

Gestión de la Calidad: Control Estadístico y Seis Sigma

*Giampaolo Orlandoni Merli**

Introducción

El concepto de Calidad se refiere a ciertos atributos mensurables de un producto, servicio o proceso, para los que se ha establecido cierto estándar. Un producto o servicio es de calidad cuando satisface las necesidades de los usuarios en cuanto a seguridad, fiabilidad y servicio (Juran y Godfrey, 2001). Los conceptos de calidad y productividad están estrechamente relacionados. La baja calidad de un producto, manifestada como defectos de fabricación y baja confiabilidad en su rendimiento, afecta la productividad global de la empresa y su rentabilidad final.

El mejoramiento de la calidad contribuye a eliminar el desperdicio, reducir el retrabajo de piezas, minimizar los requerimientos para inspección y prueba, y las pérdidas por garantía, además de mejorar la satisfacción del consumidor, logrando así que la empresa se convierta en un productor confiable, de alta calidad y bajo costo para la industria. Cuando la gerencia incluye como inicio de su plan estratégico los requerimientos de consumidores y la calidad de los productos competidores, entonces la calidad se convierte en estrategia de competitividad.

La estadística es un componente fundamental en el mejoramiento de la calidad, y sus técnicas pueden emplearse para describir, comprender y controlar la *variabilidad* que se presenta entre objetos, aún si se han producido presumiblemente bajo las mismas condiciones (Montgomery, 2004).

La variabilidad de un proceso puede deberse a causas *aleatorias* y a causas *asignables*. Las causas *aleatorias o comunes* corresponden al efecto acumulado de muchas causas inevitables y pequeñas; se originan por la variabilidad de la materia prima, maquinaria distinta, diferente eficacia de la mano de obra; ellas definen un sistema estable de causas de variabilidad. Si un proceso opera únicamente con causas aleatorias de variación, entonces él está bajo control estadístico, por ser estas causas parte inherente del proceso (Montgomery, 2004).

Causas asignables pueden estar presentes a la salida de un proceso, generando variaciones irregulares no predecibles. Son de naturaleza no aleatoria y cuando actúan producen efectos que se pueden identificar y que persisten hasta que se eliminan las causas que los generan. Son ejemplos, el desgaste de la máquina, herra-

* Universidad de Los Andes. Mérida, Venezuela. Correo electrónico: orlandon@ula.ve

mientas en mal estado, la enfermedad de operarios. Un proceso que opera en presencia de causas asignables está fuera de control estadístico. Un objetivo esencial del sistema de control es que éste reaccione sólo cuando una causa tiene suficiente impacto, y es práctico y económico eliminarla para mejorar la calidad del proceso, producto o servicio (Montgomery, 2004).

El Control Estadístico de Procesos (CEP) y sus Herramientas Estadísticas Básicas

El CEP comprende un conjunto de herramientas estadísticas que permiten descubrir causas especiales de variación, que no son parte del proceso permanentemente, y surgen debido a condiciones específicas. Esas herramientas son útiles para lograr la estabilidad de los procesos productivos y mejorar su capacidad mediante la reducción de la variabilidad.

El CEP se aplica a cualquier tipo de proceso. Se basa en el uso eficiente de herramientas estadísticas, tales como: histograma de frecuencias, hoja de verificación, gráfica de Pareto, diagrama de causa y efecto, diagrama de concentración de efectos, diagrama de dispersión, cartas de control. Aún cuando estas herramientas son una parte importante de CEP, ellas constituyen tan sólo su aspecto técnico. El CEP crea un ambiente en el que todos los entes de la organización desean el mejoramiento continuo de la calidad y la productividad. Este ambiente se desarrolla mejor cuando la dirección se involucra en el proceso de mejoramiento continuo de la calidad.

Un programa de control estadístico para un proceso requiere la ejecución de dos fases (Montgomery, 2004):

Fase I: En esta fase se toman las mediciones para estimar parámetros fundamentales, como el promedio y la desviación estándar de la variable que se está midiendo. Con base en dichas estimaciones se calculan los límites de control, bajo el supuesto de que las observaciones siguen distribución normal: $LSC = \mu + L\sigma$; $LIC = \mu - L\sigma$, siendo L la distancia de los límites de control a la línea central. En general, se utilizan límites de tres sigmas alrededor del promedio. Luego se construye un gráfico de prueba conformado por una línea recta horizontal a la altura del promedio y otras dos líneas rectas a la altura de los límites de control.

En este gráfico inicial se representan las observaciones usadas en el cálculo de los límites de control. Se analiza para verificar si se cumple la hipótesis de que la variabilidad del proceso se debe sólo a causas aleatorias o si, por el contrario, existen causas asignables de variación. Esto se puede verificar, ya que cuando la fluctuación de las mediciones se debe a un sistema constante de causas aleatorias, las observaciones siguen una distribución normal. Si no se descubren causas asignables, entonces se adoptan los límites de control calculados como definitivos, y se construyen cartas de control con esos límites. En caso de existir pocos puntos fuera de control, se eliminan, se reestiman los parámetros y los límites de control con los restantes puntos, y se construye un nuevo gráfico de prueba.

