

Statistiques appliquées avec R

SCI 1018

Blocs complets aléatoires

Marc J. Mazerolle

Département des sciences du bois et de la forêt, Université Laval

Avec révisions mineures de

Élise Filotas, *Département science et technologie, Université TÉLUQ*, et

Marc-Olivier Martin-Guay, *Département des sciences biologiques, Université du Québec à Montréal*



Table des matières

1	Introduction	2
2	Effets fixes et effets aléatoires	2
2.1	ANOVA à un critère	4
2.2	ANOVA à deux critères	4
3	Blocs complets aléatoires	5
3.1	Révision de l'échantillonnage	5
3.2	Origine du dispositif en blocs complets aléatoires	9
3.3	Particularité du dispositif en blocs complets aléatoires	9
3.4	Applications moins conventionnelles	21
3.5	Conditions d'application	22
3.6	Avantages du dispositif en blocs complets aléatoires	23
3.7	Désavantages du dispositif en blocs complets aléatoires	23
	Conclusion	24
	Index	25

1 Introduction

Dans la leçon précédente, nous avons présenté l'ANOVA à deux critères avec répétitions. Nous avons vu en détail les concepts d'interaction et d'additivité. Rappelons qu'en présence d'une interaction entre deux facteurs, on ne peut pas interpréter les effets principaux séparément. En effet, une interaction implique que l'effet d'un facteur sur la variable réponse dépend du niveau de l'autre. On dit dans ce cas que les effets ne sont pas additifs.

2 Effets fixes et effets aléatoires

Dans les leçons précédentes, nous avons vu l'ANOVA à un critère et l'ANOVA à deux critères. Dans tous les exemples que nous avons présentés, les effets étudiés étaient fixés par l'expérimentateur. Par exemple, on s'intéressait à l'effet du sexe ou d'un traitement hormonal sur la concentration en calcium dans le plasma sanguin. Lorsque les niveaux du facteur sont spécifiquement choisis par l'expérimentateur, nous dirons que c'est un **effet fixe**. Ce genre d'effet est le plus commun.

Exemple 8.1 Dans une expérience sur l'effet de la zoothérapie sur des personnes atteintes d'un stress chronique, on sélectionne aléatoirement 30 personnes atteintes. L'expérimentateur s'intéresse à la différence entre les traitements suivants : thérapie accompagnée d'un chat, thérapie accompagnée d'un chien et thérapie conventionnelle. On attribue aléatoirement l'un des traitements à chaque personne atteinte, de façon à avoir 10 répétitions pour chaque traitement. Puisque l'expérimentateur s'intéresse spécifiquement à ces trois traitements, l'effet de la zoothérapie est un effet fixe.

En contrepartie, lorsque les niveaux d'un facteur sont déterminés aléatoirement, nous dirons que c'est un **effet aléatoire**. Autrement dit, l'expérimentateur choisit aléatoirement les niveaux qui seront inclus dans l'expérience parmi tous les niveaux disponibles. L'effet aléatoire permet d'étudier l'effet global d'un facteur sur la variable réponse en considérant tous les traitements possibles plutôt que les différences entre des traitements spécifiques.

Exemple 8.2 On veut connaître s'il y a un effet du fournisseur de batteries de téléphone portable sur leur durée de vie. Dans un catalogue de fournisseurs de batteries, on sélectionne aléatoirement 3 fournisseurs de batteries. On commande 10 batteries de chacun des 3 fournisseurs. Puisque les trois niveaux du facteur « fournisseur de batteries » ont été sélectionnés aléatoirement, nous dirons que c'est un effet aléatoire.

La différence entre l'effet fixe et l'effet aléatoire est au niveau de l'interprétation des résultats. Avec un effet fixe, on s'intéresse à la différence entre les niveaux spécifiques du facteur et notre conclusion se rapporte directement à ces niveaux. Pour l'effet aléatoire, la conclusion est pertinente à l'effet global du facteur et à tous les niveaux possibles du facteur. Dans le dernier cas, il peut exister une multitude de niveaux qui n'ont pas nécessairement été inclus dans l'analyse. L'exemple suivant distingue les deux types d'effets.

Exemple 8.3 On veut déterminer la variabilité géographique de la concentration en cadmium dans le sol au Québec. Si l'expérimentateur identifie différentes régions du Québec et qu'il sélectionne aléatoirement trois régions, on considérera l'effet comme aléatoire. À l'opposé, si l'expérimentateur s'intéresse spécifiquement

aux régions de l’Abitibi, du Lac St-Jean et de la Côte-Nord, on considérera l’effet comme fixe.

2.1 ANOVA à un critère

Lorsqu’on effectue une ANOVA à un critère, le calcul de la somme des carrés, des carrés moyens et du F demeure le même, peu importe que l’effet soit fixe ou aléatoire. Seule l’interprétation va changer.

2.2 ANOVA à deux critères

En ce qui concerne l’ANOVA à deux critères (A et B) avec répétition, c’est-à-dire avec plus d’une observation, on distingue trois scénarios d’effets fixes et aléatoires possibles :

- Modèle I : A fixe et B fixe
- Modèle II : A aléatoire et B aléatoire
- Modèle III : A aléatoire et B fixe¹

Le type d’effet (fixe *vs* aléatoire) détermine le terme d’erreur à utiliser dans le tableau d’ANOVA. Le tableau 1 présente la formulation des différents ratios des carrés moyens pour tester les effets de A, de B et de leur interaction selon la nature de ces effets.

Le modèle III est aussi appelé **modèle mixte** puisqu’il contient des effets fixes et des effets aléatoires. Les modèles mixtes englobent une variété d’approches qui tiennent compte de la structure de données complexes, telles que celles provenant d’un dispositif dit niché ou emboité (p.ex., sondages auprès d’individus emboîtés dans des quartiers, ces derniers emboîtés eux-mêmes dans différentes villes) ou encore avec des mesures répétées dans le

1. Le cas opposé, avec A fixe et B aléatoire, est bien sûr équivalent.

Tableau 1 – Termes d’erreurs utilisés pour tester les effets de A, de B et de leur interaction selon la nature des effets (fixes ou aléatoires) en présence de répétitions.

