Capitolo VI

INFERENZA STATISTICA

1. Cosa si intende per inferenza statistica

Nelle sezioni precedenti si sono considerati gli aspetti della statistica connessi con la descrizione dei fenomeni e con la stima di certe caratteristiche della *popolazione* o *universo* dei dati.

Quando, però, si deve risolvere un problema di carattere statistico, in genere si opera su un insieme di dati che non sempre costituisce la totalità di quegli stessi dati relativi al fenomeno studiato. Per questa ragione, nella Statistica sono stati introdotti i concetti di *popolazione* e *campione* tra i quali vi è una differenza sostanziale: mentre con il primo termine si indica l'insieme di tutte le unità statistiche nelle quali è presente il fenomeno che si vuole studiare, con il secondo termine ci si riferisce solamente ad una parte dell'intero insieme di quelle unità o popolazione, che sia stata selezionata secondo certi criteri di estrazione o *metodi di campionamento*. (estrazione casuale, campione sistematico, campionamento semplice, a grappolo, a uno o più stadi, ecc.).

Le popolazioni da cui sono estratti uno o più campioni possono contenere un numero di elementi finito oppure infinitamente grande. In questi casi si parla di *popolazioni finite* o di *popolazioni infinite*.

A loro volta, i campioni possono essere formati da un diverso numero di unità statistiche elementari tratte dalla popolazione, dando luogo a *grandi campioni* o a *piccoli campioni*. Un campione che contiene meno di 50 elementi (o talvolta anche meno di 30) si considera che sia un "piccolo

campione", mentre un campione composto da più di 50 (o 30) elementi è chiamato "grande campione". La distinzione tra campioni grandi e piccoli ha rilevanza, come si vedrà, soprattutto per determinare quale sia la distribuzione specifica da considerare per la scelta del test statistico di significatività delle stime.

La selezione delle unità della popolazione che entrano a far parte del campione da studiare può essere *casuale* o *non casuale*. Nel primo caso si ammette che l'unica motivazione della eventuale differenziazione tra le caratteristiche del campione e quelle della popolazione o universo di origine sia la accidentalità o casualità della scelta. Nel secondo caso, invece, generalmente si opera una scelta dettata da criteri definiti che nella gran parte dei casi conducono a risultati campionari "*distorti*" rispetto a quelli "veri" relativi all'intera popolazione di unità statistiche.

Nel seguito sarà preso in considerazione solo il campionamento casuale, cioè la selezione degli elementi della popolazione facendo in modo che ogni elemento abbia la stessa probabilità di essere scelto.

Un campione di n unità estratte da una data popolazione è un campione casuale quando tutti gli altri campioni possibili formati anch'essi da n unità e provenienti dalla stessa popolazione hanno la stessa probabilità di essere scelti.

Dunque, nella maggior parte dei casi, quando si ricerca la media aritmetica, la varianza ed altre misure caratteristiche della distribuzione di un dato fenomeno *di solito si utilizzano dati statistici riferiti ad un campione di unità tratto dalla popolazione di dati che caratterizzano il fenomeno che si vuole analizzare*. Se vi fosse la capacità di ottenere tutte le singole misure che formano una popolazione di dati, la media o gli altri parametri statistici caratteristici della distribuzione del fenomeno rappresenterebbero la "vera media", la "vera varianza" e così via, dell'intera popolazione di dati. Ma spesso, per ragioni di natura diversa (costo eccessivo, irraggiungibilità di tutte le unità statistiche della popolazione studiata, carenze di tempo per le rilevazioni e così via), è impossibile considerare l'intera popolazione e la maggior parte delle volte ci si deve accontentare di calcolare le misure caratteristiche relative ad un campione di unità statistiche

tratto da essa. Queste misure caratteristiche sono chiamate *statistiche* *campionarie*, mentre le vere misure sono chiamate *parametri della popolazione*. *Le statistiche campionarie sono stime dei parametri della popolazione*.

L'attendibilità di una misura ottenuta da un campione dipende dall'accuratezza di queste stime. La media o la deviazione standard calcolate su un campione casuale non forniscono elementi di conoscenza sufficienti per *trovare i valori della vera media e della vera deviazione standard relative alla popolazione*. Tuttavia, con l'aiuto di queste statistiche campionarie ed utilizzando anche certe *proprietà dei campioni casuali*, si è in condizione di trovare entro quali *limiti* ci si può attendere che siano contenuti i parametri della popolazione. Tali limiti possono essere determinati solo con un certo grado di *confidenza* o *precisione* o *accuratezza* o *attendibilità*. *Più i limiti sono ristretti, più è elevata la precisione e più è attendibile la stima*.

E' anche possibile *determinare la significatività della differenza tra i valori che una stessa statistica assume in campioni diversi*, a condizione di conoscere come varia quella statistica al variare del campione, ossia conoscere in qual modo si distribuiscono le statistiche campionarie. Per trovare, ad esempio, quale percentuale delle medie di tutti i campioni casuali che possono essere estratti da una popolazione, ci si può attendere che cada entro limiti definiti, dobbiamo conoscere quale sia la *distribuzione di frequenza delle medie* campionarie. Nello stesso modo per essere capaci di giudicare della attendibilità di una deviazione standard campionaria si deve conoscere la *distribuzione di frequenza delle deviazioni standard* campionarie. E così via anche per le altre *statistiche* possibili.

2. Uso della statistica per le decisioni: l'ipotesi statistica

Tutto l'insieme delle considerazioni ora svolte costituisce la materia di cui si occupa la teoria dell'*inferenza statistica*. Ma ora è il caso di considerare come l'insieme degli elementi discussi fino a questo punto possa essere utilizzato per assumere determinate decisioni operative.

In geometria si può affermare per ipotesi che la somma degli angoli di un triangolo è 180°. Con procedure di verifica accettate, si riesce anche a decidere se questa ipotesi sia vera o falsa. In casi come questo si ha una verifica matematica dell'ipotesi e una volta completata la verifica si potrà essere certi che l'ipotesi si dimostrerà vera oppure falsa. Anche per altre discipline diverse da quelle matematiche può essere proposta un'ipotesi o una teoria concernente un dato universo o insieme o popolazione di unità statistiche. Ma in questo caso la formulazione viene chiamata "*ipotesi statistica*" ed il solo modo di essere assolutamente certi della sua verità o falsità può essere quello di esaminare l'intera popolazione interessata. Ciò è generalmente difficile o impossibile ed in tali casi siamo forzati ad analizzare solo un *campione* di unità estratto dalla popolazione e ad usare i risultati che da esso possono ricavarsi, per decidere se l'ipotesi sia vera o falsa. L'operazione per mezzo della quale si usa il campione per "verificare" se l'ipotesi sia vera (o falsa) è chiamata *"verifica statistica della validità dell'ipotesi.*"

Sfortunatamente non vi è "certezza" che non venga commesso un errore. Infatti, in generale, si possono commettere due diversi tipi di errori:

- *se l'ipotesi ammessa è vera, essa potrebbe essere definita falsa (errore del I tipo o errore α)* oppure
- *se l'ipotesi assunta è falsa, essa potrebbe essere definita vera (errore del II tipo o errore β)*.

