La misura della forma per variabili qualitative ordinali

Lo studio della forma per variabili qualitative ordinali può essere condotto servendosi dell'indice di dispersione introdotto per tale categoria di variabili.

Nel caso di una variabile ordinale ha infatti senso studiare la disposizione delle frequenze rispetto alle modalità ordinate e chiedersi se tali frequenze si distribuiscono in maniera simmetrica. In particolare, nel caso di k pari una distribuzione simmetrica dovrebbe essere caratterizzata da un andamento perfettamente speculare delle frequenze sulle prime $\frac{k}{2}$ modalità e sulle ultime $\frac{k}{2}$ modalità della variabile. Nel caso di k dispari, tale simmetria può essere individuata considerando le frequenze corrispondenti alle $\frac{k-1}{2}$ modalità che si trovano a sinistra della modalità centrale e le frequenze corrispondenti alle $\frac{k-1}{2}$ modalità che si trovano a destra della modalità centrale.

E' possibile pertanto definire una distribuzione simmetrica se:

$$n_1 = n_k, \quad n_2 = n_{k-1}, \quad n_3 = n_{k-3}, \ldots$$

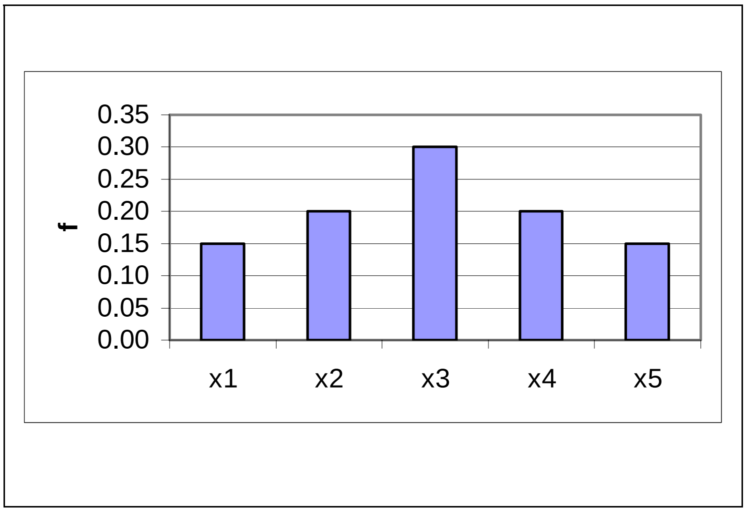
Considerando la generica i-esima modalità, si avrà:

$$n_i = n_{k-i+1}$$

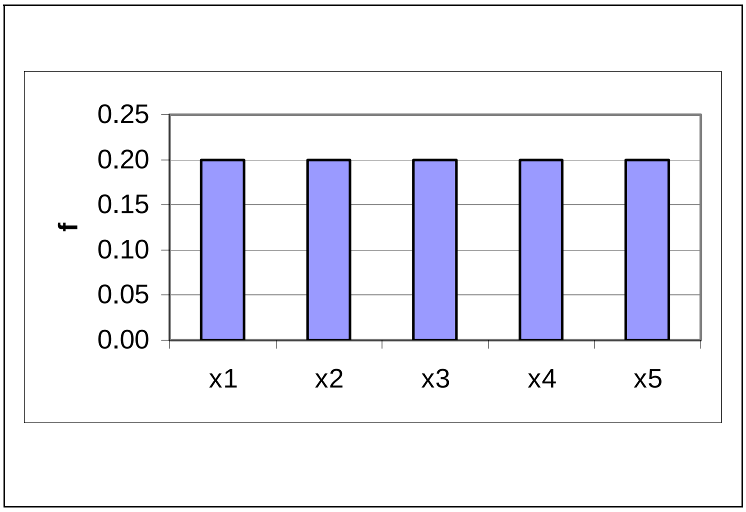
facendo variare i fino a $\frac{k}{2}$ se k è pari e fino a $\frac{k-1}{2}$ se k è dispari.

Di seguito si riportano alcuni esempi di distribuzione simmetrica in forma tabellare e grafica facendo riferimento ad una variabile qualitativa ordinale che assume 5 modalità.

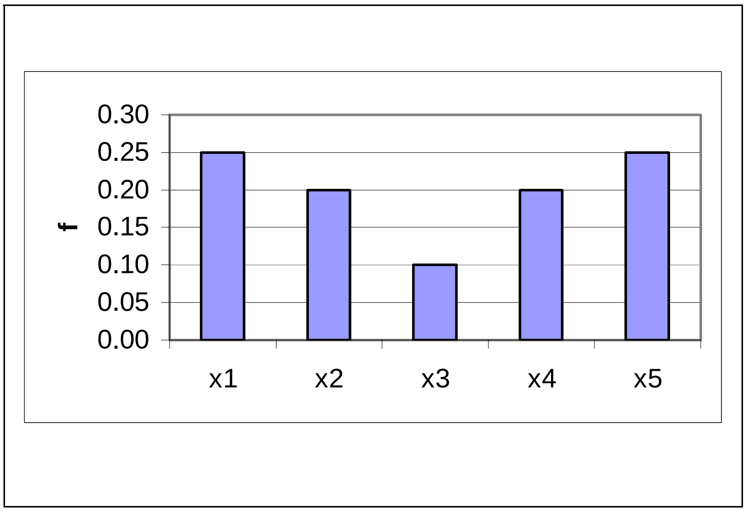
CASO 1		
X	n	f
$x_{(1)}$	15	0.15
$x_{(2)}$	20	0.20
$x_{(3)}$	30	0.30
$x_{(4)}$	20	0.20
$x_{(5)}$	15	0.15
	100	1.00



