

Mesures de distance

Introduction

Véronique Tremblay

Distance, dissemblance et similarité

Avant tout

1. Quelles caractéristiques (variables) des individus sont importantes?
2. Est-ce que toutes ses caractéristiques (variables) ont la même importance?
3. Quelles caractéristiques (variables) serviront à construire les groupes et lesquelles serviront à les décrire?
4. Comment mesurer la le niveau de similarité entre deux individus?

Une mesure de distance

d doit satisfaire

- $d(i, j) \geq 0$
- $d(i, i) = 0$
- $d(i, j) = d(j, i)$
- $d(i, l) \leq d(i, j) + d(j, l)$



Un *indice de similarité* s entre des objets doit satisfaire les propriétés suivantes pour tout $i, j \in \{1, \dots, n\}$:

- $0 \leq s(i, j) \leq 1$
- $s(i, j) = s(j, i)$
- $s(i, i) = 1$

Une distance peut se transformer en similarité en posant

$$s_{ij} = \frac{1}{1 + d_{ij}}.$$

La relation inverse n'est pas vraie en raison de l'inégalité du triangle.

On peut aussi définir la dissemblance (*dissimilarity* en anglais) entre deux objets, soit

$$d_{ij}^* = 1 - s_{ij}.$$

- Distance, similarité, dissemblance
- Varie selon le type de variable