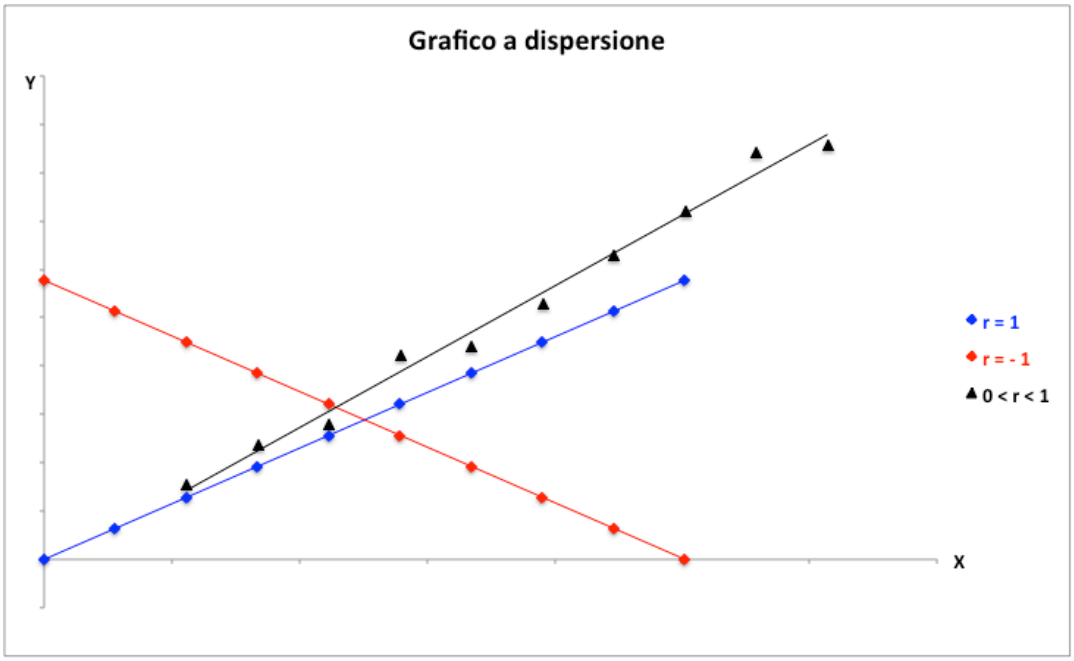
Grado di correlazione lineare

Si dice comunemente che esiste una correlazione fra due grandezze quando c'é una dipendenza dell'una dall'altra. Ad esempio c'é correlazione tra l'intensità di corrente circolante in un conduttore e la differenza di potenziale elettrico, fra la quantità di calore che si trasferisce tra due corpi a contatto e la loro differenza di temperatura e così via.
Gli esempi appena menzionati costituiscono, appunto, una correlazione nel senso statistico del termine, ovvero una corrispondenza biunivoca fra grandezze variabili.
Un modo elementare di stimare la correlazione consiste (come già sappiamo) nel "visualizzare" il valore delle grandezze variabili in un diagramma cartesiano, ottenendo punti le cui coordinate rappresentano la misura sperimentale della variabile indipendente (sull'asse x) e quelli corrispondenti della variabile dipendente (sull'asse delle y): questo tipo di diagramma é chiamato a dispersione.
Ovviamente non é possibile analizzare in modo soggettivo un diagramma a dispersione; per decidere se i dati sono linearmente correlati o meno, é necessario utilizzare una statistica che valuti l'entità della correlazione tra le due grandezze variabili.

Il "grado di correlazione" tra due grandezze qualsiasi x e y, misurate sperimentalmente, può essere valutato in modo alquanto raffinato, senza ricorrere al calcolo della deviazione standard. Salvo che non sia necessario esprimere il risultato della misura della costante di proporzionalità.
Si tratta di calcolare il **Coefficiente di correlazione lineare di Pearson r** (alcuni testi indicano il simbolo R o C_c) con la seguente formula (ovviamente il foglio elettronico e alcune calcolatrici tascabili "fanno i conti" per noi):

$$r = \frac{\sum\left(x_i - \bar{x}\right)\cdot\left(y_i - \bar{y}\right)}{\sqrt{\sum\left(x_i - \bar{x}\right)^2 \cdot \sum\left(y_i - \bar{y}\right)^2}}$$

Il coefficiente r può assumere tutti valori compresi fra -1 e +1. Si distinguono i seguenti casi particolari (vedi grafico sotto):



- ✓ **r = 1**: si ha la massima *correlazione positiva* tra le due grandezze, vale a dire una relazione in cui all'aumentare di X aumenta linearmente anche Y, ovvero si è alla presenza della **proporzionalità diretta**.
- ✓ **r = -1**: massima *correlazione negativa*, cioè all'aumentare di X diminuisce linearmente Y.
- ✓ **r = 0**: non si ha alcuna correlazione lineare tra le grandezze che, tuttavia, possono essere legate da una relazione matematica diversa da quella lineare.

