

Biostatistics club: Linear regression with R

Claire Burny

2017.02.28

Work on an R pre-loaded dataframe

(A dataframe is a matrix with columns of different types but of same length.)

help(ToothGrowth) # informations about dataset

The response is the length of odontoblasts (teeth) in each of 10 guinea pigs at each of three dose levels of Vitamin C (0.5, 1, and 2 mg) with each of two delivery methods (orange juice or ascorbic acid).

str(ToothGrowth) # on what do I work?

```
'data.frame': 60 obs. of 3 variables:  
 $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...  
 $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...  
 $ dose: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

are there links between variables?

pairs(ToothGrowth)

ANalysis Of VAriance

ANOVA à 1 facteur = généralisation du t-test pour la comparaison de 2 moyennes

$$\begin{cases} Y_1 \sim N(\mu_1, \sigma^2) \\ Y_2 \sim N(\mu_2, \sigma^2) \end{cases}$$

→ Si $\bar{Y}_1 - \bar{Y}_2$ est trop grande, on rejette H_0 , c-à-d l'égalité des moyennes.

Quelle est la loi de la différence $\bar{Y}_1 - \bar{Y}_2$?

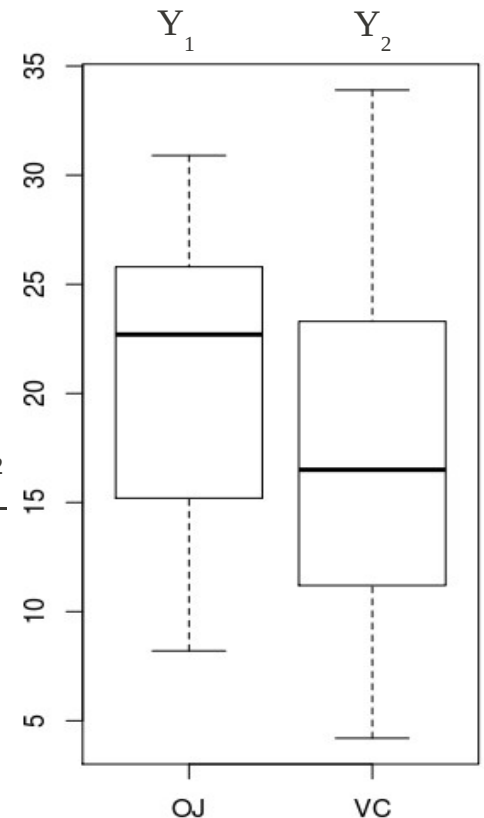
$$\begin{cases} E(\bar{Y}_1 - \bar{Y}_2) = \mu_1 - \mu_2 \\ \text{Var}(\bar{Y}_1 - \bar{Y}_2) = \text{var}(\bar{Y}_1) + \text{var}(\bar{Y}_2) - 2 \text{cov}(Y_1, Y_2) = \text{var}(\bar{Y}_1) + \text{var}(\bar{Y}_2) = \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2} \end{cases}$$

$$\rightarrow \bar{Y}_1 - \bar{Y}_2 \sim N(\mu_1 - \mu_2, \sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}})$$

$$\leftrightarrow Z = \frac{(\bar{Y}_1 - \bar{Y}_2) - (\mu_1 - \mu_2)}{\sqrt{\sigma^2 \times (\frac{1}{n_1} + \frac{1}{n_2})}} \sim N(0, 1)$$

Sous H_0 , $\mu_1 - \mu_2 = 0$

$$\leftrightarrow T = \frac{(\bar{Y}_1 - \bar{Y}_2) - (\mu_1 - \mu_2)}{\sqrt{S^2 \times (\frac{1}{n_1} + \frac{1}{n_2})}} \sim T(n-2) \text{ with } n = n_1 + n_2$$



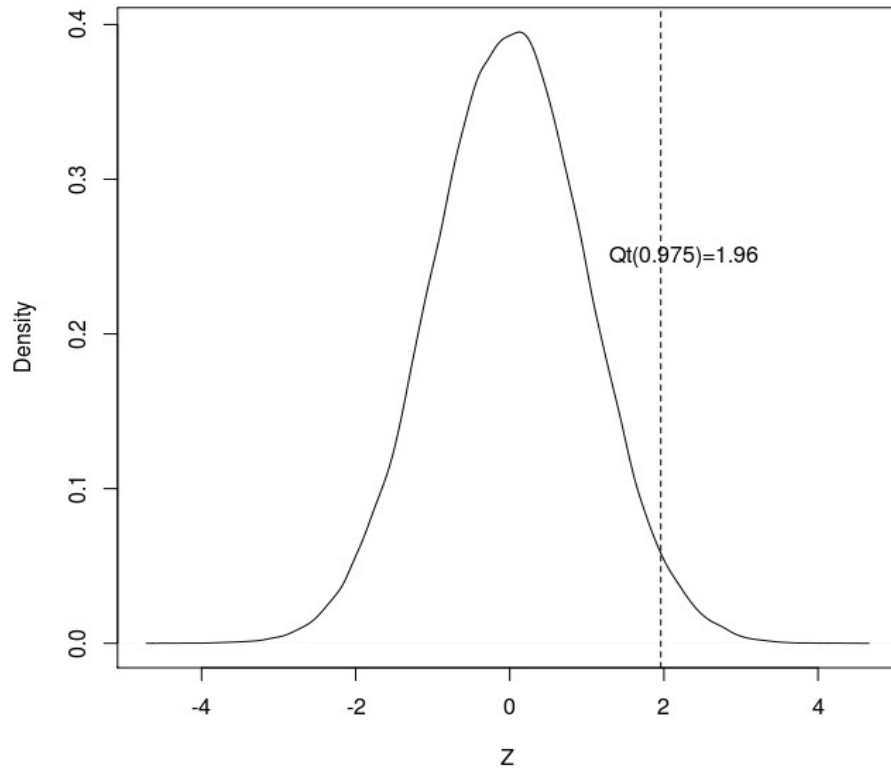
ANOVA

Under H0:

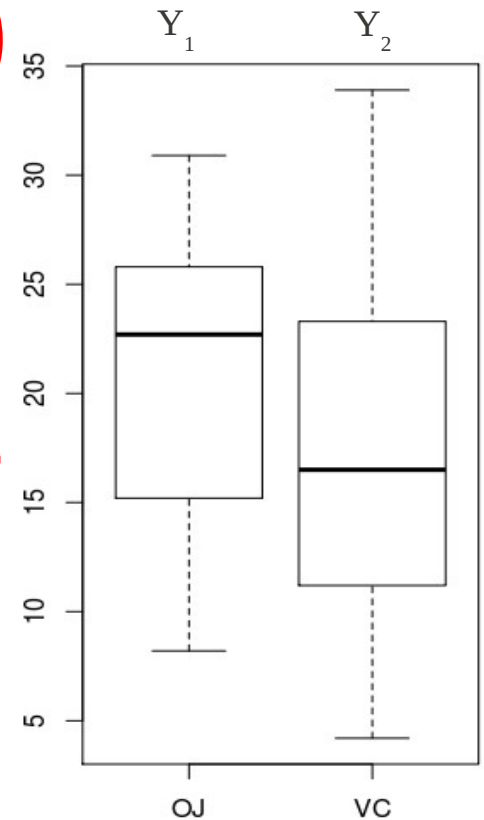
$$Z = \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{\sigma^2 \times \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = N(0, 1)$$

$$|Z| > z_{1-\alpha/2}$$

Z law



$$T = \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{S^2 \times \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = t(n-2)$$
$$|T| > t(n-2)_{1-\alpha/2}$$



Ok but where is the
Variance (ANALYSIS OF
Variance)?



ANOVA

$$SCET = SCEM + SCER = \sum_J n_j (\bar{Y}_j - \bar{Y})^2 + \sum_J \sum_{n_j} (y_{ij} - \bar{Y}_j)^2 \text{ with } \bar{Y} = \frac{n_1 \bar{Y}_1 + n_2 \bar{Y}_2}{n_1 + n_2}$$

Test de F de l'analyse de la variance:

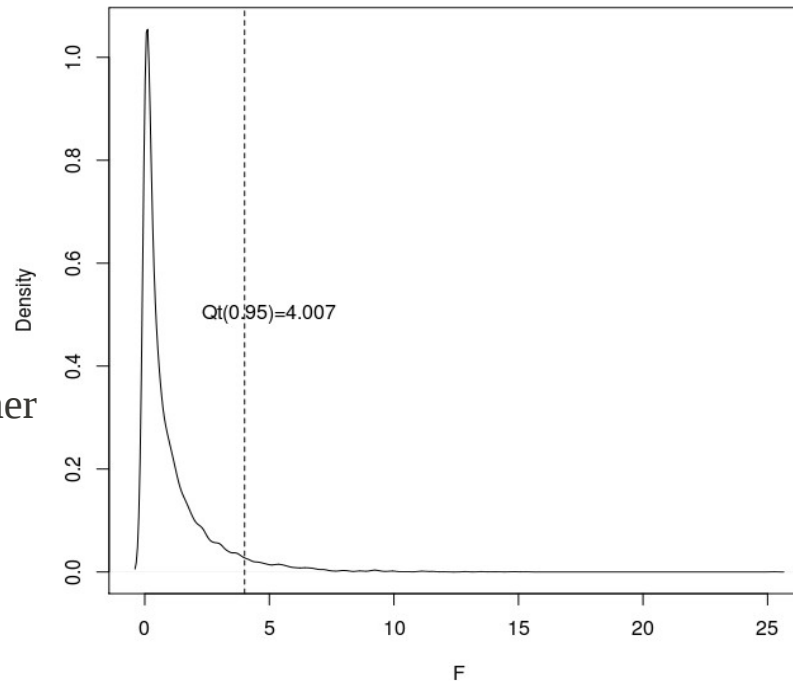
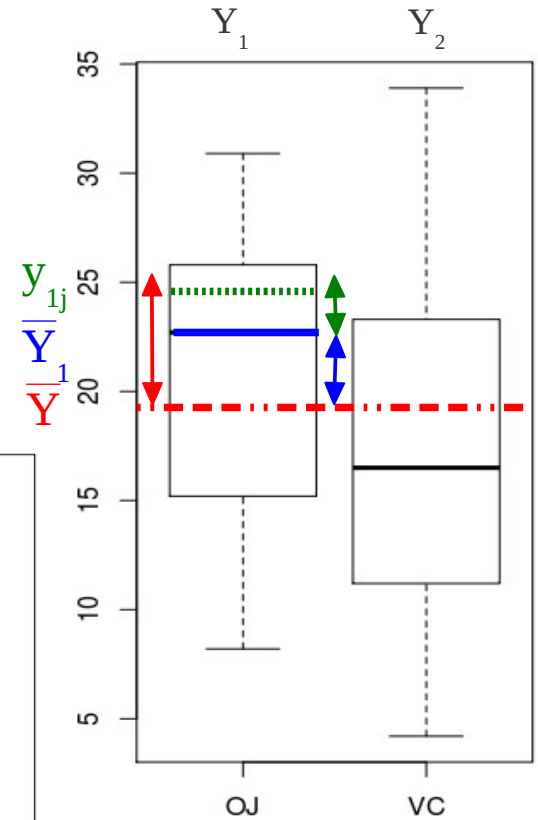
$$F = \frac{\frac{SCEM}{p-1}}{\frac{SCER}{n-p}}$$

$$F(1, n-2) = t(n-2)^2$$

$$T = \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{S^2 \times \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

F(1, 58) law

$$F > f(p-1, n-p)_{1-\alpha}$$



ddl = info réellement indépendante
= nb observations – nb paramètres à estimer

Eventually run the t-test...

