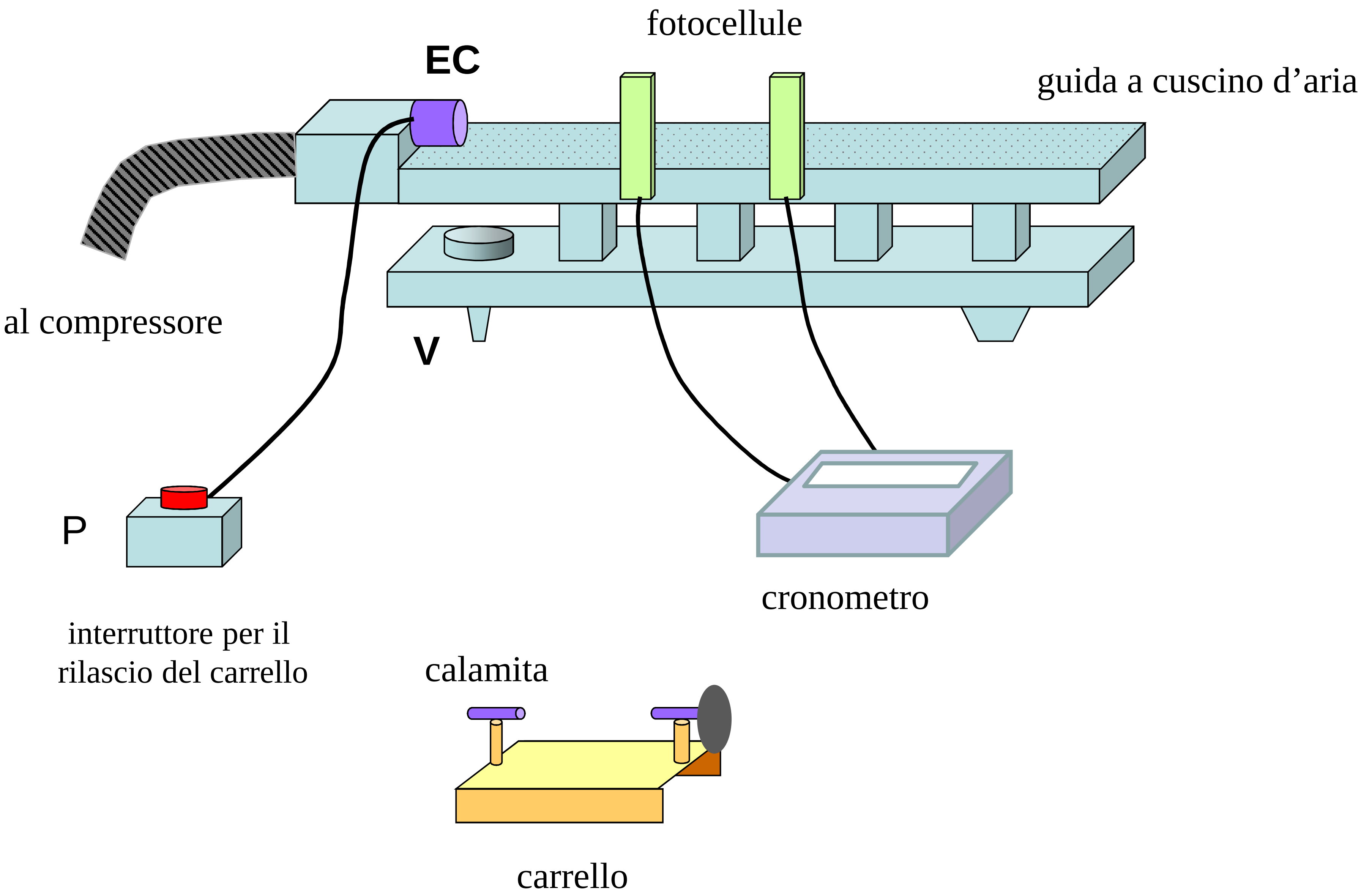
Verifica della distribuzione gaussiana degli errori



Operazioni preliminari e misura

1. accendere l'interruttore principale del banco di misura
2. accendere il cronometro e selezionare la massima risoluzione (10^4)
3. si accende il compressore, regolandone la portata (tipicamente posiz. 2,5 o 3) (non modificare la portata del compressore durante la misura)
4. si definisce la posizione orizzontale della guida agendo sulla vite V. Una volta trovata, si blocca la ghiera alla base della vite micrometrica come riferimento.
5. portare la slitta all'inizio della guida, in modo che aderisca all'elettrocalamita, e lanciarla premendo (o rilasciando) il pulsante P
6. verificare il corretto funzionamento del cronometro e la riproducibilità della misura
7. iniziare la serie di «almeno» 200 misure

