

Annexe 20. Analyse des correspondances multiples et classification ascendante hiérarchique

L'analyse des correspondances multiples (ACM) et la classification ascendante hiérarchique (CAH) sont deux des méthodes descriptives de l'analyse multivariée, qui comprend également des méthodes explicatives. Outre l'ACM et la CAH, les méthodes descriptives comportent l'analyse en composante principale (ACP) et l'analyse factorielle des correspondances (AFC) ; les méthodes explicatives se composent de la régression multiple, de l'analyse discriminante et de la segmentation.

L'analyse multivariée est utilisée pour synthétiser et expliquer l'information donnée par plusieurs variables, à les résumer en minimisant la déperdition d'information. Le but des différentes méthodes qui la composent est de mettre en lumière des dimensions cachées contenues dans les réponses aux variables sélectionnées.

L'analyse des correspondances multiples

L'Analyse des Correspondances Multiples est une extension de l'Analyse Factorielle des Correspondances (AFC) pour l'étude simultanée de plusieurs variables. Elle peut être considérée comme une AFC sur un tableau de Burt – ce dernier étant un tableau de contingence symétrique qui croise en un seul tableau toutes les modalités de chaque caractère les unes avec les autres.

L'objectif des méthodes factorielles, dont fait partie l'ACM, est de déterminer un sous-espace de dimension réduite, « compréhensible » par l'oeil, sur lequel projeter un nuage de variables. Ces méthodes constituent donc l'extension de la statistique descriptive au cas multidimensionnel et autorisent une analyse simultanée de plusieurs variables. Pour obtenir un tel sous-espace, la méthode consiste à chercher, dans un premier temps, l'axe sur lequel le nuage se déforme le moins en projection c'est-à-dire qui restitue la part la plus importante de l'inertie totale du nuage ou autrement dit de l'information qu'il contient. Une fois ce premier axe déterminé, il s'agit alors de chercher le second axe sur lequel le nuage se déforme le moins, tout

en étant orthogonal au premier. Il suffira ensuite de réitérer le processus jusqu'à restituer l'intégralité de l'inertie du nuage. L'analyse des correspondances multiples permet d'observer les combinaisons et les liens entre les variables actives, c'est un outil visuellement puissant et très intuitif. La visualisation de plans factoriels dans l'espace des variables permet d'analyser l'information contenue sur deux axes factoriels.

La classification ascendante hiérarchique

La classification ascendante hiérarchique (CAH) est une des méthodes de classification qui a pour objectif d'obtenir des classes d'individus les plus cohérentes possibles : soit en constituant les groupes les plus homogènes ou, à l'inverse, les groupes qui se distinguent le plus les uns des autres. C'est la raison pour laquelle, la mesure communément utilisée pour juger de la qualité d'une CAH est le quotient de l'inertie inter-classe sur l'inertie totale du nuage des individus : plus elle est élevée et plus les individus regroupés forment des groupes homogènes nettement différenciés les uns des autres. Il faut donc arbitrer entre nombre restreint de groupes et qualité des classifications, c'est à dire entre interprétabilité et fiabilité de l'information.