Università di Udine, Facoltà di Scienze della Formazione
Corso di Laurea in Scienze e Tecnologie Multimediali

Corso di Matematica e Statistica

(Giorgio T. Bagni)

Appunti: elementi di Statistica

1. INTENSITÀ, FREQUENZA ASSOLUTA E RELATIVA

1.1. L'analisi statistica di un fenomeno

Immaginiamo, in una ricerca antropologica, di dover rispondere alla domanda seguente: *qual è, mediamente, l'altezza delle donne europee?*

Per rispondere bisogna ad esempio precisare che significa "mediamente": l'altezza alla quale la domanda è riferita non è quella di un ben determinato individuo, ma dipende da una vasta moltitudine di casi. A tale proposito, appare poco pratico informarsi, singolarmente, sull'altezza di *tutte* le donne europee: si tratterebbe di esaminare decine di milioni di persone!

Un metodo di indagine conveniente si basa sulla raccolta di *informazioni* che spesso indicheremo con il termine *dati*: quante donne vivono in Europa, come sono distribuite nei paesi europei... Dopo aver raccolto le informazioni necessarie, saremo chiamati ad una loro *elaborazione*: cercheremo, forse, di stimare l'altezza media delle donne francesi, spagnole... Da questa elaborazione, dovremo infine giungere ad un risultato: la risposta alla domanda formulata.

Problemi come quello ora proposto e i metodi per affrontare e risolvere tali problemi fanno parte della *statistica*. Chiameremo *statistica descrittiva* il settore della statistica che si occupa della raccolta e dell'interpretazione dei dati.

1.2. Intensità e frequenza assoluta

I dati statistici sono espressi da numeri (reali, interi, ...). Inizieremo la presentazione della statistica descrittiva proponendo l'interpretazione di tali dati numerici, ovvero definendo ed illustrando alcuni generi fondamentali di dati. Riferiamoci ancora alla situazione descritta nel paragrafo precedente: un primo rilevamento di dati potrebbe riguardare l'altezza di alcuni soggetti (che dovranno essere scelti seguendo criteri precisi: di questo ci occuperemo nel seguito). Avremo così a disposizione, come dato iniziale, un insieme di misure, numeri (reali o razionali) ciascuno dei quali esprime il rapporto dell'altezza del soggetto prescelto rispetto ad una grandezza campione, l'unità di misura (il metro).

Dati come questi sono indicati dal termine *intensità.*

Definizione 1. Si dice *intensità* $I \in \mathbf{R}$ di ogni accadimento osservato la misura di esso rispetto ad una fissata unità.

Ma il descritto modo di procedere non è l'unico possibile per iniziare una ricerca statistica: ad esempio, potremmo individuare un campione di soggetti e contare quanti hanno altezza maggiore di m 1,70. Il dato iniziale, in questo caso, non è costituito da una misura (o da un insieme di misure), bensì da un numero naturale, che esprime *quante volte* si è verificata una situazione. Per classificare dati come quest'ultimo, introduciamo il concetto di *frequenza assoluta*.

Definizione 2. Si dice *frequenza assoluta* $f \in \mathbf{N}$ di un accadimento il numero di volte che esso si verifica.

I dati necessari per un'indagine statistica possono essere di entrambi i tipi ora presentati; anzi, una stessa indagine potrà essere condotta attraverso un rilevamento di intensità o di frequenza.

Esempio 1. Si voglia analizzare la richiesta del mercato di automobili nella nazione X nello scorso anno. Suddivisa l'area da esaminare in quattro regioni, possiamo occuparci della spesa complessiva che, in ciascuna regione, è stata dedicata all'acquisto di auto, nell'anno. Otteniamo la tabella seguente:

Regioni esaminate	*Spesa per l'acquisto di auto nell'anno (in milioni di euro)*
Regione X-A	35
Regione X-B	87
Regione X-C	73
Regione X-D	173

Ciascuno dei dati riportati è un'intensità: esprime infatti la misura di una grandezza (la spesa complessiva annua per l'acquisto di auto) rispetto ad una determinata unità di misura (il milione di euro).

Esempio 2. Si voglia analizzare la richiesta del mercato di automobili nella nazione X durante lo scorso anno. Suddivisa l'area da esaminare in quattro regioni, come nell'esempio precedente, possiamo occuparci del numero di auto vendute, in ciascuna regione, nell'anno. Otteniamo la tabella seguente:

Regioni esaminate	*Numero auto vendute nell'anno*
Regione X-A	2309
Regione X-B	5477
Regione X-C	4201
Regione X-D	9732

Ciascuno dei dati sopra riportati è una frequenza: esprime infatti quante volte si è verificato un ben preciso accadimento (l'acquisto di un'auto).

La frequenza assoluta è un dato statistico importante. Talvolta, però, la sua interpretazione può non essere immediata: ad esempio, se la rottura di un apparecchio, diciamo di tipo A, si presenta n volte, mentre la rottura di un apparecchio di tipo B si presenta m volte, con $m>n$, siamo autorizzati a dire che l'apparecchio di tipo A è più affidabile di quello di tipo B?

Prima di rispondere, il lettore esamini l'esempio seguente.

Esempio 3. Desideriamo studiare l'affidabilità di tre apparecchi, che indichiamo con i tipi A, B, C.

Tipo di apparecchio	*Apparecchi esaminati*	*Guasti rilevati*
Apparecchio tipo A	54	7
Apparecchio tipo B	96	9
Apparecchio tipo C	32	6

Se ci limitassimo a considerare le frequenze assolute dei guasti, sopra riportate, saremmo portati a concludere che l'apparecchio più affidabile è quello di tipo C (che fa riscontrare il minimo numero di guasti), seguito da quello di tipo A ed infine da quello di tipo C. Ma il lettore si sarà reso conto che l'esame delle frequenze assolute è, in questo caso, insufficiente, addirittura fuorviante: i dati raccolti, infatti, *non sono omogenei*, ovvero non sono riferiti a campioni di apparecchi egualmente numerosi!

Dobbiamo allora tener conto che l'apparecchio di tipo C presenta un numero minimo di guasti (6), ma riferiti a soli 32 casi esaminati. L'apparecchio di tipo A, pur facendo riscontrare un maggior numero di guasti (7), è stato esaminato in 54 casi. E l'apparecchio di tipo B (con 9 guasti), addirittura per 96 casi, il triplo dei casi riferiti all'apparecchio di tipo C...

Prima di pronunciarsi sull'affidabilità dei tre apparecchi, quindi, è opportuno elaborare i dati raccolti più a fondo.

1.3. Frequenza relativa

Definizione 3. Si dice *frequenza relativa* $p \in \mathbf{Q}$ di un determinato accadimento il rapporto tra la frequenza assoluta di tale accadimento ed il totale dei casi osservati.

Essendo il numero di accadimenti (ovvero la frequenza assoluta) non maggiore del numero totale di casi osservati, la frequenza relativa p è sempre tale che:

$$0 \le p \le 1$$

Esempio 4. Ad un convegno di studi partecipano 89 ricercatori di cinque nazioni; ecco la loro distribuzione, con le frequenze assolute e relative:

Nazioni	*Numero ricercatori (frequenze assolute)*	*Frequenze relative*
Belgio	12	0,135
Francia	22	0,247
Italia	14	0,157
Germania	31	0,349
Svizzera	10	0,112

Ogni frequenza relativa è stata ottenuta dividendo la corrispondente frequenza assoluta per 89 (numero totale dei partecipanti); ad esempio:

(Belgio) 12 (frequenza assoluta) : 89 (totale partecipanti) =
= 0,13483... (frequenza relativa) =
= 0,135 (valore approssimato, con tre decimali)

Osservazione. Esaminando il precedente esempio, sottolineiamo un'importante caratteristica: la somma di tutte le frequenze relative è 1.

