Antoine FONTANEZ

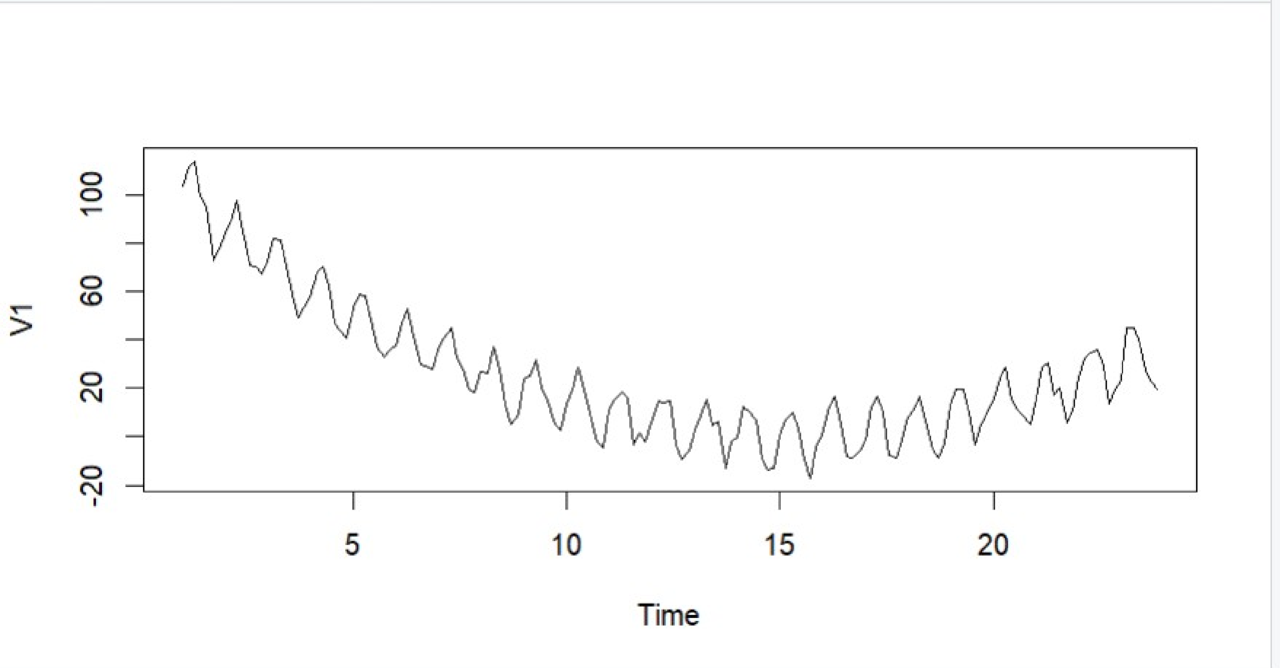
TP Statistiques

1) On commence par charger la série temporelle. Elle est de fréquence 7, car la période est en semaine (donc 7 jours) :

```
> data.ts=ts(data, frequency=7)
```

On dessine ensuite le graphique :

```
> plot(data.ts)
```



2) On isole la composante saisonnière et on en affiche ses valeurs :

```
> model=decompose(data.ts, type="additive")
> sais=model$seasonal
> sais
```

