LEZIONI DI STATISTICA SANITARIA

Dott. SIMONE ACCORDINI

Lezione n.11

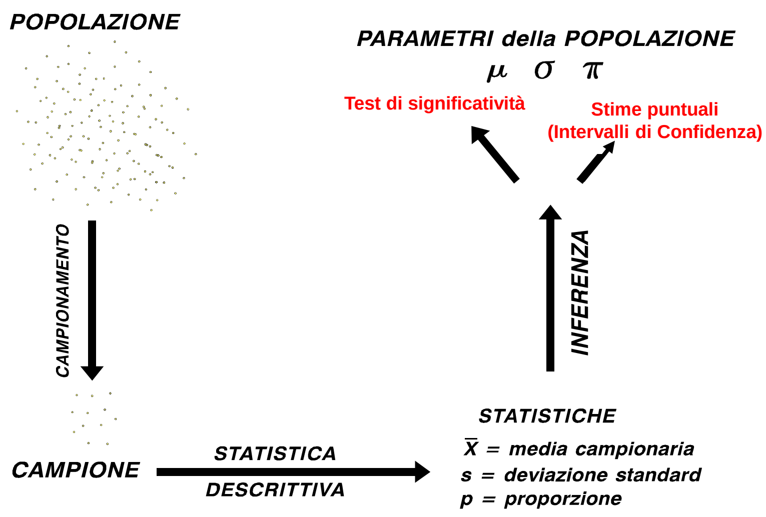
- Principi dell'inferenza statistica

- Test di significatività

- Intervallo di confidenza

Sezione di Epidemiologia & Statistica Medica
Università degli Studi di Verona

INFERENZA



METODI STATISTICI DELL'INFERENZA

1. **Verificare se il parametro di occorrenza varia tra i gruppi ovvero se vi è associazione statistica → le differenze osservate non sono dovute al caso**

   ⇒ TEST DI SIGNIFICATIVITA'

2. **Stimare il parametro di occorrenza separatamente nei gruppi e la misura di effetto ...**

   ⇒ STIMA PUNTUALE

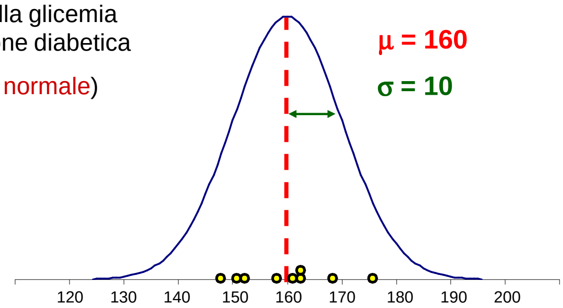
3. **… e associare alla stima puntuale una misura di incertezza**

   ⇒ INTERVALLO DI CONFIDENZA

SESM

Distribuzione della glicemia
in una popolazione diabetica

(modello teorico normale)

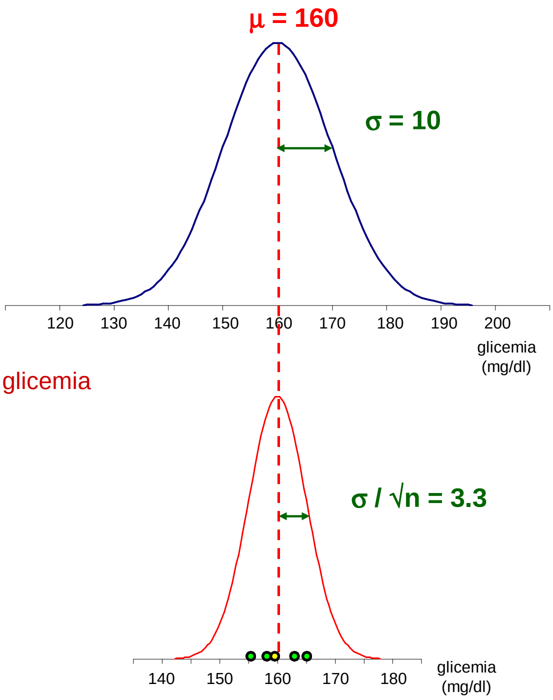


I° campione:	**162 152 158 162 168 161 148 176 150**	→	**$\bar{x}$ = 159.7**	**s = 8.9**
II° campione:	**152 164 157 180 156 163 165 166 178**	→	**$\bar{x}$ = 164.6**	**s = 9.4**
III° campione:	**157 142 163 162 152 149 152 180 151**	→	**$\bar{x}$ = 156.4**	**s = 11.0**
IV° campione:	**162 154 168 160 155 172 162 152 140**	→	**$\bar{x}$ = 158.3**	**s = 9.5**
V° campione:	**163 169 152 147 158 163 173 160 181**	→	**$\bar{x}$ = 162.8**	**s = 10.2**

