

Guide d'utilisation du logiciel R pour l'ajustement de modèles linéaires à 2 facteurs qualitatifs fixes ou aléatoires.

Introduction aux modèles linéaires mixtes

Marie Laure Delignette-Muller

26 novembre 2020

Table des matières

1	Introduction	2
1.1	Cadre, objectifs et choix méthodologiques de ce guide	2
1.2	Petit rappel sur le codage des variables ou facteurs	2
2	Modèle croisé fixe avec répétitions	2
2.1	Exemple	3
2.2	Importation et représentation des données	3
2.3	Estimation des effets fixes	5
2.4	Réalisation du tableau d'analyse de variance dans le cas d'un plan équilibré	6
2.5	Vérification des conditions d'utilisation du modèle	6
3	Modèle croisé aléatoire avec répétitions	8
3.1	Exemple	8
3.2	Importation et représentation des données	8
3.3	Estimation des composantes de la variance	10
3.4	Vérification des conditions d'utilisation du modèle	11
4	Modèle croisé mixte avec répétitions	12
4.1	Exemple	12
4.2	Importation et représentation des données	13
4.3	Estimation des effets fixes et des composantes de la variance	14
4.4	Réalisation du tableau d'analyse de variance dans le cas d'un plan équilibré	15
4.5	Vérification des conditions d'utilisation du modèle	15
5	Modèle croisé mixte sans répétition	17
5.1	Exemple de modèle mixte sans répétition	17
5.2	Importation et représentation des données	17
5.3	Estimation des effets fixes et des composantes de la variance	18
5.4	Réalisation du tableau d'analyse de variance dans le cas d'un plan équilibré	18
5.5	Vérification des conditions d'utilisation du modèle	19
6	Modèle croisé aléatoire sans répétition	20
7	Modèle hiérarchisé aléatoire	21
7.1	Exemple de modèle hiérarchisé aléatoire	21
7.2	Importation et représentation des données	21
7.3	Estimation des composantes de la variance	22
7.4	Vérification des conditions d'utilisation du modèle	23
8	Modèle hiérarchisé mixte	25
8.1	Exemple de modèle hiérarchisé mixte	25
8.2	Importation et représentation des données	25
8.3	Estimation des composantes de la variance	27
8.4	Vérification des conditions d'utilisation du modèle	27
8.5	Réalisation du tableau d'analyse de variance dans le cas d'un plan équilibré	28

1 Introduction

1.1 Cadre, objectifs et choix méthodologiques de ce guide

Ce guide a pour objectif de présenter les principales fonctions utiles pour ajuster et étudier des modèles linéaires mixtes. Il présente l'analyse de divers exemples correspondant aux divers cas que l'on peut rencontrer avec seulement deux facteurs qualitatifs (y compris l'exemple plus classique de modèle fixe déjà traité dans le guide sur le modèle linéaire). L'analyse d'exemples à plus de deux facteurs est bien entendu plus délicate, mais suit les mêmes principes que ceux présentés dans les exemples à deux facteurs. L'analyse de variance classique, associant à chaque facteur une valeur de p (p -value) calculés à partir de la décomposition de la somme des carrés des écarts, est délicate dans le cas de deux facteurs, encore plus délicate dans le cas de plus de 2 facteurs, et impossible dans le cas d'un dispositif déséquilibré dès qu'on a plus d'un facteur. J'ai donc fait le choix de présenter en priorité la méthode basée sur l'ajustement d'un modèle linéaire mixte, utilisable même si le dispositif est déséquilibré et par ailleurs de plus en plus utilisée dans la littérature scientifique. Les deux principaux packages permettant de réaliser ces ajustements sont les packages `nlme` et `lme4`. J'ai fait le choix de présenter les fonctionnalités du package le plus récent, `lme4` du fait de son cadre d'utilisation plus large et de sa manipulation qui me semble plus facile et de nature à limiter les erreurs d'implémentation possibles avec le package plus ancien `nlme`.

Le test des effets aléatoires des modèles présentés n'est pas abordé dans ce document. Outre le fait qu'il n'est pas trivial à réaliser, il serait inapproprié de choisir d'incorporer ou non des effets aléatoires dans un modèle sur la base d'un tel test. Le test des effets fixes, potentiellement plus intéressant, pose encore des problèmes méthodologiques actuellement non résolus par les statisticiens hormis dans le cas de plans d'expérience équilibrés pour lesquels une analyse de variance appropriée peut être réalisée. Il n'est donc pas implémenté dans le package `lmer` qui offre des outils valables que le plan d'expérience soit équilibré ou non. Dans le cas général, on se basera sur les intervalles de confiance associés aux effets estimés pour juger de leur significativité. Des explications plus détaillées sur ce point par les auteurs du package `lme4` peuvent être trouvées dans un article publié dans *Journal of Statistical Software* en 2015 et disponible au lien suivant : <https://www.jstatsoft.org/article/view/v067i01>.

1.2 Petit rappel sur le codage des variables ou facteurs

Par défaut, lorsque qu'une variable est codée numériquement, elle est considérée comme quantitative par R. Pour qu'un vecteur codé numériquement (par des 0 et de 1 par exemple) soit considéré comme un facteur qualitatif, il faut utiliser la fonction `as.factor` devant le nom de ce vecteur lorsqu'on l'utilise. Mais le mieux est encore de s'efforcer de coder de façon non numérique les variables ou facteurs qualitatifs. Dans tous les cas il est prudent de vérifier la structure du jeu de données et le codage de chaque variable avant toute analyse en utilisant la fonction `str` (cf. exemple suivant).

Par défaut **R** classe les modalités d'un facteur par ordre alphabétique. Si l'on veut changer cet ordre, on peut utiliser un code de ce type :

```
# créons une variable qualitative (facteur) pour l'exemple
sexe <- as.factor(c("G", "F", "F", "G", "F", "F", "G", "G", "F", "F"))
# regardons l'ordre de ses modalités par défaut
levels(sexe)

## [1] "F" "G"

# changeons cet ordre
sexe <- factor(sexe, levels = c("G", "F"))
sexe

## [1] G F F G F F G G F F
## Levels: G F
```

On peut aussi changer les noms des modalités de la façon suivante :

```
levels(sexe) <- c("garçon", "fille")
sexe

## [1] garçon fille fille garçon fille fille garçon garçon fille fille
## Levels: garçon fille
```

2 Modèle croisé fixe avec répétitions

Ce modèle est le seul de ce document à ne pas faire partie de la famille des modèles mixtes et fait déjà l'objet d'un paragraphe dans le guide concernant les modèles linéaires. Nous avons néanmoins choisi de le reprendre ici afin de

traiter dans un même document tous les cas que l'on peut rencontrer avec deux facteurs qualitatifs.

2.1 Exemple

On veut étudier l'action de l'adrénaline et de la pitressine sur l'équilibre hydrique du têtard du Crapaud sud-américain *Bufo arenarum*. Vingt animaux en tout ont été utilisés, soit 4 groupes de 5. Les données sont dans le fichier `ADREPITRE.txt`.

2.2 Importation et représentation des données

```
d1 <- read.table("ADREPITRE.txt", header = TRUE,
                 stringsAsFactors = TRUE)
```

On peut visualiser les données brutes en affichant la totalité du tableau ou en utilisant la fonction `head` qui affiche les premières lignes ou encore avec la fonction `str` qui affiche la structure du jeu de données.

```
d1
##      adrenaline pitressine reponse
## 1          non          non    -25
## 2          non          non     -3
## 3          non          non    -12
## 4          non          non     37
## 5          non          non      3
## 6          oui          non    143
## 7          oui          non     41
## 8          oui          non     58
## 9          oui          non     80
## 10         oui          non     79
## 11         non          oui    132
## 12         non          oui    138
## 13         non          oui    112
## 14         non          oui    104
## 15         non          oui    112
## 16         oui          oui    211
## 17         oui          oui    263
## 18         oui          oui    258
## 19         oui          oui    229
## 20         oui          oui    213

head(d1)
##      adrenaline pitressine reponse
## 1          non          non    -25
## 2          non          non     -3
## 3          non          non    -12
## 4          non          non     37
## 5          non          non      3
## 6          oui          non    143

str(d1)
## 'data.frame': 20 obs. of  3 variables:
## $ adrenaline: Factor w/ 2 levels "non","oui": 1 1 1 1 1 2 2 2 2 2 ...
## $ pitressine: Factor w/ 2 levels "non","oui": 1 1 1 1 1 1 1 1 1 1 ...
## $ reponse   : int  -25 -3 -12 37 3 143 41 58 80 79 ...
```

Il peut être utile de visualiser le plan d'expérience pour voir le nombre de répétitions et vérifier qu'il n'y a pas d'erreur de codage.

```
xtabs(~ adrenaline + pitressine, data = d1)
##           pitressine
## adrenaline non oui
```

```
##      non  5  5
##      oui  5  5
```

Si besoin les moyennes et écarts type peuvent être calculés à partir de la fonction `tapply`.

```
tapply(d1$reponse, d1$adrenaline:d1$pitressine, mean)

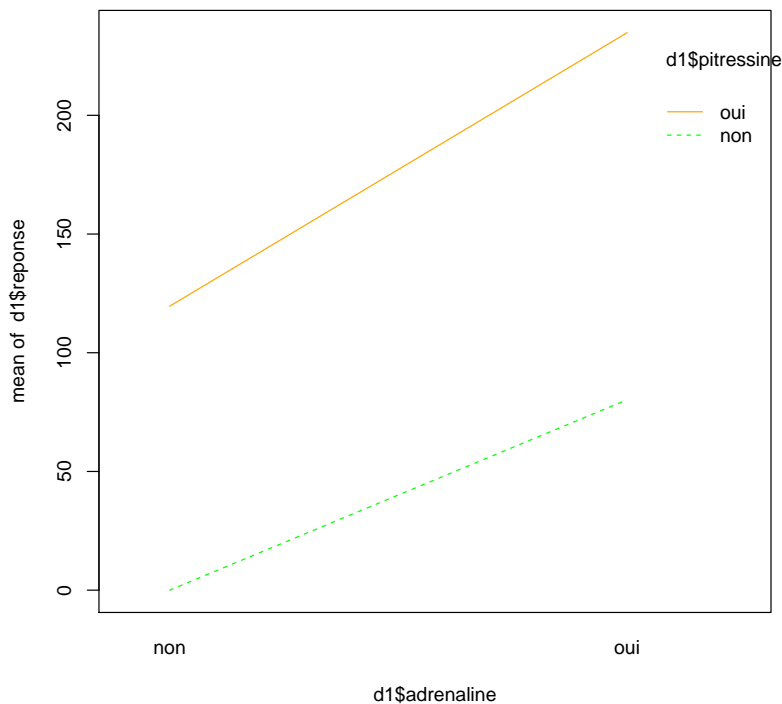
## non:non non:oui oui:non oui:oui
##      0.0  119.6   80.2  234.8

tapply(d1$reponse, d1$adrenaline:d1$pitressine, sd)

## non:non non:oui oui:non oui:oui
##      23.2   14.6   38.6   24.5
```

