Riassunto Statistica principi e metodi di Cicchitelli

Statistica (Sapienza - Università di Roma)

STATISTICA: PRINCIPI E METODI

CICCHITELLI, D'URSO, MINOZZO

CAPITOLO 1 NOZIONI INTRODUTTIVE

1. INTRODUZIONE

Statistica: disciplina che elabora i principi e le metodologie che dirigono al processo di rilevazione e raccolta dei dati, alla rappresentazione sintetica e alla interpretazione dei dati stessi e, laddove ve ne siano le condizioni, alla generalizzazione delle evidenze osservate.

La statistica è un ramo della matematica i cui obiettivi sono:

- ✓ analizzare e interpretare i dati osservati (descrittiva);
- ✓ studiare i fenomeni aleatori/casuali (probabilità);
- ✓ dedurre certe proprietà dei dati analizzando il margine d'errore (inferenza).

2. TERMINOLOGIA ESSENZIALE

Unità statistica: il singolo caso o la singola osservazione.

Collettivo statistico: insieme delle unità statistiche (insieme dei casi o osservazioni). Può indicare:

- la totalità dei casi osservati (intero insieme delle unità statistiche osservate) che in questo caso chiameremo popolazione;
- la parte della totalità dei casi osservati (sottoinsieme di unità statistiche osservate) che chiameremo campione.

Carattere: ogni caratteristica delle unità statistiche.

Modalità: i diversi modi con cui il carattere si manifesta nelle unità statistiche del collettivo. Ad ogni unità statistica di può associare una sola modalità.

I caratteri si distinguono:

- A seconda dell'insieme delle modalità:

✓ **Qualitativi:** che hanno modalità costituite da singole parole o da espressioni verbali; a loro volta si distinguono in:

a) **Sconnessi:** quando le modalità non presentano un ordine naturale, ossia se date due sue modalità distinte è possibile affermare solo se queste sono uguali o distinte (esempio il sesso e il luogo di nascita); si distinguono in:
   - **dicotomici:** quando possono assumere due sole modalità (esempio sesso: maschio o femmina);
   - **politomici:** quando assumono un numero finito di modalità distinte (esempio colore di una stanza);

b) **Ordinabili:** quando le modalità presentano un ordine naturale, ossia se date due modalità è possibile dare un ordine, specificando che una precede l'altra (esempio posizioni in graduatoria); a loro volta si distinguono in:
   - **Rettilinei:** se possiedono una modalità iniziale e una finale (esempio grado di soddisfazione);
   - **Ciclici:** se non hanno vere e proprie modalità iniziali e finali (esempio giorno della settimana);

✓ **Quantitativi:** che hanno modalità costituite da numeri; questi a loro volta si distinguono in:

a) **Discreti:** le loro modalità sono quantità distinte e sono quasi sempre espresse da numeri interi;

b) **Continui:** possono assumere tutti i valori in un certo intervallo di numeri reali;

c) **Trasferibili o non trasferibili:** trasferimento di parte del carattere da una unità all'altra.

- A seconda della relazione con il tempo:

1. **Caratteri di stato (o fenomeni di stato):** subiscono piccole variazioni nel tempo ed esso è un fattore di disturbo (esempio statura e peso di persone adulte);
2. **Caratteri di movimento (o fenomeni di movimento):** subiscono ampie variazioni nel tempo ed esso è un elemento indispensabile (esempio l'opinione dei cittadini su un determinato servizio o il valore di un titolo in borsa).

3. MISURAZIONE DEI CARATTERI

1. MISURAZIONE DEI CARATTERI QUALITATIVI

Per i *caratteri sconnessi* si parla di **scala nominale** che consente di classificare le unità statistiche in tanti gruppi distinti quante sono le modalità del carattere, tali gruppi al loro interno presentano omogeneità rispetto al carattere considerato

(esempio scala del carattere stato civile è costituita dalle modalità celibe/nubile, coniugato/coniugata, divorziato/divorziata, ecc.).

Per i *caratteri ordinabili rettilinei* si parla di **scala ordinale** che consente di classificare le unità statistiche in gruppi omogenei e in più permette di graduare i gruppi in base all'ordine che le modalità presentano (esempio scala del carattere grado di soddisfazione del cliente è costituita da basso, medio, alto e molto alto).

2. MISURAZIONE DEI CARATTERI QUANTITATIVI

Per i *caratteri discreti* si utilizza la **scala proporzionale** in cui le modalità del carattere sono espresse da numeri e queste consentono di classificare le unità statistiche in gruppi omogenei, di graduare i gruppi secondo il valore della modalità e di misurare la differenza tra i gruppi tramite la differenza o il rapporto tra le modalità.

Per i *caratteri continui* si utilizza la **scala a intervalli**, la misurazione di un carattere continuo comporta necessariamente un'approssimazione e quindi la rilevazione sottende sempre un intervallo di valori possibili detto intervallo di tolleranza.

4. ALCUNI CALCOLI STATISTICI ELEMENTARI

1. DIFFERENZE ASSOLUTE, RELATIVE E PERCENTUALI

Se *a* e *b* sono i livelli di uno stesso fenomeno, espressi nella stessa unità di misura, ma riferiti a situazioni diverse, il confronto tra le quantità *a* e *b* può essere effettuato tramite la **differenza assoluta** *b* – *a*.

Dalla differenza assoluta si passa alla **differenza relativa** dividendo per *a* o per *b*.

Assumendo come riferimento a avremo $\frac{b-a}{a}$ oppure assumendo come riferimento b avremo $\frac{b-a}{b}$.

Moltiplicando la differenza relativa per 100, si ottiene la **differenza percentuale**.

$$\frac{b-a}{a}\cdot 100 \quad \text{oppure} \quad \frac{b-a}{b}\cdot 100$$

2. I RAPPORTI STATISTICI

Il confronto tra le grandezze *a* e *b* può essere effettuato anche tramite i rapporti

I rapporti possono essere di:

✓ **Composizione:** si ottengono dividendo una quantità parziale alla quantità totale.

Dato un insieme di grandezze omogenee (cioè espresse nella stessa unità di misura) $a_1, a_2, \ldots, a_k$ si vogliono effettuare confronti con le singole grandezze con il totale $A = a_1 + a_2 + \ldots + a_k$ e tali rapporti trasformano le quantità in percentuale del totale

$\frac{a_1}{A}\cdot 100, \frac{a_2}{A}\cdot 100, \ldots, \frac{a_k}{A}\cdot 100$ e sono utili per mettere meglio in luce l'importanza delle singole unità rispetto al totale (la somma di tali rapporti è uguale a 100)

Importanza relativa dei consumi di olio combustibile nelle diverse aree geografiche nell'anno 2004

Area geografica	Milioni di tonnellate	Rapporto di composizione	%
America del Nord	1122,4	0,2979	29,79
Centro e Sud America	221,7	0,0589	5,89
Europa e Eurasia	957,3	0,2541	25,41
Medio Oriente	250,9	0,0666	3,66
Africa	124,3	0,0330	3,30
Asia e Oceania	1090,5	0,2895	28,95
Tot.	3767,1	1,0000	100,00

✓ **Coesistenza:** si ottengono rapportando due dati statistici coesistenti per i quali si ritiene debba mantenersi una certa proporzione.

Dato un insieme di grandezze $a_1, a_2, \ldots, a_k$ queste vengono messe a confronto tra di loro.

Questi rapporti trovano larga applicazione nell'analisi dei fenomeni demografici (esempio rapporto di mascolinità = $\frac{popolazione\ residente\ di\ sesso\ maschile}{popolazione\ residente\ di\ sesso\ femminile}\cdot 100$);

Rapporti di mascolinità alla nascita nelle regioni italiane nel 2004

Regioni	Nati Maschi	Nati Femmine	Rap. %
Piemonte	19283	18130	106,36
Valle d'Aos.	591	586	100,85
Lombardia	47317	45581	103,81
Trentino AA	5632	5270	106,87
Veneto	24207	22890	105,75
Friuli VG	5132	4969	103,28
Liguria	6188	5826	106,21
Emilia Rom.	19540	18535	105,42
Toscana	16329	15366	106,26
Umbria	3844	3773	101,88

- ✓ **Derivazione:** si ottengono dividendo un dato statistico X per un altro dato statistico Y che si ritiene sia la causa o il presupposto del primo (esempio tasso di natalità $= \frac{nati\,nell'anno}{popolazione\,media\,dell'anno} \cdot 1000$, questo rapporto indica il numero dei nati dell'anno su 1000 abitanti);

Consumi pro-capite di olio combustibile in Italia, Cina e Giappone nel 2004

Nazioni	Consumi in tonnellate	Abitanti	Rapporto tonnellate per abitante
Italia	89.500.000	57.300.000	1,562
Giappone	241.500.000	127.800.000	1,890
Cina	308.600.000	1.313.300.000	0,235

- ✓ **Densità:** si tratta di rapporti in cui al numeratore è posta la misura di un dato fenomeno e al denominatore la dimensione di un secondo fenomeno a cui il primo è connesso (esempio abitanti per km^2).

Abitanti e superfici di alcuni stati (2004, dati ONU)

Nazioni	Abitanti	Superficie in Km^2	Abitanti per Km^2
Cina	1.313.300.000	9.595.960	136,85
USA	298.000.000	9.629.091	30,84
Giappone	127.800.000	377.835	338,24
Francia	60.400.000	547.030	110,41
Italia	57.300.000	301.230	190,22
Regno Unito	59.400.000	244.820	242,63
Olanda	16.200.000	41.526	390,12

CAPITOLO 2 DISTRIBUZIONI STATISTICHE

1. DISTRIBUZIONI DISAGGREGATE (O UNITARIE)

Consideriamo un collettivo statistico di N unità dove sia osservato il singolo carattere X. Si chiama **distribuzione disaggregata o unitaria**, secondo il carattere X, l'insieme delle osservazioni (rappresentate da numeri o da espressioni verbali) relative alle N unità del collettivo. La distribuzione statistica semplice disaggregata (o unitaria) sarà indicata come $x_1, x_2, \dots, x_N$ dove x_1 è l'osservazione relativa all'unità identificata dal numero 1 e così via.

i	X
1	x_1
2	x_2
⋮	⋮
N	x_N

i = indice di unità; X = carattere; $x_1, x_2, \dots, x_N$ = modalità

2. DISTRIBUZIONE DI FREQUENZE

Le unità della distribuzione unitaria vengono generalmente aggregate in gruppi omogenei sulla base di uno o più caratteri e questo richiede che si individuino preliminarmente le modalità rispetto a cui effettuare il raggruppamento.

Consideriamo un collettivo statistico di N unità, prendiamo come riferimento un generico carattere X che presenta k distinte modalità indicate con $x_1, x_2, \dots, x_k$.

Se il carattere è quantitativo o qualitativo ordinabile rettilineo le modalità si intendono ordinate cioè $x_1 < x_2 < \cdots < x_j < \cdots < x_k$.

La distribuzione di frequenza è lo schema con cui si associa a ciascuna modalità del carattere X la rispettiva frequenza.

1. Frequenze assolute

Per frequenza si intende il numero di volte (n_j per ogni $j=1,2,\ldots,k$) che una data modalità si presenta nel collettivo statistico e questo numero di volte sono chiamati **frequenze assolute** e $N=\sum_{j=1}^{k} n_j$ rappresenta la sommatoria di tutte le possibili modalità il cui risultato sarà l'intero collettivo statistico N .

j	X	Frequenze (assolute)
1	x_1	n_1
2	x_2	n_2
⋮	⋮	⋮
k	x_k	n_k
Totale		N

Dove $n_1, n_2, \ldots, n_k$ sono le frequenze delle modalità $x_1, x_2, \ldots, x_k$.

2. Frequenze relative

Accanto alle frequenze assolute vengono presentate le **frequenze relative** che si ottengono rapportando le frequenze assolute al totale delle unità N e le indicheremo con $f_1, f_2, \ldots, f_k$.

$$f_j=\frac{\text{frequenze assolute (o frequenze della modalità } x_i)}{\text{numero totale di unità}}=\frac{n_j}{N} \quad \text{con} \quad j=1,2,\ldots,k$$

Le frequenze relative sono dei particolari rapporti di composizione e consentono una valutazione rapida del peso della singola modalità nell'ambito della distribuzione di frequenze.

Valgono le proprietà:

$$f_1+f_2+\ldots+f_k=1 \qquad \sum_{j=1}^{k} f_j=\sum_{j=1}^{k} \frac{n_j}{N}=\frac{1}{N}\sum_{j=1}^{k} n_j=\frac{N}{N}=1$$

$$0 \le f_j \le 1 \quad \text{con } j = 1,2, \ldots, k$$

Distribuzione di frequenze relative

j	X	Frequenze assolute	Frequenze relative
1	x_1	n_1	f_1
2	x_2	n_2	f_2
⋮	⋮	⋮	⋮
k	x_k	n_k	f_k
Tot		N	1

3. Frequenze percentuali

Nella pratica si utilizzano le **frequenze percentuali** che si ottengono moltiplicando per 100 le frequenze relative $p_j = f_j \cdot 100 \; con \; j = 1, 2, \ldots, k$.

Valgono le proprietà:

$$p_1 + p_2 + \ldots + p_k = 100 \quad \left(\sum_{j=1}^{k} p_j = 100 \right)$$

$$0 \leq p_j \leq 100 \quad \text{con j = 1,2, ..., k}$$

Distribuzione di frequenze percentuali

j	X	Frequenze assolute	Frequenze relative	Frequenze percentuali
1	x_1	n_1	f_1	p_1
2	x_2	n_2	f_2	p_2
⋮	⋮	⋮	⋮	⋮
k	x_k	n_k	f_k	p_k
Tot		N	1	100%

4. Frequenze assolute cumulate

Consideriamo una distribuzione di frequenze secondo un carattere quantitativo ordinabile, le **frequenze assolute cumulate** sono le quantità:

$$N_j = n_1 + n_2 + \ldots + n_j \quad \text{con} \quad j = 1, 2, \ldots, k \quad \text{ovvero} \quad N_j = \sum_{i=1}^{j} n_i, \forall \, j = 1, 2, \ldots, k$$

$$N_1 = n_1, N_j = N_{j-1} + n_j \, \forall \, j = 2, 3, \ldots, k; N_k = N$$

Lo schema che associa a ciascuna modalità del carattere X la rispettiva frequenza cumulata è il seguente:

j	X	Frequenze assolute	Frequenze relative	Frequenze percentuali	Frequenze cumulate assolute
1	x_1	n_1	f_1	p_1	$N_1 = n_1$
2	x_2	n_2	f_2	p_2	$N_2 = n_1 + n_2$
⋮	⋮	⋮	⋮	⋮	⋮
k	x_k	n_k	f_k	p_k	$N_k = \sum_{j=1}^{k} n_j = N$
Tot		N	1	100	

5. Frequenze relative cumulate

Le **frequenze relative cumulate** sono i rapporti:

$$F_j = \frac{N_j}{N}, \forall\, j = 1, 2, \ldots, k$$

$$F_1 = f_1, F_j = F_{j-1} + f_j \;\forall\, j = 2, 3, \ldots, k\,;\, F_k = 1$$

j	X	Frequenze assolute	Frequenze relative	Frequenze percentuali	Frequenze cumulate assolute	Frequenze cumulate relative
1	x_1	n_1	f_1	p_1	$N_1 = n_1$	$F_1 = f_1$
2	x_2	n_2	f_2	p_2	$N_2 = n_1 + n_2$	$F_2 = f_1 + f_2$
⋮	⋮	⋮	⋮	⋮	⋮	⋮
k	x_k	n_k	f_k	p_k	$N_k = N$	$F_k = \sum_{i=1}^{j} f_j = 1$
Tot		N	1	100		

6. Frequenze percentuali cumulate

Le **frequenze percentuali cumulate** sono le quantità:

$$P_j = F_j \cdot 100 = \frac{N_j}{N} \cdot 100 \quad \text{con } j = 1, 2, \ldots, k$$

7. Raggruppamento in classi

Quando il carattere è quantitativo e il numero di osservazioni è elevato, la presentazione dei dati richiede necessariamente che le modalità contigue siano aggregate tramite la formazione di classi, cioè intervalli numerici tra loro disgiunti comprendenti più modalità.

La classe j -ma ha un estremo sinistro $l_j^{-¿}$ ($c_j^{-¿}$) e un estremo destro $l_j^{+¿}$ ($c_j^{+¿}$) per ogni $j = 1, 2, \ldots, k$, dove ora k indica il numero di classi.

È necessario stabilire la chiusura delle classi:

- una classe è chiusa a destra se include le unità che presentano modalità uguali all'estremo destro della classe $\begin{matrix} +¿ \\ -¿, l_j^{¿} \\ l_j^{¿} \\ ¿ \end{matrix}$ oppure una classe è chiusa a sinistra se include le unità che presentano modalità uguali all'estremo sinistro $\begin{matrix} +¿ \\ -¿, l_j^{¿} \\ l_j^{¿} \\ ¿ \end{matrix}$ e in questi due casi $\begin{matrix} -¿ \\ +¿ = l_j^{¿} \\ l_{j-1}^{¿} \end{matrix}$ (l'estremo destro precedente è uguale all'estremo sinistro successivo) per coprire tutte le possibili modalità di X ;
- una classe può essere chiusa ad entrambi i lati $\begin{matrix} +¿ \\ -¿, l_j^{¿} \\ l_j^{¿} \\ ¿ \end{matrix}$ e in questo caso $\begin{matrix} -¿ \\ +¿ \neq l_j^{¿} \\ l_{j-1}^{¿} \end{matrix}$ perché gli intervalli devono essere disgiunti.

La distribuzione di frequenze assolute di un carattere X suddiviso in classi è lo schema con cui si associa a ciascuna classe la rispettiva frequenza assoluta.

[illegible]	[illegible]	[illegible]
[illegible]	$c_0 - [illegible]_1$	n_1
[illegible]	$[illegible]_1 - [illegible]_2$	$[illegible]_2$
[illegible]	⋮	⋮
[illegible]	$[illegible]_{k-1} - [illegible]_k$	$[illegible]_k$
[illegible]		[illegible]

L'**ampiezza della classe** generica ($\begin{matrix} +¿ \\ -¿, l_j^{¿} \\ l_j^{¿} \end{matrix}$ oppure ($\begin{matrix} +¿ \\ -¿, c_j^{¿} \\ c_j^{¿} \end{matrix}$ è la quantità $\begin{matrix} -¿ \\ +¿ - l_j^{¿} \\ d_j = l_j^{¿} \end{matrix}$ ($\begin{matrix} -¿ \\ +¿ - c_j^{¿} \\ d_j = c_j^{¿} \end{matrix}$) (differenza tra estremo destro ed estremo sinistro).

Il **valore centrale della classe** è la quantità $\begin{matrix} -¿+\frac{l_j^{+¿}}{2} \\ l_j^{¿} \\ x_j=¿ \end{matrix}$ ($\begin{matrix} -¿+\frac{c_j^{+¿}}{2} \\ c_j^{¿} \\ x_j=¿ \end{matrix}$)

Sia $\begin{matrix} +¿ \\ -¿, l_j^{¿} \\ l_j^{¿} \end{matrix}$ la generica classe di una distribuzione di frequenze con modalità raggruppate in classi, la **densità di frequenza** delle classi è il rapporto

$$h_j=\frac{frequenza\ assoluta\ della\ classe}{ampiezza\ della\ classe}=\frac{n_j}{d_j}\ con\ j=1,2,\ldots,k$$

Essendo $\begin{matrix} -¿ \\ +¿-l_j^{¿} \\ d_j=l_j^{¿} \end{matrix}$ l'ampiezza della classe.

Casi	Ampiezza
Caratteri continui o $(l_j^-, l_j^+]$ o $[l_j^-, l_j^+)$	$d_j = l_j^+ - l_j^-$
Caratteri discreti e $[l_j^-, l_j^+]$	$d_j = l_j^+ - l_j^- + 1$

La singola frequenza viene divisa per l'ampiezza della classe a cui si riferisce ottenendo un numero che indica quanto sono in media le unità del carattere e quando le classi hanno ampiezze diverse le frequenze non sono confrontabili mentre lo sono le densità di frequenza.

All'interno della classe j -esima, il numero «atteso» di unità appartenenti ad un intervallo di ampiezza $\delta \le d_j$ è pari a

$$n(\delta; j)=\frac{n_j}{d_j}\cdot\delta=h_j\cdot\delta$$

CAPITOLO 3
RAPPRESENTAZIONI GRAFICHE

Le rappresentazioni grafiche hanno lo scopo di illustrare le distribuzioni di frequenze.

1. DISTRIBUZIONE DI FREQUENZE PER CARATTERI QUANTITATIVI

Consideriamo una distribuzione di frequenze per un carattere quantitativo

j	X	Frequenze (assolute)
1	x_1	n_1
2	x_2	n_2
⋮	⋮	⋮
k	x_k	n_k
Totale		N

In cui i caratteri sono ordinati $x_1<x_2<\cdots<x_j<\cdots<x_k$.

Vi sono due modi per rappresentare tramite un sistema di assi cartesiani tale distribuzione di frequenze:

- la **funzione di frequenza** che descrive l'associazione tra modalità e frequenze;
- la **funzione di ripartizione**, che descrive l'andamento delle frequenze cumulate.

1. Diagramma ad aste

Il **diagramma ad aste** è la rappresentazione grafica della *funzione di frequenza*.

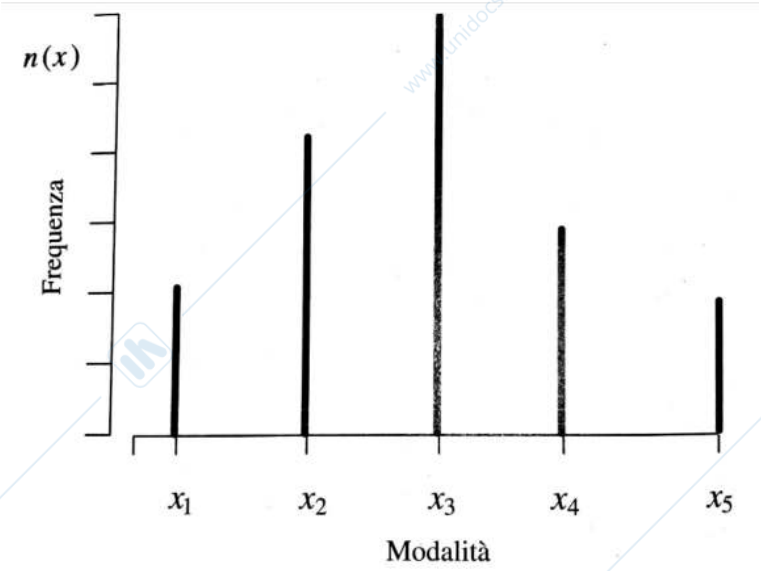
Considerando la funzione di frequenze assolute, questa è definita:

$$n(x)=\begin{cases} n_j, & \text{per } x=x_j,\ j=1,2,\ldots,k \\ 0 & \text{altrove} \end{cases}$$

La rappresentazione grafica si effettua ponendo sull'asse delle ascisse le modalità $x_1,x_2,\ldots,x_k$ e sull'asse delle ordinate le frequenze assolute corrispondenti $n_1,n_2,\ldots,n_k$ dopo aver fissato un'opportuna scala per ciascuno di essi.

Considerando la funzione di frequenze relative, questa è definita:

$$f(x)=\begin{cases} f_j, & \text{per } x=x_j,\ j=1,2,\ldots,k \\ 0 & \text{altrove} \end{cases}$$



2. Funzione di ripartizione

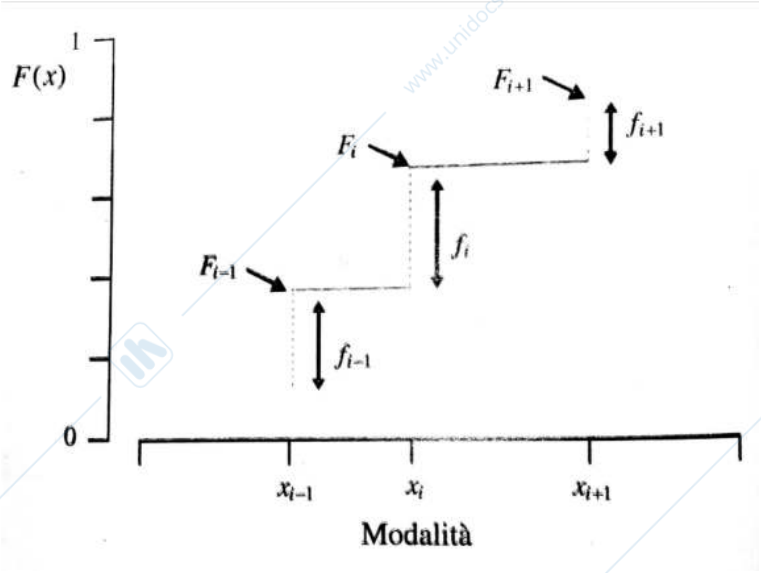
Sia X un carattere quantitativo osservato in un collettivo statistico di N unità. Si chiama **funzione di ripartizione**, $F(x)$, la funzione che associa a un qualsiasi numero, x, dell'asse reale la frequenza relativa delle unità del collettivo per cui $X \leq x$.

L'espressione della funzione di ripartizione è:

$$F(x)=\begin{cases} 0, & \text{per } x<x_1 \\ F_j, & \text{per } x_j \leq x < x_{j+1},\ j=1,2,\ldots,k-1 \\ 1 & \text{per } x \geq x_k \end{cases}$$

Il valore nullo per $x<x_1$ è dovuta al fatto che nel collettivo statistico non ci sono unità con modalità più piccole di x_1. Il valore 1 è dovuto al fatto che nel collettivo statistico tutte le unità presentano una modalità minore o uguale a x_k. Infine, la costanza della funzione nell'intervallo chiuso a sinistra si giustifica pensando che le unità del collettivo statistico con modalità minore o uguale a x rimangono ferme a N_j e quindi $F(x)=\frac{N_j}{N}=F_j$ per ogni x in tale intervallo.

La $F(x)$ è una funzione costante a tratti che presenta dei salti e questi avvengono in corrispondenza delle x_j e l'ampiezza del salto corrisponde alla frequenza relativa f_j.



2. DISTRIBUZIONE DI FREQUENZE PER CARATTERI QUANTITATIVI DIVISI IN CLASSI

Consideriamo una distribuzione di frequenze per un carattere quantitativo diviso in classi.

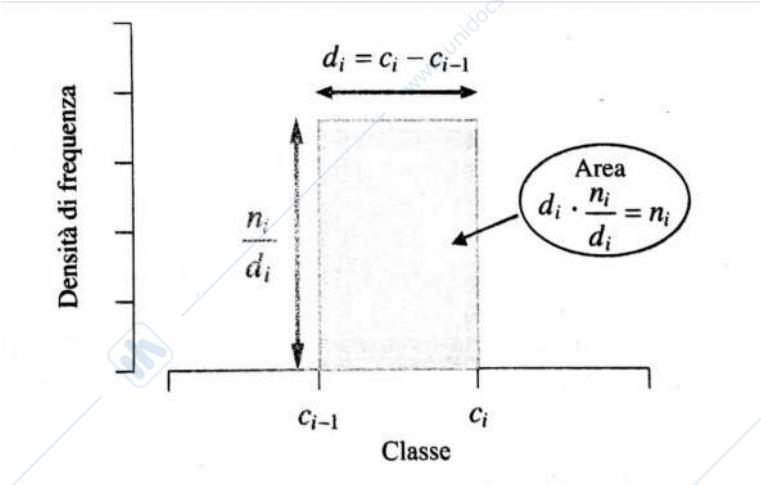
j	$Classi\ X$	Frequenze (assolute)
1	$c_0 - c_1$	n_1
2	$c_1 - c_2$	n_2
⋮	⋮	⋮
k	$c_{k-1} - c_k$	n_k
Totale		N

Vi sono due differenti rappresentazioni cartesiane:

- l'**istogramma di frequenza**, con il quale rappresentiamo l'associazione tra classi e frequenze;
- la **funzione di ripartizione** che mostra l'andamento delle *frequenze cumulate*.

1. Istogramma di frequenza

L'**istogramma di frequenza** è la rappresentazione grafica che si ottiene ponendo sull'asse delle ascisse gli estremi di classe $c_0, c_1, \ldots, c_k$ e disegnando per ogni classe (c_{i-1}, c_i) $i=1,2,\ldots,k$ un rettangolo avente per base il segmento dell'asse delle ascisse di estremi $c_{i-1} e\, c_i$ e per altezza la densità di frequenza $\frac{n_j}{d_j}$



Le frequenze sono rappresentate dalle aree dei rettangoli:

$$d_j \cdot h_j = d_j \cdot \frac{n_j}{d_j} = n_j$$

2. Funzione di ripartizione

La funzione di ripartizione $F(x)$ nel caso delle distribuzioni di frequenze con modalità raggruppate in classi è definita come frequenza relativa delle unità del collettivo statistico in cui il carattere X non supera il valore x .