Analisi dei dati. 1

Si riportano i dati in una colonna su un foglio EXCEL. Si calcola la media e la dev. standard

	A	B	C	D	E
1					
2					
3					
4		**k**	**Tk**	**Zk**	
5		1	1,2565	0,0035	
6		2	1,2606	0,0076	
7		3	1,2524	-0,0006	
8		4	1,2449	-0,0081	
9		5	1,2453	-0,0077	
10		6	1,2502	-0,0028	
11		7	1,2503	-0,0027	
12		8	1,2562	0,0032	
13		9	1,2566	0,0036	
14		10	1,2605	0,0075	
15		11	1,2552	0,0022	

= C5-C$206

deviazione dalla media (scarto)

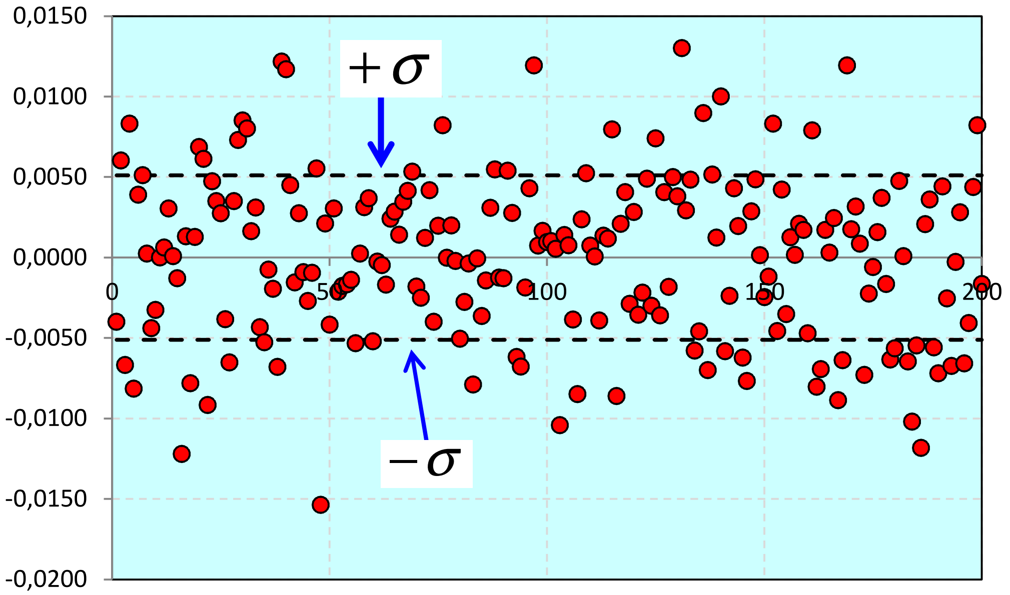
ecc. ecc. ecc.

202		198	1,2515	-0,0013	
203		199	1,2520	-0,0008	
204		200	1,2479	-0,0049	
205					
206		**<t>**	1,2528	0,0000	
207		**σ**	0,0048	0,0048	

= MEDIA(C5:C204)

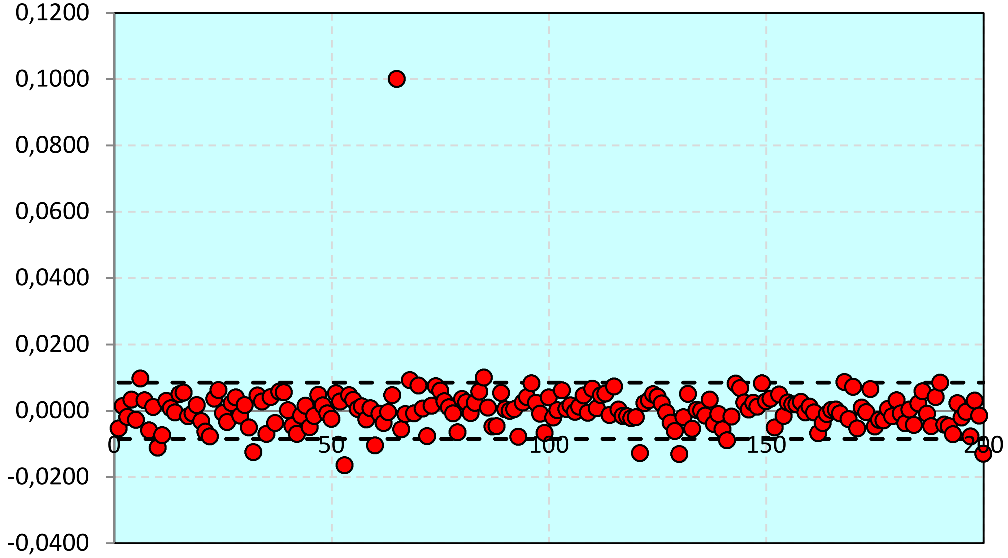
= DEV.ST.C(C5:C204)

