

# Améliorer la pertinence des indicateurs de qualité à l'aide de l'Analyse en Composantes Principales

FRANCINE DIENER

Laboratoire J.A. Dieudonné  
Université Côte d'Azur

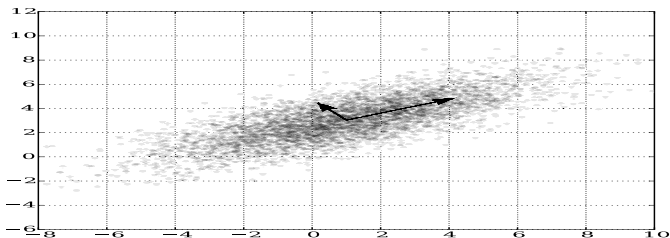
22 Février 2024

- Les mathématiques ne sont pas seulement des méthodes de calcul mais aussi des "concepts pour penser avec".
- Les chercheurs non mathématiciens peuvent comprendre ces concepts sans nécessairement connaître tous les calculs sous-jacents.
- Encore faut-il que les mathématiciens acceptent de partager ces concepts en les dépouillant de leurs aspects les plus techniques
- Un groupe de chercheurs pluridisciplinaire c'est comme un groupe de musiciens de jazz, chacun est un expert de son instrument mais, pour produire ensemble de la musique, il faut que chacun connaisse le mieux possible les possibilités des autres.

# Qu'est-ce que l'ACP?

Firefox

<https://upload.wikimedia.org/wikipedia/commons/FF5/GaussianScatt...>



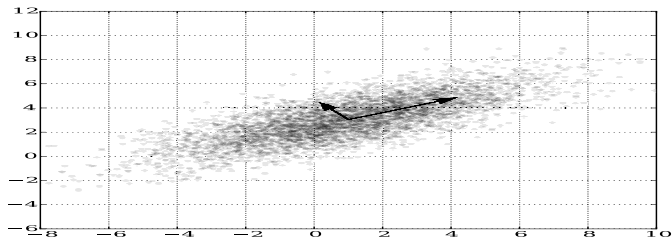
L'**analyse en composantes principales** est une méthode d'analyse des données qui consiste à transformer des variables liées entre elles (dites corrélées) en nouvelles variables décorrélées nommées **composantes principales**.

18/02/2024, 10:04

# Des axes de dispersion maximale

Firefox

<https://upload.wikimedia.org/wikipedia/commons/4/45/GaussianScatt...>



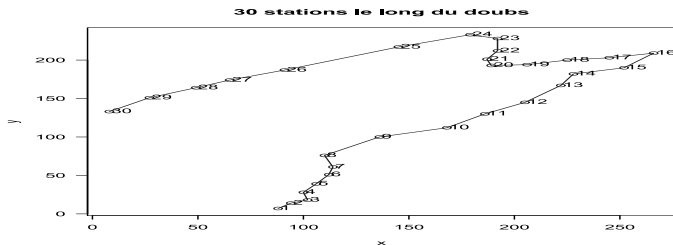
On va rechercher un premier axe qui maximise la variance du nuage, puis un second qui maximise la variance restante (dans l'espace perpendiculaire) et ainsi de suite. Les vecteurs unitaires sur ces axes sont les composantes principales. Les quelques premières capturent l'essentiel de la variabilité. On ne perd que très peu d'information en oubliant les autres.

# Un exemple

On a relevé sept données physico-chimiques (**7 variables**), le **pH**, la **dureté de l'eau**, le **phosphate**, le **nitrate**, l'**ammoniaque**, l'**oxygène** et la **demande biologique en oxygène** dans chacune des **30 stations** le long du Doubs. De l'examen des données, on voudrait tirer un **indicateur de pollution**.

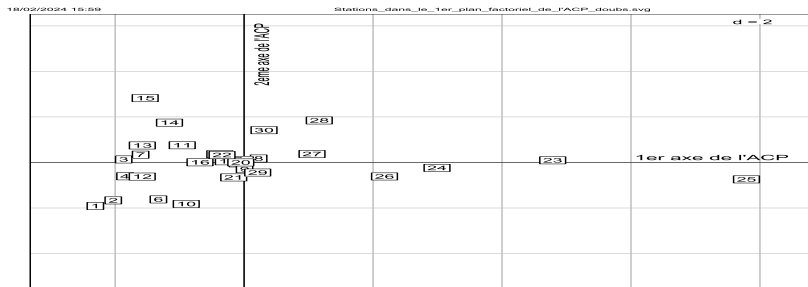
18/02/2024 15:59

Stations\_le\_long\_du\_doubs.svg



Le nuage de points comporte donc ici 30 points dans l'espace à 7 dimensions.

# Que nous apprend l'ACP?

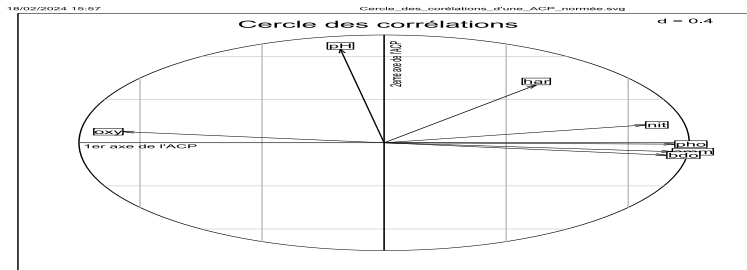


La grandeur représentée par le premier axe, qui est une combinaison linéaire des 7 variables initiales, est très grande pour les stations 25, 23 et 24 et bien plus petite pour les stations 1, 2, ou 3. Que mesure-t-elle?

