

**UNIVERSIDAD PRIVADA SAN JUAN BAUTISTA FACULTAD DE CIENCIAS
DE LA SALUD ESCUELA PROFESIONAL DE MEDICINA HUMANA**



ASIGNATURA:
SISTEMATIZACIÓN Y MÉTODOS ESTADÍSTICOS

DOCENTE:
CASTRO LOPEZ, SEGUNDO VICENTE

ESTUDIANTES:
CARDENAS SANCHEZ, ALAI ALEXIS
CHOLÁN SÁNCHEZ, ARIANA MARICIELO
PEREZ FLORES, ANDREA XIMENA
ORTIZ ALVAREZ, MARÍA FERNANDA
PINTO COAQUIRA, INES

SEDE Y TURNO
SAN BORJA - MA

2025

INSTALAR Y CARGAR

```
{r}  
install.packages("mice")  
install.packages("ggmice")
```

```
{r}  
library(mice)  
library(tidyverse)  
library(here)  
library(rio)  
library(ggmice)  
library(gtsummary)
```

1. DATOS PERDIDOS EN INVESTIGACIÓN

Los datos perdidos se refieren a valores que están ausentes en un conjunto de datos. En investigaciones en salud, esto ocurre cuando no se dispone de información para ciertas variables en algunas observaciones (pacientes, encuestados, etc.).

2. IMPUTACIÓN DE DATOS

Es una técnica estadística que se utiliza para manejar los valores perdidos en una base de datos. Consiste en reemplazar los datos faltantes con valores estimados, con el objetivo de aprovechar al máximo la información disponible y evitar la pérdida de observaciones valiosas en el análisis.

En el contexto de la salud, donde la recolección de datos puede depender de historias clínicas incompletas o encuestas con respuestas omitidas, la imputación permite mantener un mayor número de casos en los análisis, mejorando así la precisión y la validez de los resultados.

3. Dataset "Autismo"

Para ilustrar el proceso de imputación múltiple de datos, utilizaremos el conjunto de datos "data_autismo". Este dataset incluye información de 292 niños/as evaluados mediante un test para la detección de autismo. Las variables registradas comprenden: ID (Identificador del paciente), Puntuacion_P1 a Puntuacion_P10 (Ítems individuales de un test de tamizaje para autismo), Edad (en años), Genero (Masculino/Femenino), Etnicidad (Grupo étnico), Ictericia_al_nacer (Sí/No), Diagnostico_autismo (Sí/No), Pais_residencia (País de residencia), Uso_prev_aplicacion (Sí/No), Resultado (Puntaje total del test), Descripcion_edad (4–11 años), Quien_completa_prueba (Padre/Madre, etc.) y Diagnostico_ASD (Sí/No). Algunos participantes presentan valores faltantes en al menos una de estas variables.

Cargamos los datos

```
{r}
data_autismo <- import(here("data", "autismo.csv"))
```

Registered S3 method overwritten by 'data.table':
method from
print.data.table

Un vistazo a los datos

```
{r}
head(data_autismo)
```

Description: df [6 × 22]

	Puntuacion_P9 <int>	Puntuacion_P... <int>	Edad <chr>	Genero <chr>	Etnicidad <chr>
	0	0	6	Masculino	Otros
	0	0	6	Masculino	Oriente Medio
	0	0	6	Masculino	Desconocido
	0	1	5	Femenino	Desconocido
	1	1	5	Masculino	Otros
	0	1	4	Masculino	Desconocido

6 rows | 11-15 of 22 columns

4. Imputación de datos

Usamos este código porque nos mostrará los valores únicos de todas las variables de tipo character (texto) en tu dataset data_autismo.

```
{r}
sapply(data_autismo[, sapply(data_autismo, is.character)], unique)
```

\$Edad
[1] "6" "5" "4" "11" "10" "8" "7" "9" "?"

