

IMPLEMENTACIÓN DE TÉCNICAS ANÁLISIS MULTIVARIADO EN LA PLATAFORMA SOFTPROSP PARA DETERMINAR CORRELACION DE VARIABLES

IMPLEMENTATION OF TECHNIQUES MULTIVARIATE ANALYSIS IN THE SOFTPROSP PLATFORM TO DETERMINE CORRELATION OF VARIABLES

Raúl Martelo Gómez, Natividad Villabona Gómez y Heybertt Moreno Díaz

¹Programa de Ingeniería de Sistemas, Facultad de Ingeniería
Universidad de Cartagena

Recibido: Julio 10 de 2016 Aceptado: Octubre 15 de 2016

RESUMEN

El objetivo del presente proyecto es implementar la técnica análisis multivariado (Análisis de Componentes Principales ACP) para el acompañamiento de estudios prospectivos en la plataforma SoftProsp. Ésta puede ser aplicada a matrices que contienen grandes dimensiones de datos experimentales con multitud de variables (Rodríguez, Matera y Pérez, 2016). En particular, el Análisis de Componentes Principales (PCA) parte del análisis de factores, donde su finalidad es agrupar variables que tienen correlación entre sí, separándolas de las que no (Quiroga y Villalobos, 2015). En cuanto a los resultados, se pretende optimizar la utilización de ésta técnica con herramientas como el Test de Barlett, el método de rotación ortogonal Varimax y gráficos de sedimentación (Scree-plot), y además, apoyar al desarrollo de análisis multivariado con otras técnicas prospectivas implementadas en la plataforma.

Palabras Clave: Softprocesp, procesos, multivariado

ABSTRACT

The objective of the present project is to implement the multivariate analysis technique (Principal Components Analysis ACP) for the accompaniment of prospective studies in the SoftProsp platform. This can be applied to matrices that contain large dimensions of experimental data with a multitude of variables (Rodríguez, Matera and Pérez, 2016). In particular, the Principal Component Analysis (PCA) starts from the analysis of factors, where its purpose is to group variables that have correlation with each other, separating them from those that do not (Quiroga and Villalobos, 2015). As for the results, it is intended to optimize the use of this technique with tools such as the Barlett test, the Varimax orthogonal rotation method and sedimentation graphs (Scree-plot), and also support the development of multivariate analysis with other techniques prospects implemented in the platform.

Keywords: Softprocesp, processes, multivariate

1. INTRODUCCIÓN

Las técnicas estadísticas multivariadas son cada vez más utilizadas en diferentes ramas de la ciencia. Estas, se caracterizan por permitir adoptar mejores decisiones a partir de la observación y análisis de escenarios bajo condiciones de incertidumbre, aleatoriedad y variabilidad (Vertel, Cepeda y Lugo, 2014).

Existen varias técnicas multivariadas que son clasificadas en consideración de dos criterios, el primero atiende al objetivo científico principal de la investigación, ya sea para extracción de información (enfoque descriptivo) o para generación de conocimiento (enfoque inferencial). Ambos enfoques pueden vincularse a objetivos específicos de reducción de información y obtención de indicadores. El segundo criterio considera la naturaleza o tipo de variables involucradas y las relaciones entre ellas (Dependiente (D); Dependiente, criterios o explicadas(VD); independientes, predictoras o explicativas(VI); interdependiente (I)). (Quero e Inciarte, 2012)

Para el primer enfoque se distinguen las técnicas: Análisis de métodos categóricos, análisis factorial, análisis de correlación canónica, análisis de cluster y el análisis de componentes principales; y para el último: regresión lineal múltiple, análisis discriminante lineal, análisis multivariado de varianza y covarianza (Manova), ecuaciones estructurales y regresión logística (Álvarez, Caballero y Pérez, 2006).

