Argomenti

- Confronto tra le medie di due popolazioni indipendenti
- Confronto tra le medie di due popolazioni non indipendenti
- Confronto tra le proporzioni di due popolazioni
- Test *F* per la differenza tra due varianze
- Analisi della varianza (ANOVA) ad una via

Confronto tra medie di due pop. indipendenti

- Consideriamo due popolazioni indipendenti e supponiamo di estrarre un campione di ampiezza n_1 dalla prima popolazione di ampiezza n_2 dalla seconda popolazione
- Siano μ_1 e μ_2 le medie che caratterizzano rispettivamente la prima e la seconda popolazione e si assumano i due scarti quadratici medi σ_1 e σ_2 come noti
- Si vuole verificare l'ipotesi nulla che le medie delle due popolazioni (indipendenti) sono uguali tra loro

$$H_0\colon \mu_1 = \mu_2 \; (\mu_1 - \mu_2 = 0)$$

contro l'ipotesi alternativa

$$H_1\colon \mu_1 \neq \mu_2 \; (\mu_1 - \mu_2 \neq 0)$$

- A questo scopo viene definita la **statistica test *Z* per la differenza tra le due medie**

Confronto tra medie di due pop. indipendenti

Test Z per la differenza fra due medie

La statistica test in questo caso è:

$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \qquad (7.1)$$

dove:

$\overline{X}_1$ = media degli elementi del campione estratto dalla popolazione 1
μ_1 = media della popolazione 1
σ_1^2 = varianza della popolazione 1
n_1 = ampiezza del campione estratto dalla popolazione 1
$\overline{X}_2$ = media degli elementi del campione estratto dalla popolazione 2
μ_2 = media della popolazione 2
σ_2^2 = varianza della popolazione 2
n_2 = ampiezza del campione estratto dalla popolazione 2

Confronto tra medie di due pop. indipendenti

- Se si assume che i due campioni siano estratti casualmente ed indipendentemente da due popolazioni normali la statistica Z ha distribuzione normale
- Se le due popolazioni non hanno distribuzione normale il test Z può essere utilizzato con ampiezza campionarie sufficientemente elevate (in virtù del teorema del limite centrale)
- In molti casi le varianze delle due popolazioni non sono note. Se si assume l'ipotesi di omogeneità della varianze ($\sigma^2_1=\sigma^2_2$), per verificare se c'è una differenza significativa tra le medie delle due popolazioni è possibile utilizzare il **test *t* basato sulle varianze campionarie combinate**
- Il test t è appropriato se le popolazioni hanno distribuzione normale oppure, in caso contrario, se le ampiezze campionarie sono sufficientemente elevate

Confronto tra medie di due pop. indipendenti

Test t per la differenza fra due medie basato sulle varianze campionarie ponderate

La statistica test in questo caso è:

$$t = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{S_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \qquad (7.2)$$

dove

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{(n_1 - 1) + (n_2 - 1)}$$

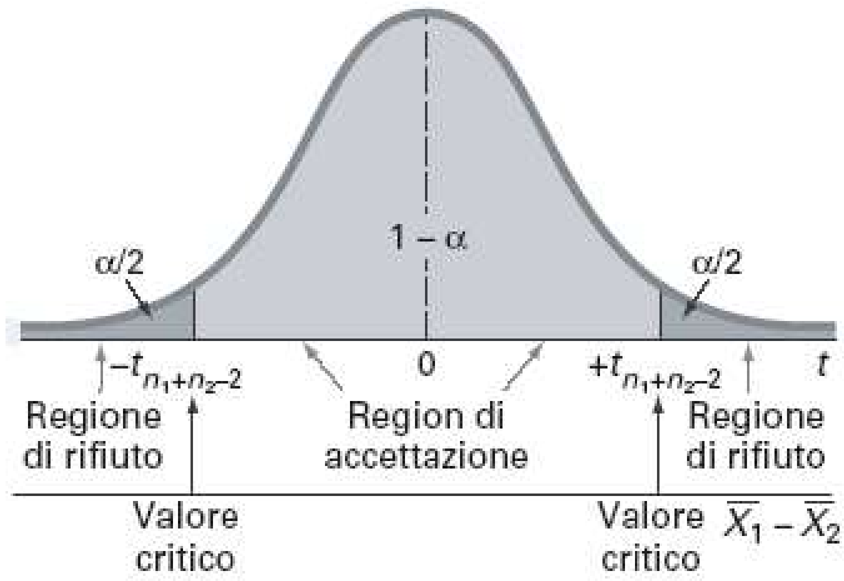
e

S_p^2 = varianza ponderata
$\overline{X}_1$ = media degli elementi del campione estratto dalla popolazione 1
S_1^2 = varianza degli elementi del campione estratto dalla popolazione 1
n_1 = ampiezza del campione estratto dalla popolazione 1
$\overline{X}_2$ = media degli elementi del campione estratto dalla popolazione 2
S_2^2 = varianza degli elementi del campione estratto dalla popolazione 2
n_2 = ampiezza del campione estratto dalla popolazione 2