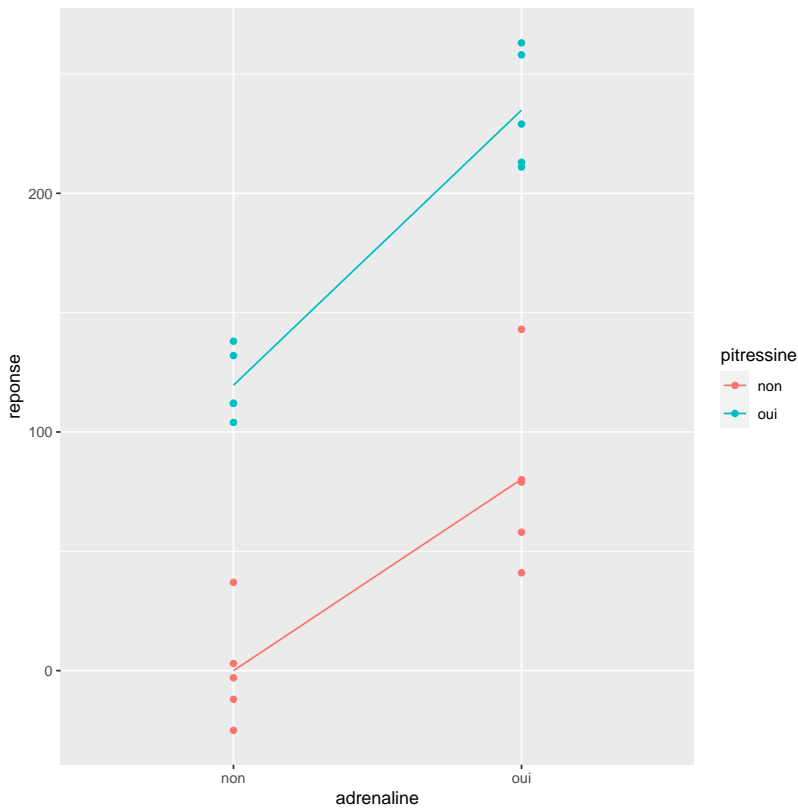
Divers graphes peuvent ensuite être réalisés pour visualiser les données. Le graphe d'interaction, représentant les moyennes pour chaque combinaison des 2 facteurs, permet en particulier de juger de l'existence ou non d'une interaction entre les 2 facteurs, et d'interpréter une interaction qui s'avèrerait significative lors des tests.

```
interaction.plot(d1$adrenaline, d1$pitressine, d1$reponse, col = c("green", "orange"))
```



On peut utiliser la fonction `qplot` du package `ggplot2` pour représenter toutes les données en fonction de chacun des 2 facteurs en y ajoutant même éventuellement les valeurs moyennes. Voici un exemple de ce que l'on peut coder facilement.

```
require(ggplot2)
qplot(adrenaline, reponse, data = d1, colour = pitressine) +
  stat_summary(fun.y = mean, geom = "line", aes(group = pitressine))
```



2.3 Estimation des effets fixes

Ce modèle ne comporte que des effets fixes qui sont estimés classiquement à l'aide de la fonction `lm`. Rappelons que par défaut dans R les effets sont estimés avec la contrainte de type "cellule de référence" (`contrasts = "contr.treatment"`), c'est-à-dire que la constante (`intercept`) correspond à la moyenne prédite pour le groupe correspondant à la première modalité de chacun des facteurs.

```
(m <- lm(reponse ~ adrenaline + pitressine + adrenaline:pitressine, data = d1))

##
## Call:
## lm(formula = reponse ~ adrenaline + pitressine + adrenaline:pitressine,
##     data = d1)
##
## Coefficients:
##           (Intercept)                adrenalineoui
##                   0.0                   80.2
##      pitressineoui adrenalineoui:pitressineoui
##                   119.6                   35.0

# les parenthèses autour de l'expression permettent la visualisation de l'objet créé
```

On peut obtenir un intervalle de confiance à 95% sur chaque effet estimé à l'aide de la fonction `confint`.

```
confint(m)

##           2.5 % 97.5 %
## (Intercept)      -25.3  25.3
## adrenalineoui    44.4 116.0
## pitressineoui    83.8 155.4
## adrenalineoui:pitressineoui -15.6  85.6
```

Une autre écriture équivalente du modèle est la suivante :

```
(m <- lm(reponse ~ adrenaline * pitressine, data = d1))

##
## Call:
```

```
## lm(formula = reponse ~ adrenaline * pitressine, data = d1)
##
## Coefficients:
##           (Intercept)                adrenalineoui
##                0.0                  80.2
##           pitressineoui adrenalineoui:pitressineoui
##                119.6                  35.0
```

Dans certains cas on peut souhaiter ajuster un modèle sans interaction (hypothèse forte à ne faire qu'avec prudence) de la façon suivante.

```
(msansint <- lm(reponse ~ adrenaline + pitressine, data = d1))
##
## Call:
## lm(formula = reponse ~ adrenaline + pitressine, data = d1)
##
## Coefficients:
## (Intercept) adrenalineoui pitressineoui
##          -8.75          97.70          137.10
```

2.4 Réalisation du tableau d'analyse de variance dans le cas d'un plan équilibré

Le tableau d'analyse de variance ne peut en théorie être réalisé que dans le cas d'un plan d'expérience équilibré. Dans un cas déséquilibré, le résultat donné par la fonction `anova` appliqué au modèle dépendra de l'ordre dans lequel sont rentrés les facteurs dans la formule du modèle.

```
anova(m)
## Analysis of Variance Table
##
## Response: reponse
##           Df Sum Sq Mean Sq F value Pr(>F)
## adrenaline    1  47726   47726   67.04 4.1e-07
## pitressine     1  93982   93982  132.01 3.9e-09
## adrenaline:pitressine 1   1531    1531    2.15  0.16
## Residuals    16  11391     712
```

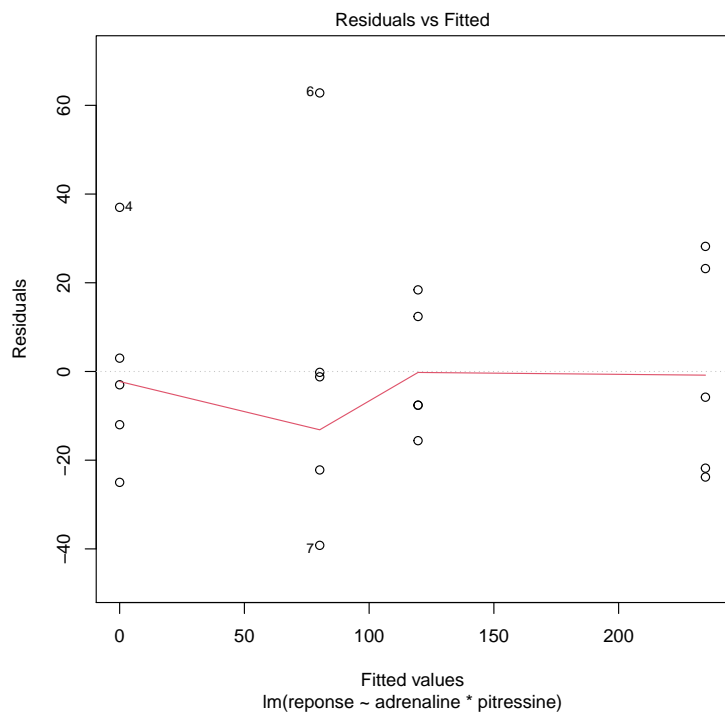
Dans le cas du modèle croisé fixe, il est impératif d'interpréter la ligne interaction en premier, et de n'interpréter les effets principaux des 2 facteurs que lorsque l'interaction est non significative.

2.5 Vérification des conditions d'utilisation du modèle

Comme pour tout modèle linéaire gaussien, il est prudent de vérifier l'hypothèse d'indépendance des résidus, de normalité de leur distribution et d'homoscédasticité à l'aide de graphes adaptés.

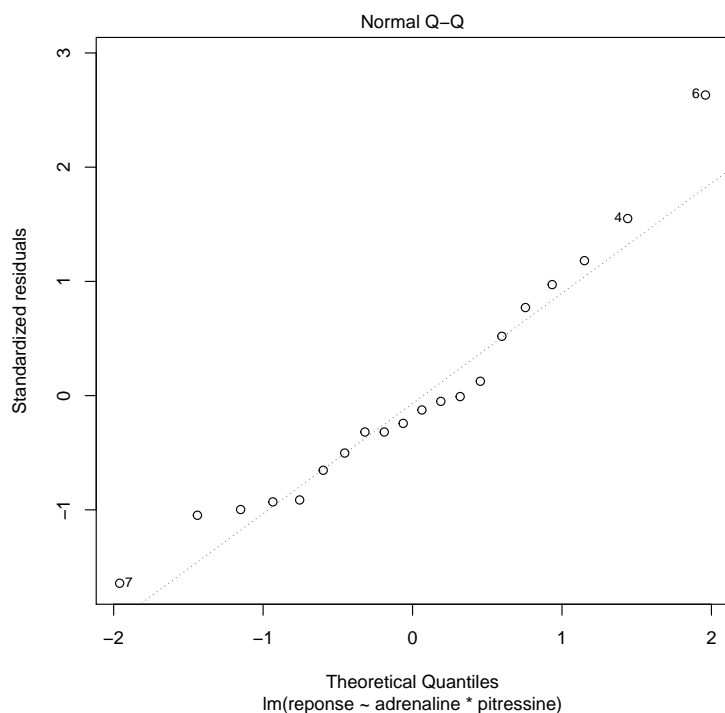
— **Graphe des résidus en fonction des valeurs prédites**

```
plot(m, 1)
```



— Graphe Quantile-Quantile des résidus

```
plot(m,2)
```



L'utilisation de tests peut compléter cette analyse des résidus **tout en gardant en tête que ces tests ne suffisent pas à vérifier les conditions d'utilisation du modèle linéaire**. le test de Shapiro-Wilk teste la normalité de la distribution globale des résidus.

```
shapiro.test(residuals(m))

##
## Shapiro-Wilk normality test
##
## data: residuals(m)
## W = 0.9, p-value = 0.3
```

Le test de Levene implémenté ci-dessous pour un plan d'expérience équilibré (cf. problème associé à l'analyse

de variance dans les cas déséquilibrés commenté dans la section précédente) teste l'homoscédasticité des résidus (hypothèse d'égalité des variances entre groupes). Dès lors que l'analyse de variance réalisée sur la valeur absolue des résidus du modèle montre un effet significatif, l'hypothèse d'homogénéité des variances est remise en cause.

```
anova(lm(abs(residuals(m)) ~ d1$adrenaline * d1$pitressine))

## Analysis of Variance Table
##
## Response: abs(residuals(m))
##
##           Df Sum Sq Mean Sq F value Pr(>F)
## d1$adrenaline      1     377      377   1.48  0.24
## d1$pitressine      1      85       85   0.33  0.57
## d1$adrenaline:d1$pitressine 1      1        1   0.00  0.95
## Residuals        16    4083      255
```

3 Modèle croisé aléatoire avec répétitions

3.1 Exemple

Pour quantifier l'effet inhibiteur de la croissance d'une souche d'*Escherichia coli* producteur de shiga-toxines (STEC) par une souche lactique (lactique), on dépose une goutte d'une suspension de la souche lactique sur une gélose préalablementensemencée avec la souche STEC et on mesure, après incubation, un diamètre d'inhibition de la croissance autour de la goutte. Cette expérience a été réalisée avec 364 souches de STEC (A : 36 modalités) et 24 souches de lactique (B : 24 modalités), en trois répétitions (n = 3). On souhaite étudier l'impact des deux facteurs aléatoires "STEC" et "lactique".

