

L'enseignement des bio-statistiques aux étudiants de licence de biologie

Un dialogue entre biologistes et mathématiciens

Judith Legrand, Nathalie Castelle

Elsa Bonnaud, Christine Dillmann, Elodie Marchadier, Christine Vassiliadis

Architecture de la formation en mathématiques dans la Licence Sciences de la Vie

...du fondamental à l'interdisciplinarité

Fonctions,
probabilités

Variables,
représentations
graphiques

Tests statistiques
Programmation R,
algorithmique

Démarche intégrative
Analyse de problèmes
concrets

Etude de dynamique temporelle
Modélisation,
Analyse graphique

L1

Mathématiques

Méthodologie

L2

Biostatistiques

Ecologie &
statistiques

Algorithmique

UEs
optionnelles :
analyse de
données, bio-
informatiques,
biomaths

L3

Biomathématiques

UEs
optionnelles :
bioinformatique

UEs obligatoires

Ues obligatoires

UEs
optionnelles

Historique de la démarche

- UE de biologie quantitative, Pierre-Henri Gouyon (MNHN) et Claire Lavigne (INRA) : « Comment les mathématiques permettent de mieux comprendre l'écologie, en particulier les cycles de vie des organismes, la génétique des populations et l'évolution des populations? »
- « En tant que biologiste, j'ai ressenti le besoin d'interagir avec des mathématiciens pour comprendre donc mieux enseigner la biologie quantitative. Le formalisme mathématique me manquait mais je sentais que c'était important que le biologiste soit aussi présent pour montrer, à travers des exemples du vivant, l'intérêt des mathématiques[...]. L'idée était que les étudiants en bio aillent au-delà de leur appréhension, qu'ils s'amuse à comprendre la biologie tout en gardant la rigueur des mathématiques. »

➤ Enseignements en binôme créés par Nathalie Castelle et Delphine Sicard



D. Sicard

Architecture de la formation en mathématiques dans la Licence Sciences de la Vie

...du fondamental à l'interdisciplinarité

Fonctions,
probabilités

Variables,
représentations
graphiques

Tests statistiques
Programmation R,
algorithmique

Démarche intégrative
Analyse de problèmes
concrets

Etude de dynamique temporelle
Modélisation,
Analyse graphique

L1

Mathématiques

Méthodologie

L2

Biostatistiques

Ecologie &
statistiques

Algorithmique

UEs
optionnelles :
analyse de
données, bio-
informatiques,
biomaths

L3

Biomathématiques

UEs
optionnelles :
bioinformatique

UEs obligatoires

Ues obligatoires

UEs
optionnelles

Des objectifs d'apprentissage progressifs

UE bio-
statistiques

UE Ecologie
& Statistiques

Représenter les données

Notions de probabilité

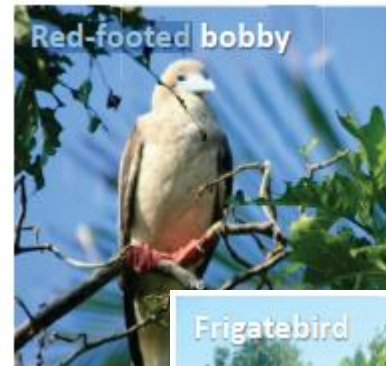
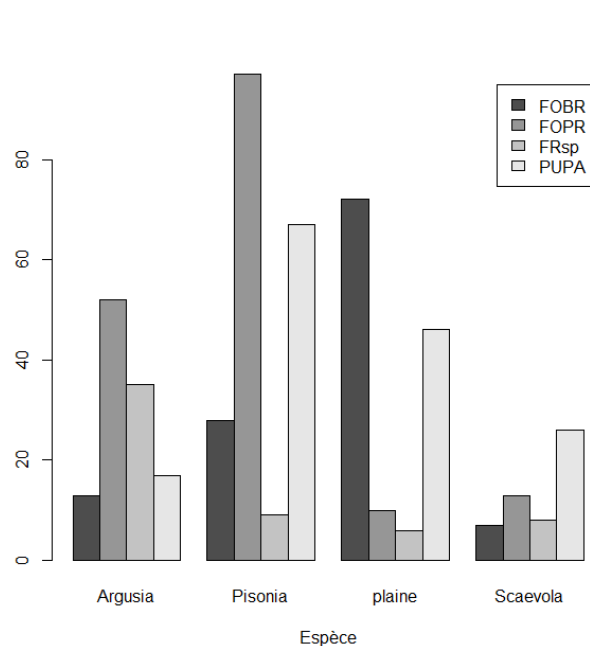
Principe des tests (7 points)

