Stima puntuale vs stima intervallare

La stima di un parametro può essere:

- **puntuale**, che restituisce un valore preciso per il parametro della popolazione con una precisione prefissata.

- **per intervallo**, in questo caso restituisce un campo di valori che, con un fissato margine di errore, contiene il valore vero del parametro.

Dato che $\hat{\theta}(x_1,\ldots,x_n)$ è una V.C. conoscere $\hat{\theta}$ non mi dà garanzie sul valore effettivo di θ. Fornisco, quindi, una stima $\hat{\theta}$ dando indicazioni sull'ampiezza dell'errore di stima $\hat{\theta}-\theta$.

Stima puntuale $\pm$ margine d'errore

Esempio

Stima puntuale: si stima che l'altezza media degli uomini olandesi sia 1.84cm.

Stima per intervallo: l'altezza media degli uomini olandesi è 1.84cm con un margine d'errore di 10cm.

Equivale a dire che c'è una buona probabilità che preso un campione a caso l'altezza media sia compresa tra 1.74 e 1.94.

Posso indicare con che **livello di confidenza (probabilità)** il parametro incognito della popolazione si trovi all'interno dell'intervallo di valori individuato.

Intervalli di confidenza

La stima per intervalli fornisce un campo di valori all'interno dei quali si presume cada il valore (non noto) della popolazione).

Indica con che livello di probabilità possiamo affermare che il parametro della popolazione incognito si trovi all'interno dell'intervallo di confidenza.

Sono degli intervalli di valori utilizzati per stimare un parametro non noto della popolazione, ricavati sommando e sottraendo alla media del campione uno o più margini d'errore, ai quali viene associato un determinato intervallo di confidenza, che indica la probabilità di estrarre un campione che contenga il parametro della popolazione.

Calcolo degli intervalli di confidenza

Per il calcolo degli intervalli di confidenza servono 4 info:

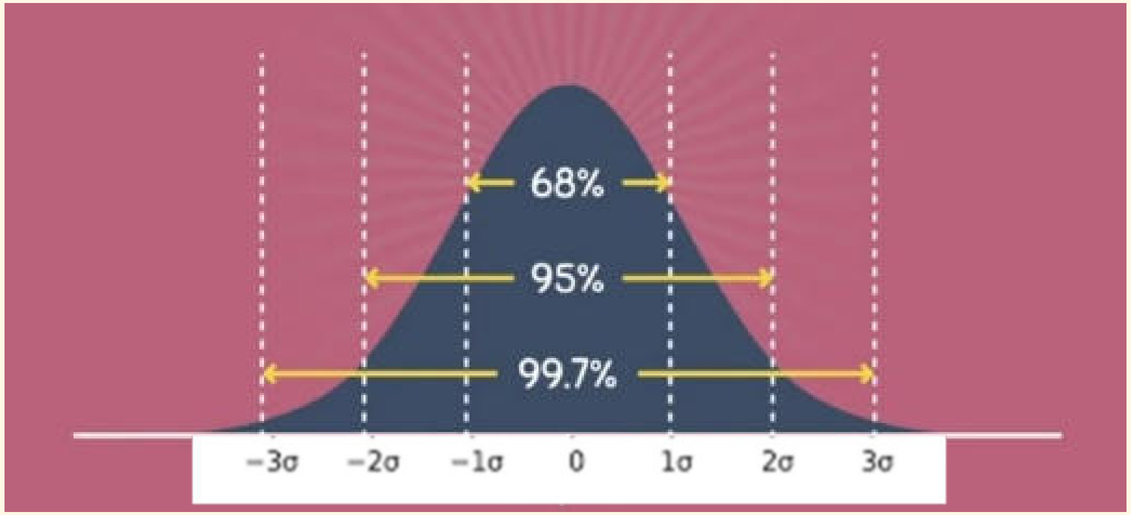
1- la media campionaria $\bar{x}$
2- l'ampiezza del campione n
3- la deviazione standard della popolazione σ, se nota
4- coefficiente di confidenza z_c (distanza dalla media in termini di deviazione standard nella normale standardizzata)

