

INFERENZA

Obiettivo: Utilizzare gli eventi osservati sotto forma di risultati campionari per giungere, nel modo migliore, alla conoscenza della popolazione che verosimilmente li ha generati.

L'informazione disponibile non può essere trattata come esaustiva e certa, ma presenta aspetti di parzialità e di casualità.

L'inferenza statistica affronta, quindi, problemi di decisione in condizioni di incertezza.

La risposta che si può ottenere sarà sempre inevitabilmente di tipo probabilistico e quindi non assolutamente certa.

Statistica descrittiva = problema diretto = deduzione

Si passa dal definire una popolazione, alla raccolta di informazioni su di essa, all'esame della distribuzione osservata, alla sintesi degli aspetti più rilevanti in termini di indicatori

Inferenza statistica = problema inverso = induzione

Dall'osservazione campionaria si giunge ad affermazioni che riguardano la popolazione e, in particolare, alla stima dei parametri o alla verifica di ipotesi su di essa

PROBLEMI INFERENZIALI

1)Stima dei parametri: stima dei valori incogniti di parametri (media,varianza, ecc.) caratterizzanti la popolazione

stima puntuale calcolo di un unico valore rappresentativo della "miglior stima" che è possibile ottenere per un parametro incognito della popolazione

stima intervallare definizione dei limiti di un intervallo entro il quale si ha fiducia che il parametro incognito della popolazione sia compreso

2)Verifica delle ipotesi: verifica della veridicità di assunzioni fatte sui parametri e/o sulla forma distribuzionale della popolazione

-test parametrici: prevedono assunzioni sulla forma della distribuzione della popolazione (ipotesi mantenute) e sottopongono a verifica ipotesi statistiche sui parametri della popolazione

es. test sulla media, sulla proporzione, sulla differenza tra medie, sulla varianza, sul confronto tra varianze

- test non parametrici non prevedono alcuna ipotesi sulla forma della distribuzione della popolazione

es.test sull'indipendenza, test sull'adattamento

Le statistiche campionarie

Sono funzioni delle osservazioni campionarie, assumono cioè valori differenti al variare dell'universo campionario, è possibile quindi costruire la loro distribuzione campionaria

sono variabili casuali

- MEDIA CAMPIONARIA

- *campionamento bernulliano*

popolazione finita, infinita: $\bar{X} \square N(\mu, \frac{\sigma^2}{n})$

- *campionamento in blocco*

popolazione infinita $\bar{X} \square N(\mu, \frac{\sigma^2}{n})$

popolazione finita $\bar{X} \square N(\mu, \frac{\sigma^2}{n} \frac{N-n}{N-1})$

- PROPORZIONE CAMPIONARIA

- *campionamento bernulliano*

popolazione finita, infinita: $P \square N(\pi, \frac{\pi(1-\pi)}{n})$

- *campionamento in blocco*

popolazione infinita $P \square N(\pi, \frac{\pi(1-\pi)}{n})$

popolazione finita $P \square N(\pi, \frac{\pi(1-\pi)}{n} \frac{N-n}{N-1})$

- VARIANZA CAMPIONARIA

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

$$nS^2 / \sigma^2 \sim \chi_n^2$$

esempi: distribuzione media campionaria e proporzione campionaria

STIMA PUNTUALE

- **STIMATORE:** statistica campionaria, $\hat{\theta}$, scelta per stimare un determinato parametro incognito, θ , della popolazione

es. la media campionaria è uno stimatore della media della popolazione

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \rightarrow \text{stimatore di } \mu$$

- **STIMA:** valore assunto dallo stimatore in un determinato campione $\hat{\theta}_i$

es. il valore assunto dalla media campionaria in un determinato campione è una stima della media incognita della popolazione da cui tale campione è stato estratto

un campione casuale di 25 laureati estratto da una popolazione di 100 mostra un reddito settimanale medio di 550 euro

$$\bar{x} \rightarrow \text{stima di } \mu \quad \bar{x} = 550 \text{ euro}$$

PROBLEMA:

PER UNO STESSO PARAMETRO INCOGNITO E' POSSIBILE

SCEGLIERE DIFFERENTI STIMATORI

La selezione della statistica campionaria più opportuna per la stima di un parametro dipende dalle caratteristiche/proprietà delle diverse statistiche campionarie in relazione al parametro da stimare.

