

BLOQUE III. ANÁLISIS MULTIVARIANTE

CAPÍTULO 7

Teoría sobre análisis multivariante.

El análisis multivariante engloba un grupo de técnicas o métodos estadísticos cuya finalidad es analizar simultáneamente conjuntos de datos multivariantes en el sentido de que hay varias variables medidas para cada individuo u objeto estudiado.

Su razón de ser radica en un mejor entendimiento del fenómeno estudiado, obteniendo información que los métodos estadísticos univariantes y bivariantes son incapaces de conseguir.

Así, como Hair et al. (1999) dicen:

"Sólo a través del análisis multivariante las relaciones múltiples de este tipo podrán ser examinadas adecuadamente para obtener un entendimiento más completo y real del entorno que permita tomar las decisiones más adecuadas." [4]

III. 7.1 Objetivos del Análisis Multivariante.

Pueden sintetizarse en dos principalmente:

1. Proporcionar métodos cuya finalidad es el estudio de conjunto de datos multivariantes, que el análisis estadístico uni y bidimensional son incapaces de conseguir.
2. 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.

III. 7.2 Tipos de técnicas multivariantes.

Se pueden clasificar en tres grandes grupos (ver esquema adjunto):

1. Métodos de dependencia

Éstos 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.

2. 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.

3. Modelos 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 cómo las variables independientes afectan a las variables dependientes, sino también cómo están relacionadas las variables de los dos grupos entre sí.

En el siguiente esquema se resumen las distintas técnicas.

