UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

Dipartimento di Comunicazione
ed Economia

Stima Intervallare

Perchà la stima puntale non basta?

Con la *stima puntuale* utilizziamo le osservazioni di un campione per stimare il parametro di interesse della popolazione mediante un singolo valore numerico!

Ciò comporta alcuni punti di debolezza:

- Il valore ottenuto può essere diverso dal "vero" valore del parametro di interesse
- Se il carattere di interesse è continuo, la probabilità di ottenere un valore della stima esattamente uguale al parametro incognito è pari a 0 $\Rightarrow$ ottenere una stima uguale al parametro è quasi impossibile!

Come fare?

Perchà la stima puntale non basta?

In molti casi si preferisce considerare (oltre alla stima puntuale) un *intervallo* di valori possibili per la stima al quale è associato un *livello di affidabilità*.

In altre parole, l'obiettivo è quello di ottenere un intervallo di valori attorno alla stima puntuale.

Inoltre, ci aspettiamo che questo intervallo contenga il vero valore del parametro incognito, con un certo *livello di fiducia*.

Intervallo di confidenza

Si definisce *intervallo di confidenza* di livello $1-\alpha$ per il parametro incognito θ un intervallo che contiene il vero valore di θ con probabilità $1-\alpha$

Resta il problema di come calcolare questo intervallo!

Innanzitutto dobbiamo definire il livello di confidenza:

Generalmente, i valori più utilizzati nella pratica per α sono 0.10, 0.05 e 0.01, cui corrispondono livelli di confidenza $(1-\alpha)$ pari a 0.90, 0.95 e 0.99, rispettivamente.

NOTA: verrà specificato in ogni esercizio il livello di confidenza da usare.

Intervallo di confidenza per la media

Logica di costruzione

Supponiamo di voler costruire un intervallo di confidenza attorno alla media incognita della popolazione.

Sappiamo che la media campionaria rappresenta una buona stima (puntuale) della vera media della popolazione.

Conosciamo, inoltre, la varianza (deviazione standard) della media campionaria, che ci indica quanto “affidabile” è la stima che abbiamo ottenuto.

Idea: Possiamo combinare la stima (media campionaria) e la rispettiva deviazione standard per ottenere un intervallo che contenga il vero valore del parametro incognito (con un certo livello di confidenza)!

Intervallo di confidenza per la media

Logica di costruzione

Ritorniamo all'esempio delle lunghezze dei pezzi prodotti da un macchinario (Borra, Di Ciaccio, p.277).

Supponiamo di sapere che la varianza della popolazione è pari a $\sigma^2 = 0.1$

Sappiamo che la media campionaria si distribuisce secondo una distribuzione Normale con media μ (incognita) e varianza $\frac{\sigma^2}{n} = \frac{0.1}{10} = 0.01$.

Intervallo di confidenza per la media

Logica di costruzione

Dunque $\bar{X} \sim N(\mu, 0.01)$

Se standardizziamo

$$Z = \frac{\bar{X} - \mu}{\sqrt{0.01}} = \frac{\bar{X} - \mu}{0.1}$$

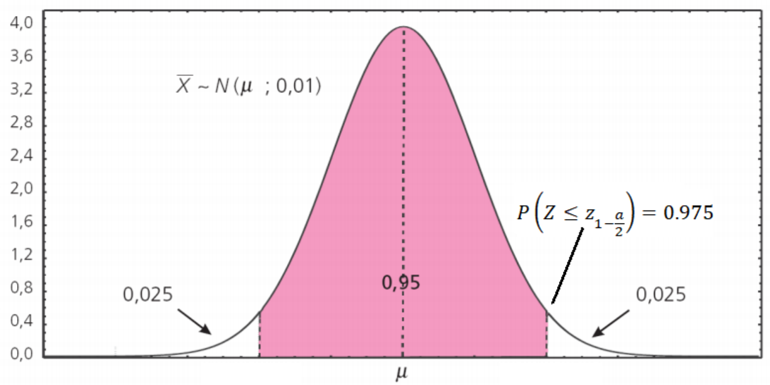
si distribuisce secondo una N(0,1).

