

Quartili

- × Primo quartile Q_1 : modalità che nella graduatoria (crescente o decrescente) bipartisce il 50% delle osservazioni con modalità più piccole o al più uguali alla Me
- × Terzo quartile Q_3 : modalità che nella graduatoria (crescente o decrescente) bipartisce il 50% delle osservazioni con modalità più grandi o al più uguali alla Me

U.S.	A	G	I	F	B	D	L	H	E	M	C
x_j	1	2	3	4	5	6	7	8	9	10	11

↑
 Q_1
↑
Me
↑
 Q_3

Pagina 60

Esempio

- × Distribuzione unitaria degli affitti settimanali in euro pagati da 19 studenti

Senza ordine		Ordine crescente			Affitto
U.S.	Affitto	Posto	U.S.	Affitto	
U1	40	1	U4	33	33
U2	43	2	U3	35	35
U3	35	3	U7	36	36
U4	33	4	U8	36	36
U5	45	5	U10	38	38
U6	40	6	U13	39	39
U7	36	7	U1	40	40
U8	36	8	U6	40	40
U9	42	9	U9	42	42
U10	38	10	U14	42	42
U11	48	11	U19	42	42
U12	51	12	U2	43	43
U13	39	13	U5	45	45
U14	42	14	U15	46	46
U15	46	15	U11	48	48
U16	59	16	U12	51	51
U17	53	17	U17	53	53
U18	55	18	U18	55	55
U19	42	19	U16	59	59

Primo Quartile è "38"

La mediana è "42"

Terzo Quartile è "48"

Pagina 61

Esempio: mediana e quartili

✘ Calcolo di mediana e quartili

MEDIANA

$$n=19$$

$$\frac{n+1}{2} = \frac{19+1}{2} = \frac{20}{2} = 10 \quad \rightarrow \quad Me = x_{\frac{n+1}{2}} = x_{10} = 42$$

QUARTILI

$$\frac{n}{4} = \frac{19}{4} = 4.75 \approx 5 \quad \rightarrow \quad \left\{ \begin{array}{l} Q_1 = x_{\frac{n}{4}} = x_5 = 38 \\ 3 \times \frac{n}{4} = 3 \times 5 = 15 \rightarrow Q_3 = x_{3\left(\frac{n}{4}\right)} = x_{15} = 48 \end{array} \right.$$

Medie Analitiche

✘ *Media aritmetica*

✘ Si possono calcolare solo per caratteri quantitativi (variabili continue o discrete)

✘ Sono funzioni matematiche di tutti i dati osservati

- Cambiano se si cambia anche un solo dato
- Sfruttano completamente l'informazione statistica disponibile

Media Aritmetica: il calcolo

1) Distribuzione unitaria semplice del carattere X

$$x_1, x_2, \dots, x_i, \dots, x_n$$

$$M = \frac{x_1 + x_2 + \dots + x_i + \dots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n}$$

Ammontare del carattere

U.S.	Voto Maturità
1	98
2	100
3	70
4	72
5	70
6	100
7	85
8	65
9	60
10	88

$$n = 10$$

$$\sum_{i=1}^{10} x_i = 98 + 100 + 70 + 72 + 70 + 100 + 85 + 65 + 60 + 88 = 808$$

$$M = \frac{98 + 100 + 70 + 72 + 70 + 100 + 85 + 65 + 60 + 88}{10} = \frac{808}{10} = 80.8$$

Media Aritmetica: il calcolo

2) Distribuzione semplice di frequenze assolute del carattere X

Car. X	FREQUENZE ASSOLUTE	FREQUENZE RELATIVE
x_1	n_1	f_1
x_2	n_2	f_2
...
x_i	n_i	f_i
...
x_K	n_K	f_K
Totale	n	1

$$M = \frac{x_1 n_1 + x_2 n_2 + \dots + x_i n_i + \dots + x_K n_K}{n} = \frac{\sum_{i=1}^K x_i n_i}{n} = \sum_{i=1}^K x_i f_i$$

