

Pratiques en langage R

Pierre Legendre
Département de sciences biologiques
Université de Montréal

Octobre 2004

1. Calcul des statistiques de base en langage R: exemple Merle

```
# Il faut d'abord indiquer à R quel est votre dossier de travail.  
# Menu File => Change dir...  
  
# Importer le fichier 'Merle' dans l'objet de type 'data frame' 'merle':  
merle <- read.table("Merle.txt")  
# ou encore  
merle = read.table("Merle.txt")  
  
# Vérification pour savoir si les données ont été lues correctement:  
merle  
  
# Saisie des valeurs de la première colonne: la longueur de l'aile.  
# La longueur est recopiée dans l'objet 'aile':  
aile=merle[,1]  
# Vérification du contenu de l'objet 'aile':  
aile  
  
# Transformation du vecteur 'aile' en un objet de type matrice 'aile.mat':  
aile.mat=as.matrix(aile)  
# Vérification du contenu de la matrice 'aile.mat':  
aile.mat  
  
# Calcul de la moyenne des longueurs d'aile (commande 'mean'):  
moyenne=mean(aile.mat)  
# Vérification de la valeur de la moyenne:  
moyenne  
  
# Calcul de la médiane des longueurs d'aile (commande 'median'):  
mediane=median(aile.mat)  
# Vérification de la valeur de la médiane:  
mediane  
  
# Calcul de la variance des longueurs d'aile (commande 'var'):  
variance=var(aile.mat)  
# Vérification de la variance:  
variance  
  
# Calcul de la taille de l'échantillon 'n':  
n=nrow(aile.mat)  
# Vérification de la valeur de 'n':  
n
```

```

# Calcul du coefficient d'asymétrie A3.
#
# Calcul préparatoire: estimation non biaisée du moment d'ordre 3, k3:
  k3=(n*sum((aile.mat-mean(aile.mat))^3))/((n-1)*(n-2))
# Vérification de k3:
  k3
#
# Calcul du coefficient d'asymétrie, noté A3:
  A3=k3/((sqrt(variance))^3)
# Vérification de A3:
  A3

# Calcul du coefficient d'aplatissement A4.
#
# Calcul préparatoire: estimation non biaisée du moment d'ordre 4, k4:
  k4=(n*(n+1)*(sum((aile.mat-mean(aile.mat))^4))-3*(n-1)*((sum((aile.mat-
mean(aile.mat))^2))^2))/((n-1)*(n-2)*(n-3))
# Vérification de k4:
  k4
#
# Calcul du coefficient d'aplatissement, noté A4:
  A4=k4/((sqrt(variance))^4)
# Vérification de A4:
  A4

# Calcul de l'étendue de la variation (ou plage de variation):
  etendue=max(aile.mat)-min(aile.mat)
# Vérification de l'étendue:
  etendue

# Calcul de l'écart type, noté Sx:
  sx=sd(aile.mat)
# Vérification de l'écart type:
  sx

# =====

# Préparation en vue de l'histogramme.

# Calcul du nombre de classes selon la règle de Sturge:
  classes=1+(3.3*log10(n))
# Vérification du nombre de classes:
  classes

# Il faut arrondir le nombre de classes donné par la règle de Sturge
# à l'entier le plus proche. Dans le cas présent, la règle de Sturge
# donne 6,2868. J'arrondis donc à 6 classes.
# Les données vont de 72 mm à 81 mm. L'étendue de variation est de 9 mm.
# Les classes peuvent couvrir chacune 2 mm de longueur d'aile.
# Je fixe la première borne à 71 mm.

```

```

# Voici les limites de classes choisies: 71,73,75,77,79,81,83.

# Indiquer à R les 7 bornes de classes choisies pour l'histogramme:
coupe=c(71,73,75,77,79,81,83)
# La commande 'c' a combiné les 7 valeurs en un vecteur appelé 'coupe'.

# Titre du graphique. Comme il est long, je l'ai coupé en deux parties
# pour qu'il soit affiché sur deux lignes.
titre=c("Figure 1: Histogramme de fréquences de la longueur d'aile (mm) de
merles", "d'Amérique mesurées au Jardin botanique de Montréal en août 2000.")
#
# Je recopie ensuite le titre dans un objet de type matrice:
titremat=matrix(titre)
# Comment sont placés les deux lignes de texte dans la matrice?
titremat

# Création de l'histogramme:
hist(aile.mat,breaks=coupe,freq=T,right=F,xlab=NULL,ylab="Fréquences
absolues",axes=T,main=NULL)

# Positionnement du titre sur l'histogramme:
mtext(text="Longueur de l'aile (mm)",side=1,line=2,cex=0.9,font=2)
mtext(text=titremat[1],side=1,line=3,cex=0.7,font=2)
mtext(text=titremat[2],side=1,line=4,cex=0.7,font=2)

# L'histogramme apparaît dans une fenêtre qu'on peut imprimer ou sauvegarder.
# La classe modale est celle qui a 80 mm comme indice de classe.

# =====

# Reprendre l'exercice pour la variable Masse(kg), Hauteur(cm) ou Longueur(cm) du fichier
'Ours.txt'. Attention à la lecture de ce fichier de données!

# =====

# Essayez les commandes suivantes pour les données d'ours:

summary(nom-du-data-frame)
plot(nom-du-data-frame)

# =====

```