ALCUNE RIFLESSIONI SULL'UTILITA' DELLA STATISTICA NELLA RICERCA E NEL CONTROLLO

Franco Taggi
Lab. Epidemiologia e Biostatistica, Istituto Superiore di Sanità

La Statistica è la Scienza nella quale, partendo da premesse dubbie, si imbastisce un ragionamento esatto per giungere ad un falsa conclusione che deve poi essere interpretata in base ad improbabili fattori di confronto.

(Anonimo)

... me spiego: da li conti che se fanno
secondo le statistiche d'adesso
risulta che te tocca un pollo all'anno;
e, se non c'entra nelle spese tue
t'entra ne la statistica lo stesso
perché c'è un antro che ne magna due.
(Trilussa, "La statistica")

A proposito di statistica...

Nonostante i successi ottenuti dalla Statistica negli ultimi cento anni, successi che hanno praticamente riguardato tutti i campi dello scibile, l'opinione che l'uomo della strada ha ancora oggi di questa disciplina è ben riassunta dal pensiero dell'Anonimo e dal frammento di sonetto del poeta Trilussa: la statistica è un imbroglio, un pasticcio, qualcosa che serve a dimostrare tutto e il contrario di tutto, un marchingegno usato da persone furbastre per convincere la gente che il bianco è nero e il nero è bianco. Tutto questo è falso: la statistica è un eccellente strumento che, se usato correttamente, aiuta a comprendere quanto effettivamente conosciamo di un fenomeno, dandoci talora preziosi suggerimenti sulla struttura latente del fenomeno stesso. La dichiarazione dell'Anonimo (anonimo perché, lo confesso, non ricordo il nome dell'autore della citazione...) è piena di livore ed inesatta: pura calunnia, anche se scherzosa; i versi di Trilussa, sia pur dissacratori, potrebbero invece avere un senso profondo se il loro scopo fosse quello di mettere in guardia contro idilliache conclusioni che possono essere tratte da valutazioni basate su piccoli campioni. In questo, Trilussa si affiancherebbe idealmente a von Mises, grande statistico del novecento che molto si è occupato di piccoli campioni e che non ha mai mancato occasione per allertare i ricercatori sulla pericolosità di estrapolazioni basate su un numero troppo limitato di dati.
Questa visione, a dir poco scettica, delle possibilità offerte dalla Statistica, tuttavia, non è propria soltanto dell'uomo della strada: anche nella tecnica e nella ricerca, la Statistica è ancor oggi da non pochi percepita come una medicina amara, la cui utilità è primariamente quella di permettere la pubblicazione di un articolo (altrimenti, la rivista non lo accetta...) o di sintetizzare i dati, principalmente per facilitarne la lettura.

In questo mio intervento mi propongo di mostrare, discutendo alcuni aspetti di interesse della materia, come la Statistica sia simile ad una fanciulla di belle forme e buon carattere che, per la sua modestia e riservatezza, vien poco notata; una ragazza "difficile", ma che conosciuta più da vicino può fare innamorare profondamente.
Quanto nel seguito trattato avrà carattere prettamente euristico e riguarderà l'intervallo di confidenza di una media, due test statistici non parametrici per l'analisi di due campioni ed un criterio di dimensionamento di un campione. Nello sviluppare questi temi, si è ipotizzato che il lettore abbia un minimo di familiarità con le tecniche statistiche.

Statistica induttiva e calcolo delle probabilità

La statistica induttiva e il calcolo delle probabilità sono, per molti versi, speculari. Nel calcolo delle probabilità, infatti, noi conosciamo (supponiamo di conoscere...) perfettamente la struttura del fenomeno in studio e cerchiamo di valutare quanto verosimilmente possiamo attenderci il verificarsi di certi eventi. Ad esempio, immaginiamo di lanciare un dado "onesto" (ovvero, un dado non truccato, in cui ogni faccia abbia la stessa probabilità di presentarsi: si osservi che in questo modo noi stiamo dichiarando di conoscere perfettamente la realtà, cioè la struttura interna del fenomeno studiato): quanto è verosimile ottenere in 7 lanci consecutivi sette volte "quattro"?
Dunque, in questo caso noi sappiamo tutto del dado e diamo per scontato che il lancio non venga influenzato da una particolare destrezza del giocatore: ci chiediamo cosa avverrà quando il dado verrà lanciato un certo numero di volte. Nella statistica induttiva, di contro, noi conosciamo i risultati di certe prove e da queste cerchiamo di immaginarci quale struttura può ragionevolmente essere sottesa dall'osservato. Riferendoci all'esempio del dado, conoscendo un certo numero di risultati di lanci successivi, abbiamo elementi obiettivi per sospettare che il dado sia truccato?
A conti fatti, lo studioso di calcolo delle probabilità conosce la struttura della popolazione (del fenomeno) e cerca di assegnare delle probabilità ai possibili campioni; lo studioso di statistica induttiva conosce, invece, delle osservazioni su campioni e da queste cerca di inferire le caratteristiche della popolazione da cui questi campioni provengono (Fig 1).

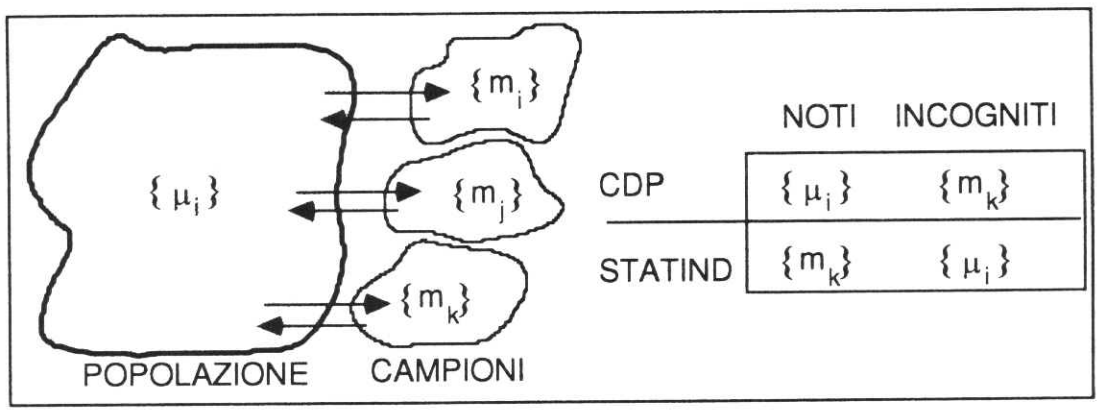


Figura 1 - *Relazioni tra la statistica induttiva e il calcolo delle probabilità.*

In conclusione, il calcolo delle probabilità ci dice:"Il mondo è fatto così; e, se il mondo è questo, allora i fatti che ne risulteranno dovranno essere sostanzialmente questi"; la statistica induttiva, viceversa, afferma:"I fatti sono questi; e, allora, se i fatti sono questi, il mondo da cui essi provengono dovrebbe, più o meno, essere questo". Come queste affermazioni possano diventare enunciati utilizzabili per prendere decisioni convenienti, avremo modo di vederlo nel seguito considerando, come esempio, la variabilità delle medie.

