

**SOFTWARE PARA LA IDENTIFICACIÓN DE BIOMARCADORES EN MEDIDAS  
OBTENIDAS MEDIANTE CROMATOGRAFÍA DE GASES (GCMS)**

**DANIELA PATRICIA MÁRQUEZ SANDOVAL  
DANIELA LUCIA MOVILLA ECHEVERRI**

**UNIVERSIDAD AUTÓNOMA DEL CARIBE  
FACULTAD DE INGENIERÍA  
PROGRAMA DE INGENIERÍA MECATRÓNICA  
BARRANQUILLA - COLOMBIA  
2020**

**SOFTWARE PARA LA IDENTIFICACIÓN DE BIOMARCADORES EN MEDIDAS  
OBTENIDAS MEDIANTE CROMATOGRFÍA DE GASES (GCMS)**

**DANIELA PATRICIA MÁRQUEZ SANDOVAL  
DANIELA LUCIA MOVILLA ECHEVERRI**

**Trabajo de grado presentado para optar al título de  
Ingeniero Mecatrónico**

**ASESORES DISCIPLINARES:  
ING. KELVIN BELEÑO SÁENZ, MSc.  
ING. CARLOS DÍAZ SÁENZ, MSc.**

**UNIVERSIDAD AUTÓNOMA DEL CARIBE  
FACULTAD DE INGENIERÍA  
PROGRAMA DE INGENIERÍA MECATRÓNICA  
BARRANQUILLA - COLOMBIA**

**2020**

**Nota de aceptación**

---

---

---

---

---

---

---

Firma del jurado 1

---

Firma del jurado 2

## **DEDICATORIA**

El presente trabajo de grado es dedicado a nuestra familia y nuestra Universidad Autónoma del Caribe, junto a todos los profesores que nos han forjado para llegar a este momento, que, aunque no ha sido sencillo, su motivación y apoyo nos ayudó a seguir adelante con el proyecto. Muchísimas gracias por acompañarnos en este proceso y ayudarnos a cumplir nuestras metas y objetivos tanto como estudiante y persona.

## TABLA DE CONTENIDO

LISTA DE FIGURAS	7
LISTA DE TABLAS	8
LISTA DE ECUACIONES	9
GLOSARIO	10
RESUMEN	11
ABSTRACT	12
1. INTRODUCCIÓN	13
2. PLANTEAMIENTO DEL PROBLEMA	144
<u>2.1.</u> ANTECEDENTES	144
2.2. FORMULACIÓN DEL PROBLEMA	188
2.3. JUSTIFICACIÓN Y ALCANCE	188
3. OBJETIVOS	20
3.1. OBJETIVO GENERAL	20
3.2. OBJETIVOS ESPECÍFICOS	20
4. MARCO DE REFERENCIA	211
4.1. ESTADO DEL ARTE	211
<u>4.2.</u> MARCO TEÓRICO	266
4.2.1. TUBERCULOSIS BOVINA	266
4.2.2. CROMATOGRAFÍA DE GASES	288
4.2.3. ESPECTROMETRÍA DE MASAS	299
4.2.4. CROMATOGRAFÍA DE GASES EN CONJUNTO CON LA ESPECTROMETRÍA DE MASAS	29
4.2.5. ANÁLISIS ESTADÍSTICO	322
4.2.5.1. TEST DE NORMALIDAD	333
4.2.5.1.1. KOLMOGOROV-SMIRNOV	344
4.2.5.1.2. SHAPIRO- WILK	344
4.2.6. DIFERENCIA ENTRE PRUEBAS PARAMÉTRICAS Y NO PARAMÉTRICA	355
4.2.7. WILCOXON	355
4.2.8. U MANN-WHITNEY	366
5. MARCO LEGAL	377
6. PROCEDIMIENTO METODOLÓGICO	399
6.1. METODOLOGÍA	399
6.2. TIPO DE ESTUDIO	39
6.3. POBLACIÓN Y MUESTRA	39

6.4.	FASES DEL DISEÑO	40
6.4.1.	FASE UNO	40
6.4.2.	FASE DOS	41
6.4.3.	FASE TRES	46
6.5.	CRONOGRAMA - PLAN DE TRABAJO	48
7.	PRESUPUESTO	50
7.1.	PRESUPUESTO GENERAL	50
7.2.	PERSONAL CIENTÍFICO Y DE APOYO	51
7.3.	MATERIALES, INSUMOS Y EQUIPOS	52
8.	PRESENTACIÓN Y ANÁLISIS DE RESULTADOS	54
8.1.	DISEÑO DEL PROTOTIPO	54
8.2.	ANÁLISIS DE RESULTADOS	55
	CONCLUSIONES Y RECOMENDACIONES	78
	BIBLIOGRAFÍA	81
	ANEXOS	90

## LISTA DE FIGURAS

Figura 1. Diagrama de proceso de Cromatografía de gases (GCMS).	30
Figura 2. Cromatograma de datos obtenidos (propia autoría).	31
Figura 3. Diagrama de flujo de test de normalidad (propia autoría).	42
Figura 4. Diagrama de flujo de pruebas no paramétricas (comparación de dos muestras) (propia autoría).	44
Figura 5. Interfaz del Software Prueba (Guide 1 en MATLAB) (propia autoría).	45
Figura 6. Interfaz del Software Prueba (Guide 2 en MATLAB) (propia autoría).	46
Figura 7. Resultados obtenidos y exportados a Excel.	56
Figura 8. Boxplot Tiempo de Retención vs p-value, Wilcoxon Control II y Control I.	58
Figura 9. Boxplot Tiempo de Retención vs p-value, Wilcoxon BCG II y BCG I.	59
Figura 10. Boxplot Tiempo de Retención vs p-value, Wilcoxon Michigan II y Michigan I.	60
Figura 11. Boxplot Tiempo de Retención vs p-value, Wilcoxon BCG III y BCG IV.	62
Figura 12. Resultados obtenidos por Prueba no paramétrica U Mann-Whitney exportados a Excel.	63
Figura 13. Boxplot Tiempo de Retención vs p-value, U Mann-Whitney Control II y BCG II.	64
Figura 14. Boxplot Tiempo de Retención vs p-value, U Mann-Whitney Control II y Michigan II.	65
Figura 15. Boxplot Tiempo de Retención vs p-value, U Mann-Whitney Control I y BCG I.	67
Figura 16. Boxplot Tiempo de Retención vs p-value, U Mann-Whitney Control I y Michigan I.	68
Figura 17. Boxplot Tiempo de Retención vs p-value, U Mann-Whitney Control I y Spanish I.	69
Figura 18. Boxplot Tiempo de Retención vs p-value, U Mann-Whitney Control III y BCG III.	70
Figura 19. Boxplot Tiempo de Retención vs p-value, U Mann-Whitney Control IV y BCG IV.	71
Figura 20. Boxplot Tiempo de Retención vs p-value, U Mann-Whitney Control IV y Michigan IV.	72
Figura 21. Boxplot Tiempo de Retención vs p-value, U Mann-Whitney Control IV y Spanish IV.	74

## LISTA DE TABLAS

Tabla 1. Prueba estadística de acuerdo con los objetivos de estudio y la escala medición de las variables.	32
Tabla 2. Presupuesto general.	50
Tabla 3. Costo personal científico.	51
Tabla 4. Costo personal de apoyo.	51
Tabla 5. Costo materiales e insumos.	52
Tabla 6. Costo equipos usados	53
Tabla 7. Comparación de número de datos antes y después de aplicar la prueba de Wilcoxon.	57
Tabla 8. Características de boxplot de la Figura 8 según p-value.	58
Tabla 9. Características de boxplot de la Figura 9 según p-value.	59
Tabla 10. Características de boxplot de la Figura 10 según p-value.	61
Tabla 11. Características de boxplot de la Figura 11 según p-value.	62
Tabla 12. Comparación de número de datos antes y después de aplicar la prueba de U Mann-Whitney.	64
Tabla 13. Características de boxplot de la Figura 13 según p-value.	65
Tabla 14. Características de boxplot de la Figura 14 según p-value.	66
Tabla 15. Características de boxplot de la Figura 15 según p-value.	67
Tabla 16. Características de boxplot de la Figura 16 según p-value.	68
Tabla 17. Características de boxplot de la Figura 17 según p-value.	69
Tabla 18. Características de boxplot de la Figura 18 según p-value.	70
Tabla 19. Características de boxplot de la Figura 19 según p-value.	72
Tabla 20. Características de boxplot de la Figura 20 según p-value.	73
Tabla 21. Características de boxplot de la Figura 21 según p-value.	74
Tabla 22. Resultados obtenidos para la vacuna Control en grupo BREATH.	76
Tabla 23. Resultados obtenidos para la vacuna Michigan en grupo BREATH.	76
Tabla 24. Resultados obtenidos para la vacuna BCG en grupo BREATH.	77
Tabla 25. Resultados obtenidos para la vacuna BCG en grupo FECES.	77

## LISTA DE ECUACIONES

Ecuación 1. Estadístico de prueba Kolmogórov-Smirnov	34
Ecuación 2. Estadístico de prueba Shapiro-Wilk	34
Ecuación 3. Estadístico de prueba Shapiro-Wilk	35
Ecuación 4. Distribución normal Estándar Wilcoxon	35
Ecuación 5. Media estadística Wilcoxon	35
Ecuación 6. Desviación estándar Wilcoxon	36
Ecuación 7. Valor estadístico 1 U Mann-Whitney	36
Ecuación 8. Valor estadístico 2 U Mann-Whitney	36
Ecuación 9. Estadístico de prueba U Mann-Whitney	36

## GLOSARIO

**Cromatografía de gases:** Técnica de separación de componentes volátiles, cada componente tiene un tiempo de paso característico a través del sistema, llamado tiempo de retención [41]

**Prueba de hipótesis:** Evalúa si una determinada premisa es realmente cierta para su conjunto de datos o población [43].

**Media aritmética:** Valor promedio de las muestras, es independiente de las amplitudes de los datos. Solo es utilizado en variables cuantitativas [44].

**Distribución normal:** distribución con forma de campana donde las desviaciones estándar sucesivas con respecto a la media establecen valores de referencia para estimar el porcentaje de observaciones de los datos [46].

**Pruebas Paramétricas:** Es una prueba de hipótesis que no requiere que la distribución de la población sea caracterizada por ciertos parámetros. Hace referencia a una parte de la inferencia estadística que utiliza estadísticos y criterios de resolución fundamentados en distribuciones conocidas [47, 53].

**Pruebas no paramétricas:** Son una herramienta estadística que se utiliza para el análisis de los factores de la población. Esta muestra debe cumplir ciertos requisitos como el tamaño. Se trata de una rama de la inferencia estadística cuyos cálculos y procedimientos están fundamentados en distribuciones desconocidas [48, 53].

**Varianza:** Medida de dispersión que representa la variabilidad de una serie de datos respecto a su media. Se calcula como la suma de los residuos al cuadrado divididos entre el total de observaciones. También se puede calcular como la desviación estándar al cuadrado [49].

**Variable dependiente nominal:** Variables cualitativa cuyas categorías no siguen ningún orden [50]

**Desviación estándar:** Indica qué tan dispersos están los datos con respecto a la media [45].

## RESUMEN

El presente documento obedece al cumplimiento de un software identificador de biomarcadores que consta de un procedimiento diverso en donde se aplican aspectos variados como la cromatografía de gases el cual es utilizado para la obtención de muestras de datos, biomarcadores que serán hallados para determinar si se es portador de una enfermedad o no, entre otros de los dichos aspectos se encuentran las pruebas estadísticas que serán la base del funcionamiento que cumplirá el software, estas pruebas proporciona un mecanismo para tomar decisiones cuantitativas sobre un proceso o procesos con la intención es determinar si hay suficiente evidencia para "rechazar" una conjetura o hipótesis sobre el proceso. Una vez que el software finalice el cálculo correspondiente en tiempo récord total de 8.96 segundos por grupo seleccionado o ingresado utilizando la prueba no paramétrica, en cada hoja se encontrarán los grupos de probabilidad de rechazo a la hipótesis nula menores o iguales a 0.05.

**Palabras claves:** Biomarcadores, Cromatografía de Gases, Pruebas Estadísticas, Software.

## ABSTRACT

This document complies with a biomarker identification software that consists of a diverse procedure where various aspects are applied, such as gas chromatography which is used to obtain data samples, biomarkers that will be found to determine the carrying of a disease or not, among other aspects such as statistical tests that will be the basis of the operation that the software will perform, these tests provide a mechanism to make quantitative decisions about a process or processes with the intention of determining if there is sufficient evidence to "reject" a conjecture or hypothesis about the process. Once the software completes the corresponding calculation in a total record time of 8.96 seconds per selected or entered group using the non-parametric test, each sheet will find the groups of probability of rejection of the null hypothesis less than or equal to 0.05.

**Keyword:** Biomarkers, Gas Chromatography, Statistical Tests, Software.

## 1. INTRODUCCIÓN

Con el fin de tomar decisiones más seguras el ser humano utiliza un estudio denominado estadística, el cual es encargado de recolectar, organizar y analizar datos, con el objetivo de establecer conclusiones, decisiones o acciones respecto a la probabilidad de sucesos de los datos presentados, este estudio puede encontrarse en la rama de la salud para ver los patrones de enfermedades en grupos de pacientes, en donde los datos son adquiridos por medio de pruebas invasivas y/o no invasivas. Las pruebas invasivas han tenido un nivel de confiabilidad alto por brindar respuestas acertadas de acuerdo al estudio que se esté realizando dentro del campo médico para la detección de enfermedades, encontrándose como una de las pruebas más seguras si lo que se buscaba era tener la certeza de que un fenómeno estaba ocurriendo en una población estudiada.

Sin embargo, estas pruebas poseen contraindicaciones dentro de las cuales puede verse afectada la muestra, en este sentido, fueron desarrollándose técnicas y/o métodos no invasivos para la extracción de muestras y que representen el mismo nivel de confianza que las pruebas invasivas basando su estudio en el análisis estadístico de la misma.

Una vez identificada cierta enfermedad por las pruebas realizadas se debe proceder a la lectura de datos y tratamiento de estos, de tal forma que, estos puedan ser analizados según el comportamiento de las variables que afecten u originen dicha enfermedad. Por lo tanto, es necesario un estudio riguroso mediante la hipótesis según la selección de la prueba estadística complementaria para el análisis de datos.

## **2. PLANTEAMIENTO DEL PROBLEMA**

Actualmente para la detección de la Tuberculosis Bovina (TB) existen métodos invasivos que pueden ser definidos como aquellos procedimientos en donde es necesario la inserción de algún objeto extraño al organismo que será estudiado y diagnosticado posteriormente. Así mismo, es un método muy preciso, pero con mayores riesgos de uso, ya que una mal práctica podría ocasionar daños en el organismo vivo. La enfermedad crónica puede ser detectada por medio de la prueba de Tuberculina, la cual consiste en una inyección intradérmica de una cantidad considerable de antígeno, con esto, se mide la reacción inmunitaria posterior a la inyección y se requieren alrededor de ocho (8) semanas para su proceso de diagnóstico.

Además, esta enfermedad se transmite vía inhalación que produce infección de las gotículas (o gotas óticas) infectadas al momento de que el animal infectado expulsa al toser.

### **2.1. ANTECEDENTES**

La tuberculosis bovina es una de las enfermedades más infecciosas en el ganado. La presencia de esta entidad plantea un impedimento en el desarrollo de la ganadería ya que amenaza la comercialización del ganado y sus subproductos entre regiones de un mismo país, así como entre países. Esta enfermedad es el origen de pérdidas masivas en el área agropecuaria, y constituye también una amenaza para la salud de la población humana. Es una enfermedad de incidencia variable dentro de Latinoamérica, de acuerdo con la reactividad de bovinos a la prueba de hipersensibilidad de tipo retardado a la tuberculina o derivado proteico purificado (PPD), existen regiones con muy baja o nula prevalencia de TB bovina, tales como algunas islas del Caribe, países de prevalencia intermedia, como México y varios integrantes de Centroamérica, y países con una prevalencia alta, como Argentina o Brasil. Gil, A. (2012). Tuberculosis Bovina: Enfermedad Reemergente en poblaciones bovinas de América. Presentation [1]. Desafortunadamente, los países productores del 70% de ganado en toda la región son los de mayor prevalencia de TB bovina. Las pérdidas económicas debidas a la TB son,

por lo tanto, sumamente elevadas, y la zoonosis representa un foco de infección para humanos que no puede ser despreciado López Marín, L., Díaz Otero, F., Vallecillo Maza, A., Esquivel Solís, H., & Gutiérrez Pabello, J. (2006). Tuberculosis humana y bovina en Latinoamérica: De estudios sobre virulencia hacia herramientas para su control [Ebook] (48th ed., p. 1). Los factores más importantes informados para este aumento son: infecciones con el virus del VIH, aparición de TB en personas nacidas en el extranjero, aparición de cepas resistentes a los medicamentos, el aumento de individuos susceptibles a la infección *M. Bovis* debido al SIDA y la transmisión de Tuberculosis que puede ser causada tanto para humanos como por bovinos por la bacteria *M. tuberculosis*, *M. bovis* [2].

No existen mayores diferencias en la patogenia de la infección por *M. bovis* entre humanos y bovinos. Una vez que infecta al hospedero, la micobacteria genera una pequeña lesión granulomatosa en el sitio de ingreso y en el linfonódulo regional, lo que en conjunto se conoce como complejo primario de la infección. Debido a que en el ser humano la tuberculosis provocada por *M. bovis* es clínica y radiológicamente indistinguible de la que causa *M. tuberculosis*, el tratamiento es similar en la mayoría de los casos. Sin embargo, como *M. bovis* es naturalmente resistente a la droga pirazinamida, ésta debe ser omitida. Abalos, P., & Retamal, P. (2004). Tuberculosis: ¿una zoonosis reemergente? *Revue Scientifique et Technique-Office International des Epizooties*, 23(2), 583-594 [3].

Es necesario indicar que los importantes avances en el conocimiento de la respuesta del huésped y sus diversos componentes han permitido el surgimiento del diagnóstico por compuestos orgánicos volátiles. “Este tipo de diagnóstico identifica animales infectados y vacunados, a través de los perfiles de compuestos específicos (aldehídos, alcanos, cetonas, alcoholes, y derivados del benceno)” Peled, N., Ionescu, R., Nol, P., Barash, O., McCollum, M., & VerCauteren, K. et al. (2012). Detection of volatile organic compounds in cattle naturally infected with *Mycobacterium bovis*. Lo no invasivo de su práctica, fácil muestreo (heces y aliento exhalado), potencialidad DIVA y posibilidad de uso en animales silvestres son las principales fortalezas que lo caracterizan [4].

El estudio realizado por MILIÁN, F., SÁNCHEZ, L., TOLEDO, P., RAMÍREZ, C., & SANTILLÁN, M. (2000). Descriptive Study of Human and Bovine Tuberculosis in Querétaro, México [Ebook] (pp. 1-2). Querétaro: Revista Latinoamericana de Microbiología. Establece los siguientes datos relacionados a las muertes por tuberculosis” En 1995, la TB mató a 3 millones de personas, de las cuales el 6% tenía menos de 15 años.7,15 En los países desarrollados, la incidencia de TB disminuyó hasta hace poco; sin embargo, desde 1985 hasta 1993 en los Estados Unidos, la morbilidad de la tuberculosis aumentó un 14%” [5]. A su vez, de acuerdo con cifras de la Organización Mundial de la Salud (OMS),4 en 1990 existían 8 millones de casos activos en el mundo, de los cuales 7.6 millones se hallaban en países en desarrollo y sólo 400 mil en países industrializados.2 En ese mismo año, las muertes por tuberculosis representaron casi 2.9 millones de personas, la mayoría de ellas radicadas en países del tercer mundo. Pérez-Guerrero, L., Milián-Suazo, F., Arriaga-Díaz, C., Romero-Torres, C., & Escartín-Chávez, M. (2008). Epidemiología molecular de las tuberculosis bovina y humana en una zona endémica de Querétaro, México [6].

El siguiente artículo Retamal, P. (2000). Tuberculosis bovina: Una breve actualización., describe algunos de los diagnósticos utilizados para la detección de la enfermedad, como lo es la prueba de hipersensibilidad retardada y el ensayo de reacción en Cadena de la Polimerasa, el artículo a su vez describe la enfermedad como “ Una enfermedad infectocontagiosa que afecta al ganado bo-vino produciendo un cuadro crónico, que genera pérdidas económicas por muerte de los animales, decomisos a nivel de matade-ros, menor productividad y valoración de la leche” [7]. Desde su introducción en 1890, la prueba cutánea de tuberculina (TST) se ha adoptado ampliamente para detectar la infección con el complejo Mycobacterium tuberculosis (MTBC), tanto para ayudar en el diagnóstico de la enfermedad de tuberculosis como para identificar a las personas con infección latente de tuberculosis que podrían beneficiarse del tratamiento. para prevenir la progresión a la enfermedad de TB. Un resultado TST positivo se basa en la induración de la piel medible en respuesta a los antígenos derivados de proteínas purificadas inyectadas por vía intradérmica y se basa en las respuestas inmunes de las células de

memoria a los antígenos micobacterianos. Sin embargo, aproximadamente el 10-25% de los pacientes con tuberculosis activa microbiológicamente confirmada no responden a la tuberculina y tienen un resultado negativo de TST. Miramontes, R., Auld, S., Click, E., Heilig, C., Cain, K., Bisson, G., & Mac. Kenzie, W. (2013). Tuberculin Skin Test Result and Risk of Death among Persons with Active TB. Jyothi Rengarajan [8].

Youssef, E., & Woollorton, E. (2005). Serious allergic reactions following tuberculin skin tests. 5 unidades de tuberculina del derivado proteico purificado (PPD) de Mycobacterium tuberculosis por vía intradérmica, a menudo en la piel volar del antebrazo. Una reacción de hipersensibilidad retardada generalmente comienza dentro de las 24 horas, alcanzando su pico entre las 48 y 72 horas, que es cuando se lee el resultado de la prueba. Característicamente, un área circular de eritema ocurre alrededor de un área más central de induración. En 1% –2% de los resultados positivos de la prueba, pueden producirse ampollas o incluso necrosis local, pero esto generalmente es autolimitado. Las reacciones locales como la linfangitis regional y la adenitis pueden También ocurren en raras ocasiones [9]. Actualmente, las técnicas de amplificación de ADN se han convertido en importantes herramientas de detección. Sin embargo, la extrema sensibilidad de tales técnicas puede resultar fácilmente en contaminación. Contaminación cruzada entre muestras; Contaminación cruzada entre ácidos nucleicos o Contaminación por arrastre de producto de PCR. Hu, Y. (2016). Regulatory Concern of Polymerase Chain Reaction (PCR) Carryover Contamination (pp. 2-4) [10].

