



Universidad Nacional Experimental del Táchira
Vicerrectorado Académico
Decanato de Docencia
Departamento de Ingeniería en Informática
Trabajo de Aplicación Profesional
Proyecto Especial de Grado

DESARROLLO DE UNA APLICACIÓN WEB PARA EL ESTUDIO DE DATOS LONGITUDINALES
PARA EL SEGUIMIENTO DE LA INFECCIÓN POR EL VIRUS DE INMUNODEFICIENCIA
HUMANA, CASO DE ESTUDIO: INSTITUTO AUTÓNOMO HOSPITAL UNIVERSITARIO DE LOS
ANDES

Autor: Stephanie Katherine Correa Alcántara
C.I.: V-19.598.574
stephanie.correa24@gmail.com
Tutor: Rossana Timaure García. PhD.
rttg@unet.edu.ve

San Cristóbal, Julio de 2018



Universidad Nacional Experimental del Táchira
Vicerrectorado Académico
Decanato de Docencia
Departamento de Ingeniería en Informática
Trabajo de Aplicación Profesional
Proyecto Especial de Grado

DESARROLLO DE UNA APLICACIÓN WEB PARA EL ESTUDIO DE DATOS LONGITUDINALES
PARA EL SEGUIMIENTO DE LA INFECCIÓN POR EL VIRUS DE INMUNODEFICIENCIA
HUMANA, CASO DE ESTUDIO: INSTITUTO AUTÓNOMO HOSPITAL UNIVERSITARIO DE LOS
ANDES

Proyecto especial de grado presentado como requisito parcial para optar al título de
Ingeniero en Informática

Autor: Stephanie Katherine Correa Alcántara
C.I.: V-19.598.574
stephanie.correa24@gmail.com
Tutor: Rossana Timaure García. PhD.
rttg@unet.edu.ve

San Cristóbal, Julio de 2018

Dedicatoria

A Dios todopoderoso.

A mis padres.

A mi pequeño Martín valiente.

A Daniel.

Y a la UNET

Agradecimientos

Primeramente a Dios, por darme la salud y fortaleza, porque me ha guiado e iluminado a lo largo de mi vida, especialmente para alcanzar esta meta.

A mis padres, Magalys y Rafael, por todo el indudable apoyo económico y moral en todo momento, por escucharme, por confiar y tener esperanza en mí.

A Daniel, por su amor, comprensión y paciencia en los momentos en los que más he necesitado.

A mi hijo Martín, el cual representa mi orgullo y mi motivación principal para ser una mejor persona y servirle de ejemplo.

A toda mi familia en general, por el apoyo brindado a lo largo de los años.

A la tutora, Rossana Timaure, por darme la oportunidad, apoyo y conocimientos para la realización de este trabajo de grado.

A todos los profesores que a lo largo de la carrera me brindaron la guía necesaria y el conocimiento.

A la ilustre Universidad Nacional Experimental del Táchira, por permitirme entrar en sus aulas y formarme como profesional.

Universidad Nacional Experimental del Táchira
Vicerrectorado Académico
Decanato de Docencia
Departamento de Ingeniería Informática

**Desarrollo de una aplicación Web para el estudio de datos longitudinales
para el seguimiento de la infección por el Virus de Inmunodeficiencia
Humana, caso de estudio: Instituto Autónomo Hospital Universitario de
los Andes**

Autor: Correa Alcántara, Stephanie Katherine

Tutor: Timaure García, Rossana. PhD.

Fecha: Julio 2018.

RESUMEN

Se describe el desarrollo de una aplicación web como paquete del lenguaje de programación R, para el seguimiento de la infección por el virus de inmunodeficiencia humana (VIH), haciendo uso de datos longitudinales y el modelo lineal mixto. La metodología de desarrollo de *software* empleada fue la espiral, mediante la cual se realizó el filtrado, descripción y exploración de la base de datos del Laboratorio de Investigaciones Hormonales de Instituto Autónomo Hospital Universitario de los Andes, en la cual se almacena la información de seguimiento del comportamiento de los biomarcadores de los pacientes pertenecientes al Programa Nacional SIDA/ITS, de aquí se obtuvieron las vistas minables que sirvieron para determinar el formato de los datos de entrada a la aplicación; de igual forma se contruyeron y codificaron los modelos de funciones que integran el paquete, utilizando la dependencia con el paquete *Shiny* de R para el despliegue gráfico sobre el mapa de Venezuela de las características demográficas y densidad de los pacientes registrados en la base de datos, así como se implementó una función para el ajuste y estimación de un modelo lineal de efectos mixtos, que permita explorar de forma descriptiva, que factores resultan significativos para explicar la variabilidad de la carga viral plasmática ó el conteo de células T^+CD4 .

Palabras clave: Datos longitudinales, modelo lineal mixto, VIH, R, Shiny.

Índice general

| | |
|---|----------|
| Dedicatoria | I |
| Agradecimientos | II |
| Resumen | III |
| Índice General | IV |
| Índice de Cuadros | VII |
| Índice de Figuras | VIII |
| Introducción | 1 |
| 1. El problema | 3 |
| 1.1. Planteamiento y formulación | 3 |
| 1.2. Objetivos | 5 |
| 1.2.1. Objetivo general | 5 |
| 1.2.2. Objetivos específicos | 5 |
| 1.3. Justificación e importancia | 6 |
| 1.4. Alcance y limitaciones | 6 |
| 2. Fundamentos teóricos | 8 |
| 2.1. Antecedentes | 8 |
| 2.2. Bases teóricas | 11 |
| 2.2.1. Estudios longitudinales | 11 |
| 2.2.2. Análisis de datos longitudinales | 12 |
| 2.2.3. Datos longitudinales en el VIH | 12 |
| 2.2.4. Modelos lineales mixtos | 14 |

| | | |
|-----------|---|-----------|
| 2.2.5. | Módulos lineales mixtos en el análisis de datos longitudinales | 15 |
| 2.2.6. | VIH | 16 |
| 2.2.7. | SIDA | 16 |
| 2.2.8. | Relación VIH - $T^+CD4 - T^+CD8$ | 17 |
| 2.2.9. | Clasificación establecida de la infección por el VIH y enfermedades re- lacionadas | 18 |
| 2.2.10. | Clasificación clínica de la infección establecida por el VIH | 18 |
| 2.2.11. | Clasificación inmunológica para la infección establecida por el VIH . . . | 18 |
| 2.2.12. | VIH en Venezuela | 19 |
| 2.2.13. | Lenguaje R para el cómputo estadístico | 21 |
| 2.2.14. | Adquisición y licencia | 22 |
| 2.2.15. | Interfaz de usuario | 22 |
| 2.2.16. | Lenguaje de programación | 23 |
| 2.2.17. | Shiny | 24 |
| 3. | Fundamentos metodológicos | 25 |
| 3.1. | Enfoque de la investigación | 25 |
| 3.2. | Tipo o nivel de investigación | 25 |
| 3.3. | Diseño de investigación | 26 |
| 3.4. | Población y muestra | 27 |
| 3.5. | Técnicas e instrumentación para la recolección de datos | 28 |
| 3.6. | Entorno de trabajo | 28 |
| 3.6.1. | <i>Software</i> de desarrollo | 28 |
| 3.6.2. | <i>Hardware</i> utilizado | 29 |
| 3.7. | Metodología | 29 |
| 4. | Desarrollo de la investigación | 32 |
| 4.1. | Desarrollo | 32 |
| 4.2. | Etapas 1 | 33 |
| 4.3. | Etapas 2 | 34 |
| 4.4. | Etapas 3 | 35 |
| 4.4.1. | Manejo de versiones | 36 |
| 4.4.2. | Filtrado de los datos | 36 |
| 4.4.3. | Exploración de datos | 37 |
| 4.4.4. | Ajuste del modelo | 39 |
| 4.4.5. | Validación del modelo | 40 |

| | |
|---|-----------|
| 4.4.6. Generación del reporte | 40 |
| 4.4.7. Visualización en el mapa | 40 |
| 4.5. Etapa 4 | 41 |
| 5. Conclusiones y recomendaciones | 43 |
| 5.1. Recomendaciones | 44 |
| A. Definición de Términos | 45 |
| B. Manual de usuario | 49 |
| Bibliografía | 59 |

Índice de Cuadros

| | |
|--|----|
| 2.1. Media de CD4 y el error estándar en el tiempo. Se dan resúmenes separados para los grupos definidos por el nivel basal de la carga viral. Belle, G. (2004). | 13 |
| 2.2. Clasificación Clínica de la OMS de la infección por el VIH establecida. OMS (2009). | 18 |
| 2.3. Clasificación inmunológica propuesta por la OMS para la infección establecida por el VIH. OMS (2009). | 19 |
| 2.4. Mortalidad por VIH/SIDA según año y sexo. Venezuela. 2002-2011. República Bolivariana de Venezuela (2014). | 20 |
| 3.1. Características de los computadores a usar | 29 |

Índice de figuras

| | |
|--|----|
| 2.1. Una muestra de trayectorias CD4 individuales de los datos del Estudio Multi-Centro de Cohorte contra el SIDA. Belle, G. (2004). | 13 |
| 2.2. Las trayectorias T^+CD4 individuales de los datos de la carga viral. Belle, G. (2004). | 16 |
| 2.3. Mapa del Estado Mérida con sus municipios. Alcaldia del Estado Mérida (2016) | 21 |
| 2.4. Interfaz de RStudio (Versión 1.1.453) | 23 |
| 2.5. Aplicación en Shiny | 24 |
| 3.1. Modelo espiral del proceso de <i>software</i> Boehm B. (1988). | 30 |
| 3.2. Wireframe base de la interfaz gráfica de usuario | 31 |
| 4.1. Configuración básica de la aplicación | 34 |
| 4.2. Creación de los manuales del paquete | 35 |
| 4.3. Cuenta alojada de GitHub de la aplicación | 36 |
| 4.4. Implementación en R de la carga de archivo | 37 |
| 4.5. Algoritmo usado para filtrar la información | 38 |
| 4.6. Implementación en R del gráfico de torta | 38 |
| 4.7. Implementación en R del histograma | 38 |
| 4.8. Implementación en R de la selección de las cargas | 39 |
| 4.9. Fórmula utilizada en el modelo lineal mixto | 39 |
| 4.10. Implementación de los residuos de la validación del modelo | 40 |
| 4.11. Estructura básica del archivo reporte.Rmd | 40 |
| 4.12. Implementación en R de la visualización del mapa | 41 |
| 4.13. Paquete cargado en las librerías de R | 42 |
| B.1. Pantalla de inicio de HIVmlm | 49 |
| B.2. Pantalla de la carga del archivo | 50 |

| | |
|---|----|
| B.3. Carga de archivo en la aplicación | 50 |
| B.4. Selección del archivo a cargar | 51 |
| B.5. Información del archivo | 51 |
| B.6. Pantalla de la visualización del mapa | 52 |
| B.7. Selección de un punto azul sobre el mapa | 52 |
| B.8. Visualización del módulo de Exploración de datos | 53 |
| B.9. Pantalla de los 3 paneles para visualizar graficamente los datos | 53 |
| B.10. Visualización de la gráfica de torta por género | 53 |
| B.11. Pantalla con el histograma de edades | 54 |
| B.12. Pantalla del selector de la sección cargas | 54 |
| B.13. Visualización del gráfico de puntos | 55 |
| B.14. Pantalla con el módulo ajuste del modelo | 55 |
| B.15. Pantalla con el selector del módulo de ajuste del modelo | 56 |
| B.16. Pantalla con la información del modelo lineal mixto | 56 |
| B.17. Pantalla selección de las variables de la validación del modelo | 57 |
| B.18. Gráfica de la validación del modelo | 57 |
| B.19. Pantalla del botón de descarga del reporte | 57 |
| B.20. Pantalla de guardar el reporte | 58 |

Introducción

El *software* R es un entorno de código abierto para la computación y graficación estadística. El *software* se compila y ejecuta en Windows, Mac OS X y Linux, distribuyéndose usualmente en formato binario para facilitar su instalación. El proyecto de *software* R fue iniciado por Robert Gentleman y Ross Ihaka. En R, la unidad fundamental de código compartible es el paquete o la librería, el cual agrupa código, datos, documentación y pruebas, y resulta simple de compartir con otros. Para enero del 2015 ya habían más de 6.000 paquetes disponibles en la Red Integral de Archivos de R, conocido comúnmente por su acrónimo CRAN, el cual es el repositorio de paquetes. Esta gran variedad de paquetes es una de las razones por las cuales R es tan exitoso, pues es probable que algún investigador o académico ya haya resuelto un problema en su propio campo usando esta herramienta, por lo que otros usuarios simplemente podrán recurrir a ella para su uso directo o para llamarla en un nuevo código.

El diagnóstico de una enfermedad basado en la historia clínica presentada por el paciente es un proceso complejo, involucra síntomas específicos así como estudios realizados y afecciones presentadas en el tiempo. Una de las enfermedades con más incidencia en la humanidad ha sido el Virus de Inmunodeficiencia Humana (VIH), el agente al entrar en el torrente sanguíneo por cualquiera de las formas de transmisión, provoca un proceso de infección, en cierto tipo de células del sistema inmunológico. Estas células generan anticuerpos producidos como una reacción de contraataque ante la presencia del VIH, como los linfocitos T^+CD4 ; según los análisis clínicos dichas células debén estar en un rango normal de 500 a 1600 células por milímetro cubico de sangre (mm^3). Cuando estas cifras bajan significativamente a 200 células de T^+CD4 por cada mm^3 de sangre, se dice que el sistema inmunológico está muy débil indicando el desarrollo del Síndrome de Inmunodeficiencia Adquirida (SIDA).

En ese orden de ideas, la epidemia del VIH presenta muchos factores de gran interés para ser estudiados y analizados bajo un enfoque temporal. En tal sentido, en el estudio epidemiológico se tiene como objetivo principal descifrar las relaciones que representan las

variables persona y tiempo; este último muchas veces no es considerado importante por lo tanto implica su estudio, dichas razones dan espacio para el análisis del patrón temporal en eventos de salud sea fundamental para entender la exposición al virus y prevenir eventos en el futuro, más allá de si el proceso es contagioso, influenciado por el ambiente o relacionado a la variabilidad genotípica.

Desde el punto de vista temporal, la epidemiología aporta grandes aspectos como lo son la determinación de los factores y los marcadores biológicos asociados a la progresión y la efectividad de los tratamientos en la enfermedad, a lo largo del tiempo, por ende es de gran importancia en el enfoque para el estudio de datos longitudinales para el seguimiento de la infección

La metodología de investigación a aplicar es la de desarrollo de software en espiral, la cual consiste en la evaluación de los riesgos establecidos previamente en la planificación y el análisis de los datos por utilizar, en esta ocasión son representativos de cada uno de los valores expuestos y obtenidos a través de exámenes clínicos realizados a cada uno de los pacientes en un determinado período, dicha investigación refleja un gran aporte científico y médico, con un enorme grado de importancia para la medicina, por lo tanto actualmente no se utiliza ningún sistema para el manejo y análisis de datos.

No obstante en cada uno de los capítulos que conforman esta investigación, se expone y describe de forma detallada el problema del caso de estudio, el objetivo principal y específicos, la justificación y su alcance, acompañados con el marco teórico, englobando todos aquellos conceptos y teorías indispensables para la mejor comprensión del tema, como también el apartado metodológico empleado y el cual se llevó a cabo para la completa resolución de la investigación, por último el desarrollo y ejecución del proyecto.