Fissiamo un livello di confidenza pari a 0.95 ($\alpha = 0.05$).
L'intervallo cercato sarà

$$P(-z \leq Z \leq z) = P(-z \leq \frac{\bar{X} - \mu}{0.1} \leq z) = 0.95$$

Intervallo di confidenza per la media

Logica di costruzione



Il valore di $z_{1-\frac{\alpha}{2}}$ cercato è 1.96

Intervallo di confidenza per la media

Logica di costruzione

Dunque

$$P\left(-z \leq \frac{\bar{X} - \mu}{0.1} \leq z\right) = 0.95$$

equivale a:

$$P\left(-1.96 \leq \frac{\bar{X} - \mu}{0.1} \leq 1.96\right) = 0.95$$

e quindi

$$P(-1.96 \times 0.1 \leq \bar{X} - \mu \leq 1.96 \times 0.1) = 0.95$$

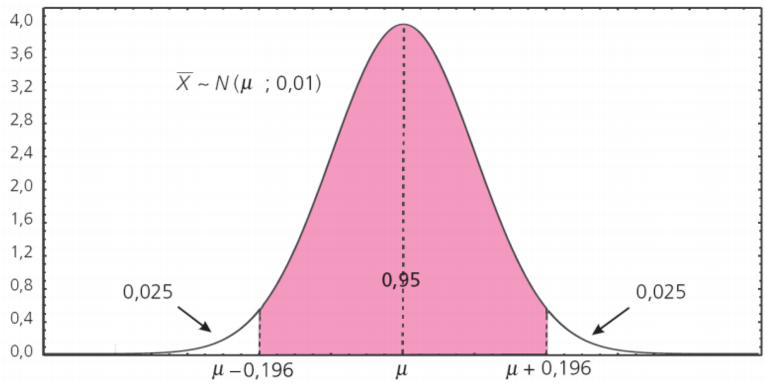
cioè

$$P(\bar{X} - 1.96 \times 0.1 \leq \mu \leq \bar{X} + 1.96 \times 0.1) = 0.95$$

$$P(\bar{X} - 0.196 \leq \mu \leq \bar{X} + 0.196) = 0.95$$

Intervallo di confidenza per la media

Logica di costruzione



(Borra, Di Ciaccio, p.288).

Intervallo di confidenza per la media

Varianza della popolazione nota

In generale, se la varianza della popolazione è nota, l'intervallo di confidenza per la media della popolazione al livello di confidenza $1-\alpha$ è dato da:

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

da cui:

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

Possiamo indicare l'intervallo di confidenza semplicemente con

$$\left[\bar{X} - z_{1-\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}}; \bar{X} + z_{1-\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}}\right]$$

Se vogliamo cambiare il livello di confidenza dell'intervallo, basta sostituire il valore di $z_{1-\frac{\alpha}{2}}$ con il valore desiderato.

Di seguito, i valori più comunemente utilizzati:

$1-\alpha$	0.90	0.95	0.99
$z_{1-\frac{\alpha}{2}}$	1.645	1.96	2.575

Esempio

Borra, Di Ciaccio, p.277

Su un ccs di 10 pezzi prodotti da un certo macchinario vengono rilevate le lunghezze, di seguito riportate:

$$x_1 = 1.18 \quad x_2 = 1.56 \quad x_3 = 1.86 \quad x_4 = 1.27 \quad x_5 = 2.03$$

$$x_6 = 1.39 \quad x_7 = 1.92 \quad x_8 = 1.26 \quad x_9 = 1.52 \quad x_{10} = 1.74$$

Supponendo che la varianza della popolazione sia nota e pari a $\sigma^2 = 0.5$, si costruisca un intervallo di confidenza al livello $1 - \alpha = 0.95$.

Esempio

Soluzione

Sappiamo che per il campione $\bar{X} = 1.57$ e la varianza è $\frac{\sigma^2}{n} = 0.05$. (da cui $\frac{\sigma}{\sqrt{n}} = 0.224$).

L'intervallo di confidenza cercato è:

$$\left[\bar{X} - 1.96 \cdot \frac{\sigma}{\sqrt{n}}; \bar{X} + 1.96 \cdot \frac{\sigma}{\sqrt{n}}\right]$$

Sostituendo i valori:

$$[1.57 - 1.96 \cdot 0.224; 1.57 + 1.96 \cdot 0.224] = [1.131; 2.009]$$

Intervallo di confidenza per la media

Varianza della popolazione non nota

E se la varianza della popolazione non è nota?

Il procedimento per ottenere l'intervallo di confidenza è del tutto analogo al caso di varianza della popolazione nota. Tuttavia è necessario *sostituire* σ^2 con una sua stima.

