

M1 MABS

Contrôle continu de Traitement des données biologiques – octobre 2012

Instructions

Commencez par enregistrer ce fichier sur le bureau en lui donnant un nom de la forme
NOM_Prenom_Groupe_CC_TDB_oct_2012.doc ;

Exemple si je m'appelle Jules-Édouard Moustic et que je suis dans le groupe 2, mon fichier aura le nom MOUSTIC_Jules-Edouard_Groupe2_CC_TDB_oct_2012.doc

Tout manquement à cette règle sera sanctionné à hauteur de 2 points sur la note finale.

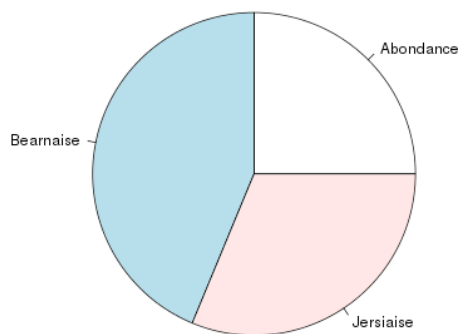
Pour chaque question, s'il y a lieu :

1. donnez la ou les commandes permettant d'y répondre
2. copiez/collez la réponse obtenue dans R (marche aussi bien pour les graphiques que le texte)
3. ajoutez un commentaire ou une explication, et soyez concis.

Exemple :

Question 0 : Réalisez un camembert des mesures effectuées sur les races de vaches.

```
# Question 0  
pie(summary(vaches))
```



```
# Dans un premier temps, la fonction summary appliquée à vaches  
# permet d'obtenir les effectifs par race de vaches  
# ensuite, ces effectifs sont passés à la fonction pie qui trace un  
# camembert
```

Envoyez 5 minutes avant la fin de l'examen votre fichier de réponse aux adresses :

barriot@biotoul.fr

bonhomme@lrsv.ups-tlse.fr

Conseil :

- mettez-vous en copie du mail afin de garder une trace de votre travail
- laissez le fichier sur le « bureau » de l'ordinateur ET laissez l'ordinateur allumé

Vous êtes responsables de l'envoi du fichier en fin d'examen. Etant donné qu'il est facile d'antidater un fichier ainsi que son envoi, tout fichier non reçu en fin de séance recevra une note de 0/20.

Problème 1

Un ingénieur agronome a mesuré la quantité de sucre (en mg par grammes de poids sec de fruit) contenu dans deux variétés bien distinctes de pommes, la Rajka et la Rubinola. Voici les valeurs :

Mesures sur Rajka :

10.94, 5.66, 10.88, 10.27, 8.40, 9.55, 14.17, 10.18, 9.36, 9.20, 12.44, 11.23, 10.74, 10.88, 9.74, 10.99, 10.67, 9.09, 9.75, 7.77, 4.24, 6.55, 10.91, 9.61, 11.84

Mesures sur Rubinola :

20.76, 21.58, 19.51, 20.87, 19.74, 18.11, 22.29, 17.38, 21.70, 16.88, 18.83, 21.90, 21.27, 21.63, 18.80, 19.30, 18.44, 20.51, 23.34, 17.21, 17.37, 22.25, 21.30, 20.92, 19.39

Question 1 : Donnez le nombre de variables et leur nature (quantitative, qualitative). Combien y a-t-il d'observations ?

Question 2 : Donnez des représentations graphiques de ces données (et incluez-les dans ce fichier) : histogrammes (pour lequel vous choisirez un nombre de classes vous semblant le plus approprié), courbes de densité (affichez notamment les courbes de densité sur le même graphique, mais avec des couleurs différentes), et boîtes à moustaches.

Question 3 : Calculez la moyenne, la médiane et la variance de ces données. Commentez ces valeurs.