Para el desarrollo de estudios prospectivos son utilizadas técnicas con amplia aplicabilidad en distintas disciplinas científicas, con el propósito de identificar los escenarios futuros más probables y deseables hacia los cuales debe dirigirse una organización, región o país. (Miklos y Arroyo, 2008)

Existen más de 120 técnicas y dentro de estas se resaltan las de análisis multivariado. En Barahona y otros (2016), analizan prospectivamente que factores relacionados con las características del paciente, la cirugía y la litiasis, constituyen efectivamente factores de riesgo preoperatorios e interoperatorios para el Síndrome de respuesta inflamatoria sistémica (SIRS) y sepsis de foco urinario en pacientes sometidos a Nefrolitotomía Percutánea(NLP). Por otro lado, Herrera, García, Suárez y Boirivant (2015) proponen un procedimiento integrador para el análisis retrospectivo y prospectivo de sistemas lecheros. Por último, Torres, Méndez-Fajardo, López-Kleine, Galarza-Molina y Oviedo (2013) evalúan que variables caracterizan el nivel de desarrollo y calidad de vida de las comunidades que habitan las localidades de Bogotá, en el que se realizaron pruebas de correlación y análisis de componentes principales para evaluar el indicador Condiciones sanitarias deficientes y necesidades básicas insatisfechas(NBI).

No obstante, para la elaboración de estudios prospectivos se requiere de un alto presupuesto, debido a la cantidad de recursos y dificultades que se presentan (Cabarcas, Martelo y Tovar, 2013). Por estas razones, entidades como 1 LIPSOR, 2 3IE y 3 EPITA han desarrollado softwares con el propósito de mejorar la forma en que son realizados. Estas herramientas no cubren todas las falencias que se presentan en la forma tradicional de realizar prospectiva, por lo que la Universidad de Cartagena desarrollaron la plataforma SoftProsp con el propósito de cubrir algunas necesidades faltantes (Martelo, Moncaris y Velez, 2016).

SoftProsp utiliza recursos de Ingeniería de Software, Inteligencia Computacional, Sistemas de Información Geográficos y Tecnologías de la Información y las Comunicaciones, por tanto, el propósito del presente documento es informar acerca de los beneficios de la implementación de estadísticas multivariadas en esta plataforma. Ahora, teniendo en cuenta la multiplicidad de técnicas existentes para este tipo de análisis, se consideró implementar el Análisis de componentes principales, esta resume la información original (varianza) en una cantidad mínima de factores con propósitos de predicción.

II. MATERIALES Y MÉTODOS

El presente proyecto se orienta bajo una investigación aplicada y documental. Aplicada porque se estudia un problema que conlleva al desarrollo de un aporte innovador y se están destinando esfuerzos para resolver una de tantas necesidades de la sociedad (Vargas, 2009). Documental porque en la búsqueda de respuestas específicas, se indagan teorías, libros, publicaciones científicas, entre otros documentos asociados a la técnica estadística bayesiana (Baena, 2014).

2.1 Procedimiento de la investigación

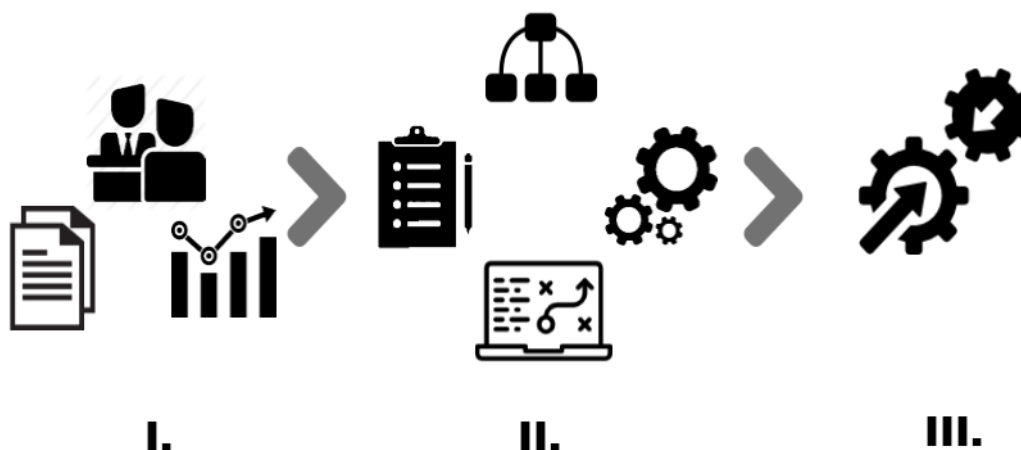


Fig. 1: Procedimiento de la investigación.

