

---

# Probabilità e Statistica Esercitazioni

a.a. 2009/2010

C.d.L.: Ingegneria Elettronica e delle Telecomunicazioni, Ingegneria Informatica

## ***Statistica descrittiva III***

Ines Campa

# Campione bivariato

Nell'analisi di un fenomeno, spesso sono posti sotto osservazione più caratteri, per esempio il reddito e la spesa per i beni di lusso, l'età e gli incidenti stradali, per stabilire se esiste una relazione tra di essi.

Supponiamo di analizzare due caratteri quantitativi. Indichiamo con  $(x_i, y_i)$  la coppia di valori che costituisce  $i$ -esima osservazione rilevata sul campione in esame (**campione bivariato**). Uno strumento utile a visualizzare campioni bivariati è il diagramma di dispersione.

*Esercizio 1.* (Sheldon M. Ross "Probabilità e statistica per l'ingegneria e le scienze")

La tabella di seguito riporta la frequenza cardiaca a riposo in battiti al minuto ( $y$ ) e gli anni complessivi di istruzione ( $x$ ) di dieci individui. Costruire il diagramma di dispersione.

$i$	1	2	3	4	5	6	7	8	9	10
$x_i$	12	16	13	18	19	12	18	19	12	14
$y_i$	73	67	74	63	73	84	60	62	76	71



Dobbiamo stabilire se esiste una **correlazione** tra i valori di  $x$  e  $y$ , ovvero se a valori elevati della  $x$  corrispondono valori elevati della  $y$  e a valori bassi corrispondono valori bassi (correlazione positiva) oppure a valori alti della  $x$  corrispondono valori bassi della  $y$  o viceversa (correlazione negativa). Per ottenere una misura quantitativa di questa relazione costruiamo una nuova statistica.

Consideriamo un campione bivariato  $(x_i, y_i)$  con  $i = 1, 2 \dots n$ . Siano  $\bar{x}_n$  e  $\bar{y}_n$  le medie campionarie relative ai valori di  $x$  e  $y$  e siano  $s_x$  e  $s_y$  le deviazioni standard campionarie relative ai valori di  $x$  e  $y$ . Si definisce **coefficiente di correlazione campionaria**

$$r := \frac{\sum_{i=1}^n (x_i - \bar{x}_n)(y_i - \bar{y}_n)}{(n-1)s_x s_y}$$

Quando  $r > 0$  i dati sono correlati positivamente, quando  $r < 0$  sono correlati negativamente.

La  $\sum$  è stata normalizzata dividendo per  $(n-1)s_x s_y \implies -1 \leq r \leq 1$

Calcolando il coefficiente di correlazione dell'esercizio precedente risulta  $r = -0,7638$ , quindi vi è una forte associazione tra una scolarizzazione lunga e una bassa frequenza cardiaca e viceversa. L'associazione tra le due diverse caratteristiche non significa rapporto causa-effetto.

*Esercizio 2.* Costruire il diagramma di dispersione per il carattere 1 e il carattere 2 di un campione di 10 unità i cui dati sono riportati nella tabella sottostante.

$i$	1	2	3	4	5	6	7	8	9	10
$x_i$	1	2	3	4	5	6	7	8	9	10
$y_i$	18	19	23	24	27	29	32	34	38	40



---

Il coefficiente di correlazione è  $r = 0.9961 \approx 1$ . Quando  $|r| = 1$  vi è relazione lineare perfetta, e i punti del diagramma di dispersione stanno tutti su una retta. Valori di  $r$  intorno a 0.8 indicano già una correlazione molto intensa e anche se i punti del grafico non stanno su una retta, esiste la retta interpolante che non passa troppo lontano da nessuno di essi. Valori di  $|r|$  intorno a 0.3 denotano relazione debole.



*Esercizio 3.* Costruire il diagramma di dispersione per il carattere 1 e il carattere 2 di un campione di 10 unità i cui dati sono riportati nella tabella sottostante.

$i$	1	2	3	4	5	6	7	8	9	10
$x_i$	1	2	3	4	5	6	7	8	9	10
$y_i$	11	10	7	5	3	0	-2	-5	-6	-10

e calcolare il coefficiente di correlazione campionaria. (risulta  $r=-0,9968$ )