

ESERCITAZIONE 1: VARIABILI E DISTRIBUZIONI

1. TIPOLOGIA DEI DATI
2. CALCOLO DI FREQUENZE
3. RAPPRESENTAZIONE GRAFICA DI UNA VARIABILE

A Roma nel 2006 è stata effettuata un'indagine, tramite questionario, per conoscere gli atteggiamenti dei titolari di esercizi commerciali nei confronti dell'immigrazione straniera. Di seguito si riportano alcune delle informazioni rilevate limitatamente ai primi 40 questionari raccolti.

Data set relativo a 40 titolari di esercizi commerciali

Variabile	Descrizione Data set
Collettività indicata	Tipo di collettività
Sesso	Maschio (= 1), Femmina (= 2)
Anno di nascita	Anno di nascita
Titolo di studio	Formazione scolastica ricevuta, dove 2 = Elementare, 3 = Media inferiore, 4 = Media superiore e 5 = Laurea
Durata dell'intervista	Numero di minuti in cui si è svolta l'intervista
Durata del percorso casa-lavoro	Durata del percorso casa lavoro (in minuti)
Numero addetti	Numero dipendenti che lavorano nel negozio

Data set relativo a 40 titolari di pubblici esercizi. Roma, 2006

Numero questionario	Collettività indicata	Sesso	Anno di nascita	Titolo di studio	Durata dell'intervista	Durata percorso casa-lavoro	Numero addetti
1	tutte	1	1972	4	18	26	3
2	romeni	2	1943	3	30	11	13
3	cinesi	2	1952	4	20	4	5
4	cinesi	2	1944	4	20	10	5
5	cinesi	2	1970	3	15	14	1
6	romeni	1	1975	4	10	18	1
7	africani	2	1959	4	10	8	4
8	albanesi	2	1939	3	15	12	2
9	cinesi	2	1935	4	15	20	3
10	tutte	1	1951	3	20	32	5
11	cinesi	2	1977	4	10	15	5
12	cinesi	2	1945	3	15	14	4
13	cinesi	1	1939	4	20	21	4
14	polacchi	2	1945	3	15	26	6
15	tutte	2	1964	4	25	18	4
16	cinesi	1	1956	4	20	12	3
17	cinesi	2	1942	4	20	4	5
18	senegalesi	1	1977	4	15	10	1
19	cinesi	2	1930	3	10	6	3
20	asiatici	1	1977	4	15	15	1
21	marocchini	1	1946	4	20	30	2
22	africani	1	1940	3	15	14	6
23	tutte	1	1949	5	30	22	4
24	tutte	2	1962	4	45	8	4
25	africani	2	1962	4	15	15	4
26	tutte	2	1967	3	20	12	4
27	romeni	2	1970	4	20	16	1
28	cinesi	2	1972	4	10	14	2
29	romeni	1	1974	4	20	28	4
30	cinesi	2	1951	4	10	15	4
31	tutte	1	1967	4	20	22	3
32	romeni	1	1940	3	20	15	5
33	tutte	1	1973	4	18	20	1
34	africani	2	1940	2	16	4	4
35	tutte	1	1977	3	16	30	5
36	marocchini	1	1968	4	15	44	6
37	romeni	1	1962	4	15	28	1
38	tutte	2	1938	2	12	15	8
39	tutte	2	1943	2	15	15	8
40	romeni	2	1947	3	10	11	6

Fonte: indagine sull'atteggiamento dei commercianti romani nei confronti degli immigrati, giugno-luglio 2006.

ESERCIZIO 1

La tabella sopra presentata contiene i dati relativi a 40 titolari di esercizi commerciali

- 1) Commentare brevemente la tabella (unità statistiche, tipo di tabella, tipo di variabile);
- 2) Costruire le distribuzioni di frequenze assolute, relative e le percentuali corrispondenti ed eventualmente cumulate (assolute, relative e le percentuali corrispondenti) per le Variabili:
 - Sesso
 - Titolo di studio
 - Numero addetti
- 3) Rappresentare graficamente le variabili del punto 2.

Soluzione

1)

La tabella è una classica matrice dei dati (**unità x variabili**). In particolare la matrice ha dimensioni **40 righe x 7 colonne**. La prima colonna contiene un identificativo numerico dei negozianti intervistati e non può essere utilizzato ai fini di analisi successive dei dati). Le unità statistiche su cui è effettuata la rilevazione sono i 40 titolari di pubblici esercizi .

