STATISTICA A – K (63 ore)

Marco Riani

mriani@unipr.it

http://www.riani.it

STIMA PUNTUALE (p. 55)

- Il parametro è stimato con un unico valore
  - Esempio: stima della share di un programma TV = % di spettatori nel campione AUDITEL (ad es. 24%)
- **Vantaggio**: semplicità
- **Svantaggio**: non si hanno informazioni su quanto la specifica stima ottenuta differisce dal valore del parametro (errore di stima)

STIMA PER INTERVALLO

- Il parametro è stimato con un intervallo di valori
  - Esempio: share di un programma TV compresa (con probabilità elevata) tra il 22% ed il 26%

Vantaggio: è possibile valutare l'incertezza (in termini di probabilità) associata alla stima → intervallo di confidenza

SIMBOLOGIA

- θ =parametro incognito della popolazione
- t= stima campionaria di θ
- t è funzione degli elementi del campione➔ prima dell’estrazione il valore di t è ignoto

- VARIABILIE ALEATORIA stimatore (T)
- Il valore t osservato nel campione è la realizzazione della v.a. T

Esempi

- μ= media della popolazione
- $\overline{X}$ = stimatore (v.a. media campionaria)
- $\overline{x}$ = media osservata nel campione

- π= frequenza relativa della popolazione
- P=stimatore (v.a. freq. rel. campionaria)
- p= frequenza relativa del campione (stima)

ESEMPI

- σ^2= varianza della popolazione
- S^2= stimatore (v.a. varianza campionaria)
- s^2= stima della varianza campionaria

Principio del campionamento ripetuto

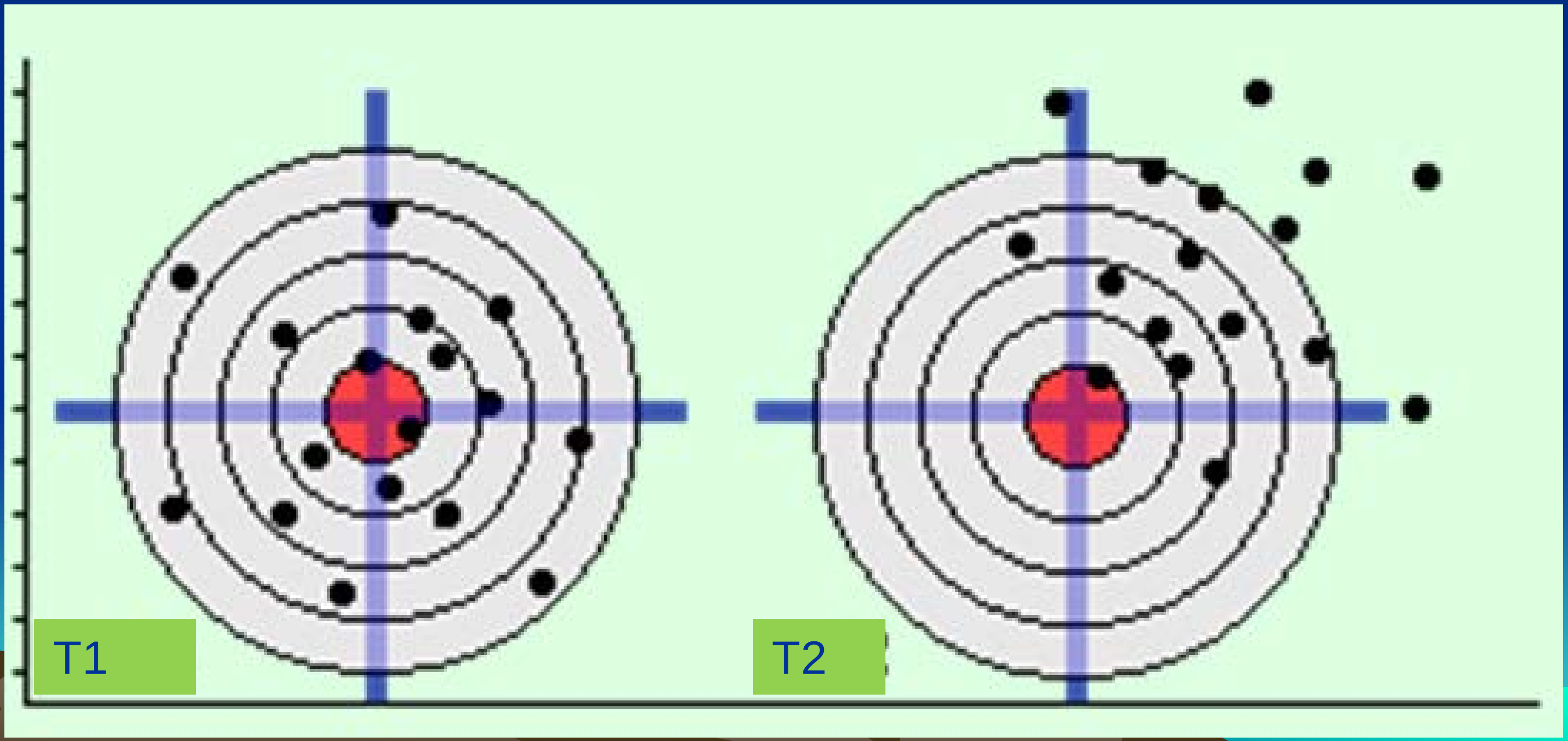
- Il campione osservato (di numerosità n) è uno dei possibili campioni che si otterrebbero ripetendo moltissime (al limite infinite) volte il campionamento (spazio dei campioni)
- La replicazione del campione è virtuale, non reale (in pratica si estrae un solo campione)

Principio del campionamento ripetuto

- Le proprietà delle procedure inferenziali sono valutate su tali replicazioni, non con riferimento allo specifico campione estratto
- Le proprietà di un metodo di stima derivano dalle proprietà della corrispondente **v.a. stimatore**