La variabilità delle medie

Studiando come variano le medie di campioni diversi estratti da una stessa popolazione secondo una procedura casuale, si comprenderà meglio quanto detto e le importanti conseguenze che da questo approccio discendono. Immaginiamo, allo scopo, di rilevare una certa caratteristica (misurata su una scala almeno intervallare) sui soggetti che fanno parte di un nostro campione, che supponiamo sia stato estratto da una popolazione normale di media 100 e deviazione standard 10 (tralasciamo, ora e nel seguito, le unità di misura).
(N.B.: si osservi che, nella realtà, noi non conosciamo né la media né la deviazione standard della popolazione; anzi, un esperimento viene in genere eseguito per stimare questi parametri; è possibile, tuttavia, generare, con opportuni algoritmi, delle misure casuali provenienti da una distribuzione normale di media e deviazione standard prefissate. I dati degli esempi che seguiranno sono di questo tipo: il vantaggio della simulazione è che in questo modo noi conosciamo sia i risultati del campione sia i parametri della popolazione da cui il campione proviene. Consideriamo, quindi, attentamente il vantaggio che ci offre questa tecnica, ma non dimentichiamoci che, all'atto pratico, conosciamo solo i dati campionari e, al più, qualche parametro della popolazione).
Siano i soggetti esaminati in numero di 5 (dunque, il campione è di numerosità 5). Avremo per la caratteristica X misurata:

no. soggetto	Valori di X
1	92.5
2	101.1
3	85.4
4	105.0
5	95.7

Su questi dati possiamo calcolarci la media e la deviazione standard campionarie. Sarà:

$m_1 = 95.9$ e $s_1 = 7.6$.

Ora, poiché noi non sappiamo che la media e la deviazione standard della popolazione sono rispettivamente 100 e 10, avremo che le nostre migliori stime per questi due parametri saranno 95.9 per la media e 7.61 per la deviazione standard.
D'altra parte, è chiaro che se estrarremo dalla popolazione un nuovo campione di 5 soggetti non otterremo quasi mai valori di media e deviazione standard campionarie coincidenti con quelli precedentemente ottenuti. Infatti, i risultati ottenuti con un secondo campione, come può vedersi appresso, forniscono ora i seguenti valori:

no. soggetto	Valore di X
6	79.5
7	94.7
8	112.6
9	89.9
10	96.9

$m_2 = 94.7$ e $s_2 = 12.04$.

Prendiamo un terzo campione di cinque soggetti:

no. soggetto	Valore di X
11	114.6
12	100.1
13	108.2
14	109.4
15	94.5

Avremo ora:
$m_3 = 105.4$ e $s_3 = 7.99$.

Dunque, tre esperimenti simili, tre risultati diversi. Ma, allora, cosa si può dire, con queste diverse stime, della media e della deviazione standard della popolazione dalla quale provengono i campioni esaminati (che noi sappiamo, per costruzione, essere rimasta invariata)?
Per rispondere a questo non banale quesito, simuliamo ulteriori esperimenti. I risultati di queste nuove osservazioni sono riportati, con i precedenti, qui appresso:

n° campione	Valori di X osservati sui 5 soggetti				
1	92.5	101.1	85.4	105.0	95.7
2	79.5	94.7	112.6	89.9	96.9
3	114.6	100.1	108.2	109.4	94.5
4	109.4	84.1	108.6	114.8	110.3
5	98.7	94.2	102.3	90.7	89.5
6	82.8	100.9	110.0	107.4	102.7
7	80.4	95.2	115.9	100.1	95.2
8	99.3	87.6	112.5	90.8	79.5
9	94.2	92.6	114.7	118.9	102.8
10	92.6	93.5	97.7	93.8	103.4

n° campione	media	deviazione standard
1	95.9	7.61
2	94.7	12.04
3	105.4	7.99
4	105.3	12.17
5	95.1	5.39
6	100.8	10.68
7	97.4	12.73
8	93.9	12.57
9	104.6	4.48
10	96.2	4.48

Come si rileva facilmente, le medie campionarie continuano ad essere sempre diverse e nessuna di queste, peraltro, coincide con la media della popolazione, anche se, per caso qualche media campionaria potrebbe venire esattamente pari a 100: di certo, tuttavia, i valori ottenuti oscillano intorno al valore della media della popolazione.

C'è, inoltre, un ulteriore risultato notevole: se si osservano le 10 medie ricavate dai campioni studiati, queste appaiono meno disperse che non le singole misure. In altre parole, le medie sembrano fluttuare meno delle misure.

Se calcoliamo la media delle 10 medie e la corrispondente deviazione standard, otteniamo $m_m = 98.9$ e $s_m = 4.67$, ed il valore ottenuto per la deviazione standard delle medie conferma l'impressione che si aveva da un'ispezione "ad occhio" dei dati.

Il risultato ottenuto in questa nostra simulazione è di carattere generale: può dimostrarsi, infatti, che se i dati provengono da una popolazione normale di media μ e deviazione standard σ, allora le medie di campioni n-numerosi seguono una distribuzione normale di media μ e di deviazione standard $\sigma_m = \frac{\sigma}{\sqrt{n}}$.

Quindi, nel caso del precedente, dove $\sigma = 10$ e $n = 5$, in base a questo teorema dovremo aspettarci una distribuzione delle medie con media pari a 100 e deviazione standard pari a $\frac{10}{\sqrt{5}} = 4.472$, valori in buon accordo con quelli osservati (98.9 e 4.67).
La deviazione standard della distribuzione delle medie campionarie si chiama solitamente "errore standard della media".
Come è intuitivo, più i campioni sono numerosi, meno fluttuano le medie che da questi si ottengono (al crescere di n la distribuzione delle medie dei campioni è sempre più concentrata intorno alla media della distribuzione d'origine).
Per esempio, se ripetessimo l'esperimento con campioni di numerosità 3 dovremmo attenderci un errore standard pari a 5.77; mentre, con campioni di numerosità 15 avremmo un valore di 2.58.
Dunque, più è grande il campione, migliore è la nostra conoscenza della media della popolazione. Questo risultato, a pensarci bene, è intuitivo: più grande è il numero di elementi che costituiscono il campione, maggiore è l'informazione di cui disponiamo per farci un'idea della media della popolazione da cui il campione proviene. Notevole è, invece, il fatto che la dispersione delle medie dei campioni sia connessa a quella della popolazione tramite una relazione molto semplice, quale la $\sigma_m = \frac{\sigma}{\sqrt{n}}$ Oltre a questo, c'è un altro fatto notevole, assolutamente non intuitivo: sotto condizioni molto generali, quale che sia la distribuzione d'origine dei campioni, al crescere della loro numerosità, la distribuzione delle medie tende ad essere sempre normale. Questo fatto è davvero sorprendente, inatteso. Si rifletta attentamente sulla sua portata: ad esempio, prendendo campioni da una distribuzione asimmetrica, per n abbastanza grande, la distribuzione delle medie campionarie sarà simmetrica (e, come detto, normale).
Al fine di sottolineare l'importanza di quanto finora discusso, e di mettere a frutto le sorprendenti caratteristiche ora segnalate, introdurremo un concetto fondamentale sia dal punto di vista teorico, sia applicativo: l'intervallo di confidenza di una media.

