Corso integrato di informatica, statistica e analisi dei dati sperimentali
Esercitazione VI

Esercizio 1) Presentazione dei risultati: cifre significative, errore.

a) Un risultato sperimentale X deve sempre essere riportato insieme alla sua incertezza (errore) δX. Nel calcolo di grandezze (X) ed errori (δX) si ottengono spesso numeri con molte cifre *significative*: ad esempio da una misura di lunghezza si ottiene X = 12.345 678 183 mm, δX = 0.025 679 357 mm. Nel riportare un risultato le cifre significative indicano la precisione con cui si conosce un valore. Dal momento che l'errore deriva da una stima e indica a sua volta l'incertezza con cui si conosce una grandezza, non ha senso indicarlo con una precisione maggiore di una cifra significative (Tuttavia se la prima cifra è 1, allora è ammesso riportare l'errore con due cifre significative). Se l'errore si ottiene da una procedura statistica di propagazione degli errori, il numero di cifre significative da usare è sempre 2. Quindi, il valore dell'errore (incertezza) deve essere sempre arrotondato a 1 o massimo due cifre significative. Per ragioni pratiche (e a volte di leggibilità delle tabelle) è ammesso riportare sempre l'incertezza con due cifre significative. Stabilite le cifre significative dell'errore, le cifre con cui si riporta il valore della grandezza devono essere accordate con quelle dell'errore, quindi:

$$X = (12.345 \pm 0.026)mm \qquad \text{oppure} \qquad X = (12.35 \pm 0.03)mm$$

Esistono diverse convenzioni per riportare l'errore:

i)

$$X = (12.345 \pm 0.026)mm;$$

Per numeri molto piccoli o molto grandi si usa la notazione esponenziale:
X = (0.000 23± 0. 0000 03)*mm* diventa:

$$X = (2.3 \pm 0.3)\cdot 10^{-4}\ mm.$$

Nota: utilizzando un foglio elettronico la notazione esponenziale è 2.3E-4

ii)

$$X=12.345(26)\ mm$$

in questo caso (suggerito dalle norme ISO) il numero tra parentesi rappresenta l'incertezza sul valore della misura. In notazione esponenziale:

$$X = 2.3(3)\cdot 10^{-4}\ mm$$

iii) Si può utilizzare l'errore relativo: $\varepsilon_r = \delta X/X$ e l'errore relativo percentuale $\varepsilon_r\% = 100\ \varepsilon_r$.
Per X = (12.35 ± 0.03) si ha: $\varepsilon_r = 0.026/12.345 = 0.002$ e $\varepsilon_r\% = 0.2\%$, quindi:

$$X = 12.35\ mm \pm 0.2\% \quad \text{oppure } X = 12.35\ (1 \pm 0.002)\ mm$$

Presentare i seguenti valori e i relativi errori su una tabella Excel o in un file (Word, notepad, wordpad etc…)

X	δX	u.m.	X	δX	u.m.
0.000541272	0.00000234	g	8.807696	1.698535	mg
6.963879	0.345670	km	13.05671	$4.2883\cdot 10^{-2}$	V
23.54947	0.0345	A	$9.28836\cdot 10^{-2}$	4.28836E-07	mF
156.9618	6.9348765	mm	696.4838	15.77133	m^2
656.3443	0.23445	N	0.00084373	0.0001254	cm^3

In Excel si possono combinare i comandi CONCATENA e TESTO per combinare testo e valori in una singola cella, (vedere: **Cifre_sgnificative.xls** nel foglio **Tabella** per un paio di esempi)

b) Se un risultato si ottiene come media di N valori: $\bar{X} = \frac{1}{N}\sum_{i=1}^{N} x_i$

l'errore sulla media si calcola usando la stima campionaria della varianza: $s_x^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{X})^2$

L'errore standard della media è: $\delta_x = s_{\bar{X}} = \frac{s_x}{\sqrt{N}}$

Il file **Cifre_sgnificative.xls** (foglio **DATI**) contiene una tabella con misure ripetute di diverse grandezze. Per ognuna di esse calcolare media ed errore standard della media. Riportare in una tabella (Excel e/o Word e/o wordpad e/o notepad, etc…) valori medi ed errore. Anche se (a mio avviso) un foglio EXCEL non è lo strumento ideale per costruire tabelle formattate, Il foglio **Risultati** mostra un possibile metodo per il calcolo dei risultati e due modi per formattare correttamente i risultati della tabella.

