

Statistica

Lezione 2

Bice Cavallo



Indice

- ◆ Dalle distribuzioni unitarie alle distribuzioni di frequenza
- ◆ Frequenze relative e frequenze percentuali
- ◆ Frequenze cumulate
- ◆ Serie storica e serie territoriale

Dalle distribuzioni unitarie alle distribuzioni di frequenza

- ◆ La **distribuzione unitaria semplice di un carattere** è l'elencazione delle modalità osservate, unità per unità, nel collettivo preso in considerazione
- ◆ Si parla di **distribuzione unitaria multipla** quando tale elencazione si riferisce a più di un carattere

Dalle distribuzioni unitarie alle distribuzioni di frequenza

TABELLA 2.1.1 Esempio di distribuzione unitaria multipla

Unità	Sesso	Età	Reddito (migliaia di €)	Statura (cm)	Colore degli occhi
1	Maschio	22	0,7	173	nero
2	Femmina	18	0,2	168	marrone
3	Femmina	34	1,6	165	marrone
4	Maschio	42	2,5	180	nero
5	Femmina	50	3,2	163	azzurro
6	Femmina	12	0,1	160	nero
7	Maschio	46	3,8	177	marrone
8	Maschio	72	1,3	164	verde
9	Femmina	27	1,2	158	azzurro
10	Femmina	48	1,7	170	nero
11	Femmina	35	1,9	167	nero
12	Maschio	84	0,8	159	marrone
13	Femmina	21	0,4	174	azzurro
14	Femmina	44	1,8	164	verde
15	Maschio	56	1,9	177	nero
16	Femmina	58	3,2	172	nero
17	Femmina	37	2,1	166	marrone
18	Femmina	16	0,1	160	marrone
19	Maschio	73	1,6	170	azzurro
20	Maschio	64	2,2	184	verde

Dalle distribuzioni unitarie alle distribuzioni di frequenza

- ◆ La distribuzione unitaria, pur descrivendo fedelmente la situazione osservata, non consente di cogliere in maniera sintetica le caratteristiche del fenomeno. Es.
 - ◆ Ci sono più maschi o più femmine?
 - ◆ Ci sono più giovani o più anziani?
 - ◆ Quale è il colore degli occhi che prevale?

Dalle distribuzioni unitarie alle distribuzioni di frequenza

- La **frequenza assoluta** di una modalità di un carattere è il numero di volte che questa viene osservata nel collettivo
- La **distribuzione di frequenze semplice** associa alle modalità che può assumere un carattere, qualitativo o quantitativo, le corrispondenze frequenze assolute

X	Freq.
x_1	n_1
x_2	n_2
...	...
x_j	n_j
...	...
x_K	n_K
Totale	n

Dalle distribuzioni unitarie alle distribuzioni di frequenza

TABELLA 2.1.2

Sesso	Freq.
maschi	8
femmine	12
Totale	20

- ◆ Si nota ad esempio che ci sono più femmine che uomini

Dalle distribuzioni unitarie alle distribuzioni di frequenza

TABELLA 2.1.3		TABELLA 2.1.4		TABELLA 2.1.5		TABELLA 2.1.6	
Età	Freq.	Reddito	Freq.	Colore occhi	Freq.	Statura	Freq.
da 10 a 30	6	0-1	6	nero	7	da 150 a 160	2
da 30 a 50	7	1-2	8	marrone	6	da 160 a 170	9
da 50 a 70	4	2-3	3	azzurro	4	da 170 a 180	7
da 70 a 90	3	3-4	3	verde	3	da 180 a 190	2
Totale	20	Totale	20	Totale	20	Totale	20

- ◆ Per i dati quantitativi ed ordinabili, la sequenza di modalità viene ordinata in senso crescente o decrescente
- ◆ I caratteri Età, Reddito e Statura sono suddivisi in classi considerando intervalli aperti a destra
- ◆ Dalla distribuzione per Età si nota che ci sono più giovani (10-30) che anziani (70-90)
- ◆ Dalla distribuzione per Colore degli occhi si può notare che i colori più frequenti sono nero e marrone
- ◆ Dalla distribuzione del reddito si può osservare che il reddito più frequente è compreso tra 1000 ed 2000 euro
- ◆ Perdita di informazione

