ECONOMIA

METODI STATISTICI PER L'ANALISI ECONOMICA E AZIENDALE

DALLA STIMA PUNTUALE ALLA STIMA INTERVALLARE

FABRIZIO ANTOLINI
fantolini@unite.it

LA STIMA PUNTUALE E STIMATORE

Dato un campione statistico, immaginiamo di voler calcolare la media di un determinato valore attraverso il calcolo della media campionaria. Il valore che otterremo può essere definito una stima puntuale, dal momento che attribuisce a un'incognita un determinato valore. Questa è una definizione semplice e intuitiva per comprendere cosa sia la stima puntuale, ma cerchiamo di andare un po' oltre.

- Sia X una v.c. che rappresenta un carattere osservato su una popolazione. Supponiamo che la v.c. sia definita da una funzione di probabilità dipendente dal parametro incognito θ.
- Sia $X_1,\dots,X_n$ la realizzazione campionaria di dimensione n.

L'obiettivo consiste in:

Ottenere attraverso un'opportuna funzione delle osservazioni campionarie $t = t(x_1, \dots, x_n)$ una stima di θ.

La stima $t = t(x_1, \dots, x_n)$

può essere considerata come una realizzazione della variabile casuale $T = t(X_1, \dots, X_n)$

chiamata stimatore di θ.

- Uno stimatore è una v.c. utilizzata per stimare un determinato parametro θ della popolazione e risulta essere una funzione delle v.c. campionarie
- Il valore assunto dallo stimatore in corrispondenza di un particolare campione viene detto stima puntuale.

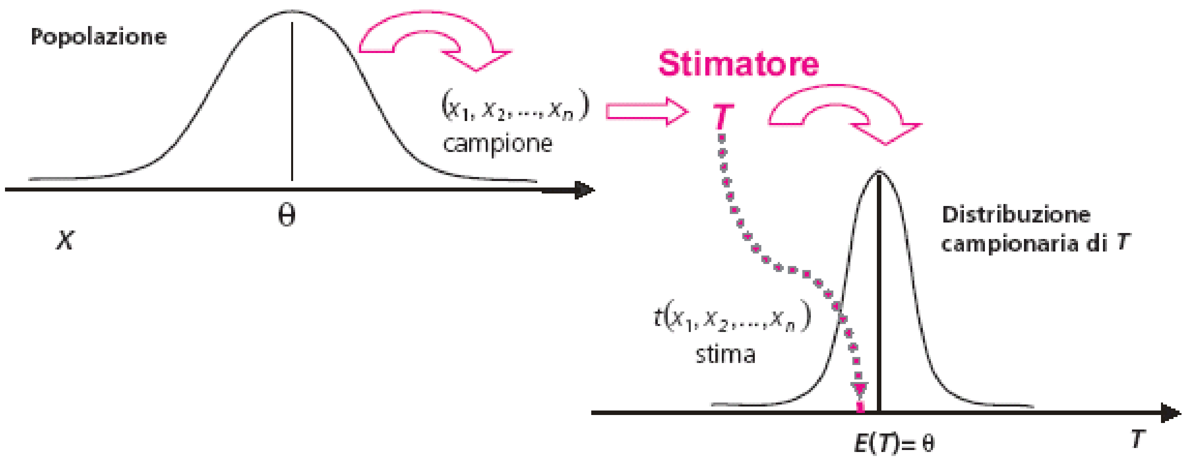
LA STIMA PUNTUALE E STIMATORE

Esempio

Se abbiamo un campione osservato composto da $X_1 = 10; X_2 = 22; X_3 = 40$

- Lo stimatore $\boldsymbol{T}$ della media campionaria sarà espresso da $\bar{X} = \frac{1}{3}\sum_{i=1}^{3} X_i$
- La stima puntuale $\boldsymbol{t}$ della media campionaria sarà 24

Lo stimatore, dipendendo dal campione, è una variabile casuale e quindi possiede una distribuzione campionaria la cui conoscenza permette di capire se lo stimatore scelto produrrà con elevata probabilità stime “vicine al valore vero del parametro.