	Misure			
A	B	C	D	E
mm	A	kg	V	g
0.53947334	12.16367685	37.44687296	534.5187998	0.00045129
0.18813654	4.982749996	2.323192103	158.7667736	0.000449743
0.094677681	0.571813353	2.272135142	62.60807965	0.000451205
0.431099087	8.196187687	22.79399235	432.8293555	0.000447903
0.737510572	7.231732146	19.3346109	712.2755047	0.000448572
0.662503966	17.01755586	51.03118001	699.5691494	0.000447054
0.527759089	10.84082011	11.24592596	520.6907689	0.000448802
0.04212186	6.710294239	3.057497897	47.70686193	0.000445294
0.933375598	17.26836611	38.2932469	959.1032936	0.000452558
0.082763182	4.889910192	2.650997973	80.34376713	0.00044644
0.51700542	6.989039963	10.68801888	514.0657509	0.000449614
0.647861216	9.862046986	52.82313355	637.5673958	0.000450975
0.820801463	12.40075956	27.81716318	818.1506957	0.00045086
0.948642283	15.91903286	64.232884	956.949517	0.000451791
0.850429938	9.24629048	14.15151493	822.6918956	0.000449613
0.494717296	6.624110868	37.46936801	472.1845189	0.000444024
	11.66822279	-3.348113751		0.00045219
	2.045773326	0.678078322		0.000450324
	-2.63514682	1.872076609		0.000452299
	8.369214506	-4.217308742		0.00044941
	4.708225409			0.000454706
	20.72407419			0.000452544
	10.13398811			0.000447444
	7.268794472			0.000452467
	12.15947835			
	10.21993939			

	A	B	C	D	E
N	16	26	20	16	24
media	0.532429908	9.060651961	19.63082336	526.876383	4.49880E-04
varianza	0.090356046	27.67371195	434.399577	94181.33632	6.67098E-12
dev. st	0.300592824	5.260580952	20.84225461	306.8897788	2.58282E-06
err.media	0.075148206	1.031684805	4.66046981	76.72244469	5.27217E-07

Misure				
A	B	C	D	E
mm	A	kg	V	m
0.53(8)	9(1)	19(5)	5.3(8)E+2	4.499(5)E-4

Misure				
A	B	C	D	E
mm	A	kg	V	m
0.53	9	20	5.3E+02	4.499E-04
0.08	1	5	8E+01	5E-07

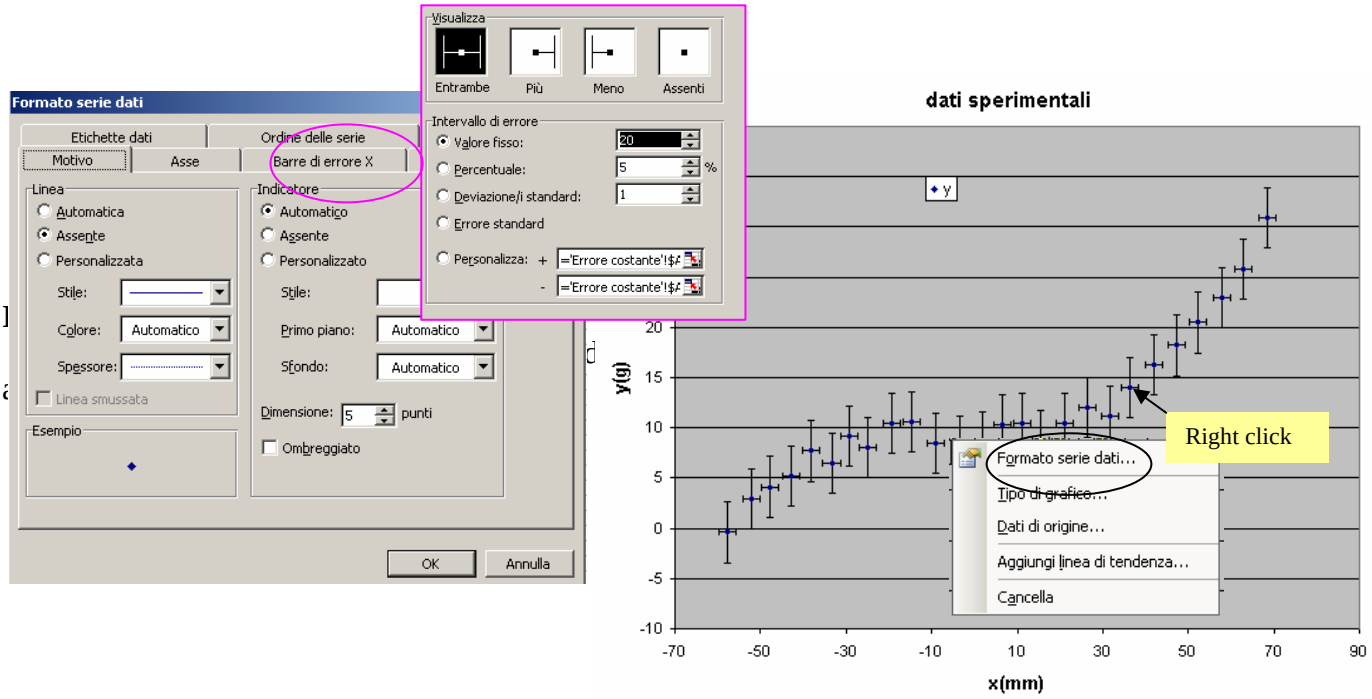
Formattazione automatica usando le funzioni TESTO(...) e CONCATENA(...)