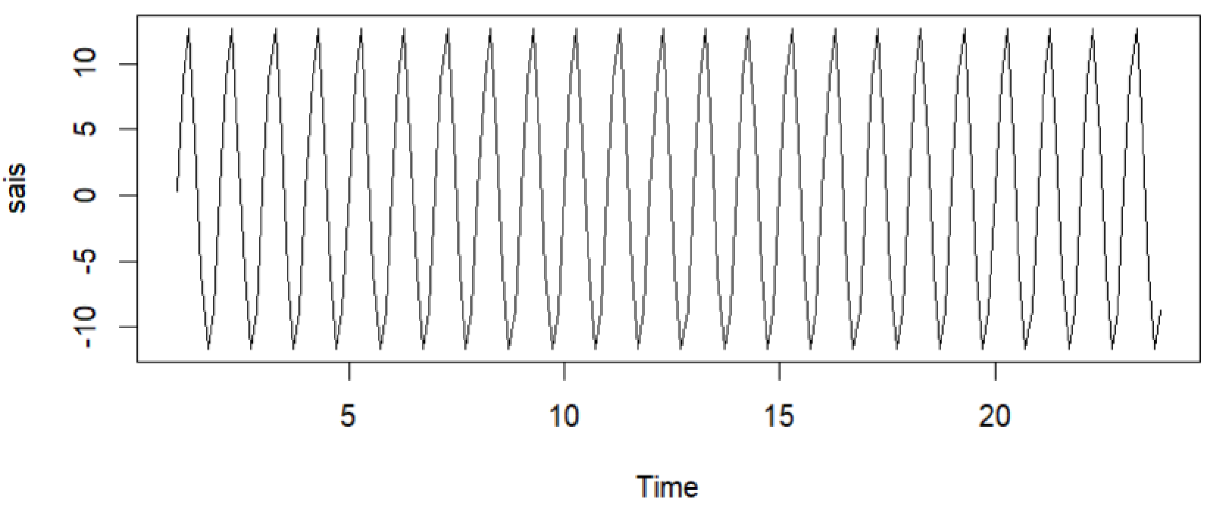
Et on obtient :

```
Time Series:
Start = c(1, 1)
End = c(23, 7)
Frequency = 7
  [1]   0.372154   8.745153  12.687328   4.360603  -5.794943 -11.670811
  [7]  -8.699485   0.372154   8.745153  12.687328   4.360603  -5.794943
 [13] -11.670811  -8.699485   0.372154   8.745153  12.687328   4.360603
 [19]  -5.794943 -11.670811  -8.699485   0.372154   8.745153  12.687328
 [25]   4.360603  -5.794943 -11.670811  -8.699485   0.372154   8.745153
 [31]  12.687328   4.360603  -5.794943 -11.670811  -8.699485   0.372154
```

```
 [37]   8.745153  12.687328   4.360603  -5.794943 -11.670811  -8.699485
 [43]   0.372154   8.745153  12.687328   4.360603  -5.794943 -11.670811
 [49]  -8.699485   0.372154   8.745153  12.687328   4.360603  -5.794943
 [55] -11.670811  -8.699485   0.372154   8.745153  12.687328   4.360603
 [61]  -5.794943 -11.670811  -8.699485   0.372154   8.745153  12.687328
 [67]   4.360603  -5.794943 -11.670811  -8.699485   0.372154   8.745153
 [73]  12.687328   4.360603  -5.794943 -11.670811  -8.699485   0.372154
 [79]   8.745153  12.687328   4.360603  -5.794943 -11.670811  -8.699485
 [85]   0.372154   8.745153  12.687328   4.360603  -5.794943 -11.670811
 [91]  -8.699485   0.372154   8.745153  12.687328   4.360603  -5.794943
 [97] -11.670811  -8.699485   0.372154   8.745153  12.687328   4.360603
[103]  -5.794943 -11.670811  -8.699485   0.372154   8.745153  12.687328
[109]   4.360603  -5.794943 -11.670811  -8.699485   0.372154   8.745153
[115]  12.687328   4.360603  -5.794943 -11.670811  -8.699485   0.372154
[121]   8.745153  12.687328   4.360603  -5.794943 -11.670811  -8.699485
[127]   0.372154   8.745153  12.687328   4.360603  -5.794943 -11.670811
[133]  -8.699485   0.372154   8.745153  12.687328   4.360603  -5.794943
[139] -11.670811  -8.699485   0.372154   8.745153  12.687328   4.360603
[145]  -5.794943 -11.670811  -8.699485   0.372154   8.745153  12.687328
[151]   4.360603  -5.794943 -11.670811  -8.699485   0.372154   8.745153
[157]  12.687328   4.360603  -5.794943 -11.670811  -8.699485
```

Et on peut également en afficher un graphique :

```
> plot(sais)
```



3) On modélise la tendance par un polynôme et on estime la valeur de ses coefficients :

```
> t=1:161
> t2=t^2
> t3=t^3
> desais=data.ts-sais
> data.lm=lm(desais~t+t2+t3)
> summary(data.lm)
```

