# STATISTICA INFERENZIALE (II): stimatori e stima intervallare

## QUALE E' L'AFFIDABILITA' DELLE STIME OTTENUTE CON IL CCS ? [9]

L'**affidabilità** delle stime ottenute come sopra è conseguenza dalle proprietà **della v.a. (o STIMATORE) di cui ciascuna stima è il valore osservato**

p. es. $\bar{x}_n$ e $s_n^2$ sono i valori osservati delle v.a. (Stimatori) $\bar{X}_n$ e $S_n^2$ **(vedi riquadri [10] e [11])**

Un effetto dell'affidabilità delle stime $\hat{p}(x)$, $\bar{x}_n = \hat{\mu}_X$, $\hat{E}(X^2)$ e $s_n^2 = \hat{\sigma}_X^2$ è che,

se la v.a. $X$ di riferimento è effettivamente sempre la stessa $\forall t = 1, 2, \dots n$, allora **per** $n \to \infty$

si osserva che **le stime "tendono a stabilizzarsi"** (vedi la colonna con "$\to$" nel riquadro [6])

**attorno ad un certo valore** che si assume essere il valore approssimato del parametro considerato

## STIMATORE e STIMA di $\mu_X$ [10]
## AFFIDABILITA' della STIMA

(*Schema CCS per la stima di* $\mu_X$)

$$\begin{array}{cccccccc}
X = X_1, & X = X_2, & & X = X_{n-1}, & X = X_n, & & \\
(X_1, & X_2, & \cdots & X_{n-1}, & X_n) & \Rightarrow & \boxed{\frac{1}{n}\sum_{t=1}^{n} X_t = \bar{X}_n} \\
\Downarrow & \Downarrow & \cdots & \Downarrow & \Downarrow & & \Downarrow \\
(x_1, & x_2, & \cdots & x_{n-1}, & x_n) & \Rightarrow & \frac{1}{n}\sum_{t=1}^{n} x_t = \bar{x}_n
\end{array}$$

regola di stima di $\mu_X$: $\boxed{\bar{x}_n = \hat{\mu}_X}$

La v.a. $\bar{X}_n$ si chiama **media campionaria** perché

ognuno dei suoi valori $\bar{x}_n$ è la media dei valori $x_t$ di una realizzazione campionaria $(x_1, x_2, \dots x_n)$

La media campionaria $\bar{X}_n$ si dice che è lo **STIMATORE** del parametro $\mu_X$ perché

la stima di $\mu_X$ è data dal valore osservato $\bar{x}_n$ della media campionaria $\bar{X}_n$ stessa

La stima di $\mu_X$ data da $\boxed{\bar{x}_n = \hat{\mu}_X}$ è **AFFIDABILE** perché

$\bar{X}_n$ è STIMATORE **CONSISTENTE** di $\mu_X$

che vuol dire che $\bar{X}_n$ ha le seguenti due proprietà (a) e (b)

$$(a)\ \lim_{n\to\infty} E(\bar{X}_n) = \mu_X \quad \left(\text{in particolare: } E(\bar{X}_n) = \mu_X \ \forall n\right)$$

$$(b)\ \lim_{n\to\infty} V(\bar{X}_n) = 0 \quad \left(\text{in particolare: } \lim_{n\to\infty} \frac{\sigma_X^2}{n} = 0\right)$$

**(a)** di cui sopra è la proprietà di **NON-DISTORSIONE ASINTOTICA** ed il suo caso particolare tra parentesi è la proprietà di **NON-DISTORSIONE**

**(b)** di cui sopra significa che per $n \to \infty$

**la dispersione** dei valori $\bar{x}_n$ **rispetto a** $\mu_X$ **diminuisce (tende a 0)** e quindi

**la probabilità** di osservare valori $\bar{x}_n$ **poco lontani (poco diversi) da** $\mu_X$ **aumenta (tende ad 1)**

Conclusione: da (a) e (b) segue dunque l'**AFFIDABILITÀ** della stima di $\mu_X$ data da $\boxed{\bar{x}_n = \hat{\mu}_X}$