\$Genero
[1] "Masculino" "Femenino"

\$Etnicidad
[1] "Otros" "Oriente Medio" "Desconocido" "Blanco-
Europeo" "Negro"
[6] "Asiático del sur" "Asiático" "Pacífico" "Hispano"
"Turco"
[11] "Latino"

\$Ictericia_al_nacer
[1] "No" "Sí"

\$Diagnostico_autismo
[1] "No" "Sí"

INTERPRETACIÓN:

- La edad se encuentra codificada como texto. Aparecen edades entre 4 y 11 años, pero también se detecta el valor "?", lo cual sugiere un valor problemático.

Luego:

```
{r}  
data_autismo[data_autismo == "?"] <- NA
```

Esto convierte todos los valores "?" en el dataset "data_autismo" a NA, sin importar en qué columna estén.

4.1 ¿Dónde están los valores perdidos?

función "(colSums(is.na()))", para saber en qué variables se encuentran los datos antes de iniciar la imputación.

```
{r}  
colSums(is.na(data_autismo))
```

ID	Puntuacion_P1	Puntuacion_P2	Puntuacion_P3	Puntuacion_P4
0	0	0	0	0
Puntuacion_P5	Puntuacion_P6	Puntuacion_P7	Puntuacion_P8	Puntuacion_P9
0	0	0	0	0
Puntuacion_P10	Edad	Genero	Etnicidad	Ictericia_al_nacer
0	4	0	0	0
Diagnostico_autismo	Pais_residencia	Uso_prev_aplicacion	Resultado	Descripcion_edad
0	0	0	0	0
Quien_completa_prueba	Diagnostico_ASD			
0	0			

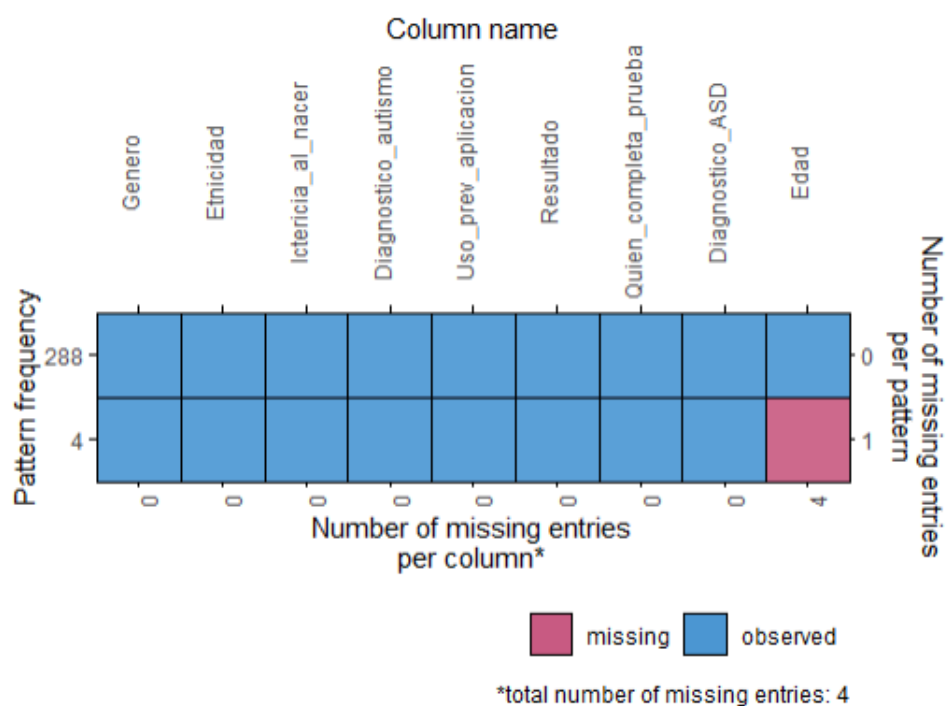
INTERPRETACIÓN:

La mayoría de las variables están completas, lo que es positivo para el análisis, solo la variable "Edad" contiene datos faltantes.

Las puntuaciones (P1 a P10), información demográfica (Género, Etnicidad), y variables relacionadas con el diagnóstico no presentan valores perdidos.

