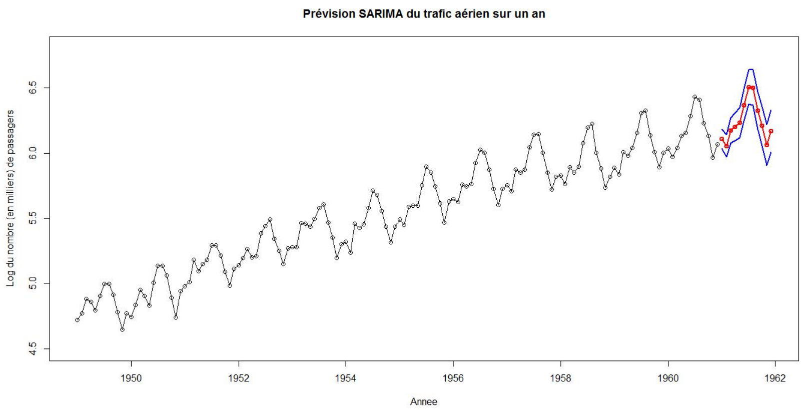
Rapport Série Temporelle TP3

WANG Andi

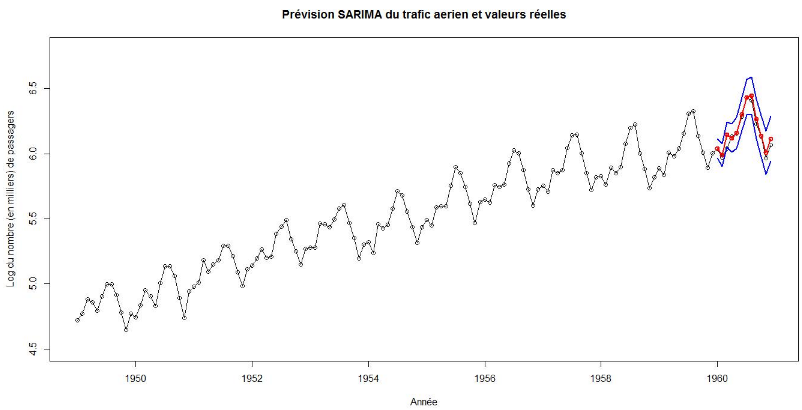
Partie 1 :

Partie exploratoire :

On fait une prévision avec tous les données pour les prochaine années et obtenir une image, ce sont tous prédiction et on n'a pas de données exactes pour les confirmer.



Et puis on utilise des codes pour faire une Back-testing sur log du nombre, et on enlève les dernières valeurs pour tester :

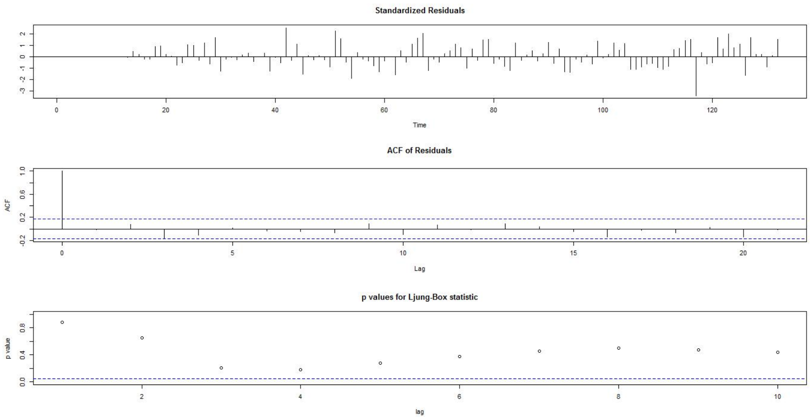


D'après cette image, on peut dire que les codes marche bien pour des prédiction sur des données de log du nombre. Il y a seulement un point de suspension des 12 points, ce point est sur les bornées de intervalles de confiance.