Esempio 5. Con riferimento all'esempio precedente, è possibile aggiungere la riga riferita al totale dei partecipanti, somma delle righe riferite ai partecipanti suddivisi per nazione:

Totale	89	1,000

Esempio 6. Con riferimento all'esempio 3, analizziamo l'affidabilità di tre apparecchi, che indichiamo con i tipi A, B, C. Esaminiamo singolarmente i tre apparecchi, e calcoliamo per ciascuno di essi la frequenza relativa dei guasti.

- *Apparecchio di tipo A (esaminati 54 casi)*

	Frequenza assoluta	*Frequenza relativa*
Guasti rilevati	7	0,130
Funzionamento corretto	47	0,870
Totale	*54*	*1,000*

- *Apparecchio di tipo B (esaminati 96 casi)*

	Frequenza assoluta	*Frequenza relativa*
Guasti rilevati	9	0,094
Funzionamento corretto	87	0,906
Totale	*96*	*1,000*

- *Apparecchio di tipo C (esaminati 32 casi)*

	Frequenza assoluta	*Frequenza relativa*
Guasti rilevati	26	0,188
Funzionamento corretto	6	0.812
Totale	*32*	*1,000*

Quanto ora ottenuto ribalta le conclusioni frettolosamente desunte dall'esame della sola frequenza assoluta dei guasti (il lettore è invitato a rivedere l'esempio 3). Infatti, confrontando le frequenze *relative* di guasto, notiamo che la minima di esse (0,094) è quella calcolata per l'apparecchio di tipo B, il quale deve dunque essere considerato di massima affidabilità; ad esso segue l'apparecchio di tipo A (frequenza relativa di guasto: 0,130) ed infine l'apparecchio di tipo C (frequenza relativa di guasto: 0,188), che si rivela pertanto il meno affidabile dei tre considerati.

----

2. VARIABILI STATISTICHE

2.1. Che cos'è una variabile statistica?

Le considerazioni introdotte, e in particolare i concetti di frequenza assoluta e di frequenza relativa, consentono di introdurre la *variabile statistica*.

Riflettiamo sul procedimento presentato nella sezione 1 per definire la frequenza assoluta e la frequenza relativa: abbiamo considerato inizialmente un insieme A, costituito dagli elementi interessati dalla nostra analisi (ad esempio: l'insieme degli allievi di un certo Istituto); abbiamo quindi raggruppato questi elementi secondo un criterio prestabilito, individuando così una partizione di A in classi (ad esempio: il sottoinsieme degli allievi di sesso maschile ed il sottoinsieme degli allievi di sesso femminile); infine, abbiamo contato gli elementi di ciascun sottoinsieme, ottenendo le frequenze assolute. Per passare alle rispettive frequenze relative, è sufficiente dividere le frequenze assolute per il numero degli elementi di A.

Possiamo riassumere quanto ora notato nella definizione seguente.

Definizione 4. Una *variabile statistica* si definisce, in un insieme A, assegnando in esso una relazione di equivalenza $R \subseteq A \times A$ e considerando l'insieme quoziente A/R; ad ogni $X \in A/R$ può essere associata la cardinalità di X (ovvero il numero di elementi di A appartenenti a X), detta frequenza assoluta $f(X)$ del fenomeno (o della situazione) X.

Alternativamente, ad ogni $X \in A/R$ può essere associata la cardinalità di X divisa per la cardinalità di A (ovvero il numero di elementi di A appartenenti a X diviso per il numero degli elementi di A), detta frequenza relativa $p(X)$ del fenomeno (o della situazione) X.

Esempio 7. Consideriamo l'insieme A dei naturali primi minori di 50, la cui rappresentazione tabulare è:

$$A = \{2; 3; 5; 7; 11; 13; 17; 19; 23; 29; 31; 37; 41; 43; 47\}$$

Indichiamo con $|A|$ la cardinalità di A; risulta: $|A| = 15$.

Il nostro scopo è di analizzare la distribuzione dei numeri naturali primi minori di 50 nelle decine tra 0 e 10, tra 11 e 20, tra 21 e 30, tra 31 e 40 e tra 41 e 50, e di ricavare le corrispondenti frequenze assolute e relative. Per fare ciò, dobbiamo, innanzitutto, suddividere gli elementi di A in 5 sottoinsiemi, costituiti dai numeri naturali primi compresi tra 0 e 10, tra 11 e 20, tra 21 e 30, tra 31 e 40 e tra 41 e 50. Operazione praticamente non difficile, ma che cercheremo di formalizzare sulla base di quanto indicato nella definizione 4.

Introduciamo (almeno intuitivamente) la *funzione parte intera*:

$$a \to [a] = \text{il massimo intero non maggiore di } a$$

Ad esempio, risulta:

$$[0{,}2] = 0 \qquad [1{,}1] = 1 \qquad [2{,}3] = 2 \qquad \ldots$$

Possiamo ora assegnare in A la seguente relazione di equivalenza:

$$R = \{(x; y) \in A \times A: [x/10] = [y/10]\}$$

Scrivere $[x/10]$ significa considerare la (sola) cifra delle decine di x; dunque:

$$\begin{array}{llllll} x = 2 & \Rightarrow & x/10 = 0{,}2 & \Rightarrow & [x/10] = 0 \\ x = 11 & \Rightarrow & x/10 = 1{,}1 & \Rightarrow & [x/10] = 1 \\ x = 23 & \Rightarrow & x/10 = 2{,}3 & \Rightarrow & [x/10] = 2 \end{array}$$

Pertanto la condizione $[x/10] = [y/10]$ significa:

la cifra delle decine di x è uguale alla cifra delle decine di y

La relazione R è di equivalenza (lasciamo la verifica al lettore) e in base ad essa possiamo ricavare l'insieme quoziente A/R, ovvero le classi di equivalenza in cui viene suddiviso A, ciascuna delle quali è costituita dagli elementi di A che hanno la stessa cifra delle decine:

$$\{2; 3; 5; 7\} \quad = \quad \{x \in A: \text{la cifra delle decine di } x \text{ è } 0\}$$

(brevemente) = X_0

{11; 13; 17; 19}	=	$\{x \in A$: la cifra delle decine di x è 1} (brevemente) = X_1
{23; 29}	=	$\{x \in A$: la cifra delle decine di x è 2} (brevemente) = X_2
{31; 37}	=	$\{x \in A$: la cifra delle decine di x è 3} (brevemente) = X_3
{41; 43; 47}	=	$\{x \in A$: la cifra delle decine di x è 4} (brevemente) = X_4

Per ricavare le frequenze assolute, è sufficiente "contare" gli elementi di ciascuna classe di equivalenza:

$f(X_0) = |X_0| = 4$
$f(X_1) = |X_1| = 4$
$f(X_2) = |X_2| = 2$
$f(X_3) = |X_3| = 2$
$f(X_4) = |X_4| = 3$

Si osservi che risulta: $\sum_{i=0}^{4} f(X_i) = |A| = 15$

Le frequenze relative si trovano dividendo le frequenze assolute per |A| = 15:

$p(X_0) = f(X_0)/|A| = |X_0|/|A| = 4/15 = 0{,}267$
$p(X_1) = f(X_1)/|A| = |X_1|/|A| = 4/15 = 0{,}267$
$p(X_2) = f(X_2)/|A| = |X_2|/|A| = 2/15 = 0{,}133$
$p(X_3) = f(X_3)/|A| = |X_3|/|A| = 2/15 = 0{,}133$
$p(X_4) = f(X_4)/|A| = |X_4|/|A| = 3/15 = 0{,}200$

Si osservi che risulta (com'era peraltro prevedibile): $\sum_{i=0}^{4} p(X_i) = 1{,}000$

2.2. La rappresentazione grafica

Utile e importante, in molte applicazioni, è la rappresentazione grafica di una variabile statistica. Ciò consente una immediata valutazione delle caratteristiche del fenomeno, alla quale può far seguito un più dettagliato elenco delle informazioni elaborate (ad esempio, in termini di frequenze assolute e relative).

