



CLASIFICACIÓN DE MÉTODOS MULTIVARIADOS

Juan Carlos Flores L.

1.-INTRODUCCION

La investigación de la gestión 2005 sobre la **Clasificación de métodos del Análisis Multivariante**, es con el fin de contribuir con la comunidad estudiantil y todos lo que tengan interés en conocer estos métodos que es de gran interés en las aplicaciones reales en el análisis de datos en las distintas disciplinas, estos métodos pueden ser aplicados en ciencias, administración de empresas, ingeniería, psicología, área salud, en el área sociológico, económico, la industria, centros de investigación de ámbito universitario, etc.

El trabajo de investigación consta de Capitulo I que es la introducción, el Capitulo II desarrolla los métodos de dependencia, el Capitulo III desarrolla métodos de Interdependencia, luego las Conclusiones, Recomendaciones y la Bibliografía.

El Análisis Multivariante es el conjunto de métodos estadísticos cuya finalidad es analizar simultáneamente conjuntos de datos multivariantes en el sentido de que existen varias variables medidas para cada individuo ú objeto estudiado. Su razón de ser radica en un mejor entendimiento del fenómeno objeto de estudio obteniendo información que los métodos estadísticos univariantes y bivariantes no lo podrían conseguir. Estos métodos hacen posible plantear preguntas específicas y precisas de considerable complejidad en marcos idóneos, lo que posibilita llevar a cabo investigaciones teóricamente significativas y evaluar los efectos de las variaciones paramétricas ocurridas de forma natural en el contexto en que normalmente ocurren. De esta forma, se pueden preservar las correlaciones naturales entre las múltiples influencias sobre el comportamiento y se pueden estudiar estadísticamente los efectos aislados de esas influencias sin provocar el típico aislamiento de esos individuos o variables.

2.- Definición del análisis multivariante

El análisis multivariante es un conjunto de técnicas de análisis de datos en expansión. Entre las técnicas más conocidas se tiene a (1) regresión múltiple y correlación múltiple; (2) análisis discriminante múltiple; (3) componentes principales y análisis factorial común; (4) análisis multivariante de varianza y covarianza; (5) correlación canónica; (6) análisis cluster; (7) análisis multidimensional y (8) análisis conjunto. Entre las técnicas emergentes también incluidas están (9) análisis de correspondencias; (10) modelos de probabilidad lineal como logit y probit; y (11) modelos de ecuaciones



simultáneas/estructurales. (*En este resumen se, desarrolla cada una de las técnicas multivariantes, definiendo brevemente la técnica y el objetivo de su aplicación.*)

Los objetivos del Análisis Multivariante pueden sintetizarse en dos:

- i) Proporcionar métodos cuya finalidad es el estudio conjunto de datos multivariantes que el análisis estadístico uni y bidimensional es incapaz de conseguir.
- ii) Ayudar al analista o investigador a tomar decisiones óptimas en el contexto en el que se encuentre teniendo en cuenta la información disponible por el conjunto de datos analizado.

3.-Tipos de Técnicas Multivariantes

Se pueden clasificar en tres grandes grupos (ver grafica 1):

Métodos de dependencia

Suponen que las variables analizadas están divididas en dos grupos: las variables dependientes y las variables independientes. El objetivo de los métodos de dependencia consiste en determinar si el conjunto de variables independientes afecta al conjunto de variables dependientes y de qué forma.

Métodos de interdependencia

Estos métodos no distinguen entre variables dependientes e independientes y su objetivo consiste en identificar qué variables están relacionadas, cómo lo están y por qué.

Métodos de dependencia

Se pueden clasificar en dos grandes subgrupos según que la variable (s) dependiente (s) sea (n) cuantitativas o cualitativas.