un primo controllo: «ideogramma» degli scarti

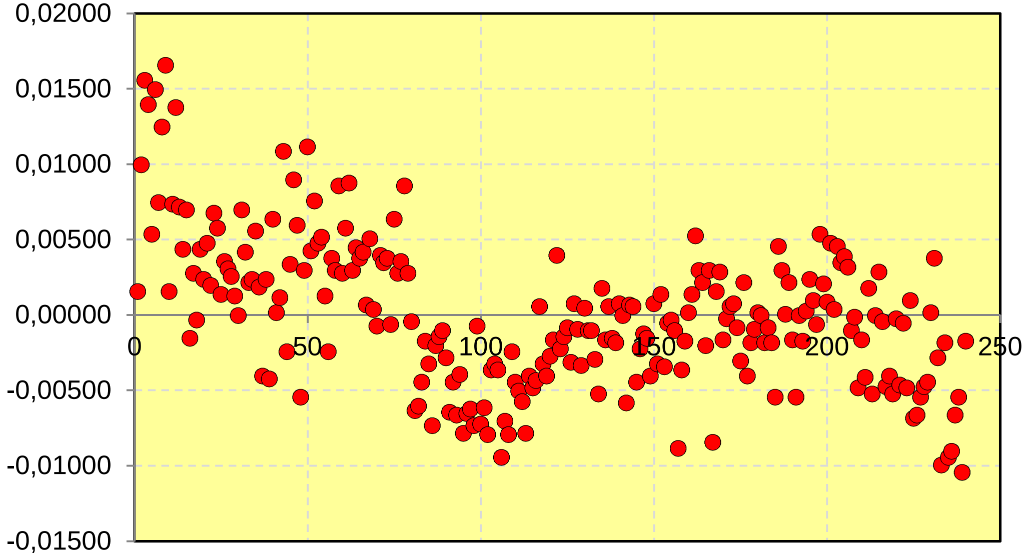


Analisi dei dati. 1

L'ideogramma può evidenziare situazioni problematiche



Uno o più punti <u>nettamente</u> «fuori dalla media».
Sono quasi certamente errori di trascrizione. Meglio eliminarli e ricalcolare media e dev. standard



Il valore medio cambia nel tempo.
E' un problema della strumentazione
Non c'è niente da fare.

Analisi dei dati. 2

Bisogna costruire l'istogramma degli scarti. Per far questo

1. si sceglie un passo opportuno Δz
2. e tanti intervalli di larghezza Δz, ovvero tanti punti z_1, z_2 ... con passo Δz
3. scelto ad es. l'intervallo $[z_1, z_2]$ si conta il numero di scarti compresi fra z_1 e z_2, lo chiamiamo N_1
4. si ripete per tutti gli altri intervalli; troveremo i numeri N_2, N_3, ...
5. si rappresentano questi valori tramite un «istogramma»

un valore suggerito di Δz è metà della deviazione standard , o non molto diverso

Il conteggio dei valori compresi nei diversi intervalli potrebbe essere assai noioso, ma per fortuna c'è una funzione EXCEL che lo fa per noi

Analisi dei dati. 2

	....	F	G	H	I
1					
2			σ	0,0048	
3			Δz	0,0024	
4			A	0,48	
5					
6					0
7			-10	-0,0240	0
8			-9	-0,0216	0
9			-8	-0,0192	0
10			-7	-0,0168	0
11			-6	-0,0144	1
12			-5	-0,0120	5
13			-4	-0,0096	7
14			-3	-0,0072	16
15			-2	-0,0048	32
16			-1	-0,0024	33
17			0	0,0000	44
18			1	0,0024	31
19			2	0,0048	17
20			3	0,0072	12
21			4	0,0096	1
22			5	0,0120	1
23			6	0,0144	0
24			7	0,0168	0
25			8	0,0192	0
26			9	0,0216	0
27			10	0,0240	0
28					

= DEV.ST.C(C5:C204)

= H2/2

= G7*H$3

= FREQUENZA(D5:D204;H7:H27)

si può costruire una colonna di valori z_k.

