

Estimatori statistici e istogrammi

Statistica 1

- Mediante una quantità limitata di informazioni aiuta a dedurre una descrizione semplice di un insieme di unità o “universo”.
- Interpretazione dei caratteri osservati per dedurre caratteristiche generali dell'insieme osservato.
- L'intero insieme di cui si vogliono dedurre le caratteristiche si chiama “popolazione”, mentre l'insieme delle nostre misure, che ne è un sottoinsieme, si chiama “campione”

Campioni e popolazioni

- Per esempio vogliamo determinare il peso medio di tutti gli scorfani del mediterraneo:
 - la popolazione è costituita da TUTTI gli scorfani del mediterraneo
 - Il campione è costituito dagli scorfani che si riescono a pescare per poi pesarli (e farci una zuppa...)
- L'analisi statistica permette di trarre conclusioni sulle caratteristiche della popolazione sulla base delle caratteristiche del campione.

Esempio

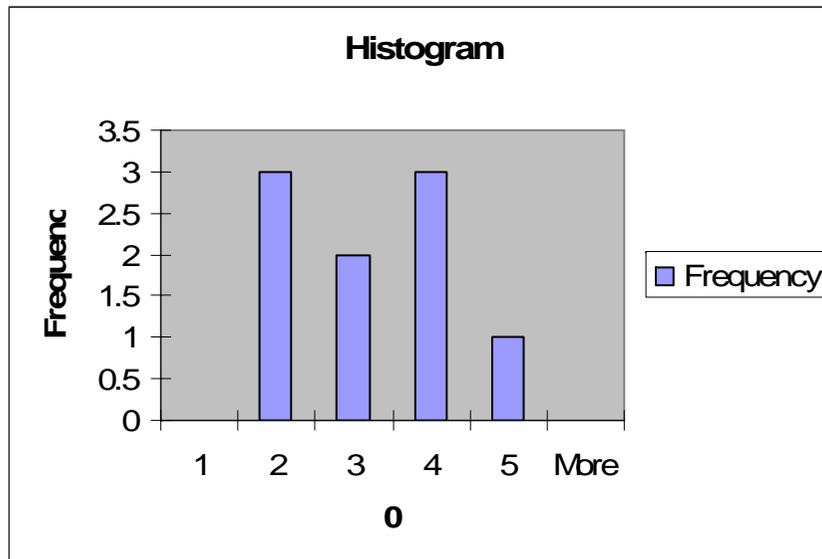
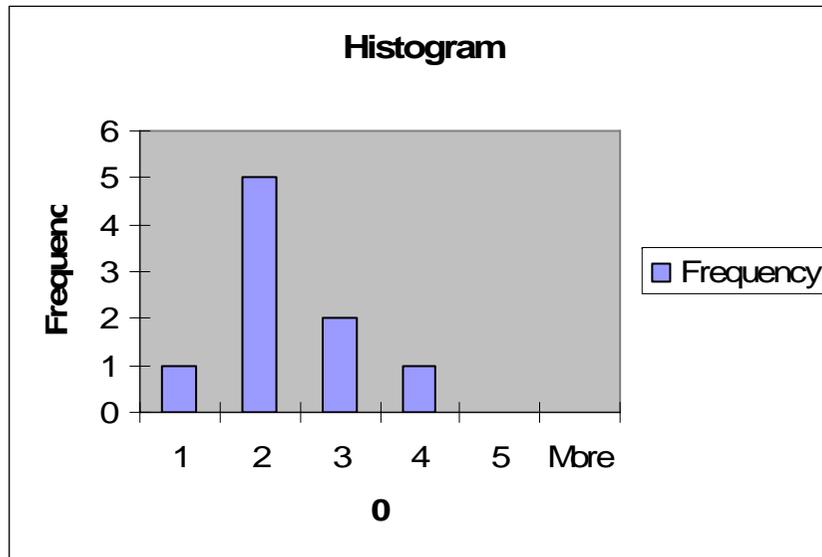
- Supponiamo di aver effettuato una serie di misure di succinato di deidrogenasi nelle cellule per determinare la frazione nucleare e quella mitocondriale ed avere ottenuto i risultati in tabella
- I valori hanno variazioni praticamente continue e questa è una caratteristica dei dati di misura

CAMP.	CONC. DI SUCCINATO DEIDROGENASI FRAZ. NUC (mg/ml)	CONC. DI SUCCINATO DEIDROGENASI FRAZ. MIT (mg/ml)
1	1.23	2.37
2	1.87	3.45
3	2.04	1.91
4	1.56	4.02
5	1.21	1.42
6	2.67	3.78
7	1.58	2.51
8	3.09	3.13
9	0.74	2.85
10	1.46	1.98

Organizzazione dei dati

- Graficamente possiamo organizzare i dati in modo da avere informazioni anche “a colpo d’occhio”
- Possiamo fare per esempio un istogramma dei dati, ovvero, suddividere l’intervallo di variazione dei dati in tanti intervallini e contare il numero di dati il cui valore cade in ciascun intervallino. L’istogramma sarà una rappresentazione grafica di queste “frequenze”.

