CORSO DI STATISTICA (parte 2) - ESERCITAZIONE 6

Dott.ssa Antonella Costanzo

a.costanzo@unicas.it

Esercizio 1. Stima puntuale per la proporzione

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 affinché lo scarto quadratico medio dello stimatore utilizzato sia inferiore a 0.02 tenendo conto dell'informazione campionaria disponibile

Sol.

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. Inoltre sappiamo che:
$E(X) = np \qquad Var(X) = np(1-p)$

A partire della variabile casuale X, possiamo ricavare lo stimatore:

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

che indica la proporzione di arance con peso maggiore di 150 grammi tra quelle estratte. Quindi:

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

$$Var(\hat{P}) = Var\left(\frac{X}{n}\right) = \frac{1}{n^2}Var(X) = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n}$$

Il valore atteso di $\hat{P}$ è:

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

B)

Per determinare la numerosità campionaria minima affinchè lo scarto quadratico medio dello stimatore utilizzato sia inferiore a 0.02 occorre definire:

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

Ossia:

$$\sqrt{\frac{p(1-p)}{n}} < 0.02$$

da cui risolvendo rispetto a n otteniamo:

$$\frac{p(1-p)}{n} < (0.02)^2$$

$$\frac{0.45 \times 0.55}{n} > (0.02)^2$$

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

Quindi $n > 618.75$ la numerosità minima richiesta è pari a 619 arance

Esercizio 2. Stima intervallare: IC per la media incognita (varianza nota)

Una partita di bulloni presenta un diametro medio μ incognito; la varianza del diametro dei bulloni è invece nota e pari a 0.01 cm. Si estrae un campione di n = 1000 bulloni, sui quali si osserva un diametro medio pari a 1.2 cm.

a) Si determini l'intervallo di confidenza per μ avendo fissato un livello di confidenza del 99%.

b) Si determini l'ampiezza di tale intervallo.

Sol.

a)Per determinare l'intervallo di confidenza (I.C.) per μ a livello di confidenza pari al 99%, bisogna innanzitutto ricavare il quantile di ordine $1-\frac{\alpha}{2}$ della distribuzione normale standard.

Quindi, poiché:

$$1-\alpha=0.99$$

si ha che:

$$\alpha=1-0.99=0.01$$

perciò:

$$\frac{\alpha}{2}=\frac{0.01}{2}=0.005$$

e di conseguenza:

$$1-\frac{\alpha}{2}=1-0.005=0.995$$

Consultando le tavole della distribuzione normale standard si ricava che $z_{0.995}=2.576$

Indicando con α l'area nelle code e con 1-α il livello di confidenza desiderato, allora:

Intervallo casuale:

$$P\left(\bar{X}-z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\leq\mu\leq\bar{X}+z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right)=1-\alpha$$

Intervallo di confidenza:

$$\left[\bar{X}-z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}};\bar{X}+z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right]$$

dove:

- $\bar{X}$ media campionaria
- $z_{\frac{\alpha}{2}}$ è il percentile di ordine $1-\frac{\alpha}{2}$ della distribuzione normale standard
- σ è lo scarto quadratico medio della popolazione di riferimento
- n ampiezza campionaria

Sostituendo quindi i valori avremo l'IC per la media incognita della popolazione a livello di confidenza di 0.99:

$$\left[1.2 - 2.576\frac{\sqrt{0.01}}{\sqrt{1000}}; 1.2 + 2.576\frac{\sqrt{0.01}}{\sqrt{1000}}\right]$$

$$[1.1918; 1.2081]$$

b) L'ampiezza di tale intervallo è data da:

1.2081-1.1918=0.01629 (cm)

In generale comunque l'ampiezza dell'intervallo di confidenza a livello 1-α per la media μ nel caso di varianza nota è pari a:

$$a = 2 * z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$$

Infatti nel caso in esame

$$a = 2 * 2.576\frac{\sqrt{0.01}}{\sqrt{1000}} = 0.01629$$

Esercizio 3. IC per la media incognita (varianza non nota)

A partire da un campione casuale di pezzi prodotti da una macchina si sono rilevate le seguenti lunghezze in mm:

201,200,205, 209, 212, 214, 208, 210, 207

E' noto che la lunghezza dei pezzi in questione si distribuisce normalmente.

a) Calcolare l' intervallo di confidenza al livello del 98% per la media
b) Supponendo che la varianza sia nota e pari a 20 (mm^2) si calcoli la numerosità campionaria necessaria affinché l'intervallo di confidenza per la media al livello del 95% non sia più lungo di 1.5 mm

Sol.

a)

$(X_1, X_2, \dots, X_n)$ è un campione casuale estratto da una v.c. $X \sim N(\mu, \sigma^2)$con varianza σ^2 incognita, inoltre la media campionaria è

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

e la varianza campionaria corretta è

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

Da ciò segue che:

$$\frac{\sqrt{n}(\bar{X} - \mu)}{\sqrt{S^2}} \sim t_{n-1}$$

$t_{n-1,\alpha}$ indica il quantile α della distribuzione t con n-1 g.d.l.

Intervallo casuale

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

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

Intervallo di confidenza:

sapendo che:

- $\bar{X} = 207.33$
- $\sqrt{S^2} = 4.69$
- $t_{n-1,{}^{\alpha}/_{2}} = 2.896$ con α=0.02, n=9

$$\left[207.33 - 2.896\frac{4.69}{\sqrt{9}}; 207.33 + 2.896\frac{4.69}{\sqrt{9}}\right]$$

IC al 98% per la media incognita è quindi pari a:

$$[202.81; 211.86]$$

b)

Supponendo che la varianza sia nota allora risulta:

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

Da cui l'intervallo di confidenza risulta pari a

$$\left[\bar{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} ; \bar{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right]$$

La lunghezza (ampiezza) dell'intervallo di confidenza è data da:

$$a = 2 * z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

Occorre determinare la numerosità campionaria affinchè l'intervallo di confidenza per la media al 95% non sia più lungo di un valore c=1.5 mm. In particolare, si richiede che: $a \leq c$ quindi

$$\sqrt{n} \geq 2 * z_{\frac{\alpha}{2}} \frac{\sigma}{c}$$

Ossia n deve essere almeno pari a $4z_{\alpha/2}^2 \frac{\sigma^2}{c^2}$

Sostituendo nel nostro caso $z_{\frac{\alpha}{2}} = 1.96$, $\sigma^2 = 20$, c=1.5 si ottiene

$$n > 136.6 \approx 137$$