Los presentes artículos científicos ayudaron a determinar la gravedad de la enfermedad que ha sido estudiada durante muchos años por científicos y doctores de todo el mundo, los cuales van buscando nuevas formas para poder detectar la tuberculosis bovina de forma más rápida y con menos falsos positivos y falsos negativos. Esta enfermedad no sólo afecta a los bovinos sino también a los humanos quienes consumen productos derivados de los bovinos que padecen tuberculosis bovina, es por ello que nos esforzamos en la creación de este software identificador de biomarcadores mediante cromatografía de gases con el fin de agilizar el proceso de análisis de resultados en

muestras tomadas por pruebas de aliento, con la finalidad de reemplazar las pruebas invasivas por las pruebas no invasivas.

## **2.2. FORMULACIÓN DEL PROBLEMA**

¿Cómo agilizar el procesamiento de tamaños muestrales grandes para el reconocimiento de biomarcadores potenciales mediante la caracterización de mezclas complejas de gases obtenidas a partir de cromatografía de gases?

## **2.3. JUSTIFICACIÓN Y ALCANCE**

Por medio de este trabajo se buscó una alternativa no invasiva basada en pruebas de hipótesis mediante muestras de aliento y muestras fecales con la misma veracidad que se obtiene del método invasivo y en menos tiempo de ejecución. Por lo tanto, se propuso el método no invasivo como opción para evitar riesgos, y minimizar el tiempo de determinación de diagnósticos y obtención de resultados, con la finalidad de disminuir los casos de sacrificios en el sector ganadero, a su vez reducir el número de portadores de la enfermedad en todos los mamíferos, en los que se provoca en términos generales un deterioro del estado general de salud.

Esto se fundamentó en el “Software para la identificación de biomarcadores obtenidos mediante cromatografía de gases (GCMS)”, con el cual se buscó simplificar el procesamiento de datos obtenidos para el estudio de potenciales biomarcadores a partir de las pruebas estadísticas para muestras que presenten normalidad o no se ajusten a algún tipo de distribución con el propósito de comprobar la veracidad de las hipótesis (nula y alternativa) con las cuales se buscó la obtención de potenciales biomarcadores, permitiendo así caracterizar la presencia de compuestos relevantes existentes en un organismo vivo. Así mismo, agiliza el hallazgo de posibles compuestos orgánicos característicos de la enfermedad de Tuberculosis Bovina por medio de factores químicos presentes en el sistema.

Esto se desenvuelve por medio del estudio de las sustancias marcadas que se encuentran por medio de cromatografía de gases mediante el uso de un programa ya adquirido con ayuda del asesor Kelvin Beleño Sáenz e inicialmente financiado por un equipo de trabajo. A lo que se refiere a un método químico centrado en la separación para la caracterización de mezclas complejas.

Con esto en mente, se concluyó que para la identificación de biomarcadores cuya finalidad es personalizar el tratamiento de enfermedades, puede obtenerse de forma rápida y sencilla utilizando las pruebas estadísticas, las cuales pueden tener diferentes tamaños muestrales y, aun así, tener la probabilidad de estos datos de estudio junto a su validación y la confirmación independiente de cada uno de estos. Por lo que puede afirmarse que es importante reducir el tiempo entre la identificación y la validación de biomarcadores frente a criterios de valoración clínicos claros en diferentes poblaciones para atacar la enfermedad, es por esto que se decide realizar esta investigación de desarrollo tecnológico en el área de la salud.

### **3. OBJETIVOS**

#### **3.1. OBJETIVO GENERAL**

Desarrollar software para la identificación de biomarcadores en medidas obtenidas mediante cromatografía de gases (GCMS).

#### **3.2. OBJETIVOS ESPECÍFICOS**

1. Caracterizar las diferentes pruebas estadísticas para el análisis de datos en grupos de estudio con enfermedades y grupos controles que permita la identificación de potenciales biomarcadores.
2. Implementar software prototipo que permita el estudio del comportamiento de los pacientes mediante la lectura de datos para la detección de compuestos característicos de cada enfermedad.
3. Validar software para pruebas.

## 4. MARCO DE REFERENCIA

### 4.1. ESTADO DEL ARTE

La teoría empleada se deriva de conceptualizaciones basadas en el desarrollo de la enfermedad Tuberculosis Bovina, de los métodos estadísticos que se pueden realizar para el estudio de muestras de pacientes enfermos y que no poseen la enfermedad con el fin propuesto.

- **Nuevas estrategias en el diagnóstico y la epidemiología de las infecciones por Mycobacterium Bovis:** Describe como se ramifica la Mycobacterium Tuberculosis acorde a la especie, aunque de igual manera puede causar enfermedades en otros hospedadores. Debido a esto aumenta la posibilidad de contraer este patógeno por medios habituales de esta como por la propagación de aire, sino también al consumir algún producto derivado de uno de los animales portadores del patógeno [11].
- **Asociación entre el diagnóstico de Tuberculosis Bovina por intradermorreacción, la anatomopatología, la bacteriología y la posible interferencia con micobacterias ambientales:** El diagnóstico de la tuberculosis es realizado por medio de administración cutánea de tuberculina que es el diagnóstico más utilizado, sin embargo, genera resultados falsos positivos y falsos negativos como indica el estudio realizado en el siguiente artículo [12].
- **Validating a breath collection and analysis system for the new tuberculosis breath test. Journal of breath research:** Tiene por objetivo el análisis y validación de las pruebas de aliento respecto a la enfermedad tuberculina, citando lo siguiente: “Mycobacterium tuberculosis libera cuatro compuestos volátiles, metil fenil acetato, metil nicotinato, metil p-anisato y o-fenilanol, algunos de los cuales previamente se ha detectado en el aliento de pacientes con tuberculosis. Estos

volátiles tienen el potencial de ofrecer una prueba de aliento no invasiva y sensible para la detección de infección de tuberculosis” [13].

- **Clasificación de compuestos químicos usando un sistema multisensorial (nariz electrónica) desarrollada sobre un dispositivo hardware (fpga):** En los sistemas multisensoriales, como el cromatógrafo de gases o la nariz electrónica, el funcionamiento se basa en el uso de un computador para realizar el procesamiento y la clasificación de los datos provenientes de una matriz de sensores de gases químicos obtenidos por el cromatógrafo. El objetivo principal del presente estudio de la nariz electrónica es sustituir la computadora habitual mediante el uso de un dispositivo hardware, con el fin de obtener un sistema que sea más portable y que proporcione resultados en tiempo real a través de la adquisición, procesamiento y clasificación de las muestras [14].
- **Multisensor system for toxic gases detection generated on indoor environments. In IOP Conference Series: Materials Science and Engineering:** El sistema artificial multisensorial propuesto en este estudio se desarrolló a través de un conjunto de seis sensores de gases químicos (MQ) de bajo costo con sensibilidades superpuestas para detectar gases peligrosos en el aire. Se implementó un parámetro estadístico para el conjunto de datos y se utilizaron dos métodos de reconocimiento de patrones, como el Análisis de componentes principales (PCA) y el Análisis de función discriminante (DFA) para la selección de características. Las categorías de gases tóxicos se clasificaron en el trabajo con una red neuronal probabilística (PNN) para validar los resultados obtenidos. Las pruebas se llevaron a cabo en este estudio para verificar la viabilidad de la aplicación a través de un modelo de comunicación inalámbrica que permitió monitorear y almacenar la información de las señales del sensor para el análisis apropiado [15].
- **Las pruebas de aliento; un nuevo horizonte en el laboratorio clínico:** Con la denominación de pruebas de aliento se definen todos los métodos que a partir del

aire espirado pueden proporcionar información sobre la producción de gas intestinal, la absorción de grasas e hidratos de carbono a nivel intestinal, el funcionamiento de órganos como el hígado, el páncreas, el estómago y el intestino, la presencia de infecciones por *Helicobacter pylori* y la tuberculosis pulmonar, el cáncer de pulmón y la esquizofrenia, entre otros usos de las pruebas de aliento en la práctica médica [16].

- **Implementación de una nariz electrónica para detectar pacientes con EPOC desde el aliento exhalado:** En este artículo se presenta la implementación de una Nariz Electrónica (NE), la cual fue desarrollada a partir de una cámara de medida, compuesta por un conjunto de 6 sensores de gases de óxidos metálicos con sensibilidades parcialmente solapadas, con el objetivo de identificar y clasificar los volátiles emitidos a través del aliento exhalado obtenido de un grupo de personas sanas y pacientes con síntomas de la Enfermedad Pulmonar Obstructiva Crónica (EPOC). En este trabajo se implementaron un conjunto de algoritmos para el preprocesamiento y procesamiento de señales, con base en técnicas de extracción característica y métodos estadísticos, como el Análisis de Componentes Principales (ACP), para el reconocimiento de patrones del conjunto de datos. De las muestras de aliento exhalado de los pacientes se obtuvieron resultados importantes, debido a la clasificación de pacientes con la enfermedad, así como la de un grupo de voluntarios sanos [17].
- **Artificial olfactory system based on nanomaterials-based sensors for early detection of tuberculosis via breath test:** Indica un estudio de casos y controles. El aliento exhalado se muestreó y analizó utilizando una matriz de sensores químicamente modificados, basados en nanomateriales y detectores de olores [18].
- **Detection of precancerous gastric lesions and gastric cancer through exhaled breath:** La investigación recolectó 968 muestras de aliento de 484 pacientes (incluyendo 99 con cromatografía de gases (GC)) para dos análisis

diferentes. La primera muestra se analizó por cromatografía de gases vinculada a la espectrometría de masas (GCMS) mientras se aplicaba la prueba t con múltiples correcciones (valor de  $p < 0,017$ ); el segundo por nanoarrays de reacción cruzada combinados con reconocimiento de patrones. Para este último, el 70% de las muestras fueron seleccionadas al azar y utilizadas en el conjunto de entrenamiento, mientras que el 30% restante constituía el conjunto de validación. El enlace operativo en el sistema de estadificación de evaluación de metaplasia gástrica intestinal (OLGIM) se utilizó para estratificar la presencia sobre la ausencia y el nivel de riesgo de lesiones precancerosas. “Se consideró que los pacientes con OLGIM estadios III-IV tenían un alto riesgo. Según los resultados de GCMS, los pacientes con cáncer, así como aquellos en alto riesgo, tenían composiciones distintivas de impresión de aliento, de los cuales se detectaron ocho compuestos orgánicos volátiles significativos [19].

- **A novel breath analysis system based on electronic olfaction:** Se sabe que ciertos gases en el aliento son indicadores de la presencia de enfermedades y condiciones clínicas. Estos gases se han identificado como biomarcadores que utilizan equipos, como la cromatografía de gases y la nariz electrónica. GC es muy preciso, pero es costoso, consume mucho tiempo y no es portátil. E-nose tiene las ventajas de bajo costo y fácil operación, pero no es particular para analizar el olor del aliento y, por lo tanto, tiene una aplicación limitada en el diagnóstico de enfermedades. Este artículo propone un sistema novedoso que es especial para el análisis de la respiración. Fueron seleccionados sensores químicos que son sensibles a los biomarcadores y las composiciones en el aliento humano, desarrollaron el sistema e introdujeron el método de preprocesamiento y clasificación de la señal de olor. Para evaluar el rendimiento del sistema, capturaron muestras de aliento de personas sanas y pacientes que se sabe que padecen diabetes, enfermedad renal e inflamación de las vías respiratorias, respectivamente, y realizamos experimentos sobre evaluación de tratamiento médico e identificación de enfermedades. Los resultados mostrados en el artículo indican que el sistema no solo es capaz de distinguir entre muestras de aliento de

sujetos que padecen diversas enfermedades o afecciones (diabetes, enfermedad renal e inflamación de las vías respiratorias) y muestras de aliento de sujetos sanos, sino que también es útil en el caso de insuficiencia renal en la evaluación de la eficacia de la hemodiálisis [20].

- **Manual de las pruebas de diagnóstico y de las vacunas para los animales terrestres (mamíferos, aves y abejas):** En este artículo se trata la manera correcta de seguir un procedimiento para la recolección de muestras de una determinada población, se debe tener claro cuáles son los fines de estudio, como también el número de muestras pertinentes que permitan realizar estudios estadísticos y que estos proporcionen resultados validos ante la investigación. Este documento propone la identificación de variables para el análisis estadístico de los datos hallados permitiendo la identificación de los parámetros como lo son el tipo de distribución, el número muestral y la escala de medición de los datos anteriormente mencionados [57].

Una vez realizado el análisis del estudio del estado del arte, se determina que a partir de los aportes de los trabajos anteriormente mencionados se obtienen los aspectos más importantes como complemento de la investigación basados en características tales como la propagación de la enfermedad, las formas del diagnóstico tradicional para la enfermedad. Así mismo, ofrece un enfoque sobre el estudio realizado para la validación de las pruebas de aliento para la enfermedad y los resultados obtenidos por este método no invasivo. Además, se explica el proceso de cromatografía de gases (GCMS) para un estudio de detección de cáncer y lesiones precancerosas, con lo cual se detalla el número de pacientes y lo que lograron alcanzar con la investigación. Por último, teniendo en cuenta lo anterior se recopiló la información necesaria sobre cómo debe seguirse un procedimiento para estudios clínicos, lo cual engloba los estudios anteriores haciendo énfasis en la importancia de seguir un protocolo para la validación de los datos que serán estudiados mediante las pruebas estadísticas, determinando así la veracidad de los datos obtenidos por medio del proceso de cromatografía de gases.

## **4.2. MARCO TEÓRICO**

A continuación, se presentan los fundamentos teóricos que se deben tener en cuenta al momento de abordar el estudio. La teoría empleada se deriva de conceptualizaciones basadas en el desarrollo de la enfermedad Tuberculosis Bovina, de los métodos estadísticos que se pueden realizar para el estudio de muestras de pacientes enfermos y que no poseen la enfermedad con el fin propuesto.

### **4.2.1. TUBERCULOSIS BOVINA**

La tuberculosis proviene de los nódulos, formados en los ganglios linfáticos del animal afectado. Se trata de una enfermedad infecciosa producida por la presencia de bacterias que poseen características más complejas y poco comunes en comparación con las convencionales [21,23]. La Tuberculosis Bovina es una enfermedad que se presenta en los animales de forma crónica, y que, además, es provocada por una bacteria llamada *Mycobacterium bovis* (*M. bovis*), dicha bacteria está fuertemente relacionada con las bacterias que causan la tuberculosis en humanos. Por esta razón la bacteria que provoca la Tuberculosis Bovina es la segunda causa más común de tuberculosis en las personas. La *M. bovis* se puede encontrar con mayor frecuencia en ganado vacuno, como también en otros animales como búfalos, alces y venados. Esta enfermedad afecta en su mayoría a los mamíferos, provocando un deterioro en la salud y en muchos casos la muerte [21,22,27].

La transmisión de la enfermedad varía según las especies. La *M. bovis* puede transmitirse en animales sanos por la fuente más importante que son aquellos que están enfermos o con los síntomas presentes y que eliminan la bacteria al exterior por medio de secreciones y excreciones, contaminando objetos como suelo, agua, pasto, alimento. El modo de transmisión en el hombre ocurre por inhalación de aerosol, ingestión (al comer o beber productos no pasteurizados), o por heridas de la piel. Sabiendo que es una enfermedad contagiosa, puede propagarse por contacto entre animales [26,39]. En cuanto al diagnóstico e identificación, se han desarrollado pruebas para la detección de

respuesta inmune contra diversos antígenos microbacteriales (ESAT-6, CFP-10, MPB70, entre otros), para la detección de interferón bovino y para la detección directa del ácido desoxirribonucleico (ADN) bacteriano, que deben ser evaluadas para las condiciones epidemiológicas de las áreas de control [24, 25].

Un método tradicional es la prueba de tuberculina, se realiza por medio de una inyección intradérmica de una cantidad considerable de antígeno, esto para medir la reacción inmunitaria que posee tras colocar la inyección. Esta prueba se fundamenta en medir el grosor de la piel posterior a la aplicación. El diagnóstico definitivo para este método requiere de tiempo para el cultivo de bacterias en un laboratorio [21, 40].

Por otro lado, *Mycobacterium tuberculosis* libera cuatro compuestos volátiles, metil fenil acetato, metil nicotinato, metil p-anisato y o-fenilanol, algunos de los cuales previamente se ha detectado en el aliento de pacientes con tuberculosis. Estos volátiles tienen el potencial de ofrecer una prueba de aliento no invasiva y sensible para la detección de infección de tuberculosis.

Existen dos técnicas que pueden ser implementadas, según el tipo de estudio clínico y el objeto de investigación, las cuales se denominan pruebas invasivas y pruebas no invasivas. Las pruebas invasivas pueden ser definidas como toda intervención realizada con fines de investigación, la cual implica un riesgo físico o psíquico para el sujeto afectado. El riesgo que está presente en este tipo de pruebas viene determinado por diferentes factores que pueden estar directamente e indirectamente relacionados con el tipo de prueba que se esté realizando, como también podría depender de la fragilidad en la que se encuentre la población en estudio. Sin embargo, el término “invasivo” ya supone una práctica clínica que puede poseer riesgos adicionales [59]. Las pruebas no invasivas son aquellas que no tocan el cuerpo humano, es decir, son aquellos procedimientos que no involucran instrumentos que rompen la piel o que penetran físicamente el cuerpo [60].

En lo que respecta a la prevención de enfermedades, los biomarcadores representan un factor importante, ya que son usados como herramienta de información para evaluar aquellos riesgos o los factores que envuelven a estos relacionados a la exposición de

agentes ambientales. Los biomarcadores se utilizan para relacionar un agente de tipo químico, biológico o fisiológico con un sistema biológico, este tipo de relación se mide a través de una interacción de datos característicos. Este tipo de respuestas frente a los agentes puede darse a nivel celular o molecular, y está estrechamente relacionada con la probabilidad de que una enfermedad esté en desarrollo. Las consideraciones que se deben tener presentes para el estudio de un biomarcador son la dosis, la vía de contacto, la duración y la frecuencia que estuvo el agente expuesto [32].

La relación entre la exposición y la enfermedad determina un diagnóstico adecuado para la prevención, permitiendo una evaluación para los tratamientos e identificación de individuos infectados o propensos a la infección [33, 34].

Así mismo, dentro de los factores que se requieren para el proceso de validación y selección de un biomarcador se encuentran los análisis de datos con el fin de establecer una relación entre las consideraciones con respecto a la respuesta y la variación ocasionada a nivel individual, se debe evaluar la variación que existe para aquella población que no ha sido expuesta, como también la predicción del riesgo que ocasiona para la población [35]. De este modo, los biomarcadores representan un papel fundamental para realizar el estudio de aquellos compuestos que son característicos de la enfermedad.

La identificación del riesgo por medio del uso que se le otorga al biomarcador es vista como una síntesis que recoge información de tipo cualitativa y cuantitativa que describe cómo ha sido el efecto que ha tenido la salud antes de que el organismo fuera expuesto, esto se hace por medio de la evaluación de la relación entre la dosis y la respuesta del sistema biológico [36,37].

#### **4.2.2. CROMATOGRAFÍA DE GASES**

La cromatografía de gases es un método no invasivo y fiable, que permite la identificación y medición de estos en muestras de aire espirado. La muestra es inyectada en una fase móvil, la cual posee gas inerte. Durante la fase, los componentes pasan a través de una

fase estacionaria que se encuentra fija en columnas capilares (actualmente las más empleadas). La velocidad de cada componente, teniendo en cuenta su tiempo de retención, será función entre la fase móvil y la fase estacionaria.

La separación de los solutos dependerá de la afinidad que presenten con la fase estacionaria, es decir, aquellos componentes que sean fuertemente retenidos, su velocidad será lenta en la fase móvil, y los que sean levemente retenidos tendrán una velocidad rápida en esta fase. Esto quiere decir que de la diferencia que existe entre los movimientos de cada uno resulta la separación de estos y analizarse de modo cuantitativo o cualitativo por medio de los detectores seleccionados [28].

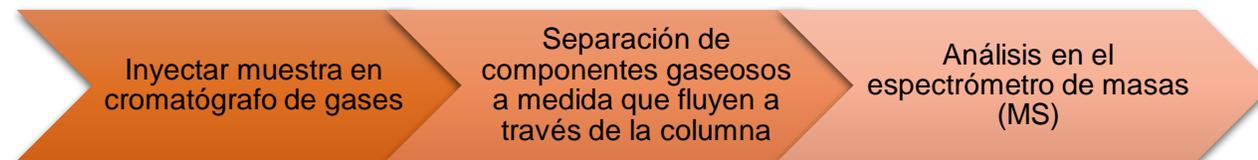
#### **4.2.3. ESPECTROMETRÍA DE MASAS**

La espectrometría de masas es considerada como una de las técnicas más completas que existen para el análisis en procesos industriales. Dentro de sus cualidades está la capacidad de identificar un espectro por cada molécula, permite medir la concentración de sustancias, es una técnica universal y específica ya que proporciona estructura de moléculas analizadas, y, además, se obtiene información detallada sobre la composición de una mezcla de gases [28].

#### **4.2.4. CROMATOGRAFÍA DE GASES EN CONJUNTO CON LA ESPECTROMETRÍA DE MASAS**

El acoplamiento que existe entre la Cromatografía de Gases y la Espectrometría de Masas se fundamenta en la separación, identificación y cuantificación de mezclas de sustancias volátiles y semivolátiles por medio de una técnica analítica [29]. Una vez realizada esta técnica, el proceso arroja una gráfica en donde se pueden observar picos característicos de cada compuesto que fue detectado. Esta gráfica (figura 2) se compone de dos ejes; uno vertical donde aparece un rango de masas según la medida, y uno horizontal donde se observa el tiempo de retención de cada compuesto.

