

## Statistica Descrittiva

### Soluzioni 1. Caratteri e distribuzioni statistiche

#### Introduzione

Sia  $X$  il carattere rilevato su un numero  $N$  di unità statistiche. Distinguiamo la **natura qualitativa** o **quantitativa** del carattere a seconda che esso sia individuato tramite parole, attributi, proprietà o, in alternativa, che se ne possa cogliere l'intensità tramite un'unità di misura. Nel caso di un carattere qualitativo, distinguiamo tra **sconnesso** e **ordinale** a seconda che le sue modalità assumano o meno valori ordinabili in senso logico. Nel caso di un carattere quantitativo, distinguiamo tra **discreto** e **continuo** a seconda che i valori assumibili dalle modalità siano o meno definiti in un insieme numerabile. Infine, ricordiamo la possibilità di distinguere tra un carattere di stato o un carattere di movimento e, nel caso di carattere quantitativo, la possibilità di specificare se esso sia trasferibile o meno (cedibile ad altre unità statistiche).

Siano  $x_j, j = 1, \dots, J$ , le modalità assunte dal carattere  $X$ . Indichiamo con  $f_j$  la **frequenza assoluta** della modalità  $x_j$ , vale a dire il numero di determinazioni di  $X$  che corrispondono a quella modalità. Indichiamo con  $p_j$  la **frequenza relativa** della modalità  $x_j$ , ottenuta come  $f_j/N$ . Ricordiamo che, mentre le frequenze assolute sommano a  $N$ , le frequenze relative sommano a  $N/N = 1$ . La determinazione delle frequenze assolute/relative per ogni modalità del carattere  $X$  in oggetto consente di definire la distribuzione assoluta/relativa di  $X$ .

Modalità	$f_j$	$p_j$	$F_j$	$P_j$
$x_1$	$f_1$	$p_1$	$f_1 = F_1$	$p_1 = P_1$
$x_2$	$f_2$	$p_2$	$f_1 + f_2 = F_2$	$p_1 + p_2 = P_2$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$x_J$	$f_J$	$p_J$	$f_1 + f_2 + \dots + f_J = F_J$	$p_1 + p_2 + \dots + p_J = P_J$

Le quantità  $F_j$  e  $P_j$  nella tabella indicano rispettivamente la **frequenza assoluta cumulata** e la **frequenza relativa cumulata**, definite come la somma della frequenza assoluta/relativa di  $x_j$  e di tutte le modalità ad esse precedenti. Si ricorda che  $F_J = N$  e  $P_J = 1$ .

Infine, ricordiamo che per un carattere è di interesse valutare la **funzione di ripartizione**. Essa, se valutata in corrispondenza della modalità  $x_j$ , viene indicata come  $F_X(x_j)$  ed è definita come il rapporto tra il numero delle determinazioni di  $X$  (frequenze assolute) minori o uguali di  $x_j$  e il totale delle determinazioni ( $N$ ).

#### Esercizio A.

1. Ammontare popolazione residente (Comuni): quantitativo discreto, di stato, trasferibile;
2. Nazionalità (Cittadini U.E.): sconnesso, di stato;
3. Ammontare delitti (Regioni): quantitativo discreto, di movimento, trasferibile;
4. Reddito pro-capite (Regioni): quantitativo continuo, di movimento, trasferibile;

5. Angolo di incidenza dei raggi solari con la superficie terrestre (Incidenti stradali verificatisi): quantitativo continuo, di stato, non trasferibile;
6. Anno di corso (Studenti universitari): ordinale, di stato;
7. Quintali di merci sbarcate (Aeroporti): quantitativo continuo, di movimento, trasferibile;
8. Percentuale di occupati per settore (Regione): quantitativo continuo, di stato, non trasferibile.
9. Livello di istruzione: qualitativo ordinale, di stato.
10. Colore degli occhi: qualitativo sconnesso, di stato.
11. Numero di figli: quantitativo discreto, di stato, non trasferibile.