Fra -5σ e +5σ si troveranno praticamente tutti i punti.

stabiliti gli intervalli, si usa la funzione FREQUENZA di EXCEL per contare gli eventi compresi. Gli argomenti sono:

- colonna degli scarti da analizzare: D5:D204
- colonna degli intervalli: H7:H27

- definire la funz. FREQUENZA una casella prima di -10*Δz
- selezionare (non trascinare) le caselle fino a 10*Δz.
- premere F2
- premere insieme CTRL - SHIFT – ENTER

Analisi dei dati. 2

= 200*H3

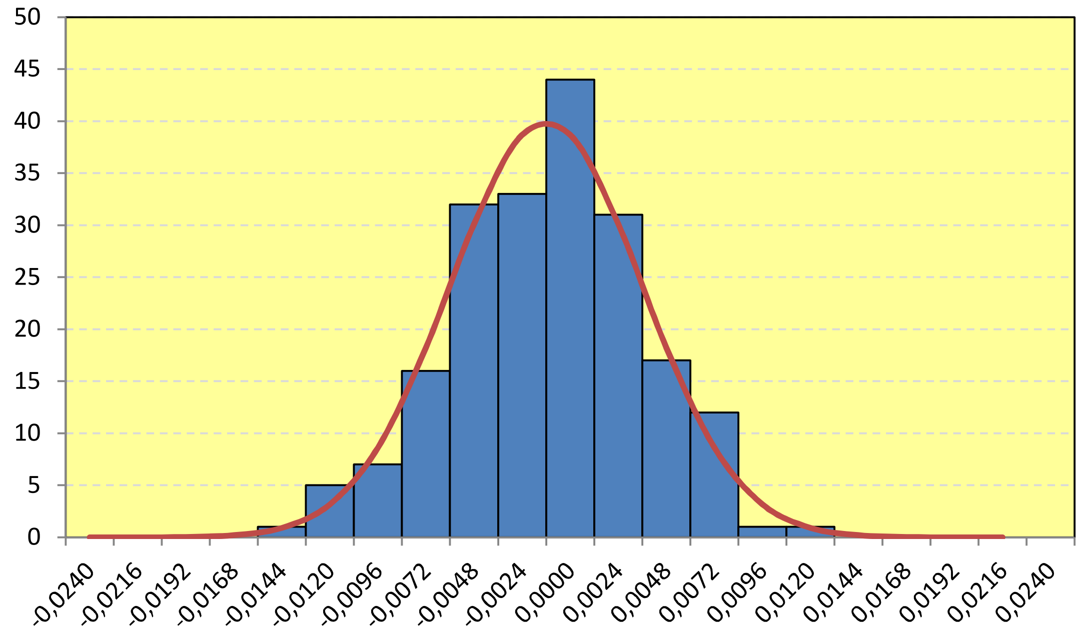
.....

	F	G	H	I	J	K
1						
2		**σ**	0,0048			
3		**Δz**	0,0024			
4		**A**	0,48			
5				freq.	z med	gauss
6				0		
7		-10	-0,0240	0	-0,0228	0,00
8		-9	-0,0216	0	-0,0204	0,00
9		-8	-0,0192	0	-0,0180	0,04
10		-7	-0,0168	0	-0,0156	0,20
11		-6	-0,0144	1	-0,0132	0,91
12		-5	-0,0120	5	-0,0108	3,17
13		-4	-0,0096	7	-0,0084	8,63
14		-3	**-0,0072**	**16**	**-0,0060**	**18,26**
15		-2	**-0,0048**	32	-0,0036	30,11
16		-1	-0,0024	33	-0,0012	38,67
17		0	0,0000	44	0,0012	38,67
18		1	0,0024	31	0,0036	30,11
19		2	0,0048	17	0,0060	18,26
20		3	0,0072	12	0,0084	8,63
21		4	0,0096	1	0,0108	3,17
22		5	0,0120	1	0,0132	0,91
23		6	0,0144	0	0,0156	0,20
24		7	0,0168	0	0,0180	0,04
25		8	0,0192	0	0,0204	0,00
26		9	0,0216	0	0,0228	0,00
27		10	0,0240	0	0,0120	1,75
28						

= (H26+H27)/2

= H$4*DISTRIB.NORM.N(J26;0;H$3)

con i valori di frequenza si traccia l'istogramma degli scarti



la frequenza in rosso (16) è il numero di scarti compresi fra i due estermi in rosso; la loro media è il valore blu della colonna «z med». In questo punto si calcolerà la gaussiana

gaussiana con lo stesso centro (0), stessa larghezza (**σ**) e la stessa area. Da calcolare «al centro degli intervalli»

