

## Test d'ipotesi: confronto fra medie

- Prof. Giuseppe Verlato
- Sezione di Epidemiologia e Statistica Medica, Università di Verona

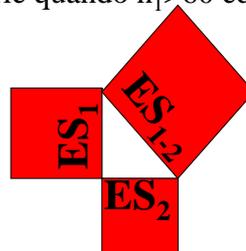
### CONFRONTO FRA MEDIE

1) confronto fra una media campionaria e una media di popolazione

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

2) confronto fra due medie campionarie quando  $n_1 > 60$  ed  $n_2 > 60$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$



Serve per capire il ragionamento teorico, ma non si utilizza nella pratica, si utilizza la formula 3

3) confronto fra due medie campionarie con piccoli campioni ( $n_1 < 60$  ed  $n_2 < 60$ )

$$s_p^2 = s_{pooled}^2 = \frac{dev_1 + dev_2}{n_1 + n_2 - 2}$$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{dev_1 + dev_2}{n_1 + n_2 - 2} \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

4) confronto per dati appaiati

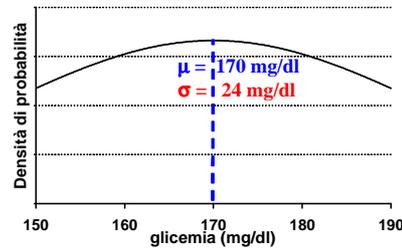
$$\left. \begin{array}{l} x_{11} - x_{21} = d_1 \\ x_{12} - x_{22} = d_2 \\ x_{13} - x_{23} = d_3 \\ x_{14} - x_{24} = d_4 \\ x_{15} - x_{25} = d_5 \end{array} \right\} \begin{array}{l} \bar{d} \\ s_d \end{array} \quad t = \frac{\bar{d} - 0}{s_d / \sqrt{n}}$$

Test d'ipotesi: confronto fra la media di un campione ( $\bar{x}$ ) e la media della popolazione ( $\mu$ )

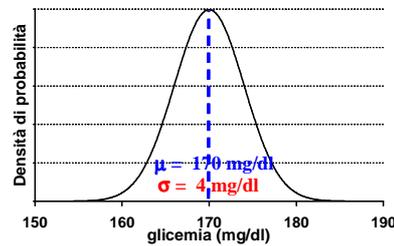
Assumiamo che nel Nord-Europa la glicemia dei diabetici di tipo 2 sia pari a  $170 \pm 24$  mg/dl ( $\mu \pm \sigma$ ).

In 36 diabetici italiani troviamo una glicemia media di 161 mg/dl.

La glicemia dei diabetici italiani è uguale alla glicemia dei diabetici del Nord-Europa?



Distribuzione della glicemia nei diabetici del Nord-Europa



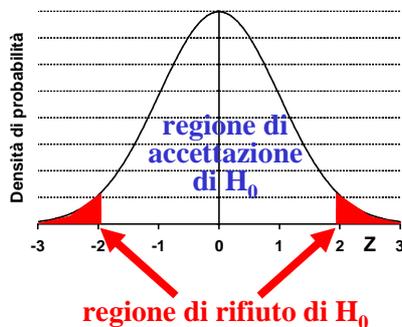
Corrispondente distribuzione della media campionaria per  $n=36$

( $\sigma = 24/\sqrt{36} = 24/6 = 4$  mg/dl)

N.B. Dal momento che  $n > 30$ , la distribuzione della media campionaria è normale indipendentemente dalla distribuzione della variabile originaria.

Ipotesi  $\left\{ \begin{array}{l} H_0: \mu_{it} = \mu_0 \\ H_1: \mu_{it} \neq \mu_0 \end{array} \right.$

Livello di significatività  
[P(errore del I tipo)] = 5%



Utilizzo il test z, basato sulla distribuzione z (deviata normale standardizzata)

Livello di significatività	10%	5%	1%
Soglia critica (valore assoluto)	1,645	1,96	2,576

**Il test è a 2 code, perché la regione di rifiuto è nelle 2 code della distribuzione.**

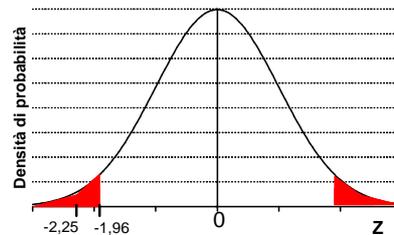
$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{161 - 170}{24 / \sqrt{36}} = \frac{-9}{4} = -2,25$$

| z osservato | > soglia critica  
2,25 > 1,96



Rifiuto  $H_0$

La glicemia dei diabetici italiani è diversa  
dalla glicemia dei diabetici del Nord-Europa  
( $P=2,4\%$ ).



