**L3 EBO**

**TP3-4 Exploration et analyse des données environnementales**

**Mise en œuvre des tests statistiques**

**Exercice 1 :**

Des relevés d’oiseaux ont été effectués dans 67 sites durant 2 années : en 2001 et en 2003. Vous disposez d’un fichier « birds.txt » contenant les données de richesse spécifique pour chacun des sites.

Vous allez tester si la richesse spécifique a varié de façon significative entre ces 2 années. Pour cela, vous vérifierez les conditions pour faire un test paramétrique. Si les conditions ne sont pas remplies, vous tenterez de modifier les données pour rendre possible un test paramétrique. Dans tous les cas, si c’est possible, vous comparerez les résultats d’un test paramétrique avec un non paramétrique (Donnez les noms des tests que vous utiliserez).

Vous illustrerez vos résultats à l’aide d’un graphique complet.

**Exercice 2 :**

Le domaine vital (ha) de l’élan a été étudié pour 6 individus en hiver et au printemps. Les résultats obtenus sont fournis dans le fichier « elan.csv ».

Ces données confirment-elles l’hypothèse d’une réduction du domaine vital de l’élan durant la période hivernale ? Donnez le nom du test que vous utiliserez.

**Exercice 3 :**

Dans un parc naturel, on étudie une population d’amphibiens (*Bufo bufo*). Pour cela, des individus sont capturés autour de leurs sites de reproduction, sont sexés et pesés. Les résultats sont donnés dans le fichier « toads.txt ».

Existe-t-il une différence significative de poids entre les mâles et les femelles ? Donnez le nom du test que vous utiliserez.

**Exercice 4 :**

Des mesures de concentration en pesticides dans des chairs de poissons carnivores (perches, sandres et brochets) ont été effectuées dans un lac A. Les résultats sont exprimés en mg/g. Six catégories de poissons sont considérées :

- a2 -> a6 sont les concentrations mesurées chez des poissons âgés de 2 (a2) à 6 ans (a6) dans un lac,

- ax correspond à des mesures faites dans un autre lac (lac B), sur des poissons de 2 ans.

Les données sont situées dans le fichier « pesticides.txt ».

La moyenne nationale pour ces espèces est également connue pour des poissons de 2 ans et est de 0.204 +/- 0.025 mg/g.

1. Comparer la concentration moyenne en pesticides des poissons de 2 ans du lac A à la moyenne nationale.

2. Calculer la concentration moyenne en pesticides des poissons de 2 ans du lac A. Calculer ensuite celle des poissons de 2 ans du lac B. Quelle est la moyenne la plus élevée entre les 2 ?

3. Vous allez tester s’il existe une différence significative entre les 2 moyennes calculées à la question 2 en utilisant un test unilatéral.

4. Comparer la concentration moyenne en pesticides des poissons de 4 ans à la moyenne nationale.

5. Y-a-t-il une différence significative entre la concentration en pesticides dans les poissons de 4 ans du lac A et celle des poissons de 2 ans du lac B?

**Exercice 5 :**

Vous disposez d’un jeu de données nommé « mais.txt » caractérisant la récolte de M. Léon en 1999. Différents paramètres ont été considérés (les noms pour qualifier les variables dans le fichier sont donnés entre parenthèses):

- l’identifiant d’un pied de maïs (*Individu*)

- sa hauteur (*Hauteur*)

- sa masse *(Masse*)

- le nombre de grains qu’il porte (*Nb.grains*)

- la masse de grains (*Masse.grains*)

- la couleur des grains (*Couleur*)

- la germination des grains sur l’épi (*Germination.epi*)

- l’importance de l’enracinement (*Enracinement*)

- le fait que le pied soit penché ou tombé (*Verse*)

- le fait qu’il soit endommagé par les insectes (*Attaque*)

- la parcelle dont il est issu (*Parcelle*)

- la hauteur du pied à J7 après récolte (*Hauteur.J7*)

a) Ouvrez le fichier « mais.txt » dans R.

b) Créez un nouvel objet dans R que vous nommerez « mais2 » et qui ne comportera que les colonnes : « Hauteur », « Masse », « Parcelle » et « Hauteur.J7 ».

Vous supprimerez les lignes avec des « NA » en testant 2 méthodes :

- en utilisant la syntaxe matricielle : « [] » ;

- en utilisant la commande suivante : > na.omit(mais2)

Vous vérifierez que le résultat est le même et nommerez l’objet final « mais3 ».

c) Vous allez tracer un graphe (commande « plot() ») représentant la masse des pieds de maïs en fonction de leur hauteur. Caractéristiques de ce graphe :

- un axe des abscisses entre 100 et 400,

- un axe des ordonnées entre 1000 et 3000,

- un titre pour chacun des axes,

- les symboles devront être des cercles, pleins et de couleur bleu.

