

Problème pratique de statistique n° pps089  
Nombre de naissances en France

J.R. Lobry

19 février 2017

Les données sont le nombre de naissances par mois en France métropolitaine de 1975 à 2007, plus quelques premières estimations pour 2008.

## 1 Origine des données

Les données sont tirées du site de l'INSEE, la description est conservée en commentaire en début du fichier :

```
cat(readLines("http://pbil.univ-lyon1.fr/R/donnees/tab000436391.csv", n = 9), sep = "\n")
# Nombre de naissances en France métropolitaine -- Identifiant : 000436391
#
# Zone géographique;France Métropolitaine
# Unité de mesure;Individu
# Durée ou moment d'observation;Observation sur un mois
# Périodicité des valeurs;Mensuelle
# Origine des données;Institut national de la statistique et des études
# économiques (Insee) - Département de la démographie
#
```

## 2 Importation des données

Importer les données dans  :

```
read.table("http://pbil.univ-lyon1.fr/R/donnees/tab000436391.csv", sep = ";", header = TRUE) -> tab
head(tab)
```

	Numéro	Période	Année	Valeur	Remarque
1	4	Avril	2008	66200	Estimé
2	3	Mars	2008	65100	Estimé
3	2	Février	2008	62000	Estimé
4	1	Janvier	2008	67100	Estimé
5	12	Décembre	2007	66724	
6	11	Novembre	2007	64811	

```
summary(tab)
```

	Numéro	Période	Année	Valeur	Remarque
Min.	: 1.00	Avril : 34	Min. :1975	Min. :52901	:396
1st Qu.:	3.00	Février : 34	1st Qu.:1983	1st Qu.:60884	Estimé : 4
Median :	6.00	Janvier : 34	Median :1991	Median :63340	
Mean :	6.46	Mars : 34	Mean :1991	Mean :63208	
3rd Qu.:	9.00	Août : 33	3rd Qu.:1999	3rd Qu.:65700	
Max. :	12.00	Décembre: 33	Max. :2008	Max. :73984	
		(Other) :198			

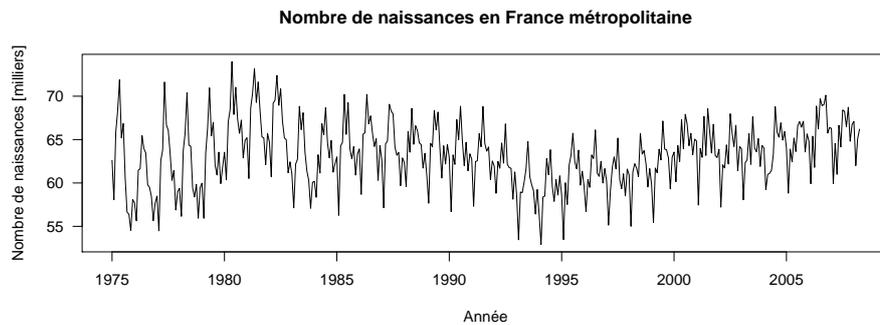
### 3 Série temporelle

Convertir les données en un objet de la classe `ts` qui est adapté à la représentation des séries temporelles. On décide d'exprimer le nombre de naissances par milliers :

```
myts <- ts(rev(tab$Valeur/1000), start = 1975, frequency = 12)
summary(myts)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  52.90  60.88   63.34   63.21  65.70   73.98
tsp(myts)
[1] 1975.00 2008.25  12.00
```

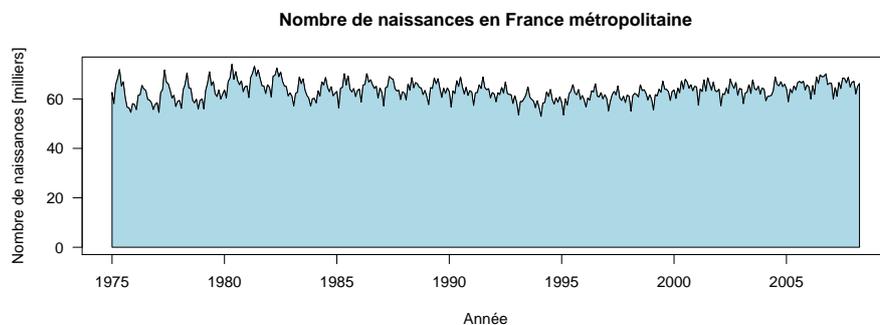
### 4 Représentations graphiques

```
plot(myts, xlab = "Année", las = 1,
     ylab = "Nombre de naissances [milliers]",
     main = "Nombre de naissances en France métropolitaine")
```



