

## Examen

Durée de l'examen: 2 heures. Seules les notes de cours et les planches de TP sont autorisées. Il est nécessaire de récupérer au préalable les fichiers *loyer.txt* et *1euro.txt* sur le bureau virtuel.

La copie devra contenir les commandes nécessaires à la réalisation des différentes questions (dans leur intégralité). Tous les commentaires permettant d'éclairer et d'appuyer le raisonnement sont également les bienvenus.

### Exercice 1 (Régression)

Cet exercice utilise le fichier *loyer.txt*

1. Construire un `data.frame` contenant l'ensemble des informations du fichier. On prendra soin de conserver le nom des colonnes.
2. Calculer la moyenne de l'indice de référence des loyers entre le premier trimestre 2000 et le troisième trimestre 2007.
3. Construire un graphique rendant compte de l'évolution de cet indice en fonction du temps. Indiquer les commandes permettant de donner un titre à cette figure.
4. Construire et afficher la droite de régression de l'indice des loyers en fonction du temps. Donner un intervalle de confiance pour la valeur de l'indice au premier trimestre 2009.

### Exercice 2 (Ressources humaines)

Le directeur des ressources humaines d'une grande chaîne de distribution effectue un sondage auprès de 200 de ses cadres choisis au hasard dans son fichier pour connaître leur niveau de satisfaction vis à vis de leur travail. Ce caractère est réparti en trois modalités: élevé, moyen et faible. On note aussi la catégorie sociale de l'un des sondés sous une des formes suivantes: A (30 000 euros et moins), B (entre 30 000 et 40 000 euros) et C (plus de 40 000 euros). Les résultats de l'enquête sont rassemblés dans le tableau suivant:

	Elevé	Moyen	Faible
A	13	19	25
B	28	29	28
C	24	18	16

Le DRH souhaite savoir si la satisfaction d'un cadre est dépendante de son niveau social.

1. Formuler le problème précédent en termes de tests d'hypothèses.
2. Rappeler brièvement le principe du test utilisé.
3. Que peut-on dire au vu de cet échantillon?
4. Afficher le tableau des résidus.

### Exercice 3 (Peine de mort)

Afin d'évaluer l'impact d'un exposé sur l'opinion face à la peine de mort, on demande à un échantillon de cinq individus de se positionner par rapport au sujet sur une échelle ordinale d'attitude (de 1='tout à fait contre' à 7='tout à fait pour') avant et après l'exposé. Voici les résultats obtenus:

Avant	2	4	5	3	4
Après	1	2	6	6	4

1. Formuler le problème précédent en termes de tests d'hypothèses.
2. Expliquer pourquoi il n'est pas raisonnable d'utiliser le test de Student-Fisher dans cette situation.

Dans cet exercice, nous allons nous intéresser à un autre type de test permettant de répondre à la problématique: le test de Mann-Whitney. Considérons (de manière générale) deux échantillons  $(x_1, \dots, x_n)$  et  $(y_1, \dots, y_m)$  dont on souhaite tester l'égalité des moyennes. Si on mélange les deux séries et si on ordonne le tout par valeurs croissantes, on doit obtenir un mélange homogène. Le test de Mann-Whitney repose sur ce principe. Pour cela les suites étant réordonnées, on compte le nombre total de couples  $(x_i, y_i)$  pour lesquels  $x_i$  a un rang plus grand que  $y_i$ . Soit  $U$  ce nombre. Il est facile de voir que  $U$  varie de 0 à  $nm$ . Le cas  $U = 0$  correspond à la situation suivante:

$$x_1, x_2, \dots, x_n, y_1, \dots, y_m.$$

Il y a dans ce cas deux phases différentes: les moyennes des deux échantillons sont clairement différentes. Si au contraire  $U$  est proche de  $nm/2$ , on pourra 'raisonnablement' affirmer que les deux échantillons sont homogènes.

Il est possible de calculer la loi de la statistique  $U$ . En particulier, si les deux distributions sont issues de la même population,  $\mathbb{E}[U] = nm/2$ . Le test de Mann-Whitney est donc tout naturellement basée sur l'idée suivante:

- Si  $U$  est 'proche' de  $nm/2$ , on acceptera l'hypothèse d'égalités des moyennes.
- Si  $U$  est significativement différent de  $nm/2$ , on pourra affirmer que les échantillons ne proviennent pas de la même population.

Ce test peut être réalisé sous **R** à l'aide de la commande `wilcox.test`. Cette dernière prend simplement comme argument les deux échantillons  $x$  et  $y$  et renvoie les informations usuelles.

3. Effectuer un test de Mann-Whitney à l'aide de la commande `wilcox.test`. Que vaut la p-valeur? Même question pour la statistique de test.
4. Au vu des observations, peut-on affirmer que l'exposé a une influence sur l'opinion face à la peine de mort?

#### Exercice 4 (Adéquation)

Cet exercice utilise le fichier `1euro.txt`. Ce dernier contient le poids en grammes de 30 pièces de un euro (2ème colonne). Le but est de déterminer si le poids d'une pièce prise au hasard suit ou non une loi gaussienne. Afin de répondre à cette question, un étudiant en troisième année de Licence MASHS a utilisé les commandes suivantes:

```
> euro <- read.table("1euro.txt")
> poids <- euro[,2]
> hist <- hist(poids,proba=TRUE)
> eff <- hist$counts
> classes <- hist$breaks
> th <- pnorm(classes,mean(poids),sd(poids))
> th[10]<- 1
> th[1] <-0
> proba_theo <- diff(th)
> chisq.test(eff,p=proba_theo)
```

Commentez chacune des commandes utilisées. Expliquez pourquoi les conclusions de cet étudiant (acceptation de l'hypothèse nulle) sont discutables.