Capítulo 1

El problema

1.1. Planteamiento y formulación

Actualmente todas las organizaciones se ven forzadas al uso de la tecnología como herramienta para desenvolverse con mayor eficiencia y eficacia en el desempeño de sus funciones. En la gestión de la salud, la evaluación clínica de los pacientes genera una serie de información médica y administrativa sobre los mismos. Esta información a su vez, se registra en varios documentos físicos o en formato digital, siendo el conjunto de éstos lo que constituye la historia clínica.

En los hospitales y centros clínicos ambulatorios del Estado Mérida, así como en los diversos establecimientos de salud del país, la historia médica es un componente básico en el registro y seguimiento de los pacientes, así pues se tratan en dichas instituciones; por ende este componente debe ser una herramienta de trabajo, garantizando el adecuado tratamiento de todos los pacientes.

No obstante, tanto el incremento del uso de la tecnología y su evolución no son motivo suficiente para justificar el uso de esta aplicación, hay razones importantes, como lo es el optimizar la búsqueda de información registrada en el tiempo, acerca de una enfermedad o síntoma entre otros tópicos, ya pues en muchas ocasiones la información obtenida no se encuentra almacenada en varios formatos o medios para facilitar su uso. Sino se encuentran solo en archivos originales o no se han registrado de manera formal.

Para los pacientes con VIH en Venezuela, se debe realizar una historia clínica completa,

el cual debe contemplar la fecha de la primera serología positiva para el VIH y tipo de prueba realizada, conteo de linfocitos T^+CD4 y la carga viral, identificación de coinfecciones, describir la sintomatología presente y revisar algunos antecedentes familiares con respecto a enfermedades cardiovasculares; todo ello se realiza de manera manual en la mayoría de los casos, o disponen de un sistema no muy actualizado, y no se sabe con exactitud cuál es el estado actual de los pacientes, si están siguiendo el tratamiento, si han presentado una mejora y no volvieron al tratamiento o en el peor de los casos, abandonaron el tratamiento por presentar un desmejoramiento.

El tratamiento ha evolucionado con respecto al tiempo. Con relación a esto [MPPS \(2014\)](#) indica que:

Los conocimientos científicos han evolucionado de manera sorprendente en los últimos años en relación a la estructura molecular, la etiopatogenia y el tratamiento de la infección por el virus de inmunodeficiencia humana (VIH), lo que ha traído con consecuencia al desarrollo de nuevas tecnologías incluyendo nuevos medicamentos con diferentes mecanismos de acción que son capaces de inhibir la replicación viral y con ello detener la progresión de la enfermedad y disminuir la mortalidad, así como, coformulaciones que permiten mejor prescripción y mejor calidad de vida de las personas con VIH.

Esta evolución de la ciencia, ha traído consigo una tasa de cambio, entre crecer y declinar diversas transiciones entre estados de salud, una correlación del efecto en el tiempo.

Para tener un mejor seguimiento del tratamiento, por parte de los científicos, es indispensable determinar la progresión del tratamiento y la enfermedad, pudiendo predecir el avance de la misma, lo cual, no se dispone de tales herramientas resultando un poco engorroso por la cantidad de información proporcionada por los pacientes, con la necesidad de hacer seguimiento de varios años para adecuar los tratamientos, según los estadios en que se encuentren los pacientes en ese momento.

Con base en esto, se propone el desarrollo de un sistema para permitir hacer un análisis temporal haciendo uso de técnicas de predicciones, el cual incluyen aspectos como la infección inmunológica y virológica, la fecha del diagnóstico, el comportamiento de la infección en determinados años, el aspecto de la zona de vivienda, tanto si es urbano o rural, y en que

influye en su comportamiento, las comorbilidades con otras infecciones; resulta interesante realizar este enfoque temporal para comprender, intervenir y prevenir la diseminación de esta enfermedad.

En este caso se quiere realizar un análisis temporal, en el Estado Mérida, representado o confeccionado por la fecha de las revisiones, la fecha del tratamiento y su diagnóstico; también permita determinar los factores y los marcadores biológicos asociados a la progresión de la enfermedad y evaluar la efectividad de los tratamientos, determinando el método de supervivencia, el periodo de incubación del SIDA y la seroconversión al VIH; lo cual propone una solución a la problemática anteriormente expresada.

Es por esta razón, en el presente trabajo se propone la implementación de una aplicación web haciendo uso del análisis temporal para el estudio de pacientes con VIH, como caso de estudio particular pacientes con VIH del programa para ITS/VIH/SIDA del Instituto Autónomo Hospital Universitario de los Andes (IAHULA); desde el año 2007, ubicado en la ciudad de Mérida, Estado Mérida, Venezuela, por criterios de ética y marco legal, el autor y tutor de este trabajo declararán desconocer información para facilitar la identificación particular de ningún paciente, poniendo de alguna forma en riesgo su privacidad.

Adicionalmente, esta investigación podrá servir de apoyo para el desarrollo de futuras aplicaciones tecnológicas en el campo de los vectores virales, no solo será representado por pacientes de VIH sino demás enfermedades ya pues representan cargas virales, partiendo de los procesos de detección y clasificación aquí expuestos.

1.2. Objetivos

1.2.1. Objetivo general

Desarrollar una aplicación web para el estudio de datos longitudinales para el seguimiento de la infección por el Virus de Inmunodeficiencia Humana. Caso de estudio: Instituto Autónomo Hospital Universitario de los Andes.

1.2.2. Objetivos específicos

- Diagnosticar la información obtenida en la base de datos suministrada por el Instituto Autónomo Hospital Universitario de los Andes.

- Analizar de manera exploratoria y descriptiva los datos para la selección de las variables a utilizar en el modelo a implementar.
- Implementar un modelo para el manejo de datos longitudinales en una aplicación web.
- Aplicar pruebas de funcionalidad en la aplicación.

1.3. Justificación e importancia

El desarrollo de un sistema para realizar un análisis temporal en pacientes con VIH representa una herramienta para el apoyo adecuado a expertos en la epidemiología, inmunología o cualquier área de la medicina, proporcionando una mejor atención a pacientes ya diagnosticados de VIH.

Además, el gobierno nacional no presenta ningún informe desde el año 2014 y ofrece cifras no reales con respecto a la incidencia y por ende crecimiento de la enfermedad y su mortalidad en el país, los datos de la prevalencia del VIH/SIDA en Venezuela, en población general son escasos, y resulta desalentador porque se destaca la carencia de estudios epidemiológicos *Oleta et al. (2012)* y en dicho informe se designa una cantidad exorbitante de dinero para tratamiento del cual no se sabe a ciencia cierta donde se dirige, es por ello que resulta conveniente esclarecer estos datos, y por ello es indispensable.

Las técnicas utilizadas en este trabajo especial de grado también constituirán un gran aporte para la Universidad Nacional Experimental del Táchira y el país, ya que dichas técnicas no han sido ampliamente usadas en el área de la salud, lo cual aporta un gran avance en la investigación de esta área en el país.

El estudio realizado podrá ser usado en futuras investigaciones, viendo cómo influyen algunos factores en el desplazamiento de la infección y su propagación, y no solo será para pacientes con VIH sino cualquier vector viral.

1.4. Alcance y limitaciones

La meta final es desarrollar un sistema informático, puesto sea capaz de determinar de forma automática, a través de un seguimiento de los chequeos médicos periódicos, como la enfermedad ha variado, si el tratamiento está haciendo efecto, y como influyen los factores

transversales y temporales en los pacientes, para lograr un mejor entendimiento de la progresión del virus en el huésped.

Una vez definido el alcance, es necesario denotar las limitaciones de la investigación, solo se limitará al estudio de los casos registrados de pacientes diagnosticados previamente de la infección, no abarcando otros posibles enfoques como lo es la selección de un mejor modelo para agrupar los datos asociados a los pacientes ni la detección de otras anomalías asociadas a la infección, lo cuales no están contemplados en los objetivos propuestos.

Capítulo 2

Fundamentos teóricos

2.1. Antecedentes

Los estudios de datos longitudinales han cobrado cada vez más importancia en el estudio de la epidemiología, dando énfasis en los modelos lineales mixtos, facilitando el desarrollo de múltiples investigaciones y aplicaciones, buscando aprovechar el alcance para generar soluciones a problemas reales y modelos de dichas técnicas para diversas aplicaciones. Entre algunas de las investigaciones que sirvieron de base para este trabajo de grado son:

Rúa et al. (2013) realizaron un artículo titulado: *Modelado del efecto de la variabilidad climática local sobre la transmisión de dengue en Medellín (Colombia) mediante análisis de series temporales* para la revista Biomédica, Instituto Nacional de Salud, Colombia; en esta investigación se empleó la incidencia del dengue como variable dependiente y como variables independientes, los factores climáticos registrados a escala semanal, utilizaron el programa expert modeler desarrollando un modelo, explicando mejor el comportamiento de la enfermedad mediante el modelo ARIMA, se seleccionaron las variables climáticas que tuvieron una relación significativa con la variable dependiente. Se evidenció una fuerte asociación entre el dengue y la precipitación permitiendo construir un modelo que ayuda a comprender la dinámica de transmisión.

León y Jaimes (2015) realizaron un artículo titulado: *Ronda Clínica y epidemiológica. Análisis de datos longitudinales* para la revista Latreia, Universidad de Antioquia, Medellín, Colombia; hacen un estudio de los datos longitudinales para evaluar de manera apropiada las medidas de un mismo sujeto que se repiten en el tiempo, indica que con esta herramienta es

más fácil entender los indicadores de cambio en procesos de salud y enfermedades, para la evaluación del efecto de diversas intervenciones terapéuticas, se estudian los modelos lineales mixtos.

[Muñoz et al. \(2009\)](#) realizaron un artículo titulado: *Tuberculosis en Barcelona: Modelo predictivo basado en series temporales* para la Revista Española de Salud Pública, Ministerio de Sanidad, servicios sociales e igualdad, Madrid, España; resaltan los modelos predictivos basados en series temporales que fueron utilizados en Cataluña con la gripe, usándolos como base en el estudio de la tuberculosis, se han asignado a partir de 1987 hasta 2008 la fecha de inicio del tratamiento, agregándose la incidencia cada 4 semanas, formando dos series temporales, una para autóctonos y otra para inmigrantes con una longitud cada una de 287 observaciones. El análisis estadístico lo realizaron con el paquete estadístico R, lo que arroja como resultados que en la población autóctona no presentó una media constante a lo largo de los años y en la población inmigrante presentó una tendencia creciente a lo largo de todo el estudio. La conclusión a la que llegan en el estudio es que las tendencias de nuevos casos de tuberculosis presentaron patrones diferentes en la población de autóctonos y la de inmigrantes. El procedimiento usado es sencillo de implementar y muy útil para predecir futuros valores de series temporales con diferentes tendencias.

[Tran \(2016\)](#) realizó un trabajo especial de grado titulado: *Comparative causal effect estimation and robust variance for longitudinal data structures with applications to observational HIV treatment* para la Universidad de California, Berkeley, Estados Unidos; esta investigación discute la aplicación y el desempeño comparativo de los estimadores robustos dobles para estimar el resultado medio específico de la intervención en ajustes longitudinales sin confundir la función del tiempo así como las variaciones correspondientes del estimador. En concreto, se centra en definir cuidadosamente los parámetros causales para evitar problemas de positividad conocidos, estimando estos parámetros utilizando la estimación basada en la pérdida mínima, comparándola con otros estimadores dobles robustos del mismo parámetro causal y estimando las variaciones correspondientes de una forma que demuestre errores válidos de Tipo I, manteniendo al mismo tiempo la potencia estadística. Se utilizó la estimación mínima basada en pérdidas para estimar el resultado medio específico de la intervención utilizando estimadores de candidatos adaptativos de aprendizaje de máquina de datos para los parámetros de molestia. Se abordan cuestiones prácticas, que incluyen definir cuidadosamente los parámetros causales y permanecer dentro de los límites implicados por el modelo estadístico mientras se utilizan los algoritmos de aprendizaje automático.

Lu (2016) realizó un trabajo especial de grado titulado: *Handling incomplete high-dimensional multivariate longitudinal data with mixed data types by multiple imputation using a longitudinal factor analysis model* para la Universidad de California, Los Angeles, Estados Unidos; desarrolló un modelo de imputación que resuelve el problema de datos faltantes basado en un análisis de factores y un modelo de efectos mixtos lineales. La cadena de Markov Monte Carlo se utiliza para ajustar el modelo, dibujar parámetros, variables latentes y valores perdidos iterativamente. El modelo de imputación fue escrito en un paquete R, probando el modelo de imputación desarrollado utilizando conjuntos de datos simulados en 32 escenarios y 2 hipotéticos, falta de datos de los mecanismos. Dos modelos competitivos PAN (Imputación Múltiple para Panel Multivariado o Datos en Clúster) y MICE (Imputación Múltiple usando Ecuaciones Encadenadas) también probados de la misma manera para comparación, para mostrar la necesidad de abordar el tipo de datos de alta dimensión y mixto continuo. En comparación con la ejecución de la simulación, la evaluación del programa R en una sola computadora, el programa utilizado para la evaluación de la simulación se ejecutó aproximadamente 600 veces más rápido. En el escenario más desfavorable probado, el coeficiente cuadrático subyacente es tan grande como 0,8 del coeficiente lineal, las tasas reales de cobertura de las estimaciones de intervalos del 95 % empiezan a descender por debajo del 90 %. Una aplicación a los datos de una odontología se muestran, en comparación con el PAN, NORM y un corredor anterior del método recientemente desarrollado.

Welch (2015) realizó un trabajo especial de grado titulado: *Implementation, evaluation and application of multiple imputation for missing data in longitudinal electronic health record research* para la Universidad de Londres, Reino Unido; los registros de salud electrónicos longitudinales son un recurso valioso para la investigación porque contienen información sobre muchos pacientes durante largos períodos de seguimiento. Lo cual se propuso adaptar, evaluar e implementar el doble algoritmo para imputar datos faltantes de grandes base de datos de atención. Para lograr esto, primero investigaron la extensión y los patrones de los datos faltantes en un estudio clínico longitudinal para los indicadores de salud asociados con el riesgo de enfermedad cardiovascular para determinar si es plausible. Además, desarrolló métodos para identificar y eliminar los valores atípicos, es decir, de datos con mediciones repetidas antes de la imputación. Finalmente, aplicó el doble algoritmo en THIN para modelar el riesgo de enfermedad cardiovascular y entender los factores asociados con una mayor reducción del colesterol total en pacientes con diabetes tipo II.

[Sharrow \(2013\)](#) realizó un paquete de R titulado: *HIV calibrated model life tables for countries with generalized HIV epidemics* para el repositorio de R *CRAN*; las funciones en este paquete producen un conjunto completo de tasas de mortalidad como una función de una combinación de la prevalencia del VIH y la esperanza de vida al nacer, la mortalidad infantil, o mortalidad infantil con mortalidad adulta.