SESM

Distribuzione della glicemia
in una popolazione diabetica

Distribuzione della media della glicemia
in campioni di numerosità = 9



DISTRIBUZIONE CAMPIONARIA DI UNA MEDIA

Sia $\bar{x}$ la media campionaria di un campione casuale di dimensione *n* selezionato da una popolazione con media μ e deviazione standard σ:

1-2) la distribuzione campionaria di $\bar{X}$ ha:

$E[\bar{X}] = \mu$

$DS[\bar{X}] = ES[\bar{X}] = \sigma/\sqrt{n}$

ERRORE STANDARD della media $\rightarrow$ misura della precisione della stima

3) TEOREMA DEL LIMITE CENTRALE: se la dimensione campionaria è sufficientemente grande ($n \geq 30$), allora la distribuzione campionaria di $\bar{X}$ è *approssimativamente normale*, indipendentemente dalla distribuzione della variabile nella popolazione

Esempio:

Cockburn et al (1980) riportano una ricerca clinica per la prevenzione dell'ipocalcemia infantile, nella quale donne in gravidanza che ricevevano un supplemento di vitamina D venivano messe a confronto con donne non trattate.

Calcemia del bambino misurata 6 giorni dopo la nascita:

	Numero di pazienti	Media (mg /100 ml)	DS (mg /100 ml)
Vitamina D (D_1)	233	9.36	1.15
Controllo (D_0)	394	9.01	1.33

Confronto della media di una variabile quantitativa tra 2 popolazioni - campioni indipendenti

⇩

Test T di Student

GRUPPO NON ESPOSTO (D_0):

$x_{01}, x_{02}, \ldots, x_{0n_0}$ → n_0 **determinazioni indipendenti della v.c.** $X_0 \sim Norm(\mu_0, \sigma_0^2)$

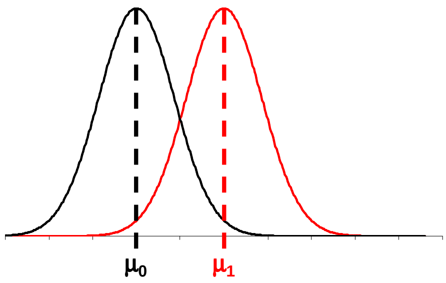
GRUPPO ESPOSTO (D_1):

$x_{11}, x_{12}, \ldots, x_{1n_1}$ → n_1 **determinazioni indipendenti della v.c.** $X_1 \sim Norm(\mu_1, \sigma_1^2)$

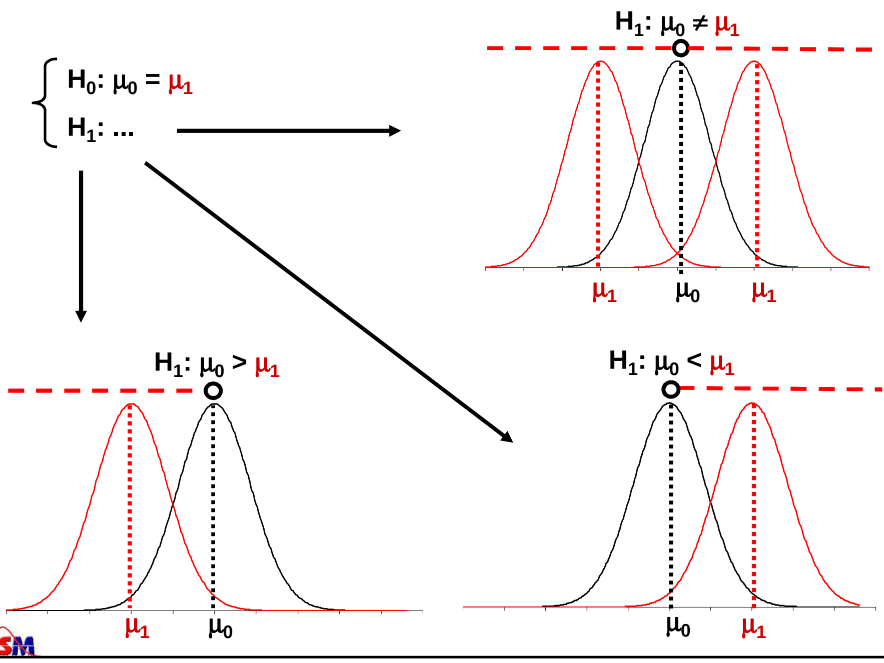
ASSUNZIONI SU CUI SI BASA IL TEST T DI STUDENT:

indipendendenza → L'esito della misurazione in un soggetto non condiziona l'esito della misurazione negli altri soggetti.

normalità



I° STEP: **definire il sistema di ipotesi da verificare**



II° STEP: definire la statistica test

$$\begin{cases} H_0: \mu_0 = \mu_1 & \text{(ipotesi nulla)} \\ H_1: \mu_0 \neq \mu_1 & \text{(ipotesi alternativa)} \end{cases}$$

differenza tra le medie = stima dell'effetto

$$T = \frac{\overline{X}_1 - \overline{X}_0}{ES[\overline{X}_1 - \overline{X}_0]}$$

errore standard della differenza = misura della precisione della stima

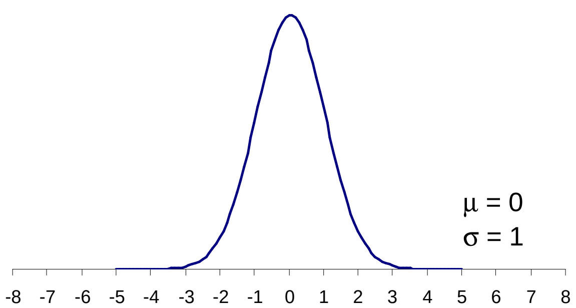
$\sigma_0^2 \neq \sigma_1^2$

$$T = \frac{\overline{X}_1 - \overline{X}_0}{ES[\overline{X}_1 - \overline{X}_0]} = \frac{\overline{X}_1 - \overline{X}_0}{\sqrt{S_1^2/n_1 + S_0^2/n_0}}$$

$\sigma_0^2 = \sigma_1^2 = \sigma^2$

$$T = \frac{\overline{X}_1 - \overline{X}_0}{ES[\overline{X}_1 - \overline{X}_0]} = \frac{\overline{X}_1 - \overline{X}_0}{S\sqrt{1/n_1 + 1/n_0}}$$

con campioni 'grandi' $\Rightarrow$ **distribuzione normale standard**
($n_1 + n_0 - 2 \geq 30$) **(sotto l'ipotesi nulla)**



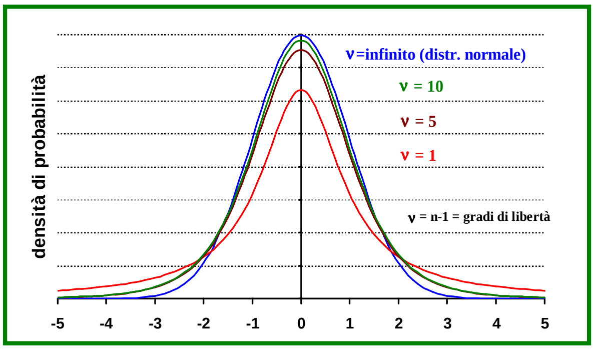
$\sigma_0^2 \neq \sigma_1^2$

$$T = \frac{\bar{X}_1 - \bar{X}_0}{ES[\bar{X}_1 - \bar{X}_0]} = \frac{\bar{X}_1 - \bar{X}_0}{\sqrt{S_1^2/n_1 + S_0^2/n_0}}$$

$\sigma_0^2 = \sigma_1^2 = \sigma^2$

$$T = \frac{\bar{X}_1 - \bar{X}_0}{ES[\bar{X}_1 - \bar{X}_0]} = \frac{\bar{X}_1 - \bar{X}_0}{S\sqrt{1/n_1 + 1/n_0}}$$

con campioni 'piccoli' ($n_1 + n_0 - 2 < 30$) $\Rightarrow$ distribuzione t di Student (sotto l'ipotesi nulla)



SESM

Esempio (continua):

