Annexe 1

Jean-Yves Barnagaud - Olivier Gimenez

juin 2023

ATTENTION : MODIFIER LE CHEMIN D’ACCES VERS VOTRE REPERTOIRE COURANT AVANT DE LANCER CE SCRIPT

# librairies utiles et options du script

## librairies

# génériques  
library(tidyverse)

> ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
> ✔ dplyr 1.1.1 ✔ readr 2.1.4  
> ✔ forcats 1.0.0 ✔ stringr 1.5.0  
> ✔ ggplot2 3.4.2 ✔ tibble 3.2.1  
> ✔ lubridate 1.9.2 ✔ tidyr 1.3.0  
> ✔ purrr 1.0.1   
> ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
> ✖ dplyr::filter() masks stats::filter()  
> ✖ dplyr::lag() masks stats::lag()  
> ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(formatR)  
library(jtools)  
library(reshape2)

>   
> Attachement du package : 'reshape2'  
>   
> L'objet suivant est masqué depuis 'package:tidyr':  
>   
> smiths

# graphiques  
library(patchwork)  
library(viridis)

> Le chargement a nécessité le package : viridisLite

library(ggplot2)  
library(cowplot)

>   
> Attachement du package : 'cowplot'  
>   
> L'objet suivant est masqué depuis 'package:patchwork':  
>   
> align\_plots  
>   
> L'objet suivant est masqué depuis 'package:lubridate':  
>   
> stamp

## répertoire courant

Attention : le répertoire de travail est changé dans l’en-tête du script Markdown.

setwd("A:/Biotope\_analyse\_donnees/annexes")

## options graphiques

Thème par défaut des graphiques

theme\_set(theme\_light(base\_size = 16))

Reproductibilité des simulations

set.seed(150)

# 2- Comparer deux échantillons au regard d’une variable quantitative

données :

mesanges <-  
 read.csv2("donnees/mesanges\_Caro\_et\_al\_PRSLB\_2019.csv", dec = ".")  
mesanges

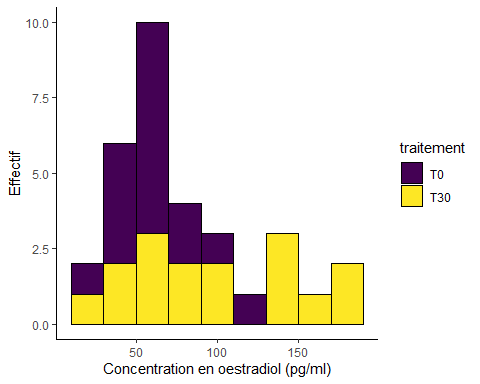
> individu T0 T30  
> 1 1 57.56761 108.74910  
> 2 2 48.59002 30.72740  
> 3 3 59.04845 186.49313  
> 4 4 28.78380 51.82936  
> 5 5 85.05568 156.50615  
> 6 6 99.77151 134.10846  
> 7 7 40.63051 86.62907  
> 8 8 64.32393 134.47867  
> 9 9 55.90166 180.93999  
> 10 10 59.04845 93.66305  
> 11 11 114.57990 146.41793  
> 12 12 56.91974 54.51338  
> 13 13 35.44758 42.01880  
> 14 14 39.61244 26.37744  
> 15 15 70.52495 87.36949  
> 16 16 51.36659 61.73247

on repasse ce jeu de données dans un format “long” plus facile à manipuler

mesanges.long <- pivot\_longer(mesanges, cols = c("T0", "T30"))  
colnames(mesanges.long) <- c("individu", "traitement", "concentration")

histogramme des données :

ggplot(mesanges.long)+  
 aes(x = concentration,fill = traitement)+  
 geom\_histogram(bins=9,color = 'black')+  
 labs(x="Concentration en oestradiol (pg/ml)",y="Effectif") +   
 scale\_fill\_viridis\_d()+  
 theme\_classic()



visualisation des liens entre groupes

png(  
 filename = "outputs/A1F3.png",  
 width = 20,  
 height = 20,  
 units = "cm",  
 res = 300,  
 bg = 'white'  
)  
plot(  
 c(rep(0.5, 16), rep(1.5, 16)),  
 c(mesanges[, 2], mesanges[, 3]),  
 xlim = c(0, 2),  
 xaxt = "n",  
 bty = "l",  
 xlab = "échantillon",  
 ylab = "concentration en oestradiol (pg/ml)",  
 pch = 21,  
 bg = "black",  
 cex.lab = 1.5  
)  
axis(  
 side = 1,  
 at = c(0.5, 1.5),  
 labels = c("T=0 min", "T=30min")  
)  
segments(  
 x0 = rep(0.5, 16),  
 x1 = rep(1.5, 16),  
 y0 = mesanges[, 2],  
 y1 = mesanges[, 3]  
)  
dev.off()

