5. La distribuzione *t*

5.1 La distribuzione *t*

5. 1. 1. Il concetto di errore standard; errore standard della media

La (3.2) informa circa il valore della deviazione standard delle medie campionarie. Tale deviazione standard è indicata col termine di *errore standard della media*, ed è indicata col simbolo $\sigma_{\overline{Y}}$.

Il concetto di errore standard può essere applicato non solo alla media, ma a qualunque altra statistica *x*. In generale la deviazione standard della distribuzione di una qualsiasi statistica campionaria *x* è detta errore standard di quella stessa distribuzione ed è indicata col simbolo σ_x.

5. 1. 2. La statistica *t* e la sua distribuzione

Consideriamo una popolazione statistica *normalmente distribuita*, di cui sia noto il parametro μ. Da tale popolazione estraiamo un campione di dimensione *n*. Di tale campione calcoliamo la media $\overline{Y}$ e la deviazione standard s. Sulla base di questi due valori, calcoliamo la nuova statistica *t* definita dalla formula seguente:

$$t = \frac{\overline{Y} - \mu}{s/\sqrt{n}} \qquad (5.1)$$

Si noti che la nuova statistica *t* ha una formulazione analoga a quella per la statistica *z* definita attraverso la (3.4); tuttavia al suo denominatore non compare la deviazione standard parametrica della popolazione σ, bensì la sua stima campionaria *s*, calcolata sulla base del campione. La frazione $s/\sqrt{n}$ esprime a sua volta una stima campionaria dell'errore standard della media $\sigma_{\overline{Y}}$ definito nella (3.2), e pertanto può essere espressa attraverso il corrispondente simbolo campionario $s_{\overline{Y}}$:

$$s_{\overline{Y}} = s/\sqrt{n} \qquad (5.2)$$

La (5.1) può pertanto ammette una formulazione equivalente:

$$t = \frac{\overline{Y} - \mu}{s_{\overline{Y}}} \qquad (5.3)$$

Ora, supponiamo di estrarre dalla popolazione in tutti i possibili differenti campioni di dimensione *n*, e di calcolare per ciascuno di essi la nuova statistica *t* come descritto sopra. In tal modo otteniamo la distribuzione campionaria dei valori della statistica *t* basata su campioni di *n* elementi relativa alla popolazione di partenza.

Si dimostra che la statistica *t* non è distribuita normalmente. Ciò è dovuto al fatto che *s* (la deviazione standard campionaria) è una stima distorta della deviazione standard parametrica σ (vedi conclusione di § 3. 3. 2.). Tuttavia, ricordando che per *n* grande l'effetto della distorsione si affievolisce, dobbiamo attenderci che in questo caso la distribuzione dei valori della statistica *t* si normalizzi.

Le considerazioni appena fatte spiegano in parte come mai la distribuzione *t* dipenda da *n*; cioè valori differenti di *n* implicano distribuzioni di *t* differenti. Più precisamente, anziché far riferimento direttamente ai valori di *n* è consuetudine riferirsi al valore *n* – 1, che in § 3. 3. 3. abbiamo etichettato nella (3.11) con la lettera greca ν e definito come *gradi di libertà* della varianza (e della deviazione standard). Utilizzando questa terminologia diciamo che la distribuzione del *t* dipende dai gradi di libertà della statistica.

Il primo studio sulla distribuzione della statistica *t* si deve al chimico Gosset, che pubblicò i risultati sotto lo pseudonimo di Student. Per questo motivo la distribuzione del *t* è spesso riferita come distribuzione del *t di Student*. Semplificando un po' le cose, la formulazione matematica della distribuzione del *t* può essere scritta così:

$$f = \frac{Y_0}{\left(1 + \frac{t^2}{\nu}\right)^{\frac{\nu+1}{2}}} \qquad (5.4)$$

dove Y_0 è una costante che dipende da ν. Come per l'equazione della distribuzione normale anche questa equazione non verrà nel seguito utilizzata, ed è riportata solo per evidenziare la dipendenza della distribuzione dai gradi di libertà ν della statistica.

Si è detto che la curva di distribuzione del *t* si differenzia dalla distribuzione normale per valori piccoli di *n* (e quindi di ν), mentre per $n \to \infty$ (cioè per $\nu \to \infty$) la distribuzione tende alla normalità. Ai fini pratici già quando *n* vale 30 la distribuzione del *t* è praticamente uguale alla normale. Per questo motivo la distribuzione *t* ed i test basati su di essa che presenteremo nel seguito sono particolarmente adatti per i cosiddetti *piccoli campioni* (cioè con $n < 30$).

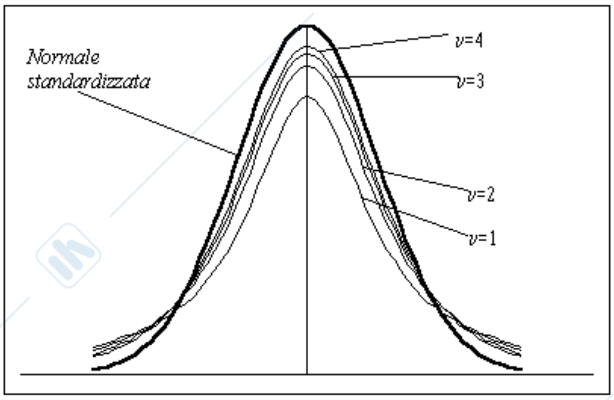

Fig. 5.1

Fig. 5.1 mostra il grafico delle curve di distribuzione del *t* corrispondenti a 1, 2, 3 e 4 gradi di libertà della statistica. Le curve sono confrontate con la normale standardizzata.