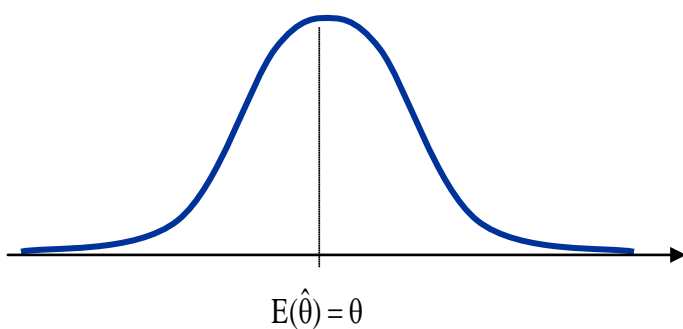
Proprietà auspicabili per uno stimatore:

- **Correttezza:** Uno stimatore è corretto (non distorto) se il valore atteso della sua distribuzione è uguale al parametro che si intende stimare.

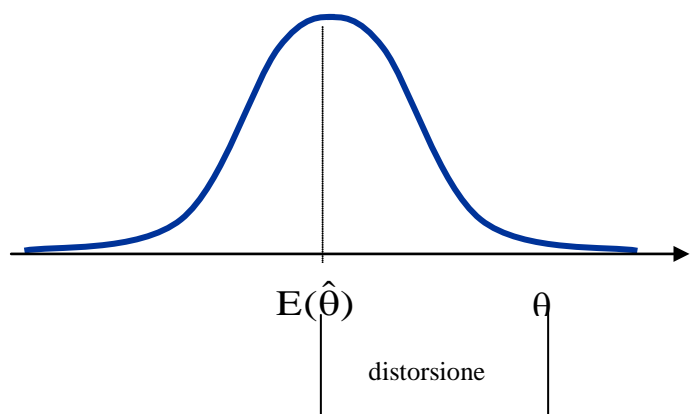
$$E(\hat{\theta}) = \theta$$

N.B. $\Delta = E(\hat{\theta}) - \theta$ distorsione

stimatore corretto



stimatore distorto



\bar{X} → stimatore corretto di μ

Me → stimatore corretto di μ (per popolazioni simmetriche)

S^2 è uno stimatore distorto di σ^2

$$E(S^2) \neq \sigma^2 = \sigma^2 \left(\frac{n-1}{n} \right)$$

$$\hat{S}^2 = S^2 \frac{n}{n-1} = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$$

$$E(\hat{S}^2) = \sigma^2$$

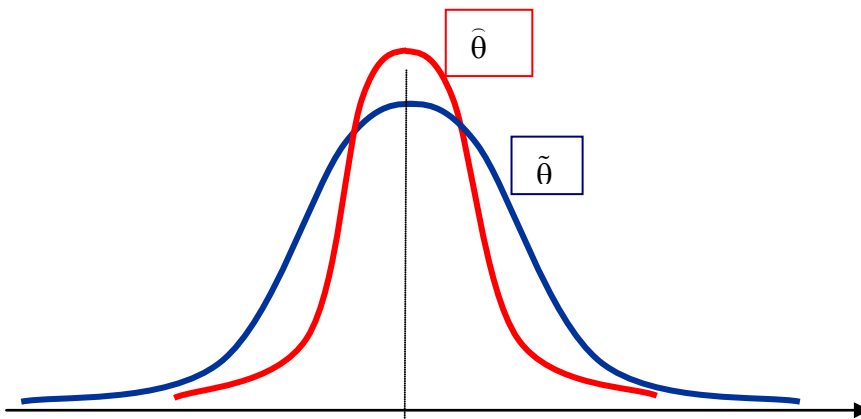
\hat{S}^2 è uno stimatore corretto di σ^2

- **Efficienza relativa:** Dati due stimatori non distorti, lo stimatore con varianza più piccola è più efficiente rispetto a quello con varianza maggiore

$$E(\hat{\theta}) = \theta$$

$$E(\tilde{\theta}) = \theta$$

$$\frac{\text{Var}(\hat{\theta})}{\text{Var}(\tilde{\theta})} < 1 \quad \hat{\theta} \text{ è più efficiente di } \tilde{\theta}$$