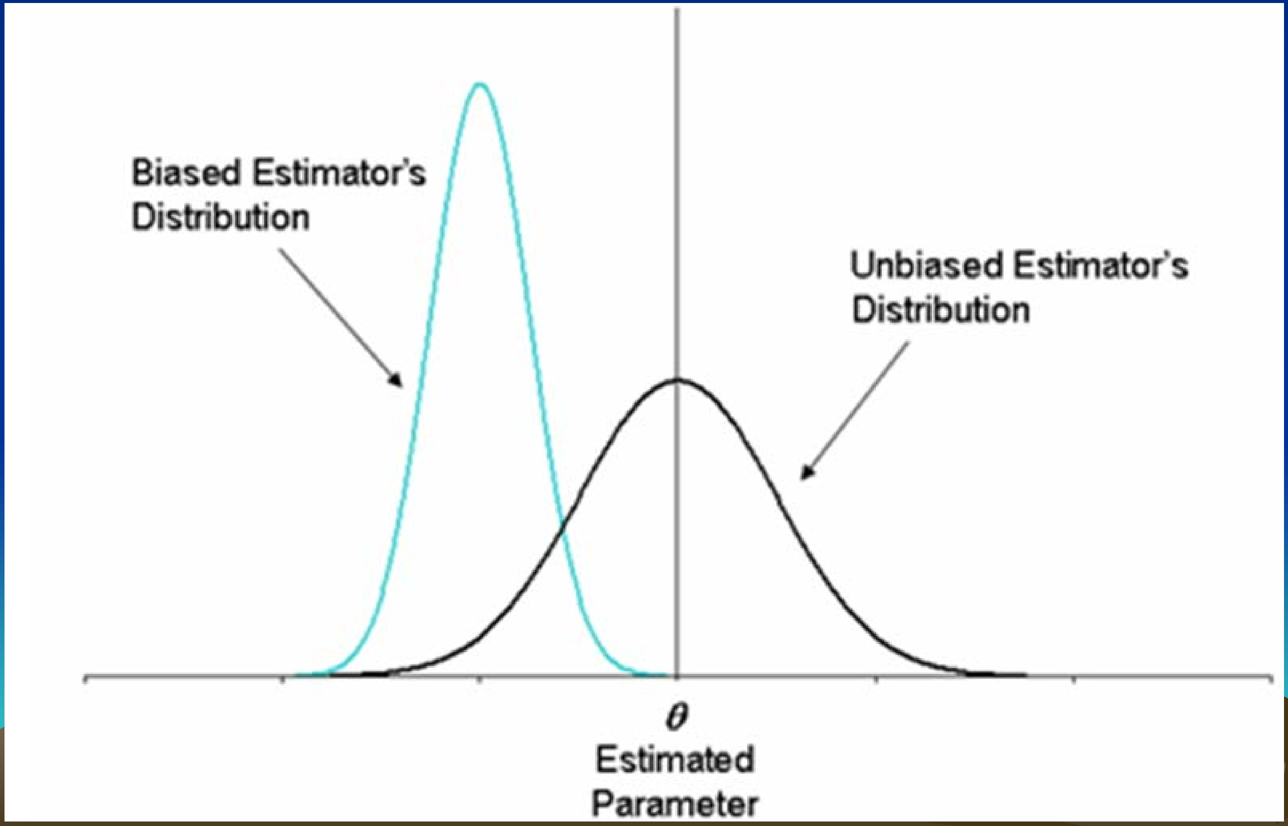
Proprietà degli stimatori: correttezza (p. 56)

- Correttezza: E(T)=θ (assenza di errore sistematico)



Proprietà degli stimatori: correttezza

- Correttezza (Unbiasness): E(T)=θ (assenza di errore sistematico)



$$E(\overline{X}) = \mu$$

- E(P)=π

Stimatore distorto

- E(T) diverso θ
- E(T)- θ= distorsione (bias)
- Esempio (stima di σ^2)

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

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

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

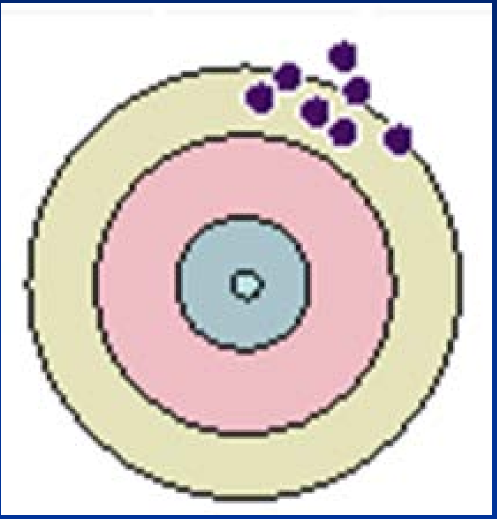
$$E(S^2_{corr}) = \sigma^2$$

Precisione (efficienza) di uno stimatore (p. 58)

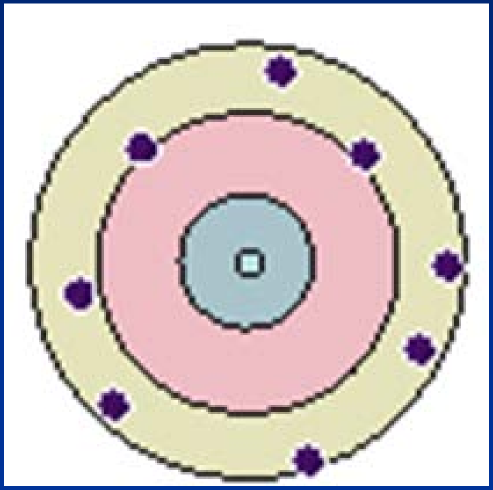
- Tanto minore è la variabilità d'uno stimatore tanto maggiore è la sua precisione
- Misure di precisione VAR(T)

Bias e precisione delle stime

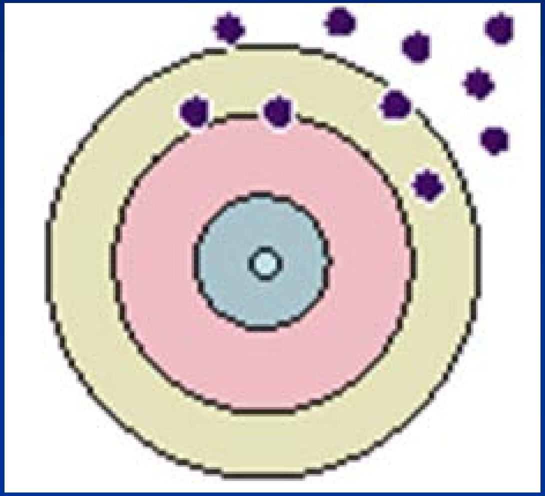
Bassa variabilità delle stime ma elevato bias

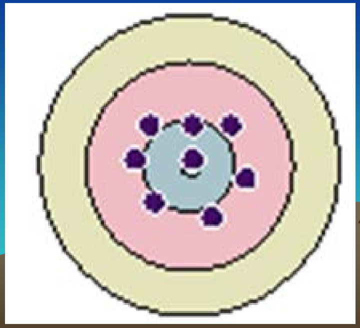


Elevata variabilità delle stime ma piccolo bias



Elevata variabilità ed elevato bias





Bassa variabilità delle stime e assenza di bias

- Uno stimatore efficiente è quello che ha la più piccola varianza (nella classe degli stimatori non distorti)

Esempio: stima di μ

- Confronto tra gli stimatori media campionaria e mediana campionaria

$$E(\overline{X}) = \mu$$

$$VAR(\overline{X}) = \frac{\sigma^2}{n}$$

$$E(Me) = \mu$$

$$VAR(Me) = \frac{\sigma^2}{n}\frac{\pi}{2} \cong \frac{\sigma^2}{n}1{,}57$$

$$VAR(Me) > VAR(\overline{X})$$

- La media campionaria è uno stimatore più preciso della mediana campionaria

Misure della precisione (variabilità campionaria)

$$\sqrt{VAR(\overline{X})} = \sigma(\overline{X}) = \sqrt{\frac{\sigma^2}{n}} \qquad \sqrt{VAR(P)} = \sigma(P) = \sqrt{\frac{\pi(1-\pi)}{n}}$$

- Dipendono da un parametro incognito della popolazione (σ^2 o π) che occorre stimare

$$\sqrt{VAR(\overline{X})} = \sigma(\overline{X}) = \sqrt{\frac{\sigma^2}{n}}$$

$$\sqrt{VAR(P)} = \sigma(P) = \sqrt{\frac{\pi(1-\pi)}{n}}$$

Errore standard di uno stimatore

$$s(\overline{X}) = \sqrt{\frac{s^2_{cor}}{n}} = \frac{s_{cor}}{\sqrt{n}}$$

$$s(P) = \sqrt{\frac{p(1-p)}{n}}$$

- Stimo σ^2 con il suo stimatore corretto s^2_{cor} e π con p

Esercizio: si consideri una generica popolazione X con media μ e varianza σ^2

- Siano $T_1=(X_1+X_2+X_3+X_4)/4$ e $T_2=(3X_1+4X_2+X_3+2X_4)/10$ due stimatori di μ per campioni di ampiezza n=4
- Si effettuino le seguenti operazioni:
  - Si verifichi che lo stimatore T_2 è non distorto
  - Si determini la varianza dei due stimatori e si stabilisca quale dei due stimatori è più efficiente

