

LEZIONE 2

Riassumere le informazioni:

- MEDIA ARITMETICA
- MEDIANA, MODA, QUANTILI

LE MEDIE

MEDIE ALGEBRICHE: calcolate con operazioni algebriche sui valori del carattere (*media aritmetica*) – per variabili

MEDIE LASCHE: determinate in base alla loro frequenza o alla posizione occupata nella graduatoria delle osservazioni individuali (*moda, mediana, quantili*),



Le medie lasche non risentono della presenza di valori estremi.

La media aritmetica

MEDIE ALGEBRICHE: calcolate con operazioni algebriche sui valori del carattere – solo per variabili

Media aritmetica

= equiripartire il carattere tra tutte le unità del collettivo

- viene indicata con \bar{x} o μ_x

Protocollo elementare

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_N}{N} = \frac{1}{N} \sum_{j=1}^N x_j$$

Distrib. di frequenze assolute

$$\bar{x} = \frac{n_1 x_1 + n_2 x_2 + \dots + n_k x_k}{N} = \frac{1}{N} \sum_{i=1}^k x_i n_i$$

Distrib. di frequenze relative

$$\bar{x} = \sum_{i=1}^k x_i f_i$$

$$\bar{x} = \frac{1}{N} \sum_{i=1}^k \hat{x}_i n_i$$

Distrib. per intervalli

MEDIA ARITMETICA – protocollo elementare

Protocollo elementare

anni	precipitazioni in millimetri
1981	608,6
1982	694,0
1983	726,4
1984	1.128,6
1985	760,9
1986	887,6
1987	904,6
	5.710,7

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_N}{N} = \frac{1}{N} \sum_{j=1}^N x_j$$

$$\bar{x} = \frac{608,6 + 694,0 + \dots + 904,6}{7} = \frac{1}{7} \sum_{j=1}^7 x_j = 815,8$$

MEDIA ARITMETICA – distrib. di frequenze assolute

NUM. FIGLI	FREQ.
X	n
0	18
1	33
2	43
3	22
4	11
5	3
	130

$$\bar{x} = \frac{n_1x_1 + n_2x_2 + \dots + n_kx_k}{N} = \frac{1}{N} \sum_{i=1}^k x_i n_i$$

$$\bar{x} = \frac{0 \cdot 18 + 1 \cdot 33 + \dots + 5 \cdot 3}{130} = \frac{1}{130} \sum_{i=1}^6 x_i \cdot n_i = 1,877$$

5

MEDIA ARITMETICA – distrib. di frequenze relative

NUM. FIGLI	freq.ass. n	f. relativ f
X	n	f
0	18	0,138
1	33	0,254
2	43	0,331
3	22	0,169
4	11	0,085
5	3	0,023
	130	1

$$f_i = \frac{n_i}{N}$$

$$\bar{x} = \sum_{i=1}^k x_i f_i$$

$$\bar{x} = 0 \cdot 0,138 + 1 \cdot 0,254 + \dots + 5 \cdot 0,023 = \sum_{i=1}^6 x_i \cdot f_i = 1,877$$

6

MEDIA ARITMETICA – distribuzione per intervalli

classi di consumo	frequenze	valore centrale	
Kw/h	n	\hat{x}_i	
10 -- 15	3	12,5	$= (15+10)/2$
15 -- 20	9	17,5	
20 -- 30	12	25	
30 -- 50	18	40	
50 -- 75	9	62,5	
75 -- 100	6	87,5	
100 -- 200	3	150	
	60		

$$\bar{x} = \frac{1}{N} \sum_{i=1}^k \hat{x}_i n_i \Rightarrow \bar{x} = \frac{12,5 \cdot 3 + 17,5 \cdot 9 + \dots + 150 \cdot 3}{60} = 45,87$$

7

proprietà MEDIA ARITMETICA – a, b, c

(a) *Identità di somma*

$$\sum_{j=1}^N x_j = \bar{x} + \bar{x} + \dots + \bar{x} = N\bar{x}$$

(b) *Nullità della somma algebrica degli scarti*

$$\sum_{j=1}^N (x_j - \bar{x}) = 0$$

(c) *Minimo della somma dei quadrati degli scarti*

$$\sum_{j=1}^N (x_j - \bar{x})^2 = \min$$

Verifica proprietà a

E' data la seguente distribuzione relativa al il numero di errori commessi da 6 studenti in una serie di test di memoria:

X
2
3
5
7
18
25

Calcolare la media aritmetica e verificarne le **proprietà fondamentali**, evidenziando graficamente che la **somma dei quadrati degli scarti dalla media aritmetica è un minimo**.

