

Università di Cassino
Corso di Statistica 1
Esercitazione del 14/01/2008
Dott. Alfonso Piscitelli

Esercizio 1

Supponiamo che il collettivo che si vuole studiare sia composto da un gruppo di amici. La seguente tabella raccoglie alcuni dati su di essi.

Soggetto	Età	Residenza	Cittadinanza	Reddito (Migliaia di €)	Auto Possedute	Punteggio quiz	Grado di soddisfazione
1	22	Cantagallo	Estera	0,7	3	173	Poco
2	18	Cantagallo	Italiana	0,2	4	168	Abbastanza
3	34	Poggio a Caiano	Italiana	1,6	2	165	Molto
4	42	Carmignano	Estera	2,5	5	180	Poco
5	50	Poggio a Caiano	Italiana	3,2	3	163	Poco
6	12	Montemurlo	Italiana	0,1	4	160	Abbastanza
7	46	Carmignano	Estera	3,8	4	177	Molto
8	72	Montemurlo	Estera	1,3	2	164	Molto
9	27	Montemurlo	Italiana	1,2	3	158	Abbastanza
10	48	Carmignano	Italiana	1,7	5	170	Poco
11	35	Montemurlo	Italiana	1,9	1	167	Abbastanza
12	84	Cantagallo	Estera	0,8	1	159	Molto
13	21	Montemurlo	Italiana	0,4	5	174	Abbastanza
14	44	Carmignano	Italiana	1,8	4	164	Poco
15	56	Carmignano	Estera	1,9	2	177	Poco
16	58	Montemurlo	Italiana	3,2	3	172	Abbastanza
17	37	Cantagallo	Italiana	2,1	1	166	Poco
18	16	Montemurlo	Italiana	0,1	4	160	Abbastanza
19	73	Carmignano	Estera	1,6	2	170	Abbastanza
20	64	Poggio a Caiano	Estera	2,2	3	184	Poco

a. Qual è l'unità statistica? Quante unità statistiche sono presenti?

b. Quali sono i caratteri? Per ogni carattere indicare se è quantitativo, discreto o continuo, su scala ad intervalli o su scala di rapporti, oppure qualitativo sconnesso o ordinato.

c. Costruire le distribuzioni di frequenza per i caratteri **Residenza**, **Cittadinanza** e **Auto Possedute**.

d. Rappresentare graficamente le distribuzioni di frequenza dei caratteri **Residenza** e **Cittadinanza** utilizzando un diagramma a barre e un diagramma di Pareto per il carattere **Residenza** e un aerogramma per il carattere **Cittadinanza**.

e. Costruire la distribuzione di frequenza per il carattere **Età** suddividendo la distribuzione in 6 classi equiampie.

f. Rappresentare graficamente la distribuzione equiampia.

Soluzioni

- a) L'unità statistica è rappresentata da ogni individuo, appartenente al collettivo oggetto di studio, sul quale sono state registrate le informazioni oggetto dell'indagine. Sono presenti 20 unità statistiche.
- b) I caratteri sono l'oggetto dell'indagine, le *caratteristiche* che si intendono rilevare sul collettivo di individui. I caratteri possono essere **qualitativi** (sconnesso o ordinabile) o **quantitativi** (discreto o continuo).

Carattere	Tipo	Livello di misura	MODALITÀ-INTENSITÀ
Cittadinanza	Qualitativa sconnessa	Scala Nominale	{Italiana, Estera}
Età	Quantitativa continua	Scala di Rapporti	[12; 84]
Reddito	Quantitativa continua	Scala di Rapporti	[0,1; 3,8]
Grado di soddisfazione	Qualitativa ordinabile	Scala Ordinale	{Poco, Abbastanza, Molto}
Auto Possedute	Quantitativa discreta	Scala di Rapporti	[1; 5]
Punteggio quiz	Quantitativa continua	Scala di Rapporti	[158; 184]
Residenza	Qualitativa sconnessa	Scala Nominale	{Cantagallo, Carmignano, Montemurlo, Poggio a Caiano}

- c) La **frequenza assoluta** non è altro che il numero di volte che una determinata intensità o modalità si presenta nel collettivo statistico. In altre parole la frequenza n_i associata alla i -esima intensità (modalità) è uguale al numero di unità statistiche che presentano quella intensità (modalità).

