

Tabelle doppie e analisi della relazione tra variabili



SAPIENZA
UNIVERSITÀ DI ROMA

annarita.vestri@uniroma1.it

In generale una tabella 2x2 riassume la frequenza di eventi sanitari (o altro) tra differenti gruppi.

Es gruppo 1 rappresenta pazienti che ricevono terapia standard e gruppo 3 pazienti che ricevono terapia sperimentale.

	Event		
	NO	YES	ALL
Group 1	a	b	a+b
Group 2	c	d	c+d
	a+c	b+d	N

La variabile di outcome primario nel trial Kawasaki trials è lo sviluppo di anomalie delle arterie coronariche (CA), variabile dicotomica. Un obiettivo dello studio è confrontare la probabilità di sviluppare anomalie CA dato il trattamento con aspirina (ASA) o gamma globuline (GG).

Anormalità CA

trattamento	No	Si	totale
ASA	63	21	84
GG	78	5	83
totale	141	26	167

Possiamo essere interessati a conoscere se la frequenza di anomalie alle coronarie è differente tra i due gruppi, cioè se uno di questi è associato con minori anomalie.

Anormalità CA

trattamento	No	Si	totale
ASA	63	21	84
GG	78	5	83
totale	141	26	167

MODALITA'

TABELLA DI CONTINGENZA

Anormalità CA

trattamento	No	Si	totale
ASA	63	21	84
GG	78	5	83
totale	141	26	167

Celle contenenti le frequenze osservate

TABELLA DI CONTINGENZA

Anormalità CA

trattamento	No	Si	totale
ASA	63	21	84
GG	78	5	83
totale	141	26	167

Totali marginali di riga e di colonna

TABELLA DI CONTINGENZA

Anormalità CA

trattamento	No	Si	totale
ASA	63	21	84
GG	78	5	83
totale	141	26	167

Totale generale

TABELLA DI CONTINGENZA

Frequenze relative percentuali di riga

Anormalità CA

trattamento	No	Si	totale
ASA	0,75 (63/84)	0,25	84
GG	0,94 (78/83)	0,06	83
totale	141	26	167

TABELLA DI CONTINGENZA

Frequenze relative percentuali di colonna

Anormalità CA

trattamento	No	Si	totale
ASA	0,45 (63/141)	0,81	84
GG	0,55 (78/141)	0,19	83
totale	141	26	167

TABELLA DI CONTINGENZA

Frequenze relative percentuali totali

Anormalità CA

trattamento	No	Si	totale
ASA	0,38 (63/167)	0,13 (21/167)	84
GG	0,47	0,03	83
totale	141	26	167

Esempio: distribuzione dell'abitudine al fumo e della broncopneumopatia cronico-ostruttiva (BPCO) in adulti italiani di età 20-44 anni (indagine ISAYA).

Fumo	n_i	p_i
non fumatore	9667	51.9%
ex fumatore	2743	14.7%
fumatore	6228	33.4%
Totale	18638	100.0%

BPCO	n_i	p_i (%)
assente	16622	89.2%
presente	2016	10.8%
Totale	18638	100.0%

distribuzione bivariata: DISTRIBUZIONE CONGIUNTA ASSOLUTA

distribuzione congiunta
del fumo e della BPCO
(n_{ij})

FUMO	BPCO		TOTALE
	assente	presente	
non fumatore	9042	625	9667
ex fumatore	2472	271	2743
fumatore	5108	1120	6228
TOTALE	16622	2016	18638

distribuzione
marginale del fumo
($n_{i\cdot}$)

dimensione
campionaria (n)

distribuzione
marginale della BPCO
($n_{\cdot j}$)

DISTRIBUZIONE BIVARIATA

(CROSS-TABULATION)



Permette la rappresentazione congiunta della distribuzione di frequenza di due variabili qualitative



Permette di capire la relazione tra le due variabili

DISTRIBUZIONE CONGIUNTA RELATIVA (%)

FUMO	BPCO		TOTALE
	assente	presente	
non fumatore	9042	625	9667
ex fumatore	2472	271	2743
fumatore	5108	1120	6228
TOTALE	16622	2016	18638

non fumatori con BPCO (n_{12})

dimensione campionaria (n)

$(625 / 18638) * 100$



$(n_{ij} / n) * 100$

FUMO	BPCO		TOTALE
	assente	presente	
non fumatore	48.5%	3.4%	51.9%
ex fumatore	13.3%	1.4%	14.7%
fumatore	27.4%	6.0%	33.4%
TOTALE	89.2%	10.8%	100.0%

distribuzione bivariata:
DISTRIBUZIONI CONDIZIONALI



*Rappresentano la **distribuzione** di
una variabile all'interno delle modalità
dell'altra variabile*

