

# Corso di Statistica

## Facoltà di Economia

**francesco mola**

a.a. 2010-2011

### Lezione n° 2

## Sommario

- Distribuzioni di frequenza
- Rappresentazioni grafiche
  - Diagrammi a barre
  - Istogrammi
- Funzione di ripartizione empirica

lez2\_2010-2011

statistica-francesco mola

2

## Distribuzioni

### Distribuzioni di quantità

Vale solo per caratteri trasferibili

Auto Vendute	
Marca 1	15643
Marca 2	277765
Marca 3	235016
Marca 4	209339
Marca 5	183456
Marca 6	235642

Si vede come una  
quantità (il  
numero di auto  
vendute) è  
ripartita tra  
modalità!!

Perché  
trasferibili?  
Quali sono i  
non  
trasferibili?

lez2\_2010-2011

statistica-francesco mola

3

## Distribuzioni ... (cont.)

### Distribuzioni di frequenza

Vale sia per caratteri quantitativi che qualitativi

#### Variabili discrete

$$\begin{aligned}n_1 + n_2 + \dots + n_i + \dots + n_k &= \\&= \sum_{i=1}^k n_i = n \quad 0 \leq n_i \leq n \\&\sum_{i=1}^k f_i = 1 \quad 0 \leq f_i \leq 1\end{aligned}$$

X	n <sub>i</sub>	f <sub>i</sub> =n <sub>i</sub> /n
x <sub>1</sub>	n <sub>1</sub>	f <sub>1</sub>
x <sub>2</sub>	n <sub>2</sub>	f <sub>2</sub>
•	•	•
x <sub>i</sub>	n <sub>i</sub>	f <sub>i</sub>
•	•	•
x <sub>k</sub>	n <sub>k</sub>	f <sub>k</sub>
n	1	

