|  |  |
| --- | --- |
| Examen  Semestre : 1 2  Session : Principale Rattrapage | logo-esprit |
| Module : Factor Analysis  Enseignante :Zouhour Hammouda  Classe(s) : 4-DS | |
| Documents autorisés  : OUI NON Nombre de pages : 8  Calculatrice autorisée  : OUI NON Internet autorisée : OUI NON | |
| Date : 05/01/2018 Heure:9h Durée : 1H:30 | |

**Note :** les parties sont indépendantes. On s’attachera à justifier autant que possible les réponses, tout en évitant de faire du « remplissage » ; la note tiendra compte de la qualité de la rédaction.

**Partie 1 : Analyse en composantes principales**

Il s’agit d’une enquête sur les budgets-temps (temps passé dans différentes activités au cours de la journée). Le tableau suivant comprend 10 variables quantitatives et 4 variables catégorielles. Les 10 variables quantitatives sont le temps passé en : PROFession, TRANsport, MENAge, ENFAnts, COURses, TOILette, REPAs, SOMMeil, TELE et LOISirs. Les temps sont en centièmes d’heure et le total d’une ligne (sur ces 10 variables quantitatives) est 2400 (24 heures). Les 4 variables qualitatives sont :

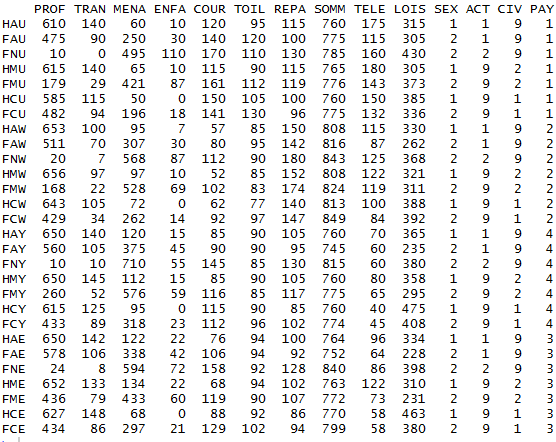
— le SEXe (1=Hommes, 2=Femmes),

— l’ACTivité (1=Actifs, 2=Non Act , 9=Non précisé),

— l’état CIVil (1=Célibataires, 2=Mariés, 9=Non précisé),

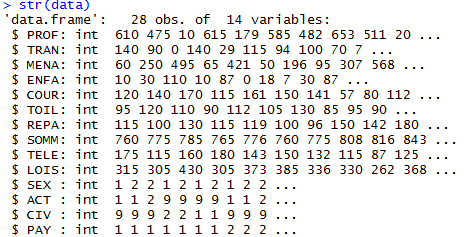
— le PAYs (1=USA, 2=Pays de l’Ouest, 3=Pays de l’Est, 4=Yougoslavie).

Par exemple, La première case en haut à gauche du tableau (HAU) indique que les Hommes Actifs des USA passent en moyenne 6 heures et 6 minutes (6 heures + 10/100 d’heure) en activité PROFessionnelle. Le total d’une ligne (sur ces 10 variables quantitatives) est 2400 (24 heures).



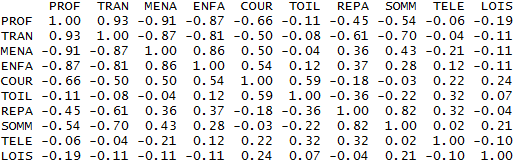
**Figure 1: Jeu de données « budget-temps »**

**Question 1** : A partir de la description citée ci-dessus et en examinant les figures 1 et 2, faire les modifications nécessaires aux données avant d’effectuer une ACP (expliquer théoriquement et donner la commande R appropriée)



**Figure 2: Description des variables**

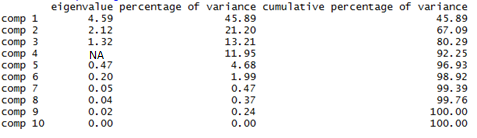
**Question 2** :



