CORSO DI LAUREA MAGISTRALE IN INGEGNERIA GESTIONALE

Gestione della Produzione e della Qualità

Correlazione

Prof. Antonella Petrillo

COSA E' UN DIAGRAMMA DI CORRELAZIONE

E' un metodo grafico di analisi, che permette di evidenziare eventuali legami esistenti tra due grandezze in esame.

ESEMPI:

- *Dimensione di un pezzo in relazione alla velocità del tornio*
- *Concentrazione di una sostanza ed il suo peso specifico*
- *Tenore d'umidità dei filati ed il loro allungamento*
- *Velocità di taglio e variazioni in lunghezza dei pezzi*

Le due variabili con cui abbiamo a che fare possono essere:

a) Una caratteristica di **qualità** ed un **fattore** che la influenza

b) Due caratteristiche di **qualità** correlate tra loro

c) Due **fattori** che si riferiscono ad una sola caratteristica di qualità

Correlazione

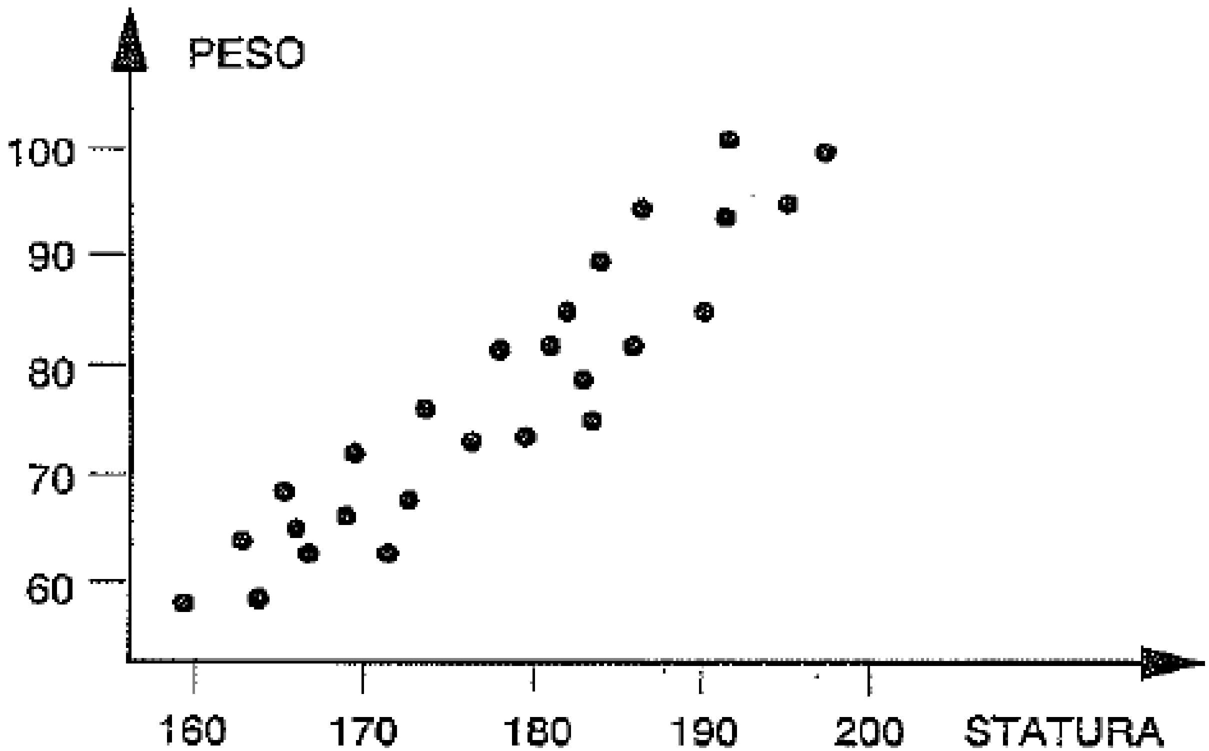
Il diagramma di correlazione consente di verificare eventuali legami esistenti tra una **causa** ed un **effetto**.
E’ uno strumento fondamentale da utilizzare dopo aver identificato con il diagramma causa-effetto tutte le cause possibili ed aver selezionato tra di esse le più probabili per individuare la vera causa.

In definitiva consente di rispondere alla seguente domanda:

AL VARIARE DELLA CAUSA PROBABILE, COME VARIA L’EFFETTO?

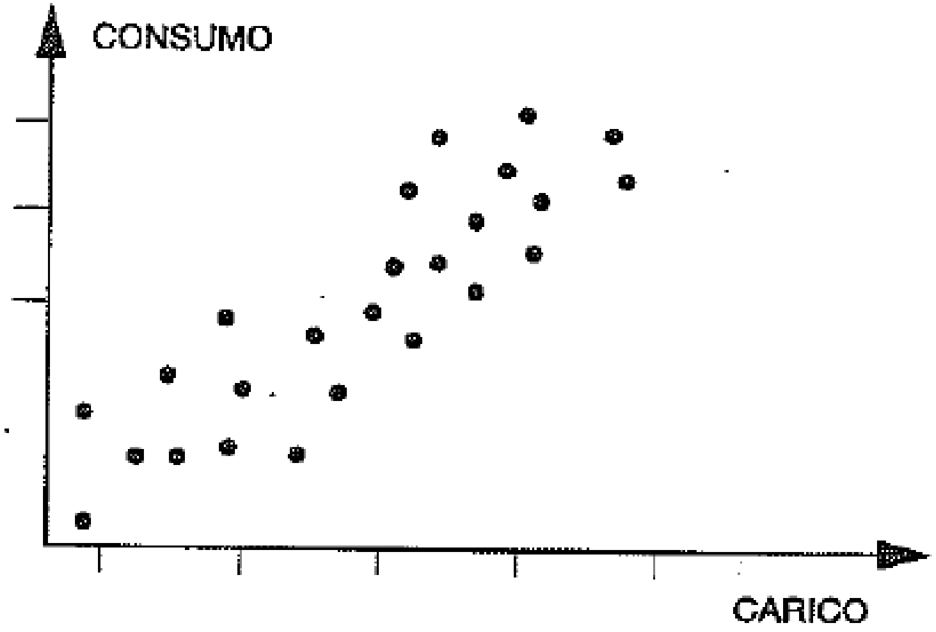
Correlazione

ESEMPI: Legame tra STATURA E PESO di 26 persone facenti parti dello stesso reparto



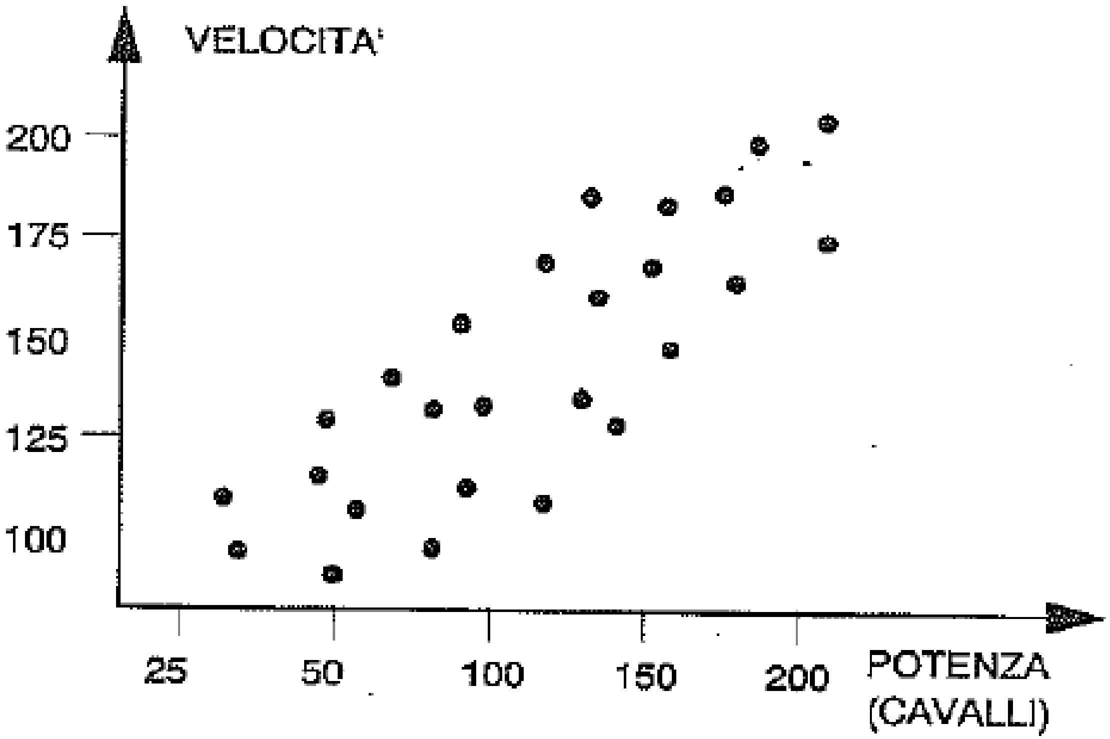
Correlazione

ESEMPI: Il consumo di CARBURANTE per km di un autocarro dipende dal carico trasportato



Correlazione

ESEMPI: Relazione tra la VELOCITÀ di un'autovettura e la sua potenza



Correlazione

COME COSTRUIRE UN DIAGRAMMA DI CORRELAZIONE

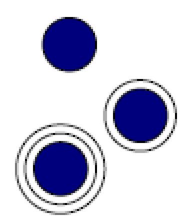
La costruzione di un diagramma di correlazione può avvenire nel seguente modo:

1. Raccogliere almeno **30 coppie** di dati per i quali si desidera indagare sulla relazione, e riportarli su di un ***foglio raccolta dati***.