N.B.: si ricordi che la popolazione da cui estrarre i campioni di dimensione *n*, in base ai quali calcolare i valori della statistica *t*, deve essere distribuita normalmente. In difetto di tale assunzione vengono a mancare i presupposti essenziali su cui si basa tutto il ragionamento sin qui condotto. Tutte le applicazioni della statistica *t* e della sua distribuzione, presentate nei paragrafi successivi, perdono di significato e di valore qualora applicate ad una popolazione che non sia normalmente distribuita. Per tali casi il lettore è rinviato al successivo § 7.5., in cui vengono presentate alternative di calcolo che non postulano una distribuzione normale.

5. 1. 3. Le tavole del t a due code

Come per tutte le curve di distribuzione, l'area totale sotto la curva di distribuzione vale 1.

Anche le aree sotto la curva di distribuzione del *t* rappresentano delle probabilità; più precisamente:

l'area sotto la curva di distribuzione del t con ν gradi di libertà compresa fra due valori a e b rappresenta la probabilità di ottenere un valore di t compreso fra a e b campionando a caso con un campione di $\nu + 1$ elementi.

Non essendoci *una* curva di distribuzione *t* ma infinite (corrispondenti alle infinite possibilità di scegliere ν), non è materialmente possibile approntare una tavola per i valori sotto le aree della curva analoga alla **Tavola 2** per la distribuzione normale standardizzata (che invece è unica).

La **Tavola 3** per la distribuzione *t* è pertanto organizzata in modo differente. Essa non contiene le aree corrispondenti ai diversi valori di *t*, ma i suoi valori critici relativi a certi livelli di probabilità standard.

La Fig. 5.2 ci aiuterà a capire.

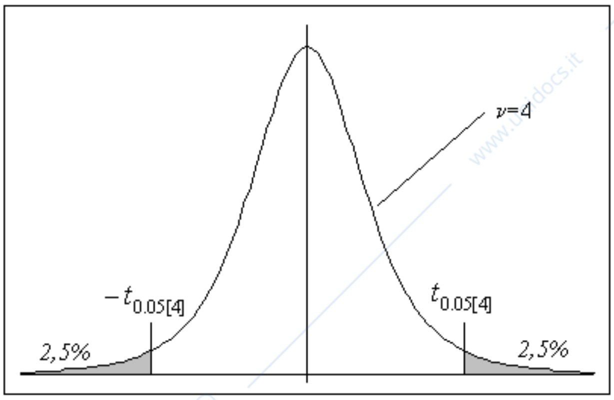

Fig. 5.2

In essa è rappresentata la curva di distribuzione di un *t* con ν = 4. Ciascuna delle due code evidenziate copre un'area pari al 2.5%, e quindi complessivamente l'area sotto le due code è pari al 5%. Il valore critico corrispondente è (analogamente a quanto detto in generale in § 4. 2. 5.) quel valore della statistica *t* che delimita le due code. La simbologia attraverso cui viene indicato è $t_{0.05[4]}$, dove il pedice 0.05 indica il livello di significatività della statistica, mentre sempre a pedice e fra parentesi quadre vengono indicati i gradi di libertà. In linea generale la simbologia corrispondente al valore critico del *t* con ν gradi di libertà ad un livello α di significatività è $t_{\alpha[\nu]}$.

La ricerca del valore critico nella **Tavola 3** è molto semplice: nella prima colonna sono riportati i diversi gradi di libertà ν (nel nostro esempio è 4); una volta trovato il valore desiderato ci si sposta in orizzontale nella colonna del livello α desiderato (nel nostro caso 0.05); otteniamo: $t_{0.05[4]} = 2.776$. A titolo di esempio è facile verificare che $t_{0.01[12]} = 3.055$, o che $t_{0.001[25]} = 3.725$.

L'ultima riga delle tavole del t è per il caso di infiniti gradi di libertà; abbiamo già detto che in questo caso la distribuzione del t non è altro che la normale standardizzata, e quindi dobbiamo aspettarci di trovare per i valori critici dei numeri noti. E infatti troviamo $t_{0.05[\infty]} = 1.960$ che equivale al valore critico $z_{0.05} = 1.96$ che abbiamo ricavato per un test a due code in § 4. 2. 5.
Si noti che nella prima colonna della **Tavola 3** per valori di ν superiori a 40 sono riportate sommariamente solo alcune righe. Questo è ragionevole visto che la distribuzione t è specifica per piccoli campioni, mentre per i grandi vale la pena di utilizzare direttamente la distribuzione normale, e quindi la **Tavola 2**.
Come per la **Tavola 2** anche per la **Tavola 3** vengono riportati i valori relativi alla sola parte positiva della distribuzione per ovvi motivi di simmetria.
Per concludere, occorre richiamare l'attenzione su un fatto importante; come si è detto si tratta di una tavola *a due code*; nel caso di test ad una sola coda occorre porre in atto un semplice ed intuitivo espediente: i valori critici al livello α di significatività per test a una coda sono uguali ai corrispondenti valori critici a livello 2α per i test a due code (si riguardino in proposito § 4. 2. 5., nonché Fig. 4.4 e Fig. 4.5). In pratica, se si dispone di tavole a due code, come la **Tavola 3**, per i valori critici a livello 0.05, 0.01 e 0.001 in un test ad una coda occorre cercare rispettivamente i valori critici a livello 0.10, 0.02 e 0.002.

