Stima puntuale

- ***Stimare***: attribuire un valore plausibile a un parametro incognito.
- ***Stimatore del parametro*** θ: ogni statistica

$$T = t(X_1, X_2, \ldots, X_n)$$

utilizzata per stimare θ. Come ogni altra statistica, lo stimatore è una variabile aleatoria in quanto funzione della variabile aleatoria multipla (X_1, X_2 ,…, X_n) che rappresenta il campione.

- ***Stima***: la singola determinazione dello stimatore,

$$t = t(x_1, x_2, \ldots, x_n)$$

cioè il valore che lo stimatore assume in corrispondenza di un campione osservato (x_1, x_2 ,…, x_n).

- Per stimare un parametro θ possiamo utilizzare *diversi stimatori* (es. per stimare la media della popolazione potremmo usare la mediana, anziché la media, campionaria). E' allora importante trovare dei *criteri per scegliere* lo stimatore più appropriato.

Criteri di valutazione di uno stimatore

- Uno dei criteri principali è l'***Errore quadratico medio*** (MSE) che per lo stimatore

$$T = t(X_1, X_2, \ldots, X_n)$$

di θ è definito come

$$MSE(T) = E\left[(T - \theta)^2\right]$$

- *Esempio*: si assuma che $X \sim$ Bin(*1, p*) e che i campioni hanno dimensione n = 5. Se p = 0,3 si ha

$\bar{x}$	$f(\bar{x})$	$(\bar{x}-0{,}3)^2$	$(\bar{x}-0{,}3)^2 f(\bar{x})$
0,0	0,168	0,09	0,0151
0,2	0,360	0,01	0,0036
0,4	0,309	0,01	0,0031
0,6	0,132	0,09	0,0119
0,8	0,028	0,25	0,0071
1,0	0,002	0,49	0,0012
	1,000		0,0420

$$MSE(\bar{X}) = 0{,}042$$

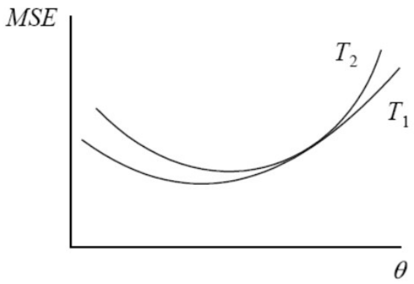
- MSE è una funzione di θ che misura l'*errore che si commette* in media utilizzando T per stimare θ. Tra due stimatori è preferibile quello che ha *uniformemente* (per ogni θ) l'MSE più piccolo.

$$T = t(X_1, X_2, \ldots, X_n)$$

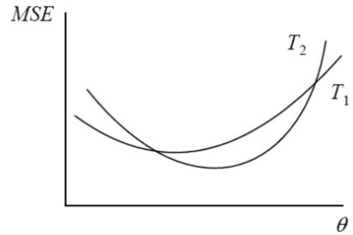
- Dati due stimatori T_1 e T_2 di θ, T_1 è ***più efficiente*** di T_2 se

$$MSE_\theta(T_1) \le MSE_\theta(T_2), \quad \forall \theta$$

e la disuguaglianza vale in senso stretto per almeno un valore di θ.



- In generale non esiste uno stimatore con MSE *uniformemente minimo* quindi usualmente ci si restringe alla classe degli stimatori non distorti.



- Lo stimatore T = t(X_1, X_2 ,…, X_n) di θ si dice ***non distorto*** se

$$E_\theta(T) = \theta, \quad \forall \theta$$

- Si chiama ***distorsione*** (o *bias*) di uno stimatore T di θ:

$$B_\theta(T) = E_\theta(T) - \theta$$

- La distorsione può essere positiva (in media T sovrastima θ) o negativa (in media T sottostima θ).

- L'errore quadratico medio può essere decomposto come segue:

$$MSE(T) = Var(T) + [B(T)]^2$$

- Se T è uno stimatore non distorto, si ha

$$MSE(T) = Var(T)$$

- Dati due stimatori non distorti T_1 e T_2 di θ, T_1 è ***più efficiente*** di T_2 se

$$Var_\theta(T_1) \le Var_\theta(T_2), \quad \forall \theta$$

e la disuguaglianza vale in senso stretto per almeno un valore di θ.

