

TRABAJO 3 RESEÑA

KEVIN STEVEN GARCÍA^a, CESAR ANDRES SAAVEDRA^b

1. Un contraste de hipótesis para datos funcionales

2. Análisis de correspondencias múltiples en presencia de datos faltantes: el principio de datos disponibles del algoritmo NIPALS (ACMpdd)

El objetivo principal de esta tesis de maestría es proponer un método de análisis de correspondencias múltiples para matrices con datos faltantes. Para ello, se propone un método denominado ACMpdd, el cuál trabaja bajo el principio de datos disponibles del algoritmo NIPALS.

Primero, debemos saber que es el análisis de correspondencias múltiples y que es el algoritmo NIPALS, para tener una idea general sobre que tipo de problemas se pueden abordar bajo este método propuesto.

El análisis de correspondencias múltiples ACM, es un método estadístico multivariado (estudia tres o más variables al mismo tiempo) que sirve para analizar variables cualitativas (variables que expresan distintas características, cualidades o modalidades), el problema fundamental que posee este método es que para realizarlo se necesitan las bases de datos totalmente completas, es decir, no admite datos faltantes, lo cuál ocurre con demasiada frecuencia en la práctica, sobre todo en las encuestas; esto hace que este método solo se pueda usar en las condiciones ideales (datos completos). Aunque se han propuesto soluciones a este problema como eliminar los registros donde se produzca el dato faltante o la imputación del dato faltante (básicamente consiste en estimar los datos faltantes a partir de los datos disponibles), estos tienen muchas críticas y muchos problemas, el primero (eliminar registros), ha sido criticado porque significaría eliminar toda una fila o una columna de datos, lo cuál no solo generaría pérdida de información sino también pérdida de recursos; y el segundo (imputación del dato faltante), al hacer las estimaciones de los datos faltantes a partir de los datos disponibles, 'hala hacia el montón', es decir, no son buenas estimaciones, ya que estas se ven muy afectadas por las categorías con mayor frecuencia (más repeticiones).

La otra solución es la que se trabaja en esta tesis que es utilizar un algoritmo que permita trabajar con la presencia de éste tipo de datos (principio de datos disponibles), ese algoritmo es el algoritmo NIPALS, el cuál realiza una descomposición singular de la matriz de datos, mediante secuencias iterativas de proyecciones ortogonales. Esta solución bajo nuestra perspectiva es la mejor, ya que se trata de 'exprimir' al máximo los datos, dando uso a cada uno de ellos. Al aplicar este algoritmo a matrices de datos completas, el resultado es equivalente al ACM, pero, su mayor virtud es que se puede utilizar con datos faltantes y obtener sus estimaciones a partir de una reconstrucción de la matriz de datos. Además de esto, el algoritmo NIPALS es de vital importancia en la regresión PLS (minimos cuadrados parciales), ya que este sirve para descomponer la matriz y así obtener las componentes ortogonales sobre las cuales se realizara la regresión. Entonces, indirectamente el algoritmo NIPALS es de vital importancia en el proceso de estimar modelos de regresión con presencia de multicolinealidad, matrices sobredefinidas (con más variables que datos), y matrices con datos faltantes.

El

^aCódigo: 1533173. E-mail: kevin.chica@correounivalle.edu.co

^bCódigo: 1628466. E-mail: cesar.saavedra@correounivalle.edu.co