LE PROPRIETÀ DELLO STIMATORE

Per valutare la "bontà" di uno stimatore T si può guardare alle sue proprietà:

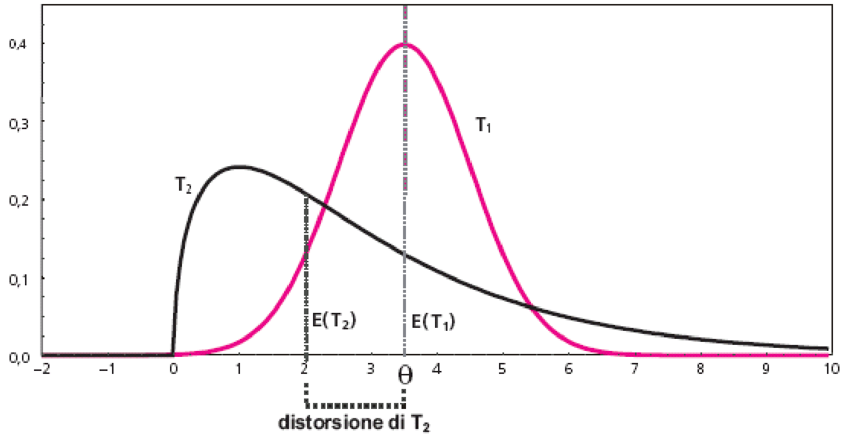
- Correttezza
- Efficienza
- Consistenza

Lo stimatore T è uno stimatore corretto di θ se:

$$E(T) = \theta, \text{ per tutti i possibili valori di } \theta$$

Ne deriva che una misura della distorsione di uno stimatore può essere espressa da:

$$B(T) = (ET) - \theta$$

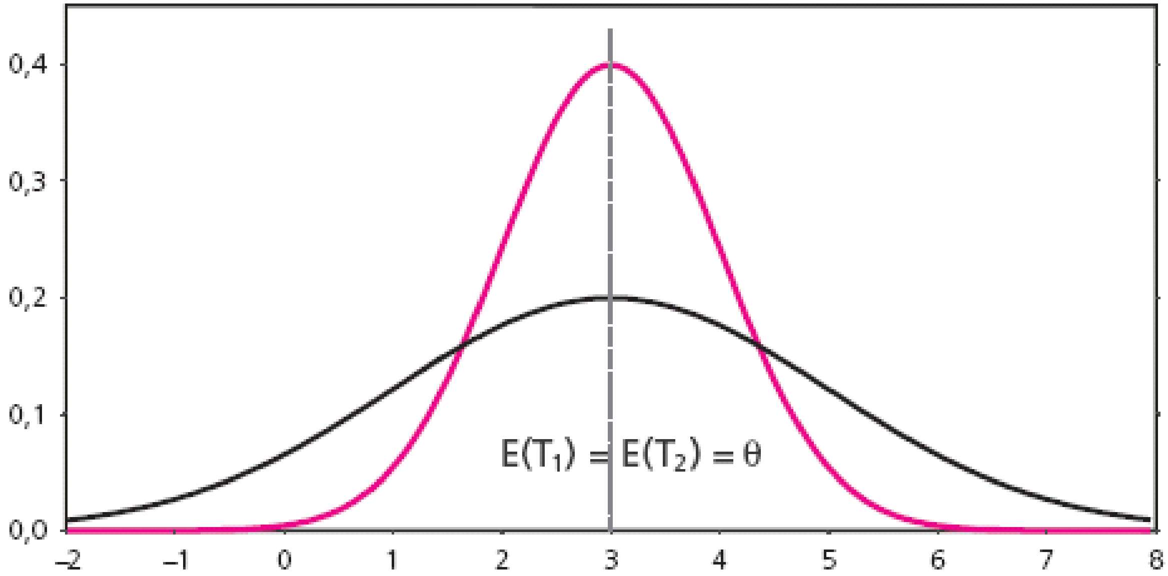


LE PROPRIETÀ DELLO STIMATORE

La proprietà dell'efficienza è una proprietà relativa. Ossia, considerati due stimatori corretti T_1 e T_2 del parametro θ, lo stimatore più efficiente sarà quello che presenta varianza minore e quindi si dirà che T_1 è più efficiente di T_2 se:

$$VAR(T_1) < VAR(T_2)$$

Nella figura sono riportate le distribuzioni campionarie di due stimatori corretti. lo stimatore T_1 (linea rossa) possiede una varianza più piccola di T_1 (linea nera).