Esempio

- A prescindere dalla popolazione di riferimento, la *media campionaria* è uno stimatore non distorto della media della popolazione:

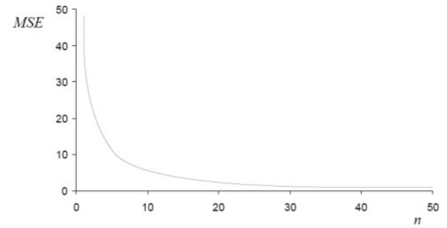
$$E(\overline{X}) = \mu$$

- Di conseguenza l'errore quadratico medio coincide con la varianza:

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

e quindi diminuisce (la sua precisione aumenta) al crescere della numerosità del campione.

- Ad esempio, se la varianza della popolazione è $\sigma^2 = 48{,}5$ si ha



Esempio

- Si consideri lo *stimatore della varianza*

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

- Questo stimatore è proporzionale alla varianza campionaria

$$\hat{\sigma}^2 = \frac{n-1}{n}S^2$$

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

e la sua distorsione è pari a

$$B(\hat{\sigma}^2) = E(\hat{\sigma}^2) - \sigma^2 = \frac{n-1}{n}\sigma^2 - \sigma^2 = -\frac{\sigma^2}{n} < 0$$

Lo stimatore *sottostima* la varianza della popolazione. La distorsione tende a 0 al crescere della dimensione campionaria.

- L'errore quadratico medio di questo stimatore non coincide con la sua varianza.

Proprietà asintotiche di uno stimatore

- Le *proprietà asintotiche* riguardano il comportamento di uno stimatore *T* quando la dimensione del campione tende a infinito.
- In questo caso si usa la notazione T_n per indicare che lo stimatore è applicato a un campione di dimensione *n*.
- Uno stimatore T_n di θ è ***consistente*** se per ogni $\varepsilon > 0$

$$\lim_{n\to\infty} P(|(T_n - \theta)| \le \varepsilon) = 1, \quad \forall \theta$$

- Uno stimatore T_n di θ è ***consistente in media quadratica*** se

$$\lim_{n\to\infty} E\left[(T_n - \theta)^2\right] = 0, \quad \forall \theta$$

La consistenza in media quadratica implica quella semplice.

- Uno stimatore T_n di θ è ***asintoticamente non distorto*** se

$$\lim_{n\to\infty} B(T_n) = 0, \quad \forall \theta$$

Metodi per la determinazione di uno stimatore

- Metodo della massima verosimiglianza
- Metodo dei momenti
- Metodo dei minimi quadrati

Metodo della massima verosimiglianza

- Dato un campione ($x_1, x_2, \ldots, x_n$) proveniente da $X \sim f(x;\theta)$, si definisce ***funzione di verosimiglianza*** la probabilità (densità) del campione interpretata come funzione del parametro θ,

$$L(\theta) = f(x_1;\theta) \cdot f(x_2;\theta) \cdot \cdots \cdot f(x_n;\theta) = \prod_{i=1}^{n} f(x_i;\theta)$$

- Quanto più è elevata la funzione di verosimiglianza per un certo valore di θ, tanto più questo valore è plausibile. Ciò permette di stabilire qual è il valore più plausibile del parametro.

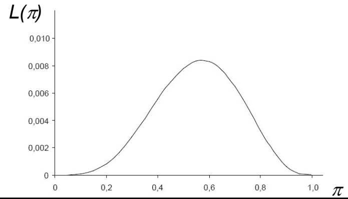
Esempio

- Si consideri un campione ($x_1, x_2, \ldots, x_n$) estratto da una *popolazione Bernoulliana* con parametro π, cioè $X \sim \mathrm{Bin}(1,\pi)$.
- La *funzione di verosimiglianza* è pari a

$$L(\pi) = \prod_{i=1}^{n} \pi^{x_i} (1-\pi)^{1-x_i} = \pi^{\sum_{i=1}^{n} x_i} (1-\pi)^{\sum_{i=1}^{n}(1-x_i)} = \pi^{s}(1-\pi)^{n-s}$$

dove $s = \sum_{i=1}^{n} x_i$ è il numero di successi e n-s di insuccessi

- Se il campione osservato è (1,0,1,0,0,1,1) si ha n = 7, s = 4 e n-s = 3 da cui $L(\pi) = \pi^4(1-\pi)^3$.



