

## CONCETTI CHIAVE DELLA STATISTICA

Il nuovo curriculum fa espresso riferimento ad alcuni concetti chiave della statistica, è pertanto opportuno riferirsi ad essi. Questa breve nota non può essere esaustiva. Per una trattazione più ampia si veda ad esempio: G. Leti, *Statistica descrittiva*, il Mulino, 1983.

### Elementi di base: fenomeni collettivi e fenomeni individuali

La **statistica** è il mezzo che consente la conoscenza quantitativa dei fenomeni collettivi ossia il metodo appropriato, per indagare tali fenomeni.

Tra i fenomeni che si manifestano alla nostra esperienza distinguiamo quelli che possiamo percepire con **un'unica osservazione**, da quelli che, invece, possiamo conoscere soltanto se eseguiamo **una massa di osservazioni**. I secondi sono chiamati **fenomeni collettivi**.

### Unità statistica e carattere

**L'unità statistica** è l'oggetto dell'osservazione di ogni fenomeno individuale che costituisce il fenomeno collettivo.

**I caratteri** di una unità statistica sono le caratteristiche presenti in essa.

Ciascun carattere è presente in ogni unità con una **modalità**.

### Classificazione dei caratteri

La classificazione dei caratteri, ossia la classificazione del tipo di informazione raccolta sulle unità è fondamentale, perché fa da guida alle elaborazioni possibili sui dati. In particolare la rappresentazione grafica dei dati e le sintesi da utilizzare dipendono dal tipo di carattere.

Classificazione dei caratteri

| Carattere                                 |                     |              |          |
|---|---------------------|--------------|----------|
| qualitativo                               |                     | quantitativo |          |
| a modalità non ordinate<br>o<br>sconnesso | a modalità ordinate | discreto     | continuo |

Relazioni e operazioni tra le modalità a seconda del tipo di carattere

| Operazioni o relazioni tra le modalità | Carattere   |          |              |
|--|-------------|----------|--------------|
|  | qualitativo |          | quantitativo |
|  | sconnesso   | ordinato |              |
| uguaglianza e disuguaglianza           | SI          | SI       | SI           |
| ordinamento                            | NO          | SI       | SI           |
| addizione sottrazione                  | NO          | NO       | SI           |

Collettivo

statistico

•L'insieme di tutte le unità statistiche che sono omogenee rispetto ad uno o più caratteri è **un collettivo statistico** (o soltanto **un collettivo**) o **una popolazione**.

•Ognuno di tali caratteri è dunque presente in tutte le unità statistiche con la stessa modalità. (esempi: il collettivo degli italiani residenti al 1 gennaio 2006; il collettivo degli studenti iscritti alla nostra scuola il 15 settembre 2005; i giocatori della squadra di calcio XYZ ad una certa data; i libri di una certa biblioteca ad una certa data; ecc.)

Dal collettivo alle sue immagini:

*1° fase: schematizzazione*

•La prima fase di una qualunque ricerca nel campo delle scienze naturali o umane, che sia relativa ad un fenomeno collettivo, è la schematizzazione.

La **schematizzazione** consiste:

- ✓ nella **definizione** del fenomeno e degli obiettivi della ricerca
- ✓ nell'**individuazione** del collettivo in cui esso si realizza e delle unità statistiche che si intendono rilevare ed osservare;
- ✓ nella **scelta delle caratteristiche** del collettivo che interessano la ricerca.

*2° fase: osservazione*

•Alla schematizzazione segue l'**osservazione**; il suo scopo è quello di ricavare l'immagine del collettivo statistico concreto prescelta e di renderla percepibile a tutti.

• Tale immagine è il **collettivo statistico derivato**, ossia un insieme di dati statistici unitari ottenuti determinando, per i caratteri considerati, le modalità presenti in ciascuna unità del collettivo (un insieme dei questionari compilati e raccolti è un esempio di collettivo statistico derivato).

Il collettivo statistico derivato

•Un **dato statistico unitario** è il risultato dell'operazione di *determinazione della modalità con cui un carattere è presente in ciascuna unità del collettivo*.

•Un **collettivo statistico derivato** è il risultato delle operazioni di:

- ✓ schematizzazione del collettivo statistico concreto (di riferimento);
- ✓ determinazione, per quei caratteri che si sono considerati, delle modalità presenti in ciascuna unità del collettivo stesso.

Sistemazione dei dati

I dati raccolti possono essere sistemati mediante:

✓ **tabelle**

✓ **grafici.**

- Entrambi costituiscono un'immagine, una rappresentazione, del collettivo che è stato rilevato.
- Le rappresentazioni hanno lo scopo di esporre in forma chiara i risultati della rilevazione.