El proceso antes mencionado puede ser apreciado en el siguiente diagrama:



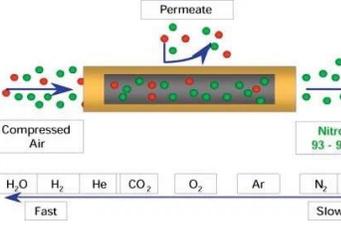
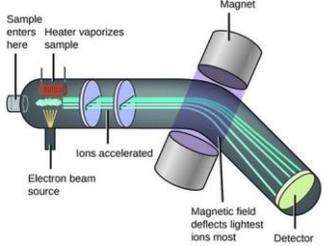
 <p>La muestra se inyecta en un puerto que se calienta hasta 300 ° C, donde el material se volatiliza.</p>	 <p>La superficie de la columna está recubierta con un material que separará los diversos compuestos químicos en la muestra en función del tamaño y / o polaridad.</p>	 <ul style="list-style-type: none"> <li>• Fuente de ionización</li> <li>• Filtro</li> <li>• Detector</li> </ul>
---	---	--

Figura 1. Diagrama de proceso de Cromatografía de gases (GCMS).

Para este acoplamiento se requieren sistemas especiales de conexión. Se trata de técnicas que trabajan en fase gaseosa por lo que resultan muy compatibles. Sin embargo, presentan un obstáculo que es la diferencia de presión a la cual está sometida cada técnica, es decir, en la columna cromatográfica sale a presión atmosférica, mientras que en el espectro debe introducirse en vacío [28]. En este tipo de estudio los datos obtenidos a partir de la técnica empleada (GCMS), para sustancias complejas la identificación de varios compuestos con mayor sensibilidad puede realizarse por medio de la técnica de detección SIR (Selected Ion Recording), con la cual se pueden detectar ciertas masas importantes con las que se trabajarán posterior al reconocimiento.

El espectro arrojado por esta técnica brinda una serie de datos que pueden ser exportados y visualizados para ser analizados según el objetivo de estudio. Esto puede ser realizado por medio de pruebas estadísticas, las cuales permiten determinar la probabilidad de que exista un compuesto característico en cada pico y sea significativo, es decir, exista un potencial biomarcador determinado por la probabilidad dada [28].

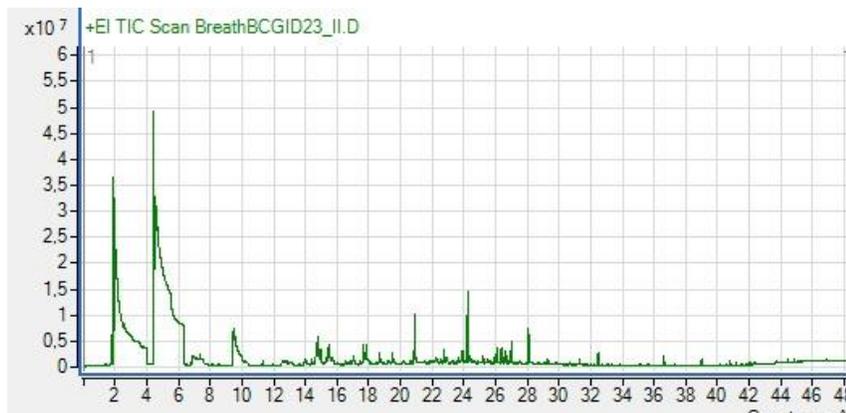


Figura 2. Cromatograma de datos obtenidos (propia autoría).

La estadística puede ser definida como una serie de diferentes métodos que se ocupan de la recolección de información, organización, presentación, análisis e interpretación de datos numéricos. Puede ser dividida en estadística descriptiva, utilizada para describir la frecuencia y distribución que presenta la variable del objeto de estudio por medio de características, y estadística inferencial, la cual está encargada del proceso de obtención de conclusiones válidas de una muestra que puedan considerarse significativas dentro de la población que se quiera estudiar [61,62].

Para hacer un análisis estadístico hay que tener en cuenta tres aspectos principales, los cuales son: diseño de investigación, número de mediciones, y escala de medición de las variables [30].

En el escenario de que aún no se presente con claridad el objetivo de estudio, debe tenerse en cuenta un análisis global de los datos, es decir, debe incluirse un análisis descriptivo de los datos que se quieren estudiar, esto se realiza con el resumen de variables estudiadas en medidas de tendencia central y de dispersión.

La tabla 1 muestra una descripción de las pruebas existentes con las características de cada una de ellas para seleccionar la prueba más adecuada según el tipo de estudio que se requiera implementar.

Tabla 1. Prueba estadística de acuerdo con los objetivos de estudio y la escala medición de las variables.

Objetivos de estudio	Variables y distribución	Tipos de muestra	Prueba recomendada
Comparar 2 promedios	Cuantitativas, distribución normal	<ul style="list-style-type: none"> <li>● Muestras relacionadas</li> <li>● Muestras independientes</li> </ul>	<ul style="list-style-type: none"> <li>● t de Student</li> <li>● t de Student</li> </ul>
	Cuantitativas discontinuas y continuas sin distribución normal	<ul style="list-style-type: none"> <li>● Muestras relacionadas</li> <li>● Muestras independientes</li> </ul>	<ul style="list-style-type: none"> <li>● Wilcoxon</li> <li>● U Mann Whitney</li> </ul>
Comparar 3 o más promedios	Cuantitativas, distribución normal	<ul style="list-style-type: none"> <li>● Muestras relacionadas</li> <li>● Muestras independientes</li> </ul>	<ul style="list-style-type: none"> <li>● Análisis de varianza (ANOVA) de 2 vías</li> <li>● ANOVA de una vía</li> </ul>
	Cuantitativas discontinuas y continuas sin distribución normal	<ul style="list-style-type: none"> <li>● Muestras relacionadas</li> <li>● Muestras independientes</li> </ul>	<ul style="list-style-type: none"> <li>● Friedman</li> <li>● Kruskal-Wallis</li> </ul>
Comparar 2 grupos	Cualitativas nominales y ordinales	<ul style="list-style-type: none"> <li>● Muestras relacionadas</li> <li>● Muestras independientes</li> </ul>	<ul style="list-style-type: none"> <li>● McNemar</li> <li>● Chi-cuadrada</li> </ul>
Comparar 3 o más grupos	Cualitativas nominales y ordinales	<ul style="list-style-type: none"> <li>● Muestras relacionadas</li> <li>● Muestras independientes</li> </ul>	<ul style="list-style-type: none"> <li>● Q de Co chran</li> <li>● Chi cuadrada</li> </ul>
Correlación de 2 variables	Cuantitativas, distribución normal	<ul style="list-style-type: none"> <li>● Muestras independientes</li> </ul>	<ul style="list-style-type: none"> <li>● Coeficiente de correlación de Pearson</li> </ul>
	Cuantitativas discontinuas y continuas sin distribución normal	<ul style="list-style-type: none"> <li>● Muestras independientes</li> </ul>	<ul style="list-style-type: none"> <li>● Coeficiente de correlación de Spearman</li> </ul>

#### 4.2.5. ANÁLISIS ESTADÍSTICO

Es definido como un componente del análisis de datos. En el contexto de la inteligencia de negocios (BI), el análisis estadístico requiere recoger y escudriñar cada muestra de datos individual en una serie de artículos desde los cuales se puede extraer las muestras. Estos datos pueden ser básicamente de 2 tipos, datos continuos y datos discretos. La

información continua es la que no puede ser contada, es decir. pueden tomar cualquier valor siempre y cuando sea dentro de un rango específico. La información discreta es la que puede ser contada como número o valor completo [42].

Para la investigación y sus variables de estudio, se tuvo en cuenta lo siguiente para la selección de las pruebas:

¿Cuál es la hipótesis planteada?: Al utilizar la hipótesis nula, se está suponiendo que se desaprobara o negara alguna afirmación "negativa" o "sin efecto", al aceptar esta hipótesis se está confirmando la probabilidad estadística de que el enunciado planteado sea no cierto, en cambio cuando esta hipótesis es rechazada, se convierte en verdadera la hipótesis alternativa o, dicho de otra forma, lo que verdaderamente se desea concluir

Las características de las muestras: Las muestras pueden ser de dos tipos, relacionadas para grupos que pueden afectarse entre sí, por ejemplo, grupos antes y después de un suceso, e independientes para grupos que no tengan relación entre sí, por ejemplo, datos tomados al azar.

La escala de medición de las variables: se define la naturaleza de la variable de forma cuantitativa o cualitativa, las variables cuantitativas tienen como atributo una cantidad y pueden ser continuas si se miden con decimales o discretas si solamente se encuentran números enteros, mientras que las variables cualitativas son categóricas o dicho de otra se clasifica por características, pueden ser nominales en donde se clasifican en categorías pero no se les da un orden jerárquico mientras que los ordinales permite establecer una secuencia lógica de la clasificación

#### **4.2.5.1. TEST DE NORMALIDAD**

Comparan los puntajes en la muestra con un conjunto de puntajes distribuidos normalmente (distribución gaussiana) con la misma media y desviación estándar; La hipótesis nula es que "la distribución de la muestra es normal". Si la prueba es significativa, la distribución no es normal [63].

#### 4.2.5.1.1. KOLMOGOROV-SMIRNOV

Esta prueba compara la función de distribución acumulada empírica de los datos de la muestra con la distribución esperada si los datos fueran normales. Si esta diferencia observada es adecuadamente grande, la prueba rechazará la hipótesis nula de normalidad de la población. Si el valor p de esta prueba es menor que el nivel de significancia ( $\alpha$  o Alpha) elegido, se rechaza la hipótesis nula y se concluye que se trata de una población no normal [65].

El estadístico de prueba es el siguiente:

$$T = \sup_{1 \leq i \leq n} |F_n(x_i) - F_o(x_i)| \quad (\text{Ec. 1})$$

Donde:

- $x_i$  = es el  $i$ -ésimo valor observado en la muestra (cuyos valores se han ordenado previamente de menor a mayor).
- $F_n(x_i)$  = es un estimador de la probabilidad de observar valores menores o iguales que  $x_i$
- $F_o(x_i)$  = es la probabilidad de observar valores menores o iguales que  $x_i$  cuando  $H_0$  es cierta [66].

#### 4.2.5.1.2. SHAPIRO- WILK

Esta prueba de normalidad trabaja con media y varianza no especificadas. La prueba Shapiro-Wilk es más apropiada para muestras pequeñas (<50 muestras), pero también puede manejar muestras de hasta 5000.

El estadístico de prueba es la siguiente:

$$W_c = \frac{b^2}{\sum_{i=1}^n (X_i - \bar{X})^2} \quad (\text{Ec. 2})$$

$$b = \sum_{i=1}^k a_i [X_{(n-i+1)} - X_i] \quad (\text{Ec. 3})$$

Donde:

- $a_i$  = el valor de un coeficiente que se encuentra tabulado para cada tamaño de muestra y la posición  $i$  de cada observación.
- $[X_{(n-i+1)} - X_i]$  = diferencias sucesivas que se obtienen al restar el primer valor al último valor, el segundo al penúltimo, el tercero al antepenúltimo y así hasta llegar a restar el último al primer valor [64].

#### 4.2.6. DIFERENCIA ENTRE PRUEBAS PARAMÉTRICAS Y NO PARAMÉTRICA

Las pruebas paramétricas tienen distribuciones estadísticas que pertenecen a los datos en estudio. Por lo tanto, deben cumplirse ciertas condiciones que son propias de estas distribuciones, de tal modo que esta prueba sea fiable con respecto a la información brindada.

Las pruebas no paramétricas no presentan un tipo de distribución, por lo que pueden aplicarse sin necesidad que cumplan algún tipo de condición. Se consideran que estas pruebas son la alternativa a las pruebas paramétricas [31,38].

#### 4.2.7. WILCOXON

Permite comparar poblaciones cuando sus distribuciones (normalmente interpretadas a partir de las muestras) no satisfacen las condiciones necesarias para otros test paramétricos [51].

$$Z = \frac{R - \mu}{\sigma} \quad (\text{Ec. 4})$$

$$\mu = \frac{n_1(n_1 + n_2 + 1)}{2} \quad (\text{Ec. 5})$$

$$\sigma = \sqrt{\frac{n1*n2(n1+n2+1)}{12}} \quad (\text{Ec. 6})$$

Donde

- R= es la suma de rangos del menor grupo
- n1= tamaño de muestra del menor grupo
- n2= tamaño de muestra del mayor grupo

#### 4.2.8. U MANN-WHITNEY

Es una prueba no paramétrica con la cual se identifican diferencias entre dos poblaciones basadas en el análisis de dos muestras independientes, cuyos datos han sido medidos al menos en una escala de nivel ordinal [52].

$$U1 = n1 * n2 + \frac{n1(n1+1)}{2} - R1 \quad (\text{Ec. 7})$$

$$U2 = n1 * n2 + \frac{n2(n2+1)}{2} - R2 \quad (\text{Ec. 8})$$

$$Z = \frac{U - \frac{n1*n2}{2}}{\sqrt{\frac{n1*n2(n1+n2+1)}{12}}} \quad (\text{Ec. 9})$$

Donde:

- n1= Tamaño de la muestra del grupo 1
- n2= Tamaño de la muestra del grupo 2
- R1= Sumatoria de rangos del grupo 1
- R2= Sumatoria de rangos del grupo 2
- U= El menor valor estadístico entre U1 y U2

## 5. MARCO LEGAL

El reglamento para las enfermedades generadas por MTC es amplio y se debe tener en cuenta que la normativa se aplica a diferentes niveles administrativos: autonómico, nacional, comunitario e internacional. En la Unión Europea, la Directiva del Consejo 391/77/CEE, de 17 de mayo de 1977 por la que se establece una acción de la Comunidad para la erradicación de la brucelosis, de la tuberculosis y de la leucosis de los bovinos, la Directiva 78/52/CEE, de 13 de Diciembre de 1977, por la que se establecen los criterios comunitarios aplicables a los planes nacionales de erradicación acelerada de la brucelosis, de la tuberculosis y la leucosis enzoótica de los bovinos y la Directiva 64/432/CEE, de 26 de junio de 1964, y sus modificaciones, relativa a problemas de policía sanitaria en materia de intercambios intracomunitarios de animales de la especie bovina y porcina, establece las pautas generales de actuación en los intercambios intracomunitarios de los animales de reproducción, producción o abasto de las especies bovina y porcina [54,56].

A escala nacional, el Real Decreto 2611/1996, de 20 de diciembre, y sus modificaciones, por el que se regulan los programas nacionales de erradicación de enfermedades de los animales; así como el Real Decreto 1440/2001 [54].

Las medidas a través de normas y documentos son la estrategia utilizada para la erradicación de la Tuberculosis bovina tanto a nivel de la Unión Europea como a nivel nacional, apoyados sobre una base científica sólida [54].

El ICA mediante la Resolución 1513 del 15 de julio de 2004, estableció las medidas necesarias sanitarias, para la prevención, control y la erradicación de la Tuberculosis bovina en Colombia, esta Resolución requiere ser actualizada con el objetivo de fortalecer cada una de las actividades mencionadas anteriormente [55].

Para realizar el muestreo de aquellos pacientes o animales que fueron infectados y se encuentran enfermos, existen métodos estadísticos necesarios con el fin de realizar estudios científicos especificados en el Código Sanitario para los Animales Terrestres de la OIE. Es posible calcular el número de animales de una manada de un tamaño determinado que se deben muestrear, para que la detección de la infección, que se

asume que está presente en un determinado porcentaje de animales, alcance una probabilidad del 95% [57].

Artículo 2.13.6.4.1. Campaña Nacional de Control y Erradicación de la Tuberculosis Bovina. Establéese la Campaña Nacional de Control y Erradicación de la Tuberculosis Bovina bajo la dirección del ICA y con la colaboración del Ministerio de Salud y Protección Social. (Decreto 438 de 1979, art. 1) [58].

Artículo 2.13.6.4.3. Control Sanitario Especial. Los predios, los animales y sus productos que sean declarados como infectados por tuberculosis bovina, serán sometidos a un control sanitario especial por el Ministerio de Salud y Protección Social y el Instituto Colombiano Agropecuario, ICA. (Decreto 438 de 1979, art. 3) [58].

De acuerdo con lo anterior, las normas o decretos que estén vigentes representan un aporte significativo, ya que, al momento de investigar una enfermedad clínica, las muestras que se obtienen de los pacientes enfermos son representativas para la comparación con pacientes sanos y realizar el respectivo análisis. Teniendo esto en cuenta, deben seguirse ciertos procesos que ya están establecidos. Así mismo, se conocen métodos de diagnóstico para la enfermedad que han sido reglamentados con la serie de pasos a seguir, como también se menciona la erradicación de las enfermedades en los animales.

## **6. PROCEDIMIENTO METODOLÓGICO**

### **6.1. METODOLOGÍA**

El método de investigación utilizado es el sistémico-analítico, debido a que se estudió el comportamiento de los biomarcadores por medio de la identificación de compuestos que es una variable que afectó directamente el análisis por su composición. Además, se buscó determinar la relación de las variables obtenidas por el proceso de cromatografía de gases (GCMS).

### **6.2. TIPO DE ESTUDIO**

El tipo de estudio que se llevó a cabo es de tipo exploratorio, ya que se determinaron las características y/o variables que identifican una permisible relación para el desarrollo de un potencial biomarcador por medio de estudios estadísticos que permitieron observar la tendencia producida por el tipo de dato a estudiar o analizar.

### **6.3. POBLACIÓN Y MUESTRA**

Esta investigación va dirigida a aquellos animales que poseen la enfermedad para ser analizada, en este caso Tuberculosis Bovina. Las muestras se extrajeron en el Sur de España en el Parque Nacional de Doñana y otras muestras en Rumania con un total de 270 animales. Para la evaluación del software se utilizaron muestras de aliento (BREATH) y muestras fecales (FECES), tanto en animales infectados y no infectados en periodos de tiempos diferentes.

Estas muestras fueron tomadas en el marco de un proyecto de investigación europeo adscritos a la convocatoria Horizonte 2020 de la unión europea. Los animales que fueron objeto de estudio se le extrajeron en dos instantes de tiempo, una en el instante de tiempo cuando fueron diagnosticados y la segunda, un tiempo después de haber aplicado la vacuna a algunos grupos de animales para evidenciar la evolución de la enfermedad en cada grupo.

Los subgrupos se detallan a continuación:

Control

BCG

Michigan

Spanish

Los animales en el grupo control, fueron animales que no se les aplicó vacunas, solo se les hizo el seguimiento en el tiempo.

Los animales dentro de los otros tres grupos (BCG, Michigan, Spanish), fueron animales que recibieron la vacuna justo después de la primera toma de muestras y se les hizo el seguimiento para analizar su evolución en el tiempo.

#### **6.4. FASES DEL DISEÑO**

En esta sección se definirán las fases para el desarrollo, implementación y validación del software, las cuales fueron necesarias para el cumplimiento de las actividades propuestas en el cronograma y plan de trabajo.

##### **6.4.1. FASE UNO**

La primera fase consistió en la caracterización de los datos obtenidos, por lo cual se realizaron diferentes investigaciones que permitieron un análisis del tipo de enfermedad estudiada con el fin de obtener los potenciales biomarcadores. Esto teniendo en cuenta que la información debe contener desde las características generales hasta las específicas, para así determinar finalizada la fase 3, la incidencia de los compuestos en los pacientes estudiados. Una vez realizada la investigación, se procedió con la identificación de las pruebas estadísticas existentes y se planteó como debe seleccionarse la prueba más adecuada según el objetivo de estudio y criterios establecidos en un protocolo de investigación.

## 6.4.2. FASE DOS

En la fase dos inicialmente se crearon plantillas en Excel con los cálculos manuales de las pruebas escogidas, las cuales fueron U de Mann-Whitney para pruebas independientes y Wilcoxon para pruebas dependientes con el fin de aplicar un primer análisis de los datos obtenidos. Una vez realizadas las pruebas en Excel, se verificaron los resultados y posterior a ello se determinaron los requisitos necesarios para el diseño de la arquitectura del software mediante los diagramas de flujo para seguir un orden en el desarrollo de las pruebas estadísticas por cada grupo de acuerdo con los datos estudiados. Sin embargo, se hizo necesario la realización de pruebas de normalidad para comprobar que los datos que se estaban analizando no se ajustaran a una distribución normal, por ello se anexo al esquema del software la aplicación de los tests de normalidad para la verificación de que se seleccione una prueba estadística adecuada antes de escogerla.

Para describir gráficamente el funcionamiento del software se crearon los siguientes diagramas de flujos (figura 3 y figura 4) para establecer los procedimientos y puntos de decisiones a seguir de acuerdo a las opciones o requerimientos, como, por ejemplo, la cantidad de grupos o muestras con las que se trabajará.

