



Propuesta metodológica para estimar niveles de mercurio en agua superficial

Luis Samuel Wallis Mosquera
Sandra Milena Gómez Gómez

Universidad del Valle sede Meléndez
Facultad de Ingeniería
Escuela de Estadística
Santiago de Cali
2020

Propuesta metodológica para estimar niveles de mercurio en agua superficial

Luis Samuel Wallis Mosquera
Sandra Milena Gómez Gómez

Trabajo de grado presentado como requisito parcial para optar al título de:
Estadístico

Director:

Jose Rafael Tovar

Universidad del Valle sede Meléndez

Facultad de Ingeniería, Escuela de Estadística

Santiago de Cali, Colombia

2020

Dedicatoria

Luis Samuel Wallis

A Dios por darme salud y sabiduría para culminar mi carrera satisfactoriamente.

A mis padres Sonia y Jhon, mis hermanas Linda Yulitza y Luisa Fernanda, a mi novia Ennit Valeria a estos por su apoyo, amor incondicional y sus múltiples enseñanzas. A mis tías y primos por creer en mí, en especial a María Doris quien siempre ha estado orgullosa de la persona que soy. A todos estos gracias por hacer parte de este logro.

Sandra Milena Gómez

Dedico este trabajo de grado a Dios por la Salud y la oportunidad, a toda mi familia y amigos, especialmente a mi madre y mi padrastro que han sido un pilar fundamental en mi formación como profesional, por brindarme la confianza, consejos, oportunidad y recursos para lograrlo, a aquellas personas que pasaron por mi vida en algún momento y que me brindaron su cariño, su paciencia, apoyo y comprensión y por último a esos verdaderos amigos con los que hemos compartimos años juntos, a mi compañero de tesis que ha estado conmigo desde el colegio y un orgullo presentar este trabajo en conjunto con él.

Agradecimientos

Agradecemos a nuestros docentes de la Universidad del Valle por haber compartido sus conocimientos a lo largo de la preparación de nuestra profesión, de manera especial, al profesor José Rafael Tovar Director de nuestro proyecto de grado quien ha guiado con su paciencia, y su rectitud como docente, así mismo a mi compañero(a) por su paciencia, compromiso y dedicación. Y a los habitantes de la comunidad de Yolombó, corregimiento La Toma, Suarez (Cauca) por su valioso aporte para nuestro trabajo de grado.

Resumen

El presente trabajo de grado propone una metodología estadística, que permita clasificar niveles de mercurio en aguas superficiales, mediante el uso de cualquier método de tirillas de detección y dos métodos de análisis de laboratorio. Dado que se contó con una muestra de observaciones reducida debido a las limitaciones logísticas y económicas del proyecto de investigación que dio origen a las mediciones, fue necesario desarrollar una metodología de análisis de datos enmarcada dentro del paradigma Bayesiano de la Estadística. Para ello, se propuso la distribución Multinomial para clasificar los diferentes niveles de mercurio (Hg), que pudiera contener las muestras de agua, y como distribución a priori se utilizó la distribución Dirichlet, cuyos hiperparámetros se obtuvieron mediante el método empírico de Bayes, de esta manera los resultados indicaron que las tirillas (Sensafe) pueden ser un instrumento de alerta temprana para detectar la contaminación de mercurio (Hg) en las fuentes hídricas cuando se toma una cantidad de muestra considerable, y el método de colorimetría para los diferentes niveles de mercurio no fueron estadísticamente significativa. Mientras, el método de espectrofotometría de vapor frío que es una prueba patrón de oro, evidencio que existe una cantidad de mercurio superior a lo permitido en la resolución 0883 de 2018, en dos de las 10 muestras analizadas con este método.

Palabras clave: Mercurio iónico (Hg), Tirilla (Sensafe), Colorimetría, Espectrofotometría, Información a priori, Métodos Empíricos de Bayes, distribución multinomial.

Abstract

This degree work proposes a statistical methodology that allows classifying mercury levels in superficial waters, this is achieved through the use of any detection strips method and two laboratory analysis methods. Due to the reduced sample observations that derive from the logistic and economic limitations of the project that gave rise to the measurements, it was necessary to develop a data analysis methodology framed with the Bayesian paradigm. For this, the multinomial distribution was proposed to classify different mercury levels (Hg) that the water samples can contain, and as a priory, the Dirichlet distribution was used, which hyperparameters were obtained using the empirical Bayes method, in this way the results indicated that the Sensafe strips can be an early warning instrument to detect mercury (Hg) contamination in the water sources when a considerable sample amount is taken, and when the colorimetry method for the different mercury levels was not statistically significant. While the cold vapor spectrophotometry method, which is a gold standard test, evidence that there is a bigger amount of mercury than the one allowed in the resolución 0883 de 2018, in 2 of the 10 samples that were analyzed with this method.

key Words: Ionic mercury (Hg), Tirilla (Sensafe), Colorimetry, Spectrophotometry, A priori information, Bayes Empirical Methods, multinomial distribution.

Contenido

Lista de Figuras	VIII
Lista de Tablas	IX
1. Introducción	1
2. Planteamiento del problema	2
2.1. Descripción	2
2.2. Justificación	3
2.3. Objetivos	3
2.3.1. Objetivo General	3
2.3.2. Objetivos Específicos	3
3. Antecedentes y Marco Teórico	4
3.1. Antecedentes	4
3.1.1. Antecedentes Contextuales	4
3.1.2. Antecedentes Estadísticos	6
3.2. Marco Teórico	8
3.2.1. Marco Teórico contextual	8
4. Metodología	15
4.1. Modelo estadístico asociado al problema	15
4.2. Análisis de los datos de colorimetría: modelo estadístico	21
4.3. Datos de espectrofotometría de absorción atómica por vapor frío.	22
5. Resultados	26
5.1. Validación de las tirillas (Sensafe)	27
5.2. Colorimetría	32
5.3. Espectrofotometría de absorción atómica por vapor frío.	33
6. Conclusiones y recomendaciones	36
A. Anexo: Código R	38

Bibliografía	40
Referencias	40

Lista de Figuras

3-1. Representación esquemática del modo de empleo de los sensores para el monitoreo de Mercurio.(Granada y Zertuche, 2019).	10
5-1. Estimación a posteriori de Hg para cada punto de muestreo.	31

Lista de Tablas

4-1. Número de muestras en el punto k , en la j -ésima fecha de muestreo.	17
4-2. Tamaño muestral n con un nivel α asociado al experto y un error tipo I igual a d para probabilidades multinomiales, Propuesto por Bromaghin (1993) . .	19
4-3. Patrón de oro	24
5-1. Resultados de concentración de Hg con las tirillas Sensafe, en ocho puntos de muestreo y por seis fechas en el corregimiento de La Toma.	27
5-2. Resultados obtenidos en laboratorio a partir de la variable Y	27
5-3. Distribución de las combinaciones de resultados por sitio de muestreo	28
5-4. Estimación de máxima verosimilitud para cada punto de muestro	29
5-5. Hiperparámetros de la distribución Dirichlet, para los diferentes niveles de experto.	29
5-6. Estimación a posteriori de la probabilidad conjunta de observar niveles de Mercurio asociado a una de las tres categorías en cada punto de muestreo . .	30
5-7. Región de credibilidad para las estimaciones a posteriori de Hg para cada punto de muestreo.	31
5-8. Clasificación para los colores de las celdas	32
5-9. Valores de colorimetría con clasificación de colores para cada W_i	32
5-10. Estadísticas descriptivas grupos 1 y 2	33
5-11. Resultados del patrón de oro vs las tirillas (Sensafe)	34

Capítulo 1

Introducción

Para el desarrollo de este trabajo de grado se consideró que la contaminación por Mercurio (Hg) en las fuentes hídricas tiene un impacto adverso en lo ecológico (fauna y flora) y en la salud humana, debido a que suele acumularse en órganos como los riñones y el cerebro; a largo plazo puede causar daños neurológicos y otras consecuencias. Por tal motivo, es importante conocer el estado en que se encuentran las fuentes de abastecimiento antes de ser ingerida o consumida por las familias que están expuestas a este elemento. En este sentido, el presente trabajo propone una metodología estadística, que permita clasificar niveles de Hg en aguas superficiales, mediante el uso de cualquier método de tirillas de detección. Dado que se contó con una muestra de observaciones bastante reducida debido a las limitaciones logísticas y económicas del proyecto de investigación que dio origen a las mediciones, fue necesario desarrollar una metodología de análisis de datos enmarcada dentro del paradigma Bayesiano de la Estadística.

