



Dong Juan

Master 2 de Mathématiques, parcours :  
Modélisation mathématique et analyse statistique  
Université de Poitiers

---

Rapport de stage de fin de master 2  
Février 2013 – Juin 2013

---

**Intitulé du stage :** Modèles linéaires mixtes et analyses multi-factorielles pour l'analyse des données d'un dispositif de recherche agronomique visant à détecter des variétés de ray-grass anglais économes en azote et en eau

**Lieu du stage :** Centre INRA Poitou-Charentes, Unité de Recherche Pluridisciplinaire Prairies et Plantes Fourragères, 86600 Lusignan

**Maître de stage :** Jean-Paul Sampoux, ingénieur de recherche, animateur de l'équipe génétique de l'URP3F à Lusignan, responsable du Centre de Ressources Génétiques des espèces fourragères et à gazon hébergé par l'UR P3F.

**Tuteurs :** Anthony Phan, Maître de Conférences, Julien Michel, Maître de Conférences.

## Table des matières

1.Introduction.....	5
2.Présentation de l'INRA.....	6
2.1.L'INRA, un acteur majeur de la recherche agronomique dans le monde.....	6
2.2.Le centre INRA Poitou-Charentes.....	8
2.3.Présentation de l'URP3F de Lusignan.....	8
3.Présentation du projet et du stage.....	10
3.1.Le contexte du stage.....	10
3.2.Les objectifs du projet.....	10
L'indice de nutrition azotée (INN).....	11
L'indice de la conductance stomatique.....	11
3.3.Présentation de l'étude.....	11
3.4.Les objectifs du stage.....	12
3.5.Mise en place des essais et présentation des données.....	13
Matières azotées totales : mat.....	15
Le rendement en matière sèche : rdtsec.....	15
4.Quelques remarques sur le logiciel R.....	17
4.1.R et l'INRA.....	17
4.2.Programmer avec R.....	18
4.3.Les principales fonctions de R utilisées pendant le stage.....	18
a. La fonction « lm » (base).....	18
b. La fonction « lmer » (bibliothèque : lme4) .....	18
c. La fonction Anova (bibliothèque car).....	19
Analyses de variance de type I , II ou III.....	19
Quelques remarques concernant la fonction Anova (bibliothèque « car »).....	19
d. La fonction lsmeans de la bibliothèque lsmeans.....	20
e. La fonction SNK.test de la bibliothèque agricolae.....	20
f. ACP et R.....	20
4.4.Ecriture des modèles.....	20
5.Analyse de la variance des données et comparaison des moyennes.....	21
5.1.Première série d'analyses des données par lieu et par date.....	21
a. Vérification des conditions.....	21
b. Construction des modèles et explications.....	21
c. Interprétation des résultats.....	22
d. Retour sur la question de la vérification des hypothèses.....	24
5.2.Deuxième série de tests d'analyses des données par lieu.....	26
5.3.Troisième série d'analyses des données, tous lieux confondus .....	29
a. Facteurs hiérarchisés.....	30

b. Construction des modèles et explications.....	31
5.4.Comparaisons multiples et calcul des lsmeans.....	34
a. La méthode de la plus petite différence significative.....	35
b. La méthode de Newman et Keuls.....	35
c. Calcul des moyennes ajustées.....	36
d. Les résultats.....	37
e. Remarque sur la fonction SNK.test.....	41
6.ACP.....	42
6.1.Choix des variables.....	42
6.2.ACP et R.....	46
6.3.Résultats.....	46
a. ACP par lieu (voir annexe 8).....	46
b. ACP rdtsec/INN et rdtsec/mat sur plusieurs lieux (annexe 9).....	47
c. ACP multilocale (annexe 10).....	48
.....	50
7.Conclusion.....	51
8.Bibliographie/webographie.....	52
9.Annexes.....	53
9.1.Annexe 1 : le fichier « initialisation.R ».....	53
9.2.Annexe 2 : analyse de la base de donnée.....	55
9.3.Annexe 3 : les fonctions « analyse1() » et « analyse2() » et traitement.....	57
9.4.Annexe 4 : test des résidus de la première série d'analyse de variance .....	60
9.5.Annexe 5 : la fonction « analyse3() » et traitement.....	62
9.6.Annexe 6 : analyse de la variance dans le cas multilocal (rdtsec, mat et INN ).....	66
9.7.Annexe 7 : analyse de la variance dans le cas multilocal (deltaC13, teneurN et teneurC).....	74
9.8.Annexe 8 : ACP par lieu.....	80
9.9.Annexe 9 : ACP rdtsec/INN et rdtsec/mat.....	82
9.10.Annexe 10 : ACP multilocale.....	83

## Remerciements

Je tiens à remercier dans un premier temps, toute l'équipe pédagogique de l'université de mathématiques de Poitiers qui m'a apporté durant deux années la formation théorique sans laquelle je n'aurais pas pu faire ce stage dans de bonnes conditions.

Je remercie également toutes les personnes de l'URP3F à Lusignan qui m'ont apporté de l'aide durant ces quatre mois passés à l'INRA : Stéphane Fourtier (AI) et Jean-François Bourcier (TR) qui m'ont aidé dans la partie « base de données », Rémi Dupuis (TR) qui m'a fait visiter l'INRA, les serres et les champs, Jérôme Chargelègue (TR) qui m'a fait visiter les laboratoires de l'INRA, Vincent Furstoss formateur sur R et Nathalie Bonnet qui m'a aidé dans les démarches administratives.

Je remercie également Laurent Delineau, professeur de mathématiques au Lycée Pilote Innovant de Jaunay-Clan pour son aide dans la rédaction de ce rapport.

Je remercie enfin Monsieur Sampoux, Ingénieur de recherche et animateur de l'équipe génétique de l'URP3F à Lusignan pour son aide et ses conseils précieux tout au long de ce stage, ainsi que pour la patience dont il a du faire preuve pour répondre à toutes mes questions.

## 1. Introduction

Quand j'ai entrepris de faire un master dans le domaine des statistiques il y a deux ans, mon but était de pouvoir aller plus loin dans les mathématiques et surtout de pouvoir utiliser mes connaissances dans ce domaine comme un outil au service d'un travail plus concret. Le master sur la modélisation mathématique et l'analyse statistique, qui forme des ingénieurs mathématiciens polyvalents, convenait parfaitement à ce que j'attendais : utiliser des outils mathématiques pour résoudre des problèmes réels.

En deuxième année de master, un professeur de l'université m'a parlé d'une proposition de stage à L'INRA. Arrivée en France il y a deux ans, je ne connaissais pas cet organisme et encore moins le domaine agronomique. Mais, en lisant l'intitulé du stage, j'ai très vite compris qu'il correspondait parfaitement à ce que j'attendais : approfondir mes connaissances en statistiques et m'en servir pour répondre à des questions concrètes.

Le premier entretien avec Monsieur Sampoux a confirmé cette impression. Mais, ne connaissant rien au domaine de l'agronomie, c'est avec une certaine appréhension que j'ai débuté ce stage. Heureusement, Monsieur Sampoux et son équipe m'ont très bien accueilli et guidé dans mes premiers pas à l'INRA.

Ensuite, les choses se sont enchaînées, le travail proprement dit a commencé, les difficultés aussi. Mais c'est toujours avec beaucoup d'enthousiasme du premier jour au dernier jour que j'ai effectué ce stage. Dans ce rapport, après avoir rapidement présenté l'INRA, j'exposerai mon travail durant ce stage, mes missions et tâches à effectuer, les résultats obtenus, les difficultés rencontrées et les solutions apportées.

## 2. Présentation de l'INRA

### 2.1. L'INRA, un acteur majeur de la recherche agronomique dans le monde

L'Institut National de la Recherche Agronomique (INRA) est un organisme français de recherche en agronomie fondé en 1946, placé sous la double tutelle du ministère chargé de la Recherche et du ministère chargé de l'Agriculture.

L'INRA est un acteur majeur de la recherche agronomique dans le monde. Le nombre d'articles scientifiques publiés chaque année en sciences agricoles et en sciences de la plante et de l'animal le place depuis plusieurs années dans le top des institutions européennes et mondiales comme le montre le tableau ci-dessous :

Champ disciplinaire	D'après le nombre de citations reçues		D'après le nombre de papiers	
	Rang mondial	Rang français	Rang mondial	Rang français
Agronomie	2ème	1er	3ème	1er
Biologie végétale et animale	3ème	1er	5ème	1er
Microbiologie	18ème	3ème	13ème	3ème
Environnement / Écologie	34ème	2ème	15ème	1er

Positions mondiales et européennes de l'INRA (chiffres 2011)<sup>1</sup>

Les grandes missions confiées à l'INRA sont les suivantes :

- Produire et diffuser des connaissances scientifiques ;
- Concevoir des innovations et des savoir-faire pour la société ;
- Éclairer, par son expertise, les décisions des acteurs publics et privés en apportant notamment un avis éthique et déontologique ;
- Développer la culture scientifique et technique et participer au débat science-société ;
- Former à la recherche.

En 2013, L'INRA emploie près de 8 500 agents titulaires : 20 % de chercheurs dans le domaine des sciences du vivant, des sciences de la matière et des sciences humaines, 30 % d'ingénieurs et 50 % de techniciens et administratifs.

Par ailleurs chaque année l'INRA accueille également un grand nombre de stagiaires (environ 2400

<sup>1</sup> <http://institut.inra.fr/Reperes/Chiffres>

en 2011) et des doctorants (plus de 500 en 2011) ainsi que de nombreux chercheurs étrangers.

Pour répondre à ses missions, l'INRA a structuré ses activités autour de six « métaprogrammes » :

- Adaptation de l'agriculture et de la forêt au changement climatique
- Déterminants et impact de la diète, interactions et transitions
- Gestion durable de la santé des cultures
- Gestion intégrée de la santé des animaux
- Méta-omiques des écosystèmes microbiens
- Sélection génomique

Ces « métaprogrammes » sont développés au sein de « départements scientifiques ». Les activités au sein de l'INRA sont ainsi organisées en 13 « départements scientifiques » :

- Environnement et agronomie
- Écologie des forêts, prairies et milieux aquatiques
- Alimentation humaine
- Caractérisation et élaboration des produits issus de l'agriculture
- Génétique animale
- Biologie et amélioration des plantes
- Mathématiques et informatique appliquées
- Microbiologie et chaîne alimentaire
- Physiologie animale et systèmes d'élevage
- Santé animale
- Santé des plantes et environnement
- Sciences pour l'action et le développement
- Sciences sociales, agriculture et alimentation, espace et environnement

Ces départements scientifiques sont répartis dans **18 centres régionaux** comprenant en tout, plus de **150 sites**.



Carte des centres INRA en France métropolitaine

## 2.2. Le centre INRA Poitou-Charentes

Le centre INRA Poitou-Charentes étudie la gestion durable des prairies, des systèmes fourragers et des territoires, et celle des productions animales.

Ce centre compte 240 techniciens, ingénieurs et chercheurs, s'étend sur 600 hectares et compte plus de 60 000 m<sup>2</sup> de bâtiments.

Il regroupe dix unités (deux unités de recherche, sept unités expérimentales et une unité d'appui à la recherche) réparties sur quatre sites géographiques (Lusignan-Rouillé dans la Vienne, le Magneraud et Saint Laurent-de-la-Prée en Charente-Maritime, Chizé dans les Deux-Sèvres) :

- sept unités INRA à Lusignan, Le Magneraud et Saint Laurent-de-la-Prée
- une unité sous contrat INRA/CNRS à Chizé
- deux unités du Groupement d'intérêt public Geves (Groupe d'étude et de contrôle des variétés et des semences) au Magneraud.

Parmi ces unités, figure l'URP3F<sup>2</sup> (Unité de Recherche Pluridisciplinaire Prairies et Plantes Fourragères) implantée à Lusignan.

## 2.3. Présentation de l'URP3F<sup>3</sup> de Lusignan

Cette unité rassemble 51 agents permanents ainsi que plusieurs stagiaires et doctorants. Elle dispose des moyens de travail suivants :

- un Centre de Ressources Génétiques (espèces fourragères et à gazon)
- un laboratoire de biochimie pour évaluer la valeur alimentaire des fourrages, la composition de mélanges prairiaux (mesure des teneurs en protéines, sucres, cendres, fractions des parois ligno-cellulosiques) et la composition élémentaire en carbone, azote et leurs isotopes,
- un laboratoire de biotechnologies (extraction d'ADN, génotypage, ...)
- des installations expérimentales diverses : parcelles, pépinières, serres et chambres de culture, chambres froides pour conserver les graines,
- des équipements d'écophysiologie (équipements de caractérisation du milieu, de l'architecture des plantes, de l'acquisition et de l'allocation des ressources dans les plantes),

L'URP3F est composée de deux équipes de recherche (une équipe en écophysiologie et une équipe en génétique) ainsi que de trois équipes techniques transversales.

C'est au sein de l'équipe de génétique que j'ai effectué mon stage de fin de master. Cette équipe est animée par Monsieur Jean-Paul Sampoux, Ingénieur de recherche.

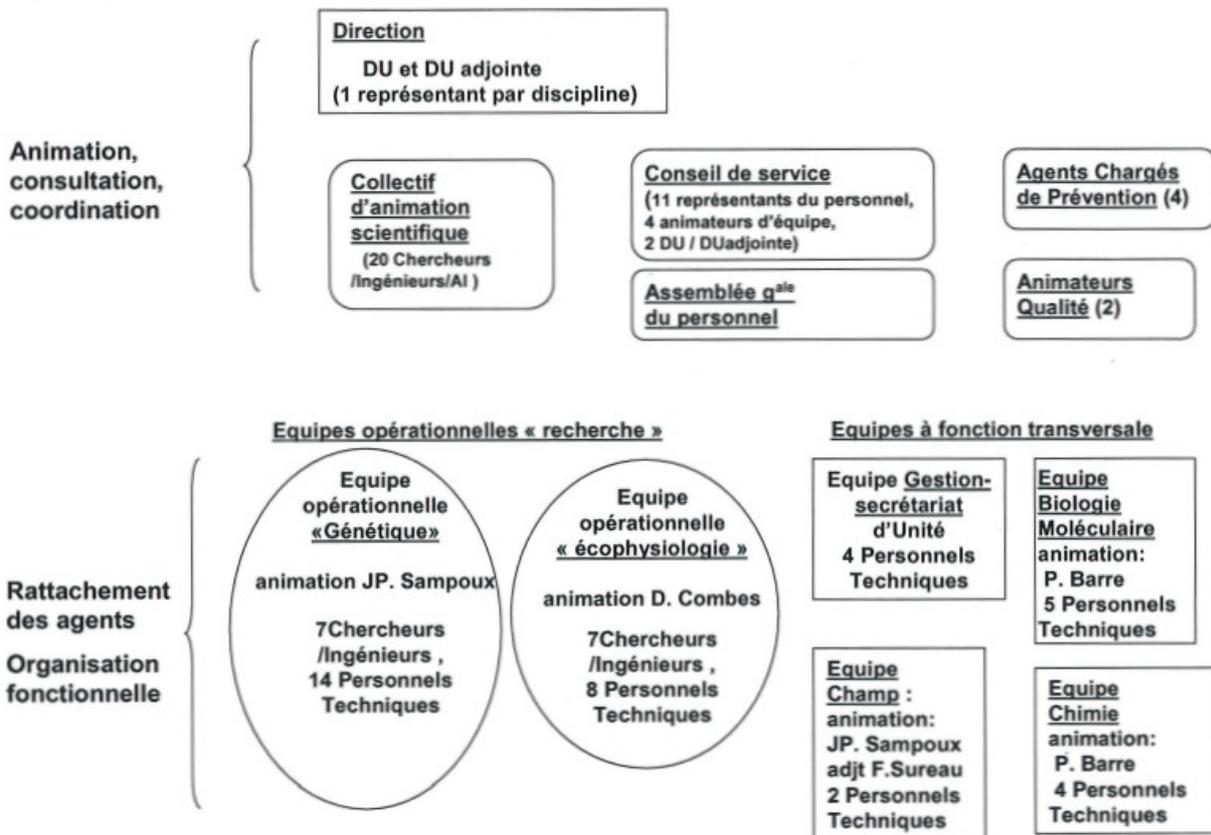
---

2 Site de l'URP3F : <http://www6.poitou-charentes.inra.fr/urp3f>

3 <http://www6.poitou-charentes.inra.fr/urp3f>

<http://www.poitou-charentes.inra.fr/Le-centre-Les-recherches/Prairies-semees>

<http://www.poitou-charentes.inra.fr/layout/set/print/Le-centre-Les-recherches/Unites-du-centre/URP3F/>



Organigramme fonctionnel de l'unité

## 3. Présentation du projet et du stage

### 3.1. Le contexte du stage

Le stage que j'ai effectué se situe dans le cadre d'un projet de collaboration entre l'INRA et un groupe d'entreprises privées intervenant dans le domaine de la création de variétés de graminées fourragères destinées aux semis de prairies.

En effet, la mission<sup>4</sup> de l'URP3F est de produire des connaissances pour améliorer la durabilité des prairies semées, la productivité et la qualité du fourrage en économisant les intrants (engrais, pesticides, eau) dans le respect de l'environnement et la prise en compte des changements climatiques.

Les prairies sont un atout majeur pour développer des systèmes d'élevage durables. Elles permettent d'améliorer l'autonomie des élevages et apportent des services écologiques variés. Sur les 10 millions d'hectares de prairies et cultures fourragères françaises, les prairies semées occupent près de 3 millions d'hectares.



Dans un contexte économique et environnemental contraignant, la prairie semée doit, aujourd'hui, répondre à plusieurs exigences : nourrir les animaux d'élevage mais aussi préserver l'environnement, fournir des services éco-systémiques et s'adapter au changement climatique (des températures plus chaudes et une disponibilité en eau réduite).

Lusignan est le principal site de l'INRA d'études des prairies semées destinées à l'alimentation des animaux d'élevage.

Les scientifiques de l'unité cherchent à analyser, modéliser et comprendre le fonctionnement de différentes espèces fourragères en fonction de contraintes diverses (eau, azote, lumière, sol, climat, ...) et des pratiques agricoles.

Pour répondre à ces objectifs, les pistes de recherche sont multiples : choix des espèces et des composition de mélanges, amélioration des variétés, adaptation des pratiques de gestion. L'unité associe ainsi des compétences multiples en génétique, écophysiologie, écologie, agronomie et statistiques.

### 3.2. Les objectifs du projet

L'approche classique de sélection de variétés fourragères consiste à sélectionner des plantes en fonction de leur rendement maximum et ceci dans des situations où il y a peu de contraintes sur les apports en eau ou azote.

Mais les nouvelles préoccupations environnementales liées au développement durable, aux changements climatiques conduisent notamment à vouloir optimiser les apports en intrants notamment l'eau et l'azote (principal constituant des engrais).

Dans cette optique, on peut par exemple réaliser des essais dans des conditions de faible apport en

<sup>4</sup> <http://www6.poitou-charentes.inra.fr/urp3f/Recherches/Axe-3-Gestion-et-valorisation-des-ressources-genetiques>

azote et mettre en évidence des variétés pour lesquelles on obtient un fort rendement. Mais le niveau de production dépend bien sûr de beaucoup d'autres paramètres et de tels essais ne permettrait pas de caractériser l'aptitude d'une espèce à prélever et valoriser l'azote du sol.

Il est donc important de pouvoir caractériser cette aptitude dans certains géotypes, indépendamment des autres facteurs qui contribuent à un bon rendement.

De cette façon, un sélectionneur pourrait ensuite recombinaison génétiquement une bonne aptitude à prélever l'azote avec d'autres facteurs favorables à la productivité fourragère.

Il serait donc intéressant de disposer de critères biologiques prédictifs pour évaluer l'efficacité de l'utilisation de l'azote et également de l'eau par les différentes variétés de plantes. **C'est un des objectifs du projet que d'évaluer l'intérêt de deux indices comme critères « intégratifs<sup>5</sup> » permettant de mesurer respectivement l'efficacité de l'utilisation de l'azote et de l'eau : l'INN et l'indice de la conductance stomatique ( $\Delta^{13}\text{C}$ ) :**

### ***L'indice de nutrition azotée (INN)***

l'indice de nutrition azotée (INN) permet de quantifier un état de carence ou de suralimentation en azote, en cours de végétation. L'INN renseigne donc sur la contrainte azotée subie par la plante. Il permet d'estimer la satisfaction des besoins en azote des plantes : s'il est inférieur à 1, il traduit un état de carence des plantes et, s'il est supérieur ou égal à 1, l'azote n'est pas facteur limitant de la croissance de la culture.

### ***L'indice de la conductance stomatique***

Cet indice représente la teneur en isotope  $^{13}\text{C}$  de la biomasse aérienne et permet de quantifier la conductance stomatique. Ainsi, cet indice permet de mesurer l'efficacité photosynthétique (la capacité à produire de la matière végétale par photosynthèse) et révèle également pour la plante étudiée, une situation de déficit en eau.

## ***3.3. Présentation de l'étude***

Pour ce projet, un essai comprenant **24 variétés<sup>6</sup> de ray-grass anglais d'origine variées** a été installé en 2011 dans cinq lieux d'essais aux conditions climatiques variées.

Dans ces différents lieux, les variétés sont testées avec **deux niveaux d'azote** (conditions de faible et fort apport d'engrais azoté).

Le premier niveau de fertilisation azotée (360 kg N/ha par an) a pour but de tester les variétés dans une situation de disponibilité en azote non limitante. Le second niveau correspond à un niveau de disponibilité en azote assez nettement limitant (160 kg N/ha par an).

Différents caractères liés à la productivité fourragère ont été mesurés en 2012. L'INN et le  $\Delta^{13}\text{C}$  ont été mesurés en laboratoire à partir d'échantillons prélevés sur chaque variété dans les différents lieux d'essais au moment de chaque coupe.



5 Ici, le terme « intégratif » s'entend au sens de cumulatif de toute une série d'informations, dans le temps et dans l'ensemble de la matière végétale.

6 Panel composé de variétés d'Europe océanique, d'Europe continentale, de Nouvelle-Zélande et de populations naturelles d'origine variées.

L'objectif de l'étude est de conduire l'analyse statistique des données de productivité fourragère, d'INN et de  $\Delta^{13}\text{C}$  ainsi recueillies afin de tenter ainsi de répondre aux questions suivantes :

- Existe-t-il, selon les différentes variétés d'une graminée fourragère, des variations significatives pour l'INN ?
- Si oui, les valeurs les plus élevées de l'INN sont-elles associées à une bonne productivité fourragère en conditions de disponibilité en azote limitante ?
- Existe-t-il, selon les différentes variétés d'une graminée fourragère, des variations significatives pour le  $\Delta^{13}\text{C}$  (efficacité photosynthétique) ?
- Si oui, est-il possible d'établir une relation entre efficacité photosynthétique et maintien de la productivité fourragère sous contrainte hydrique ?

### 3.4. Les objectifs du stage

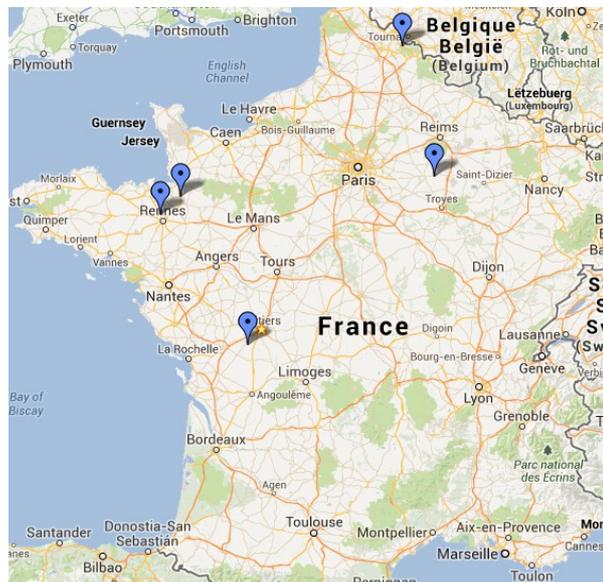
Cette étude s'inscrit sur plusieurs années. Les premières données ont été recueillies en 2011 mais c'est seulement pour 2012 qu'on dispose de données complètes sur une année. Le principal objectif du stage est de mener des analyses statistiques de ces données afin de préparer l'étude proprement dite qui intégrera des données supplémentaires. Il s'agit donc d'un travail préliminaire destiné à tester la qualité des données fournies par les différents établissements et qui doit permettre de conclure quant à leur exploitation possible. Le but du stage est également de mettre en évidence des premières tendances afin de voir si elle vont bien dans le sens attendu.

Mon stage a donc comporté plusieurs étapes :

- Mise en place de requêtes Access permettant d'extraire les données de la base Access qui héberge les données du projet, ceci en vue de leur utilisation par les logiciels statistiques.
- Tests de la validité des données recueillies (détection de valeurs aberrantes, contrôle de la forme des distributions empiriques).
- Conduite d'analyses statistiques à l'aide de modèles mixtes d'analyse de la variance intégrant les facteurs génétique (variété), environnementaux (lieux d'essai et dose d'azote) et temporel (dates de mesure).
- Utilisation de tests de comparaisons multiples de moyennes pour tester la précision des indices INN et  $\Delta^{13}\text{C}$ .
- Analyse d'associations entre caractères mesurés à l'aide de méthodes d'ACP.
- Réalisation des analyses statistiques avec le logiciel R : écriture de programmes, recherche des fonctions et des bibliothèques adaptées, etc.

### 3.5. Mise en place des essais et présentation des données

Comme cela a été dit plus haut, les données portent sur 24 variétés de ray-grass anglais. Ces variétés sont testées dans 5 lieux différents (voir la carte ci-contre) correspondant chacun à un établissement partenaire du projet.



Abréviation utilisée dans ce rapport	Nom de l'établissement	Lieu de l'essai
barenbrug	Barenbrug Tourneur Recherches 82600 Mas-grenier	Connantres (51230)
carneau	CARNEAU Frères 59358 Orchies cedex	Orchies (59358)
dlf	DLF – TRIFOLIUM France SARL 49320 Les Alleuds	La Mézière (35520)
gie Grass	GIE GRASS	86600 Saint-Sauvant
r2n	R2N - 12033 Rodez	Montours (35460)

Sur ces 24 variétés, on a mesuré entre 19 et 35 variables différentes (selon les lieux) et ceci, à différentes dates. Selon le lieu et selon la variable étudiée, le nombre de dates varie entre 1 et 10. Enfin, les dates d'observation des différentes variables sont différentes selon les lieux.

On observe un certain nombre d'anomalies dans les données fournies par certains établissements : données manquantes pour certaines variables : il y a en tout **47194** données non vides et **1046** données vides (soit environ 2,2 % de données vides).

Par ailleurs, parmi les données, il y a deux types d'essai : type « ligne » et type « parcelle », avec une minorité de données de type « ligne » (528 sur 47194). Les cinq établissements ont fourni des données de type « parcelle » et parmi eux, deux (« barenbrug » et « carneau ») ont fourni des données de type « ligne ». Le fait d'utiliser un dispositif en lignes plutôt qu'en parcelles facilite notamment le comptage des épis lors de la date d'épiaison<sup>7</sup>. C'est pourquoi deux établissements ont

<sup>7</sup> La précocité d'épiaison ou date de début d'épiaison correspond à la date à partir de laquelle les premiers épis sont visibles (environ 10 épis par mètre linéaire). C'est à partir de ce stade que la qualité alimentaire du ray-grass anglais et des autres graminées en général diminue rapidement.

fait le choix d'utiliser ce dispositif pour étudier l'épiaison et ont également profité du même dispositif pour étudier d'autres variables.



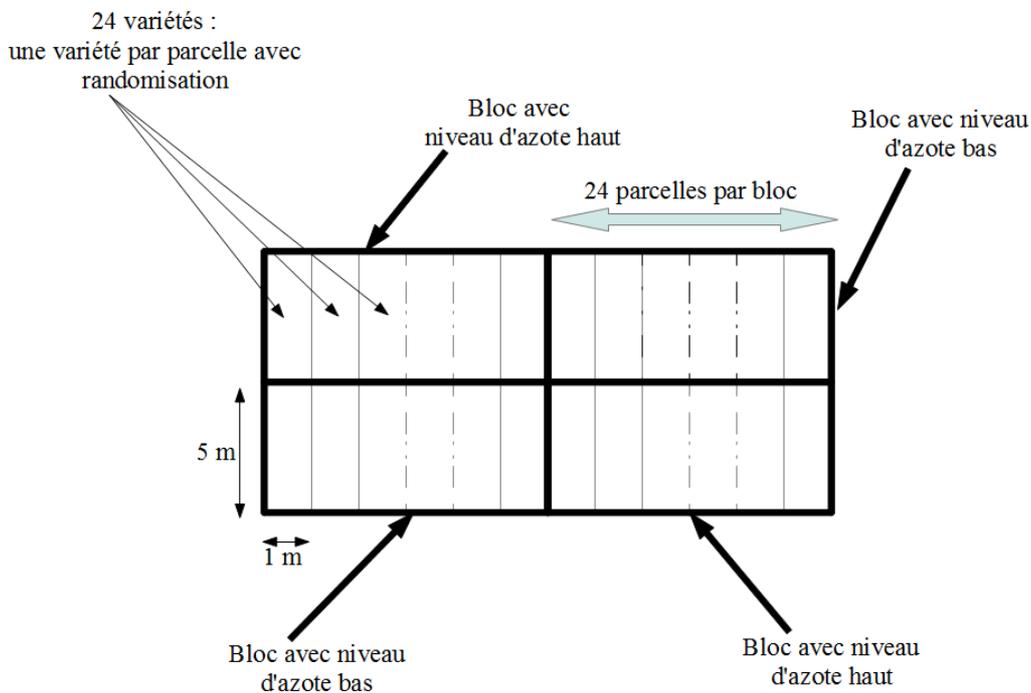
*La taille de chaque parcelle est de 5 m<sup>2</sup>*

### Les essais de type parcelle :

Dans quatre lieux sur cinq, les variétés ont été testées en parcelles, avec **deux niveaux d'azote** (conditions de faible et fort apport d'engrais azoté) et **deux répétitions** (blocs) pour chaque niveau d'azote.