5.2. Limiti di confidenza della media parametrica

5. 2. 1. Il concetto di limiti di confidenza

Supponiamo di essere interessati a conoscere una determinata popolazione scolastica, ad esempio gli studenti delle classi seconde di un dato istituto, rispetto ad una determinata abilità; supponiamo inoltre che vi sia l'impossibilità pratica (per motivi logistici) di passare in rassegna l'intera popolazione in modo esaustivo. Si procede dunque all'estrazione di un campione in rappresentanza dell'intera popolazione e si somministra la prova oggettiva. La media $\overline{Y}$ dei punteggi grezzi stima la media μ dell'intera popolazione, cioè abbiamo motivo di pensare, sulla base delle argomentazioni espresse nel Cap. 3, che il valore (incognito) della media parametrica μ sia *vicino* al valore della media campionaria $\overline{Y}$. Ma esattamente, quanto *vicino*?
I limiti di confidenza di un parametro sono due valori L_1 e L_2 al cui interno abbiamo fiducia, espressa in termini di probabilità, che si trovi il parametro.
Nel caso dell'esempio introduttivo, fissare i limiti di confidenza del parametro μ significa trovare due valori L_1 e L_2 al cui interno stia il valore incognito μ con una probabilità assegnata. Calcolare i limiti di confidenza per μ equivale dunque a chiarire in quali termini il valore $\overline{Y}$ stimi il parametro μ, e, si noti bene, senza conoscere μ stesso.

5. 2. 2. Formule di calcolo per i limiti di confidenza della media parametrica μ

Torniamo all'esempio introduttivo. Dopo aver somministrato la prova oggettiva e aver raccolto tutti i punteggi grezzi, conosciamo i valori di n (numerosità del campione), $\overline{Y}$ (media campionaria dei punteggi grezzi) ed s (deviazione standard campionaria dei punteggi grezzi).
A questo punto, facciamo riferimento alla Fig. 5.3 che generalizza la Fig. 5.2:

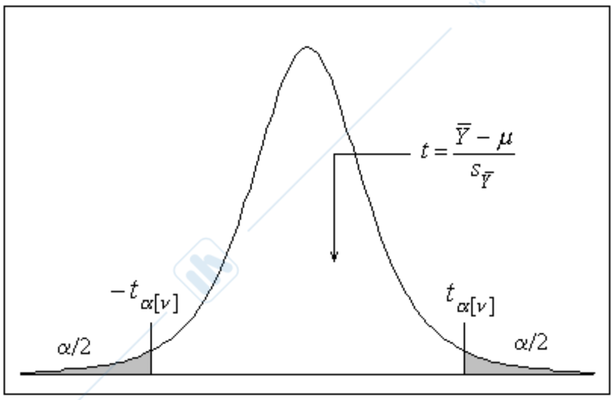

Fig. 5.3

i gradi di libertà sono ν invece di 4; al posto di una area totale sotto le code pari a 0.05 abbiamo un'area pari ad α (α/2 per ciascuna delle due code); infine, al posto dei valori critici $\pm t_{0.05[4]}$ abbiamo $\pm t_{\alpha[\nu]}$.
Ora, se la probabilità che t cada sotto le code della distribuzione è α, allora la probabilità che t cada internamente ai due valori critici sarà 1 – α. In simboli, la relazione $-t_{\alpha[\nu]} < t < t_{\alpha[\nu]}$ vale con una probabilità pari ad 1 – α.
Ora ricordando la (5.3) che definisce la statistica t possiamo scrivere:

$$-t_{\alpha[\nu]} < \frac{\overline{Y} - \mu}{s_{\overline{Y}}} < t_{\alpha[\nu]} \qquad (5.5)$$

da cui, attraverso semplici passaggi algebrici si ottiene:

$$\overline{Y} - t_{\alpha[\nu]} \cdot s_{\overline{Y}} < \mu < \overline{Y} + t_{\alpha[\nu]} \cdot s_{\overline{Y}} \qquad (5.6)$$

ricordando sempre che la sua validità non è assoluta ma probabilistica; cioè la (5.6) è valida con una probabilità pari ad $1 - \alpha$.

I due membri estremi della catena di disuguaglianze (5.6) danno la formulazione matematica dei limiti di confidenza di μ. In pratica abbiamo:

$$\begin{aligned} L_1 &= \overline{Y} - t_{\alpha[\nu]} \cdot s_{\overline{Y}} \\ L_2 &= \overline{Y} + t_{\alpha[\nu]} \cdot s_{\overline{Y}} \end{aligned} \qquad (5.7)$$

Come si vede dalle (5.7) i limiti di confidenza per μ sono simmetrici rispetto a $\overline{Y}$; la situazione può essere visualizzata graficamente come in Fig. 5.4.

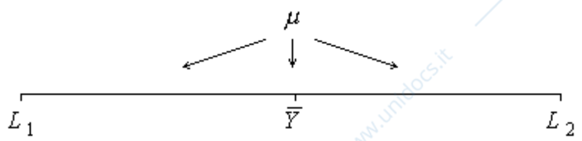

Fig. 5.4

Gli aspetti computazionali più pratici del sono illustrati nel Box 5.1, in cui sono calcolati i limiti di confidenza di un parametro μ per ciascuno dei tre livelli convenzionali di α.

5. 2. 3. Come variano i limiti di confidenza al variare di α, s e n

Come si vede bene dai risultati del Box 5.1, al diminuire di α i limiti di confidenza di μ si allargano. Per capire questo fatto occorre pensare che diminuendo α aumenta la probabilità $1 - \alpha$ che μ cada effettivamente entro i limiti calcolati; in altre parole, diminuire α equivale ad aumentare la nostra sicurezza sui limiti calcolati. Se vogliamo essere più sicuri, dobbiamo pagare in termini di allargamento dei limiti. Con una similitudine prosaica ma efficace, è un po' come succede ai principianti del tennis, che per essere sicuri di prendere la palla hanno bisogno di una racchetta più larga; cioè: se vogliamo essere più sicuri di prendere la pallina μ, la racchetta dei limiti di confidenza deve essere più larga.

