Université Claude Bernard Lyon 1

M2IGAPAS (Semestre 3) — Statistiques
Session 1 — 6 Décembre 2010
Durée : 2 h.

Examen CCF2 de Statistiques

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

AVERTISSEMENT
L'ensemble des fichiers de données nécessaires pour cet examen ('`VictBudgL1.txt`', '`vagueappel.txt`', '`notesbacadmis.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 '`\\teraetu\Enseignants`'),
  - puis '`jerome.bastien`',
  - enfin sur '`M2IGAPAS\examen\CCF2`'.

Pour les utilisateurs de Rcmdr, on prendra bien garde à utiliser la version 2.7 de R et non la version 2.9 ; on trouvera cette version, comme d'habitude, en faisant "démarrer", puis "programmes", puis "R" puis "R 2.7". Si cette version n'est pas installée sur votre ordinateur, il faut le redémarrer !

Exercice 1.
On étudie le fichier de données '`VictBudgL1.txt`'. Pour l'ouvrir, il faut procéder comme dans le cas 2 page 155 de l'annexe J, c'est-à-dire :

- *Avec Rcmdr :*

  (1) Comme d'habitude aller dans le menu déroulant "Données" de Rcmdr, choisir l'option "Importer des données" puis "Depuis un fichier texte ou le presse-papier...". Dans la fenêtre de dialogue qui s'ouvre, donner un nom au jeu de données (à la place de Dataset, choisi par défaut), le nom du fichier texte sans extension, c'est-à-dire : '`VictBudgL1`'. *Il faudra choisir dans "séparateur de champs", "Autre" et dans le champ "Spécifiez", tapez point-virgule, c'est-à-dire ;* Laisser les autres champs avec les valeurs choisies par défaut.

  (2) Employer la fenêtre qui s'ouvre alors pour retrouver le fichier à importer ('`VictBudgL1.txt`').

- *Sans Rcmdr :*
  Vous utiliserez la commande vue en section D.2.2, mais en spécifiant la valeur de l'argument optionnel 'sep' égal à ";" (par défaut il vaut l'espace " ") en tapant :

```
VictBudgL1 <- read.table("VictBudgL1.txt", h = T, sep = ";")
```

Ce jeu de données contient neuf informations concernant les clubs de la Ligue 1 de football française pour les saisons 2003-2004 à 2007-2008 (qui sont notées sous la forme 'S3-4' à 'S7-8'). Il s'agit à la fois des résultats sportifs de ces clubs et de leurs budget (en milliers d'euros) ainsi que le logarithme de ce budget. Notons que 3 clubs n'ont pas communiqué leur budget en 2003-2004.
L'objectif général de cette étude est de relier les performances sportives aux budget des clubs.

(1) (a) Analyser la variable '`Budget`'.

  (b) La variable '`LogBudget`' est le logarithme (en base 10) de ces budgets. Analyser cette variable et commentez les différences !

(2) (a) Analyser graphiquement si les budgets (variable '`Budget`') dépendent de la saison (variable '`Saison`'). De quelle façon ?

  (b) Analyser si la variable'`LogBudget`' dépend de la saison.

(3) Analyser la relation entre la variable '`Victoire`' et la variable '`LogBudget`'.

(4) *Question facultative*

  (a) On va créer une nouvelle variable nommée '`Y`' qui sera la somme du nombre de victoires (variable '`Victoire`') et de la moitié du nombre de match nuls (variable '`Nul`'). On pourra procéder comme dans l'exercice E.13 de l'annexe E du cours.

  Quelle est la signification de cette variable '`Y`' ?

  (b) Analyser si cette variable dépend de la saison. Expliquer pourquoi en moyenne ce n'est absolument pas le cas (sauf pour l'année 2003-2004 où 3 clubs ne sont pas comptabilisés).

  (c) Realiser un graphique représentant, *pour chaque saison séparément*, la relation entre '`Y`' et le logarithme du budget des clubs.

