello noto o da una distribuzione che descrive un fenomeno, permette di calcolare la probabilità di osservare determinati esiti. È il caso di un'urna dalla composizione nota: possiamo determinare con precisione la probabilità che un'estrazione casuale produca un certo risultato. Qui il ragionamento è incerto, ma il modello è dato a priori e costituisce la base delle conclusioni.

L'inferenza induttiva inversa, oggi conosciuta come **inferenza statistica**, ribalta questa prospettiva. Non ci si chiede quali osservazioni aspettarsi da un modello noto, ma quale modello, quali parametri o quali proprietà della popolazione possano spiegare i dati osservati. È il problema dell'urna dalla composizione incognita: data una serie di estrazioni, come possiamo inferire la composizione complessiva? Questo tipo di inferenza è intrinsecamente probabilistico e dipende fortemente dal contesto e dalle ipotesi di partenza.

Ciò che rende l'inferenza statistica così straordinaria è proprio la sua capacità di trasformare osservazioni parziali, influenzate dal caso, in una comprensione più ampia e generalizzabile. Ma ciò che la rende possibile è il linguaggio formale che l'uomo ha costruito per descriverla. Questo linguaggio non solo rende espliciti i passaggi impliciti del ragionamento, ma permette di comunicare e replicare i risultati, superando i limiti del contesto individuale.

In definitiva, l'inferenza statistica è l'espressione più raffinata di una capacità condivisa con molte specie, ma che l'uomo ha elevato a strumento consapevole e rigoroso. Essa rappresenta un ponte tra l'istinto naturale e la razionalità formale, dimostrando come un processo profondamente radicato nella biologia possa essere trasformato in un metodo universale per comprendere il mondo.

Per comprendere la differenza tra inferenza induttiva diretta e inversa, consideriamo il classico esempio di un'urna contenente bussolotti di due colori: rosso e blu.

10.1 Inferenza da popolazioni finite

Le **popolazioni finite** rappresentano il contesto più vicino all'intuizione comune. Si tratta di insiemi chiusi e completamente enumerabili di unità, come le persone registrate in un censimento o gli studenti iscritti a un corso. In questi casi, ogni elemento della popolazione può essere associato a una lista, e il campione rappresenta una frazione precisa della popolazione. Questo contesto è tipico delle **statistiche ufficiali**, come quelle prodotte dall'**ISTAT**, dall'**Eurostat** e dall'**OCSE**.

L'inferenza da popolazioni finite è quindi apparentemente intuitiva, ma richiede un impianto operativo articolato: occorre disporre di un **registro completo della popolazione**, progettare un **disegno di campionamento** adatto (spesso stratificato o a più stadi), costruire e testare **strumenti di rilevazione standardizzati** (come questionari), e garantire **qualità e controllo dei dati raccolti**. Si tratta di operazioni costose e complesse, che solo istituzioni pubbliche di grande dimensione e con mandato istituzionale riescono a realizzare su scala nazionale o internazionale.

Ecco alcuni esempi concreti di indagini campionarie su popolazioni finite:

- **L'Indagine sulle forze di lavoro (ISTAT)** ha l'obiettivo di stimare il tasso di occupazione, disoccupazione e inattività della popolazione residente in Italia. La popolazione di riferimento è costituita da tutte le persone residenti in famiglie, escluse le collettività. L'indagine è continua e si basa su un campione a rotazione di circa 77.000 famiglie ogni trimestre, selezionate casualmente. I questionari raccolgono informazioni su attività lavorativa, ricerca di lavoro, orari, contratti e condizioni lavorative.

- **L'Indagine sui consumi delle famiglie (ISTAT)** rileva le spese sostenute per beni e servizi, allo scopo di descrivere i comportamenti di consumo e aggiornare il paniere per l'indice dei prezzi. La popolazione è costituita dalle famiglie residenti in Italia, e il campione include oltre 30.000 famiglie distribuite lungo l'anno. Le famiglie selezionate devono compilare un diario giornaliero delle spese e rispondere a interviste dettagliate, il che rende l'indagine particolarmente impegnativa.

- **L'indagine EU-SILC (coordinata da Eurostat, realizzata in Italia da ISTAT)** fornisce informazioni su redditi, condizioni abitative, povertà e disuguaglianza. La popolazione è costituita dalle famiglie residenti nei paesi europei. In Italia, il campione supera le 20.000 famiglie, con interviste condotte annualmente e in parte replicate su base panel. I questionari sono armonizzati a livello europeo per consentire il confronto tra paesi.

- **L'indagine PISA (OCSE)** valuta le competenze degli studenti quindicenni in lettura, matematica e scienze, con l'obiettivo di confrontare i sistemi educativi dei paesi partecipanti. La popolazione di riferimento è costituita dagli studenti iscritti al secondo ciclo dell'istruzione secondaria, indipendentemente dall'anno frequentato. In Italia, il campione coinvolge circa 11.000 studenti ogni tre anni, selezionati da un insieme rappresentativo di scuole. I questionari comprendono sia prove cognitive standardizzate sia sezioni dedicate al contesto scolastico, familiare e motivazionale, permettendo analisi multilivello sul rendimento e le disuguaglianze educative.

- **L'indagine PIAAC (OCSE)** valuta le competenze fondamentali degli adulti tra i 16 e i 65 anni, in particolare la capacità di comprendere testi, usare strumenti numerici e risolvere problemi. In Italia, la popolazione campionata è estratta dai registri anagrafici e comprende circa 5.000 individui. Le interviste includono sia un modulo socio-demografico sia prove individuali computerizzate, somministrate in centri di test.

- **L'indagine TALIS (OCSE)** raccoglie informazioni sulle condizioni di lavoro degli insegnanti e dei dirigenti scolastici, esplorando aspetti come la formazione, la soddisfazione professionale, le pratiche didattiche e il clima scolastico. La popolazione di riferimento è costituita dagli insegnanti delle scuole secondarie inferiori. Anche qui il campione è probabilistico e i questionari, somministrati in modo standardizzato, permettono confronti tra paesi.

Queste analisi si basano su **campioni estratti da popolazioni note e completamente enumerate**. Il disegno campionario, il calcolo dei pesi e l'analisi dei dati tengono esplicitamente conto della struttura finita della popolazione e della complessità del piano di campionamento. La qualità dell'inferenza dipende in larga parte dalla precisione con cui questi aspetti sono progettati e implementati.

10.2 Inferenza da popolazioni infinite

Nel lavoro statistico reale, capita spesso di trovarsi in contesti in cui la popolazione non è elencabile né finita, o non è nemmeno definibile in modo operativo. In questi casi si adotta un approccio diverso: si assume che i dati osservati derivino da una **popolazione concettualmente infinita**, e si modellano come **realizzazioni di una variabile aleatoria**. Si tratta di una rappresentazione idealizzata, ma estremamente potente e spesso necessaria per dare senso e struttura all'analisi dei dati.

Esempi classici includono:

- **Processi fisici e ambientali**, come il livello di un fiume, la temperatura media di una regione, o la concentrazione di un inquinante nell'aria. Non possiamo elencare tutti i valori possibili, né osservare tutti i momenti futuri: costruiamo un modello e lavoriamo con i dati come se fossero campioni da una variabile continua.
- **Processi industriali e produttivi**, in cui i pezzi prodotti in futuro non esistono ancora, ma si assume che seguano la stessa legge dei pezzi già osservati.
- **Eventi ripetibili**, come lanci di monete, click su un sito o richieste a un server: non ci interessa una popolazione finita, ma una regolarità stocastica nel lungo periodo.

- **Sperimentazioni controllate**, come i *clinical trials*, in cui si valuta l'efficacia di un farmaco o di una terapia. La popolazione di riferimento è concettualmente infinita (tutti i possibili pazienti), e si ipotizza che l'effetto del trattamento segua una distribuzione probabilistica.
- **Psicologia sperimentale e scienze comportamentali**, dove si testano ipotesi su preferenze, scelte, reazioni, sotto vincoli controllati. Anche in questi casi, si modellano le risposte come esiti stocastici, spesso su scala continua o discreta.
- **Economia sperimentale**, in cui si studiano decisioni individuali o interazioni strategiche (come giochi o aste) in laboratorio. I dati sono trattati come osservazioni da una popolazione astratta, con l'obiettivo di descrivere regolarità comportamentali generalizzabili.

Anche quando i dati provengono da una popolazione finita, può essere utile (o necessario) adottare il punto di vista **modellistico**, trattando i dati come esiti casuali di un esperimento teorico. È una scelta metodologica che consente di rispondere a domande su effetti, relazioni, rischi, scenari ipotetici.

10.3 Inferenza non parametrica e inferenza parametrica

Una volta accettato il quadro modellistico, si aprono due grandi strade a seconda del grado di struttura imposto al modello probabilistico:

- Nell'inferenza **non parametrica** (o *distribution-free*), si assume solo che le osservazioni siano **IID** (indipendenti e identicamente distribuite), ma non si specifica la forma della distribuzione. L'obiettivo è stimare quantità come media, mediana o varianza con il minimo numero di ipotesi. È un approccio flessibile, adatto a situazioni esplorative o con pochi dati.

- Nell'inferenza **parametrica**, si ipotizza che i dati seguano una certa **famiglia di distribuzioni** (ad esempio normale, binomiale, esponenziale…), a meno di pochi parametri ignoti. Questo approccio consente di ottenere inferenze più precise, stimare probabilità di eventi complessi e costruire modelli predittivi, a patto che l'ipotesi sul modello sia ragionevole.

Per esempio:

- Un analista può stimare la **mediana del tempo di percorrenza** su una tratta, senza assumere nulla sulla forma della distribuzione: in questo caso si opera in ambito non parametrico.
- Un gestore del traffico può usare un **modello di Poisson** per stimare la probabilità che un incrocio abbia più di 10 auto in coda a mezzogiorno.
- Un epidemiologo può usare un **modello binomiale** per stimare la probabilità che almeno 3 persone in un piccolo gruppo siano infette.

L'adozione di un modello più o meno strutturato comporta sempre un **compromesso tra generalità e precisione**: meno ipotesi permettono maggiore robustezza, ma richiedono strumenti più cauti; più ipotesi rendono le conclusioni più forti, ma più sensibili a eventuali deviazioni dalla realtà.

10.4 Sintesi dei contesti

Questa classificazione dei contesti di inferenza aiuta a chiarire il tipo di domande che lo statistico può affrontare. Mentre l'inferenza da popolazioni finite si concentra su contesti pratici e limitati, tipici delle statistiche ufficiali, l'inferenza da popolazioni infinite e da modelli probabilistici consente di affrontare problemi più astratti e complessi, con applicazioni che spaziano dalla scienza alla finanza, dall'industria alla ricerca sperimentale.

10.5 Dalle popolazioni ai modelli: la metafora dell'urna

Per fissare le idee sui diversi schemi inferenziali che incontreremo nel corso, possiamo fare ricorso alla metafora dell'urna, già utilizzata nello studio della probabilità. In ciascun caso, l'urna rappresenta un meccanismo stocastico, idealmente realizzabile o puramente concettuale, da cui si ottengono osservazioni mediante un campionamento probabilistico. La metafora non suggerisce un'applicazione, ma si concentra sul **meccanismo probabilistico** alla base di ciascun tipo di inferenza.

10.5.1 Campionamento da un'urna di dimensione nota, senza reinserimento (popolazioni finite)

Un'urna contiene $N = 100$ bussolotti, un numero ignoto R di rossi e $B = N - R$ bianchi. Vengono estratti $n = 10$ bussolotti **senza reinserimento**, e si osservano $s = 6$ bianchi.

Domanda inferenziale:
Quanti bussolotti bianchi sono contenuti nell'urna? Possiamo stimare B, o almeno approssimare un intervallo plausibile per la proporzione $\frac{B}{N}$?

10.5.2 Campionamento da un'urna a composizione ignota, con reinserimento (popolazioni infinite, modello binomiale)

L'urna ha una composizione fissa ma ignota: una proporzione π di bianchi e $1 - \pi$ di rossi. Si eseguono $n = 10$ estrazioni **con reinserimento**, e si osservano $s = 6$ bianchi.

Domanda inferenziale:
Qual è la stima di π, cioè della proporzione di bianchi nell'urna? Possiamo fornire un intervallo credibile per π, sulla base del campione osservato?

10.5.3 Campionamento da un'urna che genera numeri reali con legge ignota (popolazioni infinite, inferenza non parametrica)

L'urna genera numeri reali, secondo una distribuzione ignota. Si estraggono $n = 10$ valori:
$X = \{2.3,\ 1.8,\ 3.1,\ 2.6,\ 2.0,\ 2.5,\ 2.9,\ 1.7,\ 2.1,\ 2.4\}$
con media campionaria $\bar{x} = 2.34$ e varianza campionaria $s^2 = 0.19$.

Domande inferenziali:
Cosa possiamo concludere sul valore medio μ della distribuzione da cui provengono questi dati? E sulla variabilità σ^2?
Quali margini di incertezza sono associati a queste stime, pur non conoscendo la forma della distribuzione?

10.5.4 Campionamento da un'urna che genera conteggi secondo un modello di Poisson (popolazioni infinite, inferenza parametrica discreta)

L'urna genera conteggi secondo una legge di Poisson con parametro λ ignoto. Si osservano
$Y = \{3,\ 1,\ 2,\ 4,\ 2,\ 0,\ 1,\ 3,\ 2,\ 1\}$
con media campionaria $\bar{y} = 1.9$.

Domanda inferenziale:
Possiamo stimare il tasso medio λ? Quanto è plausibile che il valore vero di λ sia maggiore di 2? Quale margine di errore possiamo associare alla stima ottenuta?

10.5.5 Campionamento da un'urna che genera valori reali secondo una distribuzione normale

(popolazioni infinite, inferenza parametrica continua)

L'urna genera numeri reali secondo una distribuzione normale $\mathcal{N}(\mu, \sigma^2)$. Si osservano:
$\mathbf{x} = \{175.2,\ 179.3,\ 173.5,\ 176.1,\ 178.7,\ 174.0,\ 175.9,\ 177.4,\ 176.8,\ 178.1\}$,
con media campionaria $\bar{x} = 176.5$ e deviazione standard campionaria $s = 1.8$.

Domanda inferenziale:
Possiamo stimare μ e σ, e sulla base di tali stime calcolare la probabilità che un nuovo valore superi una certa soglia? Qual è la probabilità che la prossima estrazione sia superiore a 180?

Nel proseguire, ci concentreremo sull'inferenza da popolazioni infinite e da modelli probabilistici, che sono centrali per lo studio dell'economia e della scienza dei dati. Tuttavia, l'approccio sarà sempre lo stesso: estrarre il massimo di informazioni dai dati osservati, bilanciando rigore e comprensibilità.

10.6 Statistica Classica

In questo libro adotteremo il paradigma della **statistica classica**, un approccio che si fonda sull'idea che i dati osservati siano manifestazioni imperfette di una realtà ideale descritta da modelli probabilistici. Pur essendo un sostenitore della prospettiva bayesiana, mi concentrerò su questo approccio tradizionale per il suo valore didattico e per la sua centralità in molti contesti applicativi.

Alla base della statistica classica vi è un'assunzione fondamentale: i parametri di un modello, come μ, σ^2 o λ, **esistono** in un senso ideale. Questi parametri non sono osservabili, ma rappresentano la struttura profonda e immutabile della popolazione. Ogni osservazione x_i e ogni statistica campionaria, come $\bar{x}$ o s^2, non sono altro che **ombre** di questa verità ideale, un'emanazione del parametro inosservabile θ.

La distinzione tra ciò che è ideale e ciò che è osservabile richiama direttamente la metafora platonica della caverna. I parametri, rappresentati da lettere greche, abitano l'**iperuranio**, il regno delle verità perfette e immutabili. In questo regno, esiste una media ideale, una varianza ideale, una probabilità ideale. I dati osservati e le statistiche campionarie, invece, sono come le **ombre proiettate sulle pareti della caverna**: imperfette, corrotte dalla variabilità del campionamento, ma comunque legate a quella verità profonda che tentiamo di comprendere.

Ad esempio:

- Una singola osservazione x_i è un'ombra della struttura imposta dal parametro θ, ma la sua forma concreta è influenzata dalla variabilità intrinseca del processo probabilistico.
- La media campionaria $\bar{x}$ è un'approssimazione imperfetta della media ideale μ, che abita l'iperuranio.
- La frequenza relativa osservata m/n riflette solo parzialmente la probabilità ideale π in una distribuzione Bernoulli.

Questo approccio funziona perfettamente come un **gioco mentale rigoroso**, a condizione che le assunzioni del modello siano rispettate. In particolare:

1. Le forme distributive delle variabili casuali devono essere corrette. Se assumiamo che una variabile segua una distribuzione normale, tale ipotesi deve essere giustificata dai dati o dal contesto.
2. I legami di dipendenza o indipendenza tra le osservazioni devono essere realistici. Ad esempio, l'assunzione che i dati siano indipendenti e identicamente distribuiti (IID) è cruciale in molti modelli, ma deve essere verificata.

Queste assunzioni creano il legame tra il mondo ideale e quello osservato. Senza di esse, la connessione tra le ombre (x_i e le statistiche campionarie) e la verità (θ) si spezza, e le inferenze statistiche rischiano di essere prive di fondamento.

Ecco una revisione della chiosa, che sottolinea il carattere filosofico e potenzialmente contestabile di questa impostazione, in relazione alle radici empiriste della statistica:

L'idea che θ esista in un regno ideale, come verità perfetta e inosservabile, riflette una visione filosofica di derivazione platonica, che può sembrare in contrasto con le radici empiriste su cui si fonda la statistica moderna. La disciplina, infatti, nasce dalla necessità di descrivere e comprendere il mondo reale attraverso dati osservabili, più vicina all'approccio britannico della conoscenza basata sull'esperienza.

Questa apparente tensione tra la ricerca della verità ideale e l'attenzione ai dati concreti non è priva di contraddizioni. La statistica classica si sviluppa su una sottile linea di compromesso: accetta l'esistenza di parametri teorici per costruire modelli potenti e rigorosi, ma dipende integralmente dalle osservazioni empiriche per avvicinarsi a quelle verità astratte. È un modello mentale che funziona all'interno delle sue assunzioni, ma che rimane soggetto a critica e revisione, specialmente quando queste assunzioni vengono meno.

Nel nostro percorso, accetteremo questo approccio come base metodologica, riconoscendo che esso rappresenta un sistema coerente, anche se non l'unica via possibile per fare inferenza. L'idea di parametri perfetti che governano i dati può essere vista come un artificio concettuale, utile per l'analisi e la modellazione, ma non necessariamente una descrizione della realtà ultima. In questo senso, la statistica classica non è solo uno strumento matematico, ma anche un prodotto di una specifica eredità filosofica, che merita di essere compresa e, quando necessario, interrogata.

10.6.1 Un'altra via: la statistica bayesiana

Accanto alla statistica classica, esiste un altro approccio all'inferenza, fondato su una concezione diversa della probabilità: la **statistica bayesiana**. In questa impostazione, il parametro θ non è una costante fissa e ignota che governa i dati, ma una **variabile aleatoria** che rappresenta lo **stato di conoscenza** dell'osservatore. La probabilità non misura una frequenza nel lungo periodo, ma **esprime il grado di fiducia** o credenza razionale che si attribuisce a un certo valore di θ, alla luce delle informazioni disponibili.

La statistica bayesiana si fonda su un principio semplice e potente: aggiornare le proprie convinzioni quando si osservano nuovi dati. Questo aggiornamento avviene attraverso il **teorema di Bayes**, che combina l'**informazione a priori** (la distribuzione di probabilità assegnata a θ prima di osservare i dati) con l'**informazione contenuta nel campione** (attraverso la verosimiglianza) per ottenere una **distribuzione a posteriori** su θ, che rappresenta lo stato di conoscenza aggiornato.

Da questo punto di vista, l'inferenza non è più una stima di qualcosa che esiste "là fuori", ma un processo di revisione coerente delle proprie credenze alla luce dell'evidenza. Questo rende l'approccio bayesiano particolarmente adatto in contesti in cui:

- l'esperienza pregressa o la letteratura suggerisce una conoscenza parziale ma strutturata del fenomeno;
- l'incertezza ha un ruolo centrale nella decisione (come nella teoria delle decisioni o nella valutazione dei rischi);
- il numero di osservazioni è limitato, o l'inferenza richiede uno sforzo di integrazione tra fonti diverse.

L'impostazione bayesiana è concettualmente più vicina all'**empirismo epistemico** e ha trovato crescente applicazione grazie alla disponibilità di strumenti computazionali per approssimare le distribuzioni a posteriori (come i metodi Monte Carlo via Markov Chain). Tuttavia, essa richiede anche una maggiore responsabilità soggettiva da parte dell'analista, che deve esplicitare le proprie ipotesi a priori e accettare che l'inferenza dipenda, almeno in parte, da esse.

Nel presente corso non utilizzeremo formalmente il quadro bayesiano, ma ne terremo conto a livello concettuale. In alcuni esercizi verrà proposta l'interpretazione soggettiva della probabilità come alternativa alla frequenza relativa. Più avanti, sarà utile confrontare le risposte classiche e quelle bayesiane a problemi semplici, per cogliere il significato profondo delle differenze tra i due approcci.

Elementi di Teoria della Stima

11

Il problema della **stima** nasce quando si vuole assegnare un valore numerico a una quantità incognita, chiamata **parametro**, a partire da un insieme limitato di osservazioni. In statistica, questa operazione si fonda sull'idea che i dati provengano da un **meccanismo aleatorio** e che il parametro rappresenti una proprietà stabile della popolazione o del modello da cui i dati sono stati generati.

Nel nostro contesto, un parametro può essere, ad esempio, la media μ, la varianza σ^2, una proporzione π, un tasso λ, o più in generale un elemento θ di uno spazio dei parametri Θ. La stima si basa sull'uso di un **campione**, ossia un insieme di dati osservati, che contiene informazione parziale sulla popolazione o sul modello.

11.1 Campionamento

Un campione casuale è un certo numero di n di osservazioni prese a caso dalla popolazione $\mathscr{P}$. Nel quadro dell'inferenza statistica, il **campionamento** è il processo attraverso cui si seleziona un sottoinsieme di unità da una popolazione $\mathscr{P}$, al fine di trarne informazioni utili. Il tipo di inferenza che si può condurre dipende in larga parte dal modo in cui è stato ottenuto il campione.

11.1.1 Campioni casuali

La **casualità nella selezione** non è un espediente tecnico, ma un presupposto fondamentale per garantire che le conclusioni tratte dal campione siano generalizzabili alla popolazione. Solo un campione ottenuto in modo casuale consente di utilizzare il linguaggio probabilistico per quantificare l'incertezza delle stime.

In assenza di casualità, è difficile (o impossibile) distinguere ciò che è strutturale da ciò che è frutto di una selezione distorta. Le ipotesi matematiche, come l'indipendenza tra osservazioni o l'identica distribuzione, si fondano sull'idea che ogni unità della popolazione abbia avuto la stessa probabilità di essere inclusa, o comunque una probabilità nota e controllata.

11.1.2 Campionamento da popolazioni finite e note

Quando la popolazione è **finita e nota**, il disegno di campionamento può essere esplicitamente definito e controllato. Questo è il caso tipico delle **statistiche ufficiali**, dove la popolazione di riferimento è costituita, ad esempio, da tutti i residenti, le famiglie, le imprese, gli studenti iscritti, ecc., elencati in una lista completa.

In tali contesti, la selezione del campione non avviene semplicemente in modo casuale puro, ma attraverso un **piano di campionamento** (o **disegno**) strutturato, il cui obiettivo è garantire:

- **buona rappresentatività**, anche su sottogruppi della popolazione;
- **precisione accettabile** delle stime, cioè varianza bassa;
- **contenimento del costo** e dell'ampiezza del campione n.

Per ottenere questi obiettivi, si ricorre a tecniche come il **campionamento stratificato**, il **campionamento a grappoli**, o il **campionamento multistadio**. In questi casi, la struttura stessa del disegno campionario influisce sul calcolo delle stime e sull'analisi dell'errore, ed è parte integrante del modello inferenziale.

11.1.3 Campionamento da un'urna con e senza reintroduzione

Ogni meccanismo di generazione casuale può essere rappresentato da un opportuno **sistema d'urne**. La metafora dell'urna consente di astrarre la nozione di popolazione e di modellare sia fenomeni concreti sia schemi probabilistici più generali.

Un'urna può contenere elementi **in numero finito**, noti o ignoti, e l'estrazione può avvenire **con o senza reintroduzione**, modellando così diversi schemi di campionamento. Ma possiamo spingerci oltre: urne ideali possono rappresentare anche **meccanismi continui** (come una normale) o **conteggi** (come una Poisson), facendo dell'urna un **dispositivo concettuale** che cattura l'idea stessa di casualità controllata.

In questa prospettiva:

- le urne finite senza reintroduzione modellano il campionamento da **popolazioni finite note**, come nei censimenti o nelle statistiche ufficiali;
- le urne con reintroduzione, eventualmente a composizione ignota, rappresentano **popolazioni concettualmente infinite**, come nei modelli Bernoulli o Poisson;
- un'urna che genera numeri reali con legge normale o gamma rappresenta un **modello statistico continuo**, in cui ogni estrazione è una realizzazione di una variabile aleatoria.

La potenza di questa rappresentazione sta nella sua **neutralità applicativa**: possiamo pensare a urne di oggetti fisici, ma anche a urne astratte che sintetizzano comportamenti, eventi, preferenze, misure. In ogni caso, descrivere bene l'urna equivale a **definire un modello statistico**.

Nel nostro percorso considereremo due situazioni fondamentali:

- Il **campionamento con reinserimento** (o con restituzione), dove ogni osservazione è estratta in modo indipendente dalla popolazione o dal modello.
- Il **campionamento senza reinserimento**, dove le osservazioni non sono indipendenti e provengono da una popolazione finita di dimensione N, da cui vengono estratti n elementi senza ripetizioni.

Nel primo caso, $X_1, ..., X_n$ sono **IID** (indipendenti e identicamente distribuite) e rappresentano il caso classico di inferenza da popolazioni infinite o da modelli. Nel secondo, si ha una struttura di **dipendenza debole** tra le unità campionarie, e l'inferenza deve tener conto della dimensione della popolazione e della frazione di campionamento n/N.

11.1.4 Lessico

Un campione in essere (prima di essere osservato) è una sequenza di VC

$$X_1, ..., X_n$$

Tabella 11.1: Campioni SR $n = 2$

	0 ; 1/3	1 ; 1/3	3 ; 1/3	7 ; 1/3
0 ; 1/4	-	$(0,1)$; 1/12	$(0,3)$; 1/12	$(0,7)$; 1/12
1 ; 1/4	$(1,0)$; 1/12	-	$(1,3)$; 1/12	$(1,7)$; 1/12
3 ; 1/4	$(3,0)$; 1/12	$(3,1)$; 1/12	-	$(3,7)$; 1/12
7 ; 1/4	$(7,0)$; 1/12	$(7,1)$; 1/12	$(7,3)$; 1/12	-

un campione osservato è un insieme di numeri

$$\mathbf{x} = (x_1, ..., x_n)$$

Lo *Spazio dei Campioni* $\mathcal{S}$ è il supporto della VC multipla $X_1, ..., X_n$ è l'insieme di tutti i possibili campioni di ampiezza n: $(x_1, ..., x_n)$ che possiamo osservare.

11.1.5 Esempio al finito

Si dispone di un'urna contente 4 bussolotti

$$\{0, 1, 3, 7\}$$

$$\begin{aligned}
\mu &= E(X_i) = \sum_{x \in S_X} x P(X = x) \\
&= 0\frac{1}{4} + 1\frac{1}{4} + 3\frac{1}{4} + 7\frac{1}{4} \\
&= 2.75 \\
\sigma^2 &= V(X_i) = \sum_{x \in S_X} x^2 P(X = x) - \mu^2 \\
&= \left(0^2\frac{1}{4} + 1^2\frac{1}{4} + 3^2\frac{1}{4} + 7^2\frac{1}{4}\right) - (2.75)^2 \\
&= 7.188
\end{aligned}$$

Se estraiamo $n = 2$ volte Senza Reinserimento, lo spazio dei campioni è di dimensione $\#\mathcal{S} = 4 \times 3$ ed è descritto in tabella (11.1).

Se invece estraiamo CR, lo spazio dei campioni è di dimensione $\#\mathcal{S} = 4^2 = 16$ ed è descritto in tabella (11.2).

11.2 Il Modello Statistico

Un **modello statistico** descrive formalmente il processo che ha generato i dati osservati. È composto da due elementi principali:

Tabella 11.2: Campionamento CR $n = 2$

	0 ; 1/4	1 ; 1/4	3 ; 1/4	7 ; 1/4
0 ; 1/4	$(0,0)$; 1/16	$(0,1)$; 1/16	$(0,3)$; 1/16	$(0,7)$; 1/16
1 ; 1/4	$(1,0)$; 1/16	$(1,1)$; 1/16	$(1,3)$; 1/16	$(1,7)$; 1/16
3 ; 1/4	$(3,0)$; 1/16	$(3,1)$; 1/16	$(3,3)$; 1/16	$(3,7)$; 1/16
7 ; 1/4	$(7,0)$; 1/16	$(7,1)$; 1/16	$(7,3)$; 1/16	$(7,7)$; 1/16

1. un **modello probabilistico**, che specifica la distribuzione delle osservazioni in funzione di uno o più parametri incogniti;
2. un **piano di campionamento**, che determina le modalità con cui le osservazioni sono state ottenute dalla popolazione di riferimento.

In queste pagine assumeremo sempre che le osservazioni $X_1, ..., X_n$ siano **indipendenti e identicamente distribuite** (IID), ossia ottenute tramite **campionamento casuale semplice con reinserimento**. Questo tipo di campionamento è coerente con l'ipotesi che ogni unità del campione sia stata estratta in modo indipendente, e che tutte le unità della popolazione abbiano la stessa probabilità di essere selezionate.

Nel nostro contesto, il modello probabilistico scriveremo con $X_1, .., X_n$, n VC IID, replicazioni della stessa variabile $X \sim \mathscr{L}(\theta)$, $\theta \in \Theta$

$$X_i \sim \mathscr{L}(\theta), \forall i = 1, ..., n \qquad \theta \in \Theta,$$

dove $\mathscr{L}(\theta)$ è una famiglia di distribuzioni parametrica e θ è il parametro (o vettore di parametri) che descrive la popolazione e Θ lo spazio in cui θ è definito.

Useremo le lettere greche per i parametri (incogniti) della popolazione e lettere latine per le osservazioni.

11.2.1 Esempi

- $X_i \sim \text{Ber}(\pi)$, con $\pi \in [0, 1]$: modello per variabili dicotomiche, adatto per rappresentare esiti binari (successo/insuccesso). In questo caso $\theta \equiv \pi$ e $\Theta \equiv [0, 1]$.
- $X_i \sim \text{Pois}(\lambda)$, con $\lambda \in \mathbb{R}^+$: modello per conteggi, adatto a descrivere eventi rari su un intervallo fissato. In questo caso $\theta \equiv \lambda$ e $\Theta \equiv \mathbb{R}^+$.
- $X_i \sim N(\mu, \sigma^2)$, con $\mu \in \mathbb{R}, \sigma^2 \in \mathbb{R}^+$: modello per variabili continue, con distribuzione simmetrica e forma a campana. In questo caso $\theta \equiv (\mu, \sigma^2)$ e $\Theta \equiv \mathbb{R} \times \mathbb{R}^+$.
- $X_i \sim \mathscr{L}(\theta)$: notazione generica per una famiglia distribuzionale qualsiasi, dove $\theta \in \Theta$ rappresenta i parametri che vogliamo stimare.

Nel *paradigma classico* la probabilità si assegna alle X_i, in quanto risultato di un sorteggio casuale

$$P(X_1, ..., X_n; \theta)$$

Ma non è consentito trattare con lo strumento della probabilità l'incertezza sul parametro θ che governa la popolazione. Perché θ è incognito ma non è il frutto di una selezione casuale.

Nel *paradigma Bayesiano* l'incertezza sul parametro viene trattata con gli stessi strumenti dell'incertezza sui dati, dando vita ad un teoria coerente e molto utile per alcune applicazioni particolari.

11.2.2 Scopo del modello

L'obiettivo dell'inferenza statistica è trarre conclusioni sul parametro θ, sulla base del campione osservato $\mathbf{x} = (x_1, ..., x_n)$. Lo stesso campione può essere interpretato in modo diverso a seconda del modello adottato: per questo, esplicitare il modello è sempre il primo passo di un'analisi statistica. Perfetto, possiamo riformulare e ampliare il paragrafo sul campionamento, tenendo conto della necessità di collegarlo bene al discorso sull'inferenza e sul modello statistico. Ti propongo una bozza con tono sobrio e coerente con il resto del libro:

11.3 Gli stimatori

Stimare, in statistica, significa scegliere un punto (stima puntuale) o una regione (stima intervallare) dello spazio dei parametri Θ alla luce dei dati $x_1, ..., x_n$.

Uno **stimatore** puntuale (point estimator) è una statistica $\hat{\theta}$ che trasforma il campione $X_1, ..., X_n$ in un punto dello spazio dei parametri:

$$\hat{\theta} : \mathcal{S} \to \Theta$$

Il campione $X_1, ..., X_n$ casuale viene trasformato attraverso $\hat{\theta}$ in un punto specifico di Θ

$$\hat{\theta}(X_1, ..., X_n) = \hat{\theta} \in \Theta$$

Uno **stimatore** è una variabile casuale in quanto funzione di valori casuali.

Esempio 11.3.1. Da una popolazione che ha $E(X_i) = \mu$ incognita, potremmo proporre di *stimare* μ con la media dei dati che *otterremo*:

$$\hat{\mu}(X_1, ..., X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X} = \hat{\mu}$$

Esempio 11.3.2. Da una popolazione di Poisson che ha $X_i \sim \text{Poiss}(\lambda)$, λ incognita, potremmo proporre di *stimare* λ con la mediana dei dati che *otterremo*:

$$\hat{\lambda}(X_1, ..., X_n) = X_{0,5} = \hat{\lambda}$$

11.3.1 Stimatori e Stime

Nota

- Uno **Stimatore**: $\hat{\theta}(X_1, ..., X_n)$ è funzione di $X_1, ..., X_n$ è una **VC**
- Una **Stima**: $\hat{\theta}(x_1, ..., x_n)$ è funzione di $x_1, ..., x_n$ e dunque è un **numero**

Esempio 11.3.3. Da una popolazione che ha $E(X_i) = \mu$ incognita, potremmo proporre di *stimare* μ con la media dei dati che *otterremo*:

$$\hat{\mu}(X_1, ..., X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i = \hat{\mu}$$

Estraiamo $n = 5$ individui dalla popolazione $x_1 = 2.1$, $x_2 = 2.4$, $x_3 = 3.2$, $x_4 = 1.7$, $x_5 = 3.0$,

Per ottenere $\hat{\mu}$ la stima di μ applichiamo $\hat{\theta}$ ai dati e otteniamo

$$\hat{\mu}(2.1, 2.4, 3.2, 7.1, 3.0) = \frac{1}{5}\sum_{i=1}^{5} x_i = \frac{1}{5}17.8 = 3.56$$

Esempio 11.3.4. Da una popolazione di Poisson che ha $X_i \sim \text{Poiss}(\lambda)$, λ incognita, potremmo proporre di *stimare* λ con la mediana dei dati che *otterremo*:

$$\hat{\lambda}(X_1, ..., X_n) = X_{0,5} = \hat{\lambda}$$

Osserviamo $n = 7$ valori (già riordinati) $x_{(1)} = 0$, $x_{(2)} = 0$, $x_{(3)} = 2$, $x_{(4)} = 2$, $x_{(5)} = 3$, $x_{(6)} = 4$, $x_{(7)} = 7$,

Per ottenere $\hat{\lambda}$ la stima di λ applichiamo $\hat{\theta}$ ai dati e otteniamo

$$\hat{\lambda}(x_1, ..., x_7) = x_{0.5} = x_{(4)} = 2$$

11.3.2 Come scegliere uno stimatore

La definizione non offre un criterio per la scelta. Gli stimatori vengono costruiti per avere la miglior *precisione possibile*. La precisione non si può valutare sulla singola *stima* ma studiando, prima di osservare i dati, le proprietà probabilistiche dello *stimatore*.

Le proprietà auspicabili per uno stimatore sono di due tipi

- Esatte (per n finito)
- Asintotiche (per n che diverge)

11.3.3 Proprietà Auspicabili di uno stimatore (per n finito)

Definizione 11.3.1 (Correttezza di uno stimatore). Siano $X_1, ..., X_n$, n VC, IID, replicazioni della stessa $X \sim \mathscr{L}(\theta)$, sia $\hat{\theta}$ uno stimatore per θ. Lo stimatore $\hat{\theta}$ si dice **corretto** se

$$E(\hat{\theta}(X_1, ..., X_n)) = E(\hat{\theta}) = \theta$$

Definizione 11.3.2 (Mean Squared Error di uno stimatore). Si definisce **Errore Quadratico Medio** (*Mean Squared Error*) la quantità

$$MSE(\hat{\theta}) = E((\hat{\theta} - \theta)^2) = V(\hat{\theta}) + B^2(\hat{\theta})$$

dove

$$B(\hat{\theta}) = |E(\hat{\theta}) - \theta|$$

se $\hat{\theta}$ è corretto allora

$$MSE(\hat{\theta}) = V(\hat{\theta})$$

Definizione 11.3.3 (Efficienza di uno stimatore). Siano $\hat{\theta}_1$ e $\hat{\theta}_2$ due stimatori per θ, si dice che $\hat{\theta}_1$ è **più efficiente** di $\hat{\theta}_2$ se e solo se

$$MSE(\hat{\theta}_1) < MSE(\hat{\theta}_2)$$

Nota

Se $\hat{\theta}_1$ e $\hat{\theta}_2$ sono entrambi corretti, allora, $\hat{\theta}_1$ è **più efficiente** di $\hat{\theta}_2$ se e solo se

$$V(\hat{\theta}_1) < V(\hat{\theta}_2)$$

L'**errore** di uno stimatore è l'*inverso* della sua **precisione**.

11.3.4 Media aritmetica e varianza campionaria caso IID

Siano $X_1, ..., X_n$, n VC, IID, replicazioni della stessa X tale che $E(X) = \mu$ e $V(X) = \sigma^2$, sia $\hat{\theta} \equiv \hat{\mu}$ uno stimatore per μ

$$\hat{\mu} = \bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

Dai risultati che già conosciamo sappiamo che

$$E(\hat{\mu}) = \mu$$

e dunque $\hat{\mu}$ è sempre uno stimatore corretto per μ. Essendo $\hat{\mu}$ corretto per μ allora

$$MSE(\hat{\mu}) = V(\hat{\mu}) = \frac{\sigma^2}{n}$$

Si consideri la varianza campionaria:

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \hat{\mu})^2$$

Si può dimostrare che

$$E(\hat{\sigma}^2) = \frac{n-1}{n}\sigma^2 < \sigma^2$$

11.3.5 Media aritmetica campionamento SR (popolazioni finite)

Siano $X_1, ..., X_n$, n VC, osservazioni estratte SR da una popolazione X di N individui, tale che $E(X) = \mu$ e $V(X) = \sigma^2$, sia $\hat{\theta} \equiv \hat{\mu}$ uno stimatore per μ

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Allora

$$E(\hat{\mu}) = \mu$$

e dunque $\hat{\mu}$ è sempre uno stimatore corretto per μ. Essendo $\hat{\mu}$ corretto per μ allora

$$MSE(\hat{\mu}) = V(\hat{\mu})$$

Per calcolare $V(\hat{\mu})$ dobbiamo tenere conto della *frazione di campionamento* n/N

$$MSE(\hat{\mu}) = \frac{N-n}{N-1}\frac{\sigma^2}{n}$$

dove $\frac{N-n}{N-1}$ è chiamato *coefficiente di correzione per popolazioni finite*. Osserviamo che più alto è n più il coefficiente tende ad 1. Se $n = N$ il coefficiente diventa zero il campione è diventato l'intera popolazione e l'incertezza sulla media è zero.

11.3.6 Esempio al finito

Riprendiamo il nostro esempio al finito con una popolazione di $N = 4$

$$\{0, 1, 3, 7\}$$

$$\begin{aligned}
\mu &= E(X_i) = \sum_{x \in S_X} xP(X = x) \\
&= 0\frac{1}{4} + 1\frac{1}{4} + 3\frac{1}{4} + 7\frac{1}{4}
\end{aligned}$$

Tabella 11.3: Valori medi dei campioni SR senza reinserimento di due elementi

	0 ; 1/3	1 ; 1/3	3 ; 1/3	7 ; 1/3
0 ; 1/4	-	0.5 ; 1/12	1.5 ; 1/12	3.5 ; 1/12
1 ; 1/4	0.5 ; 1/12	-	2 ; 1/12	4 ; 1/12
3 ; 1/4	1.5 ; 1/12	2 ; 1/12	-	5 ; 1/12
7 ; 1/4	3.5 ; 1/12	4 ; 1/12	5 ; 1/12	-

$$
\begin{aligned}
&= 2.75 \\
\sigma^2 &= V(X_i) = \sum_{x \in S_X} x^2 P(X = x) - \mu^2 \\
&= \left(0^2\frac{1}{4} + 1^2\frac{1}{4} + 3^2\frac{1}{4} + 7^2\frac{1}{4}\right) - (2.75)^2 \\
&= 7.188
\end{aligned}
$$

Se estraiamo $n = 2$ volte Senza Reinserimento, lo spazio dei campioni è di dimensione $\#\mathcal{S} = 4 \times 3$ ed è descritto in tabella (11.1). Qui scegliamo

$$\hat{\mu} = \frac{1}{2}(X_1 + X_2)$$

In tabella (11.3) la distribuzione della VC media aritmetica campionaria.

Ricaviamo la distribuzione di $\hat{\mu}$

$\hat{\mu}$	0.5	1.5	2	3.5	4	5
$P(\hat{\mu})$	$\frac{2}{12}$	$\frac{2}{12}$	$\frac{2}{12}$	$\frac{2}{12}$	$\frac{2}{12}$	$\frac{2}{12}$

il suo valore atteso

$$E(\hat{\mu}) = \frac{2}{12} \cdot (0.5 + 1.5 + 2 + 3.5 + 4 + 5) = \frac{16.5}{6} = 2.75 = \mu$$

lo stimatore è corretto. Ricaviamo il suo MSE

$$
\begin{aligned}
MSE(\hat{\mu}) &= V(\hat{\mu}) + B^2(\hat{\mu}) \\
&= V(\hat{\mu}) \qquad \text{in virtù della correttezza} \\
&= \Big(0.5^2\frac{1}{6} + 1.5^2\frac{1}{6} + 2^2\frac{1}{6} + \\
&\quad +3.5^2\frac{1}{6} + 4^2\frac{1}{6} + 5^2\frac{1}{6}\Big) - (2.75)^2 \\
&= 2.3958
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{4-2}{4-1}\frac{7.1875}{2} \\
&= \frac{N-n}{N-1}\frac{\sigma^2}{n}
\end{aligned}
$$

Se estraiamo $n = 2$ volte **con reinserimento**, lo spazio dei campioni è di dimensione $\#\mathcal{S} = 4 \times 4$ ed è descritto in tabella (11.2). Anche qui scegliamo:

$$\hat{\mu} = \frac{1}{2}(X_1 + X_2)$$

In tabella qui sotto è riportata la distribuzione della variabile casuale media aritmetica campionaria.

	$0;\ \frac{1}{4}$	$1;\ \frac{1}{4}$	$3;\ \frac{1}{4}$	$7;\ \frac{1}{4}$
$0; 1/4$	$0;\ \frac{1}{16}$	$0.5;\ \frac{1}{16}$	$1.5;\ \frac{1}{16}$	$3.5;\ \frac{1}{16}$
$1; 1/4$	$0.5;\ \frac{1}{16}$	$1;\ \frac{1}{16}$	$2;\ \frac{1}{16}$	$4;\ \frac{1}{16}$
$3; 1/4$	$1.5;\ \frac{1}{16}$	$2;\ \frac{1}{16}$	$3;\ \frac{1}{16}$	$5;\ \frac{1}{16}$
$7; 1/4$	$3.5;\ \frac{1}{16}$	$4;\ \frac{1}{16}$	$5;\ \frac{1}{16}$	$7;\ \frac{1}{16}$

E ricaviamo la distribuzione di, $\hat{\mu}$

$\hat{\mu}$	0	0.5	1	1.5	2	3	3.5	4	5	7
$P(\hat{\mu})$	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{2}{16}$	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{2}{16}$	$\frac{2}{16}$	$\frac{1}{16}$

Calcoliamo valore atteso e MSE.

$$
\begin{aligned}
E(\hat{\mu}) &= \sum_{\hat{\mu}} \hat{\mu} P(\hat{\mu}) \\
&= 0\frac{1}{16} + 0.5\frac{2}{16} + 1\frac{1}{16} + 1.5\frac{2}{16} + 2\frac{2}{16} + \\
&\quad +3\frac{1}{16} + 3.5\frac{2}{16} + 4\frac{2}{16} + 5\frac{2}{16} + 7\frac{1}{16} \\
&= 2.75 = \mu \qquad \text{lo stimatore è corretto} \\
MSE(\hat{\mu}) &= V(\hat{\mu}) + B^2(\hat{\mu}) \\
&= V(\hat{\mu}) \\
&= \left(0^2\frac{1}{16} + 0.5^2\frac{2}{16} + 1^2\frac{1}{16} + 1.5^2\frac{2}{16} + 2^2\frac{2}{16} + \right. \\
&+ \left. 3^2\frac{1}{16} + 3.5^2\frac{2}{16} + 4^2\frac{2}{16} + 5^2\frac{2}{16} + 7^2\frac{1}{16}\right) - (2.75)^2 \\
&= 3.59375 \\
&= \frac{7.1875}{2}
\end{aligned}
$$

$$= \frac{\sigma^2}{n}$$

Consideriamo la stima della varianza

$$\hat{\sigma}^2 = \frac{1}{2}\left((X_1 - \hat{\mu})^2 + (X_2 - \hat{\mu})^2\right)$$

	0; $\frac{1}{4}$	1; $\frac{1}{4}$	3; $\frac{1}{4}$	7; $\frac{1}{4}$
0; 1/4	0; $\frac{1}{16}$	0.25; $\frac{1}{16}$	2.25; $\frac{1}{16}$	12.25; $\frac{1}{16}$
1; 1/4	0.25; $\frac{1}{16}$	0; $\frac{1}{16}$	1; $\frac{1}{16}$	9; $\frac{1}{16}$
3; 1/4	2.25; $\frac{1}{16}$	1; $\frac{1}{16}$	0; $\frac{1}{16}$	4; $\frac{1}{16}$
7; 1/4	12.25; $\frac{1}{16}$	9; $\frac{1}{16}$	4; $\frac{1}{16}$	0; $\frac{1}{16}$

E ricaviamo la distribuzione di, $\hat{\sigma}^2$

$\hat{\sigma}^2$	0	0.25	1	2.25	4	9	12.25
$P(\hat{\sigma}^2)$	$\frac{4}{16}$	$\frac{2}{16}$	$\frac{2}{16}$	$\frac{2}{16}$	$\frac{2}{16}$	$\frac{2}{16}$	$\frac{2}{16}$

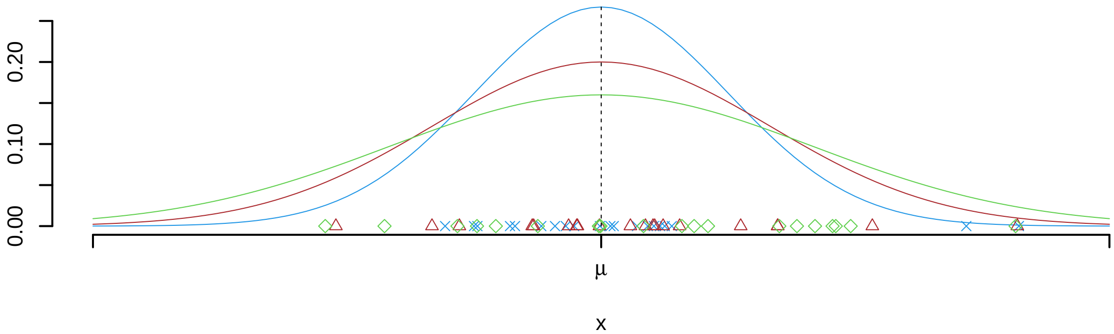
e osserviamo che

$$\begin{aligned}
E(\hat{\sigma}^2) &= \sum_{\hat{\sigma}^2} \hat{\sigma}^2 P(\hat{\sigma}^2) \\
&= 0\frac{4}{16} + 0.25\frac{2}{16} + 1\frac{2}{16} + 2.25\frac{2}{16} + \\
&+ 4\frac{2}{16} + 9\frac{2}{16} + 12.25\frac{2}{16} \\
&= 3.59375 \\
&= \frac{1}{2}7.1875 \\
&= \frac{n-1}{n}\sigma^2
\end{aligned}$$

11.3.7 Distribuzione delle statistiche

Supponiamo di estrarre campioni di ampiezza n CR da una popolazione con media μ e varianza σ^2. Osservo tre diversi stimatori per μ: $\hat{\mu}_1$, $\hat{\mu}_2$ e $\hat{\mu}_3$

Se si ripetesse l'estrazione un grande numero di volte potremmo vedere i tre stimatori nel grafico qui di seguito. Lo stimatore in con la distribuzione in blu è il più efficiente dei tre: la probabilità che si avveri lontano dal vero parametro è minore che per gli altri due. Mentre lo stimatore in con la distribuzione in verde è il meno efficiente dei tre: la probabilità che si avveri lontano dal vero parametro è maggiore che per gli altri due.



11.3.8 Proprietà Auspicabili di uno stimatore (per $n \to \infty$)

Siano $X_1, ..., X_n$, n VC, IID, replicazioni della stessa $X \sim \mathscr{L}(\theta)$, sia $\hat{\theta}$ uno stimatore per θ.

Definizione 11.3.4 (Correttezza Asintotica). Lo stimatore $\hat{\theta}$ si dice **asintoticamente corretto** se

$$\lim_{n\to\infty} E(\hat{\theta}(X_1, ..., X_n)) = E(\hat{\theta}) = \theta$$

Esempio 11.3.5.

$$\lim_{n\to\infty} E(\hat{\sigma}^2) = \lim_{n\to\infty} \frac{n-1}{n}\sigma^2 = \sigma^2$$

Definizione 11.3.5 (Correttezza Asintotica). Lo stimatore $\hat{\theta}$ si dice **consistente** (in media quadratica) se e solo se

$$\lim_{n\to\infty} MSE(\hat{\theta}(X_1, ..., X_n)) = \lim_{n\to\infty} MSE(\hat{\theta}) = 0$$

Essendo

$$MSE(\hat{\theta}) = V(\hat{\theta}) + B^2(\hat{\theta})$$

allora

$$\lim_{n\to\infty} MSE(\hat{\theta}) = 0, \text{ se e solo se } \lim_{n\to\infty} V(\hat{\theta}) = 0 \text{ e } \lim_{n\to\infty} B^2(\hat{\theta}) = 0$$

Esempio 11.3.6 (Consistenza). Siano $X_1, ..., X_n$, n VC, IID, replicazioni della stessa VC X con $E(X) = \mu$ e $V(X) = \sigma^2$ Usiamo $\hat{\mu}$ per stimare μ:

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{S_n}{n}$$

Siccome $\hat{\mu}$ è stimatore corretto per μ:

$$E(\hat{\mu}) = E\left(\frac{X_1 + ... + X_n}{n}\right) = \frac{1}{n}(E(X_1) + ... + E(X_n)) = \frac{1}{n}(\mu + ... + \mu) = \mu$$

Allora

$$MSE(\hat{\mu}) = V(\hat{\mu}) = \frac{\sigma^2}{n}$$

Al divergere di n

$$\lim_{n\to\infty} MSE(\hat{\mu}) = \lim_{n\to\infty} \frac{\sigma^2}{n} = 0$$

Lo stimatore $\hat{\mu}$ per μ è stimatore **corretto** e **consistente**.

11.4 La SD e lo SE

La *standard deviation* (SD) σ, rappresenta la dispersione degli individui dalla media, è un indicatore di *variabilità* della *popolazione*, per esempio in una popolazione finita di N individui:

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2},$$

la *deviazione standard* σ è la radice della varianza della popolazione σ^2.
Lo *standard error* $SE(\hat{\theta})$ di uno stimatore $\hat{\theta}$ per θ è un indicatore della *variabilità* dello stimatore nello *spazio dei parametri*

$$SE(\hat{\theta}) = \sqrt{V(\hat{\theta})}$$

Lo *standard error* $SE(\hat{\theta})$ di uno stimatore $\hat{\theta}$ per θ è la radice della varianza della VC $\hat{\theta}$.
La *standard deviation stimata* σ, rappresenta la dispersione degli individui *del campione* dalla media *del campione*, è un indicatore di *variabilità* del *campione*:

$$\hat{\sigma} = \sqrt{\hat{\sigma}^2} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\mu})^2}$$

La *deviazione standard stimata* $\hat{\sigma}$ è la radice della varianza del campione $\hat{\sigma}^2$.

Esempio 11.4.1. Lo standard error dello stimatore media aritmetica campionaria $\hat{\mu}$ per μ

$$SE(\hat{\mu}) = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$$

L'errore che si commette nello stimare una media dipende da due fattori

- la *standard deviation* σ che indica la variabilità degli individui tra di loro

- $1/\sqrt{n}$ che è l'inverso dell'ampiezza del campione

Se σ è incognito viene stimato da (come vedremo nel paragrafo (12.9.4))

$$S = \sqrt{S^2} = \sqrt{\frac{n}{n-1}\hat{\sigma}^2}$$

Ottenendo

$$\widehat{SE(\hat{\mu})} = \frac{S}{\sqrt{n}}$$

Teoria della Verosimiglianza

12

12.1 Il Modello Statistico

Un modello statistico è l'insieme di un modello probabilistico $X_i \sim \mathscr{L}(\theta)$ e di un piano di campionamento dalla popolazione $\mathscr{P}$. In queste pagine considereremo come unico piano di campionamento il campionamento casuale semplice **con** reintroduzione, ovvero assumeremo sempre le ipotesi IID. Per esempio sono un modello statistico:

- Le $X_1, ..., X_n$ sono IID, replicazioni di $X \sim \text{Ber}(\pi)$, $\pi \in [0, 1]$.
- Le $X_1, ..., X_n$ sono IID, replicazioni di $X \sim \text{Pois}(\lambda)$, $\lambda \in \mathbb{R}^+$.
- Le $X_1, ..., X_n$ sono IID, replicazioni di $X \sim N(\mu, \sigma^2)$, $(\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}^+$.
- Le $X_1, ..., X_n$ sono IID, replicazioni di $X \sim \mathscr{L}(\theta)$, $\theta \in \Theta$-

12.1.1 Esiste lo stimatore più efficiente?

Dipende dalle informazioni che abbiamo sulla Popolazione $\mathscr{P}$ In contesti *distribution free* e ipotesi IID, senza alcuna ulteriore conoscenza della popolazione, non esistono procedure che possano essere dimostrate *ottimali*, il ricercatore valuta da caso a caso, previa un'attenta analisi descrittiva preliminare. In ipotesi *distribution free* l'intera distribuzione della variabile X è incognita. Se spostiamo l'attenzione all'*inferenza da modello* ipotizziamo di conoscere la forma della distribuzione di probabilità delle X a meno dei suoi parametri.

Sotto alcune condizioni di regolarità gli stimatori più efficienti sono gli stimatori di Massima Verosimiglianza. Lo stimatore di massima verosimiglianza parte dell'assunto che tutta l'informazione che un campione porge nella comprensione della popolazione risieda in una misura chiamata *Verosimiglianza*.

12.2 La Verosimiglianza

La *Verosimiglianza* è una misura di incertezza **non** sul risultato di un esperimento casuale, ma sui meccanismi che generano una sequenza casuale. Nella teoria della verosimiglianza si può parlare di probabilità solo per il campione, ma non per i meccanismi che lo hanno generato. Una volta osservati i dati la conoscenza di questi meccanismi diventa più o meno *verosimile* agli occhi del ricercatore alla luce dell'osservazione.

Nella teoria della verosimiglianza, dunque, si usano due termini diversi: *probabilità* per indicare la misura dell'incertezza sui risultati dell'estrazione del campione e *verosimiglianza* per indicare la misura di incertezza sui meccanismi che hanno prodotto il campione.

La *teoria della verosimiglianza* presuppone la totale ignoranza del ricercatore che esplora un sistema casuale di cui sta cercando di comprendere i parametri.

Se per esempio voglio conoscere la probabilità π di una moneta truccata di porgere Testa, la teoria della verosimiglianza presuppone che per me, prima di osservare il campione, tutti i possibili valori di π siano equamente verosimili, compresi quelli più estremi.

Questa totale ignoranza non è sempre giustificata e per allargare la teoria della verosimiglianza rimando il lettore su testi di statistica Bayesiana che sfruttano il teorema di Bayes per costruire una misura alternativa (più ampia) della verosimiglianza, basata solo sul concetto allargato di probabilità.

> Donovan, T. M., and Mickey, R. M. (2019). Bayesian Statistics for Beginners: A Step-by-Step Approach. Oxford: Oxford University Press.

La *funzione di verosimiglianza* è la una funzione di probabilità dei dati, fissata sul campione osservato, in cui la variabile è il parametro. La verosimiglianza è indicata con la lettera L (*Likelihood*) e si scrive

$$L(\theta; \text{Dati}) \propto P(\text{Dati}; \theta)$$

e si legge che la *verosimiglianza* L di θ è proporzionale $\propto$ alla probabilità di osservare i dati osservati nell'ipotesi che θ sia vera. Il simbolo proporzionale $\propto$ significa che:

$$L(\theta; \text{Dati}) = Const. \cdot P(\text{Dati}; \theta)$$

dove $Const.$ è una constante qualunque che non dipende da θ. Il valore di $L(\theta; \text{Dati})$ per un θ fissato, non ha alcune significato se non è confrontato con altri valori. Infatti $L(\theta; \text{Dati})$ **non** è una probabilità, ma se

$$L(\theta_1; \text{Dati}) > L(\theta_2; \text{Dati})$$

Significa che l'ipotesi che sia stato $\theta = \theta_1$ il valore del parametro del modello che ha generato i dati è più verosimile dell'ipotesi che sia stato $\theta = \theta_2$.

12.2.1 La Verosimiglianza attraverso un esempio

Supponiamo di avere un'urna che ha solo $N = 10$ bussolotti alcuni bianchi B e i rimanenti non bianchi $\overline{B} = N - B$, ma non conosciamo B. Il numero di bianchi B potrà essere $0, 1, ..., 10$. La VC X che registra l'evento bianco o nero di una estrazione è chiaramente Bernoulli $X \sim \text{Ber}(\pi)$ di parametro:

$$\pi = \frac{B}{10}$$

π è la proporzione di bussolotti bianchi nell'urna. In questo specifico esempio:

$$\pi \in \left\{ \frac{0}{10} = 0.0, \frac{1}{10} = 0.1, ..., \frac{9}{10} = 0.9, \frac{10}{10} = 1.0 \right\}$$

Lo spazio dei parametri ha dimensione 11. Uno stimatore ha il compito di scegliere uno di questi 11 valori.

Estraiamo $n = 5$ bussolotti CR (IID) e otteniamo

$$x_1 = 0, x_2 = 1, x_3 = 1, x_4 = 0, x_5 = 1$$

Se conoscessi π attraverso il calcolo delle probabilità saprei calcolare la probabilità della sequenza (ordinata) 0,1,1,0,1 proveniente da 5 esperimenti di Bernoulli IID

$$
\begin{aligned}
P(X_1 = 0 \cap X_2 = 1 \cap X_3 = 1 \cap X_4 = 0 \cap X_5 = 1; \pi) = \\
&= P(X_1 = 0; \pi)P(X_2 = 1; \pi)P(X_3 = 1; \pi)P(X_4 = 0; \pi)P(X_5 = 1; \pi) \\
&= (1 - \pi)\pi\pi(1 - \pi)\pi \\
&= \pi^3(1 - \pi)^{5-3}
\end{aligned}
$$

La posso calcolare per ogni possibile valore di $\pi \in \{0.0, 0.1, ..., 1.0\}$.

12.2.2 Se π fosse...

A questo punto graduare decidere i valori di π tra più e meno *verosimili* alla luce dei dati $\mathbf{x} = (0, 1, 1, 0, 1)$. Questo si fa sostituendo π con i suoi possibili valori e calcolando la *ipotetica* probabilità.

Se fosse $\pi = 0$ con quale probabilità avrei osservato la sequenza $0, 1, 1, 0, 1$?

$$0^3 \cdot (1 - 0)^2 = 0 \qquad \text{l'ipotesi } \pi = 0 \text{ ha verosimiglianza proporzionale a } 0$$

Se fosse $\pi = 0.1$ con quale probabilità avrei osservato la sequenza $0, 1, 1, 0, 1$?

$$0.1^3 \cdot (1 - 0.1)^2 = 0.00081 \qquad \text{l'ipotesi } \pi = 0.1 \text{ ha verosimiglianza proporzionale a } 0.00081$$

Se fosse $\pi = 0.2$ con quale probabilità avrei osservato la sequenza $0, 1, 1, 0, 1$?

$$0.2^3 \cdot (1 - 0.2)^2 = 0.00512 \qquad \text{l'ipotesi } \pi = 0.2 \text{ ha verosimiglianza proporzionale a } 0.00512$$

Se fosse $\pi = 0.3$ con quale probabilità avrei osservato la sequenza $0, 1, 1, 0, 1$?

$$0.3^3 \cdot (1 - 0.3)^2 = 0.01323 \qquad \text{l'ipotesi } \pi = 0.3 \text{ ha verosimiglianza proporzionale a } 0.01323$$

Se fosse $\pi = 0.4$ con quale probabilità avrei osservato la sequenza $0, 1, 1, 0, 1$?

$$0.4^3 \cdot (1 - 0.4)^2 = 0.02304 \qquad \text{l'ipotesi } \pi = 0.4 \text{ ha verosimiglianza proporzionale a } 0.02304$$

Se fosse $\pi = 0.5$ con quale probabilità avrei osservato la sequenza $0, 1, 1, 0, 1$?

$$0.5^3 \cdot (1 - 0.5)^2 = 0.03125 \qquad \text{l'ipotesi } \pi = 0.5 \text{ ha verosimiglianza proporzionale a } 0.03125$$

Se fosse $\pi = 0.6$ con quale probabilità avrei osservato la sequenza $0, 1, 1, 0, 1$?

$$0.6^3 \cdot (1 - 0.6)^2 = 0.03456 \qquad \text{l'ipotesi } \pi = 0.6 \text{ ha verosimiglianza proporzionale a } 0.03456$$

Se fosse $\pi = 0.7$ con quale probabilità avrei osservato la sequenza $0, 1, 1, 0, 1$?

$$0.7^3 \cdot (1 - 0.7)^2 = 0.03087 \qquad \text{l'ipotesi } \pi = 0.7 \text{ ha verosimiglianza proporzionale a } 0.03087$$

Se fosse $\pi = 0.8$ con quale probabilità avrei osservato la sequenza $0, 1, 1, 0, 1$?

$$0.8^3 \cdot (1 - 0.8)^2 = 0.02048 \qquad \text{l'ipotesi } \pi = 0.8 \text{ ha verosimiglianza proporzionale a } 0.02048$$

Se fosse $\pi = 0.9$ con quale probabilità avrei osservato la sequenza $0, 1, 1, 0, 1$?

$$0.9^3 \cdot (1-0.9)^2 = 0.00729 \qquad \text{l'ipotesi } \pi = 0.9 \text{ ha verosimiglianza proporzionale a } 0.00729$$

Se fosse $\pi = 1$ con quale probabilità avrei osservato la sequenza $0, 1, 1, 0, 1$?

$$1^3 \cdot (1-1)^2 = 0 \qquad \text{l'ipotesi } \pi = 1 \text{ ha verosimiglianza proporzionale a } 0$$

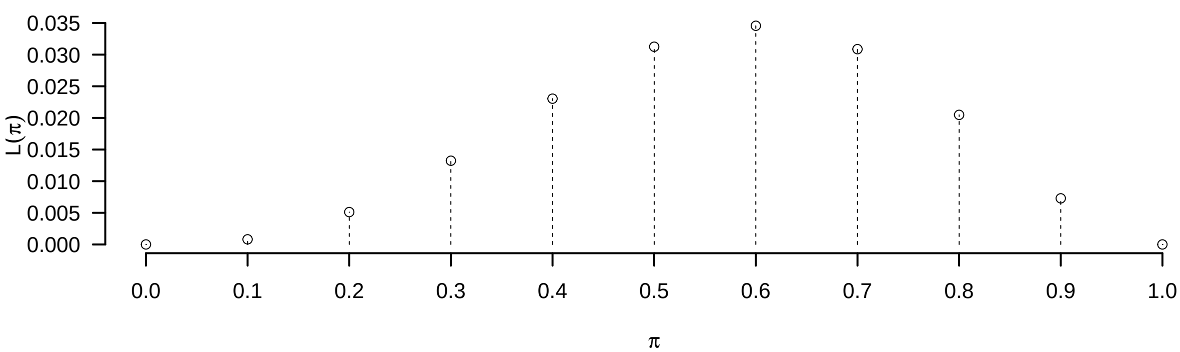
Definiamo la **funzione di verosimiglianza** (Likelihood), la funzione L del parametro incognito π alla luce dei dati $X_1 = 0, X_2 = 1, X_3 = 1, X_4 = 0, X_5 = 1$ osservati:

$$\begin{aligned} L(\pi; X_1 = 0, X_2 = 1, X_3 = 1, X_4 = 0, X_5 = 1) = & \\ = & \ L(\pi) \\ = & \ K \cdot P(X_1 = 0 \cap X_2 = 1 \cap X_3 = 1 \cap X_4 = 0 \cap X_5 = 1; \pi) \qquad \text{con } K > 0 \\ \propto & \ P(X_1 = 0 \cap X_2 = 1 \cap X_3 = 1 \cap X_4 = 0 \cap X_5 = 1; \pi) \\ \propto & \ \pi^3 (1-\pi)^2 \end{aligned}$$

La verosimiglianza gradua quanto un certo valore di π è compatibile con i dati osservati. Per esempio l'ipotesi $\pi = 0.5$ è più verosimile dell'ipotesi $\pi = 0.4$, alla luce dei dati $x_1 = 0, x_2 = 1, x_3 = 1, x_4 = 0, x_5 = 1$,

$$L(0.5) = 0.0312 > L(0.4) = 0.023$$

Se mettiamo π in ascissa e $L(\pi)$ in ordinata, otteniamo il grafico della verosimiglianza di π, alla luce dei dati osservati.

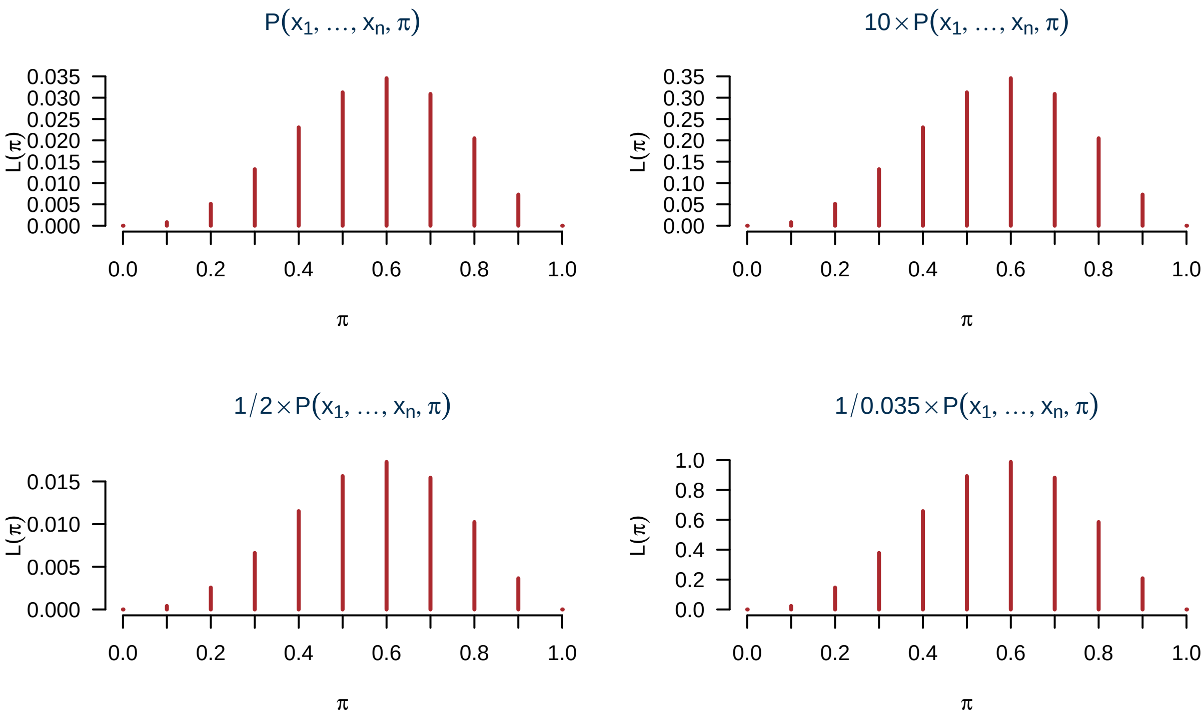


12.2.3 La verosimiglianza non è una probabilità

Notiamo che

$$\sum_{\pi \in \{0.0, 0.1, \ldots, 1.0\}} L(\pi) = 0 + 0.0008 + \ldots + 0 = 0.1666 \neq 1$$

La possiamo moltiplicare per un numero qualunque



12.2.4 La stima di massima verosimiglianza

$\hat{\pi}_{ML} = \hat{\pi}$ è

$$\hat{\pi} \in \{0.0, 0.1, 0.2, ..., 1.0\} : L(\hat{\pi}) > L(\pi), \forall \pi \neq \hat{\pi}$$

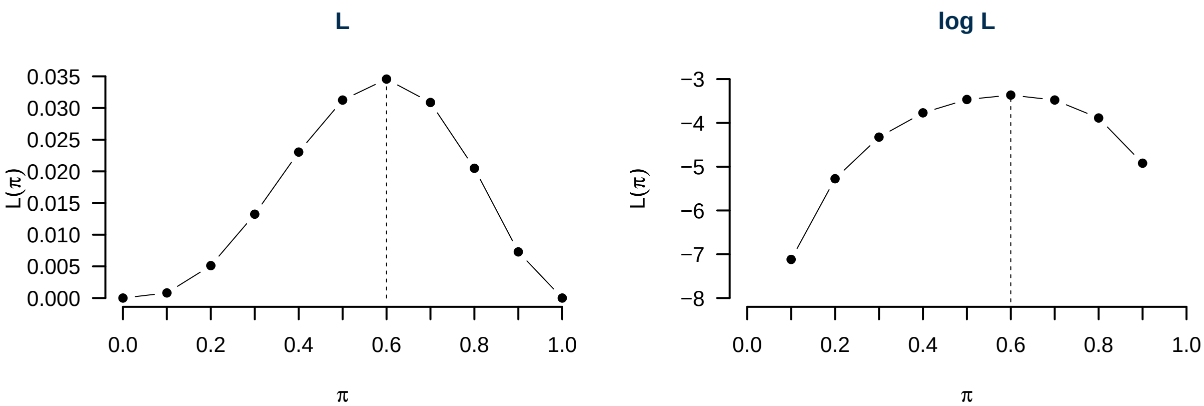
E quindi:

$$\hat{\pi} = 0.6 = \frac{3}{5}$$

Consideriamo ℓ, il logaritmo di L

$$\ell(\pi) = \log L(\pi)$$

$\pi =$	0	0.1000	0.2000	0.3000	0.400	0.5000	0.6000	0.7000	0.8000	0.9000	1
$L(\pi)$	0	0.0008	0.0051	0.0132	0.023	0.0312	0.0346	0.0309	0.0205	0.0073	0
$\ell(\pi)$	-Inf	-7.1185	-5.2746	-4.3253	-3.771	-3.4657	-3.3651	-3.4780	-3.8883	-4.9213	-Inf



12.2.5 Esempio IID da popolazione finita (parte due)

Riprendiamo l'esempio di prima: un'urna che ha solo $N = 10$ bussolotti alcuni bianchi B altri neri N, ma non conosciamo B ed N. Il numero di bianchi B potrà essere $0, 1, ..., 10$ e π è la proporzione di bussolotti bianchi nell'urna

$$\pi = \frac{B}{10}$$

Estraiamo $n = 5$ bussolotti CR (IID) e otteniamo 3 *successi* (bussolotto bianco) e 2 *insuccessi* (bussolotto nero). Non conosciamo l'ordine.

$X \sim \text{Binom}(5, \pi)$, il mio campione è un'estrazione dalla binomiale con $n = 5$. Speculiamo su π

Se conoscessi π attraverso il calcolo delle probabilità saprei calcolare la probabilità $P(X) = 3$, con $X \sim \text{Binom}(5, \pi)$

$$P(X = 3; \pi) = \binom{5}{3} \pi^3 (1 - \pi)^{5-3} = 10 \cdot \pi^3 (1 - \pi)^2$$

Se fosse $\pi = 0$ ($B = 0$) con quale probabilità avrei osservato la sequenza $X = 3$?

$$\binom{5}{3} 0^3 \cdot (1 - 0)^2 = 0$$

- l'ipotesi $\pi = 0$ ha **verosimiglianza** proporzionale a zero

Se fosse $\pi = 0.1$ ($B = 1$) con quale probabilità avrei osservato la sequenza $X = 3$?

$$\binom{5}{3} 0.1^3 \cdot (1 - 0.1)^2 = 0.0081$$

- l'ipotesi $\pi = 0.1$ ha **verosimiglianza** proporzionale a 0.0081

...

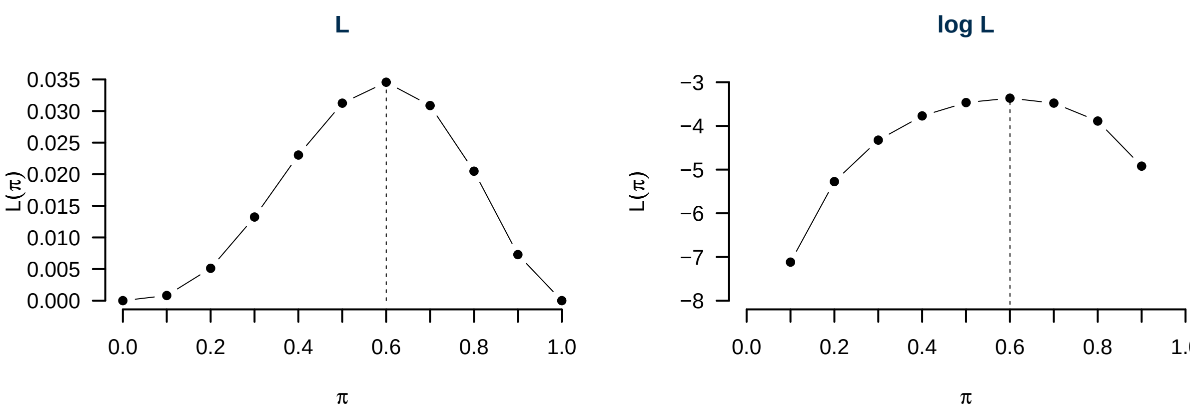
Definiamo la funzione di verosimiglianza (Likelihood), la funzione L del parametro incognito π alla luce dei dati $x = 3$ osservati:

$$\begin{aligned} L(\pi; x = 3) &= L(\pi) \\ &\propto P(x = 3; \pi) \\ &= \binom{5}{3}\pi^3(1-\pi)^2 \\ &\propto \pi^3(1-\pi)^2 \end{aligned}$$

La tabella

$\pi =$	0	0.1000	0.2000	0.3000	0.400	0.5000	0.6000	0.7000	0.8000	0.9000	1
$L(\pi)$	0	0.0008	0.0051	0.0132	0.023	0.0312	0.0346	0.0309	0.0205	0.0073	0
$\log(L(\pi))$	-Inf	-7.1185	-5.2746	-4.3253	-3.771	-3.4657	-3.3651	-3.4780	-3.8883	-4.9213	-Inf

e il grafico



12.2.6 Abbiamo trovato il vero π?

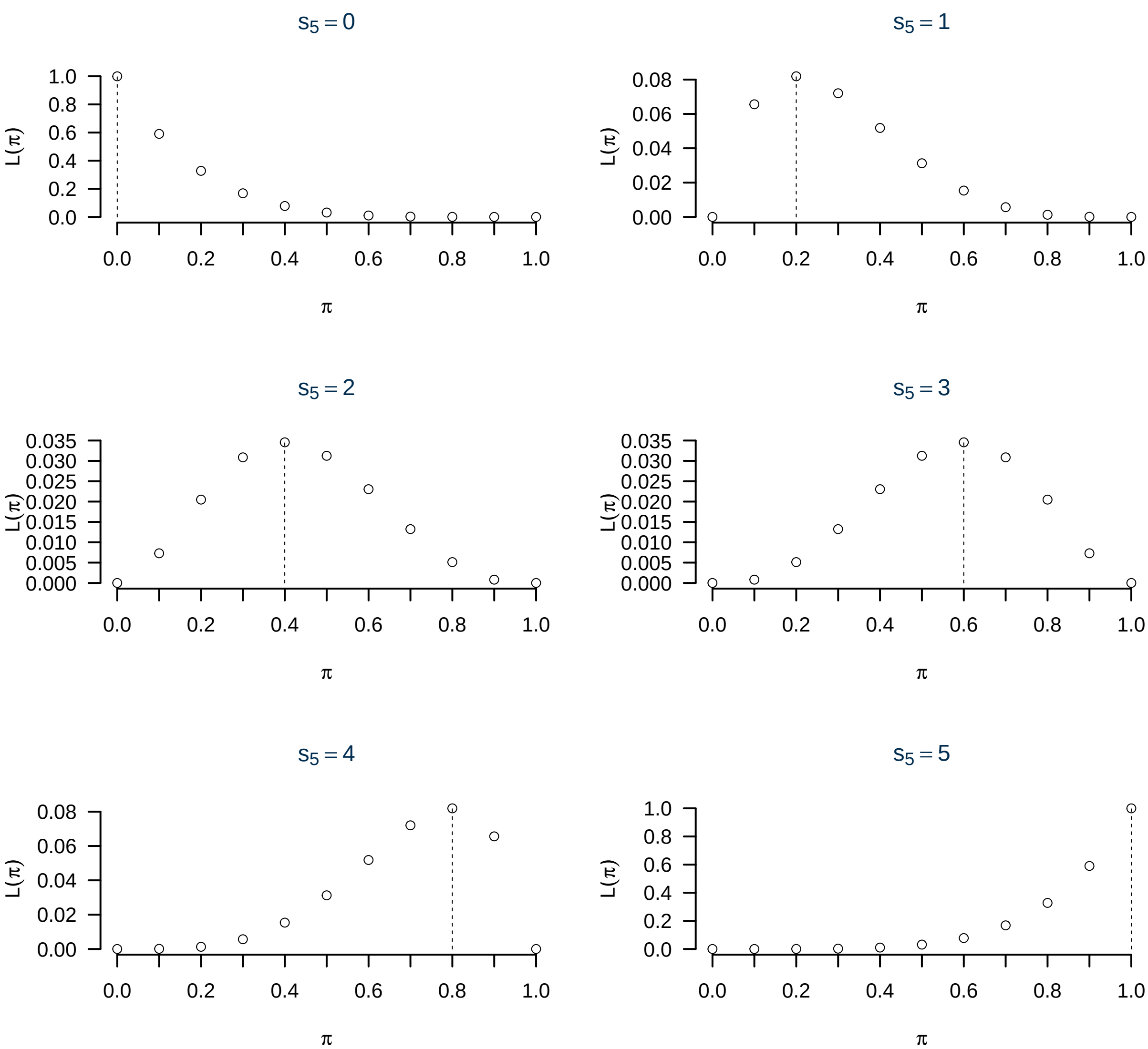
Ovviamente $\hat{\pi}$ non è π che non conosceremo mai, $\hat{\pi} = 0.6$ è il valore *più verosimile* tra tutti i possibili valori di π, ma non è π. Ci possiamo chiedere se, per esempio, l'ipotesi $\pi = 0.5$ è *"impossibile"*. Anche in questo caso la risposta è negativa, $\pi = 0.5$ è solo, alla luce dei dati, *meno verosimile* dell'ipotesi $\pi = 0.6$. E possiamo anche calcolare di di quanto:

$$\frac{L(\hat{\pi} = 0.6)}{L(\pi = 0.5)} = \frac{0.3456}{0.3125} = 1.1059$$

Alla luce dei dati (3 successi su 5 estrazioni) il valore $\hat{\pi} = 0.6$ è il 10.592% *più verosimile* di $\pi = 0.5$.

$$(1.1059 - 1) \times 100\% = 10.592\%$$

12.2.7 Muoviamo anche S_n



In sintesi, lo spazio $X \times \Theta$ è l'incrocio tra tutti i possibili π e tutti i possibili s_n, ne esce una matrice con 10 righe e 10 colonne dove le righe rappresentano s_n e le colonne π.

