

Analisi dei dati

Statistica Descrittiva

Angelica Lo Duca
angelica.loduca@iit.cnr.it

Obiettivo

L'obiettivo dell'analisi dei dati consiste nello scoprire **trend**, **pattern** e **relazioni** nascosti nei dati.



Tipo di analisi

Analisi Quantitativa

Analizza una grande quantità di dati al fine di estrarne il comportamento e l'andamento.

Analisi Qualitativa

Analizza una piccola quantità di dati al fine di estrarne alcuni dettagli.

Metodo di analisi

Manuale

E' molto preciso, ma richiede molto tempo. Inoltre può essere fatto solo su una piccola quantità di dati.

Semiautomatico

Utilizza degli strumenti (ad esempio basati su Visual Analytics). Necessita comunque della supervisione del data journalist.

Automatico

Basato sulla scrittura di codice. Permette di analizzare grandi quantità di dati, ma non è preciso come i due casi precedenti.

Tipi di **Statistica**

Descrittiva

Dato un insieme di dati (campione di dati), cerca di descrivere il campione stesso.

Inferenziale

Considera il campione di dati come un sottoinsieme di una popolazione. Cerca di capire il comportamento della popolazione a partire dal campione. L'obiettivo è di costruire modelli predittivi per capire l'andamento di un fenomeno.

Statistica descrittiva

Si basa sul calcolo di alcune **metriche** o **indici**

- Indici di frequenza
- Indici di tendenza centrale
- Indici di variabilità

Indici di Frequenza

*Descrivere una singola
variabile nel dataset*

1

COUNT

Data una variabile, contare quante
volte appare una certa categoria

2

PERCENTUALE

percentuale relativa al conteggio
precedente

Indici di frequenza

Presentazione grafica

grafici a barre, linee, diagrammi a torta, distribuzione della frequenza

Indici di Tendenza Centrale

*Descrivere i dati con un
solo valore*

1

MEDIA ARITMETICA

Somma dei dati

Numero dei dati

2

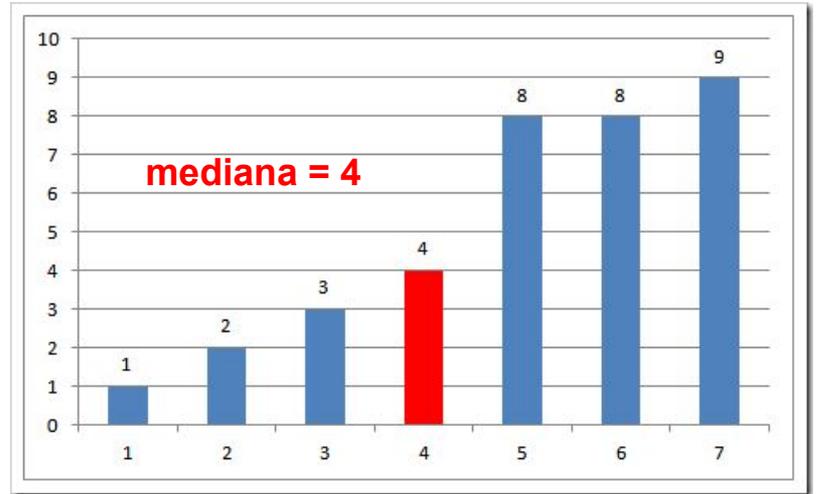
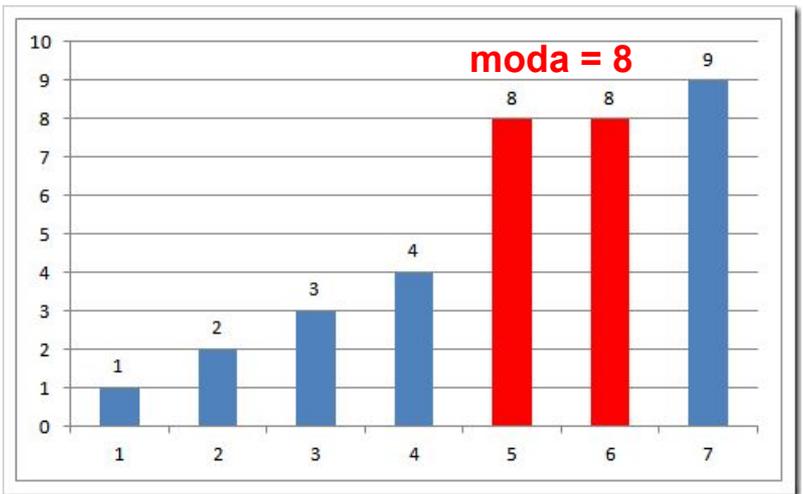
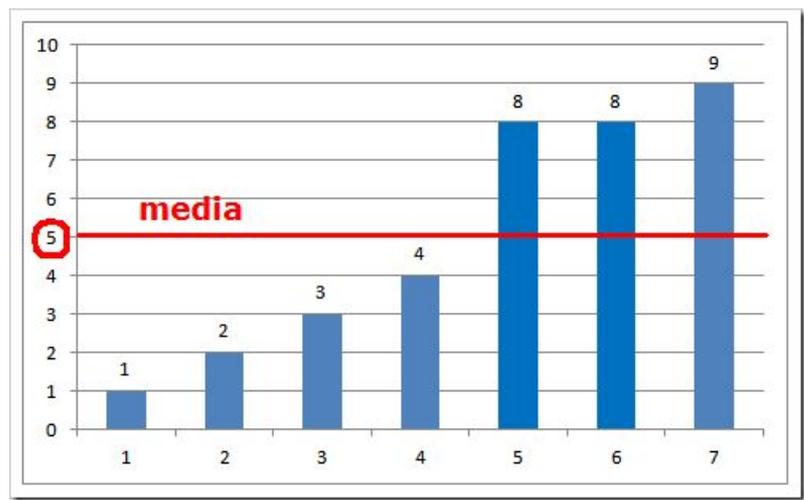
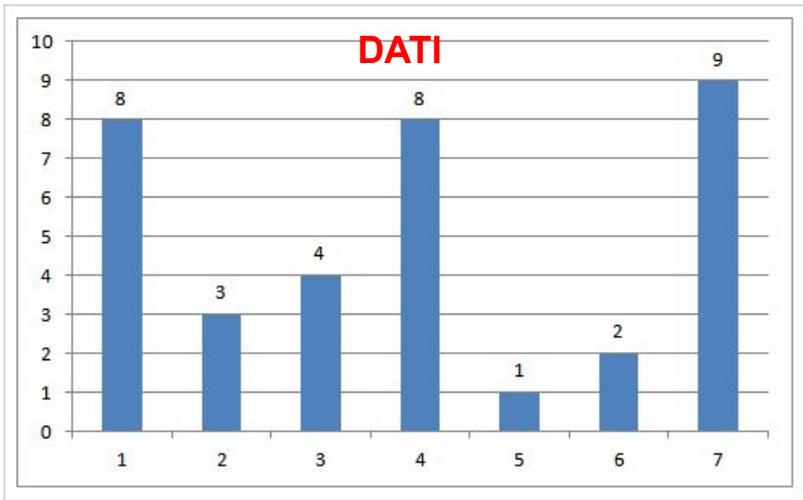
MEDIANA (o 50° percentile)

