



Examen CCF2 de Statistique

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

IMPORTANT : Les étudiants doivent choisir un groupe d'exercices : soit 1-3 (J. Bastien), soit A-D (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 1-3

L'ensemble des fichiers de données nécessaires pour cet examen (L3APA06.txt, qualitatifbis.txt et javelot.txt) 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 APA\examen\CCF2').

AVERTISSEMENT pour les exercices A-D

L'ensemble des données se trouve dans le fichier `m1data0809.txt` sur le site pédagogique 'Enseignement des statistiques en biologie' à l'adresse <http://pbil.univ-lyon1.fr/R/donnees/>.

Exercice 1 (J. Bastien).

(1) Créez le vecteur suivant

```
notes <- c(19, 8, 3, 0, 15, 4, 16, 9, 14, 8, 7, 3, 0, 8, 4, 19)
```

et analysez-le graphiquement et par le calcul.

(2) Chargez le fichier 'L3APA06.txt'; Analysez-en graphiquement et par le calcul les sports pratiqués (variable 'sport').

Exercice 2 (J. Bastien).

Considérons le jeu de données portant sur 592 étudiants (extrait de [Sne74]) Pour chaque étudiant on a observé 3 variables qualitatives : la couleur des cheveux, la couleur des yeux et le sexe. Les données se trouvent dans le fichier 'qualitatifbis.txt'.

- (1) On étudie indépendamment les unes des autres les deux variables 'cheveux' et 'yeux'.
 - (a) Tracer à main levée quelques graphiques importants.
 - (b) Déterminer quelques indicateurs pour chacune de ces variables.
- (2) On s'intéresse au croisement entre la variable 'cheveux' et la variable 'yeux'. Ces deux variables sont-elles dépendantes ?
- (3) On s'intéresse maintenant à une nouvelle variable qualitative de nom 'teinte.cheveux' qui aura la modalité
 - 'clair' pour les cheveux de couleurs 'Blond' et 'Roux'
 - 'foncé' pour les cheveux de couleurs 'Marron' et 'Noir'.

De même, on définit une nouvelle variable qualitative de nom 'teinte.yeux' qui aura la modalité

- 'clair' pour les yeux de couleurs 'Bleu', 'Noisette' et 'Vert'
- 'foncé' pour les yeux de couleurs 'Marron'.

- (a) Voir de nouveau le fichier 'qualitatifbis.txt' qui contient ces variables. On s'intéresse au croisement entre la variable 'teinte.cheveux' et la variable 'teinte.yeux'. Ces deux variables sont-elles dépendantes ?
- (b) Comparez avec la question 2 et critiquez la méthodologie de cette question.

Exercice 3 (J. Bastien).

Le fichier de données 'javelot.txt' comporte

- les longueurs (en mètres),
- les diamètres (en mètre à l'endroit où il est le plus important),
- les masses (en kilogrammes)
- et les types (2 types différents "A" et "B").

d'un certains nombre de javelots pour des épreuves masculines.

- (1) On étudie indépendamment les unes des autres les deux variables 'longueur' et 'masse'.
 - (a) Tracer à main levée quelques graphiques importants.
 - (b) Déterminer quelques indicateurs statistiques pour chacune de ces variables.
- (2) On s'intéresse au croisement entre la variable 'longueur' et la variable 'type'. Ces deux variables sont-elles dépendantes ?
- (3) Un texte réglementaire précise les éléments suivants : "Chez les hommes, le javelot réglementaire doit peser au moins 0.8 kg. Sa longueur peut varier de 2.6 m à 2.7 m et son diamètre de 25 mm à 30 mm au point le plus gros."
 - (a) Si 'javelot' désigne le nom de la variable dans laquelle vous avez enregistré le fichier 'javelot.txt', que font les commandes suivantes dans \mathbb{R} :

```
indl<-(javelot$longueur>=2.6)&(javelot$longueur<=2.7)
javelot$longueur[indl]
indm<-javelot$masse>=0.8)
javelot$longueur[indm]
```

```
inddia<-(javelot$diametre>=0.025)&(javelot$diametre<=0.03)
javelot$longueur[inddia]
```

- (b) Calculer la moyenne et l'écart-type des longueurs des javelots qui ont une longueur réglementaire.
- (c) Calculer la moyenne et l'écart-type des longueurs des javelots qui ont à la fois une longueur, un diamètre et une masse réglementaires.
- (d) Pourriez-vous créer un data.frame des javelots qui ont à la fois une longueur, un diamètre et une masse réglementaires ?
- (4) (a) On s'intéresse au croisement entre la variable 'longueur' et la variable 'masse'. Ces deux variables sont-elles dépendantes ?
- (b) Introduisez maintenant une variable 'auxi' égale au produit 'longueur×(diametre)^2', c'est-à-dire

```
auxi<-javelot$longueur*(javelot$diametre)^2
```

On s'intéresse au croisement entre la variable 'masse' et la variable 'auxi'. Ces deux variables sont-elles dépendantes ?

