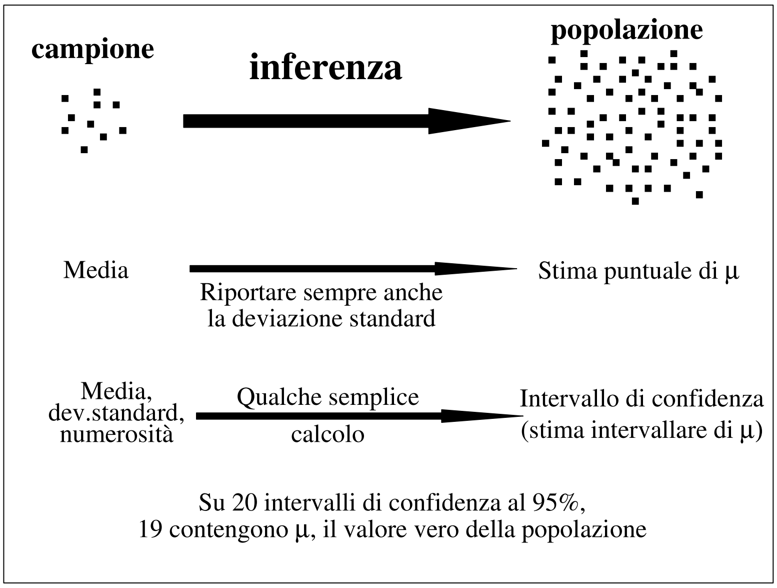
Brevi cenni all'intervallo di confidenza



Dal momento che il campione viene estratto casualmente dalla popolazione, le conclusioni tratte da un campione possono essere errate.

L'inferenza statistica viene fatta "con umiltà":

1) si cerca di stimare la probabilità di commettere errori

2) si cerca di limitare la probabilità di commettere errori

INTERVALLO di CONFIDENZA

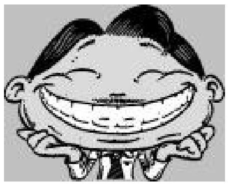
Lo scopo dell'inferenza statistica è la conoscenza dei **parametri** che caratterizzano una popolazione.

Per conoscere il parametro, però, dovremmo prendere in esame **tutte** le unità statistiche che costituiscono la popolazione; questo spesso è impossibile perché:

1. numerosità molto elevata
2. spesso la popolazione obiettivo è infinita

impossibile conoscere il **parametro**

Non potendo calcolare con esattezza il parametro, **ricorriamo ad una sua stima**.



La **statistica** (es $\overline{x}$, s) calcolata su un campione estratto dalla popolazione obiettivo è una **stima puntuale** del parametro della popolazione.

Questa stima puntuale del parametro <u>non sarà mai identica</u> al vero parametro della popolazione, ma sarà affetta da un **errore** per eccesso o per difetto.

In molte situazioni è preferibile **una stima intervallare** (cioè è preferibile indicare come stima del parametro un intervallo al posto di un *singolo punto* sull'asse dei valori) che esprima anche l'**errore associato alla stima** (precisione).

INTERVALLO di CONFIDENZA: DEFINIZIONE

Per intervallo di confidenza di un parametro Θ (ad es. della media μ) della popolazione, intendiamo un intervallo delimitato da due limiti L_{inf} (limite inferiore) ed L_{sup} (limite superiore) che abbia una definita probabilità ($1-\alpha$) (ad es. $(1-0.05)=0.95$) di contenere il vero parametro della popolazione:

$$p(L_{inf} < \Theta < L_{sup}) = 1-\alpha$$

$$p(L_{inf} < \mu < L_{sup}) = 1-0.05 = 0.95$$

dove:

$1-\alpha$ = grado di confidenza

α = probabilità di errore

Esempio

- Vogliamo stimare il livello medio di glicemia nei diabetici italiani:
- prendiamo un campione di 36 soggetti; la media della glicemia in questo gruppo risulta 155 mg/dl *(stima puntuale).*
- Calcoliamo l'intervallo di confidenza al 95% che risulta: 147,2-162,8 mg/dl *(stima intervallare)*

questo intervallo ha una probabilità del 95% di contenere la vera media della popolazione dei diabetici italiani

Esempio - *continua*

- L'intervallo di confidenza al 95% del lucido precedente è stato ottenuto nel seguente modo:

$$\bar{x} = 155 mg/dl$$

$$s = 24 mg/dl$$

$$n = 36$$

$$\bar{x} \pm 1{,}96 * \frac{s}{\sqrt{n}} = 155 \pm 1{,}96 * \frac{24}{\sqrt{36}}$$

L_{inf}=147,2

L_{sup}=162,8

RIASSUMENDO…

La **stima puntuale** fornisce un <u>singolo valore</u>. Tuttavia:

1. questo valore non coincide quasi mai con il valore vero (parametro) della popolazione;
2. campioni diversi forniscono stime puntuali diverse.

La **stima intervallare** fornisce <u>un intervallo</u>:

1. quest'intervallo ha una determinata probabilità (in genere, il 95%) di contenere il valore vero (parametro) della popolazione;
2. Il metodo generale per la costruzione dell'intervallo di confidenza al (1-α) è:

$$\bar{x} \pm z_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

$$\bar{x} \pm z_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

la **probabilità d'errore** α determina il valore del coefficiente z:

1-α	**α/2**	**$z_{\alpha/2}$**
0.90	**0.05**	**1.64**
0.95	**0.025**	**1.96**
0.99	**0.005**	**2.58**

SESM