

# Quand le contrôle des facteurs de confusion est plus compliqué qu'on ne l'imagine

Marie Laure Delignette-Muller

06 janvier, 2022

## Un exemple "jouet" réaliste

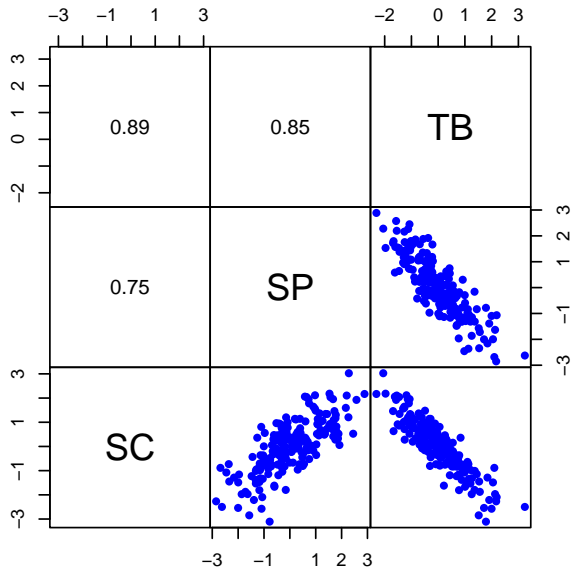
- ▶ **Surpoids du chien** (normalisé) :  $SC$
- ▶ **Surpoids du propriétaire** (normalisé) :  $SP$
- ▶ **Temps de ballade commun** (normalisé) :  $TB$

**Avec comme relations causales**  $TB \longrightarrow SC$  et  $TB \longrightarrow SP$

Simulation de données à partir de

- ▶  $TB \sim N(0, 1)$
- ▶  $SP \sim N(\alpha_{SP} + \beta_{SP} \times TB, \sigma_{SP})$
- ▶  $SC \sim N(\alpha_{SC} + \beta_{SC} \times TB, \sigma_{SC})$

## Visualisation des données simulées



## Modèle linéaire : SC en fonction de SP

```
mSP <- lm(SC ~ SP)
summary(mSP)$coefficients
```

##	Estimate	Std. Error	t value	Pr(> t )
## (Intercept)	0.00051	0.051	0.01	9.9e-01
## SP	0.70752	0.045	15.76	8.6e-37

**Bien sûr, si on ne prend pas en compte le facteur TB dans le modèle, on met en évidence une corrélation entre SC et SP liée ici au facteur causal commun TB.**

## Modèle linéaire : SC en fonction de SP et du facteur de confusion TB

```
mSPTB <- lm(SC ~ SP + TB)
summary(mSPTB)$coefficients
```

##	Estimate	Std. Error	t value	Pr(> t )
## (Intercept)	0.017	0.034	0.49	6.2e-01
## SP	-0.042	0.057	-0.74	4.6e-01
## TB	-1.058	0.069	-15.38	1.4e-35

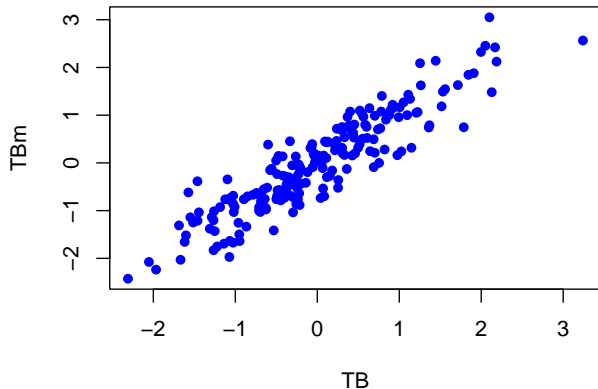
**En prenant en compte le facteur de confusion TB, comme attendu on ne met plus en évidence d'effet significatif de SP sur SC.**

*Alors ça marche ?*

*Oui mais qu'en est-il dans la vraie vie ?*

Mais que se passe-t-il si la mesure de TB disponible est bruitée (réaliste dans la vraie vie) ?