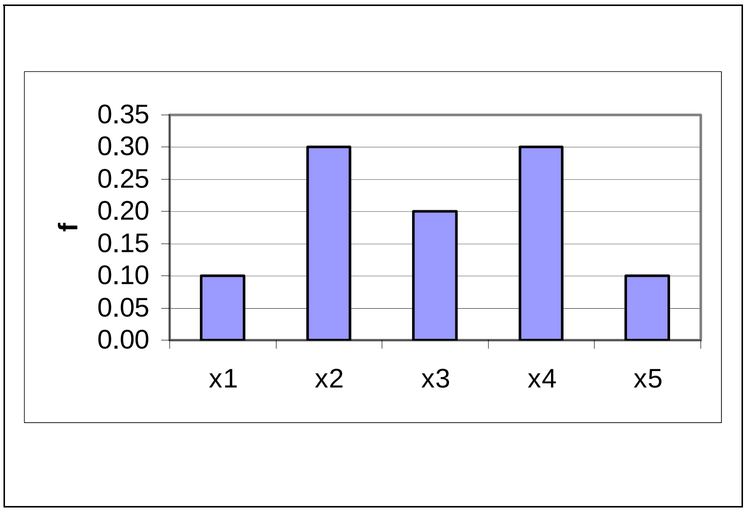
CASO 2		
X	n	f
$x_{(1)}$	20	0.20
$x_{(2)}$	20	0.20
$x_{(3)}$	20	0.20
$x_{(4)}$	20	0.20
$x_{(5)}$	20	0.20
	100	1.00



CASO 3		
X	n	f
$x_{(1)}$	25	0.25
$x_{(2)}$	20	0.20
$x_{(3)}$	10	0.10
$x_{(4)}$	20	0.20
$x_{(5)}$	25	0.25
	100	1.00



CASO 4		
X	n	f
$x_{(1)}$	10	0.10
$x_{(2)}$	30	0.30
$x_{(3)}$	20	0.20
$x_{(4)}$	30	0.30
$x_{(5)}$	10	0.10
	100	1.00

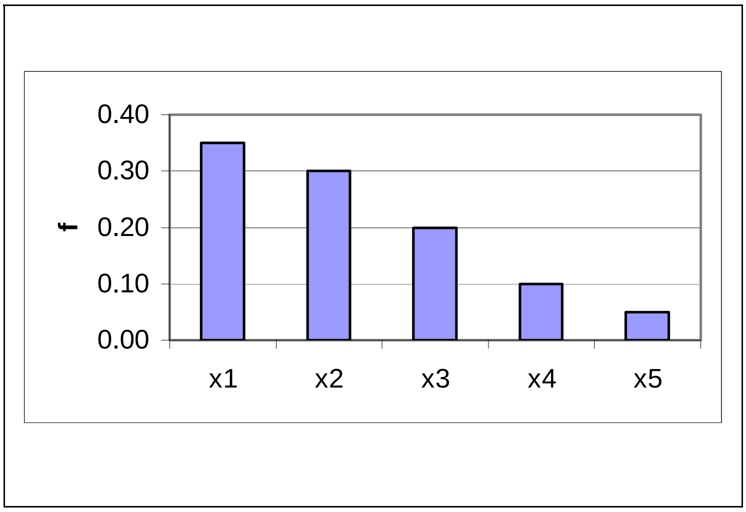


I quattro casi riportati, sebbene caratterizzati da una forma diversa l'uno dall'altro, come del resto è evidente dal confronto dei rispettivi grafici a barra, soddisfano tutti la condizione di simmetria fornita in precedenza. In particolare i casi si distinguono per una maggiore frequenza sulla modalità centrale e una frequenza decrescente sulla modalità esterna (caso 1); una frequenza uniforme per tutte le modalità della variabile (caso 2); una frequenza maggiore sulle modalità esterne e via via decrescente quando si va verso la modalità centrale (caso 3) ed infine una disposizione altalenante delle frequenze rispetto all'ordinamento delle modalità (caso 4). Si tratta naturalmente solo di alcune delle possibili distribuzioni che soddisfano la condizione di simmetria.

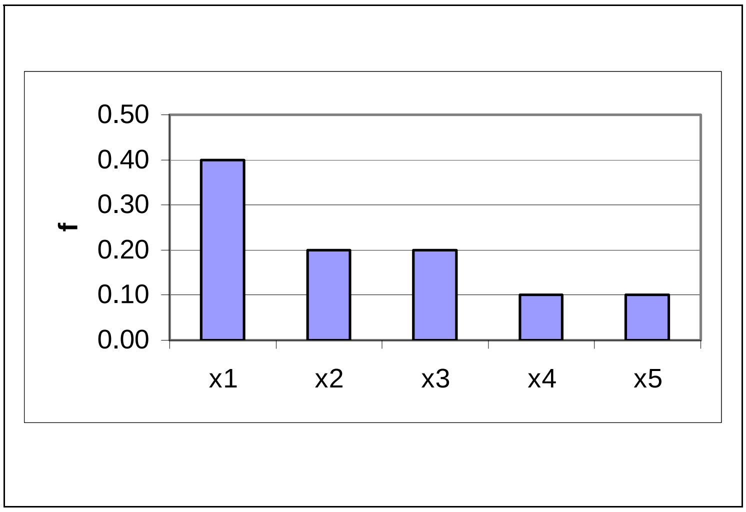
Se tale condizione non è soddisfatta, si dice allora che la distribuzione è asimmetrica, distinguendo i due casi di asimmetria positiva (o destra) e di asimmetria negativa (o sinistra). Si parla di asimmetria positiva se la distribuzione è caratterizzata da un maggior numero di unità che presentano le modalità più basse e, viceversa, di asimmetria negativa se sono le modalità più alte ad essere più ricorrenti.

Di seguito sono riportati tre casi di distribuzioni caratterizzate da asimmetria positiva.