  Attention

- *Avec Rcmdr :*
  Il faut utiliser le menu déroulant "Graphes" du Rcmdr, puis l'option "Nuage de points". Choisir dans le bouton "graphe par groupe" la variable "Saison" et décocher l'option "courbe de lissage".
- *Sans Rcmdr :*
  Il faut taper dans "Rgui" une suite de commandes du type

```
library(car)
scatterplot(Y ~ LogBudget | Saison, smooth = FALSE, data = VictBudgL1)
```

Représenter grossièrement ce graphique sur votre copie. Qu'observe-t-on ?

Exercice 2.

(1) On étudie une loi binomiale de paramètres $n = 10$ et $p = 0.01$.
Calculer les probabilités :
(a) $P(X \leq 2)$
(b) $P(X > 3)$
(c) $P(1 \leq X \leq 9)$
(d) $P(X \leq -1)$

(2) On étudie une loi normale centrée réduite (c'est-à-dire, de moyenne nulle et d'écart-type égal à 1).
Calculer les probabilités :
(a) $P(X \leq -2.4)$
(b) $P(X > 2.8)$
(c) $P(0.2 \leq X \leq 1.2)$

Exercice 3.

(1) (a) On étudie les moyennes des notes d'un groupe d'étudianst ayant eu le baccalauréat. Voir fichier '`notesbacadmis.txt`'. Calculer la moyenne et l'écart-type des notes et en déduire un intervalle de confiance au niveau de confiance 0.95 de la moyenne.
(b) Peut-on affirmer que la moyenne du groupe est supérieure à 10 ?
(c) Peut-on affirmer (dans 95 % des cas) que la population théorique dont proviendrait ce groupe est une loi normale de moyenne supérieure à 10 ?
(d) Conclure.

(2) Répondre aux 4 mêmes questions si on suppose cette fois-ci que l'on étudie un groupe de $n = 80$ étudiants et que seuls la moyenne $\mu = 11.2$ et l'écart-type $\sigma = 3.5$ sont connus.

Exercice 4.

Nous reprenons l'exercice 2.33 page 14 (chapitre 2) (*numéroté 2.29 page 14 dans la version papier distribuée en cours !*) du polycopié de cours, qui porte sur les vagues d'appels au téléphone.
On rappelle que l'on utilise alors une loi binomiale de paramètres $n = 200$ et $p = 0.6$.
Soucieux d'optimiser son nouveau centre téléphonique, installé a plus de 3000 km de la métropole, le directeur étudie les dernier résultats des vagues de sondages ; le fichier '`vagueappel.txt`' contient, pour les 40 dernières vagues d'appels de $n = 200$ personnes, la proportion et le nombre de personnes réellement contactées.

(1) (a) Analyser la variable '`proportion`'.

(b) Commentez !

(2) Rappelez le nombre moyen de personnes contactées !

(3) Peut-on garantir au directeur que ce nombre moyen de personnes contactées sera atteint dès la première vague d'appel ?

(4) *Question facultative*

(a) On suppose maintenant que la variable aléatoire suit une loi binomiale de paramètres $p = 0.6$ et n quelconque (plus nécessairement égal à 200). Pour $N \in \{0, ..., n\}$, rappeler comment calculer $P(X \geq N)$.

(b) Pouvez-vous déterminer une valeur de n garantissant que $P(X \geq 120) = NC$ pour $NC = 0.95$? Même question pour $NC = 0.999$

(c) Retrouver ces valeurs en supposant que n est grand et que l'on fait l'approximation usuelle :

$$\frac{\frac{X}{n} - p}{\sigma} \rightsquigarrow \mathcal{N}(0,1), \text{ où } \sigma = \sqrt{\frac{p(1-p)}{n}}.$$

On pourra montrer en introduisant le quantile q tel que

$$P(z \geq q) = NC$$

où z suit une loi normale centrée réduite, que l'on a

$$q = \frac{\frac{N}{n} - p}{\sigma}$$

puis que l'on obtient l'équation du second degré

$$M^2 p + q\sqrt{p(1-p)}M - N = 0.$$

(d) Conclusion !

Corrigé

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