

## CORRECTION DE L'INTERROGATION

### Analyse Exploratoire de Données

On étudie les données fournies par le Centre National de Cinématographie (CNC) sur la fréquentation des salles de cinéma en 2003. Pour 94 départements de France métropolitaine, on mesure les 10 variables décrites ci-dessous, et on obtient un tableau dont un extrait est donné ci-dessous.

ندرس البيانات التي قدمها المركز الوطني للسينما (CNC) حول حضور السينما في عام 2003. بالنسبة لـ 94 مقاطعة من فرنسا الحضرية، نقيس المتغيرات العشرة الموضحة أدناه، ونحصل على جدول يتم تقديم مقتطف منه أدناه.

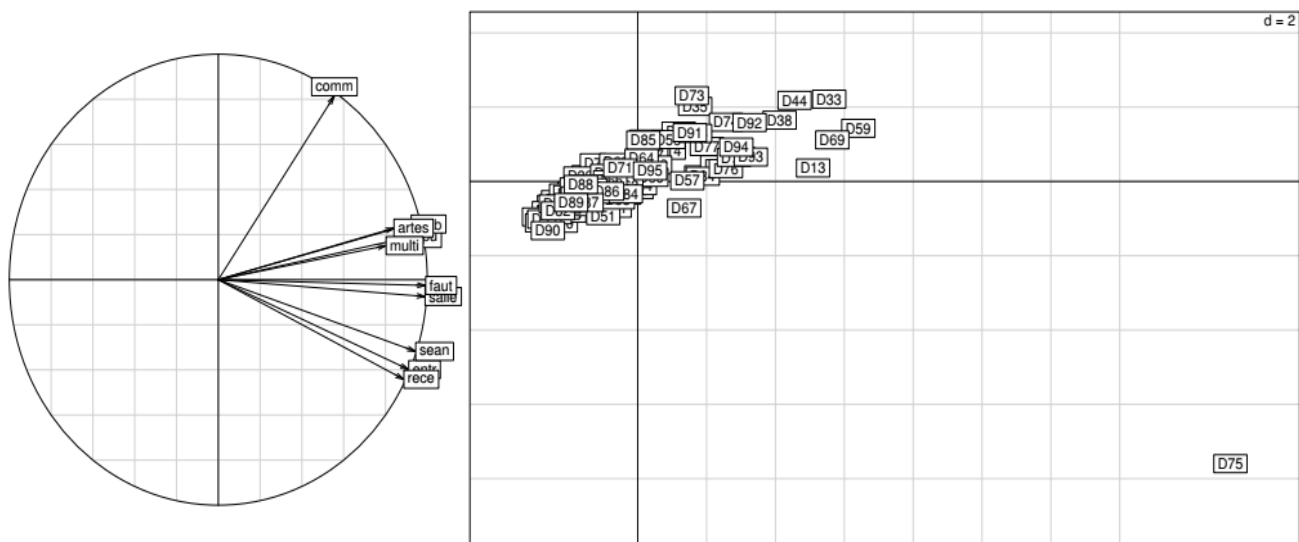
|                                                                  | popu      | entr   | rece    | sean | comm | etab | salle | faut  | artes | multi |
|------------------------------------------------------------------|-----------|--------|---------|------|------|------|-------|-------|-------|-------|
| <b>popu:</b> population du département (en millions d'habitants) | D71 0.545 | 0.878  | 4.704   | 36   | 18   | 24   | 50    | 9728  | 13    | 0     |
| <b>entr:</b> nombres d'entrées réalisées (en millions)           | D72 0.53  | 1.119  | 5.617   | 40   | 12   | 15   | 39    | 7561  | 7     | 1     |
| <b>rece:</b> recettes (en millions d'euros)                      | D73 0.373 | 1.704  | 10.264  | 53   | 42   | 59   | 89    | 17870 | 10    | 1     |
| <b>sean:</b> nombre de séances (en milliers)                     | D74 0.632 | 2.981  | 17.616  | 74   | 33   | 51   | 98    | 21201 | 15    | 3     |
| <b>comm:</b> nombre de communes équipées de salles de cinéma     | D75 2.125 | 30.439 | 192.244 | 698  | 1    | 92   | 368   | 72752 | 38    | 5     |
| <b>etab:</b> nombre de cinémas en activité                       | D76 1.239 | 3.366  | 19.172  | 128  | 23   | 30   | 108   | 23436 | 11    | 3     |
| <b>salle:</b> nombre de salles en activité                       | D77 1.194 | 3.298  | 19.964  | 87   | 30   | 33   | 89    | 19029 | 10    | 2     |
| <b>faut:</b> nombre de fauteuils disponibles                     | D78 1.354 | 4.625  | 26.7    | 130  | 29   | 35   | 95    | 19302 | 14    | 2     |
| <b>artes:</b> nombre de salles d'art et essai                    | D79 0.344 | 0.522  | 2.58    | 17   | 13   | 16   | 26    | 5266  | 11    | 0     |
| <b>multi:</b> nombre de multiplexes                              | D80 0.556 | 0.889  | 5.082   | 27   | 15   | 18   | 32    | 6622  | 4     | 0     |
|                                                                  | D81 0.343 | 0.601  | 2.818   | 21   | 11   | 13   | 26    | 4798  | 8     | 0     |

**Extrait des données**

On effectue une ACP sur les données centrées-réduites, et on obtient les valeurs propres suivantes:

7.77 1.30 0.53 0.20 0.16 0.02 0.01 0.01 0.00 0.00

Le premier plan principal est représenté par le cercle des corrélations et la projection des individus ci-dessous:



**Question 1:** Combien de composantes principales peut-on retenir? Quelle inertie expliquée peut-on obtenir?

**Réponse:** Il y a deux valeurs propres supérieures à 1 (Règle de Kaiser). L'inertie représentée par le premier plan principal est 9,07, ce qui fait 90,7 % de l'inertie totale (puisque le total est 10).