Assumiamo che nel Nord-Europa la glicemia dei diabetici di tipo 2 si distribuisca normalmente con media 170 mg/dl. **La variabilità della glicemia è ignota.**

In 25 diabetici italiani troviamo una glicemia media di  $155 \pm 20$  mg/dl (media  $\pm$  deviazione standard).

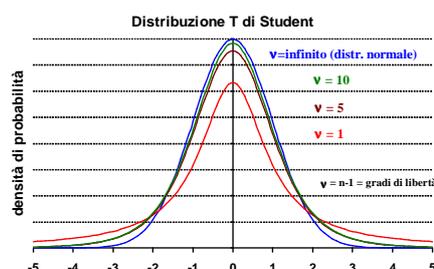
**La glicemia dei diabetici italiani è uguale alla glicemia dei diabetici del Nord-Europa?**

Stimiamo  $\sigma$  (deviazione standard della popolazione) a partire da  $s$  (deviazione standard del campione):

$$\text{Errore Standard} = s / \sqrt{n} = 20 / \sqrt{25} = 20 / 5 = 4$$

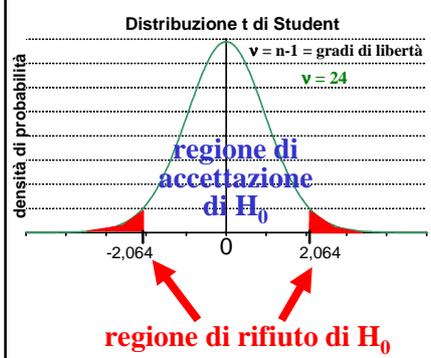
Tuttavia così facendo, introduciamo un'ulteriore fonte di variabilità campionaria: oltre a stimare la media, stimiamo anche la deviazione standard.

Per ovviare al problema, sostituiamo alla distribuzione  $z$  la distribuzione  $t$  di Student, che presenta una maggiore dispersione.



Ipotesi  $\begin{cases} H_0: \mu_{it} = \mu_0 \\ H_1: \mu_{it} \neq \mu_0 \end{cases}$

Livello di significatività  
[P(errore del I tipo)] = 5%



Con  $\alpha = 5\%$ ,  $n-1 = 24$   
soglia critica = 2,064

Il test è a 2 code, perché la regione di rifiuto è nelle 2 code della distribuzione.

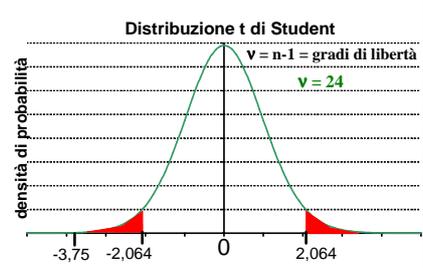
$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{155 - 170}{20/\sqrt{25}} = \frac{-15}{4} = -3,75$$

$|t \text{ osservato}| > t \text{ tabulato}$   
3,75 > 2,064



Rifiuto  $H_0$

La glicemia dei diabetici italiani è diversa  
(minore) dalla glicemia dei diabetici del  
Nord-Europa (P=0,001).