L'intervallo di confidenza di una media

La media aritmetica $m = \frac{\sum_{i=1}^{n} x_i}{n}$ è, sotto molti punti di vista, la migliore stima che si può avere della media incognita μ della popolazione da cui proviene il campione. Una stima di questo genere viene detta "puntuale", perchè è rappresentata da un solo valore.
I limiti di questo approccio sono evidenti. Si supponga, infatti, di avere tre ricercatori, ugualmente bravi ed attrezzati, che studiano uno stesso fenomeno: il dott. Tizio esegue 10 misure e ottiene una media pari a 7.2; il prof. Caio ne esegue 50 e trova un valor medio ancora pari a 7.2; l'ing. Sempronio, ricco di personale e risorse, ottiene anch'egli una media di 7.2, ma effettuando ben 1000 misure. Se i tre studiosi riferiscono i loro risultati ad un congresso cui siamo presenti, non potremo che prendere atto che la media

del fenomeno studiato è intorno a 7.2, ma di certo la nostra fiducia sarà massima per il risultato di Sempronio e minima per quello di Tizio. Tuttavia, se non troviamo modo di includere nella valutazione della media l'informazione "numerosità delle prove effettuate" (che, si noti bene, a parità di tutte le altre condizioni, determina il nostro grado di fiducia), i tre risultati saranno indistinguibili tra loro. D'altra parte, abbiamo visto che la numerosità del campione "pilota" la dispersione delle medie, nel senso che più i campioni sono numerosi più è "stretta" la relativa distribuzione delle medie: addirittura, se la numerosità del nostro campione fosse infinita (o, nel caso finito, pari al numero delle unità statistiche che costituiscono la popolazione indagata), avremmo una conoscenza completa del valore μ (la nostra stima puntuale sarebbe esatta, ed ogni campione di tale numerosità darebbe sempre come risultato μ , essendo nulla la variabilità delle medie); se il nostro campione fosse invece costituito da una sola misura, la nostra stima puntuale della media incognita sarebbe data proprio dalla misura stessa e la dispersione della distribuzione delle medie provenienti da campioni uno-numerosi sarebbe quella della popolazione. Tra questi due estremi abbiamo infinite distribuzioni delle medie (o tante quant'è la numerosità della popolazione, se questa è finita), sempre più strette all'aumentare di n, secondo la formula già vista.

C'è da chiedersi, dunque, se non sarebbe possibile affiancare la nostra stima puntuale con un intervallo di possibili valori di stima della media incognita, intervallo in qualche modo relato alla numerosità del campione (nel seguito lo chiameremo "intervallo di confidenza").

Dato che si è visto che la distribuzione delle medie ricavate da campioni n-numerosi è normale e la sua deviazione standard è data da $\sigma_m = \frac{\sigma}{\sqrt{n}}$, se fossero noti μ, σ ed n potremmo costituire intervalli del tipo $\mu \pm \frac{\sigma}{\sqrt{n}}$, $\mu \pm 2\frac{\sigma}{\sqrt{n}}$,ecc., dove cadrebbero rispettivamente il 68.4%, il 95.4%, e così via, delle medie.

In termini più teorici che intuitivi, nel caso dell'intervallo fiduciale di una media di una distribuzione normale $N(\mu, \sigma)$, quando sia nota la deviazione standard, può essere così sintetizzato:

supponiamo di avere a che fare con una $N(\mu, \sigma)$ di cui si conosca σ ma non μ. Se si estrae da questa distribuzione un campione rappresentativo di n elementi $\{x_i$, potremo calcolarci la media campionaria $m = \sum_{i=1}^{n} \frac{x_i}{n}$. Ora, poiché sappiamo che la distribuzione delle medie di campioni n-numerosi provenienti da una $N(\mu, \sigma)$ è anch'essa normale e del tipo $N(\mu, \frac{\sigma}{\sqrt{n}})$, questo comporterà che la variabile standardizzata $\frac{m-\mu}{\sigma/\sqrt{n}}$ segua la $N(0,1)$ e soddisfi la relazione:

$$\Pr(-z_{\alpha/2} < \frac{m-\mu}{\sigma/\sqrt{n}} < z_{\alpha/2}) = 1-\alpha \qquad (1).$$

Si osservi che la precedente espressione è una dichiarazione in termini di calcolo delle probabilità perché, noti μ, σ e n, ci porta ad affermare qualcosa sulle caratteristiche di un possibile campione (e cioè che la media osservata sul campione, diminuita della media della popolazione e divisa per l'errore standard, fornirà con una probabilità pari a $(1-\alpha)$ un valore compreso tra $\pm z_{\alpha/2}$.

Sviluppiamo ora, su un caso particolare, un'interessante riformulazione della (1).

Come si ricorderà, data la normale standardizzata N(0,1), è $z_{0.025}=1.96$. Allora, per la (1) sarà:

$$\Pr(-1.96 < \frac{m-\mu}{\sigma/\sqrt{n}} < 1.96) = 0.95 \qquad \text{(1 bis)}.$$

La prima delle due disequazioni tra parentesi, risolta rispetto a μ, fornisce come risultato $m + 1.96\frac{\sigma}{\sqrt{n}} > \mu$; la seconda, analogamente, $m - 1.96\frac{\sigma}{\sqrt{n}} < \mu$. La (1 bis) diviene perciò

$$\Pr(m - 1.96\frac{\sigma}{\sqrt{n}} < \mu < m + 1.96\frac{\sigma}{\sqrt{n}}) = 0.95$$

Nel caso generale, la precedente espressione diventa

$$\Pr(m - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < m + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}) = 1-\alpha \qquad (2)$$

La (2) è particolarmente interessante perché mette in relazione un elemento aleatorio (il parametro m, che varia in funzione del particolare campione estratto) con il parametro incognito μ che si desidera stimare. Nei fatti, la (2) fornisce la probabilità che l'intervallo stocastico $m \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ contenga la media incognita μ della distribuzione normale considerata. Questa espressione, dunque, non formula previsioni su un risultato campionario, ma dal risultato campionario inferisce qualcosa di relativo ad un parametro della popolazione da cui il campione proviene.

ESEMPIO: sappiamo che un certo fenomeno si distribuisce normalmente con deviazione standard pari a 5. Abbiamo effettuato sette determinazioni, ottenendo i seguenti valori: 21.1, 16.5, 12.9, 28.8, 23.1, 19.7, 14.1 (in realtà, questi valori sono stati ottenuti per simulazione da una distribuzione $N(20,5)$). Cosa possiamo dire della media della popolazione da cui il campione proviene?

In primo luogo, la nostra stima puntuale sarà data dalla media aritmetica, pari a 19.44; scegliendo poi come livello di rischio il 5% (ovvero, il 95% di probabilità), faremo riferimento al quantile della distribuzione normale standardizzata z=1.96. La quantità da sommare e da sottrarre alla media campionaria sarà perciò: $.96\frac{5}{\sqrt{7}} = 3.70$.