Funcion "plot_pattern()" de ggmic", para visualizar los datos perdidos en un mapa de calor

```
{r}
data_autismo |>
  select(
    Edad,
    Genero,
    Etnicidad,
    Ictericia_al_nacer,
    Diagnostico_autismo,
    Uso_prev_aplicacion,
    Resultado,
    Quien_completa_prueba,
    Diagnostico_ASD
  ) |>
  ggml::plot_pattern(
    square = TRUE,
    rotate = TRUE
  )
```



INTERPRETACIÓN:

- En este caso, se observa que solo la variable “Edad” presenta valores perdidos, específicamente 4 datos ausentes, mientras que el resto de variables no tiene valores faltantes. Esto también se confirma en el eje derecho del gráfico, que muestra que hay un único patrón de datos faltantes (1 fila con 4 casos), todos correspondientes a “Edad”. Hay 4 pacientes (niños) quienes tienen valores perdidos en una variable.

4.2 Comparación de participantes con y sin valores perdidos

Antes de realizar la imputación de datos es evaluar si existen diferencias sistemáticas entre los participantes con valores perdidos y aquellos sin datos perdidos. Esto es importante porque

permite valorar si los datos están faltando al azar o si su ausencia podría estar relacionada con otras variables del estudio.

En nuestro caso, la única variable con datos faltantes es "Edad". Por lo tanto, evaluamos si las demás variables, como: género, ictericia al nacer, diagnóstico de autismo, puntaje total del test (Resultado), entre otras, difieren significativamente entre los participantes con "Edad" conocida y aquellos con "Edad" faltante.

Si no existen diferencias importantes entre estos grupos, es razonable asumir que los datos faltan al azar, y podría utilizarse un análisis de casos completos sin comprometer los resultados.

```
{r}
tabla_edad <- data_autismo |>
  dplyr::select(
    Edad,
    Genero,
    Etnicidad,
    Ictericia_al_nacer,
    Diagnostico_autismo,
    Pais_residencia,
    Uso_prev_aplicacion,
    Resultado,
    Descripcion_edad,
    Quien_completa_prueba,
    Diagnostico_ASD
  ) |>
  mutate(
    missing = factor(
      is.na(Edad),
      levels = c(FALSE, TRUE),
      labels = c("Sin valores perdidos", "Con valores perdidos")
    )
  ) |>
  tbl_summary(
    by = missing,
    statistic = list(
      all_continuous() ~ "{mean} ({sd})",
      all_categorical() ~ "{n} ({p}%)"
    )
  ) |>
  modify_header(
    label = "***variable***",
    all_stat_cols() ~ "***{level}***<br>N = {n} ({style_percent(p, digits = 1)}%"
  )"
  ) |>
  modify_caption("***Características de los participantes según disponibilidad
del dato Edad***") |>
  bold_labels()
```

tabla_edad	Large tbl_summary (5 elements, 12.2 MB)	🔍
------------	---	---

```
{r}
tabla_edad
```

Características de los participantes según disponibilidad del dato
Edad

Variable	Sin valores perdidos	Con valores perdidos
	N = 288 (98.6%) [†]	N = 4 (1.37%) [†]
Edad		
10	18 (6.3%)	0 (NA%)
11	26 (9.0%)	0 (NA%)
4	92 (32%)	0 (NA%)
5	45 (16%)	0 (NA%)
6	39 (14%)	0 (NA%)
7	27 (9.4%)	0 (NA%)
8	21 (7.3%)	0 (NA%)

9	20 (6.9%)	0 (NA%)
Unknown	0	4
Genero		
Femenino	84 (29%)	0 (0%)
Masculino	204 (71%)	4 (100%)
Etnicidad		
Asiático	46 (16%)	0 (0%)
Asiático del Sur	21 (7.3%)	0 (0%)
Blanco-Europeo	108 (38%)	0 (0%)
Desconocido	40 (14%)	3 (75%)
Hispano	7 (2.4%)	0 (0%)

Ictericia_al_nacer

No	210 (73%)	2 (50%)
Sí	78 (27%)	2 (50%)
Diagnostico_autismo		
No	239 (83%)	4 (100%)
Sí	49 (17%)	0 (0%)

