

S6_2_series_chronologiques

Table of contents

1	Introduction aux séries chronologiques	1
1.A	Modèle auto-régressif d'ordre 1 :	2
1.A.1	Qu'est-ce qu'une régression (dans ce contexte) ?	2
1.A.2	Pourquoi dit-on souvent que « $a = 0$ » ?	2
1.A.3	Variance : définition générale	3
1.A.4	Variance du bruit : ce que c'est exactement	3
1.A.5	Propagation de la variance dans un AR(1)	3
1.A.6	Variance stationnaire : définition explicite	4
1.A.7	Corrélation temporelle : définition explicite	4
1.A.8	Exemple chiffré complet	4
1.B	Simulation courte	5
1.C	Corrélation temporelle	5
2	Modèle MA = Moving Average	5
3	Autocorrélogramme partiel	6
3.A	Exemple 1	7
3.B	Exemple 2	7
3.B.1	Modèle autorégressif AR(1)	7
3.B.2	Modèle moyenne mobile MA(1)	8
3.B.3	Autocorrélogramme	8
3.B.4	Autocorrélogramme partiel	8
3.B.5	Processus stationnaire	9
4	Différents modèles : AR, MA, ARMA, SARMA, SARIMA, SARIMAX	9
5	Exemple d'analyse de données avec R	9
5.A	Méthode graphique	9
5.B	Méthode automatique avec fonction <code>auto.arima</code>	13
6	Fonction <code>stl()</code>	16

1 Introduction aux séries chronologiques

Méthodes SARIMAX / ARMA / ARIMA

- SARIMAX : Seasonal Autoregressive Integrated Moving Average with eXogenous regressors
- ARMA : Autoregressive Moving Average
- ARIMA : Autoregressive Integrated Moving Average

1.A Modèle auto-régressif d'ordre 1 :

Soit une variable Y mesurée au temps i

- Régression de Y sur ce qui se passe au temps $i - 1$: $Y(i) = (a) + bY(i - 1) + \epsilon$
 - Ce qui se passe au temps i = une fonction linéaire de ce qui se passe au temps $i - 1$
+ un bruit aléatoire
 - Processus auto-régressif d'ordre 1 (AR(1))

On observe une variable Y dans le temps : Y_1, Y_2, Y_3 .

Le modèle AR(1) écrit :

$$Y_i = a + bY_{i-1} + \varepsilon_i$$

où ε_i est un bruit aléatoire indépendant d'un instant à l'autre, c'est à dire qu'il n'y a pas de corrélation entre les ε_i entre eux

1.A.1 Qu'est-ce qu'une régression (dans ce contexte) ?

Une régression est un modèle qui cherche à expliquer une variable par une autre, en supposant une relation linéaire.

Dans un AR(1), la variable explicative n'est pas une autre variable, mais **la même variable observée juste avant**, c'est-à-dire Y_{i-1} .

C'est donc une régression temporelle, ou auto-régression :

- Y_{i-1} est l'information disponible au temps précédent,
- Y_i est ce qu'on cherche à prédire,
- ε_i représente la part de Y_i qui n'est pas expliquée par Y_{i-1} .

Autrement dit : le présent dépend linéairement du passé (= *c'est une fonction linéaire*), plus un terme aléatoire.

1.A.2 Pourquoi dit-on souvent que « $a = 0$ » ?

Le coefficient a est une constante, comme l'intercept dans une régression classique.

Cependant, dans la majorité des cours et logiciels :

1. La série est centrée avant modélisation.
 - Centrer consiste à soustraire la moyenne : $Y_i \leftarrow Y_i - \bar{Y}$.
 - Une série centrée a une moyenne égale à 0.
2. Si la série est centrée, alors le terme constant devient inutile.
 - En effet, **la constante sert justement à ajuster une moyenne non nulle**.
 - Donc, dans une série centrée : $Y_i = bY_{i-1} + \varepsilon_i$
 - et on peut prendre $a = 0$ par convention.

1.A.3 Variance : définition générale

La variance est une mesure de la dispersion d'une variable autour de sa moyenne.

- Une grande variance = la variable varie beaucoup.
- Une petite variance = la variable varie peu.

Mathématiquement :

$$\text{Var}(Y) = E[(Y - E[Y])^2]$$

Dans un AR(1), il y a deux sources de variance :

1. la variance du bruit ε_i
2. la propagation de ce bruit à travers le coefficient b

1.A.4 Variance du bruit : ce que c'est exactement

Le bruit ε_i est un terme aléatoire supposé :

- d'espérance 0 (= moyenne, c'est à dire que le bruit n'a pas de tendance moyenne à être positif ou négatif),
- indépendant d'un instant à l'autre,
- de variance constante σ^2 .