Vediamo ora come influisce s sul calcolo dei limiti di confidenza.

Matematicamente parlando, quando s diminuisce anche l'errore standard della media diminuisce in forza della (5.2); e analogamente un aumento di s causa un aumento dell'errore standard della media. Per questo motivo se ora guardiamo le (5.7) ci accorgiamo che all'aumentare di s i limiti si allontanano maggiormente da $\overline{Y}$ e dunque si allargano, mentre diminuendo s i limiti si avvicinano a $\overline{Y}$ e dunque si stringono. In sintesi i limiti di confidenza sono ampi quando s ha un valore alto, e sono stretti quando s ha un valore basso.

Quanto dedotto matematicamente ha una sua intuibilità: se s è piccolo i dati hanno una scarsa variabilità e ci aspettiamo che la media parametrica non sia tanto lontana da $\overline{Y}$; se viceversa i dati sono caratterizzati da ampia variabilità siamo rassegnati a pensare che la media parametrica possa essere anche lontana $\overline{Y}$.

Per concludere vediamo cosa succede al variare di n.

Anche in questo caso tutto è molto intuitivo: se il mio campione è costituito da parecchi elementi nutriamo una buona speranza che il valore di $\overline{Y}$ sia vicino a μ, mentre all'opposto se il campione è costituito da un basso numero di elementi la media $\overline{Y}$ avrà meno probabilità di avvicinarsi a μ. Dunque, ad intuito ci aspettiamo che al diminuire di n i limiti di confidenza si allarghino, e viceversa.

Ciò è facile da vedere anche matematicamente, in quanto dalla (5.2) diminuendo n aumenta l'errore standard della media e viceversa; ma non basta: il valore di n influisce anche sul valore critico $t_{\alpha[\nu]}$: è facile vedere che diminuendo i gradi di libertà i valori critici del t aumentano.

5. 3. Confronto di una media campionaria con una media parametrica

Il problema è simile a quello già trattato in § 4. 1. e sintetizzato nel Box 4.1.

Là dovevamo confrontare una popolazione di cui era nota una media campionaria $\overline{Y}$ con una seconda popolazione di cui erano noti i parametri μ e σ. Nel caso trattato ora invece conosciamo le stime campionarie $\overline{Y}$ e s di una popolazione da confrontare con una seconda popolazione di cui è noto μ. Faremo uso della statistica t, e dunque il test è particolarmente adatto per piccoli campioni.

Un gruppo di ricerca ha elaborato e sperimentato su una popolazione locale un determinato reattivo, supponiamo un questionario per la misurazione di una determinata attitudine scolastica. La sperimentazione per la validazione del reattivo è stata fatta su una popolazione vasta ma locale. In una pubblicazione i ricercatori descrivono il reattivo ed illustrano i risultati della somministrazione alla popolazione locale, indicando il punteggio medio raggiunto.
Consideriamo questo dato come un parametro della popolazione studentesca locale. Indichiamolo con μ.
La pubblicazione riscuote l'interesse di un gruppo di lavoro che intende verificarne l'efficacia nella propria realtà scolastica. Il reattivo è dunque somministrato ad un piccolo campione di n studenti; il punteggio medio raggiunto è $\overline{Y}$ e la deviazione standard s.
Il gruppo di lavoro è ora interessato a confrontare i propri risultati con quelli della pubblicazione.
L'ipotesi nulla H_0 è quella di omogeneità delle due popolazioni studentesche. Non avendo particolari motivi per sospettare che le eventuali differenze abbiano un verso piuttosto che un altro, l'ipotesi alternativa H_1 è che le due popolazioni abbiano medie parametriche differenti. Si tratta dunque di un test a due code.
Per questa situazione la statistica z definita dalla (3.3) non è appropriata, in quanto non disponiamo del parametro σ della popolazione indagata dai ricercatori, ma della deviazione standard campionaria della popolazione indagata dal gruppo di lavoro.
Ricorriamo quindi alla statistica t definita dalla (5.3), di cui possiamo calcolare il valore. Una volta ottenuto, confrontiamo col competente valore critico al livello α desiderato.
I dettagli computazionali sono mostrati nel Box 5.2.
Notiamo che il valore di t è negativo; per la simmetria della distribuzione possiamo considerare il suo valore assoluto e confrontarlo con i valori critici, ottenendo $t_{0.01[23]} < |t| < t_{0.05[23]}$ (in quanto $2.807 < 2.886 < 3.767$). Secondo le convenzioni fissate in § 4. 2. 4. diciamo che il valore del t è significativo a livello 0.01, e dunque lo etichettiamo col doppio asterisco: $t = -2.886$ $**$.
In generale a seconda della posizione del t rispetto ai valori critici dei livelli standard, la simbologia è conforme a quanto illustrato in Fig. 5.5, che a sua volta è conforme alle convenzioni generali enunciate in § 4. 2. 4.

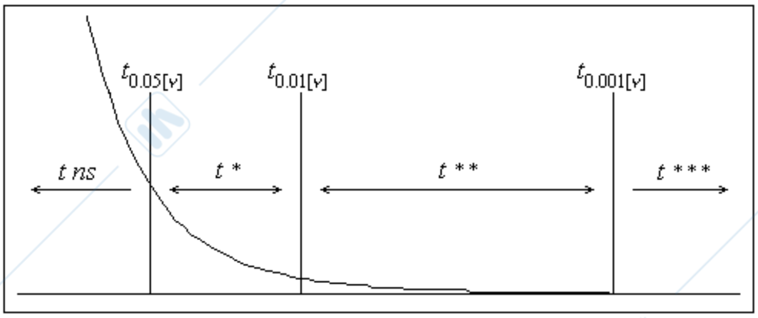

Fig. 5.5

In Fig. 5.4 è visualizzata una coda della distribuzione t, la posizione dei tre valori critici e l'opportuno simbolismo per etichettare i valori del t.

