

24-38/000306 IFCD104

Inteligencia Artificial y Big Data aplicado al ámbito biosanitario

ESCALA DE CALIFICACIÓN E₁

INSTRUCCIONES PARA EL ALUMNADO

MÓDULO:	Módulo 4
ACTIVIDAD:	E ₁ : Creación de gráficos exploratorios en R
FECHA:	16/04/2025
DURACIÓN:	2 horas
NOMBRE Y APELLIDOS	

Descripción de la práctica

En esta práctica, los estudiantes aplicarán sus conocimientos en R y visualización de datos para explorar un conjunto de datos biosanitarios. Usando gráficos adecuados, deberán analizar tanto variables cualitativas como cuantitativas, interpretando los patrones observados y describiendo la distribución de los datos. A partir de un dataset ficticio de pacientes, los estudiantes demostrarán su capacidad para:

- Configurar el entorno de trabajo en RStudio.
- Utilizar librerías esenciales como ggplot2, dplyr y tidyverse para la manipulación y exploración de datos.
- Crear gráficos adecuados para diferentes tipos de variables:
 - Diagrama de sectores para datos cualitativos.
 - Histogramas para datos cuantitativos discretos.
 - Diagramas de caja para datos cuantitativos continuos.
- Analizar e interpretar las visualizaciones generadas.

Este ejercicio tiene como objetivo evaluar la habilidad del estudiante para trabajar con datos en R, desde su carga y limpieza hasta la visualización e interpretación. Se espera que los participantes estructuren su código de manera clara y utilicen buenas prácticas de programación, añadiendo comentarios explicativos.

Instrucciones específicas

Configuración del entorno:

- Verifica que las librerías necesarias (ggplot2, dplyr, tidyverse) estén instaladas en tu entorno. Si no lo están, instálalas usando `install.packages()`.
- Carga las librerías necesarias usando `library()`.
- Muestra la versión de R y de las librerías cargadas.

2. Carga y exploración de datos:

- Descarga y carga el archivo `datos_pacientes.csv` en un dataframe usando la función `read.csv()`.
- Muestra las primeras 5 filas del dataset usando la función `head()`.
- Obtén información sobre los tipos de datos y verifica la existencia de valores nulos usando `str()`, `summary()` y `anyNA()`.

3. Visualización de datos:

- **Gráfico de sectores (Pie chart):** Representa la distribución de una variable cualitativa, como el género o el estado civil de los pacientes. Asegúrate de personalizar el gráfico con título, etiquetas y colores adecuados.
- **Histograma:** Crea un histograma para una variable cuantitativa discreta, como la edad de los pacientes. Ajusta el número de intervalos y añade etiquetas y título.
- **Diagrama de caja (Boxplot):** Representa la distribución de una variable continua, como la presión arterial o el IMC (Índice de Masa Corporal). Añade título y etiquetas para facilitar la interpretación.

4. Análisis e interpretación:

- Describe los patrones observados en cada gráfico. Por ejemplo: ¿La variable cualitativa está equilibrada? ¿Cómo es la distribución de la edad? ¿Existen valores atípicos en el IMC?
- Explica la relevancia de estos patrones en el contexto biosanitario.

Equipo y material

Equipo:

- Computadora o laptop con conexión a Internet.
- R y RStudio instalados como entorno de desarrollo.

Materiales y recursos:

- Dataset: Archivo CSV llamado `datos_pacientes.csv` con información ficticia de pacientes.
- Librerías necesarias:
 - `ggplot2`
 - `dplyr`
 - `tidyverse`

Criterios de evaluación (100 puntos)

1. **Configuración del entorno (10 puntos):**
 - Instalación y carga adecuada de librerías.
 - Verificación de versiones.
2. **Carga y exploración de datos (20 puntos):**
 - Carga correcta del dataset.
 - Exploración inicial usando funciones adecuadas (`head()`, `str()`, `summary()`).
 - Identificación de valores nulos.
3. **Creación de gráficos (40 puntos):**
 - Gráfico de sectores: presentación correcta de una variable cualitativa (10 puntos).
 - Histograma: visualización adecuada de una variable cuantitativa discreta (15 puntos).
 - Diagrama de caja: correcta representación de una variable continua (15 puntos).
4. **Análisis e interpretación (20 puntos):**
 - Descripción clara y precisa de los patrones observados.
 - Interpretación de la distribución de los datos en el contexto biosanitario.
5. **Documentación y claridad del código (10 puntos):**
 - Código limpio, bien estructurado y con comentarios explicativos.
 - Buenas prácticas de programación en R.