lez2\_2010-2011

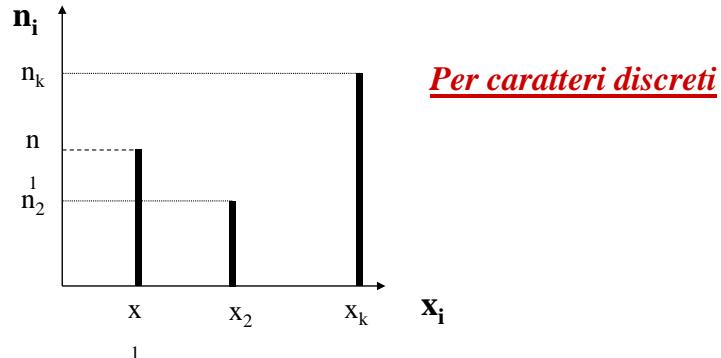
statistica-francesco mola

4

## Prime rappresentazioni grafiche

### Diagrammi a barre

*È una prima sintesi sulla forma della distribuzione*



lez2\_2010-2011

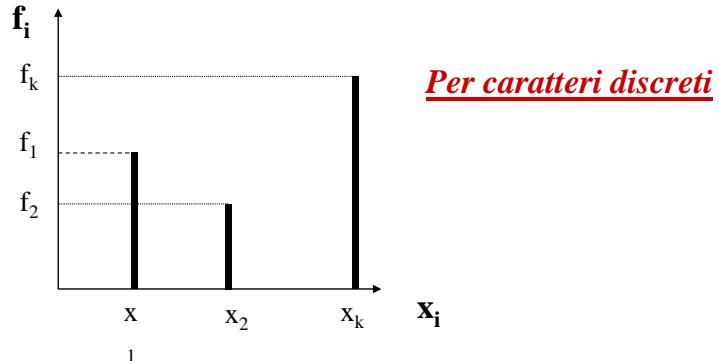
statistica-francesco mola

5

## Prime rappresentazioni grafiche

### Diagrammi a barre

*Utilizziamo questa volta le frequenze relative*

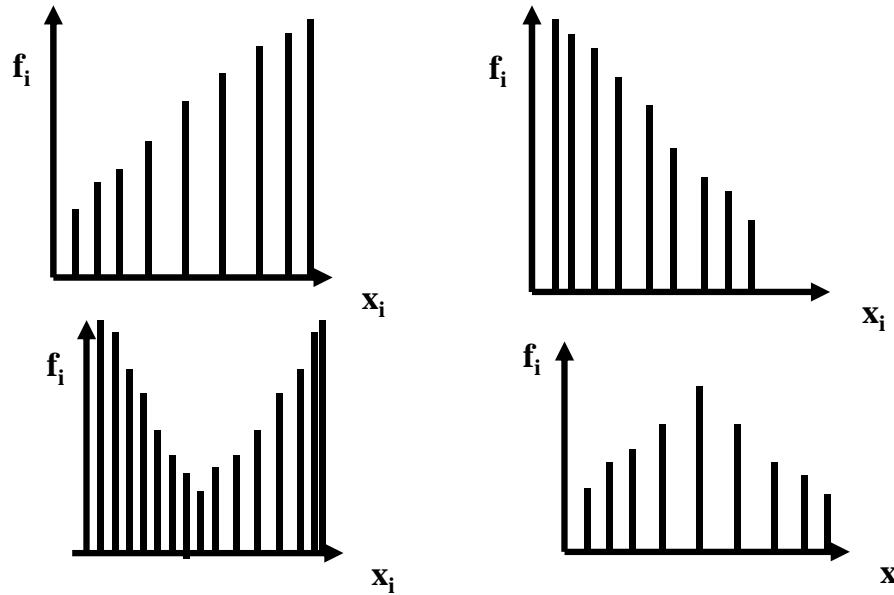


lez2\_2010-2011

statistica-francesco mola

6

### Alcuni esempi di grafici interessanti

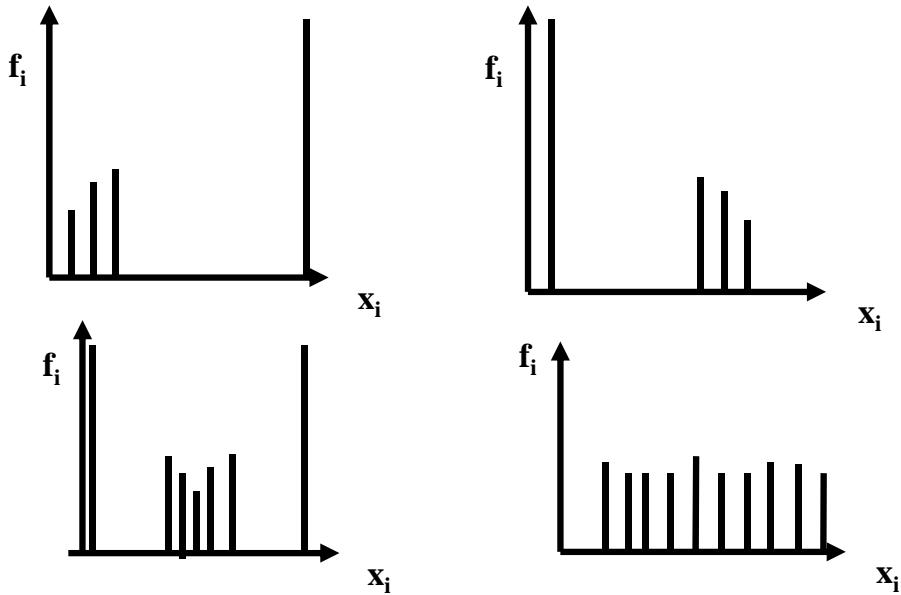


lez2\_2010-2011

statistica-francesco mola

7

### Alcuni esempi di grafici interessanti (cont.)



lez2\_2010-2011

statistica-francesco mola

8

## ...ricapitolando....

- I diagrammi a barre permettono di.....
  - Analizzare la “forma” di una distribuzione
  - Analizzare **picchi**
  - Analizzare **gap**
  - Visualizzare il **range** (minimo e massimo)
  - Sintetizzare la distribuzione

lez2\_2010-2011

statistica-francesco mola

9

## Distribuzioni



### Distribuzioni di frequenza

In genere prevede la suddivisione in classi (accorpamento)

X	$n_i$	X	$n_i$	Permette una prima sintesi ed una migliore leggibilità!
$x_1$	1	$x_0-x_1$	$n_1$	
$x_2$	1	$x_1-x_2$	$n_2$	
•	•	•	•	
$x_i$	1	$x_{i-1}-x_i$	$n_i$	$x_{i-1} - x_i = \text{classe}$
•	•	•	•	$i = 1, 2, \dots, k$
$x_n$	1	$x_{k-1}-x_k$	$n_k$	$b_i = x_i - x_{i-1}$ base

lez2\_2010-2011

statistica-francesco mola

10

## Distribuzioni di frequenza

- Per frequenza di una modalità si intende il numero di volte che essa è stata osservata
  - Frequenze assolute
  - Frequenze relative
  - Frequenze assolute cumulate
  - Frequenze relative cumulate
  - Frequenze retro cumulate

lez2\_2010-2011

statistica-francesco mola

11

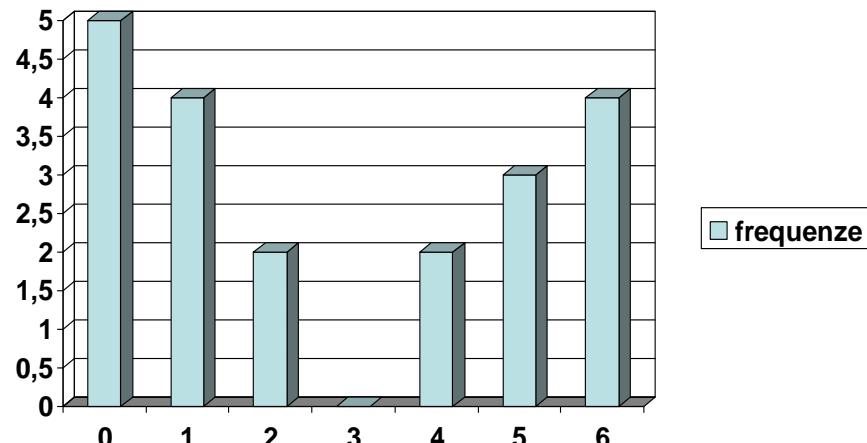
Puntate viste	frequenze	Cumulate	Relative (%)	Relative cumulate (%)
0	5	5	25	25
1	4	9	20	45
2	2	11	10	55
3	0	11	0	55
4	2	13	10	65
5	3	16	15	80
6	4	20	20	100
<b>Totale</b>	<b>20</b>		<b>100</b>	

