STATISTICA esercizi svolti sulla STIMA PUNTUALE

1 STIMA PUNTUALE

1. Siano A, B e C tre stimatori con i quali si vuole stimare la media Q di una popolazione. Tali stimatori presentano le seguenti distribuzioni:

A	$P(A)$	B	$P(B)$	C	$P(C)$
95	0.10	96	0.25	98	0.10
100	0.80	97	0.25	99	0.40
105	0.10	98	0.25	101	0.40
		99	0.25	102	0.10

Supponendo che la vera media Q sia pari a 100:
a) valutare la correttezza dei tre stimatori;
b) scegliere, tra gli stimatori corretti, il più efficiente.

Svolgimento

a) Lo stimatore A risulta essere uno stimatore corretto per Q, in quanto il suo valore atteso coincide con il vero valore di Q:

$$E(A) = \sum_{i=1}^{3} a_i \cdot P(A = a_i) = 95 \cdot 0.10 + 100 \cdot 0.80 + 105 \cdot 0.10 = 100.$$

Lo stimatore B non è corretto (per il parametro Q), infatti

$$E(B) = \sum_{i=1}^{4} b_i \cdot P(B = b_i) = 96 \cdot 0.25 + 97 \cdot 0.25 + 98 \cdot 0.25 + 99 \cdot 0.25 = 97.5$$

e quindi $E(B) \neq 100$.

Lo stimatore C è corretto (per il parametro Q), infatti

$$E(C) = \sum_{i=1}^{4} c_i \cdot P(C = c_i) = 98 \cdot 0.10 + 99 \cdot 0.40 + 101 \cdot 0.40 + 102 \cdot 0.10 = 100$$

e quindi il suo valore atteso coincide con il vero valore di Q.

b) Per poter determinare lo stimatore più efficiente è necessario calcolare la varianza degli stimatori non distorti.

$$\begin{aligned} Var(A) &= \sum_{i=1}^{3} a_i^2 \cdot P(A = a_i) - [E(A)]^2 \\ &= 95^2 \cdot 0.10 + 100^2 \cdot 0.80 + 105^2 \cdot 0.10 - 100^2 = 5. \end{aligned}$$

$$\begin{aligned} Var(C) &= \sum_{i=1}^{4} c_i^2 \cdot P(C = c_i) - [E(C)]^2 \\ &= 98^2 \cdot 0.10 + 99^2 \cdot 0.40 + 101^2 \cdot 0.40 + 102^2 \cdot 0.10 - 100^2 = 1.6. \end{aligned}$$

Confrontando i valori della varianza degli stimatori A e C (che abbiamo visto essere corretti), dal momento che

$$Var(A) = 5 > 1.6 = Var(C)$$

possiamo concludere che lo stimatore più efficiente tra i corretti è lo stimatore C (perchè ha varianza minore).

2. Dato un campione di tre elementi estratti con riposizione da una popolazione di media μ e varianza σ^2, scegliere quale tra i seguenti due stimatori è preferibile:

$$T_1 = \frac{X_1 + 2X_2 + 2X_3}{5} \qquad T_2 = \frac{X_1 + 2X_2 + X_3}{4}.$$

Svolgimento
É noto che il campione (X_1, X_2, X_3), essendo estratto con riposizione, ha componenti indipendenti e identicamente distribuite, quindi:

$$E(X_i) = \mu \qquad \text{e} \quad Var(X_i) = \sigma^2 \qquad i = 1, 2, 3.$$

Analizziamo la correttezza degli stimatori T_1 e T_2, utilizzando le proprietà del valore atteso:

$$\begin{aligned} E(T_1) &= E\left[\frac{X_1 + 2X_2 + 2X_3}{5}\right] = \frac{1}{5} \cdot E[X_1 + 2X_2 + 2X_3] = \\ &= \frac{1}{5} \cdot [E(X_1) + 2E(X_2) + 2E(X_3)] = \frac{1}{5} \cdot [\mu + 2\mu + 2\mu] = \mu. \end{aligned}$$

$$\begin{aligned} E(T_2) &= E\left[\frac{X_1 + 2X_2 + X_3}{4}\right] = \frac{1}{4} \cdot E[X_1 + 2X_2 + X_3] = \\ &= \frac{1}{4} \cdot [E(X_1) + 2E(X_2) + E(X_3)] = \frac{1}{4} \cdot [\mu + 2\mu + \mu] = \mu. \end{aligned}$$

Entrambi gli stimatori T_1 e T_2 risultano corretti per l'ignota media μ (della popolazione di riferimento).

