

PROJET

Consignes du projet

- Avant toute chose, n'hésitez pas à venir me voir ou à me contacter par mail si vous avez un doute, un problème ou une question sur le projet.
- Mon adresse mail : emmanuel.caron-parte@univ-avignon.fr.
- **Ne restez pas bloqué !**
- Tout le matériel concernant cette SAE est disponible sur la page e-uapv correspondante.

Introduction

- Le projet se fait par groupes de 2 étudiants (ou seul) et consiste en la rédaction d'un rapport écrit (pas de soutenance orale).
- Vous devez faire l'analyse de la série temporelle qui vous a été attribuée et mener à bien toutes les étapes de cette analyse, en suivant le plan du cours.
- Le rapport doit présenter le travail que vous avez effectué, avec vos conclusions.
- La constitution des groupes et l'attribution des jeux de données a été faite lors de la séance du 19 novembre. Concentrez-vous uniquement sur votre jeu de données !

Consignes pour le rapport

- Le rapport doit être structuré, lisible et avec suffisamment d'explications. Une attention particulière sera accordée sur la manière dont vous présentez les résultats.
Attention aux fautes d'orthographe et de grammaire ! Ces derniers points feront partie de la note.
- Le rapport devra comprendre une introduction pour présenter rapidement les données et le travail à effectuer, et une conclusion afin de résumer les résultats obtenus.
- En annexe du rapport, vous mettrez vos codes R **commentés**. **Attention** : Le code R sera aussi noté ! Les codes doivent compiler sans erreurs. Ils doivent notamment être cohérents avec les figures présentées. **Attention : Un code qui ne compile pas entraîne la division de la note par 2 !**
- **ATTENTION** : Chaque rapport/code doit être unique ! Il est interdit de copier entre vous. Cela se voit tout de suite et a déjà entraîné des sanctions les années passées...
- **En outre** : pour le code et l'étude des données, utilisez uniquement les commandes R vues en cours et présentes dans les TP (et les corrections). L'utilisation de toute autre commande devra être dûment justifiée, et apparaître comme nécessaire à l'étude.
- **Deadline pour rendre le projet : vendredi 9 janvier 2026 à 23h59.**

Présentation des données du projet

Il y a 6 jeux de données :

- `monthly_car_sales.csv` : nombre de voitures vendues par mois chez un concessionnaire automobile pendant 16 ans.
- `monthly_co2_pollution_in_ug_by_m3.csv` : taux de CO2 (dioxyde de carbone) dans l'air (en $\mu g/m^3$) mesuré mensuellement par un observatoire météorologique à Hawaï pendant 39 ans.
- `monthly_no2_pollution_in_ug_by_m3.csv` : taux de NO2 (dioxyde d'azote) dans l'air (en $\mu g/m^3$) mesuré mensuellement par un observatoire météorologique en Chine pendant 18 ans.
- `monthly_electric_production_in_MWh.csv` : production électrique mensuelle, en Mégawatt heure, d'une centrale électrique en France, mesurée pendant 20 ans.
- `monthly_oyster_production_in_tons.csv` : production mensuelle d'huîtres françaises, en tonnes, pendant 27 ans.
- `monthly_total_vehicles_miles_traveled_of_car_rental_agency.csv` : nombre de miles parcourus par l'ensemble du parc automobile d'une agence de location de voitures, mesuré mensuellement pendant 18 ans.

Le projet

L'objectif du projet est de traiter un jeu de données en utilisant tout ce que vous avez vu pendant ce cours. Il s'agit en quelque sorte de reprendre tous les TP et de les appliquer, point par point, en suivant la structure du cours et la logique de la modélisation des données temporelles.

Nous donnons ici les points principaux sur lesquels vous devez travailler, et que vous devez expliciter dans le rapport.

- 1) Importer les données dans R et les préparer pour l'étude.
 - Faire une première analyse descriptive et graphique : nombre de points, mesure du temps, modèle à utiliser, période, forme apparente de la tendance, de la composante saisonnière, etc.
 - Etudier la fonction d'autocorrélation et donner vos conclusions.
- 2) Estimer la tendance. Commencer par la méthode des moyennes mobiles en choisissant un ordre adapté. En déduire la forme de la tendance. Ajuster un modèle paramétrique sur la tendance.
- 3) Estimer la composante saisonnière. Donner la période, repérer le nombre de périodes. En déduire le nombre de coefficients saisonniers et les estimer. Construire la composante saisonnière complète à partir de cela.
- 4) Calculer la série CVS (Corrigée des Variations Saisonnières). Retrouve-t-on à peu près la forme de la tendance ?
- 5) En déduire la série ajustée. Calculer le critère MSE. Donner vos conclusions.
- 6) Ensuite, étudier les résidus. Regarder leur fonction d'autocorrélation. Ajuster un modèle stochastique (modèle ARMA) sur les résidus. On peut s'aider de la fonction d'autocorrélation (`acf(x)`) et la fonction d'autocorrélation partielle (`pacf(x)`) afin de choisir les ordres p et q à utiliser pour le modèle ARMA.

- 7) Après avoir estimé les résidus, construire le modèle ajusté total (pour toutes les données depuis le début, à partir des 3 composantes estimées). Calculer le critère MSE. Donner vos conclusions.
- 8) Utiliser votre modèle pour faire des prévisions à court terme. Prévoir les valeurs sur la période entière suivant votre dernière donnée.

Les points ci-dessus ne sont que les étapes principales de l'étude d'une série temporelle et servent à vous guider. N'hésitez pas à utiliser toutes vos connaissances afin d'étudier la série temporelle, et à aller plus loin.