Tipologia delle variabili presenti nel dataset

	Qualitativo nominale	Qualitativo ordinale	Quantitativo discreto	Quantitativo continuo
Sesso	■	□	□	□
Durata percorso casa- lavoro	□	□	□	■
Durata intervista	□	□	□	■
Numero addetti	□	□	■	□
Collettività indicata	■	□	□	□
Titolo di studio conseguito	□	■	□	□

2)

Le frequenze assolute permettono di osservare le variazioni numeriche di un fenomeno di interesse, ma ciascuna frequenza dipende in primo luogo dall'ammontare complessivo delle unità statistiche osservate. Al fine di confrontare le variazioni nella composizione della variabile è necessario costruire le frequenze relative o le percentuali. Queste misure esprimono le frequenze assolute normalizzate in modo che il totale valga 1 (oppure 100). In particolare esse vengono calcolate, rispettivamente, come:

$$\text{frequenza relativa} = \text{frequenza assoluta} / \text{totale}$$
$$\text{percentuale} = \text{frequenza assoluta} / \text{totale} \times 100$$

Per calcolare le distribuzioni di frequenze cumulate, la variabile deve essere almeno misurata su una scala di tipo ordinale. Per variabili qualitative nominali è, invece, possibile calcolare solo le frequenze assolute, relative e le percentuali.

Per variabili quantitative continue (e a volte anche per variabili quantitative discrete) le distribuzioni di frequenze semplici possono consistere nell'elenco delle modalità osservate della variabile. Pertanto è opportuno ricorrere ad una distribuzione in classi della variabile specificando se i valori estremi della classe sono o meno compresi nella classe stessa.

Non esiste un unico criterio per raggruppare in classi le modalità di una variabile quantitativa continua. Le classi possono infatti presentarsi:

- **equi-ampie**

- si stabilisce l'ampiezza delle classi; in tal caso il numero delle classi dipende dall'ampiezza scelta;
- si stabilisce il numero di classi; in tal caso il numero delle classi dipende dal numero delle classi.

$$\text{Ampiezza} = (\text{Valore massimo} - \text{Valore minimo}) / \text{Numero classi}$$

- **equi-frequenti**

si stabilisce una frequenza uguale per tutte le classi; in tal caso è necessario procedere all'ordinamento dei dati per poi passare agevolmente alla suddivisione in classi.

N.B. Per alcune distribuzioni, tali criteri possono risultare inadatti perché non riescono a rappresentare al meglio la variabile misurata. Ciò si verifica in presenza di distribuzioni in cui le frequenze sono concentrate su poche modalità. In tal caso, è più opportuno definire classi di diversa ampiezza, non necessariamente equi-frequenti.

3)

Il tipo di grafico più adatto per rappresentare la distribuzione di una variabile dipende dal livello di misurazione della variabile in esame.

Per variabili qualitative, si possono costruire i diagrammi a barre (a nastri) utilizzando i dati originari o le frequenze percentuali. In alternativa si utilizzano i grafici a torta.

Per variabili quantitative discrete, con un ridotto numero di modalità, è possibile costruire un diagramma a barre. In generale, per variabili quantitative distribuite in classi è possibile costruire un istogramma.

Esempio 1

Sesso: Variabile qualitativa nominale

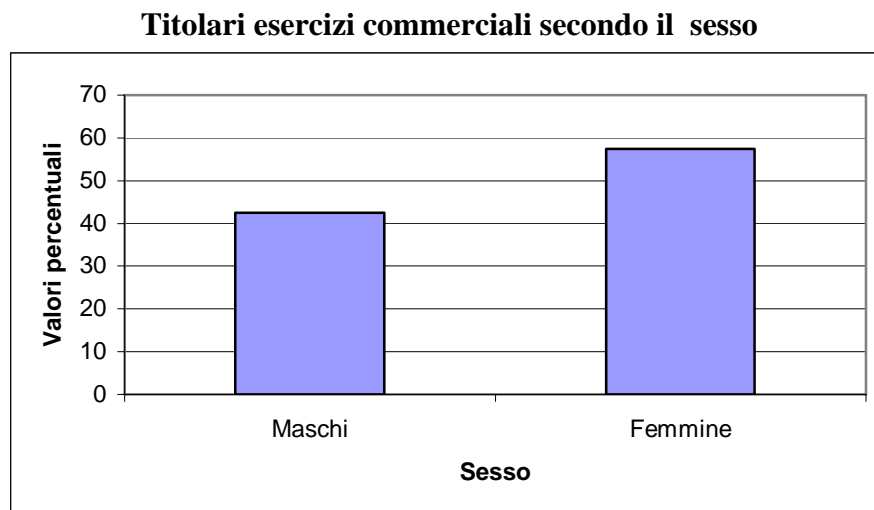
Calcolo Distribuzioni di frequenze assolute, relative e le percentuali

Individuiamo le $k = 2$ modalità assunte dal carattere Sesso, a cui associamo le **frequenze assolute**, le **frequenze relative** (dividere quelle assolute per il totale calcolato), le **percentuali** corrispondenti (moltiplicando per 100 le frequenze relative).

