

Le frequenze cumulate

Si consideri una variabile qualitativa ordinale X . Per essa, oltre alle frequenze assolute, relative e percentuali, è possibile calcolare anche le frequenze cumulate e retrocumulate (assolute, relative e percentuali) avendo senso ordinare le modalità della variabile.

In particolare la frequenza cumulata in corrispondenza di una data modalità del carattere, indica il numero (in caso di frequenze assolute) o la frazione (in caso di frequenze relative o percentuali) delle unità della popolazione considerata che presentano un valore della variabile minore o uguale (ovvero non superiore) alla modalità in questione.

La seguente tabella riporta la definizione delle frequenze cumulate nel caso di una generica variabile che può assumere k modalità:

X	n	f	p	N	F	P
$x_{(1)}$	n_1	f_1	p_1	$N_1 = n_1$	$F_1 = f_1$	$P_1 = p_1$
$x_{(2)}$	n_2	f_2	p_2	$N_2 = n_1 + n_2$	$F_2 = f_1 + f_2$	$P_2 = p_1 + p_2$
·	·	·	·	·	·	·
·	·	·	·	·	·	·
·	·	·	·	·	·	·
$x_{(i-1)}$	n_{i-1}	f_{i-1}	p_{i-1}	$N_{i-1} = n_1 + \dots + n_{i-1}$	$F_{i-1} = f_1 + \dots + f_{i-1}$	$P_{i-1} = p_1 + \dots + p_{i-1}$
$x_{(i)}$	n_i	f_i	p_i	$N_i = n_1 + \dots + n_i$	$F_i = f_1 + \dots + f_i$	$P_i = p_1 + \dots + p_i$
$x_{(i+1)}$	n_{i+1}	f_{i+1}	p_{i+1}	$N_{i+1} = n_1 + \dots + n_{i+1}$	$F_{i+1} = f_1 + \dots + f_{i+1}$	$P_{i+1} = p_1 + \dots + p_{i+1}$
·	·	·	·	·	·	·
·	·	·	·	·	·	·
·	·	·	·	·	·	·
$x_{(k-1)}$	n_{k-1}	f_{k-1}	p_{k-1}	$N_{k-1} = n_1 + \dots + n_{k-1}$	$F_{k-1} = f_1 + \dots + f_{k-1}$	$P_{k-1} = p_1 + \dots + p_{k-1}$
$x_{(k)}$	n_k	f_k	p_k	$N_k = n_1 + \dots + n_k = N$	$F_k = f_1 + \dots + f_k = 1$	$P_k = p_1 + \dots + p_k = 100$
	N	1	100			

dove con N_i , F_i e P_i si sono indicate, rispettivamente, le frequenze cumulate assolute, relative e percentuali in corrispondenza della generica i -esima modalità.

Indicando con $\#(\cdot)$ la cardinalità di un insieme, per esse si ha, rispettivamente:

$$N_i = \#(X \leq x_i) = \sum_{j=1}^i n_j$$

$$F_i = \frac{\#(X \leq x_i)}{N} = \frac{\sum_{j=1}^i n_j}{N} = \sum_{j=1}^i \frac{n_j}{N} = \sum_{j=1}^i f_j$$

$$P_i = \frac{\#(X \leq x_i)}{N} \times 100 = \frac{\sum_{j=1}^i n_j}{N} \times 100 = \sum_{j=1}^i \frac{n_j}{N} \times 100 = \sum_{j=1}^i f_j \times 100 = \sum_{j=1}^i p_j$$

Le frequenze retrocumulate

E' possibile definire la frequenza retrocumulata in corrispondenza di una data modalità come il numero (nel caso di frequenze assolute) o la frazione (nel caso di frequenze relative o percentuali) di unità che hanno un valore del carattere maggiore o uguale (ovvero non minore) della modalità in questione.