Si assume che le unità siano uniformemente distribuite all'interno di ciascuna classe e su questa base la funzione di ripartizione è espressa da:

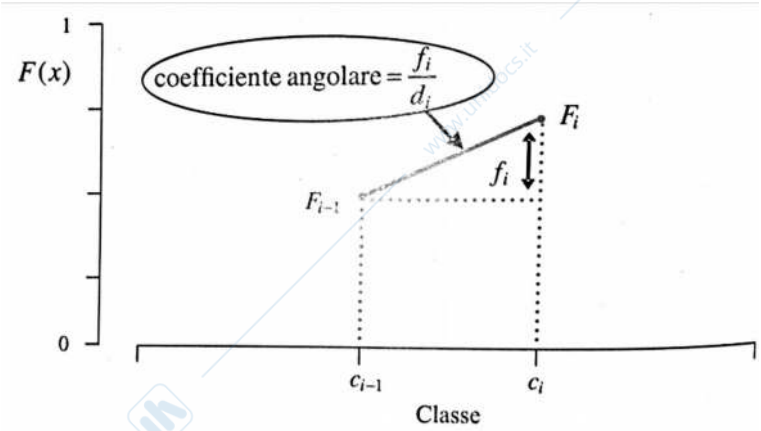
$$F(x) = \begin{cases} 0, & -\infty \le x < l_1 \\ F_{j-1} + \frac{h_j}{N}\left(x - l_j\right), & l_j \le x < l_{j+1},\ j = 1, 2, \ldots, k \\ 1, & \text{per } x \ge l_k \end{cases}$$

$F(x) = 0$ per $x < l_1$ e $F(x) = 1$ per $x \ge l_k$ sono dovuti alla non presenza nel collettivo statistico rispettivamente di unità con modalità minori di l_1 e maggiori di l_k .

$$F(x)=F_{j-1}+\frac{h_j}{N}\cdot\frac{x-l_j}{l_j}$$ è l'equazione del segmento di retta che congiunge i punti del piano di coordinate (c_{i-1}, F_{i-1}) e (c_i, F_i), ne segue che la funzione di ripartizione assume la forma di linea spezzata: ad ogni classe corrisponde un segmento della retta con una inclinazione data dalla densità di frequenza relativa.

Individuiamo nel piano cartesiano i punti di coordinate $(c_0,0),(c_1,F_1),\ldots,(c_k,1)$.

Si uniscono poi a due a due i punti consecutivi con segmenti di retta.



CAPITOLO 4 MEDIE

Le **medie** costituiscono uno strumento con cui si sintetizzano le distribuzioni statistiche.

Le medie possono essere raggruppate in due grandi classi:

- medie analitiche;
- medie lasche.

Le **medie analitiche** sono quelle che si ottengono dall'applicazione di opportune operazioni matematiche a tutti i valori del carattere che formano la distribuzione statistica considerata. Rientrano in questa categoria la **media aritmetica**, **la media geometrica**, **la media armonica** e **la media quadratica**. Queste sono applicabili solo a distribuzioni statistiche di caratteri quantitativi

Le **medie lasche** sono caratterizzate dal fatto che nel loro calcolo intervengono solo alcuni valori specifici della distribuzione, tipicamente quelli che occupano particolari posizioni nella graduatoria dei valori osservati. Rientrano in questa categoria la **mediana**, i **quartili**, i **decili** (e più in generale in **quantili**), denominati anche medie di posizione, il **valore centrale** e la **moda**. Le medie di posizione sono determinabili anche nel caso di distribuzioni statistiche di caratteri qualitativi a modalità ordinabili mentre la moda è definibile per qualsiasi carattere.

Le medie sono indici di posizione, ovvero sono un singolo valore atto a sintetizzare l'intera distribuzione (necessariamente con della perdita di informazione).

Indichiamo con M una generica media e supponiamo di riferirci a una distribuzione statistica disaggregata di un carattere quantitativo $x_1, x_2, \ldots, x_N$ e indichiamo con $x_{(1)}$ e $x_{(N)}$ rispettivamente il termine più piccolo e quello più grande della distribuzione.

Sintetizzare una distribuzione statistica con una media equivale ad assegnare a ogni unità del collettivo o ad ogni osservazione la costante M, quindi si passando dalla distribuzione statistica $x_1, x_2, \ldots, x_N$ a quella costituita da N termini tutti uguali M $M, M, \ldots, M$ e questo è l'effetto livellamento.

Si conviene che la media sia interna e quindi compresa tra $x_{(1)}$ e $x_{(N)}$:

$$x_{(1)} \leq M \leq \quad x_{(N)}$$

Un primo criterio per una definizione operazionale di un insieme di medie è il **criterio di invarianza o di Chisini** in base al quale, indicata con $f(x_1, x_2, \ldots, x_N)$ un'operazione matematica sui termini della distribuzione si definisce come media, secondo questa operazione matematica, la quantità M che sostituita ai termini della distribuzione fa sì che l'operazione matematica applicata alla nuova distribuzione $f(M, M, \ldots, M)$ e a quella originaria dia luogo allo stesso risultato.

$$f(x_1, x_2, \ldots, x_N) = f(M, M, \ldots, M)$$

Una secondo criterio per una definizione operazionale di un insieme di medie è il **criterio della minima perdita** in base al quale, fissata un'operazione matematica che misuri la perdita di informazione connessa alla sostituzione ai vari termini della distribuzione della media M si definisce come media la quantità che minimizza tale funzione di perdita. Generalmente la perdita viene rappresentata con gli scostamenti in valore assoluto tra i valori osservati e la media che stiamo cercando.

$$\sum_{i=1}^{N} |x_i - M|^r$$

1. MEDIA ARITMETICA

La **media aritmetica** di una distribuzione statistica disaggregata è la somma dei termini $x_1, x_2, \ldots, x_N$ divisa per N

$$\mu = \frac{x_1 + x_2 + \cdots x_N}{N} = \frac{1}{N} \sum_{i=1}^{N} x_i$$

Nella presentazione delle proprietà della media aritmetica ci avvarremo del concetto di scarto o scostamento della media aritmetica inteso come differenza tra il singolo termine della distribuzione e la media aritmetica

$$scarto\,della\,media\,aritmetica = x_i - \mu \; con \; i = 1,2,\ldots,N$$

La media aritmetica presenta le proprietà di seguito indicate:

a) è interna, essendo compresa tra il minimo e il massimo dei termini della distribuzione: $x_{(1)} \leq \mu \leq x_{(N)}$. Questa indica la circostanza ovvia che la media aritmetica assume un valore intermedio tra minimo e massimo dei termini della distribuzione e coincide con questi se minimo e massimo sono uguali e quindi tutte le quantità sono uguali $x_1 = x_2 = \ldots = x_N$;

b) la somma dei termini della distribuzione è uguale alla media aritmetica moltiplicata per il numero di unità: $\sum_{i=1}^{N} x_i = N\mu$. In altre parole, la media aritmetica rispecchia il criterio di invarianza per la funzione matematica "somma dei termini". Questa deriva direttamente dalla definizione di media aritmetica moltiplicando la media aritmetica per il numero dei termini otteniamo $x_1 + x_2 + \cdots x_N$ ovvero la somma dei termini della distribuzione ed essa può essere interpretata anche come proprietà di invarianza rispetto all'operazione somma dei termini se si sostituisce la media aritmetica a ogni termine della distribuzione, la somma dei termini della nuova distribuzione è uguale alla somma dei termini della distribuzione originaria;

a) la somma algebrica degli scarti della media aritmetica è nulla: $\sum_{i=1}^{N} (x_i - \mu) = 0$. Essa mette in luce il fatto che μ è un valore intermedio rispetto alle quantità che sintetizza complessivamente gli scarti negativi si compensano con quelli positivi;

b) La somma dei quadrati degli scarti dei termini della distribuzione da una costante c è minima quando c è uguale alla media aritmetica: $\sum_{i=1}^{N}(x_i-\mu)^2=\min_c \sum_{i=1}^{N}(x_i-c)^2$. La somma dei quadrati degli scarti della media aritmetica è più piccola della somma dei quadrati degli scarti da un qualsiasi altro valore c . Se si trasformano i termini $x_1, x_2, \ldots, x_N$ secondo la funzione $y_i=a+b\,x_i$ con i=1, 2, ..., N con a e b costanti qualsiasi, la media aritmetica μ_y dei termini trasformati è legata alla media μ_X dei termini originati dalla medesima trasformazione ossia $\mu_Y=a+b\mu_X$. Tale proprietà è chiamata proprietà di linearità.

c) se un collettivo statistico di N unità è suddiviso in L sottoinsiemi disgiunti aventi numerosità $N^{(1)}, N^{(2)}, \ldots, N^{(L)}$ e medie aritmetiche $\mu^{(1)}, \mu^{(2)}, \ldots, \mu^{(L)}$, la media aritmetica del collettivo può essere calcolata nel modo seguente:

$$\mu=\frac{\mu^{(1)}N^{(1)}+\mu^{(2)}N^{(2)}+\cdots+\mu^{(L)}N^{(L)}}{N^{(1)}+N^{(2)}+\cdots+N^{(L)}}$$

Con riferimento a questa proprietà si dice che la media aritmetica è associativa.

2. MEDIA ARMONICA

La **media armonica** di una distribuzione statistica disaggregata $x_1, x_2, \ldots, x_N$ i cui termini sono tutti diversi da 0, è data dal rapporto tra N e la somma dei reciproci dei termini

$$\mu_a=\frac{N}{\frac{1}{x_1}+\frac{1}{x_2}+\ldots+\frac{1}{x_N}}=\sum_{i=1}^{N}\frac{1}{x_i}$$

Da questa si comprende come si arriva alla media armonica calcolando prima la media aritmetica delle quantità trasformate $\frac{1}{x_i}$ e applicando poi al risultato la trasformazione inversa.

3. MEDIA GEOMETRICA

La **media geometrica** di una distribuzione statistica disaggregata $x_1, x_2, \ldots, x_N$ in cui i termini sono tutti maggiori di 0, è data dalla radice N-esima del prodotto dei termini:

$$\mu_g=\sqrt[N]{x_1\cdot x_2\cdot\ldots\cdot x_N}=\sqrt[N]{\prod_{i=1}^{N}x_i} \quad \text{dove} \quad \prod_{i=1}^{N}x_i=x_1\cdot x_2\cdot\ldots\cdot x_N \ .$$

4. MEDIA QUADRATICA

La **media quadratica** di una distribuzione statistica disaggregata $x_1, x_2, \ldots, x_N$ è la radice quadrata della media aritmetica dei quadrati dei termini della distribuzione

$$\mu_q = \sqrt{\frac{x_1^2 + x_2^2 + \ldots + x_N^2}{N}} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2}$$

5. MEDIA DI POTENZA

La **media di potenza** di ordine r di una distribuzione statistica disaggregata $x_1, x_2, \ldots, x_N$ è data da:

$$\mu_r = \sqrt[r]{\frac{1}{N}\sum_{i=1}^{N} x_i^r}$$

Dove r è un numero reale qualsiasi nell'intervallo $(-\infty, \infty)$.

È facile riconoscere che:

$con\, r = -1\, allora\, \mu^{(r)} = \mu_a (media\, armonica);$

$con\, r = 1\, allora\, \mu^{(r)} = \mu (media\, aritmetica);$

$con\, r = 2\, allora\, \mu^{(r)} = \mu_q (media\, quadratica).$

Si dimostra che:

$per\, r \to -\infty\, allora\, \mu^{(r)} = x_1 (valore\, minimo);$

$per\, r \to 0\, allora\, \mu^{(r)} = \mu_g (media\, geometrica);$

$per\, r \to +\infty\, allora\, \mu^{(r)} = x_N (valore\, massimo).$

6. MEDIE ANALITICHE PER LA DISTRIBUZIONE DI FREQUENZE

Adesso andiamo a adattare le formule precedentemente considerate nel caso della distribuzione di frequenze.

j	x_j	n_j
1	x_1	n_1
2	x_2	n_2
⋮	⋮	⋮
$k\ (\neq N)$	x_k	n_k
	Totale	N

1. Media aritmetica

Poiché la somma dei primi n_1 termini tutti uguali a x_1 è data da $x_1 \cdot n_1$, ecc. la formula $\mu = \frac{x_1 + x_2 + \cdots x_N}{N} = \frac{1}{N}\sum_{i=1}^{N} x_i$

Diviene

$$\mu = \frac{x_1 n_1 + x_2 n_2 + \cdots + x_k n_k}{N} = \frac{1}{N}\sum_{j=1}^{k} x_j n_j$$

Essendo $f_j = \frac{n_j}{N}$ la frequenza relativa della modalità x_j la formula può essere scritta nel seguente modo

$$\mu = \sum_{j=1}^{k} x_j \frac{n_j}{N} = \sum_{i=1}^{k} x_j f_j = x_1 f_1 + x_2 f_2 + \cdots + x_k f_k$$

2. Media armonica, media geometrica e media quadratica

Con le stesse considerazioni svolte per la media aritmetica, è facile stabilire che:

$$\mu_a = \frac{N}{\frac{1}{x_1} + \frac{1}{x_2} + \ldots + \frac{1}{x_N}} = \sum_{i=1}^{N} \frac{1}{x_i}$$

Diviene

$$\mu_a = \frac{N}{\sum_{i=1}^{k} \frac{n_i}{x_i}}$$

$$\mu_g = \sqrt[N]{x_1 \cdot x_2 \cdot \ldots \cdot x_N} = \sqrt[N]{\prod_{i=1}^{N} x_i}$$

Diviene

$$\mu_g = \sqrt[N]{x_1^{n_1} \cdot x_2^{n_2} \cdot \ldots \cdot x_k^{n_k}} = \sqrt[N]{\prod_{i=1}^{N} x_i^{n_i}}$$

$$\mu_q = \sqrt{\frac{x_1^2 + x_2^2 + \ldots + x_N^2}{N}} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} x_i^2}$$

Diviene

$$\mu_q = \sqrt{\frac{x_1^2 \cdot n_1 + x_2^2 \cdot n_2 + \ldots + x_k^2 \cdot n_k}{N}} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} x_i^2 \cdot n_i}$$

7. IL CASO DELLE DISTRIBUZIONI DI FREQUENZE CON MODALITÀ RAGGRUPPATE IN CLASSI

Quando le modalità sono raggruppate in classi il calcolo della media è necessariamente approssimato.

j	$Classi\ X$	Frequenze (assolute)
1	$c_0 - c_1$	n_1
2	$c_1 - c_2$	n_2
⋮	⋮	⋮
k	$c_{k-1} - c_k$	n_k
Totale		N

Sia μ_1 la media del carattere nelle n_1 unità appartenenti alla classe (c_0, c_1), μ_2 la media del carattere nelle n_2 unità appartenenti alla classe (c_1, c_2), e così via.

Il prodotto $\mu_i \cdot n_i$ fornisce il totale del carattere posseduto dalle unità della classe (c_{i-1}, c_i) con $i = 1,2,\ldots,k$ e quindi la somma $\mu_1 \cdot n_1 + \mu_2 \cdot n_2 + \ldots + \mu_k \cdot n_k = \sum_{i=1}^{k} \mu_i n_i$ è il totale del carattere dell'intera distribuzione. Ma le quantità μ_i non sono generalmente conosciute. Per questo assumiamo che essere siano approssimativamente uguali ai valori centrali delle classi assumiamo cioè che

$$\mu_i \approx \frac{c_{i-1} + c_i}{2}$$

Posto allora $\acute{x}_i = \frac{c_{i-1} + c_i}{2}$ la media aritmetica è data da $\mu = \frac{1}{N} \sum_{i=1}^{k} \acute{x}_i \cdot n_j$

È immediato estendere quanto detto per la media aritmetica alle altre medie analitiche basta sostituire a x_i i valori centrali di classe x_i .

$$\mu_a = \frac{N}{\sum_{i=1}^{k} \frac{n_i}{x_i}}$$

$$\mu_g = \sqrt[N]{\prod_{i=1}^{N} \acute{x}_i^{\,n_i}}$$

$$\mu_q = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \acute{x}_i^{2} \cdot n_i}$$

8. MEDIANA

Il calcolo della mediana richiede che sia formata preliminarmente la graduatoria dei termini della distribuzione.

Sia $x_1, x_2, \ldots, x_N$ una distribuzione statistica disaggregata, sia $x_{(1)}, x_{(2)}, \ldots, x_{(N)}$ con $x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(N)}$ la corrispondente distribuzione dei termini ordinati in senso non decrescente:

- Se N è dispari chiama mediana della distribuzione la quantità m che occupa il posto centrale cioè il posto $\frac{(N+1)}{2}$ Della graduatoria dei termini ordinati;
- Se N è pari si assume come mediana la media aritmetica dei termini che occupano i due posti centrali della graduatoria dei termini ordinari ossia le posizioni $\frac{N}{2}$ e $\frac{N}{2}+1$.

In sintesi

$$m = \begin{cases} x_{\left(\frac{N+1}{2}\right)}, & \text{se } N \text{ è dispari} \\ \frac{1}{2}\left[x_{\left(\frac{N}{2}\right)} + x_{\left(\frac{N}{2}+1\right)}\right], & \text{se } N \text{ è pari} \end{cases}$$

Osserviamo che la mediana come quantità che si colloca al centro della graduatoria dei termini della distribuzione è una misura sintetica dell'ordine di grandezza del fenomeno oggetto di studio.

La mediana presenta le proprietà di seguito indicate:

- È interna tra il minimo e il massimo dei termini della distribuzione: $x_{(1)} \leq m \leq x_{(N)}$
- Se si trasformano i dati secondo la funzione $y_i = a + bx$ con $i = 1,2,\ldots,N$ la mediana dei dati così trasformati m_Y è legata a quella dei dati iniziali e m_X dalla relazione $m_Y = a + b\, m_X$
- Tale proprietà è chiamata proprietà di linearità che implica come casi particolari ponendo una funzione di trasformazione $b = 1$ nel primo caso e $a = 0$ nel secondo caso:
  - la proprietà di traslatività: se si aggiunge o si sottrae una portante la mediana della nuova distribuzione è uguale alla mediana della distribuzione iniziale aumentata o diminuita di tale quantità;
  - la proprietà di omogeneità: se i termini della distribuzione sono moltiplicati per una costante la mediana della nuova distribuzione e la costante on volte la mediana della distribuzione iniziale.

La mediana è il valore *c* che minimizza la somma dei valori assoluti degli scarti:

$$|x_i - m| = \min_c \sum_{i=1}^{N} ¿\, x_i - c \vee ¿ \sum_{i=1}^{N} ¿$$

9. QUARTILI

I quartili sono le tre quantità che suddividono la graduatoria dei termini della distribuzione in quattro parti.

Sia $x_1, x_2, \ldots, x_N$ una distribuzione statistica disaggregata; sia $x_{(1)}, x_{(2)}, \ldots, x_{(N)}$ la corrispondente distribuzione dei termini ordinati in senso non decrescente, cioè tali che $x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(N)}$. Presa la quantità $N \cdot \frac{l}{4}$ con $l = 1,2,3$ e individuati i due interi consecutivi $h-1$ e h , tali che $h - 1 \leq N \cdot \frac{l}{4} < h$, definiamo come *l*-esimo quartile della distribuzione la quantità

$$q_l = \begin{cases} x_{(h)}, & \text{se } N \cdot \frac{l}{4} > h-1 \\ \frac{1}{2}\left[x_{(h-1)} + x_{(h)}\right], & \text{se } N \cdot \frac{l}{4} = h-1 \end{cases}$$

N.B. il secondo quartile coincide con la mediana.

Quando N è un multiplo di 4:

1. il primo quartile può essere visto come l'elemento separatore tra il ¼ dei termini ordinati e i restanti ¾;
2. il secondo quartile come l'elemento separatore tra la prima metà dei termini ordinati e la restante metà;
3. il terzo quartile come l'elemento separatore tra i primi ¾ dei termini ordinati e il restante ¼.

Detto questo si può affermare che esattamente:

1. ¼ dei termini della distribuzione assume un valore minore o uguale a q_1;
2. La metà dei termini della distribuzione presenta un valore minore o uguale di q_2;
3. ¾ dei termini della distribuzione assume un valore minore o uguale di q_3.

Se N non è multiplo di 4 si può affermare che almeno:

1. ¼ dei termini della distribuzione assume un valore minore o uguale a q_1;
2. La metà dei termini della distribuzione presenta un valore minore o uguale di q_2;
3. ¾ dei termini della distribuzione assume un valore minore o uguale di q_3.

10. QUANTILI

L'idea dei quartili può essere estesa:

- Per definire i 9 **decili** come le quantità che suddividono la graduatoria dei termini della distribuzione in 10 parti e per calcolare il decile *l*-esimo ($l=1,2,\ldots,9$), d_l , si ricercano i due interi consecutivi, $h-1$ e h , tali che

$$h-1\leq N\cdot\frac{l}{10}<h$$ e si assegna a d_l il valore $$d_l=\begin{cases} x_{(h)}, & \text{se } N\cdot\frac{l}{10}>h-1 \\ \frac{1}{2}\left[x_{(h-1)}+x_{(h)}\right], & \text{se } N\cdot\frac{l}{10}=h-1 \end{cases}$$

- Per definire i 99 **centili** come le quantità che suddividono la graduatoria dei termini della distribuzione in 100 parti e per calcolare il decile *l*-esimo ($l=1,2,\ldots,99$), c_l , si ricercano i due interi consecutivi, $h-1$ e h , tali che

$$h-1\leq N\cdot\frac{l}{100}<h$$ e si assegna a c_l il valore $$d_l=\begin{cases} x_{(h)}, & \text{se } N\cdot\frac{l}{100}>h-1 \\ \frac{1}{2}\left[x_{(h-1)}+x_{(h)}\right], & \text{se } N\cdot\frac{l}{100}=h-1 \end{cases}$$

- Ecc.

11. MEDIANA, QUARTILI E QUANTILI NEL CASO DELLA DISTRIBUZIONE DI FREQUENZE

Dobbiamo indicare il procedimento per il calcolo della mediana, dei quartili e dei quantili quando si opera in caso di distribuzione di frequenze, cioè su distribuzioni con modalità $x_1, x_2, \ldots, x_k$ a cui sono associate le frequenze $n_1, n_2, \ldots, n_k$, dove le modalità sono disposte in ordine crescente $x_1 < x_2 < \ldots < x_k$

1. Mediana

Si determinano prima di tutto le frequenze cumulate $N_1, N_2, \ldots, N_k$. Dopodiché si individuano le due frequenze cumulate consecutive, N_{h-1} e N_h, tali che $N_{h-1} \leq N \cdot \frac{1}{2} < N_h$.

Infine, si assegna alla mediana il valore

$$m = \begin{cases} x_h, & \text{se } N \cdot \frac{1}{2} > N_{h-1} \\ \frac{1}{2}\left(x_{h-1} + x_h\right), & \text{se } N \cdot \frac{1}{2} = N_{h-1} \end{cases}$$

Dove x_{h-1} e x_h sono le modalità del carattere a cui corrispondono le frequenze cumulate N_{h-1} e N_h.

2. Quartili

Una volta determinate le frequenze cumulate, si individuano le due frequenze cumulate consecutive N_{h-1} e N_h tali che $N_{h-1} \leq N \cdot \frac{l}{4} < N_h$ con $l = 1,2,3$, dopodiché:

$$q_l = \begin{cases} x_h, & \text{se } N \cdot \frac{l}{4} > N_{h-1} \\ \frac{1}{2}\left(x_{h-1} + x_h\right), & \text{se } N \cdot \frac{l}{4} = N_{h-1} \end{cases}$$

3. Quantili

Sia v il numero di gruppi in cui è ripartita la popolazione e sia $l=1,\dots,(v-1)$ l'indice del quantile scelto. Sia $x_1, x_2, \dots, x_k$ una distribuzione con frequenze $n_1, n_2, \dots, n_k$ si cerca $h: N_{h-1} \leq N \cdot \frac{l}{v} < N_h$. Il quantile $q\left(\frac{l}{v}\right)$ è definito come segue

$$q\left(\frac{l}{v}\right)=\begin{cases} x_h, & \text{se } N \cdot \frac{l}{v} > N_{h-1} \\ \frac{1}{2}\left(x_{h-1}+x_h\right), & \text{se } N \cdot \frac{l}{v} = N_{h-1} \end{cases}$$

Confronto tra funzione di ripartizione e quantili per una distribuzione di frequenze

Sia v il numero di gruppi in cui è ripartita la popolazione e sia $l=1,\dots,(v-1)$ l'indice del quantile scelto. Sia $x_1, x_2, \dots, x_k$ una distribuzione con frequenze $n_1, n_2, \dots, n_k$ si cerca $h: F_{h-1} \leq \frac{l}{v} < F_h$. Il quantile $q\left(\frac{l}{v}\right)$ è definito come segue

$$q\left(\frac{l}{v}\right)=\begin{cases} x_h, & \text{se } \frac{l}{v} > F_{h-1} \\ \frac{1}{2}\left(x_{h-1}+x_h\right), & \text{se } \frac{l}{v} = F_{h-1} \end{cases}$$

Il caso delle distribuzioni di frequenze con modalità raggruppate in classi

Mediana

Quando le modalità sono raggruppate in classi, il calcolo della mediana avviene in due fasi:

- Si individua la classe mediana;
- All'interno della classe individuata di calcola il valore della mediana.

Sia $[c_{h(sx)}, c_{h(dx)}]$ la classe mediana. Nell'ipotesi di uniforme distribuzione delle unità della classe la mediana è data da

$$m = c_{h(sx)} + \frac{N \cdot \frac{1}{2} - N_{h-1}}{N_h - N_{h-1}} \left[c_{h(dx)} - c_{h(sx)}\right]$$

Si può generalizzare a un qualsiasi quantile

$$q\left(\frac{l}{v}\right)=c_{h(sx)}+\frac{N\cdot\frac{l}{v}-N_{h-1}}{N_h-N_{h-1}}\left[c_{h(dx)}-c_{h(sx)}\right]$$

la proporzione di incremento può essere calcolata anche tramite le frequenze relative cumulate

$$q\left(\frac{l}{v}\right)=c_{h(sx)}+\frac{\frac{l}{v}-F_{h-1}}{F_h-F_{h-1}}\left[c_{h(dx)}-c_{h(sx)}\right]$$

Valore centrale

Può essere utilizzato per caratteri quantitativi.

Sia data una distribuzione disaggregata $x_1,x_2,\ldots,x_i,\ldots,x_N;$ sia $x_{(1)},x_{(2)},\ldots,x_{(i)},\ldots,x_{(N)}$ la corrispondente distribuzione dei termini ordinati, con $x_{(1)}\leq x_{(2)}\leq\ldots,\leq x_{(i)}\leq\ldots,\leq x_{(N)}$. Dove $x_{(1)}$ e $x_{(N)}$ sono rispettivamente il minimo e il massimo. Il valore centrale della distribuzione è la media aritmetica dei valori estremi:

$$\mu_c=\frac{x_{(1)}+x_{(N)}}{2}$$

Moda

Può essere utilizzata per caratteri qualitativi su scala nominale, su scala ordinale, caratteri quantitativi.

La moda di una distribuzione di frequenze è la modalità che presenta la frequenza più alta

Nel caso di caratteri quantitativi raggruppati in classi si deve parlare di classe modale che è la classe che presenta la frequenza specifica più alta.

CAPITOLO 5 INDICI DI VARIABILITÀ

Introduzione

Per variabilità si intende l'**attitudine** dei fenomeni a manifestarsi in modi **differenti.**

Se il fenomeno considerato è rappresentato tramite una variabile quantitativa X, la variabilità di X è l'attitudine di X ad assumere **valori** differenti.

Considerare solo una **media** vuol dire considerare solo l'informazione relativa all'**ordine di grandezza** del fenomeno, ma ciò non vuol dire che tutte le osservazioni siano uguali alla media. Al contrario una distribuzione è composta da diversi valori e un indice di variabilità è un'informazione aggiuntiva circa la loro diversità.

Ma distribuzioni diverse possono avere stessa media e stesso indice di variabilità, quindi qualche informazione si perde comunque. Medie e indici di variabilità rappresentano una **sintesi**.

Classificazione degli indici di variabilità

Un problema preliminare è quello di individuare i criteri per la costruzione di indici idonei a misurare il fenomeno della variabilità, a questo fine considerando una generica distribuzione disaggregata si possono individuare 3 categorie di indici di variabilità:

1. Indici di dispersione o **scostamenti** medi;
2. Indici di disuguaglianza o **differenze** medie;
3. Indici basati sugli **intervalli** di variazione.

A questi si aggiungono

1. Indici di **concentrazione** per caratteri **trasferibili**;
2. Indici di **eterogeneità** per caratteri **qualitativi**.

Scostamenti medi

Gli scostamenti sono rappresentati da $|X-c|$, dato il carattere X si opera una traslazione $-c$ calcolando così quanto i valori di X si allontanano dal valore centrale c. Vogliamo calcolare delle **distanze**, quindi non ci interessa il segno, ma solo l'intensità, il valore assoluto.

Dal momento che parliamo di scostamenti medi scegliamo una media M e l'applichiamo ai nuovi valori.

Facendo un'opportuna sintesi di tali distanze tramite una media, si ottiene una misura di variabilità che si annulla in caso di assenza di variabilità ed è tanto maggiore quanto maggiore è la distanza media dei termini della distribuzione della loro media.

Il valore centrale di riferimento per calcolare gli scostamenti è la media aritmetica $c=\mu$.

Quindi data la distribuzione statistica disaggregata $x_1, x_2, \ldots, x_N$ consideriamo le distanze dei termini x_i dalla media aritmetica

$$|x_1-\mu|, |x_2-\mu|, \ldots, |x_N-\mu|$$

Queste quantità indicano il grado di diversità dei termini della distribuzione rispetto alla media aritmetica e sono nulle quando $x_i=\mu$.

Una prima misura di variabilità è lo scostamento semplice medio.

Data la distribuzione statistica disaggregata $x_1, x_2, \ldots, x_N$ si chiama scostamento semplice medio la media aritmetica degli scarti dalla media presi in valore assoluto (una volta calcolati gli scostamenti ne valutiamo la media aritmetica):

$$S_\mu = \frac{|x_1-\mu|+|x_2-\mu|+\ldots+|x_N-\mu|}{N} = \frac{1}{N}\sum_{i=1}^{N} \text{¿}\, x_i-\mu \vee \text{¿}$$

Una seconda misura di variabilità è lo scostamento quadratico medio o deviazione standard.