```
# supplement covariable
```

```
len_OJ <- ToothGrowth[which(ToothGrowth$supp == "OJ"), ]$len
```

```
len_VC <- ToothGrowth[which(ToothGrowth$supp == "VC"), ]$len
```

→ Are the distributions gaussian? Are the variances equal?

```
shapiro.test(len_OJ); shapiro.test(len_VC)
```

```
var.test(len_OJ, len_VC)
```

→ perform the t-test

```
t.test(len_OJ, len_VC, var.equal = TRUE)
```

Two Sample t-test

data: len_OJ and len_VC

$t = 1.9153$, $df = 58$, $p\text{-value} = 0.06039$ / $t = 1.9153$, $df = 55.309$, $p\text{-value} = 0.06063$

alternative hypothesis: true difference in means is not equal to 0

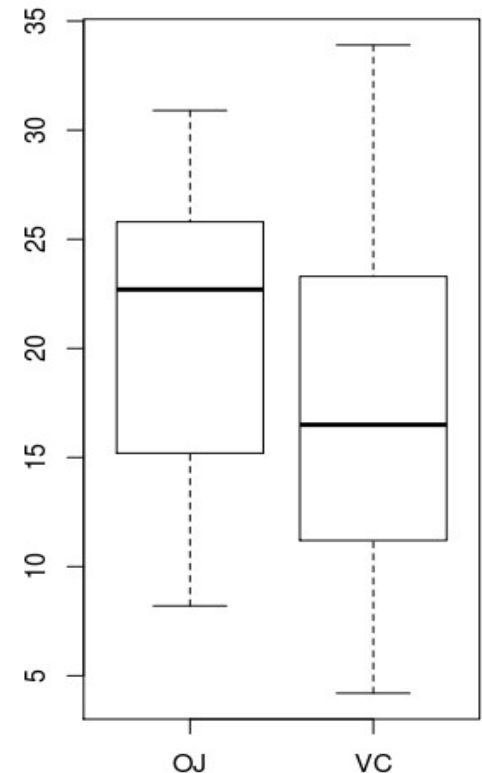
95 percent confidence interval:

-0.1670064 7.5670064

sample estimates:

mean of x mean of y

20.66333 16.96333



Régression linéaire

On veut expliquer une variable, la longueur Y , en fonction des variables „traitement“ et „dose“.

→ effet traitement (yes → output du t-test)

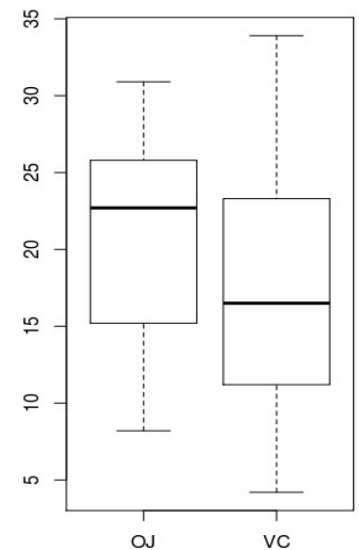
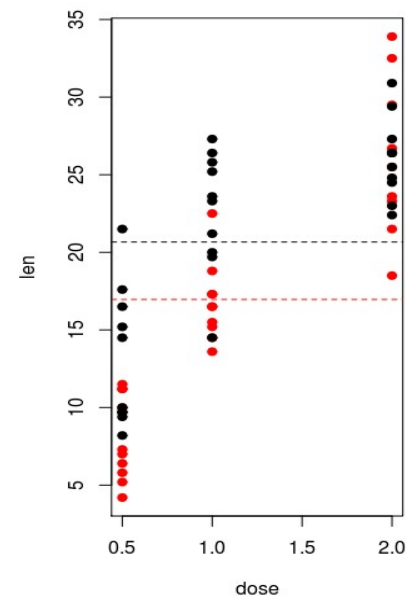
→ effet linéaire de la dose: en théorie, $y_i = \beta_0 + \beta_1 \text{dose}_i$

Or Y est influencée par plusieurs facteurs latents, on peut commettre des erreurs de mesure, une même valeur de dose ne conduit pas à une même longueur... d'où le modèle aléatoire:

$$y_i = \beta_0 + \beta_1 \text{dose}_i + \varepsilon_i = \text{partie déterministe} + \text{bruit}$$

Aucune hypothèse n'est nécessaire pour estimer les valeurs des coefficients par la méthode des **moindres carrés** mais certaines sont requises pour tester leur valeur:

- indépendance des mesures
- variables explicatrices contrôlées
- *relation linéaire supposée vraie*
- *résidus gaussiens*
- *résidus indépendants des variables explicatrices*
- *pas de mesure aberrante*



Construction du modèle 1

`# model with dose covariable`

`lm1 <- lm(len ~ dose, data = ToothGrowth)` → On construit $y_i = \beta_0 + \beta_1 \text{dose}_i + \varepsilon_i$

`summary(lm1)`

`lm(formula = len ~ dose, data = ToothGrowth)`

Residuals:

Min	1Q	Median	3Q	Max
-8.4496	-2.7406	-0.7452	2.8344	10.1139

$$T_{\bar{\beta}_0} = \frac{\bar{\beta}_0 - 0}{\sqrt{\frac{\sigma^2}{n}}}$$

$$T_{\bar{\beta}_1} = \frac{\bar{\beta}_1 - 0}{\sqrt{\frac{\sigma^2}{n}}}$$

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	7.4225	1.2601	5.89	2.06e-07 ***
dose	9.7636	0.9525	10.25	1.23e-14 ***

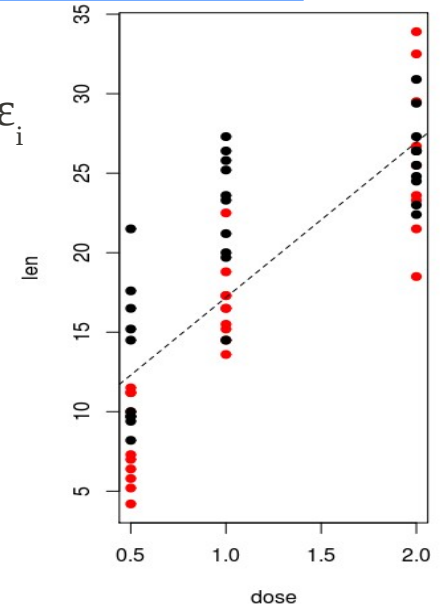
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.601 on 58 degrees of freedom

Multiple R-squared: 0.6443, Adjusted R-squared: 0.6382

F-statistic: 105.1 on 1 and 58 DF, p-value: 1.233e-14

`abline(lm1, col = 1, lty = 2)`



Quelques remarques:

MCO: $(\bar{\beta}_0, \bar{\beta}_1)$ approxime (β_0, β_1) et minimise $\sum (y_i - (\beta_0 + \beta_1 x_i))^2$