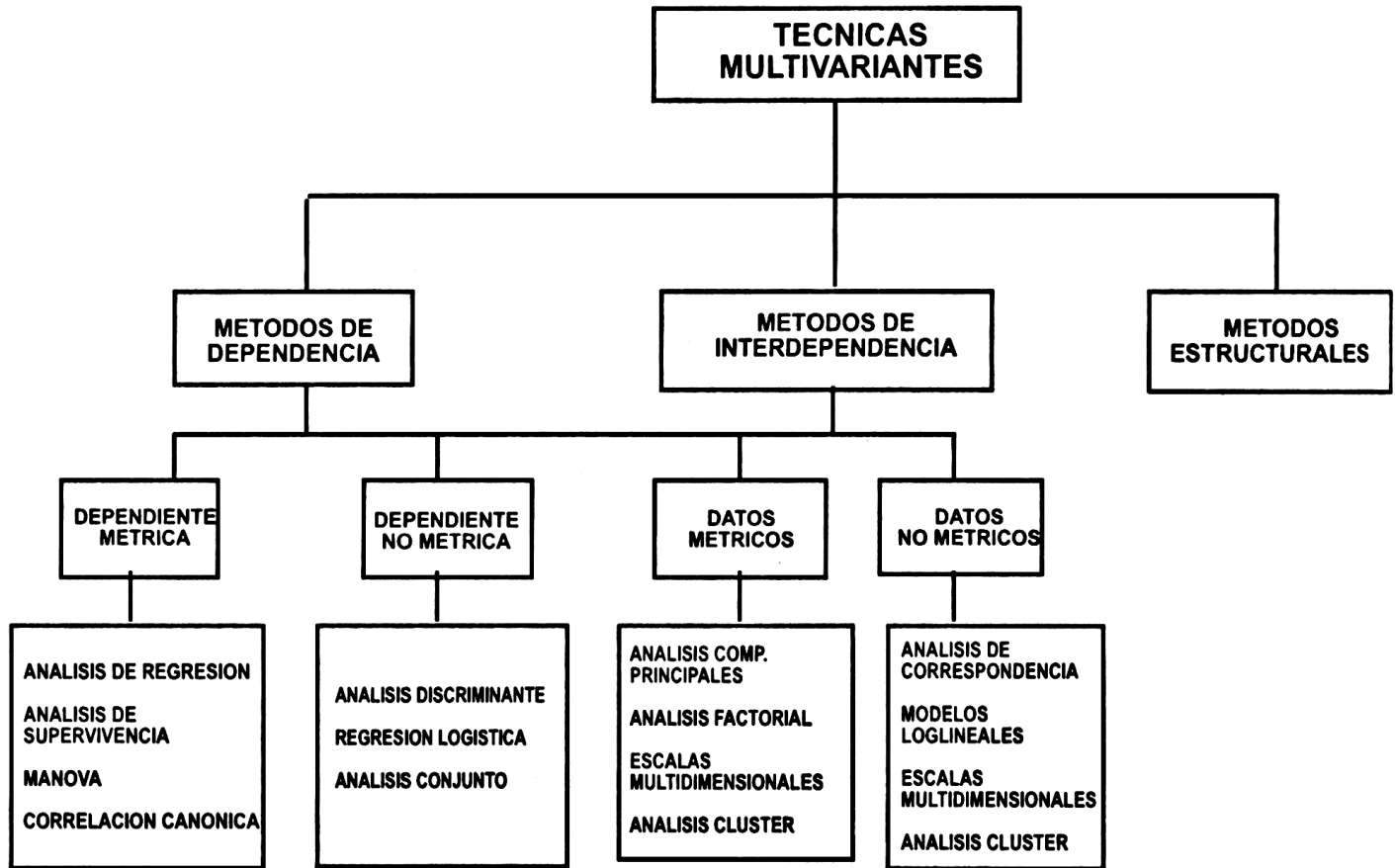
Si la variable dependiente es cuantitativa, algunas de las técnicas que se pueden aplicar son las siguientes:

Métodos estructurales

Suponen que las variables están divididas en dos grupos: el de las variables dependientes y el de las independientes. El objetivo de estos métodos es analizar, no sólo como las variables independientes afectan a las variables dependientes, sino también cómo están relacionadas las variables de los dos grupos entre sí. En resumen los mostramos en la siguiente grafica 1



Grafica 1



Análisis de Regresión

Es la técnica adecuada si en el análisis hay una o varias variables dependientes métricas cuyo valor depende de una o varias variables independientes métricas. Por ejemplo, 1 intentar predecir el gasto anual en cine de una persona a partir de su nivel de ingresos, nivel educativo, sexo y edad.