Data la distribuzione statistica disaggregata $x_1, x_2, \ldots, x_N$, si chiama scostamento quadratico medio o deviazione standard la media quadratica degli scarti $(x_1-\mu), (x_2-\mu), \ldots, (x_N-\mu)$ (una volta calcolati gli scostamenti ne valutiamo la media quadratica):

$$\sigma = \sqrt{\frac{(x_1-\mu)^2+(x_2-\mu)^2+\ldots+(x_N-\mu)^2}{N}} \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2}$$

La suddetta formula può essere anche espressa nella forma:

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2-\mu^2} = \sqrt{\mu_q^2-\mu^2}$$

Che chiameremo formula operativa perché non coinvolgendo gli scarti della media rende più rapido il calcolo.

Gli indici di variabilità appena illustrati hanno le proprietà si seguito indicate:

1. Assumono il valore 0 nel caso di assenza di variabilità, cioè quando tutti i termini della distribuzione sono uguali:

2. $x_1=x_2=\cdots=x_N=c \qquad \mu=\frac{1}{N}\sum_{i=1}^{N} c=\frac{Nc}{N}=c \qquad \text{¿}\, x_i-\mu \vee \text{¿}=\sum_{i=1}^{N}|c-c|=0$

$\sum_{i=1}^{N} \text{¿}$

3. Non cambiano se a ciascun termine della distribuzione si aggiunge una quantità costante positiva o negativa (n*on importa se la distribuzione viene spostata a sinistra o a destra sull'asse dei numeri reali, l'attitudine ad assumere valori differenti è la stessa*);
4. La moltiplicazione di ciascun termine della distribuzione per una costante, positiva o negativa, ha come conseguenza la moltiplicazione degli indici per il valore assoluto della costante (l*a costante moltiplicativa cambia la scala, quindi anche l'ampiezza degli scarti dalla media aritmetica. Il valore negativo ruota le osservazioni sull'asse dei numeri reali ma le distanze sono valutate in valore assoluto*).

Il quadrato dello scostamento quadratico medio

$$\sigma^2=\left(\sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2}\right)^2=\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2$$

Ossia la media aritmetica dei quadrati degli scarti è chiamata **varianza**.

Un'altra quantità è la **devianza**, si tratta della somma dei quadrati degli scarti della media ovvero il numeratore della varianza:

$$D=(x_1-\mu)^2+(x_2-\mu)^2+\ldots+(x_N-\mu)^2=\sum_{i=1}^{N}(x_i-\mu)^2$$

Il caso delle distribuzioni di frequenze

Per una distribuzione di frequenze le formule dello scostamento semplice medio e dello scostamento quadratico medio vanno opportunamente adattate.

Scostamento semplice medio

$\text{¿}\, x_j-\mu \vee \text{¿}\, n_j$

$S_\mu=\frac{1}{N}\sum_{i=1}^{N} \text{¿}$

Scostamento quadratico medio in caso di distribuzione di frequenze

$$\sigma=\sqrt{\frac{1}{N}\sum_{j=1}^{k}(x_j-\mu)^2 n_j}$$

La quale può essere riscritta nella seguente forma

$$\sigma=\sqrt{\sum_{j=1}^{k}(x_j-\mu)^2 f_j}$$

Dove $f_j=\frac{n_j}{N}$ è la frequenza relativa alla generica modalità x_j .

La formula operativa diviene:

$$¿\sqrt{\frac{1}{N}\sum_{j=1}^{k} x_i{}^2 n_j-\mu^2}$$

Il caso delle distribuzioni di frequenze con modalità raggruppate in classi

Scostamento semplice medio

$$¿\acute{x}_j-\mu\vee¿n_j$$
$$S_\mu=\frac{1}{N}\sum_{i=1}^{N}¿$$

Scostamento quadratico medio

$$\sigma=\sqrt{\frac{1}{N}\sum_{j=1}^{k}(\acute{x}_j-\mu)^2 n_j}$$

Dove $\acute{x}_j$ rappresenta i valori centrali delle classi $\acute{x}_j=\frac{c_{i-1}+c_i}{2}$ con $j=1,2,\ldots,k$

Differenze medie

Un secondo indice di variabilità è la differenza semplice media.

Data una distribuzione statica disaggregata possiamo pensare di costruire un indice di variabilità confrontando il singolo termine con tutti gli altri termini della distribuzione. In questo modo il termine x_1 dà luogo alle $N-1$ differenze

$|x_1-x_2|,|x_1-x_3|,\ldots,|x_1-x_N|$

E così via.

Si ottengono così $N(N-1)$ differenze, ciascuna delle quali indica il grado di diversità tra gli elementi di una coppia di termini.

Una misura di variabilità può allora essere costruita facendo una opportuna sintesi si queste $N(N-1)$ misure di diversità.

Sia $x_1,x_2,\ldots,x_N$ una distribuzione statistica disaggregata. Si chiama differenza media semplice della distribuzione, la media aritmetica delle differenze in valore assoluto, $|x_i-x_j|, i,j=1,2,\ldots,N$, tra le $N(N-1)$ coppie di termini della distribuzione:

$$\Delta=\frac{1}{N(N-1)}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}|x_i-x_j|$$

Come media delle $N(N-1)$ distanze $|x_i-x_j|, i,j=1,2,\ldots,N$ la differenza semplice media può essere interpretata come distanza media.

Le differenze sono a due a due uguali in quanto per ogni coppia $i,j, |x_i-x_j|=|x_j-x_i|$, ne segue che possiamo riscrivere la formula nella forma equivalente

$$\Delta=\frac{2}{N(N-1)}\sum_{i=2}^{N}\sum_{j=1}^{i-1}|x_i-x_j|$$

La differenza semplice media ha le stesse proprietà dello scostamento quadratico medio:

1. Assume valore 0 nel caso di assenza di variabilità, cioè quando tutti i termini della distribuzione sono uguali;
2. Non cambia se a ciascun termine della distribuzione si aggiunge una quantità costante positiva o negativa;
3. La moltiplicazione di ciascun termine della distribuzione per una costante, positiva o negativa, ha come conseguenza la moltiplicazione degli indici per il valore assoluto della costante.

Il caso delle distribuzioni di frequenze

$$\Delta=\frac{2}{N(N-1)}\sum_{i=2}^{k}\sum_{j=1}^{i-1}|x_i-x_j|n_i n_j$$

Nel caso di dati raggruppati in classi è necessario individuare un valore rappresentativo (valore centrale in caso di assenza di altre informazioni, valore medio nel caso in cui si conoscano i totali di classe).

Campo di variazione e differenza interquartile

Esistono indicatori di variabilità di una distribuzione dove vengono talvolta assunte le differenze tra coppie di valori che occupano particolari posizioni nella graduatoria non decrescente dei termini della distribuzione stessa, e sono:

- Campo di variazione;
- Differenza interquartile.

Sia $x_1, x_2, \ldots, x_N$ una distribuzione statistica disaggregata, sia $x_{(1)}, x_{(2)}, \ldots, x_{(N)}$, con $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(N)}$, la corrispondente distribuzione dei termini disposti in senso non decrescente. Si chiama **campo di variazione** la differenza tra il valore più grande e quello più piccolo della distribuzione: $\Delta_c = x_{(N)} - x_{(1)}$.

Si tratta dell'ampiezza dell'intervallo entro cui sono contenuti tutti i termini della distribuzione. L'indice dipende unicamente dai valori estremi e assume lo stesso valore per tutte le distribuzioni che hanno gli stessi estremi indipendentemente dai valori degli altri N-2 termini.

Sia $x_1, x_2, \ldots, x_N$ una distribuzione statistica disaggregata. Si chiama **differenza interquartile** la differenza tra il terzo e il primo quartile della distribuzione: $\Delta_q = q_3 - q_1$ essendo q_1 e q_3 il primo e il terzo quartile.

Nell'intervallo delimitato da q_1 e q_3 è compresa la metà dei termini della distribuzione, cosicché, a parità di altre condizioni, la variabilità è tanto più grande quanto maggiore è la differenza tra q_3 e q_1.

Indici percentuali di variabilità

Gli indici finora esaminati, denominati tradizionalmente indici assoluti di variabilità, sono espressi nell'unità di misura in cui è espresso il carattere e risentono dell'ordine di grandezza della distribuzione. Ma questi indici non sono utilizzabili per il confronto di più distribuzioni quando queste presentano medie sensibilmente diverse oppure quando le unità di misura sono diverse. Un modo per risolvere il problema consiste nell'esprimere gli indici come percentuali della media e in questo modo otteniamo i cosiddetti **indici percentuali di variabilità**.

Si chiama indice percentuale di variabilità il rapporto, moltiplicato per 100, tra gli indici assoluti di variabilità introdotti in precedenza e la media aritmetica. Si ha così la *deviazione standard percentuale*, la *differenza media percentuale*, il *campo di variazione percentuale* e la *media interquartile percentuale*.

Particolare rilievo per le applicazioni è il coefficiente di variazione

$$CV = \frac{\sigma}{\mu} \cdot 100$$

$$S_{\mu_Y} = \left| \frac{100}{\mu} \right| S_{\mu_X};$$

$$CV = \sigma_Y = \left| \frac{100}{\mu} \right| \sigma_X;$$

$$\Delta_Y = \left| \frac{100}{\mu} \right| \Delta_X;$$

$$\Delta_{c_Y} = \left| \frac{100}{\mu} \right| \Delta_{c_X};$$

$$\Delta_{q_Y} = \left| \frac{100}{\mu} \right| \Delta_{q_X}$$

Concentrazione: variabilità per distribuzione secondo caratteri trasferibili

I caratteri trasferibili sono caratteri, come il reddito e il patrimonio delle famiglie, per i quali ha senso parlare di trasferibilità cioè della possibilità che una unità ceda una parte di carattere a un'altra unità. Sono caratteri quantitativi su scala di rapporti (Tizio può possedere il doppio di Caio e se Sempronio non possiede nulla si ha assenza del fenomeno) che assumono valori non negativi (non si può possedere una risorsa in quantità negativa, a meno che non si considerino i debiti... ma in tal caso si creerebbero delle eccezioni sulle proprietà elencate in seguito).

La **concentrazione** si può intendere come l'attitudine di un carattere a essere posseduto da un numero ristretto di unità, quindi una distribuzione presenta una concentrazione tanto più elevata quanto maggiore è la frazione (del totale) del carattere posseduta dalle unità con le modalità più alte o in modo equivalente quanto minore è la frazione (del totale) del carattere posseduta dalle unità con modalità più basse.

Gli indici di concentrazione sono indici di variabilità che godono delle proprietà per distribuzione secondo caratteri trasferibili.

Sia $x_1, x_2, \ldots, x_N$ una distribuzione disaggregata secondo un carattere, X, trasferibile. Sia $x_{(1)}, x_{(2)}, \ldots, x_{(N)}$, con $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(N)}$ la stessa distribuzione dei termini disposti in senso non decrescente. Immaginiamo l'insieme delle possibili distribuzioni di N termini aventi lo stesso totale $N \cdot \mu$.

Due distribuzioni particolari di questo insieme sono quella in cui tutti i termini sono uguali tra loro $x_{(1)} = x_{(2)} = \ldots = x_{(N)}$ e si parla di equidistribuzione, e quella in cui il totale del carattere è posseduto da una sola unità $x_{(1)} = x_{(2)} = \ldots = x_{(N-1)} = 0$ e $x_{(N)} = N \cdot \mu$ e si parla di massima concentrazione.

Misura della concentrazione nel caso di distribuzioni disaggregate

È necessario definire preliminarmente le grandezze che sono alla base della misura della misura della concentrazione. Data una distribuzione statistica disaggregata di termini ordinati $x_{(1)}, x_{(2)}, \ldots, x_{(N)}$, consideriamo la somma $A_i = x_{(1)} + x_{(2)} + \ldots + x_{(i)}$, quantità che rappresenta l'ammontare del carattere posseduto dalle i unità meno dotate, cioè dalle unità aventi modalità non superiore a $x_{(i)}$. Ovviamente, $A_N = N \cdot \mu$.

Chiameremo *i*-esima frazione di carattere la quantità $Q_i = \frac{A_i}{A_N}$ con $i = 1, 2, \ldots, N$.

Naturalmente $Q_N = 1$. Nel caso di equidistribuzione la precedente formula assume la forma $Q_i = \frac{A_i}{A_N} = \frac{i \cdot \mu}{N \cdot \mu} = \frac{i}{N} = P_i$. Nel seguito la quantità $P_i = \frac{i}{N}$ sarà chiamata *i*-esima frazione di unità. Quindi si comprende facilmente che la concentrazione è tanto maggiore quanto più le Q_i differiscono dalle P_i. Queste quantità sono alla base del **rapporto di concentrazione di Gini**. Facciamo osservare che nel caso di equidistribuzione $P_i = Q_i \quad i = 1, 2, \ldots, N$, mentre nel caso di massima concentrazione $Q_1 = Q_2 = \ldots = Q_{N-1} = 0$.

Sia X un carattere trasferibile e sia $x_1, x_2, \ldots, x_N$ una distribuzione statistica disaggregata secondo il carattere X. Una misura della concentrazione di tale distribuzione è data dal rapporto di concentrazione di Gini:

$$G=\frac{\sum_{i=1}^{N-1}\left(P_i-Q_i\right)}{\sum_{i=1}^{N-1}P_i}$$

Osserviamo che vale la disuguaglianza

$$\sum_{i=1}^{N-1}\left(P_i-Q_i\right)\leq\sum_{i=1}^{N-1}P_i$$

Dove l'espressione a sinistra è un'ovvia misura di concentrazione che è nulla nel caso di equidistribuzione ed è massima, cioè pari a $\sum_{i=1}^{N-1}P_i$, nel caso di massima concentrazione. Quindi il rapporto di Gini è un indice di concentrazione che varia tra 0 e 1, essendo uguale a 0 nel caso di equidistribuzione e uguale a 1 nel caso di massima concentrazione. Per questa proprietà, l'indice G, diversamente dagli indici di variabilità assoluti, è idoneo per il confronto della concentrazione di due o più distribuzioni.

Una formula equivalente al rapporto di Gini è data da:

$$\frac{2}{N-1}\sum_{i=1}^{N-1}\left(P_i-Q_i\right)$$

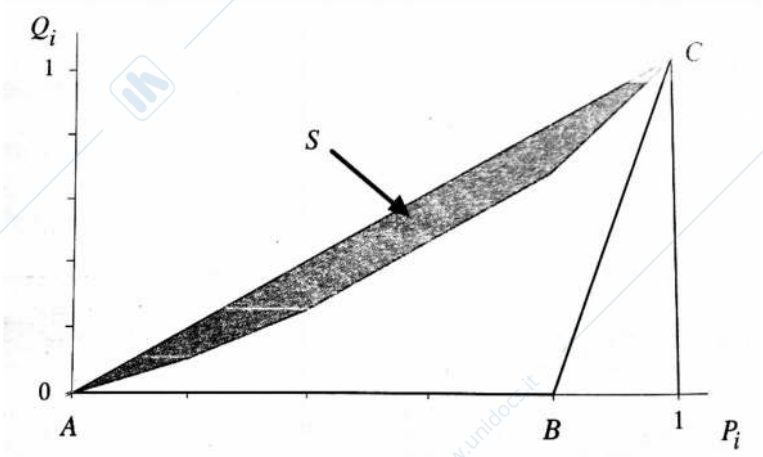
L'indice di Gini gode delle seguenti proprietà:

1. Assume valori nell'intervallo [0, 1]; è uguale a 0 nel caso di equidistribuzione, è uguale a 1 nel caso di massima concentrazione;
2. Diminuisce se si aggiunge a ogni termine della distribuzione una quantità positiva (aumenta se si sottrae a ogni termine della distribuzione una quantità positiva);
3. Non cambia se ogni termine della distribuzione è moltiplicato per una costante positiva;
4. Aumenta a seguito di un trasferimento di intensità da una data unità ad un'unità che presenta modalità non inferiore (diminuisce a seguito di un trasferimento di intensità da una data unità ad un'unità che presenta modalità non superiore).

Interpretazione geometrica del rapporto di concentrazione

La Formula del rapporto di Gini si presta a una interessante interpretazione geometrica tramite la curva di concentrazione nota anche come curva di Lorenz.

Data una distribuzione disaggregata $x_1, x_2, \ldots, x_N$ rappresentiamo su un piano cartesiano i punti di coordinate (P_i, Q_i) con $i = 1,2,\ldots,N$. Si chiama curva di concentrazione la spezzata che si ottiene congiungendo con segmenti di retta le coppie consecutive di punti di coordinate $(0,0),(P_1,Q_1),(P_2,Q_2),\ldots,(P_N,Q_N)$



Da un punto di vista geometrico l'indice di Gini viene reinterpretato in termini geometrici.

Nel caso di equidistribuzione, la curva si riduce al segmento AC detto segmento di equidistribuzione, aventi come estremi i punti $(0,0)$ e $(1,1)$.

Nel caso di massima concentrazione, la curva coincide con la spezzata ABC che unisce i punti seguenti $A=(0,0), B=\left(\frac{N-1}{N},0\right), C=(1,1)$ che viene chiamata curva di massima concentrazione.

Sia S la superficie racchiusa tra il segmento di equidistribuzione e la curva di concentrazione e sia $max\,S$ il massimo di S , allora il rapporto di concentrazione di Gini può essere espresso come $G=\frac{S}{max\,S}$ dove $max\,S=\frac{N-1}{2N}$

La quantità $max\,S$ non è altro che l'area del triangolo ABC che ha altezza 1 e base $\frac{N-1}{N}$ (lunghezza del segmento AB).

L'espressione $G=\frac{S}{max\,S}$ aiuta a capire il significato dell'indice: poiché $max\,S$ è costante, possiamo dire che G cresce con S e quindi è tanto più grande quanto più la curva di concentrazione si allontana dal segmento di equidistribuzione.

Quando N è sufficientemente grande, l'area di massima concentrazione è approssimativamente uguale a 1/2, essendo $\frac{N-1}{N}\approx 1$ ($N-1\approx N$¿ , con questa approssimazione $G=\frac{2}{N-1}\sum_{i=1}^{N-1}\left(P_i-Q_i\right)$ diviene

$$R=\frac{2}{N}\sum_{i=1}^{N-1}\left(P_i-Q_i\right)$$

Espressione che posto $P_0=Q_0=0$ possiamo scrivere nella forma

$$\frac{1}{N}[\left(P_{i-1}-Q_{i-1}\right)+¿\left(P_i-Q_i\right)]$$

$$R=\sum_{i=1}^{N} ¿$$

Tra i due indici G e R sussiste una relazione di proporzionalità

$$R=\frac{N-1}{N}G$$

Misura della concentrazione nel caso di distribuzione di frequenze

Sia data la distribuzione di frequenze secondo un carattere trasferibile. Sia N_i la generica frequenza cumulata e $A'_i=x_1\cdot n_1+x_2\cdot n_2+\ldots+x_i\cdot n_i$ la quantità di carattere posseduta complessivamente dalle unità aventi modalità non superiori a x_i .

Consideriamo i rapporti

$$P'_i=\frac{N_i}{N}, Q'_i=\frac{A'_i}{A'_k}$$

Essendo $A'_k=\sum_{j=1}^{k} x_i n_i=N\cdot\mu$, rapporti che indicano rispettivamente la frazione sul totale delle unità e la frazione sul totale del carattere. Allora il rapporto di concentrazione R assume la forma

$$R=\sum_{i=1}^{k}\left[\left(P'_{i-1}-Q'_{i-1}\right)+\left(P'_i-Q'_i\right)\right]\frac{n_i}{N}$$

Dove $P'_0=Q'_0=0$

Variabili nelle distribuzioni secondo caratteri qualitativi: eterogeneità

Per le distribuzioni di frequenze secondo un carattere qualitativo è possibile introdurre indici che misurino il grado di diversità tra le unità del collettivo. È opportuno partire dalla definizione di **omogeneità** ed **eterogeneità**.

Una distribuzione di frequenze secondo un carattere qualitativo presenta la **minima** eterogeneità, ovvero la massima omogeneità, quando *tutte le unità del collettivo hanno la stessa modalità* del carattere; per contro, la distribuzione presenta **massima** eterogeneità quando *le modalità hanno tutte la stessa frequenza*.

Ovvero quando tutte le unità del collettivo presentano la stessa modalità del carattere non vi è diversità tra le unità e quindi vi è massima omogeneità. A mano a mano che ci si allontana da questa situazione cresce l'eterogeneità tra le unità, la situazione di massima eterogeneità si realizza quando le unità del collettivo sono ripartite tra le modalità in parti uguali.

Modalit	Frequenz
à	a
x_1	0
x_2	0
…	…
x_i	N
…	…
x_k	0
TOTALE	N

Schema di distribuzione di frequenza con omogeneità massima

Modalit	Frequenz
à	a
x_1	N/k
x_2	N/k
…	…
x_i	N/k
…	…
x_k	N/k
TOTALE	N

Schema di distribuzione di frequenza con eterogeneità massima

I due indici idonei a misurare il grado di eterogeneità di una distribuzione sono noti come indice di eterogeneità di Gini (e_1) ed entropia (e_2):

$$e_1 = 1 - \sum_{j=1}^{k} f_j^2$$

$$e_2 = -\sum_{i=1}^{k} f_j \ln(f_j)$$

Dove f_i con $i = 1,2,\ldots,k$ sono le frequenze relative.

L'indice di eterogeneità di Gini raggiunge il minimo (0) nel caso di omogeneità e raggiunge il massimo ($k-1/k$) nel caso di massima eterogeneità.

L'entropia raggiunge il minimo (0) nel caso di omogeneità e raggiunge il massimo ($\ln(k)$) nel caso di massima eterogeneità.

CAPITOLO 6 INDICI DI FORMA

Le medie fanno l'idea dell'origine di grandezza del fenomeno studiato.

Gli indici di variabilità segnalano il grado di diversità tra le singole manifestazioni del fenomeno.

Gli indici di forma completano il quadro delle tecniche per l'analisi e la comprensione delle caratteristiche di una distribuzione statistica secondo un carattere quantitativo.

Due sono gli aspetti che con questi indici si studiano:

1. Il primo riguarda il maggiore o minore allontanamento della distribuzione dalla forma simmetrica;
2. Il secondo è relativo al grado di deviazione della distribuzione da una particolare distribuzione teorica, la distribuzione normale.

Asimmetria

Definiamo la simmetria.

Sia data una distribuzione di frequenze secondo un carattere quantitativo discreto

Modalità	Frequenza

x_1	n_1
x_2	n_2
…	…
x_i	n_i
…	…
x_k	n_k
TOTALE	N

Avente mediana m .

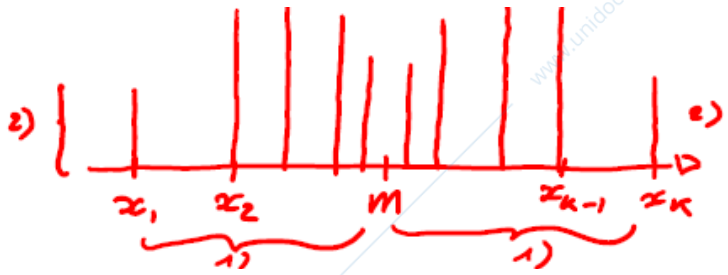
Consideriamo le seguenti coppe di modalità:

- La prima e l'ultima (x_1, x_k) ;
- La seconda e la penultima (x_2, x_{k-1}) ;
- La terza e la terzultima (x_3, x_{k-2}) ;
- Ecc… (x_j, x_{k-j+1}) .

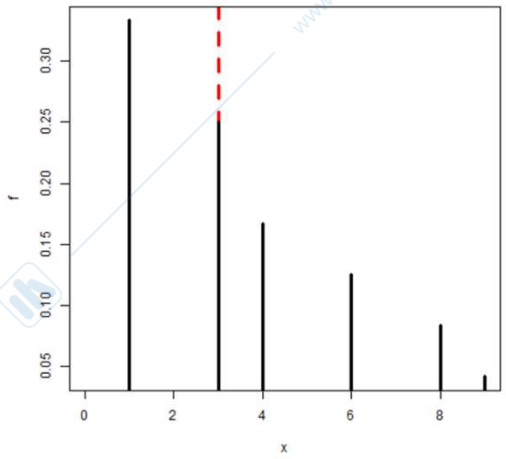
La distribuzione si dice simmetrica se per ciascuna coppia le modalità sono equidistanti dalla mediana e hanno stessa frequenza.

In termini formali si dice che la distribuzione è simmetrica se valgono simultaneamente le seguenti identità:

1. $¿ x_j - m \vee ¿ ¿ x_{k-j+1} - m \vee ¿$;
2. $n_j = n_{k-j+1}$ e $f_j = f_{k-j+1}$.

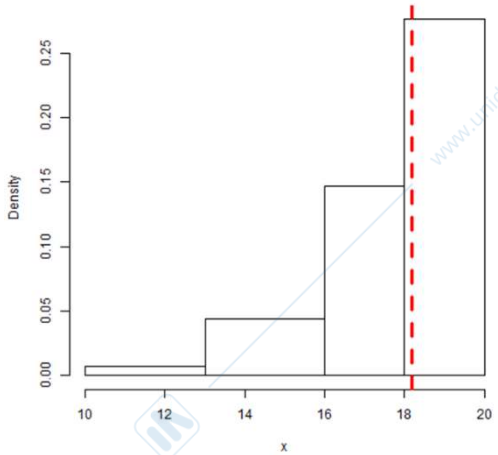


Con linguaggio geometrico, possiamo anche dire che una distribuzione è simmetrica quando l'asse di simmetria suddivide il suo grafico in due parti tali che la parte destra è l'immagine speculare di quella sinistra.



Per i caratteri divisi in intervalli la definizione di simmetria va adattata sostituendo al termine modalità quello di classe e al termine frequenza quello di densità di frequenza.

1. $\begin{matrix} +¿-m\vee¿ \\ -¿-m\vee¿¿l_{k-j+1}^{¿} \\ ¿l_j^{¿} \end{matrix}$ e $\begin{matrix} -¿-m\vee¿ \\ +¿-m\vee¿\vee l_{k-j+1}^{¿} \\ ¿l_j^{¿} \end{matrix}$;
2. $h_j=h_{k-j+1}$ e $\frac{h_j}{N}=\frac{h_{k-j+1}}{N}$.



Una distribuzione simmetrica ha le seguenti proprietà:

a) La media aritmetica coincide con la mediana $\mu=m$;

b) La somma degli scarti dalla media aritmetica elevati a una potenza dispari è uguale a 0 , per esempio: $\sum_{i=1}^{k}\left(x_i-\mu\right)^3 n_i=0$;

c) Il primo e il terzo quartile hanno la stessa distanza dalla mediana $|q_1-m|=|q_3-m|$ ossia (poiché $q_1 \leq m \leq q_3$ ¿ $(m-q_1)-(q_3-m)=0$.

In assenza di simmetria le distribuzioni si dicono asimmetriche.

Per introdurre il concetto degli indici di asimmetria dobbiamo parlare di asimmetria positiva e asimmetria negativa.

Un criterio qualitativo per distinguere l'asimmetria positiva da quella negativa si richiama alla rappresentazione grafica della distribuzione, preso come riferimento il punto dell'asse delle ascisse che rappresenta la mediana, si ha asimmetria positiva se il grafico tende a prolungarsi a destra di tale punto che non a sinistra e si ha asimmetria negativa nel caso opposto.

A questo punto presentiamo gli indici di asimmetria premettendo che questi indici sono nulli in caso di simmetria, assumono valori positivi in caso di asimmetria positiva e negativi in caso di asimmetria negativa.

1. $a_1=\frac{\mu-m}{\sigma}$;

Se vi è simmetria l'indice si annulla per la proprietà a) delle distribuzioni simmetriche. Il segno dell'indice tende a essere positivo quando la distribuzione presenta asimmetria positiva perché in questo caso la media aritmetica tende a superare la mediana; viceversa, l'indice assume valori negativi nel caso di asimmetria negativa.

2. $a_2=\frac{1}{\sigma^3}\left[\frac{1}{N}\sum_{j=1}^{k}(x_j-\mu)^3 n_j\right]$;

Rappresenta la media aritmetica degli scarti standardizzati elevati al cubo e come è tale è un numero puro e non risente della variabilità della distribuzione. Il valore che esso assume non consente di valutare se l'asimmetria è più o meno forte. È molto sensibile alla presenza nella distribuzione di valori eccezionalmente piccoli o grandi a causa dell'elevazione al cubo degli scarti corrispondenti a questi valori.

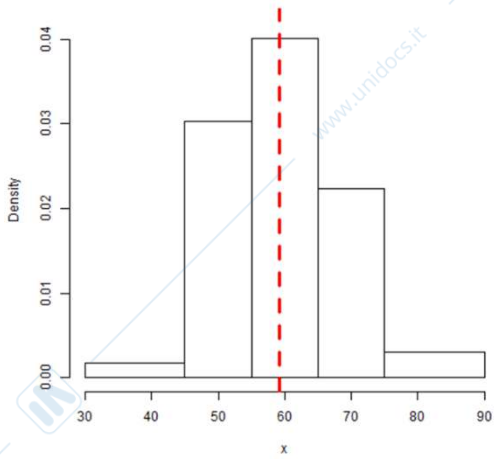
3. $a_3=\frac{(q_3-m)-(m-q_1)}{(q_3-m)+(m-q_1)}=\frac{q_3+q_1-2m}{q_3-q_1}$.