	$s_5 = 0$	$s_5 = 1$	$s_5 = 2$	$s_5 = 3$	$s_5 = 4$	$s_5 = 5$
$\pi = 0$	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000
$\pi = 0.1$	0.5905	0.3280	0.0729	0.0081	0.0005	0.0000
$\pi = 0.2$	0.3277	0.4096	0.2048	0.0512	0.0064	0.0003
$\pi = 0.3$	0.1681	0.3601	0.3087	0.1323	0.0284	0.0024
$\pi = 0.4$	0.0778	0.2592	0.3456	0.2304	0.0768	0.0102
$\pi = 0.5$	0.0312	0.1562	0.3125	0.3125	0.1562	0.0312
$\pi = 0.6$	0.0102	0.0768	0.2304	0.3456	0.2592	0.0778
$\pi = 0.7$	0.0024	0.0284	0.1323	0.3087	0.3601	0.1681
$\pi = 0.8$	0.0003	0.0064	0.0512	0.2048	0.4096	0.3277
$\pi = 0.9$	0.0000	0.0004	0.0081	0.0729	0.3280	0.5905
$\pi = 1$	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000

questa tabella, letta per righe ci indica la probabilità, letta per colonne ci indica la *verosimiglianza*.

12.3 La Funzione di Verosimiglianza

Definizione 12.3.1 (Funzione di Verosimiglianza). Siano $x_1, ..., x_n$ n osservazioni di $X \sim \mathscr{L}(\theta)$, $\theta \in \Theta$, si definisce la verosimiglianza L di θ la funzione:

$$L(\theta; x_1, ..., x_n) = L(\theta) \propto P(X_1 = x_1, ..., X_n = x_n; \theta)$$

La funzione di verosimiglianza è una funzione in θ (la variabile) per $x_1, ..., x_n$ fissi. Indica quanto un particolare valore di θ è supportato dai dati. Più alta è la verosimiglianza più i valori di θ che la rendono alta sono supportati dall'evidenza campionaria. Se $x_1, .., x_n$ sono osservazioni IID otteniamo

$$\begin{aligned} L(\theta) &\propto P(X_1 = x_1; \theta) \cdot ... \cdot P(X_n = x_n; \theta) \\ &\propto f(x_1; \theta) \cdot ... \cdot f(x_n; \theta) \\ &\propto \prod_{i=1}^{n} f(x_i; \theta) \end{aligned}$$

Definizione 12.3.2 (Log Verosimiglianza). Si definisce la log-verosimiglianza ℓ:

$$\begin{aligned} \ell(\theta) &= \log L(\theta) \\ &= \log \prod_{i=1}^{n} f(x_i; \theta) \\ &= \sum_{i=1}^{n} \log f(x_i; \theta) \end{aligned}$$

12.4 La Stimatore di massima Verosimiglianza

Definizione 12.4.1 (Stimatore du Massima Verosimiglianza). Lo stimatore di *massima verosimiglianza* per θ è

$$\begin{aligned}\hat{\theta} &= \underset{\theta\in\Theta}{\operatorname{argmax}}\, L(\theta) \\ &= \underset{\theta\in\Theta}{\operatorname{argmax}}\, \ell(\theta)\end{aligned}$$

$$\hat{\theta} : L(\hat{\theta}) > L(\theta), \forall\theta \neq \hat{\theta}, \qquad \ell(\hat{\theta}) > \ell(\theta), \forall\theta \neq \hat{\theta}$$

12.5 Il Principio di Verosimiglianza

Secondo la teoria della verosimiglianza, dato un modello statistico tutta l'informazione che un campione $\mathbf{x} = (x_1, ..., x_n)$ porge a θ è contenuta nella sua funzione di verosimiglianza.

12.6 Verosimiglianza e Statistiche Sufficienti

da scrivere

12.7 Caso Bernoulli urna infinita.

Se l'urna è infinita $N \to \infty$, allora $\pi \in [0, 1]$. Le variabili $X_1, ..., X_n$ tutte replicazioni IID di $X \sim \text{Ber}(\pi)$, si realizzano in $x_1, ..., x_n$.

Esempio. $n = 5$, $x_1 = 0, x_2 = 1, x_3 = 1, x_4 = 0, x_5 = 1$, La probabilità della singola estrazione è

$$P(X_i = x_i; \pi) = f(x_i; \pi) = \pi^{x_i}(1 - \pi)^{1-x_i}$$

La verosimiglianza è

$$\begin{aligned}L(\pi) &\propto \prod_{i=1}^{n} f(x_i; \pi) \\ &= \prod_{i=1}^{n} \pi^{x_i}(1 - \pi)^{1-x_i} \\ &= \pi^{x_1}(1 - \pi)^{1-x_1}\pi^{x_2}(1 - \pi)^{1-x_2}...\pi^{x_n}(1 - \pi)^{1-x_n} \\ &= \pi^{x_1}\pi^{x_2}...\pi^{x_n} \quad (1 - \pi)^{1-x_1}(1 - \pi)^{1-x_2}...(1 - \pi)^{1-x_n} \\ &= \pi^{x_1+x_2+...+x_n}(1 - \pi)^{1-x_1+1-x_2+...+1-x_n} \\ &= \pi^{\sum_{i=1}^{n} x_i}(1 - \pi)^{n-\sum_{i=1}^{n} x_i} \\ &= \pi^{s_n}(1 - \pi)^{n-s_n}, \qquad s_n = \sum_{i=1}^{n} x_i\end{aligned}$$

La statistica s_n contiene **tutta** l'informazione del campione $x_1, ..., x_n$. La log-verosimiglianza è

$$\begin{aligned}
\ell(\pi) &= \log L(\pi) \\
&= \log \pi^{s_n}(1-\pi)^{n-s_n} \\
&= \log \pi^{s_n} + \log(1-\pi)^{n-s_n} \\
&= s_n \log \pi + (n - s_n)\log(1-\pi)
\end{aligned}$$

Per derivare il π che rende massima la verosimiglianza si deve derivare la funzione ℓ ed uguagliare a zero la derivata prima:

$$\ell'(\pi) = \frac{s_n}{\pi} + (-1)\frac{n-s_n}{1-\pi} = \frac{s_n}{\pi} - \frac{n-s_n}{1-\pi}$$

$\hat{\pi}$ è dunque quel valore tale che

$$\ell'(\hat{\pi}) = 0$$

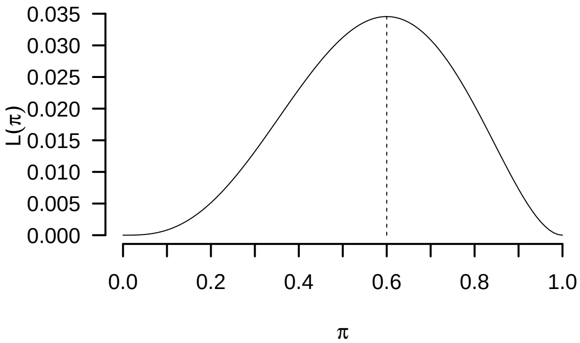
Eguagliamo a zero la derivata prima della log verosimiglianza:

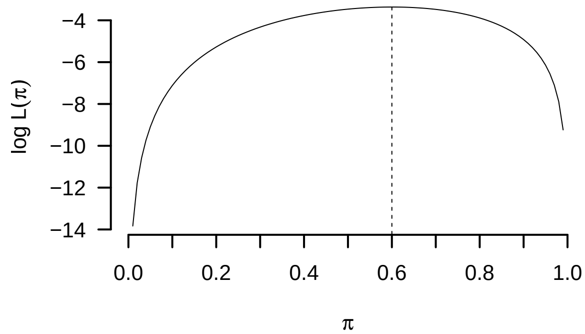
$$\begin{aligned}
\ell'(\pi) &= 0 \\
\frac{s_n}{\pi} - \frac{n-s_n}{1-\pi} &= 0 \\
\frac{s_n(1-\pi) - (n-s_n)\pi}{\pi(1-\pi)} &= 0 \qquad \text{il denominatore è ininfluente} \\
s_n - s_n\pi - n\pi + s_n\pi &= 0 \\
s_n - n\pi &= 0 \\
n\pi &= s_n \\
\hat{\pi} &= \frac{s_n}{n} \\
&= \frac{\sum_{i=1}^{n} x_i}{n}
\end{aligned}$$

Se $n = 5$, $s_5 = 3$ allora:

$$\hat{\pi} = \frac{3}{5} = 0.6$$

$L(\pi; s_5 = 3)$, $\ell(\pi; s_5 = 3)$.





12.7.1 Calcolo delle proprietà di $\hat{\pi}$

Dunque

Siano $X_1, ...X_n$ n VC IID, tali che $X_i \sim \text{Ber}(\pi)$ lo stimatore di massima verosimiglianza per π è

$$\hat{\pi} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Il *vero* valore di π è incognito ma sappiamo che:

$\hat{\pi}$ è corretto per π, infatti

$$E(\hat{\pi}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{\pi + ... + \pi}{n} = \frac{n}{n}\pi = \pi$$

E quindi

$$MSE(\hat{\pi}) = V(\hat{\pi}) = \frac{\pi(1-\pi)}{n}$$

che è ancora funzione di π.

Lo stimatore $\hat{\pi}$ per π è *consistente*, infatti

$$\lim_{n\to+\infty} MSE(\hat{\pi}) = \lim_{n\to+\infty} \frac{\pi(1-\pi)}{n} = 0$$

$\hat{\pi}$ è *corretto* e *consistente* per π.

Osserviamo che:

$$SE(\hat{\pi}) = \sqrt{\frac{\pi(1-\pi)}{n}}$$

È un risultato teorico che dipende dal *vero* π, che non conosciamo.

L'errore di stima si stima sostituendo a π la sua stima $\hat{\pi}$

$$\widehat{SE(\hat{\pi})} = \sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}$$

Se $\hat{\pi} = 0.6$ e $n = 5$

$$\widehat{SE(\hat{\pi})} = \sqrt{\frac{0.6(1-0.6)}{5}} = 0.2191$$

Nota

Lo Standard Error è l'ordine di grandezza dell'errore commesso.

12.7.2 Se n aumenta e $\hat{\pi} = 0.6$

Se $n = 10$ e $s_{10} = 6$, allora anche in questo caso

$$\hat{\pi} = \frac{6}{10} = 0.6$$

ma

$$\widehat{SE(\hat{\pi})} = \sqrt{\frac{0.6(1-0.6)}{10}} = 0.1549.$$

Se $n = 20$ e $s_{10} = 12$, allora anche in questo caso

$$\hat{\pi} = \frac{12}{20} = 0.6$$

e

$$\widehat{SE(\hat{\pi})} = \sqrt{\frac{0.6(1-0.6)}{20}} = 0.1095.$$

Se $n = 100$ e $s_{100} = 60$, allora anche in questo caso

$$\hat{\pi} = \frac{60}{100} = 0.6$$

e

$$\widehat{SE(\hat{\pi})} = \sqrt{\frac{0.6(1-0.6)}{100}} = 0.049.$$

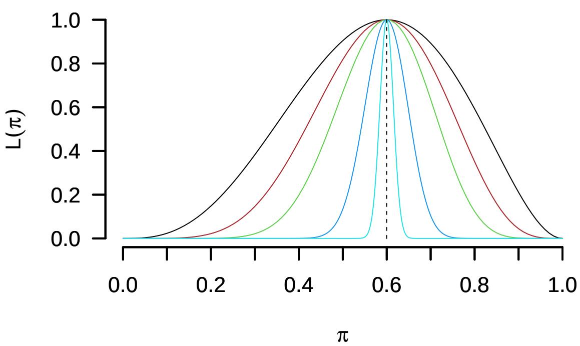
Se $n = 1000$ e $s_{10} = 600$, allora anche in questo caso

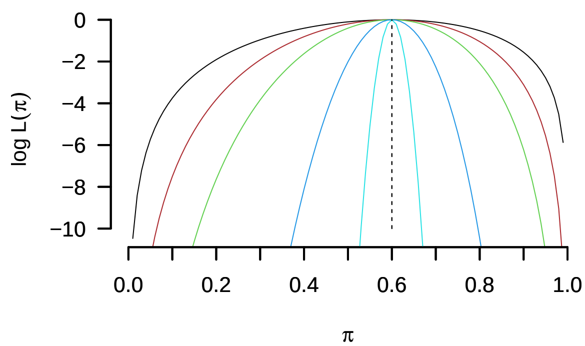
$$\hat{\pi} = \frac{600}{1\ 000} = 0.6$$

e

$$\widehat{SE(\hat{\pi})} = \sqrt{\frac{0.6(1-0.6)}{1000}} = 0.0155.$$

Osserviamo nel grafico $L(\pi; s_n = 0.6 \cdot n)$ e $\ell(\pi; s_n = 0.6 \cdot n)$ per $n = 5, 10, 20, 100, 1\ 000$





12.7.3 L'ipotesi $\pi = 0.5$

Se $n = 5$, $s_n = 3$ il valore $\hat{\pi} = 0.6$ è il 10.592% *più verosimile* di $\pi = 0.5$.

$$\frac{L(0.6; s_5 = 3)}{L(0.5; s_5 = 3)} = \frac{0.6^3(1-0.6)^2}{0.5^3(1-0.5)^2} = 1.1059$$

Se $n = 10$, $s_n = 6$ il valore $\hat{\pi} = 0.6$ è il 22.3059% *più verosimile* di $\pi = 0.5$.

$$\frac{L(0.6; s_5 = 3)}{L(0.5; s_5 = 3)} = \frac{0.6^6(1-0.6)^4}{0.5^6(1-0.5)^4} = 1.2231$$

Se $n = 20$, $s_n = 12$ il valore $\hat{\pi} = 0.6$ è il 49.5873% *più verosimile* di $\pi = 0.5$.

$$\frac{L(0.6; s_5 = 3)}{L(0.5; s_5 = 3)} = \frac{0.6^{12}(1-0.6)^8}{0.5^{12}(1-0.5)^8} = 1.4959$$

Se $n = 100$, $s_n = 60$ il valore $\hat{\pi} = 0.6$ è il 648.9869% *più verosimile* di $\pi = 0.5$.

$$\frac{L(0.6; s_5 = 3)}{L(0.5; s_5 = 3)} = \frac{0.6^{60}(1-0.6)^{40}}{0.5^{60}(1-0.5)^{40}} = 7.4899$$

Se $n = 1000$, $s_n = 600$ il valore $\hat{\pi} = 0.6$ è il 55557465413.0872% *più verosimile* di $\pi = 0.5$.

$$\frac{L(0.6; s_5 = 3)}{L(0.5; s_5 = 3)} = \frac{0.6^{600}(1-0.6)^{400}}{0.5^{600}(1-0.5)^{400}} = 555574655.1309$$

12.8 Il modello Poisson

Siano $X_1, ..., X_n$ n VC IID, replicazioni della stessa $X \sim \text{Pois}(\lambda)$, e dunque con funzione di probabilità:

$$f(x_i; \lambda) = \frac{\lambda^{x_i}}{x_i!} e^{-\lambda}$$

La verosimiglianza per λ è

$$\begin{aligned}
L(\lambda) &= \prod_{i=1}^{n} \frac{\lambda^{x_i}}{x_i!} e^{-\lambda} \\
&= \frac{\lambda^{x_1}}{x_1!} e^{-\lambda} \cdot \frac{\lambda^{x_2}}{x_2!} e^{-\lambda} \cdot \ldots \cdot \frac{\lambda^{x_n}}{x_n!} e^{-\lambda} \\
&= \frac{1}{x_1! x_2! ... x_n!} \; \lambda^{x_1} \lambda^{x_2} ... \lambda^{x_n} \; e^{-\lambda} e^{-\lambda} ... e^{-\lambda} \\
&= \frac{1}{\prod_{i=1}^{n} x_i!} \lambda^{x_1+\ldots+x_n} e^{-\lambda-\ldots-\lambda} \\
&\propto \lambda^{\sum_{i=1}^{n} x_i} e^{-n\lambda} \\
&\propto \lambda^{s_n} e^{-n\lambda}, \qquad s_n = \sum_{i=1}^{n} x_i
\end{aligned}$$

Tutta l'**informazione** sulla Poisson è contenuta nella statistica s_n.

12.8.1 La log-verosimiglianza della Poisson

Essendo

$$L(\lambda) \propto \lambda^{s_n} e^{-n\lambda}$$

Allora

$$\begin{aligned}
\ell(\lambda) &= \log \lambda^{s_n} e^{-n\lambda} \\
&= \log \lambda^{s_n} + \log e^{-n\lambda} \\
&= s_n \log \lambda - n\lambda, \qquad \text{in quanto } \log e^a = a
\end{aligned}$$

12.8.2 La stima di massima verosimiglianza della Poisson

Essendo

$$\ell(\lambda) = s_n \log \lambda - n\lambda$$

Allora

$$\ell'(\lambda) = \frac{s_n}{\lambda} - n$$

E dunque

$$\ell'(\lambda) \;=\; 0$$

$$\begin{aligned}
\frac{s_n}{\lambda} - n &= 0 \\
\frac{s_n}{\lambda} &= n \\
n\lambda &= s_n \\
\hat{\lambda} &= \frac{s_n}{n} \\
\hat{\lambda} &= \frac{1}{n}\sum_{i=1}^{n} x_i
\end{aligned}$$

12.8.3 Proprietà dello stimatore di massima verosimiglianza della Poisson $\hat{\lambda}$

Dunque

Siano $X_1, ...X_n$ n VC IID, tali che $X_i \sim \text{Pois}(\lambda)$ lo stimatore di massima verosimiglianza per π è

$$\hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Correttezza:

$$E(\hat{\lambda}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\sum_{i=1}^{n} \lambda = \lambda$$

Mean Squared Error:

$$MSE(\hat{\lambda}) = V(\hat{\lambda}) = V\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} V(X_i) = \frac{n}{n^2}\lambda = \frac{\lambda}{n}$$

Consistenza:

$$\lim_{n\to+\infty} MSE(\hat{\lambda}) = \lim_{n\to+\infty} \frac{\lambda}{n} = 0$$

Standard Error

$$SE(\hat{\lambda}) = \sqrt{\frac{\lambda}{n}}$$

Standard Error stimato

$$\widehat{SE(\hat{\lambda})} = \sqrt{\frac{\hat{\lambda}}{n}}$$

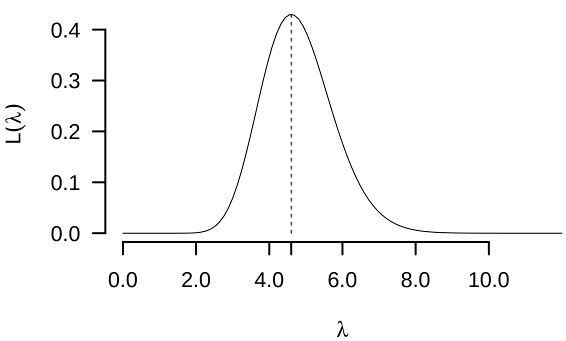
12.8.4 Esempio $n = 5$

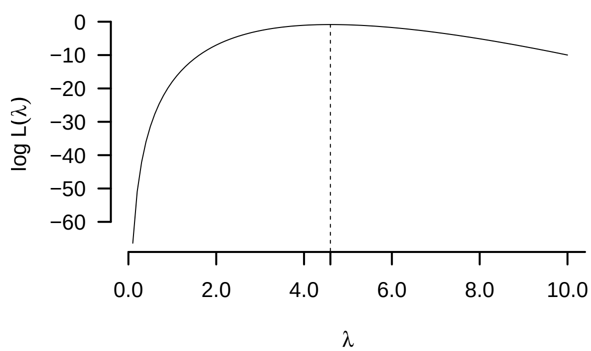
Il numero di clienti del negozio A è distribuito come una Poisson di parametro λ incognito. Dopo $n = 5$ giorni di osservazione si sono osservati i seguenti ingressi $(3, 4, 5, 8, 3)$. La stima $\hat{\lambda}$ di λ è

$$\hat{\lambda} = \frac{1}{5}23 = 4.6$$

Lo Standard Error stimato

$$\widehat{SE(\hat{\lambda})} = \sqrt{\frac{4.6}{5}} = 0.9592$$





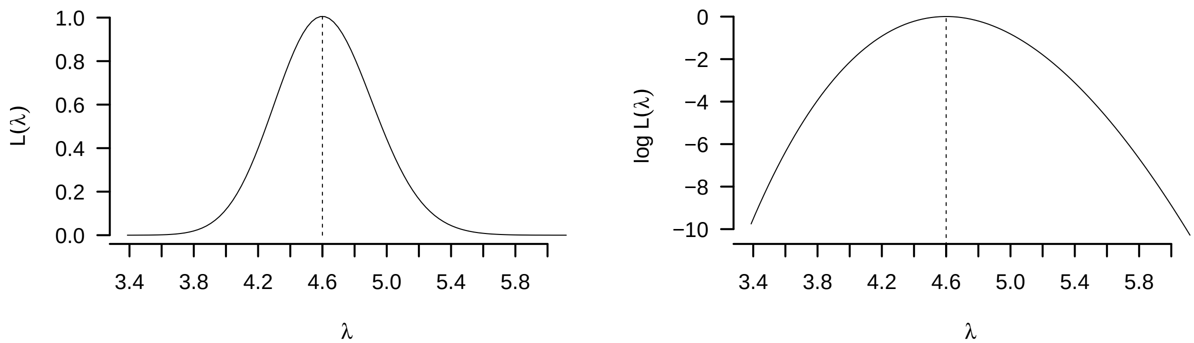
12.8.5 Esempio $n = 50$

Il numero di clienti del negozio A è distribuito come una Poisson di parametro λ incognito. Dopo $n = 50$ giorni di osservazione si è osservata una media di ingressi pari a ingressi (4.6). La stima $\hat{\lambda}$ di λ è

$$\hat{\lambda} = 4.6$$

Lo Standard Error stimato

$$\widehat{SE(\hat{\lambda})} = \sqrt{\frac{4.6}{50}} = 0.3033$$



12.9 Il modello Normale

Siano $X_1, ..., X_n$ n VC IID, replicazioni della stessa $X \sim N(\mu, \sigma^2)$, e dunque con funzione di probabilità:

$$f(x_i; \mu, \sigma^2)$$

La verosimiglianza per (μ, σ^2) è

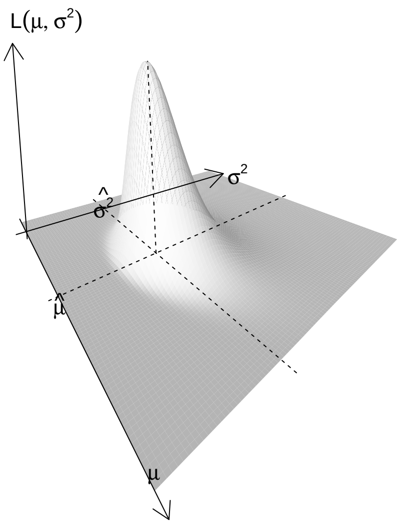
$$L(\lambda) = \prod_{i=1}^{n} f(x_i; \mu, \sigma^2)$$

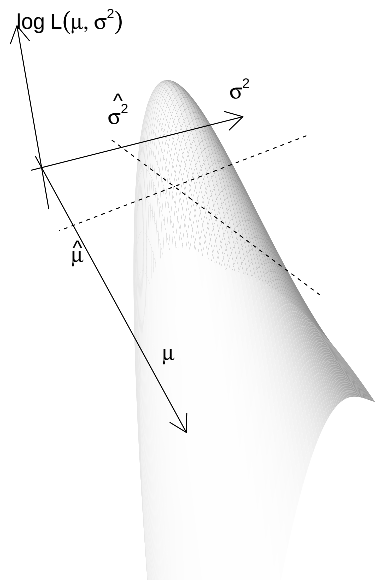
La log-verosimiglianza della Normale

Allora

$$\begin{aligned} \ell(\mu, \sigma^2) &= \log \prod_{i=1}^{n} f(x_i; \mu, \sigma^2) \\ &= \sum_{i=1}^{n} \log f(x_i; \mu, \sigma^2) \end{aligned}$$

12.9.1 Verosimiglianza e log-verosimiglianza della Normale





12.9.2 Le stime di massima verosimiglianza della Normale

Per ottenere $\hat{\mu}$ e $\hat{\sigma}^2$ bisogna eguagliare a zero il sistema di equazioni di derivate di $\ell(\mu, \sigma^2)$ rispetto a μ e σ^2

$$\begin{cases} \frac{d\ell(\mu,\sigma^2)}{d\mu} = 0 \\ \frac{d\ell(\mu,\sigma^2)}{d\sigma^2} = 0 \end{cases}$$

Proprietà 12.9.1.

$$\begin{aligned} \hat{\mu} &= \frac{1}{n}\sum_{i=1}^{n} x_i \\ \hat{\sigma}^2 &= \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\mu})^2 \end{aligned}$$

$$= \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \hat{\mu}^2$$

Tutta l'**informazione** del campione è contenuta nelle statistiche $\sum_{i=1}^{n} x_i$ e $\sum_{i=1}^{n} x_i^2$.

Dimostrazione.

$$\begin{aligned} L(\mu, \sigma^2; \mathbf{x}) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{(x_i-\mu)^2}{\sigma^2}\right)} \\ &= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2\sigma^2}((x_1-\mu)^2+(x_2-\mu)^2+\ldots+(x_n-\mu)^2)} \\ &\propto \sigma^{-2n} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\} \end{aligned}$$

$$\begin{aligned} \ell(\mu, \sigma^2) &= \log L(\mu, \sigma^2; \mathbf{x}) \\ &= \log \sigma^{-2n} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\} \\ &= -n \log \sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2 \end{aligned}$$

$$\begin{aligned} \frac{d\ell(\mu, \sigma^2)}{d\mu} &= +\frac{2}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu) \\ \frac{d\ell(\mu, \sigma^2)}{d\sigma^2} &= -\frac{n}{\sigma^2} + \frac{1}{(\sigma^2)^2}\sum_{i=1}^{n}(x_i-\mu)^2 \\ &= \frac{-n\sigma^2 + \sum_{i=1}^{n}(x_i-\mu)}{(\sigma^2)^2} \\ \frac{d\ell(\mu, \sigma^2)}{d\mu} &= 0 \Rightarrow \hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i \\ \frac{d\ell(\mu, \sigma^2)}{d\sigma^2} &= 0 \Rightarrow \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\hat{\mu})^2 \end{aligned}$$

□

12.9.3 Proprietà di $\hat{\mu}$

Correttezza per μ:

$$E(\hat{\mu}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\sum_{i=1}^{n} \mu = \mu$$

Mean Squared Error per μ:

$$MSE(\hat{\mu}) = V(\hat{\mu}) = V\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} V(X_i) = \frac{n}{n^2}\sigma^2 = \frac{\sigma^2}{n}$$

Consistenza per μ:

$$\lim_{n\to+\infty} MSE(\hat{\mu}) = \lim_{n\to+\infty} \frac{\sigma^2}{n} = 0$$

E lo Standard Error:

$$SE(\hat{\mu}) = \sqrt{\frac{\sigma^2}{n}}$$

Standard Error stimato tra poco verrà ricavato 12.9.5.

12.9.4 Proprietà di $\hat{\sigma}^2$

Correttezza per $\hat{\sigma}^2$:

$$E(\hat{\sigma}^2) = \frac{n-1}{n}\sigma^2$$

$\hat{\sigma}^2$ non è stimatore corretto per σ^2.

Correzione di $\hat{\sigma}^2$

$$S^2 = \frac{n}{n-1}\hat{\sigma}^2 = \frac{n}{n-1}\frac{1}{n}\sum_{i=1}^{n}(X_i - \hat{\mu})^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \hat{\mu})^2$$

Osserviamo che

$$E(S^2) = E\left(\frac{n}{n-1}\hat{\sigma}^2\right) = \frac{n}{n-1}E\left(\hat{\sigma}^2\right) = \frac{n}{n-1}\frac{n-1}{n}\sigma^2 = \sigma^2$$

S^2 è stimatore corretto per σ^2.

12.9.5 Lo SE di $\hat{\mu}$

Standard Error

$$SE(\hat{\mu}) = \sqrt{\frac{\sigma^2}{n}}$$

Standard Error stimato.

$$\widehat{SE(\hat{\mu})} = \sqrt{\frac{S^2}{n}} = \sqrt{\frac{\hat{\sigma}^2}{n-1}}$$

In quanto

$$\frac{S^2}{n} = \frac{1}{n}S^2 = \frac{1}{n}\frac{n}{n-1}\hat{\sigma}^2 = \frac{\hat{\sigma}^2}{n-1}$$

12.9.6 Esempio $n = 10$

Il fatturato mensile del negozio A è distribuito come una Normale di parametri μ e σ^2 incogniti. Dopo $n = 10$ mesi di osservazione si sono osservati i seguenti fatturati $(x_1 = 2.103, x_2 = 3.185, x_3 = 4.588, x_4 = 1.87, x_5 = 2.92, x_6 = 3.132, x_7 = 3.708, x_8 = 2.76, x_9 = 4.984, x_{10} = 2.861)$. La stima $\hat{\mu}$ di μ è

$$\hat{\mu} = \frac{1}{10}32.1115 = 3.2112$$

La varianza campionaria $\hat{\sigma}^2$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\mu})^2 = \frac{1}{n}\sum_{i=1}^{n}x_i^2 - \hat{\mu}^2 = \frac{111.8468}{10} - 3.2112^2 = 0.8732$$

S^2 la stima corretta di σ^2

$$S^2 = \frac{n}{n-1}\hat{\sigma}^2 = \frac{10}{9}0.8732 = 0.9702$$

Lo SE stimato di $\hat{\mu}$

$$\widehat{SE(\hat{\mu})} = \sqrt{\frac{0.9702}{10}} = 0.3115$$

12.9.7 Esempio $n = 100$

L'ammontare delle transazioni finanziarie compiute al minuto dal server A è distribuito come una Normale di parametri μ e σ^2 incogniti. Dopo $n = 100$ ore di osservazione si sono osservati $\bar{x} = 3.2112$, $\hat{\sigma} = 0.9344$. La stima $\hat{\mu}$ di μ è

$$\hat{\mu} = 3.2112$$

La varianza campionaria $\hat{\sigma}^2$

$$\hat{\sigma}^2 = 0.9344^2 = 0.8732$$

S^2 la stima corretta di σ^2

$$S^2 = \frac{n}{n-1}\hat{\sigma}^2 = \frac{100}{99}0.8732 = 0.882$$

Lo SE stimato di $\hat{\mu}$

$$\widehat{SE(\hat{\mu})} = \sqrt{\frac{0.882}{100}} = 0.0939$$

12.9.8 Perché $n - 1$

Per calcolare la varianza campionaria dobbiamo prima calcolare la media dei dati. Per calcolare la media bisogna sommare i dati, per esempio se $n = 3$: $x_1 = 7$, $x_2 = 8$ e $x_3 = 11$

$$x_1 + x_2 + x_3 = 26$$

Ma $x_1 = 7$, $x_2 = 8$ e $x_3 = 11$ non sono l'unica tripla di x che somma a 26, ma $n - 1 = 2$ valori possono essere scelti liberamente (es $x_1 = 5$ e $x_2 = 15$): Il terzo **è vincolato**:

$$x_3 = 26 - x_1 - x_2$$

Fissata la somma il sistema *ha perso un grado di libertà*.

12.10 Proprietà degli stimatori di massima verosimiglianza

Proprietà 12.10.1 (Stimatori di massima verosimgilianza). *Siano $X_1, ..., X_n$ n VC IID, replicazioni di $X \sim \mathscr{L}(\theta)$ e sia $\hat{\theta}$ lo stimatore di massima verosimiglianza per per θ, allora*

1. *$\hat{\theta}$ non è sempre stimatore corretto ma è sempre corretto asintoticamente:*

$$E(\hat{\theta}) \xrightarrow{n \to \infty} \theta$$

2. *$\hat{\theta}$ non è sempre stimatore a* massima efficienza *ma lo è sempre asintoticamente:*

$$V(\hat{\theta}) \xrightarrow{n \to \infty} I^{-1}(\theta)$$

   dove $I(\theta)$ è l'infromazione di Fisher.

3. *$\hat{\theta}$ è asintoticamente distribuito normalmente*

$$\hat{\theta} \underset{a}{\sim} N(\theta, I^{-1}(\theta))$$

4. *Lo stimatore di massima verosimiglianza è invariante alle trasformazioni monotone invertibili g:*

$$\text{se } \psi = g(\theta), \text{ allora } \hat{\psi} = g(\hat{\psi})$$

1. La proprietà uno riguarda la correttezza. Non sempre gli SMV sono corretti ma lo sono sempre asintoticamente. Esempio: lo stimatore $\hat{\sigma}^2$ di σ^2 non è corretto solo asintoticamente

$$E(\hat{\sigma}^2) = \frac{n-1}{n}\sigma^2 \xrightarrow{n \to \infty} \sigma^2$$

2. La proprietà due riguarda l'efficienza dello stimatore: non sempre lo SMV è il più efficiente per piccoli campioni, ma se il campione diventa grande, lo SMV è lo stimatore che raggiunge la varianza minima. La varianza minima è chiamata Informazione di Fisher ed è indicata con $I^{-1}(\theta)$:
$$I(\theta) = -E\left(\ell''(\theta)\right)$$
dove $\ell''(\theta)$ è la derivata seconda della log verosimiglianza calcolata in θ.

- $I(\theta)$ è la curvatura media della log verosimiglianza intorno al punto θ.
- $I^{-1}(\theta)$ è un risultato teorico ed un limite sotto al quale nessuno stimatore può scendere.
- Se esiste lo stimatore più efficiente allora è quello di *massima verosimiglianza*.
- Esempio: $\hat{\pi}$, $\hat{\lambda}$ e $\hat{\mu}$ sono stimatori a efficienza massima.

$$\begin{aligned}
I^{-1}(\pi) &= \frac{\pi(1-\pi)}{n} \\
I^{-1}(\lambda) &= \frac{\lambda}{n} \\
I^{-1}(\mu) &= \frac{\sigma^2}{n}
\end{aligned}$$

3. La proprietà tre ci garantisce che, per n sufficientemente alto, sappiamo la distribuzione degli SMV

- Esempio: lo stimatore $\hat{\pi}$ di π, dal TLC
$$\hat{\pi} \underset{a}{\sim} N\left(\pi, \frac{\pi(1-\pi)}{n}\right)$$
- Esempio: lo stimatore $\hat{\lambda}$ di λ, dal TLC
$$\hat{\lambda} \underset{a}{\sim} N\left(\lambda, \frac{\lambda}{n}\right)$$

4. La proprietà 4 garantisce che trasformazioni invertibili dei parametri non richiedono di ricalcolare la SMV.

- Esempio: $\sigma = \sqrt{\sigma^2}$ e dunque $\hat{\sigma} = \sqrt{\hat{\sigma}^2}$

Stima Intervallare

13

13.1 Obiettivo

Oltre alla stima di un punto specifico $\hat{\theta}$ dello spazio dei parametri potremmo essere interessati a trovare *regioni più verosimili*.

L'obiettivo è di stimare un intervallo nel quale pensiamo *verosimilmente* si collochi il vero θ alla luce dei dati.

Nel gergo dei sondaggisti viene chiamata forbice.

La teoria della verosimiglianza offre tutti gli strumenti per la derivazione coerente di intervalli per una classe molto ampia di modelli di probabilità. La trattazione sistematica attraverso la verosimiglianza esula dagli scopi di questo corso.

13.2 Il Contesto Probabilistico

Siano $X_1, ...X_5$, $n = 5$ VC IID, $X_i \sim N(\mu = 2.5, \sigma^2 = 2.25)$

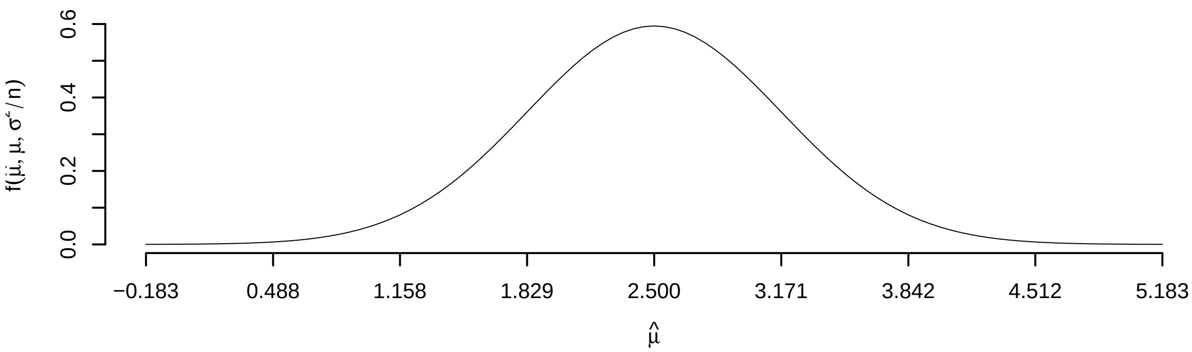
Dalle proprietà della normale

$$\frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X} = \hat{\mu} \sim N\left(\mu, \frac{\sigma^2}{n}\right), \qquad \text{ovvero } \hat{\mu} \sim N\left(\mu, SE^2(\hat{\mu})\right)$$

In questo caso

$$\begin{aligned}
\hat{\mu} &\sim N\left(2.5, \frac{2.25}{5}\right) \\
&\sim N\left(2.5, \left(\frac{1.5}{\sqrt{5}}\right)^2\right) \\
&\sim N\left(2.5, 0.6708^2\right)
\end{aligned}$$

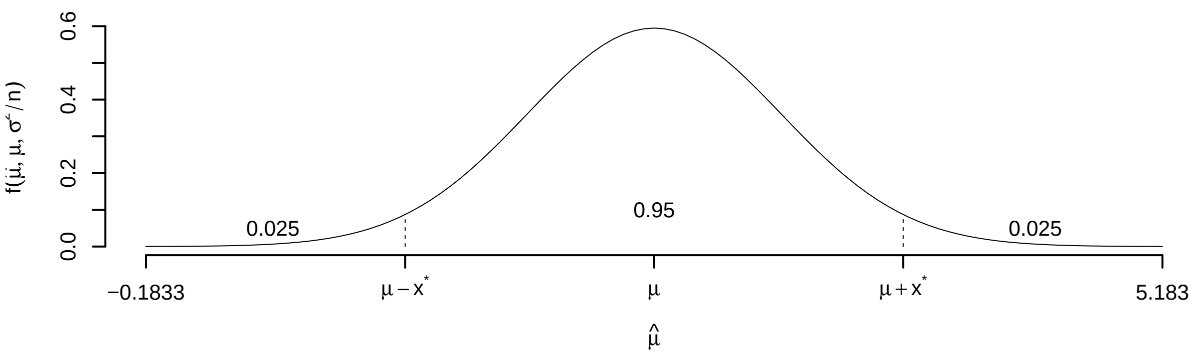
La densità di probabilità di $\hat{\mu}$.



13.2.1 Un intervallo per $\hat{\mu}$

Domanda: qual è quel valore $x^* > 0$ tale che

$$P(\mu - x^* < \hat{\mu} < \mu + x^*) = 0.95 \quad ?$$



qual è quel numero x^* tale che

$$P(\hat{\mu} > \mu + x^*) = 0.025, \qquad P(\hat{\mu} < \mu - x^*) = 0.025 \quad ?$$

Trasferiamo il problema sulla normale standard

Se $Z \sim N(0, 1)$

Qual è quel numero z^* tale che

$$P(Z > z^*) = 0.025 \quad ?$$

Dall'ultima riga delle tavole osserviamo che $z^* = 1.96$, infatti

$$P(Z > 1.96) = 1 - \Phi(1.96) = 1 - 0.975 = 0.025$$

E dunque

$$P(-19.6 < Z < +1.96) = 0.95$$

Ci serviamo delle tavole della Z per $\hat{\mu}$, ricordiamo che $\hat{\mu} \sim N(\mu, SE^2(\hat{\mu}))$, allora

$$Z = \frac{\hat{\mu} - \mu}{SE(\hat{\mu})} \sim N(0,1)$$

e quindi

$$\begin{aligned}
P(-1.96 < Z < +1.96) &= 0.95 \\
P\left(-1.96 < \frac{\hat{\mu} - \mu}{SE(\hat{\mu})} < +1.96\right) &= 0.95 \\
P\left(-1.96\ SE(\hat{\mu}) < \hat{\mu} - \mu < +1.96\ SE(\hat{\mu})\right) &= 0.95 \\
P\left(\mu - 1.96\ SE(\hat{\mu}) < \mu + \hat{\mu} - \mu < \mu + 1.96\ SE(\hat{\mu})\right) &= 0.95 \\
P(\mu - 1.96\ SE(\hat{\mu}) < \hat{\mu} < \mu + 1.96\ SE(\hat{\mu})) &= 0.95
\end{aligned}$$

Numericamente

$$\begin{aligned}
P\left(2.5 - 1.96\ \frac{1.5}{\sqrt{5}} < \hat{\mu} < 2.5 + 1.96\ \frac{1.5}{\sqrt{5}}\right) &= 0.95 \\
P\left(2.5 - 1.96 \times 0.6708 < \hat{\mu} < 2.5 + 1.96 \times 0.6708\right) &= 0.95 \\
P(1.1852 < \hat{\mu} < 3.8148) &= 0.95
\end{aligned}$$

13.2.2 n e σ^2 rimangono fissi, cambiamo μ

Se $n = 5$ e $\sigma^2 = 2.25$, ma per esempio $\mu = 1.2$ allora

$$\begin{aligned}
P(\mu - 1.96\ SE(\hat{\mu}) < \hat{\mu} < \mu + 1.96\ SE(\hat{\mu})) &= 0.95 \\
P\left(\mu - 1.96\ \frac{\sigma}{\sqrt{n}} < \hat{\mu} < \mu + 1.96\ \frac{\sigma}{\sqrt{n}}\right) &= 0.95 \\
P\left(1.2 - 1.96\ \frac{1.5}{\sqrt{5}} < \hat{\mu} < 1.2 + 1.96\ \frac{1.5}{\sqrt{5}}\right) &= 0.95 \\
P\left(1.2 - 1.96 \times 0.6708 < \hat{\mu} < 1.2 + 1.96 \times 0.6708\right) &= 0.95 \\
P(-0.1148 < \hat{\mu} < 2.5148) &= 0.95
\end{aligned}$$

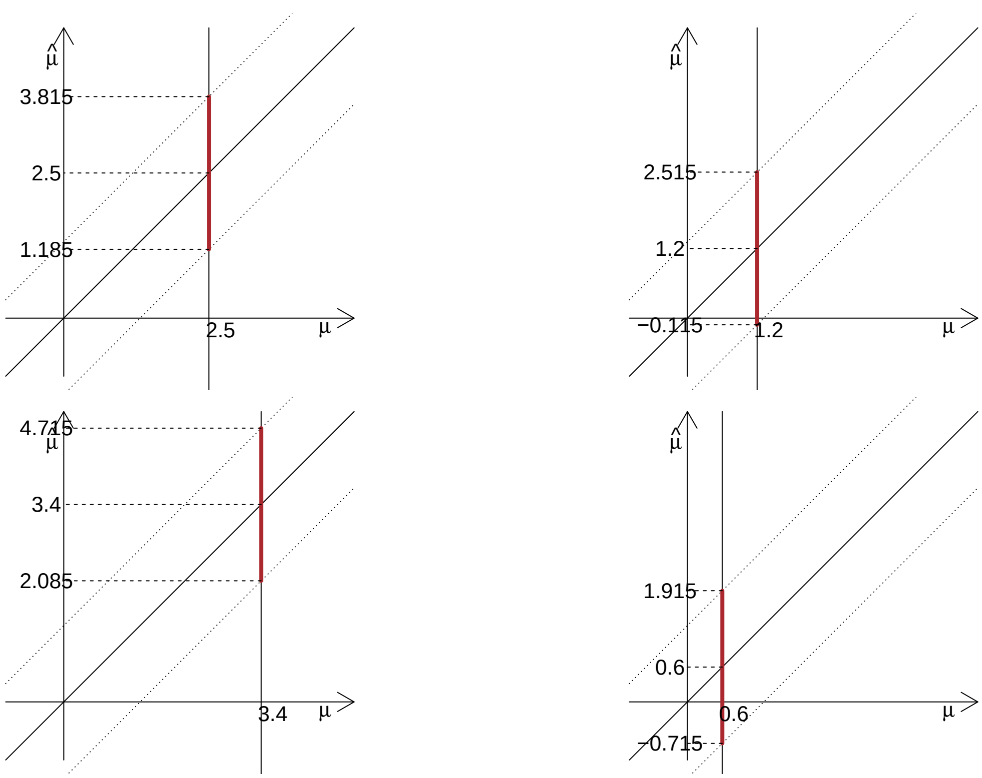
Se $n = 5$ e $\sigma^2 = 2.25$, ma per esempio $\mu = 3.4$ allora

$$\begin{aligned}
P\left(3.4 - 1.96\ \frac{1.5}{\sqrt{5}} < \hat{\mu} < 3.4 + 1.96\ \frac{1.5}{\sqrt{5}}\right) &= 0.95 \\
P\left(3.4 - 1.96 \times 0.6708 < \hat{\mu} < 3.4 + 1.96 \times 0.6708\right) &= 0.95 \\
P(2.0852 < \hat{\mu} < 4.7148) &= 0.95
\end{aligned}$$

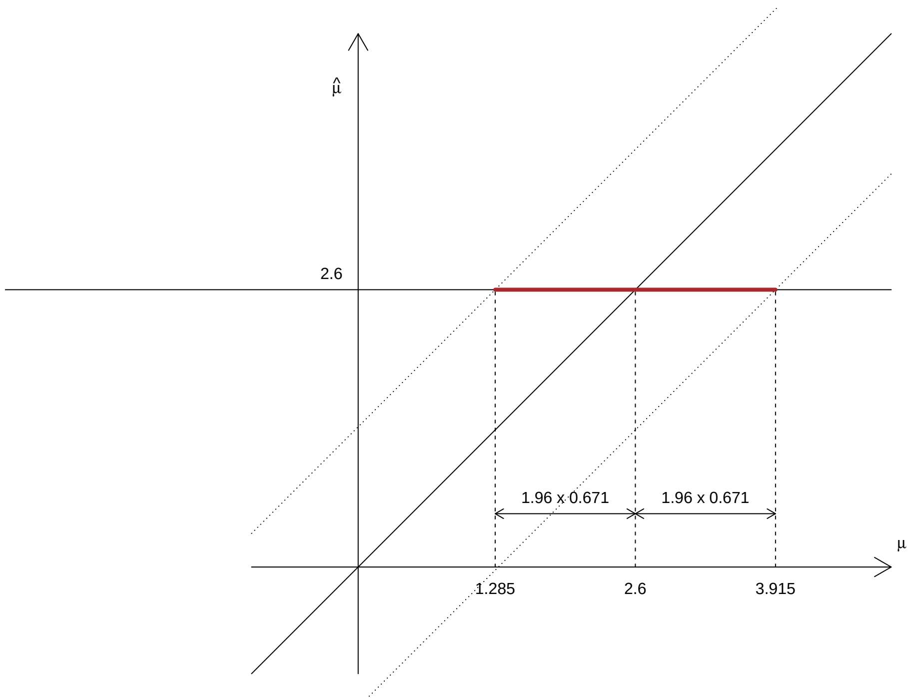
Se $n = 5$ e $\sigma^2 = 2.25$, ma per esempio $\mu = 0.6$ allora

$$\begin{aligned} P\left(0.6 - 1.96 \times 0.6708 < \hat{\mu} < 0.6 + 1.96 \times 0.6708\right) &= 0.95 \\ P(-0.7148 < \hat{\mu} < 1.9148) &= 095 \end{aligned}$$

Rimangono fissi n e σ^2 , cambiamo μ



13.2.3 n e σ^2 rimangono fissi e noti, μ incognita $\hat{\mu} = 2.6$



Algebricamente, osserviamo

$$
\begin{aligned}
P(-1.96 < Z < +1.96) &= 0.95 \\
P\left(-1.96 < \frac{\hat{\mu} - \mu}{SE(\hat{\mu})} < +1.96\right) &= 0.95 \\
P\left(-1.96\ SE(\hat{\mu}) < \hat{\mu} - \mu < +1.96\ SE(\hat{\mu})\right) &= 0.95 \\
P\left(-1.96\ \frac{\sigma}{\sqrt{n}} < \hat{\mu} - \mu < \mu + 1.96\ \frac{\sigma}{\sqrt{n}}\right) &= 0.95 \\
P\left(-\hat{\mu} - 1.96\ \frac{\sigma}{\sqrt{n}} < -\hat{\mu} + \hat{\mu} - \mu < -\hat{\mu} + 1.96\ \frac{\sigma}{\sqrt{n}}\right) &= 0.95 \qquad \text{sottraggo } \hat{\mu} \\
P\left(-\hat{\mu} - 1.96\ \frac{\sigma}{\sqrt{n}} < -\mu < -\hat{\mu} + 1.96\ \frac{\sigma}{\sqrt{n}}\right) &= 0.95 \\
\left(+\hat{\mu} + 1.96\ \frac{\sigma}{\sqrt{n}} > +\mu > +\hat{\mu} - 1.96\ \frac{\sigma}{\sqrt{n}}\right) &= 0.95 \qquad \text{cambio segno e verso} \\
P\left(\hat{\mu} - 1.96\ \frac{\sigma}{\sqrt{n}} < \mu < \hat{\mu} + 1.96\ \frac{\sigma}{\sqrt{n}}\right) &= 0.95
\end{aligned}
$$

13.3 Intervalli casuali

Attenzione

$$P\left(\hat{\mu}-1.96\,\frac{\sigma}{\sqrt{n}}<\mu<\hat{\mu}+1.96\,\frac{\sigma}{\sqrt{n}}\right)=0.95$$

non è la probabilità che μ si trovi tra $\hat{\mu}-1.96\,\frac{\sigma}{\sqrt{n}}$ e $\hat{\mu}+1.96\,\frac{\sigma}{\sqrt{n}}$,

$$P\left(\hat{\mu}-1.96\,\frac{\sigma}{\sqrt{n}}<\mu<\hat{\mu}+1.96\,\frac{\sigma}{\sqrt{n}}\right)=0.95$$

è la probabilità che **l'intervallo casuale** $\left[\hat{\mu}-1.96\,\frac{\sigma}{\sqrt{n}},\hat{\mu}+1.96\,\frac{\sigma}{\sqrt{n}}\right]$ cada su μ.

Se quindi $\hat{\mu}=2.6$ l'intervallo

$$\begin{aligned}\left[\hat{\mu}-1.96\,\frac{\sigma}{\sqrt{n}},\hat{\mu}+1.96\,\frac{\sigma}{\sqrt{n}}\right] &= \left[2.6-1.96\,\frac{1.5}{\sqrt{5}},\hat{\mu}+1.96\,\frac{1.5}{\sqrt{5}}\right]\\ &= [1.2852,3.9148]\end{aligned}$$

L'intervallo $[1.2852, 3.9148]$ è una realizzazione dell'intervallo casuale $\left[\hat{\mu}-1.96\,\frac{\sigma}{\sqrt{n}},\hat{\mu}+1.96\,\frac{\sigma}{\sqrt{n}}\right]$.

13.4 Intervallo di confidenza per μ al 95%, $n=5$ e $\sigma^2=2.25$.

Se $\hat{\mu}=2.6$ l'intervallo

$$\begin{aligned}\left[\hat{\mu}-1.96\,\frac{\sigma}{\sqrt{n}},\hat{\mu}+1.96\,\frac{\sigma}{\sqrt{n}}\right] &= \left[2.6-1.96\,\frac{1.5}{\sqrt{5}},\hat{\mu}+1.96\,\frac{1.5}{\sqrt{5}}\right]\\ &= [1.2852,3.9148]\end{aligned}$$

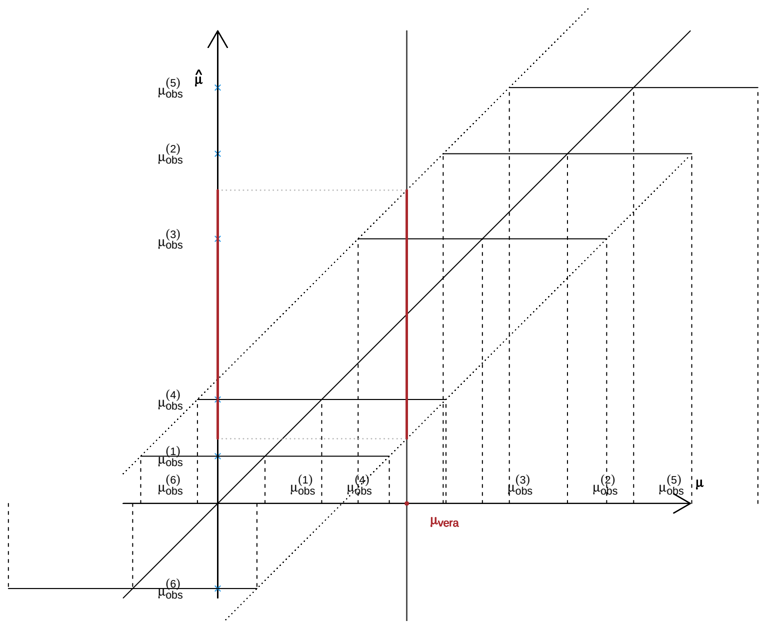
è chiamato *intervallo di confidenza* per μ al 95%.

Nota

Infelice traduzione di *Confidence Interval*,

- *to be confident that the real parameter is in* $[1.2852, 3.9148]$, *with a confidence level of 95%.*
- *Siamo fiduciosi che il vero parametro si trovi in* $[1.2852, 3.9148]$, *con un livello di fiducia del 95%*

L'intervallo è costruito con una metodologia che 95 volte su 100 produce intervalli che coprono il vero μ. Il long run:



13.5 Stimatori e intervalli di confidenza

Se θ è il parametro da stimare uno stimatore puntuale h è una VC

$$h(X_1, ..., X_n) = h = \hat{\theta}$$

Siano $L_1(X_1, ..., X_n) = L_1$ e $L_2(X_1, ..., X_n) = L_2$ due statistiche tali che $L_1 \leq L_2$ per ogni campione $X_1, ..., X_n$. L'intervallo

$$[L_1, L_2]$$

è un *intervallo casuale*.

Definizione 13.5.1. Un **intervallo di confidenza** per θ al livello $(1-\alpha) \times 100\%$ è costruito su quella coppia di statistiche L_1 e L_2 tali che

$$P(L_1 < \theta < L_2) = 1 - \alpha$$

Un *intervallo di confidenza* per θ al livello $(1-\alpha) \times 100\%$ è l'intervallo $[L_1, L_2]$ calcolato sui dati del campione.

13.6 Massima Verosimiglianza e intervalli di confidenza

Se $\hat{\theta}$ è lo stimatore di massima verosimiglianza per θ, se n è sufficientemente alto

$$\hat{\theta} \underset{a}{\sim} N(\theta, \widehat{SE^2(\hat{\theta})} \equiv I^{-1}(\theta))$$

L'*intervallo di confidenza* per θ al livello $(1-\alpha) \times 100\%$ è ricavato da:

$$P(\hat{\theta} - z_{\alpha/2}\widehat{SE(\hat{\theta})} < \theta < \hat{\theta} + z_{\alpha/2}\widehat{SE(\hat{\theta})}) = 1 - \alpha$$

Un *intervallo di confidenza* per θ al livello $(1-\alpha) \times 100\%$ è l'intervallo $[\hat{\theta} - z_{\alpha/2}\widehat{SE(\hat{\theta})}, \hat{\theta} + z_{\alpha/2}\widehat{SE(\hat{\theta})}]$ calcolato sui dati del campione.

13.7 Intervalli di Confidenza per μ al livello $(1-\alpha) \times 100$, σ^2 nota

Sia $0 < \alpha < 1$, osserviamo che

$$\begin{aligned} P(-z_{\alpha/2} < Z < +z_{\alpha/2}) &= 1 - \alpha \\ P\left(-z_{\alpha/2} < \frac{\hat{\mu} - \mu}{SE(\hat{\mu})} < +z_{\alpha/2}\right) &= 1 - \alpha \\ P\left(\hat{\mu} - z_{\alpha/2}\, \frac{\sigma}{\sqrt{n}} < \mu < \hat{\mu} + z_{\alpha/2}\, \frac{\sigma}{\sqrt{n}}\right) &= 1 - \alpha \end{aligned}$$

Definizione 13.7.1 (Intervallo di Confidenza per μ (σ^2 nota))**.** Si definisce L'IdC al livello $(1 - \alpha) \times 100\%$ per μ con σ^2 nota, l'intervallo

$$IdC: \quad \left[\hat{\mu} - z_{\alpha/2}\, \frac{\sigma}{\sqrt{n}}, \hat{\mu} + z_{\alpha/2}\, \frac{\sigma}{\sqrt{n}}\right]$$

Dove $z_{\alpha/2}$ è quel valore tale che $P(Z > z_{\alpha/2}) = \alpha/2$

$$P\left(\hat{\mu} - z_{\alpha/2}\, \frac{\sigma}{\sqrt{n}} < \mu < \hat{\mu} + z_{\alpha/2}\, \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

è la probabilità che **l'intervallo casuale** $\left[\hat{\mu} - z_{\alpha/2}\, \frac{\sigma}{\sqrt{n}}, \hat{\mu} + z_{\alpha/2}\, \frac{\sigma}{\sqrt{n}}\right]$ cada su μ.

Il valore $0 < \alpha < 1$, può essere qualunque ma solitamente si usa

- $\alpha = 0.05$: intervalli al $(1 - 0.05) \times 100\% = 0.95 \times 100\% = 95\%$
- $\alpha = 0.01$: intervalli al $(1 - 0.01) \times 100\% = 0.99 \times 100\% = 99\%$
- Gli intervalli per $\alpha = 0.1$ e $\alpha = 0.001$, intervalli al 90% e al 99.9% meno usati.

dalle tavole abbiamo

α	$\alpha/2$	$z_{\alpha/2}$		
$\alpha = 0.1$	con $\alpha/2 = 0.05$	e quindi $z_{\alpha/2} = z_{0.05}$	$= 1.6449$	Raro
$\alpha = 0.05$	con $\alpha/2 = 0.025$	e quindi $z_{\alpha/2} = z_{0.025}$	$= 1.96$	Freq.
$\alpha = 0.01$	con $\alpha/2 = 0.005$	e quindi $z_{\alpha/2} = z_{0.005}$	$= 2.5758$	Freq.
$\alpha = 0.001$	con $\alpha/2 = 0.0005$	e quindi $z_{\alpha/2} = z_{0.005}$	$= 3.2905$	Raro

Esempio 13.7.1. Intervallo al 99%, $\hat{\mu} = 2.6$, $n = 5$, $\sigma^2 = 2.25$. L'intervallo al 99% implica un $\alpha = 0.01$, infatti

$$1 - \alpha = 0.99$$

E dunque

$$\alpha/2 = 0.005$$

Dalle tavole

$$z_{\alpha/2} = z_{0.005} = 2.5758$$

Se quindi $\hat{\mu} = 2.6$ l'intervallo

$$\begin{aligned}\left[\hat{\mu} - 2.5758\,\frac{\sigma}{\sqrt{n}}, \hat{\mu} + 2.5758\,\frac{\sigma}{\sqrt{n}}\right] &= \left[2.6 - 2.5758\,\frac{1.5}{\sqrt{5}}, 2.6 + 2.5758\,\frac{1.5}{\sqrt{5}}\right] \\ &= [0.8721, 4.3279]\end{aligned}$$

È l'intervallo di confidenza per μ al 99%

Esempio 13.7.2. Intervallo al 99%, $\hat{\mu} = 2.6$, $n = 5$, $\sigma^2 = 2.25$. L'intervallo al 90% implica un $\alpha = 0.1$, infatti

$$1 - \alpha = 0.90$$

E dunque

$$\alpha/2 = 0.05$$

Dalle tavole

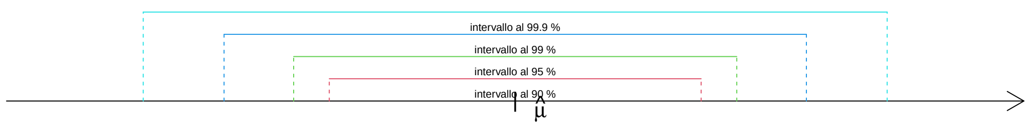
$$z_{\alpha/2} = z_{0.05} = 1.6449$$

Se quindi $\hat{\mu} = 2.6$ l'intervallo

$$\begin{aligned}\left[\hat{\mu} - 1.6449\,\frac{\sigma}{\sqrt{n}}, \hat{\mu} + 1.6449\,\frac{\sigma}{\sqrt{n}}\right] &= \left[2.6 - 1.6449\,\frac{1.5}{\sqrt{5}}, 2.6 + 1.6449\,\frac{1.5}{\sqrt{5}}\right] \\ &= [1.4966, 3.7034]\end{aligned}$$

È l'intervallo di confidenza per μ al 90%.

Osserviamo graficamente gli intervalli di confidenza per μ con $\hat{\mu} = 2.6$, $n = 5$ e $\sigma^2 = 2.25$, al 90%, 95%, 99% e 99.9%.



	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.001$
$\hat{\mu} - z_{\alpha/2} SE(\hat{\mu})$	1.497	1.285	0.8721	0.3926
$\hat{\mu} + z_{\alpha/2} SE(\hat{\mu})$	3.703	3.915	4.3279	4.8074
Ampiezza	2.207	2.630	3.4558	4.4147

Esempio 13.7.3. Un venditore di bustine di tè assicura che ogni bustina ha un peso medio pari a 20g con una SD pari a 1.5g. L'acquirente esegue 6 misure di controllo e ottiene i seguenti risultati: 19; 20; 20.5; 21; 18.5; 15. La media di questi numeri è $\hat{\mu} = 19$ e $\hat{\sigma} = 1.979$. Determinare un IdC al livello di $(1 - \alpha) = 0.99$ per μ.

$$\begin{aligned}
& \left[\hat{\mu} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}; \hat{\mu} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right] \\
= & \left[19 - 2.576\frac{1.5}{\sqrt{6}}; 19 + 2.576\frac{1.5}{\sqrt{6}}\right] \\
= & \ [19 - 1.5775; 19 + 1.5775] \\
= & \ [17.4225; \quad 20.5775]
\end{aligned}$$

Notare la modalità di indicare la SD della P e dei dati. Vi sono due informazioni su σ e bisogna scegliere quella giusta.

13.8 Intervalli di Confidenza per μ al livello $(1-\alpha) \times 100$, σ^2 incognita

Se σ^2 è incognito va stimato dai dati

Consideriamo lo stimatore S^2 di σ^2

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \hat{\mu})^2 = \frac{n}{n-1}\frac{1}{n}\sum_{i=1}^{n}(X_i - \hat{\mu})^2 = \frac{n}{n-1}\hat{\sigma}^2$$

Ricordiamo che

$$\widehat{SE(\hat{\mu})} = \sqrt{\frac{S^2}{n}} = \frac{S}{\sqrt{n}}$$

Ricordiamo infine che

$$T = \frac{\hat{\mu}-\mu}{\widehat{SE(\hat{\mu})}} \sim t_{n-1}$$

Ovvero

$$T = \frac{\hat{\mu}-\mu}{S/\sqrt{n}} \sim t_{n-1}$$

13.8.1 σ nota e σ incognita

Se σ è nota

$$Z = \frac{\hat{\mu}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

Se σ è incognita

$$T = \frac{\hat{\mu}-\mu}{S/\sqrt{n}} \sim t_{n-1}$$

Sia $0<\alpha<1$, osserviamo che

$$\begin{aligned}
P(-t_{n-1;\alpha/2} < T < +t_{n-1;\alpha/2}) &= 1-\alpha \\
P\left(-t_{n-1;\alpha/2} < \frac{\hat{\mu}-\mu}{\widehat{SE(\hat{\mu})}} < +t_{n-1;\alpha/2}\right) &= 1-\alpha \\
P\left(\hat{\mu} - t_{n-1;\alpha/2}\,\frac{S}{\sqrt{n}} < \mu < \hat{\mu} + t_{n-1;\alpha/2}\,\frac{S}{\sqrt{n}}\right) &= 1-\alpha
\end{aligned}$$

Dove $t_{n-1;\alpha/2}$ è quel valore tale che $P(T > t_{n-1;\alpha/2}) = \alpha/2, \qquad T \sim t_{n-1}$

Definizione 13.8.1 (Intervallo di Confidenza per μ (σ^2 incognita))**.** Si definisce L'IdC al livello $(1-\alpha)\times 100\%$ per μ con σ^2 incognita, l'intervallo

$$IdC: \left[\hat{\mu} - t_{n-1;\alpha/2}\,\frac{S}{\sqrt{n}}, \hat{\mu} + t_{n-1;\alpha/2}\,\frac{S}{\sqrt{n}}\right]$$

$$P\left(\hat{\mu} - t_{n-1;\alpha/2}\,\frac{S}{\sqrt{n}} < \mu < \hat{\mu} + t_{n-1;\alpha/2}\,\frac{S}{\sqrt{n}}\right) = 1-\alpha$$

è la probabilità che **l'intervallo casuale** $\left[\hat{\mu} - t_{n-1;\alpha/2}\,\frac{S}{\sqrt{n}}, \hat{\mu} + t_{n-1;\alpha/2}\,\frac{S}{\sqrt{n}}\right]$ cada su μ.

Esempio 13.8.1. Un venditore di bustine di tè assicura che ogni bustina ha un peso medio pari a 20g. L'acquirente esegue 6 misure di controllo e ottiene i seguenti risultati: 19; 20; 20.5; 21; 18.5; 15. La media di questi numeri è $\hat{\mu} = 19$ e $\hat{\sigma} = 1.979$. Determinare un IdC al livello di $(1-\alpha) = 0.99$ per μ.

$$\alpha = 0.01 \qquad \alpha/2 = 0.005, \qquad t_{6-1;0.005} = 4.0321$$

$$\begin{aligned}
& S = \sqrt{\frac{n}{n-1}}\hat{\sigma} = \sqrt{\frac{6}{6-1}}1.979 = 2.1679 \\
& \left[\hat{\mu} - t_{(n-1);\alpha/2}\frac{S}{\sqrt{n}}; \hat{\mu} + t_{(n-1);\alpha/2}\frac{S}{\sqrt{n}}\right] \\
= & \left[19 - 4.0321\frac{2.1679}{\sqrt{6}}; 19 + 4.0321\frac{2.1679}{\sqrt{6}}\right] \\
= & [19 - 3.5687;\ 19 + 3.5687] \\
= & [15.4313; 22.5687]
\end{aligned}$$

Si noti che la mancanza di informazioni sulla varianza rende più incerto il risultato; infatti, la lunghezza dell'IdC aumenta.

Esempio 13.8.2. Si sono rilevati i tempi dedicati a ciascun cliente da un impiegato di banca in 49 casi e si è ottenuta una media $\bar{x} = 3.5$ minuti con una SD pari a 0.5 minuti. Determinare un IdC al livello di $(1-\alpha) = 0.95$ per μ.

$$\alpha = 0.05 \qquad \alpha/2 = 0.025. \qquad t_{49-1;0.025} = 2.0106$$

$$\begin{aligned}
& s = \sqrt{\frac{n}{n-1}}\hat{\sigma} = \sqrt{\frac{49}{49-1}}0.5 = 0.505 \\
& \left[\bar{X} - t_{(n-1);\alpha/2}\frac{S}{\sqrt{n}}; \bar{X} + t_{(n-1);\alpha/2}\frac{S}{\sqrt{n}}\right] \\
= & \left[3.5 - 2.0106\frac{0.505}{\sqrt{49}}; 3.5 + 2.0106\frac{0.505}{\sqrt{49}}\right] \\
= & [3.5 - 0.145; 3.5 + 0.145] \\
= & [3.355; 3.645]
\end{aligned}$$

Si noti che per $n > 120$ si può approssimare con una normale.

13.9 IDC per la proporzione

$X_1, \dots, X_n$ VC IID, tutte Ber(π). Per il TLC

$$\bar{X} = \hat{\pi} \underset{a}{\sim} N\left(\pi; \frac{\pi(1-\pi)}{n}\right) \Rightarrow \hat{\pi} \underset{a}{\sim} N\left(\pi; SE^2(\hat{\pi})\right)$$

E quindi

$$Z = \frac{\hat{\pi} - \pi}{SE(\hat{\pi})} \underset{a}{\sim} N(0,1) \Rightarrow P\left(-z_{\alpha/2} < \frac{\hat{\pi} - \pi}{SE(\hat{\pi})} < z_{\alpha/2}\right) = 1 - \alpha$$

Infine

$$\begin{aligned} P\left(\hat{\pi} - z_{\alpha/2} SE(\hat{\pi}) < \pi < \hat{\pi} + z_{\alpha/2} SE(\hat{\pi})\right) &= 1 - \alpha \\ P\left(\hat{\pi} - z_{\alpha/2}\sqrt{\frac{\pi(1-\pi)}{n}} < \pi < \hat{\pi} + z_{\alpha/2}\sqrt{\frac{\pi(1-\pi)}{n}}\right) &= 1 - \alpha \end{aligned}$$

l'IdC DIPENDE da π NON NOTA. Per il TLC e n sufficientemente grande si può sostituire a π dell'IdC la sua stima $\hat{\pi}$.

$$\widehat{SE(\hat{\pi})} = \sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}$$

Condizioni per l'approssimazione

$$n\,\pi \geq 5 \quad \text{e} \quad n\,(1-\pi) \geq 5$$

e quindi l'IdC al livello $(1-\alpha) \times 100$ per $\hat{\pi}$ è:

Definizione 13.9.1 (Intervallo di Confidenza per π). Si definisce L'IdC al livello $(1-\alpha) \times 100\%$ per π l'intervallo

$$\left[\hat{\pi} - z_{\alpha/2}\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}; \hat{\pi} + z_{\alpha/2}\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}\right]$$

Esempio 13.9.1. Una indagine sulle intenzioni di voto degli italiani per lo schieramento A ha mostrato che 240 su 500 lo voterebbero. Determinare un IdC al livello di $(1-\alpha) = 0.99$ per π.

$$\alpha = 0.01, \qquad \alpha/2 = 0.005, \qquad z_{0.005} = 2.5758$$

$$\hat{\pi} = \frac{\text{favorevoli}}{n} = \frac{240}{500} = 0.48$$

$$\left[\hat{\pi} - z_{\alpha/2}\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}; \hat{\pi} + z_{\alpha/2}\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}\right]$$

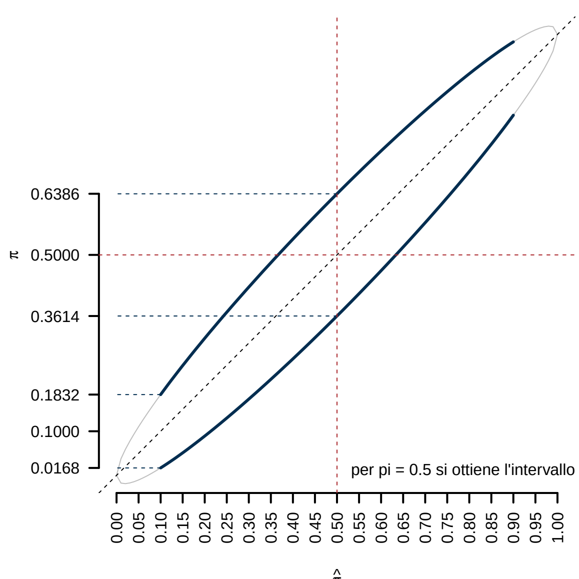
$$\begin{aligned}
&= \left[0.48 - 2.576\sqrt{\frac{0.48(1-0.48)}{500}}; 0.48 + 2.576\sqrt{\frac{0.48(1-0.48)}{500}}\right] \\
&= [0.48 - 0.0576; 0.48 + 0.0576] \\
&= [0.4224; 0.5376]
\end{aligned}$$

Esempio 13.9.2. Una indagine sulle intenzioni di voto degli italiani per lo schieramento A ha mostrato che 2400 su 5000 lo voterebbero. Determinare un IdC al livello di $(1-\alpha) = 0.99$ per π.

$$\begin{aligned}
&\hat{\pi} = \frac{\text{favorevoli}}{n} = \frac{2400}{5000} = 0.48 \\
&\left[\hat{\pi} - z_{\alpha/2}\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}; \hat{\pi} + z_{\alpha/2}\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}\right] \\
&= \left[0.48 - 2.576\sqrt{\frac{0.48(1-0.48)}{5000}}; 0.48 + 2.576\sqrt{\frac{0.48(1-0.48)}{5000}}\right] \\
&= [0.48 - 0.0182; 0.48 + 0.0182] \\
&= [0.4618; 0.4982]
\end{aligned}$$

13.9.1 IdC per π per α ed n fissati

Se fissiamo α ed n, per esempio $\alpha = 0.05$ ed $n = 50$, possiamo variare $S_n \in \{0, ..., n\}$ e quindi $\hat{\pi} \in 0/n, 1/n, ..., n/n$. Per ogni valore di $\hat{\pi}$ calcoliamo l'IdC al livello $(1-\alpha)$ e rappresentiamo graficamente



13.10 Specchietto Finale per gli IdC

- Si definisce L'IdC al livello $(1-\alpha) \times 100\%$ per μ con σ^2 nota, l'intervallo

$$IdC: \quad \left[\hat{\mu} - z_{\alpha/2}\,\frac{\sigma}{\sqrt{n}}, \hat{\mu} + z_{\alpha/2}\,\frac{\sigma}{\sqrt{n}}\right]$$

- Si definisce L'IdC al livello $(1-\alpha) \times 100\%$ per μ con σ^2 incognita, l'intervallo

$$IdC: \quad \left[\hat{\mu} - t_{n-1;\alpha/2}\,\frac{S}{\sqrt{n}}, \hat{\mu} + t_{n-1;\alpha/2}\,\frac{S}{\sqrt{n}}\right]$$

- Si definisce L'IdC al livello $(1-\alpha) \times 100\%$ per π l'intervallo

$$\left[\hat{\pi} - z_{\alpha/2}\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}; \hat{\pi} + z_{\alpha/2}\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}\right]$$

Teoria dei test 14

14.1 Le Ipotesi

Siano $X_1, ..., X_n$ n VC, replicazioni di $X \sim \mathscr{L}(\theta)$

Un *Test Statistico* è la scelta tra due ipotesi diverse su θ alla luce dei dati che osserveremo:

$$\begin{cases} H_0 : \theta \in \Theta_0, & \Theta_0 \subset \Theta \\ H_1 : \theta \in \Theta_1, & \Theta_1 \subset \Theta \end{cases}$$

- Se $\Theta_0 = \{\theta_0\}$ è un solo punto si dice che H_0 è un'ipotesi **semplice**, altrimenti è **composta**
- Se $\Theta_1 = \{\theta_1\}$ è un solo punto si dice che H_1 è un'ipotesi **semplice**, altrimenti è **composta**

14.1.1 Esempi di ipotesi

Esempio: siamo indecisi se l'urna da cui stiamo per estrarre le palline abbia l'80% di palline vincenti oppure il 40%:

$$\begin{cases} H_0 : \pi = 0.8, \\ H_1 : \pi = 0.4 \end{cases}$$

In questo caso, H_0 è un'ipotesi semplice (specifica solo un punto) e un'ipotesi H_1 è semplice serve solo per comprendere la teoria. i primi esempi saranno svolti con due ipotesi semplici.

Esempio: siamo indecisi se il reddito medio degli italiani, di cui stiamo per estrarre un campione, sia uguale a 18 mila € annui oppure minore di 18 mila € annui:

$$\begin{cases} H_0 : \mu = 18, \\ H_1 : \mu < 18 \end{cases}$$

In questo caso, H_0 è un'ipotesi semplice (specifica solo un punto) e H_1 è composta (specifica un'intera regione di Θ) è il caso più interessante nella pratica ed è il caso che svilupperemo maggiormente.

Esempio: siamo indecisi se la SD del reddito degli italiani sia maggiore o minore di 2 mila € annui:

$$\begin{cases} H_0 : \sigma \geq 2, \\ H_1 : \sigma < 2 \end{cases}$$

In questo caso, H_0 è un'ipotesi composta e H_1 è composta è un caso meno interessante e non svilupperemo

14.2 La Decisione

Preparare un test statistico significa dividere lo spazio dei campioni in due

$$\mathcal{S} = \mathcal{S}_0 \cup \mathcal{S}_1, \qquad \mathcal{S}_0 \cap \mathcal{S}_1 = \emptyset$$

Il test è una **decisione**: se il campione proverrà da $\mathcal{S}_0$ il test deciderà per H_0, se il campione proverrà da $\mathcal{S}_1$ il test deciderà per H_1.

Ogni decisione ha delle conseguenze

- Se H_0 è vera ma il campione cade in $\mathcal{S}_1$ sceglierò H_1 erroneamente
- Se H_1 è vera ma il campione cade in $\mathcal{S}_0$ sceglierò H_0 erroneamente

14.3 La tavola della verità

La *tavola della verità* è una tabella simbolica in cui sulle righe viene scritto il *vero stato di natura* cioè, se H_0 è vera o falsa. Per colonna viene scritta la decisione, cioè se scelgo di tenere H_0 o di rifiutarla in favore di H_1.

		Decisione	
		decido H_0	decido H_1
stato di natura	H_0	Corretta	Errore I tipo
	H_1	Errore II tipo	Corretta

Dunque:

Definizione 14.3.1 (Errori di primo e secondo tipo). Si definiscono

- L'**errore di primo tipo** è l'errore che si commette scegliendo H_1 quando è vera H_0.
- L'**errore di secondo tipo** è l'errore che si commette scegliendo H_0 quando è vera H_1.

Dunque ad ogni decisione corrisponde un possibile errore. Per valutare un test si devono calcolare le probabilità di errore

		Decisione	
		decido H_0	decido H_1
stato di natura	H_0	$1-\alpha$	α
	H_1	β	$1-\beta$

Dove

Definizione 14.3.2 (Probabilità degli Errori di primo e secondo tipo).

$$\alpha = P(\text{Errore I tipo}) = P(\text{Decidere } H_1; H_0) = P(X_1, ..., X_n \in \mathcal{S}_1; H_0)$$

$$\beta = P(\text{Errore II tipo}) = P(\text{Decidere } H_0; H_1) = P(X_1, ..., X_n \in \mathcal{S}_0; H_1)$$

α è il livello di **significatività** del test, α è la probabilità di scegliere H_1 quando invece è vera H_0. β è la probabilità di scegliere H_0 quando invece è vera H_1.

Infine

Definizione 14.3.3 (Potenza di un Test).

$$1 - \beta = P(\text{Decidere } H_1; H_1) = P(X_1, ..., X_n \in \mathcal{S}_1; H_1)$$

$1 - \beta$ è la **potenza del test**, $1 - \beta$ è la probabilità di scegliere H_1 quando H_1 è vera.

