Statistica con Excel

Corso di Fisica ed Elementi di Laboratorio ed Informatica

CdL Scienze Biologiche

AA 2015/2016

Errori di misura

Il risultato della misurazione di una grandezza fisica non è mai un valore numerico esatto !

$$M \pm E$$

Dove

M è la misura (o la media delle misure) effettuata

E è l'errore

Errore Assoluto:

Dato un insieme di misure $\{x_1, x_2, \ldots, x_n\}$ si ha $E_a = (x_{max} - x_{min})/2$

Errore relativo:

E_r = Errore assoluto/Risultato di misura

Errore sistematico:

Dovuto ad un difetto dello strumento

Errore casuale:

Dovuto ad un errore nella procedura di misura

Errori sistematici

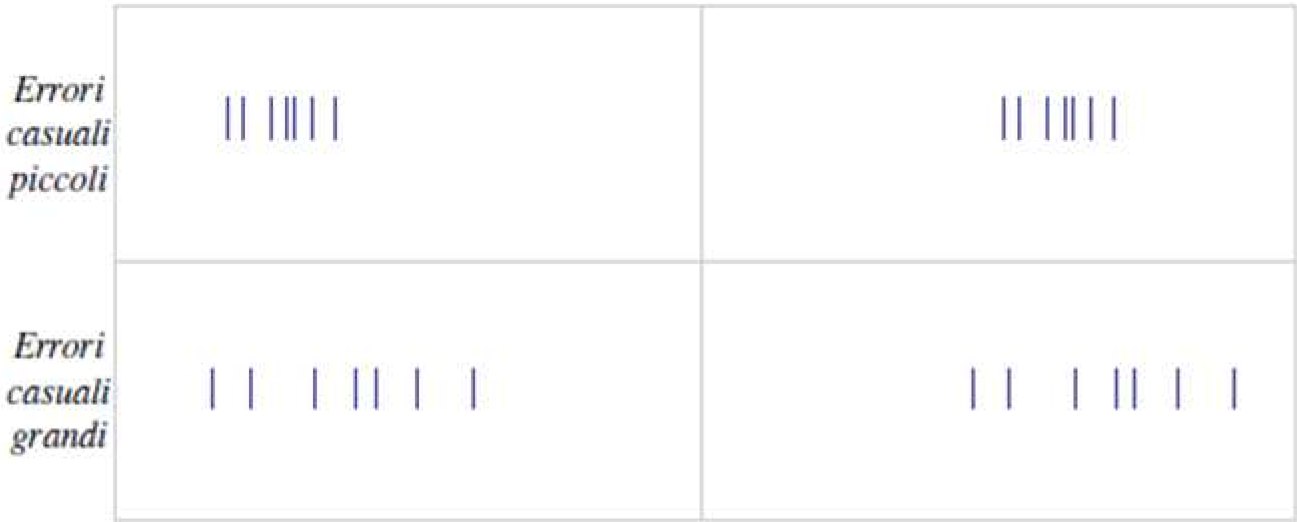
Errori sistematici: si ripetono sistematicamente ad ogni misura effettuata

- Sono legati a cause di errori intrinseche nel processo di misura
- Comprendono anche quelli dovuti ad errori umani nella definizione della procedura di misurazione e di calcolo

