

## Examen Semestriel Analyse Exploratoire de Données Correction

### Exercice 1 (10 pts):

- a) Combien de variables peut-on trouver dans le tableau ? Sont-elles qualitatives ou quantitatives ?

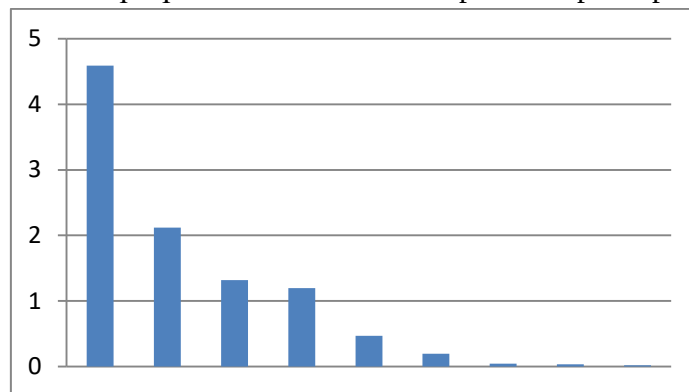
*On trouve dans le tableaux 10 variables quantitatives (ou numériques)*

- b) Calculer la valeur propre manquante (X.XXXX).

*On sait que le total des valeurs propres est égal au nombres des variables, c-à-d 10. D'où :*

$$X.XXXX = 10 - (4.5887 + 1.3210 + 1.1953 + 0.4684 + 0.1990 + 0.0468 + 0.0371 + 0.0239) = 10 - 7.8802 = 2.1198$$

- c) Tracer l'histogramme des valeurs propres. Combien de composantes principales faut-il retenir?



*La règle de Kaiser conduit à retenir les valeurs propres qui sont supérieures à 1, ce qui signifie ici les 4 premières.*

- d) Quel est le pourcentage d'inertie totale expliquée par le sous-espace principal correspondant?

$$(4.5887 + 2.1198 + 1.3210 + 1.1953)/10 = 92.25 \%$$

- e) Si on ne retient que le premier plan principal, quel est le pourcentage expliqué?

$$(4.5887 + 2.1198)/10 = 67.09 \%$$

- f) Pour chacune des 2 premières composantes principales, donner 3 individus qui déterminent l'axe de manière significative.

*On s'intéresse aux coordonnées les plus grandes en valeur absolue.*

*Pour l'axe 1 : FNW(4.3234), FNU (4.0334), FNE (3.9187)*

*Pour l'axe 2 : FCU (2.8443) négatif, HCW (2.572) positif, HMW (2.4639) positif*

### Exercice 2 (10 pts):

- a) La statistique du khi-deux vaut 8371.28. Quel est le degré de liberté de cette statistique ? Que peut-on conclure ?

*ddl = (10-1)(6-1) = 9 x 5 = 45. Selon le tableau du Khi-deux, et avec une erreur  $\alpha=0.05$ , la statistique Khi-deux est entre 63.691 et 76.154. La statistique du khi-deux des données (égale à 8371.28) est supérieure à celle du tableau. On peut conclure que la cause du suicide chez le citoyen de l'Allemagne de l'Ouest est dépendante de son sexe et son âge.*

- b) Quel pourcentage de l'inertie est expliquée par les deux premiers axes factoriels ?

*L'inertie totale est la somme des valeurs propres, elle est égale à 0.17097. Le pourcentage de l'inertie expliquée par les deux premiers axes factoriels est :  $(0.10328 + 0.05634)/0.17097 = 93.36\%$*

- c) Quel point-ligne est le moins bien représenté dans ce plan ? C'est le point-ligne F10A20 (0.8428)

- d) Les points-colonnes « suffocation » et « pendaïson » affichent de plus grandes contributions dans le deuxième axe factoriel que dans le premier. Qu'est-ce que cela signifie ?

*Que ces deux modes de suicide contribuent davantage à l'interprétation du deuxième axe factoriel que les autres.*