Una rappresentazione grafica molto usata è quella mediante *diagrammi a colonne* ed *istogrammi*. Di essa forniamo alcuni esempi.

Esempio 8. La variabile statistica illustrata nell'esempio precedente può essere rappresentata mediante un diagramma a colonne in cui in ogni colonna è rappresentata la frequenza assoluta (lasciamo al lettore la rappresentazione).

Esempio 9. La rappresentazione grafica mediante il diagramma a colonne può essere utilmente modificata in alcuni casi.

Occupiamoci di una variabile statistica descritta dai dati seguenti:

naturali primi minori di 20:	{2; 3; 5; 7; 11; 13; 17; 19}	totale: 8
naturali primi tra 20 e 30:	{23; 29}	totale: 2
naturali primi tra 30 e 50:	{31; 37; 41; 43; 47}	totale: 5

Se rappresentiamo graficamente la situazione analogamente a quanto fatto nell'esempio precedente, otteniamo un diagramma costituito da tre colonne. Un tale diagramma rappresenta correttamente la variabile statistica sopra assegnata? Apparentemente sì, ma la diversità degli intervalli in cui è stato suddiviso il campo di indagine (da 0 a 50) può ingenerare un malinteso: sulla base del diagramma, infatti, sembra che la distribuzione dei numeri primi tra 0 e 50 sia caratterizzata da una netta irregolarità, una sorta di brusca diminuzione di frequenza nella zona centrale, impressione che *non* corrisponde a realtà.

Esempio 10. Per ovviare all'inconveniente presentato nell'esempio precedente si ricorre ad un particolare diagramma a colonne, detto *istogramma*. In esso, le basi dei rettangoli formanti le colonne non sono costanti (come nell'ordinario diagramma a colonne), ma sono scelte proporzionali alle ampiezze degli intervalli di volta in volta considerati (nel nostro caso: 20; 10; 20).

Frequentemente utilizzate sono rappresentazioni con *diagrammi circolari* (detti *diagrammi a torta*) o nel piano cartesiano.

Esempio 11. Consideriamo la variabile statistica riferita alla scelta dell'indirizzo di laurea di 1045 studenti iscritti alla facoltà di ingegneria, distribuita secondo i dati seguenti:

256 studenti di ingegneria *civile*	(frequenza relativa: 0,245)
120 studenti di ingegneria *meccanica*	(frequenza relativa: 0,115)
54 studenti di ingegneria *chimica*	(frequenza relativa: 0,052)
385 studenti di ingegneria *elettronica*	(frequenza relativa: 0,368)

68 studenti di ingegneria *aeronautica*	(frequenza relativa: 0,065)
162 studenti di ingegneria *gestionale*	(frequenza relativa: 0,155)

(Totale frequenze relative: 1,000)

Il lettore realizzi il diagramma circolare che visualizza tale variabile statistica.

Esempio 12. La distribuzione degli alunni iscritti alla prima classe di un Istituto scolastico nel periodo 1985-1996 è riassunta nei dati della tabella seguente:

1985	iscritti 130 alunni	1991	iscritti 155 alunni
1986	iscritti 148 alunni	1992	iscritti 122 alunni
1987	iscritti 150 alunni	1993	iscritti 104 alunni
1988	iscritti 186 alunni	1994	iscritti 132 alunni
1989	iscritti 155 alunni	1995	iscritti 143 alunni
1990	iscritti 147 alunni	1996	iscritti 162 alunni

Una rappresentazione dell'andamento delle iscrizioni può essere realizzata in un piano cartesiano. I punti rappresentanti ogni singolo dato saranno uniti con segmenti al fine di visualizzare in modo continuo l'andamento. A tale proposito sottolineiamo che i dati rilevati sono *discreti*, stabiliti anno per anno; il modo più rigoroso di rappresentare dei dati come questi in un diagramma cartesiano, quindi, dovrebbe essere *con dei singoli punti*, non con una linea continua.

----

3. MODA E MEDIANA

3.1. Moda

Consideriamo una variabile statistica costituita da fenomeni X_i e da una corrispondente distribuzione di frequenze assolute $f(X_i)$ o relative $p(X_i)$. Spesso è interessante tenere in particolare considerazione il fenomeno X con la *massima* frequenza (si pensi ad una consultazione elettorale: il vincitore è il candidato che raggiunge la massima frequenza nella distribuzione dei voti!); in altre occasioni, può essere opportuno riferirsi al fenomeno che, pur non essendo quello di massima frequenza, si colloca in "posizione centrale" rispetto agli altri.

In questa sezione introdurremo alcune importanti nozioni per una variabile statistica che ci consentiranno di formalizzare quanto sopra accennato.

Definizione 5. Data una variabile statistica, costituita da alcuni fenomeni X_i e da una corrispondente distribuzione di frequenze assolute $f(X_i)$ o relative $p(X_i)$, si dice *moda* il fenomeno X_M al quale corrisponde la massima frequenza (assoluta o relativa).

Esempio 13. Consideriamo nuovamente la variabile statistica presentata e visualizzata nell'esempio 11, riferita alla scelta dell'indirizzo di laurea di 1045 studenti iscritti alla facoltà di ingegneria, distribuita secondo i dati seguenti:

256 studenti di ingegneria *civile* (frequenza relativa: 0,245)
120 studenti di ingegneria *meccanica* (frequenza relativa: 0,115)
54 studenti di ingegneria *chimica* (frequenza relativa: 0,052)
385 studenti di ingegneria *elettronica* (frequenza relativa: 0,368)
68 studenti di ingegneria *aeronautica* (frequenza relativa: 0,065)
162 studenti di ingegneria gestionale (frequenza relativa: 0,155)

La moda è individuata dalla massima frequenza (assoluta o relativa); essa corrisponde alla scelta dell'indirizzo di ingegneria elettronica; infatti:

385 studenti di ingegneria *elettronica* (frequenza relativa: 0,368)

e le frequenze (385, assoluta, 0,368, relativa) sono le massime nella distribuzione considerata.

Esempio 14. Talvolta la moda non è unica; la distribuzione di valori:

1, 2, 2, 2, 2, 3, 3, 4, 5, 5, 5, 5, 6, 6

ha le *due* mode 2 e 5 (entrambe caratterizzate dalla frequenza assoluta 4).

In alcune circostanze, inoltre, il concetto di moda risulta non applicabile, ovvero non significativo; ad esempio, se le frequenze (assolute) di una distribuzione sono tutte 1, come accade nella distribuzione di numeri:

2, 3, 7, 9, 10, 34, 177, 178, 190

non è possibile individuare una moda (al più, *tutti* i valori sarebbero da considerare mode, con frequenza 1).

3.2. Mediana

Il concetto di mediana, collegato alla "posizione centrale" di un valore in un insieme di valori ordinati, è introdotto dalla definizione seguente.

Definizione 6. Data una distribuzione di n valori numerici, disposti in ordine (non crescente o non decrescente), si dice *mediana* il valore X_c che si trova nel posto centrale, se n è dispari, oppure la semisomma dei due valori centrali, se n è pari.

Esempio 15. Applichiamo la definizione di mediana a due distribuzioni, rispettivamente caratterizzate da un numero dispari ed un numero pari di dati.

La distribuzione di (9) valori:

1, 2, 5, 7, 8, 9, 11, 245, 1098

ha mediana $X_c = 8$.

La distribuzione di (10) valori:

−7, −1, 0, 3, 4, 6, 9, 21, 29, 100

ha mediana $X_c = 5$ (semisomma di 4 e 6).