Análisis de Supervivencia

Es similar al análisis de regresión pero con la diferencia de que la variable independiente es el tiempo de supervivencia de un individuo ú objeto. Por ejemplo, intentar predecir el tiempo de permanencia en el desempleo de un individuo a partir de su nivel de estudios y de su edad.

Análisis de la varianza

Se utilizan en situaciones en las que la muestra total está dividida en varios grupos basados en una o varias variables independientes no métricas y las variables dependientes analizadas son métricas. Su objetivo es averiguar si hay diferencias significativas entre dichos grupos en cuanto a las variables dependientes se refiere. Por ejemplo, ¿hay diferencias en el nivel de colesterol por sexos? ¿Afecta, también, el tipo de ocupación?



Correlación Canónica

Su objetivo es relacionar simultáneamente varias variables métricas dependientes e independientes calculando combinaciones lineales de cada conjunto de variables que maximicen la correlación existente entre los dos conjuntos de variables. Por ejemplo, analizar cómo está relacionado el tiempo dedicado al trabajo y al ocio de una persona con su nivel de ingresos, su edad y su nivel de educación.

Si la variable dependiente es cualitativa, las técnicas que se pueden aplicar son las siguientes:

Análisis Discriminante

Esta técnica proporciona reglas de clasificación óptimas de nuevas observaciones de las que se desconoce su grupo de procedencia basándose en la información proporcionada los valores que en ella toman las variables independientes. Por ejemplo, determinar los ratios financieros que mejor permiten discriminar entre empresas rentables y poco rentables.

Modelos de regresión logística

Son modelos de regresión en los que la variable dependiente es no métrica. Se utilizan como una alternativa al análisis discriminante cuando no hay normalidad.

Análisis Conjunto (Conjoint Analysis)

Es una técnica que analiza el efecto de variables independientes no métricas sobre variables métricas o no métricas. La diferencia de Conjoint con el Análisis de la Varianza radica en que las variables dependientes pueden ser no métricas y los valores de las variables independientes no métricas son fijadas por el analista.

En otras disciplinas, a la técnica Conjoint se la conoce con el nombre de Diseño de Experimentos. Por ejemplo, una empresa quiere diseñar un nuevo producto y, para ello, necesita especificar la forma del envase, su precio, el contenido por envase y su composición química. Presenta diversas composiciones de estos cuatro factores. 100 clientes proporcionan un ranking de las combinaciones que se le presentan. Se quiere determinar los valores óptimos de estos 4 factores.

Métodos de Interdependencia

Se pueden clasificar en dos grandes grupos según que el tipo de datos que analicen sean métricos o no métricos.

Si los datos son métricos, se pueden utilizar, entre otras, las siguientes técnicas:

Análisis Factorial (AF) y Análisis de Componentes Principales (ACP)

Se utiliza para analizar interrelaciones entre un número elevado de variables métricas explicando dichas interrelaciones en términos de un número menor de variables denominadas factores (si son inobservables) o componentes principales (si son observables).

Así, por ejemplo, si un analista financiero quiere determinar cuál es el estado de salud financiero de una empresa a partir del conocimiento de un número de ratios financieros, construyendo varios índices numéricos que definan su situación, el problema se resolvería mediante un ACP. Si un psicólogo quiere determinar los



factores que caracterizan la inteligencia de un individuo a partir de sus respuestas a un test de inteligencia, utilizaría para resolver este problema un AF.

Escalas Multidimensionales

Su objetivo es transformar juicios de semejanza o preferencia en distancias representadas en un espacio multidimensional. Como consecuencia, se construye un mapa en el que se dibujan las posiciones de los objetos comparados de forma que aquellos percibidos como similares están cercanos unos de otros y alejados de objetos percibidos como distintos. Por ejemplo, analizar, en el mercado de refrescos, las percepciones que un grupo de consumidores tiene acerca de una lista de refrescos y marcas con el fin de estudiar qué factores subjetivos utiliza un consumidor a la hora de clasificar dichos productos.

Análisis Cluster

Su objetivo es clasificar una muestra de entidades (individuos o variables) en un número pequeño de grupos de forma que las observaciones pertenecientes a un grupo sean muy similares entre sí y muy disimilares del resto. A diferencia del Análisis Discriminante, en el Análisis Cluster se desconoce el número y la composición de dichos grupos. Por ejemplo, clasificar grupos de alimentos (pescados, carnes, vegetales y leche) en función de sus valores nutritivos.

