

BIOINFORMATIQUE, Unité libre 2010  
Master TPV, Partie "Statistiques"  
(à rédiger sur une feuille simple séparée)

Le fichier `somelea.dar`, disponible sur Internet à l'adresse :

<http://forge.info.univ-angers.fr/~gh/Datasets/somelea.dar>

contient un extrait de la base de données LEAPDB. Les protéines sont fournies en ligne : la colonne 1 du fichier est l'identifiant de la protéine, la colonne 2 est sa longueur et la colonne 3 est une indication de groupe.

Etudiez la longueur des protéines globalement puis pour chaque groupe. Peut-on considérer que les deux groupes sont équivalents en terme de longueur de protéines ?

On séparera soigneusement la partie *présentation (succincte) des résultats numériques* de la partie *rédaction en direction d'un public de biologistes*.

Extrait des données :

identi	longueur	groupe
I00051	86	2
I00107	112	2
I00173	344	1
I00260	256	1
I00318	314	1
I00319	326	1
I00365	82	2
...		
I00685	277	1
I00697	112	2
I00698	83	2
I00699	112	2
I00700	93	2
I00702	89	2