Uso_prev_aplicacion		
No	278 (97%)	3 (75%)
Sí	10 (3.5%)	1 (25%)
Resultado	6 (2)	5 (3)
Descripcion_edad		
4-11 years	288 (100%)	4 (100%)

INTERPRETACIÓN:

De los 292 participantes, 288 (98.6%) tienen la edad registrada, mientras que 4 (1.37%) presentan valores perdidos. Entre quienes tienen datos completos, las edades se distribuyen principalmente entre los 4 y 11 años, siendo la edad más frecuente los 4 años (32%).

En cuanto al género, el grupo con edad conocida está compuesto mayoritariamente por varones (71%), mientras que el grupo con edad perdida está conformado exclusivamente por varones (100%).

Respecto a la etnicidad, el grupo con datos completos presenta una mayor proporción de participantes blanco-europeos (38%), mientras que en el grupo con edad faltante, el 75% tiene etnicidad no especificada.

En relación con la ictericia al nacer, la mayoría de los participantes con edad registrada no la presentaron (73%), mientras que en el grupo con datos perdidos, la mitad tuvo ictericia. Finalmente, se observa que todos los participantes con edad faltante tienen diagnóstico de autismo, en contraste con solo el 17% del grupo con datos completos.

La imputación simple (media/mediana) podría introducir sesgos debido a estos patrones sistemáticos.

Considerar métodos de imputación múltiple que incorporen estas variables asociadas, realizar análisis de sensibilidad comparando resultados con y sin los casos incompletos e investigar posibles causas de estos patrones en la recolección de datos.

4.3 ¿Qué variables debo incluir en el proceso de imputación?

Debemos incluir todas las variables que se utilizarán en los análisis posteriores, incluso aquellas que no presentan valores perdidos. La razón es que el modelo de imputación debe ser tan complejo como el análisis que se realizará posteriormente. De lo contrario, se puede perder información relevante. Además, aunque algunas variables no tengan datos faltantes, su inclusión en el modelo mejora la estimación de los valores imputados.

En este caso, incluiremos la variable "Edad"(que tiene valores faltantes) junto con otras variables categóricas y continuas relevantes, asegurándonos de que las categóricas estén codificadas como factores.


```
{r}
input_data <- data_autismo |>
  dplyr::select(
    Edad,
    Genero,
    Etnicidad,
    Ictericia_al_nacer,
    Diagnostico_autismo,
    Pais_residencia,
    uso_prev_aplicacion,
    Resultado,
    Descripcion_edad,
    Quien_completa_prueba,
    Diagnostico_ASD,
    starts_with("Puntuacion_")
  ) |>
  mutate(
    Genero = as.factor(Genero),
    Etnicidad = as.factor(Etnicidad),
    Ictericia_al_nacer = as.factor(Ictericia_al_nacer),
    Diagnostico_autismo = as.factor(Diagnostico_autismo),
    Pais_residencia = as.factor(Pais_residencia),
    uso_prev_aplicacion = as.factor(uso_prev_aplicacion),
    Descripcion_edad = as.factor(Descripcion_edad),
    Quien_completa_prueba = as.factor(Quien_completa_prueba),
    Diagnostico_ASD = as.factor(Diagnostico_ASD)
  )
```

input_data

292 obs. of 21 variables

INTERPRETACIÓN:

Se incluyó específicamente la variable “Edad” que es la que presenta valores perdidos junto con otras variables categóricas como Género Etnicidad Diagnóstico de autismo y variables continuas como las puntuaciones y resultados del test.

Estas variables se convirtieron en factores para el modelo de imputación asegurando que el proceso refleje adecuadamente las relaciones entre ellas y mantenga la coherencia con los análisis posteriores.

El código tiene el objetivo de preparar los datos seleccionando las variables necesarias y transformando las categóricas a factores para facilitar el proceso de imputación.