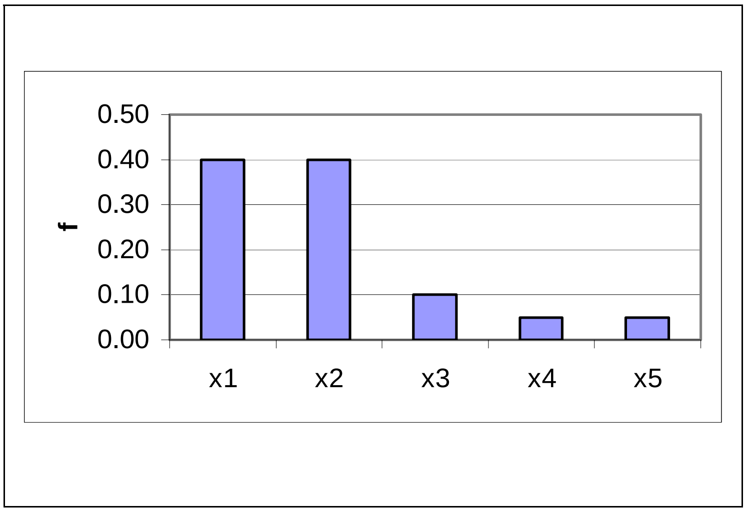
CASO 5		
X	n	f
$x_{(1)}$	35	0.35
$x_{(2)}$	30	0.30
$x_{(3)}$	20	0.20
$x_{(4)}$	10	0.10
$x_{(5)}$	5	0.05
	100	1.00



CASO 6		
X	n	f
$x_{(1)}$	40	0.40
$x_{(2)}$	20	0.20
$x_{(3)}$	20	0.20
$x_{(4)}$	10	0.10
$x_{(5)}$	10	0.10
	100	1.00



CASO 7		
X	*n*	*f*
$x_{(1)}$	40	0.40
$x_{(2)}$	40	0.40
$x_{(3)}$	10	0.10
$x_{(4)}$	5	0.10
$x_{(5)}$	5	0.05
	100	1.00



Le tre distribuzioni, infatti, sono caratterizzate da un maggior addensamento delle unità sulle modalità che occupano il posto più basso nella distribuzione pur differenziandosi per una diversa disposizione delle frequenze.

Il caso di distribuzioni con asimmetria negativa si può ottenere semplicemente invertendo le frequenze nelle tre distribuzioni appena riportate.

Un indice di forma, a differenza di un indice di variabilità, deve essere in grado di segnalare queste tre situazioni tipo, vale a dire distribuzione simmetrica, distribuzione caratterizzata da asimmetria positiva e distribuzione caratterizzata da asimmetria negativa. Per tale motivo gli indici di forma possono assumere valori positivi e negativi, segnalando, rispettivamente, la presenza di asimmetria negativa e di asimmetria positiva. Un indice pari a 0 dovrebbe segnalare una distribuzione simmetrica: tale condizione è però necessaria e non sufficiente in quanto uno dei difetti degli indici di forma è che questi possono valere 0 anche nel caso in cui la distribuzione non è perfettamente simmetrica. Per evitare conclusioni errate è quindi consigliabile supportare sempre la lettura di tali indici con l'esame della rappresentazione grafica, da cui si riesce ad evincere velocemente la presenza di eventuali situazione anomale. Naturalmente, oltre al segno dell'indice, sarà possibile interpretare la grandezza dello stesso: un indice maggiore (minore) indicherà un grado di asimmetria positiva (negativa) più accentuato.

Come detto in precedenza, un indice di forma per variabili qualitative ordinali può essere costruito servendosi dell'indice di dispersione: in particolare è possibile dividere la distribuzione in due parti di pari numerosità, calcolare per ciascuna delle due sottodistribuzioni ottenute l'indice di dispersione e confrontare i due valori. Se la distribuzione di sinistra è caratterizzata da un valore più alto dell'indice di dispersione è possibile aspettarsi un'asimmetria negativa e viceversa. Quanto maggiore sarà il divario tra i due indici tanto più accentuata sarà la presenza di asimmetria. Nel caso in cui i due indici di dispersione per le due distribuzioni siano uguali è possibile che la distribuzione sia simmetrica pur non essendo questo necessariamente vero (i due indici possono annullarsi, come detto in precedenza, anche nel caso di distribuzioni non simmetriche).

La divisione della distribuzione in due parti può essere fatta distinguendo il caso di N pari e di N dispari:

N pari	→	la distribuzione di sinistra è formata dalle prime $\frac{N}{2}$ unità e la distribuzione di destra dalle ultime $\frac{N}{2}$ unità,
N dispari	→	la distribuzione di sinistra è formata dalle prime $\frac{N+1}{2}$ unità e la distribuzione di destra dalle ultime $\frac{N+1}{2}$ unità.

Nel caso di N dispari, per avere due distribuzioni di uguale numerosità senza dover rinunciare a nessuna delle unità rilevate, si considera due volte l'unità centrale (quella che occupa il posto $\frac{N+1}{2}$), ovvero la si fa rientrare sia nella distribuzione di sinistra che in quella di destra.

Un indice assoluto di forma, per quanto detto in precedenza, si ottiene come:

$$D_d - D_s$$

dove con D_s si è indicato l'indice di dispersione relativo alla prima metà della distribuzione (distribuzione di sinistra) e con D_d si è indicato lo stesso indice per la seconda metà (distribuzione di destra).

Per costruire un indice relativo è possibile ricorrere al seguente indice:

$$A = \frac{D_d - D_s}{D_d + D_s}$$

ottenuto dividendo l'indice assoluto per la somma dei due indici di dispersione.

Tale indice è normalizzato in [-1, +1]: assume il valore -1 nel caso di massima asimmetria negativa, il valore +1 nel caso di massima asimmetria positiva. Il valore 0 dovrebbe segnalare una presenza di simmetria nella distribuzione, pur essendo possibile che tale valore si presenti anche nel caso di distribuzioni non perfettamente simmetriche.

