

## COMENTARIS DE LLIBRES

G.A.F. SEBER

MULTIVARIATE OBSERVATIONS

J. Wiley, 686 pp., New York

Año: 1984.

Comentario hecho por: CARLOS M. CUADRAS

Aunque G.A.F. Seber no es un especialista en el tema, la obra que comentamos es, junto con otras del mismo autor, una notable contribución a la estadística, que destaca por la amplitud de los temas tratados, por la buena exposición didáctica, y por la actualidad de la bibliografía utilizada.

El primer capítulo es una introducción al tema, presentando las notaciones utilizadas, el tipo de datos y el campo de aplicación del análisis multivariante, y algunos de los criterios de decisión (punto de vista clásico) en procesos de inferencia multivariante.

El segundo capítulo, aborda algunas de las distribuciones multivariantes más usuales, con especial atención a la normal y relacionadas, sobre todo la distribución de Wishart. A destacar la síntesis conseguida acerca de los valores propios asociados a matrices de dispersión muestrales.

El tercer capítulo se dedica a estudiar los diferentes problemas de estimación y contraste de hipótesis relativos a una o varias poblaciones normales. Dividido en tres partes, (inferencia sobre la media, inferencia sobre la matriz de covarianzas, comparación de poblaciones normales), expone con profundidad todos los casos importantes, con abundantes referencias bibliográficas y relacionando los temas entre sí.

Un cambio importante de orientación se refleja en el capítulo cuarto. En él se exponen aquellas técnicas de representación gráfica, que podríamos llamar informales (caras de Chernoff, curvas de Andrews, representaciones mediante árboles, etc.). También son presentados algunos tests de normalidad, como

el que está basado en la transformación de Box-Cox, incluyendo también una breve pero interesante referencia a la estimación robusta y al problema de las observaciones -- "outliers".

Los métodos multivariantes clásicos propiamente dichos aparecen en el capítulo cinco. El análisis de componentes principales está bien expuesto, que es presentado como una técnica de reducción de la dimensión, y cuyas propiedades y generalizaciones son detalladamente estudiadas. Debo destacar que es el primer libro donde se explica el método biplot para representar simultáneamente variables e individuos, que está relacionado con el análisis de correspondencias, aunque de este tema apenas se hace mención. El análisis factorial recibe un tratamiento clásico, es decir, un enfoque exploratorio, no confirmatorio, con una referencia pobre al tema de las rotaciones. El tema del análisis de escalas multidimensionales ("multidimensional scaling") se inicia con el clásico teorema de Torgerson, que da origen al análisis de coordenadas principales (caso métrico). El caso no métrico es tratado bastante informalmente, pero con numerosos ejemplos, en contraste con el análisis "procrustes" que el autor detalla bien. También es correcta la exposición del análisis de correlación canónica y sus aplicaciones (representación de poblaciones a través de ejes canónicos).

El capítulo seis, sobre análisis discriminante, es uno de los más completos. Se centra únicamente en el problema de asignar un individuo a uno de entre varios grupos conocida cierta información multivariante. En el caso de variables continuas el autor combina

hábilmente las diferentes posibilidades que pueden presentarse (dos grupos, distribución conocida, parámetros desconocidos, caso de normalidad, varios grupos, criterio Bayes, criterio minimax, etc.). También estudia el caso discreto. Otra novedad de esta obra es el tratamiento bastante completo de la discriminación logística. El análisis cluster se explica en el capítulo siete. No hay una presentación demasiado formal, lo que sería posible en el caso jerárquico (por ejemplo, utilizando las propiedades de la geometría ultramétrica, que ayudarían a entender el porque de algunas de las técnicas conocidas). Sin embargo, hay abundantes comentarios, -- ejemplos y llamadas a referencias bibliográficas.

Los capítulos ocho y nueve son los que el autor explica mejor. No en vano G.A.F. Seber es autor de diversos libros sobre modelos lineales, de los cuales aquí expone la versión multivariante. Desde luego, el capítulo ocho se apoya en el segundo y tercero. Constituye un estudio completo sobre los modelos lineales multivariantes (estimación de parámetros, test de hipótesis, intervalos de confianza simultáneos, etc.).

El siguiente capítulo es una aplicación al tema clásico del análisis de la varianza y covarianza en su versión multivariante, dedicando también una sección al análisis de curvas de crecimiento.

En el capítulo diez y último encontramos diversas técnicas matemáticas (descomposición de Cholesky, descomposición singular, diagonalización simétrica generalizada, etc.) que tienen interés como técnicas auxiliar en análisis multivariante, junto con otros tópicos especiales, difíciles de clasificar, como el análisis de datos incompletos.

Cuatro apéndices sobre álgebra matricial, -- proyecciones ortogonales, algunas distribuciones usuales y tablas estadísticas, completan la obra.

Resumiendo, se trata de una excelente obra, totalmente recomendable a iniciados en el tema. Destacan especialmente los capítulos -- que permiten ser abordados mediante un tratamiento algebraico y geométrico. Una buena co

lección de problemas añade más interés a la obra, así como las oportunas referencias a paquetes de programas relacionados con el análisis multivariante (ALLOC, ALSCAL, BMDP, CLUSTAN, GENSTAT, etc.).