Se propuso la distribución multinomial para clasificar los diferentes niveles de Hg, que pudiera contener las muestras de agua, y como distribución a priori se utilizó la distribución Dirichlet, cuyos hiperparámetros se obtuvieron mediante el Método Empírico de Bayes (MEB), dado que en la literatura no se encontraron estudios estadísticos previos sobre clasificación de Hg en aguas superficiales.

Para ilustrar los resultados, se utilizó un archivo de datos con 48 observaciones, el cual contaba con tres métodos de identificación de Mercurio que son: Tirillas Sensafe (método para el que se estaba evaluando su desempeño en campo), Colorimetría y espectrofotometría. Los resultados sugieren que las tirillas pueden ser un instrumento de alerta para la contaminación de Hg en las fuentes hídricas cuando se toma una cantidad de muestra considerable. Por su parte, los valores de la colorimetría para los diferentes niveles de Hg no presentaron una diferencia estadísticamente significativa. Mientras, el método de espectrofotometría que es una prueba patrón de oro, evidencia que existe una cantidad de Hg superior a lo permitido en la resolución 0883 de 2018, en dos de las 10 muestras analizadas con este método.

Capítulo 2

Planteamiento del problema

2.1. Descripción

Problema Contextual: Por las graves consecuencias que ocasiona en los seres humanos ingerir Mercurio (Hg), es importante establecer un control de las fuentes hídricas que abastecen a las personas que viven en lugares donde se utiliza este elemento químico. Por ello, es importante evaluar si el método de detección de tirillas puede ser un mecanismo de alerta temprana para detectar Mercurio iónico en fuentes hídricas. (Granada y Zertuche, 2019)

Problema estadístico:

Desde el punto de vista estadístico, el problema radica en la necesidad de contar con una metodología de análisis de datos, que permita obtener resultados que validen el uso de las tirillas Sensafe en agua superficial. Dadas las limitaciones logísticas y económicas del proyecto de investigación en el que se recogieron las mediciones en agua superficial, no fue posible contar con una muestra de datos lo suficientemente grande como para aplicar métodos estadísticos clásicos asociados a la validación de procedimientos de clasificación. Por lo que, se propuso el uso de procedimientos de análisis propios del paradigma Bayesiano de la Estadística, el cual considera que la muestra de datos con la que se cuenta son observaciones fijas y lo aleatorio es el conjunto de parámetros a estimar.

2.2. Justificación

Los habitantes de muchas regiones del país, especialmente las zonas rurales, se encuentran expuestos a la realización de actividades laborales que representan riesgo para salud y la vida como la minería legal o ilegal (Granada y Zertuche, 2019). En muchas zonas, el trabajo minero es desarrollado bajo condiciones que ponen en riesgo el bienestar de las personas que realizan la actividad. Una de estas condiciones de riesgo es el uso de Mercurio para la extracción de oro. El riesgo para los humanos se presenta tanto a nivel individual como ambiental, pues el lavado del lodo para extraer el oro genera descargas que van a parar a ríos, quebradas y otras fuentes hídricas como los pozos y aljibes. Siendo conscientes de esta problemática, profesores y estudiantes de las Escuelas de Ingeniería Ambiental y de Ingeniería Sanitaria de la Universidad del Valle, ejecutaron un proyecto de investigación con el objetivo de desarrollar y validar en campo un mecanismo económico y de fácil manejo, que le permita a los moradores de las zonas mineras en las que se utiliza Mercurio, monitorear sus fuentes hídricas y hacer un seguimiento de los niveles de contaminación por este elemento químico.

Durante el proyecto de investigación desarrollado por los profesionales de ingeniería, se realizaron seis muestreos a ocho fuentes hídricas ubicadas en diferentes partes de las comunidades rurales. Dada la cantidad tan limitada de datos para realizar un estudio de validación de procedimientos, se decidió recurrir a los métodos Bayesianos y a la simulación estocástica para obtener resultados que permitieran tener un grado de generabilidad aceptable.

2.3. Objetivos

2.3.1. Objetivo General

Estimar los niveles de Mercurio en fuentes de agua superficial a partir de observaciones obtenidas con tres metodologías diferentes de análisis de laboratorio.

2.3.2. Objetivos Específicos

- Proponer una metodología para evaluar la capacidad de clasificación de las tirillas Sensafe para detectar Mercurio.
- Evaluar los niveles de Mercurio (Hg) usando las observaciones obtenidas con espectrofotometría de absorción atómica por vapor frío, por las tirillas (Sensafe) y por el método de colorimetría.

Capítulo 3

Antecedentes y Marco Teórico

3.1. Antecedentes

Son diversos los artículos y trabajos de investigación en los que se habla del Mercurio y las consecuencias del contacto con el ser humano, la flora y fauna, pero son muy pocos los trabajos de investigación que usan muestras de agua superficial para validar la presencia de Mercurio asociadas a métodos estadísticos más allá de un análisis exploratorio de los datos recolectados. En este orden de ideas, se realizó una consulta y revisión bibliográfica de algunos trabajos que sirvieron como base metodológica y temática para el presente trabajo.

3.1.1. Antecedentes Contextuales

- **CRC (2017) (Corporación Regional del Cauca):** La Secretaría de Gobierno del Municipio de Suárez, Departamento del Cauca, denunció la presunta contaminación por cianuro en la quebrada Yolombó, que desemboca en el río Ovejas y posteriormente al río Cauca. Para corroborar y proceder con la denuncia interpuesta, la CRC se dirigió a las minas (La Gelima, El Silencio, La Esperanza y a tres nacimientos de agua) con el objetivo de conocer su funcionamiento, entables y localizar el lugar puntual de sus vertimientos a las quebradas. Durante la recolección de las muestras de agua se tomaron dos puntos para cada quebrada, uno antes de llegar al lugar de los vertimientos y otro justo después de éstos.

La CRC obtuvo como resultado después de enviar las muestras de agua recolectadas al laboratorio que: Las muestras tomadas antes de los vertimientos presentaban una concentración de Mercurio por debajo de la norma y las que tomaron después de éstos evidenciaron que la concentración de Mercurio se encontraban por encima de la norma establecida en la resolución 631 de 2015. Como segundo paso para determinar el grado de contaminación se usó la fórmula de “importancia”, siendo ésta una medida cualitativa del impacto a partir del grado de incidencia de la alteración producida por

los compuestos químicos presentes en el agua y de sus efectos. Esta formula tiene una clasificación de acuerdo a la puntuación con un rango de 8 a 80, el cual está dividido en cinco niveles de impacto (Irrelevante (8), leve (9-20), moderado (21-40), severo (41-60) y crítico (61-80)). Los resultados indicaron un valor de 29 por la ecuación de la importancia, de acuerdo a los rangos mencionados anteriormente afirmaron que la afectación a los recursos hídricos es moderada, por lo tanto, concluyeron que no se encontró presencia de Mercurio superior a 1ppb en las fuentes superficiales que pudieran afectar a la comunidad o al medio ambiente.

Vale la pena resaltar que la Universidad del Valle se había dirigido a la vereda Yolombó para realizar un estudio de concentración de Mercurio en fuentes hídricas superficiales, en donde fueron analizados trece (13) nacimientos de agua y los vertimientos de las minas La Gelima, El Silencio, La Esperanza, mostrando como resultado que solo uno de los yacimientos estudiados evidenciaba una concentración de Mercurio por debajo de la norma. Este informe de resultados fue entregado a la Junta de Acción Comunal (J.A.C), la CRC realizó una comparación entre ambos informes afirmando que los resultados de la Universidad del Valle carecían de validez, por lo cual no fueron tomados en cuenta.

- Como segundo antecedente y en respuesta a un llamado de la comunidad de la vereda Yolombó, La Toma, Municipio de Suárez (Cauca) por una segunda mortalidad de peces, las estudiantes del programa de Ingeniería Sanitaria y Ambiental **Granada y Zertuche** en el (2019), se dirigieron para apoyar a la comunidad. En el trabajo de grado plantearon cumplir dos objetivos; como primer objetivo, crear una conciencia colectiva partiendo del conocimiento científico sobre el impacto de la contaminación con Mercurio en la salud humana y en los ecosistemas. El segundo objetivo fue caracterizar la presencia de Mercurio en distintas fuentes de agua superficial en el territorio.

Implementaron la metodología IAP (Investigación y Acción Participativa) para el primer objetivo, dado que como método cualitativo le permitió a la comunidad participar activamente en el proceso investigativo e informativo, entendiendo los riesgos de salud que acarrea entrar en contacto con éste metal. Para el cumplimiento del segundo objetivo, después de definir un muestreo por importancia, establecieron tres procedimientos para analizar las muestras de agua que fueron, las tirillas Sensafe, prueba de laboratorio y el colorímetro (Instrumento para cuantificar el color de las tirillas).

