

Tercer reto: Análisis Numérico.

Álvarez Sánchez, Emanuel
alvarez_emanuel@javeriana.edu.co

Afanador Ochoa, Juan José Burgos Melo, Diego Andrés
j.afanador@javeriana.edu.co *burgosmd@javeriana.edu.co*

Barrera Martinez, Nicolai
barreram.n@javeriana.edu.co

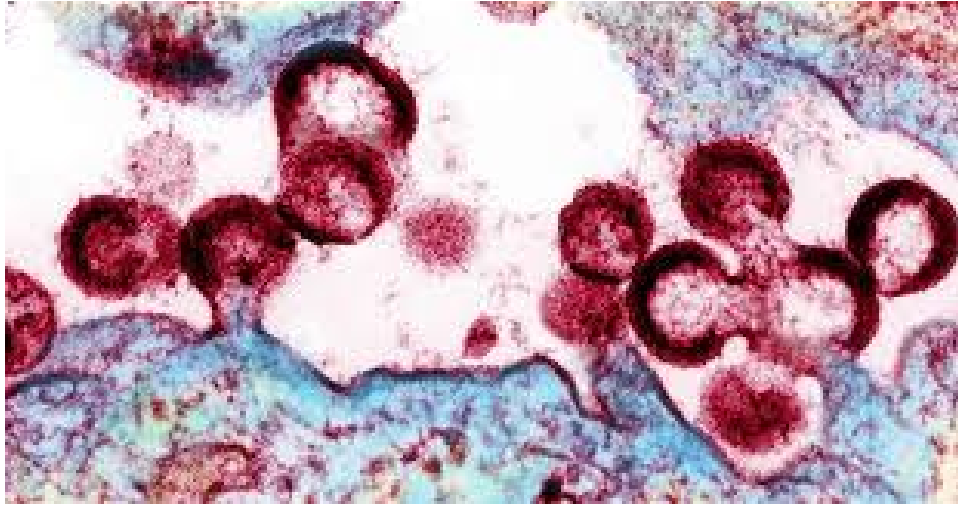
24 de mayo de 2020

1. Introduction

A continuación, veremos el respectivo documento del tercer reto de análisis numérico, donde podremos encontrar la descripción del problema escogido. El cual fue la la representación de una aplicación interactiva con el usuario que permite modelar a partir de una base de datos el comportamiento de un virus. En este caso escogimos la modelación del virus de la inmunodeficiencia humana o VIH, usando la técnica del Método iterativo de Runge-Kutta de cuarto orden y el de Adams, que facilita la aproximación a ecuaciones diferenciales ordinarias, a partir de ciertos valores iniciales.

2. Problema

El reto escogido por el grupo fue el modelamiento del virus de la inmunodeficiencia humana o VIH. el cual es una enfermedad la cual causa la falla del sistema inmune lo cual prolifera la aparición de otras enfermedades. Esta enfermedad infecta celular vitales a partir del contacto de ciertos fluidos como: sangre, semen, flujo vaginal y leche de lactancia.



1. Celular infectadas por VIH.

Teniendo en cuenta lo anterior se debía crear una aplicación que modele y calcule cifras a partir de datos proporcionados por el usuario de una forma que resulte didáctica e intuitiva para este, comparando los cálculos experimentales con las bases de datos obtenidas en la investigación. Para ello usamos el paquete "Shiny."^a adaptado para RStudio, el cual permite la creación de objetos como labels, cuadros de texto, gráficas, botones, barras de texto, entre otras, que facilitan la creación de aplicaciones.

