

La confiabilidad interevaluador (Inter-Rater Reliability, IRR) es un aspecto crítico en estudios de codificación, evaluación de intervenciones y análisis de datos observacionales. En R, el cálculo del Coeficiente de Correlación Intraclass (ICC) suele realizarse mediante la función `psych::ICC()`, que implementa un enfoque clásico basado en ANOVA. Sin embargo, este procedimiento requiere matrices completas, eliminando datos faltantes y limitando la flexibilidad del análisis.

En esta presentación mostramos una comparación paso a paso entre el cálculo clásico con `psych::ICC()` y un enfoque alternativo basado en modelos de efectos mixtos (`lme4`). Utilizamos un conjunto de datos real de codificación de mensajes por múltiples evaluadores y extraemos la varianza atribuible a mensajes, codificadores, conversaciones y error residual. Presentamos las fórmulas para el ICC de un evaluador y para el promedio de k evaluadores, discutiendo cómo se interpretan los componentes de varianza en cada método.

Los resultados evidencian que ambos enfoques producen estimaciones prácticamente equivalentes, pero que los modelos de efectos mixtos ofrecen ventajas adicionales: permiten trabajar con datos desbalanceados, proveen estimaciones mediante máxima verosimilitud y facilitan la descomposición de varianza en múltiples niveles. Concluimos con recomendaciones prácticas para investigadores que buscan reportar la confiabilidad interevaluador de manera más transparente e informativa en sus análisis con R.