3.2 Importation et représentation des données

```
d2 <- read.table("diametres.txt", header = TRUE, dec=".",
                stringsAsFactors = TRUE)
str(d2)

## 'data.frame': 2591 obs. of 5 variables:
## $ lactique : Factor w/ 24 levels "11A1","13A","17v",...: 8 8 8 2 2 2 23 23 23 9 ...
## $ STEC : Factor w/ 36 levels "05117C","1.1.1",...: 27 27 27 27 27 27 27 27 27 27 ...
## $ origine : Factor w/ 12 levels "Bouse","Camembert",...: 4 4 4 4 4 4 4 4 4 4 ...
## $ serogroupe: Factor w/ 5 levels "0103","0111",...: 4 4 4 4 4 4 4 4 4 4 ...
## $ inhibition: num 7 14 9 7 8 8 6 9 7 9 ...
```

Visualisation du plan d'expérience :

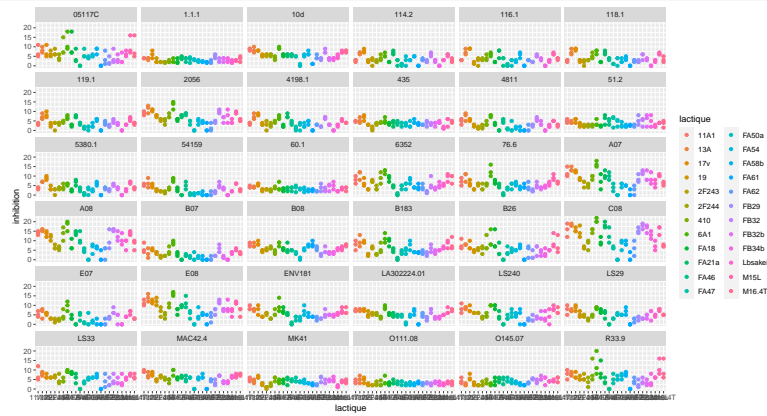
```
xtabs(~ STEC + lactique, data = d2)

##           lactique
## STEC      11A1 13A 17v 19 2F243 2F244 410 6A1 FA18 FA21a FA46 FA47 FA50a
## 05117C      3  3  3  3      3      3  3  3  3  3  3  3  3
## 1.1.1       3  3  3  3      3      3  3  3  3  3  3  3  3
## 10d         3  3  3  3      3      3  3  3  3  3  3  3  3
## 114.2       3  3  3  3      3      3  3  3  3  3  3  3  3
## 116.1       3  3  3  3      3      3  3  3  3  3  3  3  3
## 118.1       3  3  3  3      3      3  3  3  3  3  3  3  3
## 119.1       3  3  3  3      3      3  3  3  3  3  3  3  3
## 2056        3  3  3  3      3      3  3  3  3  3  3  3  3
## 4198.1      3  3  3  3      3      3  3  3  3  3  3  3  3
## 435         3  3  3  3      3      3  3  3  3  3  3  3  3
## 4811        3  3  3  3      3      3  3  3  3  3  3  3  3
## 51.2        3  3  3  3      3      3  3  3  3  3  3  3  3
## 5380.1      3  3  3  3      3      3  3  3  3  3  3  3  3
## 54159       3  3  3  3      3      3  3  3  3  3  3  3  3
## 60.1        3  3  3  3      3      3  3  3  3  3  3  3  3
## 6352        3  3  3  3      3      3  3  3  3  3  3  3  3
## 76.6        3  3  3  3      3      3  3  3  3  3  3  3  3
```


##	A07	3	3	3	3	3	3	3	3	3	3	3	3
##	A08	3	3	3	3	3	3	3	3	3	3	3	3
##	B07	3	3	3	3	3	3	3	3	3	3	3	3
##	B08	3	3	3	3	3	3	3	3	3	3	3	3
##	B183	3	3	3	3	3	3	3	3	3	3	3	3
##	B26	3	3	3	3	3	3	3	3	3	3	3	3
##	C08	3	3	3	3	3	3	3	3	3	3	3	3
##	E07	3	3	3	3	3	3	3	3	3	3	3	3
##	E08	3	3	3	3	3	3	3	3	3	3	3	3
##	ENV181	3	3	3	3	3	3	3	3	3	3	3	3
##	LA302224.01	3	3	3	3	3	3	3	3	3	3	3	3
##	LS240	3	3	3	3	3	3	3	3	3	3	3	3
##	LS29	3	3	3	3	3	3	3	3	3	3	3	3
##	LS33	3	3	3	3	3	3	3	3	3	3	3	3
##	MAC42.4	3	3	3	3	3	3	3	3	3	3	3	3
##	MK41	3	3	3	3	3	3	3	3	3	3	3	3
##	O111.08	3	3	3	3	3	3	3	3	3	3	3	3
##	O145.07	3	3	3	3	3	3	3	3	3	3	3	3
##	R33.9	3	3	3	3	3	3	3	3	3	3	3	3
##		lactique											
##	STEC	FA54	FA58b	FA61	FA62	FB29	FB32	FB32b	FB34b	Lbsakei	M15L	M16.4T	
##	05117C	3	3	3	3	3	3	3	3	3	3	3	3
##	1.1.1	3	3	3	3	3	3	3	3	3	3	3	3
##	10d	3	3	3	3	3	3	3	3	3	3	3	3
##	114.2	3	3	3	3	3	3	3	3	3	3	3	3
##	116.1	3	3	3	3	3	3	3	3	3	3	3	3
##	118.1	3	3	3	3	3	3	3	3	3	3	3	3
##	119.1	3	3	3	3	3	3	3	3	3	3	3	3
##	2056	3	3	3	3	3	3	3	3	3	3	3	3
##	4198.1	3	3	3	3	3	3	3	3	3	3	3	3
##	435	3	3	3	3	3	3	3	3	3	3	3	3
##	4811	3	3	3	3	3	3	3	3	3	3	3	3
##	51.2	3	3	3	3	3	3	3	3	3	3	3	3
##	5380.1	3	3	3	3	3	3	3	3	3	3	3	3
##	54159	3	3	3	3	3	3	3	3	3	3	3	3
##	60.1	3	3	3	3	3	3	3	3	3	3	3	3
##	6352	3	3	3	3	3	3	3	3	3	3	3	3
##	76.6	3	3	3	3	3	3	3	3	3	3	3	3
##	A07	3	3	3	3	3	3	3	3	3	3	3	3
##	A08	3	3	3	3	3	3	3	3	3	3	3	3
##	B07	3	3	3	3	3	3	3	3	3	3	3	3
##	B08	3	3	3	3	3	3	3	3	3	3	3	3
##	B183	3	3	3	3	3	3	3	3	3	3	3	3
##	B26	3	3	3	3	3	3	3	3	3	3	3	3
##	C08	3	3	3	3	3	3	3	3	3	3	3	3
##	E07	3	3	3	3	3	3	3	3	3	3	3	3
##	E08	3	3	3	3	3	3	3	3	3	3	3	3
##	ENV181	3	3	3	3	3	3	3	3	3	3	3	3
##	LA302224.01	3	3	3	3	3	3	3	3	3	3	3	3
##	LS240	3	3	3	3	3	3	3	3	3	3	3	3
##	LS29	3	3	3	3	3	3	3	3	3	3	3	3
##	LS33	3	3	3	3	3	3	3	3	3	3	3	3
##	MAC42.4	3	3	3	3	3	3	3	3	3	3	3	3
##	MK41	3	3	3	3	3	2	3	3	3	3	3	3
##	O111.08	3	3	3	3	3	3	3	3	3	3	3	3
##	O145.07	3	3	3	3	3	3	3	3	3	3	3	3
##	R33.9	3	3	3	3	3	3	3	3	3	3	3	3

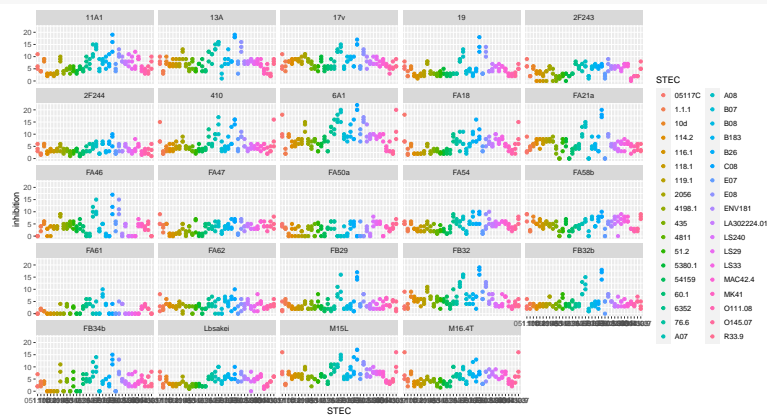
Les données peuvent être visualisées graphiquement en fonction de chaque facteur comme dans l'exemple précédent à l'aide des fonctions `stripchart` ou `qplot` (pensez à charger auparavant le package `ggplot2` pour utiliser la deuxième).
— Pour chaque STEC en fonction des lactiques (en abscisse)

```
qplot(lactique, inhibition, facets = ~ STEC, data = d2, colour = lactique)
```



— Pour chaque lactique en fonction des STEC (en abscisse)

```
qplot(STEC, inhibition, facets = ~ lactique, data = d2, colour = STEC)
```



3.3 Estimation des composantes de la variance

Ce modèle ne comportant que des effets aléatoires (hormis la constante), l'estimation des effets revient à l'estimation des composantes de la variance. On peut les calculer en reprenant Cette estimation peut être réalisée en utilisant la fonction `lmer` du package `lme4`.

```
require(lme4)
(mm <- lmer(inhibition ~ (1|STEC)+ (1|lactique), data = d2))

## Linear mixed model fit by REML ['lmerMod']
## Formula: inhibition ~ (1 | STEC) + (1 | lactique)
## Data: d2
## REML criterion at convergence: 11820
## Random effects:
## Groups Name Std.Dev.
## STEC (Intercept) 1.95
## lactique (Intercept) 1.83
## Residual 2.26
## Number of obs: 2591, groups: STEC, 36; lactique, 24
## Fixed Effects:
## (Intercept)
## 5.03
```

On peut obtenir un intervalle de confiance à 95% sur chaque effet estimé à l'aide de la fonction `confint`.

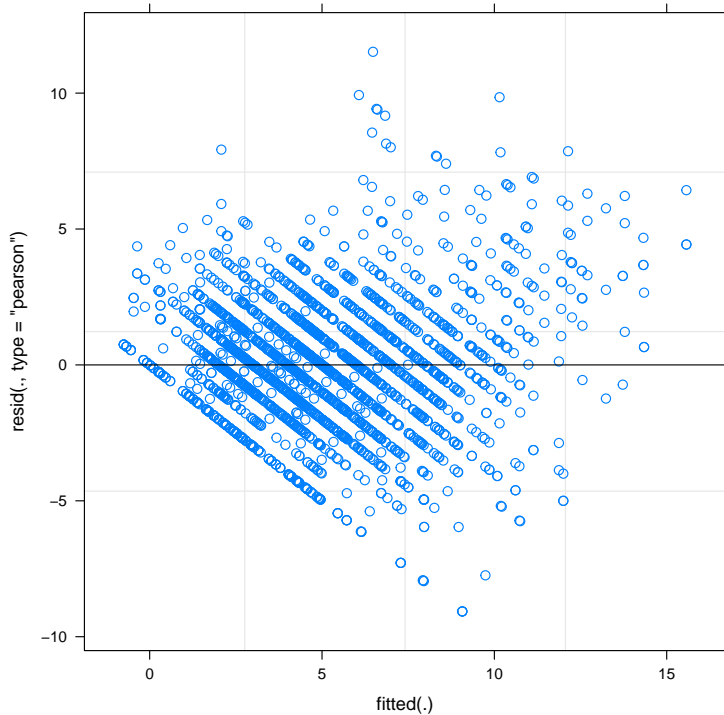
```
confint(mm)

##           2.5 % 97.5 %
## .sig01      1.56  2.51
## .sig02      1.39  2.48
## .sigma      2.20  2.32
## (Intercept) 4.04  6.01
```

3.4 Vérification des conditions d'utilisation du modèle

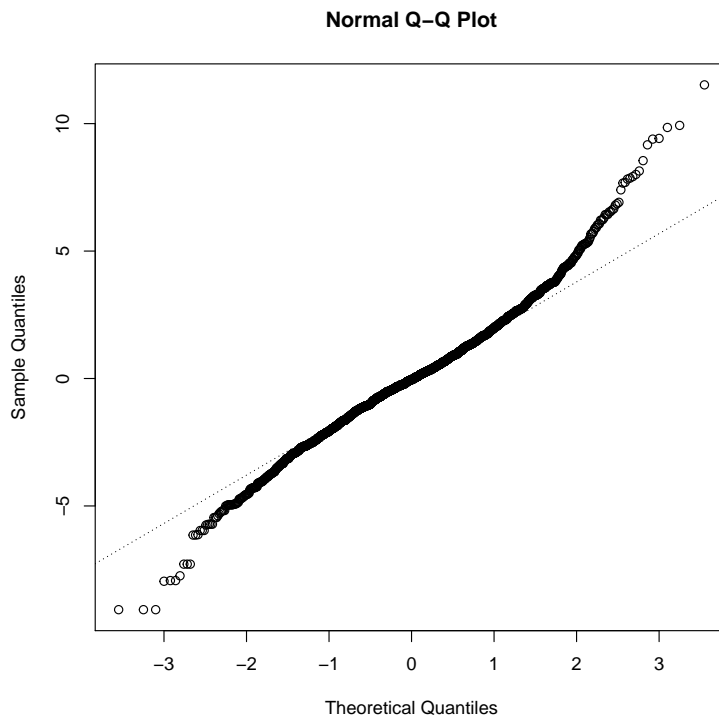
— Graphe des résidus en fonction des valeurs prédites

```
plot(mm)
```



