Probabilités et statistiques pour la biologie   
(STAT1, ENSBBAU16L) – Examen – 22 janvier 2018

Calculatrices Autorisées ; Documents Non Autorisés.

**Pondération :** cet examen compte pour 100% de la note finale.

## Question 1 (4 points)

On fait une recherche de similarités avec l’algorithme BLAST, en comparant une séquence protéique inconnue aux 25000 séquences d’un protéome de référence. On décide arbitrairement de retenir les 5 meilleures correspondances, et d’interpréter le résultat en tenant compte de leur ordre. Combien de possibilités y a-t-il ? Expliquez le raisonnement, indiquez le nom de la fonction et sa formule (avec les symboles), puis remplacez dans cette formule chacun des symboles par les valeurs numériques appropriées de l'énoncé. Il n'est pas nécessaire de fournir le résultat numérique final, dont le calcul nécessiterait un ordinateur.

## Question 2 (6 points)

On scanne la séquence d’un génome bactérien de 4 Mb pour y trouver toutes les occurrences du motif TACGATGC, en acceptant au plus 2 substitution. On considère que les nucléotides sont répartis de façon indépendante, et qu’ils sont équiprobables dans ce génome. Comment calcule-t-on la probabilité d’observer une occurrence du motif à une position donnée du génome ?

1. Expliquez le raisonnement qui permet de déterminer la distribution théorique de probabilité à utiliser.
2. Ecrivez la formule de cette distribution théorique en expliquant (dans le contexte général) ce que représente chaque symbole.
3. Ecrivez cette même formule en remplaçant les symboles par les valeurs numériques extraites de l’énoncé. Vous ne devez pas calculer le résultat final.
4. Discutez de la pertinence des hypothèses de travail.

## Question 3 (10 points)

On veut évaluer l’impact d’un médicament sur la concentration d’une protéine dans une lignée cellulaire, en mesurant la fluorescence d’un marqueur associé. On suppose que les fluctuations expérimentales suivent une loi normale. Les mesures sont indiquées ci-dessous.

Cellules traitées : 92 96 102 86 118 110 94 102

Cellules non-traitées : 125 86 76 101 95 136 108 113

1. Calculez les moyennes, écarts-types et les médianes des deux échantillons.
2. Pour chaque échantillon, estimez la moyenne et l’écart-type de la population.
3. Quel test choisiriez-vous pour évaluer si le traitement a un effet ou non ? Justifiez vos choix en vous basant sur l’énoncé, et sur les résultats de la sous-question b.
4. Ecrivez la formule de l'hypothèse nulle et décrivez-la en une phrase.
5. Ecrivez la formule de la statistique du test choisi.
6. Réécrivez cette formule en remplaçant chaque symbole par la valeur numérique correspondante.
7. Calculez la valeur de la statistique de test.
8. Calculez le nombre de degrés de liberté (*d.l.*).
9. Indiquez la p-valeur
10. Expliquez en une phrase ou deux comment on interprète cette p-valeur.
11. Quelle décision prenez-vous à l’issue du test ? Justifiez.