$\bar{\beta}_1$: effet de la dose sur la longueur!

Pour tester si $\bar{\beta}_0$ est non significativement $\neq 8$:

`2*(1-pt(abs((7.4225-8)/1.183), dim(ToothGrowth)[1]-2))`

CI_{95%} de $\bar{\beta}_1$: $P(l < \bar{\beta}_1 \leq r) = 95\%$

La loi de $\bar{\beta}_1$ étant symétrique, cela revient à chercher:

$$P(\bar{\beta}_1 \leq r) = 97.5\% - P(\bar{\beta}_1 > l) = 2.5\%$$

roit $r = \bar{\beta}_1 + qt(0.975, n-2)$

ANOVA du modèle 1

L'analyse de la variance a 2 objectifs:

→ tester si $\bar{\beta}_1$ est non significativement différent de 0 via un test de F

→ estimer le coef de détermination, pourcentage de la variabilité totale des données expliquée par le modèle

$$F = \frac{\frac{SCEM}{p-1}}{\frac{SCER}{n-p}}$$

$$F > f(p-1, n-p)_{1-\alpha}$$

`anova(lm1)`

Analysis of Variance Table

Response: len

Somme des carrés des écarts
Somme des carrés des écarts/ddl
Statistique du test F

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
dose	1	2224.3	2224.30	105.06	1.233e-14 ***
Residuals	58	1227.9	21.17		

Le test F pour savoir si $\bar{\beta}_1$ est significativement \neq de 0 est **unilatéral**

1-pf(2224.3/1227.9, 1, 58)

SCE	ddl	Source de variabilité
$SCE_T = \sum_{i=1}^n (y_i - \bar{y})^2$	$n - 1$	Variabilité totale des données
$SCE_M = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$	$p - 1$	Variabilité expliquée par le modèle
$SCE_R = \sum_{i=1}^n (y_i - \hat{y}_i)^2$	$n - p$	Variabilité résiduelle

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

Coefficient de détermination

Le coefficient de détermination est une proportion:

$$R^2 = \frac{SCEM}{SCEM + SCER} = 1 - \frac{SCER}{SCEM + SCER}$$
$$R^2_{adj} = \frac{n-1}{n-p} \frac{SCEM}{SCEM + SCER}$$

qui mesure l'adéquation du modèle aux data; plus il est élevé, meilleur est le modèle. On le retrouve dans le summary (**Multiple R-squared: 0.6443**); *la dose explique 64% de la variabilité de la longueur, lorsque l'on suppose une relation linéaire entre ces deux variables.*

Tester si R^2 est significativement $\neq 0$ revient à tester si $\bar{\beta}_1$ est significativement $\neq 0$.

$$\bar{\beta}_1 = \text{cov}(\text{dose}, \text{len}) / \text{var}(\text{dose}) = \text{cor}(\text{dose}, \text{len}) \sigma_{\text{len}} / \sigma_{\text{dose}}$$

De plus, **R^2 correspond au carré du coefficient de corrélation.** On rappelle que s'il existe une relation linéaire entre deux variables aléatoires, le coefficient de corrélation est de 1 ou -1, et le R^2 de 1. La réciproque n'est pas toujours juste.

Validation des hypothèses

Normalité des résidus: test ou méthode graphique

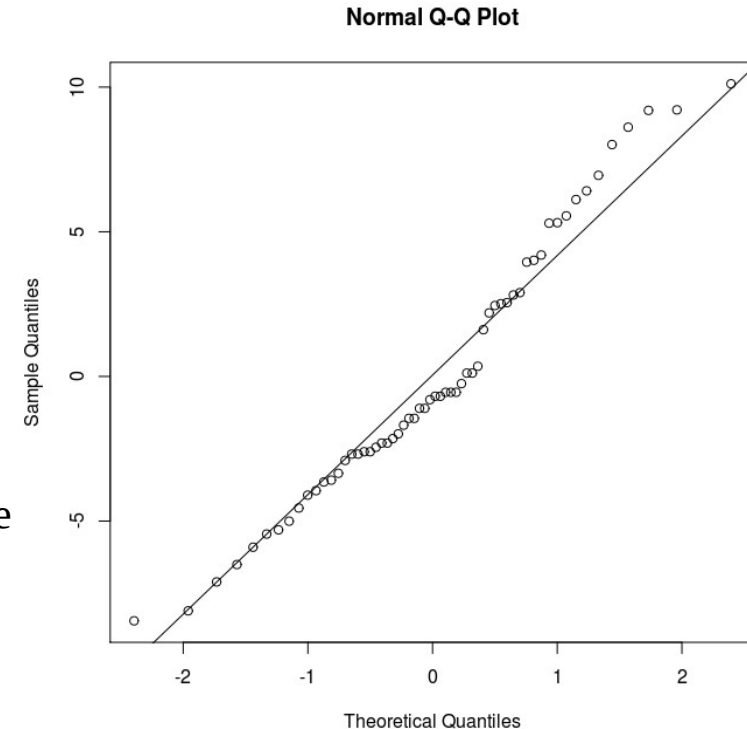
```
shapiro.wilk(residuals(lm1))
```

Shapiro-Wilk normality test