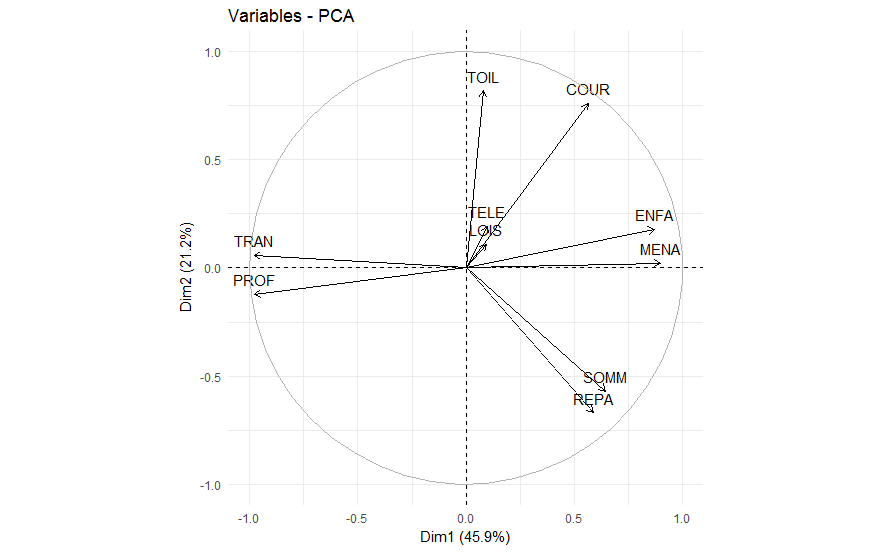
**Figure 3: ...............................**

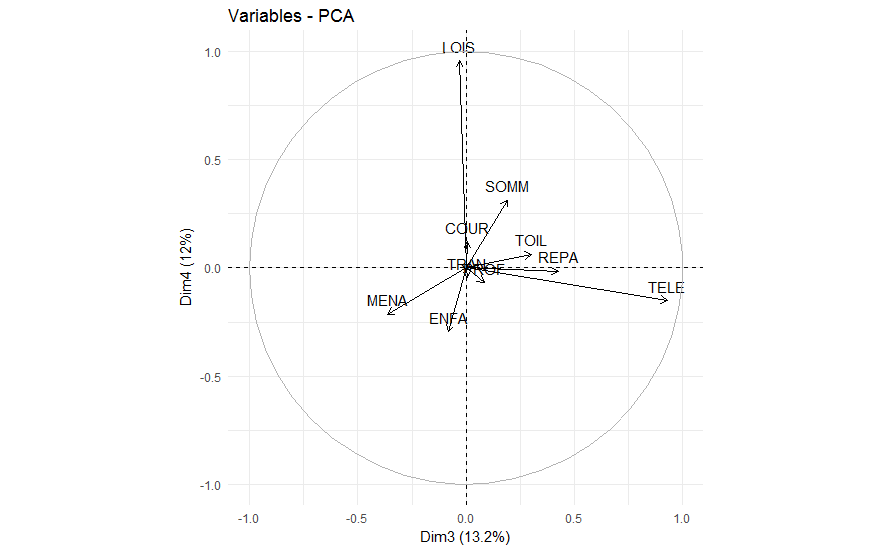
1. Nommer la figure 3
2. Donner la commande R qui nous fournit cette sortie
3. Interpréter les corrélations entre les variables ? Commenter les valeurs remarquables.
4. Justifier l'utilisation d'une ACP et décrire son objectif dans ce cas.
5. Y a-t-il des variables dans le jeu de données qui doivent être illustratives?