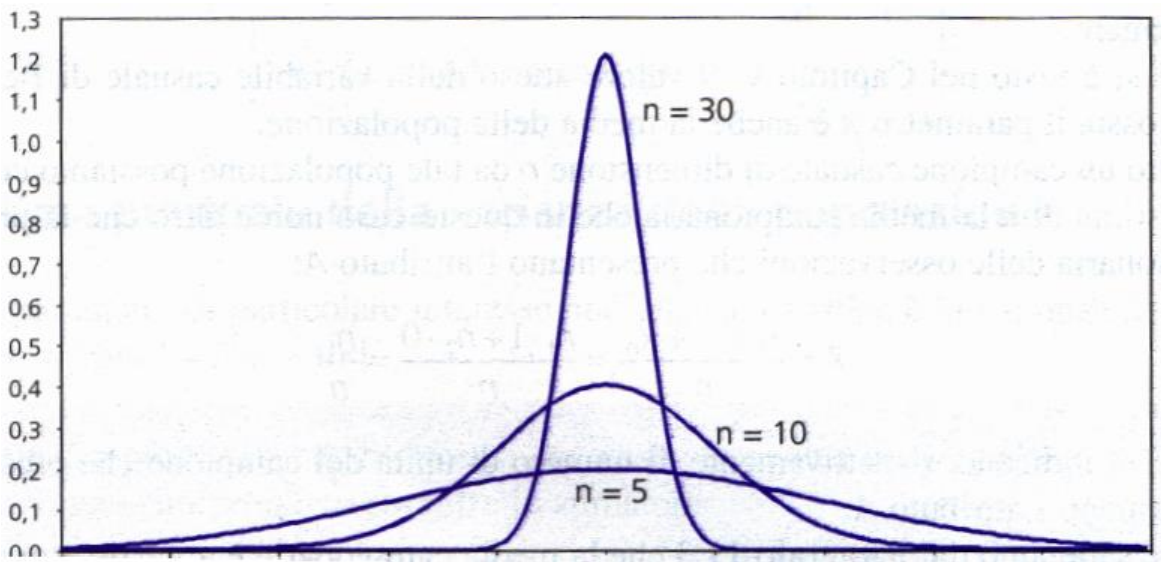


LE PROPRIETÀ DELLO STIMATORE

Si intuisce che nel momento in cui applico lo stimatore alla realizzazione campionaria, il risultato che ottengo dipende anche dalla numerosità del campione.

Un campione di dimensione n_1 fornirà ovviamente un dettaglio informativo sulla popolazione minore rispetto ad un campione di dimensione n_2 se $n_1 < n_2$

La proprietà della consistenza stabilisce dunque che all'aumentare dell'informazione contenuta nel campione, ossia all'aumentare della numerosità del campione, cresce la precisione dello stimatore:



STIMA PUNTUALE E STIMA INTERVALLARE

LA STIMA PUNTUALE

Si è parlato in precedenza della stima puntuale dei parametri ignoti di una data popolazione. Partendo dallo stimatore puntuale e considerando la sua varianza è possibile costruire una stima intervallare, cioè un intervallo di confidenza che contiene il parametro oggetto d'indagine con una probabilità assegnata $P = 1 - \alpha$ (detta livello di confidenza o di fiducia), ove α è il livello di significatività e fornisce il rischio che si corre nel confidare che l'intervallo stimato contenga il parametro incognito della popolazione.

- Sia X una v.c. che rappresenta un carattere osservato su una popolazione. Supponiamo che la v.c. sia definita da una funzione di probabilità $f(x;\theta)$ dipendente dal parametro θ incognito
- Sia $X_1,\dots,X_n$ un campione di dimensione n e $x_1,\dots,x_n$ il corrispondente campione osservato.

L'obiettivo è determinare due statistiche campionarie:

$$L_1 = L_1(X_1,\dots,X_n) \text{ e } L_2 = L_2(X_1,\dots,X_n)$$

Tali che $L_1 < L_2$ per ogni possibile campione e che l'intervallo $[L1;L2]$ contenga il parametro θ con probabilità $1-\alpha$.

L'INTERVALLO DI CONFIDENZA

L'intervallo casuale $[L_1(X_1,\ldots,X_n);\ L_2(X_1,\ldots,X_n)\ ]$ si definisce intervallo di confidenza per il parametro θ se contiene con probabilità $1-\alpha$ il parametro ignoto della popolazione, ossia:

$$Pr[L_1(X_1,\ldots,X_n) \leq \theta \leq L_2(X_1,\ldots,X_n)\,] = 1-\alpha$$

In genere si fissano valori di $1-\alpha$ pari a 0,99; 0,95 e 0,90.

