

Examen médian

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 (Tailles et poids d'enfants de 4 à 7 ans)

1. Récupérer le fichier *tauber.txt*.
2. Créer un *data.frame* contenant les données du fichier, et enregistrez le dans une variable nommée *tauberData*.
3. Exécutez et expliquez à quoi sert l'ensemble des commandes suivantes:

```
iG <- tauberData$GENRE == "M"
iF <- tauberData$GENRE == "F"
tailleG <- tauberData$TAILLE[iG]
tailleF <- tauberData$TAILLE[iF]
```

4. Donnez la moyenne et la variance des vecteurs *tailleG* et *tailleF*.
5. Dessinez l'histogramme des variables *tailleG* et *tailleF*. Quelle distribution ces vecteurs semblent-ils suivre ?
6. On veut savoir si les garçons âgés de 4 à 7 ans sont plus grands que les filles.
 - (a) Quel test proposez-vous ?
 - (b) Effectuer ce test sous R. *Remarque: par défaut, l'option alternative de la commande `t.test` vaut "two.sided" pour faire un test bivarié. On peut passer `alternative="greater"` (resp. "less") pour un test univarié à droite (resp. à gauche).*
7. Calculez l'indice de masse corporelle de chaque individu ($IMC = \frac{\text{poids en kg}}{(\text{taille en metre})^2}$). Arrondissez l'IMC grâce à la commande *round* et enregistrez le résultat dans la variable *IMC*.
8. On veut savoir si l'IMC évolue avec l'âge des enfants.
 - (a) Quelles sont les hypothèses à tester ?
 - (b) Exécutez la commande

```
table(tauberData$AGE,IMC)
```

Cette commande permet de créer le tableau de contingence associée aux variables aléatoires *IMC* et *Age* (un exemple de cette commande est donné dans l'énoncé de l'exercice 2)
 - (c) Effectuer le test adéquat et conclure.

Exercice 2 (SIDA)

La commande `table(X,Y)` permet de créer le tableau de contingence associée aux variables aléatoires qualitatives X et Y . X et Y peuvent être un vecteur de n'importe quel type (numériques, chaînes de caractères...) contenant les réalisations, mais doivent être de même taille. Exemple:

```
X <- c("bleu","vert","bleu","marron","marron","bleu","vert","marron","bleu")
Y <- c("brun","roux","blond","brun","brun","blond","blond","brun","roux")
table(X,Y)
      Y
X      blond brun roux
bleu      2    1    1
marron     0    3    0
vert      1    0    1
```

1. Récupérez le fichier *sida.txt* et créer un *data.frame* contenant les données du fichier et créez un *data.frame*.
2. De quelle(s) variable(s) peut-on dire que l'infection par le VIH est dépendante ? *On pourra utiliser la commande table décrite ci-dessus afin de construire les tableaux de contingences utiles à cette question*

Exercice 3 (Loi des grands nombres)

On rappelle le théorème de la loi de grands nombres:

Théorème 1 Soit (X_k) , $k \in \mathbb{N}$ une suite de variables aléatoires indépendantes, identiquement distribuées suivant la loi de Bernoulli de paramètre p . Alors,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n X_k = p .$$

Commenter et décrire le programme suivant (que font les différentes lignes du programme, à quoi correspondent la courbe et la ligne tracées, quel théorème cela illustre-t-il et pourquoi...):

```
LGN <- function(p){
  for(i in 1:5){
    A=rbinom(500,size=1,prob=p);
    V=cumsum(A);
    D=1:500;
    V=V/D;
    x11()
    plot(V,type="l");
    W=p*rep(1,500);
    lines(W);
  }
}
```

Le modifier pour faire apparaître la valeur de p en rouge, donner un nom aux axes, ainsi qu'un titre aux graphiques.