

Resumen

En el universo de la ciencia de datos el STATIS-R, metodología desarrollada en esta investigación, se inserta en el área de análisis exploratorio, específicamente para datos cuantitativos no simétricos. Utilizando como base el STATIS (desarrollada por Escoufier- Lavit, 1985), se propone un símil de la estrategia denominada ICI (Interestructura - Compromiso - Intraestructura) para este caso en particular (cuantitativo no simétrico).

Objetivos

Adaptar la metodología STATIS para el caso cuantitativo no simétrico.

- Definir la base conceptual que permita la construcción de la estrategia denominada ICI (Interestructura - Compromiso - Intraestructura) para este caso en particular.
- Elaborar programas de computación ad hoc para poner en práctica la propuesta desarrollada y probar la pertinencia de los resultados.

Metodología

- Para el desarrollo del fundamento teórico de la aplicación se hace la determinación del objeto representativo, luego se hace la definición el producto escalar y finalmente se realiza la conceptualización de la distancia estadística obtenida a partir del producto escalar.
- Creación de algoritmos en lenguaje R, obtención de datos, data wrangling y análisis de los resultados obtenidos.

Resultados y análisis

Al realizar análisis de redundancia individualizados (para cada ocasión), un aspecto importante a considerar, es que los ejes principales de cada análisis de redundancia no tienen el mismo significado. Por lo tanto, no es posible comparar las posiciones relativas entre las variables en cada ocasión a través de biplots individuales. Surge así la necesidad de obtener un eje de representación único que resuma la información simultanea de las K ocasiones.

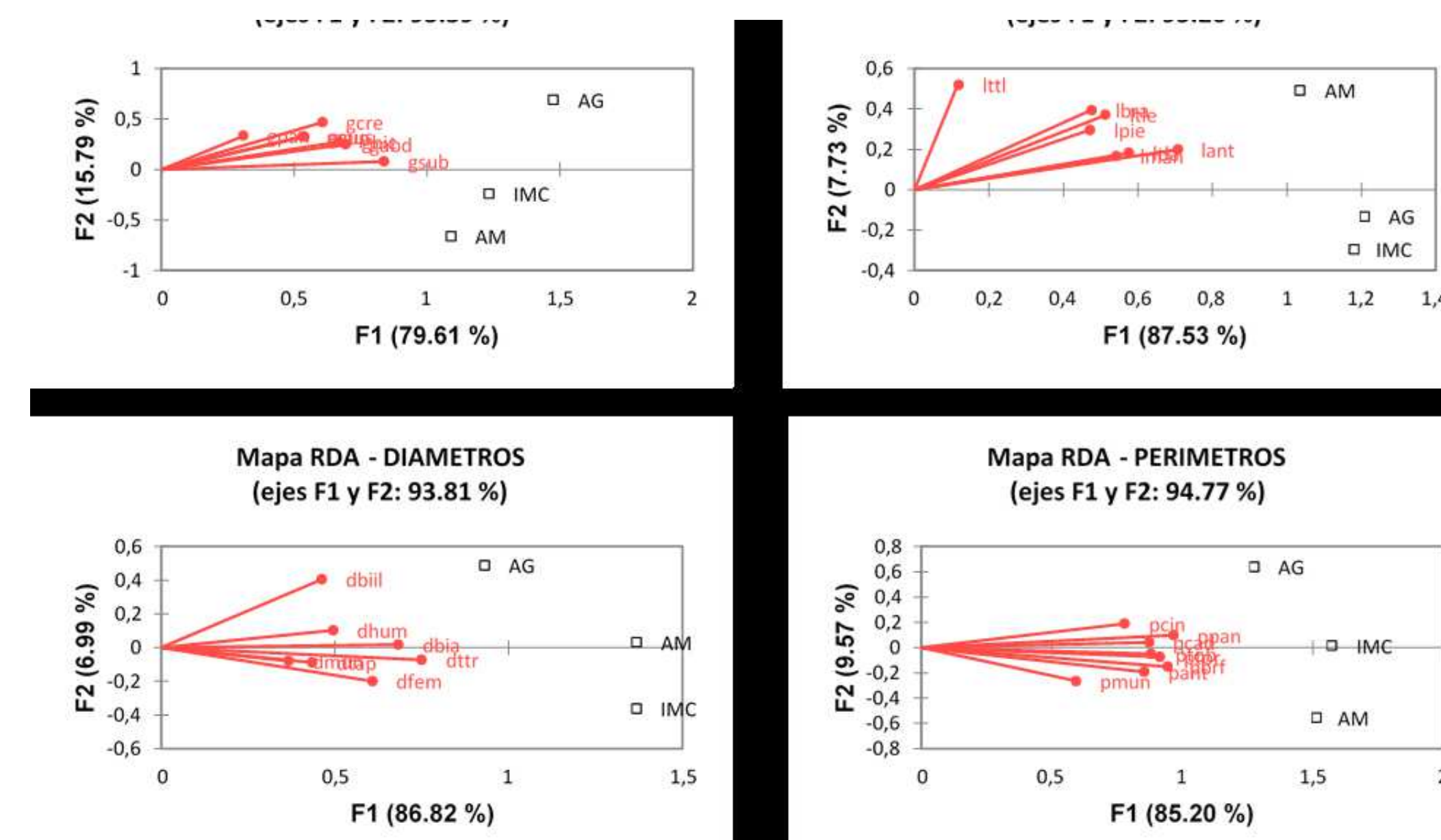


Figura 1: Comparación de cubierta verdadera y predicha

La revisión de la descomposición espectral del objeto compromiso permite garantizar que con las dos primeras direcciones se capturara el 93% de la variabilidad contenida en el mismo. Reflejando así en una sola figura, toda la información de la interrelación entre las variables en las K ocasiones.