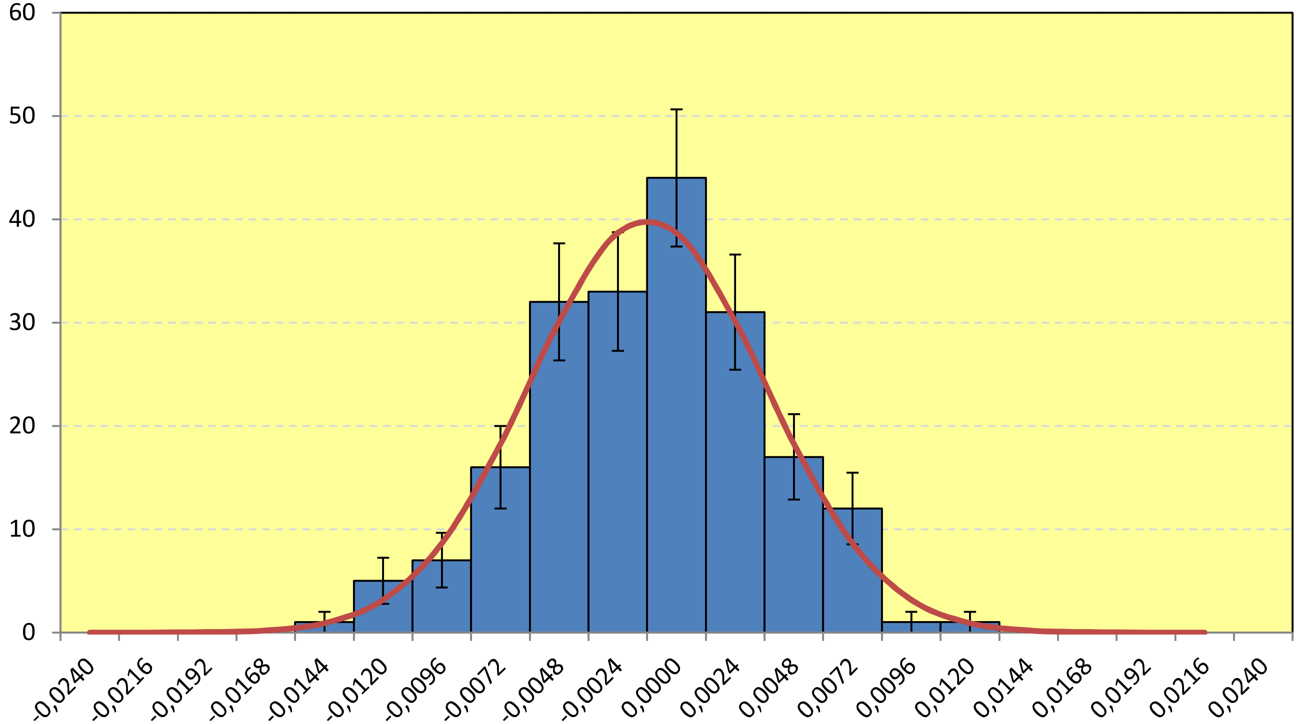
Analisi dei dati. 3

La distribuzione misurata è veramente gaussiana?

Questa domanda rientra nel vasto campo delle «decisioni statistiche», ovvero della valutazione di un'ipotesi in base a risultati che sono per forza limitati e soggetti ad incertezze.

In genere non è possibile dire che l'ipotesi è vera o falsa, ma solo la probabilità delle due alternative. Cerchiamo un criterio quantitativo.

- quanto vale l'errore dei singoli conteggi?
- come valutare la bontà dell'accordo?



Analisi dei dati. 3

bisogna confrontare le differenze istogramma-gaussiana intervallo per intervallo

in pratica si calcola il «Chi-quadro»

$$\chi^2 = \sum_k \frac{[N_k - g(z_k)]^2}{\sigma_k^2}$$

dove

N_k	è il numero di conteggi nell’intervallo k-mo
z_k	è il valore di z al centro dell’intervallo
$g(z_k)$	è la funzione gaussiana valutata in z_k
σ_k	è l’incertezza su N_k

La somma si calcolerà sui canali in cui N_k è «abbastanza grande», in pratica una decina di canali intorno al valore centrale.

per l’errore si assumerà $\sigma_k = \sqrt{N_k}$ dev. standard per distribuzione di Poisson