Cuando las observaciones no siguen un patrón aleatorio, indicando la existencia de causas asignables, es necesario investigar para descubrirlas y eliminarlas. Una vez hecho esto, se deberán tomar nuevas observaciones y calcular nuevos límites de control de prueba.

Fase 2: En esta fase las nuevas observaciones del proceso se representan en el gráfico de control, verificando que ellas estén dentro de los límites, y que no se observen patrones no aleatorios. El 99.8% de las observaciones deben estar dentro de los límites de 3.09 sigmas alrededor de la media. Esto significa que sólo una observación en 500 (0.2% de las observaciones) puede estar, por causas aleatorias, fuera de los límites de control. Cuando hay más de un punto fuera de los límites de control, significa que el sistema de causas aleatorias que provoca la variabilidad habitual de las observaciones se ha alterado por la aparición de alguna causa asignable; es necesario entonces descubrir y eliminar dicha causa para que el proceso retorne a control.

Cartas de Control (CC)

Las cartas de control son las herramientas que permiten determinar la presencia de las causas asignables. Se clasifican en CC para variables y CC para atributos.

Cartas de Control para Variables. Son herramientas cuyo objetivo es controlar estadísticamente un proceso, detectando cuándo está fuera de control. Las cartas de control más comunes son las Cartas de Shewhart. Examinando el proceso a ser controlado, se toman m muestras a intervalos regulares. De cada muestra se mide una o varias variables; las muestras correspondientes a un mismo intervalo constituyen un subgrupo. Los intervalos pueden definirse en términos de unidades de tiempo o de cantidad. Los valores medidos se comparan con los límites calculados. Los límites más comunes son los Límites de Tolerancia Natural (LTN) del proceso: límites entre los que se mueve el proceso de manera natural. Se utilizan como Límites de Control (LIC, LSC) paralelos a la línea central representativa del valor medio (Gutiérrez y Vara, 2009; Montgomery, 2004).

Cuando se analizan variables es necesario controlar tanto el promedio como la variabilidad de la característica de calidad. En cada subgrupo, se mide una cierta característica, y se estiman diversos estadísticos, como: media del subgrupo (μ), rango (R), desviación estándar (σ).

Si la característica a controlar es una variable continua, las medias de las muestras siguen aproximadamente una distribución normal $N(\mu, \sigma)$. En general, los límites de control están definidos por $LC = \mu \pm k\sigma$, donde el valor k depende del tipo de carta de control. Estos límites en general se estiman en la fase de estudio inicial, a partir de las observaciones de m subgrupos, siendo $m=25$ usualmente. Las cartas de control más frecuentes para una sola variable son las CC Media-Rango (\bar{x}, R), CC Media-Desviación Estándar (\bar{x}, S) y CC para valores individuales (X_i).

Cartas de Control Media-Rango. Para cada grupo se representan los valores (\bar{x}, R); si están comprendidos en el intervalo (LIC;LSC) entonces el proceso está

bajo control; caso contrario el proceso puede estar fuera de control. Es importante mantener bajo control tanto la media como la variabilidad del proceso en estudio (Gutiérrez y Vara, 2009; Montgomery, 2004).

Cartas de Control para Atributos. Las CCA se utilizan cuando la característica de calidad no puede ser cuantificada; sólo puede indicarse si ella está conforme o no con ciertas especificaciones (defectuoso-no defectuoso; conforme-no conforme). Así, se determina el número o porcentaje de unidades defectuosas en la muestra. Las CCA tienen gran aplicación en las actividades relacionadas con servicios y con mejoramiento de procesos fuera de la manufactura; pueden clasificarse como se indica en la Tabla 1:

Tabla 1. Cartas Control para Atributos

Carta	Características
1. Carta p	• Fracción disconforme de unidades en una muestra de n unidades
2. Carta np	• Número de disconformidades
3. Carta c	• Número total de disconformidades por unidad.
4. Carta u	• Número promedio de disconformidades por unidad

Análisis de Capacidad del Proceso

Luego que el proceso en estudio entra en control y se estabiliza, es necesario asegurar que tiene capacidad para operar dentro de ciertas especificaciones definidas. Un proceso en control indica que está operando dentro de su variabilidad natural, pues las causas especiales han sido eliminadas. Sin embargo, es posible que el proceso no tenga capacidad para cumplir con ciertas especificaciones, aunque esté en control. La capacidad de un proceso se refiere a su uniformidad, siendo la variabilidad que presenta una medida de esta uniformidad.

Usualmente se toma la dispersión seis sigmas en la distribución de la característica de calidad como una medida de la capacidad del proceso. Bajo distribución normal $N(\mu, \sigma)$, los límites de tolerancia natural del proceso son $\mu \pm 3\sigma$. Ellos incluyen el 99.73% de la variabilidad, quedando por fuera de esos límites de tolerancia el 0,27% de la salida del proceso, lo que representa 2.700 partes por millón disconformes.