Cette variance σ^2 mesure l'amplitude des fluctuations imprévisibles.

C'est la "source" de variabilité du modèle.

Si σ^2 est grande : chaque instant reçoit un choc important.

Si σ^2 est petite : les variations imprévisibles sont faibles.

1.A.5 Propagation de la variance dans un AR(1)

C'est ici que l'AR(1) devient différent d'une régression classique.

Dans une régression ordinaire, l'erreur affecte uniquement la valeur courante.

Dans un AR(1), l'erreur affecte toute la suite, car :

$$Y_{i+1} = a + bY_i + \varepsilon_{i+1}$$

Donc si ε_i est grand :

- il fait monter ou descendre Y_i ,
- ce Y_i influence Y_{i+1} via le coefficient b ,
- et influence Y_{i+2} via b^2 , etc.

Ainsi, les erreurs se propagent dans le futur, et la variance totale est un équilibre entre :

- la variabilité injectée à chaque instant (σ^2),
- la quantité de mémoire du système (le coefficient b).

1.A.6 Variance stationnaire : définition explicite

Une variance stationnaire signifie :

1. La variance de Y_i reste la même pour tous les instants : elle ne diverge pas, ne monte pas, ne descend pas.
2. Le processus garde une variabilité “stable” dans le temps.

Autrement dit : la série ne devient ni explosive, ni de plus en plus variable, ni de moins en moins variable.

Cette situation n'est possible que si $|b| < 1$.

- Si $|b| < 1$, les erreurs se propagent mais finissent par s'atténuer \square variance stable.
- Si $|b| = 1$, les erreurs se cumulent sans limite \square variance infinie (marche aléatoire).
- Si $|b| > 1$, la série explose \square variance diverge.

Dans le cas stationnaire ($|b| < 1$) :

$$\text{Var}(Y_i) = \frac{\sigma^2}{1-b^2}$$

Cette formule signifie :

- la variance totale est plus grande que la variance du bruit,
- plus b est grand, plus la variance de Y_i augmente,
- car plus b est grand, plus les erreurs anciennes continuent à influencer les valeurs futures.

1.A.7 Corrélation temporelle : définition explicite

La corrélation mesure l'intensité du lien linéaire entre deux valeurs.

Dans un AR(1) stationnaire:

$$\text{Corr}(Y_i, Y_{i-k}) = b^k$$

Donc :

- $\text{Corr}(Y_i, Y_{i-1}) = b$
- $\text{Corr}(Y_i, Y_{i-2}) = b^2$
- $\text{Corr}(Y_i, Y_{i-3}) = b^3$
- etc.

C'est une corrélation qui décroît exponentiellement avec la distance temporelle.

1.A.8 Exemple chiffré complet

Prenons un AR(1) :

$$Y_i = 0.8Y_{i-1} + \varepsilon_i, \quad \text{Var}(\varepsilon_i) = 1$$

Variance stationnaire

$$\text{Var}(Y_i) = \frac{1}{1-0.8^2} = \frac{1}{1-0.64} = \frac{1}{0.36} \approx 2.78$$

Donc :

- la variance du processus est presque 3 fois la variance du bruit,
- car $b=0.8$ propage beaucoup les erreurs passées.

1.B Simulation courte

Supposons $Y_0 = 10$.

- $\varepsilon_1 = 0.5$

$$Y_1 = 0.8 \times 10 + 0.5 = 8.5$$

- $\varepsilon_2 = -0.2$

$$Y_2 = 0.8 \times 8.5 - 0.2 = 6.6$$

- $\varepsilon_3 = 0.1$

$$Y_3 = 0.8 \times 6.6 + 0.1 = 5.38$$

La série décroît progressivement vers 0 (car $|b| < 1$) et reste influencée par chaque erreur qui se propage.