Si annulla nelle distribuzioni simmetriche per la proprietà c) delle distribuzioni simmetriche. Il segno dell'indice tende a essere positivo quando il grafico si prolunga a destra della mediana, la distanza tra il 3° quartile e la mediana tende ad essere superiore alla distanza tra il 1° e la mediana, il segno dell'indice tende ad essere

negativo in caso di fenomeno opposto. Questo indice essendo normalizzato ci consente di dire se l'asimmetria della distribuzione è più o meno forte, per fare ciò basta confrontare il valore che l'indice assume con gli estremi del suo campo di variazione (-1,1).

Curtosi

Per curtosi si intende il grado di diversità della distribuzione di frequenze effettiva rispetto al modello teorico rappresentato dalla distribuzione normale. La distribuzione normale è un modello matematico che può essere impiegato per approssimare una distribuzione reale. Consideriamo la distribuzione per classi

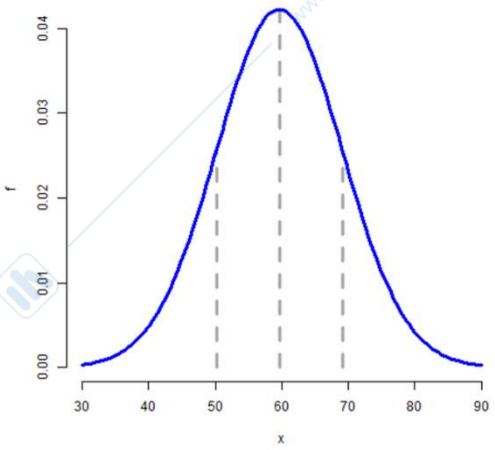


Abbiamo costruito il grafico ponendo sull'asse delle ordinate le densità di frequenza $\frac{f_i}{d_i}$ in modo tale che la superficie complessiva dell'istogramma sia uguale a 1. Esso descrive al variare di x la densità di frequenza densità che è costante all'interno di ciascuna classe.

Si voglia ora approssimare l'istogramma con la curva di equazione

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\left(\frac{x-\mu}{\sigma}\right)^2}$$

Detta funzione di densità normale dove x assume tutti i valori dell'asse reale mentre μ e σ sono due costanti numeriche che sono interpretabili come la media aritmetica e la deviazione standard della distribuzione teorica. La funzione presenta la forma a campana.



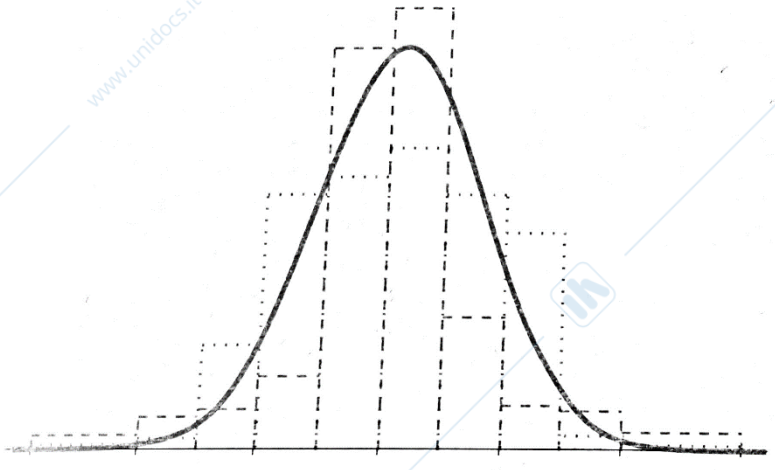
La media aritmetica coincide con il punto di massima densità mentre la deviazione standard determina il grado di concentrazione della curva attorno alla media, a valori crescenti di σ corrispondono curve sempre più schiacciate verso l'asse delle ascisse.

Appiattimento della distribuzione, con minore densità al centro e maggiore densità sulle code: distribuzione **platicurtica**.

Allungamento della distribuzione al centro, con maggiore densità al centro e minore densità sulle code: distribuzione **leptocurtica**.

Il modello **Normale** viene considerato come modello distributivo di riferimento

- nel caso di distribuzione platicurtica si dice anche distribuzione iponormale
- nel caso di distribuzione leptocurtica si dice anche distribuzione ipernormale



Normale (—),
ipernormale (- - -),
iponormale (· · ·).

Una misura della curtosi è data dall'indice

$$\gamma = \frac{1}{\sigma^4}\left[\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^4\right] \qquad \gamma = \frac{1}{\sigma^4}\left[\frac{1}{N}\sum_{j=1}^{k}(x_j-\mu)^4 n_j\right]$$

Ipotizzando una distribuzione perfettamente Normale questo indice sarebbe pari a 3.

$$\gamma=\frac{1}{\sigma^4}\left[\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^4\right]-3 \qquad \gamma=\frac{1}{\sigma^4}\left[\frac{1}{N}\sum_{j=1}^{k}(x_j-\mu)^4 n_j\right]-3$$

L'indice assume valori vicini allo 0 se la distribuzione è molto prossima alla normale.

È negativo $\gamma<0$ in caso di distribuzione platicurtica o iponormale.

È positivo $\gamma>0$ in caso di distribuzione leptocurtica o ipernormale.

Standardizzazione

Per poter confrontare, almeno qualitativamente, la forma di due distribuzioni conviene eliminare i fattori che potrebbero oscurarne le differenze. Per questo motivo è opportuno traslare e trasformare la scala di misurazione in modo da uniformare l'indicatore di posizione ed il grado di dispersione, così da effettuare confronti a parità di media (convenzionalmente posta uguale a zero) e di scarto quadratico medio (posto uguale a 1). Questa operazione si chiama standardizzazione e consiste nel trasformare le modalità di un carattere X attraverso la trasformazione lineare.

Applicando la standardizzazione a diversi insiemi di dati, ne segue che le eventuali differenze emergeranno solamente calcolando medie di ordine superiore a due.

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{\left(\frac{x-\mu}{\sigma}\right)^2}$$

La media di ordine 3 dei valori standardizzati può essere usata come indice di asimmetria.

$a_2=\frac{1}{\sigma^3}\left[\frac{1}{N}\sum_{j=1}^{k}(x_j-\mu)^3 n_j\right]=\frac{1}{N}\sum_{j=1}^{k}\left(\frac{x_j-\mu}{\sigma}\right)^3 n_j$ La media di ordine 4 dei valori standardizzati può essere usata come indice di curtosi.

$$\gamma=\frac{1}{\sigma^4}\left[\frac{1}{N}\sum_{j=1}^{k}(x_j-\mu)^4 n_j\right]-3=\gamma=\frac{1}{N}\sum_{j=1}^{k}\left(\frac{x_j-\mu}{\sigma}\right)^4 n_j-3$$

Sia $x_1,x_2,\ldots,x_N$ una distribuzione disaggregata con media μ e scarto quadratico medio σ, allora (dalle proprietà già note di media e scarto quadratico medio)

$z_i=\frac{x_i-\mu}{\sigma}=\frac{1}{\sigma}x_i-\frac{\mu}{\sigma}, i=1,2,\ldots,N$ è una distribuzione con media 0 e scarto quadratico medio 1. La standardizzazione può essere definita nel caso di distribuzioni di frequenza.

Momenti

Sia $M(\cdot)$ è la media aritmetica applicata alla trasformazione indicata tra parentesi del carattere.

Momenti dall'origine di ordine r $M(X^r)$

Le medie potenziate sono $\mu_r = \sqrt[r]{M(X^r)}$

La formula indiretta per il calcolo della varianza è $\sigma^2 = M(X^2) - M(X)^2$

Sia $\mu = M(X)$

Momenti centrali $M[(X-\mu)^r]$

la varianza è $\sigma^2 = M[(X-\mu)^2]$

Sia $\sigma = \sqrt{M[(X-\mu)^2]}$

Momenti standardizzati $M\left[\left(\frac{X-\mu}{\sigma}\right)^r\right]$

gli indici di asimmetria e curtosi sono $a_2 = M\left[\left(\frac{X-\mu}{\sigma}\right)^3\right]; \gamma = M\left[\left(\frac{X-\mu}{\sigma}\right)^4\right] - 3$

CAPITOLO 8 NUMERI INDICI

I numeri indici sono particolari rapporti statistici. I numeri indici elementari esprimono in termini percentuali le variazioni del fenomeno ponendo uguale a 100 il livello del fenomeno stesso a un tempo di riferimento (mese, anno, ...) che può essere fisso o variabile.

Faremo costante riferimento alle serie storiche che sono particolari distribuzioni statistiche costituite dalle intensità di un dato fenomeno associate a tempi successivi (anni, mesi, ...).

Una serie storica viene schematizzata:

Tempo	1	2	⋯	t	⋯	k
Livello del fenomeno	a_1	a_2	⋯	a_t	⋯	a_k

dove $1,2,\dots,t,\dots,k$ indicano i tempi e $a_1,a_2,\dots,a_t,\dots,a_k$ le intensità del fenomeno oggetto d'indagine riferite ai rispettivi tempi.

Siano $a_1,a_2,\dots,a_k$ le intensità di un dato fenomeno nei tempi $1,2,\dots,k$. Si chiamano **numeri indici a base fissa** i rapporti tra le intensità del fenomeno un una di esse. Se come base si assume il tempo h ($h=1,2,\dots,k$¿ , i numeri indici con base h sono espressi da:

$$❑_h I_1=\frac{a_1}{a_h},❑_h I_2=\frac{a_2}{a_h},\dots,❑_h I_k=\frac{a_k}{a_h}$$

Di solito si assume $h=1$.

Con i numeri indice a base fissa, le intensità del fenomeno vengono raffrontate tutte con l'intensità del tempo assunto come base. In alcune situazioni risulta più conveniente mettere a confronto la singola intensità con l'intensità del tempo immediatamente precedente, in questo modo si ottengono i numeri indici a base mobile.

Siano $a_1,a_2,\dots,a_k$ le intensità di un dato fenomeno nei tempi $1,2,\dots,k$. Si chiamano numeri indici a base mobile i rapporti tra le intensità del fenomeno e le intensità al tempo precedente:

$$❑_1 I_2=\frac{a_2}{a_1},❑_2 I_3=\frac{a_3}{a_2},\dots,❑_{t-1} I_t=\frac{a_t}{a_{t-1}},\dots,❑_h I_k=\frac{a_k}{a_h}$$

Cambiamento di base

Data una serie di numeri indici con base h , è possibile passare a una nuova base h' dividendo ogni numero indice per $❑_{h'} I_h$, cioè per il numero indice della nuova base rispetto alla vecchia. Con riferimento al generico tempo t , il procedimento è compendiato nella formula

$$❑_{h'} I_t=\frac{❑_h I_t}{❑_h I_{h'}}\cdots=\frac{\frac{a_t}{a_h}}{\frac{a_{h'}}{a_h}}=\frac{a_t}{a_h}\cdot\frac{a_h}{a_{h'}}=\frac{a_t}{a_{h'}}$$

Cambiamento da base fissa a base mobile

Dati i numeri indici a base fissa $\square_h I_1, \square_h I_2, \ldots \square_h I_t, \ldots, \square_h I_k$, i numeri indici a base mobile $\square_1 I_2, \square_2 I_3, \ldots \square_{t-1} I_t, \ldots, \square_{k-1} I_k$, si possono calcolare dividendo il singolo numero indice a base fissa per il precedente. Per il generico tempo t ($t=2,3,\ldots,k$) vale l'identità

$$\frac{\square_h I_t}{\square_h I_{t-1}} = \frac{a_t/a_h}{a_{t-1}/a_h} = \square_{t-1} I_t$$

Cambiamento da base mobile a base fissa

Data una serie di numeri indice a base mobile $\square_1 I_2, \square_2 I_3, \ldots, \square_{t-1} I_t$, ..., $\square_{k-1} I_k$, h e t , il numero indice del tempo t con base h è dato da

$$\square_h I_t = \begin{cases} \left(\prod_{i=t+1}^{h} \square_{i-1} I_i \right)^{-1} & t<h \\ 1 & t=h \\ \prod_{i=h+1}^{t} \square_{i-1} I_i & t>h \end{cases}$$

Variazioni relative

Nel caso delle serie storiche fissati due tempi h e t con $h<t$, la differenza relativa tra i livelli del fenomeno nei tempi h e t è data da

$$\square_h v_t = \frac{a_t - a_h}{a_h} = \frac{a_t}{a_h} - 1 = \square_h I_t - 1$$

È verrà chiamata variazione relativa, oppure incremento relativo, se positiva, e decremento relativo se negativa.

Variazioni relative medie

Dati due tempi h e t con $h<t$ se la differenza $t-h$ è maggiore o uguale a due ha senso parlare di variazione relativa media per unità di tempo.

Data una serie storica, fissati due tempi h e t con $h<t$ siano $i_{h+1}, i_{h+2}, \ldots, i_t$ i numeri indici a base mobili riferiti ai tempi $h+1, h+2, \ldots t$. Allora la variazione relativa media dal tempo h al tempo t è data da

$$\square_h \acute{v}_t = \square_h \acute{I}_t - 1 = \sqrt[t-h]{\prod_{i=h+1}^{t} \square_{i-1} I_i} - 1$$

Questa formula è il numero indice a base mobile medio diminuito di 1 e come tale dà la variazione relativa media dal tempo h al tempo t riferita all'unità di tempo (anno, mese, ecc.).

Si noti che

$$\prod_{i=h+1}^{t} {}_{i-1}I_i = \prod_{i=h+1}^{t} \frac{a_i}{a_{i-1}} = \frac{a_{h+1}}{a_h} \cdot \frac{a_{h+2}}{a_{h+1}} \cdots \frac{a_t}{a_{t-1}} = \frac{a_t}{a_h} = {}_hI_t$$

da cui

$${}_h\dot{I}_t = \sqrt[t-h]{{}_hI_t}$$

e

$${}_h\dot{v}_t = {}_h\dot{I}_t - 1 = \sqrt[t-h]{{}_hI_t} - 1 = \sqrt[t-h]{\frac{a_t}{a_h}} - 1$$

Dove a_h e a_t sono i livelli del fenomeno nei tempi h e t.

Questa formula consente di calcolare la variazione relativa media dei livelli assunti dal fenomeno nei tempi h e t, oppure in funzione al numero indice ${}_hI_t$.

Numeri indici complessi

Il problema che affronteremo è quello della misura della variazione relativa media di un insieme di grandezze tra due tempi successivi.

Indichiamo c il numero di variabili considerate indicate con l'indice $j = 1,2,\ldots,c$ (il testo li indica k e i ma si rischia di confonderle con il numero di rilevazioni temporali).

Siano $p_b = (p_{1b}, p_{2b}, \ldots, p_{cb})$ i prezzi di k beni e servizi riferiti al tempo b e $p_t = (p_{1t}, p_{2t}, \ldots, p_{ct})$ i prezzi degli stessi beni e servizi riferiti al tempo successivo t (con $b<t$ (successivo)).

Per semplicità i tempi b e t saranno chiamati tempo base e tempo corrente, rispettivamente.

Ai fini della costruzione dei numeri indici complessi è necessario introdurre le grandezze $q_b = (q_{1b}, q_{2b}, \ldots, q_{cb})$ che rappresentano le quantità dei k beni e servizi riferite al tempo base e $q_t = (q_{1t}, q_{2t}, \ldots, q_{ct})$ che rappresentano le stesse quantità di beni e servizi riferite al tempo corrente.

Immaginiamo che i beni e servizi vengono acquistati al tempo b e di rivenderli interamente al tempo t mentre tra i due tempi ci sono state altre transazioni che ne hanno modificato la distribuzione delle quantità.

Complessivamente al tempo b abbiamo speso $\sum_{j=1}^{c} p_{jb} \cdot q_{jb}$ ed al tempo t abbiamo incassato $\sum_{j=1}^{c} p_{jt} \cdot q_{jt}$. Il numero indice calcolato sul totale è dato da:

$$ {}_{b}^{\square}I_t = \frac{\sum_{j=1}^{c} p_{jt} \cdot q_{jt}}{\sum_{j=1}^{c} p_{jb} \cdot q_{jb}} $$

Un primo modo per sintetizzare i numeri indici elementari di ciascun bene e servizio $\frac{p_{1t}}{p_{1b}}, \frac{p_{2t}}{p_{2b}}, \ldots, \frac{p_{kt}}{p_{kb}}$ è quello di fare una media aritmetica ponderata assumendo come pesi $p_{1b} \cdot q_{1b}, p_{2b} \cdot q_{2b}, \ldots, p_{kb} \cdot q_{kb}$ ($p_{jb} q_{jb}$) come indicato nell'espressione:

$$ I_L = M\left({}_{b}^{\square}I_{t_j}\right) = \frac{\sum_{j=1}^{c} {}_{b}^{\square}I_{t_j} \cdot p_{jb} q_{jb}}{\sum_{j=1}^{c} p_{jb} q_{jb}} = \frac{\sum_{j=1}^{c} \frac{p_{jt}}{p_{jb}} \cdot p_{jb} q_{jb}}{\sum_{j=1}^{c} p_{jb} q_{jb}} = \frac{\sum_{j=1}^{c} p_{jt} q_{jb}}{\sum_{j=1}^{c} p_{jb} q_{jb}} $$

Questa formula è nota come indice di Laspeyres.

Naturalmente la variazione media dei prezzi dal tempo base b al tempo corrente t si ottiene sottraendo 1.

Un'altra possibilità è quella di far intervenire nella struttura della formula di sintesi le quantità al tempo corrente utilizzando i valori dei beni del tempo corrente dati da $p_{1t} \cdot q_{1t}, p_{2t} \cdot q_{2t}, \ldots, p_{kt} \cdot q_{kt}$ ($p_{jt} q_{jt}$) come pesi per il calcolo di una media aritmetica ponderata dei numeri indici elementari di ciascun bene e servizio $\frac{p_{1t}}{p_{1b}}, \frac{p_{2t}}{p_{2b}}, \ldots, \frac{p_{kt}}{p_{kb}}$ e si ottiene l'espressione

$$ I_P = M\left({}_{b}^{\square}I_{t_j}^{-1}\right)^{-1} = \left(\frac{\sum_{j=1}^{c} {}_{b}^{\square}I_{t_j}^{-1} \cdot p_{jt} q_{jt}}{\sum_{j=1}^{c} p_{jt} q_{jt}} \right)^{-1} = \left(\frac{\sum_{j=1}^{c} {}_{t}^{\square}I_{b_j} \cdot p_{jt} q_{jt}}{\sum_{j=1}^{c} p_{jt} q_{jt}} \right)^{-1} = ¿ $$

$$¿\left(\frac{\sum_{j=1}^{c}\frac{p_{jb}}{p_{jt}}\cdot p_{jt}q_{jt}}{\sum_{j=1}^{c}p_{jt}q_{jt}}\right)^{-1}=\left(\frac{\sum_{j=1}^{c}p_{jb}q_{jt}}{\sum_{j=1}^{c}p_{jt}q_{jt}}\right)^{-1}=\frac{\sum_{j=1}^{c}p_{jt}q_{jt}}{\sum_{j=1}^{c}p_{jb}q_{jt}}$$

Questa formula è nota come indice di Paasche.

Anche in questo caso la variazione relativa media dei prezzi si ottiene sottraendo 1.

Va osservato che ai fini dell'applicazione delle due formule non è necessario conoscere le quantità: per l'indice di Laspeyres è sufficiente disporre dei valori dei beni e servizi al tempo base e dei numeri indici elementari; per l'indice di Paasche basta conoscere i valori dei beni e servizi al tempo corrente e i numeri indici elementari.

È stata inoltre proposta la cosiddetta formula ideale di Fisher data dalla media geometrica degli indici di Laspayres e di Paasche:

$$I_F=\sqrt{I_L\cdot I_P}$$

CAPITOLO 9 ANALISI DELLE DISTRIBUZIONI DOPPIE: DIPENDENZA

Definiamo la distribuzione doppia disaggregata come l'insieme delle coppie di modalità di due caratteri X e Y osservate nelle unità del collettivo statistico.

Una distribuzione doppia disaggregata verrà rappresentata come $(x_1,y_1¿,(x_2,y_2),\ldots,(x_N,y_N)$ dove i deponenti $1,2,\ldots,N$ denotano le unità statistiche.

Una distribuzione doppia di frequenze viene rappresentata dalla seguente tabella

Carattere X	Carattere Y						Totale
	y_1	y_2	$\cdots$	y_j	$\cdots$	y_t	
x_1	n_{11}	n_{12}	$\cdots$	n_{1j}	$\cdots$	n_{1t}	n_{10}
x_2	n_{21}	n_{22}	$\cdots$	n_{2j}	$\cdots$	n_{2t}	n_{20}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_i	n_{i1}	n_{i2}	$\cdots$	n_{ij}	$\cdots$	n_{it}	n_{i0}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_s	n_{s1}	n_{s2}	$\cdots$	n_{sj}	$\cdots$	n_{st}	n_{s0}
Totale	n_{01}	n_{02}	$\cdots$	n_{0j}	$\cdots$	n_{0t}	N

Schema di distribuzione doppia di frequenze.

Dove n_{ij} (chiamata frequenza congiunta) è il numero delle unità del collettivo che presentano le modalità x_i del carattere X e la modalità y_i del carattere Y, n_{i0} denota la somma delle frequenze della riga che corrisponde alla modalità x_i e come tale rappresenta il numero delle unità del collettivo che presentano le modalità x_i del carattere X, n_{0j} denota la somma delle frequenze della colonna che corrisponde alla modalità y_i e come tale rappresenta il numero delle unità del collettivo che corrisponde alla modalità y_i del carattere Y.

Una distribuzione doppia di frequenze come quella rappresentata nella tabella viene denominata tabella a doppia entrata o tabella di contingenza.

Distribuzioni marginali, distribuzioni condizionate e indipendenza statistica

Distribuzioni marginali

Se si concentra l'attenzione su uno dei due caratteri e si trascura l'altro si configura la cosiddetta distribuzione marginale.

Data una distribuzione doppia di frequenze di N unità, si chiama distribuzione marginale del carattere X la distribuzione statica semplice (distribuzione secondo un solo carattere) delle N unità del collettivo secondo tale carattere.

$X \setminus Y$	y_1	y_2	$\cdots$	y_j	$\cdots$	y_t	Tot.
x_1	n_{11}	n_{12}	$\cdots$	n_{1j}	$\cdots$	n_{1t}	n_{10}
x_2	n_{21}	n_{22}	$\cdots$	n_{2j}	$\cdots$	n_{2t}	n_{20}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
x_i	n_{i1}	n_{i2}	$\cdots$	n_{ij}	$\cdots$	n_{it}	n_{i0}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
x_s	n_{s1}	n_{s2}	$\cdots$	n_{sj}	$\cdots$	n_{st}	n_{s0}
Tot.							N

La somma per riga della tabella a doppia entrata restituisce il numero di osservazioni per le diverse righe, ovvero le diverse modalità del carattere X senza considerare le modalità del carattere Y. Si chiamano frequenze **marginali** di X. La loro somma è pari al numero complessivo di osservazioni N.

$$\sum_{j=1}^{t} n_{ij} = n_{i0} \; \forall i = 1, \ldots, s$$

$X \setminus Y$	y_1	y_2	$\cdots$	y_j	$\cdots$	y_t	Tot.
x_1	n_{11}	n_{12}	$\cdots$	n_{1j}	$\cdots$	n_{1t}	n_{10}
x_2	n_{21}	n_{22}	$\cdots$	n_{2j}	$\cdots$	n_{2t}	n_{20}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
x_i	n_{i1}	n_{i2}	$\cdots$	n_{ij}	$\cdots$	n_{it}	n_{i0}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
x_s	n_{s1}	n_{s2}	$\cdots$	n_{sj}	$\cdots$	n_{st}	n_{s0}
Tot.	n_{01}	n_{02}	$\cdots$	n_{0j}	$\cdots$	n_{0t}	N

La somma per colonna della tabella a doppia entrata restituisce il numero di osservazioni per le diverse colonne, ovvero le diverse modalità del carattere Y senza considerare le modalità del carattere X. Si chiamano frequenze **marginali** di Y. La loro somma è pari al numero complessivo di osservazioni N.

$$\sum_{i=1}^{s} n_{ij} = n_{0j} \; \forall \, j = 1, \ldots, t$$

La corrispondente distribuzione di frequenze relative che chiameremo distribuzione marginale relativa del carattere X da un lato e distribuzione marginale del carattere Y dall'altro e che indicheremo rispettivamente con f_x e f_y si ottiene dividendo le frequenze assolute della distribuzione marginale di X e Y per N.

Quindi $f_X = \left\{ \frac{n_{i0}}{N}, i = 1, \ldots, s \right\}$ e $f_Y = \left\{ \frac{n_{0j}}{N}, j = 1, \ldots, t \right\}$

$X \setminus Y$	y_1	y_2	$\cdots$	y_j	$\cdots$	y_t	Tot.
x_1	f_{11}	f_{12}	$\cdots$	f_{1j}	$\cdots$	f_{1t}	f_{10}
x_2	f_{21}	f_{22}	$\cdots$	f_{2j}	$\cdots$	f_{2t}	f_{20}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
x_i	f_{i1}	f_{i2}	$\cdots$	f_{ij}	$\cdots$	f_{it}	f_{i0}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
x_s	f_{s1}	f_{s2}	$\cdots$	f_{sj}	$\cdots$	f_{st}	f_{s0}
Tot.	f_{01}	f_{02}	$\cdots$	f_{0j}	$\cdots$	f_{0t}	1

Dividendo per N tutti e le frequenze della tabella a doppia entrata si ottengono le rispettive frequenze relative

$$f_{ij}=\frac{n_{ij}}{N} \forall i=1,\dots,s \text{ e } j=1,\dots,t;$$

$$\sum_{j=1}^{t} f_{ij}=f_{i0} \forall i 1,\dots,s$$

$$\sum_{i=1}^{s} f_{ij}=f_{0j} \forall j=1,\dots,t$$

$$\sum_{j=1}^{t}\sum_{i=1}^{s} f_{ij}=1$$

Distribuzione condizionata

Nel calcolo di frequenze relative è possibile circoscrivere l'attenzione alle specifiche modalità del carattere X o del carattere Y .

Quindi la distribuzione condizionata del carattere Y associata alla modalità x_i del carattere X è data dalla tabella:

Modalità Y	**Frequenza**
y_1	n_{i1}
y_2	n_{i2}
...	…
y_i	n_{ij}
…	…
y_t	n_{it}
TOTALE	n_{i0}

La corrispondente distribuzione di frequenze relative che chiameremo distribuzione condizionata relativa del carattere Y e che indichiamo con $f_{Y \vee x_j}$ si ottiene dividendo le frequenze assolute della distribuzione precedente per n_{i0} .

$$f_{Y \vee x_i} = \left\{ \frac{n_{ij}}{n_{i0}}, j = 1, \ldots, t \right\}$$

Analogamente la distribuzione condizionata del carattere Y associata alla modalità x_i del carattere X è data dalla tabella:

Modalità X	Frequenza
x_1	n_{1j}
x_2	n_{2j}
...	...
x_i	n_{ij}
...	...
x_t	n_{sj}
TOTALE	n_{0j}

chiameremo distribuzione condizionata relativa al carattere X e che indichiamo con $f_{X \vee y_j}$ si ottiene dividendo le frequenze assolute della distribuzione precedente per n_{0j} .

$$f_{X \vee y_j} = \left\{ \frac{n_{ij}}{n_{0j}}, i = 1, \ldots, s \right\}$$

In conclusione, $f_{X \vee y_j}$ è la distribuzione di frequenze relative del carattere X circoscrivendo l'attenzione alle sole unità statistiche che per il carattere Y presentano la modalità y_j . Quindi ci sono tante distribuzioni condizionate di X tante quante sono le modalità di Y , cioè t . E $f_{Y \vee x_i}$ è la distribuzione di frequenze relative del carattere Y circoscrivendo l'attenzione alle sole unità statistiche che per il carattere X presentano la modalità x_i . Quindi ci sono tante distribuzioni condizionate di Y tante quante sono le modalità di X, cioè s .

Indipendenza statistica

Per stabilire se esiste una qualche forma di associazione tra i caratteri X e Y si può accertare se le distribuzioni condizionate di uno dei due caratteri dipendono dalle modalità dell'altro.

Data una tabella di contingenza si dice che i caratteri X e Y sono statisticamente indipendenti se le frequenze interne della tabella sono pari a $n_{ij}=\frac{n_{i0}\cdot n_{0j}}{N}$ per ogni coppia (i,j).

In una tabella di contingenza le seguenti tre condizioni sono equivalenti:

4. $f_{X\vee y_j}=f_X \forall j=1,\dots,t$
5. $f_{Y\vee x_i}=f_Y \forall i=1,\dots,s$
6. $n_{ij}=\frac{n_{i0}\cdot n_{0j}}{N} \forall i=1,\dots,t\ j=1,\dots,s$

Nel caso di indipendenza tra X e Y , tutte le distribuzioni condizionate relative di Y , al variare della modalità x_i di X , sono uguali alla distribuzione marginale relativa di Y , come pure tutte le distribuzioni condizionate relative di X al variare della modalità y_j di Y sono uguali alla distribuzione marginale relativa di X .

Questo mette in evidenza il ruolo completamente simmetrico che giocano i due caratteri X e Y nel concetto di indipendenza statistica. Nel caso in cui una qualunque delle tre condizioni non sia verificata si dice che esiste dipendenza statistica tra X e Y .

Misura della dipendenza

Per la costruzione di un indice che misuri il grado di dipendenza tra due caratteri, la via più immediata è quella di comparare la tabella di contingenza osservata con quella di indipendenza, cioè con la tabella che si avrebbe se vi fosse indipendenza tra X e Y .

