5.2.2 La mediana

Se il carattere è per lo meno ordinato rettilineo, si può sfruttare questa sua caratteristica per dare un ordine crescente alle unità e porsi poi la domanda: "qual è la modalità dell'unità centrale, se n è dispari, o di ciascuna delle due unità centrali, se n è pari?". Per n dispari tale modalità è la mediana (m). E' evidente che, sempre per n dispari e con tutte le unità diverse fra loro, nella distribuzione (n-1)/2 unità sono minori della mediana, (n-1)/2 sono maggiori e che la mediana occupa il posto (n+1)/2. Se fosse lecito dividere una unità statistica a metà, si potrebbe dire che la mediana è quella modalità che consente di bipartire la distribuzione ordinata. Quando n è pari, vi sono due unità centrali in corrispondenza ai posti n/2 e (n/2)+1 e di conseguenza due modalità mediane. Se tali due modalità sono uguali si parla nuovamente di mediana (m). Lasciando le due unità una a sinistra e l'altra a destra della distribuzione ordinata, essa viene bipartita.

Si osserva che la mediana può essere individuata rispetto alla distribuzione semplice sia unitaria, sia di frequenze e la sua determinazione presuppone l'ordinamento, di solito crescente, delle unità rispetto alle modalità del carattere.

Il concetto di mediana è relativamente semplice, meno semplice ed immediato è l'individuazione o il calcolo della mediana nelle distribuzioni di frequenze, che dipende dalla natura del carattere e dal modo in cui sono espresse le modalità, oltreché dall'essere n pari o dispari.

a) Carattere ordinato

Nella distribuzione della tabella 1 l'individuazione della mediana richiede alcune elaborazioni.

Tab. 1 - Studenti presenti alla lezione di statistica del 6-10-1994
per titolo di studio del padre

Titolo di studio	N. studenti
Scuola elementare	10
Scuola media inferiore	29
Scuola media superiore	48
Laurea	22
Totale	109

Se si legge la tabella tenendo conto dell'ordinamento degli studenti (le unità) in base al titolo di studio del padre (carattere) si può dire che le prime 10 unità presentano come titolo di studio la licenza elementare, le 29 unità successive - ossia dalla 11° alla 39° - hanno modalità scuola media inferiore; le 48 successive - ossia dalla 40° alla 77° unità - presentano modalità scuola media superiore ed infine le ultime 22 - ossia dal 78° alla 109° - la modalità laurea. Seguendo la definizione, per n dispari la mediana è la modalità dell'unità statistica di posto centrale. In questo caso il posto centrale è il 55° - (109+1)/2 - al quale corrisponde la modalità scuola media superiore, che è la mediana. Senza mostrare la distribuzione, potremmo dire che: al minimo il 50% del collettivo degli studenti ha un padre con titolo di studio non superiori alla scuola media superiore, o anche che meno del 50% del collettivo ha titolo di studio superiore alla scuola media superiore.

Il percorso logico esposto può essere formalizzato introducendo il concetto di frequenza assoluta cumulata N_t con $N_t = \sum_{l=1}^{t} n_l$. N_t é la frequenza delle unità con modalità $\leq x_t$. La mediana è quella modalità x_t per la quale è $N_{t-1} < n/2 < N_t$. Se N_t = n/2, la mediana non è unica ed assume le due modalità x_t e x_{t+1} . In questo caso, infatti, essendo n pari, poiché N_t è intero, vi sono due modalità mediane, da indicare entrambe, a meno che non siano uguali.

Per individuare la mediana, si costruisce la tabella 2.

Tab. 2 - Studenti presenti alla lezione di statistica del 6-10-1994
per titolo di studio del padre
(frequenze assolute e assolute cumulate)

Titolo di studio	N. studenti (n_t)	Frequenze assolute cumulate (N_t)
Scuola elementare	10	10
Scuola media inferiore	29	39
Scuola media superiore	48	87
Laurea	22	109
Totale	109	

Poiché 39<109/2<87, ossia $N_2 < 109/2 < N_3$, si ha m = scuola media superiore, risultato a cui si era già pervenuti.