```
data: residuals(lm1)
```

```
W = 0.9647, p-value = 0.07976
```

Avec 5% de chances de se tromper, on ne peut pas rejeter l'hypothèse de normalité des résidus.



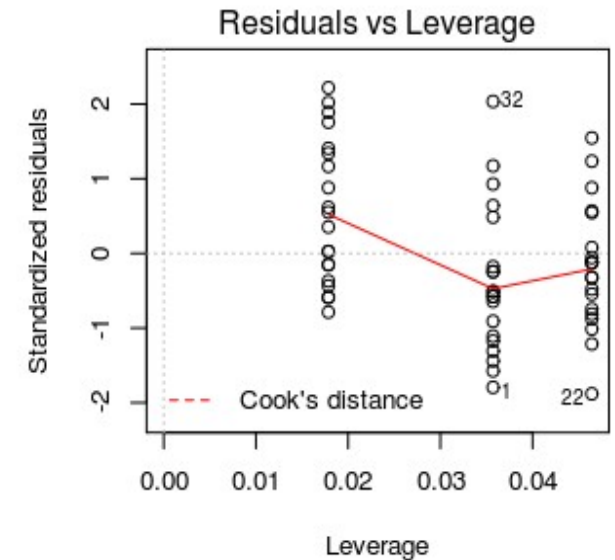
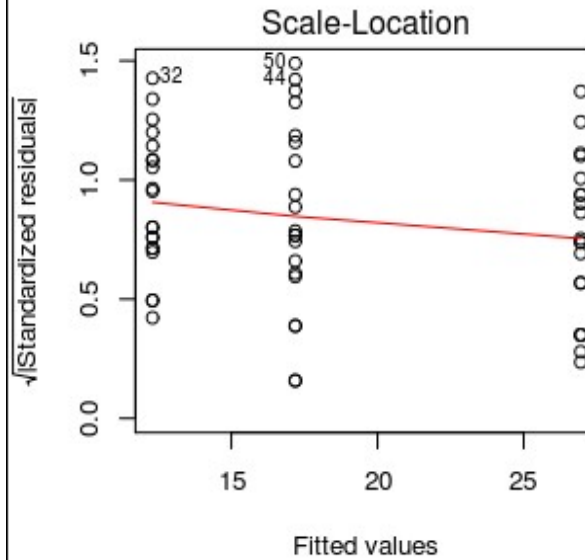
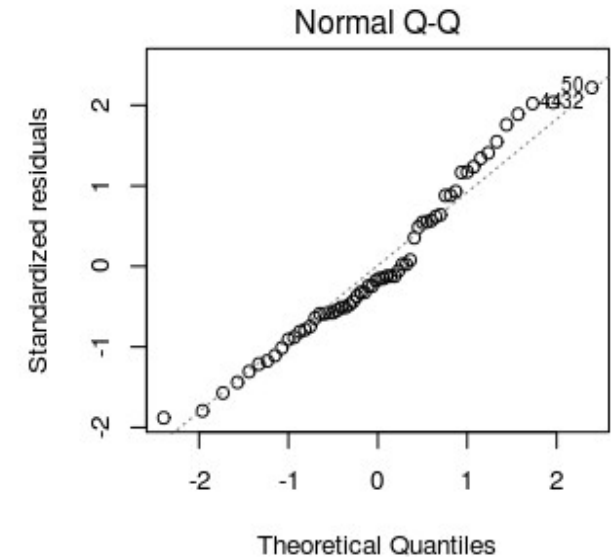
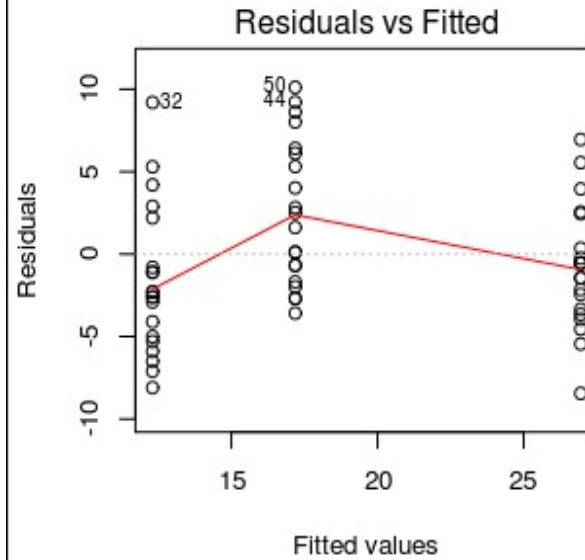
```
qqnorm(residuals(lm1)); qqline(residuals(lm1))
```

(On classe par ordre croissant les n valeurs de longueur $y(1) < y(2) < \dots < y(n)$ où $y(i)$ représente le quantile observé de rang $i/(n+1)$. $z_i = F^{-1}(i/(n+1))$ (Fonction de répartition gaussienne) représente le quantile d'ordre $i/(n+1)$ de la distribution théorique observée. Si les n observations proviennent bien d'une loi normale, alors le tracé des quantiles observés en fonction des quantiles théoriques doit donner une droite.)

Validation des hypothèses

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