Si chiama tabella di indipendenza la tabella che presenta le stesse distribuzioni marginali di X e Y della tabella osservata e frequenze congiunte date da:

$$\hat{n}_{ij}=\frac{n_{i0}\cdot n_{0j}}{N}, \text{per ogni coppia} (i,j)$$

Le frequenze teoriche $\hat{n}_{ij}$ sono sempre strettamente maggiori di 0

Le quantità che misurano la distanza relativa tra le frequenze osservate e le frequenze teoriche:

$c_{ij}=\frac{n_{ij}-\hat{n}_{ij}}{\hat{n}_{ij}}$ sono denominate contingenze e sono alla base di un primo indice di dipendenza.

Data una tabella di contingenza, una misura della dipendenza tra i caratteri X e Y è rappresentata dall'indice:

$$\psi=\sqrt{\frac{1}{N}\sum_{i=1}^{s}\sum_{j=1}^{t}\left(\frac{n_{ij}-\hat{n}_{ij}}{\hat{n}_{ij}}\right)^2\hat{n}_{ij}}=\sqrt{\frac{1}{N}\sum_{i=1}^{s}\sum_{j=1}^{t}\frac{\left(n_{ij}-\hat{n}_{ij}\right)^2}{\hat{n}_{ij}}}=\sqrt{\sum_{i=1}^{s}\sum_{j=1}^{t}\frac{n_{ij}^2}{n_{i0}\cdot n_{0j}}-1}$$

Interpretabile come media quadratica ponderata delle contingenze c_{ij} con pesi pari a $\hat{n}_{ij}$.

Questa formula è una misura di dipendenza, in quanto è una media delle deviazioni in valore assolute tra le frequenze effettive e quelle di indipendenza, deviazioni rappresentate dalle contingenze.

Essa assume un valore tanto maggiore quanto più la tabella osservata differisce dalla tabella di indipendenza. La media quadratica è adatta allo scopo perché non dipende dai segni delle contingenze ma solo dal valore assoluto.

Dipendenza in media

Nella definizione della misura della dipendenza tra X e Y finora l'attenzione è stata rivolta ad accertare se e in che misura al variare delle modalità di un carattere mutano le distribuzioni condizionate relative dall'altro carattere. Inoltre, le formule introdotte fanno uso unicamente delle frequenze della tabella di contingenza.

Quindi invece di confrontare le distribuzioni, se almeno uno dei due caratteri è quantitativo, possiamo confrontare le medie delle distribuzioni condizionate.

Se le medie condizionate possiamo dire che il carattere del quale abbiamo calcolato le medie è indipendente in media dall'altro.

Se entrambi i caratteri sono quantitativi possiamo verificare lo stesso scambiando il carattere condizionante con quello condizionato, non è detto che ci sia simmetria nella relazione.

L'indipendenza in distribuzione implica l'indipendenza in media ed essendo simmetrica l'implica in entrambe le direzioni.

Abbiamo già introdotto le distribuzioni di Y condizionate alla i -esima modalità di X

$$f_{Y\vee x_i}=\left\{\frac{n_{ij}}{n_{i0}}, j=1,\ldots,t\right\}$$

analogamente, se Y è quantitativo, si possono calcolare le medie di Y condizionate alla i -esima modalità di X .

$$\mu_Y(x_i)=\sum_{j=1}^{t} y_j \frac{n_{ij}}{n_{i0}}=\frac{1}{n_{i0}}\sum_{j=1}^{t} y_j \cdot n_{ij}, \forall\, i=1,\dots,s$$

Inoltre, attraverso le frequenze marginali di Y è possibile calcolare la media di Y

$$\mu_Y=\frac{1}{N}\sum_{j=1}^{t} y_j n_{0j}$$

Si dice che il carattere Y è indipendente in media dal carattere X se le medie aritmetiche condizionate di Y sono uguali tra loro e conseguentemente uguali alla media di Y

$$\mu_Y(x_1)=\mu_Y(x_2)=\cdots=\mu_Y(x_i)=\cdots=\mu_Y(x_s)=\mu_Y$$

In generale considerando una distribuzione doppia di frequenze la media aritmetica della generica distribuzione condizionata di Y è data da:

$$\mu_Y(x_i)=\frac{1}{n_{i0}}\sum_{j=1}^{t} y_j n_{ij} \quad \text{con} \quad i=1,2,\dots,s$$

La media della distribuzione marginale di Y è data da

$$\mu_Y=\frac{1}{N}\sum_{j=1}^{t} y_j n_{0j}$$

Ai fini della costruzione di un indice che misuri il grado di dipendenza delle medie condizionate del carattere Y dalle modalità del carattere X , è necessario definire la dipendenza in media.

Si dice che il carattere Y dipende in media dal carattere X se le medie aritmetiche delle distribuzioni condizionate di Y sono diverse tra loro. Ci; implica che siano diverse almeno due delle medie $\mu_Y(x_1),\mu_Y(x_2),\dots,\mu_Y(x_s)$.

La dipendenza in media di Y da X (o di X da Y) non implica la dipendenza in media di X da Y (o di Y da X).

Misura della dipendenza in media

Per misurare il grado di dipendenza in media di Y da X è quindi necessario valutare l'attitudine ad assumere valori diversi dalle medie aritmetiche condizionate.

Per la proprietà associativa della media aritmetica

$\mu_Y = \frac{1}{N}\sum_{i=1}^{s} \mu_Y(x_i) n_{i0}$ quindi possiamo valutare la devianza **tra** le medie (spiegata dalla dipendenza in media)

$$D_S = \sum_{i=1}^{s} \left[\mu_Y(x_i) - \mu_Y\right]^2 n_{i0}$$

Come indice di dipendenza in media di Y da X è ragionevole prendere un'appropriata misura di diversità delle medie condizionate $\mu_Y(x_i)$. Considerando che la media della distribuzione marginale di Y è la media ponderata della s medie condizionate:

$$\mu_Y = \frac{1}{N}\sum_{i=1}^{s} \mu_Y(x_i) n_{i0}$$

La quantità:

$$D_S = \sum_{i=1}^{s} \left[\mu_Y(x_i) - \mu_Y\right]^2 n_{i0}$$

È la devianza delle medie condizionate medesime e, come tale, è una misura della loro diversità. Quindi D_S ha le caratteristiche di indice di dipendenza in media: assume il valore 0 nel caso di indipendenza in media, cioè quando le medie condizionate sono tutte uguali, ed è tanto maggiore quanto più le medie condizionate sono diverse tra loro. È preferibile, tuttavia, che assuma valori nell'intervallo [0,1]. A questo fine facciamo ricorso alla scomposizione della devianza.

La devianza della distribuzione marginale di Y può essere scritta come somma di due componenti:

$$\sum_{j=1}^{t} (y_j - \mu_Y)^2 n_{0j} = \sum_{i=1}^{s} \left[\mu_Y(x_i) - \mu_Y\right]^2 n_{i0} + \sum_{i=1}^{s}\sum_{j=1}^{t} \left[y_j - \mu_Y(x_i)\right]^2 n_{ij}$$

Il primo termine del secondo membro è la devianza delle medie condizionate; il secondo termine, somma della devianza delle s distribuzioni condizionate di Y è una misura della variabilità complessiva delle distribuzioni condizionate.

Quando il collettivo è suddiviso in sottogruppi (come nel caso del condizionamento alle modalità di un altro carattere) si dimostra che la devianza complessiva è pari alla somma tra devianza **tra** i gruppi e devianza **nei** gruppi e l'effetto incrociato tra unità di diversi gruppi è nullo.

$$D=D_S+D_R \qquad \sum_{j=1}^{t}(y_j-\mu_Y)^2 n_{0j}=\sum_{i=1}^{s}\left[\mu_Y(x_i)-\mu_Y\right]^2 n_{i0}+\sum_{i=1}^{s}\sum_{j=1}^{t}\left[y_j-\mu_Y(x_i)\right]^2 n_{ij}$$

Si parlerà di devianza totale denotata con D_Y , di devianza spiegata denotata con D_S (l'espressione devianza spiegata è dovuta alla circostanza che essa è tanto maggiore quanto più le medie condizionate sono diverse, ma la diversità delle medie condizionate è legata alla diversità delle modalità di X), di devianza residua denotata con D_R , per indicare rispettivamente il primo membro, il primo termine e il secondo termina a dx.

Una misura plausibile della dipendenza in media di Y da X è data dal rapporto di correlazione.

Si chiama rapporto di correlazione il rapporto tra la devianza spiegata e la devianza totale:

$$\eta_Y^2=\frac{DS}{DT}=\frac{\sum_{i=1}^{s}\left[\mu_Y(x_i)-\mu_Y\right]^2 n_{i0}}{\sum_{j=1}^{t}(y_j-\mu_Y)^2 n_{0j}}$$

E siccome $D_S=D_Y-D_R$ la formula diventa

$$1-\frac{\sum_{i=1}^{s}\sum_{j=1}^{t}\left[y_j-\mu_Y(x_i)\right]^2 n_{ij}}{\sum_{j=1}^{t}(y_j-\mu_Y)^2 n_{0j}}=1-\frac{DR}{DT}$$

Il rapporto di correlazione è compreso nell'intervallo [0,1]. Infatti, il numeratore varia nell'intervallo $[0,D_Y]$. Il minimo 0 è assunto quanto la devianza D_S è nulla, cioè nel caso di dipendenza in media. Il massimo 1 è raggiunto quando la devianza D_R è nulla ossia quando alla generica modalità x_i è associato un solo valore di Y ripetuto n_{i0} volte, nel qual caso si ha dipendenza perfetta. Inoltre, l'indice aumenta al crescere della variabilità delle distribuzioni condizionate.

Quindi è possibile definire l'indice di grado di dipendenza in media attraverso il rapporto di composizione

$$\eta_Y^2=\frac{DS}{DT}=\frac{\sum_{i=1}^{s}\left[\mu_Y(x_i)-\mu_Y\right]^2 n_{i0}}{\sum_{j=1}^{t}(y_j-\mu_Y)^2 n_{0j}}=¿ \qquad ¿1-\frac{DR}{DT}=1-\frac{\sum_{i=1}^{s}\sum_{j=1}^{t}\left[y_j-\mu_Y(x_i)\right]^2 n_{ij}}{\sum_{j=1}^{t}(y_j-\mu_Y)^2 n_{0j}}$$

$$f_{X\vee y_j}=\left\{\frac{n_{ij}}{n_{0j}},i=1,\ldots,s\right\}\forall\, j=1,\ldots,t \qquad \mu_X(y_j)=\frac{1}{n_{0j}}\sum_{i=1}^{s}x_i\cdot n_{ij} \qquad \mu_X=\frac{1}{N}\sum_{i=1}^{s}x_i n_{i0}=\frac{1}{N}\sum_{j=1}^{t}\mu_X(y_j)n_{0j}$$

$$\eta_X^2=\frac{\sum_{j=1}^{t}\left[\mu_X(y_j)-\mu_X\right]^2 n_{0j}}{\sum_{i=1}^{s}(x_i-\mu_X)^2 n_{i0}}$$

CAPITOLO 10 ANALISI DELLE DISTRIBUZIONI DOPPIE: REGRESSIONE

In matematica una variabile y si dice funzione di x se a ciascun valore di x corrisponde uno ed un solo valore di y, valore che si determina tramite le operazioni algebriche specificate dalla relazione funzionale che lega y e x.

È facile riconoscere la grandezza che funge da variabile indipendente e quella che assume il ruolo di variabile dipendente; è anche facile comprendere come a un dato livello della variabile indipendente corrisponda una molteplicità di valori della variabile dipendete. I legami di dipendenza di questo tipo di chiamano relazioni statistiche, in esse la risposta non è determinata univocamente dal livello della variabile indipendente.

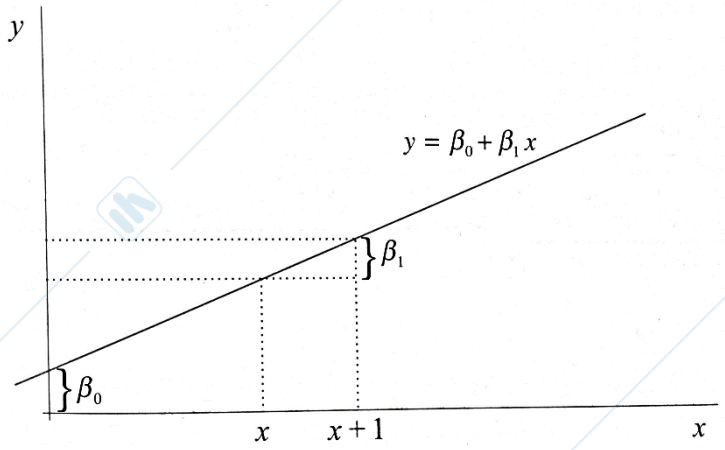
Una relazione statistica può essere descritta tramite l'equazione $y=f(x)+\varepsilon$ dove la variabile risposta y è espressa come somma di due componenti, quella rappresenta dalla funzione matematica $f(x)$ e quella residuale ε che sintetizza il contributo di tutti i fattori che potrebbero influire sulla variabile risposta e che non vengono considerati. Tale relazione definisce il **modello di regressione** di Y su X.

Utilizzeremo il caso in cui la funzione $f(x)$ è data dall'equazione della retta.

Regressione lineare

Il modello della regressione lineare consiste nell'assumere che la funzione matematica $f(x)$ che descrive la dipendenza di Y e X sia l'equazione della retta. Su questa base la relazione statistica diviene $f(x)=\beta_0+\beta_1 x$ dove β_0 e β_1 sono i parametri

della funzione. In particolare, β_0 è l'intercetta (ordinata all'origine) e β_1 il coefficiente angolare della retta.



Consideriamo ora una distribuzione doppia disaggregata costituita dalle N coppie di valori osservati $x_1, y_1 ¿, (x_2, y_2), \ldots, (x_N, y_N)$ e immaginiamo di rappresentare queste coppie di valori.

Il problema è quello di assegnare ai parametri suddetti i valori che consentano di approssimare nel miglior modo possibile la situazione reale, quindi dobbiamo determinate quella retta, tra le infinite rette del piano, che meglio si adatti ai nostri punti osservati. La soluzione è l'applicazione del metodo dei minimi quadrati.

Indicati con b_0 e b_1 due particolari valori di β_0 e β_1, siano $\hat{y}_i = b_0 + b_1 x_i \; \forall \, i = 1,2,\ldots,N$ i valori teorici di Y, ossia i valori forniti dall'equazione quando assegniamo a X i valori osservati $x_1, x_2, \ldots, x_N$.

È evidente come dai valori che si assegnano a b_0 e b_1 dipenda il grado di approssimazione dei valori teorici $\hat{y}_i$ rispetto ai valori effettivi y_i. Assumendo infatti come misura del grado di approssimazione la quantità:

$$S_q = (y_1 - b_0 - b_1 x_1)^2 + (y_2 - b_0 - b_1 x_2)^2 + \ldots + (y_N - b_0 - b_1 x_N)^2 = \sum_{i=1}^{N} (y_i - b_0 - b_1 x_i)^2$$

Si comprende che S_q è funzione di b_0 e b_1. Con il metodo dei minimi quadrati si assegnano a b_0 e b_1 i valori che rendono minima la quantità S_q, somma dei quadrati delle differenze tra i valori effettivi e i valori teorici di Y.

Data una distribuzione doppia disaggregata i valori dei parametri del modello $f(x)=\beta_0+\beta_1 x$ che minimizzano la funzione $S_q=(y_1-b_0-b_1x_1)^2+(y_2-b_0-b_1x_2)^2+\ldots+(y_N-b_0-b_1x_N)^2=\sum_{i=1}^{N}(y_i-b_0-b_1x_i)^2$ sono dati da:

$$b_1=\frac{\sum_{i=1}^{N}(x_i-\mu_X)(y_i-\mu_Y)}{\sum_{i=1}^{N}(x_i-\mu_X)^2} \quad \text{e} \quad b_0=\mu_Y-b_1\mu_X \ .$$

Essendo μ_Y e μ_X le medie aritmetiche di Y e X rispettivamente.

La quantità a numeratore della prima parte va sotto il nome di codevianza e varrà indicata con il simbolo C_{XY} (la quale può essere scritta anche nella forma $\sum_{i=1}^{N}x_i y_i-N\mu_X\mu_Y$). La quantità a denominatore è la devianza della distribuzione marginale di X e verrà indicata con il simbolo D_X . Con questa simbologia il coefficiente angolare della retta di regressione può essere espresso come $b_1=\frac{C_{XY}}{D_X}=\frac{\sigma_{XY}}{\sigma_X^2}$ dove σ_{XY} indica la cosiddetta covarianza data dalla codevianza divisa per N (la seconda espressione di b_1 si ottiene dalla prima dividendo numeratore e denominatore per N).

Una volta calcolati b_0 e b_1 l'equazione che ne risulta $\hat{y}=b_0+b_1x$ rappresenta la retta di regressione di Y su X detta anche retta interpolatrice dei minimi quadrati.

Il significato da attribuire al coefficiente della retta di regressione b_1 è poiché la retta rispecchia l'andamento medio dei dati osservati esso indica la variazione media che subisce Y quando X aumenta di una unità.

La retta di regressione passa per il baricentro della distribuzione doppia cioè per il punto di coordinate (μ_Y,μ_X), infatti essendo $b_0=\mu_Y-b_1\mu_X$ la sostituzione nella $\hat{y}=b_0+b_1x$ di μ_X a x dà $\hat{y}=\mu_Y$.

I residui

Le differenze tra i valori effettivi e i valori teorici di Y ovvero $e_i = y_i - \hat{y}_i$ sono dette residui e la loro somma è uguale a 0 : $\sum_{i=1}^{N} e_i = 0$.

I residui sono le differenze tra le osservazioni reali e le previsioni fornite dalla retta di regressione. L'avere applicato il metodo dei minimi quadrati ci consente di affermare che la somma dei quadrati dei residui è la minima possibile, nel senso che non esiste un'altra retta per cui tale somma è più piccola.

Adattamento della retta di regressione ai dati

L'analisi di regressione include come momento importante la verifica dell'idoneità del modello a rappresentare la relazione statistica tra le variabili Y e X . A questo fine introduciamo un apposito indice che misura la bontà dell'adattamento della retta di regressione ai punti osservati, per la cui costruzione ci si avvale della scomposizione della devianza.

Data una distribuzione doppia disaggregata, la devianza della variabile Y può essere così scomposta:

$$D_Y = \sum_{i=1}^{N} (y_i - \mu_Y)^2 = \sum_{i=1}^{N} (\hat{y}_i - \mu_y)^2 + \sum_{i=1}^{N} (y_i - \hat{y}_i)^2$$

Dove $\hat{y}_i$ sono i valori teorici forniti dalla retta di regressione.

La prima quantità della somma va sotto il nome di devianza spiegata (D_{SL} rappresenta la parte della variabilità totale dei valori di Y attribuite alle variabilità dei valori teorici $\hat{y}_i$ forniti dalla retta

di regressione in virtù dell'equazione $\hat{y}_i = b_0 + b_1 x_i$, alla variabilità di X) la seconda quantità è chiamata devianza residua (in quanto è la devianza della distribuzione dei residui $y_i - \hat{y}_i$ con $i = 1,2,\ldots,N$). La somma di queste due devianze è chiamata devianza totale. Quindi:

$$D_Y = D_{SL} + D_{RL}$$

Data una distribuzione doppia disaggregata una misura della bontà dell'adattamento della retta ai punti osservati, denominata indice di determinazione, è rappresentata dal rapporto:

$$r^2=\frac{\sum_{i=1}^{N}\left(\hat{y}_i-\mu_y\right)^2}{\sum_{i=1}^{N}\left(y_i-\mu_Y\right)^2}$$

In breve, $r^2=\frac{D_{SL}}{D_Y}$.

L'indice può essere scritto nella forma:

$$r^2=1-\frac{\sum_{i=1}^{N}\left(y_i-\hat{y}_i\right)^2}{\sum_{i=1}^{N}\left(y_i-\mu_Y\right)^2}$$

In breve, $r^2=1-\frac{D_{RL}}{D_Y}$.

Una terza espressione di r^2 è data da

$$r^2=\frac{\left[\sum_{i=1}^{N}\left(x_i-\mu_X\right)\left(y_i-\mu_Y\right)\right]^2}{\sum_{i=1}^{N}\left(x_i-\mu_X\right)^2\sum_{i=1}^{N}\left(y_i-\mu_Y\right)^2}$$

Quest'ultima formula è particolarmente utile per i calcoli perché è funzione delle quantità C_{XY}, D_X e D_Y.

L'indice r^2 gode delle seguenti proprietà:

- Assume valori nell'intervallo $[0,1]$;
- Assume il valore minimo, 0 , se e solo se $D_{SL}=0$, cioè se e solo se la retta di regressione è parallela all'asse delle ascisse;
- Raggiunge il valore massimo, 1 , se e solo se $D_{RL}=0$, circostanza che si verifica se e solo se tutti i residui sono uguali tra loro ed uguali alla loro media, ovvero sono uguali a 0 , pertanto se e solo se tutti i punti osservati giacciono sulla retta.

$$r^2=\frac{D_{SL}}{D_Y}=\frac{b_1^2 D_X}{D_Y}=\frac{\left(\frac{C_{XY}}{D_X}\right)^2 D_X}{D_Y}=\frac{C_{XY}^2}{D_X D_Y}=\frac{\left[\sum_{i=1}^{N}\left(x_i-\mu_X\right)\left(y_i-\mu_Y\right)\right]^2}{\sum_{i=1}^{N}\left(x_i-\mu_X\right)^2\sum_{i=1}^{N}\left(y_i-\mu_Y\right)^2}$$

La prima proprietà deriva dalla $D_Y=\sum_{i=1}^{N}(y_i-\mu_Y)^2=\sum_{i=1}^{N}(\hat{y}_i-\mu_y)^2+\sum_{i=1}^{N}(y_i-\hat{y}_i)^2$ per la quale $0\leq D_{SL}\leq D_Y$.

La seconda proprietà: dalla $r^2=\frac{\sum_{i=1}^{N}(\hat{y}_i-\mu_y)^2}{\sum_{i=1}^{N}(y_i-\mu_Y)^2}$ si desume che r^2 è nullo se e solo se la devianza spiegata è nulla, cioè se e solo se $\hat{y}_i=\mu_Y$, il che si verifica quando la retta di regressione ha il coefficiente angolare nulla (retta parallela all'asse delle ascisse).

Il caso delle distribuzioni doppie di frequenze

$X \setminus Y$	y_1	y_2	$\cdots$	y_j	$\cdots$	y_t	Tot.
x_1	n_{11}	n_{12}	$\cdots$	n_{1j}	$\cdots$	n_{1t}	n_{10}
x_2	n_{21}	n_{22}	$\cdots$	n_{2j}	$\cdots$	n_{2t}	n_{20}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
x_i	n_{i1}	n_{i2}	$\cdots$	n_{ij}	$\cdots$	n_{it}	n_{i0}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
x_s	n_{s1}	n_{s2}	$\cdots$	n_{sj}	$\cdots$	n_{st}	n_{s0}
Tot.	n_{01}	n_{02}	$\cdots$	n_{0j}	$\cdots$	n_{0t}	N

Consideriamo una distribuzione doppia di frequenze, dobbiamo estendere a questo tipo di distribuzioni tutte le formule date in precedenza, riguardanti il calcolo dei parametri della retta di regressione e la misura della bontà di adattamento: basta considerare che in luogo della generica coppia (x_i, y_i¿ della distribuzione doppia disaggregata dobbiamo considerare la stessa coppia che se presenta con una frequenza n_{ij} e che alle quantità x_i e y_i sono associate le frequenze marginali n_{i0} e n_{0j} rispettivamente.

Abbiamo allora:

$$b_1=\frac{\sum_{i=1}^{s}\sum_{j=1}^{t}(x_i-\mu_X)(y_i-\mu_Y)n_{ij}}{\sum_{i=1}^{s}(x_i-\mu_X)^2 n_{i0}} \quad \text{e} \quad b_0=\mu_Y-b_1\mu_X$$

Dove μ_X e μ_Y sono le medie aritmetiche delle distribuzioni marginali di X e di Y , rispettivamente:

$$\mu_X=\frac{1}{N}\sum_{i=1}^{s}x_i n_{i0}\,;\mu_Y=\frac{1}{N}\sum_{j=1}^{t}y_j n_{0j}$$

La devianza totale, la devianza della distribuzione marginale, la devianza spiegata e la devianza residua divengono a loro volta:

$$D_Y=\sum_{j=1}^{t}\left(y_j-\mu_Y\right)^2 n_{0j}$$

$$D_X=\sum_{i=1}^{s}\left(x_i-\mu_X\right)^2 n_{i0}$$

$$D_{SL}=\sum_{i=1}^{s}\left(\hat{y}_i-\mu_Y\right)^2 n_{i0}$$

$$D_{RL}=\sum_{i=1}^{s}\sum_{j=1}^{t}\left(y_j-\hat{y}_i\right)^2 n_{ij}$$

A questo punto otteniamo le formule di r^2 :

a) $$r^2=\frac{\sum_{i=1}^{N}\left(\hat{y}_i-\mu_y\right)^2 n_{i0}}{\sum_{i=1}^{N}\left(y_i-\mu_Y\right)^2 n_{0j}}$$

b) $$r^2=1-\frac{\sum_{i=1}^{N}\left(y_i-\hat{y}_i\right)^2 n_{ij}}{\sum_{i=1}^{N}\left(y_i-\mu_Y\right)^2 n_{0j}}$$

La curva di regressione

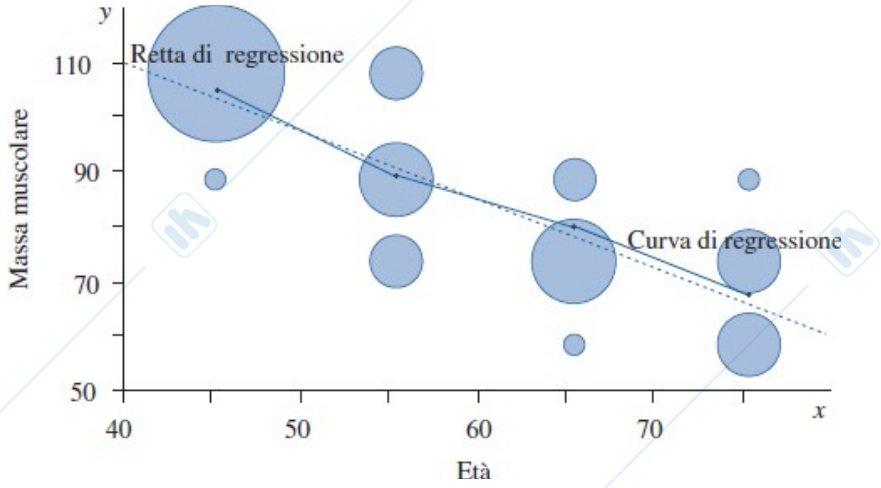
Quando si è in presenza di una distribuzione doppia di frequenze ha senso condurre un'analisi particolare associando alle modalità del carattere X le medie delle corrispondenti condizionate del carattere Y , $\mu_Y(x_i), i=1,2,\ldots,s$.

Consideriamo una distribuzione doppia di frequenze. Rappresentiamo su un piano cartesiano i punti di coordinate $\left[x_1,\mu_Y(x_1)\right],\left[x_2,\mu_Y(x_2)\right],\ldots,\left[x_s,\mu_Y(x_s)\right]$, detti punti delle medie condizionate.

Si chiama curva di regressione la funzione che assume il valore $\mu_Y(x_i)$ per $x=x_i$ e che nell'intervallo tra due modalità successive di X , x_i e x_{i+1} , si identifica con il segmento della retta che congiunge i punti di coordinate $\left[x_i,\mu_Y(x_i)\right]$ e $\left[x_{i+1},\mu_Y(x_{i+1})\right]$.

La curva di regressione è dunque formata fa un sistema di segmenti di retta, ciascuno dei quali congiunge i punti delle medie associati a due modalità contigue del carattere

X . Come tale essa mostra l'andamento delle medie condizionate del carattere Y al variare delle modalità del carattere X .



CAPITOLO 11 ANALISI DELLE DISTRIBUZIONI DOPPIE: CORRELAZIONE

Con il termine correlazione in statistica si allude alla relazione reciproca tra due caratteri quantitativi.

Indicati con X e Y i caratteri considerati, nel seguito opereremo su N coppie di valori osservati che formano una distribuzione doppia disaggregata.

