

Rapport Synthétique: Analyse Epidémiologique des Syndromes Grippaux

ÉTUDE ÉPIDÉMIOLOGIQUE ET MODELISATION DES TAUX
D'INCIDENCE DES SYNDROMES GRIPPAUX

Diego CASAS BARCENAS ; Leo Jean UNITE
IUT PARIS - RIVES DE SEINE

Introduction	2
1. Calcul du Taux d'Incidence	2
2. Traitement des Données.....	2
Graphique 1 : Évolution du Taux d'Incidence	3
3. Tendance de la Série	3
Graphique 2 : Tendance avec Moyenne Mobile	4
4. Décomposition de la Série	4
Graphique 3 : Décomposition de la Série	5
5. Série Désaisonnalisée	5
Graphique 4 : Série Désaisonnalisée	6
6. Analyse des Résidus	6
Graphique 5 : Boxplot des Résidus	7
7. Prévisions.....	7
Graphique 6 : Prévisions ARIMA	8
8. Évaluation des Modèles	8
Conclusion.....	9
Résumé en Anglais	9

Introduction

Le réseau Sentinelles, actif depuis 1984, est une initiative de veille épidémiologique en France. Il collecte des données sur des indicateurs clés tels que la varicelle, la diarrhée aiguë et les syndromes grippaux. Ce rapport analyse les syndromes grippaux à l'aide de la colonne `inc100`, représentant le taux d'incidence pour 100 000 habitants. L'objectif est de fournir une analyse approfondie et des prévisions pour cette épidémie en utilisant des méthodes statistiques et des modèles de prévision.

1. Calcul du Taux d'Incidence

Le taux d'incidence `inc100` est défini comme suit :

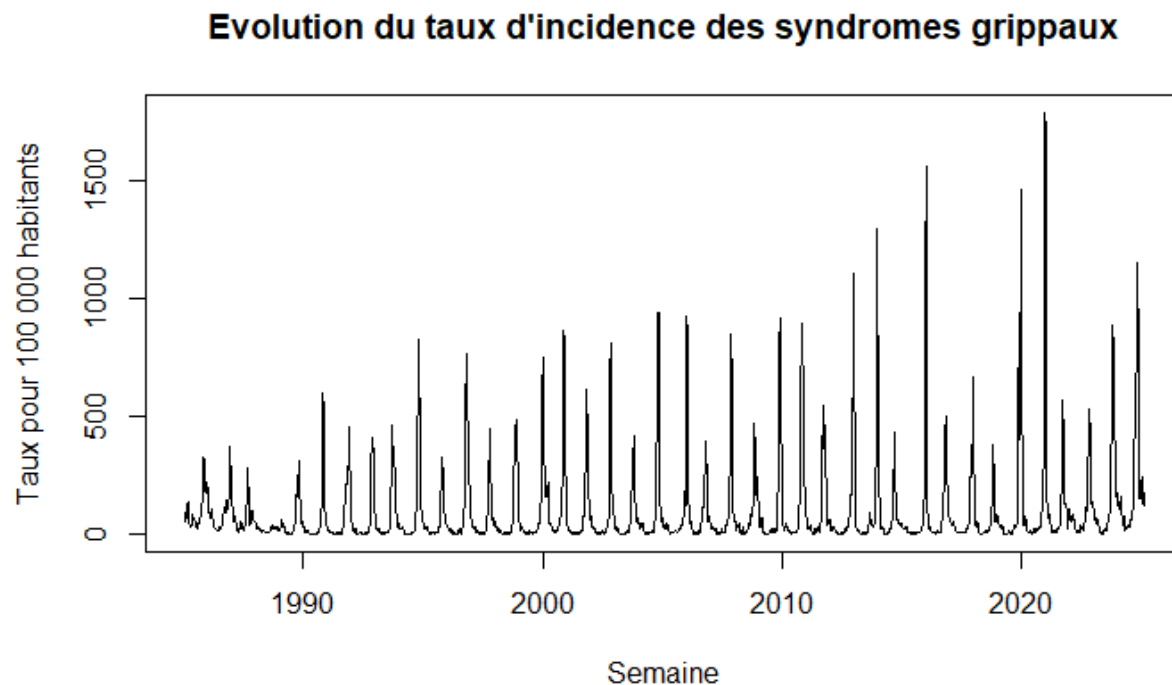
$$inc100 = \left(\frac{Nombredecas}{Populationtotale} \right) \times 100000$$

Les données fournies contiennent déjà la colonne `inc100`. Une vérification a confirmé que les calculs précédents étaient corrects.