El análisis de capacidad de procesos (ACP) se puede realizar mediante histogramas de frecuencias, cartas de control y diseños experimentales.

Análisis de la capacidad del proceso mediante histogramas. El histograma de una distribución normal tiene forma acampanada, simétrica, unimodal. Cuando se aleja de este patrón, es indicio de que en el proceso están actuando causas asignables que deforman la curva acampanada. En estos casos la predicción de la capacidad del proceso no es precisa.

Las formas no acampanadas del histograma indican distribución asimétrica, o multimodal. Las causas posibles son: las diferencias en lotes de materia prima y

diferentes proveedores, varios operadores en el proceso con criterios diferentes, mediciones hechas por operarios diferentes, con instrumentos diferentes o mal calibrados (Juran y Godfrey, 2001; Gutiérrez y Vara, 2009; Montgomery, 2004).

Análisis de la capacidad del proceso mediante Índices de Capacidad. Los índices de capacidad comparan la tolerancia de los Límites Especificación (USL-LSL) con la amplitud de la variación real del proceso (6σ). Los principales índices son: capacidad potencial del proceso C_p , capacidad real C_{pk} , centrado del proceso K , índice Taguchi C_{pm} . Estos índices pueden analizarse a corto plazo y a largo plazo.

Tópicos especiales de Control Estadístico de Procesos

El estudio de las cartas de control de Shewhart y de la capacidad de los procesos, constituye la base del control de calidad. Hay además muchos otros tópicos que incorporan particularidades propias de las mediciones hechas, así como del tipo especial de proceso en estudio (Montgomery, 2004). Así se tiene el caso de la necesidad de detectar cambios pequeños, o procesos que generan datos autocorrelacionados.

- **CC para detección de cambios pequeños:** Cartas CUSUM y Cartas EWMA (Promedios Móviles Ponderados Exponencialmente).
- **CC para procesos generadores de mediciones autocorrelacionadas:** Cartas de control ARIMA.
- **Cartas de Control Multivariadas:** para el estudio simultáneo de varias características de calidad. Cartas T2 Hotelling.

Gestión de Calidad y Metodología Seis Sigma

Para que los productos y servicios sean competitivos es necesario asegurar la calidad en los respectivos procesos productivos. La gestión de la calidad basada en la metodología seis sigma es una estrategia de mejora continua de los procesos; su objetivo es localizar y eliminar las causas raíces de los errores, defectos y retrasos en dichos procesos, centrándose en los aspectos críticos de satisfacción para el cliente. Se trata de una metodología sistemática y cuantitativa, basada en herramientas y pensamiento estadísticos, orientada a mejorar los resultados de los procesos en tres áreas: satisfacción del cliente, reducción del tiempo de ciclo y disminución de los defectos (Gutiérrez y Vara, 2009).

La metodología seis sigma se puede plantear para dos situaciones diferentes:

Metodología Seis Sigma para reducir tasa de defectos de productos, servicios y procesos existentes (DMAMC). Las fases de este enfoque son: 1) DEFINICIÓN. Definición de Problemas. Descripción de Procesos. 2) MEDICIÓN. Sistemas de Medición. Evaluación de los sistemas de medición. Repetibilidad y Reproducibilidad. 3) ANÁLISIS. a) Análisis de las variables significativas del proceso. Diseño de Experimentos. Estudios MultiVari. b) Evaluación de la Estabilidad y Capacidad del Proceso. Gráficas de Control. 4) MEJORAMIEN-

TO. Mejoramiento y Optimización del Proceso. Métodos Experimentales y Diseño Estadístico de Experimentos. Análisis de Regresión. Metodología de Superficie de Respuesta. 5) CONTROL. Control y Monitoreo del Proceso. Mejoramiento Continuo.

Las acciones de control se definen a tres niveles: proceso, documentación y monitoreo. Así, es necesario estandarizar el proceso, documentar el plan de control, monitorear el proceso, cerrar y difundir el proyecto.

Metodología Seis Sigma para la fase de diseño del producto y del proceso (DMADV). Cuando el producto es nuevo y el proceso aún no ha sido probado, la metodología seis sigma recurre al diseño estadístico. Las fases de este enfoque son: Definición, Medición, Análisis, Diseño, Verificación.

Referencias bibliográficas

- Juran, Joseph y Godfrey, Blanton. (2001). **Manual de Calidad.** Vol I. 5ta Ed. McGraw-Hill. España.
- Gutiérrez, Humberto y Vara, Román. (2009). **Control Estadístico de Calidad y Seis Sigma.** 2ª Ed. McGraw-Hill. México.
- Montgomery, Douglas. (2004). **Control Estadístico de la Calidad (Introduction to Statistical Quality Control).** 3a Ed. Limusa-Wiley. México.