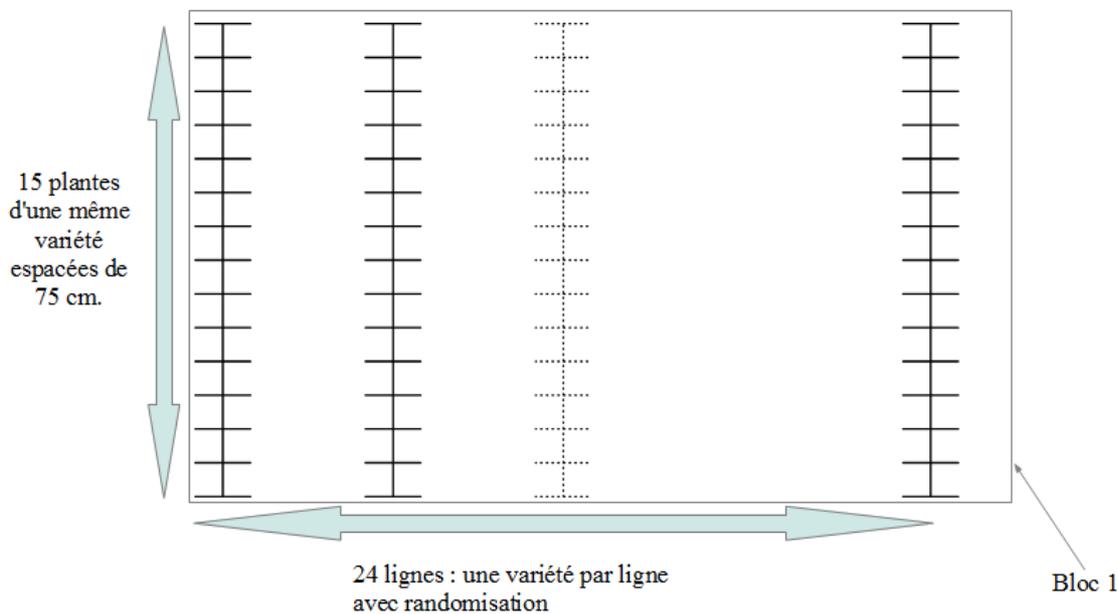
Pour le cinquième lieu (« gie grass »), le niveau faible d'azote n'a pas été testé et le niveau fort d'azote a été testé avec **quatre répétitions** (blocs).

### Mise en place des essais de type parcelle :



*Essai de type parcelle  
(sauf « gie grass » pour lequel les deux blocs sont à niveau d'azote fort)*

### Mise en place des essais de type ligne :



*Essai de type ligne : schéma d'un bloc  
(quatre blocs de ce type sont installés pour un essai ligne)*

Nous avons remarqué ci-dessus qu'il y avait un grand nombre de variables étudiées (35 en tout) et que ce nombre diffèrait selon les lieux.

Mais dans le cadre de cette étude, nous nous intéresserons plus particulièrement à quatre variables : « INN », « deltaC13 », « mat » et « rdtsec ».

Nous avons déjà évoqué les variables « INN » et « delta13 ». Passons rapidement en revue les deux autres variables :

#### **Matières azotées totales : mat**

La variable « mat » correspond à la quantité totale de matière azotée, elle s'exprime en pourcentage de la matière sèche.

#### **Le rendement en matière sèche : rdtsec**

« rdtsec » est une variable qui donne la production de matière sèche en tonnes/hectare. Les variétés à fort rendement permettent de produire le fourrage nécessaire sur une surface plus réduite. Cela permet donc de produire pour moins cher : moins de frais de semis, de récolte, d'entretien (azote...), tout en passant moins de temps sur la culture. Cela permet également de produire plus sur une surface donnée lorsque la surface disponible est limitée et par conséquent de maximiser le revenu par unité de surface. Ce critère est donc particulièrement important.

Comme le montre le tableau suivant, le nombre et la répartition des observations dans le temps et

l'espace ne sont pas les mêmes selon les variables. En particulier, seules les variables « INN », « rdtsec » et « mat » sont présentes dans les cinq lieux.

Bien qu'il y ait un grand nombre de variables mesurées, mon stage s'est focalisé sur quelques variables particulièrement importantes pour le sujet : « rdtsec, INN et « deltaC13 ».

Lieu	Variable	Nombre de données (non vides)	Nombre de dates	Dates	Nombre de niveaux d'azote (niveauN)	Nombre de blocs
barenbrug	INN	459	5	01/06/2012 - 10/05/2012 - 24/10/2012 - 26/07/2012 - 28/06/2012	2	2
barenbrug	mat	459	5	01/06/2012 - 10/05/2012 - 24/10/2012 - 26/07/2012 - 28/06/2012	2	2
barenbrug	rdtsec	459	5	01/06/2012 - 10/05/2012 - 24/10/2012 - 26/07/2012 - 28/06/2012	2	2
carneau	deltaC13	96	1	25/05/2012	2	2
carneau	INN	384	4	22/08/2012 - 25/05/2012 - 25/06/2012 - 30/04/2012	2	2
carneau	mat	384	4	22/08/2012 - 25/05/2012 - 25/06/2012 - 30/04/2012	2	2
carneau	rapportC/N	96	1	25/05/2012	2	2
carneau	rdtsec	528	6	13/11/2012 - 14/10/2011 - 22/08/2012 - 25/05/2012 - 25/06/2012 - 30/04/2012	2	2
carneau	teneurC	96	1	25/05/2012	2	2
dlf	INN	433	5	06/09/2012 - 16/05/2012 - 18/07/2012 - 19/06/2012 - 30/11/2012	2	2
dlf	mat	433	5	06/09/2012 - 16/05/2012 - 18/07/2012 - 19/06/2012 - 30/11/2012	2	2
dlf	rdtsec	576	6	06/09/2012 - 07/11/2011 - 16/05/2012 - 18/07/2012 - 19/06/2012 - 30/11/2012	2	2
gie grass	INN	480	5	03/04/2012 - 14/06/2012 - 18/07/2012 - 19/05/2012 - 26/10/2012	1	4
gie grass	mat	480	5	03/04/2012 - 14/06/2012 - 18/07/2012 - 19/05/2012 - 26/10/2012	1	4
gie grass	rdtsec	576	6	03/04/2012 - 14/06/2012 - 16/11/2011 - 18/07/2012 - 19/05/2012 - 26/10/2012	1	4
r2n	deltaC13	191	2	10/05/2012 - 28/08/2012	2	2
r2n	INN	672	7	03/07/2012 - 05/06/2012 - 10/05/2012 - 12/04/2012 - 25/07/2012 - 28/08/2012 - 28/09/2012	2	2
r2n	mat	672	7	03/07/2012 - 05/06/2012 - 10/05/2012 - 12/04/2012 - 25/07/2012 - 28/08/2012 - 28/09/2012	2	2
r2n	rapportC/N	192	2	10/05/2012 - 28/08/2012	2	2
r2n	rdtsec	960	10	03/07/2012 - 05/06/2012 - 09/11/2011 - 10/05/2012 - 12/04/2012 - 23/10/2012 - 25/07/2012 - 26/09/2011 - 28/08/2012 - 28/09/2012	2	2
r2n	teneurC	192	2	10/05/2012 - 28/08/2012	2	2

Analyse des données (réalisée avec R : voir annexe 2)

## 4. Quelques remarques sur le logiciel R

### 4.1. R et l'INRA

Durant ce stage, une très grande partie de mon travail s'est effectué sur informatique et plus précisément sur le logiciel « R ». En effet, la nature expérimentale des données nécessairement entachées d'erreurs de toute sorte impose un traitement statistique.

Pendant longtemps, à l'INRA, les calculs statistiques ont été réalisés à l'aide du logiciel commercial « SAS »<sup>8</sup>. Mais les coûts importants liés à son utilisation ont conduit plusieurs centres INRA, ces dernières années, à abandonné ce logiciel au profit d'un logiciel libre et gratuit : « R »<sup>9</sup>.

L'INRA de Lusignan a suivi ce mouvement en adoptant très récemment ce logiciel. Mais ce changement a eu des conséquences importantes à cause des différences importantes entre ces deux logiciels.

Tout d'abord, les syntaxes des commandes utilisées dans les deux logiciels sont assez différentes ainsi que les informations affichés par les commandes. Voici par exemple le code correspondant à une analyse de la variance à un facteur pour chaque logiciel :

#### Analyse de la variance à un facteur - Exemple avec SAS

```
# On importe les données dans le répertoire de travail « travail »
Proc import out=travail.foret datafile="C:\data\foret.txt"
# on déclare que la procédure va travailler sur la table « foret »
Proc glm data=travail.foret
# On déclare que la variable foret est qualitative
Class foret
# On déclare le modèle
Model hauteur=foret
# On affiche les moyennes
means foret
# on réalise une sortie des résultats de l'analyse
output out=sortie r=residu p=predite;run
# On reprend les données par gplot pour afficher un graphique
plot residu*predite
```

#### Analyse de la variance à un facteur - Exemple avec R

```
# On importe les données
foret1=read.table("C:\data\foret.txt",header=TRUE)
# On déclare que la variable foret est qualitative
foret1$foret=ac.factor(foret1$foret)
# on réalise un ajustement avec lm
lm.foret=lm(hauteur~foret,foret1)
# On installe la bibliothèque car
library(car)
# On fait une anova et on affiche les résultats
Anova(lm.foret,type=III)
# On affiche des graphiques
par(mfrow=c(2,2))
plot(lm.foret,las=1)
```

On constate que la façon de faire, la syntaxe, les commandes sont très différentes. Tout ceci pose évidemment le problème de formation du personnel à ce nouveau logiciel. Au début de mon stage, j'ai ainsi eu l'occasion d'assister à un stage d'initiation à R dans les locaux de l'INRA de Lusignan à destination du personnel.

8 Site officiel de SAS : <http://www.sas.com/>

9 Site officiel de R : <http://www.r-project.org/>

De fait, pour développer mes analyses avec R, je n'ai pas pu m'appuyer sur l'expérience des collègues et bien au contraire, une de mes missions pendant ce stage était d'apporter des solutions avec R à des problèmes qui jusqu'à maintenant n'avaient jamais été traités avec ce logiciel à l'INRA de Lusignan.

SAS est un logiciel très complet qui fait référence aujourd'hui et bien que le projet R ait déjà plusieurs années d'ancienneté, il n'a sans doute pas encore la puissance de SAS. De plus, la redevance annuelle due pour l'utilisation de SAS donne droit à un support technique (par téléphone ou e-mail), ce qui n'est pas le cas avec R. Mais la communauté qui s'est développée autour de ce logiciel libre est extrêmement active.

Le principe de R est basé sur l'existence d'un noyau de base donnant accès à de nombreuses fonctions de base auquel viennent s'ajouter des bibliothèques<sup>10</sup>. Il existe aujourd'hui plus de 45000 bibliothèques différentes !

Mais ce dynamisme a aussi des aspects négatifs : il est en effet très difficile de se repérer dans ces innombrables librairies : certaines deviennent obsolètes au profit d'autres plus récentes et plus solides, certaines fonctions présentent des buggues, parfois encore, telle fonction d'une librairie est incompatible avec telle autre d'une librairie installée.

Pour réaliser un simple test, on a souvent le choix dans R entre plusieurs fonctions disponibles dans des librairies différentes, chaque fonction ayant ses limitations et ses spécificités.

Une aide en anglais existe pour chaque librairie, rédigée par les développeurs, mais face à une situation concrète particulière, beaucoup de questions restent sans réponse. Il faut alors aller à la pêche aux informations sur internet, sur des forums de discussion spécialisés.

## 4.2. Programmer avec R

La base de donnée présente un grand nombre de paramètres et mon premier travail a consisté à réaliser toute une série d'analyse de la variance par lieu, date et variable, ce qui représentait plusieurs dizaines d'analyses.

Je me suis très vite rendu compte que pour être efficace, il me faudrait également apprendre à programmer avec R et réaliser des fonctions comprenant des boucles et des conditions, etc...

La réalisation de ces fonctions a été quelquefois longue et délicate mais heureusement, le langage de programmation de R est assez simple.

## 4.3. Les principales fonctions de R utilisées pendant le stage

### a. La fonction « lm » (base)

La fonction « lm » disponible dans le noyau de base est utilisée pour ajuster des modèles linéaires à effets fixes.

### b. La fonction « lmer » (bibliothèque : lme4)

La fonction « lmer » disponible dans la bibliothèque « lme4 » est utilisée pour ajuster des modèles à effets aléatoires ou mixtes.

---

<sup>10</sup> En anglais : « library »

Dans un premier temps, j'ai voulu utiliser la fonction « lme » de la librairie « nlme » et qui s'applique également aux modèles linéaires à effets aléatoires ou mixtes. En effet, cette librairie est déjà ancienne et de ce fait, bien documentée.

Mais, après avoir commencé à utiliser cette librairie, je me suis aperçue qu'elle ne permettait que de traiter les effets aléatoire imbriqués (hiérarchisés) et par conséquent, pas les effets aléatoires croisés.

En cherchant à nouveau sur internet, je suis tombée sur une autre bibliothèque « lme4 » mettant à disposition la fonction « lmer ».

La bibliothèque « lme4 » est plus récente que « nlme » et de fait, moins bien documentée. Mais contrairement à la fonction « nlme », la fonction « lmer » permet de traiter tous les types d'effets aléatoires.

De plus, la fonction « lmer » est plus puissante et rapide que la fonction « lme »

### c. La fonction Anova (bibliothèque car)

La fonction « Anova » réalise une analyse de la variance sur un modèle ajusté (par une fonction lm, glm, etc...)

Dans un premier temps, j'avais commencé à utiliser la fonction « anova » disponible dans le noyau de base (c'est la seule que je connaissais et c'est celle que j'avais utilisée en TP en master). Mais, en faisant des recherches sur internet, j'ai découvert que cette fonction n'était pas adaptée dans le cas de données non équirépétées, ce qui est assez souvent le cas avec nos données, du fait de données manquantes.

En effet, la fonction standard « anova » effectue une analyse de « type I » (voir ci-dessous), ce qui dans le cas de données non équilibrées, n'est pas conseillé. Il vaut mieux dans ce cas faire une analyse de type II ou III, ce que permet la fonction « Anova » de la bibliothèque « car ».

#### **Analyses de variance de type I , II ou III**

Un modèle d'analyse de la variance peut-être divisé en différents termes (les effets principaux et les différentes interactions). Dans le cas équirépété, les propriétés algébriques font que la décomposition de la variance est unique.

Mais dans le cas non-équirépété cette décomposition n'est pas unique. Il faut alors faire des choix. L'apparition des différents types I, II et III dans la littérature statistique résulte de considérations qui mélangent des éléments statistiques et informatiques. Dans le cas équirépété, les décompositions de type I, II ou III donnent les mêmes résultats.

C'est semble-t-il le logiciel SAS qui, peu à peu, a imposé comme des normes, ces différents types de décomposition. Les considérations mathématiques qui expliquent ces différents types sont assez complexes et je n'ai pas eu le temps d'approfondir cette question.

Mais il est clair que dans le cas non équirépété, c'est le type III qui est le plus souvent adopté et c'est celui que j'ai également retenu dans mes différentes analyses.

#### **Quelques remarques concernant la fonction Anova (bibliothèque « car »)**

Afin de bien utiliser la fonction Anova, il est important de maîtriser quelques paramètres :

Dans la syntaxe, on précisera toujours « Singular.ok=TRUE » (par défaut, singular.ok=FALSE). Cette option permet au programme de ne pas s'arrêter en affichant un message d'erreur quand la

matrice à analyser n'est pas régulière, ce qui est très utile dans le cadre d'une boucle effectuant un grand nombre d'analyse. Il arrive en effet que certaines analyses du fait de données manquantes n'aboutissent pas. Il est alors très gênant que dans ce cas, le programme s'arrête.

On prendra également soin de bien préciser que la statistique à afficher est la statistique de Fisher en ajoutant l'option « `test.statistic= 'F'` »

Enfin, on veillera à bien préciser le contraste avec par exemple la commande « `options(contrasts=c("contr.sum","contr.sum"))` »

#### d. La fonction `lsmeans` de la bibliothèque `lsmeans`

La fonction « `lsmeans` » calcule les moyennes ajustées (« least-squares means ») des valeurs pour les facteurs spécifiés : `lsmeans(modele_ajuste,~facteur)`

#### e. La fonction `SNK.test` de la bibliothèque `agricolae`

La fonction `SNK.test` effectue un test de comparaison des moyennes selon la méthode de la plus petite différence significative (« Least significant difference »).

#### f. ACP et R

La fonction « `princomp` » de R permet de faire des ACP. C'est cette fonction que j'ai utilisée au début mais elle s'est avérée inadaptée dans la cas de données manquantes.

J'ai alors utilisé la fonction « `PCA` » de la librairie « `FactoMineR` ». Un site internet<sup>11</sup> est dédié à cette librairie et de nombreux exemples y sont disponibles.

### 4.4. Ecriture des modèles

Les modèles dans les fonctions `lm`, `glm`, `lme`, `lmer`, ... sont spécifiés sous une forme compacte symbolique. L'opérateur `~` est fondamentale dans la formation de ces modèles.

Une expression de la forme « `y ~ modèle` » signifie que la réponse `y` est modélisé par un prédicteur linéaire spécifié symboliquement par le modèle. Un tel modèle se compose d'une série de termes séparés par des opérateurs `+`.

Les termes eux-mêmes sont constitués de noms de variables et de facteurs séparés par différents opérateurs (« `:` », « `*` », « `^` », etc..). Dans l'écriture des modèles j'ai principalement utilisé les opérateurs « `:` » et « `*` » :

- `A:B` est interprété comme l'interaction de A et B
- L'opérateur `*` est un facteur de croisement :  $A * B \Leftrightarrow A + B + A:B$
- On a également l'opérateur « `/` » pour spécifier un effet hiérarchisé dans une autre : par exemple si l'effet B est « hiérarchisé » dans l'effet A, on écrit : « `A/B` »

---

<sup>11</sup> <http://factominer.free.fr>

## 5. Analyse de la variance des données et comparaison des moyennes

### 5.1. Première série d'analyses des données par lieu et par date

La première phase de mon travail a consisté à vérifier que les facteurs « variété » et « niveau d'azote » était bien significatifs, c'est-à-dire que, pour chaque date d'observation d'une variable et dans chaque lieu, les données mettaient bien en évidence des différences entre les différentes variétés et également entre les niveaux d'azote. Dans le cas contraire, la suite de l'étude perdrait tout son intérêt.

#### a. Vérification des conditions

Pour effectuer une analyse de la variance, il est en principe nécessaire d'effectuer des tests afin de vérifier que les conditions d'application de la méthode sont remplies :

1) Les données sont indépendantes : on considère en général que cette condition est vérifiée et c'est au moment du recueil des données que toutes les précautions pour garantir cette hypothèse doivent être prises.

2) les données de chaque échantillon sont issues de populations qui ont la même variance (homoscédasticité),

3) les données de chaque échantillon sont issues de populations qui suivent toutes des lois normales.

Plutôt que de vérifier la condition de normalité sur les observations initiales des différents échantillons, on peut vérifier que les erreurs par rapport au modèle théorique de l'analyse de la variance suivent une loi normale. Concrètement, on appliquera les tests classiques de normalité aux résidus après ajustement des valeurs : il s'agit là d'une méthode empirique et indirecte mais couramment utilisée.

Cette 3ème condition n'est pas importante dans le cas d'une analyse de la variance, puisque l'on peut s'en passer quand le nombre de données est important du fait de la propriété de normalité asymptotique.

Dans le cadre du stage, dans un premier temps, j'avais commencé à étudier comment tester simplement les hypothèses de normalité et d'homoscédasticité avec R dans le cas d'une analyse de la variance.

Mais en fait, après réflexion, il s'est avéré que dans les différentes analyses de la variance dont nous allons parlé maintenant, l'hypothèse 3 n'est pas vérifiable puisque, comme nous allons l'expliquer un peu plus bas, nous considérerons à chaque fois des échantillons de taille 1 (essais sans répétition). Nous reviendrons un peu plus loin sur ce point.

#### b. Construction des modèles et explications

J'ai tout d'abord construit une première fonction « analyse1() » (voir annexe 3) qui fait des analyses de la variance par lieu, date et type de variable. Selon les cas, parcelle ou ligne, 1 ou 2 niveaux d'azote, on utilise différents modèles :

- Dans le cas où il y a un seul niveau d'azote (données « gie grass », données de type ligne, ou autres données de type parcelle avec données absentes sur l'un des niveaux), on utilise le

modèle :  $valeur = variete + bloc + \epsilon$

\* valeur : la valeur prise pour la variable étudiée

\* variete : le facteur variété (24 variétés en tout)

\* bloc : le facteur bloc (2 ou 4 selon les cas)

Le fait d'avoir choisi les variétés et le niveau d'azote comme des facteurs est nécessaire si on veut mettre en évidence une significativité de ces facteurs. Pour le facteur « bloc », normalement, le terrain sur lequel ont été délimités les 2 blocs (eux-mêmes divisés en parcelles élémentaires) présente une certaine homogénéité et par conséquent, on ne devrait pas observer d'effet significatif de ce facteur. Mais, dans la réalité, il peut arriver que les blocs soient sensiblement différents, ceci n'est pas gênant à partir du moment où on inclut ce facteur dans notre modèle.

- Par ailleurs, dans le cas où il y a deux niveaux d'azote, on teste deux modèles : un modèle fixe dans lequel l'effet de l'interaction « variété:niveauN » est fixe et un autre dans lequel l'effet de l'interaction « variété:niveauN » est aléatoire :

Un modèle à effet fixe :  $valeur = variete * niveauN + bloc + \epsilon$

Un modèle à effet mixte :  $valeur = variete + niveauN + bloc + \underbrace{variete : niveauN + \epsilon}_{\text{les effets aléatoires}}$

La formulation R des modèles est alors la suivante :

modèle fixe :  $valeur \sim variete * niveauN + bloc$

modèle mixte :  $valeur \sim variete + niveauN + bloc + (1 | variete : niveauN)$  (1)

Remarque : le modèle (1) ci-dessus n'a pas réellement de sens, il s'agit plutôt d'une astuce qui permet de calculer simultanément les statistiques de Fisher des deux modèles suivant :

$valeur \sim variete + bloc + (1 | niveauN) + (1 | variete : niveauN)$  (2)

$valeur \sim niveauN + bloc + (1 | variete) + (1 | variete : niveauN)$  (3)

Dans le modèle (2), le niveau d'azote est considéré comme aléatoire, ce qui permet de voir si le facteur variété est significatif, indépendamment des niveaux d'azote.

Dans le modèle (3), les variétés sont considérées comme aléatoire, ce qui permet de voir si le facteur « niveau d'azote » est significatif, indépendamment des « variétés ».

En utilisant le modèle (1), on calcule ainsi les deux statistiques de Fisher respectivement associées au facteur variété du modèle (2) et au facteur « niveauN » du modèle (3).

N'étant pas entièrement convaincue de cela, j'ai effectué quelques tests sur des exemples et j'ai pu vérifier la validité de cette pratique.

### c. Interprétation des résultats

La fonction « analyse1 » (voir annexe 3) permet ainsi de produire un tableau complet de résultats pour chaque lieu. A la demande de Monsieur Sampoux, j'ai ainsi affiché dans ces tableaux un certain nombre de paramètres utiles à une analyse complète des résultats.

Voici un très court extrait des résultats obtenus :

Analyse de la variance pour "barenbrug"									
Lieu	typessai	codevar	datejour	Modèle fixe : valeur=variete+bloc+ε				Modèle mixte : valeur=variete*niveauN +bloc+ε	
				Pvalue Variété	Pvalue NiveauN	Pvalue Bloc	Pvalue Variete:NiveauN	Pvalue Variété	Pvalue NiveauN
barenbrug	ligne	epiaison	-	8,20E-10	-	3,02E-01	-	-	-
barenbrug	ligne	haut	14/04/2012	1,91E-07	-	2,05E-01	-	-	-
barenbrug	ligne	resistfroid	01/04/2012	1,75E-07	-	3,64E-01	-	-	-
barenbrug	ligne	resistmal	2510/2011	4,01E-03	-	1,48E-01	-	-	-
barenbrug	ligne	resistrouill	17/09/2011	1,40E-10	-	2,51E-03	-	-	-
barenbrug	parcelle	adf	01/06/2012	6,62E-08	9,29E-01	3,34E-01	2,28E-01	1,86E-04	9,38E-01
barenbrug	parcelle	adf	10/05/2012	1,02E-10	1,62E-14	1,95E-02	5,96E-01	5,15E-07	1,27E-10
barenbrug	parcelle	adf	24/10/2012	3,41E-05	1,83E-15	4,68E-01	1,58E-01	8,57E-03	1,28E-09
barenbrug	parcelle	adf	26/07/2012	1,60E-03	7,03E-02	4,95E-01	9,22E-01	3,66E-03	5,79E-02
barenbrug	parcelle	adf	28/06/2012	1,90E-03	1,04E-11	8,30E-01	5,81E-01	8,75E-03	5,45E-09
barenbrug	parcelle	adl	01/06/2012	9,98E-10	8,46E-01	1,91E-01	3,03E-02	5,10E-04	8,90E-01
barenbrug	parcelle	adl	10/05/2012	6,10E-02	1,55E-15	8,87E-03	3,20E-01	1,91E-01	1,87E-10
barenbrug	parcelle	adl	24/10/2012	8,31E-02	9,90E-01	6,30E-02	5,47E-01	1,22E-01	9,90E-01
barenbrug	parcelle	adl	26/07/2012	4,15E-05	1,58E-05	2,10E-01	2,30E-01	5,29E-03	3,20E-04

Après lecture de l'ensemble des résultats, on observe que dans la plupart des cas, le facteur variété est très significatif (p-value très inférieure à 0,01) ainsi que le facteur «niveau d'azote ».

Dans les cas où le facteur variété ou le facteur «niveau d'azote » ne sont pas significatifs, le point de vue d'un « expert » en agronomie est nécessaire car d'autres paramètres (date, variable étudiée, ...) peuvent expliquer les résultats.

L'analyse de ces résultats d'un point de vue strictement mathématique n'est donc pas pertinente. C'est donc Monsieur Sampoux qui a analysé ces données et n'a rien relevé d'anormal.

Par souci de présentation, je n'ai pas fait figurer dans le tableau ci-dessus toutes les variables calculées. Par exemple, l'affichage des min et max des variables permet de déceler d'éventuelles anomalies.

Pour compléter cette première étude, j'ai effectué une seconde série d'analyses de la variance par lieu, date, type de variable et niveau d'azote. En effet, dans le cas d'une interaction forte avec le niveau d'azote, il est intéressant d'étudier séparément pour chaque niveau d'azote afin de mieux cerner l'effet principal de variété. C'est dans ce but que j'ai construit la fonction « analyse2 » (voir annexe 3)

Dans le tableau précédent, on remarque une interaction forte « Variete:NiveauN » pour le lieu « barenbrug », la variable « adl » à la date du 01/06/2012.

En étudiant séparément ce cas pour des niveaux faibles et forts d'azote, on obtient les résultats suivants :

Lieu	Codevar	datejour	niveauN	Pvalue Variété	Pvalue bloc	Val. min.	Val. moy.	Val. max.	Residu Mean Sq.
barenbrug	adf	01/06/2012	faible	0.0024	0.46	22	24.1	26.94	0.29
barenbrug	adf	01/06/2012	fort	0.00038	0.069	20.8	24.1	27.8	2.86

Ces résultats montrent bien un effet significatif du facteur « variété ». On a ainsi pu vérifier dans les autres cas similaires la significativité du facteur « variété ».

#### d. Retour sur la question de la vérification des hypothèses

Les analyses précédentes montrent que le facteur « bloc » est en général peu significatif mais on note un certain nombre d'exceptions notoires. Ceci révèle des situations pour lesquelles le terrain sur lequel était délimités les blocs n'était pas vraiment homogène.

Dans la suite de l'étude, on décide donc d'inclure le facteur « bloc » dans les différents modèles.

Normalement, les blocs sont équivalents à des répétitions d'un essai donné. Le fait d'inclure systématiquement le facteur « bloc » dans les modèles a une conséquence importante : il réduit systématiquement la taille de nos échantillons à 1 (pas de répétition).

Dans ces conditions, la vérification des hypothèses de normalités des populations dont sont issus les différents échantillons n'est pas vraiment possible.

Mais ces tests ont tout de même leur utilité dans le cas d'échantillons de taille 1 pour vérifier l'homoscédasticité des résidus ou des anomalies sur les observations. Dans le cas d'un grand nombre de résidus, si le test révèle que ces résidus ne suivent pas une loi normale, cela peut indiquer que certaines observations sont suspectes ou bien qu'on est en présence d'un ou plusieurs facteurs non contrôlés par le modèle.