Si esamini ora la tabella 3 che contiene una distribuzione inventata delle frequenze assolute cumulate rispetto ad un carattere ordinato con 5 modalità.

Tab. 3 - Distribuzione rispetto al carattere X
(frequenze assolute cumulate)

Modalità	N_t
I	10
II	50
III	60
IV	100
V	120

In questo caso è: 50<120/2 = 60. Poiché 60 = N_3, la prima modalità della mediana è la modalità III, e la seconda modalità è IV.

b) Carattere quantitativo discreto

L'individuazione della mediana della distribuzione degli studenti per voto alla maturità (Tab. 4) può essere condotta rapidamente seguendo quanto su esposto. La tab. 5 contiene la distribuzione delle frequenze assolute cumulate.

Tab. 4 - Studenti presenti alla lezione di statistica del 6-10-1994
per voto alla maturità

	Voto alla maturità																						
	36	37	38	40	41	42	43	44	45	46	48	50	51	52	53	54	55	56	57	58	60	n.i	Tot.
N. studenti	5	1	8	8	2	12	1	8	2	3	6	8	1	8	1	7	1	3	1	5	17	1	109

Tab. 5 - Studenti presenti alla lezione di statistica del 6-10-1994
per voto alla maturità
(frequenze assolute cumulate)

	Voto alla maturità																				
	36	37	38	40	41	42	43	44	45	46	48	50	51	52	53	54	55	56	57	58	60
N_t	5	6	14	22	24	36	37	45	47	50	56	64	65	73	74	81	82	85	86	91	108

Poiché si ha 50<108/2<56, ossia N_{10} <108/2< N_{11} , e ad N_{11} corrisponde il voto 48, segue che m=48. Senza mostrare tutta la distribuzione, si può dire che gli studenti con voto non superiore a 48 sono come minimo il 50%.

Quando il carattere è quantitativo discreto ed n è pari, se $N_t = n/2$, qualsiasi valore compreso fra x_t e x_{t+1} occupa una posizione centrale nella distribuzione ordinata, in tale caso, per convenzione si assume $m = (x_t + x_{t+1})/2$. Si consideri ad esempio la distribuzione cumulativa di frequenze contenuta nella tabella 6 e relativa ad un collettivo di 40 studenti secondo il voto alla maturità.

Tab. 6 - Studenti per voto alla maturità
(frequenze assolute cumulate)

Voto alla maturità	N_t
36	10
40	15
45	20
54	38
60	40

E' n/2=20=N_3, pertanto esistono due modalità mediane, una in corrispondenza a 45 e l'altra in corrispondenza a 54, per convenzione si ha allora m = (45+54)/2 = 49,5. In questo caso esattamente il 50% degli studenti ha un voto non superiore a 49,5 e il 50% non inferiore a 49,5.

c) Carattere quantitativo continuo

Quando il carattere è continuo la mediana è unica ed è individuabile in modo univoco. Se si vuole calcolare la mediana rispetto alla distribuzione delle femmine secondo la statura (Tab.7) è opportuno innanzitutto fare la rappresentazione grafica della distribuzione (Fig. 1).

Tab. 7 - Studenti presenti alla lezione di statistica del 6-10-1994
classificati per sesso e classe di statura

	Sesso		
Classe di statura (cm)	Maschi	Femmine	Totale
153,5 ⊣ 156,5	0	3	3
156,5 ⊣ 159,5	0	2	2
159,5 ⊣ 162,5	0	12	12
162,5 ⊣ 165,5	0	14	14
165,5 ⊣ 168,5	2	6	8
168,5 ⊣ 171,5	4	8	12
171,5 ⊣ 174,5	5	5	10
174,5 ⊣ 177,5	15	2	17
177,5 ⊣ 180,5	14	2	16
180,5 ⊣ 183,5	7	0	7
183,5 ⊣ 186,5	2	0	2
186,5 ⊣ 189,5	2	0	2
189,5 ⊣ 192,5	4	0	4
Totale	55	54	109