- Se si fissa il valore di α esistono infiniti intervalli che soddisfano la precedente condizione; si dimostra tuttavia che per distribuzioni campionarie simmetriche (anche approssimativamente) l'intervallo migliore (cioè quello con l'ampiezza minore) è quello centrato intorno alla stima puntuale

- È importante osservare a questo punto che la stima mediante intervallo di confidenza è tanto migliore quanto minore è la sua ampiezza. È evidente, infatti, che ai fini del problema di stima avrebbe poco senso dire, ad esempio, che la media delle stature (espresse in cm) degli studenti iscritti al corso di laurea in economia aziendale di una data Università è contenuta nell'intervallo [155, 195] con una probabilità pari a 0,999999

- Aumentando il livello di fiducia $P=1-\alpha$ aumenta l'ampiezza dell'intervallo e si annulla quasi il rischio di trovare un intervallo che non contenga il valore incognito del parametro della popolazione. La riduzione del rischio di commettere una stima errata paga cioè il prezzo di avere un intervallo così ampio che non dà alcuna informazione utile.

LA DETERMINAZIONE DELLA DIMENSIONE DEL CAMPIONE

Secondo l'opinione comune, la bontà dei risultati ottenibili da un campione dipende *unicamente* dal numero degli individui che compongono il campione stesso e non dal *modo* con cui essi sono stati selezionati. La debolezza di questo assunto è già stata dimostrata nelle unità precedenti.

Un'altra opinione comune prevede che la dimensione del campione debba essere proporzionata alla dimensione della popolazione in studio. Le leggi della statistica dimostrano invece che questo assunto è completamente falso. Ad esempio, il fatto di voler fare un sondaggio sugli abitanti di un capoluogo di provincia, su quelli di una grande città o addirittura su tutta la popolazione italiana non ha nessuna influenza sul numero di persone necessario per ottenere un campione rappresentativo.

Si intuisce la necessità, dunque, prima di intraprendere un’indagine, di stabilire quante unità di interesse dovranno essere esaminate per raggiungere con sufficiente attendibilità l’obiettivo desiderato.

Il calcolo della dimensione del campione, più propriamente detta numerosità, è abbastanza complicato e, soprattutto, richiede la conoscenza di informazioni diverse. I principali fattori da considerare nell'individuazione della numerosità del campione sono:

- la varianza nella popolazione del carattere in studio
- l'ampiezza desiderata dell'intervallo di confidenza.

LA DETERMINAZIONE DELLA DIMENSIONE DEL CAMPIONE

Per quanto riguarda la varianza, misura del grado di variazioni o oscillazioni presenti relativamente al parametro che vogliamo stimare nella popolazione, questa può essere derivata, almeno approssimativamente dall'esperienza o dai risultati di altre analoghe indagini effettuate in precedenza.

- Una popolazione in cui il parametro da misurare presenta ampie oscillazione ha una varianza elevata; una popolazione in cui le oscillazioni sono scarse ha una varianza bassa.
- È intuitivo che la precisione di un campione è maggiore quando la popolazione da cui è estratto è tendenzialmente omogenea; mentre è minore quando la popolazione è eterogenea.

Per quanto riguarda il livello di confidenza, questo come abbiamo visto rappresenta una misura della bontà di una stima. Un intervallo di confidenza molto ampio suggerisce che non siamo molto sicuri del punto in cui si trova il vero valore. Viceversa, un intervallo ristretto indica che siamo abbastanza sicuri che il valore trovato è piuttosto vicino al valore vero della popolazione, in questo caso la stima sarà, quindi, più precisa.

- La scelta del livello di confidenza è spesso dettata da considerazioni pratiche (quantità di risorse e di tempo disponibili) più che dalla teoria
- È intuitivo che per raggiungere un livello di confidenza elevato si dovrà esaminare un campione più grande.

All'atto pratico, la determinazione della numerosità del campione dipende da considerazioni di tipo:

- **non-statistico,** (ad esempio, le risorse disponibili, la mano d'opera e finanziamenti; etc..)
- **statistico,** consistono nella dimensione della popolazione ed appartenenza ad una distribuzione normale (garantita dal teorema del limite centrale), nel margine di errore, nel livello di confidenza del campione, nella media, varianza, e deviazione standard del carattere esaminato e lo Z-score.

DIMENSIONE DELLA POPOLAZIONE

Questo dato si riferisce al numero di persone che costituiscono l'oggetto di studio.