1.C Corrélation temporelle

$$\text{Corr}(Y_3, Y_2) = 0.8$$

$$\text{Corr}(Y_3, Y_1) = 0.8^2 = 0.64$$

$$\text{Corr}(Y_3, Y_0) = 0.8^3 = 0.512$$

Résumé

Un AR(1) est un modèle où :

- le présent est expliqué par le passé (régression),
 - le coefficient b crée une corrélation décroissante avec le passé,
 - la variance totale résulte de l'équilibre entre :
 - la variance du bruit,
 - la propagation de cette variance via b ,
 - la variance stationnaire est une variance stable, finie, identique à travers le temps,
 - la constante a disparait si on travaille sur une série centrée.
-

2 Modèle MA = Moving Average

Un modèle MA(q) exprime la valeur actuelle comme une **MOYENNE MOBILE** des **ERREURS PASSÉES**.

Les **BRUITS** (ou erreurs) sont liés les uns avec les autres.

Modèle MA(1) :

$$Y_j = \varepsilon_i + \varepsilon_{i-1}$$

$$Y_{j-1} = \varepsilon_{i-1} + \varepsilon_{i-2}$$

La corrélation entre Y_j et Y_{j-1} est donc due à la présence commune de ε_{i-1} dans les deux équations. Donc corrélation $\neq 0$ car il y a un terme commun.

Alors que si on fait la corrélation entre Y_j et Y_{j-2} :

$$Y_j = \varepsilon_i + \varepsilon_{i-1}$$

$$Y_{j-2} = \varepsilon_{i-2} + \varepsilon_{i-3}$$

Il n'y a pas de terme commun entre les deux équations, donc corrélation = 0.

Donc avec :

- Un modèle AR(1) : la corrélation diminue exponentiellement avec le temps = autocorrélogramme.
- Un modèle MA(1) : la corrélation est nulle au-delà du lag 1.



Donc à partir de la représentation de la corrélation en fonction du lag (autocorrélogramme), on peut deviner si le modèle est AR ou MA.

3 Autocorrélogramme partiel

Un autocorrélogramme partiel permet de voir la corrélation entre

- Y_i et Y_{i-p}
- ajusté sur $Y_{i-1}, Y_{i-2}, \dots, Y_{i-(p-1)}$.

Corrélation spécifique ajustée entre le passé ± lointain et le présent, en ajustant sur tout ce qui se

passe entre les deux.

Si on fait avec $p = 2$:

- On regarde la corrélation entre Y_i et Y_{i-2}

On a Y_1, Y_2, Y_3 (car $p=2$ donc on a besoin de 3 observations pour avoir un lag de 2 entre Y_3 et Y_1)

- Processus autorégressif d'ordre 1: Y_1 agit sur Y_2 (lag 1) et Y_2 agit sur Y_3 (lag 1)
- Corrélation entre Y_1 et Y_3 vaut la corrélation entre chaque Y exponentiée par le lag : b^2
- Mais si on ajuste sur Y_2 , on enlève l'effet de Y_2 sur Y_3 , donc Y_1 n'a plus d'effet sur Y_3 !
- La corrélation entre Y_1 et Y_3 ajustée sur Y_2 vaut donc 0.

Donc l'autocorrélogramme partiel va donner 0 au-delà du lag 1 pour un modèle MA(1).

3.A Exemple 1

$$Y_i = bY_{i-1} + cY_{i-2} + \varepsilon_i + \varepsilon_{i-1}$$

Modèle ARMA(2,1)

Dans un modèle ARMA(p, q),

- p = ordre autorégressif (AR) \square nombre de termes Y retardés,
- q = ordre moyenne mobile (MA) \square nombre de termes ε retardés.

1. Partie AR (autorégressive)

- Y_{i-1}
- Y_{i-2}

Donc 2 termes AR $\square p = 2$.

2. Partie MA (moyenne mobile)

- ε_i \square toujours présent, ce n'est pas compté dans q
- ε_{i-1}

Le terme ε_i est l'erreur courante : on ne le compte pas dans l'ordre.

On compte seulement les retards.

Donc il y a 1 terme MA retardé $\square q = 1$.

3.B Exemple 2

Mesure de la TA tous les matins

Y_i = TA du jour i : il y en a donc 365

3.B.1 Modèle autorégressif AR(1)

Les Y_i ont une variance = 1 c'est à dire que la TA varie peu d'un jour à l'autre.

$$\rightarrow Y_i = bY_{i-1} + \varepsilon_i$$

- On peut pas faire une régression linéaire habituelle car Y_i est à la fois à gauche et à droite de l'équation
- si variance = 1 / variance stationnaire :
 - b = corrélation entre Y_i et Y_{i-1} quelque soit i
 - avec le temps : corrélation entre Y_i et $Y_{i-2} = b^2$
 - corrélation entre Y_i et $Y_{i-k} = b^k$
 - avec le temps, la corrélation tend vers 0 de façon exponentielle