Esempio 14.3.1 (Spam e Filtri email). Il Un filtro delle email che separa lo *spam* dal *non spam* **non** è una *procedura certa*

$$P(\text{Falso Positivo}) = P(\text{filtrata spam; non è spam}) = \alpha$$
$$P(\text{Falso Negativo}) = P(\text{filtrata non spam; è spam}) = \beta$$

La tavola della verità

		Filtro	
		Non Spam	Spam
stato di natura	Non Spam	$1 - \alpha$	α
	Spam	β	$1 - \beta$

Obiettivo: costruire il filtro in modo tale che α sia fissato ad un valore arbitrariamente piccolo $1-\beta$ sia la più alta possibile, per α fissato

14.4 Esempio: Scegliere tra due ipotesi semplici

Esempio: siamo indecisi se l'urna da cui stiamo per estrarre le palline abbia l'80% di palline vincenti oppure il 40%:

$$\begin{cases} H_0 : \pi = 0.8, \\ H_1 : \pi = 0.4 \end{cases}$$

Decidiamo di estrarre 10 palline CR (IID), $X_1, ..., X_{10}$ per decidere tra H_0 e H_1. Prima di estrarre le palline possiamo calcolare la probabilità di tutti i possibili campioni *sotto ipotesi* H_0 e *sotto ipotesi* H_1. La somma

$$S_{10} = X_1 + ... + X_{10}$$

descrive tutti i possibili campioni Bernoulli IID di ampiezza $n = 10$, $S_{10} \sim \text{Binom}(\pi)$ e quindi

$$\mathcal{S} = \{0,\ 1,\ 2, ..., 10\}$$

S_n	0	1	2	3	4	5	6	7	8	9	10
$P(S_n; \pi = 0.8)$	0.0000	0.0000	0.0001	0.0008	0.0055	0.0264	0.0881	0.2013	0.3020	0.2684	0.1074
$P(S_n; \pi = 0.4)$	0.0060	0.0403	0.1209	0.2150	0.2508	0.2007	0.1115	0.0425	0.0106	0.0016	0.0001

e

$$P(S_{10} = s; \pi) = \binom{10}{s} \pi^s (1-\pi)^{n-s}, \qquad \pi \in \{0.8, 0.4\}$$

Possiamo quindi calcolare

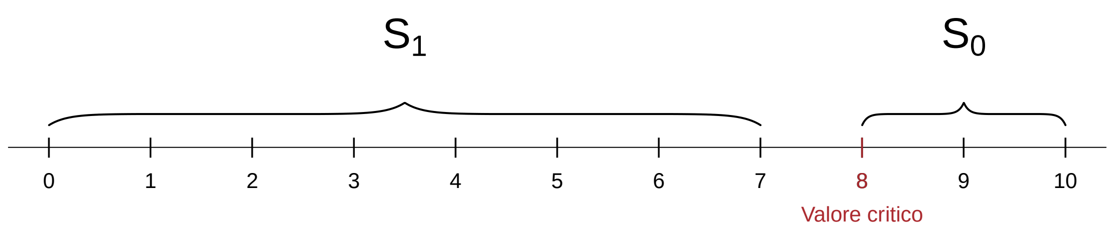
14.4.1 Tre diversi Test a confronto

Sia dato il sistema di ipotesi:

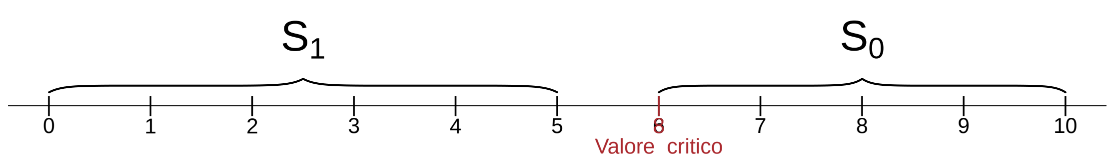
$$\begin{cases} H_0 : \pi = 0.8, \\ H_1 : \pi = 0.4 \end{cases}$$

Un test è una divisione dello spazio dei campioni, e nel modello Bernoulli IID i campioni di ampiezza $n = 10$ sono riassunti dalla loro somma S_{10}

Decisione A Se $S_{10}/10 \geq 0.8 \Rightarrow S_{10} \geq 8$ allora scelgo H_0; se $S_{10}/10 < 0.8 \Rightarrow S_{10} < 8$ allora scelgo H_1



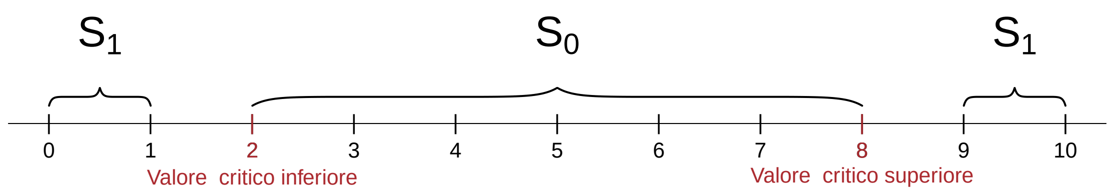
Decisione B Se $S_{10}/10 \geq 0.6 \Rightarrow S_{10} \geq 6$ allora scelgo H_0; se $S_{10}/10 < 0.6 \Rightarrow S_{10} < 6$ allora scelgo H_1



Decisione C Se $0.2 \leq S_{10}/10 \leq 0.8 \Rightarrow 2 \leq S_{10} \leq 8$ allora scelgo H_0; se $S_{10}/10 < 0.2$ oppure $S_{10} > 0.8, \Rightarrow S_{10} < 2$ oppure $S_{10} > 8$ allora scelgo H_1

S_n	0	1	2	3	4	5	6	7	8	9	10
$P(S_n; \pi = 0.8)$	0.0000	0.0000	0.0001	0.0008	0.0055	0.0264	0.0881	0.2013	0.3020	0.2684	0.1074
$P(S_n; \pi = 0.4)$	0.0060	0.0403	0.1209	0.2150	0.2508	0.2007	0.1115	0.0425	0.0106	0.0016	0.0001

S_n	0	1	2	3	4	5	6	7	8	9	10
$P(S_n; \pi = 0.8)$	0.0000	0.0000	0.0001	0.0008	0.0055	0.0264	0.0881	0.2013	0.3020	0.2684	0.1074
$P(S_n; \pi = 0.4)$	0.0060	0.0403	0.1209	0.2150	0.2508	0.2007	0.1115	0.0425	0.0106	0.0016	0.0001



14.4.2 Gli errori della decisione A

Siccome sappiamo calcolare la distribuzione di S_{10} sotto H_0 e sotto H_1:

Decisione A: $\mathcal{S}_0 = \{S_{10} \geq 8\}$ e $\mathcal{S}_1 = \{S_{10} < 8\}$

$$
\begin{aligned}
\alpha &= P(\mathcal{S}_1; H_0) = P(S_{10} = 0 \ \cup S_{10} = 1 \ \cup ... \cup \ S_{10} = 7) \\
&= 0 + 0 + 0 + 0.001 + 0.006 + 0.026 + 0.088 + 0.201 = 0.3222 \\
\beta &= P(\mathcal{S}_0; H_1) = P(S_{10} = 8 \ \cup S_{10} = 9 \ \cup \ S_{10} = 10) \\
&= 0.011 + 0.002 + 0 = 0.0123
\end{aligned}
$$

La significatività del test A è $\alpha = 0.3222$. La potenza del test A è $1 - \beta = 0.9877$

		Decisione A	
		H_0	H_1
stato di natura	$\pi = 0.8$	0.6778	0.3222
	$\pi = 0.4$	0.0123	0.9877

14.4.3 Gli errori della decisione B

Siccome sappiamo calcolare la distribuzione di S_{10} sotto H_0 e sotto H_1:

S_n	0	1	2	3	4	5	6	7	8	9	10
$P(S_n; \pi = 0.8)$	0.0000	0.0000	0.0001	0.0008	0.0055	0.0264	0.0881	0.2013	0.3020	0.2684	0.1074
$P(S_n; \pi = 0.4)$	0.0060	0.0403	0.1209	0.2150	0.2508	0.2007	0.1115	0.0425	0.0106	0.0016	0.0001

Decisione B: $\mathcal{S}_0 = \{S_{10} \geq 6\}$ e $\mathcal{S}_1 = \{S_{10} < 6\}$

$$\begin{aligned}
\alpha &= P(\mathcal{S}_1; H_0) = P(S_{10} = 0 \ \cup S_{10} = 1 \ \cup ... \cup \ S_{10} = 5) \\
&= 0 + 0 + 0 + 0.001 + 0.006 + 0.026 = 0.0328 \\
\beta &= P(\mathcal{S}_0; H_1) = P(S_{10} = 7 \ \cup ... \cup \ S_{10} = 10) \\
&= 0.111 + 0.042 + 0.011 + 0.002 + 0 = 0.1662
\end{aligned}$$

La significatività del test B è $\alpha = 0.0328$. La potenza del test B è $1 - \beta = 0.8338$

		Decisione B	
		H_0	H_1
stato di natura	$\pi = 0.8$	0.9672	0.0328
	$\pi = 0.4$	0.1662	0.8338

14.4.4 Gli errori della decisione C

Siccome sappiamo calcolare la distribuzione di S_{10} sotto H_0 e sotto H_1:

Decisione C: $\mathcal{S}_0 = \{2 \leq S_{10} \leq 8\}$ e $\mathcal{S}_1 = \{S_{10} < 2\} \cup \{S_{10} > 8\}$

$$\begin{aligned}
\alpha &= P(\mathcal{S}_1; H_0) = P(S_{10} = 0 \ \cup S_{10} = 1 \ \cup \ S_{10} = 9 \ \cup S_{10} = 10) \\
&= 0 + 0 + 0.268 + 0.107 = 0.3758 \\
\beta &= P(\mathcal{S}_0; H_1) = P(S_{10} = 2 \ \cup ... \cup \ S_{10} = 8) \\
&= 0.121 + 0.215 + 0.251 + 0.201 + 0.111 + 0.042 + 0.011 = 0.952
\end{aligned}$$

La significatività del test C è $\alpha = 0.3758$, la potenza del test C è $1 - \beta = 0.048$

		Decisione C	
		H_0	H_1
stato di natura	$\pi = 0.8$	0.6242	0.3758
	$\pi = 0.4$	0.9520	0.0480

14.4.5 Confronto

		Decisione A	
		H_0	H_1
stato di natura	$\pi = 0.8$	0.6778	0.3222
	$\pi = 0.4$	0.0123	0.9877

		Decisione B	
		H_0	H_1
stato di natura	$\pi = 0.8$	0.9672	0.0328
	$\pi = 0.4$	0.1662	0.8338

		Decisione C	
		H_0	H_1
stato di natura	$\pi = 0.8$	0.6242	0.3758
	$\pi = 0.4$	0.9520	0.0480

14.5 Ipotesi Nulla e Ipotesi Alternativa

Nell'approccio *decisionista* la statistica è di aiuto al decision manager: la decisione tra H_0 e H_1 deve tenero conto dei *costi* che una decisione sbagliata comporta e pesare i *costi* con la probabilità di sbagliare per ottenere una valutazione di *rischio*. Interessante approccio ma non verrà seguito in questo corso.

Nell'approccio *falsificazionista* (https://it.wikipedia.org/wiki/Principio_di_falsificabilit%C3%A0), la statistica è di aiuto alla conoscenza del mondo esterno e delle sue *leggi*. Seguiremo questo approccio. In molte discipline H_0 viene chiamata **Ipotesi Nulla** e ha una particolare interpretazione: H_0 è lo stato di conoscenza pregresso. Rappresenta l'ipotesi che i dati non abbiano aggiunto niente di diverso da ciò che già conoscevamo. Sotto H_0 asserisce che la **differenza** tra l'evidenza del campione e l'ipotesi nulla è solo **dovuta al caso**. H_1 è chiamata **ipotesi alternativa**, sotto H_1 la **differenza** tra l'evidenza dei dati e H_0 **non è dovuta al caso** ma ad un *fattore sistematico*. La maggior parte delle volte l'obiettivo della ricerca si conclude positivamente se H_0 viene rifiutata, e dunque i dati presentati nella ricerca hanno aggiunto *conoscenza*.

14.6 Rifiutare o non rifiutare H_0

Quando si prepara test, spesso, si vuole mostrare che i dati campionari smentiscono alcune ipotesi pregresse. H_0 viene abbandonata (**rifiutata**,**confutata**) solo se c'è una forte evidenza campionaria contro

- Se H_0 **non** viene rifiutata allora la differenza tra i dati e l'ipotesi nulla è considerata non

significativa
- Se H_0 viene rifiutata allora la differenza tra i dati e l'ipotesi è considerata significativa

Rifiutare H_0 **non** significa accettare H_1, ma sostituire una ipotesi pregressa H_0 con una nuova supportata dai dati, H_1. H_1 diventa la nuova H_0 e sarà considerata lo stato di conoscenza fino a quando non si troveranno nuovi dati che la **confutano**. Per preservare H_0 scegliamo una probabilità di significatività bassa

$$\alpha = \{0.05,\ 0.01,\ 0.005,\ 0.001\}$$

- $\alpha = 0.05$ produce un test al 5%
- $\alpha = 0.01$ produce un test all'1%
- ecc.

L'obiettivo è trovare, tra tutte i test a livello α fissato, quello con potenza più alta. Non sempre il test più potente esiste. Non tratteremo il problema del test più potente in maniera sistematica.

14.7 Test per μ: due ipotesi semplici, σ^2 nota

Stiamo per estrarre $n = 10$ VC IID, $X_1, ..., X_{10}$ da una normale $X \sim N(\mu, \sigma^2)$, di cui conosciamo $\sigma^2 = 3.5^2$. Per esempio siamo indecisi tra:

$$\begin{cases} H_0 : \mu = \mu_0 = 21 \\ H_1 : \mu = \mu_1 = 24 \end{cases}$$

Siccome $S_{10} = X_1 + ... + X_{10}$ racchiude tutta l'informazione sulla media μ che possiamo estrarre da un campione di normali IID con varianza nota, possiamo ragionare su $\hat{\mu} = S_{10}/10$ invece che su tutto lo spazio dei campioni $\mathcal{S}$.

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N\left(\mu, \frac{\sigma^2}{n} = 1.1068^2\right)$$

- Sotto H_0

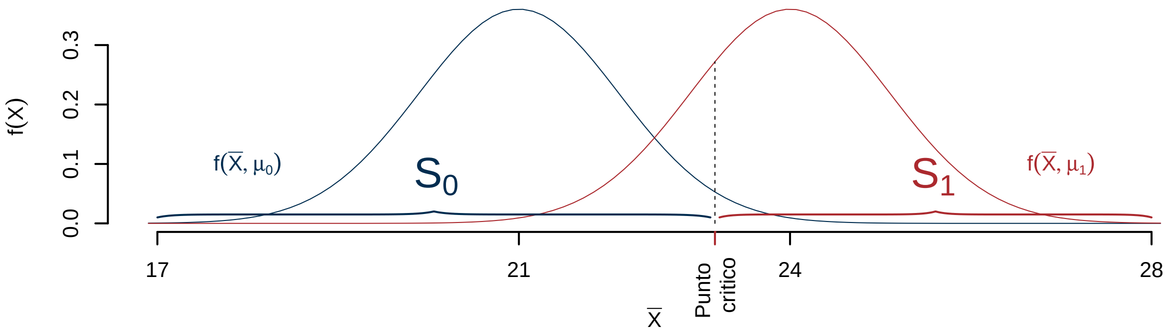
$$\hat{\mu} \sim N(21, 1.1068^2)$$

- Sotto H_1

$$\hat{\mu} \sim N(24, 1.1068^2)$$

14.7.1 Test per μ: scegliere il punto critico

Una decisione consiste nello scegliere il punto critico sullo spazio delle medie:



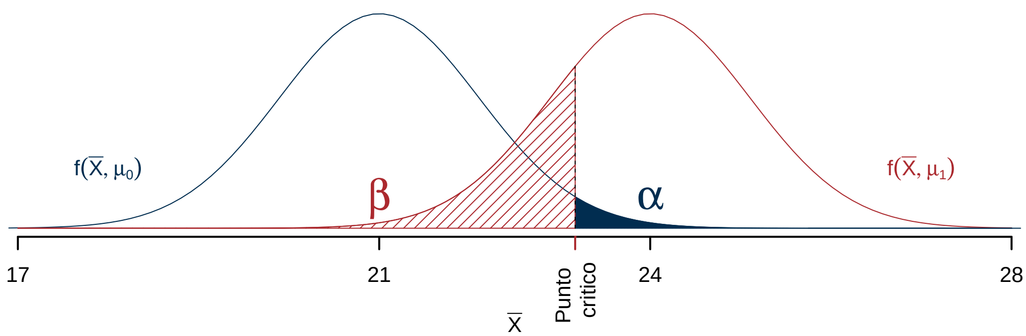
Potremmo prendere, per esempio, come punto critico $\bar{x} = 23$: se la media dei dati è minore di 23 sceglierò H_0 altrimenti sceglierò H_1.

14.7.2 Probabilità di errore di primo e di secondo tipo

Fissato il punto critico, per esempio 23, osserviamo

$$
\begin{aligned}
\alpha &= P(\text{Errore I Tipo}) \\
&= P(\text{Scegliere } H_1 \text{ quando è vera } H_0) \\
&= P(\bar{X} > 23; H_0) \qquad \text{sotto ipotesi } H_0\text{: } \mu = 21 \\
&= 1 - P(\bar{X} \leq 23; H_0) \\
&= 1 - P\left(\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \leq \frac{23 - 21}{3.5/\sqrt{10}}\right) \\
&= 1 - \Phi(1.807) \\
&= 0.0354 \\
\beta &= P(\text{Errore II Tipo}) \\
&= P(\text{Scegliere } H_0 \text{ quando è vera } H_1) \\
&= P(\bar{X} < 23; H_1) \qquad \text{sotto ipotesi } H_1\text{: } \mu = 24 \\
&= P(\bar{X} \leq 23; H_1) \\
&= P\left(\frac{\bar{X} - \mu_1}{\sigma/\sqrt{n}} \leq \frac{23 - 24}{3.5/\sqrt{10}}\right) \\
&= \Phi(-0.9035) \\
&= 0.1831
\end{aligned}
$$

Graficamente osserviamo.



14.7.3 Test per α fissato, $\alpha = 0.05$

Scegliere il punto critico in questo modo è molto arbitrario e si preferisce scegliere α. Il valore della probabilità di primo tipo viene scelto basso preservare H_0 e rifiutarla solo se l'evidenza campionaria è molto a favore di H_1. Fissiamo un valore di α abbastanza piccolo, per esempio, $\alpha = 0.05$. Trovare il **punto critico** per α fissato significa risolvere il seguente problema inverso

$$P(\hat{\mu} > \text{Punto Critico}; H_0) = 0.05$$

Sotto H_0

$$\frac{\hat{\mu} - \mu_0}{\sigma/\sqrt{n}} = \frac{\hat{\mu} - 21}{3.5/\sqrt{10}} = \frac{\hat{\mu} - 21}{1.1068} \sim N(0,1)$$

Problema inverso lo trasferiamo su Z, per esempio, e trovo il punto critico

$$\begin{aligned}
P(Z > z_{0.05}) &= 0.05 \\
z_{0.05} &= 1.6449 \qquad \text{letto dalle tavole} \\
P(Z > 1.6449) &= 0.05 \\
P\left(\frac{\hat{\mu} - 21}{1.1068} > 1.6449\right) &= 0.05 \\
P\left(\hat{\mu} > 21 + 1.6449 \cdot 1.1068\right) &= 0.05 \\
P\left(\hat{\mu} > 22.8205\right) &= 0.05 \\
\text{Punto Critico} &= \mu_0 + z_{0.05}\frac{\sigma}{\sqrt{n}} \\
&= 22.8205
\end{aligned}$$

14.7.4 La regola di decisione, $\alpha = 0.05$

- Se risulterà $\hat{\mu} \leq 22.8205$ allora decideremo di **non rifiutare** H_0
- Se risulterà $\hat{\mu} > 22.8205$ allora decideremo di **rifiutare** H_0

La **significatività** di questo test è al 5%:

$$\alpha = 0.05 = P(\hat{\mu} > 22.8205; H_0)$$

Per calcolare β dobbiamo calcolare

$$\begin{aligned}
\beta &= P(\hat{\mu} < 22.8205; H_1) \\
&= P\left(\frac{\hat{\mu}-\mu_1}{\sigma/\sqrt{n}} < \frac{22.8205-24}{1.1068}\right) \\
&= P\left(Z < 1.6449 + \frac{21-24}{1.1068}\right) \\
&= P(Z < -1.0657) \\
&= \Phi(-1.0657) \\
&= 0.1433
\end{aligned}$$

La **potenza** del test è 85.6714%

$$1 - \beta = 0.8567$$

14.7.5 Test per α fissato, $\alpha = 0.01$

Fissiamo un valore di α abbastanza piccolo, per esempio, $\alpha = 0.01$. Trovare il punto critico per α fissato significa risolvere il seguente problema inverso

$$P(\hat{\mu} > \text{Punto Critico}; H_0) = 0.01$$

Sotto H_0

$$\frac{\hat{\mu}-\mu_0}{\sigma/\sqrt{n}} = \frac{\hat{\mu}-21}{3.5/\sqrt{10}} = \frac{\hat{\mu}-21}{1.1068} \sim N(0,1)$$

Problema inverso lo trasferiamo su Z, per esempio, e trovo il punto critico

$$\begin{aligned}
P(Z > z_{0.01}) &= 0.01 \\
z_{0.01} &= 2.3263 \qquad \text{letto dalle tavole} \\
P(Z > 2.3263) &= 0.01 \\
P\left(\frac{\hat{\mu}-21}{1.1068} > 2.3263\right) &= 0.01 \\
P\left(\hat{\mu} > 21 + 2.3263 \cdot 1.1068\right) &= 0.01 \\
\text{Punto Critico} &= \mu_0 + z_{0.01}\frac{\sigma}{\sqrt{n}} \\
&= 23.5748
\end{aligned}$$

14.7.6 La regola di decisione, $\alpha = 0.01$

- Se risulterà $\hat{\mu} \leq 23.5748$ allora decideremo di **non rifiutare** H_0.
- Se risulterà $\hat{\mu} > 23.5748$ allora decideremo di **rifiutare** H_0.

La significatività di questo test è all'1%:

$$\alpha = 0.01 = P(\hat{\mu} > 23.5748; H_0)$$

Per calcolare β dobbiamo calcolare

$$\begin{aligned}
\beta &= P(\hat{\mu} < 23.5748; H_1) \\
&= P\left(\frac{\hat{\mu} - \mu_1}{\sigma/\sqrt{n}} < \frac{23.5748 - 24}{1.1068}\right) \\
&= P\left(Z < 2.3263 + \frac{21 - 24}{1.1068}\right) \\
&= P(Z < -0.3842) \\
&= \Phi(-0.3842) \\
&= 0.3504
\end{aligned}$$

La **potenza** del test è 64.9576%

$$1 - \beta = 0.6496$$

14.8 H_0 semplice e H_1 composta

Nella pratica più comune, se θ è il parametro da testare, analizzeremo i tre seguenti sistemi di ipotesi

Unilaterale destra

$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta > \theta_0 \end{cases}$$

Unilaterale sinistra

$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta < \theta_0 \end{cases}$$

Bilaterale

$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta \neq \theta_0 \end{cases}$$

14.9 La Statistica Test

La **Statistica Test** è una statistica (una funzione dei dati) che agevola il processo decisionale. Consente di non dover trovare il punto critico nello spazio di $\mathcal{S}$. Calcolata sul campione osservato consente di decidere immediatamente se il campione porta più evidenza ad H_0 oppure ad H_1.

Nota

Se $\hat{\theta}$ è stimatore per θ, con errore di stima $SE(\hat{\theta})$, una pratica comune è la statistica test

$$T = \frac{\hat{\theta} - \theta_0}{SE(\hat{\theta})}$$

come misura di allontanamento da H_0

- se $\hat{\theta} = \theta_0$, allora $T = 0$
- se $\hat{\theta} > \theta_0$, allora $T > 0$
- se $\hat{\theta} < \theta_0$, allora $T < 0$

Se l'errore di stima $SE(\hat{\theta})$ è grande, allora piccole differenze tra $\hat{\theta}$ e θ_0 **non comportano** a grandi variazioni di T intorno a zero.
Se l'errore di stima $SE(\hat{\theta})$ è piccolo, allora piccole differenze tra $\hat{\theta}$ e θ_0 **comportano** grandi variazioni di T intorno a zero.
Trovare il punto critico sullo spazio di $\mathcal{S}$ equivale a trovalo su T.

Test per una media e una proporzione

15

In questo capitolo svilupperemo il test per una media. Essendo la proporzione una particolare media su elementi 0 e 1 la costruzione verrà in automatico. In tutti i casi considereremo un solo campione $x_1, ..., x_n$ proveniente da una popolazione che ha media $E(X_i) = \mu, \forall i$ e $V(X_i) = \sigma^2, \forall i$.

Il sistema di ipotesi porrà sempre $H_0 : \mu = \mu_0$ e si leggerà: la vera media che ha generato i dati è μ_0, il fatto che la media dei dati sia $\bar{x} \neq \mu_0$ è dovuto all'effetto del campionamento e la differenza tra $\bar{x}$ e μ_0 **non è significativa**. L'ipotesi alternativa H_1, che sarà unilaterale o bilaterale, si legge al contrario: la differenza tra la media osservata e la media prescritta da H_0 **è significativa** e il campione proviene da una popolazione che non ha media μ_0.

Alla luce dei dati decideremo se scegliere di rifiutare H_0 o se non rifiutarla. Costruiremo anche una misura di *vicinanza* empirica ad H_1, il p_{value}, che ci consentirà di capire quanto l'evidenza empirica dei dati supporti supporti una delle due ipotesi.

15.1 Test sulla media, σ^2 noto

15.1.1 Test sulla media, ipotesi unilaterale destra, σ^2 noto

Siano $X_1, ...X_n$, n VC, IID: $X_i \sim N(\mu, \sigma^2)$ con σ^2 noto. Consideriamo il seguente sistema di ipotesi

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu > \mu_0 \end{cases}$$

Dalle proprietà della normale

$$\hat{\mu} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

E quindi

$$Z = \frac{\hat{\mu} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

Sotto $H_0 : \mu = \mu_0$ la statistica

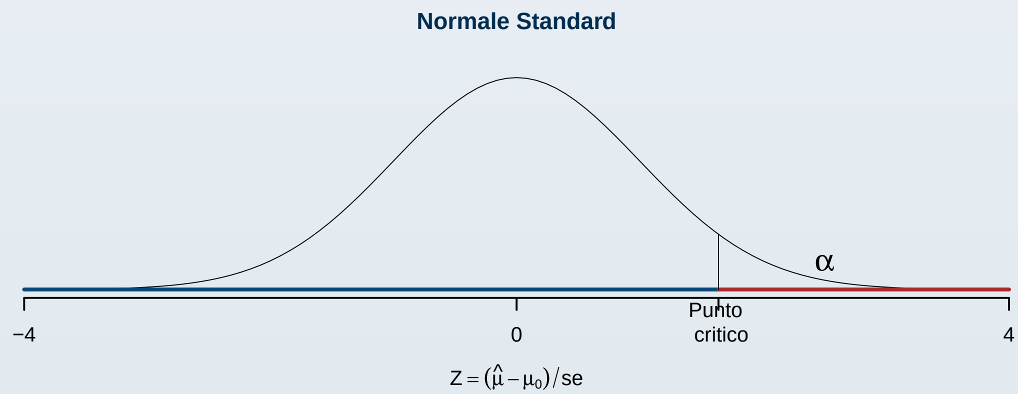
$$Z = \frac{\hat{\mu} - \mu_0}{\sigma/\sqrt{n}} \sim N(0, 1)$$

Z osservata sul campione misura l'evidenza contro H_0, tanto più è alto il valore di Z tanto più H_0 è inverosimile.

Decisione sul campione. Si decide un livello α e si ricava z_α, si estrae un campione. Lo stimatore $\hat{\mu}$ si realizza nella media osservata del campione $\bar{x}$

$$z_{\text{obs}} = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

- Se $z_{\text{obs}} < z_\alpha$ H_0 **non viene rifiutata** al livello di significatività $\alpha \times 100\%$
- Se $z_{\text{obs}} > z_\alpha$ H_0 **viene rifiutata** al livello di significatività $\alpha \times 100\%$



15.1.2 Test sulla media, σ^2 noto, vari livelli di α

Se $\alpha = 0.05$ allora

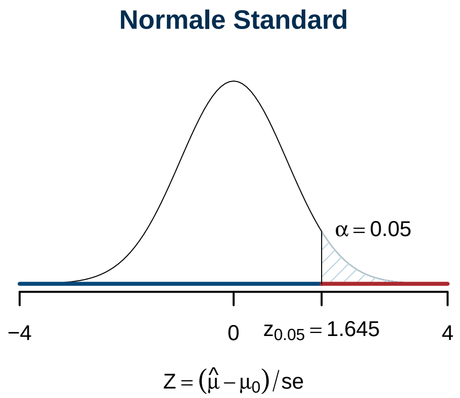
$$\text{Punto Critico} = z_{0.05}, \qquad P(Z > z_{0.05}) = 0.05$$

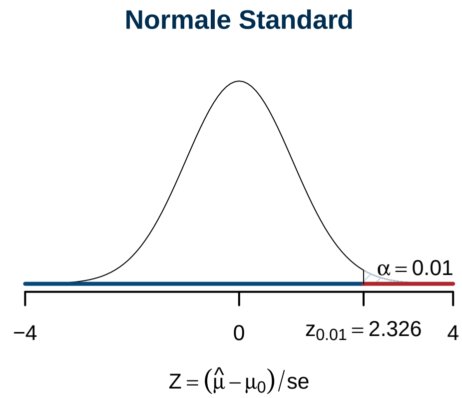
$z_{0.05} = 1.6449$

Se $\alpha = 0.01$ allora

$$\text{Punto Critico} = z_{0.01}, \qquad P(Z > z_{0.01}) = 0.01$$

$z_{0.01} = 2.3263$





Esempio 15.1.1. In un laboratorio, vi sono cavie con peso medio uguale a 30g e una DS pari a 5g. Uno studente seleziona 25 cavie e ottiene un peso medio uguale a 32g con una $\hat{\sigma}$ pari a 7g. La selezione è casuale o il valore ottenuto della media è troppo alto, a un LdS del 5%?

Test Z per una media, variazna nota

$\boxed{\text{A}}$ FORMULAZIONE DELLE IPOTESI

$$\begin{cases} H_0 : \mu = \mu_0 = 30 \\ H_1 : \mu > \mu_0 = 30 \end{cases}$$

$\boxed{\text{B}}$ SCELTA E CALCOLO STATISTICA-TEST, Z

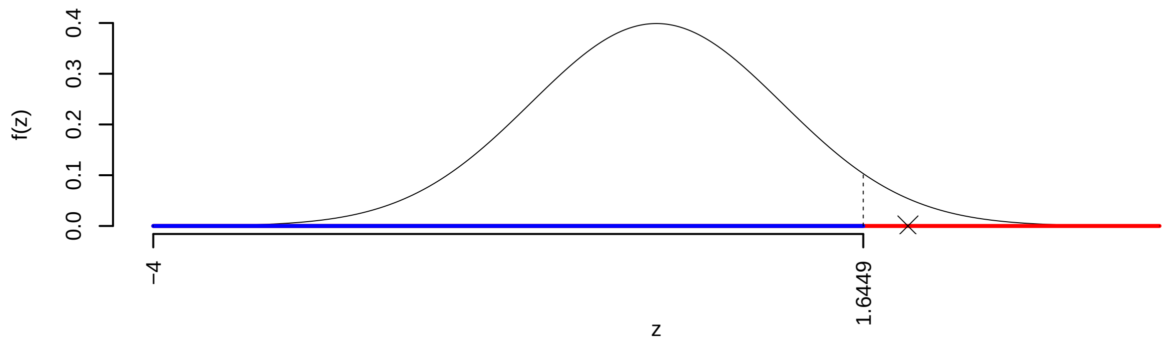
σ^2 di P è nota: $\Rightarrow$ z-Test.

$$\begin{aligned} \frac{\hat{\mu} - \mu_0}{\sigma/\sqrt{n}} &\sim N(0,1) \\ z_{\text{obs}} &= \frac{(32-30)}{5/\sqrt{25}} = 2\,. \end{aligned}$$

$\boxed{\text{C}}$ CONCLUSIONE

La siginficatitività è $\alpha = 0.05$, dalle tavole osserviamo $z_{0.05} = 1.6449$.

Essendo $z_{\text{obs}} = 2 > z_{0.05} = 1.6449$ allora **rifiuto** H_0 al 5%.



Esempio 15.1.2. In un laboratorio, vi sono cavie con peso medio uguale a 30g e una DS pari a 5g. Uno studente seleziona 25 cavie e ottiene un peso medio uguale a 32g con una $\hat{\sigma}$ pari a 7g. La selezione è casuale o il valore ottenuto della media è troppo alto, a un LdS del 1%?

Test Z per una media, variazna nota

$\boxed{\text{A}}$ FORMULAZIONE DELLE IPOTESI

$$\begin{cases} H_0 : \mu = \mu_0 = 30g \\ H_1 : \mu > \mu_0 = 30g \end{cases}$$

$\boxed{\text{B}}$ SCELTA E CALCOLO STATISTICA-TEST, Z

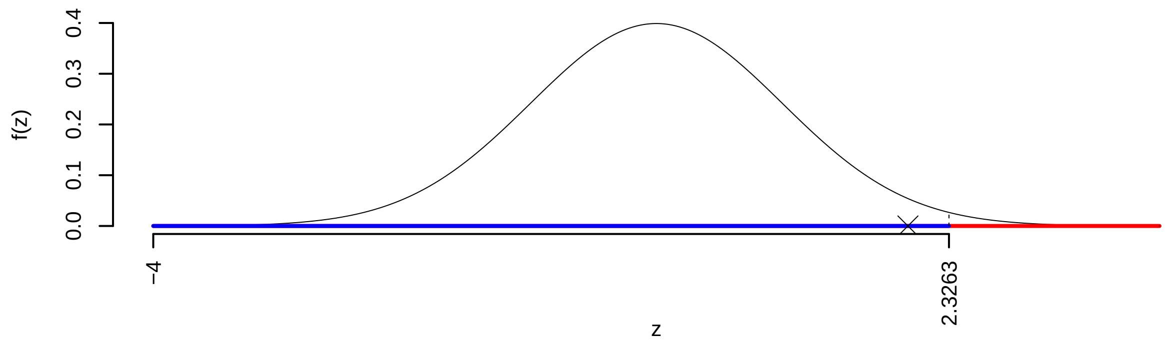
σ^2 di P è nota: $\Rightarrow$ z-Test.

$$\begin{aligned} \frac{\hat{\mu} - \mu_0}{\sigma/\sqrt{n}} &\sim N(0,1) \\ z_{\text{obs}} &= \frac{(32-30)}{5/\sqrt{25}} = 2\,. \end{aligned}$$

$\boxed{\text{C}}$ CONCLUSIONE

La siginficatitività è $\alpha = 0.01$, dalle tavole osserviamo $z_{0.01} = 2.3263$.

Essendo $z_{\text{obs}} = 2 < z_{0.01} = 2.3263$ allora **non** rifiuto H_0 al 1%.

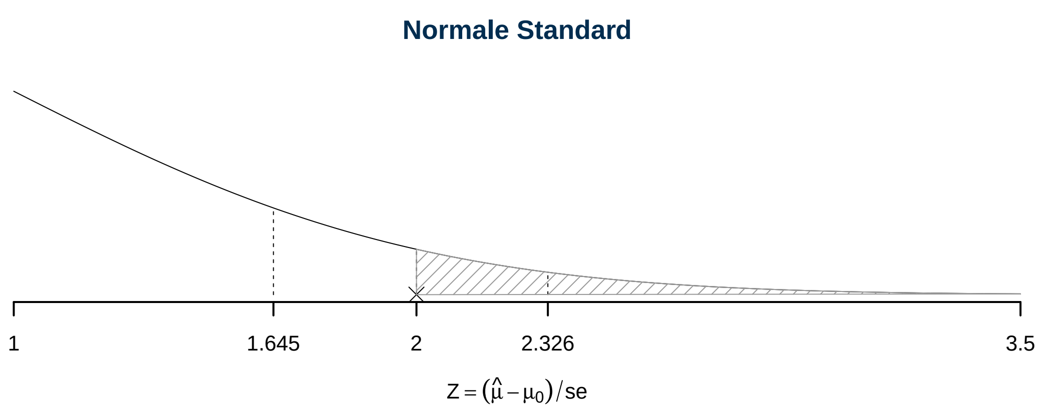


15.1.3 La probabilità di significatività osservata il p_{value}

Il p_{value} risponde a questa domanda: *Se nell'esempio precedente usassimo z_{obs} come punto critico quale sarebbe la probabilità di significatività?*

Definizione 15.1.1 (p_{value} caso unilaterale destro)**.**

$$p_{\text{value}} = P(Z > z_{\text{obs}}; H_0)$$



$$\begin{aligned} p_{\text{value}} &= P(Z > 2) \\ &= 1 - P(Z < 2) \\ &= 1 - \Phi(2) \\ &= 1 - 0.9772 \\ &= 0.0228 \end{aligned}$$

15.1.4 Lettura del p_{value}

La probabilità di significatività osservata p_{value} ci dice la probabilità, *se fosse vera* H_0 di trovare un campione ancora più in favore di H_1 di quello che ho trovato. In altre parole ci dice, *se fosse vera* H_0, quanto sarebbe improbabile il nostro campione.

Nota

Tanto più basso è il p_{value} tanto più forte è l'evidenza dei dati contro H_0:

- Se $p_{\text{value}} > 0.1 \to$ il test **non** è significativo
- Se $0.05 < p_{\text{value}} \leq 0.1 \to$ il test è *marginalmente* significativo $\boxed{\cdot}$
- Se $0.05 < p_{\text{value}} \leq 0.01 \to$ il test è significativo $\to$ significativo al 5% $\boxed{*}$
- Se $0.001 < p_{\text{value}} \leq 0.01 \to$ il test è *molto* significativo $\to$ significativo all'1% $\boxed{**}$
- Se $p_{\text{value}} \leq 0.001 \to$ il test è *estremamente* significativo $\to$ significativo sotto all'1‰ $\boxed{***}$

15.1.5 Test sulla media, ipotesi unilaterale sinistra, σ^2 noto

Siano $X_1, ...X_n$ n VC, IID: $X_i \sim N(\mu, \sigma^2)$ con σ^2 noto.

Consideriamo il seguente sistema di ipotesi

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu < \mu_0 \end{cases}$$

Dalle proprietà della normale

$$\hat{\mu} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

E quindi

$$Z = \frac{\hat{\mu} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

Sotto H_0 la statistica

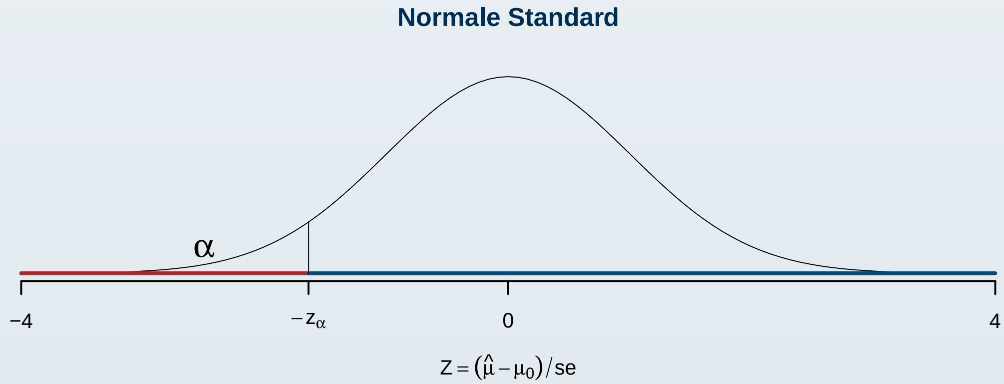
$$Z = \frac{\hat{\mu} - \mu_0}{\sigma/\sqrt{n}} \sim N(0, 1)$$

Z osservata sul campione misura l'evidenza contro H_0, tanto più è basso il valore di Z tanto più H_0 è inverosimile

Decisione sul campione. Si decide un livello α e si ricava z_α, si estrae un campione. Lo stimatore $\hat{\mu}$ si realizza nella media osservata del campione $\bar{x}$

$$z_{\text{obs}} = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

- Se $z_{\text{obs}} > -z_\alpha$ H_0 **non viene rifiutata** al livello di significatività $\alpha \times 100\%$
- Se $z_{\text{obs}} < -z_\alpha$ H_0 **viene rifiutata** al livello di significatività $\alpha \times 100\%$



Esempio 15.1.3. La durata, in ore, della resistenza di un transistor alle alte temperature sia $X \sim N(\mu, \sigma^2)$ e, in base all'esperienza, sia $\mu_0 = 6$ ore e $\sigma_0 = 0.5$ ore. Si apportano modifiche alla composizione dei materiali per fare diminuire la durata della resistenza alla temperatura. Si eseguono 16 osservazioni, dalle quali si ottiene una durata media $\bar{x} = 5.7$ e uno deviazione standard $s = 0.6$. Verificare l'ipotesi, al LdS dell'1%, che la resistenza sia rimasta invariata contro l'alternativa che sia diminuita.

Test Z per una media, variazna nota

$\boxed{\text{A}}$ FORMULAZIONE DELLE IPOTESI

$$\begin{cases} H_0 : \mu = \mu_0 = 6h \\ H_1 : \mu < \mu_0 = 6h \end{cases}$$

$\boxed{\text{B}}$ SCELTA E CALCOLO STATISTICA-TEST, Z

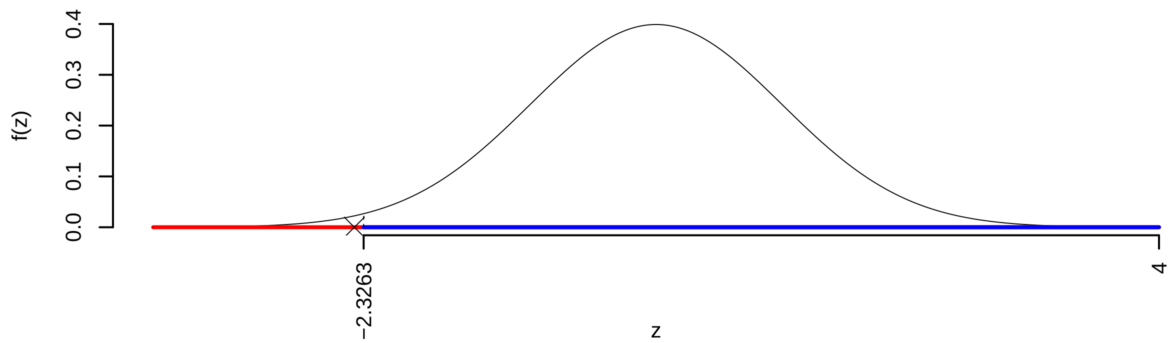
σ^2 di P è nota: $\Rightarrow$ z-Test.

$$\begin{aligned} \frac{\hat{\mu} - \mu_0}{\sigma/\sqrt{n}} &\sim N(0,1) \\ z_{\text{obs}} &= \frac{(5.7 - 6)}{0.5/\sqrt{16}} = -2.4\,. \end{aligned}$$

$\boxed{\text{C}}$ CONCLUSIONE

La siginficatitività è $\alpha = 0.01$, dalle tavole osserviamo $z_{0.01} = -2.3263$.

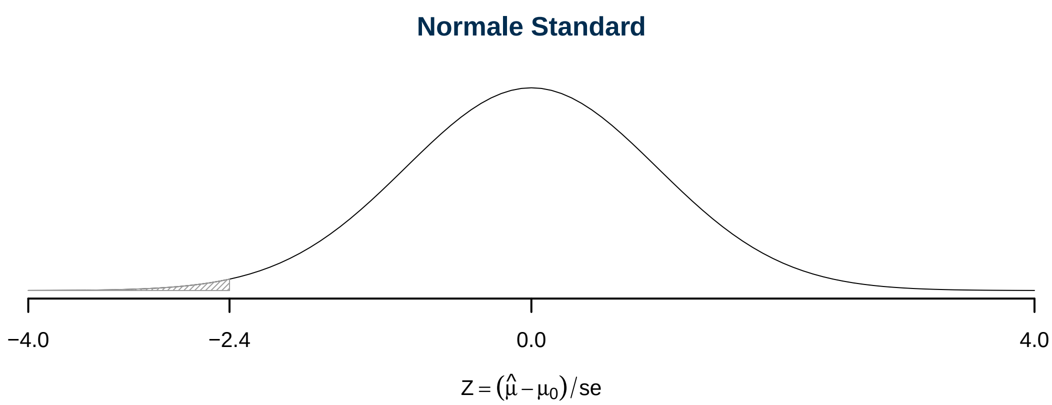
Essendo $z_{\text{obs}} = -2.4 < z_{0.01} = -2.3263$ allora **rifiuto** H_0 al 1%.



15.1.5.1 La probabilità di significatività osservata il p_{value}

Definizione 15.1.2 (p_{value} caso unilaterale sinistro)**.** Nel caso di ipotesi unilaterale sinistra abbiamo

$$p_{\text{value}} = P(Z < z_{\text{obs}}; H_0)$$



$$\begin{aligned} p_{\text{value}} &= P(Z < -2.4) \\ &= \Phi(-2.4) \\ &= 1 - \Phi(2.4) \\ &= 1 - 0.9918 \\ &= 0.0082 \end{aligned}$$

15.1.6 Test sulla media, ipotesi bilaterale, σ^2 noto

Siano $X_1, ...X_n$ n VC, IID: $X_i \sim N(\mu, \sigma^2)$ con σ^2 noto. Consideriamo il seguente sistema di ipotesi

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases}$$

Sotto H_0 la statistica

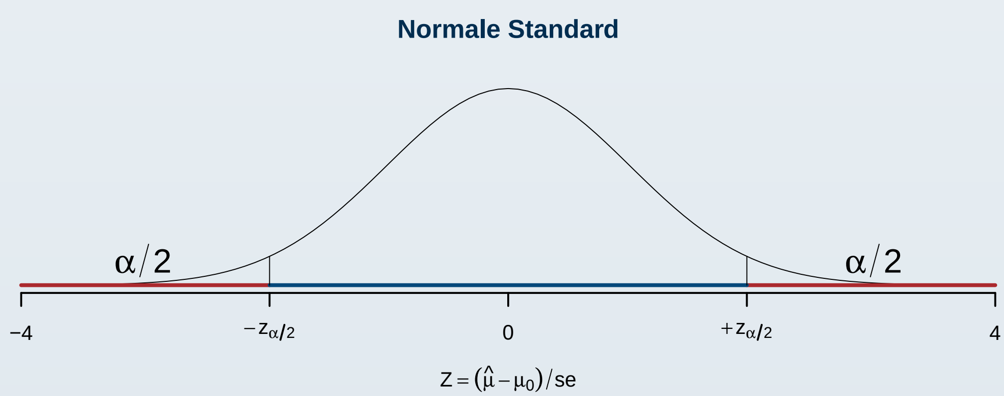
$$Z = \frac{\hat{\mu} - \mu_0}{\sigma/\sqrt{n}} \sim N(0, 1)$$

Z osservata sul campione misura l'evidenza contro H_0, tanto più è il valore di Z è **diverso** da zero tanto più H_0 è inverosimile

Decisione sul campione. Si decide un livello α e si ricava $z_{\alpha/2}$, si estrae un campione. Lo stimatore $\hat{\mu}$ si realizza nella media osservata del campione $\bar{x}$

$$z_{\text{obs}} = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

- Se $-z_{\alpha/2} \leq z_{\text{obs}} \leq z_{\alpha/2}$ H_0 **non viene rifiutata** al livello di significatività $\alpha \times 100\%$
- Se $z_{\text{obs}} < -z_{\alpha/2}$, o $z_{\text{obs}} > +z_{\alpha/2}$ H_0 **viene rifiutata** al livello di significatività $\alpha \times 100\%$



Esempio 15.1.4. Un produttore di semiconduttori afferma che la durata media dei suoi chip è di 300 ore con una deviazione standard di 3 ore. Un'azienda decide di testare la qualità dei chip e conduce un esperimento con 16 chip, ottenendo una durata media di 298 ore e una deviazione standard di 4 ore. I risultati di questo test sono coerenti con l'affermazione del produttore i chip tendono a durare in modo diverso dalla durata dichiarata, al 5%?

Test Z per una media, variazna nota

$\boxed{\text{A}}$ FORMULAZIONE DELLE IPOTESI

$$\begin{cases} H_0 : \mu = \mu_0 = 300\ h \\ H_1 : \mu \neq \mu_0 = 300\ h \end{cases}$$

$\boxed{\text{B}}$ SCELTA E CALCOLO STATISTICA-TEST, Z

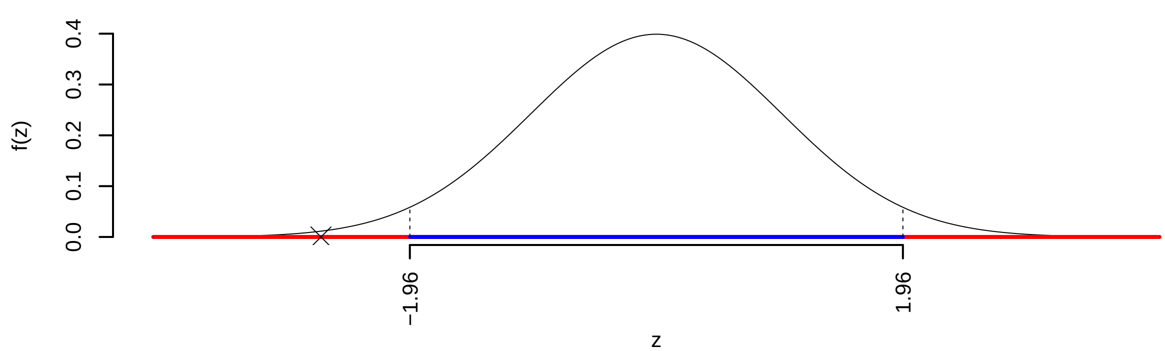
σ^2 di P è nota: $\Rightarrow$ z-Test.

$$\begin{aligned} \frac{\hat{\mu} - \mu_0}{\sigma/\sqrt{n}} &\sim N(0,1) \\ z_{\text{obs}} &= \frac{(298 - 300)}{3/\sqrt{16}} = -2.667\,. \end{aligned}$$

$\boxed{\text{C}}$ CONCLUSIONE

La siginficatitività è $\alpha = 0.05$, dalle tavole osserviamo $z_{0.025} = 1.96$.

Essendo $|z_{\text{obs}}| = 2.6667 > z_{0.025} = 1.96$ allora **rifiuto** H_0 al 5%.



Il p_{value} è

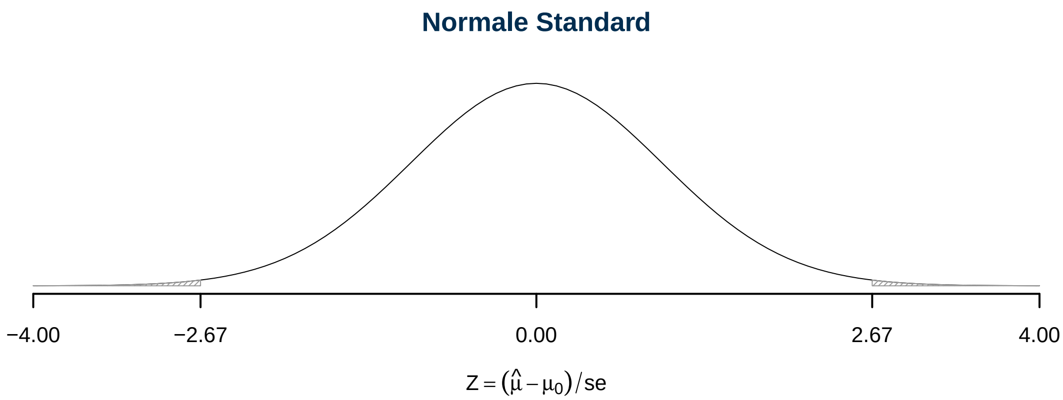
$$p_{\text{value}} = P(|Z| > |-2.67|) = 2P(Z > 2.67) = 0.007661$$

$$0.001 < p_{\text{value}} = 0.007661 \leq 0.01$$

15.1.6.1 La probabilità di significatività osservata il p_{value}

Definizione 15.1.3 (p_{value} **caso bilaterale**)**.** Nel caso di ipotesi bilaterale abbiamo

$$p_{\text{value}} = P(|Z| > |z_{\text{obs}}|; H_0)$$



$$\begin{aligned}
p_{\text{value}} &= P(|Z| > 2.67) \\
&= P(Z < -2.67) + P(Z > 2.67) \\
&= 2 \cdot P(Z > 2.67) \\
&= 2 \cdot (1 - \Phi(2.67)) \\
&= 2 \cdot 0.0038 \\
&= 0.0076
\end{aligned}$$

15.2 Significatività non fissata

Le strategie per decidere per H_0 invece di H_1 cambiano a seconda di cosa rappresentano realmente le due ipotesi. In controllo di qualità industriale H_0 significa conforme agli standard e H_1 non conforme, fissare α e di conseguenza β è frutto di un bilancio rischi benefici che richiede di esplicitare i costi che l'azienda paga in base all'errore: quando costa commettere l'errore di primo tipo? Quanto costa commettere l'errore di secondo tipo?

In altri contesti di ricerca la lontananza dei dati da H_0 può essere oggetto di dibattito e il p_{value} misura esattamente questo. Senza avere fissato in anticipo α il ricercatore di trova a discutere se il p_{value} è sufficientemente piccolo da rigettare l'ipotesi nulla. Il prossimo esempio esemplifica l'approccio

Esempio 15.2.1. Un'azienda farmaceutica afferma che la concentrazione media di principio attivo in una certa medicina è di 50 mg con una deviazione standard di 2 mg. Un laboratorio indipendente decide di verificare questa affermazione eseguendo 25 analisi e ottiene una concentrazione media di

51 mg e una deviazione standard di 2.5 mg. I risultati delle analisi sono coerenti con l'affermazione dell'azienda o la concentrazione del principio attivo è maggiore dalla quantità dichiarata?

Test Z per una media, variazna nota

$\boxed{\text{A}}$ FORMULAZIONE DELLE IPOTESI

$$\begin{cases} H_0 : \mu = \mu_0 = 50mg \\ H_1 : \mu > \mu_0 = 50mg \end{cases}$$

$\boxed{\text{B}}$ SCELTA E CALCOLO STATISTICA-TEST, Z

σ^2 di P è nota: $\Rightarrow$ z-Test.

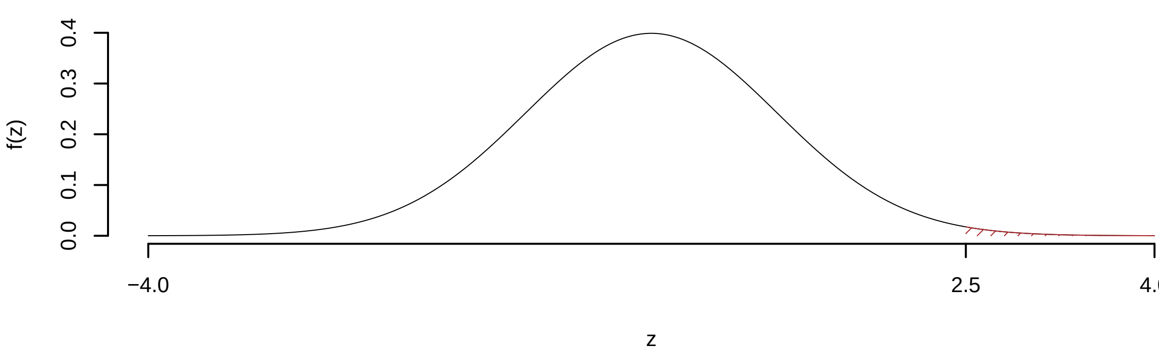
$$\begin{aligned} \frac{\hat{\mu} - \mu_0}{\sigma/\sqrt{n}} &\sim N(0,1) \\ z_{\text{obs}} &= \frac{(51-50)}{2/\sqrt{25}} = 2.5\,. \end{aligned}$$

$\boxed{\text{C}}$ CONCLUSIONE

Il p_{value} è

$$p_{\text{value}} = P(Z > 2.5) = 0.006210$$

$$0.001 < p_{\text{value}} = 0.006210 \leq 0.01$$



Rifiuto H_0 all'1%,

$0.001 < p_{\text{value}} < 0.01$, *molto significativo* $\boxed{**}$.

15.3 Test per μ, σ incognita

Se σ^2 è incognito va stimato dai dati. Consideriamo lo stimatore S^2 di σ^2

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \hat{\mu})^2 = \frac{n}{n-1}\frac{1}{n}\sum_{i=1}^{n}(X_i - \hat{\mu})^2 = \frac{n}{n-1}\hat{\sigma}^2$$

Ricordiamo che

$$\widehat{SE(\hat{\mu})} = \sqrt{\frac{S^2}{n}} = \frac{S}{\sqrt{n}}$$

Ricordiamo infine che

$$T = \frac{\hat{\mu} - \mu}{\widehat{SE(\hat{\mu})}} \sim t_{n-1}$$

Ovvero

$$T = \frac{\hat{\mu} - \mu}{S/\sqrt{n}} \sim t_{n-1}$$

E quindi sotto H_0

$$T = \frac{\hat{\mu} - \mu_0}{S/\sqrt{n}} \sim t_{n-1}$$

T è la statistica test che misura l'evidenza dei dati contro H_0

15.3.1 Test sulla media, ipotesi unilaterale destra, σ^2 incognito

Siano $X_1, ...X_n$ n VC, IID: $X_i \sim N(\mu, \sigma^2)$ con σ^2 incognito.

Consideriamo il seguente sistema di ipotesi

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu > \mu_0 \end{cases}$$

Sotto H_0 la statistica

$$T = \frac{\hat{\mu} - \mu_0}{S/\sqrt{n}} \sim t_{n-1}$$

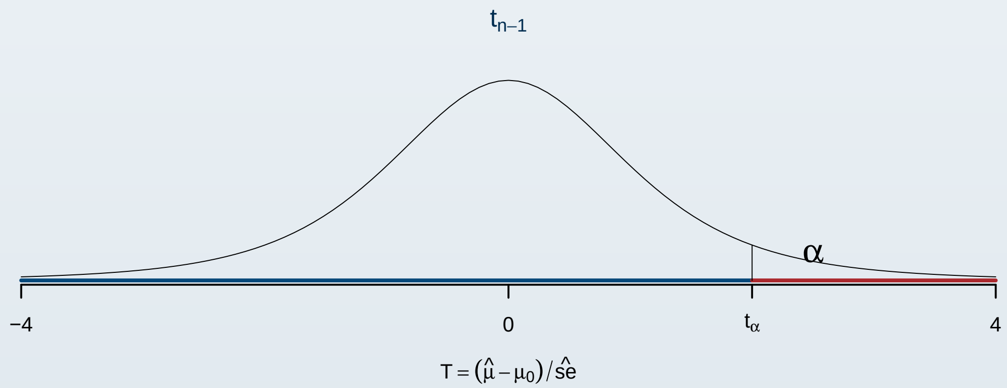
La T osservata sul campione misura l'evidenza contro H_0, tanto più è alto il valore di Z tanto più H_0 è inverosimile

Decisione sul campione Si decide un livello α e si ricava $t_{n-1;\alpha}$, si estrae un campione. Lo stimatore $\hat{\mu}$ si realizza nella media osservata del campione $\bar{x}$

$$t_{\text{obs}} = \frac{\bar{x} - \mu_0}{S/\sqrt{n}}$$

- Se $t_{\text{obs}} < t_{n-1;\alpha}$ H_0 **non viene rifiutata** al livello di significatività $\alpha \times 100\%$

- Se $t_{\text{obs}} > t_{n-1;\alpha}$ H_0 **viene rifiutata** al livello di significatività $\alpha \times 100\%$



Esempio 15.3.1. Un produttore di aspirine afferma che il prodotto lenisce il mal di testa in 30 minuti. Un campione casuale di 25 persone la usa. Risultato: $\bar{x} = 31.4$m con $\hat{\sigma} = 4.2$m. Verificare l'affermazione del produttore a un LdS del 5%.

Test t per una media, varianza incognita

$\boxed{\text{A}}$ FORMULAZIONE DELLE IPOTESI

$$\begin{cases} H_0 : \mu = \mu_0 = 30m \\ H_1 : \mu > \mu_0 = 30m \end{cases}$$

$$S = \sqrt{\frac{n}{n-1}}\,\hat{\sigma} = \sqrt{\frac{25}{25-1}} \times 4.2 = 4.287$$

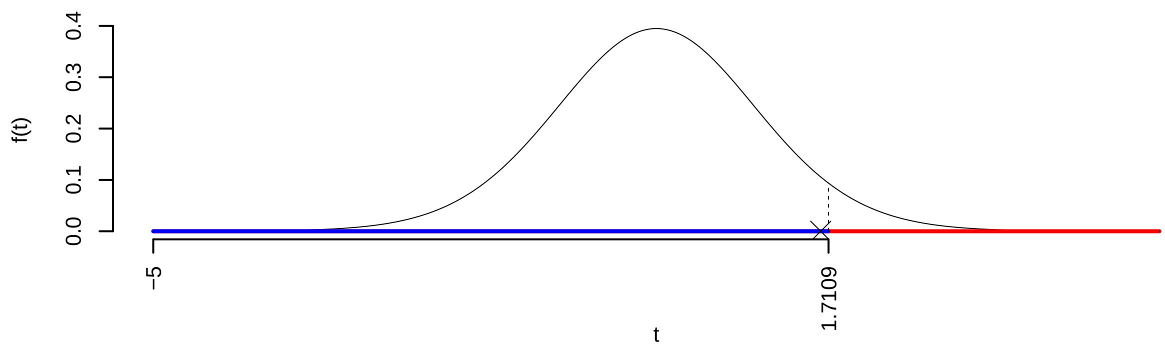
$$\frac{\hat{\mu} - \mu_0}{S/\sqrt{n}} \sim t_{n-1}$$

$$t_{\text{obs}} = \frac{(31.4 - 30)}{4.287/\sqrt{25}} = 1.633\,.$$

$\boxed{\text{C}}$ CONCLUSIONE

La siginficatitività è $\alpha = 0.05$, dalle tavole osserviamo $t_{25-1;0.05} = 1.7109$.

Essendo $t_{\text{obs}} = 1.633 < t_{25-1;0.05} = 1.7109$ allora **non** rifiuto H_0 al 5%.



Il p_{value} è

$$p_{\text{value}} = P(T_{25-1} > 1.63) = 0.057761$$

Attenzione il calcolo del p_{value} con la T è puramente illustrativo e non può essere riprodotto senza una calcolatrice statistica adeguata.

$$0.05 < p_{\text{value}} = 0.057761 \leq 0.1$$

15.3.2 Test sulla media, ipotesi unilaterale sinistra, σ^2 incognito

Siano $X_1, ...X_n$ n VC, IID: $X_i \sim N(\mu, \sigma^2)$ con σ^2 incognito. Consideriamo il seguente sistema di ipotesi

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu < \mu_0 \end{cases}$$

Sotto H_0 la statistica

$$T = \frac{\hat{\mu} - \mu_0}{S/\sqrt{n}} \sim t_{n-1}$$

Lo stimatore $\hat{\mu}$ si realizza nella media osservata del campione $\bar{x}$

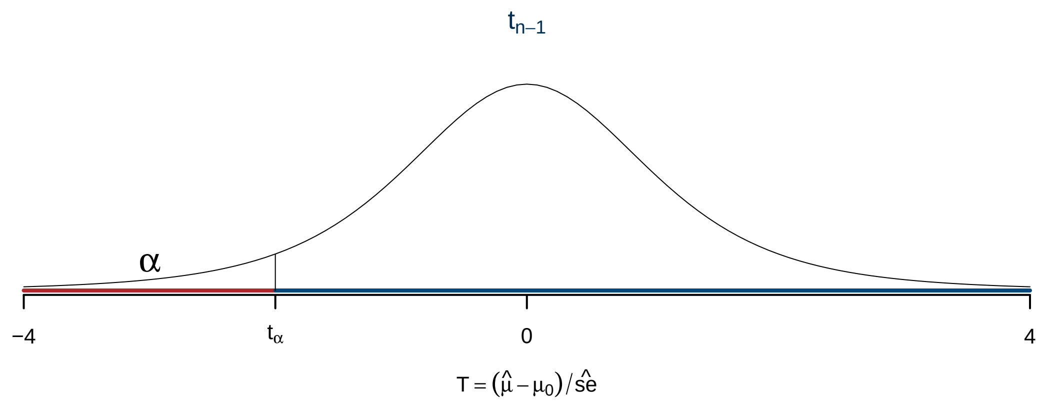
$$t_{\text{obs}} = \frac{\bar{x} - \mu_0}{S/\sqrt{n}}$$

Decisione sul campione Si decide un livello α e si ricava $t_{n-1;\alpha}$, si estrae un campione. Lo stimatore $\hat{\mu}$ si realizza nella media osservata del campione $\bar{x}$

$$t_{\text{obs}} = \frac{\bar{x} - \mu_0}{S/\sqrt{n}}$$

- Se $t_{\text{obs}} > -t_{n-1;\alpha}$ H_0 **non viene rifiutata** al livello di significatività $\alpha \times 100\%$

- Se $t_{\text{obs}} < -t_{n-1;\alpha}$ H_0 **viene rifiutata** al livello di significatività $\alpha \times 100\%$



Esempio 15.3.2. Un Centro dietetico afferma che con i loro programmi si perdono in media 2kg nella prima settimana. Si selezionano casualmente 25 soggetti, tra gli iscritti al programma.

Risultato: $\overline{\Delta x} = 1.5$kg con $\hat{\sigma} = 1.4$kg. Verificare l'affermazione del Centro al LdS del 5%.

Test t per una media, varianza incognita

$\boxed{\text{A}}$ FORMULAZIONE DELLE IPOTESI

$$\begin{cases} H_0 : \mu = \mu_0 = 2kg \\ H_1 : \mu < \mu_0 = 2kg \end{cases}$$

$$S = \sqrt{\frac{n}{n-1}}\,\hat{\sigma} = \sqrt{\frac{25}{25-1}} \times 1.4 = 1.429$$

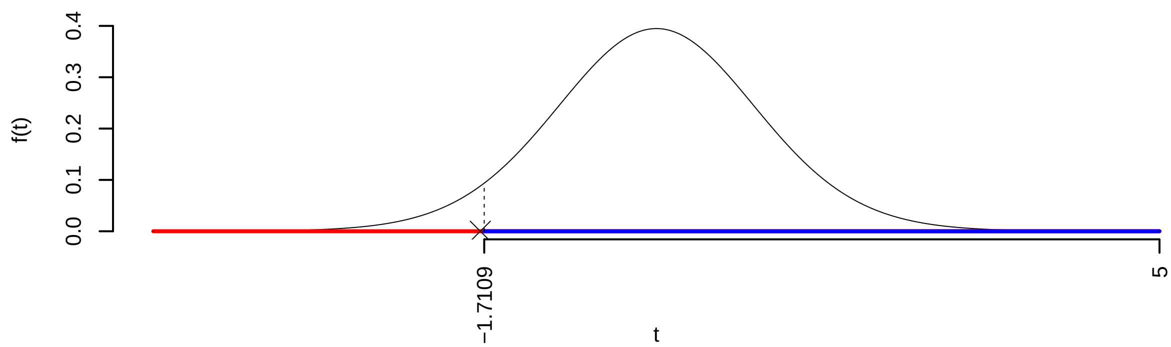
$$\frac{\hat{\mu} - \mu_0}{S/\sqrt{n}} \sim t_{n-1}$$

$$t_{\text{obs}} = \frac{(1.5-2)}{1.429/\sqrt{25}} = -1.75\,.$$

$\boxed{\text{C}}$ CONCLUSIONE

La siginficatitività è $\alpha = 0.05$, dalle tavole osserviamo $t_{25-1;0.05} = -1.7109$.

Essendo $t_{\text{obs}} = -1.7496 < t_{25-1;0.05} = -1.7109$ allora **rifiuto** H_0 al 5%.



Il p_{value} è

$$p_{\text{value}} = P(T_{25-1} < -1.75) = 0.046480$$

Attenzione il calcolo del p_{value} con la T è puramente illustrativo e non può essere riprodotto senza una calcolatrice statistica adeguata.

$$0.01 < p_{\text{value}} = 0.046480 \leq 0.05$$

15.3.3 Test sulla media, ipotesi bilaterale, σ^2 incognito

Siano $X_1, ...X_n$ n VC, IID: $X_i \sim N(\mu, \sigma^2)$ con σ^2 incognito. Consideriamo il seguente sistema di ipotesi

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases}$$

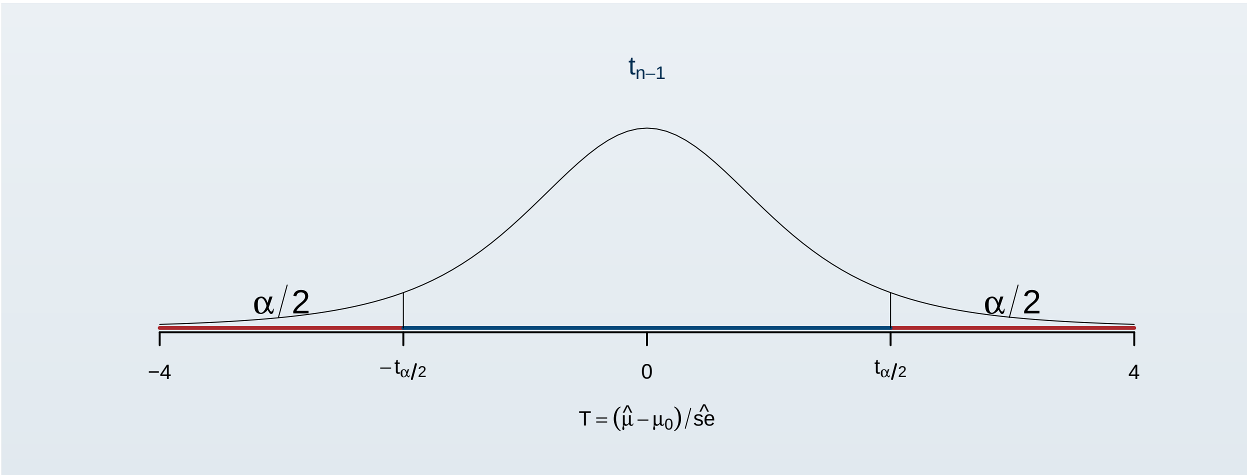
Sotto H_0 la statistica

$$T = \frac{\hat{\mu} - \mu_0}{S/\sqrt{n}} \sim t_{n-1}$$

Decisione sul campione. Si decide un livello α e si ricava $t_{n-1;\alpha}$, si estrae un campione. Lo stimatore $\hat{\mu}$ si realizza nella media osservata del campione $\bar{x}$

$$t_{\text{obs}} = \frac{\bar{x} - \mu_0}{S/\sqrt{n}}$$

- Se $-t_{n-1;\alpha/2} < t_{\text{obs}} < t_{n-1;\alpha/2}$ H_0 **non viene rifiutata** al livello di significatività $\alpha \times 100\%$
- Se $t_{\text{obs}} > t_{n-1;\alpha/2}$ o $t_{\text{obs}} < -t_{n-1;\alpha/2}$ H_0 **viene rifiutata** al livello di significatività $\alpha \times 100\%$



Esempio 15.3.3. Eseguite 17 misure di resistenza su 17 campioni di filo. Risultato: la media è 7.5N (Newton) con $\hat{\sigma} = 1.2$N.

Il filo è stato ottenuto con un nuovo procedimento, ma non si conoscono i possibili effetti sulla resistenza. Verificare, a un LdS del 5%, l'ipotesi che la resistenza media sia uguale al filo standard, che è 6.8N, contro l'alternativa che sia diverso.

Test t per una media, varianza incognita

$\boxed{\text{A}}$ FORMULAZIONE DELLE IPOTESI

$$\begin{cases} H_0 : \mu = \mu_0 = 6.8N \\ H_1 : \mu \neq \mu_0 = 6.8N \end{cases}$$

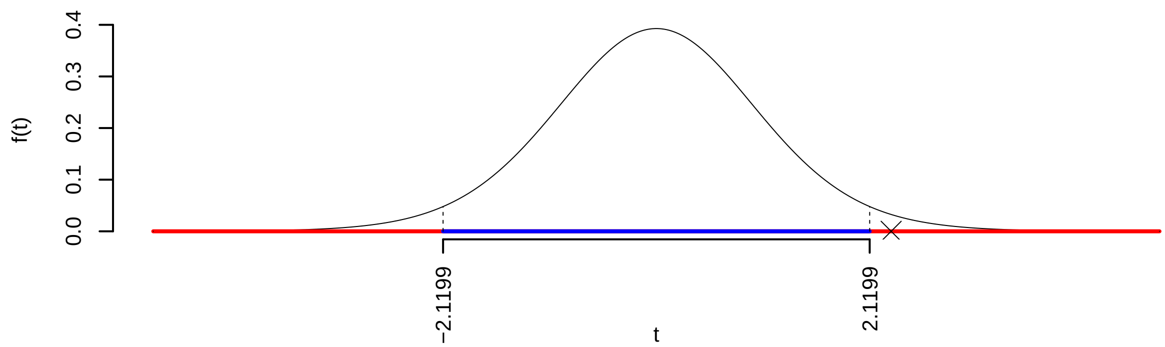
$$S = \sqrt{\frac{n}{n-1}}\,\hat{\sigma} = \sqrt{\frac{17}{17-1}} \times 1.2 = 1.237$$

$$\begin{aligned} \frac{\hat{\mu} - \mu_0}{S/\sqrt{n}} &\sim t_{n-1} \\ t_{\text{obs}} &= \frac{(7.5 - 6.8)}{1.237/\sqrt{17}} = 2.333\,. \end{aligned}$$

$\boxed{\text{C}}$ CONCLUSIONE

La siginficatitività è $\alpha = 0.05$, dalle tavole osserviamo $t_{17-1;0.025} = 2.1199$.

Essendo $|t_{\text{obs}}| = 2.3333 > t_{17-1;0.025} = 2.1199$ allora **rifiuto** H_0 al 5%.



Il p_{value} è

$$p_{\text{value}} = P(|T_{17-1}| > |2.33|) = 2P(T_{17-1} > 2.33) = 0.033005$$

Attenzione il calcolo del p_{value} con la T è puramente illustrativo e non può essere riprodotto senza una calcolatrice statistica adeguata.

$$0.01 < p_{\text{value}} = 0.033005 \leq 0.05$$

15.3.4 Significatività non fissata

Come per lo z-test se la significatività non è fissata potremmo interpretare il solo p_{value}, ma il p_{value} per la distribuzione t non è calcolabile senza un'opportuna funzione che manca nelle gran parte delle calcolatrici scientifiche.

Se non si dispone di un software adeguato si possono usare le tavole statistiche e ricavare le soglie critiche per diversi α. Una strategia comune è fissare

$$\alpha = \{1/10, 1/20, 1/100, 1/1000\} = \{0.1, 0.05, 0.01, 0.001\},$$

ricavare dalle tavole i rispettivi $t_{n-1,\alpha}$ ($t_{n-1,\alpha/2}$, se il test è bilaterale) e vedere dove cade il t_{obs}

Esempio 15.3.4. Un'azienda farmaceutica afferma che la concentrazione media di principio attivo in una certa medicina è di 50 mg. Un laboratorio indipendente decide di verificare questa affermazione eseguendo 25 analisi e ottiene una concentrazione media di 51 mg e una deviazione standard di 2.5 mg. I risultati delle analisi sono coerenti con l'affermazione dell'azienda o la concentrazione del principio attivo tende a differire dalla quantità dichiarata?

Test t per una media, varianza incognita

$\boxed{\text{A}}$ FORMULAZIONE DELLE IPOTESI

$$\begin{cases} H_0 : \mu = \mu_0 = 50\% \\ H_1 : \mu \neq \mu_0 = 50\% \end{cases}$$

$$S \quad = \quad \sqrt{\frac{n}{n-1}}\,\hat{\sigma} = \sqrt{\frac{25}{25-1}} \times 2.5 = 2.552$$

$$\begin{aligned} \frac{\hat{\mu}-\mu_0}{S/\sqrt{n}} &\sim t_{n-1} \\ t_{\text{obs}} &= \frac{(51-50)}{2.552/\sqrt{25}} = 1.96\,. \end{aligned}$$

$\boxed{\text{C}}$ CONCLUSIONE

Siccome H_1 è bilaterale, considereremo $\alpha/2$, anziché α

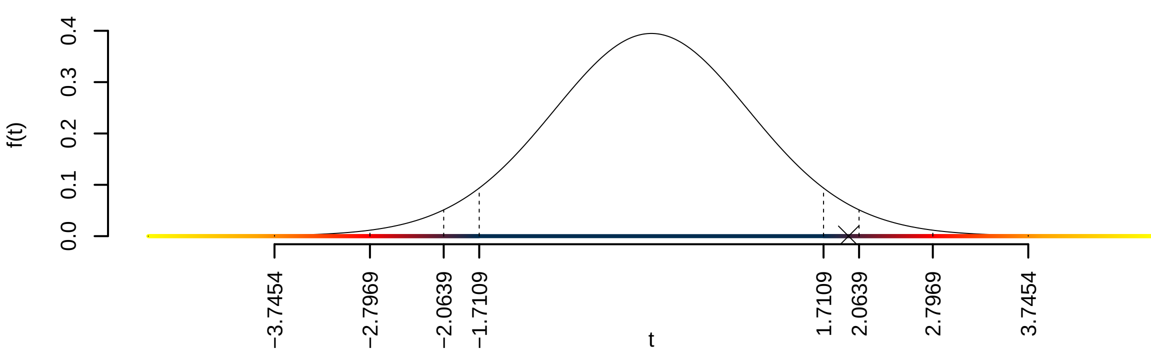
$\alpha = 0.1, 0.05, 0.01, 0.001$ e quindi $\alpha/2 = 0.05, 0.025, 0.005, 0.0005$

I valori critici sono

$t_{25-1;0.05} = 1.7109$; $t_{25-1;0.025} = 2.0639$; $t_{25-1;0.005} = 2.7969$; $t_{25-1;0.0005} = 3.7454$

Siccome $1.7109 < |t_{\text{obs}}| = 1.9596 < 2.0639$, indecisione sul rifiuto di H_0 al 10%,

$0.05 < p_{\text{value}} < 0.1$, *marginalmente significativo* $\boxed{\cdot}$.



Il p_{value} è

$$p_{\text{value}} = P(|T_{25-1}| > |1.96|) = 2P(T_{25-1} > 1.96) = 0.061756$$

Attenzione il calcolo del p_{value} con la T è puramente illustrativo e non può essere riprodotto senza una calcolatrice statistica adeguata.

$$0.05 < p_{\text{value}} = 0.061756 \leq 0.1$$

15.4 Massima verosimiglianza e test

Se $\hat{\theta}$ è stimatore di massima verosimiglianza per θ allora, per n sufficientemente grande

$$\hat{\theta} \sim N(\theta, SE^2(\hat{\theta}))$$

Sotto ipotesi H_0

$$\hat{\theta} \sim N(\theta_0, SE^2(\hat{\theta}))$$

La statistica test

$$Z = \frac{\hat{\theta} - \theta_0}{SE(\hat{\theta})} \sim N(0,1)$$

Esempio: se σ è incognita, la statistica test per μ è:

$$T = \frac{\hat{\mu} - \mu_0}{S/\sqrt{n}} \sim t_{n-1}$$

Se n diverge

$$t_{n-1} \to N(0,1)$$

Se $n > 100$ il t-test diventa lo z-test.

15.5 Test per π

Siano $X_1, ..., X_n$ n VC IID, replicazioni di $X \sim \text{Ber}(\pi)$. Lo stimatore di massima verosimiglianza

$$\hat{\pi} = \frac{1}{n}\sum_{i=1}^{n} X_i \underset{a}{\sim} N\left(\pi, \frac{\pi(1-\pi)}{n}\right)$$

Sotto H_0, $\pi = \pi_0$

$$\hat{\pi} \underset{a}{\sim} N\left(\pi_0, \frac{\pi_0(1-\pi_0)}{n}\right)$$

E quindi

$$\frac{\hat{\pi} - \pi_0}{\sqrt{\pi_0(1-\pi_0)/n}} \underset{a}{\sim} N(0,1)$$

La statistica osservata è

$$z_{\text{obs}} = \frac{\hat{\pi}_{\text{obs}} - \pi_0}{\sqrt{\pi_0(1-\pi_0)/n}}$$

A seconda di H_1 decideremo con le solite regole

Esempio 15.5.1. Lanciamo $n = 50$ un moneta di cui non siamo sicuri se è truccata oppure no. Osserviamo 30 successi su 50 lanci. Verificare l'ipotesi che la moneta sia bilanciata ($\pi = 0.5$), contro l'alternativa che sia maggiore.

Lo stimatore di massima verosimiglianza è

$$\hat{\pi} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Test Z per una proporzione

La stima

$$\hat{\pi} = \frac{30}{50} = 0.6$$

$\boxed{\text{A}}$ FORMULAZIONE DELLE IPOTESI

$$\begin{cases} H_0 : \pi = \pi_0 = 0.5 \\ H_1 : \pi > \pi_0 = 0.5 \end{cases}$$

$\boxed{\text{B}}$ SCELTA E CALCOLO STATISTICA-TEST, Z Test Binomiale per n grande: $\Rightarrow$ z-Test.

