

LABORATORIO 3

Análisis de datos asistido por IA y funciones matemáticas

Estudio de elección de métodos anticonceptivos

Escuela de Informática

Información del laboratorio

Curso: EIY403 - Introducción al análisis de datos para otras carreras

Valor: 10 % de la nota final

Modalidad: Grupal - En los grupos que han venido trabajando

Duración: 90 minutos

Entregables: Archivo .Rmd + archivo .html

Dataset: eleccion_metodo_anticonceptivo.csv (proporcionado)

Entrega: Correo institucional a jordy.alfaro.brenes@una.cr

Asunto: Lab3_NombreEquipo_Integrante1_Integrante2

1. Introducción y objetivos

Este laboratorio integra dos componentes fundamentales del curso:

1. **Funciones matemáticas en R (30 %):** Implementación de funciones matemáticas aplicadas: evaluación funcional, análisis combinatorio y métodos numéricos
2. **Análisis de datos asistido por IA (70 %):** Aplicación práctica de técnicas de prompting para análisis exploratorio y visualización de resultados

Contexto del laboratorio:

Trabjarán con datos reales de un estudio sobre elección de métodos anticonceptivos que incluye 1,473 mujeres con información demográfica y de planificación familiar. Aplicarán lo aprendido en la presentación de “IA para Análisis de Datos” para realizar un análisis completo asistido por inteligencia artificial, documentando cada paso del proceso.

Importante: El uso de IA está permitido y recomendado también para la creación de funciones matemáticas (Parte I), siempre y cuando se documente el proceso y se verifique que el código funcione correctamente.

2. PARTE I: Funciones matemáticas en R (30 puntos)

Sobre el uso de IA en esta sección:

Pueden usar herramientas de IA (ChatGPT, Claude, Gemini, etc.) para ayudarles a crear estas funciones. Sin embargo, deben:

- Documentar el prompt usado (en markdown antes del código)
- Verificar que el código funciona correctamente
- Entender qué hace cada línea de código
- Ejecutar las pruebas proporcionadas para validar

2.1. Ejercicio 1: Evaluación de funciones matemáticas (10 puntos)

Contexto matemático:

Una función matemática $f(x)$ establece una relación entre un valor de entrada (preimagen o dominio) y un valor de salida (imagen o rango). Por ejemplo, si $f(x) = x^2 + 3x - 2$, entonces:

- Para $x = 2$: $f(2) = 2^2 + 3(2) - 2 = 4 + 6 - 2 = 8$
- Para $x = -1$: $f(-1) = (-1)^2 + 3(-1) - 2 = 1 - 3 - 2 = -4$

En R, podemos definir funciones matemáticas como objetos y evaluarlas en cualquier punto.

Ejercicio 1: Crear función evaluadora

Implemente una función llamada `evaluar_funcion` que:

- **Entrada:** Reciba dos parámetros:
 1. `f`: Una función matemática de R
 2. `x`: Un valor numérico (la preimagen)
- **Salida:** Retorne el resultado de evaluar $f(x)$ (la imagen)

```
# Si usan IA, documenten aqui el prompt usado

# COMPLETE: Implementar la funcion evaluadora
evaluar_funcion <- function(f, x) {
  # Su codigo aqui
}

# Ejemplos de prueba (NO MODIFICAR - para verificacion)
f1 <- function(x) x^2 + 3*x - 2
f2 <- function(x) sqrt(x) + log(x)
f3 <- function(x) exp(x) - 2*x

# Pruebas
cat("f1(2) =", evaluar_funcion(f1, 2), "\n")      # Debe dar 8
cat("f2(4) =", evaluar_funcion(f2, 4), "\n")      # Debe dar ~3.386
cat("f3(0) =", evaluar_funcion(f3, 0), "\n")      # Debe dar 1
```

2.2. Ejercicio 2: Combinaciones y permutaciones (10 puntos)

Contexto matemático: Análisis combinatorio

En estadística y probabilidad, frecuentemente necesitamos contar el número de formas de seleccionar elementos:

Permutaciones $P(n, r)$: Número de formas de ordenar r elementos de un conjunto de n elementos (el orden SÍ importa).