Connaitre les tests classiques

Choisir le test approprié

Réaliser un test statistique dans R

Interpréter d'un point de vue biologique



Des objectifs d'apprentissage progressifs

UE bio-
statistiques

UE Ecologie
& Statistiques

Représenter les données

Notions de probabilité

Décrire une variable aléatoire

Identifier la nature d'une variable aléatoire qualitative ou quantitative

Décrire la distribution d'une variable aléatoire selon sa nature

Principe des tests (7 points)

Connaitre les tests classiques

Choisir le test approprié

Réaliser un test statistique dans R

Interpréter d'un point de vue biologique

X_i : la hauteur de la i -ème plante

Y_i : la couleur des graines de la i -ème plante



Des objectifs d'apprentissage progressifs

UE bio-
statistiques

UE Ecologie
& Statistiques

Représenter les données

Notions de probabilité

Principe des tests (7 points)

Modéliser un problème biologique

Déterminer les hypothèses

De la statistique de test à la conclusion

Connaitre les tests classiques

Choisir le test approprié

Réaliser un test statistique dans R

Interpréter d'un point de vue biologique

Le test en 7 points :

1. Le modèle
2. Les hypothèses à tester
3. La statistique de test et sa loi sous H_0

4. Risque de première espèce et zone de rejet
5. Application numérique
6. P-value
7. Conclusion statistiques et biologique

Des objectifs d'apprentissage progressifs

UE bio-statistiques

**UE Ecologie
& Statistiques**

Représenter les données			
Notions de probabilité			
Principe des tests (7 points)			
Connaitre les tests classiques			
Test gaussien sur la moyenne	(1 population)		
Test de conformité de Student	(1 population)		
Test de conformité du Khi2	(1 population)		
Test du Khi2 d'indépendance	(1 population)		
Test d'homogeneité de Student	(2 populations)		
Test de comparaison de variance	(2 populations)		
Tests approximation gaussienne	(1 ou 2 populations)		
Choisir le test approprié			
Réaliser un test statistique dans R			
Interpréter d'un point de vue biologique			

$$T \sim_{H_0} \text{Student}(n-1)$$

$$Q \sim_{H_0} \chi^2(d-1)$$

$$Z \sim_{H_0} N(0,1)$$

$$F \sim_{H_0} \text{Fisher}(n-1, m-1)$$

Des objectifs d'apprentissage progressifs

UE bio-
statistiques

UE Ecologie
& Statistiques

Représenter les données

Notions de probabilité

Principe des tests (7 points)

Connaitre les tests classiques

Choisir le test approprié

Ecrire le modèle & les hypothèses

Vérifier les conditions d'application

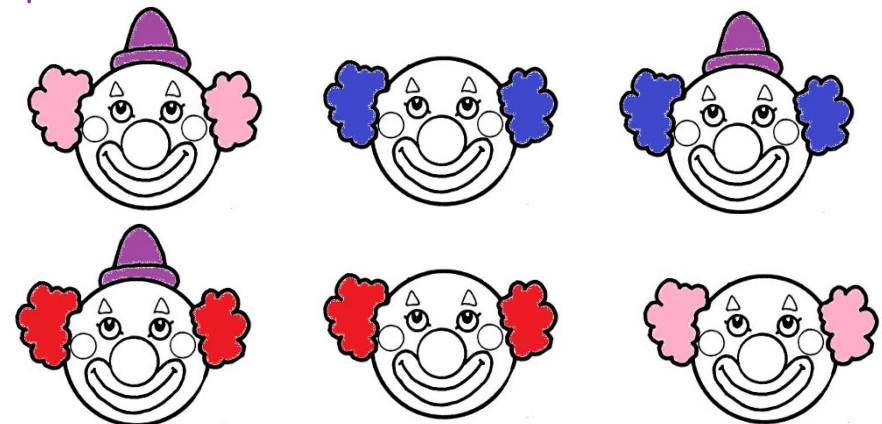
Réaliser un test statistique dans R

Interpréter d'un point de vue biologique

Pour chaque individu, je mesure deux traits quantitatifs, je veux étudier l'indépendance de ces traits, je choisis...

