STATISTICA

DIAGRAMMI DI DISPERSIONE. REGRESSIONE LINEARE

Prof. Rosario Lo Franco – Lezione 8

[redacted]

Diagramma di dispersione

Esempio 9.1.1. Per $i = 1, 2, \ldots, 10$, consideriamo le 10 coppie di valori (x_i, y_i), che legano y (il rendimento percentuale di un esperimento di laboratorio), a x (la temperatura a cui è stato condotto l'esperimento):

i	1	2	3	4	5	6	7	8	9	10
x_i	100	110	120	130	140	150	160	170	180	190
y_i	45	52	54	63	62	68	75	76	92	88

Quello rappresentato in Figura 9.1 è un *diagramma di dispersione* delle coppie di dati raccolti. In pratica, si tratta di tracciare un segno per ogni coppia, con le due coordinate pari ai valori di x e y rispettivamente (si veda anche quanto detto a proposito di statistica descrittiva nella Sezione 2.6). Poiché il grafico mostra, a meno di errori casuali, una relazione lineare tra y e x, sembra che la scelta di un modello di regressione lineare sia in questo caso appropriata. □

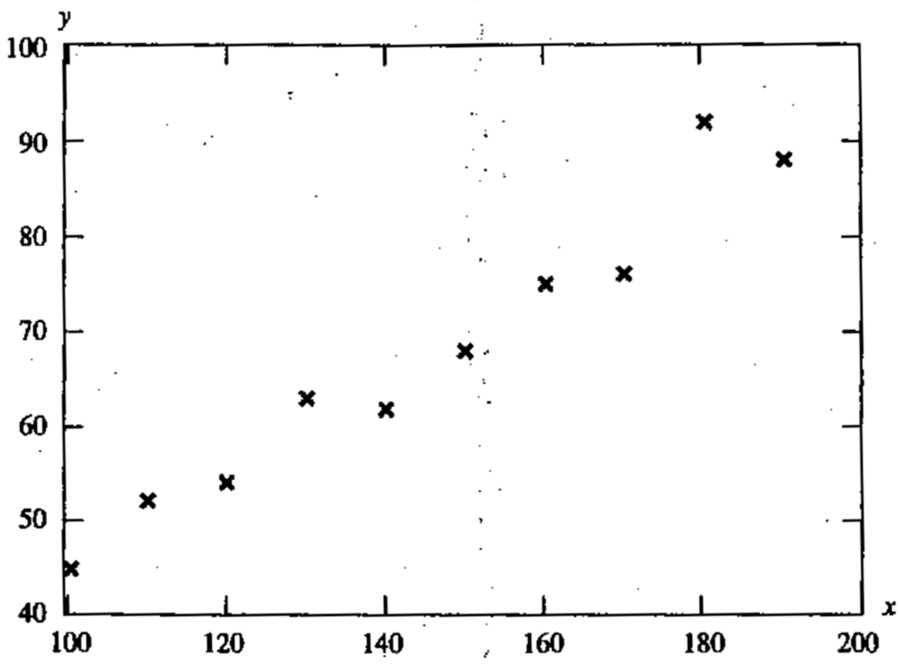

Figura 9.1 Diagramma di dispersione.

Coefficiente di correlazione

Definizione 2.6.1. Sia dato un campione bivariato (x_i, y_i), per $i = 1, 2, \ldots, n$, con medie campionarie $\bar{x}$ e $\bar{y}$ e deviazioni standard campionarie s_x e s_y, per i soli dati x e per i soli dati y rispettivamente. Allora di dice *coefficiente di correlazione campionaria* e si denota con r la quantità

$$r := \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_x s_y}$$
$$= \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{2.6.1}$$

Quando $r > 0$ i dati sono *correlati positivamente*, mentre se $r < 0$ sono *correlati negativamente*.

Proposizione 2.6.1. Di seguito diamo alcune delle proprietà del coefficiente di correlazione campionaria.

1. $-1 \leq r \leq 1$.

2. Se per oppurtune costanti a e b, con $b > 0$, sussiste la relazione lineare

$$y_i = a + bx_i, \qquad \forall i = 1, 2, \ldots, n$$

   allora $r = 1$.

3. Se per oppurtune costanti a e b, con $b < 0$ sussiste la relazione lineare

$$y_i = a + bx_i, \qquad \forall i = 1, 2, \ldots, n$$

   allora $r = -1$.

