# 12) Metodo dei minimi quadrati e linea di tendenza

Si supponga di avere una tabella di dati $\{y^{\exp}{}_i\}_{i=1,\ldots,n}$ in funzione di altri dati $\{x_i\}_{i=1,\ldots,n}$ che siano il risultato di una qualche misura sperimentale. Tanto per rimanere in campo ambientale, supponiamo, ad esempio, che si sia misurata la concentrazione di polveri sottili (*pm10*) in una qualche unità di misura, in funzione del numero medio di auto per ora che passano in una determinata via:

| *Numero di auto per ora* | *Concentrazione di pm10* |
|---|---|
| **X** | **$Y_{exp}$** |
| 25 | 65 |
| 50 | 165 |
| 75 | 187 |
| 100 | 210 |
| 125 | 248 |
| 150 | 299 |
| 175 | 315 |
| 200 | 404 |

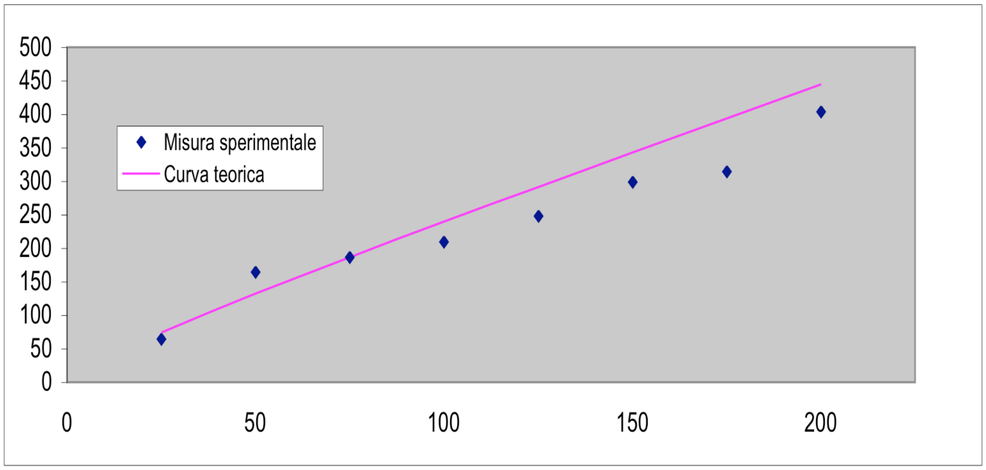
**Tabella 6**

Per quello che compete in questo corso, supporremo che le misure siano senza incertezza sperimentale. Ovviamente le misure sono sempre abbastanza casuali, in quanto sono soggette a errori di misura, complicati fattori ambientali eccetera.

Si supponga però che esista una legge teorica che leghi il valore della $x$ a quello della $y$, per esempio una legge del tipo

$$y^{\text{theo}}(x) = ax + b\sqrt{x}$$

dove $a$ e $b$ sono parametri da determinare. Per fissare le idee supponiamo di fissare il valore di $a$ pari a 1,8 e di $b$ pari a 6. Facciamo un grafico dei dati sperimentali e della curva teorica:




Come si vede, la curva sembra approssimare bene i dati sperimentali, ma non sappiamo quanto bene. Ci chiediamo ora quali sono i valori di $a$ e $b$ per cui la curva teorica meglio approssima i dati sperimentali. Per fare questo utilizziamo il cosiddetto *metodo dei* *minimi quadrati*. Definiamo *varianza* dei dati rispetto alla legge teorica, la funzione seguente:

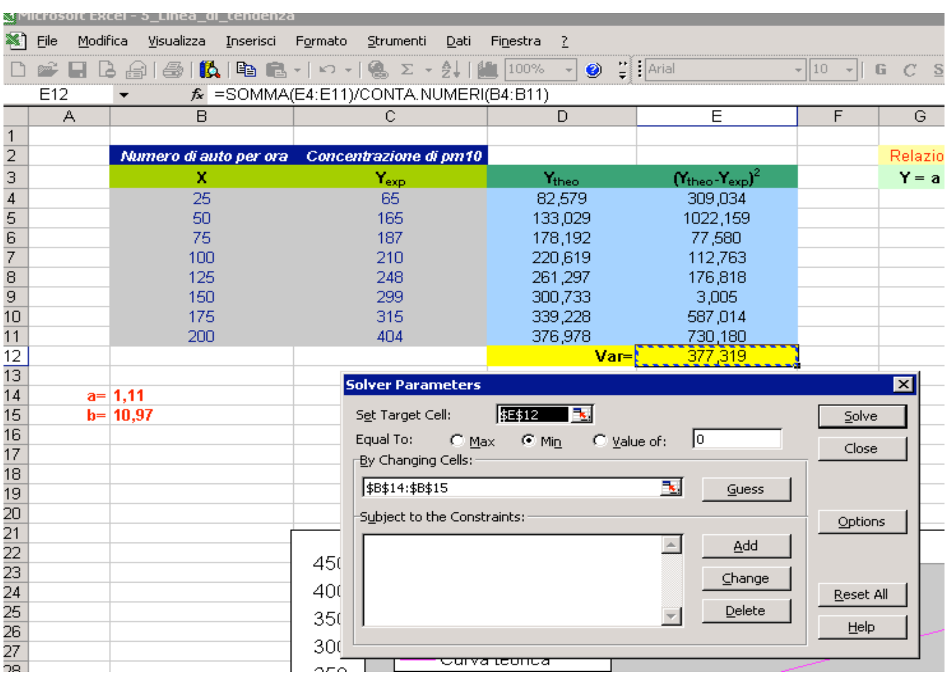
$$\mathrm{var} = \frac{1}{n}\sum_{i=1}^{n}\left[y_i^{\mathrm{theo}}(x_i) - y_i^{\mathrm{exp}}\right]^2$$

dove $n$ è il numero di dati sperimentali. La varianza è quindi la media degli scarti tra il valore teorico e quello sperimentale elevati al quadrato. Per sua natura sarà quindi sempre un numero positivo. Essa sarà funzione dei parametri incogniti $a$ e $b$ (nel nostro caso specifico). Il metodo dei minimi quadrati dice che **la curva teorica che meglio approssima i dati è quella che minimizza la varianza**. Se le misure sperimentali sono affette da errore, al posto della varianza conviene minimizzare la cosiddetta funzione $\chi^2$ definita come:

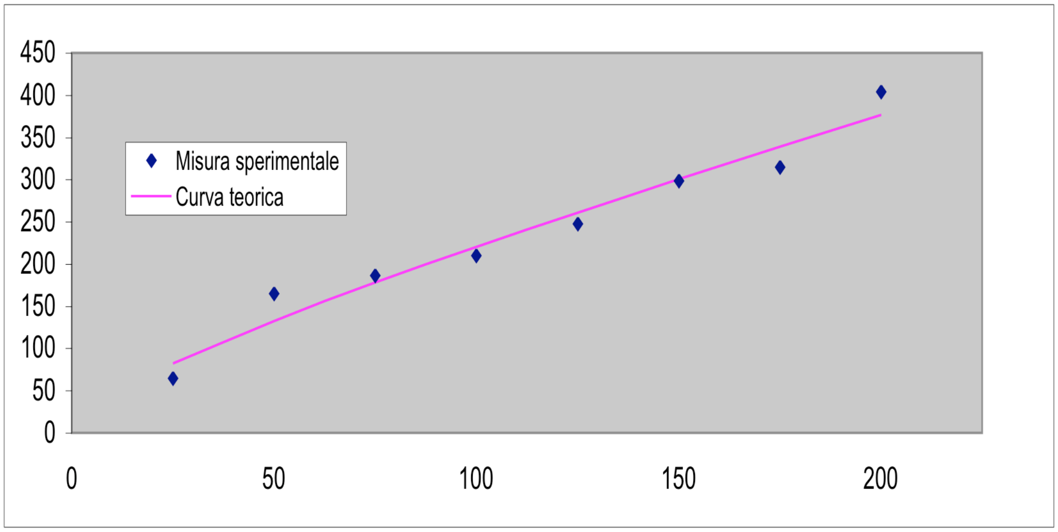
$$\chi^2 = \sum_{i=1}^{n}\left[\frac{y_i^{\mathrm{theo}}(x_i) - y_i^{\mathrm{exp}}}{\sigma_i}\right]^2$$

dove $\sigma_i$ è l'incertezza sulla misura $i$-esima. In questo modo le misure che hanno una maggiore precisione (minore $\sigma_i$) avranno un maggiore peso.

