

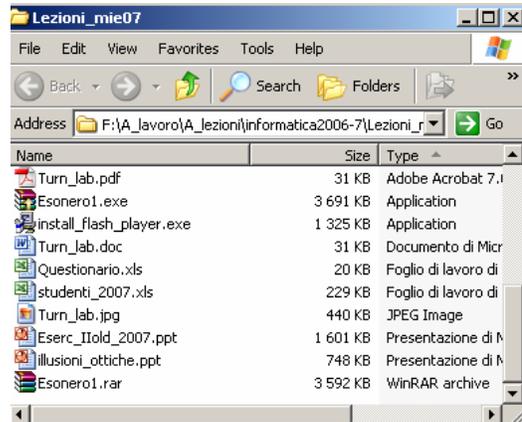
Corso integrato di informatica, statistica e analisi dei dati sperimentali
Altri esercizi_ esercitazione V

Sui PC a disposizione sono installati diversi sistemi operativi. All'accensione scegliere Windows.

Immettere Nome utente **b##** (## : numero del pc)
 Pass.: **biologia##** (## : numero del pc)
Stud_fisica

Il sistema operativo è Windows 2000. (Non molto diverso dal WXP!) base a quanto detto a lezione controllare che:

- a) le cartelle mostrino i dettagli dei files;
- b) siano visibili le estensioni dei files anche per i tipi conosciuti;
- c) la configurazione della tastiera corrisponda alla tastiera effettivamente in uso (Inglese);
- d) i numeri decimali siano codificati con il “.” e sia usato lo spazio “ ” per il raggruppamento delle cifre.



- e) Attivare **Excel** e controllare che siano installate le opzioni “analisi dati” o aggiungerle utilizzando “strumenti > opzioni aggiuntive”
- f) Creare una cartella di lavoro nella cartella dei documenti “**Eserc_A##**” dove salvare il proprio lavoro.
- g) Scaricare e installare la cartella con i dati per l’esercitazione dalla pag. web del corso (**esercitazioneA.exe**)

Esercizio 1) Statistica descrittiva Importare un file di dati in Excel, calcolo di indici di posizione, ampiezza e asimmetria. Confronto di dati, Istogramma di frequenze.

Il file di dati: **studenti.dat** è un file ASCII a tre colonne che riporta l’età e l’altezza di un gruppo di studenti divisi per sesso:

1	se	eta'	peso
2	F	21	55
3	M	20	65
4	M	20	60
5	F	21	58
6	M	19	62

A- Importare i dati in Excel utilizzando uno dei seguenti metodi:

- a) aprendo il file con Excel,
- b) aprendo il file con un editor di testo (blocco note) e utilizzando le opzioni copia e incolla
- c) aprendo il file da Excel utilizzando *File > apri > ...*
 Utilizzare l’opzione *dati > testo in colonne* per organizzare i dati nelle celle.

ATTENZIONE: quando si apre un file ASCII con EXCEL ricordarsi di salvare il lavoro utilizzando il formato xls in uscita, altrimenti si perde tutto il lavoro fatto!

B- Utilizzare le funzioni di Excel per calcolare indici di posizione e variabilità presentare i risultati in una tabella con 2 cifre significative

Statistica descrittiva		
N	tutti	
189	età	peso
	Anni	Kg
max	31	97
min	18	43
media	20.11	64.69
moda	19.00	58.00
mediana	19.00	62.00
Q1	19.00	55.00
Q3	20.00	75.00
varianza	4.43	137.63
dev.st.	2.10	11.73
interq.	1.00	20.00

C- Determinare il numero e la frazione di uomini e donne nel campione utilizzando, ad esempio, le funzioni `CONTA.VALORI(...)` e `CONTA.SE(...)`

