

Corso Integrato di Statistica Informatica e Analisi dei dati

Dr Carlo Meneghini

Dip. di Fisica “E. Amaldi”
via della Vasca Navale 84

meneghini@fis.uniroma3.it

<http://webusers.fis.uniroma3.it/~meneghini>

Sintesi Statistica

Riepilogo

Tabelle di frequenza

Istogrammi

Sintesi numerica

Funzioni avanzate di Excel

Fenomeno	Caratteri	Dati
Maturazione	peso	50 (g)
	colore	giallo
	sapore	aspro

Una variabile è una funzione che associa un valore ad una caratteristica del fenomeno studiato.

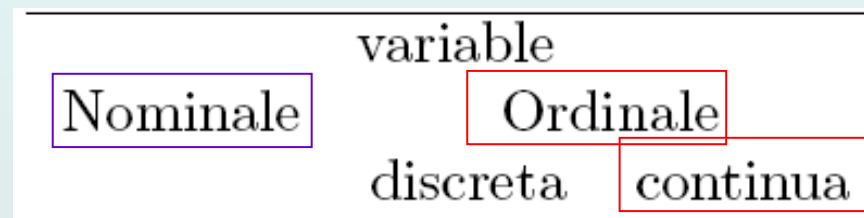


Tabelle di frequenza...

Frequenze assolute o relative?

Integrate o no?

Numero di studenti

#	X	#	X	#	X
1	96	11	92	21	91
2	90	12	98	22	73
3	101	13	93	23	90
4	85	14	95	24	87
5	65	15	84	25	96
6	101	16	69	26	82
7	102	17	87	27	79
8	95	18	98	28	80
9	98	19	101	29	89
10	99	20	103	30	86

X = Peso frutti

#	X[g]	#	X[g]	#	X[g]
1	14.1	11	12.9	21	14.0
2	11.2	12	9.9	22	13.9
3	15.7	13	9.5	23	19.0
4	18.8	14	12.1	24	14.7
5	18.6	15	12.7	25	14.4
6	20.2	16	8.6	26	13.5
7	8.4	17	13.3	27	20.9
8	14.3	18	13.8	28	17.6
9	18.3	19	15.4	29	22.1
10	11.7	20	13.9	30	13.0

Costruzione di una tabella di frequenza

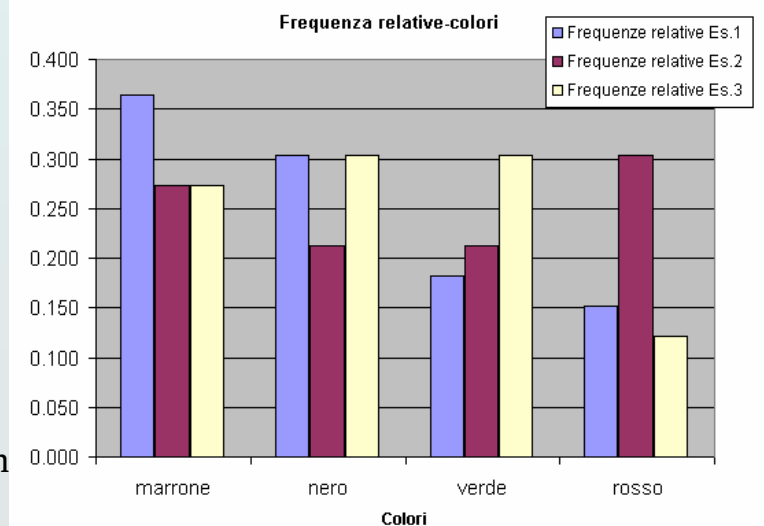
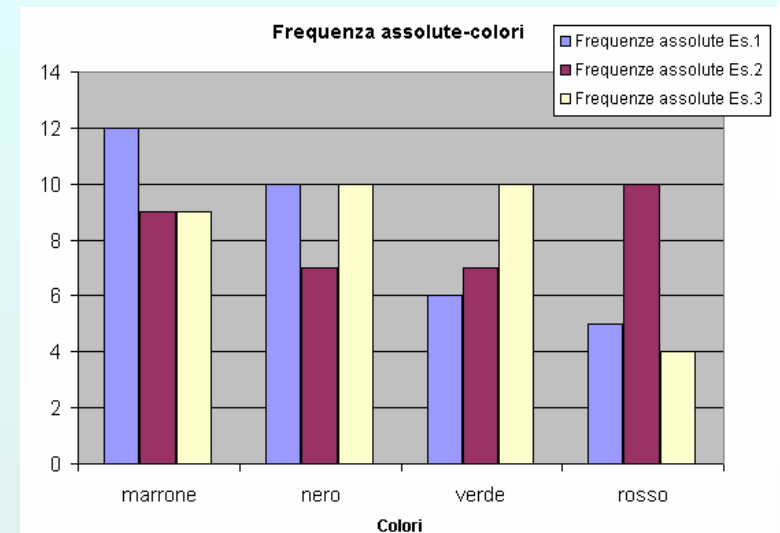
Dati nominali

facile: una classe per ogni dato

Funzione excel: **CONTA.SE**

Dati nominali		
es.1	Es.2	Es.3
marrone	marrone	marrone
marrone	marrone	marrone
nero	rosso	nero
marrone	nero	rosso
rosso	nero	marrone
nero	rosso	nero
verde	nero	rosso
verde	verde	nero
rosso	rosso	nero
rosso	verde	verde
nero	nero	marrone
marrone	marrone	marrone
rosso	rosso	rosso

	Es.1	Es.2	Es.3
N.tot	33	33	33
Frequenze assolute			
classi	Es.1	Es.2	Es.3
marrone	12	9	9
nero	10	7	10
verde	6	7	10
rosso	5	10	4
controllo	33	33	33
Frequenze relative			
classi	Es.1	Es.2	Es.3
marrone	0.364	0.273	0.273
nero	0.303	0.212	0.303
verde	0.182	0.212	0.303
rosso	0.152	0.303	0.121
controllo	1	1	1