3.B.2 Modèle moyenne mobile MA(1)

$$Y_i = \varepsilon_i + \varepsilon_{i-1}$$

- 2 bruits : régression linéaire impossible
- Le temps i et le temps $i - 1$ n'ont rien à voir l'un avec l'autre et sont liés au bruit
 - Appareil de mesure a aussi un bruit et a un défaut ponctuel à un moment donné
 - * stress du jour qui influe sur le bruit du jour aussi

Globalement : dans la mesure du jour, le bruit observé vient d'un bruit issu de la mesure du jour + du bruit d'hier qui vient d'un contexte, influençant l'erreur de mesure qu'il y a dans Y_i aujourd'hui.

- Corrélation entre Y_i et $Y_{i-1} = \varrho$ et est due au fait qu'il y a un terme commun dans les 2 équations : ε_{i-1}
- Corrélation entre Y_i et $Y_{i-2} = 0$ car pas de terme commun dans l'équation

3.B.3 Autocorrélogramme

Permet assez facilement de visualiser si le modèle est AR ou MA

3.B.4 Autocorrélogramme partiel

Corrélation entre Y_i et Y_{i-k} ajustée sur $Y_{i-1}, Y_{i-2}, \dots, Y_{i-(k-1)}$ (ajusté sur toutes les mesures qu'on a entre les 2)

Si modèle AR(1) :

- Y_1 agit sur Y_2 et Y_2 agit sur Y_3
- Mais Y_1 agit sur Y_3 uniquement via Y_2
- Donc si on ajuste sur Y_2 , la corrélation entre Y_1 et Y_3 ajustée sur $Y_2 = 0$

Donc modèle AR(1) dans autocorrélogramme partiel ressemble au modèle MA(1) dans autocorrélogramme classique.

3.B.5 Processus stationnaire

= Processus dont les propriétés statistiques (moyenne, variance, autocorrélation) ne changent pas au cours du temps.

4 Différents modèles : AR, MA, ARMA, SARMA, SARIMA, SARIMAX

On part d'un processus stationnaire.

1. **Description** : évolution au cours du temps => MOYENNE STABLE ? VARIANCE STABLE ?
 2. **Modélisation** : autocorrélogramme et autoccorrélogramme partiel => modèle AR, MA ou ARMA ?
 - voire SARMA : saisonnier (avec saisonnalité)
 - voire SARIMA : avec dérive au cours du temps (écart à la stationnarité : I = Integrated c'est à dire qu'on intègre une tendance linéaire au modèle)
 - voire SARIMAX : avec des variables explicatives (exogènes)
 - exogènes = variables explicatives externes au modèle de la série chronologique
 - $Y_i = bY_{i-1} + cY_{i-2} + \varepsilon_i + \varepsilon_{i-1} + d\vartheta$
-

5 Exemple d'analyse de données avec R

5.A Méthode graphique

Le fichier gs contient des données sur les tendances de recherche Google (sur recherche = suicide entre Juin 2017 et Juin 2022).

Voir si indice X = survenue du confinement modifie la recherche du mot “suicide”

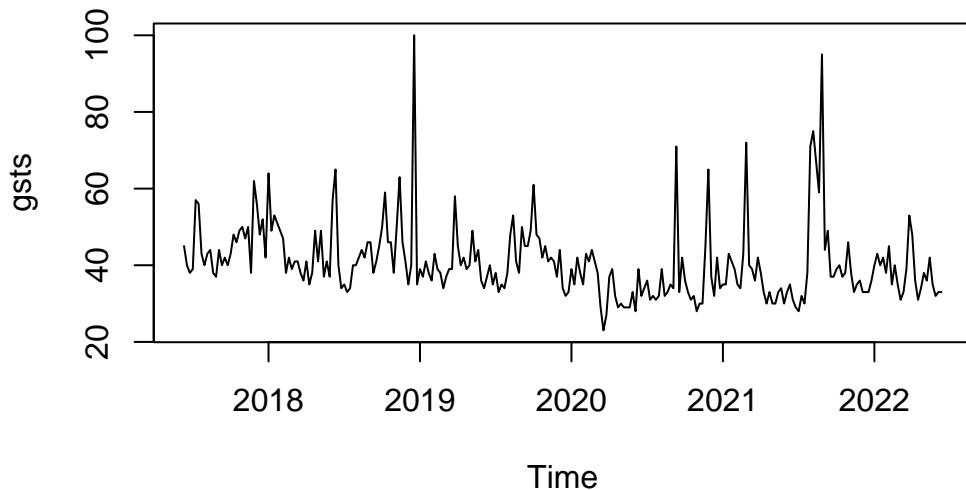
Modèle = ARMAX (avec X = facteur exogène = confinement)