1	A	B	C	D	E	F	G	H	I
	SESSO	ETA'	PESO			Statistica descrittiva			
2		Anni	Kg			N	tutti		
3	F	21	55			169	età	peso	
4	M	20	65				Anni	Kg	
5	M	19	60			max	31	97	
6	F	21	58			min	18	43	
7	M	19	82			media	20.11	64.69	
8	F	18	50			moda	19.00	58.00	
9	M	22	81			mediana	19.00	62.00	
10	M	19	70			Q1	19.00	55.00	
11	M	19	75			Q3	20.00	75.00	
12	M	20	58			varianza	4.43	137.63	
13	M	23	72			dev.st.	2.10	11.73	
14	M	19	72			interq.	1.00	20.00	
15	M	20	77						
16	M	20	65						
17	M	20	76						
18	F	21	60						
19	F	26	65						
20	F	21	54						
21	M	18	63						

`=CONTA.SE(A3:A171,"M")`

`=CONTA.VALORI(A3:A171)`

	N. stud.	Maschi	Femmine
	169	73	96
frazione		0.43	0.57
%		43.2	56.8

D- descrivere separatamente la distribuzione di altezza e peso per uomini e donne (Si suggerisce di usare un nuovo foglio). Per questo, ad esempio, si possono **ordinare** i dati (usando l'opzione `dati>ordina`) in modo da poter selezionare in modo semplice i dati per Maschi e e Femmine

Statistica descrittiva				
	età anni		peso Kg	
	F	M	F	M
N	96	73	96	73
max	31	29	80	97
min	18	18	43	58
media	20.08	20.14	57.57	74.05
moda	19.00	19.00	55.00	80.00
mediana	19.00	19.00	55.50	75.00
Q1	19.00	19.00	52.00	68.00
Q3	20.25	20.00	61.00	80.00
varianza	5.09	3.62	61.11	84.05
dev.st.	2.26	1.90	7.82	9.17
interq.	1.25	1.00	9.00	12.00

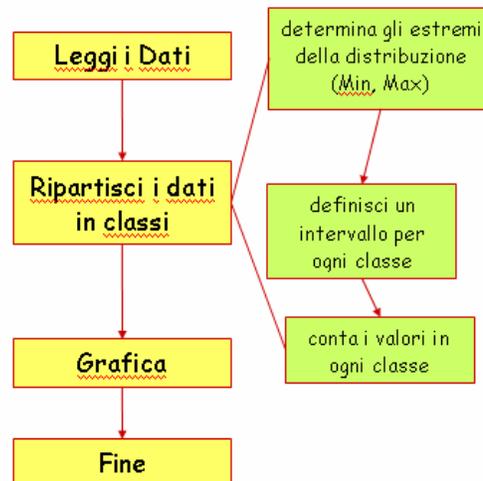
Per confrontare i valori medi ottenuti per altezza e peso si calcolano l'errore standard sulla media ($err = dev.st / N^{1/2}$). Si può vedere che la differenza di età tra la popolazione di uomini e donne mostra una differenza nell'altezza media di maschi e femmine mostra una differenza molto piccola rispetto all'errore statistico sulla media. Al contrario la differenza tra i pesi è molto maggiore degli errori statistici sulle medie. Vedremo in seguito quanto queste differenze siano significative da un punto di vista statistico.

	età anni		peso Kg	
	F	M	F	M
N	96	73	96	73
media	20.08	20.14	60.26	75.38
err.st.	0.2	0.2	0.8	1.1

Utilizzare la funzione avanzata "statistica descrittiva" (in *strumenti > analisi dati*) per riassumere i risultati. Se non compare l'opzione "analisi dati" nel menu *strumenti* utilizzare l'opzione *componenti aggiuntivi* per aggiungere le macro "analisi dati"

E- Costruire un istogramma per visualizzare i dati di peso e altezza. Durante l'esercitazione scorsa sono stati mostrati due metodi per la ripartizione dei dati in classi. Qui di seguito descriviamo un terzo metodo che consiste nell'utilizzare la **definizione di formule in forma in matrice in Excel**:

- 1) una volta definite le classi per le quali definire l'istogramma di frequenze, selezionare il gruppo di celle per le quali calcolare le frequenze,
- 2) digitare nella barra delle formule "`=frequenza(matrice_dati; matrice_classi)`" dove *matrice_dati* è l'insieme delle celle contenenti i dati (solo celle contigue) e *matrice_classi* è l'insieme delle celle contenenti le classi (corrispondente al gruppo di celle per le frequenze),
- 3) per attivare la formula utilizzare **Ctrl + Shift + Enter**, in questo modo la formula viene calcolata automaticamente per per tutte le celle selezionate.