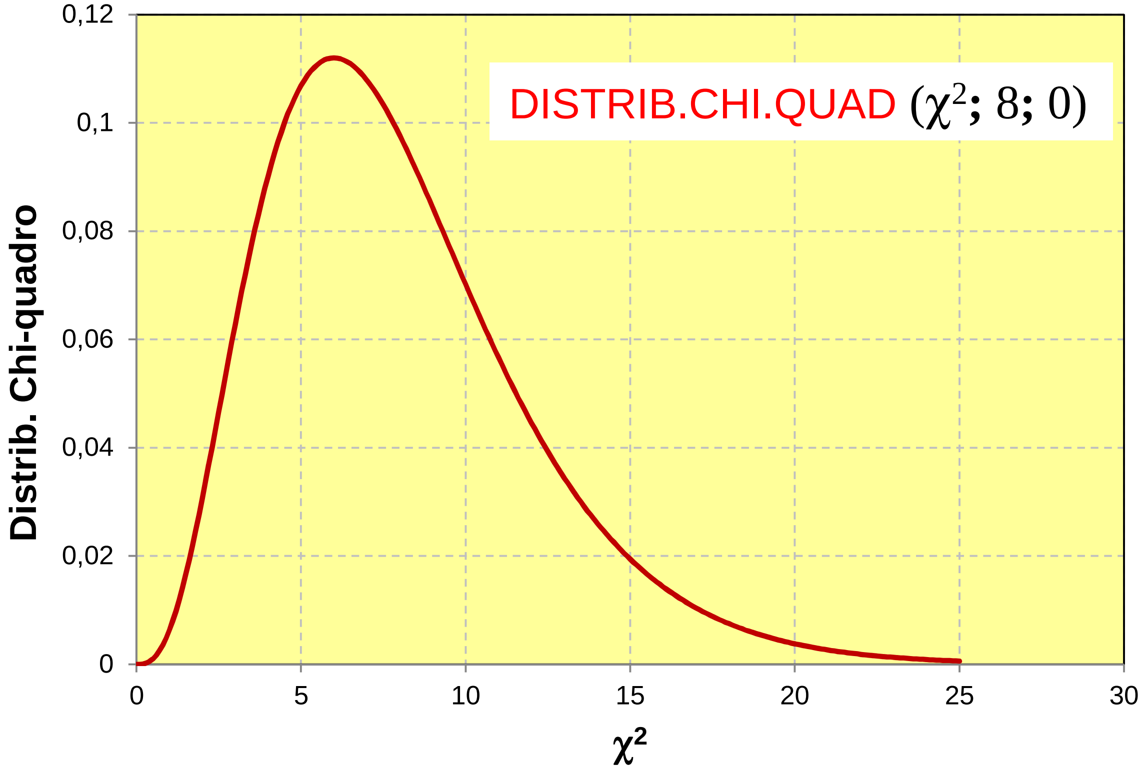
$$\chi^2 = \sum_k \frac{[N_k - g(z_k)]^2}{N_k}$$

Analisi dei dati. 3

Metodo: valutare la probabilità di ottenere la nostra distribuzione da N misure, nell'ipotesi che queste abbiano veramente una distribuzione gaussiana (ipotesi da verificare)

- se questa è particolarmente piccola, possiamo scartare l'ipotesi
- se è abbastanza grande diremo che i dati sono compatibili con l'ipotesi

per far ciò bisogna conoscere la distribuzione del Chi-quadro, usando le tavole o meglio la funzione DISTRIB.CHI.QUAD(...) di EXCEL.



