

# Statistica per la digital economy (compito 13 dic 2024)

Caricare in memoria il dataset USA tramite l'istruzione

```
load USA
```

Questo dataset contiene 3 indicatori di criminalità dei 50 stati americani (Murder, Assault, Rape)

Murder = Assassini

Assault = Aggressioni

Rape = Stupri

ed un indicatore legato alla densità della popolazione urbana (UrbanPop).

```
load USA
Xtable=USA;
X=Xtable{:, :};
nomiRig=Xtable.Properties.RowNames;
nomiVar=Xtable.Properties.VariableNames;
```

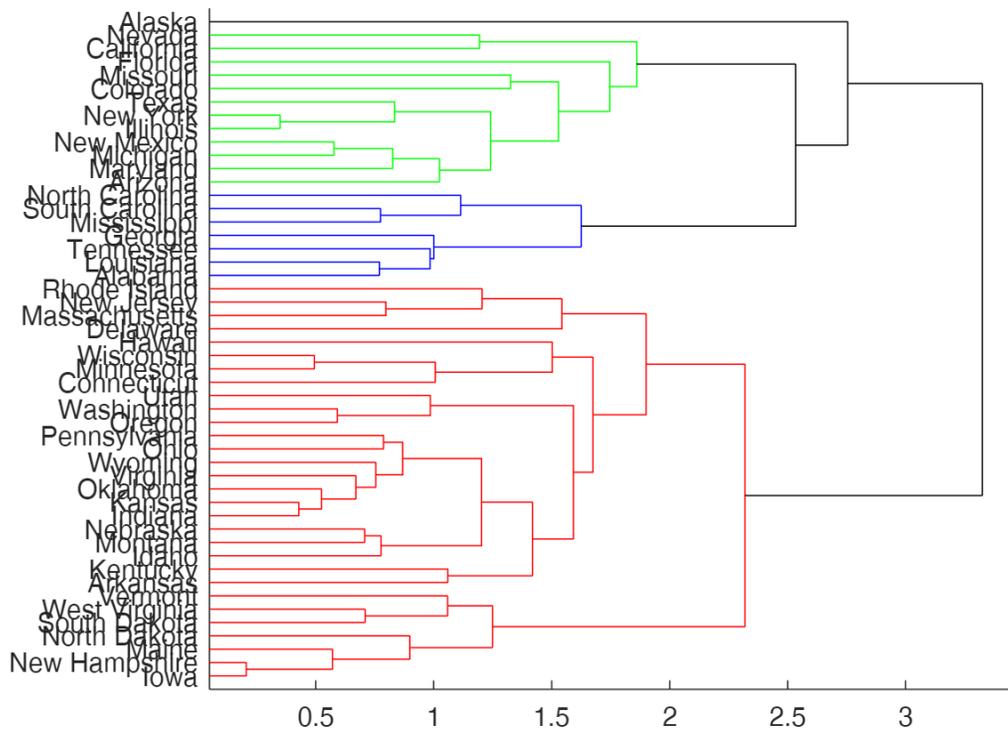
Dopo aver opportunamente trasformato i dati utilizzare la metodologia di cluster gerarchico del legame medio **(punti 4)**.

```
% Passo a valori standardizzati per eliminare la diversa unità di misura ed
% il diverso ordine di grandezza
Z=zscore(X);

% Metodo del legame medio
tree = linkage(Z, 'average');
```

Mostrare il dendrogramma visualizzando tutte le aggregazione con l'opzione 'Orientation','right' e visualizzando i nomi delle unità. Colorare nel dendrogramma i rami utilizzando la 'ColorThreshold' di default. **(punti 5)**

```
figure
dendrogram(tree,0, 'Orientation', 'right', 'ColorThreshold', 'default', 'Labels',n
nomiRig);
```



Quanti gruppi questa metodologia suggerisce se si utilizza la 'ColorThreshold' di default? **(punti 3)**

```
% Se si utilizza la 'ColorThreshold' di default questa metodologia
% suggerisce la presenza di quattro gruppi
```

Utilizzando il numero di gruppi suggerito dalla 'ColorThreshold' di default, assegnare ogni unità ad un determinato gruppo **(punti 4)**.

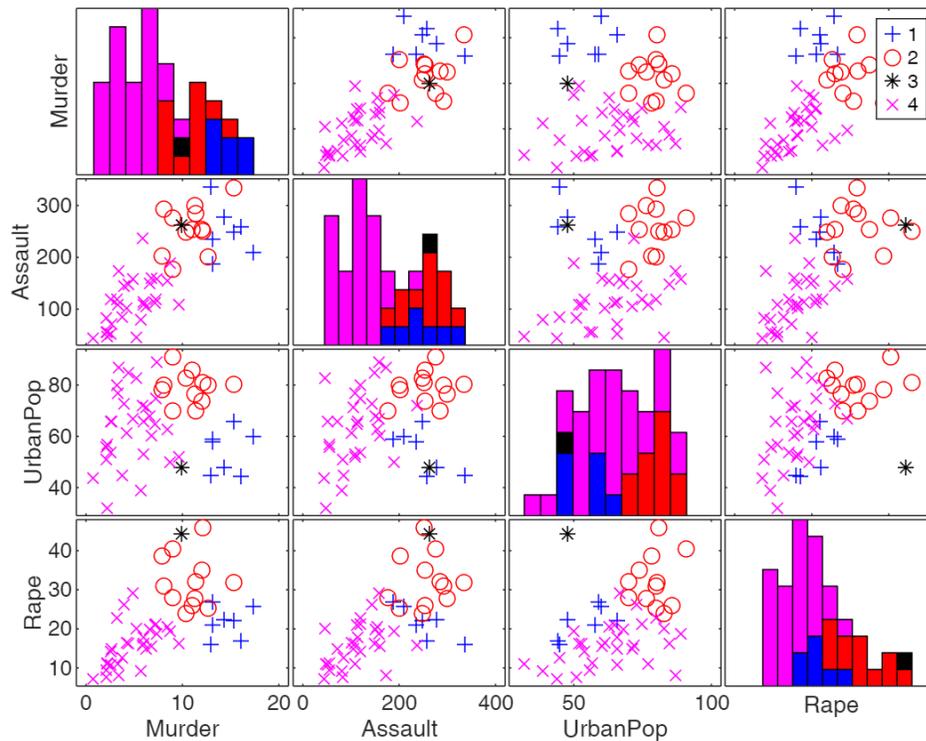
```
% La funzione cluster viene chiamata con il "Name",value 'MaxClust',4 per
% utilizzare 4 gruppi
idxLegameMedio = cluster(tree,'MaxClust',4);
```