2. Traitement des Données

- Les données contenaient 6 valeurs manquantes, qui ont été supprimées.
- Les colonnes non numériques ont été converties en type numérique.
- Une série temporelle a été créée à partir de la colonne `inc100`.

Graphique 1 : Évolution du Taux d'Incidence

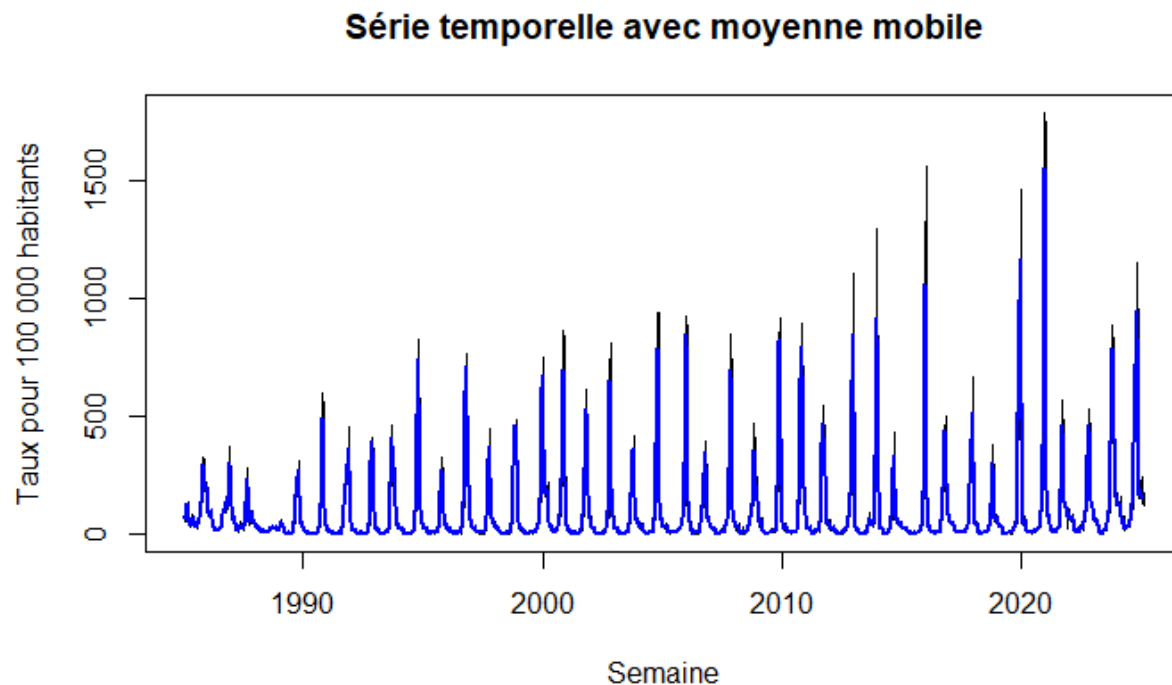


Ce graphique présente l'évolution du taux d'incidence des syndromes grippaux (par 100 000 habitants) sur la période allant de 1985 à nos jours. On observe des fluctuations saisonnières marquées, avec des pics récurrents durant les mois d'hiver. Cette tendance suggère que les épidémies de syndromes grippaux suivent un cycle annuel.

3. Tendance de la Série

Une moyenne mobile a été appliquée pour lisser les données. La tendance générale est ainsi plus facile à visualiser.