Avremo, così: 19.44 - 3.70 = 15.74; 19.44 + 3.70 = 23.14. Possiamo dunque affermare, con un rischio del 5%, che la media incognita della popolazione è contenuta nell'intervallo di confidenza 15.74 +---+ 23.14.

Si osservi che la (2) ci assicura che al crescere della numerosità del campione da cui proviene la media empirica m, la nostra conoscenza del parametro incognito diventa sempre più buona, tendendo al vero valore di μ per $n \to \infty$.

Con questa strategia, tuttavia, non risolviamo il problema in generale perchè il metodo visto è applicabile soltanto quando si conosca la vera dispersione della distribuzione oggetto del nostro interesse; e, in genere, nelle situazioni più usuali, σ non è noto, ma stimato dal campione.

Come procedere, allora? Il problema non è semplice, è di quelli che a prima vista somigliano ad un serpente che si morde la coda. Questo deve aver pensato Gossett, un geniale statistico che nei primi anni del '900 ha attentamente considerato la questione. Il suo ragionamento è il seguente: supponiamo di conoscere la distribuzione di una certa quantità che dipenda dal parametro incognito della popolazione e dalla media e deviazione standard del campione: se conosco tale distribuzione, avrò certamente gli elementi per definire un intervallo di confidenza della media vera incognita, trasformando una relazione probabilistica, tipo la (1), in un'affermazione inferenziale, come la (2).

Gossett ha scelto, per questa sua ricerca, la variabile statistica $t_{oss} = \frac{m - \mu}{\frac{s}{\sqrt{n}}}$, dove m è la media campionaria, μ la media incognita della popolazione, S la deviazione standard stimata sul campione ed n il numero di elementi costituenti il campione stesso. La t_{oss}, segue una distribuzione particolare, simmetrica e di media nulla, che dipende da n, detta oggi "t di Student", dallo pseudonimo sotto il quale, descrivendola, Gossett pubblicò il suo lavoro fondamentale.

Per comprendere meglio il concetto di intervallo di confidenza, quando non conosciamo la vera dispersione della popolazione, sarà opportuno simulare un esperimento.

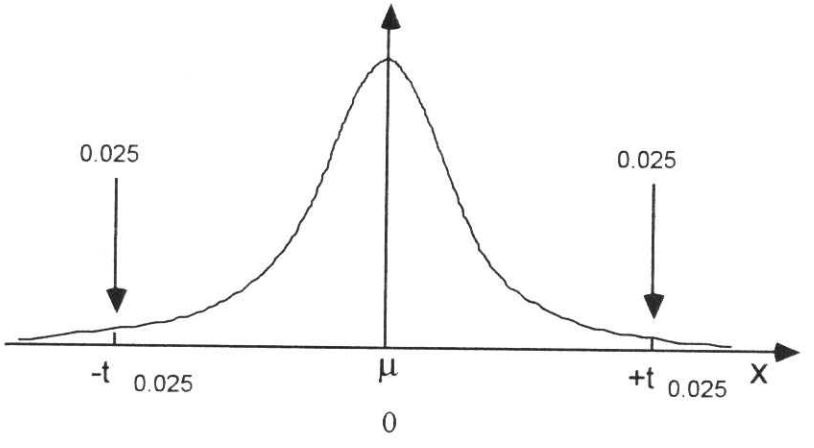
Immaginiamo che la popolazione da cui si estrae il campione sia distribuita normalmente con media 50 e deviazione standard 7. Fissiamo la numerosità del campione pari a 5 elementi. Ora, dato un certo campione, potremo calcolarci il valore di t_{oss} dalla formula precedente (N.B.: il pedice "oss" sta a ricordarci che si tratta del valore di t osservato sul nostro campione).

La tabella seguente riporta i risultati ottenuti su dieci campioni di cinque elementi, di cui si è calcolata la media, la deviazione standard e il valore di t_{oss}:

Campione	Elementi del campione					m	S	t_{oss}
1	54.7	43.9	38.0	51.9	51.8	48.06	6.916	-0.627
2	53.6	56.3	55.4	36.4	49.7	50.28	8.162	0.077
3	47.9	55.0	47.7	52.7	40.4	48.74	5.617	-0.502
4	64.9	38.8	39.7	58.5	59.9	52.36	12.206	0.432
5	49.3	43.8	53.7	50.1	56.7	50.72	4.872	0.330
6	55.5	40.3	52.4	49.3	48.9	49.28	5.683	-0.283
7	49.6	54.3	49.1	50.5	49.8	50.66	2.096	0.704
8	44.4	42.9	37.3	61.7	45.3	46.32	9.142	-0.900
9	46.4	45.8	55.1	40.5	56.4	48.84	6.729	0.385
10	48.1	48.7	48.3	51.2	67.3	52.72	8.245	0.738

Come si osserva, i diversi campioni hanno prodotto diversi valori di t_{oss}, valori che sembrano distribuiti abbastanza simmetricamente rispetto allo zero Ora, se immaginiamo di prelevare un numero grandissimo di campioni di numerosità 5, potremo costruirci empiricamente la distribuzione della variabile t_{oss} per n=5.

Chiamiamo ora con $t_{0.025}$ il quantile di t_{oss} (ovvero della distribuzione t di Student, che la t_{oss} segue) che lascia fuori il 5% della distribuzione (si veda la figura seguente).



Per costruzione sarà allora:

$$\Pr ob\left(\frac{|m-\mu|}{\frac{s}{\sqrt{n}}} < t_{0.025}\right) = 0.95$$

, cioè:

$$\Pr ob\left(|m-\mu| < t_{0.025}\cdot\frac{s}{\sqrt{n}}\right) = 0.95$$

il che comporta, analogamente a quanto visto in precedenza:

$$\Pr ob\left(m - t_{0.025}\cdot\frac{s}{\sqrt{n}} < \mu < m + t_{0.025}\cdot\frac{s}{\sqrt{n}}\right) = 0.95$$

Quest'ultimo risultato ci assicura che il 95% delle volte l'intervallo $m \pm t_{0.025}\cdot\frac{s}{\sqrt{n}}$ conterrà la media incognita della popolazione (e che, quindi, solo il 5% delle volte tale intervallo non la conterrà: questo 5% sarà, dunque, il rischio di sbagliare in cui incorreremo affermando che l'intervallo di confidenza così calcolato continene la media incognita μ). Nota la distribuzione di t , il problema è dunque risolto. Il grande merito di Gossett è stato proprio quello di determinare tale distribuzione.

E' bene riflettere su questo risultato, perchè in esso è contenuta gran parte della logica della statistica induttiva.: noi non conosceremo mai perfettamente μ, ma non sarà insensato affermare che questo valore è verosimilmente contenuto tra i due limiti calcolati con la procedura vista.

