DEFINIZIONI

POPOLAZIONE: Insieme degli elementi che costituiscono l'oggetto di una analisi statistica.

Esempio: il numero di incidenti nel 1993 di un patentato italiano.
Esempio: La durata di vita di un certo componente delle vetture prodotte da una casa automobilistica.

N.B.: In ambedue i casi precedenti la popolazione non è una popolazione fisica ma una popolazione numerica: numero incidenti subiti da un patentato nel 1993 (e non i patentati), durate di vita dei componente automobilistico (e non i componenti)

CAMPIONE: Un sottoinsieme della popolazione (opportunamente scelto) considerato rappresentativo della popolazione.
Il campione è rappresentativo se si può considerare che esso riproduca le stesse caratteristiche della popolazione e la loro distribuzione.
Il campione può anche coincidere con la popolazione (censimento) ma è più in generale un sottoinsieme proprio della popolazione.

STATISTICA INFERENZIALE: Insieme delle tecniche che permette di indagare la distribuzione nel campione delle caratteristiche analizzate e l'estensione dei dati osservati all' intera popolazione.

LA SCELTA DEL CAMPIONE

Perchè un campione sia rappresentativo di una popolazione occorre che sia un campione non distorto.
Campione distorto: Ottenuto con una scelta di campionamento che induce deformazioni sistematiche nella distribuzione nel campione delle caratteristiche da analizzare.

Esempio di campione distorto:
Scopo dell'indagine: incidentalità dei patentati italiani nel 1993.
Campione: incidenti d'auto per i degenti dei reparti traumatizzati degli ospedali italiani.

Semplici sistemi di campionamento sono:
Il campionamento percentuale ed
Il campionamento sistematico.

Campionamento percentuale: si determina opportunamente la percentuale di elementi che si vogliono estrarre dalla popolazione e gli elementi vengono scelti in base a ragionamenti empirici.

Esempio: dai dati forniti dalle compagnie di assicurazione si determinano i comuni in cui il tasso di incidentalità risulta più vicino alla media nazionale (calcolata sugli stessi dati forniti dalle assicurazioni) e si intervistano (allo scopo di sapere il numero di incidenti subiti nel 1993) i patentati di questi comuni fino a raggiungere un numero di interviste pari alla percentuale della popolazione prefissata.

Campionamento sistematico: consiste nello scandire la popolazione includendo nel campione un elemento ogni numero prefissato di elementi considerati.
Esempio: si scorrono le liste dei patentati depositate presso gli uffici della motorizzazione e si introduce nel campione il numero di incidenti occorsi nel 1993 ad un patentato ogni dieci (ogni k), la frazione di elementi della popolazione inserita nel campione è di 1/10 (l/k) della popolazione totale.

Per una più rigorosa operazione di inferenza statistica, è opportuno che siano individuabili i legami probabilistici tra popolazione e campione, quali, ad esempio, la probabilità che un elemento della popolazione entri a fare parte del campione.
Per tale motivo si ricorre generalmente a campionamenti di tipo probabilistico, dando così luogo a:

Campioni Casuali !

Campione casuale: Ogni elemento della popolazione ha una probabilità nota e non nulla di entrare a fare parte del campione.

Campione casuale semplice: Il campionamento è equiprobabi1e, il che non significa necessariamente che ogni elemento della popolazione ha la stessa probabilità di entrare a far parte del campione stesso.

Esempio: Ad ogni patentato viene associato un numero corrispondente alla sua posizione negli elenchi della motorizzazione civile, il campionamento viene eseguito estraendo da una distribuzione uniforme (equiprobabile) tanti numeri quanti sono gli elementi che devono costituire il campione ed includendo nel campione il numero di incidenti subiti nel 1993 dal patentato corrispondente al numero estratto.
Si noti che ogni patentato ha la stessa probabilità di essere estratto per andare a comporre le informazioni dei campione.

Campione casuale stratificato: La popolazione viene ripartita in un numero predeterminato di classi di numerosità nota rispetto alla numerosità dei campione ed all'interno di ognuna di queste classi si effettua un campionamento casuale.

Esempio: I patentati vengono divisi in classi di età, una prima dai 18 ai 20 anni, le altre con scansione decennale fino ai 60 anni e quinquennale oltre questa età. Ali' interno di ogni classe così definita si effettuano delle estrazioni da una distribuzione uniforme e si include nel campione il numero di incidenti subiti nel 1993 dal patentato estratto. Si ripete I' operazione per tutte le classi. In generale il numero dì estrazioni effettuate per ogni classe è proporzionale alla numerosità della classe rispetto alla numerosità dei campione.