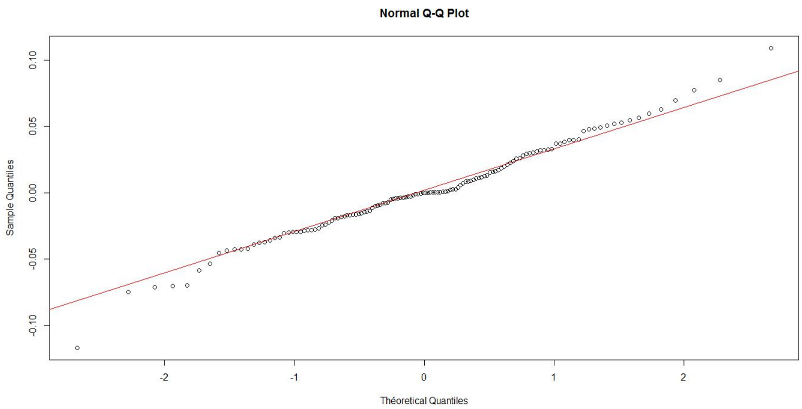
Utilisation du modèle pour faire de la prévision : on enlève les 12 dernières valeurs que l'on cherche ensuite à prédire (back-testing)

On remplit la partie «travail à faire» en utilisant la même méthode de la partie en seulement remplissant les données de log du nombre avec les données du nombre, et puis on fait une même modèle arima (p=0,d=1,q=1) $(P=0,D=1,Q=1)_{s=12}$ comme le précédent et obtient les images :

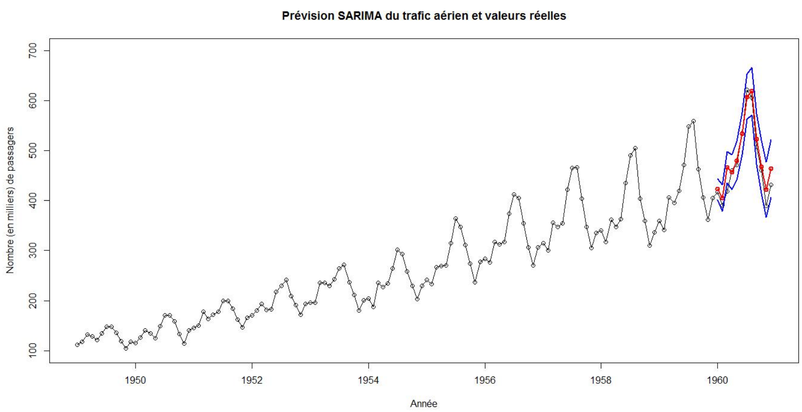
Pour le résidus, on a l'image :



Pour adapter aux quantile, on a l'image :



Et puis on obtient des résultats et en comparant avec des prédiction :



On observe l'image, et on peut trouver qu'il y a 3 points très poche de bornées des intervalles ou hors des intervalles, donc, il correspond bien le sujet que l'on a 3 points de suspension.

Modèle linéaire pour tendance et saisonnalité et comparaison avec B&J :

Pour faire une modèle linéaire, on peut utiliser le method :

$$X_t = M_t + S_t + U_t$$

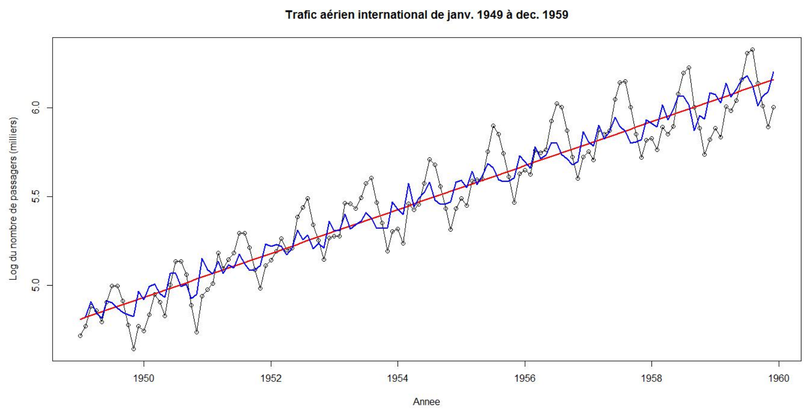
Avec $M_t = \exp(m_t)$,$S_t = \exp(s_t)$,$U_t = \exp(u_t)$

Donc pour le modèle linéaire de log(X_t), on a :