Chapeau?

Couleur des cheveux?



Des objectifs d'apprentissage progressifs

UE bio-
statistiques

UE Ecologie
& Statistiques

Représenter les données

Notions de probabilité

Principe des tests (7 points)

Connaitre les tests classiques

Choisir le test approprié

Réaliser un test statistique dans R

Adapter un script R à un nouveau jeu de données

Interpréter les sorties proposées par R

Interpréter d'un point de vue biologique

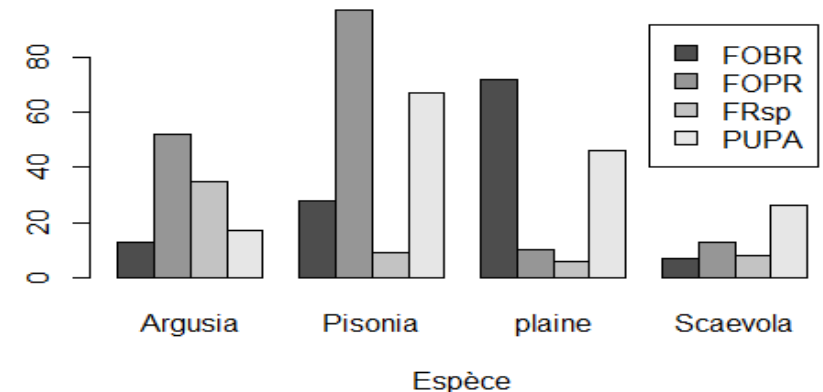


```
TP2_correction_etudiants.R* x
Source on Save
1 rm(list=ls())
2
3 #Q1
4 #chargement des données
5 tab<-read.table("oiseaux_student.csv",header=T,sep=";")
6 #affichage des 6 premières lignes
7 head(tab)
8 |
9 #on fabrique une table de contingence
10 contingence<-table(oiseaux2002[,c("espece","vg")])
11 print(contingence)
12
13 #representation de la distribution
14 x11()
15 barplot(contingence,beside=T,xlab="Espèce",legend.text =rownames(contingence))
16
17 ##réalisation du khi2 d'indépendance
18 khi2<-chisq.test(contingence)
19 print(khi2)
20 #effectifs observés et théoriques
21 khi2$observed
22
```

```
> chisq.test(contingence)

Pearson's Chi-squared test

data:  contingence
X-squared = 180.41, df = 9, p-value < 2.2e-16
> |
```



Des objectifs d'apprentissage progressifs

UE bio-
statistiques

UE Ecologie
& Statistiques

Représenter les données

Notions de probabilité

Principe des tests (7 points)

Connaitre les tests classiques

Choisir le test approprié

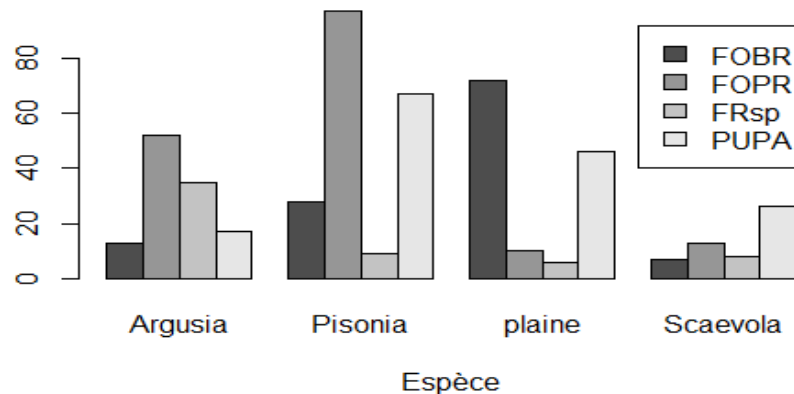
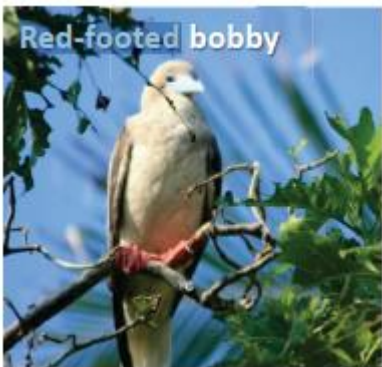
Réaliser un test statistique dans R