Per calcolare i valori del limite inferiore e superiore=

$$\bar{x} - z_c \frac{\sigma}{\sqrt{n}}$$

$$\bar{x} + z_c \frac{\sigma}{\sqrt{n}}$$

Intervalli di confidenza



Intervalli di confidenza

$z = \pm 1{,}96 = 95\%$

$z = \pm 2{,}58 = 99\%$

z	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0754
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1809	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2258	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2518	0,2549
0,7	0,2580	0,2612	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2996	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441
1,6	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,7	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1,8	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4761	0,4767
2,0	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4811
2,1	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857
2,2	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	0,4890
2,3	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	0,4911	0,4913	0,4916
2,4	0,4918	0,4920	0,4922	0,4925	0,4927	0,4929	0,4931	0,4932	0,4934	0,4936
2,5	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952
2,6	0,4953	0,4955	0,4956	0,4957	0,4959	0,4960	0,4961	0,4962	0,4963	0,4964
2,7	0,4965	0,4966	0,4967	0,4968	0,4969	0,4970	0,4971	0,4972	0,4973	0,4974
2,8	0,4974	0,4975	0,4976	0,4977	0,4977	0,4978	0,4979	0,4979	0,4980	0,4981
2,9	0,4981	0,4982	0,4982	0,4983	0,4984	0,4984	0,4985	0,4985	0,4986	0,4986

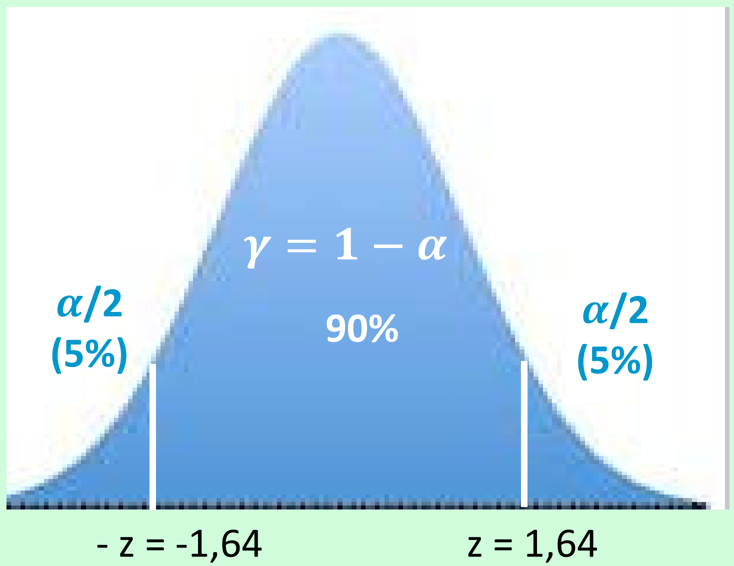
Livelli di confidenza

Livello di confidenza	99,73%	99%	95,45%	95%	90%	68,27%
z_c	3,00	2,58	2,00	1,96	1,64	1,00

Esempio, vogliamo stimare la spesa media degli italiani per l'acquisto di farmaci. Sappiamo che la deviazione standard della popolazione è di 45€. Estraiamo un campione di 2.000 famiglie e calcoliamo la media campionaria di 92€. Scegliamo un IC del 95% ossia $z_c = 1,96$

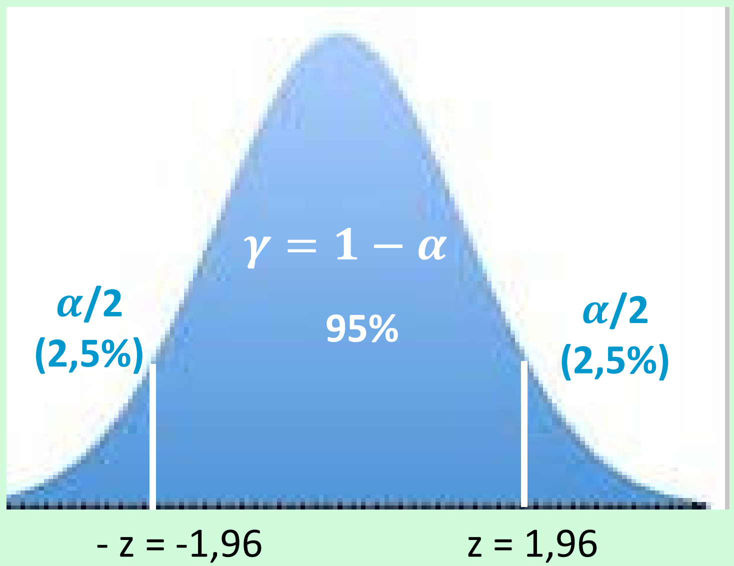
92 - 1,96*45/ √2000= 93,97 **limite inferiore**
92 + 1,96*45/ √2000= 90,03 **limite superiore**