### Esercizio B.

a) Le unità statistiche sono le entità sulle quali vengono rilevati i caratteri di interesse per l'indagine. Nel caso in esame, si tratta dei "fabbricati residenziali di nuova costruzione". I caratteri rilevati su tali unità statistiche e le relative modalità sono la *provincia* (carattere qualitativo sconnesso, modalità: Verona; Vicenza; Belluno; Treviso; Venezia; Padova; Rovigo) e il *tipo di finanziamento pubblico* (carattere qualitativo sconnesso, modalità: senza finanziamento pubblico; a sovvenzione Statale o Regionale; a contributo Statale o Regionale; a sovvenzione o contributo della Provincia o Comune).

b) La distribuzione doppia di frequenza riporta il numero di unità statistiche (abitazioni) per provincia e per tipo di finanziamento pubblico: è data dall'unione delle colonne (seconda, quarta, sesta, decima) con intestazione "Abitazioni". La distribuzione doppia di quantità invece fa riferimento alle colonne (prima, terza, quinta, nona) con intestazione "Volume".

### Esercizio C.

a) Le unità statistiche sono gli occupati del Veneto, considerati nel periodo 1996-2005. I caratteri rilevati sono l'*Anno* (carattere qualitativo ordinale, modalità: 1996, 1997, ..., 2005) e il *Settore di attività economica* (carattere qualitativo sconnesso, modalità: Agricoltura, Industria, Servizi).

b) La distribuzione doppia di frequenza riporta il numero di unità statistiche (occupati) per Anno e per Settore di attività economica: è data dalla prima colonna della Tavola, dall'unione della seconda, e terza colonna (= quarta colonna, *Totale*) e infine dalla quinta colonna della Tavola.

c) La distribuzione di frequenza assoluta del carattere Anno riporta il totale di unità statistiche per anno dal 1996 al 2005, senza distinzione per settore di attività economica in cui gli occupati sono distinti. La distribuzione di frequenza relativa si ottiene dividendo la frequenza assoluta per il totale delle unità statistiche,  $N$ . Se indichiamo con  $X$  il carattere Anno, che assume 10 modalità distinte e con  $f_j$  e  $p_j$  la frequenza assoluta e la frequenza relativa della  $j$ -esima modalità,  $j = 1, \dots, 10$ , si ottiene

$X$	$f_j$	$p_j$
1996	96 + 738 + 968 = 1802	1802/N=0,0935
1997	1825	0,0947
1998	1841	0,0955
1999	1867	0,0968
2000	1917	0,0994
2001	1942	0,1007
2002	1954	0,1013
2003	2028	0,1052
2004	2042	0,1059
2005	2063	0,1070
	$N=19281$	1

In modo analogo per il carattere  $Y$ , *Settore di attività economica*, si ha

$Y$	$f_j$	$p_j$
Agricoltura	900	0,0467
Industria	7724	0,4006
Servizi	10657	0,5527
	$N=19281$	1

#### Esercizio D.

a) La distribuzione doppia di frequenza richiesta si costruisce contando il numero di unità statistiche che cadono nelle classi individuata per Popolazione e Superficie:

Superficie	Popolazione			Totale
	0-300	300-1000	1000 e oltre	
0-50	12	16	2	30
50-150	9	2	0	11
150-300	2	0	0	2
300 e oltre	0	1	0	1
Totale	23	19	2	44

b) Per la distribuzione di frequenza richiesta si consideri la tabella:

Superficie	$f_i$	$p_i = f_i/N$	$p_i \times 100$
0-50	30	0,6818	68,18
50-150	11	0,2500	25,00
150-300	2	0,0455	4,55
300 e oltre	1	0,0227	2,27
Totale	44	1	100

c) La distribuzione di quantità richiesta si ottiene sommando le superfici dei Comuni che appartengono alle diverse classi:

Superficie	$f_i$	Superficie
0-50	30	701,65
50-150	11	1012,28
150-300	2	336,62
300 e oltre	1	415,94
Totale	44	2466,49

**Esercizio E.**

a) La distribuzione di frequenza richiesta è data da:

Età	$f_i$
meno di 20	853437
[20 – 40)	1358196
[40 – 70)	1855662
70 e oltre	632655
Totale	4699950

b) Per disegnare l'istogramma di frequenza e la funzione di ripartizione consideriamo le quantità riportate nella seguente tabella.

