

Examen

Durée de l'examen: 2 heures. Les notes de cours et les polycopiés sont autorisées, ainsi que les planches de TP (et les fichiers R associés). Il est nécessaire de récupérer au préalable les fichiers *tauber.txt*, *sida.txt* et *LGN.R* sur le bureau virtuel.

En plus d'une copie contenant les réponses aux questions, vous enverrez par mail à l'adresse kowalski@cmi.univ-mrs.fr, trois fichiers *exercice1.R*, *exercice2.R* et *exercice3.R* contenant uniquement les commandes nécessaires à la réalisation des exercices. Tous les commentaires permettant d'éclairer et d'appuyer le raisonnement sont également les bienvenus. Les remarques pertinentes donneront droit à des points de bonifications.

Les exercices sont indépendants et peuvent être traités dans l'ordre voulu. *Bien lire les énoncés des questions, jusqu'au bout !*

Exercice 1 (Influence du THC chez les rats)

Le fichier *thc.txt* contient les données relative à une expérience pour connaître l'effet du THC sur l'activité des rats. Pour cela, 5 quantités différentes de THC ont été injectés à 5 groupes de rats. La première colonne contient le nom du groupe et la deuxième colonne une mesure de l'activité de chaque rat.

1. Pour chaque groupe, calculer la moyenne de l'activité des rats, la variance, et dessinez l'histogramme.
2. Rappeler l'utilité de l'ANOVA.
3. Quelles sont les conditions pour effectuer une ANOVA ? Sont-elles vérifiées ?
4. Effectuer l'ANOVA en suivant le protocole de Fisher et conclure précisément.

Exercice 2 (Evolution de la population aux Etats Unis)

Le fichier *pop_eu.txt* contient l'évolution de la population (en million d'habitants) aux États-Unis entre 1800 et 1990 (une mesure tous les 10 ans).

1. Représentez graphiquement l'évolution de la population.
2. Testez la nullité du coefficient de corrélation linéaire et conclure.
3. Effectuer une régression linéaire et tracer la droite de régression.
4. D'après ce modèle, combien devrait-il y avoir d'habitant aux États-Unis en 2010 ? Donnez une prédiction ponctuelle ainsi qu'un intervalle de *prédiction* à 95%.
5. Représentez graphiquement l'évolution de la **racine carrée** du nombre d'habitant. (*On rappelle que la racine carrée est obtenue par la commande sqrt*)
6. Effectuer une régression linéaire et tracer la droite de régression avec le changement de variable suggéré.
7. Graphiquement, quel modèle vous semble le plus approprié pour la régression linéaire ?
8. Donner un intervalle de *prédiction* du nombre d'habitant pour l'an 2010.

Exercice 3 (Influence du choix de la partition pour le test du χ^2)

On se propose de regarder comment se comporte la *p-valeur* du test du χ^2 d'adéquation à une loi continue, en fonction de la partition choisie. On fournit le programme *partition.R*, qui, pour un échantillon gaussien donné, renvoie l'histogramme (sans le dessiner) associé à une partition admissible aléatoire (ie. qui contient plus de 5 échantillons).

L'appel se fait simplement par: *h<-partition(X)*, où X est un échantillon gaussien. On rappelle que *h* contient une liste contenant les éléments suivants:

- breaks: les classes créés;
 - counts: le nombre d'observations par classes;
1. Rappeler brièvement le principe du test du χ^2 d'adéquation pour une loi *continue*.
 2. Générer 1000 échantillons indépendants et identiquement distribués selon une loi $\mathcal{N}(0,1)$, et créer l'histogramme associé à une partition admissible.
 3. Faire une fonction permettant de calculer les effectifs théoriques associés à une partition. (*On pourra très largement s'inspirer du script de l'exercice 4 de l'examen 2008, fait en TD.*)
 4. Effectuer un test du χ^2 d'adéquation pour cette partition et noter sa *p-valeur*.
 5. Reprendre la question précédente, pour une autre partition créée aléatoirement. Que vaut la *p - valeur* à présent ?
 6. Faire un script ou une fonction, qui permet de faire un nombre arbitraire de fois l'expérience précédente, et qui enregistre les *p-valeurs*. Dessiner alors ces *p-valeurs* sur un graphique (on fera au moins une centaine d'expérience). Que remarquez-vous ? (*On rappelle que la p-valeur d'un test est donnée par l'attribut p.value*)
 7. Décrire et commenter le programme *partition.R*.