**Exercice ROADKILLS**

Cette étude vise à tester plusieurs variables paysagères sur la richesse spécifique en amphibiens tués sur la route (S.RICH). Le nombre total d’amphibiens (TOT.N) et le nombre de crapauds calamites (BufoCalamita) tués sur la route sont aussi estimés. Les inventaires ont été réalisés dans 3 secteurs différents (Sector). Réaliser la procédure pour tester l’effet de la distance d’une zone humide la plus proche (D.WAT.RES) et la distance à la limite du parc national (D.PARK) sur (i) la richesse spécifique (S.RICH) et (ii) le nombre total d’amphibiens tués (TOT.N).

Objectif : travailler sur des données de comptage

* 1 richesse spécifique qui est du nombre d’espèces donc au niveau individuel
* Et les deux suivantes sont liés au comptage

On nous demande si les deux distances ont un impact sur la richesse spécifique et le nombre total d’amphibiens tués.

Ça parle de « secteur » donc attention à cela.

* Regarder la distribution de la richesse spécifique
* Puis la distribution du nombre d’individus

On veut regarder la distribution.

D’abord je mets les secteurs en facteurs. Puis je regarde la distribution.

On insère la table

tabRK <- read.table("C:\\Users\\lizro\\Desktop\\ETUDES\\5A\\COURS\\analyses de données\\TP3\\RoadKills.txt", header = TRUE,dec=".")

head(tabRK)

names(tabRK)

summary(tabRK)

je mets les secteurs en facteurs

tabRK$fSector<-factor(tabRK$Sector)

#salle que je créé <- facteur de celle que j'ai de base

head(tabRK)

#permet de voir que j'ai bien créé la colonne

#je teste la distribution de la richesse

hist(tabRK$S.RICH)

#la variance de la distribution paraît normale

#je vérifie plot et boxplot

plot(S.RICH~D.WAT.RES, data = tabRK)

plot(S.RICH~D.PARK, data = tabRK)

boxplot(S.RICH~fSector,data=tabRK)

#je souhaite tester les variables plan d'eau et parcs mais je veux controler le secteur

#je ne veux pas que le secteur biaise mes résultats avec le plan d'eau ou le park

#quand y'a une influence avec le secteur, la pente change !

#je regarde le lien entre la richesse et les deux variables (plan d'eau et parc)

m1<-lm(S.RICH~D.WAT.RES\*D.PARK, data=tabRK)

anova(m1)

summary(m1)

#*D.WAT.RES:D.PARK 1.567e-08 6.303e-08 0.249 0.80477* il n'y a pas d'étoile alors ce n'est pas significatif

#je simplifie le modèle, au lieu de mettre \*,je mets +

m2<-lm(S.RICH~D.WAT.RES+D.PARK, data=tabRK)

anova(m2)

summary(m2)

*#D.WAT.RES 7.233e-05 5.756e-04 0.126 0.901*

*#D.PARK -1.455e-04 3.273e-05 -4.444 5.04e-05 \*\*\**

#il y a des étoiles seulement pour le park alors le parc est significatif

#je vérifie les deux variables indépendamment de l'autre (plan d'eau d'un côté et parc de l'autre)

m3<-lm(S.RICH~D.WAT.RES, data=tabRK)

anova(m3)

summary(m3)

*#D.WAT.RES -0.0013566 0.0005599 -2.423 0.0191* \* le plan d'eau est significatif indépendamment du park

m4<-lm(S.RICH~D.PARK, data=tabRK)

anova(m4)

summary(m4)

*#D.PARK -1.432e-04 2.688e-05 -5.326 2.39e-06* \*\*\* le parc est significatif indépendamment du plan d'eau

#pour donner les résultats il faut donner :

# *F-statistic: 28.37 on 1 and 50 DF, p-value: 2.389e-06* (la dernière ligne du summary)

#plus f-statistic est bas, mieux c'est

# degré de liberté = DF

#p-value : plus c'est bas, mieux c'est et quand <0,05, on peut accepter l'hypothèse qu'il y a une différence

#Hypothèse 0, il n'y a pas d'impact et hypothèse 1, il y a un impact

#avec une p-value <0,05, on rejette l'hypothèse 0

#en fonction du secteur on choisi si on fait à effet fixe ou à effet mixte

#on commence avec le modèle effet mixte

library(nlme)

m5<-lme(S.RICH~D.WAT.RES\*D.PARK,random=~1|fSector,data=tabRK)

anova(m5)

summary(m5)

#si rien n'est rejeté alors ça a fonctionné (là ça l'est)

#si c'est rejeté, il faut vérifier toutes les variables, et mettre en facteur

#je regarde l'interaction du modèle

#*D.WAT.RES:D.PARK 1 46 0.0618 0.8048* c'est pas significatif

#je refais le même modèle sans l'interraction

m6<-lme(S.RICH~D.WAT.RES+D.PARK,random=~1|fSector,data=tabRK)

anova(m6)

summary(m6)