C.I. - statistica, informatica e analisi
dei dati sperimentali

Costruzione di una tabella di frequenza

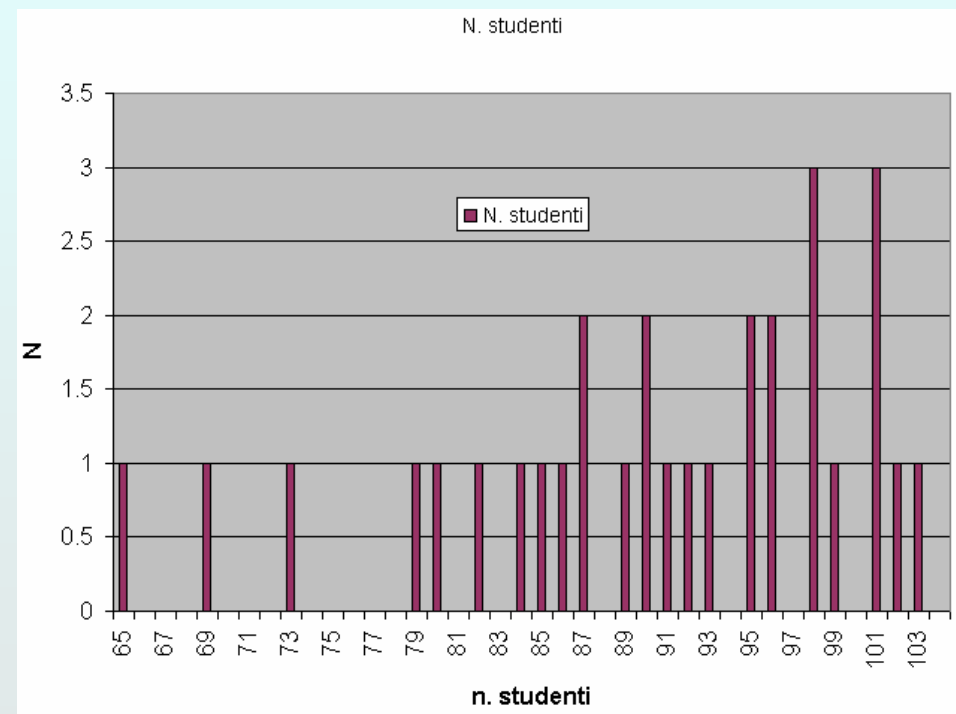
Dati ordinali

dati

Numero di studenti					
#	X	#	X	#	X
1	96	11	92	21	91
2	90	12	98	22	73
3	101	13	93	23	90
4	85	14	95	24	87
5	65	15	84	25	96
6	101	16	69	26	82
7	102	17	87	27	79
8	95	18	98	28	80
9	98	19	101	29	89
10	99	20	103	30	86

N

X = N. studenti presenti					
X	N	X	N	X	N
65	1	75	0	85	1
66	0	76	0	86	1
67	0	77	0	87	2
68	0	78	0	88	0
69	1	79	1	89	1
70	0	80	1	90	2
71	0	81	0	91	1
72	0	82	1	92	1
73	1	83	0	93	1
74	0	84	1	94	0



Costruzione di una tabella di frequenza

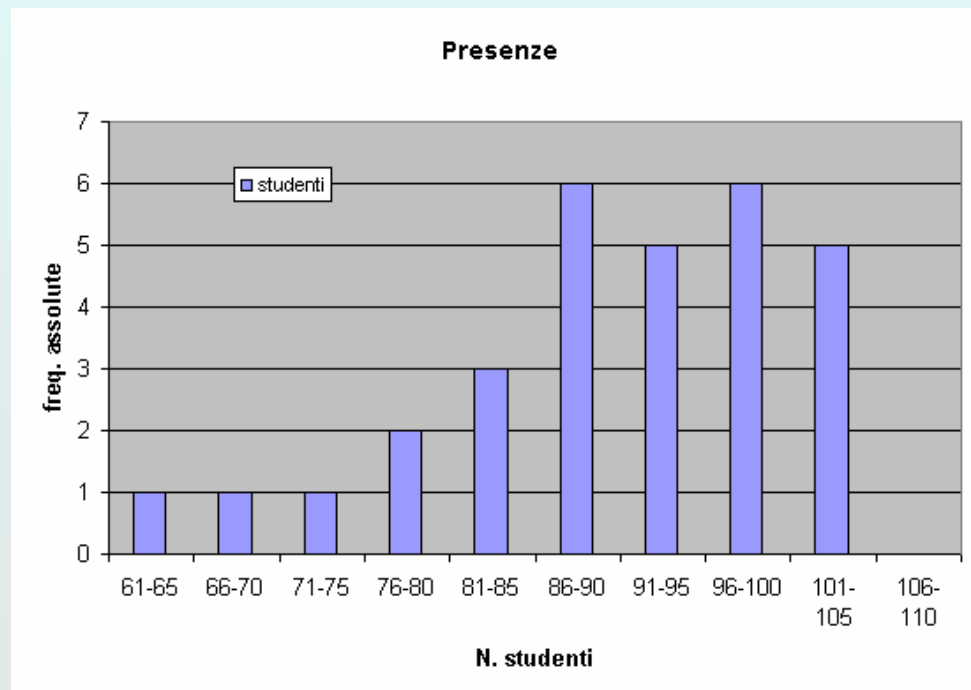
Dati ordinali

Numero di studenti					
#	X	#	X	#	X
1	96	11	92	21	91
2	90	12	98	22	73
3	101	13	93	23	90
4	85	14	95	24	87
5	65	15	84	25	96
6	101	16	69	26	82
7	102	17	87	27	79
8	95				
9	98				
10	99				

X = N. studenti presenti

X	N
61 - 65	1
66 - 70	1
71 - 75	1
76 - 80	2
81 - 85	3
86 - 90	6
91 - 95	5
96 - 100	6
101 - 105	5
106 - 110	0

Valori aggregati per le classi
=
Maggior chiarezza



Costruzione di una tabella di frequenza

Dati ordinali

Numero di studenti					
#	X	#	X	#	X
1	96	11	92	21	91
2	90	12	98	22	73
3	101	13	93	23	90
4	85	14	95	24	87
5	65	15	84	25	96
6	101	16	69	26	82
7	102	17	87	27	70
8	95				
9	98				
10	99				