## Test d'ipotesi: confronto fra due medie campionarie ( $\bar{x}_1$ e $\bar{x}_2$ )

**Esempio:** 20 pazienti ipertesi vengono assegnati casualmente a due gruppi di trattamento. Il primo gruppo viene trattato con un diuretico, mentre il secondo gruppo viene trattato con un farmaco beta-bloccante. Dopo due settimane di trattamento viene rilevata la frequenza cardiaca (in battiti al minuto):

Diuretico	80	86	88	82	88	87	77	72	88	79
Beta-bloccante	70	65	66	76	68	66	71	72	69	82

**La frequenza cardiaca differisce in modo significativo tra i 2 gruppi?**

**Test t di Student (per dati non-appaiati)**

$$\begin{cases} H_0: \mu_{\text{diur}} = \mu_{\text{B.}} \\ H_1: \mu_{\text{diur}} \neq \mu_{\text{B.}} \end{cases}$$

**test a due code**

$$\begin{cases} H_0: \mu_{\text{diur}} \leq \mu_{\text{B.}} \\ H_1: \mu_{\text{diur}} > \mu_{\text{B.}} \end{cases}$$

**test a una coda**

**Livello di significatività = 5%**

**Gradi di libertà =  $n_1 + n_2 - 2 = 10 + 10 - 2 = 18$**

**Soglia critica =  $t_{18, 0,025} = 2,101$**

$$t = \frac{|\bar{x}_1 - \bar{x}_2|}{ES_{\bar{x}_1 - \bar{x}_2}} = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{(1/n_1 + 1/n_2) (\text{dev}_1 + \text{dev}_2) / (n_1 + n_2 - 2)}}$$

	$\Sigma X$	$\Sigma X^2$	n	$\bar{x}$	Dev	Var	DS
Diuretico	827	68675	10	82,7	282,1	31,34	5,599
Beta-bloc	705	49947	10	70,5	244,5	27,17	5,212

**Assunzioni:**

- 1) la frequenza cardiaca si distribuisce normalmente
- 2) La varianza non differisce tra i due gruppi

$$t = \frac{|82,7 - 70,5|}{\sqrt{(1/10 + 1/10) (282,1 + 244,5) / (10 + 10 - 2)}} = \frac{12,2}{\sqrt{5,85}} = \frac{12,2}{2,42} = 5,043$$

t osservato (5,043) > t tabulato (2,101)



Rifiuto  $H_0$  (P < 0,001)

Test d'ipotesi:  
t di Student per dati appaiati

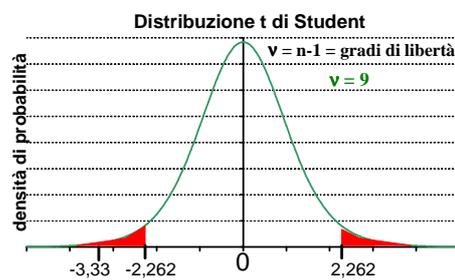


$t$  tabulato  
 $| t \text{ osservato} | > \text{soglia critica}$   
3,327                      2,262



Rifiuto  $H_0$

La glicemia dei diabetici è diminuita in modo significativo dopo somministrazione del farmaco ipoglicemizzante ( $P=0,009$ ).



Confronto fra medie:  
calcolo della numerosità campionaria

**Variabile di risposta quantitativa: Confronto fra 2 medie**

$$n > 2 \left[ \frac{(z_\alpha + z_\beta) \sigma}{\delta} \right]^2$$

dove **n** = numerosità di ciascuno dei due gruppi

**$z_\alpha = 1.96$**  per **alfa = 5%**

**$z_\beta = 0.842, 1.282, 1.645$**  per **potenza = 80, 90, 95%**

**$\sigma$**  = deviazione standard, desunta da studi pilota o dalla letteratura

**$\delta = \bar{x}_1 - \bar{x}_2$**  = differenza minima clinicamente rilevante

**Esempio:** Il farmaco di riferimento riduce la pressione sistolica di 25 mmHg, il nuovo farmaco per essere competitivo dovrebbe ridurre la pressione sistolica di almeno 30 mmHg (ovvero 5 mmHg in più). La deviazione standard della riduzione della pressione viene stimata in 10 mmHg da studi precedenti. Si adotta un alfa del 5% e una potenza del 90%, pertanto  $z_\alpha = 1.96$  e  $z_\beta = 1.282$ .

$$n > 2 \left[ \frac{(z_\alpha + z_\beta) \sigma}{\delta} \right]^2$$

$$n > 2 \left[ \frac{(1.96 + 1.28) 10}{5} \right]^2$$

$$n > 84.06 \quad n \geq 85$$

**Occorrono almeno 85 soggetti per gruppo.**