

**Correction sujet d'examen - Janvier 2020 - Durée 1h30****1 ACP**

1. Combien de dimensions retiendriez-vous pour analyser ces données ?

RÉPONSE 2PTS. En observant le graphique donnant les pourcentages d'inertie exprimés par chaque dimension (ou axe), on remarque :

- une décroissance rapide jusqu'à la dimension 3,
- une décroissance beaucoup plus lente à partir de la dimension 3,
- l'inertie cumulée sur les 2 premiers axes représentent 84.6% de l'inertie totale.

Cela indique que les 2 premières dimensions reflètent des variations beaucoup plus importantes que les axes suivants. On retiendra pour l'analyse les 2 premiers axes.

Quel pourcentage d'inertie est associé au plan factoriel (1, 2) ?

RÉPONSE 1PT. Pour répondre à cette question, il suffit d'additionner les pourcentages d'inertie associés à chacun de ces 2 premiers axes :  $59.763 + 24.852 = 84,615$ . Le plan factoriel (1,2) représente environ 84.6% de l'inertie totale.

2. Quelle est la variable la mieux représentée sur l'axe 1 ?

RÉPONSE 1PT. La qualité de représentation sur l'axe 1 d'une variable correspond au cosinus carré de l'angle formé par cette variable et sa projection sur cet axe, soit la quantité `cos2` associée à `Dim.1` dans les sorties numériques concernant les variables. La variable la mieux représentée est celle qui correspond à la plus grande quantité `cos2` qui vaut ici 0,966. Donc la variable la mieux représentée est `InterBC`

Quel est l'individu le moins bien représenté sur l'axe 2 ?

RÉPONSE 1PT. La qualité de représentation sur l'axe 2 d'un individu correspond au cosinus carré de l'angle formé par le vecteur représentant cet individu et sa projection sur cet axe, soit la quantité `cos2` associée à `Dim.2` dans les sorties numériques concernant les individus. L'individu le moins bien

représenté est celui qui correspond à la plus petite quantité **cos2** qui vaut ici 0. Donc l'individu le moins bien représenté est le mammifère n°9.

3. Quelle est la corrélation entre la variable **HeightB** et l'axe factoriel 1 ?

RÉPONSE 1PT. Lorsqu'on représente les variables dans le cercle des corrélations, leurs coordonnées s'interprètent comme la corrélation entre chacune de ces variables et l'axe factoriel correspondant. Ici, on peut lire dans la colonne **Dim.1** la valeur -0.684 correspondant à la variable **HeightB**. Ceci indique que  $Corr(\text{HeightB}, F_1) \simeq -0.684$ ; la variable **HeightB** est négativement corrélée avec l'axe 1.

Quelle est la variable la plus corrélée avec l'axe factoriel 2 en valeur absolue ?

RÉPONSE 1PT. Pour déterminer la variable la plus corrélée avec l'axe 2 en valeur absolue, il suffit de chercher dans la colonne **Dim.2** la quantité la plus proche de 1 en valeur absolue soit 0,945. La variable correspondante étant **HeightA**, c'est donc cette variable qui est la plus corrélée avec l'axe 2.

4. Un commentaire indique que les molaires des 11 individus représentés à droite sur l'axe 1 possèdent en moyenne un angle AB (numéroté 10 sur le schéma présentant les 18 mesures) plus petit que celui des 11 individus représentés à gauche de l'axe 1. Êtes-vous d'accord avec cette affirmation ?

RÉPONSE 2PTS. D'après les résultats concernant les variables, la variable **AngleAB** est fortement positivement corrélée (0.937) avec l'axe 1. Autrement dit, plus l'abscisse d'un mammifère sera grande (i.e. vers la droite), plus ce mammifère présentera un angle AB important. L'affirmation est donc fausse.

5. Préciser la méthode permettant de représenter les deux modalités *Kuehneotherium* et *Morganucodon*.

RÉPONSE 2PTS. La modalité *Kuehneotherium* est positionnée au barycentre des mammifères appartenant à cette espèce (i.e. moyenne des coordonnées des individus appartenant à cette espèce); de même, la modalité *Morganucodon* est positionnée au barycentre des mammifères relevant de cette espèce. La Figure 2 traduit cette méthode de représentation.

6. Que représente la figure 2 ?

RÉPONSE 1PT. La Figure 2 représentent les mammifères sur le plan principal (1,2); un code couleur a été utilisé pour distinguer les deux espèces en

présence (noir pour les *Kuehneotherium* et rouge pour les *Morganucodon*).

7. Quelle est la variable qui contribue le plus à la construction de l'axe 1 ?

RÉPONSE 1PT. Il suffit de regarder la variable qui possède la quantité **ctr** la plus élevée pour **Dim.1**. On obtient 8.979; c'est donc la variable correspondante, **InterBC**, qui contribue le plus à la construction de l'axe 1.