**Question 3** :



**Figure 4 :…………………………………..**

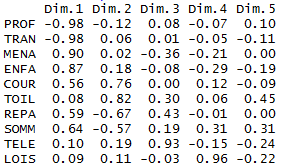
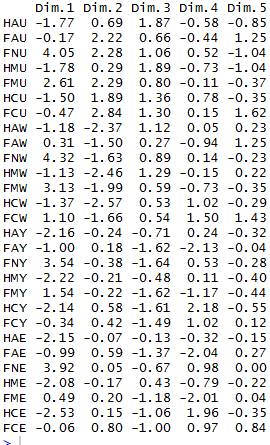
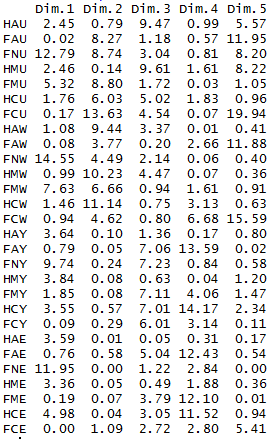


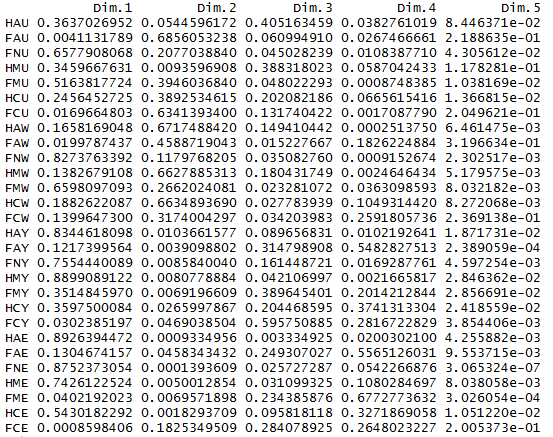


**Figure 5 :…………………………………..**

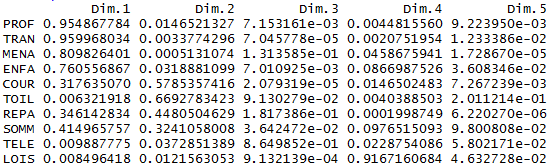
**Corrélations variables/composantes**

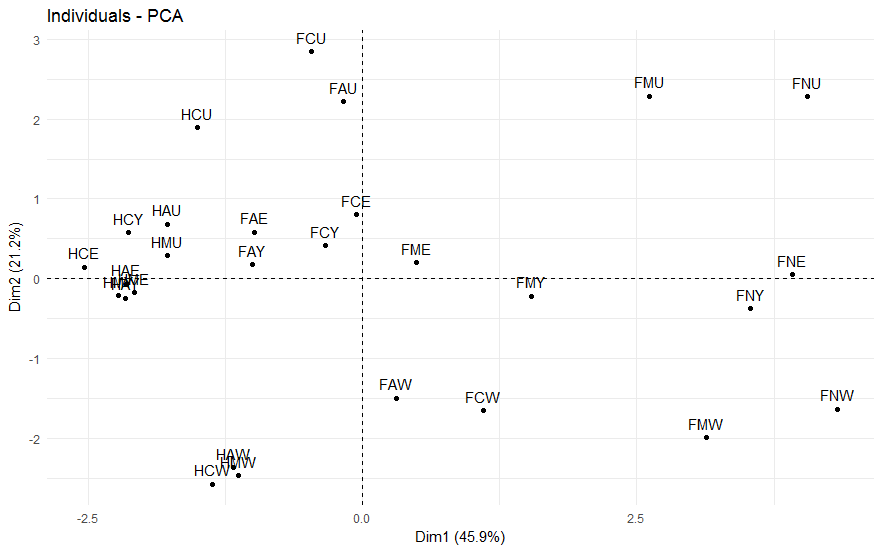
**Coordonnées des individus Contributions des individus**

