

BIOINFORMATIQUE, avril 2011
Master TPV, Partie "Statistiques"
(à rédiger sur une feuille simple séparée)

Vos enseignants de bioinformatique ont étudié la longueur des protéines dans la base de données nommée LEAPDB, globalement et par classe. Vous trouverez sur Internet à l'adresse :

<http://forge.info.univ-angers.fr/~gh/Bism/examen2011.php>

une partie des résultats de cette analyse.

1. En admettant que les données sont dans un tableau nommé LEA, qu'il y a dans ce tableau une colonne nommée cl (pour classe) et une colonne nommée lon (pour longueur), quelle instruction R, éventuellement issue de `statgh.r` fournit le tableau de résultats numériques de la page ?
2. Que peut-on déduire de ce tableau de résultats ?

On fournira une conclusion pour statisticien[ne]s et une conclusion plus lisible pour biologistes.

3. Les graphiques proposés sur la page sont des boxplots et des beanplots.

Fournissent-ils graphiquement la même information que le tableau des résultats numériques ?

En quoi ces deux types de graphiques sont complémentaires ?