

Rapport Series Temporelle

FRANCOIS Gautier, VALLEE Paul et DESVALLEES Augustin

2024-05-21

Contents

1) Introduction	1
2) Estimation de la tendance et de la saisonnalité	1
2.1) Visualisation des données	1
2.2) Recherche de la tendance	3
2.3) Recherche de la saisonnalité	3
3) Estimation par régression linéaire	4
4) Etude des résidus	6
5) Ajustement du modele complet	7
6) Conclusion	8

1) Introduction

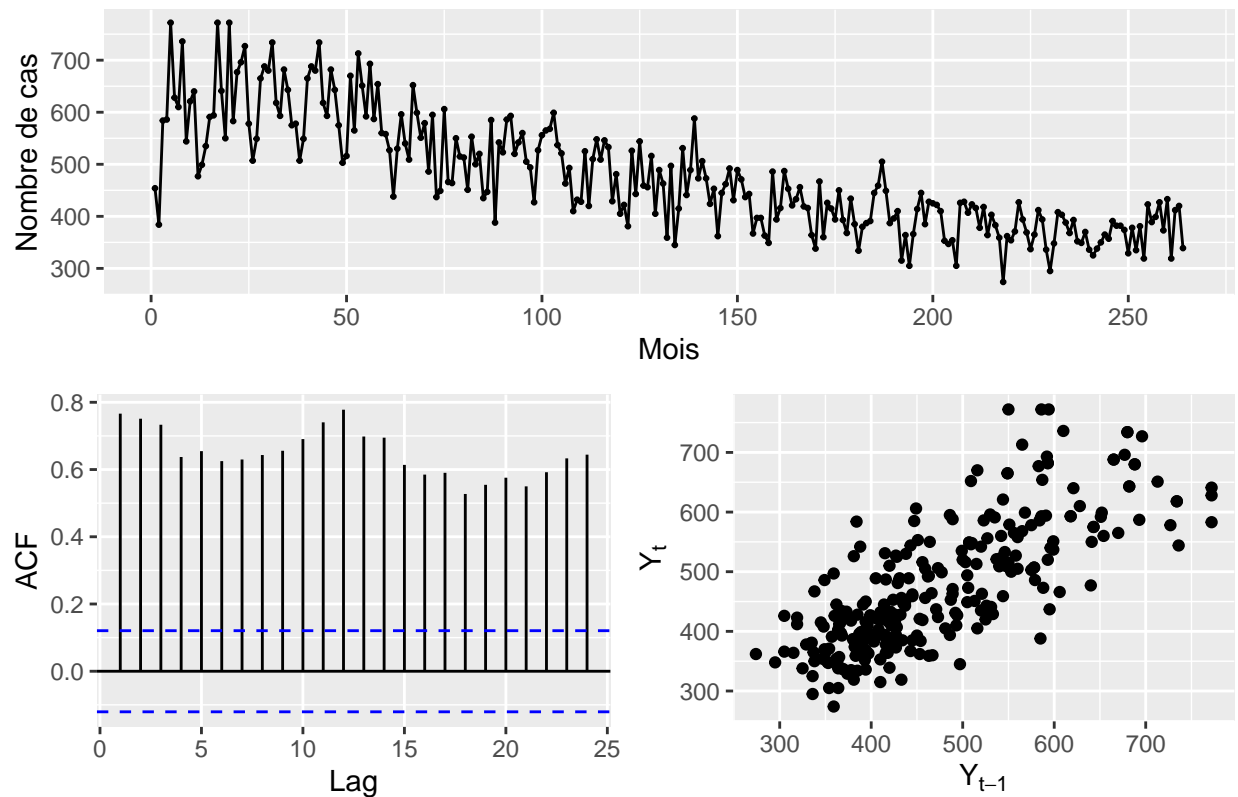
L'objectif de ce projet est d'analyser les données mensuelles des cas de pneumonie à Hong Kong entre 1997 et 2018. Cette analyse vise à identifier les tendances, la saisonnalité, et à construire un modèle prédictif qui peut aider à anticiper les futurs cas de pneumonie. Pour ce faire, nous utiliserons des techniques de séries temporelles et de régression linéaire.