Errori casuali

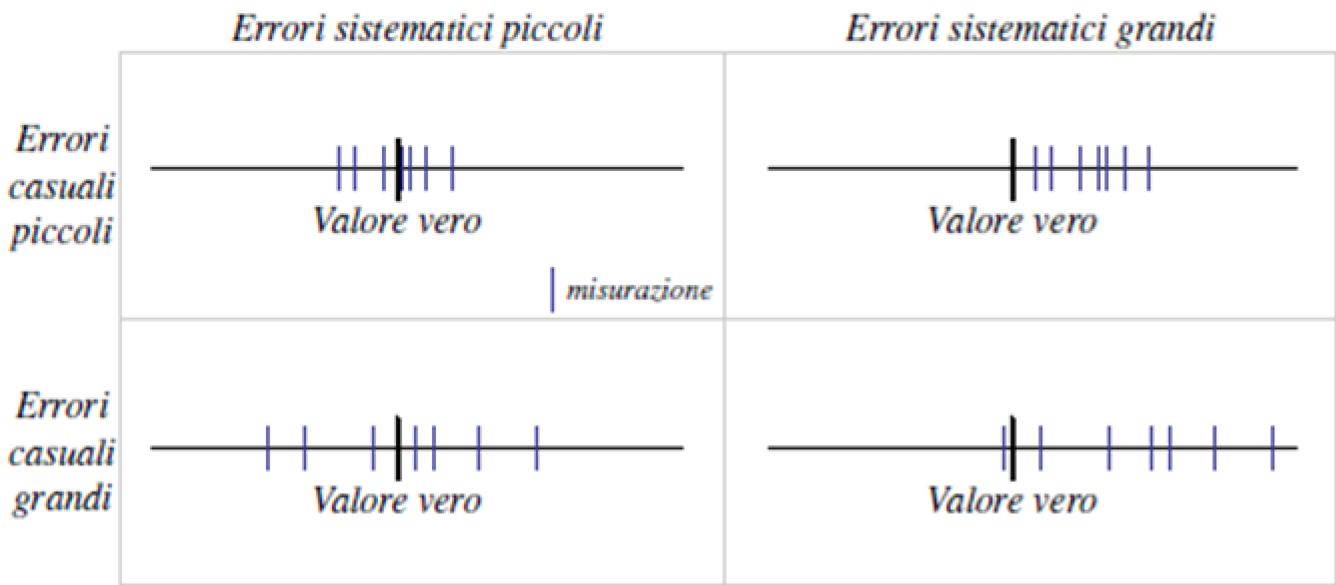
Errori casuali: variano in modo non prevedibile

- Ci si aspetta che con uguale probabilità causino sottostime e sovrastime
- Possono essere generati anche a rumori di fondo dello strumento
- Possono essere studiati con l'analisi statistica
- Possono essere minimizzati ripetendo le misure



Errori casuali e sistematici

- Conoscendo il valore atteso della misura possiamo capire se ci sono errori sistematici
- Conoscendo la distribuzione delle misure possiamo capire come sono gli errori casuali



Analisi statistica

Per minimizzare gli errori casuali effettuiamo molte misurazioni di una stessa quantità

La stima migliore è la *media* delle misurazioni

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

Nome	Espressione
Semisomma valori più lontani	$\frac{X_{min} + Xma_x}{2}$
Moda	Valore più frequente
Mediana	Valore corrispondente alla posizione centrale
Semidispersione	$\frac{X_{max} - Xm_{in}}{2}$

Analisi statistica

La stima migliore è la *media* delle misurazioni

$$\overline{x} = \frac{\sum_{i=1}^{n} X_i}{n}$$

La dispersione delle misure è data dalla *deviazione standard*

$$\sigma_x = \sqrt{\frac{\sum (X_i - \overline{x})^2}{n-1}}$$

- ✓ Grandi distanze dalla media hanno grande peso nella sommatoria
- ✓ La deviazione standard dà un'idea dello scarto complessivo dei valori intorno alla media

Analisi statistica

La variabile **SCARTO** corrisponde a

$$\boldsymbol{X_i - \overline{x}}$$

La **VARIANZA** corrisponde al quadrato della deviazione standard

$$\boldsymbol{\sigma_x}^2 = \frac{\sum(\boldsymbol{X_i - \overline{x}})^2}{\boldsymbol{n-1}}$$