MEDIA ARITMETICA PONDERATA

Esempio: distribuzione di frequenze

- × Distribuzione degli studenti di SDC frequentanti la facoltà nell'a.a. 2001/2002 per Num. Corsi Frequentati

$$n = 283 \quad K = 7$$

Num. Corsi Freq.	n_i	$x_i n_i$
1	15	1*15=15
2	43	2*43=86
3	103	3*103=309
4	80	4*80=320
5	32	5*32=160
6	8	6*8=48
7	2	7*2=14
Totale	283	952

$$\sum_{i=1}^7 x_i n_i = 15 + 86 + 309 + 320 + 160 + 48 + 14 = 952$$

$$M = \frac{(1 \times 15) + (2 \times 43) + (3 \times 103) + (4 \times 80) + (5 \times 32) + (6 \times 8) + (7 \times 2)}{283} = \frac{952}{283} = 3.36$$

$$M = \frac{\sum_{i=1}^7 x_i n_i}{n} = \frac{952}{283} = 3.36$$

Esempio

- × Distribuzione degli studenti di SDC frequentanti la facoltà nell'a.a. 2001/2002 per Voto di Maturità

$$n = 40 \quad K = 24$$

$$\sum_{i=1}^{24} x_i n_i = 3201$$

$$M = \frac{\sum_{i=1}^{24} x_i n_i}{n} = \frac{3201}{40} = 80.03$$

Voto Maturità	n_i	$x_i n_i$
60	1	60
62	2	124
66	1	66
67	3	201
68	2	136
70	2	140
71	2	142
72	1	72
73	2	146
74	1	74
75	2	150
76	1	76
79	1	79
80	1	80
81	1	81
82	1	82
83	1	83
86	1	86
87	1	87
90	2	180
92	3	276
93	2	186
94	1	94
100	5	500
Totale	40	3201

Media Aritmetica: il calcolo

3) Distribuzione semplice di frequenze assolute del carattere X raggruppato in classi

Car. X	n_i	Valori centrali
$(x_0, x_1]$	n_1	c_1
$(x_1, x_2]$	n_2	c_2
...
$(x_{i-1}, x_i]$	n_i	c_i
...
$(x_{K-1}, x_K]$	n_K	c_K
Totale	n	

$$c_i = \frac{x_{i-1} + x_i}{2}$$

$$M = \frac{c_1 n_1 + c_2 n_2 + \dots + c_i n_i + \dots + c_K n_K}{n} = \frac{\sum_{i=1}^K c_i n_i}{n} = \sum_{i=1}^K c_i f_i$$

Esempio

✘ Distribuzione degli studenti di SDC frequentanti la facoltà nell'a.a. 2001/2002 per Voto di Maturità

$$n = 283 \quad K = 5$$

Voto Maturità	n_i	c_i	$c_i n_i$
[60-70]	72	65.0	4680
(70-80]	78	75.0	5850
(80-90]	65	85.0	5525
(90-95]	18	92.5	1665
(95-100]	50	97.5	4875
Totale	283		22595

$$c_i = \frac{x_{i-1} + x_i}{2} \quad \rightarrow \quad i=1 \quad c_1 = \frac{60+70}{2} = 65$$

$$\sum_{i=1}^5 c_i n_i = 4680 + 5850 + 5525 + 1665 + 4875 = 22595$$

$$M = \frac{\sum_{i=1}^5 c_i n_i}{n} = \frac{22595}{283} = 79.84$$

Proprietà della media aritmetica

$$M_{maschi} = \frac{\sum_{i=1}^{10} x_i n_i}{n_{maschi}} = \frac{1248}{16} = 78.00$$

$$M_{femmine} = \frac{\sum_{i=1}^{19} x_i n_i}{n_{femmine}} = \frac{2596}{32} = 81.12$$

Maschi

Voto Maturità	n_i	$x_i n_i$
67	2	134
68	3	204
72	2	144
75	2	150
76	1	76
80	1	80
83	1	83
87	1	87
90	1	90
100	2	200
Totale	16	1248

