Statistica

Professoressa Barbara Pacini
Università di Pisa
a.s 2021/2022

Il processo statistico si divide in cinque fasi :

1. Definizione degli obbiettivi = descrivere le caratteristiche del fenomeno, verificare le ipotesi, esplorare le relazioni tra i vari aspetti del fenomeno etc.
2. Raccolta e organizzazione dei dati.
3. Analisi dei dati ed elaborazione metodologica.
4. Presentazione e interpretazione dei dati.
5. Utilizzo dei risultati dell'indagine.

CENSIMENTO = indagine su tutta la popolazione
(Attualmente il censimento è fatto su una parte campionaria ma è necessario poter generalizzare)

STATISTICA DESCRITTIVA	INFERENZA STATISTICA
Comprende i dati statistici elementari ottenuti grazie la sintesi numerica, le tabelle, i grafici etc. I risultati vengono descritti senza fare nessuna confusione generale sulla popolazione. I tempi sono più lunghi rispetto all'inferenza statistica ma i risultati sono più sicuri. Ex. Vengono esaminati 50 studenti e il risultato riguarda solo loro.	Vengono ricavati dei dati osservando un campione e vengono poi generalizzati sull'intera popolazione. I tempi di indagine vengono ridotti ma i risultati non saranno mai certi. Osservando un campione non si potrà conoscere con esattezza l'intero. Per essere più precisi nell'indagine deve subentrare il caso, cioè tutti devono avere la probabilità di essere selezionati in modo da non distruggere la selezione. Anche in questo caso si è soggetto comunque ad errori di campionamento. Particolare —> generale

CAMPIONE = porzione di popolazione

Tutte le volte che gli individui vengono selezionati senza variare le condizioni, l'indagine viene distorta. Anche se facciamo un indagine volontaria distorgo la valutazione perchè applico un'autovalutazione.

Può esserci anche un errore non campionario ma questo errore può essere tenuto sotto controllo.
Ex. Mancate risposte, risposte date sbagliate etc.

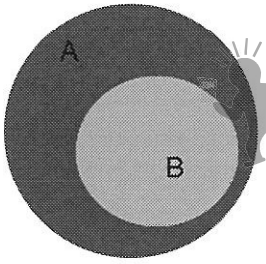
PARAMETRO
Ex. Risultati delle elezioni → Obbiettivo conoscitivo.
Dato esteso alla generalità.
Dato reale (il campione non può equivalergli perfettamente).

→ Percentuale campionaria su cui viene effettuata l'indagine vera e propria.

STIMA
Ex. Proiezione elettorale

Più allargo la porzione su cui eseguo l'indagine più la stima assomiglierà al parametro ma non posso allargare troppo perchè altrimenti i tempi sarebbero eccessivamente lunghi e non sarebbe più una stima.

Ex.



Spesa mensile media

A = spesa media mensile delle 25 milioni di famiglie italiane (parametro)

B = spesa media mensile delle 20 mila famiglie analizzate (stima)

ERRORE CAMPIONARIO = il parametro è diverso dalla stima per via delle selezione casuale.
(a seconda del campione che seleziono ho stime diverse.) qualitative sconnesse le qualità non possono essere messi in ordine di preferenza ordinarie possono essere messi in ordine di preferenza ex soddisfazione

VARIABILITÀ CAMPIONARIA	VARIABILITÀ NATURALE
Può essere arginata selezionando bene il proprio campione ma non può essere eliminata.	Vari individui avranno sempre caratteristiche diverse che mutano con il tempo quindi questa variabilità non può essere né arginata né eliminata.

VARIABILI

- Quantitative
  - Discrete = conteggio (n° naturali)
  - Continue = valori intermedi e mutabili (n° reali)
- Qualitative
  - Sconnesse = le qualità non possono essere
  - Ordinarie = possono essere messe in ordine di preferenza (ex. soddisfazione)

Come organizzare le informazioni

1. **MATRICE DEI DATI** : Ogni riga rappresenta un individuo, cioè si inseriscono le sue risposte al questionario un questionario. Ciascuna colonna rappresenta le risposte relative ad una domanda.