2. Tracciare gli **assi orizzontale** e **verticale** del diagramma, riportando la grandezza **A** sull'asse orizzontale e la grandezza **B** sull'asse verticale utilizzando opportune scale.

3. Riportare i dati sul ***diagramma***. Si suggerisce il seguente metodo grafico:

Dato unico
Due dati coincidenti
Tre dati coincidenti



4. Completare il diagramma con le necessarie note *(scopo dell'analisi, periodo di osservazione, reparto, ecc.)*

Correlazione

ESEMPIO 1

Si considerino i valori assunti dalla **velocità del trasportatore** *(causa)* e la **lunghezza del pezzo tagliato** *(effetto)*.

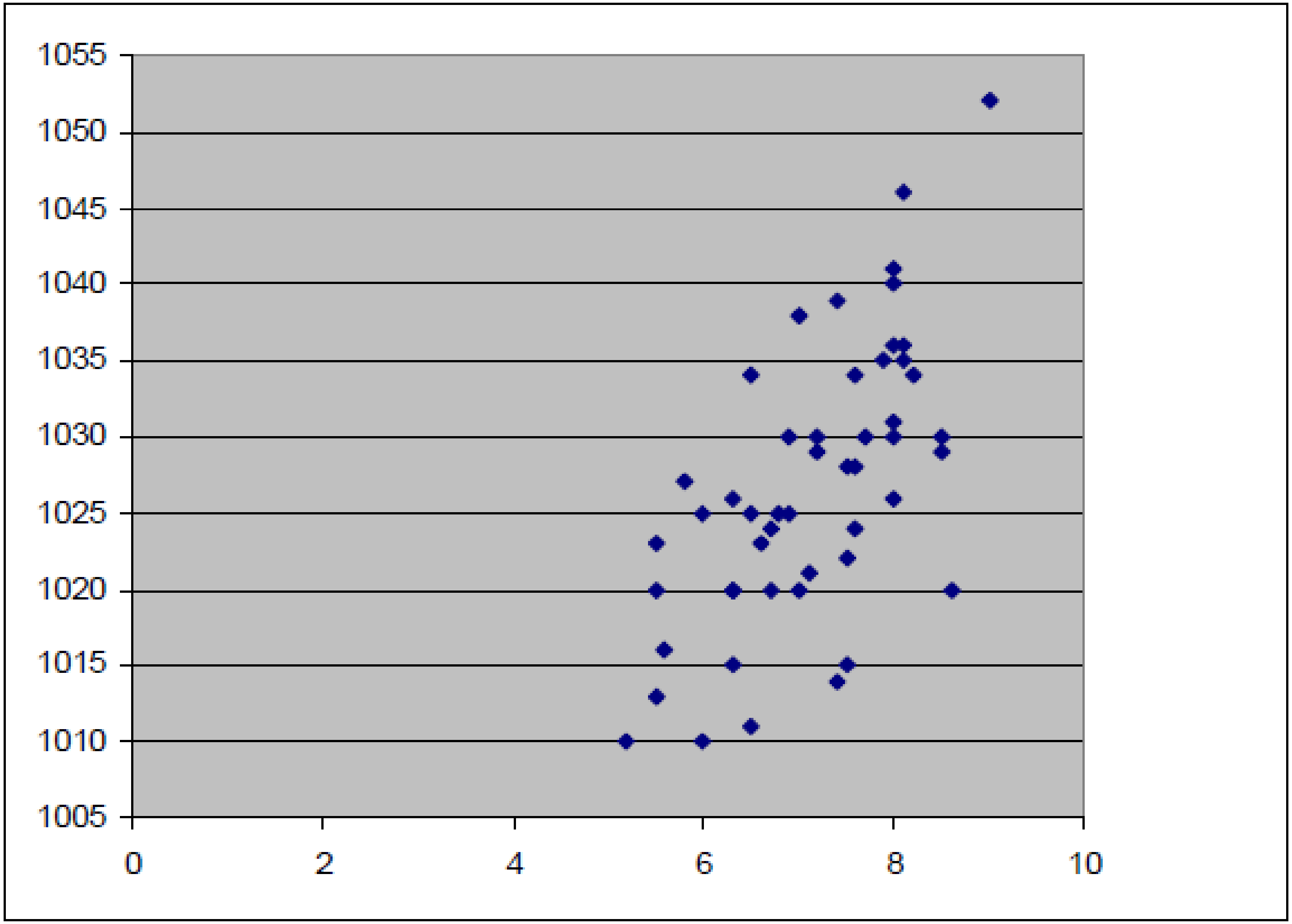
Velocità del trasportatore cm/sec

Lunghezza tagliata mm

Si riportino tali valori su di un diagramma.

1	8,1	1046	**26**	8	1040
2	7,7	1030	**27**	5,5	1013
3	7,4	1039	**28**	6,9	1025
4	5,8	1027	**29**	7	1020
5	7,6	1028	**30**	7,5	1022
6	6,8	1025	**31**	6,7	1020
7	7,9	1035	**32**	8,1	1035
8	6,3	1015	**33**	9	1052
9	7	1038	**34**	7,1	1021
10	8	1036	**35**	7,6	1024
11	8	1026	**36**	8,5	1029
12	8	1041	**37**	7,5	1015
13	7,2	1029	**38**	8	1030
14	6	1010	**39**	5,2	1010
15	6,3	1020	**40**	6,5	1025
16	6,7	1024	**41**	8	1031
17	8,2	1034	**42**	6,9	1030
18	8,1	1036	**43**	7,6	1034
19	6,6	1023	**44**	6,5	1034
20	6,5	1011	**45**	5,5	1020
21	8,5	1030	**46**	6	1025
22	7,4	1014	**47**	5,5	1023
23	7,2	1030	**48**	7,5	1028
24	5,6	1016	**49**	8,6	1020
25	6,3	1020	**50**	6,3	1026

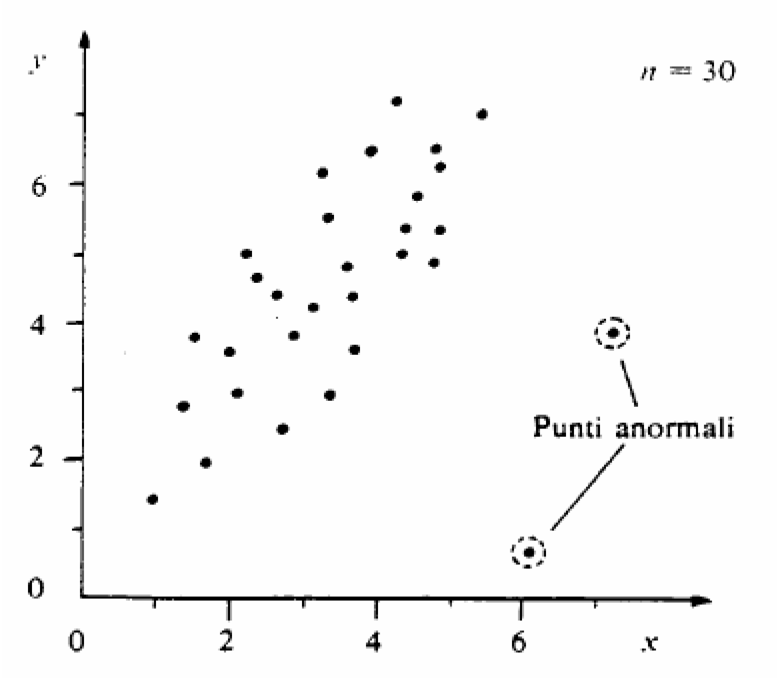
Correlazione



Correlazione

COME INTERPRETARE I DIAGRAMMI DI CORRELAZIONE

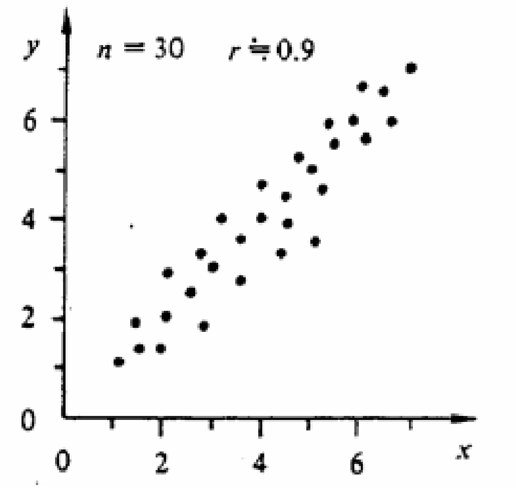
1. Verificare la presenza di **punti anomali** (*isolati*) sul diagramma. In tal caso bisogna indagare circa le loro cause ed eventualmente eliminarli prima di ulteriori elaborazioni.