Quelle est la variable qui contribue le plus à la construction de l'axe 2 ?

RÉPONSE 1PT. Il suffit de regarder la variable qui possède la quantité **ctr** la plus élevée pour **Dim.2**. On obtient 19.956; c'est donc la variable correspondante, **HeightA**, qui contribue le plus à la construction de l'axe 2.

8. Quel axe sépare le mieux les deux espèces de mammifères ?

RÉPONSE 1PT. Les abscisses des mammifères séparent parfaitement les deux espèces, ce qui n'est absolument pas le cas du point de vue des ordonnées. C'est donc l'axe 1 qui sépare le mieux les deux espèces de mammifères.

Quelle propriété doivent vérifier les variables qui distinguent le mieux les deux espèces de mammifères ?

RÉPONSE 1PT. Les variables qui distinguent le mieux les deux espèces de mammifères sont celles qui sont le plus corrélées avec l'axe 1.

9. Écrire un bref commentaire résumant les principales différences entre ces deux espèces de mammifères.

RÉPONSE 2PTS. D'après la remarque précédente et au regard des coordonnées des variables sur l'axe 1 (i.e corrélations avec l'axe 1), les principales mesures qui différencient le mieux ces deux espèces (en prenant par exemple les corrélations plus grandes que 0.9) sont **LengthA**, **LengthC**, **TotalLength**, **AngleAB**, **AngleAC**, **Boffset**, **InterAC**, **InterBC** et **CuspAngle**.

10. Quel axe sépare le mieux les deux mammifères 4 et 12 ?

RÉPONSE 1PT. D'après la représentation graphique des mammifères sur le plan principal, on remarque que la différence entre les ordonnées de ces deux mammifères est de l'ordre de 8 alors qu'elle est seulement d'environ 4 du point de vue des abscisses. C'est donc l'axe 2 qui sépare le mieux ces deux mammifères.

Écrire un bref commentaire résumant les principales différences entre ces deux mammifères.

RÉPONSE 2PTS. D'après la remarque précédente, il suffit d'examiner les variables les plus corrélées avec l'axe 2 pour obtenir les caractéristiques les plus discriminantes entre ces deux mammifères. Au regard des coordonnées des variables sur l'axe 2, les principales mesures qui différencient le mieux ces deux mammifères (en prenant par exemple les corrélations plus grandes que 0.9) sont `HeightA` et `HeightC`.

## 2 AFC et AFCM

### 2.1 Analyse Factorielle des Correspondances

1. Combien d'axes sont nécessaires pour représenter la totalité de la relation entre ces deux variables ?

RÉPONSE 1PT. En regardant les sorties fournissant les inerties (i.e. **Eigenvalues**), on constate que deux dimensions suffisent pour obtenir la totalité de l'inertie.

2. Sans faire de calculs, ordonner de la plus petite à la plus grande, les 3 quantités suivantes :  $d_{\chi^2}(\text{know\_well}, \text{POUR})$ ,  $d_{\chi^2}(\text{know\_well}, \text{CONTRE})$ ,  $d_{\chi^2}(\text{know\_well}, \text{ABS})$ .

RÉPONSE 1PT. On peut lire ces distances directement à partir de la représentation simultanée des modalités puisqu'ici le plan (1,2) représente la totalité de l'inertie. On constate que

$$d_{\chi^2}(\text{know\_well}, \text{POUR}) < d_{\chi^2}(\text{know\_well}, \text{CONTRE}) < d_{\chi^2}(\text{know\_well}, \text{ABS})$$

Que pouvez-vous en déduire sur les européens favorables au revenu universel ?

RÉPONSE 2PTS. Au regard de la remarque précédente, il semble que la part des individus qui connaissent bien le concept de revenu universel soit plus importante chez les européens enclin à voter POUR comparativement à ceux qui y sont plutôt défavorables ; cette différence est encore plus marquée avec ceux qui s'abstiennent.

3. À partir des résultats de l'AFC :  
— l'axe 1 est-il le lieu d'une association exclusive ?

RÉPONSE 1PT. Pour que l'axe 1 soit le lieu d'une association exclusive, il faut que l'inertie associée soit égale à 1. Or ici on en est loin puisque

l'inertie associée à l'axe 1 est égale à 0.106. Conclusion : non, l'axe 1 n'est pas le lieu d'une association exclusive.

- quelle est l'intensité de la liaison entre ces 2 variables ?

RÉPONSE 1PT. D'après le cours, l'intensité de la liaison entre deux variables catégorielles peut être mesurée à l'aide du  $\Phi^2 = \text{inertie totale} = 0.106 + 0.007 = 0.113$

- quelle est la modalité de la variable **vote** qui contribue le plus à l'axe 1 ?