Misure				
A	B	C	D	E
mm	A	kg	V	m
0.53 +/- 0.08	9.06 +/- 1.03	20 +/- 5	527 +/- 77	4.4988E-4 +/- 5E-7

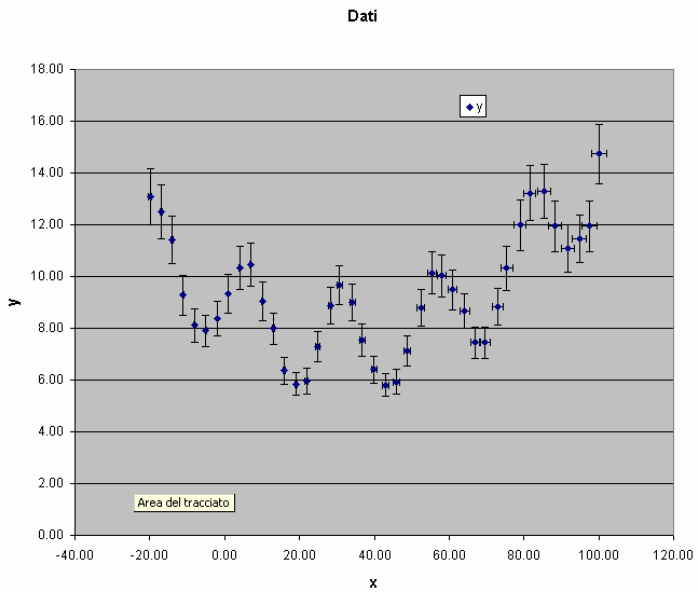
Esercizio 2) Presentazione dei dati con errore.

a) Il file **dati_xy_A.dat** contiene due colonne di dati x e y(x). Gli errori sulla variabile indipendente (x) e sulla variabile dipendente (y) sono costanti. (dx = 2 mm e dy = 3 mm). Riportare su un grafico xy i dati con le barre d'errore. Utilizzare il tasto destro del mouse su uno dei punti del grafico per attivare l'opzione "*formato serie di dati*" e inserire l'errore costante per x e y. Nota: è abbastanza complicato rispettare le regole che impongono di riportare l'errore con due sole cifre significative in tabelle di dati excel, può essere tollerato utilizzare più cifre significative su tutta la tabella, l'importante è che le cifre significative dei dati siano accordate con l'espressione dell'errore.

```
#       x       y
#(err.=2)  (err.=3)
#     [mm]      [A]
      -58        0
      -52        3
```



Il file **dati_xy_B.dat** contiene 4 colonne di dati : due colonne per la variabile indipendente x con il suo errore (per ogni punto) e due colonne per la variabile dipendente y con l'errore associato per ogni punto. Riportare su un grafico xy i dati con le barre d'errore. Utilizzare il tasto destro del mouse su uno dei punti graficati per attivare l'opzione "formato serie di dati" e inserire nel campo "errore personalizzato" i riferimenti agli errori



