

Statistica: principi e metodi



Capitolo 2

Distribuzioni statistiche

Distribuzioni statistiche

- L'effetto dell'operazione di determinazione della modalità con cui ognuno dei caratteri si presenta in ciascuna unità del collettivo determina la "*distribuzione*" del collettivo secondo i caratteri considerati.
- Il nome "*distribuzione*" deriva dal fatto che mediante essa si indica come le modalità dei caratteri si distribuiscono nelle unità del collettivo.

Distribuzioni statistiche disaggregate

Si consideri un collettivo statistico di N unità, dove si sia osservato il carattere X .

Si chiama **distribuzione statistica disaggregata** secondo il carattere X l'insieme delle osservazioni (rappresentate da numeri o da espressioni verbali) relative alle N unità del collettivo.

In simboli, la distribuzione disaggregata sarà indicata come

$$x_1, x_2, \dots, x_N$$

dove x_1 è l'osservazione relativa all'unità identificata dal numero 1, x_2 l'osservazione relativa all'unità identificata dal numero 2 e così via.

Esempi di distribuzioni statistiche disaggregate

Studente	Sesso	Livello di soddisfazione sul corso di laurea	Numero di esami sostenuti	voto medio	voto medio in classi		
Gianni	M	Abbastanza	2	24.50	24		27
Maria	F	Abbastanza	2	26.50	24		27
Lucia	F	Poco	3	25.00	24		27
Luca	M	Molto	5	24.20	24		27
Fabio	M	Per nulla	1	18.00	18		22
Mirko	M	Abbastanza	3	27.33	27		30
Angela	F	Molto	5	26.60	24		27
Michele	M	Molto	3	24.00	24		27
Marco	M	Molto	4	23.25	22		24
Antonio	M	Abbastanza	2	23.50	22		24
Luigi	M	Abbastanza	1	24.00	24		27
Paola	F	Molto	3	24.67	24		27
Sara	F	Abbastanza	2	22.50	22		24
Aldo	M	Poco	2	21.00	18		22
Monica	F	Poco	1	28.00	27		30
Giovanni	M	Abbastanza	2	30.00	27		30
Chiara	F	Molto	5	27.80	27		30
Davide	M	Poco	1	21.00	18		22
Fabrizio	M	Abbastanza	4	24.25	24		27
Manuela	F	Per nulla	3	22.00	18		22

Distribuzioni di frequenze

- L'operazione di raggruppamento delle unità statistiche viene realizzata mediante la **classificazione** o lo **spoglio dei dati**.

Per **frequenza** si intende il numero di volte che una data modalità si presenta nel collettivo statistico.

Distribuzioni di frequenze

Si chiama **distribuzione di frequenze** lo schema con cui si associa a ciascuna modalità del carattere X la rispettiva frequenza.

<i>Modalità (x_i)</i>	<i>Frequenza (n_i)</i>
x_1	n_1
x_2	n_2
\vdots	\vdots
x_k	n_k
Totale	N

dove n_1, n_2, \dots, n_k sono le frequenze delle modalità x_1, x_2, \dots, x_k

$$\sum_{i=1}^k n_i = n_1 + n_2 + \dots + n_k = N$$

Distribuzione di frequenze con modalità raggruppate in classi

Si chiama **distribuzione di frequenze** di un carattere X suddiviso in classi lo schema con cui si associa a ciascuna classe la rispettiva frequenza:

Classe ($c_{i-1}-c_i$)	Frequenza (n_i)
$c_0 - c_1$	n_1
$c_1 - c_2$	n_2
\vdots	\vdots
$c_{k-1} - c_k$	n_k
Totale	N

-	<i>aperto</i>	non comprende gli estremi
-	<i>chiuso a sx</i>	comprende l'estremo inferiore
-	<i>chiuso a dx</i>	comprende l'estremo superiore
-	<i>chiuso</i>	comprende gli estremi