Después de tomar las 48 muestras de agua y llevarlas al laboratorio, realizaron una adaptación de las tirillas Mercury check para observar las posibles cantidades de Hg presente en el agua. Cuantificaron el color de las tirillas usando el colorímetro para

observar si las variaciones de color en las tirillas estaban asociadas a la presencia de diferentes cuantificaciones de Hg. Posteriormente enviaron diez de las 48 muestras de agua al laboratorio para ser analizadas con el método de espectrofotometría de absorción atómica de vapor frío.

Como resultado obtuvieron que los rangos de las tirillas Sensafe para detectar Mercurio iónico en las muestras de agua, teniendo conocimiento del Ph, es bastante amplio, por lo tanto presenta una desventaja al momento de observación dado que el resultado de la misma puede terminar siendo subjetivo. Además de esta variabilidad, se concluyó que no son la mejor técnica de detección de Mercurio iónico para implementar in-situ, pues la variación de color se ve afectada.

En cuanto a los resultados de laboratorio se encontró presencia de Mercurio en dos de las diez mediciones. Se obtuvo un empoderamiento de la comunidad en cuanto a la lucha contra la minería ilegal y su contextualización de las consecuencias ambientales y de salud, aunque hay personas quienes no ven como opción salir del territorio para encontrar nuevas fuentes de abastecimiento. Sin embargo, la mayoría manifestaron estar dispuestos a cambiar sus prácticas cotidianas para el mejoramiento de las condiciones ambientales pensando en su salud y en la de las próximas generaciones.

3.1.2. Antecedentes Estadísticos

Al realizar una búsqueda detallada en la literatura sobre estudios estadísticos para validación de instrumento de Hg en aguas superficiales, no se encontró una metodología clara sobre este proceso, por tal motivo se consultó en la literatura sobre estudios de clasificación que pudieran ser útiles para esta investigación.

- Inicialmente tenemos a Hossain, Habib, y Andersson(2017) quienes realizaron un algoritmo para clasificar pacientes con la enfermedad del dengue basado en la capacidad de manejar incertidumbres. El dengue es una enfermedad debilitante transmitida por mosquitos, según la Organización Mundial de Salud (OMS) cada año la padecen en promedio 50 millones de personas. La forma que actualmente los médicos diagnostican a un paciente con dengue es de manera subjetiva, dado que observan los diversos signos, síntomas y factores de riesgo de esta enfermedad y clasifican en primera instancia según su conocimiento. Por lo tanto Hossain et al., proponen un sistema efectivo para diagnosticar la enfermedad del dengue, ya que es esencial para las personas que viven en zonas de alto riesgo de poseer esta enfermedad.

En esta investigación muestran el desarrollo y la aplicación de un sistema experto basado en la capacidad de manejar varios tipos de incertidumbres asociadas con los signos, síntomas y factores de riesgo del dengue. Para la ilustración del algoritmo utilizan una base de datos de pacientes ubicados en la ciudad de Chittagong y Bangladesh, la información se recolectó en un archivo de datos que fue entregado a un grupo de expertos en dengue para que diagnosticaran si el paciente tenía la enfermedad, además el mismo archivo de datos se utilizó en el algoritmo propuesto.

Como resultado mediante las curvas de características de funcionamiento del receptor (ROC), obtuvieron un valor para el algoritmo propuesto de 0.833 y un intervalo de (0.755 - 0.911) con una confianza del 95 %, y la curva ROC obtenida por la opinión de los expertos en dengue fue de 0.810 y un intervalo de (0.726 - 0.894) con una confianza del 95 %. De esta manera concluyen que los resultados generados por el sistema experto son más confiables que los del sistema basado en expertos humanos para diagnosticar a un paciente con la enfermedad del dengue.

- Por su parte **Quiñones y Suárez (2018)** realizaron una validación mediante la simulación de cuatro prototipos (algoritmos) ya establecidos por Méndez (2017) para clasificar individuos con dengue, a partir de la combinación de signos y síntomas expresados por el paciente en la consulta clínica.

Para la generación de las bases de datos se usó la simulación de variables aleatorias binarias correlacionadas mediante el algoritmo propuesto por Park, Park, y Shini (1996), Méndez (2017) estableció en su trabajo de grado un conjunto de signos y síntomas relevantes para el diagnóstico, en este trabajo Quiñones y Suárez (2018) los reagruparon debido a que contaban con una connotación similar. También habían obtenido como resultado por prototipos, unas sensibilidades y especificidades mostrada en tablas y que fueron necesarias para la asignación de los parámetros y así plantear la correlación entre la validación realizada anteriormente y la validación que se deseaba plantear como objetivo inicial en este trabajo.

Para la validación de los algoritmos se analizaron seis diferentes escenarios de simulación, los primeros cuatro escenarios están dados para los prototipos I y II y los últimos dos escenarios para los prototipos III y IV. Estos escenarios permitieron evaluar la efectividad de los algoritmos. La validación se realizó aplicando el método de bootstrap no paramétrico y calculando las probabilidades predictivas de los prototipos de Bayes discreto para los algoritmos prototipos I y II; Bayes continuo para los prototipos III y IV.

Como conclusión el prototipo III presentó un mejor desempeño en la clasificación de los pacientes, pues al momento de añadir los resultados del hemograma a los prototipos clasificadores se evidenció una disminución en la sensibilidad y un aumento en la especificidad del algoritmo. Respecto al poder discriminatorio los prototipos II y IV presentaron una disminución, caso contrario a los prototipos I y III.

Al someter los cuatro algoritmos a prueba en distintos escenarios se logró observar que el escenario I presentó una sensibilidad del 63,9 % siendo mayor que el 16,8 % del prototipo III, es decir, que cuando se tiene una prevalencia baja de la enfermedad y hay muchos casos de pacientes con resultados positivos en la categoría dolor general, el prototipo I discrimina mejor a los individuos con dengue. Una vez adicionada la información del hemograma, ambos prototipos presentan medidas de desempeño similares.

En el escenario III, se observó que el prototipo I presenta mayor sensibilidad que el prototipo III, siendo éstas del 87,4 % y 53,7 % respectivamente. Es decir que el prototipo I identifica mejor a los pacientes con dengue cuando se presenta una prevalencia baja de la enfermedad, pero alta en pacientes con la categoría “síntomas respiratorio positiva”. Con el hemograma se obtuvieron resultados similares para ambos prototipos. En cuanto al escenario IV donde se tiene un aumento de la prevalencia sin el hemograma, se obtiene un comportamiento similar entre los prototipos; una vez se adiciona el hemograma, el prototipo III presenta una sensibilidad del 73,6 % siendo ésta mayor a la del prototipo I, que pasa a ser del 65,2 %.

3.2. Marco Teórico

En este capítulo se presentan los conceptos de los aspectos más destacados para la contextualización de la minería en Colombia, sus causas y sus consecuencias además de técnicas estadísticas como la simulación de números binarios correlacionados, el método empírico de Bayes y algunas distribuciones de probabilidades, entre otros temas que son importantes comprender para llevar a cabo este trabajo de grado.

3.2.1. Marco Teórico contextual

Contaminación con Mercurio:

Como afirma Granada y Zertuche(2019):

“El Mercurio es el metal pesado más común y más ampliamente distribuido, y en particular es un potente inhibidor de muchas funciones proteicas, puede ser rápidamente absorbido en los pulmones a través de la piel o por ingestión. Muchas de las enfermedades inducidas por la absorción de compuestos de Mercurio son causadas por una exposición crónica de bajo nivel al metal o por una exposición temporal de alto nivel debido a negligencia humana y contaminación”. (p.9)

Mercurio iónico:

Es una transformación del Mercurio elemental que suele ocasionarse debido a la oxidación del vapor de Mercurio en el agua en presencia del oxígeno, generando la óxido-reducción. Este Mercurio puede sufrir dos reacciones en el medio. Una es la reducción del vapor de Mercurio metálico en condiciones apropiadas para ayudar la propagación de bacterias llamadas Pseudomonas (aeróbicas de oxidación), éste permite que al ser consumido por algún ser humano o animal pueda el cuerpo por medio de su respiración o cualquier otra secreción eliminar constantemente su concentración. La segunda es la conversión de los compuestos a la forma metil y dimetil Mercurio, siendo su forma más peligrosa dado que éste no se elimina del cuerpo sino que por el contrario se acumula donde dichas metilaciones se producen aeróbica (medio con oxígeno) y anaeróbicamente (medio sin oxígeno).