In particolare, l'indice A assume il valore -1 nel caso in cui $D_d = 0$, ovvero nel caso in cui la dispersione della seconda metà della distribuzione è nulla (ovvero se tutte le unità si concentrano in una sola modalità). In tale caso si ha infatti:

$$A = \frac{0 - D_s}{0 + D_s} = \frac{-D_s}{+D_s} = -1$$

Analogamente nel caso in cui tutte le unità della prima metà della distribuzione sono concentrate in una sola modalità della variabile, si ha $D_s = 0$, da cui:

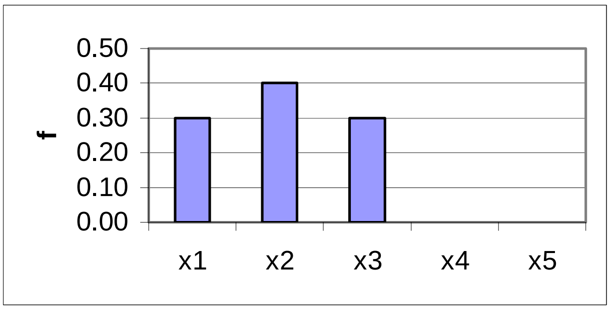
$$A = \frac{D_d - 0}{D_d + 0} = \frac{D_d}{D_d} = 1$$

Si riportano, di seguito, per ciascuna delle 7 distribuzioni di esempio riportate in precedenza, i calcoli necessari per la costruzione dell'indice A.

CASO 1

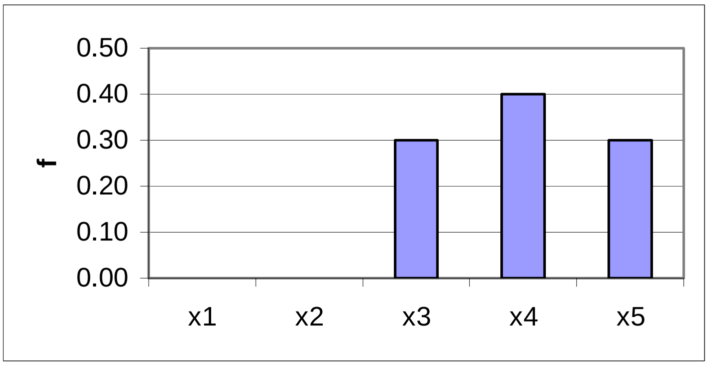
PRIMA METÀ DELLA DISTRIBUZIONE:

X	n	N	f	F	1 - F	F * (1 - F)
x1	15	15	0.30	0.30	0.70	0.21
x2	20	35	0.40	0.70	0.30	0.21
x3	15	50	0.30	1.00	0.00	0.00
x4	0	50	0.00	1.00	0.00	0.00
x5	0	50	0.00	1.00	0.00	0.00
	50		1.00			0.42
						0.84 Ds



SECONDA METÀ DELLA DISTRIBUZIONE:

X	n	N	f	F	1 - F	F * (1 - F)
x1	0	0	0.00	0.00	1.00	0.00
x2	0	0	0.00	0.00	1.00	0.00
x3	15	15	0.30	0.30	0.70	0.21
x4	20	35	0.40	0.70	0.30	0.21
x5	15	50	0.30	1.00	0.00	0.00
	50		1.00			0.42
						0.84 Dd

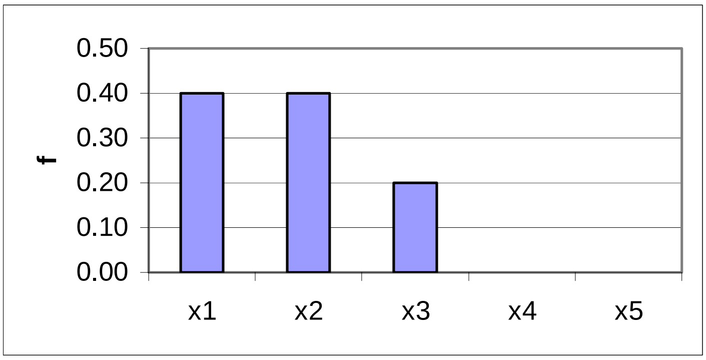


Essendo uguali i due indici di dispersione, sia l'indice assoluto che l'indice relativo di forma si annullano segnalando la presenza di una possibile simmetria, confermata del resto dall'esame del diagramma a barre relativo all'intera distribuzione.

CASO 2

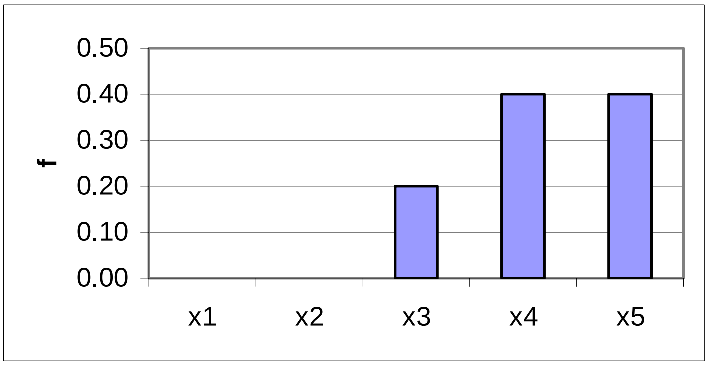
PRIMA METÀ DELLA DISTRIBUZIONE:

X	n	N	f	F	1 - F	F * (1 - F)	
x1	20	20	0.40	0.40	0.60	0.24	
x2	20	40	0.40	0.80	0.20	0.16	
x3	10	50	0.20	1.00	0.00	0.00	
x4	0	50	0.00	1.00	0.00	0.00	
x5	0	50	0.00	1.00	0.00	0.00	
	50		1.00			0.40	
						0.80	Ds



SECONDA METÀ DELLA DISTRIBUZIONE:

X	n	N	f	F	1 - F	F * (1 - F)	
x1	0	0	0.00	0.00	1.00	0.00	
x2	0	0	0.00	0.00	1.00	0.00	
x3	10	10	0.20	0.20	0.80	0.16	
x4	20	30	0.40	0.60	0.40	0.24	
x5	20	50	0.40	1.00	0.00	0.00	
	50		1.00			0.40	
						0.80	Dd

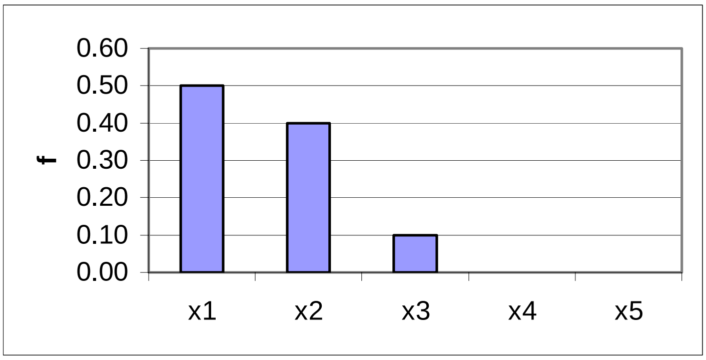


Anche in questo caso i due indici di dispersione sono uguali comportando, per entrambi gli indici di forma, un valore nullo.

CASO 3

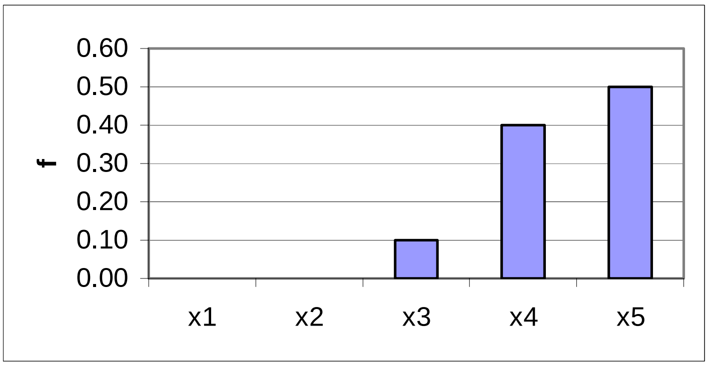
PRIMA METÀ DELLA DISTRIBUZIONE:

X	n	N	f	F	1 - F	F * (1 - F)	
x1	25	25	0.50	0.50	0.50	0.25	
x2	20	45	0.40	0.90	0.10	0.09	
x3	5	50	0.10	1.00	0.00	0.00	
x4	0	50	0.00	1.00	0.00	0.00	
x5	0	50	0.00	1.00	0.00	0.00	
	50		1.00			0.34	
						0.68	Ds



SECONDA METÀ DELLA DISTRIBUZIONE:

X	n	N	f	F	1 - F	F * (1 - F)	
x1	0	0	0.00	0.00	1.00	0.00	
x2	0	0	0.00	0.00	1.00	0.00	
x3	5	5	0.10	0.10	0.90	0.09	
x4	20	25	0.40	0.50	0.50	0.25	
x5	25	50	0.50	1.00	0.00	0.00	
	50		1.00			0.34	
						0.68	Dd

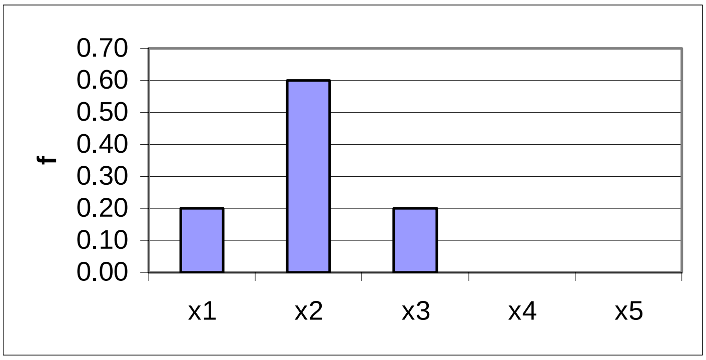


Anche in questo caso i due indici di dispersione sono uguali comportando, per entrambi gli indici di forma, un valore nullo: viene correttamente rilevata la simmetria che è immediatamente leggibile anche dalla rappresentazione grafica.

CASO 4

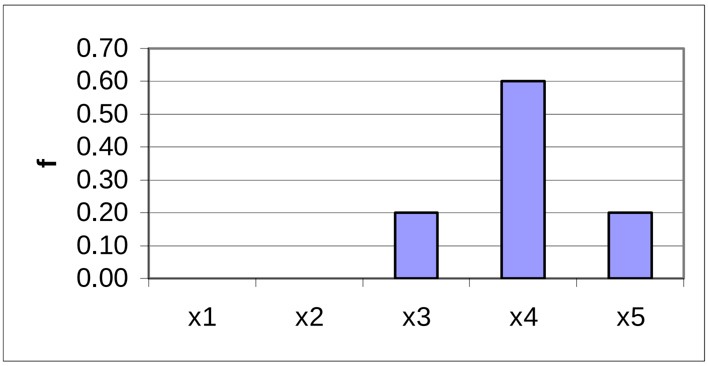
PRIMA METÀ DELLA DISTRIBUZIONE:

X	n	N	f	F	1 - F	F * (1 - F)	
x1	10	10	0.20	0.20	0.80	0.16	
x2	30	40	0.60	0.80	0.20	0.16	
x3	10	50	0.20	1.00	0.00	0.00	
x4	0	50	0.00	1.00	0.00	0.00	
x5	0	50	0.00	1.00	0.00	0.00	
	50		1.00			0.32	
						0.64	Ds



SECONDA METÀ DELLA DISTRIBUZIONE:

X	n	N	f	F	1 - F	F * (1 - F)	
x1	0	0	0.00	0.00	1.00	0.00	
x2	0	0	0.00	0.00	1.00	0.00	
x3	10	10	0.20	0.20	0.80	0.16	
x4	30	40	0.60	0.80	0.20	0.16	
x5	10	50	0.20	1.00	0.00	0.00	
	50		1.00			0.32	
						0.64	Dd

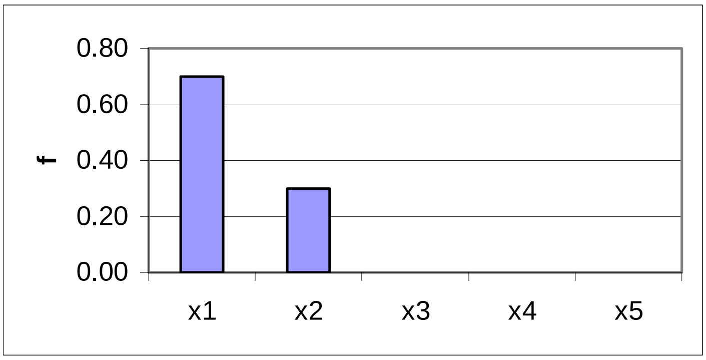


Così come per i casi precedenti, i due indici (sia quello assoluto che quello relativo) sono correttamente nulli essendo anche questa distribuzione perfettamente simmetrica.

CASO 5

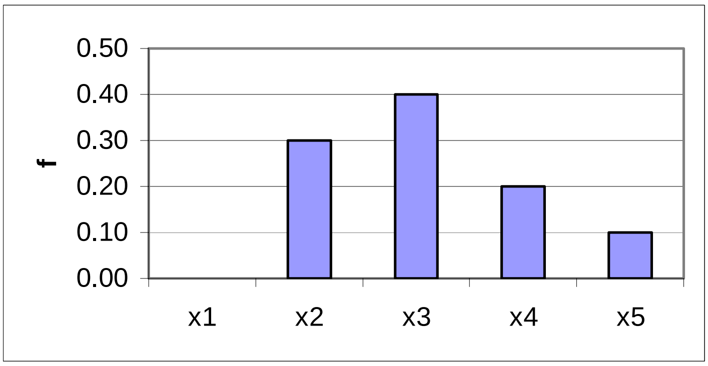
PRIMA METÀ DELLA DISTRIBUZIONE:

X	n	N	f	F	1 - F	F * (1 - F)	
x1	35	35	0.70	0.70	0.30	0.21	
x2	15	50	0.30	1.00	0.00	0.00	
x3	0	50	0.00	1.00	0.00	0.00	
x4	0	50	0.00	1.00	0.00	0.00	
x5	0	50	0.00	1.00	0.00	0.00	
	50		1.00			0.21	
						0.42	Ds



SECONDA METÀ DELLA DISTRIBUZIONE:

X	n	N	f	F	1 - F	F * (1 - F)	
x1	0	0	0.00	0.00	1.00	0.00	
x2	15	15	0.30	0.30	0.70	0.21	
x3	20	35	0.40	0.70	0.30	0.21	
x4	10	45	0.20	0.90	0.10	0.09	
x5	5	50	0.10	1.00	0.00	0.00	
	50		1.00			0.51	
						1.02	Dd



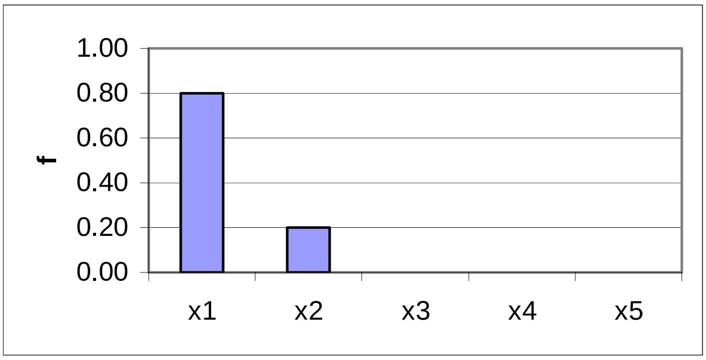
In questo caso gli indici di asimmetria segnalando correttamente la presenza di asimmetria positiva o destra, essendo:

NUMERATORE DI A	NUMERATORE DI A	A
1.02 – 0.42 = 0.60	1.02 + 0.42 = 1.44	0.42

CASO 6

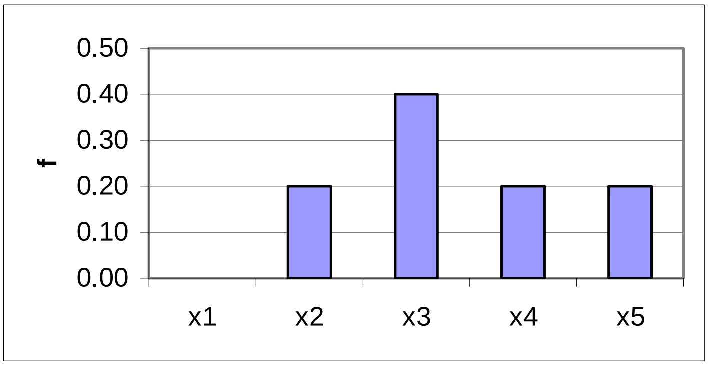
PRIMA METÀ DELLA DISTRIBUZIONE:

X	n	N	f	F	1 - F	F * (1 - F)	
x1	40	40	0.80	0.80	0.20	0.16	
x2	10	50	0.20	1.00	0.00	0.00	
x3	0	50	0.00	1.00	0.00	0.00	
x4	0	50	0.00	1.00	0.00	0.00	
x5	0	50	0.00	1.00	0.00	0.00	
	50		1.00			0.16	
						0.32	Ds



SECONDA METÀ DELLA DISTRIBUZIONE:

X	n	N	f	F	1 - F	F * (1 - F)	
x1	0	0	0.00	0.00	1.00	0.00	
x2	10	10	0.20	0.20	0.80	0.16	
x3	20	30	0.40	0.60	0.40	0.24	
x4	10	40	0.20	0.80	0.20	0.16	
x5	10	50	0.20	1.00	0.00	0.00	
	50		1.00			0.56	
						1.12	Dd



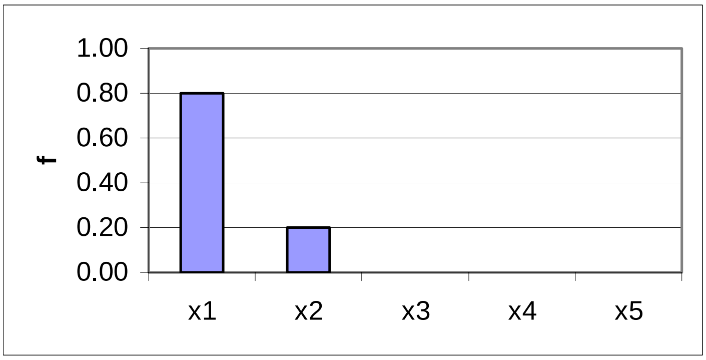
Anche per questa distribuzione gli indici di asimmetria segnalano correttamente la presenza di asimmetria positiva o destra, essendo:

NUMERATORE DI A	NUMERATORE DI A	A
$1.12 - 0.32 = 0.80$	$1.12 + 0.32 = 1.44$	0.56

CASO 7

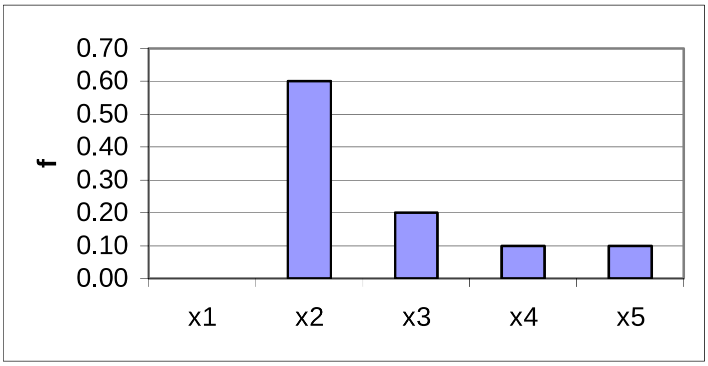
PRIMA METÀ DELLA DISTRIBUZIONE:

X	n	N	f	F	1 - F	F * (1 - F)	
x1	40	40	0.80	0.80	0.20	0.16	
x2	10	50	0.20	1.00	0.00	0.00	
x3	0	50	0.00	1.00	0.00	0.00	
x4	0	50	0.00	1.00	0.00	0.00	
x5	0	50	0.00	1.00	0.00	0.00	
	50		1.00			0.16	
						0.32	Ds



SECONDA METÀ DELLA DISTRIBUZIONE:

X	n	N	f	F	1 - F	F * (1 - F)	
x1	0	0	0.00	0.00	1.00	0.00	
x2	30	30	0.60	0.60	0.40	0.24	
x3	10	40	0.20	0.80	0.20	0.16	
x4	5	45	0.10	0.90	0.10	0.09	
x5	5	50	0.10	1.00	0.00	0.00	
	50		1.00			0.49	
						0.98	Dd



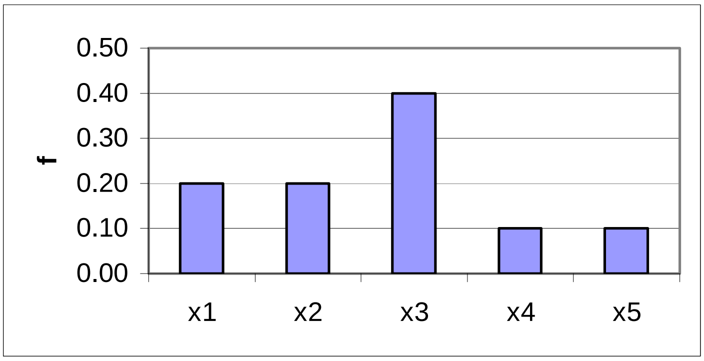
Così come per i due casi precedenti, gli indici di asimmetria segnalano correttamente la presenza di asimmetria positiva o destra, essendo:

NUMERATORE DI A	NUMERATORE DI A	A
$0.98 - 0.32 = 0.60$	$0.98 + 0.32 = 1.30$	0.51

UN CASO PARTICOLARE: A = 0 IN PRESENZA DI ASIMMETRIA

Si consideri la seguente distribuzione:

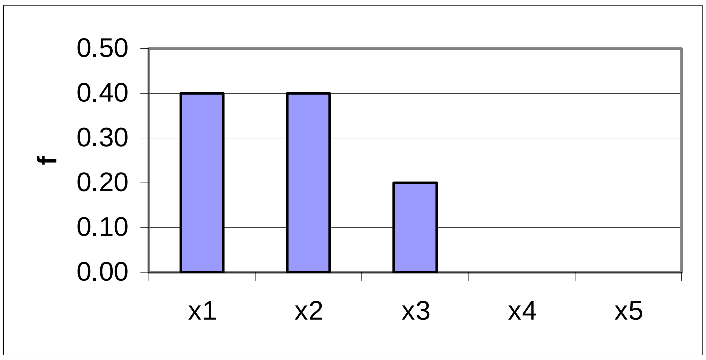
X	n	f
$x_{(1)}$	20	0.20
$x_{(2)}$	20	0.20
$x_{(3)}$	40	0.40
$x_{(4)}$	10	0.10
$x_{(5)}$	10	0.10
	100	1.00



Si può notare la presenza di una leggera asimmetria positiva. Se si procede al calcolo dell'indice di forma per tale distribuzione il valore risultante è però pari a 0: questo conferma il difetto dell'indice in questione (peraltro condiviso da tutta la categoria degli indici di forma) per cui un indice pari a 0 non implica necessariamente la presenza di simmetria, mentre vale sicuramente il contrario. In questo caso l'esame della tabella di frequenza, e ancor meglio della rappresentazione grafica, conferma che la distribuzione esaminata non è perfettamente simmetrica.
Di seguito si riportano i calcoli dell'indice A.