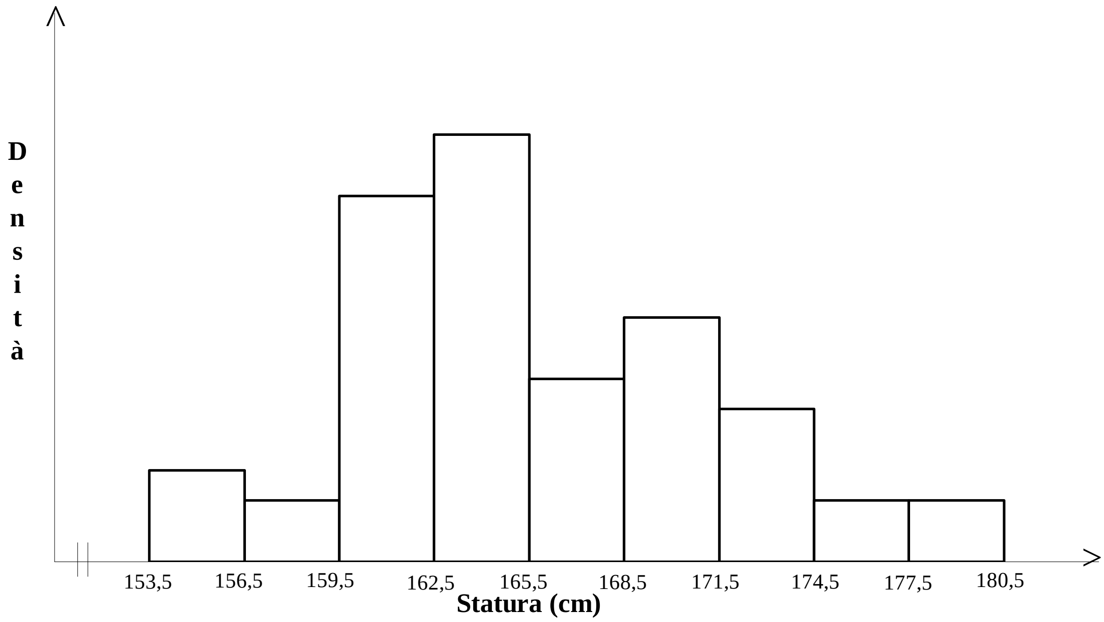

Fig. 1 - Distribuzione delle studentesse per classe di statura

La mediana, geometricamente, è quel valore m sull’asse x che consente, tracciando la retta x=m, di bipartire l’area sottesa dalla poligonale. Coi dati a disposizione è possibile individuare l’intervallo che comprende la mediana, a tal fine si costruisce la distribuzione delle frequenze cumulate (Tab. 8).

Tab. 8 - Studentesse presenti alla lezione di statistica del 6-10-1994
per classe di statura
(frequenze assolute cumulate)

Classe di statura (cm)	n_t	N_t
153,5 $\dashv$ 156,5	3	3
156,5 $\dashv$ 159,5	2	5
159,5 $\dashv$ 162,5	12	17
162,5 $\dashv$ 165,5	14	31
165,5 $\dashv$ 168,5	6	37
168,5 $\dashv$ 171,5	8	45
171,5 $\dashv$ 174,5	5	50
174,5 $\dashv$ 177,5	2	52
177,5 $\dashv$ 180,5	2	54
Totale	54	

Poiché si ha 17<54/2<31 ne consegue che 162,5<m<165,5. Per determinare m occorre trovare la quantità da aggiungere a 162,5 per far sì che l'area sottesa dalla poligonale, per x che va da 153,5 a m sia esattamente n/2=27. Si osserva che l'area per 153,5≤x≤162,5 è nota ed è 17, perciò l'area mancante per arrivare a 27 è 10=27-17. Di tale area è necessario determinare la base, dato che la sua altezza è la densità di frequenza della quarta classe, ossia 14/3=4,67. La base che si cerca è dunque 10/4,67=2,14 e quindi m = cm (162,5 +2,14) = cm 164,64.