Minería ilegal:

Como menciona Juarez (2016):

“El Código de Minas (Congreso de la República, Ley 685 de 2001, Art. 159) describe la minería ilegal como la actividad exploratoria o de extracción de minerales, bien sean propiedad de la nación o de particulares, que se desarrolla sin el correspondiente título minero vigente o sin la autorización del titular de la propiedad privada donde se ubique el proyecto. Por su parte el código penal (Congreso de la República, Ley 599 de 2000, Art. 244, según el Código de Minas) la señala como un delito, mientras que la legislación ambiental (Congreso de la República, Ley 99 de 1993) establece los mecanismos para ejercer la autoridad ambiental sobre ella, en caso de destrucción del medio ambiente, y las autoridades locales actúan, por vía policiva, para erradicarla; en este contexto, se evita hacer referencia a la actividad de los grupos armados ilegales los cuales vuelven víctimas a los propios mineros, a quienes no se puede atribuir las estrategias del victimario.” (p.137)

Sensores comerciales Sensafe (*Mercury CheckTM*):

Son dispositivos (tirillas) de uso portátil que pueden ser utilizados *in-situ*, siendo de fácil implementación en el momento que sea necesario, para detectar niveles altos y bajos de Mer-

curio en agua con un rango de 50 ppb a 1000 ppb. Bajo la verificación del Ph en la muestra de agua se procede a sumergir la tirilla como lo indica la figura 3-1 (Granada y Zertuche, 2019) obteniendo un resultado en dos minutos y medio (150 segundos), procediendo a contrastar la tirilla con los colores que vienen en el exterior del recipiente, identificando así la cantidad de Mercurio presente.

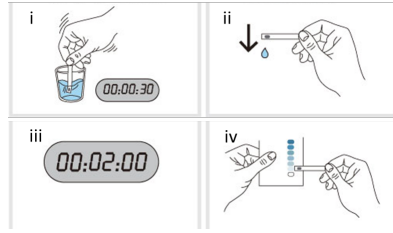


Figura 3-1: Representación esquemática del modo de empleo de los sensores para el monitoreo de Mercurio.(Granada y Zertuche, 2019).

Colorímetro:

Es un instrumento que sirve para la cuantificación del color, el procedimiento para usar este instrumento es colocar el escáner sobre la parte donde la tirilla muestra la coloración de acuerdo a la cantidad de Mercurio, obteniendo así la diferencia de color total ΔE que consta de tres valores ΔL que es la diferencia de luz y oscuridad ($+$ = *más luminoso* y $-$ = *más oscuro*), Δa diferencia en rojo y verde ($+a$ *más rojo*, $-a$ *más verde*) y Δb diferencia entre amarillo y azul ($+b$ *más amarillo*, $-b$ *más azul*). Para determinar la diferencia total de color entre las tres coordenadas se usa la fórmula 3-1. A partir de esta fórmula se obtuvieron las cuantificaciones del cambio de color de las tirillas.

$$\Delta E = \sqrt{\Delta L^2 + \Delta a^2 + \Delta b^2} \quad (3-1)$$

Espectofotometría de Absorción Atómica – Vapor Frío, SM 3112 B:

El método de detección de Mercurio más usado es la espectofotometría de absorción atómica de vapor frío, siendo esta técnica una de las mejores ya que permite la determinación directa si la muestra contiene Mercurio iónico Hg^+ que luego es transformado a Mercurio metálico (Icontec, 2017).

De acuerdo a González-E et al.(2014):

“El vapor formado se arrastra por un gas inerte hacia una celda de cuarzo en la que se produce el proceso de la absorción atómica. Las muestras sólidas se digieren antes para transformar todas las especies de *Hg* (inorgánicas y orgánicas) a *Hg2+*, ya que esta es la única especie capaz de generar el vapor de Mercurio atómico”.(p.116)

Dentro de las desventajas se puede citar que el método consume tiempo y es complicado por la posibilidad de pérdidas por volatilización o digestión incompleta, así como la contaminación de las muestras (Kuboyama, Sasaki, Nakagome, y Kataoka, 2005). Se denota que de todas las técnicas analíticas utilizadas esta es la más empleada, por realizar etapas muy sencillas de extracción, además de cuantificar satisfactoriamente el Mercurio, el cual es más usado en la detección de Mercurio en los músculos del pescado (Ruiz Chaves, 2016).

Marco Teórico Estadístico

Métodos Monte Carlo Basados en Cadenas de Markov (MCMC)

Los métodos MCMC son algoritmos iterativos que se usan para generar muestras de una distribución posterior y estimar estas cantidades, el diseño de una cadena de Markov es generada muestreando $\theta^{(t+1)} \sim p(\theta|\theta^t)$, este p es llamado el *Kernel de transición* de la cadena de Markov. La forma en que se construyen las cadenas garantiza la existencia de la distribución estacionaria. (Correa y Barrera, 2019)

La metodología de este algoritmo consiste en simular $\theta_1, \theta_2, \dots, \theta_R$, de una distribución de probabilidad dada, digamos $F(\theta)$, y estimar los valores esperados bajo esta distribución. Siendo $f(\theta)$ la densidad asociada con la distribución y esta será llamada *Densidad objetivo*, la cual puede ser la distribución a posteriori, que es la distribución de interés y esta condicionada en los datos.(Correa y Barrera, 2019)

Una de las propiedades más importante del método Monte Carlo es que la exactitud de la estimación no depende de la dimensionalidad del espacio muestreado. Si el algoritmo es implementado de la forma correcta se cumple también que el valor simulado θ_t depende únicamente de su predecesor θ_{t-1} , la convergencia de la cadena está garantizada independientemente de cuáles sean los valores iniciales. En este método es necesario simular la cadena para un número elevado de iteraciones para aproximarse a la distribución estacionaria y los primeros valores simulados (interacciones de burn-in) se eliminan si estos no se encuentran en estado estacionario.(Naranjo, 2017)

Distribución multinomial

Sea un experimento con n ensayos idénticos e independientes, en el que los resultados pueden tomar J valores distintos $(n_1, n_2, \dots, n_J)$. La probabilidad para cada valor n_j se denota por p_j ,

para $j = 1, 2, 3, \dots, J$, de modo que $0 < p_j < 1$ y $p_1 + p_2 + \dots + p_J = 1$. De esta manera, si N_j es la variable aleatoria que describe el número de veces que se observa el valor n_j en los n ensayos, para $j = 1, 2, 3, \dots, J$, entonces, el vector $N = (N_1, N_2, \dots, N_J)$ sigue una distribución multinomial, así:

$$N_1, N_2, \dots, N_J | p_1, p_2, \dots, p_J \sim Multinomial(n, p_1, p_2, \dots, p_J) \quad (3-2)$$

La función de probabilidad de dicha distribución viene dada por:

$$P(N_1 = n_1, N_2 = n_2, \dots, N_J = n_J) = \frac{n!}{n_1! n_2! \dots n_J!} p_1^{n_1} p_2^{n_2} \dots p_J^{n_J} \quad (3-3)$$

donde $\sum_{i=1}^J p_i = 1$, $0 \leq p \leq 1$, y $\sum_{i=1}^J n_i = n$.

Distribución Dirichlet

La distribución Dirichlet es una extensión multivariante de la distribución beta. Se trata de una distribución ampliamente utilizada en el contexto Bayesiano como una distribución a priori conjugada para una distribución multinomial (Connor y Mosimann, 1969)

Sea $Y = (y_1, y_2, \dots, y_j)$ un vector con j componentes, entonces diremos que este vector sigue una distribución Dirichlet con vector de parámetros $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_j)$ es decir,

$$Y | \alpha \sim Dirichlet(\alpha)$$

Si su función de densidad es:

$$P(y_1, y_2, y_3, \dots, y_j) = \frac{\Gamma(\sum_{k=1}^j \alpha_k)}{\Gamma(\alpha_1) \Gamma(\alpha_2) \dots \Gamma(\alpha_j)} y_1^{\alpha_1-1} y_2^{\alpha_2-1} y_3^{\alpha_3-1} \dots y_j^{\alpha_j-1}$$

donde $\sum_{i=1}^j y_i = 1$, $0 \leq y_i \leq 1$ con $i = 1, 2, \dots, j$. Y $\alpha_0 = \sum_{i=1}^j \alpha_i$