— Graphe Quantile-Quantile des résidus

```
qqnorm(residuals(mm))  
qqline(residuals(mm), lty = 3)
```

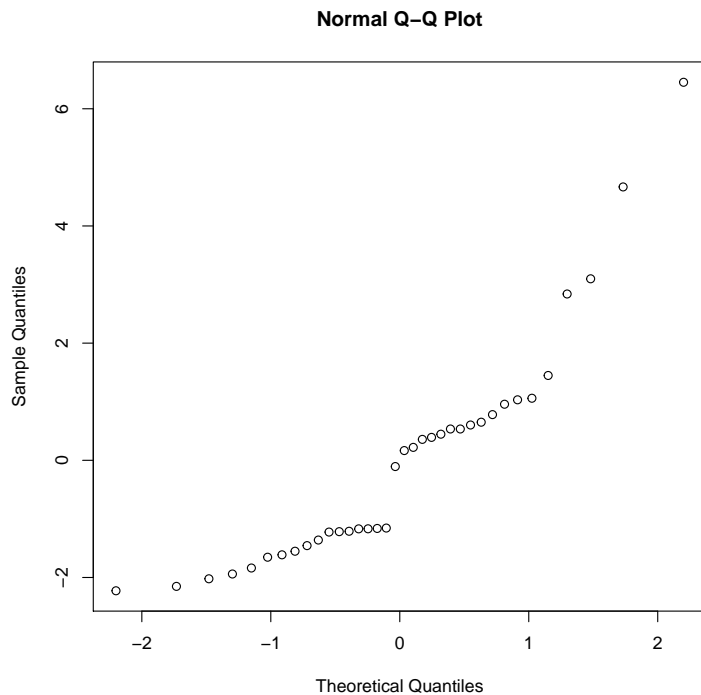


— Graphes Quantile-Quantile des effets aléatoires

Ce graphe permet de vérifier la condition de normalité des effets aléatoires présumée dans un modèle linéaire mixte. Dans cet exemple le faible nombre de modalités des facteurs aléatoires rend cette analyse peu intéressante.

— Effets STEC :

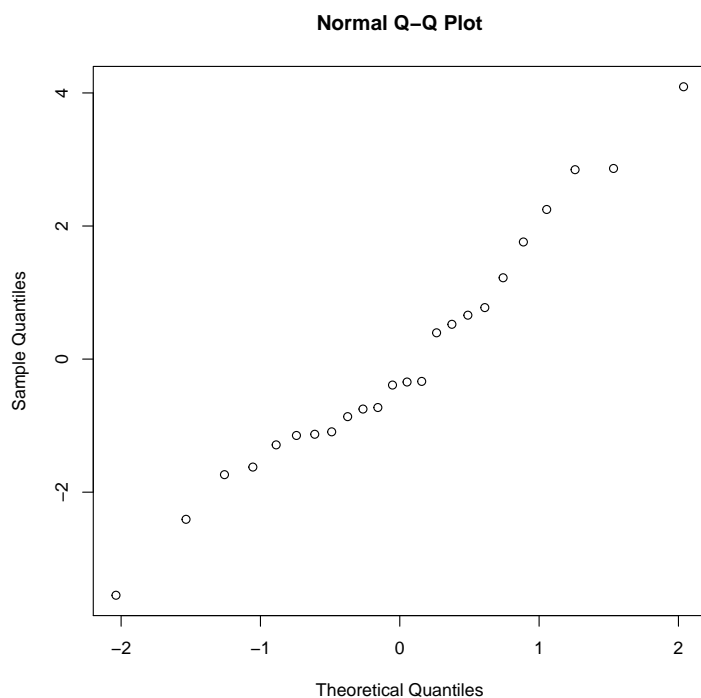
```
qqnorm(ranef(mm)$STEC[,1])
```



Dans cet exemple, la figure précédente est particulièrement intéressante : la répartition des effets liés à la souche STEC met en évidence un écart au modèle gaussien, avec l'apparition de sous-groupes de souches qui mérite d'être explorée

— Effets lactique :

```
qqnorm(ranef(mm)$lactique[,1])
```



Il est possible de faire ces 2 graphes d'un coup plus simplement (mais dans 2 fenêtres différentes) en tapant `plot(ranef(mm))`.

4 Modèle croisé mixte avec répétitions

4.1 Exemple

Une société commercialisant de nombreuses races de souris mutantes à l'usage des laboratoires décide de changer le type de nourriture de ses animaux. Une expérience est entreprise afin de comparer les performances de deux aliments. Cinq souches de souris sont tirées au sort sur le catalogue de la société. Pour chaque souche, huit animaux sont choisis, quatre recevant un des deux aliments. La variable dépendante est le gain de poids (dénommé 'poids' en grammes).

L'objectif de ce protocole est de déterminer l'aliment optimal pour l'ensemble des souches élevées et non pas seulement pour les cinq souches étudiées. Les données sont dans le fichier `NOURRITU.txt`.

4.2 Importation et représentation des données

```
d3 <- read.table("NOURRITU.txt",header=TRUE,  
                stringsAsFactors = TRUE)  
str(d3)
```

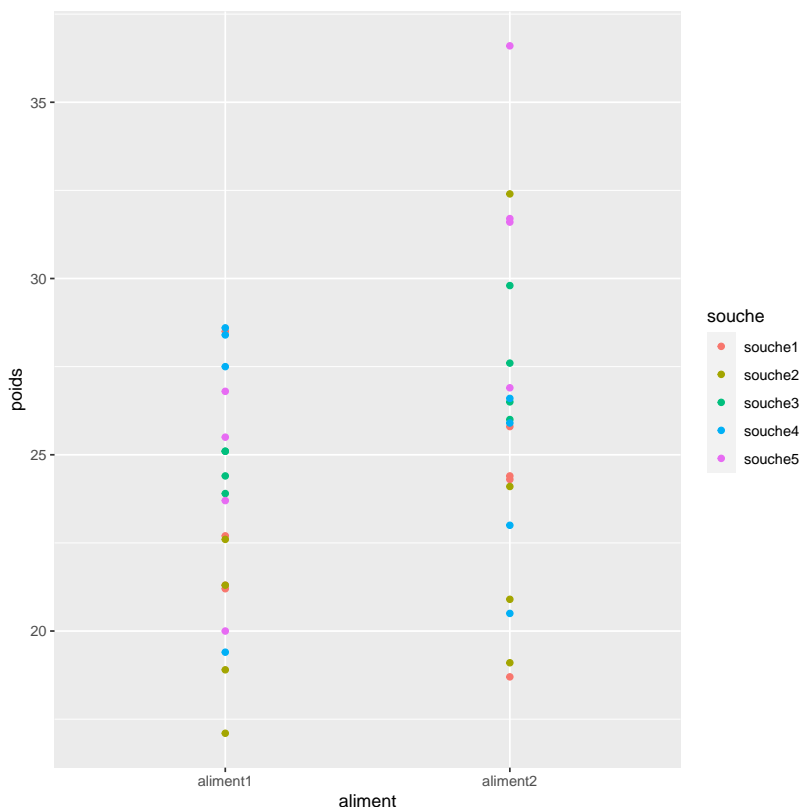
Visualisation du plan d'expérience :

```
xtabs(~ aliment + souche, data = d3)
```

```
##           souche  
## aliment  souche1 souche2 souche3 souche4 souche5  
## aliment1     4     4     4     4     4  
## aliment2     4     4     4     4     4
```

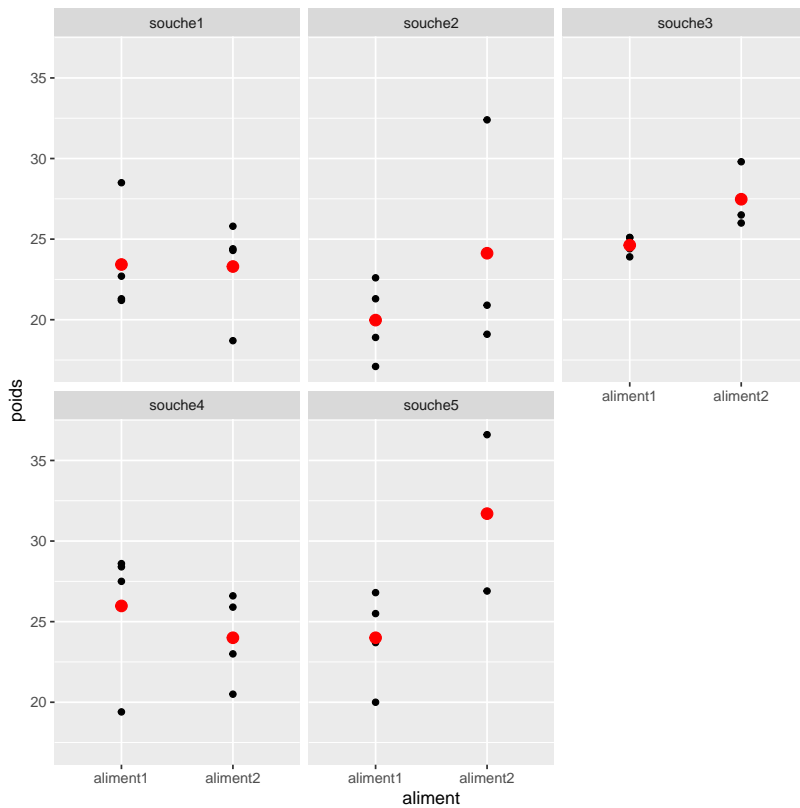
Souvent, dans ce type d'exemple, les données sont visualisées graphiquement uniquement en fonction du facteur fixe, par exemple à l'aide de la fonction `stripchart`. Mais il peut être intéressant de visualiser aussi l'effet du facteur aléatoire par exemple à l'aide de la fonction `qplot` du package `ggplot2`.

```
qplot(aliment, poids, colour = souche, data = d3)
```



On peut aussi éclater le graphe par souche et ajouter les moyennes pour chaque couple de modalités des deux facteurs :

```
qplot(aliment, poids, facets = ~ souche, data = d3) +  
  stat_summary(fun.y = mean, geom = "point", colour = "red", size = 3)
```



4.3 Estimation des effets fixes et des composantes de la variance

Les effets fixes et aléatoires peuvent être estimés en un seul temps à l'aide de la fonction `lmer` du package `lme4` (à charger auparavant).