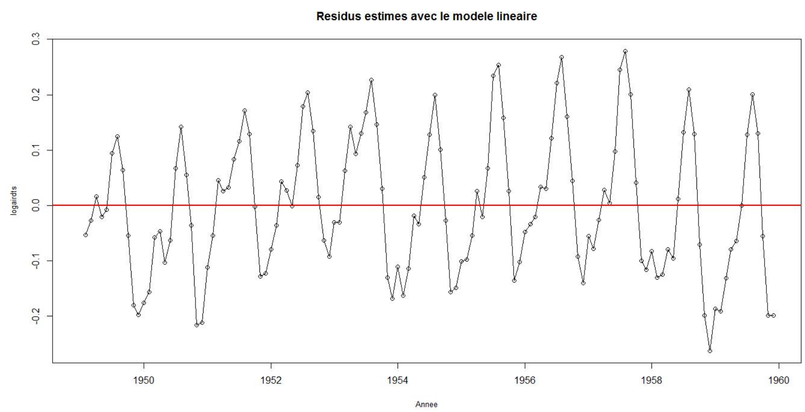
$$\log(X_t) = m_t + s_t + u_t$$

Avec m_t on peut considérer comme $m_t = \alpha t$ pour la tendance, et s_t pour la saisonnalité et peut estimer par diff(log (X_t)) et u_t et une terme aléatoire et peut être directement obtenu par rnorm(n).

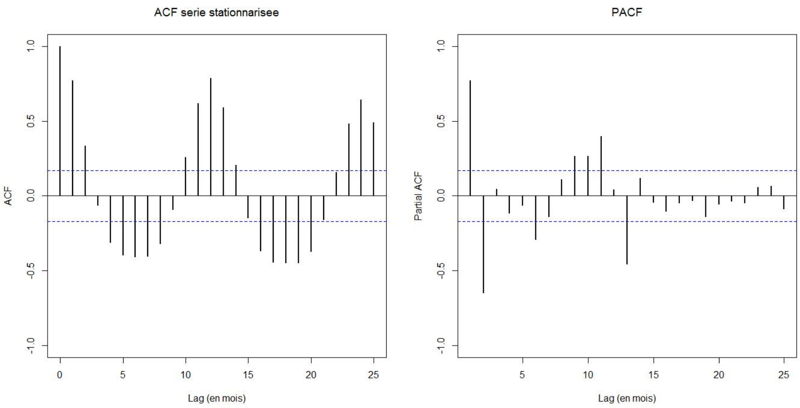
Pour capter la tendance et insérer la figure montrant le log de la série, sa tendance et sa saisonnalité :



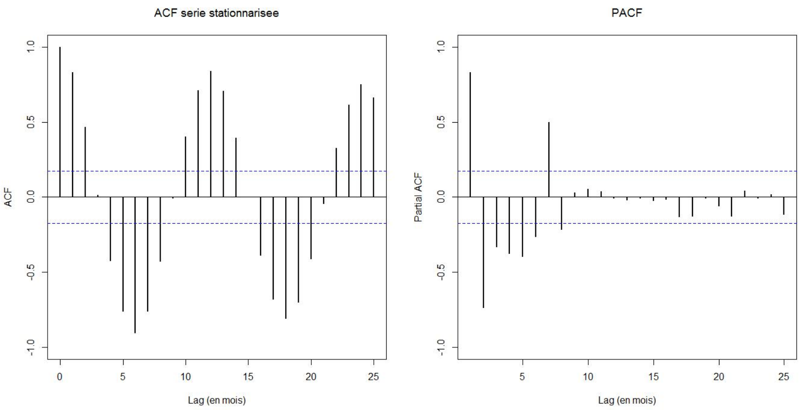
Et la figure de sans tendance et désaisonnalité :



Et puis on analyse les image ACF et PACF :