Naturalmente, è molto importante la frequenza con cui gli errori sono compiuti e sarà necessario fare in modo che questa frequenza possa essere controllata in una certa misura.

Esempi di ipotesi che possono essere soggette ad una verifica statistica sono i seguenti:

1. Il gruppo di osservazioni in esame è un campione tratto da una popolazione con media uguale a μ. Sono di questo tipo, ad esempio, le affermazioni seguenti:

a. Le lampadine elettriche di un certo stock sono di qualità standard (durata media di vita μ uguale ad uno specifico valore μ_o).
b. Il numero medio di batteri uccisi da goccie campione di un germicida è uguale ad un certo numero standard.
c. L'intelligenza media di una data classe è uguale a quella media di tutti gli studenti.

2. Il gruppo di osservazioni in esame è un campione tratto da una popolazione con varianza σ^2. Sono di questo tipo, ad esempio, le affermazioni seguenti:

a. La classe di studenti esaminata è variabile in intelligenza proprio come la classe media.
b. Una data macchina produce perni di diametro più uniforme della variazione ammessa σ^2.
c. Il periodo di crescita per un dato ibrido di mais è più variabile del periodo di crescita di altri ibridi.

3. I due gruppi in esame sono tratti da popolazioni aventi la stessa media ($\mu_1=\mu_2$). Sono di questo tipo, ad esempio, le affermazioni seguenti:

a. Il metodo A è migliore del metodo B per insegnare l'algebra.
b. L'acciaio prodotto con il metodo A è più duro dell'acciaio prodotto con il metodo B.
c. La penicillina è più efficace della streptomicina nel trattamento della malattia X.

Ipotesi di altro tipo che implichino più di due medie e varianze, quelle che riguardano le proporzioni, le ipotesi concernenti le associazioni, eccetera, saranno introdotte via via nel prosieguo del corso.

La decisione se accettare o rigettare una ipotesi si basa sulle *informazioni* che si ottengono dalle osservazioni fatte e sul livello che si ritiene sostenibile per il *rischio* che la decisione da prendere sia sbagliata.

Anzitutto si deve *definire la ipotesi di lavoro* (per esempio, stabilire un dato valore per un parametro della popolazione). Quindi *si raccoglie un certo numero di osservazioni* (il campione) e *si esaminano i risultati* ottenuti per vedere se essi siano o no simili a quelli della popolazione stabiliti nella ipotesi avanzata a priori. Se vi è una stretta concordanza, si accetta l'ipotesi. Se la concordanza è scarsa, l'ipotesi sarà rigettata. Per decidere se vi sia o no una stretta concordanza, di solito si calcola qualche *statistica* ed il valore particolare ottenuto dal campione si compara con la *distribuzione campionaria* di questa statistica, supponendo che l'ipotesi sia vera.

3. Errore α o del I tipo (rigettare come falsa un'ipotesi vera)

Il procedimento per verificare una qualsiasi delle ipotesi definite nel paragrafo precedente può essere illustrato tramite l'uso della distribuzione campionaria della media m_x. Si supponga che la media della popolazione sia μ.

La statistica m_x viene calcolata per un campione di n osservazioni. Lo studio della distribuzione campionaria di m_x mostra che una media del campione molto discosta da μ si verificherà raramente se la ipotesi è vera. Quindi, se la media calcolata sul campione è prossima a μ, si accetterà l'ipotesi che il campione provenga da una popolazione che ha la stessa media. Se la media calcolata è molto diversa da μ, l'ipotesi sarà rigettata.

Si consideri ora la distribuzione campionaria di m_x per trovare la *relazione esplicita tra la dimensione della deviazione e l'attendibilità della deviazione stessa*. Si supponga che il campione sia formato di n osservazioni tratte da una popolazione normale con deviazione standard pari a σ. Possiamo supporre che anche la distribuzione campionaria della media sia normale ed in tal caso si dimostra che la sua deviazione standard è $\sigma_m = \sigma/\sqrt{n}$. Se la media della popolazione è μ, allora anche la media della distribuzione campionaria di m_x è uguale a μ.

Si scelgano ora i valori di m_x che indurranno a rigettare il valore ipotetico $\mu=\mu_0$. Si decida, per esempio, di rigettare l'ipotesi $\mu=\mu_0$ se m_x è a più di $z_{\alpha/2}.\sigma_m$ unità da μ_0, ove:

- con z si è indicata la variabile standardizzata che corrisponde alla variabile originaria m_x, cioè si è posto $z = \dfrac{m_x - \mu}{\sigma_m}$;

- con α si è indicata la probabilità di rigetto dell'ipotesi $\mu=\mu_0$, mentre con 100α si indica il grado percentuale di rischio che si accetta di sostenere;

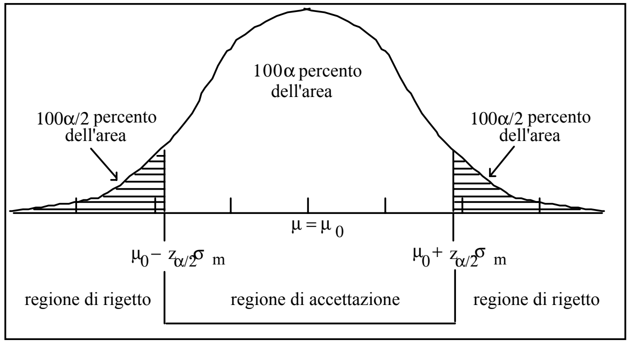
- con $z_{\alpha/2}$. si è indicato il valore che assume z quando si suppone che sia $\mu=\mu_0$ e con una probabilità pari ad α.

In altre parole, si rigetta l'ipotesi che sia $\mu=\mu_0$ se m_x è minore di $(\mu_0 - z_{\alpha/2}.\sigma_m)$ oppure se m_x è maggiore del valore $(\mu_0 + z_{\alpha/2}.\sigma_m)$ e si accetta l'ipotesi $\mu=\mu_0$ se m_x è compreso tra questi due valori.

Dalla tavola che fornisce i valori *cumulati* delle ordinate della distribuzione normale [il riferimento alla tavola è possibile considerando la relazione $z = (m_x - \mu)/\sigma_m$] si può rilevare che $100\alpha/2$ percento dell'area sotto la distribuzione campionaria corrisponde a valori di m_x che sono minori della

quantità $(\mu_0 - z_{\alpha/2}.\sigma_m)$ oppure che sono maggiori di $(\mu_0 + z_{\alpha/2}.\sigma_m)$. Quindi, se la media della popolazione è effettivamente $\mu = \mu_0$, si ha una probabilità pari ad α oppure a 100α percento che una media campionaria m_x sia minore del valore $(\mu_0 - z_{\alpha/2}.\sigma_m)$ oppure sia maggiore di $(\mu_0 + z_{\alpha/2}.\sigma_m)$ e che induca quindi a rigettare come falsa l'ipotesi vera.