- La *stima di massima verosimiglianza* è il valore di θ che massimizza la funzione di verosimiglianza, cioè

$$\hat{\theta} = t(x_1, x_2, \ldots, x_n) \quad \textit{tale che} \quad L(\hat{\theta}) = \max_{\theta \in \Theta} L(\theta)$$

- Lo *stimatore di massima verosimiglianza*, come ogni altro stimatore è una variabile casuale, ed è definito come:

$$t(X_1, X_2, \ldots, X_n)$$

- Generalmente risulta conveniente massimizzare la *log-verosimiglianza*

$$l(\theta) = \log[L(\theta)]$$

risolvendo l'equazione

$$\frac{d\, l(\theta)}{d\theta} = 0$$

e controllando che in corrispondenza del valore trovato di θ la derivata seconda di $l(\theta)$ sia negativa.

Esempio

- Nel caso della *popolazione Bernoulliana* $X \sim \mathrm{Bin}(1,\pi)$, la *funzione di log-verosimiglianza* è pari a

$$l(\pi) = \log[L(\pi)] = s\log(\pi) + (n-s)\log(1-\pi)$$

- La *derivata prima* di $l(\pi)$ rispetto a π è

$$\frac{d\, l(\pi)}{d\pi} = \frac{s}{\pi} - \frac{n-s}{1-\pi}$$

che è pari a 0 quando $\pi = \hat{\pi}$, dove

$$\hat{\pi} = \frac{s}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

è la *stima* di *massima verosimiglianza* di π. Si noti che questa coincide con la media campionaria. Quindi lo *stimatore* di *massima verosimiglianza* di π è:

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

- Per il campione osservato (1,0,1,0,0,1,1), si ha $\hat{\pi} = 4/7 = 0{,}571$

Esempio

- Si assuma che il campione è estratto da una *popolazione Normale* con media e varianza incognite, $X \sim N(\mu, \sigma^2)$.
- Lo *stimatore di massima verosimiglianza* della media, μ, coincide con la media campionaria

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

- Lo *stimatore di massima verosimiglianza* della varianza, σ^2, è dato da

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2}{n}$$

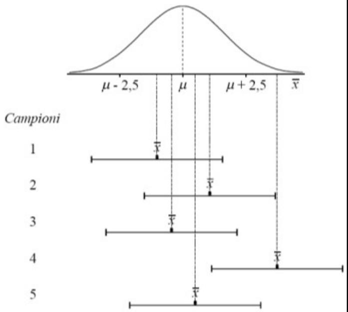
Proprietà di uno stimatore di massima verosimiglianza

- Se $\hat{\theta}$ è lo stimatore di massima verosimiglianza di θ, allora $g(\hat{\theta})$ è lo stimatore di massima verosimiglianza di $g(\theta)$. Ad esempio lo stimatore di σ nel caso di popolazione Normale è $\hat{\sigma} = \sqrt{\hat{\sigma}^2}$.
- In genere permette di ottenere *stimatori efficienti* nella classe di quelli non distorti.
- Sotto certe condizioni di regolarità, lo stimatore di massima verosimiglianza gode delle seguenti *proprietà asintotiche*:
  - è *consistente*;
  - è *asintoticamente efficiente;*
  - ha *distribuzione asintotica normale*.

Stima per intervallo

- In ogni stima è ovviamente insito un certo *margine di errore*, ad esempio la probabilità $P(\overline{X} = \mu)$ è nulla quando X è continua.
- Nell'ambito della *stima per intervallo* un parametro viene stimato tramite un *intervallo di confidenza* anziché con un singolo valore come nell'ambito della *stima puntuale*.