Osservazione. I concetti di moda e di mediana, ora introdotti, sono riferiti a particolari caratteristiche della distribuzione di valori considerata (la massima frequenza, o la posizione centrale). Ciò non significa che essi possano essere considerati, in assoluto, *i valori più significativi* della distribuzione, quelli che meglio la rappresentano o la sintetizzano: nella sezione seguente, presenteremo un procedimento alternativo, e molto importante, per ricavare un unico valore da una distribuzione di più valori numerici.

Esempio 16. Al termine dell'anno scolastico, in ciascuna materia, da un insieme di voti (valutazioni numeriche espresse da numeri interi o frazionari compresi tra 1 e 10) è necessario "estrarre" un (unico) voto finale. Come possiamo procedere? È opportuno applicare i concetti di moda e di mediana?

Consideriamo i seguenti voti, riportati da uno studente in fisica:

5 5 5 5½ 6½ 7 8

Proviamo a calcolare il voto finale in fisica applicando i concetti introdotti:

- la *moda* è 5, in quanto la frequenza assoluta del valore 5 è la massima (compare 3 volte);
- la *mediana* è 5½, in quanto il valore 5½ si trova in "posizione centrale", nell'elenco ordinato dei voti.

In entrambi i casi ora ipotizzati, il nostro studente si sentirebbe ingiustamente maltrattato: 5 è un voto troppo basso, perché nonostante i tre 5 iniziali, nella seconda parte del periodo considerato c'è stato un buon recupero! Anche 5½ sembra poco, nonostante sia il risultato "centrale"...

Rimandiamo dunque il calcolo del voto finale alla sezione seguente, dopo che avremo introdotto il concetto di *media aritmetica*.

---

4. MEDIE ARITMETICHE

4.1. Media aritmetica semplice

Poco fa, per esemplificare una tipica situazione da affrontare con i metodi della statistica, è stato proposto il seguente quesito:

qual è, mediamente, l'altezza delle donne europee?

La domanda è incentrata su di un termine importante, *"mediamente"*, e richiede una valutazione complessiva riguardante un numero molto alto di soggetti. Sarebbe certamente insensato elencare le altezze di tutte le donne europee! Né appare produttiva l'applicazione delle nozioni di moda e di mediana, introdotte nella precedente sezione. Pertanto, è necessario introdurre un nuovo procedimento per "riassumere" in un (unico) numero le informazioni derivante da numerosi rilevamenti dello stesso tipo. Un metodo assai utile in molti casi simili a quello sopra accennato è introdotto dalla definizione seguente.

Definizione 7. Assegnati n numeri a_1, a_2, a_3, ..., a_n, si dice *media aritmetica semplice* di tali numeri il numero M ottenuto addizionando tutti i numeri assegnati e dividendo la somma per n:

$$M=\frac{1}{n}\sum_{i=1}^{n}a_i=\frac{a_1+a_2+a_3+\ldots+a_n}{n}$$

Esempio 17. Riprendiamo i dati dell'esempio 16: una diretta applicazione della media aritmetica semplice ben nota agli studenti è il calcolo della "media" dei voti conseguiti, durante l'anno scolastico, in una determinata materia.

Ad esempio, ogni studente sa che se i suoi voti in fisica sono:

5 5 5 5½ 6½ 7 8

la sua "media" in fisica (la media aritmetica semplice dei numeri ora citati) è:

$$M=\frac{5+5+5+5{,}5+6{,}5+7+8}{7}=6$$

voto che sintetizza i risultati raggiunti dallo studente nel periodo considerato assai meglio di quanto ottenuto applicando la moda o la mediana (si riveda, a tale proposito, l'esempio precedente).

Esempio 18. La media aritmetica semplice si applica in geometria analitica:

- le coordinate cartesiane (x_M; y_M) del punto medio del segmento avente per estremi P_1 (x_1; y_1) e P_2 (x_2; y_2) sono:

$$x_M=\frac{x_1+x_2}{2} \qquad y_M=\frac{y_1+y_2}{2}$$

Pertanto: *l'ascissa e l'ordinata del punto medio di un segmento sono le medie aritmetiche rispettivamente delle ascisse e delle ordinate degli estremi di tale segmento*.

- le coordinate cartesiane (x_G; y_G) del baricentro del triangolo avente per vertici P_1 (x_1; y_1), P_2 (x_2; y_2) e P_3 (x_3; y_3) sono:

$$x_G=\frac{x_1+x_2+x_3}{3} \qquad y_G=\frac{y_1+y_2+y_3}{3}$$

Pertanto: *l'ascissa e l'ordinata del baricentro di un triangolo sono le medie aritmetiche rispettivamente delle ascisse e delle ordinate dei vertici di tale triangolo*.

4.2. Media aritmetica ponderata

La media aritmetica semplice è uno strumento spesso utilizzato nelle applicazioni pratiche, in molti campi. In alcuni casi, però, l'impiego della media aritmetica semplice per riassumere in un unico numero le informazioni derivanti da una pluralità di dati può risultare poco significativo, se non fuorviante.

Esempio 19. In un magazzino sono depositate 1000 taniche d'olio: alcune di esse (500) hanno la capacità di 5 litri l'una; alcune di esse (350) sono da 10 litri l'una, le rimanenti (150) sono da 15 litri l'una. Qual è la capacità media delle taniche depositate nel magazzino?

In questo caso, l'applicazione della media aritmetica semplice porterebbe al risultato (*errato*) di 10 litri:

$$\frac{5+10+15}{3}=\frac{30}{3}=10$$

Infatti, operando con la media aritmetica semplice si attribuisce ai tre dati (5 litri, 10 litri, 15 litri) la stessa importanza, ovvero la stessa influenza sul risultato finale. Ciò *non è corretto*; ad esempio, ben 500 taniche hanno la capacità di 5 litri, mentre solo 150 di 15 litri: questi due dati, dunque, non possono essere considerati allo stesso modo!

In situazioni come quella descritta nell'esempio, l'impiego della media aritmetica semplice può essere sostituito da quello della media aritmetica ponderata.

Definizione 8. Assegnata una variabile statistica costituita da m numeri a_1, a_2, ..., a_m, a ciascuno dei quali corrisponda una frequenza assoluta f_1, f_2, ..., f_m, si dice *media aritmetica ponderata* di tali numeri il numero M ottenuto addizionando tutti i numeri assegnati, ciascuno moltiplicato per la propria frequenza assoluta, e dividendo la somma per la somma delle frequenze assolute:

$$M=\frac{1}{f_1+f_2+f_3+\ldots+f_m}\sum_{i=1}^{m}a_i f_i=\frac{a_1f_1+a_2f_2+a_3f_3+\ldots+a_mf_m}{f_1+f_2+f_3+\ldots+f_m}$$

I coefficienti $f_1, f_2, f_3, \ldots, f_m$, si dicono *pesi*.

Osservazione. La definizione di media aritmetica ponderata è strettamente collegata alla definizione di media aritmetica semplice. Infatti, nella situazione prevista dalla definizione 8, è: $f_1+f_2+f_3+...+f_m = n$, numero totale dei casi considerati; ricavando la media aritmetica semplice di tutti questi n casi, otteniamo:

$$M = \frac{(\underbrace{a_1 + a_1 + \ldots + a_1}_{f_1 \text{ addendi}}) + (\underbrace{a_2 + a_2 + \ldots + a_2}_{f_2 \text{ addendi}}) + \ldots + (\underbrace{a_m + a_m + \ldots + a_m}_{f_m \text{ addendi}})}{n} =$$

$$= \frac{a_1 f_1 + a_2 f_2 + a_3 f_3 + \ldots + a_m f_m}{f_1 + f_2 + f_3 + \ldots + f_m}$$

che coincide con l'espressione della media aritmetica ponderata.