Error2.dat

```
#          dati
# X     err.x     y    err.y
#     m              V
#
-19.91   -0.40    13.10    1.08
-16.89   -0.34    12.50    1.04
-13.95   -0.28    11.42    0.91
-11.09   -0.22     9.28    0.76
-8.04    -0.16     8.11    0.65
-5.03    -0.10     7.90    0.61
-1.99    -0.04     8.39    0.66
```

Per la rappresentazione grafica dei dati si può usare il programma GNUPLOT (vedi lezioni)

- Attivare gnuplot
- Spostarsi nella directory di lavoro

Provare ad utilizzare i seguenti comandi:

```
set title 'Errore Costante'
set xlabel 'x [mm]'
set ylabel 'y [A]'
plot 'dati_xy_a.dat' using 1:2:(2.0):(3.0) t'y' with xyerror
```

Definisce il titolo del grafico

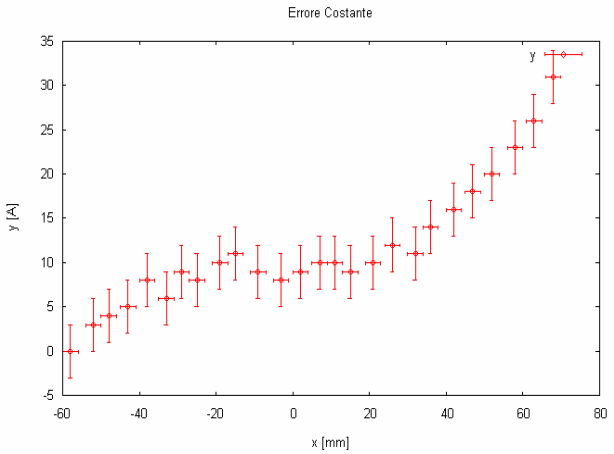
Definisce l'etichetta delle x

Definisce l'etichetta delle y

File dati

Usa i dati in colonna 1 come x, i dati in colonna 2 come y, l'errore su x è 2.0 , l'errore su y

Usa lo stile xyerror: riporta le barre di errore sia su x che su y



I comandi possono essere dati in modo abbreviato:

pl 'dati_xy_a.dat' **u** 1:2:(2):(3) t'y' **w xyer**

produce lo stesso grafico di cui sopra. I comandi possono essere scritti in un file ASCII (es. usando il blocnotes di Windows) e richiamati con il comando

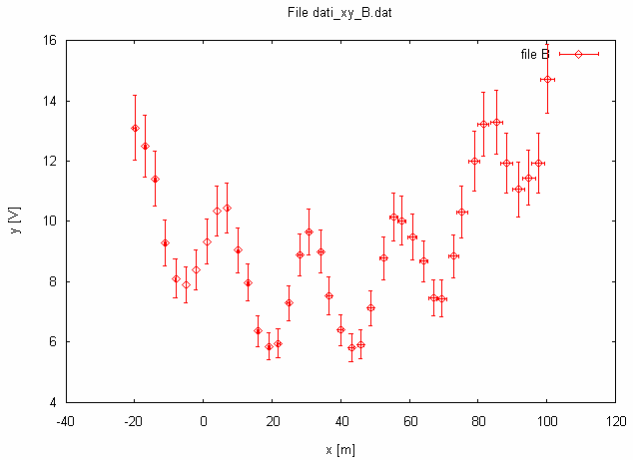
load 'file'

ad esempio provare a eseguire lo script del file **plotA.plt.**

Per graficare i dati del file dati_xy_B.dat utilizzare il comando:

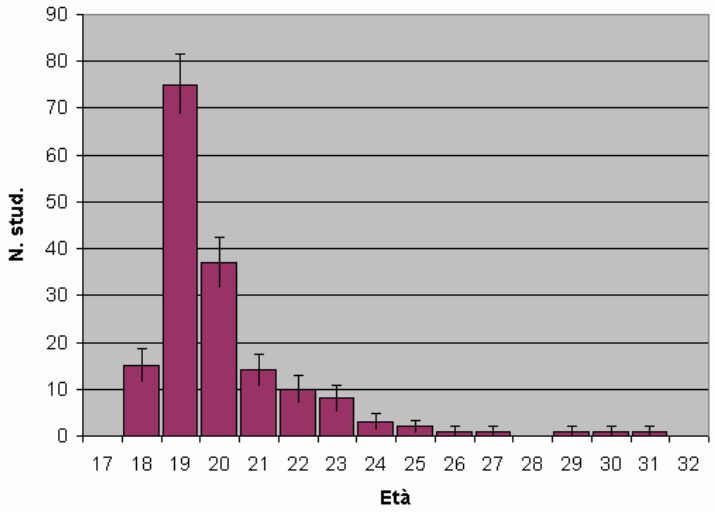
pl 'dati_xy_b.dat' **u** 1:3:2:4 t 'file B' **w xyerr**