Per determinare il più efficiente, è necessario quindi calcolare le loro varianze. Utilizzando le proprietà della varianza e ricordando che le X_i sono indipendenti, ricaviamo che:

$$\begin{aligned} Var(T_1) &= Var\left[\frac{X_1 + 2X_2 + 2X_3}{5}\right] = \left(\frac{1}{5}\right)^2 \cdot Var[X_1 + 2X_2 + 2X_3] = \\ &= \frac{1}{25} \cdot \left[Var(X_1) + 2^2 Var(X_2) + 2^2 Var(X_3)\right] = \\ &= \frac{1}{25}[\sigma^2 + 4\sigma^2 + 4\sigma^2] = \frac{9}{5} \cdot \sigma^2 = 0.366 \cdot \sigma^2 \end{aligned}$$

$$\begin{aligned} Var(T_2) &= Var\left[\frac{X_1 + 2X_2 + X_3}{4}\right] = \left(\frac{1}{4}\right)^2 \cdot Var[X_1 + 2X_2 + X_3] = \\ &= \frac{1}{16} \cdot \left[Var(X_1) + 2^2 Var(X_2) + Var(X_3)\right] = \\ &= \frac{1}{16}[\sigma^2 + 4\sigma^2 + \sigma^2] = \frac{6}{16} \cdot \sigma^2 = 0.375 \cdot \sigma^2. \end{aligned}$$

Sebbene non sia nota la quantità σ^2 confrontando i valori della varianza degli stimatori T_1 e T_2, dal momento che

$$Var(T_1) = 0.366 \cdot \sigma^2 < 0.375 \cdot \sigma^2 = Var(T_2)$$

possiamo concludere che lo stimatore più efficiente è lo stimatore T_1 (perchè ha varianza inferiore).

3. Sia θ la proporzione ignota di palline bianche contenute in un'urna. Si estraggono con riposizione due palline e si propongono i seguenti stimatori di θ:

$$T_1 = \frac{1}{3}X_1 + \frac{2}{3}X_2 \qquad T_2 = \frac{3}{4}X_1 + \frac{1}{4}X_2 \qquad T_3 = \frac{1}{2}X_1 + \frac{1}{2}X_2,$$

dove X_i $(i = 1, 2)$ è la variabile che assume valore 1 se viene estratta pallina bianca e 0 se viene estratta una pallina non bianca. Dopo aver verificato che gli stimatori sono non distorti, si individui lo stimatore più efficiente.

Svolgimento

Innanzitutto riconosciamo che le variabili casuali X_i $(i = 1, 2)$ sono variabili casuali indicatori con funzione di probabilità

$$X_i = \begin{cases} 1 & \text{con probabilità } \theta \\ & \\ 0 & \text{con probabilità } 1-\theta \end{cases} \qquad (i = 1, 2)$$

e quindi, come è noto,

$$E(X_i) = 1 \cdot \theta + 0 \cdot (1-\theta) = \theta \qquad\qquad i = 1,2$$
$$Var(X_i) = 1^2 \cdot \theta + 0^2 \cdot (1-\theta) - \theta^2 = \theta - \theta^2 = \theta \cdot (1-\theta) \qquad\qquad i = 1,2.$$