Generalmente si tratta di:
- *errori di misurazione*
- *errori di trascrizione*
- *cambio di procedura di analisi o altre anomalie*

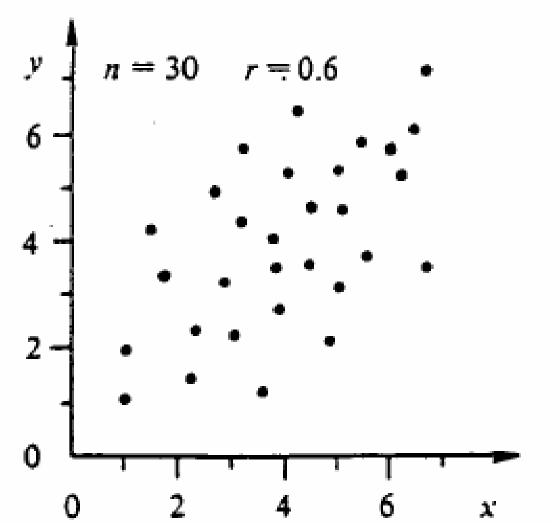
Correlazione

ESEMPI POSSIBILI DI CORRELAZIONI



A. Correlazione POSITIVA:

Quando X aumenta, anche la Y aumenta. Tenendo sotto controllo i valori della X risulteranno sotto controllo anche i valori della Y.

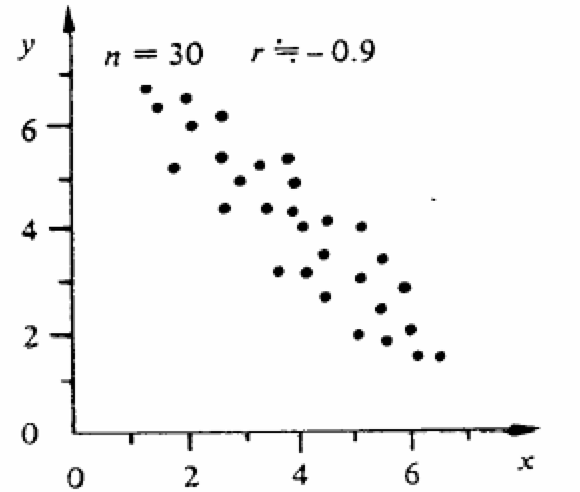


B. Correlazione Probabilmente POSITIVA:

Quando X aumenta, Y tende ad aumentare. Ciò si verifica quando altri fattori diversi da X influenzano Y.

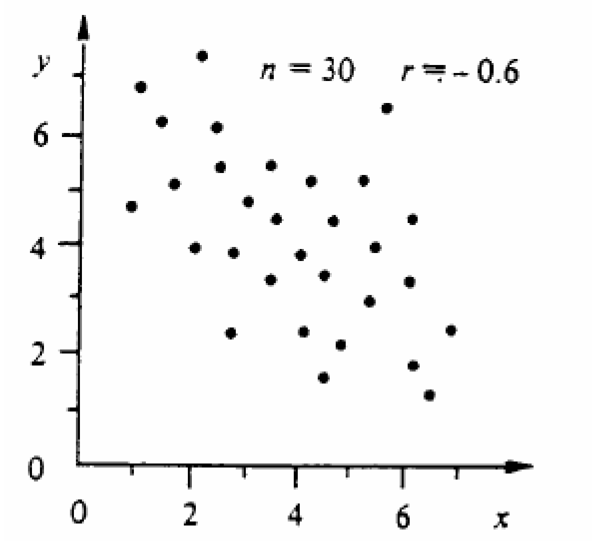
Correlazione

ESEMPI POSSIBILI DI CORRELAZIONI



C. Correlazione NEGATIVA:

Quando X aumenta, Y diminuisce. Tenendo sotto controllo i valori della X risulteranno sotto controllo anche i valori della Y.

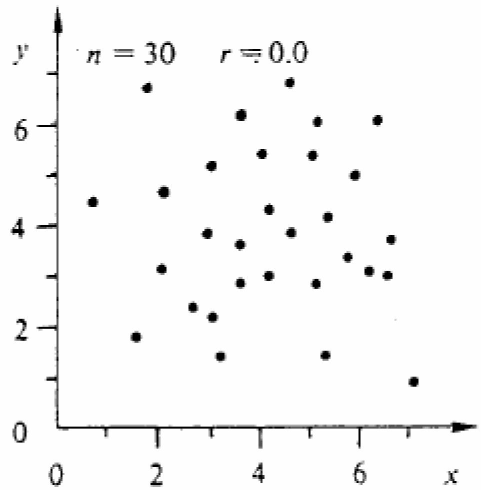


D. Correlazione Probabilmente NEGATIVA:

Quando X aumenta, Y tende a diminuire. Ciò si verifica quando altri fattori diversi da X influenzano Y.

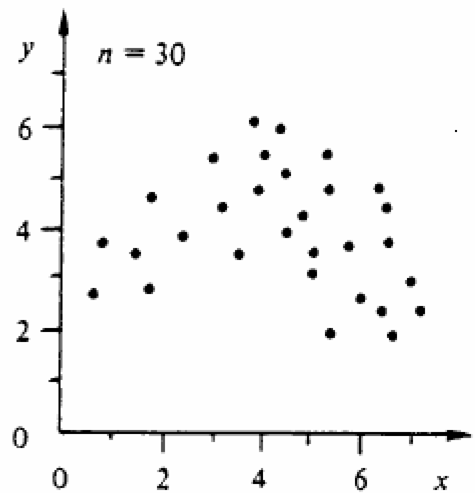
Correlazione

ESEMPI POSSIBILI DI CORRELAZIONI



E. NESSUNA Correlazione:

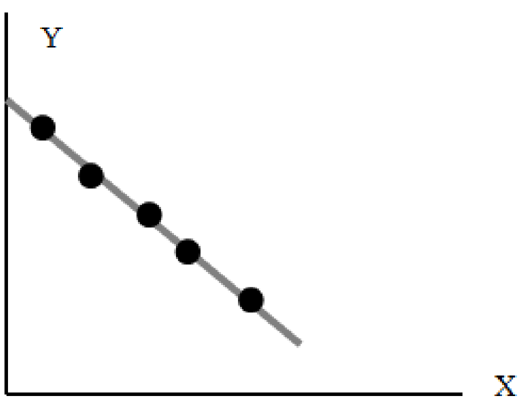
Per ogni valore di X non esiste un valore preferenziale di Y.



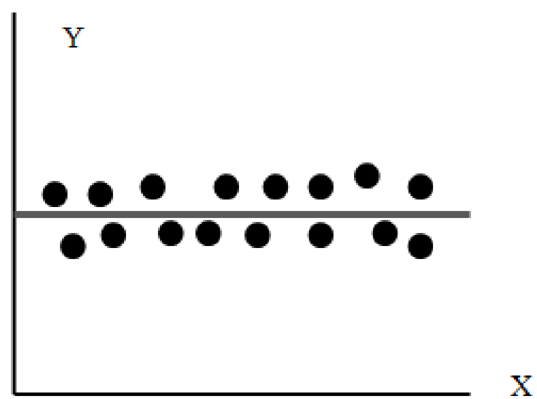
Correlazione

ESEMPI POSSIBILI DI CORRELAZIONI

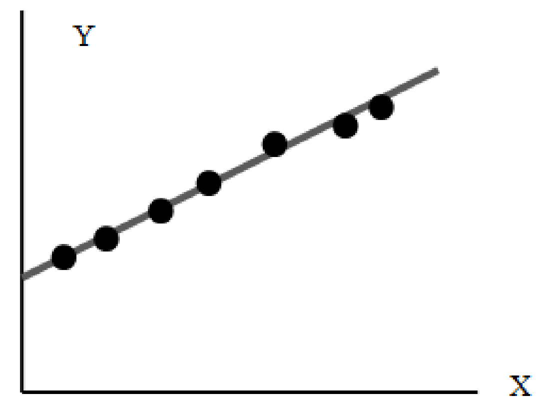
“Perfetta correlazione” significa che, se si disegnano i punti su un diagramma di dispersione, tutti i punti sono allineati, ovvero sono disposti su una retta.



