Confronto fra due medie osservate aventi varianza diversa

Se, applicando il test F, le varianze di due serie di dati risultano significativamente diverse, si imposta il test per il confronto delle medie osservate come per il caso precedente

Ipotesi

H_0: $x_A = x_B$ *test a 2 code*
H_1: $x_A \neq x_B$

Funzione discriminante

Si calcolano il t di Student e i gradi dì libertà nel modo seguente:

$$t_c = \frac{\bar{x}_A - \bar{x}_B}{\sqrt{\dfrac{s_A^2}{n_A} + \dfrac{s_B^2}{n_B}}}$$

$$\nu = \frac{\left(\dfrac{s_A^2}{n_A} + \dfrac{s_B^2}{n_B}\right)^2}{\dfrac{\left(\dfrac{s_A^2}{n_A}\right)^2}{n_A+1} + \dfrac{\left(\dfrac{s_B^2}{n_B}\right)^2}{n_B+1}} - 2$$

Anche in questo caso bisognerebbe calcolare z per i campioni con $n \geq 30$ e t per i campioni con $n < 30$, ma poiché la funzione discriminante è la stessa per entrambi i parametri e i valori di t tendono ai valori di z per $n \rightarrow \infty$, si può usare la funzione t in tutti i casi.

ESEMPIO 16 **Per la determinazione del maltosio in un campione di birra, sono stati usati il metodo di Fehling (1) e il metodo di Lane-Enyon (2), che hanno fornito i seguenti risultati.**

	$\bar{x}$	s	n
Metodo **1**	0,817	0,053	13
Metodo **2**	0,853	0,028	10

In base al test F le due varianze risultano diverse. Verificare se i due risultati possono essere considerati simili.

Si calcola t:

$$t_c = \frac{\bar{x}_A - x_B}{\sqrt{\dfrac{s_A^2}{n_A} + \dfrac{s_B^2}{n_B}}} = \frac{0{,}853 - 0{,}817}{\sqrt{\dfrac{(0{,}053)^2}{13} + \dfrac{(0{,}028)^2}{10}}} = 2{,}098$$

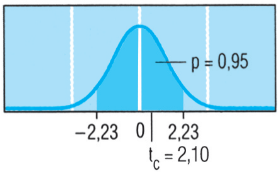
e si calcolano i gradi di libertà:

$$\nu = \frac{\left(\dfrac{s_A^2}{n_A} + \dfrac{s_B^2}{n_A}\right)^2}{\dfrac{\left(\dfrac{s_A^2}{n_A}\right)^2}{n_A+1} + \dfrac{\left(\dfrac{s_B^2}{n_B}\right)^2}{n_B+1}} - 2 = \frac{\left[\dfrac{(0{,}053)^2}{13} + \dfrac{(0{,}028)^2}{10}\right]^2}{\dfrac{\left[\dfrac{(0{,}053)^2}{13}\right]^2}{14} + \dfrac{\left[\dfrac{(0{,}028)^2}{10}\right]^2}{11}} - 2 \simeq 10$$

Il valore tabulato di t (*v.* tabella 22.6) per $\alpha = 0{,}05$ e $\nu = 10$ è 2,228; quindi, poiché:

$$t_c < t_{0,05;\,10}$$

l'ipotesi nulla è da considerarsi valida e i due metodi possono essere ritenuti simili per quanto riguarda l'accuratezza del risultato, anche se la precisione del metodo 2 è decisamente migliore di quella del metodo 1.



▶ Confronto di dati appaiati

A volte si devono confrontare due insiemi di **dati appaiati**, cioè tali che a ogni valore del primo corrisponde un valore del secondo.

Un esempio può essere quello di una variabile determinata prima e dopo un particolare trattamento effettuato su un campione.

Ipotesi

H_0: $d = 0$ *test a due code*
H_1: $d \neq 0$

ovvero:
H_0: Le differenze (d) fra le due serie non sono diverse
H_1: Le differenze sono significativamente diverse

Funzione discriminante

Anche in questo caso bisognerebbe calcolare z per i campioni con $n \geq 30$ e t per campioni con $n < 30$, ma poiché valgono le stesse considerazioni viste per il caso precedente, si confronta la media delle deviazioni con il suo errore standard, calcolando t nel modo seguente:

$$t_c = \frac{\bar{x} - \bar{d}}{\sqrt{\frac{s_d^2}{n}}}$$

dove • d è la media delle differenze fra le coppie di dati, ciascuna presa con il proprio segno • s_d^2 la varianza delle differenze • n il numero di coppie di dati.

Decisione

Anche in questo caso, si confronta il valore (t_c) calcolato con il corrispondente valore (t_t) riportato nella tabella 31.6 in base ai gradi di libertà ($\nu = n - 1$) e al valore di α prescelto. Se:

$$t_c \leq t_t$$

allora l'ipotesi nulla è da ritenersi valida, ovvero si può ritenere che non vi sia differenza fra le due serie di dati. Se invece:

$$t_c > t_t$$

allora la differenza è significativamente diversa (il trattamento effettuato sul campione fra la prima e la seconda serie di osservazioni ha dunque prodotto delle differenze significative sui risultati).

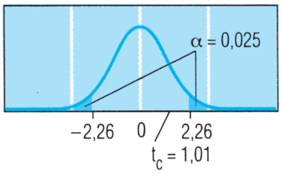
ESEMPIO 17

Si vuole stabilire se il numero di globuli rossi rimane costante da un giorno all'altro in un campione di sangue a cui è stato aggiunto EDTA. Per l'indagine, si usa il sangue prelevato da 10 individui e su ciascuno si effettuano due determinazioni a 24 ore di distanza l'una dall'altra. I dati raccolti, espressi in milioni di globuli rossi per mm3 sono riportati nella seguente tabella.

n	Globuli rossi (milioni/mm^2)		*d*
	giorno 1	giorno 2	
1	4,2	4,2	0,0
2	3,9	3,8	0,1
3	4,8	4,5	0,3
4	5,2	5,1	0,1
5	4,1	4,6	−0,5
6	4,4	4,5	−0,1
7	4,9	4,5	0,4
8	5,2	4,8	0,4
9	4,5	4,0	0,5
10	4,9	5,1	−0,2

$$\sum d_i = 1{,}0 \qquad \bar{d} = \frac{1{,}0}{10} = 0{,}10 \qquad \sum d_i^2 = 0{,}98 \qquad \frac{(\sum d_i)^2}{n} = \frac{(1{,}0)^2}{10} = 0{,}10$$

$$s_d^2 = \frac{0{,}98 - 0{,}10}{9} = 0{,}098 \qquad t_c = \frac{0{,}10}{\sqrt{\dfrac{0{,}098}{10}}} = 1{,}01$$



Il valore di *tc* calcolato è dunque minore del valore tabulato (tab. 22.6) t_t (2,262) per $\alpha = 0{,}05$ e per 9 gradi di libertà. Le differenze fra le coppie di osservazioni, quindi, non sono significative e si può affermare che il numero di globuli rossi non cambia significativamente da un giorno all'altro in seguito al trattamento con EDTA.