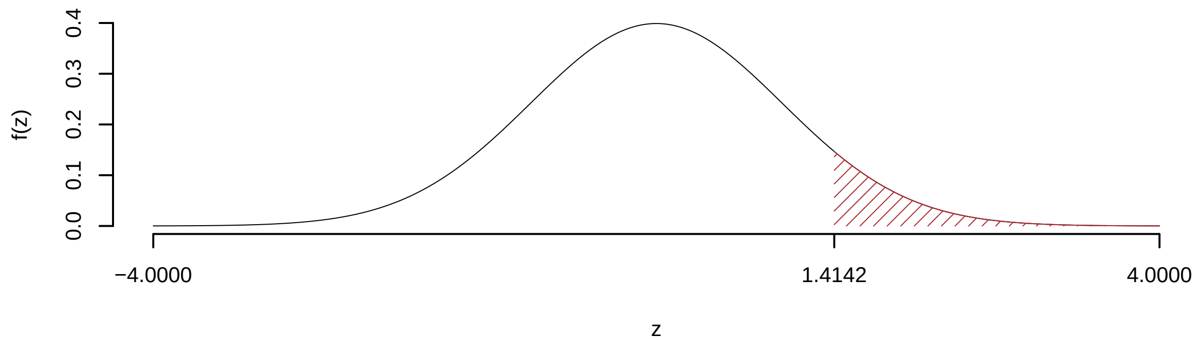
$$\begin{aligned} \frac{\hat{\pi} - \pi_0}{\sqrt{\pi_0(1-\pi_0)/\,n}} &\sim N(0,1) \\ z_{\text{obs}} &= \frac{(0.6 - 0.5)}{\sqrt{0.5(1-0.5)/50}} = 1.414\,. \end{aligned}$$

$\boxed{\text{C}}$ CONCLUSIONE

Il p_{value} è

$$p_{\text{value}} = P(Z > 1.41) = 0.078650$$

$$0.05 < p_{\text{value}} = 0.078650 \leq 0.1$$



Indecisione sul rifiuto di H_0 al 10%,

$0.05 < p_{\text{value}} < 0.1$, *marginalmente significativo* $\boxed{\cdot}$.

Esempio 15.5.2. Lanciamo $n = 100$ un moneta di cui non siamo sicuri se è truccata oppure no. Osserviamo 60 successi su 50 lanci. Verificare l'ipotesi che la moneta sia bilanciata ($\pi = 0.5$), contro l'alternativa che sia maggiore.

Lo stimatore di massima verosimiglianza è

$$\hat{\pi} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Test Z per una proporzione

La stima

$$\hat{\pi} = \frac{60}{100} = 0.6$$

$\boxed{\text{A}}$ FORMULAZIONE DELLE IPOTESI

$$\begin{cases} H_0 : \pi = \pi_0 = 0.5 \\ H_1 : \pi > \pi_0 = 0.5 \end{cases}$$

$\boxed{\text{B}}$ SCELTA E CALCOLO STATISTICA-TEST, Z Test Binomiale per n grande: $\Rightarrow$ z-Test.

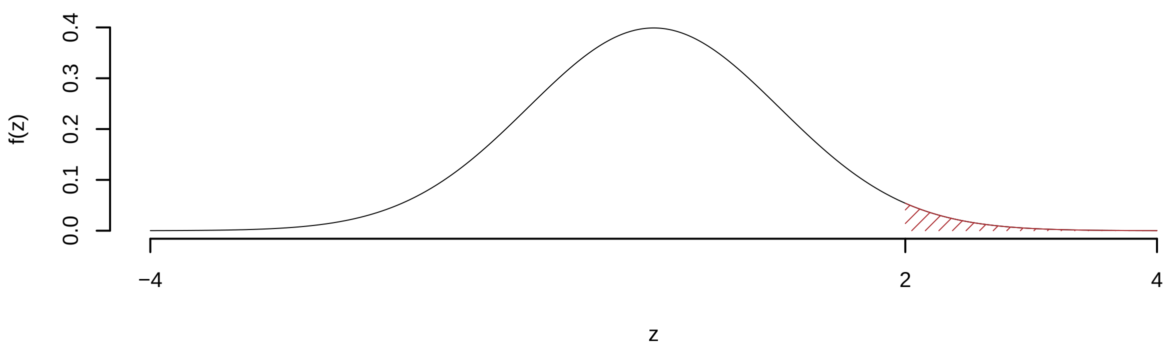
$$\begin{aligned} \frac{\hat{\pi} - \pi_0}{\sqrt{\pi_0(1-\pi_0)/n}} &\sim N(0,1) \\ z_{\text{obs}} &= \frac{(0.6-0.5)}{\sqrt{0.5(1-0.5)/100}} = 2\,. \end{aligned}$$

$\boxed{\text{C}}$ CONCLUSIONE

Il p_{value} è

$$p_{\text{value}} = P(Z > 2) = 0.022750$$

$$0.01 < p_{\text{value}} = 0.022750 \leq 0.05$$



Rifiuto H_0 al 5%,

$0.01 < p_{\text{value}} < 0.05$, *significativo* $\boxed{*}$.

Esempio 15.5.3. Lanciamo $n = 1000$ un moneta di cui non siamo sicuri se è truccata oppure no. Osserviamo 600 successi su 50 lanci. Verificare l'ipotesi che la moneta sia bilanciata ($\pi = 0.5$), contro l'alternativa che sia maggiore.

Lo stimatore di massima verosimiglianza è

$$\hat{\pi} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

Test Z per una proporzione

La stima

$$\hat{\pi} = \frac{600}{1000} = 0.6$$

$\boxed{\text{A}}$ FORMULAZIONE DELLE IPOTESI

$$\begin{cases} H_0 : \pi = \pi_0 = 0.5 \\ H_1 : \pi > \pi_0 = 0.5 \end{cases}$$

$\boxed{\text{B}}$ SCELTA E CALCOLO STATISTICA-TEST, Z Test Binomiale per n grande: $\Rightarrow$ z-Test.

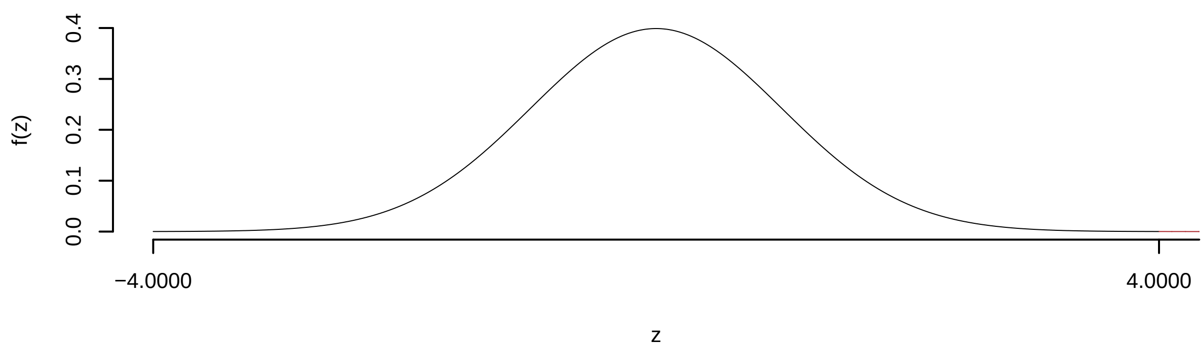
$$\frac{\hat{\pi}-\pi_0}{\sqrt{\pi_0(1-\pi_0)/\,n}} \sim N(0,1)$$
$$z_{\text{obs}} = \frac{(0.6-0.5)}{\sqrt{0.5(1-0.5)/1000}} = 6.325\,.$$

$\boxed{\text{C}}$ CONCLUSIONE

Il p_{value} è

$$p_{\text{value}} = P(Z > 6.32) = 1e-10$$

$$0 < p_{\text{value}} = 1e-10 \leq 0.001$$



Rifiuto H_0 sotto all'1‰,

$p_{\text{value}} < 0.001,$ *estremamente significativo* $\boxed{***}$.

Esempio 15.5.4. In una indagine su 100 imprese, si ha che 30 imprese decentrano la lavorazione tipo A. Il censimento precedente aveva rilevato una proporzione di decentramento, $\pi = 0.4$. Verificare l'ipotesi che il valore osservato sia dovuto al caso, contro l'alternativa che vi sia stata una diminuzione della proporzione di decentramento.

Test Z per una proporzione

La stima

$$\hat{\pi} = \frac{30}{100} = 0.3$$

$\boxed{\text{A}}$ FORMULAZIONE DELLE IPOTESI

$$\begin{cases} H_0 : \pi = \pi_0 = 0.4 \\ H_1 : \pi < \pi_0 = 0.4 \end{cases}$$

$\boxed{\text{B}}$ SCELTA E CALCOLO STATISTICA-TEST, Z Test Binomiale per n grande: $\Rightarrow$ z-Test.

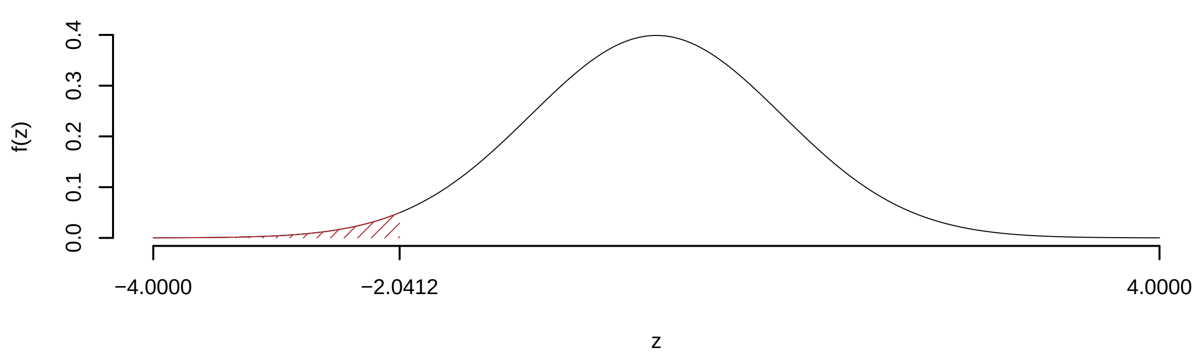
$$\begin{aligned} \frac{\hat{\pi} - \pi_0}{\sqrt{\pi_0(1-\pi_0)/n}} &\sim N(0,1) \\ z_{\text{obs}} &= \frac{(0.3-0.4)}{\sqrt{0.4(1-0.4)/100}} = -2.041\,. \end{aligned}$$

$\boxed{\text{C}}$ CONCLUSIONE

Il p_{value} è

$$p_{\text{value}} = P(Z < -2.04) = 0.020613$$

$$0.01 < p_{\text{value}} = 0.020613 \leq 0.05$$



Rifiuto H_0 al 5%,

$0.01 < p_{\text{value}} < 0.05$, *significativo* $\boxed{*}$.

Esempio 15.5.5. Una città è composta da 50000 soggetti. Si estrae un campione casuale di 100 soggetti e si trova che 20 soggetti possiedono una connessione a banda ultra-larga. La proporzione posseduta a livello nazionale è pari al 25%. Verificare che nella città la percentuale sia diversa (in più o in meno).

Test Z per una proporzione

La stima

$$\hat{\pi} = \frac{20}{100} = 0.2$$

$\boxed{\text{A}}$ FORMULAZIONE DELLE IPOTESI

$$\begin{cases} H_0 : \pi = \pi_0 = 0.25 \\ H_1 : \pi \neq \pi_0 = 0.25 \end{cases}$$

$\boxed{\text{B}}$ SCELTA E CALCOLO STATISTICA-TEST, Z Test Binomiale per n grande: $\Rightarrow$ z-Test.

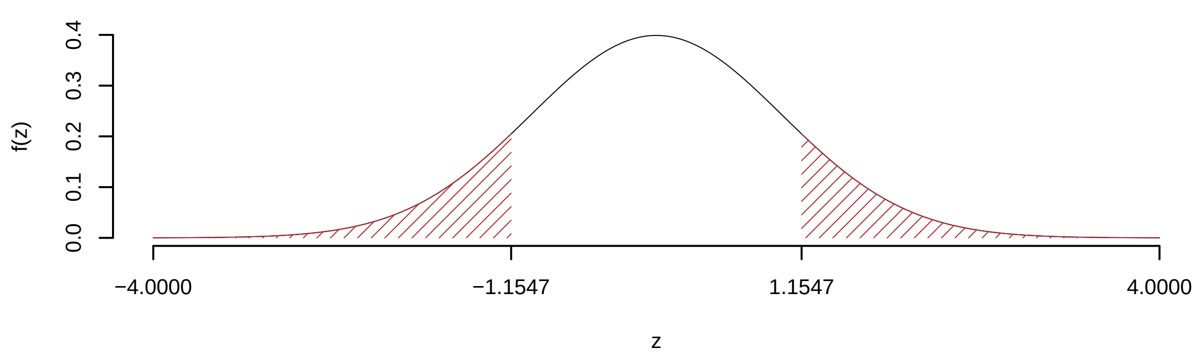
$$\begin{aligned} \frac{\hat{\pi} - \pi_0}{\sqrt{\pi_0(1-\pi_0)/\,n}} \quad &\sim \quad N(0,1) \\ z_{\text{obs}} \quad &= \quad \frac{(0.2 - 0.25)}{\sqrt{0.25(1-0.25)/100}} = -1.155\,. \end{aligned}$$

$\boxed{\text{C}}$ CONCLUSIONE

Il p_{value} è

$$p_{\text{value}} = P(|Z| > |-1.15|) = 2P(Z > 1.15) = 0.248213$$

$$0.1 < p_{\text{value}} = 0.248213 \leq 1$$



Non rifiuto H_0 a **nessun** livello di significatività,

$p_{\text{value}} > 0.1$, *non significativo*

15.6 Specchietto Finale per i Test ad un Campione

$H_0 : \mu = \mu_0$	σ^2	Dist.	Statistica Test	Zona Rifiuto	p_{value}
$H_1 : \mu > \mu_0$	Noto	Z	$z_{\text{obs}} = \dfrac{\hat{\mu} - \mu_0}{\sigma/\sqrt{n}}$	$z_{\text{obs}} > z_\alpha$	$p_{\text{value}} = P(Z > z_{\text{obs}})$
$H_1 : \mu < \mu_0$	Noto	Z	$z_{\text{obs}} = \dfrac{\hat{\mu} - \mu_0}{\sigma/\sqrt{n}}$	$z_{\text{obs}} < -z_\alpha$	$p_{\text{value}} = P(Z < z_{\text{obs}})$
$H_1 : \mu \neq \mu_0$	Noto	Z	$z_{\text{obs}} = \dfrac{\hat{\mu} - \mu_0}{\sigma/\sqrt{n}}$	$\lvert z_{\text{obs}}\rvert > \lvert z_{\alpha/2}\rvert$	$p_{\text{value}} = 2P(Z > \lvert z_{\text{obs}}\rvert)$
$H_1 : \mu > \mu_0$	Incognito	t_{n-1}	$z_{\text{obs}} = \dfrac{\hat{\mu} - \mu_0}{S/\sqrt{n}}$	$t_{\text{obs}} > t_{n-1;\ \alpha}$	$p_{\text{value}} = P(T > t_{\text{obs}})$
$H_1 : \mu < \mu_0$	Incognito	t_{n-1}	$z_{\text{obs}} = \dfrac{\hat{\mu} - \mu_0}{S/\sqrt{n}}$	$t_{\text{obs}} < -t_{n-1;\ \alpha}$	$p_{\text{value}} = P(T < t_{\text{obs}})$
$H_1 : \mu \neq \mu_0$	Incognito	t_{n-1}	$z_{\text{obs}} = \dfrac{\hat{\mu} - \mu_0}{S/\sqrt{n}}$	$\lvert t_{\text{obs}}\rvert > \lvert t_{n-1;\ \alpha/2}\rvert$	$p_{\text{value}} = 2P(T > \lvert t_{\text{obs}}\rvert)$
$H_0 : \pi = \pi_0$		Dist.	Statistica Test	Zona Rifiuto	p_{value}
$H_1 : \pi > \pi_0$		Z	$z_{\text{obs}} = \dfrac{\hat{\pi} - \pi_0}{\sqrt{\pi(/1-\pi)}/\sqrt{n}}$	$z_{\text{obs}} > z_\alpha$	$p_{\text{value}} = P(Z > z_{\text{obs}})$
$H_1 : \pi < \pi_0$		Z	$z_{\text{obs}} = \dfrac{\hat{\pi} - \pi_0}{\sqrt{\pi(/1-\pi)}/\sqrt{n}}$	$z_{\text{obs}} < -z_\alpha$	$p_{\text{value}} = P(Z < z_{\text{obs}})$
$H_1 : \pi \neq \pi_0$		Z	$z_{\text{obs}} = \dfrac{\hat{\pi} - \pi_0}{\sqrt{\pi(/1-\pi)}/\sqrt{n}}$	$\lvert z_{\text{obs}}\rvert > \lvert z_{\alpha/2}\rvert$	$p_{\text{value}} = 2P(Z > \lvert z_{\text{obs}}\rvert)$

Confronto tra due Popolazioni 16

Abbiamo trattato un gruppo di 15 pazienti con un placebo e un secondo gruppo di 10 pazienti con il farmaco A, il numero medio di giorni dei pazienti trattati con il placebo è di 10.4 giorni, con una sd di 2.3 giorni; il numero medio di giorni dei pazienti trattati con il farmaco è di 9.6 giorni, con una sd di 1.9 giorni. - La differenza tra 10.4 e 9.6 è **significativa**? - Il farmaco davvero diminuisce il numero medio di giorni oppure la differenza è colpa del caso?

In un'indagine sul reddito estraiamo un campione di 56 individui dalla città A e un secondo campione di 67 individui dalla città B. Il reddito medio del campione estratto da A è $\bar{x}_A = 15.1$mila euro lordi annui, con una sd $\hat{\sigma}_A = 3.1$mila euro; Il reddito medio del campione estratto da B è $\bar{x}_A = 18.6$mila euro lordi annui, con una sd $\hat{\sigma}_A = 5.7$mila euro; - Le due città hanno lo stesso reddito medio e la differenza nei campioni è dovuta ala caso oppure i due campioni provengono da due città con reddito medio diverso?

Facciamo un sondaggio di opinione sul gradimento di un personaggio X in due provincie, A e B. Nella provincia A 35 persone su 130 intervistate gradisce il personaggio X, nella provincia B 54 su 150 persone intervistate gradisce il personaggio X. - la differenza tra i due campioni è significativa? - nella provincia A e in quella B la proporzione di gradimento è la stessa e la differenza nei campioni è dovuta al caso oppure no?

16.1 Test per due medie

16.1.1 il contesto probabilistico

Siano $X_{1,A}, X_{2,A}, ..., X_{n_A,A}$ n_A VC IID replicazioni di $X_A \sim N(\mu_A, \sigma_A^2)$ e siano $X_{1,B}, X_{2,B}, ..., X_{n_B,B}$ n_B VC IID replicazioni di $X_B \sim N(\mu_B, \sigma_B^2)$

$X_{1,A}, X_{2,A}, ..., X_{n_A,A}$ è un campione di ampiezza n_A dalla popolazione $\mathscr{P}_A$ e $X_{1,B}, X_{2,B}, ..., X_{n_B,B}$ è un campione di ampiezza n_B dalla popolazione $\mathscr{P}_B$

Ci possiamo chiedere se:

$$\begin{cases} H_0 : \mu_A = \mu_B \\ H_1 : \mu_A > \mu_B \end{cases}$$

Test Unilaterale

Oppure se

$$\begin{cases} H_0 : \mu_A = \mu_B \\ H_1 : \mu_A \neq \mu_B \end{cases}$$

Test Bilaterale

Perché solo due possibili H_1? Perché l'ipotesi

$$\begin{cases} H_0 : \mu_A = \mu_B \\ H_1 : \mu_A < \mu_B \end{cases}$$

equivale a

$$\begin{cases} H_0 : \mu_B = \mu_A \\ H_1 : \mu_B > \mu_A \end{cases}$$

16.1.2 Derivazione della statistica test

Se $X_{1,A}, X_{2,A}, ..., X_{n_A,A}$ n_A VC IID replicazioni di $X_A \sim N(\mu_A, \sigma_A^2)$ e siano $X_{1,B}, X_{2,B}, ..., X_{n_B,B}$ n_B VC IID replicazioni di $X_B \sim N(\mu_B, \sigma_B^2)$, allora

$$\hat{\mu}_A \sim N\left(\mu_A, \frac{\sigma_A^2}{n_A}\right) \qquad \text{e} \qquad \hat{\mu}_B \sim N\left(\mu_B, \frac{\sigma_B^2}{n_B}\right)$$

E dunque, dalle proprietà delle normali

$$\hat{\mu}_A - \hat{\mu}_B \sim N\left(\mu_A - \mu_B, \frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}\right)$$

E quindi

$$\frac{(\hat{\mu}_A - \hat{\mu}_B) - (\mu_A - \mu_B)}{\sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}} \sim N(0,1)$$

Sotto $H_0 : \mu_A = \mu_B$

$$Z = \frac{\hat{\mu}_A - \hat{\mu}_B}{\sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}} \sim N(0,1)$$

Ma σ_A e σ_B sono incogniti

16.1.3 Stima di σ_A e σ_B

Dipende dalle ipotesi che abbiamo sulle due popolazioni:

Ipotesi 1: omogeneità: si ipotizza che in popolazione $\sigma_A^2 = \sigma_B^2 = \sigma^2$, e dunque sia il campione proveniente da A che quello proveniente da B contribuiscono a stimare la varianza comune di popolazione σ^2.

Ipotesi 2: eterogeneità: si ipotizza che in popolazione $\sigma_A^2 \neq \sigma_B^2$

In teoria per scegliere tra l'ipotesi di omogeneità e quella di eterogeneità dovremmo fare un test

$$\begin{cases} H_0 : \sigma_A^2 = \sigma_B^2 \\ H_1 : \sigma_A^2 \neq \sigma_B^2 \end{cases}$$

il cui sviluppo esula dagli scopi del corso.

In tutti gli esercizi che faremo l'ipotesi verrà **assunta** nel problema.

16.1.4 Ipotesi 1: omogeneità

Sotto ipotesi di omogeneità, entrambi i campioni contribuiscono alla stima della stima comune di σ^2, lo stimatore congiunto (*pooled*) è

$$S_p^2 = \frac{n_A \hat{\sigma}_A^2 + n_B \hat{\sigma}_B^2}{n_A + n_B - 2}$$

E quindi sotto H_0

$$T = \frac{\hat{\mu}_A - \hat{\mu}_B}{\sqrt{\frac{S_p^2}{n_A} + \frac{S_p^2}{n_B}}} \sim t_{n_A + n_B - 2}$$

La statistica osservata

$$t_{\text{obs}} = \frac{\bar{x}_A - \bar{x}_B}{\sqrt{\frac{S_p^2}{n_A} + \frac{S_p^2}{n_B}}}$$

Andrà letta nella direzione di H_1, sulle tavole della t con $n_A + n_B - 2$ gradi di libertà.

16.1.5 Ipotesi 2: eterogeneità

Sotto ipotesi di eterogeneità, costruiamo gli stimatori corretti per σ_A^2 e σ_B^2

$$S_A^2 = \frac{n_A}{n_A - 1} \hat{\sigma}_A^2 \qquad S_B^2 = \frac{n_B}{n_B - 1} \hat{\sigma}_B^2$$

E quindi sotto H_0

$$T = \frac{\hat{\mu}_A - \hat{\mu}_B}{\sqrt{\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}}} \sim t_{n_A + n_B - 2}$$

La statistica osservata

$$t_{\text{obs}} = \frac{\bar{x}_A - \bar{x}_B}{\sqrt{\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}}}$$

Andrà letta nella direzione di H_1, sulle tavole della t con $n_A + n_B - 2$ gradi di libertà.

16.1.6 Esempio

Si sperimentano due diete: A e B. Per la dieta A si selezionano a caso 15 soggetti. Dopo due settima- ne, si osserva una diminuzione di peso: $\bar{x}_A = 6$kg con $\hat{\sigma}_A = 1.2$kg. Per la dieta B si selezionano a caso 18 soggetti. Dopo due settimane, si osserva una diminuzione di peso: $\bar{x}_B = 5$kg con $\hat{\sigma}_B = 1.8$kg. Nell'ipotesi di varianze **eterogenee**, verificare, se le due diete sono equivalenti contro l'alternativa che la dieta A sia più efficace (maggiore) della dieta B.

Test t per due medie, (eterogeneità)

$\boxed{\text{A}}$ FORMULAZIONE DELLE IPOTESI

$$\begin{cases} H_0 : \mu_{\mathrm{A}} = \mu_{\mathrm{B}} \\ H_1 : \mu_{\mathrm{A}} > \mu_{\mathrm{B}} \end{cases}$$

$\boxed{\mathrm{B}}$ SCELTA E CALCOLO STATISTICA-TEST, T

$$S^2_{A} = \frac{n_{A}}{n_{A} - 1} \hat{\sigma}^2_{A} = \frac{15}{15 - 1} 1.2^2 = 1.543 \qquad S^2_{B} = \frac{n_{B}}{n_{B} - 1} \hat{\sigma}^2_{B} = \frac{18}{18 - 1} 1.8^2 = 3.431$$

$$\begin{aligned} \frac{\hat{\mu}_{A} - \hat{\mu}_{B}}{\sqrt{\frac{S^2_{A}}{n_{A}} + \frac{S^2_{B}}{n_{B}}}} &\sim t_{n_{A} + n_{B} - 2} \\ t_{\mathrm{obs}} &= \frac{(6 - 5)}{\sqrt{\frac{1.543}{15} + \frac{3.431}{18}}} = 1.846\,. \end{aligned}$$

$\boxed{\mathrm{C}}$ CONCLUSIONE

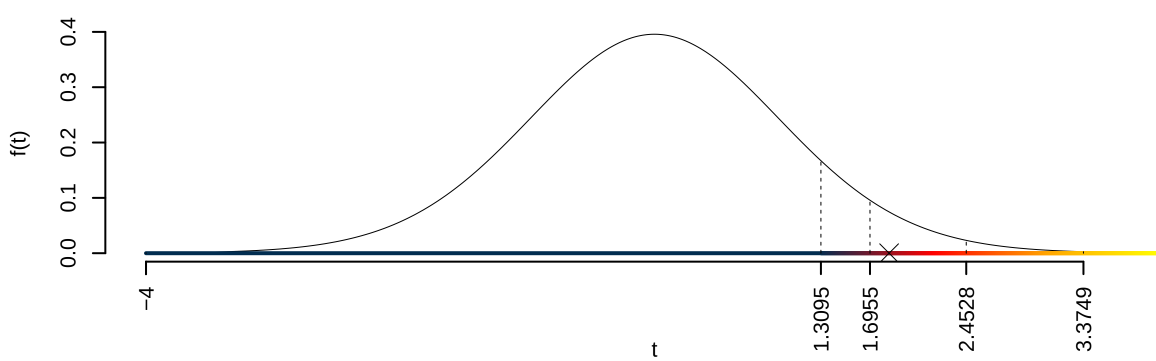
Consideriamo $\alpha = 0.1, 0.05, 0.01, 0.001$

I valori critici sono

$t_{33-2;0.1} = 1.3095$; $t_{33-2;0.05} = 1.6955$; $t_{33-2;0.01} = 2.4528$; $t_{33-2;0.001} = 3.3749$

Siccome $1.6955 < t_{\mathrm{obs}} = 1.846 < 2.4528$, quindi **rifiuto** H_0 al 5%,

$0.01 < p_{\mathrm{value}} < 0.05$, *significativo* $\boxed{*}$.



Il p_{value} è

$$p_{\mathrm{value}} = P(T_{33-2} > 1.85) = 0.037229$$

Attenzione il calcolo del p_{value} con la T è puramente illustrativo e non può essere riprodotto senza una calcolatrice statistica adeguata.

$$0.01 < p_{\text{value}} = 0.037229 \leq 0.05$$

16.1.7 Esempio

La direzione vuole verificare se l'ammontare delle vendite di due supermercati, A e B, sia la stessa. Un campione di 18 giorni per il supermercato A fornisce una vendita media giornaliera pari a $\bar{x}_A =$ 55 mila euro, con $\hat{\sigma}_A = 2.9$ mila euro. Un campione di 24 giorni per il supermercato B fornisce $\bar{x}_B = 57$ mila euro, con $\hat{\sigma}_A = 3.1$ mila euro.

Sotto assunto di **omogeneità** delle varianze verificare l'ipotesi che la vendita media del supermercato A sia uguale a quella del supermercato B, contro l'alternativa sia diversa.

Test T per due medie, (omogeneità)

$\boxed{\text{A}}$ FORMULAZIONE DELLE IPOTESI

$$\begin{cases} H_0 : \mu_A = \mu_B \\ H_1 : \mu_A \neq \mu_B \end{cases}$$

$\boxed{\text{B}}$ SCELTA E CALCOLO STATISTICA-TEST, T

L'ipotesi è di omogeneità e quindi calcoliamo:

$$S_p^2 = \frac{n_A \hat{\sigma}^2_A + n_B \hat{\sigma}^2_B}{n_A + n_B - 2} = \frac{18 \cdot 2.9^2 + 24 \cdot 3.1^2}{18 + 24 - 2} = 9.551$$

$$\frac{\hat{\mu}_A - \hat{\mu}_B}{\sqrt{\frac{S_p^2}{n_A} + \frac{S_p^2}{n_B}}} \sim t_{n_A + n_B - 2}$$

$$t_{\text{obs}} = \frac{(55 - 57)}{\sqrt{\frac{8.905}{18} + \frac{10.03}{24}}} = -2.076\,.$$

$\boxed{\text{C}}$ CONCLUSIONE

Siccome H_1 è bilaterale, considereremo $\alpha/2$, anziché α

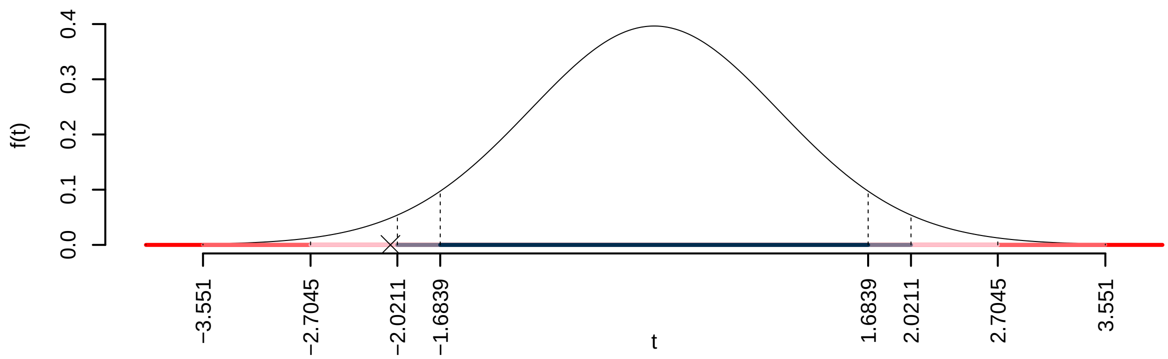
$\alpha = 0.1, 0.05, 0.01, 0.001$ e quindi $\alpha/2 = 0.05, 0.025, 0.005, 0.0005$

I valori critici sono

$t_{42-2;0.05} = 1.6839$; $t_{42-2;0.025} = 2.0211$; $t_{42-2;0.005} = 2.7045$; $t_{42-2;0.0005} = 3.551$

Siccome $2.0211 < |t_{\text{obs}}| = 2.0756 < 2.7045$, quindi **rifiuto** H_0 al 5%,

$0.01 < p_{\text{value}} < 0.05$, *significativo* $\boxed{*}$.



Il p_{value} è

$$p_{\text{value}} = P(|T_{42-2}| > |-2.08|) = 2P(T_{42-2} > 2.08) = 0.044408$$

Attenzione il calcolo del p_{value} con la T è puramente illustrativo e non può essere riprodotto senza una calcolatrice statistica adeguata.

$$0.01 < p_{\text{value}} = 0.044408 \leq 0.05$$

16.2 Test per due proporzioni

16.2.1 Il contesto probabilistico

Siano $X_{1,A}, X_{2,A}, ..., X_{n_A,A}$ n_A VC IID replicazioni di $X_A \sim \text{Ber}(\pi_A)$ e siano $X_{1,B}, X_{2,B}, ..., X_{n_B,B}$ n_B VC IID replicazioni di $X_B \sim \text{Ber}(\pi_B)$

$X_{1,A}, X_{2,A}, ..., X_{n_A,A}$ è un campione di ampiezza n_A dalla popolazione $\mathscr{P}_A$ e $X_{1,B}, X_{2,B}, ..., X_{n_B,B}$ è un campione di ampiezza n_B dalla popolazione $\mathscr{P}_B$

Ci possiamo chiedere se:

$$\begin{cases} H_0 : \pi_A = \pi_B \\ H_1 : \pi_A > \pi_B \end{cases}$$

Test Unilaterale

Oppure se

$$\begin{cases} H_0 : \pi_A = \pi_B \\ H_1 : \pi_A \neq \pi_B \end{cases}$$

Test Bilaterale

16.2.2 Derivazione della statistica test

Se $X_{1,A}, X_{2,A}, ..., X_{n_A,A}$ n_A VC IID replicazioni di $X_A \sim \text{Ber}(\pi_A)$ e siano $X_{1,B}, X_{2,B}, ..., X_{n_B,B}$ n_B VC IID replicazioni di $X_B \sim \text{Ber}(\pi_B)$, allora

$$\hat{\pi}_A \underset{a}{\sim} N\left(\pi_A, \frac{\pi_A(1-\pi_A)}{n_A}\right) \qquad \text{e} \qquad \hat{\pi}_B \underset{a}{\sim} N\left(\pi_B, \frac{\pi_B(1-\pi_B)}{n_B}\right)$$

E dunque, dalle proprietà delle normali

$$\hat{\pi}_A - \hat{\pi}_B \underset{a}{\sim} N\left(\pi_A - \pi_B, \frac{\pi_A(1-\pi_A)}{n_A} + \frac{\pi_B(1-\pi_B)}{n_B}\right)$$

E quindi

$$\frac{(\hat{\pi}_A - \hat{\pi}_B) - (\pi_A - \pi_B)}{\sqrt{\frac{\pi_A(1-\pi_A)}{n_A} + \frac{\pi_B(1-\pi_B)}{n_B}}} \sim N(0,1)$$

Sotto $H_0 : \pi_A = \pi_B = \pi_C$

$$Z = \frac{\hat{\pi}_A - \hat{\pi}_B}{\sqrt{\frac{\pi_C(1-\pi_C)}{n_A} + \frac{\pi_C(1-\pi_C)}{n_B}}} \sim N(0,1)$$

La stima di π_C è

$$\hat{\pi}_C = \frac{\#\{\text{successi nel gruppo A}\} + \#\{\text{successi nel gruppo B}\}}{n_A + n_B} = \frac{n_A\hat{\pi}_A + n_B\hat{\pi}_B}{n_A + n_B}$$

E dunque

$$Z = \frac{\hat{\pi}_A - \hat{\pi}_B}{\sqrt{\frac{\hat{\pi}_C(1-\hat{\pi}_C)}{n_A} + \frac{\hat{\pi}_C(1-\hat{\pi}_C)}{n_B}}} \underset{a}{\sim} N(0,1)$$

La statistica osservata è

$$z_{\text{obs}} = \frac{\hat{\pi}_A - \hat{\pi}_B}{\sqrt{\frac{\hat{\pi}_C(1-\hat{\pi}_C)}{n_A} + \frac{\hat{\pi}_C(1-\hat{\pi}_C)}{n_B}}}$$

e andrà letta nella direzione di H_1 sulle tavole della Z.

16.2.3 Esempio

Tra i abitanti del comune A, si intervistano 80 uomini e 100 donne per capire l'impatto che avrà la legge comunale sul divieto di fumo nei parchi pubblici. Dalle interviste risulta che 70 uomini e 70 donne si dichiarano a favore di tale legge. Verificare l'ipotesi che la nuova legge sia accolta in modo equivalente da donne e uomini, contro l'alternativa che le donne si dimostrino meno propense a accettare tale legge.

Test Z per due proporzioni

$\boxed{\text{A}}$ FORMULAZIONE DELLE IPOTESI

$$\begin{cases} H_0 : \pi_U = \pi_D \\ H_1 : \pi_U > \pi_D \end{cases}$$

$\boxed{\text{B}}$ SCELTA E CALCOLO STATISTICA-TEST, Z

$$\hat{\pi}_U = \frac{s_U}{n_U} = \frac{70}{80} = 0.875 \qquad \hat{\pi}_D = \frac{s_D}{n_D} = \frac{70}{100} = 0.7$$

Calcoliamo la proporzione comune sotto H_0

$$\pi_C = \frac{s_U + s_D}{n_U + n_D} = \frac{140}{180} = 0.7778$$

$$\frac{\hat{\pi}_U - \hat{\pi}_D}{\sqrt{\frac{\pi_C(1-\pi_C)}{n_U} + \frac{\pi_C(1-\pi_C)}{n_D}}} \sim N(0,1)$$

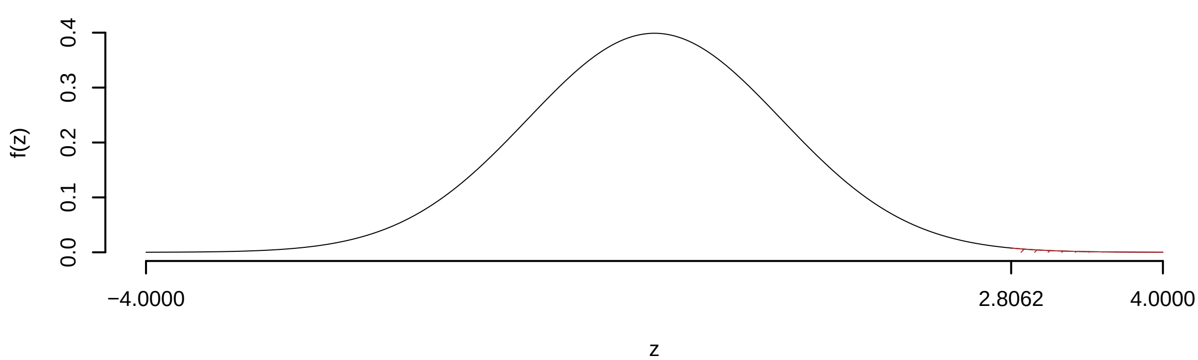
$$z_{\text{obs}} = \frac{(0.875 - 0.7)}{\sqrt{\frac{0.7778(1-0.7778)}{80} + \frac{0.7778(1-0.7778)}{100}}} = 2.806\,.$$

$\boxed{\text{C}}$ CONCLUSIONE

Il p_{value} è

$$p_{\text{value}} = P(Z > 2.81) = 0.002506$$

$$0.001 < p_{\text{value}} = 0.002506 \leq 0.01$$



Rifiuto H_0 all'1%,

$0.001 < p_{\text{value}} < 0.01$, *molto significativo* $\boxed{**}$.

16.3 Specchietto Finale per i Test ad Due Campioni

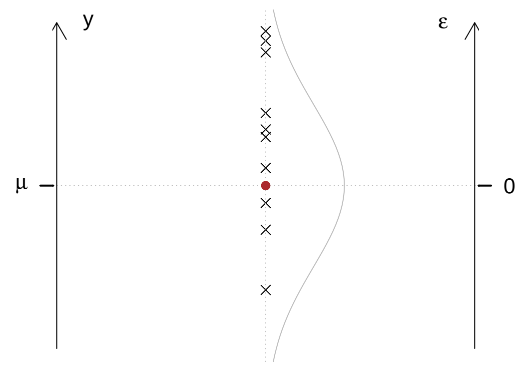
Test t, 2 Campioni Omogeneità	Test t, 2 Campioni Eterogeneità	Proporzione, 2 Campioni
$t_{\text{obs}} = \dfrac{\hat{\mu}_A - \hat{\mu}_B}{\sqrt{\frac{S_p^2}{n_A} + \frac{S_p^2}{n_B}}}$ $S_p^2 = \dfrac{n_A \hat{\sigma}_A^2 + n_B \hat{\sigma}_B^2}{n_A + n_B - 2}$	$t_{\text{obs}} = \dfrac{\hat{\mu}_A - \hat{\mu}_B}{\sqrt{\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}}}$	$z_{\text{obs}} = \dfrac{\hat{\pi}_A - \hat{\pi}_B}{\sqrt{\frac{\hat{\pi}_C(1-\hat{\pi}_C)}{n_A} + \frac{\hat{\pi}_C(1-\hat{\pi}_C)}{n_B}}}$ $\hat{\pi}_C = \dfrac{\#\{\text{successi A}\} + \#\{\text{successi B}\}}{n_A + n_B} = \dfrac{n_A \hat{\pi}_A + n_B \hat{\pi}_B}{n_A + n_B}$

Regressione Lineare 17

17.1 Il modello d'errore

Siano $Y_1, ..., Y_n$ n VC IID, replicazioni tc $Y_i \sim N(\mu, \sigma^2_\varepsilon)$, dalle proprietà della normale possiamo riscrivere:

$$Y_i = \mu + \varepsilon_i \qquad \varepsilon_i \sim N(0, \sigma^2_\varepsilon)$$



$$Y_i \sim N(\mu, \sigma^2_\varepsilon) \quad \text{è equivalente a dire} \quad Y_i = \mu + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2_\varepsilon)$$

17.1.1 Esempi

Esempio 17.1.1. Stiamo studiando la produttività media di un ettaro coltivato ad una certa varietà di riso: per prima cosa piantiamo 10 ettari **non** trattati con fertilizzante $(X = 0)$ con questa varietà e calcoliamo i quintali per ettaro, otteniamo (28.22, 27.46, 28.89, 28.6, 29.64, 28.69, 26.72, 27.79, 29.9, 29.78)

$$Y_i = \mu_{(X=0)} + \varepsilon_i \qquad \varepsilon_i \sim N(0, \sigma^2_\varepsilon)$$

Stimiamo

$$\hat{\mu}_{(X=0)} = \frac{1}{n}\sum_{i=1}^{n} y_i = 28.569$$

e

$$\hat{\sigma}^2_\varepsilon = \frac{1}{n}\sum_{i=1}^{n} y_i^2 - \hat{\mu}^2_{(X=0)}$$

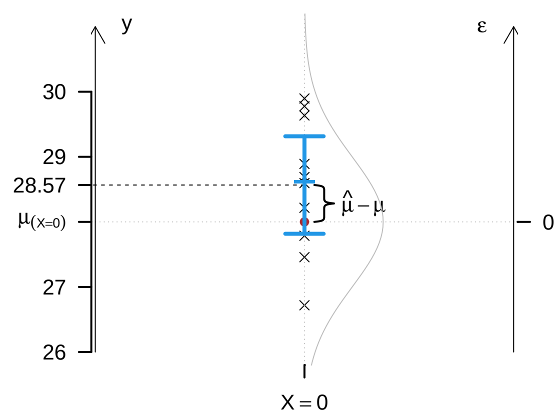
$$
\begin{aligned}
&= \frac{1}{10}(28.22^2 + 27.46^2 + 28.89^2 + ... + 29.9^2 + 29.78^2) - 28.569^2 \\
&= 0.9881 \\
\hat{\sigma}_\varepsilon &= 0.994
\end{aligned}
$$

Costruiamo dapprima un intervallo di confidenza sui dati Correggiamo $\hat{\sigma}^2_\varepsilon$

$$
\begin{aligned}
S^2_{(X=0)} &= \frac{n}{n-1}\hat{\sigma}^2_{(X=0)} \\
&= \frac{10}{10-1} \cdot 0.9881 \\
&= 1.0983 \\
S_{(X=0)} &= 1.048
\end{aligned}
$$

L'intervallo di confidenza al 95% per $\mu_{(X=0)}$ è

$$
\hat{\mu}_{(X=0)} \pm t_{n-1;\alpha/2}\frac{S}{\sqrt{n}} = 28.569 \pm 2.2622\frac{1.048}{10} = (27.8193; 29.3187)
$$



E quindi potremmo proporre la **previsione**: quanta produzione ci aspetteremo sul prossimo ettaro che pianteremo con quella varietà di riso?

$$
\begin{aligned}
E(\hat{Y}_{11}) &= E(\hat{\mu}_{(X=0)} + \varepsilon_{11}) \\
&= 28.569 + E(\varepsilon_{11}) \\
&= 28.569 \qquad \text{poiché } \varepsilon_{11} \sim N\left(0, \sigma^2_\varepsilon\right)
\end{aligned}
$$

È una *previsione* corretta?

$$E(\hat{Y}_{11}) = E(\hat{\mu} + \varepsilon_{11}) = E(\hat{\mu}) + E(\varepsilon_{11}) = \mu + 0 = \mu$$

L'errore di previsione stimato è $S_{(X=0)} = 1.048$.

Esempio 17.1.2. Supponiamo di aver piantato altri 10 ettari di questa varietà ma su terreni trattati con 1.13 hg ($X = 1.13$) di concime azotato per ettaro e osserviamo (28.16, 30.38, 27.74, 29.07, 30.71, 28.4, 28.53, 28.36, 28.71, 29.14)

$$Y_i = \mu_{(X=1.13)} + \varepsilon_i \qquad \varepsilon_i \sim N(0, \sigma_\varepsilon^2)$$

Stimiamo

$$\hat{\mu}_{(X=1.13)} = \frac{1}{n}\sum_{i=1}^{n} y_i = 28.92$$

e

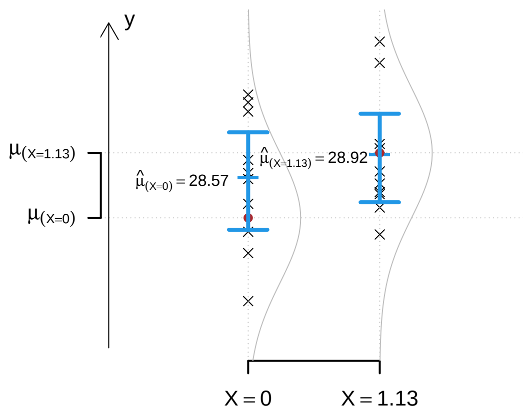
$$\begin{aligned}
\hat{\sigma}_\varepsilon^2 &= \frac{1}{n}\sum_{i=1}^{n} y_i^2 - \hat{\mu}_{(X=0)}^2 \\
&= \frac{1}{10}(28.16^2 + 30.38^2 + 27.74^2 + ... + 28.71^2 + 29.14^2) - 28.92^2 \\
&= 0.8157 \\
\hat{\sigma}_\varepsilon &= 0.9032
\end{aligned}$$

Correggiamo $\hat{\sigma}_\varepsilon^2$

$$\begin{aligned}
S_{(X=1.13)}^2 &= \frac{n}{n-1}\hat{\sigma}_{(X=0)}^2 \\
&= \frac{10}{10-1} \cdot 0.8157 \\
&= 0.9063 \\
S_{(X=1.3)} &= 0.952
\end{aligned}$$

L'intervallo di confidenza al 95% per $\mu_{(X=1.13)}$ è

$$\hat{\mu}_{(X=1.13)} \pm t_{n-1;\alpha/2}\frac{S}{\sqrt{n}} = 28.92 \pm 2.2622\frac{0.952}{10} = (28.239; 29.601)$$



Se mettiamo a test, otteniamo

Test T per due medie, (omogeneità)

$\boxed{\text{A}}$ FORMULAZIONE DELLE IPOTESI

$$\begin{cases} H_0 : \mu_{(X\,=\,1.13)} = \mu_{(X\,=\,0)} \\ H_1 : \mu_{(X\,=\,1.13)} > \mu_{(X\,=\,0)} \end{cases}$$

$\boxed{\text{B}}$ SCELTA E CALCOLO STATISTICA-TEST, T

L'ipotesi è di omogeneità e quindi calcoliamo:

$$S_p^2 = \frac{n_{\,(X\,=\,1.13)}\,\hat{\sigma}^2_{(X\,=\,1.13)} + n_{\,(X\,=\,0)}\,\hat{\sigma}^2_{(X\,=\,0)}}{n_{\,(X\,=\,1.13)} + n_{\,(X\,=\,0)} - 2} = \frac{10\cdot 0.9032^2 + 10\cdot 0.994^2}{10+10-2} = 1.002$$

$$\begin{aligned} \frac{\hat{\mu}_{\,(X\,=\,1.13)} - \hat{\mu}_{\,(X\,=\,0)}}{\sqrt{\frac{S_p^2}{n_{\,(X\,=\,1.13)}} + \frac{S_p^2}{n_{\,(X\,=\,0)}}}} &\sim t_{n_{\,(X\,=\,1.13)} + n_{\,(X\,=\,0)} - 2} \\ t_{\text{obs}} &= \frac{(28.92 - 28.57)}{\sqrt{\frac{0.9064}{10} + \frac{1.098}{10}}} = 0.784\,. \end{aligned}$$

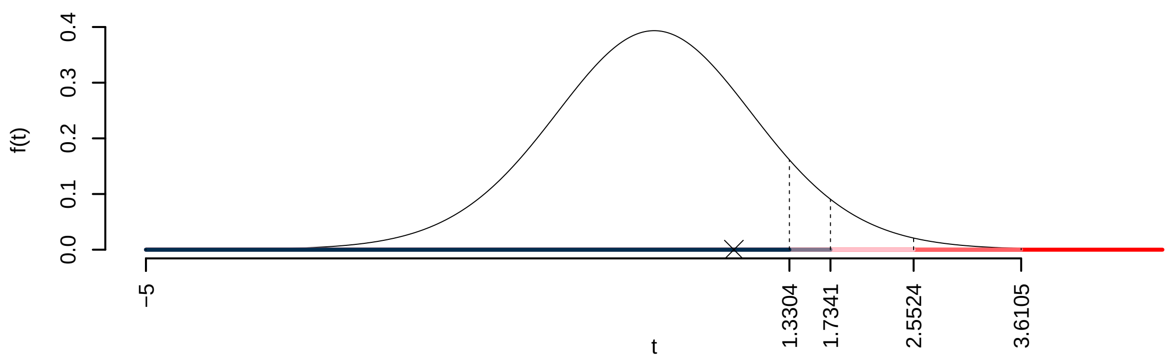
$\boxed{\text{C}}$ CONCLUSIONE

Consideriamo $\alpha = 0.1, 0.05, 0.01, 0.001$

I valori critici sono

$t_{20-2;0.1} = 1.3304$; $t_{20-2;0.05} = 1.7341$; $t_{20-2;0.01} = 2.5524$; $t_{20-2;0.001} = 3.6105$

Siccome $t_{\text{obs}} = 0.784 < t_{20-2;0.1} = 1.3304$, quindi **non** rifiuto H_0 a **nessun** livello di significatività, $p_{\text{value}} > 0.1$, *non significativo*



Il p_{value} è

$$p_{\text{value}} = P(T_{20-2} > 0.78) = 0.221609$$

Attenzione il calcolo del p_{value} con la T è puramente illustrativo e non può essere riprodotto senza una calcolatrice statistica adeguata.

$$0.1 < p_{\text{value}} = 0.221609 \leq 1$$

NULL

Esempio 17.1.3. Supponiamo di aver piantato altri 10 ettari di questa varietà ma su terreni trattati con 2.84 hg ($X = 2.84$) di concime azotato per ettaro e osserviamo (29.03, 32.09, 29.56, 29.59, 28.52, 30.24, 30.2, 29.89, 30.41, 30.94)

$$Y_i = \mu_{(X=2.84)} + \varepsilon_i \qquad \varepsilon_i \sim N(0, \sigma_\varepsilon^2)$$

Stimiamo

$$\hat{\mu}_{(X=2.84)} = \frac{1}{n}\sum_{i=1}^{n} y_i = 30.047$$

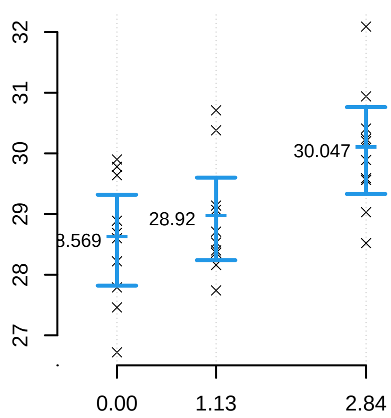
e

$$\hat{\sigma}^2_{(X=2.84)} = \frac{1}{n}\sum_{i=1}^{n} y_i^2 - \hat{\mu}^2_{(X=2.84)}$$

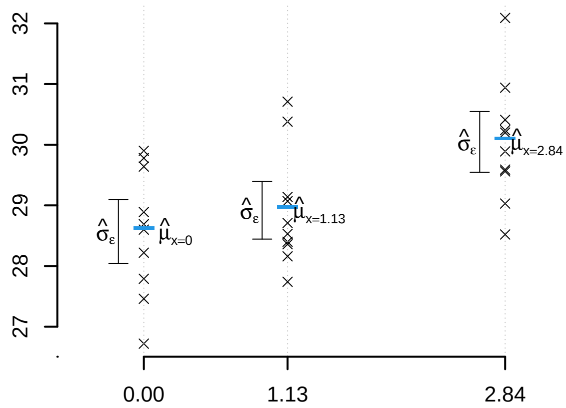
$$
\begin{aligned}
&= \frac{1}{10}(29.03^2 + 32.09^2 + 29.56^2 + ... + 30.41^2 + 30.94^2) - 30.047^2 \\
&= 0.9 \\
\hat{\sigma}_{(X=2.84)} &= 0.9487
\end{aligned}
$$

infine

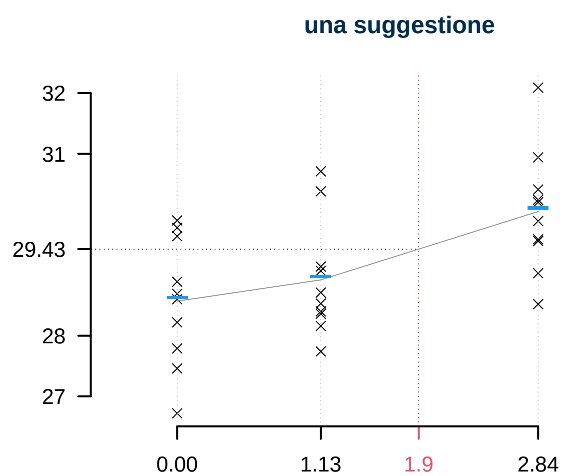
$$
\begin{aligned}
S^2_{(X=2.84)} &= \frac{n}{n-1}\hat{\sigma}^2_{(X=2.84)} \\
&= \frac{10}{10-1} \cdot 0.9 \\
&= 1 \\
S_{(X=2.84)} &= 1
\end{aligned}
$$



problema 1: abbiamo tre stime di μ e tre stime di σ ottenute come se i campioni fossero separati. Non abbiamo tenuto conto della natura metrica della X.

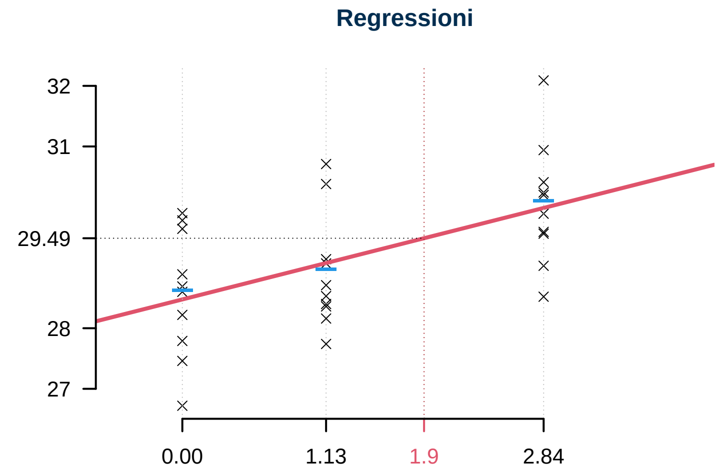


problema 2: alla luce di questi dati cosa possiamo dire sulla produzione media se usassimo 1.9 hg di concime per ettaro? Potremmo proporre di congiungere le medie con una spezzata



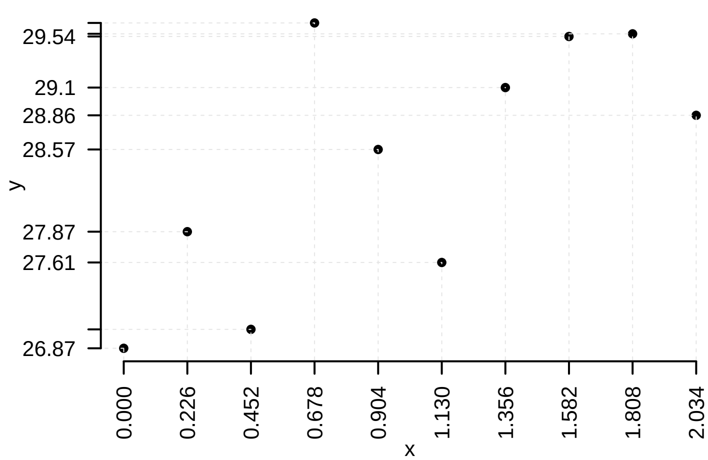
Il problema è che dobbiamo stimare 3 coppie di parametri μ_1, σ_1^2, μ_2, σ_2^2 e μ_3, σ_3^2 **senza** tenere conto della natura metrica della x.

L'idea è di sostituire a μ_1, μ_2 e μ_3 una funzione in x, $\mu(x)$ e stimarla coi dati $\hat{\mu}(x)$, per ottenere qualcosa del tipo

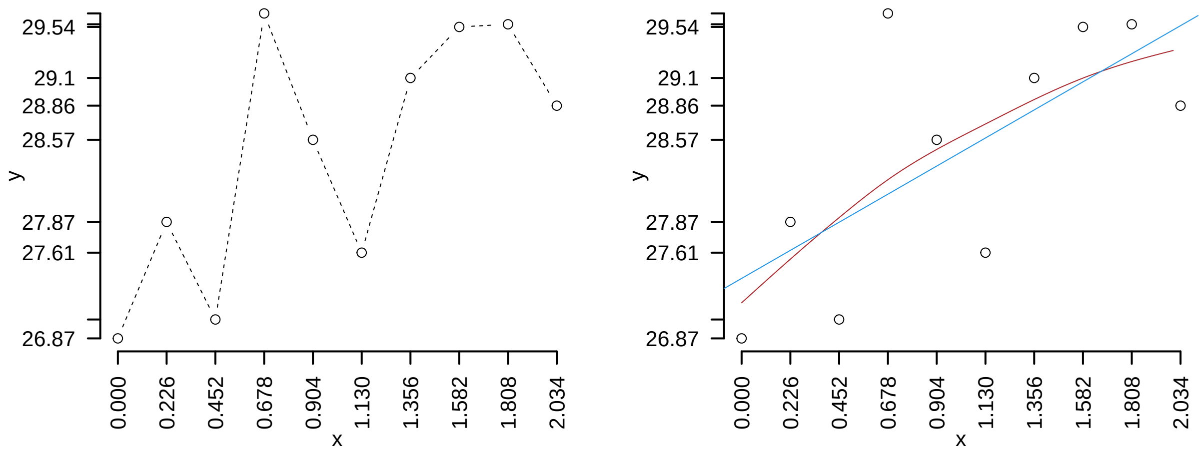


Esempio 17.1.4. **problema 3:** Dobbiamo studiare un nuova varietà di riso, ma il budget è inferiore. osserviamo 10 ettari

X = Fertilizzante	0.00	0.226	0.452	0.678	0.904	1.13	1.356	1.582	1.808	2.034
Y = Raccolto	26.87	27.869	27.035	29.651	28.571	27.61	29.099	29.536	29.558	28.863



La previsione sembra più difficile, la suggestione di prima non sembra funzionare, dobbiamo modellare il luogo delle medie μ con una funzione $\mu(x)$ e stimarla coi dati $\hat{\mu}(x)$.



17.2 Il modello di regressione

Vogliamo modellare la linea delle medie. Ci aspetteremmo una linea che passi tra i punti che che sia, per ogni x dato, il valore atteso di Y condizionato a quella x. Cioè abbiamo bisogno di un modello per la media di Y per un dato X

$$E(Y_i|X = x_i) = \mu(x_i)$$

E si legge che Y, in media, condizionato ad $X = x$ vale $f(x)$, dove f è una funzione da scegliere. L'idea di base è che

$$y_i = \mu(x_i) + \varepsilon_i$$

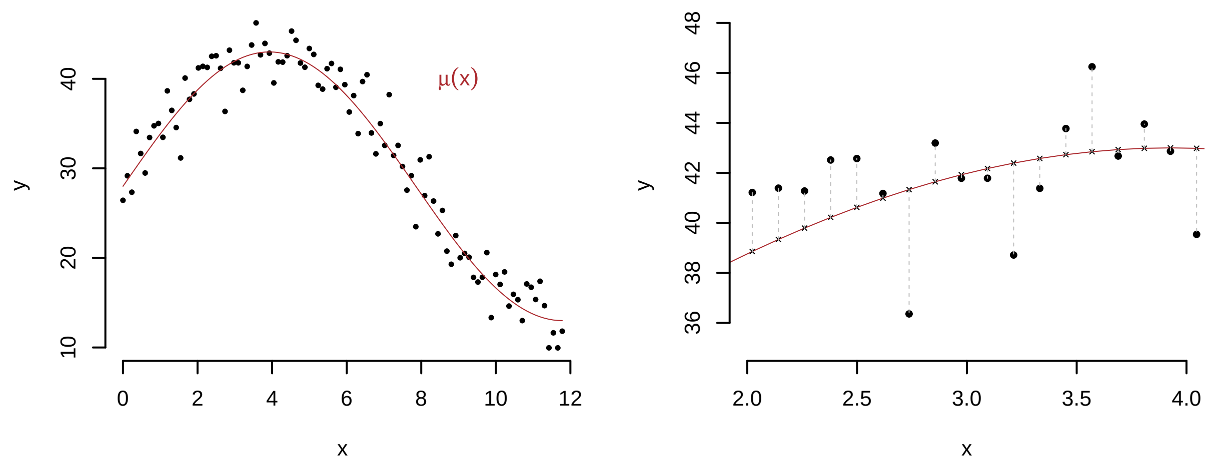
la y osservata in corrispondenza della x è pari ad $\mu(x)$ a meno di un errore casuale. Dove si assume che l'errore abbia media zero

$$E(\varepsilon_i) = 0$$

e varianza costante

$$V(\varepsilon_i) = \sigma^2_\varepsilon, \ \forall i$$

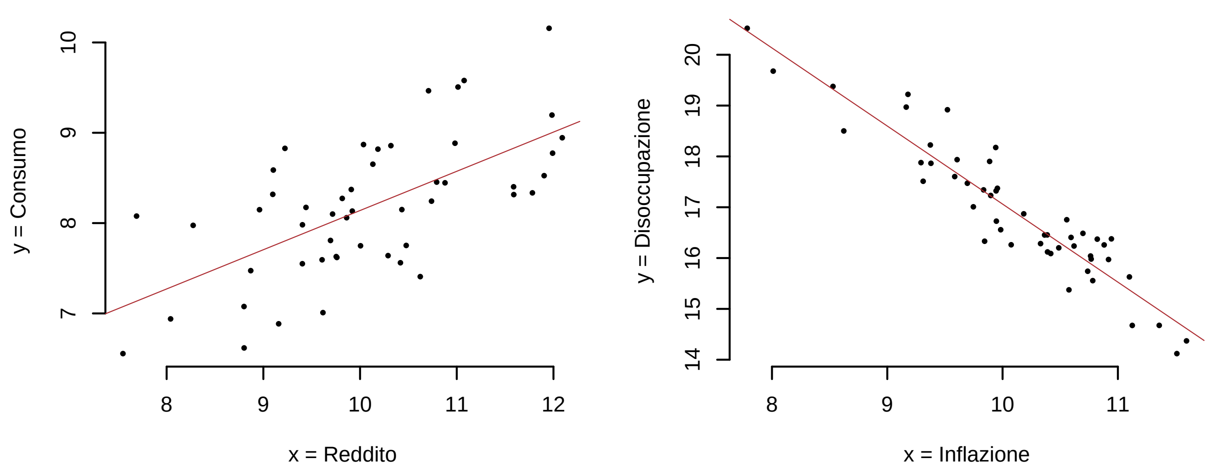
Qui di seguito un paio di esempi grafici



17.3 La Regressione Lineare

Tra le tante $\mu(x)$ limitiamo l'attenzione alle funzioni lineari

$$y_i = \mu(x_i) = \beta_0 + \beta_1 x_i + \varepsilon_i$$



17.3.1 Il modello di regressione lineare semplice

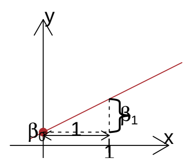
Il modello di regressione (lineare semplice) è

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \qquad E(\varepsilon_i) = 0,\ V(\varepsilon_i) = \sigma^2_{\varepsilon_i}$$

Nota

Le variabili sono chiamate:
- y dipendente, endogena, risposta, ecc.
- x indipendente, esogena, stimolo, ecc.
- ε è chiamato, residuo, o errore

I parametri di popolazione sono **3**: β_0, β_1 e σ_ε^2
- β_0 è l'intercetta della retta. β_0 rappresenta la media di y quando $x = 0$, non sempre ha significato fenomenico
- β_1 è il coefficiente angolare. β_1 rappresenta quanto la media di y aumenta all'aumentare unitario della x.
- σ_ε^2 è la varianza dell'errore ε. σ_ε^2 rappresenta la variabilità dei punti intorno alla retta.



17.3.2 La Storia del Metodo

Nel 1886, Francis Galton pubblicò un articolo su *Nature* intitolato *Regression towards mediocrity in hereditary stature*. In quell'occasione, rese noti i dati raccolti su centinaia di coppie padre-figlio, relativi alla statura (convertita poi in centimetri). Galton ipotizzò inizialmente un modello molto semplice: l'altezza del figlio y_i doveva corrispondere a quella del padre x_i, salvo errori casuali:

$$y_i = x_i + \varepsilon_i$$

In questo modello, l'errore medio sarebbe nullo, e la retta teorica che descrive la relazione padre-figlio sarebbe la bisettrice del primo quadrante. Tuttavia, una volta osservati i dati empirici, l'evidenza fu diversa.

Nel campione considerato, la relazione stimata tra statura del padre e del figlio risultò essere:

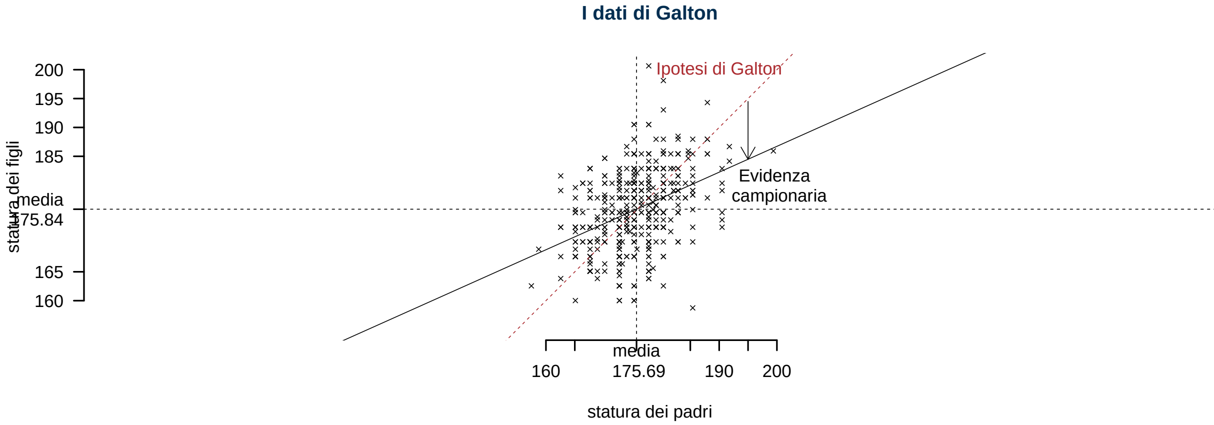
$$y_i = 97.1776 + 0.4477 \cdot x_i + \varepsilon_i$$

dove l'intercetta 97.1776 è positiva e il coefficiente angolare 0.4477 è **inferiore a 1**. Questo significa che:

- i figli di padri molto alti tendono a essere sì più alti della media ma più bassi dei loro padri,

- mentre i figli di padri molto bassi tendono a essere sì più bassi della media ma più alti dei loro padri.

Il fenomeno fu battezzato da Galton stesso con il nome di **regressione verso la media**, e in seguito divenne uno dei concetti fondamentali della statistica.

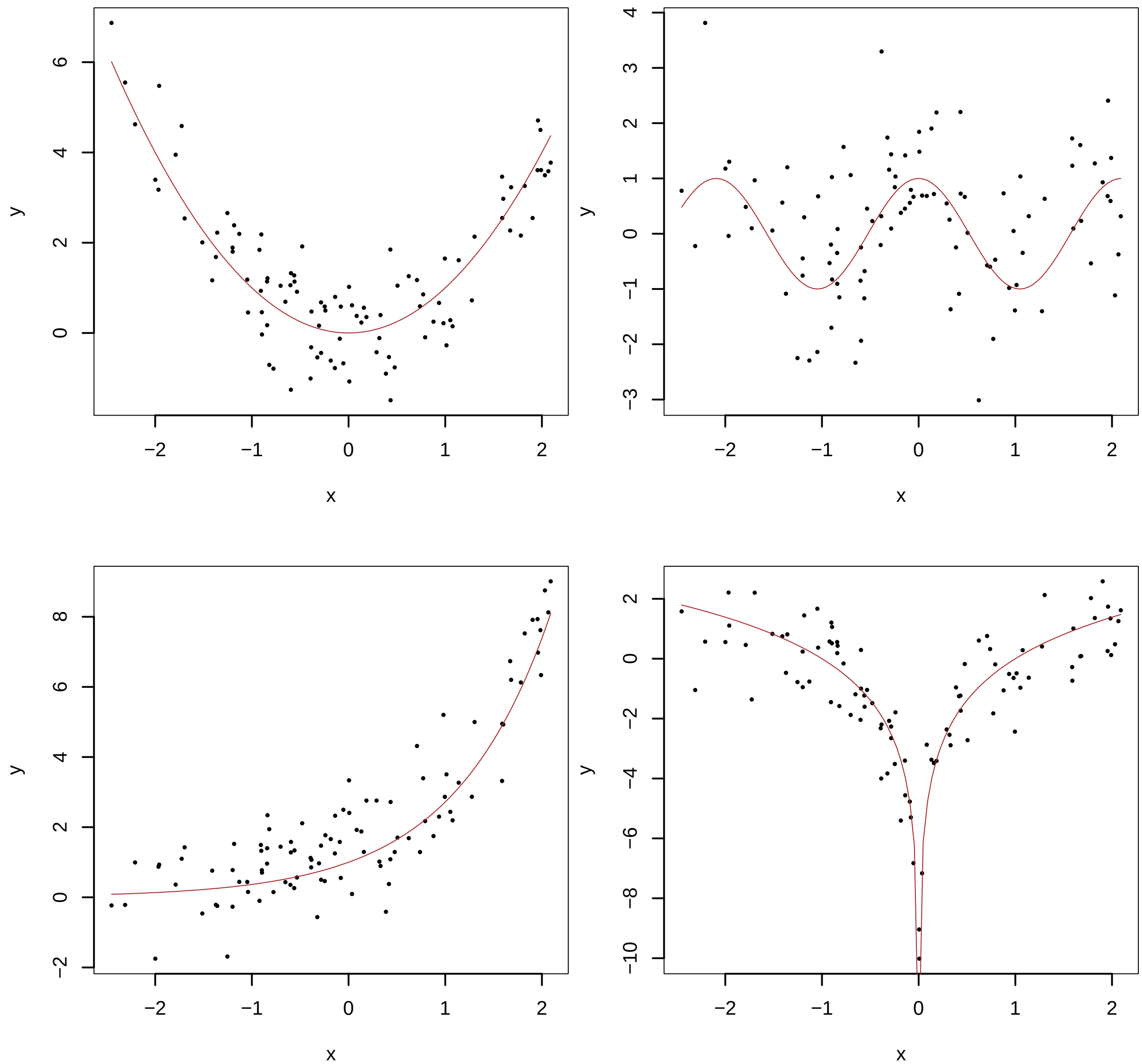


17.3.3 Gli assunti del modello di regressione

0. Dati $(x_1, y_1), ..., (x_n, y_n)$, n coppie di punti, si assume che
$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$
1. Il valore atteso dell'errore è nullo
$$E(\varepsilon_i) = 0$$
2. Omoschedasticità
$$V(\varepsilon_i) = \sigma^2_\varepsilon, \qquad \text{costante } \forall i$$
3. Indipendenza dei residui
$$\varepsilon_i \text{ è indipendente da } \varepsilon_j \quad \forall i \neq j$$
4. Indipendenza tra i residui e la X
$$X_i \text{ è indipendente da } \varepsilon_i \quad \forall i$$
5. *Esogeneità* della X: la distribuzione su X non è oggetto di inferenza
6. Normalità dei residui
$$\varepsilon_i \sim N(0, \sigma^2_\varepsilon)$$

Assunto 0 Dati $(x_1, y_1), ..., (x_n, y_n)$, n coppie di punti, si assume che
$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

Situazioni in cui l'assunto 0 è violato

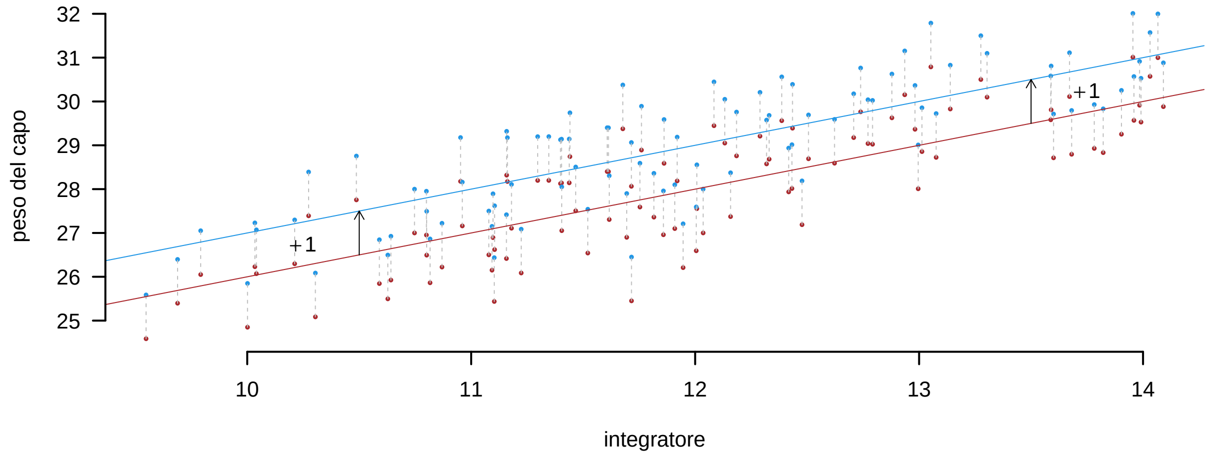


Assunto 1, il valore atteso dell'errore è nullo

$$E(\varepsilon_i) = 0$$

Inverificabile, se per esempio

$$E(\varepsilon_i) = +1, \quad \forall i, \;\; E(Y_i|x_i) = \beta_0 + \beta_1 x_i + E(\varepsilon_i) = \beta_0 + \beta_1 x_i + 1$$

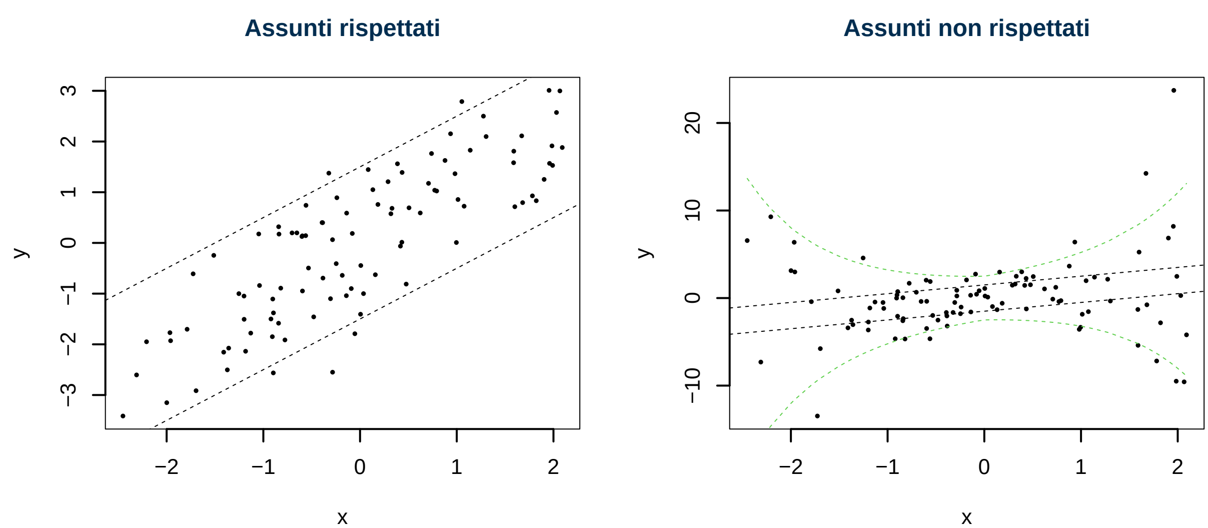


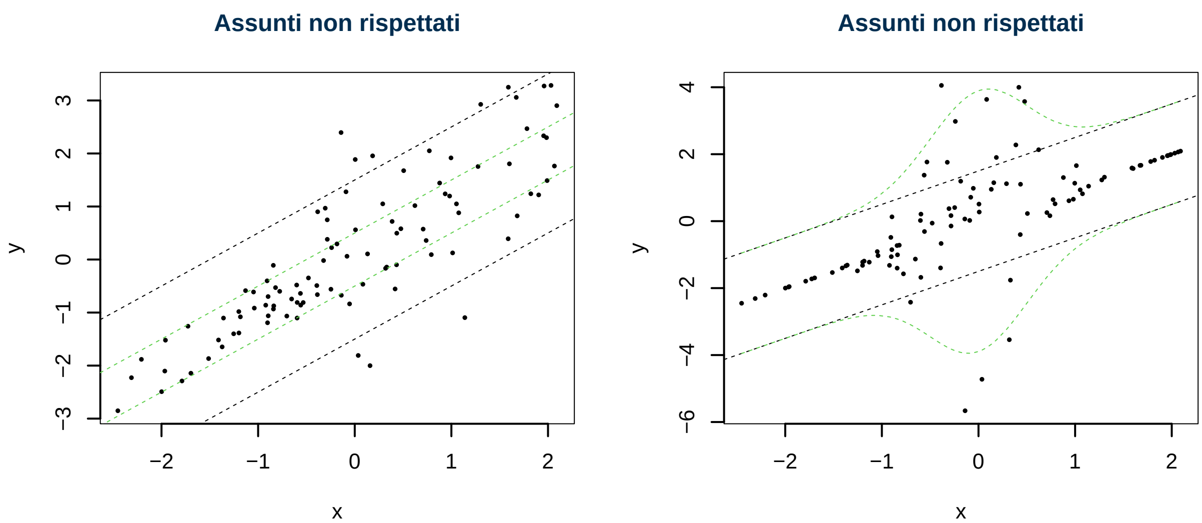
Assunti 3. e 6. Omoschedasticità e Indipendenza tra i residui e la X

$$V(\varepsilon_i) = \sigma_\varepsilon^2, \qquad \text{costante } \forall i$$

$$X_i \text{ è indipendente da } \varepsilon_i \;\; \forall i$$

La varianza non cambia con le osservazioni

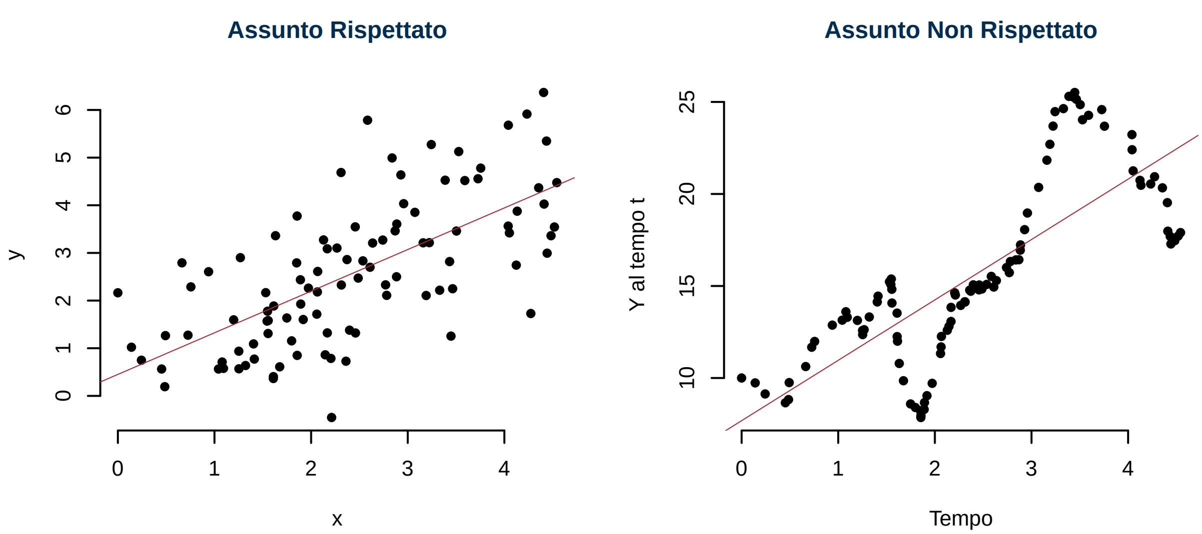




4. Indipendenza dei residui

$$\varepsilon_i \text{ è indipendente da } \varepsilon_j \quad \forall i \neq j$$

L'assunto vine tipicamente violato nelle serie temporali



5. Esogeneità della X: la distribuzione su X non è oggetto di inferenza. In contesto di sperimentazione la X viene fissata dal ricercatore. In un contesto di dati sul campo la X non può essere fissata, ma viene usata *come se fosse* fissata. L'obiettivo di ricerca rimane la Y e la relazione tra X ed Y, **non** la distribuzione di X.

