UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

Dipartimento di Comunicazione
ed Economia

Le misure di variabilità

Le misure di centralità non bastano

Può accadere che due o più popolazioni presentino lo stesso centro, ma che il livello di sintesi sia completamente differente.

Occorrono altre misure che consentano di valutare il grado di dispersione delle modalità e la bontà della sintesi della distribuzione operata tramite gli indici di centralità.

$$\Downarrow$$

Misure di variabilità

Le misure di centralità non bastano - Esempio

Due gruppi di individui, la variabile è il peso (in *kg*)

I gruppo (*kg*)

68, 74, 81, 86, 91

II gruppo (*kg*)

45, 50, 66, 110, 129

Le misure di centralità non bastano - Esempio

Due gruppi di individui, la variabile è il peso (in *kg*)

I gruppo (*kg*)

68, 74, 81, 86, 91

II gruppo (*kg*)

45, 50, 66, 110, 129

La media è 80*kg* per entrambi i gruppi.

Attenzione! Il primo gruppo è molto più omogeneo quanto al peso dei singoli.

Guardando oltre al centro della distribuzione

Ci interessa avere anche un'idea di quanto diversi siano i valori assunti dalla variabile, ossia ci interessa avere una idea della variabilità del carattere.

Guardando oltre al centro della distribuzione

Ci interessa avere anche un'idea di quanto diversi siano i valori assunti dalla variabile, ossia ci interessa avere una idea della variabilità del carattere.

Per farlo, possiamo vedere come si muovono le osservazioni intorno al centro della distribuzione.

L'esigenza di una misura della dispersione

Studiare la variabilità di un carattere è importante perchè:

- Analizzare e misurare l'attitudine a variare di un fenomeno, rappresenta una delle finalità che si vuole perseguire con l'analisi statistica $\Rightarrow$ Valore Intrinseco

- L'impiego delle medie non è sufficiente a sintetizzare le informazioni rilevate sulla popolazione oggetto di studio, specialmente quando occorre confrontare tra di loro popolazioni diverse $\Rightarrow$ Accuratezze della Sintesi dei Dati

Requisiti per gli indici di variabilità

- Se almeno due osservazioni sono diverse tra di loro, allora dev'essere > 0

Requisiti per gli indici di variabilità

- Se almeno due osservazioni sono diverse tra di loro, allora dev'essere > 0
- Se tutte le osservazioni sono uguali tra di loro (carattere degenere), allora deve assumere il suo valore minimo

Requisiti per gli indici di variabilità

- Se almeno due osservazioni sono diverse tra di loro, allora dev'essere > 0
- Se tutte le osservazioni sono uguali tra di loro (carattere degenere), allora deve assumere il suo valore minimo
- Se il carattere X è più variabile del carattere Y, allora l'indice di variabilità di X dev'essere maggiore dell'indice di variabilità di Y

Indice

Variabilità e variabili quantitative

Indice

Variabilità e variabili quantitative

- Campo di variazione
- Differenza interquartile
- Box plot
- Varianza

Campo di variazione

Dato un insieme di n valori osservati $x_1, x_2, \ldots, x_n$ si definisce **campo di variazione (range)** la differenza tra il valore più grande e il valore più piccolo

$$R = x_n - x_1$$

Campo di variazione

Dato un insieme di n valori osservati $x_1, x_2, \dots, x_n$ si definisce **campo di variazione (range)** la differenza tra il valore più grande e il valore più piccolo

$$R = x_n - x_1$$

Pregi:

- È semplice da calcolare
- È di immediata interpretazione (rappresenta l'ampiezza dell'intervallo in cui si è manifestato il fenomeno).

Difetti:

- Dipende solo da due osservazioni e non tiene conto delle altre;
- È poco stabile in quanto estremamente sensibile agli outliers;

$\Rightarrow$ impiego limitato a poche applicazioni (esempio: controllo statistico della produzione)

Esempio: (distribuzione unitaria)

Dati i seguenti valori

3.1 5.6 1.3 4.1 1.8 2.0 1.3 1.1

calcolare il campo di variazione

Esempio: (distribuzione unitaria)

Dati i seguenti valori

3.1 5.6 1.3 4.1 1.8 2.0 1.3 1.1

calcolare il campo di variazione

Svolgimento:

Innanzitutto ordiniamo i valori:

1.1 1.3 1.3 1.8 2.0 3.1 4.1 5.6

Esempio: (distribuzione unitaria)