2) Estimation de la tendance et de la saisonnalité

2.1) Visualisation des données

La première étape a consisté à visualiser ces données pour obtenir un aperçu général de l'évolution des cas de pneumonie au fil du temps.

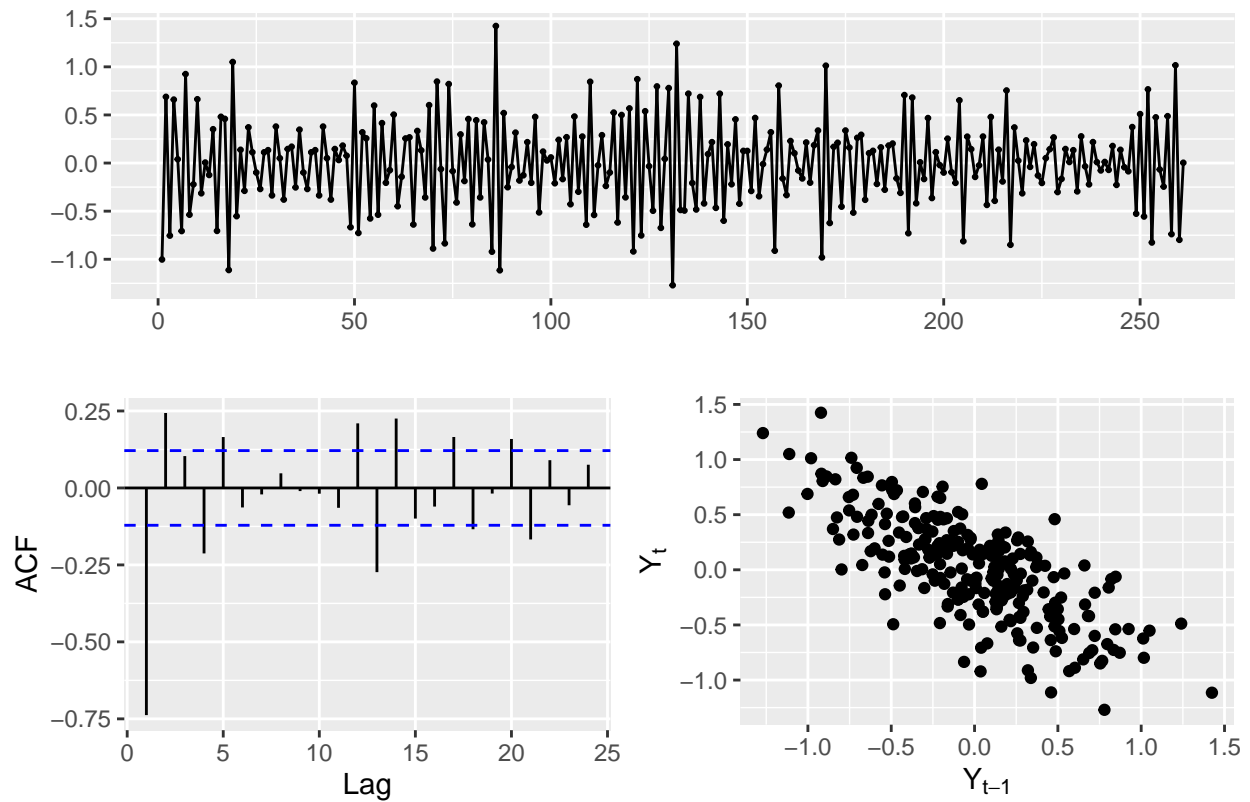
Chronogramme des cas de pneumonie à Hong Kong



Après avoir examiné le chronogramme, deux observations principales ressortent : une tendance générale à la baisse et une saisonnalité annuelle avec des pics marqués pendant les mois d'été. Ces caractéristiques se reflètent également dans les autocorrélations, montrant des pics positifs forts et un schéma périodique avec une période de 12 mois. De plus, il semble y avoir une réduction de l'amplitude des pics, ce qui peut nécessiter une transformation logarithmique des données pour corriger ce phénomène.