Utilizzando tali risultati, possiamo ora verificare la correttezza dei tre stimatori T_1, T_2 e T_3:

$$E(T_1) = E\left[\frac{1}{3}X_1 + \frac{2}{3}X_2\right] = E\left[\frac{1}{3}X_1\right] + E\left[\frac{2}{3}X_2\right] = \frac{1}{3}\cdot\theta + \frac{2}{3}\cdot\theta = \theta$$

$$E(T_2) = E\left[\frac{3}{4}X_1 + \frac{1}{4}X_2\right] = E\left[\frac{3}{4}X_1\right] + E\left[\frac{1}{4}X_2\right] = \frac{3}{4}\cdot\theta + \frac{1}{4}\cdot\theta = \theta$$

$$E(T_3) = E\left[\frac{1}{2}X_1 + \frac{1}{2}X_2\right] = E\left[\frac{1}{2}X_1\right] + E\left[\frac{1}{2}X_2\right] = \frac{1}{2}\cdot\theta + \frac{1}{2}\cdot\theta = \theta.$$

Per valutare l'efficienza degli stimatori, calcoliamo la varianza di ognuno di essi, ricordando che le due variabili casuali X_1 e X_2 sono indipendenti (poichè le estrazioni avvengono con riposizione):

$$\begin{aligned}
Var(T_1) &= Var\left[\frac{1}{3}X_1 + \frac{2}{3}X_2\right] = Var\left[\frac{1}{3}X_1\right] + Var\left[\frac{2}{3}X_2\right] = \\
&= \left(\frac{1}{3}\right)^2 Var(X_1) + \left(\frac{2}{3}\right)^2 Var(X_2) = \\
&= \frac{1}{9}\cdot\theta(1-\theta) + \frac{4}{9}\cdot\theta(1-\theta) = \frac{5}{9}\cdot\theta(1-\theta).
\end{aligned}$$

$$\begin{aligned}
Var(T_2) &= Var\left[\frac{3}{4}X_1 + \frac{1}{4}X_2\right] = Var\left[\frac{3}{4}X_1\right] + Var\left[\frac{1}{4}X_2\right] = \\
&= \left(\frac{3}{4}\right)^2 Var(X_1) + \left(\frac{1}{4}\right)^2 Var(X_2) = \\
&= \frac{9}{16}\cdot\theta(1-\theta) + \frac{1}{16}\cdot\theta(1-\theta) = \frac{5}{8}\cdot\theta(1-\theta).
\end{aligned}$$

$$\begin{aligned}
Var(T_3) &= Var\left[\frac{1}{2}X_1 + \frac{1}{2}X_2\right] = Var\left[\frac{1}{2}X_1\right] + Var\left[\frac{1}{2}X_2\right] = \\
&= \left(\frac{1}{2}\right)^2 Var(X_1) + \left(\frac{1}{2}\right)^2 Var(X_2) = \\
&= \frac{1}{4}\cdot\theta(1-\theta) + \frac{1}{4}\cdot\theta(1-\theta) = \frac{1}{2}\cdot\theta(1-\theta).
\end{aligned}$$

Riassumendo, si ha:

$$\begin{aligned}
Var(T_1) &= \frac{5}{9} \cdot \theta(1-\theta) = 0.\bar{5} \cdot \theta(1-\theta) \\
Var(T_2) &= \frac{5}{8} \cdot \theta(1-\theta) = 0.625 \cdot \theta(1-\theta) \\
Var(T_3) &= \frac{1}{2} \cdot \theta(1-\theta) = 0.5 \cdot \theta(1-\theta).
\end{aligned}$$

Sebbene la quantità $\theta(1-\theta)$ non sia nota (si sa solo che è una quantità compresa tra 0 e 0.25), è comunque possibile affermare che lo stimatore T_3 è lo stimatore con varianza minore (dato che $0.5 < 0.\bar{5} < 0.625$) e quindi è il più efficiente dei tre.

4. Una popolazione di quattro unità presenta i seguenti valori del carattere X:

$$5 \quad 9 \quad 11 \quad 15.$$

a) Costruire la distribuzione dello stimatore "media aritmetica campionaria" relativa ai campioni di ampiezza 2 estratti casualmente e con ripetizione;
b) verificare che tale stimatore è corretto per la media della popolazione;
c) verificare che relazione sussiste tra la varianza del carattere X e la varianza dello stimatore proposto.