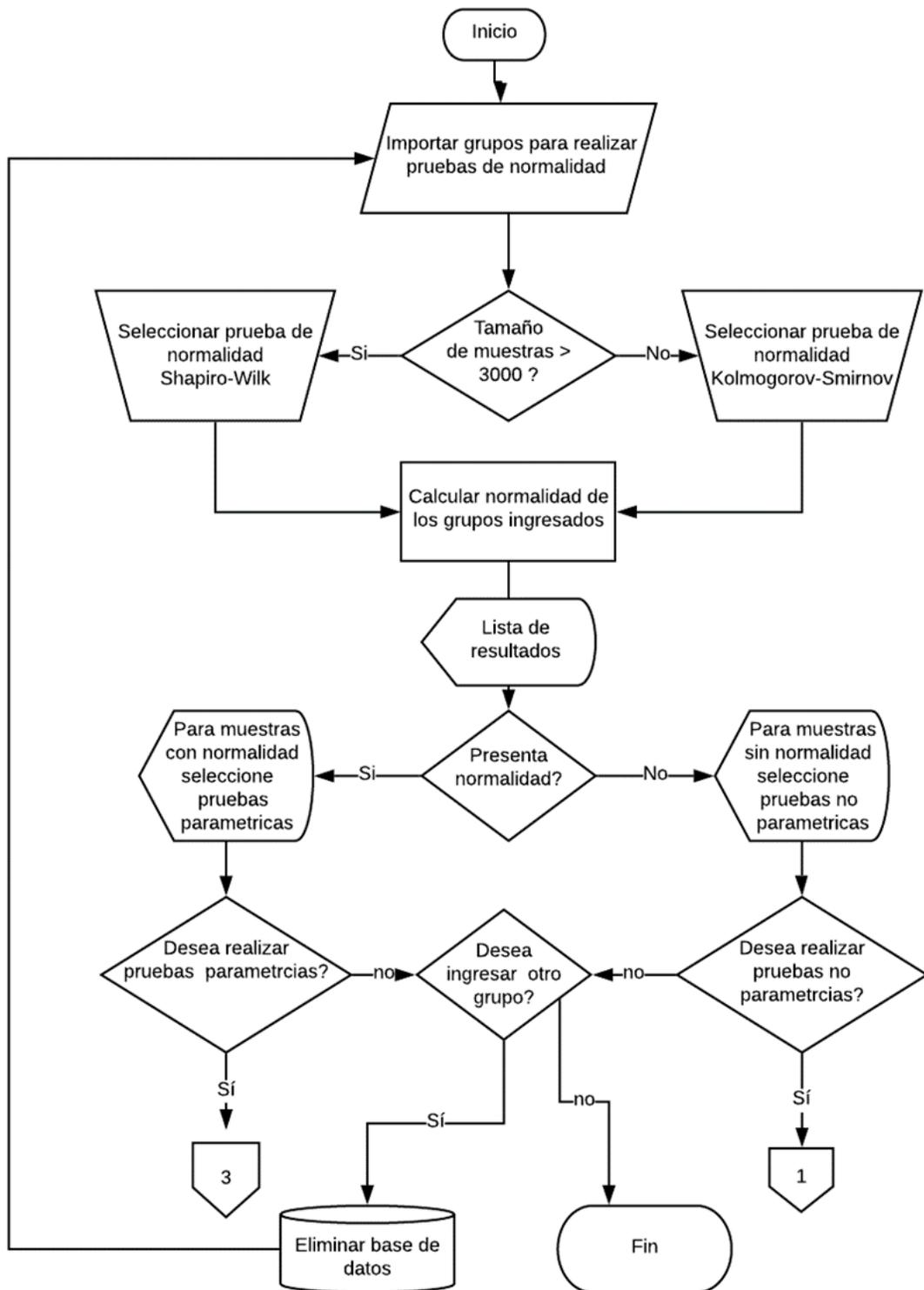


Figura 3. Diagrama de flujo de test de normalidad.

La figura 3 corresponde al diagrama de flujo, el cual representa la primera interfaz que se encontrara el usuario. Como puede observarse al inicio del programa el usuario deberá ingresar los grupos a los cuales desea realizar la prueba de normalidad, dependiendo del tamaño muestral podrá escoger entre Shapiro-Wilk o Kolmogórov-Smirnov. El software realizará una serie de cálculos dependiendo de la prueba y después mostrará los resultados e indicará si presenta distribución normal o no. Luego, el usuario tendrá la opción de trabajar con pruebas no paramétricas (continua el programa en el cuadro 1) o paramétricas (continua el programa en el cuadro 3), y en caso de que aún no desee seleccionar esas pruebas, podrá ingresar otro grupo o finalizar el programa.

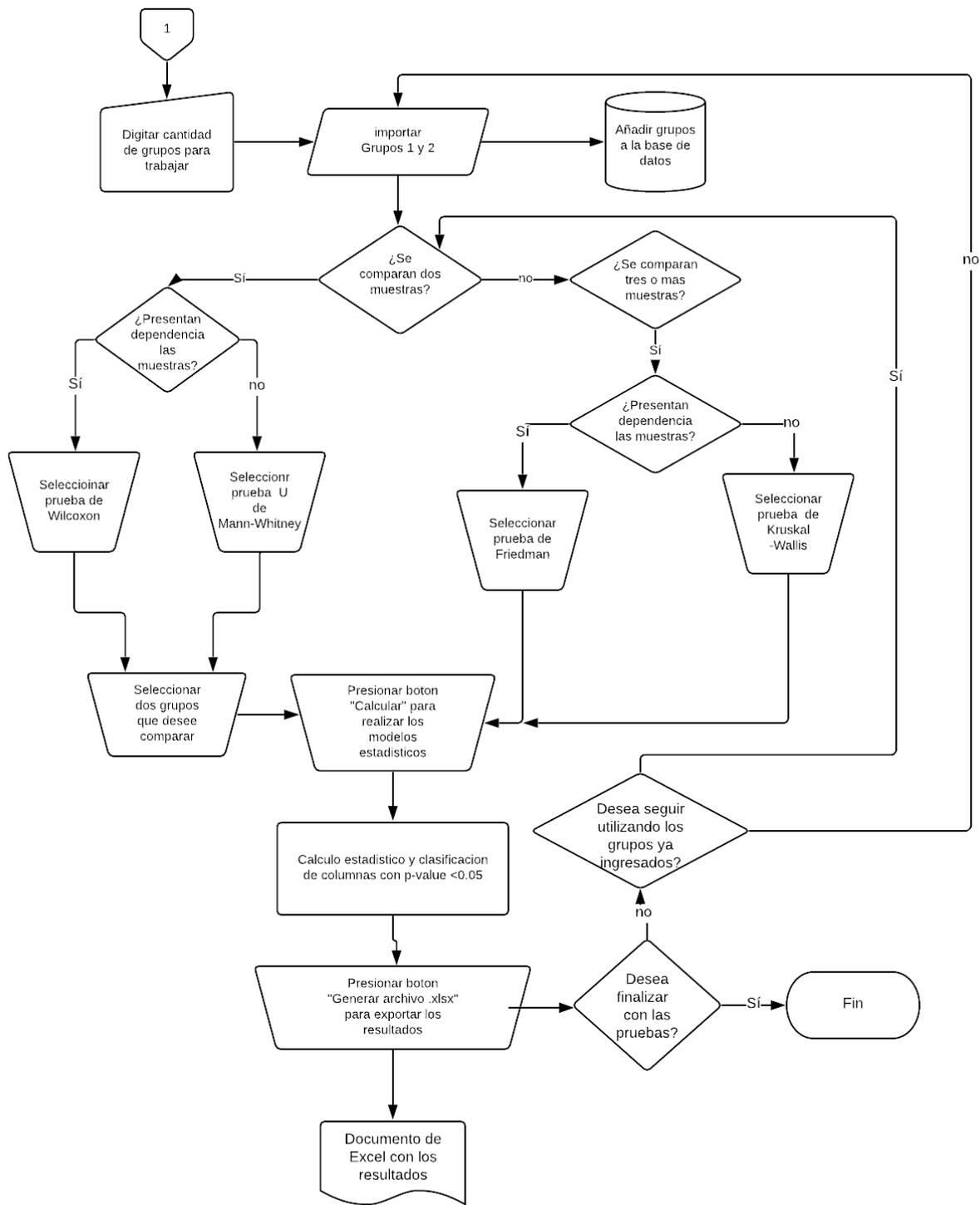


Figura 4. Diagrama de flujo de pruebas no paramétricas (comparación de dos muestras).

La figura 4 representa la interfaz donde se encontrarán las pruebas no paramétricas, después de ingresar los grupos si se desea comparar entre dos grupos, deberá escoger la prueba Mann-Whitney si los grupos no presentan dependencia entre ellos, en caso de que si, se utilizará la prueba de Wilcoxon. En caso de que el usuario desee comparar entre tres o más grupos se selecciona Friedman si presentan dependencia y en caso de que no se elegirá Kruskal-Wallis. Independientemente de la prueba seleccionada, los resultados serán exportados en Excel en donde solo se enviarán los tiempos o columnas que cuyo resultado de probabilidad fue menor o igual que 0.05

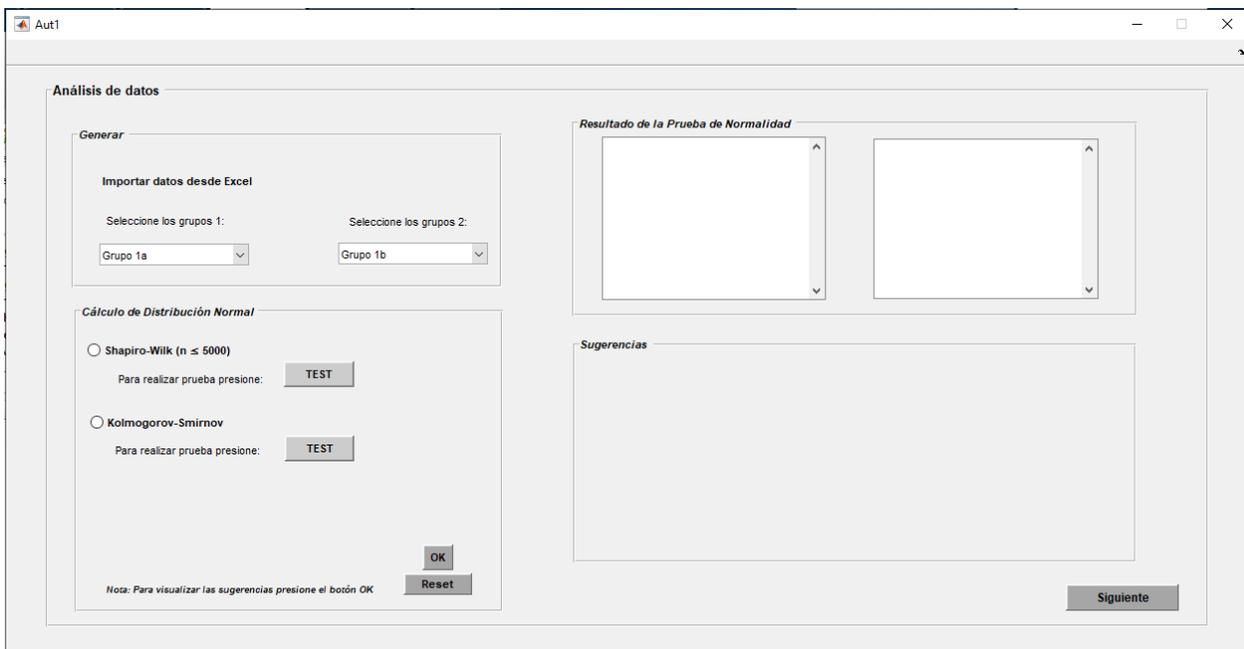


Figura 5. Interfaz del Software Prueba (Guide 1 en MATLAB).

En la figura 5 se puede observar la interfaz donde se aplicarán las pruebas de normalidad, y en la parte derecha se mostrarán los resultados de las pruebas por pacientes de estudio. Además, se sugieren diferentes tipos de pruebas según el análisis que este realizando (si se comparan dos o más muestras).

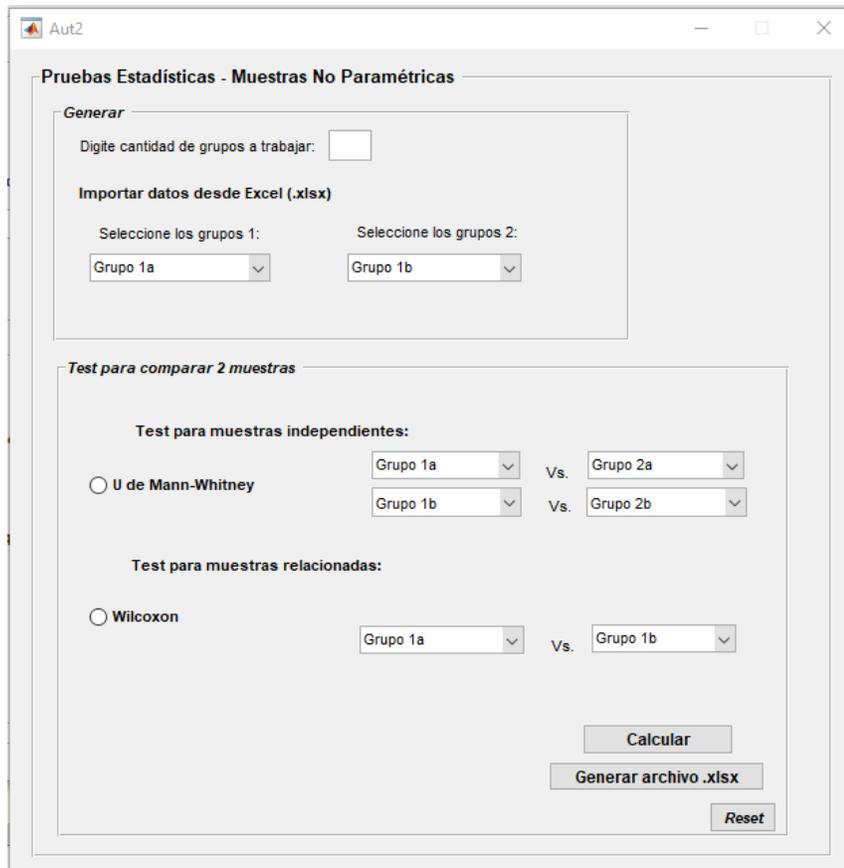


Figura 6. Interfaz del Software Prueba (Guide 2 en MATLAB).

La figura 6 hace referencia a la interfaz donde se realizarán las pruebas estadísticas luego de que el usuario aplique las pruebas de normalidad. En esta interfaz, luego de hacer los cálculos, se exportarán las columnas en Excel que posean una probabilidad menor o igual que 0.05.

### 6.4.3. FASE TRES

En la fase 3, para la validación del software, se compararon las probabilidades obtenidas en Matlab con respecto a los resultados obtenidos manualmente en Excel, como estos últimos ya estaban verificados, lo que se hizo fue comparar aquellas columnas cuya probabilidad fue menor o igual que 0.05 en ambos resultados. Teniendo esto en cuenta, se comparó el tiempo y si la probabilidad resultante fue la misma, en caso de que no fuera así, se hizo una revisión completa de todas las configuraciones realizadas en Matlab para determinar el problema y realizar las mejoras. Además, para la identificación de los

compuestos posterior a la prueba que se exportó en Matlab, se usó el entorno de MassHunter con los programas de Qualitative y Unknown Analysis, con los cuales a partir del tiempo de retención y la masa del compuesto se identificó qué compuestos pertenecían a cada tiempo arrojado por el software según la probabilidad obtenida.

## 6.5. CRONOGRAMA – PLAN DE TRABAJO

Universidad Autónoma del Caribe Proyecto de Grado - Ingeniería Mecatrónica Cronograma de actividades				
Componentes	Descripción	Fecha Inicio	Fecha Final	Duración (Días)
<b>OBJETIVO 1</b>	<b>Caracterizar las diferentes pruebas estadísticas para el análisis de datos en grupos de estudio con enfermedades y grupos controles que permita la identificación de potenciales biomarcadores.</b>	11/1/2019	12/15/2019	26
<b>Entregable # 1</b>	<b>Identificar las características de los datos obtenidos en los pacientes estudiados.</b>	11/1/2019	11/23/2019	22
Actividad 1	Investigar el tipo de enfermedad a tratar.	11/1/2019	11/11/2019	10
Actividad 2	Determinar la cantidad de muestras obtenidas a partir de la población estudiada.	11/12/2019	11/15/2019	3
Actividad 3	Establecer grado de dependencia entre el conjunto de muestras	11/17/2019	11/23/2019	6
<b>Entregable # 2:</b>	<b>Pruebas de validación estadística necesarias para la implementación y programación del dispositivo.</b>	11/24/2019	12/15/2019	21
Actividad 5	Investigar las diferentes clases de pruebas estadísticas	11/24/2019	11/28/2019	4
Actividad 6	Clasificar las pruebas estadísticas según los tipos de datos analizados	11/29/2019	12/4/2019	5
Actividad 7	Relacionar los modelos estadísticos con las características de los datos obtenidos	12/4/2019	12/14/2019	10
Actividad 8	Selección de las pruebas de validación estadística	12/14/2019	12/15/2019	1
<b>OBJETIVO 2</b>	<b>Implementar software prototipo que permita el estudio del comportamiento de los pacientes mediante la lectura de datos para la detección de compuestos característicos de cada enfermedad.</b>	1/4/2020	2/17/2020	44

<b>Entregable # 3</b>	<b>Elaborar bases de datos con respecto a la enfermedad de Tuberculosis Bovina en Animales.</b>	<b>1/4/2020</b>	<b>1/26/2020</b>	<b>22</b>
Actividad 9	Identificar los compuestos característicos de la Tuberculosis Bovina	1/4/2020	1/18/2020	14
Actividad 10	Crear tabla de bases de datos de la enfermedad identificada (Tuberculosis Bovina)	1/19/2020	1/24/2020	5
Actividad 11	Ingresar los registros de la enfermedad a la tabla de base de datos para su posterior análisis	1/25/2020	1/29/2020	4
<b>Entregable # 4</b>	<b>Diseñar un prototipo de software adecuado que permita el hallazgo de potenciales biomarcadores.</b>	<b>1/30/2020</b>	<b>2/23/2020</b>	<b>24</b>
Actividad 12	Diseño de la arquitectura del software	1/30/2020	2/10/2020	11
Actividad 13	Implementar la prueba estadística de Wilcoxon para muestras dependientes	2/11/2020	2/14/2020	3
Actividad 14	Implementar la prueba estadística de Mann-Whitney-Wilcoxon para muestras independientes	2/15/2020	2/18/2020	3
Actividad 15	Validación de resultados obtenidos a partir de las pruebas estadísticas implementadas en software existente	2/19/2020	2/23/2020	4
<b>OBJETIVO 3</b>	<b>Validar software para pruebas</b>	<b>2/23/2020</b>	<b>4/3/2020</b>	<b>40</b>
<b>Entregable # 5</b>	<b>Pruebas de los parámetros necesarios del software para la implementación del análisis estadístico.</b>	<b>2/25/2020</b>	<b>3/20/2020</b>	<b>24</b>
Actividad 16	Describir el resultado de las pruebas acorde a las hipótesis planteadas para determinar su nivel de confiabilidad	2/25/2020	3/9/2020	13
Actividad 17	Comprobar la información a través de la implementación en el Software Masshunter	3/10/2020	3/20/2020	10
<b>Entregable # 6</b>	<b>Pruebas de seguridad y funcionalidad del software para la detección de potenciales biomarcadores.</b>	<b>3/23/2020</b>	<b>4/3/2020</b>	<b>11</b>
Actividad 18	Evaluar la conformidad con la especificación de requisitos.	3/23/2020	3/29/2020	6
Actividad 19	Verificación del software y funcionalidad del mismo a partir del diseño creado	3/30/2020	4/3/2020	4

## 7. PRESUPUESTO

A continuación, se presenta el cómputo estimado de los costos y gastos, respecto a personal científico, personal de apoyo, equipos y materiales e insumos, presentados en el proceso de efectuar software.

### 7.1. PRESUPUESTO GENERAL

Tabla 2. Presupuesto general.

 <b>UNIVERSIDAD AUTÓNOMA DEL CARIBE</b> <small>Barranquilla, Atlántico</small>	<b>PRESUPUESTO PARA PROYECTOS DE INVESTIGACIÓN</b>				GC-IV-PR-05-03
					Versión 4
					16/10/2018
<b>PRESUPUESTO GENERAL DEL PROYECTO</b>					
RUBROS	Fuentes de Financiamiento				Total
	Vicerrectoría de Investigación y Transferencia UAC	Facultad / Programa	Otras fuentes Externas	Contrapartida UAC	
1. Personal Científico	\$ 0	\$ 0	\$ 0	\$ 29.769.936	\$ 29.769.936
2. Personal de Apoyo	\$ 0	\$ 0	\$ 0	\$ 3.426.816	\$ 3.426.816
3. Consultoría Especializada y Servicios Técnicos	\$ 0	\$ 0	\$ 0	\$ 0	\$ 0
4. Materiales e Insumos	\$ 0	\$ 0	\$ 28.417.451	\$ 0	\$ 28.417.451
5. Salidas de Campo	\$ 0	\$ 0	\$ 0	\$ 0	\$ 0
6. Equipos	\$ 0	\$ 0	\$ 3.400.000	\$ 0	\$ 3.400.000
7. Bibliografía	\$ 0	\$ 0	\$ 0	\$ 0	\$ 0
8. Difusión de Resultados	\$ 0	\$ 0	\$ 0	\$ 0	\$ 0
9. Viajes	\$ 0	\$ 0	\$ 0	\$ 0	\$ 0
<b>TOTAL PRESUPUESTO DEL PROYECTO</b>	<b>\$ 0</b>	<b>\$ 0</b>	<b>\$ 31.817.451</b>	<b>\$ 33.196.752</b>	<b>\$ 65.014.203</b>

## 7.2. PERSONAL CIENTÍFICO Y DE APOYO

El presupuesto invertido en este rubro consiste en el costo del tiempo empleado por el personal de investigación vinculados a este proyecto, que incluye a los directores y a los auxiliares de investigación.

Tabla 3. Costo personal científico.

1. PERSONAL CIENTIFICO										
Nombres y Apellidos	Función dentro del Proyecto	Tipo de Contrato	Valor Hora (\$)	Dedicación Horas/semana	No. de Semanas	Fuentes de Financiamiento				
						Vicerrectoría de Investigación y Transferencia	Facultad / Programa	Otras Fuentes Externas	Contrapartida UAC	SUB-TOTAL
1. MSc. Kelvin Beleño Sáenz	Invest. Principal	Titular	\$ 46.666	7	48				\$ 15.679.776	\$ 15.679.776
2. MSc. Carlos Díaz Sáenz	Co-Investigador	Asociado	\$ 41.935	7	48				\$ 14.090.160	\$ 14.090.160
<b>SUB-TOTAL</b>						<b>\$ 0</b>	<b>\$ 0</b>	<b>\$ 0</b>	<b>\$ 29.769.936</b>	<b>\$ 29.769.936</b>

Tabla 4. Costo personal de apoyo.