Hint: X_1 X_2 X_3 X_4 are random variables IID (independent and identically distributed) with the same distribution of X

Soluzione $U\sim(\mu\ \sigma^2)$

- $T_2=(3X_1+4X_2+X_3+2X_4)/10$

  Verifica che T_2 è non distorto

$E(T_2)=(1/10)E(3X_1+4X_2+X_3+2X_4)$

$=(1/10)\ [3\ E(X_1)+4E(X_2)+E(X_3)+2E(X_4)]$

$=(1/10)\ [\ 3\mu+4\mu+\mu+2\mu]=\mu$

Soluzione $U \sim (\mu\ \sigma^2)$

- $T_1=(X_1+X_2+X_3+X_4)/4$
  $T_2=(3X_1+4X_2+X_3+2X_4)/10$

- Calcolo della varianza dei due stimatori

$VAR(T_1)=?$ $VAR(T_2)=?$

- Calcolo della varianza dei due stimatori

$VAR(T_1)=\sigma^2/4=0{,}25\sigma^2$

$VAR(T_2)= (1/100)\ [\ 9\sigma^2 +16\sigma^2 + \sigma^2+4\sigma^2]$

$=(30/100)\ \sigma^2=0{,}3\sigma^2$

Dato che $VAR(T_1)<VAR(T_2)$

T_1 è più efficiente e quindi preferibile

Esercizio

- Dato un universo con media 6.12, varianza 46 e indice di asimmetria di Fisher pari a 3, calcolare
  - il valore atteso la varianza e l'indice di asimmetria di Fisher della v.a. secondo elemento del campione.
  - il valore atteso e la varianza dello stimatore $T=(X_1+2X_2)/3$.

Soluzione

- U~ (6.12 46)
- Indice di asimmetria di Fisher (γ)= 3

- Dato che $X_1 \ldots X_n$ are random variables IID (independent and identically distributed) with the same distribution of X
- $E(X_2)$=? $VAR(X_2)$=? $\gamma(X_2)$?

- $E(X_2)$=6.12 $VAR(X_2)$=46
- Indice di asimmetria di Fisher di X_2 = 3

Soluzione

- U~ (6.12 46)
- $T=(X_1+2X_2)/3$ E(T)?

$E(T)=E((X_1+2X_2)/3)$

$=(1/3)\,[E(X_1)+2\,E(X_2)]=$

$(1/3)(6.12+2\ 6.12)= 6.12 = \mu$

Soluzione

- U~ (6.12 46)
- $T=(X_1+2X_2)/3$.   VAR(T)?

- $VAR(T)= VAR((X_1+2X_2)/3)$

$= (1/9)\,[VAR(X_1)+4\,VAR(X_2)]$

$= (5/9)\,46$

STIMA PER INTERVALLO

Stima per intervallo

- Intervallo di confidenza di livello 1- α =intervallo che contiene il vero (ma ignoto) valore del parametro dell’universo con probabilità 1-α

- 1-α= livello di confidenza

Stima della media dell'universo (grandi campioni n>100)

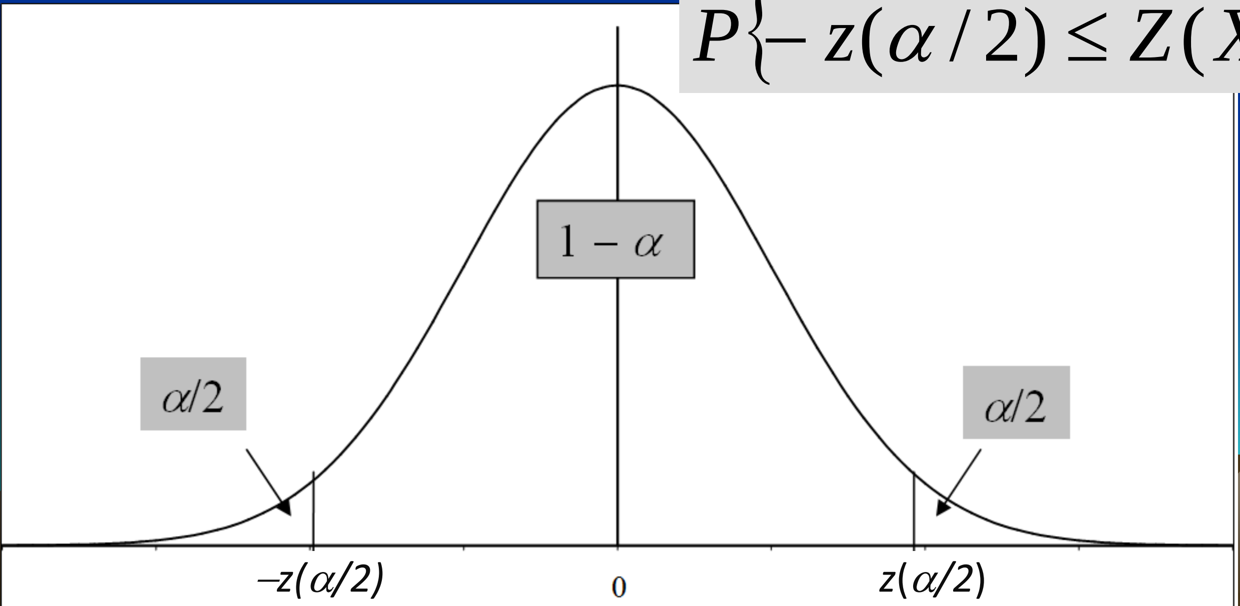
$$E(\overline{X}) = \mu \qquad VAR(\overline{X}) = \frac{\sigma^2}{n}$$

- Teorema centrale del limite

$$\overline{X} \sim \mathrm{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

$$z(\overline{X}) = \frac{\overline{X} - E(\overline{X})}{\sqrt{\mathrm{var}(\overline{X})}} = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}} \xrightarrow{Legge} \mathrm{N}(0,1)$$

