



Examen CT de Statistique
--------------------------

Document autorisés : tout type (papier ou numérique), voire ordinateur personnel

**IMPORTANT :** Les étudiants doivent choisir un groupe d'exercices : soit 1-2 (J. Bastien), soit A-C (AB. Dufour). Il est conseillé de choisir ceux proposés par votre enseignant. Noter le nom de votre enseignant en début de copie. Merci

**AVERTISSEMENT pour les exercices 1-2**

L'ensemble des fichiers de données nécessaires pour cet examen (`morphosportbis.txt` et `file138.txt`) est normalement disponible sur le réseau de l'université Lyon I : il faut aller sur :

- 'Poste de travail',
- puis sur le répertoire 'P:' (appelé aussi : enseignants sur '`\Univ-lyon1\enseignement\homes`'),
- puis '`jerome.bastien`',
- enfin sur '`M1 APA\examen\CT`').

**Exercice 1** (J. Bastien).

Chargez le fichier '`morphosportbis.txt`'.

- (1) Analyser rapidement par le calcul et un graphique la variable 'sport'.
- (2) Les variables 'tde' (taille debout) et 'sport' sont-elles liées ? Y'a-t-il des sports où les individus sont plus grands que dans les autres sports ? Commentez ?
- (3) (a) La variable 'pratique.bask' est une variable qualitative qui contient la modalité 'oui' si le sport pratiqué est le basket et 'non' sinon.  
Les variables 'tde' (taille debout) et 'pratique.bask' sont-elles liées ?  
(b) Commentez les résultats obtenus par rapport aux résultats de la question 2.  
(c) Combien y'a-t-il de sportifs qui pratiquent le basket ? Cette méthodologie est-elle vraiment pertinente ?

**Exercice 2** (J. Bastien).

On étudie le fichier de données '`file138.txt`' qui comporte 56 notes d'étudiants, répartis dans plusieurs groupes de TD (`g1`, `g2` et `g3`).

- (1) La colonne 'note.brute' contient les notes brutes, obtenues dans chacun des groupes de TD, où les modalités de notations (sujets de contrôle, barèmes,...) ne sont pas identiques. Analysez les notes brutes graphiquement et par le calcul.
- (2) Les variables 'note.brute' et 'groupe' sont-elles liées ? Interprétez !
- (3) Pour tenter de réharmoniser les notes, l'enseignant responsable de l'UE a décidé de faire subir à chaque note une "dilatation-translation" (chaque note  $n_i$  est remplacé par  $an_i + b$ ) de telle sorte que, pour chaque groupe de TD, la moyenne et l'écart-type du groupe soient identiques à la moyenne et l'écart-type de l'ensemble des notes brutes. Les notes ainsi calculées sont indiquées dans la colonne 'note'.
  - (a) Quelles sont la moyenne et l'écart-type de l'ensemble des notes brutes ?
  - (b) En utilisant par exemple les commandes : ('file138' désigne le nom de la variable dans laquelle vous avez enregistré le fichier 'file138.txt')
 

```
indg1<-file138$groupe=="g1"
file138$note[indg1]
```

 vérifier que les moyennes et les écart-types des nouvelles notes, groupe par groupe, sont bien égales à la moyenne et à l'écart-type de l'ensemble des notes brutes.
  - (c) Les variables 'note' et 'groupe' sont-elles liées ? Interprétez !
  - (d) Concluez sur l'harmonisation faite. Cela vous semble-t-il juste comment façon d'harmoniser ? Que proposeriez-vous ?