- avec X
1. Transformation de la variable `gs$suicide` en une série chronologique grâce à la fonction `ts` = timeserie
 - si pas de fonction `ts`: on peut pas faire d'analyses sur séries chronologiques avec R.
 - syntaxe de la fonction `ts` :
 - `ts(data, frequency, start)`
 - `data` = données,
 - `frequency` = fréquence (ex: 52 semaines dans une année),
 - * permet de dire que la semaine i et la semaine $i + 52$ sont comparables (même semaine de l'année)
 - `start` = date de début (ex: `c(2017,24)`)

2. Affichage de la série chronologique avec la fonction `plot`

```
#construction du dataset pour série chronologique
gsts <- ts(data=gs$suicide,frequency=52,start=c(2017,24))
#affichage du graphique qui montre l'évolution de la série chronologique
plot(
  gsts,
  main = "Évolution temporelle du nombre standardisé de requêtes \nsur le moteur de recherche Google autour du mot « suicide »")
```

Évolution temporelle du nombre standardisé de requêtes sur le moteur de recherche Google autour du mot « suicide »

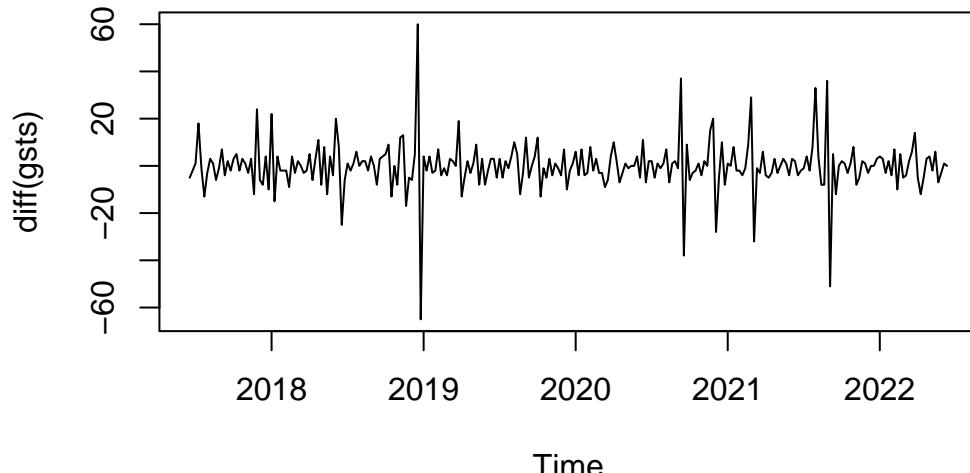


- On considère “à la louche” que la variance est stationnaire au cours du temps.
- Il y a peut-être une tendance à la baisse (le processus ne serait pas strictement stationnaire)
- 2 pics d’incidence sont observés, respectivement vers la fin des années 2018 et 2021 : cela peut faire craindre des problèmes d’écart à l’hypothèse de normalité dans les analyses à venir.

Pour obtenir un processus stationnaire : une solution classique est de modéliser la différence :
 $\Delta Y_t = Y_t - Y_{t-1}$

```
plot.ts(diff(gsts), main = "Évolution temporelle de la différence Delta Y_t = Y_t - Y_{t-1}.\n\n")
```

Évolution temporelle de la différence $\Delta Y_t = Y_t - Y_{t-1}$. Le processus est rendu stationnaire, mais les écarts à la normalité



- Le processus est beaucoup plus stationnaire
- Il persiste un pic en 2019, qui fait que la variance est pas complètement constante, et la normalité des résidues ε_i ne sera pas parfaite.