Frequenze relative

~

Frequenze assolute

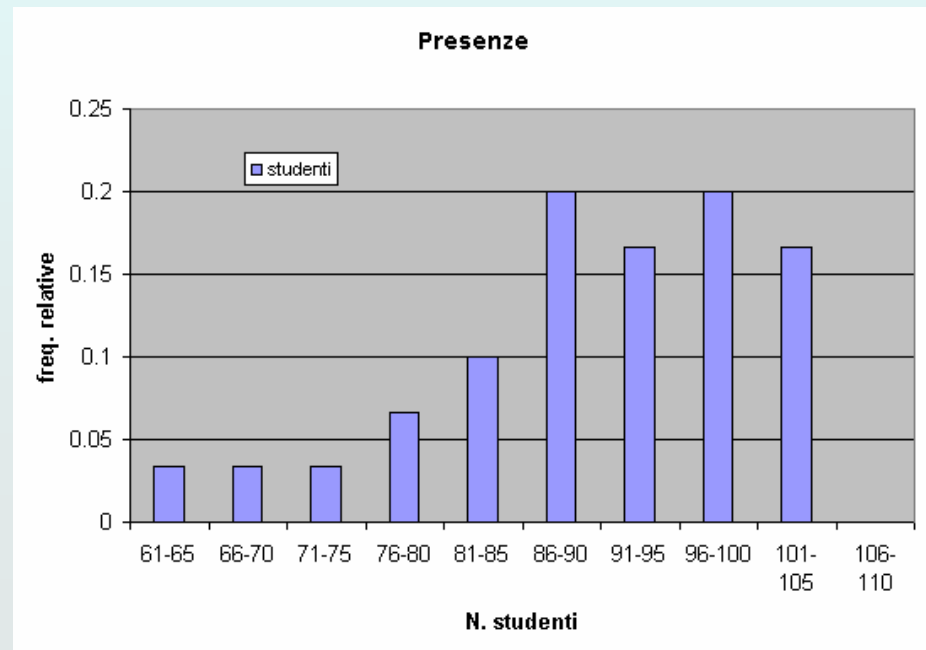
$$f_i = \frac{N_i}{N_T}$$

$$i) \quad 0 \leq f_i \leq 1$$

$$ii) \quad \sum_{i=1}^m f_i = 1$$

X = N. studenti presenti

X	f
61 - 65	0.033
66 - 70	0.033
71 - 75	0.033
76 - 80	0.067
81 - 85	0.100
86 - 90	0.200
91 - 95	0.167
96 - 100	0.200
101 - 105	0.167
106 - 110	0.000

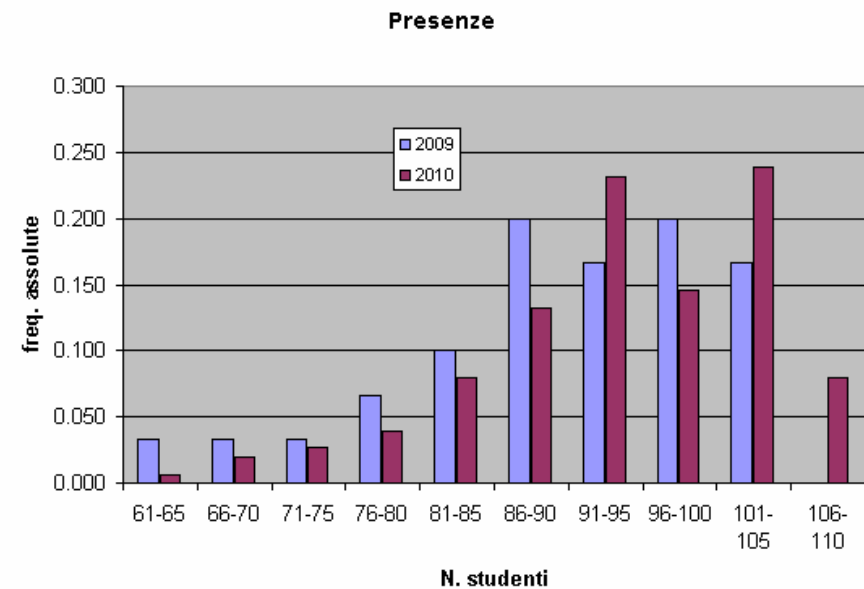
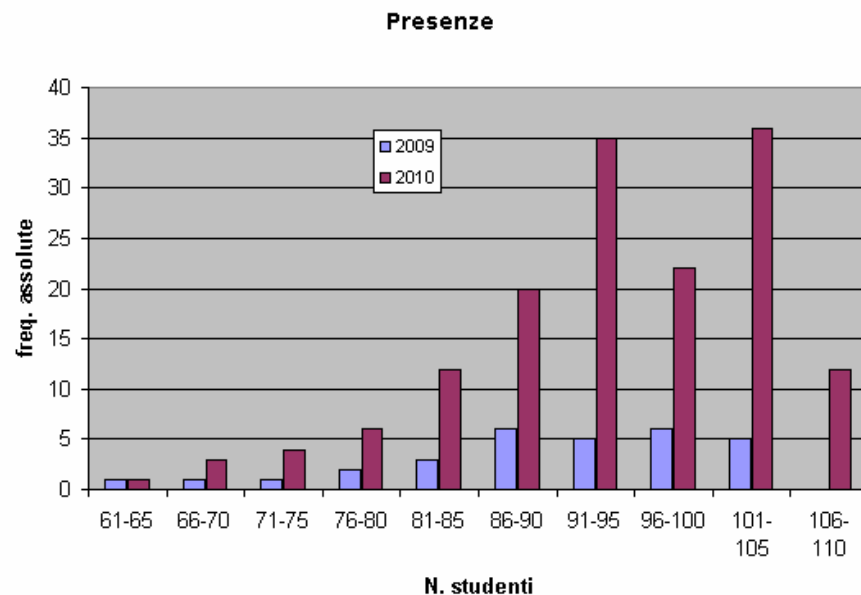