Interpréter d'un point de vue biologique

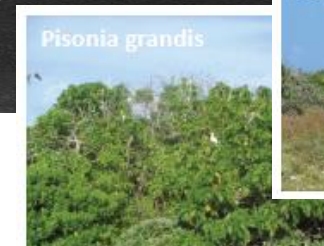
```
> chisq.test(contingence)

Pearson's Chi-squared test

data:  contingence
X-squared = 180.41, df = 9, p-value < 2.2e-16
> |
```



Les espèces ne colonisent pas les différents habitats de la même manière.



UE biostatistiques (S3)

Co-responsables Nathalie Castelle, Elodie Marchadier

Biologistes



Mathématiciens



5 cours magistraux (8 h)



2 TD de 3 heures de statistiques



4 TD de
1h30 de statistiques



1h30 de bio-statistiques



Exercice 1. Ecologie : estimation de l'abondance d'une population

Une des méthodes utilisées pour estimer l'abondance (nombre total d'individus présents) d'une espèce dans une zone géographique est la méthode de capture-marquage-recapture (CMR). Le principe est le suivant : on capture un nombre m d'individus que l'on marque, puis que l'on relâche. Dans un second temps, on capture au hasard n individus de la population, et on compte le nombre d'individus marqués, parmi ceux re-capturés. La connaissance de m , la proportion d'individus marqués dans l'échantillon permet d'estimer l'abondance a de la population.

Une étude réalisée de cette façon sur les écureuils de la forêt de Sénart a abouti à estimer leur effectif total à 10000 animaux, chiffre que l'on supposera connu par la suite. On pose donc $a = 10000$. On envisage le marquage de 3000 écureuils ($m = 3000$), leur relâche, puis la recapture de n écureuils.

1. Quelle sera la fréquence des écureuils marqués dans la population de la forêt de Sénart ? De façon générale, comment s'écrit cette fréquence p en fonction de m et a ?
2. Pour simuler une étape de recapture de $n = 10$ écureuils, vous allez tirer au hasard des écureuils dans une *population infinie* contenant 30 % d'individus marqués. On utilisera soit une table de

Exercice 1. Ecologie : estimation de l'abondance d'une population

Une des méthodes utilisées pour estimer l'abondance (nombre total d'individus présents) d'une espèce dans une zone est la méthode de capture-marquage-recapture. Le principe est le suivant : on capture une première fois une population d'individus. Dans un second temps, on capture une nouvelle fois la même population. Parmi les individus capturés, on compte le nombre d'individus marqués, dans l'échantillon. On peut alors estimer leur effectif total à 10000. On envisage le même exercice avec $a = 10000$ écureuils.

1. Quelle sera la façon générale de procéder ?

2. Pour simuler la capture-marquage-recapture, on utilise une table de nombres au hasard.



On utilise une table de nombres au hasard des écureuils.

Aide à la réussite et Modalités de contrôle des connaissances

- Statistiques
 - Deux devoirs WIMS notés, à la maison
 - Classe ouverte pendant une semaine pour l'entraînement
 - Un partiel de 2 heures
 - Un examen de 2 heures
- Biostatistiques
 - Exercices des feuilles de TD à rendre, facultatif
 - Un devoir sur table
 - Un examen d'1 heure

UE Ecologie & statistiques appliqués (S4)

Co-responsables Elsa Bonnaud, Nathalie Castelle, Judith Legrand



Écologues, biologistes



Mathématiciens

5 cours magistraux d'écologie (10 h)
Les menaces principales pour la
biodiversité*