	Numero di pazienti	Media (mg/100 ml)	DS (mg/100 ml)
Vitamina D (D_1)	233	9.36	1.15
Controllo (D_0)	394	9.01	1.33

$$\begin{cases} H_0: \mu_0 = \mu_1 \\ H_1: \mu_0 \neq \mu_1 \end{cases} \leftarrow \boxed{\text{I° STEP}}$$

$$t = \frac{\bar{x}_1 - \bar{x}_0}{\sqrt{S_1^2/n_1 + S_0^2/n_0}} = \frac{9.36 - 9.01}{0.1008} = 3.47 \leftarrow \boxed{\text{II° STEP}}$$

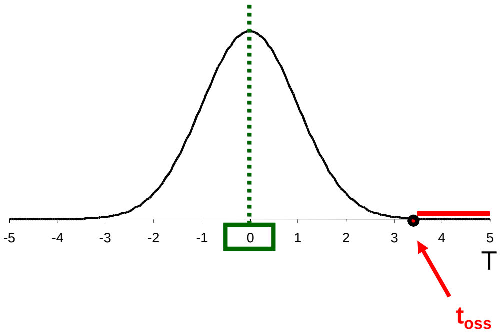
SESM

III° STEP: calcolare il p-value

misura continua del livello di consistenza dei dati con l'ipotesi nulla, ovvero di quanto sono plausibili i dati sotto H_0

ipotesi alternativa unilaterale

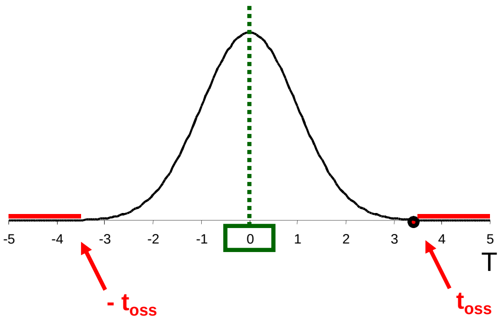
$$\begin{cases} H_0: \mu_0 = \mu_1 \\ H_1: \mu_0 < \mu_1 \end{cases}$$



upper one-sided p-value $(p_+) = \Pr(T \geq t_{oss} \mid H_0 \text{ è vera})$

ipotesi alternativa bilaterale

$$\begin{cases} H_0: \mu_0 = \mu_1 \\ H_1: \mu_0 \neq \mu_1 \end{cases}$$



upper one-sided p-value $(p_+) = \Pr(T \geq t_{oss} \mid H_0 \text{ è vera})$

lower one-sided p-value $(p_-) = \Pr(T \leq t_{oss} \mid H_0 \text{ è vera})$

two-sided p-value $(p) = 2 \min\{p_+, p_-\}$

Esempio (continua):

	Numero di pazienti	Media (mg/100 ml)	DS (mg/100 ml)
Vitamina D (D_1)	233	9.36	1.15
Controllo (D_0)	394	9.01	1.33

$$\begin{cases} H_0: \mu_0 = \mu_1 \\ H_1: \mu_0 \neq \mu_1 \end{cases}$$ I° STEP

$$t = \frac{\bar{x}_1 - \bar{x}_0}{\sqrt{s_1^2/n_1 + s_0^2/n_0}} = \frac{9.36 - 9.01}{0.1008} = 3.47$$ II° STEP

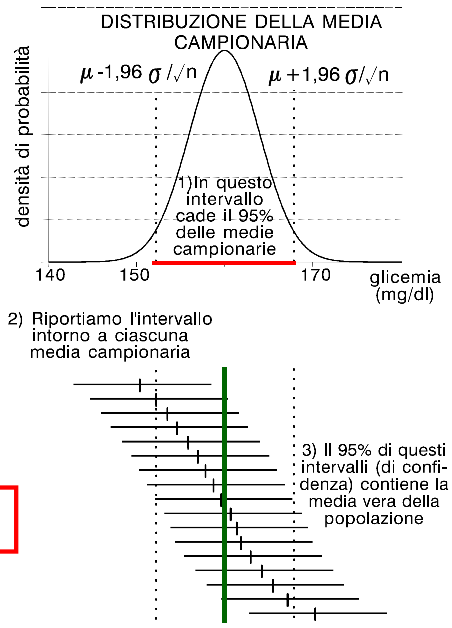
p-value = 2 prob ($T \geq 3.47 \mid H_0$ è vera) < 0.001 III° STEP