**N.B. Rispondere alla domanda (3b) dell'Es. nei riquadri [8] e [14] (la risposta è nel riquadro [14bis])**

# STIMATORI e STIMA di $\sigma_X^2$ [11]
# AFFIDABILITA' della STIMA

(*Schema CCS per la stima di* $\sigma_X^2$)

$$\begin{array}{ccccccc} X \equiv X_1, & X = X_2, & \dots & X = X_{n-1}, & X = X_n, & & \\ (X_1, & X_2, & \dots & X_{n-1}, & X_n) & \Rightarrow & \boxed{\left[\frac{1}{n}\sum_{t=1}^{n} X_t^2 - \left(\bar{X}_n\right)^2\right]\frac{n}{n-1} = S_n^2\frac{n}{n-1} = S_c^2} \\ \Downarrow & \Downarrow & \dots & \Downarrow & \Downarrow & & \Downarrow \\ (x_1, & x_2, & & x_{n-1}, & x_n) & \Rightarrow & \left[\frac{1}{n}\sum_{t=1}^{n} x_t^2 - \left(\bar{x}_n\right)^2\right]\frac{n}{n-1} = s_n^2\frac{n}{n-1} = s_c^2 \end{array}$$

regola di stima di $\sigma_X^2$: $\boxed{s_c^2 = s_n^2\frac{n}{n-1} = \hat{\sigma}_X^2}$ e (con $n \geq 30$) anche: $\boxed{s_n^2 = \hat{\sigma}_X^2}$

La v.a. $S_n^2$ si chiama **VARIANZA CAMPIONARIA** perché

ognuno dei suoi valori $s_n^2$ è la **varianza** dei valori $x_t$ di una realizzazione **campionaria** $\left(x_1, x_2, \dots x_n\right)$

$S_n^2$ ha due formule equivalenti e similmente suoi valori possibili $s_n^2$

$$\boxed{S_n^2} = \boxed{\frac{1}{n}\sum_{t=1}^{n} X_t^2 - \left(\bar{X}_n\right)^2} = \boxed{\frac{1}{n}\sum_{t=1}^{n}\left(X_t - \bar{X}_n\right)^2}, \quad \boxed{s_n^2} = \frac{1}{n}\sum_{t=1}^{n} x_t^2 - \left(\bar{x}_n\right)^2 = \frac{1}{n}\sum_{t=1}^{n}\left(x_t - \bar{x}_n\right)^2$$

La varianza campionaria $S_n^2$ si dice che è lo **STIMATORE** del parametro $\sigma_X^2$ perché

la stima di $\sigma_X^2$ è data dal valore osservato $s_n^2$ della varianza campionaria $S_n^2$ stessa

$S_n^2$ è stimatore **DISTORTO** $\boxed{\textbf{ma}}$ **ASINTOTICAMENTE NON-DISTORTO** di $\sigma_X^2$ perché

$$\boxed{E\left(S_n^2\right) = \sigma_X^2\frac{n-1}{n} \neq \sigma_X^2 \quad \boxed{\textbf{ma}} \quad \lim_{n\to\infty} E\left(S_n^2\right) = \sigma_X^2\frac{n-1}{n} = \sigma_X^2}$$

ne segue che $S_n^2$ si usa per stimare $\sigma_X^2$ con **grandi campioni** ($n \geq 30$)

La v.a. $\boxed{S_c^2} = \boxed{S_n^2\frac{n}{n-1} = \frac{1}{n-1}\sum_{t=1}^{n}\left(X_t - \bar{X}_n\right)^2}$ con valori possibili $\boxed{s_c^2} = s_n^2\frac{n}{n-1} = \frac{1}{n-1}\sum_{t=1}^{n}\left(x_t - \bar{x}_n\right)^2$

si chiama VARIANZA CAMPIONARIA **CORRETTA** perché

è data dalla varianza campionaria $S_n^2$ moltiplicata per il **fattore di correzione** $\frac{n}{n-1}$ **che rende**