Si è visto che t dipende da n: nota la struttura della distribuzione t, mediante tavole od opportuni algoritmi, è possibile ricavare i quantili che ci interessano (ovvero, i quantili che ci permetteranno di affermare che l'intervallo calcolato continene la media incognita, sapendo di sbagliare in questa affermazione un certo numero di volte su cento) e determinare, quindi, gli intervalli di confidenza adeguati alle nostre esigenze (se accetteremo il rischio di sbagliare una volta su dieci, sceglieremo il quantile $t_{0.05}$; se una volta su cento, il quantile $t_{0.005}$; ecc. ecc.).

ESEMPIO: sappiamo che un certo fenomeno si distribuisce normalmente, ma non sappiamo con quale dispersione. Abbiamo effettuato sette determinazioni, ottenendo i seguenti valori: 21.1, 16.5, 12.9, 28.8, 23.1, 19.7 e 14.1, cui corrispondono una media m=19.44 e una deviazione standard s=5.522. Cosa possiamo dire della media della popolazione da cui il campione proviene? In primo luogo, la nostra stima puntuale sarà data dalla media aritmetica, pari a 19.44; scegliendo poi come livello di rischio il 5% (ovvero, il 95% di probabilità), faremo riferimento al quantile della distribuzione t di Student con (7-1) gradi di libertà, $t_{6,0.025} = 2.446$. La quantità da sommare e da sottrarre alla media campionaria sarà perciò: $2.4469 \frac{5.522}{\sqrt{7}} = 5.11$.

Avremo, così: 19.44 - 5.11 = 14.33; 19.44 + 5.11 = 24.55.

Possiamo dunque affermare, con un rischio del 5%, che la media incognita della popolazione è contenuta nell'intervallo di confidenza 14.33 +---+ 24.55.

Calcoliamoci ora, come ulteriore esercizio, gli intervalli di confidenza al 5% di rischio (95% di probabilita') sui 10 campioni precedentemente estratti, ricordando che $n = 5$ e sapendo che è $t_{4,0.025} = 2.7764$.

n° campione	intervallo fiduciale della media (p=0.95)
1	39.5 - 56.7
2	40.1 - 60.4
3	41.8 - 55.7
4	37.2 - 67.5
5	44.7 - 56.8
6	42.2 - 56.3
7	48.1 - 53.3
8	35.0 - 57.7
9	40.5 - 57.2
10	42.5 - 63.0

Come può osservarsi, tutti gli intervalli calcolati contengono il valore della media della distribuzione di origine dei campioni ($\mu = 50$). D'altra parte, il rischio scelto (5%,

ovvero $0.05 = \frac{1}{20}$) è tale che, mediamente, solo un campione su 20 darà luogo ad un intervallo che non conterrà il valore incognito della media della popolazione.
In altre parole, se un esperimento analogo a quello del nostro esercizio venisse eseguito da 20 sperimentatori, ognuno di essi, calcolando l'intervallo di confidenza della media incognita, affermera': "Non conosco la media della popolazione da cui proviene il mio campione, ma ho ragione di ritenere (e so che posso sbagliare il 5% delle volte) che questo valore sarà compreso tra gli estremi dell'intervallo fiduciale da me calcolato sulla base dei dati del campione".
Tra questi 20 sperimentatori, ce ne sarà verosimilmente uno che otterrà un intervallo che non contiene la media incognita. Per esempio, un ricercatore che avesse ottenuto dal suo campione m=54.8 e s=3.72, avrebbe un intervallo di estremi 50.2 +-------+ 59.4, intervallo che, come si vede, non contiene il valore incognito 50, valore della media della popolazione d'origine del campione stesso.
Ma allora, dirà qualcuno: non si avrà mai la certezza di conoscere la media incognita! In effetti, è proprio cosi'. La statistica induttiva non può fornire certezze, bensi ragionevoli indicazioni per effettuare scelte razionali (anche se, sotto condizioni particolarmente fortunate, tali indicazioni corrispondono, in pratica, ad una certezza per quanto riguarda le decisioni da prendere).
Ragionamenti analoghi a quanto visto per la media, possono essere fatti per la deviazione standard e per altri parametri: la struttura logica del ragionamento, tuttavia, è sempre la stessa.

I test di ipotesi

Molte volte, nel corso del proprio lavoro, il ricercatore è indotto a controllare in base ai dati sperimentali raccolti la validità di una sua ipotesi. In realtà, egli cerca di valutare se i dati prodotti possano in qualche modo invalidare quella che viene indicata come "ipotesi nulla" (H_0), che nei fatti è poi la negazione della sua ipotesi, "l'ipotesi di lavoro". In altre parole, se sto studiando l'effetto di un certo farmaco, la mia ipotesi di lavoro sarà che il farmaco manifesterà un certo effetto; la mia ipotesi nulla, che il farmaco non abbia effetto: conoscendo la distribuzione di certe quantità statistiche nel caso in cui sia vera l'ipotesi nulla, posso sperare di mettere in evidenza un eventuale disaccordo tra i dati osservati e quanto mi dovrei aspettare quando nella realtà l'ipotesi nulla è verificata (si osservi che in questo caso non valutiamo la probabilità dell'ipotesi nulla $\Pr(H_0)$, bensì la probabilità condizionata $\Pr(\{x_i\} / H_0)$ di osservare un campione come quello estratto nel caso in cui l'ipotesi nulla sia vera. La Statistica Induttiva mette a disposizione del ricercatore numerosi strumenti per questo tipo di valutazione. In genere, tuttavia, quando si voglia controllare un'ipotesi su due campioni di dati espressi su scala almeno intervallare, si fa uso del solo test t di Student, test che è soggetto a numerose limitazioni. Nel seguito esamineremo due test non-parametrici, di grande applicabilità su

piccoli campioni, che non sono soggetti ad alcuna limitazione e che, ben compresi, permetteranno di entrare meglio nella logica alla base dei test statistici d'ipotesi.

Il test di casualizzazione per due campioni indipendenti - Il test di casualizzazione per due campioni indipendenti è il test più potente per analisi di questo tipo ed è anche molto elegante. Per applicarlo è necessario che i dati siano misurati almeno su una scala intervallare, in quanto, come vedremo, per il suo calcolo sarà necessario effettuare delle somme che dovranno poi essere tra loro confrontate. L'idea centrale del test è semplice, ma non banale: supponiamo di avere due gruppi di soggetti, rispettivamente di numerosità n_1 e n_2 , con $n = n_1 + n_2$, il primo di controllo e il secondo a trattamento. Ora, se è vera l'ipotesi nulla (assenza di effetti dovuti al trattamento), un generico soggetto assegnato al primo gruppo potrebbe certamente scambiarsi con un soggetto qualsiasi del secondo gruppo, senza che la nostra valutazione subisca particolari modifiche.