Esempio 20. La capacità media delle taniche depositate nel magazzino, secondo i dati dell'esempio precedente, può essere calcolata mediante la media aritmetica ponderata:

$$M = \frac{500 \cdot 5 + 350 \cdot 10 + 150 \cdot 15}{1000} = 8,25$$

La capacità media è dunque di 8,25 litri.

La proposizione seguente introdurrà una definizione alternativa di media aritmetica ponderata.

Proposizione 1. Data una variabile statistica costituita da m numeri a_1, a_2, a_3, ..., a_m, a ciascuno dei quali corrisponda una frequenza relativa p_1, p_2, p_3, ..., p_m, la *media aritmetica ponderata M* di tali numeri si ottiene addizionando tutti i numeri assegnati, ciascuno moltiplicato per la propria frequenza relativa:

$$M = \sum_{i=1}^{m} a_i p_i = a_1p_1+a_2p_2+a_3p_3+...+a_mp_m$$

Dimostrazione. Dalla definizione di media aritmetica ponderata:

$$M = \frac{a_1 f_1 + a_2 f_2 + a_3 f_3 + \ldots + a_m f_m}{f_1 + f_2 + f_3 + \ldots + f_m}$$

ricordando che $f_1+f_2+f_3+...+f_m = n$ (numero totale dei casi considerati) e che, per ogni indice i: $p_i = \frac{f_i}{n}$, segue direttamente la tesi:

$$M = \frac{a_1 f_1 + a_2 f_2 + a_3 f_3 + \ldots + a_m f_m}{n} = a_1p_1+a_2p_2+a_3p_3+\ldots+a_mp_m \quad \blacksquare$$

Esempio 21. Sempre con riferimento all'esempio 19, i valori (in litri):

$c_1 = 5$ $\qquad$ $c_2 = 10$ $\qquad$ $c_3 = 15$

compaiono con frequenze assolute:

$f_1 = 500$ $\qquad$ $f_2 = 350$ $\qquad$ $f_3 = 150$

e con frequenze relative (essendo $n = 1000$):

$p_1 = 0{,}500$ $\qquad$ $p_2 = 0{,}350$ $\qquad$ $p_3 = 0{,}150$

Applicando la proposizione precedente, possiamo calcolare la media aritmetica ponderata nella forma:

$$M = 5\cdot 0{,}500+10\cdot 0{,}350+15\cdot 0{,}150 = 8{,}25$$

4.3. Scarti

Definizione 9. Assegnati n numeri $a_1, a_2, a_3, \ldots, a_n$, sia α un valore compreso tra il minimo ed il massimo di tali numeri; i valori:

$$s_1 = a_1-\alpha, \qquad s_2 = a_2-\alpha, \qquad s_3 = a_3-\alpha, \qquad \ldots, \qquad s_n = a_n-\alpha$$

si dicono *scarti* dei rispettivi numeri $a_1, a_2, a_3, \ldots, a_n$ considerati da α.

Spesso considereremo gli scarti di alcuni valori rispetto alla loro media aritmetica M (semplice o ponderata), come nell'esempio seguente.

Esempio 22. Riferiamoci ancora una volta all'esempio 19; gli scarti di:

$c_1 = 5$ $\qquad$ $c_2 = 10$ $\qquad$ $c_3 = 15$

dalla media (ponderata) $M = 8{,}25$ sono rispettivamente:

$s_1 = -3{,}25$ $s_2 = 1{,}75$ $s_3 = 6{,}75$

Proposizione 2. Se M è la media aritmetica semplice di n numeri $a_1, a_2, a_3, ..., a_n$, allora la somma degli scarti da M di $a_1, a_2, a_3, ..., a_n$ è 0:

$$\sum_{i=1}^{n}(a_i - M) = (a_1 - M) + (a_2 - M) + \ldots + (a_n - M) = 0$$

Dimostrazione. In base alla definizione di media aritmetica semplice:

$$\sum_{i=1}^{n}(a_i - M) = (a_1 - M) + (a_2 - M) + \ldots + (a_n - M) =$$
$$= a_1 + a_2 + a_3 + \ldots + a_n - n \cdot M =$$
$$= a_1 + a_2 + a_3 + \ldots + a_n - n \cdot \frac{a_1 + a_2 + a_3 + \ldots + a_n}{n} = 0 \quad \blacksquare$$

Analogamente può essere dimostrato il risultato seguente.

Proposizione 3. Se M è la media aritmetica ponderata di $a_1, a_2, a_3, ..., a_m$ con pesi rispettivi $f_1, f_2, f_3, ..., f_m$, allora:

$$\sum_{i=1}^{m}(a_i - M)f_i = (a_1 - M)f_1 + (a_2 - M)f_2 + \ldots + (a_m - M)f_m = 0$$

Se ai valori $a_1, a_2, a_3, ..., a_m$ sono associate le frequenze relative $p_1, p_2, p_3, ..., p_m$, analogamente:

$$\sum_{i=1}^{m}(a_i - M)p_i = (a_1 - M)p_1 + (a_2 - M)p_2 + \ldots + (a_m - M)p_m = 0$$

Esempio 23. Con riferimento all'esempio 19, ai valori (in litri):

$c_1 = 5$ $c_2 = 10$ $c_3 = 15$

sono associate le frequenze assolute:

$f_1 = 500$ $f_2 = 350$ $f_3 = 150$

La media aritmetica ponderata è: $M = 8{,}25$. Calcoliamo ora:

$$\sum_{i=1}^{3}(c_i - M)f_i = (5-8{,}25)\cdot 500+(10-8{,}25)\cdot 350+(15-8{,}25)\cdot 150 = 0$$

come previsto dalla proposizione precedente.

Ci limitiamo ad enunciare le proposizioni seguenti.

Proposizione 4. Se M è la media aritmetica semplice di n numeri a_1, a_2, a_3, ..., a_n, allora la somma dei quadrati degli scarti da $\alpha \in \mathbf{R}$ di a_1, a_2, a_3, ..., a_n:

$$\sum_{i=1}^{n}(a_i - \alpha)^2 = (a_1 - \alpha)^2 + (a_2 - \alpha)^2 + \ldots + (a_n - \alpha)^2$$

è minima quando $\alpha = M$.

Proposizione 5. Se M è la media aritmetica ponderata di a_1, a_2 , ..., a_m, allora:

$$\sum_{i=1}^{m}(a_i - \alpha)^2 f_i = (a_1 - \alpha)^2 f_1 + (a_2 - \alpha)^2 f_2 + \ldots + (a_m - \alpha)^2 f_m$$

è minima quando $\alpha = M$. Se ai valori a_1, a_2, ..., a_m sono associate le frequenze relative p_1, p_2, ..., p_m, analogamente:

$$\sum_{i=1}^{m}(a_i - \alpha)^2 p_i = (a_1 - \alpha)^2 p_1 + (a_2 - \alpha)^2 p_2 + \ldots + (a_m - \alpha)^2 p_m$$

è minima quando $\alpha = M$.

Le dimostrazioni delle proposizioni ora enunciate possono essere ottenute generalizzando quanto illustrato nell'esempio seguente.