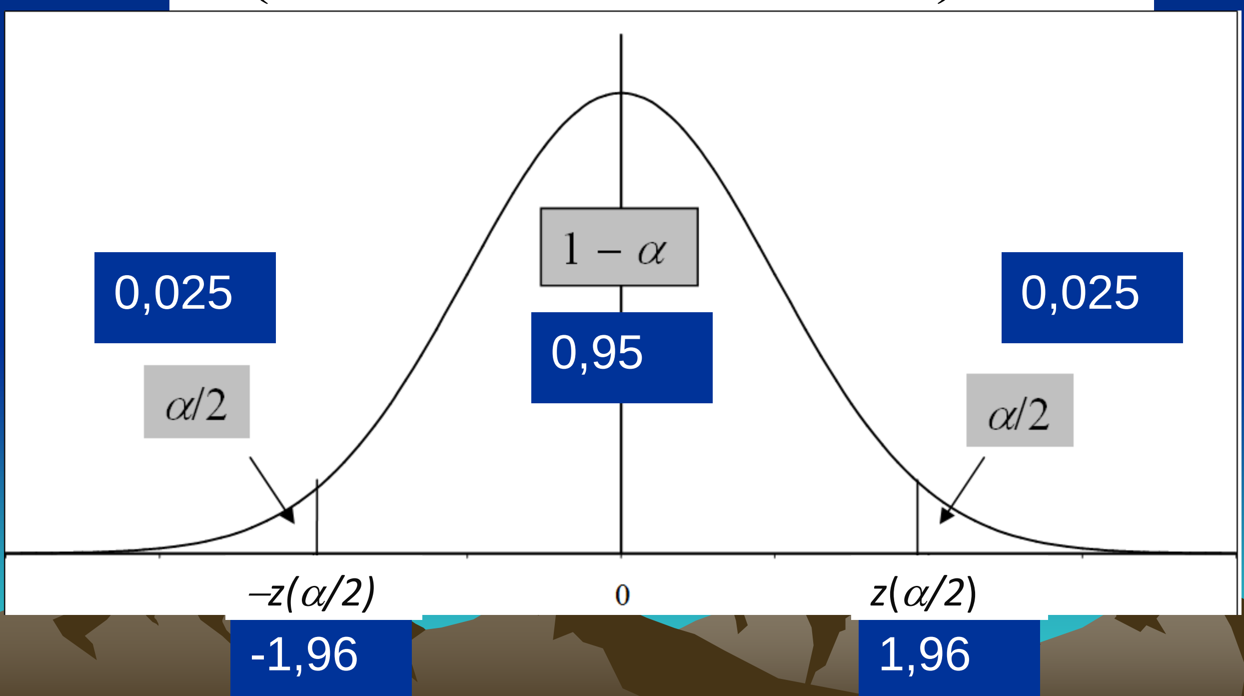
$$P\left\{- z(\alpha/2) \le Z(\overline{X}) \le +z(\alpha/2)\right\} = 1 - \alpha$$



Costruzione dell’ int. di confidenza per la media campionaria al 95%

$$P\left\{-z(\alpha/2) \leq Z(\bar{X}) \leq +z(\alpha/2)\right\} = 1-\alpha$$

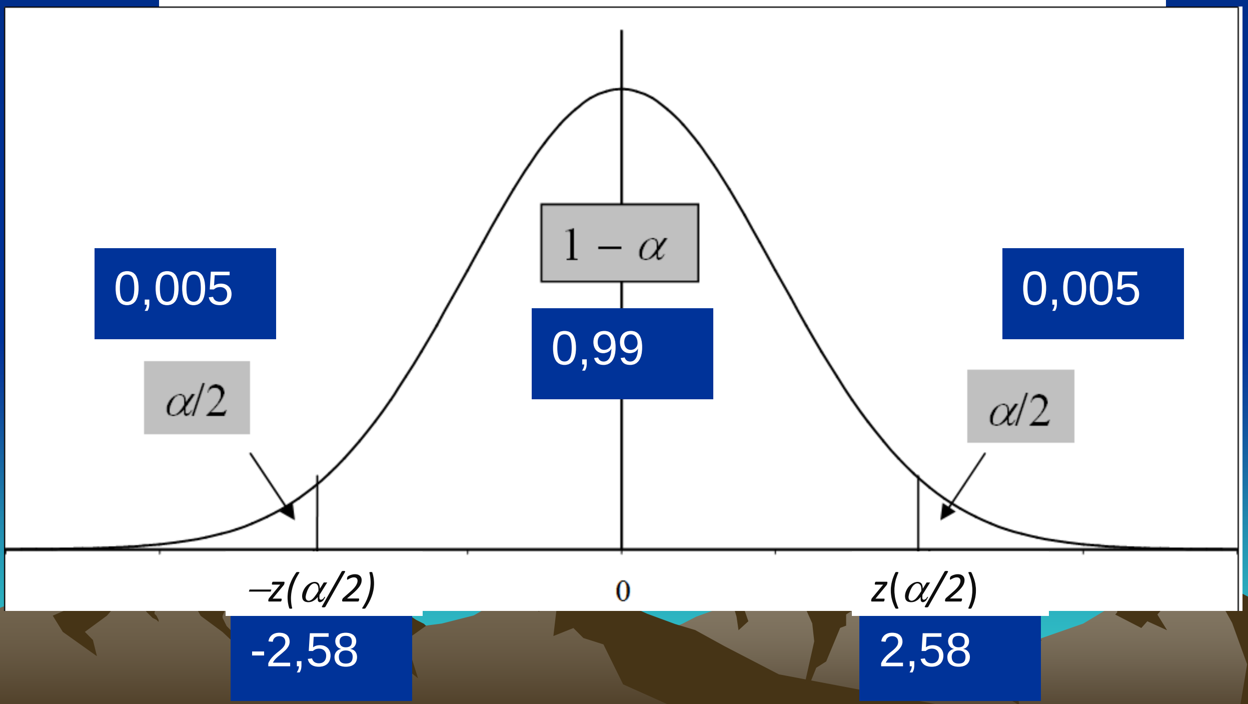
$$\Pr\left(-1.96 < \frac{\bar{X} - E(\bar{X})}{\sqrt{\operatorname{var}(\bar{X})}} < 1.96\right) = 0.95$$



Costruzione dell’ int. di confidenza per la media campionaria al 99%

$$P\left\{-z(\alpha/2) \leq Z(\overline{X}) \leq +z(\alpha/2)\right\} = 1-\alpha$$

$$\Pr\left(-2.58 < \frac{\overline{X} - E(\overline{X})}{\sqrt{\operatorname{var}(\overline{X})}} < 2.58\right) = 0.99$$



Costruzione dell' int. di confidenza per la media campionaria

$$z(\bar{X}) = \frac{\bar{X} - E(\bar{X})}{\sqrt{\operatorname{var}(\bar{X})}} = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \xrightarrow{Legge} \mathrm{N}(0,1)$$

$$\Pr\left(-z(\alpha/2) < \frac{\bar{X} - E(\bar{X})}{\sqrt{\operatorname{var}(\bar{X})}} < z(\alpha/2) \right) = 1 - \alpha$$

$$\Pr\left(-z(\alpha/2) < \frac{\bar{X} - \mu}{\sqrt{\operatorname{var}(\bar{X})}} < z(\alpha/2) \right) = 1 - \alpha$$

$$\Pr\left(\mu - z(\alpha/2)\sqrt{\operatorname{var}(\bar{X})} < \bar{X} < \mu + z(\alpha/2)\sqrt{\operatorname{var}(\bar{X})} \right) = 1 - \alpha$$

Costruzione dell’ int. di confidenza per la media campionaria al 95%

$$\Pr\left(\mu - z(\alpha/2)\sqrt{\mathrm{var}(\overline{X})} < \overline{X} < \mu + z(\alpha/2)\sqrt{\mathrm{var}(\overline{X})}\right) = 1-\alpha$$

$$\Pr\left(\mu - 1.96\sigma(\overline{X}) < \overline{X} < \mu + 1.96\sigma(\overline{X})\right) = 0.95$$

Interpretazione: intervallo (simmetrico rispetto a X medio) entro il quale è compresa, con probabilità 0,95, la media d’un campione estratto a caso da un universo di cui si conoscono la media μ e la varianza σ^2.