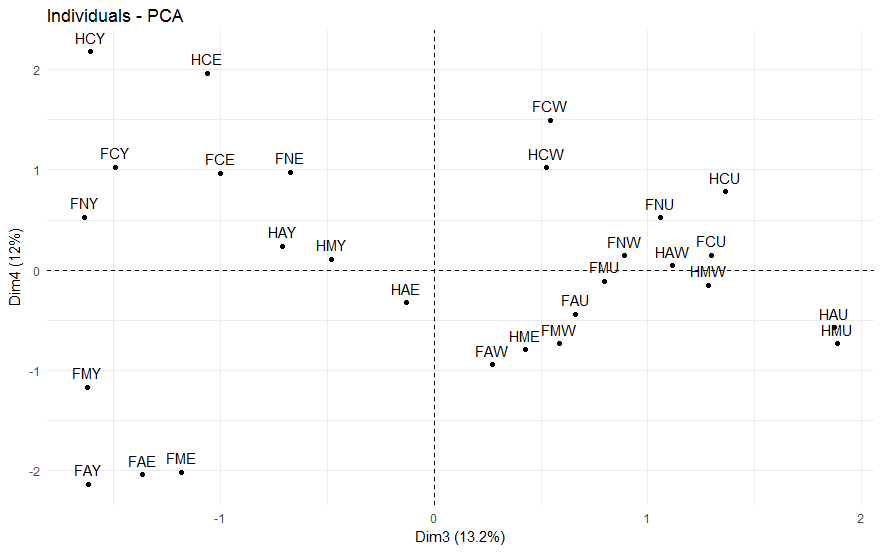
   
 **Qualité de représentation des individus sur les composantes**



**Qualité de représentation des variables**







**Figure 6 : la projection des individus sur le premier plan principal (en haut) et sur le plan (3,4) (en bas)**

1. Nommer les figures 4 et 5
2. Calculer la valeur propre manquante NA dans la figure 4.Justifier théoriquement ce calcul.
3. Commenter la répartition de l'inertie. Combien d'axes principaux retient-on? Justifier votre réponse
4. Proposer une autre méthode afin de déterminer le nombre d’axes principaux à retenir.
5. Pour chacune des 4 premières composantes principales, donner la liste des individus qui contribuent à l’axe de manière significative.
6. En utilisant la description des noms des individus donnée en introduction et à partir uniquement des sorties numériques et graphiques des individus (figure 6), donner une première interprétation des axes (au moins des deux premiers).
7. Interpréter minutieusement la figure 5.
8. Peut-on interpréter toutes les variables ? justifier votre réponse
9. En croisant ces résultats avec ceux des individus, donner une nouvelle interprétation des axes.
10. Est-il intéressant de refaire l’analyse en considérant certains individus comme supplémentaires ? Justifier votre réponse.

**Partie 2 : Analyse de la variance à un seul facteur**

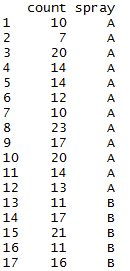
Les insecticides sont des [substances actives](https://fr.wikipedia.org/wiki/Substance_active_d%27un_produit_phytopharmaceutique) ou des [préparations phytosanitaires](https://fr.wikipedia.org/wiki/Pr%C3%A9paration_phytopharmaceutique) ayant la propriété de tuer les [insectes](https://fr.wikipedia.org/wiki/Insecte), leurs larves et/ou leurs œufs. Six (A,B,C,D,E,F) insecticides (spray) ont été testés chacun sur 12 cultures. La réponse observée (count) est le nombre d'insectes.