Dati i seguenti valori

$$3.1 \quad 5.6 \quad 1.3 \quad 4.1 \quad 1.8 \quad 2.0 \quad 1.3 \quad 1.1$$

calcolare il campo di variazione

Svolgimento:

Innanzitutto ordiniamo i valori:

$$1.1 \quad 1.3 \quad 1.3 \quad 1.8 \quad 2.0 \quad 3.1 \quad 4.1 \quad 5.6$$

Quindi:

$$R = 5.6 - 1.1 = 4.5$$

Esempio: (distribuzione di frequenze)

Data la distribuzione del numero medio di figli

X	n_j
0	5
1	12
2	19
3	9
4	4
5	1
Totale	**50**

calcolare il campo di variazione

Esempio: (distribuzione di frequenze)

Data la distribuzione del numero medio di figli

X	n_j
0	5
1	12
2	19
3	9
4	4
5	1
Totale	**50**

calcolare il campo di variazione

Svolgimento:

$$R = 5 - 0 = 5$$

Esempio: (distribuzione in classi)

Data la distribuzione delle altezze

X	n_j
(70;100]	20
(100;120]	7
(120;140]	18
(140;170]	65
(170;180]	21
(180;200]	45
(200;220]	24
Totale	**200**

calcolare il campo di variazione

Esempio: (distribuzione in classi)

Data la distribuzione delle altezze

X	n_j
(70;100]	20
(100;120]	7
(120;140]	18
(140;170]	65
(170;180]	21
(180;200]	45
(200;220]	24
Totale	**200**

calcolare il campo di variazione

Svolgimento:

$$R = 220 - 70 = 150$$

Indice

Variabilità e variabili quantitative

- Campo di variazione
- Differenza interquartile
- Box plot
- Varianza

Differenza interquartile

Rappresenta l'ampiezza dell'intervallo centrale (intorno alla mediana) nel quale si collocano il 50% dei valori.
$\Rightarrow$ Tanto più è piccola tanto più la metà delle osservazioni risulterà addensata intorno alla mediana.

Dato un insieme di n valori osservati $x_1, x_2, \ldots, x_n$ si definisce **differenza interquartile** la differenza tra il terzo e il primo quartile

$$DI = Q_3 - Q_1$$

Differenza interquartile

Rappresenta l'ampiezza dell'intervallo centrale (intorno alla mediana) nel quale si collocano il 50% dei valori.
$\Rightarrow$ Tanto più è piccola tanto più la metà delle osservazioni risulterà addensata intorno alla mediana.

Dato un insieme di n valori osservati $x_1, x_2, \ldots, x_n$ si definisce **differenza interquartile** la differenza tra il terzo e il primo quartile

$$DI = Q_3 - Q_1$$

Pregi:

- È più stabile del campo di variazione perchè non si basa sulle osservazioni estreme;

Difetti:

- Potrebbe essere nulla senza che il carattere sia degenere

Esempio: (distribuzioni unitarie)

Dati i seguenti valori (ordinati)

1.1 1.3 1.3 1.8 2.0 3.1 4.1 5.6

calcolare la differenza interquartile

Esempio: (distribuzioni unitarie)

Dati i seguenti valori (ordinati)

$$1.1 \quad 1.3 \quad 1.3 \quad 1.8 \quad 2.0 \quad 3.1 \quad 4.1 \quad 5.6$$

calcolare la differenza interquartile

Svolgimento:

$$i = 0.25 \times 8 = 2$$

$$Q_1 = \frac{x_{(2)} + x_{(3)}}{2} = \frac{1.3 + 1.3}{2} = 1.3$$

$$i = 0.75 \times 8 = 6$$

$$Q_3 = \frac{x_{(6)} + x_{(7)}}{2} = \frac{3.1 + 4.1}{2} = 3.6$$

Esempio: (distribuzioni unitarie)

Dati i seguenti valori (ordinati)

$$1.1 \quad 1.3 \quad 1.3 \quad 1.8 \quad 2.0 \quad 3.1 \quad 4.1 \quad 5.6$$

calcolare la differenza interquartile

Svolgimento:

$$i = 0.25 \times 8 = 2$$

$$Q_1 = \frac{x_{(2)} + x_{(3)}}{2} = \frac{1.3 + 1.3}{2} = 1.3$$

$$i = 0.75 \times 8 = 6$$

$$Q_3 = \frac{x_{(6)} + x_{(7)}}{2} = \frac{3.1 + 4.1}{2} = 3.6$$

Quindi:

$$DI = Q_3 - Q_1 = 3.6 - 1.3 = 2.3$$

Esempio: (distribuzioni di frequenza)