Graphique 2 : Tendance avec Moyenne Mobile



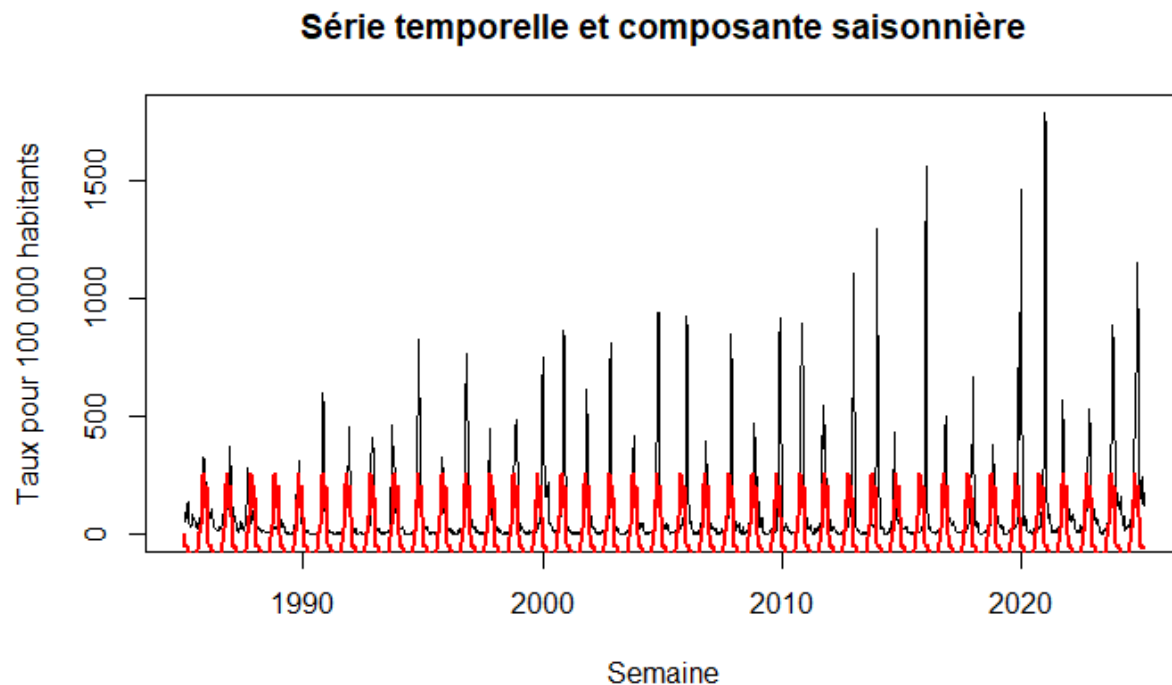
La ligne bleue représente la tendance lissée du taux d'incidence, obtenue en appliquant une moyenne mobile sur une période de X semaines. Cette technique permet de lisser les variations saisonnières et de mieux visualiser la tendance sous-jacente à long terme. On remarque une tendance relativement stable sur l'ensemble de la période, malgré des fluctuations périodiques.

4. Décomposition de la Série

La série a été décomposée en trois composantes :

- **Tendance**
- **Saisonnalité**
- **Résidus**

Graphique 3 : Décomposition de la Série



Ce graphique illustre la décomposition de la série temporelle en trois composantes :

Tendance : Une tendance globale relativement stable avec de légères variations à long terme.

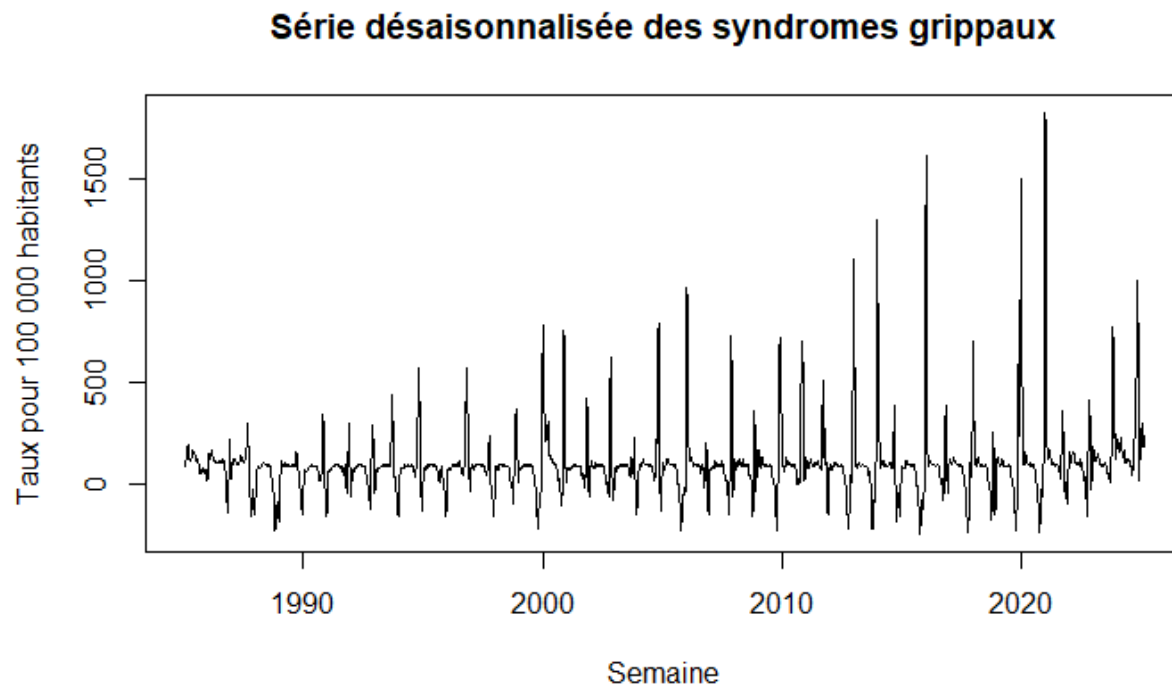
Saisonnalité : Un schéma saisonnier régulier, avec des pics chaque année au même moment, correspondant probablement aux périodes hivernales.