lez2\_2010-2011

statistica-francesco mola

12

## Diagramma a barre

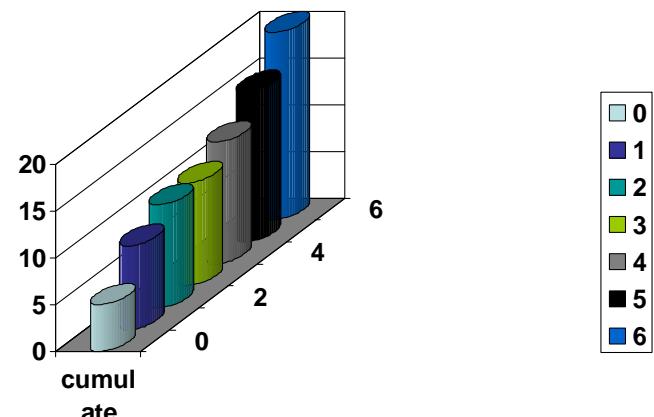


lez2\_2010-2011

statistica-francesco mola

13

## Diagramma a barre con frequenze cumulate



lez2\_2010-2011

statistica-francesco mola

14

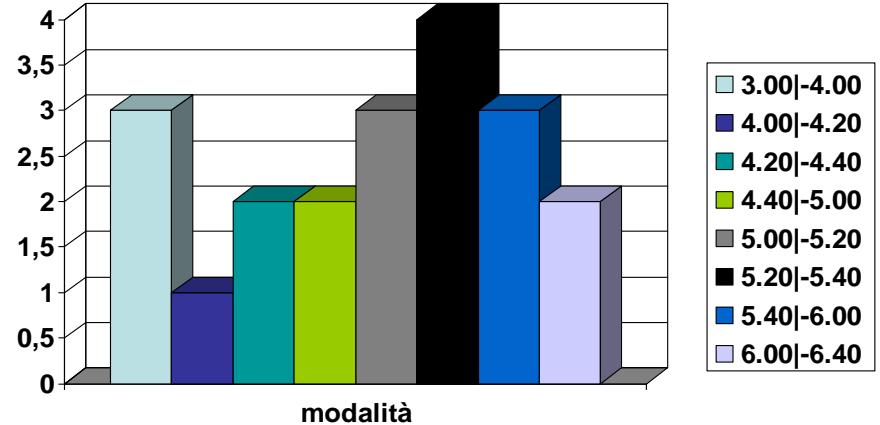
Tempo di accensione	frequenze	Cumulate	Relative (%)	Relative cumulate (%)
3.00 4.00	3	3	15	15
4.00 4.20	1	4	5	20
4.20 4.40	2	6	10	30
4.40 5.00	2	8	10	40
5.00 5.20	3	11	15	55
5.20 5.40	4	15	20	75
5.40 6.00	3	18	15	90
6.00 6.40	2	20	10	100
<b>Totale</b>	<b>20</b>		<b>100</b>	

lez2\_2010-2011

statistica-francesco mola

15

## Istogramma frequenze



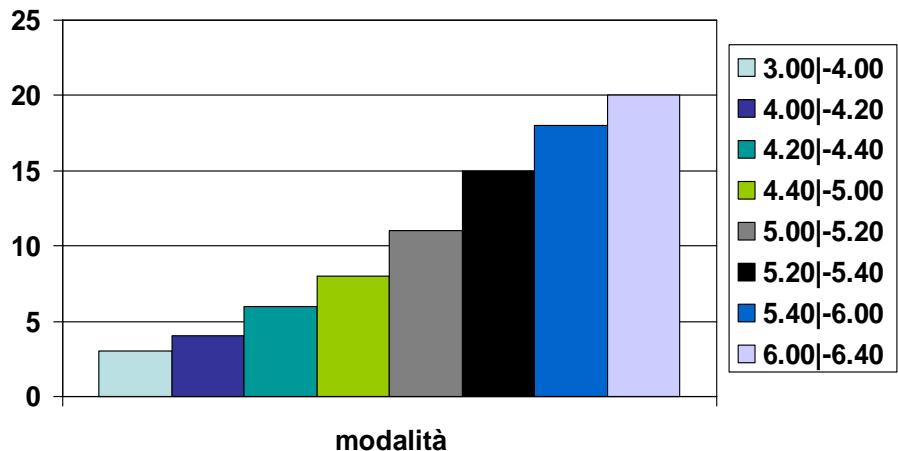
In questa rappresentazione grafica c'è un errore: qual è?

lez2\_2010-2011

statistica-francesco mola

16

# Istogramma frequenze cumulate



In questa rappresentazione grafica c'è un errore: qual è?

lez2\_2010-2011

statistica-francesco mola

17

# Riprendiamo l'esempio del tabacco

- Come posso rappresentare graficamente i dati?
- Esiste un sol modo?
- Qual è quello più efficace?

lez2\_2010-2011

statistica-francesco mola

18

## Popolazione di 25 marche di sigarette

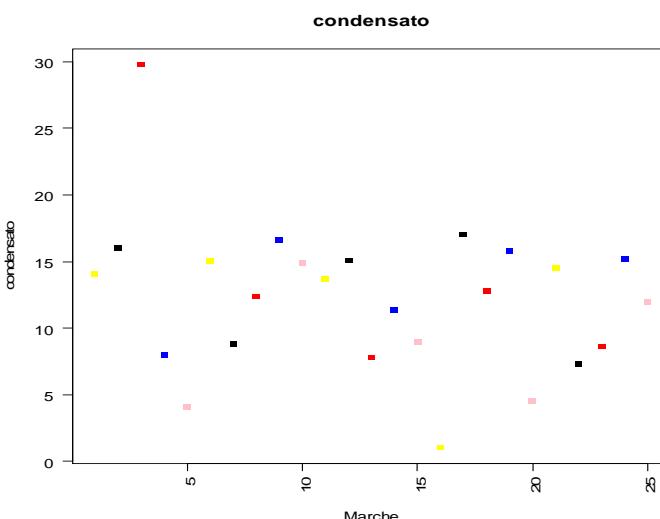
MARCA	condensato	peso	monossido	prezzo
Alpine		14.1	0.86	13.6
Benson&Hedges		16	1.06	16.6
BullDurham		29.8	2.03	23.5
CamelLights		8	0.67	10.2
Carlton		4.1	0.4	5.4
Chesterfield		15	1.04	15
Goldentlights		8.8	0.76	9
Kent		12.4	0.95	12.3
Kool		16.6	1.12	16.3
L&M		14.9	1.02	15.4
LarkLights		13.7	1.01	13
Marlboro		15.1	0.9	14.4
Merit		7.8	0.57	10
MultiFilter		11.4	0.78	10.2
NewportLights		9	0.74	9.5
Now		1	0.13	1.5
OldGold		17	1.26	18.5
PallMallLight		12.8	1.08	12.6
Raleigh		15.8	0.96	17.5
SalemUltra		4.5	0.42	4.9
Tareyton		14.5	1.01	15.9
True		7.3	0.61	8.5
ViceroyRichLight		8.6	0.69	10.6
VirginiaSlims		15.2	1.02	13.9
WinstonLights		12	0.82	14.9