***Pour vous aider :***

***- les limites des axes se définissent en utilisant les options « xlim=c(limite\_inf, limite\_sup) » et « ylim=c(limite\_inf, limite\_sup) » dans la commande « plot »,***

***- la nature des symboles d’un graphe se définit avec l’option « pch=un nombre entier », chaque type de symbole étant préexistant dans R sous un nombre entier : essayez les options : « pch=0 », « pch=1 », « pch=18 », « pch=19 » et « pch=16 »,***

***- la couleur des symboles d’un graphe se définit avec l’option « col= « couleur en anglais » ».***

d) Monsieur Léon souhaiterait savoir si les moyennes de masse des pieds de maïs sont similaires entre les parcelles Nord et Sud.

e) Monsieur Léon voudrait savoir si la moyenne des hauteurs des pieds de maïs de la parcelle nord est significativement plus petite que celle de la parcelle ouest.

Ensuite, vous allez illustrer votre résultat en suivant les étapes suivantes:

- créer un objet que vous nommerez « dataH » comprenant, dans cet ordre, la hauteur moyenne des pieds de maïs de la parcelle nord et celle de la parcelle ouest.

- créer un objet que vous nommerez « datasdH » comprenant, dans cet ordre, l’écart-type des pieds de maïs de la parcelle nord et celui de la parcelle ouest.

- tapez :

> labx<-c("Parcelle nord", "Parcelle ouest")

> graph<-barplot(dataH, names=labx,space=0.5, ylab="Hauteur en cm",col="blue", border="blue", axes=T, ylim=c(0,350), cex.names=1, las=1)

*Essayez de changer l’option « las » et de la faire varier entre 0 et 3. Que fait cette option ?*

* En laissant votre graphe ouvert, tapez :

> segments(graph,dataH-datasdH, graph, dataH+datasdH, lwd=2)

> arrows(graph, dataH-datasdH, graph, dataH + datasdH, lwd=2, angle=90, code=3, length=0.05)

*Que font ces 2 commandes?*

- Fermez la fenêtre graphique, puis tapez :

> graph<-barplot(dataH, names=labx,space=0.5, ylab="Hauteur en cm",col="green", border="black", density=7, angle=45, axes=T, ylim=c(0,350), cex.names=1, las=1)

- En laissant votre graphe ouvert, tapez :

> segments(graph, dataH-0, graph, dataH+datasdH, lwd=2)

> arrows(graph, dataH-0, graph, dataH + datasdH, lwd=2, angle=90, code=2, length=0.05)

*Quelles sont les differences avec le graphe précédent?*

- vous allez placer la significativité de votre test en respectant les étapes suivantes :

tapez :

> text(locator(1), "a"), *puis placez votre curseur de souris au-dessus du premier barplot là où vous souhaiterez voir apparaître la lettre « a », cliquez une première fois pour fixer le curseur et cliquer une deuxième fois pour afficher la lettre définitivement.*

Tapez :

> text(locator(1), "b"), *puis même chose pour placer la lettre « b ».*

f) Des études contradictoires indiquent que la moyenne de hauteur des pieds de maïs n’est plus la même une semaine après récolte: pour certaines, elle serait supérieure, pour d’autres inférieure. Comparez la hauteur avec la hauteur à J7 après récolte en considérant l’ensemble des parcelles.

**Exercice 6 :**

Afin de tester la toxicité d’une variété de maïs OGM, 3 groupes de 10 rats a été nourri avec ce maïs. Le maïs OGM représentait 11% de la ration alimentaire dans le premier groupe, 22% dans le second, et 33% dans le troisième. Un quatrième groupe témoin de 60 rats a été nourri avec du maïs non-OGM. Après 90 jours, on mesure le poids du foie de chaque rat.

1. Le tableau de données est enregistré dans le fichier « ogm.csv ». Charger ce fichier.

2. Représenter graphiquement la taille du foie en fonction de la quantité d’OGM présent dans l’alimentation des rats.

3. Créer les variables g0, g11, g22 et g33 contenant les lignes du tableau correspondant aux rats nourris avec 0%, 11%, 22% et 33% d’OGM, respectivement.

4. Comparer le poids moyen du foie des rats du groupe sans OGM, avec celui des rats du groupe à 11% d’OGM, puis à 22% et à 33%.

5. Pourquoi l’étude n’a-t-elle pas utilisé un protocole avec appariement des rats, type “avant consommation d’OGM” versus “après consommation d’OGM”?