Si dimostra che la statistica test t sotto l'ipotesi nulla si distribuisce secondo una t di Student con $n_1 + n_2 - 2$ gradi di libertà.

Confronto tra medie di due pop. indipendenti

Regione di rifiuto e di accettazione per la differenza tra due medie utilizzando la statistica test *t* basata sulle varianze combinate (test a due code)

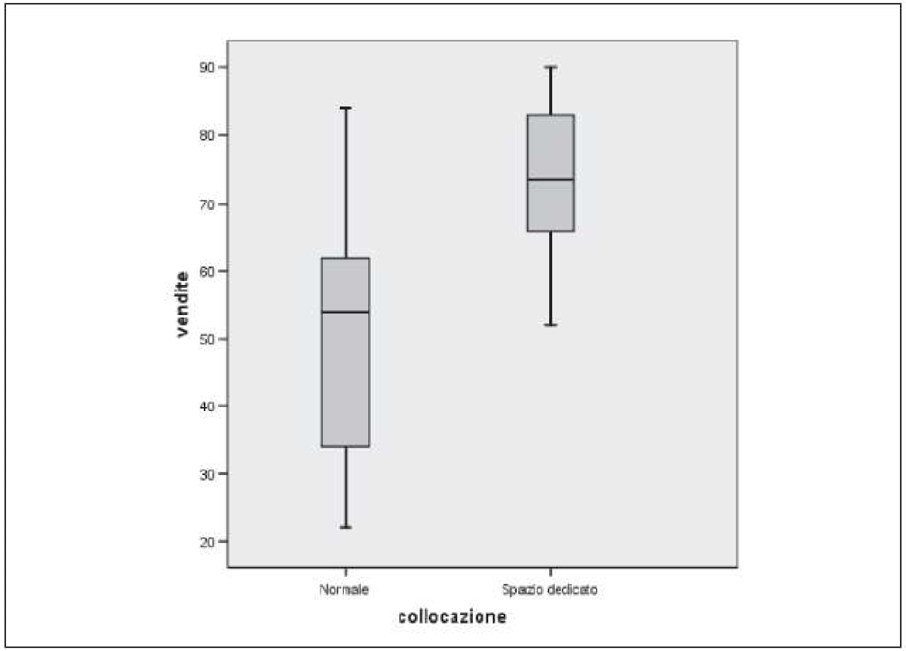


Quando l'assunzione dell'omogeneità delle varianze non è plausibile occorre fare riferimento al **test *t* con varianze diverse** (ricorrendo all'Excel o ad altri software statistici)

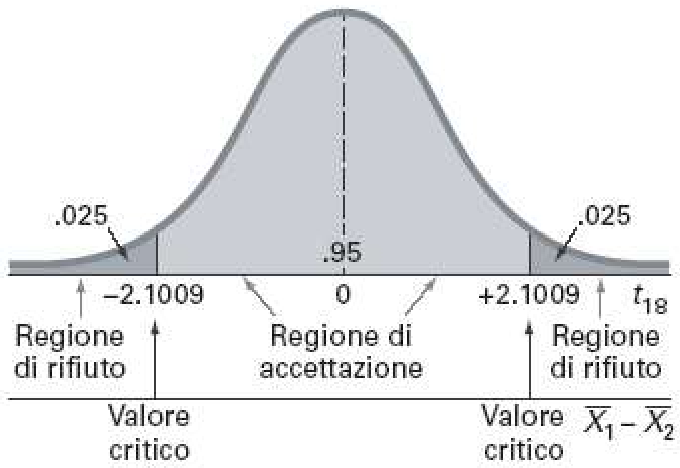
Confronto tra medie di due pop. indipendenti

Esempio: confronto tra le vendite settimanali (numero di pezzi venduti) della BLK cola in due gruppi di supermercati, dove il primo adotta la collocazione a scaffale mentre il secondo utilizza uno spazio dedicato