Con el propósito de implementar la técnica estadística bayesiana en la plataforma SoftProsp, se ha diseñado el modelo de la Figura 1, el cual dispone de las siguientes etapas:

2.2 Recolección de información

Para la recolección de información se han utilizado las técnicas: entrevista con el propósito de obtener información en relación con un tema determinado, buscar que la información recabada fuera lo más precisa posible y aclarar los significados atribuidos por los informantes en los temas en cuestión (Díaz-Bravo, Torruco-García, Martínez-Hernández y Varela-Ruiz, 2013); y revisión y análisis documental para buscar, seleccionar, organizar y analizar un conjunto de materiales escritos que den respuesta a una o varias incógnitas (Bermeo-Yaffar, Hernández-Mosqueda y Tobón-Tobón, 2016).

2.3 Aplicación de la metodología de desarrollo RUP

Como metodología de desarrollo software se utilizará RUP (Proceso Unificado de Rational), el cual proporciona un acercamiento disciplinado a la asignación de tareas y responsabilidades en una organización de desarrollo y tiene como objetivo principal, asegurar la producción de software de alta calidad que se ajuste a las necesidades de los usuarios finales con costos y calendarios predecibles. Además, se caracteriza por ser una metodología iterativa e incremental y por enfocarse en los casos de uso y la arquitectura de la solución (Martínez y Martínez, 2002).

Las fases de esta metodología de desarrollo son:

- **Inicial**
Se obtienen los requerimientos de la técnica estadística bayesiana. Para esto se realizan revisiones bibliográficas y entrevistas a expertos.
- **Elaboración.**
Abarca el diseño de modelos y diagramas UML a partir de los requerimientos establecidos.
- **Construcción.**
En esta fase se realiza el desarrollo de las funcionalidades, componentes y estructura de datos; también se realiza la documentación técnica a partir de los diagramas, modelos y patrones arquitectónicos establecidos en la fase anterior.
- **Transición**
En esta última fase se realizan las pruebas acorde a los atributos de calidad que más se ajustan al sistema y finalmente se elabora la documentación final del proyecto.

2.4 Integración de estadísticas bayesianas a SoftProsp

Culminado el proceso de desarrollo de la técnica estadística bayesiana se procede a integrar ésta con el grupo de técnicas disponibles en la plataforma SoftProsp. Para verificar el adecuado funcionamiento de estadísticas bayesianas con los otros componentes o módulos del sistema, se realizaron pruebas de integración.

III. RESULTADOS Y DISCUSIÓN

A continuación se presentan los resultados obtenidos a partir de los pasos establecidos en el procedimiento de la investigación:

3.1 Recolección de Información

En virtud de los resultados obtenidos en esta etapa, se pudo definir que las técnicas de análisis multivariado difieren de las uni y bivariantes, en las que se analizan covarianzas o correlaciones que reflejan relaciones entre tres o más variables. Los objetivos que persiguen éstas son proporcionar métodos para el estudio conjunto de datos multivariantes y apoyar en la toma de decisiones teniendo en cuenta la información que aporta el conjunto de datos analizado. (Closas, Arriola, Kuc, Amarilla y Jovanovich, 2013)

Con relación a lo expuesto previamente, Peña (2002) añade que el análisis de datos multivariantes tiene por objeto el estudio estadístico de varias variables medidas en elementos de una población. Además, agrega que estas pretenden resumir un conjunto de variables sin perder la mínima información, encontrar posibles grupos en los datos, clasificar las observaciones en grupos definidos y relacionar dos conjuntos de variables.