5.4. Confronto fra due gruppi di dati indipendenti

5. 4. 1. Una estensione della statistica *t*

In § 5. 1. 2. abbiamo definito la statistica t attraverso la (5.3); in essa compare a numeratore la differenza fra una statistica campionaria ($\overline{Y}$) e il corrispondente parametro (μ); a denominatore abbiamo una stima campionaria dell'errore standard di quella statistica ($s_{\overline{Y}}$).
Consideriamo ora una nuova statistica: la *differenza di due medie indipendenti* (derivate cioè da gruppi di dati che non hanno nessi di dipendenza l'uno dall'altro). Il valore campionario della statistica è ovviamente $\overline{Y}_1 - \overline{Y}_2$. Il corrispondente parametro è in modo ancora ovvio $\mu_1 - \mu_2$. Meno semplice è l'espressione per $s_{\overline{Y}_1 - \overline{Y}_2}$, la stima campionaria dell'errore standard di questa nuova statistica; l'espressione non viene dimostrata, ed è quindi da prendere per buona (sebbene in § 8. 4. 2. verrà giustificata nel caso particolare in cui $n_1 = n_2 = n$):

$$s_{\overline{Y}_1 - \overline{Y}_2} = \sqrt{\frac{(n_1 - 1)s_1^{\ 2} + (n_2 - 1)s_2^{\ 2}}{n_1 + n_2 - 2} \cdot \frac{n_1 + n_2}{n_1 \cdot n_2}} \qquad (5.8)$$

Se ora montiamo questi tre componenti in una formula con struttura analoga alla (5.3) otteniamo:

$$t=\frac{\left(\overline{Y}_1-\overline{Y}_2\right)-\left(\mu_1-\mu_2\right)}{\sqrt{\frac{(n_1-1)s_1^{\,2}+(n_2-1)s_2^{\,2}}{n_1+n_2-2}\cdot\frac{n_1+n_2}{n_1\cdot n_2}}} \qquad (5.9)$$

Si dimostra che la nuova statistica così costruita è distribuita come un *t* con $\nu = n_1 + n_2 - 2$ gradi di libertà.

5. 4. 2. Il *t* – test per il confronto fra due medie indipendenti

La (5.9) permette di sottoporre a test la differenza fra due medie campionarie indipendenti. Facciamo un esempio.
Si vuole attuare una certa sperimentazione secondo un disegno a due gruppi equivalenti (vedi § 4.2.1. della Parte metodologica). Per valutare l'equivalenza dei due gruppi viene somministrato un pre–test.
I risultati dei due gruppi sono sintetizzati come in tabella:

	Numero di elementi	*Media dei punteggi*	*Varianza dei punteggi*
Gruppo sperimentale	n_1	$\overline{Y}_1$	$s_1^{\,2}$
Gruppo di controllo	n_2	$\overline{Y}_2$	$s_2^{\,2}$

Posto che il pre–test sia strutturato in modo tale da informare in modo affidabile ed esauriente circa le abilità di partenza dei due gruppi, si tratta di valutare se questi ultimi possano essere considerati equivalenti oppure no confrontandone le due medie.
L'ipotesi nulla H_0 è quella dell'equivalenza dei gruppi ($\mu_1=\mu_2$); l'ipotesi alternativa è quella della non equivalenza ($\mu_1\neq\mu_2$). Si tratta dunque di un test a due code.
Ora si deve calcolare il valore della statistica *t* definita dalla (5.9). Si sarà notato che non disponiamo delle due medie parametriche, ma questo non è un problema in quanto se vale l'ipotesi nulla H_0 (ed è in questa ipotesi che calcoliamo la statistica *t*) la differenza $\mu_1 - \mu_2$ vale zero. Se dunque si utilizza per sottoporre a test la differenza fra due medie campionarie la (5.9) si semplifica nella seguente:

$$t=\frac{\overline{Y}_1-\overline{Y}_2}{\sqrt{\frac{(n_1-1)s_1^{\,2}+(n_2-1)s_2^{\,2}}{n_1+n_2-2}\cdot\frac{n_1+n_2}{n_1\cdot n_2}}} \qquad (5.10)$$

I dettagli del calcolo sono mostrati nel Box 5.3.
Essendo *t* significativo a livello 0.05 rifiutiamo l'ipotesi nulla e concludiamo che i due gruppi non sono equivalenti.
Prima di procedere nella sperimentazione occorrerà dunque procedere o a una ridistribuzione casuale dei soggetti fra i due gruppi oppure si potrà tentare una riassegnazione dei soggetti secondo qualche tecnica di bilanciamento (vedi § 1.4.6. della Parte metodologica). Se ciò non fosse possibile occorrerà infine dirottare su un corrispondente disegno quasi sperimentale (vedi § 5.2 della Parte metodologica).
La procedura statistica del confronto fra due medie campionarie indipendenti può naturalmente essere utilizzata anche per valutare gli esiti del post – test al termine della sperimentazione.