Svolgimento

a) Come è noto, lo spazio campionario è dato da tutte le $4^2 = 16$ coppie ordinate (con ripetizione) che si possono ottenere con i numeri 5,9,11,15. Ciascun campione ha la stessa probabilità di essere estratto.
Completiamo perciò la seguente tabella in cui riportiamo nella prima colonna, tutti i campioni possibili, nella seconda il valore che lo stimatore "media campionaria" assume e nella terza la probabilità che tale campione venga estratto.

Campioni	$\bar{x}_i$	*Prob.*
$(5,5)$	5	$\frac{1}{16}$
$(5,9)$	7	$\frac{1}{16}$
$(5,11)$	8	$\frac{1}{16}$
$(5,15)$	10	$\frac{1}{16}$
$(9,5)$	7	$\frac{1}{16}$
$(9,9)$	9	$\frac{1}{16}$
$(9,11)$	10	$\frac{1}{16}$
$(9,15)$	12	$\frac{1}{16}$
$(11,5)$	8	$\frac{1}{16}$
$(11,9)$	10	$\frac{1}{16}$
$(11,11)$	11	$\frac{1}{16}$
$(11,15)$	13	$\frac{1}{16}$
$(15,5)$	10	$\frac{1}{16}$
$(15,9)$	12	$\frac{1}{16}$
$(15,11)$	13	$\frac{1}{16}$
$(15,15)$	15	$\frac{1}{16}$
		1

É possibile ora sintetizzare la precedente tabella, indicando i valori che può assumere lo stimatore "media campionaria" e le probabilità corrispondenti. Esplicitiamo quindi la distribuzione dello stimatore "media campionaria":

$$\bar{X} = \begin{cases} 5 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 15 \\ \frac{1}{16} & \frac{2}{16} & \frac{2}{16} & \frac{1}{16} & \frac{4}{16} & \frac{1}{16} & \frac{2}{16} & \frac{2}{16} & \frac{1}{16} \end{cases}$$

b) Verifichiamo la correttezza di $\bar{X}$.
Per prima cosa, calcoliamo la vera media della popolazione di riferimento, indicandola con μ:

$$\mu = \frac{5+9+11+15}{4} = 10.$$

Calcoliamo ora il valore atteso di $\bar{X}$:

$$E(\bar{X}) = \sum_{i=1}^{9} \bar{x}_i \cdot P(\bar{X} = \bar{x}_i) =$$

$$= 5\cdot\frac{1}{16}+7\cdot\frac{2}{16}+8\cdot\frac{2}{16}+9\cdot\frac{1}{16}+10\cdot\frac{4}{16}+11\cdot\frac{1}{16}+12\cdot\frac{2}{16}+13\cdot\frac{2}{16}+15\cdot\frac{1}{16} = 10.$$

Poichè tale valore atteso coincide con il vero valore μ, possiamo concludere che lo stimatore "media campionaria" $\bar{X}$ è uno stimatore non distorto per la media della popolazione.

c) Verifichiamo che sussiste la relazione

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

dove si è indicato con σ^2 la varianza della popolazione di riferimento e con n l'ampiezza campionaria.

La varianza dello stimatore $\bar{X}$ è data da:

$$Var(\bar{X}) = E(\bar{X}^2) - [E(\bar{X})]^2 =$$

$$= 5^2\cdot\frac{1}{16}+7^2\cdot\frac{2}{16}+8^2\cdot\frac{2}{16}+9^2\cdot\frac{1}{16}+10^2\cdot\frac{4}{16}+11^2\cdot\frac{1}{16}+12^2\cdot\frac{2}{16}+13^2\cdot\frac{2}{16}+15^2\cdot\frac{1}{16}-10^2 = 6.5.$$