Esempio

- Un'azienda ha 25000 dipendenti; la retribuzione media di tutti i dipendenti è μ=1800 Euro con σ=700
- Calcolare l'intervallo in cui è compresa con prob. 0,95 la media di un campione di 200 dipendenti

$$\Pr\left(\mu - z(\alpha/2)\sqrt{\mathrm{var}(\bar{X})} < \bar{X} < \mu + z(\alpha/2)\sqrt{\mathrm{var}(\bar{X})}\right) = 1-\alpha$$

- 1-α=0,95 ➔ z(α/2)=1,96
- μ=1800

$$\sigma(\bar{X}) = \sigma/\sqrt{n} = 700/\sqrt{200} = 49{,}497$$

$$\Pr\left(\mu - z(\alpha/2)\sqrt{\mathrm{var}(\bar{X})} < \bar{X} < \mu + z(\alpha/2)\sqrt{\mathrm{var}(\bar{X})}\right) = 1-\alpha$$

- 1-α=0,95 ➔ z(α)=1,96
- μ=1800

$$\sigma(\bar{X}) = \sigma/\sqrt{n} = 700/\sqrt{200} = 49{,}497$$

$$\Pr\left(1800 - 1{,}96 \times 49{,}497 < \bar{X} < 1800 + 1{,}96 \times 49{,}497\right) = 0{,}95$$

$$\Pr\left(1703 < \bar{X} < 1897\right) = 0{,}95$$

- Intervallo in cui è compresa con prob. 0,95 la media delle retribuzioni di un campione di 200 dipendenti

Costruzione dell' int. di confidenza per μ (p. 64)

$$z(\bar{X}) = \frac{\bar{X} - E(\bar{X})}{\sqrt{\operatorname{var}(\bar{X})}} = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \xrightarrow{Legge} \mathrm{N}(0,1)$$

$$\Pr\left(-z(\alpha/2) < \frac{\bar{X} - \mu}{\sqrt{\operatorname{var}(\bar{X})}} < z(\alpha/2) \right) = 1 - \alpha$$

$$\Pr\left(\bar{X} - z(\alpha/2)\sqrt{\operatorname{var}(\bar{X})} < \mu < \bar{X} + z(\alpha/2)\sqrt{\operatorname{var}(\bar{X})} \right) = 1 - \alpha$$

Intervallo di confidenza di μ

$$\Pr\left(\overline{X} - z(\alpha/2)\sqrt{\mathrm{var}(\overline{X})} < \mu < \overline{X} + z(\alpha/2)\sqrt{\mathrm{var}(\overline{X})}\right) = 1 - \alpha$$

- Intervallo entro cui è compresa con prob. 1-α l'ignota media dell'universo μ

$$\sigma(\overline{X}) = \sigma / \sqrt{n}$$

- Osservazione: la varianza dell'universo è solitamente ignota ➔ stimata con s^2_{cor}

$$s(\overline{X}) = \frac{s_{cor}}{\sqrt{n}}$$

- Errore standard

Intervallo di confidenza di µ ad uso operativo (p. 65)

$$P\left\{\overline{X} - z(\alpha/2)\frac{s_{cor}}{\sqrt{n}} \le \mu \le \overline{X} + z(\alpha/2)\frac{s_{cor}}{\sqrt{n}}\right\} = 1-\alpha$$

- Ipotesi: n>100 (n>30)

- Esempio: stima della durata media del funzionamento delle pile d'un certo tipo
- n=160 $\overline{X}$ =248 ore; s=26 ore
- Livello di confidenza =0,99

$$P\left\{\overline{X}-z(\alpha/2)\frac{s_{cor}}{\sqrt{n}}\le\mu\le\overline{X}+z(\alpha/2)\frac{s_{cor}}{\sqrt{n}}\right\}=1-\alpha$$

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

$$S_{corr}=26\sqrt{\frac{160}{159}}=26{,}082$$

$$P\left\{248-2{,}58\frac{26{,}082}{\sqrt{160}}\le\mu\le248+2{,}58\frac{26{,}082}{\sqrt{160}}\right\}=0{,}99$$

$$P\{242{,}68\le\mu\le253{,}32\}=0{,}99$$

Osservazione

- Nell’esempio precedente avevamo potuto applicare il teorema centrale del limite poiché n era elevato (n>30)
- Cosa faccio quando n è piccolo?

Ip. Il fenomeno presenta distribuzione normale nell'universo

X~N(μ, σ²)

- Se σ² è noto

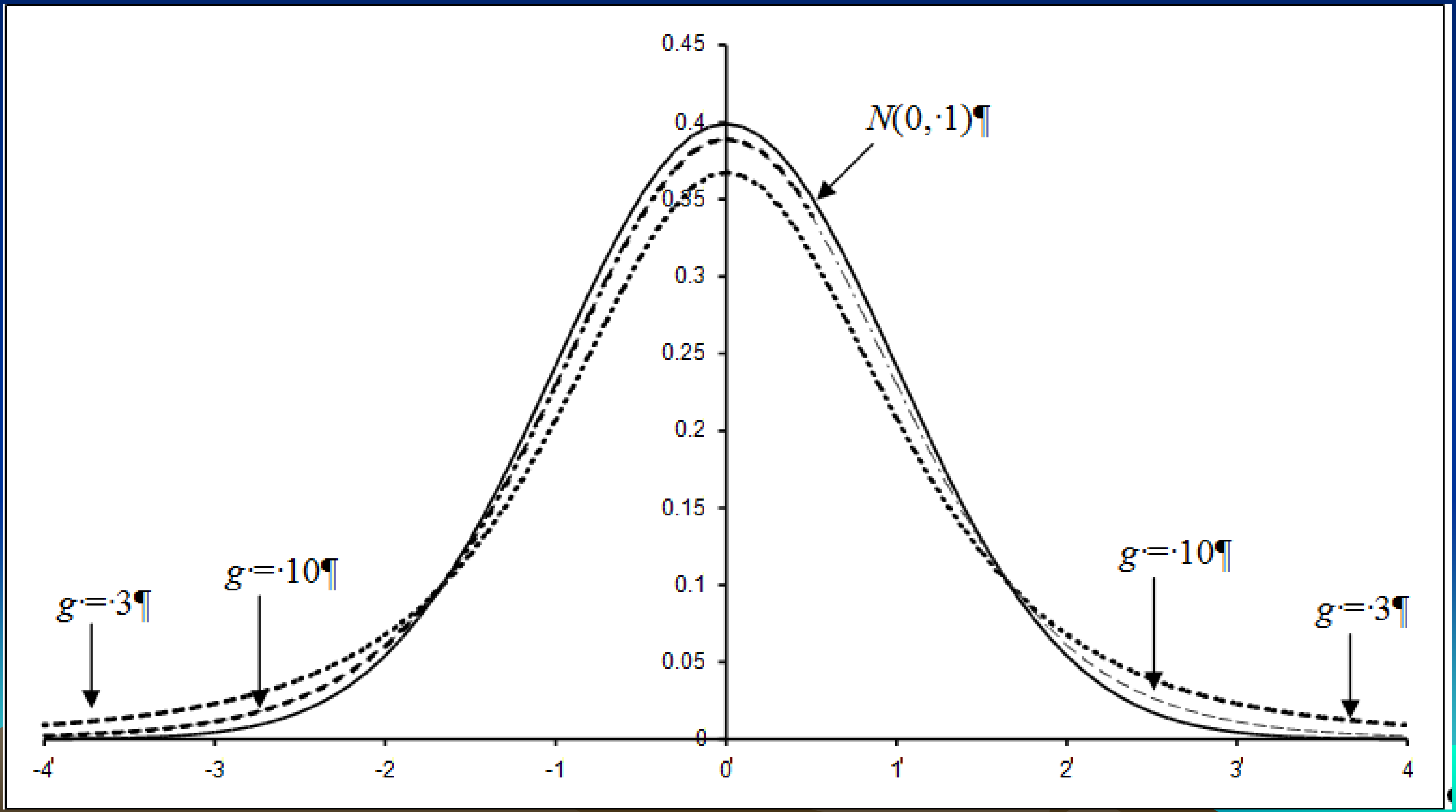
$$z(\overline{X}) = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}} \sim \mathrm{N}(0,1)$$