Facteur	Modèle I (A fixe, B fixe)	Modèle II (A aléatoire, B aléatoire)	Modèle III (A aléatoire, B fixe)
A	$\frac{MS\ A}{MS\ erreur}$	$\frac{MS\ A}{MS\ A \times B}$	$\frac{MS\ A}{MS\ erreur}$
B	$\frac{MS\ B}{MS\ erreur}$	$\frac{MS\ B}{MS\ A \times B}$	$\frac{MS\ B}{MS\ A \times B}$
A \times B	$\frac{MS\ A \times B}{MS\ erreur}$	$\frac{MS\ A \times B}{MS\ erreur}$	$\frac{MS\ A \times B}{MS\ erreur}$

temps. Le dispositif expérimental le plus simple du modèle III est le dispositif en **blocs complets aléatoires** et c’est celui que nous verrons en détail dans la prochaine section.

3 Blocs complets aléatoires

Le dispositif en **blocs complets aléatoires** consiste en deux facteurs, l’un fixe, l’autre aléatoire, mais sans répétition. Avant d’aller plus loin, revisitons les dispositifs expérimentaux menant à certaines des analyses que nous avons vues dans les leçons précédentes.

3.1 Révision de l’échantillonnage

Certains dispositifs expérimentaux permettent d’échantillonner les données de façon complètement aléatoire. Par exemple, lorsqu’on applique le test t à un échantillon, on sélectionne les unités expérimentales aléatoirement parmi la population et on compare la moyenne à celle de l’hypothèse nulle (fig. 1).

Pour ce qui est du test t sur deux échantillons indépendants, on sélectionne les unités expérimentales et on attribue aléatoirement l’un des deux traitements aux unités (fig. 2). Dans l’ANOVA à un critère, les unités expérimentales sont sélectionnées aléatoirement et on attribue les traitements de façon complètement aléatoire (fig. 3).

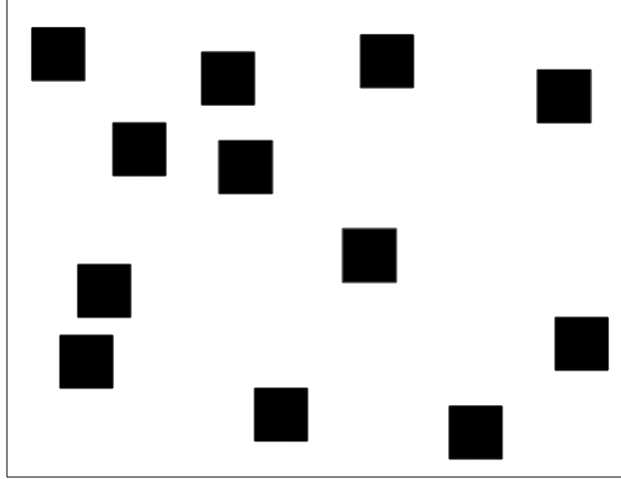


FIGURE 1 – Sélection aléatoire des unités expérimentales parmi la population pour le test t à un échantillon ($n = 12$).

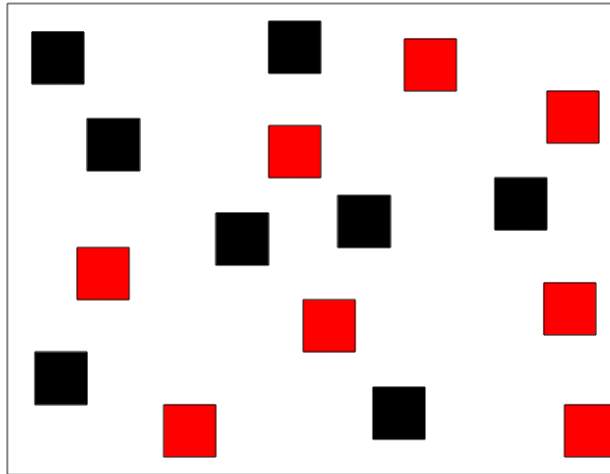


FIGURE 2 – Sélection aléatoire des unités expérimentales et attribution aléatoire des unités parmi la population pour le test t à deux groupes. À noter que les deux couleurs distinguent les deux traitements appliqués aux unités expérimentales ($n = 8$ répétitions par traitement).

D'autres dispositifs expérimentaux imposent une structure aux données : certaines unités expérimentales ne sont pas indépendantes. Par exemple, le test t pour données appariées traite les données par paire. On sélectionne aléatoirement les paires et on attribue aléatoirement l'un des traitements à chaque élément de la paire (fig. 4). Ce dispositif minimise la variabilité

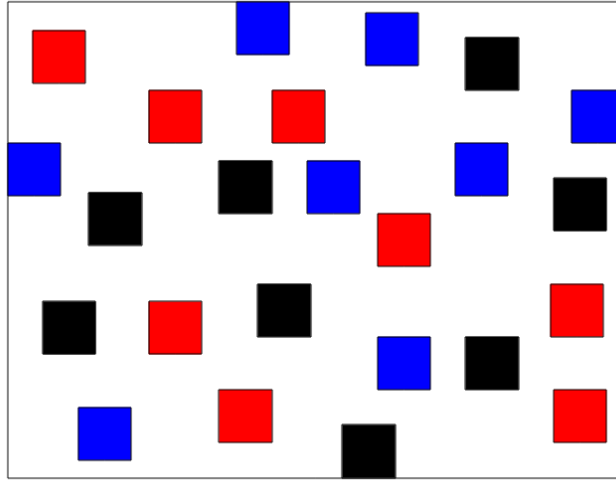


FIGURE 3 – Sélection aléatoire des unités expérimentales et assignation aléatoire de trois traitements dans un dispositif complètement aléatoire. À noter que chaque couleur correspond à un traitement appliqué à l'unité expérimentale ($n = 8$ répétitions par traitement).

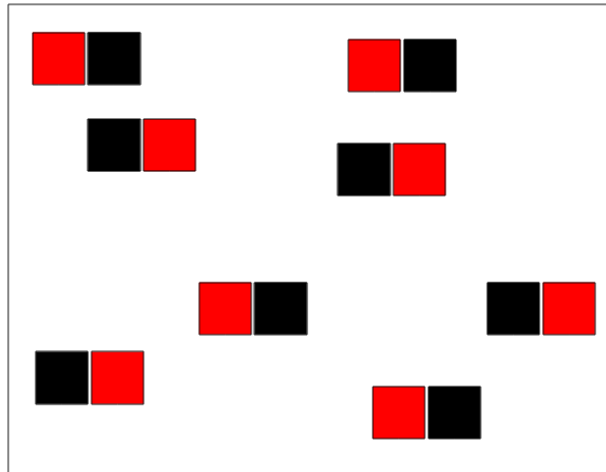


FIGURE 4 – Sélection aléatoire des paires d'unités expérimentales et assignation aléatoire de l'un des deux traitements. À noter que chaque couleur correspond à un des deux traitements appliqués à chaque unité expérimentale d'une paire ($n = 8$ paires).

intrapaire, mais permet une grande variabilité entre paires. Cette stratégie d'échantillonnage est avantageuse puisque, dans certains cas, il est logistiquement plus simple d'échantillonner des paires d'unités expérimentales que des unités expérimentales indépendantes. Par exemple,

le dispositif apparié présenté à la figure (4) nécessite huit paires d'unités expérimentales. En contrepartie, si on utilise un dispositif complètement aléatoire avec deux groupes comme pour le test t à deux groupes indépendants, on aura recours à 16 unités expérimentales indépendantes (fig. 2).