Si utilizza la *varianza campionaria corretta*

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

Intervallo di confidenza per la media

Varianza della popolazione non nota

Tuttavia, questa sostituzione non è priva di conseguenze...

Se nella standardizzazione sostituiamo σ con $S = \sqrt{S^2}$, otteniamo la v.c.

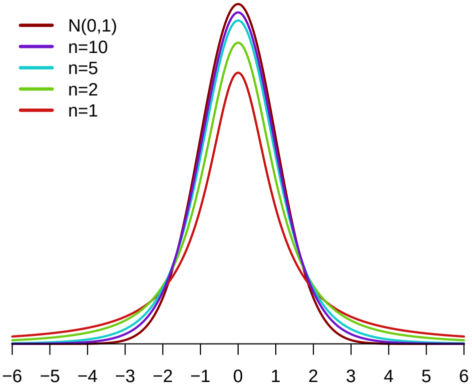
$$T = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}$$

che non sarà più distribuita come una N(0,1), bensì come una *t-Student con $n-1$ gradi di libertà*.

Questo implica che il valore di $z_{1-\frac{\alpha}{2}}$ del caso precedente dovrà essere sostituito da un nuovo valore $t_{\frac{\alpha}{2},n-1}$ da cercare sulle tavole della distribuzione t-Student.

t di Student

La t di Student viene introdotta da William Gossett (alias Student, 1876-1937), responsabile della birrifcazione per la Guinness, per confrontare mediante campioni la qualità di diverse produzioni di birra.



Ha una forma simile alla normale ma con code tanto più pesanti quanto più piccolo è n.

Tavola della t di Student

La tavola della t, diversamente da quella della normale, non riporta la FdR ma i quantili corrispondenti a particolari livelli di probabilità.

Le righe corrispondono ai gradi di libertà.

Le colonne alle probabilità.

	Probabilità					
n	0.10	0.050	0.025	0.010	0.005	0.001
1	3.08	6.31	12.71	31.82	63.66	318.30
2	1.89	2.92	4.30	6.96	9.93	22.33
3	1.64	2.35	3.18	4.54	5.84	10.21
4	1.53	2.13	2.78	3.75	4.60	7.17
5	1.48	2.02	2.57	3.37	4.03	5.89
6	1.44	1.94	2.45	3.14	3.71	5.21
7	1.42	1.90	2.37	3.00	3.50	4.79
8	1.40	1.86	2.31	2.90	3.35	4.50
9	1.38	1.83	2.26	2.82	3.25	4.30
10	1.37	1.81	2.23	2.76	3.17	4.14
11	1.36	1.80	2.20	2.72	3.11	4.03
12	1.36	1.78	2.18	2.68	3.06	3.93
13	1.35	1.77	2.16	2.65	3.01	3.85
14	1.34	1.76	2.14	2.62	2.98	3.79
15	1.34	1.75	2.13	2.60	2.95	3.73
16	1.34	1.75	2.12	2.58	2.92	3.69
17	1.33	1.74	2.11	2.57	2.90	3.65
18	1.33	1.73	2.10	2.55	2.88	3.61
19	1.33	1.73	2.09	2.54	2.86	3.58
20	1.32	1.73	2.09	2.53	2.85	3.55
21	1.32	1.72	2.08	2.52	2.83	3.53
22	1.32	1.72	2.07	2.51	2.82	3.50
23	1.32	1.71	2.07	2.50	2.81	3.48
24	1.32	1.71	2.06	2.49	2.80	3.47
25	1.32	1.71	2.06	2.48	2.79	3.45
26	1.31	1.71	2.06	2.48	2.78	3.44
27	1.31	1.70	2.05	2.47	2.77	3.42
28	1.31	1.70	2.05	2.47	2.76	3.41
29	1.31	1.70	2.04	2.46	2.76	3.40
30	1.31	1.70	2.04	2.46	2.75	3.38

Intervallo di confidenza per la media

Varianza della popolazione non nota

In generale, se la varianza della popolazione non è nota, l'intervallo di confidenza per la media della popolazione al livello di confidenza $1-\alpha$ è dato da:

$$P(-t_{\frac{\alpha}{2},n-1} \leq \frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}} \leq t_{\frac{\alpha}{2},n-1}) = 1-\alpha$$

da cui:

$$P(\bar{X} - t_{\frac{\alpha}{2},n-1} \times \frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{\frac{\alpha}{2},n-1} \times \frac{S}{\sqrt{n}}) = 1-\alpha$$