#### Distribuzione semplice

- L'operazione di determinazione delle modalità con cui ognuno dei caratteri si presenta in ciascuna unità del collettivo e di enumerazione delle unità statistiche che presentano le stesse modalità è la **distribuzione del collettivo secondo i caratteri considerati**.
- Nel curriculum della scuola secondaria di primo grado e nel biennio della secondaria di secondo grado è previsto di esaminare un carattere per volta. Perciò nei due ordini di scuola si trattano essenzialmente le **distribuzioni semplici**

#### La suddivisione in classi

- Effettuata l'enumerazione delle unità del collettivo e la determinazione delle modalità con cui il carattere considerato è presente in ciascuna unità, la popolazione viene **suddivisa in classi secondo il carattere considerato**.
- Per qualunque carattere è possibile sempre stabilire se due unità sono **uguali o diverse**: sono uguali se il carattere è presente con la stessa modalità, diverse in caso contrario. Le unità che presentano tutte la stessa modalità, possono essere considerate come un unico gruppo. E la popolazione può così essere suddivisa in sottopopolazioni, prive di elementi comuni, dette **classi di equivalenza o semplicemente classi**

#### La distribuzione di frequenze

- Il numero delle unità che appartengono ad una classe (di equivalenza) è detto **frequenza assoluta** della classe, o **frequenza assoluta delle unità che presentano la modalità x**, o più brevemente **frequenza assoluta della modalità x**.
- Il risultato della successiva esecuzione delle operazioni di suddivisione in classi di equivalenza, di enumerazione e della loro associazione, prende il nome di **distribuzione di frequenze del collettivo secondo il carattere considerato**.

### Sistemazione della distribuzione di frequenze in una tabella semplice

|   |  |
|---|--|
| Titolo: Collettivo indagato <b>secondo il</b><br>Carattere (indicazioni di tempo e luogo) |  |
| Carattere X   | Frequenza assoluta o numero di unità statistiche |
| Modalità ( $x_1$ )  | $n_1$  |
| Modalità ( $x_2$ )  | $n_2$  |
|   |  |
| Modalità ( $x_i$ )  | $n_i$  |
|   |  |
| Modalità ( $x_k$ )  | $n_k$  |
| Totale  | $n$  |

Effettuata l'operazione di suddivisione del collettivo in classi di equivalenza e la loro enumerazione, la distribuzione semplice di frequenze associa la modalità  $x_i$  con la frequenza assoluta  $n_i$ . La somma degli  $n_i$ , per  $i$  che varia da 1 a  $k$ , fornisce  $n$ , la numerosità del collettivo. La prima elaborazione che è possibile sempre fare sulla distribuzione delle frequenze assolute consiste nel passare alla distribuzione delle frequenze relative, quando si rapporta ogni  $n_i$  ad  $n$  ottenendo le quantità  $f_i$ , con la somma delle  $f_i$ , per  $i$  che varia da 1 a  $k$ , uguale a 1. E' anche possibile ottenere la distribuzione delle frequenze percentuali con la trasformazione:  $p_i = f_i \times 100$ , con la somma dei  $p_i$ , per  $i$  che varia da 1 a  $k$ , uguale a 100. La trasformazione delle frequenze da assolute a relative o percentuali è necessaria per effettuare confronti fra distribuzioni riferite allo stesso carattere, ma ottenute da collettivi diversamente numerosi.

Nel caso di distribuzioni di frequenze rispetto ad un carattere quantitativo, le modalità possono essere espresse in classi. Ciò fa perdere informazioni sulla singola unità statistica. In generale la scelta dell'ampiezza della classe è legata alle modalità presentate dalle unità e alla numerosità del collettivo. Nel caso in cui la distribuzione di frequenze dovesse presentare una frequenza uguale a zero è necessario modificare l'ampiezza della classe che ha condotto a tale risultato. In generale si consiglia che la frequenza assoluta associata ad una modalità espressa in classi non sia inferiore a 5, per non perdere troppe informazioni e non "sgranare" la distribuzione e il suo grafico

#### I grafici

• **Il passaggio dalla tabella ai grafici** può essere una operazione artistica, un momento ludico, ma **non è certo un passaggio facile**, anzi con un grafico si può involontariamente, o volontariamente, trarre in inganno chi lo guarda senza disporre della tabella da cui è stato derivato.

• Gli studenti non possono perciò essere lasciati liberi di inventare.

• Esistono indicazioni metodologiche secondo le quali abbinare i grafici ai caratteri studiati.