[Rizopoulos \(2018\)](#) realizó un paquete de R titulado: *Join modeling of longitudinal and survival data* para el repositorio de R *CRAN*; un ensayo clínico aleatorizado en el que se recopilaron datos longitudinales y de supervivencia para comparar la eficacia y seguridad de dos medicamentos antirretrovirales en el tratamiento de pacientes que fracasaron o que fueron intolerante a la terapia con zidovudina (AZT).

2.2. Bases teóricas

Este apartado presenta una fundamentación teórica de los principales conceptos y teorías usadas en esta investigación, como lo son los datos longitudinales, análisis de los datos longitudinales, análisis temporal y VIH; los cuales sirven como soporte, necesarios para lograr el desarrollo de los objetivos propuestos.

2.2.1. Estudios longitudinales

Son estudios de cohortes o seguimiento, implica mediciones repetidas, estas mediciones son tomadas en intervalos de tiempo entre la exposición y el comienzo de la enfermedad, requiere de la experiencia de la población a largo plazo. Este estudio se basa en que un mismo individuo es observado en más de una ocasión, lo cual lo diferencia de los estudios de seguimiento, en los que los individuos son seguidos hasta la ocurrencia de un suceso tal, como la muerte o una enfermedad [Delgado y Llorca \(2004\)](#)

Los estudios longitudinales se dividen en dos grandes componentes, están reflejados los datos longitudinales y los datos panel, haciendo referencia a los datos panel, han estado últimamente en muchos trabajos de investigación, esto en gran parte, al avance en las bases de datos, lo cual hace la recolección de estos datos de individuos a lo largo del tiempo. Las condiciones son sencillas, debe haber individuos al cual, se les va a realizar la observación durante un determinado tiempo, lo cual representa una dimensión temporal, pero estos datos son usados frecuentemente en estudios econométricos.

En esta investigación solo estará centrada en los datos longitudinales, puesto que implica más de dos mediciones a lo largo de un seguimiento, es necesario más de dos, lo cual es una ventaja a la hora de realizar el estudio de los datos de los individuos. Aunque hay una mayor probabilidad que existan datos un poco inconclusos, pacientes que hayan decidido abandonar el tratamiento, y estos datos son faltantes.

2.2.2. Análisis de datos longitudinales

Los análisis de estos datos, incorporan un enfoque diferente, aprovechándose de las mediciones repetidas en los sujetos a lo largo del tiempo, permitiendo una inferencia, no solo poblacional sino a nivel individual en los cambios de un proceso a lo largo del tiempo o en las transiciones entre diferentes estados de salud y la enfermedad; también puede evaluar y separar los efectos de cohorte y edad, pueden formular la dependencia en el tiempo de este efecto, y un seguimiento sobre los acumulados hasta el inicio del mismo, asociando la exposición y efecto en cualquier momento de la historia de los pacientes y la exposición actúa. [Delgado y Llorca \(2004\)](#)

2.2.3. Datos longitudinales en el VIH

En particular, los datos longitudinales han demostrado el efecto de la infección del VIH como indicador de una función inmunológica como el conteo de las células T^+CD4 . Todo ello, con unas características base como la carga viral, inmediatamente después del diagnóstico de la seroconversión, asociado con un declive del conteo de las células T^+CD4 . [Belle et al. \(2004\)](#)

El cuadro 2.1 refleja que entre más alta sea la carga viral, mayor será la reducción de las células T^+CD4 .

| Línea base de carga viral | | | |
|---------------------------|-------|-------|-------|
| Año | Bajo | Medio | Alto |
| 0-1 | 744.8 | 638.9 | 600.3 |
| 1-2 | 721.2 | 588.1 | 511.8 |
| 2-3 | 645.5 | 512.8 | 474.6 |
| 3-4 | 604.8 | 470.0 | 353.9 |

Cuadro 2.1: Media de CD4 y el error estándar en el tiempo. Se dan resúmenes separados para los grupos definidos por el nivel basal de la carga viral. Belle, G. (2004).

En cada panel del recuento de T^+CD4 para un solo sujeto se representa frente a los tiempos que obtuvieron mediciones. Tales mediciones permiten la inspección de los patrones de respuesta individuales y si existe una fuerte heterogeneidad en las trayectorias. Además, los individuos pueden ser evaluados para el cambio en el tiempo. En la Figura 2.1 indica la mayoría de los sujetos son más o menos estables en sus mediciones con el tiempo, o tienden a ser decreciente.

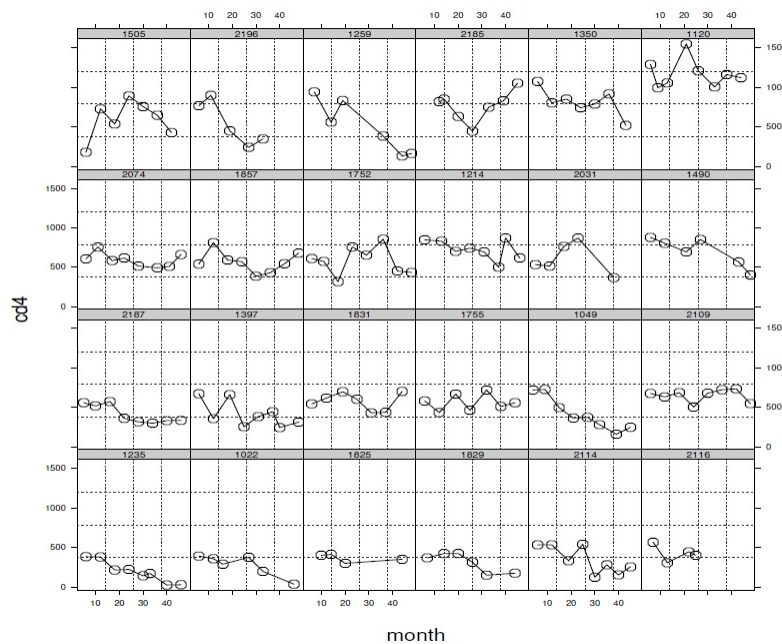


Figura 2.1: Una muestra de trayectorias CD4 individuales de los datos del Estudio Multi-Centro de Cohorte contra el SIDA. Belle, G. (2004).

En la situación común en el que se está interesado en la correlación de los resultados medidos a factores tales como el grupo de tratamiento o exposición también será útil para graficar series individuales por el grupo de covarianza. [Belle et al. \(2004\)](#)

2.2.4. Modelos lineales mixtos

Muchos modelos estadísticos se pueden expresar como modelos que incorporan dos partes: una parte son los efectos fijos, son parámetros asociados a toda la población o ciertos niveles de factores experimentales repetibles; otra parte son los efectos al azar, están asociados con los sujetos experimentales o unidades extraídas u observados al azar de la población. Los modelos con ambos, tanto el efecto fijo y el efecto al azar se denominan modelos mixtos.

Los modelos mixtos se utilizan principalmente para describir las relaciones entre una dimensión-variable de respuesta final o multidimensional y algunas covariables posiblemente relacionados con los datos observados, agrupandose de acuerdo con uno o más factores de agrupamiento. Tales datos deben incluir datos longitudinales, datos de medidas repetidas. [Zhang \(2009\)](#)

Para un solo nivel de agrupación de datos correlacionados, los modelos lineales mixtos especifican la n_i —dimensional responde al vector $Y_i = (Y_{i1}, Y_{i2}, \dots, Y_{inj})^T$ para $i = 1, 2, \dots, m$, con un número total de observaciones $\sum_{i=1}^m n_i$

$$Y_i = X_i\beta + Z_ib_i + \epsilon_i, i = 1, 2, \dots, m, \quad (2.1)$$

$$b_i \sim N(0, \Psi), \epsilon_i \sim N(0, \delta^2\Lambda_i), \quad (2.2)$$

Donde β es el vector p-dimensional de efectos fijos

los b_i son los vectores q-dimensionales de efectos al azar

X_i son $n_i \times p$ matrices regresivas de efectos fijos

Z_i son $n_i \times q$ matrices regresivas de efectos al azar

ϵ_i son las n_i —dimensionales sin un grupo de vectores de errores

ψ es una matriz definida simétrica positiva denotando la estructura de la varianza y covarianza entre los efectos al azar

b_i, λ_i son matrices definidas positivas denotando la estructura de la varianza y la covarianza dependientes sin grupo de errores.

Los efectos al azar b_i y el grupo sin errores ϵ_i se asumen que son independientes para los

diferentes grupos y para ser independientes entre sí con el mismo grupo. En los modelos lineales mixtos, se supone que la respuesta continua gaussiana es una función lineal de covariables con coeficientes de regresión que varían sobre los individuos, lo que refleja la heterogeneidad natural debido a factores no medidos. [Zhang \(2009\)](#)

2.2.5. Modelos lineales mixtos en el análisis de datos longitudinales

En medicina, se utilizan estudios longitudinales para descubrir factores predictivos de enfermedades. Los datos recogidos en un estudio longitudinal son datos longitudinales. En el análisis de datos longitudinales, los modelos lineales mixtos son ampliamente utilizados. Medidas repetidas en datos longitudinales se observan en los sujetos a lo largo del tiempo. Además de la variable de tiempo, normalmente se observan otras covariables, lo que a su vez requiere una selección más precisa para buscar el mejor modelo de una gran cantidad de modelos de candidatos. En el análisis de los datos longitudinales por modelos lineales mixtos, la selección del modelo también implica la selección de covariables y la variación de la selección de la estructura de covarianza de los efectos aleatorios y los errores dentro del grupo.

En la figura 2.2. Toma una muestra y traza líneas para cada sujeto estratificado por el nivel de carga viral de referencia. Esta cifra sugiere que la mayor carga viral del grupo tiene el menor recuento de T^+CD4 , y sugiere que la variación entre las mediciones también pueden ser menores para el grupo de carga viral basal, en comparación con los grupos medio y bajo. La figura 2.2 también se puede usar para identificar las personas que exhiben tendencias temporales que difieren notablemente de otras personas. En el grupo de alta carga viral hay un individuo que parece mejorar dramáticamente con el tiempo, y hay una sola medición inusual, donde el recuento de T^+CD4 es superior a 2000. El trazado de series individuales es un prelude exploratorio a un análisis estadístico confirmatorio más cuidadoso.

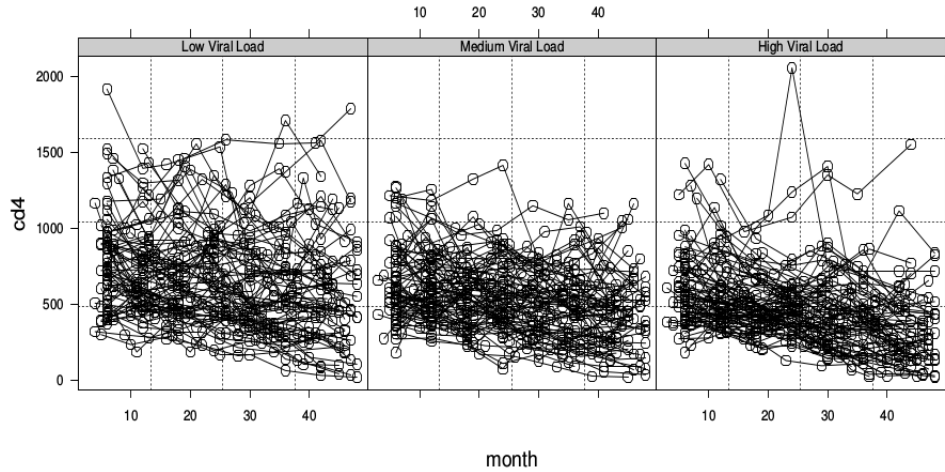


Figura 2.2: Las trayectorias T^+CD4 individuales de los datos de la carga viral. Belle, G. (2004).

2.2.6. VIH

El virus de la inmunodeficiencia humana es un retrovirus o virus de ácido ribonucleico (ARN), la conversión de ARN en ácido desoxirribonucleico (ADN) es la característica definitoria del virus, la cual se lleva a cabo mediante la acción enzimática de la transcriptasa inversa.

Según la [OMS \(2009\)](#), el VIH puede transmitirse por las relaciones sexuales vaginales, anales u orales con una persona infectada, la transfusión de sangre. Asimismo, puede transmitirse de la madre al hijo durante el embarazo, el parto y la lactancia.

La infección por VIH es silenciosa, porque no muestra ningún síntoma y suele ser diagnosticado por medio del estudio llamado: ensayo inmunoenzimático ligado a enzimas (ELISA). [Cuevas y Rebolledo \(2009\)](#)

2.2.7. SIDA

Se denomina así, al cuadro clínico que presenta los estadios más avanzados de la infección por VIH, es decir; se considera cuando la cuenta de linfocitos T^+CD4 se encuentra por debajo de las 200 células en un mililitro o cuando presenta el cuadro complejo de combinación de coinfecciones relacionadas con el VIH. [Cuevas y Rebolledo \(2009\)](#)

2.2.8. Relación VIH - T^+CD4 – T^+CD8

Dos pruebas se pueden utilizar para monitorear el avance del VIH, son los recuentos de las células T^+CD4 y las pruebas de carga viral. Los recuentos de T^+CD4 indican el estado del sistema inmunitario. Las pruebas de carga viral indican que tan activo está el VIH. Al comienzo estas pruebas deben hacerse con 2 a 4 semanas de diferencia para establecer el “punto de partida” (baseline). Luego deben repetirse aproximadamente cada 3 meses. Las pruebas de carga viral, los recuentos de T^+CD4 , proveerá una imagen más clara sobre el riesgo que la enfermedad avance, el estado del sistema inmunitario y la capacidad del organismo para combatir el VIH.

Normalmente, las pruebas sobre el recuento de los varios tipos de glóbulos blancos. Un tipo de glóbulos blancos son las células B, las cuales están involucradas en la producción de anticuerpos. Las células B también se enfrentan a las infecciones que están por fuera de las células. Otro tipo de glóbulos blancos son las células T^+CD8 , las cuales se enfrentan a las infecciones por dentro de las células. El tercer tipo de glóbulos blancos son las células T^+CD4 , las cuales ayudan a las células B y T^+CD8 a llevar a cabo su labor. Estos glóbulos blancos se denominan linfocitos.

En las personas VIH negativas, los recuentos normales de T^+CD4 fluctúan entre 600 y 1,500/ mm^3 . El recuento normal de T^+CD8 es de 300 a 800/ mm^3 . En general, una persona VIH negativa tiene 2 células T^+CD4 por cada célula T^+CD8 en la sangre. Sin embargo, en la enfermedad del VIH, el virus normalmente ocasiona una lenta disminución progresiva de las células T^+CD4 , y entre los que no están en terapia para el VIH es común ver invertida la relación entre T^+CD4 y T^+CD8 .