Media aritmetica:

$$\bar{x} = \frac{2+3+\dots+25}{6} = 10$$

a) **Identità di somma:**

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



$$2+3+\dots+25 = 6 \cdot 10$$

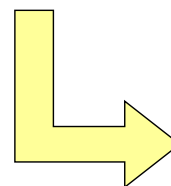
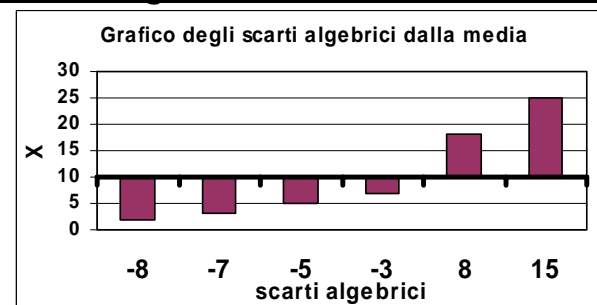
Verifica proprietà b

x
2
3
5
7
18
25

$$\bar{x} = 10 \quad \text{Media aritmetica}$$

b) **La somma algebrica degli scarti dalla media arit. è nulla:**

$$\sum_{j=1}^N (x_j - \bar{x}) = 0$$



$$(2-10) + (3-10) + \dots + (25-10) = 0$$

Verifica proprietà c

Valori di X: 2 3 5 7 18 25

$$\bar{x} = 10 \quad \text{Media aritmetica}$$

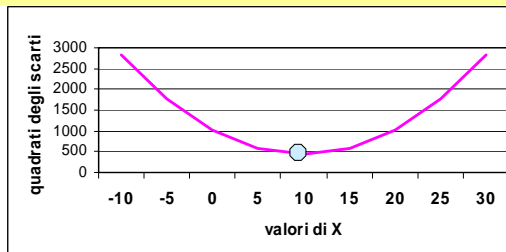
c) **Somma quadrati degli scarti è un minimo**

$$\sum_{j=1}^N (x_j - \bar{x})^2 = \min$$

$$\underbrace{(2-10)^2}_{64} + \underbrace{(3-10)^2}_{49} + \underbrace{(5-10)^2}_{25} + \underbrace{(7-10)^2}_{9} + \underbrace{(18-10)^2}_{64} + \underbrace{(25-10)^2}_{225} = \min = 436$$

$$(2-8)^2 + (3-8)^2 + (5-8)^2 + (7-8)^2 + (18-8)^2 + (25-8)^2 = 36+25+9+1+100+289 = 460$$

$$(2-12)^2 + (3-12)^2 + (5-12)^2 + (7-12)^2 + (18-12)^2 + (25-12)^2 = 100+81+49+25+36+169 = 460$$



LE MEDIE LASCHE - moda

La Moda

è la modalità che nell'insieme di osservazioni si presenta con la frequenza più alta (è possibile calcolarla per qualsiasi tipo di carattere), anche qualitativi e non ordinabili)

PROPRIETA': dà luogo al maggior numero di scarti nulli

- intervalli con ampiezza w_i **costante:**

intervallo modale: quello a cui corrisponde la più elevata freq. ass. " n_i "

- intervalli con ampiezza w_i **variabile:**

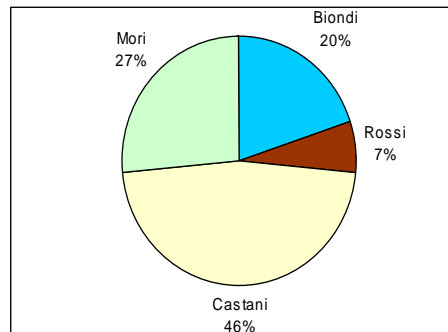
intervallo modale: quello a cui corrisponde la più elevata densità di frequenza $d_i = n_i / w_i$

Esempio calcolo valore modale

Esempio:

E' dato un collettivo di 30 studenti su cui è stato rilevato il carattere colore dei capelli. Determinare il **valore modale**, e rappresentare la distribuzione graficamente.

