INTERVALLO di CONFIDENZA

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

Per conoscere il parametro, 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 $\bar{x}$, 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).

Tale stima prende il nome di:

INTERVALLO DI CONFIDENZA:

per IC di un parametro della popolazione θ, intendiamo un intervallo delimitato da $\mathbf{L_i}$ (limite inferiore) e $\mathbf{L_s}$ (limite superiore) che abbia una definita **probabilità ($1 - \alpha$) di contenere il vero parametro della popolazione**:

$$pr(L_i \le \theta \le L_S) = 1 - \alpha$$

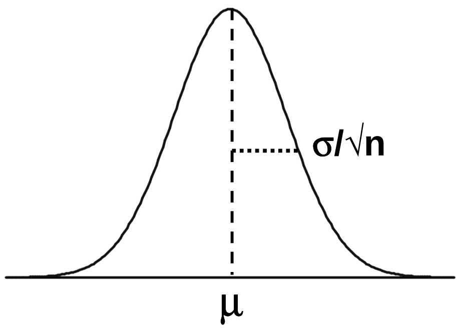
dove: $1 - \alpha$ = **grado di confidenza**

α = **probabilità di errore**

quanto più grande è l'IC tanto più imprecisa è la nostra stima!

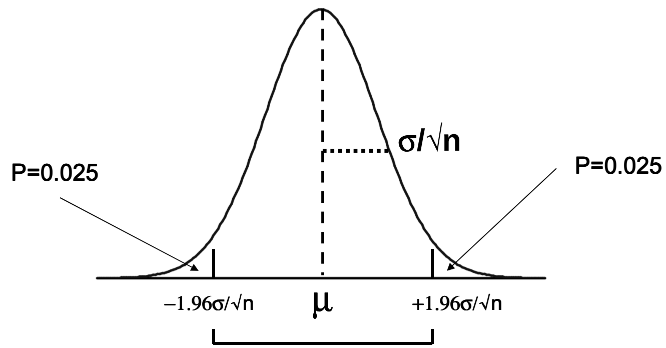
INTERVALLO di CONFIDENZA al 95% di una media

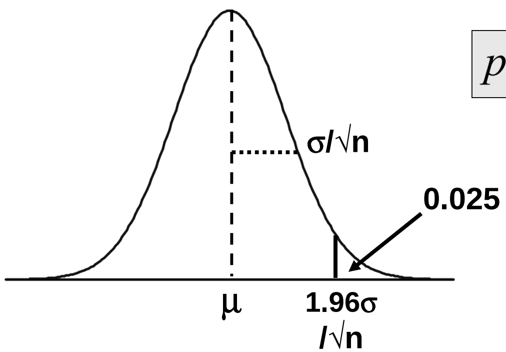
Si assuma che: $\bar{x} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$



Data la distribuzione, l'ntervallo simmetrico che comprende il 95% delle medie campionarie (p=0.95) sarà per definizione:

μ +/- 1.96 e.s.





$$\boxed{pr(L_i \leq \theta \leq L_s) = 1 - \alpha}$$

$$pr\left\{\mu - 1.96\frac{\sigma}{\sqrt{n}} \leq \bar{x} \leq \mu + 1.96\frac{\sigma}{\sqrt{n}}\right\} = 0.95$$

$$pr\left\{\mu - 1.96\frac{\sigma}{\sqrt{n}} \leq \bar{x} \leq \mu + 1.96\frac{\sigma}{\sqrt{n}}\right\} = 0.95$$

e, riarrangiando le due disuguaglianze interne alla parentesi:

$$pr\left\{\boxed{\bar{x} - 1.96\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}}\right\} = 0.95$$

INTERVALLO DI CONFIDENZA

$$pr\left\{\boxed{\bar{x}-1.96\frac{\sigma}{\sqrt{n}}}\le \mu \le \boxed{\bar{x}+1.96\frac{\sigma}{\sqrt{n}}}\right\}=0.95$$

L_s

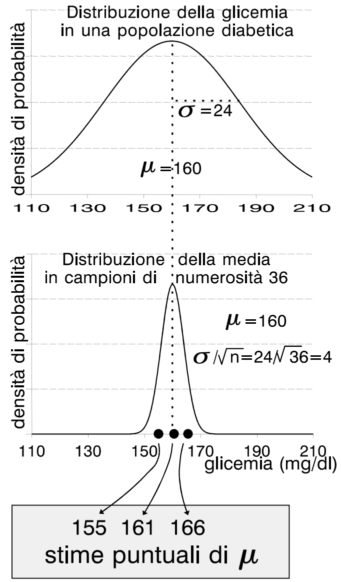
(LIMITE SUPERIORE DELL'INTERVALLO)