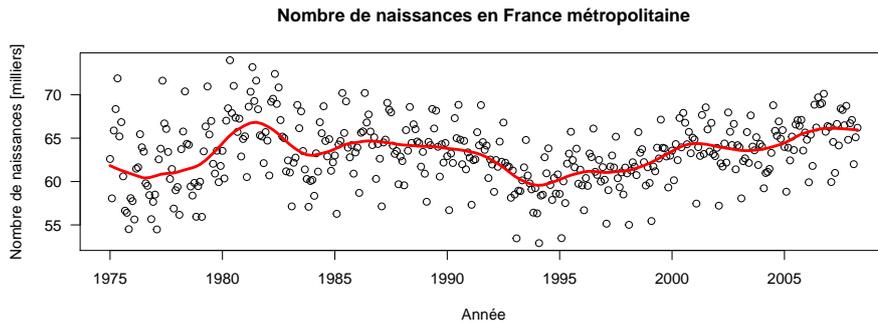
On note la présence de fluctuations saisonnières. Relativisons avec une échelle partant de zéro :

```
plot(myts, xlab = "Année", las = 1,
     ylab = "Nombre de naissances [milliers]",
     main = "Nombre de naissances en France métropolitaine",
     ylim = c(0, max(myts)))
polycurve <- function(x, y, base.y = min(y), ...) {
  polygon(x = c(min(x), x, max(x)), y = c(base.y, y, base.y),
  ...)
}
polycurve(time(myts), myts, col = "lightblue", base.y = 0)
```



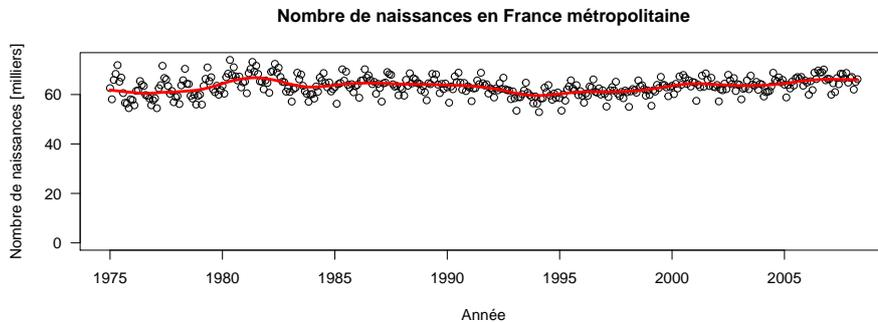
Faisons un petit lissage pour gommer les fluctuations saisonnières :

```
plot(myts, xlab = "Année", las = 1,
     ylab = "Nombre de naissances [milliers]",
     main = "Nombre de naissances en France métropolitaine", type = "p")
lissage <- lowess(myts, f=1/10)
lines(lissage, lwd = 3, col = "red")
```



Relativisons les tendances à long terme :

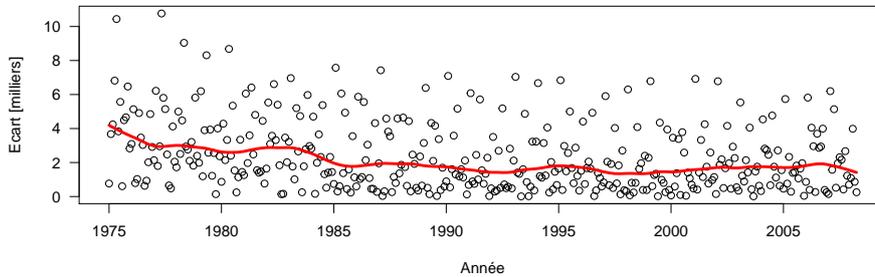
```
plot(myts, xlab = "Année", las = 1,
     ylab = "Nombre de naissances [milliers]",
     main = "Nombre de naissances en France métropolitaine", type = "p",
     ylim = c(0,max(myts)))
lissage <- lowess(myts, f=1/10)
lines(lissage, lwd = 3, col = "red")
```