La varianza della popolazione è invece data da:

$$\begin{aligned} \sigma^2 &= \frac{1}{4}\cdot\sum_{i=1}^{4} x_i^2 - \mu^2 \\ &= \frac{5^2+9^2+11^2+15^2}{4} - 10^2 = 13. \end{aligned}$$

Ricordando che l'ampiezza campionaria è pari a 2, possiamo verificare che

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

dal momento che

$$6.5 = \frac{13}{2}.$$

5. Viene rilevata la quantità (in cl) di liquido detergente immesso da un impianto in un campione di 100 contenitori destinati alla vendita:

Quantità (in cl)	$190 \vdash 195$	$195 \vdash 205$	$205 \vdash 210$
Frequenze	15	70	15

Sulla base dell'esperienza passata si ritiene che lo scarto quadratico medio della quantità di liquido immessa dall'impianto sia pari a 5 cl. Si vuole stimare la quantità media μ di liquido immessa dall'impianto in un contenitore.
a) Si stimi puntualmente μ e si indichi una misura della variabilità dello stimatore utilizzato;
b) si determini la probabilità di stimare μ con un errore inferiore a 1 cl;
c) si determini quanti contenitori occorre ancora analizzare se si vuole che lo scarto quadratico medio dello stimatore di μ sia inferiore a 0.25;
d) si determini quanti contenitori occorre ancora analizzare se si vuole che lo stimatore della quantità media del liquido immesso dall'impianto si discosti dal vero contenuto medio per meno di 1 cl con probabilità pari a 0.99;
e) considerando non nota l'informazione relativa allo scarto quadratico medio della popolazione di riferimento, si stimi la varianza σ^2 della quantità di liquido immessa dall'impianto.

Svolgimento

a) Per stimare l'ignota quantità media μ, utilizziamo lo stimatore "media campionaria" $\bar{X}$. Per calcolare il suo valore, utilizzando i valori centrali delle classi si ottiene:

$$\bar{x} = \frac{192.5 \cdot 15 + 200 \cdot 70 + 207.5 \cdot 15}{100} = 200.$$

Per misurare la variabilità dello stimatore "media campionaria", calcoliamo la sua varianza che è data da:

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

dove si è indicato con σ^2 la varianza della popolazione di riferimento e con n l'ampiezza campionaria.
Si ha quindi:

$$Var(\bar{X}) = \frac{5^2}{100} = 0.25.$$

b) La probabilità richiesta è data da:

$$P[|\bar{X} - \mu| < 1].$$

Ricordando che per $n \geq 30$ la distribuzione di $\bar{X}$ può essere approssimata, per il teorema del limite centrale, con la distribuzione normale:

$$\bar{X} \sim N(\mu; Var(\bar{X}) = 0.25)$$

e che quindi

$$\frac{\bar{X} - \mu}{\sqrt{0.25}} = \frac{\bar{X} - \mu}{0.5} = Z \sim N(0; 1),$$

possiamo scrivere che

$$\begin{aligned}
P[|\bar{X} - \mu| < 1] &= P[-1 < \bar{X} - \mu < 1] \\
&= P\left[-\frac{1}{0.5} < \frac{\bar{X} - \mu}{0.5} < \frac{1}{0.5}\right] \\
&= P\left[-\frac{1}{0.5} < Z < \frac{1}{0.5}\right] \\
&= P\left[-2 < Z < 2\right] \\
&= \Phi(2) - \Phi(-2) \\
&= \Phi(2) - [1 - \Phi(2)] \\
&= 2 \cdot \Phi(2) - 1 \\
&= 2 \cdot 0.97725 - 1 \\
&= 0.9545.
\end{aligned}$$

La probabilità di stimare μ con un errore inferiore a 1 cl è pertanto pari a 0.9545.

c) Lo scarto quadratico medio dello stimatore $\bar{X}$ è dato da:

$$\sqrt{Var(\bar{X})} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}.$$

Nel caso in esame, si ha che

$$\sigma = 5$$

e quindi imporre che lo scarto quadratico medio di $\bar{X}$ sia inferiore a 0.25, significa imporre che

$$\frac{5}{\sqrt{n}} < 0.25.$$

Risolvendo rispetto a n, si ricava che

$$n > 400$$

e quindi si può concludere che si devono ancora analizzare 400-100=300 contenitori affinchè lo scarto quadratico medio di $\bar{X}$ sia inferiore a 0.25.

d) Vogliamo ora determinare l'ampiezza campionaria n tale che $P[|\bar{X} - \mu| < 1] = 0.99$.