3. Desarrollo

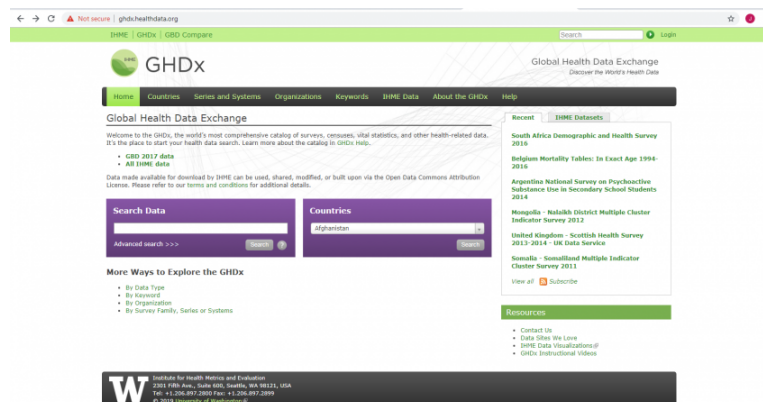
El desarrollo de este reto se llevó a cabo dividiendo las tareas necesarias para el cumplimiento de este en tres fases: 1) búsqueda y recolección de datos, 2) la compresión y modelamiento de datos y 3) el desarrollo y montaje de datos en la aplicación.

3.1. Recolección de datos.

En cuanto a la recolección de datos buscamos información sobre el tema en fuentes de alta calidad como es "Global Health Data Exchange" la cual se encarga de compilar datos a partir de encuestas, censos, publicaciones científicas, informes, entre otras. Llenado una base de datos.

Esto con el objetivo que la recolección de datos tuviera características como ser: precisos, completos, compatibles, relevantes y útiles. A su vez ad-

quirir información completa y ayudarnos a tener un mayor dominio de las variables que usamos en otras fases del reto como modelamiento y aplicación de datos.



2. Interfaz de búsqueda de "Global Health Data Exchange".

3.2. Modelamiento de datos.

En la recolección de datos pudimos obtener datos de 3 sectores que usamos como delimitador para un mayor control, los cuales son América latina y el caribe, Brasil y Colombia. Así pudimos recolectar mayor información y subdividirla en dos tipos de población, los susceptibles y los infectados.

Así mismo la información recolectada, se cruzó con mas datos para así crear, tres archivos uno por cada sector de división (América latina y el caribe, Brasil y Colombia) en Microsoft Excel los cuales contienen la información sobre el VIH, estos archivos poseen la siguiente información: País, código del país, año, Muertes VIH, nuevos casos de infección y personas vivas que tiene VIH.

Pais,Code,Year,Deaths from HIV/AIDS (Number),New infections of HIV/AIDS (new cases of
Colombia,COL,1990,429.5881894267906,2179.88218846,745.9131484210001,33102575
Colombia,COL,1991,521.3494903871102,3093.70377427,1041.70393119,33758337
Colombia,COL,1992,543.8891361162904,3875.22634517,1348.0776424800001,34422569
Colombia,COL,1993,673.4046761742588,4500.30913246,1657.5596942299999,35091275
Colombia,COL,1994,927.3024930953117,4942.83101448,1961.6539415500001,35758979
Colombia,COL,1995,1039.5034567743235,5175.98147584,2251.57245495,36421438
Colombia,COL,1996,1089.9871754225953,5100.0322598,2592.6974519600003,37076387
Colombia,COL,1997,1818.6276472206207,4734.53213925,3011.8292217,37723800
Colombia,COL,1998,1823.5345902355507,4247.152102,3440.8712044900003,38364309

3. Archivo excel con datos del VIH - Colombia.

Como podemos observarlo en la imagen 3, la información está separada mediante ',' para, realizan una tokenización con mayor facilidad y un salto de línea para separar el año de la información recolectada.