Frequenze relative e frequenze percentuali

- ◆ La **frequenza relativa** è il rapporto tra la frequenza assoluta ed il numero totale di unità osservate
- ◆ La **frequenza percentuale** è pari alla frequenza relativa moltiplicata per 100

Frequenze relative e frequenze percentuali

X	Frequenza assoluta	Frequenza relativa	Frequenza percentuale
x_1	n_1	f_1	p_1
x_2	n_2	f_2	p_2
...
x_j	n_j	f_j	p_j
...
x_K	n_K	f_K	p_K
Totale	n	1	100

$$f_i = \frac{n_i}{n} \quad p_i = f_i \cdot 100$$

Frequenze relative e frequenze percentuali

TABELLA 2.1.2

Sesso	Freq.
maschi	8
femmine	12
Totale	20



- ◆ Determinare frequenze relative e frequenze percentuali

Frequenze relative e frequenze percentuali

TABELLA 2.1.2

Sesso	Freq.
maschi	8
femmine	12
Totale	20



Sesso	Frequenze relative
maschi	0,4
femmine	0,6

Sesso	Frequenze percentuali
maschi	40
femmine	60

$$f_1 = \frac{8}{20} = 0.4 \quad p_1 = 0.4 \cdot 100 = 40$$

$$f_2 = \frac{12}{20} = 0.6 \quad p_2 = 0.6 \cdot 100 = 60$$

Frequenze relative e frequenze percentuali

- ◆ Perché frequenze relative e percentuali?
 - ◆ I valori delle frequenze assolute dipendono dalla dimensione del collettivo

Frequenze relative e frequenze percentuali

Gruppo 1	
x_1	2
x_2	4
x_3	8
Totale	14

Gruppo 2	
x_1	12
x_2	46
x_3	32
Totale	90

- x_1, x_2, x_3 sono tre modalità osservate di un carattere X su due diversi gruppi di persone
- n_1 (gruppo 1) = 2 < n_1 (gruppo 2) = 12

Frequenze relative e frequenze percentuali

Gruppo 1	
x_1	14,29
x_2	28,57
x_3	57,14
Totale	100,00

Gruppo 2	
x_1	13,33
x_2	51,11
x_3	35,56
Totale	100,00

- p_1 (gruppo 1) = 14.29 > p_1 (gruppo 2) = 13.33

Frequenze cumulate

- ◆ Nel caso in cui le modalità del carattere in esame sono ordinate (ossia il carattere è qualitativo ordinato o quantitativo) può essere interessante considerare la frequenza con cui si presentano **modalità di ordine inferiore o uguale ad una certa modalità.**

Frequenze cumulate

TABELLA 2.1.4

Reddito	Freq.
0-1	6
1-2	8
2-3	3
3-4	3
Totale	20

N.B. classi con intervalli aperti a destra

- ◆ Quanti individui guadagnano meno di 2000 euro?
- ◆ Quanti individui guadagnano meno di 3000 euro?

Frequenze cumulate

TABELLA 2.1.4

Reddito	Freq.
0-1	6
1-2	8
2-3	3
3-4	3
Totale	20

N.B. classi con intervalli aperti a destra

- ◆ Quanti individui guadagnano meno di 2000 euro? $6+8=14$
- ◆ Quanti individui guadagnano meno di 3000 euro?

Frequenze cumulate

TABELLA 2.1.4

Reddito	Freq.
0-1	6
1-2	8
2-3	3
3-4	3
Totale	20

- ◆ Quanti individui guadagnano meno di 2000 euro? $6+8=14$
- ◆ Quanti individui guadagnano meno di 3000 euro? $14+3$

Frequenze cumulate

TABELLA 2.1.4

Reddito	Freq.
0-1	6
1-2	8
2-3	3
3-4	3
Totale	20



TABELLA 2.3.1

Reddito	Frequenza assoluta cumulata
0-1	6
1-2	14
2-3	17
3-4	20

- ◆ Quanti individui guadagnano meno di 2000 euro? $6+8=14$
- ◆ Quanti individui guadagnano meno di 3000 euro? $14+3=17$