Riquadro A
Perfetta correlazione negativa
(ρ = - 1)



Riquadro B
Assenza di correlazione
(ρ = 0)



Riquadro C
Perfetta correlazione positiva
(ρ = + 1)

Correlazione

METODI PER LA DETERMINAZIONE DELL'ESISTENZA DI CORRELAZIONE IN UN DIAGRAMMA DI CORRELAZIONE

Il diagramma di correlazione serve per dare risposta a due domande fondamentali:

A. *Esiste o no correlazione tra due grandezze prese in esame ?*

B. *Se esiste correlazione, come può essere espresso il legame tra le due grandezze ?*

Per determinare l'esistenza di una correlazione e l'esistenza della stessa, è possibile utilizzare uno dei seguenti metodi:

1. Calcolo del ***coefficiente di correlazione***
2. Grafico di ***probabilità binomiale***
3. ***Sign test*** (test del segno)

METODI PER LA DETERMINAZIONE DELL'ESISTENZA DI CORRELAZIONE IN UN DIAGRAMMA DI CORRELAZIONE

… in definitiva …

Non sempre l'esame visivo di un diagramma consente di definire il tipo di correlazione.

E' pertanto necessario utilizzare **supporti statistici** per valutare l'esistenza di correlazione.

Uno strumento semplice, ma rigoroso è il cosi detto **sign-test** che consente di stabilire se esiste o no una correlazione tra due grandezze confrontando la numerosità dei punti presenti nei quadranti individuati dalle mediane con i dati riportati su di una tabella opportunamente precostituita.

1. COEFFICIENTE DI CORRELAZIONE

Si definisce *coefficiente di correlazione* il rapporto:

$$r = \frac{S(xy)}{\sqrt{S(xx) * S(yy)}}$$

ove:

x rappresenta la *causa*

y = D(x) rappresenta l'*effetto*

Il coefficiente di correlazione r varia nell'intervallo :

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

I termini S(xx) e S(yy) sono le *varianze* delle due variabili; il termine S(xy) è la ***covarianza*** tra le due variabili così definite.

Correlazione

varianza

$$S(xx) = \sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}$$

$$S(yy) = \sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}$$

covarianza

$$S(xy) = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{n} x_i y_i - \frac{\left(\sum_{i=1}^{n} x_i\right) * \left(\sum_{i=1}^{n} y_i\right)}{n}$$

Nel caso di forte correlazione positiva si ottiene un valore prossimo a 1; nel caso di correlazione negativa si ottiene un valore di r prossimo a -1.
Pertanto quando il valore assoluto di r è prossimo a ± 1 si ha forte correlazione tra x e y; invece quando r è prossimo a 0 si ha debole correlazione.

Correlazione

ESEMPIO 2.

In tabella sono riportati i licenziamenti attuati presso la ***FONDERSUD*** negli ultimi mesi, nonché le domande di richiesta di lavoro inoltrate all'ufficio *CERCA LAVORO*. Verificare l'esistenza di correlazione determinando il valore del coefficiente r.

Licenziamenti (X)	0	0	25	20	50	60	60	25	20	35	45	0	
	0	0	0	0	20	0	30	30	40	0	0	10	470
Domande (Y)	50	60	80	65	110	145	115	125	120	120	110	70	
	60	65	90	55	70	45	100	105	105	70	95	100	2130

Correlazione

SOLUZIONE

Il coefficiente di correlazione si determina applicando la relazione:

$$r = \frac{S(xy)}{\sqrt{S(xx) * S(yy)}}$$

$$S(xy) = \sum_{i=1}^{n} x_i y_i - \frac{\left(\sum_{i=1}^{n} x_i\right) * \left(\sum_{i=1}^{n} y_i\right)}{n} = 51825 - \frac{(2130)\cdot(470)}{24} = 10112{,}5$$

$$S(xx) = \sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n} = 18900 - \frac{(470)^2}{24} = 9695{,}8$$

$$S(yy) = \sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n} = 206350 - \frac{(2130)^2}{24} = 17312{,}5$$

$$r = \frac{S(xy)}{\sqrt{S(xx) * S(yy)}} = \frac{10112{,}5}{\sqrt{9695{,}8 \cdot 17312{,}5}} = 0{,}78$$

Y	X	Y^2	X^2	X•Y
50	0	2500	0	0
60	0	3600	0	0
80	25	6400	625	2000
65	20	4225	400	1300
110	50	12100	2500	5500
145	60	21025	3600	8700
115	60	13225	3600	6900
125	25	15625	625	3125
120	20	14400	400	2400
120	35	14400	1225	4200
110	45	12100	2025	4950
70	0	4900	0	0
60	0	3600	0	0
65	0	4225	0	0
90	0	8100	0	0
55	0	3025	0	0
70	20	4900	400	1400
45	0	2025	0	0
100	30	10000	900	3000
105	30	11025	900	3150
105	40	11025	1600	4200
70	0	4900	0	0
95	0	9025	0	0
100	10	10000	100	1000
2130	470	206350	18900	51825

2. TEST DEL SEGNO *(Sign Test)*

1. Trovare il ***valore mediano*** delle x (x) e delle y (y). Tracciare le ***mediane*** parallele agli assi cartesiani.

2. Identificare i ***4 quadranti*** (**I, II, III, IV**) partendo dalla zona alta a destra e procedendo in senso antiorario.

3. ***Contare i punti*** che si trovano in ogni quadrante.

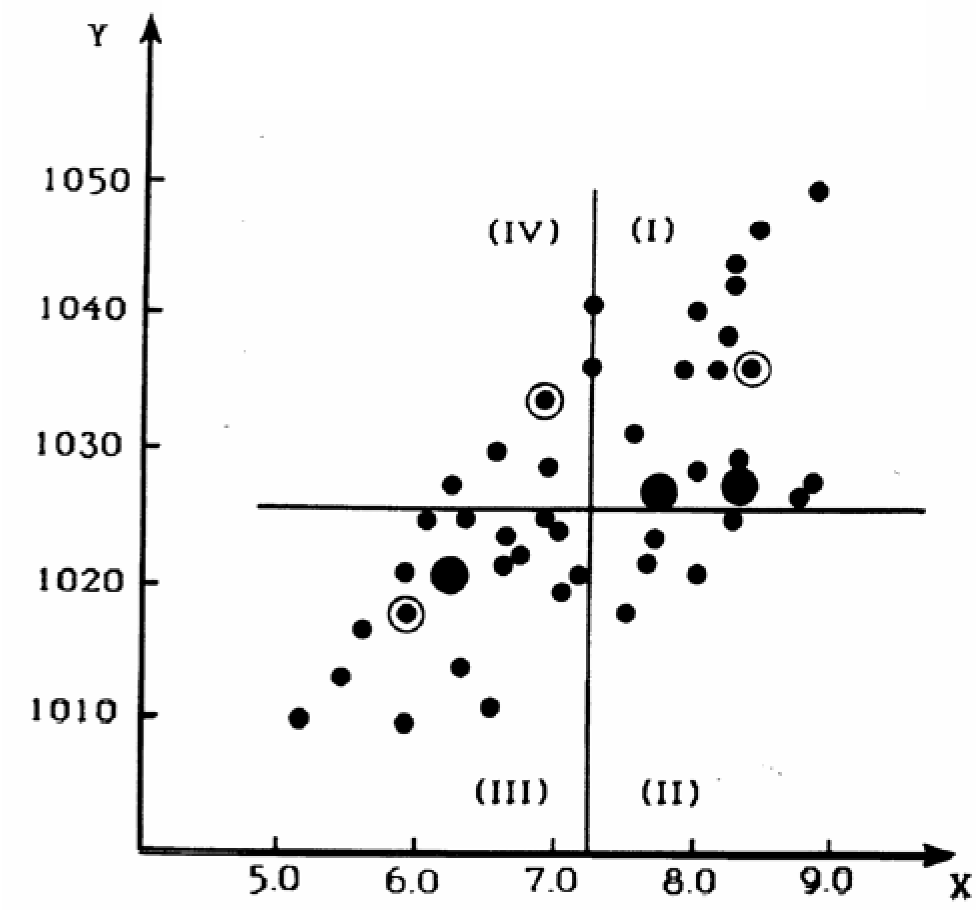
4. ***Sommare*** i punti contenuti nei ***quadranti opposti*** (**I + III e II + IV**). Prendere in considerazione la coppia di quadranti con minore numerosità (sia **M** questo valore)

5. Calcolare **N** = *Totale dei punti - punti sulle mediane*

6. Confrontare il ***numero di punti*** **M** con il ***numero limite*** (**NL**) tabulato in corrispondenza di N

7. Se **M < NL** *esiste correlazione*

Correlazione



ZONA	Punti
(I)	19
(II)	4
(III)	20
(IV)	5
Sulle linee	2
Totale	50

Numerosità campione = 50

M = 4+5 = 9
N = 9-2 = 7

TABELLA DEL TEST DEL SEGNO

Si consideri un campione composto da una serie di **misure**, ciascuna delle quali ha segno **più** e segno **meno**.