Il nostro esempio specifico è trattato nel file "`Linea_di_tendenza.xls`" nel foglio "minimi quadrati". La varianza è minimizzata tramite il risolutore. Si vede che i valori ottimali di $a$ e $b$ sono $a = 1{,}11$ e $b = 10{,}97$.




Il grafico della funzione con i valori ottimali dei parametri è il seguente:



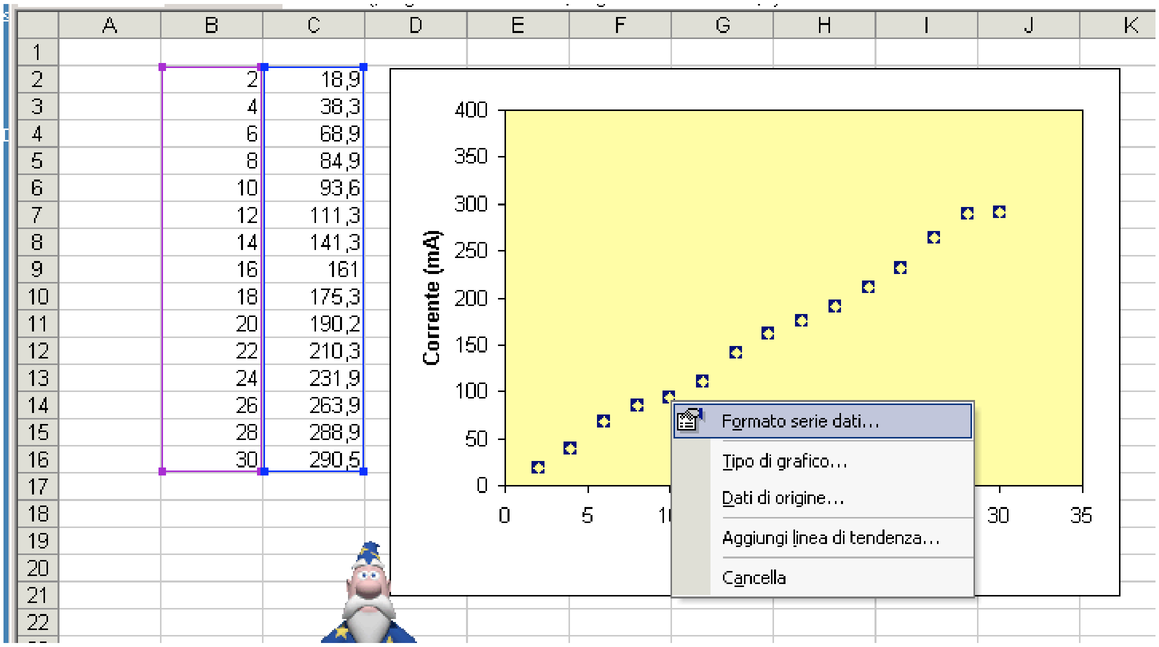

Come si vede la curva approssima molto meglio i dati di origine. La curva teorica assume anche i nomi di *linea di tendenza* o *di regressione* o *di best fit*.

Quando la linea di tendenza ha una forma semplice (ad esempio, una retta o un polinomio), la linea di tendenza è già implementata in EXCEL®, senza bisogno di calcolare la varianza. Facciamo un esempio: si supponga di aver misurato la corrente circolante in una resistenza in funzione della tensione applicata ai suoi capi:

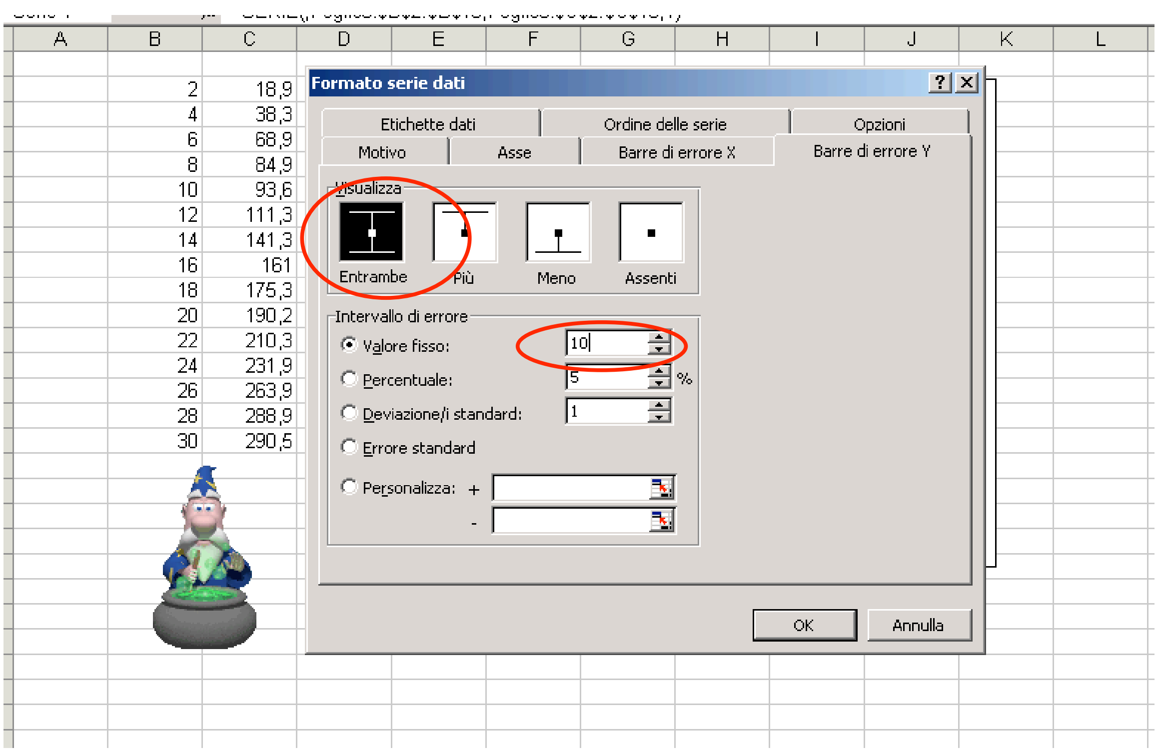
| Tensione (V) | Corrente (mA) |
|---|---|
| 2 | 18,9 |
| 4 | 38,3 |
| 6 | 68,9 |
| 8 | 84,9 |
| 10 | 93,6 |
| 12 | 111,3 |
| 14 | 141,3 |
| 16 | 161,0 |
| 18 | 175,3 |
| 20 | 190,2 |
| 22 | 210,3 |
| 24 | 231,9 |
| 26 | 263,9 |
| 28 | 288,9 |
| 30 | 290,5 |

**Tabella 7**

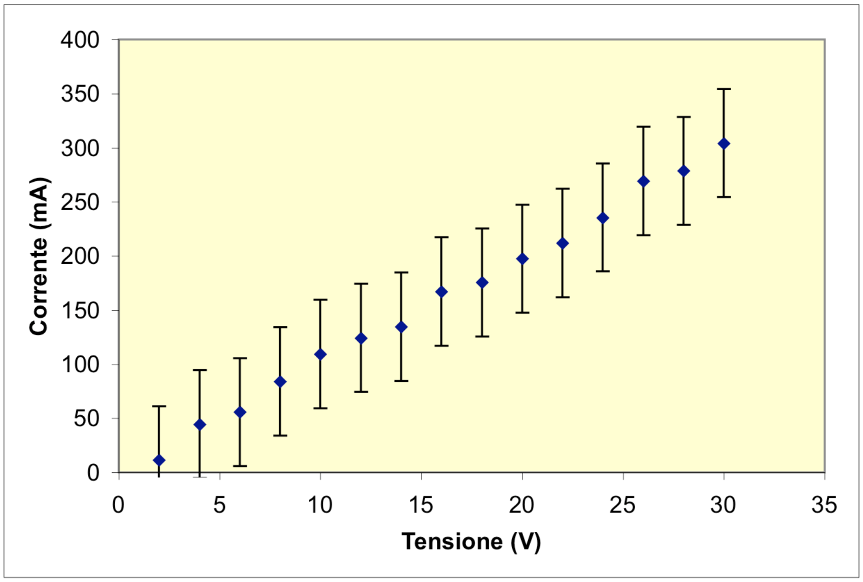
In questo esempio abbiamo supposto che il valore vero della resistenza sia 100 kΩ e abbiamo simulato un'incertezza di misura dovuta, ad esempio, ad una tolleranza dello strumento di misura della corrente, pari a ±10 mA per mezzo della funzione CASUALE descritta nel capitolo 6. Per evidenziare l'incertezza di misura occorre aggiungere delle barre di errore sui dati. Ciò è possibile selezionando uno qualunque dei punti sul grafico e premendo il tasto destro del mouse, selezionando successivamente "formato serie dati":



