

Série de TD N°1

Clustering 1 (Types des variables)

Exercice 1 : Les variables de données peuvent être de type : numérique, ordinale, nominale (ou catégoriale), asymétrique binaire, intervalle ou vecteurs.

- a- Donner un exemple pour chaque type de variables.
- b- Décrire brièvement (en donnant un exemple) comment mesurer la similarité entre deux éléments de chacune des variables citées plus haut.
- c- Comment procéder pour mesurer la similarité entre des éléments de types différents ?

Exercice 2 : Soient les mesures suivantes de la variable âge : 18, 22, 25, 42, 28, 43, 33, 35, 56, 28. Standardiser cette variable par l'écart absolu moyen ensuite par le Z-score.

Exercice 3 : Soient deux objets représentés par les tuples (22, 1, 42, 10) et (20, 0, 36, 8) :

- a- Calculer la distance euclidienne entre ces deux objets
- b- Calculer la distance de Manhattan entre ces deux objets
- c- Calculer la distance de Minkowski entre ces deux objets, en utilisant $q=3$.

Exercice 4 : L'observation de quatre variables binaires sur quatre individus différents a conduit aux résultats suivants.

Construire la matrice des similitudes en utilisant l'indice de Jaccard.

Individu	Variable 1	Variable 2	Variable 3	Variable 4
A	1	1	0	1
B	1	0	0	0
C	0	1	1	0
D	0	1	1	1