

Bioinformatique, partie Statistiques (L3) TD1 : Etude des longueurs de chaines polypeptidiques

Présentation des données

La base de données DBDB disponible à l'adresse

<http://www.info.univ-angers.fr/pub/richer/rec/bio/dbdb/>

contient de nombreuses protéines avec des ponts disulfure. On s'intéresse ici à l'ensemble de ces protéines sous l'angle de "chaines polypeptidiques". Le fichier `chp.fasta` contient ces chaines au format *Fasta*, le fichier `chp.lng` contient les longueurs de ces chaines et le fichier `chp.cnt` dénombre de plus les acides aminés par chaine.

Les fichiers `chpi.*` avec $i=1, 2$ ou 4 contiennent les mêmes informations pour des sous-populations choisies de la DBDB :

- $i = 1$: protéines avec ponts inter et intra ;
- $i = 2$: protéines avec ponts inter sans pont intra ;
- $i = 4$: toxines avec ponts intra sans pont inter.

Série de Questions 1

Combien y a-t-il de chaines polypeptidiques dans le fichier `chp.fasta` ?

Quelle est la nature de la variable statistique LNG ?

Quels calculs statistiques descriptifs faut-il effectuer globalement sur l'ensemble de la population pour cette variable LNG ? et par sous-population ?

Les effectuer avec *Rstat* puis commenter les résultats sans oublier de réaliser les graphiques correspondant. On discutera de l'automatisation des calculs. Pourquoi ne faut-il pas utiliser *Excel* ou *OpenOffice/Calc* ?

Série de Questions 2

Quelle est la nature des 20 variables statistique CntA, CntC... ?

Quels calculs statistiques descriptifs faut-il effectuer globalement sur l'ensemble de la population pour ces variables ? et par sous-population ?

Les effectuer avec *Excel* et *Rstat* puis commenter les résultats sans oublier de réaliser les graphiques correspondant. On discutera de l'automatisation des calculs.

Rappel : Les fichiers de données sont disponibles à l'adresse

`http://www.info.univ-angers.fr/pub/gh/Bis/bis.htm`

Pour le logiciel R, on charge l'archive des programmes par

```
source("http://www.info.univ-angers.fr/pub/gh/wstat/statgh.r")
```

el la lecture des données (fichier *.chp) peut se faire par

```
chp <- lit.dar("http://www.info.univ-angers.fr/pub/gh/Bis/chp.lng")
names(chp)
attach(chp)
lng <- LNG
```