

■ **Statistica descrittiva: le misure della tendenza centrale**

4 ottobre 2011

Statistica sociale

1

Le misure della tendenza centrale

- Vediamo i passi che abbiamo fatto finora:
- 1) Abbiamo stabilito che cosa è un dato statistico (valore di una variabile associato a una unità statistica);
- 2) Abbiamo visto i diversi tipi di variabili, a seconda delle scale di misura;
- 3) Abbiamo organizzato i dati in un "archivio" (la matrice "caso X variabile"), per poterli analizzare;
- Il successivo passo:
- 4) Costruire una misura, un "indice" che **SINTETIZZI** l'informazione presente in una popolazione (una classe, una scuola, un comune, un qualunque insieme di unità statistiche)

4 ottobre 2011

Statistica sociale

2

Le misure della tendenza centrale

- Come si può realizzare questo obiettivo?
- Dipende, ancora una volta, dalla scala di misura
- Ogni scala di misura ha il proprio indice della tendenza centrale
- Più è complessa la scala, più complesso sarà il relativo indice di tendenza centrale
- Se una scala è più complessa, ammette anche gli indici di tendenza centrale delle scale meno complesse

4 ottobre 2011

Statistica sociale

3

In termini schematici:

Scala di misura	Nominale	Ordinale	Ad intervalli	A rapporti
Indici di tendenza centrale ammissibili	Moda (la modalità con la frequenza più alta)	Idem + Mediana (la modalità che divide in due parti uguali la distribuzione dei soggetti, ordinati sulla base della proprietà in questione)	Idem + Media aritmetica (la somma delle modalità della variabile in questione per i tutti i soggetti diviso il numero dei soggetti)	Idem + Media geometrica

4 ottobre 2011

Statistica sociale

4

Torniamo alla nostra matrice CxV (n=200)

VARIABILI	SCUOLA	PREP	ANNONASC	VOTO
CASI				
Giacomo	L	Ottima	1988	90
Giuseppe	P	Sufficiente	1986	80
Maria	L	Buona	1988	94
Laura	L	Ottima	1989	96
Filippo	T	Insufficiente	1987	70
....
Federica	T	Sufficiente	1987	72

4 ottobre 2011

Statistica sociale

5

Variabile 1: la scuola frequentata

Dal conteggio dei dati relativi alle 200 matricole, sappiamo che la distribuzione, secondo il tipo di scuola frequentata, è la seguente:

<i>Scuola</i>	<i>n</i>	<i>%</i>
Licei	50	25%
Ist.tecnici	70	35%
Ist.professionali	80	40%
TOTALE	200	100%

4 ottobre 2011

Statistica sociale

6

La moda

- La moda (o valore modale) è l'indice di tendenza centrale più semplice
- È il valore della variabile considerata più frequente nella popolazione
- Nel nostro caso abbiamo:
- Frequenza(istituto professionale)=80
- Quindi il valore modale è: ISTITUTO PROFESSIONALE

4 ottobre 2011

Statistica sociale

7

La moda

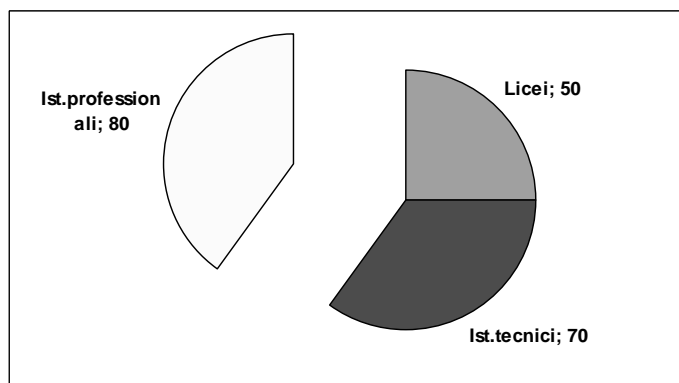
- La moda è l'unico indice di tendenza centrale che si può ricavare per dati su scala NOMINALE;
- La moda si può ricavare in tutte le scale di misura, anche in quelle più complesse