$S_c^2$ stimatore **NON-DISTORTO** di $\sigma_X^2$, ovvero si ha: $\boxed{E\left(S_c^2\right) = \sigma_X^2 \;\; \forall n}$

La stima di $\sigma_X^2$ data

$$\boxed{\text{da } s_c^2 = s_n^2\frac{n}{n-1} = \hat{\sigma}_X^2 \;\; \forall n, \text{ o da } s_n^2 = \hat{\sigma}_X^2 \text{ con } n \geq 30}$$

è **AFFIDABILE** perché

$\boxed{S_c^2 \text{ e } S_n^2 \text{ sono } \textbf{entrambi} \text{ STIMATORI } \textbf{CONSISTENTI} \text{ di } \sigma_X^2}$

che vuol dire che hanno le seguenti due proprietà (a) e (b)

$$\boxed{\begin{array}{l} (a)\; \lim_{n\to\infty} E\left(S_n^2\right) = \sigma_X^2, \; \lim_{n\to\infty} E\left(S_c^2\right) = \sigma_X^2 \; \left(\text{in particolare: } E\left(S_c^2\right) = \sigma_X^2 \;\; \forall n\right) \\ (b)\; \lim_{n\to\infty} V\left(S_n^2\right) = 0, \; \lim_{n\to\infty} V\left(S_c^2\right) = 0 \end{array}}$$

**N.B. Rispondere alla domanda (4b) dell'Es. nei riquadri [8] e [14] (la risposta è nel riquadro [14bis]).**

## DEFINIZIONE GENERALE DELLE PROPRIETA' DI **CONSISTENZA ED EFFICENZA** DEGLI STIMATORI [12]

Data una popolazione statistica $X$ con un parametro non noto $\theta$, un campione statistico $(X_1, X_2, \ldots X_n)$ ed uno stimatore $T_n = g(X_1, X_2, \ldots X_n)$, si dice che lo stimatore $T_n$ **è stimatore consistente per il parametro** $\theta$ se valgono le seguenti due proprietà (a) e (b):

$$\boxed{\begin{array}{l}(a)\ \lim\limits_{n\to\infty} E(T_n) = \theta \ \left(\text{in particolare se } E(T_n) = \theta\ \forall n\right) \\ (b)\ \lim\limits_{n\to\infty} V(T_n) = 0\end{array}}$$

dove (a) è la proprietà di **non-distorsione asintotica** ed il suo caso particolare è la **non-distorsione**.
Dato un altro stimatore $T'_n = g'(X_1, X_2, \ldots X_n)$ dello stesso parametro $\theta$, $T_n$ **è più efficiente di** $T'_n$ **se**

$$\boxed{EQM_{T_n}(\theta) = V(T_n) + \left[D_{T_n}(\theta)\right]^2 < EQM_{T'_n}(\theta) = V(T'_n) + \left[D_{T'_n}(\theta)\right]^2}$$

dove $EQM$ è l'**errore quadratico medio** dello stimatore ovvero **la somma di due addendi** che sono: la **varianza** dello stimatore e la **distorsione al quadrato** dello stimatore. La **distorsione** è data da

$$\boxed{D_{T_n}(\theta) = E(T_n) - \theta, \text{ od anche } D_{T_n}(\theta) = \theta - E(T_n)}$$

(ovviamente **la distorsione è zero** se lo stimatore è **non-distorto** per $\theta$ )