**Question 1 :** Rappeler le principe de l’analyse de la variance à un seul facteur (ANOVA).Dire à quoi sert cette méthode dans notre cas

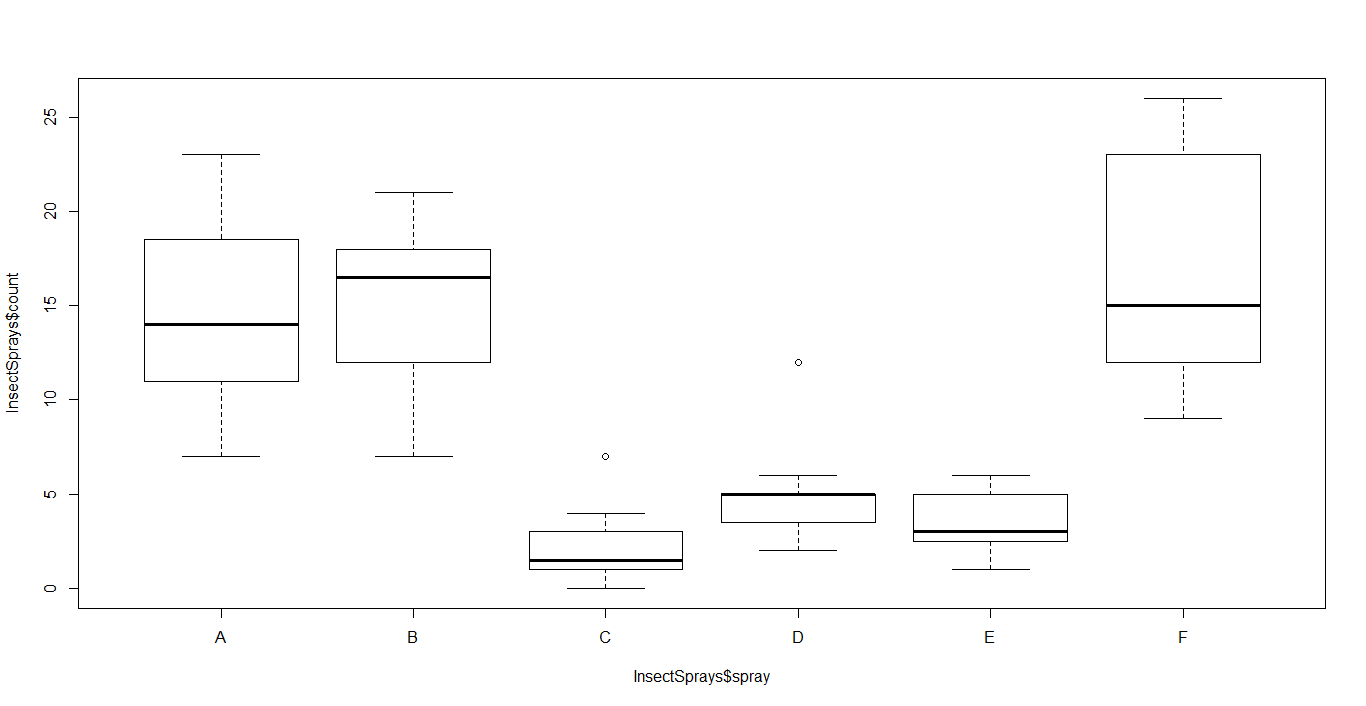
**Question 2 :** Nommer la figure 8.Donner la commande R qui permet de donner ce graphe. Interpréter.

**Question 3 :** Rappeler les conditions d’applicationde l’ANOVA.

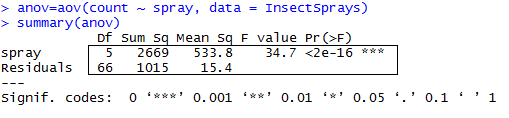
**Question 4 :** Interpréter le résultat de l’ANOVA. Donner la signification de toutes les valeurs encadrées dans la figure 9