L'**ANOVA en blocs complets aléatoires** est l'extension du test t pour données appariées lorsqu'on a plus de deux traitements. Le **bloc** est l'analogue à la paire dans le test t : alors que la paire était constituée de deux observations liées entre elles, le bloc est constitué de plus de deux observations reliées entre elles (fig. 5). Tout comme le test t pour données appariées, le bloc minimise la variabilité à l'intérieur d'un bloc, mais la variabilité entre blocs est grande. En fait, l'ANOVA en blocs complets aléatoires effectuée sur des paires est identique au test t pour données appariées.

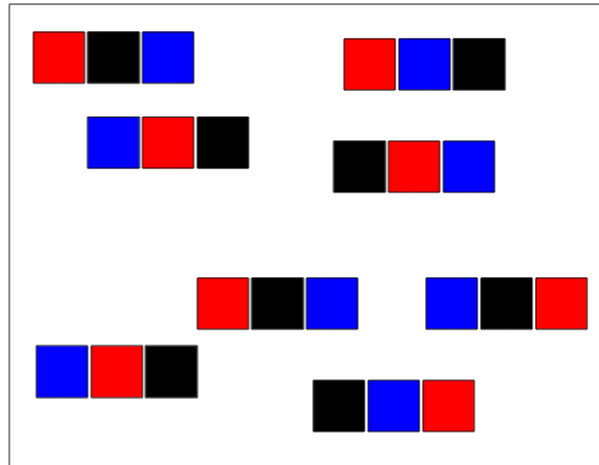


FIGURE 5 – Sélection aléatoire des blocs et assignation aléatoire de l'un des trois traitements aux unités expérimentales. À noter que chaque couleur correspond à l'un des trois traitements appliqués à chaque unité expérimentale d'un bloc ($n = 8$ blocs).

Dans la figure 5, l'ANOVA en blocs complets aléatoires utilise huit blocs pour tester l'effet de trois traitements, alors que l'ANOVA à un critère classique aurait nécessité 24 unités expérimentales pour tester ces mêmes effets (fig. 3). Étant donné ses avantages logistiques (p.ex., coût en temps et en argent), le dispositif en blocs complets aléatoires est beaucoup

utilisé en sciences.

3.2 Origine du dispositif en blocs complets aléatoires

Le dispositif en blocs complets aléatoires a été développé en agronomie pour les cas où un **gradient** est présent sur les parcelles étudiées. On entend par gradient, un ensemble de conditions telles que l'humidité, la température, le pH ou l'exposition au soleil, qui varient sur un site d'étude ou dans un laboratoire. Les blocs sont disposés perpendiculairement au gradient (fig. 6), de façon à ce que les conditions à l'intérieur d'un bloc soient homogènes.

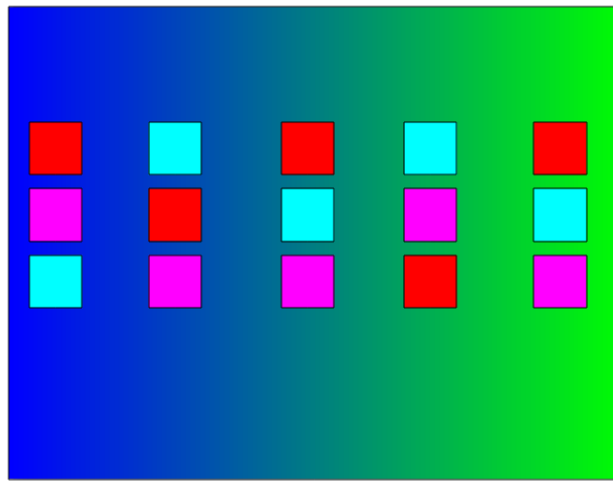


FIGURE 6 – Les blocs sont perpendiculaires au gradient (ici, le gradient va de gauche à droite). Cinq blocs ont été utilisés pour comparer trois traitements.

3.3 Particularité du dispositif en blocs complets aléatoires

Comme il est illustré aux figures 5 et 6, chaque bloc reçoit une seule répétition de chaque traitement, ce qui explique le terme « bloc complet » (chaque bloc reçoit tous les traitements). Les traitements sont assignés aléatoirement à chaque unité dans le bloc. Cette approche s'applique également en laboratoire lorsque l'espace est limité. Par exemple, on pourrait

considérer des essais dans une chambre de croissance² comme des blocs, où chaque essai consisterait à placer un plat de Petri pour chacun des traitements d'intérêt dans la chambre de croissance pendant une période donnée. Nous pourrions répéter la procédure plusieurs fois pour obtenir plusieurs essais. Comme autre exemple, on pourrait penser à des aquariums divisés en trois sections avec les mêmes densités de têtards de grenouilles, mais où chaque section recevrait différents prédateurs, afin d'estimer la survie des têtards. Ici, les aquariums représenteraient les blocs, puisque chaque aquarium donnerait trois observations (le nombre de survivants), soit une observation dans chaque section. Un dernier exemple serait une compagnie de raffinerie de pétrole qui s'intéresse à la performance en termes d'économie d'essence de 4 types d'essence qu'ils ont développés. Elle pourrait tester les 4 types d'essence sur 8 modèles de voitures sélectionnées aléatoirement.

Dans tous ces exemples, il y a donc toujours un facteur dont l'effet nous intéresse (effet des prédateurs sur les têtards, effet des types d'essence sur l'économie) et un facteur qui ne nous intéresse pas ou peu (effet des différents aquariums, effet des différents modèles de voitures). Le premier facteur représente l'effet fixe et le second représente l'effet aléatoire.

