

Analyse de données
M1 Statistique et économétrie - 2012-2013
V. Monbet
Analyse discriminante décisionnelle

Un médecin cherche à mettre au point un moyen de prévoir si un enfant risque de développer un problème de surpoids lors de ses prochaines années de vie. Pour cela il a relevé l'indice de masse corporelle de 13 garçons de huit ans. Trois ans plus tard, six des garçons sont en surpoids, tandis que ça n'est pas le cas des sept autres. Les enfants en surpoids sont regroupés dans le groupe A, tandis que les enfants ne souffrant pas de surpoids sont regroupés dans le groupe B. Le tableau suivant répertorie l'indice de masse corporelle des 13 enfants :

groupe A	21.7	20.4	17.0	17.3	19.3	20.0	
groupe B	16.2	15.2	13.4	16.8	11.7	12.9	13.9

On suppose que la proportion d'enfants en surpoids dans la population totale est de 20%.

Dans un premier temps, on suppose de plus que, dans les deux populations (surpoids/non surpoids), la variable *indice de masse corporelle* (IMC) suit une loi normale et que les avariances dans les deux sous population sont égales à 3.4.

1. Estimer la moyenne des IMC de chacun des groupes.
2. On suppose que si un enfant a un IMC de moins de 16, on le classe dans le groupe B. Calculer le *risque de Bayes* associé à cette règle (i.e. la probabilité de se tromper).
3. Déterminer la *règle de décision de Bayes* associée à l'IMC. Par définition, la règle de Bayes est la règle qui minimise la probabilité de mauvais classement. Classifier un enfant dont l'IMC est égal à 17.5 suivant cette règle.
4. Calculer le risque de Bayes associé à cette nouvelle règle de décision. Que remarque t'on ?
5. On suppose maintenant que la proportion des enfants en surpoids dans la population totale est égale à 1/8. Estimer la règle de Bayes associée. Selon cette règle, enfant dont l'IMC est égal à 17.5 doit-il être suivi par un diététicien ?
6. On considère que l'erreur qui consiste à ne pas détecter un enfant qui développera un surpoids est 1.5 fois plus coûteuse (en terme de santé publique) que celle qui consiste à classer à tort un garçon dans le groupe des enfants à risque. Estimer la règle de Bayes associée à cette matrice de coût. Selon cette règle, enfant dont l'IMC est égal à 17.5 doit-il être suivi par un diététicien ?
7. L'hypothèse selon laquelle les variances dans les 2 groupes sont égales est abusive. Supposons que les variances estimées sont de bonnes approximations des vraies variances. Donner la nouvelle règle de Bayes. Que devient la décision concernant l'enfant dont l'IMC est égal à 17.5 ?
8. Comparer les risques de Bayes associés aux différentes règles de décision énoncées ci-dessus.