Su un gruppo di scommettitori è stata rilevata la variabile X: *scommesse effettuate nell'ultima settimana*

X	n_i
1	8
2	44
3	63
4	70
	185

calcolare la differenza interquartile

Esempio: (distribuzioni di frequenza)

Su un gruppo di scommettitori è stata rilevata la variabile X: *scommesse effettuate nell'ultima settimana*

X	n_i
1	8
2	44
3	63
4	70
	185

calcolare la differenza interquartile

Svolgimento:

X	n_i	f_i	F_i
1	8	0.04	0.04
2	44	0.24	0.28
3	63	0.34	0.62
4	70	0.38	1.00
	185	1	

$Q_1 = 2 \qquad Q_3 = 4$

Esempio: (distribuzioni di frequenza)

Su un gruppo di scommettitori è stata rilevata la variabile X: *scommesse effettuate nell'ultima settimana*

X	n_i
1	8
2	44
3	63
4	70
	185

calcolare la differenza interquartile

Svolgimento:

X	n_i	f_i	F_i
1	8	0.04	0.04
2	44	0.24	0.28
3	63	0.34	0.62
4	70	0.38	1.00
	185	1	

$$Q_1 = 2 \qquad Q_3 = 4$$

Quindi:

$$DI = Q_3 - Q_1 = 4 - 2 = 2$$

Esempio: (distribuzioni in classi)

Sullo stesso gruppo di scommettitori è stata rilevata la variabile: *vincita ottenuta in €*

X	n_i
(0-50]	10
(50-100]	70
(100-200]	55
(200-500]	50
Totale	185

calcolare la differenza interquartile

Esempio: (distribuzioni in classi)

Sullo stesso gruppo di scommettitori è stata rilevata la variabile: *vincita ottenuta in €*

X	n_i
(0-50]	10
(50-100]	70
(100-200]	55
(200-500]	50
Totale	185

calcolare la differenza interquartile

Svolgimento:

X	n_i	f_i	F_i	a_i
(0-50]	10	0.05	0.05	50
(50-100]	70	0.38	0.43	50
(100-200]	55	0.30	0.73	100
(200-500]	50	0.27	1.00	300
Totale	185	1		

$$Q_1 = 50 + \left(\frac{0,25 - 0,05}{0,43 - 0,05}\right)50 = 76,32$$

$$Q_3 = 200 + \left(\frac{0,75 - 0,73}{1 - 0,73}\right)300 = 222,22$$

Esempio: (distribuzioni in classi)

Sullo stesso gruppo di scommettitori è stata rilevata la variabile: *vincita ottenuta in €*

X	n_i
(0-50]	10
(50-100]	70
(100-200]	55
(200-500]	50
Totale	185

calcolare la differenza interquartile

Svolgimento:

X	n_i	f_i	F_i	a_i
(0-50]	10	0.05	0.05	50
(50-100]	70	0.38	0.43	50
(100-200]	55	0.30	0.73	100
(200-500]	50	0.27	1.00	300
Totale	185	1		

$$Q_1 = 50 + \left(\frac{0,25 - 0,05}{0,43 - 0,05}\right) 50 = 76,32$$

$$Q_3 = 200 + \left(\frac{0,75 - 0,73}{1 - 0,73}\right) 300 = 222,22$$

$$DI = Q_3 - Q_1 = 222,22 - 76,32 = 145,9$$

Indice

Variabilità e variabili quantitative

Campo di variazione

Differenza interquartile

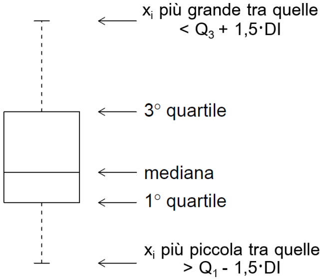
Box plot

Varianza

Box plot

Fornisce una idea schematica di un insieme di dati (di una distribuzione) basata sui quartili.

È costituito da una *scatola* e due *baffi* costruiti come segue



Eventuali osservazioni anomale (outliers) sono disegnate oltre i baffi (ad. esempio utilizzando un pallino).