5. 5. Confronto fra due gruppi di dati dipendenti o appaiati

5. 5. 1. Una nuova estensione della statistica *t*

Per comprendere il senso esatto di quanto segue, il lettore deve fare riferimento al § 4.3.1. della Parte metodologica.
Partiamo da un esempio. In un disegno all'interno dei soggetti, vogliamo verificare l'efficacia di un determinato trattamento per migliorare una determinata prestazione cognitiva. A tale scopo somministriamo ad un gruppo di soggetti un pre–test prima dell'applicazione del trattamento ed un post–test al termine della sperimentazione. Dunque abbiamo a disposizione due gruppi di punteggi. Per valutare la differenza fra i punteggi prima e quelli dopo l'applicazione del trattamento sperimentale *non* è né appropriato né conveniente ricorrere al test presentato in § 5. 4.
Non appropriato: in § 5. 4. i due gruppi di dati sono *indipendenti*, in quanto prodotti da gruppi indipendenti (differenti) in contesti indipendenti l'uno dall'altro. In questa nuova situazione invece i due gruppi di dati non sono indipendenti, in quanto ogni singolo punteggio raccolto dopo il trattamento dipende dal corrispondente punteggio ottenuto dallo stesso soggetto prima del trattamento. Si ricordi invece che in § 5. 4. 1. i tre elementi che costituiscono la (5.9) sono stati calcolati sotto l'ipotesi di indipendenza dei gruppi.
Non conveniente: usando la (5.9) di fatto confrontiamo i risultati *globali* della prima prova con quelli *globali* della seconda prova. Non sfruttiamo quindi informazioni preziose che invece sono disponibili in un disegno fra i soggetti: cioè un confronto tra prima e dopo *soggetto per soggetto*. Questa perdita di informazione si traduce in una riduzione della potenza del test, come verificheremo in § 5. 5. 3.
Nel disegno sperimentale all'interno dei soggetti che abbiamo ipotizzato disponiamo di dati strutturati al modo seguente:

Soggetto	*Punteggio prima (P)*	*Punteggio dopo (D)*	*Differenza di punteggio (d)*
1	P_1	D_1	$d_1 = D_1 - P_1$
2	P_2	D_2	$d_2 = D_2 - P_2$
3	P_3	D_3	$d_3 = D_3 - P_3$
...	...	...	...
n	P_n	D_n	$d_n = D_n - P_n$

Consideriamo ora la media delle differenze dei due gruppi dipendenti *P* e *D*:

il valore campionario della statistica è $\bar{d} = \dfrac{\sum d}{n}$;

il valore del parametro corrispondente è $\delta = \mu_D - \mu_P$ (essendo μ_D e μ_P le medie parametriche dei due gruppi *P* e *D*;

l'errore standard stimato della media delle differenze è $s_{\bar{d}} = s_d / \sqrt{n}$.

Ancora una volta possiamo montare questi tre elementi nella formula

$$t = \frac{\bar{d} - \delta}{s_d / \sqrt{n}} \qquad (5.11)$$

La nuova statistica è distribuita come un *t* con $\nu = n - 1$ gradi di libertà.

5. 5. 2. Il *t* – test per dati appaiati

La statistica (5.11) permette di sottoporre a test la significatività delle differenze in un disegno sperimentale all'interno dei soggetti.

Occorre inizialmente precisare quale sia l'ipotesi nulla H_0 e l'ipotesi alternativa H_1.

L'ipotesi nulla è quella che afferma una sostanziale inefficacia del trattamento sperimentale. Se così vanno le cose non dovranno evidenziarsi sostanziali differenze, all'interno dei soggetti, fra prima e dopo l'applicazione del trattamento sperimentale. In altri termini le differenze *d* all'interno dei soggetti tenderanno ad essere nulle, così come la loro media parametrica δ. Per quanto riguarda l'ipotesi alternativa, vogliamo verificare non una generica differenza fra il prima e il dopo; se il trattamento è didatticamente efficace noi ci aspettiamo che le differenze *d* siano tendenzialmente positive. Questo conduce ad ipotizzare che la loro media parametrica δ sia positiva.

Dunque, formalizzando le due ipotesi:

H_0: $\delta = 0$

H_1: $\delta > 0$

Si tratta dunque di un test ad una sola coda.

Ricordiamo che in un test di ipotesi si vuole falsificare l'ipotesi nulla; quindi nel nostro caso la statistica *t* va calcolata nell'ipotesi di inefficacia del trattamento; ponendo dunque $\delta = 0$ la (5.11) si semplifica nella:

$$t = \frac{\bar{d}}{s_d / \sqrt{n}} \qquad (5.12)$$

dove precisiamo che con s_d indichiamo la stima campionaria della deviazione standard delle differenze *d*, e dunque con la consueta formula per la deviazione standard abbiamo:

$$s_d = \sqrt{\frac{\sum d^2 - \left(\sum d\right)^2 / n}{n - 1}} \qquad (5.13)$$

Il Box 5.4 illustra i dettagli del calcolo, tediosi ma lineari.

In particolare si osservi che trattandosi di un test ad una coda (ed essendo la **Tavola 3** a due code) i valori critici utilizzati per i livelli di significatività 0.05, 0.01 e 0.001 sono quelli corrispondenti ai livelli 0.10, 0.02 e 0.002, conformemente a quanto indicato in § 5. 1. 3.

5. 5. 3. Confronto fra i due test sulle medie campionarie

In § 5. 5. 1. abbiamo detto che se i due gruppi di dati non sono indipendenti il test per il confronto di medie è inappropriato ed inefficace.

Per chiarire questo punto fondamentale vale la pena di spendere qualche parola in più sul problema affrontato nel Box 5.4.

Supponiamo per un momento di effettuare l'analisi dei dati attraverso il *t* – test per gruppi indipendenti.