En este caso se tiene una distribución conocida para las marginales, que son distribuciones betas con parámetros $a = \alpha_j$ y $b = \alpha_0 - \alpha_j$. Para cada una se tiene:

$$E(Y_j) = \frac{\alpha_j}{\alpha_0}, \quad Var(Y_j) = \frac{\alpha_j(\alpha_0 - \alpha_j)}{\alpha_0^2(\alpha_0 + 1)}$$

Teorema de Bayes

De acuerdo a lo escrito en el informe de Barajas (2018), definió matemáticamente el teorema de Bayes como:

Sea $Y = \{y_1, y_2, \dots, y_n\}'$ un vector de n observaciones cuya distribución de probabilidad $Pr(y|\theta)$ depende de k parámetros involucrados en el vector $\theta = \{\theta_1, \theta_2, \dots, \theta_n\}'$. Supóngase también que θ tiene una distribución de probabilidades $Pr(\theta)$. Entonces, la distribución conjunta de θ e Y es:

$$Pr(y | \theta) = Pr(y | \theta)Pr(\theta) = Pr(\theta | y)Pr(y) \quad (3-4)$$

donde la distribución de probabilidad condicional de θ dado el vector de observaciones Y resulta:

$$Pr(\theta | y) = \frac{Pr(y | \theta)Pr(\theta)}{Pr(y)} \quad (3-5)$$

Con $Pr(y) \neq 0$

A esta ecuación se lo conoce como el teorema de Bayes, donde $Pr(y)$ es la distribución de probabilidad marginal de Y y puede ser expresada como:

$$Pr(y) = \begin{cases} \int Pr(y|\theta)Pr(\theta)d\theta & \text{si } \theta \text{ es continuo} \\ \sum Pr(y|\theta)Pr(\theta) & \text{si } \theta \text{ es discreto} \end{cases} \quad (3-6)$$

donde la suma o integral es tomada sobre el espacio paramétrico de θ . De este modo, el teorema de Bayes puede ser escrito como:

$$Pr(\theta | y) = cPr(y|\theta)Pr(\theta) \quad (3-7)$$

donde: $Pr(\theta)$ representa lo que es conocido de θ antes de recolectar los datos y es llamada la distribución a priori de θ , y $Pr(\theta|y)$ representa lo que se conoce de θ después de recolectar los datos y es llamada la distribución posterior de θ dado Y , y c es una constante normalizadora necesaria para que $Pr(\theta|y)$ sume o integre uno.

Dado que el vector de datos Y es conocido a través de la muestra, $Pr(y|\theta)$ es una función de θ y no de Y . En este caso a $Pr(y|\theta)$ se le denomina función de verosimilitud de θ dado Y y se le denota por $l(\theta|y)$. Entonces la formula de Bayes puede ser expresada como:

$$Pr(\theta|y) \approx l(\theta|y)Pr(\theta) \quad (3-8)$$

Prueba U de Mann withney

En estadística la prueba U de Mann-Whitney (también llamada de MannWhitney-Wilcoxon, prueba de suma de rangos Wilcoxon, o prueba de Wilcoxon-Mann-Whitney) es una prueba no paramétrica aplicada a dos muestras independientes. Es la versión no paramétrica de la habitual prueba t de Student, con la siguiente hipótesis.

H_o : la distribuciones de las poblaciones z_1 y z_2 son idénticas.

H_a : la distribución de la población z_1 es diferente a la distribución de la población z_2 .

se define el estadístico de prueba como

$$U = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - W$$

donde n_1, n_2 son los números de observaciones de las muestras z_1 y z_2 respectivamente; y W , es el estadístico calculado a partir de la suma de los rangos de la población de menor tamaño.

Sin embargo, cuando las muestras son grandes, por el teorema central del limite la prueba se aproxima bastante bien a la distribución normal, entonces se tiene que:

$$E(U) = \frac{n_1 n_2}{2}$$

$$\sigma_U^2 = \frac{n_1 n_2 (n_1 + n_2 + 1)}{12}$$

de donde se obtiene el siguiente estadístico de prueba.

$$Z = \frac{U - E(U)}{\sigma_U} \sim N(0, 1)$$

Y su región de rechazo viene dada por: $(Z < Z_\alpha)$

Fue propuesto inicialmente en 1945 por Wilcoxon para muestras de igual tamaño y extendido a muestras de tamaño arbitrario como en otros sentidos por Mann y Whitney (1947).

La prueba se usa para constatar si dos muestras independientes han sido seleccionadas de la misma población o de poblaciones diferentes que poseen el mismo comportamiento, bajo una distribución de probabilidad, es decir, que sus medias son iguales. La prueba, a diferencia de la de Wilcoxon, no se fundamenta en muestras pareadas, es decir, muestras en las que se obtienen dos observaciones para el mismo individuo o fenómeno.

Capítulo 4

Metodología

En esta sección se presenta la metodología propuesta para validar si las tirillas de detección son un instrumento que puede o no servir como mecanismo de alerta temprana, evaluando la capacidad que tienen estas para detectar la presencia de Mercurio en fuentes hídricas de abastecimiento. Dicha evaluación, se lleva a cabo utilizando un instrumento de colorimetría para cuantificar el cambio de las tirillas y un método de referencia gold standard, con muestras tomadas en varios puntos y distintas fechas.

4.1. Modelo estadístico asociado al problema

Parte de las componentes importantes del modelo estadístico es identificar el proceso generador de datos y la situación de la cual los datos están ofreciendo información. Se plantea entonces describir un experimento estadístico que es un procedimiento que genera los resultados de interés con los que se plantea el modelo estadístico.

Experimento estadístico: Tomar un registro de un triplicado de tirilla después de ser introducida en una muestra de agua tomada en el k -ésimo punto, de la j -ésima fecha de muestreo y observar el nivel de Mercurio marcado.

Variables aleatorias: Para la clasificación de las cantidades de Mercurio que fueron obtenidas por las tirillas, se crea la variable X , categórica de nivel ordinal, que representa el nivel de Hg considerado en la p -ésima observación, como se muestra a continuación, donde a_1 es el menor nivel de Mercurio y a_f el nivel más alto de Mercurio que puede marcar la tirilla.

$$X_{kjp} = \begin{cases} a_1 \\ a_2 \\ \vdots \\ a_c \\ \vdots \\ a_f \end{cases} \quad p = 1, 2, \dots, r, \quad j = 1, 2, \dots, m, \quad k = 1, 2, \dots, l.$$

donde $p = 1, 2, \dots, r$ hace referencia a la p -ésima observación, en la j -ésima fecha, para el k -ésimo punto de muestreo, con $j = 1, 2, \dots, m$ y $k = 1, 2, \dots, l$.

Al clasificar los resultados de las tirillas con la variable X , se crea una variable aleatoria Y que tomará valores de 1 o 2 según sea su clasificación de X , como se presenta a continuación:

$$Y_{kjp} = \begin{cases} 1; & \text{si } X \in \{a_1, a_2, a_3, \dots, a_c\} \rightarrow \text{cantidad despreciable de Mercurio.} \\ 2; & \text{si } X \in \{a_{c+1}, \dots, a_f\} \rightarrow \text{cantidad considerable de Mercurio.} \end{cases} \quad (4-1)$$

donde y_{kjp} es la cantidad de Mercurio clasificado en el punto $k = 1, 2, \dots, l$, en la fecha $j = 1, 2, \dots, m$. Y el termino a_c indica el nivel máximo tolerable de Mercurio para el consumo humano y se establece con ayuda de un experto.

Lo anterior establece que para un punto de muestreo en particular en una fecha dada, se tienen r observaciones en la variable Y_{kj} (tabla **4-1**). Partiendo de ello, en general para cualquier punto de muestreo en cualquier fecha, existen 2^r combinaciones posibles de las r observaciones. Por ejemplo, si se recolectaron $r = 3$ muestras (triplicados), es decir, se tomó una muestra de agua en el punto de muestreo uno en la fecha uno, y la misma se dividió en tres submuestras de agua para hacer los análisis, de modo que se tiene $2^3 = 8$ combinaciones posibles de la prueba de laboratorio en un triplicado. Dado que cada triplicado es llevado al laboratorio y es evaluado por un mismo analista, el orden en que aparece los resultados no es relevante, que para efectos del análisis son permutables porque son de una misma muestras de agua tomada en el mismo punto en la misma fecha, esto quiere decir que la combinación (1, 2, 1) es igual a la combinación (2, 1, 1) ó (1, 1, 2).

Tabla 4-1: Número de muestras en el punto k , en la j -ésima fecha de muestreo.