Costruzione di una tabella di frequenza

Dati ordinali

	2009			2010	
61-65	1	0.033		1	0.007
66-70	1	0.033		3	0.020
71-75	1	0.033		4	0.026
76-80	2	0.067		6	0.040
81-85	3	0.100		12	0.079
86-90	6	0.200		20	0.132
91-95	5	0.167		35	0.232
96-100	6	0.200		22	0.146
101-105	5	0.167		36	0.238
106-110	0	0.000		12	0.079
	30			151	

Frequenze relative
=
utili per i confronti



dei dati sperimentali

Costruzione di una tabella di frequenza

Dati ordinali

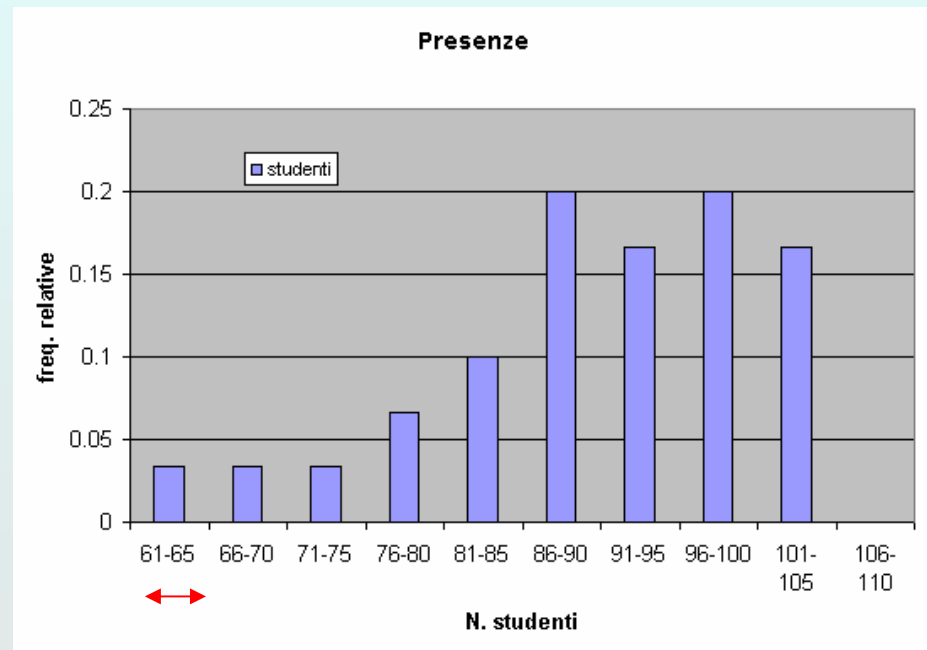
Quante classi?

$$K \sim \sqrt{N}$$

$$K \sim 1 + \log_2 N$$

$$\Delta \sim \frac{x_{max} - x_{min}}{K}$$

Buon senso e
chiarezza



Costruzione di una tabella di frequenza

Versione semplificata: solo l'estremo superiore della classe (N oppure f)

X = N. studenti presenti	
X	f
61 - 65	0.033
66 - 70	0.033
71 - 75	0.033
76 - 80	0.067
81 - 85	0.100
86 - 90	0.200
91 - 95	0.167
96 - 100	0.200
101 - 105	0.167
106 - 110	0.000

X = N. studenti presenti		
X	N	f
61	0	0.000
65	1	0.033
70	1	0.033
75	1	0.033
80	2	0.067
85	3	0.100
90	6	0.200
95	5	0.167
100	6	0.200
105	5	0.167
110	0	0.000