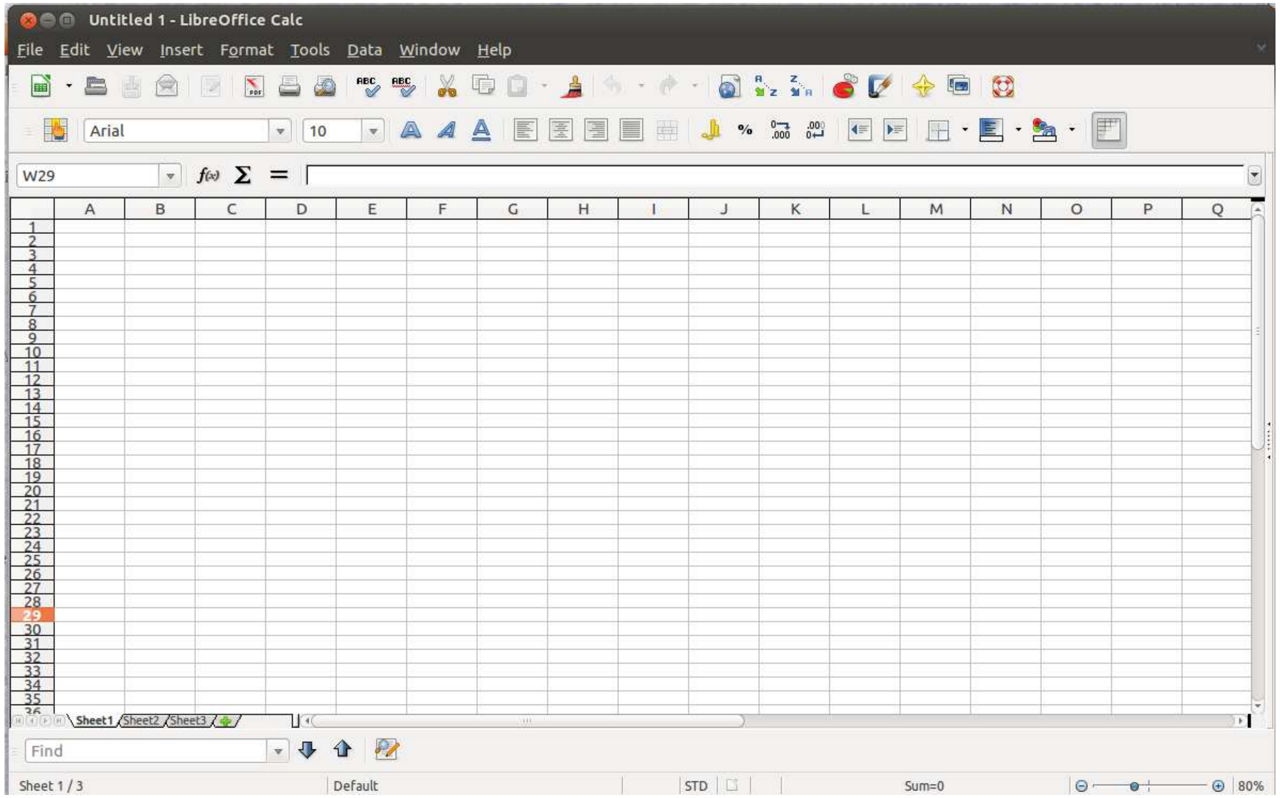
L'**errore standard** di una misura è definito come la stima della variabilità dello stimatore. Se abbiamo n campioni indipendenti con deviazione standard S_x l'**ERRORE STANDARD** è

$$\boldsymbol{S_x = \frac{\sigma_X}{\sqrt{n}}}$$

LibreOffice

Software libero che comprende una serie di programmi per l'elaborazione di testi, fogli di calcolo, presentazioni, grafici e disegni, database e formule matematiche.

LibreOffice CALC equivalente di Office Excel



Funzioni

Function	Syntax	Description
LOG	=LOG(n, base)	Returns the log of the number n to the specified base
LOG10	=LOG10(n)	Returns the log of the number n to the base 10
LN	=LN(n)	Returns the natural logarithm of n
EXP	=EXP(n)	Returns e^n

Funzioni trigonometriche

Function	Syntax	Description
SIN	=SIN(angle)	Returns the sine of the given angle, where angle is in radians.
COS	=COS(angle)	Returns the cosine of the given angle, where angle is in radians.
TAN	=TAN(angle)	Returns the tangent of the given angle, where angle is in radians.
ASIN	=ASIN(n)	Returns the inverse sine of the given number, n, where n must be within the range -1 to 1. The returned angle is in radians and within the range -p/2 to p/2.
ACOS	=ACOS(n)	Returns the inverse cosine of the given number, n, where n must be within the range -1 to 1. The returned angle is in radians and within the range 0 to p.
ATAN	=ATAN(n)	Returns the inverse tangent of the given number, n. The returned angle is in radians and within the range -p/2 to p/2.

Esercizi

1. Eseguire, utilizzando la proprietà di "trascinamento" la somma dei primi 131 numeri interi positivi.

2. Eseguire, dei primi 131 numeri interi, la seguente somma a segni alterni:1 -2 + 3 -4 + 5 -6 + … -130 +131.

MODO ALTERNATIVO

=Somma(A1:A10) restituisce la somma dei numeri contenuti nelle caselle che vanno da A1 ad A10

Ripetiamo gli esercizi utilizzando questa funzione!!!