L'ANOVA en blocs complets aléatoires est en fait une ANOVA à deux critères sans répétitions : il n'y a qu'une observation pour chaque combinaison de traitement et de bloc. Le facteur d'intérêt est un facteur fixe, alors que le bloc peut être considéré comme un facteur aléatoire. On suppose que les blocs sont différents entre eux. Afin d'optimiser ce dispositif, il faut s'assurer que les conditions soient homogènes à l'intérieur d'un bloc donné, et que les blocs diffèrent entre eux. Ceci a pour effet de minimiser la variabilité à l'intérieur d'un bloc et de maximiser la variabilité entre les blocs ($\sigma_{\text{intra-bloc}}^2 < \sigma_{\text{inter-bloc}}^2$). Puisqu'il n'y a aucune répétition à l'intérieur d'un même bloc (p.ex., une observation par traitement), il est impossible de tester le terme d'interaction bloc \times traitement.

À l'instar de l'ANOVA à deux critères, le rejet de l'hypothèse nulle pour le facteur d'intérêt peut être suivi de comparaisons multiples pour déterminer où se trouvent les différences.

2. Une chambre de croissance est un contenant isotherme dans lequel les conditions sont contrôlées afin d'optimiser la croissance de plantes ou de cultures.

Toutefois, si le facteur bloc explique une bonne partie de la variance et l'hypothèse nulle est rejetée pour le bloc (c.-à-d., $P < \alpha$), nous ne testons pas les différences entre blocs. Le facteur bloc n'est pas directement d'intérêt et il reflète simplement le dispositif qui a été utilisé. Le prochain exemple illustre l'application de l'ANOVA à ce dispositif expérimental.

Exemple 8.4 On veut déterminer l'effet de quatre diètes différentes sur le gain de masse en grammes de cochons d'Inde. Chaque individu est placé dans une cage individuelle située dans un grand entrepôt. Étant donné qu'il n'y a qu'une source de chaleur à une extrémité de l'entrepôt, on pourrait s'attendre à avoir un gradient de température. On dispose les cages en blocs complets aléatoires de façon à avoir des groupes de quatre cages à différentes distances de la source de chaleur. On trouve ces données dans le fichier `cochons.txt`.

```
> ##importation
> cochons <- read.table("cochons.txt", header = TRUE)
> ##premières observations
> head(cochons)

  Bloc Diete Masse
1     1     1  1.5
2     1     2  2.7
3     1     3  2.1
4     1     4  1.3
5     2     1  1.4
6     2     2  2.9

> ##structure du jeu de données
> str(cochons)

'data.frame':      20 obs. of  3 variables:
```

```
$ Bloc : int  1 1 1 1 2 2 2 2 3 3 ...
$ Diete: int  1 2 3 4 1 2 3 4 1 2 ...
$ Masse: num  1.5 2.7 2.1 1.3 1.4 2.9 2.2 1 1.4 2.1 ...
```

Dans notre expérience, les deux facteurs sont **Bloc** et **Diete**. Toutefois, R considère ces deux variables comme numérique (des entiers). À chaque fois qu'un facteur est codé avec des valeurs numériques, comme ici 1, 2, 3, 4, il faut recoder soi-même la variable en facteur.

```
> ##convertir en facteur
> cochons$Bloc <- as.factor(cochons$Bloc)
> cochons$Diete <- as.factor(cochons$Diete)
> ##reconnus maintenant comme facteurs
> str(cochons)

'data.frame':      20 obs. of  3 variables:
 $ Bloc : Factor w/ 5 levels "1","2","3","4",...: 1 1 1 1 2 2 2 2 3 3 ...
 $ Diete: Factor w/ 4 levels "1","2","3","4": 1 2 3 4 1 2 3 4 1 2 ...
 $ Masse: num  1.5 2.7 2.1 1.3 1.4 2.9 2.2 1 1.4 2.1 ...
```

On peut visualiser les données à l'aide d'un diagramme de boîtes et moustaches (fig. 7).

```
> boxplot(Masse ~ Diete, data = cochons, xlab = "Diète",
          ylab = "Gain de massse (g)")
```

Pour réaliser l'ANOVA à deux critères sans répétition, on utilise encore une fois `aov()` :

```
> ##ANOVA à deux critères sans répétition
> bloc.aov <- aov(Masse ~ Diete + Bloc, data = cochons)
```

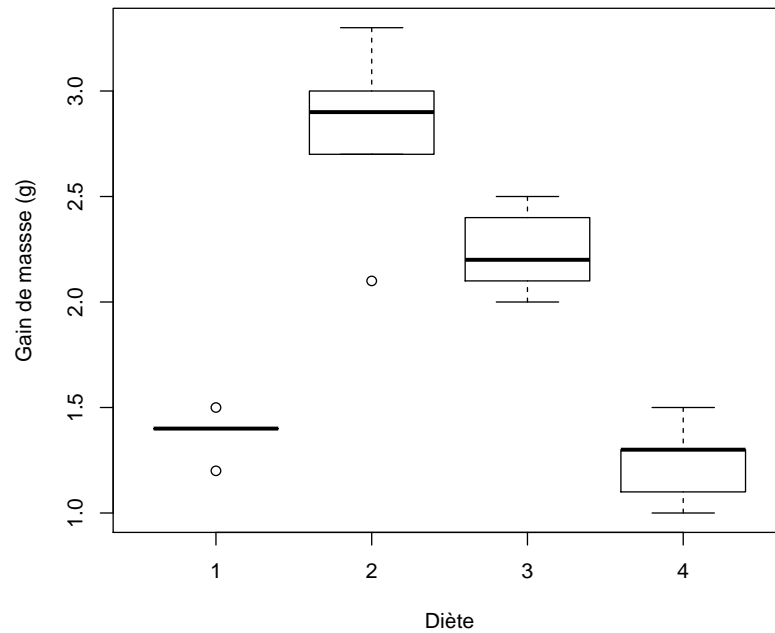


FIGURE 7 – Diagramme de boîtes et moustaches à partir des données de gain de masse de cochons d’Inde.

