

Lezione 2

La statistica inferenziale: concetti di base

Argomenti della lezione:

- Campionamento
- Stima
- Distribuzione campionaria

Popolazione (o universo)

Insieme di tutti gli elementi cui si rivolge il ricercatore nell'indagine

Campione

Sottoinsieme degli elementi dell'insieme oggetto dell'indagine

- Campione rappresentativo
- Campione casuale

Campionamento casuale semplice

- Estrarre a caso da una popolazione
- Tutti gli elementi della popolazione hanno la stessa probabilità di essere estratti

Campionamento casuale stratificato

Suddivisione della popolazione in sub-popolazioni omogenee. Estrazione casuale dalle sub-popolazioni

Parametri e indicatori

Caratteristica studiata riferita:

alla popolazione ⇒ parametro

al campione ⇒ indicatore

Stima dei parametri

Conoscere parametri della popolazione tramite dati del campione

- **Stima puntuale o del valore**
- **Stima intervallare o dell'intervallo**

Stimatore o statistica

- **Metodo di calcolo impiegato per ottenere una stima**
- **La media campionaria $\Sigma X_i/n$ è uno stimatore della media μ della popolazione**
- **Funzione dei valori osservati**

Proprietà degli stimatori

Correttezza

$E(t) = \theta$

Consistenza

$\lim_{n \to \infty} E(t) = \theta$ $\lim_{n \to \infty} VAR(t) = 0$

Efficienza

Varianza campionaria minima

Distribuzione campionaria

- **Estrazione da una popolazione di N elementi, di tutti i possibili campioni di n < N elementi**
- **Calcolo dell'indicatore (per es. $\bar{x}$) in ogni campione**
- **Distribuzione di frequenza degli indicatori**

Distribuzione campionaria della media

Popolazione: 5 7 9

Parametri:

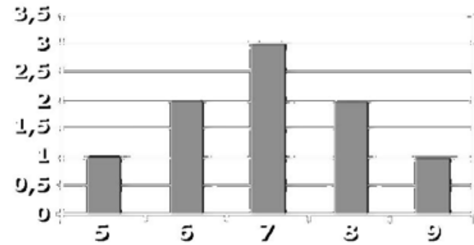
$\mu = 7$, $\sigma^2 = 2.66$, $\sigma = 1.63$
$N = 3$, $n= 2$

Campioni di ampiezza n = 2

Campioni	5-5	5-7	5-9	7-5	7-7	7-9	9-5	9-7	9-9
Medie	5	6	7	6	7	8	7	8	9

Distr. Campionaria della media

$\bar{x}$	5	6	7	8	9
Frequenza	1	2	3	2	1



$\mu_{\bar{x}} = 7 \Rightarrow \mu_{\bar{x}} = \mu$

$\sigma^2_{\bar{x}} = 1.33 \Rightarrow \sigma^2_{\bar{x}} = \frac{\sigma^2}{n}$; $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

Numerosità del campione (n)

Influenza la forma; la precisione della media; l'efficienza della stima

Al crescere di n (n>30):

Forma ⇒ Normale

Media ⇒ Media della popolazione

Errore standard ⇒ Minimo

Spesso la varianza della popolazione (σ^2) non è nota

La varianza della DCM può essere stimata dai dati del campione nel modo seguente:

$$\hat{\sigma}^2_{\bar{x}} = \frac{s^2}{n-1} \qquad \hat{\sigma}_{\bar{x}} = \frac{s}{\sqrt{(n-1)}}$$

Uso della distribuzione campionaria della media (DCM)

Campione: n, $\bar{x}$, s^2
Media μ della popolazione ?

$\mu_{\bar{x}} = \mu$

"quale sarà $\mu_{\bar{x}}$?"

- Intervallo di confidenza nel quale ricade μ con una probabilità elevata (es., p=.95)

$$P(\bar{X} - c \leq \mu \leq \bar{X} + c) = .95$$

- Utilizzare le proprietà della DCM per definire l'intervallo

Proprietà della DCM:

Forma normale (per n>30)