### Exercice A (AB. Dufour).

- (1) Un enseignant fait passer un examen à 18 étudiants. L'expérience a montré que pour ce type de sujet, la probabilité de réussite est  $p = 0.8$ . Soit  $X$  la variable 'nombre de réussites à l'examen' (on reconnaît ici une loi binomiale).
  - (a) Donner les paramètres de cette loi binomiale  $X$ .
  - (b) Donner la moyenne et la variance associées à cette loi.
  - (c) Calculer la probabilité qu'exactly 16 étudiants réussissent l'examen  $P(X = 16)$ .
  - (d) Calculer la probabilité qu'au moins 16 étudiants réussissent l'examen  $P(X \leq 16)$ .
- (2) On sait que quand  $n$  est grand,  $X$  peut être approchée par une **loi normale** de moyenne 14.4 et de variance 2.88. Il est rappelé que la variable centrée réduite  $Z = \frac{X-14.4}{\sqrt{2.88}}$  suit alors une loi normale centrée réduite.
  - (a) On pose  $X = 17$ , donner la valeur de  $z$  et  $P(Z < z)$ .
  - (b) On pose  $X = 10$ , donner la valeur de  $z$  et  $P(Z < z)$ .
  - (c) A l'aide des informations ci-dessus, calculer  $P(10 < X < 17)$ .

### Exercice B (AB. Dufour).

Dans un sondage réalisé pour l'université de Montréal, Sylvie Beauchamp et Pierre Bouchard se sont interrogés sur les habitudes de lecture d'une population adulte. Chacune des personnes interrogées devait répondre à la question : avez-vous lu au moins un livre au cours des trois mois précédant

l'interview ?

Les données :

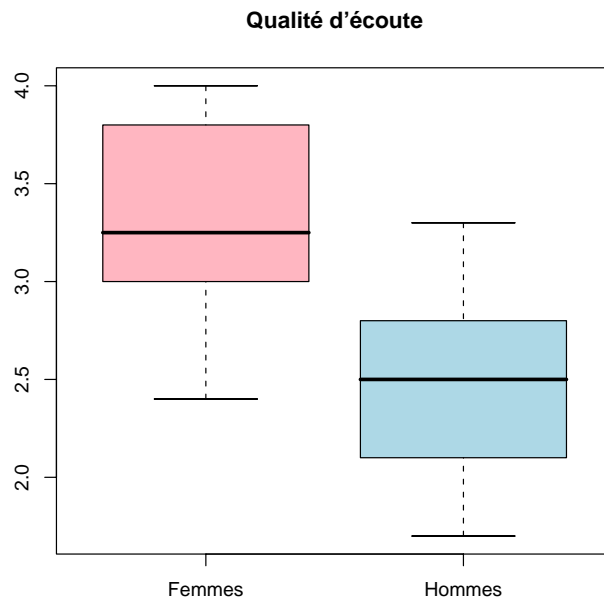
	hommes	femmes
oui	265	406
non	319	264

- (1) Combien y-a-t-il eu de personnes interrogées dans cette étude ?
- (2) Quelles sont les fréquences observées d'hommes et de femmes ayant lu au moins un livre au cours des trois mois précédant l'interview ?
- (3) Existe-t-il une différence de lecture, en proportions, entre les hommes et les femmes ?

**Exercice C** (AB. Dufour). La bibliothèque de l'Université organise une enquête de satisfaction (enquête fictive). La variable étudiée concerne une évaluation de l'écoute du personnel à la demande de l'étudiant. Une note sur 5 points est attribuée (de 1. très satisfait à 5. très insatisfait). On distingue deux groupes selon le sexe. Les réponses sont stockées dans deux vecteurs `repH` pour les hommes et `repF` pour les femmes.

```
repH <- c(2.7, 2.2, 2.1, 3.3, 2.8, 2.4, 2.8, 2.6, 2.9, 1.7, 1.9,  
2.1)  
repF <- c(2.9, 4, 3.2, 3.2, 3, 2.4, 3.8, 3.4, 3.2, 4, 3.3, 3.5,  
3.8, 2.4)
```

- (1) Calculer la moyenne et la variance descriptives des réponses des étudiants `repH`.
- (2) Donner l'intervalle de confiance de la moyenne de la qualité de l'écoute chez les étudiants, ceci au niveau 0.95.
- (3) Commenter la représentation graphique ci-dessous.



- (4) Quelle hypothèse peut-on émettre quant à la satisfaction de la qualité d'écoute selon le sexe ?
- (5) En utilisant les compétences statistiques acquises durant ce semestre, répondre à l'hypothèse proposée ci-dessus (Rédiger dans le sens d'un rapport de recherche).