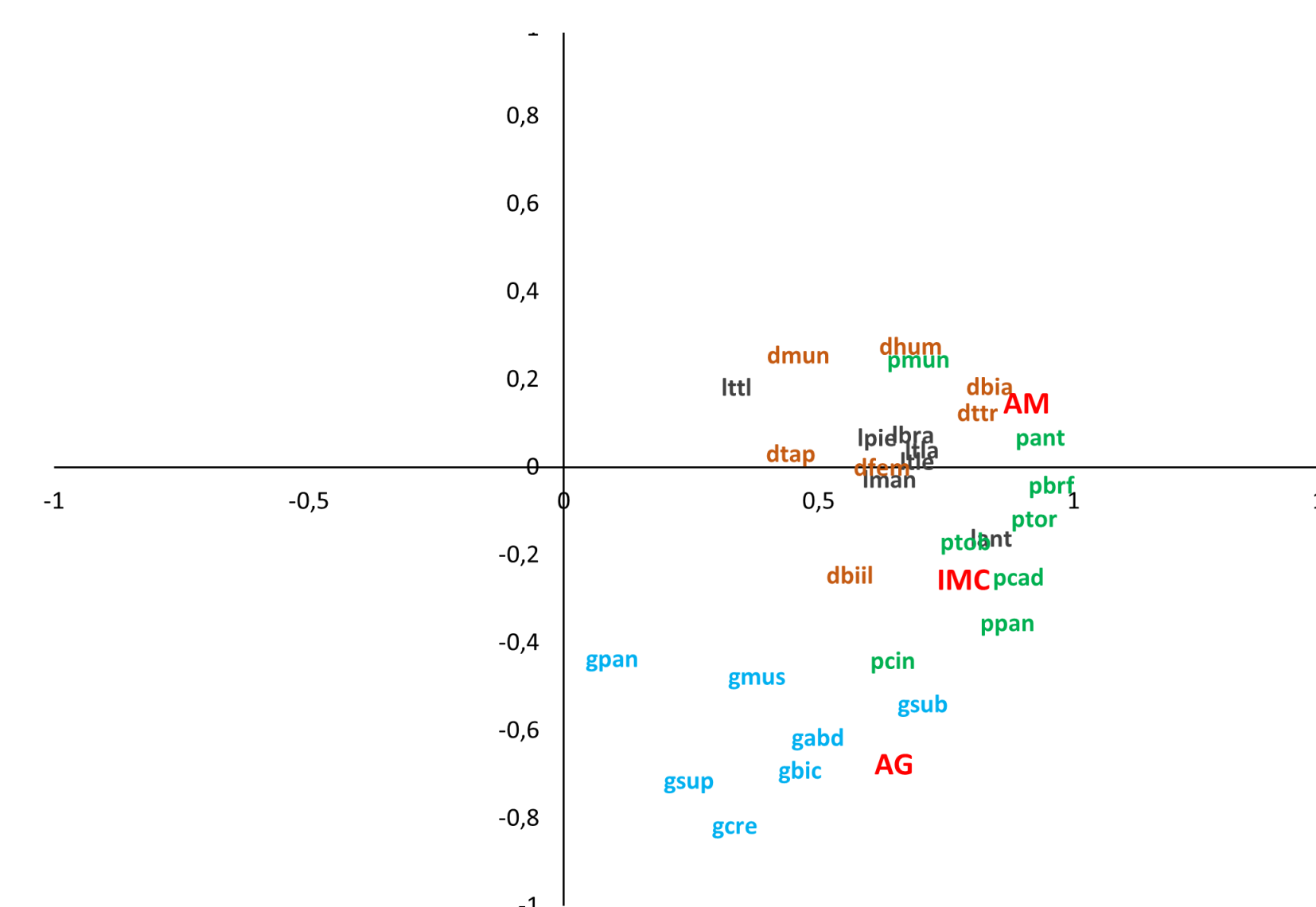


Figura 2: Espacio compromiso obtenido con la metodología STATIS-R

Conclusiones

La propuesta STATIS-R (STATIS de Redundancia), permite representar en un espacio de baja dimensión la información contenida en múltiples bloques pareados, respetando su carácter asimétrico, para comparar las condiciones que generan la redundancia de unas variables sobre otras.

Se explora la posibilidad de crear un espacio compromiso de representación común para todas las variables de los diversos bloques explicativos y bloques criterio crea un dispositivo grafico que permite comparar la forma en que se interrelacionan todas las variables según un criterio como el índice de redundancia, facilitando su interpretación simultánea, pero respetando su rol asimétrico.

Como en cualquier representación factorial en espacio reducido, es necesario tomar en consideración la calidad de representación de cada objeto en el plano, para evitar conclusiones erróneas.

Referencias

- Artículo Statis-R
- Datos utilizados en la investigación
- Código STATIS-R