¿ES NECESARIO UTILIZAR DYNAMIC TIME WARPING PARA LA AGRUPACIÓN DE DATOS LONGITUDINALES? Un estudio de simulación de Montecarlo

BELTRÁN, Celina[1].; BARBONA, Ivana[1]; BOCADI, Eugenia[1]; CIMINARI, Jesica[2]; ORTEGA, Agustina[1]; SUÁREZ, Josefina[1]

[1]Cátedra de Estadística. Facultad de Ciencias Agrarias. UNR
[2]Cátedra de Bioestadística. Facultad de Ciencias Veterinarias. UNR

Palabras claves: Clustering, Machine learning, distancia.

INTRODUCCIÓN

La elección de una medida de distancia es clave en el clustering de datos longitudinales. Dynamic Time Warping (DTW) es muy popular para alinear trayectorias desfasadas, pero se desconoce su ventaja real cuando el número de mediciones es reducido (T ≤ 8).

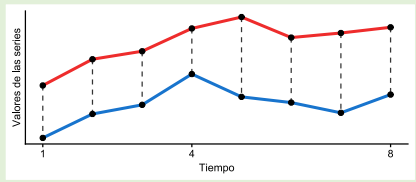
OBJETIVO

Evaluar si la distancia DTW ofrece ventajas frente a la distancia Euclídea en el agrupamiento jerárquico de series longitudinales cortas.

DISEÑO DE LA SIMULACIÓN

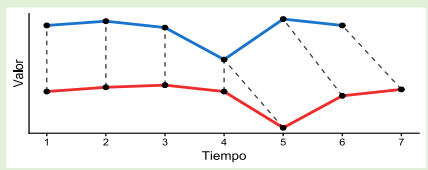
- **3 grupos subyacentes** (k = 3)
- **3 tamaños muestrales** n = 30, 50, 100
- **Número de mediciones** T = 4, 6, 8
- **Estructura de covariancia intra-sujeto** CS, AR(1), UN
- **Patrón temporal** Lineal, no lineal, con desplazamiento temporal
- **Clustering jerárquico**
- **Distancias:** Euclídea DTW
- **Métodos de enlace:** Promedio Completo
- **Evaluación:** ARI AMI

¿CÓMO SE COMPARAN LAS TRAYECTORIAS?



Distancia euclídea

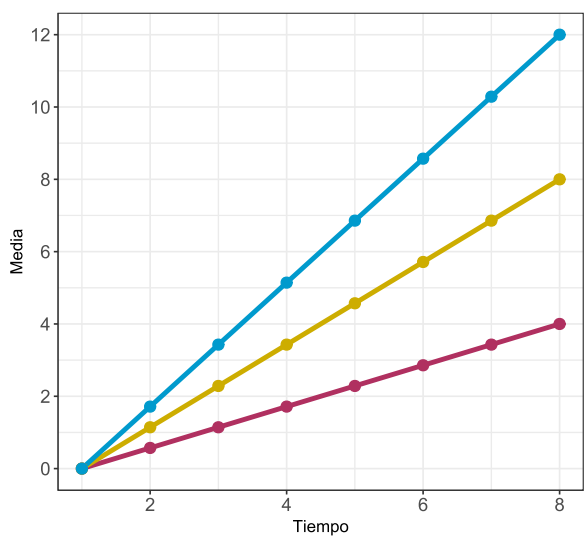
Compara los valores en los mismos momentos de tiempo



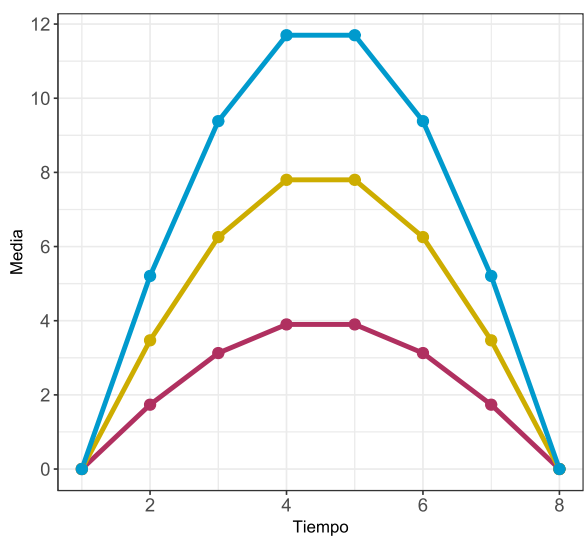
Dynamic Time Warping (DTW)