lez2\_2010-2011

statistica-francesco mola

19

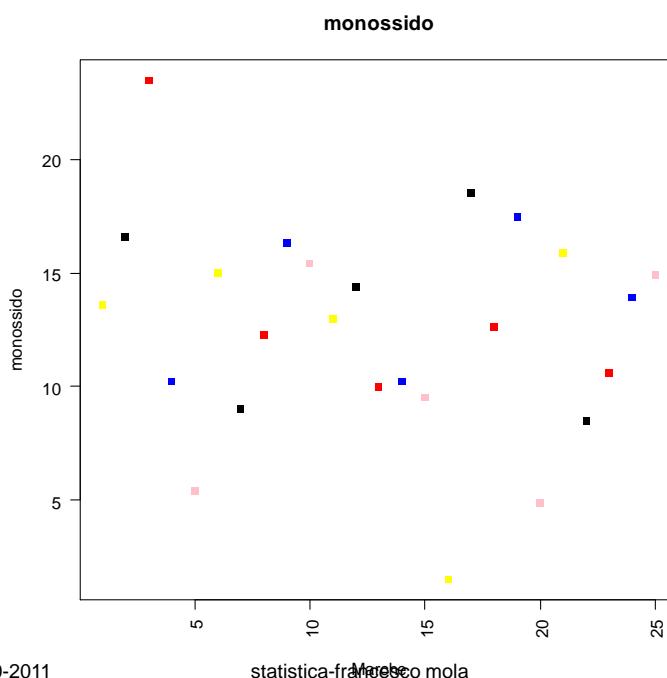
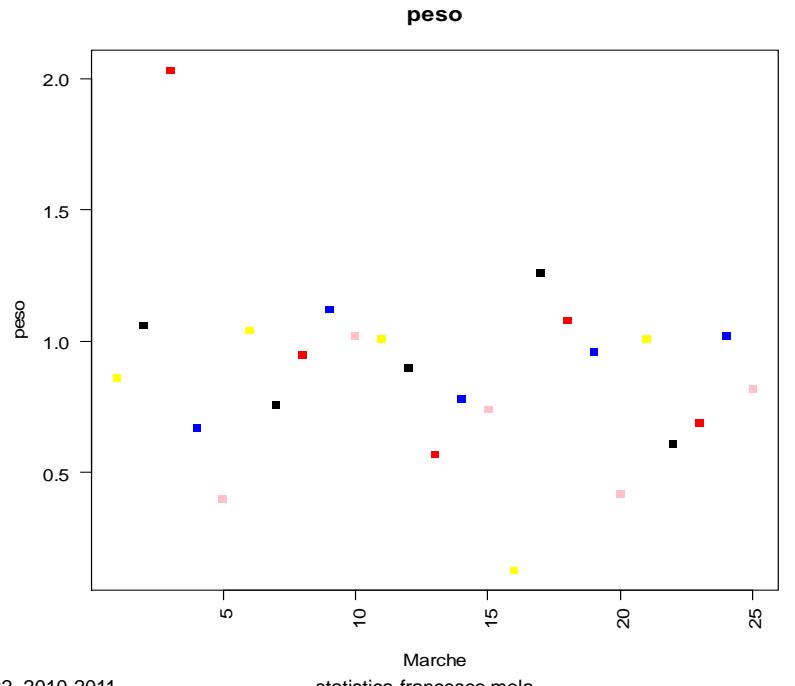
## Un semplice plot



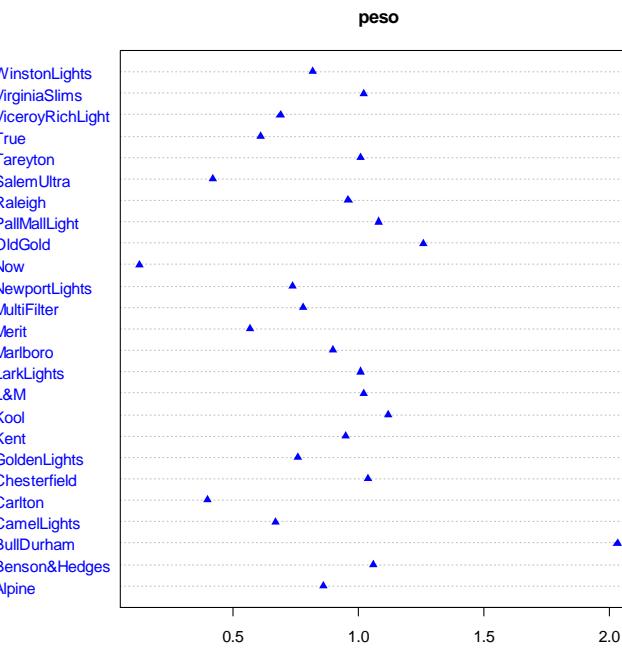
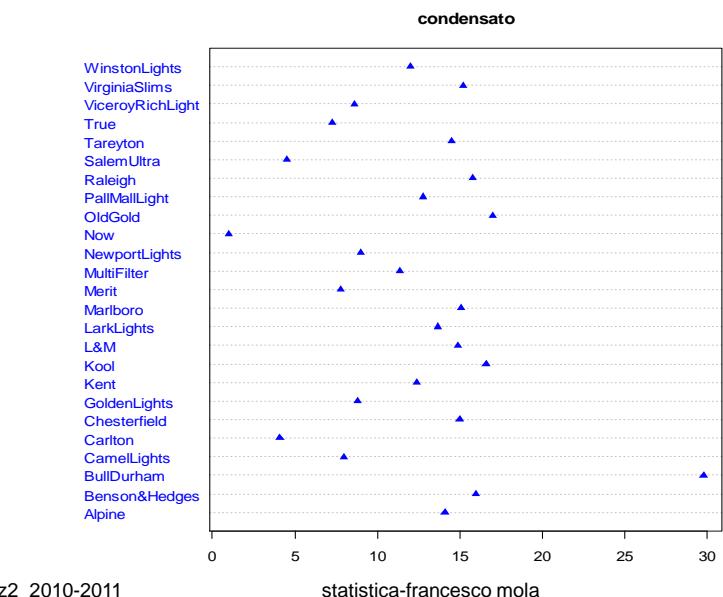
lez2\_2010-2011