Possiamo indicare l'intervallo di confidenza semplicemente con

$$\left[\bar{X} - t_{\frac{\alpha}{2},n-1} \times \frac{S}{\sqrt{n}}; \bar{X} + t_{\frac{\alpha}{2},n-1} \times \frac{S}{\sqrt{n}}\right]$$

Esempio

Borra, Di Ciaccio, p.295

Sia X la quantità di calcio (in mg) contenuta in un litro di acqua minerale. Supponiamo di aver estratto un ccs di 15 unità osservando le seguenti quantità di calcio:

$$x_1 = 2.41 \quad x_2 = 2.75 \quad x_3 = 3.27 \quad x_4 = 2.11 \quad x_5 = 2.07$$

$$x_6 = 3.02 \quad x_7 = 2.45 \quad x_8 = 2.96 \quad x_9 = 2.15 \quad x_{10} = 2.54$$

$$x_{11} = 2.86 \quad x_{12} = 2.15 \quad x_{13} = 3.01 \quad x_{14} = 2.82 \quad x_{15} = 2.68$$

Si costruisca un intervallo di confidenza al livello $1 - \alpha = 0.99$.

Esempio

Soluzione

Innanzitutto, calcoliamo media e varianza campionaria ottenendo:

$$\bar{X} = 2.62 \text{ e } S^2 = 0.147$$

da cui $\frac{S}{\sqrt{n}} = \frac{0.3834}{\sqrt{15}} = 0.099.$

Osservando che $1 - \alpha = 0.99$ da cui $\alpha = 1 - 0.99 = 0.01$ dunque $\alpha/2 = 0.005$

Dalle tavole otteniamo che $t_{\frac{\alpha}{2},n-1} = t_{0.005,14} = 2.9768$ L'intervallo di confidenza cercato è:

$$\left[\bar{X} - 2.9768 \cdot \frac{S}{\sqrt{n}}; \bar{X} + 2.9768 \cdot \frac{S}{\sqrt{n}}\right]$$

Sostituendo i valori:

$$[2.62 - 2.9768 \cdot 0.099; 2.62 + 2.9768 \cdot 0.099] = [2.325; 2.915]$$

Intervallo di confidenza per la proporzione

Logica di costruzione

Supponiamo di essere interessati a costruire un intervallo di confidenza per la proporzione di una popolazione.

Il procedimento per ottenere l'intervallo di confidenza è del tutto analogo ai casi precedenti (ovviamente con i dovuti accorgimenti!)

Abbiamo visto che la proporzione puè essere stimata mediante la media campionaria

$$\bar{x} = \frac{x_1 + \cdots + x_n}{n}$$

e che

$$E(\bar{x}) = \pi \text{ e } Var(\bar{x}) = \frac{\pi(1-\pi)}{n}$$

Intervallo di confidenza per la proporzione

Logica di costruzione

Si può dimostrare che, al crescere di n, la distribuzione di
$\bar{x} \sim N(\pi, \frac{\pi(1-\pi)}{n})$

Se standardizziamo

$$Z = \frac{\bar{x} - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}}$$

si distribuisce secondo una N(0,1).

Intervallo di confidenza per la proporzione

Logica di costruzione

Analogamente a quanto visto in precedenza, l'intervallo di confidenza per proporzione al livello di confidenza $1 - \alpha$ è dato da:

$$P\left(-z_{1-\frac{\alpha}{2}} \leq \frac{\bar{x} - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}} \leq z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha$$

da cui:

$$P\left(\bar{x} - z_{1-\frac{\alpha}{2}} \times \sqrt{\frac{\pi(1-\pi)}{n}} \leq \pi \leq \bar{x} + z_{1-\frac{\alpha}{2}} \times \sqrt{\frac{\pi(1-\pi)}{n}}\right) = 1 - \alpha$$

Intervallo di confidenza per la proporzione

Logica di costruzione

Possiamo sostituire a $\sqrt{\frac{\pi(1-\pi)}{n}}$ una sua stima data da $\sqrt{\frac{\bar{x}(1-\bar{x})}{n}}$,

ottenendo così l'intervallo di confidenza

$$\left[\bar{x} - z_{1-\frac{\alpha}{2}} \times \sqrt{\frac{\bar{x}(1-\bar{x})}{n}} ; \bar{x} + z_{1-\frac{\alpha}{2}} \times \sqrt{\frac{\bar{x}(1-\bar{x})}{n}}\right]$$

Esempio

Per stimare la proporzione di individui che posseggono uno smartphone, si intervista un campione casuale di 1500 persone. Di queste, 890 posseggono uno smartphone.

a. Qual è la stima della proporzione di possessori di smartphone nella popolazione?
b. Si trovi un i.c. al 95% per la proporzione di individui che posseggono uno smartphone nella popolazione.