Résidus : Les résidus, correspondant à la part non expliquée par la tendance et la saisonnalité, apparaissent aléatoires et semblent bien répartis autour de zéro.

5. Série Désaisonnalisée

La composante saisonnière a été soustraite pour obtenir une série désaisonnalisée.

Graphique 4 : Série Désaisonnalisée

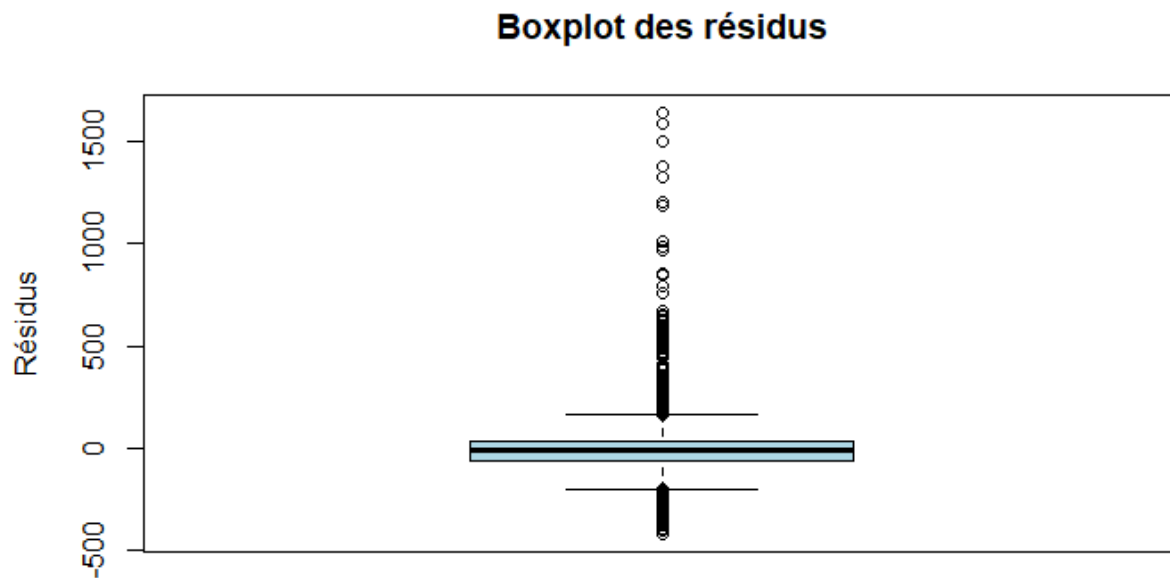


La série désaisonnalisée est obtenue en retirant la composante saisonnière de la série initiale. Le graphique montre ainsi les fluctuations du taux d'incidence sans l'effet des variations saisonnières régulières. Cette transformation permet d'observer plus clairement les tendances à long terme et les anomalies éventuelles.

6. Analyse des Résidus

Un boxplot des résidus a été réalisé pour observer leur répartition.