Istogrammi



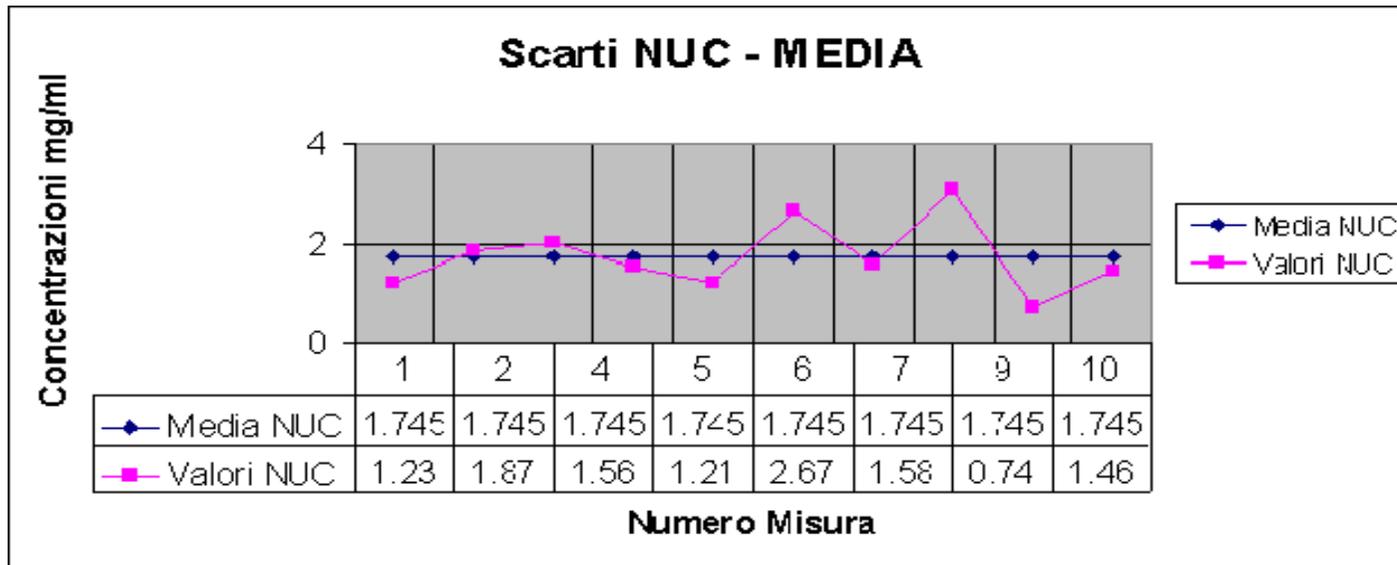
- Si vede che il primo (nucleare) è spostato più a sinistra del secondo (mit.) e che è più stretto
- Come dedurre queste caratteristiche senza ricorrere ai grafici?

Media e scarto

- Possiamo definire un estimatore della media della popolazione come la somma degli elementi del campione divisa per il numero di campioni

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

Se consideriamo gli scarti dalla media: $(x - x_m)$, la loro somma si può verificare che è nulla, perché rispetto alla media gli scarti sono in più o in meno



Media e scarto quadratico

- Dunque gli scarti non ci danno informazioni. Dobbiamo ricorrere alla somma del quadrato degli scarti, che essendo tutti positivi ci danno l'informazione di quanto i punti si disperdano attorno alla media.
- Definiamo dunque lo Scarto Quadratico Medio come:

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

...che definisce il quadrato della “varianza” del campione.

Il termine a denominatore N-1, deriva dal fatto che l'ultimo scarto è determinato univocamente dagli N-1 precedenti per la condizione che la somma deve essere nulla.

La varianza della popolazione non ha il -1.

Deviazione Standard

- La deviazione standard (Root Mean Square – RMS) è la radice dello scarto quadratico medio.
- Ci sono anche altri estimatori che ci danno informazioni, per esempio su quanto sia asimmetrica la distribuzione dei dati rispetto alla media (skewness – somma del cubo degli scarti divisa per la varianza)