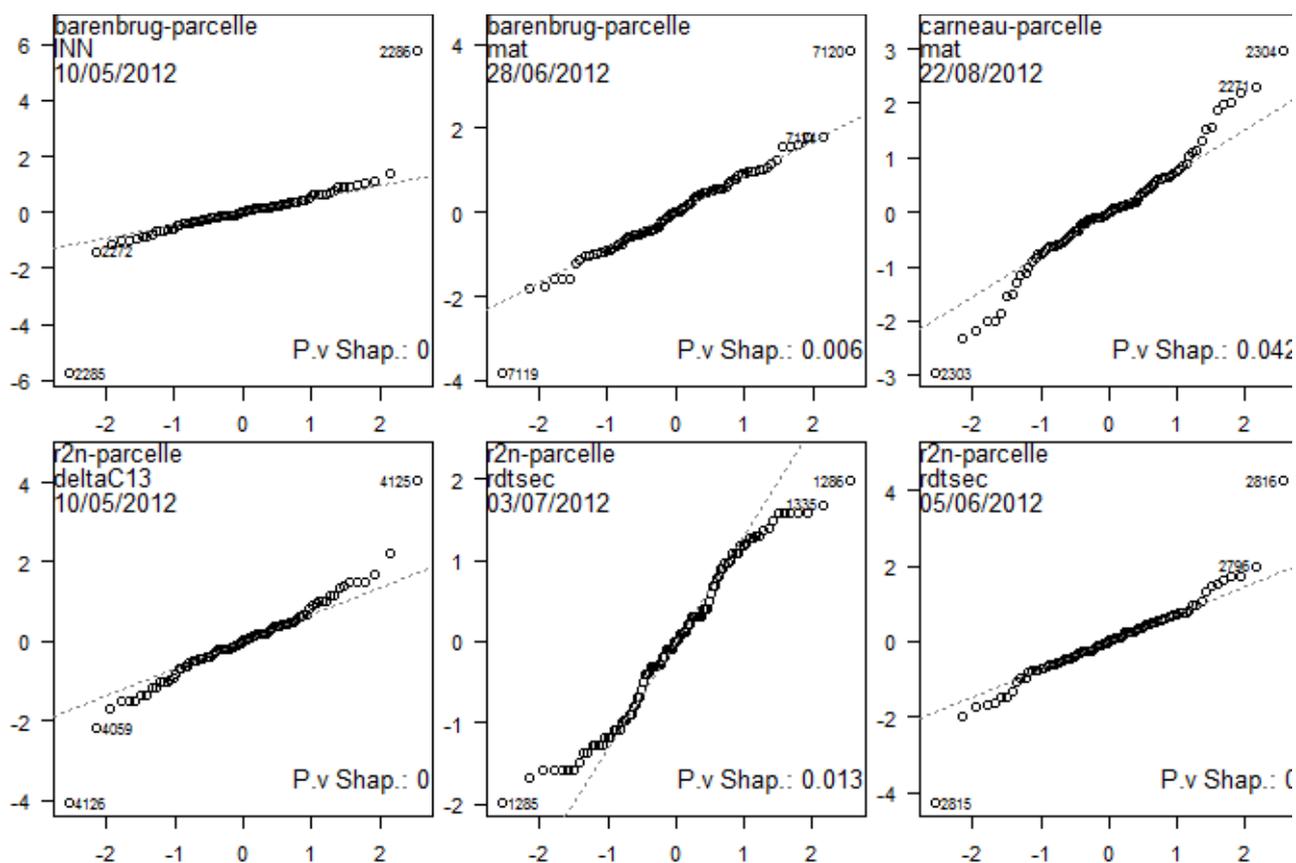
Une méthode visuelle pour détecter ces observations suspectes est alors d'afficher le QQ-plot des résidus qui peut mettre en évidence un écart important d'une erreur par rapport à 0. Il faut alors s'interroger sur la validité de cette mesure, mais la suppression d'un point atypique doit toujours être justifiée par des considérations non-statistiques (sauf cas évident).

Suite à cette discussion, j'ai mené toute une série de tests par rapport aux analyses précédentes en utilisant le test de Shapiro (test de normalité) ainsi qu'un examen visuel du QQ-plot.

J'ai limité ces tests aux variables « rdtsec », « mat », « INN » et « deltaC13 ».

On peut trouver en annexe 4 le code R permettant l'affichage des cas pour lesquels les résidus ne sont pas issus d'une population normale (au seuil de 0.05). J'ai ainsi mis en évidence six cas « suspects » :

Lieu	Type essai	Variable	Ddate
Barenbrug	parcelle	INN	10/05/2012
Barenbrug	parcelle	mat	28/06/2012
Carneau	parcelle	mat	22/08/2012
R2n	parcelle	deltaC13	10/05/2012
R2n	parcelle	rdtsec	03/07/2012
R2n	parcelle	rdtsec	05/06/2012



QQ-plot et p-value du test de Shapiro des cas suspects

Résultats obtenus avec la fonction « tests » (annexe 4)

L'affichage des QQ-plot de ces cas met en évidence dans 5 cas un couple de données visiblement atypiques. En examinant ces données et en les comparant aux autres données de la série, j'ai effectivement déceler des anomalies :

Cas suspect	Anomalie probable
Barenbrug/INN/10/05/2012	Lignes suspectes : 2285 et 2286 Explication possible : la valeur de la ligne 2286 est nulle (erreur probable de saisie)
Barenbrug/mat/28/06/2012	Lignes suspectes : 7119 et 7120 Explication possible : la différence des valeurs pour les blocs 1 et 2 est anormalement élevée (3) comparée aux autres différences globalement inférieures à 1.
Carneau/mat/22/08/2012	Lignes suspectes : 2303 et 2304 Explication possible : la différence des valeurs pour les blocs 1 et 2 est anormalement élevée (5,2) comparée aux autres différences globalement inférieures à 0,5.

Cas suspect	Anomalie probable
R2n/deltaC13/10/05/2012	Lignes suspectes : 4125 et 4126 La valeur de la ligne 4126 anormalement basse par rapport aux autres valeurs de la série.
R2n/rdtsec/03/07/2012	Pas d'explication. Peut-être est-on en présence d'un facteur non contrôlé ?
R2n/rdtsec/05/06/2012	Lignes suspectes : 2815 et 2816 La valeur de la ligne 2816 (2,577) anormalement élevée comparée aux autres leurs. Peut-être une erreur de saisie (2,577 au lieu de 1,577).

J'ai alors effectué de nouveaux tests en supprimant la ou les données suspectes et tous les problèmes ont disparu à l'exception du cinquième pour lequel je n'ai pas d'explication.

Mais ces simples constatations ne suffisent pas pour justifier la suppression de ces observations suspectes de la base de donnée. D'autres justifications (conditions d'expérimentation, considérations biologiques, etc...) seraient nécessaires pour confirmer ou infirmer leur suppression.

Malheureusement, tous ces tests ont été effectués en fin de stage, au moment de la rédaction de ce rapport. Je n'ai donc pas eu le temps de pousser plus loin l'étude de ces cas et éventuellement de reprendre les analyses sans ces données.

Etant donné le faible nombre de cas suspects, il est probable que, même en supprimant ces données, les résultats obtenus seraient pratiquement identiques. Néanmoins, pour les analyses qui seront menées dans les prochaines années, il sera important de tenir compte de ces tests.

## 5.2. Deuxième série de tests d'analyses des données par lieu

Dans cette deuxième série de tests, nous nous sommes intéressés plus particulièrement à trois variables : « INN », « mat » et « rdtsec ».

Par ailleurs, nous avons concentré l'étude sur les seules dates de l'année 2012 (année d'observation complète).

Comme précédemment, nous allons tester, pour chaque lieu, sur une variable donnée, deux modèles différents.

Les facteurs étudiés sont :

- le facteur variete (24 variétés différentes)
- le facteur niveauN : le niveau d'azote, faible ou fort
- le facteur bloc : la répétition dans l'essai (2 ou 4 selon les cas)
- le facteur datejour : la date à laquelle a été effectuée la mesure.

Selon les cas, parcelle ou ligne, 1 ou 2 niveaux d'azote, on utilise différents modèles :

- Dans le cas où il y a un seul niveau d'azote (données « gie grass », données de type ligne, ou autres données de type parcelle avec données absentes sur l'un des niveaux), on utilise les deux modèles :

Un modèle à effet fixe :  $\text{valeur} = \text{variete} * \text{datejour} + \text{bloc} + \varepsilon$

Un modèle à effet mixte :  $valeur = variete + bloc + \underbrace{datejour + variete : datejour + \epsilon}_{\text{les effets aléatoires}}$

- Dans le cas où il y a deux niveaux d'azote, on teste deux modèles :  
 Un modèle à effet fixe :  $valeur = variete * niveauN * datejour + bloc + \epsilon$

Un modèle à effet mixte :  
 $valeur = variete * niveauN + bloc + \underbrace{datejour + variete : niveauN : datejour + \epsilon}_{\text{les effets aléatoires}}$

Dans le modèle à effets fixes, on teste la significativité des facteurs « variété », « niveauN », « datejour » et « bloc ».

Dans le modèle à effets aléatoires, on teste la significativité des facteurs « variété », « niveauN » et « bloc » en considérant que « datejour » est un effet aléatoire.

Tout comme précédemment, le script R (voir annexe 5) que j'ai écrit permet de produire directement un tableau de résultats directement analysable. Voici, un extrait des résultats obtenus sur la variable « rendement sec » (par soucis de présentation, je n'ai pas fait figurer dans le tableau ci-dessus toutes les variables calculées) :

Analyse de la variance pour la variable "rdtsec" pour toute l'année 2012								
Lieu	Pvalue Variété (date fixe)	Pvalue NiveauN (date fixe)	Pvalue Date (date fixe)	Pvalue Variete:NiveauN (date fixe)	Pvalue Variete:Date (date fixe)	Pvalue NiveauN:Date (date fixe)	Pvalue Variété (date aléa.)	Pvalue NiveauN (date aléa.)
gie grass	1,87E-05	-	9,55E-227	-	5,15E-03	-	1,62E-02	-
dlf	4,06E-21	1,94E-106	9,10E-198	3,78E-02	1,08E-29	1,70E-60	1,10E-01	3,16E-36
barenbrug	1,15E-18	1,67E-90	2,59E-115	1,06E-02	2,12E-36	2,27E-43	1,94E-01	8,92E-29
carneau	3,66E-05	3,89E-17	1,13E-114	5,83E-01	1,23E-12	2,16E-17	4,69E-01	2,16E-07
r2n	6,77E-44	2,45E-122	6,33E-200	1,53E-05	4,71E-30	7,67E-26	1,27E-08	1,57E-50

Les colonnes 2, 3 et 4 montrent clairement que les facteurs « variété », « niveauN » et « datejour » sont très fortement significatifs.

On constate également une très forte interaction « variété:date » (colonnes 6). Cela signifie que le facteur « variétés » n'intervient pas de la même manière au cours des différents moments observés. Par exemple, les différences de rendement observées entre deux variétés sont plus ou moins importantes selon la date d'observation.

L'interaction variété:niveauN (colonnes 5) est également assez marquée (sauf pour carneau).

Le deuxième modèle dans lequel le facteur « datejour » est considéré comme aléatoire permet d'apprécier le rapport des ordres de grandeur entre l'effet principal « variété » et l'interaction « variété:date:niveauN ». En effet, dans le cas du modèle mixte, la statistique de Fisher du facteur « variété » est égale au rapport du carré moyen du facteur « variété » et du carré moyen de l'interaction « variété:date:niveauN ».

Si ce rapport est grand par rapport à 1 (p-valeur faible), cela signifie que l'effet « variété » est plus important que l'effet de l'interaction « variété:date:niveauN ». l'effet « variété » peut être considéré comme non dépendant des dates choisies.

Au contraire, si ce rapport est proche de 1 (p-valeur grande), cela signifie que l'effet « variété » n'est

pas plus important que l'effet de l'interaction « variété:date:niveauN » : l'effet « variété » est très sensible aux conditions environnementales du temps.

Dans notre exemple plus haut, il est clair (colonne 8) que l'effet « variété » est très sensible au facteur temps, sauf pour « r2n » et « gie grass ». L'effet « niveau d'azote » quant à lui, n'est pas sensible au facteur « temps ».

Voici une partie des résultats obtenus pour les variables « INN » et « mat » :

<b>Analyse de la variance pour la variable "mat" pour toute l'année 2012</b>								
Lieu	Pvalue Variété (date fixe)	Pvalue NiveauN (date fixe)	Pvalue Date (date fixe)	Pvalue Variete:NiveauN (date fixe)	Pvalue Variete:Date (date fixe)	Pvalue NiveauN:Date (date fixe)	Pvalue Variété (date aléa.)	Pvalue NiveauN (date aléa.)
gie grass	1.27E-03	-	3.00E-102	-	3.22E-01	-	8.23E-03	-
dlf	5.37E-02	7.60E-03	7.60E-150	6.78E-01	2.75E-03	2.74E-48	3.16E-01	7.42E-31
barenbrug	3.33E-06	6.05E-126	2.75E-177	6.19E-02	2.70E-08	1.65E-38	4.16E-01	1.06E-68
carneau	3.25E-20	4.53E-105	9.01E-137	1.07E-01	1.02E-13	7.16E-16	3.44E-05	1.44E-56
r2n	1.14E-48	7.40E-100	5.92E-232	2.21E-02	2.69E-19	4.50E-16	7.18E-19	1.51E-53

Les constatations pour la variable « mat » (matière azotée totale) reste à peu près les mêmes que pour la variable « rdtsec ».

<b>Analyse de la variance pour la variable "INN" pour toute l'année 2012</b>								
Lieu	Pvalue Variété (date fixe)	Pvalue NiveauN (date fixe)	Pvalue Date (date fixe)	Pvalue Variete:NiveauN (date fixe)	Pvalue Variete:Date (date fixe)	Pvalue NiveauN:Date (date fixe)	Pvalue Variété (date aléa.)	Pvalue NiveauN (date aléa.)
gie grass	1.04E-01	-	1.08E-114	-	8.74E-01	-	1.06E-01	-
dlf	4.80E-01	5.93E-05	1.16E-63	9.89E-01	5.15E-01	2.79E-32	5.60E-01	4.43E-65
barenbrug	1.40E-02	1.15E-133	2.09E-118	4.41E-01	3.40E-05	9.30E-32	8.20E-01	9.32E-81
carneau	9.74E-11	1.67E-108	3.97E-72	2.78E-02	1.51E-19	7.41E-28	2.84E-01	7.62E-50
r2n	5.07E-26	4.88E-164	7.20E-123	2.49E-05	1.30E-11	7.28E-25	2.09E-07	1.05E-96

On constate que pour l'INN, concernant les lieux « gie grass » et « dlf », le facteur « variété » n'est pas très significatif.

Les mêmes études ont ensuite été menées sur les mêmes variables mais en ne retenant que les dates d'avril à juin correspondant à des mois pluvieux.

<b>Analyse de la variance pour "mat" pour le printemps 2012</b>								
Lieu	Pvalue Variété (date fixe)	Pvalue NiveauN (date fixe)	Pvalue Date (date fixe)	Pvalue Variete:NiveauN (date fixe)	Pvalue Variete:Date (date fixe)	Pvalue NiveauN:Date (date fixe)	Pvalue Variété (date aléa.)	Pvalue NiveauN (date aléa.)
gie grass	6.28E-02	-	8.89E-08	-	6.00E-01	-	1.02E-01	-
dlf	7.65E-04	3.99E-62	7.02E-105	1.71E-01	3.91E-02	5.91E-42	9.86E-01	8.50E-22
barenbrug	1.32E-07	1.13E-74	1.11E-90	9.73E-02	4.73E-06	3.90E-33	6.49E-01	1.11E-30
carneau	4.91E-19	3.98E-70	1.76E-87	1.26E-01	1.51E-08	1.26E-03	1.27E-07	1.39E-40
r2n	1.96E-14	8.43E-60	3.14E-75	9.29E-01	9.51E-04	3.34E-07	1.09E-06	4.40E-37

Analyse de la variance pour "rdtsec" pour le printemps 2012								
Lieu	Pvalue Variété (date fixe)	Pvalue NiveauN (date fixe)	Pvalue Date (date fixe)	Pvalue Variete:NiveauN (date fixe)	Pvalue Variete:Date (date fixe)	Pvalue NiveauN:Date (date fixe)	Pvalue Variété (date aléa.)	Pvalue NiveauN (date aléa.)
gie grass	6.67E-03	-	9.20E-127	-	1.11E-02	-	2.86E-01	-
dlf	3.40E-20	4.77E-73	6.28E-127	1.14E-01	5.19E-26	7.23E-48	5.11E-01	1.31E-18
barenbrug	3.64E-20	7.12E-55	2.91E-46	6.65E-02	1.45E-30	4.43E-08	1.08E-01	1.44E-15
carneau	2.09E-07	2.03E-05	1.06E-97	8.96E-01	2.16E-18	7.82E-06	4.41E-01	2.65E-02
r2n	2.66E-12	1.89E-58	3.20E-72	8.57E-04	5.65E-14	9.84E-02	1.71E-02	4.52E-27

Analyse de la variance pour "INN" pour le printemps 2012								
Lieu	Pvalue Variété (date fixe)	Pvalue NiveauN (date fixe)	Pvalue Date (date fixe)	Pvalue Variete:NiveauN (date fixe)	Pvalue Variete:Date (date fixe)	Pvalue NiveauN:Date (date fixe)	Pvalue Variété (date aléa.)	Pvalue NiveauN (date aléa.)
gie grass	0.236	-	9.67E-57	-	0.861	-	0.243	-
dlf	0.058	1.32E-84	1.83E-54	0.880	0.436	5.42E-33	0.990	8.02E-40
barenbrug	0.002	1.82E-75	1.02E-67	0.604	0.0004	4.37E-23	0.827	5.02E-37
carneau	5.54E-09	1.44E-75	9.86E-54	0.011	7.61E-22	0.007	0.408	5.42E-32
r2n	4.92E-07	4.40E-85	3.47E-74	0.033	0.664	9.46E-11	0.0009	8.97E-57

### 5.3. Troisième série d'analyses des données, tous lieux confondus

Dans cette troisième série de tests, en plus des trois variables étudiées précédemment, nous nous sommes intéressés à trois autres variables, portant à six le nombre total : « INN », « deltaC13 », « mat », « rdtsec », « rapportCN », « teneurC » et « teneur N ».

Par ailleurs, nous avons également limité l'étude aux données de l'année 2012.

L'étude précédente a montré que, pour un lieu donné, les facteurs « variété », « niveauN » et « datejour » étaient très fortement significatifs et qu'en général, que l'effet « variété » était très sensible au facteur temps. En revanche, le facteur « niveau d'azote » (niveauN), également très significatif, n'est pas sensible au facteur d'environnement « temps ». Dans l'ensemble, les résultats locaux sont relativement homogènes et justifient une approche multilocale.

De façon générale, les influences des facteurs temps et lieu doivent être étudiées avec attention dans le cas d'études portant sur le domaine agronomique. Par exemple, une interaction d'un facteur principal avec le facteur temps peut indiquer que le facteur principal n'intervient pas de la même manière aux cours des différentes périodes. De même, une interaction d'un facteur principal avec le facteur lieu peut indiquer des différences de comportement aux différents endroits. Une telle situation peut avoir pour conséquence de devoir formuler des conclusions différentes pour les différents lieux (ou années ou périodes).

Pour une approche toujours plus globale, nous allons maintenant effectuer des analyses tous lieux confondus dites « multilocales ».

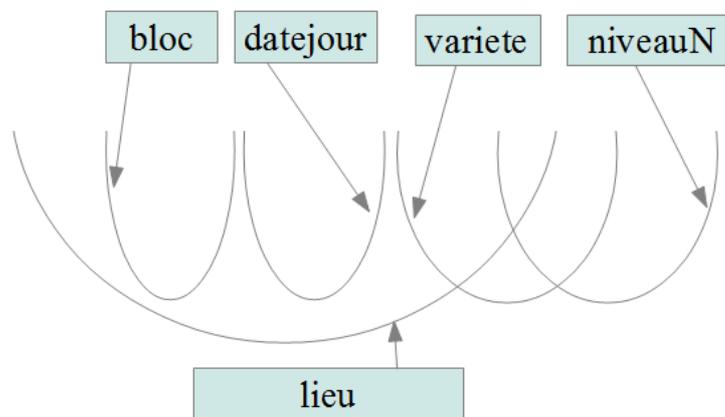
**Remarque sur le nombre de lieux étudiés et le nombre de répétitions** : de façon évidente, la qualité des résultats des analyses est d'autant meilleure que le nombre de lieux étudiés et le nombre de répétitions sont importants. Mais des raisons pratiques (notamment financières) empêchent la multiplication des expériences. Des études ont montré qu'il est préférable de privilégier le nombre de lieux au nombre de répétitions (nombre de blocs) en chacun des lieux. En pratique, le nombre de

répétitions par expérience individuelle, c'est-à-dire le nombre de blocs par expérience, est souvent limité à 2.

### a. Facteurs hiérarchisés

Les facteurs date et lieu sont normalement hiérarchisés l'un par rapport à l'autre : l'effet « date » est hiérarchisé dans l'effet « lieu ». En effet, il ne fait pas le même temps partout en France et par conséquent, une date donnée n'a pas la même signification dans un lieu ou un autre.

De même, le facteur bloc est hiérarchisé dans le facteur lieu : les caractéristiques de chaque bloc ne sont pas les mêmes dans chaque lieu. Le fait de parler du bloc n°1 et du bloc n°2 n'a pas de sens dans l'absolu : le bloc n°1 n'est pas le même dans un lieu ou dans un autre. Le modèle ainsi mis en évidence est dit partiellement hiérarchisé car par ailleurs, il y a des interactions faisant intervenir les autres facteurs. On peut représenter schématiquement la situation de la façon suivante :



Représentation schématique du modèle partiellement hiérarchisé.

La mise en œuvre dans R de modèle avec effets hiérarchisés donne lieu à énormément de discussions sur les sites spécialisés de R.

En théorie, dans R, pour de tenir compte d'effets hiérarchisés dans un modèle, on utilise l'opérateur « / ». Par exemple, si l'effet B est « hiérarchisé » dans l'effet A, on écrit : « A/B ». Mais une question qui revient assez souvent sur les sites spécialisés est la suivante : si l'effet hiérarchisé du facteur B dans le facteur A est « implicite » dans la structure des données, obtient-on les mêmes résultats que si les données sont structurées de façon à rendre « explicite » l'effet hiérarchisé ?

Exemple : A prend les modalités a1 et a2 et B les modalités b1 et b2. On considère les données suivantes :

A	B	y
a1	b1	4
a1	b2	5
a1	b1	6
a1	b2	4
a2	b1	5
a2	b2	6
a2	b1	5
a2	b2	4

Ici, le fait que l'effet B est hiérarchisé dans l'effet A est « implicite » : c'est l'utilisateur qui sait que le facteur B est hiérarchisé dans A.

Dans le tableau suivant, on a rendu explicite l'effet hiérarchisé de B dans A en remplaçant la colonne B par une colonne concaténée de A et B :

A	B	y
a1	a1-b1	4
a1	a1-b2	5
a1	a1-b1	6
a1	a1-b2	4
a2	a2-b1	5
a2	a2-b2	6
a2	a2-b1	5
a2	a2-b2	4

Le facteur B comporte alors 4 niveaux (au lieu de 2 dans une formulation implicite).

J'ai effectué plusieurs tests avec les fonctions « anova » et « Anova » (bibliothèque « car ») appliquées à des modèles calculés avec « lm » ou « lmer ».

Concernant les résultats avec « anova », il sont identiques dans les deux cas, mais avec la fonction « Anova » de type III, des différences numériques très faibles apparaissent. Cela s'explique sans doute par le fait que les modèles calculés par R dans les deux cas, même s'ils sont équivalents, ne s'écrivent pas de la même manière.

De nombreuses personnes sur les sites spécialisés sur R recommandent de rendre explicite l'effet hiérarchisé de B dans A, par exemple en remplaçant dans les données la colonne B par une colonne concaténée de A et B. Dans notre étude, nous avons suivi cette recommandation pour le facteur « bloc ».

Pour le facteur « datejour », l'effet hiérarchisé de ce facteur est déjà explicite car les dates choisies par les différents lieux sont toutes différentes les unes des autres.

J'ai néanmoins voulu comparer sur quelques exemples les résultats obtenus avec des données brutes sans formulation explicite de l'effet hiérarchisé de bloc et les résultats obtenus avec des données modifiées faisant apparaître une formulation explicite de l'effet hiérarchisé de bloc. On constate alors que les résultats obtenus sont numériquement différents mais les différences sont très très faibles, c'est-à-dire inférieures ou très inférieures à  $10^{-5}$  %.

## b. Construction des modèles et explications

Nous avons testé des modèles différents pour tenir compte de l'effet aléatoire ou non des facteurs « datejour » et « lieu ». Par ailleurs, les analyses ont été menées pour deux types de dates : année complète 2012 et mois pluvieux d'avril à juin 2012.

Les facteurs étudiés sont :

- le facteur lieu (5 lieux différents)
- le facteur variete (24 variétés différentes)
- le facteur niveauN : le niveau d'azote, faible ou fort
- le facteur bloc : la répétition dans l'essai (2 ou 4 selon les cas)

- le facteur datejour : la date à laquelle a été effectuée la mesure.

**Premier modèle dans lequel tous les modèles sont fixes :**

- Effets principaux étudiés : variete, niveauN et lieu
- Interactions étudiées : variete-niveauN, variete-lieu, niveauN-lieu, lieu-datejour, lieu-bloc, variete-niveauN-lieu, variete-lieu-datejour, niveauN-lieu-datejour, variete-niveauN-lieu-datejour

Avec R, le modèle peut alors s'écrire de la façon suivante :

```
valeur~variete*niveauN*lieu + lieu/datejour + lieu/bloc + variete:lieu:datejour +
niveauN:lieu:datejour + variete:niveauN:lieu:datejour
```

**Deuxième modèle dans lequel « lieu », « datejour » et « bloc » sont des facteurs aléatoires :**

- Effets principaux étudiés : variete et niveauN
- Interactions étudiées : variete-niveauN
- Effets aléatoires portant sur : « lieu », « datejour » et « bloc »

Avec R, le modèle peut alors s'écrire de la façon suivante :

```
valeur~variete*niveauN+(1|lieu)+(1|lieu/bloc)+(1|lieu/datejour)+ (1|datejour:variete:niveauN:lieu)
```

Concernant le premier modèle (modèle fixe), on fait appel à la fonction « lm ». Les 5 lieux ont été pris en compte. En revanche, pour le deuxième modèle avec effets aléatoires, on utilise la fonction « lmer ». Pour ce deuxième modèle, le lieu « Gie grass » n'a pas été pris en compte. En effet, les données « Gie grass » ne comportent qu'un seul niveau d'azote ce qui entraînait une erreur lors de l'exécution de la fonction « lmer ». De façon générale, la fonction « lmer » est moins sophistiquée que la fonction « lm » qui est capable de s'adapter sans renvoyer une erreur à un plus grand nombre de cas « atypiques ».

Voici quelques résultats obtenus pour les différentes variables étudiées (voir annexe 6 pour le code R utilisé) :

Variable “rdtsec”										
	Modèle Fixe							Modèle aléatoire		
	P.val Variété	P.val niveauN	P.val lieu	P.val Variété:Lieu	P.val niveauN:Lieu	P.val date:Lieu	P.val Bloc:Lieu	P.val Variété	P.val niveauN	P.val Variete:niveauN
Toute l'année 2012	0.04	~0	0	~0	~0	0	~0	0.19	~0	0.99
Printemps 2012	0.99	~0	0	0.13	0.13	0	4.28e-06	0.48	~0	0.99

Variable “INN”										
	Modèle Fixe							Modèle aléatoire		
	P.val Variété	P.val niveauN	P.val lieu	P.val Variété:Lieu	P.val niveauN:Lieu	P.val date:Lieu	P.val Bloc:Lieu	P.val Variété	P.val niveauN	P.val Variete:niveauN
Toute l'année 2012	0.035	~0	~0	~0	~0	~0	~0	0.48	~0	0.99
Printemps 2012	0.99	~0	~0	0.99	~0	~0	~0	0.82	~0	0.99

Variable “rdtsec”										
	Modèle Fixe							Modèle aléatoire		
	P.val Variété	P.val niveauN	P.val lieu	P.val Variété:Lieu	P.val niveauN:Lieu	P.val date:Lieu	P.val Bloc:Lieu	P.val Variété	P.val niveauN	P.val Variete:niveauN
Variable “mat”										
	Modèle Fixe							Modèle aléatoire		
	P.val Variété	P.val niveauN	P.val lieu	P.val Variété:Lieu	P.val niveauN:Lieu	P.val date:Lieu	P.val Bloc:Lieu	P.val Variété	P.val niveauN	P.val Variete:niveauN
Toute l'année 2012	~0	~0	~0	~0	~0	~0	1.5e-04	4.8e-03	~0	0.99
Printemps 2012	0.95	1.51e-05	~0	0.99	~0	~0	~0	6.0e-03	~0	0.99

Concernant les résultats du modèle fixe, on constate que dans tous les cas, le facteur « niveau d'azote » est très fortement significatif. Il est de même du facteur lieux. Le facteur variété est significatif sur l'année entière (moins marqué pour rdtsec et INN). Mais sur les périodes pluvieuses du printemps, ce facteur n'est plus significatif.

Plusieurs interactions (Variété:lieu, niveauN:lieu) sont également très significatives, ce qui montre que les facteurs « variétés » et « niveauN » n'interviennent pas de la même manière dans différents lieux.

Comme cela a été expliqué plus haut, le deuxième modèle dans lequel les facteurs « datejour », « bloc » et « lieu » sont considérés comme aléatoires permet d'apprécier le rapport des ordres de grandeur entre les effets principaux et les interactions.

Les valeurs élevées des p-valeur de « variete » (statistique de Fisher petite) pour « rdtsec » et « INN » indiquent que l'effet « variété » est de même ordre de grandeur que les effets d'interaction.

Les valeurs très faibles des p-valeur de « niveauN » (statistique de Fisher grande) pour « rdtsec », « INN » et « mat » indiquent au contraire que l'effet « niveauN » est plus important que les effets d'interaction. L'effet « niveauN » est peu sensible à l'effet environnemental.