La distribuzione campionaria, le probabilità percentuali di rigetto e di accettazione di una ipotesi vera ed i limiti della regione critica possono essere rappresentati come segue:

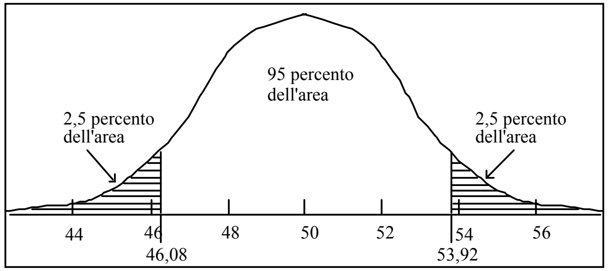


Come applicazione, si supponga che la media della popolazione sia 50, cioè $\mu = 50$, che sia a = 0,05 e quindi in corrispondenza di a/2 = 0,025 dalla tavola dei valori cumulati della curva normale sarà $z_{a/2} = 1{,}960$.

La statistica m_x sia calcolata per un campione di n = 25 osservazioni tratto da una popolazione normale con deviazione standard pari a 10. Inoltre, sappiamo che pure la distribuzione campionaria della media è normale, con deviazione standard $\sigma_m = \sigma/\sqrt{n} = 10/5 = 2$.

La distribuzione campionaria suddetta sarà ben rappresentata nel grafico che segue, ove sono indicati gli estremi del campo che definisce l'intervallo di confidenza dei risultati, determinato così come è stato indicato più sopra, nella parte teorica, e nel ragionamento specifico che segue.

Si scelgano i valori della media m_x che indurranno a rigettare il valore ipotetico $\mu = 50$. Si supponga, per esempio, che si decida di rigettare l'ipotesi $\mu = 50$ se m_x è a più di $1{,}960.\sigma_m = 1{,}960 \times 2 = 3{,}920$ unità da 50. Cioè, si rigetta l'ipotesi $\mu = 50$ se m_x è minore di 46,08 oppure se m_x è maggiore di 53,92 e si accetta l'ipotesi $\mu = 50$ se m_x è compreso tra questi due valori.



Nella tavola della cumulata delle ordinate della distribuzione normale si vede che il 2,5 percento dell'area sotto la distribuzione campionaria rappresentata nel grafico sopra riportato corrisponde a valori di m_x minori di 46,08 (il riferimento alla tavola si effettua ponendo z = (46,08 - 50)/2 = - 1,960) ed il 2,5 percento corrisponde a valori di m_x maggiori di 53,92. Quindi, se μ = 50, si ha una probabilità del 5 percento (α = 0,05) che una media campionaria m_x sia minore di 46,08 o maggiore di 53,92 e che induca quindi a rigettare come falsa l'ipotesi vera.

Riepilogando, si consideri un esperimento campionario nel quale siano estratti campioni successivi di n osservazioni ciascuno da una popolazione di unità statistiche avente media μ e deviazione standard σ. Approssimativamente il 100α percento dei campioni avrà una media m_x minore di $(\mu_0 - z_{\alpha/2} . \sigma_m)$ o maggiore di $(\mu_0 + z_{\alpha/2} . \sigma_m)$ e per questi la regola di decisione sopra indicata conduce all'errore di *rigettare* $\mu = \mu_0$ *come la media vera* della popolazione. Per circa il 100(1-α)percento dei campioni la media m_x sarà compresa tra $(\mu_0 - z_{\alpha/2} . \sigma_m)$ e $(\mu_0 + z_{\alpha/2} . \sigma_m)$ e per questi casi correttamente si deciderà di *accettare l'ipotesi che sia* $\mu = \mu_0$. Si deve riconoscere che per questi casi l'ipotesi è corretta, ma nondimeno ci si aspetta comunque di sbagliare nelle conclusioni relative all'ipotesi un certo numero percentuale di volte, in questo esempio nel 100α percento dei casi. Naturalmente, nelle reali applicazioni ai problemi di solito si deve decidere di accettare o rigettare l'ipotesi sulla base di un solo campione di osservazioni e quindi di correre un 100α percento di rischio in ogni singola conclusione.

I valori possibili di m_x che determinano il rigetto dell'ipotesi formano la *regione di rigetto* o *regione critica*. In questo caso la regione critica della verifica dell'ipotesi $\mu = \mu_0$ è $m_x < (\mu_0 - z_{\alpha/2} . \sigma_m)$ e $m_x > (\mu_0 + z_{\alpha/2} . \sigma_m)$. La

probabilità di trovare un singolo campione che ha m_x nella regione critica *se l'ipotesi* $\mu = \mu_0$ *è vera* è pari al 100α percento. Questa probabilità è chiamata *livello di significatività* del test.

Poiché il valore di α può essere scelto da chi esegue l'esperimento e la sua scelta determinerà in parte la *regione di accettazione o di rigetto dell'ipotesi*, esso deve essere fissato prima che l'esperimento venga iniziato. *Se può essere oggetto di serie preoccupazioni il rigetto di un'ipotesi vera, il rischio* α *di compiere questo errore deve essere reso piccolo. Se è di grande interesse che una ipotesi sia rigettata anche se vi è solo qualche dubbio sulla sua veridicità, potremo scegliere un* α *più grande*. Una convenzione frequentemente adottata consiste nel ritenere il risultato *significativo* se l'ipotesi viene rigettata con α= 0,05 e *molto significativo* se essa viene respinta con α= 0,01.

4. Errore β o del II tipo (accettare come vera un'ipotesi falsa)

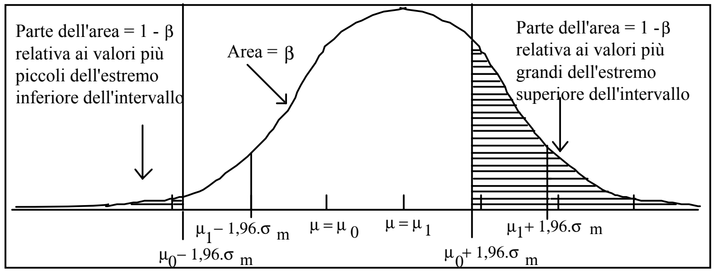
Oltre al possibile errore α di rigettare l'ipotesi quando essa è vera, vi è anche la possibilità di commettere un errore β, accettando per buona l'ipotesi quando essa è falsa. La notazione β viene qui usata sia per indicare l'errore stesso sia per indicare la probabilità che si verifichi l'errore, come in precedenza, quando si è usata la notazione α sia per indicare l'errore che può essere commesso sia per indicare la probabilità di commettere l'errore stesso.