Se c'è un valore intermedio si mette **-/** o **/-** a seconda del valore che vogliamo escludere.

Nelle tabelle ci sono tante righe quante sono le unità statistiche e tante colonne quante sono le variabili.

Ex.

ID	Età	Ore di studio	Esami
1	20	20	4
2	19	10	6

Schema:

U (unità statistica)	X (variabile 1)	Y (variabile 2)	Z (variabile 3)
1	X1 (modalità della variabile X assunta dalla prima unità statistica)	Y1	Z1
2	X2	Y2	Z2
3	X3	Y3	Z3
N (intera popolazione)	XN	YN	ZN

Se inserisco " i " come unità statistica intendo un termine generico

X_i = variabile X in termini generici

2. **DISTRIBUZIONI DI FREQUENZA UNIVARIATA** : È una tabella che sintetizza una matrice dei dati considerando una variabile alla volta, detta anche "distribuzione di frequenza semplice".

Ex. Variabile del genere –> variabile qualitativa nominale dicotomica (M o F).
É una variabile sconnessa, non importa l'ordine della variabile considerata.

Matrice :

ID	Genere	Anni
1	M	20
2	F	18
3	F	19

ID	Genere	Anni
4	F	20
5	M	18

Posso analizzare tutta la matrice o anche un singolo sottogruppo di mio interesse (ex. Ragazzi tra i 20 e i 19 anni) e si parlerà comunque di frequenza assoluta e relativa anche se entrambe riferite ad un sottogruppo.

Distribuzione di frequenza semplice :

Modalità di genere	Frequenza assoluta
M	2
F	3

master COPY

3 + 2= 5 –> numerosità complessiva (somma delle unità statistiche della matrice)

Si può calcolare anche una frequenza relativa che deriva da quella assoluta

Fr = Fa / Nc ex. Fr (M) = 2/5 = 0,4 (40%) Fr (F) = 3/5 = 0,6 (60%)

Fr (X) + Fr (Y) = 1 (100%) ex. Fr (M) + Fr (F) = 0,4 + 0,6 = 1

Dispositivo (variabile)	F.A	F.R
Computer	4	0,33
Smartphone	4	0,33
Tablet	2	0,17
Vari	2	0,17

Trattandosi di una distribuzione di frequenza semplice la variabile considerate sempre una.

Giudizio	F.A	F.R	F.C.A	F.C.R
Insoddisfatto	3	0,25	3	0,25
Abbastanza insoddisfatto	4	0,33	7	0,58
Abbastanza soddisfatto	2	0,17	9	0,75
Soddisfatto	3	0,25	12	1
Totale	12	1		

master COPY

Può essere costruita anche una frequenza cumulata che si divide in assoluta e relativa in base a quella da cui parte (si somma la frequenza considerata a quelle precedenti).

master COPY COPISTERIA

É possibile fare una tabella di distribuzione di frequenza unitaria anche con variabili quantitative se non ci sono troppi dati altrimenti sarebbe inutile.
Per questo motivo non si può fare con le variabili quantitative continue altrimenti che non vengono divise in classi.

Ex. Variabile quantitativa discreta :

N. Esami	F.A
1	1
2	2
3	1
4	3
5	1
6	4

master COPY

Ex. Variabile quantitativa continua :

Ore di studio	F.A
5 -/ 10	5
10 -/ 15	2
15 -/ 20	4
20 -/ 25	1

Se sono interessata più ad una fascia che ad un'altra posso suddividerla ulteriormente in classi più piccole, anche se questo complica i calcoli.

AMPIEZZA = differenza tra estremo superiore e inferiore
Ex. 10 - 5 = 5 15 - 10 = 5. (in questo caso le classi sono equiampie)

Ore di lezione	F.A	F.R	F.C.A	F.C.R
0 -/ 5	1	0,08	1	0,08
5 -/ 10	3	0,25	4	0,33
10 -/ 20	4	0,33	8	0,66
20 -/ 30	3	0,25	11	0,92
30 —>	1	0,08	12	1

master COPY

(L'ultima classe non può essere aperta quindi se voglio utilizzare il tasto devo ipotizzare un massimo)

Schema :

master COPY COPISTERIA