

Module 3 - Analyse statistique avec R - Séance 2

DUBii 2019

Travail maison

2019-02-07

Contents

Consignes	1
Exercice 1 : Comparaison de l'âge moyen des patients malades et sains	1
Exercice 2 : ACP sur le jeu de données <code>microbiota.abundance</code>	1

Consignes

Le travail maison qui vous est demandé porte sur les données étudiées en TP de la séance 2. Ce travail comporte deux exercices. On vous demande de rédiger un compte rendu pour chaque exercice où vous devez répondre aux questions, commenter et interpréter les résultats. Le rapport de chaque exercice doit faire au maximum une page.

Pensez à consulter les diapos de cours pour vous aider avec les commandes R.

Exercice 1 : Comparaison de l'âge moyen des patients malades et sains

On cherche à savoir si l'âge des patients est significativement différent de celui des contrôles sains ? Pour cela, répondez aux questions suivantes :

1. Identifiez la variable aléatoire étudiée.
2. Identifiez le test que vous allez réaliser.
3. Représentez, sur le même graphique, la distribution de la variable aléatoire chez les individus sains et les patients malades en utilisant des boîtes à moustaches (fonction `boxplot()`).
4. Calculez la moyenne et l'écart-type de la variable chez les individus sains et malades (fonctions `by()`, `mean()` et `sd()`).
5. Calculez l'intervalle de confiance à 95% de l'espérance de la variable aléatoire pour les individus sains et les individus malades (fonctions `qnorm()` et `var()`).
6. Effectuez le test statistique pour répondre à la question. Posez les hypothèses et concluez au test. (fonctions `t.test()`)

Exercice 2 : ACP sur le jeu de données `microbiota.abundance`

Réalisez une ACP centrée réduite sur ce jeu de données `microbiota.abundance`.

- a. Quel pourcentage de la variance expliquée est représentée dans le premier plan factoriel ?
- b. Proposez une représentation graphique des individus avec les individus malades d'une couleur et les individus sains d'une autre. La variable renseignant sur le statut des individus (*healthy*, *liver*) se trouve dans le jeu `metadata`. Dans les deux fichiers de départ, les individus sont rangés dans le même ordre.
- c. Proposez une représentation graphique des individus avec les enterotypes de trois couleurs différentes. Que constatez-vous ?