

# BIOINFORMATIQUE, mars 2016

## Master BTV, Partie "Statistiques"

(à rédiger sur une copie séparée)

Pour répondre aux questions posées, vous utiliserez les données à l'adresse :

`http://forge.info.univ-angers.fr/~gh/Bism/shsp4609.dar`

Le fichier correspondant contient un peu plus de quatre mille six cent protéines considérées comme des "petites" protéines HSP. Vous pouvez charger ces données à l'aide de la fonction `lit.dar()` vue en cours.

1. Combien y a-t il de protéines exactement dans ce *data frame* ?  
Vous indiquerez la commande R qui permet d'obtenir ce nombre.
2. On s'intéresse à la colonne `Foldindex` de ces données. Quelle est *a priori* la plage de valeurs de cette variable ? Vous pouvez consulter la page <http://bip.weizmann.ac.il/fldbin/findex> pour essayer de trouver la réponse.  
Quelle est la valeur du minimum et du maximum de cette variable pour les données ?
3. Essayez ensuite de décrire complètement cette variable `Foldindex` avec des indicateurs statistiques numériques. Vous complétez cette liste d'indicateurs par la rédaction détaillée et soignée d'un unique paragraphe. On ne demande pas de reproduire de graphique.

*(suite au dos)*

4. On voudrait savoir si globalement on peut considérer que la moyenne de **Foldindex** est égale à la valeur  $\mu = 0$  à l'aide du test  $t$  de *Student*, soit la fonction *t.test()*.  
Donnez le détail de la commande R qui permet de tester cette hypothèse.
5. Indiquez clairement la conclusion statistique du test avant de rédiger une conclusion plus "métier" donc facilement compréhensible par tout biologiste.
6. Est-ce que la fonction *wilcox.test()* permet de tester si globalement on peut considérer que la médiane de **Foldindex** est égale à 0 ?  
Si oui, réalisez le test. Sinon, trouvez comment on peut tester cette hypothèse en R et indiquez quelle fonction il faut utiliser.  
Rédigez ensuite soigneusement la conclusion de ce test.

– FIN DU SUJET –