In generale il valore assoluto del coefficiente di correlazione è minore di 1 e quanto più esso si avvicina a 1, tanto più le grandezze variabili in esame tendono a essere correlate linearmente. Più precisamente per la *correlazione positiva* si ha che:

- $0 < r \leq 0,3$: la correlazione è <u>debole</u>;
- $0,3 < r \leq 0,7$: la correlazione è <u>moderata</u>;
- $r > 0,7$: la correlazione è <u>forte</u>.

Al fine di precisare il livello della correlazione, è inoltre possibile determinare la **probabilità $P_N(r)$**, vale a dire *la probabilità che, per una data coppia di valori del numero di prove N e del coefficiente di correlazione r, le due grandezze fisiche sono correlate linearmente*.
La funzione $P_N(r)$ è alquanto complessa ma si ha a disposizione una tabella (di seguito riprodotta) in cui è indicato il valore percentuale della probabilità complementare $P_N^* = 100 - P_N$, cioè la probabilità che le due variabili <u>non</u> siano correlate linearmente.

PROBABILITÀ PER I COEFFICIENTI DI CORRELAZIONE (Tab. 13)

$C_C \rightarrow$ N ↓	0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1
3	100	94	87	81	74	67	59	51	41	29	0
4	100	90	80	70	60	50	40	30	20	10	0
5	100	87	75	62	50	39	28	19	10	3,7	0
6	100	85	70	56	43	31	21	12	5,6	1,4	0
7	100	83	67	51	37	25	15	8,0	3,1	0,6	0
8	100	81	63	47	33	21	12	5,3	1,7	0,2	0
9	100	80	61	43	29	17	8,8	3,6	1,0	0,1	0
10	100	78	58	40	25	14	6,7	2,4	0,5	-	0
11	100	77	56	37	22	12	5,1	1,6	0,3	-	0
12	100	76	53	34	20	9,8	3,9	1,1	0,2	-	0
13	100	75	51	32	18	8,2	3,0	0,8	0,1	-	0
14	100	73	49	30	16	6,9	2,3	0,5	0,1	-	0
15	100	72	47	28	14	5,8	1,8	0,4	-	-	0
16	100	71	46	26	12	4,9	1,4	0,3	-	-	0
17	100	70	44	24	11	4,1	1,1	0,2	-	-	0
18	100	69	43	23	10	3,5	0,8	0,1	-	-	0
19	100	68	41	21	9,0	2,9	0,7	0,1	-	-	0
20	100	67	40	20	8,1	2,5	0,5	0,1	-	-	0
25	100	63	34	15	4,8	1,1	0,2	-	-	-	0
30	100	60	29	11	2,9	0,5	-	-	-	-	0
40	100	54	22	6,0	1,1	0,1	-	-	-	-	0
45	100	51	19	4,5	0,6	-	-	-	-	-	0
50	100	49	16	3,4	0,4	-	-	-	-	-	0
60	100	45	13	2,0	0,2	-	-	-	-	-	0
70	100	41	9,7	1,2	0,1	-	-	-	-	-	0
80	100	38	7,5	0,7	-	-	-	-	-	-	0
90	100	35	5,9	0,4	-	-	-	-	-	-	0
100	100	32	4,6	0,2	-	-	-	-	-	-	0

Utilizzo della tabella

Facciamo un esempio: sono state svolte 5 prove ed è stato calcolato un valore assoluto del coefficiente di correlazione pari a 0,8.

"Incrociando" i parametri nella tabella, ho che:

$$\begin{cases} N = 5 \\ r = 0{,}8 \end{cases} \Rightarrow P_5^*(0{,}8) = 10\%$$

Ciò significa che:

$$P_5(0{,}8) = 100 - 10 = 90\%$$

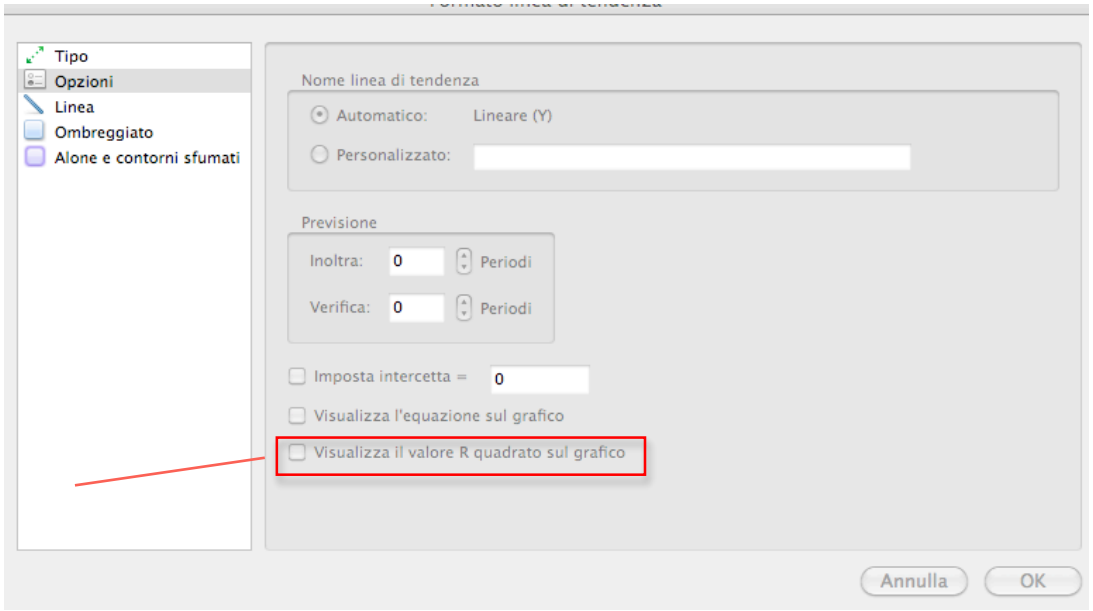
Vale a dire, la probabilità che le due grandezze sono linearmente correlate è del 90%.

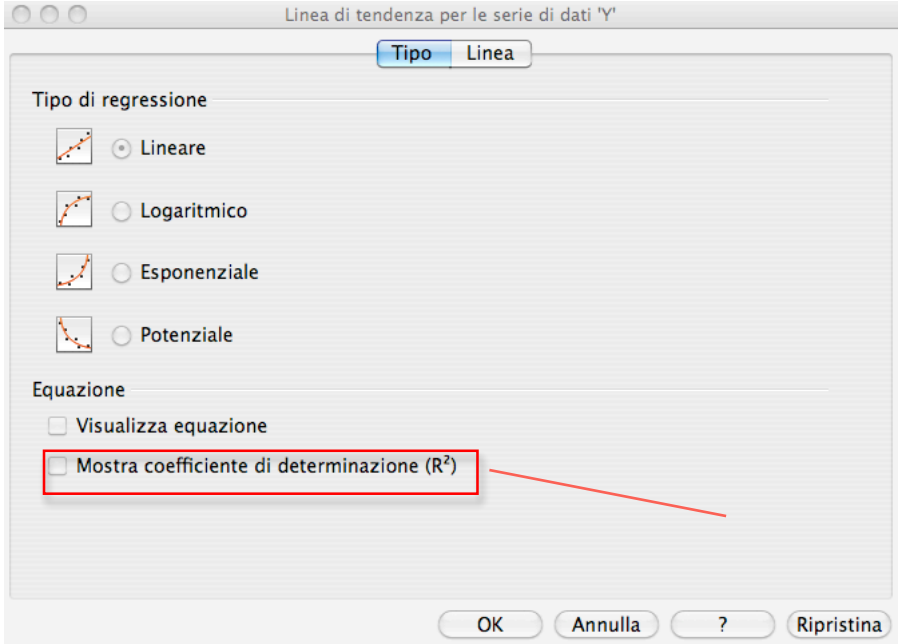
Va notato, infine, il "trattino" che compare nelle ultime celle a destra della tabella: esso sta a significare che il valore di P^* è molto piccolo, come dire è <<0,1, ovvero P>>99,9%.