6. Normalità dei residui Gli errori sono normalmente distribuiti

$$\varepsilon_i \sim N\left(0, \sigma_\varepsilon^2\right)$$

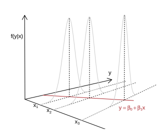
Essendo

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad \varepsilon_i \sim N\left(0, \sigma_\varepsilon^2\right)$$

allora, dalle proprietà delle normali

$$Y_i \sim N\left(\beta_0 + \beta_1 x_i, \sigma_\varepsilon^2\right)$$

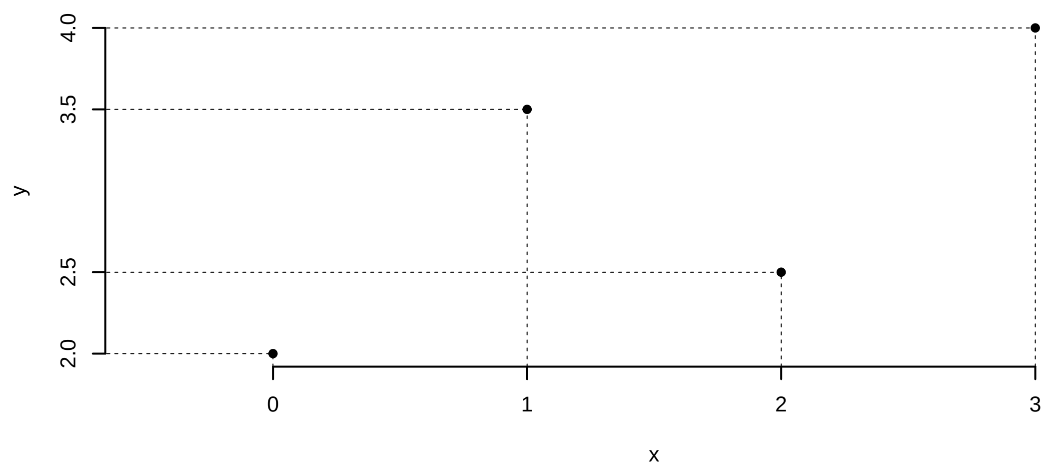
Y_i è normale con media $\beta_0 + \beta_1 x_i$ e varianza σ_ε^2



17.3.4 Il metodo dei minimi quadrati

È un metodo di stima per β_0 e β_1, si può dimostrare che, in virtù dell'assunto 6. (la normalità dei residui) il metodo dei minimi quadrati produce le stime di massima verosimiglianza. Svilupperemo la teoria attraverso un esempio su 4 punti.

i	x_i	y_i
1	0	2.0
2	1	3.5
3	2	2.5
4	3	4.0



17.3.5 La distanza di una retta dai punti (il metodo dei minimi quadrati)

Si tratta di cercare la retta che rende minima la somma dei quadrati delle distanza della retta misurate nella scala di misura della y. Per ogni retta candidata (β_0^*, β_1^*), costruiamo la previsione

$$\hat{y}_i^* = \beta_0^* + \beta_1^* x_i$$

e quindi i residui

$$\hat{\varepsilon}_i^* = y_i - \hat{y}_i$$

Il criterio dei **minimi quadrati** è

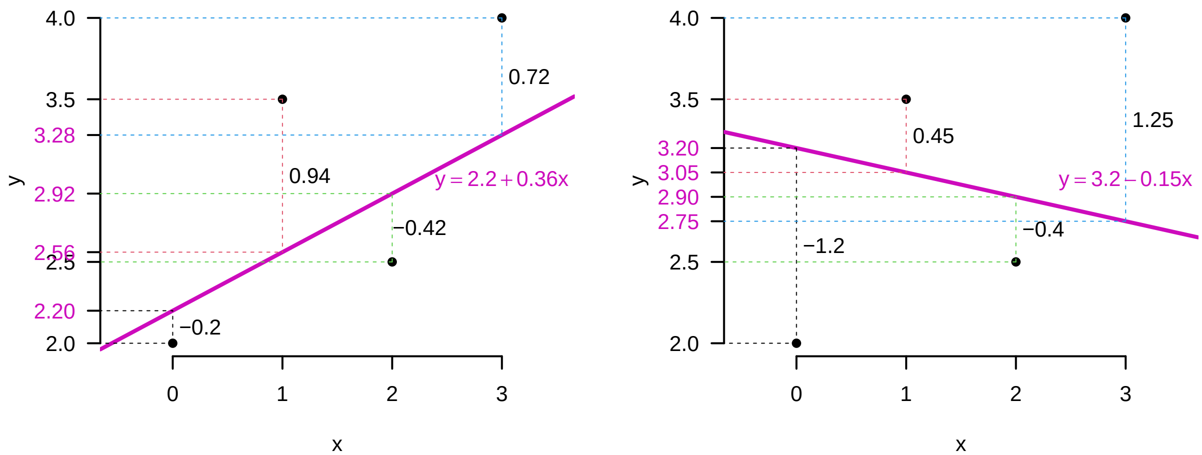
$$G(\beta_0^*, \beta_1^*) = \sum_{i=1}^{n} (\hat{\varepsilon}_i^*)^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i^*)^2 = \sum_{i=1}^{n} (y_i - (\beta_0^* + \beta_1^* x_i))^2$$

La retta dei minimi quadrati è quella coppia di $\hat{\beta}_0$ e $\hat{\beta}_1$ tali che

$$G(\hat{\beta}_0, \hat{\beta}_1) < G(\beta_0^*, \beta_1^*), \quad \forall (\beta_0^*, \beta_1^*) \neq (\hat{\beta}_0, \hat{\beta}_1)$$

Se, per esempio, scegliessimo $\beta_0^* = 2.2,\ \beta_1^* = 0.36$, otterremmo

i	x_i	y_i	$\hat{y}_i^* = \beta_0^* + \beta_1^* x_i$	$\hat{\varepsilon}_i^* = y_i - \hat{y}^*$	$(\hat{\varepsilon}_i^*)^2$
1	0	2.0	2.20	-0.20	0.0400
2	1	3.5	2.56	0.94	0.8836
3	2	2.5	2.92	-0.42	0.1764
4	3	4.0	3.28	0.72	0.5184



i	x_i	y_i	$\widehat{y}_i^* = \beta_0^* + \beta_1^* x_i$	$\widehat{\varepsilon}_i^* = y_i - \widehat{y}^*$	$(\widehat{\varepsilon}_i^*)^2$
1	0	2.0	3.20	-1.20	1.4400
2	1	3.5	3.05	0.45	0.2025
3	2	2.5	2.90	-0.40	0.1600
4	3	4.0	2.75	1.25	1.5625

$$G(2.2, 0.36) = 3.365$$

Se invece scegliessimo $\beta_0^* = 3.2,\ \beta_1^* = -1.5$, otterremmo

$$G(3.2, -1.15) = 3.365$$

Siccome

$$G(2.2, 0.36) = 1.6184 < G(3.2, -1.15) = 3.365$$

allora diremo che la retta $\beta_0^* = 2.2,\ \beta_1^* = 0.36$ è *più vicina*, nel senso dei minimi quadrati, della retta $\beta_0^* = 3.2,\ \beta_1^* = -1.5$.

17.3.6 Soluzioni dei minimi quadrati

La retta dei minimi quadrati è quella coppia di $\widehat{\beta}_0$ e $\widehat{\beta}_1$ tali che

$$G(\widehat{\beta}_0, \widehat{\beta}_1) < G(\beta_0, \beta_1), \quad \forall (\beta_0, \beta_1) \neq (\widehat{\beta}_0, \widehat{\beta}_1)$$

ovvero

$$(\widehat{\beta}_0, \widehat{\beta}_1) = \underset{(\beta_0, \beta_1) \in \mathbb{R}^2}{\operatorname{argmin}} G(\beta_0, \beta_1)$$

Proprietà 17.3.1 (Stimatori dei Minimi Quadrati). *Gli stimatori dei minimi quadrati $\hat{\beta}_0$ e $\hat{\beta}_1$ sono*

$$\begin{aligned}\hat{\beta}_1 &= \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2} = \frac{cov(x,y)}{\hat{\sigma}_X^2}\\ \hat{\beta}_0 &= \bar{y}-\hat{\beta}_1\bar{x}\end{aligned}$$

dove

$$\bar{y}=\frac{1}{n}\sum_{i=1}^{n}y_i, \qquad \bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$$

e

$$cov(x,y)=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})$$

Dimostrazione.

$$\begin{aligned}G(\beta_0,\beta_1) &= \frac{1}{n}\sum_{i=1}^{n}\left(y_i-(\beta_0+\beta_1x_i)\right)^2\\ \frac{\partial G(\beta_0,\beta_1)}{\partial\beta_0} &= -\frac{2}{n}\sum_{i=1}^{n}\left(y_i-(\beta_0+\beta_1x_i)\right)\\ &= -\frac{2}{n}\left(\sum_{i=1}^{n}y_i-\sum_{i=1}^{n}\beta_0-\sum_{i=1}^{n}\beta_1x_i\right)\\ &= -\frac{2}{n}\left(n\bar{y}-n\beta_0-n\beta_1\bar{x}\right)\\ &= -2\left(\bar{y}-\beta_0-\beta_1\bar{x}\right)\\ \frac{\partial G(\beta_0,\beta_1)}{\partial\beta_1} &= -\frac{2}{n}\sum_{i=1}^{n}x_i\left(y_i-(\beta_0+\beta_1x_i)\right)\\ &= -\frac{2}{n}\left(\sum_{i=1}^{n}x_iy_i\right)-\left(\sum_{i=1}^{n}\beta_0x_i\right)-\left(\sum_{i=1}^{n}\beta_1x_i^2\right)\\ &= -2\left(\left(\frac{1}{n}\sum_{i=1}^{n}x_iy_i\right)-\beta_0\bar{x}-\beta_1\left(\frac{1}{n}\sum_{i=1}^{n}x_i^2\right)\right)\end{aligned}$$

E otteniamo il **sistema di equazioni normali**

$$\begin{cases}\frac{\partial G(\beta_0,\beta_1)}{\partial\beta_0} &= 0\\ \frac{\partial G(\beta_0,\beta_1)}{\partial\beta_1} &= 0\end{cases}$$

Dividendo per 2 e cambiando di segno, si ha

$$\begin{cases} (\bar{y} - \beta_0 - \beta_1 \bar{x}) & = 0 \\ \left(\frac{1}{n}\sum_{i=1}^n x_i y_i\right) - \beta_0 \bar{x} - \beta_1 \left(\frac{1}{n}\sum_{i=1}^n x_i^2\right) & = 0 \end{cases}$$

da cui

$$\hat{\beta}_0 = \bar{y} - \beta_1 \bar{x}$$

E sostituiamo

$$\begin{aligned} \left(\frac{1}{n}\sum_{i=1}^n x_i y_i\right) - (\bar{y} - \beta_1 \bar{x})\bar{x} - \beta_1 \left(\frac{1}{n}\sum_{i=1}^n x_i^2\right) &= 0 \\ \beta_1 \left(\frac{1}{n}\sum_{i=1}^n x_i^2 - \bar{x}^2\right) &= \frac{1}{n}\sum_{i=1}^n x_i y_i - \bar{x}\cdot\bar{y} \\ \hat{\beta}_1 &= \frac{\frac{1}{n}\sum_{i=1}^n x_i y_i - \bar{x}\cdot\bar{y}}{\frac{1}{n}\sum_{i=1}^n x_i^2 - \bar{x}^2} \end{aligned}$$

Siccome

$$\begin{aligned} \operatorname{cov}(x,y) &= \frac{1}{n}\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \\ &= \frac{1}{n}\sum_{i=1}^n (x_i\ y_i - y_i\ \bar{x} - x_i\ \bar{y} + \bar{x}\bar{y}) \\ &= \frac{1}{n}\sum_{i=1}^n x_i\ y_i - \frac{\bar{x}}{n}\sum_{i=1}^n y_i - \frac{\bar{y}}{n}\sum_{i=1}^n x_i + \frac{1}{n}\sum_{i=1}^n \bar{x}\bar{y} \\ &= \frac{1}{n}\sum_{i=1}^n x_i\ y_i - \bar{x}\bar{y} - \bar{x}\bar{y} + -\bar{x}\bar{y} \\ &= \frac{1}{n}\sum_{i=1}^n x_i\ y_i - \bar{x}\bar{y} \end{aligned}$$

Concludendo la prova. □

17.4 La covarianza

Definizione 17.4.1. La Covarianza $\operatorname{cov}(x,y)$ tra due variabili x e y è una misura della loro *covariazione*

$$\operatorname{cov}(x,y) = \frac{1}{n}\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

La covarianza è la *media degli scarti incrociati dalla media* Come per la varianza esiste una formula semplificata:

Proprietà 17.4.1.

$$cov(x, y) = \frac{1}{n}\sum_{i=1}^{n} x_i\ y_i - \bar{x}\bar{y}$$

Nota

La covarianza è *la media dei prodotti degli scarti dalla media* può essere riletta come *la media dei prodotti* meno *il prodotto delle medie.*

17.4.1 Calcolo della covarianza

Nel caso in esame

$$\bar{x} = \frac{1}{4}(0 + 1 + 2 + 3) = \frac{6}{4} = 1.5$$

$$\bar{y} = \frac{1}{4}(2 + 3.5 + 2.5 + 4) = \frac{12}{4} = 3$$

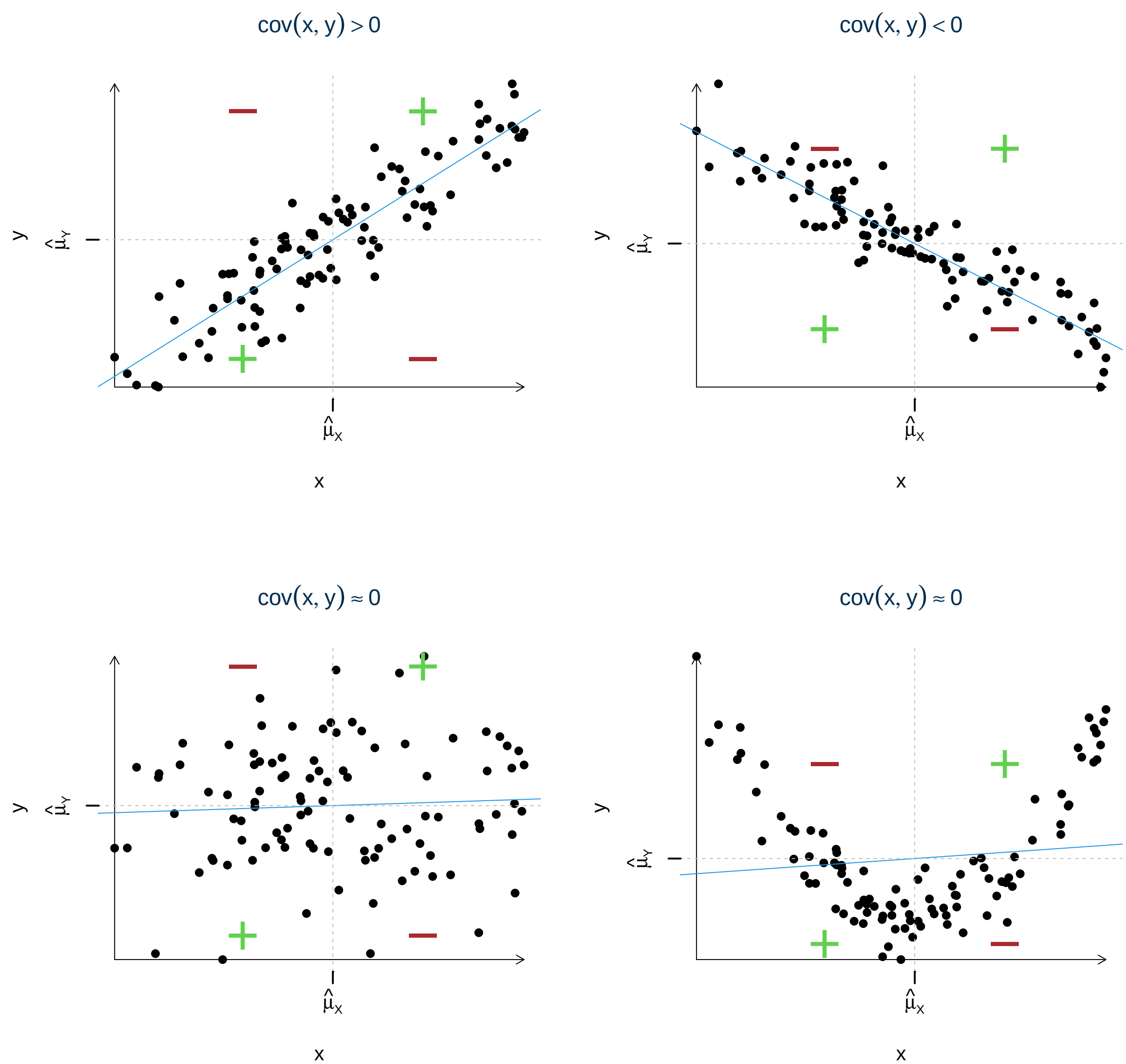
Per la covarianza:

$$\begin{aligned}
\text{cov}(x, y) &= \frac{1}{4}\Big((0-1.5)(2-3) + (1-1.5)(3.5-3) + (2-1.5)(2.5-3) + (3-1.5)(4-3)\Big) \\
&= \frac{1}{4}(0 \times 2 + 1 \times 3.5 + 2 \times 2.5 + 3 \times 4) - 1.5 \times 3 \\
&= \frac{1}{4}(20.5) - 4.5 \\
&= 0.625
\end{aligned}$$

17.4.2 Interpretazione della Covarianza

La covarianza non è direttamente leggibile perché ha un'unità di misura mista, prodotto della unità di misura della x e quella della y.

È interessante il segno della covarianza perché indica la pendenza della retta se la covarianza è positiva c'è **concordanza dei segni**, ad x maggiori del loro baricentro $((x_i - \bar{x}) > 0)$ corrispondo, *in media*, y maggiori del loro baricentro $(y_i - \bar{y}) > 0$. Mentre ad x minori del loro baricentro $((x_i - \bar{x}) < 0)$ corrispondo, *in media*, y minori del loro baricentro $(y_i - \bar{y}) < 0$. Se invece la covarianza è negativa c'è **discordanza dei segni**, cioè ad x maggiori del loro baricentro $((x_i - \bar{x}) > 0)$ corrispondo, *in media*, y minori del loro baricentro $(y_i - \bar{y}) < 0$. Mentre ad x minori del loro baricentro $((x_i - \bar{x}) < 0)$ corrispondo, *in media*, y maggiori del loro baricentro $(y_i - \bar{y}) > 0$.



17.4.3 Altre proprietà della covarianza

Simmetria

$$\text{cov}(x, y) = \text{cov}(y, x)$$

la dimostrazione è immediata

$$\text{cov}(x, y) = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{n} \sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x}) = \text{cov}(y, x)$$

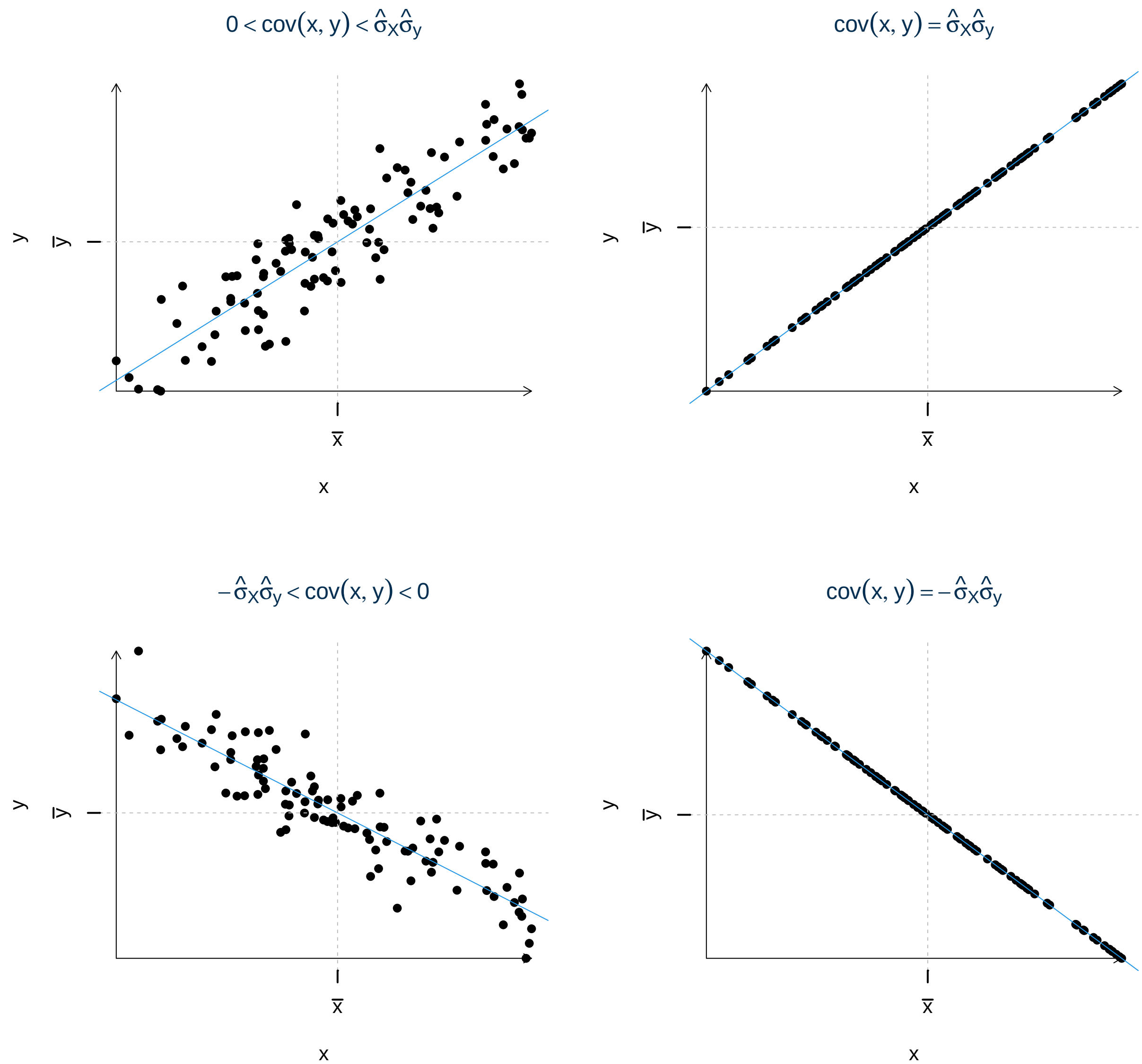
Caso particolare

$$\operatorname{cov}(x,x) = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(x_i-\bar{x}) = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2 = \hat{\sigma}_X^2$$

la covarianza di una variabile con se stessa è la varianza. La covarianza estende il concetto di varianza quando le variabili sono due.

17.4.4 Campo di variazione della covarianza

$$-\hat{\sigma}_X\hat{\sigma}_Y \le \operatorname{cov}(x,y) \le +\hat{\sigma}_X\hat{\sigma}_Y$$



Stimiamo di σ_X e σ_Y

Calcolo di $\hat{\sigma}_X^2$

$$\begin{aligned}\hat{\sigma}_X^2 &= \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \hat{\mu}_X^2 \\ &= \frac{1}{4}(0^2+1^2+2^2+3^2) - 1.5^2 \\ &= 1.25 \\ \hat{\sigma}_X &= 1.118\end{aligned}$$

Calcolo di $\hat{\sigma}_X^2$

$$\begin{aligned}\hat{\sigma}_Y^2 &= \frac{1}{n}\sum_{i=1}^{n} y_i^2 - \hat{\mu}_Y^2 \\ &= \frac{1}{4}(2^2+3.5^2+2.5^2+4^2) - 3^2 \\ &= 0.625 \\ \hat{\sigma}_Y &= 0.7906\end{aligned}$$

17.4.5 Calcolo in colonna

Nota

Molto più comodamente in colonna

i	x_i	y_i	x_i^2	y_i^2	$x_i \cdot y_i$
1	0.0	2.0	0.0	4.000	0.000
2	1.0	3.5	1.0	12.250	3.500
3	2.0	2.5	4.0	6.250	5.000
4	3.0	4.0	9.0	16.000	12.000
Totale	**6.0**	**12.0**	**14.0**	**38.500**	**20.500**
Totale/n	1.5	3.0	3.5	9.625	5.125

$$\begin{aligned}\hat{\sigma}_X^2 &= \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2 = 3.5 - 1.5^2 = 1.25 \\ \hat{\sigma}_Y^2 &= \frac{1}{n}\sum_{i=1}^{n} y_i^2 - \bar{y}^2 = 9.625 - 3^2 = 0.625 \\ \mathrm{cov}(x,y) &= \frac{1}{n}\sum_{i=1}^{n} x_i\ y_i - \bar{x}\bar{y} = 5.125 - 1.5 \cdot 3 = 0.625\end{aligned}$$

17.4.6 Calcolo di $\hat{\beta}_0$ e $\hat{\beta}_1$

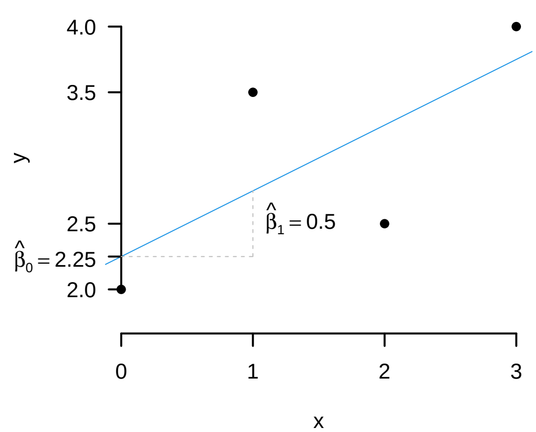
La stima del coefficiente angolare è

$$\hat{\beta}_1 = \frac{\text{cov}(x, y)}{\hat{\sigma}_X^2} = \frac{0.625}{1.25} = 0.5$$

La stima dell'intercetta è

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 3 - 0.5 \cdot 1.5 = 2.25$$

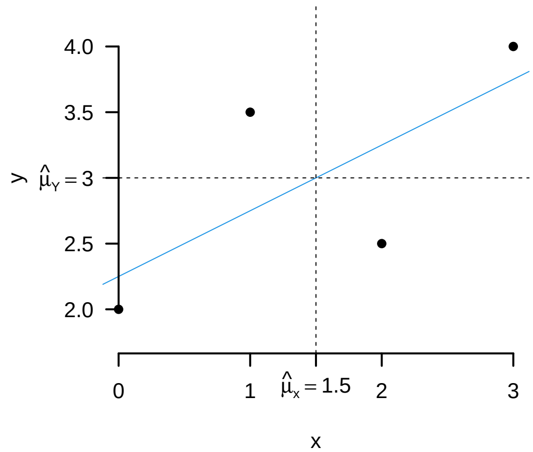
La retta stimata



17.5 Proprietà della retta dei minimi quadrati

La retta dei minimi quadrati passa per il baricentro di (x,y) che è ($\bar{x}$, $\bar{y}$)

$$\begin{aligned}\hat{\beta}_0 + \hat{\beta}_1 \bar{x} &= (\bar{y} - \hat{\beta}_1 \bar{x}) + \hat{\beta}_1 \bar{x} \\ &= \bar{y}\end{aligned}$$



Le previsioni, sulle x osservate

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

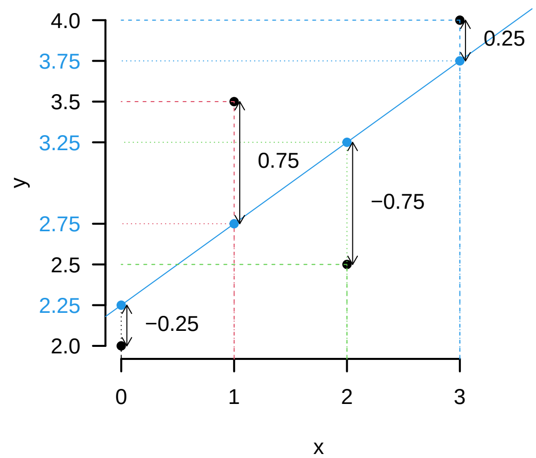
Le stime degli errori

$$\hat{\varepsilon}_i = y_i - \hat{y}_i$$

Valgono le seguenti proprietà:

$$\begin{aligned}
& y_i, && \text{le } y \text{ osservate} \\
& \hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i, && \text{le } y \text{ stimate} \\
& \hat{\varepsilon}_i = y_i - \hat{y}_i, && \text{gli errori stimati} \\
& \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i, && \text{la media degli } y \\
& \bar{y} = \frac{1}{n}\sum_{i=1}^{n} \hat{y}_i, && \text{la media degli } \hat{y} \text{ coince con qeulla degli } y \\
& 0 = \frac{1}{n}\sum_{i=1}^{n} \hat{\varepsilon}_i, && \text{la media degli scarti dalla retta è zero}
\end{aligned}$$

17.5.1 Calcolo di $\hat{y}_i$ e $\hat{\varepsilon}_i$



i	x_i	y_i	$\hat{y}_i = 2.25 + 0.5x_i$	$\hat{\varepsilon}_i$
1	0.0	2.0	2.25	-0.25
2	1.0	3.5	2.75	0.75
3	2.0	2.5	3.25	-0.75
4	3.0	4.0	3.75	0.25
Totale	**6.0**	**12.0**	**12.00**	**0.00**
Totale/n	1.5	3.0	3.00	0.00

17.6 Il coefficiente di Correlazione

Quanto si adatta bene la retta ai punti?

$$\begin{aligned} -\hat{\sigma}_X\hat{\sigma}_Y \leq &\ \operatorname{cov}(x,y) \leq +\hat{\sigma}_X\hat{\sigma}_Y \\ -\frac{\hat{\sigma}_X\hat{\sigma}_Y}{\hat{\sigma}_X\hat{\sigma}_Y} \leq &\ \frac{\operatorname{cov}(x,y)}{\hat{\sigma}_X\hat{\sigma}_Y} \leq +\frac{\hat{\sigma}_X\hat{\sigma}_Y}{\hat{\sigma}_X\hat{\sigma}_Y} \\ -1 \leq &\quad r \quad \leq +1 \end{aligned}$$

Definizione 17.6.1. Il coefficiente r

$$r = \frac{\operatorname{cov}(x,y)}{\hat{\sigma}_X\hat{\sigma}_Y}$$

è chiamato *coefficiente di correlazione.*

Nel nostro esempio

$$r = \frac{0.625}{1.118 \cdot 0.7906} = 0.7071$$

r misura *l'associazione lineare* ovvero CRESCE in valore assoluto quando i punti SI ADDENSANO intorno alla retta.

17.6.1 Proprietà di r

1. $-1 \leq r \leq 1$. Il segno indica la direzione della relazione;
   - $r > 0$, al crescere di X, *in media*, cresce Y;
   - $r < 0$, al crescere di X, *in media*, decresce Y;
   - $r = 1$, associazione perfetta diretta;
   - $r = -1$, associazione perfetta indiretta.
2. r è un numero puro, ovvero è privo di unità di misura
3. è simmetrico: $r_{XY} = r_{YX} = r$
4. è invariante per cambiamenti di scala: Se $W = a + bY$, allora

$$r_{XW} = \text{sign}(b) r_{XY}$$

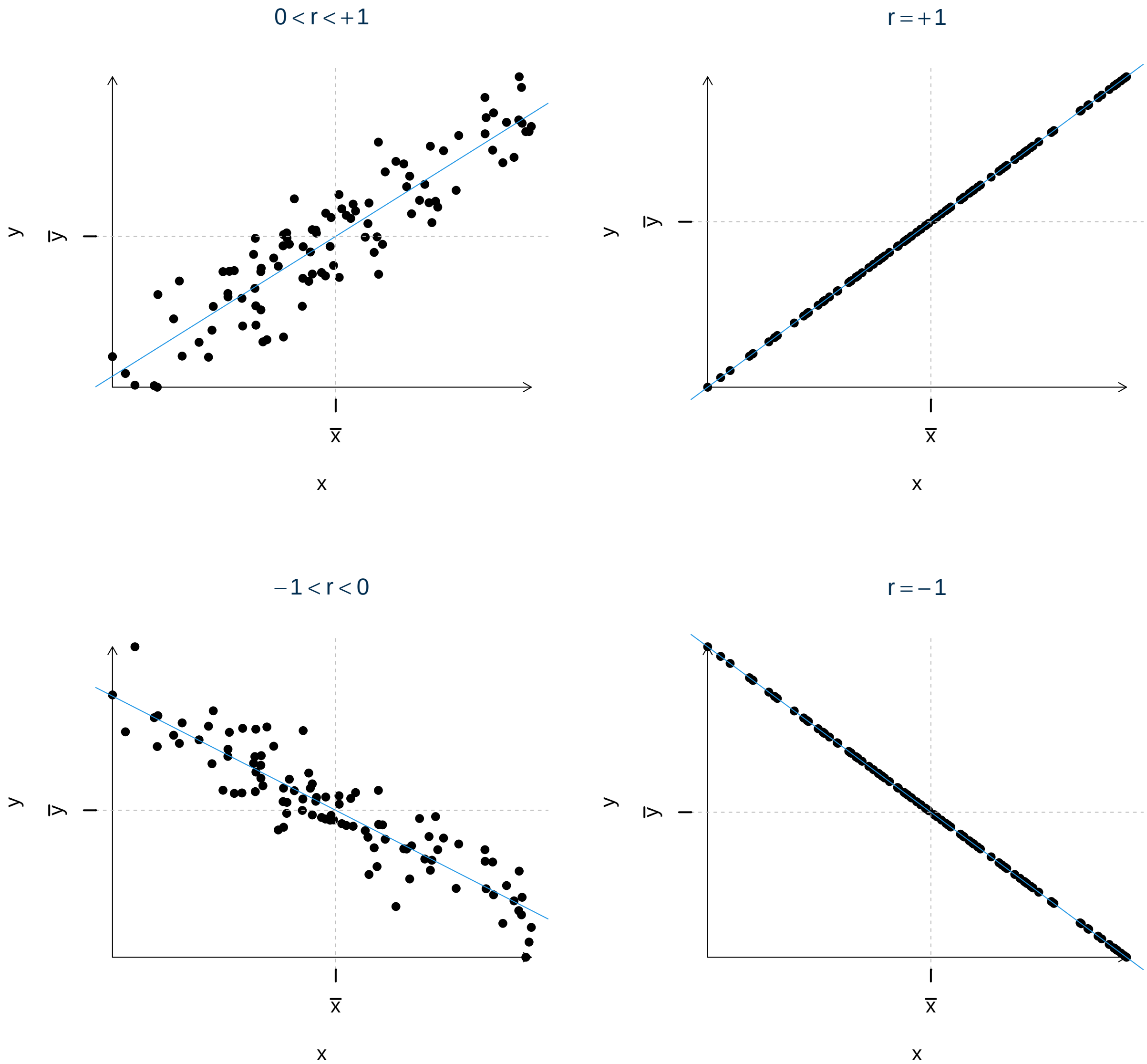
dove

$$\text{sign}(b) = \begin{cases} +1, & \text{se } b > 0 \\ -1, & \text{se } b < 0 \end{cases}$$

5. r misura l'associazione lineare:
   - r misura come i punti si addensano intorno alla retta.
   - $f(x)$ **non lineare** r è parzialmente inutile
   - il valore di r, da solo, non è in grado di descrivere tutte le possibili relazioni che si possono realizzare tra due variabili.
6. r è più elevato se i dati sono aggregati in medie o percentuali

1. La prima proprietà ci dice che r non può mai essere minore di -1 e mai maggiore di $+1$ per costruzione

$$-1 \leq r \leq 1$$



2. La seconda proprietà ci dice che r non dipende dalla scala di misura delle variabili X e Y. Mentre la covarianza ha un'unità di misura spuria, prodotto dell'unità di misura della x (u_X) e della y (u_Y), r non ha unità di misura perché è un rapporto tra la covarianza che porta l'unità di misura della X moltiplicata quella della Y e le standard deviation delle stesse.

$$\begin{aligned} \mathrm{cov}(x,y) &= \frac{1}{n}\sum_{i=1}^{n}(x_i \cdot u_X - \bar{x} \cdot u_X)(y_i \cdot u_Y - \bar{y} \cdot u_Y) \\ &= \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})u_X(y_i - \bar{y})u_Y \end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})u_X\cdot u_Y \\
&= \mathrm{cov}(x,y)u_X\cdot u_Y
\end{aligned}
$$

Le standard deviation sono espresse nell'unità di misura delle variabili

$$
\begin{aligned}
\hat{\sigma}_Y &= \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i\cdot u_Y-\bar{y}\cdot u_Y)^2} \\
&= \sqrt{\frac{1}{n}\sum_{i=1}^{n}((y_i-\bar{y})u_Y)^2} \\
&= \left(\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y})^2}\right)u_Y \\
&= \hat{\sigma}_Y u_Y \\
\hat{\sigma}_X &= \hat{\sigma}_X u_X
\end{aligned}
$$

E quindi r

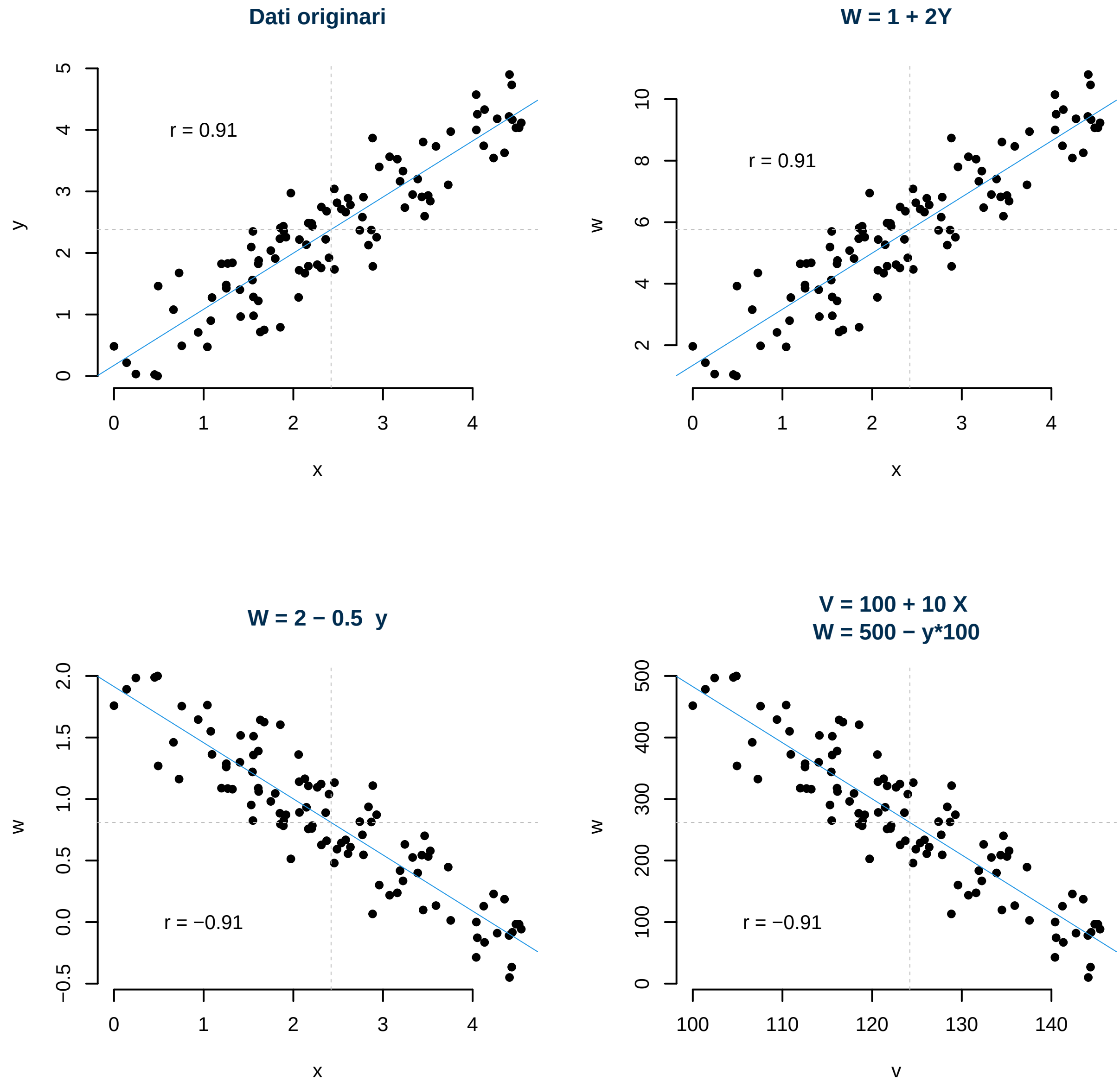
$$
r=\frac{\mathrm{cov}(x,y)}{\hat{\sigma}_X\hat{\sigma}_Y}=\frac{\mathrm{cov}(x,y)u_X\cdot u_Y}{\hat{\sigma}_X\cdot u_X\hat{\sigma}_Y\cdot u_Y}=\frac{\mathrm{cov}(x,y)u_X\cdot u_Y}{\hat{\sigma}_X\cdot u_X\hat{\sigma}_Y\cdot u_Y}
$$

3. La terza proprietà deriva direttamente dalla simmetria della covarianza:

$$
r_{XY}=\frac{\mathrm{cov}(x,y)}{\hat{\sigma}_X\hat{\sigma}_Y}=\frac{\mathrm{cov}(y,x)}{\hat{\sigma}_Y\hat{\sigma}_X}=r_{YX}
$$

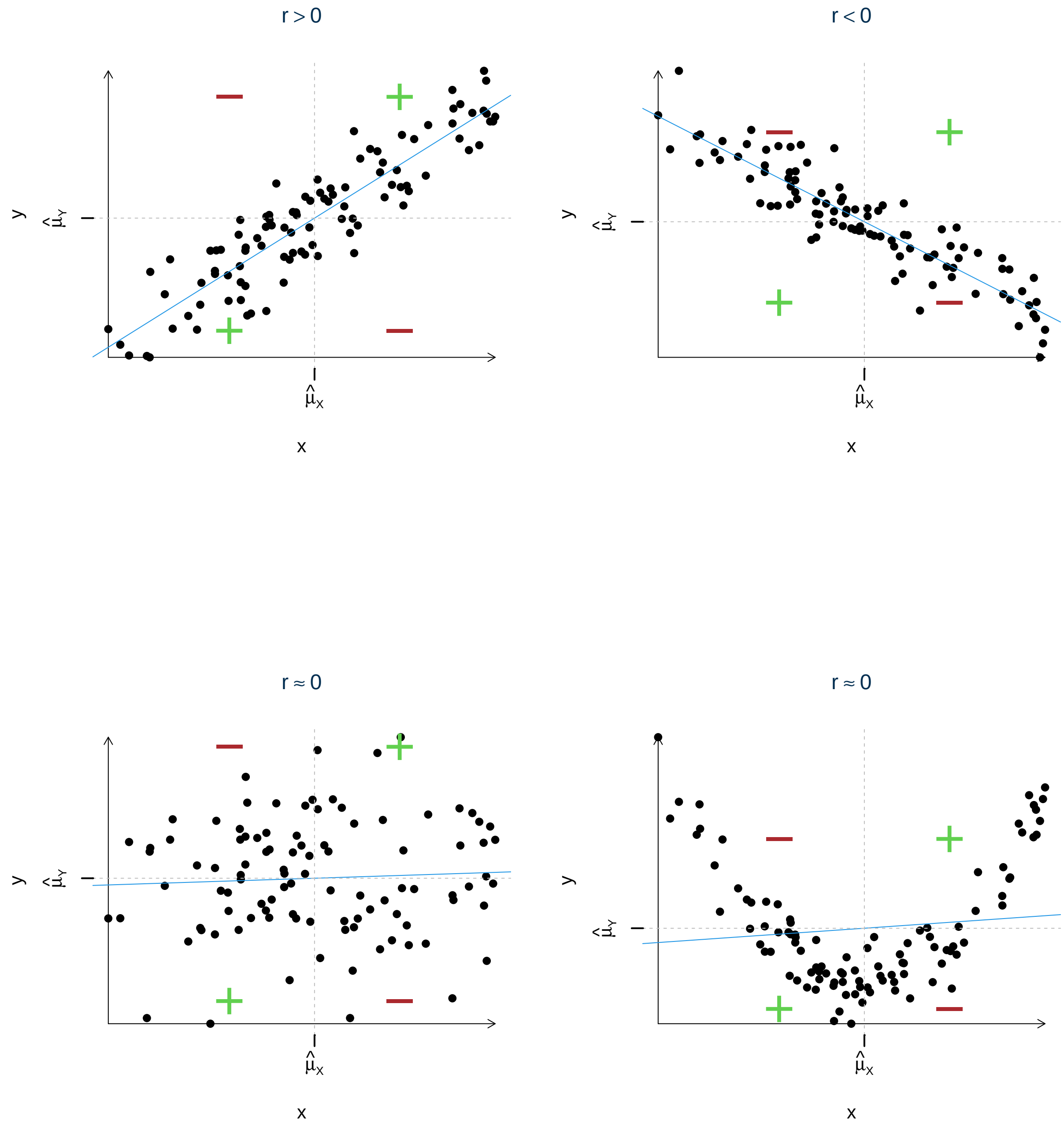
4. La quarta proprietà ci dice che, essendo un numero puro, r, non dipende dalle unità di misura di X ed Y e quindi cambiarla non comporta alterazioni su r.

$$
\text{se } W=a+bY,\text{allora } r_{X,W}=\mathrm{sign}(b)r_{XY},\ \text{dove la funzione } \mathrm{sign}(b)=\begin{cases}+1, & \text{se } b>0\\ -1, & \text{se } b<0\end{cases}
$$

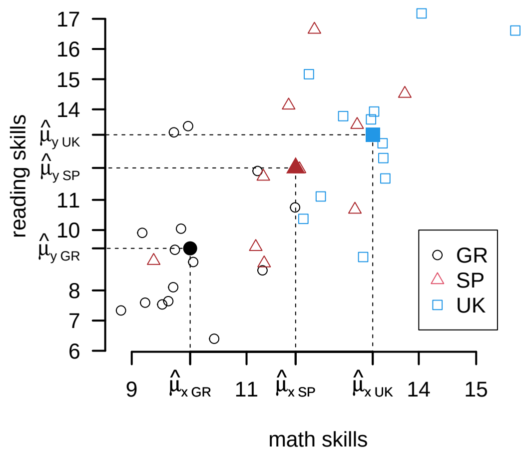


5. La quinta proprietà dice che r misura l'associazione lineare, ovvero Il coefficiente di correlazione r è una misura della distanza dei punti da una retta.

- r misura come i punti si addensano intorno alla retta.
- $f(x)$ **non lineare** r è parzialmente inutile
- il valore di r, da solo, non è in grado di descrivere tutte le possibili relazioni che si possono realizzare tra due variabili.



6. Infine r è più elevato se i dati sono aggregati in medie o percentuali. Infatti se aggreghiamo, come nell'esempio qui sotto, i dati di tre paesi nelle loro medie, diminuiamo la variabilità.



L'indice r calcolato su tutti i dati è

$$r = 0.7056$$

L'indice r calcolato sulle tre medie è

$$r_{\text{Medie}} = 0.9893$$

17.7 Scomposizione della varianza

La scomposizione della varianza ci offre un quadro teorico per comprendere come la variabilità iniziale della variabile di interesse Y sia scomponibile in due pezzi: la variabilità spiegata dal modello e quella residua (casuale).

Ricordiamo che:

$$\begin{array}{lll} y_i \quad \text{dati} & \hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i, & \hat{\varepsilon}_i = y_i - \hat{y}_i \\ \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i, & \bar{y} = \frac{1}{n}\sum_{i=1}^{n} \hat{y}_i & 0 = \frac{1}{n}\sum_{i=1}^{n} \hat{\varepsilon}_i \end{array}$$

Definizione 17.7.1 (Varianza Totale, Varianza Spiegata e Varianza Residua)**.** La varianza di y (senza osservare x) è

$$\hat{\sigma}^2_Y = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n} = \frac{TSS}{n}, \qquad TSS \text{ Total Sum of Squares}$$

La varianza di $\hat{y}$ è

$$\hat{\sigma}^2_{\hat{Y}} = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{n} = \frac{ESS}{n}, \qquad ESS \text{ Explained Sum of Squares}$$

La varianza di $\hat{\varepsilon}$ è

$$\hat{\sigma}^2_{\varepsilon} = \frac{\sum_{i=1}^{n}(\hat{\varepsilon}_i - 0)^2}{n} = \frac{\sum_{i=1}^{n}\hat{\varepsilon}_i^2}{n} = \frac{RSS}{n}, \qquad RSS \text{ Residual Sum of Squares}$$

Proprietà 17.7.1. *Vale la seguente relazione*

$$TSS = ESS + RSS$$

La variabilità totale di y (quella osservata senza x) e scomponibile nella somma di due parti

$$\left\{ \begin{array}{c} \text{varibilità di } y \\ \text{intorno alla sua media} \end{array} \right\} = \left\{ \begin{array}{c} \text{varibilità della retta} \\ \text{intorno alla media} \end{array} \right\} + \left\{ \begin{array}{c} \text{varibilità delle } y \\ \text{intorno alla retta} \end{array} \right\}$$

Dividendo ogni membro per TSS, si ottiene

$$\begin{aligned} TSS &= ESS + RSS \\ \frac{TSS}{TSS} &= \frac{ESS}{TSS} + \frac{RSS}{TSS} \\ 1 &= \frac{ESS}{TSS} + \frac{RSS}{TSS} \\ \frac{RSS}{TSS} &= 1 - \frac{ESS}{TSS} \\ \frac{RSS}{TSS} &= 1 - r^2, \quad r^2 = \frac{ESS}{TSS} \end{aligned}$$

17.8 Il coefficiente di determinazione lineare R^2

È un indicatore sintetico che indica la quota di varianza spiegata dal modello. Nel caso della regressione lineare semplice

Definizione 17.8.1 (Indice di Determinazione Lineare)**.** Si definisce

$$R^2 = \left(\frac{ESS}{TSS}\right) = r^2 = \left(\frac{\text{cov}(x,y)}{\hat{\sigma}_x \hat{\sigma}_y}\right)^2$$

l'indice di determinazione lineare ed è, nel contesto della regressione lineare semplice, il quadrato dell'indice di correlazione

$$0 \le R^2 \le 1$$

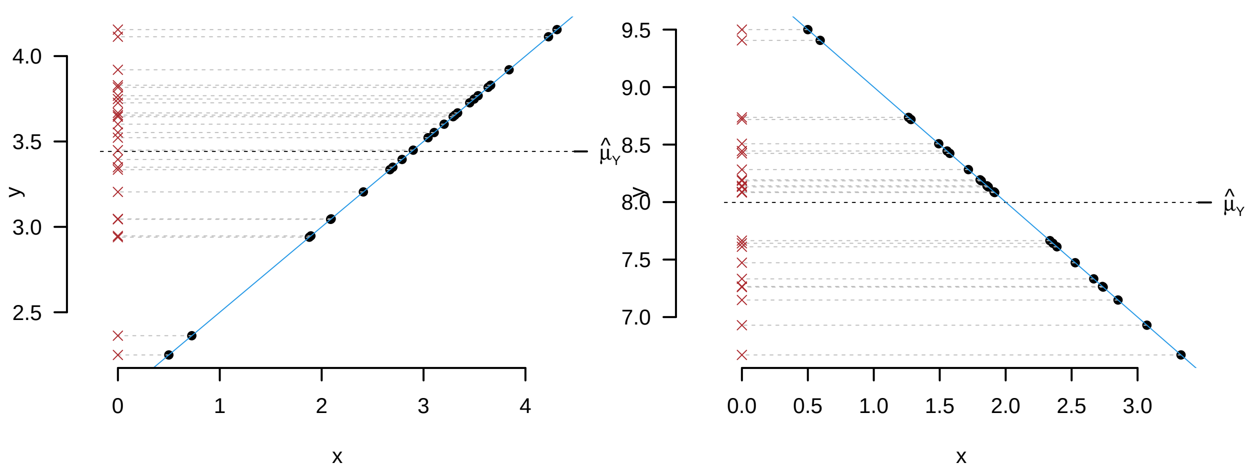
Se $R^2 = 1$, allora $r = -1$ oppure $r = +1$, associazione lineare perfetta, 100% della variabilità spiegata, se $R^2 = 0$, allora $r = 0$ associazione lineare nulla, 0% della variabilità spiegata, se $R^2 > 0.75$, allora considereremo l'associazione lineare *soddisfacente*.

17.8.1 Interpretazione di r^2

La variabilità di y viene scomposta in due, la componente spiegata dalla retta e della residua. Caso limite uno: la retta spiega tutta la variabilità di y:

$$TSS = ESS \Rightarrow RSS = 0 \Rightarrow r^2 = 1$$

i punti sono allineati su una retta

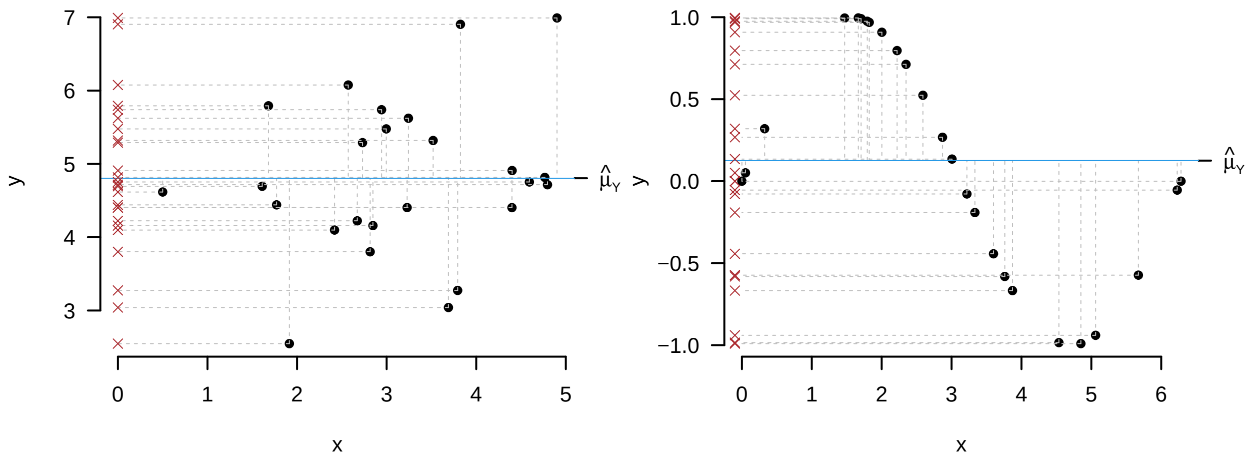


Se la retta è in grado di assorbire tutta la variabilità di y significa che tutti i punti sono allineati sul di essa.

Caso limite due: la retta **non** spiega nulla della variabilità di y

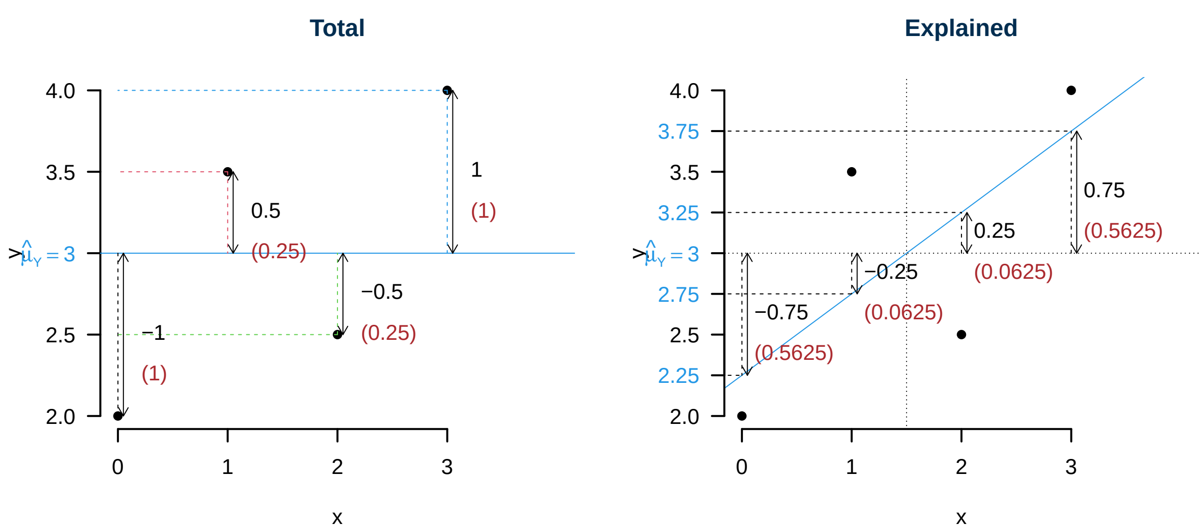
$$TSS = RSS \Rightarrow ESS = 0 \Rightarrow r^2 = 0$$

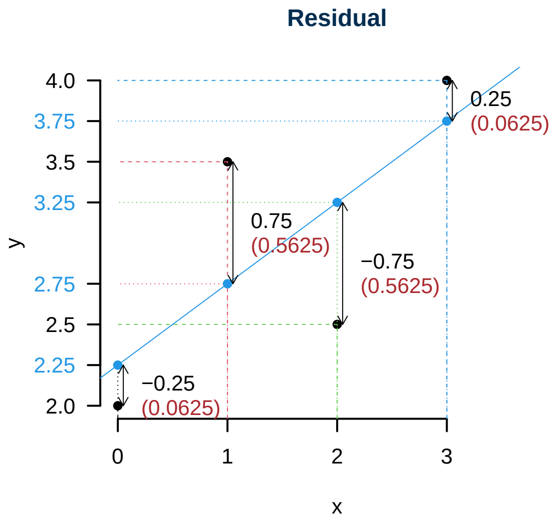
la retta è orizzontale e coincide con $\hat{\mu}_Y$



Se la retta è parallela all'asse dell x non spiega nulla della variabilità di y.

17.8.2 Scomposizione della varianza sui dati di esempio





i	$(y_i - \bar{y})^2$	$(\hat{y}_i - \bar{y})^2$	$\hat{\varepsilon}_i^2$
1	1.00	0.5625	0.0625
2	0.25	0.0625	0.5625
3	0.25	0.0625	0.5625
4	1.00	0.5625	0.0625
Totale	**2.50**	**1.2500**	**1.2500**

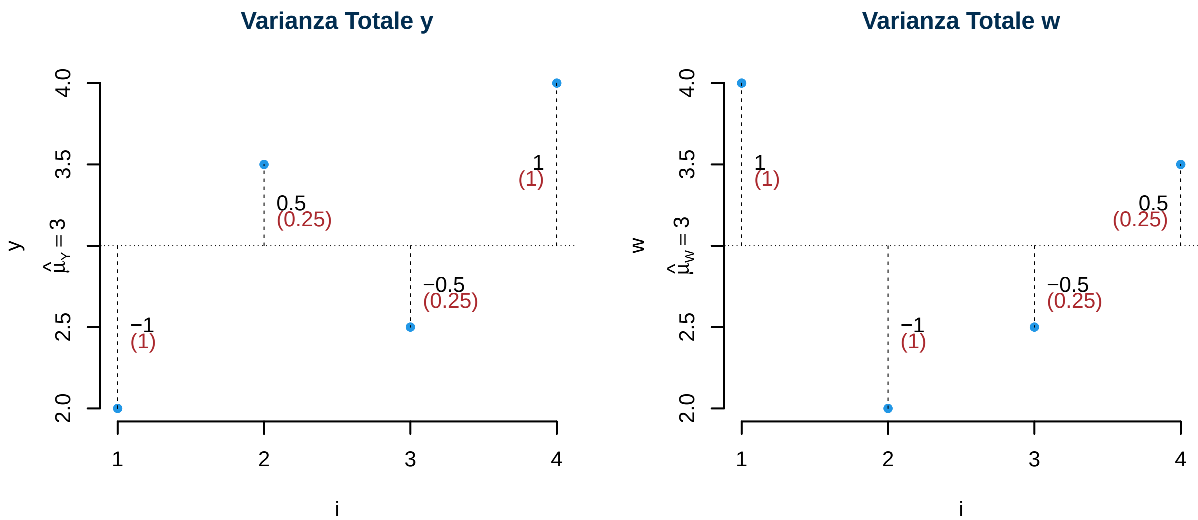
$$\begin{aligned} TSS &= ESS + RSS \\ 2.5 &= 1.25 + 1.25 \\ \frac{RSS}{TSS} &= 1 - r^2 \\ \frac{1.25}{2.5} &= 1 - 0.7071^2 \\ 0.5 &= 1 - 0.5 \end{aligned}$$

Ovvero la retta di regressione di y dato x spiega il 50% della variabilità totale della y.

Osservazione. Se volessi studiare la relazione tra X e W nella tabella qui sotto:

i	x_i	y_i	w_i
1	0	2.0	4.0
2	1	3.5	2.0
3	2	2.5	2.5
4	3	4.0	3.5
Totale	**6**	**12.0**	**12.0**

Osservo dapprima Le variabili y e w hanno la stessa media e la stessa varianza, infatti $\sum_i y_i = \sum_i w_i = 12$ e $\sum_i y_i^2 = \sum_i w_i^2 = 38.5$. Se osservati rispetto ai soli valori, in ipotesi IID, sono fenomeni indistinguibili.



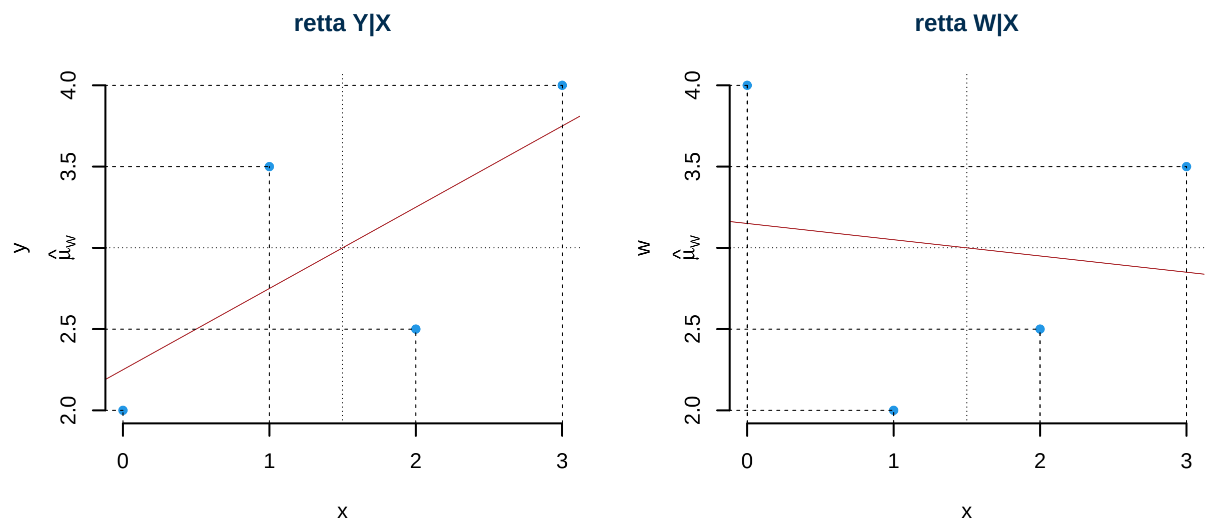
La varianza di y è dunque

$$\hat{\sigma}_Y^2 = \frac{1}{n}\sum_i (y_i - \bar{y})^2 = \frac{1}{n}\sum_i y_i^2 - \bar{y}^2 = 9.625 - (3)^2 = 0.625$$

e quella di w

$$\hat{\sigma}_W^2 = \frac{1}{n}\sum_i (w_i - \bar{w})^2 = \frac{1}{n}\sum_i w_i^2 - \bar{w}^2 = 9.625 - (3)^2 = 0.625$$

Ma la variabile X non spiega nello stesso modo Y e W. Infatti se osserviamo la relazione tra x ed Y e la relazione tra x e w ci accorgiamo che i fenomeni sono diversi.



- La retta $y|x$ assorbe il 50% della variabilità di y
- La retta $w|x$ assorbe il 2% della variabilità di w

17.9 Stima di σ_ε^2

Il parametro σ_ε^2 rappresenta la variabilità dei punti intorno alla retta. La stima di σ_ε^2 deriva dalla scomposizione della varianza

$$\begin{aligned}
\frac{RSS}{TSS} &= 1 - r^2 \\
RSS &= TSS(1 - r^2) \\
\frac{RSS}{n} &= \frac{TSS}{n}(1 - r^2) \\
\hat{\sigma}_\varepsilon^2 &= \hat{\sigma}_Y^2(1 - r^2)
\end{aligned}$$

In modo analogo alla stima di σ^2 in un modello normale, $\hat{\sigma}_\varepsilon^2$ **non** è stima corretta di σ_ε^2 e si dimostra che

$$E(\hat{\sigma}_\varepsilon^2) = \frac{n-2}{n}\sigma_\varepsilon^2$$

si quindi corregge con:

$$S_\varepsilon^2 = \frac{n}{n-2}\hat{\sigma}_\varepsilon^2$$

17.10 Statistiche Sufficienti del Modello di Reegressione

Se l'ipotesi di normalità dei residui viene ritenuta valida si può dimostrare che le stime di massima verosimiglianza per β_0, β_1 e σ_ε^2 coincidono con quelle dei minimi quadrati $\hat{\beta}_0$, $\hat{\beta}_1$ e $\hat{\sigma}_\varepsilon^2$. Tutta l'informazione sul modello di regressione lineare semplice è contenuta nelle seguenti statistiche

$$\sum_{i=1}^{n} x_i, \quad \sum_{i=1}^{n} y_i, \quad \sum_{i=1}^{n} x_i^2, \quad \sum_{i=1}^{n} y_i^2, \quad \sum_{i=1}^{n} x_i y_i$$

i	x_i	y_i	x_i^2	y_i^2	$x_i \cdot y_i$
1	0.0	2.0	0.0	4.000	0.000
2	1.0	3.5	1.0	12.250	3.500
3	2.0	2.5	4.0	6.250	5.000
4	3.0	4.0	9.0	16.000	12.000
Totale	**6.0**	**12.0**	**14.0**	**38.500**	**20.500**
Totale/n	1.5	3.0	3.5	9.625	5.125

Ricapitolando nel nostro esempio:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 1.5 \qquad \hat{\sigma}_X^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2 = 1.25$$

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = 3 \qquad \hat{\sigma}_Y^2 = \frac{1}{n}\sum_{i=1}^{n} y_i^2 - \bar{y}^2 = 0.625$$

$$\mathrm{cov}(x,y) = \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \bar{x}\bar{y} = 0.625 \qquad r = \frac{\mathrm{cov}(x,y)}{\hat{\sigma}_X \hat{\sigma}_Y} = 0.7071$$

$$\hat{\beta}_1 = \frac{\mathrm{cov}(x,y)}{\hat{\sigma}_X^2} = 0.5 \qquad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 2.25.$$

$$\hat{\sigma}_\varepsilon^2 = \hat{\sigma}_Y^2 (1 - r^2) = 0.3125 \qquad S_\varepsilon^2 = \frac{n}{n-2}\hat{\sigma}_\varepsilon^2 = 0.625$$

$$\hat{\sigma}_\varepsilon = \hat{\sigma}_Y \sqrt{(1 - r^2)} = 0.559 \qquad S_\varepsilon = \sqrt{\frac{n}{n-2}}\hat{\sigma}_\varepsilon = 0.7906$$

Inferenza e Diagnostica sul Modello di Regressione Lineare

18

18.1 Teorema di Gauss-Markov

Teorema 18.1.1 (Gauss-Markov). *Sotto gli assunti dallo 0 al 5, gli stimatori dei minimi quadrati sono corretti*

$$E(\hat{\beta}_1) = \beta_1, \qquad E(\hat{\beta}_0) = \beta_0$$

La loro varianza è:

$$\begin{aligned} V(\hat{\beta}_1) &= \frac{\sigma_\varepsilon^2}{n\hat{\sigma}_X^2} \\ V(\hat{\beta}_0) &= \sigma_\varepsilon^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{n\hat{\sigma}_X^2} \right) \\ cov(\hat{\beta}_0, \hat{\beta}_1) &= -\sigma_\varepsilon^2 \frac{\bar{x}}{n\hat{\sigma}_X^2} = -\bar{x}V(\hat{\beta}_1) \end{aligned}$$

Gli stimatori $\hat{\beta}_0$ e $\hat{\beta}_1$ di β_0 e β_1 sono BLUE (Best Linear Unbiased Estimators).

18.2 La previsione $\hat{Y}$

Definiamo la previsione di Y per un valore x qualunque

$$\hat{Y}_{(X=x)} = \hat{\beta}_0 + \hat{\beta}_1 x$$

Se $X = x_i$ allora

$$\hat{Y}_{(X=x_i)} = \hat{Y}_i$$

La previsione è *corretta*

$$E\left(\hat{Y}_{(X=x)}\right) = E(\hat{\beta}_0 + \hat{\beta}_1 x) = E(\hat{\beta}_0) + E(\hat{\beta}_1)x = \beta_0 + \beta_1 x$$

La varianza

$$V(\widehat{Y}_{(X=x)}) = \sigma_\varepsilon^2 \left(\frac{1}{n} + \frac{(x-\bar{x})^2}{n\hat{\sigma}_X^2} \right)$$

18.3 Standard Errors e Stima degli SE

Gli Errori Standard (ES) delle stime sono le radici quadrate delle varianze degli stimatori

$$SE(\hat{\beta}_0) \quad = \quad \sqrt{\hat{\beta}_0}$$

$$
\begin{aligned}
&= \sqrt{\sigma_\varepsilon^2\left(\frac{1}{n}+\frac{\bar{x}^2}{n\hat{\sigma}_X^2}\right)} \\
SE(\hat{\beta}_1) &= \sqrt{V(\hat{\beta}_1)} \\
&= \sqrt{\frac{\sigma_\varepsilon^2}{n\hat{\sigma}_X^2}} \\
SE(\widehat{Y}_{(X=x)}) &= \sqrt{V(\widehat{Y}_{(X=x)})} \\
&= \sqrt{\sigma_\varepsilon^2\left(\frac{1}{n}+\frac{(x-\bar{x})^2}{n\hat{\sigma}_X^2}\right)}
\end{aligned}
$$

Ricordando che

$$S_\epsilon = \sqrt{\frac{n}{n-2}}\,\hat{\sigma}_Y\sqrt{1-r^2}$$

Otteniamo le stime campionarie per gli Standard Errors

$$
\begin{aligned}
\widehat{SE(\hat{\beta}_0)} &= \sqrt{S_\varepsilon^2\left(\frac{1}{n}+\frac{\bar{x}^2}{n\hat{\sigma}_X^2}\right)} \\
\widehat{SE(\hat{\beta}_1)} &= \sqrt{\frac{S_\varepsilon^2}{n\hat{\sigma}_X^2}} \\
\widehat{SE(\widehat{Y}_{X=x})} &= \sqrt{S_\varepsilon^2\left(\frac{1}{n}+\frac{(x-\bar{x})^2}{n\hat{\sigma}_X^2}\right)}
\end{aligned}
$$

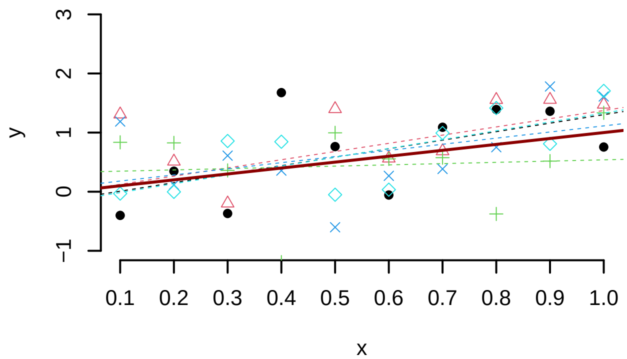
18.4 Inferenza su β_0 e β_1 e su $\hat{Y}$

Se l'assunto 6. (normalità dei residui) è rispettato, allora

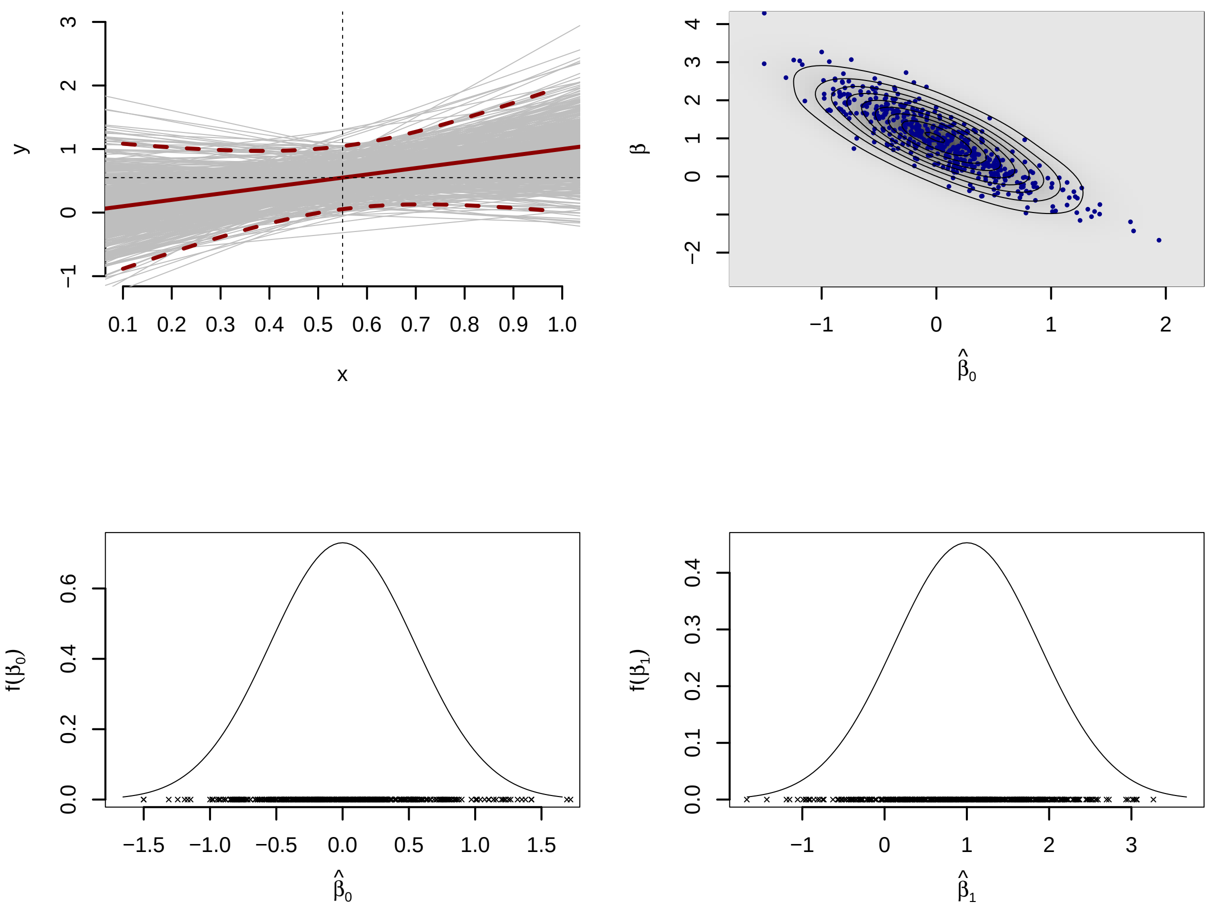
$$\hat{\beta}_1 \sim N(\beta_1, V(\hat{\beta}_1)),\quad \hat{\beta}_0 \sim N(\beta_0, V(\hat{\beta}_0)),\quad \hat{Y}_{(X=x)} \sim N\left(\beta_0+\beta_1 x, V(\hat{Y}_{(X=x)})\right)$$

Fisso $n=10$, $x_1=0.1, x_2=0.2, ..., x_{10}=1.0$, $\beta_0=0$, $\beta_1=1$ e $\sigma_\varepsilon^2=0.64$. Simulo 5 volte $n=10$ punti da

$$Y_i = 0 + 1\cdot x_i + \varepsilon_i,\quad \varepsilon_i \sim N(0, 0.64)$$



E ripeto le simulazioni ripetute 500 volte:



La varianza di $\hat{Y}$ è l'*errore di previsione*

$$V(\widehat{Y}_{(X=x)}) = \sigma^2_\varepsilon \left(\frac{1}{n} + \frac{(x-\bar{x})^2}{n\hat{\sigma}^2_X} \right)$$

18.4.1 Interpolazione e Estrapolazione

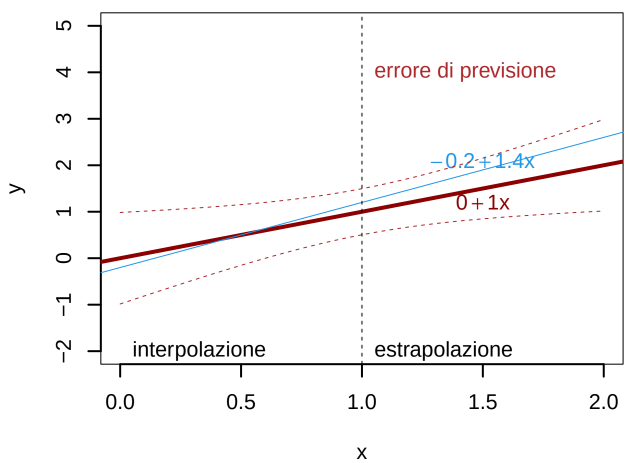
Parliamo di *interpolazione* dei punti se $\hat{Y}_{(X=x)}$ è calcolato per

$$\min\{x_i\} \le x \le \max\{x_i\}$$

Parliamo di *estrapolazione* dei punti se $\hat{Y}_{(X=x)}$ è calcolato per

$$x < \min\{x_i\} \quad \text{oppure} \quad x > \max\{x_i\}$$

Voglio le vedere le previsioni per $x > 1$



18.4.2 Intervalli di Confidenza per β_0, β_1 e $\hat{Y}$

Si dimostra che

$$\frac{\hat{\beta}_0 - \beta_0}{\widehat{SE(\hat{\beta}_0)}} \sim t_{n-2}$$

$$\frac{\hat{\beta}_1 - \beta_1}{\widehat{SE(\hat{\beta}_1)}} \sim t_{n-2}$$

$$\frac{\hat{Y}_{(X=x)} - (\beta_0 + \beta_1 x)}{\widehat{SE(\hat{Y}_{(X=x)})}} \sim t_{n-2}$$

E dunque gli IdC al livello $(1-\alpha)$ sono dati da

$$\begin{array}{rcl} \hat{\beta}_0 & \pm & t_{n-2;\alpha/2}\widehat{SE(\hat{\beta}_0)} \\ \hat{\beta}_1 & \pm & t_{n-2;\alpha/2}\widehat{SE(\hat{\beta}_1)} \\ \hat{Y}_{(X=x)} & \pm & t_{n-2;\alpha/2}\widehat{SE(\hat{Y}_{(X=x)})} \end{array}$$

18.4.3 Test per β_0, e β_1

Consideriamo i due seguenti sistemi di ipotesi

$$\begin{cases} H_0 : \beta_0 = \beta_{0,H_0} \\ H_1 : \text{da scegliere } \neq, > \text{ o } < \end{cases} \qquad \begin{cases} H_0 : \beta_1 = \beta_{1,H_0} \\ H_1 : \text{da scegliere } \neq, > \text{ o } < \end{cases}$$

Sapendo che

$$\widehat{SE(\hat{\beta}_0)} = \sqrt{S_\varepsilon^2\left(\frac{1}{n} + \frac{\bar{x}^2}{n\hat{\sigma}_X^2}\right)} \qquad \widehat{SE(\hat{\beta}_1)} = \sqrt{\frac{S_\varepsilon^2}{n\hat{\sigma}_X^2}}$$

Sotto H_0

$$\begin{array}{rcl} \dfrac{\hat{\beta}_0 - \beta_{0,H_0}}{\widehat{SE(\hat{\beta}_0)}} & \sim & t_{n-2} \\ \dfrac{\hat{\beta}_1 - \beta_{1,H_0}}{\widehat{SE(\hat{\beta}_1)}} & \sim & t_{n-2} \end{array}$$

otteniamo

$$\begin{array}{rcl} t_{0,\text{obs}} & = & \dfrac{\hat{\beta}_0 - \beta_{0,H_0}}{\widehat{SE(\hat{\beta}_0)}} \\ t_{1,\text{obs}} & = & \dfrac{\hat{\beta}_1 - \beta_{1,H_0}}{\widehat{SE(\hat{\beta}_1)}} \end{array}$$

Che andranno lette nella direzione di H_1 con le solite regole

18.4.4 Esempio sui 4 punti

Tipicamente quando si fa un modello di regressione per prima cosa si testa la significatività dei coefficienti:

$$\begin{cases} H_0 : \beta_0 = 0 \\ H_1 : \beta_0 \neq 0 \end{cases} \qquad \begin{cases} H_0 : \beta_1 = 0 \\ H_1 : \beta_1 \neq 0 \end{cases}$$

Calcolo gli standard errors stimati

$$V(\hat\beta_1)=\frac{\sigma^2_\varepsilon}{n\sigma^2_X}\qquad \widehat{SE(\hat\beta_1)}=\sqrt{\frac{S^2_\varepsilon}{n\hat\sigma^2_X}}=\sqrt{\frac{0.625}{4\times 1.25}}=0.3536$$

$$V(\hat\beta_0)=\sigma^2_\varepsilon\left(\frac{1}{n}+\frac{\bar x^2}{n\sigma^2_X}\right)\qquad \widehat{SE(\hat\beta_0)}=\sqrt{S^2_\varepsilon\left(\frac{1}{n}+\frac{\bar x^2}{n\sigma^2_X}\right)}=\sqrt{0.625\left(\frac{1}{4}+\frac{1.5^2}{4\times 1.25}\right)}=0.6614$$

18.4.5 Calcolo dei valori osservati e dei valori critici

otteniamo

$$t_{0,\text{obs}}=\frac{\hat\beta_0-0}{\widehat{SE(\hat\beta_0)}}=3.4017$$

$$t_{1,\text{obs}}=\frac{\hat\beta_1-0}{\widehat{SE(\hat\beta_1)}}=1.4142$$

dalle tavole ricaviamo - $t_{n-2;0.025}=4.3027$ - $t_{n-2;0.005}=9.9248$

e osserviamo che, per il coefficiente β_0:

- $|t_{0,\text{obs}}|=|3.4017|<4.3027=t_{n-2;0.025}$ e quindi non rifiuto H_0 al livello di significatività $\alpha=0.05$.
- $|t_{0,\text{obs}}|=|3.4017|<9.9248=t_{n-2;0.005}$ e quindi non rifiuto H_0 al livello di significatività $\alpha=0.01$.

per il coefficiente β_1

- $|t_{1,\text{obs}}|=|1.4142|<4.3027=t_{n-2;0.025}$ e quindi non rifiuto H_0 al livello di significatività $\alpha=0.05$.
- $|t_{1,\text{obs}}|=|1.4142|<9.9248=t_{n-2;0.005}$ e quindi non rifiuto H_0 al livello di significatività $\alpha=0.01$.