La **distribuzione di frequenza semplice** associa alle modalità o intensità che può assumere un carattere X , le corrispondenti frequenze assolute. Per costruire una distribuzione di frequenza semplice partendo da una successione di osservazioni è sufficiente assegnare ad ogni intensità (o modalità, nel caso di variabili qualitative) la propria frequenza assoluta.

Le **frequenze relative** f_i , invece, sono date dal rapporto tra la frequenza assoluta, n_i , ed il totale delle osservazioni, N .

Quindi la distribuzione di frequenze del carattere (variabile) **Cittadinanza** è:

Cittadinanza	n_i	f_i
Estera	8	0,4
Italiana	12	0,6

Tot: 20 1

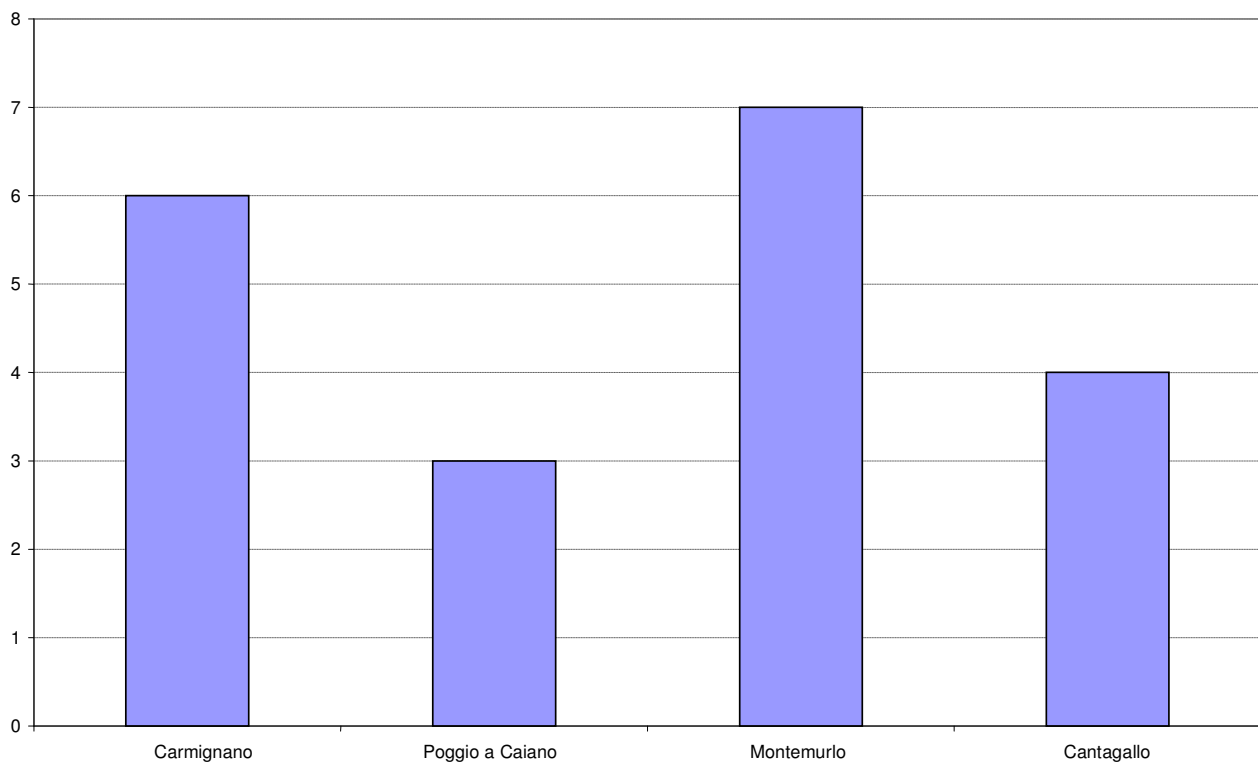
Quella associata alla variabile **Residenza** è:

Residenza	n_i	f_i
Carmignano	6	0,3
Poggio a Caiano	3	0,15
Montemurlo	7	0,35
Cantagallo	4	0,2
Tot:	20	1

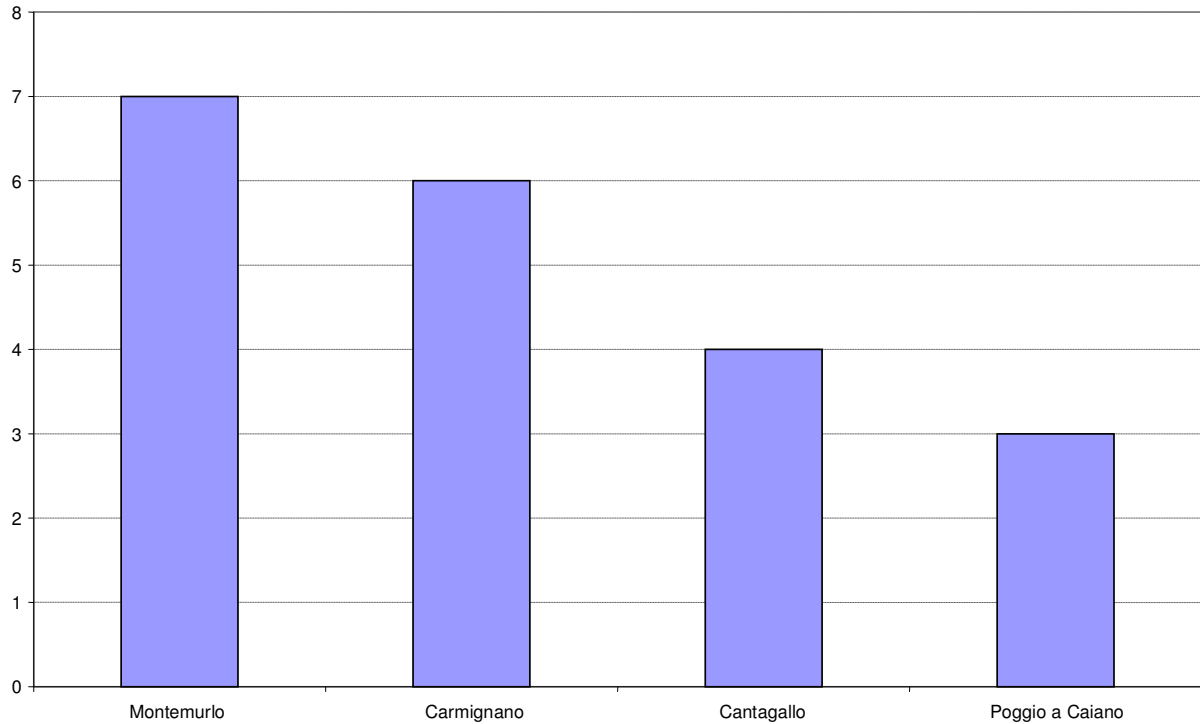
Infine, la distribuzione di frequenze della variabile **Auto Possedute** è:

Auto Possedute	n_i	f_i
1	3	0,15
2	4	0,2
3	5	0,25
4	5	0,25
5	3	0,15
Tot:	20	1

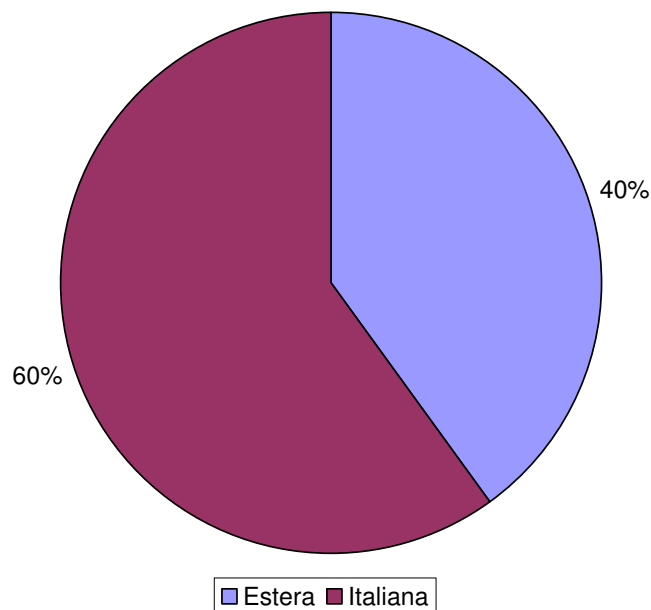
d) Il **diagramma a barre** rappresenta la frequenza o la frequenza relativa di una tabella di frequenza sottoforma di un rettangolo o barra o colonna.



Il **diagramma di Pareto** presenta le stesse caratteristiche del diagramma a barre con l'unica differenza nell'ordine delle barre: le modalità sono ordinate in senso decrescente rispetto alle frequenze



Un **aerogramma** o **diagramma a torta** rappresenta i dati sottoforma di fette o sezioni di un cerchio. Ogni fetta rappresenta una categoria e la dimensione della fetta è proporzionale alla frequenza relativa della categoria



e) Nella tabella sono riportati i valori della variabile **Età**.

Osservazione	Età
1	22
2	18
3	34
4	42
5	50
6	12
7	46
8	72
9	27
10	48
11	35
12	84
13	21
14	44
15	56
16	58
17	37
18	16
19	73
20	64

Al fine di pervenire a delle classi equampie è necessario individuare il campo di variazione della variabile che si vuole suddividere in classi ed in seguito dividerlo per il numero di classi desiderate.

In questo caso, l'ampiezza delle classi è ottenuta come il rapporto tra il campo di variazione (il valore massimo della distribuzione - il valore minimo) e il numero di classi:

$$d = \frac{X_{max} - X_{min}}{K}$$

dove **K** = numero di classi.

Nel nostro caso: $x_{max} = 84$ $X_{min} = 12$ **K** = 6 \Rightarrow **d** = 12

Una volta individuato quanto devono essere ampie le classi si può procedere alla costruzione della distribuzione per classi.

Partendo dal valore minimo (X_{min}) si aggiunge l'ampiezza di classe

Età	d_i	n_i	f_i	F_i	h_i
12 - 24	12	5	0,25	0,25	0,42
24 - 36	12	3	0,15	0,4	0,25
36 - 48	12	4	0,2	0,6	0,33
48 - 60	12	4	0,2	0,8	0,33
60 - 72	12	1	0,05	0,85	0,08
72 - 84	12	3	0,15	1	0,25

Tot: 20 1

Dove:

n_i = frequenze assolute

f_i = frequenze relative

F_i = frequenze cumulate relative

h_i = densità di frequenza, cioè il rapporto tra la frequenza della i -esima

classe e la sua ampiezza: $h_i = \frac{n_i}{d_i}$.

Si precisa che nel caso di classi equampie l'informazione ottenuta dalla densità di frequenza è la stessa che avremmo analizzando la distribuzione di frequenze semplice.

f) Sulla base dei risultati ottenuti nella precedente tabella è possibile costruire **l'istogramma** associato alla nostra distribuzione di frequenze.

L'istogramma normalizzato presenta sulle ascisse gli estremi delle classi e sulle ordinate la densità di frequenza. Essendo le classi di ampiezza uguale le basi dei rettangoli associati alle diverse classi avranno tutte la stessa lunghezza.

