

Statistique descriptive et exploratoire pour données fonctionnelles

Na'amane LA IB

Introduction. Les progrès récents en matière de stockage et de traitement des données se traduisent de plus en plus fréquemment dans de nombreux domaines scientifiques par la présence de données de type fonctionnel. On s'intéresse à l'étude statistique des observations qui ne sont pas des réalisations de variables aléatoires réelles ou vectorielles, mais des fonctions aléatoires : [courbes](#), [images](#), etc...

- Courbe des températures relevées en un point donnée à différents instants.
- Courbe des cumuls mensuels de précipitations en un point donnée
- Courbe de croissance d'un individu (ou d'une plante) au cours du temps.

Il s'agit de données appartenant à un espace de dimension infinie et qui rentrant dans le champ de la "très grande dimension" (Big data).

Problématique. Les problèmes statistiques qui peuvent se poser sont les mêmes qu'en statistique multivariée classique.

1. Statistique descriptive et exploratoire.

- On s'intéresse dans un premier temps aux mesures résumées de statistique descriptive spécifiques à l'analyse de données fonctionnelles.
- Puis à l'extension des techniques d'Analyse en Composantes Principales (ACP) au cas fonctionnel.
 - Son premier intérêt sera, comme en statistique multivariée de fournir un outil pour visualiser la répartition des données, elle sera également très utilisée pour réduire la dimension des données, par exemple dans le cas du modèle linéaire fonctionnel.

2. Il peut s'agir de problèmes de prédiction

- Peut-on prédire à partir de la courbe X une autre caractéristique réelle Y ?
 - dans ce cas des techniques de régression devront être utilisées si Y est quantitative.

Objectif. L'objectif de ce projet est d'utiliser ce point de vue fonctionnel pour illustrer les points évoqués ci-dessus à travers un jeu de données de votre choix.

Protocole à suivre.

- Une recherche bibliographique sur ce type de données et leurs applications. • Le jeu de données:

- Il est constitué d'un échantillon issu des données considérées comme une réalisation d'une v.a. fonctionnelle. Bien identifier qui sont les individus et qui sont les variables

- Une introduction avec:

- Description du jeu de données (source, année,...)
- Présentation de chaque variable avec leur nature, représentation graphique appropriée
- Objectif de l'étude (à quelle(s) interrogation(s) souhaitez-vous répondre avec votre étude statistique ?)

- Une analyse des résultats et comparaison de différentes méthodes utilisées :

Il ne faut pas se contenter d'un listing de méthodes mais apporter une analyse pertinente des résultats. Il faut aller au-delà des résultats numériques.

- Une conclusion avec une partie sur l'intérêt (ou non intérêt, ou simplement commentaires) de cet exercice.

- Rendu. Vous devrez rendre

- un pdf de l'analyse de votre jeu de données,
- le jeu de données brute (Excel/csv) en donnant la référence (cette référence doit être facilement accessible en cliquant sur l'adresse du site),
- le script (commenté) avec vos traitements sous R.

- Le nom de l'archive doit contenir les noms des auteurs. N'oubliez pas de mettre vos noms en entête de vos documents.

- Seul le pdf sera corrigé. Cela signifie que votre rapport doit être suffisamment détaillé avec [lignes de codes](#), [graphiques](#), [analyses et commentaires](#). [Les autres documents sont aussi nécessaires et vous devez les faire avec beaucoup de soin, ils seront notés.](#)

Référence. <http://modulad.fr/archives/numero-43/VIEU/2-Vieu.pdf> 2