$$P(n, r) = \frac{n!}{(n-r)!}$$

Combinaciones $C(n, r)$: Número de formas de seleccionar r elementos de un conjunto de n elementos (el orden NO importa).

$$C(n, r) = \frac{n!}{r!(n-r)!}$$

Ejemplo práctico: Si tenemos 5 medicamentos y queremos seleccionar 2:

- **Permutaciones** $P(5, 2) = \frac{5!}{3!} = \frac{120}{6} = 20$: Si importa el orden (medicamento A primero, luego B \neq B primero, luego A)
- **Combinaciones** $C(5, 2) = \frac{5!}{2!3!} = \frac{120}{2 \cdot 6} = 10$: Si NO importa el orden (A+B = B+A)

Ejercicio 2: Implementar funciones combinatorias

Cree dos funciones separadas:

a) Función para permutaciones:

```
# Si usan IA, documenten aqui el prompt usado

calcular_permutaciones <- function(n, r) {
  # COMPLETE: Implementar  $P(n,r) = n! / (n-r)!$ 
  # Pueden usar la funcion factorial() de R
}
```

b) Función para combinaciones:

```
# Si usan IA, documenten aqui el prompt usado

calcular_combinaciones <- function(n, r) {
  # COMPLETE: Implementar  $C(n,r) = n! / (r! * (n-r)!)$ 
}
```

Pruebas de verificación:

```
# Ejemplo: De 8 pacientes, seleccionar 3 para un estudio
n <- 8
r <- 3

cat("De", n, "pacientes, seleccionar", r, ":\n")
cat("Permutaciones (orden importa):", calcular_permutaciones(n, r),
    "\n")
cat("Combinaciones (orden NO importa):", calcular_combinaciones(n, r
), "\n")

# Valores esperados:  $P(8,3) = 336$ ,  $C(8,3) = 56$ 
```

2.3. Ejercicio 3: Método de Newton-Raphson (10 puntos)**Contexto matemático: Métodos numéricos**

El **método de Newton-Raphson** es un algoritmo iterativo para encontrar raíces (ceros) de funciones. Es decir, valores x donde $f(x) = 0$.

Fórmula iterativa:

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$

donde $f'(x)$ es la derivada de $f(x)$.

Ejemplo paso a paso: Encontrar la raíz de $f(x) = x^2 - 2$ (es decir, calcular $\sqrt{2}$)

Derivada: $f'(x) = 2x$

Empezando con $x_0 = 1,5$:

$$x_1 = 1,5 - \frac{(1,5)^2 - 2}{2(1,5)} = 1,5 - \frac{0,25}{3} = 1,4167$$

$$x_2 = 1,4167 - \frac{(1,4167)^2 - 2}{2(1,4167)} = 1,4142$$

$$x_3 \approx 1,4142135... \quad (\text{muy cercano a } \sqrt{2})$$

Ejercicio 3: Implementar Newton-Raphson

Cree una función que encuentre la raíz de una función usando el método de Newton-Raphson:

```
# Si usan IA, documenten aqui el prompt usado

newton_raphson <- function(f, df, x0, tol = 1e-6, max_iter = 100) {
  # PARAMETROS:
  # f: funcion objetivo
  # df: derivada de la funcion
  # x0: valor inicial (aproximacion inicial)
  # tol: tolerancia (cuando parar)
  # max_iter: maximo numero de iteraciones

  x <- x0 # Valor inicial
  iter <- 0 # Contador de iteraciones

  # COMPLETE: Implementar el algoritmo iterativo
  # while (condicion de continuacion) {
  #   1. Calcular f(x) y df(x)
  #   2. Aplicar formula: x_nuevo = x - f(x)/df(x)
  #   3. Verificar convergencia: si |f(x_nuevo)| < tol, parar
  #   4. Actualizar x y contador
  # }

  # Retornar lista con resultado e informacion
  return(list(
    raiz = x,
    iteraciones = iter,
    valor_funcion = f(x)
  ))
}

# Ejemplo de prueba: Encontrar raiz cuadrada de 2
f <- function(x) x^2 - 2
df <- function(x) 2*x

resultado <- newton_raphson(f, df, x0 = 1.5)
cat("Raiz encontrada:", resultado$raiz, "\n")
cat("Iteraciones:", resultado$iteraciones, "\n")
cat("f(raiz) =", resultado$valor_funcion, "\n")
cat("Verificacion: sqrt(2) =", sqrt(2), "\n")
```