Indicare se ci sono unità (stati americani) che non si aggregano ad altri **(punti 2)**

```
% Lo stato dell'Alaska (gruppo 3) non si aggrega con nessun altro stato
```

Creare la matrice dei diagrammi di dispersione utilizzando simboli diversi per le unità appartenenti ai diversi gruppi **(punti 3)**

```
spmpplot(Xtable,idxLegameMedio);
```

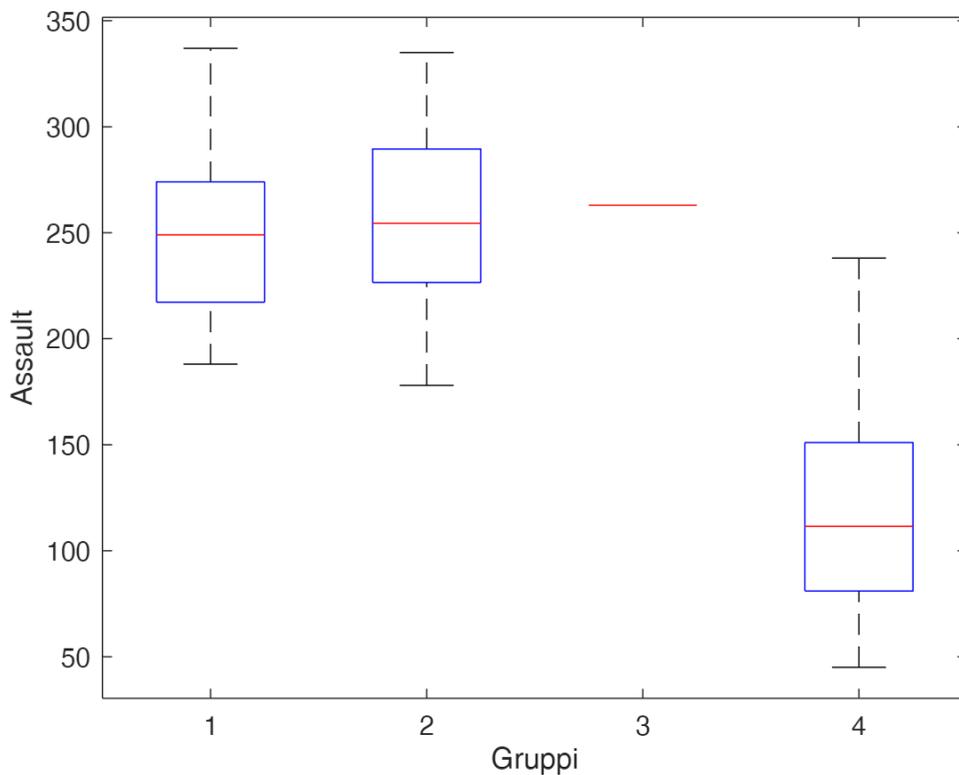


Che caratteristiche hanno le città che appartengono al gruppo 1 rispetto alla variabile murder? **(punti 3)**

% Dall'esame della matrice dei diagrammi di dispersione emerge che gli  
 % stati che appartengono al gruppo 1 sono caratterizzati da valori  
 % generalmente molto  
 % elevati di murder (v. ad es. diagrammi nella prima riga della spm)

Mostrare il boxplot della variabile "Assault" separato per ciascun gruppo che è stato trovato e commentare le caratteristiche del gruppo 4 **(punti 3)**

```
figure
boxplot(Xtable{:,"Assault"},idxLegameMedio)
ylabel('Assault')
xlabel('Gruppi')
```



% Le unità che appartengono al gruppo 4 presentano valori di Assault  
 % generalmente più bassi di quello dei gruppi 1 e 2 ed inferiori a quelli  
 % del gruppo 3 (Alaska)

Calcolare il secondo decile per tutte e 4 le variabili del dataset originario e mostrare nella Command Window il risultato in formato table (**punti 3**)

```
quan=quantile(X,0.2);
quanT=array2table(quan,"RowNames","X0.2","VariableNames",nomiVar);
disp(quanT)
```

	<u>Murder</u>	<u>Assault</u>	<u>UrbanPop</u>	<u>Rape</u>
<b>X0.2</b>	3.35	94	51.5	13.5