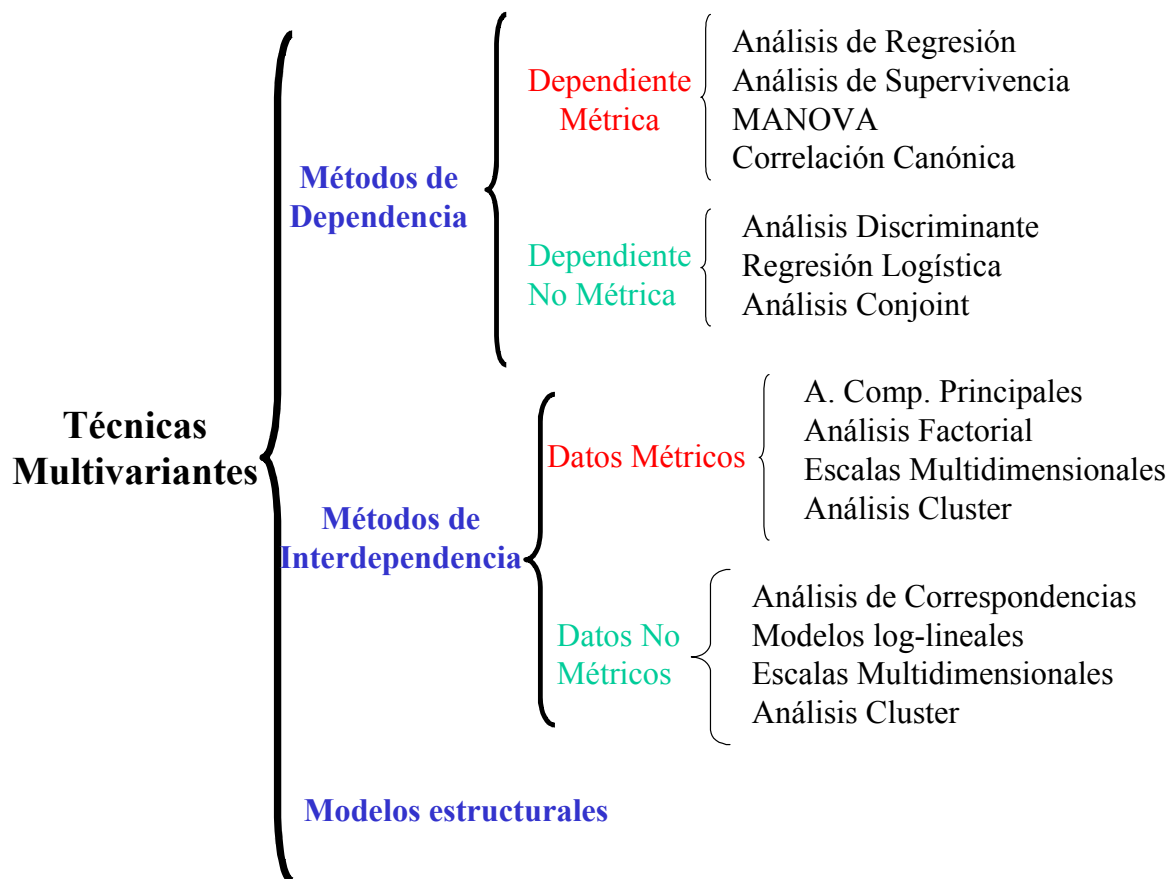


Figura 20. Técnicas de análisis multivariante.

III. 7.2.1 Métodos de dependencia.

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

III. 7.2.1.1 Dependencia métrica.

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, intentar predecir el gasto anual en cine de una persona a partir de su nivel de ingresos, nivel educativo, sexo y edad.

2. 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 u 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.

3. 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?

4. 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.

III. 7.2.1.2 Dependencia no métrica.

1. 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 por 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.

El objetivo del análisis discriminante es, por tanto, identificar cuáles son las características distintivas de los individuos en cada grupo y, posteriormente, poder utilizarlas para estimar el grupo al que pertenecen otros individuos de los que se conocen dichas características, pero no su grupo de pertenencia.

Este método de análisis permite, en primer lugar, determinar cuáles son las variables (de entre una serie de variables seleccionadas previamente por el investigador) que mejor explican la pertenencia de un individuo a un determinado grupo.

En segundo lugar, también permite determinar el grupo al que pertenecerá un individuo pendiente de clasificación sobre la base de las respuestas de dicho individuo a las variables que más explican la pertenencia a cada grupo.

Es una técnica muy potente puesto que permite considerar muchas variables de forma simultánea.

2. 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.

3. Análisis Conjoint.

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

III. 7.2.2 Métodos de Interdependencia.

Se pueden clasificar en dos grandes grupos según los tipos de datos que analicen, que al igual que en los métodos de dependencia, pueden ser métricos o no métricos.

III. 7.2.2.1 Dependencia métrica.

1. Análisis Factorial y Análisis de Componentes Principales.

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 Análisis de Componentes Principales.

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 Análisis Factorial.

2. Escalado Multidimensional (MDS).

Esta técnica ha sido una de las utilizadas para la realización de este proyecto y aquí sólo se menciona brevemente. Se hará un desarrollo más exhaustivo en un apartado posterior.

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 aquéllos 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.

3. Análisis Cluster.

Al igual que ocurría con el escalamiento multidimensional, el análisis cluster es la segunda técnica empleada en el proyecto.

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

distintas del resto. A diferencia del Análisis Discriminante se desconoce el número y la composición de dichos grupos. Precisamente, es la utilización de escalamiento multidimensional, así como del análisis cluster, lo que permite dar respuesta a esas dos incógnitas.

Por ejemplo, clasificar grupos de alimentos (pescados, carnes, vegetales y leche) en función de sus valores nutritivos.

III. 7.2.2.2 Dependencia no métrica.

1. 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 en Aragón teniendo en cuenta la provincia, sexo, edad y nivel de estudios del parado.

2. Modelos log-lineales.

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

III. 7.3 Etapas de un análisis multivariante.

Las etapas necesarias para realizar un análisis multivariante pueden sintetizarse en 6:

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 perdidos (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 re-especificaciones adicionales de las variables o del modelo con lo cual se puede volver de nuevo a los pasos 3) y 4).

6. 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).