Esempi di distribuzioni di frequenze

Sesso		
	x_i	n_i
x_1	Femmine	8 n_1
x_2	Maschi	12 n_2
	Totale	20 N

Numero di esami sostenuti		
	x_i	n_i
x_1	1	4 n_1
x_2	2	6 n_2
x_3	3	5 n_3
x_4	4	2 n_4
x_5	5	3 n_5
	Totale	20 N

Livello di soddisfazione sul corso di laurea		
	x_i	n_i
x_1	Per nulla	2 n_1
x_2	Poco	4 n_2
x_3	Abbastanza	8 n_3
x_4	Molto	6 n_4
	Totale	20 N

voto medio		
	$c_{i-1} \vdash c_i$	n_i
$c_0 \vdash c_1$	18 -22	4 n_1
$c_1 \vdash c_2$	22 -24	3 n_2
$c_2 \vdash c_3$	24 -27	9 n_3
$c_3 \vdash c_4$	27 - 30	4 n_4
	Totale	20 N

Frequenze relative e percentuali

- ▣ frequenze **relative o proporzioni**: si ottengono rapportando le frequenze assolute al totale delle unità, N . Indicheremo con f_1, f_2, \dots, f_k tali quantità, essendo

$$f_i = \frac{\text{frequenza della modalità } x_i}{\text{numero totale di osservazioni}} = \frac{n_i}{N}, i = 1, 2, \dots, k.$$

- ▣ frequenze **percentuali** p_1, p_2, \dots, p_k : si ottengono moltiplicando per 100 le frequenze relative.

$$\sum_{i=1}^k f_i = f_1 + f_2 + \dots + f_k = 1$$

$$\sum_{i=1}^k p_i = p_1 + p_2 + \dots + p_k = 100$$

Esempio di distribuzioni di frequenze assolute, relative e percentuali

Voto in Microeconomia	Frequenza	Frequenza relativa	Frequenza percentuale
18	5	0.016	1.6
19	12	0.038	3.8
20	20	0.063	6.3
21	20	0.063	6.3
22	15	0.048	4.8
23	32	0.102	10.2
24	53	0.168	16.8
25	41	0.130	13.0
26	43	0.137	13.7
27	42	0.133	13.3
28	15	0.048	4.8
29	5	0.016	1.6
30	12	0.038	3.8
Totale	315	1.000	100

Esempi di distribuzioni di frequenze relative e percentuali

Sesso

	x_i	f_i	p_i
x_1	Femmine	0.4 f_1	40 p_1
x_2	Maschi	0.6 f_2	60 p_2
	Totale	1	100

Livello di soddisfazione

	x_i	f_i	p_i
x_1	Per nulla	0.1 f_1	10 p_1
x_2	Poco	0.2 f_2	20 p_2
x_3	Abbastanza	0.4 f_3	40 p_3
x_4	Molto	0.3 f_4	30 p_4
	Totale	1	100

Numero di esami

	x_i	f_i	p_i
x_1	1	0.20 f_1	20 p_1
x_2	2	0.30 f_2	30 p_2
x_3	3	0.25 f_3	25 p_3
x_4	4	0.10 f_4	10 p_4
x_5	5	0.15 f_5	15 p_5
	Totale	1	100

Voto medio

	$c_{i-1}+c_i$	f_i	p_i
c_0+c_1	18 -22	0.20 f_1	20 p_1
c_1+c_2	22 -24	0.15 f_2	15 p_2
c_2+c_3	24 -27	0.45 f_3	45 p_3
c_3+c_4	27 -30	0.20 f_4	20 p_4
	Totale	1	100

Frequenze cumulate

Consideriamo una distribuzione di frequenze secondo un carattere a modalità ordinabili.

- Si chiamano **frequenze cumulate** le quantità

$$N_i = n_1 + n_2 + \dots + n_i, \quad i = 1, 2, \dots, k.$$

Per ogni dato i , N_i rappresenta il numero delle unità del collettivo nelle quali il carattere X assume un valore non superiore a x_i .