Esempio 24. Riprendiamo in considerazione i dati dell'esempio 19: ai valori:

$c_1 = 5$ $\qquad\qquad$ $c_2 = 10$ $\qquad\qquad$ $c_3 = 15$

(in litri) sono associate le frequenze assolute:

$f_1 = 500$ $\qquad\qquad$ $f_2 = 350$ $\qquad\qquad$ $f_3 = 150$

La media aritmetica ponderata è: $M = 8{,}25$. Calcoliamo ora:

$$\sum_{i=1}^{3}(c_i - M)^2 f_i =$$
$$= (5-8{,}25)^2 \cdot 500 + (10-8{,}25)^2 \cdot 350 + (15-8{,}25)^2 \cdot 150 = 13187{,}5$$

Tentiamo ora di ottenere un valore inferiore dell'espressione precedente, sostituendo a M il valore $M+\delta$, con $\delta \neq 0$:

$$\sum_{i=1}^{3}(c_i - M - \delta)^2 f_i =$$
$$= (5-8{,}25-\delta)^2 \cdot 500 + (10-8{,}25-\delta)^2 \cdot 350 + (15-8{,}25-\delta)^2 \cdot 150 =$$
$$= 13187{,}5 + 1000\delta^2$$

Questo valore *non* è inferiore al precedente (in accordo con quanto affermato nella proposizione 5): per ogni scelta di $\delta \neq 0$, infatti, abbiamo aggiunto al valore precedente la quantità (positiva) $1000\delta^2$.

---

5. MEDIE GEOMETRICHE

5.1. Media geometrica semplice

La media aritmetica (semplice e ponderata) non è l'unico concetto mediante il quale si descrivono in un (unico) numero le informazioni derivante da più rilevamenti dello stesso tipo. Un concetto per molti versi analogo alla media studiata nella sezione precedente è introdotto dalla definizione seguente.

Definizione 10. Assegnati n numeri $a_1, a_2, a_3, ..., a_n$, si dice *media geometrica semplice* di tali numeri il numero M_g ottenuto moltiplicando tutti i numeri assegnati ed estraendo la radice n-esima del prodotto ottenuto:

$$M_g = \sqrt[n]{\prod_{i=1}^{n} a_i} = \sqrt[n]{a_1 \cdot a_2 \cdot a_3 \cdot \ldots \cdot a_n}$$

Esempio 25. La media geometrica semplice dei numeri 2, 4, 64 è:

$$M_g = \sqrt[3]{2 \cdot 4 \cdot 64} = \sqrt[3]{512} = 8$$

5.2. Media geometrica ponderata

Analogamente a quanto fatto nel caso delle medie aritmetiche, introduciamo la media geometrica ponderata.

Definizione 11. Assegnata una variabile statistica costituita da *m* numeri a_1, a_2, ..., a_m, a ciascuno dei quali corrisponda una frequenza assoluta f_1, f_2, ..., f_m, si dice *media geometrica ponderata* di tali numeri il numero M_g ottenuto moltiplicando tutti i numeri assegnati, ciascuno elevato alla propria frequenza assoluta, ed estraendo la radice $(f_1+f_2+...+f_m)$-esima del prodotto ottenuto:

$$M_g = \sqrt[f_1+f_2+f_3+...+f_m]{\prod_{i=1}^{m} a_i^{f_i}} = \sqrt[f_1+f_2+f_3+...+f_m]{a_1^{f_1} \cdot a_2^{f_2} \cdot a_3^{f_3} \cdot ... \cdot a_m^{f_m}}$$

I coefficienti f_1, f_2, f_3, ..., f_m, si dicono *pesi*.

Esempio 26. La media geometrica ponderata dei numeri 2, 3, 5 con pesi rispettivi 4, 4, 2 è:

$$M_g = \sqrt[4+4+2]{2^4 \cdot 3^4 \cdot 5^2} = \sqrt[10]{32400} = 2{,}825$$

Ci limitiamo ad enunciare la proposizione seguente.

Proposizione 6. Assegnata una variabile statistica costituita da *m* numeri a_1, a_2, a_3, ..., a_m, a ciascuno dei quali corrisponda una frequenza relativa p_1, p_2, p_3, ..., p_m, la *media geometrica ponderata* M_g di tali numeri si ottiene moltiplicando tutti i numeri assegnati, ciascuno elevato alla propria frequenza relativa:

$$M_g = \prod_{i=1}^{m} a_i^{p_i} = a_1^{p_1} \cdot a_2^{p_2} \cdot a_3^{p_3} \cdot ... \cdot a_m^{p_m}$$

Esempio 27. La media geometrica ponderata dei numeri 2, 3, 5 con pesi rispettivi 0,4, 0,4, 0,2 è:

$$M_g = 2^{0,4} \cdot 3^{0,4} \cdot 5^{0.2} = 2{,}825$$

risultato che coincide con quello ottenuto nell'esempio precedente (il lettore si renderà conto che la situazione esaminata è la stessa).
----

6. MEDIE ARMONICHE

6.1. Media armonica semplice

Definizione 12. Assegnati n numeri a_1, a_2, a_3, ..., a_n, si dice *media armonica semplice* M_h di tali numeri il reciproco della media aritmetica dei reciproci degli n numeri dati:

$$M_h = \frac{n}{\sum_{i=1}^{n} \frac{1}{a_i}} = \frac{n}{\frac{1}{a_1} + \frac{1}{a_2} + \frac{1}{a_3} + \ldots \frac{1}{a_n}}$$

Esempio 28. La media armonica semplice dei numeri $\frac{1}{3}$ e $\frac{1}{7}$ è:

$$M_h = \frac{2}{3+7} = \frac{1}{5} = 0{,}2$$

(Il lettore è invitato a notare che 5 è la media aritmetica semplice di 3 e 7, come previsto dalla precedente definizione).

6.2. Media armonica ponderata

Analogamente a quanto fatto nel caso delle medie aritmetiche e geometriche, introduciamo la media armonica ponderata.

Definizione 13. Assegnata una variabile statistica costituita da m numeri a_1, a_2, a_3, ..., a_m, a ciascuno dei quali corrisponda una frequenza assoluta f_1, f_2, f_3, ..., f_m, si dice *media armonica ponderata* di tali numeri il numero M_h:

$$M_h = \frac{f_1 + f_2 + f_3 + \ldots + f_m}{\sum_{i=1}^{m} \frac{f_i}{a_i}} = \frac{f_1 + f_2 + f_3 + \ldots + f_m}{\frac{f_1}{a_1} + \frac{f_2}{a_2} + \frac{f_3}{a_3} + \ldots \frac{f_m}{a_m}}$$

I coefficienti $f_1, f_2, f_3, ..., f_m$, si dicono *pesi*.

Esempio 29. La media armonica ponderata dei numeri $\frac{1}{2}, \frac{1}{3}, \frac{1}{4}$ con pesi rispettivi 2, 2, 1 è:

$$M_h = \frac{5}{2 \cdot 2 + 3 \cdot 2 + 4 \cdot 1} = \frac{5}{14} = 0{,}357$$

Ci limitiamo ad enunciare la proposizione seguente.

Proposizione 7. Assegnata una variabile statistica costituita da m numeri a_1, a_2, a_3, ..., a_m, a ciascuno dei quali corrisponda una frequenza relativa p_1, p_2, p_3, ..., p_m, la *media armonica ponderata* M_h di tali numeri si ottiene:

$$M_h = \frac{1}{\sum_{i=1}^{m} \frac{p_i}{a_i}} = \frac{1}{\frac{p_1}{a_1} + \frac{p_2}{a_2} + \frac{p_3}{a_3} + \ldots \frac{p_m}{a_m}}$$

Esempio 30. La media armonica ponderata dei numeri $\frac{1}{2}, \frac{1}{3}, \frac{1}{4}$ con pesi rispettivi 0,4, 0,4, 0,2 è:

$$M_h = \frac{1}{2 \cdot 0{,}4 + 3 \cdot 0{,}4 + 4 \cdot 0{,}2} = 0{,}357$$

risultato che coincide con quello ottenuto nell'esempio precedente.