<u>Quando si lavora in grande scala si possono utilizzare dati approssimati, anziché numeri precisi</u>.

- La precisione ha un impatto statistico importante viene preso in considerazione un piccolo gruppo; per esempio, se si vuole condurre un'indagine fra i membri di un'organizzazione locale o fra gli impiegati di una piccola società, la numerosità della popolazione deve essere il più precisa possibile.
- Gli studi che considerano popolazioni statistiche enormi ammettono una deviazione maggiore rispetto alla realtà; per esempio, se il gruppo è composto da tutti gli abitanti degli Stati Uniti, si potrebbe utilizzare la stima grossolana di 320 milioni di individui.

MARGINE DI ERRORE

Valore α che indica con quanta probabilità i risultati dell'indagine rifletteranno il punto di vista della popolazione complessiva.

- I margini di errore piccoli forniscono risposte più precise, ma in tal caso è necessario utilizzare un campione molto grande

LIVELLO DI CONFIDENZA DEL CAMPIONE

$(1-\alpha)$ Questo valore misura il grado di certezza in merito al fatto che il campione rappresenta correttamente la popolazione all'interno di un margine di errore definito:

- In altre parole, scegliendo un livello di confidenza pari al 95%, si può affermare che si è certo al 95% che i risultati ricadano all'interno dell'intervallo stabilito dal margine di errore
- Un elevato livello di confidenza indica una grande precisione, ma richiede anche un campione numeroso.

DEVIAZIONE STANDARD E COEFFICIENTE DI VARIAZIONE

La deviazione standard viene utilizzata per quantificare la variazione attesa fra le varie risposte. Come abbiamo visto in precedenza questa informazione è molto importante nel calcolo della numerosità campionaria:

- Dato che questo valore è difficile da determinare prima dell'indagine, la maggior parte dei ricercatori lo imposta a 0,5.

Il coefficiente di variazione è ignoto. Per superare questo problema si hanno due soluzioni possibilità:

- considerare il valore del coefficiente di variazione di un'altra indagine fatta da terzi
- effettuare un'indagine pilota di poche unità (20-30) e calcolare il coefficiente di variazione.

Z-SCORE

Si tratta di un valore costante che si imposta automaticamente sul livello di confidenza. Indica il "risultato standard" o il numero di deviazioni standard tra un qualsiasi valore selezionato e la media della popolazione. I valori **z** della variabile normale standardizzata sono tabulati nella tabella Z e messi a disposizione. La maggior parte dei ricercatori memorizza semplicemente gli Z-score necessari per i livelli di confidenza più utilizzati:

- 90% di confidenza > 1,65 Z-score
- 95% di confidenza > 1,96 Z-score
- 99% di confidenza > 2,58 Z-score.

FORMULA STANDARD

Per popolazione di piccola o media dimensione e noti tutti i valori chiave, la formula per calcolare la dimensione del campione è data da:

$$\boldsymbol{sample\ size} = \frac{\dfrac{\boldsymbol{z^2 * p(1-p)}}{\boldsymbol{e^2}}}{\boldsymbol{1 + (\dfrac{z^2 * p(1-p)}{e^2 N})}}$$

Dove:

- $\boldsymbol{N}$ è la dimensione del campione
- $\boldsymbol{e}$ è il margine di errore
- $\boldsymbol{z}$ è il punteggio z
- $\boldsymbol{p}$ è la deviazione standard.

Per una popolazione molto numerosa o sconosciuta, la formula è:

$$\boldsymbol{sample\ size} = \frac{\boldsymbol{z^2 * p(1-p)}}{\boldsymbol{e^2}}$$

FORMULA STANDARD

Per popolazioni finite distribuite normalmente e con varianza ignota la formula per il calcolo della numerosità campionaria è:

$$\boldsymbol{sample\ size = \frac{z^2 * cv^2 * N}{(N-1)e^2 + z^2 * cv^2}}$$

Dove:

- $\boldsymbol{cv}$ è il coefficiente di variazione.

INTERVALLO DI CONFIDENZA PER LA MEDIA (VARIANZA NOTA)

Dato un campione casuale estratto da una popolazione Normale con media ignota e varianza **nota**, l'intervallo di confidenza per la media della popolazione al livello di confidenza $(1-\alpha)$ è:

$$[\bar{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}; \bar{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}]$$

Dove i valori $z_{\frac{\alpha}{2}}$ della variabile normale standardizzata sono tabulati.