In altre parole, i risultati ottenuti con gli n_1 soggetti allocati nel gruppo di controllo e con i restanti n_2 assegnati al gruppo a trattamento non rappresentano altro che una delle tante possibilità, sostanzialmente equivalenti se è vera H_0 , che avremmo potuto osservare con una diversa allocazione dei soggetti nei due gruppi. Le diverse situazioni che potevano presentarsi distribuendo nei due gruppi gli n soggetti sono date, d'altra parte, dal numero di combinazioni di n oggetti a n_1 a n_1 , ovvero di n oggetti a n_2 a n_2 , numero pari a tutte le diverse configurazioni sperimentali che potevano ottenersi distribuendo a caso i soggetti nei due gruppi. Questo numero di possibili configurazioni sarà dato da:

$$n_{camp} = \binom{n}{n_1} = \binom{n}{n_2} = \frac{n!}{(n-n_1)!\,n_1!} = \frac{n!}{(n-n_2)!\,n_2!}$$

Consideriamo, ora, la somma delle risposte di uno dei due gruppi: è chiaro che detto valore dipenderà dai particolari soggetti assegnati a detto gruppo; se uno o più soggetti fossero scambiati con altri del secondo gruppo, tale somma sarebbe probabilmente diversa. Dunque, la somma dei dati relativi ad un gruppo può costituire un criterio per ordinare i possibili diversi campioni che si sarebbero potuti ottenere distribuendo nello schema sperimentale i soggetti selezionati.

D'altra parte, una volta che tutte le possibili somme sono state calcolate, potremmo anche individuare in che punto si inserisce la somma risultante dal nostro esperimento.

Se moltiplichiamo il numero delle possibili somme per un certo valore di rischio di primo tipo, potremmo stabilire una regione critica di rifiuto (unilaterale o bilaterale), all'interno della quale rifiutare l'ipotesi nulla al livello di rischio scelto.

Riassumiamo, dunque, i vari passi necessari per l'effetuazione del test:

a) riferirsi ad uno dei gruppi e calcolare la somma dei dati relativi ai soggetti ivi assegnati;

b) calcolare il numero di possibili campioni, n_{camp} , e definire quali somme delimitano la regione critica, unilaterale ($\alpha \cdot n_{camp}$) o bilaterale ($\frac{\alpha}{2} \cdot n_{camp}$);
c) calcolare tutte le possibili somme ed ordinarle;
d) controllare dove cade la somma che osserviamo nell'esperimento;
e) decidere se rifiutare o meno l'ipotesi nulla.
Vale la pena sottolineare che non sempre è necessario esplicitare tutti questi passi. Infatti, quando il risultato della somma è estremo, o quasi estremo, la conclusione è praticamente immediata.
Ad esempio, si immagini un esperimento in cui si confrontino un gruppo di cinque soggetti di controllo e un gruppo di sei soggetti trattati, esperimento in cui i risultati del gruppo trattato siano tutti superiori a quelli del gruppo di controllo: in questo caso, la somma osservata sarà certamente estrema e la sua significatività sarà data (nel caso di test unilaterale) da $\frac{1}{\binom{11}{6}} = \frac{1}{\binom{11}{5}} = \frac{1}{\frac{11!}{6!5!}} = \frac{1}{462} = 0.0022$
Si osservi che è rilevante il fatto che la conclusione cui si è giunti (i campioni provengono da popolazioni diverse, con $p<0.0022$) è, data la natura del test, indipendente da qualunque ipotesi relativa alle popolazioni di provenienza (es. normalità, omoschedasticità, ecc.). In altre parole, mentre il risultato del test t per due campioni indipendenti è valido solo sotto certe condizioni, il risultato ottenuto col test di casualizzazione è, per sua natura, sempre valido.
L'ulteriore esempio numerico, nel seguito riportato, aiuterà a rinforzare la comprensione dei concetti esposti e chiarirà molti dubbi.

Test di casualizzazione per due campioni indipendenti: esempio

Concentrazione ematica della sostanza XYZ: dati in mg/l00mi di sangue

Gruppo **A** (Controllo)	N= 9	Gruppo **B** (Trattato)
8		5
12		7
11		9
15		8
12		Nb=4
Na= 5		Somma B= 29
Somma A= 58		

$$Ncamp.= \binom{9}{5} = \binom{9}{4} = 126$$

$$\frac{9!}{5!4!} = \frac{9x8x7x6}{4x3x2} = 126$$

N camp = 126

Ipotesi nulla - il farmaco non ha alcun effetto sulla concentrazione ematica della XYZ

Ipotesi altemativa - il farmaco abbassa i livelli ematici della XYZ

Calcolo della regione critica - il test è evidentemente ad una coda; se stabiliamo un livello di errore alfa pari a 0.05, avremo che la somma critica sarà data da: 126x0.05= 6.3. Arrotondando per difetto, diremo che se la somma osservata per il gruppo trattato è tra le prime sei somme (le sei somme più piccole), allora rifiuteremo l'ipotesi nulla ad un livello di rischio di errore di prima specie pari a 0,05.

Calcolo esaustivo (tramite programma per elaboratore elettronico) - il programma calcola le 126 possibili somme e determina quante superano la somma osservata, quante sono a questa equivalenti e quante sono superate dalla stessa. Su questa base di risultati si decide o meno di rifiutare l'ipotesi nulla.

Calcolo abbreviato - la somma della colonna a trattamento è certamente una somma estrema, in quanto i dati dei trattati sono tutti inferiori a quelli dei controlli, ad eccezione dei valore pari ad 8 (primo valore dei controlli). La somma più bassa possibile la si ottiene scambiando il dato 8 dei controlli con il dato 9 dei trattati (Sb= 28). Scambiando il dato dato 8 dei controlli col dato 8 dei trattati si ottiene la somma immediatamente superiore, pari a 29, valore che coincide con la somma osservata nella sperimentazione effettuata.
Tutte le altre possibili somme che possono costruirsi risulteranno superiori al valore 29 perché importeremo nel gruppo dei trattati dati superiori a quelli che verranno spostati nel gruppo dei controlli. Rifiutiamo, perciò H° al livello $p<0,05$, in quanto il valore osservato della somma, pari a 29, entra nella coda delle prime tre somme più piccole (28, 29 e 29)

N.B.: si osservi che se dividiamo la coda determinata dalla somma osservata nell'esperimento effettuato per il numero dei possibili campioni, abbiamo la probabilità esatta di errore di tipo alfa nel rifiutare l'ipotesi nulla. Nel nostro caso si avrà: 3/126= 0,024.

Il test di casualizzazione per due campioni dipendenti

Il test di casualizzazione per due campioni dipendenti è il test più potente per analisi di questo tipo e, come l'analogo già visto nel caso di due campioni indipendenti, è assai elegante ed istruttivo. E', nella sua essenzialità, bello, decisamente. Per applicarlo è necessario che i dati siano misurati almeno su una scala intervallare, in quanto, come vedremo, sarà necessario anche in questo caso effettuare delle somme che dovranno essere tra loro confrontate. L'idea centrale del test è questa: supponiamo di avere n soggetti sui quali viene effettuato un esperimento tipo prima-dopo, per esempio misurando una certa variabile prima e dopo la somministrazione di un farmaco. Orbene, focalizzando l'attenzione su un singolo soggetto, se è vera l'ipotesi nulla (cioè, che il farmaco non abbia alcun effetto), il fatto che la prima misura sia risultata pari ad x_i e la seconda a y_i è puramente casuale, dovuto a fluttuazioni e non certamente all'effetto del farmaco. Nell'ipotesi nulla, una situazione che vedesse "prima" un risultato pari ad y_i e

"dopo" pari a x_i , sarebbe altrettanto probabile. Se questo è vero, allora la configurazione di risultati che noi osserviamo sull'insieme dei soggetti è una delle 2^n configurazioni possibili nell'ipotesi nulla (sono 2^n perché ognuno degli n soggetti fornisce due possibilità che, se vale l'ipotesi nulla, sono ugualmente probabili). Si osservi che, detta d_i la differenza osservata tra i risultati riscontrati per un generico soggetto, lo scambio dei due valori porta al cambiamento del segno della d_i , che diventa così $-d_i$.