Sitio ID	Fecha de muestreo			
	Fecha 1	Fecha 2	...	Fecha m
Punto 1	$y_{111}, y_{112}, \dots, y_{11r}$	$y_{121}, y_{122}, \dots, y_{12r}$	...	$y_{1m1}, y_{1m2}, \dots, y_{1mr}$
Punto 2	$y_{211}, y_{212}, \dots, y_{21r}$	$y_{221}, y_{222}, \dots, y_{22r}$	...	$y_{2m1}, y_{2m2}, \dots, y_{2mr}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
Punto l	$y_{l11}, y_{l12}, \dots, y_{l1r}$	$y_{l21}, y_{l22}, \dots, y_{l2r}$	...	$y_{lm1}, y_{lm2}, \dots, y_{lmr}$

Después de realizar las diferentes combinación se concluyen que existen $3+1$ ($r+1$) resultados posibles de interés, lo cual lleva a tener las siguientes combinaciones: $(1, 1, 1), (1, 1, 2), (1, 2, 2), (2, 2, 2)$. Y de manera general los siguientes eventos de interés W_t , con $t = 1, 2, \dots, r, r+1$, donde:

$$\begin{cases} W_1 = & \text{Existen } r \text{ muestras con } Y = 1 \\ W_2 = & \text{Existen } r-1 \text{ muestras con } Y = 1 \\ \vdots = & \vdots \\ W_t = & \text{Existen } (r+1) - t \text{ muestras con } Y = 1 \end{cases} \quad (4-2)$$

Esto quiere decir que, W_1 es el menor nivel de Hg marcado por la tirilla y W_{r+1} el mayor nivel de Mercurio marcado por la tirilla, en el punto k , de la j -ésima fecha de muestreo.

Para representar de una mejor manera lo que sucede en el k -ésimo punto de muestreo, se crea una variable aleatoria D_{ktj} , que identifica el evento W_t , tal que:

$$D_{ktj} = \begin{cases} 1 & \text{si ocurre } w_t \\ 0 & \text{si no ocurre } w_t \end{cases} \quad t = 1, 2, 3, \dots, r, r+1 \quad (4-3)$$

donde $D_{ktj} \sim \text{Bernoulli}(\theta)$ con función de masa de probabilidad:

$$P[D_{ktj} = d_{ktj} | \theta_{kt}] = \theta_{kt}^{d_{ktj}} (1 - \theta_{kt}^{1-d_{ktj}}) \quad ; \quad D_{ktj} = \{0, 1\} \quad y \quad 0 \leq \theta_{kt} \leq 1$$

θ_{kt} = La probabilidad de que ocurra la observación d_t en el punto k .

Posteriormente se crea la variable Q_{kt} , la cual representa el número de éxitos de interés del evento definido por la variable D_{kt} , esto es:

$$Q_{kt} = \sum_{j=1}^{m_k} D_{ktj} \quad \text{para } t = 1, 2, \dots, r, r+1; \quad k = 1, \dots, l \quad (4-4)$$

donde m_k es la cantidad de muestras obtenida en las j -ésimas fechas de muestreo en el punto k .

Entonces, la distribución de masa conjunta es una multinomial, y está dada por:

$$P(Q_{k1} = q_{k1}, Q_{k2} = q_{k2}, \dots, Q_{kr} = q_{kr}, Q_{k(r+1)} = q_{k(r+1)} | \theta_{k1}, \theta_{k2}, \dots, \theta_{kr}, \theta_{k(r+1)}) = \frac{m_k!}{q_{k1}! q_{k2}! \dots q_{k(r+1)}!} \theta_{k1}^{q_{k1}} \theta_{k2}^{q_{k2}} \dots \theta_{k(r+1)}^{q_{k(r+1)}}$$

donde $\sum_{t=1}^{r+1} \theta_{kt} = 1$, para todo k y $m_k = \sum_{t=1}^{r+1} q_{kt}$ Entonces

$$Q_k \sim \text{multinomial} (m_k, \theta_{k1}, \theta_{k2}, \theta_{k3}, \dots, \theta_{kr}, \theta_{k(r+1)}) \quad (4-5)$$

donde θ_{kt} es la probabilidad de que, en el punto k , se presente el evento t .

Su estimación por máxima verosimilitud, viene dada por:

$$\hat{\theta}_{kj} = \frac{q_{kj}}{m_k} \quad (4-6)$$

Inferencia Bayesiana

Al estar en un escenario de pocos datos, se hace uso de la inferencia Bayesiana para estimar los parámetros de interés $\vec{\theta}_k = (\theta_{k1}, \theta_{k2}, \theta_{k3}, \dots, \theta_{kr}, \theta_{k(r+1)})$. Esta combina la información de los datos más información subjetiva, sea de un experto o información de estudios previos. Dado que en la revisión de la literatura no se encontraron estudios con características similares a éste, que pudieran servir como fuente para obtener información a priori y tampoco se tuvo acceso a un especialista en el tema que pudiera ser considerado experto, se decidió utilizar los métodos empíricos de Bayes (MEB) para obtener las distribuciones a priori (Gelman et al., 2004). Al asumir una distribución multinomial como mejor modelo para representar las observaciones obtenidas en campo una elección natural para la distribución a priori es la distribución Dirichlet, la cual es la distribución a priori conjugada para una verosimilitud multinomial. Se tiene entonces que la distribución posterior del vector de parámetros de interés viene dada por la siguiente expresión:

$$f(\theta_{k1}, \theta_{k2}, \dots, \theta_{kr}, \theta_{k(r+1)} | \alpha_{k1}, \alpha_{k2}, \dots, \alpha_{k(r+1)}) = C \theta_{k1}^{\alpha_{k1}-1} \theta_{k2}^{\alpha_{k2}-1} \dots \theta_{k(r+1)}^{\alpha_{k(r+1)}-1} \quad (4-7)$$

$$\text{donde } C = \frac{\Gamma(\sum_{t=1}^{r+1} \alpha_{kt})}{\Gamma(\alpha_{k1}) \Gamma(\alpha_{k2}) \dots \Gamma(\alpha_{k(r+1)})}, \quad \sum_{t=1}^{r+1} \theta_{kt} = 1, \quad 0 \leq \theta_k \leq 1, \quad \text{y} \quad \alpha_0 = \sum_{t=1}^{r+1} \alpha_t.$$

Para obtener los hiperparámetros de la distribución a priori Dirichlet, se hizo una adaptación del método propuesto por Flórez (2014), usando información obtenida a partir de los mismos datos (MEB).

Obtención de los hiperparámetros de la distribución a priori Dirichlet

El método propuesto por Flórez (2014) consiste en tres procedimientos. Primero, encontrar el n equivalente, el cual depende de la experiencia o credibilidad que el facilitador asigna al

experto, de tal manera que dicho tamaño de muestra represente la opinión de éste. Dado que no se tiene información de un experto, Bromaghin (1993) propone que el tamaño de muestra n se puede obtener de la siguiente manera ver tabla **4-2**, donde α toma valores de 0.6, 0.4 y 0.2 lo que indica el conocimiento que el facilitador asigna al experto, poco conocimiento, mediano conocimiento y alto conocimiento del tema respectivamente y d es el máximo error tipo I asumido al considerar el tamaño de muestra asociado al número de categorías de la variable multinomial (v).

Tabla 4-2: Tamaño muestral n con un nivel α asociado al experto y un error tipo I igual a d para probabilidades multinomiales, Propuesto por Bromaghin (1993)

v	α								
	0,6			0,4			0,2		
	d			d			d		
	0,05	0,08	0,1	0,05	0,08	0,1	0,05	0,08	0,1
2	268	118	65	163	72	40	107	47	26
3	333	147	81	224	98	55	163	72	40
4	381	167	93	268	118	65	206	91	50
5	418	184	102	304	134	74	240	106	59
6	449	197	109	333	147	81	268	118	65
7	475	209	116	359	158	87	293	129	71
8	498	219	121	381	167	93	314	138	77
9	518	228	126	400	176	97	333	147	81
10	536	236	130	418	184	102	351	154	85

En el segundo procedimiento, se aplica el método de la Ruleta propuesto por Oakley, Darneshkhah, y O'Hagan (2010) que le permite al analista estimar el vector de probabilidades de ocurrencia de cada parámetro $\vec{\theta}_k = (\theta_{k1}, \theta_{k2}, \theta_{k3}, \dots, \theta_{kr}, \theta_{k(r+1)})$ dialogando con el experto. Como no se tiene información de dicha proporción se tomaron las estimaciones globales de las muestras, obtenidas por el método de máxima verosimilitud como el vector de probabilidades de ocurrencia.