----

7. SCARTO SEMPLICE MEDIO VARIANZA E SCARTO QUADRATICO MEDIO

7.1. Scarto semplice medio

Abbiamo poco fa introdotto il concetto di *scarto*. Riprenderemo ora tale nozione, proponendo alcune importanti definizioni. La prima di esse riguarda n numeri a_1, a_2, a_3, ..., a_n, e la loro media aritmetica semplice M; prenderemo in considerazione le quantità:

$$|a_1-M|, \qquad |a_2-M|, \qquad |a_3-M|, \qquad ..., \qquad |a_n-M|$$

cioè i valori assoluti degli scarti dalla media M.

Definizione 14. Assegnati n numeri a_1, a_2, ..., a_n, sia M la loro media aritmetica semplice. Si dice *scarto semplice medio della media aritmetica* s_M dei valori considerati la media aritmetica semplice dei valori assoluti degli scarti da M:

$$s_M = \frac{1}{n}\sum_{i=1}^{n}|a_i - M| = \frac{|a_1 - M| + |a_2 - M| + |a_3 - M| + \ldots + |a_n - M|}{n}$$

Esempio 31. Sono assegnati i 5 valori (ripresi dall'esempio 16):

5 5 5 5,5 6,5 7 8

La loro media aritmetica semplice (esempio 17) è: $M = 6$. Lo scarto semplice medio della media aritmetica dei valori sopra riportati è:

$$s_M = \frac{|5-6| + |5-6| + |5-6| + |5,5-6| + |6,5-6| + |7-6| + |8-6|}{7} =$$
$$= \frac{1+1+1+0,5+0,5+1+2}{7} = 1$$

Lo scarto semplice medio della media aritmetica può riferirsi alla media aritmetica ponderata se i valori in gioco costituiscono una variabile statistica.

Definizione 15. Assegnata una variabile statistica costituita da m numeri a_1, a_2, a_3, ..., a_m, a ciascuno dei quali corrisponda una frequenza assoluta f_1, f_2, f_3, ..., f_m, sia M la loro media aritmetica ponderata. Si dice *scarto semplice medio della media aritmetica* s_M dei valori considerati la media aritmetica ponderata dei valori assoluti degli scarti da M:

$$s_M = \frac{1}{f_1 + f_2 + f_3 + \ldots + f_m}\sum_{i=1}^{m}|a_i - M|f_i =$$
$$= \frac{|a_1 - M|f_1 + |a_2 - M|f_2 + |a_3 - M|f_3 + \ldots + |a_m - M|f_m}{f_1 + f_2 + f_3 + \ldots + f_m}$$

Esempio 32. Riprendiamo i dati dell'esempio 19; in un magazzino sono depositate 1000 taniche d'olio: alcune di esse (500) hanno la capacità di 5 litri l'una; alcune (350) sono da 10 litri l'una, le rimanenti (150) sono da 15 litri l'una. La media aritmetica ponderata delle capacità (esempio 20) è: $M = 8{,}25$ (in litri). Lo scarto semplice medio della media aritmetica dei valori riportati è:

$$s_M = \frac{|5-8{,}25|\cdot 500 + |10-8{,}25|\cdot 350 + |15-8{,}25|\cdot 150}{1000} =$$
$$= \frac{3{,}25\cdot 500 + 1{,}75\cdot 350 + 6{,}75\cdot 150}{1000} = 3{,}25$$

Osservazione. Ci limitiamo ad accennare che se una variabile statistica è costituita da m numeri $a_1, a_2, a_3, ..., a_m$, a ciascuno dei quali corrisponde una frequenza *relativa* $p_1, p_2, p_3, ..., p_m$, essendo M la loro media aritmetica ponderata, allora lo scarto semplice medio della media aritmetica dei valori considerati è:

$$s_M = \sum_{i=1}^{m} |a_i - M| p_i =$$
$$= |a_1 - M| p_1 + |a_2 - M| p_2 + |a_3 - M| p_3 + \ldots + |a_m - M| p_m$$

(analogamente a quanto affermato nella proposizione 1).

7.2. Varianza

Definizione 16. Assegnati n numeri $a_1, a_2, a_3, ..., a_n$, sia M la loro media aritmetica semplice. Si dice *varianza* σ^2 dei valori considerati la media aritmetica semplice dei quadrati degli scarti da M:

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(a_i - M)^2 =$$
$$= \frac{(a_1 - M)^2 + (a_2 - M)^2 + (a_3 - M)^2 + \ldots + (a_n - M)^2}{n}$$

Esempio 33. Consideriamo nuovamente i 5 valori (ripresi dall'esempio 16):

5 5 5 5,5 6,5 7 8

La loro media aritmetica semplice (esempio 17) è: $M = 6$. La varianza è:

$$\sigma^2 = \frac{(5-6)^2 + (5-6)^2 + (5-6)^2 + (5{,}5-6)^2 + (6{,}5-6)^2 + (7-6)^2 + (8-6)^2}{7}$$
$$= \frac{1^2 + 1^2 + 1^2 + 0{,}5^2 + 0{,}5^2 + 1^2 + 2^2}{7} = 1{,}214$$

Anche la varianza può essere riferita alla media aritmetica ponderata, nel caso in cui i valori a disposizione costituiscano una variabile statistica.

Definizione 17. Assegnata una variabile statistica costituita da m numeri a_1, a_2, a_3, ..., a_m, a ciascuno dei quali corrisponda una frequenza assoluta f_1, f_2, f_3, ..., f_m, sia M la loro media aritmetica ponderata. Si dice *varianza* σ^2 dei valori considerati la media aritmetica ponderata dei quadrati degli scarti da M:

$$\sigma^2 = \frac{1}{f_1 + f_2 + f_3 + ... + f_m} \sum_{i=1}^{m} (a_i - M)^2 f_i =$$
$$= \frac{(a_1 - M)^2 f_1 + (a_2 - M)^2 f_2 + (a_3 - M)^2 f_3 + ... + (a_m - M)^2 f_m}{f_1 + f_2 + f_3 + ... + f_m}$$

Esempio 34. Riprendiamo i dati dell'esempio 19, che ricordiamo ancora una volta: in un magazzino sono depositate 1000 taniche d'olio: alcune di esse (500) hanno la capacità di 5 litri l'una; alcune (350) sono da 10 litri l'una, le rimanenti (150) sono da 15 litri l'una. Sappiamo che la media aritmetica ponderata delle capacità (calcolata nell'esempio 20) è (in litri):

$$M = 8{,}25$$

La varianza dei valori sopra riportati è:

$$\sigma^2 = \frac{(5-8{,}25)^2 \cdot 500 + (10-8{,}25)^2 \cdot 350 + (15-8{,}25)^2 \cdot 150}{1000} =$$
$$= \frac{3{,}25^2 \cdot 500 + 1{,}75^2 \cdot 350 + 6{,}75^2 \cdot 150}{1000} = 13{,}187$$

Osservazione. Ci limitiamo ad accennare che se una variabile statistica è costituita da m numeri a_1, a_2, a_3, ..., a_m, a ciascuno dei quali corrisponde una frequenza *relativa* $p_1, p_2, p_3, ..., p_m$, essendo M la loro media aritmetica ponderata, allora la varianza dei valori considerati è:

$$\sigma^2 = \sum_{i=1}^{m} (a_i - M)^2 p_i =$$
$$= (a_1 - M)^2 p_1 + (a_2 - M)^2 p_2 + (a_3 - M)^2 p_3 + ... + (a_m - M)^2 p_m$$

(analogamente a quanto affermato nella proposizione 1 e nell'osservazione al termine del paragrafo precedente).

7.3. Scarto quadratico medio

Definizione 18. Assegnati n numeri $a_1, a_2, a_3, ..., a_n$, sia M la loro media aritmetica semplice. Si dice *scarto quadratico medio* σ dei valori considerati la radice quadrata della media aritmetica semplice dei quadrati degli scarti da M:

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (a_i - M)^2} =$$
$$= \sqrt{\frac{(a_1 - M)^2 + (a_2 - M)^2 + (a_3 - M)^2 + ... + (a_n - M)^2}{n}}$$

Osservazione. Confrontando le definizioni di scarto quadratico medio e di varianza, possiamo concludere che lo scarto quadratico è la radice quadrata (aritmetica) della varianza.