N.B. Se le distribuzioni condizionali sono differenti, si può supporre che esista una relazione tra le due variabili

DISTRIBUZIONI CONDIZIONALI AI MARGINALI DI RIGA: (distribuzione della BPCO per livello di fumo)

FUMO	BPCO		TOTALE
	assente	presente	
non fumatore	9042	625	9667
ex fumatore	2472	271	2743
fumatore	5108	1120	6228
TOTALE	16622	2016	18638

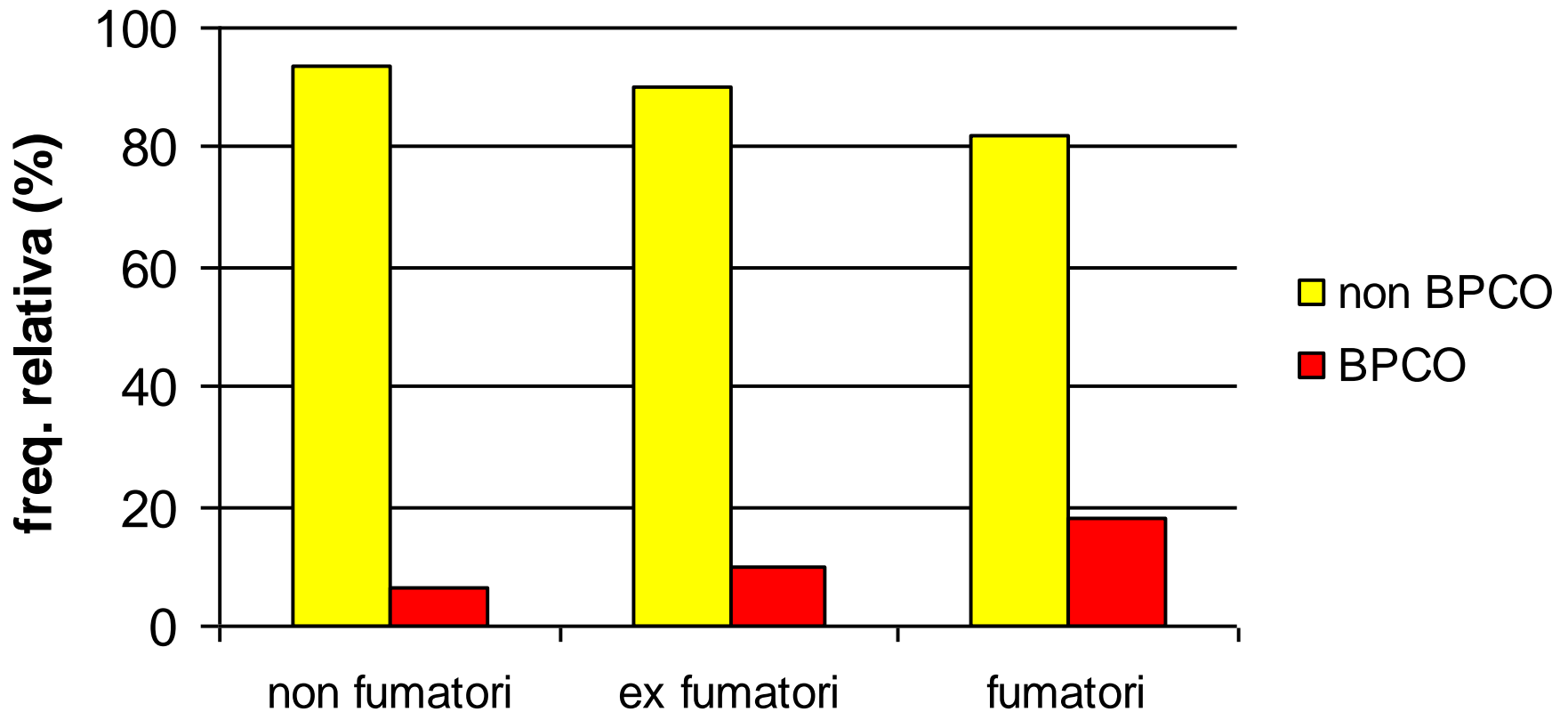
marginali di riga (n_i)

*$(625 / 9667) * 100$*

FUMO	BPCO		TOTALE
	assente	presente	
non fumatore	93.5%	6.5%	100.0%
ex fumatore	90.1%	9.9%	100.0%
fumatore	82.0%	18.0%	100.0%
TOTALE	89.2%	10.8%	100.0%

*$(n_{ij} / n_i) * 100$*

distribuzioni condizionali ai marginali di riga:
Distribuzione della BPCO in gruppi di
soggetti esposti a vari livelli di fumo attivo