2 cours magistraux de statistiques (3 h)



4 TP en salle informatique



* les changements climatiques, surexploitation, pollution, invasions biologiques, destruction/fragmentation des habitats

TP 3 : effet du changement d'usage des terres sur la diversité des coléoptères

1. Contexte général sur la thématique du changement d'usage des terres

2. Question biologique posée par les chercheurs

3. Découverte des données collectées et du plan d'expérience

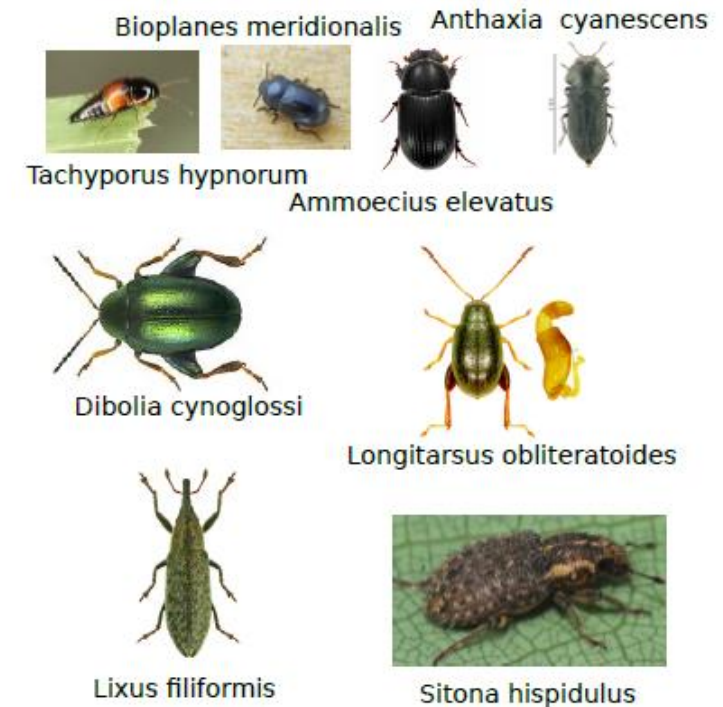
4. Représentation graphique des données

5. Modélisation et choix du test

6. Mise en œuvre sous R

7. Lecture des sorties et conclusion statistique

8. Interprétation biologique



TP 3 : effet du changement d'usage des terres sur la diversité des coléoptères

1. Contexte général sur la thématique du changement d'usage des terres

2. Question biologique posée par les chercheurs

