

4. LE PRIME TRE FASI DI UN'INDAGINE STATISTICA; TERMINOLOGIA

PRIMA FASE: LA SCELTA DEL "FENOMENO COLLETTIVO" DA ANALIZZARE,

vale a dire la scelta del "CARATTERE" che si vuole studiare, e della "POPOLAZIONE" di cui ci si vuole occupare.

Come abbiamo già accennato, a volte interesserebbe una "popolazione" nella sua interezza, ma per motivi di tempi, di costi, di fattibilità se ne prende solo una parte, un "campione" (basti pensare ai sondaggi elettorali ...); certo, si porrà poi il problema di valutare in che misura l'indagine fatta sul campione possa essere rappresentativa a riguardo della popolazione intera ... di questo si occupa la "statistica inferenziale".

Si sceglie, per il "carattere", un insieme di "MODALITÀ".

Il carattere "sesso" fra i bambini di una scuola elementare si può manifestare in sole due modalità: Maschile o Femminile, ma sovente c'è invece una certa **discrezionalità**:

ad esempio, il carattere "soddisfazione dell'utenza rispetto a un certo prodotto" potrebbe essere analizzato

- nelle 3 modalità "Poco soddisfatto/Sufficientemente soddisfatto/Molto soddisfatto",
- oppure nelle 5 modalità: "Per niente/Poco/Sufficientemente/Molto/Moltissimo",
- oppure ancora domandando di esprimere la propria soddisfazione con un voto, che so, da 0 a 10.

Le fasi successive dell'indagine statistica sono finalizzate innanzitutto a stabilire ed annotare in quante, fra le "unità statistiche", si presenta ciascuna delle "modalità".

SECONDA FASE: LA RILEVAZIONE DEI DATI

Si può effettuare:

- con l'**osservazione o misurazione diretta** (ad es. per il colore degli occhi, o per il peso ..., o per rilevare le condizioni di salute di un ammalato al quale sia stato somministrato un dato farmaco, ...);
- tramite un'**intervista**;
- tramite un **questionario**, che potrà essere:
 - I) a **risposta chiusa**; II) a **risposta aperta** (più laborioso, in questo caso, lo spoglio dei dati, e meno facile l'interpretazione delle risposte ...); III) "**semistrutturato**" (è un "misto" fra le due tipologie precedenti).

TERZA FASE: LO SPOGLIO DEI DATI

Si conta, per ciascuna delle "modalità" del "carattere", quante fra le "unità statistiche" presentano quella modalità.

Si annotano questi conteggi in una tabella che prenderà il nome di "**DISTRIBUZIONE DI FREQUENZA**".

Ad esempio: "Qual è lo sport che pratici con maggiore divertimento?" (in una popolazione di 480 giovani)

	Frequenza assoluta	Frequenza relativa (appross.)	Freq. rel. percentuale (appross.)
Calcio	120	0,25	25 %
Pallavolo o Basket	92	0,19	19 %
Footing, Atletica leggera	62	0,13	13 %
Nuoto	57	0,12	12 %
Altro sport	100	0,21	21 %
Nessuno	49	0,10	10 %
TOTALE	480	1,00	100 %

A volte si annota solo la "**FREQUENZA ASSOLUTA**" di ciascuna modalità, ossia il **NUMERO DI UNITÀ STATISTICHE NELLE QUALI QUELLA MODALITÀ SI È PRESENTATA**.

Altre volte (come abbiamo fatto nella tabella precedente) si va a calcolare anche la "**FREQUENZA RELATIVA**" ossia il **rapporto, il quoziente, fra la frequenza assoluta e il numero totale di unità statistiche**:

$$\text{frequenza relativa} = \frac{\text{frequenza assoluta}}{\text{numero delle unità statistiche}}$$

La freq. relativa esprime *quale parte, quale frazione* delle unità statistiche presenta quella determinata modalità.

Dire che la frequenza relativa della modalità "calcio" è stata $\frac{120}{480} = \frac{1}{4} = 0,25$ significa affermare

che $\frac{1}{4}$ dei giovani interpellati ritiene, fra gli sport praticati, il calcio come il più divertente.

Se la frequenza relativa viene poi moltiplicata per 100, si avrà la "**FREQUENZA PERCENTUALE**" che ci dice **quante unità statistiche su 100 hanno presentato quella modalità**:

$$\text{frequenza relativa percentuale} = \text{frequenza relativa} \cdot 100 = \frac{\text{frequenza assoluta}}{\text{numero delle unità statistiche}} \cdot 100$$

Nel nostro esempio, il calcio ha come frequenza percentuale $0,25 \cdot 100 = 25$: lo predilige il 25% degli interpellati.