Clasificación de las Técnicas Estadísticas Multivariantes

Milton Quero Virla y Karla Inciarte Pirela***

Introducción

En esta nota se presenta una clasificación de algunas técnicas de análisis estadístico multivariante. En sintonía con una realidad socio-económica-empresarial compleja actual, las técnicas multivariantes encuentran un oportuno y conveniente campo de aplicación, no obstante los orígenes de dichas técnicas se remontan a fines del siglo XIX y primeras décadas del siglo XX con estudios en áreas como psicología y antropología (Peña, 2002, p. 10) (Landeroy y González, 2006, p. 391-393).

No siendo un término fácil de definir, en su sentido amplio, el análisis multivariante se refiere a todos los métodos estadísticos que analizan simultáneamente medidas múltiples de cada individuo u objeto sometido a investigación; y en otras, palabras, cualquier análisis simultáneo de más de dos variables puede ser considerado aproximadamente como un análisis multivariante (Hair y Taham, 2008, p.4). Coincidiendo con este criterio de “simultaneidad”, también puede definirse como el conjunto de técnicas estadísticas que de forma simultánea miden, explican y predicen todas las relaciones existentes entre los elementos que conforman una tabla de datos, proporcionando un resultado que debe ser interpretado minuciosamente por el analista (Lévy y Varela, 2005, p.6).

Clasificación de las Técnicas Multivariantes

Para la clasificación de las técnicas recopiladas en esta Nota se han considerado dos criterios no excluyentes: el primero, que atiende al objetivo científico principal de una investigación: extracción de información (enfoque descriptivo)

* Licenciado en Educación Mención Matemática y Física. MSc Estadística Aplicada. Estudios Doctorales en Ciencias de la Educación. Profesor de Educación Integral Especialidad Matemática. Profesor Agregado del Departamento de Matemáticas de la Facultad Experimental de Ciencias, LUZ. Maracaibo, Venezuela. Correo electrónico: miltonqv@hotmail.com

** Técnico Superior Universitario en Diseño Gráfico del Colegio Universitario Dr. Rafael Beloso Chacín, CUNIBE. Estudios de Pregrado en la Licenciatura en Matemáticas de la Universidad del Zulia, LUZ. Maracaibo, Venezuela. Correo electrónico: karlavip_3107@hotmail.com

por una parte, y generación de conocimiento (enfoque inferencial), por la otra. Ambos enfoques pueden vincularse a objetivos específicos de reducción de la información y obtención de indicadores; clasificación y construcción de grupos; y, relación-predicción de variables (Peña, 2002, p.7) (Pérez, 2004, p. 1).

Un segundo criterio, considera la naturaleza o tipo de variables involucradas y las relaciones entre estas: técnicas de análisis de la dependencia (D), con variables dependientes, criterios o explicadas (VD) y variables independientes, predictoras o explicativas (VI); y las técnicas de análisis de la interdependencia (I), que no discriminan entre VD y VI. Los resultados se muestran en los cuadros siguientes.

Cuadro 1
Clasificación de los Métodos Multivariantes
según enfoque y objetivos

Objetivos	Enfoque descriptivo (Información) Métodos de exploración de datos Objetivo: Extraer información	Enfoque Inferencial (Conocimiento) Métodos de Inferencia Objetivo: Generar conocimiento
Obtener Indicadores	<ul style="list-style-type: none"> • Componentes Principales (I) • Correspondencias (I) 	<ul style="list-style-type: none"> • Factorial (I)
Clasificar	<ul style="list-style-type: none"> • Conglomerados (I) 	<ul style="list-style-type: none"> • Discriminante (D) • Elección Discreta (D)
Construir grupos	<ul style="list-style-type: none"> • Conglomerados (I) 	
Relacionar variables	<ul style="list-style-type: none"> • Regresión Múltiple (D) 	<ul style="list-style-type: none"> • Correlación Canónica (D) • MANOVA (D) • Ecuaciones Estructurales (D)

Fuente: Elaboración propia a partir de la original de Peña (2002, p. 7).

Cuadro 2
Técnicas de Análisis de la Dependencia

Nombre (S) y Objetivo (S)	Supuestos y naturaleza de las variables	Comentarios Adicionales
<p>Análisis Multivariable de la Varianza (Manova).</p> <p>Analizar una relación de dependencia representada como las diferencias en un conjunto de medidas dependientes a través de una serie de grupos formados por una o más medidas independientes categóricas (Hair y Taham, 2008, p.355).</p> <p>Contrastar si los valores no métricos de VI determinarán la igualdad de vectores de medidas de una serie de grupos determinados por ellos en las VD (Pérez, 2004, p. 8).</p>	<p>VD: Métricas VI: No métricas.</p> $\mathcal{Y}_1 + \mathcal{Y}_2 + \dots + \mathcal{Y}_m = X_1 + X_2 + \dots + X_n$ <p>Supuestos:</p> <ol style="list-style-type: none"> 1. Independencia de las observaciones. 2. Homogeneidad de las matrices de varianzas-covarianzas (homocedasticidad). 3. Normalidad multivariante del conjunto de VD. 4. Linealidad-multicolinealidad del valor teórico de las VI. 5. Sensibilidad a casos atípicos (Hair y Taham, 2008, pp. 355 y 362). <p>Adicionalmente (entre otras cosas), consideraciones referidas al tamaño muestral deben tomarse en cuenta (Hair y Taham, 2008, pp. 356-357).</p>	<p>Como construcción teórica fue introducido hace varias décadas según la formulación original de Wilks (Hair y Taham, 2008, p. 345).</p> <p>La técnica del Manova es una generalización del Anova (univariante) para un vector de variables dependientes (Lévy y Varela, 2005, p. 217).</p> <p>Más aún, el Manova es la extensión multivariante de las técnicas univariantes para valorar las diferencias entre las medias de los grupos: el contraste t para situaciones de dos grupos, y el Anova en situaciones con tres o más grupos definidos por dos o más VI. Tanto el Anova como el Manova son particularmente útiles cuando se utilizan conjuntamente con diseños de experimentos (Hair y Taham, 2008, p. 346).</p>