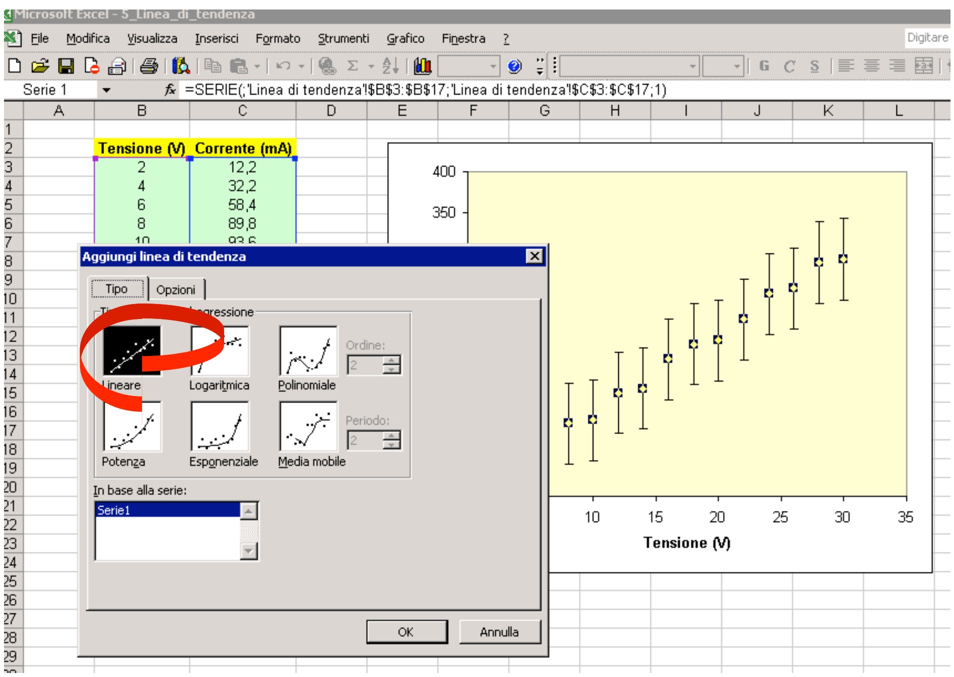

Nel sottomenù “Barre di errore Y” selezioniamo “Visualizza Entrambe” e “Valore fisso” pari a 10:



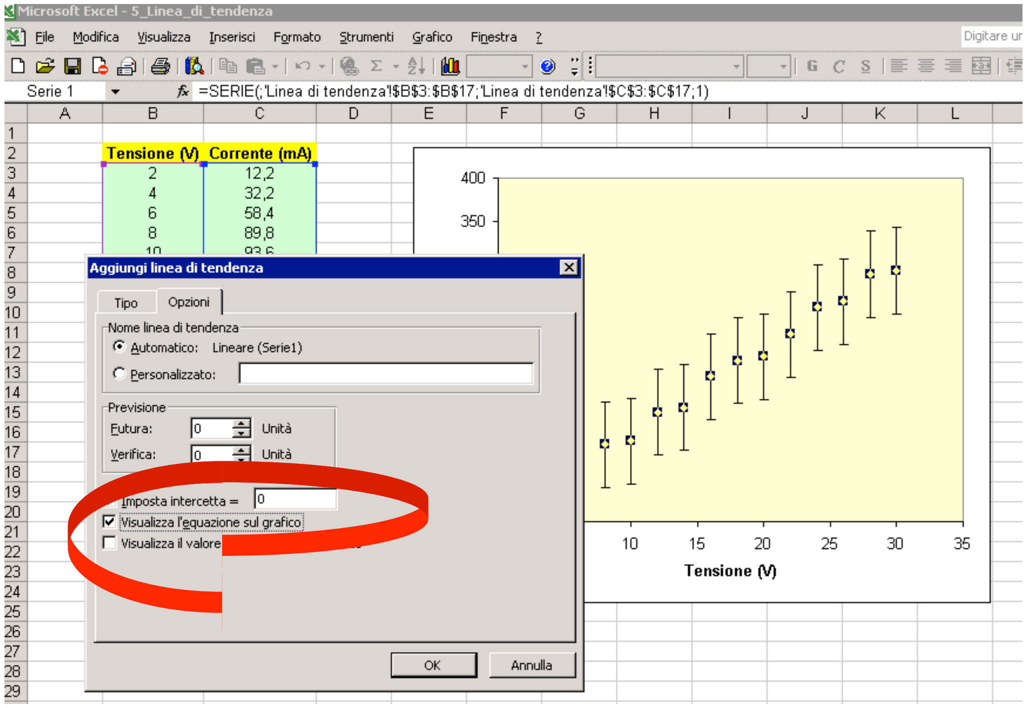

Premendo OK, sul grafico compariranno le barre di errore corrispondenti:



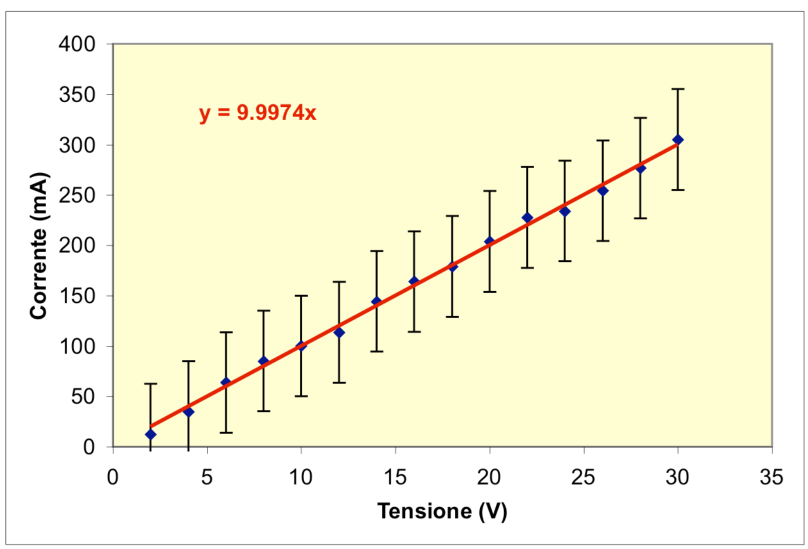

A questo punto, poiché sappiamo che tra la tensione applicata ai capi di una resistenza e la corrente circolante in essa vi è la relazione I = V/R, possiamo supporre che tra la $y$ e la $x$ vi sia una relazione lineare del tipo $y = ax$. Selezionando la serie di dati, con il tasto destro del mouse scegliamo l'opzione "aggiungi linea di tendenza"




Scegliamo ovviamente il tipo "lineare". Inoltre in "opzioni" impostiamo l'intercetta uguale a zero (poiché già sappiamo che l'equazione della retta è del tipo $y = ax$ piuttosto che $y = ax + b$, e quindi la retta deve passare per l'origine). Infine scegliamo anche di visualizzare l'equazione sul grafico.