El tercer paso se desarrolla un estudio de simulación con el cual se obtienen los valores de los hiperparámetros de la distribución Dirichlet utilizando como insumo la información obtenida en los pasos 1 y 2, es decir, los tamaños de muestras asociados a valores predefinidos de (α , d y v) y al vector de parámetros θ_k . Para cada combinación de valores se simulan N muestras (para nuestro caso $N=1000$) de la distribución multinomial con vector de parámetros θ_k , se obtuvieron los 1000 vectores de estimaciones del vector θ_k y para cada componente se obtuvo la media y la varianza, asumiendo la media como $E[Y_i]$ y la varianza como $Var[Y_i]$.

Luego de tener los resultados del tercer componente se procedió a calcular los valores de

los hiperparámetros de la distribución a priori Dirichlet, haciendo uso de los dos primeros momentos de dicha distribución, que son los siguientes:

$$E[Y_i] = \frac{\alpha_i}{\alpha_0} \quad (4-8)$$

$$Var[Y_i] = \frac{\alpha_i(\alpha_0 - \alpha_i)}{\alpha_0^2(\alpha_0 + 1)} \quad (4-9)$$

donde $\alpha_0 = \sum_{t=1}^{r+1} \alpha_t$ y t representa el número de categorías de la distribución.

Las estimaciones de la ecuaciones 4-8 y 4-9 son fáciles de obtener conociendo los resultados del tercer paso. Sin embargo, el procedimiento algebraico está en despejar α_0 y α_i de las expresiones anteriores. Las cuales quedan expresadas de la siguiente manera:

$$\alpha_0 = \frac{(E[Y_i] - (E[Y_i])^2)}{Var[Y_i]} - 1 \quad (4-10)$$

$$\alpha_t = \left(\frac{(E[Y_i] - (E[Y_i])^2)}{Var[Y_i]} - 1 \right) E[Y_i] \quad (4-11)$$

De esta forma se obtienen los hiperparámetros de la distribución Dirichlet, (ecuaciones 4-10 y 4-11).

Ahora es posible obtener la distribución posterior conjunta para el vector θ_k usando la formula de Bayes, es decir podemos obtener la distribución posterior de la siguiente manera:

$$f(\vec{\theta}_{kt} | \vec{q}_{kt}) = hL(\vec{\theta}_k | \vec{q}_{kt})f(\vec{\theta}_k) \propto \prod_{t=1}^{r+1} \theta_{kt}^{q_{kt} + \alpha_{kt} - 1} \quad (4-12)$$

Donde $L(\vec{\theta}_k | \vec{q}_{kt})$ es la verosimilitud de la distribución multinomial y $f(\vec{\theta}_k)$, la información a priori Dirichlet, indexada por los hiperparámetros obtenidos en la ecuación 4-10, 4-11, y h una constante de normalización. La $f(\vec{\theta}_{kt} | \vec{q}_{kt})$ contiene toda la información posible sobre el vector de parámetros θ_k de modo que, se pueden obtener medidas de resumen de la misma para estimar la probabilidad de observar una cantidad q de tirillas en la t -ésima categoría de nivel de Mercurio. Asumiendo una función de perdida cuadrática se puede tomar la media de la distribución posterior como estimador de Bayes. Es decir;

$$\hat{\theta}_{kt} = E(\theta_{kt} | q_{kt}) = \frac{q_{kt} + \alpha_{kt}}{m_k + \sum_{t=1}^{r+1} \alpha_t} \quad (4-13)$$

Estas estimaciones de los $\vec{\theta}_k$ darán una alerta sobre cómo es la condición del agua antes de ser ingerida. El método de la tirillas no permite hacer una cuantificación exacta de la cantidad de Mercurio en el k -ésimo punto, en la j -ésima fecha de muestreo. Las cantidades de Hg en ppb marcadas por las tirillas son números enteros que identifican límites superiores para posibles cantidades de Hg (0, 50, 100, 500, 700 y 1000 ppb) de modo que, si en el agua hay una cantidad de Hg igual a un valor ubicado en el interior de un intervalo, la coloración de la tirilla va a ser interpretada (de manera visual) como el límite inferior del intervalo. La forma como los ingenieros han manejado esta limitación ha sido utilizar el método de la colorimetría, el cual consiste en medir la intensidad de color a través de un procedimiento de laboratorio.

4.2. Análisis de los datos de colorimetría: modelo estadístico

Para cuantificar el valor colorimétrico se utiliza la siguiente formula.

$$\Delta E = \sqrt{\Delta L^2 + \Delta a^2 + \Delta b^2} \quad (4-14)$$

Dicha ecuación 4-14, se le realiza a las tiras reactivas después de ser introducidas en muestras de agua, con el fin de tener un valor cuantitativo que represente la cantidad de Mercurio en las fuentes hídricas. Para ello se tiene el siguiente modelo estadístico asociado.

Experimento estadístico: Observar el resultado de colorimetría, luego de cuantificar el nivel de Mercurio de la observación d_{kt} en la fecha de muestreo j .

Variable aleatoria: Sea Z_t la variable aleatoria que toma los valores de cuantificación colorimétrica de la observación d_t , donde $t = 1, 2, 3, \dots, r, r + 1$ haciendo referencia a la categoría de la distribución multinomial.

Por lo tanto, para cada categoría se tendrá en total u_t observaciones con $u = 1, 2, \dots, n_t$, asociadas a la cantidad de tirillas en la categoría t , a estas se les realiza un análisis descriptivo y posteriormente la prueba no paramétrica de Kruskal-Wallis si $r + 1 > 2$, ó U de Mann Withney si $r + 1 = 2$, para evidenciar estadísticamente si los valores de Z_t medidos en la colorimetría, difieren entre las $(r + 1)$ categorías. De esta manera concluir si los resultados arrojados por la colorimetría tienen un soporte para aportar a la validación de las tirillas. Sin embargo, es necesario resaltar que las tirillas se tendrían que comparar con un método de patrón de oro para concluir sobre la sensibilidad de esta para detectar Mercurio en agua superficial.

4.3. Datos de espectrofotometría de absorción atómica por vapor frío.

Esta prueba se considera como patrón de oro, ya que puede clasificar correctamente si una muestra de agua tiene una cantidad de Hg superior a 1ppb (una parte por billón), que es lo permitido por la norma de salud pública en Colombia. Por ende, las muestras a las que se les realice dicha prueba serán presentadas en una tabla junto a los resultados de las tirillas asociados a dichas muestras del patrón de oro. En consecuencia, se plantea el siguiente modelo estadístico:

Experimento estadístico: Observar el resultado que se obtiene al aplicar el patrón de oro a una muestra de agua. Además observar la clasificación de las tirillas asociada a esa muestra con la metodología de la sección 4.1. (ecuación 4-2).

Teniendo en cuenta que para cada país la norma sobre las cantidades permitidas de Mercurio en agua consumible para el ser humano es diferente, se crea un punto de corte b . Luego, con ayuda de un experto se define cuáles observaciones de W_t se consideran tolerables para el consumo humano. De esta manera, se crean los siguientes eventos de interés:

Evento H_1 : El patrón de oro obtiene un resultado mayor que b , y la observación de W_t no es tolerable para el consumo humano.

Evento H_2 : El patrón de oro obtiene un resultado mayor que b , y la observación de W_t es tolerable para el consumo humano.

Evento H_3 : El patrón de oro obtiene un resultado menor que b , y la observación de W_t no es tolerable para el consumo humano.

Evento H_4 : El patrón de oro obtiene un resultado menor que b , y la observación de W_t es tolerable para el consumo humano.

