

# Tooth Growth

*Sophie Donnet*

*Printemps 2020*

## Contexte

Nous considérons le jeu de données `ToothGrowth`. La réponse  $Y$  est la longueur des odontoblastes (cellules intervenant dans la croissance des dents) chez  $n = 60$  cochons de Guinée. Chaque animal a reçu une des trois doses possibles de vitamine C (0.5, 1, and 2 mg/day) par le biais d'une des deux méthodes d'administration (jus d'orange OJ ou acide ascorbique VC). On s'intéresse à l'influence de ces facteurs sur la croissance dentaire. Le jeu de données est représenté par le box-plot suivant. (Pour l'exercice, nous avons transformé la variable quantitative `dose` en un facteur `doselevel`).

## Statistiques descriptives

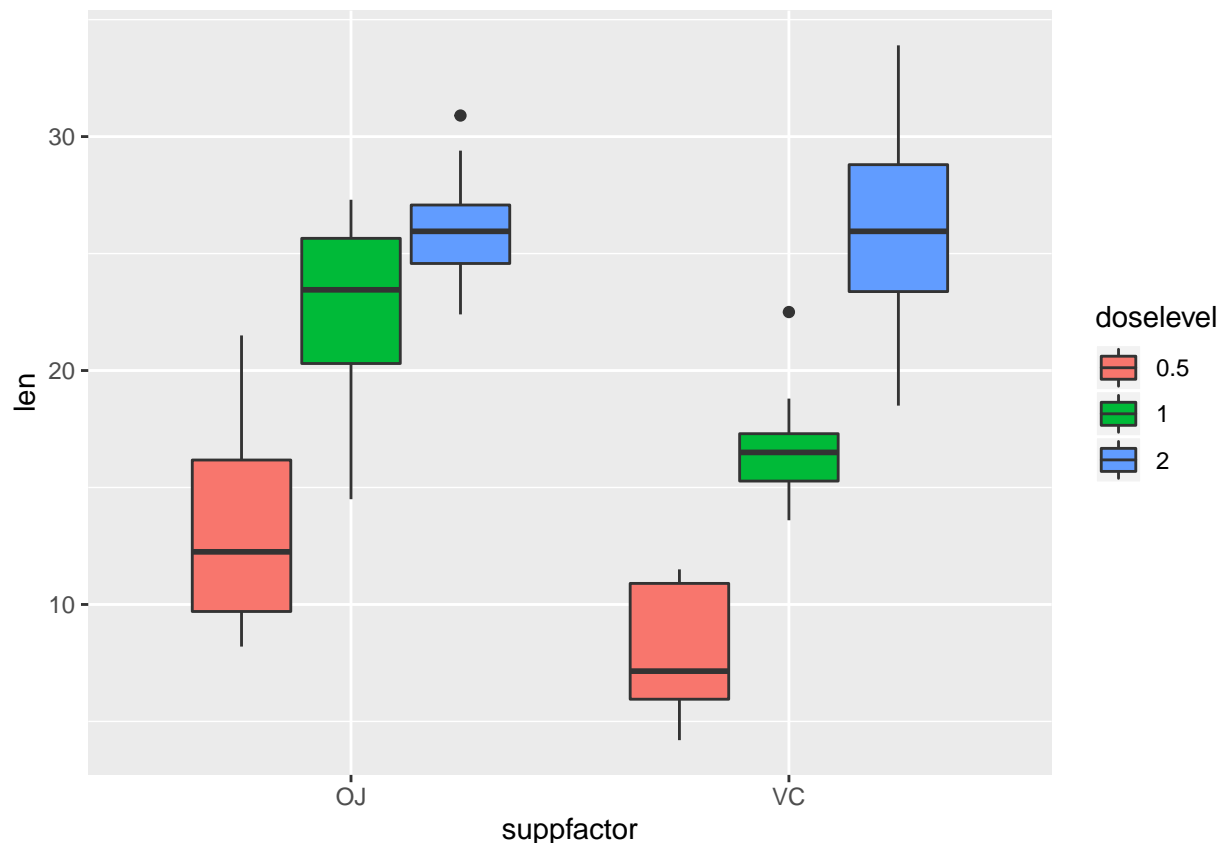
```
ToothGrowth$doselevel = as.factor(ToothGrowth$dose)
names(ToothGrowth)=c('len', 'suppfactor', 'dose', 'doselevel')
summary(ToothGrowth)
```

##	len	suppfactor	dose	doselevel
##	Min. : 4.20	OJ:30	Min. :0.500	0.5:20
##	1st Qu.:13.07	VC:30	1st Qu.:0.500	1 :20
##	Median :19.25		Median :1.000	2 :20
##	Mean :18.81		Mean :1.167	
##	3rd Qu.:25.27		3rd Qu.:2.000	
##	Max. :33.90		Max. :2.000	

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.5.3
```

```
ggplot(ToothGrowth, aes(y = len,x=suppfactor,fill = doselevel))+ geom_boxplot()
```



## Modélisation et inférence

1- Ecrire le modèle correspondant aux instruction suivantes (sans oublier les hypothèses et les gammes de variation des indices).

