

La scelta del campione

Il primo problema che si pone quando si deve selezionare un campione è come scegliere gli individui in modo che essi siano rappresentativi della collettività, e quanto grande debba essere tale campione per far sì che i risultati che si ricavano dal suo studio siano sufficientemente precisi da essere adattabili all'intera popolazione. La più semplice forma di scelta è il **campionamento casuale**, in cui ogni possibile campione di date dimensioni estratto dalla popolazione ha la stessa probabilità di essere scelto. Per selezionare un campione casuale si potrebbe associare ad ogni elemento della popolazione un numero e , mediante un processo di scelta casuale, estrarne alcuni. Ad esempio, se si deve selezionare un campione di ampiezza 1000 fra gli elettori che votano in una certa città, ci si può riferire alle liste elettorali, in cui ogni cittadino elettore ha un numero progressivo di lista, e fare la selezione in base a dei numeri casuali generati da un computer. Se il computer avesse generato la lista di numeri che segue (li abbiamo raggruppati in gruppi di 25 numeri ciascuno), si potrebbe pensare di suddividere le liste elettorali in gruppi di 100 e scegliere 25 elementi per gruppo secondo la numerazione di ogni blocco.

In questo modo i primi 25 elementi del nostro campione sarebbero gli elettori contrassegnati con i numeri 3, 47, 43, 73, 86, ecc. per il primo gruppo, i successivi 25 quelli individuati dai numeri 36, 96, 47, ecc. per il secondo gruppo e così via.

03 47 43 73 86	36 96 47 36 61	46 98 63 71 62	33 26 16 80 45	60 11 14 10 95
97 74 24 67 62	42 81 14 57 20	42 53 32 37 32	27 07 36 07 51	24 51 79 89 73
16 76 62 27 66	56 50 26 71 08	32 90 79 78 53	13 55 38 58 59	88 97 54 14 10
12 56 85 99 26	96 96 68 27 31	05 03 72 93 15	57 12 10 14 21	88 26 49 81 76
55 59 56 35 64	38 54 82 46 22	31 62 43 09 90	06 18 44 32 53	23 83 01 30 30

Un'altra forma di campionamento è detta di **campionamento sistematico**. Questa modalità si usa soprattutto quando si ha a che fare con popolazioni ordinate di elementi. Essa consiste nello scegliere a caso il primo elemento del campione dalla lista ordinata e nell'estrarre poi gli altri a distanze regolari dal primo. Riprendendo l'esempio precedente delle liste elettorali, per campionare nel rapporto di 1 a 100 (scegliendo cioè un elemento ogni 100 della lista), si selezionerebbe il primo estratto mediante la generazione di un numero casuale, ad esempio 43; gli altri estratti cadrebbero a distanze regolari di 100 unità dalla scelta del primo, sarebbero cioè gli elettori corrispondenti ai numeri 143, 243, 343, ecc. della lista.

Nelle indagini statistiche su vasta scala, spesso la popolazione può ritenersi naturalmente suddivisa in gruppi; è il caso, ad esempio, di una indagine riguardante il reddito pro capite, in cui la popolazione può ritenersi suddivisa in modo naturale nelle regioni di residenza; oppure quando, volendo analizzare l'effetto di un certo vaccino antinfluenzale, si suddivide naturalmente la popolazione in fasce di età. Si parla in questi casi di **campionamento stratificato** in cui ogni gruppo costituisce uno strato. Da ogni strato si estrae poi un campione parziale che, insieme agli altri, concorre a formare il campione da analizzare. È importante per questo tipo di campionamento che la variabilità fra gli strati sia molto alta, cioè che i diversi gruppi siano del tutto eterogenei fra loro rispetto al carattere su cui si indaga. Se si vuole valutare l'efficacia di un vaccino verificando l'ipotesi che essa vari al variare dell'età del vaccinato, occorre che la popolazione venga divisa in fasce di età, in modo da poter scegliere un campione con elementi che provengono da ogni fascia; un campionamento casuale, in questo caso, non darebbe garanzie sulla bontà del campione, nel senso che potrebbero essere estratti individui per la maggior parte giovani o per la maggior parte anziani, con il conseguente rischio di falsare i risultati dell'indagine.