Ora, se consideriamo la quantità $d_T = \sum_{i=1}^{n} d_i$, somma delle differenze col segno, possiamo giustamente ritenere che nell'ipotesi nulla il suo valore oscillerà intorno a zero, nel senso che la somma algebrica delle $\{d_i$ conterrà più o meno un numero equivalente di differenze positive e negative, di modulo più o meno analogo. Solo poche di queste somme si allontaneranno sostanzialmente dallo zero (quelle in cui la gran parte dei segni è positiva o la gran parte negativa). Nel caso in cui l'ipotesi nulla H_0 non fosse la reale situazione sottostante, ci aspetteremmo, naturalmente, una "tendenza" delle differenze e, quindi, la prevalenza di un segno, cui conseguirebbe un valore sostanzialmente lontano da zero della d_T . Se noi calcolassimo, in base ai risultati ottenuti nella nostra sperimentazione, le 2^n somme possibili, ordinandole potremmo stabilire una regione critica per il rifiuto di H_0. Tale regione sarà certo data dalle prime $\alpha \cdot 2^n$ somme o dalle ultime $2^n \cdot (1-\alpha)$ somme, nel caso di test unilaterale, oppure dalle prime $\frac{\alpha}{2} \cdot 2^n$ somme e dalle ultime $2^n \cdot (1-\frac{\alpha}{2})$ somme nel caso di un test a due code.

Esempio - abbiamo effettuato uno studio prima-dopo su 7 soggetti, misurando una certa variabile prima e dopo la somministrazione di un certo farmaco. I risultati ottenuti sono i seguenti:

Soggetto	Prima	Dopo	Differenza D - P
1	14	16	+2
2	19	23	+4
3	18	17	-1
4	24	25	+1
5	8	12	+4
6	13	14	+1
7	27	32	+5

Somma delle differenze osservata nell'esperimento = +16

Ora, nell'ipotesi nulla, abbiamo $2^7 = 12$ possibili configurazioni, ad ognuna delle quali corrisponderà una certa somma delle differenze col segno. Scegliendo un rischio del 5%, nell'ipotesi che la somministrazione del farmaco o non abbia effetto o aumenti il valore della variabile misurata (ipotesi unilaterale), avremo una regione critica di somme pari a

$28 \cdot 0.05 = 6.4$. In termini conservativi, considereremo significativo, con $p < 0.0$, il risultato dell'esperimento se la somma osservata delle differenze cadrà nella regione critica costituita dalle 6 somme estreme. Come si verifica facilmente, accade proprio questo: infatti, la somma massima possibile si ottiene invertendo i risultati ottenuti sul terzo soggetto; la somma osservata è quella immediatamente inferiore, ex-aequo con altre due somme (ottenute l'una invertendo i risultati del quarto soggetto, l'altra quelli del sesto soggetto). Il risultato è, quindi significativo al 5% di rischio in quanto la somma osservata cade nella regione critica predefinita.

Dimensionamento

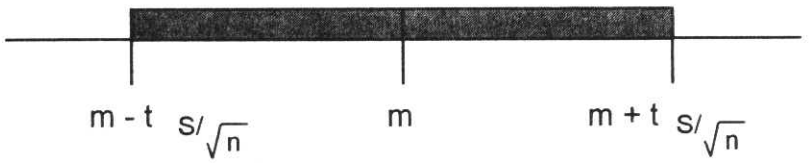
Una domanda che spesso ci si pone nell'impostazione di una ricerca è la seguente: "Le dimensioni scelte per il campione sono sufficienti per stimare il parametro che ci interessa con un'approssimazione che renda possibile una sua vantaggiosa utilizzazione in successive applicazioni?"
La domanda non è banale. Infatti, il parametro che interessa viene stimato mediante campione e, come abbiamo visto, la media campionaria fluttua intorno al valore incognito del parametro. L'espressione dell'intervallo fiduciale di una media può, però, suggerirci una strada per risolvere il problema.
Abbiamo visto che risulta:

$$\Pr(m - t_{\nu,\alpha/2} \frac{s}{\sqrt{n}} < \mu < m + t_{\nu,\alpha/2} \frac{s}{\sqrt{n}}) = 1 - \alpha$$

e che, dunque, l'intervallo

$$m - t_{\nu,\alpha/2} \frac{s}{\sqrt{n}} + - - - - + m + t_{\nu,\alpha/2} \frac{s}{\sqrt{n}}$$

contiene con probabilità pari a $(1 - \alpha)$ la media incognita μ della popolazione da cui è stato tratto il campione (campione di n elementi, con media m e deviazione standard s). Dunque, la "vera" media è entro $\pm t_{\nu,\alpha/2} \frac{s}{\sqrt{n}}$ dalla media m del campione; e facendo questa affermazione sappiamo che abbiamo un rischio α di sbagliare.
Si osservi, (v. figura), che questo equivale a dire che, sempre con rischio α , la media incognita μ della popolazione non disterà dalla media del campione più di $\left| t_{\nu,\alpha/2} \frac{s}{\sqrt{n}} \right|$.



Se questo è vero, stabilita in base ai nostri obiettivi pratici la massima incertezza che siamo disposti ad accettare sulla media della popolazione, potremmo utilizzare la quantità $t_{\nu,\alpha/2}\frac{s}{\sqrt{n}}$, vincolandola a tale incertezza, al fine di derivare il valore di n corrispondente.

Vediamo meglio il tutto su un esempio. Si immagini di avere a che fare con misure di concentrazione di colesterolo in un certo campione di siero: può darsi che in base a conoscenze medico-laboratoristiche, se il risultato deve essere utilizzato a fini clinici (diagnostici, prognostici o terapeutici che siano) un'inaccuratezza del 10% sulla vera concentrazione possa essere considerata accettabile; se l'uso che se ne fa è, invece, quello di realizzare un siero di riferimento per un programma di controllo di qualità tra laboratori, tale inaccuratezza potrebbe essere ritenuta eccessiva. Chiaramente, il numero di replicati necessari per rispondere alla prima esigenza sarà inferiore al numero di quelli necessari per fornire una risposta valida nel secondo caso.