Una vez teniendo estos datos en los Excel, fueron montados al archivo en RStudio donde posteriormente van a ser tratados, usando métodos iterativos de Runge-Kutta de cuarto orden y el de Adams. Estos métodos necesitan de cierta información para resolver la ecuación que a continuación veremos y explicaremos:

A partir de la tesis “Modelos matemáticos del sida” de la universidad autónoma de nuevo león, tomamos como referencia la ecuación que nos ayuda a modelar el comportamiento de la población infectada y susceptible (susceptibles-infecciosos) al virus del VIH la cual fue (Figura 4) para los infectados y para los susceptibles (Figura 5):

$$\dot{I} = \alpha SI$$

4. Ecuación para calcular los infectados.

$$\dot{S} = -\alpha SI$$

5. Ecuación para calcular los susceptibles.

Con estas ecuaciones se nos generó la necesidad de hallar un valor en alfa (que corresponde a la probabilidad de que una persona sea infectada, mayor a cero), I (cantidad de personas infectadas) y S (cantidad de personas susceptibles).

3.3. Creación del aplicativo.

En cuanto a la creación del aplicativo se indago a partir de archivos como “An introduction to R Shiny” y la documentación “Package Shiny” proporcionadas por la profesora. La cual nos ayudo a dar una noción de las posibilidades a las que podíamos llegar a plasmas en la interfaz y los datos que iban a ser pedidos por pantalla como lo es, el alfa (probabilidad de que una persona sea infectada) y la cantidad de personas infectadas o personas susceptibles, depende cual sea el caso.

4. Implementacion de codigo

4.1. Descripción de librerías requeridas para el uso del programa.

```
library(shiny)
library(readr)
require(pracma)
require(ggplot2)
require(dplyr)
```

Estas librerías son utilizadas para dos cosas principalmente, para el cálculo de los métodos iterativos como lo son el Runge-Kutta de cuarto orden y el

de Adams facilitando operaciones numéricas y así mismo el cálculo de las aproximaciones a comparar. Y por otra parte se usan para la creación de botones y deslizadores los cuales nos van a proporcionar información para el cálculo de las operaciones.

4.2. Importación de datos a partir de archivos.

En esta sección se lee la información compactada y clasificada en un solo archivo que se encuentra en un Excel llamado "Data", para que luego se proceda a el cálculo de variables como susceptibles, infectados y total de la población .

```
Data <- read_csv("Data.csv")

suceptibles <- (Data$Population - Data$'New infections of HIV/AIDS
(new cases of HIV infection)')
infectados <- ((Data$'Number of people living with HIV (x10) (tens of
people living with HIV)'*10) + Data$'New infections of HIV/AIDS (new
cases of HIV infection)')
total <- (suceptibles+infectados)
```

4.3. Creacion interfaz.

A continuación veremos como se crean los botones y deslizadores, con valores y nombres predeterminados.

```
sidebarLayout(
  sidebarPanel(

    selectInput(inputId = "pais", label = strong("Indice_de_paises"),
               choices = unique(Data$Pais),
               selected = "Colombia"),

    selectInput(inputId = "poblacion", label = strong
("Indice_de_tipo_de_poblacion"),
               choices = c("Infectados", "Susceptibles"),
               selected = "Infectados"),

    selectInput(inputId = "metodo", label = strong("Indice_de_metodos"),
               choices = c("Runge-Kutta", "Adams"),
               selected = "Runge-Kutta"),
```

```

sliderInput("alfa",
            "Valor_de_alfa:",
            min = 0,
            max = 0.01,
            value = 0),

sliderInput("inicial",
            "Poblacion_infectada_inicial:",
            min = round(min(infectados),0),
            max = round(max(infectados),0),
            value = 9000),
sliderInput("inicialS",
            "Poblacion_susceptible_inicial:",
            min = round(min(susceptibles),0),
            max = round(max(susceptibles),0),
            value = 2)
),

# Show a plot of the generated distribution
mainPanel(
  plotOutput(outputId = "linePlot"),
  tableOutput(outputId = "table")
)

```

En el apartado "mainPanel" junto con la generación de graficas en tiempo real de datos experimentales, datos teóricos y errores calculados.