Intervalli di confidenza

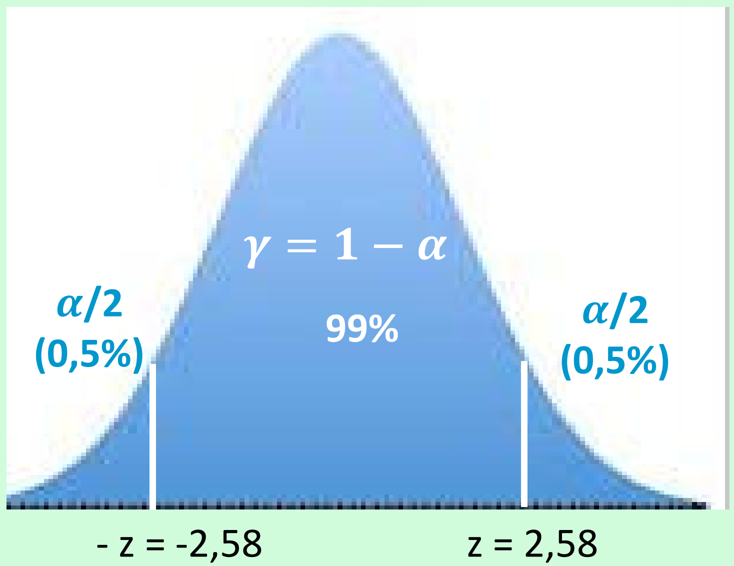


Il **livello di confidenza** è indicato con (1−α)% dove α è la probabilità che l'intervallo di confidenza non contenga il parametro della popolazione.

Intervalli di confidenza



Intervalli di confidenza



Intervalli di confidenza

Supponiamo di avere un campione casuale sufficientemente grande. La distribuzione delle medie campionarie è una gaussiana, quindi:

- per il 99% dei campioni: $\mu - 2.58\ \sigma x \leq x \leq \mu + 2.58\ \sigma x$
- per il 95% dei campioni: $\mu - 1.96\ \sigma x \leq x \leq \mu + 1.96\ \sigma x$

In termini di μ, le disuguaglianze precedenti definiscono gli intervalli di confidenza per la media μ della popolazione:
intervallo di confidenza al 99%: $x - 2.58\ \sigma x \leq \mu \leq x + 2.58\ \sigma x$
intervallo di confidenza al 95%: $x - 1.96\ \sigma x \leq \mu \leq x + 1.96\ \sigma x$

Interpretazione

Le espressioni $(1\text{-}\alpha)$ e $\frac{\alpha}{2}$ indicano il livello prefissato di probabilità dell'intervallo, ovvero il livello di confidenza che viene scelto da chi effettua l'analisi.

Indica la probabilità di estrarre un campione che contenga il parametro della popolazione.

Esempio: se estraiamo 100 campioni e creiamo intervalli di confidenza IC al 95%, 95 intervalli dovrebbero contenere la media reale della popolazione, mentre 5 saranno troppo spostati rispetto agli altri. Nel 95% dei casi il campione estratto rappresenta bene la popolazione, mentre nel 5% dei casi seleziona campioni che non so indicativi.

NON È CORRETTO affermare, invece, che c'è il 95% di probabilità che la media della popolazione sia compresa nell'IC.

ORIGINAL ARTICLE

Evaluation of mRNA-1273 SARS-CoV-2 Vaccine in Adolescents

Kashif Ali, M.D., Gary Berman, M.D., Honghong Zhou, Ph.D., Weiping Deng, Ph.D., Veronica Faughnan, B.S., Maria Coronado-Voges, M.S., Baoyu Ding, M.S., Jacqueline Dooley, B.A., Bethany Girard, Ph.D., William Hillebrand, M.S., Rolando Pajon, Ph.D., Jacqueline M. Miller, M.D., Brett Leav, M.D., and Roderick McPhee, M.D., Ph.D.

ABSTRACT

BACKGROUND

The incidence of coronavirus disease 2019 (Covid-19) among adolescents between 12 and 17 years of age was approximately 900 per 100,000 population from April 1 through June 11, 2021. The safety, immunogenicity, and efficacy of the mRNA-1273 vaccine in adolescents are unknown.