Los recuentos normales de T^+CD4 en las personas con VIH son de 350 a 800. Además de observar estos recuentos de células, es también útil observar los porcentajes relativos de células T^+CD4 y T^+CD8 . El porcentaje de T^+CD4 es el porcentaje de células T^+CD4 del total del recuento de glóbulos blancos. El rango normal es del 28 al 58 %. Un porcentaje de T^+CD4 por debajo del 14 % indica un diagnóstico de SIDA. [Ramduth \(2011\)](#)

2.2.9. Clasificación establecida de la infección por el VIH y enfermedades relacionadas

Según la Organización Mundial de la Salud, en 1990 se desarrolló un sistema de estadificación clínica de 4 estadios a efectos clínicos que solo está basado en adultos. Las clasificaciones clínicas de los centros para el control y la prevención de enfermedades de los Estados Unidos y de la OMS, reconocen la infección primaria por el VIH, también proponen la aparición de eventos nuevos o recurrentes de estadificación clínica y la clasificación inmunológica, sean usados para evaluar a los individuos cuando ya están recibiendo antirretrovirales. [OMS \(2009\)](#)

2.2.10. Clasificación clínica de la infección establecida por el VIH

Los eventos clínicos, se aplican para clasificar la enfermedad por el VIH en adultos afectados por el VIH se dividen en aquellos donde pueda hacerse un diagnóstico clínico presuntivo y aquellos que requieren un diagnóstico definitivo. En la Tabla se muestra el estadio clínico con su relación en términos simplificados para describir la variedad de síntomas relacionados con la infección por el VIH: asintomático, síntomas leves, síntomas avanzados y síntomas graves. La Tabla resume los eventos de estadificación clínica. [OMS \(2009\)](#)

| Síntomas asociados a la infección por el VIH | Estadio Clínico de la OMS |
|--|---------------------------|
| Asintomático | 1 |
| Síntomas leves | 2 |
| Síntomas avanzados | 3 |
| Síntomas graves | 4 |

Cuadro 2.2: Clasificación Clínica de la OMS de la infección por el VIH establecida. OMS (2009).

2.2.11. Clasificación inmunológica para la infección establecida por el VIH

Según sea la clasificación definida por el estado clínico o por el estado inmunitario, el paciente siempre requiere tratamiento antirretroviral, especialmente cuando la enfermedad es avanzada desde el punto de vista clínico, pero a veces puede retrasarse el inicio del tratamiento,

si el estado inmunitario indica que solo hay una inmunodeficiencia leve o insignificante y el paciente esta asintomático o solo tiene síntomas leves. [OMS \(2009\)](#)

| Valores de T^+CD4 relacionados con la edad | | | | |
|--|-----------------|-------------|-------------|---------------|
| Inmunodeficiencia asociada al VIH | ≤ 11 meses | 12-35 meses | 36-59 meses | ≥ 5 años |
| Ninguna | > 35 | > 30 | > 25 | > 500 |
| Leve | 30-35 | 25-30 | 20-25 | 350-499 |
| Avanzada | 25-29 | 20-24 | 15-19 | 200-349 |
| Grave | < 25 | < 20 | < 15 | $< 200mm^3$ |

Cuadro 2.3: Clasificación inmunológica propuesta por la OMS para la infección establecida por el VIH. [OMS \(2009\)](#).

2.2.12. VIH en Venezuela

En Venezuela se estima que la epidemia de VIH es concentrada, las personas que viven con el virus son aproximadamente 101.871, de las cuales 64 % corresponde al sexo masculino, pero se ha evidenciado no poseer mucha data correspondiente a la epidemia, solo se muestran casos muy antiguos y no se ha actualizado la información, los reportes avalados por el gobierno, no poseen información seria y confiable, no se dispone de mucha más información, solo la obtenida en 2012 sobre las características generales de la epidemia. [de Venezuela \(2014\)](#)

En la siguiente tabla se muestra la mortalidad en Venezuela desde el periodo 2007-2011, haciendo referencia que la tasa de mortalidad específica por cada 100 mil habitantes por causa para VIH/SIDA, aumentó y se incrementó en ambos sexos.

| Año | Hombres | Tasa | Mujeres | Tasa | Total | Tasa |
|------|---------|-------|---------|------|-------|------|
| 2007 | 1288 | 9.34 | 382 | 2.79 | 1670 | 6.08 |
| 2008 | 1223 | 8.73 | 409 | 2.94 | 1632 | 5.84 |
| 2009 | 1327 | 9.32 | 408 | 2.88 | 1735 | 6.11 |
| 2010 | 1380 | 9.55 | 450 | 3.13 | 1830 | 6.35 |
| 2011 | 1612 | 10.99 | 554 | 3.79 | 2166 | 7.40 |

Cuadro 2.4: Mortalidad por VIH/SIDA según año y sexo. Venezuela. 2002-2011. República Bolivariana de Venezuela (2014).

Sigue siendo una aspiración de los diversos actores involucrados en la prevención y control de la enfermedad y una prioridad para la comunidad nacional, poder disponer de la estimación y proyección de la situación epidemiológica del VIH/SIDA actualizada, paso indispensable para poder cumplir la Meta 7 de los Objetivos del Milenio de las Naciones Unidas: Haber detenido y empezado a revertir la incidencia del VIH-SIDA en el año 2015. La mortalidad por VIH/SIDA, se caracterizó por un aumento progresivo. La tendencia en el lapso de 21 años es al ascenso. *Oleta et al. (2012)*

Las fallas y deficiencias identificadas en los sistemas de vigilancia epidemiológica de los casos de VIH/SIDA, no permiten identificar oportunamente los cambios recientes en las tendencias de infección por el VIH, debido entre otras razones, al largo período de latencia de la enfermedad. La introducción de tratamientos antirretrovirales ha contribuido sustancialmente a la sobrevida de los pacientes.

Tomando a Venezuela como un todo, algunos estudios demuestran la alarmante situación del VIH en el país. En particular, porque no se dispone de mucha información oficial y verídica, permitiendo aclarar la situación, en este caso solo se toma en cuenta el caso particular del estado Mérida, haciendo énfasis en su geografía por municipios, para poder estudiar mejor los cambios que ha devenido en el tiempo.



Figura 2.3: Mapa del Estado Mérida con sus municipios. Alcaldía del Estado Mérida (2016)

2.2.13. Lenguaje R para el computo estadístico

El entorno de programación R es un clon de los lenguajes S y S-plus, de tal forma que muchos programas escritos en S y S-plus pueden ejecutarse en R sin modificaciones. S y S-plus son lenguajes de muy alto nivel diseñados para: La exploración y visualización de datos; el modelado estadístico; y la programación con datos.

R es un entorno de programación para el análisis estadístico y gráfico de datos, que cada vez se hace más popular entre los investigadores de todas las disciplinas. Tiene muchas ventajas y es oportuno y pertinente para los investigadores de cualquier área del saber.

Como software libre es aprobado por varios motivos: transmite valores socialmente positivos (libertad individual, conocimiento compartido, solidaridad y cooperación); se aproxima al método científico, porque permite el examen y mejora del código desarrollado por otros usuarios y la reproducibilidad de los resultados obtenidos; pueden adquirirse de manera legal y gratuita copias del programa, sin necesidad de licencias personales o académicas.

Aparte de su faceta de software libre, R tiene algunas ventajas específicas: por ejemplo, su sintaxis básica es sencilla e intuitiva, lo que se traduce en un aprendizaje rápido y cómodo; además, tiene una enorme comunidad de usuarios, estructurada alrededor de la *Comprehensive R Archive Network*, CRAN, lo que desarrolla cada día nuevos paquetes que extienden sus funcionalidades y cubren casi todas las necesidades computacionales y estadísticas de un científico.

2.2.14. Adquisición y licencia

El entorno R es un software libre en código fuente bajo la definición dada en la licencia GNU (General Public License) de la FSF (Free software foundation), el cual puede ser descargado ya sea como código fuente o como un ejecutable para los sistemas operativos Linux, Windows o MacOS.

2.2.15. Interfaz de usuario

La interacción con el usuario se basa en una interfaz de usuario llamada RStudio, ambiente de desarrollo integrado que permite una interacción rápida y amigable con R, además del desarrollo de código de forma interactiva.

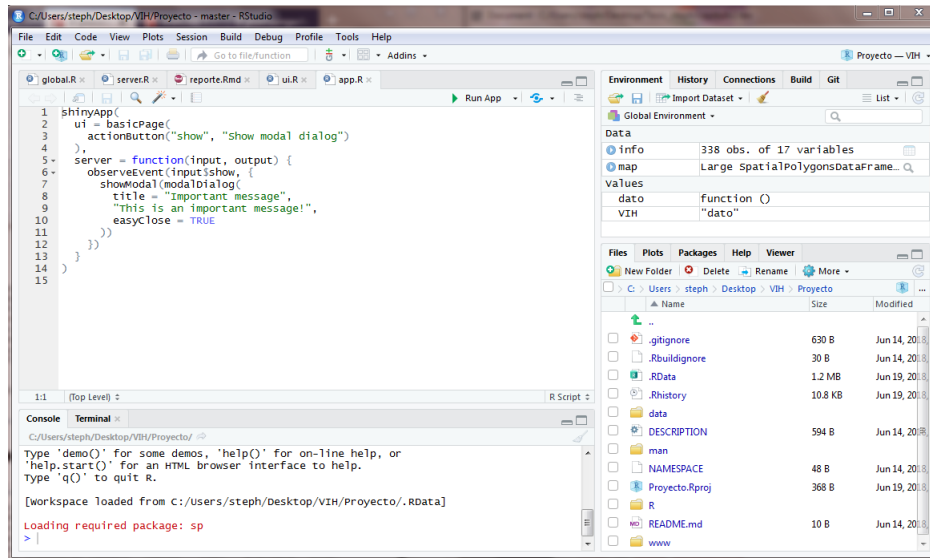


Figura 2.4: Interfaz de RStudio (Versión 1.1.453)

2.2.16. Lenguaje de programación

En su gramática, la sintaxis del lenguaje R es similar a la de C y C++, pero su semántica sigue los paradigmas de la programación funcional y la programación orientada a objetos, lo que implica que el lenguaje tiene la capacidad de manipular directamente los objetos del lenguaje, aplicar reglas de sustitución y evaluar expresiones. [Grunsky \(2002\)](#)

R es un lenguaje orientado a objetos, tal que, inclusive los tipos más básicos de datos, tales como: booleanos, enteros, reales, caracteres, vectores, matrices, listas y hojas de datos son objetos mismos. Esta característica permite que el usuario interactúe de forma transparente, ya que las llamadas se realizan a funciones genéricas, como print, summary o plot, las cuales determinan internamente que método debe ser llamado dependiendo de la clase de objetos a las que pertenecen sus argumentos.

R soporta internamente dos implementaciones para la programación orientada a objetos llamadas S3, que fue diseñado para uso interactivo, y S4, el cual supera las deficiencias de S3, y adiciona nuevos elementos. Adicionalmente, el usuario puede definir sus propias clases y los métodos asociados a ellas.

2.2.17. Shiny

Shiny es un paquete R que facilita la creación de aplicaciones web interactivas (aplicaciones) directamente desde R. Las aplicaciones Shiny están contenidas en un solo script llamado `app.R`. La secuencia de comandos `app.R` ubicado en un directorio (por ejemplo, `nuevo/`) y la aplicación se puede ejecutar con `runApp("nuevo")`.

`app.R` tiene tres componentes:

- un objeto de interfaz de usuario
- una función de servidor
- una llamada a la función `shinyApp`

El objeto de la interfaz de usuario (`ui`) controla el diseño y la apariencia de la aplicación. La función del servidor contiene las instrucciones que la computadora necesita para construir la aplicación. Finalmente, la función `shinyApp` crea objetos de la aplicación Shiny desde un par de UI / servidor explícito.



Figura 2.5: Aplicación en Shiny

Capítulo 3

Fundamentos metodológicos

Se planteó la estructura a seguir por el trabajo, detallando el enfoque, tipo, nivel, diseño de la investigación y la metodología implementada entre otros, el cual fue de vital importancia para el desarrollo de la misma.

3.1. Enfoque de la investigación

La investigación se desarrollo siguiendo un enfoque cuantitativo, como lo indican [Pallela y Martins \(2012\)](#), “la investigación cuantitativa requiere el uso de instrumentos de medición y comparación, proporcionando datos cuyo estudio necesita la aplicación de modelos matemáticos y estadísticos, el conocimiento está basado en hechos”.

3.2. Tipo o nivel de investigación

Este proyecto planteó un tipo de investigación de campo, según como lo indican [Pallela y Martins \(2012\)](#), la investigación de campo “consiste en la recolección directamente de la realidad donde ocurren los hechos, sin manipular o controlar variables” ya que permite indagar los efectos de la interrelación entre los diferentes tipos de variable en lugar de los hechos.

En este punto se debe determinar la profundidad que abarca esta investigación, teniendo en cuenta que de acuerdo con [Arias \(2012\)](#) el nivel de la investigación es definido como “grado de profundidad con que se aborda un fenómeno u objeto de estudio”

En este sentido, se tiene que dadas las características del proyecto, se asocia con un nivel descriptivo, tal como lo indican [Pallela y Martins \(2012\)](#), “hace énfasis sobre conclusiones dominantes o sobre como una persona, grupo o cosa se conduce o funciona en el presente” esto debido a que se medirán los datos extraídos sin alterarlos para ser mostrados en el sistema.

Cuando se habla de un nivel descriptivo junto con una investigación de tipo de campo, en ella no se formulan hipótesis y las variables se enuncian en los objetivos de la investigación desarrollada.

3.3. Diseño de investigación

Según [Arias \(2012\)](#), el diseño de la investigación es “la estrategia general que adopta el investigador para responder al problema planteado” puesto es vital establecer una correcta secuencia de pasos para elaborar el prototipo de *software* dando solución a la problemática principal de la investigación.

Con este enfoque, es evidente entonces, el trabajo siguió un diseño no experimental para datos longitudinales retrospectivos, enfocado en el uso de información existente, de acuerdo con lo dicho por [Pallela y Martins \(2012\)](#) al definir el diseño no experimental como:

Es el que se realiza sin manipular en forma deliberada ninguna variable. El investigador no sustituye intencionalmente las variables independientes. Se observan los hechos tal y como se presentan en su contexto real y en un tiempo determinado o no, para luego analizarlos. Por lo tanto, este diseño no se construye una situación específica sino que se observan las que existen. Las variables independientes ya han ocurrido y no pueden ser manipuladas, lo que impide influir sobre ellas para modificarlas.

Esto indica que no hubo manipulación de variables. Esta investigación presenta una modalidad de proyecto especial que, como lo indican [Pallela y Martins \(2012\)](#), los proyectos especiales “destinados a la creación de productos que puedan solucionar deficiencias evidenciadas, se caracterizan por su valor innovador y aporte significativo”, ya que se creará un *software* aplicable al área de estudio.

3.4. Población y muestra

Según [Morles \(1994\)](#) la población o universo al conjunto para el cual fueron validas las conclusiones que se obtuvieron de los elementos o las unidades (personas, instituciones o cosas) involucradas en la investigación.

En este sentido, se tiene que para las fases de elaboración y pruebas del prototipo de *software*, se planteó utilizar archivos con extensión .csv, como formato de datos de entrada a la aplicación.

La información para los archivos de prueba proviene de vistas minables de la base de datos de tipo relacional, desarrollada en *PostgreSQL* del Laboratorio de Investigaciones Hormonales del Instituto Autónomo Hospital Universitario de los Andes, que alberga al programa para pacientes con VIH, que cuenta con una población registrada de 45.649

Esta información fue filtrada, par considerando los siguientes criterios:

- La base de datos del Laboratorio de Investigaciones Hormonales IAHULA, no sólo almacena información del Programa de Pacientes con VIH, las vistas minables estan conformadas por pacientes confirmados por las pruebas *Elisa* y Western Blot para VIH
- En las vistas minables para las pruebas sólo se categorizarán pacientes con residencia en el Estado Mérida
- Las vistas minables sólo las integraran pacientes con más de tres años en seguimiento. Dado que los modelos son para datos longitudinales.
- Se categorizaron los pacientes por semestres y en cada lapso en medicados y no medicados.

