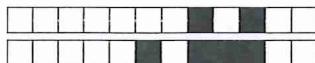


# Contrôle continu le 17 novembre

© Théo Jalabert

Test Blanc  
APP



+20/1/28+

|   |   |   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 |
| 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |
| 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 |
| 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 |
| 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 |
| 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 |
| 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 |

QCM

← Codez votre numéro d'étudiant ci-contre en noir cissant (ne pas entourer ni cocher) au stylo l'intérieur des cases correspondantes (colonne 1  $\Leftrightarrow$  1 chiffre, etc), et écrivez vos nom et prénom ci-dessous.

Nom et prénom :

.....

## M1 ISFA - Analyse de données Test blanc du 16/10/2017

Documents de cours et TD autorisés

Durée 1h15

Toutes les feuilles sont à rendre en fin d'épreuve.

Les questions faisant apparaître le symbole ☺ peuvent présenter une ou plusieurs bonnes réponses.  
Les autres ont une unique bonne réponse.

### Autour de la santé

Les données proviennent du site gouvernemental de la Santé (Sources - DREES, rétropélation des comptes de la santé - base 2005). Elles contiennent la part de la sécurité sociale (respectivement de la mutuelle) dans la consommation de soins et de biens médicaux pour les postes suivants : soins hospitaliers (SH), soins de ville (SdV), transports de malade (T), médicaments (M) et autres biens médicaux (ABM) sur une période s'étalant de 1950 à 2010.

**Question 1** Détaillez les commandes R suivantes :

```
> secumut <- read.table("SecuMut.txt", h=TRUE, dec=",", row.names=1)
> mut <- secumut[, 6:10]
> names(mut) <- c("SH", "SdV", "T", "M", "ABM")
```

? la colonne 1 contient le nom des lignes

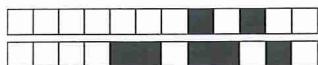
F P A B J

on lit le tableau en format txt, en précisant que les colonnes ont des entêtes, que les nb décimales sont écrit avec des virgules  
On garde seulement les colonnes 6 à 10 auxquels on donne des noms.

**Question 2** On construit le graphe des boîtes à moustaches sur les différents postes liés à la part des mutuelles (Cf figure 1). Au vu de ce graphique, quel type d'analyse en composantes principales vous paraît le plus judicieux ?

- ACP centrée, réduite       ACP centrée, non réduite  
 C ACP non centrée, non réduite

/



+20/3/26+

**Question 5** Explicitez le critère de Kaiser dans le cas de données non-réduites :

F  P  A  B  J

La valeur propre doit être supérieur à l'inertie totale du nuage de pts divisé par le nb de variable

$$\lambda \geq \frac{\text{Tr}}{p} = \frac{\sum \text{vp}}{p}$$

≥ aux moy

**Question 6** On réalise une analyse en composantes principales sur les postes liés à la part des mutuelles. Les valeurs propres obtenues sont les suivantes :

[1] 3.62861933 0.91933369 0.30188241 0.10399598 0.04616859

Cochez toutes les réponses correctes :

- A Selon le critère adouci de Kaiser, deux axes sont à conserver.
- B L'inertie projetée sur l'axe 2 est égale à 18,4%.
- C L'inertie totale est égale à 32.482.
- D Aucune de ces réponses n'est correcte.

**Question 7** Détaillez les commandes R suivantes et le résultat produit par la dernière commande :

```
> acp<-dudi.pca(mut, scannf=FALSE, nf=5)
> attach(acp)
> eig[eig>=1]
```

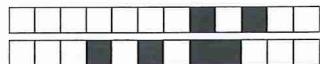
F  P  A  B  J

On réalise une acp centrée réduite, en gardant 5 axes ; sans afficher le graphique des vp. attach(acp) nous permet d'appeler les elem d'acp sans préciser à chaque fois acp\$

Affiche les vp  $\geq 1$ , vecteur - /

**Question 8** Nous considérons la représentation simultanée des lignes et des colonnes, donnée par la figure 2. Cochez toutes les réponses correctes :

- A L'axe 2 est un axe de temporalité.
- B On observe un effet taille des postes dans la part liée aux mutuelles.
- C Les parts des soins hospitaliers et des transports sont corrélées sur le premier plan factoriel.
- D Aucune de ces réponses n'est correcte.

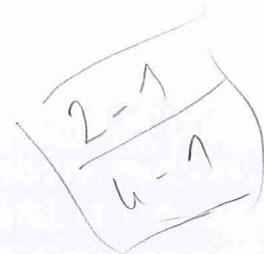


(+20/5/24+)

| catégorie            | jour | nuit |
|----------------------|------|------|
| piétons              | 258  | 210  |
| bicyclettes          | 117  | 82   |
| cycles $\leq 50cm^3$ | 73   | 62   |
| cycles $> 50cm^3$    | 433  | 281  |

(N-1/K-1)

TABLE 20 – Sécurité routière



**Question 9** Nous disposons des informations analogues pour les postes dans la part liée à la sécurité sociale. Nous réalisons sur la figure 3 les représentations graphiques pour quatre postes. Les courbes en trait plein représentent les données centrées réduites pour les mutuelles ; les courbes en pointillés représentent les données centrées réduites pour la sécurité sociale. Cochez toutes les réponses correctes :

- A Le poste transport se compense entre la part de la sécurité sociale et la part de la mutuelle.
- B Centrer et réduire les variables permet de comparer l'évolution temporelle.
- C La part de la sécurité sociale sur les soins hospitaliers augmente au cours du temps.
- D Aucune de ces réponses n'est correcte.

## Exercice

La table 20 présente le nombre de tués sur les routes en France Métropolitaine, pour l'année 2015 (Site de la Sécurité Routière) selon 4 catégories d'usagers et selon le moment de la journée.

**Question 10** Proposez des commandes pour créer une table de contingence contenant les valeurs et les noms des modalités rassemblés dans la table 20.

 F  P  A  B  J

matix  
names

**Question 11** Citez deux fonctions permettant de construire une représentation graphique de la table 20 :

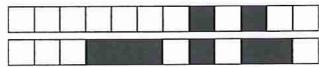
 F  P  A  B  J

braiplot  
hist

ballonplot

**Question 12** Considérons le test du  $\chi^2$ . Quel est le degré de liberté associé à la table ?

- A 6
- B 2
- C 3
- D 1
- E 8



+20/7/22+

**Question 18 ☺** Cochez toutes les réponses correctes :

- Afin d'avoir une idée des couples de modalités sur-représentés - par rapport à l'hypothèse d'indépendance des deux variables - dans les données, la routine `scatter` est bien adaptée.
- À partir de notre tableau, nous pourrions facilement construire le tableau de Burt correspondant à nos deux variables.
- Si nous disposons du tableau disjonctif complet sur nos deux variables, nous pourrions facilement construire la table de contingence pour notre couple de variables.
- Afin d'avoir une idée des couples de modalités sur-représentés - par rapport à l'hypothèse d'indépendance des deux variables - dans les données, la routine `mosaicplot` est bien adaptée.
- La fusion de deux modalités d'une même variable conduit à une perte d'inertie.
- Les modalités des variables sont pondérées par l'inverse de leurs fréquences d'apparition.
- Aucune de ces réponses n'est correcte.*