Per un fissato numero n di osservazioni, se si sceglie a piacere α*, automaticamente sarà determinato anche* β. Inoltre, *per una data dimensione n del campione, una diminuzione in* α *aumenterà* β. *Se si vuole diminuire sia* α *che* β*, si deve aumentare il numero delle osservazioni*.

Come illustrazione dell'errore β, si consideri il test del grafico relativo all'ultimo esempio fatto, dove si è convenuto di rigettare il valore $\mu = \mu_0$ se m_x era minore di $(\mu_0 - 1{,}960 . \sigma_m)$ o maggiore di $(\mu_0 + 1{,}960 . \sigma_m)$. L'area tratteggiata in tale grafico corrisponde al valore di α ed è il 5 percento dell'area complessiva sotto la curva.

Ora si supponga che l'ipotesi non sia vera (cioè il campione non proviene da una popolazione avente media μ) e quindi si ammetta che sia $\mu = \mu_0$, mentre la media della distribuzione campionaria di m_x è $\mu = \mu_1$ con deviazione standard, come in precedenza, uguale a σ_m.

Il grafico che segue mostra questa distribuzione campionaria di m_x . In esso l'area tratteggiata rappresenta la probabilità di rigetto dell'ipotesi $\mu=\mu_0$. Poiché i campioni considerati derivano da una popolazione con $\mu = \mu_1$, il rigetto dell'ipotesi $\mu = \mu_0$ non sarà un errore. La dimensione dell'area tratteggiata è $(1 - \beta)$ e naturalmente β è la dimensione dell'area non tratteggiata.



La dimensione dell'area non tratteggiata viene determinata con le due fasi che seguono:

- *calcolo dei valori delle ascisse* con le espressioni seguenti:

$$z_1 = \frac{(\mu_o - 1{,}96.\sigma_m) - \mu_1}{\sigma_m} \quad \text{e} \quad z_2 = \frac{(\mu_o + 1{,}96.\sigma_m) - \mu_1}{\sigma_m}$$

- *ricerca dei valori che corrispondono a queste ascisse nella tavola dei valori cumulati della distribuzione normale* e quindi si trova il valore di β.

Per illustrazione, si consideri un esperimento campionario, su una popolazione avente media vera 52, media ipotetica 50 e σ = 10, dalla quale si estraggano successivi campioni di numerosità n = 25. Approssimativamente l'83 percento dei campioni avrà m_x compreso tra 46,08 e 53,92 e in questi casi si potrà fare l'errore di accettare μ = 50. Approssimativamente il 17 percento dei campioni avrà m_x minore di 46,08 o maggiore di 53,92 e in questi casi si potrà correttamente concludere che la media non è 50. Le affermazioni delle due conclusioni alternative sono le seguenti:

1. Se m_x è maggiore di 53,92 o minore di 46,08, si deve rigettare l'ipotesi che sia μ = 50. Se μ è veramente uguale a 50 si deve prendere la decisione di rigettare l'ipotesi (naturalmente sbagliata) il 5 percento delle volte. Quindi α = 0,05.
2. Se m_x è compreso tra 46,08 e 53,92, si deve accettare l'ipotesi μ = 50. Se in realtà è μ = 52, si accetterà μ = 50 (ancora una decisione errata) l'83 percento delle volte. Quindi β = 0,83.

Si deve notare che nell'illustrazione precedente avrebbero potuto essere usati anche altri valori diversi da 52. La discussione di questo punto è rinviata al paragrafo 8 che segue.

Generalmente si indica con il simbolo H_0 l'ipotesi nulla e con il simbolo H_1 l'ipotesi contraria. L'ipotesi nulla equivale ad ammettere che la statistica del campione segua una distribuzione con media uguale a quella della popolazione da cui si pensa che il campione sia stato estratto. Ciò che si dimostra effettivamente quando un campione è estratto proprio dalla popolazione dalla quale si considera estratto.

Il procedimento che si adotta per verificare statisticamente un'ipotesi può essere riepilogato come segue:

1. Non si conosce la media della popolazione μ .
2. Si formula l'ipotesi nulla, che, cioè, la media sia $\mu = \mu_0$.
3. Si decide di rigettare l'ipotesi quando vi sono risultati la cui probabilità di verificarsi quando l'ipotesi è vera sia solo del 100α percento. (Nell'esempio considerato si rigetta l'ipotesi quando $m_x < \mu_0 - 1{,}96\ \sigma_m$ oppure quando $m_x > \mu_0 + 1{,}96\ \sigma_m$).
4. Si calcola m_x da un campione di n osservazioni per decidere se accettare o rigettare l'ipotesi avanzata.
5. Se si osserva che il campione ha condotto al risultato m_x minore di $(\mu_0 - z_{\alpha/2}\ \sigma_m)$ o maggiore di $(\mu_0 + z_{\alpha/2}\ \sigma_m)$, si rigetta l'ipotesi; se si osserva un valore di m_x compreso tra questi valori, si accetta l'ipotesi avanzata.

In molte applicazioni statistiche l'errore del II tipo o errore β non viene controllato, ma anche in tale eventualità lo sperimentatore non solo deve essere cosciente che questo errore esiste, ma deve anche avere una idea di quanto grande esso possa essere.

5. Verifica di una ipotesi statistica

Dato che la procedura usata per l'esperimento sulle medie è una procedura standard, essa sarà qui riassunta di nuovo, con riferimento al tema generale della *verifica di una ipotesi statistica*. Le tappe da sviluppare sono:

1. *Formulazione dell'ipotesi nulla.*

2. *Specificazione dei livelli di rischio o errori α e β.* Questi due parametri determineranno il numero di osservazioni che si devono considerare per calcolare la statistica che è stata scelta. Nella pratica corrente più di frequente sono specificati i valori di α e di n.

3. *Si specificano i valori di una data statistica che definiscono la regione critica*, stabilendo quali di essi indurranno a rigettare l'ipotesi e quali invece indurranno ad accettarla.

4. *Si calcola il valore della statistica* sulla base delle osservazioni sperimentali.

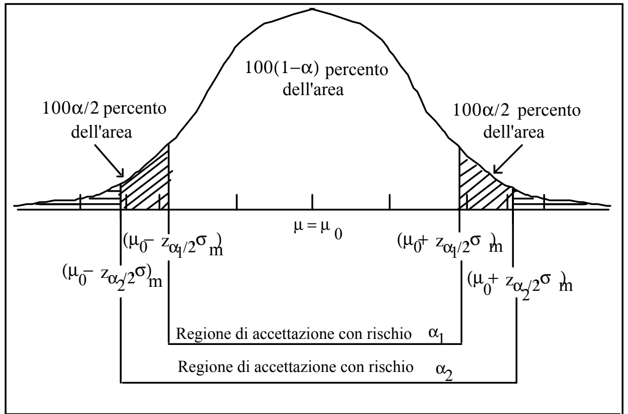
5. *Si accetta o si rigetta l'ipotesi*, a seconda che il valore ottenuto per la statistica cada dentro o al di fuori della regione critica.

