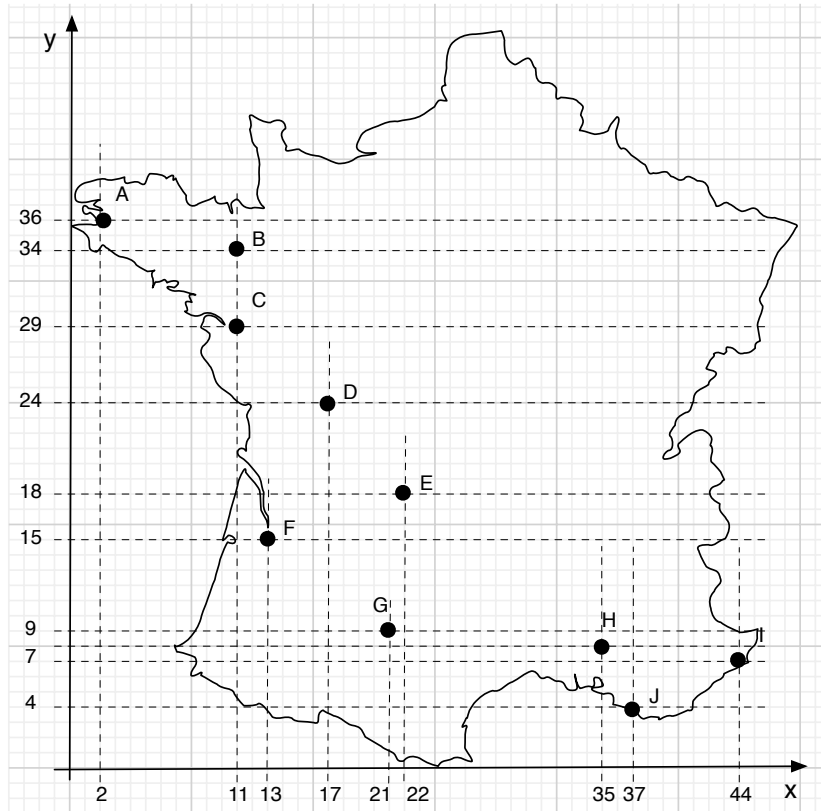


TD2 : K-Mean

Objectif : faire exécuter l'algorithme K-Mena à la main puis avec le logiciel R. Un compte rendu est à rendre en fin de TD.

Données : la carte ci-dessous, les points de A à I représentent des villes. Les axes x et y fournissent les coordonnées des villes sur la carte.



Questions :

- 1- Calculer le centre de gravité et l'inertie des individus

$$G = \frac{1}{N} \sum_{i=1}^N x_i \quad \mathcal{I} = \sum_{i=1}^N d(x_i, G)^2$$

- 2- Avec Excel, faire les étapes de l'algorithme ci-dessous avec k=2. On utilisera les points A et B pour l'initialisation. Les fonctions Excel suivantes seront utiles :

- SI(test; valeur V; valeur F) rend la valeur V si le test est vrai, sinon valeur F ;
- NB.SI(plage ; valeur) donne le nombre de cellules égales à la valeur ;
- MOYENNE.SI.ENS(plage ; plage de test; valeur) calcule la moyenne de plage pour les cellules égales à valeur de la plage de test.
- SOMME.SI.ENS(plage ; plage de test; valeur) calcule la somme de plage pour les cellules égales à valeur de la plage de test.
-

Algorithme :

- Choisir k individus qui représentent le centre de gravité des classes. On les note :

$$g_1^{(1)} \dots g_k^{(1)} \dots g_K^{(1)}$$

$$t = 0 ; I_W^{(0)} = \infty$$

- Répéter

- $t = t + 1$

- calculer la distance au carrée entre chaque individu et chaque centre de gravités

- assigner chaque individu à la classe la plus proche

$$C_k^{(t)} = \{x_i : d(x_i, G_k^{(t)}) < d(x_i, G_j^{(t)}), \forall j = 1 \dots K\} \quad \forall k = 1 \dots K$$

- mettre à jour les centres de gravité

- calculer l'inertie intra classe avec les nouveau centres de gravité

$$I_W^{(t)}$$

- tant que

$$I_W^{(t)} - I_W^{(t-1)} > \delta$$

- 5- Pour étape, calculer les inerties interclasse. Vérifier que $I = I^W + I^B$

$$I^W = \sum_{l=1}^{|C|} \mathcal{I}_l$$

$$I^B = \sum_{l=1}^{|C|} |C_l| d(G_l, G)^2$$