4.4 La función "mice ()" para imputar datos

La función "mice" (Multivariate Imputation by Chained Equations), se utiliza para realizar imputación múltiple de datos, una técnica para manejar valores perdidos.

En nuestro caso, solo la variable "Edad" presenta valores faltantes. Sin embargo, incluimos otras variables relevantes como "Genero, Ictericia_al_nacer, Diagnostico_autismo, Resultado y otros en el modelo de imputación. Aunque estas variables no tengan datos perdidos, su inclusión permite mejorar la precisión de las imputaciones al aportar información útil para predecir los valores faltantes de "Edad".

Para imputar "Edad", utilizamos el método "pmm" (Predictive Mean Matching), adecuado para variables continuas. Las variables que no presentan valores perdidos son simplemente indicadas con comillas vacías (""), lo que informa a "mice()" que no deben ser imputadas.

Este enfoque permite conservar todos los casos y realizar análisis más robustos, evitando el sesgo que puede introducirse al eliminar observaciones con datos faltantes.

Aunque la variable con datos perdidos es únicamente "Edad", todas las demás se incluyen en el proceso de imputación porque aportan información valiosa que puede ayudar a estimar con mayor precisión los valores faltantes.

```
{r}
names(input_data)
```

[1]	"Edad"	"Genero"	"Etnicidad"
[4]	"Ictericia_al_nacer"	"Diagnostico_autismo"	"Pais_residencia"
[7]	"Uso_prev_aplicacion"	"Resultado"	"Descripcion_edad"
[10]	"Quien_completa_prueba"	"Diagnostico_ASD"	"Puntuacion_P1"
[13]	"Puntuacion_P2"	"Puntuacion_P3"	"Puntuacion_P4"
[16]	"Puntuacion_P5"	"Puntuacion_P6"	"Puntuacion_P7"
[19]	"Puntuacion_P8"	"Puntuacion_P9"	"Puntuacion_P10"

INTERPRETACIÓN:

- Nos muestra el resultado de ejecutar la función "names(input_data)" en R, lo que nos da una lista de 20 variables presentes en el conjunto de datos input_data.
- El conjunto de datos input_data contiene información demográfica (Edad, Género, Etnicidad, País de residencia), historial médico o de desarrollo (Ictericia al nacer), detalles sobre la administración de pruebas (uso_prev_aplicacion, Quien_completa_prueba, Descripcion_edad), y resultados específicos relacionados con el diagnóstico y puntuaciones de autismo.

La función "mice" se emplea para imputar datos faltantes mediante ecuaciones encadenadas lo cual es útil para manejar valores perdidos de manera robusta.

En este caso solo la variable "Edad" tiene valores faltantes pero se incluyen otras variables como "Género" "Ictericia al nacer" "Diagnóstico de autismo" y "Resultado" para mejorar la precisión de la imputación ya que aportan información relevante.

El método utilizado es "pmm" (Predictive Mean Matching) ideal para variables continuas como "Edad" mientras que las variables sin datos faltantes se indican con comillas vacías para que no sean imputadas.

Este enfoque permite mantener todos los casos en el análisis evitando sesgos por eliminación de observaciones incompletas y asegurando que las imputaciones reflejan patrones reales en los datos.

El método de imputación la indicaremos con el argumento "method" en el mismo orden que aparecen las variables en el dataset.

```
{r}
data_imputada <- mice(
  input_data,
  m = 20,
  method = c(
    "pmm", # Edad
    "",    # Genero
    "",    # Etnicidad
    "",    # Ictericia_al_nacer
    "",    # Diagnostico_autismo
    "",    # Pais_residencia
    "",    # Uso_prev_aplicacion
    "",    # Resultado
    "",    # Descripcion_edad
    "",    # Quien_completa_prueba
    "",    # Diagnostico_ASD
    "",    # Puntuacion_P1
    "",    # Puntuacion_P2
    "",    # Puntuacion_P3
    "",    # Puntuacion_P4
    "",    # Puntuacion_P5
    "",    # Puntuacion_P6
    "",    # Puntuacion_P7
    "",    # Puntuacion_P8
    "",    # Puntuacion_P9
    "",    # Puntuacion_P10
  )