Cuadro 2 (Continuación)

Nombre (S) y Objetivo (S)	Supuestos y naturaleza de las variables	Comentarios Adicionales
<p>Regresión Lineal Múltiple</p> <p>Predecir los cambios (cantidad o magnitud) en la VD en respuesta a cambios en varias de las VI (Hair y Taham, 2008, p. 11).</p> <p>Analizar la relación entre una VD (endógena) métrica y varias VI (exógenas) también métricas (Pérez, 2004, p. 5).</p>	<p>VD: métrica VI: métricas. $Y_1 = X_1 + X_2 + \dots + X_n$</p> <p>Supuestos:</p> <ol style="list-style-type: none"> 1. Linealidad del fenómeno medido. 2. Homocedasticidad del término del error. 3. Independencia de los términos de error. 4. Normalidad de la distribución del término del error (Hair y Taham, 2008, p.166). <p>Además, ciertas condiciones del tamaño muestral deben considerarse (Hair y Taham, 2008, p. 159).</p>	<p>Es la técnica de dependencia más versátil y ampliamente utilizada (Hair y Taham, 2008, p.143).</p> <p>Admite la posibilidad de trabajar con variables independientes no métricas si se emplean variables ficticias para transformarlas en métricas (Pérez, 2004, p. 5) (Hair y Taham, 2008, p. 159).</p> <p>Las crecientes aplicaciones de la regresión múltiple se agrupan en dos amplias clases de problemas de investigación (no mutuamente excluyentes): predicción y explicación (Hair y Taham, 2008, p. 155).</p>
<p>Correlación Canónica, Análisis Canónico.</p> <p>Analizar la relación entre múltiples VD (endógenas) métricas y varias VI (exógenas) también métricas (Pérez, 2004, p. 5).</p> <p>El objetivo es correlacionar simultáneamente varias VD métricas y varias VI métricas (Hair y Taham, 2008, p. 12).</p> <p>También puede extenderse al caso de VD no métricas y de VI no métricas (Pérez, 2004, p. 5).</p>	<p>VD: métricas y no métricas VI: métricas y no métricas $Y_1 + Y_2 + \dots + Y_m = X_1 + X_2 + \dots + X_n$</p> <p>Supuestos:</p> <ol style="list-style-type: none"> 1. Linealidad de correlaciones entre cualesquiera dos variables. 2. Linealidad de la relación. 3. Normalidad multivariante. 4. Homocedasticidad. 5. No multicolinealidad entre variables (Hair y Taham, 2008, p. 474). 	<p>Su creación se atribuye a Hotelling en 1936, como una extensión de la idea de componentes principales (Peña, 2002, p. 486).</p> <p>Estudia las interrelaciones entre múltiples variables criterio (dependientes) y múltiples variables predictoras (independientes). Mientras la regresión múltiple predice una única VD, la correlación canónica predice simultáneamente múltiples VD (p. 470). Aunque es el método más extendido está restringido a la identificación de relaciones lineales (Hair y Taham, 2008, p. 474).</p>

Cuadro 2 (Continuación)

Nombre (S) y Objetivo (S)	Supuestos y naturaleza de las variables	Comentarios Adicionales
<p>Análisis Discriminante Múltiple</p> <p>Pretende dos posibles fines: uno descriptivo y otro predictivo (Lévy y Varela, 2005, p. 19).</p> <p>Persigue los objetivos primarios de entender las diferencias de los grupos y predecir la verosimilitud de que una entidad (persona u objeto) pertenezca a una clase o grupo particular basándose en varias VI métricas (Hair y Taham, 2008, p. 11).</p> <p>Analizar la relación entre una VD (endógena) no métrica y varias VI (exógenas) métricas (Pérez, 2004, p. 5).</p>	<p>VD: No métrica VI: Métricas</p> $\chi^2 = X_1 + X_2 + \dots + X_n$ <p>Supuestos:</p> <ol style="list-style-type: none"> 1. Normalidad multivariante de las VI. 2. Linealidad de las relaciones. 3. No multicolinealidad entre las VI. 4. Igualdad en las matrices de covarianzas de los grupos. <p>Además, consideraciones respecto a tamaños muestrales y valores atípicos, también deben ser tomados en cuenta (Hair y Taham, 2008, p. 260 y 264).</p>	<p>Su creación se atribuye a Fisher en la década de 1930, cuando resolvió un problema de discriminación en antropología basado en Anova (Peña, 2002, p. 10).</p> <p>Es una técnica de análisis multivariante de la dependencia, y también se considera como una técnica de clasificación (Pérez, 2004, pp. 4 y 6).</p> <p>Tiene la capacidad de tratar tanto dos grupos como grupos múltiples (tres o más) (Hair y Taham, 2008, p. 250).</p>