3. PARTE II: Análisis de datos asistido por IA (70 puntos)

Dataset: Elección de métodos anticonceptivos

El archivo `eleccion_metodo_anticonceptivo.csv` contiene datos de 1,473 mujeres con las siguientes variables:

- **edad:** Edad de la mujer (16-49 años)
- **educacion:** Nivel educativo (muy.baja, baja, media, alta)
- **cantidad.hijos:** Número de hijos (0-16)
- **trabaja:** Si trabaja o no (si, no)
- **tamano.familia:** Tamaño del núcleo familiar (3-18 personas)
- **metodo:** Método anticonceptivo usado (no.usa, corto.plazo, largo.plazo)

Separador de columnas: Punto y coma (;)

Total de observaciones: 1,473 mujeres

Instrucciones generales para TODA la Parte II:

Para cada ejercicio de esta sección:

1. **DOCUMENTE EL PROMPT:** Antes de cada chunk de código generado por IA, incluya una celda de texto markdown con:
 - El prompt completo que usó
 - El modelo de IA utilizado (ChatGPT, Claude, Gemini, etc.)
2. **EJECUTE EL CÓDIGO:** Copie el código generado y ejecútelo
3. **ADAPTE SI ES NECESARIO:** Si el código tiene errores, documéntelo y corrija
4. **INTERPRETE:** Agregue interpretación en texto después de cada resultado

Ejemplo de formato esperado:

```
### Prompt utilizado
**Modelo:** ChatGPT 4
**Prompt:** "Genera código R para..."

```{r}
Código generado por IA
```

**Interpretación:** Los resultados muestran que...
```


3.1. Ejercicio 4: Carga y exploración inicial con IA (20 puntos)

Ejercicio 4a: Carga de datos (4 puntos)

Use IA para generar código que:

- Cargue el dataset correctamente (separador es ;)
- Muestre las primeras 10 filas
- Verifique dimensiones (debe dar 1473 filas y 6 columnas)
- Muestre tipos de variables usando `str()`

Prompt sugerido base:

“Soy estudiante analizando datos de elección de métodos anticonceptivos. Genera código R para cargar el archivo 'eleccion_metodo_anticonceptivo.csv' que usa punto y coma como separador. El dataset tiene 1,473 observaciones y estas variables: edad (numérica), educacion (categórica: muy.baja, baja, media, alta), cantidad.hijos (numérica), trabaja (categórica: si/no), tamano.familia (numérica), metodo (categórica: no.usa, corto.plazo, largo.plazo). Muestra las primeras 10 filas, dimensiones del dataset, y la estructura de los datos usando str(). Incluye comentarios explicativos.”

Ejercicio 4b: Estadísticas descriptivas (8 puntos)

Use IA para generar código que calcule:

- Media, mediana, mínimo, máximo de `edad`
- Distribución de frecuencias de `educacion` (tabla con conteos y porcentajes)
- Distribución de frecuencias de `metodo` (tabla con conteos y porcentajes)
- Media de `cantidad.hijos` por nivel de `educacion`

Prompt sugerido:

“Genera código R que calcule: 1) estadísticas descriptivas completas (media, mediana, min, max) de la variable edad, 2) tabla de frecuencias absolutas y porcentajes para la variable educacion, 3) tabla de frecuencias para metodo, 4) promedio de cantidad.hijos agrupado por nivel de educacion. Usa funciones de R base o dplyr. Incluye comentarios.”

Ejercicio 4c: Identificación de patrones iniciales (8 puntos)

Solicite a la IA que genere código para:

- Tabla cruzada entre `educacion` y `metodo` con porcentajes por fila
- Calcular la edad promedio por cada tipo de `metodo`
- Identificar cuántas mujeres que trabajan vs no trabajan usan cada método

Prompt sugerido:

“Genera código R para crear: 1) tabla cruzada entre educacion y metodo mostrando porcentajes por fila, 2) edad promedio para cada tipo de metodo anticonceptivo, 3) tabla de frecuencias cruzadas entre trabaja y metodo. Usa funciones como table(), prop.table(), y aggregate() o dplyr. Incluye comentarios explicativos.”