Ad esempio

Siano n = 10; σ = 3; $1-\alpha$ = 0,99

Dalle tavole della Normale standardizzata si ottiene:

$z_{\frac{\alpha}{2}}$ =2,576

Se $\bar{X}$ = 4,924; si ottiene

$[4{,}924 - 2{,}576 \frac{3}{\sqrt{10}};\ 4{,}924 + 2{,}576 \frac{3}{\sqrt{10}}]$ ovvero $[2{,}4802;\ 7{,}3678]$

INTERVALLO DI CONFIDENZA PER LA MEDIA (VARIANZA NOTA)

La lunghezza dell'intervallo di confidenza si ricava dalla differenza tra estremo superiore ed estremo inferiore:

$$\boldsymbol{lunghezza = 2z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}}$$

Dipende da:

- la dimensione del campione $\boldsymbol{n}$
- il livello di confidenza $\boldsymbol{\alpha}$
- la varianza della popolazione $\boldsymbol{\sigma}$.

N.B: Intervenendo sulla dimensione del campione o sul livello di confidenza si può aumentare o diminuire la lunghezza dell'intervallo. Una volta fissati questi due elementi, al variare dei campioni estratti, la lunghezza degli intervalli corrispondenti rimane costante.

INTERVALLO DI CONFIDENZA PER LA MEDIA (VARIANZA IGNOTA)

Sia X una ***v.c.*** che rappresenta un carattere osservato su una popolazione.

Supponiamo che la ***v.c.*** sia distribuita come una Normale con media e varianza ignota.

Per stimare la varianza della popolazione si utilizza lo stimatore varianza campionaria corretta:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

Pertanto, la ***v.c***.

$$T = \frac{(\bar{X} - \mu)}{\frac{S}{\sqrt{n}}}$$

si distribuisce come una ***v.c. t-Student*** con n-1 gradi di libertà.

I TEST DI IPOTESI

Nell'ambito dell'inferenza statistica capita spesso di trovare problemi di verifica delle ipotesi.

Un'ipotesi statistica è una congettura sulla forma della distribuzione di probabilità di una variabile casuale ovvero sul valore del parametro incognito. Nel primo caso si parla di ipotesi funzionale, mentre nel secondo si parla di ipotesi parametrica.

L'ipotesi che si vuole sottoporre a verifica, denotata con H_0 , è detta ipotesi nulla o di base, mentre l'ipotesi alternativa è indicata con H_1.

L'impostazione data da J.Neyman e E.S.Pearson prevede che le ipotesi vengono formulate in base ad informazioni che si possiedono del fenomeno in esame. Esse possono essere:

- semplici, quando specificano completamente la popolazione, ovvero, $H: \theta = \theta_0$
- composte, quando non specificano completamente la popolazione, ovvero $H: \theta \neq \theta_0$

Ponendo l'ipotesi nulla $H_0: \theta = \theta_0$, il test si dice:

- unidirezionale, quando specifica un intervallo di valori, ovvero $H_1: \theta < \theta_0$, o $H_1: \theta > \theta_0$,
- bilaterale, quando specifica due intervalli di valori, ovvero $H_1: \theta \neq \theta_0$

L'obiettivo è, attraverso un campione di osservazioni, stabilire, con un certo grado di attendibilità, se poter rifiutare o meno l'ipotesi nulla a favore dell'ipotesi alternativa.

I sistemi di ipotesi più frequentemente utilizzati sono:

$$\begin{cases} H_0: & \theta = \theta_0 \\ H_1: & \theta > \theta_0 \end{cases} \qquad \begin{cases} H_0: & \theta = \theta_0 \\ H_1: & \theta \neq \theta_0 \end{cases} \qquad \begin{cases} H_0: & \theta = \theta_0 \\ H_1: & \theta < \theta_0 \end{cases}$$

Dove θ_0 è un valore fissato del parametro.

REGIONE DI ACCETTAZIONE E DI RIFIUTO

Un test statistico è una regola che permette di discriminare i campioni che portano all'accettazione dell'ipotesi nulla da quelli che portano al suo rifiuto.

- Il test si basa sul valore assunto da una statistica test. La statistica test è una statistica campionaria la cui distribuzione deve essere completamente nota sotto l'ipotesi nulla
- L'insieme dei valori della statistica test che portano all'accettazione dell'ipotesi nulla è chiamata regione di accettazione.
- L'insieme dei valori della statistica test che portano al rifiuto dell'ipotesi nulla è chiamata regione di rifiuto.