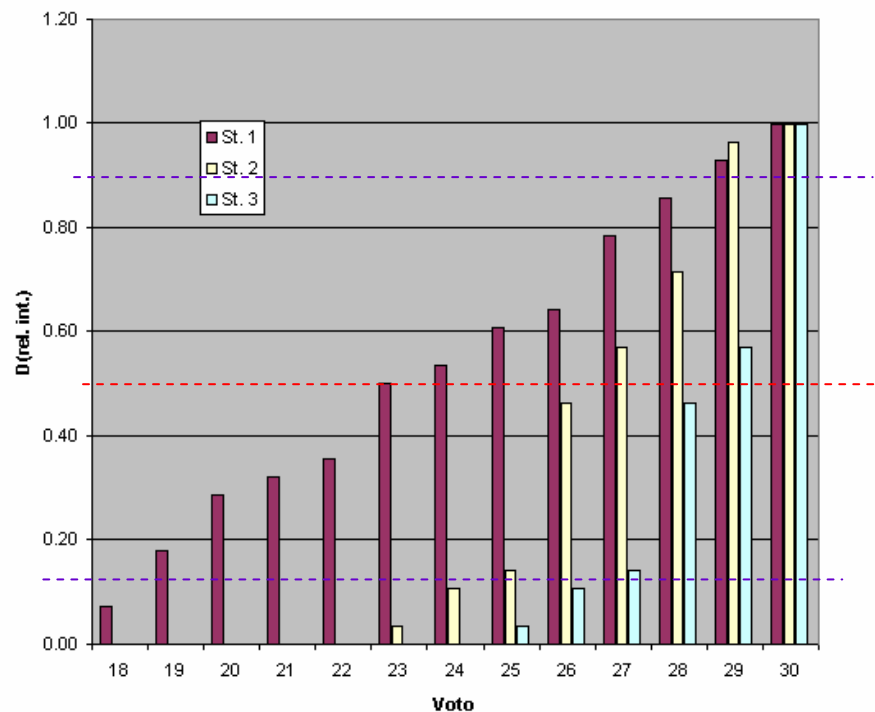
Costruzione di una tabella di frequenza

Dati ordinali

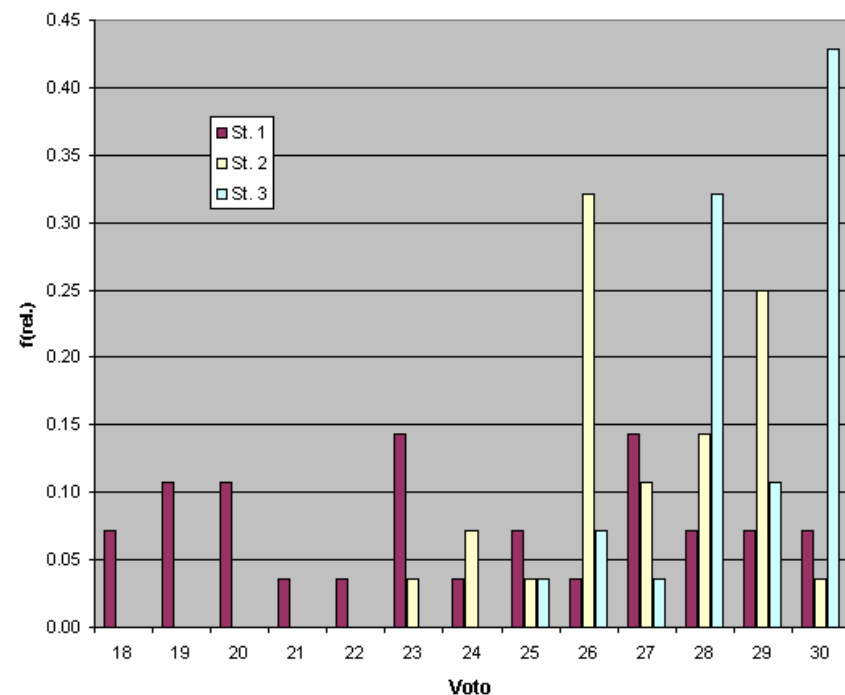
Frequenze o frequenze integrate?

Dipende dalle informazioni che si vogliono mettere in risalto

Frequenze relative integrate



Frequenze relative



info
i sper

Costruzione di una tabella di frequenza

Dati ordinali

Le frequenze integrate assolute sono
facili da calcolare dai dati ordinati

dati

X = Peso frutti					
#	X[g]	#	X[g]	#	X[g]
1	14.1	11	12.9	21	14.0
2	11.2	12	9.9	22	13.9
3	15.7	13	9.5	23	19.0
4	18.8	14	12.1	24	14.7
5	18.6	15	12.7	25	14.4
6	20.2	16	8.6	26	13.5
7	8.4	17	13.3	27	20.9
8	14.3	18	13.8	28	17.6
9	18.3	19	15.4	29	22.1
10	11.7	20	13.9	30	13.0

$T(x) = \text{n. di osservazioni con valore minore o uguale a } x$

dati ordinati

X = Peso frutti					
X[g]	T	X[g]	T	X[g]	T
8.4	1	13.3	11	15.4	21
8.6	2	13.5	12	15.7	22
9.5	3	13.8	13	17.6	23
9.9	4	13.9	14	18.3	24
11.2	5	13.9	15	18.6	25
11.7	6	14.0	16	18.8	26
12.1	7	14.1	17	19.0	27
12.7	8	14.3	18	20.2	28
12.9	9	14.4	19	20.9	29
13.0	10	14.7	20	22.1	30

$T(x)$ è una
funzione continua

Costruzione di una tabella di frequenza

Dati ordinali

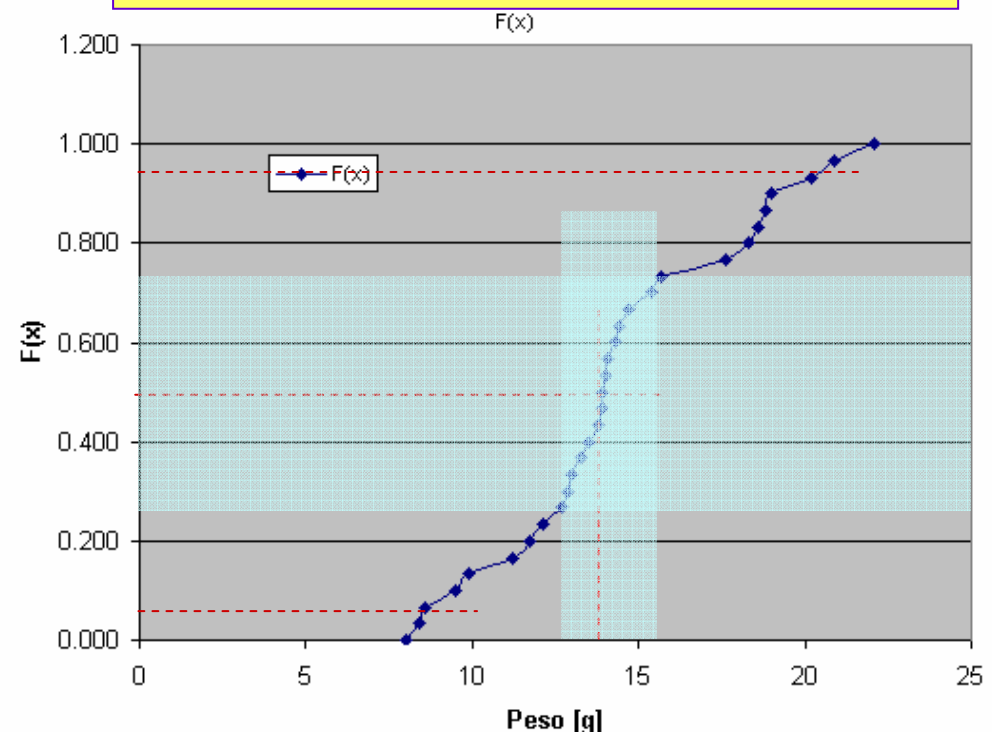
Le frequenze integrate relative $F(x)$

$$F(x_i) = \frac{T(x)}{N_T}$$

X = Peso frutti [g]					
X[g]	F	X[g]	F	X[g]	F
8.4	0.033	13.3	0.367	15.4	0.700
8.6	0.067	13.5	0.400	15.7	0.733
9.5	0.100	13.8	0.433	17.6	0.767
9.9	0.133	13.9	0.467	18.3	0.800
11.2	0.167	13.9	0.500	18.6	0.833
11.7	0.200	14.0	0.533	18.8	0.867
12.1	0.233	14.1	0.567	19.0	0.900
12.7	0.267	14.3	0.600	20.2	0.933
12.9	0.300	14.4	0.633	20.9	0.967
13.0	0.333	14.7	0.667	22.1	1.000

