

# Statistica Descrittiva I

## Serie statistiche monovariate

- Rappresentazioni tabellari
- Principali rappresentazioni grafiche
- Nozioni di Frequenza

1

## Organizzazione dei dati - I

Scuola di provenienza dei 99 studenti del corso

T L S T C S C S T C S L T S T L L S L S  
S T T L S C S L C T L L C T C T T S T S  
T L T L L L C T L L T L S S T T S L S L S  
S L T T C C T S L S C S T L C S T S L  
S C C L L S L T T T S S T T T S T L C S

L:liceo classico; S:liceo scientifico; T:tecnico ; C:commerciale

- Dato qualitativo non ordinabile
- $M = 4$ ;  $N = 99$
- Rappresentazione verbosa difficile per  $N$  elevata.

2

## Frequenze assolute

Raccolgo le  $M$  modalità (serie statistica)

| $i$ | $m_i$         | $n_i$     |
|-----|---------------|-----------|
| 1   | S             | 28        |
| 2   | L             | 26        |
| 3   | T             | 30        |
| 4   | C             | 15        |
|     |               |           |
|     | <b>totale</b> | <b>99</b> |

(tabella ad entrata semplice)

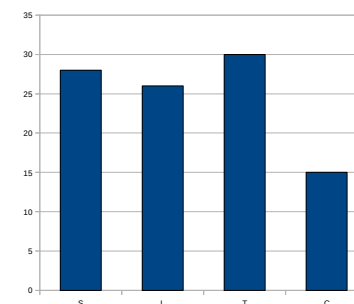
- **Frequenze assolute  $n_i$** : osservazioni della modalità  $m_i$ .
- Se ogni unità fornisce un'osservazione, si ha che

$$\sum_{i=1}^M n_i = N$$

3

## Diagramma a barre

| $i$ | $m_i$         | $n_i$     |
|-----|---------------|-----------|
| 1   | S             | 28        |
| 2   | L             | 26        |
| 3   | T             | 30        |
| 4   | C             | 15        |
|     |               |           |
|     | <b>totale</b> | <b>99</b> |



- La posizione, l'ordine e la larghezza delle colonne è irrilevante
- Ordinata legata alla dimensione  $N$

4

## Frequenze relative

Evitare la dipendenza da  $N$

| $i$ | $m_i$  | $n_i$ | $f_i$ |
|-----|--------|-------|-------|
| 1   | S      | 28    | 0,283 |
| 2   | L      | 26    | 0,263 |
| 3   | T      | 30    | 0,303 |
| 4   | C      | 15    | 0,152 |
|     |        |       |       |
|     | totale | 99    | 1     |

- **F. relative:** *Frazione delle osservazioni con modalità  $m_i$*

$$f_i = \frac{n_i}{N}$$

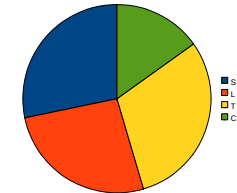
- Si ha che

$$\sum_{i=1}^M f_i = 1$$

5

## Diagramma a torta

| $i$ | $m_i$  | $n_i$ | $f_i$ |
|-----|--------|-------|-------|
| 1   | S      | 28    | 0,283 |
| 2   | L      | 26    | 0,263 |
| 3   | T      | 30    | 0,303 |
| 4   | C      | 15    | 0,152 |
|     |        |       |       |
|     | totale | 99    | 1     |



- Ampiezza del settore della  $i$ -sima modalità proporzionale alla frequenza relativa

$$\varphi_i = 360 f_i = 360 \frac{n_i}{N}$$

6

## Organizzazione dei dati - II

Carattere: Gradimento del corso di Matematica

Popolazione: 77 studenti di Biotechnologie

M S I I M S M A I I M S S S I A S I S S  
M I M S M I S S I M S A S M A S S M S  
M I I A M M S M S M S S S M M M S I M  
A A S M S S S M S M M I I S M S M S I

A: alto M: medio S: sufficiente I: insufficiente

- Dato qualitativo ordinabile

$$A > M > S > I$$

7

## Rango

Tabella a entrata semplice

- Rango: intero usato per ordinare le modalità

| $r_i$ | $m_i$  | $n_i$ | $f_i$ |
|-------|--------|-------|-------|
| 1     | I      | 15    | 0,195 |
| 2     | S      | 30    | 0,390 |
| 3     | M      | 25    | 0,325 |
| 4     | A      | 7     | 0,091 |
|       |        |       |       |
|       | totale | 77    | 1     |

- Diagrammi possibili
  - Diagramma a barre (l'ordine delle modalità è dato dal rango)
  - Torta. Nessuna modifica