DISTRIBUZIONI CONDIZIONALI AI MARGINALI DI COLONNA:
(distribuzione del fumo per livello della BPCO)

FUMO	BPCO		TOTALE
	assente	presente	
non fumatore	9042	625	9667
ex fumatore	2472	271	2743
fumatore	5108	1120	6228
TOTALE	16622	2016	18638



$$(625 / 2016) * 100$$

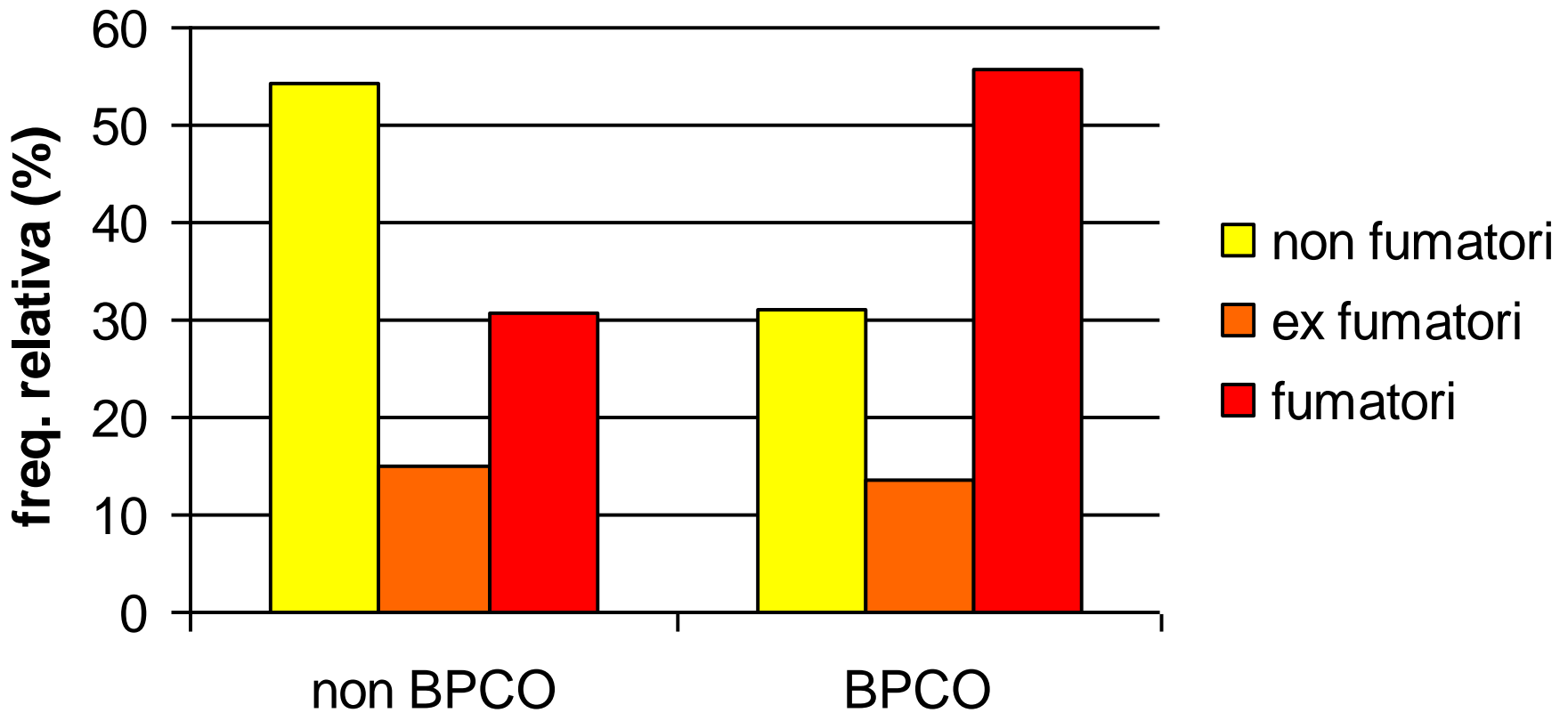
marginali di colonna
 (n_j)

FUMO	BPCO		TOTALE
	assente	presente	
non fumatore	54.4%	31.0%	51.9%
ex fumatore	14.9%	13.4%	14.7%
fumatore	30.7%	55.6%	33.4%
TOTALE	100.0%	100.0%	100.0%

$$(n_{ij} / n_j) * 100$$



distribuzioni condizionali ai marginali di colonna:
*Distribuzione dell'abitudine al fumo in soggetti
sani e in soggetti affetti da BPCO*



ESERCIZIO



In un'indagine è stato chiesto ad un gruppo di 101 consumatori e ad un gruppo di 124 dentisti se erano favorevoli alla pubblicità fatta dai dentisti per attrarre nuovi pazienti.