statistica-francesco mola

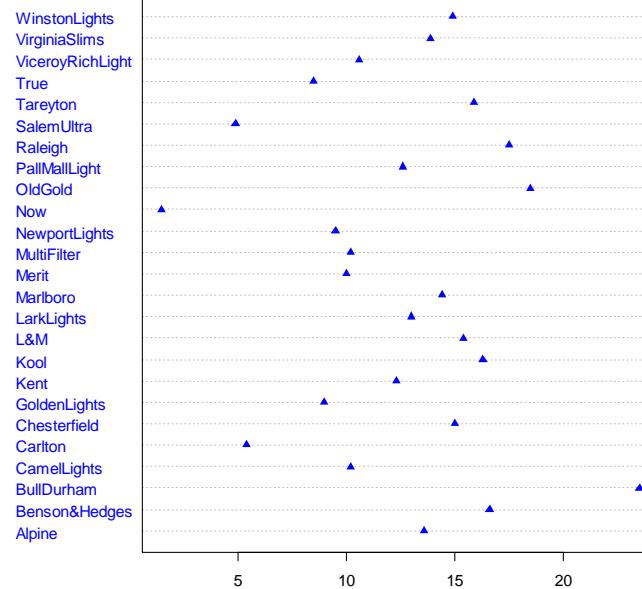
20



## Un modo alternativo



monossido



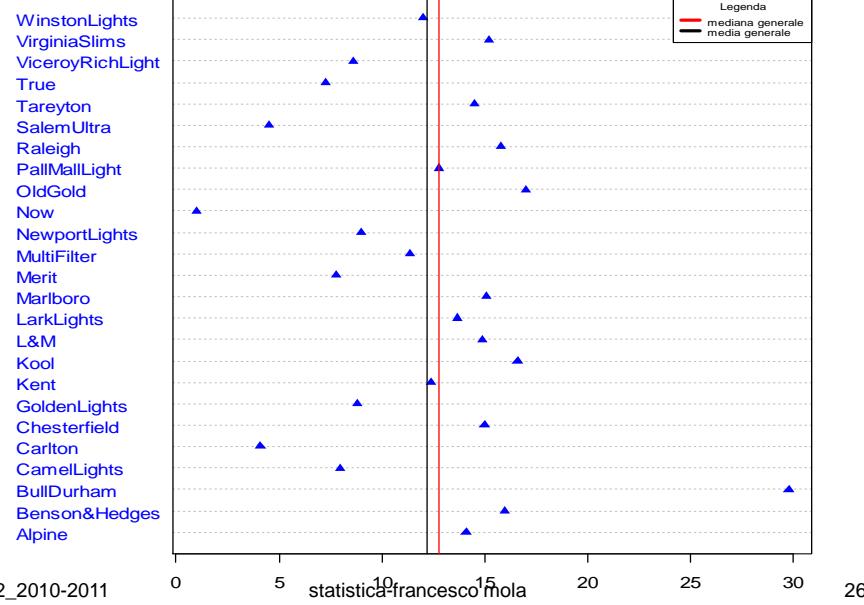
lez2\_2010-2011

statistica-francesco mola

25

## Ancora più informazioni

condensato

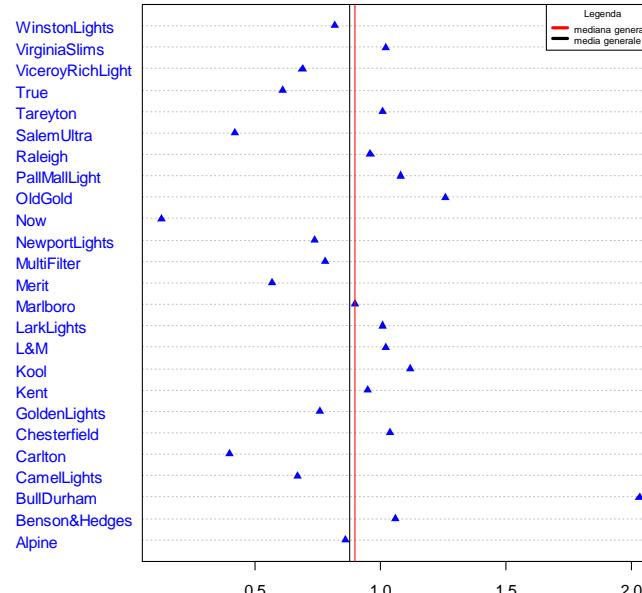


lez2\_2010-2011

statistica-francesco mola

26

peso

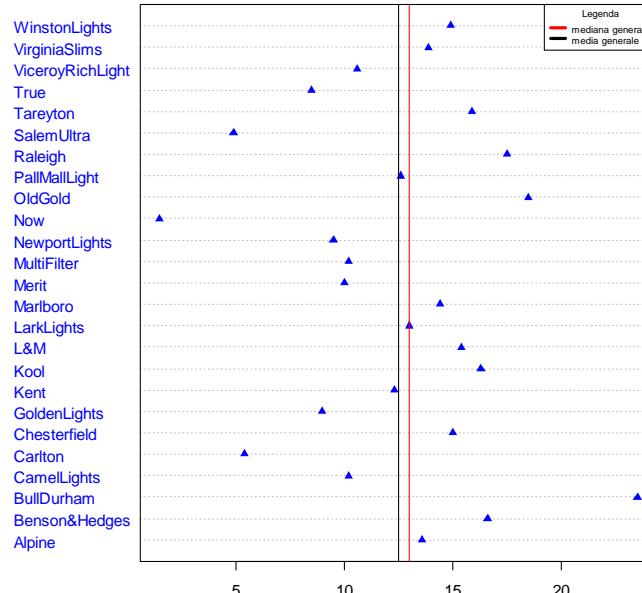


lez2\_2010-2011

statistica-francesco mola

27

monossido



lez2\_2010-2011

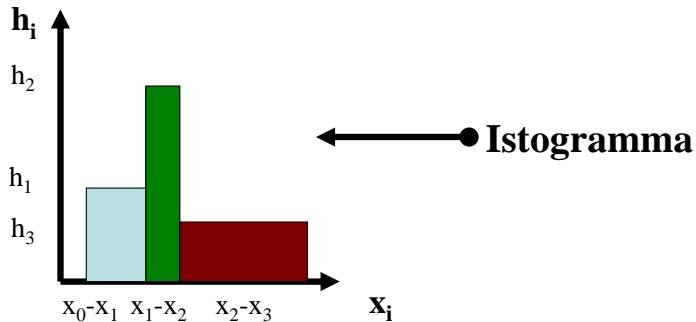
statistica-francesco mola

28

## Costruzione di un istogramma

$$b_i = x_i - x_{i-1} \Rightarrow \text{base}$$

$$h_i = \frac{n_i}{b_i} \Rightarrow \text{densità di frequenza}$$



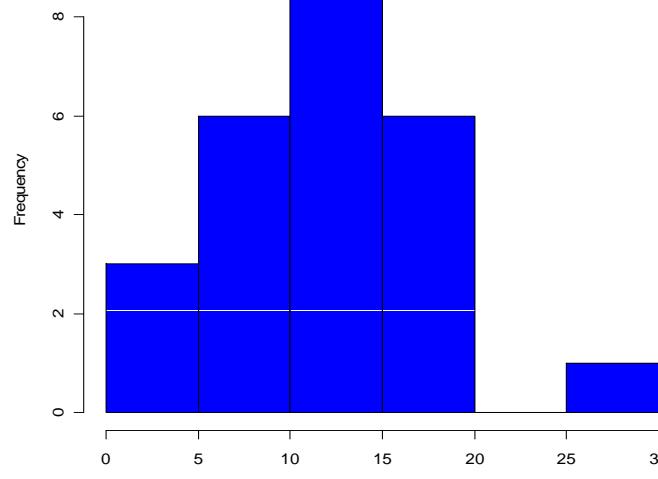