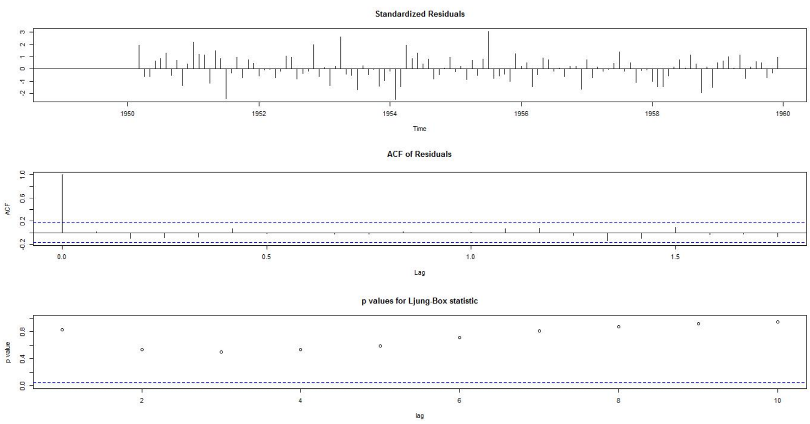


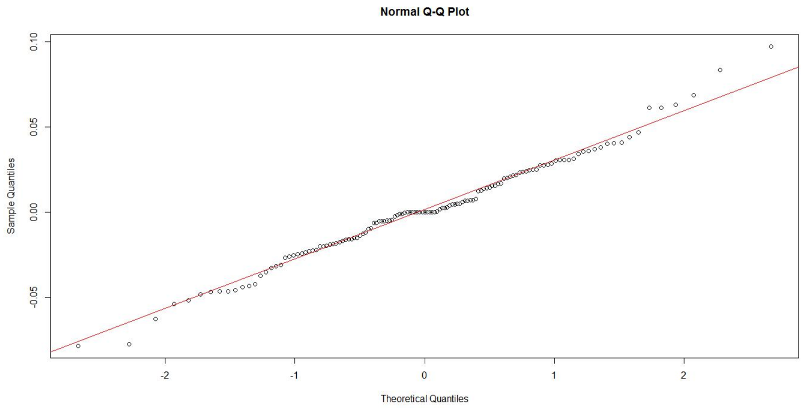
Et on peut trouver que sur des image, on a $\boldsymbol{\pi(1)}, \boldsymbol{\pi(2)}, \boldsymbol{\pi(6)}, \boldsymbol{\pi(9)}, \boldsymbol{\pi(10)}, \boldsymbol{\pi(11)}, \boldsymbol{\pi(13)}$ sont significativement non nuls, et puis on fait une diff de lag=6 pour la série :

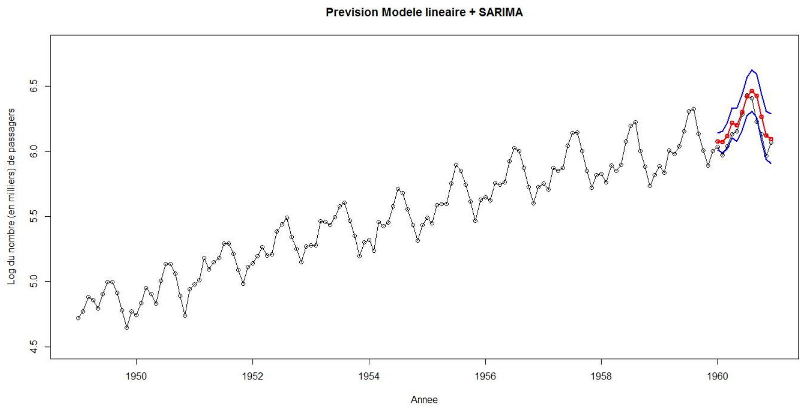


Il y a aussi une variation saisonnaire de période =12.

Et d'après ces images, on peut établir une modèle de arima (p=0,d=1,q=1)(P=0,D=1,Q=2)$_{s=12}$ pour estimer la prédiction et on a des images :

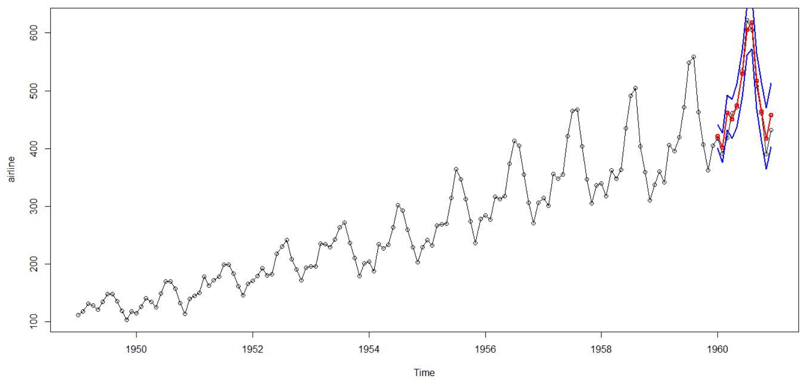






Il y a 3 points de suspension mais ce sont proche des intervalles de confiance 95%, donc la prédiction d'après cette modèle peut bien marcher.