VIH Simulator

Indice de paises

Latin America and Caribbean

Indice de tipo de poblacion

Susceptibles

Indice de metodos

Runge-Kutta

Valor de alfa:

0

0.0038

0.01



Poblacion infectada inicial:

9,639

2,144,957



Poblacion suceptible inicial:

33,100,395

441,217,603

635,210,536



6. Interfaz del usuario.

4.4. Cálculo de datos por cada método.

```
choose_method <- reactive({  
  
  alfa <- input$alfa  
  yInicial <- input$inicial  
  func <- choose_poblacion()  
  
  if(input$poblacion == "Susceptibles"){  
    yInicial <- input$inicialS  
  }  
  
  if(input$metodo == "Runge-Kutta"){  
    res <- rk4(func, 1, 28, yInicial, 27)  
    res  
  } else {  
    res <- abm3pc(func, 1, 28, yInicial, 27)  
    res  
  }  
})
```

Podemos observar el método `reactive(...)`, el cual a partir de las entradas que nos proporciona el usuario se guardan en las variables; “alfa” y “yInicial”, que serán pasadas como parámetro a los métodos que el usuario haya escogido, como opción en la variables “metodo”. Nota: el proceso para los infectados es el mismo, solo cambia la ecuación a solucionar, deja de ser positiva y se torna negativa.

4.5. Creación de gráficas de comportamiento y tablas de error

```
output$linePlot <- renderPlot({  
  # generate bins based on input$bins from ui.R  
  
  #func <- choose_poblacion()  
  res <- choose_method()  
  
  df <- data.frame((res$x+1989), res$y, "experimental")
```

```

names(df)<- c("anio", "infectados", "tipo")

DataFiltered <- (Data %>% filter(Pais == input$pais))
infectados <- choose_dataP()
df2 <- data.frame((res$x+1989), infectados, "Exacto")
names(df2)<- c("anio", "infectados", "tipo")

dfd <- rbind(df, df2)

p <- ggplot(dfd, aes(x = anio, y = infectados, colour = tipo))
p <- p +geom_point(size = 3) + facet_grid(tipo ~ .) + theme_light
()
p
})

output$table <- renderTable({

  res <- choose_method()

  infectadosExactos <- (update_table() %>% filter(tipo == "Exacto"))
  infectadosExperi <- (update_table() %>% filter(
    tipo == "experimental"))

  errorAbs <- abs(infectadosExactos$infectados -
    infectadosExperi$infectados)
  errorRel <- abs(infectadosExactos$infectados -
    infectadosExperi$infectados)/infectadosExactos$infectados

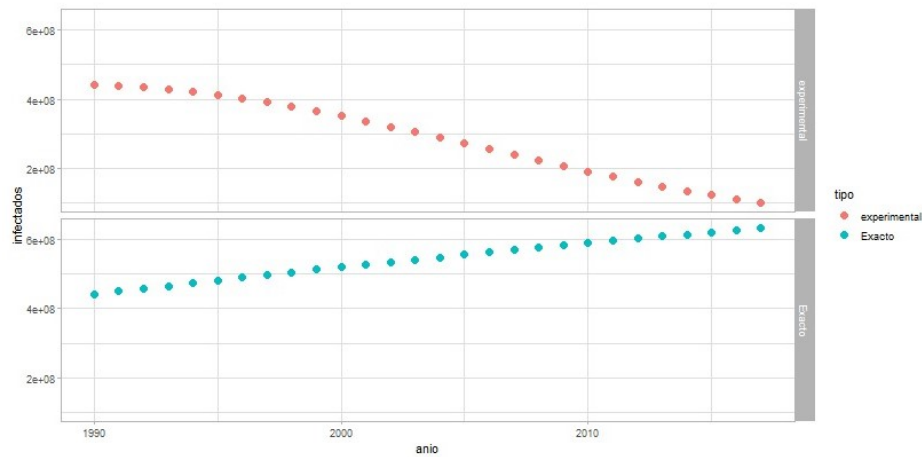
  error <- data.frame((res$x+1989), infectadosExactos$infectados,
    infectadosExperi$infectados, errorAbs, errorRel)
  names(error) <- c("anio", paste(input$poblacion," reales _
    tomados_de_Global_Health_Data_Exchange"), sep = "_"), paste
    (input$poblacion, "experimentales_(calculados)", sep = "_"),
    "error_absoluto", "error_relativo")
  error
})
}

