

U.E. Biologie et Modélisation – Évaluation

L. Palmeira, A. Necşulea

Décembre 2005

Tous documents autorisés.

Vous pouvez avoir besoin d'autres commandes que celles suggérées.

Quelques fonctions utiles:

- * *? pour obtenir de l'aide;*
 - * *par() pour modifier les paramètres graphiques;*
 - * *plot() pour tracer un nuage de points;*
 - * *boxplot() pour tracer une boîte à moustaches;*
 - * *pie() pour tracer un "camembert";*
 - * *table() pour compter le nombre d'individus pour les modalités d'un facteur;*
 - * *cor() pour calculer un coefficient de corrélation;*
 - * *lm() pour ajuster un modèle linéaire à un nuage des points;*
- Les fonctions suggérées ne sont là que pour vous aider, et ne sont en aucun cas les seules fonctions utiles.*

Nom :

Prénom :

Table des matières

1	Poulets, nutrition et représentations graphiques	2
2	Piratage des logiciels et bénéfices du secteur IT	3
3	Trois variables - mystère, deux corrélations	4
4	Étude d'une population de chats de la Croix-Rousse	5
5	Groupes sanguins et loi de Hardy-Weinberg	6

1 Poulets, nutrition et représentations graphiques

1. Chargez les données `ChickWeight` à l'aide de la fonction `data()`. Qu'y a-t-il dans ces données?

2. Donnez les commandes nécessaires pour, dans une même fenêtre graphique, tracer les quatre graphiques suivants :
 - ★ Le nuage de points du poids des poulets en fonction du temps, en spécifiant un nom pour l'axe des x et un nom pour l'axe des y, ainsi qu'un titre.
 - ★ L'histogramme du poids des poulets, en spécifiant un nom pour l'axe des x et un nom pour l'axe des y, ainsi qu'un titre.
 - ★ Les boîtes à moustaches des poids des poulets selon le régime alimentaire, en donnant une couleur pour chaque modalité du facteur (rouge, bleu, vert et jaune).
 - ★ Le “camembert” des régimes alimentaires, en modifiant les couleurs de manière à utiliser les mêmes que précédemment (pensez à vérifier l'ordre des couleurs).

Il n'est pas nécessaire de recopier les graphiques.

2 Piratage des logiciels et bénéfices du secteur IT

En décembre 2005, les organisations BSA (Business Software Alliance) et IDC (International Data Corporation) ont réalisé une étude sur l'impact économique du secteur de la technologie de l'information (IT), face au problème du piratage des logiciels.

Une des conclusions de cette étude est que la diminution du taux de piratage provoque une hausse des bénéfices apportés par ce secteur.

Les auteurs de l'étude ont mesuré, pour 69 pays, le taux de piratage des logiciels pour l'année 2004, ainsi que la part de PIB représenté par le secteur de la technologie de l'information (les deux valeurs sont données en pourcentages). Vous pouvez retrouver ces données à l'adresse suivante :

<http://pbil.univ-lyon1.fr/R/donnees/it2004.txt>

1. Importez ces données dans l'environnement R et mettez-les dans une nouvelle variable appelée `it`. Ecrivez la commande nécessaire à l'importation des données sous R.

2. Tracez le graphique représentant la dépendance entre le taux de piratage et la part de PIB du secteur IT. Superposez sur ce graphique le modèle linéaire correspondant. Recopiez ci-dessous les commandes correspondantes et le graphique obtenu.

3. Quel est le coefficient de corrélation linéaire entre les deux variables ?

4. Au vu des informations précédentes, est-ce que vous confirmez la conclusion de l'étude ?

3 Trois variables - mystère, deux corrélations

L'objectif de cet exercice est de comparer deux régressions linéaires. On vous rappelle que le coefficient de corrélation (calculé avec la fonction `cor()` dans R) est une mesure de la dépendance linéaire entre deux variables.

Vous disposez de trois variables x , y et z , sur lesquelles vous n'avez aucune autre information. Les données se trouvent à l'adresse :

`http://pbil.univ-lyon1.fr/R/donnees/compcor.txt`.

1. Importez ces données dans l'environnement R et mettez-les dans une nouvelle variable appelée `compcor`. Ecrivez la commande nécessaire à l'importation des données sous R.

2. Calculez les coefficients de corrélation entre les variables x et y d'une part, et x et z d'autre part.

3. Pour lequel des deux cas ($x - y$ et $x - z$) trouvez-vous que la dépendance linéaire est la plus forte ?

3. A l'aide d'un test de Chi-Deux, comparez les effectifs observés pour les six génotypes aux fréquences attendues sous la loi de Hardy-Weinberg. Conclure.