Avant d’interpréter les résultats de l’ANOVA, nous devons vérifier les suppositions habituelles.

```
> ##vérification des suppositions
> par(mfrow = c(2, 2))
> plot(bloc.aov)
```

On remarque que les variances sont légèrement hétérogènes (fig. 8). Essayons une transformation logarithmique afin d’homogénéiser les variances :

```
> ##ANOVA à deux critères sans répétition
> ##en utilisant le logarithme de la Masse
> bloc.aov.log <- aov(log(Masse) ~ Diete + Bloc, data = cochons)

> ##vérification des suppositions
> par(mfrow = c(2, 2))
```

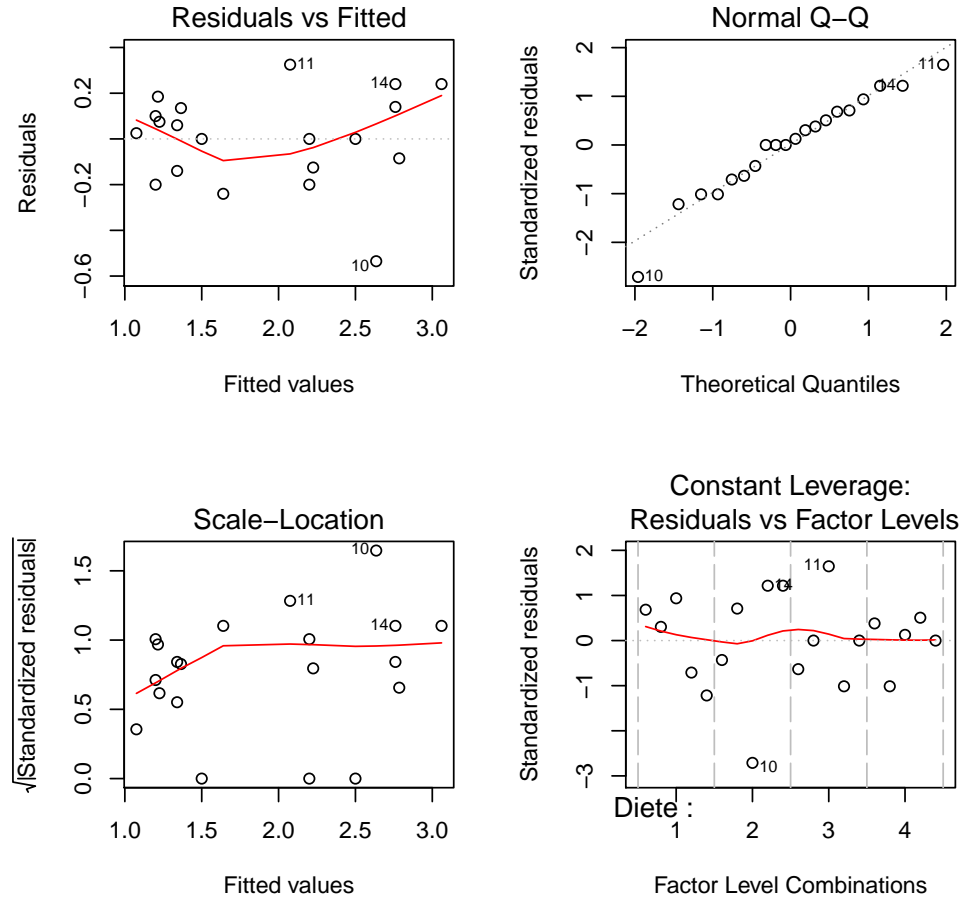


FIGURE 8 – Vérifications des suppositions d’homogénéité des variances et de la normalité des résidus.

```
> plot(bloc.aov.log)
```

On observe que la transformation logarithmique a permis de mieux respecter les suppositions (fig. 9).

Comparons maintenant l’ANOVA en blocs complets aléatoires et l’ANOVA à un critère avec ces mêmes données.

```
> ##ANOVA à 1 critère
> aov1.log <- aov(log(Masse) ~ Diete, data = cochons)
> ##comparaisons ANOVA à 1 critère vs blocs
```

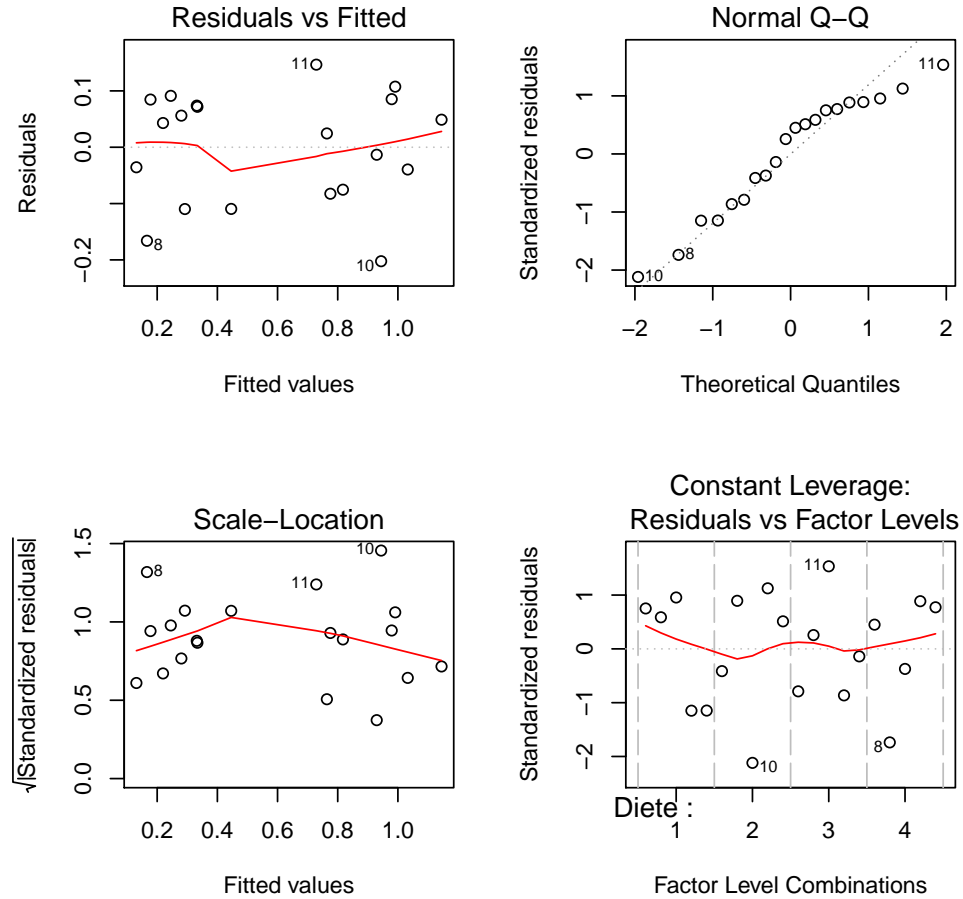


FIGURE 9 – Vérifications des suppositions d’homogénéité des variances et de la normalité des résidus avec les données log-transformées.

```
> summary(aov1.log)

              Df Sum Sq Mean Sq F value    Pr(>F)
Diete           3  2.2515   0.7505   43.12 6.83e-08 ***
Residuals       16  0.2784   0.0174
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> summary(bloc.aov.log)
```