Permite alinear trayectorias desfasadas en el tiempo

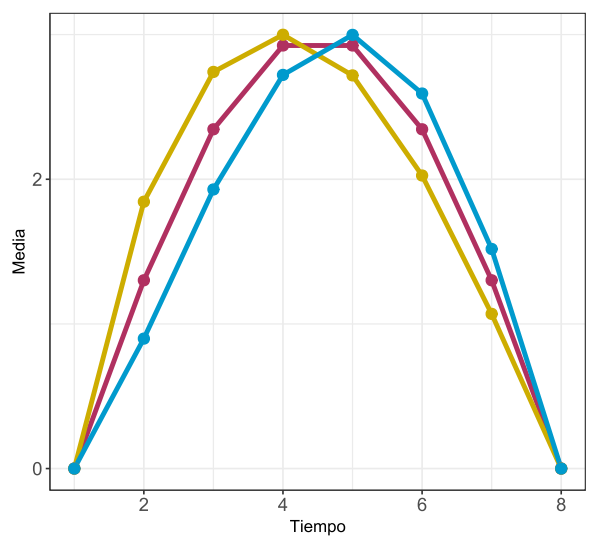
PATRONES DE TRAYECTORIAS SIMULADAS



DTW mostró una ventaja consistente



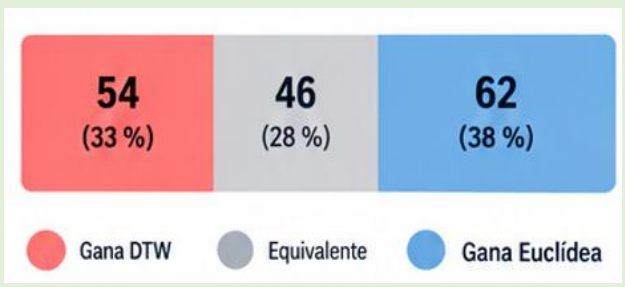
La distancia euclídea resultó más adecuada



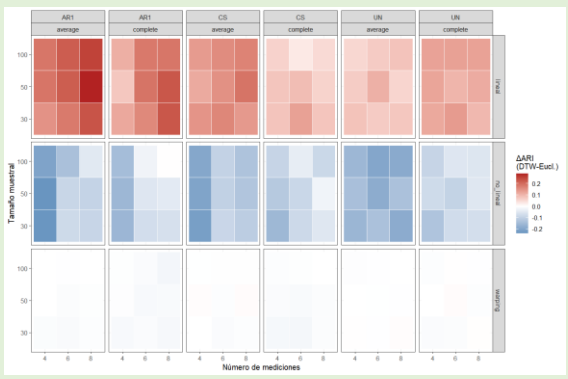
Con sólo 4 – 8 mediciones, ambos métodos resultaron prácticamente equivalentes

RESUMEN GENERAL DE RESULTADOS

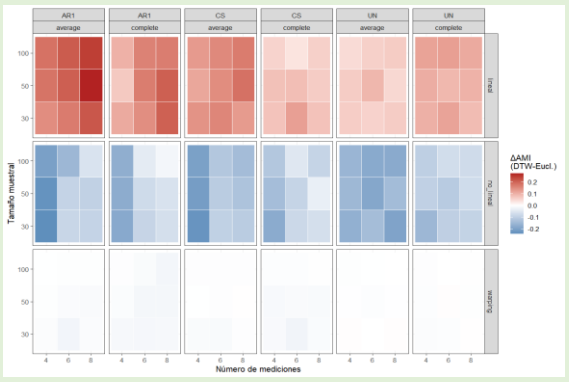
De los 162 escenarios evaluados:



RESULTADOS DETALLADOS

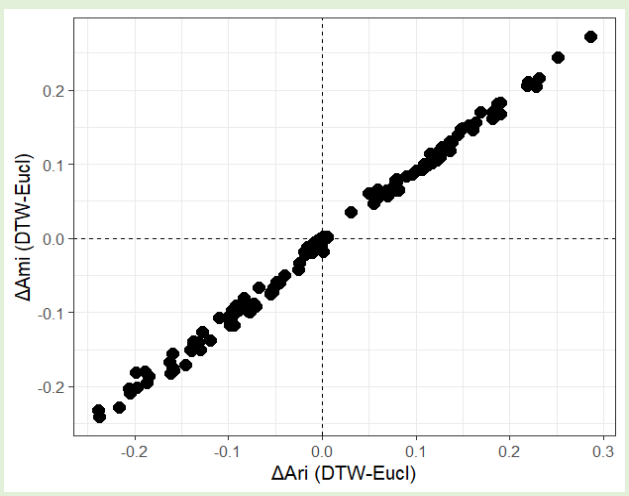


Diferencia en ARI (DTW – Euclídea)



Diferencia en AMI (DTW – Euclídea)

CONSISTENCIA ENTRE ÍNDICES



Cada punto representa un escenario simulado. ARI y AMI muestran **resultados consistentes** a lo largo de todos los escenarios.

CONCLUSIONES

- El desempeño relativo de ambas distancias depende principalmente del patrón temporal de las trayectorias.
- DTW muestra un mejor desempeño para trayectorias lineales, la distancia euclídea para trayectorias no lineales, y en presencia de desalineamiento temporal ambos métodos presentan desempeños prácticamente equivalentes.
- Para datos longitudinales con pocas mediciones, la elección entre DTW y la distancia euclídea depende del patrón temporal de las trayectorias, siendo limitada la ventaja práctica de DTW en las condiciones evaluadas.

En conjunto, los resultados indican que no es imprescindible utilizar DTW para la agrupación de datos longitudinales cortos.

REFERENCIAS