```
plot(lm1)
```



Validation des hypothèses

HG:

- Les résidus sont indépendants des x ($\beta_0 + \beta_1 x_i$ proportionnelles aux x , donc si les résidus sont indépendants des valeurs prédites, alors ils sont également indépendants des x);
- les valeurs des résidus tournent autour de 0;
- le modèle linéaire est vrai; si la relation n'avait pas été linéaire, alors les résidus n'auraient pas été indépendants des valeurs prédites.

Si on fitte $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, alors que le vrai modèle est $y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \varepsilon_i$, on aurait eu $\varepsilon_i = \beta_2 x_i^2$ (modèle quadratique).

HD:

Normalité des résidus

BG:

Si on observe une relation entre „Standardized residuals“ et les valeurs prédites, la variance des résidus n'est pas constante.

BD:

Influence des points aberrants

Validation des hypothèses

`plot(lm1, 4)`

$$\bar{\varepsilon}_i = y_i - \bar{y}_i \text{ or } \bar{Y} = X \bar{\beta} = HY \text{ car } \bar{\beta} = (X'X)^{-1} X'Y$$

donc $\bar{\varepsilon}_i = y_i - h y_i$ avec $\text{var}(\bar{\varepsilon}_i) = (\text{Identité} - h) \sigma^2$

Résidus standardisés: $t_i = \bar{\varepsilon}_i / \sigma \sqrt{(1 - h_i)}$ si les résidus sont de variance constante

Résidus studentisés: $t_i^* = \bar{\varepsilon}_i / \sigma(i) \sqrt{(1 - h_i)}$

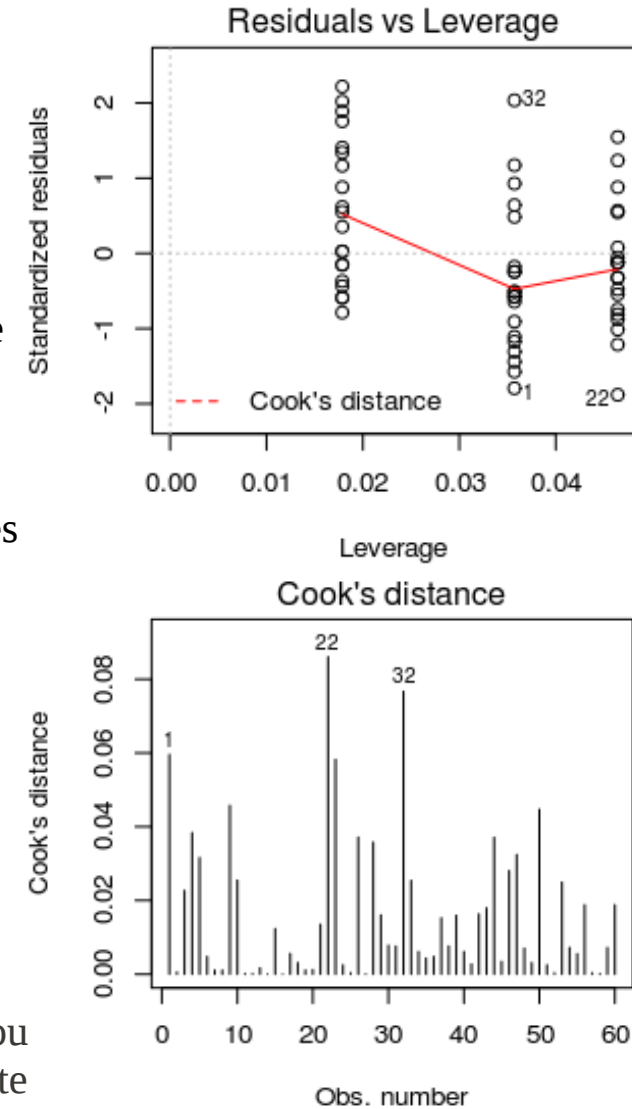
Distance de Cook = influence de la $j^{\text{ème}}$ observation sur l'ensemble des paramètres

$$D_{\text{Cook } j} = \frac{\sum_i (\bar{y}_i - \bar{y}_i(j))^2}{\sigma^2 p} = \frac{1}{p} \frac{h_j}{1 - h_j} t_j^*$$

$< F(p, n-p) (0.1)$, préoccupant si $> F(p, n-p) (0.5)$

Point levier: la valeur x est éloignée de la moyenne générale \bar{x} et a un fort poids dans l'estimation de la pente

Point aberrant: échec du modèle ou erreur de mesure pour x importante

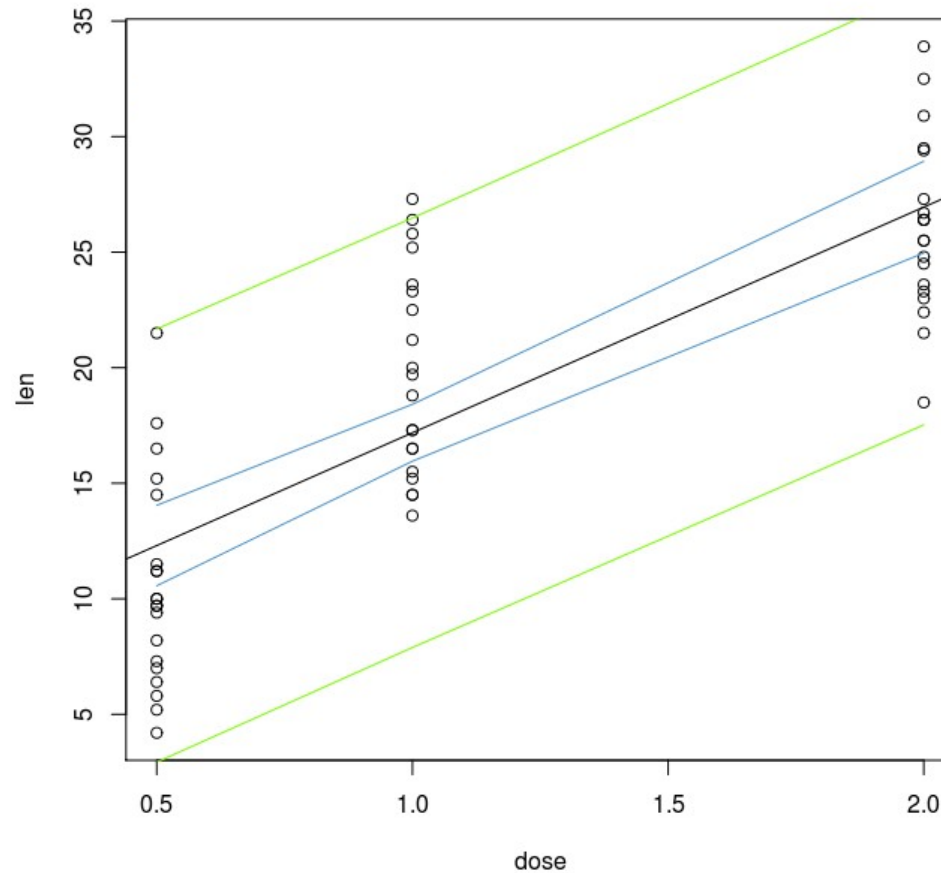


IC des valeurs prédites

→ confidence: IC des valeurs prédites par le modèle pour les valeurs de x des données initiales.

→ predict: IC de la valeur prédite par le modèle pour une nouvelle valeur de x (non présente dans le jeu de données ayant permis l'estimation des paramètres du modèle).

L'IC de type predict est plus large que celui de type confidence.



Construction du modèle 2: variable qualitative

model with treatment covariable

```
lm2 <- lm(len ~ supp, data = ToothGrowth)
```

contrasts(ToothGrowth\$supp) *#matrice de contraste pour variable qualitative*: autant de lignes que de niveaux de la variable qualitative, et autant de colonnes - 1. Chaque ligne correspond aux valeurs prédites par le modèle/niveau (dans l'ordre donné par levels) et chaque colonne correspond à l'un des paramètres - β_0 .

VC

OJ 0 *#réf quand contraste de type traitement par défaut dans R*

VC 1

$$\mu_{OJ} = \beta_0 + 0 \beta_1$$

$$\mu_{VC} = \beta_0 + \beta_1$$

```
supp2 <- ToothGrowth$supp; contrasts(supp2) <- contr.sum
```

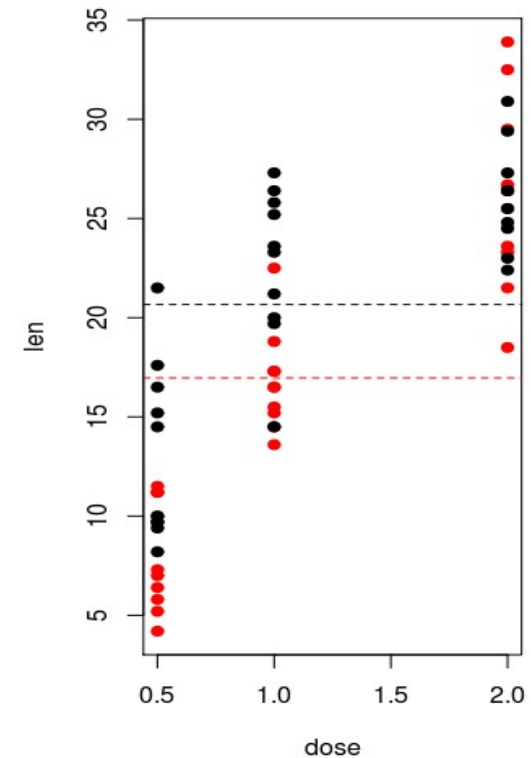
```
contrasts(supp2)
```

OJ 1

VC -1

$$\mu_{OJ} = \beta_0 + \beta_1$$

$$\mu_{VC} = \beta_0 - \beta_1 \rightarrow \beta_0 = (\mu_{OJ} + \mu_{VC}) / 2$$



Ce contraste met en référence la moyenne générale des mesures sur l'ensemble des traitements, puis effectue les comparaisons „moyenne par traitement par rapport à la moyenne générale des traitements“.

Régression linéaire: variable qualitative (contr.treatment)

model with treatment covariable