Xi	Ni
Biondi	6
Rossi	2
Castani	14
Mori	8
TOT.	30



VALORE MODALE: Castani

13

Moda per distribuzioni per intervalli

concetto di **classe modale** (classe a cui corrisponde la massima densità di frequenza) → non è detto corrisponda alla classe con massima frequenza (a meno che le classi non abbiano uguale ampiezza)

Densità di frequenza: $\implies = n_i / (x_i - x_{i-1})$

<u>Esempio 1:</u>				<u>Esempio 2:</u>			
Stature	freq. Ass	densità		Stature	freq. Ass	densità	ampiezza
160 — 165	10	2		20 — 25	10	2	5
165 — 170	40	8		25 — 35	60	6	10
170 — 175	60	12		35 — 37	30	15	2
175 — 180	10	2		37 — 42	10	2	5
	120				110		

Moda è: 170 —| 175

Moda è: 35 —| 37

14

La mediana

E' il valore del carattere portato dall'unità statistica che occupa il **POSTO CENTRALE** nella successione ordinata degli N valori individuali; è definita per **mutabili ordinate rettilineari** e per **caratteri quantitativi**.

L'unità centrale

Se N è dispari → occupa il posto $(N+1)/2$

Se N è pari → esistono due posti centrali: $(N)/2$ ed $(N/2)+1$



se le osservazioni a essi corrispondenti non coincidono, esse individuano due mediane; in questo caso, se il carattere è quantitativo, si conviene di prendere come valore mediano la semisomma dei due valori mediani.

$$Me = \frac{X_{\frac{N}{2}} + X_{\frac{N}{2}+1}}{2}$$

15

la mediana – protocollo elementare

E' il valore che occupa il posto centrale nella successione ordinata (non decrescente o non crescente) delle N osservazioni individuali.

anni	precipitazioni in millimetri
1981	608,6
1982	694,0
1983	726,4
1984	1.128,6
1985	760,9
1986	887,6
1987	904,6
	5.710,7

→ ordinare

precipitazioni in millimetri
608,6
694,0
726,4
760,9
887,6
904,6
1.128,6

$N=7$ (collettivo dispari) → la mediana è il carattere portato dall' unità che occupato posto $(7+1)/2$ ossia la 4° unità ed è pari a **760,9**

16

la mediana – distribuz. di frequenza

In una distribuzione di frequenza, per individuare la classe in cui è compresa l'unità statistica che nella graduatoria occupa la posizione centrale (*classe mediana*), è necessario passare alle frequenze cumulate

Esempio:

Sia data la seguente distribuzione delle famiglie bolognesi secondo il numero dei componenti:

Si determini la moda e la mediana.

N° componenti	N° famiglie (n _i)	freq. Ass.
1	69.397	69.397
2	97.944	167.341
3	93.081	260.422
4	58.590	319.012
5	18.511	337.523
6	5.264	342.787
7 e oltre	2.390	345.177
		345.177

17

Mediana – carattere discreto

N° Componenti	N° famiglie	freq. cumulate	
X _i	n _i	N _i	
1	69.397	69.397	← = n ₁
2	97.944	167.341	← = n ₁ + n ₂
3	93.081	260.422	← = n ₁ + n ₂ + n ₃
4	58.590	319.012	...
5	18.511	337.523	...
6	5.264	342.787	...
7 e oltre	2.390	345.177	...
	345.177	← = N	
Unita' centrale	172.589	← = (n+1)/2	
Mediana	3		
Moda	2		

N=345.177 è dispari → l'unità che occupa il posto centrale è 172.589 che cade nella freq. Cum. della modalità del carattere 3

