

BIOINFORMATIQUE, août 2012
Master TPV, Partie "Statistiques"
(à rédiger sur une feuille simple séparée)

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

<http://forge.info.univ-angers.fr/~gh/Bism/hsp1952.dar>

Le fichier correspondant contient un peu moins de deux mille protéines considérées comme des protéines **HSP**.

1. Si on devait fournir un indicateur de tendance centrale pour la variable **Length** (longueur des protéines, en nombre d'acides aminés), que choisiriez-vous ? La moyenne ou la médiane ? Et surtout, pourquoi ?
Essayez ensuite de décrire complètement cette variable **Length**. On ne demande pas de reproduire de graphique.
2. Les protéines **HSP** sont nommées ainsi parce qu'elles sont en général des « petites » protéines. Comparer le poids moléculaire (variable **MW**) de ces 1952 protéines à celui des 765 protéines du fichier **lea4.dar** (variable **mw4**). On fournira une conclusion "propre" sur la présence d'une différence significative ou non pour ces poids moléculaires et sur le fait que ces protéines sont ou non plus petites que celles de la **LEADB**.