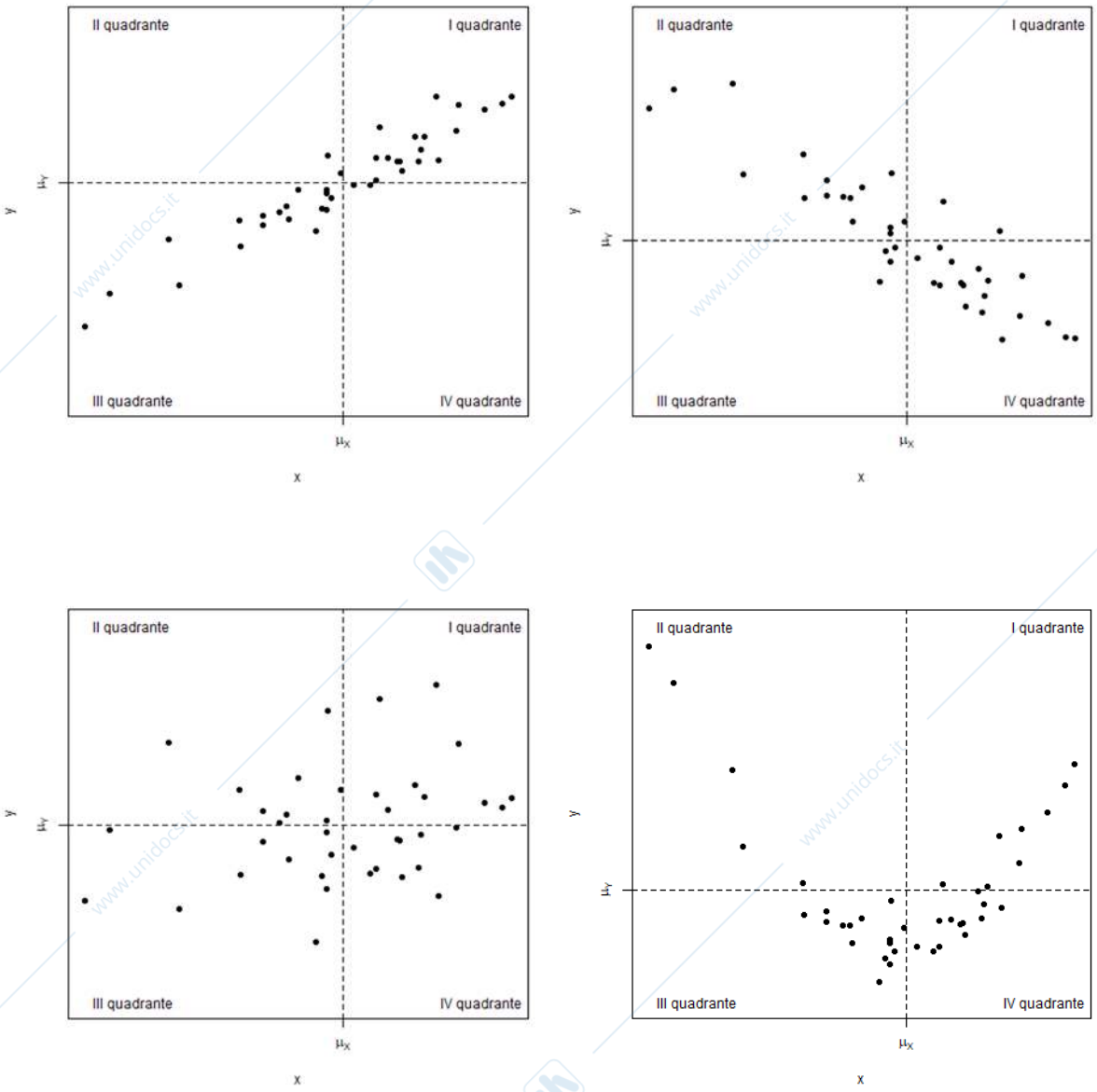
Data una distribuzione doppia in forma disaggregata si dice che tra i due caratteri X e Y vi è correlazione positiva o concordanza quando essi tendono a crescere (decrescere) insieme. Si ha una correlazione negativa o discordanza quando al crescere di un carattere l'altro tende a decrescere.

Misura della correlazione

In una distribuzione doppia disaggregata, siano μ_X e μ_Y le medie di X e Y e siano σ_X e σ_X le deviazioni standard dei medesimi caratteri e siano $z_{x_i}=\left(\frac{x_i-\mu_X}{\sigma_X}\right)$ e $z_{y_i}=\left(\frac{y_i-\mu_Y}{\sigma_Y}\right)$ gli scarti standardizzati corrispondenti alle modalità di x_i di X e y_i di Y .

Più precisamente la quantità z_{x_i} è la misura della deviazione di x_i da μ_X (negativa se $x_i<\mu_X$ e positiva nel caso opposto) rapportata alla distanza media delle modalità di X da μ_X e analogo il significato delle quantità z_{y_i} .

Ovviamente il prodotto $c_i=z_{x_i}\cdot z_{y_i}$ è positivo se i due fattori hanno lo stesso segno e negativo in caso contrario.



Nelle figure sono disegnati nuovi assi cartesiani ottenuti assumendo come origine delle ascisse e delle ordinate rispettivamente la media μ_X e la media μ_Y . Su tale piano cartesiano sono poi indicati i nuovi 4 settori.

Per le unità rappresentate dai punti che si collocano nei settori I e III le quantità di c_i sono positive mentre per le unità rappresentate dai punti che si collocano nei settori II e IV le quantità di c_i sono negative.

Adesso possiamo definire il coefficiente di correlazione lineare di Bravais come misura della correlazione tra X e Y.

Data una distribuzione doppia disaggregata, il coefficiente di correlazione lineare di Bravais è la media aritmetica delle quantità c_i definite dalla $c_i = z_{x_i} \cdot z_{y_i}$ ed è quindi dato da:

$$r = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{x_i-\mu_X}{\sigma_X}\right)\left(\frac{y_i-\mu_Y}{\sigma_Y}\right) = \frac{\sum_{i=1}^{N}(x_i-\mu_X)(y_i-\mu_Y)}{\sqrt{\sum_{i=1}^{N}(x_i-\mu_X)^2\sum_{i=1}^{N}(y_i-\mu_Y)^2}}$$

In breve, $r = \frac{C_{XY}}{\sqrt{D_X \cdot D_Y}}$

$$C_{XY} = \sum_{i=1}^{N}(x_i-\mu_X)(y_i-\mu_Y) = \sum_{i=1}^{N} x_i y_i - N\mu_X\mu_Y$$

$$D_X = \sum_{i=1}^{N}(x_i-\mu_X)^2 = \sum_{i=1}^{N} x_i^2 - N\mu_X^2 ; D_Y = \sum_{i=1}^{N}(y_i-\mu_Y)^2 = \sum_{i=1}^{N} y_i^2 - N\mu_Y^2$$

O anche

$$r = \frac{\sum_{i=1}^{N}(x_i-\mu_X)(y_i-\mu_Y)}{\sqrt{\sum_{i=1}^{N}(x_i-\mu_X)^2\sum_{i=1}^{N}(y_i-\mu_Y)^2}} = \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu_X)(y_i-\mu_Y)}{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu_X)^2}\sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i-\mu_Y)^2}} = \frac{\sigma_{XY}}{\sigma_X \cdot \sigma_Y}$$

$$\sigma_{XY} = \frac{1}{N}\sum_{i=1}^{N}(x_i-\mu_X)(y_i-\mu_Y) = \frac{1}{N}\sum_{i=1}^{N} x_i y_i - \mu_X\mu_Y$$

$$\sigma_X = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu_X)^2} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2 - \mu_X^2} ; \sigma_Y = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i-\mu_Y)^2} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} y_i^2 - \mu_Y^2}$$

Il coefficiente di correlazione lineare di Bravais è caratterizzato dalle seguenti proprietà:

- Varia nell'intervallo $[-1,1]$; i valori estremi -1 e 1 vengono raggiunti se e solo se tra le variabili X e Y c'è perfetta relazione lineare; in particolare è pari a -1 quando tutti i punti (x_i, y_i) si trovano su una retta con coefficiente angolare negativo, mentre è uguale a 1 quando tutti i punti si trovano su una retta con coefficiente angolare positivo (In quanto è la radice quadrata dell'indice di determinazione r^2 riferito alla retta di regressione di Y su X, indice che assume valori nell'intervallo $[0,1]$. Dalla stessa circostanza, si

deduce che l'indice assume i valori estremi -1 e 1 se e solo se tra X e Y c'è perfetta linearità;

- È positivo quando la retta di regressione Y su X ha coefficiente angolare positivo; è negativo nel caso opposto. La relazione tra il segno dell'indice e la pendenza della retta di regressione si desume dal fatto che il segno di r è determinato dal numeratore dell'indice, la codevianza, quantità che si trova anche a numeratore del coefficiente angolare della retta di regressione e che ne determina il segno.

  Il numeratore di r ed il numeratore di b_1 coincidono, mentre i denominatori sono sempre positivi

$$r=\frac{C_{XY}}{\sqrt{D_X D_Y}}=\frac{\sum_{i=1}^{N}(x_i-\mu_X)(y_i-\mu_Y)}{\sqrt{\sum_{i=1}^{N}(x_i-\mu_X)^2\sum_{i=1}^{N}(y_i-\mu_Y)^2}}$$

$$b_1=\frac{C_{XY}}{D_X}=\frac{\sum_{i=1}^{N}(x_i-\mu_X)(y_i-\mu_Y)}{\sum_{i=1}^{N}(x_i-\mu_X)^2}$$

- Non cambia se le modalità della singola variabile vengono moltiplicate per una costante positiva, o se vengono aumentate (diminuite) di una costante positiva.

$$V=a+bX\,;W=c+dY$$

$$r_{VW}=\frac{\frac{1}{N}\sum_{i=1}^{N}(v_i-\mu_V)(w_i-\mu_W)}{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(v_i-\mu_V)^2}\sqrt{\frac{1}{N}\sum_{i=1}^{N}(w_i-\mu_W)^2}}=¿$$

$$|b|\vee d\vee ¿\, r_{XY}$$

$$\frac{b\,d\,\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu_X)(y_i-\mu_Y)}{|b|\sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu_X)^2}\vee d\vee\sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i-\mu_Y)^2}}=\frac{b\,d}{¿}$$

Le proprietà dell'indice r sopra esposte ci consentono di affermare che la correlazione tra X e Y è:

- Negativa se $-1 \leq r < 0$;
- Nulla se $r = 0$;
- Positiva se $0 < r \leq 1$.

Il caso delle distribuzioni doppie di frequenze

In una distribuzione doppia di frequenze l'indice r assume la forma:

$$r = \frac{C_{XY}}{\sqrt{D_X \cdot D_Y}} = \frac{\sum_{i=1}^{s}\sum_{j=1}^{t}(x_i - \mu_X)(y_i - \mu_Y)n_{ij}}{\sqrt{\sum_{i=1}^{s}(x_i - \mu_X)^2 n_{i0} \sum_{j=1}^{t}(y_j - \mu_Y)^2 n_{0j}}}$$

Che può essere posta nella forma alternativa che ne facilità il calcolo

$$r = \frac{\sum_{i=1}^{s}\sum_{j=1}^{t} x_i y_j n_{ij} - N\mu_X\mu_Y}{\sqrt{\left(\sum_{i=1}^{s} x_i^2 n_{i0} - N\mu_X^2\right)\left(\sum_{j=1}^{t} y_t^2 n_{0j} - N\mu_Y^2\right)}}$$

$$C_{XY} = \sum_{i=1}^{s}\sum_{j=1}^{t}(x_i - \mu_X)(y_j - \mu_Y)n_{ij} = \sum_{i=1}^{s}\sum_{j=1}^{t} x_i y_j n_{ij} - N\mu_X\mu_Y$$

$$D_X = \sum_{i=1}^{s}(x_i - \mu_X)^2 n_{i0} = \sum_{i=1}^{s} x_i^2 n_{i0} - N\mu_X^2 \; ; \; D_Y = \sum_{j=1}^{t}(y_j - \mu_Y)^2 n_{0j} = \sum_{j=1}^{t} y_j^2 n_{0j} - N\mu_Y^2$$

CAPITOLO 12 LA PROBABILITÀ

La **statistica** è la disciplina che elabora i principi e le metodologie che presiedono al processo di rilevazione e raccolta dei dati, alla rappresentazione sintetica e alla interpretazione dei dati stessi e, laddove ve ne siano le condizioni, alla generalizzazione delle evidenze osservate.

Fino ad ora abbiamo **descritto** ciò che abbiamo osservato ma l'obiettivo non è solo descrivere ma anche *generalizzare* le evidenze osservate. I dati osservati sono solo un campione e da essi vogliamo ottenere informazioni circa l'intera popolazione e

l'**inferenza** si occupa della generalizzazione ma per fare ciò, l'inferenza necessita della **probabilità**.

Esperimenti casuali

Un **esperimento casuale** è un esperimento che può avere più esiti possibili non prevedibili con certezza. Esperimento causale indicherà ogni atto o processo idealizzato la cui singola esecuzione (prova) dà luogo a un risultato non prevedibile. La nozione di singola esecuzione implica che l'esperimento sia ipoteticamente ripetibile.

Spazio campionario ed eventi

Per configurare un esperimento casuale è necessario enumerare tutti i possibili risultati dell'esperimento stesso.

I diversi esiti rappresentano gli **eventi elementari** (generalmente indicati con la lettera greca minuscola ω).

L'insieme di tutti gli eventi elementari è lo **spazio campionario** (indicato con la lettera greca maiuscola Ω). Lo spazio campionario è l'insieme di tutti i possibili risultati dell'esperimento casuale.

Un evento è un insieme di più eventi elementari, ossia un sottoinsieme ($\subseteq$) dello spazio campionario Ω .

In molti casi per individuare gli eventi elementari e lo spazio campionario associati a un fato esperimento casuale ci avvarremo di rappresentazioni grafiche chiamate alberi degli eventi.

Lo spazio campionario può contenere un numero finito, una infinità numerabile p una infinità non numerabile di eventi elementari.

Spazi campionari costituiti da un numero finito o da una infinità numerabile di elementi si chiamano discreti mentre quelli costituiti da una infinità non numerabile di elementi si dicono continui.

Va considerato che gli spazi campionari non numerabili sono solo una idealizzazione matematica di situazioni reali.

Uno spazio campionario è sempre riconducibile a un insieme di punti.

Probabilità

Ricordando che un generico evento A è un sottoinsieme dello spazio campionario: $A \subseteq \Omega$.

Ad ogni possibile evento A vogliamo attribuire una misura di probabilità, pertanto la probabilità è una funzione definita su una famiglia di sottoinsiemi di Ω (algebra o σ -algebra) ovvero, ad ogni evento associamo una probabilità.

Due eventi $A, B \subseteq \Omega$ si dicono disgiunti se $A \cap B = \varnothing$.

Vogliamo ora una funzione $P(\cdot)$ che assegni a ogni elemento di una famiglia di sottoinsiemi di Ω un livello di probabilità.

Il dominio della funzione $P(\cdot)$ è costituito da una collezione di insiemi per questo viene chiamata funzione d'insieme.

Per dare una definizione di $P(\cdot)$ dobbiamo partire dalla nozione di frequenza relativa, immaginiamo di aver ripetuto n volte un esperimento: indicando con A un dato esito dell'esperimento la frequenza relativa dell'esito A è un numero compreso nell'intervallo [0,1] dato dal rapporto $f(A) = \frac{n(A)}{n}$ dove $n(A)$ indica la frequenza assoluta di A, cioè quante volte si è verificato A. inoltre se A_1 e A_2 sono due esiti distinti dell'esperimento la frequenza relativa con cui si è verificato o l'uno o l'altro è data dalla somma delle frequenze relative di A_1 e A_2 . Infine, la frequenza relativa con cui si è verificato uno qualsiasi dei possibili esiti dell'esperimento è uguale a 1.

La probabilità è una funzione $P(\cdot)$ su una famiglia di sottoinsiemi di Ω che gode delle seguenti proprietà:

- $P(\Omega) = 1$;
- $P(A) \geq 0$, per ogni evento A ;
- $P(A_1 \cup A_2 \cup \cdots) = P(A_1) + P(A_2) + \cdots$, per ogni successione di eventi a due a due disgiunti.

Le proprietà delle probabilità enunciate vanno sotto il nome di assiomi della probabilità di Kolmogorov. Con la prima proprietà si assegna probabilità 1 all'evento certo che si verifichi uno qualsiasi degli eventi elementari dello spazio campionario. La seconda proprietà afferma che la probabilità non è negativa. La terza stabilisce che si hanno due o più eventi che non presentano risultati in comune, la probabilità che si verifichi uno qualsiasi di essi è pari alla somma delle loro probabilità.

Uno spazio campionario al quale è stata associata una funzione di probabilità dà luogo ad uno spazio di probabilità.

Dalle proprietà sopra enunciate si deducono le seguenti ulteriori proprietà:

- $P(\acute{A})=1-P(A)$, (regola dell'evento complementare) per ogni evento A
- $A \cap \acute{A}=\varnothing$ quindi sono disgiunti, pertanto per il terzo assioma $P(A \cup \acute{A})=P(A)+P(\acute{A})$;
- $A \cup \acute{A}=\Omega$, pertanto, per il primo assioma $P(A \cup \acute{A})=P(\Omega)=1$;
- Si osserva quindi che $P(A)+P(\acute{A})=1$;
- $P(A)\leq 1$, per ogni evento A. Si deduce dalla proprietà precedente e dal secondo assioma $P(\acute{A})\geq 0$;
- $P(\varnothing)=0$, si ricava dalla prima proprietà, osservando che $\acute{\Omega}=\varnothing$ viene chiamato evento impossibile;
- $P(A)\leq P(B)$, per ogni coppia di eventi per cui $A \subseteq B$. Per il terzo assioma $P(B)=P(A)+P(B \cap \acute{A})$. Per il secondo assioma $P(B \cap \acute{A})\geq 0$;
- $P(A_1 \cup A_2)=P(A_1)+P(A_2)-P(A_1 \cap A_2)$, (regola della somma) per ogni coppia A_1 e A_2 di eventi qualsiasi. A_1 e A_2 non sono necessariamente disgiunti. Definiamo $B=A_2\{A¿_1$ che ci permette di definire la probabilità di A_2 come $P(A_2)=P(B)+P(A_1 \cap A_2)$ essendo i due disgiunti, da cui si ricava $P(B)=P(A_2)-P(A_1 \cap A_2)$. A_1 e B sono disgiunti, pertanto $P(A_1 \cup A_2)=P(A_1 \cup B)=P(A_1)+P(B)=P(A_1)+P(A_2)-P(A_1 \cap A_2)$.

Interpretazione della probabilità

La definizione classica ci dice che: la probabilità dell'evento A è il rapporto tra il numero di casi favorevoli al verificarsi di A e il numero totale di eventi elementari, ammesso che questi siano ugualmente probabili.

La definizione frequentista ci dice che: il limite a cui tende la frequenza relativa con cui si presenta l'evento A , aumentando il numero di prove e assumendo che le prove non si influenzino a vicenda e che le condizioni rimangano immutate.

La definizione soggettivista ci dice che: il grado di fiducia che un individuo nutre nel verificarsi dell'evento A .

La definizione assiomatica di Kolmogorov è quella su cui si basano i modelli probabilistici moderni.

Assegnazione delle probabilità agli eventi

Dobbiamo occuparci di come assegnare le probabilità agli eventi dello spazio campionario, ma bisogna distinguere la trattazione gli spazi campionari discreti da un lato e per quelli continui dall'altro.

Esamineremo adesso il problema con riferimento ai soli spazi campionari discreti esaminando dapprima quelli composti da un numero finito N di eventi elementari nel caso di informazione completa (si conosce la probabilità di ogni singolo evento elementare) e poi quelli composti da un'infinità numerabile (la cui procedura è analoga).

Nel caso di Ω finito, pasta assegnare una probabilità a ciascun punto dello spazio campionario, ossia a ciascun evento elementare, dopodiché la probabilità del generico evento A si ottiene sommando la probabilità degli eventi elementari appartenenti ad A. Se p_i è la probabilità del generico evento elementare ω_i , la probabilità dell'evento A è:

$$P(A)=\sum_{i:\omega_i \in A} p_i$$

Dove la somma è estesa a tutti i punti di Ω contenuti in A.

Quando gli eventi elementari sono ugualmente probabili ovvero $p_i=\frac{1}{N}$ per $i=1,2,\ldots,N$ la suddetta formula diventa $P(A)=\frac{n(A)}{N}$, dove $n(A)$ è il numero di eventi elementari contenuti in A e in questo particolare caso la probabilità dell'evento A è la proporzione degli eventi elementari in esso contenuti. Quindi per calcolare $P(A)$ occorre determinare il numero $n(A)$ e per questa operazione è utile il calcolo combinatorio.

Calcolo combinatorio

Dato un insieme di oggetto il calcolo combinatorio fornisce i criteri per configurare i raggruppamenti che possiamo formare con tali oggetti. La costruzione dei gruppi presuppone la scelta di un criterio in base al quale si possa dire se due gruppi sono uguali oppure diversi. Possiamo decidere che i gruppi sono da considerare distinti in quanto alcuni differiscono per l'ordine in cui si presentano oppure possiamo convenire viceversa di considerare raggruppamenti distinti solo quelli che differiscono per almeno un elemento,

Disposizioni

Oggetti distinti

Da n oggetti distinti, si chiamano disposizioni semplici a m a m , o di classe m , i gruppi di m oggetti (con $m \leq n$) che possiamo formare con gli n oggetti

dati (da n oggetti distinti si estraggono m oggetti), seguendo il criterio di considerare distinti due gruppi, sia se differiscono per qualche elemento sia se differiscono per l'ordine con cui gli elementi si presentano.

Per stabilire qual è il numero delle possibili disposizioni, conviene determinare innanzitutto una relazione ricorsiva (relazione tra una entità e se stessa), a questo fine immaginiamo di aver formato con gli n oggetti tutte le disposizioni di classe $m-1$, sia $D_{n,m-1}$ il numero delle disposizioni così formate, da tali disposizioni possiamo formare quelle di classe m aggiungendo a ognuna di queste uno degli $n-m+1$ elementi non presenti in essa.

Denotando con $D_{n,m}$ il numero delle disposizione di classe m abbiamo

$$D_{n,m}=(n-m+1)\,D_{n,m-1}$$

Dalla relazione ricorsiva assegnano ad m i valori $2,3,\ldots,$ abbiamo:

$$D_{n,2}=(n-1)\,D_{n,1}$$

$$D_{n,3}=(n-2)\,D_{n,2}$$

$$\vdots$$

$$D_{n,m}=(n-m+1)\,D_{n,m-1}$$

Poiché $D_{n,1}=n$ abbiamo

$$D_{n,m}=n\cdot(n-1)\cdots(n-m+1)=\prod_{i=1}^{m}(n-i+1)=\frac{n!}{(n-m)!}$$

Il numero delle disposizioni di classe m è dunque dato dal prodotto dei primi m numeri naturali decrescenti a partire da n.

Oggetti ripetitivi

Dati n oggetti ripetitivi in un singolo gruppo, per determinare il numero delle disposizioni con ripetizione dobbiamo indicarlo con $D^{(r)}_{n,m}$ e vale la seguente relazione ricorsiva (relazione tra un'entità e se stessa)

$$D^{(r)}_{n,m}=n\,D^{(r)}_{n,m-1}$$

Tale formula si giustifica pensando che dalle disposizioni con ripetizione di classe $m-1$ si passa a quelle di classe m aggiungendo a ciascuna delle prime, a uno a uno, tutti gli n oggetti e applicando ripetutamente la formula per $m-1,m-2,\ldots,1$ otteniamo

$$D_{n,m}^{(r)}=n^m$$

Permutazioni

Le permutazioni di n oggetti sono un caso particolare delle disposizioni semplici e precisamente sono le disposizioni di n oggetti di classe n (è come se estraessimo n oggetti da un insieme di n oggetti distinguendo gli esiti per l'ordine).

Il numero delle permutazioni di n elementi è dato da

$$P_n=D_{n,n}=n(n-1)\cdot(n-2)\cdots 2\cdot 1=n!$$

Cioè dal prodotto dei primi n naturali. Tale prodotto è indicato convenzionalmente con il simbolo $n!$ Che si legge n fattoriale.

Combinazioni

Le combinazioni di n oggetti di classe m sono tutti i gruppi di m oggetti che possiamo formare con gli n oggetti, considerando distinti due gruppi solo se differiscono per almeno un oggetto.

Il numero delle combinazioni semplici di n oggetti di classe m, che si indica con $C_{n,m}$, si ricava considerando che da ogni combinazione possiamo ottenere $m!$ Disposizioni permutando in tutti i modi possibili le m componenti (ovvero avendo estratto $m\leq n$ oggetti distinti tra loro, il numero delle loro permutazioni, $m!$, indica quante volte l'insieme delle disposizioni è più grande dell'insieme delle combinazioni). Vale allora la relazione

$$D_{n,m}=m!\,C_{n,m}$$

Da cui

$$C_{n,m}=\frac{D_{n,m}}{m!}=\frac{n\cdot(n-1)\cdots(n-m+1)}{m!}=\frac{\prod_{i=1}^{m}(n-i+1)}{m!}=\frac{n!}{(n-m)!\,m!}=\binom{n}{m}$$

Il numero delle combinazioni con ripetizione (con m che può essere maggiore di n) è dato da:

$$C_{n,m}^{(r)}=C_{n+m-1,m}=\binom{n+m-1}{m}$$

Probabilità condizionata

Dati due eventi A e B di Ω, interessa talvolta valutare la probabilità di B supponendo che A si sia già verificato.

La probabilità in questione, che indicheremo con $P(B|A)$, è chiamata probabilità condizionata.

Se A e B sono due eventi dello spazio campionario Ω e $P(A)>0$, allora la probabilità condizionata di B dato A è determinata come

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

Se A è un evento dello spazio campionario Ω e $P(A)>0$, allora valgono le seguenti proprietà:

- $P(\Omega|A)=\frac{P(\Omega\cap A)}{P(A)}=\frac{P(A)}{P(A)}=1$;
- Per ogni evento B, $P(B|A)=\frac{P(A\cap B)}{P(A)}\geq 0$;
- Per ogni successione di eventi $B_1, B_2, \ldots$ a due a due disgiunti
-

$$P(B_1\cup B_2\cup\cdots|A)=\frac{P[(B_1\cup B_2\cup\cdots)\cap A]}{P(A)}=\frac{P[(B_1\cap A)\cup(B_2\cap A)\cdots]}{P(A)}=\frac{P(B_1\cap A)+P(B_2\cap A)+\cdots}{P(A)}=P(B_1|A)$$

Le tre proprietà ricalcano i tre assiomi di Kolmogorov.

Riprendendo la formula:

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

Mettendo in evidenza il numeratore a destra dell'uguale possiamo scrivere

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

Quest'ultima è nota come regola del prodotto. Notiamo che se $P(B)>0$ scambiando i ruoli di due eventi, possiamo anche scrivere

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

Definiamo la partizione dello spazio campionario.

I k eventi $C_1, C_2, \ldots, C_k$ costituiscono una partizione dello spazio campionario Ω se sono a due a due disgiunti e la loro unione dà Ω, cioè se:

- $C_i\cap C_j=\varnothing$ per ogni $i\neq j$;
- $C_1\cup C_2\cup\cdots\cup C_k=\Omega$.

Se gli insiemi $C_1, C_2, \ldots, C_k$ costituiscono una **partizione** dello spazio campionario Ω per ogni evento A possiamo scrivere

$$A = A \cap \Omega = A \cap (C_1 \cup C_2 \cup \ldots \cup C_k) = (A \cap C_1) \cup (A \cap C_2) \cup \ldots \cup (A \cap C_k)$$

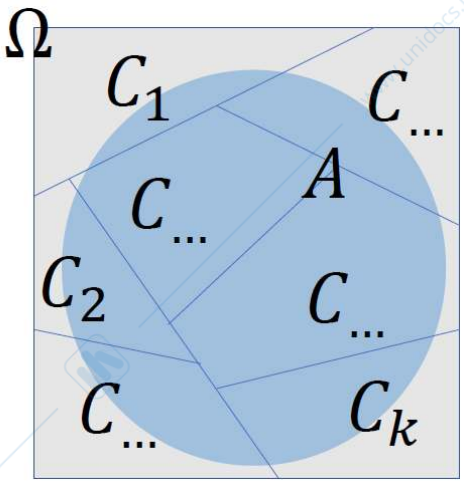
Da cui, considerando che gli eventi sono incompatibili $A \cap C_1, A \cap C_2, \ldots, A \cap C_k$ in quanto $C_1, C_2, \ldots, C_k$ sono incompatibili, per il terzo assioma della probabilità si ha:

$$P(A) = P(A \cap C_1) + P(A \cap C_2) + \ldots + P(A \cap C_k)$$

L'applicazione della regola del prodotto a ognuno dei termini alla destra dell'uguale dà infine:

$$P(A) = \sum_{j=1}^{k} P(A \cap C_j) = \sum_{j=1}^{k} P(A|C_j) P(C_j)$$

Questa formula si chiama formula delle probabilità totali ed esprime la probabilità dell'evento A come una media aritmetica ponderata delle probabilità condizionate $P(A|C_j)$, con pesi parti alle probabilità $P(C_j)$.



Formula di Bayes

Quando un evento A può essere visto come risultato ("effetto") di uno tra k possibili eventi ("cause") $C_1, C_2, \ldots, C_k$, incompatibili e tali che uno di essi si deve verificare interessa valutare la probabilità che, ammesso che A si sia verificato, sia

C_i la causa che l'ha prodotto. Tale probabilità viene indicata con il simbolo $P(C_i \vee A)$.

Quindi sia $C_1, C_2, \ldots, C_k$ una partizione dello spazio campionario, allora si osserva che

$$P(C_i \vee A) = \frac{P(A \cap C_i)}{P(A)} = \frac{P(A|C_i)P(C_i)}{\sum_{j=1}^{k} P(A|C_j)P(C_j)}$$

Dove $P(C_i \vee A)$ è la probabilità condizionata dell'evento A ammesso che sia verificato l'evento C_j .

La formula si ricava immediatamente dalla definizione di probabilità condizionata, applicando al numeratore la regola del prodotto:

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

E al denominatore la formula delle probabilità totali ($P(A) = \sum_{j=1}^{k} P(A|C_j)P(C_j)$). La formula è nota come formula di Bayes.

$P(C_i)$ e $P(C_i \vee A)$ sono rispettivamente le probabilità a priori e la probabilità a posteriori dell'evento C_i che ha causato l'evento A.