> png   
> 2

La différence entre les deux groupes :

mesanges$delta <- apply(mesanges[, -1], FUN = "diff", MARGIN = 1)  
mesanges

> individu T0 T30 delta  
> 1 1 57.56761 108.74910 51.181490  
> 2 2 48.59002 30.72740 -17.862617  
> 3 3 59.04845 186.49313 127.444686  
> 4 4 28.78380 51.82936 23.045553  
> 5 5 85.05568 156.50615 71.450470  
> 6 6 99.77151 134.10846 34.336949  
> 7 7 40.63051 86.62907 45.998554  
> 8 8 64.32393 134.47867 70.154736  
> 9 9 55.90166 180.93999 125.038322  
> 10 10 59.04845 93.66305 34.614606  
> 11 11 114.57990 146.41793 31.838033  
> 12 12 56.91974 54.51338 -2.406363  
> 13 13 35.44758 42.01880 6.571222  
> 14 14 39.61244 26.37744 -13.234996  
> 15 15 70.52495 87.36949 16.844541  
> 16 16 51.36659 61.73247 10.365871

On range les valeurs

mesanges <- mesanges[order(abs(mesanges$delta)), ]  
mesanges$rang <- rank(abs(mesanges$delta))  
mesanges

> individu T0 T30 delta rang  
> 12 12 56.91974 54.51338 -2.406363 1  
> 13 13 35.44758 42.01880 6.571222 2  
> 16 16 51.36659 61.73247 10.365871 3  
> 14 14 39.61244 26.37744 -13.234996 4  
> 15 15 70.52495 87.36949 16.844541 5  
> 2 2 48.59002 30.72740 -17.862617 6  
> 4 4 28.78380 51.82936 23.045553 7  
> 11 11 114.57990 146.41793 31.838033 8  
> 6 6 99.77151 134.10846 34.336949 9  
> 10 10 59.04845 93.66305 34.614606 10  
> 7 7 40.63051 86.62907 45.998554 11  
> 1 1 57.56761 108.74910 51.181490 12  
> 8 8 64.32393 134.47867 70.154736 13  
> 5 5 85.05568 156.50615 71.450470 14  
> 9 9 55.90166 180.93999 125.038322 15  
> 3 3 59.04845 186.49313 127.444686 16

sommes des rangs

sum(mesanges[which(mesanges$delta > 0), "rang"])

> [1] 125

sum(mesanges[which(mesanges$delta < 0), "rang"])

> [1] 11

test de Wilcoxon

wilcox.test(mesanges[,2],mesanges[,3],paired=T)

>   
> Wilcoxon signed rank exact test  
>   
> data: mesanges[, 2] and mesanges[, 3]  
> V = 11, p-value = 0.001678  
> alternative hypothesis: true location shift is not equal to 0

test de Mann Whitney

wilcox.test(échantillon 1,échantillon 2,paired=F)

# 3- Comparer des fréquences

données

proc.migr <- read.csv2("donnees/procellariiformes.csv", dec = ".")  
  
proc.migr$species <- factor(proc.migr$species)  
proc.migr$Migratory.mode <- factor(proc.migr$Migratory.mode)  
proc.migr$Nest.type <- factor(proc.migr$Nest.type)  
  
head(proc.migr)

> species Migratory.mode Nest.type  
> 1 Nesofregetta fuliginosa nonmigrant cavity  
> 2 Fregetta tropica nonmigrant cavity  
> 3 Fregetta grallaria nonmigrant cavity  
> 4 Oceanites oceanicus nonmigrant cavity  
> 5 Pelagodroma marina migrant cavity  
> 6 Garrodia nereis nonmigrant cavity

dim(proc.migr)

> [1] 68 3

summary(proc.migr)

> species Migratory.mode Nest.type   
> Bulweria bulwerii : 1 migrant :30 cavity:46   
> Calonectris diomedea : 1 nonmigrant:38 open :22   
> Calonectris leucomelas: 1   
> Daption capense : 1   
> Diomedea bulleri : 1   
> Diomedea cauta : 1   
> (Other) :62

table de contingence en effectifs

table(proc.migr[,-1])