4. Se r è il coefficiente di correlazione del campione (x_i, y_1), $i = 1, \ldots, n$, allora lo è anche per il campione

$$(a + bx_i, c + dy_i) \qquad \forall i = 1, 2, \ldots, n$$

   purché le costanti b e d abbiano lo stesso segno.

Regressione Lineare

Relazione lineare semplice tra risposta Y e predittore x, con un errore non osservabile *e*:

$$Y = \alpha + \beta x + e$$

Metodo dei minimi quadrati per stimare α e β con A e B.

$$SS := \sum_{i=1}^{n} (Y_i - A - Bx_i)^2$$

Si minimizzano i residui SS per determinare A e B, derivando e uguagliando a zero rispetto ad A e B la precedente. Si ottiene:

$$B = \frac{\sum_i x_i Y_i - \bar{x} \sum_i Y_i}{\sum_i x_i^2 - n\bar{x}^2}$$

$$A = \overline{Y} - B\bar{x}$$

La retta $y = A + Bx$ è la *stima della retta di regressione*,

Regressione Lineare

Esempio 9.2.1. Il materiale grezzo usato per la produzione di una particolare fibra sintetica è immagazzinato in un ambiente che non dispone di controllo dell'umidità. Per 15 giorni vengono prese misurazioni abbinate dell'umidità atmosferica e dell'acqua assorbita dal materiale, ottenendo i risultati seguenti (in punti percentuali),

Umidità atmosferica	46	53	29	61	36	39	47	49	52	38	55	32	57	54	44
Acqua assorbita	12	15	7	17	10	11	11	12	14	9	16	8	18	14	12

Questi dati sono rappresentati nella Figura 9.2. Per calcolare gli stimatori dei minimi quadrati e la stima della retta di regressione utilizziamo il Programma 9.2, ottenendo la schermata che compare in Figura 9.3. □

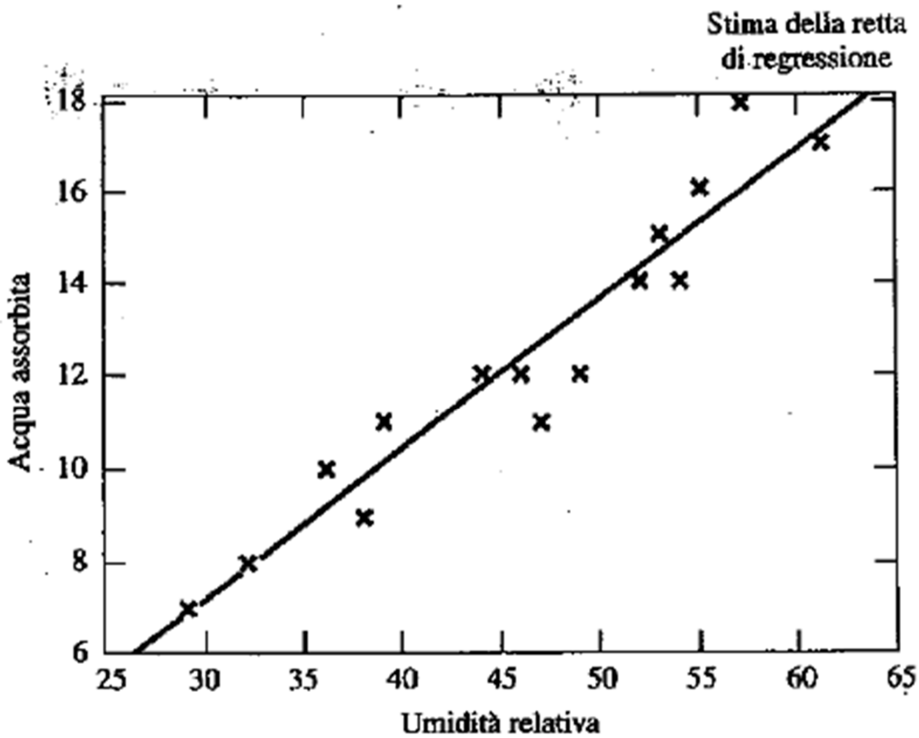

Figura 9.2 Diagramma di dispersione dei dati dell'Esempio 9.2.1.