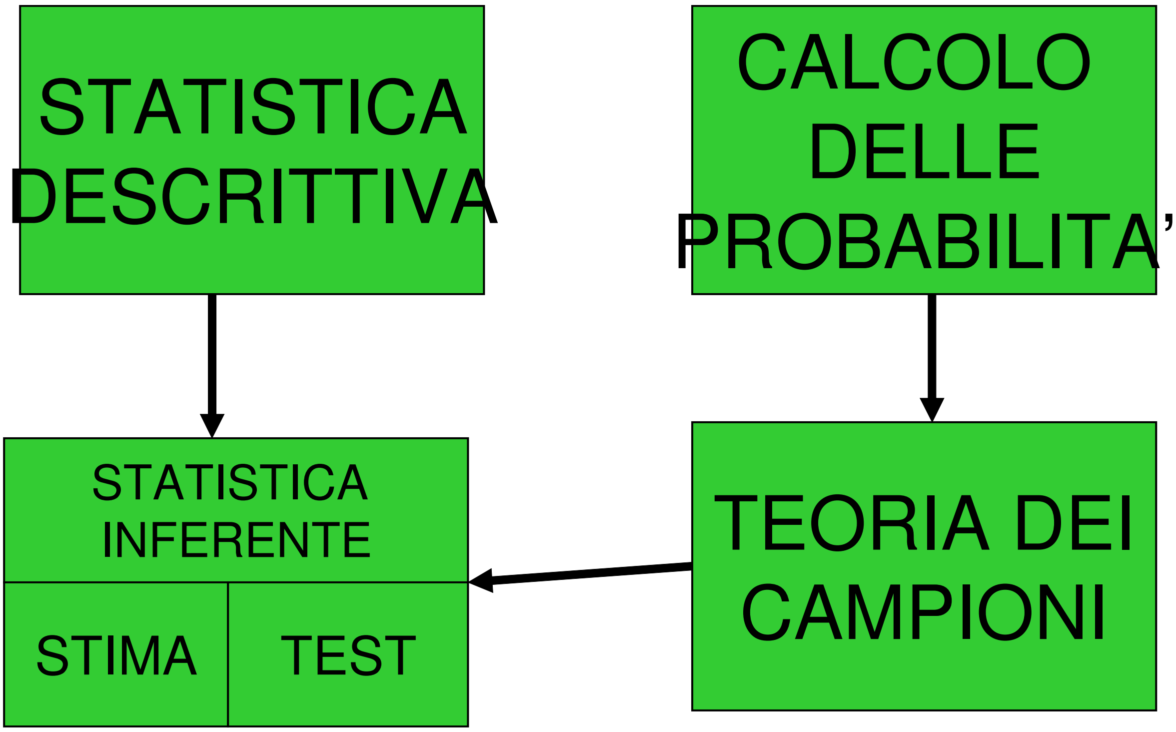
INFERENZA o STATISTICA INFERENTE

Le informazioni sui parametri della *popolazione* si possono ottenere sia mediante una rilevazione totale (o rilevazione censuaria) sia mediante una rilevazione parziale (o rilevazione campionaria).

Rilevazione totale: STATISTICA DESCRITTIVA
Rilevazione campionaria: STATISTICA INFERENTE

L'INFERENZA è l'insieme dei metodi e delle tecniche statistiche con cui si riescono ad ottenere informazioni sui parametri che caratterizzano la *popolazione* d'origine, utilizzando i dati di un campione casuale.



I metodo inferenziali consistono in procedimenti di tipo induttivo, che in quanto tali portano a commettere degli errori: errore di campionamento ed errore di stima.

