STATISTICA (MODULO II - INFERENZA STATISTICA)
Soluzione esercitazione 4

A. La v.c. $X_i \sim \mathrm{N}(12; 9)$ esprime il consumo giornaliero di acqua potabile da parte della i-esima famiglia. Il consumo medio di un campione di 100 famiglie può essere espresso dalla v.c. $\bar{X} \sim \mathrm{N}(12; 9/100)$.

1.
$$P(\bar{X} > 12{,}6) = P\left(Z > \frac{12{,}6 - 12}{\sqrt{9/100}}\right) = 1 - \Phi(2) = 0{,}02275$$

2. Se si rimuove l'ipotesi di normalità della distribuzione delle v.c. X_i la probabilità calcolata al punto precedente rimane approssimativamente valida grazie al teorema del limite centrale. Infatti, la numerosità campionaria $n = 100$ è sufficientemente grande da garantire che il teorema possa applicarsi.

B. Sulla base della distribuzione della v.c. X è immediato calcolare

$$\mu = \mathrm{E}(X) = 1 \times 0{,}2 + 2 \times 0{,}1 + 3 \times 0{,}3 + 4 \times 0{,}4 = 2{,}9$$
$$\sigma^2 = \mathrm{Var}(X) = (1 - 2{,}9)^2 0{,}2 + (2 - 2{,}9)^2 0{,}1 + (3 - 2{,}9)^2 0{,}3 + (4 - 2{,}9)^2 0{,}4 = 1{,}29$$

1. I possibili campioni di dimensione $n = 2$ sono $4^2 = 16$. Le probabilità di ciascuno di essi e le corrispondenti media e varianza campionaria sono riportare nella seguente tabella:

x_1	x_2	$f(x_1)$	$f(x_2)$	$f(x_1,x_2)$	$\bar{x}$	s^2
1	1	0,2	0,2	0,04	1,0	0,0
1	2	0,2	0,1	0,02	1,5	0,5
1	3	0,2	0,3	0,06	2,0	2,0
1	4	0,2	0,4	0,08	2,5	4,5
2	1	0,1	0,2	0,02	1,5	0,5
2	2	0,1	0,1	0,01	2,0	0,0
2	3	0,1	0,3	0,03	2,5	0,5
2	4	0,1	0,4	0,04	3,0	2,0
3	1	0,3	0,2	0,06	2,0	2,0
3	2	0,3	0,1	0,03	2,5	0,5
3	3	0,3	0,3	0,09	3,0	0,0
3	4	0,3	0,4	0,12	3,5	0,5
4	1	0,4	0,2	0,08	2,5	4,5
4	2	0,4	0,1	0,04	3,0	2,0
4	3	0,4	0,3	0,12	3,5	0,5
4	4	0,4	0,4	0,16	4,0	0,0
				1,00		

2 e 3. Distribuzione della media campionaria

$\bar{x}$	$f(\bar{x})$	$\bar{x}f(\bar{x})$	$[\bar{x} - \mathrm{E}(\bar{X})]^2 f(\bar{x})$
1,0	0,04	0,04	0,1444
1,5	0,04	0,06	0,0784
2,0	0,13	0,26	0,1053
2,5	0,22	0,55	0,0352
3,0	0,17	0,51	0,0017
3,5	0,24	0,84	0,0864
4,0	0,16	0,64	0,1936
	1,00	2,90	0,6450

Le prime due colonne riportano la distribuzione della media campionaria.

La media campionaria è uno stimatore non distorto della media della popolazione in quanto

$$\mathrm{D}(\bar{X}) = \mathrm{E}(\bar{X}) - \mu = 2{,}90 - 2{,}90 = 0$$

dove $\mathrm{E}(\bar{X}) = \sum \bar{x} f(\bar{x}) = 2{,}90$.

L'errore quadratico medio si calcola come segue:

$$\mathrm{EQM}(\bar{X}) = \mathrm{Var}(\bar{X}) + \mathrm{D}(\bar{X})^2 = \mathrm{Var}(\bar{X}) = \sum [\bar{x} - \mathrm{E}(\bar{X})]^2 f(\bar{x}) = 0{,}645$$

Da notare che la varianza della media campionaria è pari alla varianza nella popolazione divisa per la dimensione campionaria, cioè $1{,}29/2 = 0{,}645$.

4 e 5. Distribuzione della varianza campionaria

s^2	$f(s^2)$	$s^2 f(s^2)$
0,0	0,30	0,00
0,5	0,34	0,17
2,0	0,20	0,40
4,5	0,16	0,72
	1,00	1,29

Le prime due colonne riportano la distribuzione della varianza campionaria.

La varianza campionaria è uno stimatore non distorto della varianza della popolazione in quanto

$$\mathrm{D}(S^2) = \mathrm{E}(S^2) - \sigma^2 = 1{,}29 - 1{,}29 = 0$$

dove $\mathrm{E}(S^2) = \sum s^2 f(s^2) = 1{,}29$.

C. Le variabili casuali X_i $(i = 1, \ldots, 5)$ sono indipendenti e tali per cui $\mathrm{E}(X_i) = \mu$, $\mathrm{Var}(X_i) = \sigma^2$.

1. Il valore atteso dello stimatore $T = 0{,}2X_1 + 0{,}3X_2 + 0{,}1X_3 + 0{,}1X_4 + 0{,}3X_5$ è pari a

$$\mathrm{E}(T) = 0{,}2\mu + 0{,}3\mu + 0{,}1\mu + 0{,}1\mu + 0{,}3\mu = (0{,}2 + 0{,}3 + 0{,}1 + 0{,}1 + 0{,}3)\mu = \mu$$

Quindi lo stimatore è non distorto, cioè $\mathrm{D}(T) = \mathrm{E}(T) - \mu = 0$.