•Esempio: Se $X \sim N(\mu,44)$, n = 27 e si usano intervalli del tipo $\left(\overline{X} - 2{,}5\,;\, \overline{X} + 2{,}5\right)$ anziché $\overline{X}$ per stimare μ, si ottiene



Intervallo di confidenza

- Dato un campione (X_1, X_2,…, X_n) estratto da una popolazione di cui interessa stimare il parametro θ, e date due statistiche campionarie $L_1 = L_1(X_1, X_2, \ldots, X_n)$ e $L_2 = L_2(X_1, X_2, \ldots, X_n)$ con $L_1 \leq L_2$ per ogni possibile campione, l'intervallo (L_1,L_2) è un *intervallo di confidenza* per θ al (1-α)100% se

$$P(L_1 \leq \theta \leq L_2) = 1 - \alpha$$

- Il *coefficiente fiduciario* 1-α è solitamente pari al 90%, 95% o 99%.
- L'*ampiezza di un intervallo di confidenza* è data da

$$A = L_2 - L_1$$

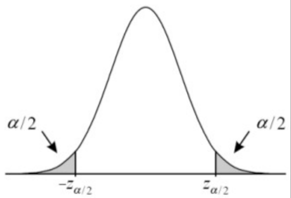
Intervalli di confidenza per la media
(popolazione normale, varianza nota)

- Se la popolazione ha *distribuzione normale*, $X \sim N(\mu,\sigma^2)$, si ha

$$\frac{\overline{X}-\mu}{\sqrt{\sigma^2/n}} \sim N(0,1)$$

- Di conseguenza

$$P\left(-z_{\alpha/2} \le \frac{\overline{X}-\mu}{\sqrt{\sigma^2/n}} \le z_{\alpha/2}\right) = 1-\alpha$$



dove $z_{\alpha/2}$ è il quantile tale che $P(Z \ge z_{\alpha/2}) = \alpha/2$ e $Z \sim N(0,1)$

- L'*intervallo di confidenza* per μ al $(1-\alpha)100\%$ è dato da

$$\left(\overline{X} - z_{\alpha/2}\sqrt{\sigma^2/n}\ ,\ \overline{X} + z_{\alpha/2}\sqrt{\sigma^2/n}\right)$$

la cui ampiezza è

$$A = 2z_{\alpha/2}\sqrt{\sigma^2/n}$$

Esempio

- Si supponga di voler costruire l'*intervallo di confidenza* al 95% per μ sulla base del campione
  170,75 186,14 173,39 185,12 173,39
  sotto l'assunzione $X \sim N(\mu,\sigma^2)$, con varianza $\sigma^2 = 50$.

- La *media campionaria* è pari a $\bar{x} = 177{,}758$

- Dato che $1-\alpha = 0{,}95$, si ha che $\alpha = 0{,}05$ e il corrispondente *quantile* della distribuzione Normale standard è pari a $z_{\alpha/2} = z_{0,025} = 1{,}96$. Si noti che $\Phi(z_{0,025}) = 0{,}975 = 1-\alpha/2$.

- L'*intervallo di confidenza* è quindi

$$\left(177{,}758 - 1{,}96\sqrt{50/5}\ ,\ 177{,}758 + 1{,}96\sqrt{50/5}\right) = (171{,}560\ ,\ 183{,}956)$$

che ha ampiezza pari a

$$A = 2 \cdot 1{,}96\sqrt{50/5} = 12{,}396$$

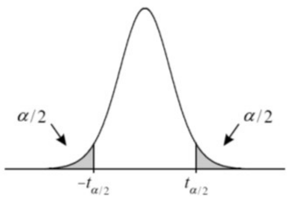
Intervalli di confidenza per la media
(popolazione normale, varianza non nota)

- Se $X \sim N(\mu,\sigma^2)$ ma σ^2 *non è nota,* si utilizza il risultato

$$\frac{\overline{X}-\mu}{\sqrt{S^2/n}} \sim t(n-1)$$

- Di conseguenza

$$P\left(-t_{\alpha/2} \le \frac{\overline{X}-\mu}{\sqrt{S^2/n}} \le t_{\alpha/2}\right) = 1-\alpha$$



dove $t_{\alpha/2}$ è il quantile tale che $P(T \ge t_{\alpha/2}) = \alpha/2$ e $T \sim t(n-1)$

- L'*intervallo di confidenza* per μ al $(1-\alpha)100\%$ è dato da

$$\left(\overline{X} - t_{\alpha/2}\sqrt{S^2/n}\ ,\ \overline{X} + t_{\alpha/2}\sqrt{S^2/n}\right)$$

la cui ampiezza è $A = 2t_{\alpha/2}\sqrt{S^2/n}$

Esempio

- Si supponga di voler costruire l'*intervallo di confidenza* al 95% per μ sulla base del campione
  170,75 186,14 173,39 185,12 173,39
  sotto l'assunzione $X \sim N(\mu,\sigma^2)$, con varianza σ^2 non nota.