La seguente tabella riporta la definizione delle frequenze retrocumulate nel caso di una generica variabile che può assumere k modalità:

X	n	f	p	RN	RF	RP
$x_{(1)}$	n_1	f_1	p_1	$RN_1 = n_1 + \dots + n_k = N$	$RF_1 = f_1 + \dots + f_k = 1$	$RP_1 = p_1 + \dots + p_k = 100$
$x_{(2)}$	n_2	f_2	p_2	$RN_2 = n_2 + \dots + n_k$	$RF_2 = f_2 + \dots + f_k$	$RP_2 = p_2 + \dots + p_k$
.
.
.
$x_{(i-1)}$	n_{i-1}	f_{i-1}	p_{i-1}	$RN_{i-1} = n_{i-1} + \dots + n_k$	$RF_{i-1} = f_{i-1} + \dots + f_k$	$RP_{i-1} = p_{i-1} + \dots + p_k$
$x_{(i)}$	n_i	f_i	p_i	$RN_i = n_i + \dots + n_k$	$RF_i = f_i + \dots + f_k$	$RP_i = p_i + \dots + p_k$
$x_{(i+1)}$	n_{i+1}	f_{i+1}	p_{i+1}	$RN_{i+1} = n_{i+1} + \dots + n_k$	$RF_{i+1} = f_{i+1} + \dots + f_k$	$RP_{i+1} = p_{i+1} + \dots + p_k$
.
.
.
$x_{(k-1)}$	n_k	f_k	p_k	$RN_k = n_k$	$RF_k = f_{k-1} + f_k$	$RP_{k-1} = p_{k-1} + p_k$
$x_{(k)}$	n_{k-1}	f_{k-1}	p_{k-1}	$RN_{k-1} = n_{k-1} + n_k$	$RF_k = f_k$	$RP_k = p_k$
	N	1	100			

dove con RN_i , RF_i e RP_i si sono indicate, rispettivamente, le frequenze retrocumulate assolute, relative e percentuali in corrispondenza dell'i-esima modalità. Per esse si ha, rispettivamente:

$$N_i = \#(X \geq x_i) = \sum_{j=i}^k n_j$$

$$F_i = \frac{\#(X \geq x_i)}{N} = \frac{\sum_{j=i}^k n_j}{N} = \sum_{j=i}^k \frac{n_j}{N} = \sum_{j=i}^k f_j$$

$$P_i = \frac{\#(X \geq x_i)}{N} \times 100 = \frac{\sum_{j=i}^k n_j}{N} \times 100 = \sum_{j=i}^k \frac{n_j}{N} \times 100 = \sum_{j=i}^k f_j \times 100 = \sum_{j=i}^k p_j$$

Dalle definizioni di frequenze retrocumulate e cumulate è possibile ricavare la seguente relazione che permette di calcolare le prime una volta note le seconde:

$$RN_i = n_i + \dots + n_k = N - (n_1 + \dots + n_{i-1}) = N - N_{i-1}$$

ovvero di calcolare la frequenza retrocumulata in corrispondenza dell'i-esima modalità sottraendo da N il valore della frequenza cumulata in corrispondenza della modalità precedente. Valgono ovviamente anche le seguenti due relazioni per le frequenze retrocumulate relative e percentuali:

$$RF_i = 1 - F_{i-1}$$

$$RP_i = 100 - P_{i-1}$$

E' chiaro che in corrispondenza della prima modalità si ha sempre:

$$RN_1 = N$$

$$RF_1 = 1$$

$$RP_1 = 100$$

Le frequenze cumulate e retrocumulate possono essere calcolate, chiaramente, anche per una variabile quantitativa. Non risultano invece applicabili nel caso di una variabile qualitativa nominale o sconnessa, anche se dal punto di vista meramente algebrico risulta possibile calcolarle¹.

ESEMPIO

Si consideri la seguente tabella contenente i dati relativi al rendimento all'esame di Statistica, di 350 studenti della Facoltà di Economia:

VOTO ESAME	Frequenza Assoluta
Negativo (< 18)	80
Appena sufficiente (18)	70
Basso (da 19 a 21)	110
Medio (da 22 a 25)	55
Alto (> 25)	35
	350

Il calcolo delle frequenze assolute cumulate è riportato nella seguente tabella:

VOTO ESAME	Frequenza Assoluta	Frequenza Assoluta Cumulata
Negativo (< 18)	80	80
Appena sufficiente (18)	70	(80 + 70=) 150
Basso (da 19 a 21)	110	(80 + 70 + 110) 260
Medio (da 22 a 25)	55	(80 + 70 + 110 + 55) 315
Alto (> 25)	35	(80 + 70 + 110 + 55 + 35) 350
	350	

¹ Un caso in cui si effettua il ricorso alle frequenze cumulate anche per caratteri qualitativi nominali è per la costruzione del diagramma di Pareto.