RÉPONSE 1PT. Il suffit de chercher la plus grande valeur dans la colonne **ctr** associée à Dim.1 soit 84,402. C'est donc la modalité **ABS** de la variable **vote** qui participe le plus à la construction de l'axe 1.

- quelle est la modalité de la variable **awareness** qui contribue le plus à l'axe 1 ?

RÉPONSE 1PT. Il suffit de chercher la plus grande valeur dans la colonne **ctr** associée à Dim.1 soit 74,656. C'est donc la modalité **know\_nothing** de la variable **awareness** qui participe le plus à la construction de l'axe 1.

4. À partir de la représentation simultanée des modalités, commentez brièvement la relation entre ces 2 variables catégorielles.

RÉPONSE 3PTS. En lisant les proximités entre les modalités représentées, on peut dire que 1) la part des européens ne connaissant rien au concept de revenu universel est plus importante chez les abstentionnistes, 2) la part des européens connaissant bien le sujet est plus importante parmi ceux qui y sont favorables, 3) on remarque que  $d_{\chi^2}(\text{know\_little}, \text{CONTRE})$  est plus petite que  $d_{\chi^2}(\text{know\_little}, \text{POUR})$ , ce qui indique que la proportion d'européens ayant une faible connaissance du concept de revenu universel serait plus importante chez ceux qui s'y opposent.

## 2.2 Analyse Factorielle des Correspondances Multiples

1. Est-ce que toutes les variables ont participé à la construction des axes factoriels ?

RÉPONSE 1PT. Les sorties numériques **Supplementary categorical variables (eta2)** nous indiquent qu'il y a deux variables supplémentaires : **country** et **age\_group**.

Ces deux variables ne participent pas à la construction des axes factoriels.

2. Que vaut l'inertie moyenne ?

RÉPONSE 1PT. D'après le cours, l'inertie moyenne est l'inertie totale divisée par le nombre total d'axes qui est encore égale à  $1/p$  où  $p$  est le nombre total de variables participant à la construction des axes. Ici  $p = 7$  ce qui implique que l'inertie moyenne  $= 1/7 \simeq 0,14$ .

3. Dans la représentation des variables catégorielles, quelle est l'abscisse de la variable **job** ?

RÉPONSE 1PT. L'abscisse de la variable **job** se lit dans les sorties

**Categorical variables (eta2)** à l'intersection entre la colonne **Dim.1** et la ligne **job** soit 0,548.

Que représente cette quantité ?

RÉPONSE 1PT. 0.548 représente le rapport de corrélation au carré entre la variable **job** et les coordonnées des individus sur l'axe 1.

4. Donner deux pays pour lesquels on observe une différence marquée relativement à la variable **job**.

RÉPONSE 1PT. Il suffit de prendre deux pays éloignés relativement à l'axe 1 puisque la variable **job** possède un rapport de corrélation assez grand avec l'axe 1. Par exemple, à partir de la représentation graphique des modalités, on peut dire que SI (i.e. Slovénie) et RO (Roumanie) sont susceptibles de posséder une différence marquée du point de vue de la variable **job**.

5. Donner deux pays pour lesquels on observe une différence marquée relativement à la variable **vote**.

RÉPONSE 1PT. D'après la Figure 3, la variable **vote** possède un rapport de corrélation au carré non négligeable avec le facteur 2. Ceci est confirmé en lisant dans les sorties numériques **Categorical variables (eta2)** à l'intersection entre la colonne **Dim.2** et la ligne **vote** soit 0,454. On va donc chercher deux pays éloignés selon l'axe 2 ; on peut prendre par exemple CZ (i.e. République tchèque) et CY (i.e. Chypre).

6. À partir des représentations graphiques, qu'est-ce qui caractérise principalement les individus qui voteraient en faveur (i.e. **POUR**) du revenu universel de ceux qui y sont opposés (i.e. vote **CONTRE**) ?

RÉPONSE 3PTS. D'après la représentation des modalités, on peut dire que 1) la part des européens bien informés sur le revenu universel est plus importante chez ceux qui y sont favorables que chez ceux qui y sont opposés, 2) la part des hommes semble plus importante chez les européens qui sont pour le revenu universel, 3) la part des européens vivant dans un milieu urbain est plus importante chez ceux qui sont favorables au revenu universel. De plus, il semble que la proportion de jeunes soit plus grande chez les européens favorables au revenu universel, ce qui est corroboré par la proximité de la modalité 14\_25 avec POUR (ce qui peut expliquer aussi la proximité de POUR avec child\_no). Enfin, on peut remarquer que le plan (1,2) représente pas plus d'environ 24% de l'inertie totale et que les sorties fournies ne permettent pas de connaître la qualité de représentation des différentes modalités. Il faut donc être prudent sur ces interprétations qui peuvent être confirmés ou infirmés par des enquêtes plus ciblées.