Femmine

Voto Maturità	n_i	$x_i n_i$
60	1	60
62	2	124
66	1	66
67	1	67
68	1	68
70	2	140
71	2	142
73	2	146
74	1	74
75	1	75
79	1	79
81	1	81
82	1	82
86	2	172
90	3	270
92	3	276
93	2	186
94	2	188
100	3	300
Totale	32	2596

$$\frac{M_{maschi} + M_{femmine}}{2} = \frac{78.00 + 81.12}{2} = 79.56 \neq \text{voto medio} = \frac{\sum_{i=1}^{19} x_i n_i}{n} = \frac{3844}{48} = 80.08$$

$$M_{TOT} = \frac{M_{maschi} n_{maschi} + M_{femmine} n_{femmine}}{n} = \frac{1248 + 2596}{48} = 80.08$$

...segue proprietà della media aritmetica

- ✓ La media di un collettivo è la media aritmetica delle medie dei sottogruppi in cui si può ripartire il collettivo stesso ponderata per le numerosità dei sottogruppi

Gruppo	1	2	3	h	L
Medie	M_1	M_2	M_3	M_h	M_L
Numerosità	n_1	n_2	n_3	n_h	n_L

$$M = \frac{\sum_{h=1}^L M_h n_h}{n} \quad \text{con} \quad n = \sum_{h=1}^L n_h$$

Punti deboli della media aritmetica

- ✓ Robustezza: sensibilità ai valori estremi
- ✓ Rappresentatività nei confronti di distribuzioni asimmetriche. La media aritmetica è un valore rappresentativo nei confronti di distribuzioni simmetriche

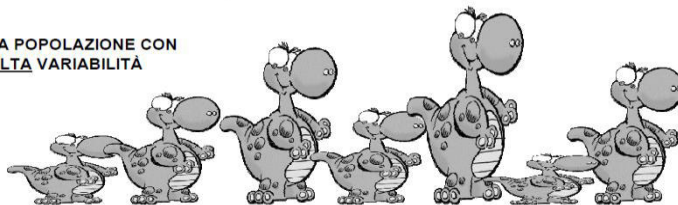
Variabilità o Dispersione

Definizione

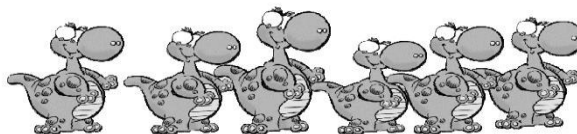
Attitudine di un fenomeno ad assumere diverse modalità

la variabile d'interesse è l'ALTEZZA

UNA POPOLAZIONE CON
MOLTA VARIABILITÀ



UNA
POPOLAZIONE
CON POCHE
VARIABILITÀ



Le medie non bastano ...

➤ Esempio: caratteri quantitativi

Condominio A u.s.	Numero televisori	Condominio B u.s.	Numero televisori
u1	8	u1	3
u2	8	u2	5
u3	8	u3	8
u4	8	u4	10
u5	8	u5	14
Me=M=8		Me=M=8	

✗ Misure di sintesi

- ✓ Medie
- ✓ Indici di variabilità/dispersione (Car. Quantitativi)
- ✓ Indici di mutabilità (Car. Qualitativi)

Requisiti degli indici di variabilità e dispersione

- ✗ Assume valore minimo se e sole se tutte le u.s. presentano la stessa modalità
- ✗ Positivo se c'è variabilità o dispersione
- ✗ Aumenta all'aumentare della diversità tra le modalità assunte dalle u.s.
- ✗ Non cambia se le frequenze vengono moltiplicate per una costante positiva

Indici di variabilità e dispersione

✗ Indici di *variabilità*

- La variabilità si misura considerando tutte le differenze tra le modalità della distribuzione presentate dalle u.s. prese due a due → *Differenze Medie*