Se si considera, a titolo esemplificativo, la frequenza cumulata in corrispondenza della classe di rendimento “BASSO”, questa esprime il numero di studenti che presentano un rendimento non superiore a “BASSO” (ovvero un rendimento minore o al più uguale a BASSO). In termini di voto ottenuto, tale richiesta equivale a contare gli studenti che non hanno conseguito più di 21 (andando a contare, cioè, quelli che non hanno superato l’esame e quelli che hanno avuto come voto massimo 21).

Nel caso in questione si ha che 260 (dei 350) studenti analizzati presentano tale caratteristica.

Con riferimento agli stessi dati, si riporta, nella seguente tabella, il calcolo delle frequenze relative e delle frequenze cumulate relative. E’ chiaro che analoghi calcoli vanno applicati nel caso si vogliano ottenere le frequenze percentuali.

VOTO ESAME	Frequenza Relativa	Frequenza Relativa Cumulata
Negativo (< 18)	0.23	0.23
Appena sufficiente (18)	0.20	(0.23 + 0.20) 0.43
Basso (da 19 a 21)	0.31	(0.23 + 0.20 + 0.31) 0.74
Medio (da 22 a 25)	0.16	(0.23 + 0.20 + 0.31 + 0.16) 0.90
Alto (> 25)	0.10	(0.23 + 0.20 + 0.31 + 0.16 + 0.10) 1.00
	1.00	

Facendo riferimento alla stessa modalità “BASSO” del carattere esaminata in precedenza, in questo caso il valore 0.74 della frequenza relativa cumulata indica che il 74% degli studenti analizzati presenta un voto non superiore (ovvero minore o al più uguale) a 21.

In tale caso, quindi, la frequenza relativa cumulata esprime non più il numero quanto la frazione di unità (di studenti) che presentano un valore del carattere (voto) non superiore a quello considerato (voto uguale a 21).

Sullo stesso insieme di dati è naturalmente possibile calcolare le frequenze retrocumulate. Si riportano, nella seguente tabella, le frequenze assolute retrocumulate:

VOTO ESAME	Frequenza Assoluta	Frequenza Assoluta Retrocumulata
Negativo (< 18)	80	(80 + 70 + 110 + 55 + 35) 350
Appena sufficiente (18)	70	(70 + 110 + 55 + 35) 270
Basso (da 19 a 21)	110	(110 + 55 + 35) 200
Medio (da 22 a 25)	55	(55 + 35) 90
Alto (> 25)	35	(35) 35
	350	

Con riferimento a tali frequenze, il valore di 200 corrispondente alla modalità “BASSO” del carattere, è da intendersi come il numero di studenti che presentano un voto non inferiore (ovvero maggiore o uguale) a BASSO; in altri termini un voto maggiore o uguale a 19.

Per il calcolo delle frequenze relative e percentuali retrocumulate basta semplicemente cambiare la colonna di partenza.

Nella successiva tabella è riportato il caso delle frequenze relative retrocumulate per lo stesso insieme di dati:

VOTO ESAME	Frequenza Relativa	Frequenza Relativa Retrocumulata
Negativo (< 18)	0.23	(0.23 + 0.20 + 0.31 + 0.16 + 0.10) 1.00
Appena sufficiente (18)	0.20	(0.20 + 0.31 + 0.16 + 0.10) 0.77
Basso (da 19 a 21)	0.31	(0.31 + 0.16 + 0.10) 0.57
Medio (da 22 a 25)	0.16	(0.16 + 0.10) 0.26
Alto (> 25)	0.10	(0.10) 0.10
	1.00	

In termini di interpretazione, facendo riferimento sempre alla modalità “BASSO” del carattere “VOTO”, il valore 0.57 ci dice che solo il 57% degli studenti analizzati presenta un voto almeno (non inferiore ovvero maggiore o uguale) BASSO.

Si riporta infine, nella seguente tabella, per il solo caso delle frequenze assolute, una modalità alternativa di calcolo delle frequenze retrocumulate che sfrutta la relazione che lega le stesse alle frequenze cumulate:

VOTO ESAME	Frequenza Assoluta	Frequenza Assoluta Cumulata	Frequenza Assoluta Retrocumulata
Negativo (< 18)	80	80	(350) 350
Appena sufficiente (18)	70	(80 + 70) 150	(350 - 80) 270
Basso (da 19 a 21)	110	(80 + 70 + 110) 260	(350 - 150) 200
Medio (da 22 a 25)	55	(80 + 70 + 110 + 55) 315	(350 - 260) 90
Alto (> 25)	35	(80 + 70 + 110 + 55 + 35) 350	(350 - 315) 35
	350		