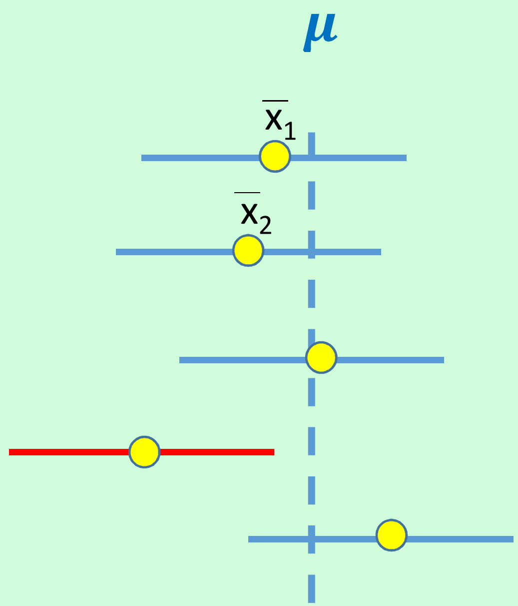
METHODS

In this ongoing phase 2–3, placebo-controlled trial, we randomly assigned healthy adolescents (12 to 17 years of age) in a 2:1 ratio to receive two injections of the mRNA-1273 vaccine (100 μg in each) or placebo, administered 28 days apart. The primary objectives were evaluation of the safety of mRNA-1273 in adolescents and the noninferiority of the immune response in adolescents as compared with that in young adults (18 to 25 years of age) in a phase 3 trial. Secondary objectives included the efficacy of mRNA-1273 in preventing Covid-19 or asymptomatic severe acute respiratory syndrome coronavirus 2 infection.

RESULTS

A total of 3732 participants were randomly assigned to receive mRNA-1273 (2489 participants) or placebo (1243 participants). In the mRNA-1273 group, the most common solicited adverse reactions after the first or second injections were injection-site pain (in 93.1% and 92.4%, respectively), headache (in 44.6% and 70.2%, respectively), and fatigue (in 47.9% and 67.8%, respectively); in the placebo group, the most common solicited adverse reactions after the first or second injections were injection-site pain (in 34.8% or 30.3%, respectively), headache (in 38.5% and 30.2%, respectively), and fatigue (in 36.6% and 28.9%, respectively). No serious adverse events related to mRNA-1273 or placebo were noted. The geometric mean titer ratio of pseudovirus neutralizing antibody titers in adolescents relative to young adults was 1.08 (95% confidence interval [CI], 0.94 to 1.24), and the absolute difference in serologic response was 0.2 percentage points (95% CI, −1.8 to 2.4), which met the noninferiority criterion. No cases of Covid-19 with an onset of 14 days after the second injection were reported in the mRNA-1273 group, and four cases occurred in the placebo group.

Rappresentazione grafica IC

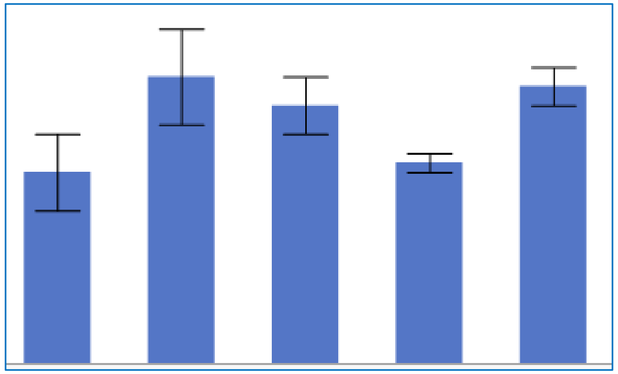


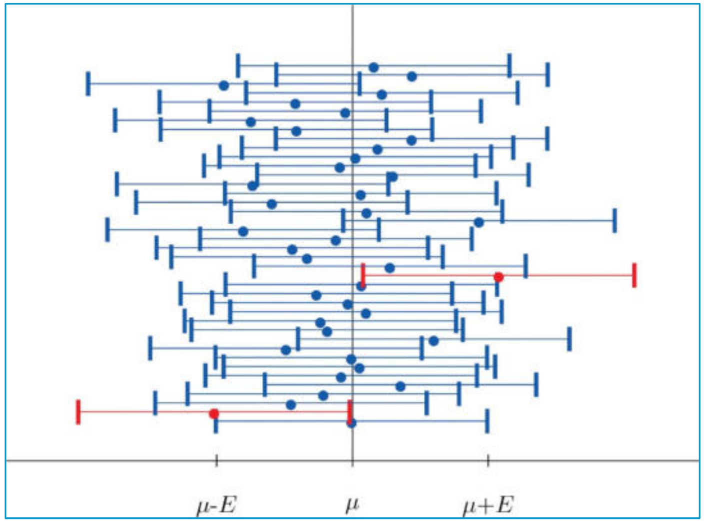
Il (1-α)% degli intervalli costruiti contengono μ

Si considerano i valori di $\bar{x}$ che si potrebbero ottenere se prendessimo campioni diversi, ciascuno con uguale numerosità.