	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Diete	3	2.2515	0.7505	49.363	5.03e-07 ***
Bloc	4	0.0960	0.0240	1.579	0.243
Residuals	12	0.1824	0.0152		

Signif. codes:

0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

On remarque que la somme des carrés du facteur **Diete** est la même pour les deux analyses (2.251). La différence entre l'ANOVA à un critère et l'ANOVA en blocs complets aléatoires est au niveau du terme d'erreur. Dans l'ANOVA en blocs complets aléatoires, la somme des carrés des erreurs (SSE) est de 0.182, alors que celle de l'ANOVA à un critère est de 0.278. On constate que la somme des carrés du terme **Bloc** est de 0.096 et que cette valeur ajoutée à la somme de carrés des erreurs de l'ANOVA en blocs complets aléatoires donne un total de 0.278 ($0.096 + 0.182 = 0.278$).

Le terme **Bloc** explique une partie de la variabilité de la variable réponse (**Masse**) et réduit le terme d'erreur. Ceci a pour conséquence de réduire le carré moyen de l'erreur (MSE). Puisque le F est le ratio entre le carré moyen de **Diete** et le carré moyen de l'erreur, la valeur du F est plus grande en présence du bloc.

On constate que la diète influence fortement le gain de masse ($F_{3,12} = 49.36$, $P < 0.0001$) et que les blocs ne diffèrent pas entre eux ($F_{4,12} = 1.58$, $P = 0.24$). On peut procéder à des comparaisons multiples afin d'identifier les groupes définis par la diète.

```
> ##moyennes des groupes
> tapply(X = cochons$Masse, INDEX = cochons$Diete, FUN = mean)

1      2      3      4
```

```
1.38 2.80 2.24 1.24
```

```
> ##comparaisons multiples
```

```
> TukeyHSD(x = bloc.aov.log, which = "Diete")
```

```
Tukey multiple comparisons of means
```

```
95% family-wise confidence level
```

```
Fit: aov(formula = log(Masse) ~ Diete + Bloc, data = cochons)
```

```
$Diete
```

	diff	lwr	upr	p adj
2-1	0.6990462	0.4675197	0.93057281	0.0000060
3-1	0.4836196	0.2520930	0.71514616	0.0002317
4-1	-0.1143399	-0.3458665	0.11718666	0.4856391
3-2	-0.2154267	-0.4469532	0.01609992	0.0712144
4-2	-0.8133862	-1.0449127	-0.58185959	0.0000012
4-3	-0.5979595	-0.8294861	-0.36643294	0.0000299

Les comparaisons multiples identifient deux groupes, le premier est constitué des diètes 4 et 1, et le deuxième est composé des diètes 2 et 3 :

4	1	3	2
1.24	1.38	2.24	2.80

On peut présenter les résultats graphiquement.

```
> ##Définissons l'axe des x de la figure, c'est à dire les diètes
```

```
> diete <- 1:4
```

```
> ##Puis, calculons les moyennes sur les donnees transformees
```

```

> means.log <- tapply(X = log(cochons$Masse), INDEX = cochons$Diete, FUN = mean)
> ##Il nous faut maintenant trouver le MSE.
> ##Pour se faire, nous devons nous référer au résumé de l'ANOVA
> summary(bloc.aov.log)

              Df Sum Sq Mean Sq F value    Pr(>F)
Diete           3  2.2515   0.7505   49.363 5.03e-07 ***
Bloc            4  0.0960   0.0240    1.579    0.243
Residuals      12  0.1824   0.0152
---
Signif. codes:
  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> ##Le MSE est la valeur données à la ligne Residuals de la colonne Mean Sq
> MSE <- 0.0152
> ##On peut aussi le trouver de cette façon:
> S<-summary(bloc.aov.log)
> MSE<-S[[1]][[3]][[3]]
> ##Calculons maintenant les barres d'erreur
> low.log <- means.log - sqrt(MSE)
> upp.log <- means.log + sqrt(MSE)
> ##On affiche les moyennes
> plot(means.log ~ diete, ylab = "Log du gain de masse (g)", xlab = "Diète",
       type = "p", ylim = c(min(low.log), max(upp.log)),
       xlim = c(0, 5), cex.lab = 1.2, xaxt = "n",
       main = "Log des Moyennes ± racine carrée de MSE")
> ##Ajoutons l'axe des x
> axis(side = 1, at = c(1, 2, 3, 4))
> ##Ajoutons les barres d'erreur

```

```

> arrows(x0 = diete, x1 = diete, y0 = low.log, y1 = upp.log,
         angle = 90, code = 3, length = 0.05)

> ##Ajoutons les lettres qui designent les groupes

> text(x = 1, y = 0.18, labels = "a", cex = 1.2)
> text(x = 4, y = 0.07, labels = "a", cex = 1.2)
> text(x = 2, y = 0.88, labels = "b", cex = 1.2)
> text(x = 3, y = 0.66, labels = "b", cex = 1.2)