Le modèle le plus général considère un effet aléatoire à la fois sur le terme constant et sur l'effet fixe lié ici au facteur aliment, c'est-à-dire qu'il prend en compte une interaction potentielle entre le facteur fixe et le facteur aléatoire.

```
(mm <- lmer(poids ~ aliment + (aliment|souche), data = d3))
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: poids ~ aliment + (aliment | souche)
## Data: d3
## REML criterion at convergence: 215
## Random effects:
## Groups Name Std.Dev. Corr
## souche (Intercept) 1.44
## aliment1 aliment2 2.89 -0.12
## Residual 3.43
## Number of obs: 40, groups: souche, 5
## Fixed Effects:
## (Intercept) aliment1 aliment2
## 23.60 2.52
```

Les intervalles de confiance sur chaque effet estimé sont obtenus à l'aide de la fonction `confint` :

```
confint(mm)
##           2.5 % 97.5 %
## .sig01      0.00  4.01
## .sig02     -1.00  1.00
## .sig03      0.00  6.94
## .sigma      2.71  4.48
## (Intercept) 21.45 25.75
## aliment1    -1.10  6.14
```

Un modèle plus simple est parfois utilisé, avec un effet aléatoire uniquement sur le terme constant, c'est-à-dire sans prise en compte de l'interaction potentielle entre le facteur fixe et le facteur aléatoire. Le code suivant correspond à l'ajustement de ce modèle plus simple :

```
(mmsansinteraction <- lmer(poids ~ aliment + (1|souche), data = d3))
```

4.4 Réalisation du tableau d'analyse de variance dans le cas d'un plan équilibré

Dans le cas d'un plan d'expérience équilibré, il est possible de tester l'effet du facteur fixe à l'aide d'une analyse de variance en utilisant le code suivant :

```
summary(aov(poids ~ aliment + souche + Error(aliment:souche), data = d3))

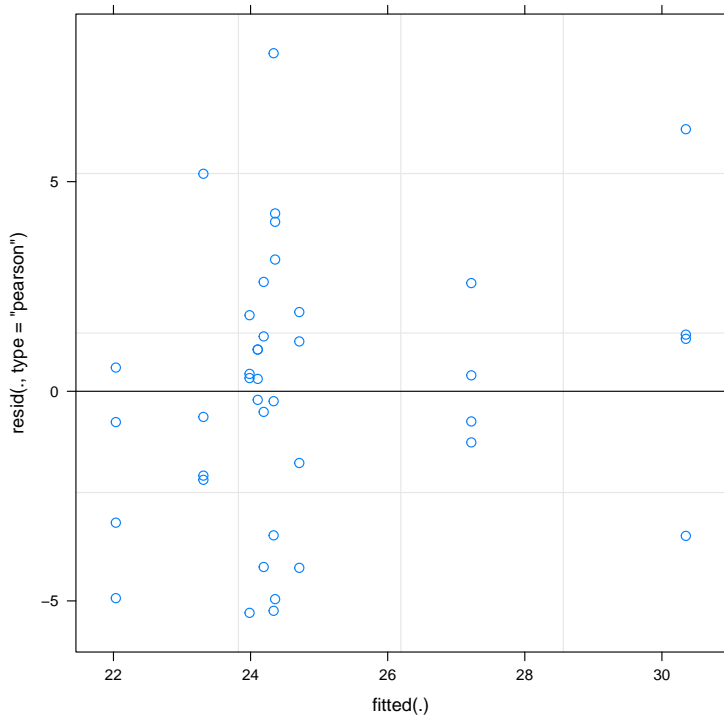
##
## Error: aliment:souche
##           Df Sum Sq Mean Sq F value Pr(>F)
## aliment    1   63.5    63.5    2.24  0.21
## souche     4  164.1    41.0    1.44  0.37
## Residuals  4  113.6    28.4
##
## Error: Within
##           Df Sum Sq Mean Sq F value Pr(>F)
## Residuals 30   352    11.7
```

Il ne faut lire dans la table précédente que la valeur de p associée au facteur fixe, mais attention, les résultats obtenus avec cette analyse sont différents de ceux que l'on obtiendrait en ANOVA 1, c'est-à-dire en ne prenant pas en compte l'effet souche.

4.5 Vérification des conditions d'utilisation du modèle

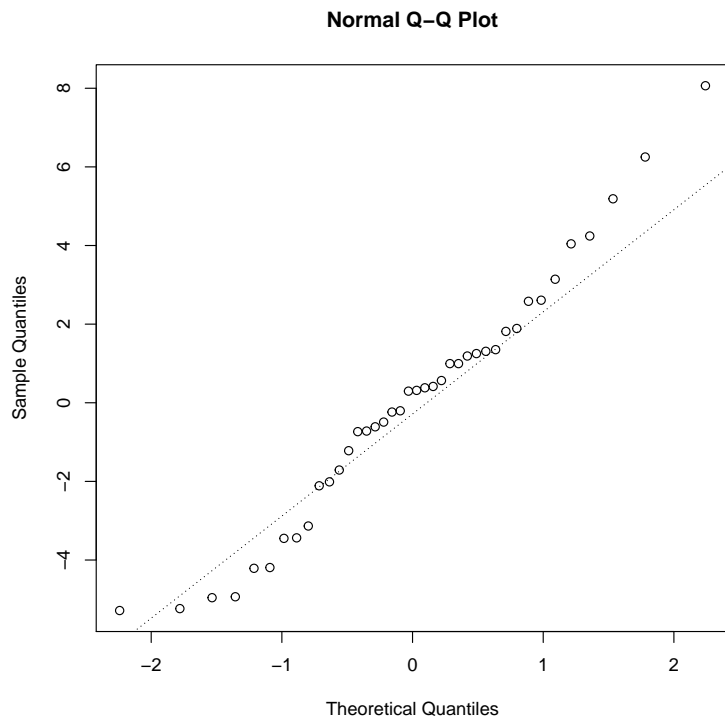
— Graphe des résidus en fonction des valeurs prédites

```
plot(mm)
```



— Graphe Quantile-Quantile des résidus

```
qqnorm(residuals(mm))
qqline(residuals(mm), lty = 3)
```

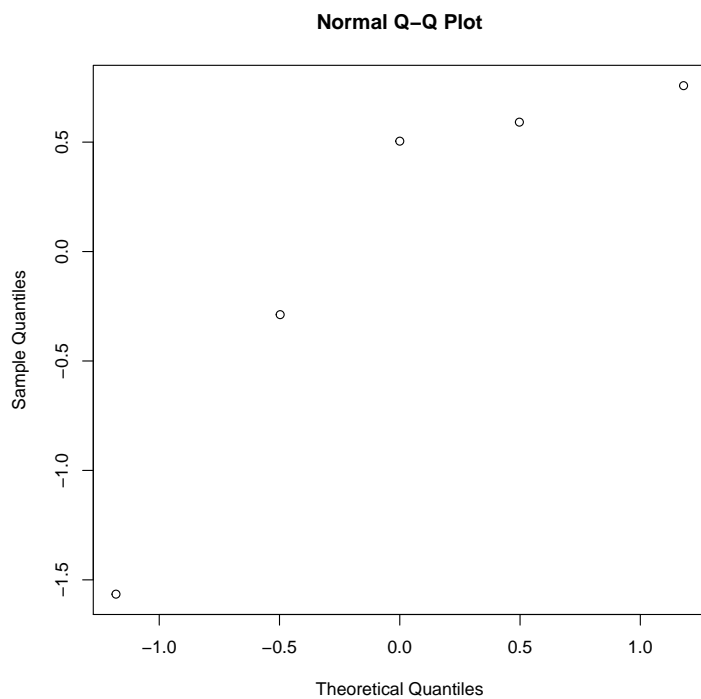


— **Graphes Quantile-Quantile des effets aléatoires**

Ce graphe permet de vérifier la condition de normalité des effets aléatoires présumée dans un modèle linéaire mixte. Dans cet exemple le faible nombre de modalités des facteurs aléatoires rend cette analyse peu intéressante.

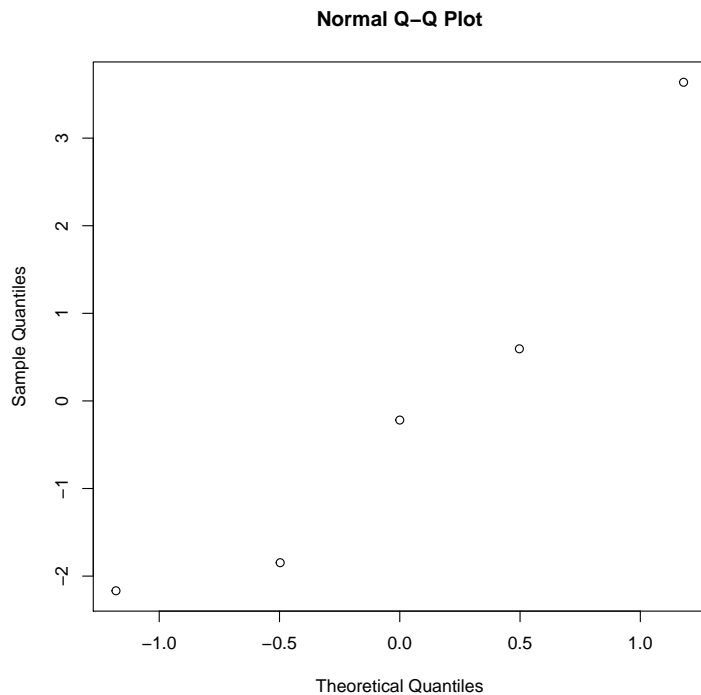
— Effet souche sur le terme constant :

```
qqnorm(ranef(mm)$souche[,1])
```



— Effet souche sur l'effet fixe :

```
qqnorm(ranef(mm)$souche[,2])
```

5 Modèle croisé mixte sans répétition

5.1 Exemple de modèle mixte sans répétition

On désire étudier comparativement l'influence de 3 traitements sur le développement d'une tumeur greffée et de ses métastases. on dispose pour cela de 15 animaux. Le critère de jugement choisi est le poids de la tumeur et des métastases après 2 mois de traitement. Cette détermination étant longue et minutieuse, on ne peut sacrifier et examiner que 3 animaux par jour en fin d'expérience, et on décide de ce fait de constituer 5 blocs de 3 animaux recevant les 3 traitements différents. Les résultats (en g) sont dans le fichier `tumeur.txt`.

5.2 Importation et représentation des données

```
d4 <- read.table("tumeur.txt", header = TRUE,
                 stringsAsFactors = TRUE)
str(d4)
```

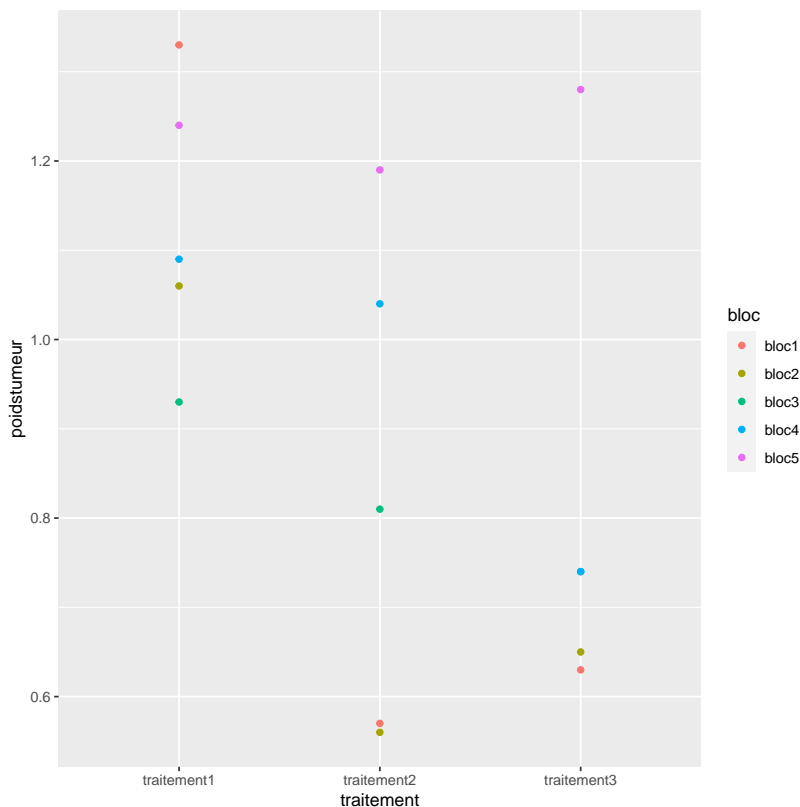
Visualisation du plan d'expérience :

```
xtabs(~ traitement + bloc, data = d4)

##           bloc
## traitement bloc1 bloc2 bloc3 bloc4 bloc5
## traitement1     1     1     1     1     1
## traitement2     1     1     1     1     1
## traitement3     1     1     1     1     1
```

Dans ce type d'exemple les données sont souvent visualisées graphiquement en fonction du facteur fixe à l'aide de la fonction `stripchart` mais il est pertinent de faire apparaître le facteur par exemple par un codage de couleur avec le package `ggplot2`.

```
qplot(traitement, poids_tumeur, data = d4, colour = bloc)
```



5.3 Estimation des effets fixes et des composantes de la variance

Les effets fixes et aléatoires peuvent être estimés en un seul temps à l'aide de la fonction `lmer` du package `lme4` (à charger auparavant).