Costruzione di un istogramma: diagramma di flusso dell'algoritmo.

	A	B	C
1	SESSO	ETA'	PESO
2		Anni	Kg
3	F	21	55
4	F	21	58
5	F	18	50
6	F	21	60
7	F	26	65
8	F	21	54
9	F	20	50
0	F	20	52
1	F	23	52
2	F	20	51
3	F	18	55
4	F	19	54
5	F	21	55
6	F	21	53
7	F	19	56
8	F	20	57
9	F	20	75
0	F	19	50
1	F	19	57

1

min	18
max	31

Istogramma	
classi	totali
17	
18	
19	
20	
21	
22	
23	
24	
25	
26	
27	
28	
29	
30	
31	
32	

Definire le classi tra il valore minimo e il valore massimo

Selezionare il gruppo di celle per le quali calcolare la frequenza

2

fx = `=frequenza(B3:B171,F7:F22)`

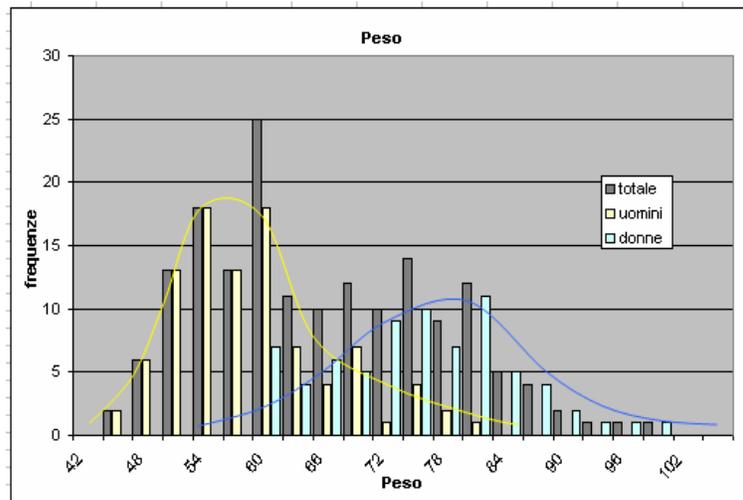
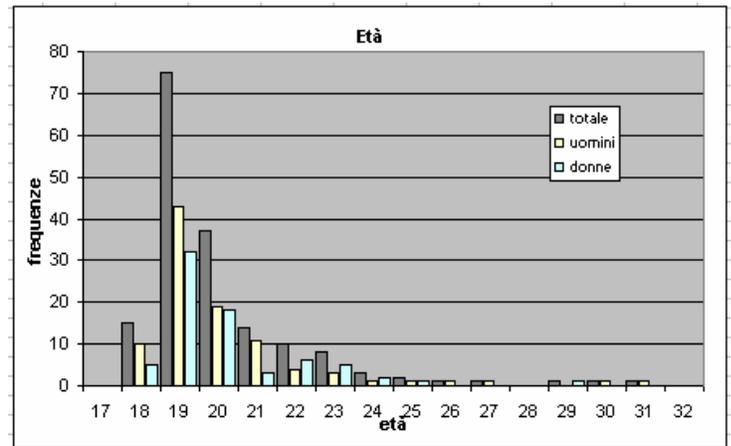
`FREQUENZA(matrice_dati, matrice_classi)`

3

Istogramma	
classi	totali
17	0
18	15
19	75
20	37
21	14
22	10
23	8
24	3
25	2
26	1
27	1
28	0
29	1
30	1
31	1
32	0

Confrontare graficamente le distribuzioni di altezza e peso per uomini e donne.