La Tabella del *test del segno* è la tabella con cui verificare se questo campione può essere considerato come estratto da una popolazione avente:

N.ro di segni + = N.ro segni -

Composizione della TABELLA

La ***colonna di sinistra*** riporta la grandezza **n** del campione.

Pr indica il livello di rischio (1%; 5%)

La Tabella fornisce i valori del ***limite inferiore*** e del ***limite superiore*** per un campione di n elementi estratto da una popolazione avente frazione difettosa del 50%.

Correlazione

TABELLA per il test del segno

Data la numerosità del campione n ....ponendoci sulle probabilità sia 1% sia 5% e individuando i rispettivi limiti superiore ed inferiore la somma dei due valori deve ridare n......ovvero la numerosità del campione

Esempio:

Numerosità campione = 50

Per 1% si ha 15+35 = 50

Per 5% si ha 17+33 = 50

Pr / n	Limite Infer.		Limite Super.		Pr / n	Limite Infer.		Limite Super.		Pr / n	Limite Infer.		Limite Super.	
	1%	[redacted]	5%	1%		1%	5%	5%	1%		1%	5%	5%	1%
1					31	7	9	22	24	61	20	22	39	41
2					32	8	9	23	24	62	20	22	40	42
3				3	33	8	10	23	25	63	20	23	40	43
4				4	34	9	10	24	25	64	21	23	41	43
5			5	5	35	9	11	24	26	65	21	24	41	44
6		0	6	6	36	9	11	25	27	66	22	24	42	44
7		0	7	7	37	10	12	25	27	67	22	25	42	45
8	0	0	8	8	38	10	12	26	28	68	22	25	43	46
9	0	1	8	9	39	11	12	27	28	69	23	25	44	46
10	0	1	9	10	40	11	13	27	29	70	23	26	44	47
11	0	1	10	11	41	11	13	28	30	71	24	26	45	47
12	1	2	10	11	42	12	14	28	30	72	24	27	45	48
13	1	2	11	12	43	12	14	29	31	73	25	27	46	48
14	1	2	12	13	44	13	15	29	31	74	25	28	46	49
15	2	3	12	13	45	13	15	30	32	75	25	28	47	50
16	2	3	13	14	46	13	15	31	33	76	26	28	48	50
17	2	4	13	15	47	14	16	31	33	77	26	29	48	51
18	3	4	14	15	48	14	16	32	34	78	27	29	49	51
19	3	4	15	16	49	15	17	32	34	79	27	30	49	52
20	3	5	15	17	50	15	17	33	35	80	28	30	50	52
21	4	5	16	17	51	15	18	33	36	81	28	31	50	53
22	4	5	17	18	52	16	18	34	36	82	28	31	51	54
23	4	6	17	19	53	16	18	35	37	83	29	32	51	54
24	5	6	18	19	54	17	19	35	37	84	29	32	52	55
25	5	7	18	20	55	17	19	36	38	85	30	32	53	55
26	6	7	19	20	56	17	20	36	39	86	30	33	53	56
27	6	7	20	21	57	18	20	37	39	87	31	33	54	56
28	6	8	20	22	58	18	21	37	40	88	31	34	54	57
29	7	8	21	22	59	19	21	38	40	89	31	34	55	58
30	7	9	21	23	60	19	21	39	41	90	32	35	55	58

CHE COSA E' UNA ANALISI DI REGRESSIONE

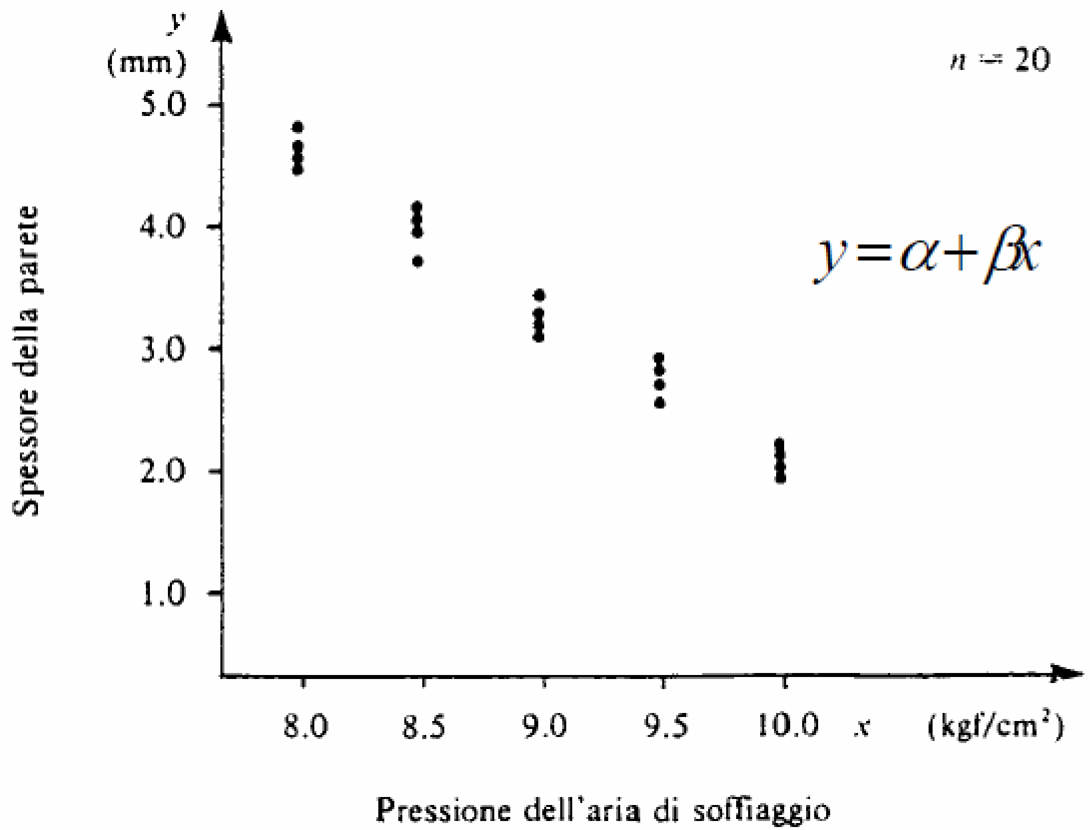
Dopo avere identificato l'esistenza di una correlazione, è necessaria una sua **quantificazione**, ossia sapere di ***quanto varia Y per una determinata variazione di X***.

Si considerino i seguenti dati, relativi allo spessore di un contenitore al variare della pressione dell'aria:

Pressione dell'aria (kgf/cm^2)	**8.0**	**8.5**	**9.0**	**9.5**	**10.0**
Spessore della parete (mm)	4.62	4.12	3.21	2.86	1.83
	4.50	3.88	3.05	2.53	2.02
	4.43	4.01	3.16	2.71	2.24
	4.81	3.67	3.30	2.62	1.95

Si consideri il diagramma di correlazione relativo a tali dati.

Correlazione



STIMA DELLA RETTA DI REGRESSIONE

E' possibile definire la *retta di regressione* con l'espressione:

$$\mathbf{y} = \alpha + \beta \mathbf{x}$$

y = variabile dipendente
x = variabile indipendente
α = costante
β = coefficiente di regressione

Metodo dei minimi quadrati

Considero un insieme di *n* coppie di dati osservati (x_i, y_i) con $1 < i < n$.
Si consideri il termine ε_i (*errore*):

$$\varepsilon_i = y_i - (\alpha^* + \beta^* x_i)$$

Con il metodo dei minimi quadrati, α^* e β^* rappresentano i valori che minimizzano il termine:

$$\sum_{i=1}^{n} \varepsilon_i^2$$

Correlazione

PROCEDURA

1. Calcolare i valori medi $\bar{x}; \bar{y}$
2. Calcolare S(xx) ed S(xy)
3. Calcolare β mediante la formula

$$\hat{\beta} = \frac{S(xy)}{S(xx)}$$

4. Calcolare α mediante la formula

$$\hat{\alpha} = \bar{y} - \hat{\beta} \cdot \bar{x}$$

ESEMPIO 3.

Dopo avere variato la pressione dell’aria, si è proceduto alla misurazione dello spessore dei contenitori ai diversi valori di pressione. I valori ottenuti sono i seguenti.

Pressione dell’aria (kgf/cm^2)	**8.0**	**8.5**	**9.0**	**9.5**	**10.0**
Spessore della parete (mm)	4.62	4.12	3.21	2.86	1.83
	4.50	3.88	3.05	2.53	2.02
	4.43	4.01	3.16	2.71	2.24
	4.81	3.67	3.30	2.62	1.95

Nell’ipotesi in cui esista una correlazione tra pressione dell’aria e spessore della parete del recipiente, si calcoli il valore della retta di regressione.