Il percorso logico compiuto può essere formalizzato. Indicando con ${}_mc_{t-1} \dashv {}_mc_t$ la classe che contiene m, con ${}_mn_t$ la corrispondente frequenza assoluta e con ${}_mN_{t-1}$ l'accumulo della classe che precede quella contenente la mediana, si ha:

$$m={}_mc_{t-1}+\frac{\frac{n}{2}-{}_mN_{t-1}}{\frac{{}_mn_t}{{}_mc_t-{}_mc_{t-1}}} = {}_mc_{t-1}+\frac{\frac{n}{2}-{}_mN_{t-1}}{{}_mn_t}\left({}_mc_t-{}_mc_{t-1}\right) \qquad \text{(i)}$$

Applicando la (i) si ottiene immediatamente

$$m = \text{cm}\left(162{,}5 - \frac{27-17}{14}3\right) = \text{cm } 164{,}64$$

Applicando la formula alla distribuzione della statura dei maschi (Tab.7), si ottiene rapidamente che 177,5<m_M<180,5, poiché 26<55/2<40 e quindi

$$m_M = \text{cm}\left(177{,}5+\frac{\frac{55}{2}-26}{14}3\right) = \text{cm}\left(177{,}5+0{,}32\right) = \text{cm } 177{,}82$$

5.2.3. I quantili

Se anziché cercare la modalità dell'unità statistica che bipartisce la distribuzione ordinata, si cerca, più in generale, la modalità che lascia alla sua destra np unità con 0≤p≤1 e alla sua sinistra n(1-p) unità, si individuano i quantili che in generale si possono indicare con $x_{(p)}$.

Ponendo p=1/2=0,5 si ritrova la mediana: $x_{(0,5)}$ = m.

Ponendo p_1 =1/4; p_2 =2/4; p_3 =3/4, si hanno rispettivamente il 1° quartile: $x_{(0,25)} = Q_1$, il 2° quartile: $x_{(0,5)}$ = m, e il 3° quartile: $x_{(0,75)} = Q_3$, dove la lettera Q è l'iniziale della parola quartile.

Ponendo p_1 = 0,10; p_2 = 0,20;…,p_9 = 0,90 si individuano i decili: $x_{(0,10)}$, $x_{(0,20)}$, … ,$x_{(0,90)}$.

L'individuazione dei quantili segue il percorso logico già visto per la mediana, utilizzando il valore p opportuno.

a) e b) Carattere ordinato e quantitativo discreto

Rispetto alla distribuzione degli studenti secondo il titolo di studio del padre (tab. 2) per calcolare Q_1 si considera che 10<109/4<39, perciò Q_1 = scuola media inferiore. Per calcolare Q_3, si tiene conto della diseguaglianza 39<109/4×3<87 e perciò Q_3 = scuola media superiore.

Rispetto ai voti alla maturità (tab. 5) si ha che Q_1= 42, perché 24<108/4<36; mentre $Q_3 = \frac{54+55}{2} = 54,5$ perché (108/4)×3=81 e quindi Q_3 può assumere qualsiasi valore tra 54 e 55.

Il primo decile nella stessa distribuzione è 38, poiché 6<108/10<14. Il 9° decile è 60, perché 91<(108/10)×9<108.

c) Carattere quantitativo continuo

Rispetto alla statura delle studentesse (tab. 8), il calcolo di Q_1 utilizza la (i), tenendo conto che il p di interesse è ora 1/4. Perciò si ha 159,5<Q_1<162,5.

$$Q_1 = \text{ cm } \left(159{,}5 + \frac{\frac{54}{4} - 5}{12}\, 3\right) = \text{cm} \left(159{,}5 + \frac{13{,}5 - 5}{12}\, 3\right) = \text{cm} \left(159{,}5 + 2{,}12\right) = \text{cm } 161{,}62\ .$$

Analogamente per Q_3 si ha:

$$Q_3 = \text{ cm } \left(168{,}5 + \frac{40{,}5 - 37}{8}\, 3\right) = \text{cm} \left(168{,}5 + 1{,}32\right) = \text{cm } 169{,}82\ .$$

BIBLIOGRAFIA

M. FRAIRE (1994), *Metodi di analisi multidimensionale dei dati. Aspetti statistici ed applicazioni informatiche*, CISU, Roma.

G. LETI (1983), *Statistica descrittiva*, il Mulino, Bologna, 1983.

M. G. OTTAVIANI (1993), I libri di testo e la statistica: metodi di analisi e confronto di testi. *INDUZIONI*, **6**, pp.93-120.