✗ Indici di *dispersione*

- La dispersione si misura con gli scarti tra le modalità presentate dalle u.s. e un indice di dimensione della distribuzione (M o Me)
 - ✓ *Varianza*
 - ✓ *Scostamento (scarto) quadratico medio*

Esempio: calcolo della varianza

U.S.	Temperatura Minima	$(x_i - M)$	$(x_i - M)^2$
A	9	6.9	47.61
B	-2	-4.1	16.81
C	4	1.9	3.61
D	-3	-5.1	26.01
E	-2	-4.1	16.81
F	0	-2.1	4.41
G	6	3.9	15.21
H	4	1.9	3.61
I	-4	-6.1	37.21
J	9	6.9	47.61
			218.9

$$n = 10$$

$$M = \frac{\sum_{i=1}^{10} x_i}{n} = \frac{21}{10} = 2.1$$

$$\text{Var}(X) = \frac{\sum_{i=1}^{10} (x_i - M)^2}{n} = \frac{218.9}{10} = 21.89$$

$$\sigma(X) = \sqrt{\text{Var}(X)} = \sqrt{21.89} = 4.679$$

Esempio: calcolo della varianza

- ✗ Distribuzione degli studenti di SDC frequentanti la facoltà nell'a.a. 2001/2002 per Num. Corsi Frequentati

Num. Corsi Freq.	n_i	$x_i n_i$	$(x_i - M)^2$	$(x_i - M)^2 n_i$
1	15	15	5.57	83.55
2	43	86	1.85	79.55
3	103	309	0.13	13.39
4	80	320	0.41	32.80
5	32	160	2.69	86.08
6	8	48	6.97	55.76
7	2	14	13.25	26.50
Totale	283	952		377.63

$$n = 283 \quad K = 7$$

$$M = \frac{\sum_{i=1}^7 x_i n_i}{n} = \frac{952}{283} = 3.36$$

$$Var(X) = \frac{\sum_{i=1}^7 (x_i - M)^2 n_i}{n} = \frac{377.63}{283} = 1.33$$

$$\sigma(X) = \sqrt{Var(X)} = \sqrt{1.33} = 1.15$$

Scarto quadratico medio e varianza

- ✗ Distribuzione unitaria

➤ Scostamento quadratico medio (deviazione standard)

$$\sigma = \sqrt{\frac{(x_1 - M)^2 + (x_2 - M)^2 + \dots + (x_n - M)^2}{n}}$$

➤ Varianza (non ha la stessa unità di misura del carattere)

$$Var(X) = \frac{(x_1 - M)^2 + (x_2 - M)^2 + \dots + (x_n - M)^2}{n}$$

* Devianza

$$Dev(X) = (x_1 - M)^2 + (x_2 - M)^2 + \dots + (x_n - M)^2$$

- ✗ Distribuzione semplice di frequenze

➤ Scostamento quadratico medio (deviazione standard)

$$\sigma = \sqrt{\frac{(x_1 - M)^2 n_1 + \dots + (x_i - M)^2 n_i + \dots + (x_K - M)^2 n_K}{n}}$$

➤ Varianza (non ha la stessa unità di misura del carattere)

$$Var(X) = \frac{(x_1 - M)^2 n_1 + \dots + (x_i - M)^2 n_i + \dots + (x_K - M)^2 n_K}{n}$$

* Devianza

$$Dev(X) = (x_1 - M)^2 n_1 + \dots + (x_i - M)^2 n_i + \dots + (x_K - M)^2 n_K$$

Indici di variabilità relativa

- ✘ Consentono di effettuare confronti sulla variabilità di fenomeni che
 - presentano unità di misura differenti
 - pur avendo la stessa unità di misura hanno valori medi differenti e quindi distribuzioni differenti
- ✘ In alcune situazioni è fuorviante utilizzare la deviazione standard per il confronto:
 - della variabilità di una variabile osservata su due collettivi differenti di u.s.
 - della variabilità di due o più variabili osservate sul medesimo collettivo di u.s.