Si sono ottenuti i seguenti risultati:

GIUDIZIO

CATEGORIA	GIUDIZIO					TOTALE
	molto favorevole	abbastanza favorevole	indifferente	abbastanza sfavorevole	molto sfavorevole	
consumatore	34	49	9	4	5	101
dentista	9	18	23	28	46	124
TOTALE	43	67	32	32	51	225

- 1. C'è differenza tra il giudizio espresso dai consumatori e dai dentisti? C'è relazione tra la categoria e il giudizio?*
- 2. Risolvete l'esercizio calcolando le opportune distribuzioni condizionali e rappresentatele graficamente*

SOLUZIONE - I



→ confrontare le distribuzioni condizionali della variabile 'giudizio' tra i consumatori e i dentisti:

$$(34 / 101) * 100$$

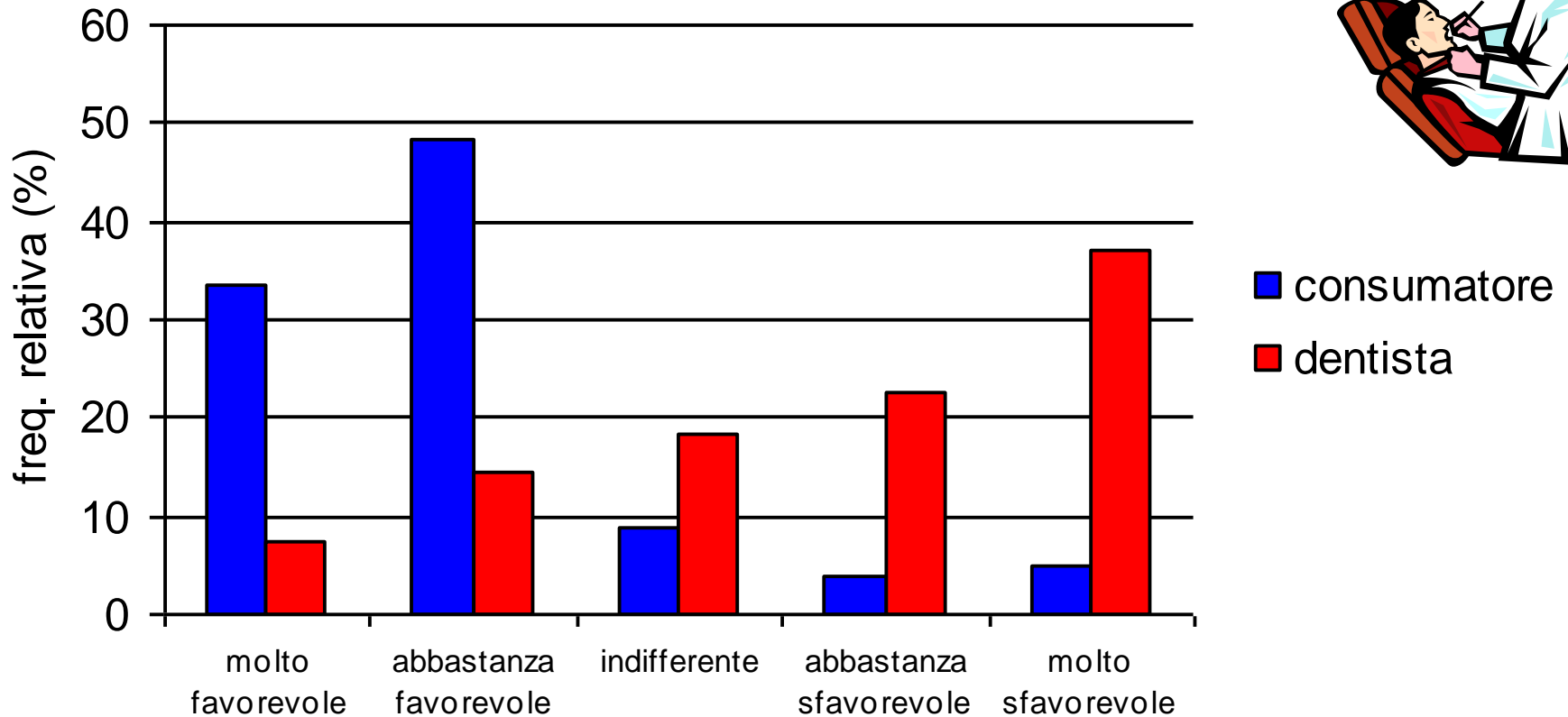
GIUDIZIO

CATEGORIA	GIUDIZIO					TOTALE
	molto favorevole	abbastanza favorevole	indifferente	abbastanza sfavorevole	molto sfavorevole	
consumatore	33.7%	48.5%	8.9%	4.0%	4.9%	100.0%
dentista	7.3%	14.5%	18.5%	22.6%	37.1%	100.0%
TOTALE	19.1%	29.8%	14.2%	14.2%	22.7%	100.0%