L_i

(LIMITE INFERIORE DELL'INTERVALLO)

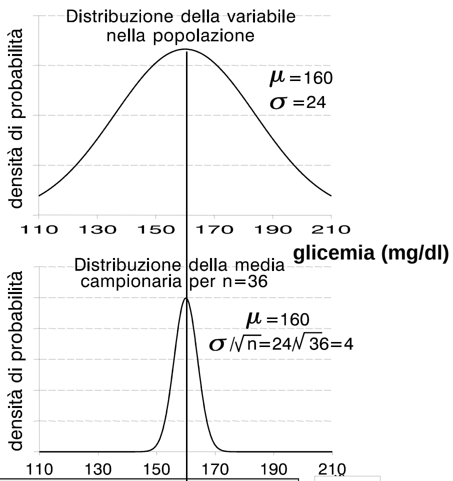
GLICEMIA nella POPOLAZIONE DIABETICA

(stime puntuali)



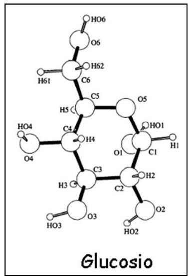
GLICEMIA nella POPOLAZIONE DIABETICA

(stime intervallari)

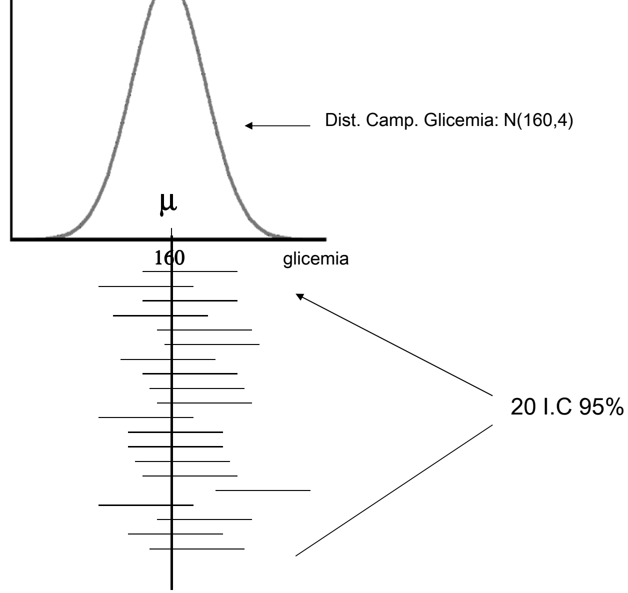


stime intervallari di μ

155±1,96*4	147,2	162,8
161±1,96*4	153,2	168,8
166±1,96*4	158,2	173,8



Errore nella previsione di μ con l'utilizzo dell'intervallo di confidenza al 95%



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 di una media al (1-a) è:

$$\bar{x} \pm z_{\alpha/2} \cdot ES(\bar{x})$$

da cosa dipende l'ampiezza dell'IC?

$$\bar{x} \pm z_{\alpha/2} \cdot ES(\bar{x})$$

1. la **probabilità d'errore α** che determina il valore del coefficiente del limite fiduciale (z):

1-α	**α/2**	**$z_{\alpha/2}$**
0.90	**0.05**	**1.64**
0.95	**0.025**	**1.96**
0.98	**0.01**	**2.33**
0.99	**0.005**	**2.58**

2. la **dimensione del campione (n)**

3. la **variabilità della variabile nella popolazione (σ)**

NOTA BENE!

Nel calcolare l'intervallo di confidenza di una media si è supposto che la deviazione standard della popolazione fosse nota.

Infatti è stata usata
la deviata standardizzata:

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

Molto spesso, però, σ è ignoto. In questo caso, se il campione è sufficientemente grande, **s può essere utilizzata come stima di σ** per il calcolo dell'intervallo di confidenza

INTERVALLO di CONFIDENZA di una PROPORZIONE

Per N > 30: $p \sim N\left(\pi, \sqrt{\frac{\pi(1-\pi)}{n}}\right)$

In analogia con quanto visto per la media, segue che:

π sarà stimato da p

E che:

$$p \pm z_{\alpha/2} \cdot \sqrt{\frac{p(1-p)}{n}}$$