3. Autocorrélogramme

Fonction `acf` (auto correlation function) permet de tracer l'autocorrélogramme

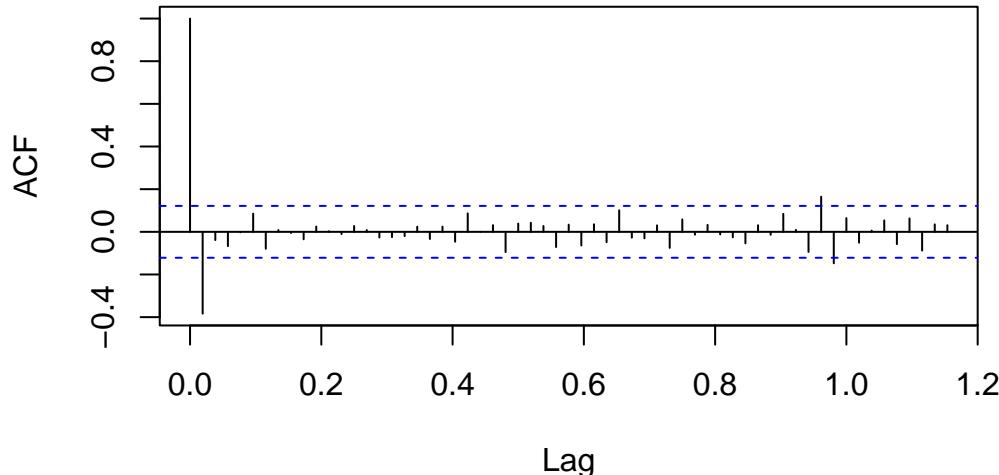
Faite sur la différence de la série chronologique pour rendre le processus stationnaire

Syntaxe :

- `acf(x, lag.max, main)`
 - `x` = série chronologique
 - `lag.max` = nombre maximum de lags à afficher
 - `main` = titre du graphique

```
acf(diff(gsts), lag.max= 60, main="Autocorrélogramme de la différence de la série chronologique")
```

Autocorrélogramme de la différence de la série chronologique



- 1ère mesure est fortement corrélée avec la mesure précédente (lag 1) (corrélation entre Y_i et Y_{i-1} donc on s'en fiche un peu)
- 2e mesure : corrélation négative Y_i et Y_{i-1} largement au dessus du seuil de significativité (pointillés bleus)
- 3e mesure : aucune corrélation

Il y a un décrochage assez net pour $k > 1$ (lag > 1), ce qui est en faveur d'un modèle MA(1).

Donc au moins une compensante MA(1) !

4. Autocorrélogramme partiel

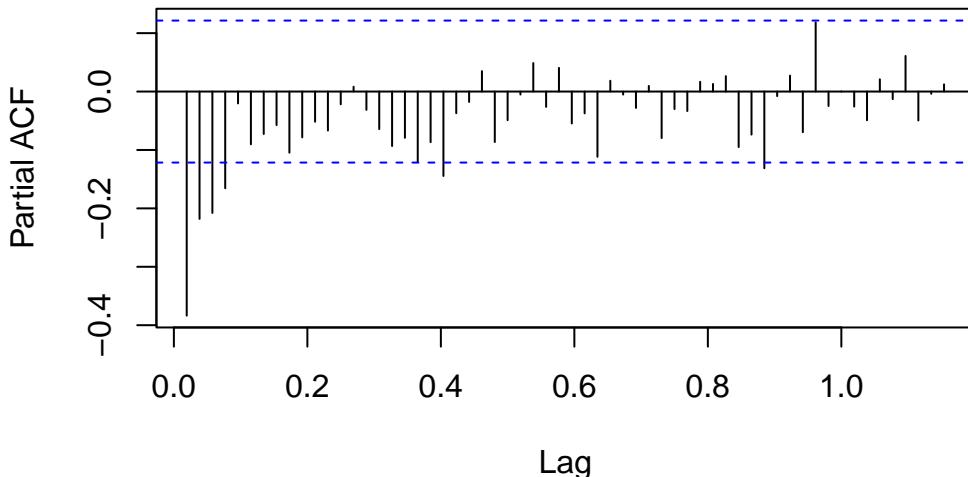
Fonction `pacf` (partial auto correlation function) permet de tracer l'autocorrélogramme partiel

Syntaxe :

- `pacf(x, lag.max, main)`
 - `x` = série chronologique
 - `lag.max` = nombre maximum de lags à afficher
 - `main` = titre du graphique

```
pacf(diff(gsts), lag.max= 60, main="Autocorrélogramme partiel de la différence de la série chronologique")
```

Partocorrélogramme partiel de la différence de la série chronologique



- 1ère corrélation partielle est élevée (lag 1)
- 2e corrélation partielle est statistiquement significative (lag 2)
- Mais chute énorme de corrélation entre 1ère, 2e...

