

RAPPORT : SÉRIES TEMPORELLES

Jonathan El Methni

jonathan.el-methni@parisdescartes.fr

Instructions :

Commencez par télécharger et installer les packages suivants :

- *caschrono*
- *forecast*

Vous me rendrez votre rapport au format PDF ou HTML exclusivement en me l'envoyant à mon adresse mail. Veillez à bien utiliser votre adresse de Dauphine. Je vous répondrai pour vous signifier que je l'ai bien reçu. Si jamais ce n'est pas le cas renvoyez le moi s'il vous plaît. Sur MyCourse vous trouverez un fichier Rmarkdown. Je vous expliquerai au prochain cours comment l'utiliser. Ce n'est pas obligatoire mais il vous sera d'une grande aide pour rédiger votre rapport.

Vous devez traiter chacun au moins 2 jeux de données dont au moins un nouveau jeu de données. Par traiter j'entends mettre en application ce que l'on a vu tout au long des 3 cours à l'aide de R. Vous pouvez aller plus loin et cela sera valorisé.

Il ne suffit pas de faire *auto.arima* sur la série ou *ets*. J'attends de vous que vous "dépliez" la série petit à petit en regardant les résidus obtenus à chaque étape et en commentant. Cela vous permettra de justifier votre choix pour les ordres de vos modèles que vous pourrez comparer avec ceux obtenus grâce aux fonctions R.

Attention le modèle obtenu par une sélection R n'est pas nécessairement le meilleur ! Il est possible d'avoir plusieurs modèles concurrents. Auquel cas il serait intéressant de les comparer ainsi que leurs prévisions. Il vaut mieux que vous mettiez moins de graphiques mais que vous les commentiez tous ainsi que toutes vos sorties R.

Le rapport n'a pas besoin d'être très long. Essayez d'être synthétique. Pas besoin de tout faire figurer (graphiques et sorties R). Vous pouvez vous contenter d'expliquer ce que vous avez obtenu. N'oubliez pas d'enlever la dernière année afin de pouvoir comparer les différentes méthodes de prédictions (lissage et modélisation) avec les vraies valeurs.

Tous les jeux de données se trouvent soit sur R soit sur MyCourse soit en ligne. En cas de problème n'hésitez pas à m'écrire un mail. Enfin je vous encourage vivement à travailler en binôme auquel cas je vous demanderai de me rendre 3 jeux de données traités.

Jeux de données :

Les nouveaux jeux de données sont :

- *sncf*
- *multiTimeline*
- *c7ex2*

Les anciens jeux de données sont :

- *co2*
- *USAccDeaths*
- *trendstatistique*

Autres jeux de données :

Je vous encourage à aller sur <https://trends.google.fr/>. Sur ce site il vous suffit de taper un mot dans la barre de recherche. Vous obtiendrez alors diverses statistiques dont la série temporelle de l'occurrence de ce mot dans les recherches Google depuis 2004. Vous pouvez alors créer votre propre jeu de données. Le jeu de données *multiTimeline* a été obtenu de cette façon.

1 *sncf*

Nous nous intéressons dans ce jeu de données à l'étude du nombre de voyageurs sur le réseau SNCF. Les données sont issues du site <https://freakonometrics.hypotheses.org>. On commencera par le charger .

```
> sncf=read.table("http://freakonometrics.free.fr/sncf.csv",header=TRUE,sep=";")
> train=as.vector(t(as.matrix(sncf[,2:13])))
> X=ts(train,start = c(1963, 1), frequency = 12)
```

2 *multiTimeline*

On s'intéresse à l'évolution de l'intérêt pour le mot-clef football sur Google trend entre le 1er janvier 2004 et le 23 janvier 2019.

```
> data <- read.csv('multiTimeline.csv')
> X <- ts(data,start=c(2004,1),frequency=12)
```

3 *c7ex2*

Nous nous intéressons dans ce jeu de données à l'évolution sur 10 ans du nombre d'immatriculations de voitures particulières en France.

```
> library(readxl)
> immat <- read_excel("c7ex2.xls")
> X <- ts(immat[!is.na(immat[,2]),2],frequency = 12)
```