Collocazione									
Scaffale					Spazio dedicato				
22	34	52	62	30	52	71	76	54	67
40	64	84	56	59	83	66	90	77	84



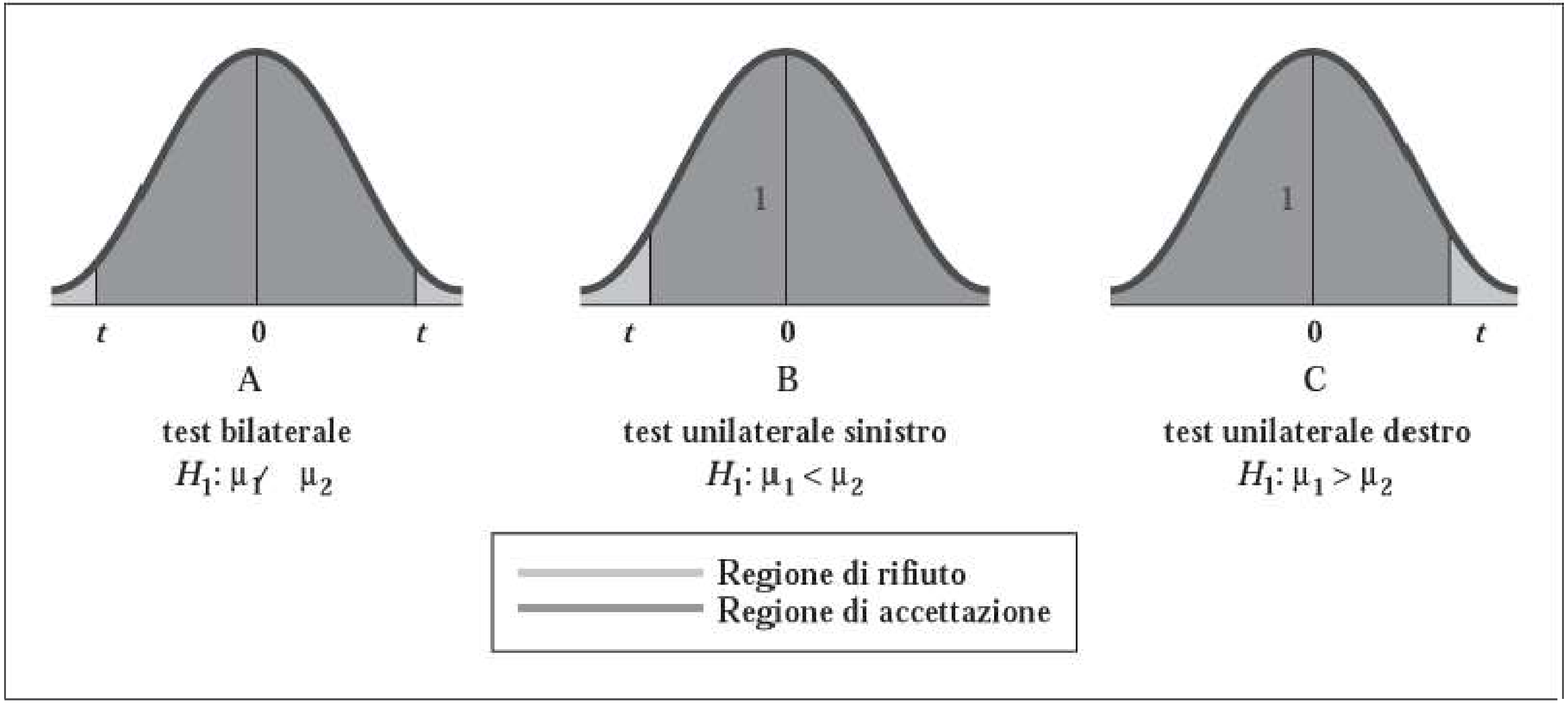
Confronto tra medie di due pop. indipendenti



	A	B	C
1	Test t: due campioni assumendo uguale varianza		
2			
3		*Scaffale*	*Spazio dedicato*
4	Media	50,3	72
5	Varianza	350,6778	157,3333
6	Osservazioni	10	10
7	Varianza complessiva	254,0056	
8	Differenza ipotizzata per le r	0	
9	gdl	18	
10	Stat t	-3,0446	
11	P(T<=t) una coda	0,0035	
12	t critico una coda	1,7341	
13	P(T<=t) due code	0,0070	
14	t critico due code	2,1009	