Interpretación requerida:

- ¿Qué nivel educativo es más común en la muestra?
- ¿Qué método anticonceptivo predomina?
- ¿Hay diferencias en el uso de métodos según nivel educativo?
- ¿Las mujeres que trabajan usan métodos de forma diferente?

3.2. Ejercicio 5: Visualizaciones específicas con IA (50 puntos)

Ejercicio 5a: Histograma de edades (12 puntos)

Solicite a la IA código para crear:

- Histograma de la variable `edad`
- Con 20 bins (barras)
- Coloreado o con facetas separadas por tipo de `metodo`
- Títulos y etiquetas de ejes en español
- Tema limpio (`theme_minimal` o `theme_bw`)
- Leyenda clara y visible

Prompt sugerido:

“Genera código R usando ggplot2 para crear un histograma de la variable edad con 20 bins. Separa los datos por tipo de metodo anticonceptivo usando facetas (facet_wrap). Usa títulos y etiquetas en español. Aplica theme_minimal(). El gráfico debe mostrar claramente las diferencias en la distribución de edades según el método usado.”

Interpretación requerida:

- ¿En qué rango de edad se concentra la muestra?
- ¿Hay diferencias en la distribución de edades según el método?
- ¿Qué método es más común en mujeres jóvenes vs mayores?

Ejercicio 5b: Gráfico de barras: Educación vs Método (13 puntos)

Solicite código para:

- Gráfico de barras apiladas al 100 %
- Eje X: `educacion`
- Barras coloreadas por `metodo`
- Mostrar porcentajes relativos (`position = "fill"`)
- Títulos descriptivos y leyenda clara
- Etiquetas de porcentaje en las barras (opcional pero valorado)

Prompt sugerido:

“Genera código R con ggplot2 para un gráfico de barras apiladas que muestre la relación entre educacion (eje X) y metodo anticonceptivo (colores de relleno). Usa position='fill' para mostrar proporciones del 100 %. Incluye títulos en español, leyenda clara, y theme_minimal(). El gráfico debe permitir comparar fácilmente qué métodos predominan en cada nivel educativo.”

Interpretación requerida:

- ¿Qué relación observa entre nivel educativo y método?
- ¿En qué nivel es más común no usar métodos?
- ¿El uso de métodos de largo plazo aumenta con la educación?

Ejercicio 5c: Boxplots comparativos (12 puntos)

Solicite código para:

- Boxplot de `cantidad.hijos` agrupado por `metodo`
- Colores diferenciados por grupo
- Mostrar outliers claramente
- Títulos en español
- Etiquetas de ejes descriptivas

Prompt sugerido:

“Genera código R con ggplot2 para crear boxplots de cantidad.hijos agrupados por metodo anticonceptivo. Usa colores diferentes para cada método. Incluye títulos y etiquetas en español. Usa theme_minimal(). El gráfico debe mostrar claramente las medianas, rangos intercuartílicos y valores atípicos para cada grupo.”

Interpretación requerida:

- ¿Hay diferencias en cantidad de hijos según el método usado?
- ¿Cuál grupo tiene mayor variabilidad?
- ¿Las mujeres con más hijos tienden a usar métodos específicos?
- ¿Hay valores atípicos? ¿Qué representan?

Ejercicio 5d: Gráfico de dispersión con tendencia (13 puntos)

Solicite código para:

- Scatter plot de `edad` (eje X) vs `cantidad.hijos` (eje Y)
- Puntos coloreados por `metodo`
- Línea de tendencia suavizada (`geom_smooth`)
- Facetas separadas por `trabaja` (dos paneles: si trabaja, no trabaja)
- Transparencia en los puntos para ver superposición
- Títulos descriptivos en español

Prompt sugerido:

“Genera código R con ggplot2 para un gráfico de dispersión que muestre edad (eje X) vs cantidad.hijos (eje Y). Colorea los puntos según metodo anticonceptivo. Agrega líneas de tendencia suavizadas con geom_smooth(). Separa el gráfico en dos facetas según la variable trabaja (si/no). Usa transparencia (alpha) en los puntos para visualizar superposición. Incluye títulos y etiquetas en español, y theme_minimal().”

