

Devoir de « Logiciels statistiques » à réaliser avec et R Commander

Pour réaliser cet exercice, commencez par les étapes suivantes. Respectez **scrupuleusement** les consignes sous peine de perte de données et/ou fichiers qui entraînerait une note de 0.

- Créez, dans votre répertoire **P:\Travail**, un dossier nommé **DevoirLogiciel**. Y copiez le fichier trouvé dans le répertoire commun (répertoire villa) : c'est la version électronique de ce document. Ouvrez-le puis enregistrez-le à votre nom sous votre répertoire **P:\Travail\DevoirLogiciel** : *Fichier -> Enregistrer sous ...* ; nommez-le avec votre nom et prénom selon le schéma : **Devoir-NathalieVilla.doc**. **5 minutes avant la fin du devoir**, enregistrez le fichier **correctement nommé** : je le récupérerai directement dans votre dossier. Ne sortez qu'après vous être assuré que j'ai bien récupéré votre devoir.
- Sous un terminal serveur (Morta/Décima/Nona), lancer le logiciel R puis la librairie Commander : `> library(Rcmdr)`
Changer le répertoire de travail de manière à vous trouver sous le bon répertoire de votre **P:** puis enregistrez le script sous **Devoir.R** et le répertoire de travail sous **Devoir.Rdata**.
Attention !! Pensez à régulièrement sauver le script et le répertoire de travail. En cas de problème et de nécessité de redémarrer, la perte d'information sera considérée comme une erreur !
- Pour charger le fichier de données, utilisez la commande *Données -> Données dans les packages -> Lire des données depuis un package attaché*. Dans le menu de gauche, choisir « datasets » (double cliquez) puis, dans celui de droite, choisir « CO2 », comme indiqué dans la Figure 1.

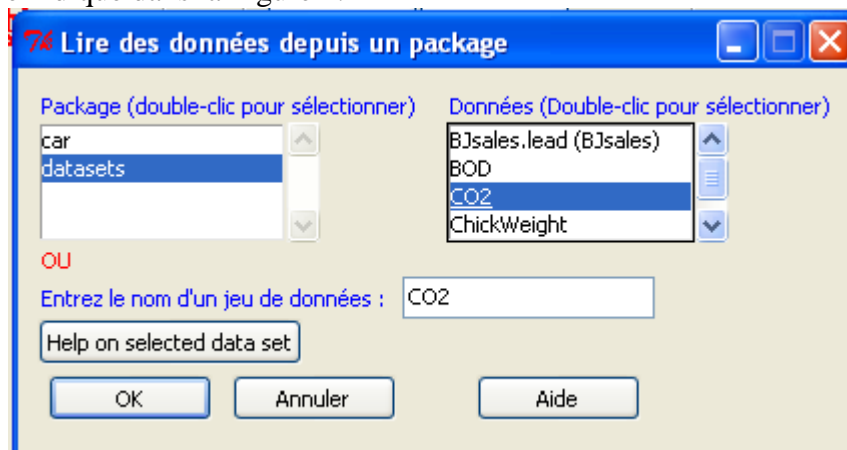


Figure 1 : Comment charger le jeu de données "CO2" ?

En cas de problème, vous pouvez aussi exécuter la commande :

```
data(CO2,package="datasets")
```

Une fois le fichier de données **co2** chargé et sélectionné, vous pouvez le visualiser. Il est composé de 84 observations correspondant à une espèce d'herbe dénommée *Echinochloa crus-galli* (appelée communément *Panic pied-de-coq*, cf Figure 2) et de 5 variables :

- **Plant** : l'identifiant de la plante (chaque plante est observée plusieurs fois) ;
- **type** : l'origine de la plante (Mississippi ou Quebec) ;

- **treatment** : une variable indiquant si la plante a été arrosée (chilled) ou non (nonchilled) ;
- **conc** : une variable numérique donnant la concentration de CO₂ dans l'air ambiant ;
- **uptake** : le taux de CO₂ consommé par la plante.



Figure 2 : Echinochloa crus-galli

La problématique posée par ces données est celle de l'influence du climat (par le biais de l'origine de la plante) sur la vivacité de celle-ci (par le biais de la consommation de CO₂).