Bajo estas condiciones se generó una vista minable (muestra) con 115 registros de pacientes, con una sección transversal conformada por la edad el género y el municipio de residencia y una sección longitudinal con información entre Enero del año 2007 y Diciembre del año 2013. sobre el Log_{10} de la carga viral plasmática y las sub-poblaciones linfocitaria de células T^+ CD4 y T^+ CD8.

3.5. Técnicas e instrumentación para la recolección de datos

En el desarrollo de este prototipo de *software* es de vital importancia contar con una correcta técnica de recolección de datos y un instrumento adecuado para dicha técnica, pues fueron la base para lograr un efectivo análisis estadístico, y así como eficaces resultados en la fase de pruebas. Todo esto, teniendo en cuenta que, según [Hurtado \(2000\)](#) la selección de técnicas e instrumentos de recolección de datos implica determinar por cuáles medios o procedimientos el investigador obtendrá la información necesaria para alcanzar los objetivos de la investigación.

Para efectos de esta investigación, se tuvo que la técnica de recolección de datos seleccionada es la observación no participante, indirecta y estructurada, la cual se ajusta a este de investigación, ya que como lo indican [Pallela y Martins \(2012\)](#) la observación es indirecta:

Cuando el investigador entra en conocimiento del hecho o fenómeno a través de las observaciones realizadas anteriormente por otra persona. Esto último ocurre cuando se utilizan libros, revistas, informes, grabaciones, fotografías, realizadas con lo que se esta investigando, los cuales han sido obtenidos o elaborados por personas que antes se ocuparon de lo mismo.

Esto debido, a que los datos usados provienen de la data existente del IAHULA, las cuales fueron tomadas previamente por otro investigador, por lo que no se intervino directamente en su recolección. Para esta fase de la investigación se realizó una exportación de las tablas que conforman la base de datos que almacenaban información sobre: Paciente, Pruebas, Resultados a archivos de texto plano, luego estos archivos fueron exportados a *Microsoft Office Excel*, con formato .csv donde los registros por pacientes fueron filtrados para conformar muestras anónimas y su dirección categorizarla por municipios.

3.6. Entorno de trabajo

3.6.1. *Software* de desarrollo

Para la realización del prototipo de *software*, fue necesario definir el entorno de desarrollo usado, se selecciono el lenguaje de programación R, dado que es un lenguaje de código abierto bajo el paradigma de desarrollo colaborativo. Como herramienta de desarrollo integrada

Integrated Development Environment, *IDE* para lenguaje R, se seleccionó RStudio, dado su bran abanico de paquetes e integración con repositorios Git que permite llevar un control de versiones del sistema apoyado en el servicio de alojamiento gratuito de GitHub.

3.6.2. *Hardware* utilizado

En la parte de hardware usado, se contó con dos computadoras (laptop y de PC de escritorio), con las características descritas en la tabla:

| Componente | Laptop | PC Escritorio |
|------------|--------------------|--------------------|
| CPU | Intel Core i3-380 | Intel Core i5-2310 |
| RAM | 4GB | 4GB |
| S.O | Windows 7 Ultimate | Windows 10 |

Cuadro 3.1: Características de los computadores a usar

3.7. Metodología

Para el desarrollo de este trabajo se implementó una metodología para el desarrollo del *software* en espiral la es un modelo del ciclo de vida del *software* donde el esfuerzo del desarrollo es iterativo, tan pronto termina un esfuerzo, ahí mismo comienza el otro; además en cada ejecución del desarrollo se sigue cuatro pasos principales:

1. Determinar o fijar los objetivos: se definen los objetivos específicos para posteriormente identificar las limitaciones del proceso y del sistema de *software*; además se diseña una planificación detallada de gestión y se identifican los riesgos.
2. Análisis de riesgo: se efectúa un análisis detallado para cada uno de los riesgos identificados del proyecto, definiendo los pasos a seguir para reducir los riesgos y luego analizarlos para planear las estrategias alternativas.
3. Desarrollar, verificar y validar: después de realizar el análisis de riesgo, se elige un paradigma para el desarrollo del sistema de *software* y se desarrolla.
4. Planificar: el proyecto se revisa, del cual se toma la decisión si se debe continuar con un ciclo posterior al de la espiral. Si se decide continuar, se desarrollan los planes para la siguiente fase del proyecto. [Boehm \(1988\)](#)

lo que hace que la aplicación fuera poco portable. Así que se toma la desición de cambiar el ambiente de desarrollo para la GUI.

Un segundo diseño de *software* fue desarrollado por un procedimiento cuádruple: Evaluación del primer prototipo en términos de sus fuerzas, debilidades, y riesgos; Definir los requisitos del segundo prototipo; Planeando y desarrollando el segundo prototipo; Construyendo y probando el segundo prototipo.

El segundo diseño de *software* mantuvo las características seleccionadas para el primer diseño sobre el cálculos y las herramientas gráficas pero la GUI se desarrolla empleando el paquete *Shiny* del IDE RStudio, lo que permite tener un servidor local para cada usuario.

Además, se construyó el sistema final al que se denominó HIVmlm (Human Immunology Virus linear model mixed effects), basado en el segundo diseño pero mejorando las características de la GUI en la Figura 3.3, se muestra el *wireframe* siguiendo el modelo basado en experiencia de usuario. El sistema final se evaluó y se probó con metodos de caja blanca y caja negra.

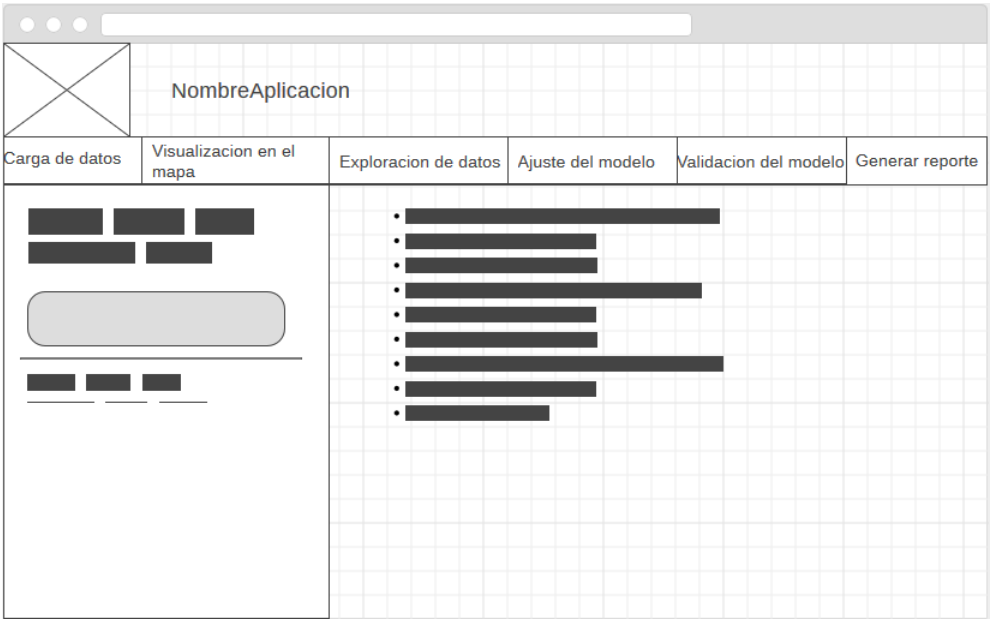


Figura 3.2: Wireframe base de la interfaz gráfica de usuario

Capítulo 4

Desarrollo de la investigación

Este capítulo comprende la exposición detallada del desarrollo de versión final del *software* siguiendo una metodología en espiral; se describe las diferentes iteraciones del ciclo y las fases de la elaboración del *software* al que se denominó HIVmlm (Human Immunology Virus linear model mixed effects).

4.1. Desarrollo

Siguiendo la metodología de trabajo en espiral, se construyó una lista de los módulos a desarrollar en la aplicación acorde a los objetivos planteados en la investigación, y siguiendo los paradigmas para el desarrollo de paquetes en lenguaje R, presentados por [Wickham \(2017\)](#) y por el creador del paquete *formatR* [Xie \(2017\)](#). Los módulos desarrollados fueron:

1. Carga de datos
2. Exploración de los datos
3. Visualización en el mapa
4. Ajuste del modelo
5. Validación del modelo
6. Generar reporte

Cada módulo paso por las etapas del modelo en espiral, en este caso, 4 etapas planificadas, estas etapas se realizaron dentro del tiempo establecido, con una revision semanal por parte de la tutora, dichas etapas se detallarán a continuación:

4.2. Etapa 1

Siguiendo cada una de las etapas de la metodologóa en espiral el desarrollo de *software*, se comienza por el análisis de riesgo y el diseño de prototipos. Para el desarrollo de la aplicación se seleccionó Lenguaje R.

Los métodos para el diseño de las funciones primarias en R serán los diagramas de flujo; y para su codificación se seguirán las normas de estilo para codificación en R, sugeridas por [Wickham \(2017\)](#) y por el creador del paquete *formatR* [Xie \(2017\)](#), además se establecerán la dependencia con las funciones de código base y las recomendadas para desarrollo en R.

Se desarrolló el módulo de Carga de Datos asociados a los pacientes, primero se filtró la información para el caso de estudio, la cual pertenece a la Base de Datos original del Laboratorio de Investigaciones Hormonales del Instituto Autónomo Hospital Universitario de Los Andes, esto con el objetivo de dar cumplimiento a las Leyes de protección para personas con VIH en Venezuela y las recomendadas por las Organizacion Panamericana (PAHO) y Mundial de la Salud (OMS), obteniendo información relevante para su representación en la aplicación y careciendo de forma absoluta de información de tipo personal o familiar.

Se procedió a desarrollar el analisis de riesgos, que incluye, en este caso, determinar las herramientas usadas para construir el prototipo de *software*, se utilizó para la elaboración del prototipo, el lenguaje de programación R y la herramienta *RStudio*, que es un entorno de desarrollo integrado (IDE) para el lenguaje R, dedicado al estudio de la computación estadística y gráfica, también se utilizo el paquete de R, *Shiny*.

El desarrollo del prototipo de *software* HIVmlm (Human Immunology Virus linear model mixed effects) partió de la lista de módulos antes expuestos.

4.3. Etapa 2

En la etapa 2, se decidió realizar, la creación del paquete desde *RStudio*, y así realizar la estructura que debe tener el paquete, adaptándose a *Shiny*. La creación del paquete de R, dadas sus ventajas siendo una de las principales, permitir probar el paquete sobre varias plataformas distintas (Linux, Mac y Windows) automáticamente, el código está públicamente disponible en la cuenta *GitHub*.

Estructura del paquete

La estructura es generada por `package.skeleton`, y sigue la siguiente: los archivos `NAMESPACE` y `DESCRIPTION`, y los directorios `R` y `data`, todo esto para mantener una colección coherente del código.

Archivo DESCRIPTION

El archivo `DESCRIPTION` contiene la siguiente información básica:

```
Package: Proyecto
Title: Aplicacion web para el seguimiento de la infeccion VIH
Version: 1.1
Authors@R: person("stephanie", "correa", email = "stephanie.correa24@gmail.com",
                  role = c("aut", "cre"))
Description: Desarrollo de una aplicacion web para el estudio de datos longitudi
Depends: R (>= 3.1.0)
License: MIT
Type: Package
LazyData: true
RoxygenNote: 6.0.1
Imports: ggplot2,
        plotly,
        influence.ME
```

Figura 4.1: Configuración básica de la aplicación

Archivo NAMESPACE

El archivo `NAMESPACE` lo crea *RStudio*, en base al archivo `DESCRIPTION` previamente creado, es de vital importancia este archivo, porque representa la autenticidad del paquete ante otros paquetes. Se genera al utilizar la siguiente expresión:

```
> library(devtools)
> library(roxygen2)
> load_all()
Loading Proyecto
```

Figura 4.2: Creación de los manuales del paquete

Se realizaron 3 archivos .R:

- global.R
- ui.R
- server.R

Por último, se realizó el reporte que debe seguir la extensión .Rmd, siguiendo la metodología del paquete Knitr, paquete que ayuda a la implementación del reporte en varios formatos.

El archivo global.R contiene todas las librerías asociadas para que funcione la aplicación, y la carga de los archivos asociados a la carga y visualización del mapa.

El archivo ui.R contiene la estructura gráfica que muestra la aplicación, con los módulos antes expuestos, se procedió a crear la estructura básica, con el diseño planteado por el wireframe, basado en la experiencia de usuario; conteniendo una estructura sencilla y funcional.

El archivo server.R contiene todos los cálculos y análisis estadístico, relativos a la aplicación, cada módulo realizado en el archivo ui.R tiene una función relacionada, con la elaboración y construcción de cada módulo. server.R contiene una función que recibe todas las entradas y salidas relativas del archivo ui.R.

Se realizó el plan de desarrollo para realizar la codificación de la aplicación.

4.4. Etapa 3

Se desarrollo la construcción de los archivos antes mencionados, con la elaboración de las funciones pertinentes para elaborar los diferentes modulos.

4.4.1. Manejo de versiones

Se desarrolló el *software* en *RStudio*, y se procedió a realizar un control de versiones de la aplicación por medio de las herramientas *Git* y *GitHub*, lo que permitió un mejor manejo de lo realizado en cada etapa de la investigación.

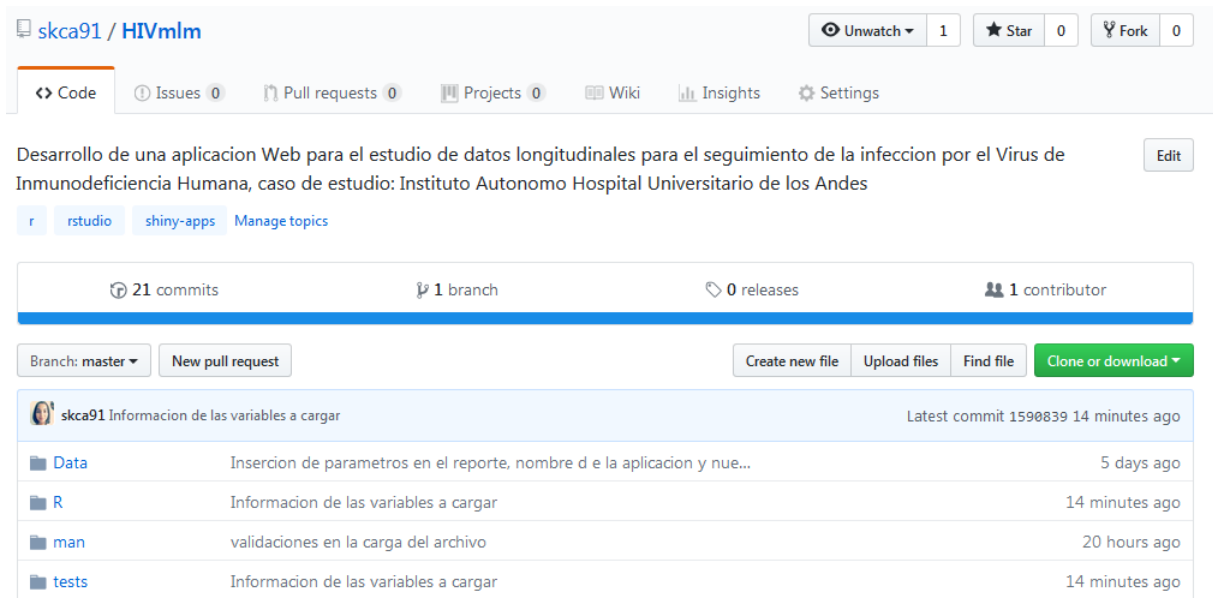


Figura 4.3: Cuenta alojada de GitHub de la aplicación

4.4.2. Filtrado de los datos

Se ha trabajado con datos recogidos por el Instituto Autónomo Hospital Universitario de los Andes del estado Mérida, desde el período 2007-1 hasta el 2013-2. Los datos registrados, en el período de estudio, son 1610. Los casos de VIH se han asignado a partir de la fecha de inicio del tratamiento, y en caso de no saberse, por la fecha de declaración. Se ha agregado la incidencia cada 6 meses (semestres), formando así dos series temporales por año.