TECNICHE Di ESTRAZIONE PER CAMPIONI CASUALI
Estrazione del campione **con reintroduzione**: Ogni elemento estratto dalla popolazione per andare a formare il campione viene '"reintrodotto" nella popolazione, può essere dunque riestratto una seconda (o una terza ...) volta e la possibilità di tali successive estrazioni è la stessa della prima. Con questa tecnica la composizione della popolazione non viene alterata dal processo di campionamento.
Estrazione dei campione **senza reintroduzione**: Ogni elemento può essere estratto una sola volta dalla popolazione, se estratto viene eliminato dalla popolazione la cui composizione risulta dunque alterata. Nel corso di tale procedura gli elementi residui della popolazione hanno una probabilità non nulla di essere estratti mentre gli elementi già estratti hanno probabilità di estrazione nulla. La probabilità di estrazione varia al procedere dei campionamento ed è influenzata delle estrazioni precedenti.

N.B.: Il risultato di un campionamento casuale è una realizzazione di una variabile aleatoria multidimensionale

1 . Il risultato di una estrazione di un elemento del campione da una popolazione non può essere predetto a priori, è dunque una v.a.

2. La probabilità che un elemento dei campione assuma un certo valore dipende dalla distribuzione degli elementi della popolazione e dalla tecnica di campionamento utilizzata (con reintroduzione o senza).

3. L'insieme dei valori dei campione è dunque il risultato di successive estrazioni dalla popolazione. Se e solo se le estrazioni avvengono con reintroduzione, esse sono indipendenti tra loro.

4. Il campione ottenuto è la realizzazione di una v.a. multidimensionale. Se la tecnica di estrazione è con reintroduzione gli elementi dei vettore aleatorio sono identicamente ed indipendentemente distribuiti, con distribuzione identica a quella della popolazione.

5. La ripetizione dello stesso processo di campionemento sulla stessa popolazione ha come risultato una diversa realizzazione della v.a. campione.

LE STATISTICHE CAMPIONARIE

(campione casuale con reintroduzione)

Ogni funzione di un campione casuale è essa stessa una v.a, detta statistica campionaría. Si definiscono in particolare i due eguenti momenti campionari:

Media aritmetica campionaria:

$$\bar{x} = \frac{1}{n} \cdot \sum_i x_i \qquad \bar{x} = \sum_k \frac{N_k}{N} \cdot \frac{1}{n_k} \cdot \sum_i x_{i,k}$$

N.B.: La media campionaria è una v.a., la media della popolazione è un parametro.

Varianza campionaria:

$$s^2 = \sum_k \frac{N_k}{N} \cdot \frac{1}{n_k - 1} \cdot \sum_i \left(x_{i,k} - \bar{x}_k\right)^2 \qquad s^2 = \frac{1}{n-1} \cdot \sum_i (x_i - \bar{x})^2$$

Si ha:

$$E[\bar{x}] = E\left[\sum_k \frac{N_k}{N} \cdot \frac{1}{n_k} \sum_i x_{i,k}\right] = \frac{1}{N} \cdot \sum_k \frac{N_k}{n_k} \cdot \frac{1}{n_k} \sum_i \mu = \mu$$

Analogamente si può dimostrare che risulta (per un campione casuale semplice):

$$E\left[s^2\right] = \sigma^2$$

	CAMP. CASUALE SEMPLICE	CAMP. CASUALE STRATIFICATO
$\bar{x}$	$\frac{1}{n} \cdot \sum_i x_i$	$\sum_k \frac{N_k}{N} \cdot \frac{1}{n_k} \cdot \sum_i x_{i,k}$
s^2	$\frac{1}{n-1} \cdot \sum_i (x_i - \bar{x})^2$	$\sum_k \frac{N_k}{N} \cdot \frac{1}{n_k - 1} \cdot \sum_i (x_{i,k} - \bar{x}_k)^2$
$E[\bar{x}]$	μ	μ
$Var[\bar{x}]$	$\frac{\sigma^2}{n}$	$\frac{\sigma^2}{N^2} \cdot \sum_k \frac{N_k^2}{n_k}$
$E[s^2]$	σ^2	σ^2

CAMPIONAMENTO CASUALE SENZA REIMISSIONE

In questo caso però non è più possibile scrivere:

$$f_Y(x_1, x_2, \ldots, x_n) = f_X(x_1) f_X(x_2) \ldots f_X(x_2) \qquad \text{con } f_X(x_i) \approx f_X(x)$$

giacchè le variabili aleatorie costituenti il campione non sono più indipendenti.

E' possibile però dimostrare che:

$$E[\bar{x}] = E\left[\frac{1}{n} \cdot \sum_i x_i\right] = \mu \qquad Var[\bar{x}] = Var\left[\frac{1}{n} \cdot \sum_i x_i\right] = \frac{N-n}{N-1} \cdot \frac{\sigma^2}{n}$$

$$E\left[s^2\right] = E\left[\frac{1}{n-1} \cdot \sum_i (x_i - \bar{x})^2\right] = \frac{N}{N-1} \cdot \sigma^2$$

L'utilizzazione di un campionamento di tipo casuale permette dunque di porre i parametri della popolazione in relazione con alcuni parametri del campione.