

Trabajo práctico 1

Validación de datos

Cátedra de Bioestadística

25 de marzo de 2024

Introducción

En un ensayo clínico aleatorizado se comparan dos drogas para tratar la hemorragia posparto en mujeres embarazadas. Durante el período de reclutamiento, todas las mujeres en trabajo de parto temprano admitidas en alguno de los hospitales participantes son invitadas a formar parte del estudio. Si la mujer consiente a participar y cumple con todos los criterios de elegibilidad, se le asigna aleatoriamente uno de los dos tratamientos y se la sigue por 24 horas o hasta que es dada de alta del hospital (lo que ocurra primero).

La información necesaria para llevar a cabo el análisis se recolecta a lo largo de nueve formularios en papel. En particular, el formulario de parto (DEL) registra las características basales de la mujer e información adicional sobre el nacimiento y administración del tratamiento experimental (ver **Formulario de parto.pdf**).

Consigna

Tiene en su poder la base de datos correspondiente a la primera sección del formulario de parto (*Información de la mujer*) de las primeras 50 mujeres reclutadas. La base se encuentra disponible en el archivo **parto_sec1_n50.xlsx**. Su objetivo es llevar a cabo un proceso exhaustivo de validación de datos, empleando las técnicas desarrolladas en clase para evaluar la calidad de los datos ingresados y detectar errores sistemáticos en el llenado de los formularios.

1. En base al formulario, plantee un conjunto de reglas de validación que le permita detectar inconsistencias en los datos ingresados
2. Pruebe el set de validación contra la base de participantes y registre los errores detectados
3. Presente un reporte de los resultados obtenidos. Además, informe:
 - Participantes limpios (sin inconsistencias)
 - Participantes con más inconsistencias
 - Errores más frecuentes

Sobre los datos

- La base está en formato crudo (variables codificadas, sin etiquetas)
- Los nombres de las variables se componen por las tres letras que identifican el formulario, seguido de la numeración de la pregunta (*delo3* es la pregunta 3 del formulario de parto)
- *part_id* es el ID único que identifica a la participante