2. PERSONAL DE APOYO										
Nombres y Apellidos	Función dentro del Proyecto	Tipo de Vinculación	Valor Hora (\$)	Dedicación Horas/semana	No. de Semanas	Fuentes de Financiamiento				
						Vicerrectoría de Investigación y Transferencia	Facultad / Programa	Otras Fuentes Externas	Contrapartida UAC	SUB-TOTAL
1. Daniela Lucia Movilla Echeverri	Aux. Investigación	Semillero	\$ 2.231	16	48				\$ 1.713.408	\$ 1.713.408
2. Daniela Patricia Márquez Sandoval	Aux. Investigación	Semillero	\$ 2.231	16	48				\$ 1.713.408	\$ 1.713.408
<b>SUB-TOTAL</b>						<b>\$ 0</b>	<b>\$ 0</b>	<b>\$ 0</b>	<b>\$ 3.426.816</b>	<b>\$ 3.426.816</b>

### 7.3. MATERIALES, INSUMOS Y EQUIPOS

El presupuesto dedicado a esta sección incluye las licencias de programas utilizados para desarrollar y verificar el funcionamiento del software y la base de datos con la cual se hallarán los biomarcadores.

Tabla 5. Costo materiales e insumos.

4. MATERIALES E INSUMOS						
Descripción	Justificación	Fuentes de Financiamiento				
		Vicerrec toría de Investig ación y Transfer encia	Facultad / Program a	Otras Fuentes Externas	Contraparti da UAC	SUB-TOTAL
1. Bases de datos	Librería con los compuestos existentes			\$ 9.974.570		\$ 9.974.570
2. Licencia: Programa OriginPro	Alineación de los datos para diferentes tamaños muestrales			\$ 3.567.140		\$ 3.567.140
3. Licencia: Programa SPSS	Validación de hipótesis estadística y probabilidad			\$ 5.146.878		\$ 5.146.878
4. Licencia: Masshunter	Programa para identificar los compuestos			\$ 4.713.989		\$ 4.713.989
5. Licencia Matlab	Programa para el diseño de interfaz y programación del software			\$ 3.811.399		\$ 3.811.399
6. QIMacros	Validación de hipótesis estadística			\$ 1.203.475		\$ 1.203.475
<b>SUB-TOTAL</b>		<b>\$ 0</b>	<b>\$ 0</b>	<b>\$ 28.417.451</b>	<b>\$ 0</b>	<b>\$ 28.417.451</b>

Tabla 6. Costo equipos usados

6. EQUIPOS							
Descripción	Justificación	Cantidad	Fuentes de Financiamiento				
			Vicerrectoría de Investigación y Transferencia	Facultad / Programa	Otras Fuentes Externas	Contrapartida UAC	SUB-TOTAL
1. Computador	Investigaciones y diseño del software	2			\$ 3.400.000		\$ 3.400.000
<b>SUB-TOTAL</b>			\$ 0	\$ 0	\$ 3.400.000	\$ 0	\$ 3.400.000

## 8. PRESENTACIÓN Y ANÁLISIS DE RESULTADOS

### 8.1. DISEÑO DEL PROTOTIPO

En el presente se encuentran los diagramas de flujos que fueron usados como guía acorde como debe funcionar el software y las variables a tener en cuenta, a su vez se encuentra un diseño prototipo de la interfaz el cual fue creado en Matlab. Los pasos que sigue el prototipo son los siguientes

1. Pedir los datos de los grupos para realizar uno de los test de normalidad propuestos, acorde al número de grupos importados.
2. Se selecciona un test de normalidad y se escoge uno de los grupos
3. Si el resultado fue que presenta normalidad se debe usar pruebas paramétricas, de lo contrario se usan pruebas no paramétricas.
4. Se realizan los pasos 2 y 3 hasta usar todos los grupos.
5. Una vez finalizado el test de normalidad, y al oprimir el botón “OK” saldrán unas sugerencias guía, para indicar que tipo de pruebas se utiliza acorde a los datos. Se procede a la segunda fase del software.
6. Se abrirá una segunda ventana, en la cual se ingresará el número de grupos con el que se trabajará, después de ingresar el número e importar los grupos acorde al número, se seleccionará entre las pruebas disponibles, acorde el número de muestras a trabajar.
7. Una vez seleccionada la prueba, se escogerá a qué grupo se le hará la prueba y se pulsará en el botón de “CALCULAR”, y después al botón de “GENERAR ARCHIVO .XLSX”.
8. Con esto se generará un archivo en Excel con los cálculos y los resultados.

## 8.2. ANÁLISIS DE RESULTADOS

Los datos tratados corresponden a dos tipos de pruebas que fueron realizadas en los pacientes para verificar la similitud de compuestos presentes en la enfermedad y si esto genera algún cambio. Por lo tanto, se tienen dos grupos denominados BREATH y FECES que corresponden a las pruebas realizadas, cada uno de estos con los mismos pacientes sometidos a procedimientos diferentes.

Donde Control I, II, III y IV hace referencia a muestras tomadas en animales sanos, es decir, son el punto de referencia a comparar con los animales infectados.

Los grupos II (para muestras de aliento (BREATH II)) y grupos III (para muestras fecales (FECES III)) fueron tomados de animales infectados de tuberculosis bovina, antes de aplicarse una vacuna.

Los grupos I (para muestras de aliento (BREATH I)) y grupo IV (para muestras fecales (FECES IV)) fueron tomados de los animales infectados de los grupos II y III respectivamente, tiempo después de aplicárseles una vacuna.

En el desarrollo del análisis estadístico presentado se tuvo en cuenta la implementación de estas pruebas en otro software estadístico que permitieran el cálculo del test de Wilcoxon y el test U Mann-Whitney como lo es SPSS. Se puso a prueba el tiempo en que se realiza cada test, incluyendo el procesamiento que debe existir previo, como lo es la organización de los datos, si estos están en una escala ordinal o nominal (según el manual para el uso de SPSS) y cómo se comparan los promedios entre los grupos de estudio. Además, los resultados obtenidos se tuvieron en pares, es decir, el estudio aplicado se realiza entre dos muestras. Si se trabaja con una gran cantidad de datos (expresado en columnas), el software investigado para evaluar dichas pruebas, sólo valorará columna por columna, donde el usuario es quien elige con cuales desea trabajar, pero este limita a que sean dos (2) columnas. Como se desean comparar dos promedios, se volvió repetitivo el proceso debido a la cantidad de datos con los que se trabajó y poco automático.

En la implementación realizada por medio del software MATLAB, con el desarrollo del programa aplicado se logra determinar la probabilidad de encontrar un potencial biomarcador dentro de los grupos ingresados teniendo en cuenta la organización previa que deben tener según el modelo estadístico que fue aplicado. En la Figura 7, se puede observar una tabla, la cual fue realizada por el software una vez haya finalizado el cálculo correspondiente por grupo seleccionado o ingresado utilizando la prueba de Wilcoxon, en cada hoja se encuentran los cálculos por grupo indicando según la nomenclatura que se usó en el desarrollo de software.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	43,05404	59,03078	57,07022	82,94602	84,94268	65,03887	205,01534	92,06377	230,19383	82,94628	43,05411	207,03062	207,03036	78,04635	78,04621	178,99374	209,02682	116,90512
2	1,4275384	1,4605334	1,754404	1,85725	1,9008286	4,507953	4,5419664	4,555846	4,811233	5,585059	6,5309715	7,361733	7,4655004	9,3813	9,471493	11,685688	13,031499	13,537666
3																		
4	1	260448,5	4614054,5	1	11774209	8655969	306520,3	336626,94	239907,25	1	167095,7	1	1	1	83222520	1	2009639,4	1
5	1	139891,7	4055088	1	19188752	225122,9	2068190,1	50839828	312855,12	1	1	1	1	1	106618280	1	1649842,5	1
6	118178,39	99611,78	5014796,5	1	13749619	9053980	75354,44	11810689	250004,12	1	202692,53	1	1	1	79555272	1	455920,47	1
7	1	217982,62	4613404,5	4324166,5	7124996,5	3373270,5	290988,88	529421632	111931,07	1	1	1237275,8	1	1	91010496	1	2634875,2	1
8	73010,68	1	4083361	1	1	6248879,5	460560,56	36683080	199567,3	1	1	1750855,4	1	1	49836440	1	2168054,5	1
9	1	256932,77	5431866,5	1	19203962	1	282171,4	1	109555,49	1	1	136829,3	1	1	55007212	1	256568,45	1
10																		
11	286225,25	1	6801248,5	6906556	24598450	24045050	1	1	1	1	200487,62	1719594,1	45139,61	8323307	56081372	422415,4	1	635390,4
12	324143,06	1	31417682	13488059	21205394	100397,57	1	658401,94	181914,81	1772005,5	1	548095,1	1	4772183,5	21729690	1	1	622873,25
13	486555,47	1	5767496,5	10116372	26867024	101233800	1	1	1	1048751,5	432563,94	1495299,2	88284,24	6426300	1174855	383330,44	1	1
14	441392,75	1	5444713,5	15241585	38834292	176466112	320742,84	1	1	773620	245981,66	918503,9	51456,9	7136415,5	45489656	122109,94	177290,81	867106,9
15	304171,03	1	7092246	1	12805501	42795876	1	1	1	325817,84	451333,06	2202966	51258,46	1	58649596	284880	1	1979897,5
16	1	1	8189088	6484238	17808822	100897248	1	1	1	710791,06	275061,7	1803253,5	5556689,5	4700097,5	1	299367,53	1	1126283,9
17																		
18	20	-20	21	20	19	19	-19	-20	-21	20	20	19	20	20	-19	20	-21	20
19	2,0965697	-2,0965697	2,2013982	2,0965697	1,9917412	1,9917412	-1,9917412	-2,0965697	-2,2013982	2,0965697	2,0965697	1,9917412	2,0965697	2,0965697	-1,9917412	2,0965697	-2,2013982	2,0965697
20																		
21	0,0360317	0,0360317	0,0277078	0,0360317	0,0463995	0,0463995	0,0463995	0,0360317	0,0277078	0,0360317	0,0360317	0,0463995	0,0360317	0,0360317	0,0463995	0,0360317	0,0277078	0,0360317
22																		
23																		

Figura 7. Resultados obtenidos por Prueba no paramétrica de Wilcoxon exportados a Excel.

En la hoja de Excel puede observarse el resumen de datos para la prueba de Wilcoxon en muestras dependientes. Para lo anterior, se realizó un preprocesamiento de los datos, es decir, aquellos pacientes que aparecieran en un grupo y en el otro grupo no (prueba realizada para dos muestras), era eliminado manualmente con el fin de que las dimensiones de las muestras coincidieran tanto en fila como en columnas, ya que es uno de los requisitos previos para realizar el test probabilístico.

Lo anterior puede evidenciarse por medio de la tabla 7.

En el punto A (fila 1 y 2, hasta que termine número de columnas que tienen datos), se encuentra la masa (Mz) y el tiempo de retención (RT) de cada compuesto detectado. En la sección B, de la fila 4 a la fila 16, están los grupos a los cuales se les realizó la prueba.

De la fila 18 a la 19 (sección C), son los resultados de los cálculos realizados para hallar la probabilidad (valor en la fila 21, sección D) de rechazar la hipótesis nula, si la probabilidad es mayor que 0.05, se acepta la hipótesis nula y no se guarda en el Excel, en cambio si la probabilidad es menor que 0.05, se niega la hipótesis nula y se guarda en el Excel (esta condición también se cumple para la prueba de Mann-Whitney). Con esto, se tiene un número reducido de columnas para identificar los posibles compuestos característicos que están determinados por la enfermedad, minimizando el tiempo y trabajo en el desarrollo de los cálculos y análisis por columna cuando se encuentren con muestras muy grandes entre los grupos de estudios seleccionados.

Tabla 7. Comparación de número de datos antes y después de aplicar la prueba de Wilcoxon.

Grupos		Número de pacientes (inicial)	Número de pacientes después de preprocesamiento	Número de columnas (inicial)	Número de columnas luego de realizar la prueba
BREATH II	Control II	6	6	1834	153
	Bcg II	7	5	1834	125
	Michigan II	8	7	1834	143
BREATH I	Control I	7	6	1834	153
	Bcg I	5	5	1834	125
	Michigan I	7	7	1834	143
FECES III	Control III	7	3	1834	-
	Bcg III	7	6	1834	43
FECES IV	Control IV	3	3	1834	-
	Bcg IV	6	6	1834	43

Se presenta una visualización gráfica por medio del esquema cajas y bigotes para la representación de los datos obtenidos para la prueba de Wilcoxon, según la tabla 7:

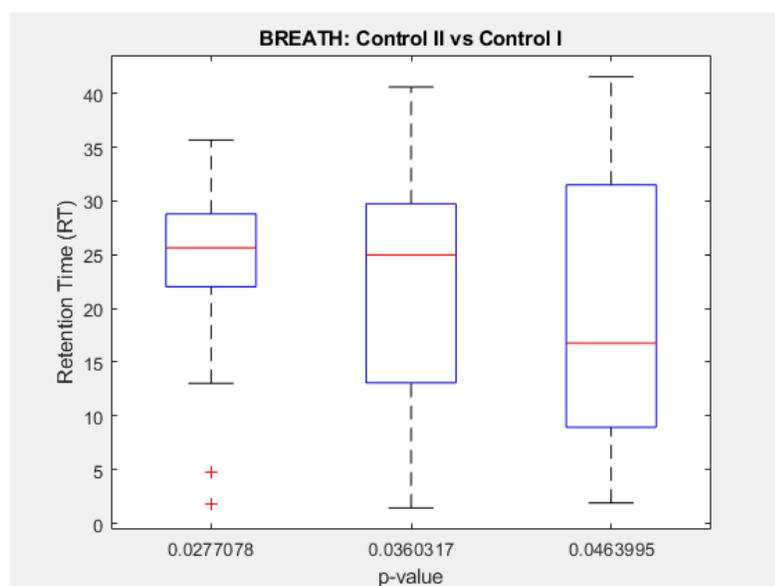


Figura 8. Boxplot Tiempo de Retención vs p-value, Wilcoxon Control II y Control I.

Tabla 8. Características de boxplot de la Figura 8 según p-value.

	p-value		
	0,027707849	0,03603169	0,04639946
Min	1,754404	1,4275384	1,9008286
1er Cuartil	22,006462	12,611667	8,416613
Mediana	25,639421	24,978384	16,785398
3er Cuartil	28,81983	29,778224	31,615901
Max	35,66482	40,59565	41,5565
Caja inferior	3,632959	12,366717	8,368785
Caja superior	3,180409	4,79984	14,830503
Bigote superior	6,84499	10,817426	9,940599
Bigote inferior	20,252058	11,1841286	6,5157844

En la figura 8 puede observarse en el eje de abscisas las diferentes probabilidades que corresponden a los dos grupos estudiados (Control II y Control I) en su totalidad de datos, y el eje de ordenadas está conformado por el tiempo de retención. En la tabla 8, se evidencian las características de la figura 8 por cada probabilidad que arrojaron los cálculos realizados por el software dentro del mismo grupo de estudio.

El recuadro de color azul o “caja” representa el recorrido intercuartil, es decir, el rango de valores encontrados entre el primer cuartil y el tercer cuartil, donde el primer cuartil

corresponde al 25% de los datos y el tercer cuartil indica el 75%. La línea roja ubicada en el interior de la caja hace referencia a la mediana, lo que es traducido en que la mitad de las observaciones es menor o igual que este valor (primer cuartil) y la otra mitad es mayor o igual que la mediana (tercer cuartil). Las líneas extendidas desde los cuartiles superior e inferior se denominan bigotes, estos marcan los datos cuyo valor numérico está por encima o por debajo del 75% y 25% respectivamente, pero que no llegan a ser valores 1.5 veces mayores ni valores 1.5 veces menores que el rango del cuartil. Los puntos o asteriscos (de color rojo) representados en la gráfica indican los valores que llegan a ser 1.5 veces más grandes o más pequeños que el rango intercuartil, estos puntos son denominados valores atípicos.

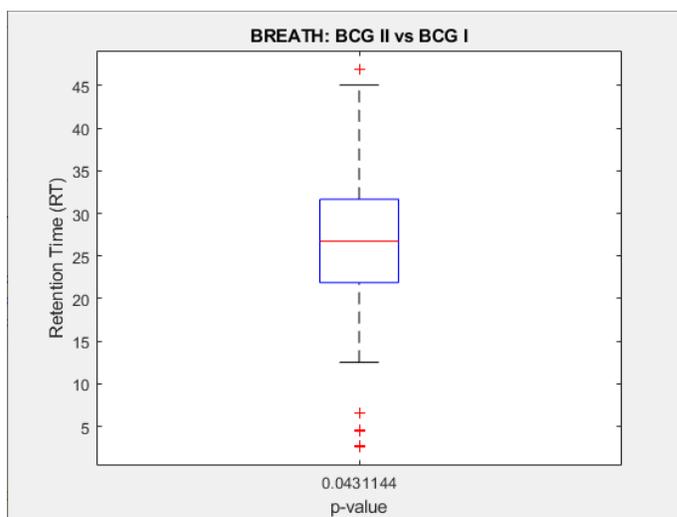


Figura 9. Boxplot Tiempo de Retención vs p-value, Wilcoxon BCG II y BCG I.

Tabla 9. Características de boxplot de la Figura 9 según p-value.

	<b>p-value</b>
	<b>0,04311445</b>
Min	2,607085
1er Cuartil	21,839298
Mediana	26,74317
3er Cuartil	31,719204
Max	46,894043
Caja inferior	4,903872
Caja superior	4,976034
Bigote superior	15,174839
Bigote inferior	19,232213

Teniendo en cuenta la figura 9, solo se presenta un valor de probabilidad de 0,04311445 dentro del conjunto de datos entre BCG II y BCG I en el eje de abscisas. En la tabla 9, se pueden observar los valores que hacen referencia a la construcción del gráfico de caja y bigotes representado en la figura 9. Se pueden observar valores atípicos tanto superiores e inferiores situados en los extremos de cada bigote. Los cuatros valores atípicos son respectivamente de menor a mayor: 2,607085-4,483109-6,5309715-46,894043, del eje de ordenadas donde se ve representado el tiempo de retención. La mediana de los datos se encuentra en el valor de 26,74317 (línea roja), con el 25% es decir, el primer cuartil en 21,839298 y el 75% (tercer cuartil) en 31,719204.

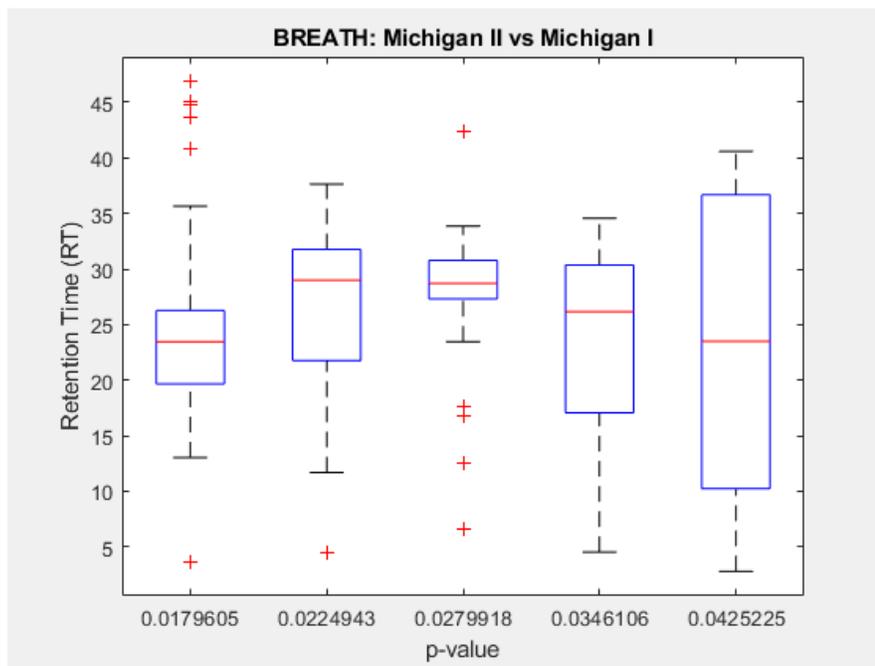


Figura 10. Boxplot Tiempo de Retención vs p-value, Wilcoxon Michigan II y Michigan I.

Tabla 10. Características de boxplot de la Figura 10 según p-value.

	p-value				
	<b>0,01796048</b>	<b>0,02249427</b>	<b>0,02799182</b>	<b>0,03461056</b>	<b>0,04252248</b>
Min	3,611707	4,507953	6,5309715	4,5328665	2,7901905
1er Cuartil	19,4180055	19,935192	27,305862	16,511362	8,38093763
Mediana	23,4578705	28,997585	28,70845	26,16657	23,504241
3er Cuartil	26,301983	31,799586	30,821999	30,6352825	37,6642053
Max	46,894043	37,644127	42,33882	34,55809	40,59565
Caja inferior	4,039865	9,062393	1,402588	9,655208	15,1233034
Caja superior	2,8441125	2,802001	2,113549	4,4687125	14,1599643
Bigote superior	20,59206	5,844541	11,516821	3,9228075	2,93144475
Bigote inferior	15,8062985	15,427239	20,7748905	11,9784955	5,59074713

Como se puede apreciar en la figura 10 consta de 5 valores de probabilidad, donde mediante la tabla 10 se puede evidenciar la mediana con un valor (de izquierda a derecha) de 23,4578705- 28,997585- 28,70845- 26,16657- 23,504241 respectivamente y los rangos intercuartiles de cada uno de estos.

Para la probabilidad de 0,01796048 sus valores atípicos (de arriba hacia abajo) son: 46,894043- 44,813198- 43,685806- 40,79933- 3,611707. En la probabilidad de 0,02249427 su valor atípico es: 4,507953. Para la probabilidad de 0,02799182 sus valores atípicos son: 42,33882- 23,469841- 17,660606- 16,785398- 12,523952- 6,5309715. Estos valores relacionados con los tiempos en el eje "y".