Interpretación requerida:

- ¿Cómo se relaciona la edad con la cantidad de hijos?
- ¿Esta relación difiere entre mujeres que trabajan y las que no?
- ¿Hay diferencias en patrones de uso de métodos entre ambos grupos?
- ¿Qué tendencias generales observa en el gráfico?

4. Especificaciones técnicas de entrega

Configuración obligatoria del YAML:

```
---
title: "Laboratorio 3: An lisis con IA y Funciones Matem ticas"
author: "Equipo [nombre] - [Integrante1], [Integrante2], [
  Integrante3]"
date: "'r Sys.Date()' "
output:
  html_document:
    toc: true
    toc_float: true
    toc_depth: 3
    number_sections: true
    theme: cosmo
    highlight: tango
    code_folding: show
    df_print: paged
---
```

Estructura del documento:

1. Sección de carga de librerías (con `include=FALSE`)
2. Parte I: Funciones Matemáticas (Ejercicios 1-3)
3. Parte II: Análisis con IA (Ejercicios 4-5)
4. Cada prompt debe documentarse en markdown ANTES del chunk de código
5. Interpretaciones en markdown DESPUÉS de cada resultado

Archivos a entregar:

- `laboratorio3.NombreEquipo_Integrante1_Integrante2.Rmd`
- `laboratorio3.NombreEquipo_Integrante1_Integrante2.html`
- Ambos archivos deben ser funcionales e independientes

Correo de entrega:

- **Destinatario:** `jordy.alfaro.brenes@una.cr`
- **Asunto:** `Lab3.NombreEquipo_Integrante1_Integrante2`
- **Ejemplo:** `Lab3_Equipo1_Maria_Rodriguez_Carlos_Perez`
- **Cuerpo:** Lista de integrantes del equipo y breve descripción de principales hallazgos
- **Adjuntos:** Los 2 archivos mencionados arriba

5. Criterios de evaluación detallados

| Criterio | Puntos | Descripción |
|--|--|--|
| PARTE I: FUNCIONES MATEMÁTICAS (30 pts) | | |
| Ejercicio 1: Evaluación funcional | 10 | |
| - Implementación correcta | 6 | Función evalúa correctamente $f(x)$ |
| - Pruebas verificadas | 3 | Los 3 casos de prueba dan resultados correctos |
| - Documentación (si usa IA) | 1 | Prompt documentado apropiadamente |
| Ejercicio 2: Combinatoria | 10 | |
| - Función permutaciones | 4 | $P(n, r)$ correctamente implementada |
| - Función combinaciones | 4 | $C(n, r)$ correctamente implementada |
| - Verificación de resultados | 1 | Pruebas dan valores esperados |
| - Documentación (si usa IA) | 1 | Prompts documentados |
| Ejercicio 3: Newton-Raphson | 10 | |
| - Algoritmo iterativo correcto | 5 | Bucle while implementado apropiadamente |
| - Condición de convergencia | 3 | Verifica tolerancia correctamente |
| - Resultado verificado | 1 | Encuentra raíz con precisión adecuada |
| - Documentación (si usa IA) | 1 | Prompt documentado |
| PARTE II: ANÁLISIS CON IA (70 pts) | | |
| Ejercicio 4: EDA con IA | 20 | |
| - 4a: Carga de datos | 4 | Prompt + código + verificación de 1473 filas |
| - 4b: Estadísticas descriptivas | 8 | Todas las métricas + prompts documentados |
| - 4c: Patrones iniciales | 4 | Tablas cruzadas + prompts |
| - Interpretación global | 4 | Responde las preguntas planteadas |
| Ejercicio 5: Visualizaciones | 50 | |
| - 5a: Histograma de edades | 12 | Prompt + gráfico correcto + interpretación |
| - 5b: Barras Educación vs Método | 13 | Prompt + gráfico apilado + interpretación |
| - 5c: Boxplots comparativos | 12 | Prompt + gráfico + interpretación completa |
| - 5d: Scatter con tendencia | 13 | Prompt + gráfico con facetas + interpretación |
| ASPECTOS GENERALES | | |
| Documentación de prompts | Evaluado en cada ejercicio de Parte II | |
| Formato y estructura | YAML correcto, secciones organizadas | |
| Reproducibilidad | Código ejecutable sin errores | |
| Calidad de interpretaciones | Análisis profundo, no superficial | |
| TOTAL | 100 | |