Tipologie \_\_\_\_\_ di \_\_\_\_\_ grafici

|   |
|---|
| La rappresentazione grafica è legata al tipo di carattere rispetto al quale si classifica il collettivo |
|---|

| Tipo di carattere     | Tipo di grafico  |
|-----------------------|--|
| Qualitativo sconnesso | A barre verticali (colonne)<br>A nastri orizzontali<br>Ad aree suddivise |
| Qualitativo ordinato  | A colonne (con orientamento dell'asse di appoggio dei rettangoli)        |
| Quantitativo discreto | Diagramma ad aste  |
| Quantitativo continuo | Istogramma di frequenze (classi di uguale o diversa ampiezza)            |

Si osserva che per un qualsiasi carattere sono possibili i tipi di grafico ad esso abbinati sulla stessa riga, ed anche quelli delle righe precedenti. Ne risulta che il carattere sconnesso non consente altri tipi di grafico oltre a quelli indicati nella sua riga.

*Grafico a colonne, a nastri orizzontali, e diagramma a settori circolari*

- Quando si deve rappresentare graficamente una tabella che si riferisce ad un carattere qualitativo (come ad esempio il genere) è bene far osservare agli studenti che gli unici elementi numerici rappresentabili nella coppia (modalità, frequenza) sono le frequenze, assolute o relative o percentuali.

- Quanto alla scelta del grafico, si può utilizzare un **grafico a colonne o a nastri orizzontali**, sia per le frequenze assolute che per quelle relative o percentuali. Il principio guida in questo caso è di usare un supporto, rispettivamente orizzontale o verticale, su cui collocare le modalità rappresentate da segmenti tutti uguali ed equidistanti, mentre le frequenze sono rappresentate dall'altezza delle colonne o dalla lunghezza dei nastri in base alla scala di misura adottata.

- La rappresentazione mediante **diagramma a settori circolari** è invece una rappresentazione areale, in cui l'area del cerchio rappresenta la numerosità totale che viene ripartita in parti proporzionali alle frequenze delle singole modalità. Il grafico ha il vantaggio di mostrare la composizione del fenomeno, ma ciò è utile in pratica solo quando le modalità sono poche.

*Grafico ad aste*

- Quando si deve rappresentare una distribuzione rispetto ad un carattere quantitativo discreto, le cui modalità sono numeri interi perché derivano da conteggi, si deve tener conto che entrambi gli elementi della coppia (modalità, frequenza) sono numeri, sicché per ogni modalità è possibile individuare un punto nel piano cartesiano ortogonale, riportando le modalità del carattere sull'asse orizzontale e le frequenze sull'asse verticale; usando scale di misura che possono anche essere diverse sui due assi.

- Per rendere la distribuzione più evidente, si usa tracciare la perpendicolare dal punto (modalità, frequenza) all'asse orizzontale. Il grafico ottenuto ha un nome che ricorda il modo con cui è stato costruito: **diagramma ad aste**.

*Istogramma di frequenze*

Intervalli di uguale ampiezza

Se si vuole rappresentare una distribuzione rispetto ad un carattere le cui modalità sono classi: carattere continuo aventi per modalità degli intervalli (ad esempio: peso, statura, tempo,...) o discreti con modalità raggruppate in classi (ad esempio: numero di addetti delle imprese, numero di residenti dei comuni, reddito in €, ecc...) e le classi o gli intervalli sono tutti di

uguale ampiezza, la rappresentazione grafica viene effettuata mediante l'istogramma di frequenze a basi uguali. Il grafico utilizza un sistema di riferimento cartesiano ortogonale: sull'asse delle ascisse, si riportano i valori degli estremi degli intervalli mentre, su quello delle ordinate, vanno riportate le frequenze assolute negli intervalli stessi.

L'**istogramma di frequenze a basi uguali**, è una figura formata da rettangoli contigui, le cui basi sono gli intervalli tutti uguali e le altezze le frequenze di ogni intervallo **(questo, si ribadisce, se le classi sono di uguale ampiezza)**.

Dunque le aree dei singoli rettangoli sono proporzionali alle frequenze e l'area formata dalla somma delle aree dei rettangoli è proporzionale alla numerosità del collettivo.

#### Intervalli di diversa ampiezza

Nel caso in cui le classi fossero di ampiezza variabile, tenuto conto che le aree dei rettangoli sono proporzionali alle frequenze, per costruire correttamente l'**istogramma di frequenze** sull'asse delle ascisse si riportano gli estremi degli intervalli (di ampiezza diversa) e sull'asse delle ordinate si riporta la densità di frequenza ottenuta dal rapporto tra la frequenza e l'ampiezza dell'intervallo ad essa associata.