```

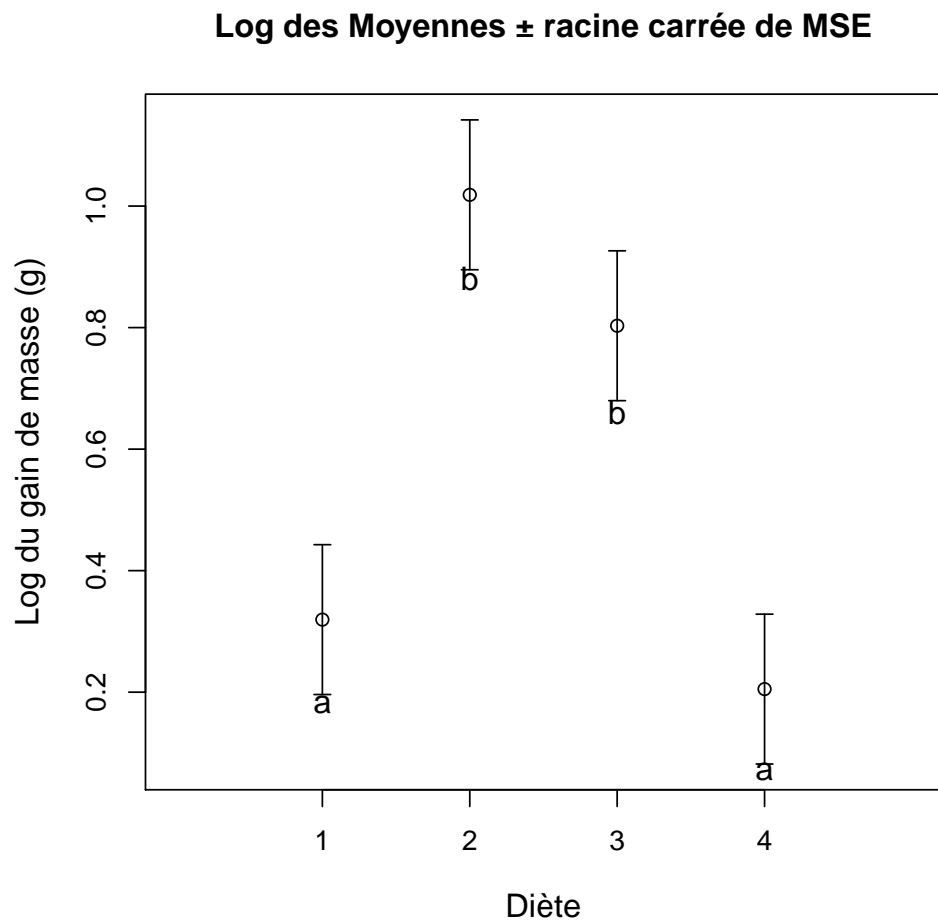


FIGURE 10 – Effet de la diète sur le logarithme du gain de masse. Les lettres différentes indiquent les groupes.

Pour faciliter l'interprétation de notre analyse représentons les données originales plutôt que les données transformées.

```

> ##Calculons les moyennes sur les donnees originales
> means <- tapply(X = cochons$Masse, INDEX = cochons$Diete, FUN = mean)
> ##Les barres d'erreurs sur les donnees originales se determinent en prenant la
> ## sur les barres d'erreurs des donnees transformees
> low <- exp(low.log)
> upp <- exp(upp.log)
> ##On cree la figure
> plot(means ~ diete, ylab = "Gain de masse (g)", xlab = "Diète",
       type = "p", ylim = c((min(low)-0.2), max(upp)),
       xlim = c(0, 5), cex.lab = 1.2, xaxt = "n",
       main = "Moyennes ± racine carrée de MSE")
> ##Ajoutons l'axe des x
> axis(side = 1, at = c(1, 2, 3, 4))
> ##Ajoutons les barres d'erreur
> arrows(x0 = diete, x1 = diete, y0 = low, y1 = upp,
       angle = 90, code = 3, length = 0.05)
> ##Ajoutons les lettres qui designent les groupes
> text(x = 1, y = 1.19, labels = "a", cex = 1.2)
> text(x = 4, y = 1.0, labels = "a", cex = 1.2)
> text(x = 2, y = 2.4, labels = "b", cex = 1.2)
> text(x = 3, y = 1.9, labels = "b", cex = 1.2)

```

Sur la figure 11, on constate la présence de deux groupes de diètes. Le premier groupe est composé des diètes 4 et 1, et le deuxième est constitué des groupes 3 et 2. De plus, on peut conclure que les diètes 4 et 1 augmentent moins le gain de masse que les diètes 3 et 2.

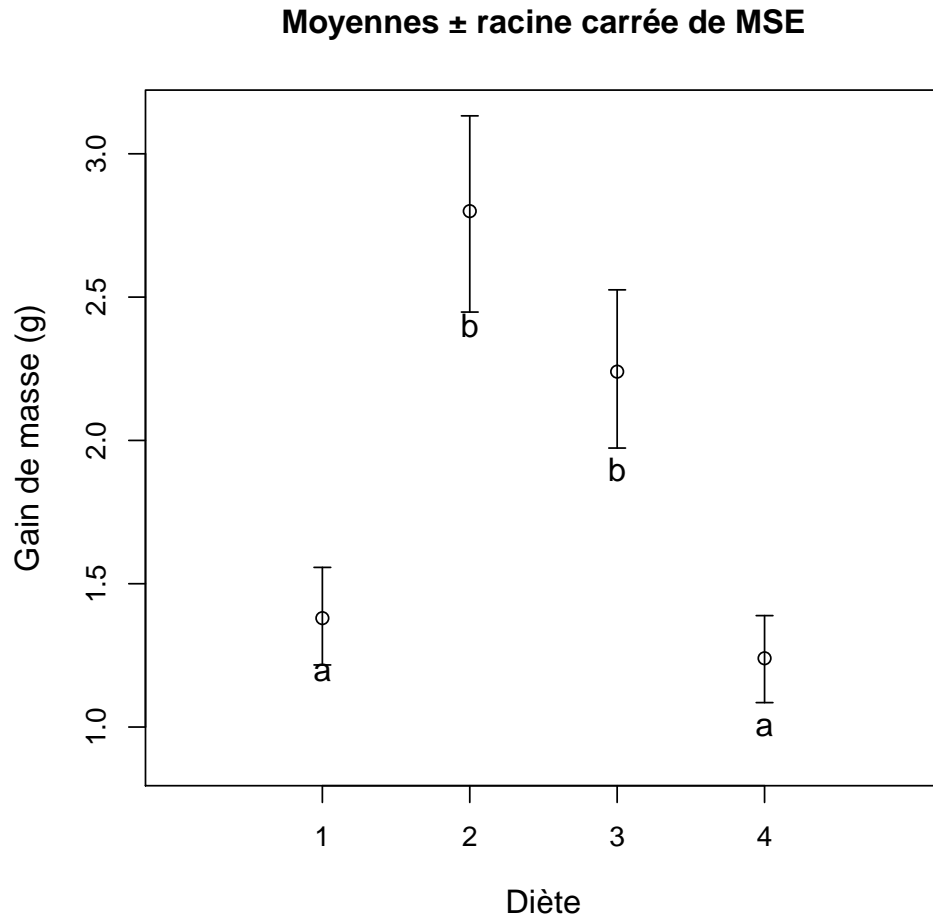


FIGURE 11 – Effet de la diète sur le gain de masse. Les lettres différentes indiquent les groupes.

3.4 Applications moins conventionnelles

Typiquement en agronomie ou en écologie forestière, les blocs sont disposés dans l'espace et perpendiculaires à un gradient. Toutefois, on peut aussi disposer les blocs dans le temps. Par exemple, considérons une expérience de germination de graines en serre (croissance en une semaine), où l'on ne peut pas répéter tous les traitements d'intérêt dû à un manque d'espace. Pour pallier le problème, on peut créer des blocs temporels de façon à avoir une répétition de chaque traitement et à effectuer un bloc par semaine.