Usa i dati in colonna 1 come x, i dati in colonna 3 come y, i dati in colonna 2 come errore su x e i dati in colonna 4 come errore su



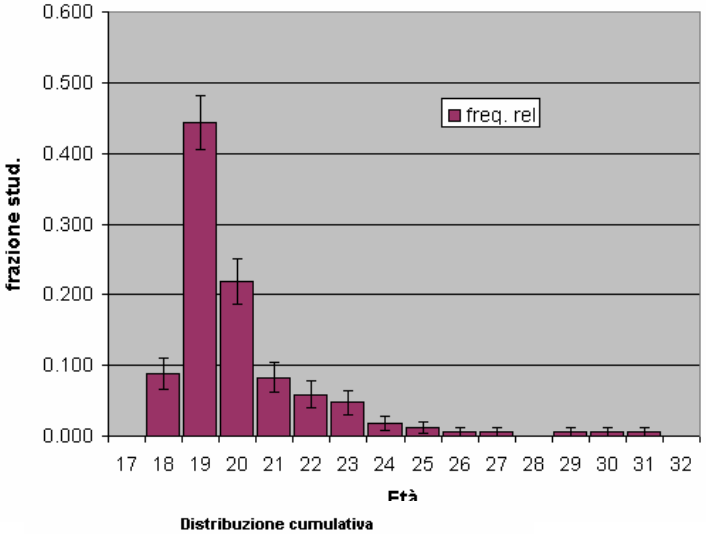
Provate a utilizzare (ed eventualmente modificare) lo script nel file **plotB.plt.**

b) Il file **Error3.dat** contiene l'età di un gruppo di studenti che frequentano le lezioni del II anno di un dato corso di laurea: la prima colonna riporta i dati grezzi, nella seconda e terza colonna sono i dati che riportano il numero di studenti per fascia d'età. Importare i dati in un file Excel e graficare l'istogramma. Calcolare gli errori sulle frequenze osservate e riportarli sul grafico.
Per il calcolo degli errori usiamo il fatto che il numero di osservazioni in ciascuna classe segue una distribuzione Binomiale (attenzione la distribuzione degli studenti nelle varie classi d'età non segue una distribuzione Binomiale): si hanno N prove (numero totale di studenti), la probabilità di successo (osservazione di uno studente nella classe di età i-esima) stimata è $p \sim f_i = n_i/N$, dove n_i è il numero di studenti osservato nella classe i-esima. La varianza di una distribuzione Binomiale è $\sigma^2 = Np(1-p) = n_i(N-n_i)/N$, l'errore sul numero di studenti osservato è $\varepsilon=\sigma$.
D'altronde la distribuzione Binomiale tende ad una distribuzione di Poisson se $p<<1$ (quindi per $n_i << N$). Per una distribuzione di Poisson il valor medio e la varianza coincidono, in questo caso l'errore è $\varepsilon= (n_i)^{1/2}$



classi	frequenze	Err. Bin.	Err. Poiss.
17	0.0	0.0	0.0
18	15.0	3.7	3.9
19	75.0	6.5	8.7
20	37.0	5.4	6.1
21	14.0	3.6	3.7
22	10.0	3.1	3.2
23	8.0	2.8	2.8
24	3.0	1.7	1.7

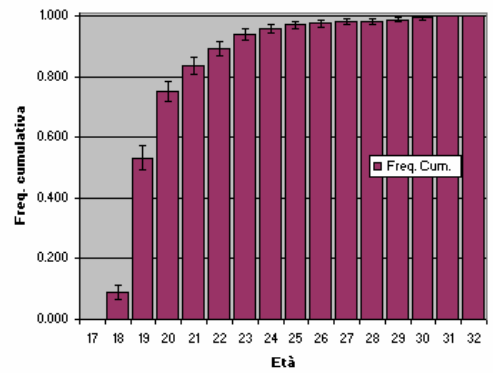
Si può notare che gli errori calcolati usando una distribuzione di Poisson o una distribuzione Binomiale coincidono quando n_i è piccolo ma sono molto diversi se n_i è grande. (notare che, in tabella, i dati e gli errori sono riportati accordando il numero di cifre significative).