```
mod_compl = lm(len~suppfactor*doselevel,data=ToothGrowth)
summary(mod_compl)
```

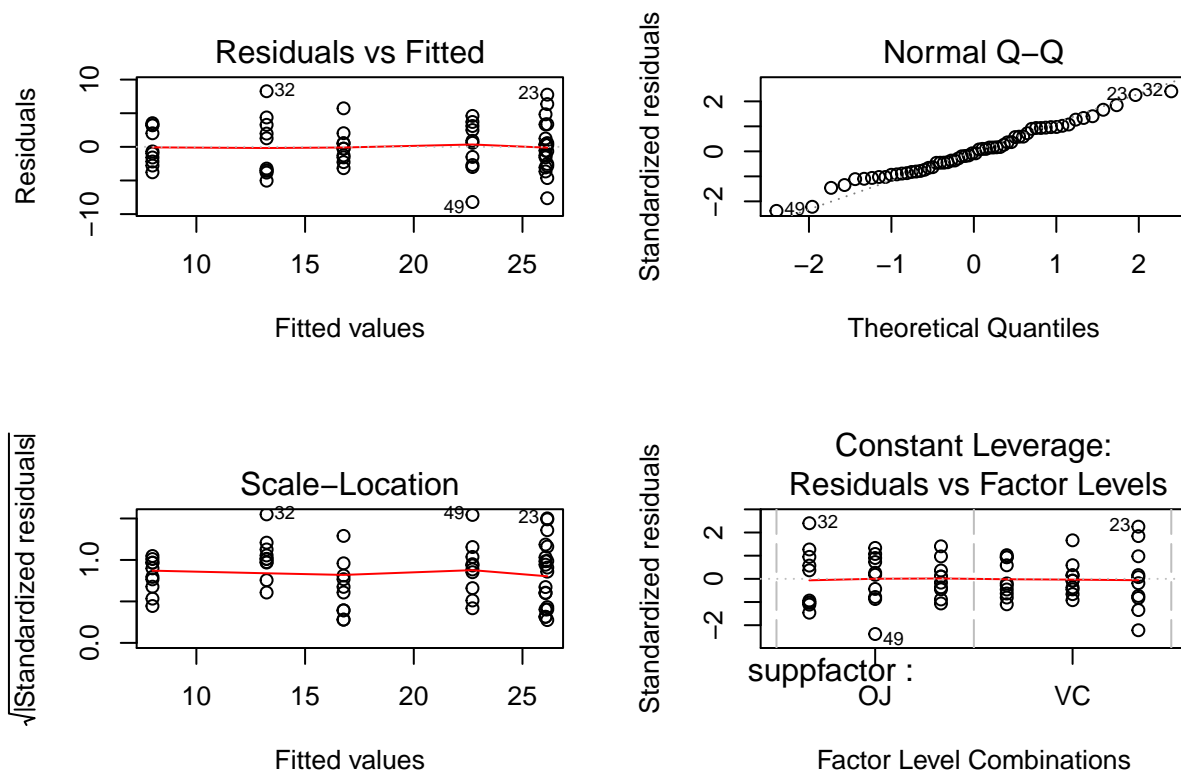
```
##
## Call:
## lm(formula = len ~ suppfactor * doselevel, data = ToothGrowth)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -8.20  -2.72  -0.27   2.65   8.27
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      13.230      1.148  11.521 3.60e-16 ***
## suppfactorVC       -5.250      1.624  -3.233 0.00209 **
## doselevel1         9.470      1.624   5.831 3.18e-07 ***
## doselevel2        12.830      1.624   7.900 1.43e-10 ***
## suppfactorVC:doselevel1 -0.680      2.297  -0.296 0.76831
## suppfactorVC:doselevel2  5.330      2.297   2.321 0.02411 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 3.631 on 54 degrees of freedom
## Multiple R-squared:  0.7937, Adjusted R-squared:  0.7746
## F-statistic: 41.56 on 5 and 54 DF,  p-value: < 2.2e-16
```

2- Quelles sont les contraintes utilisées par le logiciel R ? (retrouver ces contraintes dans les sorties précédentes)

3- Les hypothèses du modèle linéaire sont elles vérifiées?

```
par(mfrow=c(2,2))
plot(mod_compl)
```



4- Doit-on garder les interactions?.

```
mod_add = lm(lm(len~supfactor + doselevel,data=ToothGrowth))
anova(mod_add, mod_compl)
```

```
## Analysis of Variance Table
##
## Model 1: len ~ supfactor + doselevel
## Model 2: len ~ supfactor * doselevel
##   Res.Df    RSS Df Sum of Sq    F Pr(>F)
## 1      56 820.43
## 2      54 712.11  2   108.32 4.107 0.02186 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

5- A-t-on un effet significatif des facteurs sur la croissance?

```
anova(mod_compl)
```

```
## Analysis of Variance Table
```

```
##
```

```
## Response: len
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
## suppfactor	1	205.35	205.35	15.572	0.0002312	***
## doselevel	2	2426.43	1213.22	92.000	< 2.2e-16	***
## suppfactor:doselevel	2	108.32	54.16	4.107	0.0218603	*
## Residuals	54	712.11	13.19			

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```