

STID 2

Analyse des Correspondances

T. Dhorne

TD2

1 Tableaux de travail

On dispose de la table de contingence suivante qui représente le croisement de la religion et de la préférence politique d'américains.

Religion	Préférence politique		
	Dém	Ind	Rép
Cat	93	14	61
Jui	69	27	38
Pro	71	19	126

1.1 Quelles sont les dimensions du tableau de données de départ ?

Correction

Les dimensions du tableau de départ sont :

- le nombre d'individus en lignes : ici la somme du tableau de contingence 518
- le nombre de variables en colonnes : ici 2 variables qualitatives

soit 518×2 .

1.2 Quelles sont les dimensions du tableau disjonctif correspondant. Donner l'allure du tableau disjonctif en représentant au moins chacune des lignes différentes.

Correction

Les dimensions du tableau disjonctif sont :

- le nombre d'individus en lignes : ici la somme du tableau de contingence 518
- le nombre de modalités des variables en colonnes : ici la somme de 3 et 3 = 6

soit 518×6 .

1.3 Quelles sont les dimensions du tableau de Burt ? Donner le tableau de Burt associé.

Correction

Le tableau de Burt est une matrice carrée égale à tXX où X est le tableau disjonctif $X = X_1|X_2$, il a donc pour taille le nombre de colonnes de X au carré, soit 6×6 .

2 Lien entre les variables qualitatives

2.1 Quelle est la dimension de l'espace du lien global entre les deux variables ? Quel vecteur trivial (de variance nulle) appartient à cet espace ? Quelle est la dimension de l'espace (non trivial) du lien.

Correction

L'espace du lien est a priori de dimension 3 (3 dimensions pour chaque variable). Mais dans chaque espace le vecteur des 1 somme des colonnes disjonctives d'une variable est une variable triviale de variance nulle. La dimension de l'espace non trivial est donc 2, c'est encore la valeur de l'indice de Cramér : $\min(I - 1, J - 1)$

2.2 Donner la forme d'une combinaison linéaire des colonnes (indicatrices) d'une des variables issues du tableau disjonctif (cf cours d'algèbre linéaire)

Correction

Si le tableau disjonctif est $X = X_1|X_2$, une combinaison linéaire de X_1 (resp de X_2) est $X_1\alpha$, resp $X_2\beta$.

2.3 Choisir deux combinaisons linéaires simples (les plus simples possibles mais non triviales) des indicatrices des deux variables qualitatives (tableau disjonctif) et calculer leur corrélation.

Correction

On peut choisir des combinaisons visant à sélectionner certaines colonnes : 100, 010 ou 001.

2.4 Essayer d'optimiser manuellement pour trouver les deux combinaisons linéaires maximisant la corrélation. Conclure.