file:///C:/Travail/F/TEX/slides/Stations\_dans\_le\_1er\_plan\_factoriel\_de\_l'ACP\_doubs.svg

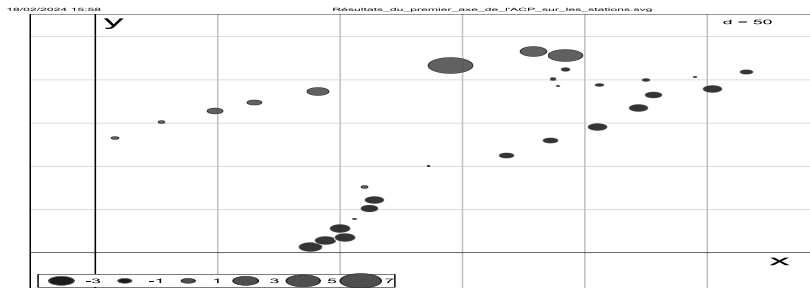
1/1

# Un indicateur de pollution?



Dans le cercle des corrélations, le premier axe montre une corrélation négative entre l'oxygène et 3 polluants connus : nitrates (nit), phosphates (pho), ammoniacque (amm). On peut dire que cet axe est un **indicateur de pollution**. Le deuxième axe est un axe caractérisé essentiellement par le pH.

# Un indicateur de pollution



Chaque station est remplacée par un cercle prenant la valeur du résultat de l'ACP sur le premier axe de l'ACP. L'indicateur pourra être ré-utilisé pour des études futures ou pour d'autres rivières et fleuves.

file:///C:/Travail/F/TEX/slides/Résultats du premier axe de l'ACP sur les stations.svg

1/1



# Mais que peut l'ACP pour l'intervention sociale?

- L'ACP est une **méthode de base**, il en existe beaucoup d'autres plus élaborées, donc sans doute plus pertinentes mais plus difficiles à comprendre.
- l'ACP est l'exemple type d'une méthode d'**apprentissage non supervisé** au sens de l'IA
- La plupart du temps, quand on fait une ACP, on reste **insatisfait** du résultat: on y voit plus clair, mais ce n'est pas si nette que ça, ...
- Pour le mathématicien, c'est **une opération triviale**, un simple changement de bases dans l'espace des variables
- Pour le chercheur non mathématicien, c'est une manière ultra simplifiée de voir les données, **sans contenu conceptuel**
- La seule façon d'en tirer quelque chose est d'**y réfléchir de façon interdisciplinaire**, en partageant à la fois la compréhension de la méthode pour savoir ce qu'elle peut montrer et à la fois la nature des données et ce qu'elles pourraient révéler pour savoir interpréter.

J'ai mis l'accent sur la formation d'indicateurs, bien sur, l'ACP peut beaucoup plus!

Mais le fait d'avoir de bons indicateurs me semble important pour l'intervention sociale: en amont pour être en mesure de se **fixer des objectifs quantifiés**, en aval pour permettre l'**évaluation** de l'intervention.

Avec des indicateurs pertinents, on peut imaginer, un peu comme le préconise Ester Duflo, de réaliser **des "expériences" qui permettent de comparer** l'efficacité d'interventions sociales différentes sur des populations (échantillons) semblables.

Pour m'amuser, je termine en vous proposant la version Chatjpt de mon exposé ... à vous de juger ...

L'analyse en composantes principales (ACP) est une technique statistique largement utilisée dans divers domaines de recherche, y compris la recherche pour l'intervention sociale. Voici quelques façons dont l'ACP peut être utile dans ce contexte

1. **\*\*Réduction de la dimensionnalité des données\*\*** : La recherche en intervention sociale peut souvent impliquer des ensembles de données complexes avec de nombreuses variables. L'ACP peut aider à réduire la dimensionnalité de ces données en identifiant les combinaisons linéaires de variables qui expliquent le plus de variance dans l'ensemble des données, ce qui permet de simplifier l'analyse et d'identifier les principaux facteurs sous-jacents.
2. **\*\*Identification des structures cachées\*\*** : L'ACP peut aider à identifier les structures cachées ou les modèles sous-jacents dans les données sociales. Par exemple, dans une étude sur les facteurs contribuant à la réussite scolaire des élèves, l'ACP pourrait révéler des clusters de variables liées aux ressources familiales, à l'environnement scolaire, ou à d'autres facteurs socio-économiques.
3. **\*\*Visualisation des données\*\*** : L'ACP peut être utilisée pour visualiser les données dans un espace de dimension réduite, ce qui facilite l'interprétation des relations entre les variables. Par exemple, en réduisant les données à deux dimensions, on peut créer des graphiques de dispersion ou des cartes de chaleur pour explorer les relations entre différents aspects de l'intervention sociale.
4. **\*\*Identification des outliers et des observations atypiques\*\*** : L'ACP peut aider à détecter les valeurs aberrantes ou les observations atypiques dans les données, ce qui peut être important pour comprendre les exceptions ou les cas spéciaux dans le contexte de l'intervention sociale.
5. **\*\*Analyse des facteurs de risque et de protection\*\*** : L'ACP peut être utilisée pour identifier les facteurs de risque et de protection associés à un problème social spécifique. En examinant les relations entre les variables, on peut identifier les facteurs qui sont corrélés avec des résultats positifs ou négatifs, ce qui peut orienter la conception des interventions.

En résumé, l'ACP peut être un outil puissant pour explorer et comprendre les données complexes en recherche pour l'intervention sociale, en permettant une analyse plus approfondie des relations entre les variables et en facilitant l'identification des facteurs clés qui influent sur les résultats sociaux.