Esempio 35. Consideriamo nuovamente i 5 valori (ripresi dall'esempio 16):

5 5 5 5,5 6,5 7 8

La loro media aritmetica semplice (esempio 17) è: $M = 6$. Lo scarto quadratico medio dei valori sopra riportati è:

$$\sigma = \sqrt{\frac{(5-6)^2 + (5-6)^2 + (5-6)^2 + (5{,}5-6)^2 + (6{,}5-6)^2 + (7-6)^2 + (8-6)^2}{7}}$$
$$= \sqrt{\frac{1^2 + 1^2 + 1^2 + 0{,}5^2 + 0{,}5^2 + 1^2 + 2^2}{7}} = \sqrt{1{,}2142...} = 1{,}102$$

Anche lo scarto quadratico medio può essere riferito alla media aritmetica ponderata, nel caso in cui i valori costituiscano una variabile statistica.

Definizione 19. Assegnata una variabile statistica costituita da m numeri a_1, a_2, a_3, ..., a_m, a ciascuno dei quali corrisponda una frequenza assoluta f_1, f_2, f_3, ..., f_m, sia M la loro media aritmetica ponderata. Si dice *scarto quadratico medio* σ dei valori considerati la radice quadrata della media aritmetica ponderata dei quadrati degli scarti da M:

$$\sigma = \sqrt{\frac{1}{f_1+f_2+f_3+\ldots+f_m}\sum_{i=1}^{m}(a_i-M)^2 f_i} =$$

$$= \sqrt{\frac{(a_1-M)^2 f_1+(a_2-M)^2 f_2+(a_3-M)^2 f_3+\ldots+(a_m-M)^2 f_m}{f_1+f_2+f_3+\ldots+f_m}}$$

Esempio 36. Riferiamoci sempre ai dati dell'esempio 19 che ricordiamo per comodità: in un magazzino sono depositate 1000 taniche d'olio: alcune di esse (500) hanno la capacità di 5 litri l'una; alcune (350) sono da 10 litri l'una, le rimanenti (150) sono da 15 litri l'una. La media aritmetica ponderata delle capacità (esempio 20) è (in litri): $M = 8{,}25$. Lo scarto quadratico medio è:

$$\sigma = \sqrt{\frac{(5-8{,}25)^2\cdot 500+(10-8{,}25)^2\cdot 350+(15-8{,}25)^2\cdot 150}{1000}} =$$

$$= \sqrt{\frac{3{,}25^2\cdot 500+1{,}75^2\cdot 350+6{,}75^2\cdot 150}{1000}} = \sqrt{13{,}1875} = 3{,}631$$

Osservazione. Ci limitiamo infine ad accennare che se una variabile statistica è costituita da m numeri a_1, a_2, a_3, ..., a_m, a ciascuno dei quali corrisponde una frequenza *relativa* p_1, p_2, p_3, ..., p_m, essendo M la loro media aritmetica ponderata, allora lo scarto quadratico medio dei valori considerati è:

$$\sigma = \sqrt{\sum_{i=1}^{m}(a_i-M)^2 p_i} =$$

$$= \sqrt{(a_1-M)^2 p_1+(a_2-M)^2 p_2+(a_3-M)^2 p_3+\ldots+(a_m-M)^2 p_m}$$

8. CONCENTRAZIONE

8.1. Il concetto di concentrazione

Il concetto di *concentrazione* è riferito specificamente a beni (entità materiali, economiche...) di uno stesso tipo che possono essere sommate tra di loro e che possono essere considerate di proprietà di un soggetto. Ad esempio, è possibile parlare di concentrazione di capitali, di redditi, di mezzi di produzione…

Indichiamo con a_1, a_2, a_3, ..., a_n le quantità di un bene rispettivamente appartenenti agli individui I_1, I_2, I_3, ..., I_n. Sia $A = a_1+a_2+a_3+...+a_n$ il totale dei beni a cui viene fatto riferimento. Con il termine *concentrazione* ci riferiamo alle caratteristiche con cui tale bene è distribuito tra gli individui esaminati.

- *Se i beni sono distribuiti uniformemente (se ogni individuo possiede una stessa quantità del bene totale) diciamo che la concentrazione è nulla.*
- *Se un solo individuo possiede interamente tutto il bene, diciamo che la concentrazione è massima.*

Esempio 37. Consideriamo le aziende A, B, C, D, che operano nella stessa regione in un determinato settore. Il fatturato annuo di tali aziende è:

Azienda A:	10	(in milioni di euro)
Azienda B:	15	(in milioni di euro)
Azienda C:	35	(in milioni di euro)
Azienda D:	65	(in milioni di euro)

Complessivamente, il fatturato di quel settore in quella regione è 125 (in milioni di euro). Esso si divide nelle quattro aziende secondo le percentuali:

Azienda A:	8 %
Azienda B:	12 %
Azienda C:	28 %
Azienda D:	52 %
Totale	*100 %*

I dati ora riportati possono essere rappresentati addizionando (*cumulando*) ordinatamente le percentuali delle quattro aziende (ciascuna di esse, dunque, rappresenta il 25 % del numero delle aziende operanti). Si ottengono allora le seguenti *percentuali cumulate*:

Percentuale aziende	Percentuale fatturato	
25 %	8 %	
50 %	20 %	(8 % + 12 %)
75 %	48 %	(8 % + 12 % + 28 %)
100 %	100 %	(8 % + 12 % + 28 % + 52 %)

8.2. Cenno sulla rappresentazione di Lorenz

Generalizzando quanto presentato nell'esempio, indichiamo con a_1, a_2, a_3, ..., a_n le quantità di un bene appartenenti agli individui I_1, I_2, I_3, ..., I_n; sia A = $a_1+a_2+a_3+...+a_n$ il totale dei beni a cui viene fatto riferimento. Siano p_1, p_2, p_3, ..., p_n le percentuali cumulate calcolate riferite ai possessori e b_1, b_2, b_3, ..., b_n le percentuali cumulate riferite alle quantità di bene posseduto.

Riportando in un riferimento cartesiano tali percentuali cumulate, possiamo determinare una curva, detta *curva di Lorenz*.

Si noti che in caso di concentrazione nulla (ovvero nel caso in cui a ciascuno degli n possessori spetti esattamente $1/n$ del bene totale A disponibile), la curva di Lorenz si riduce ad un segmento di retta, denominata *retta di concentrazione nulla* (o *retta di equidistribuzione*).

Nel caso invece in cui tutto il bene disponibile sia in possesso di un solo soggetto, la curva di Lorenz si riduce all'unione dei due segmenti ortogonali (*spezzata di concentrazione massima*).

L'area della parte di piano compresa tra la curva di Lorenz e la retta di concentrazione nulla è detta *area di concentrazione*. Il rapporto tra l'area di concentrazione e l'area della parte di piano compresa tra la retta di concentrazione nulla e la spezzata di concentrazione massima è detto *rapporto di concentrazione*. Può variare tra 0 (concentrazione nulla) e 1 (concentrazione massima).

Esempio 39. Nel caso esaminato nell'esempio precedente il lettore è invitato a costruire la rappresentazione grafica. Si calcola per l'area di concentrazione:

$$\frac{1}{2}-\frac{0{,}08\cdot 0{,}25+(0{,}08+0{,}20)\cdot 0{,}25+(0{,}20+0{,}48)\cdot 0{,}25+(0{,}48+1)\cdot 0{,}25}{2}=$$

$$=0{,}185$$

Rapporto di concentrazione: $\frac{0{,}185}{0{,}5}=0{,}37$