- La *media campionaria* e la *varianza campionaria* sono pari a

$$\bar{x} = 177{,}758 \quad e \quad s^2 = 52{,}932$$

- Dato che $1-\alpha = 0{,}95$, si ha che $\alpha = 0{,}05$ e il corrispondente *quantile* della distribuzione $t(4)$ è pari a $t_{\alpha/2} = t_{0,025} = 2{,}776$.

- L'*intervallo di confidenza* è quindi

$$\left(177{,}758 - 2{,}776\sqrt{52{,}932/5}\ ,\ 177{,}758 + 2{,}776\sqrt{52{,}932/5}\right) = (168{,}726\ ,\ 186{,}790)$$

che ha ampiezza pari a $A = 2 \cdot 2{,}776\sqrt{52{,}932/5} = 18{,}064$
maggiore di quella calcolata per il caso della varianza nota.

Intervalli di confidenza per la media

(popolazione qualsiasi, varianza non nota, grandi campioni)

- Per una *popolazione qualsiasi*, nel caso di *grandi campioni* (n≥30) si ha che
$$\frac{\overline{X}-\mu}{\sqrt{S^2/n}}$$
ha approssimativamente distribuzione Normale standard.
- Di conseguenza
$$P\left(-z_{\alpha/2} \le \frac{\overline{X}-\mu}{\sqrt{S^2/n}} \le z_{\alpha/2}\right) = 1-\alpha$$
dove $z_{\alpha/2}$ è il quantile tale che $P(Z \ge z_{\alpha/2}) = \alpha/2$ e $Z \sim N(0,1)$
- L'*intervallo di confidenza* per μ al $(1-\alpha)100\%$ è dato da
$$\left(\overline{X}-z_{\alpha/2}\sqrt{S^2/n}\ ,\ \overline{X}+z_{\alpha/2}\sqrt{S^2/n}\right)$$
la cui ampiezza è $A = 2z_{\alpha/2}\sqrt{S^2/n}$

Esempio

- Si supponga di voler costruire l'*intervallo di confidenza* al 90% per μ sulla base di un campione di dimensione n = 100 che ha media e varianza campionarie pari rispettivamente a
$$\bar{x} = 15{,}52 \quad e \quad s^2 = 24{,}47$$
- Dato che $1-\alpha = 0{,}90$, si ha che $\alpha = 0{,}10$ e il corrispondente *quantile* della distribuzione Normale standard è pari a $z_{\alpha/2} = z_{0,05} = 1{,}645$. Si noti che $\Phi(z_{0,05}) = 0{,}95 = 1-\alpha/2$.
- L'*intervallo di confidenza* è quindi
$$\left(15{,}52-1{,}645\sqrt{24{,}47/100}\ ,\ 15{,}52+1{,}645\sqrt{24{,}47/100}\right) = (14{,}706\ ,\ 16{,}334)$$
che ha ampiezza pari a
$$A = 2\cdot 1{,}645\sqrt{24{,}47/100} = 1{,}627$$

Intervalli di confidenza per la media

(popolazione Bernoulliana, grandi campioni)

- Per una *popolazione Bernoulliana*, nel caso di *grandi campioni* (n≥30) si ha che
$$\frac{\overline{X}-\pi}{\sqrt{\overline{X}(1-\overline{X})/n}}$$
ha approssimativamente distribuzione Normale standard.
- Di conseguenza
$$P\left(-z_{\alpha/2} \le \frac{\overline{X}-\pi}{\sqrt{\overline{X}(1-\overline{X})/n}} \le z_{\alpha/2}\right) = 1-\alpha$$
dove $z_{\alpha/2}$ è il quantile tale che $P(Z \ge z_{\alpha/2}) = \alpha/2$ e $Z \sim N(0,1)$
- L'*intervallo di confidenza* per π al $(1-\alpha)100\%$ è dato da
$$\left(\overline{X}-z_{\alpha/2}\sqrt{\overline{X}(1-\overline{X})/n}\ ,\ \overline{X}+z_{\alpha/2}\sqrt{\overline{X}(1-\overline{X})/n}\right)$$
la cui ampiezza è
$$A = 2z_{\alpha/2}\sqrt{\overline{X}(1-\overline{X})/n}$$