L'absence de répétition rend impossible la prise en compte dans le modèle d'une éventuelle interaction entre le facteur fixe et le facteur aléatoire. L'analyse de telles données fait donc d'emblée l'hypothèse d'absence d'une telle interaction.

```
(mm <- lmer(poidstumeur ~ traitement + (1|bloc), data = d4))

## Linear mixed model fit by REML ['lmerMod']
## Formula: poidstumeur ~ traitement + (1 | bloc)
## Data: d4
## REML criterion at convergence: 2.94
## Random effects:
## Groups Name Std.Dev.
## bloc (Intercept) 0.156
## Residual 0.185
## Number of obs: 15, groups: bloc, 5
## Fixed Effects:
## (Intercept) traitementtraitement2 traitementtraitement3
## 1.130 -0.296 -0.322
```

Les intervalles de confiance sur chaque effet estimé sont obtenus à l'aide de la fonction `confint` :

```
confint(mm)

## 2.5 % 97.5 %
## .sig01 0.000 0.3558
## .sigma 0.113 0.2762
## (Intercept) 0.920 1.3398
## traitementtraitement2 -0.522 -0.0697
## traitementtraitement3 -0.548 -0.0957
```

5.4 Réalisation du tableau d'analyse de variance dans le cas d'un plan équilibré

Dans le cas d'un plan d'expérience équilibré, il est possible de tester l'effet du facteur fixe à l'aide d'une analyse de variance en utilisant le code suivant :

```
summary(aov(poidstumeur ~ traitement + bloc, data = d4))
```

```
##           Df Sum Sq Mean Sq F value Pr(>F)
## traitement  2  0.320  0.1600   4.68  0.045
## bloc        4  0.428  0.1071   3.13  0.079
## Residuals   8  0.273  0.0342
```

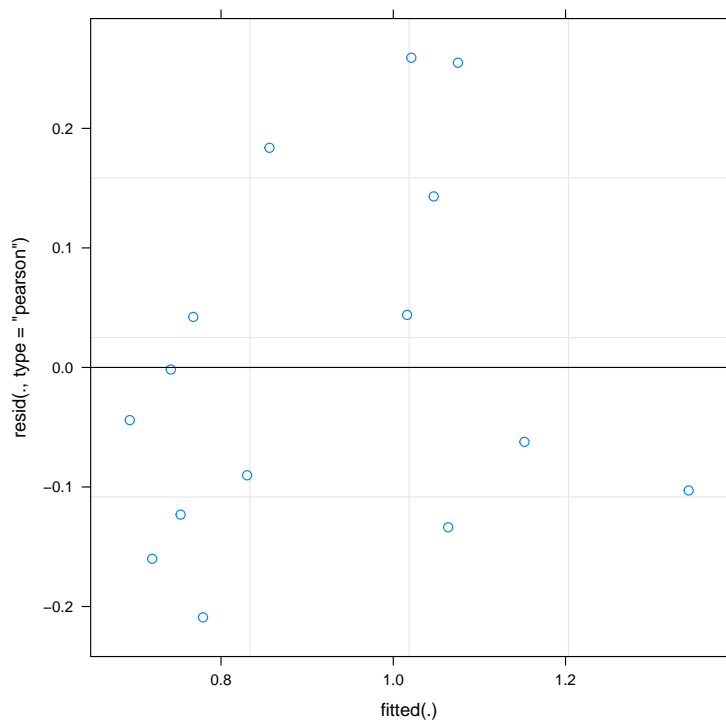
équivalent au code

```
anova(lm(poidstumeur ~ traitement + bloc, data = d4))
```

5.5 Vérification des conditions d'utilisation du modèle

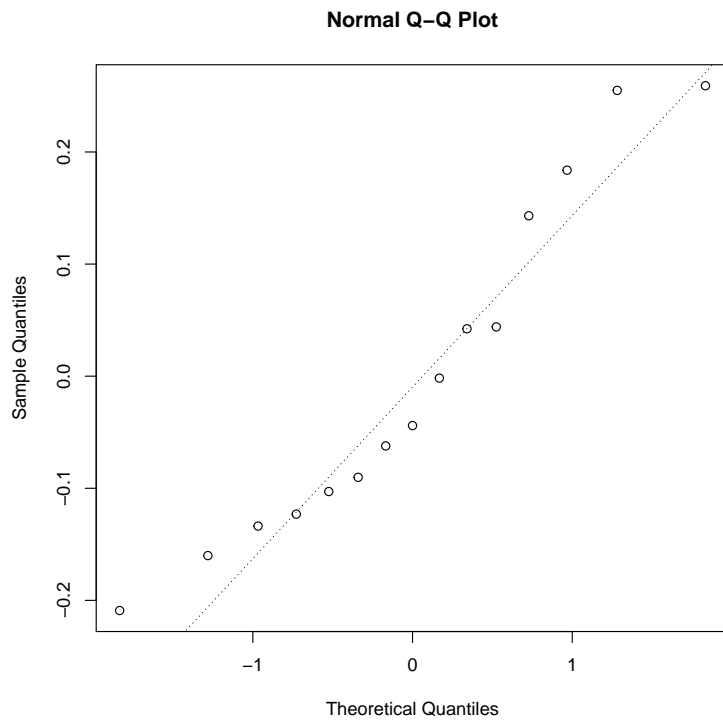
— Graphe des résidus en fonction des valeurs prédites

```
plot(mm)
```



— Graphe Quantile-Quantile des résidus

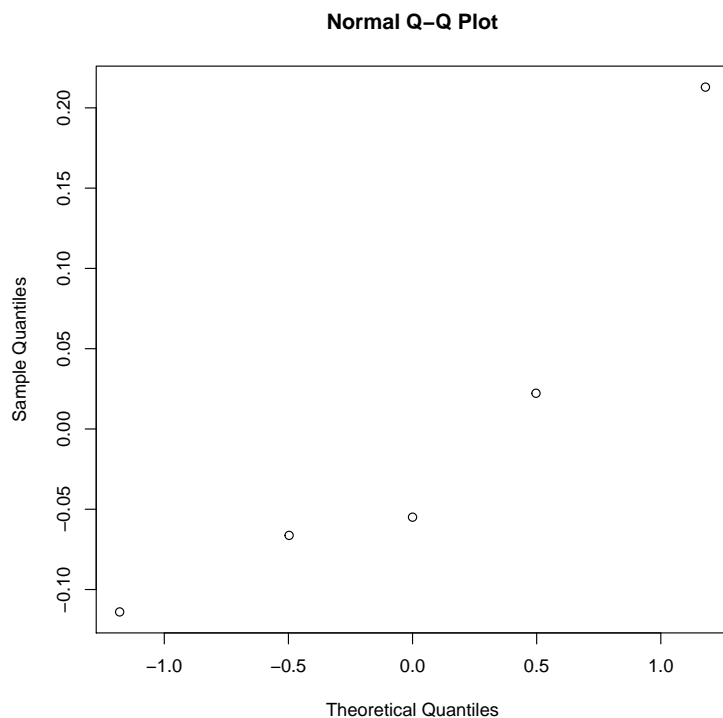
```
qqnorm(residuals(mm))
qqline(residuals(mm), lty = 3)
```



— **Graphes Quantile-Quantile des effets aléatoires**

Ce graphe permet de vérifier la condition de normalité des effets aléatoires présumée dans un modèle linéaire mixte. Dans cet exemple le faible nombre de modalités du facteur aléatoire rend cette analyse peu intéressante.

```
qqnorm(ranef(mm)$bloc[,1])
```



6 Modèle croisé aléatoire sans répétition

L'analyse de données correspondant à un modèle croisé aléatoire sans répétition suit la même logique que celle d'un modèle croisé avec répétitions (cf. chapitre 3).

7 Modèle hiérarchisé aléatoire

7.1 Exemple de modèle hiérarchisé aléatoire

Le tableau de données du fichier `PDSSOURI.txt` répertorie les poids à 6 semaines de souris femelles issues de différentes portées issues de 18 femelles qui ont été croisées avec 6 mâles différents (à raison de 3 femelles par mâle). Deux souris femelles ont ensuite été tirées au sort par portée et pesées. On voudrait quantifier la variabilité des poids des femelles à 6 mois due respectivement au facteur “père” et au facteur “mère”.

7.2 Importation et représentation des données

```
d5 <- read.table("PDSSOURI.txt", header = TRUE,
                stringsAsFactors = TRUE)
```

Il est très important, lorsqu'on a un modèle hiérarchisé, de vérifier que le codage a été réalisé correctement, et en particulier que chaque modalité du facteur de niveau inférieur porte bien un nom qui lui est propre : par exemple ici que chaque femelle porte un nom différent et que l'on n'a pas redonné des noms identiques aux femelles des mâles différents.

```
str(d5)

## 'data.frame': 36 obs. of 3 variables:
## $ pere : Factor w/ 6 levels "A","B","C","D",...: 1 1 1 1 1 1 2 2 2 2 ...
## $ mere : Factor w/ 18 levels "m1","m10","m11",...: 1 1 11 11 12 12 13 13 14 14 ...
## $ poids: num 19.3 21.9 22.7 24.6 21 19.1 19.3 21.9 20.6 17.9 ...
```

Visualisation du plan d'expérience :

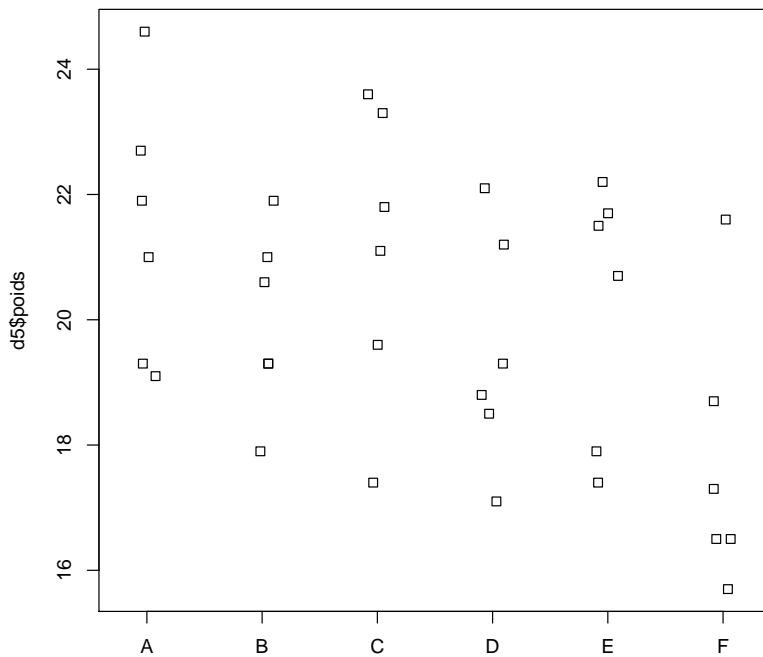
```
xtabs(~ pere + mere, data = d5)

##      mere
## pere m1 m10 m11 m12 m13 m14 m15 m16 m17 m18 m2 m3 m4 m5 m6 m7 m8 m9
##  A   2  0  0  0  0  0  0  0  0  0  2  2  0  0  0  0  0  0
##  B   0  0  0  0  0  0  0  0  0  0  0  0  0  2  2  2  0  0  0
##  C   0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  2  2  2
##  D   0  2  2  2  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##  E   0  0  0  0  2  2  2  0  0  0  0  0  0  0  0  0  0  0  0
##  F   0  0  0  0  0  0  0  0  2  2  2  0  0  0  0  0  0  0  0
```

Un modèle hiérarchisé à 2 facteurs ne peut être ajusté que s'il y a des répétitions, sinon on est ramené à ajuster un modèle à un seul facteur