Las técnicas multivariantes se pueden clasificar según los siguientes criterios: El primer criterio de clasificación depende del modelo de análisis de la investigación y se establece teniendo en cuenta la finalidad analítica explicativa; el siguiente, es algebraico y técnico-instrumental, como condición previa a toda finalidad analítica; y un último, que conjuga los dos anteriores con la métrica de las variables (técnicas de interdependencia). (López-Roldán y Fachelli, 2015)

De acuerdo a lo anterior, las técnicas multivariantes pueden ser clasificadas en:

- Técnicas de exploración de datos: regresión múltiple, análisis discriminante, métodos log-lineales y logit, análisis de correlación canónica, análisis multivariante de la varianza.
- Técnicas de Inferencia: análisis de componentes principales, análisis factorial, multidimensional scaling, análisis de correspondencias, análisis de cluster.

A partir de la información recopilada, se consideró necesario limitar el alcance en el proceso de implementación e integración de las técnicas Análisis multivariado en la plataforma SoftProsp, debido a la variedad de técnicas existentes. Por ello, el alcance del proyecto abarcaría solo la técnica Análisis de Componentes principales (ACP).

Esta técnica apoya el estudio e interpretación de amplios datos muestrales, se caracteriza por no distinguir entre variables dependientes e independientes y su objetivo principal consiste en identificar las variables y/o objetos de estudio que están relacionados, centrando su atención en cómo lo están. Además, combina las variables buscando elementos comunes y eliminando las redundancias, de forma que las variables iniciales se convierten en un pequeño número de variables artificiales o factores que explican un alto porcentaje de la información. (Unceta, Gutiérrez-Goiria y Goitisoló, 2014)

De Carmona (2014) se pudieron obtener las etapas que conforman la técnica Análisis de Componentes Principales (ver Figura 2). Estas son: a) Cálculo de componentes principales; b) Extracción de factores; c) Cálculo de porcentajes de variabilidad; d) Cálculo de componentes principales a partir de la matriz de correlaciones; e) Matriz factorial; f) Cálculo de covarianzas y correlaciones entre variables originales y los factores; y g) Cambios de escalas e identificación de componentes.

3.2 Aplicación de la metodología de desarrollo RUP

Teniendo en cuenta lo anterior, se consideró establecer para el proceso de Análisis de Componentes Principales (ACP) las características únicas de ésta en la plataforma SoftProsp, en donde se determinaron las siguientes mejoras frente a las soluciones existentes:

- Aplicación del test de Barlett previo al ACP, con el fin de comprobar si la correlación entre las variables analizadas son lo suficientemente grande como para factorizar la matriz de coeficientes de correlación.

- Representación con gráficos de sedimentación (Scree-plot) en la etapa de cálculo de componentes principales. Con el fin de evaluar visualmente cuáles componentes o factores explican la mayor parte de variabilidad de los datos.
 - Implementación del método de rotación ortogonal Varimax en la etapa de extracción de factores, para minimizar el número de variables que tienen saturaciones altas en cada factor y simplificar la interpretación de los factores optimizando la solución por columna.
- Apoyo de la herramienta MICMAC de SoftProsp para contribuir en la comprensión e identificación de problemas y representar el sistema de variables, reduciendo así su complejidad. Ésta, con la ayuda de una matriz de análisis estructural identifica las variables y analiza las relaciones directas entre ellas.
- El módulo Encuesta de la plataforma, puede ser utilizada como una herramienta de recolección de información para, a partir de esta, disponer de la valoración media de una población.