```

```
),
maxit = 20,
seed = 3,
print = FALSE
)
```

Aviso: Number of logged events: 801

INTERPRETACIÓN:

- Solo Edad tiene "pmm" como método, porque es continua y tiene valores perdidos.
- Las demás se marcan con "" porque no se imputan, pero se usan para predecir.
- Las variables categóricas fueron transformadas a factor, lo cual es necesario para que "mice()" las interprete correctamente.

data_imputada

List of 23



{r}

data_imputada

Descripción: df [6 × 5]

	it <dbl>	im <dbl>	dep <chr>	meth <chr>
1	0	0		constant
2	1	1	Edad	pmm
3	1	1	Edad	pmm
4	1	2	Edad	pmm
5	1	2	Edad	pmm
6	1	3	Edad	pmm

6 rows | 1-5 of 5 columns

INTERPRETACIÓN:

- Los datos muestran que se han generado al menos tres conjuntos de datos imputados para la variable "Edad" (im = 1, 2, 3). Todos ellos fueron imputados utilizando el método Predictive Mean Matching (pmm), y este proceso se llevó a cabo dentro de la primera iteración (it = 1) del algoritmo de imputación. La primera fila (im=0) sirve como referencia o el estado inicial.
- Las repeticiones de (it=1, im=1) y (it=1, im=2) son detalles que podrían indicar una forma de cómo se organiza la información internamente.

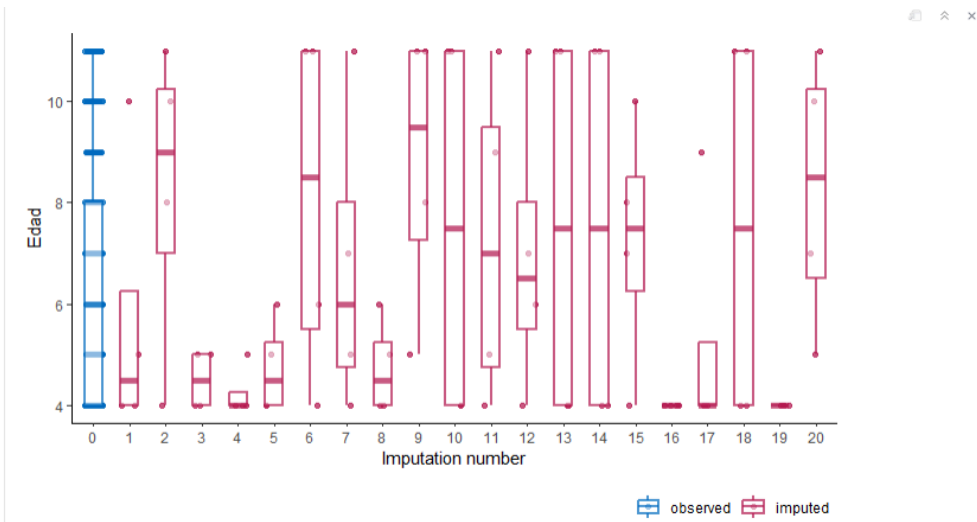
El resultado de la imputación se ha guardado en el objeto data_imputada y muestra que es un objeto de clase mids (multiply imputed dataset), el número de imputaciones (20), el método de imputación para todas las variables, y en una matriz, cuales variables han sido usadas para predecir otras.

5. ANALIZANDO DATOS IMPUTADOS

Antes de realizar análisis adicionales, es importante examinar si los valores imputados para la variable "Edad" son razonables en comparación con los valores originalmente observados. Un gráfico de cajas y bigotes permite comparar la distribución de los valores imputados frente a los valores reales, ayudando a identificar si el modelo de imputación ha generado datos plausibles. Si las distribuciones son similares, entonces podemos confiar en que la imputación no ha introducido sesgos artificiales.