Indipendenza tra eventi

Uno dei concetti più importanti della teoria della probabilità è quello di indipendenza.

Due eventi A e B si dicono indipendenti se le probabilità di intersezione $A \cap B$ può essere scritta come il prodotto tra la probabilità di A e la probabilità di B .

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

(L'evento B è indipendente dall'evento A se il verificarsi dell'evento A non muta la probabilità che si verifichi l'evento B : $P(B|A) = P(B) > 0$. È una relazione simmetrica quindi se B è indipendente da A anche A è indipendente da B , è opportuno quindi dire che i due eventi A e B sono indipendenti).

Nei casi non degeneri (nei casi in cui la probabilità non è concentrata su un unico valore) l'indipendenza può essere definita facendo uso delle probabilità condizionate.

Dati due eventi A e B con $P(A)>0$ e $P(B)>0$ le seguenti tre condizioni sono equivalenti (se i due eventi sono indipendenti le seguenti tre condizioni con $P(A)>0$ e $P(B)>0$ sono equivalenti):

- $P(B|A)=P(B)$
- $P(A|B)=P(A)$
- $P(A\cap B)=P(A)P(B)$, poiché è sempre vero che $P(A\cap B)=P(A|B)P(B)$ ma in caso di indipendenza $P(A|B)=P(A)$

Indipendenza e incompatibilità (eventi disgiunti) sono due concetti opposti, poiché in caso di incompatibilità $P(B|A)=\frac{P(A\cap B)}{P(A)}=\frac{P(\varnothing)}{P(A)}=0$ sebbene $P(B)>0$.

CAPITOLO 13 VARIABILI CAUSALI

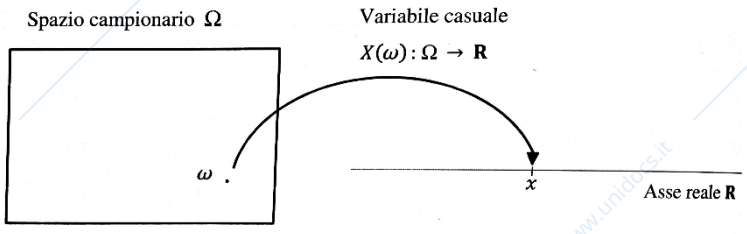
L'espressione **variabile causale** indica una quantità il cui valore dipende dall'esito di un esperimento casuale.

In generale dato lo spazio campionario Ω di un qualsiasi esperimento casuale, possiamo pensare di associare a ogni evento elementare ω di Ω un numero x secondo una regola stabilita.

Una variabile casuale $X(\omega)$, indicata per brevità con X , è una funzione definita dallo spazio campionario Ω che associa un numero reale x ad ogni evento elementare ω di Ω .

$$X(\omega):\Omega\rightarrow R$$

Con l'introduzione della variabile causale X operiamo la trasformazione degli eventi elementari di Ω in punti dell'asse reale.



Non tutte le funzioni reali sono variabili casuali ma solo quelle per le quali è possibile calcolare la probabilità.

La funzione ha supporto Ω non il suo insieme delle parti.

La funzione non è necessariamente iniettiva, ad eventi elementari diversi può corrispondere lo stesso numero reale.

Attraverso i valori assunti dalla variabile casuale possiamo definire degli eventi e calcolarne le rispettive probabilità.

Variabili casuali discrete

Una variabile casuale X si dice **discreta** se può assumere un numero finito o un'infinità numerabile di valori. Per poter operare con una variabile casuale X discreta è opportuno ordinare in senso crescente i suoi valori e assegnare a ciascuno di essi il livello di probabilità che gli compete. A questo fine è del tutto logico assegnare al singolo valore x di X la probabilità dell'unione degli eventi elementari a cui è associato questo valore.

La regola generale con cui si assegnano le probabilità ai valori di una qualsiasi variabile casuale discreta può essere così compendiata. Consideriamo l'evento $X=x$. Indichiamo con A_x l'unione degli eventi elementari ω di Ω per i quali $X(\omega)=x$, ossia l'insieme:

$$A_x=\{\omega : X(\omega)=x\}, \forall\, x \in S_X$$

(L'insieme di tutti possibili valori assunti da X si chiama **supporto**: $S_X \subseteq R$. Per ognuno dei valori possiamo definire un evento in Ω e associarvi quindi una **probabilità)**

Allora la probabilità dell'evento $X=x$ è data da:

$$P(X=x)=P(A_x), \forall\, x \in S_X$$

In altri termini, possiamo dire che la probabilità che $X=x$ è data dalla probabilità complessiva degli eventi elementari (dallo spazio campionario associato all'esperimento casuale) ai quali è associato il valore x assegnato.

Lo schema dei valori assunti da X e le rispettive probabilità prende il nome di **distribuzione di probabilità**. Per una variabile casuale discreta avente k valori dove $x_1, x_2, \ldots, x_k$ sono i possibili valori della variabile casuale X e $p_1, p_2, \ldots, p_k$ le rispettive probabilità.

x	x_1	x_2	$\cdots$	x_k
$P(X=x)$	p_1	p_2	$\cdots$	p_k

Questo schema può essere convenientemente riassunto nella funzione

$$f(x)=P(X=x), x=x_1, x_2, \ldots, x_k$$

Che va sotto il nome di funzione di probabilità.

(La funzione $f(x): R \to [0,1]$ che associa ad ogni valore dell'asse reale la rispettiva probabilità prende il nome di **funzione di probabilità** $f(x)=\begin{cases} P(X=x) \text{ se } x \in S_X \\ 0 \text{ altrimenti} \end{cases}$)

La funzione di probabilità gode delle seguenti **proprietà**

$$f(x) \geq 0 \text{ ; per ogni } x, \sum_{x \in S_x} f(x)=1$$

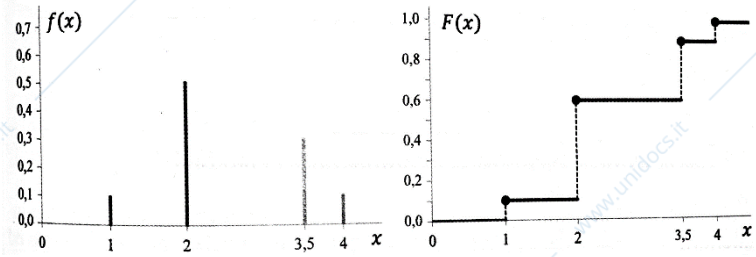
Un altro modo per descrivere una variabile casuale è la funzione di ripartizione, che si ottiene associando a ogni valore x la somma delle probabilità corrispondenti a x e a tutti i valori inferiori:

$$F(x)=P(X \leq x)= \sum_{t \in S_x : t \leq x} f(t)$$

$F(x)$ è definita per qualsiasi intervallo $-\infty, +\infty$) ma essendo una probabilità assume valori nell'intervallo $[0,1]$ ($F(x): R \to [0,1]$ associa ad ogni valore dell'asse reale una probabilità cumulata compresa nell'intervallo $[0,1]$). È una funzione non decrescente, ovvero presi $a<b$ allora $F(a) \leq F(b)$ ed è continua da destra.

La rappresentazione grafica della funzione di probabilità $f(x)$ si effettua mediante un sistema di assi cartesiani ponendo sull'asse delle ascisse i valori di X e su quello delle ordinate le corrispondenti probabilità, usualmente raffigurate come aste.

La rappresentazione grafica della funzione di ripartizione $F(x)$ dà luogo invece a un grafico a gradini.



Media e varianza

Per una distribuzione di frequenze la media aritmetica può essere scritta come somma dei prodotti tra modalità e frequenze relative:

$$\mu = x_1 f_1 + x_2 f_2 + \ldots + x_k f_k = \sum_{i=1}^{k} x_i f_i$$

La media di una variabile casuale discreta X si ottiene come somma dei prodotti tra i calori della variabile causale e i rispettivi livelli di probabilità.

Si chiama media o valore atteso della variabile casuale discreta X e la si denota con μ o con $E(X)$ il numero dato da

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

Dove la somma è estesa a tutti i valori della variabile casuale X .

Per una variabile casuale (come per la distribuzioni di frequenze) la media indica l'ordine di grandezza del fenomeno rappresentato dalla variabile casuale ed è espressa nella stessa unità di misura di questa. L'espressione valore atteso è legata al fatto che la probabilità associata a un dato valore x è interpretabile come la frequenza relativa con cui tale valore x si manifesta se l'esperimento è ripetuto un numero elevato di volte, per cui il calcolo in questione è interpretabile come la media dei valori osservati su un gran numero di prove.

La media gode della proprietà della linearità: se si effettua la trasformazione $Y = a + bX$ dove a e b sono due costanti numeriche, la media di Y , μ_Y , è legata alla media di X , μ_X , dalla medesima relazione: $\mu_Y = a + b\mu_X$.

Un'altra caratteristica di rilievo per la variabile casuale è la varianza.

Per la distribuzione di frequenze la varianza è data dalla somma dei quadrati degli scostamenti delle modalità dalla media aritmetica, moltiplicati per le rispettive frequenze relative. Per una variabile casuale discreta, la struttura è la stessa, con la sostituzione delle frequenze relative con le probabilità.

Sia X una variabile casuale discreta e sia $\mu = E(X)$ la sua media. Si chiama varianza di X e la si denota con σ^2 o $Var(X)$ la quantità:

$$\sigma^2 = Var(X) = \sum_{x \in S_x} (x - \mu)^2 f(x)$$

Si tratta del valore atteso del quadrato della variabile casuale scarto della media, $X-\mu$.

Quest'ultima formula può esse posta nella forma:

$$\sigma^2=\sum_{x\in S_x} x^2 f(x)-\mu^2=E(X^2)-E(X)^2$$

La varianza è sempre non negativa ed è nulla se e solo se X assume un unico valore di probabilità 1; è tanto più grande quanto maggiore è la dispersione dei valori di X attorno alla media e gode della seguente proprietà. Se si effettua la trasformazione $Y=a+bX$ con a e b costanti numeriche, tra la varianza di Y , σ_Y^2 , e la varianza di X , σ_X^2 , sussiste la relazione

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

Variabili casuali continue

Se indichiamo con $X(\omega)=\omega$ una durata la variabile casuale X può assumere a priori un qualsiasi intervallo $[0,T]$ della retta dei numeri reali, essendo T il massimo desumibile.

In prima approssimazione, possiamo dire che una variabile casuale è continua se può assumere tutti i valori di un determinato intervallo (o insiemi di intervalli) di numeri reali.

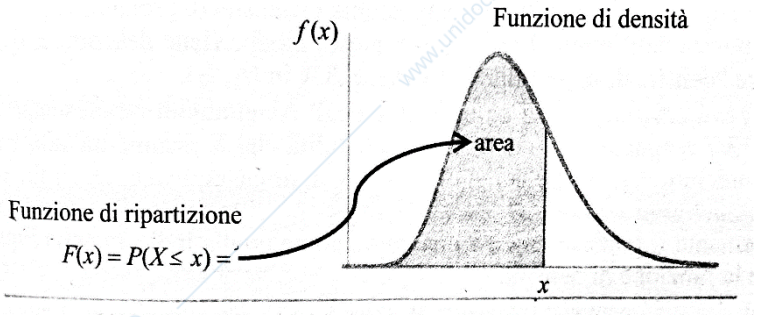
Per definizione formale partiamo dalla funzione di ripartizione che per ogni assegnato valore x dà la probabilità che la variabile casuale assuma un valore minore o uguale a $x: F(X)=P(X\le x)$.

Una variabile casuale X è detta continua se esiste una funzione $f(x)$ tale che la funzione di ripartizione $F(x)=P(X\le x)$ è data dall'area sottesa a $f(x)$ a sinistra di x . La funzione $f(x)$ è denominata **funzione di densità di probabilità** o, più brevemente, funzione di densità

$$F(x)=P(X\le x)=\int_{-\infty}^{x} f(t)\,dt\,,\forall x\in R$$

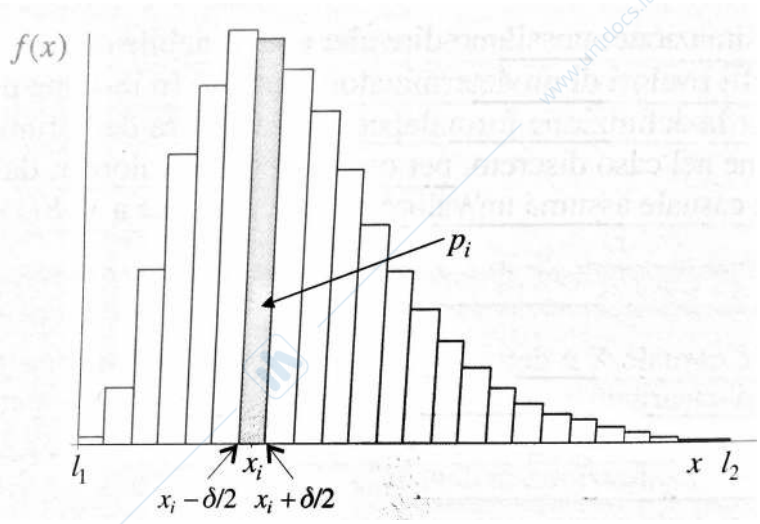
$$f(x)=\frac{d}{dx}F(x)$$

La relazione tra funzione di ripartizione e funzione di densità è



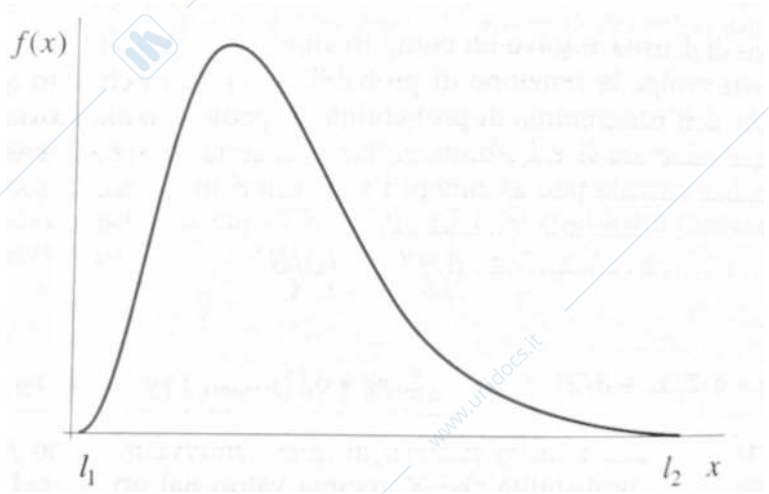
(Il fatto che $F(x)$ sia l'area sottesa di $f(x)$ a sinistra di x e che $f(x)\geq 0$ garantisce che la **funzione di ripartizione sia non decrescente**, il vincolo analitico affinché una generica funzione non negativa possa rappresentare una funzione di densità è che **l'area sottesa ad essa lungo tutto l'asse reale sia pari a** 1 . Ciò fa sì che $0\leq F(x)\leq 1$).

Supponiamo che la variabile casuale X assuma valori tra l_1 e l_2 e immaginiamo di suddividere l'intero intervallo in cui la variabile casuale può assumere i suoi valori in N piccoli sottointervalli di ampiezza $\delta=\frac{l_2-l_1}{N}$ dati da $\left(x_1-\frac{\delta}{2},x_1+\frac{\delta}{2}\right),\left(x_2-\frac{\delta}{2},x_2+\frac{\delta}{2}\right),\ldots,\left(x_N-\frac{\delta}{2},x_N+\frac{\delta}{2}\right)$ essendo $x_1,x_2,\ldots,x_N$ i valori centrali di questi intervalli. Siano $p_1,p_2,\ldots,p_N$, rispettivamente, le probabilità che X assuma valori nel primo, nel secondo, ..., nell' N -esimo intervallo. Rappresentiamo la distribuzione di probabilità così costruita con un istogramma



Dove il generico rettangolo ha area uguale a p_i e altezza uguale a $\frac{p_i}{\delta}$, quantità interpretabile come densità di probabilità per unità di ampiezza della base del rettangolo. In questo modo viene descritta la maggiore o minore densità di probabilità

al variare di X in (l_1, l_2) . L'area complessiva degli N rettangoli è uguale a 1, perché è la probabilità che X assuma un valore qualsiasi nell'intervallo (l_1, l_2) . Se aumentiamo N o diminuiamo δ il profilo del grafico (ossia il sistema di segmenti di retta paralleli che delimita superiormente il grafico) tende a una curva che p la funzione di densità.



Una funzione $f(x)$ può assumere la veste di funzione di densità se è non negativa e se l'area che esse sottende è uguale a 1.

Media e varianza

Si chiama media o **valore atteso** della variabile casuale continua X il numero dato da;

$$\mu = E(X) = \int_{-\infty}^{\infty} x \cdot f(x)\, dx$$

Si chiama **varianza** della variabile casule discreta X la quantità:

$$\sigma^2 = Var(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx$$

La varianza può essere calcolata anche come segue:

$$\sigma^2 = \int_{-\infty}^{\infty} x^2 f(x)\, dx - \mu^2 = E(X^2) - E(X)^2$$

Sia $Y = a + bX$ valore atteso e varianza di Y sono pari a $\mu_Y = a + b\mu_X$ e $\sigma_Y^2 = b^2 \sigma_X^2$

Quantili

Considerata la funzione di probabilità di una variabile casuale continua X descritta dalla funzione di ripartizione $F(x)$, fissato un livello di probabilità p , si chiama

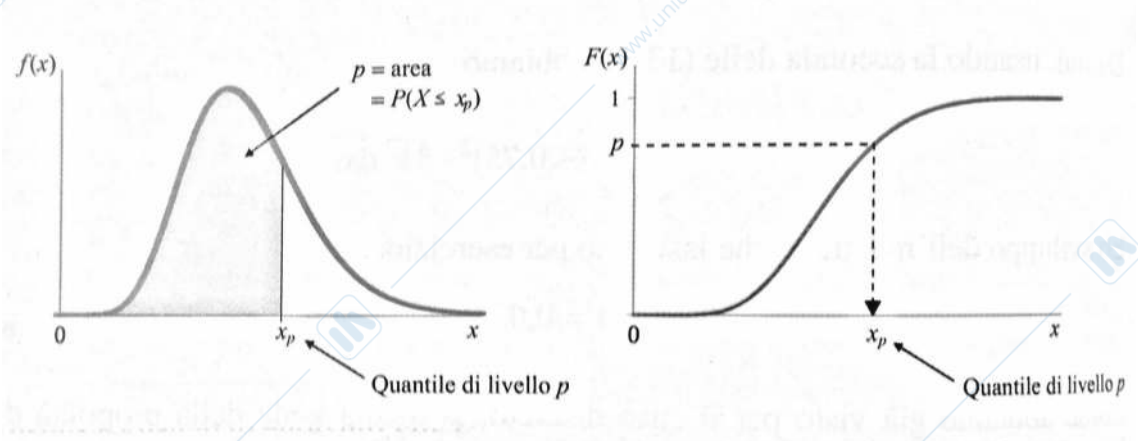
quantile di livello p la quantità x_p in corrispondenza della quale la funzione di ripartizione assume il valore p :

$$F(x_p)=P(X\leq x_p)=p$$

Indicando con $F^{-1}(\cdot)$ inversa della funzione di ripartizione $F(x)$, nel caso in cui questa inversa esista ovvero nel caso in cui non ci sono tratti nei quali $F(x)$ è costante possiamo scrivere

$$x_p=F^{-1}(p)$$

Poiché $F(x_p)$ è interpretabile come area sottesa alla funzione di densità a sinistra di x_p possiamo dire che il quantile di livello p e il punto dell'asse delle ascisse alla cui sinistra l'area sottesa alla funzione di densità è uguale a p .



Ulteriori proprietà delle variabili causali

Data una variabile casuale X , discreta o continua, avente media μ e varianza σ^2 , si chiama variabile casuale standardizzata la quantità

$$Y=\frac{X-\mu}{\sigma}$$

La variabile casuale così definita presenta le seguenti proprietà: ogni variabile causale standardizzata ha media 0 e varianza 1 . Basta considerare che Y è una trasformazione lineare di X del tipo $Y=a+bX$ con $a=\frac{-\mu}{\sigma}$ e $b=\frac{1}{\sigma}$. E come tale per la proprietà della linearità della media si ha $E(Y)=\frac{-\mu}{\sigma}+\frac{1}{\sigma}\mu=0$ mentre la formula della varianza di una trasformazione lineare di una variabile casuale si ha $Var(Y)=\frac{1}{\sigma^2}\sigma^2=1$.

Definiamo momento di ordine r di una distribuzione di probabilità, si tratta di una generalizzazione delle costanti caratteristiche media e varianza.

Sia X una variabile casuale e sia r un intero positivi. Si chiama momento non centrato di ordine r di X la quantità $\mu_r = E(X^r)$ e si chiama momento centrato di ordine r di X la quantità $\acute{\mu}_r = E\left[(X-\mu)^r\right]$ dove $\mu = E(X)$. È immediato stabilire che per una generica variabile casuale X il momento non centrato di ordine 1 è la media $\mu_1 = E(X^1) = E(X)$ e che il momento centrato di ordine 2 è la varianza $\acute{\mu}_2 = E\left[(X-\mu)^2\right] = Var(X)$.

CAPITOLO 14 ALCUNE PARTICOLARI DISTRIBUZIONI DI PROBABILITÀ

Una distribuzione di probabilità espressa da una funzione matematica (di probabilità o di densità) che dipende da uno o più parametri viene chiamata modello probabilistico.

Modelli probabilistici più comuni:

- Distribuzione uniforme discreta
- Distribuzione di Bernoulli
- Distribuzione binomiale
- Distribuzione di Poisson
- Distribuzione uniforme continua
- Distribuzione normale

Distribuzione uniforme discreta

Il supporto è rappresentato dalla coppia di numeri 0 e 1 e viene usata per modellare il realizzarsi o meno di un evento che ha probabilità di verificarsi pari a p .

La distribuzione di probabilità uniforme discreta assegna la stessa probabilità ai primi n numeri naturali.

Una variabile casuale discreta X ha una distribuzione uniforme sui primi n numeri naturali se la sua funzione di probabilità è data da

$$f(x)=\frac{1}{n}, x=1,2,\ldots,n$$

Dove n numero di valori presenti nel supporto (intero positivo).

La media e la varianza di una variabile casuale X avente distribuzione uniforme sui primi n numeri naturali sono date da:

$$E(X)=\frac{n+1}{2}, Var(X)=\frac{n^2-1}{12}$$

Distribuzione di Bernoulli

Il supporto è rappresentato dalla coppia di numeri 0 e 1 e viene usata per modellare il realizzarsi o meno di un evento che ha probabilità di verificarsi pari a p .

Una variabile causale discreta X ha distribuzione di Bernoulli se la sua funzione di probabilità è data da:

$$f(x)=p^x(1-p)^{1-x}, x=0,1$$

dove p è la probabilità che l'evento si verifichi (numero reale $0<p<1$).

La media e la varianza di una variabile casuale X avente distribuzione di Bernoulli sono date da:

$$E(X)=p, Var(X)=p(1-p)$$

Distribuzione binomiale

È un modello probabilistico usato per il conteggio del numero di volte che si realizza un evento in n prove indipendenti in cui la probabilità di verificarsi non cambia mai ed è pari a p .

Una variabile casuale discreta X ha distribuzione binomiale se la sua funzione di probabilità è espressa da ($f(x)=\binom{n}{x}p^x(1-p)^{n-x}, x=0,1,\ldots,n$)

$$f(x)=\frac{n!}{x!(n-x)!}p^x(1-p)^{n-x}, x=0,1,\ldots,n$$

Dove:

- p : probabilità che l'evento si verifichi (numero reale $0<p<1$);

- n : numero di prove indipendenti (intero positivo).

Nota: trattandosi di un conteggio non siamo interessati all'ordine con cui l'evento si verifica nelle diverse prove indipendenti.

La media e la varianza di una variabile casuale X avente distribuzione binomiale sono date da:

$$E(X)=np, Var(X)=np(1-p)$$

Distribuzione uniforme continua

Tra le distribuzioni per variabili casuali continue troviamo quella uniforme.

Il supporto è rappresentato da tutti in numeri reali compresi nell'intervallo $[a,b]$ e la densità di probabilità è costante in tutto l'intervallo.

Una variabile casuale X ha distribuzioni uniforme continua se la sua funzione di densità è espressa da:

$$f(x)=\begin{cases}\frac{1}{b-a}, a\le x\le b\\ 0, altrove\end{cases}$$

Dove a e b sono due costati reali e $a<b$.

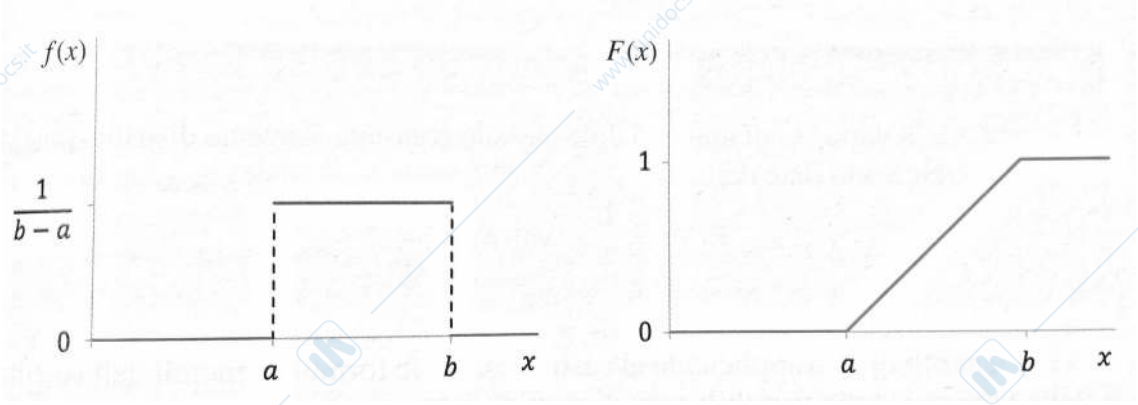
La media e la varianza di una variabile casuale X avente distribuzione uniforme continua sono date da:

$$E(X)=\frac{a+b}{2}, Var(X)=\frac{(b-a)^2}{12}$$

La funzione di ripartizione è pari a zero per $x<a$ è pari a 1 per $x>b$ e per $a\le x\le b$ è data da

$$F(x)=\frac{x-a}{b-a}$$

La formula si ricava considerando che $F(x)$ è l'area sottesa a $f(x)$ nell'intervallo (a,x) ossia l'area del rettangolo di base $x-a$ e altezza $\frac{1}{b-a}$.



Funzione di densità e funzione di ripartizione di una variabile casuale uniforme continua

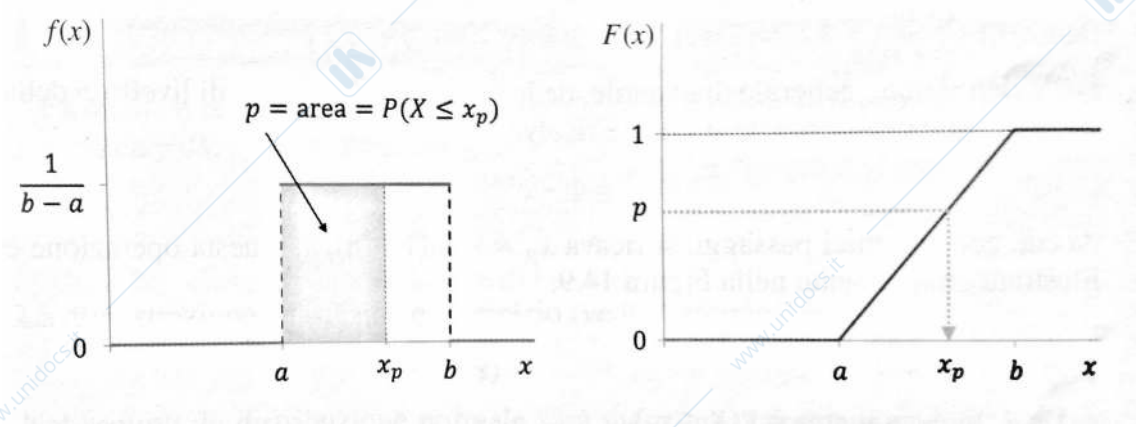
Quantili

Il quantile di livello p si determina risolvendo in x l'equazione

$$\frac{x-a}{b-a}=p$$

da cui

$$x_p=a+p(b-a)$$



Distribuzione normale

Viene spesso usata per modellare delle misure ripetute che mediamente sono pari a μ ma presentano una varianza pari a σ^2 per effetto di errori di misura.