Calcolare per ogni classe la distribuzione di frequenze relative $f_i = n_i/N$, gli errori sulle frequenze relative (questi sono: $\varepsilon_{fi}=(f_i(1-f_i)/N)^{1/2}$ … giustificare la formula), riportare l'errore in una colonna e ripetere il grafico utilizzando le frequenze relative con gli errori dati.

La distribuzione cumulativa F(x) rappresenta il numero di studenti con età minore di X. Calcolare la distribuzione cumulativa, l'errore associato per ogni F(x) e riportarla su un grafico.

Distribuzione cumulativa

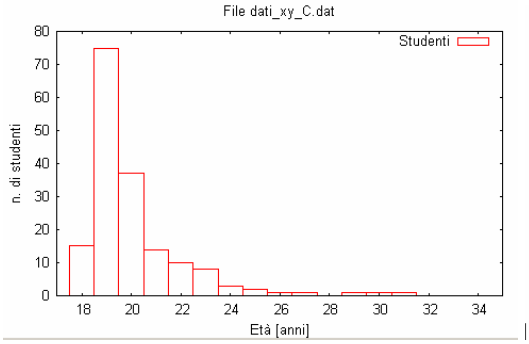


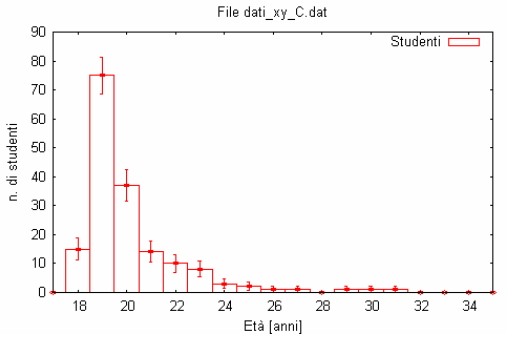
La funzione FREQUENZA(**Dati**,C) di Excel conta nella matrice **Dati** il numero di dati tali che $x_i < C$. Attenzione a fissare in modo opportuno la matrice di dati quando si copia la formula su più celle. (il file **Istogramma3.xls** contiene un esempio di soluzione)

Il file **plotC.plt** contiene uno script per graficare di dati usando Gnuplot.
Il comando:

pl 'dati_xy_C.dat' u 2:3 t 'Studenti' w boxes

permette di graficare un istogramma. Nel file **plot3.plt** ci sono una serie di script che mostrano come graficare istogrammi con errore.

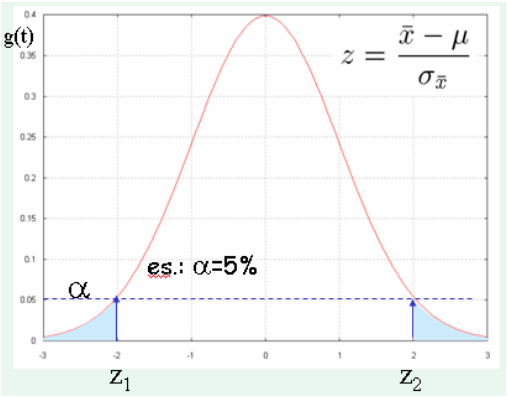




Per graficare la F(x) con Gnuplot salvare la tabella ottenuta con Excel in un file ASCII e utilizzare i comandi mostrati in precedenza.