lez2\_2010-2011

statistica-francesco mola

29

## Iistogramma di frequenze

condensato

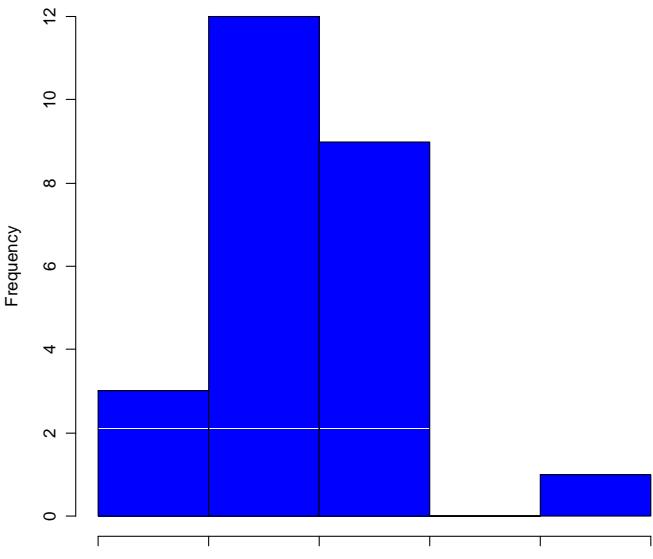


lez2\_2010-2011

statistica-francesco mola

30

**peso**

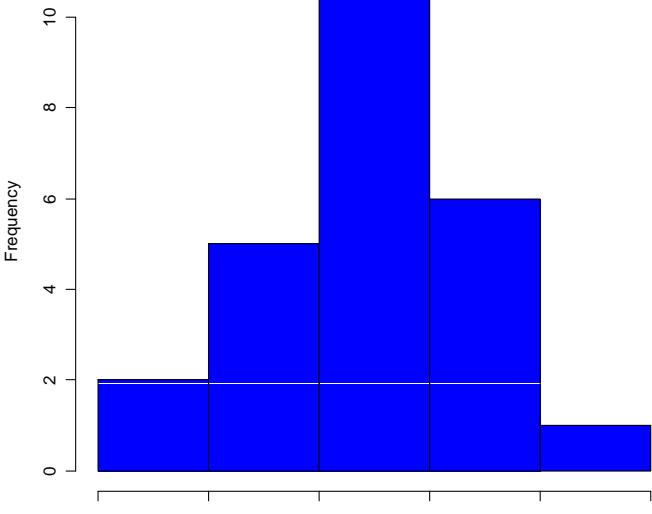


lez2\_2010-2011

statistica-francesco mola

31

**monossido**



lez2\_2010-2011

statistica-francesco mola

32

# Il problema dell'ampiezza delle classi

- Conoscenza del fenomeno
- Analisi grafica della distribuzione
- Classi equi ampie
- Classi equi frequenti

lez2\_2010-2011

statistica-francesco mola

33

## Classi equi frequenti

L'ampiezza di ogni classe

deve essere tale da contenere  $\frac{n}{k}$  frequenze

$k = \# \text{classi}$  (è scelto a priori!)