Cuadro 2 (Continuación)

Nombre(S) y Objetivo(S)	Naturaleza de las variables y supuestos	Comentarios Adicionales
<p>Análisis Conjunto: Analizar la relación lineal–no lineal entre una VD (endógena) generalmente ordinal (aun- que también métrica) y varias VI (exógenas) no métricas (Pérez, 2004, p. 9).</p> <p>Entender cómo los encuestados desarrollan preferencias acerca de los productos o servi- cios, basándose en la simple premisa de que los consumidores evalúan el valor del produc- to/servicio/idea (real o hipotética) combinan- do cantidades separadas de valor que propor- ciona cada atributo (Hair y Taham, 2008, p. 408).</p>	<p>VD : No métrica o métrica VI: No métricas $Y_i = X_{i1} + X_{i2} + \dots + X_{in}$ Supuestos: 1. Propiedades del modelo. 2. Representatividad de la muestra (Hair y Taham, 2008, p. 418).</p> <p>Tiene pocos supuestos estadísticos. Está de- terminado por la teoría en su diseño, estima- ción e interpretación. Los supuestos concep- tuales son quizá mayores que con cualquier otra técnica multivariante. Esto hace innecesarios la mayoría de los test realizados en otros métodos de dependencia: normalidad, homocedasticidad e independencia (Hair y Taham, 2008, p. 433).</p>	<p>Desde los años 70 ha suscitado una atención considerable, aumentando su utilidad hasta 10 veces en los años 80. Durante los 90 au- mentó aún más, extendiéndose a amplios campos de estudio (Hair y Taham, 2008, p. 407).</p> <p>Es una técnica descomposicional (descompo- ne las preferencias), mientras que el Análisis Discriminante y el Análisis de Regresión son técnicas composicionales (Pérez, 2004, p. 9).</p> <p>Está muy relacionado con la experimentación tradicional (diseño y análisis de experimen- tos) (Hair y Taham, 2008, p.408).</p> <p>Su desarrollo lo ha experimentado en el seno del marketing (Lévy y Varela, 2005, p. 21).</p>

Cuadro 2 (Continuación)

<p>Ecuaciones Estructurales (Modelo de o Análisis con), Análisis de estructuras de covarianzas, Lisrel.</p> <p>Potente técnica que permite analizar varias relaciones de dependencia que se presentan simultáneamente (Lévy y Varela, 2005, p. 23).</p>	<p>VD: Métricas VI: Métricas y no métricas</p> $\mathcal{X}_1 = X_{11} + X_{12} + \dots + X_{1n}$ $\mathcal{X}_2 = X_{21} + X_{22} + \dots + X_{2n}$ $\mathcal{X}_m = X_{m1} + X_{m2} + \dots + X_{mn}$	<p>Sus antecedentes se remontan desde finales de los años 60 y la década de los 70 (Lévy y Varela, 2005, p.769). Es considerada una técnica avanzada emergente (Hair y Taham, 2008, p. 609).</p>
<p>Permite separar las relaciones para conjuntos de VD (Hair y Taham, 2008, p. 14).</p> <p>Todas las técnicas asociadas a este modelo se distinguen por: 1. La estimación de relaciones de dependencia múltiples y cruzadas. 2. Capacidad de representar conceptos no observados en estas relaciones y tener en cuenta el error de medida en el proceso de estimación (Hair y Taham, 2008, p. 612).</p>	<p>Supuestos: Comparte 3 supuestos con los otros métodos: observaciones independientes, muestra aleatoria de los encuestados y linealidad de las relaciones. Adicional a esto, es más sensible al incumplimiento de la normalidad multivariante o una fuerte curtosis. Como complemento, además de la ausencia de datos y tamaño muestral, asuntos como covarianzas versus correlaciones y los tipos de correlaciones o covarianzas utilizadas deben tomarse muy en cuenta (Hair y Taham, 2008, p. 630-631).</p>	<p>Abarca una familia entera de modelos conocidos con muchos nombres: análisis de variables latentes, análisis de factor confirmatorio, además de los ya mencionados. Se ha utilizado en casi todos los campos de estudio concebibles (Hair y Taham, 2008, p. 612).</p> <p>Es característico el empleo de variables no observables directamente (variables latentes, factoriales o constructos) que resumen la información contenida en las variables observadas (manifiestos o indicadores) (Lévy y Varela, 2005, p. 23).</p>