Per fare ciò, ricordiamo che

$$\bar{X} \sim N\left(\mu; \frac{\sigma^2}{n}\right)$$

e che quindi

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{\bar{X} - \mu}{5/\sqrt{n}} = Z \sim N(0; 1).$$

Dobbiamo quindi individuare un particolare valore n tale che:

$$\begin{aligned}
&P[|\bar{X} - \mu| < 1] = 0.99\\
&P[-1 < \bar{X} - \mu < 1] = 0.99\\
&P\left[-\frac{1}{\sigma/\sqrt{n}} < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < \frac{1}{\sigma/\sqrt{n}}\right] = 0.99\\
&P\left[-\frac{1}{5/\sqrt{n}} < \frac{\bar{X} - \mu}{5/\sqrt{n}} < \frac{1}{5/\sqrt{n}}\right] = 0.99\\
&P\left[-\frac{1}{5/\sqrt{n}} < Z < \frac{1}{5/\sqrt{n}}\right] = 0.99\\
&P\left[-\frac{\sqrt{n}}{5} < Z < \frac{\sqrt{n}}{5}\right] = 0.99\\
&\Phi\left(\frac{\sqrt{n}}{5}\right) - \Phi\left(-\frac{\sqrt{n}}{5}\right) = 0.99\\
&\Phi\left(\frac{\sqrt{n}}{5}\right) - \left[1 - \Phi\left(\frac{\sqrt{n}}{5}\right)\right] = 0.99\\
&2 \cdot \Phi\left(\frac{\sqrt{n}}{5}\right) - 1 = 0.99\\
&\Phi\left(\frac{\sqrt{n}}{5}\right) = 0.995.
\end{aligned}$$

A questo punto, utilizzando le tavole della normale, ricaviamo che

$$\Phi(2.575) = 0.995$$

e quindi, imponendo che

$$\frac{\sqrt{n}}{5} = 2.575$$

possiamo ricavare n e ottenere

$$n = 165.76 \qquad \text{(arrotondando } n = 166\text{)}.$$

Possiamo perciò affermare che occorre ancora analizzare 166-100=66 contenitori affinchè $P[|\bar{X} - \mu| < 1] = 0.99$.

e) Per stimare la varianza σ^2 con uno stimatore non distorto, supponendo non nota l'informazione relativa allo scarto quadratico medio, è necessario utilizzare lo stimatore "varianza campionaria corretta" la cui espressione è data da:

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

Nel caso in esame, si ha (indicando con x_j i valori centrali delle classi):

$$\begin{aligned} s_c^2 &= \frac{1}{n-1} \cdot \sum_{j=1}^{k} (x_j - \bar{x})^2 \cdot n_j \\ &= \frac{1}{100-1} \cdot \sum_{j=1}^{3} (x_j - 200)^2 \cdot n_j \\ &= \frac{1}{(100-1)} \cdot [(192.5 - 200)^2 \cdot 15 + (200 - 200)^2 \cdot 70 + (207.5 - 200)^2 \cdot 15] \\ &= 17.045. \end{aligned}$$

6. Da un lotto di arance se ne estraggono 400, e di queste 180 risultano avere un peso superiore a 150 grammi.
a) Stimare la percentuale di arance nel lotto con peso superiore a 150 grammi;
b) indicare quale dovrebbe essere la numerosità campionaria minima affinchè lo scarto quadratico medio dello stimatore utilizzato sia inferiore a 0.02 sia tenendo conto dell'informazione campionaria che non tenendone conto;
c) si calcoli la probabilità che lo stimatore differisca dall'ignota proporzione p di arance con peso superiore a 150 grammi, per meno di 0.015.