  - per qualunque n (anche n=1)
  - Se σ ignoto viene stimato con s_{cor} allora

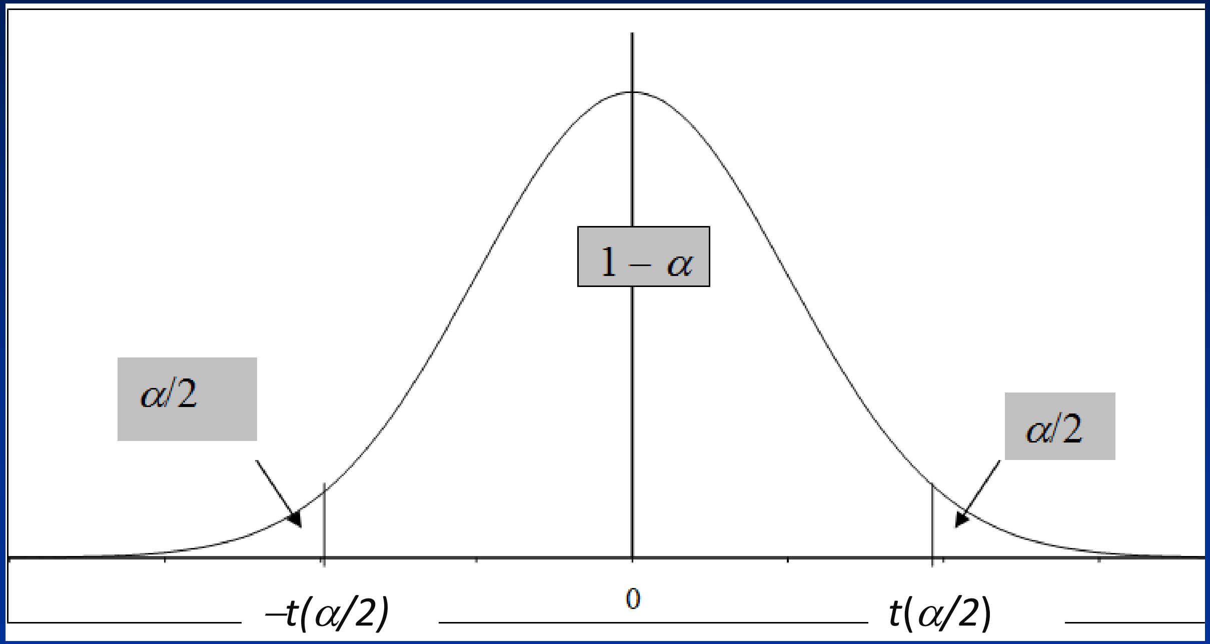
$$z(\overline{X}) = \frac{\overline{X} - \mu}{s_{cor} / \sqrt{n}} \sim t(n-1)$$

Distribuzione "t di Student" con n-1 gradi di libertà

Confronto tra una v.a. t di Student con g gradi di libertà ed una v.a. N(0,1)



t(α) valori critici ("percentili") nella v.a. *t* con g gradi di libertà

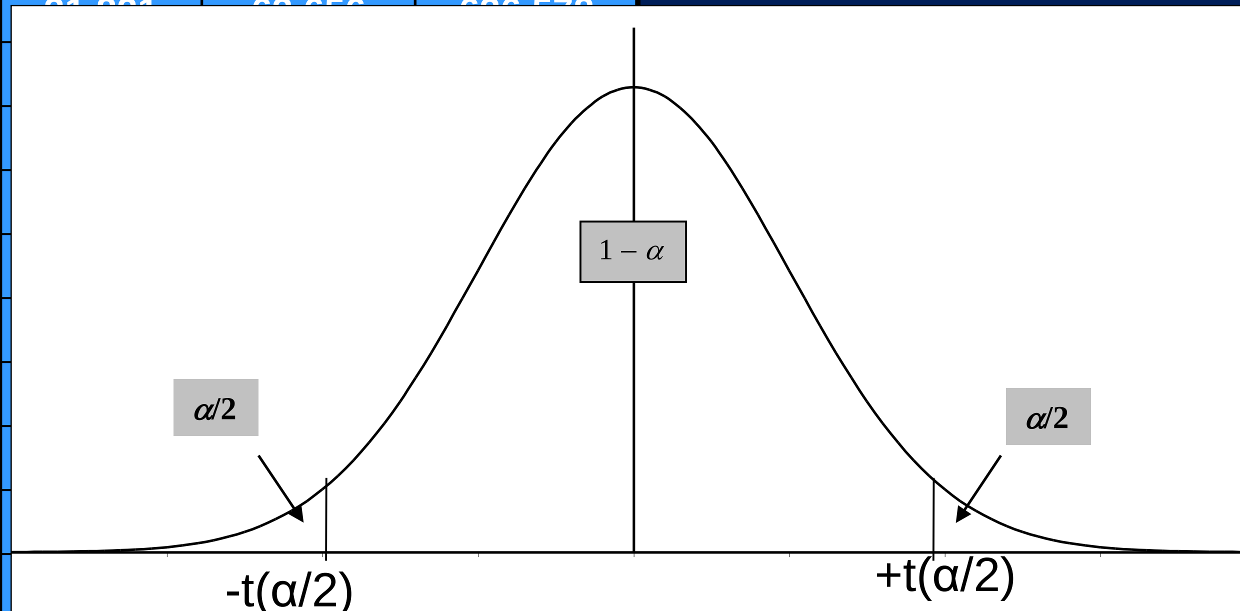


- $F[-t(\alpha/2)] = \alpha/2$ $\qquad$ $F[t(\alpha/2)] = 1-\alpha/2$

- Tavola in appendice: non riporta F(*t*) ma i "percentili" *t*(α) per α e g prefissati

VALORI CRITICI t(α) SU DUE CODE DELLA VARIABILE ALEATORIA T DI STUDENT PER g GRADI DI LIBERTA' ED AL LIVELLO DI SIGNIFICATIVITA' α

