

TP4 : Tests sur données dans R

Les fichiers `data.txt` et `databis.txt` sont disponibles dans le dossier TP4 à l'adresse suivante <http://perso.math.univ-toulouse.fr/smourare/Enseignement/>

R permet de définir des fonctions à l'aide de la commande `function`. Voici un exemple simple qui calcule la norme euclidienne du vecteur (x, y) :

```
norme=function(x,y){  
z=x^2+y^2  
sqrt(z)  
}
```

Taper `help("function")` pour plus de détails.

Exercice 1 Le fichier `data.txt` contient un ensemble de dix échantillons de taille 100. Chaque échantillon est rangé en colonne. Chaque échantillon a été tiré suivant une loi exponentielle de paramètre θ où $\theta = 1$ ou $\theta = 1.5$ suivant les échantillons.

1. A l'aide de la fonction `read.table` et de la fonction `as.matrix`, créer une fonction `rmean(file)` qui lit le fichier `file`, le transforme en matrice puis retourne la moyenne de chaque colonne de la matrice.
2. Utiliser la fonction `rmean` ainsi créée pour lire le fichier `data.txt` et calculer la moyenne de chaque colonne.
3. Déterminer le test du rapport de vraisemblance de $H_0 : \theta = 1$ contre $H_1 : \theta = 1.5$ de niveau $\alpha = 0.05$ pour $n = 100$.
4. Créer une fonction `test(file)` qui lit les données du fichier `file` et retourne pour chaque colonne le résultat du test.
5. On considère maintenant le fichier `databis.txt` qui contient un ensemble de dix échantillons de taille 100. Chaque échantillon est rangé en colonne. Chaque échantillon a été tiré suivant une loi normale d'espérance μ et de variance 1, où $\mu = 1$ ou $\mu = 1.5$. Déterminer un intervalle de confiance pour μ de niveau $\alpha = 0.05$ pour $n = 100$.
6. Créer une fonction `intervalle(file)` qui lit les données du fichier `file` et retourne pour chaque colonne, 1 si 1.5 appartient à l'intervalle de confiance, et 0 sinon.

Exercice 2

1. Simuler mille échantillons de taille 100 de loi exponentielle de paramètre $\theta = 1.5$ dans une matrice 100 lignes et 1000 colonnes. A l'aide de la fonction `write.table`, enregistrer votre matrice dans un fichier `simul.txt`.
2. A l'aide de la fonction `test` de l'exercice 1, compter le nombre de fois que le test choisit H_1 sur les simulations. En déduire une valeur approchée de la puissance du test.

Exercice 3

1. Simuler mille échantillons de taille 100 de loi normale d'espérance $\mu = 1.5$ et de variance 1, dans une matrice de 100 lignes et 1000 colonnes. À l'aide de la fonction `write.table`, enregistrer la matrice dans un fichier `simul.txt`.
2. À l'aide de la fonction `intervalle` de l'exercice 1, compter le nombre de fois que l'intervalle de confiance contient 1.5 sur les simulations. De quelle valeur théorique ce nombre devrait être proche ?