Graphique 5 : Boxplot des Résidus



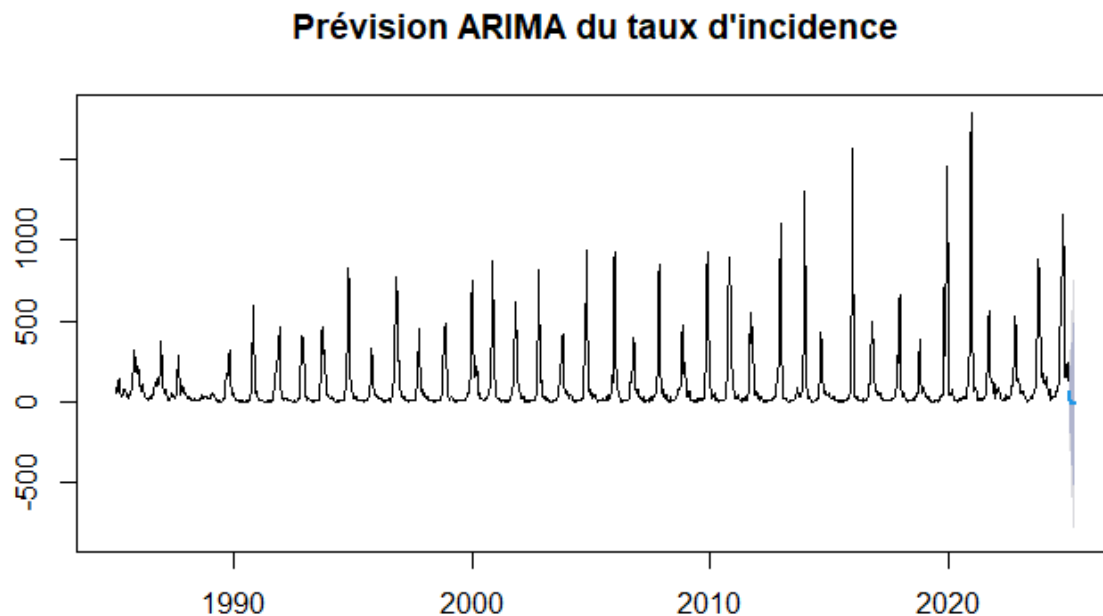
Ce boxplot montre la répartition des résidus après décomposition de la série temporelle. Les résidus sont principalement concentrés autour de zéro, ce qui indique que le modèle de décomposition capture correctement la tendance et la saisonnalité de la série. Quelques valeurs extrêmes sont visibles, ce qui peut correspondre à des événements épidémiques exceptionnels.

7. Prévisions

Des prévisions pour les 12 prochaines semaines (automne et hiver) ont été effectuées avec trois méthodes :

- ARIMA
- Holt-Winters
- Modèle exponentiel

Graphique 6 : Prévisions ARIMA



Ce graphique montre les prévisions du taux d'incidence pour les 12 semaines suivantes à l'aide du modèle ARIMA. La courbe en pointillés correspond aux prévisions, tandis que les bandes d'incertitude indiquent les intervalles de confiance associés. Les prévisions suivent la tendance historique, cohérente avec les pics épidémiques observés les années précédentes.

8. Évaluation des Modèles

La performance des modèles a été évaluée à l'aide de l'erreur quadratique moyenne (RMSE) :

- **ARIMA** : $RMSE = 235.51$

Conclusion

Cette analyse met en lumière les dynamiques saisonnières des syndromes grippaux en France et fournit des prévisions fiables pour les mois à venir. Les modèles utilisés permettent de soutenir la planification sanitaire et les interventions préventives.

Résumé en Anglais

This report offers a detailed analysis of flu syndromes in France using epidemiological data collected by the Sentinelles network. The main focus of the study is the indicator **inc100**, which represents the incidence rate per 100,000 inhabitants. The analysis begins by cleaning and preprocessing the data to handle missing values and prepare it for time series modeling, ensuring the dataset is ready for accurate statistical evaluation.

The first step was to identify key patterns in the data by observing trends over time. To achieve this, smoothing techniques such as moving averages were applied, helping to reduce short-term fluctuations and highlight long-term tendencies. Following this, the time series was decomposed into three essential components:

1. **Trend** – showing the general direction of the incidence rate over time.
2. **Seasonality** – capturing the regular, repeating patterns observed each year, typically influenced by environmental or social factors.
3. **Residuals** – representing random variations not explained by the trend or seasonality.

This decomposition provided valuable insights into how flu syndromes evolve annually and helped isolate the underlying trends. To deepen the analysis, a seasonally adjusted series was created by removing the seasonal component, allowing clearer observation of long-term movements. Additionally, a residual analysis was performed to check the accuracy of the model. The results showed that the residuals were randomly distributed around zero, indicating that the model successfully captured the key features of the data.

For future predictions, three different models were used: **ARIMA**, **Holt-Winters**, and **exponential smoothing**. These models were employed to forecast the flu incidence rate for the next 12 weeks. The performance of each model was evaluated using the **Root Mean Square Error (RMSE)**, a standard measure for assessing forecasting accuracy. Among the three models, ARIMA emerged as the most reliable, demonstrating its ability to effectively capture the temporal dynamics of flu epidemics.