*# Value Std.Error DF t-value p-value*

*#(Intercept) 6.486248 0.4328291 47 14.985703 0.0000*

*#D.WAT.RES 0.000072 0.0005756 47 0.125650 0.9005*

*#D.PARK -0.000145 0.0000327 47 -4.444404 0.0001*

##on voit pour D.WAT.RES que la p-value est énorme et que la "value" est plus petite que l'erreur => très bof

*#StdDev: 7.257618e-05 1.420767*

#voici la variance de l'intercept de l'effet secteur

#et ici elle est très faible

#pour le moment le modèle n'est pas stabilisé, la procédure n'est pas qualibrée. On ne l'abandonne pas mais on essaye autre chose

#on fait le modèle à effet fixe

m7<-lm(S.RICH~D.WAT.RES\*D.PARK\*fSector,data=tabRK)

anova(m7)

summary(m7)

#dans le tableau, je regarde seulement la dernière ligne avec l'interaction triple.

#il n'y a aucun effet significatif sur cette interaction alors je la sors du modèle

#je fais un modèle sans l'interaction triple

#je prends chaque variable seule et j'ajoute leur interaction par deux avec chacune d'elles

m8<-lm(S.RICH~D.WAT.RES+D.PARK+fSector+D.WAT.RES\*D.PARK+D.WAT.RES\*fSector+D.PARK\*fSector,data=tabRK)

anova(m8)

*#D.WAT.RES:D.PARK 1 0.297 0.297 0.2532 0.6174937*

*#D.WAT.RES:fSector 2 8.271 4.135 3.5278 0.0383498 \**

*# D.PARK:fSector 2 38.690 19.345 16.5032 5.142e-06 \*\*\**

#on peut enlever celle où ce n'est pas significatif et refaire le modèle

m9<-lm(S.RICH~D.WAT.RES+D.PARK+fSector+D.WAT.RES\*fSector+D.PARK\*fSector,data=tabRK)

anova(m9)

#on regarde le résultat pour les interactions doubles

#ces deux variables (eau et park) ne peuvent donc pas être traitées sans être en lien avec le secteur (fSector)

#comme secteur intervient sur les 2, on ne peut pas s'en sortir avec un modèle à effet mixte

#je pars donc sur le modèle à effet fixe

summary(m9)

#le secteur a n'apparait pas et est pris comme référence

#le secteur a est une surface, avec distance plan d'eau et distance au park

#les surfaces a,b,c doivent se couper car elles n'ont pas la même pente

#voir le cours pour comprendre tout le summary !!

#maintenant on fait le même procédé pour le nbre total d'individus = variable type comptage

#je teste la distribution de la richesse

hist(tabRK$TOT.N)

#la distribution n'est pas normale

#je regarde les plot

plot(TOT.N~D.WAT.RES,data=tabRK)

plot(TOT.N~D.PARK,data=tabRK)

#la procédure que l'on va utiliser est hyper classique ##masterécologie

#sur une échelle de temps courte, on obtient beaucoup de fois un nombre faible d'individus plutot qu'un nombre élevé d'individus

#c'est comme ça, ce sont des stats

#ici, l'acrt-type ne nous donne rien car la distribution est nulle

#theta va être l'élément qui permettra d'observer la dispersion des points autour de la moyenne = paramêtre de dispersion des points

#plus le paramêtre de dispersion est faible, plus la variance est forte et inversement

#modèle simple entre D?PARK et TOT?N sur une distribution binomiale négative

#il faut un package

library(MASS)

#quand je fais un modèle linéaire généralisé, là je prédis une distribution paramêtré différemment d'une distribution normale

#ce paramêtre je dois en parler

m10<-glm.nb(TOT.N~D.PARK,link="log", data=tabRK)

#binomiale négative = fonction de variance

#log = lien

#dès qu'on a eu une binomiale négative, on part sur le lien log

anova(m10)

*# Df Deviance Resid. Df Resid. Dev Pr(>Chi)*

*#NULL 51 155.445*

*#D.PARK 1 100.7 50 54.742 < 2.2e-16 \*\*\**

#on voit que le chi2 est significatif et donc que la distance du park par rapport à la base (aka NULL) est significative

#la défiance est à 100,7 et donc c'est beaucoup et cela montre l'impact de la distance du park

*#Negative Binomial(3.681) = theta*

#plus theta est fort et plus la variance diminue

summary(m10)

#on écrase plus d'inidvidus quand on est proche du park

###SUITE

#la binomiale négative fonctionne, cela prédit des choses correctes

#on peut donc repartir sur l'interraction triple en sortant l'anova

#interraction triple n'est pas significative, je l'enlève

#pareil pour la suite avec deux interractions doubles

#si le theta est supérieur, la variance est réduite et le modèle est donc meilleur

#il faut savoir tracer les prédis sur un graphe avec les intervalles de confiance

# et binomiale négative avec modèle fixe