Si rifletta sul significato di questa incertezza, che nel seguito chiameremo "inaccuratezza accettata": essa rappresenta il massimo "rumore" che siamo disposti a tollerare nella nostra conoscenza della media della popolazione a fronte di quello che poi in base a questa conoscenza dovremo decidere o fare. Questo "rumore", questo "grigio", non è una grandezza di carattere statistico, bensì un limite che colui che sta conducendo la ricerca, o affrontando uno specifico problema, può tentare di stabilire. In altre parole, dovendo decidere se tizio ha o non ha il diabete, che la sua glicemia risulti 230 mg/100ml o 190 mg/100 ml non mi cambia molto le cose; se, invece, sto studiando un fenomeno fisiopatologico in cui modeste variazioni di glicemia giocano un ruolo importante, il fatto che questa risulti 190 mg/100 ml invece di 230 mg/100ml può togliere qualunque validità alla mia ricerca.

Supponiamo, dunque, che sia possibile stabilire questa inaccuratezza massima accettabile, diciamola $2 \cdot D$.

Ora, dire che vogliamo che la vera media sia contenuta con rischio α, entro $\pm D$ dalla media osservata nel campione, equivale a dire che la grandezza dell'intervallo fiduciale della media deve risultare al massimo pari a due volte il valore D

In formule, questo significa imporre: $D = t_{\nu,\alpha/2}\frac{s}{\sqrt{n}}$;

e, risolvendo rispetto ad n, si ha: $n = \left(\frac{t_{\nu,\alpha/2} \cdot s}{D}\right)^2$, relazione che ci permette di stabilire quanti elementi bisogna estrarre dalla popolazione per conoscere la media "vera" con una incertezza di $\pm D$ dalla media campionaria.

Si osservi che l'applicazione di questa formula implica la conoscenza di una stima della deviazione standard della popolazione: quando questa non è direttamente disponibile (né ricavabile da dati della letteratura) è necessario, per stimarla, effettuare un campionamento pilota di ragionevoli dimensioni (per es., di 30 elementi).
C'è, inoltre, un'ulteriore complicazione, in quanto nel membro di destra della precedente formula, anche il termine $t_{\nu,\alpha/2} = t_{(n-1),\alpha/2}$ dipende da n e quindi la relazione non è immediatamente risolvibile.
Un metodo per aggirare questa difficoltà è quello di utilizzare la formula ponendo al posto di $t_{\nu,\alpha/2}$ il quantile $z_{\alpha/2}$ della distribuzione normale standardizzata (stiamo facendo l'ipotesi che t abbia infiniti gradi di libertà). Avremo così: $n_0 = \left(\frac{z_{\alpha/2} \cdot s}{D}\right)^2$.
Detto n_0^* l'intero più vicino a n_0, si può calcolare un nuovo valore di n, diciamolo n_1, rientrando poi nella formula vista, ma sostituendo al posto di $z_{\alpha/2}$, che ha assolto la sua funzione di valore di ingresso nel processo ricorsivo, la quantità $t_{(n_0^*-1),\alpha/2}$, ovvero il quantile della distribuzione t di Student con i gradi di libertà definiti dal valore n_0^* ricavato nel primo ciclo: $n_1 = \left(\frac{t_{(n_0^*-1),\alpha/2} \cdot s}{D}\right)^2$.
Il processo mostrato può essere ripetuto utilizzando ora il valore n_1^* (l'intero più vicino a n_1) che ci consentirà di rientrare nella formula con un $t_{(n_1^*-1),\alpha/2}$, ecc. ecc..
Generalmente, i due membri della relazione si stabilizzano dopo tre-quattro cicli (in altre parole, il valore di n non cambia più o oscilla di una unità).

Esempio - vogliamo conoscere la concentrazione di una certa sostanza in un liquido biologico con un'inaccuratezza pari a 2 unità. Sappiamo da studi precedenti che la deviazione standard di misure replicate è, nell'intorno della concentrazione presunta, pari a 4 unità. Quante misure sono necessarie al 5% di rischio?
Osserviamo innanzitutto che se l'inaccuratezza accettata è di due unità, allora accetteremo al massimo intorno alla media campionaria un'oscillazione di più o meno una unità. Sarà:

$$n_0 = \left(\frac{z_{\alpha/2} \cdot s}{D}\right)^2 = \left(\frac{1.96 \cdot 4}{1}\right)^2 = 61.5$$

;

approssimando a 62, rientriamo nella formula con 61 gradi di libertà:

$$n_1 = \left(\frac{t_{61,0.025} \cdot s}{D}\right)^2 = \left(\frac{1.9997 \cdot 4}{1}\right)^2 = 63.8$$

;

approssimando a 64, rientriamo nella formula con 63 gradi di libertà:

$$n_2 = \left(\frac{t_{63,0.025} \cdot s}{D} \right)^2 = \left(\frac{1.9984 \cdot 4}{1} \right)^2 = 63.9$$

;

il valore si è così stabilizzato (rientreremmo sempre con 63 gradi di libertà e, quindi, possiamo perciò concludere che per soddisfare i termini del problema sono necessarie almeno 64 misure. Tenendo conto di possibili dati aberranti, misurazioni da scartare, ecc., sembra ragionevole prevedere almeno 70 misure.

Poiché, alcune volte la variabilità del fenomeno è conosciuta in termini di coefficiente di variazione, può essere utile modificare l'equazione già vista per un uso più diretto.

Se, infatti, dividiamo numeratore e denominatore del membro di destra dell'equazione per m (dove m è la media campionaria) otteniamo:

$$n = \left(\frac{t_{\nu,\alpha/2} \cdot \frac{s}{m}}{\frac{D}{m}} \right)^2$$

, ovvero:

$$n = \left(\frac{t_{\nu,\alpha/2} \cdot CV\%}{D\%} \right)^2$$

,

dove $CV\%$ è ora la nostra stima del coefficiente di variazione della popolazione espresso in percentuale e $D\%$ l'inaccuratezza percentuale nei due sensi che siamo disposti ad accettare sulla media vera.

Quest'ultima formula è quella generalmente utilizzata e permette di rispondere a problemi tipo: "Data una popolazione (un metodo) con coefficiente di variazione del 20%, quanti elementi deve contenere un campione per stimare la media della popolazione con un'inaccuratezza del 10% ad un rischio del 5%?". In questo caso si avrebbe:

$$n_0 = \left(\frac{z_{\alpha/2} \cdot s}{D} \right)^2 = \left(\frac{1.96 \cdot 20}{5} \right)^2 = 61.5$$

, ed il processo è nel seguito analogo a quanto visto in precedenza.

Si osservi, infine, che se vogliamo dimezzare l'inaccuratezza non basta raddoppiare il numero di dati, ma è necessario quadruplicarli. Infatti, se per avere un "rumore" massimo pari a D ho bisogno di n misure, date da $n = \left(\frac{t_{\nu,\alpha/2} \cdot s}{D} \right)^2$, per un "rumore" pari a $\frac{D}{2}$ saranno necessarie n^* misure, date da:

$$n^* = \left(\frac{t_{\nu,\alpha/2} \cdot s}{\frac{D}{2}} \right)^2 = 4 \left(\frac{t_{\nu,\alpha/2} \cdot s}{D} \right)^2 = 4n$$

.