Cuadro 2 (Continuación)

<p>Modelos de Elección Discreta, Modelos de Probabilidad Lineal, Regresión Logística.</p> <p>Analizar la relación entre una VD (endógena) no métrica, y varias VI (exógenas) métricas (Pérez, 2004, p. 6).</p> <p>Analiza la relación causal existente entre una VD dicotómica y varias VI métricas o no métricas, estimar la probabilidad de que se produzca el suceso definido por la VD, si las VI toman determinado valor (Lévy y Varela, 2005, p. 20).</p>	<p>VD: No métrica. VI: Métricas</p> <p>Supuestos: Muchas decisiones del análisis discriminante son idénticas al de la regresión logística. Así, los supuestos considerados para aquel, son también aplicables acá. Sin embargo, se tiene una robustez del modelo de regresión logística respecto al incumplimiento del supuesto de igualdad de las matrices de varianzas-covarianzas entre grupos, lo cual hace que la regresión logística resulte bastante adecuada en esta situación (Hair y Taham, 2008, p. 316).</p>	<p>Es un tipo especial de regresión que se utiliza para predecir y explicar una variable categórica binaria (Hair y Taham, 2008, p. 251). Consiste en una combinación de regresión múltiple y análisis discriminante múltiple (Hair y Taham, 2008, p. 14).</p> <p>Tiene la misma naturaleza que el análisis discriminante pero lo que predice es la probabilidad de pertenencia a una categoría (clase) para los valores dados de las VI (Pérez, 2004, p. 6).</p> <p>La expresión analítica de la probabilidad estimada recibe el nombre de modelo logit cuando dicha estimación se hace como una función logística. Si se plantea como función normal estándar se tiene el modelo probit (Lévy y Varela, 2005, p. 20-21).</p>
---	--	--

Fuente: Elaboración propia, a partir de los autores citados.

Cuadro 3
Técnicas de Análisis de la Interdependencia

Nombre(S) y Objetivo(S)	Naturaleza de las Variables y Supuestos	Comentarios Adicionales
<p>Análisis Factorial (Común). Persigue reducir la dimensión de una tabla de datos excesivamente grande por el número de variables que contiene y quedarse con unas cuantas variables ficticias que, aunque no observadas, sean combinación de las reales (Pérez, 2004, p.12). Identificar factores subyacentes o las dimensiones que reflejan qué es lo que las variables comparten en común (Hair y Taham, 2008, p.89).</p>	<p>Variables: métricas. Supuestos: Los supuestos básicos subyacentes son más de tipo conceptual que estadístico. Desde este segundo punto de vista pueden obviarse algunos, pero con la previsión de las implicaciones que cada una en específico puede tener. Los supuestos conceptuales se relacionan con la serie de variables seleccionadas y la muestra elegida. El investigador debe asegurarse de que las pautas observadas sean válidas y conceptualmente apropiadas para usar la técnica (Hair y Taham, 2008, pp. 88,89 y p. 108).</p>	<p>Los antecedentes se remontan a finales del siglo XIX con Galton y sus estudios de medida de la inteligencia humana. Pearson en 1901 crea el primer procedimiento de cálculo para este análisis, y Spearman (1904) fue el primero en aplicar el análisis, tiempos en los que por la complejidad solía extraerse un solo factor. Ya para los años 50 se había racionalizado y extendido su uso por la aparición de las computadoras y los programas estadísticos (Lanero y González, 2006, pp.391-393). Son varias las relaciones y comparaciones que pueden establecerse entre esta técnica y otras (Pérez, 2004, pp. 12-13). (Hair y Taham, 2008, p.80) (Lévy y Varela, 2005, p. 25). Esta técnica puede lograr sus propósitos desde una perspectiva exploratoria o confirmatoria. (Hair y Taham, 2008, p. 81).</p>

Cuadro 3 (Continuación)