Donc probablement une composante AR(1)

Au total : probablement un modèle ARMA(1,1)

5.B Méthode automatique avec fonction `auto.arima`

1. Sans variable exogène

Fonction `auto.arima` du package `forecast` permet de faire une sélection automatique du modèle ARIMA le plus adapté aux données.

Pour Falissard : c'est un piège car automatique et n'oblige pas à réfléchir !

Syntaxe :

- `auto.arima(y, xreg, seasonal, trace)`
 - `y` = série chronologique
 - `xreg` = matrice de variables explicatives exogènes (optionnel)
 - `seasonal` = si `TRUE`, recherche de composantes saisonnières (optionnel)
 - `trace` = si `TRUE`, affiche les modèles testés (optionnel)

```
auto.arima(diff(gsts), trace=TRUE)
```

Fitting models using approximations to speed things up...

```
ARIMA(2,0,2)(1,0,1) [52] with non-zero mean : Inf
ARIMA(0,0,0)           with non-zero mean : 1969.555
ARIMA(1,0,0)(1,0,0) [52] with non-zero mean : 1953.522
```

```

ARIMA(0,0,1)(0,0,1)[52] with non-zero mean : 1901.823
ARIMA(0,0,0)           with zero mean      : 1967.529
ARIMA(0,0,1)           with non-zero mean : 1899.763
ARIMA(0,0,1)(1,0,0)[52] with non-zero mean : 1937.627
ARIMA(0,0,1)(1,0,1)[52] with non-zero mean : 1939.683
ARIMA(1,0,1)           with non-zero mean : 1889.213
ARIMA(1,0,1)(1,0,0)[52] with non-zero mean : 1915.516
ARIMA(1,0,1)(0,0,1)[52] with non-zero mean : 1890.644
ARIMA(1,0,1)(1,0,1)[52] with non-zero mean : 1917.15
ARIMA(1,0,0)           with non-zero mean : 1930.944
ARIMA(2,0,1)           with non-zero mean : 1890.456
ARIMA(1,0,2)           with non-zero mean : 1890.673
ARIMA(0,0,2)           with non-zero mean : 1892.568
ARIMA(2,0,0)           with non-zero mean : 1921.135
ARIMA(2,0,2)           with non-zero mean : 1892.543
ARIMA(1,0,1)           with zero mean     : 1887.24
ARIMA(1,0,1)(1,0,0)[52] with zero mean    : 1913.441
ARIMA(1,0,1)(0,0,1)[52] with zero mean    : 1888.642
ARIMA(1,0,1)(1,0,1)[52] with zero mean    : 1915.07
ARIMA(0,0,1)           with zero mean     : 1897.757
ARIMA(1,0,0)           with zero mean     : 1928.903
ARIMA(2,0,1)           with zero mean     : 1888.431
ARIMA(1,0,2)           with zero mean     : 1888.639
ARIMA(0,0,2)           with zero mean     : 1891.134
ARIMA(2,0,0)           with zero mean     : 1919.077
ARIMA(2,0,2)           with zero mean     : 1890.504

```

Now re-fitting the best model(s) without approximations...

```
ARIMA(1,0,1)           with zero mean     : 1884.453
```

```
Best model: ARIMA(1,0,1)           with zero mean
```

```
Series: diff(gsts)
ARIMA(1,0,1) with zero mean
```

Coefficients:

	ar1	ma1
0.3803	-0.9698	
s.e.	0.0622	0.0196

```
sigma^2 = 80.35: log likelihood = -939.18
AIC=1884.36   AICc=1884.45   BIC=1895.04
```

Ici : Modèle ARIMA(1,0,1) est sélectionné donc équivalent à ARMA(1,1)

- Avec l pour Integrated = 0 (= dérive au cours du temps) (car on a différencié la série chronologique initiale)
- (1,0,1) car :
 - p = 1 (AR(1))
 - d = 0 (Integrated)
 - * d=1 : ça veut dire qu'il y a une tendance linéaire à la baisse, ce qui aurait été le cas si on avait pris 'gsts'
 - * d=2 : tendance quadratique, c'est à dire courbe en U ou en cloche
 - q = 1 (MA(1))
- sans composante saisonnière (car pas de saisonnalité dans les données) (s'il y en avait une, ça aurait été noté ARIMA(p,d,q)(P,D,Q)[s] avec s = saisonnalité)

2. Avec variable exogène

On regarde s'il y a une rupture le 17/03/20 = confinement

1. Création de la variable exogène

Variable x construite :