- (c) *Question facultative*

En admettant que le nombre défini par

$$\text{rapport} = \frac{\text{masse}}{\text{longueur} \times \text{diamètre}^2}$$

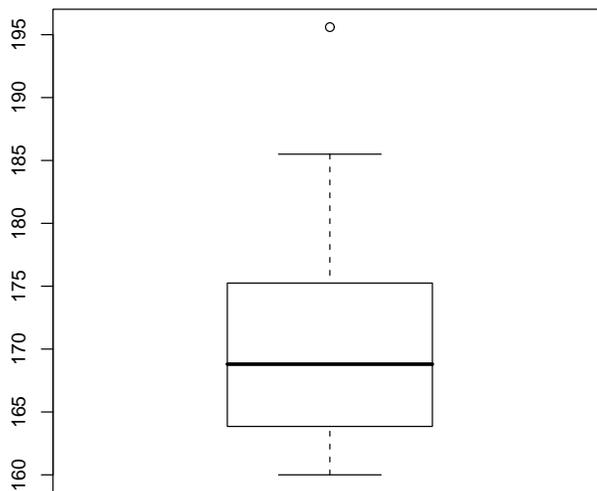
est proportionnel à la masse volumique du matériau qui constitue le javelot, expliquer le résultat de la question 4b.

Exercice A (AB. Dufour). On connaît la taille de 16 criminels de la fin du XIX^{ème} siècle en Grande-Bretagne. Les données sont les suivantes : la première ligne contient les valeurs de taille (en cm), la deuxième ligne les répétitions.

160	162.6	165.1	167.4	170.2	172.7	177.8	185.5	195.6
2	2	1	3	2	2	2	1	1

- (1) Donner la moyenne et la variance descriptives de la taille de ces 16 individus, taille rangée dans un vecteur `taicrim` (site pédagogique `m1data0809.txt`).
- (2) Quelle commande a permis d'obtenir la boîte à moustaches ci-dessous ?

Taille des 16 individus



- (3) Discuter cette représentation graphique.
- (4) Wright (1884) définit une valeur aberrante comme étant une valeur supérieure à trois fois l'écart-type (ici descriptif). Combien y-a-t-il d'individus dans cette distribution dont la valeur de la taille semble aberrante ?
- (5) Donner l'intervalle de confiance au niveau 0.95 de la taille des criminels dans la Grande-Bretagne du XIX^{ème} siècle.

Exercice B (AB. Dufour).

Une **loi de Poisson** est une variable aléatoire désignant le temps d'attente entre deux événements, dont la caractéristique principale est l'égalité de la moyenne et de la variance. X varie de 0 à l'infini. Soit X définie par son unique paramètre $\lambda = 1$. Nous donnons les valeurs de 0 à 10.

```
dpois(0:10, 1)
```

```
[1] 3.679e-01 3.679e-01 1.839e-01 6.131e-02 1.533e-02 3.066e-03 5.109e-04
[8] 7.299e-05 9.124e-06 1.014e-06 1.014e-07
```

`ppois(0:10, 1)`

[1] 0.3679 0.7358 0.9197 0.9810 0.9963 0.9994 0.9999 1.0000 1.0000 1.0000 1.0000

(1) À l'aide des informations ci-dessus, calculer les probabilités :

- (a) $P(X = 2)$
- (b) $P(3 \leq X \leq 4)$
- (c) $P(X \leq 6)$
- (d) $P(X \geq 5)$

(2) On donne quelques valeurs liées à une **loi normale centrée réduite**.

`axex <- seq(-3, 3, by = 0.3)`

`axex`

[1] -3.0 -2.7 -2.4 -2.1 -1.8 -1.5 -1.2 -0.9 -0.6 -0.3 0.0 0.3 0.6 0.9 1.2

[16] 1.5 1.8 2.1 2.4 2.7 3.0

`pnorm(axex)`

[1] 0.001350 0.003467 0.008198 0.017864 0.035930 0.066807 0.115070 0.184060

[9] 0.274253 0.382089 0.500000 0.617911 0.725747 0.815940 0.884930 0.933193

[17] 0.964070 0.982136 0.991802 0.996533 0.998650

À l'aide des informations ci-dessus, calculer les probabilités :

- (a) $P(X \leq -2.4)$
- (b) $P(X > 2.8)$
- (c) $P(0.2 \leq X \leq 1.2)$

Exercice C (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 :

langue maternelle				
français		anglais		
	hommes	femmes	hommes	femmes
oui	265	406	40	65
non	319	264	20	22

- (1) Donner la table de répartition des réponses oui-non selon le sexe, quelle que soit la langue maternelle d'origine.
- (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 D (AB. Dufour).

Une étude porte sur l'atmosphère familiale lorsque la fratrie comprend soit un enfant présentant un désordre bipolaire (DBP), soit un enfant présentant une dépression récurrente profonde (DRE). Les données se présentent sous la forme d'une note allant de 0 (atmosphère sereine) à 5 (atmosphère très perturbée). Elles se trouvent sur le site pédagogique `m1apa0809.txt` sous la forme de deux vecteurs. Existe-t-il une différence d'atmosphère familiale selon la pathologie ?

(réponse argumentée dans le sens d'un rapport de recherche : statistiques élémentaires, tests, graphiques, ...)

Références

[Sne74] R.D Snee. Graphical display of two-way contingency tables. *The American Statistician*, 28(9–12), 1974.