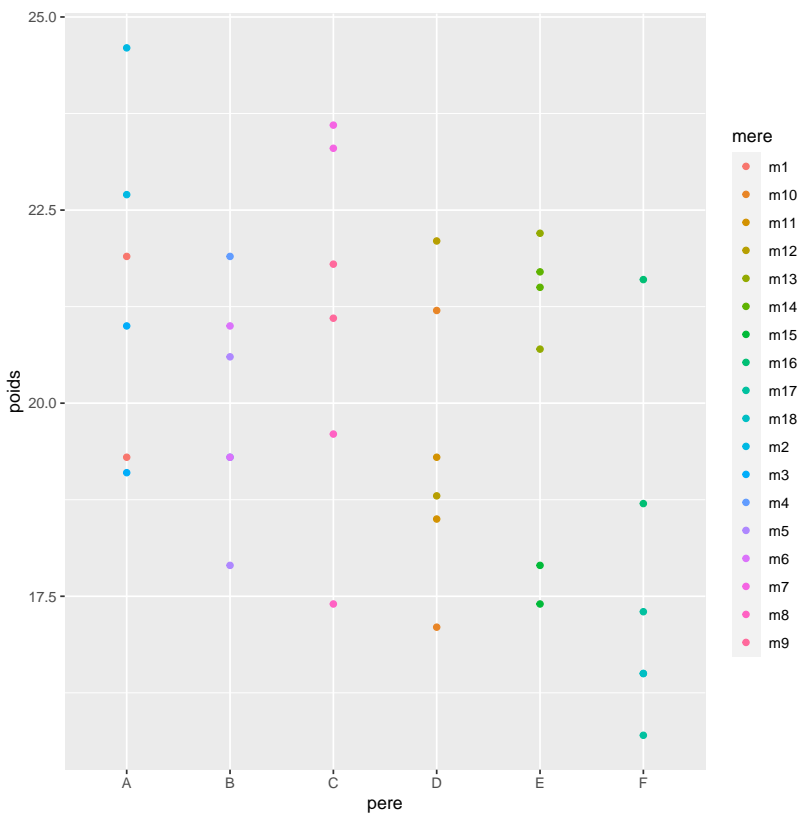
Dans ce type d'exemple, on visualisera classiquement les données en fonction du facteur du niveau supérieur, à l'aide de la fonction `stripchart` par exemple.

```
stripchart(d5$poids ~ d5$pere, vertical = TRUE, method = "jitter")
```



On peut rendre plus informatif ce graphe en utilisant la couleur et/ou le type de points pour différencier les modalités sur facteur de niveau inférieur, par exemple à l'aide de la fonction `qplot` du package `ggplot2`.

```
qplot(pere, poids, data = d5, colour = mere)
```



7.3 Estimation des composantes de la variance

On peut estimer les composantes de la variance à l'aide de la fonction `lmer` du package `lme4` préalablement chargé.

```
(mm <- lmer(poids ~ (1|pere/mere), data = d5))

## Linear mixed model fit by REML ['lmerMod']
## Formula: poids ~ (1 | pere/mere)
## Data: d5
## REML criterion at convergence: 152
## Random effects:
## Groups Name Std.Dev.
## mere:pere (Intercept) 1.531
## pere (Intercept) 0.793
## Residual 1.474
## Number of obs: 36, groups: mere:pere, 18; pere, 6
## Fixed Effects:
## (Intercept)
## 20
```

Les intervalles de confiance associés sont obtenus à l'aide de la fonction `confint` :

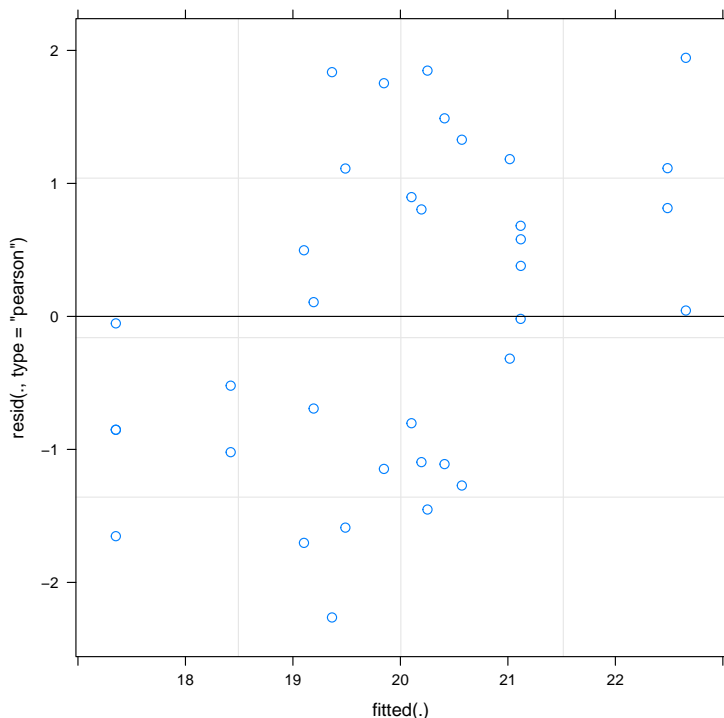
```
confint(mm)

##           2.5 % 97.5 %
## .sig01      0.504  2.58
## .sig02      0.000  2.20
## .sigma      1.098  2.13
## (Intercept) 18.852 21.15
```

7.4 Vérification des conditions d'utilisation du modèle

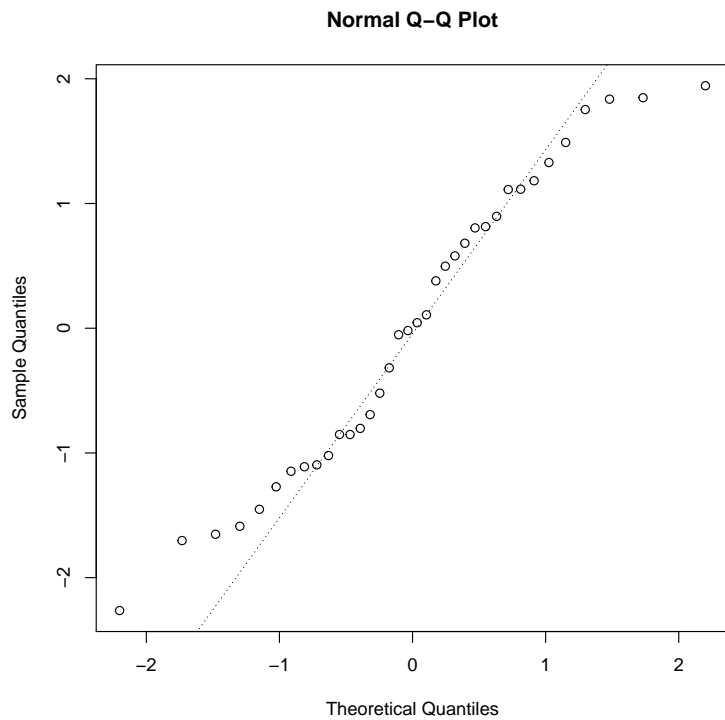
— Graphe des résidus en fonction des valeurs prédites

```
plot(mm)
```



— Graphe Quantile-Quantile des résidus

```
qqnorm(residuals(mm))
qqline(residuals(mm), lty = 3)
```

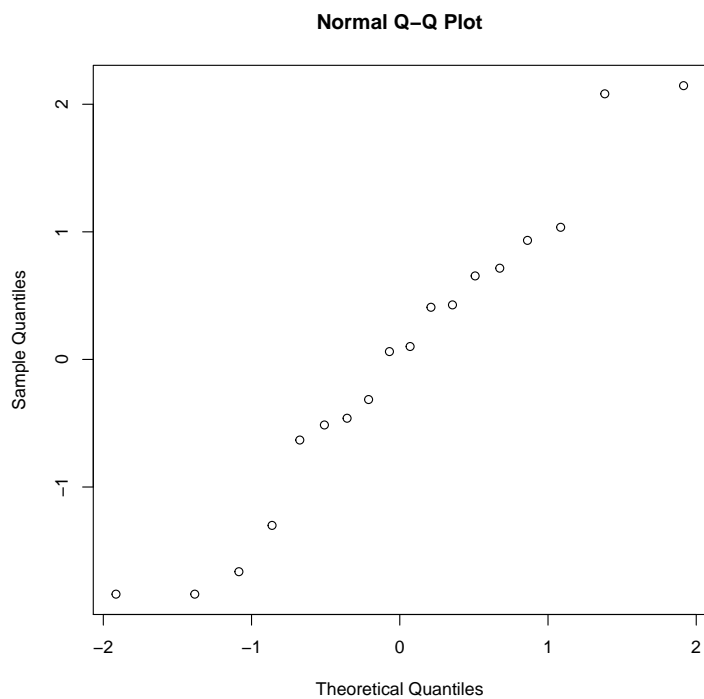


— **Graphes Quantile-Quantile des effets aléatoires**

Ce graphe permet de vérifier la condition de normalité des effets aléatoires présumée dans un modèle linéaire mixte.

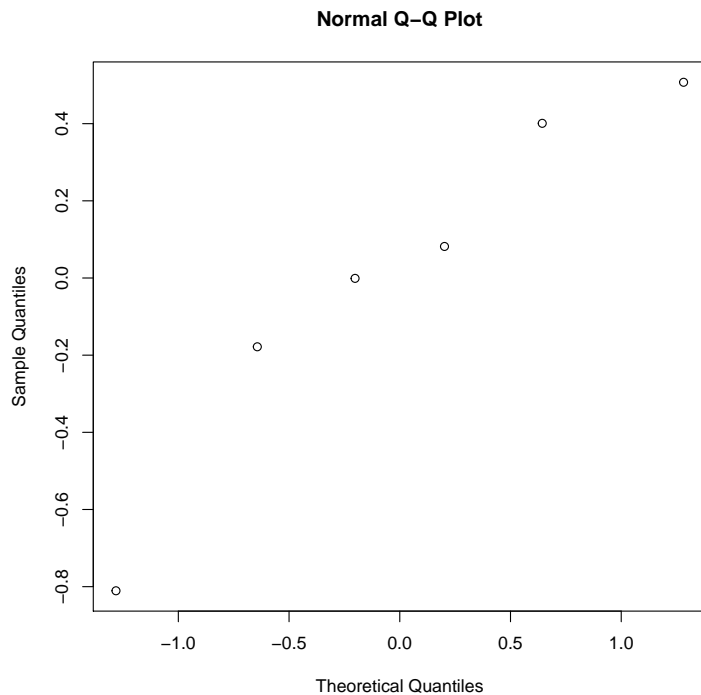
— Effet du facteur au niveau inférieur (mère ici)

```
qqnorm(ranef(mm)$mere[,1])
```



— Effet du facteur au niveau supérieur (père ici)

```
qqnorm(ranef(mm)$pere[,1])
```

8 Modèle hiérarchisé mixte

8.1 Exemple de modèle hiérarchisé mixte

A partir de 6 souris A, B, C, D, E, F, dites "donneuses" (de cellules cancéreuses), on injecte 200 000 cellules à 24 souris "receveuses". Au bout d'un certain temps on compte le nombre de nodules pulmonaires qui apparaissent chez chacune des 24 souris. Les souris "donneuses" A, B, C ont subi une irradiation aux rayons X, les souris "donneuses" D, E, F, ont subi une irradiation au rayon γ . Les résultats de l'expérience sont stockés dans le fichier `souris.txt`. On voudrait savoir si le traitement reçu par les donneuses modifie l'avenir des receveuses. Il utilisera un modèle hiérarchisé mixte avec un facteur fixe (l'irradiation) et un facteur aléatoire (la souris donneuse) emboîté (ou niché) dans le facteur fixe.