```

Usando la función `renderPlot()` que se encuentra en la primera parte del código, calculamos dinámicamente las gráficas de los valores calculados

experimentalmente vs los valores recuperados de las bases de datos.

En la segunda parte, encontramos la generación de la tabla de errores (absoluto y relativo) que resulta de la comparación entre infectados exactos que es extraído de la base de datos y infectadosExperi” los cuales son los datos generados a partir del método escogido por el usuario.



7.Graficas comparativas entre datos teoricos y experimentales.

anio	Susceptibles reales (tomados de Global Health Data Exchange)	Susceptibles experimentales (calculados)	error absoluto	error relativo
1990.00	441607352.99	441217603.00	389749.99	0.00
1991.00	449626128.01	438709816.64	10916311.37	0.02
1992.00	457613553.11	434561807.62	23051745.49	0.05
1993.00	465579644.91	428820400.56	36759244.35	0.08
1994.00	473522063.80	421549911.54	51972152.26	0.11
1995.00	481436511.90	412830948.91	68605563.00	0.14
1996.00	489328249.17	402758926.01	86569323.16	0.18
1997.00	497165723.83	391442324.19	105723399.64	0.21
1998.00	504943579.87	379000750.25	125942829.62	0.25
1999.00	512574493.67	365562836.96	147011656.71	0.29
2000.00	520060845.07	351264037.96	168796807.11	0.32
2001.00	527394437.31	336244369.79	191150067.52	0.36
2002.00	534581880.89	320646153.12	213935727.76	0.40
2003.00	541644087.46	304611803.67	237032283.79	0.44

8.tabla comparativa entre datos teoricos y experimentales, junto con sus errores.

5. Conclusiones

En conclusión, podemos observar que los datos calculados de los infectados, tienen un error que va creciendo a medida que los datos van aumentando con respecto a los datos adquiridos en la base de datos. Esto ocurre a causa de que los datos teóricos, se tiene en cuenta todas las campañas preventivas que se realizan en determinado lugar y modifican la tendencia de manera atípica lo cual implica que el modelo que nosotros usamos sirva en condiciones ideales.

Por ello se llega a la conclusión que la correcta investigación del modelo matemático del VIH y la buena aplicación de los métodos vistos en clase lograron emular satisfactoriamente la tendencia de contagio que a lo largo de 28 años ha tenido el virus en lugares como América latina, Brasil y Colombia, BAJO ciertas condiciones.

Por otra parte, los datos arrojados por la aplicación en cuanto a los susceptibles difiere de las bases de datos investigadas, esto a causa que nuestros datos calculados con los métodos iterativos no tienen en cuenta la variable que implica el crecimiento exponencial de la población en este periodo de tiempo contemplado. por ende, la gráfica de los datos experimentales de la población susceptible tiene una pendiente negativa a causa de la reducción de la población no infectada. Lo contrario a la gráfica de los datos teóricos de la población susceptible que posee tiene una pendiente positiva.

existen métodos como el de Runge-Kutta y el de Adams, que junto con una ecuación correcta y cierta cantidad limitada de datos, nos permiten modelar el comportamiento de situaciones. En este caso los congios por VIH, pero que pueden ser aplicados en muchas áreas de conocimiento trayendo beneficios para nosotros, al poder predecir, entender y aprender sobre problemas que aquejan actualmente a nuestra sociedad como lo es el COVID-19.