Questa procedura è *standardizzata*, nel senso che essa si può considerare valida ed identica per tutte le analisi tendenti ad accertare il livello di accettabilità del risultato del confronto tra una determinata statistica relativa al campione e la stessa statistica relativa all'intera popolazione da cui si ritiene che quel campione sia stato estratto.

6. Effetto delle variazioni di α sulla regione critica e su β.

Si consideri una popolazione normale, con deviazione standard σ nota e si voglia verificare una ipotetica media $\mu = \mu_0$. Se si usa un campione casuale di n osservazioni e si rigetta l'ipotesi $\mu = \mu_0$. se m_x è a più di $z_{\alpha/2}$ unità di deviazione standard da μ_0, si compie un errore pari ad α e si ha un livello di significatività di 100α percento. Una volta scelto $z_{\alpha/2}$, si trova α dalla tavola delle cumulate della distribuzione normale o, per converso, se si sceglie α la regione critica (m_x a più di $z_{\alpha/2}$ unità di deviazione standard da μ_0) viene trovata

usando la stessa tavola. Poiché si può scegliere qualunque valore di α si voglia, si dovranno considerare gli effetti che differenti scelte di α possono determinare sia sulla regione critica sia su β.



Differenza tra le regioni critiche corrispondenti a due livelli di rischio, con $\alpha_2 < \alpha_1$

Se scegliamo due valori α_1 e α_2 tali che sia $\alpha_2 < \alpha_1$ con il grado di rischio α_2 i valori critici di m_x saranno lontani da μ_0 più di quanto lo sono con un grado di rischio pari ad α_1 , come ben si vede nel grafico precedente.

Si ricorda che la regione critica è complementare rispetto alla regione di accettazione. Pertanto, dal grafico precedente risulta che al diminuire dell'errore del I tipo α (probabilità di rigettare come falsa una ipotesi vera) cresce la regione di accettazione e quindi diminuisce la regione critica.

Se ci si riferisce agli esempi fatti sinora, si può vedere che se α_1 è scelto minore del 5 percento i valori critici di m_x saranno lontani da μ_0 più di $z_{\alpha_1/2}$ = 1,96 unità di deviazione standard, mentre se si sceglie α_2 minore dell'1 percento i valori critici di m_x saranno lontani da μ_0 più di $z_{\alpha_2/2}$ = 2,576 unità di deviazione standard. Infatti, se, come si è fatto già per l'ipotesi $\alpha_1 = 0,05$, si sceglie un livello di errore $\alpha_2 = 0,01$ si determinano due valori tali che l'area sotto la distribuzione campionaria di m_x a sinistra di un valore è 0,005 e a destra dell'altro è pure 0,005.

Utilizzando la tavola della distribuzione normale cumulativa, si trova che i due valori cercati sono $(\mu_0 - 2,576\ \sigma_m)$ il primo e $(\mu_0 + 2,576\ \sigma_m)$ il secondo. In questo caso, la regione critica relativa ad un livello di rischio dell'uno percento è qualunque m_x minore di $(\mu_0 - 2,576.\sigma_m)$ o maggiore di $(\mu_0 + 2,576\ \sigma_m)$. Questi valori sono indicati nel grafico che precede.

Per il momento, si può restringere la discussione solo ai due valori del livello di significatività $\alpha_1 = 0,05$ ed $\alpha_2 = 0,01$, ma è chiaro che essa resta valida anche per qualunque altro livello.

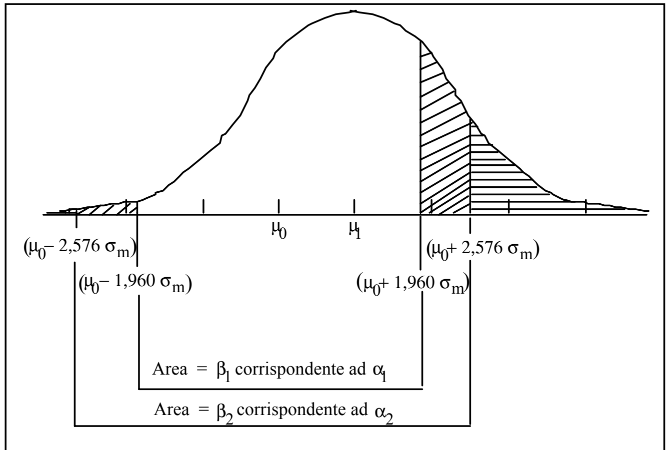
In generale, si può affermare che: *più è piccolo il rischio α che si vuole correre, più la regione critica sarà lontana dalla media suggerita* (cioè si allarga la regione di accettazione).

Ora di nuovo, come nel paragrafo 3, si supponga che il campione proviene in realtà da una popolazione con media μ_1 . Utilizzando il test $\alpha = 0,01$ dell'ipotesi $\mu = \mu_0$, si accetti $\mu = \mu_0$, quindi commettendo un errore β, se m_x è compreso tra $(\mu_0 - 2,576\ \sigma_m)$ e $(\mu_0 + 2,576\ \sigma_m)$. La distribuzione campionaria di m_x è ora considerata normale con media μ_1 e deviazione standard σ_m ; cosicché β è l'area sotto la curva normale che risulta compresa tra $(\mu_0 - 2,576\ \sigma_m)$ e $(\mu_0 + 2,576\ \sigma_m)$. Per trovare β, si calcolino i due valori seguenti:

$$z_1 = \frac{(\mu_o - 2,576.\sigma_m) - \mu_1}{\sigma_m} \quad \text{e} \quad z_2 = \frac{(\mu_o + 2,576.\sigma_m) - \mu_1}{\sigma_m}$$

e si faccia riferimento alla tavola della distribuzione normale cumulativa per ottenere il valore di β.

Nel grafico che segue è indicato come varia l'area corrispondente a β , al variare del livello di α. Si noti che un rischio inferiore (α_2=0,01 invece di α_1=0,05) di rigettare una ipotesi vera ($\mu = \mu_0$) determina un rischio maggiore di non riconoscere l'ipotesi come falsa quando qualche ipotesi alternativa ($\mu = \mu_1$ in questo caso) è vera ($\beta_2 > \beta_1$).