Esercizio: costruzione di un boxplot

Dati (già ordinati):

1.1 1.3 1.4 1.6 1.8 2.0 2.5 2.6 2.8 4.1 5.6

Esercizio: costruzione di un boxplot

Dati (già ordinati):

$$1.1 \;\; 1.3 \;\; 1.4 \;\; 1.6 \;\; 1.8 \;\; 2.0 \;\; 2.5 \;\; 2.6 \;\; 2.8 \;\; 4.1 \;\; 5.6$$

Perciò

$Q_1 = 1.4, \quad Me = 2, \quad Q_3 = 2.8, \quad DI = 1.4, \quad 1.5DI = 1.5 \times 1.4 = 2.1$

Esercizio: costruzione di un boxplot

Dati (già ordinati):

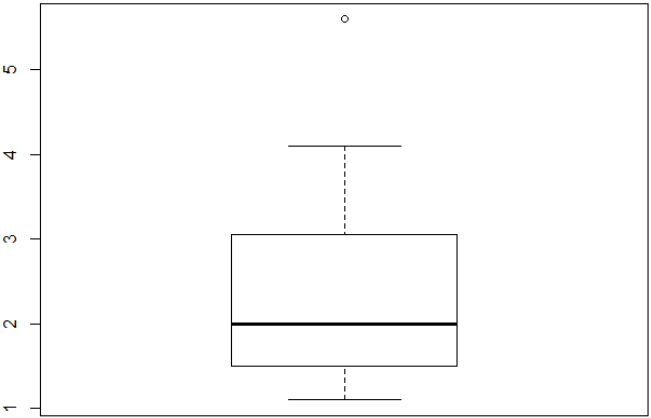
$$1.1 \;\; 1.3 \;\; 1.4 \;\; 1.6 \;\; 1.8 \;\; 2.0 \;\; 2.5 \;\; 2.6 \;\; 2.8 \;\; 4.1 \;\; 5.6$$

Perciò

$$Q_1 = 1.4, \quad Me = 2, \quad Q_3 = 2.8, \quad DI = 1.4, \quad 1.5DI = 1.5 \times 1.4 = 2.1$$

1. *scatola*: da 1.4 a 2.8 con la mediana indicata da una linea a 2;
2. *baffo inferiore*: fino all'osservazione più piccola tra quelle maggiori di $Q_1 - 2.1 = -0.7$, ovvero fino a 1.1;
3. *baffo superiore*: fino all'osservazione più grande tra quelle minori di $Q_3 + 2.1 = 4.9$, ovvero fino a 4.1;
4. vanno disegnate esplicitamente le osservazioni più piccole di 1.1 o più grandi di 4.9 $\Rightarrow$ in questo caso solamente 5.6.

Box plot (esempio precedente)



Esercizio:

Data la seguente distribuzione di frequenze, disegnare il boxplot

X	n_i
1	3
2	8
10	30
20	45
30	22
40	12
50	10
70	5
90	2
100	1
Totale	138

Esercizio:

X	n_i	f_i	F_i
1	3	0.022	0.022
2	8	0.058	0.080
10	30	0.217	0.297
20	45	0.326	0.623
30	22	0.159	0.783
40	12	0.087	0.870
50	10	0.072	0.942
70	5	0.036	0.978
90	2	0.014	0.993
100	1	0.007	1.000
Totale	138	1	

Esercizio:

X	n_i	f_i	F_i
1	3	0.022	0.022
2	8	0.058	0.080
10	30	0.217	0.297
20	45	0.326	0.623
30	22	0.159	0.783
40	12	0.087	0.870
50	10	0.072	0.942
70	5	0.036	0.978
90	2	0.014	0.993
100	1	0.007	1.000
Totale	138	1	

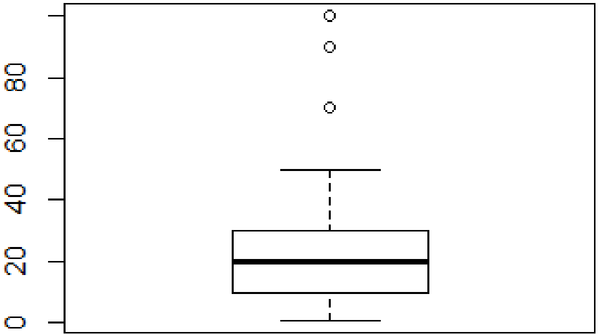
Perciò

$$Q_1 = 10, \quad Me = 20, \quad Q_3 = 30, \quad DI = 20, \quad 1.5DI = 1.5 \times 20 = 30$$

Esercizio:

1. *scatola*: da 10 a 30 con la mediana indicata da una linea a 20;
2. *baffo inferiore*: fino all'osservazione più piccola tra quelle maggiori di $Q_1 - 30 = -20$, ovvero fino a 1;
3. *baffo superiore*: fino all'osservazione più grande tra quelle minori di $Q_3 + 30 = 60$, ovvero fino a 50;
4. vanno disegnate esplicitamente le osservazioni più piccole di -20 o più grandi di 60 $\Rightarrow$ in questo caso $70, 90, 100$.