8

## Organizzazione dei dati - III

Carattere: Cellulari posseduti

Popolazione: 78 studenti di Biotecnologie

1 2 2 1 1 1 2 4 1 2 1 2 3 1 1 4 3 4 1  
0 0 1 4 2 2 0 1 0 2 2 2 2 2 1 1 2 0 2  
2 1 4 1 3 1 1 0 2 1 1 2 1 1 1 1 2 0 3  
1 2 2 1 2 0 1 1 2 1 1 2 2 2 4 1 3 2

- Dato quantitativo discreto

9

## Rappresentazioni

- Tabella singola entrata

| $i$ | $m_i$         | $n_i$ | $f_i$ |  |
|-----|---------------|-------|-------|--|
| 1   | 0             | 8     | 0,107 |  |
| 2   | 1             | 30    | 0,400 |  |
| 3   | 2             | 26    | 0,347 |  |
| 4   | 3             | 5     | 0,067 |  |
| 5   | 4             | 6     | 0,080 |  |
|     |               |       |       |  |
|     | <b>Totali</b> | 75    | 1     |  |

- Rappresentazioni grafiche
  - Digramma a barre: ordine importante
  - Torta: non più utilissimo ma possibile

10

## Frequenze Cumulate

Torna utile la seguente definizione

| $i$ | $m_i$         | $n_i$ | $f_i$ | $F_i$ |
|-----|---------------|-------|-------|-------|
| 1   | 0             | 8     | 0,107 | 0,107 |
| 2   | 1             | 30    | 0,400 | 0,507 |
| 3   | 2             | 26    | 0,347 | 0,853 |
| 4   | 3             | 5     | 0,067 | 0,920 |
| 5   | 4             | 6     | 0,080 | 1,000 |
|     |               |       |       |       |
|     | <b>Totali</b> | 75    | 1     |       |

- **F. cumulate:** osservazioni minori o pari ad  $m_i$ ,  
(non più di..)

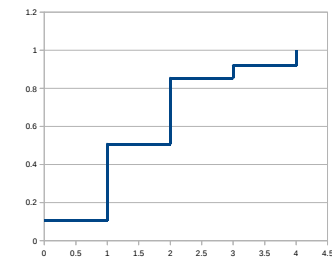
$$F_i = \sum_{j=1}^M f_j$$

- Si ha che  $F_M = 1$

11

## Funzione di ripartizione (empirica)

| $i$ | $m_i$         | $n_i$ | $f_i$ | $F_i$ |
|-----|---------------|-------|-------|-------|
| 1   | 0             | 8     | 0,107 | 0,107 |
| 2   | 1             | 30    | 0,400 | 0,507 |
| 3   | 2             | 26    | 0,347 | 0,853 |
| 4   | 3             | 5     | 0,067 | 0,920 |
| 5   | 4             | 6     | 0,080 | 1,000 |
|     |               |       |       |       |
|     | <b>Totali</b> | 75    | 1     |       |



- Grafico a gradini

12

## Organizzazione dei dati - IV

Carattere: Altezza

Popolazione: 50 studenti di Biotechnologie

1.652 1.554 1.636 1.531 1.658 1.523 1.708 1.674 1.704  
1.554 1.679 1.486 1.583 1.521 1.802 1.702 1.471 1.696  
1.599 1.711 1.768 1.639 1.561 1.621 1.694 1.644 1.732  
1.711 1.647 1.645 1.624 1.561 1.601 1.772 1.581 1.603  
1.480 1.626 1.663 1.812 1.848 1.590 1.621 1.794 1.408  
1.673 1.506 1.447 1.586 1.631

- Dato quantitativo continuo
- Definizione legata al carattere e non all'unità di misura: <sub>13</sub> sarebbe stato continuo anche se misurato in mm!

## Classi di modalità

- Impossibile elencare le singole modalità.
- Introduco delle classi  $c_i$

| $c_i$ | $inf_i$ | $sup_i$       | $n_i$ | $f_i$ | $F_i$ |
|-------|---------|---------------|-------|-------|-------|
| 1     | 1,40    | 1,45          | 2     | 0,04  | 0,04  |
| 2     | 1,45    | 1,50          | 3     | 0,06  | 0,10  |
| 3     | 1,50    | 1,55          | 4     | 0,08  | 0,18  |
| 4     | 1,55    | 1,60          | 9     | 0,18  | 0,36  |
| 5     | 1,60    | 1,65          | 12    | 0,24  | 0,6   |
| 6     | 1,65    | 1,70          | 8     | 0,16  | 0,76  |
| 7     | 1,70    | 1,75          | 6     | 0,12  | 0,88  |
| 8     | 1,75    | 1,80          | 3     | 0,06  | 0,94  |
| 9     | 1,80    | 1,85          | 3     | 0,06  | 1     |
|       |         | <b>totale</b> | 50    | 1     |       |

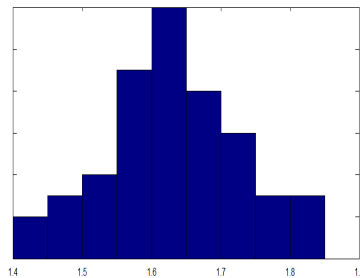
- un'osservazione  $o$  appartiene alla  $i$ -sima classe se  $inf_i < o \leq sup_i$