Differenza tra gli errori di II tipo β_1 e β_2 corrispondenti ad errori del I tipo α_1 e α_2

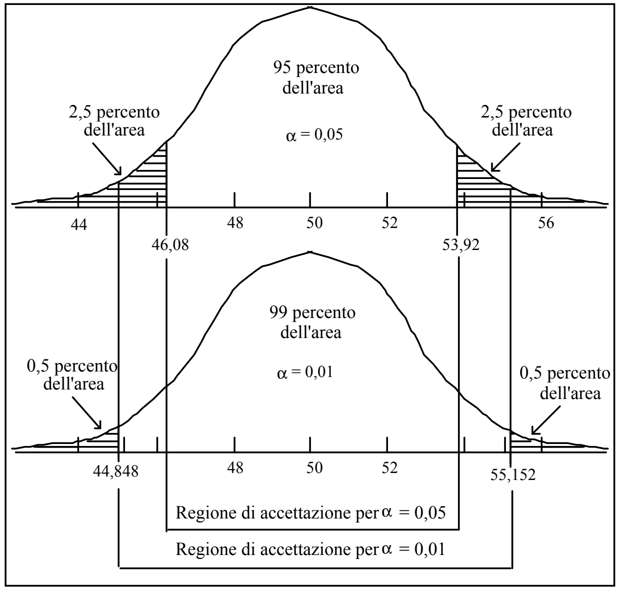
Come applicazione, si supponga che la popolazione sia normale, con deviazione standard $\sigma = 10$ e che si voglia verificare una ipotetica media $\mu = 50$. Si è visto che se si ha un livello di significatività del 5 percento e si usa un campione casuale di $n = 25$ osservazioni , per cui è $\sigma_m=\sigma/\sqrt{n}=10/5=2$, si rigetta l'ipotesi $\mu = 50$ se m_x è a più di 1,96 unità di deviazione standard, ad esempio 3,92 unità, da 50. Una volta scelto 1,96, si trova $\alpha = 0{,}05$ dalla tavola delle cumulate della distribuzione normale o, per converso, se si sceglie $\alpha = 0{,}05$ la regione critica (m_x a più di 1,96 unità di deviazione standard da 50) è trovata usando la stessa tavola. Poiché si può scegliere qualunque valore di α si voglia, dovranno essere considerati quali sono gli effetti che possono essere determinati da differenti scelte di α sul livello dell'errore di II tipo.

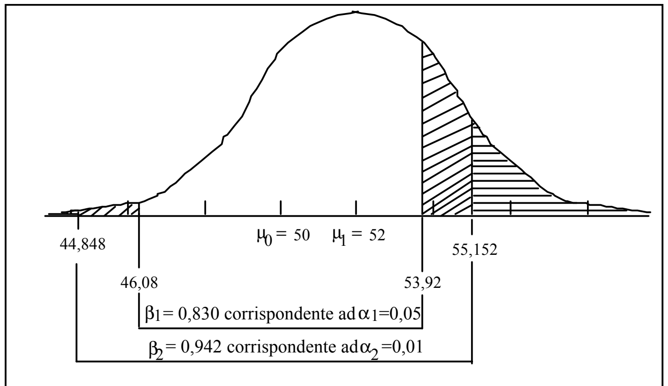
Nel grafico riportato nella pagina che segue si vede che se si sceglie α minore del 5 percento i valori critici di m_x saranno lontani da 50 più di 1,96 unità di deviazione standard e quindi coincideranno con i valori 46,08 (=50-1,96.2) e 53,92 (=50+1,96.2). Se invece si sceglie un livello $\alpha = 0{,}01$ si determinano due valori tali che l'area sotto la distribuzione campionaria di m_x a sinistra di un valore è 0,005 e a destra dell'altro è 0,005. Dalla tavola della distribuzione normale cumulativa si vede che i due valori cercati sono $(\mu_0 - 2{,}576\ \sigma_m) = 50 - 2{,}576.2 = 44{,}848$ e $(\mu_0 + 2{,}576\sigma_m) = 55{,}152$. Allora la regione critica dell'uno percento è qualunque m_x minore di 44,848 o maggiore di 55,152.

Ora di nuovo, come nel paragrafo .3, si supponga che il campione proviene in realtà da una popolazione con media 52 . Usando il test $\alpha = 0{,}01$ dell'ipotesi $\mu = 50$, si accetti $\mu = 50$, quindi commettendo un errore β, se m_x è compreso tra 44,848 e 55,152. La distribuzione campionaria di m_x è ora considerata normale con media 52 e deviazione standard 2; cosicché β è l'area sotto la curva normale compresa tra 44,848 e 55,152. Per trovare β, si calcolino

$$z_1 = \frac{44{,}848 - 52}{2} = -3{,}576 \qquad \text{e} \qquad z_2 = \frac{55{,}152 - 52}{2} = +1{,}576$$

e si faccia riferimento alla tavola della distribuzione normale cumulativa per ottenere il valore di β= 0,942 - 0,000 = 0,942. Nel grafico che segue è indicato come varia l'area corrispondente a β, al variare del livello di α.





Differenza tra gli errori di II tipo β_1 e β_2 corrispondenti ad errori del I tipo α_1 e α_2

Si noti che un rischio inferiore (0,01 invece di 0,05) di rigettare una ipotesi vera (μ = 50) determina un rischio maggiore (0,942 invece di 0,830) di non riconoscere l'ipotesi come falsa quando qualche ipotesi alternativa (μ = 52 in questo caso) è vera.

L'osservazione che β aumenta man mano che α diminuisce è vera in generale e lo sperimentatore può scegliere di variare il livello di significatività α del suo test per determinare un corrispondente cambiamento in β.

Un valore piccolo di α è certamente desiderabile, ma il prendere α troppo piccolo può avere per conseguenza un β così grande che raramente si riconosce come falsa l'ipotesi quando essa è veramente falsa.

La scelta del livello accettabile sia per l'errore del I tipo sia per quello del II tipo ovviamente dipende dalla sensibilità e dall'esperienza particolare del ricercatore, tenuto conto delle caratteristiche del fenomeno studiato.

7. Effetto delle variazioni di n sulla regione critica

Finora, in questo capitolo si è considerato come unico esempio quello di campioni formati da n = 25 unità. Ma, dato che $\sigma_m = \sigma/\sqrt{n}$, al variare di n cambia pure la regione critica e, poiché n figura nella formula di σ_m al denominatore, *un aumento nella dimensione n del campione causerà una diminuzione in* σ_m e quindi un aumento della regione critica o di rigetto.

Nell'esempio con σ = 10 e la media ipotetica uguale a 50, si è deciso di accettare l'ipotesi al livello α = 0,05 se la media campionaria per un campione di 25 osservazioni era compresa tra 46,08 e 53,92, cioè, se m_x era entro 1,96.σ_m dalla media suggerita 50.

Se si fosse considerato un campione di dimensione n = 100, allora si sarebbe avuto un valore $\sigma_m = 10/\sqrt{100} = 1$ e ci si sarebbe potuto aspettare che il 95 percento dei campioni avesse una media m_x compresa entro un valore di 1,96. σ_m = 1,96 unità dalla media μ della popolazione. Allora, per α = 0,05 la regione critica corrispondente avrebbe avuto valori di m_x minori di 50 - 1,96 = 48,04 e maggiori di valore 50 + 1,96 = 51,96.