Correlazione

SOLUZIONE

Fase 1:

$$\bar{x} = (8{,}0 + 8{,}5 + 9{,}0 + 9{,}5 + 10{,}0) \cdot 4 / 20 = 9{,}00$$

$$\bar{y} = (4{,}62 + 4{,}50 + \ldots + 1{,}95) / 20 = 3{,}276$$

Fase 2:

$$S(xx) = \sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n} = 1630 - \frac{(180)^2}{20} = 10{,}0$$

$$S(xy) = \sum_{i=1}^{n} x_i y_i - \frac{\left(\sum_{i=1}^{n} x_i\right) * \left(\sum_{i=1}^{n} y_i\right)}{n} = 576{,}88 - \frac{(180) \cdot (65{,}52)}{20} = -12{,}8$$

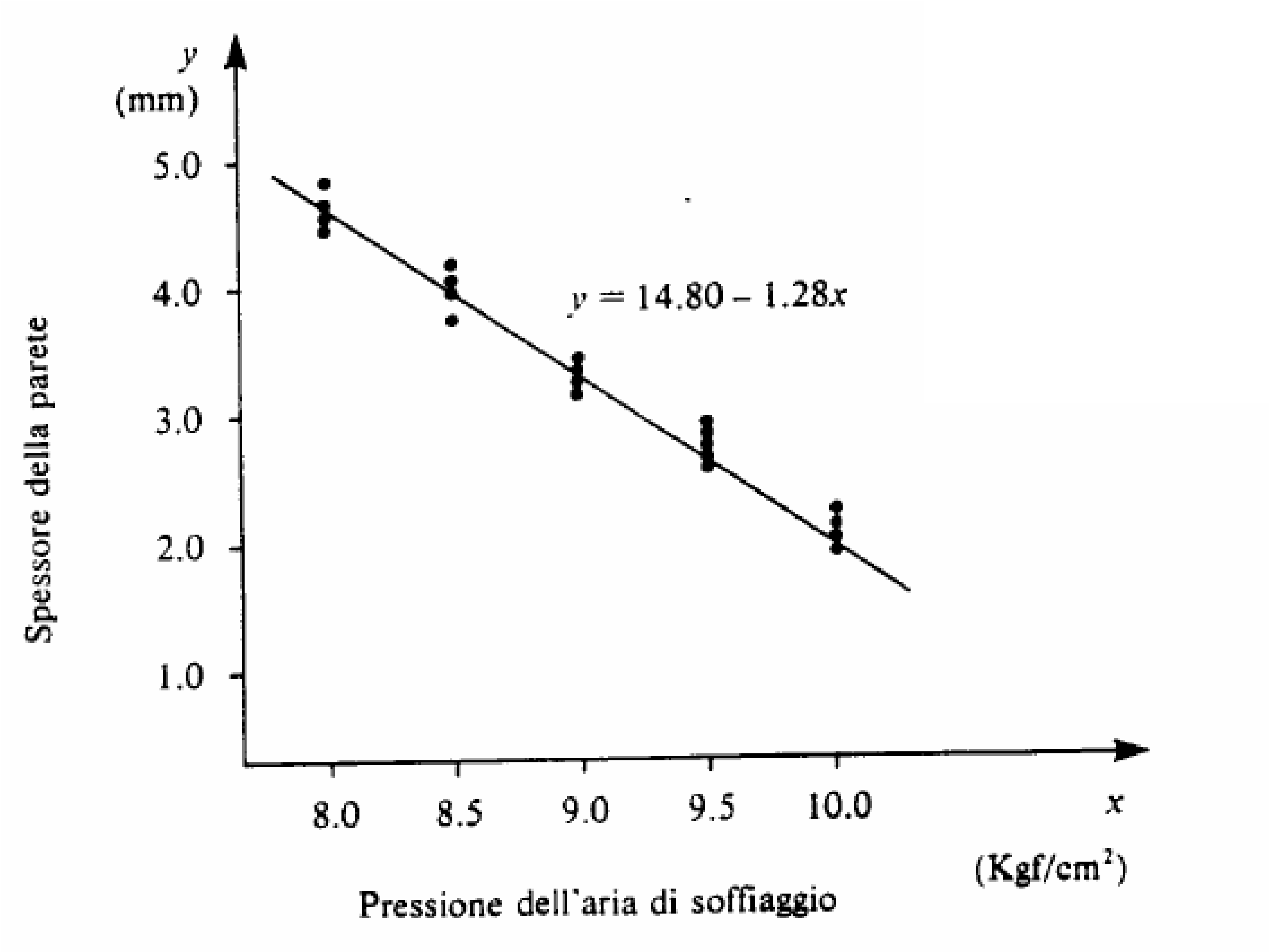
Fase 3:

$$\hat{\beta} = -12{,}8 / 10{,}0 = -1{,}28$$

$$\hat{\alpha} = 3{,}276 - (-1{,}28) \cdot 9{,}00 = 14{,}80$$

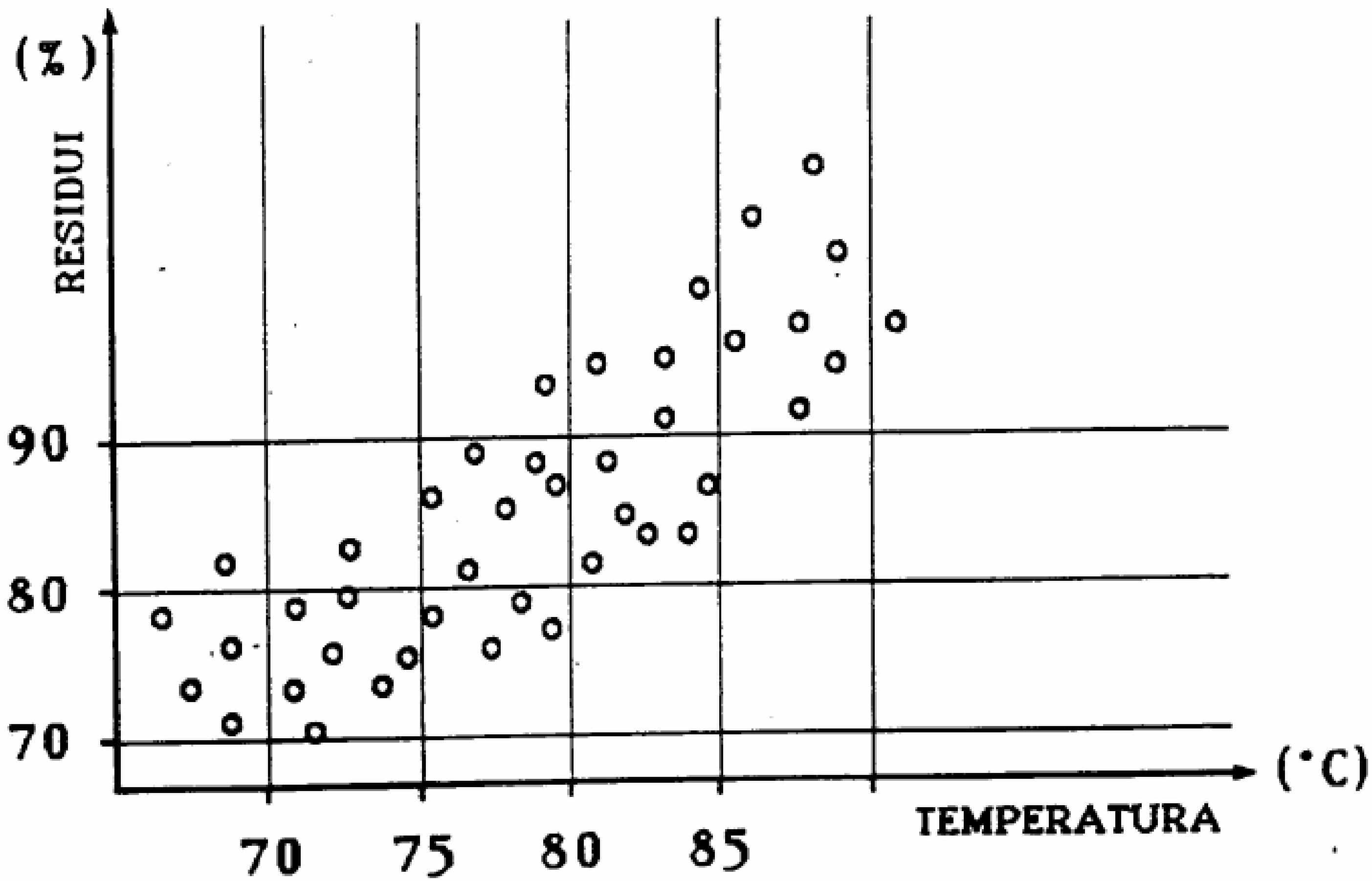
$$y = 14{,}80 - 1{,}28 \cdot x$$

Correlazione



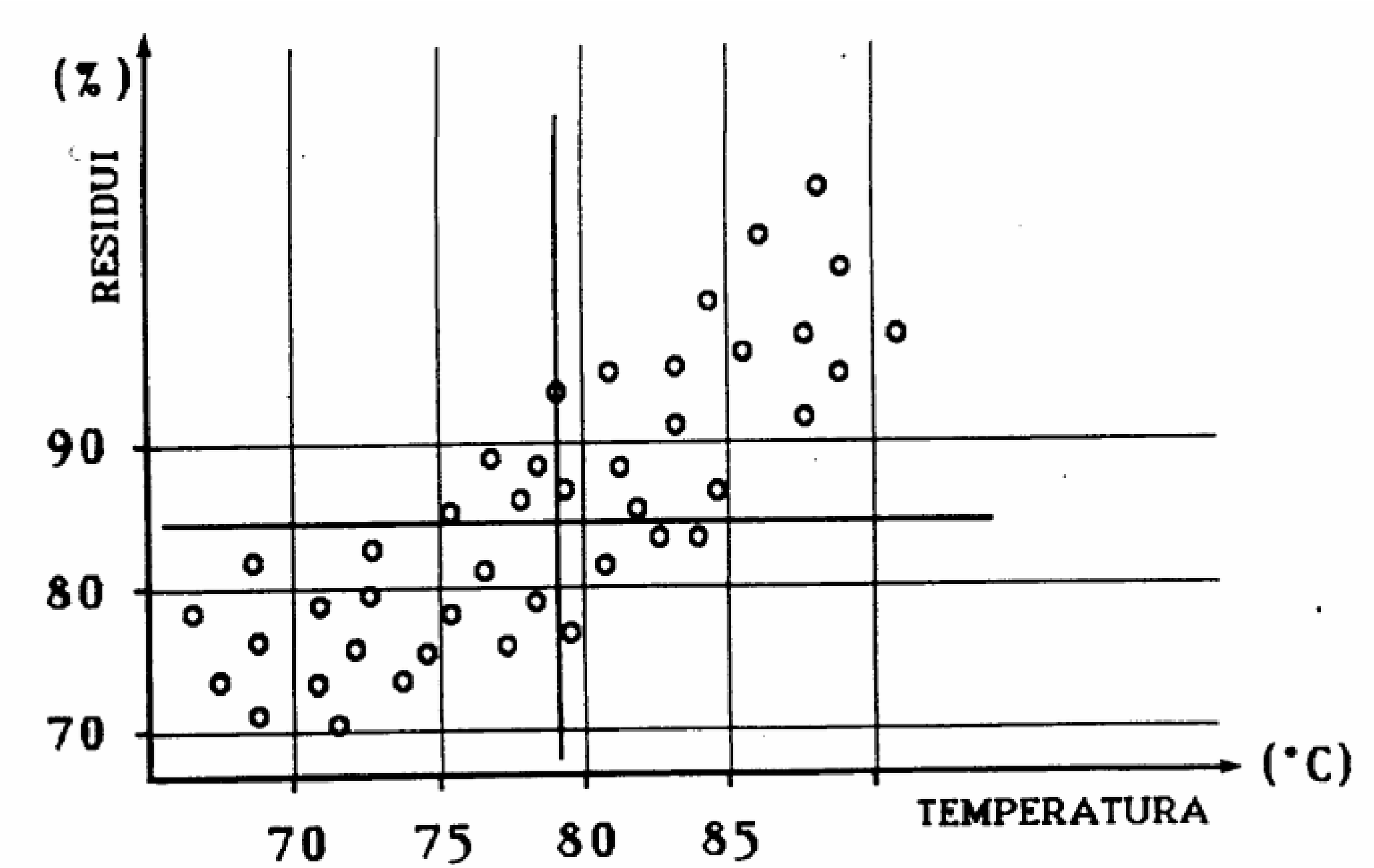
Correlazione

DETERMINAZIONE GRAFICA DELLA RETTA DI REGRESSIONE



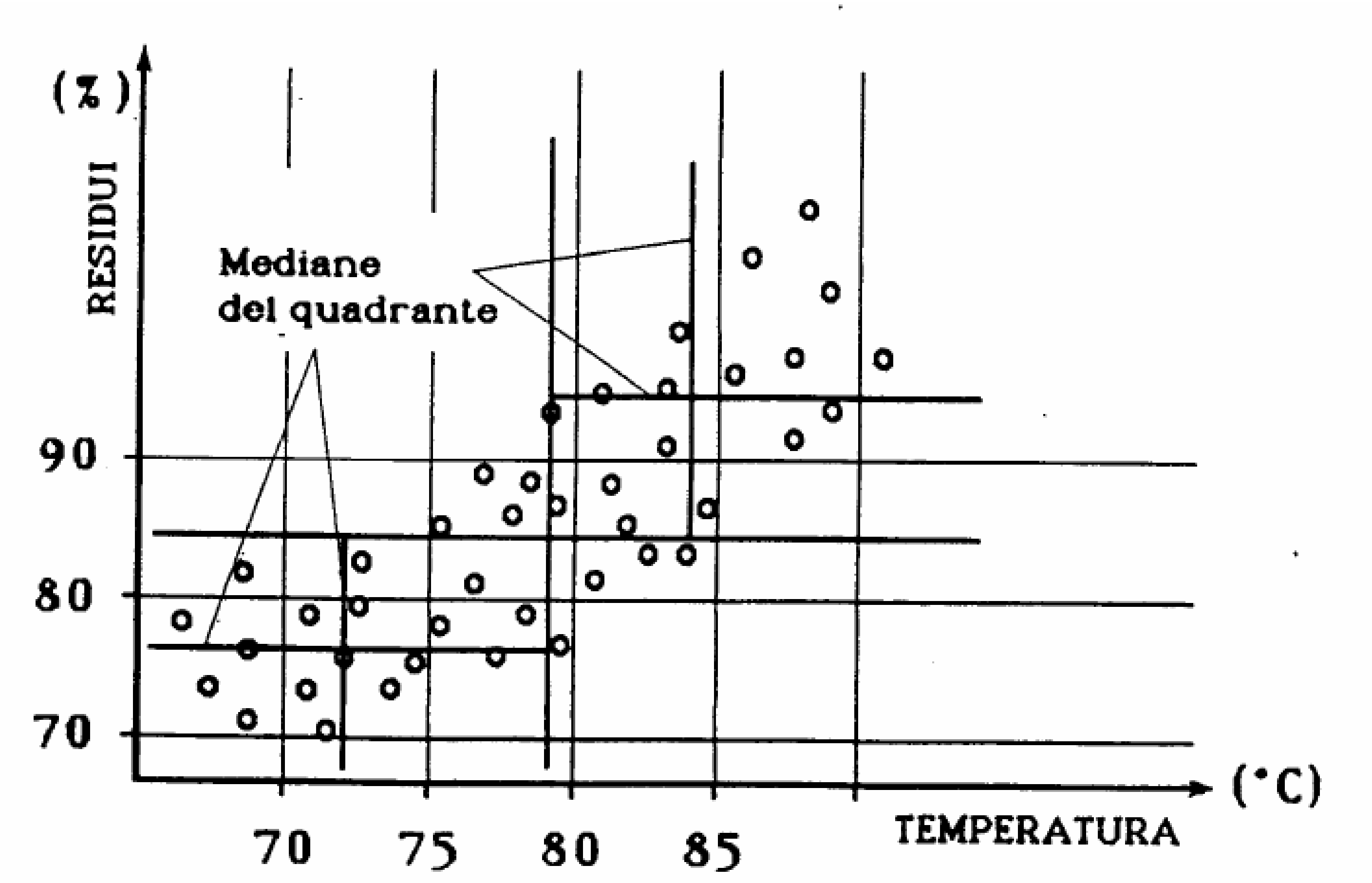
Correlazione

DETERMINAZIONE GRAFICA DELLA RETTA DI REGRESSIONE



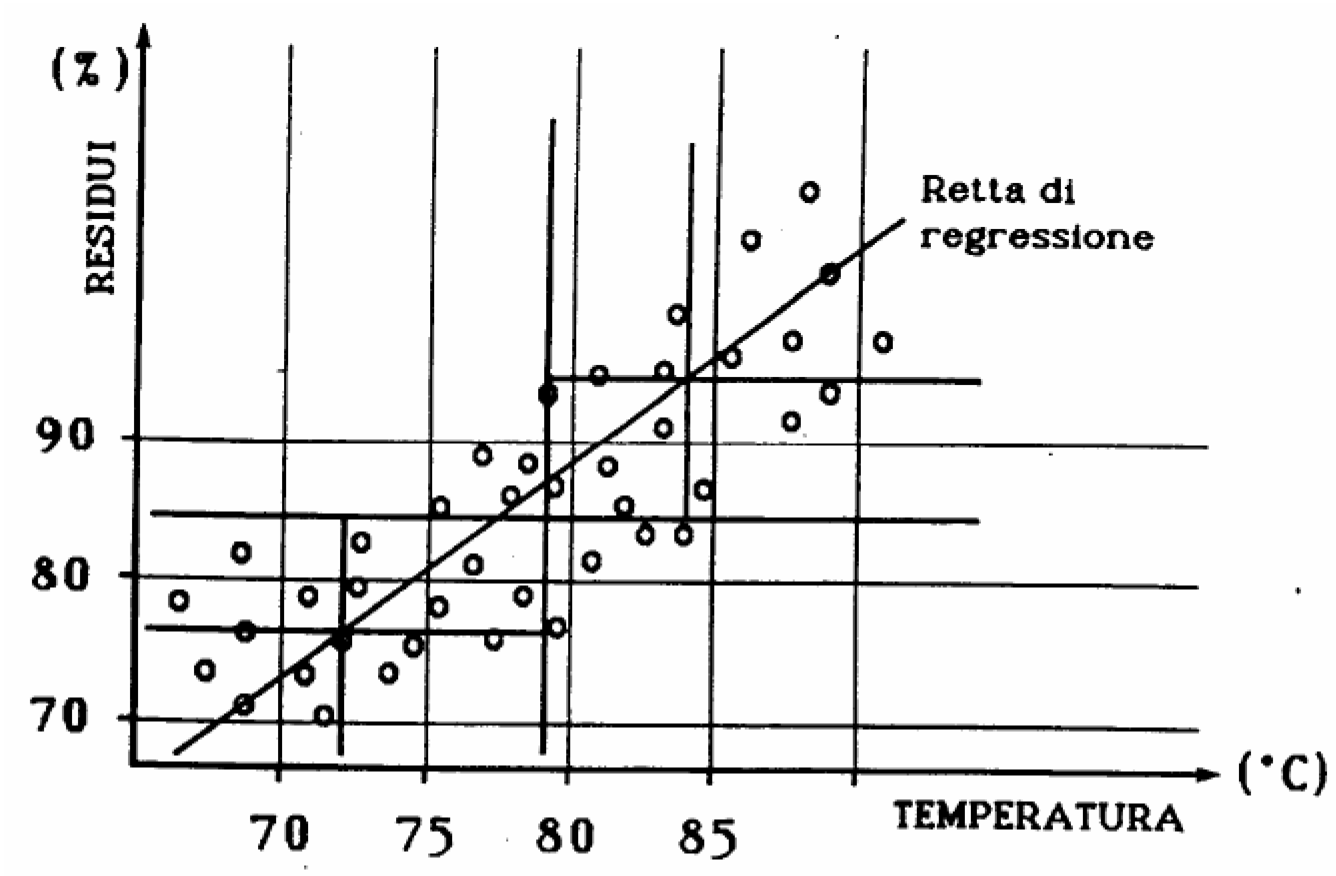
Correlazione

DETERMINAZIONE GRAFICA DELLA RETTA DI REGRESSIONE



Correlazione

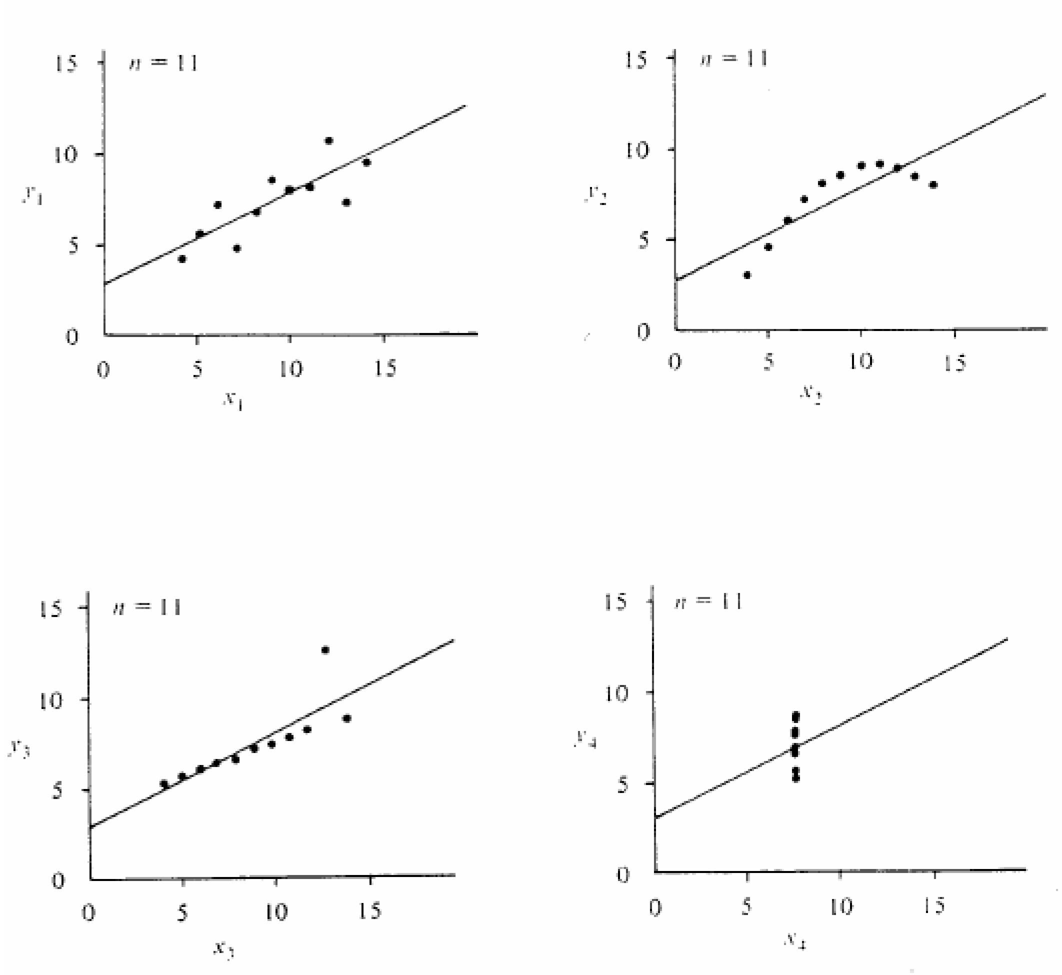
DETERMINAZIONE GRAFICA DELLA RETTA DI REGRESSIONE



Correlazione

NOTE SULL'ANALISI DI REGRESSIONE

ESEMPIO



Correlazione

Non effettuare mai una analisi di regressione se prima non si è disegnato il diagramma di correlazione.

N.	x_1	y_1	x_2	y_2	x_3	y_3	x_4	y_4
1	10	8.04	10	9.14	10	7.46	8	6.58
2	8	6.95	8	8.14	8	6.77	8	5.76
3	13	7.58	13	8.74	13	12.74	8	7.71
4	9	8.81	9	8.77	9	7.11	8	8.84
5	11	8.33	11	9.26	11	7.81	8	8.47
6	14	9.96	14	8.10	14	8.84	8	7.04
7	6	7.24	6	6.13	6	6.08	8	5.25
8	4	4.26	4	3.10	4	5.39	19	12.50