Nombre(S) y Objetivo(S)	Naturaleza de las Variables y Supuestos	Comentarios Adicionales
<p>Análisis de Componentes Principales Resumir la mayoría de la información original (varianza) en una cantidad mínima de factores con propósitos de predicción. (Hair y Taham, 2008, p.89) Reducir la dimensión de una tabla de datos excesivamente grande por el elevado número de variables que contiene X_1, X_2, \dots, X_n y quedarse con unas cuantas variables C_1, C_2, \dots, C_p, combinaciones de las iniciales (componentes principales) perfectamente calculables (Pérez, 2004, 12).</p>	<p>Variabes: métricas. Supuestos: Interdependencia: fuerte correlación lineal entre las variables. Multicolinealidad: que unas variables puedan expresarse como combinación lineal de las demás. Normalidad Multivariada: ajuste de cada una de las variables a una distribución normal o al menos la combinación lineal por pares de las mismas. (Landroero y González, 2006, p. 394)</p>	<p>Fue creado por Hotelling (1933) (L y G, p. 389), aunque sus orígenes se encuentran en los ajustes ortogonales por mínimos cuadrados introducidos por Pearson en 1901 (Pérez, 2004, p.134) La principal diferencia entre el análisis factorial común y este análisis radica en el método de extracción de factores. (Lévy y Varela, 2005, p.25). Además, en el factorial se trata de encontrar variables sintéticas, latentes, inobservables, aún no medidas, y cuya existencia se sospecha en las variables originales, mientras que acá las variables sintéticas son combinación de las originales (Pérez, 2004, p.12).</p>
<p>Análisis de Correspondencias Establecer relaciones entre variables no métricas. Estudia la asociación entre las categorías de múltiples variables no métricas pudiendo obtener un mapa perceptual que ponga de manifiesto esta asociación en modo gráfico. (Pérez, 2004, p.14) Persigue dos objetivos básicos: asociación entre categorías de la columna o fila, y asociación entre categorías de filas y columnas (Hair y Taham, 2008, p.250)</p>	<p>Variabes: no métricas. Supuestos: Comparte con otras técnicas de análisis multidimensional una relativa libertad respecto a sus supuestos básicos. Sus datos no métricos representan igualmente bien relaciones lineales y no lineales. Además de ciertos principios relativos a la percepción, el investigador debe esforzarse por conseguir la comparabilidad de los objetos y la generalidad de los atributos utilizados (Hair y Taham, 2008, pp. 553, 561 y 576)</p>	<p>Creado por Benzecri (1973) y extendido por sus estudiantes como Lebart y Greenacre, entre otros (Pérez, 2004, p. 195). Sin embargo, el antecedente más lejano se remonta a Kuder y Richardson (1933) (Lévy y Varela, 2005, p.363). Difiere de otras técnicas de interdependencia en su capacidad para acomodar tantos datos no métricos como relaciones no lineales (Hair y Taham, 2008, p. 13)</p>

Cuadro 3 (Continuación)

Nombre(S) y Objetivo(S)	Naturaleza de las Variables y Supuestos	Comentarios Adicionales
Análisis de Conglomerados (Clusters), Clasificación automática o no supervisada, y de reconocimiento de patrones sin supervisión (Pérez, 2004, p. 220); Análisis Q, Construcción de Tipología, Análisis de Clasificación y Taxonomía Numérica. (Hair y Taham, 2008, p. 492) Cuantifica las características estructurales de un conjunto de observaciones (Hair y Taham, 2008, p.507). Clarificar objetos-sujetos en función de ciertas características (Lévy y Varela, 2005, p.26)	Variables: métricas y no métricas. Supuestos: Por su definición, tiene fuertes propiedades matemáticas más no fundamentos estadísticos. Tienen poco peso acá las exigencias de normalidad, linealidad y homocedasticidad. Sin embargo, e investigador debe centrarse en dos asuntos críticos: la representatividad de la muestra y la multicolinealidad. (Hair y Taham, 2008, p. 508)	Es una técnica que clasifica una muestra de cantidades (personas u objetos) en un número pequeño de grupos de mutuamente excluyentes basados en similitudes entre las entidades. A diferencia del análisis discriminante, los grupos no están predefinidos. (Hair y Taham, 2008, p.13). Es similar el análisis factorial, sin embargo éste solo se emplea para agrupar variables, mientras que el cluster fundamentalmente agrupa objetos, aunque también variables (Lévy y Varela, 2005, p. 26)

Fuente: Elaboración propia, a partir de los autores citados.

Comentarios finales

Tal como se mencionó al inicio, no se clasificaron acá obviamente todas las técnicas existentes, aunque quizás si una buena parte de ellas. Por ejemplo, no se mencionaron algunas calificadas como nuevas técnicas o técnicas emergentes: búsqueda y extracción o minería de datos, redes neuronales, muestra repetida o remuestreo, entre otras.

El estudioso podrá conseguir seguramente similitudes y discrepancias de las clasificaciones aquí presentadas con las de otras fuentes; lo cual es natural, y debería servir para contrastar con buena disposición otros criterios y puntos de vista, ampliando así su espectro dentro del análisis multivariante. Esta Nota pudiera constituir, con todo, una modesta pero significativa referencia.

Referencias bibliográficas

- Hair, Anderson y Taham; Black. (2008). **Análisis Multivariante**. 5^{ta} edición. Pearson Prentice Hall. España.
- Landero, René y González, Mónica. (2006). **Estadística con SPSS y Metodología de la Investigación**. Editorial Trillas. México.
- Lévy, Jean Pierre y Varela, Jesús. (2005). **Análisis Multivariable para las Ciencias Sociales**. Pearson Prentice Hall. Madrid.
- Peña, Daniel. (2002). **Análisis de Datos Multivariante**. Mc Graw-Hill. España.
- Pérez, César. (2004). **Técnicas de Análisis Multivariante de Datos. Aplicaciones con SPSS**. Pearson Prentice Hall. Madrid.