J'ai ensuite mené une seconde série de tests sur les variables « deltaC13 », « rapportCN », « teneurC » et « teneur N ». Les modèles utilisés étaient les mêmes mais les lieux concernés étaient restreints à « carneau » et « r2n » car les données concernées n'étaient pas encore disponibles pour les autres lieux. Voici une partie des résultats obtenus (voir annexe 7 pour le code R utilisé) :

Variable “deltaC13”										
	Modèle Fixe							Modèle aléatoire		
	P.val Variété	P.val niveauN	P.val lieu	P.val Variété:Lieu	P.val niveauN:Lieu	P.val date:Lieu	P.val Bloc:Lieu	P.val Variété	P.val niveauN	P.val Variete:niveauN
Toute l'année 2012	~0	0.08	~0	~0	~0	-	0.019	~0	0.41	0.99

Variable “rapportC/N”										
	Modèle Fixe							Modèle aléatoire		
	P.val Variété	P.val niveauN	P.val lieu	P.val Variété:Lieu	P.val niveauN:Lieu	P.val date:Lieu	P.val Bloc:Lieu	P.val Variété	P.val niveauN	P.val Variete:niveauN
Toute l'année 2012	~0	~0	~0	~0	~0	-	0.019	0.192	~0	0.98

Variable “teneurC”										
	Modèle Fixe							Modèle aléatoire		
	P.val Variété	P.val niveauN	P.val lieu	P.val VariétéLieu	P.val niveauN:Lieu	P.val date:Lieu	P.val Bloc:Lieu	P.val Variété	P.val niveauN	P.val Variete:niveauN
Toute l'année 2012	0.88	0.89	3.9e-05	0.98	~0	-	~0	0.88	0.89	0.99

Variable “teneurN”										
	Modèle Fixe							Modèle aléatoire		
	P.val Variété	P.val niveauN	P.val lieu	P.val VariétéLieu	P.val niveauN:Lieu	P.val date:Lieu	P.val Bloc:Lieu	P.val Variété	P.val niveauN	P.val Variete:niveauN
Toute l'année 2012	~0	~0	~0	~0	~0	-	0.033	0.65	1.04	0.999

Parmi ces variables, la variable « deltaC13 » est celle à laquelle nous accordons le plus d'attention. Il apparaît que le facteur « variété » est très fortement significatif, de même que l'interaction « variete:lieu ». De plus la valeur faible de la p-valeur de « variete » dans le cas du modèle avec effets aléatoires (statistique de Fisher grande) indique que l'effet « variété » est plus important que l'effet d'interaction.

#### 5.4. Comparaisons multiples et calcul des lsmeans

Suite aux différentes études que nous avons menées, il en résulte, dans le cas du modèle à effets fixes, que le facteur variété est significatif sur l'année 2012 pour les variables qui nous intéressent plus particulièrement : « INN », « deltaC13 », « mat » et « rdtsec ». Mais le rejet de l'hypothèse d'égalité des moyennes soulève alors la question d'identifier les différentes variétés qui s'écartent de la moyenne générale.

De nombreuses méthodes existent connues sous le nom de tests de comparaisons des moyennes, chacune présentant des avantages et des inconvénients.

Parmi les méthodes de comparaisons de moyenne, certaines ne sont valables que dans le cas de données équilibrées d'échantillons de même effectif, elles n'ont donc pas retenu mon attention.

Par ailleurs, dans notre cas, nous ne disposons pas de « témoin » par rapport auquel se référer et toutes les modalités sont sur un même pied d'égalité. Nous nous orientons donc vers des méthodes dites de comparaisons multiples où on procède à toutes les comparaisons deux à deux des moyennes.

Même avec ces restrictions, il y a encore de nombreuses méthodes possibles mais deux méthodes reviennent assez souvent dans les différentes lectures : la méthode de la plus petite différence significative et la méthode de Newman et Keuls.

##### a. La méthode de la plus petite différence significative

On présente ici la méthode dans le cas d'une analyse de la variance à effets fixes. On suppose que le facteur à étudier présente  $p$  modalités. On suppose également que l'hypothèse nulle  $H_0: m_1 = m_2 = \dots = m_p$  a été rejetée. Pour déterminer les égalités à éliminer, on compare les moyennes deux à deux grâce au test de Student. Cela nécessite donc pour les  $p(p-1)/2$  couples

$(i, i')$ , l'utilisation de la formule : 
$$t_{obs} = \frac{|\bar{x}_i| - |\bar{x}_{i'}|}{\sqrt{\frac{SCE_i + SCE_{i'}}{n_i + n_{i'} - 2} \left( \frac{1}{n_i} + \frac{1}{n_{i'}} \right)}}$$
 (ou bien, si les effectifs sont égaux : 
$$t_{obs} = \frac{|\bar{x}_i| - |\bar{x}_{i'}|}{\sqrt{\frac{SCE_i + SCE_{i'}}{n(n-1)}}}$$
 ).

Mais on a supposé que les variances de toutes les populations sont égales, au lieu de calculer à chaque fois la variance  $\hat{\sigma}^2 = \frac{SCE_i + SCE_{i'}}{n_i + n_{i'} - 2}$ , on peut utiliser dans tous les cas l'estimation globale fournie par le carré moyen résiduel :  $\hat{\sigma}^2 = CM_r$ .

On rejette donc l'hypothèse d'égalité de deux moyennes chaque fois que

$$t_{obs} = \frac{|\bar{x}_i - \bar{x}_{i'}|}{\sqrt{CM_r \left( \frac{1}{n_i} + \frac{1}{n_{i'}} \right)}} \geq t_{1-\alpha/2}$$

De plus, lorsque les effectifs sont égaux, on a  $\frac{1}{n_i} + \frac{1}{n_{i'}} = \frac{2}{n}$ . On peut donc calculer une fois pour toutes la quantité  $t_{1-\alpha/2} \sqrt{2CM_r/n}$  et rejeter l'hypothèse d'égalité des moyennes chaque fois que la différence  $|\bar{x}_i - \bar{x}_{i'}|$  est supérieur ou égale à cette quantité. Cette expression est appelée la plus petite différence significative.

Dans R, cette méthode est disponible avec la fonction « `LSD.test()` » dans la bibliothèque « `agricolae` ».

Cette méthode est très simple et s'effectue en une seule étape mais elle présente un inconvénient important : comme on effectue  $p(p-1)/2$  tests d'égalité de deux moyennes et que chacun de ces tests a un risque de première espèce égale à  $\alpha$ , le risque global de première espèce devient très supérieur à  $\alpha$ . Après discussion avec Monsieur Sampoux, il s'avère préférable d'utiliser une autre méthode qui permettent de maîtriser ce risque d'erreur.

C'est pourquoi je me suis orienté vers la méthode de Newman et Keuls. Contrairement à la méthode précédente, elle s'applique en plusieurs étapes successives.

## b. La méthode de Newman et Keuls

On calcule tout d'abord des valeurs (plus petites amplitudes significatives) relatives à des groupes de deux moyennes, de trois moyennes, . . . , de  $p$  moyennes. Par exemple, dans les cas simples où les effectifs sont égaux, ces amplitudes sont définies par  $t'_{1-\alpha/2} \sqrt{2CM_r/n}$  où les valeurs  $t'_{1-\alpha/2}$  peuvent être trouvées dans des tables disponibles sur Internet. Ces tables donne les valeurs de  $t'_{1-\alpha/2}$  en fonction du nombre de populations prises en considération et du nombre de degrés de liberté du carré moyen résiduel. Elle sont disponibles avec  $\alpha=0.01$  ou  $\alpha=0.05$ .

Dans le groupe de  $p$  moyennes, on compare ensuite l'amplitude observée avec la valeur de la « plus petite amplitude significative » correspondante, et on procède de la même manière pour les différents groupes de  $p-1$  moyennes, de  $p-2$  moyennes, etc., les moyennes étant toujours rangées soit par ordre croissant, soit par ordre décroissant. Pour les groupe pour lesquels l'amplitude est inférieure à la plus petite amplitude significative correspondante, le processus s'arrête et le groupe

est considéré comme homogène (et les moyennes théoriques correspondantes sont considérées comme toutes égales). Ainsi les comparaisons pour des groupes de  $p$ ,  $p - 1$ ,  $p - 2$ , ... moyennes sont effectivement réalisées avec un risque d'erreur  $\alpha$ . Ainsi la méthode de Newman et Keuls garantit un risque global de première espèce égal à  $\alpha$ . En revanche, elle s'applique plus difficilement « à la main » aux cas des échantillons d'effectifs inégaux. Heureusement, cette méthode est disponible dans R avec la fonction «SNK.test()».

L'inconvénient général des méthodes de comparaison des moyenne concerne le fait qu'elles conduisent fréquemment à des conclusions difficiles à interpréter du fait que les groupes mis en évidence par ces méthodes peuvent se chevaucher.

### c. Calcul des moyennes ajustées

Pour compléter ces tests et préparer l'ACP, il m'a également été demandé de calculer les « moyennes ajustées » (« lsmean » : least squares means ) pour chacune des variables « INN », « deltaC13 », « mat », « rdtsec », « rapportCN », « teneurC » et « teneur N ». Pour ce faire, dans R, on dispose par exemple de la fonction lsmeans() de la bibliothèque lsmeans().

Dans le cas où il n'y a pas de données manquantes, le calcul des moyennes ajustées aboutit au même résultat qu'une moyenne arithmétique classique. Dans le cas contraire, il est préférable d'utiliser les moyennes ajustées qui reflètent mieux le modèle qui est ajusté aux données.

L'exemple suivant illustre dans un cas très simple la différence entre les deux moyennes dans le cas de données manquantes : considérons les données suivantes à un facteur A avec 3 niveaux (1, 2 et 3) et 3 répétitions de chaque.

<b>Facteur A</b>			
Répétitions\niveau	<b>1</b>	<b>2</b>	<b>3</b>
1	<b>4</b>	-	<b>4</b>
2	<b>6</b>	<b>3</b>	<b>2</b>
3	<b>2</b>	<b>5</b>	<b>3</b>
Sommes	12	8	9
Moyenne	4	4	3

La moyenne arithmétique classique est égale à :  $\frac{4+ 6+ 2+ 3+ 5+ 4+ 2+ 3}{8} = \frac{29}{8} = 3.625$

La moyenne ajustée est égale à :  $\frac{4+ 4+ 3}{3} = \frac{11}{3} = 3.667$

Dans le cas général, plus complexe, la moyenne ajustée est « reconstituée » à partie des estimations des paramètres du modèle linéaire.

### d. Les résultats

J'ai mené toute une série de tests sous forme de programmes R permettant le calcul des moyennes ajustées et la comparaison des moyennes dans divers cas : les tests portaient sur les variables « INN », « deltaC13 », « mat », « rdtsec », « rapportCN », « teneurC » et « teneur N », sur deux types de périodes (année 2012 entière et printemps 2012) et également dans les trois cas suivants :

apports d'azote faibles et forts confondus, apport faible d'azote et apport fort d'azote.

Voici par exemple les résultats obtenus pour la variable « mat » (quantité totale de matière azotée exprimée en pourcentage de la matière sèche) sur l'année 2012 :

### Variable « mat » - Année 2012

Résultats pour niveaux azote faible et fort confondus			Résultats pour niveau azote faible			Résultats pour niveau azote fort		
Variété	Moyennes ajustées	Comparaison moyennes	Variété	Moyennes ajustées	Comparaison moyennes	Variété	Moyennes ajustées	Comparaison moyennes
centre	18.11371	a	centre	15.80973	a	centre	19.5345	a
danemark	17.88847	ab	cantalou	15.37838	b	danemark	19.5165	a
barlet	17.77867	abc	danemark	15.31789	bc	barlet	19.35783	ab
delphin	17.71561	abcd	bretagne	15.29263	bcd	delphin	19.29217	abc
bretagne	17.63163	bcde	barlet	15.28526	bcd	bartango	19.16167	abcd
bartango	17.59102	bcdef	delphin	15.22632	bcde	bretagne	19.113	abcde
cantalou	17.57804	bcdef	loporello	15.19053	bcdef	hongrie	19.10322	abcde
lactal	17.46788	bcdefg	bartango	15.11105	bcdef	lactal	19.0675	abcdef
loporello	17.46724	bcdefg	gagny	15.08368	bcdefg	cantalou	18.9345	abcdef
hongrie	17.45639	bcdefg	roy	15.08237	bcdefg	milca	18.93067	abcdef
milca	17.4248	bcdefg	milca	15.04711	bcdefgh	loporello	18.90917	abcdef
roy	17.285	cdefg	lactal	15.00692	bcdefgh	roy	18.68	bcdef
oceal	17.24122	cdefg	oceal	15.00368	bcdefgh	oceal	18.65833	bcdef
gagny	17.19816	defg	aberavon	14.91053	bcdefgh	clerpin	18.59433	bcdef
aberavon	17.11714	efg	hongrie	14.89947	bcdefgh	gagny	18.53733	bcdef
clerpin	17.06778	fg	ponant	14.85026	bcdefgh	stefani	18.532	bcdef
ponant	17.04724	fg	alutus	14.78553	cdefgh	aberavon	18.51467	bcdef
stefani	16.99434	g	limbos	14.77211	cdefgh	barutti	18.49167	cdef
mara	16.98122	g	clerpin	14.71923	defgh	ponant	18.43867	cdef
base	16.98064	g	mara	14.70553	defgh	mara	18.4225	def
limbos	16.97245	g	base	14.67393	efgh	expo	18.39217	def
barutti	16.93867	g	stefani	14.62872	fgh	limbos	18.366	def
alutus	16.89051	g	expo	14.50474	gh	base	18.2724	ef
expo	16.8848	g	barutti	14.48658	h	alutus	18.22367	f

Dans le tableau ci-dessus, chaque lettre représente un groupe de variétés homogènes. Bien qu'on ait utilisé les mêmes lettres dans les différentes colonnes (niveau d'azote faible et fort confondus, niveau faible et niveau fort), ils ne s'agit pas des mêmes groupes.

On constate que les résultats en situation de faible apport d'azote sont différents de ceux en situation de fort apport d'azote. Par exemple, la variété « centre » se distingue des autres variétés en faible apport d'azote, alors qu'en fort apport d'azote, le groupe « a » (« centre », « danemark », « barlet »,

« delphin », « bartango », « bretagne », « hongrie », « lactal », « cantalou », « milca », « loporello ») forme un groupe homogène d'un point de vue statistique (au seuil de 5%).

La comparaison des moyennes est évidemment particulièrement intéressante en situation de faible apport d'azote puisque qu'on est alors dans une démarche de développement durable.

Voici sans commentaire les résultats concernant la variable « Rdtsec » (rendement en matière sèche) :

**Variable « rdtsec » - Année 2012**

Résultats pour niveaux azote faible et fort confondus		
Variété	Moyennes ajustées	Comparaison moyennes
barutti	2.06691	a
base	2.058038	a
clerpin	2.03162	ab
mara	2.01412	ab
gagny	2.013051	ab
ponant	2.01087	ab
roy	2.00093	abc
stefani	1.99412	abc
alutus	1.98897	abc
limbos	1.98857	abc
expo	1.98619	abc
cantalou	1.96628	abc
aberavon	1.96567	abc
lactal	1.95926	abc
milca	1.95693	abc
bartango	1.95653	abc
loporello	1.94811	abc
delphin	1.94231	abc
barlet	1.92025	bcd
oceal	1.91623	bcd
hongrie	1.87544	cd
centre	1.83571	d
bretagne	1.83333	d
danemark	1.71927	e

Résultats pour niveau azote faible		
Variété	Moyennes ajustées	Comparaison moyennes
base	1.9141	a
barutti	1.848175	ab
expo	1.821025	ab
bartango	1.798525	abc
ponant	1.7917	abc
mara	1.777975	abcd
alutus	1.766575	abcd
stefani	1.765725	abcd
roy	1.757475	abcd
delphin	1.74855	abcde
milca	1.74055	abcde
clerpin	1.7396	abcde
limbos	1.73605	abcde
loporello	1.72355	bcde
aberavon	1.717825	bcde
cantalou	1.716675	bcde
gagny	1.710075	bcde
lactal	1.70605	bcde
barlet	1.670675	bcdef
hongrie	1.613	cdef
oceal	1.6002	def
bretagne	1.569625	ef
centre	1.5322	f
danemark	1.421425	g

Résultats pour niveau azote fort		
Variété	Moyennes ajustées	Comparaison moyennes
clerpin	2.2263	a
gagny	2.218458	a
barutti	2.212733	a
mara	2.17155	ab
roy	2.163233	ab
ponant	2.156983	ab
limbos	2.156917	ab
stefani	2.146383	ab
base	2.1444	ab
alutus	2.137233	ab
cantalou	2.132683	ab
aberavon	2.1309	ab
lactal	2.128067	ab
oceal	2.126917	ab
milca	2.101183	ab
loporello	2.097817	ab
expo	2.0963	ab
barlet	2.086633	ab
delphin	2.071483	ab
bartango	2.061867	ab
hongrie	2.0504	abc
centre	2.03805	abc
bretagne	2.009133	bc
danemark	1.917833	c

Voici à présent les résultats concernant la variable « INN » (indice de nutrition azotée)

## Variable « INN» - Année 2012

Résultats par niveaux azote faibles et forts		
Variété	Moyennes ajustées	Comparaison moyennes
centre	0.70585	a
barlet	0.69897	ab
bartango	0.69470	ab
milca	0.69433	abc
gagny	0.69020	abcd
clerpin	0.6882	abcd
delphin	0.68572	abcd
cantalou	0.68487	abcd
lactal	0.68425	abcd
loporello	0.68396	abcd
oceal	0.68313	abcd
aberavon	0.6818	abcd
barutti	0.68085	abcd
stefani	0.6807	abcd
ponant	0.68064	abcd
mara	0.67915	abcd
danemark	0.67785	abcd
roy	0.6758	bcd
bretagne	0.67573	bcd
hongrie	0.67559	bcd
alutus	0.67521	bcd
limbos	0.66582	cd
expo	0.66413	d
base	0.64	e

Résultats par niveaux azote faibles		
Variété	Moyennes ajustées	Comparaison moyennes
bartango	0.57192	a
gagny	0.56989	a
centre	0.56927	a
milca	0.56921	a
ponant	0.56760	a
barlet	0.56473	a
stefani	0.56107	a
clerpin	0.56079	a
alutus	0.55994	a
mara	0.559	a
aberavon	0.55928	a
loporello	0.55892	a
barutti	0.55744	a
roy	0.55707	a
oceal	0.55536	a
delphin	0.55468	a
cantalou	0.55378	a
lactal	0.55258	a
expo	0.54871	a
bretagne	0.54644	a
limbos	0.54547	a
hongrie	0.54536	a
danemark	0.54297	a
base	0.51714	b

Résultats par niveaux azote forts		
Variété	Moyennes ajustées	Comparaison moyennes
centre	0.79008	a
barlet	0.784	a
milca	0.77353	ab
bartango	0.77246	ab
clerpin	0.77101	ab
lactal	0.76983	ab
delphin	0.76871	ab
cantalou	0.7679	ab
gagny	0.76769	ab
oceal	0.76405	ab
danemark	0.76328	ab
loporello	0.76316	ab
aberavon	0.75955	ab
hongrie	0.75947	ab
barutti	0.75901	ab
stefani	0.75861	ab
bretagne	0.75761	ab
mara	0.75493	ab
ponant	0.75223	ab
roy	0.75115	ab
alutus	0.74821	ab
limbos	0.74205	b
expo	0.73723	b
base	0.7088	c

Rappelons que l'INN permet de quantifier un état de carence ou de suralimentation en azote, en cours de végétation. L'INN renseigne donc sur la contrainte azotée subie par la plante.

En situation de faible apport ou de fort apport, on constate qu'il n'existe pas selon les différentes variétés des variations significatives pour l'INN sur l'année 2012 (quand on regarde uniquement les résultats pour le printemps 2012, les variations sont un peu plus significatives). Néanmoins ceci apporte une première réponse dans le cas multilocal à une des quatre questions posées au début de l'étude.

Le calcul de l'INN fait intervenir le rendement sec (« rdtsec ») et la matière azotée (« mat »). L'INN cumule donc les erreurs expérimentales affectant ces deux caractères. C'est pourquoi l'erreur sur l'INN est importante, ce qui explique que le test SNK ne montre pas de différences importantes.

Voici à présent les résultats concernant la variable « deltaC13 » (indice de nutrition azotée)

### Variable « deltaC13 » - Année 2012

Résultats par niveaux azote faibles et forts			Résultats par niveaux azote faibles			Résultats par niveaux azote forts		
Variété	Moyennes ajustées	Comparaison moyennes	Variété	Moyennes ajustées	Comparaison moyennes	Variété	Moyennes ajustées	Comparaison moyennes
clerpin	-30.09475	a	clerpin	-30.02025	a	base	-30.0977	a
base	-30.107	a	base	-30.11625	ab	lactal	-30.131	ab
lactal	-30.20786	ab	limbos	-30.15475	ab	clerpin	-30.1692	ab
roy	-30.23412	ab	cantalou	-30.203	ab	roy	-30.1952	ab
limbos	-30.2665	ab	lactal	-30.2655	ab	loporello	-30.3487	abc
cantalou	-30.29137	ab	roy	-30.273	ab	delphin	-30.349	abc
delphin	-30.34925	abc	delphin	-30.3495	abc	limbos	-30.3782	abcd
loporello	-30.42687	bc	loporello	-30.505	bcd	cantalou	-30.3797	abcd
gagny	-30.54525	cd	barlet	-30.662	cde	gagny	-30.4242	abcd
barlet	-30.64487	de	gagny	-30.66625	cde	stefani	-30.515	bcde
stefani	-30.71025	def	alutus	-30.74325	de	barlet	-30.6277	cde
alutus	-30.7145	def	ponant	-30.7675	de	bretagne	-30.65	cdef
oceal	-30.766	def	oceal	-30.79975	de	alutus	-30.6857	cdefg
bartango	-30.79563	def	aberavon	-30.822	de	oceal	-30.7322	cdefgh
aberavon	-30.79713	def	bartango	-30.841	de	bartango	-30.7502	cdefgh
ponant	-30.821	def	danemark	-30.8855	de	aberavon	-30.7722	defgh
bretagne	-30.82525	def	centre	-30.8955	de	centre	-30.7762	defgh
centre	-30.83588	ef	stefani	-30.9055	de	expo	-30.8677	efgh
danemark	-30.8815	efg	expo	-30.9205	de	ponant	-30.8745	efgh
expo	-30.89412	efg	barutti	-30.96	de	danemark	-30.877	efgh
barutti	-30.93137	efg	milca	-30.96775	de	barutti	-30.9027	efgh
milca	-31.0085	fg	bretagne	-31.0005	e	milca	-31.0492	fgh
hongrie	-31.1025	g	mara	-31.07	e	hongrie	-31.0887	gh
mara	-31.10675	g	hongrie	-31.11625	e	mara	-31.1435	h

En situation de faible apport ou de fort apport, on constate qu'il existe selon les différentes variétés des variations assez significatives pour le deltaC13 sur l'année 2012.

#### e. Remarque sur la fonction SNK.test

Pour les différents tests qui précèdent je me suis basé sur le modèle fixe. J'ai également fait des tests de comparaison sur un modèle avec effets aléatoire mais j'ai rencontré le problème suivant : la fonction SNK.test() ne fonctionne pas avec des modèles mixtes. J'ai contourné le problème en installant la bibliothèque « multcomp » et en utilisant les fonctions glkt() et cld() (voir la documentation de R).

## 6. ACP

La dernière partie de mon stage a consisté à faire des ACP (analyse en composantes principales) sur plusieurs ensembles de données, ceci afin de mettre en évidence des liens entre certains paramètres de l'étude.

L'objectif général d'une ACP est de résumer un tableau de données individus  $\times$  variables et d'étudier les ressemblances entre individus du point de vue de l'ensemble des variables afin de dégager des profils d'individus. Comme cela a déjà été dit, comme nous disposons de peu de données sur la variable « deltaC13 », nous avons surtout porté notre attention sur l'INN.

Ici, les individus sont les 24 variétés sur lesquelles portent l'étude.

### 6.1. Choix des variables

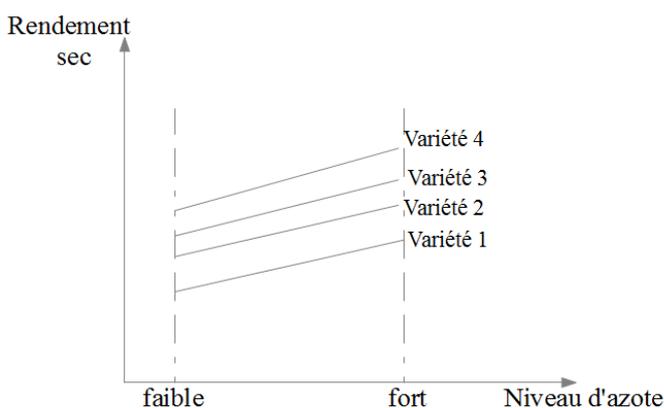
Les variables étudiées sont de deux types : d'une part des moyennes ajustées calculées dans la partie précédente et d'autre part des coefficients d'interaction (issus des modèles précédemment étudiés) « variété/niveau d'azote faible » dans le cas des trois variables « rdtsec », « INN » et « mat ».

L'intérêt porté aux rendements moyens ajustés est évident : plus la moyenne ajustée d'une variété est importante, plus le rendement moyen de la variété sera élevé.

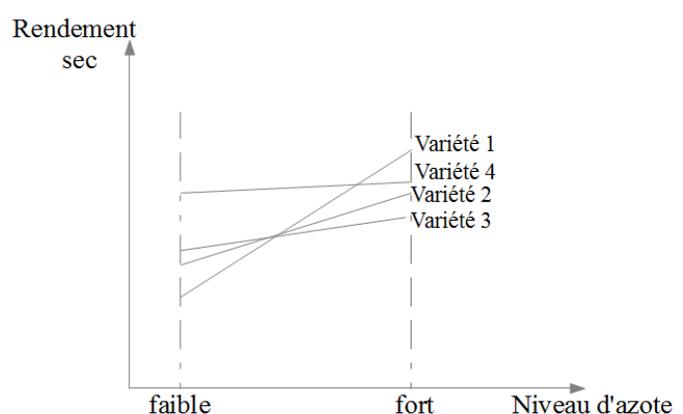
Mais les agronomes sont également intéressés par d'autres critères de sélection : une variété ne présentant pas trop de différences de rendement selon le niveau d'azote apporté est préférée à une variété très sensible au niveau d'azote. C'est pourquoi on s'intéresse aux coefficients d'interaction « variété:niveau d'azote » car ils mettent en évidence des différences de comportement entre les faibles apports et les forts apports d'azote.

Par exemple, dans le cas de « rdtsec », si les coefficients d'interaction « variété:niveau d'azote » sont proches de 0, cela signifie que les différences de rendements observées selon les variétés ne dépendent pas (ou peu) du niveau d'azote et seront à peu-près égales dans les deux cas. Au contraire, si les coefficients sont importants, les différences de rendements observées seront très inégales.

On peut illustrer la situation de la façon suivante :



Cas de coefficients d'interaction faibles



Cas de coefficients d'interaction forts  
(sauf variété 3)

Dans R, j'ai utilisé une paramétrisation (contraste) de sorte que la somme des coefficients pour un facteur donné est nulle. Le coefficient d'interaction « variété:niveau d'azote faible » est donc

l'opposé du coefficient « variété:niveau d'azote fort ». Il est donc inutile de prendre comme variables l'un et l'autre de ces coefficients. C'est donc le choix du coefficient d'interaction « variété:niveau d'azote faible » qui a été fait pour les différentes ACP.

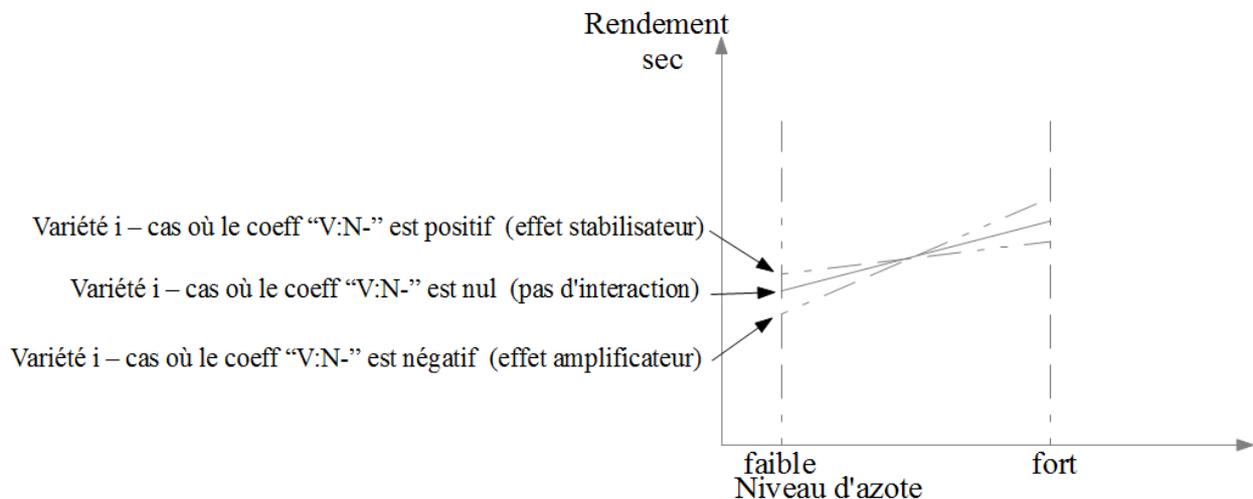
Pour une variété dont le coefficient d'interaction « variété:niveau d'azote » est proche de zéro, son rendement peut être prédit par la somme des effets additifs principaux des facteurs « variétés » et « niveau d'azote ». Lorsque le coefficient d'interaction « variété:niveau d'azote » est différents de zéro pour une variété, cette variété présentera des différences de rendement entre les deux niveaux d'azote, plus ou moins importantes que celles prédites par la somme des effets principaux selon le signe du coefficient d'interaction.

Nous allons maintenant expliquer pourquoi des valeurs positives du coefficient d'interaction « variété:niveau d'azote faible » indiquent un effet stabilisateur de l'interaction alors que des valeurs négatives révèlent un effet amplificateur :

Pour une variété donnée, un niveau faible d'azote aura pour conséquence un rendement moins important que le niveau fort (effet principal de l'azote), mais si le coefficient d'interaction « variété:niveau d'azote faible » est positif et élevé, cela signifie que l'interaction a un effet « bonus » sur le rendement qui vient en partie compenser l'effet principal de baisse. Au final, on a un effet stabilisateur de l'interaction.

Inversement, si le coefficient d'interaction « variété:niveau d'azote faible » est négatif, l'interaction aura un effet négatif sur le rendement, amplifiant ainsi l'effet principal de baisse.

On peut illustrer ce phénomène sur le schéma suivant :



Sur notre dessin, pour la variété « i » donnée, un coefficient d'interaction « variété:niveau d'azote faible » positif correspond à une pente faible. Plus ce coefficient est grand, plus la pente est faible. Au contraire, coefficient négatif correspond à une pente forte.