Para la variable Edad

```
{r}  
ggmice(data_imputada, aes(x = .imp, y = Edad)) +  
  geom_jitter(height = 0, width = 0.25) +  
  geom_boxplot(width = 0.5, size = 1, alpha = 0.55, outlier.shape = NA) +  
  labs(x = "Imputation number")
```



INTERPRETACIÓN:

- Número de Imputaciones: Se muestran 20 imputaciones (del 1 al 20).
- Comparación de Distribuciones:
 - Para la mayoría de los números de imputación (por ejemplo, 1, 2, 3, 4, 5, 7, 8, 9, 10, 11, 12, 13, 14, 15, 17, 18, 19, 20), la distribución de los valores imputados (cajas rojas) parece similar a la distribución observada (caja azul en 0). Las medianas, los RIC y la dispersión general se alinean bien. Esto indica que el método de imputación (PMM, como vimos en la imagen anterior) está haciendo un buen trabajo al generar valores imputados que se asemejan a los datos observados.
- Posibles Problemas/Atípicos:
 - Podemos observar que los diagramas de caja rojos son muy estrechos para los números de imputación como 6 y 16. Estos están inusualmente comprimidos y tienen muy pocos, o aparentemente constantes, valores imputados. Esto podría indicar un problema con el modelo de imputación para estas imputaciones específicas, o quizás que muy pocos valores faltaban para “Edad” en esas ejecuciones particulares de imputación, lo que lleva a una pequeña muestra de valores imputados para el diagrama de caja.
 - También hay algunos puntos individuales dispersos (valores atípicos) que caen fuera del rango principal de los diagramas de caja.

Con la función “ggmice()”, podemos visualizar la distribución de la variable imputada (Edad) a través de gráficos de cajas. Los datos observados aparecen primero (generalmente en azul), y los boxplots siguientes representan las imputaciones múltiples (en nuestro caso, 20). Si los valores imputados se encuentran dentro del rango de los valores observados, entonces podemos considerarlos plausibles.

Para las variables categóricas, una buena práctica es verificar si la distribución de estas variables se mantiene similar antes y después de la imputación. Para ello, convertimos el conjunto de datos imputados en formato “long” con complete (data_imputada, "long") y comparamos las proporciones entre los datos originales e imputados mediante tablas de contingencia

```
{r}
data_imputada_l <- complete(data_imputada, "long", include = TRUE)
```

data_imputada_l 6132 obs. of 23 variables

```
{r}
data_imputada_l <- data_imputada_l %>%
  mutate(imputed = .imp > 0,
         imputed = factor(imputed,
                        levels = c(FALSE, TRUE),
                        labels = c("Observado", "Imputado")))

prop.table(table(data_imputada_l$Diagnostico_autismo,
                data_imputada_l$imputed),
          margin = 2)
```

	Observado	Imputado
No	0.8321918	0.8321918
Sí	0.1678082	0.1678082

INTERPRETACIÓN:

Interpretación de la tabla:

- La tabla muestra las proporciones de la variable “Diagnostico_autismo” para los datos “Observados” y para los datos “Imputados”.
- Columna “Observado”:
 - No: 83.21918% de las observaciones originales (no imputadas) tienen un “Diagnostico_autismo” de “No”.
 - Sí: 16.78082% de las observaciones originales (no imputadas) tienen un “Diagnostico_autismo” de “Sí”.
- Columna “Imputado”:
 - No: 83.21918% de las observaciones en los conjuntos de datos imputados tienen un “Diagnostico_autismo” de “No”.
 - Sí: 16.78082% de las observaciones en los conjuntos de datos imputados tienen un “Diagnostico_autismo” de “Sí”.

Conclusión de la tabla:

- La tabla muestra que las proporciones de “Diagnostico_autismo” son idénticas entre los datos originales observados y el conjunto combinado de datos imputados.
- Esto es un resultado muy importante en el diagnóstico de imputación. Nos indica que el proceso de imputación no ha alterado la distribución marginal de la variable “Diagnostico_autismo”. Es decir, la prevalencia o proporción de "Sí" y "No" para el diagnóstico de autismo es la misma antes y después de la imputación (cuando se promedian todas las imputaciones para el caso "Imputado"). Esto es una buena señal de que la imputación es consistente con la información presente en los datos originales.

