

## Corrigé de l'examen de statistique.

## Exercice – I.V.G.

1 La tranche d'âge pour la quelle on observe la plus faible proportion d'IVG est celle des 45-49 ans, quelque soit l'année. On observe cela dans le tableau des pourcentages colonnes.

2. On observe une augmentation de la proportion d'IVG, en 1998 pour la tranche 45-49 ans ; en 1999 pour la tranche 18-19 ans.

3. Première analyse multidimensionnelle

a) Le tableau sur lequel porte l'analyse est un tableau de contingence. Il s'agit donc d'une analyse des correspondances.

b) Le  $\Phi^2$  du tableau est égal à 0.00134. Le tableau analysé a 8 lignes et 11 colonnes. La plus petite dimension du tableau est donc 8. En excluant la première valeur propre « triviale », il reste donc 7 axes possibles. Si tous les axes contribuaient également, cette contribution moyenne serait égale à  $\Phi^2/7$ , soit  $0.00134/7=0.00019$ . On retiendra donc les axes pour lesquels la valeur propre est supérieure à 0.00019, soit les deux premiers axes, ce qui représente 83% de la variance totale.

c) Pour les tranches d'âge on a retenu celles ayant une contribution forte au premier axe : soit une contribution supérieure à  $1000/8=125$ . le premier axe oppose les 15-17 ans, 18-19ans et 45-49 ans aux 25-29 ans et 30-34 ans.

On pourrait dire que le premier axe oppose les tranches d'âge extrême (très jeunes et plus âgées) aux tranches d'âge intermédiaire.

d) Sur la base du même critère, le second axe oppose la tranche d'âge 45-49 ans à la tranche d'âge 18-19 ans.

e) L'axe 2 est essentiellement constitué par la tranche d'âge 45-49 ans. Celle-ci a une contribution relative de  $799/1000$ , soit quasiment 80%. Sur le graphique on voit que cette tranche d'âge est très excentrée. Il faut remarquer que c'est la tranche d'âge pour laquelle il y a le moins d'IVG. C'est donc une tranche d'âge atypique, pour ce qui concerne le nombre d'IVG au cours de la période considérée.

f) On interprétera les années sur la base du même raisonnement que celui appliqué aux tranches d'âge. Comme il y a 11 années, on retiendra les années qui ont une contribution supérieure à  $1000/11=91$ . Ainsi le premier axe oppose l'année 1990 aux années 1998, 1999 et 2001. Le deuxième axe oppose l'année 1998 aux années 1999 et 2001. On remarque, sur le graphique, que l'année 1998 est très excentrée, contribuant fortement aussi bien à l'axe 1 qu'à l'axe 2. L'année 1998 est l'année où pour la tranche des 45-49 ans on observe un doublement du nombre d'IVG.

4. Deuxième analyse multidimensionnelle : la classification.

a) La partition en trois classes correspond à un taux de variance inter égal à 72.5%. Les trois classes sont les classes 13, 11 et 8.

b) La classe 13 correspond aux tranches d'âge 20-24 ans, 25-29 ans, 30-34 ans, 35-39 ans et 40-44 ans.

La classe 11 correspond aux plus jeunes (15-17 ans et 18-19 ans) et la classe 8 n'est autre que la tranche des 45-49 ans.

c) En regardant le graphique des tranches d'âge de l'analyse des correspondances, on retrouve les trois classes, avec la tranche 45-49 ans en haut à gauche du plan 1-2, les deux tranches jeunes (15-17 ans et 18-19 ans) en bas à gauche et les tranches intermédiaires groupées autour de l'origine. Comme le plan 1-2 représente 83% de la variance du nuage, il est normal de retrouver les trois classes, assez bien distinguées dans ce plan.

5. Troisième analyse multidimensionnelle.

a) En ne retenant que les tranches d'âge ayant une forte contribution, l'axe 1 oppose les « jeunes » (15-17 ans et 18-19 ans) à la tranche 25-29 ans. L'axe 2, lui, oppose les 20-24 ans aux 25-29 ans. Ces deux dernières tranches sont celles qui ont les plus forts effectifs d'IVG.

b) De façon un peu plus marquée que dans la première analyse, l'axe 1 peut être considéré comme l'axe du temps. Il ordonne presque complètement les années de 1990 à 2001. Il oppose les années 1990 et 1994 aux années 1999 et 2001, c'est-à-dire qu'il oppose le début de la décennie à la fin de la décennie. L'axe 2 oppose les années 1990 et 1999 aux années 1995 et 2001.

c) Cette deuxième analyse, dans laquelle la tranche 45-49 ans a été mise en élément supplémentaire, apporte une vision plus claire. Les tranches d'âge sont plus dispersées dans le plan 1-2, et l'opposition entre les jeunes et les intermédiaires est plus marquée.

En fait, seuls deux axes ont été retenus dans la première analyse. Mais on remarquera que le troisième axe de la première analyse est devenu le deuxième axe de la deuxième analyse. Le fait d'avoir évacué la tranche 45-49 ans a renforcé la structure du tableau et le plan 1-2 représente 89% de la variance totale.