On pourrait bien sûr faire les mêmes remarques dans le cas des variables « mat » et « INN ».

En combinant les variables de type « moyenne » et celles de type « coefficients d'interaction » dans les différentes ACP, on espère par exemple mettre en évidence des critères permettant de distinguer des variétés ayant un bon rendement et/ou ayant une bonne stabilité.

A terme, il s'agit de recombinaison génétiquement des variétés ayant à la fois un bon rendement et une bonne stabilité par rapport à l'apport d'azote.

J'ai ainsi mené plusieurs types d'ACP : par lieu, par variable et « multilocales ».

Dans le cas des ACP par lieu (« Barenbrug », « Carneau » et « Dlf »), on s'est intéressé aux liens entre le rendement sec et l'INN. les variables retenues étaient les suivantes :

- Les moyennes ajustées de « rdtsec » des différentes variétés pour chaque niveau d'azote,
- Les moyennes ajustées d'« INN » des différentes variétés pour chaque niveau d'azote,
- Les coefficients de l'interaction « variété:niveau d'azote faible » pour le modèle ajusté de l'anova sur la variable « rdtsec » .
- Les coefficients de l'interaction « variété:niveau d'azote faible » pour le modèle ajusté de l'anova sur la variable « INN » .

Selon les lieux, on a retenu certaines dates et dans plusieurs cas, il a fallu faire des choix car le nombre de variables devait rester inférieur au nombre d'individus étudiés (les variétés). Ces choix ont été réfléchis en examinant les corrélations entre les variables et en éliminant les variétés peu corrélées aux autres.

	Moyennes ajustées « rdtsec » « niveau d'azote fort » et « niveau d'azote faible»	Moyennes ajustées « INN » « niveau d'azote fort » et « niveau d'azote faible»	Coefficients de l'anova « rdtsec » pour l'interaction « variété:niveau d'azote faible ».	Coefficients de l'anova « INN » pour l'interaction « variété:niveau d'azote faible »
« Barenbrug » (23 variables)	10/05/2012 28/06/2012 26/07/2012 24/10/2012	10/05/2012 01/06/2012 28/06/2012 26/07/2012 24/10/2012		10/05/2012 01/06/2012 28/06/2012 26/07/2012 24/10/2012
« Carneau » (20 variables)	30/04/2012 25/05/2012 25/06/2012 22/08/2012	30/04/2012 25/05/2012 25/06/2012 22/08/2012		30/04/2012 25/05/2012 25/06/2012 22/08/2012
« dlf » (21 variables)	16/05/2012 19/06/2012 18/07/2012 30/11/2012	16/05/2012 19/06/2012 18/07/2012	16/05/2012 19/06/2012 18/07/2012 30/11/2012	16/05/2012 19/06/2012 18/07/2012
« R2n » (22 variables)	26/09/2011 10/05/2012 03/07/2012 25/07/2012 28/08/2012 28/09/2012	10/05/2012 03/07/2012 25/07/2012 28/08/2012 28/09/2012		

Dans le cas de l'ACP pour « Gie grass » les variables retenues étaient les suivantes :

	Moyennes ajustées « rdtsec »	Moyennes ajustées « INN »
« Gie grass » (10 variables)	03/04/2012 19/05/2012 14/06/2012 18/07/2012 26/10/2012	03/04/2012 19/05/2012 14/06/2012 18/07/2012 26/10/2012

Pour le lieu « r2n », on a également procédé à une ACP prenant en compte la variable deltaC13 :

	Moyennes ajustées « rdtsec » « niveau d'azote fort » et « niveau d'azote faible »	Moyennes ajustées « deltaC13 » « niveau d'azote fort » et « niveau d'azote faible »
« R2n » (22 variables)	26/09/2011 12/04/2012 10/05/2012 05/06/2012 03/07/2012 25/07/2012 28/08/2012 28/09/2012 23/10/2012	10/05/2012 28/08/2012

J'ai également effectué des ACP avec plusieurs lieux pour étudier les associations entre caractères observés sur des lieux différents :

« ACP INN 2012 » et « ACP INN printemps 2012 » : ces ACP ont pour objectif d'étudier les relations entre rendement sec et INN, d'une part sur l'ensemble de l'année, d'autre part sur observations du printemps seulement. Les variables retenues étaient (22 variables) :

- les moyennes ajustées de « rdtsec » pour gie grass et les moyennes ajustées de « rdtsec » pour les quatre autres lieux par niveau faible d'azote, soit 9 variables.
- Les moyennes ajustées d'« INN » pour gie grass et les moyennes ajustées de « INN » pour les quatre autres lieux par niveau faible d'azote, soit 9 variables.
- Les coefficients du modèle ajusté de l'anova sur la variable « INN » pour l'interaction « variété:niveau d'azote faible » pour les lieux sauf gie grass, soit 4 variables

« ACP mat 2012 » et « ACP mat printemps 2012 » : ces ACP ont pour objectif d'étudier les relations entre rendement sec et matière azotée, d'une part sur l'ensemble de l'année, d'autre part sur observations du printemps seulement. Les variables retenues étaient (22 variables) :

- les moyennes ajustées de « rdtsec » pour gie grass et les moyennes ajustées de « rdtsec » pour les quatre autres lieux par niveau faible d'azote, soit 9 variables.
- Les moyennes ajustées de « mat » pour gie grass pour le niveau d'azote fort et les moyennes ajustées de « mat » pour les quatre autres lieux par niveau faible d'azote, soit 9 variables.
- Les coefficients du modèle ajusté de l'anova sur la variable « mat » pour l'interaction « variété:niveau d'azote faible » pour les lieux sauf gie grass, soit 4 variables

Enfin, j'ai réalisé une dernière ACP multilocale destinées à étudier les relations entre variables à partir des moyennes multilocales des variétés. Les variables retenues étaient :

- pour l'année 2012, les moyennes ajustées de « rdtsec », « INN », « mat » et « deltaC13 » pour tous les lieux sauf gie grass par niveau faible d'azote et par niveau fort d'azote, soit huit variables,
- la même chose pour le printemps 2012, soit huit variables,
- pour l'année 2012, les moyennes ajustées de « deltaC13 » (niveaux forts et faibles d'azote),
- Les coefficients du modèle ajusté de l'« anova multilocale » pour l'année 2012 pour les variables « rdtsec », « mat » et « INN », soit 3 variables,
- la même chose pour le printemps 2012, soit 3 autres variables.

## 6.2. ACP et R

La fonction « princomp » de R permet de faire des ACP. C'est cette fonction que j'ai utilisée au début mais elle s'est avérée inadaptée dans le cas de données manquantes.

J'ai alors utilisé la fonction « PCA » de la librairie « FactoMineR ». Un site internet<sup>12</sup> est dédié à cette librairie et de nombreux exemples y sont disponibles.

## 6.3. Résultats

### a. ACP par lieu (voir annexe 8)

Pour que les résultats d'une ACP soient intéressants, il faut choisir un nombre d'axes factoriels suffisants pour conserver un maximum d'inertie mais pas trop grand pour qu'on puisse interpréter les résultats.

Dans le cas des ACP par lieu, les résultats obtenus ne permettent pas de dégager des interprétations solides car, pour conserver une inertie suffisante, il est nécessaire de conserver un grand nombre d'axes.

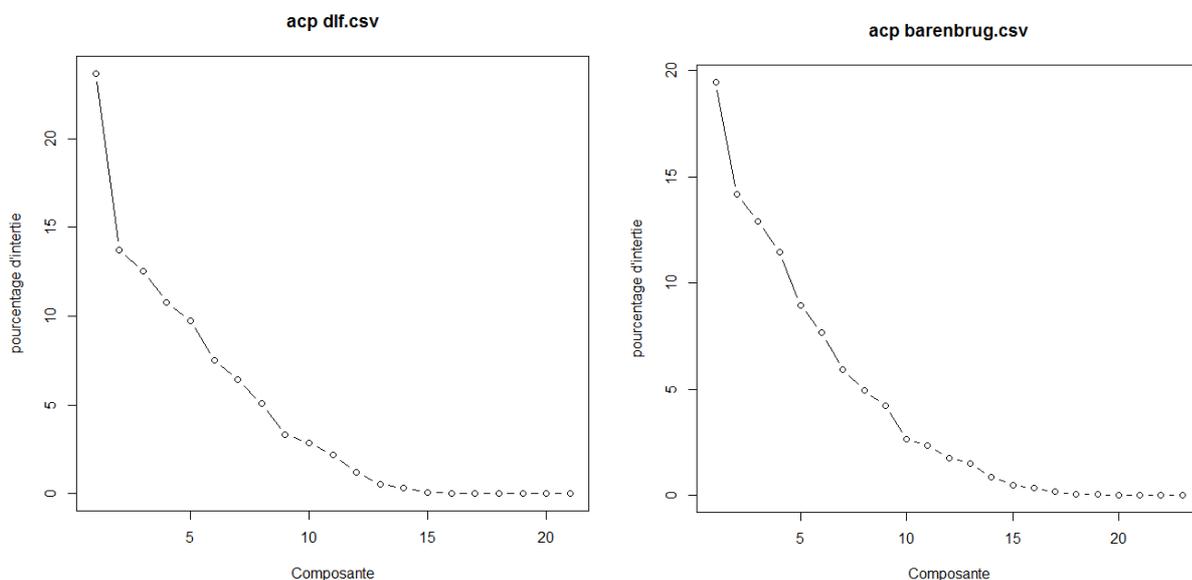
Ainsi, dans la plupart des cas étudiés, pour conserver au moins 80% de l'inertie, il est nécessaire de prendre en compte au moins les 6 ou 7 premiers axes :

	1 axe	2 axes	3 axes	4 axes	5 axes
Dlf	23 %	37,4 %	49,9 %	60,7 %	70,4 %
Carneau	38,4 %	55,1 %	67,4 %	76,4 %	83 %
Barenbrug	19,4 %	33,6 %	46,5 %	58,01 %	67 %
Gie Grass	25,5 %	47,5 %	63,3 %	73,5 %	82,2 %
R2n	32,9 %	49 %	59,9 %	68,5 %	76,4 %
R2n (deltaC13)	31,4 %	53 %	62,9 %	70,2 %	76,5 %

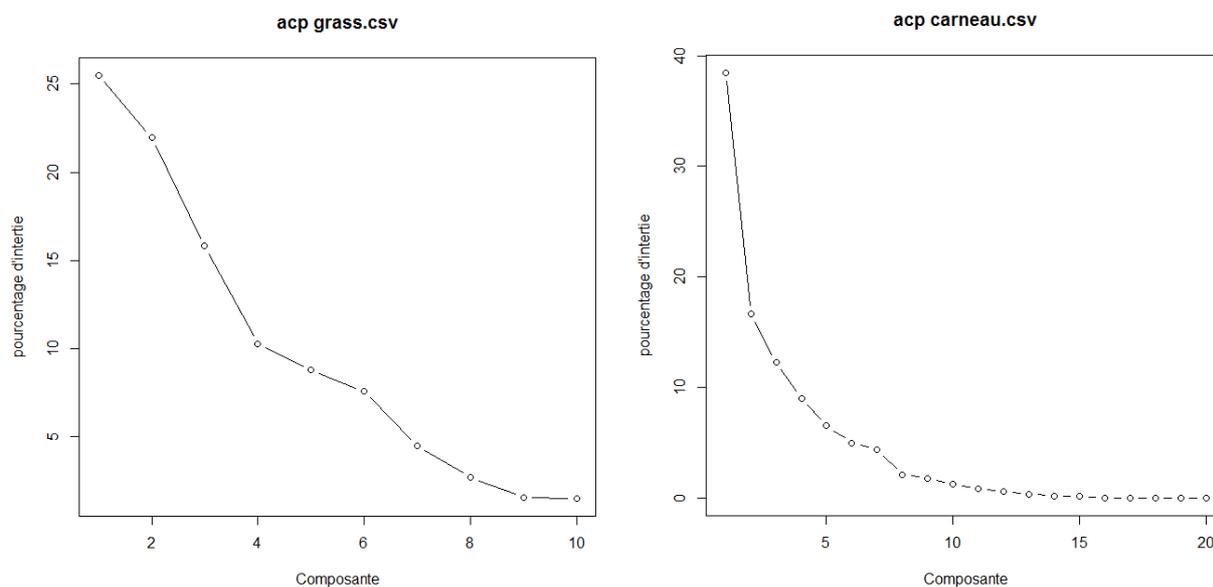
Tableau indiquant le pourcentage d'inertie conservée selon le nombre d'axes

<sup>12</sup> <http://factominer.free.fr>

En analysant le tableau des contributions des variables à chaque axe, il est alors difficile, voire impossible, d'interpréter les axes de façon simple.



Pourcentage d'inertie pour chaque composante – Exemples de « dlf » et « Barenbrug »



Pourcentage d'inertie pour chaque composante – Exemples de « Gie Grass » et « Carneau »

### b. ACP rdtsec/INN et rdtsec/mat sur plusieurs lieux (annexe 9)

Pas plus que dans les cas précédents, les ACP réalisées avec plusieurs lieux pour étudier les liens possibles rdtsec/INN et rdtsec/mat ne permettent pas d'apporter des conclusions claires et de dégager des profils parmi les variétés.

En effet, comme précédemment, pour conserver une inertie suffisante, il est nécessaire de conserver

un grand nombre d'axes et chaque axe ne représente qu'une faible partie de l'inertie.

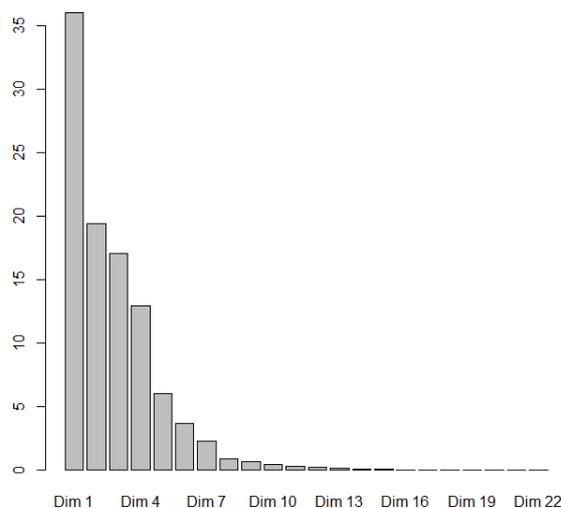
	1 axe	2 axes	3 axes	4 axes	5 axes
ACP rdtsec/mat année 2012	32,3 %	54,4 %	64 %	72 %	78 %
ACP rdtsec/INN année 2012	23,3 %	39,13 %	52,6 %	62,5 %	70 %
ACP rdtsec/mat printemps 2012	27,7 %	49,6 %	60,7 %	69,1 %	76 %
ACP rdtsec/INN printemps 2012	22,7 %	38,4 %	51,1 %	61,6 %	70,5 %

Tableau indiquant le pourcentage d'inertie conservée selon le nombre d'axes

### c. ACP multilocale (annexe 10)

L'ACP multilocale contrairement aux deux autres ACP s'avère plus intéressante car les 3 premiers axes calculés par R permettent de conserver 72 % de l'inertie :

	1 axe	2 axes	3 axes	4 axes
Modèle mixte	36,03 %	55,5 %	72.5 %	85.4 %



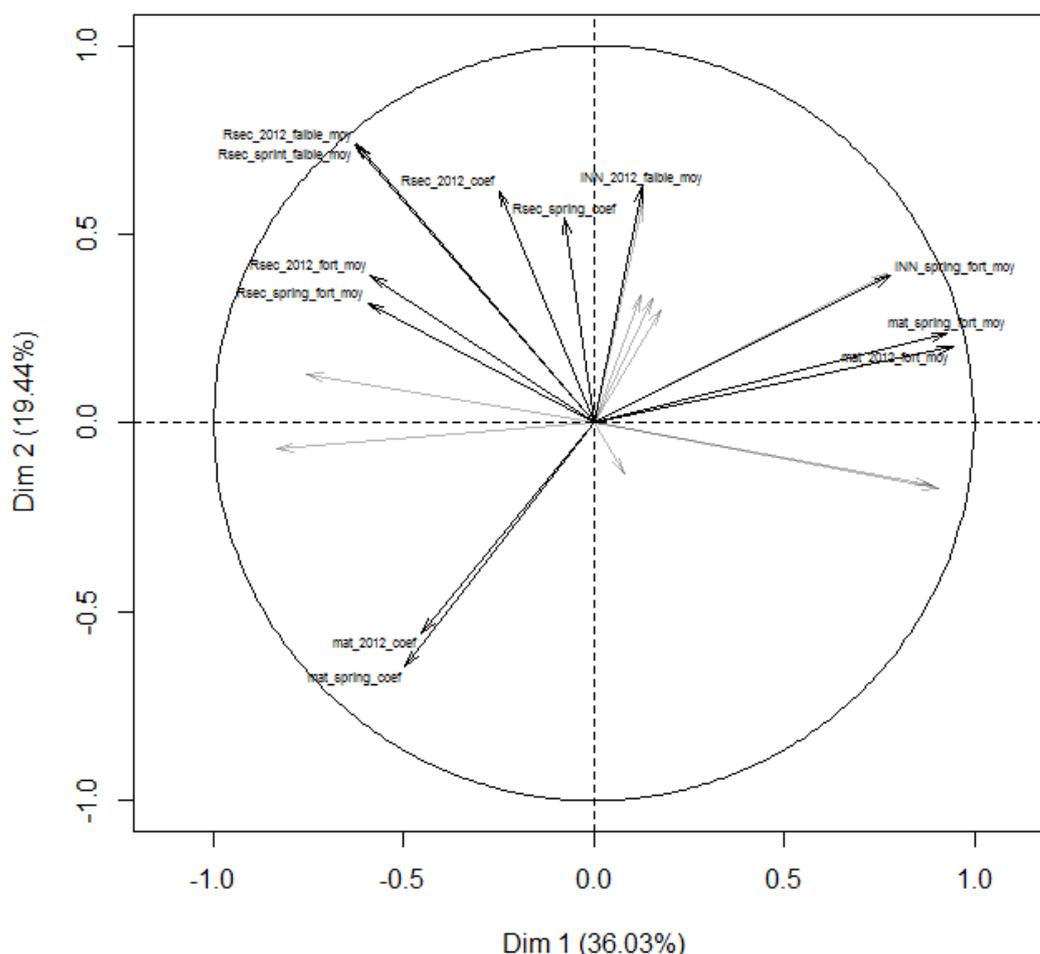
Pourcentage d'inertie pour chaque composante

Bien que l'ACP multilocale soit plus intéressante que les précédentes, son interprétation reste délicate. En effet, les résultats ne concernent actuellement que la seule année 2012 et sont par conséquent encore « instables ». Il sera nécessaire d'avoir d'autres données sur plusieurs années pour tirer des conclusions définitives.

Néanmoins, les résultats obtenus sont encourageants car ils montrent des tendances qui vont tout-à-fait dans le sens attendu.

Une première analyse du cercle des corrélations permet déjà d'entrevoir des résultats prometteurs :

### Variables factor map (PCA)

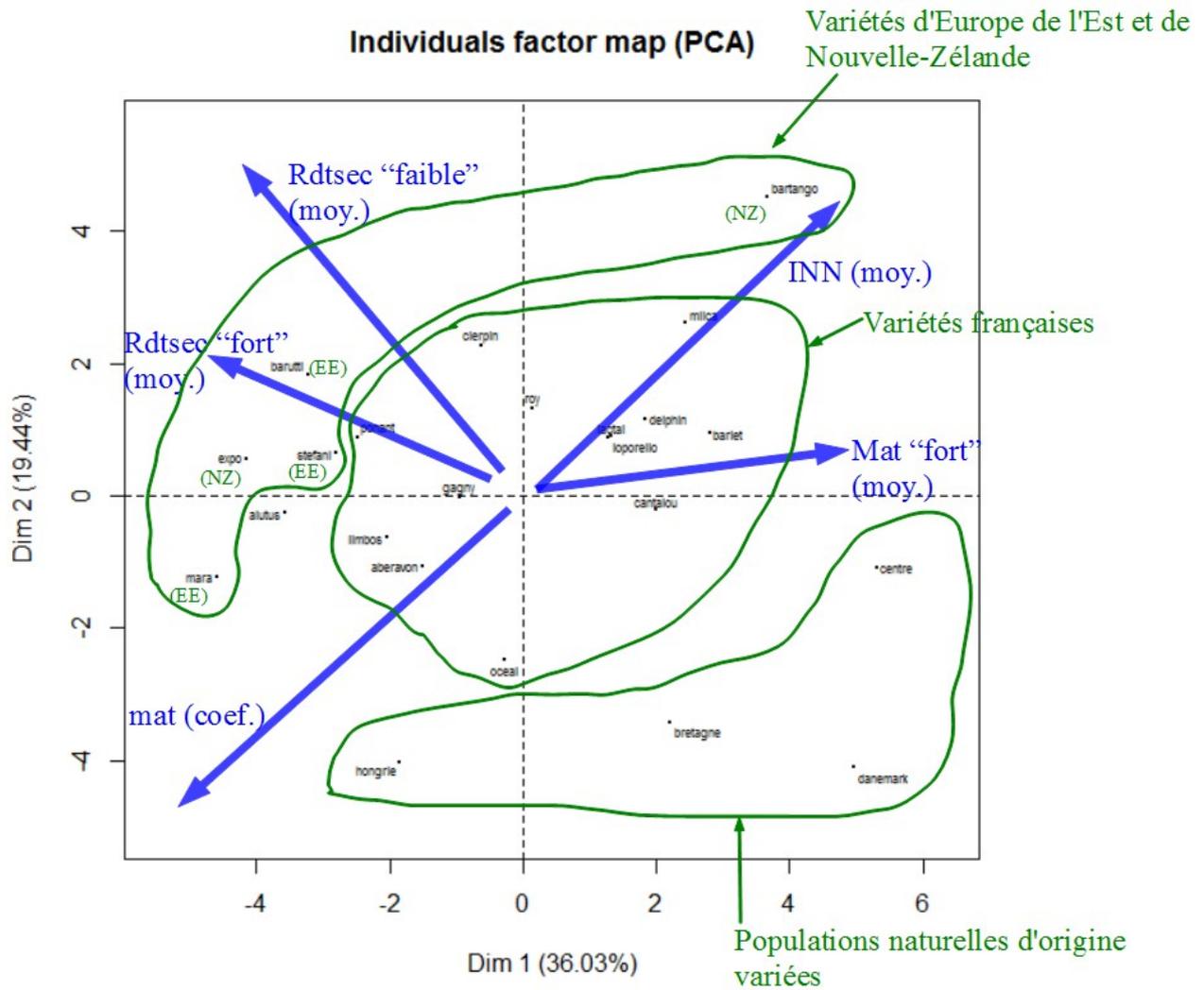


Cercle des corrélations – Axes 1 et 2

Sur la figure ci-dessus, on peut observer une corrélation positive entre les moyennes d'INN et les moyennes de rendement sec. On constate également une corrélation positive entre les moyennes d'INN et les coefficients d'interaction « variété:niveau d'azote faible » pour le rendement. Ces constatations vont dans le sens espéré, c'est-à-dire être en mesure au final de pouvoir « piloter » le rendement d'une variété grâce à l'INN.

Par ailleurs, l'analyse des projections des diverses variétés dans le plan formé par les axes principaux apportent déjà des informations exploitables. Dans le graphique ci-dessous, j'ai superposé en bleu les grandes tendances observées sur le cercle des corrélations.

La répartition des variétés apparaît alors de façon cohérente aux yeux des biologistes qui peut reconnaître des groupes de variétés avec des caractères génétiques plus ou moins marqués :



Projection des variétés dans le plan formé par les deux axes principaux

## 7. Conclusion

L'objectif de ce stage n'était pas d'apporter des réponses définitives aux grandes questions posées par l'étude mais de « préparer le terrain » en vue de l'arrivée de nouvelles données 2013.

Les travaux effectués durant ce stage ont ainsi permis de tester la qualité des données fournies par les différents établissements et de conclure quant à leur exploitation possible. En ce sens, le stage a parfaitement atteint ses objectifs.

Il s'agissait également de mettre en place des méthodes de travail avec le logiciel R, déterminer les différentes bibliothèques de fonctions qui peuvent être utilisées, découvrir les possibilités du logiciel R et ses limites, disposer de scripts qui pourront être exploités dans le futur.

Enfin, les premières analyses réalisées donnent un aperçu de ce qu'on pourra tirer de l'analyse de l'ensemble des données 2012 et 2013 et de ce point de vue, les résultats sont très encourageants et prometteurs.

Sur un plan personnel, ce stage m'a énormément apporté. J'ai pu consolider mes connaissances dans le domaine des statistiques et confronter la théorie à la pratique.

J'ai pu également me rendre compte à quel point les statistiques sont un outil au service des autres disciplines et qu'une approche purement mathématique n'est pas suffisante. Pour réaliser de bonnes analyses, il est nécessaire de connaître parfaitement le champ disciplinaire de l'étude. Ceci est indispensable pour savoir dans quelles directions orienter ses analyses : par exemple choisir des modèles pertinents pour l'analyse de la variance, choisir les bonnes variables pour les ACP, etc...

Etant de nationalité chinoise et ne maîtrisant pas encore suffisamment le français, la barrière linguistique a également été une difficulté supplémentaire à certains moments du stage. Lors de discussions avec Monsieur Sampoux, il est arrivé que certaines explications un peu trop techniques et « pointues » échappent à ma compréhension, ce qui a pu gêner un peu l'avancée de mon travail.

## 8. Bibliographie/webographie

- Le modèle linéaire par l'exemple - Régression, analyse de la variance et plans d'expérience illustrés avec R et SAS - Jean-Marc Azaïs, Jean-Marc Bardet – Chez Dunod
- Statistique théorique et appliquée , T2 Inférence statistique - Pierre Dagnelie - Chez De Boeck
- <http://cran.r-project.org/web/packages/lsmmeans/lsmmeans.pdf> : manuel de référence de la bibliothèque « lsmmeans »
- <http://cran.r-project.org/web/packages/car/car.pdf> : manuel de référence de la bibliothèque « car »
- <http://cran.r-project.org/web/packages/lme4/lme4.pdf> : manuel de référence de la bibliothèque « lme4 »
- <http://cran.r-project.org/web/packages/FactoMineR/FactoMineR.pdf> : manuel de référence de la bibliothèque « FactoMineR »

## 9. Annexes

### 9.1. Annexe 1 : le fichier « initialisation.R »

Ce fichier est appelé par les autres scripts et contient des fonctions d'importation et d'exportation des données.

```
# Definition des chemins de base pour lire et enregistrer
baseread<-"/Users/INRA/data/"
basewrite<-"/Users/INRA/resultats/"

# Chargement de la librairie xlsx pour utiliser write_xlsx
library(xlsx)

#definir deux fonction pour lire et sortir des tableaux
#Attention: on utilise des "<<" pour indiquer que les variables sont globales
#De cette facon, on peut les utiliser en dehors de la fonction

#####
# read #
#####
# Cette fonction permet de lire un fichier de donnees au format CSV
# et d'appliquer la fonction factor pour convertir certaines variables en facteurs
# Entree
# -----
# baseread : le chemin d'accès au dossier contenant les fichiers
# objet : le fichier a ouvrir dans R
# Sortie
# -----
# Variable globale "lieu" : tableau des donnees

read<-function(baseread,objet) {
  chemin<-paste(baseread,objet, ".csv", sep="")
  lieu<<-read.csv2(chemin)
  lieu$variete<<-factor(lieu$variete)
  lieu$niveauN<<-factor(lieu$niveauN)
  lieu$bloc<<-factor(lieu$bloc)
  lieu$typessai<<-factor(lieu$typessai)
  lieu$datejour<<-factor(lieu$datejour)
  lieu$codevar<<-factor(lieu$codevar)
  #print(summary(lieu))
  return(lieu)
}

#####
# write #
#####
# Cette fonction permet d'ecrire un fichier de donnees au format CSV
# Entree
# -----
# tb : le tableau de donnees a enregistrer
# basewrite : le chemin d'accès au dossier ou sont enregistres les résultats
# nom : le nom du fichier de résultats

write<-function(tb,basewrite,nom) {
  cheminwrite<-paste(basewrite,nom, ".csv", sep="")
  write.table(tb,cheminwrite,sep=";",dec=" ",row.names=FALSE,col.names=FALSE)
  #row.names=FALSE,col.names=FALSE ==> Enlever les titres dex lignex et des colonnes que R a cree
}

#####
# write_xlsx #
#####
# Cette fonction permet d'ecrire un fichier de donnees au format XLSX
# Entree
# -----
# tb : le tableau de donnees a enregistrer
# basewrite : le chemin d'accès au dossier ou sont enregistres les résultats
# nom : le nom du fichier de résultats
write_xlsx<-function(tb,basewrite,nom) {
  cheminwrite<-paste(basewrite,nom, ".xlsx", sep="")
  write.xlsx2(tb,cheminwrite,row.names=FALSE,col.names=FALSE)
  #row.names=FALSE,col.names=FALSE ==> Enlever les titres dex lignex et des colonnes que R a cree
}
```