18.4.6 Se $n=10$

Si supponga che a parità di statistiche $\bar x$, $\bar y$, $\hat\sigma_X$, $\hat\sigma_Y$, e $\text{cov}(X,Y)$, ma per $n=10$, si ottiene:

Calcolo gli standard errors stimati

$$V(\hat{\beta}_1) = \frac{\sigma_\varepsilon^2}{n\sigma_X^2} \qquad \widehat{SE(\hat{\beta}_1)} = \sqrt{\frac{S_\varepsilon^2}{n\hat{\sigma}_X^2}} = \sqrt{\frac{0.3906}{10 \times 1.25}} = 0.1768$$

$$V(\hat{\beta}_0) = \sigma_\varepsilon^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{n\sigma_X^2}\right) \quad \widehat{SE(\hat{\beta}_0)} = \sqrt{S_\varepsilon^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{n\sigma_X^2}\right)} = \sqrt{0.3906 \left(\frac{1}{10} + \frac{1.5^2}{10 \times 1.25}\right)} = 0.3307$$

Calcolo dei valori osservati e dei valori critici e ottengo

$$t_{0,\text{obs}} = \frac{\hat{\beta}_0 - 0}{\widehat{SE(\hat{\beta}_0)}} = 6.8034$$

$$t_{1,\text{obs}} = \frac{\hat{\beta}_1 - 0}{\widehat{SE(\hat{\beta}_1)}} = 2.8284$$

dalle tavole ricaviamo $t_{n-2;0.025} = 2.306$ e $t_{n-2;0.005} = 3.3554$

e osserviamo che, per il coefficiente β_0:

- $|t_{0,\text{obs}}| = |6.8034| > 2.306 = t_{n-2;0.025}$ e quindi rifiuto H_0 al livello di significatività $\alpha = 0.05$.
- $|t_{0,\text{obs}}| = |6.8034| > 3.3554 = t_{n-2;0.005}$ e quindi rifiuto H_0 al livello di significatività $\alpha = 0.01$.

per il coefficiente β_1

- $|t_{1,\text{obs}}| = |2.8284| > 2.306 = t_{n-2;0.025}$ e quindi rifiuto H_0 al livello di significatività $\alpha = 0.05$.
- $|t_{1,\text{obs}}| = |2.8284| < 3.3554 = t_{n-2;0.005}$ e quindi non rifiuto H_0 al livello di significatività $\alpha = 0.01$.

18.5 Il modello di regressione lineare multiplo

Si tratta di un'estensione del modello semplice ma con più variabili indipendenti

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + ... + \beta_k X_{ik} + \varepsilon_i$$

La nube dei punti non può più essere rappresentata per $k > 2$ e le soluzioni dei minimi quadrati tanto meno. Il coefficiente di determinazione lineare R^2 soffre di grossi limiti all'aumentare del numero di k. Per valutare quanto bene il modello si adatta ai dati si devo compiere analisi statistiche ulteriori.

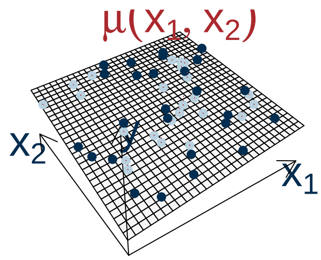


Figura 18.1: Nube in 3D, i punti tali per cui $\hat{y}_i > y_i$ sono colorati in blu scuro, quelli per cui $\hat{y}_i < y_i$, sono colorati in azzuro chiaro

18.6 Analisi dei Residui

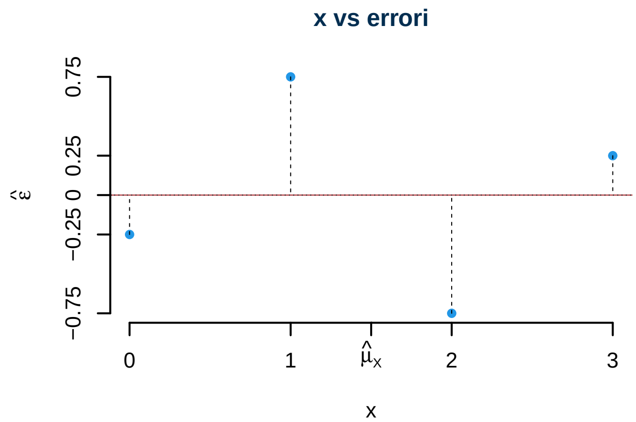
Il modello lineare poggia sugli assunti che abbiamo elencato. La violazione degli assunti invalida le procedure inferenziali La **analisi dei residui** è una serie di procedure diagnostiche per controllare che gli assunti siano rispettati. Le procedure consistono nel produrre statistiche e grafici sui residui osservati $\hat{\varepsilon}_i$

18.6.1 Diagramma dei residui e retta dei residui

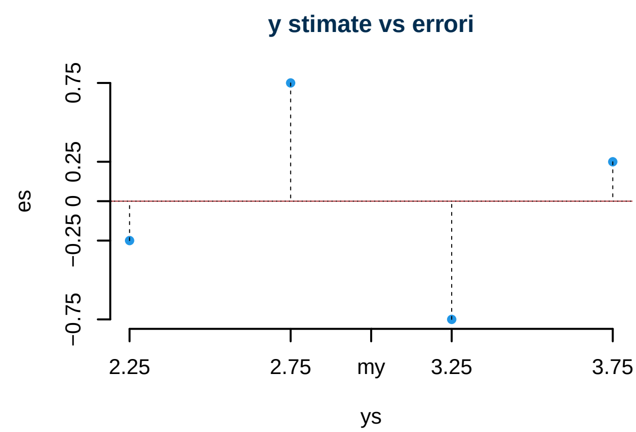
Un modo comune per visualizzare i residui consiste nel mettere in ordinata le x_i e in ascissa i $\hat{\varepsilon}_i$. Per costruzione

$$\mathrm{cov}(x, \hat{\varepsilon}) = 0$$

e dunque la **retta dei residui**, che è la retta di regressione tra x e $\hat{\varepsilon}$ è parallela all'asse delle x e coincide con esso. Esempio sui 4 punti

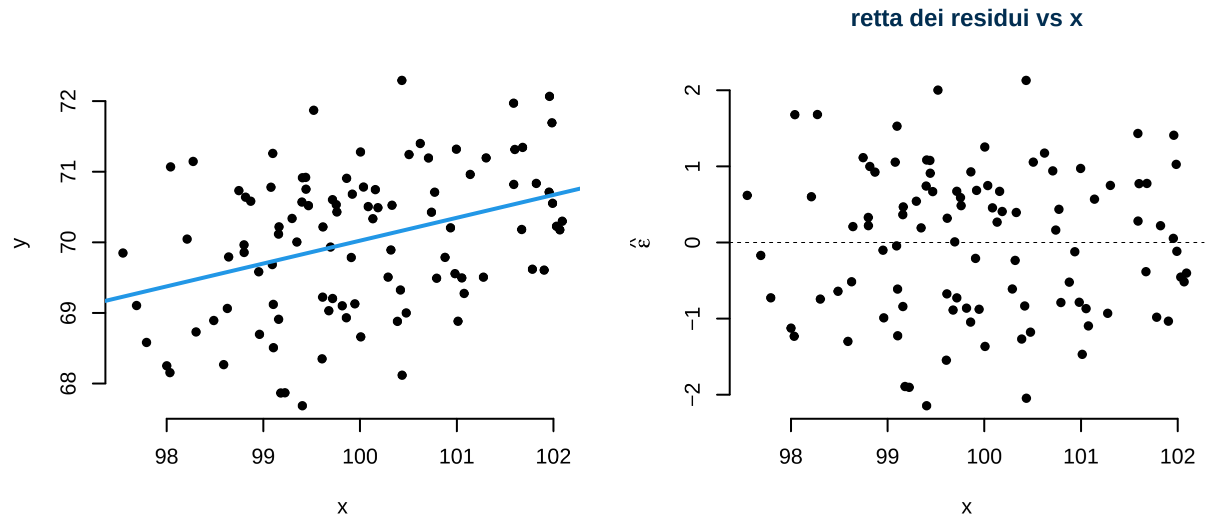


essendo $\hat{y}_i$ combinazioni lineari degli x_i se mettiamo gli $\hat{y}_i$ in ordinata e gli ε_i in ascissa otteniamo lo stesso grafico

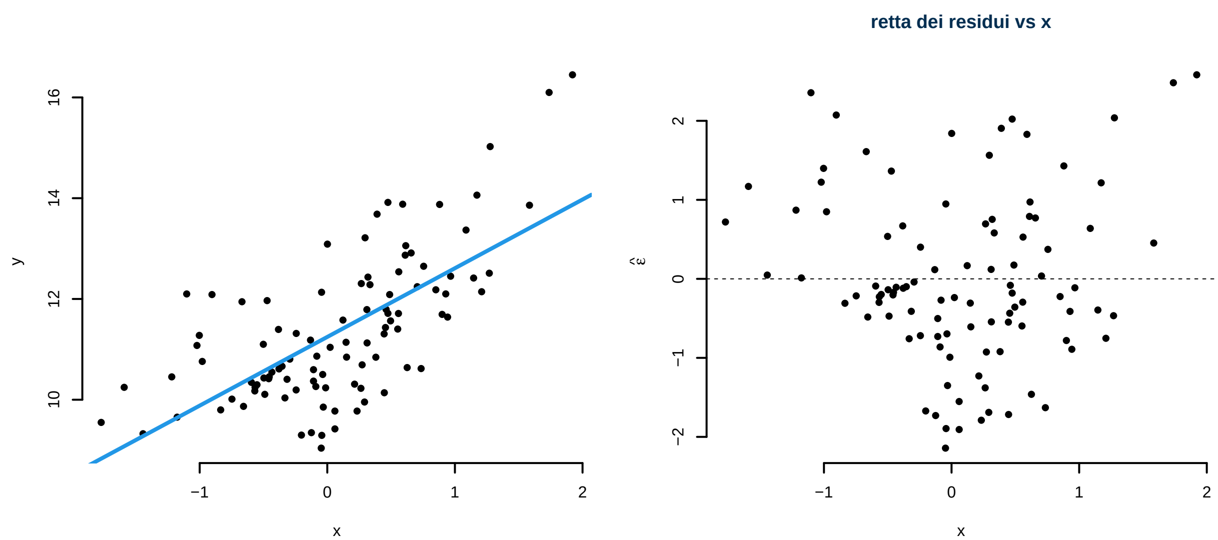


18.6.2 Lettura del Diagramma dei residui

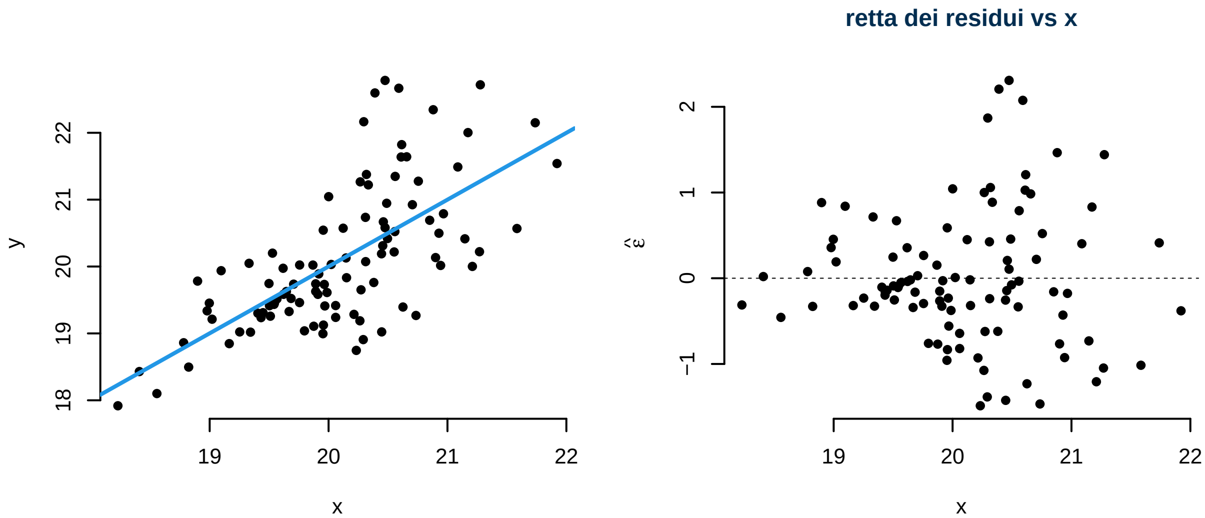
Se tutte le assunzioni sono rispettate i residui devono essere distribuiti in modo uniforme intorno alla retta dei residui



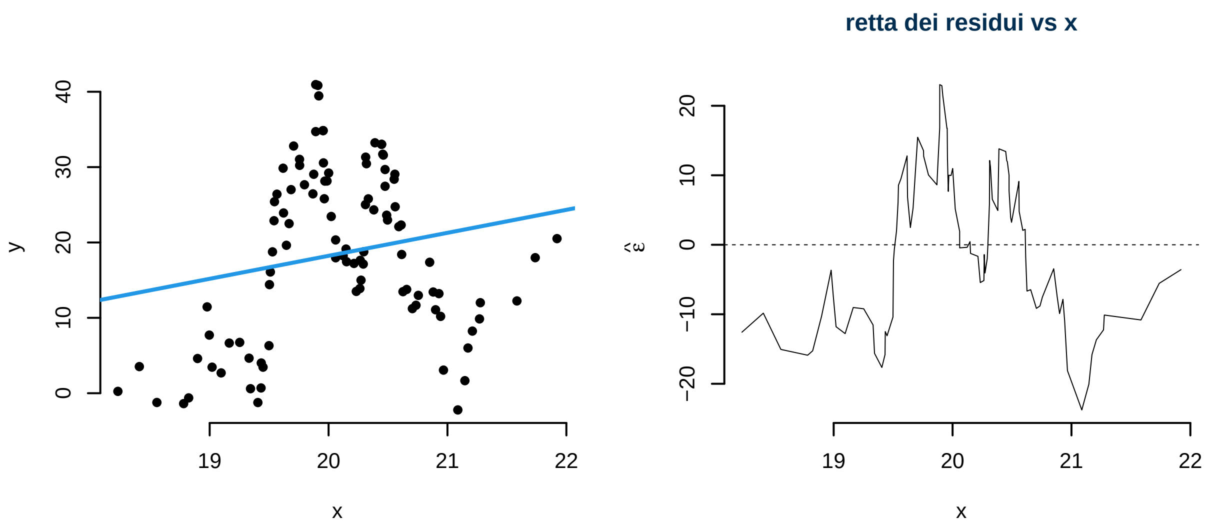
Violazione dell'assunto 0 (i punti provengono da una relazione lineare)



Violazione degli assunti 2. e 4. (omoschedasticità e indipendenza tra x e ε)

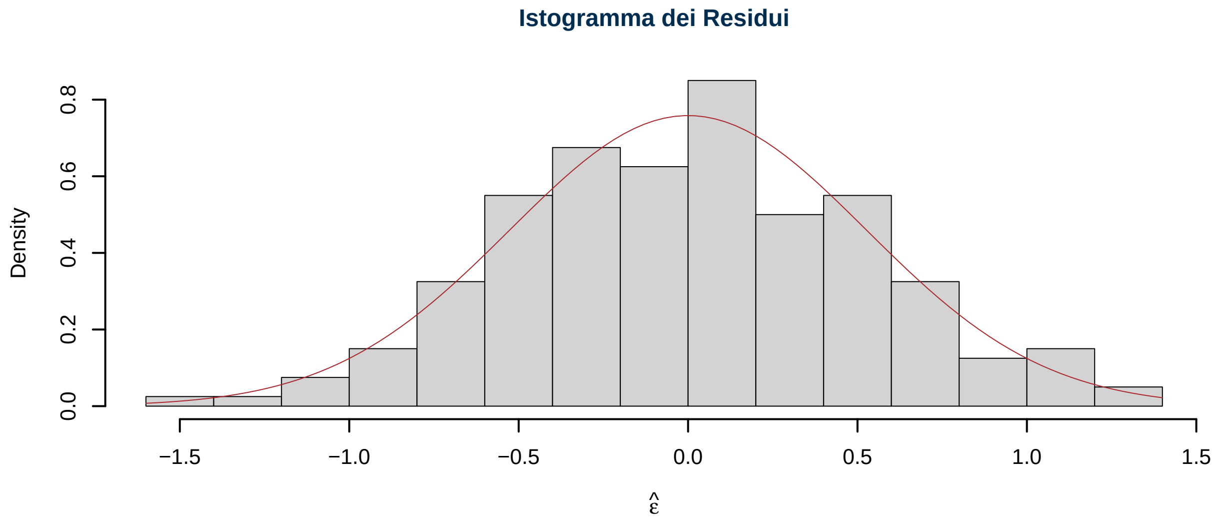


Violazione dell'assunto 3. (indipendenza gli ε)

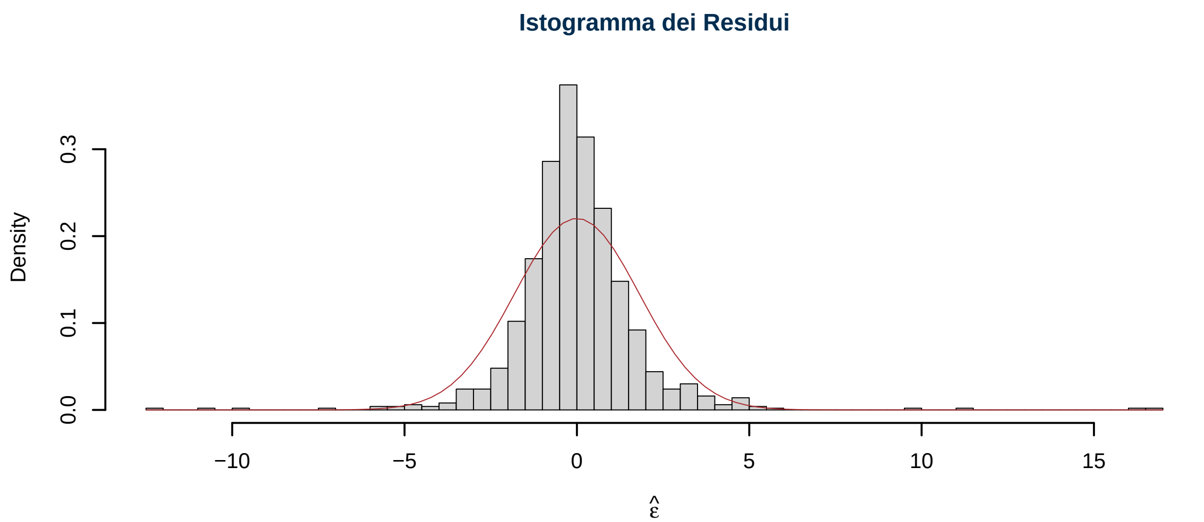


Violazioni dell'assunto 6. (normalità dei residui)

Per diagnosticare se i residui provengono da una normale ci sono diverse tecniche. Per esempio l'istogramma dei residui: si costruisce l'istogramma di frequenza e lo si compara con la normale. Se gli assunti sono rispettati ci aspetteremmo una situazione del genere



Esempio di assunto **non** rispettato



18.6.3 Normal QQ plot

Si tratta di un grafico che mette sull'asse delle x i *quantile* (percentili in inglese) teorici della normale e in ordinata i *quantile* osservati dei residui sul campione.

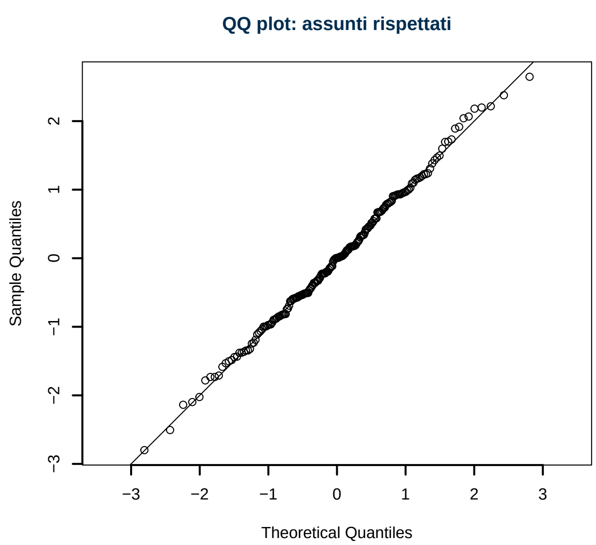
Si crea una tabella dei percentili degli $\hat{\varepsilon}_{(1)}$

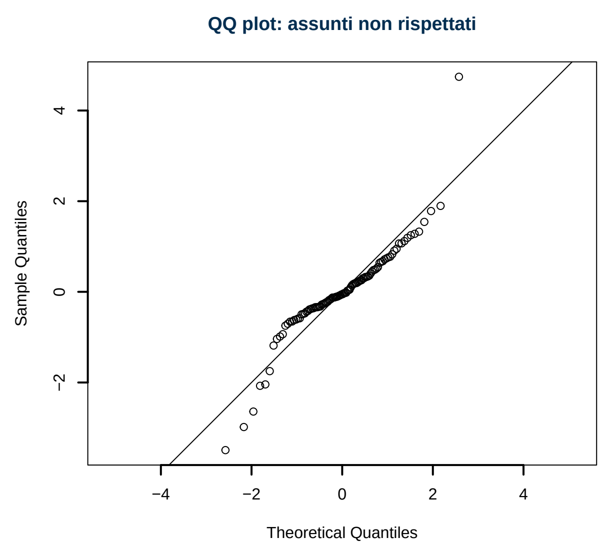
errori ordinati	ordine del percentile	percentile teorico
$\hat{\varepsilon}_{(1)}$	$1/n$	$z_{1/n}$
$\hat{\varepsilon}_{(2)}$	$2/n$	$z_{2/n}$
...	...	...

dove $z_{i/n}$ è il percentile di un $Z \sim N(0,1)$:

$$z_{i/n} : P(Z \leq z_{i/n}) = i/n$$

Se le $z_{i/n}$ e le $\hat{\varepsilon}_i$ giacciono su una retta, allora gli errori si possono assumere normali, tanto più i punti si allontanano tanto più l'ipotesi è violata.

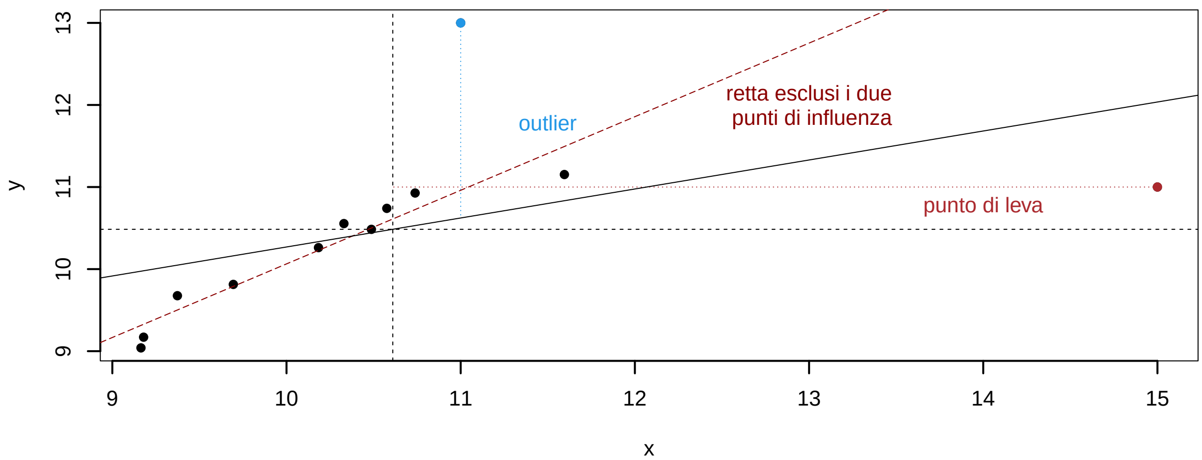




18.7 Punti di leva, Outliers e punti influenti

Ci sono tre tipi di dati anomali, in particolare

- Outlier: osservazione con residuo anomale (sulle y)
- Leverage: (punto di leva), valore anomalo (sulle x)
- Influence Points: (punti influenti) osservazioni con comportamento anomalo che influenzano notevolmente i risultati



18.7.1 Punti di leva

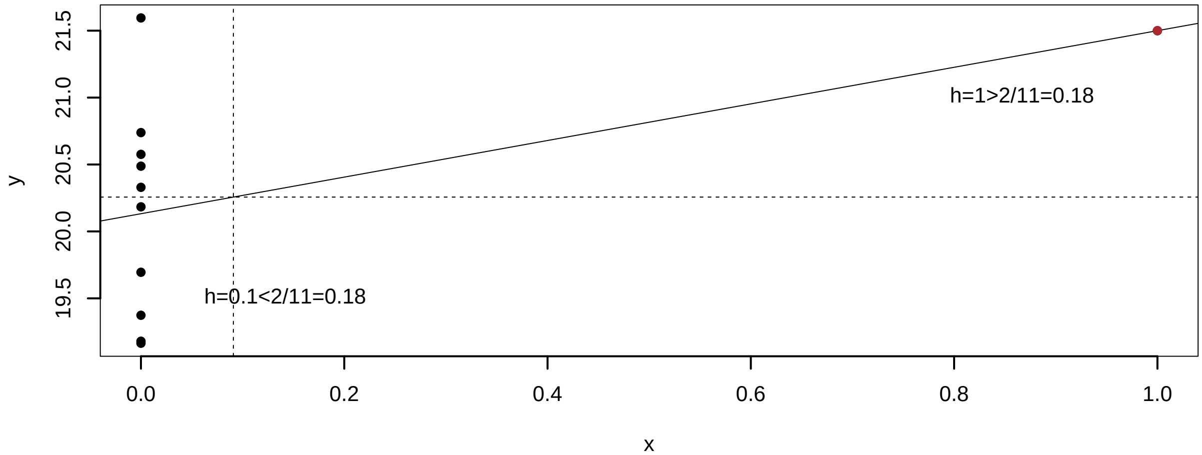
Possiamo misurare la distanza di ogni singola x_i dalla propria media $\bar{x}$ con la seguente misura chiamata *leva*

$$h_i = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{n\hat{\sigma}_X^2}$$

Valori di x con indice di leva alto sono *lontani* dal centro. In particolare se

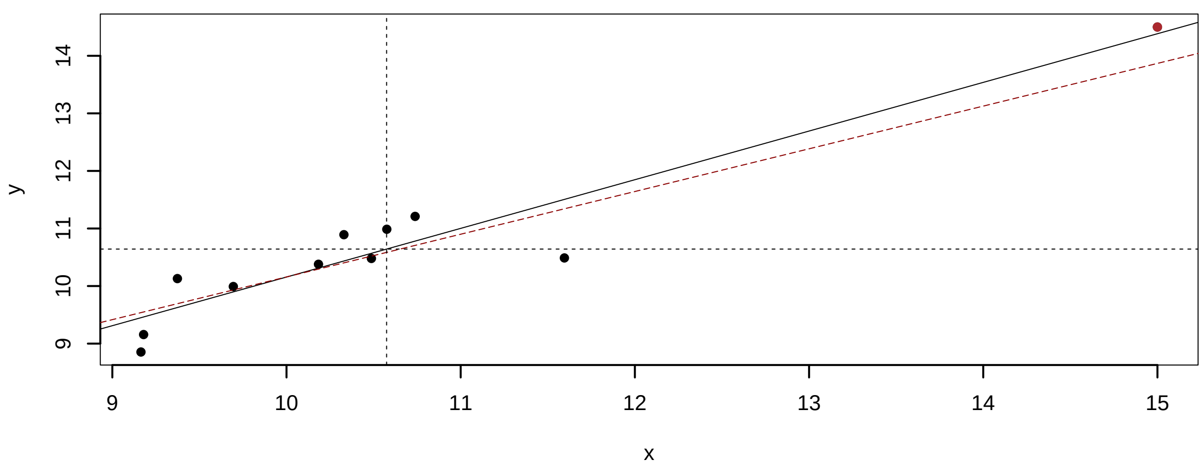
$$h_i > \frac{2}{n}$$

allora x_i è un *punto di leva*. I Punti di leva possono avere effetto sul calcolo dei coefficienti di regressione. I punti a alta leva (con $h_i > 2/n$) sono nei valori estremi della variabile esplicativa e sono potenzialmente influenti, nel senso che possono influenzare in misura rilevante la pendenza della RdR. Infatti, i punti di leva possono portare a risultati forvianti per esempio



x_i	0.00	0.00	0.00	0.0	0.00	0.00	0.00	0.00	0.00	0.00	1.0
y_i	19.37	20.18	19.16	21.6	20.33	19.18	20.49	20.74	20.58	19.69	21.5
h_i	0.10	0.10	0.10	0.1	0.10	0.10	0.10	0.10	0.10	0.10	1.0

Ma non sempre un punto a leva alta è un punto influente



x_i	9.37	10.18	9.16	11.60	10.33	9.18	10.49	10.74	10.58	9.69	15.00
y_i	10.13	10.38	8.85	10.49	10.89	9.16	10.48	11.21	10.99	9.99	14.50
h_i	0.14	0.10	0.16	0.13	0.09	0.16	0.09	0.09	0.09	0.12	0.82

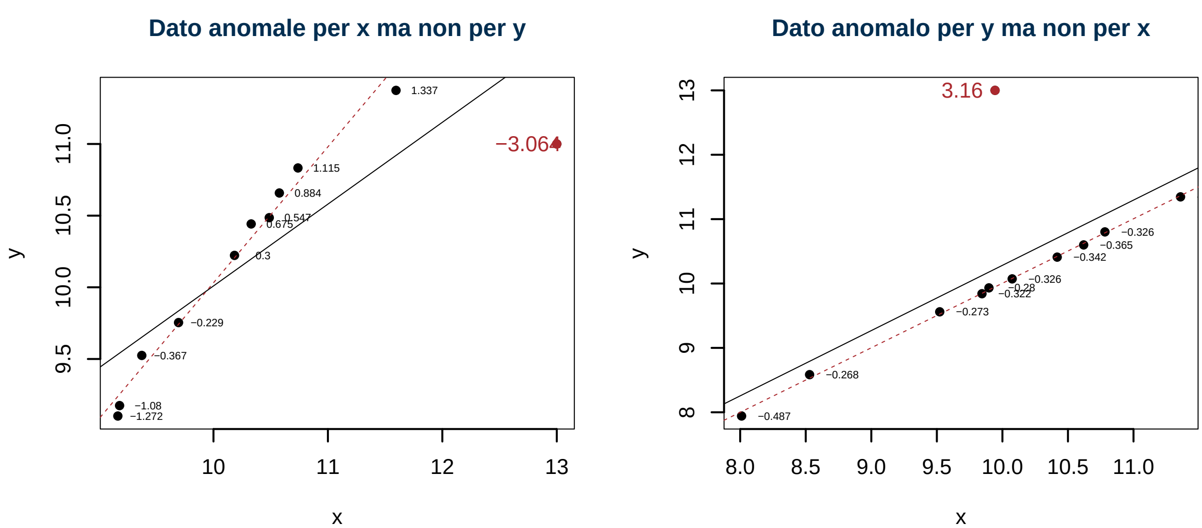
18.7.2 I residui Studentizzati

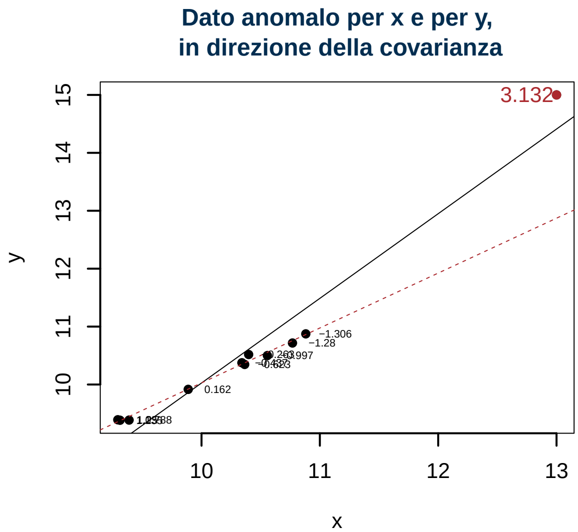
La studentizzazione è una specie di standardizzazione nella quale si tiene conto anche dei valori di leva. I residui studentizzati sono dati da:

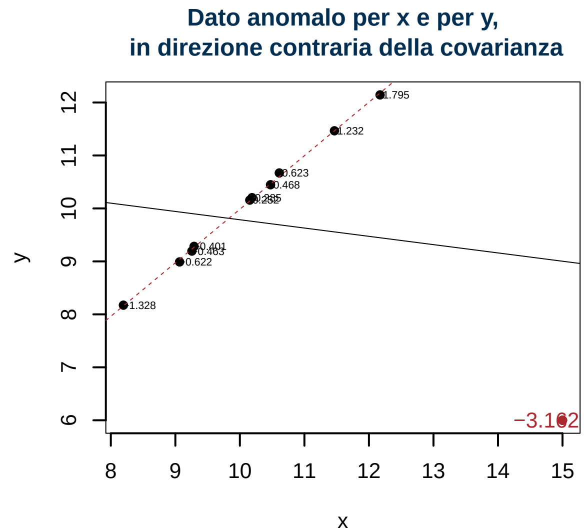
$$\tilde{\varepsilon}_i = \frac{\hat{\varepsilon}_i}{S_\varepsilon \sqrt{1-h_i}} \sim t_{n-2}$$

Si preferiscono i residui studentizzati perché incorporano le leve e sono più confrontabili. La distribuzione è t con $n-2$ gradi di libertà, se per qualche i, $|\tilde{\varepsilon}_i| > t_{\alpha;n-2}$ allora siamo in presenza di punti anomali che diventano **punti influenti** per il calcolo di $\hat{\beta}_0$ e$\hat{\beta}_1$.

Esempi

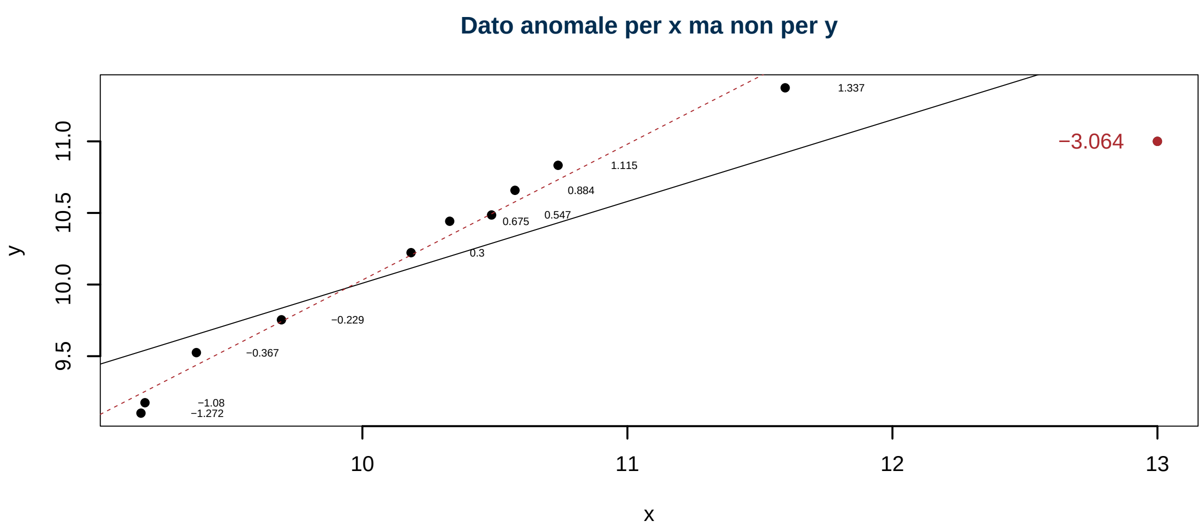






Esempio numerico

x_i	9.37	10.18	9.16	11.60	10.33	9.18	10.49	10.74	10.58	9.69	13.00
y_i	9.52	10.22	9.10	11.37	10.44	9.18	10.49	10.83	10.66	9.75	11.00
h_i	0.17	0.09	0.21	0.20	0.09	0.20	0.09	0.10	0.09	0.13	0.62
$\tilde{\varepsilon}_i$	-0.37	0.30	-1.27	1.34	0.68	-1.08	0.55	1.11	0.88	-0.23	-3.06



18.8 Relazione tra $Y|X$ e $X|Y$

Fin'ora abbiamo considerato il modello

$$y_i = \beta_0 + \beta_1 + \varepsilon_i$$

Supponiamo di invertire il ruolo della x con la y

$$x_i = \alpha_0 + \alpha_1 y_i + \delta_i$$

Le stime dei minimi quadrati sono analoghe

$$\begin{array}{ll} \hat{\beta}_1 = \dfrac{\mathrm{cov}(x,y)}{\hat{\sigma}_X^2} & \hat{\alpha}_1 = \dfrac{\mathrm{cov}(x,y)}{\hat{\sigma}_Y^2} \\ \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} & \hat{\alpha}_0 = \bar{x} - \hat{\alpha}_1 \bar{y} \end{array}$$

In generale

$$\hat{\beta}_0 \neq \hat{\alpha}_0, \qquad \hat{\beta}_1 \neq \hat{\alpha}_1$$

In particolare

$$\hat{\beta}_1 = \hat{\alpha}_1, \text{ se e solo se } \hat{\sigma}_X^2 = \hat{\sigma}_Y^2$$

Mentre

$$\hat{\beta}_0 = \hat{\alpha}_0, \text{ se e solo se } \hat{\beta}_1 = \hat{\alpha}_1, \text{ e se } \bar{y} = \bar{x}$$

18.8.1 Relazione tra gli α i β ed r

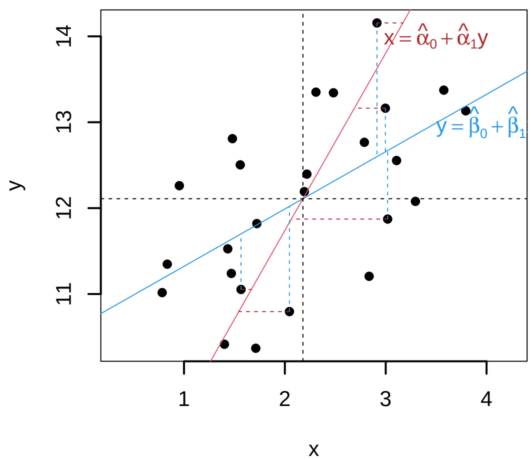
Essendo:

$$\begin{aligned} \hat{\beta}_1 &= \frac{\mathrm{cov}(x,y)}{\hat{\sigma}_X^2} \\ \hat{\beta}_1 &= \frac{\hat{\sigma}_Y}{\hat{\sigma}_Y} \frac{\mathrm{cov}(x,y)}{\hat{\sigma}_X^2} \\ \hat{\beta}_1 &= \frac{\hat{\sigma}_Y}{\hat{\sigma}_X} \frac{\mathrm{cov}(x,y)}{\hat{\sigma}_Y \hat{\sigma}_X} \\ \hat{\beta}_1 &= \frac{\hat{\sigma}_Y}{\hat{\sigma}_X} r \end{aligned}$$

Quindi:

$$\begin{aligned} \hat{\alpha}_1 &= \frac{\mathrm{cov}(x,y)}{\hat{\sigma}_Y^2} \\ \hat{\alpha}_1 &= \frac{\hat{\sigma}_X}{\hat{\sigma}_Y} r \end{aligned}$$

Graficamente

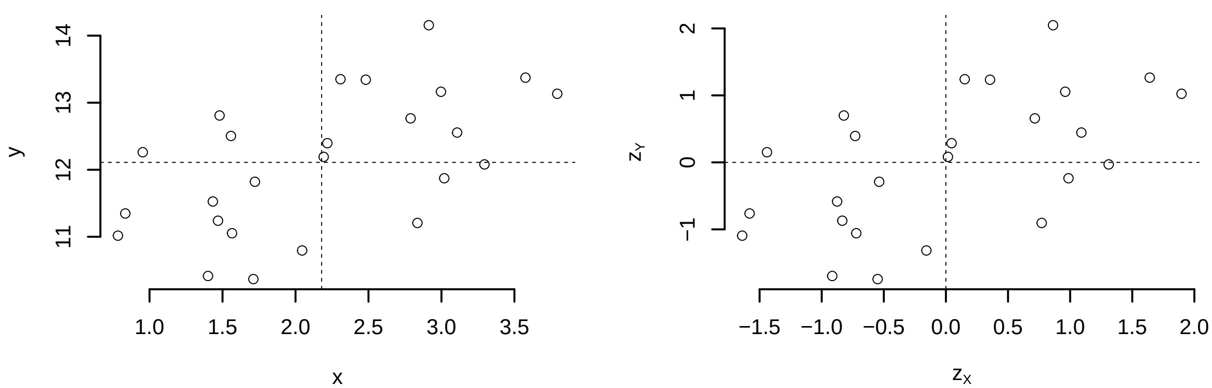


18.8.2 Regressione sulle variabili standardizzate

Se standardizziamo sia x che y, otteniamo

$$z_{Xi} = \frac{x_i - \bar{x}}{\hat{\sigma}_X} \qquad z_{Yi} = \frac{y_i - \bar{y}}{\hat{\sigma}_Y}$$

Abbiamo eliminato l'unità di misura sia da x che da y e centrato la nube dei dati



I dati standardizzati hanno media zero e varianza 1

$$\frac{1}{n}\sum_{i=1}^{n} z_{Xi} = 0 \quad \frac{1}{n}\sum_{i=1}^{n} z_{Xi}^2 = 1$$
$$\frac{1}{n}\sum_{i=1}^{n} z_{Yi} = 0 \quad \frac{1}{n}\sum_{i=1}^{n} z_{Yi}^2 = 1$$

Dalle proprietà del coefficiente di correlazione

$$r_{Z_X,Z_Y} = r_{X,Y} = r$$

E dunque

$$\begin{aligned} r &= \frac{\mathrm{cov}(z_X, z_Y)}{\hat\sigma_{Z_X}\hat\sigma_{Z_Y}} \\ &= \frac{\mathrm{cov}(z_X, z_Y)}{1 \times 1} \\ &= \mathrm{cov}(z_X, z_Y) \end{aligned}$$

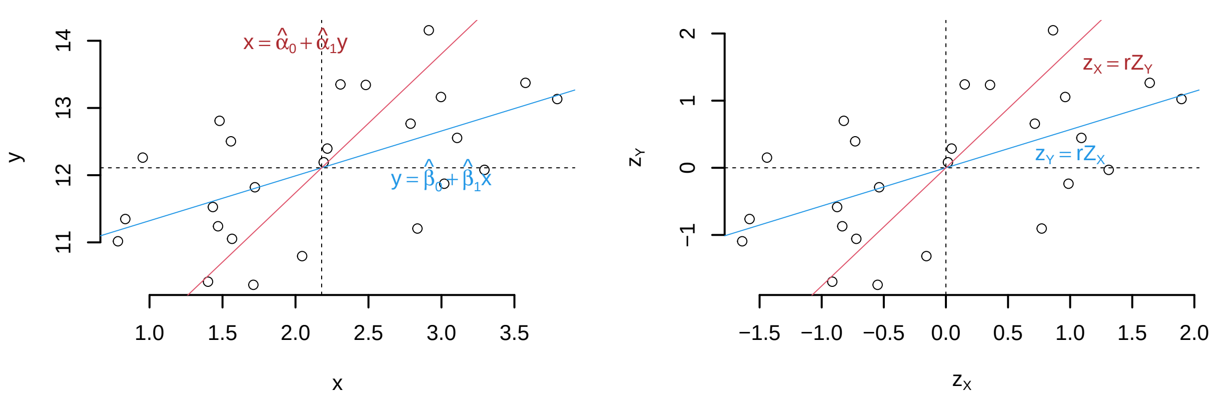
Si considerino i due modelli

$$z_{Yi} = \beta_{0Z} + \beta_{1Z} \cdot z_{Xi} + \varepsilon_{Zi}, \qquad z_{Xi} = \alpha_{0Z} + \alpha_{1Z} \cdot z_{Yi} + \delta_{Zi}$$

Allora

$$\begin{aligned} \hat\beta_{1Z} &= \frac{\mathrm{cov}(z_X, z_Y)}{\hat\sigma^2_{Z_X}} & \hat\alpha_{1Z} &= \frac{\mathrm{cov}(z_X, z_Y)}{\hat\sigma^2_{Z_Y}} \\ &= \frac{r}{1^2} = r & &= \frac{r}{1^2} = r \\ \hat\beta_{0Z} &= \bar z_Y - \hat\beta_{1Z}\bar z_X & \hat\alpha_{0Z} &= \bar z_X - \hat\alpha_{1Z}\bar z_Y \\ &= 0 + r \cdot 0 = 0 & &= 0 + r \cdot 0 = 0 \end{aligned}$$

Graficamente



Il Test Chi-Quadro 19

19.1 Test di Significatività pura

I test che abbiamo visto fin'ora prevedono l'ipotesi su un parametro θ. Ovvero tutti i test che abbiamo visto fin'ora sono del tipo

$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta \in \Theta_1 \end{cases}$$

I test di significatività pura rispondono a domande diverse, come per esempio testare se due variabili, alla luce dei dati, sono o non sono indipendenti o se la distribuzione osservata di una variabile è compatibile con un modello probabilistico. In un test di significatività pure solo l'ipotesi H_0 viene elicitata, mentre l'ipotesi H_1 non ha molto senso. Il test del χ^2 consente di risolvere diversi test di significatività pura

19.2 Associazione tra due variabili

Abbiamo visto nel modello di regressione le statistiche che misurano *l'associazione lineare* tra due variabili quantitative. Abbiamo osservato che l'associazione lineare **non** l'unico tipo di associazione osservabile tra x ed y. Il modello di regressione lineare funziona solo se x e y sono quantitative. Come misurare l'associazione tra x e y se sono entrambe categoriali?

19.2.1 Le tavole di contingenza

Sono tabelle che consentono di vedere le *frequenze congiunte* delle osservazioni a coppie

Esempio 19.2.1. 100 individui dalla città A, 100 dalla città B e 100 dalla città C. Abbiamo misurato X = Genere (M, F), Y = abbonamento allo stadio (Sì, No). In tutte e tre le città abbiamo osservato 50 M e 50 F, e abbiamo osservato 50 abbonati e 50 non abbonati.

	Città A		
	Abbonato	Non Abbonato	Tot
M	50	0	**50**
F	0	50	**50**
Tot	**50**	**50**	**100**

	Città B		
	Abbonato	Non Abbonato	Tot
M	35	15	**50**
F	15	35	**50**
Tot	**50**	**50**	**100**

	Città C		
	Abbonato	Non Abbonato	Tot
M	25	25	**50**
F	25	25	**50**
Tot	**50**	**50**	**100**

Nella città A solo i maschi hanno l'abbonamento allo stadio **perfetta** associazione tra Genere e Abbonamento. Nella città B molti maschi hanno l'abbonamento allo stadio ma non solo c'è associazione tra Genere e Abbonamento. Nella città C non c'è distinzione tra genere e passione per il pallone **Non** c'è associazione tra Genere e Abbonamento.

Ma, come vedremo nel prossimo esempio, non è sempre facile individuare dove non c'è associazione.

Esempio 19.2.2 (Popolazione finita, estrazioni SR (Efficienza))**.** 100 individui dalla città A, 100 dalla città B e 100 dalla città C. Abbiamo misurato X = Genere (M, F), Y = abbonamento allo stadio

(Sì, No). In tutte e tre le città abbiamo osservato 60 M e 40 F, e abbiamo osservato 60 abbonati e 40 non abbonati.

	Città A		
	Abbonato	Non Abbonato	Tot
M	60	0	**60**
F	0	40	**40**
Tot	**60**	**40**	**100**

	Città B		
	Abbonato	Non Abbonato	Tot
M	40	20	**60**
F	20	20	**50**
Tot	**60**	**40**	**100**

	Città C		
	Abbonato	Non Abbonato	Tot
M	36	24	**60**
F	24	16	**40**
Tot	**60**	**40**	**100**

Nella città A solo i maschi hanno l'abbonamento allo stadio **perfetta** associazione tra Genere e Abbonamento. Nella città B molti maschi hanno l'abbonamento allo stadio ma non solo,
c'è associazione tra Genere e Abbonamento. Nella città C non c'è distinzione tra genere e passione per il pallone **non** c'è associazione tra Genere e Abbonamento ma non è evidente da vedere.

19.2.2 Un passo indietro: il concetto di indipendenza

Se X e Y sono due VC **INDIPENDENTI** allora

$$P(X = x \ \cap \ Y = y) = P(X = x)P(Y = y)$$

Se per esempio

$$S_X = \{0, 1\}, \qquad S_Y = \{0, 1\}$$

e

$$P(X = 1) = \pi_X \qquad P(Y = 1) = \pi_y$$

Sotto regime di indipendenza

	$Y = 0$	$Y = 1$	
$X = 0$	$\pi_{11} = P(X = 0 \cap Y = 0) = (1 - \pi_X)(1 - \pi_Y)$	$\pi_{12} = P(X = 0 \cap Y = 1) = (1 - \pi_X)\pi_Y$	$1 - \pi_X$
$X = 1$	$\pi_{21} = P(X = 1 \cap Y = 0) = \pi_X(1 - \pi_Y)$	$\pi_{22} = P(X = 1 \cap Y = 1) = \pi_X\pi_Y$	π_X
	$1 - \pi_Y$	π_Y	1

Sia X la VC che registra testa nel lancio di una moneta perfetta e Y una VC che registra la faccia numero 6 dal lancio di un dado perfetto

e

$$P(X = 1) = \frac{1}{2} \qquad P(Y = 1) = \frac{1}{6}$$

Sotto regime di indipendenza

	$Y = 0$	$Y = 1$	
$X = 0$	0.4167	0.0833	$\frac{1}{2}$
$X = 1$	0.4167	0.0833	$\frac{1}{2}$
	$\frac{5}{6}$	$\frac{1}{6}$	1

L'indipendenza è imposta dal meccanismo di generazione di casualità Ovvero, se io lancio 100 volte il dado e la moneta mi aspetto, in media di osservare 41.7 volte $X = 0$ e $Y = 0$, 41.7 volte $X = 1$ e $Y = 0$, 8.3 volte $X = 0$ e $Y = 1$ e 8.3 volte $X = 1$ e $Y = 1$. Se dopo 100 lanci le frequenze osservate sono molto diverse dalle frequenze teoriche in regime di indipendenza, potrei dubitare dell'ipotesi di indipendenza

19.2.3 Estensione a più di due modalità

Se abbiamo due VC categoriali X e Y con supporto

$$S_X = \{x_1, ..., x_I\} \qquad S_Y = \{y_1, ..., y_J\}$$

definiamo le **probabilità congiunte**

$$\pi_{ij} = \text{la probabilità che } X = x_i \text{ e } Y = y_j$$

La distribuzione di probabilità doppia è

	y_1	y_2	...	y_j	...	y_J	
x_1	π_{11}	π_{12}	...	π_{1j}	...	π_{1J}	$\pi_{1\bullet}$
x_2	π_{21}	π_{22}	...	π_{2j}	...	π_{2J}	$\pi_{2\bullet}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$		$\vdots$	$\vdots$
x_i	π_{i1}	π_{i2}	...	π_{ij}	...	π_{iJ}	$\pi_{i\bullet}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\ddots$	$\vdots$	$\vdots$
x_I	π_{I1}	π_{I2}	...	π_{Ij}	...	π_{IJ}	$\pi_{I\bullet}$
	$\pi_{\bullet 1}$	$\pi_{\bullet 2}$	...	$\pi_{\bullet j}$	...	$\pi_{\bullet J}$	1

Le **probabilità marginali** sono

$$\pi_{i\bullet} = \pi_{i1} + ... + \pi_{iJ}, \qquad \pi_{\bullet j} = \pi_{1j} + ... + \pi_{Ij}$$

19.2.4 Esempio

Nella città A ci sono 27 individui occupati con la licenza media , 45 individui occupati con la licenza superiore, 18 individui occupati laureati, 18 individui disoccupati con la licenza media , 30 individui disoccupati con la licenza superiore e 12 laureati

	M	S	L	
Occupato	27	45	18	90
Disoccupato	18	30	12	60
	45	75	30	150

Qual è la probabilità di estrarre un soggetto che abbia un dato titolo di studio e una data condizione occupazionale?

	M	S	L	
Occupato	$\pi_{11} = \frac{27}{150} = 0.18$	$\pi_{12} = \frac{45}{150} = 0.3$	$\pi_{13} = \frac{18}{150} = 0.12$	$\pi_{1\bullet} = \frac{90}{150} = 0.6$
Disoccupato	$\pi_{21} = \frac{18}{150} = 0.12$	$\pi_{22} = \frac{30}{150} = 0.2$	$\pi_{23} = \frac{12}{150} = 0.08$	$\pi_{2\bullet} = \frac{60}{150} = 0.4$
	$\pi_{\bullet 1} = \frac{45}{150} = 0.3$	$\pi_{\bullet 2} = \frac{75}{150} = 0.5$	$\pi_{\bullet 3} = \frac{30}{150} = 0.2$	$\frac{150}{150} = 1.0$

Titolo di studio e occupazione sono **indipendenti**?

Sì, in quanto

$$\pi_{ij} = \pi_{i\bullet}\pi_{\bullet j}, \ \forall i, j$$

Ad esempio

$$\pi_{11} = 0.18 = 0.6 \times 0.3 = \pi_{1\bullet}\pi_{\bullet 1}$$

Ad esempio

$$\pi_{23} = 0.08 = 0.4 \times 0.2 = \pi_{2\bullet}\pi_{\bullet 3}$$

Nella città B osserviamo

	M	S	L	
Occupato	0	60	30	90
Disoccupato	45	15	0	60
	45	75	30	150

Qual è la probabilità di estrarre un oggetto che abbia un dato titolo di studio e una data condizione occupazionale?

	M	S	L	
Occupato	$\pi_{11} = \frac{0}{150} = 0$	$\pi_{12} = \frac{60}{150} = 0.4$	$\pi_{13} = \frac{30}{150} = 0.2$	$\pi_{1\bullet} = \frac{90}{150} = 0.6$
Disoccupato	$\pi_{21} = \frac{45}{150} = 0.3$	$\pi_{22} = \frac{15}{150} = 0.1$	$\pi_{23} = \frac{0}{150} = 0$	$\pi_{2\bullet} = \frac{60}{150} = 0.4$
	$\pi_{\bullet 1} = \frac{45}{150} = 0.3$	$\pi_{\bullet 2} = \frac{75}{150} = 0.5$	$\pi_{\bullet 3} = \frac{30}{150} = 0.2$	$\frac{150}{150} = 1.0$

Titolo di studio e occupazione sono **indipendenti**?

No, in quanto

$$\pi_{ij} \neq \pi_{i\bullet}\pi_{\bullet j}$$

Ad esempio

$$\pi_{11} = 0 \neq 0.6 \times 0.3 = \pi_{1\bullet}\pi_{\bullet 1}$$

Ad esempio

$$\pi_{23} = 0 \neq 0.4 \times 0.2 = \pi_{2\bullet}\pi_{\bullet 3}$$

19.2.5 Dalla popolazione al campione

Abbiamo usato le due città come se fossero l'intera popolazione. Non sempre possiamo valutare l'intera popolazione. Spesso abbiamo l'osservazione di n coppie di VC categoriali estratte dalla popolazione di riferimento. Da una popolazione dove X ed Y sono indipendenti, ci aspettiamo che n loro realizzazioni fotografino l'indipendenza che c'è in popolazione.

19.2.6 Notazione formale per le tavole di contingenza

Se abbiamo due variabili categoriali X e Y con supporto

$$S_X = \{x_1, ..., x_k\} \qquad S_Y = \{y_1, ..., y_J\}$$

definiamo le **frequenze assolute congiunte**

$$n_{ij} = \text{il numero di volte che } X = x_i \text{ e } Y = y_j$$

La tabella di contingenza si presenta

$$\begin{array}{c|cccccc|c}
 & y_1 & y_2 & ... & y_j & ... & y_J & \\
\hline
x_1 & n_{11} & n_{12} & ... & n_{1j} & ... & n_{1J} & n_{1\bullet} \\
x_2 & n_{21} & n_{22} & ... & n_{2j} & ... & n_{2J} & n_{2\bullet} \\
\vdots & \vdots & \vdots & \ddots & \vdots & & \vdots & \vdots \\
x_i & n_{i1} & n_{i2} & ... & n_{ij} & ... & n_{iJ} & n_{i\bullet} \\
\vdots & \vdots & \vdots & & \vdots & \ddots & \vdots & \vdots \\
x_I & n_{I1} & n_{I2} & ... & n_{Ij} & ... & n_{IJ} & n_{I\bullet} \\
\hline
 & n_{\bullet 1} & n_{\bullet 2} & ... & n_{\bullet j} & ... & n_{\bullet J} & n
\end{array}$$

Le **frequenze assolute marginali** sono

$$n_{i\bullet} = n_{i1} + ... + n_{iJ}$$

$$n_{\bullet j} = n_{1j} + ... + n_{Ij}$$

Definiamo le **frequenze relative congiunte**

$$\hat{\pi}_{ij} = \frac{n_{ij}}{n}$$

$$\begin{array}{c|cccccc|c}
 & y_1 & y_2 & ... & y_j & ... & y_J & \\
\hline
x_1 & \hat{\pi}_{11} = \frac{n_{11}}{n} & \hat{\pi}_{12} = \frac{n_{12}}{n} & ... & \hat{\pi}_{1j} = \frac{n_{1j}}{n} & ... & \hat{\pi}_{1J} = \frac{n_{1J}}{n} & \hat{\pi}_{1\bullet} = \frac{n_{1\bullet}}{n} \\
x_2 & \hat{\pi}_{21} = \frac{n_{21}}{n} & \hat{\pi}_{22} = \frac{n_{22}}{n} & ... & \hat{\pi}_{2j} = \frac{n_{2j}}{n} & ... & \hat{\pi}_{2J} = \frac{n_{2J}}{n} & \hat{\pi}_{2\bullet} = \frac{n_{2\bullet}}{n} \\
\vdots & \vdots & \vdots & \ddots & \vdots & & \vdots & \vdots \\
x_i & \hat{\pi}_{i1} = \frac{n_{i1}}{n} & \hat{\pi}_{i2} = \frac{n_{i2}}{n} & ... & \hat{\pi}_{ij} = \frac{n_{ij}}{n} & ... & \hat{\pi}_{iJ} = \frac{n_{iJ}}{n} & \hat{\pi}_{i\bullet} = \frac{n_{i\bullet}}{n} \\
\vdots & \vdots & \vdots & & \vdots & \ddots & \vdots & \vdots \\
x_I & \hat{\pi}_{I1} = \frac{n_{I1}}{n} & \hat{\pi}_{I2} = \frac{n_{I2}}{n} & ... & \hat{\pi}_{Ij} = \frac{n_{Ij}}{n} & ... & \hat{\pi}_{IJ} = \frac{n_{IJ}}{n} & \hat{\pi}_{I\bullet} = \frac{n_{I\bullet}}{n} \\
\hline
 & \hat{\pi}_{\bullet 1} = \frac{n_{\bullet 1}}{n} & \hat{\pi}_{\bullet 2} = \frac{n_{\bullet 2}}{n} & ... & \hat{\pi}_{\bullet j} = \frac{n_{\bullet j}}{n} & ... & \hat{\pi}_{\bullet J} = \frac{n_{\bullet J}}{n} & 1
\end{array}$$

Le **frequenze relative marginali** sono

$$\hat{\pi}_{i\bullet} = \hat{\pi}_{i1} + ... + \hat{\pi}_{iJ}$$

$$\hat{\pi}_{\bullet j} = \hat{\pi}_{1j} + ... + \hat{\pi}_{Ij}$$

19.2.7 Le frequenze sono stime dei π

$$\hat{\pi}_{ij} = \frac{n_{ij}}{n}, \qquad \hat{\pi}_{i\bullet} = \frac{n_{i\bullet}}{n}, \qquad \hat{\pi}_{\bullet j} = \frac{n_{\bullet j}}{n}$$

$\hat{\pi}_{ij}$ è stima di π_{ij}, $\forall i, j$

$\hat{\pi}_{i\bullet}$ è stima di $\pi_{i\bullet}$, $\forall i$

$\hat{\pi}_{\bullet j}$ è stima di $\pi_{\bullet j}$, $\forall j$

Se in popolazione x ed y sono indipendenti, allora

$$\pi_{ij} = \pi_{i\bullet}\pi_{\bullet j}$$

E quindi, mi aspetterei

$$\hat{\pi}_{ij} \approx \hat{\pi}_{i\bullet}\hat{\pi}_{\bullet j}$$

Ovvero

$$\begin{aligned} n\hat{\pi}_{ij} &\approx n\ \hat{\pi}_{i\bullet}\hat{\pi}_{\bullet j} \\ n_{ij} &\approx n\frac{n_{i\bullet}}{n}\frac{n_{\bullet j}}{n} \\ &\approx \frac{n_{i\bullet}n_{\bullet j}}{n} \\ &\approx n^*_{ij} \end{aligned}$$

n^*_{ij} è la frequenza assolute attese in caso di indipendenza. La domanda è: quanto si discostano le frequenze assolute osservate n_{ij} dalle frequenze attese n^*_{ij}?

19.2.8 Esempio (continua)

Ripartiamo dall'esempio di partenza, immaginando che nelle tre città quei 100 individui siano un campione

	Città A		
	Abbonato	Non Abbonato	Tot
M	60	0	**60**
F	0	40	**40**
Tot	**60**	**40**	**100**

	Città B		
	Abbonato	Non Abbonato	Tot
M	40	20	**60**
F	20	20	**50**
Tot	**60**	**40**	**100**

	Città C		
	Abbonato	Non Abbonato	Tot
M	36	24	**60**
F	24	16	**40**
Tot	**60**	**40**	**100**

Costruiamo le n^*_{ij}

	Freq. th		
	Abbonato	Non Abbonato	Tot
M	$\frac{60\cdot 60}{100} = 36$	$\frac{60\cdot 40}{100} = 24$	**60**
F	$\frac{40\cdot 60}{100} = 24$	$\frac{40\cdot 40}{100} = 16$	**40**
Tot	**60**	**40**	**100**

Nella città C le frequenza osservate coincidono con quelle attese, **non** c'è associazione tra genere e abbonamento allo stadio. Quanto si discostano le città A e B dalla situazione di indipendenza?

19.3 L'indice χ^2

Il χ^2_{obs}, che abbiamo già incontrato, è utilizzato qui per nominare un misura di distanza

$$\chi^2_{\text{obs}} = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{\left(n_{ij} - n^*_{ij}\right)^2}{n^*_{ij}}$$

	Città A		
	Abbonato	Non Abbonato	Tot
M	$\frac{(60-36)^2}{36} = 36$	$\frac{(0-24)^2}{24} = 24$	40
F	$\frac{(0-24)^2}{24} = 24$	$\frac{(40-16)^2}{16} = 36$	60
Tot	40	60	$\chi^2_{\text{obs},A} = 100$

	Città B		
	Abbonato	Non Abbonato	Tot
M	$\frac{(40-36)^2}{36} = 0.4444$	$\frac{(20-24)^2}{24} = 0.6667$	1.1111
F	$\frac{(20-24)^2}{24} = 0.6667$	$\frac{(20-16)^2}{16} = 1$	1.6667
Tot	1.1111	1.6667	$\chi^2_{\text{obs},B} = 2.7778$

Ovviamente per la città C:

	Città C		
	Abbonato	Non Abbonato	Tot
M	$\frac{(36-36)^2}{36} = 0$	$\frac{(24-24)^2}{24} = 0$	0
F	$\frac{(24-24)^2}{24} = 0$	$\frac{(16-16)^2}{16} = 0$	0
Tot	0	0	$\chi^2_{\text{obs},C} = 0$

Nella città A c'è una evidente associazione tra genere e abbonamento. Nella città C c'è evidente indipendenza tra genere e abbonamento. Nella città B lo scostamento dalla situazione di indipendenza, è dovuto al caso oppure effettivamente nella città B c'è associazione?

19.4 Test per l'ipotesi di indipendenza

Chiedersi se x e y sono indipendenti significa fare un test in cui è facile scrivere H_0 ma è **superfluo scrivere** H_1. Se H_0 prescrive l'indipendenza, allora

$$H_0 : \pi_{ij} = \pi_{i\bullet}\pi_{\bullet j}, \forall i, j$$

H_1 non la scriviamo perché non può altro che essere la negazione di H_0. Questi sono *Test di significatività pura.*

19.4.1 La statistica test χ^2

Se H_0 prescrive l'indipendenza, allora

$$H_0 : \pi_{ij} = \pi_{i\bullet}\pi_{\bullet j}, \forall i, j$$

Si dimostra che, sotto H_0, prima di osservare i dati, l'indicatore χ^2 vista come VC è:

$$\chi^2_{VC} \sim \chi^2_{gdl}$$

I gradi di libertà sono

$$gdl = (I - 1) \times (J - 1)$$

Si fissa α e si cerca

$$\chi^2_{gdl;\alpha} : P(\chi^2_{gdl} > \chi^2_{gdl;\alpha}) = \alpha$$

Esempio $\alpha = 0.05$, nell'esempio genere/abbonamento i gdl sono

$$gdl = (2-1) \times (2-1) = 1$$

$$\chi^2_{1;0.05} = 3.8415$$

$$\begin{array}{ccc} \chi^2_{\text{obs};A} = 100 > 3.8415; & \chi^2_{\text{obs};B} = 2.7778 < 3.8415; & \chi^2_{\text{obs};C} = 0 < 3.8415 \\ \text{Rifiuto } H_0 & \text{non Rifiuto} H_0 & \text{non Rifiuto} H_0 \end{array}$$

19.4.2 Esempio

In una città sono stati estratti 20 individui occupati con la licenza media , 50 individui occupati con la licenza superiore, 20 individui occupati laureati, 20 individui disoccupati con la licenza media , 25 individui disoccupati con la licenza superiore e 10 laureati

	M	S	L	
Occupato	20	50	20	90
Disoccupato	25	25	10	60
	45	75	30	150

Vogliamo testare

$$H_0 : \pi_{ij} = \pi_{i\bullet}\pi_{\bullet j}$$

Per vari livelli di α Costruiamo i valori attesi sotto H_0

$$n^*_{ij} = \frac{n_{i\bullet} n_{\bullet j}}{n}$$

e quindi

Osservati	M	S	L	
Occupato	20	50	20	90
Disoccupato	25	25	10	60
	45	75	30	150

Teorici	M	S	L	
Occupato	27	45	18	90
Disoccupato	18	30	12	60
	45	75	30	150

Costruiamo le distanze

$$\frac{\left(n_{ij} - n^*_{ij}\right)^2}{n^*_{ij}}$$

e quindi

	M	*S*	*L*	
Occupato	1.8148	0.5556	0.2222	2.5926
Disoccupato	2.7222	0.8333	0.3333	3.8889
	4.537	1.3889	0.5556	6.4815

$gdl = (2-1) \times (3-1) = 2$

Dalle tavole

$$\chi^2_{2;0.05} = 5.9915, \qquad \chi^2_{2;0.01} = 9.2103$$

E osserviamo che $\chi^2_{\text{obs}} = 6.4815 > \chi^2_{2;0.05} = 5.9915$ e quindi rifiuto H_0 al 5% e che $\chi^2_{\text{obs}} = 6.4815 < \chi^2_{2;0.01} = 9.2103$ e quindi non rifiuto H_0 al 1%.

19.4.3 I gradi di libertà

Esempio 19.4.1 (Tabella 2×2)**.** Supponiamo di voler riempire la tabella di contingenza coi totali fissati

	Abbonato	*Non Abbonato*	
M	n_{11}	n_{12}	60
F	n_{21}	n_{22}	40
	60	40	100

abbiamo $gdl = (2-1) \times (2-1) = 1$. Ho molte scelte per il valore n_{11}, per esempio

$$n_{11} = 35$$

ho usato il mio unico grado di libertà, tutti gli altri n_{ij} sono vincolati

	Abbonato	*Non Abbonato*	
M	35	25	60
F	25	15	40
	60	40	100

Esempio 19.4.2 (Tabella 2×3)**.** Se la tabella è

	M	*S*	*L*	
Occupato	n_{11}	n_{12}	n_{13}	90
Disoccupato	n_{21}	n_{22}	n_{23}	60
	45	75	30	150

abbiamo $gdl = (2-1) \times (3-1) = 2$

Ho molte scelte per il valore n_{11}, per esempio

$$n_{11} = 30$$

ho usato il primo grado di libertà, $n_{21} = 15$ è obbligato

	M	S	L	
Occupato	30	n_{12}	n_{13}	90
Disoccupato	15	n_{22}	n_{23}	60
	45	75	30	150

Fisso arbitrariamente n_{12}, per esempio

$$n_{12} = 40$$

e ho usato il secondo grado di libertà, tutti gli altri valori sono vincolati

	M	S	L	
Occupato	30	40	20	90
Disoccupato	15	35	10	60
	45	75	30	150

19.5 Misure di Conformità

Spesso, abbiamo assunto che $X \sim \mathscr{L}(\theta)$, senza questionare la scelta di $\mathscr{L}$. Ad esempio quando assumiamo $X \sim \text{Pois}(\lambda)$, sappiamo fare inferenza su λ, ma tutti i risultati sono validi se l'assunzione di partenza è valida. Oppure quando assumiamo $X \sim N(\mu, \sigma^2)$, sappiamo fare inferenza su μ, σ^2, ma tutti i risultati sono validi se l'assunzione di partenza è valida.

IL χ^2 misura la *conformità* dei dati al modello scelto. Supponiamo, le X_i, $i = 1, ..., n$ VC categoriale con supporto

$$S_X = \{x_1, ..., x_K\}$$

La $\mathscr{L}(\pi_1, ..., \pi_K)$ tale che

$$P(X_i = x_j) = \pi_j, \quad i = 1, ..., n \quad j = 1, ..., K$$

Se estraggo un campione da IID da $\mathscr{L}(\pi_1, ..., \pi_K)$, ovvero estraggo n volte X IID ottengo:

modalità	freq. assoluta	freq. relativa
x_1	n_2	$\hat{\pi}_1 = n_1/n$
x_2	n_2	$\hat{\pi}_2 = n_2/n$
$\vdots$	$\vdots$	$\vdots$
x_K	n_K	$\hat{\pi}_K = n_K/n$

Le $\hat{\pi}_j$ sono stime delle π_j. Se $\mathscr{L}(\pi_1, ..., \pi_K)$ è vera mi aspetto

$$\hat{\pi}_j \approx \pi_j$$

Ovvero

$$n \cdot \hat{\pi}_j \quad \approx \quad n \cdot \pi_j$$

$$n_j \quad \approx \quad n_j^*$$

Dove le n_j^* sono le frequenze attese

Esempio 19.5.1. Lanciamo un dado 100 volte e otteniamo

	1	2	3	4	5	6	tot.
n_j	16	15	18	20	14	17	100
$\hat{\pi}_j$	0.16	0.15	0.18	0.20	0.14	0.17	1.00

Se X è un dado perfetto, allora

$$P(X_i = j) = \frac{1}{6}, \ j = 1, ..., 6$$

Da un dado perfetto mi aspetterei, in media

$$\begin{aligned} \hat{\pi}_j &\approx \frac{1}{6} & n \cdot \hat{\pi}_j &\approx n \cdot \frac{1}{6} \\ &= 0.1667 & n_j &\approx 100\frac{1}{6} \\ & & &\approx 16.67 \end{aligned}$$

E quindi

	1	2	3	4	5	6	tot.
n_j	16	15	18	20	14	17	100
$\hat{\pi}_j$	0.16	0.15	0.18	0.20	0.14	0.17	1.00
$\hat{\pi}_j^*$	0.1667	0.1667	0.1667	0.1667	0.1667	0.1667	1.0000
n_j^*	16.667	16.667	16.667	16.667	16.667	16.667	100

Esempio 19.5.2. Faccio lanciare il dado a un'altra persona e osservo su 100 lanci

	1	2	3	4	5	6	tot.
n_j	1	2	2	5	40	50	100
$\hat{\pi}_j$	0.01	0.02	0.02	0.05	0.40	0.50	1.00
$\hat{\pi}_j^*$	0.1667	0.1667	0.1667	0.1667	0.1667	0.1667	1.0000
n_j^*	16.667	16.667	16.667	16.667	16.667	16.667	100

Che è molto diversa da quella di prima

	1	2	3	4	5	6	tot.
n_j	16	15	18	20	14	17	100
$\hat{\pi}_j$	0.16	0.15	0.18	0.20	0.14	0.17	1.00
$\hat{\pi}_j^*$	0.1667	0.1667	0.1667	0.1667	0.1667	0.1667	1.0000
n_j^*	16.667	16.667	16.667	16.667	16.667	16.667	100

19.6 Il χ^2 come misura di conformità

Misuriamo la distanza tra le frequenze attese con quelle osservate con la misura del chi-quadro

$$\chi^2_{\text{obs}} = \sum_{j=1}^{K} \frac{\left(n_{ij} - n^*_{ij}\right)^2}{n^*_{ij}}$$

Nel primo caso

	1	2	3	4	5	6	tot.
n_j	16	15	18	20	14	17	100
n^*_j	16.667	16.667	16.667	16.667	16.667	16.667	100
$\frac{(n_j - n^*_j)^2}{n^*_j}$	0.027	0.167	0.107	0.667	0.427	0.007	1.402

$$\chi^2_{\text{obs}} = 1.402$$

Nel secondo caso

	1	2	3	4	5	6	tot.
n_j	1	2	2	5	40	50	100
n^*_j	16.667	16.667	16.667	16.667	16.667	16.667	100
$\frac{(n_j - n^*_j)^2}{n^*_j}$	14.727	12.907	12.907	8.167	32.667	66.667	148.042

$$\chi^2_{\text{obs}} = 148.042$$

Ci chiediamo se il modello che stiamo usando sia supportato dai dati

$$H_0 : X \sim \mathscr{L}$$

Se X è discreta

$$H_0 : \pi_j = \pi_{j0}, \forall j$$

Sotto H_0, prima di osservare i dati, l'indice di conformità χ^2_{VC}, visto come VC è distribuito come

$$\chi^2_{VC} \sim \chi^2_{K-1}$$

Osservo i dati e calcolo

$$\chi^2_{\text{obs}} = \sum_{j=1}^{K} \frac{\left(n_{ij} - n^*_{ij}\right)^2}{n^*_{ij}}$$

Scelgo α, trovo $\chi^2_{K-1;\alpha}$ sulle tavole del chi-quadro e lo confronto con χ^2_{obs}

- se $\chi^2_{\text{obs}} > \chi^2_{K-1;\alpha}$ **rifiuto** H_0 al lds α
- se $\chi^2_{\text{obs}} < \chi^2_{K-1;\alpha}$ **non rifiuto** H_0 al lds α

19.6.1 Esempio: Scostamento da una uniforme

Lanciamo un dado 60 volte

	1	2	3	4	5	6	tot.
n_j	5	14	6	17	13	5	60
n_j^*	10	10	10	10	10	10	60
$\frac{(n_j-n_j^*)^2}{n_j^*}$	2.5	1.6	1.6	4.9	0.9	2.5	14.0

$$H_0 : \pi_j = \frac{1}{6}, j = 1, ..., 6$$

$$\chi^2_{\text{obs}} = 14$$

Se $\alpha = 0.05$, $\chi^2_{6-1;0.05} = 11.0705$, allora $\chi^2_{\text{obs}} = 14 > 11.0705 = \chi^2_{6-1;0.05}$ **rifiuto** H_0 al 5% Se $\alpha = 0.01$, $\chi^2_{6-1;0.01} = 15.0863$, allora $\chi^2_{\text{obs}} = 14 < 15.0863 = \chi^2_{6-1;0.01}$ **non rifiuto** H_0 al 1%

19.6.2 Esempio: Scostamento da una popolazione

L'hotel A ha a servizio 50 addetti: 15 addetti alle pulizie, 9 receptionist, 24 personale di sala e 2 cuochi. L'hotel fa parte di una catena alberghiera. Al controllo di gestione del personale il gruppo registra che nei loro hotel il 30% sono addetti alle pulizie, il 30% sono receptionist, il 30% personale di sala e il 10% cuochi. L'hotel A ha una gestione del personale conforme a quella del gruppo, al lds del 5%?

$$H_0 : \pi_{jA} = \pi_{jG}$$

Costruiamo le tabelle

	A	R	S	C	tot.
n_j	15	9	24	2	50
π_j^*	0.3	0.3	0.3	0.1	1.0
$n_j^* = n\pi_j^*$	15	15	15	5	50
$\frac{(n_j-n_j^*)^2}{n_j^*}$	0.0	2.4	5.4	1.8	9.6

Se $\alpha = 0.05$, $\chi^2_{4-1;0.05} = 7.8147$, allora $\chi^2_{\text{obs}} = 9.6 > 7.8147 = \chi^2_{4-1;0.05}$ **rifiuto** H_0 al 5%

19.6.3 Esempio: scostamento da una Poisson

Si sono registrati i clienti in coda presso la cassa di un negozio in 200 occasioni scelte a caso in vari momenti del giorno per un mese e si sono ottenuti i risultati riportati in tabella.