distribuzione del χ^2
con 8 gradi libertà

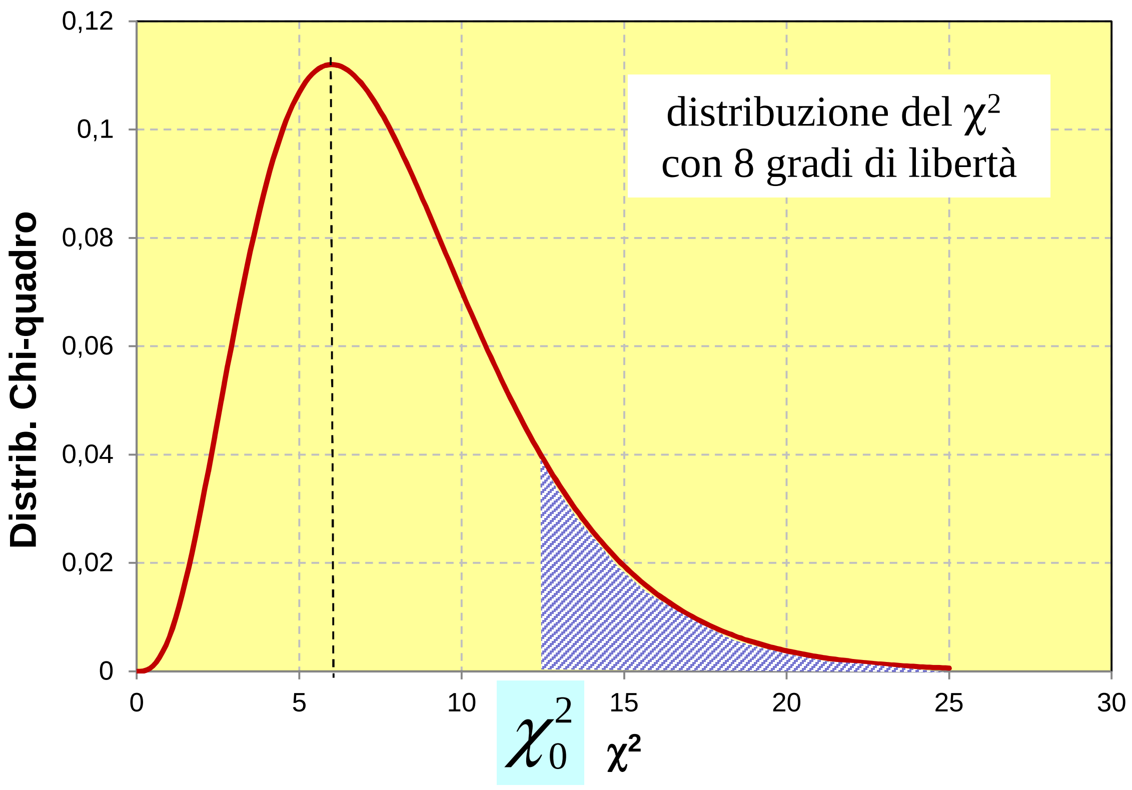
Analisi dei dati. 3

«gradi di libertà» = N° di punti (dell'istogramma) - N° parametri utilizzati

3 nel nostro caso

se i punti sono 11 i GdL sono 8. Il valore di χ^2 più probabile è 5,7.

supponiamo di aver trovato il valore χ_0^2



si calcola l'area della distribuzione fra χ_0^2 e infinito (probabilità che il χ^2 abbia valore superiore a χ_0^2)

se è «abbastanza grande» diremo che l'ipotesi è «confermata»

se è «troppo piccola» diremo che l'ipotesi è «smentita»

Analisi dei dati. 3

L'area fra χ_0^2 e infinito si calcola con la funzione DISTRIB.CHI.QUAD(χ_0^2**;** GdL**;** 1)

il valore discriminante è arbitrario. In molti casi si utilziza il 5% (0.05) oppure l' 1% (0.01).
Non bisogna però dimenticare il significato puramente statistico.

Per esempio, se 100 studenti misurano una distribuzione autenticamente gaussiana, ci aspettiamo che uno di loro (in media) trovi un χ^2 «troppo grande» a livello dell' 1%.

Con 8 Gradi di Libertà come nell'esempio :

una probabilità del 5% corrisponde a $\chi^2 = 15{,}5$
una probabilità dell' 1% corrisponde a $\chi^2 = 20$
una probabilità di 0,1% corrisponde a $\chi^2 = 26$

Analisi dei dati. 3

	F	G	H	I	J	K	L	M	N
1									
2		**σ**	0,0048						
3		**Δz**	0,0024						
4		**A**	0,48						
5									
6				0					
7		-10	-0,0240	0	-0,0228	0,00			
8		-9	-0,0216	0	-0,0204	0,00			
9		-8	-0,0192	0	-0,0180	0,04			
10		-7	-0,0168	0	-0,0156	0,20			
11		-6	-0,0144	1	-0,0132	0,91			
12		-5	-0,0120	5	-0,0108	3,17	-0,3652		
13		-4	-0,0096	7	-0,0084	8,63	0,2325		
14		-3	-0,0072	16	-0,0060	18,26	0,1416		
15		-2	-0,0048	32	-0,0036	30,11	-0,0589		
16		-1	-0,0024	33	-0,0012	38,67	0,1717		
17		0	0,0000	44	0,0012	38,67	-0,1212		
18		1	0,0024	31	0,0036	30,11	-0,0286		
19		2	0,0048	17	0,0060	18,26	0,0744		
20		3	0,0072	12	0,0084	8,63	-0,2810		
21		4	0,0096	1	0,0108	3,17	2,1740		
22		5	0,0120	1	0,0132	0,91	-0,0906		
23		6	0,0144	0	0,0156	0,20			
24		7	0,0168	0	0,0180	0,04	**χ2**	**8,88**	
25		8	0,0192	0	0,0204	0,00			
26		9	0,0216	0	0,0228	0,00	**prob.$**	**26,1**	
27		10	0,0240	0	0,0120	1,75			
28									

.....

selezioniamo un insieme di valori diversi da zero (azzurro) e, utilizzando i corrispondenti valori della gaussiana (arancio) si calcola il chi-quadro.

= (I12 - K12)^2 / I12

si calcola la probabilità che il χ^2 abbia valore superiore a quello trovato.