$\Rightarrow$ **risultato '*significativo*' e si rifiuta H_0**

Intervallo di confidenza della media in 2 popolazioni - campioni indipendenti

Per intervallo di confidenza della media μ, si intende un intervallo delimitato da due limiti L_{inf} (limite inferiore) ed L_{sup} (limite superiore) che abbia una definita probabilità (livello di confidenza) di contenere il vero valore (ignoto) del parametro nella popolazione:

$$\text{prob}(L_{inf} < \mu < L_{sup}) = 0.95$$

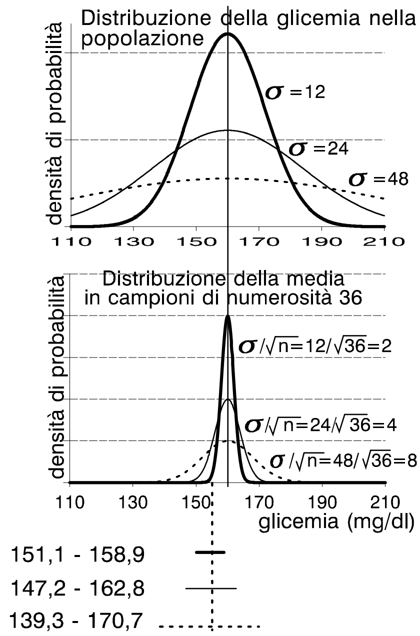


$\overline{x}$ **- 1.96 σ/√n ,** $\overline{x}$ **+ 1.96 σ/√n**

SESM

L'IC diminuisce se diminuisce la variabilità nella popolazione (σ)

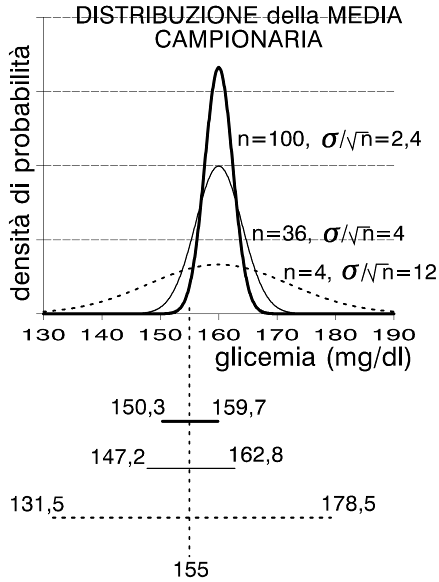
$IC_{95\%}(\mu) = \overline{x} \pm 1.96\ \sigma/\sqrt{n}$



SESM

L'IC diminuisce se aumenta la numerosità del campione (n)

$IC_{95\%}(\mu) = \bar{x} \pm 1.96\ \sigma / \sqrt{n}$



Esempio (continua):

	Numero di pazienti	Media (mg/100 ml)	DS (mg/100 ml)
Vitamina D **(D_1)**	**233**	**9.36**	**1.15**
Controllo **(D_0)**	**394**	**9.01**	**1.33**

$$95\%CI\ (\mu_0) = \bar{x}_0 \pm 1.96\ s_0 / \sqrt{n_0} = 9.01 \pm 1.96 * 1.33 / \sqrt{394}$$

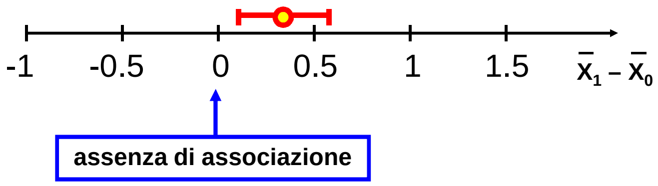
$$95\%CI\ (\mu_1) = \bar{x}_1 \pm 1.96\ s_1 / \sqrt{n_1} = 9.36 \pm 1.96 * 1.15 / \sqrt{233}$$

	controllo (n = 394)	vitamina D (n = 233)	p-value
media [95%CI]	**9.01 [8.88, 9.14]**	**9.36 [9.21, 9.51]**	**<0.001**

Esempio (continua):

	Numero di pazienti	Media (mg/100 ml)	DS (mg/100 ml)
Vitamina D (D_1)	233	9.36	1.15
Controllo (D_0)	394	9.01	1.33

differenza delle medie [95%CI]	0.35 [0.15, 0.55]



Intervalli di confidenza che indicano cinque possibili interpretazioni in termini di significatività statistica e importanza pratica