$$(9 / 124) * 100$$



SOLUZIONE - II



- 1. il giudizio sembra dipendere dalla categoria di appartenenza (le distribuzioni condizionali sono differenti)*
- 2. i consumatori sembrano essere favorevoli, mentre i dentisti no*

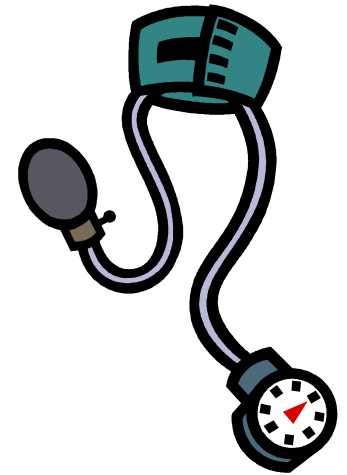
DISTRIBUZIONE CONGIUNTA di due VARIABILI QUANTITATIVE

Peso	Pressione arteriosa sistolica
55	100
58	90
60	98
61	95
61	108
61	89
61	90
62	97
62	96
62	110
64	95
64	97
65	105
65	104
65	113
66	98
66	101
66	91
67	107
67	112
67	94
68	119
69	102
69	102
69	118
69	109
70	114
70	106
71	100

Peso	Pressione Arteriosa sistolica
72	121
74	120
75	104
75	102
75	109
75	107
76	115
77	111
78	108
78	107
79	112
81	110
82	100
83	105
83	118
85	114
87	105
87	110
90	110
90	120
95	121
95	119
99	110
102	117
104	131
104	132
111	127
118	133

Esempio:

Consideriamo i dati relativi alla pressione sistolica arteriosa e al peso di 59 soggetti:

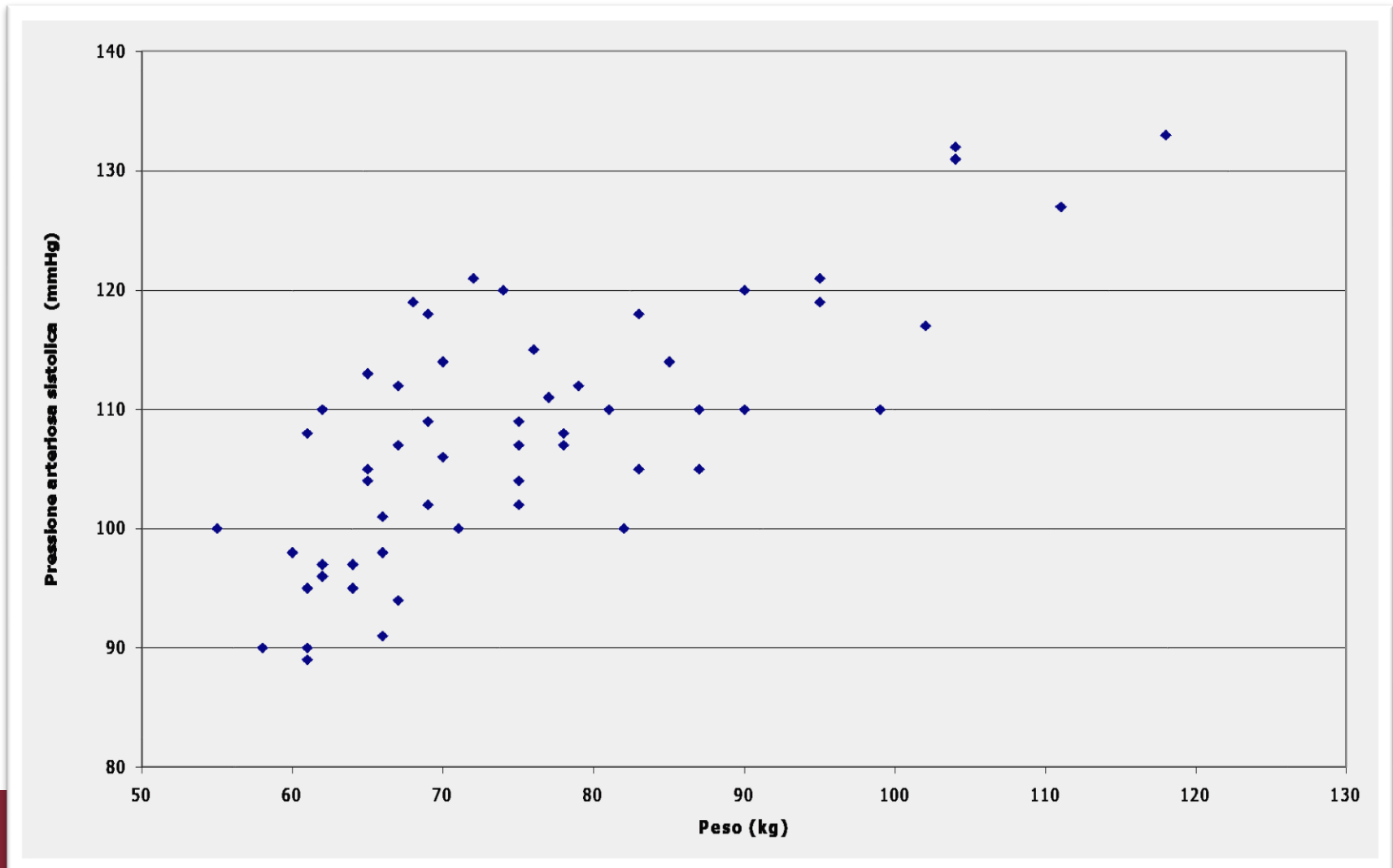


Quale relazione tra i dati?