Funzione SE

La funzione SE esegue delle operazioni condizionali, ovvero, esegue un'operazione o un'altra a seconda che il risultato di un test logico sia vero o falso.

=SE(test; se_vero; se_falso)

- **test** è l'espressione che noi vogliamo verificare
- **se_vero** è l'operazione da compiere se il risultato del test è vero
- **se_falso** è l'operazione da compiere se il risultato del test è falso

Esempio: Supponiamo di volere il valore assoluto del numero contenuto nella casella B3

=SE(B3>0,B3,-B3)

Ripetiamo l'esercizio 2 utilizzando questa funzione:
Eseguire, dei primi 131 numeri interi, la seguente somma a segni alterni:1 -2 + 3 -4 + 5 -6 + … -130 +131

Esercizio (1)

L (mm)	
9.87	9.81
9.76	9.79
9.78	9.82
9.83	9.80
9.81	9.84
9.82	9.82
9.90	9.77
9.75	9.83
9.83	9.82
9.79	9.80

I risultati della misura dello spigolo di un cubo con un calibro Palmer con sensibilità di 0.01 mm sono riportati nella tabella a sinistra

Calcolare:
1. La semisomma dei valori più lontani
2. La mediana
3. La moda
4. La media aritmetica
5. La media pesata
6. La semidispersione
7. Gli scarti
8. Lo scarto quadratico medio
9. Errore sul valor medio

Istogrammi

L'istogramma è una rappresentazione grafica dei valori di una misura ripetuta che ha dato risultati diversi, o dei valori di un campione estratto casualmente da una popolazione.

Per costruire un istogramma da un insieme di misure, è necessario organizzare le stesse in ***classi***, ovvero dividere l'intervallo dei possibili risultati in sotto-intervalli opportuni.

Normalmente le classi vengono scelte di uguale grandezza.

Il numero di classi da definire e la loro estensione vanno decisi di volta in volta, e non esiste una regola univoca per determinarle.

Istogrammi

Supponiamo di avere misurato N=20 lunghezze e le nostre misure sono comprese fra i valori:

L_{min}= 8 cm ed L_{max}= 12 cm

Se vogliamo definire C=4 classi sarà sufficiente considerare l'intervallo dei valori:

$\Delta L = L_{max} - L_{min} = 4$ cm

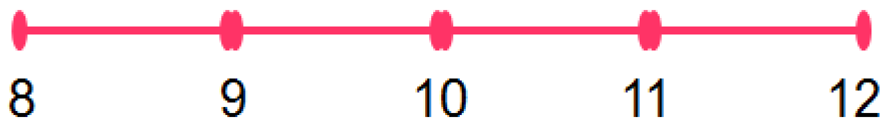
e dividerlo per il numero di classi per ottenere l'ampiezza di ciascuna classe:

$\Delta C = \Delta L / C = 4$ cm$/4 = 1$ cm

Istogrammi

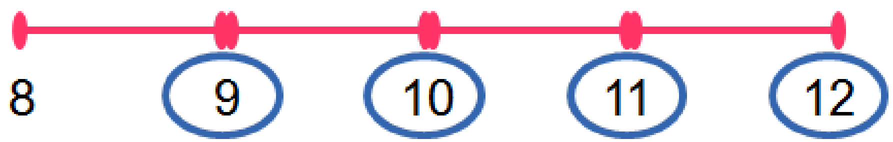
In questo modo avremo definito 4 classi:

1 – da 8 cm a 9 cm
2 – da 9 cm a 10 cm
3 – da 10 cm a 11 cm
4 – da 11 cm a 12 cm



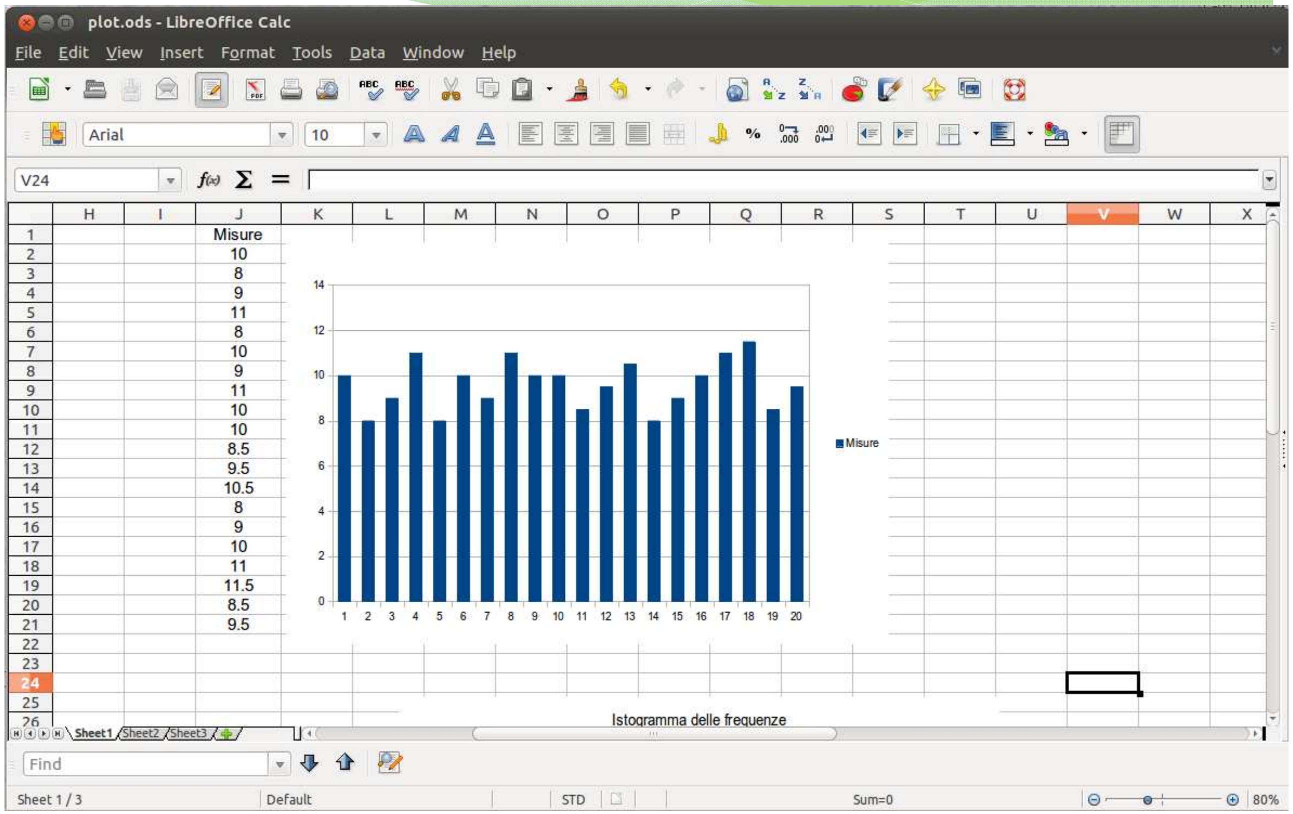
N.B.
Definendo C=4 classi otteniamo però C+1=5 valori, poiché ogni classe ha un estremo inferiore ed uno superiore!!!

Nel foglio elettronico, per la definizione delle C=4 classi, ci occorrono, invece, solo 4 numeri.
Considereremo quindi solo gli estremi superiori:



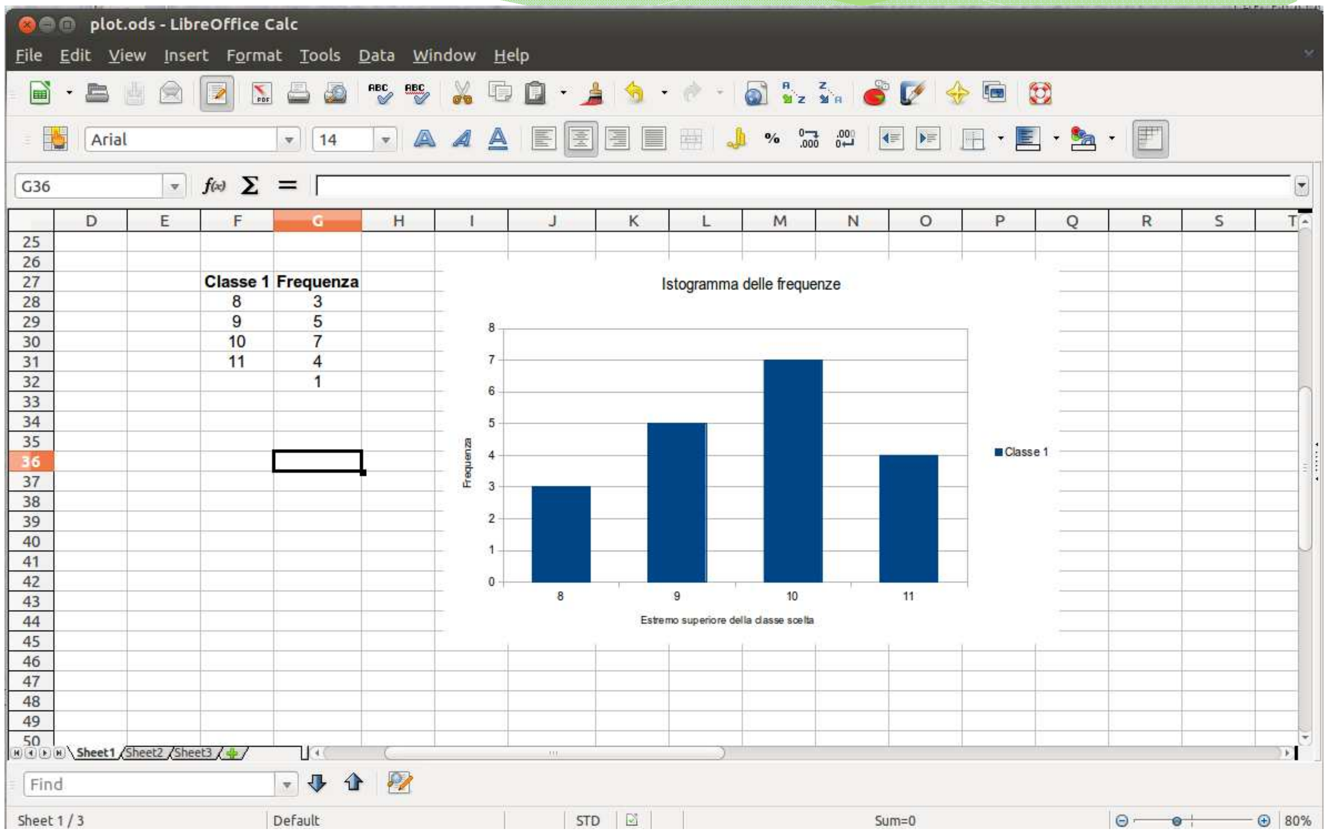
Istogrammi

Set Misure 1	
10,00	8,50
8,00	9,50
9,00	10,50
11,00	8,00
8,00	9,00
10,00	10,00
9,00	11,00
11,00	11,50
10,00	8,50
10,00	9,50



