



<b>Examen CFF2 de Statistique</b>
-----------------------------------

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

**AVERTISSEMENT**

L'ensemble des fichiers de données nécessaires pour cet examen (noteMT40A04.xls, L3APA06.xls, sondages\_sortie\_urnes.xls, STRIKES.xls) est normalement disponible à la fois

- en ligne sur <http://utbmjb.chez-alice.fr/UFRSTAPS/index.html> à la rubrique habituelle (voir 'examen', en bas de la page) ;
- en cas de problème internet, 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 PPMR\examen\CFF2'.

**Exercice 1.**

Dans cet exercice, on étudie les données disponibles dans le fichier `noteMT40A04.xls`. Elles correspondent à 56 notes d'un groupes d'étudiants Belfortains.

- (1) Quels graphiques pouvez-vous tracer pour ces données ? Dessinez en quelques-uns à main levée sur votre copie.
- (2) Déterminer les indicateurs statistiques usuels pour ces données.
- (3) Calculer la taille d'effet relative par rapport à la norme 10. Conclure.

**Exercice 2.**

Chargez le fichier `L3APA06.xls` qui contient un certains nombre de données pour 58 étudiants en APA.

Analysez-en graphiquement et par le calcul les sports pratiqués (variable 'sport').

**Exercice 3.**

Dans cet exercice, on étudie les données disponibles dans le fichier `sondages_sortie_urnes.xls`. Il comporte des données recueillies<sup>1</sup> lors de sondages "à la sorties des urnes" pour le second tour de l'élection présidentielle de 2007. Ces sondages ont été réalisés dans 50 bureau de vote ; pour chacun

---

<sup>1</sup>données simplifiées en fait par rapport à la réalité.

de ces bureaux de vote, on a interrogé environ 100 électeurs. Compte tenu des votes blancs et nuls (non comptés) et des refus de réponse, le nombre réel de réponses n'est pas exactement de 100. Dans le fichier `sondages_sortie_urnes.xls`, se trouvent pour chaque bureau de vote considéré :

- 'total' : le nombre total d'électeurs ayant répondu ;
- 'nombre' : le nombre d'électeurs ayant déclaré avoir voté pour le (futur) gagnant, Monsieur N.S. ;
- 'pr' : la proportion du nombre d'électeurs ayant déclaré avoir voté pour le (futur) gagnant, Monsieur N.S., soit  $\text{nombre}/\text{total}$ .

(1)

- (a) Pourquoi peut-on considérer que la loi gouvernant la variable 'nombre' est la loi binomiale. Pour chaque bureau de vote, quels sont ses paramètres ?
- (b) Calculer les intervalles de fréquence au niveau de confiance  $NC=0.95$  correspondant aux bureaux de vote numéro 6, 10 et 14. Commentez !
- (c) On cherche maintenant à calculer automatiquement tous les intervalles de confiance au niveau de confiance  $NC = 0.95$ . Dans Rcmdr, créer dans l'ordre trois nouvelles variables
  - de nom `SEP` et égale à `sqrt((pr*(1-pr))/total)`
  - de nom `prmin` et égale à `pr-1.959964*SEP`
  - de nom `prmax` et égale à `pr+1.959964*SEP`
 Vérifier que les deux dernières colonnes créées contiennent la borne inférieure et supérieure des intervalles de confiance au seuil  $NC = 0.95$ .
- (d) Ces renseignements sont-ils véritablement intéressants pour l'institut de sondage qui les a collectés ?

(2)

- (a) Si on enregistre les données avec le nom 'Dataset', que font les commandes suivantes (dans Rgui ou dans la fenêtre de script de Rcmdr) :
 

```
sum(Dataset$total)
sum(Dataset$nombre)
sum(Dataset$nombre)/sum(Dataset$total)
```
- (b) Déduisez-en la proportion d'électeurs pour N.S. sur l'ensemble des bureaux de votes.
- (c) Déterminez les intervalles de confiance aux niveaux 0.95, 0.99 et 0.999999 de la proportion d'électeurs pour N.S.. Commentez !
- (d) On donne finalement les résultats finaux officiels :

	nombre
M. Nicolas SARKOZY	18983138
Mme Ségolène ROYAL	16790440

Commentez ces chiffres par rapport aux résultats que vous avez obtenus.

(3) *Question facultative*

- (a) Si on enregistre les données avec le nom 'Dataset', que font les commandes suivantes :

```
p0<-18983138/(18983138+16790440)
(Dataset$prmin<=p0)&(Dataset$prmax>=p0)
100*sum((Dataset$prmin<=p0)&(Dataset$prmax>=p0))/50
```

(b) Commentez-en les résultats.

#### Exercice 4.

Reprenez les données du fichier `noteMT40A04.xls`.

- (1) Peut on considérer que ces notes proviennent d'une loi normale ?
- (2) Tracer un graphe mettant cette propriété en évidence.
- (3) On souhaite tester la moyenne par rapport à la norme  $\mu_0 = 10$ .
  - (a) On teste l'hypothèse nulle  $H_0 : \mu = \mu_0$  contre  $H_1 : \mu < \mu_0$ . Justifier ce choix.
  - (b) Réalisez ce test avec  $\mathbb{R}$  et comparez avec les résultat de la question 3 de l'exercice 1.

#### Exercice 5.

Une publicité pour une vidéo de formation promet que les techniques proposées vont améliorer l'adresse de lanceurs au baseball pour réaliser des "strikes" et que, après cet entraînement, ils seront capables de réaliser des strikes sur au moins 60% de leurs lancers. Pour tester ces affirmations, 20 lanceurs ont été enrôlés, ils ont réalisé avant de voir la vidéo et de s'entraîner, 50 lancers chacun et le nombre de strikes a été relevé. Suite à l'entraînement, ils ont a nouveau réalisé 50 lancers et le nombre de strikes a été relevé. Ces informations sont disponibles pour les 20 lanceurs dans le fichier "STRIKES.xls" qui comprend deux variables :

- 'Before' : nombre de strikes réalisés (sur 50 essais) avant la vidéo et l'entraînement
  - 'After' : nombre de strikes réalisés (sur 50 essais) après.
- (1) Le nombre de strikes a-t-il augmenté significativement suite à la vision de la vidéo et à l'entraînement ?
  - (2) Sur l'ensemble des 20 individus, il y a eu après la vision de la vidéo 663 strikes réalisés sur  $20 \times 50 = 1000$  essais. Ont-il comme l'affirme la publicité, dépassé de façon significative le seuil de 60% de strikes ?

#### Corrigé

Un corrigé sera disponible sur <http://utbmjb.chez-alice.fr/UFRSTAPS/index.html>