Maintenant, vous allez explorer les séries différenciées. Cela implique de calculer les variations entre les observations successives afin de rendre les données stationnaires, éliminant ainsi les tendances et les composantes saisonnières, ce qui permet une analyse plus claire des motifs résiduels.

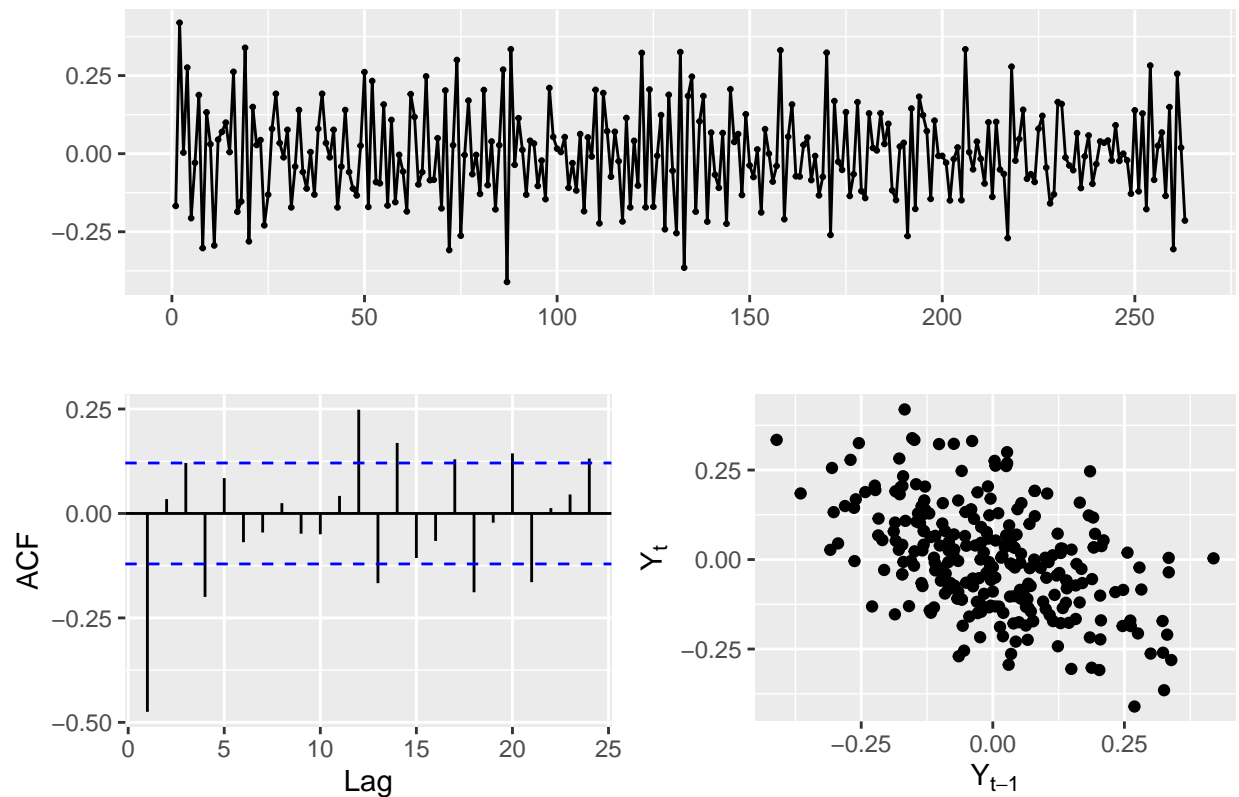
2.2) Recherche de la tendance



Après avoir effectué plusieurs tests pour trouver la meilleure différenciation, nous avons choisi de différencier les données trois fois pour obtenir une série stationnaire. Cela a permis de réduire la tendance et de rendre les données plus stables.