9	12	10.84	12	9.13	12	8.15	8	5.56
10	7	4.82	7	7.26	7	6.42	8	7.91
11	5	5.68	5	4.74	5	5.73	8	6.89
$\bar{x}$	9.0		9.0		9.0		9.0	
$\bar{y}$	7.50		7.50		7.50		7.50	
$S(xx)$	110.0		110.0		110.0		110.0	
$S(yy)$	41.27		41.27		41.23		41.23	
$S(xy)$	55.01		55.00		54.97		54.99	

Fonte: Anscombe F.J., "Graphs in Statistical Analysis", in *American Statistician*, 27, pp. 17-21 (1973).

IN SINTESI

A COSA SERVE
✓ **A verificare cause reali, individuare e misurare i legami esistenti tra due variabili**

COME SI APPLICA
✓ Bisogna definire con chiarezza le variabili
✓ Bisogna sempre ricordare che lo studio delle relazioni esistenti tra le variabili costituisce la base della comprensione dei problemi
✓ Raccogliere i dati in condizioni omogenee

DOVE SI APPLICA
✓ Nella verifica di cause a valle del diagramma causa-effetto: da cause probabili a cause vere
✓ Nello studio delle interazioni tra cause e effetti

QUANDO SI APPLICA
✓ Dopo l’analisi causa-effetto

ERRORI DA EVITARE
✓ **Trarre conclusioni sulla base di pochi dati**
✓ **Dimenticarsi di utilizzare questo strumento nella verifica delle cause probabili**