Dans la suite, on demande, pour chaque question :

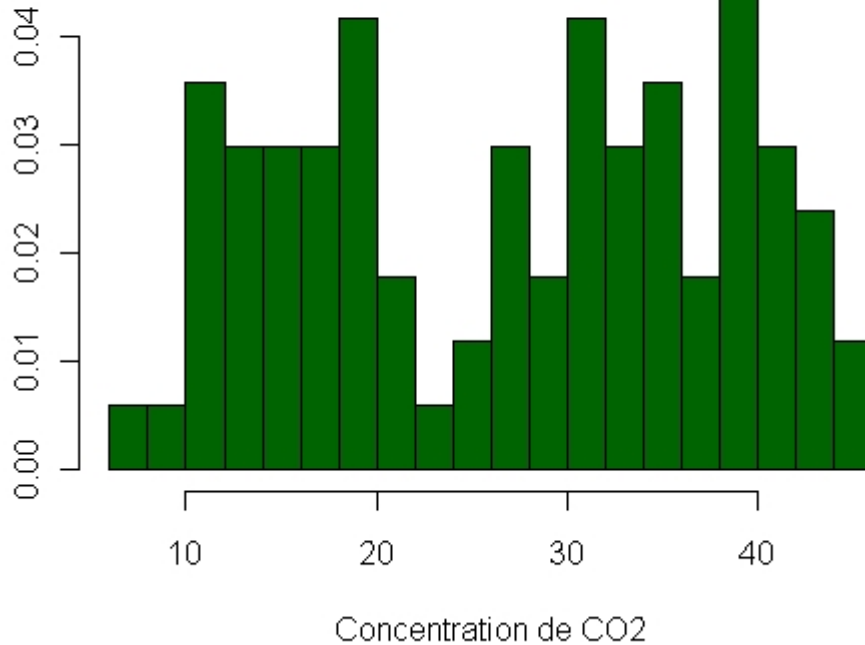
- de réaliser le travail demandé avec R et d'insérer les sorties correspondant à ce travail dans le présent document ;
- de copier la ou les commandes R (celles qui apparaissent dans la fenêtre de script) qui ont été nécessaires pour obtenir les sorties dans le cadre appelé : « Commandes R » ;
- éventuellement, de commenter les sorties.

1) Analyse univariée de la consommation de CO₂ (uptake)

Question 1

Effectuez l'histogramme de la consommation de CO₂ (uptake) en 20 classes de même amplitude. Au vu de cette figure, commentez la distribution de cette variable.

Distribution de la consommation de CO₂



Commentaires sur la distribution de la variable :

La distribution de la concentration de CO₂ est bimodale : elle contient deux pics qui peuvent, par exemple, correspondre aux deux origines des plantes étudiées.

Commandes R

```
hist(CO2$uptake,breaks=20,col="darkgreen",main="Distribution de la  
consommation de CO2",xlab="Concentration de CO2",ylab="",freq=F)
```

2) Analyse de l'influence de l'origine de la plante sur sa consommation de CO₂

Question 1

Donnez les statistiques conditionnelles (effectifs, moyennes et écarts type) de la consommation de CO₂ (`uptake`) selon l'origine de la plante (`type`). Commentez les résultats.

Statistiques conditionnelles :

	Québec	Mississippi
Effectifs	42	42
Moyennes	33,54	20,88
Écarts type	9,67	7,82

Commentaires : L'échantillonnage est équilibré entre les deux provenances (42 observations pour chacune des deux origines). On remarque que la consommation moyenne de CO₂ est plus élevée pour les plantes venant de Québec que pour celles venant du Mississippi. De même, la dispersion des valeurs de consommations est sensiblement plus élevée pour les plantes venant de Québec.

Commandes R

Le menu Statistiques -> Résumés -> Statistiques descriptives... rempli comme sur la Figure 3 permet d'obtenir les statistiques conditionnelles en générant le script suivant :

```
library(abind, pos=4)
numSummary(CO2[, "uptake"], groups=CO2$Type, statistics=c("mean", "sd"))
```