In sintesi, calcoliamo le medie e le varianze dei due gruppi, ottenendo i risultati in tabella:

	Prima	*Dopo*
n	11	11
$\overline{Y}$	39.18	47.82
s^2	209.36	204.16

I gradi di libertà sono $\nu = n_1 + n_2 - 2 = 11 + 11 - 2 = 20$.
Attraverso la (5.10) otteniamo:

$$t = \frac{\overline{Y}_1 - \overline{Y}_2}{\sqrt{\frac{(n_1 - 1)s_1^{\,2} + (n_2 - 1)s_2^{\,2}}{n_1 + n_2 - 2} \cdot \frac{n_1 + n_2}{n_1 \cdot n_2}}} = \frac{47.82 - 39.18}{\sqrt{\frac{10 \cdot 209.36 + 10 \cdot 204.16}{11 + 11 - 1} \cdot \frac{11 + 11}{11 \cdot 11}}} = 1.409 \; ns$$

Come mai il risultato di questo test è radicalmente diverso da quello appropriato per dati appaiati? La circostanza può essere ben spiegata confrontando Fig. 5.6 e Fig. 5.7:

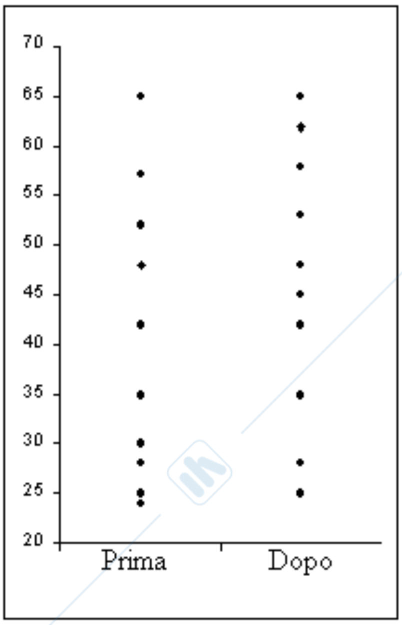

Fig. 5.6

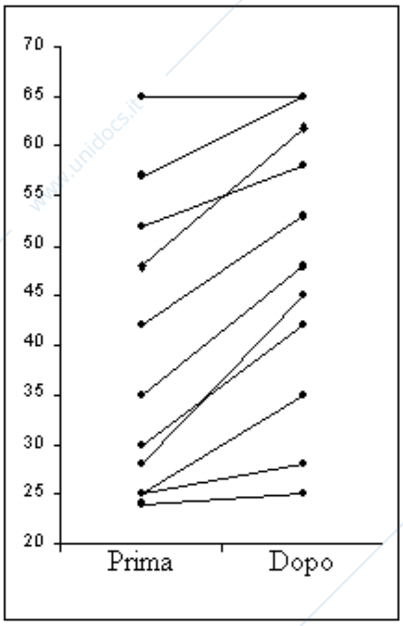

Fig. 5.7

In Fig. 5.6 i punteggi degli studenti sono rappresentati da punti posti ad una altezza corrispondente ai loro valori numerici; i punti incolonnati a sinistra riguardano l'esito del test prima del trattamento, mentre quelli a destra riguardano i punteggi del test effettuato dopo. Il colpo d'occhio che abbiamo presenta due gruppi di dati distribuiti in modo piuttosto omogeneo con medie forse non uguali ma comunque non troppo dissimili. Questa rappresentazione mostra i due gruppi di dati in modo globale. Le sensazioni soggettive che abbiamo descritto sopra sono confermate dai numeri prodotti nei calcoli del test per dati indipendenti (varianze simili, medie leggermente differenti, t non significativo).
In Fig. 5.7 è ripresa la rappresentazione precedente, ma è aggiunta l'informazione decisiva che collega ogni dato della prima colonna con il corrispondente nella seconda. Questo modo di vedere i dati è profondamente differente dal primo e l'informazione che ne deriva è più ricca: qui vediamo chiaramente che c'è un netto miglioramento, evidenziato dai tratti obliqui che sono tutti ascendenti. Non a caso dunque il test del Box 5.4 ha dato esito altamente significativo.
Sotto un profilo più strettamente statistico i due test analizzano i dati sotto due aspetti differenti, espressi chiaramente dalle rispettive ipotesi nulla e alternativa: nel caso di test per dati indipendenti si indaga sulla differenza fra le due medie μ_1 e μ_2 (quindi, ancora una volta, sui gruppi analizzati globalmente), mentre nel secondo caso si valuta la differenza media δ fra i dati del primo gruppo ed i corrispondenti del secondo.

Box 5.1. Limiti di confidenza della media parametrica
Punteggi grezzi di una prova oggettiva di verifica (dati dal Box 3.2):

24	25	11	38	34	28	27	22	11	21

Nel Box 3.2 abbiamo ottenuto:

$$n = 10$$

$$s = 8.65$$

da cui, per la (5.2) ricaviamo:

$$s_{\bar{Y}} = s/\sqrt{n} = 8.65/\sqrt{10} = 2.735$$

Per la media campionaria otteniamo:

$$\bar{Y} = \frac{24 + 25 + \cdots + 11 + 21}{10} = 24.1$$

Per i gradi di libertà della statistica abbiamo:

$$\nu = n - 1 = 9$$

Limiti di confidenza al 95% (α = 0.05)

Dalla **Tavola 3** ricaviamo:

$$t_{0.05[9]} = 2.262$$

Quindi, utilizzando le (5.7):

$$L_1 = \bar{Y} - t_{\alpha[\nu]} \cdot s_{\bar{Y}} = 2.41 - 2.262 \cdot 2.735 = 17.91$$

$$L_1 = \bar{Y} - t_{\alpha[\nu]} \cdot s_{\bar{Y}} = 2.41 + 2.262 \cdot 2.735 = 30.28$$