Rappresentazione grafica IC

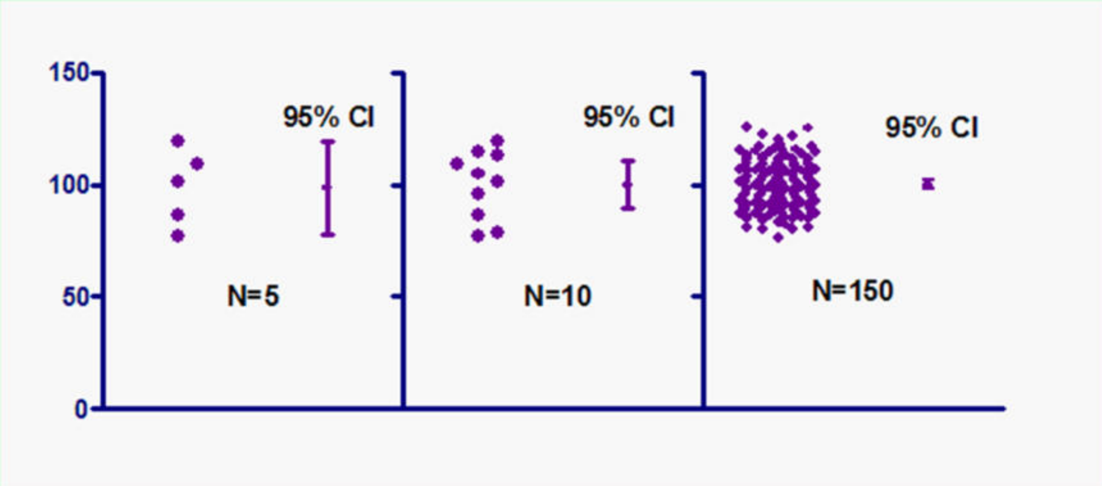
L'estensione dell'intervallo di confidenza viene raffigurato mediante dei baffi





Rappresentazione grafica IC

All’aumentare della numerosità del campione n, l’intervallo di confidenza si riduce.



Se una particolare dimensione del campione fornisce un intervallo troppo grande per essere utilizzato, si può considerare di aumentare la dimensione del campione. Si avrà così un margine d’errore più piccolo, un intervallo più stretto e una maggiore precisione.

Esercizio

Un campione casuale semplice di 50 unità ha una media campionaria di 42. La deviazione standard della popolazione è 6.

1- Fornire un IC del 95%
2- Fornire un IC del 99%

Soluzione

Un campione casuale semplice di 50 unità ha una media campionaria di 42. La deviazione standard della popolazione è 6.

1- Fornire un IC del 95%

$42 - 1{,}96 * 6/\sqrt{50} = 40{,}34$

$42 + 1{,}96 * 6/\sqrt{50} = 43{,}66$

2- Fornire un IC del 99%

$42 - 2{,}58 * 6/\sqrt{50} = 39{,}81$

$42 + 2{,}58 * 6/\sqrt{50} = 44{,}19$

Esercizio

Un IC al 95% per la media della popolazione è 152, 160.
Se σ = 15 quanto è n?

Soluzione

Un IC al 95% per la media della popolazione è 152, 160.
Se σ = 15 quanto è n?

4= 1,96*15/√ ….= 4*√ …=1,96*15= 29,4/4= 7,35^^{2} = 54

La determinazione della dimensione del campione

Si può scegliere una dimensione del campione sufficientemente grande da fornire il margine d'errore desiderato.

Viene stabilito a priori un margine d'errore rispetto al campionamento.

$$\pm z_c \frac{\sigma}{\sqrt{n}}$$ **MARGINE D'ERRORE**

Scelto il livello di confidenza 1-α possiamo determinare z_c.

La determinazione della dimensione del campione

Sia E il margine d'errore desiderato

$$E = z_c \frac{\sigma}{\sqrt{n}} \quad \Rightarrow \quad \sqrt{n} = z_c \frac{\sigma}{E}$$

La dimensione del campione per la stima intervallare della media di una popolazione è:

$$n = \frac{z_c^2 \, \sigma^2}{E^2}$$

Esercizio

Si vuole stimare il rapporto prezzo/utile medio per i titoli quotati presso la borsa di Milano nel corso del primo trimestre del 2022.