C.I. - statis
de

$F(x)$ è una funzione continua



Costruzione di una tabella di frequenza

Dati ordinali

$$N(x_{i-1}, x_i) = T(x_i) - T(x_{i-1})$$

$$f(x_i) = \frac{T(x_i) - T(x_{i-1})}{N_T} = F(x_i) - F(x_{i-1})$$

$f(x)$ come densità di frequenza

$$f(x) = \lim_{dx \rightarrow 0} \frac{F(x) - F(x - dx)}{dx} = \frac{dF(x)}{dx}$$

$$F(x) = \int_{x_{min}}^x f(x) dx \quad T(x) = N_T \int_{x_{min}}^x f(x) dx$$

X = Peso frutti [g]		
X[g]	N	f
8.0	0	0.000
10.0	4	0.133
12.0	2	0.067
14.0	9	0.300
16.0	7	0.233
18.0	1	0.033
20.0	4	0.133
22.0	2	0.067
24.0	1	0.033
26.0	0	0.000

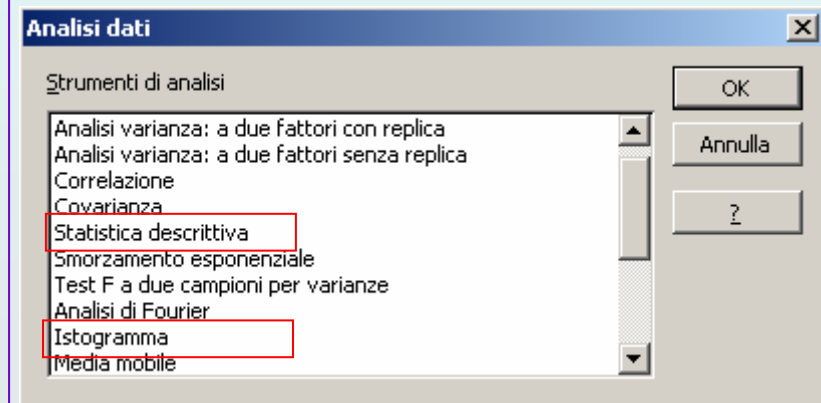
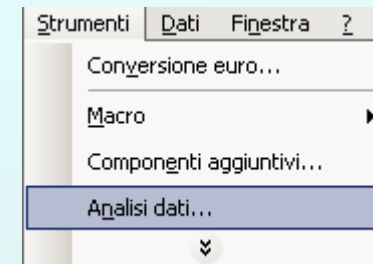
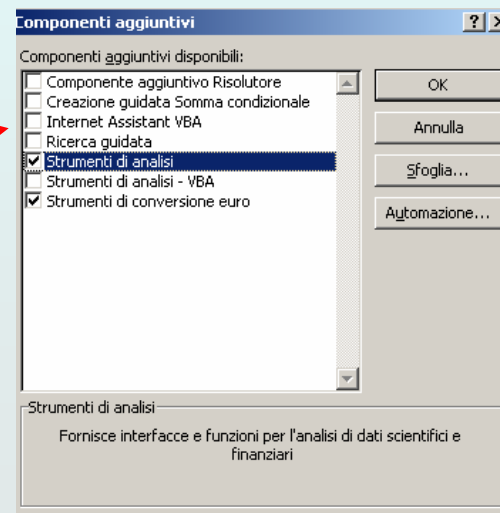
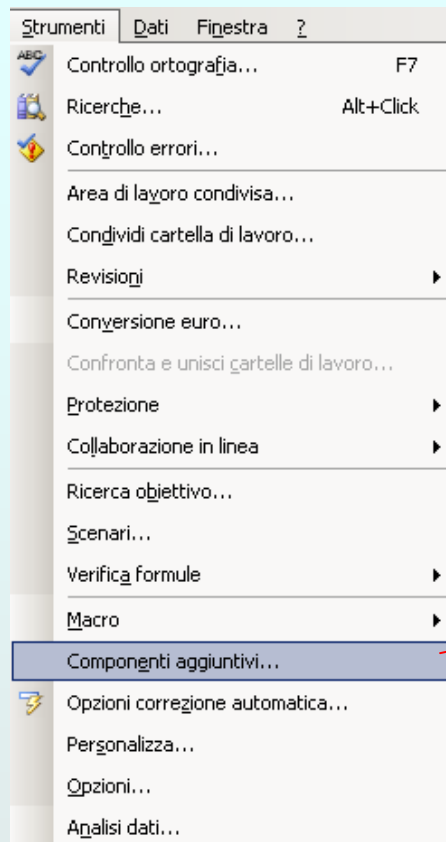
Costruzione di una tabella di frequenza

Dati ordinali Ricetta:

- 1) calcolare i valori estremi dei valori ottenuti: $x_{\min}$, $x_{\max}$, l'intervallo dei valori ($R = x_{\max} - x_{\min}$) e il numero totale di osservazioni N_T ;
- 2) Definire l'intervallo delle classi Δ ;
- 3) Calcolare gli intervalli delle classi;
- 4) Calcolare $T(x)$
- 5) Calcolare F o f
- 6) Preparare la tabella
- 7) Indicare il tipo di frequenza, indicare N_T se si usano frequenze relative, indicare l'unità di misura dei valori

