Confronti fra grandezze statistiche

Le grandezze statistiche possono essere numeri, quantità, valori che si ottengono in seguito alla rilevazione di fenomeni. Ad esempio: numero è la frequenza della popolazione residente in Italia al 1° gennaio 2008: 59.619.290; quantità è la produzione in migliaia di quintali di uva da tavola in Italia nel 2007: 13.544; valore è l'ammontare in milioni di euro delle importazioni italiane di prodotti dell'agricoltura, silvicoltura e pesca nel 2007: 10.389 (da ISTAT, L'Italia in cifre 2009).

Ogni grandezza statistica è identificata oltre che dal fenomeno rilevato, dal luogo e dal tempo al quale la rilevazione si riferisce. E' perciò possibile confrontare:

- lo stesso fenomeno in luoghi diversi e nello stesso tempo (attenzione però che se si tratta di valori l'unità di misura monetaria deve essere identica);
- lo stesso fenomeno nello stesso luogo, ma in tempi diversi;
- fenomeni diversi nello stesso luogo e nello stesso tempo

Se fra due grandezze statistiche rilevate a e b ha senso effettuare un confronto, le operazioni che l'aritmetica offre allo scopo sono due: il rapporto e la differenza.

Confronti mediante rapporti
Essendo $a>0$ e $b>0$ si possono costruire i rapporti:

$$\frac{a}{b} \qquad \frac{a}{b}\cdot 100 \qquad \frac{b}{a} \qquad \frac{b}{a}\cdot 100$$

Il primo è un rapporto relativo ed indica quante volte b è contenuto a, o anche quante unità di a vi sono per ogni unità di b; il secondo è un rapporto percentuale e si ottiene moltiplicando per 100 il rapporto relativo. In modo analogo si interpretano gli altri due rapporti.
Se a e b sono espressi nella stessa unità di misura il loro rapporto è un numero adimensionale, mentre se a e b sono grandezze espresse in unità di misura diverse il risultato del loro rapporto è espresso nel rapporto dell'unità di misura del numeratore rispetto all'unità di misura del denominatore.
Alcuni rapporti sono detti tassi, ad esempio: tasso di occupazione, tasso di scolarità; altri indici: indice di vecchia, indice di dipendenza (cfr ISTAT, L'Italia in cifre 2009).

Confronti mediante differenze o variazioni
Se a e b sono due grandezze espresse nella stessa unità di misura, le differenze o variazioni che è possibile calcolare sono:

$$a-b \qquad b-a \qquad |a-b|$$

Esse indicano rispettivamente di quanto la prima quantità è maggiore della seconda, di quanto la seconda è maggiore della prima, di quanto le due sono diverse fra loro.
Le prime due sono differenze assolute o variazioni assolute, la terza è il valore assoluto della differenza.

Confronti mediante differenze o variazioni relative
Essendo a e b grandezze statistiche, quindi dati che vivono in un contesto fenomenico reale, ci possono essere situazioni in cui la loro differenza non riesce a spiegare compiutamente la variazione

la cui entità può dipendere da *a* , da *b* o da entrambe. Tenuto conto di ciò è possibile costruire le differenze o variazioni relative. Rispetto alla differenza assoluta *a-b* si possono calcolare:

$$\frac{a-b}{a} \qquad \frac{a-b}{b} \qquad \frac{a-b}{\frac{a+b}{2}}$$

In modo analogo si possono ottenere le differenze relative con riguardo a *b-a* e $|a-b|$. Le corrispondenti differenze percentuali si ottengono da quelle relative moltiplicandole per 100.
Le differenze o variazioni relative sono adimensionali.

Il confronto fra 2 grandezze statistiche effettuato in modi differenti porta a risultati differenti. Occorrerà perciò interpretarli con attenzione. Quando poi si conosce solo il risultato finale (così come viene spesso riportato dai media), ma non l'operazione che l'ha originato può essere opportuno esimersi dalla sua interpretazione.

Consideriamo a titolo di esempio la seguente distribuzione

Occupati in Italia per posizione professionale
Anno 2008, migliaia di persone

Posizione professionale	Occupati
Dipendente	17.446
Indipendente	5.959
Totale	23.405

Fonte: Istat, Italia in cifre 2009

La domanda "Di quanto sono diversi il numero degli occupati dipendenti e il numero degli occupati indipendenti?" può avere diverse risposte tutte egualmente valide come:

- gli occupati dipendenti sono 11.487 migliaia in più degli indipendenti [differenza assoluta];
- per ogni occupato indipendente vi sono 2,93 occupati dipendenti [rapporto relativo];
- per 100 occupati indipendenti vi sono 293 occupati dipendenti [rapporto percentuale];
- per ogni occupato indipendente vi sono 1,93 occupati dipendenti in più [differenza relativa];
- per 100 occupati indipendenti vi sono 193 occupati dipendenti in più [differenza percentuale].