El procedimiento estándar para realizar un análisis de regresión después de la imputación consiste en utilizar la función `with()` para ajustar el modelo de regresión al objeto `mids` (en nuestro caso, `data_imputada`). Posteriormente, se emplea la función `pool()` para obtener los resultados combinados, siguiendo las reglas de Rubin, que permiten obtener estimaciones válidas considerando la variabilidad entre imputaciones.

Sin embargo, si se hace uso del paquete `gtsummary`, este y sus funciones como `tbl_regression()` manejan internamente el agrupamiento de las imputaciones, por lo que solo es necesario utilizar la función `with()`. A continuación, se puede aplicar `tbl_regression()` directamente al objeto resultante, sin necesidad de usar `pool()`.

En nuestro caso, se imputaron valores faltantes de la variable `Edad`, y luego se evaluó su asociación con el diagnóstico clínico de trastorno del espectro autista (`Diagnostico_ASD`), ajustando por otras variables como `Género`, `Resultado del test` y presencia de `Ictericia al nacer`. Este enfoque permite obtener estimaciones más robustas sin excluir casos con información parcialmente faltante.

```
{r}
tabla_multi <-
  with(data_imputada,
    glm(Diagnostico_autismo ~ Edad + Genero + Resultado +
Ictericia_al_nacer,
      family = binomial(link = "logit"))
  ) |>
  tbl_regression(
    exponentiate = TRUE,
    label = list(
      Edad ~ "Edad",
      Genero ~ "Género",
      Resultado ~ "Puntaje total del test",
      Ictericia_al_nacer ~ "Ictericia al nacer"
    )
  ) |>
  bold_p(t = 0.05) |>
  modify_header(
    estimate = "***OR ajustado***",
    p.value = "***p valor***"
  )
```

```
{r}
tabla_multi
```

Characteristic	OR ajustado	95% CI	p valor
Edad	0.97	0.85, 1.11	0.7
Género			
Femenino	—	—	
Masculino	0.63	0.33, 1.22	0.2
Puntaje total del test	0.93	0.81, 1.06	0.3
Ictericia al nacer			
No	—	—	
Sí	0.72	0.34, 1.50	0.4

INTERPRETACIÓN: Se realizó un análisis de regresión logística multivariable para evaluar la asociación entre varias variables (edad, género, puntaje total del test e ictericia al nacer) y el diagnóstico de autismo

- **Edad:** El odds ratio (OR) ajustado fue de 0.97, lo que indica que, por cada año adicional de edad, las probabilidades de ser diagnosticado con autismo disminuyen en un 3%. Sin embargo, este resultado no fue estadísticamente significativo, ya que el valor de p fue de 0.7 y el intervalo de confianza del 95% (entre 0.85 y 1.11) incluye el valor 1.
- **Género:** Se utilizó al sexo femenino como categoría de referencia. Para el sexo masculino, el OR ajustado fue de 0.63, lo que sugiere que los varones tendrían un 37% menos de probabilidades de ser diagnosticados con autismo en comparación con las mujeres. Este resultado tampoco fue estadísticamente significativo (valor de p = 0.2) y el intervalo de confianza del 95% (de 0.33 a 1.22) incluye el 1.
- **Puntaje total del test:** El OR ajustado fue de 0.93. Esto indica que un mayor puntaje en el test se asocia con una menor probabilidad de diagnóstico de autismo. Sin embargo, este resultado no fue significativo (valor de p = 0.3) y su intervalo de confianza del 95% (de 0.81 a 1.06) también incluye el valor 1.
- **Ictericia al nacer:** Tomando como referencia a los pacientes que no presentaron ictericia, los que sí la presentaron tuvieron un OR ajustado de 0.72. Esto indica que haber tenido ictericia al nacer se asocia con una menor probabilidad de diagnóstico de autismo. Sin embargo, este hallazgo tampoco fue significativo (valor de p = 0.4), y el intervalo de confianza del 95% (de 0.34 a 1.50) incluye el 1.