Si riportano di seguito le fasi da seguire per realizzare una verifica delle ipotesi:

i) individuazione della statistica test
ii) definizione della regola di decisione, ovvero della partizione dei valori assunti dalla statistica test in regione critica o di rifiuto di H_0 ed in regione di non rifiuto di H_0
iii) determinazione del valore empirico del test, attraverso la sostituzione dei dati campionari nella statistica test già individuata
iv) decisione del test, in cui si verifica se il valore empirico di cui al punto cade o meno nella zona di rifiuto. Nel primo caso si rifiuta l'ipotesi nulla e il test è detto significativo, mentre nel secondo non si rifiuta H_0.

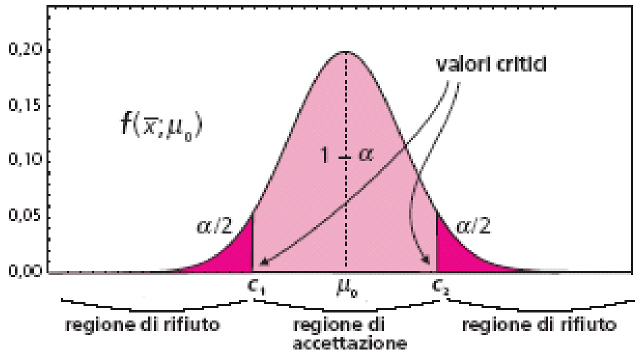
REGIONE DI ACCETTAZIONE E DI RIFIUTO

Esempio

Supponiamo che la popolazione sia Normale con media μ incognita e varianza σ^2 nota. Si vuole verificare:

$$\begin{cases} H_0: & \mu = \mu_0 \\ H_1: & \mu \neq \mu_0 \end{cases}$$

Considerando come statistica test la media campionaria $\bar{X}$ sappiamo che sotto l'ipotesi nulla questa si distribuisce come una Normale con media $\mu = \mu_0$ e varianza σ^2/n.



Dalla figura si può vedere che i valori critici definiscono la zona di accettazione e che dipendono dal livello di significatività α: maggiore è il suo valore, più ampia sarà la regione di rifiuto.

IL P-VALUE ED ERRORI DI I E II TIPO

Un altro modo per evidenziare il risultato del test è quello di calcolare il p-value, ovvero probabilità di osservare un valore della statistica test uguale o più estremo del valore ottenuto dal campione, sotto l'ipotesi nulla.

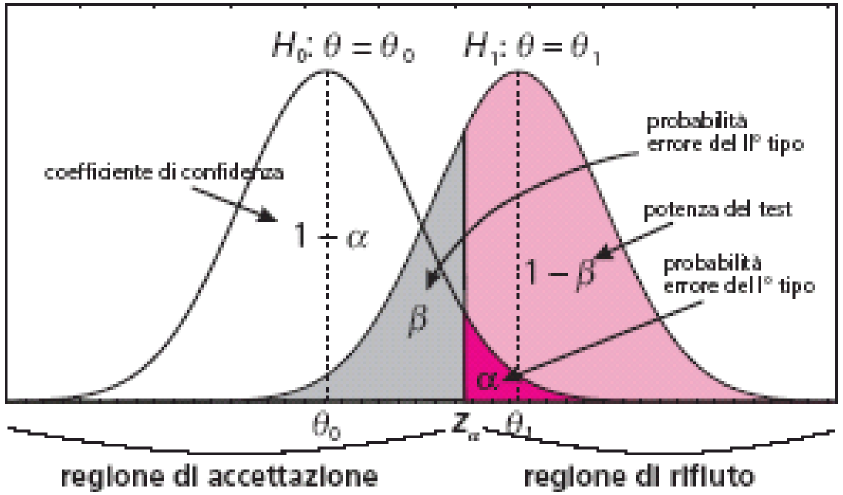
- È una quantità che misura l'evidenza fornita dai dati contro l'ipotesi nulla: minore è il valore del p-value, più è forte l'evidenza contro l'ipotesi nulla.
- Se $p\text{-value} < \alpha$ si rifiuta l'ipotesi nulla H_0.