- Si chiamano **frequenze relative cumulate** i rapporti

$$F_i = \frac{\text{frequenza cumulata fino a } x_i}{\text{numero totale di osservazioni}} = \frac{N_i}{N}, \quad i = 1, 2, \dots, k.$$

- **frequenze percentuali cumulate** : si ottengono moltiplicando per 100 le frequenze relative cumulate

Esempio di distribuzioni di frequenze cumulate assolute, relative e percentuali

Voto in Microeconomia	Frequenza	Frequenza cumulata	Frequenza relativa cumulata	Frequenza percentuale cumulata
18	5	5	0.016	1.6
19	12	17	0.054	5.4
20	20	37	0.117	11.7
21	20	57	0.181	18.1
22	15	72	0.229	22.9
23	32	104	0.330	33.0
24	53	157	0.498	49.8
25	41	198	0.629	62.9
26	43	241	0.765	76.5
27	42	283	0.898	89.8
28	15	298	0.946	94.6
29	5	303	0.962	96.2
30	12	315	1.000	100
Totale	315			

Esempi di distribuzioni di frequenze cumulate assolute, relative e percentuali

Livello di soddisfazione

x_i	n_i	N_i	F_i	P_i
Per nulla	2	2	0.10	10
Poco	4	6	0.30	30
Abbastanza	8	14	0.70	70
Molto	6	20	1.00	100
Totale	20			

Voto medio

$c_{i-1} - c_i$	n_i	N_i	F_i	P_i
18 -22	4	4	0.20	20
22 -24	3	7	0.35	35
24 -27	9	16	0.80	80
27 -30	4	20	1.00	100
Totale	20			

Numero di esami sostenuti

x_i	n_i	N_i	F_i	P_i
1	4	4	0.20	20
2	6	10	0.50	50
3	5	15	0.75	75
4	2	17	0.85	85
5	3	20	1.00	100
Totale	20			

Densità di frequenza per una distribuzione in classi

Sia (c_{i-1}, c_i) la generica classe di una distribuzione di frequenze con modalità raggruppate in classi.

Si chiama densità di frequenza della classe il rapporto

$$h_i = \frac{\text{frequenza della classe}}{\text{ampiezza della classe}} = \frac{n_i}{d_i}$$

essendo $d_i = c_i - c_{i-1}$ l'ampiezza della classe.

Distribuzione di frequenze con modalità raggruppate in classi

Lunghezza avambraccio (cm)	Frequenza	Ampiezza della classe	Densità di frequenza
40.5-42.5	5	2	2.5
42.5-44.5	17	2	8.5
44.5-46.5	25	2	12.5
46.5-48.5	35	2	17.5
48.5-50.5	32	2	16.0
50.5-52.5	20	2	10.0
52.5-54.5	6	2	3.0
Totale	140		

Distribuzioni doppie

- Considerando congiuntamente due colonne della matrice dei dati, l'insieme delle coppie di modalità dei due caratteri che così si osservano costituisce una **distribuzione doppia disaggregata**
- le distribuzioni doppie di **frequenze** sono il risultato dello spoglio dei dati basato su una preliminare definizione delle modalità e delle eventuali classi per entrambi i caratteri.

Distribuzioni doppie

	<i>Sesso</i>		
<i>Livello di soddisfazione</i>	Femmine	Maschi	Totale
Per nulla	1	1	2
Poco	2	2	4
Abbastanza	2	6	8
Molto	3	3	6
Totale	8	12	20

Studenti per sesso e livello di soddisfazione
rispetto al corso di laurea

Distribuzioni doppie

Sesso	Scuola superiore				Totale
	C	S	T	A	
M	16	75	62	14	167
F	14	41	82	11	148
Totale	30	116	144	25	315

Laureati per sesso e scuola superiore

Distribuzioni multiple

In generale si parla di:

- ▣ **distribuzione tripla**, se si considerano congiuntamente tre caratteri (tre colonne della matrice dei dati); **distribuzione quadrupla**, se si considerano congiuntamente quattro caratteri, e così via
- ▣ anche in questo caso, le distribuzioni si distinguono in disaggregate e di frequenze, a seconda che i dati si considerino allo stato grezzo o che si sia proceduto allo spoglio

Distribuzione di quantità

Si chiama **distribuzione di quantità** lo schema con cui si associa a ogni modalità del carattere X il totale dello stesso o di un altro carattere posseduto dalle unità che presentano quella data modalità di X . Lo schema è presentato di seguito per il caso di modalità non raggruppate (parte sinistra) e per il caso opposto.

