

# Examen - M1 Maths-fi - 06/01/2010

Dans tous les tests demandés, le seuil sera 5%. Pour la lecture des tests, on donnera les hypothèses nulle et alternative, la valeur numérique de la statistique de test, la  $p$ -valeur, la conclusion du test.

Formulaires, notes et programmes de TD, poly de R et calculatrices autorisés.

## Exercice 1 (Etude de rendements)

Nous étudions les données de A.E. Freeny (voir *The New S Language* par R.A. Becker, J.M. Chambers et A.R. Wilks). Les données sont contenues dans une table `freeny`, qui se situe dans la librairie `malib`. La variable explicative est le rendement trimestriel  $Y$  d'une action, observé entre le second trimestre 1962 et fin 1971. Les explicatives sont la valeur au trimestre précédent `lagY`, l'indice des prix `price`, le niveau global des revenus `income` et le potentiel du marché `market`.

Le code et les sorties SAS sont en annexes 1 à 3 et il faut s'y référer pour répondre aux questions. Il n'y a pas de programmation à faire dans cet exercice.

On estime les coefficients de la régression suivante par MCO :

$$Y = \beta_0 + \beta_1 \text{lag}Y + \beta_2 \text{price} + \beta_3 \text{income} + \beta_4 \text{market} + \varepsilon. \quad (1)$$

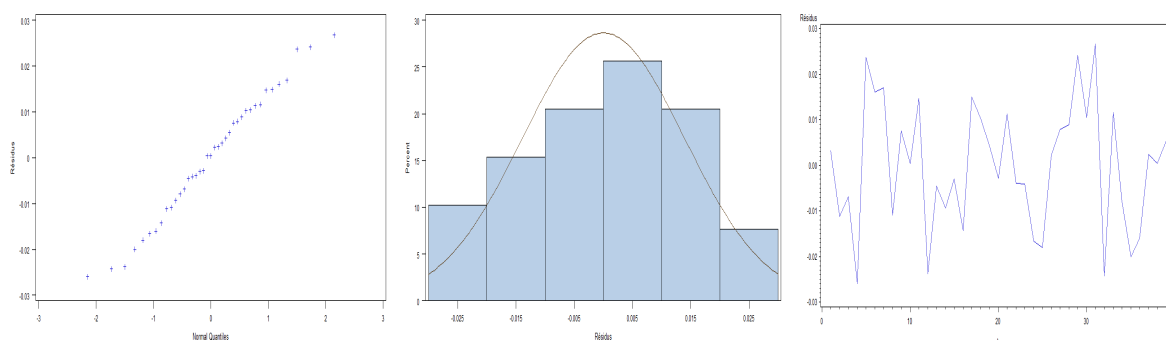


FIGURE 1 – Etude des résidus de la régression (1). Gauche : graphique quantile-quantile avec la loi normale. Centre : Histogramme des résidus. Droite : Résidus en fonction du temps.

1. Combien d'observations  $n$  a-t-on ? Combien d'explicatives  $K + 1$  a-t-on (constante comprise) ?
2. Calculer l'estimateur MCO de la variance  $\sigma^2$  des résidus.
3. Calculer les valeurs du coefficient de détermination  $R^2$  et du  $R^2_{adj}$ . La régression est-elle de bonne qualité ?
4. La régression est-elle globalement significative ?
5. Le coefficient de la variable `lagY` peut-il être considéré comme nul ? Commenter.
6. Qu'en est-il de la significativité des autres coefficients ?

7. Si l'indice des prix vaut 4, si le niveau global des revenus est de 6 et si le potentiel de marché de 13, quelle est le rendement  $Y$  prédit ?
8. Etablir un intervalle de confiance à 95% pour  $\beta_4$ . Le quantile de la loi de Student à 97,5 degrés de liberté est 2,03.
9. On souhaite tester si  $\beta_2 = -\beta_3$ . Lire le test correspondant dans les sorties SAS et commenter.
10. On récupère et on étudie les résidus de la régression. La procédure `univariate` nous fournit les graphiques en Figure 1. Commentez-les.
11. Les résidus peuvent-ils être considérés comme Gaussiens ? En quoi cette hypothèse est-elle utile ?
12. Par quelques courtes phrases, faire un bilan de la régression.

### Exercice 2 (Etude sur $\mathbf{R}$ : taches solaires)

*Cet exercice sera traité sur une copie séparée. Il faudra également rendre un fichier de programme sous la forme `Nom_Exam.R`, à envoyer à [laurence.marsalle@univ-lille1.fr](mailto:laurence.marsalle@univ-lille1.fr)*

Les astronomes ont remarqué très tôt (depuis au moins 2000 ans) des taches sombres sur la surface du soleil (attention, il ne faut jamais regarder directement le soleil sans lunette adaptée!). Elles correspondent à des zones moins chaudes et sont des manifestations de l'activité solaire, dues à des variations de champ magnétique à la surface de celui-ci. Une tache peut avoir une durée de vie de quelques semaines.

Nous disposons d'un relevé du nombre mensuel moyen de taches solaires entre 1749 et 1983, soit 2820 observations (les données entre 1749 et 1960 ont été collectées par l'Observatoire Fédéral Suisse de Zurich, et celles après 1960, par l'Observatoire Astronomique de Tokyo). L'échantillon est noté  $(x_t)_{t \in \llbracket 1, T \rrbracket}$ , où  $T = 2820$ . Il est stocké dans  $\mathbf{R}$  sous forme d'un vecteur appelé `sunspots`.

Nous considérerons dans cet exercice que les données  $(x_t)_{t \in \llbracket 1, T \rrbracket}$  sont des réalisations indépendantes de variables aléatoires de même loi.

1. Tracer les  $x_t$  en fonction de  $t$ . Au vu du graphique, que pensez-vous de l'hypothèse d'indépendance des observations ? En particulier, pourquoi l'indépendance entre les données relatives à deux mois consécutifs est-elle critiquable ?
2. Calculer la moyenne, l'écart-type, le minimum, le maximum et la médiane des observations. Commenter.
3. Explorer graphiquement la distribution des  $(x_t)_{t \in \llbracket 1, T \rrbracket}$ . Est-il raisonnable de modéliser cette loi par une loi exponentielle ?
4. Déterminer l'estimateur du maximum de vraisemblance  $\hat{\lambda}_T$  de  $\lambda$ , paramètre de cette loi exponentielle (*théorique*), puis donner sa valeur numérique.
5. Cet estimateur  $\hat{\lambda}_T$  est-il convergent lorsque  $T \rightarrow +\infty$  ? Si oui, préciser le mode de convergence, et justifier la réponse.
6. Au moyen d'une  $\delta$ -méthode, donner la loi asymptotique de  $\sqrt{T}(\hat{\lambda}_T - \lambda)$ .
7. En déduire un intervalle de confiance pour  $\lambda$  (*théorique*), et faire l'application numérique.
8. Donner une approximation de la probabilité que le nombre de taches solaires observées dépasse 300 (valeur non observée sur les données pour lesquelles le maximum est 253.8). On obtient ainsi une

estimation d'un événement très rare ...

**9.** On souhaite savoir s'il y a des variations d'un siècle à l'autre. Dans ce but, créer un tableau de contingence avec en ligne le siècle (18ème, 19ème ou 20ème) et en colonne la classe  $[0, 100)$ ,  $[100, 150)$  et  $[150, 300)$  à laquelle appartiennent les observations.

**10.** Faire un test d'indépendance du  $\chi^2$ . Détailler les résultats de **R** et commenter.