Nel test delle ipotesi si possono commettere due tipi di errori:

- di prima specie, se rifiuto l'ipotesi H_0 quando in realtà essa è vera. La probabilità di commettere tale errore si denota con α. $(1-\alpha)$ viene detto coefficiente di confidenza del test
- di seconda specie, se non rifiuto l'ipotesi H_0 quando in realtà essa è falsa. La probabilità di commettere tale errore si denota con β. $(1-\beta)$ viene detto potenza del test.

IL P-VALUE ED ERRORI DI I E II TIPO

Sarebbe opportuno ridurre congiuntamente entrambe gli errori α e β, ma purtroppo si dimostra che ciò non è possibile; fissata la dimensione campionaria n tra i due errori esiste una relazione inversa, cioè all'aumentare dell'uno diminuisce l'altro. La riduzione di entrambi gli errori si potrebbe avere soltanto aumentando la dimensione del campione.



In definitiva, una volta stabilite le ipotesi "nulla" e "alternativa", a seconda del tipo di problema, si fissa il valore di α, si sceglie la statistica test appropriata al caso in esame e si individua la regione critica di dimensione α in modo da rendere minimo β (cioè il test più potente).

TEST PER LE MEDIE

Data una variabile casuale con distribuzione Normale e varianza nota, la statistica test e sua distribuzione sotto l'ipotesi nulla è data da:

$$Z = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

Con μ_0 che indica il valore della media ipotizza in H_0.

Esempio:

Si vuole verificare se dopo una campagna pubblicitaria il fatturato medio sia aumentato rispetto a quello dell'anno precedente pari a μ = 2500. Il fatturato è una ***v.c. Normale*** con varianza nota pari a σ^2 =1296.

1) <u>Definizione del sistema di ipotesi:</u> H_0:: μ = 2500 contro H_1:: μ > 2500
2) <u>Scelta della statistica test:</u>

$$Z = \frac{\bar{X} - 2500}{\frac{36}{\sqrt{n}}}$$

3) <u>Scelta del livello di significatività e della numerosità campionaria</u>: fissiamo un α = 0,05 (e quindi un valore critico pari a 1,645) e una numerosità n = 81.

TEST PER LE MEDIE

4) Definizione della regione di rifiuto: $R\{z > 1{,}645\}$
5) Estrazione del campione: si estrae un campione casuale di 81 clienti
6) Calcolo della statistica test: la media campionaria risulta essere $\bar{X}$= 2510 e il valore della statistica test = 2,5

$$Z = \frac{2510 - 2500}{\frac{36}{\sqrt{9}}}$$

7) Decisione: poiché il valore della statistica test cade nella regione di rifiuto $\{2{,}5 > 1{,}645\}$ l'ipotesi nulla viene rifiutata.

Se il fenomeno in esame (cioè la variabile X) si distribuisce secondo una curva normale, ma non è noto il valore di σ^2, la statistica test da utilizzare è la *t-di Student* con *(n-1)* gradi di libertà:

$$T = \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}}$$

TEST PER LE MEDIE

Ad esempio, se il problema è così formulato (test bilaterale):

$$\begin{cases} H_0: & \mu = \mu_0 \\ H_1: & \mu \neq \mu_0 \end{cases}$$

la regione critica, fissato il valore di $\boldsymbol{\alpha}$, è fornita da:

$$\boldsymbol{t} \in \boldsymbol{R}: \boldsymbol{t} < -\boldsymbol{t}_{\frac{\alpha}{2}};\ \boldsymbol{t} > \boldsymbol{t}_{\frac{\alpha}{2}}$$

Dove i valori di $\boldsymbol{t}_{\frac{\alpha}{2}}$ sono tabulati al variare dei gradi di libertà.

Anche in questo, ovviamente, si rifiuta l'ipotesi nulla $\boldsymbol{H_0}$ se il valore empirico del test, calcolato con i dati campionari, cade nella regione critica.

TEST PER LE MEDIE

Esempio

Test sulla statura media in un collettivo:

H_0:: μ = 175 contro H_1:: μ > 175

Si assume che la statura sia una variabile casuale normale con varianza ignota. Si estrae un campione di 10 giovani e si trova:

$\bar{X}$= 181,5 e S= 95,067

$$t = \frac{181,5 - 175}{\sqrt{\frac{95,5067}{10}}}$$

$$t = 2{,}103$$

Ponendo α = 0,05 si ottiene dalla t-Student con 9 gradi di libertà:

$$t_{0,05} = 1{,}8331$$

e quindi si rifiuterà l'ipotesi nulla poiché $t \geq$ **1,8331**