Confronto tra medie di due pop. indipendenti

In base al fatto che l'ipotesi alternativa sia nella forma A: H_1:$\mu_1 \neq \mu_2$ oppure B: H_1:$\mu_1 < \mu_2$ o C: H_1:$\mu_1 > \mu_2$ si parla di test ad una coda e test a due code



Intervallo di confidenza per la differenza tra le medie di due pop. indipendenti

Anziché (o oltre a) sottoporre a verifica l'ipotesi nulla secondo la quale due medie sono uguali, possiamo utilizzare l'equazione (10.3) per ottenere un intervallo di confidenza per la differenza tra le medie μ_1 e μ_2 delle due popolazioni:

Intervallo di confidenza per la differenza ($\mu_1 - \mu_2$)

$$(\bar{X}_1 - \bar{X}_2) - t_{n_1 - n_2 - 1; \alpha/2} \cdot \sqrt{S_p^2 \left(1/n_1 + 1/n_2\right)} \leq \mu_1 - \mu_2 \leq$$
$$\leq (\bar{X}_1 - \bar{X}_2) + t_{n_1 - n_2 - 1; \alpha/2} \cdot \sqrt{S_p^2 \left(1/n_1 + 1/n_2\right)} \qquad \textbf{(10.3)}$$

dove $t_{n1-n2-2;\alpha/2}$ è il valore critico a cui corrisponde un'area cumulata pari a $(1-\alpha/2)$ della distribuzione t di Student con $(n_1 - n_2 - 2)$ gradi di libertà.

Confronto tra medie di 2 pop. non indipendenti

Ci sono situazioni in cui le due popolazioni poste a confronto non sono indipendenti di modo che il campione estratto dalla prima popolazione non è indipendente dal campione estratto dalla seconda:

1. **campioni appaiati** (individui o casi che condividono una stessa caratteristica)
2. **misurazioni ripetute** (stesso insieme di individui o casi)

L'attenzione si sposta sulla differenze tra i valori nei due campioni:

	Gruppo		
Valore	1	2	Differenza
1	X_{11}	X_{21}	$D_1 = X_{11} - X_{21}$
2	X_{12}	X_{22}	$D_2 = X_{12} - X_{22}$
⋮	⋮	⋮	⋮
i	X_{1i}	X_{2i}	$D_i = X_{1i} - X_{2i}$
⋮	⋮	⋮	⋮
n	X_{1n}	X_{2n}	$D_n = X_{1n} - X_{2n}$

Confronto tra medie di 2 pop. non indipendenti

Quindi verificare l'ipotesi di uguaglianza delle medie μ_1 e μ_2 di due popolazioni non indipendenti equivale a verificare ipotesi di uguaglia a zero della media della differenza D tra le due popolazioni, cioè H_0: $\mu_D=0$. Se lo scarto quadratico medio della popolazione delle differenze σ_D è noto, allora il test di riferimento è basato sulla statistica Z. In caso σ_D sia ignoto si può fare ricorso al **test *t* su campioni appaiati**.

Statistica test *Z* per la media delle differenze

$$Z = \frac{\bar{D} - \mu_D}{\sigma_D / \sqrt{n}}, \text{con } \bar{D} = 1/n \sum_{i=1}^{n} D_i \qquad \textbf{(10.4)}$$

Statistica test *t* per la media delle differenze

$$t = \frac{\bar{D} - \mu_D}{S_D / \sqrt{n}}, \text{con } \bar{D} = \frac{\sum_{i=1}^{n} D_i}{n} \text{ e } S_D = \sqrt{\frac{\sum_{i=1}^{n} (D_i - \bar{D})^2}{(n-1)}} \qquad \textbf{(10.5)}$$