La statistique est d'une redoutable stabilité sur plus de 30 ans d'observations, c'est assez remarquable. On note une diminution de l'amplitude des fluctuations saisonnières, essayons de la visualiser.

```
residus <- abs(myts - lissage$y)
plot(time(myts), residus,
     xlab = "Année", las = 1,
     ylab = "Ecart [milliers]",
     main = "Diminution de l'amplitude des fluctuations saisonnières")
lis2 <- lowess(residus, f=1/10)
lines(lis2, lwd = 3, col = "red")
```

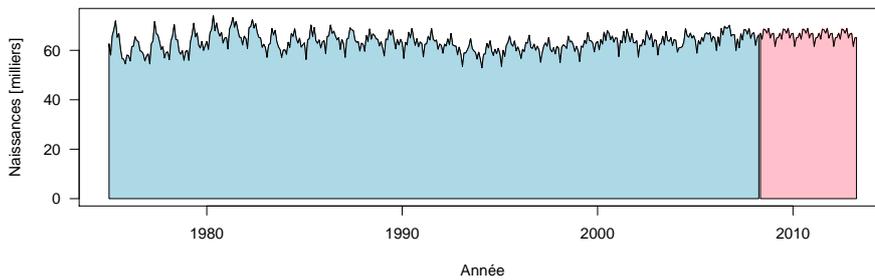
**Diminution de l'amplitude des fluctuations saisonnières**



On peut s'amuser à extrapoler avec un modèle ARIMA <sup>1</sup> :

```
fit <- arima(myts, seasonal = list(order = c(0, 1, 1), period = 12))
pred <- predict(fit, n.ahead = 12*5)
ts.plot(myts, pred$pred, col = c("black","red"),
main = "Extrapolation de la tendance avec un modèle ARIMA",
gpar = list(las = 1, xlab = "Année",
ylab = "Naissances [milliers]", ylim = c(0,max(myts)))
polycurve(time(myts), myts, base.y = 0, col = "lightblue")
polycurve(time(pred$pred), pred$pred, base.y = 0, col = "pink")
```

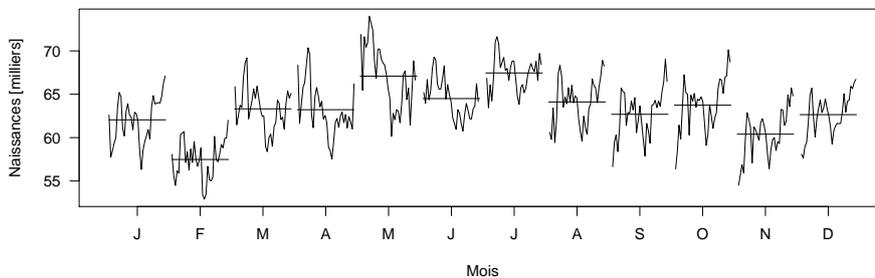
**Extrapolation de la tendance avec un modèle ARIMA**



La fonction `monthplot()` permet de visualiser simplement les fluctuations mensuelles :

```
monthplot(myts, xlab = "Mois", ylab = "Naissances [milliers]",
main = "Fluctuations mensuelles des naissances", las = 1)
```

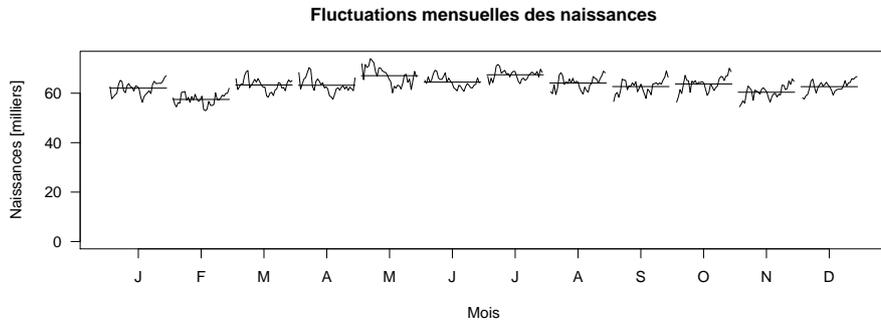
**Fluctuations mensuelles des naissances**



Si on relativise en partant de zéro :

1. Autoregressive Integrated Moving Average

```
monthplot(myts, xlab = "Mois", ylab = "Naissances [milliers]",  
main = "Fluctuations mensuelles des naissances", las = 1,  
ylim = c(0,max(myts)))
```



Les fluctuations mensuelles du nombre de naissances en France métropolitaine sont donc peu marquées depuis 1975, même si l'on note une prédilection pour les naissances au printemps, assez commune chez les mammifères.