

STATISTICA CAMPIONATORIA

2022-11-21

STATISTICA CAMPIONATORIA

Data una popolazione reale, possibilmente infinita, essa è analizzabile solo parzialmente, tramite estrazioni casuali di campioni

INDAGINI CAMPIONARIE

- Vincoli di tempo/costo
- Popolazione di interesse virtuale o reale infinita
- La rilevazione potrebbe distruggere le unità statistiche
- La precisione dei risultati ottenuti potrebbe non essere ottimale

CAMPIONE CASUALE

Sono i dati che vengono presi di riferimento in seguito a una estrazione casuale

CASUALE SEMPLICE

Può venire interpretato come una serie di n realizzazioni indipendenti di un dato esperimento

DATI OSSERVATI

$$x = (x_1, \dots, x_n), n \geq 1$$

Si riferiscono tutti alla stessa caratteristica di interesse

x_i rappresenta il valore osservato sulla i -esima unità statistica

IPOTESI FONDAMENTALE

I dati campionari x si possono interpretare come un vettore di variabili all'interno di un **vettore di variabili casuali**

$$X = (X_1, \dots, X_n)$$

X_i rappresenterebbe la variabile casuale associata a x_i , quindi x_i , un generico valore ottenuto dal campione, apparterebbe al supporto della variabile X_i

In sintesi ogni valore misurato fa parte di una variabile casuale che avrebbe potuto fornire valori diversi da quello rilevato, sempre però all'interno del suo supporto

La distribuzione di probabilità di X è ignota, o almeno in parte, ed è necessario stimarla

$$X = (X_1, \dots, X_n)$$

X rappresenta un vettore di variabili casuali, in cui ogni variabile è **indipendente e identicamente distribuita** rispetto alle altre, quindi possiedono lo stesso **modello probabilistico**

MODELLI STATISTICI PARAMETRICI

Dato un campione casuale semplice X_1, \dots, X_n , la distribuzione di probabilità delle singole variabili dipende dalla loro natura e dal fenomeno di interesse

TIPOLOGIA DI FENOMENO

BINARI

$$X_i \sim Ber(p) \quad p \in (0, 1)$$

MISURAZIONI

$$X_i \sim N(\mu, \sigma^2)$$

CONTEGGI

$$X_i \sim P(\lambda)$$

TEMPI DI FUNZIONAMENTO

$$X_i \sim Esp(\lambda)$$

PARAMETRI

Ogni modello è caratterizzato da uno o più parametri, i quali dipendono ovviamente dalla variabile di interesse

Nella statistica inferenziale parametrica si presuppone di conoscere il modello di appartenenza, ma non si conoscono i parametri, i quali andranno opportunamente stimati

MODELLO PARAMETRICO

- X_1, \dots, X_n sono indipendenti
- Le variabili possono anche non appartenere allo stesso modello
- i parametri da stimare sono detti $\theta = (\theta_1, \dots, \theta_d)$, $d \geq 1$

OBIETTIVO

Usare i dati rilevati $x = (x_1, \dots, x_n)$ per stimare opportunamente i valori di θ

$$\theta \in \Theta \subseteq R^d$$

Θ è chiamato spazio parametrico

SUPPORTO CONGIUNTO

Viene definito anche come **spazio campionario**

$$X_1, \dots, X_n$$

Come spiegato in precedenza esso corrisponde allo spazio occupato da tutti i possibili campioni x_1, \dots, x_n

SCELTA DEL MODELLO

- Capire la natura dei dati
- Aspetti/caratteristiche notevoli
- Informazioni sul meccanismo generatore dei dati della popolazione di interesse

ESEMPIO

Controllo di qualità: si analizzano n oggetti

Campione osservato $x = (x_1, \dots, x_n)$

Ogni valore $x_i \in X_i$ con $X_i \sim Ber(p)$ in quanto l'evento singolo è di tipo binario: valido(1) o non valido (0)

PARAMETRI

Una volta individuato il modello di appartenenza è necessario trovare θ , cioè il vettore dei parametri da stimare.

Nel caso in esame $\theta = p$, quindi la probabilità di successo dell'evento bernoulliano

SPAZIO PARAMETRICO

$$\theta \in \Theta = \{0, 1\}$$

SPAZIO CAMPIONARIO

$$S_X = \{0, 1\} \times \dots \times \{0, 1\} = \{0, 1\}^n$$

L'insieme di tutti i vettori di n dimensioni costituiti dai valori 0 e 1, esiti dell'esperimento bernoulliano

ESEMPIO 2

Misurazioni

Campione osservato $x = (x_1, \dots, x_n)$ è costituito da una serie di valori numerici associati alla variabile X di riferimento

$$X_i \sim N(\mu, \sigma^2)$$

PARAMETRI

Dopo aver determinato il modello è necessario stimare i parametri $\theta = (\theta_1, \theta_2) = (\mu, \sigma^2)$

SPAZIO PARAMETRICO

$$\theta \in \Theta = R \times R^+$$

R^+ perchè $V(X) = \sigma^2 \geq 0$

SPAZIO CAMPIONARIO

$$S_X = R \times \dots \times R = R^n$$

Ogni variabile $X_i \sim N(\mu, \sigma^2)$ ha come supporto R

VERIFICA DEL MODELLO

Per verificare la correttezza del modello usato è utile

- Sovrapporre istogramma con funzione di densità associata al modello scelto
- Confrontare i grafici dei quantili

PROCEDURE INFERENZIALI

STIMA PUNTUALE

Si vuole ottenere un valore numerico del parametro in base ai dati disponibili

STIMA INTERVALLARE

Si vuole determinare un intervallo di valori all'interno del quale trovare il valore vero del parametro da stimare

VERIFICA IPOTESI

Si parte da un valore associato al parametro e si dimostra la sua validità e correttezza

SATISTICHE CAMPIONARIE

Dato un campione casuale X_1, \dots, X_n ,