= SOMMA (L12:L22)

= 1- DISTRIB.CHI.QUAD (M26; 8; 1)

i dati sono compatibili con una distribuzione normale o gaussiana

Analisi dei dati. 3

per chi avesse problemi con la funzione di distribuzione del chi-quadro, ecco una tabella i valori di interesse

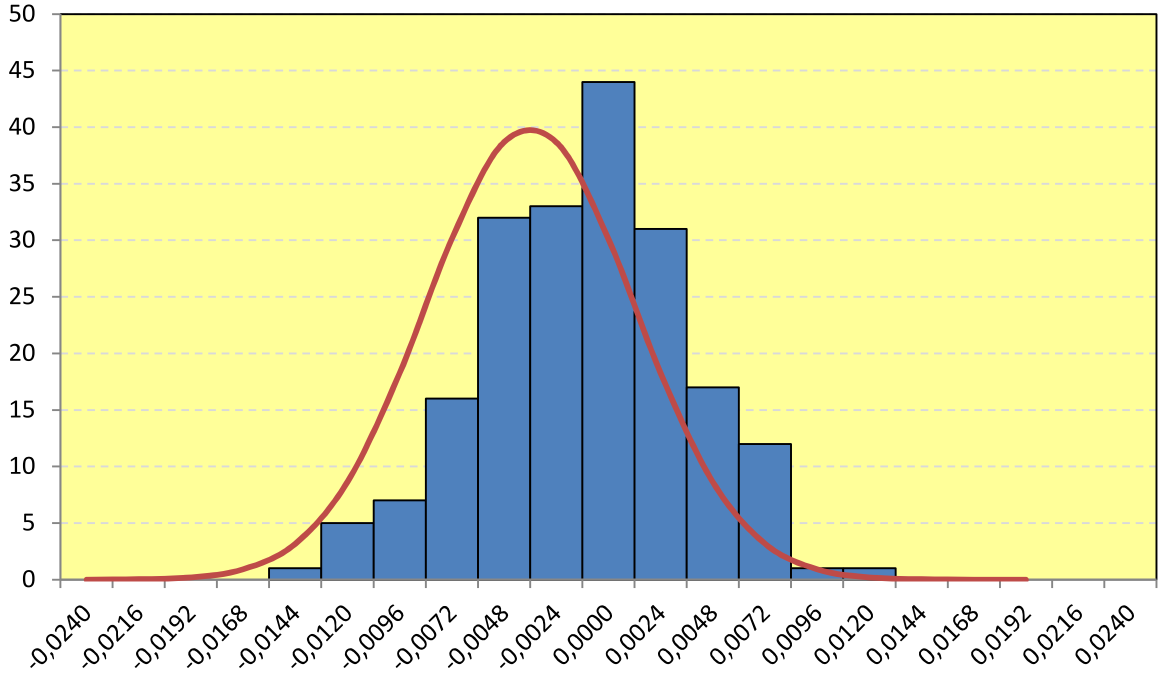
probabilità che il χ^2 superi il valore in tabella

gradi di libertà

	10%	5%	2%	1%	0,5%	0,2%	0,1%
n=**1**	2,71	3,84	5,41	6,63	7,88	9,55	10,83
2	4,61	5,99	7,82	9,21	10,60	12,43	13,82
3	6,25	7,82	9,84	11,34	12,84	14,80	16,27
4	7,78	9,49	11,67	13,28	14,86	16,92	18,47
5	9,24	11,07	13,39	15,09	16,75	18,91	20,52
6	10,64	12,59	15,03	16,81	18,55	20,79	22,46
7	12,02	14,07	16,62	18,47	20,28	22,60	24,32
8	13,36	15,51	18,17	20,09	21,96	24,35	26,12
9	14,08	16,92	19,68	21,67	23,59	26,06	27,88
10	15,99	18,31	21,16	23,21	25,19	27,72	29,59
11	17,27	19,68	22,62	24,72	26,76	29,35	31,27
12	18,55	21,03	24,05	26,22	28,30	30,96	32,91
13	19,81	22,36	25,47	27,69	29,82	32,54	32,53
14	21,06	23,68	26,87	29,14	31,32	34,09	36,12
15	22,31	25,00	28,26	30,58	32,80	35,63	37,70

Analisi dei dati. 2 (+ 3)

L'**errore tipico** nel confronto con la distribuzione normale è lo «sfasamento» della curva rispetto ai dati, come nell'esempio qui sotto. In generale questo sfasamento si vede chiaramente «ad occhio»



quali le cause? Possono essercene molte, ma due sono ricorrenti:

1. nell'esempio, si sono graficati i punti I**7**:I27 per l'istogramma, K**8**:K26 per la gaussiana (il χ^2 però è corretto)
2. un altro motivo è non aver calcolato la gaussiana nel punto giusto (e allora è sbagliato anche il χ^2)