(Teoria dei campioni) CAMPIONAMENTO

Campioni a scelta ragionata o Campioni non probabilistici
↕
Campioni probabilistici:
- campioni stratificati
- campioni a grappoli
- campioni puramente casuali o **campioni casuali**

ecc…

Se ogni unità appartenente alla popolazione ha la stessa probabilità delle altre di presentarsi nel campione, allora si parla di **campionamento casuale semplice** e si ottiene un **campione casuale**. Tale campionamento rispecchia la situazione dell'estrazione di palline da un'urna: si supponga di estrarre ***n*** palline da un'urna contenente ***N*** palline uguali, ovvero perfettamente calibrate e senza caratteri distintivi, numerate da 1 ad N. Allora ogni pallina ha la stessa probabilità delle altre di essere estratta (*Tavole dei numeri casuali*).

Se l'estrazione delle ***n*** palline dall'urna viene effettuata CON reinserimento allora si ha un **campione casuale bernoulliano** o **campione casuale con ripetizione**.
Se l'estrazione delle ***n*** palline dall'urna viene effettuata SENZA reinserimento allora si ha un **campione casuale senza ripetizione** o **c.c. in blocco** o **c.c. esaustivo**.

Data una popolazione di ***N*** elementi, si consideri lo <u>*Spazio campionario*</u> dato da tutti i possibili campioni casuali di n elementi che si possono formare estraendoli da quella popolazione. Allora

- se il campione casuale è bernoulliano, lo <u>*Spazio campionario*</u> è costituito da $\boldsymbol{N^n}$ campioni di dimensione n, dove

$$N^n = D^r_{N,n}$$

- se il campione casuale è senza ripetizione, lo <u>*Spazio campionario*</u> è costituito da ***N(N-1)(N-2)…(N-n+1)*** campioni di dimensione n, dove

$$N(N-1)(N-2)\cdots(N-n+1) = D^s_{N,n}$$

In particolare, se il campione casuale è senza ripetizione e l'ordine di estrazione degli elementi non conta, allora lo <u>*Spazio campionario*</u> è costituito da $\binom{N}{n}$ campioni di dimensione n, dove

$$\binom{N}{n} = C^s_{N,n}$$

Si supponga di voler studiare un certo carattere *X* della popolazione, di dimensione *N*. Si può ipotizzare che il carattere *X* si comporti come una v.c. *X* con una certa funzione di probabilità o funzione di densità di probabilità: si indichi con

$$p_X(x,\theta)$$

il *modello descrittivo* della v.c. *X* oggetto di studio, dove θ indica il *parametro* di definizione della v.c. *X*. In particolare siano $\boldsymbol{\mu}$=E(X) e $\boldsymbol{\sigma}^2$=Var(X).

Si supponga che della v.c. *X* interessi conoscere uno o più parametri. Allora dalla popolazione si estrae, secondo un certo criterio, un campione casuale di *n* elementi:

$$X_1, X_2, \ldots, X_n$$

L'insieme delle *n* v.c. $X_1, X_2, \ldots, X_n$ costituisce un **campione casuale** di *X*. Ogni v.c. X_i del campione casuale si distribuisce come la v.c. *X*, con $E(X_i)=\boldsymbol{\mu}$ e $Var(X_i)=\boldsymbol{\sigma}^2$ (per ogni i, con i=1,2,…,n).

Si definisce ***campione casuale*** *di ampiezza n la n-upla di v.c.* $X_1, X_2, \ldots, X_n$ *identicamente distribuite come la v.c. X oggetto di studio.*

Una generica realizzazione del campione casuale $\{X_1, X_2, \ldots, X_n\}$ è costituita da n numeri: $(x_1, x_2, \ldots, x_n)$. L'insieme degli *n* numeri $x_1, x_2, \ldots, x_n$ costituisce un **campione casuale osservabile**.

I procedimenti dell'Inferenza statistica richiedono che i dati del campione vengano elaborati. Sia $(x_1, x_2, \ldots, x_n)$ il campione casuale osservato e sia $t(\bullet)$ una funzione matematica a n variabili. Applicando $t(\bullet)$ al campione casuale osservato si ottiene un valore:

$$t(\bullet) = t(x_1, x_2, \ldots, x_n)$$

Applicando $t(\bullet)$ ad un campione casuale di *X* si ottiene una variabile casuale che, in quanto tale, ha una propria Distribuzione di probabilità:

$$t(\bullet) = t(X_1, X_2, \ldots, X_n)$$

Si definisce ***STATISTICA CAMPIONARIA*** o ***V.C. CAMPIONARIA*** una qualsiasi funzione delle v.c. $X_1, X_2, \ldots, X_n$ che compongono il campione casuale.