3. Découverte des données collectées et du plan d'expérience

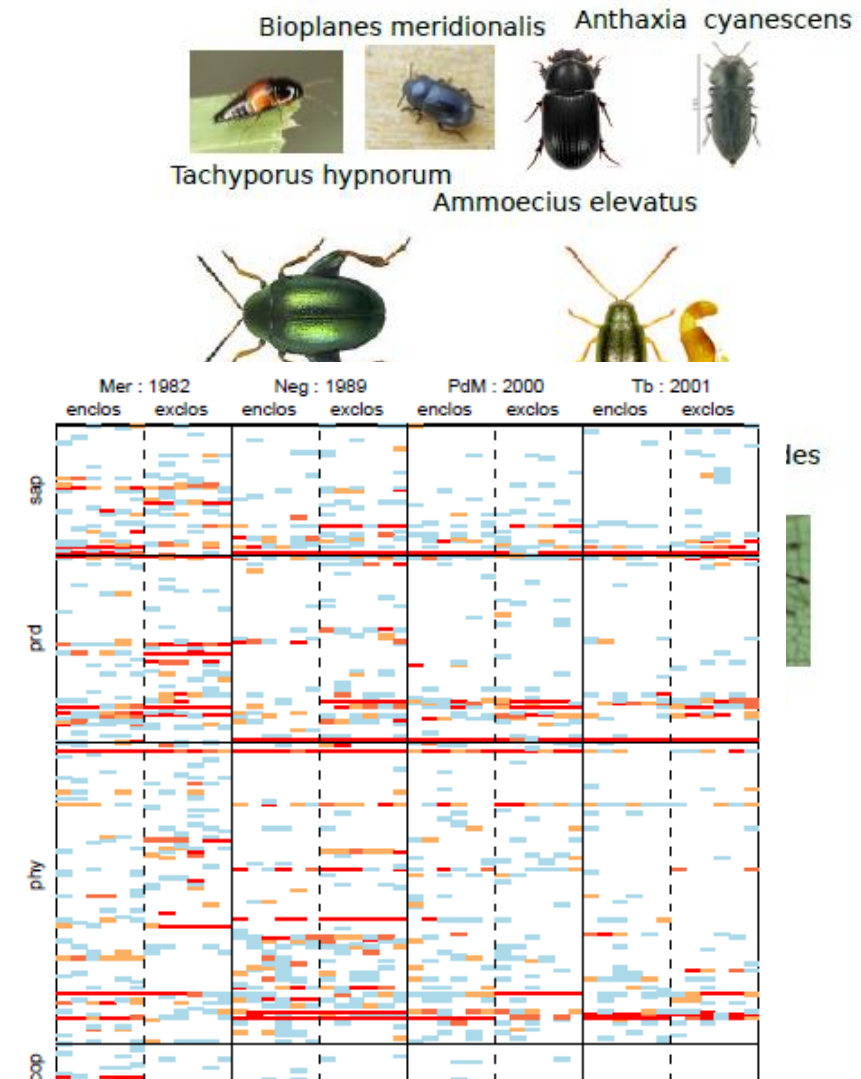
4. Représentation graphique des données

5. Modélisation et choix du test

6. Mise en œuvre sous R

7. Lecture des sorties et conclusion statistique

8. Interprétation biologique

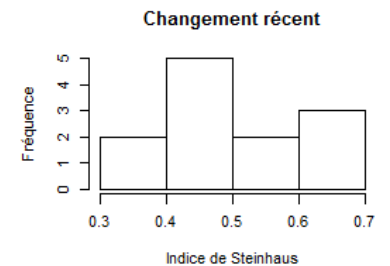
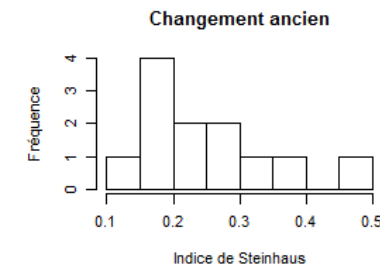
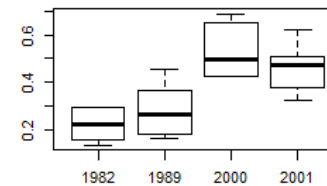


TP 3 : effet du changement d'usage des terres sur la diversité des coléoptères

1. Contexte général sur la thématique du changement d'usage des terres

2. Question biologique posée par les chercheurs

3. Découverte des données collectées et du plan d'expérience



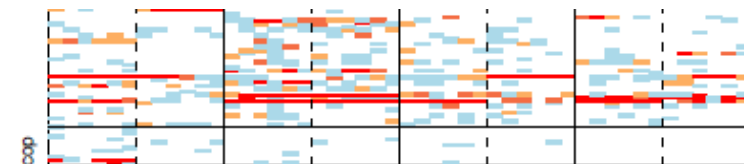
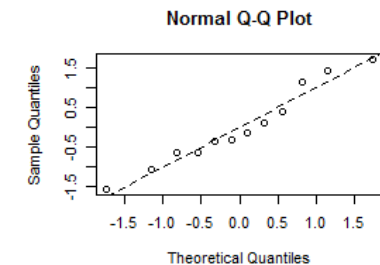
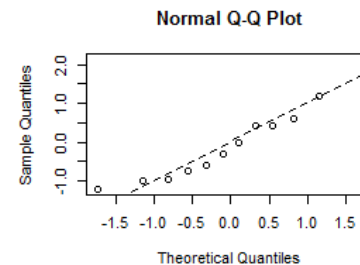
4. Représentation graphique des données avec R