GLOSSARIO, SINONIMI (felici e meno felici)

Ribadiamo che prende il nome di
DISTRIBUZIONE DI FREQUENZA
L'INSIEME DELLE COPPIE ORDINATE
(MODALITÀ, FREQUENZA ASSOLUTA);



in altre parole,

LA TABELLA CHE A CIASCUNA MODALITÀ ASSOCIA LA SUA FREQUENZA ASSOLUTA, ossia il numero delle unità statistiche che presentano quella modalità.

Se il carattere è quantitativo, e le sue modalità - espresse in questo caso da valori numerici - sono ripartite per comodità in intervalli, questi intervalli vengono chiamati le "CLASSI DI FREQUENZA".

Ad esempio, le modalità del carattere quantitativo

"estensione S in metri quadrati dell'appartamento in cui si risiede"

potrebbero essere riunite, all'atto di compilare la tabella delle coppie (*modalità, frequenza*), negli intervalli seguenti:

$$S < 40m^2;$$

$$40m^2 \leq S < 60m^2;$$

$$60m^2 \leq S < 80m^2;$$

$$80m^2 \leq S < 100m^2;$$

$$100m^2 \leq S < 120m^2;$$

$$120m^2 \leq S < 140m^2;$$

$$S \geq 140m^2$$

Avremmo allora 7 "classi di frequenza".

Ovviamente, al momento di scegliere le classi di frequenza dobbiamo fare in modo che

- la loro unione dia tutto l'intervallone delle possibili modalità
- e che, prese due qualsiasi di esse, la loro intersezione sia vuota.

Di una distribuzione di frequenza si dice che è:

- una "SERIE", se il carattere al quale si riferisce è qualitativo
- una "SERIAZIONE", se il carattere è quantitativo.

Una **distribuzione di frequenza** può anche essere chiamata

- "MUTABILE statistica" se il carattere di riferimento è qualitativo
- "VARIABILE statistica" se il carattere di riferimento è quantitativo.

Talvolta gli stessi sostantivi "mutabile", "variabile"

vengono impiegati con riferimento al *carattere* più che alla *distribuzione*.

Questa terminologia si può benissimo ignorare ... dimenticala pure ... tuttavia, te l'ho citata per avvertirti che consultando un testo o un sito di statistica, ci si deve rassegnare a "digerire" parole che a volte sembrano fatte apposta per complicare inutilmente le cose.

Avrai osservato che, in statistica, sovente si hanno più possibilità diverse per dare nomi ai concetti.

A dire il vero, alcuni termini danno l'impressione di non essere molto "azzeccati", o di venire impiegati più per mettere in sudditanza psicologica il lettore che per agevolarlo ...

In queste lezioni cercheremo sempre di adottare la terminologia a nostro modesto avviso più chiara, informando lo studente sui possibili sinonimi (almeno, su alcuni fra i tanti).

E' noto che **PER SEPARARE LA PARTE INTERA DA QUELLA DECIMALE** in un numero si può utilizzare la *virgola*, o in alternativa il *punto decimale*.

Noi nel nostro corso scegliamo di norma la prima strada - più diffusa in Italia - ma a volte preferiamo invece la seconda, per motivi vari di opportunità (e un po' anche per abituare il lettore alla doppia possibilità).

♥ **In questo capitolo dedicato alla statistica il separatore sarà la virgola,**

anche perché nel capitolo si invita sovente a servirsi di un *foglio elettronico*

(programma per computer in grado di visualizzare tabelle, effettuare calcoli e tracciare grafici),

e in un foglio elettronico (versione italiana) occorre forzatamente fare uso della virgola, in quanto

un numero scritto con un punto verrebbe interpretato dal programma come finalizzato a indicare un'ora del giorno.

Le **FASI SUCCESSIVE** dell'indagine statistica, di cui andremo ad occuparci nel seguito, consistono

- nella **RAPPRESENTAZIONE GRAFICA** dei risultati
- e nella loro **ELABORAZIONE STATISTICA**.

Ma a questo punto ti direi di fermarti per fare qualche facile (e divertente) esercizio.