valore al di sotto del quale cade la
metà dei dati (valore centrale)

3

MODA

valore che ricorre con maggiore
frequenza



Indici di Variabilità

Descrivere la variabilità dei dati

1

MAXIMUM valore massimo

MINIMUM valore minimo

RANGE Differenza tra il valore massimo e il valore minimo

2

QUARTILE

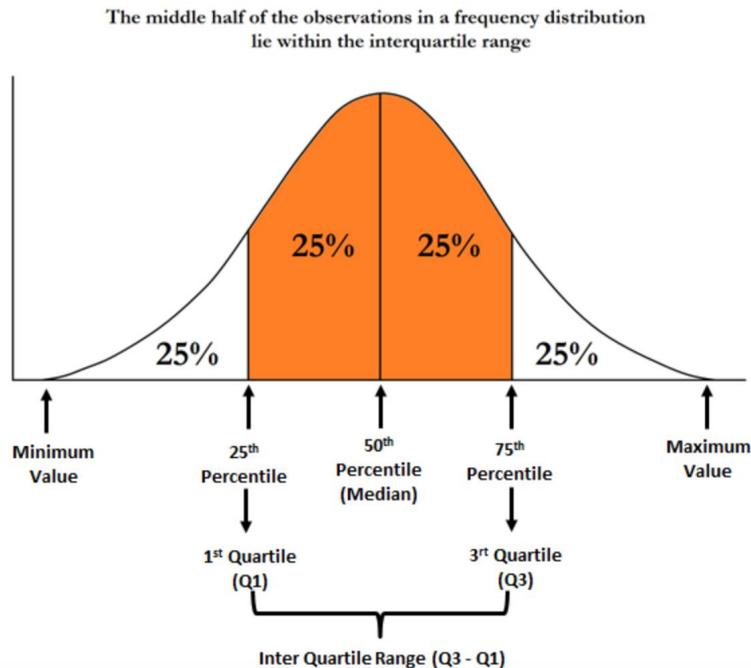
3

VARIANZA

dispersione dei valori del dataset attorno al valor medio. La deviazione standard è la radice quadrata della varianza

Quartile

I quartili dividono un set di dati in 4 parti uguali e si riferiscono ai valori del punto tra i quarti. Il Quartile inferiore (Q1) è il punto tra il 25% più basso di valori e il 75% più alto di valori. È anche chiamato il 25 ° percentile. Il secondo quartile (Q2) è il centro del set di dati. È anche chiamato 50 ° percentile, o mediana. Il quartile superiore (Q3) è il punto tra il 75% più basso e il 25% più alto di valori. È anche chiamato il 75 ° percentile.



Approfondimento

<https://www.datavedas.com/descriptive-statistics/>