Svolgimento

a) Per stimare la percentuale di arance nel lotto con peso superiore a 150 grammi, introduciamo la variabile casuale X="numero di arance estratte con peso maggiore di 150 grammi".
É noto che

$$X \sim Bin(n, p)$$

dove si è indicato con n il numero di arance estratte e con p la probabilità di estrarre un'arancia con peso maggiore di 150 grammi.
É altresì noto che:

$$E(X) = np \qquad Var(X) = npq$$

dove si è indicato con q la quantità $1 - p$.

A partire dalla variabile casuale X, possiamo ricavare lo stimatore

$$\hat{P} = \frac{X}{n}$$

che quindi indicherà la proporzione di arance con peso maggiore di 150 grammi tra quelle estratte. É facile ricavare

$$E(\hat{P}) = E\left[\frac{X}{n}\right] = \frac{1}{n} \cdot E(X) = \frac{np}{n} = p$$

$$Var(\hat{P}) = Var\left[\frac{X}{n}\right] = \frac{1}{n^2} \cdot Var(X) = \frac{npq}{n^2} = \frac{pq}{n}.$$

Calcoliamo il valore di $\hat{P}$ nel caso in esame:

$$\hat{p} = \frac{180}{400} = 0.45.$$

La stima della percentuale di arance con peso maggiore di 150 grammi è data da 0.45.

b) Per determinare la numerosità campionaria minima affinchè lo scarto quadratico medio dello stimatore utilizzato sia inferiore a 0.02, è necessario imporre che

$$\sigma(\hat{P}) = \sqrt{Var(\hat{P})} < 0.02$$

cioè, per quello che si è visto in precedenza:

$$\sqrt{\frac{pq}{n}} < 0.02$$

e ricavare n:

$$\frac{pq}{n} < (0.02)^2$$

$$n > \frac{pq}{0.0004}.$$

Caso I: consideriamo l'informazione campionaria.
Sfruttando l'informazione campionaria, si ha che $p = \hat{p} = 0.45$ e di conseguenza che $q = 1 - \hat{p} = 0.55$. Sostituendo tali valori si ottiene

$$n > \frac{0.45 \cdot 0.55}{0.0004}$$

vale a dire

$$n > 618.75.$$

Considerando l'informazione campionaria, la numerosità minima richiesta è pari a 619 arance.

Caso II: non consideriamo l'informazione campionaria.
In questo secondo caso, il procedimento è analogo: l'unica differenza sta nel fatto che p e q non possono essere stimati dal campione. Per trovare la numerosità minima richiesta è quind necessario considerare il caso più sfavorevole che corrisponde a scegliere $p = q = 0.5$. Sostituendo tali valori nella relazione

$$n > \frac{pq}{0.0004}$$

abbiamo quindi:

$$n > \frac{0.5 \cdot 0.5}{0.0004}$$

da cui si ottiene

$$n > 625.$$

Non considerando l'informazione campionaria, la numerosità minima richiesta è pari a 625 arance.

c) La probabilità richiesta è data da:

$$P[|\hat{P} - p| < 0.015].$$

Ricordando che per $n \geq 30$ la distribuzione di $\hat{P}$ può essere approssimata, per il teorema del limite centrale, con la distribuzione normale:

$$\hat{P} \sim N\left(p\,;\frac{pq}{n}\right)$$

e che quindi

$$\frac{\hat{P} - p}{\sqrt{pq/n}} = Z \sim N(0;1),$$

possiamo scrivere che

$$\begin{aligned} P[|\hat{P} - p| < 0.015] &= P[-0.015 < \hat{P} - p < 0.015] \\ &= P\left[-\frac{0.015}{\sqrt{pq/n}} < \frac{\hat{P} - p}{\sqrt{pq/n}} < \frac{0.015}{\sqrt{pq/n}}\right] \\ &= P\left[-\frac{0.015}{\sqrt{pq/n}} < Z < \frac{0.015}{\sqrt{pq/n}}\right] \\ &= P\left[-\frac{0.015 \cdot \sqrt{n}}{\sqrt{pq}} < Z < \frac{0.015 \cdot \sqrt{n}}{\sqrt{pq}}\right]. \end{aligned}$$