Dans d'autres cas, on peut avoir plusieurs mesures sur un même individu. L'individu

constituera le bloc. Pour illustrer ce concept, imaginons une expérience sur la production de lait par des vaches avec différents types de fourrage (p.ex., luzerne, trèfle, blé, maïs). Si on soupçonne que la production de lait varie beaucoup entre les vaches, on pourrait traiter chaque vache comme un bloc. Chaque vache recevrait chacun des traitements, dans une séquence aléatoire et avec suffisamment de temps entre les traitements pour éviter de contaminer les réponses des vaches par l'effet du fourrage précédent. De cette façon, on peut mieux contrôler les différences individuelles en comparant la réponse d'un même individu à différents traitements.

Un autre exemple serait de suivre les résultats d'élèves à différentes périodes lors desquelles ils auraient utilisé différentes méthodes d'apprentissage. Chaque élève représenterait un bloc complet et utiliserait toutes les méthodes à travers le temps dans un ordre aléatoire. De cette façon, on peut contrôler pour la grande variabilité inter-élève et se concentrer sur l'effet des méthodes d'apprentissage.

3.5 Conditions d'application

En plus des conditions nécessaires à l'ANOVA classique (homogénéité des variances, normalité des résidus), il faut respecter les conditions suivantes :

- chaque bloc doit être suffisamment petit pour assurer des conditions homogènes à l'intérieur du bloc ;
- chaque bloc doit être suffisamment grand pour accommoder une répétition de chaque traitement : il doit y avoir assez de place pour que les répétitions de chaque traitement soient espacées afin d'être indépendantes. En d'autres mots, l'observation provenant d'un traitement dans un bloc ne doit pas être influencée par un traitement adjacent du même bloc (p.ex., l'application de différentes concentrations d'engrais à des parcelles d'un bloc et le ruissellement contaminant le traitement d'une parcelle adjacent du même bloc) ;
- les blocs doivent être suffisamment espacés afin d'être indépendants les uns des autres ;

- il ne doit y avoir aucune interaction entre les blocs et les traitements (effets additifs seulement) : l'effet d'un traitement doit être constant pour tous les blocs.

3.6 Avantages du dispositif en blocs complets aléatoires

Le dispositif en blocs complets aléatoires comporte plusieurs avantages, notamment :

- il contrôle l'hétérogénéité environnementale (tient compte des gradients) qui autrement ferait augmenter le carré moyen des erreurs (MSE) et rendrait plus difficile de rejeter l'hypothèse nulle concernant le facteur d'intérêt ;
- en présence de gradients, l'ANOVA en blocs complets aléatoires est plus puissante que l'ANOVA complètement aléatoire ;
- il est logistiquement intéressant lorsque la répétition est contrainte dans l'espace ou le temps ;
- il peut être utilisé avec un design apparié (p.ex., > 2 observations par bloc) pour tenir compte de facteurs additionnels tels que la variabilité entre les individus d'une même famille.

3.7 Désavantages du dispositif en blocs complets aléatoires

Le dispositif en blocs complets aléatoires comporte certains désavantages, notamment :

- il y a un coût statistique à utiliser les blocs. Si l'effectif (n) est petit et si l'effet du bloc est faible, l'ANOVA à un critère est plus puissante dans ce cas.
- avec des blocs trop petits, on risque d'introduire une dépendance des observations soumis aux différents traitements d'un même bloc, si les traitements sont physiquement près les uns des autres ;
- il faut un dispositif complètement équilibré : si on a au moins une donnée manquante dans un bloc, on doit se débarrasser du bloc ou estimer les valeurs des données manquantes³ ;

3. Différentes techniques variant en complexité permettent d'estimer des valeurs manquantes, mais elles vont au-delà des objectifs du cours.

- s'il y a une interaction entre les blocs et les traitements (effets non-additifs), on a un problème puisqu'on ne peut tester l'interaction en l'absence de répétitions. Si on soupçonne un potentiel effet interactif, il faut prévoir récolter plus d'une observation d'un même traitement dans chaque bloc pour le tester formellement. Le cas échéant, on utilisera une analyse plus complexe, telle que l'ANOVA en tiroirs (*split-plot*)⁴. Une autre option est de mesurer directement la variable qui est responsable du gradient (p.ex., humidité, ensoleillement, position dans la pente).

Conclusion

Dans cette leçon, nous avons revisité différents types de dispositifs associés aux analyses vues jusqu'à présent dans le cours. Nous avons présenté les concepts d'effets fixes et d'effets aléatoires ainsi que de modèle mixte. Nous avons présenté l'ANOVA en blocs complets aléatoires comme étant le cas le plus simple de modèle mixte et comme l'extension du test t pour les données appariées. Le dispositif en blocs complets aléatoires se caractérise par la disposition de blocs perpendiculaires à un gradient. Ce gradient n'est pas mesuré directement, mais les blocs sont une mesure indirecte de la variabilité associée à ce gradient. Ce dispositif comporte des avantages importants, notamment sur le plan logistique comparé aux dispositifs complètement aléatoires (p.ex., ANOVA à un critère avec répétitions, ANOVA à deux critères avec répétitions). Bien qu'à l'origine le dispositif en blocs complets aléatoires ait été développé en agronomie, il est maintenant utilisé dans des disciplines variées incluant les sciences environnementales, la médecine, la chimie et le marketing. La prochaine leçon traitera de façon plus générale de l'élaboration de dispositifs expérimentaux et de recommandations afin d'arriver à des conclusions solides.

4. Ce dispositif plus complexe ne sera pas vu dans le cours.

Index

ANOVA à deux critères, [4](#)

ANOVA à un critère, [4](#), [5](#)

blocs complets aléatoires, [5](#)

 autres applications, [21](#)

 avantages, [23](#)

 comparaisons multiples, [10](#)

 conditions d'application, [22](#)

 désavantages, [23](#)

 dispositif, [9](#)

 gradient, [9](#), [11](#)

 origine, [9](#)

vs test t pour données appariées, [8](#)

échantillonnage, [5](#)

effets aléatoires, [2](#)

effets fixes, [2](#)

modèle I, [4](#)

modèle II, [4](#)

modèle III, [4](#)

modèle mixte, [4](#)

test t pour données appariées, [6](#)

test t sur deux échantillons indépendants, [5](#)

test t sur un échantillon, [5](#)