Et on obtient :

```
Call:
lm(formula = desais ~ t + t2 + t3)

Residuals:
    Min      1Q  Median      3Q     Max 
-7.4594 -2.5021 -0.2893  2.3854 10.4289 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  1.027e+02  1.141e+00  89.959   <2e-16 ***
t           -2.093e+00  6.082e-02 -34.416   <2e-16 ***
t2           1.117e-02  8.710e-04  12.826   <2e-16 ***
t3          -4.587e-06  3.535e-06  -1.298    0.196    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.536 on 157 degrees of freedom
Multiple R-squared:  0.9831,	Adjusted R-squared:  0.9828 
F-statistic:  3042 on 3 and 157 DF,  p-value: < 2.2e-16
```

C’est à dire :

a = 1.027e+02
b = -2.093e+00
c = 1.117e-02
d = -4.587e-06

On fait un test sur les coefs pour voir si ils sont tous significativement non-nuls en prenant un seuil d’erreur alpha de 0.05. Ils sont tous significativement non-nuls à l’exception de d (t3) dont la Pr(>|t|) dépasse 0.05, ce qui nous fait valider l’hypothèse qu’il n’est pas significativement non-nul. On peut l’éliminer.

On peut simplifier f en

$f_t = a + bt + ct^2$

Donc on redéfini les coefs (sans t3) :

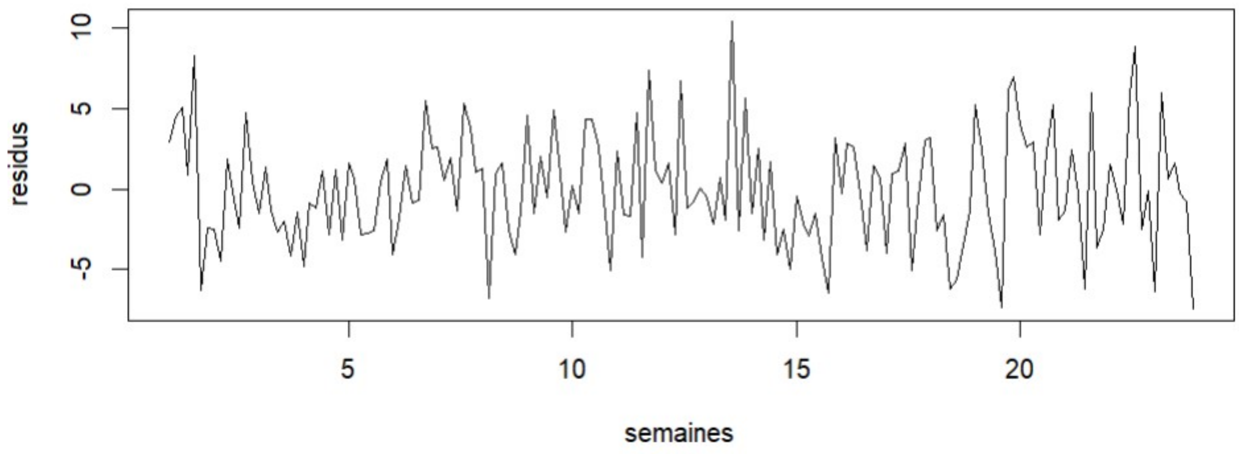
```
data.lm=lm(desais~t+t2)

> summary(data.lm)

Call:
lm(formula = desais ~ t + t2)

Residuals:
    Min      1Q  Median      3Q     Max 
-8.3812 -2.4467 -0.2373  2.2073 10.5692 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  1.017e+02  8.485e-01  119.84   <2e-16 ***
t           -2.021e+00  2.418e-02  -83.57   <2e-16 ***
t2           1.006e-02  1.446e-04   69.56   <2e-16 ***
---
Signif. codes:  
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
Residual standard error: 3.544 on 158 degrees of freedom
Multiple R-squared:  0.9829,	Adjusted R-squared:  0.9827
F-statistic:  4542 on 2 and 158 DF,  p-value: < 2.2e-16
```

Donc :
a = 1.017e+02
b = -2.021
c = 1.006e-02

4) On trace le graphique des résidus :