2.3) Recherche de la saisonnalité

Maintenant, nous allons rechercher la saisonnalité dans les données différenciées.



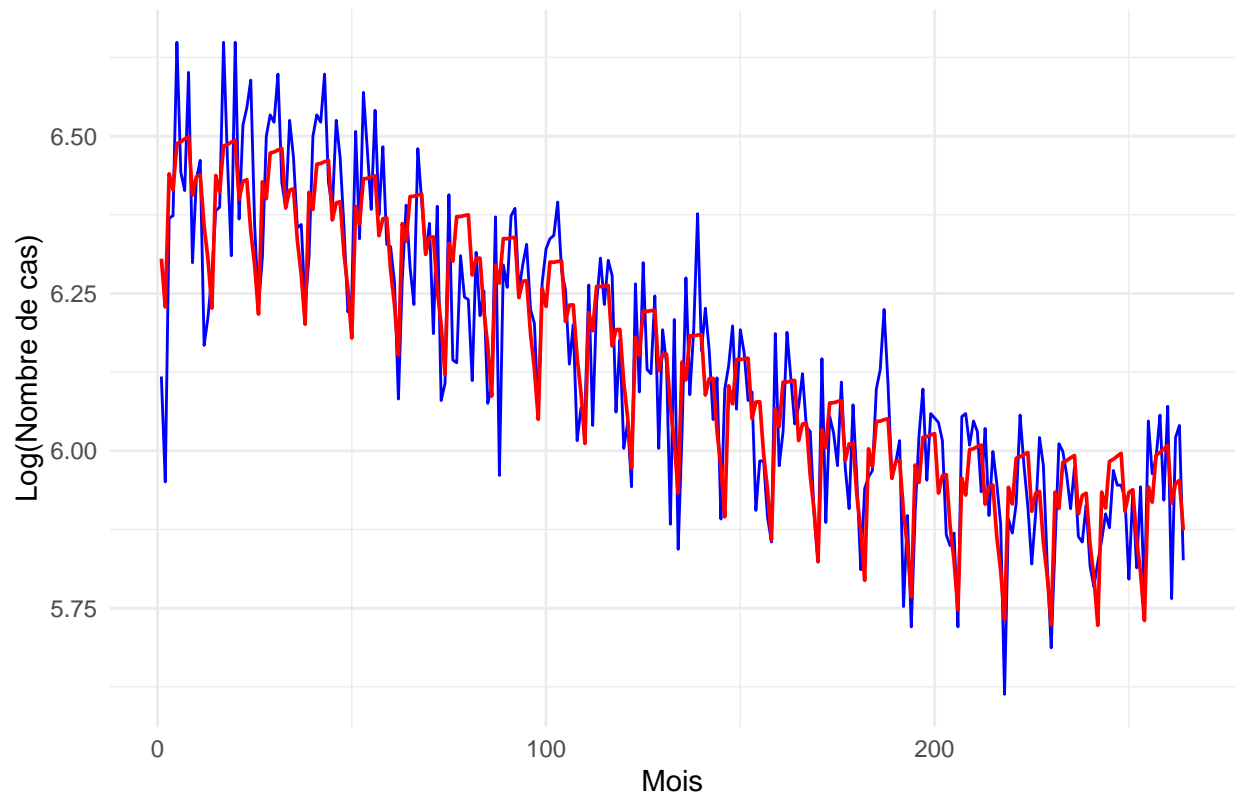
Après avoir différencié les données par un décalage de 12 mois, nous avons pu observer une série stationnaire avec une saisonnalité claire. Cela confirme la présence d'une composante saisonnière annuelle dans les cas de pneumonie à Hong Kong.