3.3 Integración de estadísticas bayesianas a SoftProsp

En correlación a la última etapa del procedimiento de investigación, donde se procede a integrar el módulo análisis multivariado una vez desarrollada y probada de manera individual a la plataforma SoftProsp, con el fin de garantizar que los cambios no conlleven comportamientos no planeados o aparezcan errores adicionales, se utilizará la prueba de regresión, la cual será realizada manualmente usando herramientas de captura/reproducción automatizadas, permitiendo obtener casos de prueba y resultados para una posterior reproducción y comparación.

En las últimas décadas se ha producido un gran crecimiento en la utilización de estas técnicas en todos los campos de la investigación científica. Algunas de las razones de este fenómeno son: (a) Análisis de relaciones simultáneas entre tres o más variables, (b) Modificación de la hipótesis a partir de la eliminación y adición continua de variables y (c) Desarrollo de programas estadísticos fáciles de usar y con mayor prestación en el análisis de los datos. (Closas, Arriola, Kuc, Amarilla y Jovanovich, 2013)

La importancia de la utilización del análisis de componentes principales, yace de que uno de los problemas principales que afectan al estudio de grandes masas de datos, es que las variables explicativas suelen ser muy parecidas: contienen información equivalente. En efecto, cuando un investigador reúne información sobre cualquier fenómeno tiende a incorporar diversas variables que son semejantes pero no iguales, de modo que el análisis resulta complejo y surgen graves problemas de colinealidad entre las variables X. Así, por ejemplo en regresión múltiple, cuando existe multicolinealidad, no queda más remedio que eliminar algunas variables. Pero eso implica una pequeña pérdida de información. (Góngora, 2010)

IV. CONCLUSIONES

Del ítem anterior se pudieron obtener las siguientes conclusiones:

Disminución de inconvenientes presentados en la realización de estudios prospectivos. Resultados con facilidad de interpretación con Scree-plot. Implementación de

herramientas para la optimización de algunas etapas del Análisis de Componentes Principales. Por último, posibilidad de complementar el PCA con otras técnicas de SoftProsp, MICMAC para la identificación del problema, variables y relaciones directas entre las variables y Encuesta para la recopilación de información para disponer la valoración media de una población.

V. REFERENCIAS BIBLIOGRAFICAS

- Álvarez Suárez, M., Caballero, A., & Pérez Lechuga, G. (2006). Análisis Multivariante: Clasificación, Organización y Validación de los resultados. *Fourth LACCEI International Latin American and Caribbean Conference for Engineering and Technology (LACCET 2006)*, (págs. 21-23). Mayagüez.
- Baena Paz, G. (2014). *Metodología de la investigación*. México: Grupo Editorial Patria.
- Barahona, J., Vinay, J., Martínez, C., Daels, P., Gonzáles, M., Díaz, P., y otros. (2016). Estudio prospectivo y análisis multivariado para detectar factores de riesgo en sepsis de foco urinario y nefrolitotomía. *Revista Chilena de Urología*, 81(3).
- Bermeo-Yaffar, F., Hernández-Mosqueda, J. S., & Tobón-Tobón, S. (2016). Análisis documental de la V Heurística mediante la Cartografía conceptual. *Ra Ximhai*, 12(6), 103-121.
- Cabarcas Álvarez, A., Martelo Gómez, R. J., & Tovar Garrido, L. C. (2013). Software para mejorar la aplicación de técnicas cuantitativas en estudios prospectivos. *Cuadernos de Administración (Universidad del Valle)*, 29(49), 64-74.
- Carmona, F. (2014). *Un ejemplo de ACP paso a paso*. Universitat De Barcelona.
- Closas, A. H., Arriola, E. A., Kuc, C. I., Amarilla, M. R., & Jovanovich, E. C. (2013). Análisis multivariante, conceptos y aplicaciones en Psicología Educativa y Psicometría. *Enfoques*, 25(1), 65-92.
- Coronel de Renolfi, M., & Cardona, G. (2013). Tipificación de Pymes mediante Técnicas de Análisis multivariado. *TEC empresarial*, 3(1-2), 45-54.
- Díaz-Bravo, L., Torruco-García, U., Martínez-Hernández, M., & Varela-Ruiz, M. (2013). La entrevista, recurso flexible y dinámico. *Investigación en educación médica*, 2(7), 162-167.
- Góngora Salazar, P. (2010). *Determinantes de la votalidad en el producto: Evidencia Empírica*. Universidad Javeriana, Bogotá.
- Gutiérrez Pulido, H., Gutiérrez González, P., Garibay López, C., & Díaz Caldera, L. (2012). Análisis multivariado y QFD como herramientas para escuchar la voz del cliente y mejorar la calidad del servicio. *Ingeniare. Revista chilena de ingeniería*, 22(1), 62-73.
- Herrera, J. A., García, A., Suárez, J., & Boirivant, J. A. (2015). Procedimiento para el análisis retrospectivo y prospectivo de sistemas lecheros. *Pastos y Forrajes*, 38(1), 38-45.
- López-Roldán, P., & Fachelli, S. (2015). *Metodología de la Investigación social Cuantitativa*. Barcelona: Bellaterra (Cerdanyola del Vallès): Dipòsit Digital de Documents, Universitat Autònoma de Barcelona.
- Martelo, R., Moncaris, L., & Velez, L. (2016). Integración del Ábaco de Régnier, Encuestas y Lluvia de Ideas en la Definición de Variables Claves en Estudios Prospectivos. *Información tecnológica*, 27(5), 243-250.
- Martínez, A., & Martínez, R. (17 de Diciembre de 2002). *Guía a Rational Unified Process*. Obtenido de Profesora Anaylen López: <https://anaylenlopez.files.wordpress.com/2011/03/trabajo-guia20rup.pdf>

