CORSO DI STATISTICA (parte 2) - ESERCITAZIONE 5

Dott.ssa Antonella Costanzo

a.costanzo@unicas.it

Esercizio 1. Approssimazione normale della Poisson (TLC)

In un determinato tratto di strada il numero di incidenti si distribuisce secondo un modello di Poisson con 36 incidenti in media all'anno. Si determini la probabilità che si verifichino al più 40 incidenti in un dato anno.

Sol.

Sia X il numero di incidenti in un anno. Si richiede di calcolare $P(X \leq 40)$ che è pari a:

$$P(X \leq 40) = \sum_{i=0}^{40} P(X = x_i) = \sum_{i=0}^{40} e^{-\lambda} \frac{\lambda^i}{i!}$$

La distribuzione di Poisson per λ elevato può tuttavia essere approssimato ad una variabile casuale normale con media λ e varianza λ (risultato del TLC) Quindi, approssimativamente, se Z è distribuita secondo una normale standard si ha:

$$P(X \leq 40) = P\left(Z \leq \frac{40 - \lambda}{\sqrt{\lambda}}\right) = P\left(Z \leq \frac{40 - 36}{6}\right) = P(Z \leq 0.66) = \varphi(0.66) = 0.7475$$

Esercizio 2. Verifica delle proprietà degli stimatori: efficienza e non distorsione

Si estragga da una popolazione normale di media μ e varianza 1 un campione casuale (X_1, X_2) di due unità. Volendo stimare il parametro μ si considerino i seguenti stimatori:

$$T_1 = \frac{2}{3}X_1 + \frac{1}{3}X_2$$

$$T_2 = \frac{1}{2}X_1 + \frac{1}{2}X_2$$

Verificare che tali stimatori siano corretti e determinare quello più efficiente.

Sol.

Uno stimatore T_n si dice non distorto per θ se $E(T_n) = \theta$ ossia quando si escludono deviazioni sistematiche nella stima di θ per cui la media dello stimatore coincide col parametro.

Le medie dei due stimatori sono:

$$E(T_1) = E\left(\frac{2}{3}X_1 + \frac{1}{3}X_2\right) = \frac{2}{3}\mu + \frac{1}{3}\mu = \mu$$

$$E(T_2) = E\left(\frac{1}{2}X_1 + \frac{1}{2}X_2\right) = \frac{1}{2}\mu + \frac{1}{2}\mu = \mu$$

I due stimatori sono entrambi corretti. La proprietà di correttezza (o non distorsione) ci dice che, in media, il nostro stimatore centra il bersaglio: cioè è uguale al vero parametro nella popolazione.
In generale non ci interessa soltanto centrare il bersaglio, vogliamo anche evitare di andarci troppo lontano quando sbagliamo. E quindi utile guardare alla varianza dello stimatore e in particolare alla proprietà dell'efficienza:

L'efficienza riguarda la precisione di uno stimatore: più è efficiente, più le stime ottenute saranno - in media - vicine al parametro. Tiene conto sia della distorsione che della variabilità di uno stimatore

La varianza dei due stimatori è:

$$Var(T_1) = Var\left(\frac{2}{3}X_1 + \frac{1}{3}X_2\right) = \left(\frac{2}{3}\right)^2 Var(X_1) + \left(\frac{1}{3}\right)^2 Var(X_2) = \frac{4}{9} + \frac{1}{9} = \frac{5}{9}$$

$$Var(T_2) = Var\left(\frac{1}{2}X_1 + \frac{1}{2}X_2\right) = \left(\frac{1}{2}\right)^2 Var(X_1) + \left(\frac{1}{2}\right)^2 Var(X_2) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

T_2 è lo stimatore più efficiente (criterio dell'efficienza relativa). Nota: dato che i due stimatori sono corretti, l'errore quadratico medio coincide con la varianza dello stimatore.

Esercizio 3. Stima puntuale per la media della popolazione

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

Quantità (cl)	[190,195)	[195,205)	[205,210)
Valori centrali	192.5	200	207.5
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 della quantità di liquido immessa dall'impianto.

Sol.

a)Per stimare l'ignota quantità media μ, utilizziamo lo stimatore media campionaria $\bar{X}$. Per il TLC sappiamo che:

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

Valore atteso della media campionaria:

$$E(\bar{X}) = \frac{192.5 \times 15 + 200 \times 70 + 207.5 \times 15}{100} = 200$$

La varianza dello stimatore media campionaria:

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

b) Sapendo che

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{n} = 0.25)$$

e che $Z = \frac{\bar{X}-\mu}{\sigma} \sim N(0,1)$

La probabilità richiesta è data da:

$$P(|\bar{X} - \mu| < 1) = P(-1 < \bar{X} - \mu < 1) =$$

$$= P\left(-\frac{1}{0.5} < Z < \frac{1}{0.5}\right) = P(-2 < Z < 2) = P(Z < 2) - P(Z < -2)$$
$$= P(Z < 2) - \big(1 - P(Z < 2)\big) = 0.9772 - (1 - 0.9772) = 0.9545$$

c) Lo scarto quadratico medio dello stimatore media campionaria è il seguente:

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

Nel nostro caso: $\sigma = 5$

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

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

E risolvendo rispetto ad n si ricava che n>400. 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) dobbiamo determinare l'ampiezza campionaria tale che $P(|\bar{X} - \mu| < 1) = 0.99$. Anche qui ricordiamo che:

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

Dobbiamo quindi individuare quel particolare valore n tale che:

$$P(|\bar{X} - \mu| < 1) = 0.99$$

$$P(-1 < \bar{X} - \mu < 1) = 0.99$$

$$P\left(-\frac{1}{\sigma/\sqrt{n}} < Z < \frac{1}{\sigma/\sqrt{n}}\right) = 0.99$$

$$P\left(-\frac{1}{5/\sqrt{n}} < Z < \frac{1}{5/\sqrt{n}}\right) = 0.99$$

$$P\left(-\frac{\sqrt{n}}{5} < Z < \frac{\sqrt{n}}{5}\right) = 0.99$$

$$P\left(Z < \frac{\sqrt{n}}{5}\right) - P\left(Z < -\frac{\sqrt{n}}{5}\right) = 0.99$$

$$P\left(Z < \frac{\sqrt{n}}{5}\right) - \left(1 - P\left(Z < \frac{\sqrt{n}}{5}\right)\right) = 0.99$$

Occorre individuare il percentile sulla tavola della funzione di ripartizione Normale che lascia a sinistra circa il 99%.

$z_{0.99} = 2.33$

Per cui imponendo che $\frac{\sqrt{n}}{5} = 2.33$ otteniamo che n=135.72 (circa 133). Pertanto occorre analizzare 133-100=33 contenitori.

e) per stimare la varianza campionaria in modo non distorto:

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

Nota $E(\bar{X}) = 200$, n=100

$$S^2 = \frac{1}{100-1}(192.5 - 200)^2 15 + (200 - 200)^2 70 + (207.5 - 200)^2 15 = 17.04$$