Media = media della popolazione

$$(\mu_{\bar{x}} = \mu)$$

Errore standard:

$$\sigma_{\bar{x}} = \sigma/\sqrt{n}$$

$$\hat{\sigma}_{\bar{x}} = s/\sqrt{(n-1)}$$

$$z_{\bar{x}} = (\bar{X} - \mu_{\bar{x}})/\sigma_{\bar{x}}$$

Poiché la DCM è normale:

$$P(-1.96 \leq z_{\bar{x}} \leq 1.96) = .95$$

ovvero

$$P(-1.96 \leq (\bar{X} - \mu_{\bar{x}})/\sigma_{\bar{x}} \leq 1.96) = .95$$

Tramite alcuni passaggi algebrici:

$$P(-1.96\sigma_{\bar{x}} \leq \bar{X} - \mu_{\bar{x}} \leq 1.96\sigma_{\bar{x}})=.95$$

$$P(-\bar{X} - 1.96\sigma_{\bar{x}} \leq -\mu_{\bar{x}} \leq -\bar{X} + 1.96\sigma_{\bar{x}})=.95$$

$$P(\bar{X} + 1.96\sigma_{\bar{x}} \geq \mu_{\bar{x}} \geq \bar{X} - 1.96\sigma_{\bar{x}})=.95$$

$$P(\bar{X} - 1.96\sigma_{\bar{x}} \leq \mu_{\bar{x}} \leq \bar{X} + 1.96\sigma_{\bar{x}})=.95$$

Formula generale per il calcolo dell'intervallo di confidenza

$$\bar{X} - z_{area}\,\sigma_{\bar{x}} \leq \mu_{\bar{x}} \leq \bar{X} + z_{area}\,\sigma_{\bar{x}}$$

Esempio

Campione:

$$\bar{x} = 19,\ s = 1.8,\ n = 50$$

p =.95

$$\bar{X} - 1.96\sigma_{\bar{x}} \leq \mu_{\bar{x}} \leq \bar{X} + 1.96\sigma_{\bar{x}}$$

Stima dell'errore standard tramite il campione

$$\hat{\sigma}_{\bar{x}} = s/\sqrt{(n-1)}$$

Intervallo di fiducia al 95%:

$$\bar{X} - 1.96\frac{s}{\sqrt{(n-1)}} \leq \mu_x \leq \bar{X} + 1.96\frac{s}{\sqrt{(n-1)}}$$

Limite inferiore:

$$19 - 1.96\left(\frac{1.8}{\sqrt{50-1)}}\right) = 18.49$$

Limite superiore:

$$19 + 1.96\left(\frac{1.8}{\sqrt{50-1)}}\right) = 19.51$$

$$P(18.49 \leq \mu_{\bar{x}} \leq 19.51 = .95$$

Intervallo di fiducia al 99%:

$$\bar{X} - 2.58\sigma_{\bar{x}} \leq \mu_{\bar{x}} \leq \bar{X} + 2.58\sigma_{\bar{x}}$$

$$P(18.33 \leq \mu_{\bar{x}} \leq 19.67) = .99$$

Campioni piccoli ($n \leq 30$)

Se $n \leq 30$, la DCM non si approssima alla normale

Al posto di z si utilizza la t di Student per n-1 gradi di libertà

$$\bar{X} - t_{area}\,\sigma_{\bar{x}} \leq \mu_{\bar{x}} \leq \bar{X} + t_{area}\,\sigma_{\bar{x}}$$

Campione:

$n=30$, $\bar{X} = 4.03$, $s=1.58$

Intervallo di fiducia al 95%

La t di Student corrispondente alla probabilità .05 con (n-1) = 29 gdl è: 2.045

$$\sigma_{\bar{x}} = 1.58/\sqrt{30-1} = .29$$

Lim. Inf = 4.03–2.045*.29 = 3.43

Lim. Sup = 4.03+2.045*.29 = 4.63

$$3.43 \leq \mu \leq 4.63$$

Intervallo di fiducia al 99%

La t di Student corrispondente alla probabilità .01 con (n-1) = 29 gdl è: 2.756

$\sigma_{\bar{x}} = 1.58/\sqrt{30-1} = .29$

Lim. Inf = 4.03–2.756*.29 = 3.22

Lim. Sup = 4.03+2. 756*.29 = 4.84

$3.22 \leq \mu \leq 4.84$

Gradi di libertà

Siano N osservazioni indipendenti, libere di assumere qualsiasi valore

$x_1\ x_2\ x_3\ \ldots\ x_i\ \ldots\ x_N$

i gradi di libertà (gdl) sono = N

Se si impongono vincoli i gdl sono N meno il numero di vincoli

Sia N = 3, ovvero $x_1\ x_2\ x_3$ e $(x_1+x_2+x_3)= 10$

due degli x possono assumere qualsiasi valore, il terzo è vincolato a riportare la somma a 10

I gdl sono 2

7 + 2 + (1) = 10
100 - 99 + (9) = 10
-50 + 50 + (10) = 10

gdl = numero di elementi liberi di variare - numero di vincoli

N - 1 = 3 - 1 = 2

GDL: numero di valori indipendenti che generano una distribuzione

CONCLUSIONE

- → Concetti fondamentali per la statistica inferenziale
- → Distribuzione Campionaria
- → Intervallo di confidenza
- → Gradi di libertà