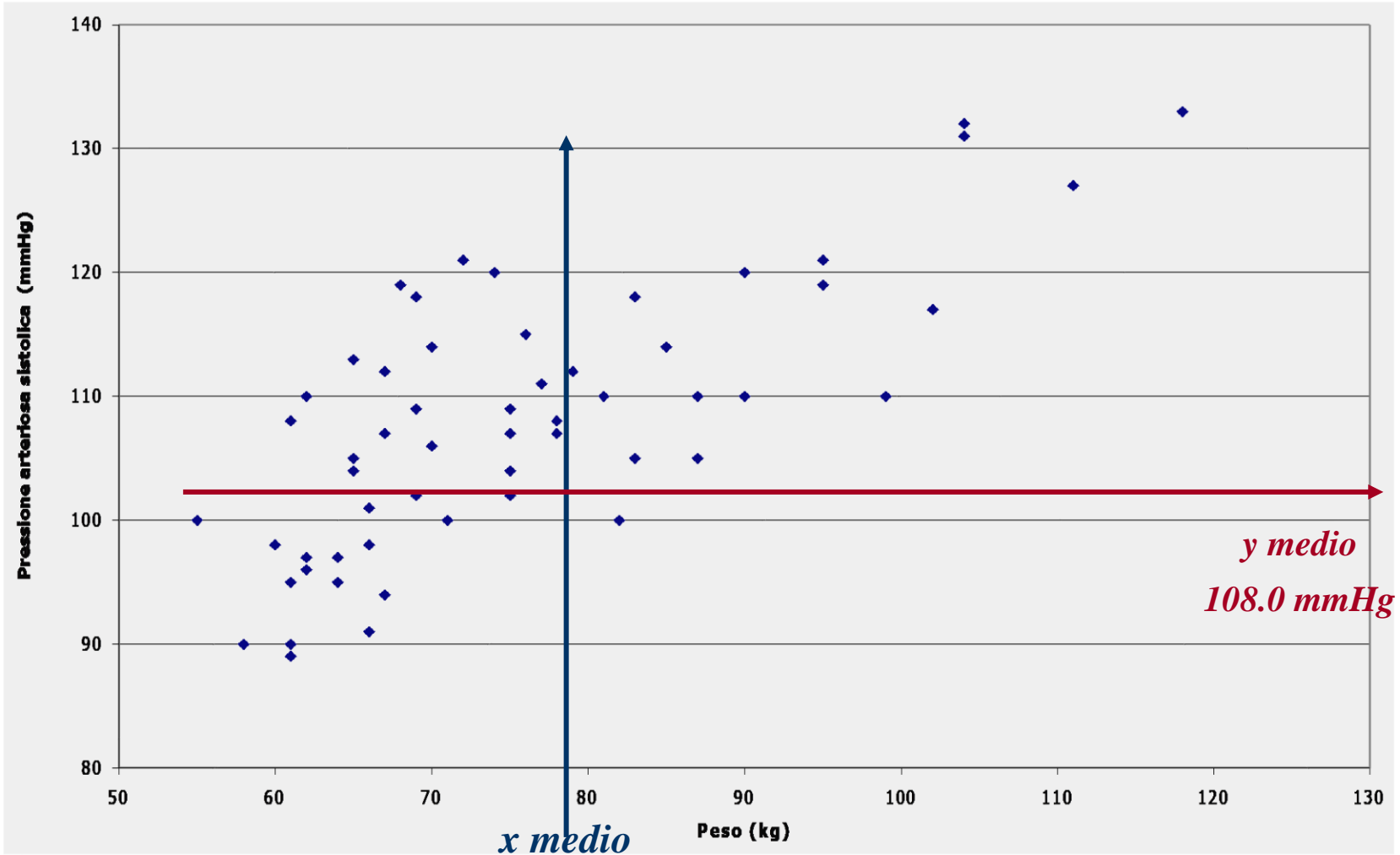
DIAGRAMMA DI DISPERSIONE

(scatterplot)

Riportiamo su un diagramma cartesiano in **ascissa (X)** i valori del peso e in **ordinata (Y)** i valori della pressione arteriosa



All'aumentare di x , y aumenta (associazione positiva)?
Diminuisce (associazione negativa)? Oppure tra x e y non sembra esserci alcuna relazione?



75.6 Kg

La maggior parte dei punti è concentrata nel I e nel III quadrante: l'associazione tra le due variabili è positiva.