3) Estimation par régression linéaire

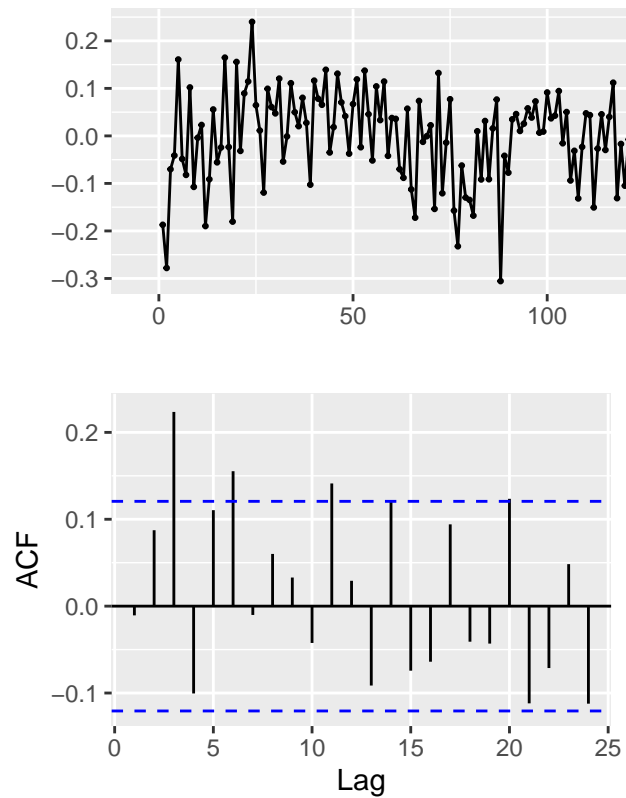
Enfin, nous allons construire un modèle prédictif en utilisant une régression linéaire multiple. Nous allons inclure des variables explicatives telles que le temps, le temps au carré, et des fonctions trigonométriques pour capturer les tendances et la saisonnalité.

```
## Warning: Using 'size' aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use 'linewidth' instead.
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.
```