Costruzione di una tabella di frequenza

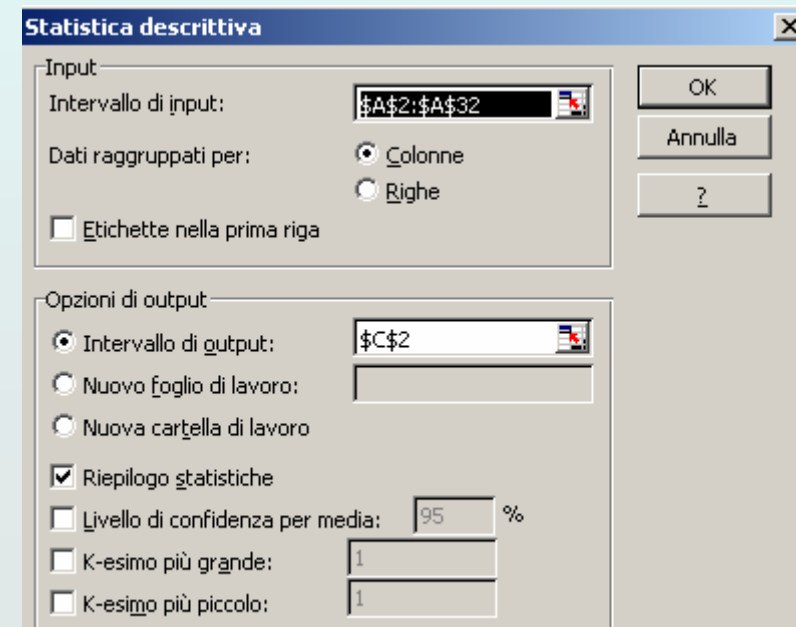
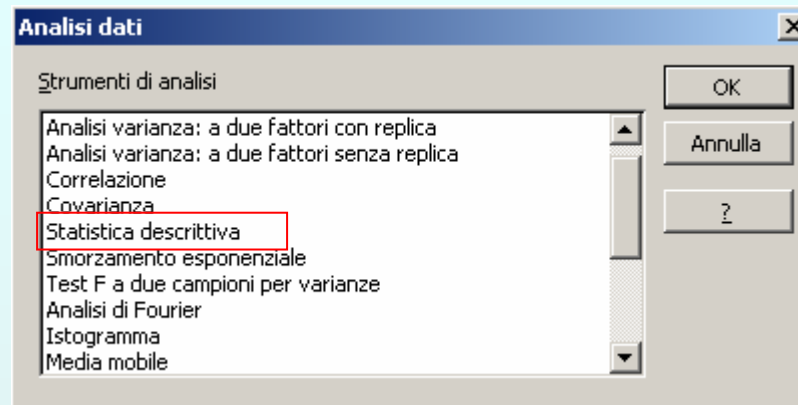
Componenti avanzati



Costruzione di una tabella di frequenza

Statistica descrittiva

Peso[g]			
20.2			Colonna1
8.4			
11.7	Media	14.33871	
9.5	Errore star	0.663466	
9.9	Mediana	13.9	
11.2	Moda	13.9	
14.3	Deviazione	3.69402	
12.1	Varianza c	13.64578	
12.7	Curtosi	-0.41856	
12.9	Asimmetri	0.277923	
14	Intervallo	14.1	
13.3	Minimo	8	
13.5	Massimo	22.1	
13.8	Somma	444.5	
14.4	Conteggio	31	
13.9			
15.7			
14.1			
8			
22.1			
14.7			
15.4			
8.6			
17.6			
18.3			
18.6			
18.8			
19			
13			
20.9			
13.9			



C.I. - statistica, informati
dei dati sperimentali

Costruzione di una tabella di frequenza

Frequenze assolute

Peso[g]
20.2
8.4
11.7
9.5
9.9
11.2
14.3
12.1
12.7
12.9
14
13.3
13.5
13.8
14.4
13.9
15.7
14.1
8
22.1
14.7
15.4
8.6
17.6
18.3
18.6
18.8
19
13
20.9
13.9

classi
8
10
12
14
16
18
20
22
24

Analisi dati

Strumenti di analisi

- Analisi varianza: ad un fattore
- Analisi varianza: a due fattori con replica
- Analisi varianza: a due fattori senza replica
- Correlazione
- Covarianza
- Statistica descrittiva
- Smorzamento esponenziale
- Test F a due campioni per varianze
- Analisi di Fourier
- Istogramma**

OK Annulla ?

Istogramma

Input

Intervallo di input:

Intervallo della classe:

☐ Etichette

Opzioni di output

☒ Intervallo di output:

☐ Nuovo foglio di lavoro:

☐ Nuova cartella di lavoro

☐ Pareto (istogramma ordinato)

☐ Percentuale cumulativa

☐ Grafico in output

OK Annulla ?

Classe	Frequenza
8	1
10	4
12	2
14	10
16	6
18	1
20	4
22	2
24	1
Altro	0

Costruzione di una tabella di frequenza

Frequenze assolute: espressione in forma di matrice

PENDENZA ✖ ✔ fx =frequenza(A2:A32,C3:C11)						
	A	B	C	D	E	F
1	Peso[g]					
2	20.2		classi	freq.		
3	8.4		8			
4	11.7		10			
5	9.5		12			
6	9.9		14			
7	11.2		16			
8	14.3		18			
9	12.1		20			
10	12.7		22			
11	12.9		24			
12	14					
13	13.3					
14	13.5					
15	13.8					
16	14.4					
17	13.9					
18	15.7					
19	14.1					
20	8					
21	22.1					
22	14.7					
23	15.4					
24	8.6					
25	17.6					
26	18.3					
27	18.6					
28	18.8					
29	19					
30	13					
31	20.9					
32	13.9					

D4 ✖ fx {=FREQUENZA(A2:A32,C3:C11)}					
	A	B	C	D	E
1	Peso[g]				
2	20.2		classi	freq.	
3	8.4		8	1	
4	11.7		10	4	
5	9.5		12	2	
6	9.9		14	10	
7	11.2		16	6	
8	14.3		18	1	
9	12.1		20	4	
10	12.7		22	2	
11	12.9		24	1	
12					
13					
14	13.5				
15	13.8				
16	14.4				
17	13.9				