## 9.2. Annexe 2 : analyse de la base de donnée

```

# Definition des chemins de base pour lire et enregistrer
basescripts<-"~/Users/INRA/scripts/"

# execute le fichier initialisation.R
source(paste(basescripts,"initialisation.R",sep=""))

#####
# Analyse des données #
#####
# baseread : le chemin d'accès au dossier contenant les fichiers
# basewrite : le chemin d'accès au dossier ou sont enregistrés les résultats
# nom : le nom du fichier enregistré
# tab : le vecteur des variables à analyser. si l'argument n'est pas spécifié
#       on analyse toutes les variables
#
# Sortie
# -----
# un tableau de données

analyse_bd<-function(baseread,basewrite,nom,tab=NULL) {
  # Première ligne du tableau de résultats
  T<-c("organisme","codevar","Nb. data non vides","Nb. data vides","types essai","Nb. variétés","Nb.
dates","Dates","Nb. niveauN","Nb. bloc")
  for (k in c("gie grass","dlf","barenbrug","carneau","r2n")){
    read(baseread,k)
    # La fonction read() crée une variable globale "lieu" : tableau des données
    for (i in (levels(lieu$codevar))){
      if ((length(which(i==tab))!=0)|(length(tab)==0)) {
        print(".....")
        print(paste("Code variable : ",i))

        # On extrait le tableau des variétés (sans les valeurs vides)
        trait2<-subset(lieu,subset=((valeur!="") & (codevar==i)))

        # On extrait le tableau des variétés
        trait<-subset(lieu,subset=(codevar==i))

        print(paste("Nombre de données non nulles : ",dim(trait2)[1]))
        print(paste("Il y a ",dim(trait)[1]-dim(trait2)[1],"données nulles"))

        # On affiche les différents types d'essais présents
        trait2$typessai<-factor(trait2$typessai) # On met à jour les niveaux
        type=""
        for (j in (levels(trait2$typessai))){
          if (type=="") type<-j else type<-paste(type,j)
        }
        print(type)
        # On affiche les différentes dates présentes
        trait2$datejour<-factor(trait2$datejour) # On met à jour les niveaux
        date=""
        nbdate<-0
        for (j in (levels(trait2$datejour))){
          if (date=="") date<-j else date<-paste(date,j,sep=" - ")
          nbdate<-nbdate+1
        }
        print(paste("Nombre de dates présentes : ",nbdate))
        print(date)

        # On détermine le nombre de variétés présentes
        trait2$codevar<-factor(trait2$codevar) # On met à jour les niveaux
        nbvar<-0
        for (j in (levels(trait2$variété ))){
          nbvar<-nbvar+1
        }
        print(paste("Nombre de variétés présentes : ",nbvar))

        # On détermine le nombre de niveau d'azote présents
        trait2$niveauN<-factor(trait2$niveauN) # On met à jour les niveaux
        nbN<-0

```

```
for (j in (levels(trait2$niveauN))){
  nbN<-nbN+1
}
print(paste("Nombre de niveau d'azote presents : ",nbN))

# On determine le nombre de blocs presents
trait2$bloc<-factor(trait2$bloc) # On met a jour les niveaux
nbB<-0
for (j in (levels(trait2$bloc))){
  nbB<-nbB+1
}
print(paste("Nombre de blocs presents : ",nbB))

# On ajoute une ligne au tableau de resultats

T<-rbind(T,c(k,i,dim(trait2)[1],dim(trait)[1]-dim(trait2)[1],type,nbvar,nbdate,date,nbN,nbB))
}
}
write(T,basewrite,nom) # On enregistre les résultats dans un fichier
}

analyse_bd(baseread,basewrite,"analyse_bd_all")
analyse_bd(baseread,basewrite,"analyse_bd",c("INN","mat","deltaC13","rdtsec","teneurC","rapportC/N")
)
```

### 9.3. Annexe 3 : les fonctions « analyse1() » et « analyse2() » et traitement

```

# On supprime tous les objets existants dans R
rm(list=ls())

# Chemin d'accès aux scripts R
basescripts<-" /Users/INRA/scripts/"

# execute le fichier initialisation.R
source(paste(basescripts,"initialisation.R",sep=""))

# Chargement de la librairie lme4 pour la fonction lmer
library(lme4)

# Chargement de la librairie car pour la fonction Anova
library(car)

# definir une fonction pour faire des analyse de variance pour chaque variable par lieu par date
# Ici, on n'analyse pas les variable "cdr", "psech" , "pvech" et "pvparcelle"
# On utilise toujours le contrasts="contr.sum" pour chaque facteur dans toute cette section
options(contrasts=c("contr.sum","contr.sum"))

#####
# analyse1 #
#####
# Analyse de la variance par lieu, date et type de variable
# Entree
# -----
# baseread : le chemin d'accès au dossier contenant les fichiers
# basewrite : le chemin d'accès au dossier ou sont enregistrés les résultats
# objet : le nom de l'organisme a analyser
# Sortie
# -----
# un tableau de données

analyse1<-function(baseread,basewrite,objet) {
  read(baseread,objet)
  # Boucle pour balayer les typessai, les codevar et les dates
  # Premiere ligne du tableau de resultats
  T<-
  c("organisme","typessai","codevar","datejour","Fv","Fn","Fb","Fv*n","val.min","val.mean","val.max","
  residu.Mean sq","coeffi de variete","Fvm","Fnm")
  # On separe les deux typessai "ligne" et "parcelle" car on applique des modeles differents
  for (i in (levels(lieu$typessai))){
    for (j in (levels(lieu$codevar))){
      for (k in (levels(lieu$datejour))){
        # Traiter toutes les combinaisons de variables et dates
        trait<-subset(lieu,subset=((valeur!="") & (typessai==i) & (codevar==j) & (datejour==k) &
        (codevar=="cdr") & !(codevar=="psech") & !(codevar=="pvech") & !(codevar=="pvparcelle")))
        # On teste si le traitement est vide
        if (dim(trait)[1]>0){
          if (i=="ligne"|objet=="gie grass"){
            a<-Anova(lm(valeur~variete+bloc,data=trait),type=3)
            # On ajoute une ligne au tableau de resultats
            T<-
            rbind(T,c(objet,i,j,k,a[2,4],NA,a[3,4],NA,min(trait[,8]),mean(trait[,8]),max(trait[,8]),a[4,1]/a[4,2
            ],sqrt(a[4,1]/a[4,2])/mean(trait[,8]),NA,NA))
          } else {
            # On teste s'il y a deux niveaux d'azote niveauN
            # car la fonction "lmer" ne peut s'appliquer que s'il y a au moins deux niveaux,
            # On traite d'abord le cas ou il y a deux niveaux
            if ((dim(subset(trait,subset=(niveauN=="fort")))[1]>0) &
            (dim(subset(trait,subset=(niveauN=="faible")))[1]>0)) {
              # modele pour l'effet fixe
              a<-Anova(lm(valeur~variete*niveauN+bloc,data=trait),type=3,singular.ok=TRUE)
              # modele pour l'effet mixte
              a2<-Anova(lmer(valeur ~ variete+niveauN+bloc+(1|
              variete:niveauN),trait),type=3,test.statistic="F")
              # On ajoute une ligne au tableau de resultats

```



```
print(T2) # On affiche les resultats a l'ecran
write(T2,basewrite,paste("analyse2_",objet,sep="")) # On enregistre les résultats dans un fichier
}

#analyse de variance par lieu et dates
for (i in c("gie grass","dlf","barenbrug","carneau","r2n")){
  analyse1(baseread,basewrite,i)
}

#analyse de variance par lieu, dates et niveau d'azote
for (i in c("gie grass","dlf","barenbrug","carneau","r2n")){
  analyse2(baseread,basewrite,i)
}
```

## 9.4. Annexe 4 : test des résidus de la première série d'analyse de variance

```

# On supprime tous les objets existants dans R
rm(list=ls())

# Chemin d'accès aux scripts R
basescripts<-"Users/INRA/scripts/"

# exécute le fichier initialisation.R
source(paste(basescripts,"initialisation.R",sep=""))

# Chargement de la librairie lme4 pour la fonction lmer
library(lme4)

# Chargement de la librairie car pour la fonction Anova
library(car)

#####
# tests #
#####
# Test de normalité des résidus
# Test d'homoscédasticité (homogénéité des variances)
# Affichage des QQ-plot

tests<-function(baseread) {
  par(mfrow=c(3,3),mai=c(0.3,0.3,0,0)) # On affiche 9 graphiques par page
  for (objet in c("barenbrug","carneau","r2n","dlf","gie grass")) {
    compteur<-1 # Compteur pour tester l'ouverture d'une nouvelle page
    read(baseread,objet)
    # Boucle pour balayer les typessai, les codevar et les dates
    # Première ligne du tableau de résultats
    # On sépare les deux typessai "ligne" et "parcelle" car on applique des modèles différents
    for (i in (levels(lieu$typessai))){
      for (j in (levels(lieu$codevar))){
        if ((j=="rdtsec")|(j=="INN")|(j=="deltaC13")|(j=="mat")) {
          for (k in (levels(lieu$datejour))){
            # Traiter toutes les combinaisons de variables et dates
            trait<-subset(lieu,subset=(valeur!="") & (typessai==i) & (codevar==j) & (datejour==k) &
              (codevar=="cdr") & !(codevar=="psech") & !(codevar=="pvech") & !(codevar=="pvparcelle"))
            # On teste si le traitement est vide
            if (dim(trait)[1]>0){
              if (i=="ligne"|objet=="gie grass"){
                modele<-as.formula(valeur~variete+bloc)
                model.lm<-lm(modele,data=trait)
              } else {
                # On teste s'il y a deux niveaux d'azote niveauN
                # car la fonction "lmer" ne peut s'appliquer que s'il y a au moins deux niveaux,
                # On traite d'abord le cas où il y a deux niveaux
                if ((dim(subset(trait,subset=(niveauN=="fort")))[1]>0) &
                  (dim(subset(trait,subset=(niveauN=="faible")))[1]>0)) {
                  # modèle pour l'effet fixe
                  modele<-as.formula(valeur~variete*niveauN+bloc)
                  model.lm=lm(modele,data=trait)
                } else{
                  modele<-as.formula(valeur~variete+bloc)
                  model.lm=lm(modele,data=trait)
                }
              }
            }
            st<-try(shapiro.test(residuals(model.lm)),silent=TRUE) # Test de Shapiro (test de
normalité)
            if (class(st)!="try-error") st.value<-round(st$p.value,3) else st.value=0 # Calcul de
la p-valeur du test de Shapiro

            # bt<-try(bartlett.test(modele,data=trait),silent=TRUE) # Test de Bartlett sur le modèle
(homoscédasticité)
            # if (class(bt)!="try-error") bt.value<-round(bt$p.value,3) else bt.value=0 # Calcul de
la p-valeur du test de Bartlett
            # Affichage du QQ-Plot
            # if (st.value<0.05|bt.value<0.05) { # On ne teste pas l'homoscédasticité ici car cela
n'a pas de sens
            if (st.value<0.05) {

```



## 9.5. Annexe 5 : la fonction « analyse3() » et traitement

### Script R permettant l'analyse de la variance par lieux et par variable

```

# On supprime tous les objets existants dans R
rm(list=ls())

# Chemin d'accès aux scripts R
basescripts<-"/Users/INRA/scripts/"

# execute le fichier initialisation.R
source(paste(basescripts,"initialisation.R",sep=""))

# Chargement de la librairie lme4 pour la fonction lmer et la fonction fixef
library(lme4)

# Chargement de la librairie car pour la fonction Anova
library(car)

#####
# analyse de variance #
#####

# On utilise toujours les contrasts="contr.sum" pour chaque facteur dans toute cette section
options(contrasts=c("contr.sum","contr.sum"))

analyse3<-function(gie,d1,d12,bar,car,r2,basewrite,nom_variable) {
  # Construire le modele lineaire fixe en utilisant la fonction "lm"
  gie.lm<-lm(valeur~variete*datejour+bloc,data=gie)
  d1.lm<-lm(valeur~variete*niveauN*datejour+bloc,data=d1)
  d12.lm<-lm(valeur~variete*niveauN*datejour+bloc,data=d12)
  bar.lm<-lm(valeur~variete*niveauN*datejour+bloc,data=bar)
  car.lm<-lm(valeur~variete*niveauN*datejour+bloc,data=car)
  r2.lm<-lm(valeur~variete*niveauN*datejour+bloc,data=r2)

  # Construire le modele lineaire mixte en utilisant la fonction "lmer"
  gie.lmer<-lmer(valeur~variete+bloc+(1|datejour)+(1|variete:datejour),gie)
  d1.lmer<-lmer(valeur ~ variete+niveauN+bloc+variete:niveauN+(1|datejour)+(1|
variete:niveauN:datejour),d1)
  d12.lmer<-lmer(valeur ~ variete+niveauN+bloc+variete:niveauN+(1|datejour)+(1|
variete:niveauN:datejour),d12)
  bar.lmer<-lmer(valeur ~ variete+niveauN+bloc+variete:niveauN+(1|datejour)+(1|
variete:niveauN:datejour),bar)
  car.lmer<-lmer(valeur ~ variete+niveauN+bloc+variete:niveauN+(1|datejour)+(1|
variete:niveauN:datejour),car)
  r2.lmer<-lmer(valeur ~ variete+niveauN+bloc+variete:niveauN+(1|datejour)+(1|
variete:niveauN:datejour),r2)

  # Analyse de la variance avec "Anova(type=3)" pour le modele fixe
  gfixe<-Anova(gie.lm,type=3)
  d1fixe<-Anova(d1.lm,type=3,singular.ok=TRUE)
  d2fixe<-Anova(d12.lm,type=3)
  bfixe<-Anova(bar.lm,type=3)
  cfixe<-Anova(car.lm,type=3)
  rfixe<-Anova(r2.lm,type=3)

  # Analyse de la variance avec "Anova(type=3)" pour le modele mixte,
  # Il faut bien preciser quel test on utilise. Ici on utilise le test de Fisher
  gmixte<-Anova(gie.lmer,type=3,test.statistic="F")
  dmixte<-Anova(d1.lmer,type=3,test.statistic="F")
  d2mixte<-Anova(d12.lmer,type=3,test.statistic="F")
  bmixte<-Anova(bar.lmer,type=3,test.statistic="F")
  cmixte<-Anova(car.lmer,type=3,test.statistic="F")
  rmixte<-Anova(r2.lmer,type=3,test.statistic="F")

  # On construit un tableau avec les colonnes suivantes

  # organisme
  # P-value Fisher Variete
  # P-value Fisher Azote
  # P-value Fisher date
  # P-value Fisher Bloc

```

```

# P-value Fisher interaction variete*azote
# P-value Fisher interaction variete*date
# P-value Fisher interaction azote*date
# P-value Fisher interaction variete*azote*date
# P-value Fisher variete modele mixte
# P-value Fisher azote modele mixte
# P-value Fisher bloc modele mixte
# P-value Fisher int. variete*azote mod. mixte

# Le tableau
g<-c("gie
grass",gfixe[2,4],NA,gfixe[3,4],gfixe[4,4],NA,gfixe[5,4],NA,NA,gmixte[2,4],NA,gmixte[3,4],NA)
d<-
c("dlf",dfixe[2,4],dfixe[3,4],dfixe[4,4],dfixe[5,4],dfixe[6,4],dfixe[7,4],dfixe[8,4],dfixe[9,4],dmix
te[2,4],dmixte[3,4],dmixte[4,4],dmixte[5,4])
d2<-c("dlf sans
0609",d2fixe[2,4],d2fixe[3,4],d2fixe[4,4],d2fixe[5,4],d2fixe[6,4],d2fixe[7,4],d2fixe[8,4],d2fixe[9,4
],d2mixte[2,4],d2mixte[3,4],d2mixte[4,4],d2mixte[5,4])
b<-
c("barenbrug",bfixe[2,4],bfixe[3,4],bfixe[4,4],bfixe[5,4],bfixe[6,4],bfixe[7,4],bfixe[8,4],bfixe[9,4
],bmixte[2,4],bmixte[3,4],bmixte[4,4],bmixte[5,4])
c<-
c("carneau",cfixe[2,4],cfixe[3,4],cfixe[4,4],cfixe[5,4],cfixe[6,4],cfixe[7,4],cfixe[8,4],cfixe[9,4],
cmixte[2,4],cmixte[3,4],cmixte[4,4],cmixte[5,4])
r<-
c("r2n",rfixe[2,4],rfixe[3,4],rfixe[4,4],rfixe[5,4],rfixe[6,4],rfixe[7,4],rfixe[8,4],rfixe[9,4],rmix
te[2,4],rmixte[3,4],rmixte[4,4],rmixte[5,4])
G<-c("organisme","P-val Variete (date fixe)","P-val NiveauN (date fixe)","P-val Datejour (date
fixe)","P-val bloc (date fixe)","P-val V*N (date fixe)","P-val V*D (date fixe)","P-val N*D (date
fixe)","P-val V*N*D (date fixe)","P-val Variete (date alea.)","P-val N (date alea.)","P-val bloc
(date alea.)","P-val V*N (date alea.)")
G<-rbind(G,g,d,d2,b,c,r)

# sortir le tableau
print(G)
# write_xlsx(G,basewrite,paste("analyse3_",nom_variable,sep=""))
write(G,basewrite,paste("analyse3_",nom_variable,sep="")) # On enregistre les résultats dans un
fichier
}

#####
#importer et traiter les donnees #
#####

#importer les donnees
gie.grass<-read(baseread,"gie grass")
dlf<-read(baseread,"dlf")
barenbrug<-read(baseread,"barenbrug")
carneau<-read(baseread,"carneau")
r2n<-read(baseread,"r2n")

#####
# Traitement des donnees pour la variable "rdtsec" - ANNEE 2012 #
#####
# On ne retient que les dates en 2012
gie<-subset(gie.grass,subset=((valeur!="") & (codevar=="rdtsec") & ((datejour=="03/04/2012") |
(datejour=="19/05/2012") | (datejour=="14/06/2012") | (datejour=="18/07/2012") |
(datejour=="26/10/2012"))))
# Pour "dlf", beaucoup de donnees manquante pour le 06/09/2012. Donc on teste avec et sans cette
date
dl<-subset(dlf,subset=((valeur!="") & (codevar=="rdtsec") & ((datejour=="16/05/2012") |
(datejour=="19/06/2012") | (datejour=="18/07/2012") | (datejour=="06/09/2012") |
(datejour=="30/11/2012"))))
dl2<-subset(dlf,subset=((valeur!="") & (codevar=="rdtsec") & ((datejour=="16/05/2012") |
(datejour=="19/06/2012") | (datejour=="18/07/2012") | (datejour=="30/11/2012"))))
bar<-subset(barenbrug,subset=((valeur!="") & (datejour!
="") & (codevar=="rdtsec") & ((datejour=="10/05/2012") | (datejour=="01/06/2012") |
(datejour=="28/06/2012") | (datejour=="26/07/2012") | (datejour=="24/10/2012"))))
car<-subset(carneau,subset=((valeur!="") & (datejour!
="") & (codevar=="rdtsec") & ((datejour=="30/04/2012") | (datejour=="25/05/2012") |
(datejour=="25/06/2012") | (datejour=="22/08/2012"))))

```

```
r2<-subset(r2n, subset=( (valeur!="") & (datejour!="") & (codevar=="rdtsec") & ((datejour=="12/04/2012") |
(datejour=="10/05/2012") | (datejour=="05/06/2012") | (datejour=="25/07/2012") | (datejour=="30/07/2012") |
(datejour=="28/08/2012") | (datejour=="28/09/2012"))))
```

```
analyse3(gie, dl, dl2, bar, car, r2, basewrite, "rdtsec_2012")
```

```
#####
# Traitement des donnees pour la variable "mat" - ANNEE 2012 #
#####
# On ne retient que les dates en 2012
gie<-subset(gie.grass, subset=( (valeur!="") & (codevar=="mat") & ((datejour=="03/04/2012") |
(datejour=="19/05/2012") | (datejour=="14/06/2012") | (datejour=="18/07/2012") |
(datejour=="26/10/2012"))))
dl<-subset(dlf, subset=( (valeur!="") & (codevar=="mat") & ((datejour=="16/05/2012") |
(datejour=="19/06/2012") | (datejour=="18/07/2012") | (datejour=="06/09/2012") |
(datejour=="30/11/2012"))))
dl2<-subset(dlf, subset=( (valeur!="") & (codevar=="mat") & ((datejour=="16/05/2012") |
(datejour=="19/06/2012") | (datejour=="18/07/2012") | (datejour=="30/11/2012"))))
bar<-subset(barenbrug, subset=( (valeur!="") & (datejour!
="") & (codevar=="mat") & ((datejour=="10/05/2012") | (datejour=="01/06/2012") | (datejour=="28/06/2012") |
(datejour=="26/07/2012") | (datejour=="24/10/2012"))))
car<-subset(carneau, subset=( (valeur!="") & (datejour!="") & (codevar=="mat") & ((datejour=="30/04/2012") |
(datejour=="25/05/2012") | (datejour=="25/06/2012") | (datejour=="22/08/2012"))))
r2<-subset(r2n, subset=( (valeur!="") & (datejour!="") & (codevar=="mat") & ((datejour=="12/04/2012") |
(datejour=="10/05/2012") | (datejour=="05/06/2012") | (datejour=="25/07/2012") | (datejour=="30/07/2012") |
(datejour=="28/08/2012") | (datejour=="28/09/2012"))))
```

```
analyse3(gie, dl, dl2, bar, car, r2, basewrite, "mat_2012")
```

```
#####
# Traitement des donnees pour la variable "INN" - ANNEE 2012 #
#####
gie<-subset(gie.grass, subset=( (valeur!="") & (codevar=="INN") & ((datejour=="03/04/2012") |
(datejour=="19/05/2012") | (datejour=="14/06/2012") | (datejour=="18/07/2012") |
(datejour=="26/10/2012"))))
dl<-subset(dlf, subset=( (valeur!="") & (codevar=="INN") & ((datejour=="16/05/2012") |
(datejour=="19/06/2012") | (datejour=="18/07/2012") | (datejour=="06/09/2012") |
(datejour=="30/11/2012"))))
dl2<-subset(dlf, subset=( (valeur!="") & (codevar=="INN") & ((datejour=="16/05/2012") |
(datejour=="19/06/2012") | (datejour=="18/07/2012") | (datejour=="30/11/2012"))))
bar<-subset(barenbrug, subset=( (valeur!="") & (datejour!
="") & (codevar=="INN") & ((datejour=="10/05/2012") | (datejour=="01/06/2012") | (datejour=="28/06/2012") |
(datejour=="26/07/2012") | (datejour=="24/10/2012"))))
car<-subset(carneau, subset=( (valeur!="") & (datejour!="") & (codevar=="INN") & ((datejour=="30/04/2012") |
(datejour=="25/05/2012") | (datejour=="25/06/2012") | (datejour=="22/08/2012"))))
r2<-subset(r2n, subset=( (valeur!="") & (datejour!="") & (codevar=="INN") & ((datejour=="12/04/2012") |
(datejour=="10/05/2012") | (datejour=="05/06/2012") | (datejour=="25/07/2012") | (datejour=="30/07/2012") |
(datejour=="28/08/2012") | (datejour=="28/09/2012"))))
```

```
analyse3(gie, dl, dl2, bar, car, r2, basewrite, "INN_2012")
```

```
#####
# Traitement des donnees pour la variable "mat" - ETE 2012 #
#####
gie<-subset(gie.grass, subset=( (valeur!="") & (codevar=="mat") & ((datejour=="03/04/2012") |
(datejour=="19/05/2012") | (datejour=="14/06/2012"))))
dl<-subset(dlf, subset=( (valeur!="") & (codevar=="mat") & ((datejour=="16/05/2012") |
(datejour=="19/06/2012") | (datejour=="18/07/2012"))))
bar<-subset(barenbrug, subset=( (valeur!="") & (datejour!
="") & (codevar=="mat") & ((datejour=="10/05/2012") | (datejour=="01/06/2012") | (datejour=="28/06/2012"))))
car<-subset(carneau, subset=( (valeur!="") & (datejour!="") & (codevar=="mat") & ((datejour=="30/04/2012") |
(datejour=="25/05/2012") | (datejour=="25/06/2012"))))
r2<-subset(r2n, subset=( (valeur!="") & (datejour!="") & (codevar=="mat") & ((datejour=="12/04/2012") |
(datejour=="10/05/2012") | (datejour=="05/06/2012"))))
```

```
analyse3(gie, dl, dl2, bar, car, r2, basewrite, "mat_ETE_2012")
```

```
#####
# Traitement des donnees pour la variable "INN" - ETE 2012 #
#####
gie<-subset(gie.grass, subset=( (valeur!="") & (codevar=="INN") & ((datejour=="03/04/2012") |
(datejour=="19/05/2012") | (datejour=="14/06/2012"))))
dl<-subset(dlf, subset=( (valeur!="") & (codevar=="INN") & ((datejour=="16/05/2012") |
```

```
(datejour=="19/06/2012") | (datejour=="18/07/2012"))))
bar<-subset(barenbrug, subset=( (valeur!="") & (datejour!
=="") & (codevar=="INN") & ((datejour=="10/05/2012") | (datejour=="01/06/2012") | (datejour=="28/06/2012"))))
car<-subset(carneau, subset=( (valeur!="") & (datejour!="") & (codevar=="INN") & ((datejour=="30/04/2012") |
(datejour=="25/05/2012") | (datejour=="25/06/2012"))))
r2<-subset(r2n, subset=( (valeur!="") & (datejour!="") & (codevar=="INN") & ((datejour=="12/04/2012") |
(datejour=="10/05/2012") | (datejour=="05/06/2012"))))

analyse3(gie, dl, dl2, bar, car, r2, basewrite, "INN_ETE_2012")

#####
# Traitement des donnees pour la variable "rdtsec" - ETE 2012 #
#####
gie<-subset(gie.grass, subset=( (valeur!="") & (codevar=="rdtsec") & ((datejour=="03/04/2012") |
(datejour=="19/05/2012") | (datejour=="14/06/2012") | (datejour=="18/07/2012") |
(datejour=="26/10/2012"))))
dl<-subset(dlf, subset=( (valeur!="") & (codevar=="rdtsec") & ((datejour=="16/05/2012") |
(datejour=="19/06/2012") | (datejour=="18/07/2012") | (datejour=="06/09/2012") |
(datejour=="30/11/2012"))))
dl2<-subset(dlf, subset=( (valeur!="") & (codevar=="rdtsec") & ((datejour=="16/05/2012") |
(datejour=="19/06/2012") | (datejour=="18/07/2012") | (datejour=="30/11/2012"))))
bar<-subset(barenbrug, subset=( (valeur!="") & (datejour!
=="") & (codevar=="rdtsec") & ((datejour=="10/05/2012") | (datejour=="01/06/2012") |
(datejour=="28/06/2012") | (datejour=="26/07/2012") | (datejour=="24/10/2012"))))
car<-subset(carneau, subset=( (valeur!="") & (datejour!
=="") & (codevar=="rdtsec") & ((datejour=="30/04/2012") | (datejour=="25/05/2012") |
(datejour=="25/06/2012") | (datejour=="22/08/2012"))))
r2<-subset(r2n, subset=( (valeur!="") & (datejour!="") & (codevar=="rdtsec") & ((datejour=="12/04/2012") |
(datejour=="10/05/2012") | (datejour=="05/06/2012") | (datejour=="25/07/2012") | (datejour=="30/07/2012") |
(datejour=="28/08/2012") | (datejour=="28/09/2012"))))

analyse3(gie, dl, dl2, bar, car, r2, basewrite, "rdtsec_ETE_2012")
```

## 9.6. Annexe 6 : analyse de la variance dans le cas multilocal (rdtsec, mat et INN )

```

# On supprime tous les objets existants dans R
rm(list=ls())

# Chemin d'accès aux scripts R
basescripts<-"~/Users/INRA/scripts/"

# execute le fichier initialisation.R
source(paste(basescripts,"initialisation.R",sep=""))

# Chargement de la librairie lme4 pour la fonction lmer et la fonction fixef
library(lme4)

# Chargement de la librairie car pour la fonction Anova
library(car)

# Chargement de la librairie lsmeans pour la fonction lsmeans
library(lsmeans)

# Chargement de la librairie agricolae pour le test "SNK.test"
library(agricolae)

#####
# analyse de variance #
#####

# On utilise toujours les contrasts="contr.sum" pour chaque facteur dans toute cette section
options(contrasts=c("contr.sum","contr.sum"))

#####
# Fonction qui cree les ajustements avec un modele specifie
# Les variables creees anmixtel.lmer, anmixte2.lmer, anmete1.lmer, anmete2.lmer sont globales
# De meme anfixe.lm, anfixe_ete.lm, anfixe.lm, anfixe_ete.lm sont globales
# pour pouvoir etre utilisees dans les fonctions qui suivent
#####

ajustement_multilocal<-function() {
  # Construire le modele lineaire fixe en utilisant la fonction "lm"
  anfixe.lm<<-
lm(valeur~variete*niveauN*organisme+organisme/datejour+organisme/bloc+variete:organisme:datejour+niv
eauN:organisme:datejour+variete:niveauN:organisme:datejour,data=D_annee2012)
  anfixe_ete.lm<<-
lm(valeur~variete*niveauN*organisme+organisme/datejour+organisme/bloc+variete:organisme:datejour+niv
eauN:organisme:datejour+variete:niveauN:organisme:datejour,data=D_ete2012)

  # construire des modeles lineaires mixtes
  anmixte2.lmer<<-lmer(valeur~variete*niveauN+(1|organisme)+(1|organisme/bloc)+(1|
organisme/datejour)+(1|datejour:variete:niveauN:organisme),D_mixte_annee2012)
  anmete2.lmer<<-lmer(valeur~variete*niveauN+(1|organisme)+(1|organisme/bloc)+(1|organisme/datejour)
+(1|datejour:variete:niveauN:organisme),D_mixte_ete2012)
}

#####
# Fonction qui fait des anova sur l'ensemble des lieux
#
#####
analyse_multilocale<-function(basewrite,nom_variable) {

  # Analyse de la variance avec "Anova(type=3)"
  anfixe<-Anova(anfixe.lm,type=3,singular.ok=TRUE)
  anfixe_ete<-Anova(anfixe_ete.lm,type=3,singular.ok=TRUE)

  # analyse de la variance avec "Anova(type=3)" pour le modele mixte
  # il faut bien preciser quel test utiliser, ici on utilise le test de Fisher
  anmixte2<-Anova(anmixte2.lmer,type=3,singular.ok=TRUE,test.statistic="F")
  anmete2<-Anova(anmete2.lmer,type=3,singular.ok=TRUE,test.statistic="F")
}