Età	$p_i$	$d_i$	$h_i$	$P_i$
(0 – 20)	0,1816	20	0,00908	0,1816
[20 – 40)	0,2890	20	0,01445	0,2890 + 0,1816 = 0,4706
[40 – 70)	0,3948	30	0,01316	0,3948 + 0,2890 + 0,1816 = 0,8654
[70 – 101)	0,1346	31	0,00434	1
Totale	1	–	–	–

Nella tabella sono riportate le modalità del carattere  $Età$ , la frequenza relativa  $p_i$ , la numerosità dell'intervallo di  $Età$  considerato  $d_i$ , la densità di frequenza per l'intervallo considerato  $h_i$  calcolata come  $p_i/d_i$  e infine la frequenza relativa cumulata  $P_i$ .

Per disegnare l'istogramma di frequenza rappresentiamo 4 rettangoli (tanti quante sono le modalità del carattere  $Età$ ), ciascuno di larghezza pari a  $d_i$  e altezza pari a  $h_i$ .

Consideriamo ora la rappresentazione della funzione di ripartizione del carattere  $X=Età$ . Indicando con  $x_j$  le (4) modalità del carattere  $Età$ , la funzione di ripartizione  $F_X(x_j)$  è data dal numero delle determinazioni di  $X$  minori o uguali di  $x_j$ , rapportato al totale delle frequenze. Ciò significa che assume i valori riportati nella colonna  $P_i$ . Per rappresentare graficamente la funzione di ripartizione basta congiungere i seguenti punti con dei segmenti di retta: (0; 0), (20; 0,1816), (40; 0,4706), (70; 0,8654), (101; 1). Ovviamente, la funzione di ripartizione è nulla per valori di  $X$  minori di 0, ed uguale a 1 per valori di  $X$  maggiori di 101.

**Esercizio F.** Consideriamo la distribuzione di frequenza assoluta e relativa per il carattere di interesse (carattere qualitativo sconnesso)

Località	$f_i$	$p_i$
Balneare	805	0,2619
Lacuale	441	0,1435
Termale	156	0,0507
Montane	596	0,1939
Città d'arte	1076	0,3500
	$N=3074$	1

a) L'areogramma si costruisce dividendo un cerchio in settori circolari, in numero pari alle modalità del carattere in esame (5). Ogni settore circolare ha un angolo calcolabile come  $a_i = p_i * 360$ . Nel caso in esame: 94, 284; 51, 660; 18, 252; 69, 804; 126, 000.

b) L'istogramma di frequenza si rappresenta tramite rettangoli (di uguale ampiezza) o barre o bastoncini in numero pari alle modalità del carattere (5) e di altezza pari alla frequenza assoluta della modalità.

**Esercizio G.**

Utilizzando le proprietà delle sommatorie si ottiene:

$$\text{a) } \sum_{i=1}^5 (a + b \cdot i) = \sum_{i=1}^5 a + \sum_{i=1}^5 b \cdot i = 5a + b \sum_{i=1}^5 i = 5a + b(1 + 2 + 3 + 4 + 5) = 5a + 15b;$$

$$\text{b) } \sum_{i=1}^4 (a + b \cdot i + c \cdot i^2) = \sum_{i=1}^4 a + b \sum_{i=1}^4 i + c \sum_{i=1}^4 i^2 = 4a + 10b + 30c;$$

$$\text{c) } \sum_{i=1}^5 \sum_{j=1}^4 (a \cdot i + b \cdot j) = \sum_{i=1}^5 \left[ \sum_{j=1}^4 a \cdot i + \sum_{j=1}^4 b \cdot j \right] = \sum_{i=1}^5 [4ai + 10b] = 60a + 50b;$$

$$\text{d) } \sum_{i=1}^4 \sum_{j=1}^i i \cdot j = \sum_{i=1}^4 i \sum_{j=1}^i j = 1(1) + 2(1 + 2) + 3(1 + 2 + 3) + 4(1 + 2 + 3 + 4) = 65.$$