14

## Istogramma

- Simile al grafico a barre: un rettangolo per ogni modalità
  - Base: estremi della classe
  - Altezza: densità di frequenza  $\frac{f_i}{sup_i - inf_i}$
  - Area unitaria

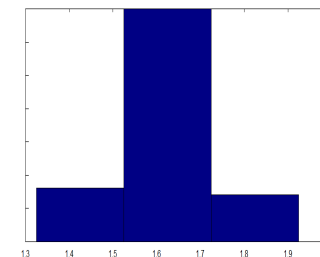
| $c_i$ | $inf_i$ | $sup_i$       | $n_i$ | $f_i$ | $d_i$ |
|-------|---------|---------------|-------|-------|-------|
| 1     | 1,40    | 1,45          | 2     | 0,04  | 0,80  |
| 2     | 1,45    | 1,50          | 3     | 0,06  | 1,20  |
| 3     | 1,50    | 1,55          | 4     | 0,08  | 1,60  |
| 4     | 1,55    | 1,60          | 9     | 0,18  | 3,60  |
| 5     | 1,60    | 1,65          | 12    | 0,24  | 4,80  |
| 6     | 1,65    | 1,70          | 8     | 0,16  | 3,20  |
| 7     | 1,70    | 1,75          | 6     | 0,12  | 2,40  |
| 8     | 1,75    | 1,80          | 3     | 0,06  | 1,20  |
| 9     | 1,80    | 1,85          | 3     | 0,06  | 1,20  |
|       |         | <b>totale</b> | 50    | 1     |       |



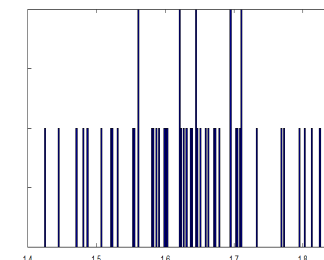
## Classi di modalità - II

- Quante classi fare ?

– Poche



– Troppe



- Prima regola empirica

$$1 + \log_2(N)$$

16

## Classi di modalità - III

Che ampiezza dare alle classi?

- Uniforme

| $c_i$ | $inf_i$ | $sup_i$       | $n_i$ | $f_i$ | $d_i$ |
|-------|---------|---------------|-------|-------|-------|
| 1     | 1,40    | 1,45          | 2     | 0,04  | 0,80  |
| 2     | 1,45    | 1,50          | 3     | 0,06  | 1,20  |
| 3     | 1,50    | 1,55          | 4     | 0,08  | 1,60  |
| 4     | 1,55    | 1,60          | 9     | 0,18  | 3,60  |
| 5     | 1,60    | 1,65          | 12    | 0,24  | 4,80  |
| 6     | 1,65    | 1,70          | 8     | 0,16  | 3,20  |
| 7     | 1,70    | 1,75          | 6     | 0,12  | 2,40  |
| 8     | 1,75    | 1,80          | 3     | 0,06  | 1,20  |
| 9     | 1,80    | 1,85          | 3     | 0,06  | 1,20  |
|       |         | <b>totale</b> | 50    | 1     |       |

- Non uniforme

| $c_i$ | $inf_i$ | $sup_i$       | $n_i$ | $f_i$ | $d_i$ |
|-------|---------|---------------|-------|-------|-------|
|       |         |               |       |       |       |
| 1     | 1,40    | 1,50          | 5     | 0,10  | 1,00  |
| 2     | 1,50    | 1,55          | 4     | 0,08  | 1,60  |
| 3     | 1,55    | 1,60          | 9     | 0,18  | 3,60  |
| 4     | 1,60    | 1,65          | 12    | 0,24  | 4,80  |
| 5     | 1,65    | 1,70          | 8     | 0,16  | 3,20  |
| 6     | 1,70    | 1,75          | 6     | 0,12  | 2,40  |
| 7     | 1,75    | 1,85          | 6     | 0,12  | 1,20  |
|       |         |               |       |       |       |
|       |         | <b>totale</b> | 50    | 1     |       |

Non esiste una regola fissa. Dipende dalla sensibilità

17

## Ricapitolando

- Dato Qualitativo non ordinabile
  - Frequenza assoluta, relativa
  - Diagramma a barre, torta
- Dato Qualitativo ordinabile
  - Frequenza assoluta, relativa, rango
  - Diagramma a barre (ordinato), Torta
- Dato Quantitativo discreto
  - Frequenza assoluta, relativa, cumulata
  - Diagramma a barre, Torta, Funzione di ripartizione
- Dato Quantitativo continuo → classi
  - Frequenza assoluta, relativa, cumulata
  - Torta, Funzione di ripartizione, Istogramma

18