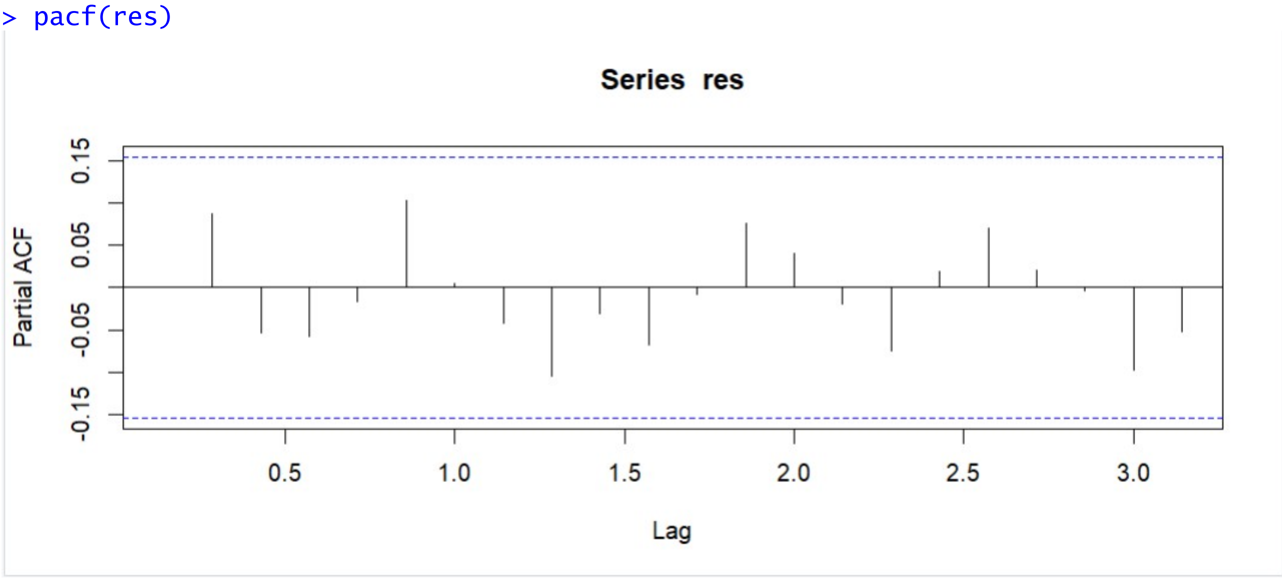
```
> trend=data.lm$fitted.values
> res=data.ts-(trend+sais)
> plot(res, xlab="semaines", ylab="residus")
```



Les résidus semblent tourner autour de 0, sans dépendance au passé. Cela ressemble à du bruit.

5) On trace le graphique des autocorrélations partielles de la série des résidus :

```
> pacf(res)
```



Un modèle AR(2) n’est pas convenable, aucun « trait » ne dépasse les pointillés bleus. Les résidus consistent simplement en du bruit gaussien.

6) $X_t = e_t$

Où e_t correspond au bruit gaussien à l’instant t.
On est sur un modèle AR(0)

7) La prédiction correspond à la tendance additionnée à la composante saisonnière. On fait la prédiction sur 24 semaines :

```
> fit<-arima(trend+sais, order=c(0,0,0))
> predict(fit, n.ahead = 24)
$pred
Time Series:
Start = c(24, 1)
End = c(27, 3)
Frequency = 7
 [1] 25.69679 25.69679 25.69679 25.69679 25.69679
 [6] 25.69679 25.69679 25.69679 25.69679 25.69679
[11] 25.69679 25.69679 25.69679 25.69679 25.69679
[16] 25.69679 25.69679 25.69679 25.69679 25.69679
[21] 25.69679 25.69679 25.69679 25.69679

$se
Time Series:
Start = c(24, 1)
End = c(27, 3)
Frequency = 7
 [1] 28.10579 28.10579 28.10579 28.10579 28.10579
 [6] 28.10579 28.10579 28.10579 28.10579 28.10579
[11] 28.10579 28.10579 28.10579 28.10579 28.10579
[16] 28.10579 28.10579 28.10579 28.10579 28.10579
[21] 28.10579 28.10579 28.10579 28.10579
```