Box plot



Esercizio 2: distribuzione in classi

Di seguito sono riportate le ore di studio settimanale per un gruppo di studenti di SCO:

X	n_i
$(0; 4]$	10
$(4; 8]$	53
$(8; 12]$	51
$(12; 16]$	30
$(16; 20]$	16
Totale	160

Disegnare il boxplot della distribuzione

X	n_i	f_i	F_i
(0; 4]	10	0,0625	0,0625
(4; 8]	53	0,33125	0,39375
(8; 12]	51	0,31875	0,7125
(12; 16]	30	0,1875	0,9
(16; 20]	16	0,1	1
Totale	160	1	

$$Q_1 = 4 + \left(\frac{0,25 - 0,0625}{0,39375 - 0,0625}\right)4 = 6,26$$

$$Me = 8 + \left(\frac{0,5 - 0,39375}{0,7125 - 0,39375}\right)4 = 9,33$$

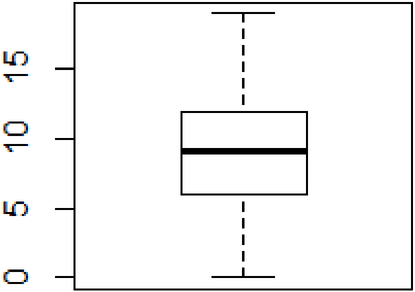
$$Q_3 = 12 + \left(\frac{0,75 - 0,7125}{0,9 - 0,7125}\right)4 = 12,8$$

$$DI = 6,54, \quad 1,5DI = 1,5 \times 6,54 = 9,81$$

Esercizio 2: distribuzione in classi

1. *scatola*: da $6,54$ a $12,8$ con la mediana indicata da una linea a $9,33$;
2. *baffo inferiore*: fino all'osservazione più piccola tra quelle maggiori di $Q_1 - 9,81 = -3,55$, ovvero fino a 0;
3. *baffo superiore*: fino all'osservazione più grande tra quelle minori di $Q_3 + 9,81 = 22,61$, ovvero fino a 20;
4. non ci sono outliers.

Esercizio 2: distribuzione in classi



Box plot per confrontare due distribuzioni

Nella tabella sono riportate le distribuzioni di frequenze del carattere "Numero di figli" per due collettivi di famiglie

	Collettivo A	Collettivo B
X	n_i^A	n_i^B
0	5	10
1	12	16
2	19	36
3	9	10
4	4	17
5	1	11
Totale	50	100

Box plot per confrontare due distribuzioni

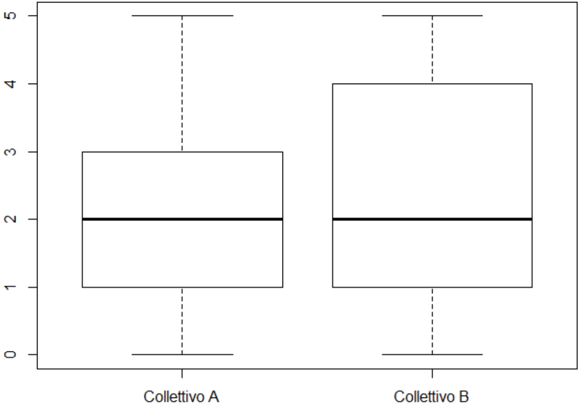
Nella tabella sono riportate le distribuzioni di frequenze del carattere "Numero di figli" per due collettivi di famiglie

	Collettivo A			Collettivo B		
X	n_i^A	f_i^A	F_i^A	n_i^B	f_i^B	F_i^B
0	5	0.1	0.1	10	0.1	0.1
1	12	0.24	0.34	16	0.16	0.26
2	19	0.38	0.72	36	0.36	0.62
3	9	0.18	0.9	10	0.1	0.72
4	4	0.08	0.98	17	0.17	0.89
5	1	0.02	1	11	0.11	1
Totale	50	1		100	1	

