

IV. ANALISIS DE DATOS (A. Moragas y J.M. Rivera-Pomar)

Sea cual fuere el método de análisis de imagen utilizado, el resultado será una matriz de datos, cuyas filas corresponderán a cada caso estudiado y las columnas a las variables o características evaluadas. Su estudio requerirá esencialmente métodos de análisis estadístico multivariado^{1, 2}. Diversos modelos de aparatos de análisis de imagen poseen ya programas de esta índole; en caso contrario cabe acudir a un centro de cálculo, o lo más aconsejable a nuestro juicio es recurrir a un ordenador personal, lo que permite el "diálogo" del investigador con la máquina. Los listados de los programas en Fortran o en Basic figuran en diversos libros³⁻⁶ y son fácilmente implementables. La estrategia del análisis estadístico dependerá evidentemente del propósito del estudio, por lo que sólo es posible aquí dar una idea esquemática.

1. Simplificación de los datos. En muchas ocasiones se valoran de entrada múltiples características de la imagen. Es evidente que gran parte de esta información será redundante y muchos de los datos superfluos, de ahí que proceda reducir el número de variables tomando solo las más representativas, lo cual se desconoce evidentemente a priori. Una técnica adecuada es la de análisis de los componentes principales de aquélla, a través del cálculo de los valores y vectores propios de la matriz de datos. El resultado será una ordenación por orden de significación de las variables transformadas, que nos permitirá su selección.

2. En el caso de que el propósito del estudio sea la diferenciación entre grupos preestablecidos (p. ej. células benignas y malignas, grados de atipia, etc.) cabe recurrir a dos métodos alternativos y en parte complementarios. Así puede establecerse un "índice intermedio" a partir del coeficiente de ambigüedad de Genchi y Mori, que va de 1 (indicando ambi-

güedad total) a 0 (perfecto valor discriminante), y el valor d' o medida de la detectabilidad⁷, que va de 0 a 0.5 (perfecta capacidad para detectar si un caso pertenece a uno u otro grupo).

El denominado análisis discriminante permite una selección de las variables en función de su grado de significación para establecer una diferencia "diagnóstica" entre grupos preestablecidos. En último término la función discriminante obtenida nos permitirá una representación gráfica bidimensional de los casos, evidenciando su separación y agrupamiento.

3. No debe confundirse el análisis discriminante con el de agrupamiento (clustering) o taxonómico, extensamente utilizado en ciencias biológicas. En este caso no se conocen unos criterios previos de clasificación y el objetivo del estudio es determinar, ante una población de casos si ésta es homogénea o por el contrario está formada por grupos definidos.

4. Para estudiar la correspondencia existente entre las variables, y de que forma se distribuyen los casos en función de éstas cabe recurrir al análisis factorial de correspondencias⁸. Podemos imaginar aquí la matriz de datos como una especie de nube de puntos en un espacio multidimensional, tratándose de hallar los distintos planos que corten esta nube y que traduzcan el máximo de separación entre las variables o los casos, representados gráficamente en cada plano. La valoración de los resultados se efectúa teniendo en cuenta la proximidad de las variables, o por el contrario su oposición en el eje de coordenadas (ejes de inercia) que definen el plano.

5. Mediante el análisis de regresión se pretende determinar de qué forma una variable (representada en el eje de las ordenadas) depende de otra, representada en las abscisas, que se toma como explicativa (p. ej. presión arterial y

edad). En ocasiones la dependencia no viene dada por una línea recta sino por una curva de distinto orden y que puede ser compleja. Así en un estudio evolutivo de la grasa parda ajustamos una curva de tipo logístico basada en tres parámetros⁹.

En el caso de la regresión múltiple se trata de determinar cómo una variable depende de otras varias que se supone que actúan al mismo tiempo (p. ej. grado de fibrosis hepática observada en una segunda biopsia, y diversas variables como esteatosis, colestasis, componente inflamatorio, etc.). Un problema especial es el que plantean los datos binarios (sobrevivencia o fallecimiento, recidiva o no recidiva). En este caso se suele recurrir al denominado método de Cox, extensamente utilizado en la valoración de factores pronósticos, tanto clínicos como morfológicos.

Es conveniente advertir, antes de aplicar cualquier método estadístico, que en ocasiones éste presupone una distribución "normal" de las variables, condición que habrá de determinarse previamente mediante diversos test.

REFERENCIAS:

1. Chatfield, C. y Collins, A.J.: Introduction to multivariate analysis. Chapman y Hall, Londres 1980.
2. Mardia, K.V.; Kent, T.J. y Bibby, J.M.: Multivariate analysis. Academic Press, Londres 1979.
3. Lebart, L.; Moriman, A. y Fénelon, J.P.: Traitement des données statistiques. Methodes et programmes. Dunod, Paris 1982.
4. Foucart, T.: Analyse factorielle. Programation sur microordinateurs. Masson, Paris, 1982.
5. Foucart, T. y Lafaye, J.Y.: Régression lineaire sur microordinateurs. Masson, Paris, 1983.

6. Foucart, T.: Analyse factorielle de tableaux multiples= Masson, Paris, 1984.
7. Bartels, P.H.: Numerical evaluation of cytologic data. III. Selection of features for discrimination. Analyt Quantit Cytol. 3: 153-159, 1979.
8. Greenacre, M.J.: Theory and applications of correspondence analysis. Academic Press, Londres, 1984.
9. Moragas, A. y Toran, N.: Prenatal development of brown adipose tissue in man. Biol Neonate, 43: 80-85, 1983.