```

```

# On construit un tableau des resultats
# Le tableau
# premiere lignedu tableau
T<-c("", "P.val Var.", "P.val N", "P.val lieu", "P.val V*N", "P.val V*L", "P.val N*L", "P.val L*D", "P.val
L*B", "P.val V*N*L", "P.val V*L*D", "P.val N*L*D", "P.val V*N*L*D",
      "P.val M2-v", "P.val M2-n", "P.val M2-v*n")

# concatener toutes les parties
C<-c("toute l'annee
2012", anfixe[2,4], anfixe[3,4], anfixe[4,4], anfixe[5,4], anfixe[6,4], anfixe[7,4], anfixe[8,4], anfixe[9,4
], anfixe[10,4], anfixe[11,4], anfixe[12,4], anfixe[13,4],
      anmixte2[2,4], anmixte2[3,4], anmixte2[4,4])
Cete<-c("Ete
2012", anfixe_ete[2,4], anfixe_ete[3,4], anfixe_ete[4,4], anfixe_ete[5,4], anfixe_ete[6,4], anfixe_ete[7,4
], anfixe_ete[8,4], anfixe_ete[9,4], anfixe_ete[10,4], anfixe_ete[11,4], anfixe_ete[12,4], anfixe_ete[13,4
],
      anm2ete[2,4], anm2ete[3,4], anm2ete[4,4])
T<-rbind(T,C,Cete)

# sortir le tableau
# write_xlsx(G,basewrite,paste("analyse_niche_",nom_variable,sep=""))
write(T,basewrite,paste("analyse_multilocale_niche_",nom_variable,sep=""))
}

#####
# Fonction permettant la fabrication d'un tableau complet #
# De moyennes ajustees par variete,variete:niveau pour modele mixte #
# a partir des ajustements anmixte2.lmer, anmete2.lmer #
# (qui sont des variables globales) #
#####
lsmeans_Variete<-function(basewrite,nom_variable) {
  # calcul des lsmeans par variete
  lsv<-lsmeans(anmixte2.lmer,~variete)
  lsvete<-lsmeans(anmete2.lmer,~variete)

  # calcul des lsmeans par variete:niveau
  lsvn<-lsmeans(anmixte2.lmer,~variete:niveauN)

  # calcul des coefficients de variete:niveauN pour le modele mixte
  # Titre de chaque colonne de tableau
  W<-c("variete", "lsm v ", "lsm v*n-", "lsm v*n+", "v*n-", "v*n+", "lsmete v", "lsmete v*n-", "lsmete
v*n+", "mv*n-", "mv*n+")

  # Traiter la partie fixe pour la classe "mer"
  fixeef(anmixte2.lmer)
  fixeef(anmete2.lmer)

  # coefficients de variete:niveauN
  coeff<-c(fixeef(anmixte2.lmer) [25:46], 0-sum(fixeef(anmixte2.lmer) [25:46]))
  coeffete<-c(fixeef(anmete2.lmer) [25:46], 0-sum(fixeef(anmete2.lmer) [25:46]))

  # toutes les varietes dans le tableau
  V<-
c("aberavon", "alutus", "barlet", "bartango", "barutti", "bretagne", "cantalous", "centre", "clerpin", "danema
rk", "delphin", "expo", "gagny", "hongrie", "lactal", "limbos", "loporello", "mara", "milca", "ocean", "ponant"
, "roy", "stefani")

  # concatener toutes les parties du tableau
  P<-cbind(V,lsv$"variete lsmeans"[,2],lsvn$"variete:niveauN lsmeans"[1:23,3],lsvn$"variete:niveauN
lsmeans"[24:46,3],coeff,(0-coeff),
          lsvete$"variete lsmeans"[,2],"-","-",coeffete,(0-coeffete))
  P<-rbind(W,P)

  # sortir le tableau
  write(P,basewrite,paste("lsmeans_Multilocal_niche_Mmixte_Variete_",nom_variable,sep="")) # On
enregistre les resultats dans un fichier
}

#####
# Fonction permettant la fabrication d'un tableau complet #
# De moyennes ajustees par niveauN pour modele mixte #

```

```

# a partir des ajustements anmixte1.lmer, anmete1.lmer, anmixte2.lmer, anmete2.lmer
# (qui sont des variables globales)
#####
lsmeans_NiveauN<-function(basewrite,nom_variable) {

  # calcul des lsmeans de niveauN
  lsn2<-c("mixte2",lsmeans(anmixte2.lmer,~niveauN)$"niveauN lsmeans"[,2])
  lsnete2<-c("mixte2 ete",lsmeans(anmete2.lmer,~niveauN)$"niveauN lsmeans"[,2])

  # sortir le tableau
  H<-c("modele", "N-", "N+")
  H<-rbind(H,lsn2,lsnete2)
  write(H,basewrite,paste("lsmeans_Multilocal_niche_Mmixte_niveauN_",nom_variable,sep="")) # On
  enregistre les resultats dans un fichier
}

#####
# Fonction permettant le rangement dans un tableau des coefficients
# par variété du modèle fixe
# a partir des ajustements anfixe.lm et anfixe_ete.lm
# (qui sont des variables globales)
#####

coef_fixe<-function(basewrite,nom_variable) {
  coef1<-c(summary.lm(anfixe.lm)$coefficients[30:52,1],0-sum(summary.lm(anfixe.lm)
  $coefficients[30:52,1]))
  coef2<-c(summary.lm(anfixe_ete.lm)$coefficients[30:52,1],0-sum(summary.lm(anfixe_ete.lm)
  $coefficients[30:52,1]))
  C<-rbind(paste(nom_variable,"coef v:n-",sep="_"),cbind(coef1,coef2))
  write(C,basewrite,paste(nom_variable,"coef",sep=" "))
}

#####
# Importer et traiter les données #
#####

#importer les données
gie.grass<-read(baseread,"gie grass")
dlf<-read(baseread,"dlf")
barenbrug<-read(baseread,"barenbrug")
carneau<-read(baseread,"carneau")
r2n<-read(baseread,"r2n")

#####
# Traitement des données pour la variable "rdtsec" - ANNEE 2012 #
#####

# Année 2012
gie<-subset(gie.grass,subset=((valeur!="") & (codevar=="rdtsec") & ((datejour=="03/04/2012") |
(datejour=="19/05/2012") | (datejour=="14/06/2012") | (datejour=="18/07/2012") |
(datejour=="26/10/2012"))))
dlf<-subset(dlf,subset=((valeur!="") & (codevar=="rdtsec") & ((datejour=="16/05/2012") |
(datejour=="19/06/2012") | (datejour=="18/07/2012") | (datejour=="06/09/2012") |
(datejour=="30/11/2012"))))
bar<-subset(barenbrug,subset=((valeur!="") & (codevar=="rdtsec") & ((datejour=="10/05/2012") |
(datejour=="01/06/2012") | (datejour=="28/06/2012") | (datejour=="26/07/2012") |
(datejour=="24/10/2012"))))
car<-subset(carneau,subset=((valeur!="") & (codevar=="rdtsec") & ((datejour=="30/04/2012") |
(datejour=="25/05/2012") | (datejour=="25/06/2012") | (datejour=="22/08/2012"))))
r2<-subset(r2n,subset=((valeur!="") & (codevar=="rdtsec") & ((datejour=="12/04/2012") |
(datejour=="10/05/2012") | (datejour=="05/06/2012") | (datejour=="25/07/2012") | (datejour=="30/07/2012") |
(datejour=="28/08/2012") | (datejour=="28/09/2012"))))

# Ete 2012
gieete<-subset(gie.grass,subset=((valeur!="") & (codevar=="rdtsec") & ((datejour=="03/04/2012") |
(datejour=="19/05/2012") | (datejour=="14/06/2012"))))
dlete<-subset(dlf,subset=((valeur!="") & (codevar=="rdtsec") & ((datejour=="16/05/2012") |
(datejour=="19/06/2012") | (datejour=="18/07/2012"))))
barete<-subset(barenbrug,subset=((valeur!="") & (codevar=="rdtsec") & ((datejour=="10/05/2012") |
(datejour=="01/06/2012") | (datejour=="28/06/2012"))))
r2ete<-subset(r2n,subset=((valeur!="") & (codevar=="rdtsec") & ((datejour=="12/04/2012") |
(datejour=="10/05/2012") | (datejour=="05/06/2012"))))
carete<-subset(carneau,subset=((valeur!="") & (codevar=="rdtsec") & ((datejour=="30/04/2012") |
(datejour=="25/05/2012") | (datejour=="25/06/2012"))))

```

```

# Annee 2012 sans la variete "base"
dlmixte<-subset(dlf, subset=( (valeur!="") & (codevar=="rdtsec") & (variete!
=="base") & ((datejour=="16/05/2012") | (datejour=="19/06/2012") | (datejour=="18/07/2012") |
(datejour=="06/09/2012") | (datejour=="30/11/2012"))))
carmixte<-subset(carneau, subset=( (valeur!="") & (codevar=="rdtsec") & (variete!
=="base") & ((datejour=="30/04/2012") | (datejour=="25/05/2012") | (datejour=="25/06/2012") |
(datejour=="22/08/2012"))))
carmete<-subset(carneau, subset=( (valeur!="") & (codevar=="rdtsec") & (variete!
=="base") & ((datejour=="30/04/2012") | (datejour=="25/05/2012") | (datejour=="25/06/2012"))))

# ETE 2012 sans la variete "base"
r2mixte<-subset(r2n, subset=( (valeur!="") & (codevar=="rdtsec") & (variete!
=="base") & ((datejour=="12/04/2012") | (datejour=="10/05/2012") | (datejour=="05/06/2012") |
(datejour=="25/07/2012") | (datejour=="30/07/2012") | (datejour=="28/08/2012") |
(datejour=="28/09/2012"))))
r2mete<-subset(r2n, subset=( (valeur!="") & (codevar=="rdtsec") & (variete!
=="base") & ((datejour=="12/04/2012") | (datejour=="10/05/2012") | (datejour=="05/06/2012"))))
dlmete<-subset(dlf, subset=( (valeur!="") & (codevar=="rdtsec") & (variete!
=="base") & ((datejour=="16/05/2012") | (datejour=="19/06/2012") | (datejour=="18/07/2012"))))

# construire un tableau
D_annee2012<-rbind(gie, dl, bar, car, r2)
D_ete2012<-rbind(gieete, dlete, barete, r2ete, carete)
D_mixte_annee2012<-rbind(dlmixte, bar, carmixte, r2mixte)
D_mixte_ete2012<-rbind(dlmete, barete, carmete, r2mete)

# On rend explicite l'effet niche de bloc dans organisme
D_annee2012$bloc<-paste(D_annee2012$bloc, D_annee2012$organisme, sep=":")
D_ete2012$bloc<-paste(D_ete2012$bloc, D_ete2012$organisme, sep=":")
D_mixte_annee2012$bloc<-paste(D_mixte_annee2012$bloc, D_mixte_annee2012$organisme, sep=":")
D_mixte_ete2012$bloc<-paste(D_mixte_ete2012$bloc, D_mixte_ete2012$organisme, sep=":")

D_annee2012$bloc<-factor(D_annee2012$bloc)
D_ete2012$bloc<-factor(D_ete2012$bloc)
D_mixte_annee2012$bloc<-factor(D_mixte_annee2012$bloc)
D_mixte_ete2012$bloc<-factor(D_mixte_ete2012$bloc)

# Sortie des resultats de l'analyse
ajustement_multilocal()
analyse_multilocale(basewrite, "rdtsec_2012")
lsmeans_Variete(basewrite, "rdtsec_2012")
lsmeans_NiveauN(basewrite, "rdtsec_2012")
coef_fixe(basewrite, "rdtsec_2012")

# sortie des resultats de l'ajustement avec modele fixe sous forme ".txt"
#####
LstM <- list(annmixte2.lmer, anmete2.lmer)
LstNames <- list("v2 mat", "vete2 mat")
c<-1
for (i in LstM) {
chemin=paste(basewrite, "resultats_anova_multilocale__niche_rdtsec_2012_", LstNames[[c]], ".txt", sep=""
)
write.table(capture.output(i), file=chemin)
c<-c+1
}

# Calcul des Lsmeans
#####

# On extrait les tableaux par niveaux forts et faibles
D_annee2012_Nplus<-subset(D_annee2012, subset=( (niveauN=="fort")))
D_annee2012_Nmoins<-subset(D_annee2012, subset=( (niveauN=="faible")))
D_ete2012_Nplus<-subset(D_ete2012, subset=( (niveauN=="fort")))
D_ete2012_Nmoins<-subset(D_ete2012, subset=( (niveauN=="faible")))

# On calcule les lsmeans par l'intermediaire du test SNK.test qui affiche les lsmeans
rdtsec_2012.snk<-SNK.test(anfixe.lm, "variete")
rdtsec_ete2012.snk<-SNK.test(anfixe_ete.lm, "variete")
anfixe_faible.lm<-
lm(valeur~variete*organisme+organisme/datejour+organisme/bloc+variete:organisme:datejour+organisme:d
atejour+variete:organisme:datejour, data=D_annee2012_Nmoins)
anfixe_fort.lm<-

```

```

lm(valeur~variete*organisme+organisme/datejour+organisme/bloc+variete:organisme:datejour+organisme:d
atejour+variete:organisme:datejour,data=D_annee2012_Nplus)
anfixe_ete_faible.lm<-
lm(valeur~variete*organisme+organisme/datejour+organisme/bloc+variete:organisme:datejour+organisme:d
atejour+variete:organisme:datejour,data=D_ete2012_Nmoins)
anfixe_ete_fort.lm<-
lm(valeur~variete*organisme+organisme/datejour+organisme/bloc+variete:organisme:datejour+organisme:d
atejour+variete:organisme:datejour,data=D_ete2012_Nplus)
rdtsec_2012_Nfaible.snk<-SNK.test(anfixe_faible.lm,"variete")
rdtsec_2012_Nfort.snk<-SNK.test(anfixe_fort.lm,"variete")
rdtsec_ete2012_Nfaible.snk<-SNK.test(anfixe_ete_faible.lm,"variete")
rdtsec_ete2012_Nfort.snk<-SNK.test(anfixe_ete_fort.lm,"variete")

# On enregistre les resultats des SNK.test dans des fichiers texte
# On utilisera ensuite ces fichiers texte pour recuperer les lsmeans et les coller dans un tableur
write.table(capture.output(rdtsec_2012.snk),file=paste(basewrite,"snk_rdtsec_2012",".txt",sep=""))
write.table(capture.output(rdtsec_ete2012.snk),file=paste(basewrite,"snk_rdtsec_ete2012",".txt",sep=
""))
write.table(capture.output(rdtsec_2012_Nfaible.snk),file=paste(basewrite,"snk_rdtsec_2012_Nfaible","
.txt",sep=""))
write.table(capture.output(rdtsec_2012_Nfort.snk),file=paste(basewrite,"snk_rdtsec_2012_Nfort",".txt
",sep=""))
write.table(capture.output(rdtsec_ete2012_Nfaible.snk),file=paste(basewrite,"snk_rdtsec_ete2012_Nfai
ble",".txt",sep=""))
write.table(capture.output(rdtsec_ete2012_Nfort.snk),file=paste(basewrite,"snk_rdtsec_ete2012_Nfort
",".txt",sep=""))

#####
# Traitement des donnees pour la variable "INN" - ANNEE 2012 #
#####

# Annee 2012
gie<-subset(gie.grass,subset=((valeur!="") & (codevar=="INN") & ((datejour=="03/04/2012") |
(datejour=="19/05/2012") | (datejour=="14/06/2012") | (datejour=="18/07/2012") |
(datejour=="26/10/2012"))))
dlf<-subset(dlf,subset=((valeur!="") & (datejour!="") & (codevar=="INN") & ((datejour=="16/05/2012") |
(datejour=="19/06/2012") | (datejour=="18/07/2012") | (datejour=="06/09/2012") |
(datejour=="30/11/2012"))))
bar<-subset(barenbrug,subset=((valeur!="") & (datejour!
=="") & (codevar=="INN") & ((datejour=="10/05/2012") | (datejour=="01/06/2012") | (datejour=="24/10/2012") |
(datejour=="28/06/2012") | (datejour=="26/07/2012"))))
car<-subset(carneau,subset=((valeur!="") & (datejour!="") & (codevar=="INN") & ((datejour=="30/04/2012") |
(datejour=="25/05/2012") | (datejour=="25/06/2012") | (datejour=="22/08/2012"))))
r2<-subset(r2n,subset=((valeur!="") & (datejour!="") & (codevar=="INN") & ((datejour=="12/04/2012") |
(datejour=="10/05/2012") | (datejour=="05/06/2012") | (datejour=="25/07/2012") | (datejour=="30/07/2012") |
(datejour=="28/08/2012") | (datejour=="28/09/2012"))))

# Ete 2012
gieete<-subset(gie.grass,subset=((valeur!="") & (codevar=="INN") & ((datejour=="03/04/2012") |
(datejour=="19/05/2012") | (datejour=="14/06/2012"))))
dlete<-subset(dlf,subset=((valeur!="") & (datejour!="") & (codevar=="INN") & ((datejour=="16/05/2012") |
(datejour=="19/06/2012") | (datejour=="18/07/2012"))))
barete<-subset(barenbrug,subset=((valeur!="") & (datejour!
=="") & (codevar=="INN") & ((datejour=="10/05/2012") | (datejour=="01/06/2012") | (datejour=="28/06/2012"))))
r2ete<-subset(r2n,subset=((valeur!="") & (datejour!="") & (codevar=="INN") & ((datejour=="12/04/2012") |
(datejour=="10/05/2012") | (datejour=="05/06/2012"))))
carete<-subset(carneau,subset=((valeur!="") & (datejour!
=="") & (codevar=="INN") & ((datejour=="30/04/2012") | (datejour=="25/05/2012") | (datejour=="25/06/2012"))))

# Annee 2012 sans la variete "base"
dlmixte<-subset(dlf,subset=((valeur!="") & (datejour!="") & (codevar=="INN") & (variete!
=="base") & ((datejour=="16/05/2012") | (datejour=="19/06/2012") | (datejour=="18/07/2012") |
(datejour=="06/09/2012") | (datejour=="30/11/2012"))))
carmixte<-subset(carneau,subset=((valeur!="") & (datejour!="") & (codevar=="INN") & (variete!
=="base") & ((datejour=="30/04/2012") | (datejour=="25/05/2012") | (datejour=="25/06/2012") |
(datejour=="22/08/2012"))))
carmete<-subset(carneau,subset=((valeur!="") & (datejour!="") & (codevar=="INN") & (variete!
=="base") & ((datejour=="30/04/2012") | (datejour=="25/05/2012") | (datejour=="25/06/2012"))))

# ETE 2012 sans la variete "base"
r2mixte<-subset(r2n,subset=((valeur!="") & (datejour!="") & (codevar=="INN") & (variete!
=="base") & ((datejour=="12/04/2012") | (datejour=="10/05/2012") | (datejour=="05/06/2012") |
(datejour=="25/07/2012") | (datejour=="30/07/2012") | (datejour=="28/08/2012") |
(datejour=="28/09/2012"))))

```

```

r2mete<-subset(r2n, subset=((valeur!="") & (datejour!="") & (codevar=="INN") & (variete!
=="base") & ((datejour=="12/04/2012") | (datejour=="10/05/2012") | (datejour=="05/06/2012"))))
dlmete<-subset(dlf, subset=((valeur!="") & (datejour!="") & (codevar=="INN") & (variete!
=="base") & ((datejour=="16/05/2012") | (datejour=="19/06/2012") | (datejour=="18/07/2012"))))

# construire un tableau
D_annee2012<<-rbind(gie, dl, bar, car, r2)
D_ete2012<<-rbind(gieete, dlete, barete, r2ete, carete)
D_mixte_annee2012<<-rbind(dlmixte, bar, carmixte, r2mixte)
D_mixte_ete2012<<-rbind(dlmete, barete, carmete, r2mete)

# On rend explicite l'effet niche de bloc dans organisme
D_annee2012$bloc<-paste(D_annee2012$bloc, D_annee2012$organisme, sep=":")
D_ete2012$bloc<-paste(D_ete2012$bloc, D_ete2012$organisme, sep=":")
D_mixte_annee2012$bloc<-paste(D_mixte_annee2012$bloc, D_mixte_annee2012$organisme, sep=":")
D_mixte_ete2012$bloc<-paste(D_mixte_ete2012$bloc, D_mixte_ete2012$organisme, sep=":")

D_annee2012$bloc<<-factor(D_annee2012$bloc)
D_ete2012$bloc<<-factor(D_ete2012$bloc)
D_mixte_annee2012$bloc<<-factor(D_mixte_annee2012$bloc)
D_mixte_ete2012$bloc<<-factor(D_mixte_ete2012$bloc)

# Sortie des resultats de l'analyse
ajustement_multilocal()
analyse_multilocale(basewrite, "INN_2012")
lsmeans_Variete(basewrite, "INN_2012")
lsmeans_NiveauN(basewrite, "INN_2012")
coef_fixe(basewrite, "INN_2012")

# sortie des resultats de l'ajustement avec modele fixe sous forme ".txt"
#####
LstM <- list(anmixte2.lmer, anmete2.lmer)
LstNames <- list("v2 mat", "vete2 mat")
c<-1
for (i in LstM) {
  chemin=paste(basewrite, "resultats_anova_multilocale_INN_2012_", LstNames[[c]], ".txt", sep="")
  write.table(capture.output(i), file=chemin)
  c<-c+1
}

# Calcul des Lsmeans
#####
# On extrait les tableaux par niveaux forts et faibles
D_annee2012_Nplus<-subset(D_annee2012, subset=((niveauN=="fort")))
D_annee2012_Nmoins<-subset(D_annee2012, subset=((niveauN=="faible")))
D_ete2012_Nplus<-subset(D_ete2012, subset=((niveauN=="fort")))
D_ete2012_Nmoins<-subset(D_ete2012, subset=((niveauN=="faible")))

# On calcule les lsmeans par l'intermediaire du test SNK.test qui affiche les lsmeans
INN_2012.snk<-SNK.test(anfixe.lm, "variete")
INN_ete2012.snk<-SNK.test(anfixe_ete.lm, "variete")
anfixe_faible.lm<-
lm(valeur~variete*organisme+organisme/datejour+organisme/bloc+variete:organisme:datejour+organisme:d
atejour+variete:organisme:datejour, data=D_annee2012_Nmoins)
anfixe_fort.lm<-
lm(valeur~variete*organisme+organisme/datejour+organisme/bloc+variete:organisme:datejour+organisme:d
atejour+variete:organisme:datejour, data=D_annee2012_Nplus)
anfixe_ete_faible.lm<-
lm(valeur~variete*organisme+organisme/datejour+organisme/bloc+variete:organisme:datejour+organisme:d
atejour+variete:organisme:datejour, data=D_ete2012_Nmoins)
anfixe_ete_fort.lm<-
lm(valeur~variete*organisme+organisme/datejour+organisme/bloc+variete:organisme:datejour+organisme:d
atejour+variete:organisme:datejour, data=D_ete2012_Nplus)
INN_2012_Nfaible.snk<-SNK.test(anfixe_faible.lm, "variete")
INN_2012_Nfort.snk<-SNK.test(anfixe_fort.lm, "variete")
INN_ete2012_Nfaible.snk<-SNK.test(anfixe_ete_faible.lm, "variete")
INN_ete2012_Nfort.snk<-SNK.test(anfixe_ete_fort.lm, "variete")

# On enregistre les resultats des SNK.test dans des fichiers texte
# On utilisera ensuite ces fichiers texte pour recuperer les lsmeans et les coller dans un tableur
write.table(capture.output(INN_2012.snk), file=paste(basewrite, "snk_INN_2012", ".txt", sep=""))
write.table(capture.output(INN_ete2012.snk), file=paste(basewrite, "snk_INN_ete2012", ".txt", sep=""))
write.table(capture.output(INN_2012_Nfaible.snk), file=paste(basewrite, "snk_INN_2012_Nfaible", ".txt",
sep=""))

```

```

write.table(capture.output(INN_2012_Nfort.snk), file=paste(basewrite,"snk_INN_2012_Nfort", ".txt", sep=
""))
write.table(capture.output(INN_ete2012_Nfaible.snk), file=paste(basewrite,"snk_INN_ete2012_Nfaible", ".txt", sep=""))
write.table(capture.output(INN_ete2012_Nfort.snk), file=paste(basewrite,"snk_INN_ete2012_Nfort", ".txt", sep=""))

#####
# Traitement des donnees pour la variable "mat" - ANNEE 2012 #
#####

# Annee 2012
gie<-subset(gie.grass, subset=( (valeur!="") & (codevar=="mat") & ( (datejour=="03/04/2012") |
(datejour=="19/05/2012") | (datejour=="14/06/2012") | (datejour=="18/07/2012") |
(datejour=="26/10/2012") )))
dl<-subset(dlf, subset=( (valeur!="") & (datejour!="") & (codevar=="mat") & ( (datejour=="16/05/2012") |
(datejour=="19/06/2012") | (datejour=="18/07/2012") | (datejour=="06/09/2012") |
(datejour=="30/11/2012") )))
bar<-subset(barenbrug, subset=( (valeur!="") & (datejour!
=="") & (codevar=="mat") & ( (datejour=="10/05/2012") | (datejour=="01/06/2012") | (datejour=="24/10/2012") |
(datejour=="28/06/2012") | (datejour=="26/07/2012") )))
car<-subset(carneau, subset=( (valeur!="") & (datejour!="") & (codevar=="mat") & ( (datejour=="30/04/2012") |
(datejour=="25/05/2012") | (datejour=="25/06/2012") | (datejour=="22/08/2012") )))
r2<-subset(r2n, subset=( (valeur!="") & (datejour!="") & (codevar=="mat") & ( (datejour=="12/04/2012") |
(datejour=="10/05/2012") | (datejour=="05/06/2012") | (datejour=="25/07/2012") | (datejour=="30/07/2012") |
(datejour=="28/08/2012") | (datejour=="28/09/2012") )))

# Ete 2012
gieete<-subset(gie.grass, subset=( (valeur!="") & (codevar=="mat") & ( (datejour=="03/04/2012") |
(datejour=="19/05/2012") | (datejour=="14/06/2012") )))
dlete<-subset(dlf, subset=( (valeur!="") & (datejour!="") & (codevar=="mat") & ( (datejour=="16/05/2012") |
(datejour=="19/06/2012") | (datejour=="18/07/2012") )))
barete<-subset(barenbrug, subset=( (valeur!="") & (datejour!
=="") & (codevar=="mat") & ( (datejour=="10/05/2012") | (datejour=="01/06/2012") | (datejour=="28/06/2012") )))
r2ete<-subset(r2n, subset=( (valeur!="") & (datejour!="") & (codevar=="mat") & ( (datejour=="12/04/2012") |
(datejour=="10/05/2012") | (datejour=="05/06/2012") )))
carete<-subset(carneau, subset=( (valeur!="") & (datejour!
=="") & (codevar=="mat") & ( (datejour=="30/04/2012") | (datejour=="25/05/2012") | (datejour=="25/06/2012") )))

# Annee 2012 sans la variete "base"
dlmixte<-subset(dlf, subset=( (valeur!="") & (datejour!="") & (codevar=="mat") & (variete!
=="base") & ( (datejour=="16/05/2012") | (datejour=="19/06/2012") | (datejour=="18/07/2012") |
(datejour=="06/09/2012") | (datejour=="30/11/2012") )))
carmixte<-subset(carneau, subset=( (valeur!="") & (datejour!="") & (codevar=="mat") & (variete!
=="base") & ( (datejour=="30/04/2012") | (datejour=="25/05/2012") | (datejour=="25/06/2012") |
(datejour=="22/08/2012") )))
carmete<-subset(carneau, subset=( (valeur!="") & (datejour!="") & (codevar=="mat") & (variete!
=="base") & ( (datejour=="30/04/2012") | (datejour=="25/05/2012") | (datejour=="25/06/2012") )))

# ETE 2012 sans la variete "base"
r2mixte<-subset(r2n, subset=( (valeur!="") & (datejour!="") & (codevar=="mat") & (variete!
=="base") & ( (datejour=="12/04/2012") | (datejour=="10/05/2012") | (datejour=="05/06/2012") |
(datejour=="25/07/2012") | (datejour=="30/07/2012") | (datejour=="28/08/2012") |
(datejour=="28/09/2012") )))
r2mete<-subset(r2n, subset=( (valeur!="") & (datejour!="") & (codevar=="mat") & (variete!
=="base") & ( (datejour=="12/04/2012") | (datejour=="10/05/2012") | (datejour=="05/06/2012") )))
dlmete<-subset(dlf, subset=( (valeur!="") & (datejour!="") & (codevar=="mat") & (variete!
=="base") & ( (datejour=="16/05/2012") | (datejour=="19/06/2012") | (datejour=="18/07/2012") )))

# construire un tableau
D_annee2012<<-rbind(gie, dl, bar, car, r2)
D_ete2012<<-rbind(gieete, dlete, barete, r2ete, carete)
D_mixte_annee2012<<-rbind(dlmixte, bar, carmixte, r2mixte)
D_mixte_ete2012<<-rbind(dlmete, barete, carmete, r2mete)

# On rend explicite l'effet niche de bloc dans organisme
D_annee2012$bloc<-paste(D_annee2012$bloc, D_annee2012$organisme, sep=":")
D_ete2012$bloc<-paste(D_ete2012$bloc, D_ete2012$organisme, sep=":")
D_mixte_annee2012$bloc<-paste(D_mixte_annee2012$bloc, D_mixte_annee2012$organisme, sep=":")
D_mixte_ete2012$bloc<-paste(D_mixte_ete2012$bloc, D_mixte_ete2012$organisme, sep=":")

D_annee2012$bloc<-factor(D_annee2012$bloc)
D_ete2012$bloc<-factor(D_ete2012$bloc)
D_mixte_annee2012$bloc<-factor(D_mixte_annee2012$bloc)