5. Modélisation et choix du test

6. Mise en œuvre sous R

7. Lecture des sorties et conclusion statistique

8. Interprétation biologique



TP 3 : effet du changement d'usage des terres sur la diversité des coléoptères

1. Contexte général sur la thématique du changement d'usage des terres

2. Question biologique posée par les chercheurs

3. Découverte des données collectées et du plan d'expérience

4. Représentation graphique des données

5. Modélisation et choix du test

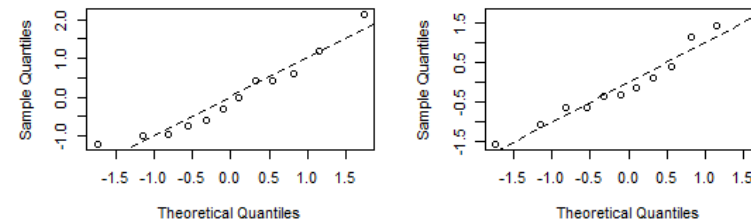
6. Mise en œuvre sous R

7. Lecture des sorties et conclusion statistique

8. Interprétation biologique



```
t.test(ancien,recent,alternative = c("less"),var.equal=TRUE)
##pour avoir la difference et l'intervalle de confiance
t.test(ancien,recent,var.equal=TRUE)
```



TP 3 : effet du changement d'usage des terres sur la diversité des coléoptères

1. Contexte général sur la thématique du changement d'usage des terres

2. Question biologique posée par les chercheurs

3. Découverte des données collectées et du plan d'expérience

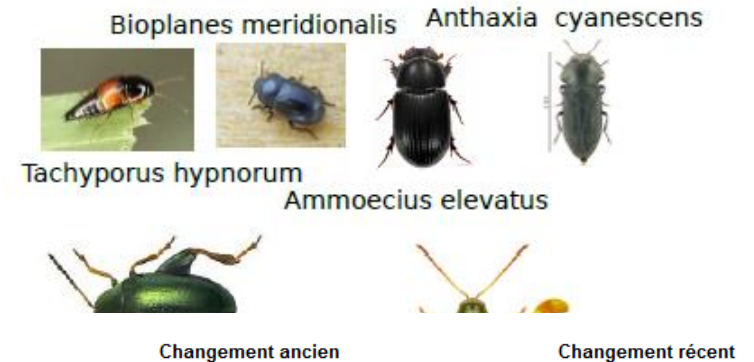
4. Représentation graphique des données

5. Modélisation et choix du test

6. Mise en œuvre sous R

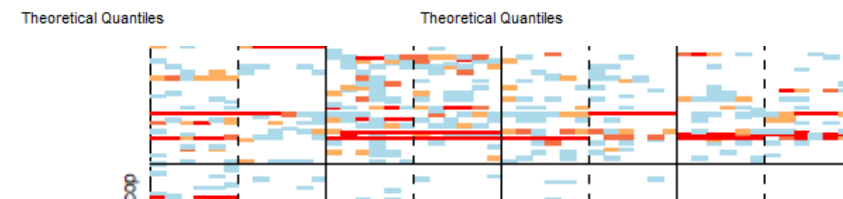
7. Lecture des sorties et conclusion statistique

8. Interprétation biologique



```
> ##pour avoir la difference et l'intervalle de confiance  
> t.test(ancien,recent,var.equal=TRUE)
```

```
t      Two Sample t-test  
# data:  ancien and recent  
t      t = -5.8131, df = 22, p-value = 7.549e-06  
alternative hypothesis: true difference in means is not equal to 0  
95 percent confidence interval:  
-0.3341015 -0.1583985  
sample estimates:  
mean of x mean of y  
0.2505833 0.4968333
```