Clienti in coda	0	1	2	>2
volte	70	85	30	15

Si presume che i clienti in coda presso la cassa del negozio sia una Poisson(1). Verificare, al livello di significatività del 5%, se i dati osservati sono coerenti con l'ipotesi.

$$H_0 : X \sim \text{Pois}(1)$$

Sotto H_0

$$\pi_j^* = P(X_i = x_j) = \frac{\lambda}{x_j!} e^{-\lambda},$$

$$\begin{aligned}
\pi_1^* = P(X_i = 0) &= \frac{1}{0!} e^{-1} = 0.3679 \\
\pi_2^* = P(X_i = 1) &= \frac{1}{1!} e^{-1} = 0.3679 \\
\pi_3^* = P(X_i = 2) &= \frac{1}{2!} e^{-1} = 0.1839 \\
\pi_4^* = P(X_i > 2) &= 1 - (\pi_1^* + \pi_2^* + \pi_3^*) = 0.0803
\end{aligned}$$

Le frequenze attese

$$n_j^* = n\pi_j^*$$

Clienti in coda	0	1	2	>2	Tot
volte	70	85	30	15	200
π_j^*	0.3679	0.3679	0.1839	0.0803	1.0000
$n_j^* = n\pi_j^*$	73.58	73.58	36.79	16.06	200.00
$\frac{(n_j - n_j^*)^2}{n_j^*}$	0.1738	1.7738	1.2525	0.0700	3.2701

$\alpha = 0.05$, quindi $\chi^2_{4-1;0.05} = 7.8147$

allora $\chi^2_{\text{obs}} = 3.2701 < 7.8147 = \chi^2_{4-1;0.05}$ **non rifiuto** H_0 al 5%.

Parte IV
Appendice

Richiami sugli Operatori Sommatoria e Produttorio

A

A.1 Operatore Sommatoria

È una forma simbolica per rappresentare somme di un numero qualunque di addendi. Si consideri un insieme di numeri indicizzati con i

$$\{a_1, ..., a_n\}$$

Si definisce la *Sommatoria per* i *che va da 1 fino ad* n

$$\sum_{i=1}^{n} a_i = a_1 + a_2 + ... + a_n$$

i ed n sono chiamati *quantificatori*.

Esempio

Si consideri l'insieme

$$S = \{a_1 = 30, a_2 = 15, a_3 = 21\}$$

allora la *Sommatoria per* i *che va da 1 fino ad 3*

$$\sum_{i=1}^{3} a_i = 30 + 15 + 21 = 66$$

Si consideri l'insieme

$$S = \{x_1 = 3.1, x_2 = 4.1, x_3 = 1.4, x_4 = 3.3, x_5 = 2.9\}$$

allora la *Sommatoria per* a *che varia in* S

$$\sum_{i=3}^{5} x_3 + x_4 + x_5 = x_3 = 1.4 + 3.3 + 2.9 = 7.6$$

Un modo alternativo per indicare i quantificatori è il seguente. Sia S un insieme di numeri

$$S = \{a_1, ..., a_n\}$$

Si definisce la *Sommatoria di tutti gli* a *in* S

$$\sum_{a \in S} a = a_1 + a_2 + ... + a_n$$

Si consideri l'insieme

$$S = \{y_1 = 3.0, y_2 = 1.5, y_3 = 2.1\}$$

allora la *Sommatoria per a che varia in S*

$$\sum_{a \in S} a = 3.0 + 1.5 + 2.1 = 6.6$$

Proprietà della Sommatoria

1. Se k è una costatante, allora
$$\sum_{1=1}^{n} kx_i = k \sum_{1=1}^{n} x_i$$

Infatti

$$\begin{aligned}\sum_{1=1}^{n} kx_i &= k \cdot x_1 + k \cdot x_n \\ &= k(x_1 + ... + x_n) \\ &= k \sum_{i=1}^{n} x_i\end{aligned}$$

Si consideri l'insieme

$$S = \{x_1 = 3.1, x_2 = 4.1, x_3 = 1.4, x_4 = 3.3, x_5 = 2.9\},$$

Posto $k = 3.6$

$$\begin{aligned}\sum_{1=1}^{n} kx_i &= 3.6 \times 3.1 + 3.6 \times 4.1 + 3.6 \times 1.4 + 3.6 \times 3.3 + 3.6 \times 2.9 \\ &= 3.6 \times (3.1 + 4.1 + 1.4 + 3.3 + 2.9) \\ &= 53.28\end{aligned}$$

2. Se consideriamo due insiemi di numeri $\{a_1, ..., a_n\}$ e $\{b_1, ..., b_n\}$, allora

$$\sum_{1=1}^{n} (a_i + b_i) = \sum_{1=1}^{n} a_i + \sum_{1=1}^{n} b_i$$

Posto $S_X = \{x_1 = 3.1, x_2 = 4.1, x_3 = 1.4\}$, $S_Y = \{y_1 = 1.9, y_2 = 6.3, y_3 = 5.1\}$

$$\begin{aligned}\sum_{1=1}^{n} kx_i &= 3.6 \times 3.1 + 3.6 \times 4.1 + 3.6 \times 1.4 + 3.6 \times 3.3 + 3.6 \times 2.9 \\ &= (3.1 + 4.1 + 1.4) + (1.9 + 6.3 + 5.1) \\ &= 21.9\end{aligned}$$

3. Se k è una costante, allora

$$\sum_{1=1}^{n} k = k + k + ... + k = n \cdot k$$

Posto $k = 3.6$ e $n = 4$, allora

$$\begin{aligned} \sum_{1=1}^{n} k &= \sum_{1=1}^{4} 3.6 \\ &= 3.6 + 3.6 + 3.6 + 3.6 \\ &= 4 \times 3.6 \\ &= 14.4 \end{aligned}$$

4. Se k e c sono due costanti, allora

$$\sum_{1=1}^{n} (c + ka_i) = n \cdot c + k \sum_{1=1}^{n} a_i$$

Posto $S_X = \{x_1 = 3.1, x_2 = 4.1, x_3 = 1.4\}$, $k = 3.6$ e $c = 0.5$, allora

$$\begin{aligned} \sum_{1=1}^{n} c + kx_i &= \sum_{1=1}^{3} 0.5 + 3.6 \times x_i \\ &= 0.5 + 3.6 \times 3.1 + 0.5 + 3.6 \times 4.1 + 0.5 + 3.6 \times 1.4 \\ &= 4 \times 0.5 + 3.6 \times (3.1 + 4.1 + 1.4) \\ &= 32.46 \end{aligned}$$

Attenzione!

$$\sum_{1=1}^{n} (a_i \cdot b_i) \neq \left(\sum_{1=1}^{n} a_i \right) \cdot \left(\sum_{1=1}^{n} b_i \right)$$

Posto $S_X = \{x_1 = 3.1, x_2 = 4.1, x_3 = 1.4\}$, $S_Y = \{y_1 = 1.9, y_2 = 6.3, y_3 = 5.1\}$

$$\begin{aligned} \sum_{1=1}^{n} x_i &= 8.6 \\ \sum_{1=1}^{n} y_i &= 13.3 \end{aligned}$$

$$\begin{aligned}\sum_{1=1}^{n} x_i y_i &= (3.1 \times 1.9) + (4.1 \times 6.3) + (1.4 \times 5.1) \\ &= 38.86 \\ &\neq 8.6 \times 13.3 = 114.38\end{aligned}$$

A.2 Operatore Produttorio

Siano $a_1, ..., a_n$, n numeri, $a_i \in \mathbb{R}$: L'operatore sommatoria somma gli elementi

$$\sum_{i=1}^{n} a_i = a_1 + a_2 + ... + a_n$$

, allo stesso modo, l'operatore *produttoria* opera il prodotto dei dati

Definizione A.2.1 (Produttoria)**.** L'operatore produttoria moltiplica gli elementi

$$\prod_{i=1}^{n} a_i = a_1 \cdot a_2 \cdot ... \cdot a_n$$

Esempio: $a_1 = 1.1$, $a_2 = 0.9$, $a_3 = 1.3$

$$\begin{aligned}\sum_{i=1}^{n} a_i &= 1.1 + 0.9 + 1.3 = 3.3 \\ \prod_{i=1}^{n} a_i &= 1.1 \times 0.9 \times 1.3 = 1.287\end{aligned}$$

Richiami di Calcolo Combinatorio B

Il **calcolo combinatorio** ha come scopo principale **il conteggio del numero di modi in cui possono verificarsi determinati eventi**, senza doverli enumerare uno per uno. Questo permette di risolvere problemi legati alla disposizione, alla scelta e alla distribuzione di oggetti in modo efficiente e rigoroso.

Un modello efficace per affrontare queste situazioni è quello delle n **scatole numerate** e delle k **palline**, che permette di classificare i problemi combinatori in base a due aspetti fondamentali:
1. Le **palline** possono essere numerate (distinguibili) o non numerate (indistinguibili).
2. Il **numero di scatole rispetto alle palline** determina se tutte possono essere sistemate senza vincoli o se vi è un limite.

A partire da questa impostazione, si sviluppano le formule fondamentali della combinatoria, che trovano applicazioni in statistica, probabilità e molte altre discipline.

Quando ogni scatola può contenere al massimo una pallina, si individuano quattro casi principali, ciascuno con una specifica formula combinatoria.

B.1 Scelte indipendenti con ripetizione: k^n

Si consideri il caso in cui abbiamo k palline numerate e **ogni pallina può essere collocata indipendentemente in una qualsiasi delle n scatole disponibili**, senza restrizioni sul numero di volte in cui una scatola può essere scelta.

Il numero totale di distribuzioni è dato da:

$$n^k$$

che rappresenta il numero di **sequenze ordinate** di n elementi scelti tra k possibilità.

Esempio B.1.1 (Valigetta)**.** Un sistema di sicurezza utilizza un codice a tre cifre, in cui ciascun numero può variare da 1 a 9. Il codice è quindi una sequenza ordinata di tre elementi, scelti tra 9 possibilità:

$$9^3 = 729$$

Esistono 729 possibili codici di apertura.

B.2 $n = k$ Palline numerate: permutazioni

Quando il numero di palline è uguale al numero di scatole e ogni pallina è distinguibile, il problema diventa quello di **ordinare n elementi distinti**. Il numero di modi in cui ciò è possibile è dato dal **fattoriale**.

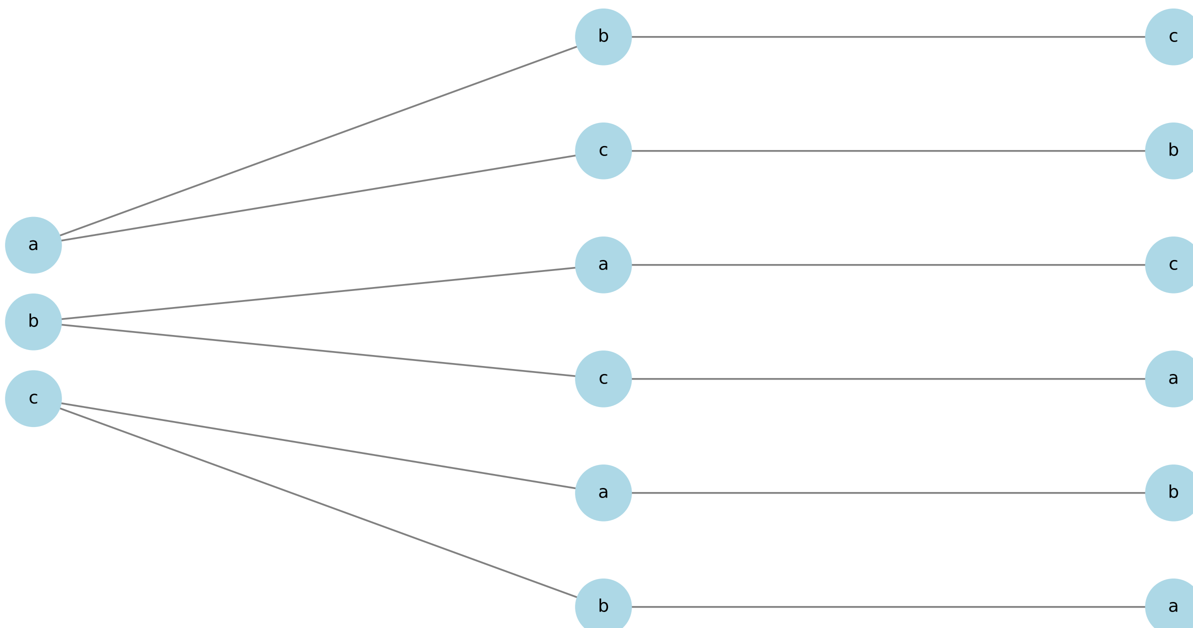
Sia $n \in \mathbb{N}$ un numero naturale, si definisce n fattoriale, il numero

$$n! = n(n-1)(n-2)...3 \cdot 2 \cdot 1$$

conta in quanti modo posso rimescolare n oggetti.

Nota. Per definizione $0! = 1$

In figura 3!.



Esempio B.2.1. $3! = 3 \cdot 2 \cdot 1 = 6$, $10! = 10 \cdot 9 \cdot ... \cdot 1 = 3628800$

Esempio B.2.2 (Mazzo di Carte)**.** Consideriamo un mazzo di 52 carte tutte distinte. Il numero di modi in cui il mazzo può essere mescolato corrisponde a tutte le possibili permutazioni delle 52 carte:

$$52! = 80658175170944942408940349866698506766127860028660283290685487972352$$

che è un numero estremamente grande, pari a circa 8.0658×10^{67}. Questo valore evidenzia come, ogni volta che un mazzo di carte viene mescolato, sia altamente improbabile che l'ordine ottenuto sia mai stato realizzato prima.

B.3 $n > k$ Palline numerate: disposizioni senza ripetizione

Se n è maggiore di k e le palline sono numerate, il problema diventa quello di scegliere k elementi distinti da un insieme di n e disporli in un ordine specifico. In questo caso, la formula è:

$$\frac{n!}{(n-k)!}$$

che rappresenta il numero di **sequenze ordinate** di k elementi scelti da un insieme di n elementi distinti.

Esempio B.3.1. Se si estrae una mano di 5 carte da un mazzo di 52 e si considera l'ordine di estrazione, il numero di possibili sequenze è:

$$\frac{52!}{(52-5)!} = \frac{52!}{47!}$$

che fornisce 311 875 200 possibili sequenze.

B.4 $n > k$ Palline non numerate: coefficienti binomiale

Se $n > k$ e le palline sono **non numerate**, il problema si riduce alla **scelta di k scatole tra n disponibili**, senza considerare l'ordine. Questo corrisponde al numero di **combinazioni** di k elementi scelti tra n, dato dal coefficiente binomiale:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Esempio B.4.1 (Sei al super enalotto)**.** Nel Superenalotto, si scelgono 6 numeri tra 90, senza che l'ordine abbia rilevanza. Il numero totale di combinazioni possibili è:

$$\binom{90}{6} = \frac{90!}{6!(90-6)!} = \frac{90!}{6!84!}$$

che fornisce 622 614 630 possibilità. Questo spiega l'estrema difficoltà di vincere il jackpot.

B.5 Il coefficiente binomiale

Il coefficiente binomiale $\binom{n}{k}$ conta in quanti modi posso disporre k oggetti indistinguibili in $n \geq k$ posti:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Proprietà utili

$$\binom{n}{k} = \binom{n}{n-k}, \qquad \text{Per esempio:}$$

$$\begin{aligned}
\binom{5}{3} &= \binom{5}{2} = 10, \\
\binom{n}{0} &= \binom{n}{n} = 1 \\
\binom{n}{1} &= \binom{n}{n-1} = n \\
\binom{n}{k} &= \binom{n-1}{k-1} + \binom{n-1}{k}
\end{aligned}$$

In matematica $\binom{n}{k}$ è il k-esimo elemento della n-esima riga del *triangolo di Tartaglia*.

	0	1	2	3	4	5	6	7
0	1							
1	1	1						
2	1	2	1					
3	1	3	3	1				
4	1	4	6	4	1			
5	1	5	10	10	5	1		
6	1	6	15	20	15	6	1	
7	1	7	21	35	35	21	7	1

Richiami di Matematica

C

C.1 Richiami sui logaritmi

Si definisce $\log x$ il logaritmo naturale di $x > 0$ - $\lim_{x \to 0} \log x = -\infty$ - $\log e = 1$, dove e è il numero di Nepero 2.7183...

Una delle utilità dei logaritmi è di trasformare il logaritmo del prodotto in somma dei logaritmi dei fattori.

Proprietà 1

$$\log a \cdot b = \log a + \log b$$

ed esprime la potenza come coefficiente moltiplicativo:

Proprietà 2

$$\log a^b = b \log a$$

Consideriamo il prodotto di logaritmi:

$$\begin{aligned} \log \prod_{i=1}^{n} a_i &= \log a_1 \cdot ... \cdot a_n \\ &= \log a_1 + ... + \log a_n \\ &= \sum_{i=1}^{n} \log a_i \end{aligned}$$

Inoltre

$$\begin{aligned} \log \prod_{i=1}^{n} a_i^{b_i} &= \log a_1^{b_1} \cdot ... \cdot a_n^{b_i} \\ &= b_1 \log a_1 + ... + a_n \log a_n \\ &= \sum_{i=1}^{n} b_i \log a_i \end{aligned}$$

C.2 Richiami di Analisi

C.2.1 Note sulla cardinalità degli insiemi

In matematica la cardinalità di un insieme indica il numero dei suoi elementi.

L'insieme $E = \{a, b, c\}$ ha *cardinalità* finita

$$\text{card}(E) = \#E = 3$$

L'insieme dei numeri $S = \{0, 1, 2, 3, ..., n\}$ ha cardinalità finita:

$$\#S = n + 1$$

L'insieme dei numeri naturali

$$\mathbb{N} = \{0, 1, 2, 3, ...\}$$

ha **cardinalità infinita numerabile**

$$\#\mathbb{N} = \aleph_0, \qquad \text{infinito numerabile}$$

L'insieme dei reali

$$\mathbb{R} = \mathbb{Q} \cup \mathbb{I}, \qquad \text{i numeri reali sono dati dall'unione dei razionali } \mathbb{Q} \text{ e degli irrazionali } \mathbb{I}$$

ha **cardinalità infinita più che numerabile**

$$\#\mathbb{R} = \aleph_1, \qquad \text{infinito più che numerabile}$$

C.2.2 Funzioni Reali e loro derivate

Solitamente in analisi matematica si studia una generica funzione f, dove la variabile è x. La maggior parte degli esercizi riguardano

$$f(x), x \in \text{Dominio di } f$$

In particolare, se $f(x) = \log x, x > 0$ sappiamo che

$$f'(x) = \frac{1}{x}$$

Nella teoria della verosimiglianza lasceremo la lettera f assegnata alle funzioni di probabilità (nel caso discreto) e le funzioni di densità (nel caso continuo) e useremo lettere differenti per indicare la funzione. Allo stesso modo le etichette x restano per individuare i dati e le variabili sono i parametri del modello.

Quindi scriveremo, per esempio,

$$g(\theta) = \log \theta, \quad \theta \in \Theta$$

e leggeremo: g è funzione di θ e se dobbiamo derivare la funzione lo facciamo rispetto a θ:

$$g'(\theta) = \frac{1}{\theta}$$

Ricordiamo qualche semplice regola di derivazione:

Se $f(\theta) = g(h(\theta))$

$$f'(\theta) = g'(h(\theta))h'(\theta)$$

Se $f(\theta) = \log h(\theta)$

$$f'(\theta) = \frac{h'(\theta)}{h(\theta)}$$

0.65 FALSE

D Com'è Realizzato il Libro

Per scrivere queste pagine mi sono avvalso di diverse tecnologie che grazie all'ambiente integrato R-Studio (cita) è possibile combinare insieme con uno sforzo relativamente basso. Tra i tanti software coinvolti ricordo

- R [1]
- R Markdown [2], [3], [4]

R è l'ambiente di calcolo col quale ho svolto tutti i calcoli e realizzato le figure. R Markdown è un software che consente di mescolare pezzi di codici HTML, pezzi di codici LaTeX e pezzi di codice R. In sostanza lo stesso documento contiene:

- il contenuto del testo;
- i codici per produrre i calcoli e i grafici;
- i codici che consentono la formattazione del testo

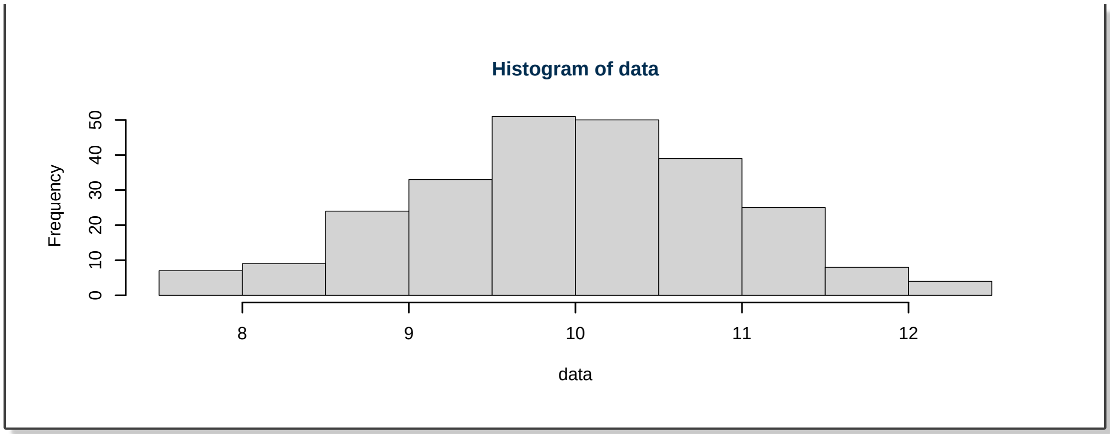
per esempio il codice

````
```{r, echo=FALSE}
# Questo è un blocco di codice in R
# fisso alcuni parametri ed eseguo alcuni conti
n <- 250 # fisso n = 250
data <- rnorm(n,10,1) # genero n dati da una normale di di media 10 e sd 1
mx <- mean(data) # chiamo mx la media dei dati
vx <- var(data) # chiamo vx la varianza dei dati
```
Abbiamo analizzato $n=`r n`$ individui, abbiamo osservato una media pari
a $\bar x =`r mx`$ e una varianza pari a $\sigma^2=`r vx`$

```{r, echo=FALSE}
# qui produco un grafico
hist(data)
```
````

produrrà il seguente risultato

> Abbiamo analizzato $n = 250$ individui, abbiamo osservato una media pari a $\bar{x} = 9.982$ e una varianza pari a $\sigma^2 = 0.9465$.



Quindi cioè che viene prodotto all'interno dei blocchi di R (chiamati R chunks), è utilizzabile nel testo richiamando `` `r nome_comando` `` il testo finale restituirà il risultato di `nome_comando`.

Le espressioni scritte tra `$` sono pezzi di codice LaTeXche è un potentissimo software di video scrittura, particolarmente adatto per scrivere formule matematiche. Per esempio il codice

```
$$
\bar x = \frac 1n \sum_{i=1}^n x_i
$$
```

produce

$$
\bar x = \frac{1}{n} \sum_{i=1}^n x_i
$$

L'obiettivo delle seguenti pagine è documentare la realizzazione del libro e non quello di approfondire passo, passo i software utilizzati. Farò solo una breve presentazione per stimolare i lettori più curiosi a cercare documentazioni più approfondite dei principali ambienti che ho utilizzato.

D.1 R: A Language and Environment for Statistical Computing

L'ambiente open source R è una collezione software per analizzare, manipolare e rappresentare dati. È sviluppato per tantissimi sistemi operativi tra cui Windows, Linux, Mac OS e tanti altri. Per maggiori informazioni su download e installazione rimando al sito. [https://www.R-project.org/][https://www.R-project.org/] nel quale si trovano diversi documenti di introduzione al software non solo inglese:

- https://cran.r-project.org/manuals.html
- https://cran.r-project.org/other-docs.html

R è un interprete interattivo, l'utente scrive un comando e una volta premuto invio R restituisce il risultato.

D.1.1 R come calcolatrice

R è anzitutto una calcolatrice

```
1+1 # uno più uno
```

```
## [1] 2
```

```
3-4 # tre meno quattro
```

```
## [1] -1
```

```
4*3 # quattro per tre
```

```
## [1] 12
```

```
4/3 # 4 diviso 3
```

```
## [1] 1.3333
```

```
5^2 # cinque elevato alla due
```

```
## [1] 25
```

```
log(2) # logaritmo naturale di due
```

```
## [1] 0.69315
```

```
exp(1) # e alla uno, il numero di Nepero
```

```
## [1] 2.7183
```

```
pi # pi greco, la costante trigonometrica
```

```
## [1] 3.1416
```

```
1/0 # 1 diviso zero restituisce  infinito
```

```
## [1] Inf
```

D.1.2 Operatori Speciali

Alcuni caratteri speciali sono

- `NA` Not Available è dedicato ai dati mancanti
- `Inf` infinito
- `NULL` non presente, nullo
- `TRUE` o `T` vero
- `FALSE` o `F` falso

```
NA+1      # sommare uno al dato mancante dà come risultato un dato mancante
```

```
## [1] NA
```

```
Inf+1     # infinito più uno è sempre infinito
```

```
## [1] Inf
```

```
NULL+1    # sommare al nulla uno non restituisce nulla
```

```
## numeric(0)
```

```
TRUE + 1  # vero + 1 fa due
```

```
## [1] 2
```

```
FALSE + 1 # falso + 1 fa uno
```

```
## [1] 1
```

- `a==b` a è uguale a b?
- `a!=b` a è diverso a b?
- `a>b` a è maggiore di b?

```
3 == 4   # 3 è uguale a 4?
```

```
## [1] FALSE
```

```
3 != 4   # 3 è diverso da 4?
```

```
## [1] TRUE
```

```
3 >  4   # 3 è maggiore di 4?
```

```
## [1] FALSE
```

Gli operatori `&` e `|` svolgono il ruolo di `and` e `or`.

```
cond1 <- TRUE
cond2 <- FALSE

cond1 & cond2 # vera se sono vere entrambe
```

```
## [1] FALSE
```

```
cond1 | cond2 # vera se almeno una delle due è vera
```

```
## [1] TRUE
```

D.1.3 Vettori e matrici

Con il comando `c` è possibile concatenare dei valori per creare un vettore, mentre si possono assegnare etichette agli oggetti con l'operatore `<-`

```
x <- c(2,1,9)
x
```

```
## [1] 2 1 9
```

```r
y <- c(2.3,1.4,2.8)
y
```

```
## [1] 2.3 1.4 2.8
```

```r
x+y # somma elemento per elemento
```

```
## [1]  4.3  2.4 11.8
```

Alcune funzioni speciali aiutano a velocizzare la creazione di vettori, per esempio `1:10` produce i numeri che vanno da uno a dieci.

```r
1:10
```

```
##  [1]  1  2  3  4  5  6  7  8  9 10
```

Alternativamente la funzione **`seq`** produce una sequenza di numeri dal minimo al massimo secondo alcuni criteri

```r
s1 <- seq(0,1,by=.1)      # sequenza da 0 ad 1 di passo 0.1
s2 <- seq(0,1,length=6)   # sequenza da 0 ad 1 di 6 numeri
s1
```

```
##  [1] 0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
```

```r
s2
```

```
## [1] 0.0 0.2 0.4 0.6 0.8 1.0
```

La lunghezza di un vettore si ricava col comando **`lenght`**

```r
length(s2)
```

```
## [1] 6
```

Per estrarre elementi da un vettore usiamo le parentesi quadre: `s2[2]` restituisce il secondo elemento di `s2` mentre `s2[1:4]` restituisce i primi 4 elementi e `s2[c(3,2,5)]` il terzo, il secondo e il quinto

```r
s2[2]
```

```
## [1] 0.2
```

```r
s2[1:4]
```

```
## [1] 0.0 0.2 0.4 0.6
```

```r
s2[c(3,2,5)]
```

```
## [1] 0.4 0.2 0.8
```

```r
x <- c(1.1,-3.4,4.2,5.1,-4.4,-3.9,2.5)
x > 0 # sequenza di TRUE e FALSE
```

```
## [1]  TRUE FALSE  TRUE  TRUE FALSE FALSE  TRUE
```

```r
which(x > 0) # in quale posizione sono gli x maggiori di zero?
```

```
## [1] 1 3 4 7
```

```r
x[x > 0] # gli x, ma solo quelli maggiori di zero
```

```
## [1] 1.1 4.2 5.1 2.5
```

```r
x[!(x > 0)] # gli x, ma solo quelli che NON sono maggiori di zero
```

```
## [1] -3.4 -4.4 -3.9
```

È possibile costruire matrici con la funzione `matrix`

```r
mat1 <- matrix(c(1,2,3,4),nrow = 2) # legge in colonna
mat1
```

```
##      [,1] [,2]
## [1,]    1    3
## [2,]    2    4
```

```r
mat2 <- matrix(c(1,2,3,4),nrow = 2,byrow = T) # legge per riga
mat2
```

```
##      [,1] [,2]
## [1,]    1    2
## [2,]    3    4
```

```r
mat1 + mat2 # restituisce la somma elemento per elemento
```

```
##      [,1] [,2]
## [1,]    2    5
## [2,]    5    8
```

Le funzioni `cbind` ed `rdbind` consentono di unire una nuova colonna o una nuova riga ad una matrice o vettore.

```r
x1 <- 1:3
x2 <- 2:4
cbind(x1,x2)
```

```
##      x1 x2
## [1,]  1  2
## [2,]  2  3
## [3,]  3  4
```

```r
rbind(x1,x2)
```

```
##    [,1] [,2] [,3]
```

```
## x1    1    2    3
## x2    2    3    4
```

Per indicizzare una matrice useremo sempre le parentesi quadre ma con 2 indici

```
mat <- matrix(1:20,nrow = 5)
mat
```

```
##      [,1] [,2] [,3] [,4]
## [1,]    1    6   11   16
## [2,]    2    7   12   17
## [3,]    3    8   13   18
## [4,]    4    9   14   19
## [5,]    5   10   15   20
```

```
mat[1,1]
```

```
## [1] 1
```

```
mat[2:3,3:4]
```

```
##      [,1] [,2]
## [1,]   12   17
## [2,]   13   18
```

I vettori e le matrici non sono solo numeriche ma possono anche contenere caratteri, ovviamente le operazioni aritmetiche non sono più consentite

```
mat <- matrix(c("testo 1","testo 2","testo 3","testo 4"),nrow = 2)
mat
```

```
##      [,1]      [,2]
## [1,] "testo 1" "testo 3"
## [2,] "testo 2" "testo 4"
```

D.1.4 Liste e dataframe

Una lista è una collezione di diversi oggetti di R

```
mat1 <- matrix(c(1,2,3,4),nrow = 2) # legge in colonna
mat <- matrix(c("testo 1","testo 2","testo 3","testo 4"),nrow = 2)
b <- c(NA,NA,NA)

lista <- list(mat1,mat,b)
lista
```

```
## [[1]]
##      [,1] [,2]
## [1,]    1    3
## [2,]    2    4
```

```
##
## [[2]]
##      [,1]     [,2]
## [1,] "testo 1" "testo 3"
## [2,] "testo 2" "testo 4"
##
## [[3]]
## [1] NA NA NA
```

Una particolare tipo di lista è il `data.frame` che consente di creare una matrice dei dati composta da colonne di diversa natura

```
sesso <- c("M","M","M","F","F")
eta   <- c(32.2,45.6,65.3,34.1,43.2)
dati <- data.frame(sesso,eta)
dati
```

```
##   sesso  eta
## 1     M 32.2
## 2     M 45.6
## 3     M 65.3
## 4     F 34.1
## 5     F 43.2
```

il simbolo del `$` aiuta a selezionare le colonne di interesse

```
dati$sesso
```

```
## [1] "M" "M" "M" "F" "F"
```

```
dati$eta
```

```
## [1] 32.2 45.6 65.3 34.1 43.2
```

D.1.5 Classi e Oggetti

R è un linguaggio funzionale, ogni elemento è un oggetto che ha un classe e metodi. Per esempio

```
x <- 3
class(x)
```

```
## [1] "numeric"
```

```
x <- matrix(1:4,nrow = 2)
class(x)
```

```
## [1] "matrix" "array"
```

```
x <- list(1:3,c("a","b"))
class(x)
```

```
## [1] "list"
```

Attenzione

Alcuni oggetti possono assomigliarsi ma essere diversi ad esempio

```
x1 <- 1;3
```

```
## [1] 3
```

```
x2 <- 2:4
mat <- cbind(x1,x2)
dat <- data.frame(x1,x2)
mat
```

```
##      x1 x2
## [1,]  1  2
## [2,]  1  3
## [3,]  1  4
```

```
dat
```

```
##   x1 x2
## 1  1  2
## 2  1  3
## 3  1  4
```

```
class(mat)
```

```
## [1] "matrix" "array"
```

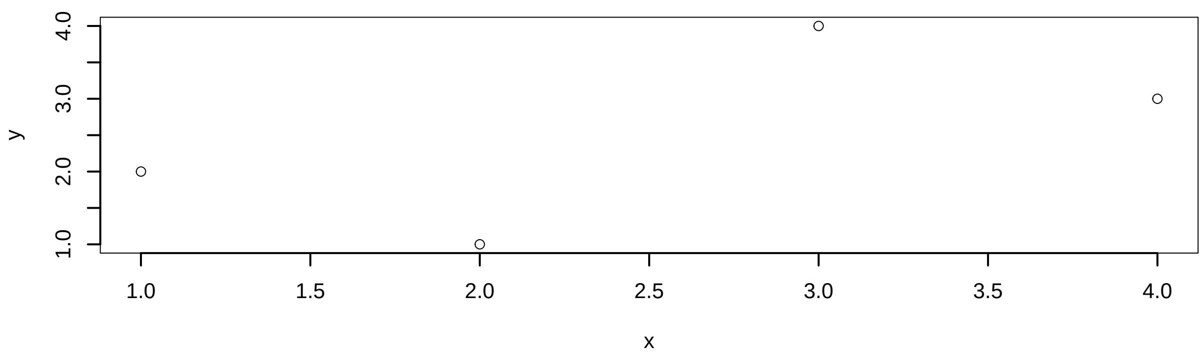
```
class(dat)
```

```
## [1] "data.frame"
```

D.1.6 I grafici

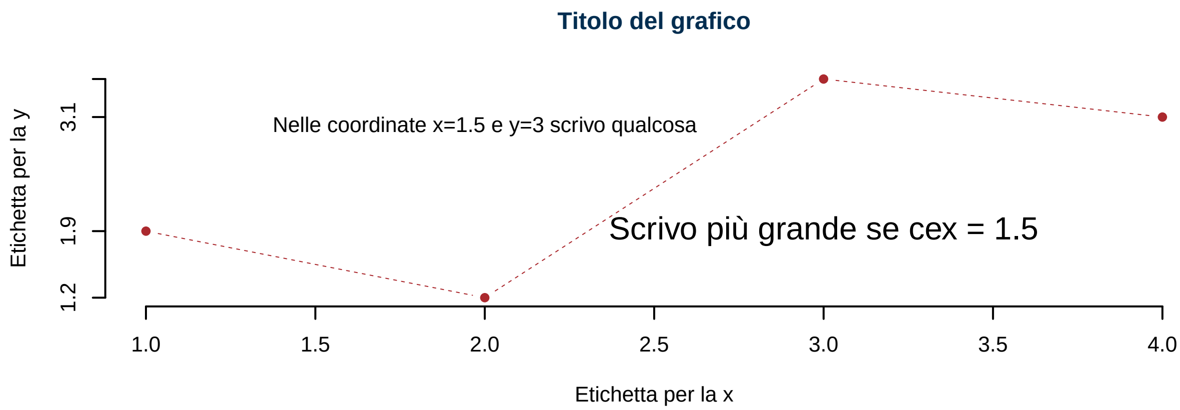
La libraria grafica di R è particolarmente ricca. La funzione di base per realizzare un grafico è la funzione `plot(x,y)`. La funzione, di default, disegna i punti di coordinate `x` e `y`.

```
x <- c(1,2,3,4)
y <- c(2,1,4,3)
plot(x,y)
```



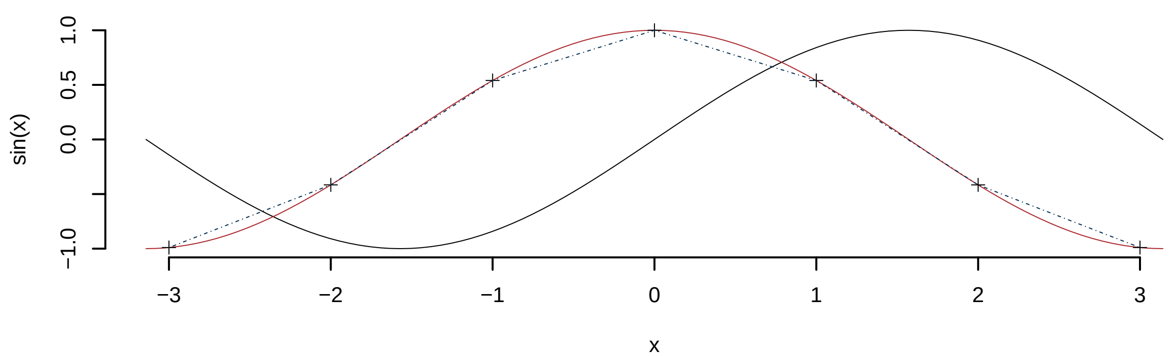
La funzione `plot` è molto flessibile e può essere arricchita con molte opzioni

```
x <- c(1,2,3,4)
y <- c(1.9,1.2,3.5,3.1)
plot(x,y,
     axes=F,      # non disegnare gli assi lo farò dopo
     pch = 16,    # codice 16, per il pallino chiuso
     col = ared, #colore rosso per i pallini
     xlab="Etichetta per la x",
     ylab="Etichetta per la y",
     type = "b",  # linea e punto
     lty  =2,     # stile di tratteggiatura
     main = "Titolo del grafico"
          )
# Un volta creato il grafico di base possiamo aggiungere
axis(1) # asse delle x
axis(2,at = y) # asse delle y sui punti osservati
text(2,3,"Nelle coordinate x=1.5 e y=3 scrivo qualcosa")
text(3,1.9,"Scrivo più grande se cex = 1.5",cex=1.5)
```



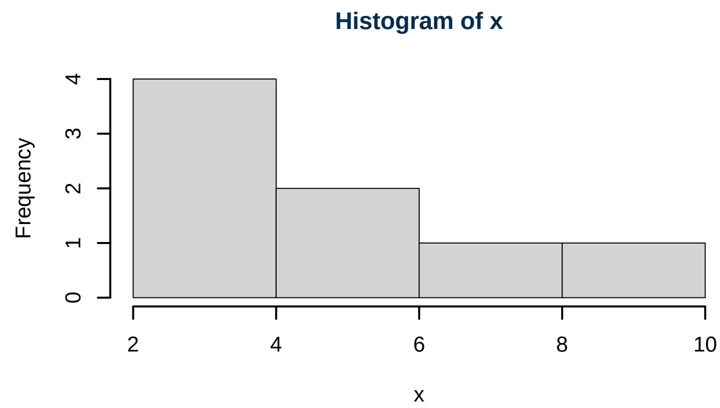
Con le funzioni `points` e `lines` è possibile aggiungere al grafico esistente punti e linee, rispettivamente. Mentre per disegnare funzioni la funzione `curves` aiuta molto

```
curve(sin(x),from = -pi,to = pi,axes=F) # disegna sin(x) tra - pi e + pi
curve(cos(x),col=ared,add=T)           # aggiunge il grafico di cos(x) in rosso
points((-3):3,cos((-3):3),pch=3)
lines((-3):3,cos((-3):3),col=iblue,lty=4)
axis(1)
axis(2)
```

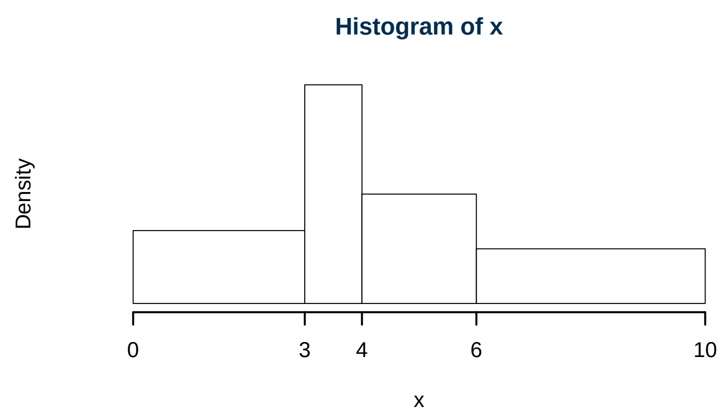


Un'altra funzione di interesse è `hist` che produce istogrammi di densità

```
x <- c(2.3,2.5,3.5,3.6,4.6,5.9,6.9,9.8)
hist(x)
```

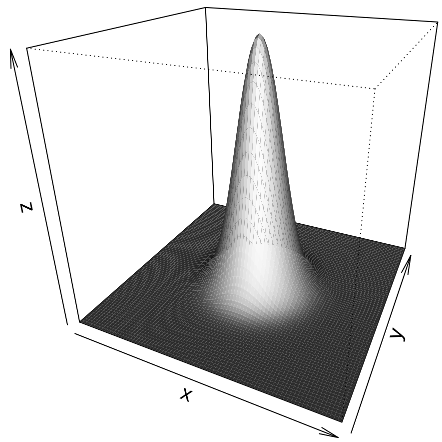



```
hist(x,breaks = c(0,3,4,6,10),probability = T,axes=F,col="white")
axis(1,at = c(0,3,4,6,10))
```



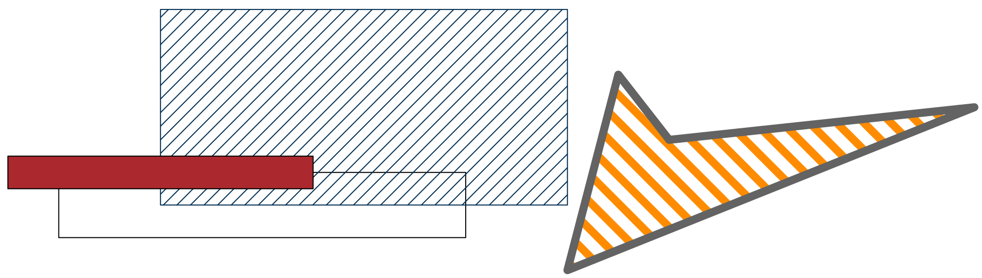
Per rappresentare dati in 3D la funzione `persp` disegna superfici in 3 dimensioni

```
x <- seq(-4,4,by=.1)
y <- seq(-4,4,by=.1)
z <- outer(x,y,function(x,y)exp(-x^2-y^2))
persp(x,y,z,theta = 25,phi = 25,ltheta = 25,border = NA,shade = 2.5)
```



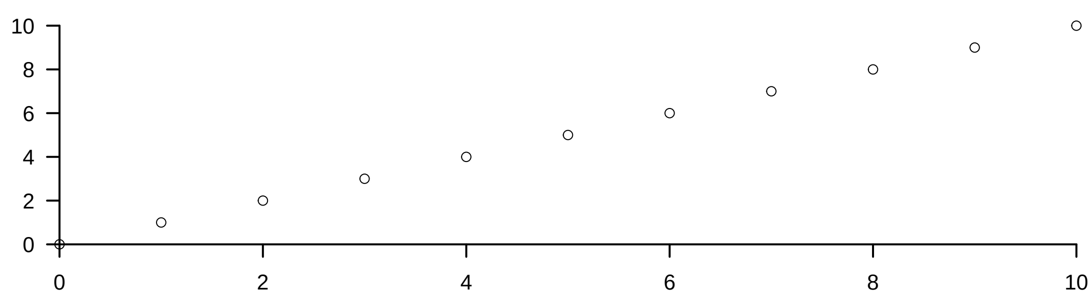
Altre funzioni grafiche interessanti sono `rect` e `polyogn` che aggiungo ai grafici rettangoli e poligoni, rispettivamente

```
plot(c(0,10),c(0,10),xlab = "", ylab="",axes=F,type="n") # inizia un grafico vuoto
rect(xleft = 1,ybottom = 2,xright = 5,ytop = 4)
rect(xleft = 2,ybottom = 3,xright = 6,ytop = 9,density = 20,col=iblue)
rect(xleft = .5,ybottom = 3.5,xright = 3.5,ytop = 4.5,col=ared)
polygon(x = c(6,6.5,7,10),y = c(1,7,5,6),density = 10,
        col="darkorange",angle = -45,border = "gray39",lwd=4)
```

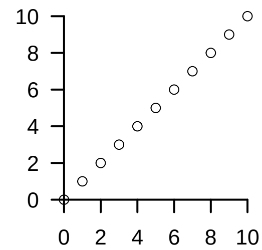


Le proporzioni tra `x` ed `y` si possono forzare con l'opzione `asp`

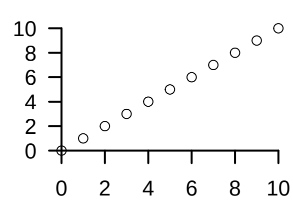
```
plot(0:10,0:10,xlab = "", ylab="",axes=F) # asp non specificato
axis(1,pos = 0)
axis(2,pos = 0,las=2)
```



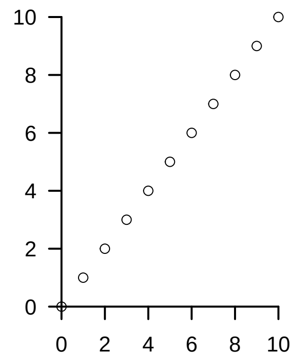

```
plot(0:10,0:10,xlab = "", ylab="",axes=F,asp=1) # asp=1
axis(1,pos = 0,at = seq(0,10,by=2))
axis(2,pos = 0,las=2)
```




```
plot(0:10,0:10,xlab = "", ylab="",axes=F,asp=9/16) # asp=9/16
axis(1,pos = 0)
axis(2,pos = 0,las=2)
```



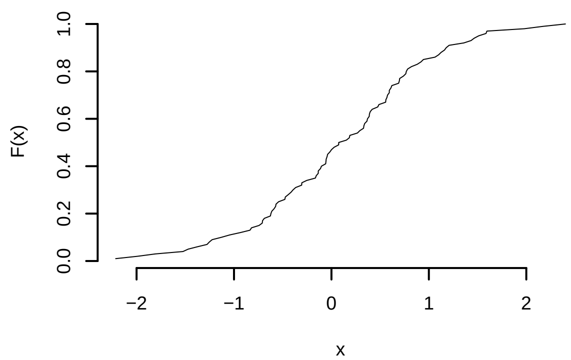

```
plot(0:10,0:10,xlab = "", ylab="",axes=F,asp=4/3) # asp=4/3
axis(1,pos = 0)
axis(2,pos = 0,las=2,at = seq(0,10,by=2))
```

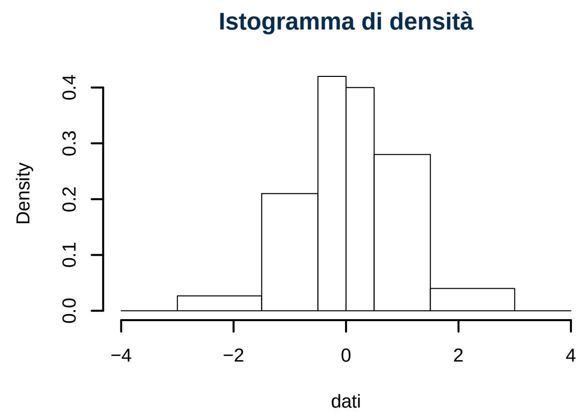


I parametri del grafico si possono settare attraverso `par`

```
set.seed(1)
par(mfrow=c(1,2),cex=0.6) # 1 riga e 2 colonne
dati <- rnorm(100) # 100 dati dalla normale
plot(
  sort(dati),(1:100)/100,xlab = "x", ylab="F(x)",axes=F,type="l"
  ) # FdR empirica
axis(1)
axis(2)

hist(
  dati,c(-4,-3,-1.5,-.5,0,.5,1.5,3,4),main="Istogramma di densità",col="white"
  ) # istogramma dei dati
```





D.1.7 Le Funzioni in R

Essendo R un linguaggio funzionale ogni funzione (che è essa stessa un oggetto) si alimenta di uno o più oggetti di determinate classi e restituisce uno o più oggetti di una determina classe. Per esempio abbiamo visto come la funzione `matrix` abbia come input un vettore e come output una matrice.

```
x <- 1:10
class(x)
```

```
## [1] "integer"
```

```
y <- matrix(x,nrow = 2)
class(y)
```

```
## [1] "matrix" "array"
```

Le funzioni di R hanno argomenti con etichette, se gli argomenti rispettano l'ordine delle etichette non c'è bisogno di chiamarli altrimenti vanno etichettati anche loro. Per esempio la funzione `matrix` si aspetta almeno due argomenti i dati `data` e il numero di righe `nrow` o il numero di colonne `ncol`

```
matrix(data,nrow,ncol,byrow = F)
```

```
x <- 1:10
matrix(x, 2)                # è identico a
```

```
##      [,1] [,2] [,3] [,4] [,5]
## [1,]    1    3    5    7    9
## [2,]    2    4    6    8   10
```

```
matrix(nrow = 2, data = x)  # ma è diverso da
```

```
##      [,1] [,2] [,3] [,4] [,5]
## [1,]    1    3    5    7    9
## [2,]    2    4    6    8   10
```

```r
matrix(data = x, ncol = 2)  #
```

```
##      [,1] [,2]
## [1,]    1    6
## [2,]    2    7
## [3,]    3    8
## [4,]    4    9
## [5,]    5   10
```

```r
matrix(2, x)                 # produce un risultato senza senso
```

```
##      [,1]
## [1,]    2
```

D.1.7.1 Le funzioni statistiche

Tra le funzioni di base ricordiamo `mean(x)` la media del vettore `x` `var(x)` la varianza **corretta**, `sd(x)` la SD **corretta** , `median(x)` la mediana e `quantile(x,p)` il percentile di ordine p. Per esempio

```r
x <- c(2.3,4.5,6.7,2.1,3.8,2.5,6.9,9.8)
mean(x)   # media
```

```
## [1] 4.825
```

```r
var(x)    # varianza corretta
```

```
## [1] 7.5621
```

```r
sd(x)     # sd corretta
```

```
## [1] 2.7499
```

```r
quantile(x,c(.25,.50,.75)) # i percentili per p=.25,.50,.75 => i quartili
```

```
##  25%  50%  75%
## 2.45 4.15 6.75
```

Per trattare le variabili doppie nella regressione ci sono strumenti appositi, le funzioni `cov(x,y)` e `cor(x,y)` calcolano la correlazione e la covarianza tra i vettori `x` ed `y`

```r
x <- c(2.3,4.5,6.7,2.1,3.8,2.5,6.9,9.8)
y <- c(2.4,2.1,5.6,7.2,6.5,7.1,4.3,9.7)
cov(x,y)   # covarianza
```

```
## [1] 2.5925
```

```r
cor(x,y)   # r
```

```
## [1] 0.36535
```

```
cor(x,y)^2 # r quadro
```

```
## [1] 0.13348
```

La funzione `lsfit` (Least Squared Fit => Stima dei Minimi Quadrati) consente di calcolare $\hat{\beta}_0$ e $\hat{\beta}_1$ rapidamente

```
x <- c(1,2,3,4)
y <- c(1.9,1.2,3.5,3.1)
modello <- lsfit(x,y)
modello$coefficients # coefficienti beta 0 e beta 1
```

```
## Intercept         X
##      0.95      0.59
```

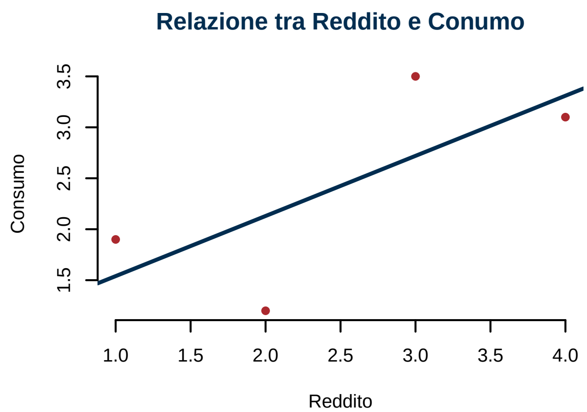
```
modello$residuals    # i residui osservati gli epsilon i
```

```
## [1]  0.36 -0.93  0.78 -0.21
```

```
plot(x,y,
     axes=F,      # non disegnare gli assi lo farò dopo
     pch = 16,    # codice 16, per il pallino chiuso
     col = ared, #colore rosso per i pallini
     xlab="Reddito",
     ylab="Consumo",
     lty  =2,     # stile di tratteggiatura
     main = "Relazione tra Reddito e Conumo"
         )
# Un volta creato il grafico di base possiamo aggiungere
axis(1) # asse delle x
axis(2) # asse delle y

# e aggiungere una retta di coefficienti beta 0 e beta 1

abline(modello$coefficients,lwd=2,col=iblue)
```



D.1.7.2 Le VC di maggiore interessa

In R sono tabulate tutte le distribuzioni di maggiore interesse, comprese, la binomiale, la Poisson, la Normale, la t di studente e il chi quadro, insieme a moltissime altre. La sintassi è relativamente semplice il prefisso `d` indica la funzione di densità o di probabilità il prefisso `p` indica la funzione di ripartizione, il prefisso `q` indica l'inverso della funzione di ripartizione e il prefisso `r` genera numeri casuali da quella distribuzione. Per esempio se $X \sim \text{Binom}(n = 5; \pi = 0.3)$

```
dbinom(x = 3,size = 5,prob = 0.3) # è la probabilità che X=3
```

[1] 0.1323

```
dbinom(x = 0:5,size = 5,prob = 0.3) #  tutte le probabilità per X=0, 1, ..., 5
```

[1] 0.16807 0.36015 0.30870 0.13230 0.02835 0.00243

```
plot(0:5,dbinom(x = 0:5,size = 5,prob = 0.3),type="h",lwd=2) # grafico
pbinom(q = 3,size = 5,prob = 0.3) # è la probabilità che X 3
```

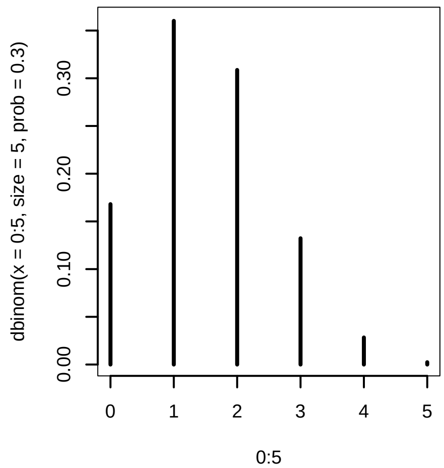
[1] 0.96922

```
qbinom(p = 0.75,size = 5,prob = 0.3) # 75-esimo percentile di X
```

[1] 2

```
rbinom(n = 10,size = 5,prob = 0.3)  # simula 10 estrazioni da X
```

[1] 1 1 1 1 1 1 2 0 1 2



Per esempio se $X \sim \text{Pois}(\lambda = 4.3)$

```
dpois(x = 3,lambda = 4.3) # è la probabilità che X=3
```

[1] 0.1798

```
dpois(x = 0:10,lambda = 4.3) #  le probabilità per X=0, 1, ..., 10
```

[1] 0.0135686 0.0583448 0.1254413 0.1797992 0.1932842 0.1662244 0.1191275 [8] 0.0731783 0.0393333 0.0187926 0.0080808

```
plot(0:10,dpois(x = 0:10,lambda = 4.3),type="h",lwd=2) # grafico
ppois(q = 3,lambda = 4.3) # è la probabilità che X 3
```

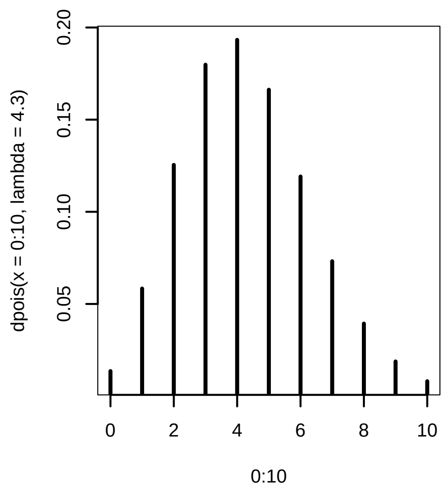
[1] 0.37715

```
qpois(p = .75,lambda = 4.3) # 75-esimo percentile di X
```

[1] 6

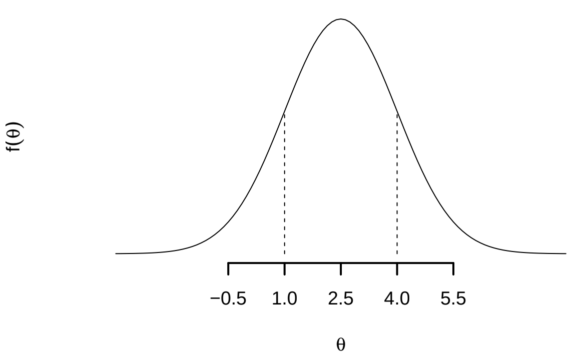
```
rpois(n = 10,lambda = 4.3)  # simula 10 estrazioni da X
```

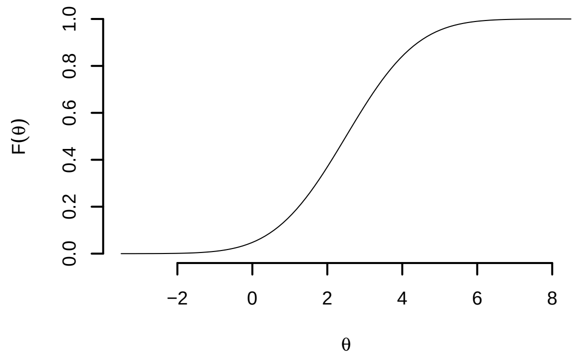
[1] 8 2 6 8 6 3 5 8 8 3



Per esempio se $X \sim N(\mu = 2.5, \sigma^2 = (1.5)^2)$

```r
par(mfrow=c(1,2),cex=0.6) # metto una figura accanto all'altra
curve(dnorm(x,mean = 2.5,sd = 1.5),
      from = 2.5-4*1.5,
      to = 2.5+4*1.5,
      axes=F,
      ylab = expression(f(theta)),
      xlab = expression(theta))
axis(1,c(2.5-2*1.5,2.5-1.5,2.5,2.5+1.5,2.5+2*1.5))
segments(x0 = c(2.5-1.5,2.5+1.5),
         y0 = 0,
         x1 = c(2.5-1.5,2.5+1.5),
         y1 = dnorm(c(2.5-1.5,2.5+1.5),2.5,1.5),
         lty = 2)
curve(pnorm(x,mean = 2.5,sd = 1.5),
      from = 2.5-4*1.5,
      to = 2.5+4*1.5,
      axes=F,
      ylab = expression(F(theta)),
      xlab = expression(theta))
axis(1)
axis(2)
```






```
pnorm(3,mean = 2.5,sd = 1.5)      # P(X<3)
```

```
## [1] 0.63056
```

```
1 - pnorm(3,mean = 2.5,sd = 1.5)  # P(X>3)
```

```
## [1] 0.36944
```

```
pnorm(3,mean = 2.5,sd = 1.5)-pnorm(0.5,mean = 2.5,sd = 1.5) #  P(0.5 < X < 3)
```

```
## [1] 0.53935
```

```
rnorm(25,2.5,1.5) # 25 estrazioni da X
```

```
##  [1] 1.54640 1.80753 4.64842 1.52396 2.18893 1.91079 2.02001 2.08133 3.24128
## [10] 2.23400 1.74106 4.51456 2.17813 2.23067 2.34971 3.56900 2.38965 2.44355
## [19] 1.47751 2.01359 2.59024 1.61666 3.29724 0.22241 2.95984
```

D.1.7.3 Funzioni tra stringhe

Una stringa è un vettore di testo, non numerico, la funzione `paste` consente di incollare testi.

```
vettore_testo1 <- c("mela","pera","pesca","banana")
vettore_testo2 <- c("rossa","verde","gialla","gialla")
paste(vettore_testo1,vettore_testo2,sep=" è ")
```

```
## [1] "mela è rossa"    "pera è verde"    "pesca è gialla"  "banana è gialla"
```

```
paste("la",paste(vettore_testo1,vettore_testo2,sep=" è "),collapse = " e ")
```

```
## [1] "la mela è rossa e la pera è verde e la pesca è gialla e la banana è gialla"
```

```
x <- round(rnorm(5,10,1),2)
x
```

```
## [1] 8.46 9.70 9.47 9.35 9.94
```

```
paste("La somma degli x è",paste(x,collapse = " + ")," = ", sum(x))
```

```
## [1] "La somma degli x è 8.46 + 9.7 + 9.47 + 9.35 + 9.94  =  46.92"
```

La funzione `cat` stampa su schermo o su file e consente l'utilizzo di alcuni caratteri speciali

```
cat(vettore_testo1,vettore_testo2)
```

```
## mela pera pesca banana rossa verde gialla gialla
```

```
cat(vettore_testo1,"\t",vettore_testo2)
```

```
## mela pera pesca banana    rossa verde gialla gialla
```

```
cat(vettore_testo1,"\n",vettore_testo2)
```

```
## mela pera pesca banana
##  rossa verde gialla gialla
```

D.1.7.4 Cicli e Condizioni

I cicli si possono fare con `for` ma come vedremo si possono aggirare in molti modi

```
for (i in 1:5){
  cat(i,"; ")
}
```

```
## 1 ; 2 ; 3 ; 4 ; 5 ;
```

Mentre le condizioni si risolvono con `if else`

```
a <- 12
if (a>10) {
  cat("Hai vinto! \n")
} else {cat("Hai perso! \n")}
```

```
## Hai vinto!
```

```
a <- 5
if (a>10) {
  cat("Hai vinto! \n")
} else {cat("Hai perso! \n")}
```

```
## Hai perso!
```

anche la funzione `ifelse` è di aiuto

```
cond <- TRUE
ifelse(cond,"VERO","FALSO")
```

```
## [1] "VERO"
```

```
cond <- FALSE
ifelse(cond,"VERO","FALSO")
```

```
## [1] "FALSO"
```

D.1.7.5 Funzioni per Ovviare ai Cicli

R è un interprete e i cicli rallentano molto il funzionamento dei programmi alcune funzioni speciali quali `apply`, `tapply`, `lapplay` e `sapply` applicano a diversi tipi di oggetti una funzione. `apply` applica una funzione alle righe o alle colonne di una matrice

```
mat <- matrix(rnorm(12),nrow = 3)
mat
```

```
##          [,1]     [,2]      [,3]     [,4]
## [1,] -1.9144 -0.46353  2.087167 -1.64061
## [2,]  1.1766 -1.11592  0.017396  0.45019
## [3,] -1.6650 -0.75082 -1.286301 -0.01856
```

```
apply(X = mat, MARGIN = 1,FUN = "sum") # applica a mat per righe la somma
```

```
## [1] -1.93133  0.52825 -3.72065
```

```
apply(X = mat, MARGIN = 2,FUN = "mean") # applica a mat per colonne la media
```

```
## [1] -0.80092 -0.77676  0.27275 -0.40299
```

`tapply` applica una funzione solo su elementi che rispettano una condizione

```
sesso <- c("M","M","M","F","F")
eta   <- c(32.2,45.6,65.3,34.1,43.2)
tapply(eta, sesso, median) # applica la mediana al gruppo M e al gruppo F
```

```
##     F     M
## 38.65 45.60
```

Le funzioni `lapplay` e `sapply` (versione semplificata delle prima) applicano una funzione ad una lista o agli elementi di un vettore

```
gruppo1 <- rnorm(5,9.4,1.1)  # gruppo1 5 estrazioni da una N( 9.4,1.1^2)
gruppo2 <- rnorm(9,12.2,1.2) # gruppo2 9 estrazioni da una N(12.2,1.2^2)
gruppo3 <- rnorm(7,11.7,0.9) # gruppo3 7 estrazioni da una N(11.7,0.9^2)
campione <- list(gruppo1,gruppo2,gruppo3)
# applica media e varianza ad ogni gruppo:
lapply(campione,FUN = function(x)c(mean(x),var(x)))
```

```
## [[1]]
## [1] 8.7818 1.1357
##
```

```
## [[2]]
## [1] 12.6776  2.1178
##
## [[3]]
## [1] 11.8613  1.0098
```

```
sapply(campione,FUN = function(x)c(mean(x),var(x))) # come sopra ma in matrice
```

```
##        [,1]    [,2]    [,3]
## [1,] 8.7818 12.6776 11.8613
## [2,] 1.1357  2.1178  1.0098
```

D.1.7.6 Funzioni personalizzate

Le funzioni si possono creare con la funzione `function`

```
funz1 <- function(x) x^2 # funz1 usa x come input e x^2 come output
funz1(3)
```

```
## [1] 9
```

```
funz2 <- function(x,y) { # funz2 usa x e y come input
  risultato <- ifelse(x>y,x+y,"x deve essere maggiore di y!")
  return(risultato)
}
funz2(2,3)
```

```
## [1] "x deve essere maggiore di y!"
```

```
funz2(3,1)
```

```
## [1] 4
```

Possiamo per esempio creare la funzione che crei la varianza del campione e la SD non corretta

```
var_pop <- function(x) {
  vpop <- mean(x^2)-mean(x)^2 # media dei quadrati meno il quadrato della media
  return(vpop)
}
n <- 10
x <- rnorm(n,mean = 10,sd = 1)
var(x)     # varianza corretta
```

```
## [1] 0.88378
```

```
var_pop(x) # varianza NON corretta
```

```
## [1] 0.7954
```

```
n/(n-1)*var_pop(x) # correzione
```

```
## [1] 0.88378
```

```
sd_pop <- function(x) sqrt(var_pop(x)) # rad. quad. della varianza
sd_pop(x)
```

```
## [1] 0.89185
```

Possiamo anche creare una funzione che calcola un intervallo di confidenza al livello $1 - \alpha$, per un α qualunque

```
## intervallo di confidenza per mu, sigma nota

idc_mu <- function(mu_obs,sigma,alpha=0.05){
z_alpha <- qnorm(1-alpha/2) # fisso z_alpha
SE      <- sigma/sqrt(n) # SE per mu = sigma diviso radice di n

cat("L'intervallo di confidenza al ",(1-alpha)*100,"% è dato da \n",
    "[",mu_obs," - ",z_alpha," x ",SE," ; ",mu_obs," + ",z_alpha," x ",SE,"] \n",
    "[",mu_obs-z_alpha*SE," ; ",mu_obs+z_alpha*SE,"]",
    sep="")
}

alpha   <- 0.01     # fisso alpha
mu_obs  <- 23.3     # fisso la media del campione
sigma   <- 2.13     # fisso il sigma di popolazione
n       <- 34       # fisso n

idc_mu(mu_obs = mu_obs, sigma = sigma, alpha = alpha)
```

```
## L'intervallo di confidenza al 99% è dato da
## [23.3 - 2.5758 x 0.36529 ; 23.3 + 2.5758 x 0.36529]
## [22.359 ; 24.241]
```

[1] R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

[2] JJ Allaire and Yihui Xie and Jonathan McPherson and Javier Luraschi and Kevin Ushey and Aron Atkins and Hadley Wickham and Joe Cheng and Winston Chang and Richard Iannone (2022). rmarkdown: Dynamic Documents for R. R package version 2.14. URL https://rmarkdown.rstudio.com.

[3] Yihui Xie and J.J. Allaire and Garrett Grolemund (2018). R Markdown: The Definitive Guide. Chapman and Hall/CRC. ISBN 9781138359338. URL https://bookdown.org/yihui/rmarkdown.

[4] Yihui Xie and Christophe Dervieux and Emily Riederer (2020). R Markdown Cookbook. Chapman and Hall/CRC. ISBN 9780367563837. URL https://bookdown.org/yihui/rmarkdown-cookbook.

Funzioni usate nel libro

E

Presento le funzioni che sono state create da me e inserite nel pacchetto `pat.book` per risolvere vari problemi di automazione, dalla creazione dei data set alla soluzione di alcuni problemi.

E.1 Installazione del pacchetto

Il pacchetto si installa da github

```
devtools::install_git("https://github.com/ix-pat/stat/",build_vignettes = T)
```

Le funzioni restituiscono pezzi di codice markdown come da esempio

```
stat_(1:4)
```

```
\begin{eqnarray*}
 \mu &=& \frac 1{ 4 }(  1 + 2 + 3 + 4  )= 2.5 \\
 \sigma^2 &=& \frac 1{ 4 }(  1 ^2+ 2 ^2+ 3 ^2+ 4 ^2 )-( 2.5 )^2= 1.25
\end{eqnarray*}
```

E.2 Statistica descrittiva

E.2.1 Media e Varianza

```
media_(1:4)                       # media dei dati 1,2,3,4
```

$$\mu = \frac{1}{4}(1+2+3+4) = 2.5$$

```
var_(1:4)                         # varianza dei dati 1,2,3,4
```

$$\sigma^2 = \frac{1}{4}(1^2+2^2+3^2+4^2) - (2.5)^2 = 1.25$$

```
stat_(1:4)                        # media e varianza insieme (new)
```

$$\begin{aligned} \mu &= \frac{1}{4}(1+2+3+4) = 2.5 \\ \sigma^2 &= \frac{1}{4}(1^2+2^2+3^2+4^2) - (2.5)^2 = 1.25 \end{aligned}$$

```
stat_(1:4,p = c(2,4,5,1))         # vettore dei pesi p
```

$$
\begin{aligned}
\mu &= 1 \cdot 0.1667 + 2 \cdot 0.3333 + 3 \cdot 0.4167 + 4 \cdot 0.0833 = 2.417 \\
\sigma^2 &= (1^2 \cdot 0.1667 + 2^2 \cdot 0.3333 + 3^2 \cdot 0.4167 + 4^2 \cdot 0.0833) - (29)^2 = 0.7431
\end{aligned}
$$

```
stat_(rep(1:4,times=c(2,4,5,1)),semp = T) # in frazione
```

$$
\begin{aligned}
\mu &= E(X_i) = \sum_{x \in S_X} xP(X = x) \\
&= 1\frac{2}{12} + 2\frac{4}{12} + 3\frac{5}{12} + 4\frac{1}{12} \\
&= 2.417 \\
\sigma^2 &= V(X_i) = \sum_{x \in S_X} x^2 P(X = x) - \mu^2 \\
&= \left(1^2\frac{2}{12} + 2^2\frac{4}{12} + 3^2\frac{5}{12} + 4^2\frac{1}{12}\right) - (2.417)^2 \\
&= 0.7431
\end{aligned}
$$

E.2.2 Istogramma

```
##############################################################################
##############################################################################
## Funzioni per generare i dati dell'esercizio 1
##
## genera_dati(brk,hhh=NULL,n,nnn=NULL,rand = T)
##
##  brk      intervalli (breaks)
##  hhh      aspetto presunto
##  n        numero totale individui
##  nnn      alternativo ad hhh, frequenza da riportare ad n
##  rand     i numeri sono casuali?
##
## tabl(x,...)                    shortcut personalizzato a kable
##
##  x        oggetto da stampare in tabella
##
## ls2e(stat_base(samp,brk))   crea diversi oggetti
##  dat2     tabella con intestazioni semplici
##  dat3     tabella con intestazioni da stampa
##  H.int(x) densità percentuale
##  F.int(x) Funzione di ripartizione
##  Q.int(p) Inversa della FdR
```

```
##    x      vettore di valori
##    p      vettore di frequenze
##  histp(axes=T,...) istogramma
##  h.int(x1,x2,...)  evidenzia istogramma
##    x1     limite inferiore
##    x2     limite superiore

set.seed(2)                      # per ottenere sempre la stessa simulazione
n <- 60                          # ampiezza campionaria

brk  <- c(0,1.5,3,5,7.5,15)      # intervalli (breaks)
hhh  <- c( 2,11,10, 2,1)         # aspetto presunto istogramma

nomex <- "Nome della X"          # nome della X
samp <- genera_dati(
  brk = brk,hhh = hhh,n = n)     # genera i dati dall'istogramma

ls2e(stat_base(samp,brk))      # crea il data set e la tabella dat3

tabl(dat3)
```

$[x_j,$	$x_{j+1})$	n_j	f_j	b_j	h_j	F_j	$\bar{x}_j$	$\bar{x}_j^2$	$\bar{x}_j n_j$	$\bar{x}_j^2 n_j$	$f_{j\%}$
0.0	1.5	3	0.0500	1.5	3.333	0.0500	0.75	0.5625	2.25	1.688	5.00
1.5	3.0	19	0.3167	1.5	21.111	0.3667	2.25	5.0625	42.75	96.188	31.67
3.0	5.0	23	0.3833	2.0	19.167	0.7500	4.00	16.0000	92.00	368.000	38.33
5.0	7.5	6	0.1000	2.5	4.000	0.8500	6.25	39.0625	37.50	234.375	10.00
7.5	15.0	9	0.1500	7.5	2.000	1.0000	11.25	126.5625	101.25	1139.062	15.00
		60	1.0000	15.0					275.75	1839.312	100.00

```
H.int(2:3)             # Calcolo della Densità percentuale
```

[1] 21.11 19.17

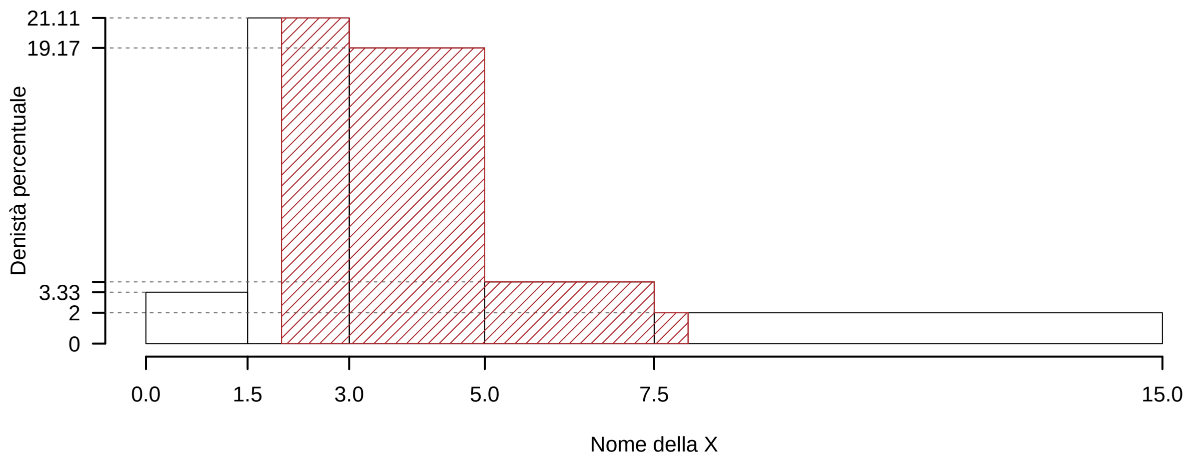
```
F.int(2:3)             # Calcolo della Ripartizione
```

[1] 0.1556 0.3667

```
Q.int((0:4)/4)         # Inverse della Ripartizione
```

[1] 0.000 2.447 3.696 5.000 15.000

```
histp(axes=T)          # Istogramma
h.int(2,8,col=ared,    # Aree selezionate
      density = 25)
```




```r
percentile(p = 0.45)        # Calcolo percentile
```

$$
\begin{aligned}
p &= 0.45, \text{essendo } F_3 = 0.75 > 0.45 \Rightarrow j_{0.45} = 3 \\
x_{0.45} &= x_{\inf;3} + \frac{0.45 - F_2}{f_3} \cdot b_3 \\
&= 3 + \frac{0.45 - 0.3667}{0.3833} \cdot 2 \\
&= 3.435
\end{aligned}
$$

```r
F_print(2,"<")              # calcolo della prop inferiore a 2
```

$$
\begin{aligned}
\%(X < 2) &= f_1 \times 100 + (2 - 1.5) \times h_2 \\
&= (0.05) \times 100 + (0.5) \times 21.11 \\
&= 0.1556 \times (100) \\
\#(X < 2) &\approx 9
\end{aligned}
$$

```r
F_print(2,">")              # calcolo della prop superiore a 2
```

$$
\begin{aligned}
\%(X > 2) &= (3 - 2) \times h_2 + f_3 \times 100 + f_4 \times 100 + f_5 \times 100 \\
&= (1) \times 21.11 + (0.3833) \times 100 + (0.1) \times 100 + (0.15) \times 100 \\
&= 0.8444 \times (100)
\end{aligned}
$$

$$\#(X > 2) \approx 51$$

```
F_print(8,">")
```

$$\begin{aligned}
\%(X > 8) &= (15-8) \times h_1 \\
&= 7 \times 2 \\
&= 0.14 \times (100) \\
\#(X > 8) &\approx 8
\end{aligned}$$

```
F_print(x = 2,verso = "",x2 = 8) # intervallo 2-8
```

$$\begin{aligned}
\%(2 < X < 8) &= (3-2) \times h_2 + f_3 \times 100 + f_4 \times 100 + (8-7.5) \times h_5 \\
&= (1) \times 21.1111 + (0.3833) \times 100 + (0.1) \times 100 + (0.5) \times 2 \\
&= 0.7044 \times (100) \\
\#(2 < X < 8) &\approx 42
\end{aligned}$$

E.3 Probabilità

E.3.1 Tavole della somma

```
# Somma di due dadi
c1 <- 6
c2 <- 6
re1 <- (two_way(S_1 = 1:c1,S_2 = 1:c2,
                num1 = rep(1,times=c1),num2 = rep(1,times=c2),
                size = "\\footnotesize"))
```

	1;	1/6	2;	1/6	3;	1/6	4;	1/6	5;	1/6	6;	1/6
1; 1/6	2;	1/36	3;	1/36	4;	1/36	5;	1/36	6;	1/36	7;	1/36
2; 1/6	3;	1/36	4;	1/36	5;	1/36	6;	1/36	7;	1/36	8;	1/36
3; 1/6	4;	1/36	5;	1/36	6;	1/36	7;	1/36	8;	1/36	9;	1/36
4; 1/6	5;	1/36	6;	1/36	7;	1/36	8;	1/36	9;	1/36	10;	1/36
5; 1/6	6;	1/36	7;	1/36	8;	1/36	9;	1/36	10;	1/36	11;	1/36
6; 1/6	7;	1/36	8;	1/36	9;	1/36	10;	1/36	11;	1/36	12;	1/36

E ricaviamo la distribuzione di, X

X	2	3	4	5	6	7	8	9	10	11	12
$P(X)$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

Calcoliamo valore atteso e varianza

$$
\begin{aligned}
\mu &= E(X_i) = \sum_{x \in S_X} xP(X=x) \\
&= 2\frac{1}{36} + 3\frac{2}{36} + 4\frac{3}{36} + 5\frac{4}{36} + 6\frac{5}{36} + 7\frac{6}{36} + 8\frac{5}{36} + 9\frac{4}{36} + 10\frac{3}{36} + 11\frac{2}{36} + 12\frac{1}{36} \\
&= 7 \\
\sigma^2 &= V(X_i) = \sum_{x \in S_X} x^2 P(X=x) - \mu^2 \\
&= \left(2^2\frac{1}{36} + 3^2\frac{2}{36} + 4^2\frac{3}{36} + 5^2\frac{4}{36} + 6^2\frac{5}{36} + 7^2\frac{6}{36} + 8^2\frac{5}{36} + 9^2\frac{4}{36} + 10^2\frac{3}{36} + 11^2\frac{2}{36} + 12^2\frac{1}{36}\right) - (7)^2 \\
&= 5.833
\end{aligned}
$$

```r
# Differenza di due dadi
res<-two_way(S_1 = 1:c1,S_2 = 1:c2,size = "\\footnotesize",
             num1 = numeric(c1)+1,num2 = numeric(c2)+1,op = `-`)
```

	1;	1/6	2;	1/6	3;	1/6	4;	1/6	5;	1/6	6;	1/6
1; 1/6	0;	1/36	−1;	1/36	−2;	1/36	−3;	1/36	−4;	1/36	−5;	1/36
2; 1/6	1;	1/36	0;	1/36	−1;	1/36	−2;	1/36	−3;	1/36	−4;	1/36
3; 1/6	2;	1/36	1;	1/36	0;	1/36	−1;	1/36	−2;	1/36	−3;	1/36
4; 1/6	3;	1/36	2;	1/36	1;	1/36	0;	1/36	−1;	1/36	−2;	1/36
5; 1/6	4;	1/36	3;	1/36	2;	1/36	1;	1/36	0;	1/36	−1;	1/36
6; 1/6	5;	1/36	4;	1/36	3;	1/36	2;	1/36	1;	1/36	0;	1/36

E ricaviamo la distribuzione di, X

X	−5	−4	−3	−2	−1	0	1	2	3	4	5
$P(X)$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

