
EXERCICE DE STATISTIQUE AVEC LE LOGICIEL R

Analyse exploratoire de données transcriptomiques

sur les régulations Gène-Métabolite dans le fruit de tomate

Les données proviennent d'une étude sur le développement du fruit de tomate (F. Mounet et al. 2009, UMR 1332 Biologie du fruit et Pathologie, Centre INRA Bordeaux). L'étude est réalisée sur deux tissus du fruit, le mésocarpe (partie charnue) et le gel (locular) à 4 stades de développement (12, 20, 35, 45 day post anthesis). La figure 6 donne le plan de l'expérience microarray. On réalisera l'analyse à partir du jeu de données fourni par les auteurs et du tableau S1 de données supplémentaires de l'article (pp.108.33967_Mounet_Supplemental_Table_S1.xls).

QUESTION 1 : Présentez dans un tableau toutes les variables prises en compte dans cette étude et leurs caractéristiques.

QUESTION 2 : Décrivez la population, l'échantillon et l'individu statistique de cette étude.

QUESTION 3 : Ouvrir le fichier de données dans un tableur, l'enregistrer dans un format compatible avec R et charger ce tableau dans le logiciel R. Vérifiez le nombre de lignes et de colonnes du tableau chargé dans R et le caractère numérique des variables quantitatives.

QUESTION 4 : Calculer les principaux paramètres de position de l'ensemble des variables quantitatives : Expression par stade/tissu, puis expression de chaque gène. Utiliser l'aide de R pour transposer le tableau de données (lignes<-> colonnes).

QUESTION 5 : On souhaite comparer l'expression des gènes entre les deux tissus au stade **12 dpa**. Déterminez la liste des gènes qui ont une **expression supérieure à 15**.

QUESTION 6 : Représenter la distribution des deux premiers gènes sélectionnés à la question 5 (gène 129 et 4485). Donner vos conclusions.

QUESTION 7 : Que faut-il vérifier avant de mettre en œuvre un test statistique ? Comment le réaliser avec R ?

QUESTION 8 : Réalisez une étude de corrélation entre les gènes sélectionnés à la question 5 (gène 129 et 4485). Présenter en détail votre démarche statistique et les conclusions de votre analyse.

QUESTION 9 : On pose l'hypothèse que les deux gènes sélectionnés à la question 5 (gène 129 et 4485) s'expriment de la même manière entre les deux tissus au stade 12 dpa. Réalisez cette comparaison en présentant en détail votre démarche statistique et les conclusions de votre analyse.