Distribuzione di quantità

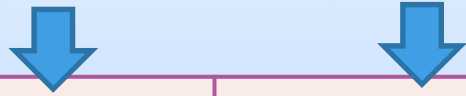
Distribuzione di quantità con modalità non raggruppate		Distribuzione di quantità con modalità raggruppate	
Modalità	Quantità	Modalità	Quantità
x_1	t_1	$c_0 - c_1$	t_1
x_2	t_2	$c_1 - c_2$	t_2
\vdots	\vdots	\vdots	\vdots
x_k	t_k	$c_{k-1} - c_k$	t_k
Totale	T	Totale	T

Esempio di distribuzione di quantità

Imprese e addetti dipendenti per classi di addetti- Anno 2011

frequenza

quantità



CLASSI DI ADDETTI	Imprese	Addetti
1	2,655,768	2,480,178
2-5	1,356,893	3,769,034
6-9	221,161	1,572,718
10-19	137,212	1,795,963
20-49	54,218	1,613,195
50-99	14,503	991,969
100-249	7,536	1,133,819
250 e più	3,646	3,520,706
Totale	4,450,937	16,877,583

*associa ad
ogni modalità
del carattere
X il totale
dello stesso
carattere*

Fonte Istat

Esempio di distribuzione di quantità

Imprese e addetti indipendenti e dipendenti per forma giuridica - Anno 2011

	frequenza ↓	quantità ↓
FORME GIURIDICHE	Imprese	Addetti
Imprese Individuali	2,813,981	4,305,867
Società di persone	767,736	2,684,667
Società di capitali	795,755	8,601,178
Società cooperative	55,928	1,102,032
Altra forma	17,537	183,839
Totale	4,450,937	16,877,583

Fonte Istat

*associa ad ogni modalità del
carattere X il totale di un altro
carattere posseduto*

Serie storiche

- Si ha una serie storica quando i dati statistici di interesse vengono associati a **modalità temporali**
- I dati statistici possono riguardare sia fenomeni di movimento (*flusso*) che fenomeni di stato (*stock*)
- A seconda della natura del fenomeno considerato, la serie storica si configura come distribuzione di frequenze oppure come distribuzione di quantità

Serie storiche

Esempio serie storica di quantità (fenomeno di movimento)

Anno	2003	2004	2005	2006	2007	2008	2009
N. di divorzi	43856	45097	47036	49534	50669	54351	54456

Divorzi in Italia dal 2003 al 2009

Esempio serie storica di quantità (fenomeno di stato)

Anno	2005	2006	2007	2008	2009
N. di residenti	58462	58752	59131	59616	60045

Popolazione residente in Italia dal 2005 al 2009

Serie territoriali

- Si ha una serie territoriale quando i dati vengono associati a modalità rappresentate da **entità territoriali**
- possono essere interpretate come distribuzioni di frequenze (risultato della classificazione di unità statistiche in base a modalità territoriali) o come distribuzioni di quantità se alle modalità territoriali vengono associati i valori di un carattere trasferibile come il PIL, la produzione industriale ecc.

Serie territoriali

Esempio serie territoriale

Ripartizione territoriale	Popolazione residente
Italia Nord-Occidentale	16120067
Italia Nord-Orientale	11643194
Italia Centrale	11950322
Italia Meridionale	14186373
Italia Insulare	6726486
Totale	60626442

Popolazione residente in Italia per ripartizione territoriale