Antes del filtrado de la información, se usaron datos de prueba para los diferentes pruebas con respecto a la carga del archivo, para esta carga se necesitaba un archivo con extension .csv, con cabecera bien definida dentro del archivo y los nombres de las variables que se necesitaban para la implementación, se conto con un archivo de extension .R, archivo usado para la creación del archivo .csv, una vez obtenido el archivo, se procedio a la carga de la información en la aplicación.

Una de las limitaciones en la carga de los archivos es la extensión, solo acepta extensión .csv, siguiendo unas reglas definidas, con respecto a los nombres de las variables.

En esta etapa se logró la elaboración del módulo de la carga de datos, luego de realizar varias pruebas, en este caso, se logró el prototipo 1 y los objetivos planteados para este prototipo, logrando así planificar la nueva etapa.

```
dato <-reactive({  
  file1 <- input$file  
  
  if(is.null(file1)){return()}  
  
  read.table(file1$datapath, sep = ",", header = TRUE)
```

Figura 4.4: Implementación en R de la carga de archivo

Para el desarrollo del módulo de Exploración de datos, se usaron las librerías de R, plotly y ggplot2.

4.4.3. Exploración de datos

Este módulo esta compuesto por 3 secciones, que comprende el género, la edad y las cargas con respecto a la carga viral plasmática, las células T^+CD4 y células T^+CD8 ; para la elaboración de la categoría por género, primero se filtraron los datos por períodos, tenemos los períodos de 20071 hasta 20132, cada período esta comprendido por semestres, es decir, un año esta compuesto por dos períodos, en este caso, se filtró la información, para omitir los datos faltantes con respecto a la muestra de la carga viral plasmática, es decir, solo se tomaron en cuenta, los pacientes que tuvieron un registro por cada período, tomando como resultado la cantidad de hombres y mujeres que asistieron al tratamiento durante determinado período; se logró con la implementación del algoritmo que se muestra en la figura 4.5.


```

idx <- with(dat, Peri == entradasInput() & is.na(CVP)==FALSE)
tre <- dat[idx,]
sexo <- c(tre[,5])
tiposex <- rep(NA,length(sexo))
tiposex[sexo==0] <- 'Masculino'
tiposex[sexo==1] <- 'Femenino'
frec.tipo.sex <- table(tiposex)

```

Figura 4.5: Algoritmo usado para filtrar la información

Se implementaron paneles para dividir los diferentes gráficos obtenidos.

Se implementó un gráfico de torta para sectorizar el género con respecto a los infectados con VIH. Se utilizó la siguiente función para su codificación.

```

plot_ly(tre, values = ~frec.tipo.sex, type = 'pie',
  textposition = 'inside',
  textinfo = 'percent',
  insidetextfont = list(color = '#FFFFFF'),
  hoverinfo = 'text',
  text = ~paste(frec.tipo.sex, ' personas'),
  marker = list(colors = colors,
    line = list(color = '#FFFFFF', width = 1)),
  showlegend = FALSE) %>%
layout(title = 'Gráfico del género por periodo',
  xaxis = list(showgrid = FALSE, zeroline = FALSE, showticklabels = FALSE),
  yaxis = list(showgrid = FALSE, zeroline = FALSE, showticklabels = FALSE))

```

Figura 4.6: Implementación en R del gráfico de torta

En la categoría por la edad, se realizó un histograma, y la debida implementación del código en R (Figura 4.7). Se realizó un filtro antes de realizar la graficación.

```

output$histogram <- renderPlotly({
  idx <- with(dato(), is.na(CVP)==FALSE)
  tre <- dato()[idx,]
  plot_ly(alpha = 0.6) %>%
  add_histogram(x = ~tre$Edad)%>%
  layout(barmode = "overlay")%>%
  layout(title = 'Frecuencia de edades',
    xaxis = list(title = 'Edad',
      zeroline = TRUE),
    yaxis = list(title = 'Frecuencia'))
})

```

Figura 4.7: Implementación en R del histograma

En la categoría por las cargas, para la mejor visualización de los datos, se utilizó un gráfico de puntos (Figura 4.8), las variables que son indispensables son la carga viral plasmática, las

células T^+CD4 y T^+CD8 .

```
output$line <- renderPlotly({
  variable <- selectcargaInput()
  plot_ly(dato(), x = ~as.character.Date(Peri), y = ~variable, type="scatter",
    mode = "markers", colors = "Set1")%>%
  layout(title = 'Comportamiento del conteo por semestre',
    yaxis = list(zeroline = FALSE,title="variable"),
    xaxis = list(zeroline = FALSE,title="Periodo"))
})
```

Figura 4.8: Implementación en R de la selección de las cargas

Las limitaciones con respecto a la exploración de datos, son los datos faltantes, por lo cual, se aplicaron los debidos filtros para omitir dichos datos.

4.4.4. Ajuste del modelo

Se planificó la elaboración de los módulos de Ajuste del modelo y Validación del modelo. El análisis de las tendencias mostradas en los módulos, se implementó mediante las librerías de R, lm4, influence.ME, nlme y lmerTest.

En el módulo Ajuste del modelo, se realizó el estudio de los datos longitudinales, las variables CVP, CD4 y CD8 son las utilizadas para aplicar el modelo lineal mixto, cada paciente presenta varias médisas de estas variables, con respecto a los años, las medidas NA representan los datos faltantes, que indican que un paciente decidio no ir mas al programa (patron monótono) o falta de forma intermitnte (patron no monótono), esta es una de las razones que permite la utilización de un modelo lineal de efectos mixtos, ya que esta técnica permite manejar, la alta heterogeneidad entre e intra pacientes, la correlación ente las medidas de un mismo paciente, así como los datos faltantes.

Se empleó la formula del modelo lineal mixto para hallar los valores asociados.

```
M0 <- lmer(LogCVP ~ 1 + Fecha + tre$CVP + (1|Pac), tre, REML = TRUE)
```

Figura 4.9: Fórmula utilizada en el modelo lineal mixto

El resultado de esta formula esta ajustado por REML.

4.4.5. Validación del modelo

El resultado gráfico corresponde a los resultados obtenidos en la formula de los modelos mixtos, y corresponde a la variación de los datos cargados en la aplicación (Figura 4.10).

```
output$valido <- renderPlot({  
  variable <- datosVinput()  
  M1RML <- lmer(variable ~ 1 + Fecha + Edad + EI + (Fecha|Pac), data = dato(), REML = FALSE)  
  qqmath (M1RML,id =0.05)|  
})
```

Figura 4.10: Implementación de los residuos de la validación del modelo

4.4.6. Generación del reporte

Se genero un archivo con extensión .Rmd llamado reporte.Rmd, con la finalidad de crear el reporte en formato que se pueda guardar y el usuario pueda revisar y analizar los datos, se utilizo la siguiente estructura básica para la generación del reporte (Figura 4.11).

```
---  
title: "Reporte sobre el seguimiento de la infeccion VIH"  
author: "Stephanie Correa"  
output: word_document()  
latex_engine: xelatex  
geometry: left = 2cm, right = 2cm, top = 2cm, bottom = 2cm  
mainfont: Arial  
params:  
  reml: NA  
  ajuste: NA  
  carga: NA  
  validacion: NA  
  periodo: NA  
---  
  
```{r global_options, include = FALSE}  
knitr::opts_chunk$set(message=FALSE,
 tidy.opts=list(width.cutoff=60))
```
```

Figura 4.11: Estructura básica del archivo reporte.Rmd

4.4.7. Visualización en el mapa

Para la visualización del mapa, se realizo la siguiente implementación de código.

```

output$mapa1 <- renderLeaflet({
  idx <- with(info, is.na(info$LNG)==FALSE)
  tre <- info[idx,]
  leaflet(map) %>% addTiles() %>% #addProviderTiles(providers$OpenStreetMap) %>%
    # addPolygons(fill = FALSE, stroke = TRUE, color = "#03F") %>% addLegend("bottomright",
    addCircleMarkers(lng = tre$LNG, lat = tre$LAT, radius = 10,color = c("blue"),
                      fillopacity = 0.3,stroke = FALSE,
                      popup = tre$NAME_2)
})

```

Figura 4.12: Implementación en R de la visualización del mapa

En esta etapa se estableció un plan de integración y pruebas, en caso de las pruebas unitarias, se realizaron en 2 fases, la primera con datos sintéticos que permitan comprobar cada estado del diagrama de flujo, que esquematiza la solución numérica que permite el cálculo de cada uno de los índices fisiológicos para entre otra herramientas pueden utilizarse los paquetes de R *RUnit* [Konig et al. \(2018\)](#) y *testthat* [Wickham \(2017\)](#), y la segunda etapa donde cada función asociada a un índice fisiológico se le realizan prueban con los conjuntos de datos de prueba que formaran parte integral del paquete y con los cuales se desarrollaran los ejemplos prácticos que conformaran la documentación que acompaña al paquete R.

4.5. Etapa 4

La etapa 4 esta compuesta solo por 3 fases, ya que en esta etapa se ha obtenido el prototipo operativo, el cual se han implementado las diferentes pruebas de caja blanca y caja negra.

En la etapa anterior, se realizó un plan de integración y pruebas, donde se estableció la utilización de las herramientas *RUnit* y *testthat*.

La implementación de la aplicación es la creación completa del paquete en R, donde el usuario puede descargar el paquete desde RStudio y utilizar la aplicación HIVmlm.

Human immunology virus linear model mixed effects



Documentation for package 'HIVmlm' version 0.1.0

- [DESCRIPTION file](#).

Help Pages

| | |
|---------------------------------|--|
| buscarCabeceras | Funcion buscarCabeceras |
| comprobar | Funcion comprobar |
| filename | Funcion filename |
| info | Carga de la informacion del mapa El 'read.csv' Permite cargar el archivo .csv para cargar la informacion del mapa |
| map | Carga del archivo .shp del mapa de venezuela El 'readOGR' Permite cargar el archivo .shp para cargar el mapa |
| server | Dato reactivo 'entradasInput' Permite crear un evento reactivo con el periodo que el usuario desee visualizar en la categoria por genero Dato reactivo 'datosVInput' |

Figura 4.13: Paquete cargado en las librerias de R

Capítulo 5

Conclusiones y recomendaciones

Se diagnosticó la información obtenida de la base de datos que fue suministrada por el Instituto Autónomo Hospital Universitario de los Andes, con el fin de filtrar los datos que allí estaban expuestos para preservar la identidad de los pacientes, y escoger los datos y variables que mejor se adaptaron al desarrollo de la aplicación. Para un mejor análisis de los resultados, se agruparon los pacientes por municipios del estado Mérida, con el fin de realizar un mejor ubicación geoespacial, con la debida visualización y representación en el mapa.

La base de datos una vez filtrada, permitió analizar de manera exploratoria y descriptiva, la selección de las variables que se implementaron en el modelo; las cuales hicieron referencia a los diferentes resultados de las pruebas con respecto a la seroconversion del VIH, muestras de datos que el paciente se hace con regularidad, es decir, el conteo de la carga viral plasmática, un dato relevante e indispensable para el modelo, así como el conteo de las células T^+CD4 y T^+CD8 . Algunos datos que fueron tomados en cuenta para el análisis fueron los demográficos: la edad, el género, el periodo y el municipio donde reside el paciente para esta investigación se trabajó como caso de estudio el estado Mérida.

Se implementó un modelo lineal mixto, porque es ampliamente utilizado en el análisis de datos longitudinales, específicamente en el área de la medicina en el seguimiento de cohortes de pacientes, además su implementación en R, permitió que pueda estimarse en un ambiente WEB, sin el uso de recursos computacionales de altas prestaciones.

Se realizaron las debidas pruebas de funcionalidad a la aplicación en cada etapa de desarrollo de la metodología, así como también las pruebas de caja blanca con las herramientas

de R antes mencionadas en el capítulo 4, permitiendo demostrar su operatividad en el manejo de los fallos.

Finalmente, destaca su uso como herramienta de análisis descriptivo de la incidencia del VIH en el estado Mérida para el estudio de datos de tipo longitudinal, el cual por estar desarrollado con una herramienta en la que se trabaja bajo el esquema colaborativo, es una herramienta de fácil mantenimiento y escalable, y aunque las pruebas de funcionabilidad se realizaron sólo considerando los marcadores inmunológicos y virológicos disponibles en las vistas minables obtenidos en el filtrado de los datos, también puede considerarse evaluar con el modelo lineal mixto, el impacto de las coinfecciones con otros virus, las condiciones de medicación, grupos de riesgo, entre otras variables.

5.1. Recomendaciones

Implementar el modelo a los diferentes estados del país, incluyendo las vistas minables de los otros estados que se encuentran almacenadas en la base de datos del Laboratorio de Investigaciones Hormonales del Instituto Autónomo Hospital Universitario de los Andes.

Unificar esfuerzos con organismos de salud públicos y privados para el desarrollo de una aplicación que unifique tanto la incidencia del virus como hacer análisis predictivos de la misma.

Apéndice A

Definición de Términos

Antirretroviral: Medicamento empleado para impedir la multiplicación de un retrovirus, como el VIH. Por lo general, el término se refiere a los medicamentos antirretrovirales contra el VIH.

Carga viral: Cantidad del VIH en una muestra de sangre. Se notifica como el número de copias de ARN del VIH por milímetro de sangre. Una meta importante del tratamiento antirretroviral (TAR) es reducir la concentración de carga viral de una persona a un nivel indetectable, que es demasiado baja para detectar el virus con una prueba de la carga viral.

Coinfección: es un término empleado cuando una persona tiene dos o más enfermedades infecciosas a la vez. Puesto que es posible infectarse con VIH y hepatitis C (VHC o Hep C para abreviar) por las mismas vías, hasta 3 personas de cada 10 infectadas con el VIH están infectadas también con el VHC.

Control de versiones: Es la gestión de los diversos cambios que se realizan sobre los elementos de algún producto o una configuración del mismo, es decir, a la gestión de los diversos cambios que se realizan sobre los elementos de algún producto o una configuración.

Comorbilidad: También conocida como "morbilidad asociada", es un término utilizado para describir dos o más trastornos o enfermedades que ocurren en la misma persona. Pueden ocurrir al mismo tiempo o uno después del otro. La comorbilidad también implica que hay una interacción entre las dos enfermedades que puede empeorar la evolución de ambas.