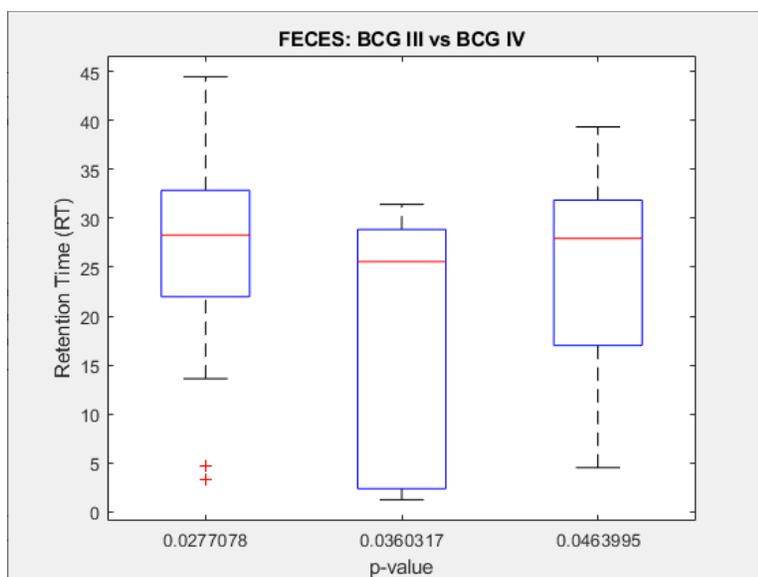


Figura 11. Boxplot Tiempo de Retención vs p-value, Wilcoxon BCG III y BCG IV.

Tabla 11. Características de boxplot de la Figura 11 según p-value.

	p-value		
	0,02770785	0,03603169	0,04639946
Min	3,365375	1,2422308	4,5181513
1er Cuartil	21,6219735	2,14541238	15,8458253
Mediana	28,288584	25,55717	27,937774
3er Cuartil	33,0153675	29,0369043	32,028752
Max	44,44808	31,415707	39,339546
Caja inferior	6,6666105	23,4117576	12,0919488
Caja superior	4,7267835	3,47973425	4,090978
Bigote superior	11,4327125	2,37880275	7,310794
Bigote inferior	18,2565985	0,90318158	11,327674

En la figura 11 solo se presentan valores atípicos en la probabilidad de 0,02770785, estos valores son de 4,659625 y 3,365375 respectivamente. Las medianas de cada caja (de izquierda a derecha) son: 28,288584- 25,55717- 27,937774 evidenciado en la figura correspondiente a la tabla 11.

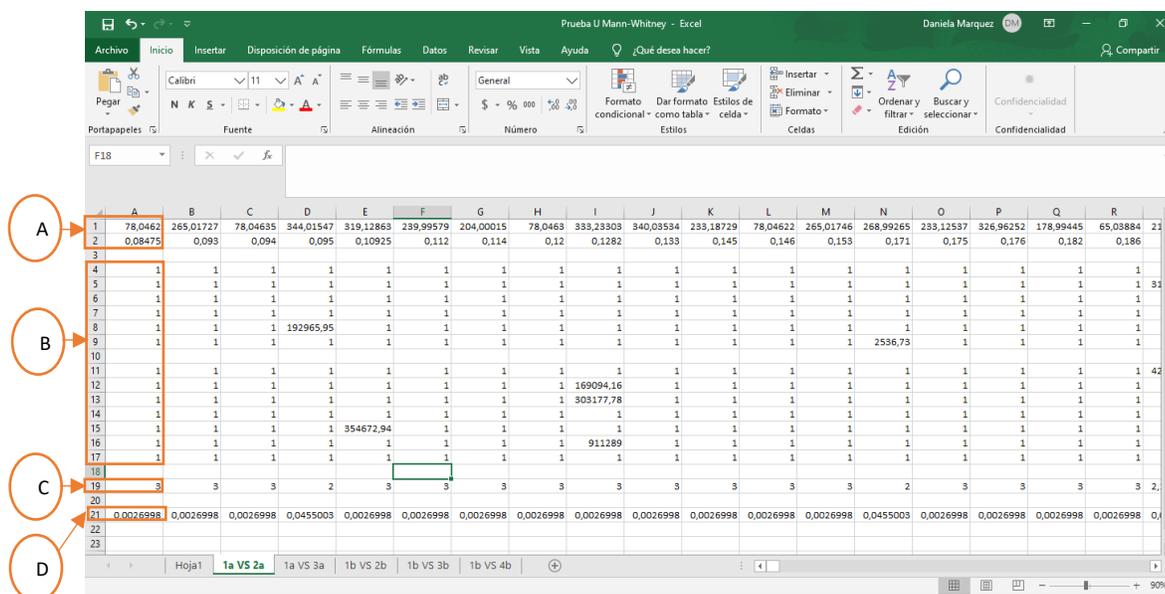


Figura 12. Resultados obtenidos por Prueba no paramétrica U Mann-Whitney exportados a Excel.

En la figura 12 se puede observar la prueba de U Mann-Whitney exportada en Excel. El punto A hace énfasis a las dos primeras filas las cuales contienen la masa del compuesto (Mz) y el tiempo de retención (RT) respectivamente. La distribución normal estándar está representada por z la cual hace referencia al punto C.

La probabilidad de que se rechace o acepte la hipótesis nula de la investigación está representada por la letra D, la cual es la fila 21 en Excel.

A continuación, se presenta una tabla con la cantidad de datos iniciales por grupo vs los datos después de haber realizado la prueba.

En la prueba U de Mann-Whitney (Muestras independientes), se tiene en cuenta lo siguiente:

Tabla 12. Comparación de número de datos antes y después de aplicar la prueba de U Mann-Whitney.

Grupos		Número de pacientes	Número de columnas (inicial)	Número de columnas luego de realizar la prueba
BREATH II	Control II	6	1834	Grupo principal, su valor cambia para cada subgrupo
	Bcg II	7	1834	1518
	Michigan II	8	1834	1532
BREATH I	Control I	7	1834	Grupo principal, su valor cambia para cada subgrupo
	Bcg I	5	1834	1479
	Michigan I	7	1834	1453
	Spanish I	5	1834	1484
FECES III	Control III	7	1834	Grupo principal, su valor cambia para cada subgrupo
	Bcg III	7	1834	1518
FECES IV	Control IV	3	1834	Grupo principal, su valor cambia para cada subgrupo
	Bcg IV	6	1834	1479
	Michigan IV	4	1834	1453
	Spanish IV	5	1834	1484

A través del diagrama cajas y bigotes se hace la representación gráfica de la tabla 12 de los resultados obtenidos para la prueba de U Mann-Whitney:

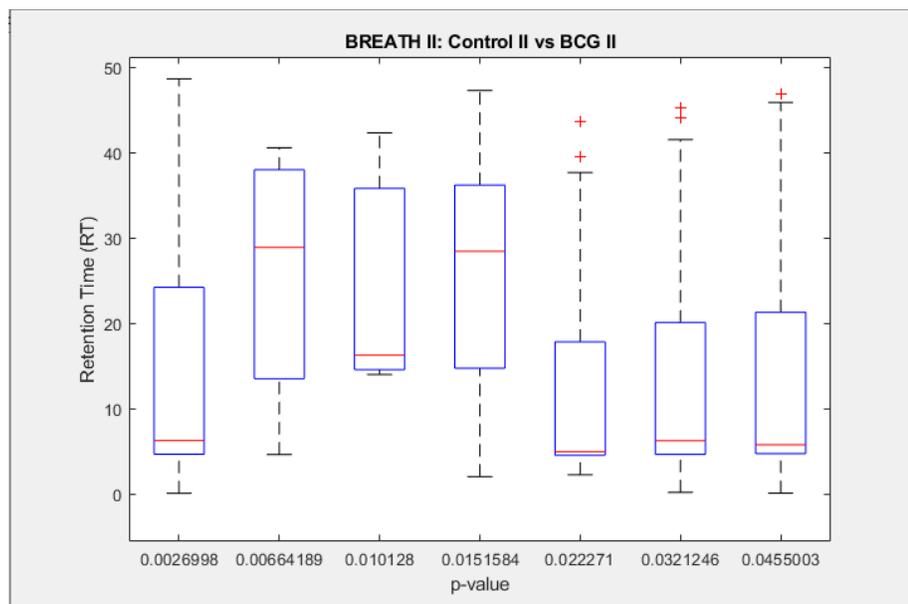


Figura 13. Boxplot Tiempo de Retención vs p-value, U Mann-Whitney Control II y BCG II.

Tabla 13. Características de boxplot de la Figura 13 según p-value.

	p-value						
	0,0026998	0,00664189	0,01012799	0,01515844	0,02227098	0,03212457	0,04550026
Min	0,08475	4,626677	14,003201	2,03185	2,2557142	0,2	0,095
1er Cuartil	4,6679165	9,06673225	14,003201	9,93002795	4,55391783	4,6335	4,73625025
Mediana	6,2705	28,909662	16,280834	28,447277	4,950889	6,2396364	5,7719
3er Cuartil	24,239	39,304844	42,33882	36,2522528	18,545554	20,272972	21,309375
Max	48,663	40,59565	42,33882	47,310074	43,685806	45,316498	46,893
Caja inferior	1,6025835	19,8429298	2,277633	18,5172491	0,39697118	1,6061364	1,03564975
Caja superior	17,9685	10,395182	26,057986	7,80497575	13,594665	14,0333356	15,537475
Bigote superior	24,424	1,290806	0	11,0578213	25,140252	25,043526	25,583625
Bigote inferior	4,5831665	4,44005525	0	7,89817795	2,29820363	4,4335	4,64125025

La mediana de estas 7 probabilidades son 6,2705- 28,909662- 16,280834- 28,447277- 4,950889- 6,2396364- 5,7719 respectivamente de izquierda a derecha de acuerdo las cajas.

Para la probabilidad de 0,02227098 sus valores atípicos de arriba hacia abajo son: 43,685806- 39,474854 respectivamente. En la probabilidad de 0,03212457 los valores atípicos representados en la figura 13 son: 45,316498- 44,1095. La última probabilidad con un valor de 0,04550026 posee un valor atípico cercano al extremo superior del bigote que se extiende desde el primer cuartil con un valor de 46,893.

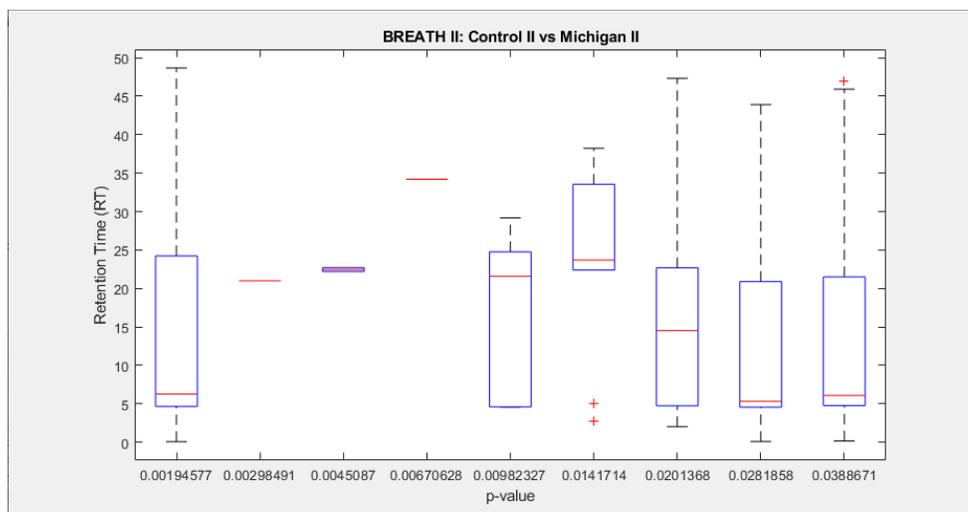


Figura 14. Boxplot Tiempo de Retención vs p-value, U Mann-Whitney Control II y Michigan II.

Tabla 14. Características de boxplot de la Figura 14 según p-value.

	p-value								
	0,0019457 7	0,00298491 4	0,004508 7	0,0067062 8	0,0098232 7	0,0141713 9	0,0201367 5	0,028185 8	0,038867 1
Min	0,08475	21,000963	22,21043	34,203	4,5667624	2,7177143	2,03185	0,095	0,171
1er Cuartil	4,6679165	21,000963	22,21043	34,203	4,597723	18,067395 7	4,7242002	4,5477	4,781
Mediana	6,289	21,000963	22,46032 7	34,203	21,588541	23,699262 5	14,515909	5,329882	6,086
3er Cuartil	24,239	21,000963	22,71022 4	34,203	26,23565	33,801508	22,902998	21,10233 3	21,554
Max	48,663	21,000963	22,71022 4	34,203	29,190998	38,21743	47,310074	43,8982	46,893
Caja inferior	1,6210835	0	0,249897	0	16,990818	5,6318668 5	9,7917088	0,782182	1,305
Caja superior	17,95	0	0,249897	0	4,647109	10,102245 5	8,387089	15,77245 1	15,468
Bigote superior	24,424	0	0	0	2,955348	4,415922	24,407076	22,79586 7	25,339
Bigote inferior	4,5831665	0	0	0	0,0309606	15,349681 4	2,6923502	4,4527	4,61

Se puede observar en la figura 14 específicamente en las probabilidades de 0,002984914 y 0,00670628 una sola línea roja, esto se debe a que solo existe un dato de tiempo para estas probabilidades, por lo que tanto la mediana, como los cuartiles y los rangos máximos y mínimos son iguales entre sí. Por lo que la caja inferior, superior y los bigotes superior e inferior son 0.

En la probabilidad con valor de 0,01417139 existen dos valores atípicos después del bigote superior, sus valores son: 5,1088886- 2,7177143. La probabilidad 0,0388671 posee un valor atípico por encima del bigote superior con un valor de: 46,893.

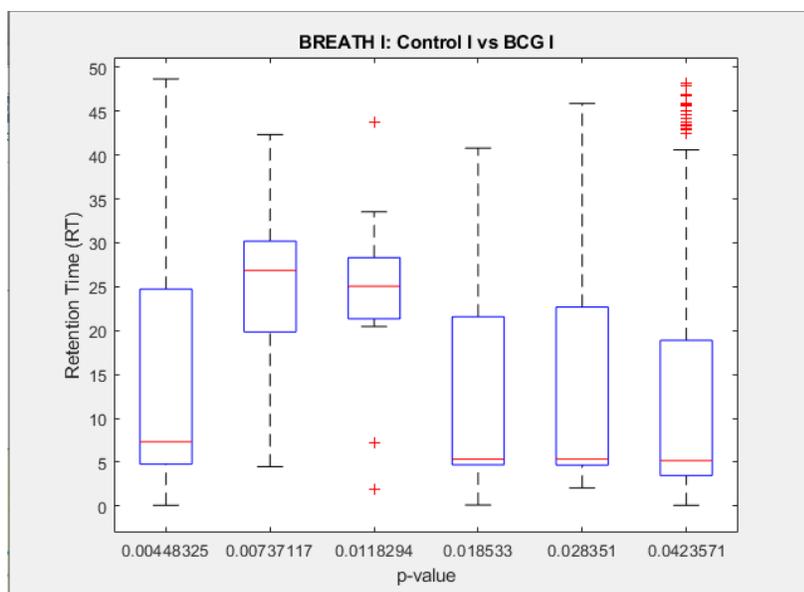


Figura 15. Boxplot Tiempo de Retención vs p-value, U Mann-Whitney Control I y BCG I.

Tabla 15. Características de boxplot de la Figura 15 según p-value.

	p-value					
	<b>0,00448325</b>	<b>0,00737117</b>	<b>0,01182938</b>	<b>0,01853297</b>	<b>0,02835104</b>	<b>0,04235706</b>
Min	0,093	4,487927	1,9008286	0,1174737	2,0662856	0,08475
1er Cuartil	4,802	18,74707	20,8966415	4,68856038	4,65891085	3,4728
Mediana	7,318667	26,8511795	25,0486175	5,35335565	5,3729997	5,1656666
3er Cuartil	24,7495	30,420856	29,0935058	40,79933	23,9012845	19,019125
Max	48,663	42,33882	43,685806	40,79933	45,8685	48,226
Caja inferior	2,516667	8,1041095	4,151976	0,66479527	0,71408885	1,6928666
Caja superior	17,430833	3,5696765	4,04488825	35,4459744	18,5282848	13,8534584
Bigote superior	23,9135	11,917964	14,5923003	0	21,9672155	29,206875
Bigote inferior	4,709	14,259143	18,9958129	4,57108668	2,59262525	3,38805

En la figura 15 puede observarse la presencia de seis valores de probabilidades diferentes, las cuales fueron encontradas luego de que el software realizara los cálculos correspondientes al modelo estadístico U Mann-Whitney.

De la probabilidad 0,01182938 pueden evidenciarse tres valores atípicos, los cuales 2 están muy por debajo del bigote inferior y son 1,9008286- 7,2391877 respectivamente y uno está encima del bigote inferior con valor de 43,685806.

Respecto a la siguiente probabilidad 0,04235706 todos sus valores atípicos, los cuales son 1.5 veces más grandes que el rango intercuartil, estas ubicados encima del bigote superior con valores (de abajo hacia arriba) de 42,514- 43,741- 44,126- 44,548- 45,057667- 45,6575- 45,748- 45,891- 46,6855- 46,743- 46,747- 47,96034- 48,226 respectivamente.

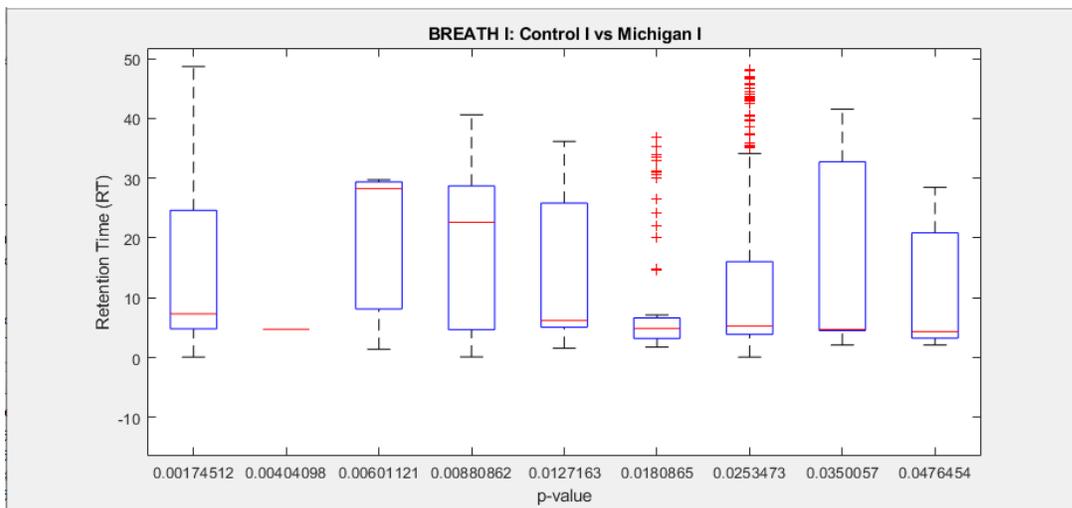


Figura 16. Boxplot Tiempo de Retención vs p-value, U Mann-Whitney Control I y Michigan I.

Tabla 16. Características de boxplot de la Figura 16 según p-value.

	p-value							
	0,00174512	0,00404098	0,00601121	0,00880862	0,01271625	0,01808646	0,02534732	0,04764543
Min	0,093	4,727728	1,4113334	0,1174737	1,577	1,7776667	0,08475	28,480692
1er Cuartil	4,800875	4,727728	1,4113334	4,67106538	5,07698105	3,11713335	3,6233535	28,480692
Mediana	7,30975	4,727728	28,252813	22,6257875	6,2028	4,8673705	5,262	28,480692
3er Cuartil	24,6262828	4,727728	29,743702	40,59565	26,4718565	6,79844665	17,4882315	28,480692
Max	48,663	4,727728	29,743702	40,59565	36,152706	36,91767	48,226	28,480692
Caja inferior	2,508875	0	26,8414796	17,9547221	1,12581895	1,75023715	1,6386465	0
Caja superior	17,3165328	0	1,490889	17,9698625	20,2690565	1,93107615	12,2262315	0
Bigote superior	24,0367173	0	0	0	9,6808495	30,1192234	30,7377685	0
Bigote inferior	4,707875	0	0	4,55359168	3,49998105	1,33946665	3,5386035	0

Como se puede evidenciar en la figura 16 en la probabilidad 0,00404098 solo existe una línea roja, lo que se traduce como que solo existe un dato de probabilidad en ese tiempo,

por lo tanto, variables como la mediana, primer y tercer cuartil y los rangos máximos y mínimos son iguales. En la probabilidad con valor 0,01808646 se encuentran los valores atípicos muy por encima del bigote superior con un total de 10 valores entre el tiempo de retención 14,712333 y 36,882748. Para el valor probabilístico de 0,02534732 sus valores atípicos se encuentran en el rango de tiempo de retención de 35,034 y 47,96034 con un total de 15 en ese intervalo.

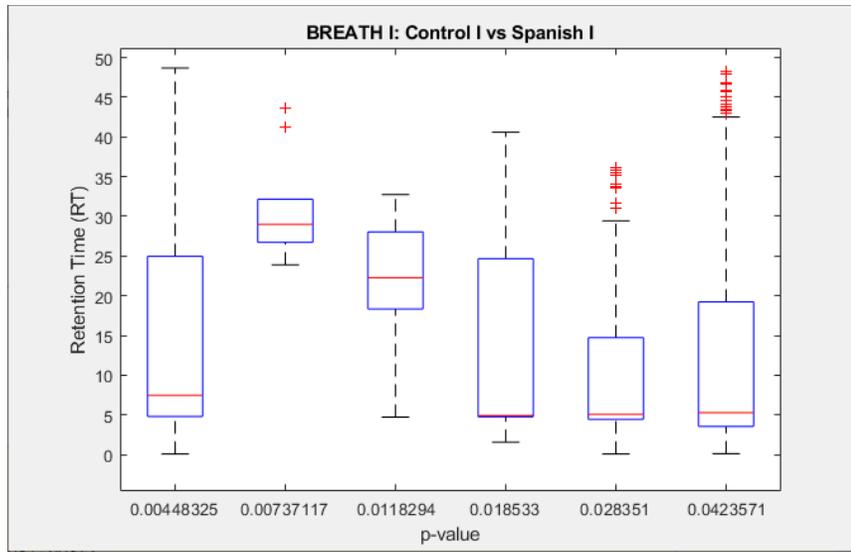


Figura 17. Boxplot Tiempo de Retención vs p-value, U Mann-Whitney Control I y Spanish I.