ESERCIZI sui concetti introduttivi alla Statistica Descrittiva (risposte a pag. 422)

- 1) Per ognuno dei seguenti fenomeni collettivi, individua una possibile “popolazione” e uno o più possibili “caratteri”, poi per ciascun carattere uno o più possibili insiemi di modalità:
 - a) le caratteristiche fisiche delle persone
 - b) il lavoro
 - c) le abitudini di spesa
 - d) l'appartamento in cui si vive
 - e) la compagnia degli animali domestici
- 2) Per ciascuno dei seguenti caratteri, stabilisci se è qualitativo *ordinato*, qualitativo *sconnesso*, quantitativo *discreto* o quantitativo *continuo*:
 - a) l'età alla quale una donna ha avuto il primo figlio
 - b) la squadra di calcio preferita
 - c) il numero di sere in cui uno studente esce abitualmente di casa in una settimana
 - d) il peso del proprio zainetto nell'entrare a scuola la mattina
 - e) il numero di libri presenti nello zainetto di uno studente all'ingresso a scuola
 - f) il gradimento di un programma televisivo
 - g) la nazionalità degli ospiti di un albergo
 - h) il numero di abitanti di un comune
- 3) Intervista telefonica a 50 persone che hanno risposto “sì” alla domanda “Possiede un gatto?”
Fra parentesi, il numero di risposte.

L'animale vive in casa o fuori?	<input type="checkbox"/> Solo in casa (14)	<input type="checkbox"/> Sia in casa che fuori (24)	<input type="checkbox"/> Quasi sempre fuori (12)
Quanto ha speso negli ultimi 30 gg per l'alimentazione del gatto?	<input type="checkbox"/> Meno di 10 euro (25)	<input type="checkbox"/> Da 10 € a 20 € (22)	<input type="checkbox"/> Più di 20 € (3)
Ha fatto ricorso al veterinario per il gatto negli ultimi 12 mesi?	<input type="checkbox"/> Mai (33)	<input type="checkbox"/> Una volta (13)	<input type="checkbox"/> Più di una volta (4)

- a) Qual è la “popolazione statistica” in questo caso? Quale il “fenomeno collettivo”? Quali i suoi “caratteri”?
 - b) Riconosci, fra i caratteri studiati, quelli “quantitativi” e quelli “qualitativi”?
 - c) Quali sono le “modalità” scelte per ciascun carattere?
 - d) Il carattere qualitativo considerato è “ordinato” o “sconnesso”?
 - e) Per il ricorso al veterinario, determina le frequenze: I) assolute II) relative III) percentuali
- 4) *Ottimo per un divertente lavoro di gruppo.*
- a) Trascrivi il questionario sottostante al computer con un *word processor* (= programma di *elaborazione testi*), ad esempio *Word* o *OpenOffice Writer*.
 - b) Stampa. Fotocopia. Distribuisci.
 - c) Raccogli i questionari compilati.
 - d) Con un *foglio elettronico* (es. *Excel* o *OpenOffice Calc*) salva gli esiti in un file.

QUESTIONARIO DEL CURIOSONE

Grazie ♥ se vorrai riempire questo questionario, RIGOROSAMENTE ANONIMO!!!

- 1) La tua altezza, in cm:
 - 2) Il tuo peso, in kg:
 - 3) Il tuo numero di scarpe:
 - 4) Il numero dei tuoi fratelli (escluso te; devono avere la stessa tua mamma e papà naturali):
 - 5) Quanti anni aveva tua mamma quando ha avuto il primo figlio?
 - 6) A che ora vai a letto, di solito, quando il giorno dopo devi andare a scuola? (è ammessa anche la “1/2 ora”, es. 22:30)
 - 7) Quanto prendi di “paghetta” mensilmente? Euro
 - 8) Qual è la tua materia preferita?
 - 9) Qual è la materia che trovi più antipatica?
 - 10) Dai un giudizio sulla tua scuola attuale (crocia la lettera corrispondente alla risposta):
Pulizia, servizi igienici, stato dell'edificio: a) Scarsissima b) Scarsa c) Sufficiente d) Buona e) Ottima
Preparazione del corpo insegnante: a) Scarsissima b) Scarsa c) Sufficiente d) Buona e) Ottima
Capacità degli insegnanti di capire i ragazzi: a) Scarsissima b) Scarsa c) Sufficiente d) Buona e) Ottima
 - 11) Quanti telefonini hai posseduto fino ad ora?
 - 12) A che età hai avuto il primo?
 - 13) Qual è il massimo numero di cm che hai saltato in alto in palestra?
 - 14) Sei iscritto a Facebook?
 - 15) Se sì, quanti “amici” hai? (pressappoco!):
- Sei maschio o sei femmina? Anno di nascita: Classe:

- 5) Compila la “distribuzione di frequenze” (assolute, relative e percentuali) per qualcuna delle voci dell'indagine statistica condotta attraverso il “Questionario del curiosone”. Esempi:

Altezza, in cm	Frequenza assoluta	Freq. relativa (appross.)	Freq. rel. perc. (appross.)
$150 \leq h < 155$	1	0,033	3,3 %
$155 \leq h < 160$	3	0,1	10 %
...
<i>Totale</i>	30	1	100 %

Materia più antipatica	Frequenza assoluta	Freq. relativa (appross.)	Freq. rel. perc. (appross.)
matematica	6	0,2	20 %
disegno	5	0,17	17 %
...
<i>Totale</i>	30	1	100 %

6) In un paese di montagna ci sono 40 coppie sposate. Il numero di figli è illustrato dalla seguente tabella.

0	2	1	2	3	1	1	0	2	2	0	0	2	1	2	1	2	1	3	2
2	1	0	2	2	4	0	1	1	2	1	1	1	3	0	1	0	1	1	1

Compila una distribuzione di frequenza, con le frequenze assolute, relative e percentuali.

7) E' stata fatta una indagine di classe:

“Sei soddisfatto della scuola scelta?”

(molto comodo tracciare un'asticella per ogni risposta, poi barrare con un tratto orizzontale i gruppi di 5!)

Ricostruisci il contenuto delle caselle cancellate.

	F. A.	F. R.	F. R. %
Moltissimo			20%
Molto	#### #### //		
Abbastanza	####		
Poco	####		
Pochissimo	//		

8) Si sa che 3 delle 4 possibili modalità di un carattere sono state osservate, su di un universo statistico, con frequenze relative 0,35; 0,4; 0,2. Determina la frequenza percentuale della modalità rimanente.

9) VERO O FALSO?

- Una volta fissato un carattere, la scelta delle sue modalità è sempre univocamente determinata.
- Un carattere è qualitativo quando non ha senso pensare a un ordinamento delle sue modalità.
- La statistica inferenziale ha come obiettivo innanzitutto di particolareggiare ad un sottoinsieme, le osservazioni generali riguardanti la popolazione.
- In medicina, quanto più il *p-value* è alto, tanto più si può essere persuasi che il farmaco sia “buono”.
- In una indagine sui costi di una notte in albergo a Roma, 80 euro può essere una unità statistica.

10) Inventa e realizza un'indagine, fra i tuoi compagni di classe, sui seguenti fenomeni collettivi:

- il tempo di permanenza davanti ad un monitor, la dipendenza dalla tecnologia
- i mezzi di trasporto posseduti in famiglia e quelli abitualmente adoperati
- educazione e maleducazione
- cibo per lo stomaco
- cibo per la mente

11) Le medie dei voti in pagella in una classe alla fine del primo quadrimestre.

Raggruppa i dati in classi di frequenza; compila la distribuzione di frequenza

6,75	5,75	6,25	8,13	7,75	6,75	7	7,63	5,5	8,5	6,25	5,63
5	7,88	4,75	8,88	5,25	6,63	6,5	8	7,75	7,5	6,25	5

12) N° di giorni in cui un libro è stato trattenuto in prestito dagli utenti di una biblioteca. Raggruppa i dati in classi di frequenza; compila la distribuz. di frequenza, calcolando anche frequenze relative e percentuali.

7	15	4	10	21	11	9	5	23	28	18	12	15	14	13	22	19	20	7	15	18	12	18
13	12	9	7	9	15	14	10	5	14	16	18	19	21	22	3	16	19	7	26	15	17	12

13) Per le modalità di quali, fra i seguenti caratteri, è opportuna una ripartizione in “classi di frequenza”?

- L'ammontare della paghetta settimanale degli adolescenti
- Il consumo annuale di acqua di una famiglia
- Il numero di quotidiani acquistati da un individuo negli ultimi 30 giorni
- La materia più amata dagli studenti
- Il numero di televisori in una casa
- Il massimo numero di centimetri realizzati nel salto in alto in palestra
- Il voto con cui uno studente è stato promosso in Terza Media (dal 6 al 10)
- Il voto con cui uno studente ha conseguito il diploma di scuola secondaria (da 60 a 100 e lode)

14) Come si potrebbe verificare se è attendibile il detto “Donne al volante, pericolo costante”?

15) Con riferimento al paragrafo “due righe di storia”, inventa una serie di 7 domande che possano andar bene per una competizione di classe “*maschi contro femmine*”. Parta poi la gara!

16) Ricerca su Internet il significato dei seguenti termini legati alla statistica:

contingenza, scala mobile, indice di ascolto, share, demografia, exit poll, polizza vita.