$\bar{X} \rightarrow$ stimatore efficiente di μ

relativamente alla mediana campionaria

- **Consistenza:** Uno stimatore si dice consistente se con l'aumentare della dimensione campionaria la probabilità che la differenza tra lo stimatore ed il parametro sia più piccola di ε (valore piccolo a piacere) è pari a 1

$$\lim_{n \rightarrow \infty} P\left[\left| \hat{\theta}_n - \theta \right| < \varepsilon \right] = 1$$

La consistenza è una proprietà importante per gli stimatori perché assicura la coerenza tra aumento della numerosità del campione e accrescimento delle informazioni in esso contenute.

Al limite uno stimatore corretto e consistente fornirà una stima esatta del parametro incognito.

n.b.

La media campionaria \bar{X} è uno stimatore corretto, efficiente e consistente della media della popolazione

STIMA PER INTERVALLI

Lo stimatore è un intervallo di valori definito sulla distribuzione della statistica campionaria, $\hat{\theta}$, scelta per stimare un determinato parametro incognito, θ , della popolazione.

L'ampiezza dell'intervallo dipende dalla prefissata probabilità di errore che si decide di accettare $1-\alpha$ =livello di confidenza

es.

$$1-\alpha=90\%$$

prima di estrarre il campione:

la probabilità di trovare campioni che contengano il parametro considerato è del 90%

dopo aver estratto il campione:

ho fiducia al 90% che il campione selezionato contenga il parametro incognito

N.B.

Si noti che uno stimatore a intervallo troppo ampio fornisce una minima informazione e che, quindi, la scelta dell'intervallo deve essere un compromesso tra la fiducia che si vuole riporre nei risultati e la precisione che si vuole ottenere dalla stima

Uno stimatore a intervalli in generale è influenzato dalla numerosità campionaria, dal livello di confidenza e dalla deviazione standard della popolazione

Stima ad intervalli per la media della popolazione

STATISTICA CAMPIONARIA: media campionaria

in particolare si utilizza la statistica campionaria standardizzata:

$$Z = \frac{\bar{X} - E(\bar{X})}{\sqrt{\text{Var}(\bar{X})}}$$

DISTRIBUZIONE DELLA STATISTICA CAMPIONARIA:

ipotesi mantenuta: popolazione normale

$$\bar{X} \square N\left(\mu, \frac{\sigma^2}{n}\right)$$

1) **nota la varianza della popolazione**

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \square Z(0,1)$$

$$P\left(-z_{\alpha/2} \leq \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq +z_{\alpha/2}\right) = 1 - \alpha$$

$$P\left(\bar{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

2) **non nota la varianza della popolazione**

$$\hat{S}^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$$

n piccolo

$$Z = \frac{\bar{X} - \mu}{\hat{S} / \sqrt{n}} \square T_{n-1}$$

$$P\left(\bar{X} - t_{\alpha/2} \frac{\hat{S}}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{\alpha/2} \frac{\hat{S}}{\sqrt{n}}\right) = 1 - \alpha$$

n grande

$$Z = \frac{\bar{X} - \mu}{\hat{S} / \sqrt{n}} \square T_{n-1} \xrightarrow{d} Z(0,1)$$

se tale ipotesi non può essere mantenuta: $\bar{X} \square N(\mu, \frac{\sigma^2}{n})$

per n grande \rightarrow teorema del limite centrale

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \square Z(0,1)$$

per n piccolo \rightarrow limite superiore di Bienaymé e Chebicheff

$$P(\bar{X} - k \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + k \frac{\sigma}{\sqrt{n}}) \geq 1 - \frac{1}{k^2}$$

$$k=2$$

$$P(\bar{X} - k \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + k \frac{\sigma}{\sqrt{n}}) \geq 0,75$$

$$k=4,47$$

$$P(\bar{X} - k \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + k \frac{\sigma}{\sqrt{n}}) \geq 0,95$$

tale intervallo è più ampio di quello ottenuto sotto l'ipotesi di normalità