Tabla 17. Características de boxplot de la Figura 17 según p-value.

	p-value					
	0,00448325	0,00737117	0,01182938	0,01853297	0,02835104	0,04235706
Min	0,093	23,890175	4,727728	1,577	0,08475	0,1174737
1er Cuartil	4,81250025	26,5437458	18,000587	4,7564335	4,4086666	3,5484167
Mediana	7,47425	28,99397	22,2739065	4,9207098	5,0926666	5,2764997
3er Cuartil	24,979038	34,4222475	29,4289778	40,59565	14,730445	19,337166
Max	48,663	43,685806	32,737514	40,59565	36,152706	48,226
Caja inferior	2,66174975	2,45022425	4,2733195	0,1642763	0,684	1,728083
Caja superior	17,504788	5,4282775	7,15507125	35,6749402	9,6377784	14,0606663
Bigote superior	23,683962	9,2635585	3,30853625	0	21,422261	28,888834
Bigote inferior	4,71950025	2,65357075	13,272859	3,1794335	4,3239166	3,430943

Como se observa en la figura 17, las probabilidades 0.00448325, 0.01182938 y 0.01853297 no presentan valores atípicos, los datos que conforman el tiempo se encuentran entre el rango de los bigotes. Las probabilidades 0.00737117, 0.02835104 y 0.04235706, presentan valores atípicos, para la probabilidad 0.00737117 se encuentran dos valores atípicos de 41.16363 y 43.685806. Para la probabilidad de 0.02835104 presentan 9 valores atípicos entre el rango de tiempos 31,047695 y 36.152706. Por último, 0.04235706 posee 19 valores atípicos entre los tiempos 48.226 y 42.514.

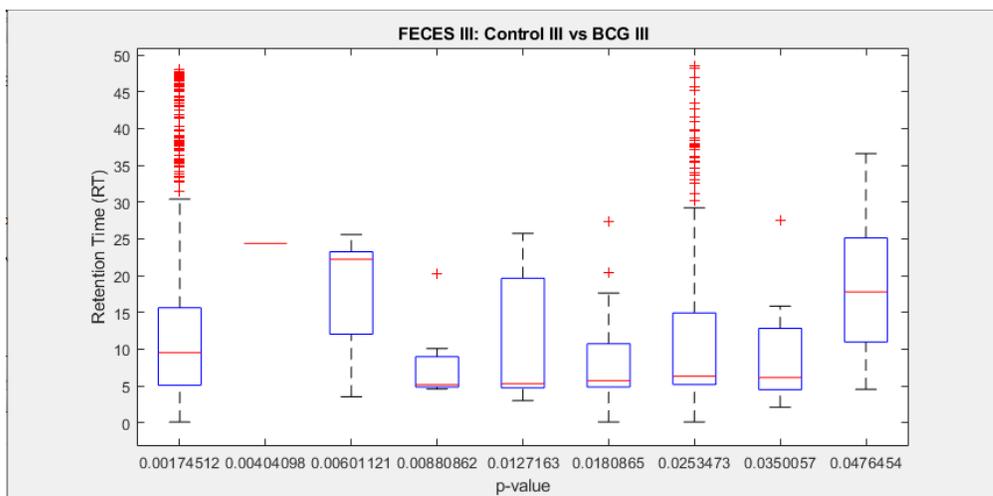


Figura 18. Boxplot Tiempo de Retención vs p-value, U Mann-Whitney Control III y BCG III.

Tabla 18. Características de boxplot de la Figura 18 según p-value.

	p-value								
	0,0017451 2	0,0040409 8	0,0060112 1	0,0088086 2	0,0127162 5	0,0180864 6	0,0253473 2	0,0350056 8	0,0476454 3
Min	0,087	24,377607	3,5301251	4,580625	3,0057998	0,0996666 6	0,09475	2,1039999	4,531818
1er Cuartil	5,0865	24,377607	9,1975625 5	4,8181305 5	4,7218423	4,85275	5,1954501	4,071066	9,4343241 3
Mediana	9,537125	24,377607	22,236092	5,1651252	5,302572	5,7	6,3220555 5	6,1281258 5	17,778007
3er Cuartil	15,65475	24,377607	24,029847 5	20,17775	19,835001	11,533514 5	14,92035	13,590957	25,144964 8
Max	48,087666	24,377607	25,583199	20,17775	25,751537	27,41061	48,513	27,57019	36,600273
Caja inferior	4,450625	0	13,038529 5	0,3469946 5	0,5807297	0,84725	1,1266054 5	2,0570598 5	8,3436828 8
Caja superior	6,117625	0	1,7937555	15,012624 8	14,532429	5,8335145	8,5982944 5	7,4628311 5	7,3669577 5
Bigote superior	32,432916	0	1,5533515	0	5,916536	15,877095 5	33,59265	13,979233	11,455308 3
Bigote inferior	4,9995	0	5,6674374 5	0,2375055 5	1,7160425	4,7530833 4	5,1007001	1,9670661	4,9025061 3

Teniendo en cuenta la figura 18, puede apreciarse 9 probabilidades, en la probabilidad 0,004040985 solo presenta un dato, por lo tanto, no se encuentran valores atípicos, y el rango intercuartil es igual a 0.

Las probabilidades 0.00601121, 0.01271625 y 0.04764543 no presentan valores atípicos.

En la probabilidad 0,00174512 se encuentran 84 valores atípicos desde el tiempo 32.74 hasta 48.087666.

Solo se presenta un valor atípico de 20,17775 en la probabilidad 0,00880862, en la probabilidad 0.03500568 también se presenta un único valor atípico en el tiempo 27.57019. La probabilidad 0,01808646 contiene dos valores atípicos, 20,487585 y 27,41061.

Por último, la probabilidad 0,02534732 presenta 29 valores atípicos, entre los tiempos 48.513 y 31.180666

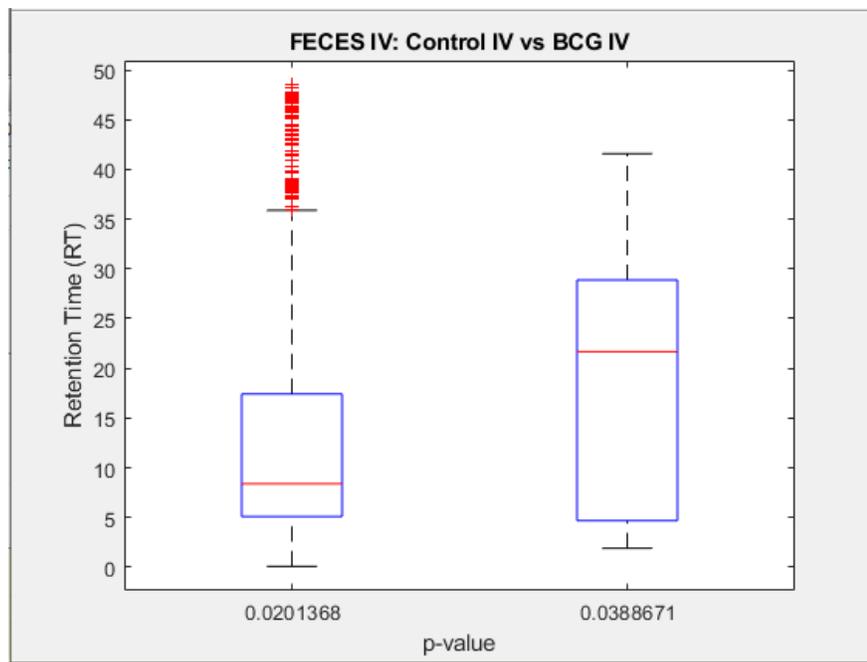


Figura 19. Boxplot Tiempo de Retención vs p-value, U Mann-Whitney Control IV y BCG IV.

Tabla 19. Características de boxplot de la Figura 19 según p-value.

	p-value	
	0,02013675	0,0388671
Min	0,087	1,9221892
1er Cuartil	5,1055	4,664786
Mediana	8,3935	21,683874
3er Cuartil	17,471899	29,308674
Max	48,513	41,580204
Caja inferior	3,288	17,019088
Caja superior	9,078399	7,6248
Bigote superior	31,041101	12,27153
Bigote inferior	5,0185	2,7425968

De acuerdo a la figura 19 en la probabilidad 0.0388671 no se encuentran valores atípicos, lo que significa que los datos de tiempo se encuentran entre el rango de los bigotes.

La probabilidad 0.02013675 tiene valores atípicos desde el tiempo 35.267 hasta el valor 48.513.

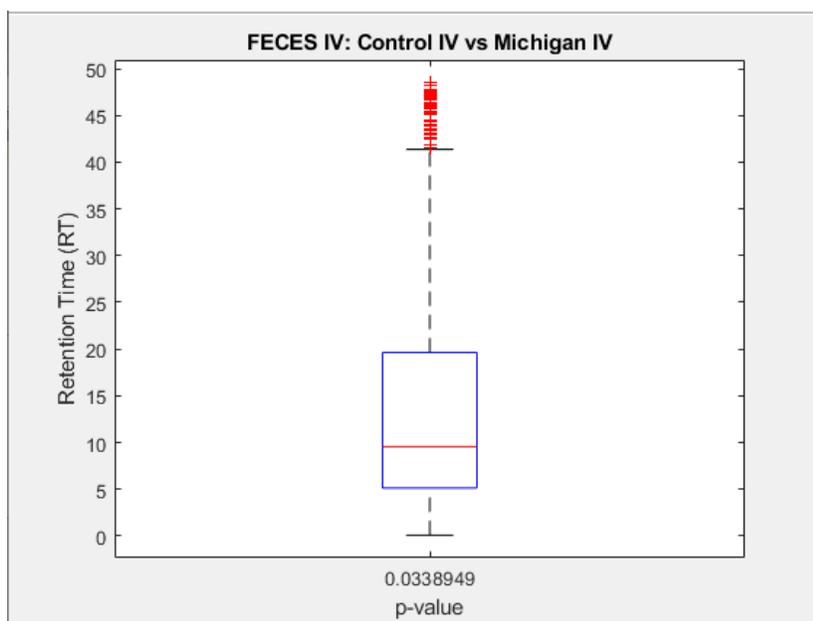


Figura 20. Boxplot Tiempo de Retención vs p-value, U Mann-Whitney Control IV y Michigan IV.

Tabla 20. Características de boxplot de la Figura 20 según p-value.

	<b>p-value</b>
	<b>0,03389485</b>
Min	0,087
1er Cuartil	5,13810008
Mediana	9,565375
3er Cuartil	19,6542503
Max	48,513
Caja inferior	4,42727493
Caja superior	10,0888753
Bigote superior	28,8587498
Bigote inferior	5,05110008

La estimación base de las ocasiones es 0.087 y su valor más extremo es 48.513.

El cuartil principal comienza con un valor de 5.13810008 y se extiende hasta llegar al medio que tiene una estimación de 9.565375. Entonces, el cuadro inferior se deja con una amplitud de 4.427274925.

El tercer cuartil comienza desde el medio y se extiende hasta el valor de 19.65425025, al restar estos valores da como resultado 10.08887525, lo que viene siendo la caja superior.

Los bigotes toman los valores menores al 25% de los datos y que no sean 1.5 veces menor que el rango de intercuartil o los valores mayores al 75% de los datos y no sean mayores al 1.5 del rango intercuartil.

El bigote superior tiene una adecuación de 28.85874975, y el bigote inferior que tiene 5.051100075.

Los valores atípicos a continuación son aquellos datos 1.5 de veces menores o mayores al rango intercuartil, estos son: 48,513- 48,145- 47,746-47,366- 46,977- 46,8905- 46,044- 45,887- 45,391- 44,456- 44,325- 43,821- 43,599575- 43,405- 42,463.

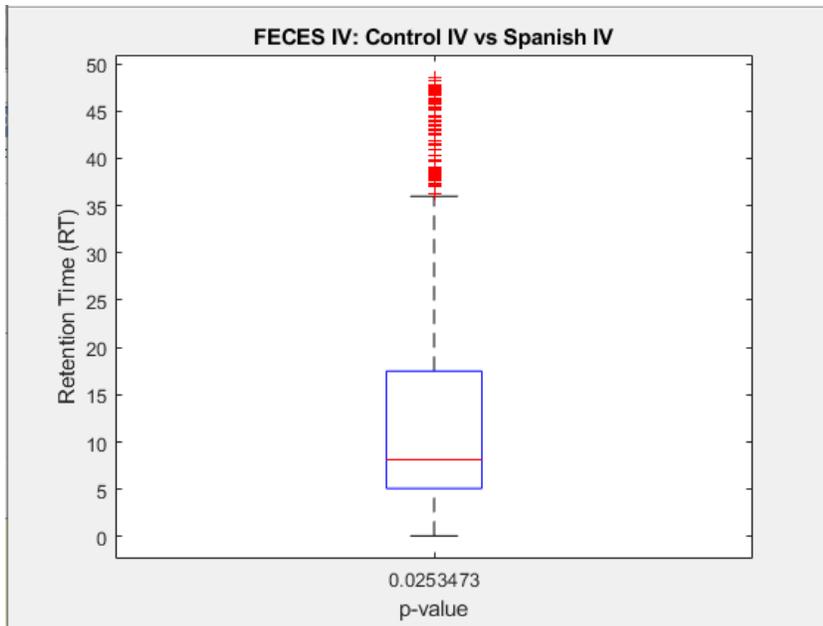


Figura 21. Boxplot Tiempo de Retención vs p-value, U Mann-Whitney Control IV y Spanish IV.

Tabla 21. Características de boxplot de la Figura 21 según p-value.

	<b>p-value</b>
	<b>0,02534732</b>
Min	0,087
1er Cuartil	5,10625
Mediana	8,155
3er Cuartil	17,5282495
Max	48,513
Caja inferior	3,04875
Caja superior	9,3732495
Bigote superior	30,9847505
Bigote inferior	5,01925

El valor mínimo de los tiempos es de 0,087 y su valor máximo es de 48,513

El primer cuartil comienza desde el valor 5,10625 y se extiende hasta llegar a la mediana que tiene un valor de 8,155. Por lo que la caja inferior queda con una amplitud de 3,0487 (valor de la resta entre la mediana y el primer cuartil)

El tercer cuartil se extiende hasta el valor 17,5282495 y comienza desde la mediana, el valor de la caja superior es la resta entre estos dos valores, lo que da como resultado 9,3732495.

Los bigotes comienzan desde donde terminan las cajas, y se extienden hasta alcanzar un valor 1.5 veces menor del rango del intercuartil en caso del bigote inferior, y en caso del bigote superior hasta 1.5 veces mayor.

Desde donde comienza el bigote superior hasta donde termina, tiene una amplitud de 30,9847505, en el caso del bigote inferior este tiene 5,01925.

Los valores atípicos de la figura 20, con la respectiva probabilidad que se puede observar, arriba hacia abajo son: 48,513-47,366-47,14-46,895-46,044-45,857-44,456-44,325-43,953-43,415-42,463-41,591-40,311-39,642998-38,62025-37,844334-36,293747-35,639835.

Se puede observar cómo se reducen las columnas en los grupos cuando se han realizado las pruebas, como también, la selección que hace internamente el programa en arrojar los datos que contengan la probabilidad menor o igual que 0.05, es decir, que rechace la hipótesis nula y aceptando la hipótesis alternativa, lo cual significa que encontró un potencial biomarcador dentro de las columnas que fueron exportadas a Excel con las especificaciones de las secciones mostradas en la Figura 7 y Figura 12.

Lo anterior es realizado en tiempo récord, ya que una vez ingresado los datos y elegido la prueba que se desea realizar, luego debe darse clic el botón calcular en el programa, y seguido de esto (sin límite de tiempo para cada acción) en el de generar archivo de Excel con la prueba que realizó. No existe límite de columnas para trabajar con el programa, como tampoco existe límite de pacientes, los cuales están representados por cada fila, es decir, no existe restricción de filas. Solo deben ingresarse los datos y el programa se encargará de entregar los resultados esperados.

Luego de obtener los resultados a través del software, donde se indica el valor del MZ y el RT correspondiente a la columna identificada como potencial biomarcador, se procedió a realizar la identificación del nombre del compuesto, esta última etapa se desarrolló mediante el uso del Software para análisis de datos obtenidos mediante cromatografía

de gases (GC-MS) cualitativos “MassHunter”, y en específico en los subprogramas de la herramienta del mismo (Qualitative y Unknown).

En la tabla 22, 23, 24 y 25 de la parte inferior se pueden apreciar algunos de los datos obtenidos e identificados mediante la herramienta desarrollada en el presente proyecto:

Tabla 22. Resultados obtenidos para la vacuna Control en grupo BREATH.

	TIPO DE VACUNA	COMPOUND	MZ @ RT	P-VALUE
Breath	CONTROL	Nonane, 2,6-dimethyl-	43.0541@14.756694	0,028
Breath	CONTROL	Carbonic acid, octyl vinyl ester	57.0697@16.545347	0,028
Breath	CONTROL	Cyclobutanone, 2,2,3-trimethyl-	70.07705@22.386898	0,028
Breath	CONTROL	3-Eicosene, (E)-	69.06969@26.334019	0,028
Breath	CONTROL	1-Heptanol, 2-propyl-	43.05472@28.60939	0,028

Tabla 23. Resultados obtenidos para la vacuna Michigan en grupo BREATH.

	TIPO DE VACUNA	COMPOUND	MZ @ RT	P-VALUE
Breath	MICHIGAN	Methane, trimethoxy-	75.02622@3.611707	0,018
Breath	MICHIGAN	Oxalic acid, isohexyl pentyl ester	43.05411@6.5309715	0,028
Breath	MICHIGAN	Carbonic acid, octyl vinyl ester	57.0697@16.545347	0,018
Breath	MICHIGAN	1-Undecyne	67.05323@18.657017	0,018
Breath	MICHIGAN	Decane, 2,4-dimethyl-	43.05472@27.374823	0,028

Tabla 24. Resultados obtenidos para la vacuna BCG en grupo BREATH.

	TIPO DE VACUNA	COMPOUND	MZ @ RT	P-VALUE
Breath	BCG	Disiloxane, hexamethyl-	147.0657@2.607085	0,043
Breath	BCG	Spiro[2.4]heptane, 1,5-dimethyl-6-methylene-	67.05318@15.439373	0,043
Breath	BCG	Cyclohexanecarboxylic acid, 4-pentyl-, 4-fluorophenyl ester	55.05427@27.305862	0,043
Breath	BCG	1-Undecene, 4-methyl-	43.05472@28.60939	0,043
Breath	BCG	1-Decanol, 2-ethyl-	57.0697@28.81983	0,043

Tabla 25. Resultados obtenidos para la vacuna BCG en grupo FECES.

	TIPO DE VACUNA	COMPOUND	MZ @ RT	P-VALUE
Feces	BCG	Butanoic acid, methyl ester	43.05459@3.365375	0,028
Feces	BCG	Cyclopentasiloxane, decamethyl-	355.07144@20.906445	0,028
Feces	BCG	1-Octanol, 2-butyl-	71.0853@22.337502	0,028
Feces	BCG	3-Hexadecene, (Z)-	55.05425@29.905725	0,028
Feces	BCG	Carbonic acid, undecyl vinyl ester	71.085@30.851112	0,046

## CONCLUSIONES Y RECOMENDACIONES

El software identificador desarrollado ofrece una solución alternativa para el análisis de los datos con respecto al análisis de datos obtenidos mediante pruebas no invasivas para detección de potenciales biomarcadores; al no requerir insertar un instrumento a través de la piel o en una abertura corporal para la comprobación de compuestos característicos de la enfermedad. El desarrollo del presente proyecto se fundamentó en el análisis de muestras de animales bovinos identificados como positivos y negativos de la enfermedad Tuberculosis Bovina o BTB. Extrayendo de cada uno de ellos muestras tales como Aliento y Fecales, con las cuales en contraste con la prueba de tuberculina, se logra una toma y análisis de datos sin el requerimiento de invadir y observar por largos periodos de tiempo el accionamiento de la tuberculina a la piel, al reducir el tiempo de respuesta de portación de la enfermedad se logran aportes significativos en el área de la salud con respecto a un mejor tiempo de respuesta contra la enfermedad al crear medicamentos en las primeras etapas de la enfermedad ya detectada.

El proceso de Cromatografía de Gases y Espectrometría de Masas (GC-MS) una vez realizado, da como resultado un archivo que puede ser leído por medio de los subprogramas Qualitative y Unknown del Software de MassHunter. En los resultados que, obtenidos luego de la prueba del proceso aplicado para la obtención de datos, se puede evidenciar una gráfica denominada cromatograma, en la que se observan diferentes picos correspondientes a la separación de la mezcla ubicado en el eje de abscisas el tiempo de retención en el que fue detectado cada compuesto y en el eje de ordenadas se observa un rango de las masas de todo el espectro cromatográfico. El cromatograma puede ser exportado en un archivo que puede ser leído por Excel, lo que se traduce en los datos que lo conforman como lo son masa y tiempo organizados en una matriz por cada muestra o paciente al que fue realizado el proceso de Cromatografía.

Una vez exportados los datos, se caracterizaron las diferentes pruebas estadísticas teniendo en cuenta que los datos obtenidos luego del proceso son muestras no paramétricas, es decir, no siguen una distribución en específico. Según lo anterior, por medio del Software que se creó, se comprobó que los datos no se ajustaran a una

distribución normal. Se realizaron dos pruebas de normalidad, Shapiro-Wilk y Kolmogórov-Smirnov siendo esta última más conservadora que la primera. Además, la distinción entre los grupos de BREATH y FECES y los subgrupos de cada uno, forman en totalidad dos subgrupos por cada grupo. Con esto en mente, se deduce que según las pruebas estadísticas existentes se pueden comparar dos promedios ya sea como muestras independientes o muestras relacionadas. Para cada caso existe un tipo de prueba estadística, en muestras independientes se utilizó la prueba U de Mann-Whitney y para las muestras relacionadas se usó Prueba de los rangos con signo de Wilcoxon, más conocida como Wilcoxon.