L'errore quadratico medio dello stimatore T è pari a

$$\begin{aligned}\mathrm{EQM}(T) &= \mathrm{Var}(T) + \mathrm{D}(T)^2 = \mathrm{Var}(T) = 0{,}2^2\sigma^2 + 0{,}3^2\sigma^2 + 0{,}1^2\sigma^2 + 0{,}1^2\sigma^2 + 0{,}3^2\sigma^2 \\ &= (0{,}2^2 + 0{,}3^2 + 0{,}1^2 + 0{,}1^2 + 0{,}3^2)\sigma^2 = 0{,}24\sigma^2\end{aligned}$$

2. Nel confronto tra stimatori l'efficienza può essere valutata confrontando i rispettivi EQM.

La media campionaria $\bar{X}$ è uno stimatore non distorto di μ, per cui l'errore quadratico medio è pari a

$$\mathrm{EQM}(\bar{X}) = \mathrm{Var}(\bar{X}) = \sigma^2/n = \sigma^2/5 = 0{,}2\sigma^2$$

Dal momento che $\mathrm{EQM}(T) > \mathrm{EQM}(\bar{X})$, lo stimatore T è uno stimatore meno efficiente di $\bar{X}$.

D. Dal momento che $\dfrac{S^2(n-1)}{\sigma^2} \sim \chi^2_{n-1}$,

$$\begin{aligned}P(17{,}04 < S^2 < 36{,}69) &= P\left(\frac{17{,}04(30-1)}{25} < \frac{S^2(n-1)}{\sigma^2} < \frac{36{,}69(30-1)}{25}\right) = P(19{,}77 < \chi^2_{30-1} < 42{,}56) \\ &= P(\chi^2_{29} < 42{,}56) - P(\chi^2_{29} < 19{,}77) = 0{,}95 - 0{,}10 = 0{,}85\end{aligned}$$

E. Sia $X \sim \mathrm{N}(74; 18^2)$ la v.c. che descrive il punteggio dell'indicatore della qualità della vita in oggetto.

1. L'errore standard della media per campioni di dimensione $n = 100$ è pari a

$$\mathrm{ES}(\bar{X}) = \sqrt{\mathrm{Var}(\bar{X})} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}} = \frac{18}{\sqrt{100}} = 1{,}8$$

   il quale indica che, in media, l'errore che si commette assegnando al punteggio medio della popolazione il valore fornito dalla media campionaria è pari a 1,8.

2. Dalla $\mathrm{ES}(\bar{X}) = \dfrac{\sigma}{\sqrt{n}}$ occorre fissare $\mathrm{ES}(\bar{X})$ al livello desiderato, cioè $1{,}8/2 = 0{,}9$, e risolvere rispetto ad n. Quindi:

$$0{,}9 = \frac{18}{\sqrt{n}} \qquad \longrightarrow \qquad n = \left(\frac{18}{0{,}9}\right)^2 = 400$$

F. Sulla base della distribuzione della v.c. $X \sim \mathrm{Bern}(0{,}4)$ è immediato calcolare $\mathrm{E}(X) = p = 0{,}4$ e $\mathrm{Var}(X) = p(1-p) = 0{,}24$.

1. I possibili campioni di dimensione $n = 3$ sono $2^3 = 8$. Le probabilità di ognuno di essi e le corrispondenti media e varianza campionarie sono riportare nella seguente tabella:

x_1	x_2	x_2	$f(x_1)$	$f(x_2)$	$f(x_3)$	$f(x_1,x_2,x_3)$	$\bar{x}$	s^2
0	0	0	0,6	0,6	0,6	0,216	0,0000	0,0000
0	0	1	0,6	0,6	0,4	0,144	0,3333	0,3333
0	1	0	0,6	0,4	0,6	0,144	0,3333	0,3333
0	1	1	0,6	0,4	0,4	0,096	0,6667	0,3333
1	0	0	0,4	0,6	0,6	0,144	0,3333	0,3333
1	0	1	0,4	0,6	0,4	0,096	0,6667	0,3333
1	1	0	0,4	0,4	0,6	0,096	0,6667	0,3333
1	1	1	0,4	0,4	0,4	0,064	1,0000	0,0000
						1,0000		

2 e 3. Dalla tabella precedente si ottiene la seguente distribuzione della media campionaria (prime due colonne):

$\bar{x}$	$f(\bar{x})$	$\bar{x}f(\bar{x})$	$[\bar{x} - \mathrm{E}(\bar{X})]^2 f(\bar{x})$
0,0000	0,216	0,000	0,03456
0,3333	0,432	0,144	0,00192
0,6667	0,288	0,192	0,02048
1,0000	0,064	0,064	0,02304
	1,000	0,400	0,08000

Le ultime due colonne contengono i calcoli che ci consentono di mostrare come il valore atteso della media campionaria è pari alla media nella popolazione, cioè

$$\mathrm{E}(\bar{X}) = \sum \bar{x} f(\bar{x}) = 0{,}4 = p$$

mentre la varianza della media campionaria è pari alla varianza nella popolazione divisa per la numerosità campionaria, cioè

$$\mathrm{Var}(\bar{X}) = \sum [\bar{x} - \mathrm{E}(\bar{X})]^2 f(\bar{x}) = 0{,}08 = \frac{0{,}24}{3} = \frac{p(1-p)}{n}$$

4.

$$\mathrm{EQM}(\bar{X}) = \mathrm{Var}(\bar{X}) + \mathrm{D}(\bar{X})^2 = 0{,}08 + (0{,}4 - 0{,}4)^2 = 0{,}08$$