Si los datos fuesen no métricos, se podrían utilizar, además de las Escalas Multidimensionales y el Análisis Cluster, las siguientes técnicas:

Análisis de Correspondencias

Se aplica a tablas de contingencia multidimensionales y persigue un objetivo similar al de las escalas multidimensionales pero representando simultáneamente las filas y columnas de las tablas de contingencia. Por ejemplo, analizar el paro teniendo en cuenta la provincia, sexo, edad y nivel de estudios del parado.

Modelos log-lineales

Se aplican a tablas de contingencia multidimensionales y modelizan relaciones de dependencia multidimensional de las variables observadas que buscan explicar las frecuencias observadas.

Métodos o Modelos Estructurales

Analizan las relaciones existentes entre un grupo de variables representadas por sistemas de ecuaciones simultáneas en las que se suponen que algunas de ellas (denominadas constructos) se miden con error a partir de otras variables observables denominadas indicadores.

Los modelos utilizados constan, por lo tanto, de dos partes:

- i) Un modelo estructural, que especifica las relaciones de dependencia existente entre los constructos latentes.
- ii) y un modelo de medida, que especifica como los indicadores se relacionan con sus correspondientes constructos.



Por ejemplo, los Modelos Estructurales permiten analizar cómo se relacionan los niveles de utilización de los servicios de una empresa con las percepciones que sus clientes tienen de ella.

4.-Etapas de un análisis multivariante

Con el fin de guiar el desarrollo de la aplicación de un análisis multivariante, podemos decir que se tienen las siguientes etapas:

1.-Objetivos del análisis

Se define el problema especificando los objetivos y las técnicas multivariantes que se van a utilizar. El investigador debe establecer el problema en términos conceptuales, definiendo los conceptos y las relaciones fundamentales que se van a investigar. Se deben establecer si dichas relaciones van a ser relaciones de dependencia o de interdependencia. Con todo esto se determinan las variables a observar.

2.- Diseño del análisis

Se determina el tamaño muestral, las ecuaciones a estimar (si procede), las distancias a calcular (si procede) y las técnicas de estimación a emplear. Una vez determinado todo esto, se proceden a observar los datos.

3.-Hipótesis del análisis

Se evalúan las hipótesis subyacentes a la técnica multivariante. Dichas hipótesis pueden ser de normalidad, linealidad, independencia, homocedasticidad, etc. También se debe decidir qué hacer con los datos missing.

4.-Realización del análisis

Se estima el modelo y se evalúa el ajuste a los datos. En este paso pueden aparecer observaciones atípicas (outliers) o influyentes cuya influencia sobre las estimaciones y la bondad de ajuste se debe analizar.

5.-Interpretación de los resultados

Dichas interpretaciones pueden llevar a reespecificaciones adicionales de las variables o del modelo con lo cual se puede volver de nuevo a los pasos 3 y 4.

5.-Validación del análisis

Consiste en establecer la validez de los resultados obtenidos analizando si los resultados obtenidos con la muestra se generalizan a la población de la que procede. Para ello se puede dividir la muestra en varias partes en las que el modelo se vuelve a estimar y se comparan los resultados. Otras técnicas que se pueden utilizar aquí son las técnicas de remuestreo (jackknife y bootstrap).



6.- BIBLIOGRAFIA

HAIR, J., ANDERSON, R., TATHAM, R. y BLACK, W. (1999). Análisis Multivariante. 5ª Edición. Prentice Hall.

T.W. ANDERSON. An Introduction to Multivariate Statistical Analysis. 2ª Edición. Wiley.

S. JAMES PRESS . Applied Multivariate Analysis . 2ª Edición . Wiley.

JHONSON WICHERN. Applied Method of Multivariate Statistical Analysis. 3ª Edición. Prentice Hall.

EVERITT, B. And GRAHAM, D. (1991). Applied Multivariate Data Analysis. Arnold.

SHARMA, S. (1998). Applied Multivariate Techniques. John Wiley and Sons.

MARDIA, K.V., KENT, J.T. y BIBBY, J.M. (1994). Multivariate Analysis. Academic Press.

FERRAN, M. (1997). SPSS para Windows. Programación y Análisis Estadístico. Mc.Graw Hill.