****

**Figure 7 : Extrait du jeu de données**



**Figure 8 :………………………………….**



**Figure 9 : Résumé du test ANOVA**

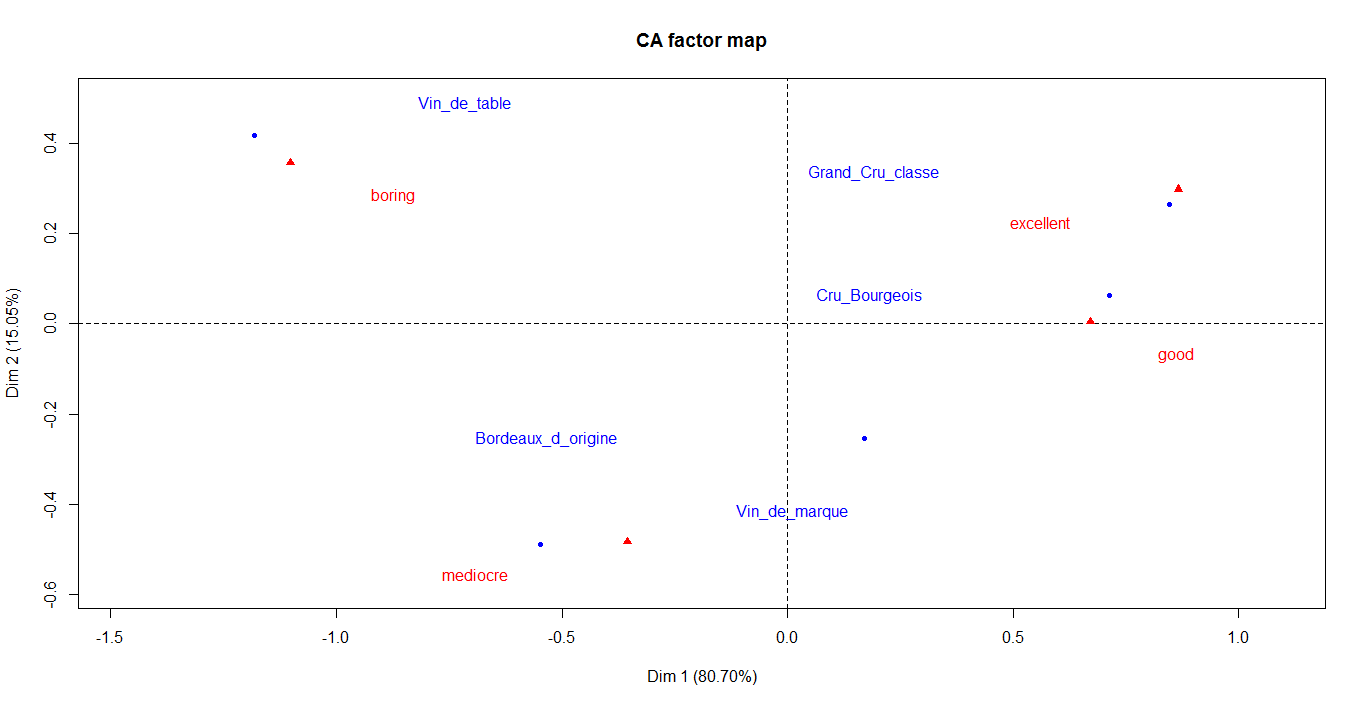
**Partie 3 : Analyse factorielle des correspondances**

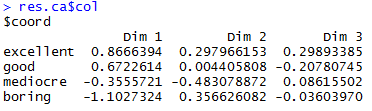
Le jeu de données bordeaux donne l'avis de 200 juges dans une dégustation à l'aveugle de cinq différents types de vin rouge à bordeaux (sud-ouest de la France)

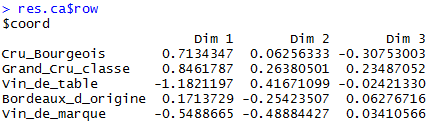
**Question 1 :** D’une manière générale, définir un tableau de contingence. En calculant les totaux des lignes et des colonnes. Décrire ce tableau de données

**Question 2 :** Que représente les variances expliquées par les deux premiers axes dans notre cas ( et )

**Question :** En utilisant le graphe et les sorties numériques ci-dessous, donner une première interprétation du premier plan factoriel.



****

****

**Bonne chance ☺**