- Elle vaut -1 pour les semaines avant la semaine 145 (date du 17 Mars 2020)
- Et vaut 0 jusqu'en Juin 2022.
- Utiliser la fonction rep pour répéter les valeurs (*c'est à dire que rep(-1,145) crée un vecteur de 145 valeurs égales à -1 et rep(0,115) crée un vecteur de 115 valeurs égales à 0*)

```
x <- c(rep(-1,145), rep(0,115))
```

Puis même modèle res mais en y intégrant la variable exogène x via l'argument xreg

Fonction arima permet de faire une modélisation ARIMA avec des variables exogènes

Syntaxe :

- arima(x, order, xreg)
 - x = série chronologique
 - order = ordre du modèle ARIMA (p,d,q) (ici, on sait que c'est (1,0,1) d'après l'analyse précédente)
 - xreg = matrice de variables explicatives exogènes

```
res <- arima(diff(gsts), order = c(1,0,1), xreg = x)
res
```

Call:

```
arima(x = diff(gsts), order = c(1, 0, 1), xreg = x)
```

Coefficients:

```

ar1      ma1   intercept      x
0.3793 -0.9999     0.0185  0.0855
s.e.  0.0576  0.0110     0.0299  0.0472

sigma^2 estimated as 77.35: log likelihood = -936.58, aic = 1883.17

```

Estimation de l'effet de `x` avec son écart-type.

Si on fait le rapport de `x` sur son écart-type, on obtient une statistique de test de Wald

On le compare à une loi normale centrée réduite pour obtenir une p-value : c'est à dire qu'on le compare à 1,96 (seuil de significativité à 5%).

Le paramètre de `x`= 0.0855

L'écart-type de `x` = 0.0472

Rapport : $0.0855 / 0.0472 = 1.81 < 1.96$ donc p-value > 0.05

Donc pas d'effet significatif du confinement sur les recherches Google autour du mot “suicide”.

6 Fonction `stl()`

Fonction graphique purement exploratoire qui permet de décomposer une série chronologique en 3 composantes :

- données originales
- saisonnalité
- tendance
- résidus

Syntaxe :

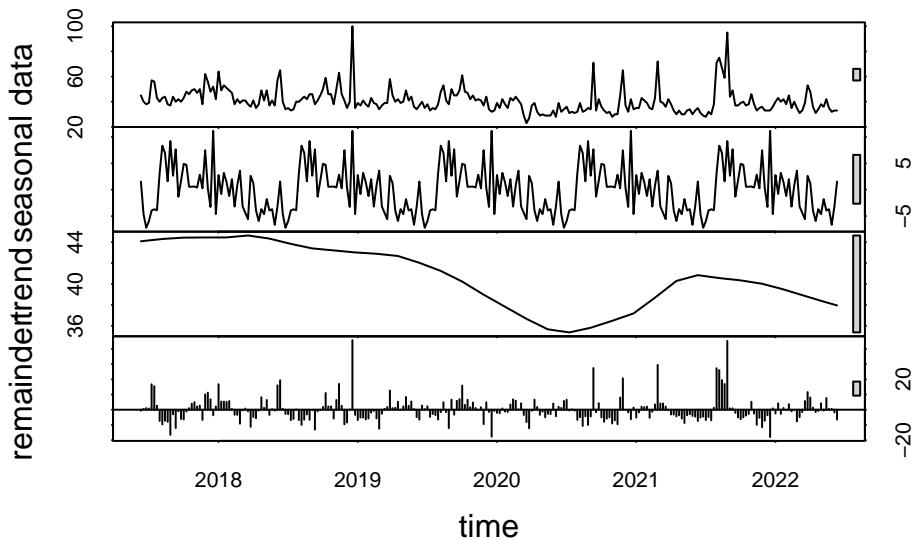
- `stl(x, s.window, t.window)`
 - `x` = série chronologique
 - `s.window` = fenêtre de lissage pour la saisonnalité (ex: “`periodic`”)
 - `t.window` = fenêtre de lissage pour la tendance (ex: 13)

```

plot(
  stl(
    gsts,
    s.window="periodic"),
    main="Analyse graphique exploratoire d'une série chronologique à l'aide de la fonction
  )

```

graphique exploratoire d'une série chronologique à l'aide de la fonction 'stl()' proposée en trois termes : une composante périodique, une composante de tendance et



- On observe une saisonnalité annuelle avec un pic fin Août (avant la rentrée scolaire)