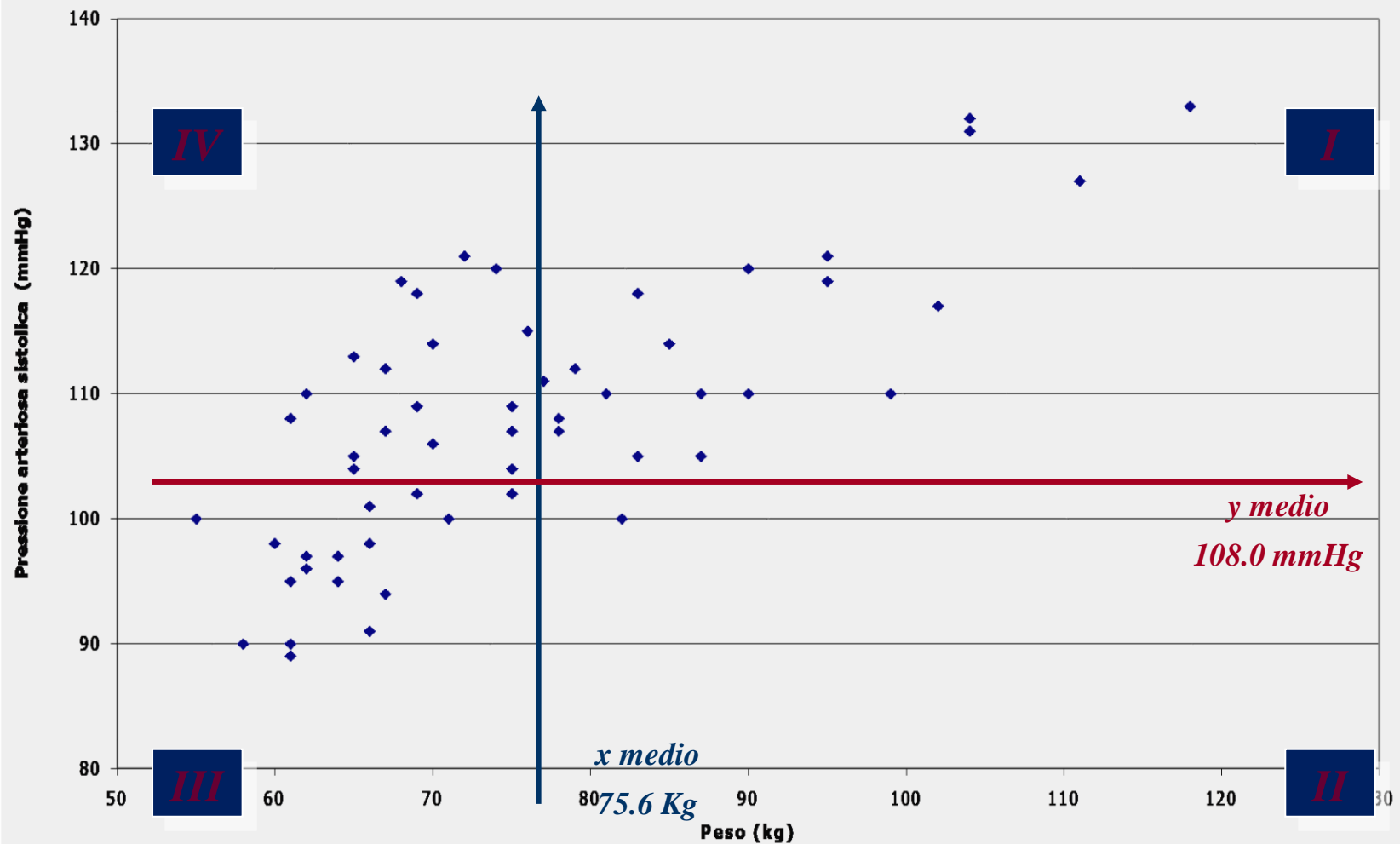
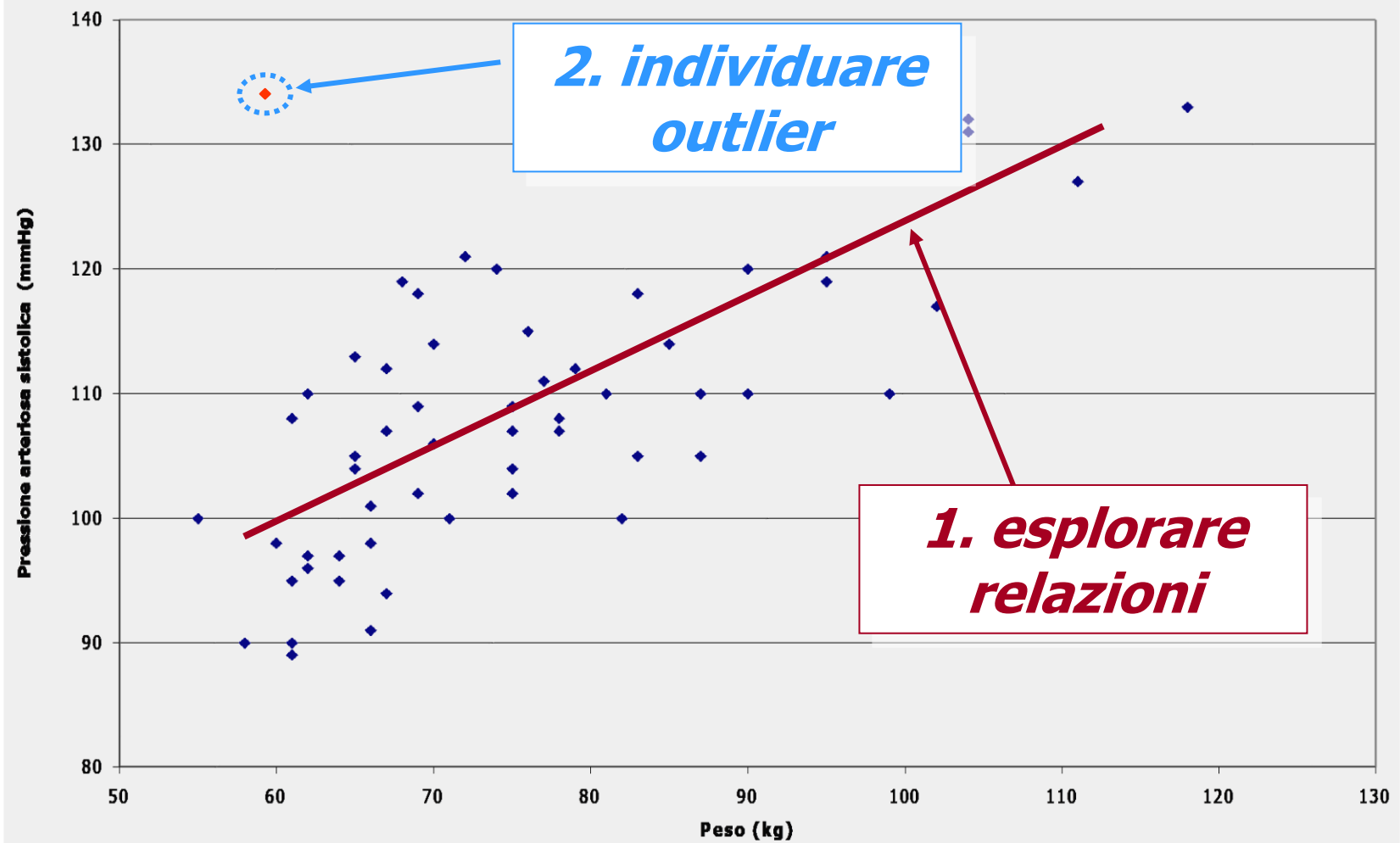


DIAGRAMMA DI DISPERSIONE (scatterplot)



commenti

Il grafico precedente ci mostra che:

- Peso e pressione sistolica arteriosa sono **positivamente associate**: i soggetti che hanno peso più elevato, hanno anche valori della pressione arteriosa maggiori
- La **relazione** tra le due variabili, ad una prima osservazione, sembra essere **lineare**