

## TP du CC2–Régressions

On considère la série statistique du nombre de nouveaux cas de Covid-19 hebdomadaire depuis le premier déconfinement. Les modèles épidémiques (les plus simples) nous informent que la dynamique suit une tendance exponentielle lorsqu’aucun élément extérieur ne vient modifier l’évolution de l’épidémie.

À partir de quand aurait-on pu prévoir l’ampleur de la seconde vague ?

### Nombre total de cas testés positifs au SARS-CoV-2

du 10 au 16/06	3125
du 17 au 23/06	3551
du 24 au 30/06	3534
du 01 au 07/07	4001
du 08 au 14/07	4494
du 15 au 21/07	4440
du 22 au 28/07	6456
du 29 au 04/08	8530
du 04 au 11/08	11838
du 12 au 18/08	17095
du 19 au 25/08	26891
du 26 au 01/09	37849
du 02 au 08/09	49517
du 09 au 15/09	59580
du 16 au 22/09	72965
du 23 au 29/09	82621
du 30 au 06/10	84073
du 07 au 13/10	121709
du 14 au 20/10	174273
du 21 au 27/10	267950
du 28 au 03/11	304068
du 04 au 10/11	326896

Données Santé Publique France

Vous enverrez par e-mail à l’adresse [julien.grepat@univ-grenoble-alpes.fr](mailto:julien.grepat@univ-grenoble-alpes.fr) un fichier contenant votre code R, vos interprétations et les éventuels graphes R.

## Correction

- On saisit la série de données covid :

```
covid=c(3125,3551,3534,4001,4494,4440,6456,8530,11838,17095,26891,  
37849,49517,59580,72965,82621,84073,121709,174273,267950,304068,326896)
```

Ainsi que la numérotation des semaines :

```
sem=seq(1:length(covid))
```

- On trace covid en fonction de sem :

```
plot(sem,covid)
```

**On s'aperçoit qu'effectivement, la dynamique semble exponentielle**

- On propose un changement de variable en supposant que la dynamique exponentielle s'ajoute à un niveau de contamination initial de 3000.

```
lncovid=log(covid-3000)
```

```
plot(sem,covid)
```

**En traçant, on s'aperçoit que le changement de variable semble bon et qu'à partir de la semaine 6, la dynamique semble installée.**

- On lance une régression linéaire de lncovid sur sem, semaine par semaine. On va calculer à chaque semaine les coefficients  $a, b, r$  issus de la régression linéaire sur les données qu'on considère jusqu'à la semaine en question.
- Calcul de  $a$  : on rappelle que de manière générale,

$$a = \frac{\text{cov}(x, y)}{\text{Var}(x)} = \frac{\frac{1}{n} \sum_{i=1}^n x_i y_i - \left(\frac{1}{n} \sum_{i=1}^n x_i\right) \times \left(\frac{1}{n} \sum_{i=1}^n y_i\right)}{\frac{1}{n} \sum_{i=1}^n x_i^2 - \left(\frac{1}{n} \sum_{i=1}^n x_i\right)^2}.$$

Notons qu'à la  $n^{\text{ieme}}$  semaine l'effectif des données est  $n$  qui correspond à la valeur de sem. Il vient que

```
a=(cumsum(sem*lncovid)/sem-cumsum(sem)*cumsum(lncovid)/(sem)^ 2)/
(cumsum(sem^ 2)/sem-(cumsum(sem)/sem)^ 2)
```

- On a également  $b$  qui vérifie

$$b = \bar{y} - a \times \bar{x} = \left(\frac{1}{n} \sum_{i=1}^n y_i\right) - a \left(\frac{1}{n} \sum_{i=1}^n x_i\right).$$

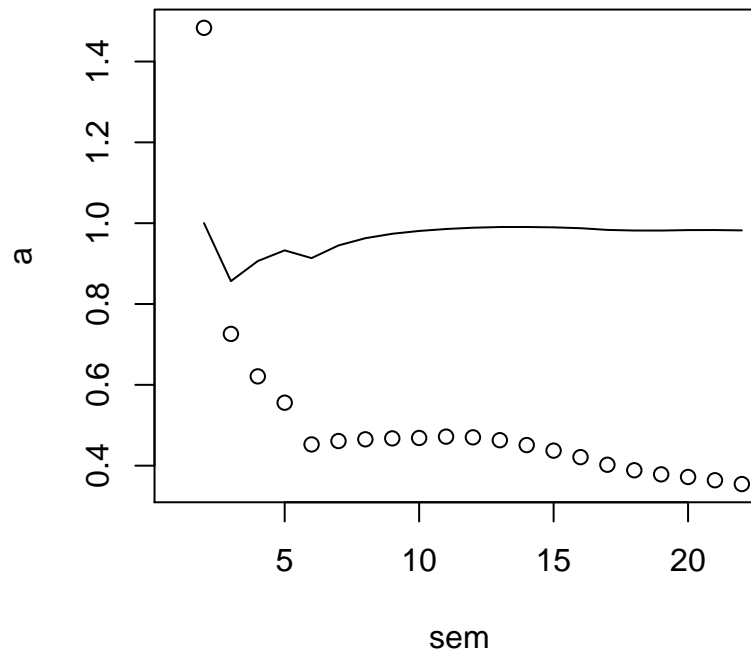
D'où :

```
b=cumsum(lncovid)/sem-a*cumsum(sem)/sem
```

- On peut (si on le souhaite) calculer le coefficient  $r$  de manière analogue à  $a$  ( $r = \text{cov}(x, y) / \sqrt{\text{var}(x)\text{var}(y)}$ ):

```
r=(cumsum(sem*lncovid)/sem-cumsum(sem)*cumsum(lncovid)/(sem)^ 2)/
((cumsum(sem^ 2)/sem-(cumsum(sem)/sem)^ 2)*
(cumsum(lncovid^ 2)/sem-(cumsum(lncovid)/sem)^ 2))^ 0.5
```

On obtient le graphe suivant :



En points, le coefficient  $a$  et en ligne continue le coefficient  $r$ .

**On s'aperçoit qu'à partir de la semaine 6 ou 7 (15-21 juillet), le coefficient  $a$  bouge peu et  $r$  est très proche de 1, donc le changement de variable est bon et la régression est proche du nuage. Ainsi, dans les semaines qui suivaient la semaine 7, on pouvait se dire que l'épidémie suivait bien une dynamique exponentielle, ce qui amène bien à une explosion des cas.**

- On évitera de faire une prévision avec le modèle de la semaine 6 pour la semaine 22 par exemple car il y a des facteurs qui ont modifié la dynamique exponentielle (comme la baisse du nombre de test hebdomadaires).
- On notera également qu'il est facile à posteriori d'affirmer qu'on pouvait savoir ce qui se passerait dès la semaine 6. Mais il est toujours difficile de prévoir l'avenir avec un modèle descriptif. Néanmoins, de proche en proche, on pouvait prévoir ce qui se passerait à une ou deux semaines.