**Análisis de Datos Clínicos en Cáncer de Tiroides: Manejo de Datos Faltantes e Imputación Múltiple**

**Introducción**

El análisis de datos clínicos requiere un enfoque riguroso, especialmente cuando se trabaja con información incompleta o con múltiples variables categóricas y continuas. En esta presentación se expone un flujo de trabajo completo aplicado a un conjunto de datos sobre cáncer de tiroides, el cual incluye información demográfica, clínica, histopatológica y de evolución de los pacientes.

El objetivo principal es evaluar asociaciones clínicas relevantes y construir modelos estadísticos válidos, manejando adecuadamente los valores perdidos mediante imputación múltiple. Para ello, se realizaron procedimientos de limpieza, diagnóstico de datos faltantes, análisis comparativos, imputación con el método MICE y análisis de regresión.

Este enfoque busca maximizar el aprovechamiento de los datos disponibles, minimizar el sesgo y obtener conclusiones robustas que puedan contribuir a la comprensión del comportamiento clínico del cáncer de tiroides.

**Carga y Preparación de los Datos**  
Antes de cualquier análisis, es fundamental cargar el conjunto de datos correctamente, revisar su estructura y asegurarnos de que las variables estén codificadas adecuadamente. Esta etapa permite identificar posibles inconsistencias en nombres, tipos de datos o formatos.

**Identificación de Valores Perdidos**

**Introducción:**

Los datos faltantes pueden sesgar los resultados si no se manejan adecuadamente. En esta fase, identificamos cuántos valores faltan por variable y visualizamos patrones de ausencia. Esto nos ayuda a decidir si es necesario imputarlos o si el análisis puede continuar con datos completos.

**Preparación para la Imputación de Datos**  
Antes de aplicar métodos de imputación, es importante seleccionar las variables relevantes para el análisis y transformar adecuadamente aquellas que son categóricas. Esta etapa asegura que el proceso de imputación sea coherente con la naturaleza de los datos.



**Imputación Usando MICE**  
Se utiliza la imputación múltiple con el paquete mice para reemplazar los valores perdidos de forma estadísticamente robusta. Este método crea varias versiones del conjunto de datos con imputaciones diferentes, lo que permite conservar la variabilidad y mejorar la validez del análisis posterior.

**Exploración de los Datos Imputados**  
Una vez imputados los datos, es necesario verificar que los valores generados sean plausibles. Mediante gráficos, comparamos la distribución de los datos observados e imputados para asegurar que no se hayan introducido sesgos artificiales.

**Análisis de Regresión con Datos Imputados**  
Con los datos imputados, realizamos un análisis de regresión para estudiar la asociación entre variables clínicas y desenlaces de interés como la recurrencia o la respuesta al tratamiento. Al combinar resultados de múltiples imputaciones, obtenemos estimaciones más precisas y confiables.

**Conclusión**

El manejo adecuado de los datos faltantes es esencial en los estudios clínicos, ya que la omisión o el tratamiento inadecuado de estos puede conducir a resultados sesgados o poco representativos. En este análisis, se implementó un flujo de trabajo riguroso que incluyó la detección de valores perdidos, la comparación entre grupos, la imputación múltiple mediante el método MICE y el análisis de regresión con datos completos.

Este enfoque permitió conservar la mayor cantidad posible de información sin comprometer la validez estadística, mejorando la precisión de las estimaciones y aumentando la potencia analítica. La visualización y validación de los datos imputados reforzaron la confianza en los resultados obtenidos.

En conjunto, este proceso demuestra cómo la integración de buenas prácticas estadísticas puede fortalecer el análisis clínico y generar evidencia más robusta para la toma de decisiones médicas en el contexto del cáncer de tiroides.