Modèle de régression linéaire multiple



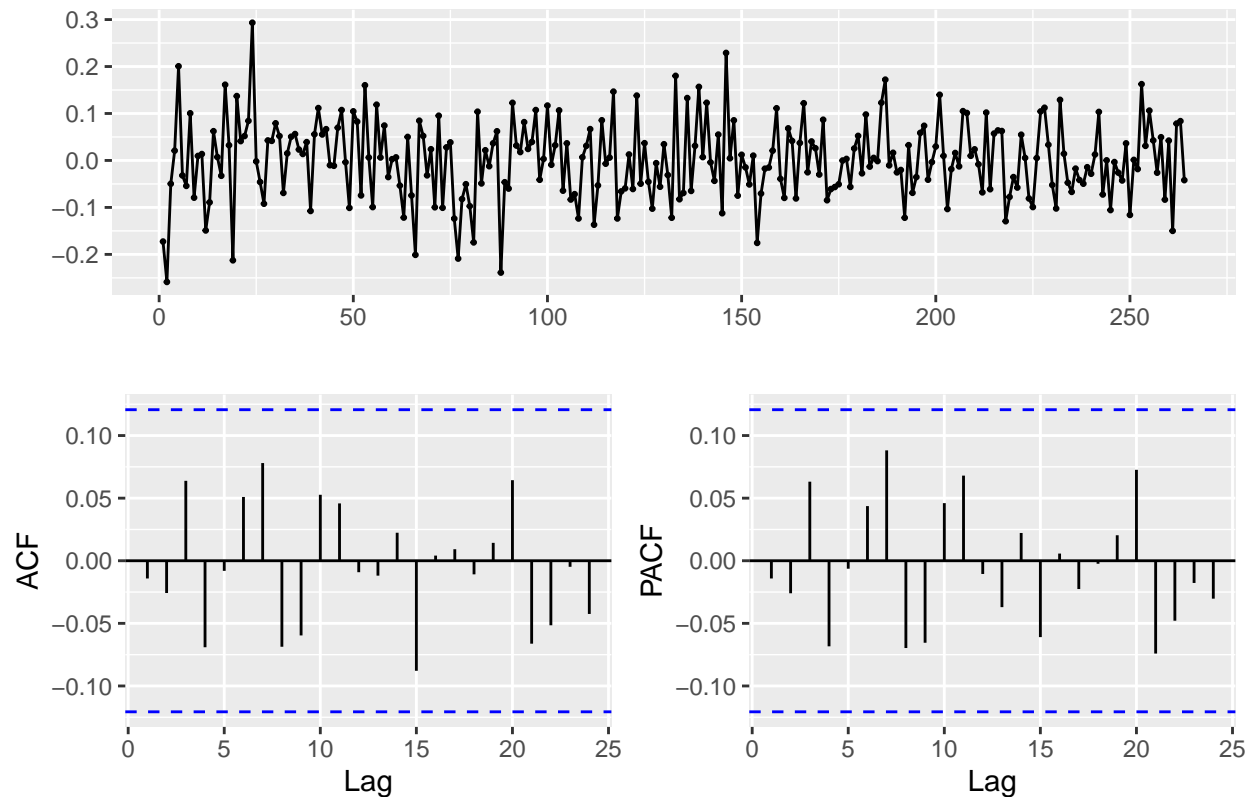
4) Etude des résidus



Observons maintenant les résidus du modèle pour affiner notre analyse.

Commençons par observer le PACF des résidus. On a un coefficient significatif pour le premier décalage, ce qui indique une autocorrélation résiduelle. On décide donc d'essayer d'ajuster un modèle $\text{ARIMA}(3,0,0)$. Après cette essai, nous nous sommes rendu compte que le modèle $\text{ARIMA}(3,0,0)$ n'était pas le plus adapté. Nous avons donc décidé de tester plusieurs modèles ARIMA pour trouver le meilleur ajustement.

Le modèle $\text{ARIMA}(3,0,3)$ nous a semblé être le plus adapté. En effet, comme on peut l'observer sur le graphique ci-dessous, les résidus sont stationnaires et ne présentent pas d'autocorrélation significative.



Ensuite, on effectue un test de blancheur sur les résidus du modèle ARIMA(3,0,3).

```
##
## Box-Pierce test
##
## data: ajustement_residus$residuals
## X-squared = 14.503, df = 25, p-value = 0.9522
```

On obtient un p-value de 0.9522, donc on conserve l'hypothèse nulle $H_0: p(1)=p(2)=\dots=p(25)=0$. Nous avons donc déterminé que le modèle ARIMA(3,0,3) serait approprié pour ajuster nos données en optimisant les paramètres via la méthode du maximum de vraisemblance.