Frequenze cumulate

X	Frequenza assoluta	Frequenza relativa	Frequenza percentuale
x_1	n_1	f_1	p_1
x_2	n_2	f_2	p_2
...
x_j	n_j	f_j	p_j
...
x_K	n_K	f_K	p_K
Totale	n	1	100



X	Frequenza assoluta cumulata	Frequenza relativa cumulata	Frequenza percentuale cumulata
x_1	N_1	F_1	P_1
x_2	N_2	F_2	P_2
...
x_j	N_j	F_j	P_j
...
x_K	N_K	F_K	P_K

Frequenze cumulate

X	Frequenza assoluta	Frequenza relativa	Frequenza percentuale
x_1	n_1	f_1	p_1
x_2	n_2	f_2	p_2
...
x_j	n_j	f_j	p_j
...
x_K	n_K	f_K	p_K
Totale	n	1	100



$$N_i = n_1 + n_2 + \dots + n_i = N_{i-1} + n_i \quad (N_1 = n_1, N_k = n)$$

$$F_i = f_1 + f_2 + \dots + f_i = F_{i-1} + f_i \quad (F_1 = f_1, F_k = 1)$$

$$P_i = p_1 + p_2 + \dots + p_i = P_{i-1} + p_i \quad (P_1 = p_1, P_k = 100)$$

X	Frequenza assoluta cumulata	Frequenza relativa cumulata	Frequenza percentuale cumulata
x_1	N_1	F_1	P_1
x_2	N_2	F_2	P_2
...
x_j	N_j	F_j	P_j
...
x_K	N_K	F_K	P_K

Frequenze cumulate

$$\begin{aligned} N_i &= n_1 + n_2 + \dots + n_i = N_{i-1} + n_i & (N_1 = n_1, N_k = n) \\ F_i &= f_1 + f_2 + \dots + f_i = F_{i-1} + f_i & (F_1 = f_1, F_k = 1) \\ P_i &= p_1 + p_2 + \dots + p_i = P_{i-1} + p_i & (P_1 = p_1, P_k = 100) \end{aligned}$$



$$\begin{aligned} n_i &= N_i - N_{i-1} \\ f_i &= F_i - F_{i-1} \\ p_i &= P_i - P_{i-1} \end{aligned}$$

Frequenze cumulate

TABELLA 2.1.3

Età	Freq.
da 10 a 30	6
da 30 a 50	7
da 50 a 70	4
da 70 a 90	3
Totale	20

- ◆ Determinare le distribuzioni di frequenze relative e frequenze relative cumulate

Frequenze cumulate

TABELLA 2.1.3

Età	Freq.
da 10 a 30	6
da 30 a 50	7
da 50 a 70	4
da 70 a 90	3
Totale	20



TABELLA 2.3.2 Distribuzioni di frequenze per il carattere Età

Età	Frequenza assoluta	Frequenza relativa	Frequenza relativa cumulata
10-30	6	0,30	0,30
30-50	7	0,35	0,65
50-70	4	0,20	0,85
70-90	3	0,15	1,00
Totale	20	1,00	

Serie storica e serie territoriale

- ◆ Quando si misura un fenomeno nel tempo, registrandolo in determinati istanti o conteggiandolo in periodi definiti, si ottiene una rappresentazione tabellare che prende il nome di **serie storica**
- ◆ Nel caso di un carattere geografico, le cui modalità rappresentano nazioni, regioni, ripartizioni geografiche, città, locazioni territoriali, ecc. la distribuzione viene detta **serie territoriale**

Serie storica e serie territoriale

TABELLA 2.3.3 Occupati in Italia, anni 2001-2011 (in migliaia, fonte ISTAT)

Anno	Occupati
2001	21.965
2002	22.240
2003	22.289
2004	22.404
2005	22.563
2006	22.988
2007	23.222
2008	23.405
2009	23.025
2010	22.872
2011	22.967

TABELLA 2.3.4 Occupati in Italia nel 2011, per area geografica (in migliaia, fonte ISTAT)

Area geografica	Occupati
Nord	11.925
Centro	4.826
Mezzogiorno	6.216