Université de Strasbourg
UFR de Mathématique et Informatique
Strasbourg
Année 2008/2009

Master Économie-Gestion, spécialité Actuariat et gestion du risque $2^{\text{ème}}$ année
Master de Statistique et Applications $2^{\text{ème}}$ année

UE : Applications de la Statistique
Session de janvier 2009 - Durée 4 heures
Enseignants Responsables : F. Bertrand, A. Guillou et M. Maumy

---

Aucun document n'est autorisé. Seules les calculatrices (autres que téléphone portable) et sans imprimante sont autorisées.
Chacune des deux parties sera rédigée sur une copie séparée. Chaque réponse devra être justifiée précisément. En annexe sont donnés le journal et la sortie d'un traitement avec le logiciel R.

Partie I : Enquêtes et Sondages.

Exercice 1 :
Dans une population $U = \{1,2,3,4,5\}$ on considère le plan de sondage suivant :

$$\mathbb{P}\left[\{1,2,4\}\right] = \frac{1}{6}, \quad \mathbb{P}\left[\{1,2,5\}\right] = \frac{1}{6}, \quad \mathbb{P}\left[\{1,4,5\}\right] = \frac{1}{6},$$

$$\mathbb{P}\left[\{2,3,4\}\right] = \frac{1}{6}, \quad \mathbb{P}\left[\{2,3,5\}\right] = \frac{1}{6}, \quad \mathbb{P}\left[\{3,4,5\}\right] = \frac{1}{6}.$$

1. Calculer les probabilités d'inclusion d'ordre 1 notées π_i. Quelle(s) relations(s) vérifient-elles ?
2. Calculer les probabilités d'inclusion d'ordre 2 notées π_{ij}.
3. De quel plan de sondage s'agit-il ? Pour répondre à cette question, servez-vous des deux questions précédentes.

Exercice 2 :
On considère une population $U = \{1,2,3\}$, sur laquelle on définit le plan de sondage suivant :

$$\mathbb{P}\left[\{1,2\}\right] = \frac{1}{2}, \quad \mathbb{P}\left[\{1,3\}\right] = \frac{1}{4}, \quad \mathbb{P}\left[\{2,3\}\right] = \frac{1}{4}.$$

Y est une variable définie sur U, telle que : $Y_1 = Y_2 = 3$, $Y_3 = 6$ dont on veut estimer le total T_Y.

1. Calculer les probabilités d'inclusion d'ordre 1 notées π_i. Quelle(s) relation(s) vérifient-elles ?
2. Calculer les probabilités d'inclusion d'ordre 2 notées π_{ij}.
3. De quel plan de sondage s'agit-il ?

4. Donner la distribution de probabilité de l'estimateur de Horvitz-Thompson du total, noté $\widehat{T}_{Y,HT}$. Nous rappelons que $\widehat{T}_{Y,HT} = \sum_{k\in S} \frac{Y_k}{\pi_k}$.
5. Calculer la variance de cet estimateur de deux manières :
   (a) directement
   (b) en appliquant la formule du cours $\left(\mathbb{V}ar\left(\widehat{T}_{Y,HT}\right) = \sum_{k\in U}\sum_{l\in U} \frac{Y_k}{\pi_k}\frac{Y_l}{\pi_l}\Delta_{kl}\right)$,
   où $\Delta_{kl} = \pi_{kl} - \pi_k\pi_l$ et $\pi_{kk} = \pi_k$.

   Que constatez-vous ?
6. Donner la distribution de probabilité de l'estimateur de la variance de $\widehat{T}_{Y,HT}$, qui est égal à $\widehat{\mathbb{V}ar}\left(\widehat{T}_{Y,HT}\right) = \sum_{k\in S}\sum_{l\in S} \frac{Y_k}{\pi_k}\frac{Y_l}{\pi_l}\frac{\Delta_{kl}}{\pi_{kl}}$.
7. Est-ce que l'estimateur de la variance de $\widehat{T}_{Y,HT}$, introduit à la question précédente, est sans biais ?

<u>Partie II : Analyse des Données.</u>

Exercice 1 :
On dispose des précipitations mensuelles (en mm) (moyennes mensuelles calculées sur 30 ans) pour 34 villes de France. Les villes choisies recouvrent à peu près uniformément le territoire français. Les données sont fournies dans le tableau 1.

1. Décrire le jeu de données (nombre d'individus, nombre de variables, nature des variables)
2. Que pouvez-vous dire à partir des données centrées-réduites (tableau 3) ?
3. On veut effectuer une ACP sur ce jeu de données : quels sont les objectifs d'une telle analyse ?
4. Les variables ont été centrées et réduites avant l'analyse. La réduction était-elle indispensable ? Justifier.
5. Les tableaux 5, 6, 7, 8 et 9 donnent les PRINCIPAUX résultats de l'ACP sur les variables et les individus. Quelle est l'inertie expliquée par le premier axe de l'ACP ? Et par le premier plan ?
6. Quelles sont les villes qui contribuent le plus à la construction des deux premiers axes ? Que signifie une contribution importante ?
7. La figure 4 donne le graphe des individus de l'ACP. La figure 5 donne le graphe des variables. Interpréter les facteurs principaux de l'ACP (à l'aide du graphe des individus et de celui des variables).
8. À partir du cercle de corrélations, que pouvez-vous dire concernant les corrélations suivantes février-mars, février-juin ?
9. VRAI ou FAUX ? Si FAUX, corriger la phrase proposée.
   – Une ville pluvieuse en juillet est également pluvieuse en octobre.
   – La variable janvier est bien représentée sur l'axe 1.
   – La ville de Vichy a joué le rôle le plus important dans la construction de l'axe 2.
   – La coordonnée d'une variable sur un axe est un indicateur de sa qualité de représentation par l'axe.

Exercice 2 :
Nous désirons étudier sur un échantillon de 100 personnes les liens entre la variable qualitative