Si definisce ***DISTRIBUZIONE CAMPIONARIA*** di una STATISTICA CAMPIONARIA la sua distribuzione di probabilità. La *DISTRIBUZIONE CAMPIONARIA* dipende dal modello descrittivo della popolazione.

STATISTICA CAMPIONARIA

MEDIA CAMPIONARIA $\overline{X}$

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

Media aritmetica o Valore atteso della MEDIA CAMPIONARIA:

$E(\overline{X}) = \mu$ (dove **μ**=E(X) è la media della popolazione)

Varianza della MEDIA CAMPIONARIA nel caso di campione casuale bernoulliano:

$$VAR(\overline{X}) = \frac{\sigma^2}{n}$$

Deviazione standard della MEDIA CAMPIONARIA nel caso di campione casuale bernoulliano:

$$\sigma_{\overline{X}} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$$

Varianza della MEDIA CAMPIONARIA nel caso di campione casuale senza ripetizione e di dimensione della popolazione finita:

$$VAR(\overline{X}) = \frac{\sigma^2}{n} \cdot \frac{N-n}{N-1}$$

Deviazione standard della MEDIA CAMPIONARIA nel caso di campione casuale senza ripetizione e di dimensione della popolazione finita:

$$\sigma_{\overline{X}} = \sqrt{\frac{\sigma^2}{n} \cdot \frac{N-n}{N-1}} = \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}$$

La **DISTRIBUZIONE CAMPIONARIA** di $\overline{X}$ dipende dal modello descrittivo della popolazione, ovvero dalla distribuzione di probabilità della v.c. *X* che descrive la popolazione.

I caso: se la v.c. *X* che descrive la popolazione è distribuita normalmente, con $\boldsymbol{\mu}$=E(X) e $\boldsymbol{\sigma}^2$=Var(X), allora $\overline{X}$ si distribuisce NORMALMENTE, qualunque sia la dimensione *n* del campione.

II caso: se la dimensione del campione è grande (: n>30), allora la distribuzione campionaria di $\overline{X}$ si avvicina alla distribuzione NORMALE, qualunque sia la distribuzione della popolazione (: conseguenza del ***Teorema del limite centrale***).

III caso: se la v.c. *X* che descrive la popolazione è distribuita normalmente, se la Varianza della popolazione è incognita e se la dimensione del campione è piccola (: n<30), allora la distribuzione campionaria di $\overline{X}$ standardizzata segue la distribuzione della v.c. ***t** di Student*.

STATISTICA CAMPIONARIA

VARIANZA CAMPIONARIA S^2

$$S^2 = \frac{\sum_{i=1}^{n} \left(X_i - \overline{X} \right)^2}{n-1}$$

Media aritmetica o Valore atteso della VARIANZA CAMPIONARIA:

$E(S^2) = \sigma^2$ (dove $\boldsymbol{\sigma}^2$=Var(X) è la varianza della popolazione)

(Varianza della VARIANZA CAMPIONARIA)

La **DISTRIBUZIONE CAMPIONARIA** di S^2 dipende dal modello descrittivo della popolazione, ovvero dalla distribuzione di probabilità della v.c. *X* che descrive la popolazione.

Caso particolare: se la v.c. *X* che descrive la popolazione è distribuita normalmente, con $\boldsymbol{\mu}$=E(X) e $\boldsymbol{\sigma}^2$=Var(X), allora la v.c. S^2 si distribuisce come una v.c. *chi-quadrato* (con $\nu = n-1$), a meno di un coefficiente di proporzionalità pari a

$$\frac{\sigma^2}{\nu}$$

Ovvero:

$$S^2 = \frac{\chi_\nu^2 \sigma^2}{\nu}$$