Calcoliamo valore atteso e varianza

$$
\begin{aligned}
\mu &= E(X_i) = \sum_{x \in S_X} xP(X=x) \\
&= (-5)\frac{1}{36} + (-4)\frac{2}{36} + (-3)\frac{3}{36} + (-2)\frac{4}{36} + (-1)\frac{5}{36} + 0\frac{6}{36} + 1\frac{5}{36} + 2\frac{4}{36} + 3\frac{3}{36} + 4\frac{2}{36} + 5\frac{1}{36} \\
&= 0 \\
\sigma^2 &= V(X_i) = \sum_{x \in S_X} x^2 P(X=x) - \mu^2 \\
&= \left((-5)^2\frac{1}{36} + (-4)^2\frac{2}{36} + (-3)^2\frac{3}{36} + (-2)^2\frac{4}{36} + (-1)^2\frac{5}{36} + 0^2\frac{6}{36} + 1^2\frac{5}{36} + 2^2\frac{4}{36} + 3^2\frac{3}{36} + 4^2\frac{2}{36} + 5^2\frac{1}{36}\right) - (0)^2 \\
&= 5.833
\end{aligned}
$$

```r
res[[1]]
```

[1] -5 -4 -3 -2 -1 0 1 2 3 4 5

```r
names(res)
```

[1] "S_3" "num3" "den3" "urn"

E.3.2 Binomiale

```r
bin_dis(x1 = 2,n = 5,pp = 0.34)
```

$$\begin{aligned}
P(X \leq 2) &= \binom{5}{0}0.34^0(1-0.34)^{5-0} + \binom{5}{1}0.34^1(1-0.34)^{5-1} + \binom{5}{2}0.34^2(1-0.34)^{5-2} \\
&= 0.1252 + 0.3226 + 0.3323 \\
&= 0.7801
\end{aligned}$$

```r
bin_dis(x1 = 4,n = 5,pp = 0.34,verso = "\\geq")
```

$$\begin{aligned}
P(X \geq 4) &= \binom{5}{4}0.34^4(1-0.34)^{5-4} + \binom{5}{5}0.34^5(1-0.34)^{5-5} \\
&= 0.0441 + 0.0045 \\
&= 0.0486
\end{aligned}$$

```r
bin_dis(x1 = 2,n = 5,pp = 0.34,comp = T)
```

$$\begin{aligned}
P(X \leq 2) &= 1 - P(X > 2) \\
&= 1 - \left(\binom{5}{3}0.34^3(1-0.34)^{5-3} + \binom{5}{4}0.34^4(1-0.34)^{5-4} + \binom{5}{5}0.34^5(1-0.34)^{5-5}\right) \\
&= 1 - (0.1712 + 0.0441 + 0.0045) \\
&= 1 - 0.2198 \\
&= 0.7802
\end{aligned}$$

```r
bin_dis(x1 = 2,n = 5,pp = 0.34,sing = T)
```

$$\begin{aligned}
P(X = 2) &= \binom{5}{2}0.34^2(1-0.34)^{5-2} \\
&= 10 \times 0.34^2(1-0.34)^3 \\
&= 0.3323
\end{aligned}$$

E.3.3 Poisson

```
pois_dis(x1 = 2,ll = 1.5)
```

$$
\begin{aligned}
P(X \leq 2) &= \frac{1.5^0}{0!}e^{-1.5} + \frac{1.5^1}{1!}e^{-1.5} + \frac{1.5^2}{2!}e^{-1.5} \\
&= 0.2231 + 0.3347 + 0.251 \\
&= 0.8088
\end{aligned}
$$

```
pois_dis(x1 = 2,ll = 1.5,verso = "\\geq")
```

$$
\begin{aligned}
P(X \geq 2) &= 1 - P(X < 2) \\
&= 1 - \left(\frac{1.5^0}{0!}e^{-1.5} + \frac{1.5^1}{1!}e^{-1.5}\right) \\
&= 1 - (0.2231 + 0.3347) \\
&= 1 - 0.5578 \\
&= 0.4422
\end{aligned}
$$

```
pois_dis(x1 = 2,ll = 1.5,sing = T)
```

$$
\begin{aligned}
P(X = 2) &= \frac{1.5^2}{2!}e^{-1.5} \\
&= 1.125 \times 0.2231 \\
&= 0.251
\end{aligned}
$$

E.3.4 Normale

```
norm_int(x1 = 1,verso = "<",mm = 3,ss = 2.2,vnam = "\\theta",
         mu = "\\mu_\\theta",sigma = "\\sigma_\\theta")
```

$$
\begin{aligned}
P(\theta < 1) &= P\left(\frac{\theta - \mu_\theta}{\sigma_\theta} < \frac{1-3}{\sqrt{2.2}}\right) \\
&= P\left(Z < -1.35\right) \\
&= 1 - \Phi(1.35) \\
&= 0.0885
\end{aligned}
$$

```r
norm_int(x1 = 4,verso = "<",mm = 3,ss = 2.2,vnam = "X",
            mu = "\\psi",sigma = "\\tau")
```

$$
\begin{aligned}
P(X < 4) &= P\left(\frac{X-\psi}{\tau} < \frac{4-3}{\sqrt{2.2}}\right) \\
&= P\left(Z < 0.67\right) \\
&= \Phi(0.67) \\
&= 0.7486
\end{aligned}
$$

```r
norm_int(x1 = 1,verso = ">",mm = 3,ss = 2.2,vnam = "Y",
            mu = "\\mu_Y",sigma = "\\sigma_Y")
```

$$
\begin{aligned}
P(Y > 1) &= P\left(\frac{Y-\mu_Y}{\sigma_Y} > \frac{1-3}{\sqrt{2.2}}\right) \\
&= P\left(Z > -1.35\right) \\
&= 1 - P(Z < -1.35) \\
&= 1 - (1 - \Phi(1.35)) \\
&= 0.9115
\end{aligned}
$$

```r
norm_int(x1 = 4,verso = ">",mm = 3,ss = 2.2,)
```

$$
\begin{aligned}
P(X > 4) &= P\left(\frac{X-\mu}{\sigma} > \frac{4-3}{\sqrt{2.2}}\right) \\
&= P\left(Z > 0.67\right) \\
&= 1 - P(Z < 0.67) \\
&= 1 - \Phi(0.67) \\
&= 0.2514
\end{aligned}
$$

```r
norm_int(x1 = 1,verso = ">",mm = -3,ss = 2.2)
```

$$
\begin{aligned}
P(X > 1) &= P\left(\frac{X-\mu}{\sigma} > \frac{1-(-3)}{\sqrt{2.2}}\right) \\
&= P\left(Z > 2.7\right) \\
&= 1 - P(Z < 2.7) \\
&= 1 - \Phi(2.7) \\
&= 0.0035
\end{aligned}
$$

```
norm_int(x1 = 1,x2=2,mm = 3,ss = 2.2,verso = NULL)
```

$$
\begin{aligned}
P(1 < X \leq 2) &= P\left(\frac{1-3}{\sqrt{2.2}} < \frac{X-\mu}{\sigma} \leq \frac{2-3}{\sqrt{2.2}}\right) \\
&= P\left(-1.35 < Z \leq -0.67\right) \\
&= \Phi(-0.67) - \Phi(-1.35) \\
&= (1 - \Phi(0.67)) - (1 - \Phi(1.35)) \\
&= (1 - 0.7486) - (1 - 0.9115) \\
&= 0.1629
\end{aligned}
$$

```
norm_int(x1 = 1,x2=2,mm = -3,ss = 2.2,verso = NULL)
```

$$
\begin{aligned}
P(1 < X \leq 2) &= P\left(\frac{1-(-3)}{\sqrt{2.2}} < \frac{X-\mu}{\sigma} \leq \frac{2-(-3)}{\sqrt{2.2}}\right) \\
&= P\left(2.7 < Z \leq 3.37\right) \\
&= \Phi(3.37) - \Phi(2.7) \\
&= 0.9996 - 0.9965 \\
&= 0.0031
\end{aligned}
$$

```
norm_int(x1 = -1,x2=2,mm = -3,ss = 2.2,verso = NULL)
```

$$
\begin{aligned}
P(-1 < X \leq 2) &= P\left(\frac{-1-(-3)}{\sqrt{2.2}} < \frac{X-\mu}{\sigma} \leq \frac{2-(-3)}{\sqrt{2.2}}\right) \\
&= P\left(1.35 < Z \leq 3.37\right) \\
&= \Phi(3.37) - \Phi(1.35) \\
&= 0.9996 - 0.9115 \\
&= 0.0881
\end{aligned}
$$

E.3.5 TLC

```
tlc(tipo = "somma",x1 = 90,x2 = 110,verso = NULL,mu = 1,s2 = 1,n = 100)
```

Teorema del Limite Centrale (somma VC qualunque)

Siano X_1,...,X_n, $n = 100$ VC IID, tc $E(X_i) = \mu = 1$ e $V(X_i) = \sigma^2 = 1, \forall i$, posto:

$$S_n = X_1 + ... + X_n$$

allora:

$$\begin{aligned} S_n &\underset{a}{\sim} N(n\mu, n\sigma^2) \\ &\sim N(100 \cdot 1, 100 \cdot 1) \\ &\sim N(100, 100) \end{aligned}$$

$$\begin{aligned} P(90 < S_n \leq 110) &= P\left(\frac{90-100}{\sqrt{100}} < \frac{S_n - n\mu}{\sqrt{n\sigma^2}} \leq \frac{110-100}{\sqrt{100}}\right) \\ &= P(-1 < Z \leq 1) \\ &= \Phi(1) - \Phi(-1) \\ &= \Phi(1) - (1 - \Phi(1)) \\ &= 0.8413 - (1 - 0.8413) \\ &= 0.6826 \end{aligned}$$

```
tlc(tipo = "media",x1 = 9,x2 = 11,verso = NULL,mu = 10,s2 = 1,n = 100)
```

Teorema del Limite Centrale (media VC qualunque)

Siano X_1,...,X_n, $n = 100$ VC IID, tc $E(X_i) = \mu = 10$ e $V(X_i) = \sigma^2 = 1, \forall i$, posto:

$$\bar{X} = \frac{S_n}{n} = \frac{X_1 + \ldots + X_n}{n}$$

allora:

$$\begin{aligned} \bar{X} &\underset{a}{\sim} N(\mu, \sigma^2/n) \\ &\sim N\left(10, \frac{1}{100}\right) \\ &\sim N(10, 0.01) \end{aligned}$$

$$\begin{aligned} P(9 < \bar{X} \leq 11) &= P\left(\frac{9-10}{\sqrt{0.01}} < \frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}} \leq \frac{11-10}{\sqrt{0.01}}\right) \\ &= P(-10 < Z \leq 10) \\ &= \Phi(10) - \Phi(-10) \\ &= \Phi(10) - (1 - \Phi(10)) \\ &= 1 - (1 - 1) \\ &= 1 \end{aligned}$$

```
tlc(tipo = "prop",x1 = .1,verso = ">",mu = .2,n = 50)
```

Teorema del Limite Centrale (proporzione)

Siano X_1,...,X_n, $n = 50$ VC IID, tc $X_i \sim \text{Ber}(\pi = 0.2), \forall i$, posto:

$$\hat{\pi} = \frac{S_n}{n} = \frac{X_1 + ... + X_n}{n}$$

allora:

$$\begin{aligned}
\hat{\pi} &\underset{a}{\sim} N(\pi, \pi(1-\pi)/n) \\
&\sim N\left(0.2, \frac{0.2 \cdot (1-0.2)}{50}\right) \\
&\sim N(0.2, 0.0032)
\end{aligned}$$

$$\begin{aligned}
P(\hat{\pi} > 0.1) &= P\left(\frac{\hat{\pi} - \pi}{\sqrt{\pi(1-\pi)/n}} > \frac{0.1 - 0.2}{\sqrt{0.0032}}\right) \\
&= P\left(Z > -1.77\right) \\
&= 1 - P(Z < -1.77) \\
&= 1 - (1 - \Phi(1.77)) \\
&= 0.9616
\end{aligned}$$

E.4 Inferenza

E.4.1 Intervalli di Confidenza

```
idc(xm = 10,sd = 1.1,alpha = .05,n = 15,dist_ = "z")
```

$1 - \alpha = 0.95$ e quindi $\alpha = 0.05 \to \alpha/2 = 0.025$

$$\begin{aligned}
Idc: \quad & \hat{\mu} \pm z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}} \\
& 10 \pm 1.96 \times \frac{1.1}{\sqrt{15}} \\
& 10 \pm 1.96 \times 0.284 \\
& [9.443, 10.56]
\end{aligned}$$

```
idc(xm = 10,sd = 1.1,alpha = .05,n = 15,dist_ = "t")
```

$1 - \alpha = 0.95$ e quindi $\alpha = 0.05 \to \alpha/2 = 0.025$

$$S = \sqrt{\frac{n}{n-1}} \cdot \hat{\sigma} = \sqrt{\frac{15}{14}} \cdot 1.1 = 1.1386$$

$$
\begin{aligned}
Idc: \qquad & \hat{\mu} \pm t_{n-1;\alpha/2} \times \frac{S}{\sqrt{n}} \\
& 10 \pm 2.145 \times \frac{1.1386}{\sqrt{15}} \\
& 10 \pm 2.145 \times 0.294 \\
& [9.369, 10.63]
\end{aligned}
$$

```
idc(xm = 10,alpha = .05,n = 15,dist_ = "z")
```

$1-\alpha = 0.95$ e quindi $\alpha = 0.05 \rightarrow \alpha/2 = 0.025$

$$
\hat{\pi} = \frac{S_n}{n} = \frac{10}{15} = 0.6667
$$

$$
\begin{aligned}
Idc: \qquad & \hat{\pi} \pm z_{\alpha/2} \times \sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}} \\
& 0.6667 \pm 1.96 \times \sqrt{\frac{0.6667(1-0.6667)}{15}} \\
& 0.6667 \pm 1.96 \times 0.1217 \\
& [0.4281, 0.9052]
\end{aligned}
$$

```
idc(xm = 7.4,sd = sqrt(7.4),alpha = .05,n = 75,dist_ = "z",mus = "\\lambda",
        ss = "\\sqrt\\lambda")
```

$1-\alpha = 0.95$ e quindi $\alpha = 0.05 \rightarrow \alpha/2 = 0.025$

$$
\begin{aligned}
Idc: \qquad & \lambda \pm z_{\alpha/2} \times \frac{\sqrt{\lambda}}{\sqrt{n}} \\
& 7.4 \pm 1.96 \times \frac{2.72029410174709}{\sqrt{75}} \\
& 7.4 \pm 1.96 \times 0.3141 \\
& [6.784, 8.016]
\end{aligned}
$$

E.4.2 Test

```
ztest_mu(muh = 0,s = 1,10,mu0 = 1,h1 = "\\neq",alpha = 0.05)
```

Test Z per una media, variazna nota

$\boxed{\text{A}}$ FORMULAZIONE DELLE IPOTESI

$$\begin{cases} H_0 : \mu = \mu_0 = 1 \\ H_1 : \mu \neq \mu_0 = 1 \end{cases}$$

$\boxed{\text{B}}$ SCELTA E CALCOLO STATISTICA-TEST, Z

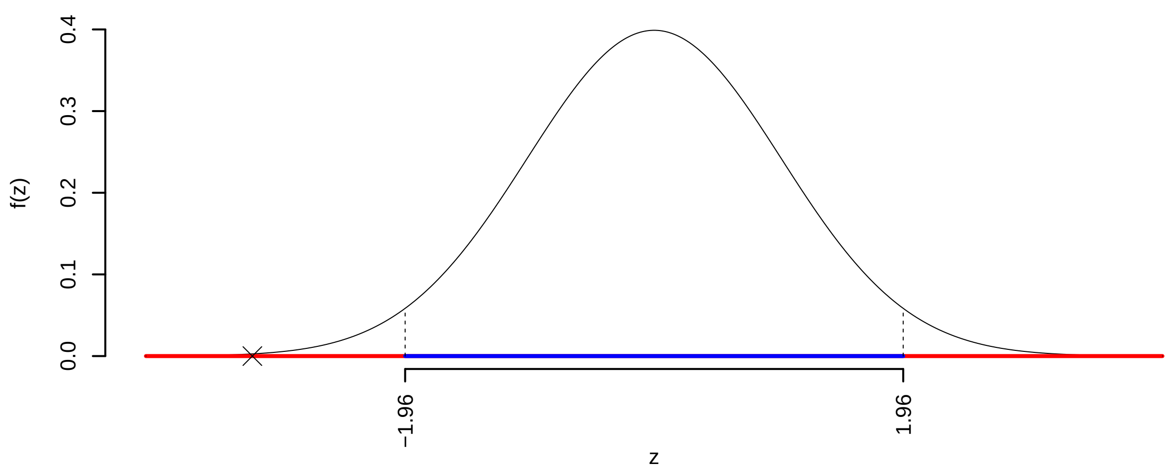
σ^2 di P è nota: $\Rightarrow$ z-Test.

$$\begin{aligned} \frac{\hat{\mu} - \mu_0}{\sigma/\sqrt{n}} &\sim N(0,1) \\ z_{\text{obs}} &= \frac{(0-1)}{1/\sqrt{10}} = -3.162\,. \end{aligned}$$

$\boxed{\text{C}}$ CONCLUSIONE

La siginficatitività è $\alpha = 0.05$, dalle tavole osserviamo $z_{0.025} = 1.96$.

Essendo $|z_{\text{obs}}| = 3.1623 > z_{0.025} = 1.96$ allora **rifiuto** H_0 al 5%.



Il p_{value} è

$$p_{\text{value}} = P(|Z| > |-3.16|) = 2P(Z > 3.16) = 0.001565$$

$$0.001 < p_{\text{value}} = 0.001565 \leq 0.01$$

```
ztest_mu(muh = 0,s = 1,10,mu0 = 1,h1 = "\\neq")
```

Test Z per una media, variazna nota

$\boxed{\text{A}}$ FORMULAZIONE DELLE IPOTESI

$$\begin{cases} H_0 : \mu = \mu_0 = 1 \\ H_1 : \mu \neq \mu_0 = 1 \end{cases}$$

$\boxed{\text{B}}$ SCELTA E CALCOLO STATISTICA-TEST, Z

σ^2 di P è nota: $\Rightarrow$ z-Test.

$$\begin{aligned} \frac{\hat{\mu} - \mu_0}{\sigma/\sqrt{n}} &\sim N(0,1) \\ z_{\text{obs}} &= \frac{(0-1)}{1/\sqrt{10}} = -3.162\,. \end{aligned}$$

$\boxed{\text{C}}$ CONCLUSIONE

Siccome H_1 è bilaterale, considereremo $\alpha/2$, anziché α

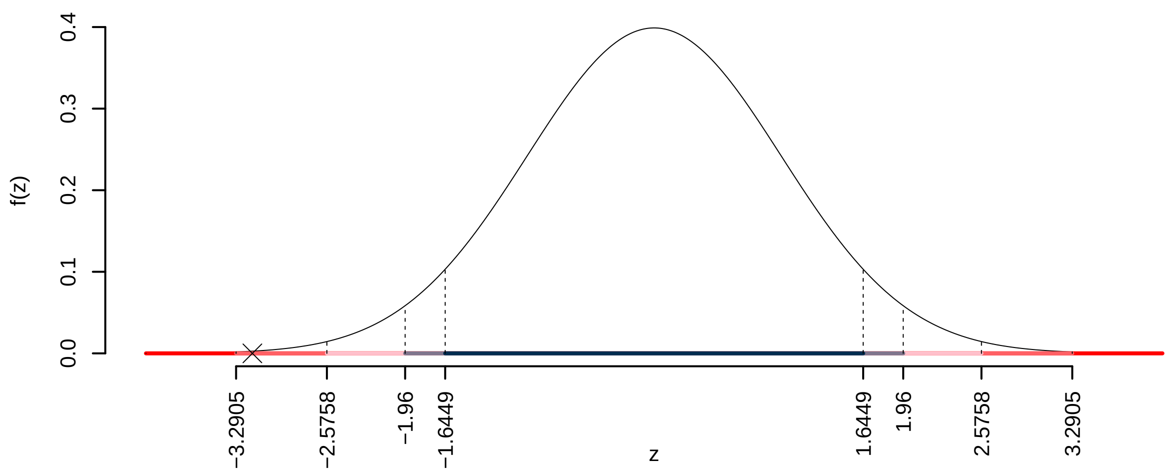
$\alpha = 0.1, 0.05, 0.01, 0.001$ e quindi $\alpha/2 = 0.05, 0.025, 0.005, 0.0005$

I valori critici sono

$z_{0.05} = 1.6449$; $z_{0.025} = 1.96$; $z_{0.005} = 2.5758$; $z_{0.0005} = 3.2905$

Siccome $2.5758 < |z_{\text{obs}}| = 3.1623 < 3.2905$, quindi **rifiuto** H_0 all'1%,

$0.001 < p_{\text{value}} < 0.01$, *molto significativo* $\boxed{**}$.



Il p_{value} è

$$p_{\text{value}} = P(|Z| > |-3.16|) = 2P(Z > 3.16) = 0.001565$$

$$0.001 < p_{\text{value}} = 0.001565 \leq 0.01$$

```
ztest_mu(muh = 0,s = 1,10,mu0 = 1,h1 = "\\neq",pv_only = T)
```

Test Z per una media, variazna nota

$\boxed{\text{A}}$ FORMULAZIONE DELLE IPOTESI

$$\begin{cases} H_0 : \mu = \mu_0 = 1 \\ H_1 : \mu \neq \mu_0 = 1 \end{cases}$$

$\boxed{\text{B}}$ SCELTA E CALCOLO STATISTICA-TEST, Z

σ^2 di P è nota: $\Rightarrow$ z-Test.

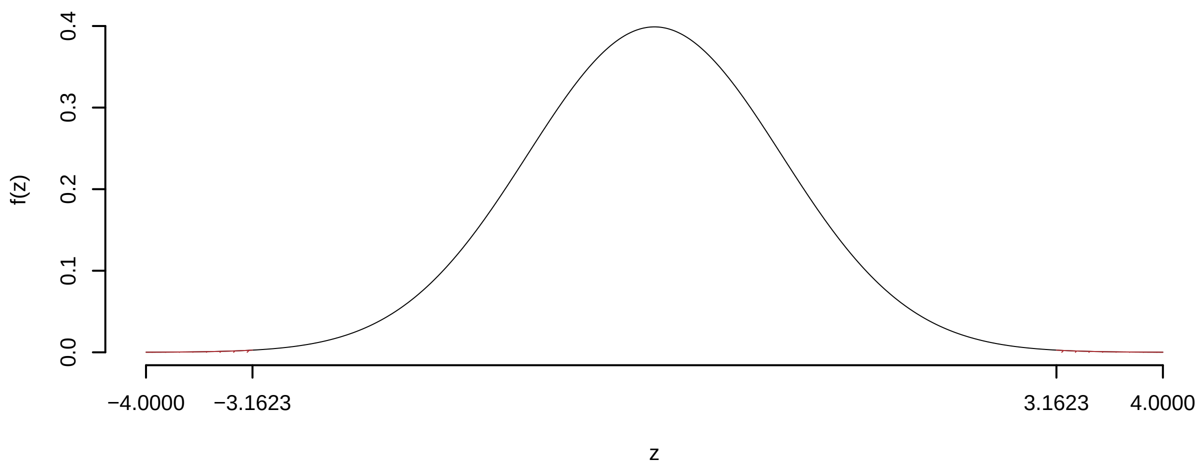
$$\begin{aligned} \frac{\hat{\mu} - \mu_0}{\sigma/\sqrt{n}} &\sim N(0,1) \\ z_{\text{obs}} &= \frac{(0-1)}{1/\sqrt{10}} = -3.162\,. \end{aligned}$$

$\boxed{\text{C}}$ CONCLUSIONE

Il p_{value} è

$$p_{\text{value}} = P(|Z| > |-3.16|) = 2P(Z > 3.16) = 0.001565$$

$$0.001 < p_{\text{value}} = 0.001565 \le 0.01$$



Rifiuto H_0 all'1%,

$0.001 < p_{\text{value}} < 0.01$, *molto significativo* $\boxed{**}$.

```
ttest_mu(muh = 0,sh = 1,n = 10,mu0 = 1,h1 = "<",alpha = 0.01)
```

Test t per una media, varianza incognita

$\boxed{\text{A}}$ FORMULAZIONE DELLE IPOTESI

$$\begin{cases} H_0 : \mu = \mu_0 = 1 \\ H_1 : \mu < \mu_0 = 1 \end{cases}$$

$$S = \sqrt{\frac{n}{n-1}}\,\hat{\sigma} = \sqrt{\frac{10}{10-1}} \times 1 = 1.054$$

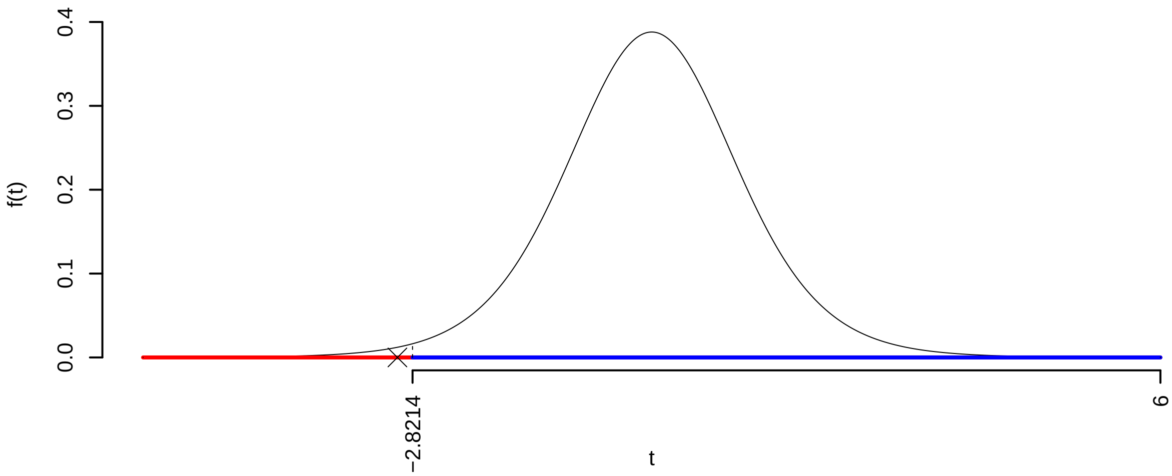
$$\frac{\hat{\mu} - \mu_0}{S/\sqrt{n}} \sim t_{n-1}$$

$$t_{\text{obs}} = \frac{(0-1)}{1.054/\sqrt{10}} = -3\,.$$

$\boxed{\text{C}}$ CONCLUSIONE

La siginficatitività è $\alpha = 0.01$, dalle tavole osserviamo $t_{10-1;0.01} = -2.8214$.

Essendo $t_{\text{obs}} = -3 < t_{10-1;0.01} = -2.8214$ allora **rifiuto** H_0 al 1%.



Il p_{value} è

$$p_{\text{value}} = P(T_{10-1} < -3) = 0.007478$$

Attenzione il calcolo del p_{value} con la T è puramente illustrativo e non può essere riprodotto senza una calcolatrice statistica adeguata.

$$0.001 < p_{\text{value}} = 0.007478 \leq 0.01$$

```
ttest_mu(muh = 0,sh = 1,n = 10,mu0 = 1,h1 = "<")
```

Test t per una media, varianza incognita

$\boxed{\text{A}}$ FORMULAZIONE DELLE IPOTESI

$$\begin{cases} H_0 : \mu = \mu_0 = 1 \\ H_1 : \mu < \mu_0 = 1 \end{cases}$$

$$S = \sqrt{\frac{n}{n-1}}\,\hat{\sigma} = \sqrt{\frac{10}{10-1}} \times 1 = 1.054$$

$$
\begin{aligned}
\frac{\hat{\mu}-\mu_0}{S/\sqrt{n}} &\sim t_{n-1} \\
t_{\text{obs}} &= \frac{(0-1)}{1.054/\sqrt{10}} = -3\,.
\end{aligned}
$$

$\boxed{\text{C}}$ CONCLUSIONE

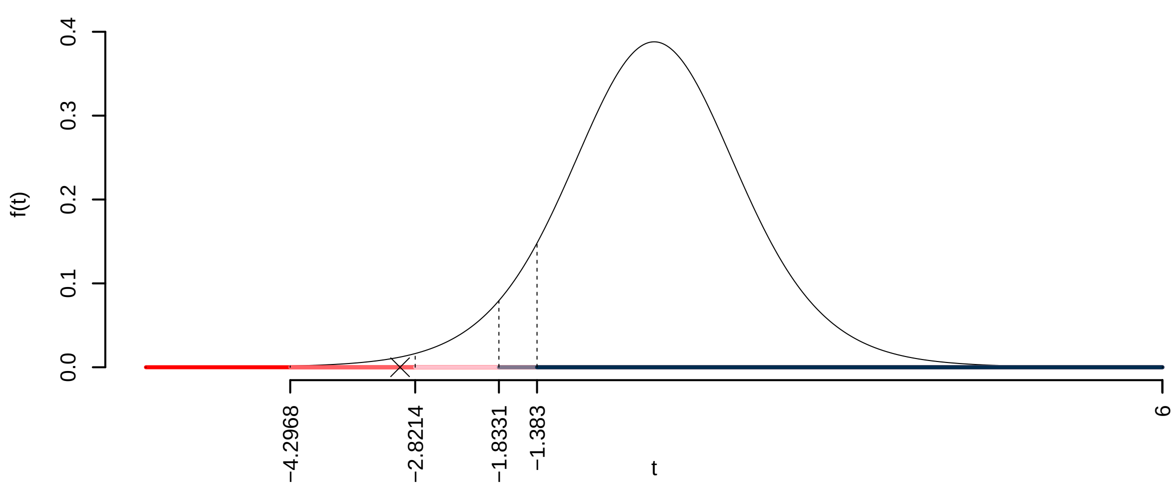
Consideriamo $\alpha = 0.1, 0.05, 0.01, 0.001$

I valori critici sono

$t_{10-1;0.1} = -1.383$; $t_{10-1;0.05} = -1.8331$; $t_{10-1;0.01} = -2.8214$; $t_{10-1;0.001} = -4.2968$

Siccome $-1.8331 < t_{\text{obs}} = -3 < -1.383$, quindi **rifiuto** H_0 all'1%,

$0.001 < p_{\text{value}} < 0.01$, *molto significativo* $\boxed{**}$.



Il p_{value} è

$$p_{\text{value}} = P(T_{10-1} < -3) = 0.007478$$

Attenzione il calcolo del p_{value} con la T è puramente illustrativo e non può essere riprodotto senza una calcolatrice statistica adeguata.

$$0.001 < p_{\text{value}} = 0.007478 \leq 0.01$$

```
ttest_mu(muh = 0,sh = 1,n = 10,mu0 = 1,h1 = "<",pv_only = T,um="cm")
```

Test t per una media, varianza incognita

$\boxed{\text{A}}$ FORMULAZIONE DELLE IPOTESI

$$\begin{cases} H_0 : \mu = \mu_0 = 1cm \\ H_1 : \mu < \mu_0 = 1cm \end{cases}$$

$$S = \sqrt{\frac{n}{n-1}}\,\hat{\sigma} = \sqrt{\frac{10}{10-1}} \times 1 = 1.054$$

$$\frac{\hat{\mu}-\mu_0}{S/\sqrt{n}} \sim t_{n-1}$$

$$t_{\text{obs}} = \frac{(0-1)}{1.054/\sqrt{10}} = -3\,.$$

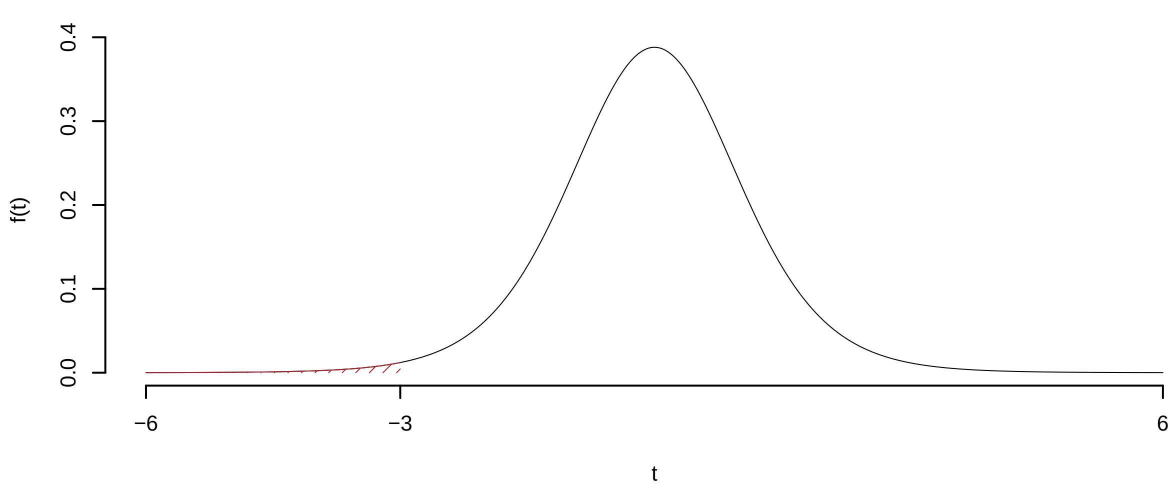
$\boxed{\text{C}}$ CONCLUSIONE

Il p_{value} è

$$p_{\text{value}} = P(T_{10-1} < -3) = 0.007478$$

Attenzione il calcolo del p_{value} con la T è puramente illustrativo e non può essere riprodotto senza una calcolatrice statistica adeguata.

$$0.001 < p_{\text{value}} = 0.007478 \leq 0.01$$



Rifiuto H_0 all'1%,

$0.001 < p_{\text{value}} < 0.01$, *molto significativo* $\boxed{**}$.

```
ztest_pi(sn = 60,n = 100,p0 = .5,h1 = ">",alpha = 0.05)
```

Test Z per una proporzione

La stima

$$\hat{\pi} = \frac{60}{100} = 0.6$$

$\boxed{\text{A}}$ FORMULAZIONE DELLE IPOTESI

$$\begin{cases} H_0 : \pi = \pi_0 = 0.5 \\ H_1 : \pi > \pi_0 = 0.5 \end{cases}$$

$\boxed{\text{B}}$ SCELTA E CALCOLO STATISTICA-TEST, Z Test Binomiale per n grande: $\Rightarrow$ z-Test.

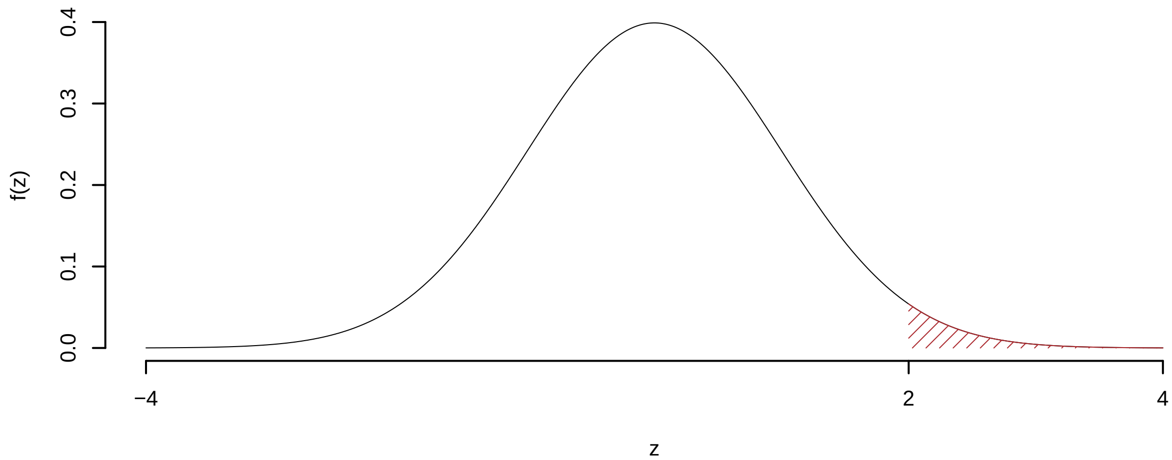
$$\begin{aligned} \frac{\hat{\pi} - \pi_0}{\sqrt{\pi_0(1 - \pi_0)/\,n}} &\sim N(0, 1) \\ z_{\text{obs}} &= \frac{(0.6 - 0.5)}{\sqrt{0.5(1 - 0.5)/100}} = 2\,. \end{aligned}$$

$\boxed{\text{C}}$ CONCLUSIONE

Il p_{value} è

$$p_{\text{value}} = P(Z > 2) = 0.022750$$

$$0.01 < p_{\text{value}} = 0.022750 \leq 0.05$$



Indecisione sul rifiuto di H_0 al 10%,

$0.05 < p_{\text{value}} < 0.1$, *marginalmente significativo* $\boxed{\cdot}$.

```
test_2c(mu1 = 11,mu2 = 12,s1h = 1.1,s2h = 1.2,n1 = 10,n2 = 12,
            h1 = "\\neq",et = T)
```

Test t per due medie, (eterogeneità)

$\boxed{\text{A}}$ FORMULAZIONE DELLE IPOTESI

$$\begin{cases} H_0 : \mu_{\text{A}} = \mu_{\text{B}} \\ H_1 : \mu_{\text{A}} \neq \mu_{\text{B}} \end{cases}$$

$\boxed{\text{B}}$ SCELTA E CALCOLO STATISTICA-TEST, T

$$S^2_{\text{A}} = \frac{n_{\text{A}}}{n_{\text{A}} - 1}\hat{\sigma}^2_{\text{A}} = \frac{10}{10-1}1.1^2 = 1.344 \qquad S^2_{\text{B}} = \frac{n_{\text{B}}}{n_{\text{B}} - 1}\hat{\sigma}^2_{\text{B}} = \frac{12}{12-1}1.2^2 = 1.571$$

$$\begin{aligned} \frac{\hat{\mu}_{\text{A}} - \hat{\mu}_{\text{B}}}{\sqrt{\frac{S^2_{\text{A}}}{n_{\text{A}}} + \frac{S^2_{\text{B}}}{n_{\text{B}}}}} &\sim t_{n_{\text{A}} + n_{\text{B}} - 2} \\ t_{\text{obs}} &= \frac{(11-12)}{\sqrt{\frac{1.344}{10} + \frac{1.571}{12}}} = -1.941\,. \end{aligned}$$

$\boxed{\text{C}}$ CONCLUSIONE

Siccome H_1 è bilaterale, considereremo $\alpha/2$, anziché α

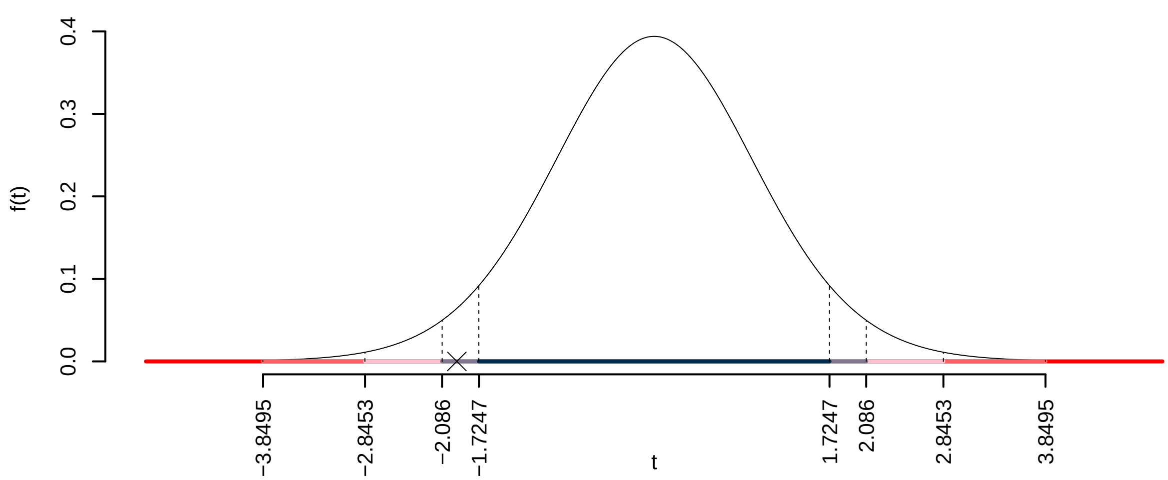
$\alpha = 0.1, 0.05, 0.01, 0.001$ e quindi $\alpha/2 = 0.05, 0.025, 0.005, 0.0005$

I valori critici sono

$t_{22-2;0.05} = 1.7247$; $t_{22-2;0.025} = 2.086$; $t_{22-2;0.005} = 2.8453$; $t_{22-2;0.0005} = 3.8495$

Siccome $1.7247 < |t_{\text{obs}}| = 1.9413 < 2.086$, indecisione sul rifiuto di H_0 al 10%,

$0.05 < p_{\text{value}} < 0.1$, *marginalmente significativo* $\boxed{\cdot}$.



Il p_{value} è

$$p_{\text{value}} = P(|T_{22-2}| > |-1.94|) = 2P(T_{22-2} > 1.94) = 0.066448$$

Attenzione il calcolo del p_{value} con la T è puramente illustrativo e non può essere riprodotto senza una calcolatrice statistica adeguata.

$$0.05 < p_{\text{value}} = 0.066448 \leq 0.1$$

```
test_2c(mu1 = 11,mu2 = 12,s1h = 1.1,s2h = 1.2,n1 = 10,n2 = 12,
            h1 = "\\neq",alpha = .05,et = F)
```

Test T per due medie, (omogeneità)

$\boxed{\text{A}}$ FORMULAZIONE DELLE IPOTESI

$$\begin{cases} H_0 : \mu_{\text{A}} = \mu_{\text{B}} \\ H_1 : \mu_{\text{A}} \neq \mu_{\text{B}} \end{cases}$$

$\boxed{\text{B}}$ SCELTA E CALCOLO STATISTICA-TEST, T

L'ipotesi è di omogeneità e quindi calcoliamo:

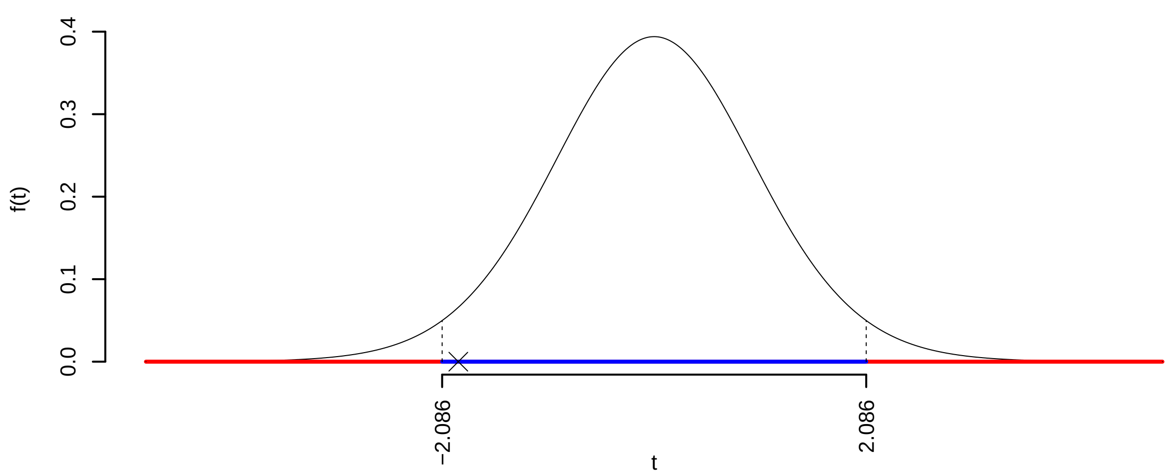
$$S_p^2 = \frac{n_{\text{ A}}\,\hat{\sigma}^2_{\text{A}} + n_{\text{ B}}\,\hat{\sigma}^2_{\text{B}}}{n_{\text{ A}} + n_{\text{ B}} - 2} = \frac{10 \cdot 1.1^2 + 12 \cdot 1.2^2}{10 + 12 - 2} = 1.469$$

$$\begin{aligned}\frac{\hat{\mu}_{\text{ A}} - \hat{\mu}_{\text{ B}}}{\sqrt{\frac{S_p^2}{n_{\text{ A}}} + \frac{S_p^2}{n_{\text{ B}}}}} &\sim t_{n_{\text{ A}} + n_{\text{ B}} - 2} \\ t_{\text{obs}} &= \frac{(11 - 12)}{\sqrt{\frac{1.344}{10} + \frac{1.571}{12}}} = -1.927\,.\end{aligned}$$

$\boxed{\text{C}}$ CONCLUSIONE

La siginficatitività è $\alpha = 0.05$, dalle tavole osserviamo $t_{22-2;0.025} = 2.086$.

Essendo $|t_{\text{obs}}| = 1.9269 < t_{22-2;0.025} = 2.086$ allora **non** rifiuto H_0 al 5%.



Il p_{value} è

$$p_{\text{value}} = P(|T_{22-2}| > |-1.93|) = 2P(T_{22-2} > 1.93) = 0.068315$$

Attenzione il calcolo del p_{value} con la T è puramente illustrativo e non può essere riprodotto senza una calcolatrice statistica adeguata.

$$0.05 < p_{\text{value}} = 0.068315 \leq 0.1$$

```
test_2c(mu1 = 11,mu2 = 12,s1h = F,s2h = NULL,n1 = 50,n2 = 60,
            h1 = "\\neq",alpha = .05,et = T)
```

Test Z per due proporzioni

$\boxed{\text{A}}$ FORMULAZIONE DELLE IPOTESI

$$\begin{cases} H_0 : \pi_{\text{A}} = \pi_{\text{B}} \\ H_1 : \pi_{\text{A}} \neq \pi_{\text{B}} \end{cases}$$

$\boxed{\text{B}}$ SCELTA E CALCOLO STATISTICA-TEST, Z

$$\hat{\pi}_{\text{ A}} = \frac{s_{\text{ A}}}{n_{\text{ A}}} = \frac{11}{50} = 0.22 \qquad \hat{\pi}_{\text{ B}} = \frac{s_{\text{ B}}}{n_{\text{ B}}} = \frac{12}{60} = 0.2$$

Calcoliamo la proporzione comune sotto H_0

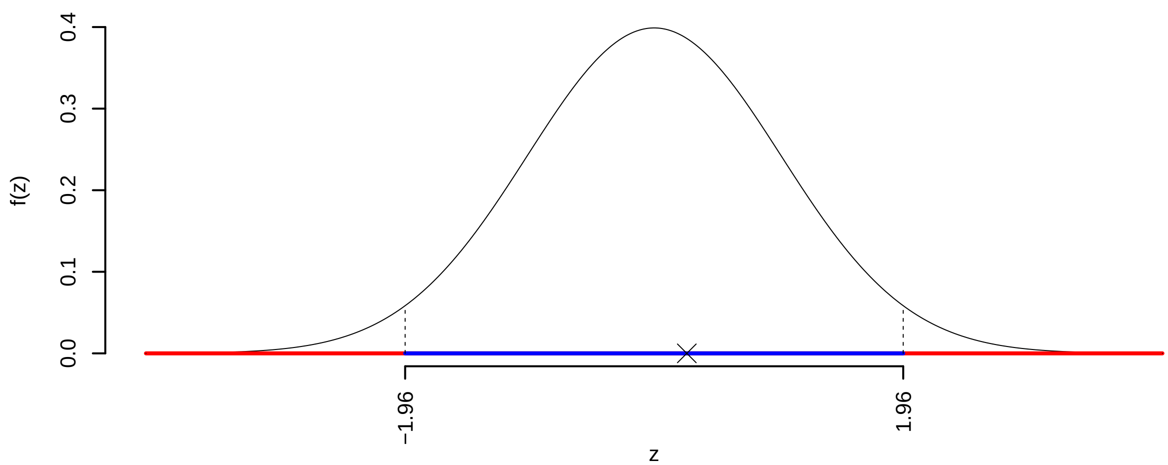
$$\pi_C = \frac{s_{\text{ A}} + s_{\text{ B}}}{n_{\text{ A}} + n_{\text{ B}}} = \frac{23}{110} = 0.2091$$

$$\begin{aligned} \frac{\hat{\pi}_{\text{ A}} - \hat{\pi}_{\text{ B}}}{\sqrt{\frac{\pi_C(1-\pi_C)}{n_{\text{ A}}} + \frac{\pi_C(1-\pi_C)}{n_{\text{ B}}}}} &\sim N(0,1) \\ z_{\text{obs}} &= \frac{(0.22 - 0.2)}{\sqrt{\frac{0.2091(1-0.2091)}{50} + \frac{0.2091(1-0.2091)}{60}}} = 0.2568\,. \end{aligned}$$

$\boxed{\text{C}}$ CONCLUSIONE

La siginficatitività è $\alpha = 0.05$, dalle tavole osserviamo $z_{0.025} = 1.96$.

Essendo $|z_{\text{obs}}| = 0.2568 < z_{0.025} = 1.96$ allora **non** rifiuto H_0 al 5%.



Il p_{value} è

$$p_{\text{value}} = P(|Z| > |0.26|) = 2P(Z > 0.26) = 0.797302$$

$$0.1 < p_{\text{value}} = 0.797302 \leq 1$$

```
ttest_2c_et(mu1 = 11,mu2 = 12,s1h = 1.1,s2h = 1.2,n1 = 10,n2 = 12,
                h1 = "\\neq",alpha = .05)
```

Test t per due medie, (eterogeneità)

$\boxed{\text{A}}$ FORMULAZIONE DELLE IPOTESI

$$\begin{cases} H_0 : \mu_{\text{A}} = \mu_{\text{B}} \\ H_1 : \mu_{\text{A}} \neq \mu_{\text{B}} \end{cases}$$

$\boxed{\text{B}}$ SCELTA E CALCOLO STATISTICA-TEST, T

$$S_1^2 = \frac{n_1}{n_1 - 1}\hat{\sigma}_1^2 = \frac{10}{10-1}1.1^2 = 1.344 \qquad S_2^2 = \frac{n_2}{n_2 - 1}\hat{\sigma}_2^2 = \frac{12}{12-1}1.2^2 = 1.571$$

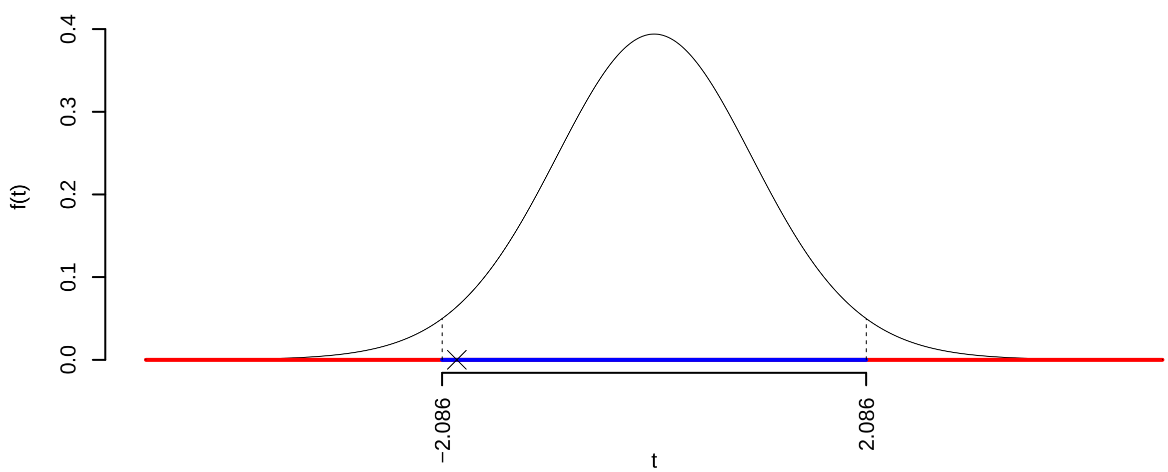
$$\frac{\hat{\mu}_1 - \hat{\mu}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim t_{n_1 + n_2 - 2}$$

$$t_{\text{obs}} = \frac{(11-12)}{\sqrt{\frac{1.344}{10}+\frac{1.571}{12}}} = -1.941\,.$$

$\boxed{\text{C}}$ CONCLUSIONE

La siginficatitività è $\alpha = 0.05$, dalle tavole osserviamo $t_{22-2;0.025} = 2.086$.

Essendo $|t_{\text{obs}}| = 1.9413 < t_{22-2;0.025} = 2.086$ allora **non** rifiuto H_0 al 5%.



Il p_{value} è

$$p_{\text{value}} = P(|T_{22-2}| > |-1.94|) = 2P(T_{22-2} > 1.94) = 0.066448$$

Attenzione il calcolo del p_{value} con la T è puramente illustrativo e non può essere riprodotto senza una calcolatrice statistica adeguata.

$$0.05 < p_{\text{value}} = 0.066448 \leq 0.1$$

```
ttest_2c_om(mu1 = 11,mu2 = 12,s1h = 1.1,s2h = 1.2,n1 = 10,n2 = 12,
                h1 = "\\neq",rbow = T)
```

Test T per due medie, (omogeneità)

$\boxed{\text{A}}$ FORMULAZIONE DELLE IPOTESI

$$\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 \neq \mu_2 \end{cases}$$

$\boxed{\text{B}}$ SCELTA E CALCOLO STATISTICA-TEST, T

L'ipotesi è di omogeneità e quindi calcoliamo:

$$S_p^2 = \frac{n_1 \hat{\sigma}_1^2 + n_2 \hat{\sigma}_2^2}{n_1 + n_2 - 2} = \frac{10 \cdot 1.1^2 + 12 \cdot 1.2^2}{10 + 12 - 2} = 1.469$$

$$\frac{\hat{\mu}_1 - \hat{\mu}_2}{\sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}} \sim t_{n_1 + n_2 - 2}$$

$$t_{\text{obs}} = \frac{(11 - 12)}{\sqrt{\frac{1.344}{10} + \frac{1.571}{12}}} = -1.927\,.$$

$\boxed{\text{C}}$ CONCLUSIONE

Siccome H_1 è bilaterale, considereremo $\alpha/2$, anziché α

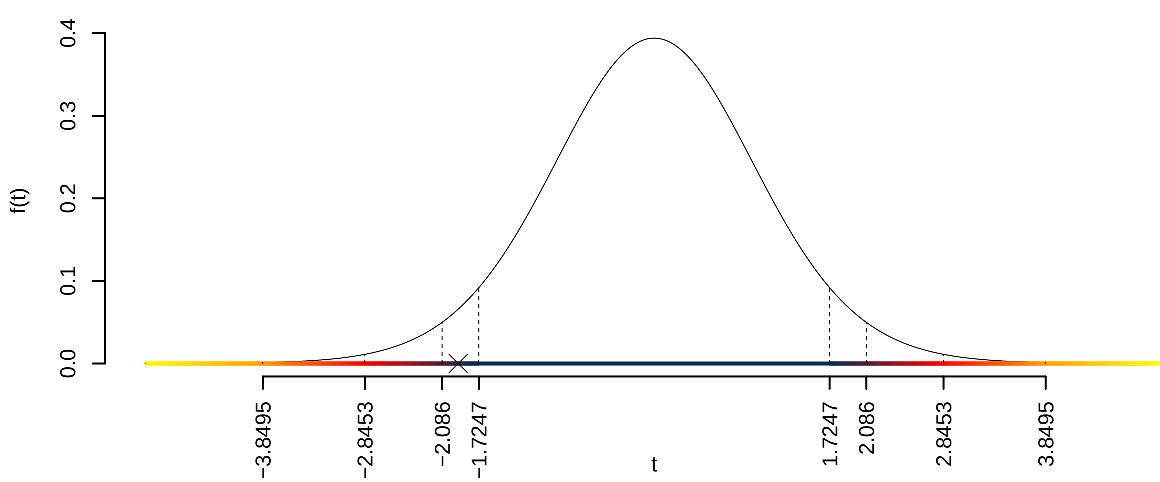
$\alpha = 0.1, 0.05, 0.01, 0.001$ e quindi $\alpha/2 = 0.05, 0.025, 0.005, 0.0005$

I valori critici sono

$t_{22-2;0.05} = 1.7247$; $t_{22-2;0.025} = 2.086$; $t_{22-2;0.005} = 2.8453$; $t_{22-2;0.0005} = 3.8495$

Siccome $1.7247 < |t_{\text{obs}}| = 1.9269 < 2.086$, indecisione sul rifiuto di H_0 al 10%,

$0.05 < p_{\text{value}} < 0.1$, *marginalmente significativo* $\boxed{\cdot}$.



Il p_{value} è

$$p_{\text{value}} = P(|T_{22-2}| > |-1.93|) = 2P(T_{22-2} > 1.93) = 0.068315$$

Attenzione il calcolo del p_{value} con la T è puramente illustrativo e non può essere riprodotto senza una calcolatrice statistica adeguata.

$$0.05 < p_{\text{value}} = 0.068315 \leq 0.1$$

```
ztest_2c_pi(s1 = 120,s2 = 130,n1 = 250,n2 = 260,h1 = "<",alpha = .01)
```

Test Z per due proporzioni

$\boxed{\text{A}}$ FORMULAZIONE DELLE IPOTESI

$$\begin{cases} H_0 : \pi_1 = \pi_2 \\ H_1 : \pi_1 < \pi_2 \end{cases}$$

$\boxed{\text{B}}$ SCELTA E CALCOLO STATISTICA-TEST, Z

$$\hat{\pi}_1 = \frac{s_1}{n_1} = \frac{120}{250} = 0.48 \qquad \hat{\pi}_2 = \frac{s_2}{n_2} = \frac{130}{260} = 0.5$$

Calcoliamo la proporzione comune sotto H_0

$$\pi_C = \frac{s_1 + s_2}{n_1 + n_2} = \frac{250}{510} = 0.4902$$

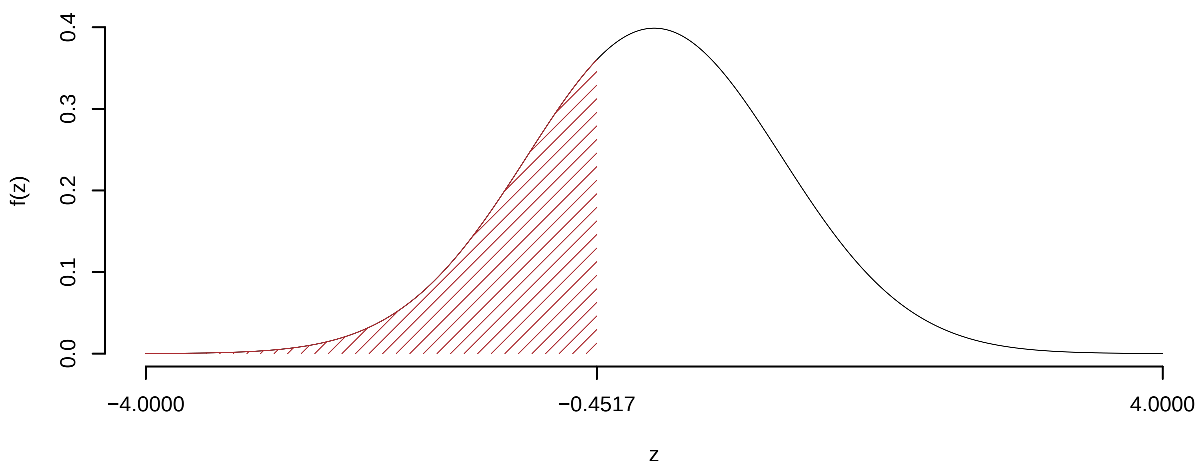
$$\begin{aligned} \frac{\hat{\pi}_1 - \hat{\pi}_2}{\sqrt{\frac{\pi_C(1-\pi_C)}{n_1} + \frac{\pi_C(1-\pi_C)}{n_2}}} &\sim N(0,1) \\ z_{\text{obs}} &= \frac{(0.48 - 0.5)}{\sqrt{\frac{0.4902(1-0.4902)}{250} + \frac{0.4902(1-0.4902)}{260}}} = -0.4517\,. \end{aligned}$$

$\boxed{\text{C}}$ CONCLUSIONE

Il p_{value} è

$$p_{\text{value}} = P(Z < -0.45) = 0.325756$$

$$0.1 < p_{\text{value}} = 0.325756 \leq 1$$



Non rifiuto H_0 a **nessun** livello di significatività,

$p_{\text{value}} > 0.1$, *non significativo*

E.4.3 Regressione

```
set.seed(12)                    # ripete le stesse generazioni casuali
n <- 100                        # fisso n
x <- rnorm(n,10)                # genero x
y <- x+rnorm(n,0,1)             # genero y
ls2e(regr(x = x,y = y))     # produco le statistiche di base

calcolo_beta()
```

$$
\begin{aligned}
\bar{x} &= \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{100}996.8831 = 9.969 \\
\bar{y} &= \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{100}997.8525 = 9.979 \\
\hat{\sigma}_X^2 &= \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2 = \frac{1}{100}10012 - 9.9688^2 = 0.7409 \\
\hat{\sigma}_Y^2 &= \frac{1}{n}\sum_{i=1}^{n} y_i^2 - \bar{y}^2 = \frac{1}{100}10133 - 9.9785^2 = 1.756 \\
\operatorname{cov}(X,Y) &= \frac{1}{n}\sum_{i=1}^{n} x_i\ y_i - \bar{x}\bar{y} = \frac{1}{100}10023 - 9.9688 \cdot 9.9785 = 0.7546
\end{aligned}
$$

$$\begin{aligned}
\hat{\beta}_1 &= \frac{\text{cov}(X,Y)}{\hat{\sigma}_X^2} \\
&= \frac{0.7546}{0.7409} = 1.018 \\
\hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \bar{x} \\
&= 9.979 - 1.0184 \times 9.9688 = -0.1736
\end{aligned}$$

```
calcolo_beta(semplice = T)
```

$$\begin{aligned}
\hat{\beta}_1 &= \frac{\text{cov}(X,Y)}{\hat{\sigma}_X^2} \\
&= \frac{0.7546}{0.7409} = 1.018 \\
\hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \bar{x} \\
&= 9.979 - 1.0184 \times 9.9688 = -0.1736
\end{aligned}$$

```
residuo(x[12],y[12])
```

$$\begin{aligned}
\hat{y}_i &= \hat{\beta}_0 + \hat{\beta}_1 x_i = \\
&= -0.1736 + 1.0184 \times 8.7061 = 8.693 \\
\hat{\varepsilon}_i &= y_i - \hat{y}_i \\
&= 8.178 - 8.693 = -0.5142
\end{aligned}$$

```
se_beta0()
```

$$\begin{aligned}
\hat{\sigma_\varepsilon}^2 &= (1 - r^2)\hat{\sigma}_Y^2 \\
&= (1 - 0.4375) \times 1.756 \\
&= 0.988 \\
S_\varepsilon^2 &= \frac{n}{n-2}\hat{\sigma_\varepsilon}^2 \\
&= \frac{100}{100-2}\hat{\sigma_\varepsilon}^2 \\
&= \frac{100}{100-2} \times 0.988 = 1.008
\end{aligned}$$

E quindi

$$V(\hat{\beta}_0) = \sigma_\varepsilon^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{n\hat{\sigma}_X^2}\right)$$

$$
\begin{aligned}
\widehat{V(\hat\beta_0)} &= S_\varepsilon^2\left(\frac{1}{n}+\frac{\bar x^2}{n\hat\sigma_X^2}\right)\\
&= 1.008\times\left(\frac{1}{100}+\frac{9.969^2}{100\times 0.7409}\right)\\
\widehat{SE(\hat\beta_0)} &= \sqrt{1.362}\\
&= 1.167
\end{aligned}
$$

```
se_beta1()
```

$$
\begin{aligned}
\hat{\sigma_\varepsilon}^2 &= (1-r^2)\hat\sigma_Y^2\\
&= (1-0.4375)\times 1.756\\
&= 0.988\\
S_\varepsilon^2 &= \frac{n}{n-2}\hat{\sigma_\varepsilon}^2\\
&= \frac{100}{100-2}\hat{\sigma_\varepsilon}^2\\
&= \frac{100}{100-2}\times 0.988 = 1.008
\end{aligned}
$$

E quindi

$$
\begin{aligned}
V(\hat\beta_1) &= \frac{\sigma_\varepsilon^2}{n\hat\sigma_X^2}\\
\widehat{V(\hat\beta_1)} &= \frac{S_\varepsilon^2}{n\hat\sigma_X^2}\\
&= \frac{1.008}{100\times 0.7409} = 0.01361\\
\widehat{SE(\hat\beta_1)} &= \sqrt{0.01361}\\
&= 0.1166
\end{aligned}
$$

```
ttest_beta(cof = 0,bj0 = 0,h1 = "<",alpha = 0.01)
```

$\boxed{\text{A}}$ FORMULAZIONE DELLE IPOTESI

$$
\begin{cases}
H_0 : \beta_0 = \beta_{0;H_0} = 0\\
H_1 : \beta_0 < \beta_{0;H_0} = 0
\end{cases}
$$

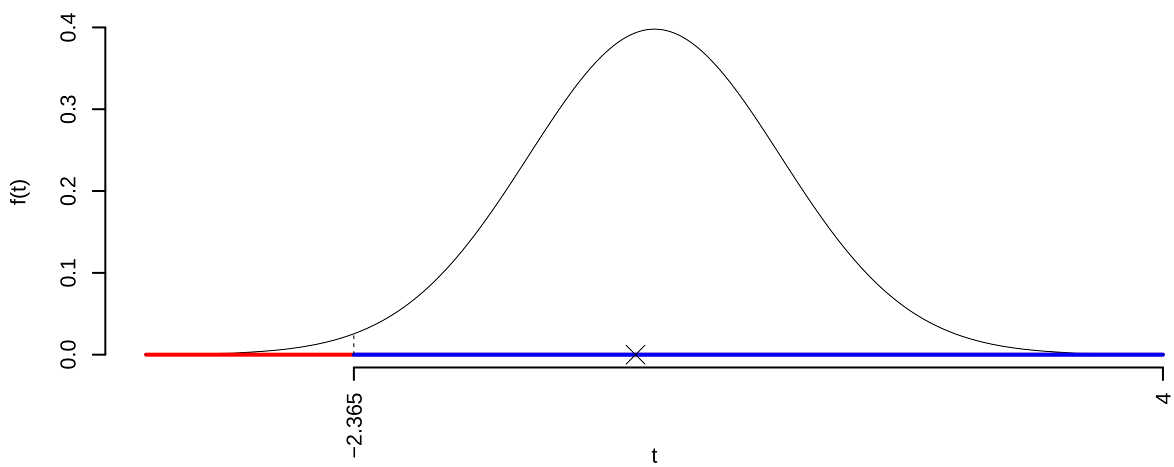
$\boxed{\text{B}}$ SCELTA E CALCOLO STATISTICA-TEST, T Test su un coefficiente di regressione: $\Rightarrow$ t-Test.

$$
\begin{aligned}
\frac{\hat\beta_0-\beta_{0;H_0}}{\widehat{SE(\hat\beta_0)}} &\sim t_{n-2}\\
t_{\text{obs}} &= \frac{(-0.1736-0)}{1.167}=-0.1488\,.
\end{aligned}
$$

$\boxed{\text{C}}$ CONCLUSIONE

La siginficatitività è $\alpha=0.01$, dalle tavole osserviamo $t_{100-2;0.01}=-2.365$.

Essendo $t_{\text{obs}}=-0.1488>t_{100-2;0.01}=-2.365$ allora **rifiuto** H_0 al 1%.



Il p_{value} è

$$p_{\text{value}}=P(T_{100-2}<-0.15)=0.441028$$

Attenzione il calcolo del p_{value} con la T è puramente illustrativo e non può essere riprodotto senza una calcolatrice statistica adeguata.

$$0.1<p_{\text{value}}=0.441028\leq 1$$

```
ttest_beta(cof = 1,bj0 = 0,h1 = "\\neq",alpha = 0.01)
```

$\boxed{\text{A}}$ FORMULAZIONE DELLE IPOTESI

$$\begin{cases}H_0:\beta_1=\beta_{1;H_0}=0\\ H_1:\beta_1\neq\beta_{1;H_0}=0\end{cases}$$

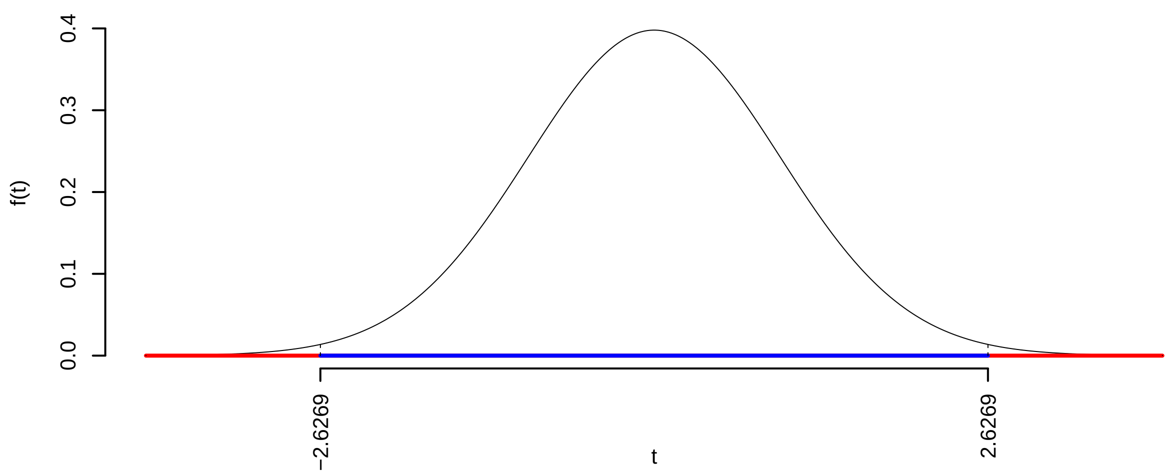
$\boxed{\text{B}}$ SCELTA E CALCOLO STATISTICA-TEST, T Test su un coefficiente di regressione: $\Rightarrow$ t-Test.

$$\begin{aligned} \frac{\hat{\beta}_1 - \beta_{1;H_0}}{\widehat{SE(\hat{\beta}_1)}} &\sim t_{n-2} \\ t_{\text{obs}} &= \frac{(1.018 - 0)}{0.1166} = 8.733\,. \end{aligned}$$

$\boxed{\text{C}}$ CONCLUSIONE

La siginficatitività è $\alpha = 0.01$, dalle tavole osserviamo $t_{100-2;0.005} = 2.6269$.

Essendo $|t_{\text{obs}}| = 8.7327 > t_{100-2;0.005} = 2.6269$ allora **rifiuto** H_0 al 1%.



Il p_{value} è

$$p_{\text{value}} = P(|T_{100-2}| > |8.73|) = 2P(T_{100-2} > 8.73) = 7e - 14$$

Attenzione il calcolo del p_{value} con la T è puramente illustrativo e non può essere riprodotto senza una calcolatrice statistica adeguata.

$$0 < p_{\text{value}} = 7e - 14 \leq 0.001$$

E.5 Esempi

E.5.0.1 Esercizio 1

````
```{r 25-test-functions-12, echo=FALSE}
set.seed(1)                         # per ottenere sempre la stessa simulazione
n <- 250                            # ampiezza campionaria

brk  <- c(0,15,30,50,100,250)          # intervalli (breaks)
hhh  <- c( 20,120,100, 50,10)       # aspetto presunto istogramma

nomex <- "Spesa"
samp <- genera_dati(brk = brk,hhh = hhh,n = n)

ls2e(stat_base(samp,brk))           # crea il data set e la tabella dat3
```

Su un campione di $`r n`$ famiglie della provincia di Modena è stato
rilevata la spesa mensile in telecomunicazioni (in euro), qui di seguito la
distribuzione delle frequenze relative:

```{r 25-test-functions-13, echo=FALSE}
kable(dat3[,c(1,2,4)]) %>%
  kable_styling(full_width = F)
```

1.a (**Punti 14**) Disegnare l'istogramma di densità
percentuale.

**Soluzione**

```{r 25-test-functions-14, echo=FALSE}
kable(dat3) %>%             # Stampa la tabella
  kable_styling(full_width = F)
histp(axes = T)
h.int(60,250,density=20)
```

1.b (**Punti 3**) Qual è la percentuale di famiglie con
spesa superiore a 60 euro?

**Soluzione**

```{r 25-test-functions-15, echo=FALSE}
F_print(60,verso=">")
```
````

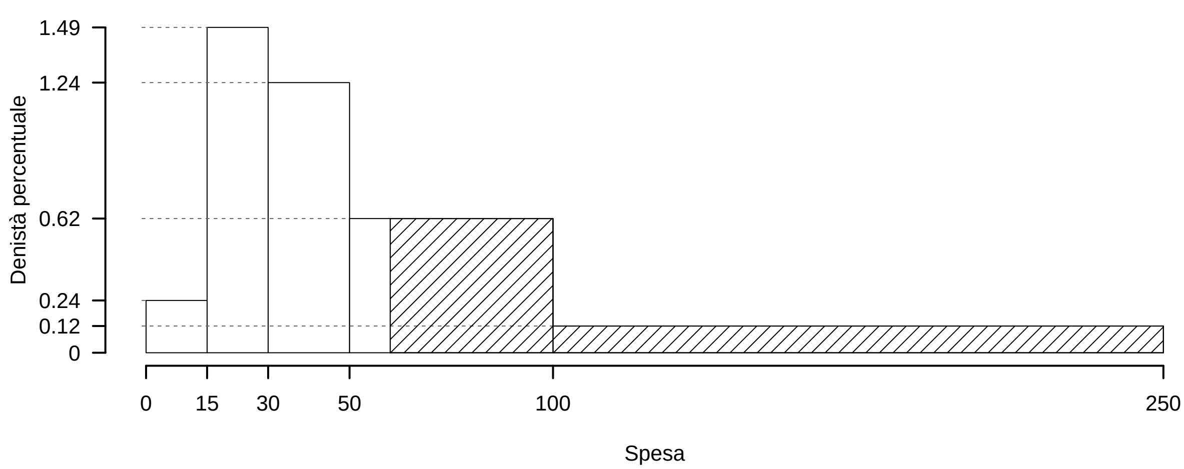
Su un campione di 250 famiglie della provincia di Modena è stato rilevata la spesa mensile in telecomunicazioni (in euro), qui di seguito la distribuzione delle frequenze relative:

$[x_j,$	$x_{j+1})$	f_j
0	15	0.036
15	30	0.224
30	50	0.248
50	100	0.308
100	250	0.184
		1.000

1.a (**Punti 14**) Disegnare l'istogramma di densità percentuale.

Soluzione

$[x_j,$	$x_{j+1})$	n_j	f_j	b_j	h_j	F_j	$f_{j\%}$
0	15	9	0.036	15	0.2400	0.036	3.6
15	30	56	0.224	15	1.4933	0.260	22.4
30	50	62	0.248	20	1.2400	0.508	24.8
50	100	77	0.308	50	0.6160	0.816	30.8
100	250	46	0.184	150	0.1227	1.000	18.4
		250	1.000	250			100.0



1.b (**Punti 3**) Qual è la percentuale di famiglie con spesa superiore a 60 euro?

Soluzione

$$\begin{aligned}
\%(X > 60) &= (100 - 60) \times h_4 + f_5 \times 100 \\
&= (40) \times 0.616 + (0.184) \times 100 \\
&= 0.4304 \times (100) \\
\#(X > 60) &\approx 108
\end{aligned}$$

````
```{r 25-test-functions-21, echo=FALSE}
# preparo i parametri
s1 <- 27
n1 <- 37
s2 <- 30
n2 <- 45
alpha <- 0.05
h1 <- "\\neq"
```

`r i2 <- i2+1;item()` Sono stati intervistati `r n1` uomini
e `r n2` donne, `r s1` su `r n1` uomini si sono
dichiarati favorevoli, mentre sono favorevoli `r s2` su `r n2`
donne. Testare al livello di significatività del 5% l'ipotesi che uomini e
donna abbiano lo stesso parare contro l'alternativa che siano diversi.

**Soluzione**

```{r 25-test-functions-22,results='asis', echo=FALSE}
ztest_2c_pi(s1 = s1,s2 = s2,n1 = n1,n2 = n2,
                h1 = h1,alpha = alpha,a = "U",b = "D")
```

`r i2 <- i2+1;item()` Costruire un intervallo di confidenza al 95% per
la proporzione di uomini favorevoli

**Soluzione**

```{r 25-test-functions-23,results='asis', echo=FALSE}
idc(xm = s1,alpha = .95,n = n1 ,dist_ = "z")
```
````

5.a Sono stati intervistati 37 uomini e 45 donne, 27 su 37 uomini si sono dichiarati favorevoli, mentre sono favorevoli 30 su 45 donne. Testare al livello di significatività del 5% l'ipotesi che uomini e donna abbiano lo stesso parare contro l'alternativa che siano diversi.

Soluzione

Test Z per due proporzioni

$\boxed{\text{A}}$ FORMULAZIONE DELLE IPOTESI

$$\begin{cases} H_0 : \pi_{\text{U}} = \pi_{\text{D}} \\ H_1 : \pi_{\text{U}} \neq \pi_{\text{D}} \end{cases}$$

$\boxed{\text{B}}$ SCELTA E CALCOLO STATISTICA-TEST, Z

$$\hat{\pi}_{\text{U}} = \frac{s_{\text{U}}}{n_{\text{U}}} = \frac{27}{37} = 0.7297 \qquad \hat{\pi}_{\text{D}} = \frac{s_{\text{D}}}{n_{\text{D}}} = \frac{30}{45} = 0.6667$$

Calcoliamo la proporzione comune sotto H_0

$$\pi_C = \frac{s_{\text{U}} + s_{\text{D}}}{n_{\text{U}} + n_{\text{D}}} = \frac{57}{82} = 0.6951$$

$$\frac{\hat{\pi}_{\text{U}} - \hat{\pi}_{\text{D}}}{\sqrt{\frac{\pi_C(1-\pi_C)}{n_{\text{U}}} + \frac{\pi_C(1-\pi_C)}{n_{\text{D}}}}} \sim N(0,1)$$

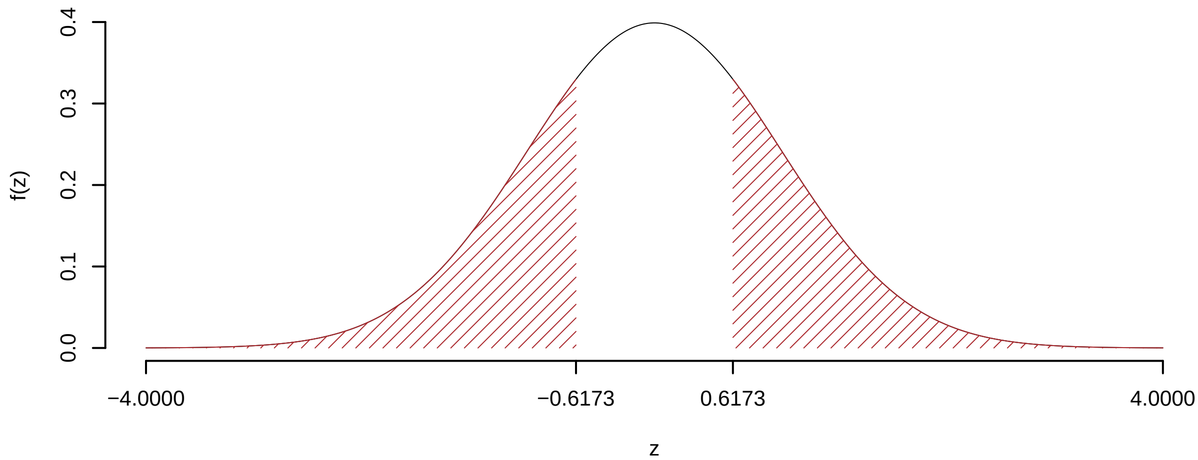
$$z_{\text{obs}} = \frac{(0.7297 - 0.6667)}{\sqrt{\frac{0.6951(1-0.6951)}{37} + \frac{0.6951(1-0.6951)}{45}}} = 0.6173\,.$$

$\boxed{\text{C}}$ CONCLUSIONE

Il p_{value} è

$$p_{\text{value}} = P(|Z| > |0.62|) = 2P(Z > 0.62) = 0.537051$$

$$0.1 < p_{\text{value}} = 0.537051 \leq 1$$



Non rifiuto H_0 a **nessun** livello di significatività,

$p_{\text{value}} > 0.1$, *non significativo*

5.b Costruire un intervallo di confidenza al 95% per la proporzione di uomini favorevoli

Soluzione

$1 - \alpha = 0.05$ e quindi $\alpha = 0.95 \rightarrow \alpha/2 = 0.475$

$$\hat{\pi} = \frac{S_n}{n} = \frac{27}{37} = 0.7297$$

$$\begin{aligned} Idc: \quad & \hat{\pi} \pm z_{\alpha/2} \times \sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}} \\ & 0.7297 \pm 0.06271 \times \sqrt{\frac{0.7297(1-0.7297)}{37}} \\ & 0.7297 \pm 0.06271 \times 0.07301 \\ & [0.7252, 0.7343] \end{aligned}$$