DESCRIPTION: Archivo que contiene información básica sobre el paquete.

Estudio longitudinal: Es un tipo de diseño de investigación que consiste en estudiar y evaluar a las mismas personas por un período prolongado de tiempo.

Git: Es un software de control de versiones diseñado por Linus Torvalds.

GitHub: Es una plataforma de desarrollo colaborativo de software para alojar proyectos

utilizando el sistema de control de versiones Git.

ggplot2: El paquete ggplot2 de R proporciona un poderoso sistema que hace que sea fácil de producir gráficos complejos de varias capas, automatiza varios aspectos tediosos del proceso de graficar manteniendo al mismo tiempo la habilidad de construir paso a paso un gráfico pues se compone de una serie de pequeños bloques de construcción independientes, esto reduce la redundancia dentro del código, y hace que sea fácil de personalizar el gráfico para obtener exactamente lo que se desea.

IDE: Una Infraestructura de Datos Espaciales (IDE) es un sistema de información integrado por un conjunto de recursos (catálogos, servidores, programas, datos, aplicaciones, páginas Web,...) dedicados a gestionar Información Geográfica (mapas, ortofotos, imágenes de satélite, topónimos,...), disponibles en Internet, que cumplen una serie de condiciones de interoperabilidad (normas, especificaciones, protocolos, interfaces,...), y que permiten que un usuario, utilizando un simple navegador, pueda utilizarlos y combinarlos según sus necesidades.

influence.ME: Calcula las medidas de influencia para modelos de efectos mixtos estimados con lme4. El razonamiento básico detrás de la medición de casos influyentes, es que cuando se omiten iterativamente unidades individuales a partir de los datos, los modelos basados en estos datos no deberían producir estimaciones sustancialmente diferentes.

ITS: son enfermedades infecciosas, que pueden transmitirse de una persona a otra durante una relación sexual vaginal, anal u oral. Las ITS afectan a todos independientemente de la orientación sexual o identidad de género. Desde el comienzo de la vida sexual puede estar expuesto/a a estas infecciones. Las producen más de 30 diferentes tipos de virus, bacterias y parásitos. Las más frecuentes son la sífilis, gonorrea, clamidia, herpes, hepatitis B y C, VIH y VPH.

IAHULA: Instituto autónomo hospital universitario de los andes.

Linfocito T^+CD4 : Los linfocitos (las células) T^+CD4 ayudan a coordinar la respuesta inmunitaria al estimular a otros inmunocitos, como los macrófagos, los linfocitos B y los linfocitos T^+CD8 para combatir la infección. El VIH debilita el sistema inmunitario al destruir los linfocitos CD4.

Linfocito T^+CD8 : Los linfocitos (las células) T^+CD8 reconocen y destruyen las células infectadas por microorganismos, como bacterias o virus.

lme4: Es un paquete de R que proporciona funciones para ajustar y analizar modelos mixtos lineales, modelos mixtos lineales generalizados y modelos mixtos no lineales. En cada uno de estos nombres, el término "mixto.", más completamente, "efectos mixtos", denota un modelo que incorpora términos de efectos fijos y aleatorios en una expresión predictiva a partir

de la cual se puede evaluar la media condicional de la respuesta.

lmerTest: Proporciona valores de p en anova de tipo I, II o III y tablas de resumen para ajustes del modelo de lmer a través del método de grados de libertad de Satterthwaite. Un método de Kenward-Roger también está disponible a través del paquete pbkrtest. Los métodos de selección de modelos incluyen tablas de pasos, drop1 y anova para efectos aleatorios (ranova). También están disponibles métodos para medios mínimos cuadrados (LS-means) y pruebas de contrastes lineales de efectos fijos.

Módulo lineal mixto: Es una herramienta flexible para ajustar otros modelos que se pueden formular como modelos lineales mixtos. Dichos modelos incluyen modelos multinivel, modelos lineales jerárquicos y modelos de coeficientes aleatorios.

NAMESPACE: Es una de las partes fundamentales de la construcción de un paquete. Tener un NAMESPACE de alta calidad ayuda a encapsular el paquete y hacerlo autónomo. Esto garantiza que otros paquetes no interfieran con el código, que el código no interfiera con otros paquetes y que el paquete funcione independientemente del entorno en el que se ejecute.

nlme: Ajusta y compara los modelos gaussianos de efectos mixtos lineales y no lineales.

Paquete R: Es una colección de funciones, datos y código R que se almacenan en una carpeta conforme a una estructura bien definida, fácilmente accesible para R.

plotly: Es una biblioteca gráfica de R, hace gráficos interactivos de calidad de publicación en línea. Ejemplos de cómo hacer gráficos de líneas, diagramas de dispersión, gráficos de áreas, gráficos de barras, barras de error, diagramas de cajas, histogramas, mapas de calor, subtramas, gráficos de ejes múltiples y gráficos 3D (basados en WebGL).

R: Es un lenguaje orientado a objetos, el cálculo estadístico y la generación de gráficos, que ofrece una gran variedad de técnicas estadísticas y gráficas. Es un entorno de análisis y programación estadísticos que, en su aspecto externo, es similar a S. Es un lenguaje de programación completo con el que se añaden nuevas técnicas mediante la definición de funciones.

Reactividad: Cuando una entrada (input) cambia, el servidor reconstruye cada salida (output) que depende de ella (también si la dependencia es indirecta).

REML: Máxima verosimilitud restringida.

Retrovirus: Tipo de virus que emplea el ARN como su material genético. Después de infectar una célula, un retrovirus emplea una enzima llamada transcriptasa inversa para convertir el ARN en ADN. Luego, el retrovirus integra su ADN en el ADN de la célula huésped, que le permite multiplicarse. El VIH, causante del SIDA, es un retrovirus.

RStudio: Es una interfaz que permite acceder de manera sencilla a toda la potencia de R, para utilizar RStudio se requiere haber instalado R previamente.

Seroconversión: Transición de la infección por el VIH a la presencia detectable de

anticuerpos contra ese virus en la sangre. Cuando ocurre seroconversión (por lo general, a las pocas semanas de la infección), el resultado de una prueba de detección de anticuerpos contra el VIH cambia de seronegativo a seropositivo.

Shiny: Es una herramienta para crear fácilmente aplicaciones web interactivas (apps) que permiten a los usuarios interactuar con sus datos sin tener que manipular el código.

SIDA: Es la enfermedad que se desarrolla como consecuencia de la destrucción progresiva del sistema inmunitario (de las defensas del organismo), producida por un virus descubierto en 1983 y denominado Virus de la Inmunodeficiencia Humana (VIH). La definen alguna de estas afecciones: ciertas infecciones, procesos tumorales, estados de desnutrición severa o una afectación importante de la inmunidad. La palabra SIDA proviene de las iniciales de Síndrome de Inmunodeficiencia Adquirida, que consiste en la incapacidad del sistema inmunitario para hacer frente a las infecciones y otros procesos patológicos. El SIDA no es consecuencia de un trastorno hereditario, sino resultado de la exposición a una infección por el VIH, que facilita el desarrollo de nuevas infecciones oportunistas, tumores y otros procesos. Este virus permanece latente y destruye un cierto tipo de linfocitos, células encargadas de la defensa del sistema inmunitario del organismo.

VIH: El virus de la inmunodeficiencia humana (VIH) infecta a las células del sistema inmunitario, alterando o anulando su función. La infección produce un deterioro progresivo del sistema inmunitario, con la consiguiente inmunodeficiencia”. Se considera que el sistema inmunitario es deficiente cuando deja de poder cumplir su función de lucha contra las infecciones y enfermedades. El síndrome de inmunodeficiencia adquirida (SIDA) es un término que se aplica a los estadios más avanzados de la infección por VIH y se define por la presencia de alguna de las más de 20 infecciones oportunistas o de cánceres relacionados con el VIH. El VIH puede transmitirse por las relaciones sexuales vaginales, anales u orales con una persona infectada, la transfusión de sangre contaminada o el uso compartido de agujas, jeringuillas u otros instrumentos punzantes. Asimismo, puede transmitirse de la madre al hijo durante el embarazo, el parto y la lactancia.

Wireframe: Un wireframe o prototipo no es más que un boceto donde se representa visualmente, de una forma muy sencilla y esquemática la estructura de una página web.

Apéndice B

Manual de usuario

Pantalla de inicio de HIVmlm



Figura B.1: Pantalla de inicio de HIVmlm

1. **Carga de datos:** Modulo de carga del archivo
2. **Visualización en el mapa:** Muestra la incidencia del virus en el mapa
3. **Exploración de datos:** Muestra el analisis grafico de los datos cargados
4. **Ajuste del Modelo:** Analisis del modelo lineal mixto
5. **Validación del modelo:** Muestra graficamente los datos obtenidos del modelo lineal mixto.

6. **Generar reporte:** Permite generar reportes con los datos de la aplicación

Módulo de carga de datos

HIVmlm

HIV

Carga de Datos

Visualizacion en el mapa

Exploracion de datos

Ajuste del modelo

Validacion del Modelo

Generar Reporte

Carga de Archivo

Browse...

Ningun archivo seleccionado

Antes de cargar la informacion, los nombres de las variables a cargar en el archivo deben coincidir con las mostradas en la tabla

Debe subir un archivo con extension .csv

| listaVariable | descripcion | ejemplos |
|---------------|--|--|
| Pac | Es el id del paciente | Es numerico: 1,2,3,4...100 |
| Sex | Es el sexo del paciente | Es un caracter: f=femenino, m=masculino |
| Genero | Es el genero del paciente | Es numerico: 0 es masculino, 1 es femenino |
| Peri | Periodo en el que se trata el paciente | Es una cadena de caracteres: 20071, 20132 |
| Edad | Edad en ese periodo de tiempo | Es numerico: 45,29 |
| CD4 | Conteo de las celulas CD4 | Es numerico: 3000,100,375 |
| CD8 | Conteo de las celulas CD8 | Es numerico: 4000,156,543 |
| CVP | Conteo de la carga viral plasmatica | Es numerico: 1000,432,20 |

Figura B.2: Pantalla de la carga del archivo

Carga de datos

Para cargar el archivo, se debe hacer click en **Browse** como lo muestra la figura B.3.

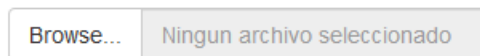


Figura B.3: Carga de archivo en la aplicación

Posteriormente se mostrará una ventana para seleccionar el archivo como se muestra en la figura B.4.

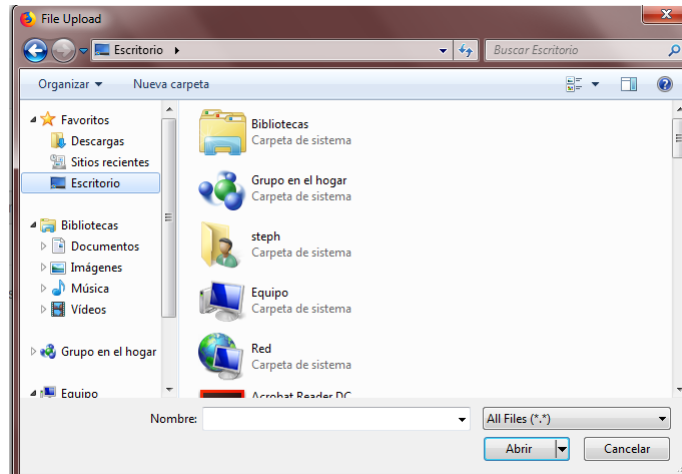


Figura B.4: Selección del archivo a cargar

Se mostrará la información en la aplicación.

[Acerca del archivo](#) [Información](#) [Resumen](#)

Show entries Search:

| X | Obs | Pac | Sex | Genero | EI | ZR | Peri | Fecha | Edad | HBSAG | ANTIHBC | CD4 |
|---|-----|-----|-----|--------|----|----|-------|-------|------|-------|---------|-----|
| 1 | 1 | 1 | m | 0 | 4 | 0 | 20071 | 1 | 45 | 0 | 0 | 362 |
| 2 | 2 | 1 | m | 0 | 4 | 0 | 20072 | 2 | 45 | 0 | 0 | 454 |
| 3 | 3 | 1 | m | 0 | 4 | 0 | 20081 | 3 | 46 | 0 | 0 | 321 |
| 4 | 4 | 1 | m | 0 | 4 | 0 | 20082 | 4 | 46 | 0 | 0 | 400 |
| 5 | 5 | 1 | m | 0 | 4 | 0 | 20091 | 5 | 47 | 0 | 0 | 413 |
| 6 | 6 | 1 | m | 0 | 4 | 0 | 20092 | 6 | 47 | 0 | 0 | 375 |
| 7 | 7 | 1 | m | 0 | 4 | 0 | 20101 | 7 | 48 | 0 | 0 | 324 |

Figura B.5: Información del archivo

Luego se puede navegar en la aplicación, con los diferentes modulos, para la exploración y el analisis de los datos.

Módulo de visualización del mapa

En este módulo, se detalla en el mapa, la incidencia de la ubicación y el contagio del virus en la localidad.

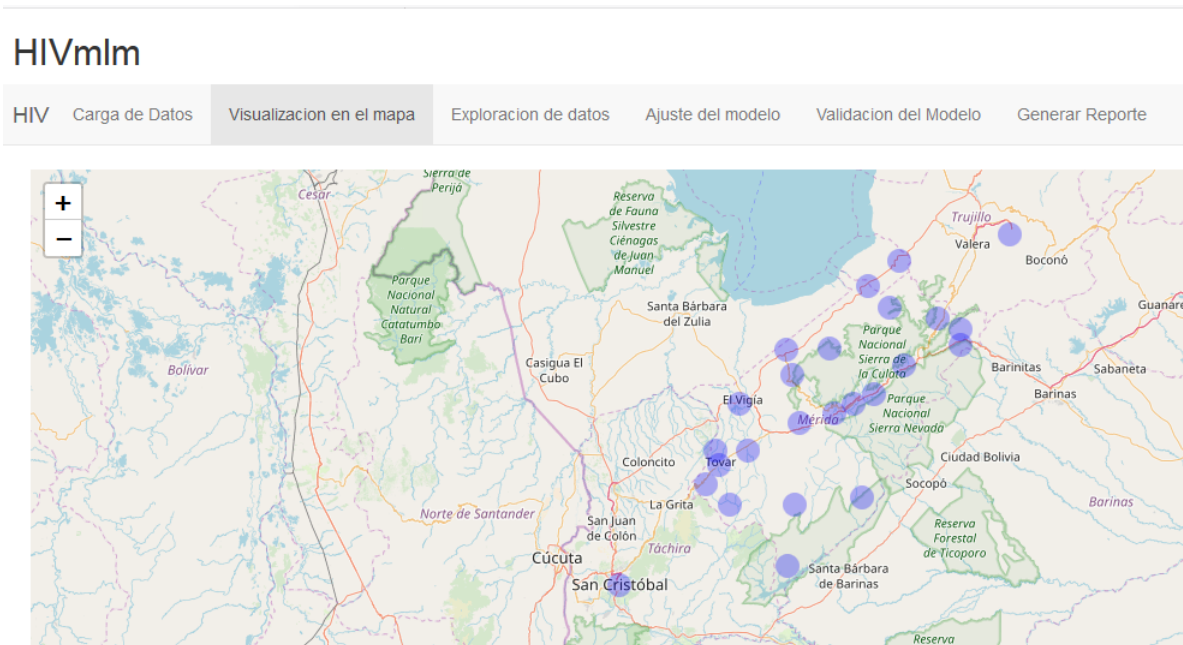


Figura B.6: Pantalla de la visualización del mapa

Al presionar o darle click a los puntos azules ubicados en el mapa, se puede visualizar la información realtiva a la media de la carga viral plasmatica y las células CD4, número de pacientes, género, desviación estandar y tiempo medio de duración.