Per lo stesso livello di significatività la regione di accettazione è più piccola per campioni di dimensione maggiore.

L'intervallo è accorciato di una quantità appropriata per tener conto del fatto che vi è una probabilità più elevata che la media m_x di un campione più grande sia più vicino a μ che non la media di un campione più piccolo.

Tavola VI.1 - Regioni critiche per verificare l'ipotesi $\mu = 50$, dato $\sigma = 10$ e rigettare l'ipotesi se m_x non è vicino a 50

	$\alpha = 0{,}05$		$\alpha = 0{,}01$	
N	$m_x <$ $50\text{-}1{,}960\sigma_m$	$m_x >$ $50+1{,}960\sigma_m$	$m_x <$ $50\text{-}2{,}576\sigma_m$	$m_x >$ $50+2{,}576\sigma_m$
1	30,40	69,60	24,24	75,76
3	38,68	61,32	35,13	64,87
5	41,23	58,77	38,48	61,52
10	43,80	56,20	41,85	58,15
25	46,08	53,92	44,85	55,15
50	47,23	52,77	46,36	53,64
100	48,04	51,96	47,42	52,58
400	49,02	50,98	48,71	51,29
1000	49,38	50,62	49,19	50,81

La tabella VI.1 mostra un insieme di regioni critiche per diversi valori di n e per $\alpha = 0{,}05$ e $\alpha = 0{,}01$. Se z_α rappresenta il percentile *100.α* della distribuzione normale unitaria definita dalla tavola della distribuzione normale cumulativa, la regione critica per verificare l'ipotesi $\mu = \mu_o$ al livello di significatività α è definita dalle due seguenti diseguaglianze che rappresentano gli estremi dell'intervallo entro il quale dovrebbe essere compreso il valore medio della popolazione:

$$M(X) < \mu_o + z_{\frac{1}{2}\alpha} \cdot \frac{\sigma}{\sqrt{N}} \qquad \text{e} \qquad M(X) > \mu_o + z_{1-\frac{1}{2}\alpha} \cdot \frac{\sigma}{\sqrt{N}}$$

8. Effetto delle variazioni di α e di n su β

Ricordando e riassumendo in parte quanto si è visto fino a questo punto, nei paragrafi precedenti si è notato che:

1. Se α, errore del I tipo, cambia da 0,05 a 0,01 e se n, dimensione del campione, rimane fisso si ha per effetto l'aumento di β, errore del II tipo.

2. Se si mantiene fisso α, aumentando n aumenta la dimensione della regione critica (cioè, si considera significativa una minore deviazione dalla ipotesi).

Ora si consideri l'effetto che ha su β l'aumento di n quando α è mantenuto fisso. Nella Tavola VI.1 sono state raccolte le regioni critiche per verificare l'ipotesi che una popolazione con $\sigma = 10$ abbia una media $\mu = 50$. Le regioni sono raccolte per diversi valori di n e per $\alpha = 0{,}05$ ed $\alpha = 0{,}01$.

Come nel paragrafo 3 di questo capitolo, si supponga di avere estratto un campione da una popolazione con media 52 al posto dell'ipotetico valore di 50. Si può allora calcolare il valore di β corrispondente a ciascuno dei 18 campioni considerati (due valori di α e nove valori di n). Per $n = 25$ si è già trovato $\beta = 0{,}830$ per $\alpha = 0{,}05$ e $\beta = 0{,}942$ per $\alpha = 0{,}01$.

La Tavola VI.2 che segue elenca i valori di β per valori di $\mu = 52$ corrispondenti alla dimensione n del campione ed ai valori di α usati nella Tavola VI.1.

Si noti che, per una determinata alternativa ed un dato livello di significatività α, la probabilità β di non dichiarare falsa una ipotesi che è falsa diminuisce con il crescere di n.

Tavola VI.2 - Valori di β per l'alternativa $\mu = 52$ nella verifica dell'ipotesi $\mu = 50$.

N	$\alpha = 0{,}05$	$\alpha = 0{,}01$
1	0,945	0,988
3	0,936	0,/985
5	0,927	0,912
10	0,903	0,973
25	0,830	0,942
50	0,707	0,877
100	0,484	0,717
400	0,021	0,077
1000	0,00001	0,0001

Da quanto fin qui esposto, è possibile tentare un riepilogo dei risultati conseguiti, nel modo che segue.

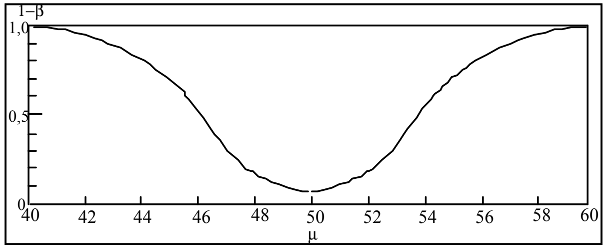
Se viene formulata una singola alternativa all'ipotesi, un test statistico avrà le due proprietà che seguono:

- per α fissato la scelta di n più piccolo aumenta β, mentre scegliere n più grande diminuisce β;

- per n fissato, la scelta di un α più piccolo aumenta β, mentre scegliere α più grande diminuisce β.

9. Dipendenza di β dall'alternativa considerata

Finora si è considerata la sola alternativa $\mu = 52$. Altre alternative avrebbero altri valori di β. Si supponga ancora di voler verificare l'ipotesi $\mu = 50$ con $\sigma = 10$ ed $\alpha = 0{,}05$ con un campione di $n = 25$ osservazioni. La regione critica dalla Tavola VI.1 richiede il rigetto dell'ipotesi se m_x non è compresa tra 46,08 e 53,92. Ora, se invece si ha $\mu = 53$, si calcoli, come nel paragrafo 3, $z_1 = (46{,}08-53)/2 = -3{,}46$ e $z_2 = (53{,}92-53)/2 = 0{,}46$ per cui nella tavola della distribuzione normale cumulativa si trova $\beta = 0{,}677 - 0{,}000 = 0{,}677$. Questo valore di β per una media $\mu = 53$ è più piccolo di $\beta = 0{,}830$ per $\mu = 52$.

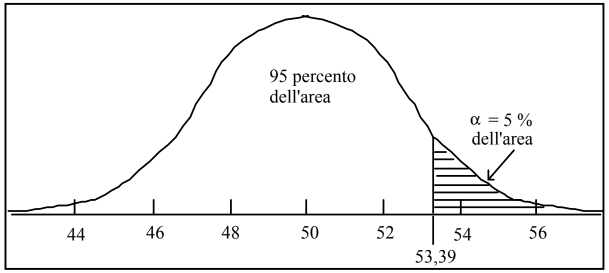


Naturalmente ci si aspetta che il test abbia una minore probabilità di fallire nel rigettare l'ipotesi $\mu = 50$ quando le osservazioni provengono da una popolazione con $\mu = 53$ che quando esse provengono da una popolazione con $\mu = 52$. In altre parole, ci si aspetta di avere una probabilità maggiore di rigettare l'ipotesi $\mu = 50$ quando un campione è tratto da una popolazione con media $\mu = 53$ che quando esso è estratto da una popolazione con media $\mu = 52$.