```
lm2 <- lm(len ~ supp, data = ToothGrowth)
```

```
summary(lm2)
```

```
anova(lm2)
```

Analysis of Variance Table

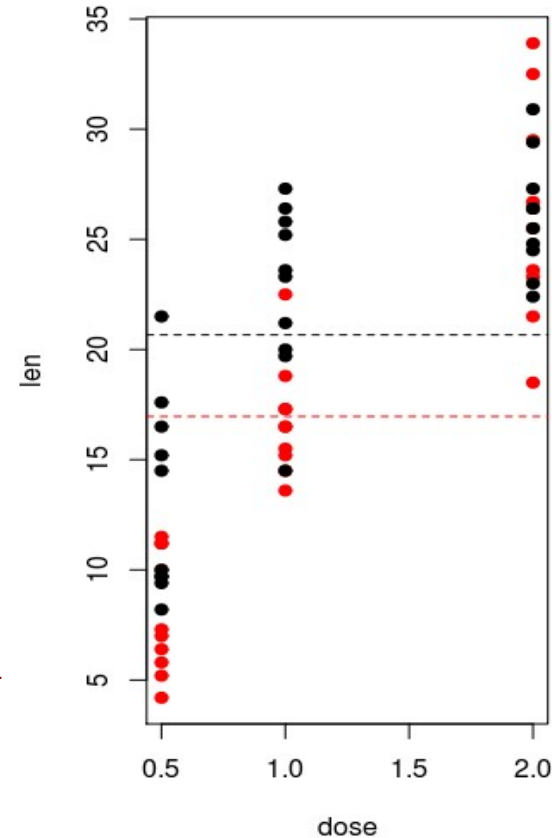
Response: len

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
supp	1	205.4	205.35	3.6683	0.06039 .

Residuals **58** 3246.9 55.98

Signif. Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

→ **t = 1.9153**



Régression linéaire: variable qualitative (contr.sum)

model with treatment covariable

```
lm2.bis <- lm(ToothGrowth$len ~ supp2)
```

```
summary(lm2.bis)
```

anova(lm2) is the same as *anova(lm2.bis)*

Call:

```
lm(formula = len ~ supp, data = ToothGrowth)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
--	----------	------------	---------	----------

(Intercept)	20.663	1.366	15.127	<2e-16 ***
suppVC	-3.700	1.932	-1.915	0.0604 .

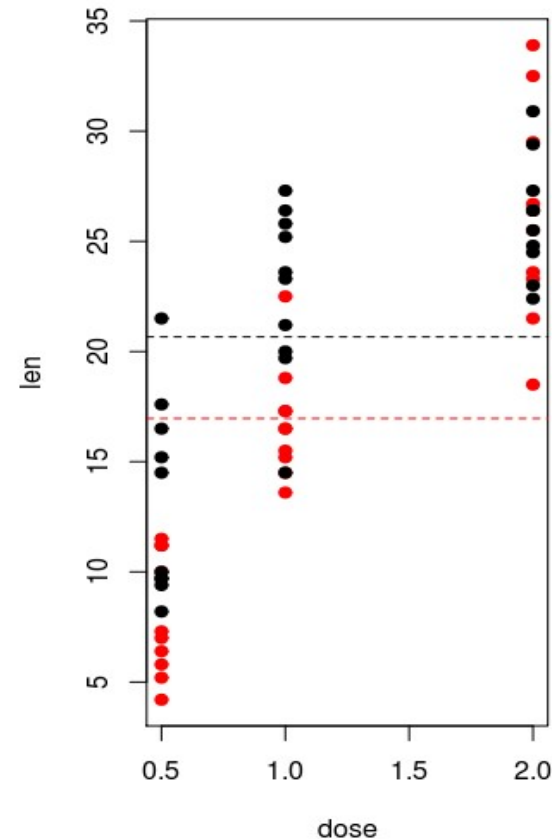
Call:

```
lm(formula = ToothGrowth$len ~ supp2)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
--	----------	------------	---------	----------

(Intercept)	18.8133	0.9659	19.477	<2e-16 ***
supp21	1.8500	0.9659	1.915	0.0604 .



Construction du modèle 3: covariance analysis

model with dose and supplement covariables

`ToothGrowth$dose.center = ToothGrowth$dose - mean(ToothGrowth$dose)`

`lm3 <- lm(len ~ dose.center + supp, data = ToothGrowth) #supp doit être en factor`

$$y_i = \beta_0 + \beta_1 \text{dose.center}_i + \beta_2 \mathbb{I}[\text{supp} == \text{„VC“}]$$

- β_0 est la moyenne des longueurs pour traitement OJ à dose moyenne

- Valeur prédite pour une dose x de traitement OJ ou VC:

$$y_{OJ} = \beta_0 + \beta_1 x_i$$

$$y_{VC} = \beta_0 + \beta_1 x_i + \beta_2$$

β_2 correspond à l'augmentation (>0) /diminution (<0) de longueur pour une même

dose en passant de OJ à VC. *Quelle que soit la dose, l'effet β_2 du traitement est le même.*

- Valeur prédite pour des doses x et x+1 de traitement OJ et VC:

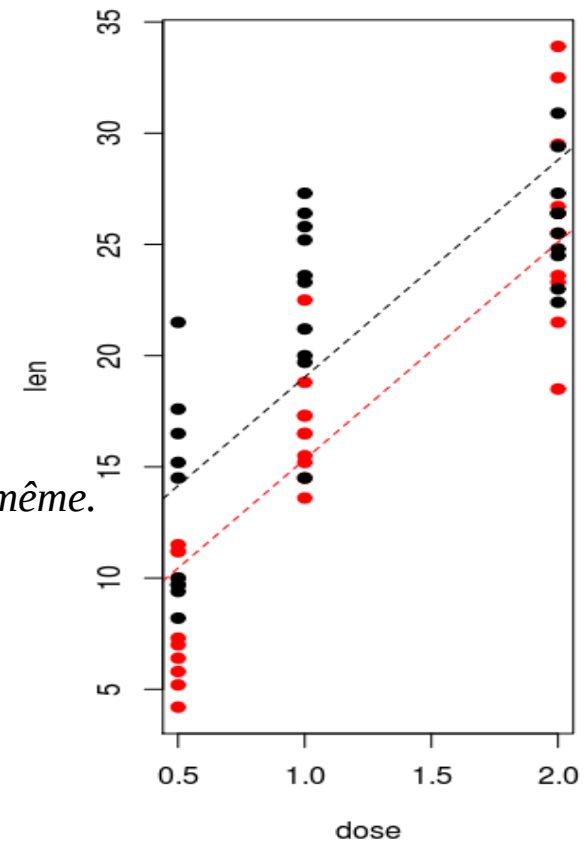
$$y(x)_{OJ} = \beta_0 + \beta_1 x_i$$

$$y(x+1)_{OJ} = \beta_0 + \beta_1 (x_i + 1)$$

$$y(x)_{VC} = \beta_0 + \beta_1 x_i + \beta_2$$

$$y(x+1)_{VC} = \beta_0 + \beta_1 (x_i + 1) + \beta_2$$

β_1 correspond à l'augmentation (>0) /diminution (<0) de longueur pour une augmentation d'une unité de dose pour un même traitement OJ. *L'effet de la dose est le même quel que soit le traitement.*



Covariance analysis

```
summary(lm3)
```

Call:

```
lm(formula = len ~ dose.center + supp, data = ToothGrowth)
```

Residuals:

```
   Min     1Q  Median     3Q    Max
-6.600 -3.700  0.373  2.116  8.800
```

Coefficients:

```
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 20.6633    0.7733  26.721 < 2e-16 ***
dose.center  9.7636    0.8768  11.135 6.31e-16 ***
suppVC      -3.7000    1.0936  -3.383  0.0013 **
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 4.236 on 57 degrees of freedom
Multiple R-squared:  0.7038,    Adjusted R-squared:  0.6934
F-statistic: 67.72 on 2 and 57 DF,  p-value: 8.716e-16
```

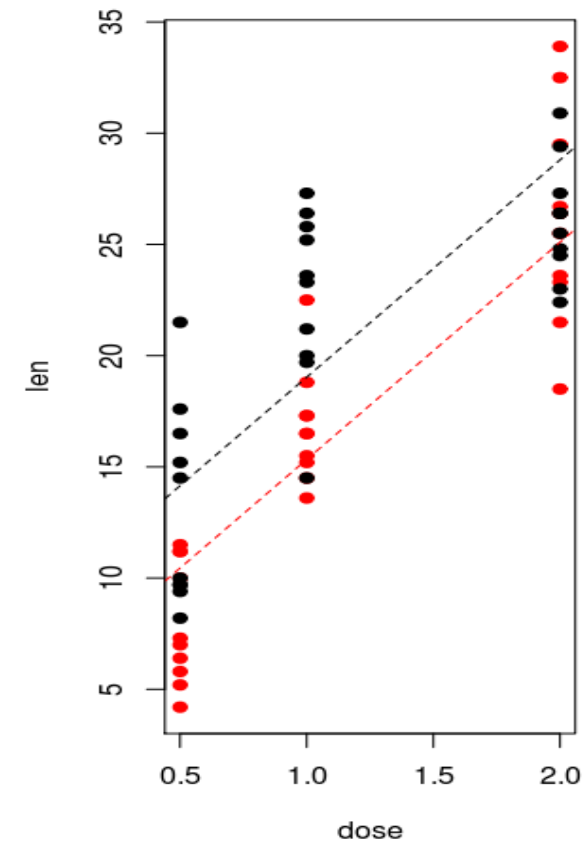
```
anova(lm3)
```

Analysis of Variance Table

Response: len

```
            Df Sum Sq Mean Sq F value    Pr(>F)
dose.center  1 2224.30 2224.30 123.989 6.314e-16 ***
supp         1  205.35  205.35  11.447  0.001301 **
Residuals   57 1022.56   17.94
```

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



Test des modèles emboîtés

`anova(lm1, lm2)`

Analysis of Variance Table

Model 1: len ~ dose.center

Model 2: len ~ supp

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	58	1227.9				
2	58	3246.9	0	-2019		

`anova(lm2, lm3)`

Analysis of Variance Table

Model 1: len ~ supp

Model 2: len ~ dose.center + supp

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	58	3246.9				
2	57	1022.6	1	2224.3	123.99	6.314e-16 ***

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

```
lm1 <- lm(len ~ dose.center, data = ToothGrowth)
lm2 <- lm(len ~ supp, data = ToothGrowth)
lm3 <- lm(len ~ dose.center + supp, data = ToothGrowth)
```

Le modèle 2 est emboîté dans le modèle 3; c'est un sous-modèle du modèle 3 (prendre $\beta_{\text{dose.center}} = 0$ dans le modèle 3 pour obtenir le modèle 2).

Lorsque deux modèles sont emboîtés, on peut tester s'ils sont statistiquement différents en comparant leur somme des carrés des écarts résiduels.

$$SCET_2 = SCET_3$$

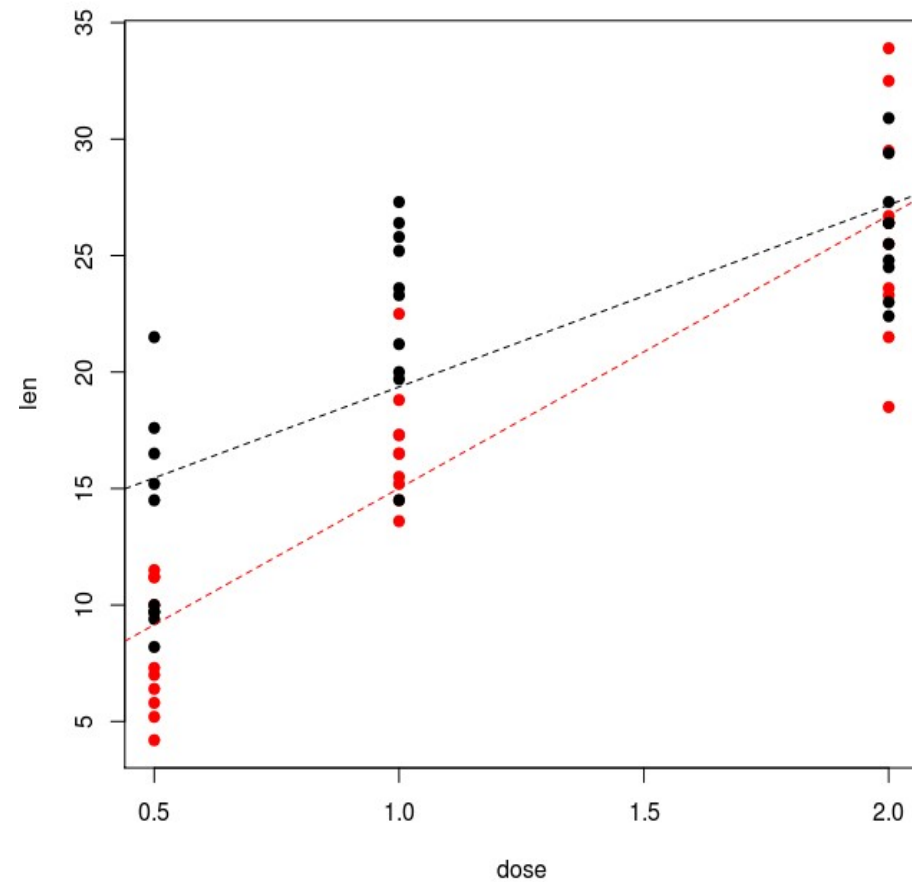
$$F\left(\frac{p_3 - p_2}{3}, \frac{n - p_3}{3}\right) = \frac{\frac{SCER_2 - SCER_3}{3}}{\frac{SCER_3}{n - p_3}}$$

Le modèle le + paramétré explique mieux les data mais n'est pas forcément le mieux statistiquement.

Intéraction

Les pentes semblent différer. Cela voudrait dire que, que l'on soit sous traitement OJ ou VC, une augmentation d'une unité de dose entraînerait une augmentation différente de la longueur. Ainsi, l'effet de la dose différerait selon le traitement.

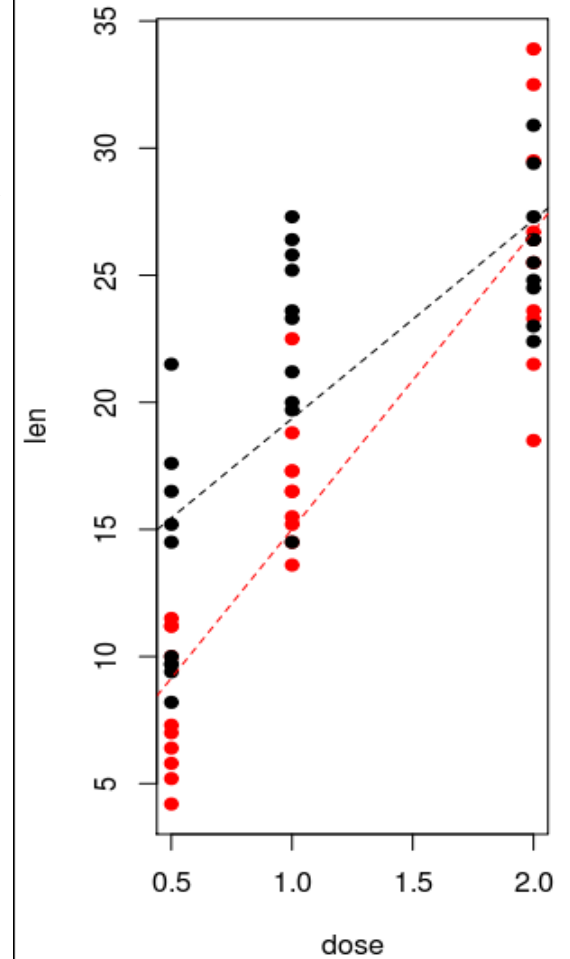
One linear regression per subset of treatment
(code not shown)



Int eraction

model with dose and treatment covariables with interaction