Esempio

- Si supponga di voler costruire l'*intervallo di confidenza* al 99% per π sulla base di un campione di dimensione n = 50 in cui ci sono $\sum_{i=1}^{n} x_i = 32$ successi, da cui
$$\bar{x} = 32/50 = 0{,}64 \quad e \quad \bar{x}(1-\bar{x}) = 0{,}64\cdot(1-0{,}64) = 0{,}23$$
- Dato che $1-\alpha = 0{,}99$, si ha che $\alpha = 0{,}01$ e il corrispondente *quantile* della distribuzione Normale standard è pari a $z_{\alpha/2} = z_{0,005} = 2{,}576$. Si noti che $\Phi(z_{0,005}) = 0{,}995 = 1-\alpha/2$.
- L'*intervallo di confidenza* è quindi
$$\left(0{,}64-2{,}576\sqrt{0{,}23/50}\ ,\ 0{,}64+2{,}576\sqrt{0{,}23/50}\right) = (0{,}465\ ,\ 0{,}815)$$
che ha ampiezza pari a
$$A = 2\cdot 2{,}576\sqrt{0{,}23/50} = 0{,}350$$

Intervalli di confidenza per la varianza

(popolazione normale, media e varianza non note)

- Se $X \sim N(\mu,\sigma^2)$ con μ *non noto,* si ha

$$\frac{(n-1)S^2}{\sigma^2} \approx \chi^2(n-1)$$

- Di conseguenza

$$P\left(\chi^2_{1-\alpha/2} \le \frac{(n-1)S^2}{\sigma^2} \le \chi^2_{\alpha/2}\right) = 1-\alpha$$

dove $\chi^2_{1-\alpha/2}$ *e* $\chi^2_{\alpha/2}$ sono i quantili della distribuzione χ^2(n-1) tali che

$$P\left(\chi^2 \ge \chi^2_{1-\alpha/2}\right) = 1-\alpha/2 \quad e \quad P\left(\chi^2 \ge \chi^2_{\alpha/2}\right) = \alpha/2$$

- L'*intervallo di confidenza* per σ^2 al (1-α)100% è dato da

$$\left((n-1)S^2 / \chi^2_{\alpha/2}\ ,\ (n-1)S^2 / \chi^2_{1-\alpha/2}\right)$$

Esempio

- Si supponga di voler costruire l'*intervallo di confidenza* al 90% per σ^2 sulla base del campione (già visto in precedenza)

170,75 186,14 173,39 185,12 173,39

sotto l'assunzione $X \sim N(\mu,\sigma^2)$ con μ *non noto.*

- La *varianza campionaria* è pari a

$$s^2 = 52{,}932;$$

dato che 1-α = 0,90, si ha che α = 0,10 e i corrispondenti *quantili* della distribuzione χ^2(4) sono pari a

$$\chi^2_{1-\alpha/2} = \chi^2_{0,95} = 0{,}71 \quad e \quad \chi^2_{\alpha/2} = \chi^2_{0,05} = 9{,}49$$

- L'*intervallo di confidenza* è quindi

$$\left(4/9{,}49 \cdot 52{,}932\ ,\ 4/0{,}71 \cdot 52{,}932\right) = \left(22{,}311\ ,\ 298{,}208\right)$$

che ha ampiezza pari a $A = 298{,}208 - 22{,}311 = 275{,}897$

Dove e come studiare

- Libro di testo: S. Borra, A. Di Ciaccio (2014), Cap. 11 e Cap. 12 (del paragrafo 12.5 solo Intervallo approssimato per grandi campioni)
- Svolgere 'Esercitazione 7'.