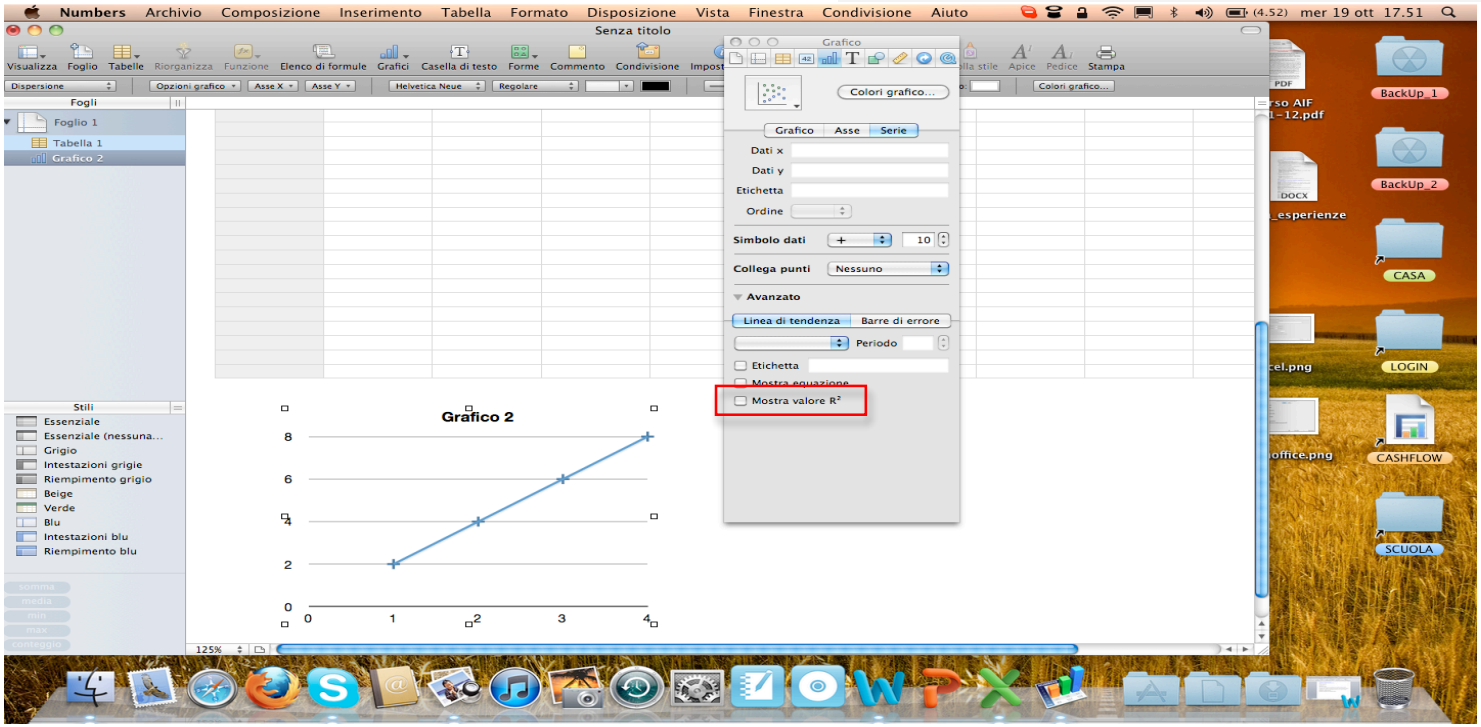
Coefficiente r e foglio elettronico

Sia *MS Excel* e *Open Office Calc*, sia *MacOSX Numbers* consentono, una volta tracciata la linea di tendenza o di andamento, di visualizzare vicino ad essa il valore del **coefficiente di correlazione al quadrato**, indicato con il simbolo $\mathbf{R^2}$ (vedi immagini alla pagina seguente). Esso è anche denominato *coefficiente di determinazione*.

Allo scopo è sufficiente "flaggare" l'opzione opportuna, come mostrato nelle immagini alla pagina successiva.


Tipo
Opzioni
Linea
Ombreggiato
Alone e contorni sfumati
Nome linea di tendenza
Automatico: Lineare (Y)
Personalizzato:
Previsione
Inoltra: 0 Periodi
Verifica: 0 Periodi
Imposta intercetta = 0
Visualizza l'equazione sul grafico
Visualizza il valore R quadrato sul grafico
Annulla
OK


Linea di tendenza per le serie di dati 'Y'
Tipo
Linea
Tipo di regressione
Lineare
Logaritmico
Esponenziale
Potenziale
Equazione
Visualizza equazione
Mostra coefficiente di determinazione (R²)
OK
Annulla
?
Ripristina


Numbers Archivio Composizione Inserimento Tabella Formato Disposizione Vista Finestra Condivisione Aiuto
Senza titolo
Grafico
Colori grafico...
Grafico Asse Serie
Dati x
Dati y
Etichetta
Ordine
Simbolo dati
Collega punti Nessuno
Avanzato
Linea di tendenza Barre di errore
Periodo
Etichetta
Mostra equazione
Mostra valore R²
Grafico 2
Fogli
Foglio 1
Tabella 1
Grafico 2
Stili
Essenziale
Essenziale (nessuna...
Grigio
Intestazioni grigie
Riempimento grigio
Beige
Verde
Blu
Intestazioni blu
Riempimento blu
BackUp_1
BackUp_2
CASA
LOGIN
CASHFLOW
SCUOLA