$Q_1^A = 1;$ $Me^A = 2;$ $Q_3^A = 3;$ $DI^A = 2$

$Q_1^B = 1;$ $Me^B = 2;$ $Q_3^B = 4;$ $DI^B = 3$

Box plot per confrontare due distribuzioni



Indice

Variabilità e variabili quantitative
- Campo di variazione
- Differenza interquartile
- Box plot
- Varianza

Ancora sulla variabilità

Abbiamo detto che per misurare la variabilità, possiamo utilizzare la "distanza" delle osservazioni dal centro della distribuzione.

Ancora sulla variabilità

Abbiamo detto che per misurare la variabilità, possiamo utilizzare la "distanza" delle osservazioni dal centro della distribuzione.

Proviamo a utilizzare la media per caratterizzare il centro della distribuzione.

Ancora sulla variabilità (cont)

Consideriamo lo scarto al quadrato:

$$(x_i - M)^2.$$

Perché il quadrato?

Ancora sulla variabilità (cont)

Consideriamo lo scarto al quadrato:

$$(x_i - M)^2.$$

Perché il quadrato?

Perché il quadrato “amplifica” le distanze grandi e “attenua” quelle piccole.

Esempio: $10^2 = 100, \qquad 0.1^2 = 0.01.$

Ancora sulla variabilità (cont)

Consideriamo lo scarto al quadrato:

$$(x_i - M)^2.$$

Perché il quadrato?

Perché il quadrato "amplifica" le distanze grandi e "attenua" quelle piccole.

Esempio: $10^2 = 100, \qquad 0.1^2 = 0.01.$

Quindi, per costruire un indice di variabilità, possiamo costruire queste n quantità (per $i = 1, \ldots, n$) e farne una media.

Varianza

La *varianza* è la media dei quadrati degli scarti di ogni osservazione dalla media aritmetica.

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - M)^2$$

Varianza

La *varianza* è la media dei quadrati degli scarti di ogni osservazione dalla media aritmetica.

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - M)^2$$

Esempio: varianza per 5 osservazioni, la media aritmetica è $M = 2.8$

Osservazioni x_i	scarti $x_i - M$	$(\text{scarti})^2$ $(x_i - M)^2$
-1	-3.80	14.44
1	-1.80	3.24
3	0.20	0.04
4	1.20	1.44
7	4.20	17.64
	Totale	36.8

La varianza è

$$\sigma^2 = \frac{36.8}{5} = 7.36$$

Varianza (formula con dist. freq.)

La *varianza* è la media dei quadrati degli scarti di ogni osservazione dalla media aritmetica per la frequenza.

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{K}(x_i - M)^2 n_i$$

Varianza

La *varianza* è la media dei quadrati degli scarti di ogni osservazione dalla media aritmetica per la frequenza.

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^{K} (x_i - M)^2 n_i$$

Esempio: ore di sonno per notte, le osservazioni sono $n = 80$, la media aritmetica è $M = 7.4$

Modalità x_i	Frequenza n_i	scarti $x_i - M$	$(\text{scarti})^2$ $(x_i - M)^2$	scarti pesati $(x_i - M)^2 n_i$
5	1	-2.40	5.76	5.76
6	10	-1.40	1.96	19.6
7	31	-0.40	0.16	4.96
8	32	0.60	0.36	11.52
9	6	1.60	2.56	15.36
			Totale	57.2

La varianza è

$$\sigma^2 = \frac{57.2}{80} = 0.715$$

Varianza: distribuzione di frequenza per classi

Per il calcolo della varianza, possiamo fare ricorso alla formula operativa utilizzando la stessa strategia adottata per il calcolo della media da distribuzioni di frequenza, ovvero utilizzando il **valore centrale** di ogni classe per rappresentare i valori della classe stessa.

Varianza: distribuzione di frequenza per classi

Per il calcolo della varianza, possiamo fare ricorso alla formula operativa utilizzando la stessa strategia adottata per il calcolo della media da distribuzioni di frequenza, ovvero utilizzando il **valore centrale** di ogni classe per rappresentare i valori della classe stessa.

La *varianza* è la media dei quadrati degli scarti di ogni osservazione dalla media aritmetica.