Esercizio 3) Calcolo degli intervalli di confidenza. Prima una breve nota sul significato di "Intervallo di confidenza". Si è soliti rappresentare il valore di una grandezza misurata tramite il suo valore X_o (eventualmente il suo valore medio $\bar{X}$ se questa deriva da una serie di misure) e il suo errore σ. Si individua cosi un intervallo attorno al valore osservato: $[X_o- \sigma, X_o- \sigma]$ il cui significato è: siamo *confidenti* nel fatto che il valore osservato si trovi vicino al valore vero (μ) e quindi, ripetendo la misura, la probabilità che il nuovo risultato ri trovi al di fuori di questo intervallo è *piccola*. Queste affermazioni: *siamo confidenti*, *la probabilità è piccola*, sono qualitative. Vediamo come renderle quantitative. Consideriamo una grandezza X che segue una distribuzione normale attorno al suo valore medio μ con varianza σ^2. La variabile "standardizzata":

$$z = \frac{x - \mu}{\sigma}$$



Segue una distribuzione normale standard, cioè con media nulla e varianza unitaria. Se si indica con α la probabilità di osservare un valore x diverso da μ (maggiore o minore), questa rappresenta il "rischio" di sbagliare e si calcola:

$$\alpha = P(z < z_1) + P(z > z_2) = \int_{-\infty}^{z_1} g(z)dz + \int_{z_2}^{\infty} g(z)dz = 2\int_{z_L}^{\infty} g(z)dz$$

dove, nell'ultimo passaggio si è sfruttata la simmetria della funzione z rispetto all'origine. Fissiamo il rischio ad un valore accettabile, es. α=5% e calcoliamo quali sono i valori di z che delimitano la regione "lontana" dall'origine $[-z_L ; z_L]$ (si ha $-z_1=z_2 = z_L$ per la simmetria della funzione z). La probabilità di osservare z all'interno dell'intervallo è:

$$P(z_1 < z < z_2) = 1 - \alpha$$

Quindi abbiamo il 95% di probabilità di osservare un valore di x nell'intervallo:

$$[z_1 < \frac{x - \mu}{\sigma} < z_2] = [\mu - z_L\sigma < x < \mu + z_L\sigma]$$

Il valore X_o osservato rappresenta la migliore stima che ho, per ora, del valore medio. Quindi l'intervallo:

$$[X_o - z_L\sigma; X_o + z_L\sigma]$$

attorno al valore osservato rappresenta l'intervallo entro cui è lecito aspettarsi (con probabilità 1-α, 95% nel nostro caso) il valore corretto. Questo rappresenta l'*intervallo di confidenza*".
La funzione Excel CONFIDENZA consente di calcolare l'intervallo di confidenza per una variabile che segue una distribuzione normale. La sintassi è:

CONFIDENZA(α,σ,N)

dove α indica il rischio di aver sbagliato (1-α indica la confidenza), σ è la dev. standard campionaria e N è la numerosità del campione. Se ho un solo valore di X_o, che mi aspetto seguire una distribuzione gaussiana, con il suo errore σ, si ha N=1.
Se ho ottenuto $X_o = \bar{X}$ come media di N valori, σ è la dev. st. campionaria. Excel utilizza l'errore standard della media:

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{N}}$$

Per calcolare l'intervallo di confidenza.

Se utilizzo $\sigma_{\bar{X}}$ come errore si deve porre N=1.

Se il valore osservato e la sua varianza sono stati determinati utilizzando un procedimento che combina diversi effetti bisogna tener conto dei gradi di liberà effettivi del procedimento e non si può utilizzare la funzione z per il calcolo dell'intervallo di confidenza. Ad esempio nel caso in cui risulta dalla somma di due grandezze A e B misurate con errori σ^2_a e σ^2_b, si ha X = A + B e $\sigma_x = (\sigma^2_a + \sigma^2_b)^{1/2}$.

In questo caso la distribuzione dei valori osservati di X attorno al valore medio seguono una distribuzione nota come t-Student:

$$t_\nu = \frac{x - \mu}{\sigma}$$

questa con ν = M-1, dove M è il numero di osservazioni che concorrono alla determinazione di X. Nel caso in esempio M = 2 e ν = 1. La distribuzione t approssima una distribuzione Gaussiana per ν grande.

t-Student	
ν	t_L
2	4.30
5	2.57
10	2.23
50	2.01
100	1.98
500	1.96
1000	1.96
	z_L
CONFIDENZA(0.05,1,1)	1.96

Per utilizzare la funzione t-Student si utilizza la funzione Excel

INV.T(α,ν)

Per calcolare i valori limite t_L dell'intervallo di confidenza**.** Da questi si determina l'intervallo di confidenza attorno al valore osservato:

$$[X_o - t_L\sigma; X_o + t_L\sigma]$$

Si può verificare che se il numero di parametri liberi è piccolo (ν<10) l'intervallo di confidenza ottenuto presupponendo una distribuzione Gaussiana sottostima l'intervallo di confidenza.