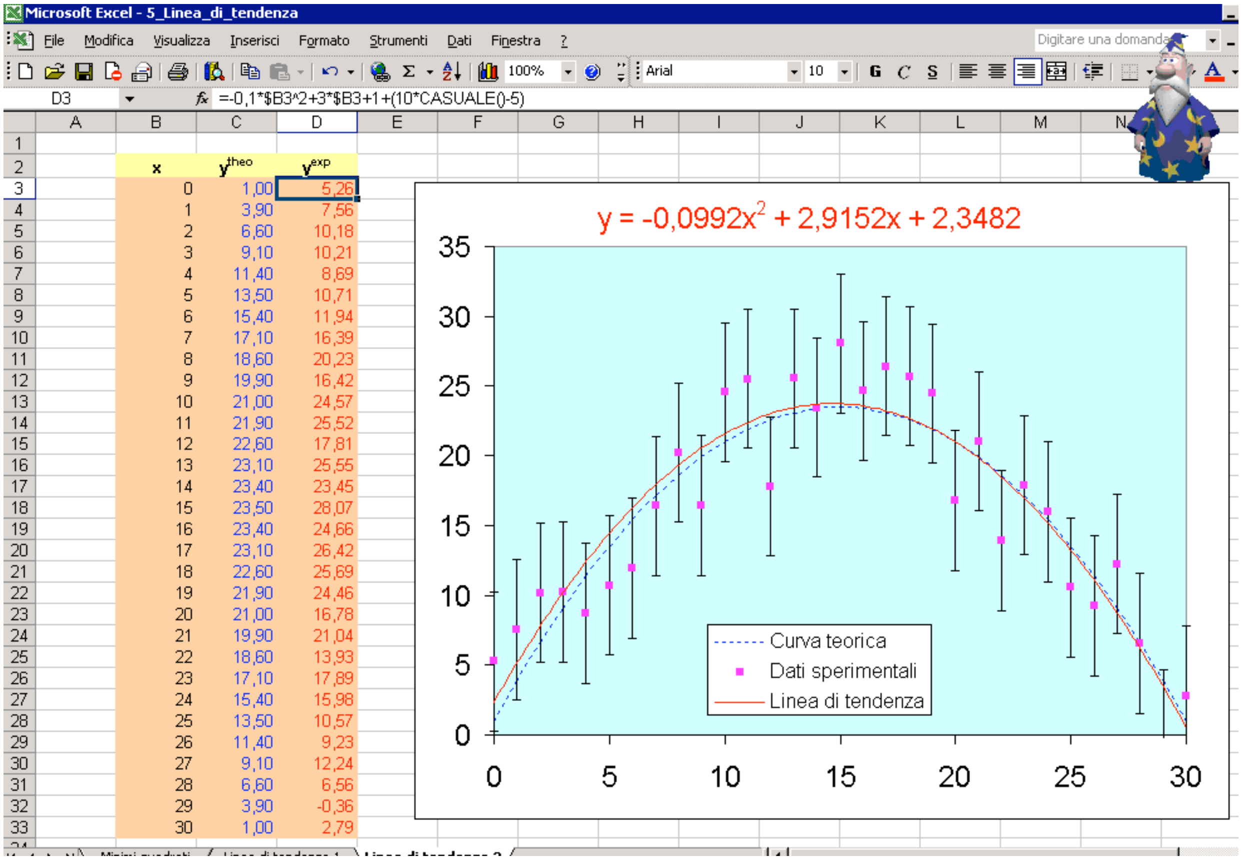

Il risultato è mostrato nella figura seguente:




Dalla relazione V = I/R deduciamo quindi il valore della resistenza

1/9,9974 V/mA = 100,02 kΩ

molto vicino al valore teorico di 100 kΩ. Nel file “`Linea_di_tendenza.xls`” è stato fatto anche un esempio di linea di tendenza di tipo polinomiale:




| x | $y^{theo}$ | $y^{exp}$ |
|---|---|---|
| 0 | 1,00 | 5,26 |
| 1 | 3,90 | 7,56 |
| 2 | 6,60 | 10,18 |
| 3 | 9,10 | 10,21 |
| 4 | 11,40 | 8,69 |
| 5 | 13,50 | 10,71 |
| 6 | 15,40 | 11,94 |
| 7 | 17,10 | 16,39 |
| 8 | 18,60 | 20,23 |
| 9 | 19,90 | 16,42 |
| 10 | 21,00 | 24,57 |
| 11 | 21,90 | 25,52 |
| 12 | 22,60 | 17,81 |
| 13 | 23,10 | 25,55 |
| 14 | 23,40 | 23,45 |
| 15 | 23,50 | 28,07 |
| 16 | 23,40 | 24,66 |
| 17 | 23,10 | 26,42 |
| 18 | 22,60 | 25,69 |
| 19 | 21,90 | 24,46 |
| 20 | 21,00 | 16,78 |
| 21 | 19,90 | 21,04 |
| 22 | 18,60 | 13,93 |
| 23 | 17,10 | 17,89 |
| 24 | 15,40 | 15,98 |
| 25 | 13,50 | 10,57 |
| 26 | 11,40 | 9,23 |
| 27 | 9,10 | 12,24 |
| 28 | 6,60 | 6,56 |
| 29 | 3,90 | -0,36 |
| 30 | 1,00 | 2,79 |

In questo caso, ad una curva teorica del tipo

$y = -0,1x^2 + 3x + 1$

è stato aggiunto un errore casuale di ±5. Interpolando i punti così ottenuti con una linea di tendenza polinomiale di grado 2 si ottiene

$y = -0,00992x^2 + 2,9152x + 2,3482$

che è molto vicina alla curva teorica di partenza.

## 13) Altre funzioni statistiche (cenni)

Si consideri un insieme di dati $\{x_i\}_{i=1...n}$. Si definisce *media* dei valori $x_i$ la quantità:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$