Question 4 : L'ingénieur se demande s'il y a une différence de concentration en sucre entre les pommes Rajka et Rubinola. Quel test statistique utiliseriez-vous pour tester cette différence ? Suppose-t-il de vérifier certaines hypothèses préalablement ? Effectuez les tests préalables si nécessaire puis effectuez le test de comparaison de ces 2 variétés. Donnez pour chaque test l'hypothèse nulle H_0 . Que concluez-vous ?

Problème 2

Une microbiologiste spécialisée en qualité et sécurité des aliments s'intéresse à l'abondance de microorganismes contenus dans un fromage affiné. Pour ce fromage, les bactéries du genre *Lactobacillus* contribuent à l'acidification du lait et les levures du genre *Penicillium* contribuent à la formation d'arômes. Selon les normes de distribution et afin d'obtenir un équilibre sanitaire et gustatifs, un prélèvement standard de fromage doit contenir des proportions relatives de 2/3 de *Lactobacillus* pour 1/3 de *Penicillium*.

Sur un prélèvement standard représentatif d'un lot de fromage prêt à être distribué, la microbiologiste a estimé par un test biochimique des abondances égales à 656 *Lactobacillus* et 334 *Penicillium*.

Question 5 : Quel test statistique utiliseriez-vous pour étudier la conformité de ces abondances vis-à-vis des proportions de *Lactobacillus* et *Penicillium* attendues ? Effectuez le test, et concluez. Ce lot de fromage peut-il être distribué ou non dans les chaînes de supermarché ?

Problème 3

Une entreprise s'intéresse à la production par voie biologique de gélatine pour la fabrication de bonbons. Les algues vertes sont de potentiels producteurs de gélatine naturelle. Afin d'améliorer le rendement, des ingénieurs ont étudié la production de gélatine par des algues vertes ayant une disponibilité variable en phosphate. Ils ont fait ces mesures sur 20 individus (fichier : « algues.txt »)

Question 6 – Combien de variables ? Quelle est la nature de cette (ces) variable(s) : quantitative(s)? qualitative(s)?

Question 7 – Faire une représentation graphique de ces données (copier l'image dans ce document).

Question 8 – Calculez le ou les paramètres permettant de mesurer la relation entre ces variables. Existe-t-il une relation entre ces deux variables? Expliquez ces paramètres et les valeurs calculées.

Question 9 – Quelle est l'équation de la droite de régression (modèle linéaire)? La valeur du coefficient de détermination (R^2) ? Ajouter la droite de régression au nuage de points.

Question 10 – Que pouvez-vous conclure de la production de gélatine par les algues en fonction de la disponibilité en phosphate? Concluez sur la stratégie à adopter par l'entreprise pour améliorer le rendement de la production de gélatine par les algues.

Problème 4

On a mesuré la robustesse mécanique du bois d'un arbre dioïque (individus mâles et femelles séparés) produisant de gros fruits. Des mâles et des femelles on été étudiés pour deux génotypes (« A » et « B »). Fichier « wood.txt ».

Question 11 – Combien de variables ? Quelle est la nature de ces variables ? Combien de « catégories » d'arbres mesurés y-a-t-il donc dans ce jeu de données ?

Question 12 – Faire une représentation graphique de ces données (copier l'image dans ce document).

Question 13 – Quel type d'analyse va permettre de tester l'effet de différents « facteurs » sur la robustesse mécanique du bois ? Certaines hypothèses doivent-elles être vérifiées au préalable ? (donner brièvement les résultats des éventuels tests préalables: nom du test, « p-valeur » et conclusion du test)

Question 14 – Y-a-t-il un effet « sexe » ? Un effet « génotype » ? Une interaction entre les deux ? Quel test statistique est effectué pour évaluer ces effets ? Pour chaque effet donner la valeur de la statistique, la p-valeur du test, et concluez quant à l'effet.

Question 15 – Quelles « catégories » présentent des différences de moyennes (en comparaison 2 à 2 ou en comparaison multiple) ? Quelle est la catégorie la plus intéressante en termes de robustesse ?