TP 3 : effet du changement d'usage des terres sur la diversité des coléoptères

1. Contexte général sur la thématique du changement d'usage des terres

2. Question biologique posée par les chercheurs

3. Découverte des données collectées et du plan d'expérience

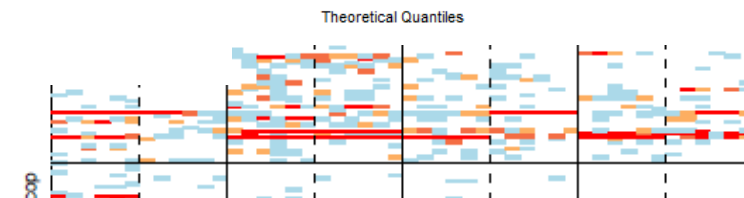
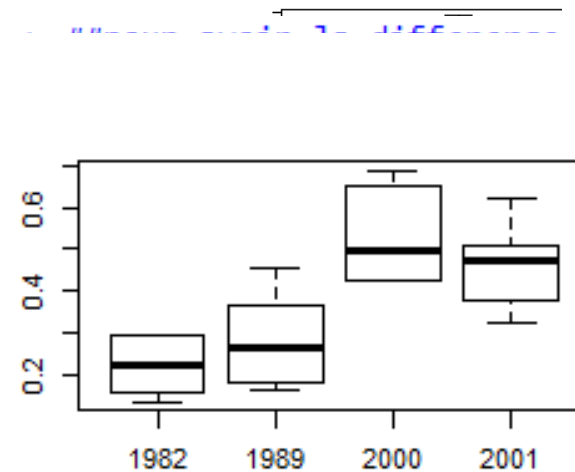
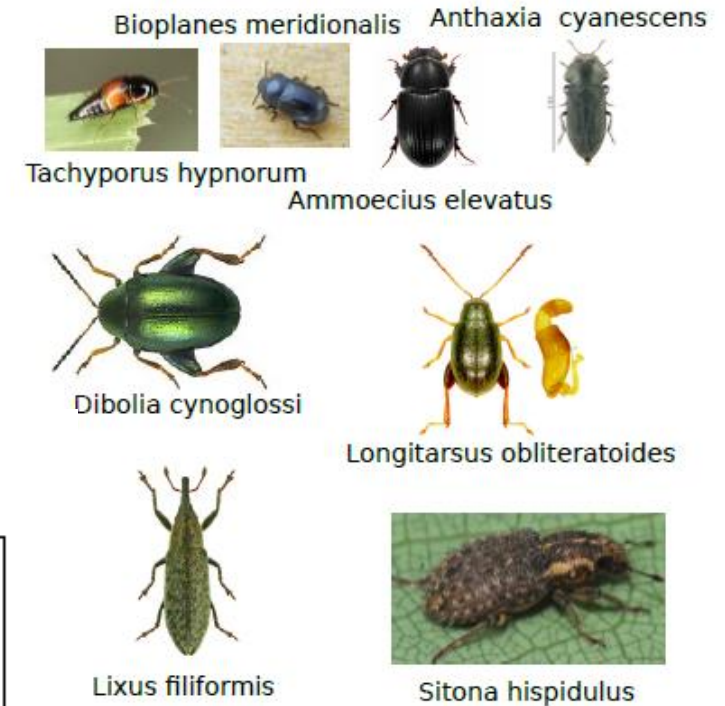
4. Représentation graphique des données

5. Modélisation et choix du test

6. Mise en œuvre sous R

7. Lecture des sorties et conclusion statistique

8. Interprétation biologique



Modalité de contrôle des connaissances et aide à la réussite

- 3 épreuves de contrôle continu
 - Un exercice de statistique sous WIMS, en TP
 - Un contrôle de cours d'écologie
 - Un contrôle sur ordinateur, analyse d'un jeu de données simple pour répondre à une question biologique
- Un examen en deux parties
 - Une partie écologie : analyses de graphiques et rédaction sur une question d'écologie
 - Une partie biostatistiques : autour d'un jeu de données d'écologie, répondre à plusieurs questions biologiques, notamment en utilisant des sorties graphiques et tests sous R

Commentaires, discussion

- Des bases à consolider sur la notion de distribution
- Poursuivre le travail sur le choix du test et les conditions d'application
- Encore plus d'intégration entre les disciplines d'appui et les enseignements de biologie

Exercice 9. Génétique des grains de maïs (ue GBMB)

Des croisements entre des plants de maïs appartenant à deux lignées pures différentes (nommées lignée I et lignée II), l'une à grains rouges lisses et l'autre à grains blancs ridés, produisant en première génération des épis à grains rouges lisses, ont donné en deuxième génération ($F_1 \times F_1$) des épis présentant les compositions suivantes :