Sulla base di studi precedenti si può considerare una deviazione standard di 9.

Quanti titoli azionari dovrei includere nel campione per mantenere un margine d’errore di 2€? Utilizzare un IC di 95%.

n= 78

Proprietà degli stimatori

Quali sono le proprietà che dovrebbe soddisfare un buon stimatore?

Caratteristiche (o proprietà) desiderabili che rendono uno stimatore migliore di un altro.

Proprietà degli stimatori

Correttezza

Se uno stimatore è corretto (non distorto), le stime che si ottengono saranno centrate attorno al valore vero del parametro; in caso contrario, le stime saranno mediamente superiori o inferiori al valore vero del parametro.

Se la media di tutte le possibili stime, calcolate con lo stimatore Θ, effettuate con i possibili campioni di dimensione n, risulta uguale al corrispondente parametro della popolazione lo stimatore è corretto.
Se invece $(E\ \Theta) \neq \theta$, allora si dice che lo stimatore è distorto.

Il bias (distorsione) $B = E\ \Theta - \theta$

Questa proprietà fornisce garanzie contro il verificarsi di errori di stima sistematici, ossia di sistematiche sovrastime o sottostime del parametro.

Proprietà degli stimatori

La media campionaria è uno stimatore corretto e consistente della media della popolazione.

La varianza campionaria che descrive le varianze di tutti i possibili campioni di ampiezza n che si possono estrarre dalla popolazione, non è uno stimatore corretto della varianza della popolazione.

Il valore medio della distribuzione della varianza campionaria non corrisponde alla varianza della popolazione.

Per avere uno stimatore corretto occorre considerare un opportuno fattore di correzione.

Proprietà degli stimatori

Efficienza

In generale, dati due possibili stimatori corretti $\hat{\Theta}_1$ e $\hat{\Theta}_2$ del parametro θ, si considera più efficiente quello che ha la minore varianza campionaria.

L'efficienza si ha quando la dispersione delle stime effettuate con lo stimatore, intorno al valore del parametro ignoto della popolazione di riferimento, al variare dei possibili campioni di dimensione n, è minore rispetto a quella ottenibile con altri stimatori.

Se le distribuzioni di 2 stimatori hanno la stessa media, scelgo quello con varianza minore.

A parità di valore medio della distribuzione è sempre meglio scegliere quello con minore volatilità, perché restituisce valori di stima più vicini a quello reale della popolazione.

Proprietà degli stimatori

Consistenza

Si analizza il comportamento dello stimatore al crescere della dimensione n del campione.

Se si verifica che all'aumentare di n cresce la probabilità che il parametro stimato coincida con quello della popolazione di riferimento, si dice che lo stimatore è consistente.

Domande multiple choice

La variabilità del campione indica?

- Quanto il valore della statistica campionaria possa variare da campione a campione
- La dispersione di una serie di misure effettuate su un campione
- Quanto il valore della statistica campionaria possa variare a seconda della numerosità del campione estratto
- Non si può applicare il concetto di variabilità ai campioni

Nel campionamento casuale stratificato il campione è composto da tanti campioni casuali semplici quanti sono gli strati?

- SI
- NO

In riferimento al bias di selezione quali delle seguenti affermazioni non è corretta?

- Compromette la validità interna di uno studio clinico randomizzato
- E' un errore sistematico
- E' un errore casuale

Quali delle seguenti affermazione è vera (sono possibili più alternative)?

- Nel rispetto del requisito di simultaneità, il censimento viene condotto nell'arco di un periodo di tempo limitato
- Dati più dettagliati possano essere ottenuti attraverso il ricorso a tecniche campionarie
- L'indagine sulle forze di lavoro utilizza un campione non probabilistico a 2 stadi
- Nella quantificazione dei costi gioca un ruolo fondamentale anche la tecnica di rilevazione adottata (es tipo di intervista)

L'estrazione di un campione da un elenco telefonico è un campione probabilistico?

- SI
- NO

Con riferimento ai campioni probabilistici quale delle seguenti affermazioni è falsa?

- Le unità statistiche sono estratte a sorte
- Eventuali scarti devono essere non significativi ed imputabili al caso
- Le unità statistiche da includere nel campione sono scelte in modo ragionato
- La scelta del campione non risulta influenzata dal ricercatore stesso