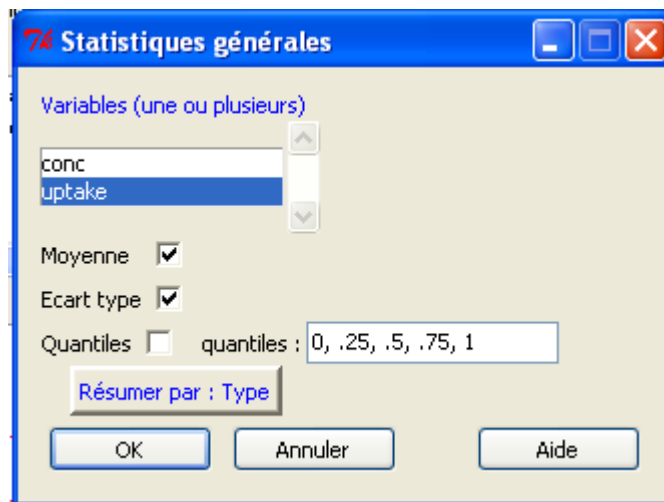
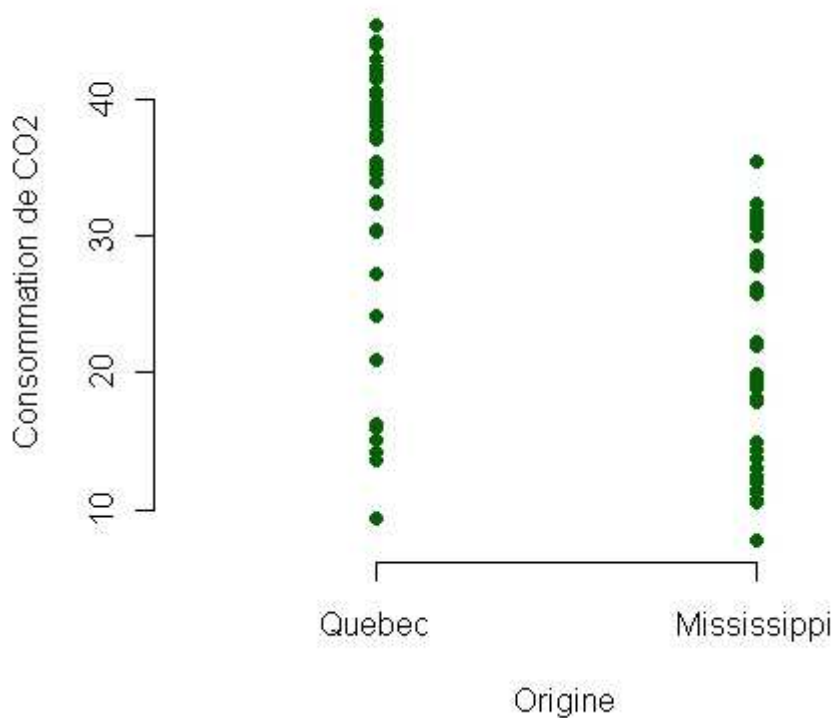


Figure 3 : Obtention des statistiques conditionnelles

Question 2

Effectuez le graphe plan de la distribution conditionnelle des variables « consommation de CO₂ » et « origine de la plante » (**uptake** et **Type**). Commentez le résultat.

Consommation de CO2 selon l'origine



Commentaires sur la distribution de la variable :

Le graphe plan confirme les conclusions de la question précédente : la distribution de la consommation de CO2 est plus importante et plus dispersée pour les plantes provenant de Québec que pour celles provenant du Mississippi.

Commandes R

Le menu *Données* -> *Gérer les variables dans le jeu de données actif* -> *Recoder des variables ...* rempli comme sur la Figure 4 permet d'obtenir un recodage de la variable `uptake` en variable numérique. Ce menu génère les commandes suivantes :

```
CO2$TypeNum<-recode(CO2$Type, '"Quebec"'=1;  
"Mississippi"'=2;', as.factor.result=FALSE)
```

Les commandes suivantes permettent ensuite d'obtenir le graphe plan :

```
plot(CO2$TypeNum, CO2$uptake, pch=19, col="darkgreen", xlab="Origine", ylab="C  
onsommation de CO2", axes=F, xlim=c(0.5, 2.5), main="Consommation de CO2  
selon l'origine")  
axis(2)  
axis(1, 1:2, c("Quebec", "Mississippi"))
```

