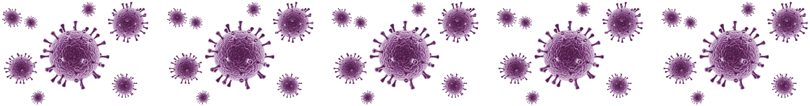
Modèles Linéaires Appliqués / Régression Poisson vs. Binomiale

Arthur Charpentier

UQAM

Hiver 2020 - COVID-19 # 21



Donnés de Mortalité

via https://www.mortality.org/

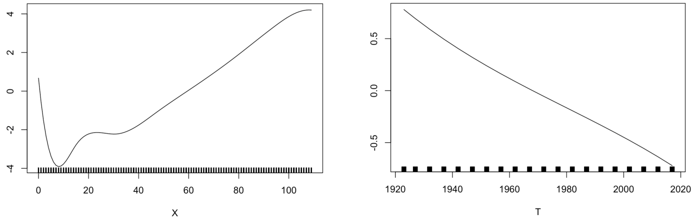
```
> loc = "http://freakonometrics.free.fr/CANmortalite.
    RData"
> download.file(loc,"CAN.RData")
> load("CAN.RData")
> str(base)
'data.frame': 2196 obs. of  4 variables:
 $ D: num  99465 14575 6671 4645 3668 ...
 $ E: num  843291 854232 866846 871320 870308 ...
 $ X: num  0 1 2 3 4 5 6 7 8 9 ...
 $ T: num  1923 1923 1923 1923 1923 ...
```

x = âge (entre 0 et 110), t = année (entre 1923 et 2017)
$D_{x,t}$ = nombre de décès observés (en 5 ans)
$E_{x,t}$ = exposition (sur 5 ans)

Modèle de Poisson

$$D_{x,t} \sim \mathcal{P}(E_{x,t} \exp(s_1(x) + s_2(t)))$$

```
> library(gam)
> reggam = gam(D~bs(X, df = 9)+bs(T)+offset(log(E)),
    data=base,family=poisson)
> plot(reggam)
```

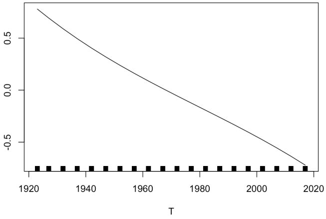


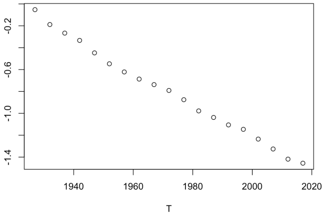
effet linéaire de t...? régresser sur t en tant que facteur

Modèle de Poisson

$$D_{x,t} \sim \mathcal{P}(E_{x,t} \exp(s_1(x) + s_2(t)))$$

```
> reg = glm(D~bs(X, df = 9)+as.factor(T)+offset(log(E)
    ),data=base,family=poisson)
> t=unique(base$T)
> plot(t[-1],coefficients(reg)[11:length(coefficients(
    reg))])
```



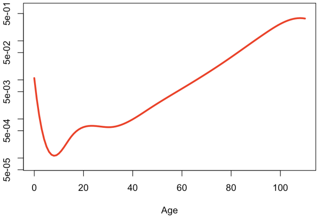


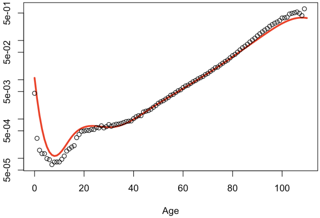
on peut régresser sur t

Modèle de Poisson

$$D_{x,t} \sim \mathcal{P}(E_{x,t} \exp(s_1(x) + s_2(t)))$$

```
> reg = glm(D~bs(X,df = 9)+T+offset(log(E)),data=base,
    family=poisson)
> newxt = data.frame(X=0:110,T=2020,E=1)
> p = predict(reg,newdata=newxt,type="response")
> B = base[base$T==2017,]
> plot(newxt$X,p,log="y",col="red")
> points(B$X,B$D/B$E)
```

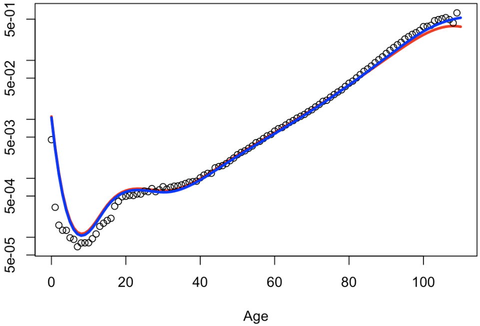




Modèle Binomial

$$D_{x,t} \sim \mathcal{B}(E_{x,t}, p_{x,t}) \text{ où } p_{x,t} = \frac{\exp[\beta_0 + \beta_1 t + s(x)]}{1 + \exp[\beta_0 + \beta_1 t + s(x)]}$$

```
> base=base[base$D<base$E,]
> reg2 = glm(D/E~bs(X,df = 9)+T,data=base,family=
    binomial,weights=E)
> p2 = predict(reg2,newdata=newxt,type="response")
> lines(newxt$X,p2,col="blue")
```



Impact sur l'Espérance de Vie

Sur le modèle de Poisson

```
> summary(reg)

Coefficients:
             Estimate  Std. Error z value Pr(>z)
(Intercept) 2.609e+01  1.908e-02  1367.7   <2e-16 ***
```

(...)

```
T          -1.514e-02  9.705e-06 -1560.3   <2e-16 ***
```

i.e. $\widehat{\beta}_1 = -0.01514312$.

En 2020, l'espérance de vie à la naissance est

```
> sum(cumprod(1-p))
[1] 81.05558
```

Impact sur l'Espérance de Vie

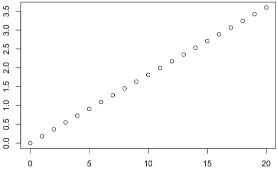
En 2020, l'espérance de vie à la naissance est

```
> sum(cumprod(1-p))
[1] 81.05558
```

et en 2020, l'espérance de vie à la naissance est

```
> sum(cumprod(1-p*exp(-0.01514312)))
[1] 81.238
```

```
> e = function(h)
sum(cumprod(1-p*exp(-1.514e-02*h)))
    -sum(cumprod(1-p))
> plot(0:20,Vectorize(e)(0:20))
```



ce qui correspond à un gain d'espérance de vie de $+0.1805$ année tous les ans.