NOTA BENE: in una distribuzione statistica, la frequenza relativa associata ad una modalità del carattere è un rapporto statistico, in particolare un rapporto di composizione.
Se il carattere è quantitativo continuo espresso in classi il rapporto fra la frequenza di un intervallo e la sua ampiezza è la densità di frequenza.

Maria Gabriella Ottaviani
"Sapienza" Università di Roma

I numeri indice

I numeri indice sono confronti mediante rapporto che vengono utilizzati quando si conosce la grandezza di un fenomeno in situazioni diverse.
Se le situazioni a cui si riferiscono i dati sono temporali, il loro insieme si chiama serie storica. Se le situazioni cui si riferiscono i dati sono geografiche si parla di serie territoriale.
Utilizzando i numeri indice si può analizzare l'andamento, l'evoluzione di un fenomeno nel tempo o nello spazio rispetto ad una situazione presa come riferimento (numeri indice a base fissa). Se le situazioni sono temporali e dunque ordinate è possibile anche confrontare l'evoluzione di un anno rispetto al precedente (numeri indice a base mobile).

Se si indica con t_i la situazione temporale generica e con a_i la grandezza assunta dal fenomeno A, per i=1,,,,,n, la tabella seguente mostra i numeri indice semplici a base fissa I_i e a base mobile i_i:

Situazione temporale	Grandezza del fenomeno	Numeri indice (base fissa t_1=1)	Numeri indice (base mobile)
t_1	a_1	1	-
t_2	a_2	$I_2 = \frac{a_2}{a_1}$	$i_2 = \frac{a_2}{a_1}$
t_i	a_i	$I_i = \frac{a_i}{a_1}$	$i_i = \frac{a_i}{a_{i-1}}$
t_{n-1}	a_{n-1}	$I_{n-1} = \frac{a_{n-1}}{a_1}$	$i_{n-1} = \frac{a_{n-1}}{a_{n-2}}$
t_n	a_n	$I_n = \frac{a_n}{a_1}$	$i_n = \frac{a_n}{a_{n-1}}$

I numeri indici sono sempre non negativi e sono numeri adimensionali.
In generale il numero indice rispetto al tempo t_i è:

- pari ad 1 se il fenomeno non si è modificato rispetto al tempo di riferimento;
- maggiore di 1 se è aumentato;
- minore di 1 se è diminuito.

Una volta calcolati i numeri indice è possibile calcolare la variazione relativa del fenomeno in esame sottraendo dall'indice il valore 1. La variazione trovata può anche essere espressa in percentuale.

Se la base è posta uguale a 100 i numeri indice sono detti percentuali e si ottengono moltiplicando per 100 i corrispondenti indici a base 1.

Si deve porre attenzione alla scelta della base in quanto anomalie riscontrate nell'epoca scelta potrebbero influire sulla determinazione dei numeri indice e dunque sulla interpretazione dell'andamento del fenomeno.

Ne “L’Italia in cifre 2009” l’ISTAT fornisce l’informazione sul numero di fumatori per 100 persone di 14 anni e più in Italia negli anni dal 1980 al 2008, i dati sono i seguenti:

Anno	Fumatori per 100 persone di 14 anni e più
1980	34,9
1983	31,1
1991	27,4
1995	25,3
2000	24,1
2008	22,2

Ci possiamo chiedere come si è modificato il numero di fumatori nel tempo prendendo come riferimento il 1980, oppure considerando un anno rispetto al precedente.

Anno	Numeri indici (base 1980=100)	Numeri indici (base variabile)
1980	100	-
1983	89	89
1991	78	88
1995	72	92
2000	69	95
2008	64	92

Come interpretare questa sequela di numeri?
Ad una lettura veloce pare che i numeri indici a base fissa e i numeri indici a base variabile ci dicano cose dissonanti, ma occorre interpretarli correttamente.
Se prendiamo il 2008, il numero 64 ci dice che i fumatori nel 2008 sono il 64% di quelli del 1980, mentre il numero 92 ci dice che i fumatori nel 2008 sono il 92% di quelli del 2000.
Dunque rispetto al 1980 il numero dei fumatori per 100 persone di 14 anni e più in Italia è diminuito con una variazione percentuale che nel tempo è stata di 11, 22, 28, 31, 36 punti percentuali, mentre da una situazione temporale all’altra la riduzione è stata: 11, 12, 8, 5, 8.

I valori indicati sono variazioni percentuali.
Rispetto ad un generico indice percentuale a base fissa la differenza o variazione relativa è:

$$I_i \cdot 100 - 100 = 100 \cdot \left(\frac{a_i - a_1}{a_1} \right);$$

rispetto ad un generico indice percentuale a base fissa la differenza o variazione relativa è:

$$i_i \cdot 100 - 100 = 100 \cdot \left(\frac{a_i - a_{i-1}}{a_{i-1}} \right).$$

Maria Gabriella Ottaviani
“Sapienza” Università di Roma