Figura B.7: Selección de un punto azul sobre el mapa

Módulo de Exploración de datos

En el siguiente módulo, esta dividido en 3 paneles, es decir, el genero, la edad y las cargas.

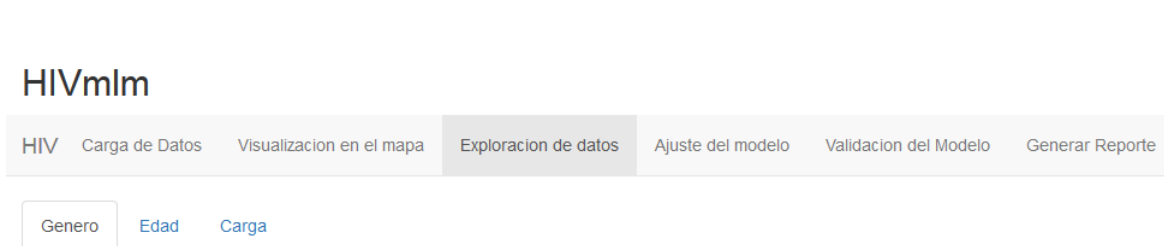


Figura B.8: Visualización del módulo de Exploración de datos

En la siguiente figura, se puede visualizar que en la sección **género** hay un selector por período, en este caso, se puede seleccionar entre 20071 hasta 20132, obteniendo el siguiente gráfico.

The screenshot shows a dropdown menu titled 'Seleccione El periodo'. The menu is open, showing a list of years from 20071 to 20132. The year 20071 is currently selected.

Figura B.9: Pantalla de los 3 paneles para visualizar graficamente los datos

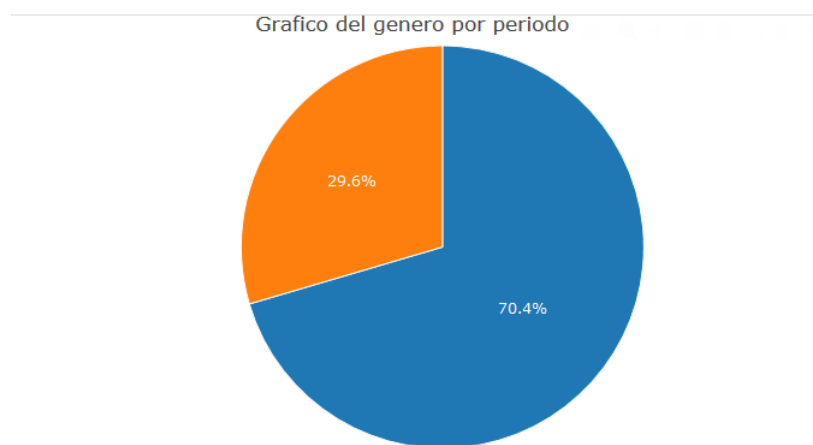


Figura B.10: Visualización de la gráfica de torta por género

En la sección **Edad** muestra un histograma como lo muestra la siguiente figura B.11

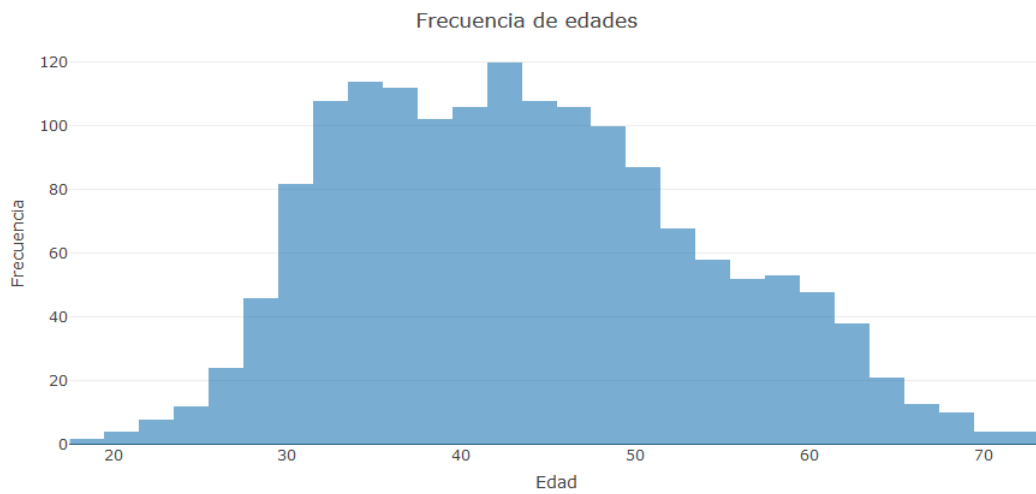


Figura B.11: Pantalla con el histograma de edades

En la sección **Cargas** muestra un selector, con las variables Carga Viral plasmatica, células T^+CD4 y T^+CD8 . Permite visualizar un gráfico de puntos.

Figure B.12 shows a screenshot of the "Carga" section selector. The interface includes three tabs: "Genero", "Edad", and "Carga". The "Carga" tab is selected. Below the tabs is a dropdown menu titled "Seleccione las Variables" with the following options: "Carga viral plasmatica", "Células CD4", and "Células CD8". The "Carga viral plasmatica" option is currently selected.

Figura B.12: Pantalla del selector de la sección cargas

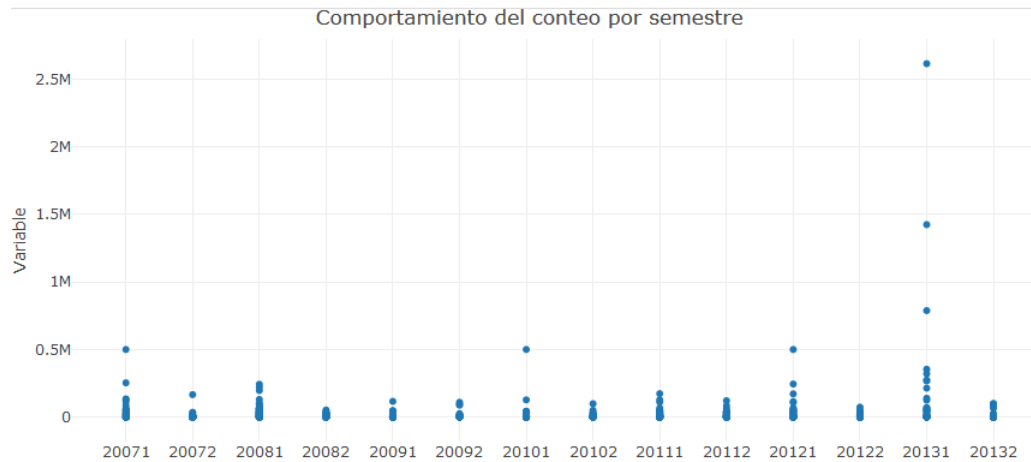


Figura B.13: Visualización del gráfico de puntos

Módulo Ajuste del modelo

En este módulo, muestra la implementación del modelo lineal mixto.

HIVmlm

HIV
Carga de Datos
Visualización en el mapa
Exploración de datos
Ajuste del modelo
Validación del Modelo
Generar Reporte

Modelo de Intercepto Aleatorio con Media Fija

Seleccione las Variables

Carga viral plasmática

☒ Restringida

Ajuste de Modelos por Maxima Verosimilitud

Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']

Formula: LogCVP ~ 1 + Fecha + variable + (1 | Pac)

Data: dato()

REML criterion at convergence: 3561.4

Scaled residuals:

| Min | 1Q | Median | 3Q | Max |
|---------|---------|---------|--------|--------|
| -6.8054 | -0.7035 | -0.2341 | 0.7065 | 2.6350 |

Figura B.14: Pantalla con el módulo ajuste del modelo

Contiene un selector que permite seleccionar entre las variables carga viral plasmática, células T^+CD4 y T^+CD8 y un cuadro de selección, con la variable Restringida, permitiendo variar el modelo matemático.

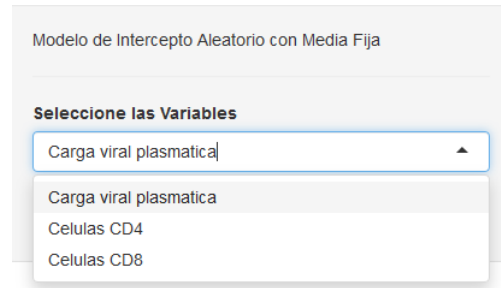


Figura B.15: Pantalla con el selector del módulo de ajuste del modelo

Luego de seleccionar las variables, sale la siguiente información como lo muestra la figura B.16.

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
Formula: LogCVP ~ 1 + Fecha + tre$CVP + (1 | Pac)
Data: tre

REML criterion at convergence: 3561.4

scaled residuals:
      Min       1Q   Median       3Q      Max
-6.8054 -0.7035 -0.2341  0.7065  2.6350

Random effects:
Groups   Name             Variance Std.Dev.
Pac      (Intercept) 0.2568    0.5067
Residual                0.8582    0.9264
Number of obs: 1251, groups: Pac, 115

Fixed effects:
              Estimate Std. Error      df t value Pr(>|t|)
(Intercept)  2.997e+00  7.200e-02  3.306e+02  41.626 < 2e-16 ***
Fecha       -5.098e-02  6.776e-03  1.158e+03  -7.524 1.06e-13 ***
tre$CVP      3.727e-06  2.931e-07  1.208e+03  12.715 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
      (Intr) Fecha
Fecha  -0.654
tre$CVP -0.024 -0.058
```

Figura B.16: Pantalla con la información del modelo lineal mixto

Módulo de validación del modelo

En este módulo, muestra el resultado de los residuos, calculados en el módulo anterior, el ajuste del modelo, lo que permite visualizar el comportamiento de los datos en el modelo lineal mixto. Permite seleccionar entre las variables Carga viral plasmática, células T^+CD4 y T^+CD8 .

Luego aparece una ventana, donde permite guardar el reporte, se le da click en ok, y lo guarda.

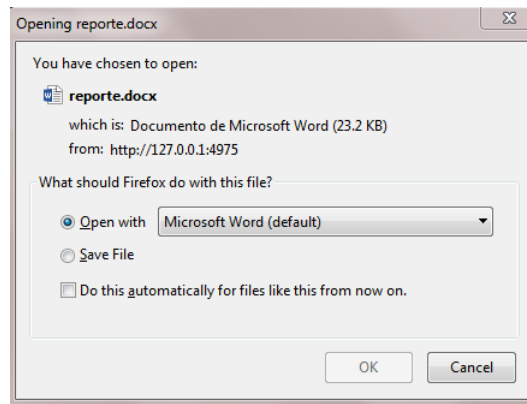


Figura B.20: Pantalla de guardar el reporte

Bibliografía

- F. Arias. *El proyecto de investigación. Introducción a la metodología científica*. Episteme, 2012.
- G. Belle, Ll. Fisher, P. Heagerty, y T. Lumley. *Biostatistics. A methodology for the health sciences*. USA, 2004.
- Barry Boehm. A spiral model of software development and enhancement. Inf. téc., TRW Defense Systems Group, 1988.
- D. Cuevas y R. Rebolledo. *Diagnóstico sobre VIH/SIDA en Baja California Sur*, 2009.
- Republica Bolivariana de Venezuela. Informe nacional de avances e la implementación de la declaración de compromisos sobre vih/sida (2001) y la declaración política sobre el vih/sida (2011). Inf. téc., Programa Nacional SIDA/ITS del Ministerio del Poder popular para la Salud, 2014.
- M. Delgado y J. Llorca. Estudios longitudinales: concepto y particularidades. *Revista Española de salud Pública*, 2004.
- E. Grunsky. *R: a data analysis and statistical programming environment – an emerging tool for the geosciences*. Computers Geosciences, 2002.
- J. Hurtado. *Metodología de la investigación holística*. Fundación SYPAL, 2000.
- T. Konig, J. Klaus, y M. Burger. Runit - a unit test framework for r. Inf. téc., CRAN, 2018.
- Alba León y Fabián Jaimes. Ronda clínica y epidemiológica. análisis de datos longitudinales. *Latreia*, 2015.
- Xiang Lu. *Handling incomplete high-dimesional multivariate longitudinal data with mixed data types by multiple imputation using a longitudinal factor analysis model*. Tesis Doctoral, University of California, Los Angeles, 2016.

- V. Morles. *Planeamiento y analisis de investigaciones*. El Dorado, 1994.
- MPPS. *Guía de tratamiento antiretroviral de las personas que viven con el VIH/SIDA en Venezuela*, 2014.
- M. Pilar Muñoz, Angels Orcau, y Joan Caylà. Tuberculosis en barcelona: Modelo predictivo basado en series temporales. *Revista Española de Salud Pública*, 2009.
- José Felix Oleta, Oswaldo Godoy, y Ana Carvajal. Alerta epidemiológica n-241, algunos aspectos epidemilógicos del vih/sida en venezuela. *Red de sociedades científicas médicas venezolanas*, 2012.
- OMS. Definición de la oms de caso de infección por el vih a efectos de vigilancia y revisión de la estadificación clínica y de la clasificación inmunológica de la enfermedad relacionada con el vih en adultos y niños. Inf. téc., Organizacion Mundial de la Salud, 2009.
- S. Pallela y F. Martins. *Metodología de la investigación cuantitativa*. Fondo Editorial de la Universidad Experimental Libertador, 2012.
- D. Ramduth. *Characterization od CD4+ and CD8+ T cell responses in HIV-1 C-clade infection*. Tesis Doctoral, Nelson Mandela school od Medicine, University of KwaZulu-Natal, 2011.
- D. Rizopoulos. Join modeling of longitudinal and survival data. Inf. téc., CRAN, 2018.
- Guillermo Rúa, Carolina Suárez-Acosta, José Chauca, Palmira Ventosilla, y Rita Almanza. Modelado del efecto de la variabilidad climática local sobre la transmisión de dengue en medellín (colombia) mediante análisis de series temporales. *Biomédica*, 2013.
- D. Sharrow. Hiv calibrated model life tables for countries with generalized hiv epidemics. Inf. téc., CRAN, 2013.
- Linh Mai Tran. *Comparative causal effect estimation and robust variance for longitudinal data structures with applications to observational HIV treatment*. Tesis Doctoral, University of California, Berkeley, 2016.
- Catherine Welch. *Implementation, evaluation and application of multiple imputation for missing data in longitudinal electronic healt record research*. Tesis Doctoral, University of London, 2015.

H. Wickham. Unit testing for r. Inf. téc., CRAN, 2017.

Y. Xie. Format r code automatically. Inf. téc., CRAN, 2017.

Bo Zhang. *Linear mixed models*. Thesis Doctoral, University of Minnesota, 2009.