Istogramma Età				Istogramma peso			
classi	totali	Femmine	Maschi	classi	totali	Femmine	Maschi
17	0	0	0	42	0	0	0
18	15	10	5	45	2	2	0
19	75	43	32	48	6	6	0
20	37	19	18	51	13	13	0
21	14	11	3	54	18	18	0
22	10	4	6	57	13	13	0
23	8	3	5	60	25	18	7
24	3	1	2	63	11	7	4
25	2	1	1	66	10	4	6
26	1	1	0	69	12	7	5
27	1	1	0	72	10	1	9
28	0	0	0	75	14	4	10
29	1	0	1	78	9	2	7
30	1	1	0	81	12	1	11
31	1	1	0	84	5	0	5
32	0	0	0	87	4	0	4
min	18			90	2	0	2
max	31			93	1	0	1
				96	1	0	1
				99	1	0	1
				102	0	0	0
				105	0	0	0
				min	43		
				max	97		



Si osserva che, mentre la distribuzione delle età degli studenti dipende poco dal sesso degli studenti, la distribuzione dei pesi è sensibilmente diversa per uomini e donne.

F- calcolare gli indici di asimmetria e curtosi per le distribuzioni totali e parziali.

Il file **studenti1.xls** contiene una possibile soluzione e può essere utile consultare le formule utilizzate.

Esercizio 2) Studio di distribuzioni

Utilizzare l'opzione *generazione di un numero casuale* di excel per generare distribuzioni di numeri casuali con legge data.

A) Generare una serie di numeri con distribuzione binomiale con probabilità data.

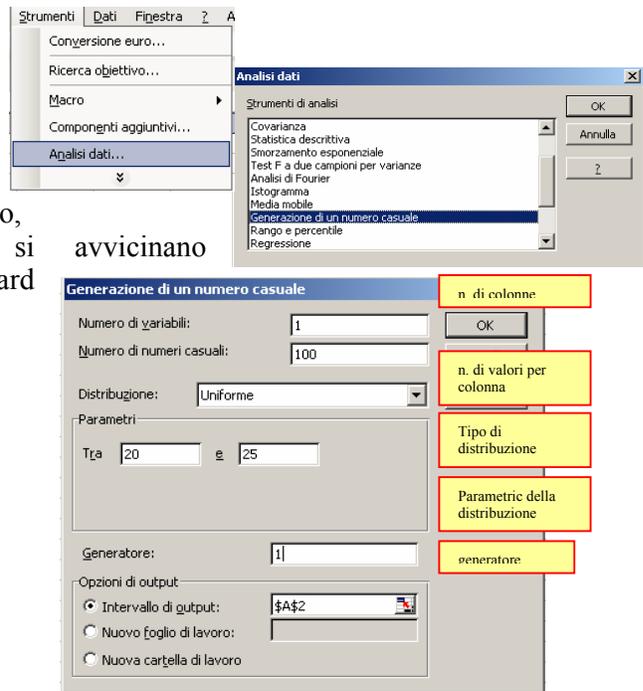
Calcolare valori medi, varianza e deviazione standard cambiando N: numero di punti.

All'aumentare del numero di punti valore medio, varianza e deviazione standard calcolate si avvicinano progressivamente ai valori teorici e l'errore standard sulla media diminuisce.

Attenzione: l'errore sulla media è, per definizione:

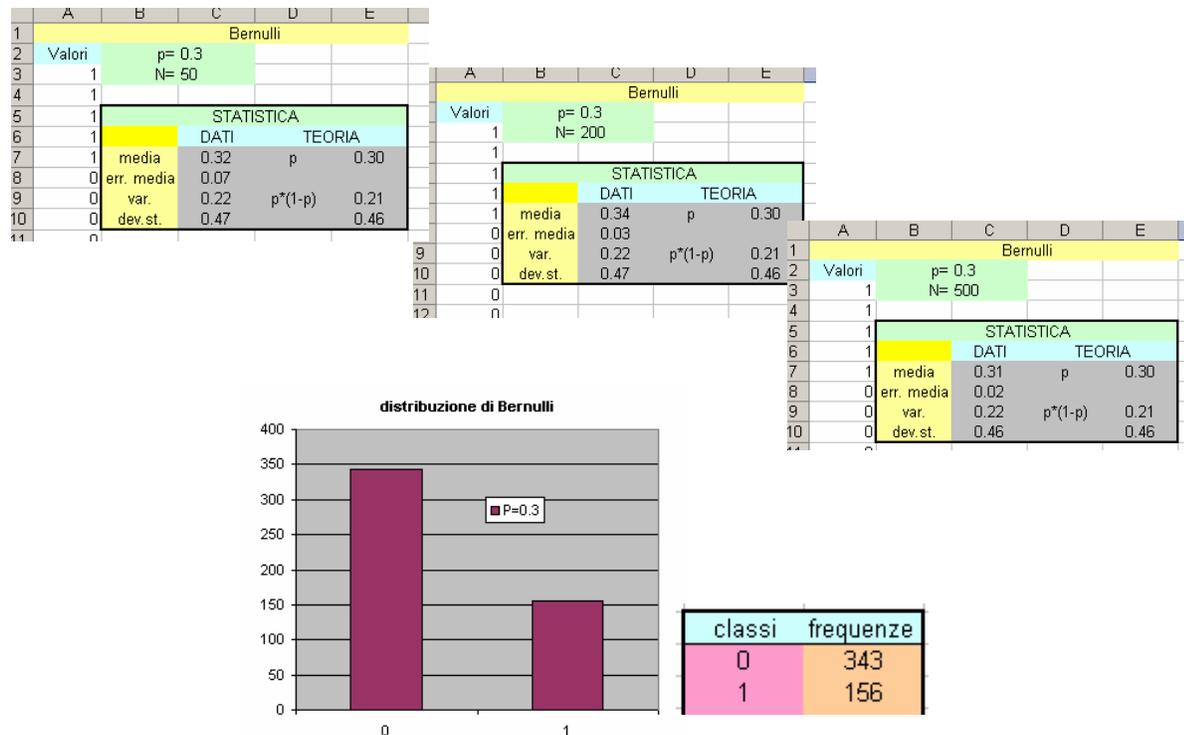
$$\epsilon_{\bar{x}} = \frac{\sigma}{\sqrt{N}}$$

Provare a cambiare il "generatore" della funzione



Visualizzare la distribuzione utilizzando un istogramma.

Il file excel **distribuzioni.xls** mostra alcuni esempi.

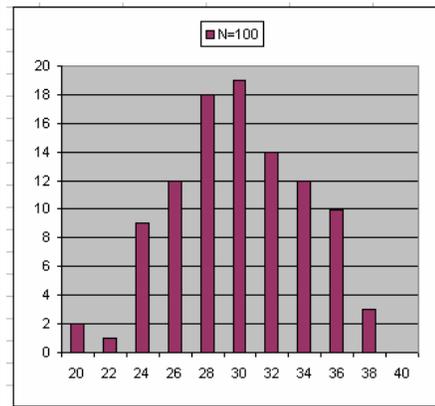


A	B	C	D	E
Binomiale				
Valori	p= 0.3			
15	N= 50			
14				
STATISTICA				
	DATI	TEORIA		
media	14.90	N*p	15.00	
err. media	0.42			
var.	8.62	N*p*(1-p)	10.50	
dev. st.	2.94		3.24	

A	B	C	D	E
Binomiale				
Valori	p= 0.3			
29	N= 100			
33				
25				
STATISTICA				
	DATI	TEORIA		
media	29.47	N*p	30.00	
err. media	0.40			
var.	16.35	N*p*(1-p)	21.00	
dev. st.	4.04		4.58	