on fait la même modèle pour la série initiale et on obtient l'image :



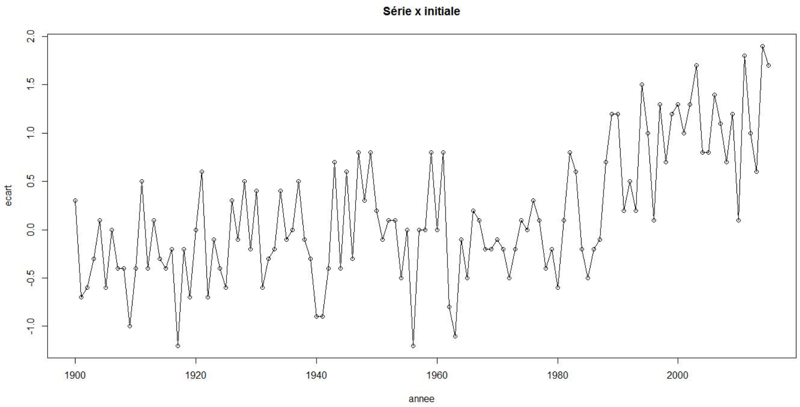
Il y a un points de suspension mais les intervalles sont plus grandes que les précédentes. Donc le modèle linéaire avec la modèle de arima peut bien marche pour la série.

Partie 2 : (analyse de données de température en France)

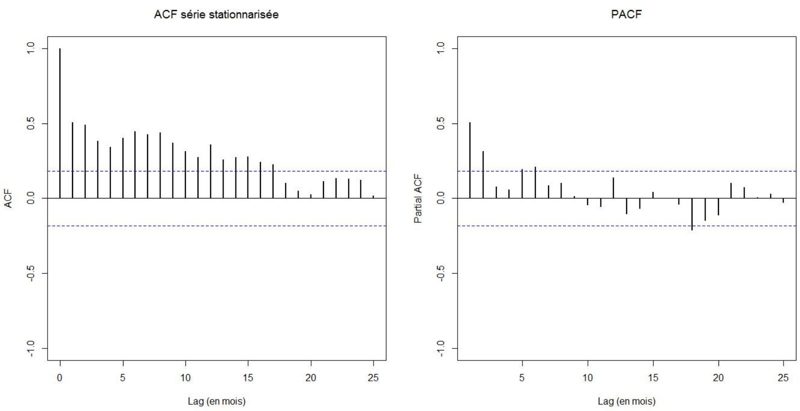
On s'intéresse à l'évolution de la température moyenne en France métropolitaine sur la période 1900-2015 (source : Météo-France) – Pour cela, on considère la série qui donne l'écart de température (en °C) avec la moyenne calculée sur la période de référence 1961-1990.

Etudier cette série (fichier ecart.txt disponible sur Campus) avec comme objectif de répondre à la question du réchauffement en France ces dernières décennies. On se demandera en particulier si le réchauffement observé est statistiquement significatif et non pas simplement le fait du hasard.

On d'abord étudier la série initiale :

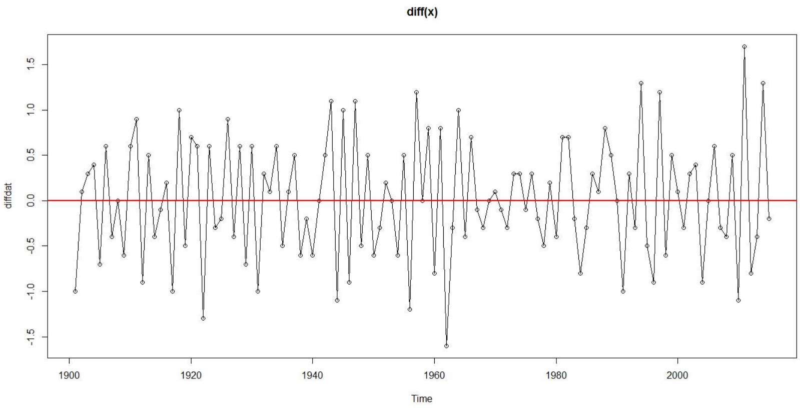


et obtenir l'ACF et le PACF :