6. Recursos de apoyo y referencias

Documentación útil:

Para Funciones Matemáticas:

- Funciones en R: `function()`, `return()`
- Factorial: `factorial(n)`
- Derivadas numéricas: Paquete `numDeriv` (opcional)
- Bucles iterativos: `while()`, condiciones de parada

Para Análisis con IA:

- Revisitar presentación: “IA para Análisis de Datos”
- Estructura de prompts: CONTEXTO + DATOS + OBJETIVO + FORMATO
- Modelos recomendados: ChatGPT, Claude, Gemini, DeepSeek
- Validación siempre necesaria: No confiar ciegamente en código generado

Paquetes de R sugeridos:

- Manipulación: `dplyr`, `tidyr`
- Visualización: `ggplot2`
- Estadística: `stats` (base)
- Tablas: `knitr::kable()`, `kableExtra`

Tips para prompts efectivos:

Componentes clave de un buen prompt:

1. **Contexto:** “Soy estudiante de [carrera] analizando [tema]...”
2. **Datos:** “Tengo un dataset con 1,473 observaciones y variables [liste variables]...”
3. **Objetivo:** “Quiero [objetivo específico]...”
4. **Método:** “Usa [técnica específica]...” (opcional)
5. **Formato:** “Genera código R con ggplot2 y comentarios...”
6. **Nivel:** “Explica como si tuviera nivel [básico/intermedio]...”

Ejemplos de mejora de prompts:

“Analiza estos datos”

“Tengo datos de 1,473 mujeres con variables edad, educacion y metodo anticonceptivo. Genera código R para crear un gráfico de barras que muestre la distribución de métodos por nivel educativo, con porcentajes en las barras y colores diferenciados. Usa ggplot2, position='fill' y theme_minimal().”

“Haz un gráfico de edad”

“Genera código R con ggplot2 para un histograma de la variable edad con 20 bins, separado por facetas según metodo. Incluye títulos en español y theme_minimal(). Debe mostrar claramente diferencias en distribuciones.”

Recordatorios importantes:

1. **Tiempo de desarrollo:** 90 minutos en clase
2. **Consultas permitidas:** Pueden preguntar al profesor durante el desarrollo
3. **Colaboración:** Este laboratorio es GRUPAL en los equipos establecidos
4. **Uso de IA:** Completamente permitido y requerido - DEBEN documentar prompts
5. **Documentación:** CRÍTICA - todos los prompts deben estar documentados en mark-down
6. **Validación:** El equipo es responsable de verificar que el código funcione
7. **Plagio:** Copiar entre equipos resulta en nota 0 para todos los involucrados
8. **Dataset:** Usar exclusivamente `eleccion_metodo_anticonceptivo.csv`
9. **Verificación:** El dataset tiene exactamente 1,473 filas - verifiquen esto al cargar

Sobre el uso responsable de IA:

Recuerden los principios de la presentación:

- La IA es una **herramienta de apoyo**, no hace el trabajo por ustedes
- Siempre **validen** que el código generado funcione correctamente
- **Entiendan** qué hace cada línea de código antes de usarlo
- Si algo falla, **itérenlo** con prompts de seguimiento más específicos
- **Documenten** honestamente el proceso, incluyendo correcciones necesarias
- Las **interpretaciones** deben ser propias, no copiadas de la IA

Este laboratorio evalúa:

1. Su capacidad de formular prompts efectivos
2. Su habilidad para validar y adaptar código generado
3. Su comprensión de los análisis realizados
4. Su capacidad de interpretar resultados críticamente

Este laboratorio integra programación matemática fundamental con aplicaciones modernas de IA, desarrollando competencias técnicas y analíticas esenciales para el análisis de datos profesional en investigación social y de salud pública.

Universidad Nacional de Costa Rica

Escuela de Informática - EIY403

II Semestre 2025