Una variabile casuale continua X ha una distribuzione normale se la sua funzione di densità è data da

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{\frac{-(x-\mu)^2}{2\sigma^2}},-\infty<x<\infty$$

Dove:

- μ : numero reale, valore atteso tale che $-\infty<\mu<\infty$

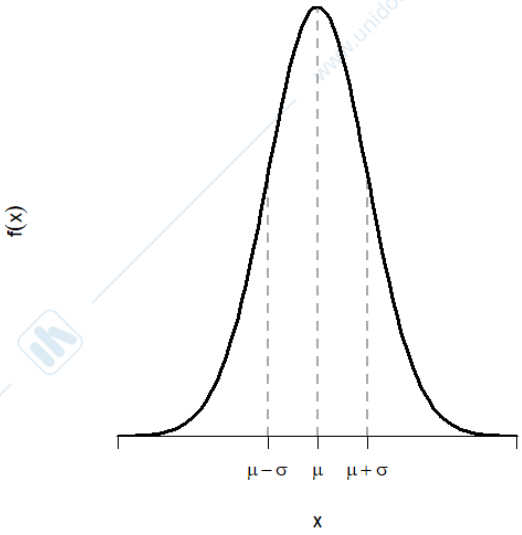
- σ^2 : numero reale positivo, varianza tale che $0<\sigma^2<\infty$

La media e la varianza di una variabile causale X avente distribuzione normale sono date da:

$$E(X)=\mu, Var(X)=\sigma^2$$

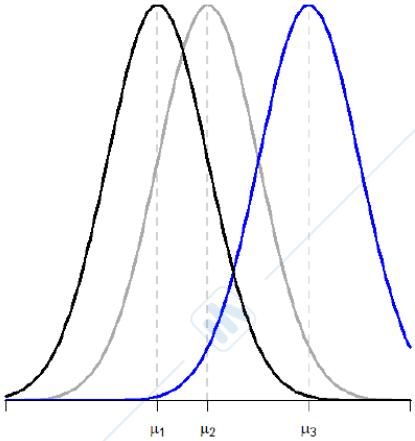
La funzione di densità ha le seguenti proprietà:

- È simmetrica, avendo come asse di simmetria la retta $x=\mu$;
- È crescente nell'intervallo $(-\infty,\mu)$ e decrescente nell'intervallo (μ,∞) ;
- Ha due punti di flesso, in $x=\mu-\sigma$ e $x=\mu+\sigma$;
- È concava (verso il basso) nell'intervallo $(\mu-\sigma,\mu+\sigma)$ e convessa altrove;
- Ha come asintoto l'asse delle x .

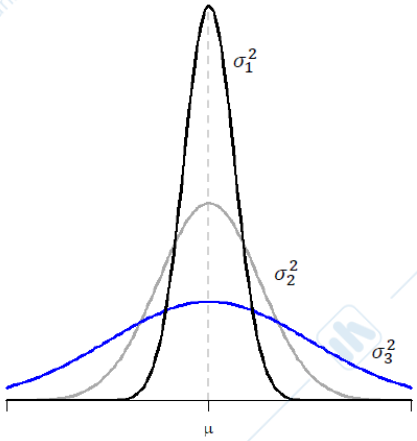


Per brevità per indicare la distribuzione normale verrà usato il simbolo $N(\mu,\sigma^2)$ per indicare una distribuzione normale con media μ e varianza σ^2 . In particolare, con la scrittura X $N(\mu,\sigma^2)$ intenderemo che la variabile casuale X ha distribuzione di probabilità normale con media μ e varianza σ^2 .

Per mostrare l'effetto sulla posizione e sulla forma della curva normale delle due costanti che appaiono nella funzione della densità, nelle immagini presentiamo sei curve normali: a destra tre curve normali aventi la stessa media e varianze diverse mentre a sinistra tre curve normali con medie diverse e varianze uguali.

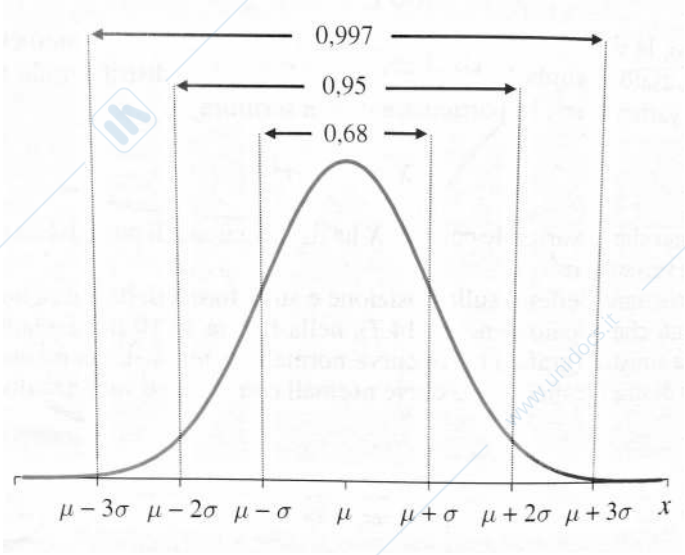


$\mu_1 < \mu_2 < \mu_3$



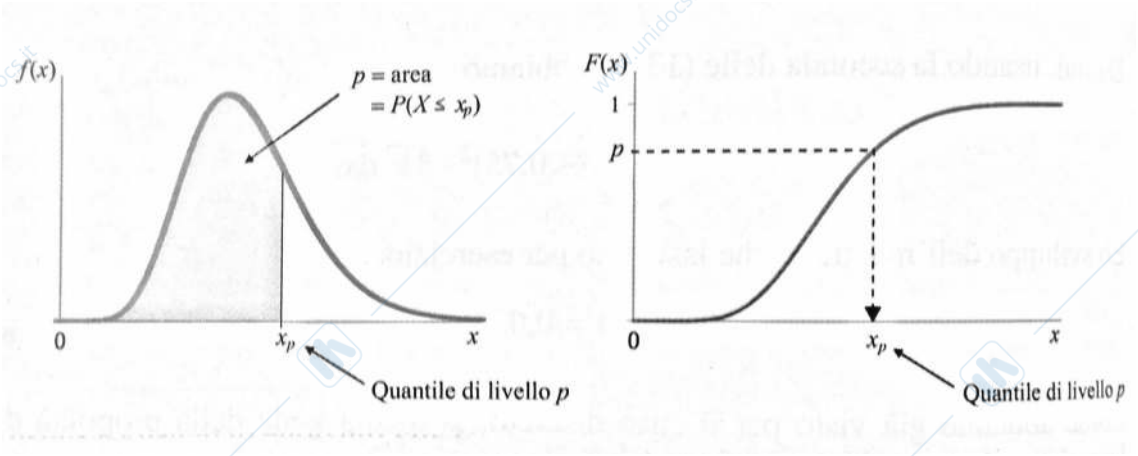
$\sigma_1^2 < \sigma_2^2 < \sigma_3^2$

Per farsi un'idea della struttura della distribuzione normale è bene considerare i seguenti tre intervalli centrali rispetto alla media: $(\mu-\sigma,\mu+\sigma)$, $(\mu-2\sigma,\mu+2\sigma)$, $(\mu-3\sigma,\mu+3\sigma)$, le probabilità associate ai tre livelli sono nell'ordine 0,68, 0,95 e 0,997.



Funzione di ripartizione

Geometricamente la funzione di ripartizione $F(x)$ di una variabile causale normale $N(\mu,\sigma^2)$ è l'area sottesa alla funzione di densità normale da $-\infty$ a x .



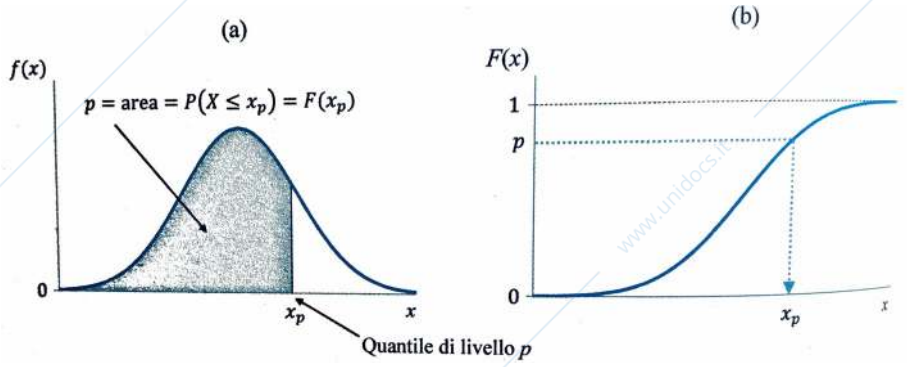
Quantili

Il quantile di livello p di una variabile causale normale $N(\mu,\sigma^2)$ è l'inversa della funzione di ripartizione $F(x)$:

$$x_p = F^{-1}(p)$$

$F(x)$ e x_p non sono esprimibili in forma esplicita.

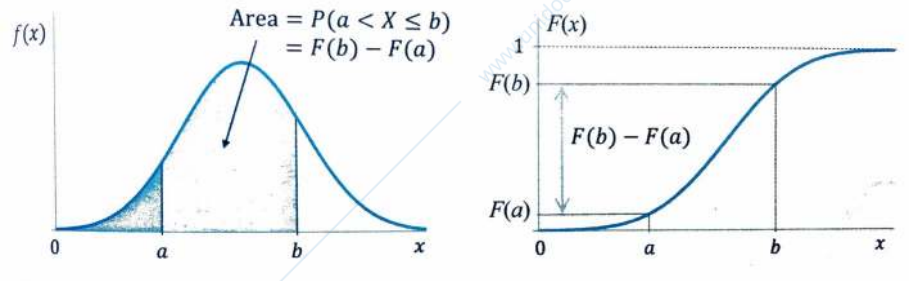
Illustriamo l'operazione di calcolo del quantile di livello p di una variabile causale normale $N(\mu,\sigma^2)$ si tratta di individuare sull'asse delle ascisse il punto a sinistra del quale l'area sottesa alla densità normale è pari a p .



Tramite la funzione di ripartizione è possibile definire la probabilità che X appartenga a un qualsiasi intervallo (a,b) :

$$P(a<X\le b)=P(X\le b)-P(X\le a)=F(b)-F(a)$$

La probabilità cercata è la differenza tra l'area sottesa alla curva normale fino a b (area azzurro scuro+ area azzurro chiaro) e l'area sottesa alla curva fino ad a (area azzurro scuro).



Distribuzione normale standard

La distribuzione normale standard è la distribuzione con media 0 e varianza 1. La sua funzione di densità è

$$f(z)=\frac{1}{\sqrt{2\pi}}e^{\frac{-z^2}{2}}$$

La distribuzione normale standard verrà indicata sinteticamente $N(0,1)$.

Sia X una variabile casuale normale $N(\mu,\sigma^2)$. Sia Z la variabile casuale standardizzata data da

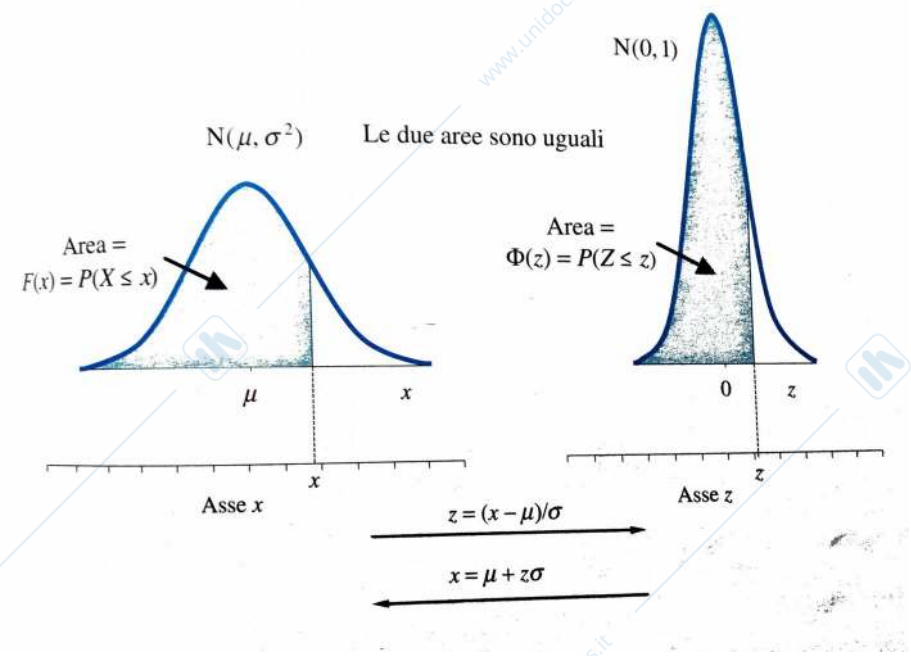
$$Z=\frac{X-\mu}{\sigma}$$

Allora la variabile casuale Z ha distribuzione normale standard con media 0 e varianza 1. Inoltre, la funzione di ripartizione della variabile casuale X può esser espressa come

$$F(x)=P(X\le x)=P\left(Z\le\frac{X-\mu}{\sigma}\right)=\Phi\left(\frac{X-\mu}{\sigma}\right)$$

Dove $\Phi(\cdot)$ è la funzione di ripartizione della distribuzione normale standard.

Quest'ultima formula ci dice che l'area sottesa alla curva normale $N(\mu,\sigma^2)$ a sinistra di un punto x assegnato è uguale all'area sottesa alla curva normale standard $N(0,1)$ a sinistra del punto $\frac{X-\mu}{\sigma}$, illustrato di seguito.



Sulla base della suddetta formula possiamo scrivere

$$P(a<x\leq b)=F(b)-F(a)=\Phi\left(\frac{b-\mu}{\sigma}\right)-\Phi\left(\frac{a-\mu}{\sigma}\right)$$

Dunque, la probabilità che una variabile casuale X con distribuzione normale $N(\mu,\sigma^2)$ assuma un valore nell'intervallo (a,b) è uguale alle probabilità che una variabile casuale standard $N(0,1)$ assuma un valore nell'intervallo di estremi $\frac{a-\mu}{\sigma}$ e $\frac{b-\mu}{\sigma}$.

Probabilità e quantili tramite la tavola della distribuzione normale standard

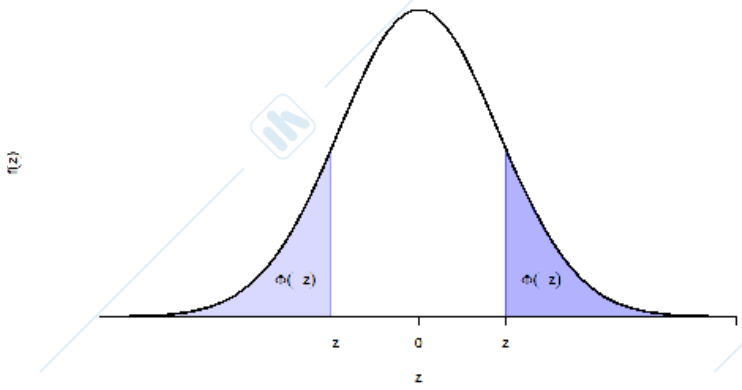
I valori $\Phi(z)$ sono confinabili nella seguente tabella a doppia entrata.

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5	0,504	0,508	0,512	0,516	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	0,591	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	0,648	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,67	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,695	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,719	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	0,758	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,791	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	0,834	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	0,877	0,879	0,881	0,883
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,898	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	0,937	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
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,975	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,983	0,9834	0,9838	0,9842	0,9846	0,985	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,989
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	0,992	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,994	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	0,996	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,997	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,998	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,999	0,999
3,1	0,999	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,2	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995
3,3	0,9995	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997
3,4	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998

Essa dà modo di calcolare con buona approssimazione le probabilità associate alla distribuzione normale $N(\mu,\sigma^2)$.

La tavola contiene i livelli della funzione di ripartizione della distribuzione normale standard per i valori di z non negativi che vanno da 0 a 3,49 con un passo di 0,01.

Ciò per due ragioni: a) per $z \geq 3{,}50$, $\Phi(z) \approx 1$; b) i valori di $\Phi(z)$ corrispondenti a valori negativi di z si possono ottenere sfruttano la relazione $\Phi(-z) = 1 - \Phi(z)$, essendo z un numero positivo.



Probabilità

Per calcolare la probabilità che una variabile casuale normale assuma valori in un dato intervallo se la variabile causale è normale standard la probabilità $P(z_1<Z\leq z_2)$ si ottiene immediatamente come

$$P(z_1<Z\leq z_2)=\Phi_{¿}(¿2)-\Phi_{¿}(¿1)$$

Dove z, z, $\Phi_{¿}(¿2)$, $\Phi_{¿}(¿1)$ si deducono dalla tavola.

Quantili

Per calcolare i quantili di una variabile casuale standard dobbiamo fissare un livello di probabilità p ed individuare il valore z che soddisfa l'equazione

$$P(X\leq x)=\Phi(z)=p$$

La soluzione in z dell'equazione cioè il quantile di livello p sarà individuato con il simbolo z_p.

Il calcolo dei quantili di una variabile casuale normale $N(\mu,\sigma^2)$ con media e varianza qualsiasi si svolge in due fasi:

1. si determina tramite la tabella il quantile z_p ;
2. si calcola il quantile x_p della variabile casuale normale mediante l'equazione

   $x_p=\sigma\cdot z_p+\mu$ (questo si evince dal fatto che se $F(x)=P(X\leq x)=\Phi\left(\frac{x-\mu}{\sigma}\right)=\Phi(z)=p$

   allora $\frac{x_p-\mu}{\sigma}=z_p$).

CAPITOLO 15 VARIABILI CASUALI DOPPIE E MULTIPLE

Combinazioni lineari di variabili casuali

Siano $X_1, X_2, \ldots, X_n$ n variabili casuali, discrete o continue. Considerando la combinazione lineare delle n variabili casuali data da $W = a_1 X_1 + a_2 X_1 + \ldots + a_n X_n$ ($W = \sum_{i=1}^{n} a_i X_i$ ¿ , dove $a_1, a_2, \ldots, a_n$ sono costanti numeriche. Allora la media e la varianza di W sono espresse da

$$\mu_W = a_1 \mu_{X_1} + a_2 \mu_{X_2} + \ldots + a_n \mu_{X_n} = \sum_{i=1}^{n} a_i \mu_{X_i} \qquad \sigma_W^2 = a_1^2 \sigma_{X_1}^2 + a_2^2 \sigma_{X_2}^2 + \ldots + a_n^2 \sigma_{X_n}^2 = ¿$$

$$¿ \sum_{i=1}^{n} a_i^2 \sigma_{X_i}^2 + \sum_{i=1}^{n} \sum_{j \neq i}^{n} a_i a_j \sigma_{X_i X_j} = \sum_{i=1}^{n} a_i^2 \sigma_{X_i}^2 + 2 \sum_{i=1}^{n} \sum_{j>i}^{n} a_i a_j \sigma_{X_i X_j} = \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j \sigma_{X_i X_j}$$

Dove μ_{X_i} e σ_{Xi}^2 sono la media e ka varianza della singola variabile casuale X_i , mentre $\sigma_{X_i X_i}$ è la covarianza tra X_i e X_j .

CAPITOLO 16 LEGGE DEI GRANDI NUMERI E TEOREMA DEL LIMITE CENTRALE

CAPITOLO 17 POPOLAZIONE, CAMPIONE, DISTRIBUZIONI CAMPIONARIE

CAPITOLO 18 STIMA PUNTUALE

Indicando con θ il parametro di interesse vediamo quali sono i criteri guida che aiutano nella scelta della statistica più appropriata per la stima del parametro.

Dato un campione casuale $(X_1, X_2, \ldots, X_n)$ proveniente da una popolazione di cui interessa stimare il parametro θ, si chiama **stimatore** la statistica campionaria $T = t(X_1, X_2, \ldots, X_n)$ utilizzata per stimare θ.

Si chiama invece **stima** la singola determinazione dello stimatore, cioè il valore $t(x_1, x_2, \ldots, x_n)$ che esso assume nel campione osservato $(x_1, x_2, \ldots, x_n)$.

I parametri che utilizzeremo di più sono la media μ di una generica popolazione e la varianza σ^2.

Per effettuare la stima puntuale dobbiamo scegliere lo stimatore più conveniente per attribuire un valore al parametro di interesse θ.

Proprietà degli stimatori

Consideriamo un campione casuale $(X_1, X_2, \ldots, X_n)$ proveniente da una popolazione di cui interessa stimare il parametro θ. Sia $T = t(X_1, X_2, \ldots, X_n)$ uno degli stimatori eleggibili per la stima di θ. Consideriamo la variabile casuale data dalla differenza tra lo stimatore e il valore del parametro incognito: $T - \theta$, che chiameremo errore di stima. Si tratta di una variabile casuale che, nel singolo campione dello spazio campionario, assumerà un valore positivo, negativo o nulla, a seconda che la stima sia al di sopra, al di sotto o sia pari al valore del parametro.

Prendiamo il valore atteso $E(T - \theta)$, questa quantità può essere negativa, positiva o nulla: se è negativa prevalgono mediamente gli errori negativi, nel senso che lo stimatore tende a sottostimare il parametro, se è positiva si verifica il fenomeno opposto, se è nulla c'è il bilanciamento tra errori negativi ed errori positivi. È evidente che è preferibile che si realizzi questa terza condizione. Pertanto, dichiamo che una proprietà desiderabile per uno stimatore è che $E(T - \theta) = 0$ per ogni θ o equivalente che E (T)=0 per ogni θ.

Sia dato un campione casuale $(X_1, X_2, \ldots, X_n)$ proveniente da una popolazione di cui interessa stimare il parametro θ. Uno stimatore $T = t(X_1, X_2, \ldots, X_n)$ del parametro θ è detto non distorto se E (T)=0 per ogni θ. La differenza $D(T) = E(T) - \theta$ è denominata distorsione.

ESEMPIO: considerando un campione casuale $(X_1, X_2, \ldots, X_n)$ proveniente da una popolazione qualsiasi con varianza σ^2 si verifichi che $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \acute{X})^2$ è uno

stimatore distorto della varianza σ^2. Poiché $\hat{\sigma}^2=\frac{n-1}{n}S^2$ possiamo scrivere $E(\hat{\sigma}^2)=E\left(\frac{n-1}{n}S^2\right)=\frac{n-1}{n}E(S^2)=\frac{n-1}{n}\sigma^2$ da cui ricaviamo la distorsione dello stimatore $D(\hat{\sigma}^2)=E(\hat{\sigma}^2)-\sigma^2=\frac{n-1}{n}\sigma^2-\sigma^2=\frac{-\sigma^2}{n}$. Possiamo perciò concludere che lo stimatore presenta una distorsione negativa, nel senso che produce stime mediamente più piccole di σ^2. Al contrario, il valore atteso della media campionaria coincide con la media della popolazione generatrice $E(\dot{X})=\mu$.

La seconda proprietà che uno stimatore dovrebbe possedere è che la quantità $E(T-\theta)^2$ sia molto piccola, cioè sia molto piccolo l'errore medio di stima associato allo stimatore.

Sia dato un campione casuale $(X_1,X_2,\ldots,X_n)$ proveniente da una popolazione di cui interessa stimare il parametro θ. Si chiama errore quadratico medio dello stimatore $T=t(X_1,X_2,\ldots,X_n)$ del parametro θ la quantità $MSE(T)=E(T-\theta)^2$.

Penando allo spazio campionario cioè all'insieme di tutti i campioni casuali di ampiezza n estraibili dalla popolazione di interesse, l'errore quadratico medio può essere interpretato come la distanza media di T da θ, distanza media misurata con la media aritmetica dei quadrati degli errori di stima $t(X_1,X_2,\ldots,X_n)-\theta$, associati ai campioni dello spazio campionario. L'errore quadratico medio è quindi la misura dell'imprecisione dello stimatore. Pertanto, lo stimatore ideale è quello che tra tutti i possibili stimatore, a parità di altre caratteristiche ha l'errore quadratico medio più piccolo.

L'errore quadratico medio dello stimatore T di θ può essere scritto nella forma

$$MSE(T)=Var(T)+\begin{¿}D(T)\end{¿}$$

dove $Var(T)=E[T-E(T)]^2$ è la varianza dello stimatore T.

Si osservi che gli stimatori non distorti hanno errore quadratico medio uguale alla varianza $MSE(T)=Var(T)$ solo se $D(T)=E(T)=\theta$ per ogni θ.

In base alla formula $MSE(T)=Var(T)+\text{¿}\ D\text{¿}\ (T)$ il grado di precisione di uno stimatore è funzione di due componenti la varianza e la distorsione: può capitare che uno stimatore distorto abbia varianza più piccola di quella di uno stimatore non distorto e abbia comunque un errore quadratico medio superiore.

Considerando una coppia di stimatori T_1 e T_2 dello stesso parametro diamo la definizione di stimatore efficiente.

Dati due stimatori dello stesso parametro θ, $T_1=t_1(X_1,X_2,\ldots,X_n)$ e $T_2=t_2(X_1,X_2,\ldots,X_n)$, si dice che il primo è **più efficiente** del secondo se

$MSE(T_1)\leq MSE(T_2)\text{per ogni}\,\theta$ in cui si assume che per almeno un valore di θ vale la diseguaglianza in senso stretto.

Poiché l'errore quadratico medio $MSE(T_1)$ può dipendere dal valore incognito di θ, lo stimatore T_1 è più efficiente di T_2 solo se il suo errore quadratico medio:

- non supera quello di T_2 per ogni valore di θ;
- è inferiore a quello di T_2 per almeno un valore di θ.

Proprietà asintotiche

Gli errori di stima tendono mediamente a diminuire al crescere della dimensione del campione, in quanto il campione più grande riesce a riprodurre più fedelmente le caratteristiche numeriche della popolazione. Pertanto, studieremo il comportamento degli stimatori quando si fa tendere n a infinito.

Preso uno stimatore T del parametro θ e fissato un numero ε positivo piccolo a piacere, consideriamo la probabilità che T differisca da θ, in valore assoluto, per meno di ε:

$$P(|T-\theta|<\varepsilon)=P(\theta-\varepsilon<T<\theta+\varepsilon)\ .$$

La domanda che ci si pone è come varia questa probabilità al crescere dell'ampiezza del campione, definiamo quindi la proprietà della consistenza e della coerenza degli stimatori.

Uno stimatore $T_n=t_n(X_1,X_2,\ldots,X_n)$ del parametro θ si dice **consistente** se per ogni ϵ positivo, piccolo a piacere, vale la relazione

$\lim_{n\to\infty} P\left(\left|T_n-\theta\right|<\varepsilon\right)=1$ per qualunque valore di θ .

Nella definizione data il deponente n che contrassegna T indica che si considera la successione di variabili casuali formata dalla stessa statistica campionaria applicata a campioni di ampiezza crescente.

Una condizione sufficiente perché uno stimatore sia consistente è che il suo errore quadratico medio tenda a 0 al tendere di n a infinito.

La *media campionaria* $\acute{X}$ (stimatore consistente per la media della popolazione), la *proporzione campionaria* $\hat{p}$ e la *varianza campionaria* S^2 (stimatore consistente per la varianza della popolazione generatrice) sono tutti e tre stimatori consistenti dei parametri rispettivamente μ , p e σ^2 .

$$\lim_{n\to\infty} P\left(\left|\acute{X}_n-\mu\right|<\varepsilon\right)=1$$

$$\lim_{n\to\infty} P\left(\left|S_n^2-\sigma^2\right|<\varepsilon\right)=1$$

$$\lim_{n\to\infty} P\left(\left|\hat{P}_n-p\right|<\varepsilon\right)=1$$

Una seconda proprietà asintotica degli stimatori è la non distorsione asintotica.

Uno stimatore $T_n=t(X_1,X_2,\ldots,X_n)$ del parametro θ si dice asintoticamente non distorto se la sua distorsione tende a 0 al tendere di n a infinito.

ESEMPIO: si verifichi che lo stimatore $\hat{\sigma}_n^2=\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\acute{X}\right)^2$ è asintoticamente non distorto. La distorsione dello stimatore è pari a $\frac{-\sigma^2}{n}$, pertanto $\lim_{n\to\infty} D\left(\hat{\sigma}_n^2\right)=\lim_{n\to\infty}\left(\frac{-\sigma^2}{n}\right)=0$. Lo stimatore è dunque asintoticamente non distorsivo.

Criteri per la scelta dello stimatore

Abbiamo visto che $\acute{X}$, $\hat{P}$ e S^2 sono stimatori non distorti e consistenti rispettivamente per i parametri μ , p e σ^2 .

È noto che sono anche gli stimatori più efficienti tra tutti gli stimatori non distorti.

Un criterio per la ricerca di uno stimatore ottimo è il **metodo della massima verosimiglianza**.

Il caso delle popolazioni descritte da modelli

Consideriamo il caso di una popolazione descritta da un modello matematico espresso dalla funzione di probabilità o di densità

Popolazione Bernoulliana

$f_X(x)=p^x(1-p)^{1-x}, x=0,1$ Probabilità di osservare un determinato campione

$$f(x_1,x_2,\dots,x_n)=\prod_{i=1}^{n} f_X(x_i)=p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i}$$

Verosimiglianza (Likelihood)

$$L(p)=p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i} \qquad \max_p L(p)$$

Possiamo passare dalla log-verosimiglianza (trasformata logaritmica, monotona, quindi il punto di massimo coinciderà)

$$l(p)=\log L(p)=\left(\sum_{i=1}^{n} x_i\right)\log(p)+\left(n-\sum_{i=1}^{n} x_i\right)\log(1-p)$$

Deriviamo in p per cercare il punto stazionario $\frac{\partial}{\partial p} l(p)=0$

Si ricordi che

$$\frac{\partial}{\partial x}\log(f(x))=\frac{f'(x)}{f(x)}$$

$$\frac{\partial}{\partial p} l(p)=\left(\sum_{i=1}^{n} x_i\right)\frac{1}{p}+\left(n-\sum_{i=1}^{n} x_i\right)\frac{-1}{1-p}$$

$$\left(\sum_{i=1}^{n} x_i\right)\frac{1}{p}+\left(n-\sum_{i=1}^{n} x_i\right)\frac{-1}{1-p}=0$$

$$\left(\sum_{i=1}^{n} x_i\right)\frac{1}{p}=\left(n-\sum_{i=1}^{n} x_i\right)\frac{1}{1-p}$$

$$(1-p)\left(\sum_{i=1}^{n} x_i\right)=\left(n-\sum_{i=1}^{n} x_i\right)p$$

$$\sum_{i=1}^{n} x_i - p\sum_{i=1}^{n} x_i = np - p\sum_{i=1}^{n} x_i$$

$$\hat{p} = \frac{\sum_{i=1}^{n} x_i}{n}$$