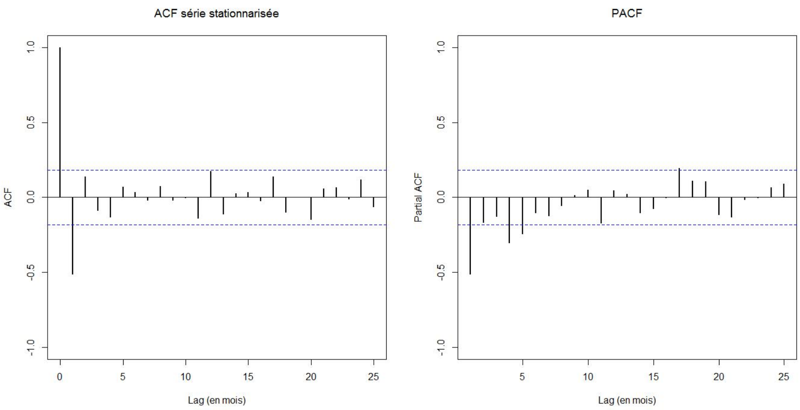


A ce moment, la série n'est pas stationnaire, donc on va traiter la série pour analyser.

Et puis on fait un diff sur la série initiale et on a des images :



On peut trouver que la série nouvelle est presque stationnaire, et puis on analyser l'ACF et le PACF et obtient :



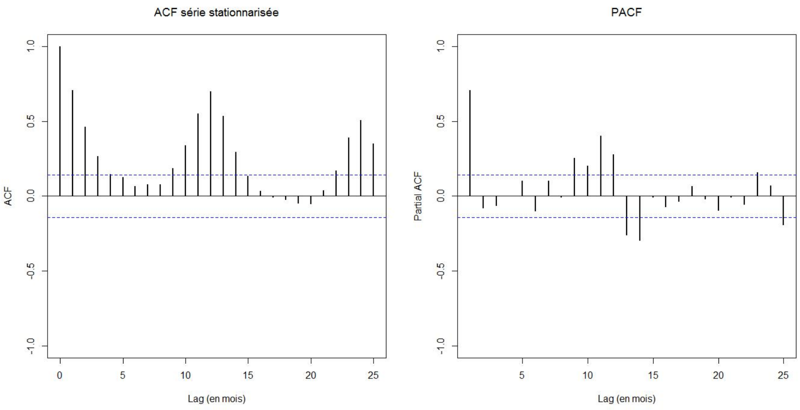
On peut trouver que $\boldsymbol{\rho(1) = \pi(1)}$, en plus $\boldsymbol{\pi(4)}$ et $\boldsymbol{\pi(5)}$ sont aussi significativement non nul, donc, on peut analyser la série par ARMA(0,1) ou ARMA(0,5). D'après la modèle on peut conclure que la série est statistiquement significatif et non pas simplement le fait du hasard.

Partie 3 : (analyse d'une série au choix) A chaque fois, on donnera les principales étapes de l'analyse : choix du modèle à partir des ACF et PACF, qualité de l'ajustement et prévision obtenue par back-testing. Commentaires. On utilisera le package « datasets » de R (sous R, exécuter la commande library(help="datasets") pour plus d'information sur les jeux de données disponibles).

Etudier la série UKDriverDeaths (Road Casualties in Great Britain 1969-84). Etudier enfin une série de votre choix.

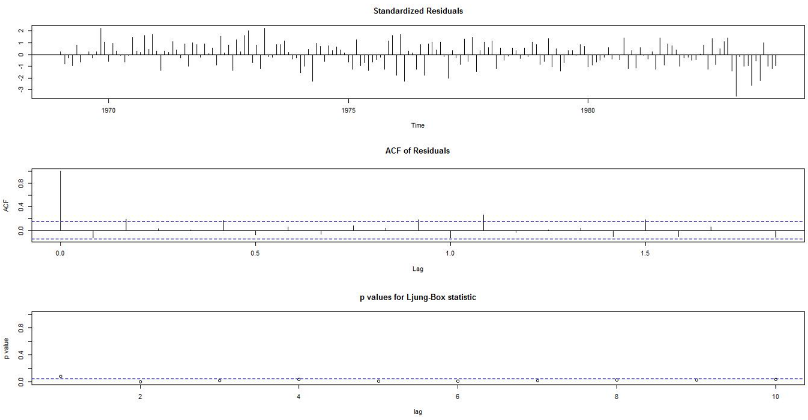
Avant d'analyser, on d'abord fait une nouvelle série avec log de la série initiale pour mieux observer et analyser des données.