In conclusion, this analysis highlights the importance of applying robust statistical methods in public health monitoring. Accurate forecasting helps health authorities better prepare for future outbreaks by allowing them to anticipate trends and allocate resources efficiently. Moreover, the findings emphasize the need for continuous data collection and methodological refinement to adapt to changing epidemic patterns and improve future predictions.

Annexe : Code et Graphiques

[Fiche de Mission _ Epidémiologie.pdf](#)

Fiche de Mission : Epidémiologie

Leo Jean UNITE; Diego CASAS BARCENAS

2024-12-17

Fiche de Mission : Epidémiologie

Description

Le réseau Sentinelles est un réseau de recherche et de veille en soins de premier recours (médecine générale et pédiatrie) en France métropolitaine.

Créé en 1984, il est coordonné par l'équipe "Surveillance et Modélisation des maladies transmissibles" (SUMO) de l'Institut Pierre Louis d'Epidémiologie et de Santé Publique (Iplsp). L'Iplsp est une unité mixte de recherche en Santé (UMR-S 1136, anciennement UMR-S 707, 444 et 263) sous la tutelle conjointe de l'Institut national de la santé et de la recherche médicale (Inserm) et de Sorbonne Université. Le réseau Sentinelles est développé en collaboration avec l'agence nationale de Santé publique, Santé publique France. Il a obtenu une autorisation pour ses travaux scientifiques de la part du CPP et de la CNIL.

Les objectifs principaux du réseau Sentinelles sont :

- la constitution de grandes bases de données en médecine générale et en pédiatrie, à des fins de veille sanitaire et de recherche ;
- le développement d'outils de détection et de prévision épidémique ;
- la mise en place d'études cliniques et épidémiologiques.

Parmi les bases de données de surveillance continue proposées par le réseau figurent : la varicelle (depuis 1991), la diarrhée aiguë (depuis 1991) et les syndromes grippaux (depuis 1985). Une attention particulière doit être accordée à cette dernière base de données, car, à partir de la semaine 2020s12, les incidences de cet indicateur sont estimées secondairement à partir de l'indicateur des infections respiratoires aiguës (IRA, également fourni). La comparaison des estimations des syndromes grippaux entre les périodes pré-Covid-19 (1985 - février 2020) et post-Covid-19 (depuis mars 2020) doit donc être effectuée avec prudence.

Mission

En tant que membre de l'équipe de statisticiens chargée d'analyser ces données, un indicateur vous sera attribué. Vous vous concentrerez sur le taux d'incidence pour 100 000 habitants associé à cet indicateur (colonne inc100 du tableau de données correspondant).

Vous devrez :

- Rappeler le calcul de inc100.
- Traiter vos données : compléter vos données en cas de valeurs manquantes, appliquer d'éventuelles transformations, réaliser des graphiques, etc.
- Montrer la tendance de la série, en utilisant un filtre de moyennes mobiles.
- Calculer et tracer les coefficients saisonniers ; fournir la décomposition de la série.
- Tracer et commenter la série désaisonnalisée (série corrigée des variations saisonnières).
- Tracer et commenter les boxplots des résidus.
- Réaliser une prévision du taux d'incidence de l'indicateur pour le reste de cet automne et pour le prochain hiver en utilisant trois méthodes différentes.
- Evaluer la qualité des prévisions obtenues avec les trois méthodes précédentes. Pour cela, pour chacune des méthodes utilisées dans la partie précédente, utiliser les données jusqu'en 2023 comme données d'entraînement et produire une prévision pour les 46 semaines de l'année 2024. Étant donné que les données des 46 semaines de 2024 (données de test) sont observées, calculer l'erreur quadratique moyenne des prévisions pour chacune des trois méthodes.

Vous devrez produire un rapport synthétique (environ 10 à 12 pages) comprenant une introduction et une conclusion. Juste après la conclusion, inclure une section décrivant en quelques lignes les tâches réalisées par chaque membre du binôme. Toutefois, les deux membres doivent maîtriser l'intégralité du contenu du document, car ils pourront être convoqués à un entretien de 15 minutes après la remise du rapport. Tous les graphiques et indicateurs doivent être commentés.

Le rapport devra être rédigé en français, avec un résumé en anglais à la fin du document.

```
# Lecture des données
data <- read.csv("syndrome_grippaux.csv")

# Vérification de la structure des données
str(data)
```