A	B	C	D	E
Binomiale				
Valori	p= 0.3			
151	N= 500			
144				
160				
158				
STATISTICA				
	DATI	TEORIA		
media	148.95	N*p	150.00	
err. media	0.45			
var.	99.80	N*p*(1-p)	105.00	
dev. st.	9.99		10.25	

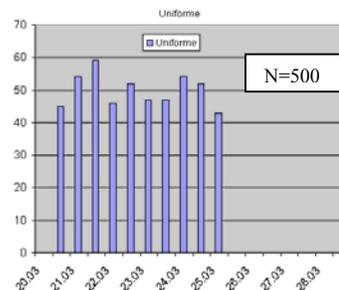
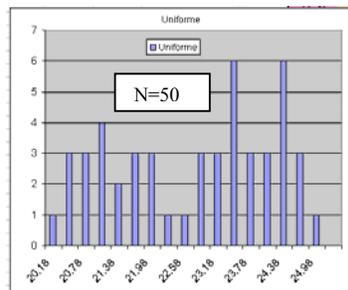
classi	frequenze
20	2
22	1
24	9
26	12
28	18
30	19
32	14
34	12
36	10
38	3
40	0
42	0
44	0
46	0
48	0
50	0
52	0
54	0
56	0
58	0
60	0
62	0
64	0
66	0
68	0
70	0



A	B	C	D	E
Uniforme				
Valori	N= 50			
20.00732				
21.09806				
21.4182				
STATISTICA				
	DATI	TEORIA		
media	22.54	22.50		
err. media	0.21			
var.	2.12	2.08		
dev. st.	1.46	1.44		

A	B	C	D	E
Uniforme				
Valori	N= 150			
20.00732				
21.09806				
21.4182				
STATISTICA				
	DATI	TEORIA		
media	22.45	22.50		
err. media	0.11			
var.	1.90	2.08		
dev. st.	1.38	1.44		

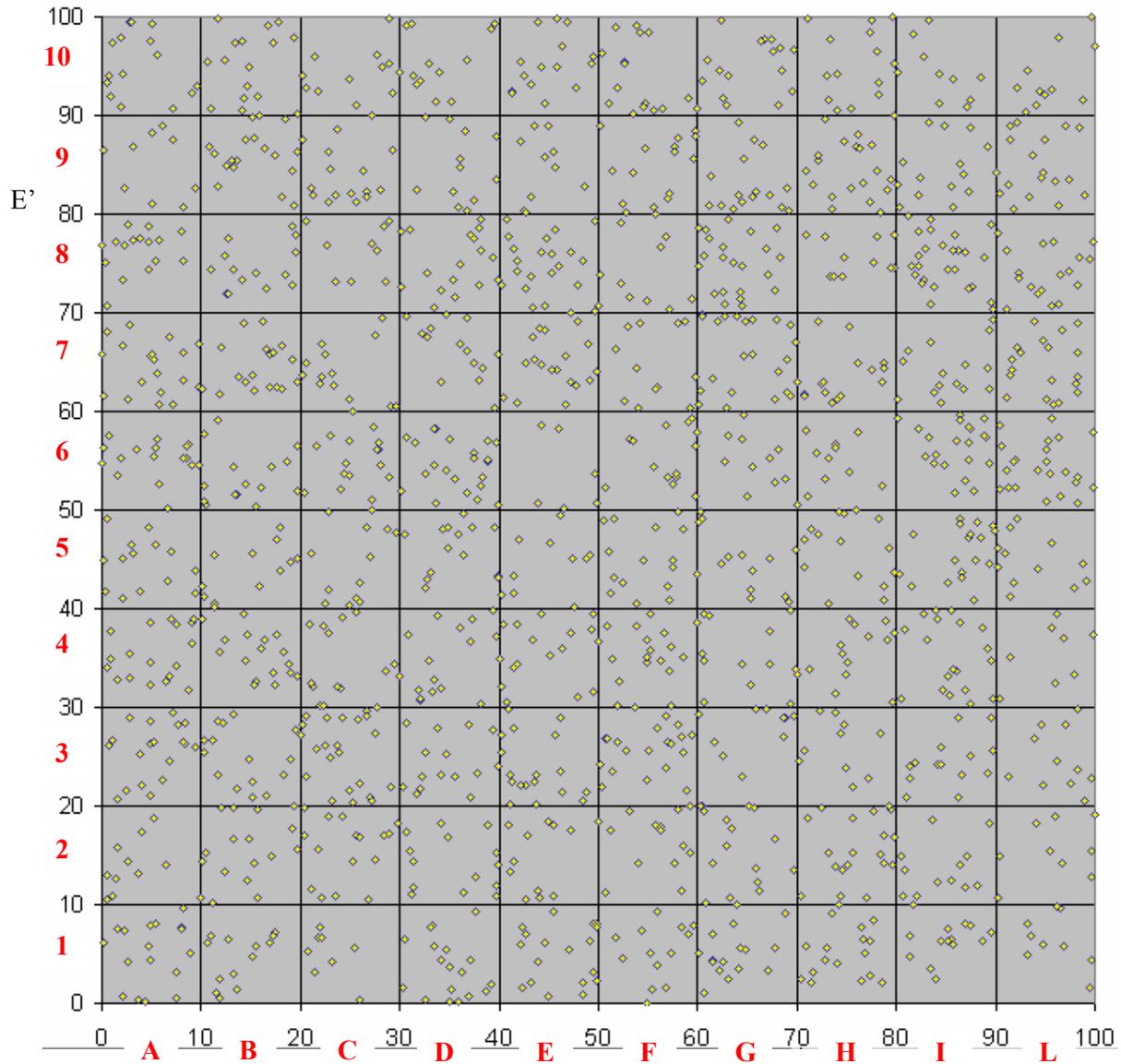
A	B	C	D	E
Uniforme				
Valori	N= 500			
20.00732				
21.09806				
21.4182				
STATISTICA				
	DATI	TEORIA		
media	22.50	22.50		
err. media	0.06			
var.	2.02	2.08		
dev. st.	1.42	1.44		