- Miklos, T., & Arroyo, M. (2008). *Prospectiva y escenarios para el cambio social*. México: Convesion Andrés Bello, Universidad Autónoma de México, Facultad de Ciencias Políticas y Sociales.
- Peña, D. (2002). *Análisis de datos multivariantes* (Vol. 24). Madrid: McGraw-Hill.
- Popper, R. (2008). Foresight methodology. En L. Georghiou, J. Cassingena, M. Keenan, I. Miles, & R. Popper, *The Handbook the technology foresigth Concepts and Practice*. Edward Elgar Publishing Limited.
- Quero Virla, M., & Inciarte Pirela, K. (2012). Clasificación de las Técnicas Estadísticas Multivariantes. *Telos*, 14(2), 275-287.
- Quiroga Juárez, C. A., & Villalobos Escobedo, A. (2015). Análisis del comportamiento bursátil de las principales bolsas financieras en el mundo usando el análisis multivariado (Análisis de componentes principales PCA) para el perioso de 2011 a 2014. *Revista CEA*, 1(2), 25-36.
- Rodríguez, E., Matera, J., & Pérez, M. (2016). Determinación de variables en problemas multivariantes. Método: SIMPLISMA. *Revista Arbitrada Venezolana del Núcleo Costa Oriental del Lago*, 11(2), 137-154.
- Torres, A., Méndez-Fajardo, S., López-Kleine, L., Galarza-Molina, S., & Oviedo, N. (2013). Calidad de vida y ciudad: análisis del nivel de desarrollo en Bogotá a través del método de necesidades básicas insatisfechas. *Estudios Gerenciales*, 29(127), 231-238.
- Unceta Satrustegui, K., Gutiérrez-Goiria, J., & Goitisoló Lezama, B. (2014). Evidencias e interrogantes sobre desarrollo, financiación externa y AOD: un análisis de componentes principales. *Revista de economía mundial*, 36.
- Vargas Cordero, Z. R. (2009). La investigación aplicada: una forma de conocer las realidades con evidencia científica. *Revista Educación*, 33(1), 155-165.
- Vertel Morinson, M. L., Cepeda Coronado, J. A., & Lugo Hernández, E. A. (2014). Análisis Multivariado de la Calidad educativa en Sucre. *Scientia et Technica*, 19(1), 96-105.