**Question 2:** Comment peut-on interpréter la première composante principale? Est-il vraiment possible d'interpréter la seconde composante principale?

**Réponse:** La première composante principale présente une corrélation positive avec les variables. On peut dire globalement que cette variable représente l'activité cinématographique du département. La seconde composante principale est plus difficile à interpréter, à cause de la variable comm (nombre de communes) qui présente une très grande influence.

**Question 3:** En observant la projection des individus sur le premier plan principal, que peut-on observer? En quoi est-ce lié aux difficultés d'analyse de la deuxième composante principale dans la question précédente ?

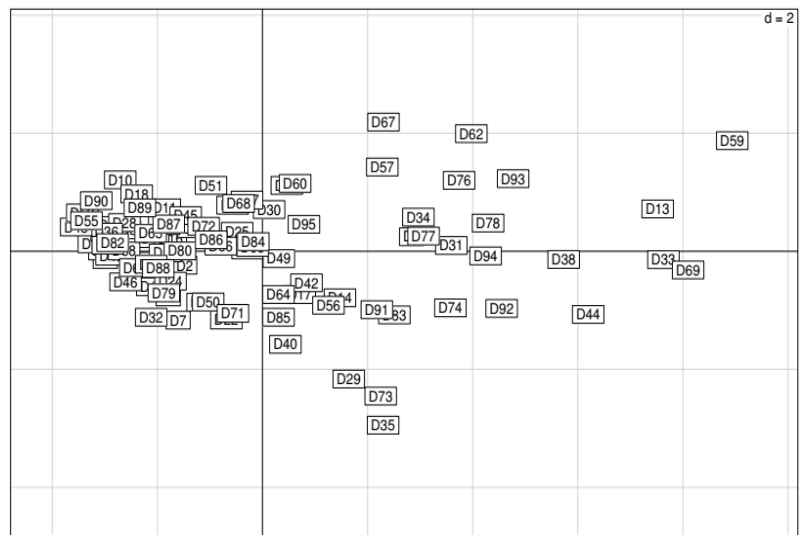
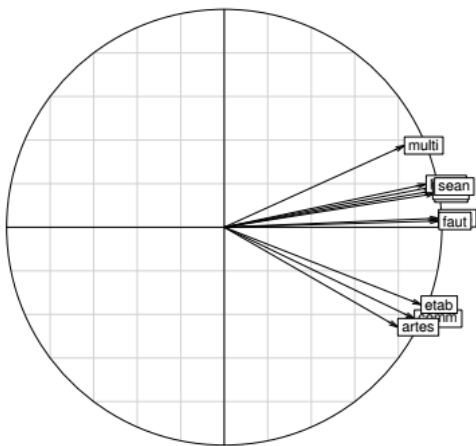
**Réponse:** On remarque tout de suite que le point D75 (Paris) est très écarté des autres. La raison est que la ville de Paris elle-même a une offre et une demande de cinéma très atypiques par rapport aux autres départements. Un facteur particulier qui saute aux yeux au vu des données fournies est que le département ne compte qu'une commune, alors que les autres chiffres sont tous très élevés. Cette entrée cause donc une grande distorsion.

**Question 4:** Que faut-il faire pour améliorer la situation?

**Réponse:** La solution est bien sûr de supprimer la ligne D75 et de refaire l'ACP sans elle.

On supprime le département de Paris (D75) et on obtient les données suivantes par ACP centrée-réduite:  
8.50 0.78 0.33 0.20 0.10 0.05 0.02 0.02 0.00 0.00

On reproduit aussi le nouveau cercle des corrélations et la projection des individus ci-dessous.



**Question 5:** Donnez les nouvelles valeurs propres retenues. Pourquoi la règle de Kaiser n'est pas appliquée? Quelle est la nouvelle inertie expliquée?

**Réponse:** La seconde composante principale a maintenant une valeur propre inférieure à 1. On doit la garder pour obtenir deux composantes principales. La règle de Kaiser n'est pas appliquée car on ne peut pas garder seulement une seule valeur propre. Avec ces deux variables, le pourcentage d'inertie totale représentée est de 92,8 %.

**Question 6:** L'interprétation de la première composante principale a-t-elle changé ? Peut-on maintenant faire une nouvelle interprétation de la seconde composante principale?

**Réponse:** *Les caractéristiques et l'interprétation de la première composante principale sont les mêmes. Pour la seconde composante principale, on peut dire que les corrélations sont trop faibles pour pouvoir apparaître clairement. Mais le nombre de communes n'est plus un point singulier (son effet a été supprimé avec la suppression du département D75)*

**Question 7:** La projection des individus sur le premier plan principal ci-dessus donne-t-elle plus d'informations sur les composantes principales?

**Réponse:** *Il y a une forte répartition horizontale des individus: à gauche, on trouve les départements ayant une faible activité en termes de projections de films. Ils forment un bloc assez compact. Par contre, les départements qui sont plus actifs (à droite) sont plus « étalés ».*