18

I QUANTILI

Sono medie di posizione che suddividono la successione ORDINATA di valori individuali in un certo numero di parti ugualmente numerose;

sono definiti per variabili e per mutabili ordinate rettilineari

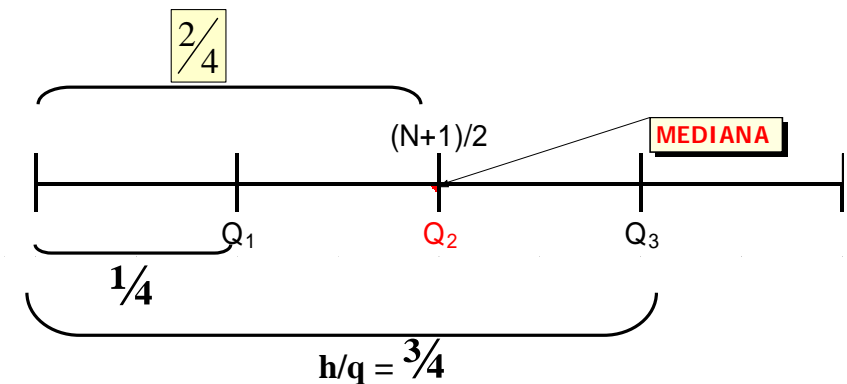
Quartili → ripartiscono in quattro gruppi di $N/4$ unità

Decili → ripartiscono in dieci gruppi di $N/10$ unità

Centili → ripartiscono in cento gruppi di $N/100$ unità

19

I QUARTILI sono 3



Q_h → indica il quartile, il pedice h , può essere 1,2 o 3 a seconda che sia il 1°, il 2° o il 3° quartile,

q → num. di gruppi in cui viene ripartito il collettivo è sempre = 4 per i quartili (mentre per i decili $q=10$ e per i centili $q=100$)

L' h -esimo quartile Q_h , dividerà il collettivo in due parti e sarà preceduto da una frazione h/q dell'intero collettivo e seguito da una frazione $1-(h/q)$

20

Calcolo quantili

Indicato con $Q_{h/q}$ il quantile dell'insieme di N osservazioni di un determinato carattere quantitativo o qualitativo ordinabile (ad esempio $Q_{1/4}$ è il primo quartile e $Q_{9/10}$ l'ultimo decile):

1) Se N è multiplo di:

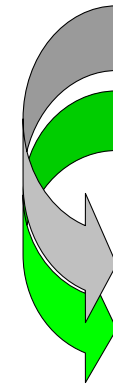
- 4 per i **quartili** \rightarrow l' h -esimo quartile è compreso tra : $(hN)/4$ ed $(hN/4)+1$
- 10 per i **decili** \rightarrow l' h -esimo decile è compreso tra : $(hN)/10$ ed $(hN/10)+1$
- 100 per i **centili** \rightarrow l' h -esimo centile è compreso tra : $(hN)/100$ ed $(hN/100)+1$

1) Se N non è multiplo di:

- 4 per i **quartili** \rightarrow l' h -esimo quartile è $(hN/4)+1$
- 10 per i **decili** \rightarrow l' h -esimo decile è $(hN/10)+1$
- 100 per i **centili** \rightarrow l' h -esimo centile è $(hN/100)+1$

Esercizio su quantili

grado istruzione	n_i	N_i
analfabeti	16	16
alfabeti senza titoli di studio	95	111
licenza di scuola elementare	213	324
licenza di scuola media inferiore	121	445
diploma di scuola media superiore	60	505
laurea	14	519
	519	
mediana corrisponde unità $(n+1)/2$	260	
90° percentile corrisponde al posto $[hN/100]+1$	468	



$(90 \cdot 519/100)+1=468$

"N=519" è dispari e non è divisibile per 100

Capitolo 2 – La statistica descrittiva: riassumere le informazioni 25

*	2.1 Introduzione	25
	2.2 Tipi di dati	26
	2.3 Scale di misura	27
	2.4 Le rappresentazioni tabellari	29
	2.5 Serie storiche e serie territoriali	34
	2.6 I dati raggruppati in classi	35
	2.7 Le rappresentazioni grafiche	37
	2.8 Misure di tendenza centrale dei dati .**.....	41

* (argomenti già trattati nella lezione 1)

** (escluso nel par. 2.8, la mediana per caratteri raggruppati in classi)