Esercizio 3) In figura sono riportate le posizioni per 1500 esemplari di piante di una specie A distribuite in una regione di superficie $S=10^4 \text{ m}^2$. Si vuole stimare la densità media di piante δ misurando il numero di piante in uno o più quadrati di prova di lato $L=10\text{m}$, di superficie s_0 . Per una densità costante di piante, il numero di piante k osservate in quadrato di lato L segue una statistica di Poisson:

$$P(k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

il cui valore medio è $\mu=\lambda$ e da la varianza è $\sigma^2=\lambda$.



Infatti possiamo assumere la distribuzione di Poisson come caso limite di una distribuzione di Bernulli con N (numero di prove= numero totale di piante nella regione S) grande e p (probabilità di successo) piccola, p è la probabilità di osservare una pianta nell'area di prova s_0 , quindi, per la distribuzione di Bernulli, la probabilità di successo è

$$p = s_0/S = 0.01$$

Il valore medio atteso dalla distribuzione di Bernulli è $\mu = Np = \delta s_0$. La varianza della distribuzione di Bernulli è $\sigma^2 = N p (1-p)$, essendo $p \ll 1$ si ha: $\sigma^2 \approx N p = \mu$.

A- contando il numero di piante k in uno dei quadrati di lato $L= 10$ m stimare la densità di piante nella regione S :

$$\delta = k / s_0$$

L'errore associato si ottiene dalla deviazione standard della distribuzione $P(k)$, quindi $\varepsilon_\delta = \sigma/s_0$.