> Nest.type  
> Migratory.mode cavity open  
> migrant 12 18  
> nonmigrant 34 4

table de contingence en fréquences:

table(proc.migr[,-1])/68

> Nest.type  
> Migratory.mode cavity open  
> migrant 0.17647059 0.26470588  
> nonmigrant 0.50000000 0.05882353

table de contingence théorique sous l’attendu de répartition aléatoire :

chisq.test(proc.migr$Migratory.mode,proc.migr$Nest.type)$expected

> proc.migr$Nest.type  
> proc.migr$Migratory.mode cavity open  
> migrant 20.29412 9.705882  
> nonmigrant 25.70588 12.294118

loi de chi² à 1 degré de liberté

png(  
 filename = "outputs/A1F5.png",  
 width = 20,  
 height = 20,  
 units = "cm",  
 res = 300,  
 bg = 'white'  
)  
  
chisq1 <- rchisq(100, 1)  
curve(  
 dchisq(x, df = 1),  
 from = 0,  
 to = 20,  
 main = "",  
 bty = "n",  
 xlab = "Chi² à 1 degré de liberté",  
 ylab = "Densité"  
)  
  
abline(  
 v = 18.7,  
 col = "steelblue",  
 lty = "dashed",  
 lwd = 2  
)  
dchisq(18.77, 1)

> [1] 7.732576e-06

dev.off()

> png   
> 2

test du chi²

chisq.test(proc.migr$Migratory.mode,proc.migr$Nest.type)

>   
> Pearson's Chi-squared test with Yates' continuity correction  
>   
> data: proc.migr$Migratory.mode and proc.migr$Nest.type  
> X-squared = 16.557, df = 1, p-value = 4.722e-05

### Le cas des échantillonnages déséquilibrés

passereaux <- read.csv2("donnees/traits\_passereaux.csv", dec = ".")  
  
passereaux$species <- factor(passereaux$species)  
passereaux$Migratory.mode <- factor(passereaux$Migratory.mode)  
passereaux$Social.system <- factor(passereaux$Social.system)  
  
head(passereaux)

> species Migratory.mode Social.system  
> 1 Attila spadiceus nonmigrant Pair  
> 2 Pitangus sulphuratus migrant Pair  
> 3 Elaenia chiriquensis migrant Pair  
> 4 Elaenia flavogaster migrant Pair  
> 5 Myiozetetes cayanensis nonmigrant Cooperative  
> 6 Sayornis nigricans nonmigrant Pair

summary(passereaux)

> species Migratory.mode Social.system  
> Acanthisitta chloris : 1 migrant :218 Cooperative: 31   
> Acridotheres tristis : 1 nonmigrant:138 Female only: 12   
> Acrocephalus arundinaceus : 1 Group : 1   
> Acrocephalus palustris : 1 Occasional : 32   
> Acrocephalus schoenobaenus: 1 Pair :269   
> Acrocephalus scirpaceus : 1 Unknown : 11   
> (Other) :350

table de contingence : on compare les modes migratoires et la socialité

table(passereaux[,c("Migratory.mode","Social.system")])

> Social.system  
> Migratory.mode Cooperative Female only Group Occasional Pair Unknown  
> migrant 11 5 0 26 171 5  
> nonmigrant 20 7 1 6 98 6

on essaie un chi2

chisq.test(passereaux$Migratory.mode,passereaux$Social.system)

> Warning in chisq.test(passereaux$Migratory.mode, passereaux$Social.system):  
> L’approximation du Chi-2 est peut-être incorrecte

>   
> Pearson's Chi-squared test  
>   
> data: passereaux$Migratory.mode and passereaux$Social.system  
> X-squared = 19.347, df = 5, p-value = 0.001656

l’avertissement vient du fait que les hypothèses du chi² ne sont pas respectée : plusieurs effectifs attendus sont très faibles

chisq.test(passereaux$Migratory.mode,passereaux$Social.system)$expected

> Warning in chisq.test(passereaux$Migratory.mode, passereaux$Social.system):  
> L’approximation du Chi-2 est peut-être incorrecte

> passereaux$Social.system  
> passereaux$Migratory.mode Cooperative Female only Group Occasional Pair  
> migrant 18.98315 7.348315 0.6123596 19.59551 164.7247  
> nonmigrant 12.01685 4.651685 0.3876404 12.40449 104.2753  
> passereaux$Social.system  
> passereaux$Migratory.mode Unknown  
> migrant 6.735955  
> nonmigrant 4.264045

on passe par un test exact de Fisher

fisher.test(passereaux$Migratory.mode,passereaux$Social.system)

>   
> Fisher's Exact Test for Count Data  
>   
> data: passereaux$Migratory.mode and passereaux$Social.system  
> p-value = 0.0008366  
> alternative hypothesis: two.sided