4 ottobre 2011

Statistica sociale

8

Un semplice grafico



4 ottobre 2011

Statistica sociale

9

Variabile 2: il giudizio sulla preparazione

Conosciamo la distribuzione; stavolta le modalità hanno anche un ORDINE:

	Freq. assoluta	%
Insufficiente	20	10
Sufficiente	100	50
Buona	60	30
Ottima	20	10
Totale	200	100

4 ottobre 2011

Statistica sociale

10

La distribuzione cumulata

Si ottiene accumulando le modalità una sull'altra, seguendo il loro ordine:

	Freq. assoluta cumulata	% cumulata
FINO A "insufficiente"	0+20	0+10=10%
FINO A "sufficiente"	100+20=120	50+10=60%
FINO A "buona"	100+20+60=180	50+10+30=90%
FINO A "ottima" = TOTALE	200	100%

4 ottobre 2011

Statistica sociale

11

La mediana

- La mediana (o valore mediano) è quel valore della variabile che "spezza" in due parti uguali (50% e 50%), almeno in modo approssimato, la distribuzione cumulata;
- Nel nostro caso, poiché si ha:
 - **% cumulata("sufficiente")= 60%**, e nessun valore più basso raggiunge il 50%, la mediana è: "sufficiente"

4 ottobre 2011

Statistica sociale

12

La mediana

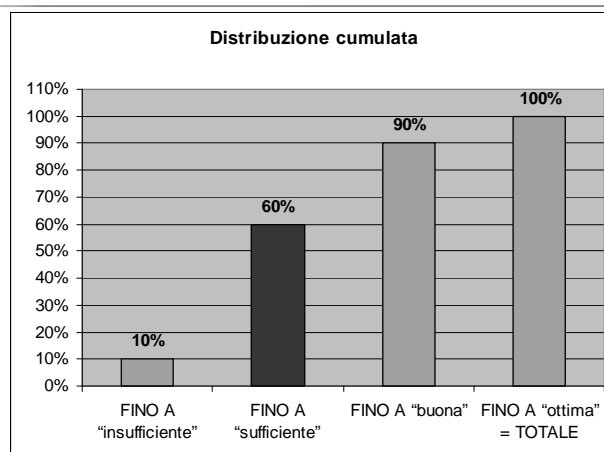
- La mediana è l'indice più complesso che sia possibile utilizzare per dati su scala ordinale;
- La mediana si può utilizzare anche in tutte le scale più complesse di quella ordinale (ad intervalli o a rapporti)

4 ottobre 2011

Statistica sociale

13

Un semplice grafico



4 ottobre 2011

Statistica sociale

14

Altri indici "di posizione"

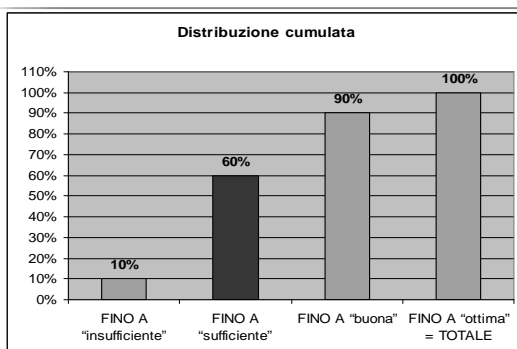
- Oltre alla mediana, può essere interessante valutare altri "punti critici" della distribuzione
- Sono importanti:
 - Il **primo quartile**: il valore in corrispondenza del quale la variabile ha accumulato il 25% delle unità statistiche;
 - Il **terzo quartile**: il valore in corrispondenza del quale la variabile ha accumulato il 75% delle unità statistiche;
- Il **10° percentile**: il valore in corrispondenza del quale la variabile ha accumulato il 10% delle unità statistiche;
- Il **90° percentile**: il valore in corrispondenza del quale la variabile ha accumulato il 90% delle unità statistiche;
- Gli indici di posizione sono importanti nella didattica e nella valutazione: si può, ad esempio, decidere che tutti gli allievi che NON RAGGIUNGONO il 10° percentile siano RESPINTI in una prova di ammissione