α / g	0,1	0,05	0,02	0,01	0,001
1	6,314	12,706	[illegible]	[illegible]	[illegible]
2	2,920	4,303	[illegible]	[illegible]	[illegible]
3	2,353	3,182	[illegible]	[illegible]	[illegible]
4	2,132	2,776	[illegible]	[illegible]	[illegible]
5	2,015	2,571	[illegible]	[illegible]	[illegible]
6	1,943	2,447	[illegible]	[illegible]	[illegible]
7	1,895	2,365	[illegible]	[illegible]	[illegible]
8	1,860	2,306	[illegible]	[illegible]	[illegible]
9	1,833	2,262	[illegible]	[illegible]	[illegible]
10	1,812	2,228	[illegible]	[illegible]	[illegible]
11	1,796	2,201	2,718	3,106	4,437
12	1,782	2,179	2,681	3,055	4,318
13	1,771	2,160	[illegible]	[illegible]	[illegible]
14	1,761	2,145	[illegible]	[illegible]	[illegible]
…..	…..	…….	…..	………..	……….
[illegible]	[illegible]	[illegible]	[illegible]	2,704	3,551
[illegible]	[illegible]	[illegible]	[illegible]	2,660	[illegible]
∞	1,645	1,960	2,326	2,587	3,291



$$F\left[-t(\alpha/2)\right]=\frac{\alpha}{2}$$

$$F\left[t(\alpha/2)\right]=1-\frac{\alpha}{2}$$

$$Es:\ g=1 \qquad F\left[6{,}314\right]=0.95$$

$$Es:\ g=12 \qquad F\left[2{,}681\right]=0.99$$

Dato che

$$z(\overline{X}) = \frac{\overline{X} - \mu}{s_{cor} / \sqrt{n}} \sim t(n-1)$$

$$\Pr\left(-t(\alpha/2) < \frac{\overline{X} - \mu}{s_{cor} / \sqrt{n}} < t(\alpha/2) \right) = 1 - \alpha$$

- F[-t(α/2)]= α/2 F[t(α/2)]= 1-α/2

Intervallo di confidenza di livello $1 - \alpha$ per la media dell'universo μ, nel caso di piccoli campioni e nell'ipotesi che $X \sim N(\mu, \sigma^2)$ con σ ignoto:

$$P\left\{ \overline{X} - t(\alpha/2)\frac{s_{cor}}{\sqrt{n}} \leq \mu \leq \overline{X} + t(\alpha/2)\frac{s_{cor}}{\sqrt{n}} \right\} = 1 - \alpha$$

- Esempio: stima della durata media del funzionamento delle pile d’un certo tipo
- n=10 $\overline{\boldsymbol{X}}$ =248 ore; s=26 ore
- Livello di confidenza =0,99

$$P\left\{\overline{X}-t(\alpha/2)\frac{s_{cor}}{\sqrt{n}}\leq\mu\leq\overline{X}+t(\alpha/2)\frac{s_{cor}}{\sqrt{n}}\right\}=1-\alpha$$

- Esempio: stima della durata media del funzionamento delle pile d'un certo tipo
- n=10 $\overline{X}$ =248 ore; s=26 ore
- Livello di confidenza =0,99

$$S_{corr} = 26\sqrt{\frac{10}{9}} = 27.4064$$

Ipotesi X~N(μ, σ^2) ragionevole

g=9 ➔F(3,250)=0,995=t(0,005)

$$P\left\{\overline{X} - t(\alpha/2)\frac{s_{cor}}{\sqrt{n}} \le \mu \le \overline{X} + t(\alpha/2)\frac{s_{cor}}{\sqrt{n}}\right\} = 1-\alpha$$

$$P\left\{248 - 3{,}25\frac{27{,}4064}{\sqrt{10}} \le \mu \le 248 + 3{,}25\frac{27{,}4064}{\sqrt{10}}\right\} = 0{,}99$$

$$P\{219{,}83 \le \mu \le 276{,}17\} = 0{,}99$$

Confronto tra i due intervalli di confidenza

- n elevato (v.a. normale standardizzata)

$$P\{242,68 \le \mu \le 253,32\} = 0,99$$

- n piccolo (v.a. T di Student)

$$P\{219,83 \le \mu \le 276,17\} = 0,99$$

Significato della probabilità associata all’intervallo di confidenza

- **Formulazione deduttiva**

$$P\left\{\mu - z(\alpha/2)\sigma(\overline{X}) \le \overline{X} \le \mu + z(\alpha/2)\sigma(\overline{X})\right\} = 1-\alpha$$

- **Principio del campionamento ripetuto $\Rightarrow$ distribuzione campionaria di** $\overline{X}$

- **Formulazione induttiva**

$$P\left\{ \overline{X} - z(\alpha/2)\frac{s_{cor}}{\sqrt{n}} \le \mu \le \overline{X} + z(\alpha/2)\frac{s_{cor}}{\sqrt{n}} \right\} = 1-\alpha$$

❖ μ **è una *costante (non una v.a.)*** $\Rightarrow$ ***come si può attribuire una probabilità ad un'affermazione che riguarda*** μ**?**

❖***Principio del campionamento ripetuto*** $\Rightarrow$ ***gli estremi dell'intervallo sono v.a. (v. esempio pp. 64-66)***

Esempio (pag. 76-78)

Illustriamo l'interpretazione della probabilità associata ad un intervallo di confidenza induttivo, mostrando i risultati di una piccola simulazione ottenuta nel caso di un universo con distribuzione *N*(10, 9).

Fissiamo il livello di confidenza 1 – α = 0,95 e supponiamo di ripetere l'operazione di campionamento da tale universo con *n* = 100.

In pratica, i parametri μ = 10 e = 3 non sono noti e devono essere stimati a partire dai dati. I risultati ottenuti con riferimento a 10 campioni, scelti tra tutti i possibili, sono riportati nel Prospetto 1.2. I corrispondenti intervalli di confidenza sono rappresentati graficamente nella Figura 1.4.

Prospetto 1.2 Risultati ottenuti in 10 campioni di numerosità $n = 100$ per la stima di μ.

Campione 1: $\bar{X} = 9{,}88$ $s_{cor} = 3{,}26$ intervallo di confidenza per μ da 9,24 a 10,52;

Campione 2: $\bar{X} = 9{,}90$ $s_{cor} = 3{,}02$ intervallo di confidenza per μ da 9,30 a 10,49;

Campione 3: $\bar{X} = 10{,}14$ $s_{cor} = 3{,}19$ intervallo di confidenza per μ da 9,52 a 10,77;

Campione 4: $\bar{X} = 9{,}99$ $s_{cor} = 2{,}98$ intervallo di confidenza per μ da 9,41 a 10,58;

Campione 5: $\bar{X} = 9{,}82$ $s_{cor} = 3{,}22$ intervallo di confidenza per μ da 9,19 a 10,45;

Campione 6: $\bar{X} = 10{,}16$ $s_{cor} = 3{,}03$ intervallo di confidenza per μ da 9,57 a 10,76;