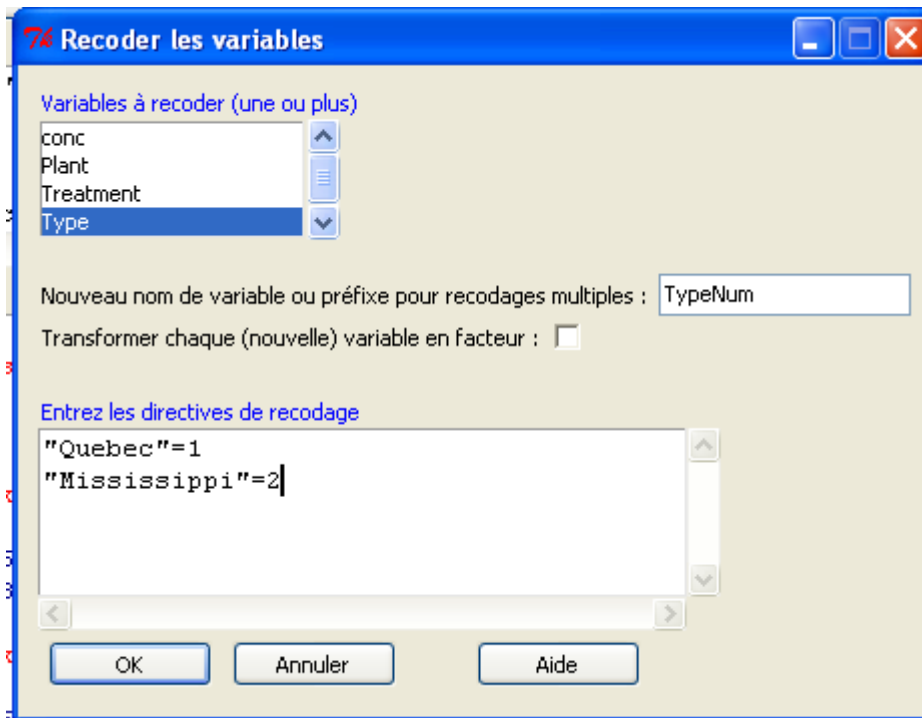


Figure 4 : Recodage de la variable Type en variable numérique

Question 3

Déterminer les variances inter-classes et intra-classes ainsi que le rapport de corrélation. Commentez le résultat.

Valeurs numériques :

Variance inter-classes : 40,066
Variance intra-classes : 75,493
Rapport de corrélation : 0,589

Commentaires sur le rapport de corrélation : L'influence de l'origine de la variable sur la consommation de CO₂ est moyenne : les plantes ont des niveaux de consommation de CO₂ différents selon leur origine mais d'autres facteurs doivent influencer le niveau de consommation de CO₂ des plantes.

Commandes R

Le menu *Statistiques -> Moyennes -> ANOVA à un facteur ...* rempli comme sur la Figure 5, permet d'obtenir les variances intra et inter classes.

```
library(multcomp, pos=4)
AnovaModel.1 <- aov(uptake~Type, data=CO2)
summary(AnovaModel.1)
numSummary(CO2$uptake, groups=CO2$Type, statistics=c("mean", "sd"))
```

Les effectifs des deux groupes puis les variances inter et intra-classes sont obtenues par les commandes suivantes :

```
Nj <- table(CO2$Type)
Vars <- summary(AnovaModel.1)[[1]]$"Sum Sq"/sum(Nj)
Vars
```

Enfin, le rapport de corrélation est obtenu par :

```
eta.sq <- Vars[1]/sum(Vars)
eta <- sqrt(eta.sq)
eta
```

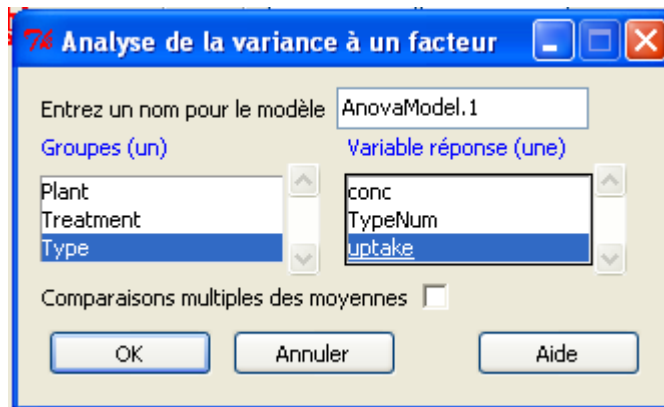


Figure 5 : Analyse de la variance