4 ottobre 2011

Statistica sociale

15

Nel nostro esempio:



10° percentile = "insufficiente"

90° percentile = "buona"

È più o meno quello che ci aspettavamo

4 ottobre 2011

Statistica sociale

16

Variabile 3: anno di nascita

- Per le nostre finalità conoscitive, conviene trasformare gli "anni di nascita" in età, sottraendo il "2011" all'anno di nascita;
- Una trasformazione di questo tipo si chiama "traslazione", e, in una scala ad intervalli, applicare una traslazione non cambia le proprietà dei dati (su questo aspetto non ci dilunghiamo)

4 ottobre 2011

Statistica sociale

17

Le età

Ricaviamo la seguente distribuzione per età delle 200 matricole:

Età anni	Freq.assolute	%
19 (n.1992)	18	9
20 (n.1991)	160	80
21 (n.1990)	18	9
22 (n.1989)	4	2
Totale	200	100

4 ottobre 2011

Statistica sociale

18

La media aritmetica

- L'età è una variabile quantitativa vera e propria;
- Sulla scala ad intervalli, come ricorderete, sono possibili le operazioni di differenza e di somma;
- La media aritmetica è quel valore della variabile che ciascuna unità statistica avrebbe se TUTTE le unità statistiche fossero uguali

4 ottobre 2011

Statistica sociale

19

Come si calcola la media aritmetica?

- Da questa definizione, segue che la media aritmetica si calcola in due passi:
 - 1) sommando i valori di tutte le unità statistiche (se sono raggruppate in classi, si fa ancora prima)
 - 2) dividendo tale somma per il numero di unità statistiche

4 ottobre 2011

Statistica sociale

20

Come si calcola la media aritmetica?

Età anni	Freq.ass.	%
19 (n.1992)	18	9
20 (n.1991)	160	80
21 (n.1990)	18	9
22 (n.1989)	4	2
Totale	200	100

Somma = $(19 \cdot 18 + 20 \cdot 160 + 21 \cdot 18 + 22 \cdot 4) = 4008$

n = 200

Media = Somma / n = $4008 / 200 = 20,04$

4 ottobre 2011

Statistica sociale

21

In simboli:

- La formula per il calcolo della media aritmetica, se gli n dati (nel nostro caso 200) sono organizzati in una distribuzione con k classi (che nel nostro caso sono 4), è la seguente:

$$\bar{x} = \frac{\sum_{i=1}^k x_i n_i}{n}$$

Dove:

$$\sum_{i=1}^k n_i = n$$

Il simbolo Σ si legge: somma di

4 ottobre 2011

Statistica sociale

22

In simboli:

- Nel caso i dati siano rappresentati da una semplice SERIE di dati (che va da 1 a n), e non da una distribuzione, la formula è ancora più semplice.

$$\bar{x} = \frac{\sum_{j=1}^n x_j}{n}$$

4 ottobre 2011

Statistica sociale

23

La media aritmetica

- Per calcolare la media aritmetica, i dati devono essere almeno su **scala ad intervalli**;
- Se i dati sono su una scala a rapporti, si può comunque calcolare la media aritmetica;
- Su una scala a rapporti, si può calcolare, oltre alla media aritmetica, anche la media geometrica; di questo aspetto, però, non ci occupiamo

4 ottobre 2011

Statistica sociale

24

Un semplice grafico



4 ottobre 2011

Statistica sociale

25

Riassumendo:

Scala di misura	Nominale	Ordinale	Ad intervalli	A rapporti
Indici di tendenza centrale ammissibili	Moda (la modalità con la frequenza più alta)	Idem + Mediana (la modalità che divide in due parti uguali la distribuzione dei soggetti, ordinati sulla base della proprietà in questione)	Idem + Media aritmetica (la somma delle modalità della variabile in questione per i tutti i soggetti diviso il numero dei soggetti)	Idem + Media geometrica

4 ottobre 2011

Statistica sociale

26