Ad esempio per il quadrato A1 il numero di piante è 17, quindi

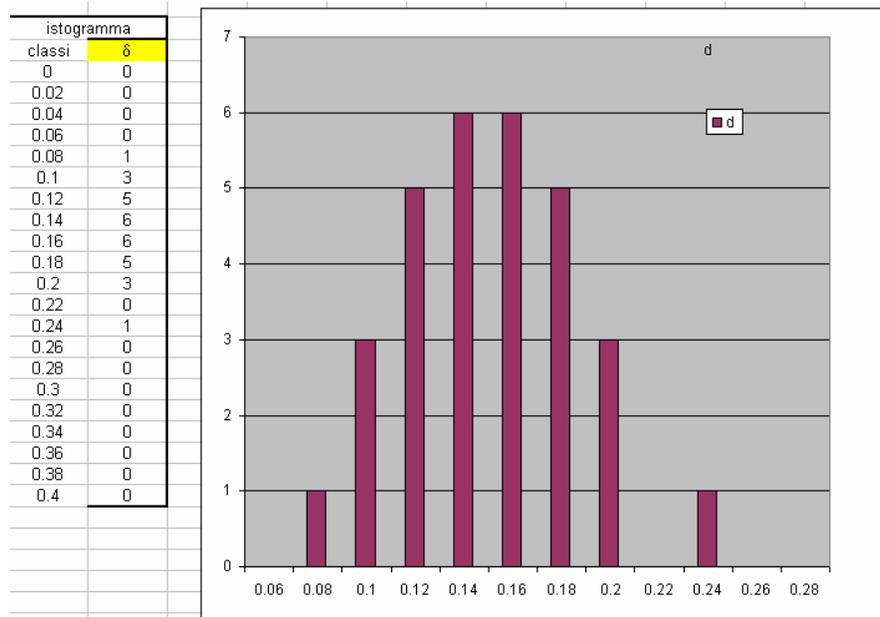
$$\delta_{A1} = 0.17 \pm 0.04$$

riquadro	K	δ	err.
A 1	17	0.17	± 0.041
B 2	13	0.13	± 0.036
C 3	8	0.08	± 0.028
D 4	18	0.18	± 0.042
E 5	18	0.18	± 0.042
F 6	13	0.13	± 0.036
G 7	16	0.16	± 0.040
H 8	17	0.17	± 0.041
I 9	14	0.14	± 0.037
L 10	9	0.09	± 0.030
A 2	11	0.11	± 0.033
B 3	16	0.16	± 0.040
C 4	16	0.16	± 0.040
D 5	14	0.14	± 0.037
E 6	12	0.12	± 0.035
F 7	13	0.13	± 0.036
G 8	15	0.15	± 0.039
H 9	19	0.19	± 0.044
I 10	12	0.12	± 0.035
L 1	9	0.09	± 0.030

B- Ripetere l'osservazione in più quadrati e riportare i dati in una tabella, eventualmente integrando i risultati ottenuti da più gruppi di studenti.

Dalla tabella si può osservare che i valori di densità osservati si distribuiscono attorno al valore medio atteso ($N / S = 0.15 \text{ m}^{-2}$)

Riportare i risultati su un istogramma:



Calcolare il valore medio, la varianza e la deviazione standard della distribuzione dei valori osservati K_i ; il valore medio e la varianza della distribuzione sono simili, come ci si aspetta per una distribuzione di Poisson.

Valori medi			
	K	δ	err.
media	14.53	0.145	± 0.038
varianza	13.36		
dev.st	3.66	0.037	

C- Ripetere la misura utilizzando riquadri di campionamento di lato L=20 m.

In questo caso l'errore sulla densità stimata è dimezzato.

Valori medi				
	K	δ		err.
media	60.00	0.150	±	0.019
varianza	87.83			
dev. st	9.37	0.023		

riquadro	K	δ		err.
A 1	57	0.1425	±	0.019
B 2	56	0.14	±	0.019
C 3	56	0.14	±	0.019
D 4	67	0.1675	±	0.020
E 5	44	0.11	±	0.017
F 6	69	0.1725	±	0.021
G 7	68	0.17	±	0.021
H 8	72	0.18	±	0.021
I 9	47	0.1175	±	0.017
L 10	52	0.13	±	0.018
A 2	58	0.145	±	0.019
B 3	65	0.1625	±	0.020
C 4	46	0.115	±	0.017
D 5	49	0.1225	±	0.018