## STIMA DI $\mu_X$ con INTERVALLO DI CONFIDENZA [13]

$IC_{1-\alpha}(\mu_X)$ (con grandi campioni: $n \geq 30$ )

$$\boxed{IC_{1-\alpha}(\mu_X) = (\hat{\mu}_X - \Delta_\alpha, \hat{\mu}_X + \Delta_\alpha) = (\overline{x}_n - \Delta_\alpha, \overline{x}_n + \Delta_\alpha)}$$

punto centrale $\boxed{\hat{\mu}_X = \overline{x}_n}$

estremo inferiore $\boxed{\overline{x}_n - \Delta_\alpha}$ ________________|_______________ $\boxed{\overline{x}_n + \Delta_\alpha}$ estremo superiore

$$\boxed{\begin{array}{c}\Delta_\alpha = \sqrt{\dfrac{\sigma_X^2}{n}} z_{1-\alpha/2} = \dfrac{\sigma_X}{\sqrt{n}} z_{1-\alpha/2} \ (\text{con } \sigma_X^2 \text{ noto}) \\ \Delta_\alpha = \sqrt{\dfrac{\hat{\sigma}_X^2}{n}} z_{1-\alpha/2} = \dfrac{\hat{\sigma}_X}{\sqrt{n}} z_{1-\alpha/2} \ (\text{con } \sigma_X^2 \text{ non noto e stimato})\end{array}}$$

**dove** $\boxed{\Delta_\alpha}$ **è tale che**

$1-\alpha$ è l'area (probabilità) sopra l'intervallo $IC_{1-\alpha}(\mu_X)$, ovvero più precisamente

$1-\alpha$ è l'area (probabilità) fra l'intervallo $IC_{1-\alpha}(\mu_X)$ e la densità

di $\overline{X}_n \sim N\left(\hat{\mu}_X; \dfrac{\sigma_X^2}{n}\right)$ (con $\sigma_X^2$ noto) o di $\overline{X}_n \sim N\left(\hat{\mu}_X; \dfrac{\hat{\sigma}_X^2}{n}\right)$ (con $\sigma_X^2$ non noto e stimato)

$\boxed{z_{1-\alpha/2}}$ **è il valore di** $Z \sim N(0;1)$ **tale che** alla sua sinistra c'è un'area (probabilità) pari a $1-\alpha/2$

$\boxed{z_{1-\alpha/2}}$ si trova con la ricerca inversa sulla tavola numerica di $N(0;1)$

$$\boxed{\begin{array}{ccccc} P\big(N(0;1) \leq & z_{1-\alpha/2}\big) & = & & 1-\alpha/2 \\ & \uparrow & & & \downarrow \\ & \nwarrow & \underset{\textit{tavola di } N(0;1)}{\longleftarrow} & & \swarrow \end{array}}$$

**N.B. Rispondere alla domanda (5) dell'Es. nei riquadri [8] e [14] (la risposta è nel riquadro [14bis])**

**[14]**

Ripresa e continuazione dell'Esercizio nel riquadro [8]. Domande:
(**3b)** si indichi inoltre lo stimatore utilizzato e le sue proprietà (si veda la domanda (3a) nel riquadro [8])
**(4b)** si indichi inoltre lo stimatore utilizzato e le sue proprietà (si veda la domanda (4a) nel riquadro [8])
**(5)** si determini, sulla base della realizzazione campionaria osservata, l'intervallo di confidenza del reddito medio atteso $\mu_X$ con coefficiente di confidenza $1-\alpha = 0.95$
**(6)** Infine, sulla base della realizzazione campionaria osservata, si esegua test statistico bilaterale per il reddito medio atteso $\mu_X$ : $H_0 : \mu_X = 43.5$ contro $H_1 : \mu_X \neq 43.5$ al livello di significatività $\alpha = 0.05$
(la risposta a questa domanda è nel riquadro [23bis] e richiede la lettura dei riquadri da [15] a [21]).

**[14bis]**

RISPOSTA ALLA DOMANDA (3b). (vedi i riquadri [9] e [10]). Lo stimatore utilizzato è la media campionaria $\overline{X}_{81} = \frac{1}{81}\sum_{t=1}^{81} X_t$ . Lo stimatore $\overline{X}_n$ è **stimatore consistente** (ha la **proprietà di consistenza**) per il parametro $\mu_X$ , cioè $\overline{X}_n$ ha le proprietà (a) e (b) esposte nel riquadro [10].