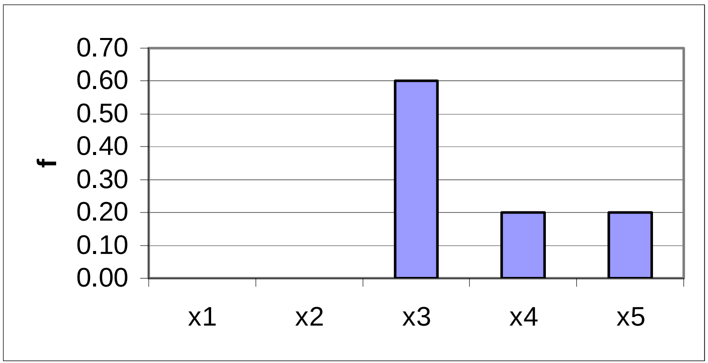
PRIMA METÀ DELLA DISTRIBUZIONE:

X	n	N	f	F	1 - F	F * (1 - F)	
x1	20	20	0.40	0.40	0.60	0.24	
x2	20	40	0.40	0.80	0.20	0.16	
x3	10	50	0.20	1.00	0.00	0.00	
x4	0	50	0.00	1.00	0.00	0.00	
x5	0	50	0.00	1.00	0.00	0.00	
	50		1.00			0.40	
						0.80	Ds



SECONDA METÀ DELLA DISTRIBUZIONE:

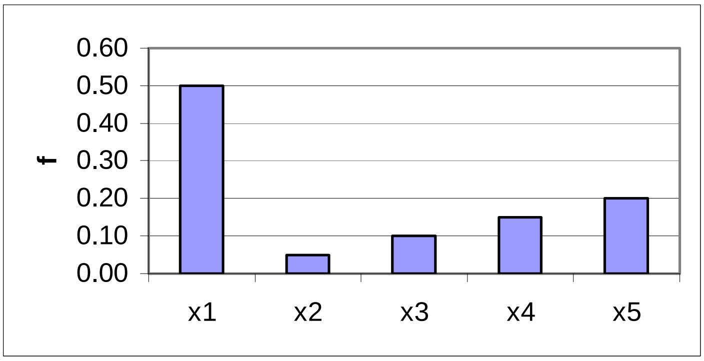
X	n	N	f	F	1 - F	F * (1 - F)	
x1	0	0	0.00	0.00	1.00	0.00	
x2	0	0	0.00	0.00	1.00	0.00	
x3	30	30	0.60	0.60	0.40	0.24	
x4	10	40	0.20	0.80	0.20	0.16	
x5	10	50	0.20	1.00	0.00	0.00	
	50		1.00			0.40	
						0.80	Dd



UN CASO PARTICOLARE: A = +1

Si considera la seguente distribuzione:

X	n	f
$x_{(1)}$	50	0.50
$x_{(2)}$	5	0.05
$x_{(3)}$	10	0.10
$x_{(4)}$	15	0.15
$x_{(5)}$	20	0.20
	100	1.00

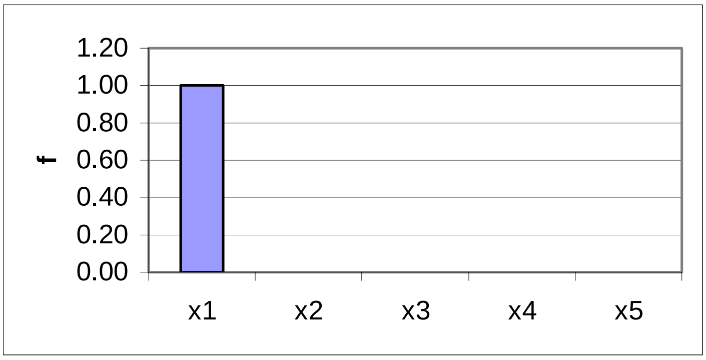


Se si calcola l'indice A per tale distribuzione si avrà un valore pari a +1: ciò è dovuto al fatto che tutte le unità della prima metà della distribuzione (le prime 50) sono concentrate sulla prima modalità della variabile. In tale caso, indipendentemente da come le altre 50 unità sono disposte sulle restanti modalità, l'indice A sarà sempre uguale ad 1 essendo $D_s = 0$. Si tratta, evidentemente, di un altro difetto di tale indice, in quanto in qualunque modo si dispongono le restanti 50 unità almeno su due modalità (altrimenti anche $D_d = 0$) l'indice non riesce a cogliere le differenze.

Si riportano di seguito, per la distribuzione in questione, i calcoli necessari per la misurazione dell'indice A.

PRIMA METÀ DELLA DISTRIBUZIONE:

X	n	N	f	F	1 - F	F * (1 - F)
x1	50	50	1.00	1.00	0.00	0.00
x2	0	50	0.00	1.00	0.00	0.00
x3	0	50	0.00	1.00	0.00	0.00
x4	0	50	0.00	1.00	0.00	0.00
x5	0	50	0.00	1.00	0.00	0.00
	50		1.00			0.00
						0.00 Ds



SECONDA METÀ DELLA DISTRIBUZIONE:

X	n	N	f	F	1 - F	F * (1 - F)
x1	0	0	0.00	0.00	1.00	0.00
x2	5	5	0.10	0.10	0.90	0.09
x3	10	15	0.20	0.30	0.70	0.21
x4	15	30	0.30	0.60	0.40	0.24
x5	20	50	0.40	1.00	0.00	0.00
	50		1.00			0.54
						1.08 Dd