Istogrammi

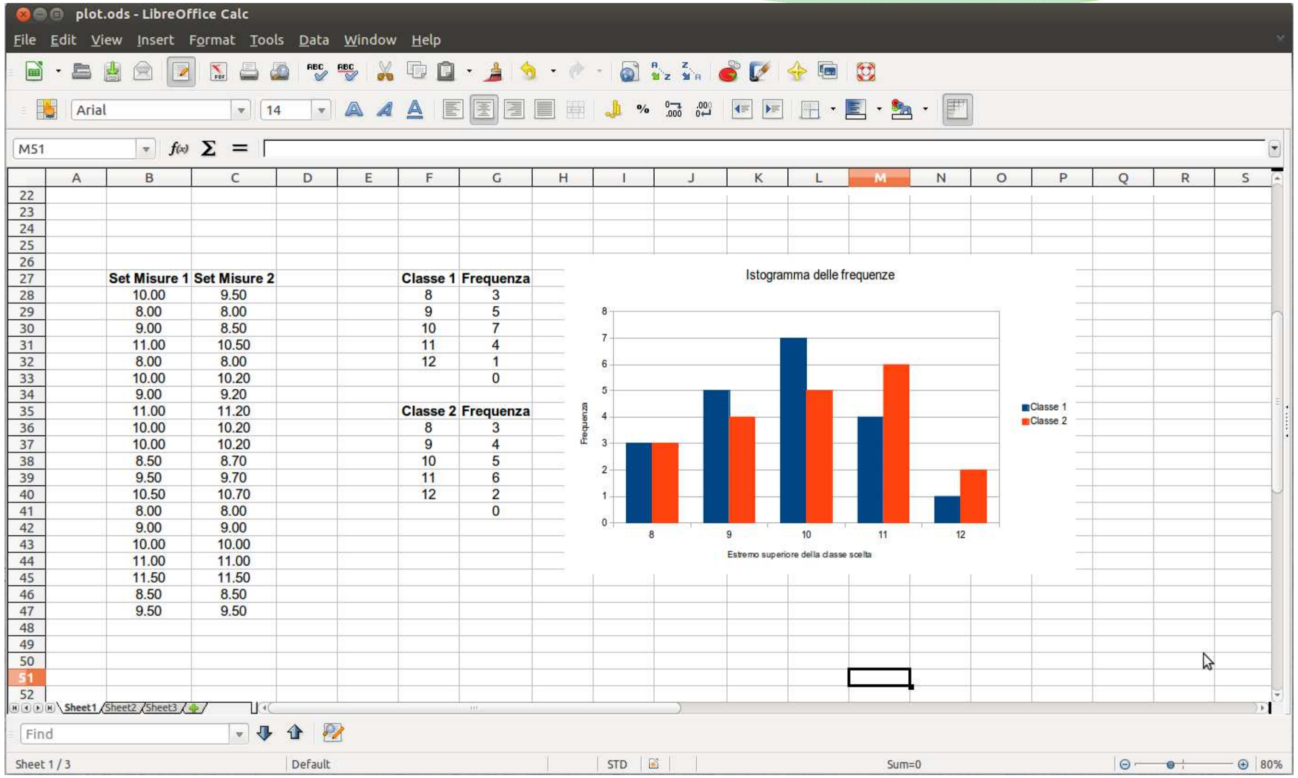
Set Misure 1	
10,00	8,50
8,00	9,50
9,00	10,50
11,00	8,00
8,00	9,00
10,00	10,00
9,00	11,00
11,00	11,50
10,00	8,50
10,00	9,50



Istogrammi

Aggiungiamo un altro set di misure

Set Misure 2	
9,50	8,70
8,00	9,70
8,50	10,70
10,50	8,00
8,00	9,00
10,20	10,00
9,20	11,00
11,20	11,50
10,20	8,50
10,20	9,50



Esercizio (2)

L (mm)	
9.87	9.81
9.76	9.79
9.78	9.82
9.83	9.80
9.81	9.84
9.82	9.82
9.90	9.77
9.75	9.83
9.83	9.82
9.79	9.80

Rappresentare la distribuzione dei risultati in un istogramma

Esercizi

1. Formare una tabella per calcolare il quadrato, il cubo, la radice quadrata e la radice cubica dei primi 15 numeri naturali; Tracciare il grafico delle varie quantità in funzione del numero.

2. Tracciare il diagramma cartesiano, con x ϵ al dominio [-3 , +3], della funzione seguente, con almeno 60 valori di x: $y = x^3 - 4x$

Esercizi

Supponiamo di avere una tabella di dati y_i in funzione di altri dati x_i che siano il risultato di una misura sperimentale.
Ad esempio, supponiamo 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

Tensione (V)	Corrente (mA)
20	190.2
22	210.3
24	231.9
26	263.9
28	288.9
30	290.5

Inserire sul grafico ottenuto le barre di errore per la corrente dove l'errore è ±10 mA

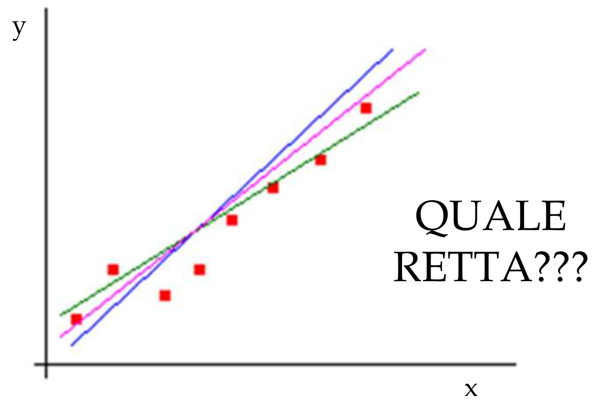
Minimi quadrati

Sappiamo che tra le due grandezze x *(Tensione)* ed y *(Corrente)* esiste una relazione

Variando opportunamente la grandezza x si osserveranno delle variazioni anche nella grandezza y.