Sesso	Freq. Ass.	Freq. Rel.	%
Maschi	17	0,425	42,5
Femmine	23	0,575	57,5
Totale	40	1,000	1000

N. B. Essendo una variabile qualitativa nominale, non è possibile calcolare le frequenze cumulate.

Rappresentazione grafica – Diagramma a barre con le percentuali corrispondenti alle 2 modalità assunte dalla variabile Sesso.



Esempio 2

Titolo di studio conseguito : **Variabile qualitativa ordinale**

Calcolo Distribuzioni di frequenze assolute, relative e le percentuali

Individuiamo le $k = 4$ modalità assunte dal carattere Titolo di studio (dove 2= Elementare, 3= Media inferiore 4 = Media superiore e 5= Laurea) a cui associamo la **frequenza assoluta, relativa e le percentuali** corrispondenti.

Titolo di studio conseguito	Freq. Ass.	Freq. Rel.	%
Elementare	3	0,075	7,5
Media inferiore	12	0,300	30
Media superiore	24	0,600	60
Laurea	1	0,025	25
<i>Totale</i>	40	1,000	100

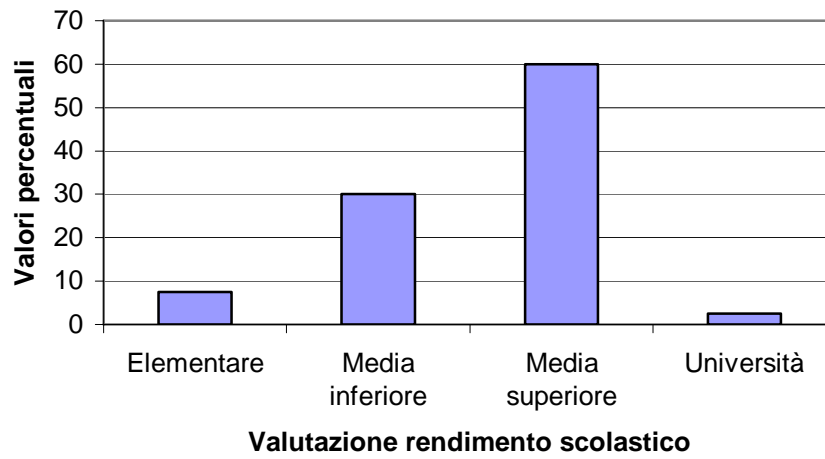
Calcolo Distribuzioni di frequenze cumulate assolute, relative e le percentuali corrispondenti.

Titolo di studio conseguito	Freq. Ass.	Freq. Ass. Cum
Elementare	3	3
Media inferiore	12	15
Media superiore	24	39
Laurea	1	40
<i>Totale</i>	40	

Titolo di studio conseguito	Freq. Rel.	Freq.Rel. Cum.	%	% Cum.
Elementare	0,075	0,075	7,5	7,5
Media inferiore	0,300	0,375	30	37,5
Media superiore	0,600	0,975	60	97,5
Laurea	0,025	1,000	25	100
<i>Totale</i>	1,000		100	

Rappresentazione grafica – Diagramma a barre con le percentuali corrispondenti alle 4 modalità assunte dalla variabile *Titolo di studio conseguito*.

Titolari di pubblici esercizi secondo il titolo di studio conseguito



Esempio 3

Numero addetti : Variabile quantitativa discreta

Calcolo Distribuzioni di frequenze assolute, relative e frequenze relative cumulate

Numero addetti	n_i	N_i	f_i	F_i
1	7	7	0,175	0,175
2	3	10	0,075	0,250
3	5	15	0,125	0,375
4	11	26	0,275	0,650
5	7	33	0,175	0,825
6	4	37	0,100	0,925
8	2	39	0,050	0,975
13	1	40	0,025	1
Totale	40		1,000	

Per il raggruppamento in classi di tale variabile si potrebbero individuare classi di diversa ampiezza (non necessariamente equi-frequenti), seguendo criteri logici nell'individuazione delle stesse.

Classi di diversa ampiezza

Numero addetti	n_i
1 — 3	15
4 — 5	18
6 — 13	7
Totale	40