8.2 Importation et représentation des données

```
d6 <- read.table("souris.txt", header = TRUE,
                stringsAsFactors = TRUE)
```

Il est très important, lorsqu'on a un modèle hiérarchisé, de vérifier que le codage a été réalisé correctement, et en particulier que chaque modalité du facteur de niveau inférieur porte bien un nom qui lui est propre : par exemple ici que chaque souris donneuse porte un nom différent et que l'on n'a pas redonné des noms identiques aux donneuses pour les deux types d'irradiation.

```
str(d6)

## 'data.frame': 24 obs. of 3 variables:
## $ irradiation: Factor w/ 2 levels "gamma","x": 2 2 2 2 2 2 2 2 2 2 ...
## $ donneuse : Factor w/ 6 levels "A","B","C","D",...: 1 1 1 1 2 2 2 2 3 3 ...
## $ nbnodules : int 10 16 21 28 15 21 26 29 25 27 ...
```

Visualisation du plan d'expérience :

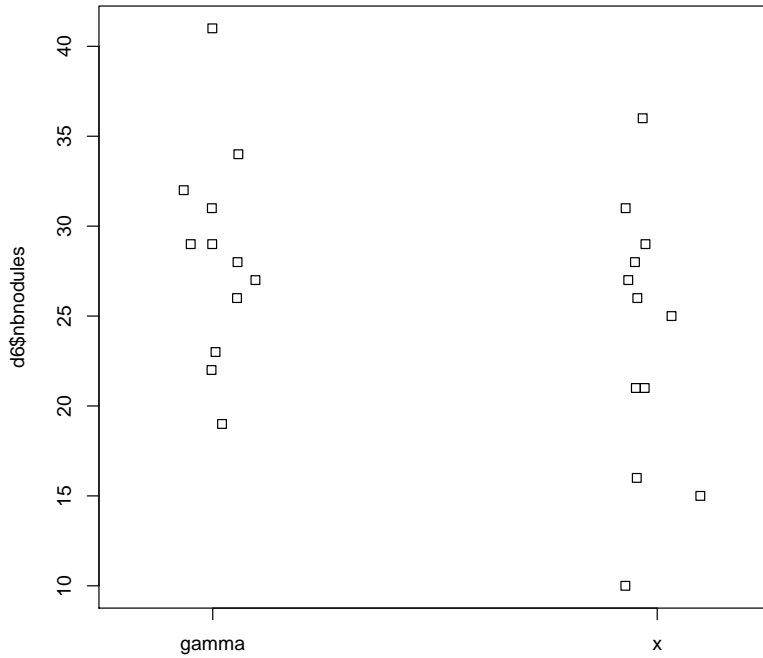
```
xtabs(~ irradiation + donneuse, data = d6)

##           donneuse
## irradiation A B C D E F
##      gamma 0 0 0 4 4 4
##      x     4 4 4 0 0 0
```

Un modèle hiérarchisé à 2 facteurs ne peut être ajusté que s'il y a des répétitions, sinon on est ramené à ajuster un modèle à un seul facteur

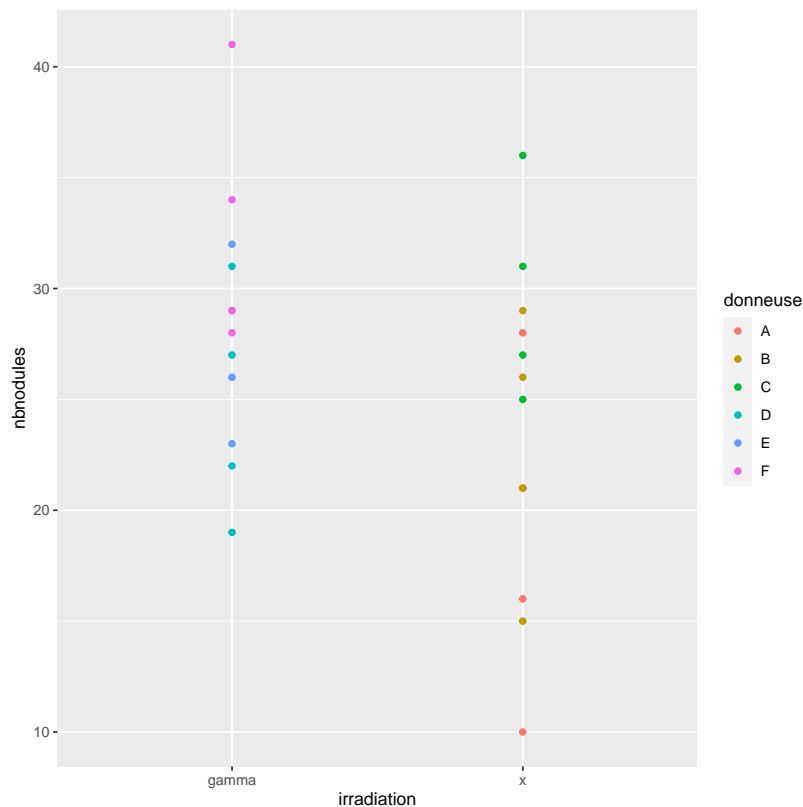
Dans ce type d'exemple, on visualisera classiquement les données en fonction du facteur du niveau supérieur, à l'aide de la fonction `stripchart` par exemple.

```
stripchart(d6$nbmodules ~ d6$irradiation, vertical = TRUE, method = "jitter")
```



On peut rendre plus informatif ce graphe en utilisant la couleur et/ou le type de points pour différencier les modalités sur facteur de niveau inférieur, à l'aide de la fonction `qplot` du package `ggplot2`.

```
qplot(irradiation, nbmodules, data = d6, colour = donneuse)
```



8.3 Estimation des composantes de la variance

On peut estimer les composantes de la variance à l'aide de la fonction `lmer` du package `lme4` préalablement chargé.

```
(mm <- lmer(nbnodes ~ irradiation + (1|donneuse), data = d6))  
  
## Linear mixed model fit by REML ['lmerMod']  
## Formula: nbnodes ~ irradiation + (1 | donneuse)  
## Data: d6  
## REML criterion at convergence: 149  
## Random effects:  
## Groups Name Std.Dev.  
## donneuse (Intercept) 4.01  
## Residual 5.74  
## Number of obs: 24, groups: donneuse, 6  
## Fixed Effects:  
## (Intercept) irradiationx  
## 28.42 -4.67
```

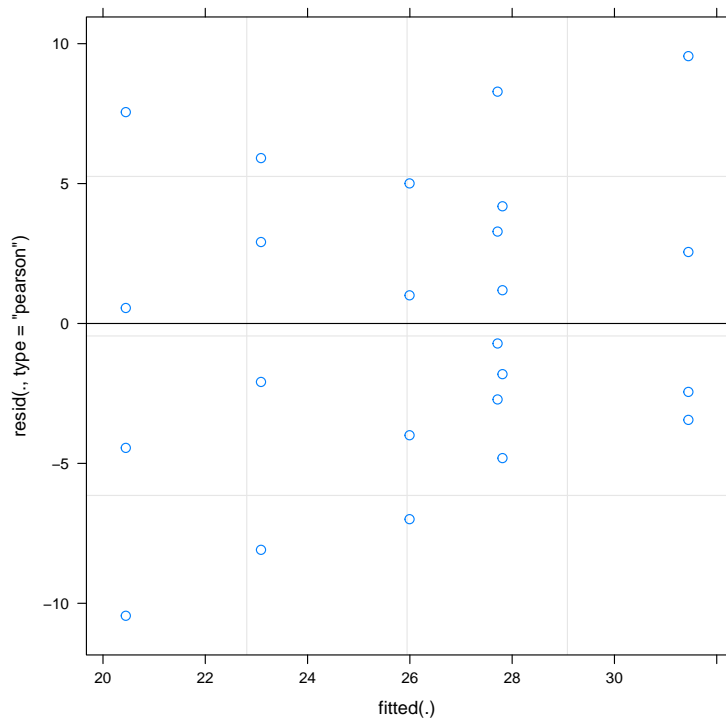
Les intervalles de confiance associés sont obtenus à l'aide de la fonction `confint` :

```
confint(mm)  
  
##           2.5 % 97.5 %  
## .sig01      0.00  7.55  
## .sigma      4.28  8.28  
## (Intercept) 23.02 33.81  
## irradiationx -12.29  2.96
```

8.4 Vérification des conditions d'utilisation du modèle

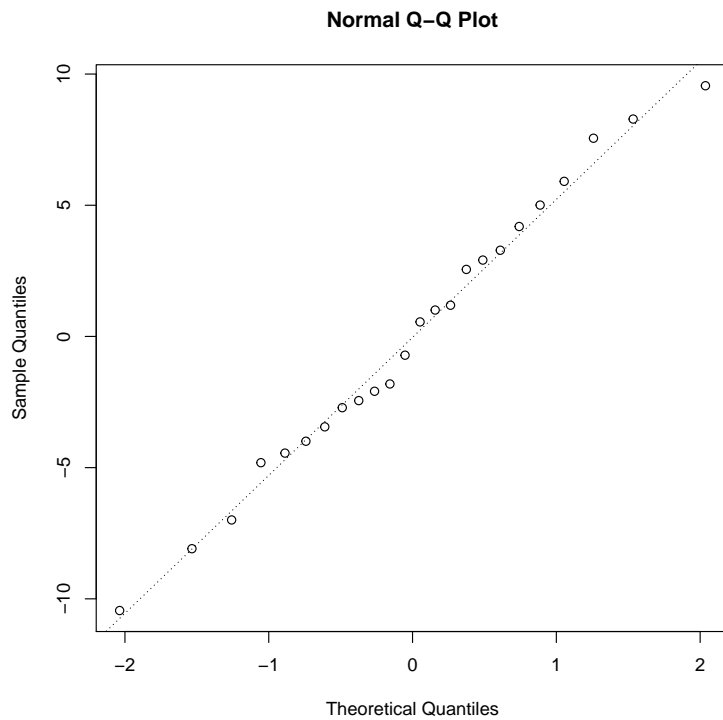
— Graphe des résidus en fonction des valeurs prédites

```
plot(mm)
```



— Graphe Quantile-Quantile des résidus

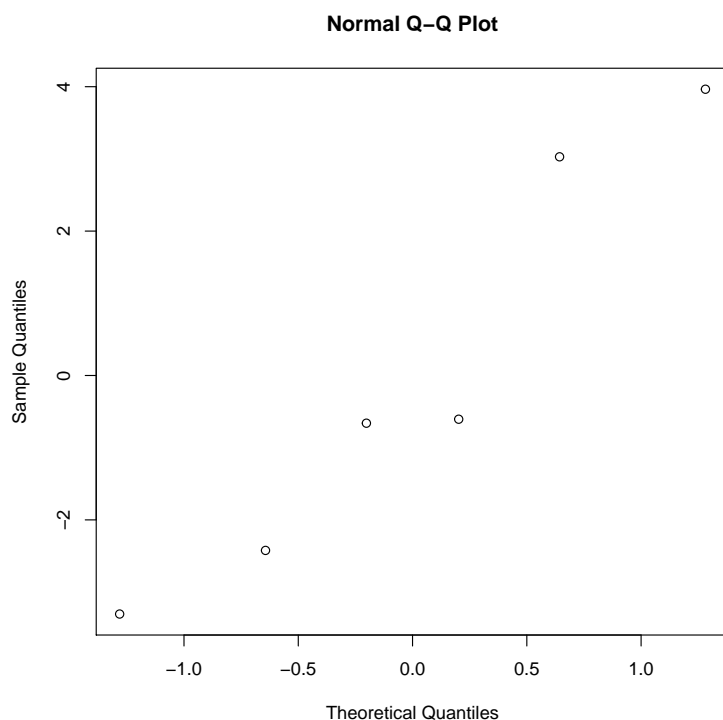
```
qqnorm(residuals(mm)); qqline(residuals(mm), lty = 3)
```



— **Graphes Quantile-Quantile de l'effet aléatoire**

Ce graphe permet de vérifier la condition de normalité des effets aléatoires présumée dans un modèle linéaire mixte.

```
qqnorm(ranef(mm)$donneuse[,1])
```



8.5 Réalisation du tableau d'analyse de variance dans le cas d'un plan équilibré

Dans le cas d'un plan d'expérience équilibré, il est possible de tester l'effet du facteur fixe à l'aide d'une analyse de variance en utilisant le code suivant :

```
summary(aov(nbnodules ~ irradiation + Error(donneuse), data = d6))
##
## Error: donneuse
##           Df Sum Sq Mean Sq F value Pr(>F)
```

```
## irradiation 1 131 130.7 1.34 0.31
## Residuals 4 389 97.3
##
## Error: Within
## Df Sum Sq Mean Sq F value Pr(>F)
## Residuals 18 594 33
```

Attention, les résultats obtenus avec cette analyse sont différents de ceux que l'on obtiendrait en ANOVA 1, c'est-à-dire en ne prenant pas en compte l'effet donneuse, et aussi différents de ceux que l'on obtiendrait en faisant par erreur une ANOVA 2 classique avec 2 facteurs fixes.