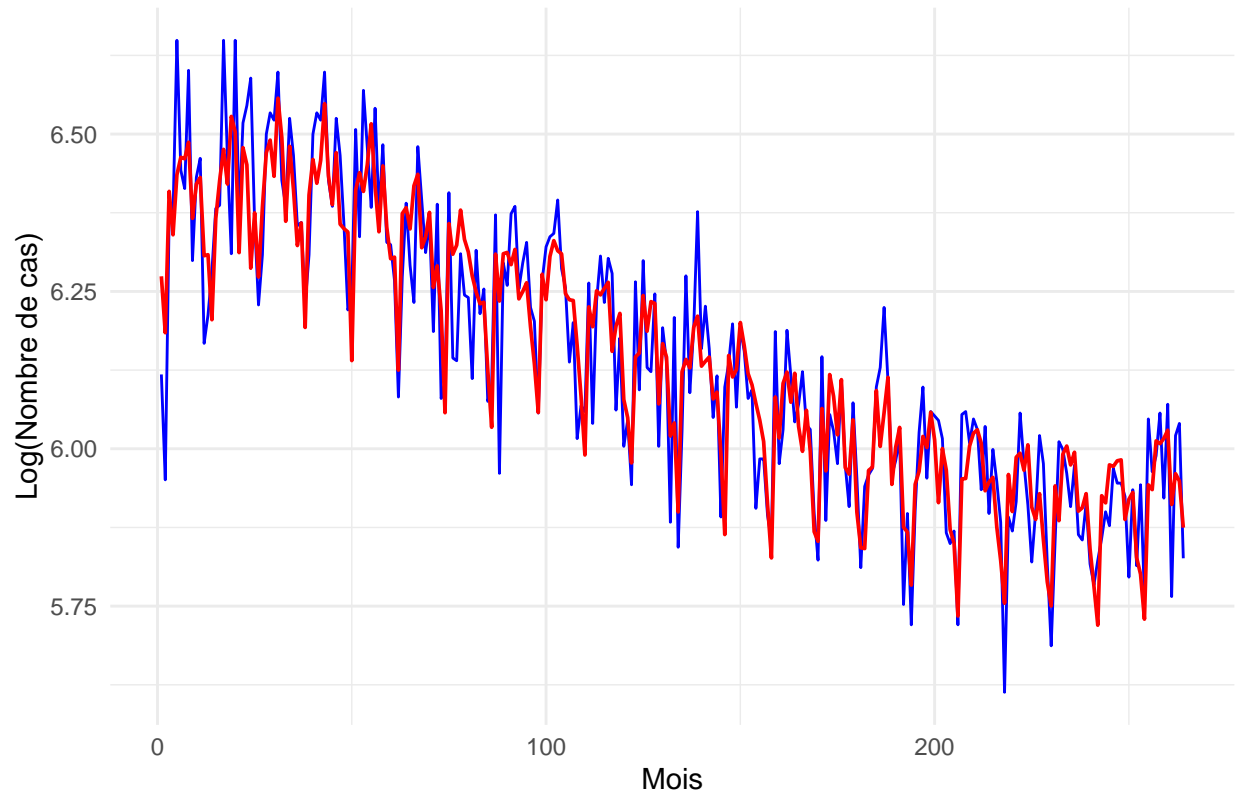
5) Ajustement du modele complet

Ajustons maintenant le modèle complet en utilisant les données différenciées et les variables explicatives. Afin de vérifier la qualité de l'ajustement, nous allons effectuer un test de blancheur sur les résidus du modèle complet.

```
##
## Box-Pierce test
##
## data: ajustement_complet$residuals
## X-squared = 16.755, df = 25, p-value = 0.8905
```

Avec une valeur p de 0.8905, nous ne rejetons pas l'hypothèse nulle au niveau de signification typique de 0.05. Cela suggère qu'il n'y a pas suffisamment de preuves pour conclure que les résidus présentent une autocorrélation significative au niveau de signification de 5 %. Par conséquent, le modèle ajusté semble être approprié pour les données de cas de pneumonie à Hong Kong.

Modèle ajusté ARIMA(3,0,3) avec variables explicatives



6) Conclusion

Grace à l'analyse des données de cas de pneumonie à Hong Kong, nous avons pu identifier une tendance et une saisonnalité claire dans les données. En utilisant une régression linéaire multiple et un modèle ARIMA(3,0,3) avec des variables explicatives, nous avons pu ajuster un modèle prédictif qui capture les tendances et la saisonnalité des cas de pneumonie.

Le modèle ajusté semble être approprié pour les données, avec des résidus stationnaires et une autocorrélation non significative. Cela suggère que le modèle peut être utilisé pour prédire les cas de pneumonie à Hong Kong avec une certaine précision.

Cependant, il convient de noter que les prévisions peuvent être influencées par d'autres facteurs non pris en compte dans le modèle, tels que les politiques de santé publique, les conditions météorologiques, et d'autres variables externes. Par conséquent, il est important de prendre en compte ces facteurs lors de l'interprétation des prévisions et des résultats du modèle.

Nous avons malheureusement pas réussi à intégrer les données de température et d'humidité dans notre modèle. Cela aurait pu nous permettre d'obtenir des prévisions plus précises en tenant compte de ces facteurs externes. Cependant, cela reste une piste d'amélioration pour de futures analyses.