Ora, considerando l'informazione campionaria, poniamo

$$p = \hat{p} = 0.45 \qquad \text{e} \qquad q = 1 - \hat{p} = 0.55$$

ed otteniamo (ricordando che $n = 400$):

$$\begin{aligned}
P[|\hat{P} - p| < 0.015] &= P\left[-\frac{0.015 \cdot \sqrt{n}}{\sqrt{pq}} < Z < \frac{0.015 \cdot \sqrt{n}}{\sqrt{pq}}\right] \\
&= P\left[-\frac{0.015 \cdot \sqrt{400}}{\sqrt{0.45 \cdot 0.55}} < Z < \frac{0.015 \cdot \sqrt{400}}{\sqrt{0.45 \cdot 0.55}}\right] \\
&= P\left[-\frac{0.015 \cdot 20}{0.497} < Z < \frac{0.015 \cdot 20}{0.497}\right] \\
&= P\left[-0.60 < Z < 0.60\right] \\
&= \Phi(0.60) - \Phi(-0.60) \\
&= \Phi(0.60) - [1 - \Phi(0.60)] \\
&= 2 \cdot \Phi(0.60) - 1 \\
&= 2 \cdot 0.72575 - 1 \\
&= 0.4515.
\end{aligned}$$

La probabilità di stimare p con un errore inferiore a 0.015, considerando l'informazione campionaria, è pertanto pari a 0.4515.

7. Al fine di stimare la spesa media mensile μ sostenuta dalle famiglie italiane per alimenti e bevande, si estrae un campione casuale di 200 famiglie. Per ciascuna di esse si rileva la spesa X (in migliaia di euro) sostenuta nell'ultimo mese per i beni sopra menzionati, ottenendo le seguenti informazioni:

$$\sum_{i=1}^{200} x_i = 167.5 \qquad \sum_{i=1}^{200} x_i^2 = 350.85.$$

a) Stimare la spesa media mensile μ.
b) Stimare la varianza σ^2 della spesa mensile utilizzando uno stimatore non distorto.

Svolgimento

a) Per stimare la spesa media mensile, si utilizza lo stimatore "media campionaria":

$$\bar{X} = \frac{1}{n} \cdot \sum_{i=1}^{n} X_i.$$

Nel caso in esame, tale stimatore assume il valore:

$$\bar{x} = \frac{1}{200} \cdot \sum_{i=1}^{200} x_i = \frac{167.5}{200} = 0.8375 \qquad \text{(migliaia di euro)}$$

b) per stimare la varianza, si utilizza lo stimatore non distorto "varianza campionaria corretta":

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

la quale è anche uguale a:

$$\begin{aligned} S_c^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(X_i^2 + \bar{X}^2 - 2X_i\bar{X})^2 \\ &= \frac{1}{n-1}\left[\sum_{i=1}^{n} X_i^2 + n \cdot \bar{X}^2 - 2\bar{X}^2 \cdot \sum_{i=1}^{n} X_i\right] \\ &= \frac{1}{n-1}\left[\sum_{i=1}^{n} X_i^2 - n \cdot \bar{X}^2\right]. \end{aligned}$$

Nel caso in esame, pertanto:

$$\begin{aligned} s_c^2 &= \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - n \cdot \bar{x}^2\right] \\ &= \frac{1}{200-1} \cdot \left[350.85 - 200 \cdot (0.8375)^2\right] \\ &= \frac{1}{199} \cdot [350.85 - 140.28125] \\ &= 1.0581. \end{aligned}$$

La stima della varianza della spesa mensile è pari a 1.0581.