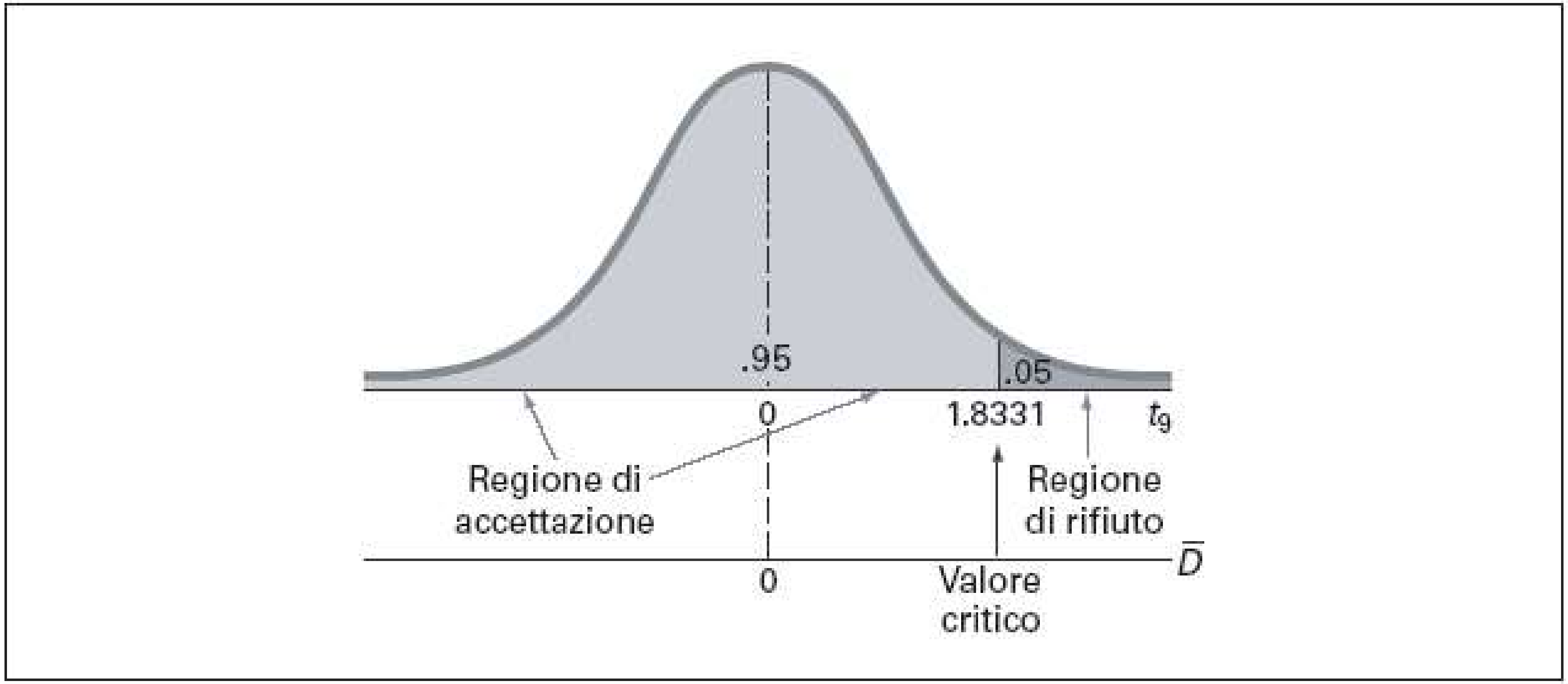
Confronto tra medie di 2 pop. non indipendenti

Esempio: Misurazioni ripetute del tempo (in secondi) di elaborazione di un progetto utilizzando due diversi software

Tempo di elaborazione (in secondi)			
Progetto	Software sul mercato	Nuovo software	Differenza (D_i)
1	9.98	9.88	+0.10
2	9.88	9.86	+0.02
3	9.84	9.75	+0.09
4	9.99	9.80	+0.19
5	9.94	9.87	+0.07
6	9.84	9.84	0.00
7	9.86	9.87	−0.01
8	10.12	9.86	+0.26
9	9.90	9.83	+0.07
10	9.91	9.86	+0.05
			+0.84

Confronto tra medie di 2 pop. non indipendenti

Test *t* a una coda per la differenza tra le medie di due popolazioni non indipendenti a un livello di significatività pari a 0.05 e con 9 gradi di libertà



Intervallo di confidenza per la differenza tra le medie di due pop. non indipendenti

Anziché (o oltre a) sottoporre a verifica l'ipotesi nulla secondo la quale due medie sono uguali, possiamo utilizzare l'equazione (10.6) per ottenere un intervallo di confidenza per la differenza μ_D:

Intervallo di confidenza per la differenza tra le medie di due popolazioni non indipendenti

$$\bar{D} - t_{n-1;\alpha/2} S_D / \sqrt{n} \le \mu_D \le \bar{D} + t_{n-1;\alpha/2} S_D / \sqrt{n} \qquad \textbf{(10.6)}$$

dove $t_{n-1;\alpha/2}$ è il valore critico a cui corrisponde un'area cumulata pari a $(1-\alpha/2)$ della distribuzione t di Student con $(n-1)$ gradi di libertà