Limiti di confidenza al 99% (α = 0.01)

Dalla **Tavola 3** ricaviamo:

$$t_{0.01[9]} = 3.250$$

Quindi, utilizzando le (5.7):

$$L_1 = \bar{Y} - t_{\alpha[\nu]} \cdot s_{\bar{Y}} = 2.41 - 3.250 \cdot 2.735 = 15.21$$

$$L_1 = \bar{Y} - t_{\alpha[\nu]} \cdot s_{\bar{Y}} = 2.41 + 3.250 \cdot 2.735 = 32.98$$

Limiti di confidenza al 99.9% (α = 0.001)

Dalla **Tavola 3** ricaviamo:

$$t_{0.001[9]} = 4.781$$

Quindi, utilizzando le (5.7):

$$L_1 = \bar{Y} - t_{\alpha[\nu]} \cdot s_{\bar{Y}} = 2.41 - 4.781 \cdot 2.735 = 11.02$$

$$L_1 = \bar{Y} - t_{\alpha[\nu]} \cdot s_{\bar{Y}} = 2.41 + 4.781 \cdot 2.735 = 37.17$$

Rif.:
§ 5. 2. 2.

Box 5.2. Confronto di una media parametrica con un campione (test a due code)
Dati relativi alla somministrazione di un reattivo in due popolazioni:

Popolazione 1: $\mu = 78.12$

Popolazione 2: $n = 24$, $\overline{Y} = 72.11$, $s = 10.20$

H_0: La Popolazione 1 e la Popolazione 2 hanno la stessa media parametrica
H_1: La Popolazione 1 e la Popolazione 2 hanno medie parametriche diverse

$$\nu = n - 1 = 23$$

$$s_{\overline{Y}} = s/\sqrt{n} = 10.20/\sqrt{24} = 2.082$$

$$t = \frac{\overline{Y} - \mu}{s_{\overline{Y}}} = \frac{72.11 - 78.12}{2.082} = -2.886 \;**$$

$$t_{0.05[23]} = 2.069$$

$$t_{0.01[23]} = 2.807$$

$$t_{0.001[23]} = 3.767$$

Conclusioni:
essendo t significativo a livello 0.01 rifiuto H_0 e concludo che le due popolazioni hanno medie parametriche differenti.

Rif.:
§ 5. 3.

Box 5.3. *t* – test per il confronto di due medie campionarie (test a due code)
Dati relativi ai punteggi della prova oggettiva iniziale per verificare l'equivalenza dei due gruppi coinvolti nella sperimentazione.

	n	$\bar{Y}$	s^2
Gruppo sperimentale (1)	15	34.5	10.2
Gruppo di controllo (2)	12	31.1	11.4

H_0: $\mu_1 = \mu_2$
H_1: $\mu_1 \neq \mu_2$

$$\nu = n_1 + n_2 - 2 = 15 + 12 - 2 = 25$$

$$t = \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{\frac{(n_1 - 1)s_1^{\,2} + (n_2 - 1)s_2^{\,2}}{n_1 + n_2 - 2} \cdot \frac{n_1 + n_2}{n_1 \cdot n_2}}} = \frac{34.5 - 31.1}{\sqrt{\frac{14 \cdot 10.2 + 11 \cdot 11.4}{15 + 12 - 2} \cdot \frac{15 + 12}{15 \cdot 12}}} = \frac{3.4}{\sqrt{1.6092}} = 2.680 \ *$$

$$t_{0.05[25]} = 2.060$$
$$t_{0.01[25]} = 2.787$$
$$t_{0.001[25]} = 3.725$$

Conclusioni:
essendo t significativo a livello 0.05 rifiuto H_0 e concludo che i due gruppi non sono equivalenti.

Rif.:
§ 5. 4. 2.

Box 5.4. *t* – test per dati appaiati (test a una coda)
Punteggi grezzi di 11 studenti in un test prima e dopo un trattamento didattico sperimentale:

Studenti	*Prima*	*Dopo*	d	d^2
Studente 1	24	25	1	1
Studente 2	25	28	3	9
Studente 3	25	35	10	100
Studente 4	28	45	17	289
Studente 5	30	42	12	144
Studente 6	35	48	13	169
Studente 7	42	53	11	121
Studente 8	52	58	6	36
Studente 9	57	65	8	64
Studente 10	48	62	14	196
Studente 11	65	65	0	0
			95	1129

H_0: Trattamento didattico inefficace; media delle differenze individuali nulla: $\delta = 0$
H_1: Trattamento didattico efficace; media delle differenze individuali positiva: $\delta > 0$
(test a una coda)

$$\nu = n - 1 = 11 - 1 = 10$$

$$\bar{d} = \frac{\sum d}{n} = \frac{95}{11} = 8.636$$

$$s_d = \sqrt{\frac{\sum d^2 - \left(\sum d\right)^2 / n}{n-1}} = \sqrt{\frac{1129 - 95^2/11}{10}} = 5.555$$

$$t = \frac{\bar{d}}{s_d/\sqrt{n}} = \frac{8.636}{5.555/\sqrt{11}} = \frac{8.636}{1.675} = 5.156 \;\; ***$$

dalla **Tavola 3** dei valori critici del *t* a due code:

$$t_{0.10[10]} = 1.812$$

$$t_{0.02[10]} = 2.764$$

$$t_{0.002[10]} = 4.144$$

Conclusioni:
il valore di *t* è significativo a livello 0.001; rifiuto l'ipotesi nulla; la media delle differenze individuali è positiva; è provata l'efficacia del trattamento didattico.

Rif.:
§ 5. 5. 2.