Indici di variabilità relativa

- ✘ Ci permettono di misurare la variabilità indipendentemente dalla grandezza e dalla scala di misura del carattere
- ✘ *Indici percentuali di variabilità o dispersione*
- ✘ *Indici di variabilità o dispersione relativa*
 - Sono numeri puri → non hanno unità di misura

Indici di variabilità relativa

✘ *Indici percentuali di variabilità o dispersione* ottenuti dividendo l'indice di variabilità (dispersione) assoluto per la media rispetto alla quale è stato calcolato

➤ Coefficiente di variazione

$$CV = \frac{\sigma}{M} \times 100$$

✘ *Indici di variabilità o dispersione relativa* ottenuti dividendo l'indice di variabilità (dispersione) assoluto per il valore massimo che esso può assumere in una situazione ipotetica

➤ Deviazione standard relativa $\frac{\sigma}{\max(\sigma)}$

★ Numero compreso tra 0 e 1

Esempio: due variabili diverse

✘ In un collettivo di 91 ragazze

➤ la media del peso è pari a 55.1 Kg e la deviazione standard è pari a 5.7 Kg

➤ La media della statura è pari a 166.1 cm e la deviazione standard è pari a 6.1 cm

✘ E' maggiore la variabilità del peso o la variabilità della statura?

✓ Le variabili misurate sono diverse perché espresse in diverse unità di misura

Esempio: due gruppi con valori molto distanti

- ✗ Tre neonati pesano rispettivamente 3, 4 e 5 Kg (media=4 Kg e $\sigma=1$ Kg)
- ✗ Tre bambini di 1 anno pesano 10, 11 e 12 Kg (media=11 Kg e $\sigma=1$ Kg)
- ✗ La variabile misurata è la stessa ma i valori medi delle osservazioni nei due gruppi sono molto distanti (le osservazioni nei due gruppi sono su diversi ordini di grandezza)

La deviazione standard è uguale nei due gruppi ma il buon senso suggerisce che la variabilità del peso sia maggiore nei neonati!!



Soluzione

	Media	Dev. Standard	CV
Peso	55.1 Kg	5.7 Kg	
Statura	166.1 cm	6.1 cm	

La variabilità del _____ è maggiore della variabilità del _____

	Media	Dev. Standard	CV
Neonati	4 Kg	1 Kg	
Bambini di 1 anno	11 Kg	1 Kg	

La variabilità del _____ è maggiore nel collettivo dei _____

Altri indici di variabilità

Campo di Variazione

$$R = \text{Max}(x_i) - \text{Min}(x_i)$$

oppure

$$R = x_{\text{max}} - x_{\text{min}}$$

- $R \geq 0$
- $R=0 \Rightarrow$ non c'è variabilità

Differenza Interquartilica

$$W = Q_3 - Q_1$$

- W rappresenta il campo di variazione per il 50% delle unità centrali

Mutabilità

- ✘ Sinonimo di *eterogeneità*
- ✘ Le misure di mutabilità si basano sull'analisi delle frequenze
- ✘ *Mutabilità nulla* = tutte le u.s. presentano la stessa modalità \rightarrow *Massima omogeneità*
 - Tutte le n_i sono uguali a zero tranne una che è pari a n
- ✘ *Mutabilità massima* = le frequenze delle varie modalità sono tutte uguali \rightarrow *Minima omogeneità*
 - $n_1=n_2=\dots=n_i=\dots=n_K=n/K$

Indici di mutabilità o eterogeneità

- ✗ Devono esser nulli \Leftrightarrow tutte le u.s. presentano la stessa modalità
- ✗ Crescere all'aumentare della mutabilità o eterogeneità
 - ✓ *Indice di Gini*

Indici di mutabilità o eterogeneità

Indice di Gini	$G = 1 - \sum_{j=1}^K \left(\frac{n_j}{n} \right)^2$	Assume valori tra 0 e $1 - (1/K)$
Indice di Gini Relativo	$\frac{G}{\max(G)} = \frac{G}{1 - \frac{1}{K}}$	Assume valori tra 0 e 1