

Introduzione alla statistica



Partite di calcio

Il calcio è lo sport più seguito e amato in Italia. Le partite del campionato appassionano migliaia di tifosi, spesso disposti ad andare anche in trasferta pur di seguire la propria squadra del cuore...

...è vero che è più facile vincere una partita in casa che in trasferta?

➔ La risposta a pag. $\alpha 20$

1. I dati statistici

■ La statistica induttiva e la statistica descrittiva

Immagina di parlare con uno sconosciuto e di raccogliere informazioni sulle sue abitudini, i suoi gusti, il suo stato di salute. Potresti dedurre un ritratto significativo di questa persona.

Se raccogliessi le stesse informazioni per molte persone, diciamo mille, potresti fare, in qualche modo, un ritratto del gruppo?

Può essere vantaggioso *raggruppare e sintetizzare i dati*: in questo modo si rinuncia a parte dell'informazione che essi contengono, ma si guadagna in leggibilità e facilità di interpretazione. In particolare si possono elaborare tanti dati relativi a individui singoli per trarne informazioni sulla popolazione nel suo complesso. A seconda poi di come questi dati vengono raggruppati è possibile studiare aspetti diversi del problema in esame.

La statistica si occupa proprio dei modi di raccogliere e analizzare **dati** relativi a un certo gruppo di persone (gli studenti di una scuola, gli abitanti di un quartiere, gli elettori di una regione ecc.) o di oggetti (le automobili, i dischi, i libri ecc.), per trarne conclusioni e fare previsioni.

► La **statistica** ha questo nome perché all'inizio essa studiava principalmente i dati utili al governo degli Stati.

► A volte anche molte informazioni possono essere inutili, se non sono ben organizzate.

► Il campione deve essere attendibile: per esempio, se si sta eseguendo un'indagine per verificare o meno il successo di una trasmissione televisiva, l'intervistato non deve essere qualcuno che lavora per quella trasmissione.

Le fasi fondamentali di un'indagine statistica sono quindi:

- il rilevamento dei dati;
- l'elaborazione dei dati.

Il gruppo preso in considerazione viene anche detto **popolazione** o **universo**. Se la rilevazione dei dati viene effettuata su tutta la popolazione, si definisce **censimento**. Spesso viene presa in esame soltanto una parte della popolazione, detta **campione**, scelta in modo che rappresenti l'intero gruppo. La raccolta di tipo globale è più significativa di quella campionaria, ma può essere molto lunga e costosa nel caso di popolazioni numerose. Per questo la maggior parte delle raccolte dati è di tipo campionario. Le tecniche utilizzate per la raccolta dei dati possono essere l'intervista diretta o indiretta. Nel caso di intervista indiretta, si possono ottenere le informazioni volute facendo compilare un questionario che viene poi spedito o consegnato a un incaricato dall'intervistato (pensa, per esempio, al censimento).

Si propongono di solito questionari anonimi con la sola richiesta dell'indicazione del sesso e dell'età.

Una volta raccolti i questionari compilati,

- li si conta per sapere il numero effettivo delle unità che costituiscono il campione;
- si contano le diverse risposte date a ciascuna domanda predisponendo tabelle di spoglio;
- si rappresentano graficamente i dati;
- si elaborano i dati con i metodi matematici più opportuni;
- si interpretano i dati e si traggono conclusioni che possano essere valide per tutta la popolazione.

I metodi per ottenere risultati soddisfacenti nel delicato procedimento di passaggio dal campione alla popolazione sono studiati da quella parte della statistica detta **statistica induttiva** (o inferenza statistica).

In questo capitolo ci limiteremo a studiare alcuni degli strumenti matematici utilizzati per descrivere i dati relativi a un certo gruppo scelto come popolazione. In questo caso si parla di **statistica descrittiva**.

■ I caratteri qualitativi e i caratteri quantitativi

Gli elementi di una popolazione si chiamano anche **unità statistiche**. È possibile studiare diverse caratteristiche di tali unità, e ogni caratteristica rappresenta un **carattere** della popolazione.

Ogni carattere viene descritto mediante le **modalità** con cui esso si può manifestare.

I caratteri possono essere di due tipi:

- **qualitativi**, se le loro modalità sono descritte da attributi;
- **quantitativi**, se le loro modalità sono descritte da numeri.

ESEMPIO

1. Il carattere «sesso» ha due modalità: «maschile» e «femminile». Si tratta di un carattere qualitativo.
2. Il carattere «mezzo di trasporto» ha più modalità: «treno», «autobus», «motorino», ... Si tratta di un carattere qualitativo.
3. Il carattere «età» ha più modalità: 14, 15, 16, ... (se espresso in anni). Si tratta di un carattere quantitativo.

DAI CENSIMENTI AI SONDAGGI D'OPINIONE

L'utilizzo di dati statistici per ottenere informazioni utili per il governo degli Stati, quali il numero di abitanti, di soldati, di addetti ai vari mestieri ecc., risale ai popoli antichi, in particolare ai Cinesi e agli Egizi.

Nella Bibbia sono descritti diversi censimenti fra gli Ebrei, tra i quali il più noto è quello di Mosè nel deserto del Sinai. Anche i Romani fecero diversi censimenti, uno di questi è quello durante il quale nacque Gesù.

Un passo avanti nell'elaborazione statistica si ebbe in Inghilterra, intorno alla metà del Seicento, con l'«aritmetica politica», principalmente a opera del matematico John Graunt (1620-1674).

A causa delle pestilenze a Londra venivano pubblicate settimanalmente le liste delle morti e quelle delle nascite. Graunt utilizzò quel materiale osser-

vando, attraverso il calcolo di percentuali, regolarità come il maggior numero di nascite maschili rispetto a quelle femminili, il legame fra suicidi e professioni, la diminuzione delle nascite nei periodi di carestia. Era la prima volta che venivano cercate *relazioni* fra i dati raccolti.

Un altro momento importante nella storia della statistica si ebbe quando, nell'Ottocento, i matematici trovarono un collegamento con la probabilità.

Infine, è dell'ultimo secolo lo sviluppo sempre più ampio della statistica come scienza matematica a sé stante. L'applicazione di tale scienza, mediante indagini a campione, investe i campi più diversi, dai fenomeni sociali a quelli meteorologici. L'Istituto nazionale di Statistica (ISTAT) è dal 1926 l'organo ufficiale italiano per le informazioni statistiche (www.istat.it).

Le tabelle di frequenza**DEFINIZIONE**

La **frequenza** di una modalità è il numero di volte in cui si presenta.

ESEMPIO In un questionario abbiamo chiesto ai 28 studenti di una classe di indicare con le seguenti lettere i mezzi di trasporto con cui vanno di solito a scuola:

A: automobile; M: motorino o scooter;
P: a piedi; C: bicicletta.
B: autobus o pullman;

Abbiamo ottenuto i seguenti risultati:

A, B, M, M, P, A, A, B, P, B, C, A, B, B, B, C, P, B, A, C, C, A, M, B, M,
B, A, C.

Contiamo quante volte si presenta ciascuna modalità, ovvero la sua frequenza. Costruiamo la seguente tabella di frequenza.