Esempio - Soluzione

a. Qual è la stima della proporzione di possessori di smartphone nella popolazione?

$$\bar{x} = \frac{890}{1500} = 0.593$$

Esempio - Soluzione

b. Si trovi un i.c. al 95% per la proporzione di individui che possegono uno smartphone nella popolazione.

L'intervallo di confidenza cercato è:

$$\left[\bar{x} - 1.96 \times \sqrt{\frac{\bar{x}(1-\bar{x})}{n}}; \bar{x} + 1.96 \times \sqrt{\frac{\bar{x}(1-\bar{x})}{n}}\right]$$

Sostituendo i valori:

$$[0.593 - 1.96 \times 0.012685; 0.593 + 1.96 \times 0.012685] = [0.5681; 0.6179]$$

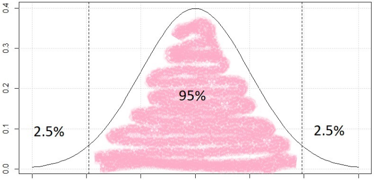
Ampiezza dell'intervallo
Come cambia

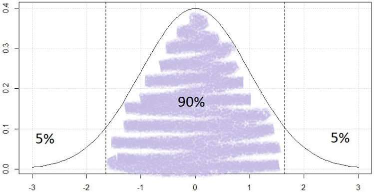
Fattori che incidono sull'ampiezza dell'i.c.:

- livello di confidenza
- numerosità campionaria
- varianza

Ampiezza dell'intervallo

Come cambia: livello di confidenza

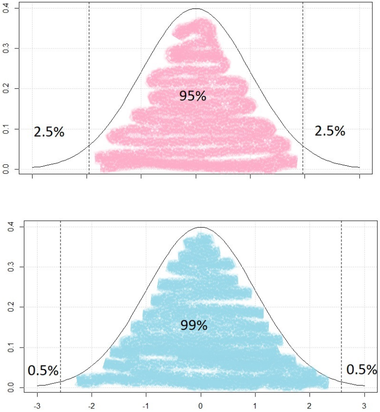




Se il livello di confidenza diminuisce: ampiezza $\downarrow$ – errore $\uparrow$

Ampiezza dell'intervallo

Come cambia: livello di confidenza



Se il livello di confidenza aumenta: ampiezza $\uparrow$ – errore $\downarrow$

Ampiezza dell'intervallo

Come cambia: numerosità campionaria

- Intervallo di confidenza per la media (Varianza nota):

$$\left[\bar{X} - z_{1-\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}}; \bar{X} + z_{1-\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}}\right]$$

- Intervallo di confidenza per la media (Varianza non nota):

$$\left[\bar{X} - t_{\frac{\alpha}{2},n-1} \times \frac{S}{\sqrt{n}}; \bar{X} + t_{\frac{\alpha}{2},n-1} \times \frac{S}{\sqrt{n}}\right]$$

- Intervallo di confidenza per la proporzione:

$$\left[\bar{x} - z_{1-\frac{\alpha}{2}} \times \sqrt{\frac{\bar{x}(1-\bar{x})}{n}}; \bar{x} + z_{1-\frac{\alpha}{2}} \times \sqrt{\frac{\bar{x}(1-\bar{x})}{n}}\right]$$

A parità di livello di confidenza se:

- dimensione campionaria $\downarrow$, ampiezza $\uparrow$
- dimensione campionaria $\uparrow$, ampiezza $\downarrow$

Ampiezza dell'intervallo

Come cambia: varianza

Quando sostituiamo σ^2 con la sua stima S^2, introduciamo incertezza!

A parità di livello di confidenza se:

- utilizziamo σ^2, ampiezza $\downarrow$
- utilizziamo S^2, ampiezza $\uparrow$