$$TB_m \sim N(TB, \sigma_{TB_m})$$



## Modèle linéaire : SC en fonction de SP et de TBm

```
mSPTBm <- lm(SC ~ SP + TBm)
summary(mSPTBm)$coefficients
```

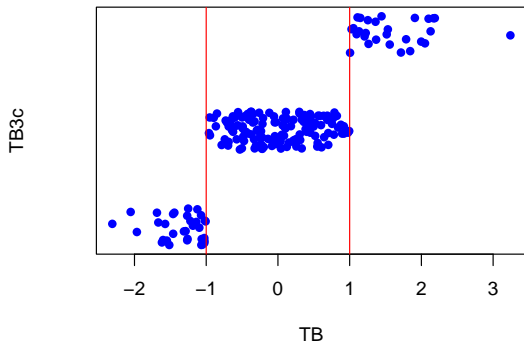
##	Estimate	Std. Error	t value	Pr(> t )
## (Intercept)	0.004	0.042	0.095	9.2e-01
## SP	0.291	0.057	5.101	7.9e-07
## TBm	-0.613	0.064	-9.600	3.8e-18

**Même en prenant en compte le facteur de confusion TBm mesurée avec une certaine erreur de mesure, on met en évidence un effet significatif de SP sur SC.**

Autre cas réaliste, si on n'a à disposition qu'une mesure qualitative de TB (par ex. en 3 classes)

TB transformé en variable qualitative (TB3c) à trois classes :

$] -4; -1]$ ,  $] -1; 1]$ ,  $] 1; 4]$ .





## Modèle linéaire : SC en fonction de SP et de TB3c

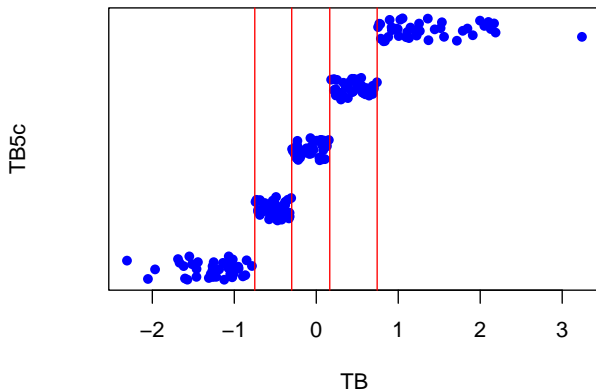
```
mSPTB3c <- lm(SC ~ SP + TB3c)
summary(mSPTB3c)$coefficients
```

##	Estimate	Std. Error	t value	Pr(> t )
## (Intercept)	0.94	0.137	6.9	7.1e-11
## SP	0.32	0.057	5.7	3.9e-08
## TB3c(-1,1]	-0.91	0.146	-6.2	3.1e-09
## TB3c(1,4]	-2.12	0.232	-9.1	7.8e-17

**On met encore en évidence un effet significatif de SP sur SC !**

## Et si la discrétisation est moins grossière et équilibrée

TB transformé en variable qualitative (TB5c) à cinq classes équilibrée, dont les limites sont définies par les quintiles.



## Modèle linéaire : SC en fonction de SP et de TB5c

```
mSPTB5c <- lm(SC ~ SP + TB5c)
summary(mSPTB5c)$coefficients
```

##	Estimate	Std. Error	t value	Pr(> t )
## (Intercept)	1.11	0.125	8.9	4.1e-16
## SP	0.14	0.061	2.3	2.2e-02
## TB5c(-0.75,-0.299]	-0.68	0.139	-4.9	2.3e-06
## TB5c(-0.299,0.165]	-0.99	0.151	-6.5	5.2e-10
## TB5c(0.165,0.743]	-1.45	0.169	-8.6	3.3e-15
## TB5c(0.743,4]	-2.34	0.212	-11.0	2.8e-22

**On met encore en évidence un effet significatif de SP sur SC !**

## Conclusion quant à la possibilité de prendre en compte des facteurs de confusion dans un modèle linéaire

**Une problématique de ce genre est des plus réalistes et je vous laisse en imaginer les conséquences !**

Une grande prudence s'impose donc lorsque qu'on interprète les résultats d'un modèle linéaire sur des données d'observationnelles (donc avec des variables explicatives non contrôlées).

### Références :

- ▶ *Christenfeld, N. J., Sloan, R. P., Carroll, D., & Greenland, S. (2004). Risk factors, confounding, and the illusion of statistical control. Psychosomatic medicine, 66(6), 868-875.*
- ▶ *Westfall, J., & Yarkoni, T. (2016). Statistically controlling for confounding constructs is harder than you think. PloS one, 11(3), e0152719.*