```
lm4 <- lm(len ~ dose.center * supp, data = ToothGrowth)
```

$$y_i = \beta_0 + \beta_1 \text{dose.center}_i + \beta_2 I[\text{supp} == \text{„VC“}] + \beta_3 I[\text{supp} == \text{„VC“}] \text{dose.center}_i$$


Int eraction

`summary(lm4)`

Call:

```
lm(formula = len ~ dose.center * supp, data = ToothGrowth)
```

Residuals:

```
    Min     1Q  Median     3Q     Max
-8.2264 -2.8463  0.0504  2.2893  7.9386
```

Coefficients:

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    20.6633     0.7455  27.718 < 2e-16 ***
dose.center      7.8114     1.1954   6.534 2.03e-08 ***
suppVC         -3.7000     1.0543  -3.510 0.000894 ***
dose.center:suppVC 3.9043     1.6906   2.309 0.024631 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 4.083 on 56 degrees of freedom
Multiple R-squared:  0.7296, Adjusted R-squared:  0.7151
F-statistic: 50.36 on 3 and 56 DF, p-value: 6.521e-16
```

`anova(lm4)`

Analysis of Variance Table

Response: len

```
              Df Sum Sq Mean Sq F value    Pr(>F)
dose.center    1 2224.30  2224.30 133.4151 < 2.2e-16 ***
supp           1  205.35   205.35  12.3170 0.0008936 ***
dose.center:supp 1   88.92    88.92   5.3335 0.0246314 *
Residuals     56  933.63   16.67
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

`anova(lm4, lm3)`

Analysis of Variance Table

Model 1: len ~ dose.center * supp

Model 2: len ~ dose.center + supp

```
      Res.Df  RSS Df Sum of Sq    F Pr(>F)
1         56 933.63
2         57 1022.56 -1   -88.92 5.3335 0.02463 *
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```


Validation des hypothèses

verification of hypothesis

```
par(mfrow = c(2, 2), pty = "m")
```

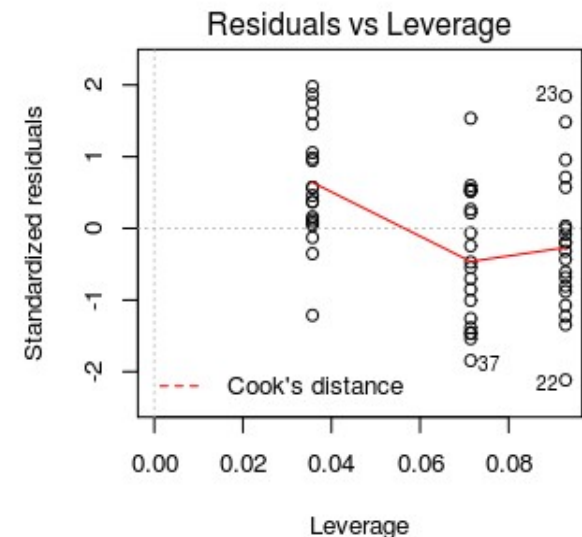
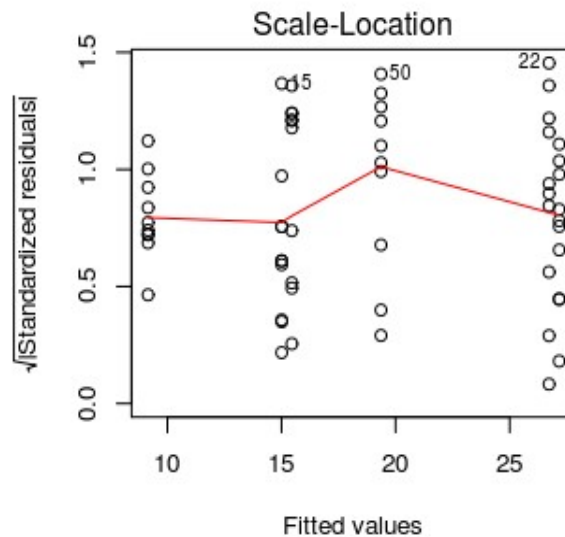
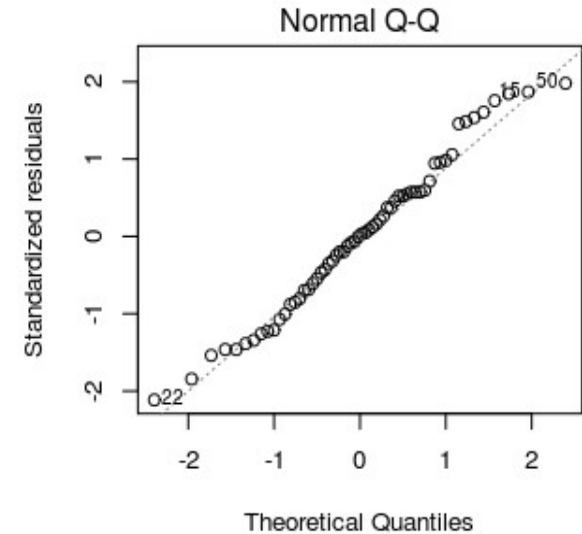
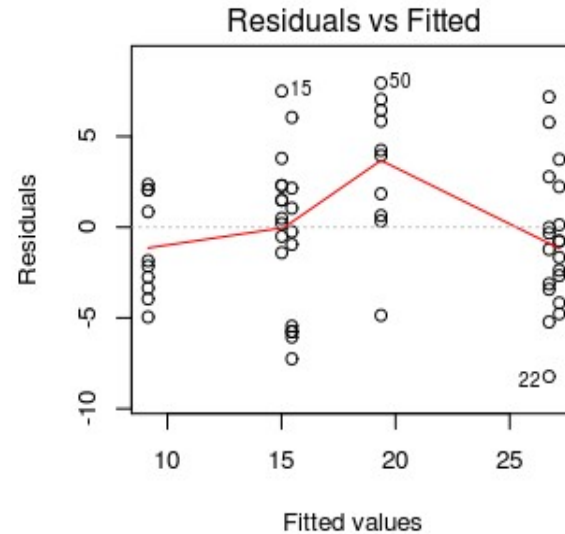
```
plot(lm4)
```

```
shapiro.test(residuals(lm4))
```

Shapiro-Wilk normality test

data: residuals(lm4)

$W = 0.9816$, $p\text{-value} = 0.5008$



Résumé des modèles construits

	p	SCER	R ²
dose	2	1228	64
supp	2	3247	6
supp+dose	3	1023	70
supp*dose	4	934	73

Merci de votre attention!

Mercredi 15 Mars 14h salle 115:

Tests statistiques en biologie par Anissa Guillemin

Soon (avril):

Laurent Modolo propose une session „Tests multiples: mieux comprendre pour mieux contrôler“ (Foodle à venir!)