

**Remarques sur l'article "Une histoire de discrétisation" par G. Celeux et C. Robert.**

M. Roux  
Université Marseille 3

1°) L'effectif total de l'échantillon est faible : 23 individus. Il est clair que le découpage en 3 classes peut provoquer des mélanges malencontreux. Personnellement je suis arrivé, empiriquement, à la conclusion qu'on ne doit pas descendre en dessous d'une dizaine d'individus par classe... encore que ceci devrait être pondéré par le nombre de variables.

2°) Dans les données de morphométrie animale l'âge des individus influe fortement sur leur taille. Il convient donc d'éliminer ce facteur taille avant toute chose. Cela est bien visible sur la figure 5, résultant de l'ACP sur les rapports à la taille de l'aile. J'ai refait le découpage en classes sur ces rapports, puis j'ai procédé à l'Analyse des Correspondances.

Le résultat est le suivant :

- Axe 1 : les points 7, 10, 12, 13, 17, 18, 19 et 21 sont tous confondus à une extrémité de l'axe, et tous les autres sont aussi confondus mais proches de l'origine.
- Axe 2 : à une extrémité on a les points 2, 4, 6, 8, 9, 14 et 16, avec seulement deux coordonnées distinctes; à l'autre extrémité on a les points 1, 5, 11, 15, 20, 22 et 23, également en deux paquets de points confondus.

Par rapport aux groupes observés en ACP, il n'y a donc qu'une seule "erreur" de classement, à savoir le point 22, erreur qui peut certainement être attribuée au découpage en classes.

3°) C'est vrai que, dans le cas de petits effectifs, le découpage en classes a tendance à brouiller les données; mais il est possible d'atténuer cet effet en utilisant un "codage flou" (où les indicatrices de classes ne sont pas strictement en 0 ou 1, mais peuvent avoir des valeurs intermédiaires entre 0 et 1, pour les individus proches des bornes des classes). J'ai essayé cela sur les données brutes mais l'amélioration n'est pas éclatante.

4°) Je suis d'accord avec les auteurs pour dire que le type de découpage (Nombre de classes, effectifs ou intervalles égaux) ne change pas grand'chose au résultat.

En conclusion je pense qu'il serait souhaitable, peut-être à l'aide de simulations basées sur un modèle multivarié simple, d'établir des critères précis pour le choix du nombre de classes. Et si de tels critères ne sont pas satisfaits on devrait s'abstenir de traiter les données de cette façon.