```

```

D_mixte_ete2012$bloc<-factor(D_mixte_ete2012$bloc)

# Sortie des resultats de l'analyse
ajustement_multilocal()
analyse_multilocale(basewrite,"mat_2012")
lsmeans_Variete(basewrite,"mat_2012")
lsmeans_NiveauN(basewrite,"mat_2012")
coef_fixe(basewrite,"mat_2012")

# sortie des resultats de l'ajustement avec modele fixe sous forme ".txt"
#####
LstM <- list(anmixte2.lmer,anmete2.lmer)
LstNames <- list("v2 mat","vete2 mat")
c<-1
for (i in LstM) {
  chemin=paste(basewrite,"resultats_anova_multilocale_mat_2012_",LstNames[[c]],".txt",sep="")
  write.table(capture.output(i),file=chemin)
  c<-c+1
}

# Calcul des Lsmeans
#####
# On extrait les tableaux par niveaux forts et faibles
D_annee2012_Nplus<-subset(D_annee2012,subset=((niveauN=="fort")))
D_annee2012_Nmoins<-subset(D_annee2012,subset=((niveauN=="faible")))
D_ete2012_Nplus<-subset(D_ete2012,subset=((niveauN=="fort")))
D_ete2012_Nmoins<-subset(D_ete2012,subset=((niveauN=="faible")))

# On calcule les lsmeans par l'intermediaire du test SNK.test qui affiche les lsmeans
mat_2012.snk<-SNK.test(anfixe.lm,"variete")
mat_ete2012.snk<-SNK.test(anfixe_ete.lm,"variete")

anfixe_faible.lm<-
lm(valeur~variete*organisme+organisme/datejour+organisme/bloc+variete:organisme:datejour+organisme:d
atejour+variete:organisme:datejour,data=D_annee2012_Nmoins)
anfixe_fort.lm<-
lm(valeur~variete*organisme+organisme/datejour+organisme/bloc+variete:organisme:datejour+organisme:d
atejour+variete:organisme:datejour,data=D_annee2012_Nplus)
anfixe_ete_faible.lm<-
lm(valeur~variete*organisme+organisme/datejour+organisme/bloc+variete:organisme:datejour+organisme:d
atejour+variete:organisme:datejour,data=D_ete2012_Nmoins)
anfixe_ete_fort.lm<-
lm(valeur~variete*organisme+organisme/datejour+organisme/bloc+variete:organisme:datejour+organisme:d
atejour+variete:organisme:datejour,data=D_ete2012_Nplus)
mat_2012_Nfaible.snk<-SNK.test(anfixe_faible.lm,"variete")
mat_2012_Nfort.snk<-SNK.test(anfixe_fort.lm,"variete")
mat_ete2012_Nfaible.snk<-SNK.test(anfixe_ete_faible.lm,"variete")
mat_ete2012_Nfort.snk<-SNK.test(anfixe_ete_fort.lm,"variete")

# On enregistre les resultats des SNK.test dans des fichiers texte
# On utilisera ensuite ces fichiers texte pour recuperer les lsmeans et les coller dans un tableur
write.table(capture.output(mat_2012.snk),file=paste(basewrite,"snk_mat_2012",".txt",sep=""))
write.table(capture.output(mat_ete2012.snk),file=paste(basewrite,"snk_mat_ete2012",".txt",sep=""))
write.table(capture.output(mat_2012_Nfaible.snk),file=paste(basewrite,"snk_mat_2012_Nfaible",".txt",
sep=""))
write.table(capture.output(mat_2012_Nfort.snk),file=paste(basewrite,"snk_mat_2012_Nfort",".txt",sep=
""))
write.table(capture.output(mat_ete2012_Nfaible.snk),file=paste(basewrite,"snk_mat_ete2012_Nfaible","
.txt",sep=""))
write.table(capture.output(mat_ete2012_Nfort.snk),file=paste(basewrite,"snk_mat_ete2012_Nfort",".txt
",sep=""))

```

## 9.7. Annexe 7 : analyse de la variance dans le cas multilocal (deltaC13, teneurN et teneurC)

```
# On supprime tous les objets existants dans R
rm(list=ls())

# Chemin d'accès aux scripts R
basescripts<-"/Users/INRA/scripts/"

# execute le fichier initialisation.R
source(paste(basescripts,"initialisation.R",sep=""))

# Chargement de la librairie lme4 pour la fonction lmer et la fonction fixef
library(lme4)

# Chargement de la librairie car pour la fonction Anova
library(car)

# Chargement de la librairie lsmeans pour la fonction lsmeans
library(lsmeans)

# Chargement de la librairie agricolae pour le test "HSD.test"
library(agricolae)

# Chargement de la librairie multcomp pour le test glht
library(multcomp)

#####
# analyse de variance #
#####

# On utilise toujours les contrasts="contr.sum" pour chaque facteur dans toute cette section
options(contrasts=c("contr.sum","contr.sum"))

#####
# Fonction qui cree les ajustements avec un modele specifie
# Les variables creees anmixte2.lmer, anmete2.lmer sont globales
# De meme anfixe.lm, anfixe_ete.lm, anfixe.lm, anfixe_ete.lm sont globales
# pour pouvoir etre utilises dans les fonctions qui suivent
#####

ajustement_multilocal<-function() {
  # Construire le modele lineaire fixe en utilisant la fonction "lm"
  anfixe.lm<-
  lm(valeur~variete*niveauN*organisme+organisme/datejour+organisme/bloc+variete:organisme:datejour+niv
eauN:organisme:datejour+variete:niveauN:organisme:datejour,data=D)
  # construire des modeles lineaires mixtes
  anmixte2.lmer<-lmer(valeur~variete*niveauN+(1|organisme)+(1|organisme/bloc)+(1|
organisme/datejour)+(1|datejour:variete:niveauN:organisme),data=D)
}

#####
# Fonction qui fait des anova sur l'ensemble des lieux
#
#####
analyse_multilocale<-function(basewrite,nom_variable) {
  # Analyse de la variance avec "Anova(type=3)"
  anfixe<-Anova(anfixe.lm,type=3,singular.ok=TRUE)

  # analyse de la variance avec "Anova(type=3)" pour le modele mixte
  # il faut bien preciser quel test utiliser, ici on utilise le test de Fisher
  anmixte2<-Anova(anmixte2.lmer,type=3,singular.ok=TRUE,test.statistic="F")

  # On construit un tableau des resultats
  # Le tableau
  # premiere lignedu tableau
  T<-c("", "Fv", "Fn", "Fl", "Fv*n", "Fv*1", "Fn*1", "Fl*d", "Fl*b", "Fv*n*1", "Fv*1*d", "Fn*1*d", "Fv*n*1*d",
"FM2-v", "FM2-n", "FM2-v*n")
}
```

```

# concatener toutes les parties
C<-c("tout
1'annee", anfixe[2,4], anfixe[3,4], anfixe[4,4], anfixe[5,4], anfixe[6,4], anfixe[7,4], anfixe[8,4], anfixe[
9,4], anfixe[10,4], anfixe[11,4], anfixe[12,4], anfixe[13,4],
  anmixte2[2,4], anmixte2[3,4], anmixte2[4,4])
T<-rbind(T,C)

# sortir le tableau
write(T,basewrite,paste("analyse_multilocale_niche_",nom_variable,sep=""))
}

#####
# Fonction permettant la fabrication d'un tableau complet #
# De moyennes ajustees par variete,variete:niveau pour modele mixte 2 #
# a partir des ajustements anmixte2.lmer
# (qui sont des variables globales)
#####
lsmeans_Variete<-function(basewrite,nom_variable) {
  # calcul des lsmeans par variete

  lsv<-lsmeans(anmixte2.lmer,~variete)
  # calcul des lsmeans par variete:niveau
  lsvn<-lsmeans(anmixte2.lmer,~variete:niveauN)

  # calcul des coefficients de variete:niveauN pour le modele mixte 2
  # Titre de chaque colonne de tableau
  W<-c("variete","lsm v ","lsm v*n-","lsm v*n+","v*n-","v*n+","lsmete v","lsmete v*n-","lsmete
v*n+","mv*n-","mv*n+")

  # Traiter la partie fixe pour la classe "mer"
  fixe f(anmixte2.lmer)

  # coefficients de variete:niveauN
  coeff<-c(fixef(anmixte2.lmer)[26:48],0-sum(fixef(anmixte2.lmer)[26:48]))

  # toutes les varietes dans le tableau
  V<-levels(D$variete)

  # concatener toutes les parties du tableau
  P<-cbind(V,lsv$"variete lsmeans"[,2],lsvn$"variete:niveauN lsmeans"[1:24,3],lsvn$"variete:niveauN
lsmeans"[25:48,3],coeff,(0-coeff))
  P<-rbind(W,P)

  # sortir le tableau
  write(P,basewrite,paste("lsmeans_Multilocal_niche_Variete_",nom_variable,sep="")) # On enregistre
les resultats dans un fichier
}

#####
# Fonction permettant la fabrication d'un tableau complet #
# De moyennes ajustees par niveauN pour modele mixte 2
# a partir des ajustements anmixte2.lmer
# (qui sont des variables globales)
#####
lsmeans_NiveauN<-function(basewrite,nom_variable) {

  # calcul des lsmeans de niveauN
  lsn2<-c("mixte2",lsmeans(anmixte2.lmer,~niveauN)$"niveauN lsmeans"[,2])

  # sortir le tableau
  H<-c("modele","N-","N+")
  H<-rbind(H,lsn2)
  write(H,basewrite,paste("lsmeans_Multilocal_niche_niveauN_",nom_variable,sep="")) # On enregistre
les resultats dans un fichier
}

# Calcul des coefficient du modele
coef_fixe<-function(basewrite,nom_variable) {
  coef1<-c(summary.lm(anfixe.lm)$coefficients[30:52,1],0-sum(summary.lm(anfixe.lm)
$coefficients[30:52,1]))
  C<<-c(paste(nom_variable,"coef v:n-",sep=" "),coef1)
  write(C,basewrite,paste(nom_variable,"coef",sep=" "))
}

```

```

}

#####
# Importer et traiter les donnees #
#####

#importer les donnees
gie.grass<-read(baseread,"gie grass")
dlf<-read(baseread,"dlf")
barenbrug<-read(baseread,"barenbrug")
carneau<-read(baseread,"carneau")
r2n<-read(baseread,"r2n")

#####
# Traitement des donnees pour la variable "rapportC/N" - ANNEE 2012 #
#####

#pour carneau
carneau<-read(baseread,"carneau")
car<-subset(carneau,subset=((valeur!="") & (datejour!
=="") & (codevar=="rapportC/N") & (datejour=="25/05/2012")))

#pour r2n
r2n<-read(baseread,"r2n")
r2<-subset(r2n,subset=((valeur!="") & (datejour!
=="") & (codevar=="rapportC/N") & (datejour=="10/05/2012")))

#construire un tableau
D<-rbind(car,r2)

# On rend explicite l'effet niche de bloc dans organisme
D$bloc<-paste(D$bloc,D$organisme,sep=":")
D$bloc<-factor(D$bloc)

# Sortie des resultats de l'analyse
ajustement_multilocal()
analyse_multilocale(basewrite,"rapportCN_2012")
lsmeans_Variete(basewrite,"rapportCN_2012")
lsmeans_NiveauN(basewrite,"rapportCN_2012")

# sortie des resultats de l'ajustement avec modele fixe sous forme ".txt"
#####
LstM <- list(anmixte2.lmer)
LstNames <- list("v2 rapportCN")
c<-1
for (i in LstM) {
  chemin=paste(basewrite,"resultats_anova_multilocale_rapportCN_2012_",LstNames[[c]],".txt",sep="")
  write.table(capture.output(i),file=chemin)
  c<-c+1
}

# Calcul des Lsmeans
D_Nplus<-subset(D,subset=((niveauN=="faible")))
D_Nmoins<-subset(D,subset=((niveauN=="fort")))

rapportCN_2012.snk<-SNK.test(anfixe.lm,"variete")
anfixe_Nplus.lm<-
lm(valeur~variete*organisme+organisme/datejour+organisme/bloc+variete:organisme:datejour+organisme:d
atejour+variete:organisme:datejour,data=D_Nplus)
rapportCN_2012_Nplus.snk<-SNK.test(anfixe_Nplus.lm,"variete")
anfixe_Nmoins.lm<-
lm(valeur~variete*organisme+organisme/datejour+organisme/bloc+variete:organisme:datejour+organisme:d
atejour+variete:organisme:datejour,data=D_Nmoins)
rapportCN_2012_Nmoins.snk<-SNK.test(anfixe_Nmoins.lm,"variete")
write.table(capture.output(rapportCN_2012.snk),file=paste(basewrite,"snk_rapportCN_2012",".txt",sep=
""))
write.table(capture.output(rapportCN_2012_Nplus.snk),file=paste(basewrite,"snk_rapportCN_2012_Nfort"
,".txt",sep=""))
write.table(capture.output(rapportCN_2012_Nmoins.snk),file=paste(basewrite,"snk_rapportCN_2012_Nfaib
le",".txt",sep=""))

```

```
#####
# Traitement des donnees pour la variable "deltaC13" - ANNEE 2012 #
#####

#pour carneau
carneau<-read(baseread,"carneau")
car<-subset(carneau,subset=((valeur!="") & (datejour!="" & (codevar=="deltaC13") & (datejour=="25/05/2012"))))

#pour r2n
r2n<-read(baseread,"r2n")
r2<-subset(r2n,subset=((valeur!="") & (datejour!="") & (codevar=="deltaC13") & (datejour=="10/05/2012")))

#construire un tableau
D<-rbind(car, r2)

# On rend explicite l'effet niche de bloc dans organisme
D$bloc<-paste(D$bloc,D$organisme,sep=":")
D$bloc<-factor(D$bloc)

# Sortie des resultats de l'analyse
ajustement_multilocal()
analyse_multilocale(basewrite,"deltaC13_2012")
lsmeans_Variete(basewrite,"deltaC13_2012")
lsmeans_NiveauN(basewrite,"deltaC13_2012")
coef_fixe(basewrite,"deltaC13_2012")

# sortie des resultats de l'ajustement avec modele fixe sous forme ".txt"
#####
LstM <- list(anmixte2.lmer)
LstNames <- list("v2 deltaC13")
c<-1
for (i in LstM) {
  chemin=paste(basewrite,"resultats_anova_multilocale_deltaC13_2012_",LstNames[[c]],".txt",sep="")
  write.table(capture.output(i),file=chemin)
  c<-c+1
}

# Calcul des Lsmeans
#####
# On extrait les tableaux par niveaux forts et faibles
D_Nplus<-subset(D,subset=((niveauN=="faible")))
D_Nmoins<-subset(D,subset=((niveauN=="fort")))

# On calcule les lsmeans par l'intermediaire du test SNK.test qui affiche les lsmeans
deltaC13_2012.snk<-SNK.test(anfixe.lm,"variete")
anfixe_Nplus.lm<-
lm(valeur~variete*organisme+organisme/datejour+organisme/bloc+variete:organisme:datejour+organisme:d
atejour+variete:organisme:datejour,data=D_Nplus)
deltaC13_2012_Nplus.snk<-SNK.test(anfixe_Nplus.lm,"variete")
anfixe_Nmoins.lm<-
lm(valeur~variete*organisme+organisme/datejour+organisme/bloc+variete:organisme:datejour+organisme:d
atejour+variete:organisme:datejour,data=D_Nmoins)
deltaC13_2012_Nmoins.snk<-SNK.test(anfixe_Nmoins.lm,"variete")

# On enregistre les resultats des SNK.test dans des fichiers texte
# On utilisera ensuite ces fichiers texte pour recuperer les lsmeans et les coller dans un tableau
write.table(capture.output(deltaC13_2012.snk),file=paste(basewrite,"snk_deltaC13_2012",".txt",sep=""))
write.table(capture.output(deltaC13_2012_Nplus.snk),file=paste(basewrite,"snk_deltaC13_2012_Nfort",".txt",sep=""))
write.table(capture.output(deltaC13_2012_Nmoins.snk),file=paste(basewrite,"snk_deltaC13_2012_Nfaible", ".txt", sep=""))

#####
# Traitement des donnees pour la variable "teneurC" - ANNEE 2012 #
#####

#pour carneau
carneau<-read(baseread,"carneau")
car<-subset(carneau,subset=((valeur!="") & (datejour!="" & (codevar=="teneurC") & (datejour=="25/05/2012"))))
```

```

#pour r2n
r2n<-read(baseread,"r2n")
r2<-subset(r2n,subset=((valeur!="") & (datejour!="") & (codevar=="teneurC") & (datejour=="10/05/2012")))

#construire un tableau
D<-rbind(car,r2)

# On rend explicite l'effet niche de bloc dans organisme
D$bloc<-paste(D$bloc,D$organisme,sep=":")
D$bloc<-factor(D$bloc)

# Sortie des resultats de l'analyse
ajustement_multilocal()
analyse_multilocale(basewrite,"teneurC_2012")
lsmeans_Variete(basewrite,"teneurC_2012")
lsmeans_NiveauN(basewrite,"teneurC_2012")
coef_fixe(basewrite,"deltaC13_2012")

# sortie des resultats de l'ajustement avec modele fixe sous forme ".txt"
#####
LstM <- list(anmixte2.lmer)
LstNames <- list("v2_teneurC")
c<-1
for (i in LstM) {
  chemin=paste(basewrite,"resultats_anova_multilocale_teneurC_2012_",LstNames[[c]],".txt",sep="")
  write.table(capture.output(i),file=chemin)
  c<-c+1
}

# Calcul des Lsmeans
#####
# On extrait les tableaux par niveaux forts et faibles
D_Nplus<-subset(D,subset=((niveauN=="faible")))
D_Nmoins<-subset(D,subset=((niveauN=="fort")))

# On calcule les lsmeans par l'intermediaire du test SNK.test qui affiche les lsmeans
teneurC_2012.snk<-SNK.test(anfixe.lm,"variete")
anfixe_Nplus.lm<-
lm(valeur~variete*organisme+organisme/datejour+organisme/bloc+variete:organisme:datejour+organisme:d
atejour+variete:organisme:datejour,data=D_Nplus)
teneurC_2012_Nplus.snk<-SNK.test(anfixe_Nplus.lm,"variete")
anfixe_Nmoins.lm<-
lm(valeur~variete*organisme+organisme/datejour+organisme/bloc+variete:organisme:datejour+organisme:d
atejour+variete:organisme:datejour,data=D_Nmoins)
teneurC_2012_Nmoins.snk<-SNK.test(anfixe_Nmoins.lm,"variete")

# On enregistre les resultats des SNK.test dans des fichiers texte
# On utilisera ensuite ces fichiers texte pour recuperer les lsmeans et les coller dans un tableur
write.table(capture.output(teneurC_2012.snk),file=paste(basewrite,"snk_teneurC_2012",".txt",sep=""))
write.table(capture.output(teneurC_2012_Nplus.snk),file=paste(basewrite,"snk_teneurC_2012_Nfort",".t
xt",sep=""))
write.table(capture.output(teneurC_2012_Nmoins.snk),file=paste(basewrite,"snk_teneurC_2012_Nfaible",
".txt",sep=""))

#####
# Traitement des donnees pour la variable "teneurN" - ANNEE 2012 #
#####

#pour carneau
carneau<-read(baseread,"carneau")
car<-subset(carneau,subset=((valeur!="") & (datejour!
=="") & (codevar=="teneurN") & (datejour=="25/05/2012")))

#pour r2n
r2n<-read(baseread,"r2n")
r2<-subset(r2n,subset=((valeur!="") & (datejour!="") & (codevar=="teneurN") & (datejour=="10/05/2012")))

#construire un tableau
D<-rbind(car,r2)

# On rend explicite l'effet niche de bloc dans organisme
D$bloc<-paste(D$bloc,D$organisme,sep=":")

```

```

D$bloc<-factor(D$bloc)

# Sortie des resultats de l'analyse
ajustement_multilocal()
analyse_multilocale(basewrite,"teneurN_2012")
lsmeans_Variete(basewrite,"teneurN_2012")
lsmeans_NiveauN(basewrite,"teneurN_2012")
coef_fixe(basewrite,"deltaC13_2012")

# sortie des resultats de l'ajustement avec modele fixe sous forme ".txt"
#####
LstM <- list(anmixte2.lmer)
LstNames <- list("v2 teneurN")
c<-1
for (i in LstM) {
  chemin=paste(basewrite,"resultats_anova_multilocale_teneurN_2012_",LstNames[[c]],".txt",sep="")
  write.table(capture.output(i),file=chemin)
  c<-c+1
}

# Calcul des Lsmeans
#####
# On extrait les tableaux par niveaux forts et faibles
D_Nplus<-subset(D,subset=((niveauN=="faible")))
D_Nmoins<-subset(D,subset=((niveauN=="fort")))

# On calcule les lsmeans par l'intermediaire du test SNK.test qui affiche les lsmeans
teneurN_2012.snk<-SNK.test(anfixe.lm,"variete")
anfixe_Nplus.lm<-
lm(valeur~variete*organisme+organisme/datejour+organisme/bloc+variete:organisme:datejour+organisme:d
atejour+variete:organisme:datejour,data=D_Nplus)
teneurN_2012_Nplus.snk<-SNK.test(anfixe_Nplus.lm,"variete")
anfixe_Nmoins.lm<-
lm(valeur~variete*organisme+organisme/datejour+organisme/bloc+variete:organisme:datejour+organisme:d
atejour+variete:organisme:datejour,data=D_Nmoins)
teneurN_2012_Nmoins.snk<-SNK.test(anfixe_Nmoins.lm,"variete")

# On enregistre les resultats des SNK.test dans des fichiers texte
# On utilisera ensuite ces fichiers texte pour recuperer les lsmeans et les coller dans un tableur
write.table(capture.output(teneurN_2012.snk),file=paste(basewrite,"snk_teneurN_2012",".txt",sep=""))
write.table(capture.output(teneurN_2012_Nplus.snk),file=paste(basewrite,"snk_teneurN_2012_Nfort",".t
xt",sep=""))
write.table(capture.output(teneurN_2012_Nmoins.snk),file=paste(basewrite,"snk_teneurN_2012_Nfaible",
".txt",sep=""))

```

## 9.8. Annexe 8 : ACP par lieu

```

# On supprime tous les objets existants dans R
rm(list=ls())

# Chemin d'accès aux scripts R
basescripts<-" /Users/INRA/scripts/"

# execute le fichier initialisation.R
source(paste(basescripts,"initialisation.R",sep=""))

# Chargement de la librairie
library(FactoMineR)

ACP_lieu<-function(nom_fichier) {
  # Importation des données
  Data<-read.csv2(paste(baseread,nom_fichier,sep=""))
  # Suppression de la première ligne (en-tête)
  Data<-Data[,-1]
  # On calcule la matrice de corrélation pour confirmer qu'il y a bien des corrélations entre
  certaines variables
  Cor<-cor(Data)
  acp<-princomp(Data,cor=TRUE,scores=TRUE)
  # On ajoute une colonne pour le nom des variables
  if (nom_fichier != "acp barenbrug.csv")
    rownames(Data)<-
c("aberavon","alutus","barlet","bartango","barutti","base","bretagne","cantalous","centre","clerpin",
"danemark","delphin","expo","gagny","hongrie","lactal","limbos","loporello","mara","milca","ocean",
"ponant","roy","stefani")
  else
    rownames(Data)<-
c("aberavon","alutus","barlet","bartango","barutti","bretagne","cantalous","centre","clerpin","danema
rk","delphin","expo","gagny","hongrie","lactal","limbos","loporello","mara","milca","ocean",
"ponant","roy","stefani")

  # Valeurs propres (le carré correspond aux variances = valeurs propres).
  val.propres<-acp$sdev^2

  print(cumsum(val.propres/length(val.propres)*100))
  # print(acp$loadings)

  # Affichage de l'inertie des valeurs propres
  plot(1:length(val.propres),val.propres/length(val.propres)*100,type="b",ylab="pourcentage
d'inertie",xlab="Composante",main=nom_fichier)

  dev.new()
  plot(1:length(val.propres),cumsum(val.propres/length(val.propres)*100),type="b",ylab="pourcentage
d'inertie cumulée",xlab="Composante",main=nom_fichier)

  #carte des individus sur les 3 premiers axes
  dev.new()
  plot(acp$scores[,1],acp$scores[,2],type="n",xlab="Comp.1",ylab="Comp.2")
  abline(h=0,v=0)
  text(acp$scores[,1],acp$scores[,2],labels=rownames(Data),cex=0.5)

  dev.new()
  plot(acp$scores[,1],acp$scores[,3],type="n",xlab="Comp.1",ylab="Comp.3")
  abline(h=0,v=0)
  text(acp$scores[,1],acp$scores[,3],labels=rownames(Data),cex=0.5)

  dev.new()
  plot(acp$scores[,2],acp$scores[,3],type="n",xlab="Comp.2",ylab="Comp.3")
  abline(h=0,v=0)
  text(acp$scores[,2],acp$scores[,3],labels=rownames(Data),cex=0.5)

  #corrélation variable-facteur
  c1 <- acp$loadings[,1]*acp$sdev[1]
  c2 <- acp$loadings[,2]*acp$sdev[2]
  c3 <- acp$loadings[,3]*acp$sdev[3]
  c4 <- acp$loadings[,4]*acp$sdev[4]
  c5 <- acp$loadings[,5]*acp$sdev[5]

```

```
c6 <- acp$loadings[,6]*acp$sdev[6]
c7 <- acp$loadings[,7]*acp$sdev[7]

#affichage
correlation <- cbind(c1,c2,c3,c4,c5,c6,c7)
*** cercle des correlations - variables actives ***
dev.new()
plot(c1,c2,xlim=c(-1,+1),ylim=c(-1,+1),type="n",main="Correlation variables-facteur")
abline(h=0,v=0)
text(c1,c2,labels=colnames(Data),cex=0.5)
symbols(0,0,circles=1,inche=FALSE,add=TRUE)

#Carte des individus et des variables
biplot(acp,cex=c(0.5,0.5),main="Carte des individus et des variables")
}

# Organisme DLF
ACP_lieu("acp dlf.csv")

# Organisme Carneau
ACP_lieu("acp carneau.csv")

# Organisme Barenbrug
ACP_lieu("acp barenbrug.csv")

# Organisme grass
ACP_lieu("acp grass.csv")

# Organisme R2n
ACP_lieu("acp r2n INN.csv")

# Organisme R2n (deltaC13)
ACP_lieu("acp r2n deltaC13.csv")
```

## 9.9. Annexe 9 : ACP rdtsec/INN et rdtsec/mat

```
# On supprime tous les objets existants dans R
rm(list=ls())

# Chemin d'accès aux scripts R
basescripts<-"~/Users/INRA/scripts/"

# execute le fichier initialisation.R
source(paste(basescripts,"initialisation.R",sep=""))

# Chargement de la librairie
library(FactoMineR)

ACP_lieu<-function(nom_fichier) {
  # Importation des donnees
  Data<-read.csv2(paste(baseread,nom_fichier,".csv",sep=""))
  # Suppression de la premiere ligne (en-tete)
  Data<-Data[,-1]
  rownames(Data)<-
c("aberavon","alutus","barlet","bartango","barutti","base","bretagne","cantalous","centre","clerpin",
"danemark","delphin","expo","gagny","hongrie","lactal","limbos","loporello","mara","milca","ocea",
"ponant","roy","stefani")
  acp<-PCA(Data,graph=FALSE,ncp=7)
  # print(acp)

  print(cumsum(acp$eig/dim(acp$eig)[1]*100))

  summary(acp,nbelements=Inf,ncp=7,file=paste(basewrite,"resultats_",nom_fichier,".txt",sep=""))
  plot(acp,choix="ind",axes=c(1,2),cex=0.5,type="n")

  dev.new()
  plot(acp,choix="ind",axes=c(1,3),cex=0.5,type="n")

  dev.new()
  plot(acp,choix="ind",axes=c(2,3),cex=0.5,type="n")

  Cor<-round(cor(Data,use="complete.obs"),3)

write.table(capture.output(Cor),file=paste(basewrite,"cor_",nom_fichier,".txt",sep=""),row.names=FALSE,col.names=FALSE,quote=FALSE)
}

ACP_lieu("acp mat")
ACP_lieu("acp INN")
ACP_lieu("acp mat ete")
ACP_lieu("acp INN ete")
```

## 9.10. Annexe 10 : ACP multilocale

```

# On supprime tous les objets existants dans R
rm(list=ls())

# Chemin d'accès aux scripts R
basescripts<-" /Users/INRA/scripts/"

# execute le fichier initialisation.R
source(paste(basescripts,"initialisation.R",sep=""))

# Chargement de la librairie
library(FactoMineR)

# Importation des données
Data<-read.csv2(paste(baseread,"acp multilocal fixe.csv",sep=""))
#Data<-read.csv2(paste(baseread,"acp multilocal mixte.csv",sep=""))

# Suppression de la première ligne (en-tête)
Data<-Data[,-1]

rownames(Data)<-
c("aberavon","alutus","barlet","bartango","barutti","base","bretagne","cantalous","centre","clerpin",
"danemark","delphin","expo","gagny","hongrie","lactal","limbos","loporello","mara","milca","ocean",
"ponant","roy","stefani")
#rownames(Data)<-
c("aberavon","alutus","barlet","bartango","barutti","bretagne","cantalous","centre","clerpin","danema
rk","delphin","expo","gagny","hongrie","lactal","limbos","loporello","mara","milca","ocean",
"ponant","roy","stefani")

acp<-PCA(Data,graph=FALSE)

print(acp)
tout<-summary(acp,nbelements=Inf)

val.propre<-round(acp$eig[1:4,],2)
cumsum(val.propre)

barplot(acp$eig[,2],names=paste("Dim",1:nrow(acp$eig)))

dimdesc(acp)

plot(acp,choix="ind",axes=c(1,2),cex=0.5)
plot(acp,choix="ind",axes=c(1,3),cex=0.5)
plot(acp,choix="ind",axes=c(2,3),cex=0.5)

plot(acp,choix="var",axes=c(1:2),cex=0.5,select="contrib 12")
plot(acp,choix="var",axes=c(1,3),cex=0.5,select="contrib 12")
plot(acp,choix="var",axes=c(2:3),cex=0.5,select="contrib 12")

round(scale(Data),2)
matrice.cor<-round(cor(Data,use="complete.obs"),2)

write.table(capture.output(summary(acp,nbelements=Inf),matrice.cor),file="/Users/Juan/Dropbox/stage
inria/analyse de la variance/resultat/acp multilocal
fixe.txt",row.names=FALSE,col.names=FALSE,quote=FALSE)

pc<-princomp(Data,cor=TRUE)
biplot(pc,cex=c(0.5,0.5),main="Carte des individus et des variables")

```