Tramite i valori delle coppie (x_i, y_i) è possibile, utilizzando metodi statistici, determinare i coefficienti a e b, ed anche i loro errori *statistici*.
In particolare con il **METODO DEI MINIMI QUADRATI** possiamo trovare la curva teorica che approssima meglio i dati sperimentali.

In particolare, in questo caso possiamo effettuare una *Regressione Lineare*
y=ax+b



Minimi quadrati

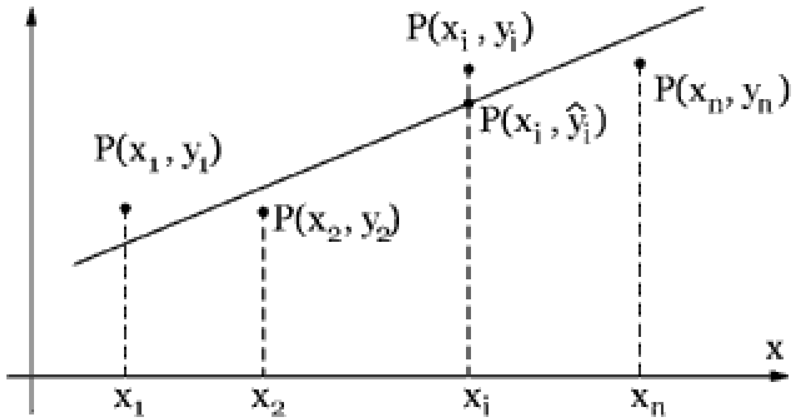
Approssimazione: Gli errori sulla variabile indipendente x sono trascurabili

Abbiamo supposto una legge di tipo lineare, per cui la coppia di valori (x_i, y_i) rispetterà questa relazione solo se $y_i = a\,x_i + b$

La quantità

$$\delta_i = y_{teo} - y_{dati} = (a\,x_i + b) - y_i$$

misura quindi la distanza tra il dato teorico e quello sperimentale.



Minimi quadrati

Consideriamo quindi la media di queste distanze (o scarti) al quadrato che rappresenterà la varianza dei dati rispetto alla legge teorica dove abbiamo n punti sperimentali

$$var = \frac{1}{n}\sum_{i=1}^{n}[y_i^{teo} - y_i^{dati}]^2$$

- **Sempre positivo !**
- **Dipende dai parametri a e b !**

Il metodo dei minimi quadrati dice che la curva teorica che meglio approssima i dati è quella che *minimizza* la varianza.

Esercizio (1)

Supponiamo di aver realizzato un'esperienza per la misura della distanza focale f di una lente. Le grandezze misurate sono state la distanza p dell'oggetto dalla lente e la distanza q dell'immagine dalla lente, che soddisfano la legge dei punti coniugati **1/p + 1/q = 1/f**. I dati registrati ed i relativi errori sono riportati nella seguente tabella:

1. Avendo posto x=1/p e y=1/q, la legge dei punti coniugati si può scrivere come y=ax+b, dove b=1/f e a=-1.
2. Rappresentare i dati x_i e y_i su un grafico. Sono allineati?
3. Determinare la linea di tendenza.
4. Effettuare una regressione lineare dei dati con Excel. Calcolare a e b con i loro errori. Il valore di a è compatibile con -1? Dal valore di b ricavare f=1/b con il suo errore $\Delta f/f=\Delta b/b$.

Esercizio (2)

p (cm)	q (cm)	Δp (cm)	Δq (cm)	x (cm^{-1})	y (cm^{-1})	Δx (cm^{-1})	Δy (cm^{-1})
15.3	29.1	0.2	0.6				
16.3	26.6	0.2	0.6				
17.5	22.8	0.2	0.7				
19.4	19.9	0.2	0.8				
20.9	19.5	0.2	1.2				
22.5	18.6	0.2	1.3				
23.5	16.4	0.2	1.3				
25.1	17.0	0.2	1.6				
26.7	16.8	0.2	1.8				
28.1	15.0	0.2	1.7				
29.2	15.2	0.2	2.0				
30.2	15.3	0.2	2.2				
31.7	14.0	0.2	2.0				
32.7	14.4	0.2	2.3				
34.2	14.0	0.2	2.5				
35.8	13.3	0.2	2.5				
37.0	13.4	0.2	2.5				
38.1	14.3	0.2	2.6				
40.0	13.0	0.2	2.8				
41.5	12.8	0.2	3.2				