RISPOSTA ALLA DOMANDA (4b). (vedi i riquadri [9] e [11]). Lo stimatore utilizzato è la varianza campionaria $S_n^2$ o la varianza campionaria corretta $S_c^2 = S_n^2 \dfrac{n}{n-1}$ . Entrambi gli stimatori hanno le proprietà (a) e (b) esposte nel riquadro [11]), sono cioè **stimatori consistenti** (hanno la **proprietà di consistenza**) per il parametro $\sigma_X^2$ .

RISPOSTA ALLA DOMANDA (5).

**DATI:** $\hat{\mu}_X = \overline{x}_{81} = 41$ (vedi risposta a domanda (3a), riquadro [8bis]),
$s_{81}^2 = 546.6667 = \hat{\sigma}_X^2$ , oppure $s_c^2 = 553.5000 = \hat{\sigma}_X^2 \left(\cong 546.6667 = s_{81}^2\right)$ (vedi risposta a domanda (4a), riquadro [8bis]), $\alpha = 0.05$, $n \geq 30$ . Infine da $X \sim N\left(\mu_X ; \sigma_X^2\right)$ segue che $\overline{X}_{81} \sim N\left(\mu_X ; \dfrac{\sigma_X^2}{81}\right)$. Ciò vale perché la media campionaria $\overline{X}_n$ è combinazione lineare di *n* = 81 v.a. $X_t$ tutte uguali alla **popolazione statistica gaussiana** $X \sim N\left(\mu_X ; \sigma_X^2\right)$, ne segue che $\overline{X}_n$ è una **gaussiana sia** con "grandi campioni" (come nel nostro caso, *n* = 81>30) **sia** con "piccoli campioni" (*n* < 30).

**CALCOLI** (vedi riquadro [12])**:**
$\alpha = 0.05$, $\alpha/2 = 0.025$, $\boxed{1-\alpha/2 = 0.975}$, $\boxed{z_{0.975} = 1.96}$, $1-\alpha = 0.95$.

Determiniamo $IC_{0.95}\left(\mu_X\right)$ prima con la stima puntuale $\hat{\sigma}_X^2 = s_{81}^2 = 546.6667$ e poi con la stima puntuale $\hat{\sigma}_X^2 = s_c^2 = 553.5000$.

Con la stima puntuale $\hat{\sigma}_X^2 = s_{81}^2$ si ottiene:

$$\sqrt{\frac{\hat{\sigma}_X^2}{n}} = \sqrt{\frac{s_{81}^2}{81}} = \sqrt{\frac{546.6667}{81}} = \sqrt{6.749} = 2.5979 \Rightarrow \boxed{\Delta_{0.05}} = 2.5979 \times 1.96 = \boxed{5.0919}$$

$$\boxed{IC_{0.95}\left(\mu_X\right) =} \left(\overline{x}_n - \Delta_\alpha, \overline{x}_n + \Delta_\alpha\right) = \left(41 - 5.0919, 41 + 5.0919\right) \boxed{= \left(35.9081, 46.0919\right)}$$

Con la stima puntuale $\hat{\sigma}_X^2 = s_c^2$ si ottiene

$$\sqrt{\frac{\hat{\sigma}_X^2}{n}} = \sqrt{\frac{s_c^2}{81}} = \sqrt{\frac{553.500}{81}} = \sqrt{6.8333} = 2.6140 \Rightarrow \boxed{\Delta_{0.05}} = 2.6140 \times 1.96 = \boxed{5.1234}$$

$$\boxed{IC_{0.95}\left(\mu_X\right) =} \left(\overline{x}_n - \Delta_\alpha, \overline{x}_n + \Delta_\alpha\right) = \left(41 - 5.1234, 41 + 5.1234\right) \boxed{= \left(35.8766, 46.1234\right)}$$