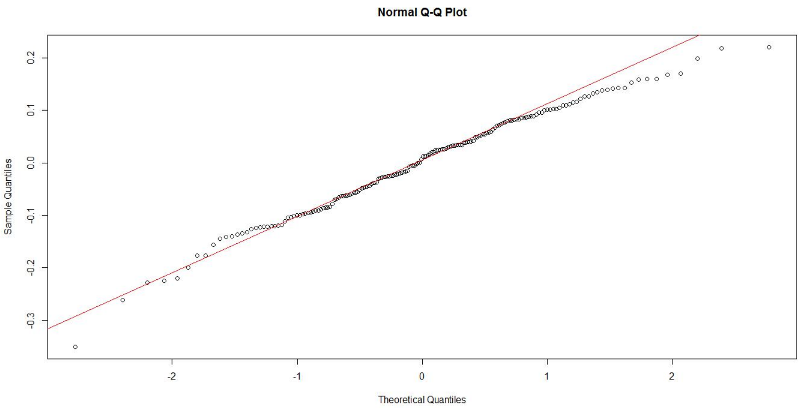
On d'abord étudie la série de UKDriverDeaths :



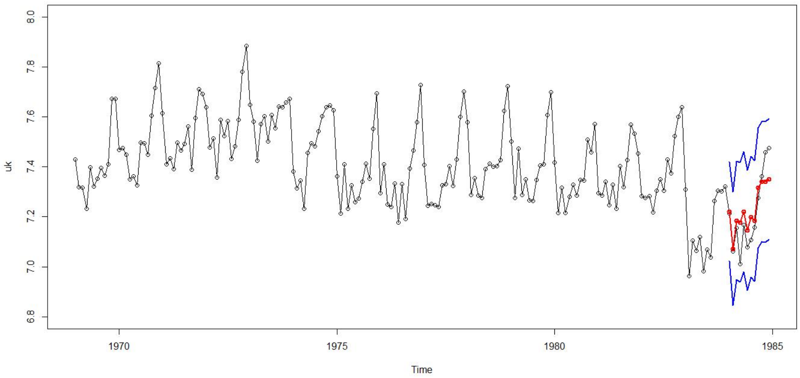
C’est évident que des données varie d’après une période de 12, et pour une modèle arima, on a évidamement que p=1 et q=0,d=0 puisque $\pi(1)$ est significativement non nul et le prochaine $\pi(n)$ significativement non nul est n=9. Donc on a le modèle :

Arima(1,0,0)(1,0,0)$_{s=12}$ et puis on va tester le performance de la modèle.





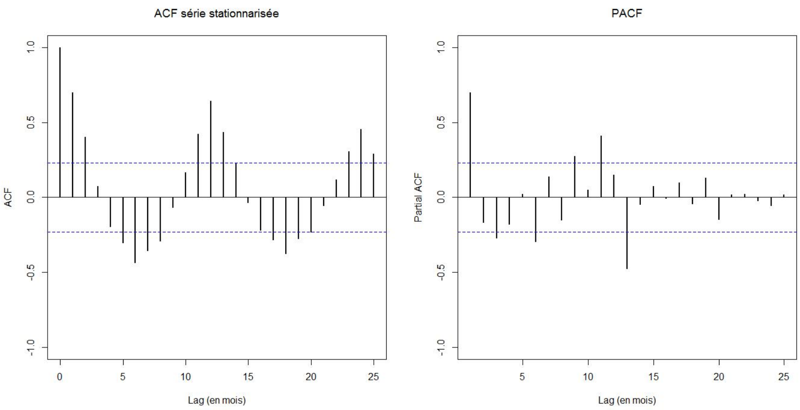
On enlève les dernières 12 valeurs pour tester, et on a des résultats de prévision :



Tous les points sont dans les intervalles de confiance=95%, il n'y a pas de point de suspension, mais on peut aussi trouve que les intervalles de prediction sont grandes comme on a déjà utilisé le log de la série.

