

TP Statistiques et Séries Chronologiques
Université de Lorraine

Introduction à R

Clément Dell'Aiera

1 Exercice 1

1. Que font les commandes *pnorm*, *qnorm*, *rnorm*, *dnorm*? Utilisez l'aide de R.
2. Tracer la fonction de densité de la loi normale. Faites varier les paramètres et afficher les différentes courbes sur le même graphique.
3. Simuler 2 vecteurs X et Y contenant chacun $N = 100$ variables indépendantes identiquement distribuées suivant une loi normale $\mathcal{N}(0, 1)$.
4. Afficher les points de coordonnées $(X[j], Y[j])$ dans le plan, pour j allant de 1 à 100.
5. Tracer la fonction de répartition empirique des $X[j]$.
6. Soit \mathcal{E} une v.a. de loi exponentielle de paramètre 1, et U une v.a. suivant une loi uniforme sur $[0, 2\pi]$. On pose

$$(X, Y) = (\sqrt{\mathcal{E}} \cos(U), \sqrt{\mathcal{E}} \sin(U)).$$

Quelle est la loi du couple (X, Y) ? (Vous pouvez le prouver, ou observer grâce à R ce qu'il se passe en simulant ces variables et en les traçant.)

2 Exercice 2

1. Une table est déjà en mémoire dans R : la table *stackloss*. Analyser la rapidement.
2. Tracer *stack.loss* en fonction de *Air.Flow*. Qu'en pensez-vous?
3. Effectuer la régression linéaire de *stack.loss* en fonction des autres variables. Quelles sont celles qui sont significatives?

3 Exercice 3

Le fichier *ozone.dta* contient les variables suivantes, pour une série de journées (qui sont ici nos individus) :

— l'identifiant de la journée,

- le maximum d’ozone (variable `maxO3`)
- l’heure à laquelle le maximum d’ozone a été obtenu (heure),
- les températures à 6h, 9h, 12h, 15h, 18h (resp. `T6` à `T18`)
- la nébulosité à 6h, 9h, 12h, 15h, 18h (resp. `Ne6` à `Ne18`)
- la projection du vent sur l’axe est-ouest à 12h (`Vx`),
- le maximum d’ozone de la veille (`maxO3v`).

Le but est de modéliser la valeur des pics d’ozone en fonction de grandeurs physiques facilement mesurables (température, heure, nébulosité, vent) afin d’avoir des approximations de la qualité de l’air faciles et rapides à obtenir.

1. Importer la table, et afficher un résumé de ce qu’elle contient.
2. Tracer `maxO3` en fonction de `T12`, puis effectuer une régression linéaire. Ajouter la droite de régression sur le graphique. Soignez la présentation.
3. Afficher les résultats de la régression.
4. Extraire les résidus et tracer leur densité estimée.
5. Effectuer la régression de `maxO3` sur toutes les variables, et supprimer récursivement celles qui ne sont pas significatives, jusqu’à ce qu’elles le soient toutes.

4 Exercice 4

1. Sur le site data.gouv.fr, vous pourrez trouver des tables de données publiques en libre accès. Choisissez un thème qui vous intéresse, puis une table en conséquence. Télécharger là.
2. Les tables sont souvent au format `.xls` : vous aurez besoin d’installer un package pour pouvoir les lire. La commande pour ce faire est `install.packages("nom du package")`. Installer le package `gdata`.
3. Analyser votre table.

5 Exercice 5

1. Télécharger et installer le package `ISwR` (Introductory Statistics with R).
2. Utiliser la commande `summary` pour analyser rapidement la table `bp.obese`. L’échantillon provient d’un échantillon de population mexicaine en Californie, et la table décrit 3 variables : le sexe (femme = 1, homme = 0), le ratio d’obésité (`obese`) et la pression sanguine systolique en *mm* de mercure (`bp`).
3. Représenter les données dans un graphe, en utilisant des symboles différents pour les hommes et les femmes.
4. Expliquer la pression sanguine en fonction du ratio d’obésité, puis du ratio d’obésité et du sexe.
5. Tracer sur un même graphe les courbes correspondant aux régressions dans les 2 modèles. Soignez la présentation (couleurs différentes, légende,...)

6 Exercice 6

1. Télécharger la librairie *MASS*.
2. Analyser rapidement la table *cats* et afficher les variables les unes en fonctions des autres par paires.
3. Effectuer une régression linéaire selon le modèle $Hwt \sim Bwt * Sex$. Cela apporte-t-il quelque-chose par rapport au modèle $Hwt \sim Bwt + Sex$?
4. Visualiser les composantes de votre régression.
5. En extraire les prédiction, les coefficients, les résidus, les résidus studentisés, et la formule du modèle.
6. Tracer le *qqplot* des résidus studentisés ainsi que la première bissectrice.
7. Tracer le graphe des résidus contre les prédictions.
8. Tracer le graphe des distance de Cook.
9. Observer les attributs que vous donne *summary*
10. Afficher le R^2 ajusté de la régression, le nombre de degrés de liberté résiduels, la matrice de variance-covariance des paramètres estimés.