Conociendo las pruebas estadísticas que pueden ser aplicadas para la identificación de potenciales biomarcadores, se evaluó cada modelo caracterizando las variables presentes y haciendo los cálculos correspondientes a través del software Matlab para verificar los resultados y el trabajo que realiza de forma matricial para integrar la totalidad de datos que son analizados. Una vez realizado esto, se utilizó el recurso de diagramas de flujo para visualizar cómo sería el funcionamiento del software lo cual puede evidenciarse en las figuras 3 y 4 en la sección de presentación y análisis de resultados. El diseño del entorno para realizar las pruebas estadísticas partiendo desde las pruebas de normalidad se visualizan en las figuras 5 y 6. Los resultados obtenidos en las figuras 7 y 12, son los datos obtenidos luego de implementar el software con las pruebas mencionadas anteriormente, notándose una considerable disminución del volumen de datos.

La principal distinción entre nuestro software y otros es que, el nuestro es capaz de trabajar con una magnitud de datos notablemente voluminosa al mismo tiempo, automáticamente y con mayor rapidez, lo que significa que el usuario no tiene que identificar o separar las columnas una por una, lo que puede ser lento e improductivo cuando se trabaja con grandes cantidades de datos, en su lugar nuestro software lo hace automáticamente. Esto puede observarse en las tablas 7 y 12, en donde se muestra la comparación del número de columnas antes y después de realizar las pruebas estadísticas para los grupos Breath I, Breath II, Feces III y Feces IV, estos grupos presentaron 1834 columnas antes de realizar las pruebas estadísticas, al momento de

exportar los resultados, en lugar de presentarnos las probabilidades para las 1834 columnas, el software se encargará de seleccionar, juntar y exportar en Excel solamente aquellas columnas que no sigan la hipótesis nula, dejando las columnas identificadas como potencial biomarcador como se indica en la última columna de las tablas 7 y 12, siendo el número de columnas para la tabla 7 donde se realizó la prueba de Wilcoxon de entre 153 y 43 columnas, y en la tabla 12 realizando la prueba de Mann-Whitney, se encuentran entre 1532 y 1453. El por qué se configuró de esa forma el software, fue con el objetivo de no abrumar y facilitar el trabajo del usuario a la hora de analizar las columnas exportadas, al reducir las columnas mostrando aquellas que cumplen con su interés.

El usuario solo necesita indicar con cuántos grupos va a trabajar y la prueba estadística, luego generará Excel con solo las columnas que rechazan la hipótesis nula, y esa es otra diferencia que otros softwares no tienen. Adicionalmente, facilita la organización de los resultados, ya que los resultados por grupo se conservarán en el mismo libro de Excel para cada prueba no paramétrica.

Para la validación e identificación de compuestos según el tiempo de retención y la masa que fue identificada en ese tiempo específico, con el valor de la probabilidad calculada por el software creado y exportada en Excel se procedió a implementar el Software MassHunter, en el cual se obtienen los compuestos presentados en las tablas 22, 23, 24 y 25. Se cuenta con el software MassHunter gracias al proyecto de investigación en el marco europeo de quien dirige este proyecto.

## BIBLIOGRAFÍA

- [1] Gil, A. (2012). Tuberculosis Bovina: Enfermedad Reemergente en poblaciones bovinas de América. Presentation.
- [2] López Marín, L., Díaz Otero, F., Vallecillo Maza, A., Esquivel Solís, H., & Gutiérrez Pabello, J. (2006). Tuberculosis humana y bovina en Latinoamérica: De estudios sobre virulencia hacia herramientas para su control [Ebook] (48th ed., p. 1). Retrieved from <https://www.medigraphic.com/pdfs/lamicro/mi-2006/mi062r.pdf> Departamento de Salud y Servicios Humanos, GobiernoUSA.gov; 2012. Mycobacterium bovis (Tuberculosis bovina) en seres humanos
- [3] Abalos, P., & Retamal, P. (2004). Tuberculosis: ¿una zoonosis reemergente? *Revue Scientifique et Technique-Office International des Epizooties*, 23(2), 583-594.
- [4] Peled, N., Ionescu, R., Nol, P., Barash, O., McCollum, M., & VerCauteren, K. et al. (2012). Detection of volatile organic compounds in cattle naturally infected with *Mycobacterium bovis*. Retrieved 18 September 2019, from <https://www.sciencedirect.com/science/article/pii/S092540051200500X?via%3Dihub>
- [5] MILIÁN, F., SÁNCHEZ, L., TOLEDO, P., RAMÍREZ, C., & SANTILLÁN, M. (2000). Descriptive Study of Human and Bovine Tuberculosis in Querétaro, México [Ebook] (pp. 1-2). Querétaro: *Revista Latinoamericana de Microbiología*. Retrieved from <https://www.medigraphic.com/pdfs/lamicro/mi-2000/mi001c.pdf>
- [6] Pérez-Guerrero, L., Milián-Suazo, F., Arriaga-Díaz, C., Romero-Torres, C., & Escartín-Chávez, M. (2008). Epidemiología molecular de las tuberculosis bovina y humana en una zona endémica de Querétaro, México. Retrieved from [http://www.scielo.org.mx/scielo.php?pid=S0036-36342008000400006&script=sci\\_arttext&tlng=pt](http://www.scielo.org.mx/scielo.php?pid=S0036-36342008000400006&script=sci_arttext&tlng=pt)
- [7] Retamal, P. (2000). Tuberculosis bovina: Una breve actualización. Retrieved from <https://web.uchile.cl/vignette/monografiasveterinaria/monografiasveterinaria.uchile.cl/CD>

A/mon\_vet\_simple/0,1420,SCID%253D18380%2526ISID%253D452%2526PRT%253D18375,00.html

[8] Miramontes, R., Auld, S., Click, E., Heilig, C., Cain, K., Bisson, G., & Mac. Kenzie, W. (2013). Tuberculin Skin Test Result and Risk of Death among Persons with Active TB. Jyothi Rengarajan. Retrieved from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3823982/>

[9] Youssef, E., & Wooltorton, E. (2005). Serious allergic reactions following tuberculin skin tests. Retrieved from <http://www.cmaj.ca/content/173/1/34>

[10] Hu, Y. (2016). Regulatory Concern of Polymerase Chain Reaction (PCR) Carryover Contamination (pp. 2-4). Retrieved from <https://www.intechopen.com/books/polymerase-chain-reaction-for-biomedical-applications/regulatory-concern-of-polymerase-chain-reaction-pcr-carryover-contamination>

[11] Liébana, Criado, Ernesto. Nuevas estrategias en el diagnóstico y la epidemiología de las infecciones por Mycobacterium Bovis, Universidad Complutense de Madrid, 2005. ProQuest Ebook Central, <https://ebookcentral.proquest.com/lib/bibliotecaucsp/detail.action?docID=3160753>

[12] Oriani, D. S., Dubarry, J. R., Errea, A., Vera, O. A., Maria, A., Cavagión, L. J., & Staskevich, A. S. (2017). Asociación entre el diagnóstico de Tuberculosis Bovina por intradermorreacción, la anatomopatología, la bacteriología y la posible interferencia con micobacterias ambientales. Ciencia Veterinaria, 13(1), 42-47, Red Universidad Nacional de La Pampa, 2011, ProQuest Ebook Central, <https://cerac.unlpam.edu.ar/index.php/veterinaria/article/view/1859>

[13] Scott-Thomas, A., Epton, M., & Chambers, S. (2013). Validating a breath collection and analysis system for the new tuberculosis breath test. Journal of breath research, 7(3), 037108, <https://www.ncbi.nlm.nih.gov/pubmed/23867751>

[14] Guerrero, O. E. G., Acevedo, C. M. D., Gallardo, J. A. A., & Sandoval, J. E. O. (2013). Clasificación de compuestos químicos usando un sistema multisensorial (nariz

electrónica) desarrollada sobre un dispositivo hardware (fpga). Ciencia, innovación y tecnología, 1, 13-21. <https://www.jdc.edu.co/revistas/index.php/rciyt/article/view/119>

[15] Durán, C. M., Monsalve, P. A. G., & Mosquera, C. J. (2016, November). Multisensor system for toxic gases detection generated on indoor environments. In IOP Conference Series: Materials Science and Engineering (Vol. 157, No. 1, p. 012029). IOP Publishing. [https://www.researchgate.net/journal/1757899X\\_IOP\\_Conference\\_Series\\_Materials\\_Science\\_and\\_Engineering](https://www.researchgate.net/journal/1757899X_IOP_Conference_Series_Materials_Science_and_Engineering)

[16] Maya, G. C. (2011). Las pruebas de aliento: un nuevo horizonte en el laboratorio clínico. Medicina & Laboratorio, 17(01-02), 7-9. <https://www.medigraphic.com/pdfs/medlab/myl-2011/myl1111-2a.pdf>

[17] Durán Acevedo, Cristhian Manuel, Velásquez Carvajal, Adriana Eugenia, & Gualdron Guerrero, Oscar Eduardo. (2012). Implementación de una nariz electrónica para detectar pacientes con EPOC desde el aliento exhalado. Ingeniería y Desarrollo, 30(2), 143-159. [http://www.scielo.org.co/scielo.php?pid=S012234612012000200002&script=sci\\_abstract&tIng=es](http://www.scielo.org.co/scielo.php?pid=S012234612012000200002&script=sci_abstract&tIng=es)

[18] Artificial olfactory system based on nanomaterials-based sensors for early detection of tuberculosis via breath test Morad Nakhleh, Raneen Jeries, Alaa Gharra, Anke Binder, Yoav Broza, Mellisa Pascoe, Keertan Dheda, Hossam Haick European Respiratory Journal Sep 2013, 42 (Suppl 57) P498. [https://erj.ersjournals.com/content/42/Suppl\\_57/P498](https://erj.ersjournals.com/content/42/Suppl_57/P498)

[19] Amal, H., Leja, M., Funka, K., Skapars, R., Sivins, A., Ancans, G., ... & Haick, H. (2016). Detection of precancerous gastric lesions and gastric cancer through exhaled breath. Gut, 65(3), 400-407. <https://www.ncbi.nlm.nih.gov/pubmed/25869737>

[20] Guo, D., Zhang, D., Li, N., Zhang, L., & Yang, J. (2010). A novel breath analysis system based on electronic olfaction. IEEE transactions on biomedical engineering, 57(11), 2753-2763. <https://www.ncbi.nlm.nih.gov/pubmed/20667805>.

[21] P. Fernandez, W. White; Ed.: 2011. Atlas de Enfermedades Animales Transfronterizas

[22] Departamento de Salud y Servicios Humanos, GobiernoUSA.gov; 2012. Mycobacterium bovis (Tuberculosis bovina) en seres humanos

[23] Bermejo, M. C., Clavera, I., Michel de la Rosa, F. J., & Marín, B. (2007). Epidemiología de la tuberculosis. In *Anales del sistema sanitario de Navarra* (Vol. 30, pp. 07-19). Gobierno de Navarra. Departamento de Salud.

[24] Buddle B., McCarthy A., Ryan T., Pollock J., Vordermeier H., Hewinson R., Andersen P. & de Lisle G. (2003). – Use of mycobacterial peptides and recombinant proteins for the diagnosis of bovine tuberculosis in skin test-positive cattle. *Vet. Rec.*, 153, 615-620.

[25] Kamerbeek J., Schouls L., Kolk A., Van Agtervaeld M., Van Soolingen D., Kuijper S., Bunschoten A., Molhuizen H., Shaw R., Goyal M. & Embden J. (1997). – Simultaneous detection and strain differentiation of Mycobacterium tuberculosis for diagnosis and epidemiology. *J. clin. Microbiol.*, 35, 907-914.

[26] Instituto Colombiano Agropecuario, Tuberculosis Bovina. Recuperado de: [https://www.ica.gov.co/getdoc/37fff3e7-2414-4129-a104-06f55f7f6c63/tuberculosis-bovina-\(1\).aspx](https://www.ica.gov.co/getdoc/37fff3e7-2414-4129-a104-06f55f7f6c63/tuberculosis-bovina-(1).aspx)

[27] Pérez-Guerrero, L., Milián-Suazo, F., Arriaga-Díaz, C., Romero-Torres, C., & Escartín-Chávez, M. (2008). Epidemiología molecular de las tuberculosis bovina y humana en una zona endémica de Querétaro, México. *salud pública de México*, 50(4), 286-291

[28] M.C. Gutiérrez, M. Droguet., LA CROMATOGRAFÍA DE GASES Y LA ESPECTROMETRÍA DE MASAS: IDENTIFICACIÓN DE COMPUESTOS CAUSANTES DE MAL OLOR. IDENTIFICACIÓN DE COMPUESTOS VOLÁTILES POR CG-MS. Recuperado de: <https://core.ac.uk/download/pdf/41780740.pdf>

[29] Oscar García Barradas, Universidad Veracruzana. (2019) Cromatografía de Gases/Espectrometría de Masas (GC/MS). Recuperado de: <https://www.uv.mx/sara/facilidades/gcms/>

[30] Eric Flores-Ruiz, María Guadalupe Miranda-Novales, Miguel Ángel Villasís-Keever, El protocolo de investigación VI: cómo elegir la prueba estadística adecuada. Estadística inferencial. Recuperado de: <http://www.scielo.org.mx/pdf/ram/v64n3/2448-9190-ram-64-03-0364.pdf>

[31] XLSTAT. ¿Cuál es la diferencia entre pruebas paramétricas y no paramétricas? Recuperado de: <https://help.xlstat.com/s/article/cul-es-la-diferencia-entre-pruebas-parametricas-y-no-parametricas?language=es>

[32] Arango, S. S. (2012). Biomarcadores para la evaluación de riesgo en la salud humana. Revista Facultad Nacional de Salud Pública, 30(1), 75-82.

[33] Albertini RJ. Biomarker responses in human populations: towards a worldwide map. Mutat. Res. 1999; 428(1-2): 217-226.

[34] Schulte PA, Hauser JE. The use of biomarkers in occupational health research, practice, and policy. Toxicology Letters. 2011 Mar.

[35] World Health Organization. Biomarkers and risk assessment: concepts and principles Environmental health criteria; N° 155I CPS 1993 [internet]. Geneva: who/ipcs; 1993 [acceso 15 de junio de 2010]. Disponible en: <http://www.inchem.org/documents/ehc/ehc/ehc155.htm>

[36] Owen R, Depledge M, Hagger J, Jones M, Galloway T. Biomarkers and environmental risk assessment: guiding principles from the human health field. Marine Pollution Bulletin. 2008; 56(1): 613-619.

[37] Schulte P. Opportunities for the development and use of biomarkers. Toxicol Lett. 1995; 77(1-3): 25-29.

- [38] Berlanga, V., & Rubio Hurtado, M. J. (2012). Clasificación de pruebas no paramétricas. Cómo aplicarlas en SPSS. *REIRE. Revista d'Innovació i Recerca en Educació*, 2012, vol. 5, num. 2, p. 101-113.
- [39] Dorronsoro, I., & Torroba, L. (2007). Microbiología de la tuberculosis. *Anales del Sistema Sanitario de Navarra*, 30(Supl. 2), 67-85. Recuperado en 30 de octubre de 2019, de [http://scielo.isciii.es/scielo.php?script=sci\\_arttext&pid=S1137-66272007000400006&lng=es&tlng=es](http://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S1137-66272007000400006&lng=es&tlng=es).
- [40] OIE - World Organisation for Animal Health (Bovine Tuberculosis. (Paris, 2014,). [ebook] Health, pp.3
- [41] ¿Qué es la Cromatografía?» TP - Laboratorio Químico", TP - Laboratorio Químico, 2019. [Online]. Available: <https://www.tplaboratorioquimico.com/laboratorio-quimico/procedimientos-basicos-de-laboratorio/que-es-la-cromatografia.html>
- [42] M. Rouse, "¿Qué es Análisis estadístico? - Definición en WhatIs.com", SearchDataCenter&nbsp;en&nbsp;Español. [Online]. Available: <https://searchdatacenter.techtarget.com/es/definicion/Analisis-estadistico>. [Accessed: 30- Oct- 2019].
- [43] Análisis de Datos Estadístico: Tipos de Datos y Medidas", *Tecnologias-informacion.com*, 2019. [Online]. Available: <https://www.tecnologias-informacion.com/analisis-estadistico.html>. [Accessed: 30- Oct- 2019].
- [44] S. Maths, "Media aritmética", *Sangakoo.com*, 2018. [Online]. Available: <https://www.sangakoo.com/es/temas/media-aritmetica>. [Accessed: 30- Oct- 2019].
- [45] ¿Qué es la desviación estándar? - Minitab", *Support.minitab.com*, 2019. [Online]. Available: <https://support.minitab.com/es-mx/minitab/18/help-and-how-to/statistics/basic-statistics/supporting-topics/data-concepts/what-is-the-standard-deviation/>. [Accessed: 30- Oct- 2019].

[46] O. ROSALES ROBLES, *Academia.edu*, 2019. [Online]. Available: [https://www.academia.edu/39490055/Qu%C3%A9\\_es\\_la\\_distribuci%C3%B3n\\_normal](https://www.academia.edu/39490055/Qu%C3%A9_es_la_distribuci%C3%B3n_normal). [Accessed: 30- Oct- 2019].

[47]"¿Qué son las pruebas paramétricas?", QuestionPro. [Online]. Available: <https://www.questionpro.com/blog/es/pruebas-parametricas/>.

[48]"Explicación de los métodos no paramétricos - Minitab", Support.minitab.com, 2018. [Online]. Available: <https://support.minitab.com/es-mx/minitab/18/help-and-how-to/statistics/nonparametrics/supporting-topics/understanding-nonparametric-methods/>. [Accessed: 24- Nov- 2019].

[49] J. López, "Varianza - Definición, qué es y concepto | Economipedia", Economipedia, 2013. [Online]. Available: <https://economipedia.com/definiciones/varianza.html>.

[50]"Variables estadísticas", Universo Formulas, 2015. [Online]. Available: <https://www.universoformulas.com/estadistica/descriptiva/variables-estadisticas/>.

[51]J. Amat Rodrigo, "RPubs - Prueba de los rangos con signo de Wilcoxon alternativa no paramétrica al t-test de medias dependientes", Rpubs.com, 2016. [Online]. Available: [https://rpubs.com/Joaquin\\_AR/218464](https://rpubs.com/Joaquin_AR/218464).

[52]M. CORREA TABORDA and S. VALLEJO FLOREZ, "La prueba U DE MANN-WHITNEY", 2010.

[53]J. López, "Diferencia entre estadística paramétrica y no paramétrica - Definición, qué es y concepto | Economipedia", Economipedia. [Online]. Available: <https://economipedia.com/definiciones/diferencia-entre-estadistica-parametrica-y-no-parametrica.html>.

[54] Gobierno de España. Sanidad animal e higiene ganadera- Enfermedades-Tuberculosis. Recuperado de: [https://www.mapa.gob.es/es/ganaderia/temas/sanidad-animal-higiene-ganadera/sanidad-animal/enfermedades/tuberculosis/Tuberculosis\\_bovina.aspx](https://www.mapa.gob.es/es/ganaderia/temas/sanidad-animal-higiene-ganadera/sanidad-animal/enfermedades/tuberculosis/Tuberculosis_bovina.aspx)

[55] ICA. Instituto Colombiano Agropecuario. Resolución No. 00017463 (15/12/3017) Recuperado de: <https://www.ica.gov.co/getattachment/d5d75826-23c4-4b35-8620-37acb8df3d5c/2017R17463.aspx>

[56] EUR-Lex, Directiva 78/52/CEE del Consejo, de 13 de diciembre de 1977, por la que se establecen los criterios comunitarios aplicables a los planes nacionales de erradicación acelerada de la brucelosis, de la tuberculosis y la leucosis enzoótica de los bovinos. *Spanish special edition: Chapter 03 Volume 013 P. 175 – 182*. Recuperado de: <https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=CELEX%3A31978L0052>

[57] OIE. Organización Mundial de Sanidad Animal. MANUAL DE LAS PRUEBAS DE DIAGNÓSTICO Y DE LAS VACUNAS PARA LOS ANIMALES TERRESTRES (mamíferos, aves y abejas) Volumen I (2004). Recuperado de: <https://www.oie.int/doc/ged/d6508.pdf>

[58] Ministerio de Agricultura y Desarrollo Rural, (Última modificación 13/08/2019). CAPÍTULO 4. Tuberculosis Bovina. Recuperado de: <https://www.minagricultura.gov.co/Normatividad/Paginas/Decreto-1071-2015/CAPITULO-4-Tuberculosis-Bovina.aspx>

[59] Riba, N., & Calvo, G. (2015). Pruebas invasivas en investigación clínica: ¿suponen siempre un incremento del riesgo. *ICB digital.*, (Noviembre 2015), [consultado 23 Feb 2018]. Disponible en: <https://se-fc.org/gestor/images/icbdigital/93aarticulo.pdf>.

[60] "No invasivo: MedlinePlus enciclopedia médica", Medlineplus.gov. [Online]. Available: <https://medlineplus.gov/spanish/ency/article/002269.htm>.

[61] Gómez-Gómez, M., Danglot-Banck, C., & Vega-Franco, L. (2003). Sinopsis de pruebas estadísticas no paramétricas. Cuando usarlas. *Revista Mexicana de Pediatría*, 70(2), 91-99.

[62] Daniel WW. Biostatistics. A foundation for analysis in the health sciences. 7th ed. New York: John Wiley and Sons Inc, 1999: 658-736.

[63] Montgomery, D. C., & Runger, G. C. (2010): —Applied statistics and probability for engineersll. John Wiley & Sons.

[64] Segnini, S. (2010). *Prueba de normalidad de Shapiro-Wilk* [E-book] (pp. 1-2). Retrieved from [http://webdelprofesor.ula.ve/ciencias/segninis/Docencia/ANEXO\\_A\\_Sahapiro-Wilks.pdf](http://webdelprofesor.ula.ve/ciencias/segninis/Docencia/ANEXO_A_Sahapiro-Wilks.pdf).

[65] Prueba de normalidad - Minitab. (2019). Retrieved from <https://support.minitab.com/es-mx/minitab/18/help-and-how-to/statistics/basic-statistics/supporting-topics/normality/test-for-normality/>

[66] Mohd Razali, Nornadiah & Yap, Bee. (2011). Power Comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling Tests. *J. Stat. Model. Analytics*. 2.

## **ANEXOS**

### **Anexo 1**

#### **Manual de usuario**