Esempio:

$$n = 600 \quad k = 6$$

$$\frac{n}{k} = \frac{600}{6} = 100$$

**Si tiene conto delle frequenze e NON delle  $x_i$ !!!**

**Ogni classe è di frequenza 100!**

X	$n_i$
$x_0-x_1$	100
$x_1-x_2$	100
$x_2-x_3$	100
$x_3-x_4$	100
$x_4-x_5$	100
$x_5-x_6$	100

lez2\_2010-2011

statistica-francesco mola

35

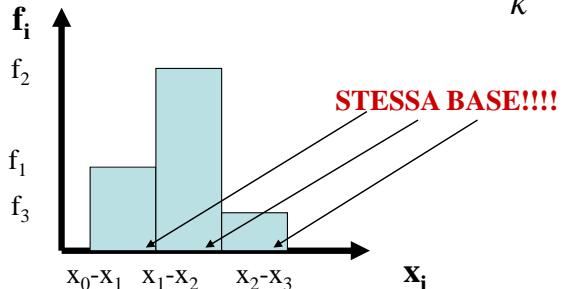
## Classi equi ampie

$$\frac{x_{\max} - x_{\min}}{k} \Rightarrow \text{ampiezza di ogni classe}$$

$k = \# \text{classi}$  (è scelto a priori!)

Esempio:

$$x_{\max} = 100 \quad x_{\min} = 20 \quad k = 4 \quad \frac{x_{\max} - x_{\min}}{k} = \frac{100 - 20}{4} = 20$$



**Ogni classe è di ampiezza 20!**

**Si tiene conto delle  $x_i$  ma NON delle frequenze!!!**

lez2\_2010-2011

statistica-francesco mola

34

Analizziamo graficamente sia con diagrammi a barre (*bar charts*) che con diagrammi a torta (*pie charts*) le tre variabili (condensato, peso, monossido) raggruppate in 5 classi.

Le rappresentazioni che seguono, sono affette da errori; il software statistico consente di rappresentare i dati in questo modo. E' compito dello "statistico" far sì che questi dati siano rappresentati in maniera corretta.

lez2\_2010-2011

statistica-francesco mola

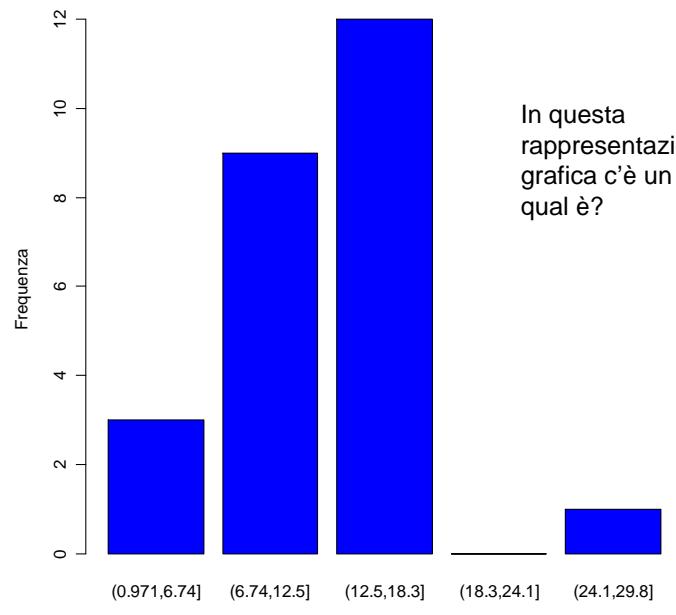
36

lez2\_2010-2011

statistica-francesco mola

36

diagramma a barre per la distribuzione in classi di condensato



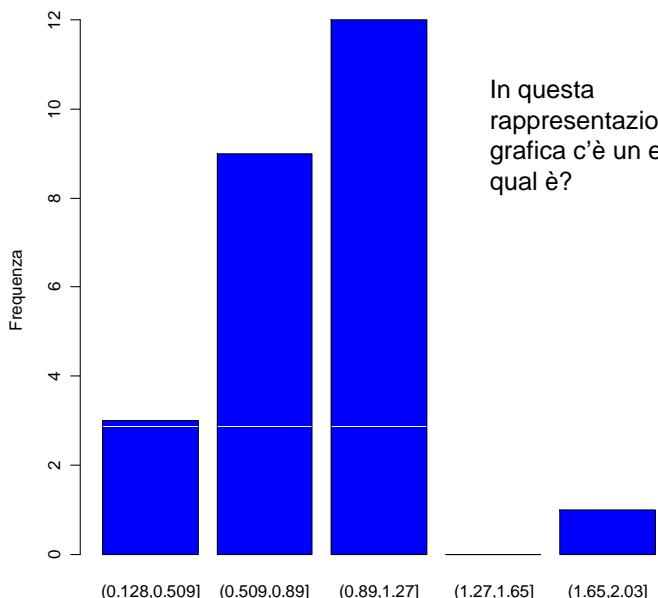
In questa rappresentazione grafica c'è un errore: qual è?