Per qualunque valore di n e di α fissati si può calcolare β per una serie di valori di μ. Per l'esempio svolto in questo capitolo con un $n = 25$ ed $\alpha =$ 0,05, i valori calcolati di β sono stati utilizzati per disegnare la curva del grafico seguente, ove μ figura sull'asse orizzontale ed $(1 - \beta)$ sull'asse verticale. La curva è chiamata la "*curva di potenza*" del test statistico usato.

10. Test ad una coda

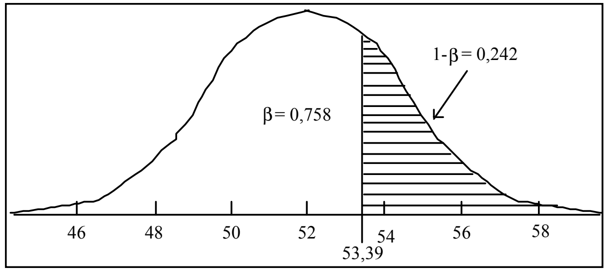
In alcuni casi può essere preferibile rigettare una media ipotetica solo se la m_x osservata è insolitamente distante in una data direzione: per esempio, se si osserva un valore della media m_x insolitamente grande. Il grafico che segue mostra la distribuzione di m_x se l'ipotesi è corretta, con la regione critica indicata.



Se l'ipotesi afferma che è $\mu = 50$, con $\sigma = 10$, $n = 25$ e l'errore del I tipo $\alpha =$ 0,05, si possono usare come regione critica i valori di m_x maggiori di $50 + 1{,}645.\sigma/\sqrt{n}$ = 50 + 1,645(2) = 53,39. Dalla tavola della distribuzione normale cumulativa si prende la quantità 1,645 dato che vi è una probabilità del 95 percento che la $z = (m_x - 50)/2$ sia minore di 1,645 se realmente m_x è tratto da una popolazione con $\mu = 50$.

La discussione degli effetti sull'errore β per variazioni in α e nella dimensione n del campione può essere svolta per questa regione critica *ad una coda* in modo simile a quella fatta nei paragrafi da 3 a 8 per la regione critica *a due code*. Infatti, se $\mu = 52$, la distribuzione campionaria di m_x ha media 52. Questa distribuzione è rappresentata nel grafico che segue in cui è evidenziata l'area che rappresenta l'errore β.

La probabilità β di accettare una media $\mu = 50$, quando nella realtà è $\mu = 52$, è l'area non tratteggiata. Il valore di β è calcolato come detto nel paragrafo 3, notando che per m_x = 53,39 si ha per la z un valore che risulta pari a $z = (53{,}39-52)/2 = 1{,}39/2 = 0{,}70$.



Dalla tavola della distribuzione normale cumulativa si vede che l'area è $\beta = 0{,}758$. Le stesse conclusioni riassunte nel paragrafo 7 possono essere ripetute per il test ad una coda.

Quando si usa la regione critica ad una coda, in realtà si considerano soltanto medie di popolazioni alternative più grandi di 50. Questo fatto può essere indicato nella definizione dell'ipotesi come $\mu \leq 50$, che si legge "*μ è minore o uguale a 50*". Si suppone qui che i valori di μ minori di 50 siano accettabili, mentre i valori di m_x che si considerano critici sono soltanto valori insolitamente più grandi di 50.

Ovviamente la discussione precedente può essere svolta di nuovo usando una regione critica che consista solo di valori piccoli di m_x per verificare l'ipotesi che sia $\mu \geq 50$.

Alcuni esempi nei quali sembra ragionevole considerare regioni critiche ad una coda sono i seguenti:

1. L'acqua potabile può essere controllata attraverso la conta dei batteri e se il risultato fornisce un numero troppo alto di batteri si può concludere con il rigetto dell'ipotesi di potabilità dell'acqua.
2. L'effetto di un nuovo fertilizzante sulla resa (produzione per ettaro) di una coltura può essere verificato e il vecchio fertilizzante può essere sostituito solo se con il nuovo si ha per risultato una resa significativamente più alta.
3. Campioni di calcestruzzo possono essere controllati per la resistenza e rifiutati solo se essi hanno una resistenza insolitamente bassa.
4. Il contenuto di vasi di burro di arachide etichettati "Peso netto otto etti" può essere pesato ed il prodotto può essere rifiutato come impropriamente imballato solo se si trova che il peso è troppo basso.

11. Verifica delle ipotesi in una popolazione dicotomica

Si consideri la distribuzione campionaria di m_x per campioni di dimensione n estratti da una popolazione dicotomica che ha individui marcati 1 nella proporzione p ed individui marcati 0 nella proporzione $q = (1 - p)$. La distribuzione campionaria di m_x in questo caso ha una media uguale a p, una deviazione standard uguale a $\sigma_m = \sigma/\sqrt{n} = \sqrt{pq/n}$ ed è approssimativamente normale per grandi valori di n. Questi risultati possono essere usati per verificare un ipotetico valore di p.

Si supponga di voler verificare l'ipotesi che una certa popolazione di mosche cresciute in un laboratorio sia formata dal 50 percento di maschi e 50 percento di femmine. Questo è un test dell'ipotesi $p = 0{,}50$. Si supponga di prendere un campione casuale di 100 mosche dalla popolazione e di contare la presenza di 40 maschi. Può questo risultato condurre al rigetto dell'ipotesi che sia $p = 0{,}50$ al livello di significatività del 5 percento?

In questo esempio si hanno i seguenti dati forniti dalle osservazioni di base: $n = 100$, $m_x = 0{,}40$, $\alpha = 0{,}05$, e $z = (m_x - \mu)/\sigma_m = (0{,}40-0{,}50)/0{,}05 = -2{,}0$ con

$$\sigma_m = \sqrt{\frac{p.q}{n}} = \sqrt{\frac{0,5x0,5}{100}} = \sqrt{\frac{0,25}{100}} = \frac{0,5}{10} = 0,05.$$

Un test a due code determinerebbe il rigetto dell'ipotesi se m_x fosse più di 1,96 unità di deviazione standard diverso dalla media ipotetica; quindi, in questo caso si rigetta l'ipotesi, perchè 40/100=0,40<0,50±0,098.

Se si adotta il test ad una coda, per una media osservata $m_x = 0{,}40$ dovrebbe essere rigettata anche un'ipotesi che sia $p \geq 0{,}50$ poiché la regione di rigetto è formata da qualunque valore della media $m_x < (\mu - 1{,}645.\sigma_m) = 0{,}50 - 1{,}645(0{,}05) = 0{,}50 - 0{,}082 = 0{,}418$.