NOM TROTIN

Prénom Benjamin

Promo L3 Alternance

Date 16/04/18

17,5

TROTIN Benjamin
L3-APP - 2017 -

MATIÈRE Analyse de données

Q1. Outils Statistiques

1.1 $\overline{x} = \dfrac{1}{n} \sum_{i=1}^{n} x_i$

1.2 $var(x) = \dfrac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2$

1.3 $ecart\text{-}type(x) = \sqrt{var(x)}$ car $var(x) = (ecart\text{-}type(x))^2$

1.4

1.5 Pour Z_{ij} la frequence theorique :

$$\left(\sum_{i=1}^{n} z_i \times \sum_{j=1}^{m} z_j \right) / \sum_{i=1}^{n} \left(\sum_{j=1}^{m} z_{ij} \right)$$

Q2. Facteur de corrélation

Un tableau de correlation permet de savoir si deux variables sont plus ou moins corrélées entre elles.

Une très forte corrélation s'exprime par une valeur proche de +1/-1. Dans la matrice la diagonale est remplie de 1 puisqu'une valeurs est obligatoirement très corrélée avec elle même.

En opposition deux valeurs sont faiblement corrélées quand la valeur dans la matrice de correlation est proche de 0.

Q3 Valeurs propres / Vecteurs propres

= 3.1 Pour calculer les valeurs propres, il faut tout d'abord calculer le déterminant pour ~~que~~ et trouver ~~des~~ lambdas qui permettent de fixer le déterminant à 0 (Voir dernière page)

3.2 Il y a autant de valeurs propres qu'il y a de colonne dans la matrice.
Par exemple ici la matrice A, a n lignes / n colonnes donc il y a n valeurs propres

Q4 Principes de l'ACP - Données quantitatives

= 4.1 «centrer» permet une meilleure lecture sur les schéma, d'utiliser les axes, puisque les valeurs ce retrouve centrées en 0.

? «reduire» permet une meilleure lisibilité des valeurs, évite les gros écarts entre celles-ci et donc d'avoir une échelle plus adaptée pour pouvoir analyser ces valeurs ?

4.2 La courbe permet de voir plus rapidement ce qu'exprime le « cumulative proportion».

+ En croisant les données des valeurs propres avec la courbe, qu'à partir de la composantes 3 90 % des valeurs sont comprises dans le résultat.

Si l'on considère que ce n'est pas assez suffisant, on peut prendre jusqu'à la composante 4, ce qui nous amène à 97%.

Prendre la suite ne ferai que complexifier la tâche pour le peu de données qu'elles apportent.

4.3 Les vecteurs propres permettent de calculer les les valeurs des composantes principales en faisant un produit matriciel :

X.P avec X matrice des variables initiales
et P matrice des vecteurs propres

Q5 AFC - Données Qualitatives

N

#

S.1 Khi 2 permet de calculer le taux d'indépendance entre deux variables selon un degré, quelque soit la matrice en question.

#

S.2 Il suffit d'abord de calculer le degré de khi-2

Si l'on reprend l'exemple de la question nous avons une matrice 3x4 donc le calcul est le suivant :

Degré = $(3-1) \times (4-1) = 6$

Donc nous allons nous placer à la 6ème lignes de la table khi-2.

Dans l'exemple l'indicateur est de 13, donc à la 6ème ligne nous cherchons ~~...~~ la valeur la plus proche inferieur à 13 (pour prendre le pire cas)

Ici nous trouvons 12,6, il suffit ensuite de remonter la colonne pour lire la valeur.

Dans l'exemple le resultat est de 0,95. J'ai donc 95 % de ~~...~~ rejeter que les variables soient indépendante

Q6 ACP - Déroulement

Tout d'abord, après avoir fait le calcul des moyennes, variances, ecart types et covariance. On calcule le tableau de corrélation.

A cette étape-ci nous pouvons déjà faire une première analyse pour voir les variables qui sont plus ou moins correlées. Mais il est un peu compliqué de faire des schémas avec les valeurs initiales.

(Ici on observe déjà que la plus grosse correlation à lieu entre IMM et EXP avec -0,9445)

Pour cette fois-ci avoir une meilleure analyse et surtout graphique on centre et on reduit les données initiales pour refaire les analyses de départ.
Grâce à ça il est plus simple maintenant de faire des cercles de corrélations. Pour être sûr de pas cette trompe à cette étape, il suffit de vérifier que notre tableau de corrélation est le même que celui obtenu précédement. (celui ne change pas malgré le centrage et la réduction)

comment? Il faut ensuite calculer les vecteurs propres et ensuite les composantes principales. On peut en deduire le coudes et calculer les valeurs avec les composantes principales.

Ici nous avons toutes les clés en main pour faire une analyse complète

3.2 (suite)

$$\begin{array}{c|c} & \begin{array}{ccccc} 1 & & j & & m \end{array} \\ \begin{array}{c} 1 \\ \\ i \\ \\ n \end{array} & \begin{bmatrix} x_{11}-\lambda & x_{12} & & & \\ x_{21} & \ddots & & & \\ & & \ddots & & \\ & & & \ddots & x_{n-1m} \\ & & & x_{nm-1} & x_{nm}-\lambda \end{bmatrix} \end{array}$$

Il faut calculer le déterminant pour en sortir une inéquation et la résoudre pour que le déterminant soit égale à 0.
Une fois les valeurs propres trouver on peut calculer les vecteurs propres