lez2\_2010-2011

statistica-francesco mola

37

diagramma a barre per la distribuzione in classi di peso



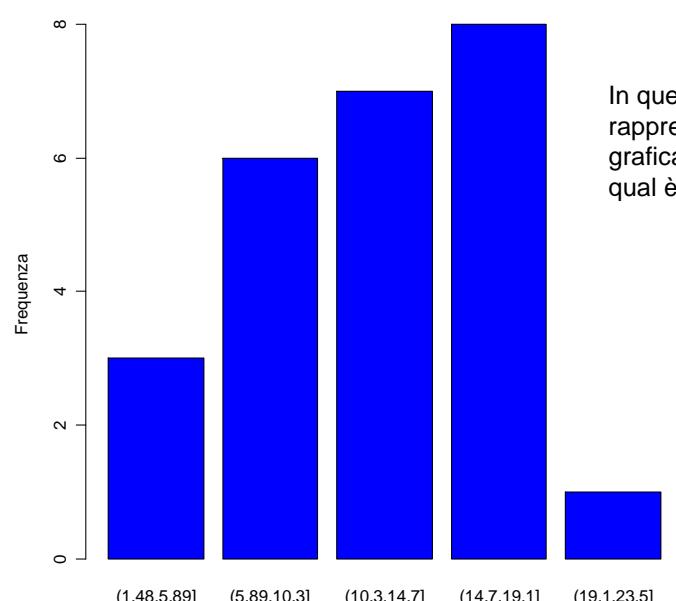
In questa rappresentazione grafica c'è un errore: qual è?

lez2\_2010-2011

statistica-francesco mola

38

diagramma a barre per la distribuzione in classi di monossido



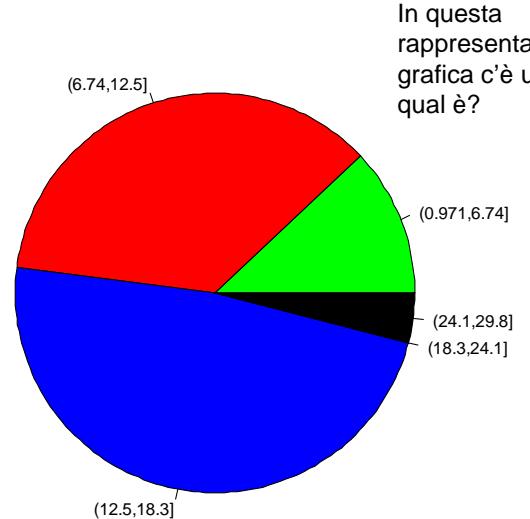
In questa rappresentazione grafica c'è un errore: qual è?

lez2\_2010-2011

statistica-francesco mola

39

pie chart di condensato

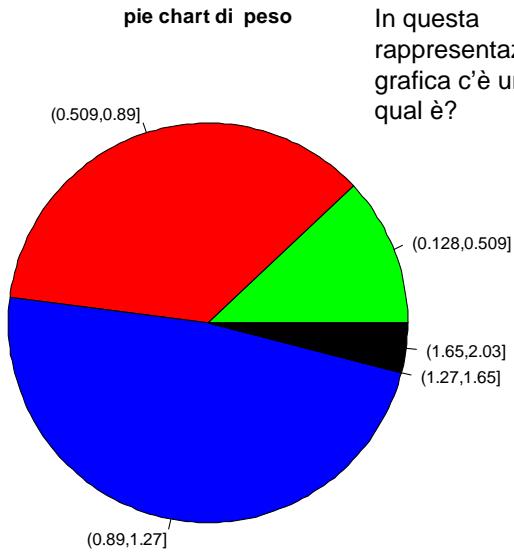


In questa rappresentazione grafica c'è un errore: qual è?

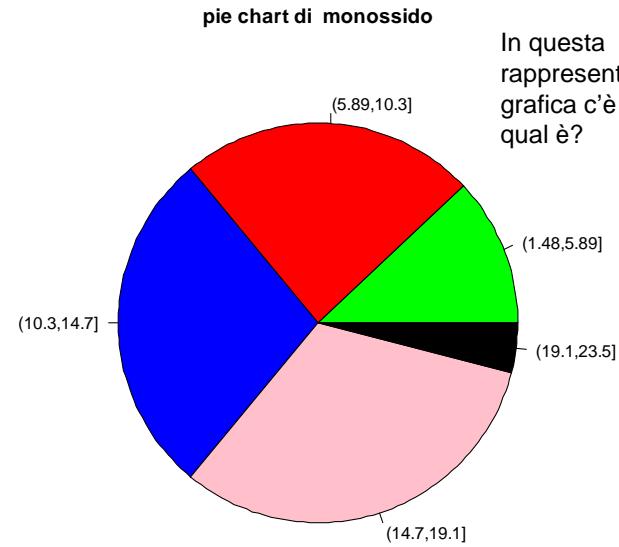
lez2\_2010-2011

statistica-francesco mola

40



In questa rappresentazione grafica c'è un errore: qual è?



In questa rappresentazione grafica c'è un errore: qual è?

## Funzione di ripartizione empirica

Consideriamo la variabile X:

$$X = x_1 \leq x_2 \leq x_3 \leq \dots \leq x_n$$

X	$n_i$	$f_i$	$F_i$
$x_1$	$n_1$	$f_1$	$F_1 = f_1$
$x_2$	$n_2$	$f_2$	$F_2 = f_1 + f_2$
•	•	•	•
$x_i$	$n_i$	$f_i$	$F_i = f_1 + f_2 + \dots + f_i$
•	•	•	•
$x_n$	$n_n$	$f_n$	$F_n = f_1 + f_2 + \dots + f_i + \dots + f_n = 1$
	n	1	

## Funzione di ripartizione empirica ..(cont.)

$F_i$  = frequenza relativa cumulata

$$F_i = \sum_{h=1}^i f_h$$

$$F(x_0) = \frac{\#(X \leq x_0)}{n}$$

# Proprietà della funzione di ripartizione empirica

$0 \leq F(x) \leq 1$

$F(x)$  è non decrescente

$F(-\infty) = 0 \quad F(+\infty) = 1$

$F(x)$  è continua da destra

# Grafico della funzione