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^{K} (c_i - M)^2 n_i$$

Esempio: distribuzione in classi

Di seguito sono riportate le ore di studio settimanale per un gruppo di studenti di SCO, la media aritmetica è M=9.725

X	n_i
(0; 4]	10
(4; 8]	53
(8; 12]	51
(12; 16]	30
(16; 20]	16
Totale	160

Esempio: distribuzione in classi

Di seguito sono riportate le ore di studio settimanale per un gruppo di studenti di SCO, la media aritmetica è M=9.725

X	n_i	c_i	$(c_i - M)^2$	$(c_i - M)^2 n_i$
(0; 4]	10	2	59.67563	596.7563
(4; 8]	53	6	13.87563	735.4081
(8; 12]	51	10	0.075625	3.856875
(12; 16]	30	14	18.27563	548.2688
(16; 20]	16	18	68.47563	1095.61
Totale	160			2979.9

La varianza è

$$\sigma^2 = \frac{2979.9}{160} = 18.624$$

Esempio: Voto di maturità

Di seguito sono riportati i voti di maturità per un gruppo di studenti, la media aritmetica è M=77.875

$(x_{i-1}, x_i]$	n_i	c_i	$c_i n_i$	$(c_i - M)^2$	$(c_i - M)^2 n_i$
[60,70]	15	65	975	165.7656	2486.484
(70,80]	35	75	2625	8.265625	289.2969
(80,90]	22	85	1870	50.76563	1116.844
(90,100]	8	95	760	293.2656	2346.125
Totale	80		6230		6238.75

Esempio: Voto di maturità

Di seguito sono riportati i voti di maturità per un gruppo di studenti, la media aritmetica è M=77.875

$(x_{i-1}, x_i]$	n_i	c_i	$c_i n_i$	$(c_i - M)^2$	$(c_i - M)^2 n_i$
[60,70]	15	65	975	165.7656	2486.484
(70,80]	35	75	2625	8.265625	289.2969
(80,90]	22	85	1870	50.76563	1116.844
(90,100]	8	95	760	293.2656	2346.125
Totale	80		6230		6238.75

La varianza è

$$\sigma^2 = \frac{6238.75}{80} = 77.98$$

Osservazioni sulla varianza

La varianza presenta alcune peculiarità:

- dipende da tutte le modalità del carattere;

Osservazioni sulla varianza

La varianza presenta alcune peculiarità:

- dipende da tutte le modalità del carattere;
- assume solo valori non negativi;

Osservazioni sulla varianza

La varianza presenta alcune peculiarità:

- dipende da tutte le modalità del carattere;
- assume solo valori non negativi;
- è nulla se e solo se il carattere è degenere;

Osservazioni sulla varianza

La varianza presenta alcune peculiarità:

- dipende da tutte le modalità del carattere;
- assume solo valori non negativi;
- è nulla se e solo se il carattere è degenere;
- è sensibile agli outliers;

Osservazioni sulla varianza

La varianza presenta alcune peculiarità:

- dipende da tutte le modalità del carattere;
- assume solo valori non negativi;
- è nulla se e solo se il carattere è degenere;
- è sensibile agli outliers;
- è espressa nel quadrato dell'unità di misura del carattere investigato.

Quest'ultima caratteristica rende non immediata la sua interpretazione

Varianza: una formula alternativa

È possibile dimostrare che una formula alternativa (di più facile utilizzo) per il calcolo della varianza è data da

$$\sigma^2 = \left(\frac{1}{n} \sum_{i=1}^{n} x_i^2 \right) - M^2$$

ovvero

$$(\text{varianza}) = \left(\begin{array}{c} \text{media dei} \\ \text{quadrati} \end{array} \right) - \left(\begin{array}{c} \text{quadrato della} \\ \text{media} \end{array} \right).$$

Formula alternativa: esempio di utilizzo

Esempio: varianza per 5 osservazioni, la media è $M = 2.8$

Osservazioni x_i	scarti $x_i - M$	(scarti)^2 $(x_i - M)^2$	(osservazioni)^2 x_i^2
-1	-3.80	14.44	1
1	-1.80	3.24	1
3	0.20	0.04	9
4	1.20	1.44	16
7	4.20	17.64	49
	Totale	36.8	76

La varianza con la "vecchia" formula è

$$\sigma^2 = \frac{36.8}{5} = 7.36$$

La varianza con la "nuova" formula è

$$\sigma^2 = \frac{76}{5} - 2.8^2 = 7.36$$