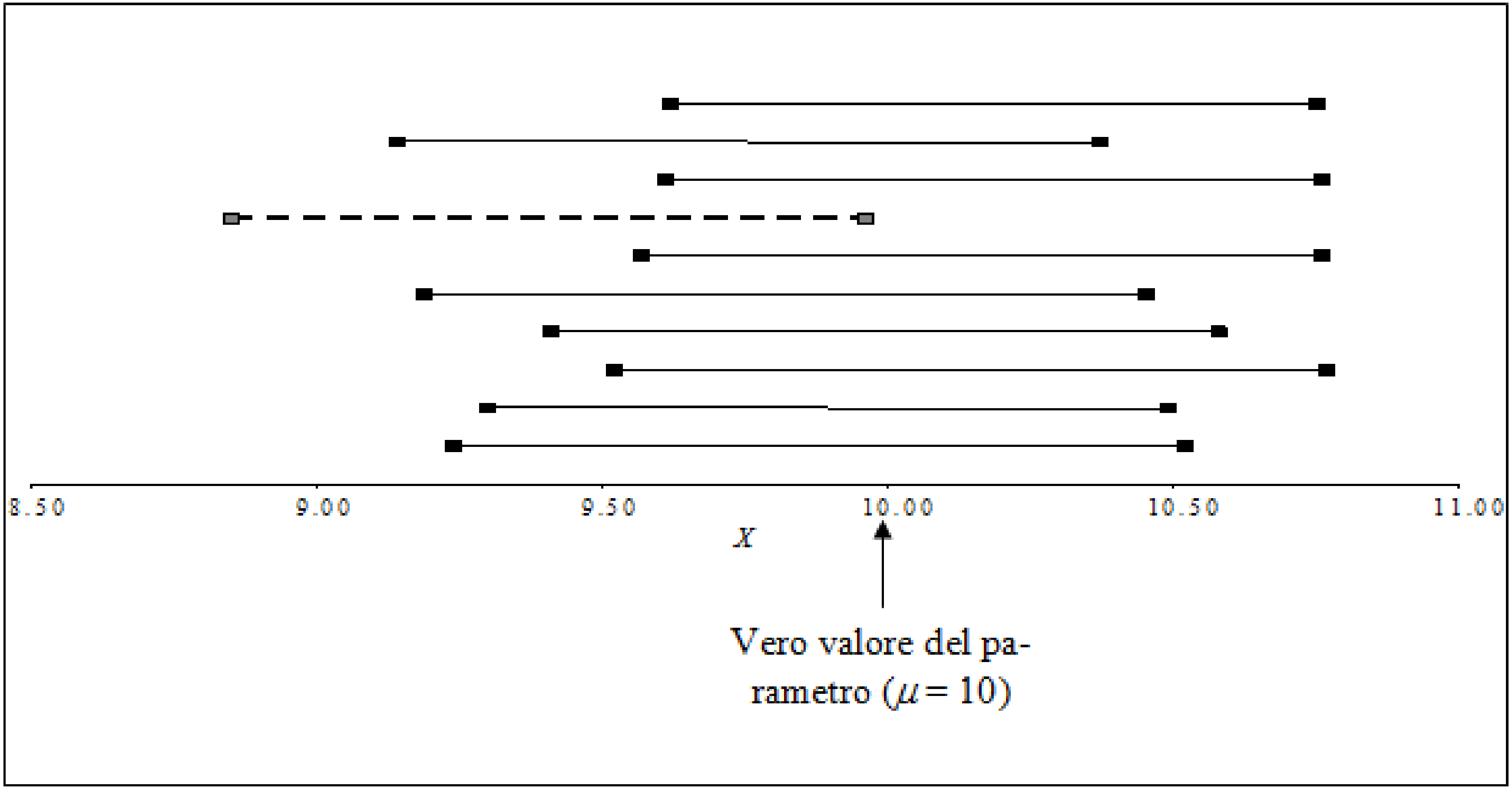
Campione 7: $\bar{X} = 9{,}41$ $s_{cor} = 2{,}83$ intervallo di confidenza per μ da 8,85 a 9,96;

Campione 8: $\bar{X} = 10{,}18$ $s_{cor} = 2{,}92$ intervallo di confidenza per μ da 9,61 a 10,76;

Campione 9: $\bar{X} = 9{,}75$ $s_{cor} = 3{,}13$ intervallo di confidenza per μ da 9,14 a 10,37;

Campione 10: $\bar{X} = 10{,}18$ $s_{cor} = 2{,}88$ intervallo di confidenza per μ da 9,62 a 10,75.

Figura 1.4 Rappresentazione dei 10 intervalli di confidenza ottenuti per la stima di μ, con probabilità $1 - \alpha = 0{,}95$. Con un tratteggio è indicato l'unico intervallo che non contiene il vero valore $\mu = 10$.



Stima della frequenza relativa (grandi campioni)

- V.a. Frequenza relativa campionaria, *P*:

  E(P) = π

$$\sigma(P) = \sqrt{\frac{\pi(1-\pi)}{n}} \quad \Rightarrow \quad s(P) = \sqrt{\frac{p(1-p)}{n}}$$

- Teorema centrale del limite

$$Z(P) = \frac{P-\pi}{s(P)} \sim N(0,1)$$

Intervallo di conf. della frequenza relativa

- **Intervallo di confidenza di livello 1 – α per la frequenza relativa dell'universo π, nel caso di grandi campioni:**

$$p \pm z(\alpha/2)s(p) = p \pm z(\alpha/2)\sqrt{\frac{p(1-p)}{n}}$$

$$P\{p - z(\alpha/2)s(p) \leq \pi \leq p + z(\alpha/2)s(p)\} = 1-\alpha$$

Esempio: stima della quota di mercato

- n = 400 consumatori;   82 acquirenti
- p = 82/400 = 0,205 $\Rightarrow$ 20,5% (stima campionaria di π)

- Calcolare l'intervallo di confidenza di π *al livello di confidenza di 0,95*

$$p \pm z(\alpha/2)s(p) = p \pm z(\alpha/2)\sqrt{\frac{p(1-p)}{n}}$$

$$P\left\{p - z(\alpha/2)s(p) \leq \pi \leq p + z(\alpha/2)s(p)\right\} = 1-\alpha$$

Esempio: stima della quota di mercato

- n = 400 consumatori;   82 acquirenti
- p = 82/400 = 0,205 $\Rightarrow$ 20,5% (stima campionaria di π)
- errore standard della v.a. P:

$$s(p) = \sqrt{\frac{0{,}205 \cdot 0{,}795}{400}} = 0{,}020$$

- $1-\alpha=0{,}95 \quad \Rightarrow \quad z(0{,}025) = 1{,}96$

$$p \pm z(\alpha/2)s(p) = p \pm z(\alpha/2)\sqrt{\frac{p(1-p)}{n}}$$

$$0{,}205 \pm 1{,}96 \cdot 0{,}020$$

$$P\{0{,}166 \leq \pi \leq 0{,}244\} = 0{,}95$$

Esempio: stima della quota di mercato

- *n* = 400 consumatori; 82 acquirenti
- *p* = 82/400 = 0,205 $\Rightarrow$ 20,5% (stima campionaria di π)

- Calcolare l'intervallo di confidenza di π *al livello di confidenza di 0,99*

$$p \pm z(\alpha/2)s(p) = p \pm z(\alpha/2)\sqrt{\frac{p(1-p)}{n}}$$

$$s(P) = \sqrt{\frac{p(1-p)}{n}}$$

- $1 - \alpha = 0{,}99 \Rightarrow \quad z(0{,}01) = 2{,}58$

$$0{,}205 \pm 2{,}58 \cdot 0{,}020$$

$$P\left\{0{,}153 \leq \pi \leq 0{,}257\right\} = 0{,}99$$

- **Intervalli ampi (stima poco precisa) $\Rightarrow$ aumentare *n***

Cosa succede se *n* è piccolo?

Piccoli campioni

- **Relazione inversa tra $\sigma(P)$ e $\sqrt{n}$:**

$$\sigma(P) = \sqrt{\frac{\pi(1-\pi)}{n}}$$

- **Il teorema centrale del limite non è applicabile $\Rightarrow$ occorre fare riferimento alla distribuzione esatta della v.a. P:**

distribuzione binomiale