

Coefficiente di correlazione e regressione lineare

1. Si vuole indagare sull'utilità di una ninfa (Plecoptera nymph) come indicatore di fattori ambientali nei corsi d'acqua. In particolare si vuole stabilire se esiste una correlazione significativa tra durezza dell'acqua, valutata in unità di CaCO_3 , e numero di ninfe. La tabella seguente riporta i dati raccolti in un campione di 13 corsi d'acqua.

x= unità CaCO_3	17	20	22	28	42	55	55
y= n ninfe	42	40	30	7	12	10	8
x= unità CaCO_3	75	80	90	145	145	170	
y= n ninfe	7	3	7	5	2	4	

- (a) Disegna il diagramma di dispersione.
- (b) Determina il coefficiente di correlazione campionaria.
- (c) Esiste una correlazione positiva/negativa? Cosa puoi concludere?
2. I dati che seguono mostrano il reddito pro capite, in dollari, in 12 città americane nel 1979 e nel 1985.

Città	Reddito 1979	Reddito 1985
Newyork	7271	11188
Baltimora	5877	8647
Denver	8553	12490
Austin	7368	11633
Cincinnati	6874	10247
Omaha	7714	12886
Detroit	6215	8852
Memphis	6466	9362
Milwaukee	7029	9765
St.Louis	5877	8799
Charlotte	7952	12259
Buffalo	5929	8840

- (a) Disegna il diagramma di dispersione.
- (b) Calcola il coefficiente di correlazione campionaria tra il reddito nel 1979 e il reddito nel 1985 in queste città.
- (c) Cosa puoi concludere?
3. In un esperimento pensato per studiare la relazione tra il numero di consumazioni alcoliche e la concentrazione di alcol nel sangue, sette individui con la stessa corporatura vengono assegnati casualmente a un certo numero di bevande alcoliche. Dopo un'attesa di un'ora, viene analizzato il sangue di ciascuno, ottenendo i seguenti risultati:

Numero di consumazioni	Concentrazione di alcol
0,5	0,01
1	0,02
2	0,05
3	0,09
4	0,10
5	0,14
6	0,20

- Disegna il diagramma di dispersione.
 - Pensi che la regressione lineare sia un buon modello per questi dati? Motiva la risposta.
 - Determina la retta di regressione dei minimi quadrati.
 - Prevedi la concentrazione di alcol di una persona, con la stessa corporatura di quelle dell'esperimento, che abbia bevuto 3 consumazioni alcoliche un'ora fa.
4. I dati seguenti mettono in relazione il consumo medio giornaliero di sigarette con il numero di radicali liberi trovati nei polmoni di 8 individui.

Numero di sigarette	Radicali liberi
0	94
10	144
14	182
5	120
18	240
20	234
30	321
40	400

- Rappresenta i dati in un diagramma di dispersione.
 - Disegna "a mano" una retta di regressione.
 - Determina la retta di regressione dei minimi quadrati, e confrontala con la precedente.
 - Prevedi il numero di radicali liberi nei polmoni di qualcuno che fuma in media 26 sigarette al giorno.
5. I dati seguenti mettono in relazione la proporzione di lavoratori delle miniere di carbone che mostrano i sintomi della pneumoconiosi con gli anni di lavoro in miniera.

Anni di lavoro	Proporzione affetta da pneumococchi
5	0
10	0,0090
15	0,0185
20	0,0672
25	0,1542
30	0,1720
35	0,1840
40	0,2105
45	0,3570
50	0,4545

- (a) Disegna il diagramma di dispersione.
- (b) Pensi che la regressione lineare sia un buon modello per questi dati? Motiva la risposta.
- (c) Determina la retta di regressione dei minimi quadrati.
- (d) Usa i dati per stimare la probabilità che una persona che abbia lavorato in miniera per 42 anni contragga la pneumoconiosi.