Crtl + SHFT + ENTER

Calcola direttamente le
frequenze assolute per un
insieme di classi

statistica, inform
dei dati sperimentali

Calcolo degli indici statistici per valori aggregati: media

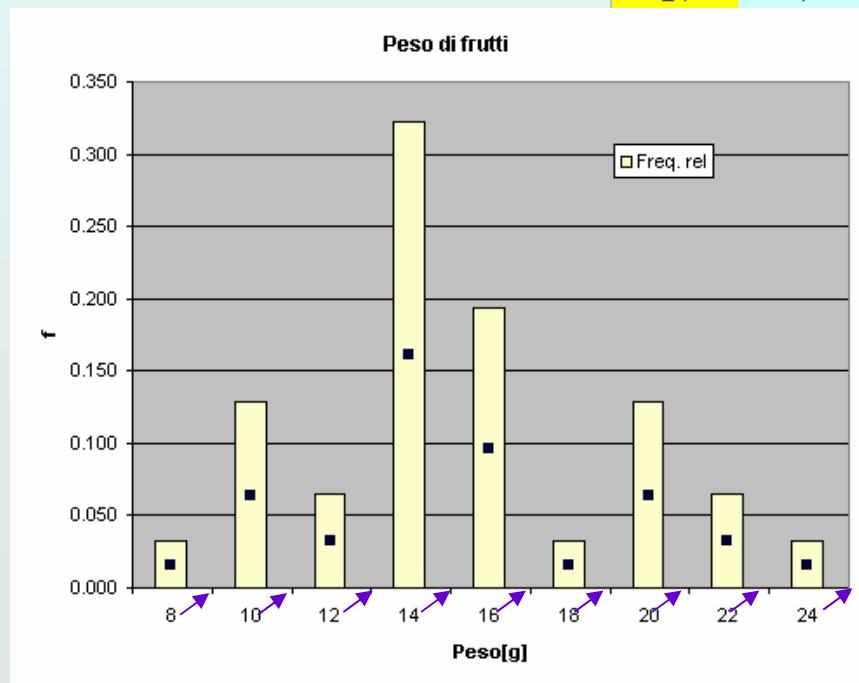
Peso[g]
20.2
8.4
11.7
9.5
9.9
11.2
14.3
12.1
12.7
12.9
14
13.3
13.5
13.8
14.4
13.9
15.7
14.1
8
22.1
14.7
15.4
8.6
17.6
18.3
18.6
18.8
19
13
20.9
13.9

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

Colonna1	
Media	14.33871

classi peso(g)	freq. Ass.	Freq. rel
8	1	0.032
10	4	0.129
12	2	0.065
14	10	0.323
16	6	0.194
18	1	0.032
20	4	0.129
22	2	0.065
24	1	0.032

	x*f
	0.2581
	1.2903
	0.7742
	4.5161
	3.0968
	0.5806
	2.5806
	1.4194
	0.7742
media agg	15.290



$$\bar{x} = \sum_{j=1}^k f_j x_j - \frac{\Delta}{2}$$

$$\bar{x} = \sum_{j=1}^k f_j x_j$$

Punto medio
dell'intervallo

Calcolo degli indici statistici per valori aggregati: varianza e dev.st.

Peso[g]	classi peso(g)	freq. Ass.	Freq. rel		x*f	f(x-media)^2
20.2						
8.4	8	1	0.032		0.2581	1.7145
11.7	10	4	0.129		1.2903	3.6113
9.5	12	2	0.065		0.7742	0.6985
9.9	14	10	0.323		4.5161	0.5371
11.2	16	6	0.194		3.0968	0.0975
14.3	18	1	0.032		0.5806	0.2369
12.1	20	4	0.129		2.5806	2.8621
12.7	22	2	0.065		1.4194	2.9045
12.9	24	1	0.032		0.7742	2.4470
14						
13.3	Colonna1			media-agg	15.290	
13.5						
13.8	Media	14.33871		Media	14.290	
14.4	Errore star	0.663466				
13.9	Mediana	13.9				
15.7	Moda	13.9				
14.1	Deviazione	3.69402		dev.st	3.95	
8	Varianza c	13.64578		varianza	15.61	
22.1	Curtosi	-0.41856				
14.7	Asimmetri	0.277923				
15.4	Intervallo	14.1				
8.6	Minimo	8				
17.6	Massimo	22.1				
18.3	Somma	444.5				
18.6	Conteggio	31				

$$s^2 = \frac{N}{N-1} \sum_{j=1}^k f_j (x_j - \bar{x})^2$$

$$s^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2$$

Peso[g]	classi peso(g)	freq. Ass.	Freq. rel		$x \cdot f$	$f(x - \text{media})^2$
20.2						
8.4	8	1	0.032		0.2581	1.7145
11.7	10	4	0.129		1.2903	3.6113
9.5	12	2	0.065		0.7742	0.6985
9.9	14	10	0.323		4.5161	0.5371
11.2	16	6	0.194		3.0968	0.0975
14.3	18	1	0.032		0.5806	0.2369
12.1	20	4	0.129		2.5806	2.8621
12.7	22	2	0.065		1.4194	2.9045
12.9	24	1	0.032		0.7742	2.4470
14						
13.3	Colonna1			media agg	15.290	
13.5						
13.8	Media	14.33871		Media	14.290	
14.4	Errore star	0.663466				
13.9	Mediana	13.9				
15.7	Moda	13.9				
14.1	Deviazione	3.69402		dev.st	3.95	
8	Varianza c	13.64578		varianza	15.61	
22.1	Curtosi	-0.41856				
14.7	Asimmetri:	0.277923				
15.4	Intervallo	14.1				
8.6	Minimo	8				
17.6	Massimo	22.1				
18.3	Somma	444.5				
18.6	Conteggio	31				
18.8						
19						
13						
20.9						
13.9						

Peso[g]	classi peso(g)	freq. Ass.	Freq. rel		$x \cdot f$	$f(x - \text{media})^2$
20.2						
8.4	7	0	0.000		0.0000	0.0000
11.7	9	3	0.097		0.8710	3.7123
9.5	11	2	0.065		0.7097	1.1346
9.9	13	6	0.194		2.5161	0.9313
11.2	15	10	0.323		4.8387	0.0121
14.3	17	2	0.065		1.0968	0.2105
12.1	19	5	0.161		3.0645	2.3369
12.7	21	2	0.065		1.3548	2.1752
12.9	23	1	0.032		0.7419	1.9658
14						
13.3	Colonna1			media agg	15.194	
13.5						
13.8	Media	14.33871		Media	14.194	
14.4	Errore star	0.663466				
13.9	Mediana	13.9				
15.7	Moda	13.9				
14.1	Deviazione	3.69402		dev.st	3.59	
8	Varianza c	13.64578		varianza	12.89	
22.1	Curtosi	-0.41856				
14.7	Asimmetri:	0.277923				
15.4	Intervallo	14.1				
8.6	Minimo	8				
17.6	Massimo	22.1				
18.3	Somma	444.5				
18.6	Conteggio	31				
18.8						
19						
13						
20.9						
13.9						