Donc on peut conclure, malgré la série statistique, elle n'est pas très bien stationnaire pour analyser et les données sont plus aléatoire en comparant avec des modèles précédentes.

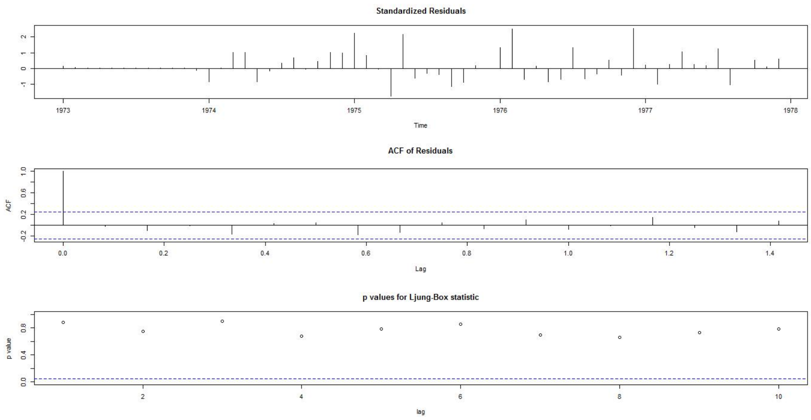
Et puis on utilise des données de ***USAccDeaths*** pour analyser les morts dans les accidents aux Etat-Unis :

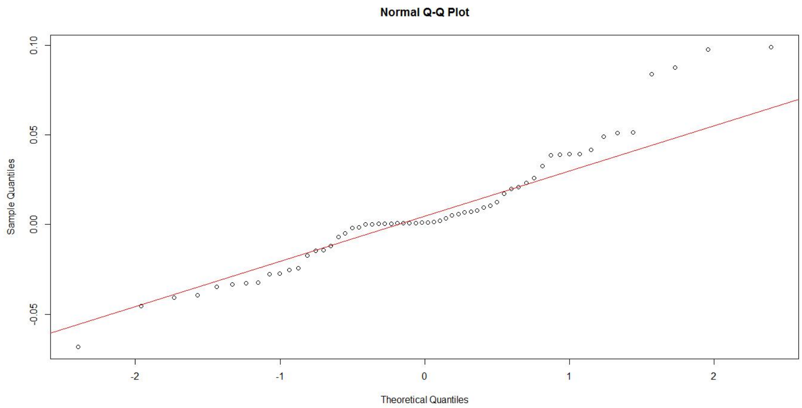


D'après cette image, on peut trouver que pour PACF, n=1, n=11 et n=13 sont significativement non nuls et donc on peut prevoir une période de 12 en visant même ACF comme il varie d'une période de 12, donc on a p=0,d=1 et q=1, et c'est une modèle similaire de la partie 1, donc on a le modèle :

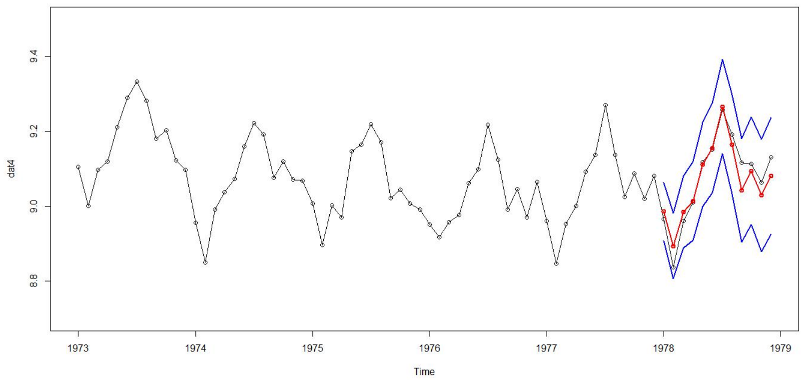
Arima(0,1,1)$(0,1,1)_{s=12}$

Et puis on va tester cette modèle :





D'après ces analyse, on peut savoir que la modèle peut être utilisée pour ces données et puis on enlève les dernière 12 valeurs pour tester la performance de prédiction et on a des résultats :



Donc la modèle est bien pour la prédiction, des résultats de prédiction sont tous bien correspondus aux données exactes.