Esercizio:
i) Calcolare gli intervalli di confidenza al 90%, 95%, 99% per i seguenti dati:

X	δX	X	δX
54.1272	0.0023	$8.802 \cdot 10^{-5}$	$1.6 \cdot 10^{-7}$
6.96	0.03	13.05	$4.\ 3 \cdot 10^{-2}$
23.5	0.05	$9.285 \cdot 10^{-2}$	4E-03
156.9	1.5	696	15
656.3	0.2	0.00084	0.00012

ii) Utilizzare i dati della tabella nel file: **Cifre_sgnificative.xls** per calcolare gli intervalli di confidenza al 90% e 95% sui valori medi

iii) propagazione degli errori, sigma e intervallo di confidenza. Nel file Excel **Confidenza.xls** (Foglio ***Dati***) sono registrati una serie di misure dei parametri **a**, **b**, **c** e dell'ascissa **x** insieme con gli errori.

a (N)			x (m)			b (Nm)			c (Nm^2)		
0.052	±	0.001	-1.00	±	0.05	1.10	±	0.02	0.40	±	0.02
0.093	±	0.005	1.30	±	0.03	1.20	±	0.01	0.20	±	0.01
0.130	±	0.010	2.00	±	0.05	0.90	±	0.04	0.10	±	0.04
0.230	±	0.020	-3.24	±	0.02	2.00	±	0.02	1.00	±	0.05
0.645	±	0.034	3.89	±	0.04	2.40	±	0.02	1.30	±	0.03
1.070	±	0.040	2.60	±	0.03	2.00	±	0.01	1.10	±	0.04
3.040	±	0.020	-2.86	±	0.06	1.20	±	0.05	1.60	±	0.01
5.150	±	0.010	4.00	±	0.30	2.00	±	0.07	1.30	±	0.03

Da questi calcolare i valori di $y_2 = ax^2 + bx + c$ con l'errore associato e gli intervalli di confidenza al 67%, 95% e 99%.

Si ricordi che per una funzione $f(p_1,p_2, \ldots p_n)$ di n parametri, ognuno caratterizzato da un errore Δp_i l'errore si calcola utilizzando la formula di propagazione degli errori:

$$\Delta y = \sqrt{\left(\frac{\partial f}{\partial p_1}\Delta p_1\right)^2 + \left(\frac{\partial f}{\partial p_2}\Delta p_2\right)^2 + \ldots \left(\frac{\partial f}{\partial p_n}\Delta p_n\right)^2}$$

calcolare i valori delle derivate parziali e l'errore su y

Derivate parziali				Risultato	
|dy/da|	|dy/dx|	|dy/db|	|dy/dc|	y	s_y
1.00	0.996	1	1	-0.65	0.06
1.69	1.4418	1.3	1	1.92	0.05
4.00	1.42	2	1	2.42	0.12
10.50	0.5096	3.24	1	-3.07	0.23
15.13	7.4181	3.89	1	20.40	0.60
6.76	7.564	2.6	1	13.5	0.4
8.18	16.1888	2.86	1	23.0	1.0
16.00	43.2	4	1	92	13

Calcolare i valori limite par la variabile t-Student con 3 gradi di libertà per una confidenza del 67%, 90% e 95%

t -lim.			
t-lim	1.16	2.35	5.84
1-α	0.67	0.9	0.99
α	0.33	0.1	0.01
ν	3		

Calcolare gli intervalli di confidenza per la y per le diverse misure e riportare su grafici diversi i valori y(x) mostrando l'errore standard o gli intervalli di confidenza al 95%

Intervalli di confidenza					
conf 33 %		conf 67 %		conf. 99%	
-0.71	-0.58	-0.78	-0.51	-0.98	-0.31
1.86	1.97	1.81	2.03	1.64	2.19
2.28	2.56	2.14	2.70	1.71	3.13
-3.33	-2.80	-3.60	-2.53	-4.38	-1.75
19.70	21.09	18.98	21.81	16.89	23.90
13.1	13.9	12.7	14.4	11.5	15.6
21.9	24.2	20.7	25.4	17.2	28.8
77	107	61	122	16	167