Sea V la variable aleatoria que identifica los resultados del evento patrón de oro, tal que:

$$V = \begin{cases} 1 & \text{si el patrón de oro obtiene un resultado mayor que } b. \\ 0 & \text{en otro caso.} \end{cases} \quad (4-15)$$

Sea C la variable aleatoria que identifica los resultados del evento tirillas, tal que:

$$C = \begin{cases} 1 & \text{si la observación de } W_t \text{ no es tolerable para el consumo humano.} \\ 0 & \text{en otro caso.} \end{cases} \quad (4-16)$$

Sean las variables E_i tales que:

$$E_i = \begin{cases} 1 & \text{si ocurre } H_i \\ 0 & \text{en otro caso.} \end{cases} \quad (4-17)$$

Entonces,

$$E_1 = \begin{cases} 1 & \text{si } V = 1 \text{ y } C = 1 \\ 0 & \text{en otro caso.} \end{cases} \quad (4-18)$$

$$E_2 = \begin{cases} 1 & \text{si } V = 1 \text{ y } C = 0 \\ 0 & \text{en otro caso.} \end{cases} \quad (4-19)$$

$$E_3 = \begin{cases} 1 & \text{si } V = 0 \text{ y } C = 1 \\ 0 & \text{en otro caso.} \end{cases} \quad (4-20)$$

$$E_4 = \begin{cases} 1 & \text{si } V = 0 \text{ y } C = 0 \\ 0 & \text{en otro caso.} \end{cases} \quad (4-21)$$

Se puede asumir que $E_i \sim \text{Bernoulli}(\theta_i)$, entonces

- $\theta_1 = P(\text{ocurra } H_1) = P(V = 1, C = 1)$
- $\theta_2 = P(\text{ocurra } H_2) = P(V = 1, C = 0)$
- $\theta_3 = P(\text{ocurra } H_3) = P(V = 0, C = 1)$
- $\theta_4 = P(\text{ocurra } H_4) = P(V = 0, C = 0)$

donde

- θ_1 es la probabilidad de que en una muestra de agua, el patrón de oro marque una cantidad de Mercurio superior a b y las tirillas marquen que no es tolerable para el consumo humano.
- θ_2 es la probabilidad de que en una muestra de agua, el patrón de oro marque una cantidad de Mercurio superior a b y las tirillas marquen que es tolerable para el consumo humano.

- θ_3 es la probabilidad de que en una muestra de agua, el patrón de oro marque una cantidad de Mercurio inferior a b y las tirillas marquen que no es tolerable para el consumo humano.
- θ_4 es la probabilidad de que en una muestra de agua, el patrón de oro marque una cantidad de Mercurio inferior a b y las tirillas marquen que es tolerable para el consumo humano.

Por consiguiente se define la variable aleatoria A_i como la suma de n_i variables aleatorias Bernoulli de parámetros θ_i . Es decir:

con $A_1|V = 1 \sim \text{Binomial}(a_1 + a_3, \lambda_1)$ y $A_4|V = 0 \sim \text{Binomial}(a_2 + a_4, \lambda_2)$ donde $\lambda_1 = P(C = 1|V = 1)$ y $\lambda_2 = P(C = 0|V = 0)$. Los datos obtenidos se pueden organizar en una tabla cruzada de la siguiente manera:

Tabla 4-3: Patrón de oro

		Gold standard		
Muestra de agua		Si	No	
Tirillas.	Si	a_1	a_2	$a_1 + a_2$
	No	a_3	a_4	$a_3 + a_4$
		$a_1 + a_3$	$a_2 + a_4$	n

donde la palabra **Si** indica que la muestra de agua es perjudicial para el consumo humano, y la palabra **No** indica que la muestra de agua no es perjudicial para el consumo humano.

Para el caso de este proyecto se utilizó como prueba gold standard la espectrofotometría de absorción atómica por vapor frío; método de laboratorio que permite identificar cantidades mayores o iguales a 1 ppb en muestras de agua.

La espectrofotometría es una técnica especializada y costosa, razón por la cual diez de las 48 muestras fueron analizadas utilizando este método, con el fin de darle una generalización a los resultados obtenidos con el gold standard se procedió a estimar la probabilidad de que una muestra de agua tomada en alguna de las fuentes seleccionadas para hacer los muestreos, este contaminada con Hg usando distribuciones beta conjugadas para un muestreo binomial bajo el principio de Laplace.

Se simuló un vector de 1000 valores de la distribución posterior y usando la función de pérdida cuadrática se obtuvo el estimador de la probabilidad de que una fuente de agua de la zona de estudio presente niveles de mercurio mayor a los aceptados. Es decir, se estimó la proporción de fuentes de agua para consumo en el área de estudio y la región de credibilidad del 95 %.

Para obtener información sobre la capacidad clasificadora de las tirillas Sensafe se tomaron las diez muestras analizadas por espectrofotometría y se comparó los resultados de este método con la clasificación arrojada por los triplicados de las tirillas evaluadas en las mismas muestras de agua para este análisis se utilizaron distribuciones apriori obtenidas bajo el principio de Laplace y verosimilitudes binomiales.

Capítulo 5

Resultados

Para ilustrar la metodología propuesta en este trabajo de grado, se utilizó un archivo de datos cribados por dos estudiantes de Ingeniería Sanitaria, quienes en su trabajo de grado estudiaron las condiciones de contaminación con Mercurio (Hg) en ocho fuentes de abastecimiento de agua (seis pozos, un acueducto y una quebrada) realizando seis jornadas de muestreo en el corregimiento La Toma y la vereda Yolombó en el municipio de Suárez, Cauca. Las 48 muestras recolectadas fueron llevadas y analizadas en el laboratorio de Ingeniería Química de la Universidad del Valle, mediante los métodos de tirillas (Sensafe), colorimetría y espectrofotometría de absorción atómica por vapor frío.

Posteriormente cada muestra se repartió en tres recipientes, y se le introdujo una tirilla marca Sensafe, para clasificar el nivel de Mercurio (Hg) que pudiera contener la muestra (esta se clasifica en 50, 100, 200, 500 ó 1000 partes por billón) de esta manera para cada fecha y punto de muestreo se tienen 3 resultados. De la misma manera, se realizó el método de colorimetría para cada tirilla, por lo cual, en cada fecha se tiene un valor de colorimetría asociado a las tirillas, y 10 de las muestras fueron enviadas al laboratorio de servicios a la comunidad en Medellín, Colombia; en el que se realizaron técnicas específicas que permitieran clasificar correctamente si la muestra de agua contenía una cantidad superior a 1ppb/L (una parte por billón por litro) que es lo permitido según la resolución 2115 del 2007 por el Ministerio de la protección social y Ministerio de ambiente, vivienda y desarrollo territorial. La Tabla **5-1** presenta los resultados de las tirillas.

Tabla 5-1: Resultados de concentración de Hg con las tirillas Sensafe, en ocho puntos de muestreo y por seis fechas en el corregimiento de La Toma.

Sitio ID	Fechas de muestreo																	
	Febrero 27			Marzo 7			Marzo 21			Abril 4			abril 18			Mayo 2		
La Cristalina	100	100	100	0	0	0	50	50	0	0	50	0	0	0	0	0	0	
Alcides	50	0	100	0	0	0	100	0	0	0	0	0	0	0	0	50	0	
Pablo Balanta	0	50	0	50	0	0	0	0	50	0	100	0	0	0	0	0	50	
Escuela	100	0	0	0	50	0	0	0	50	50	0	0	0	0	0	0	50	
Palo de Yuca	0	0	0	0	50	50	0	0	0	0	0	0	0	0	0	0	0	
Gloria	0	100	50	0	0	50	100	50	50	0	0	0	0	0	0	100	100	
Erley	50	100	50	0	0	50	100	50	50	100	100	100	0	0	0	0	0	
Tres Quebradas	0	0	0	50	0	50	100	100	50	50	50	50	0	0	0	100	100	

5.1. Validación de las tirillas (Sensafe)

Siguiendo la metodología propuesta en la sección anterior, para validar si las tirillas pueden ser o no, un instrumento de alerta temprana para detectar Mercurio en fuentes hídricas de abastecimiento para la comunidad de La Toma, Cauca. Dado que las tirillas son un método en evaluación, la persona que las desarrolló (Ingeniera sanitaria PhD. Diana Vanegas) nos informó que su capacidad para detectar niveles bajos de Hg es muy limitada y que ella consideraba que cuando la tirilla marcaba 0 o 50 ppb era posible asumir niveles muy bajos, casi despreciables de Hg en el agua. De esta manera se tiene los siguientes resultados:

Tabla 5-2: Resultados obtenidos en laboratorio a partir de la variable Y

Sitio ID	Fechas de muestreo																	
	Febrero 27			Marzo 7			Marzo 21			Abril 4			abril 18			May 2		
La Cristalina	2	2	2	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1
Alcides	1	1	2	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1
Pablo Balante	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1
Escuela	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Palo de Yuca	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Gloria	1	2	1	1	1	1	2	1	1	1	1	1	1	1	1	2	2	2
Erley	1	2	1	1	1	1	2	1	1	1	2	2	2	1	1	1	1	1
Tres Quebradas	1	1	1	1	1	1	2	2	1	1	1	1	1	1	1	2	2	2

Ahora que se tiene los resultados de la variable Y_{kj} codificada como 1 ó 2, como se muestra en la tabla 5-2, se consideran los siguientes eventos de interés W_{kjt} con $t = 1, 2, 3, 4$, que tiene las siguientes combinaciones:

- $W_1 = (1\ 1\ 1)$ → muy poca sospecha de Hg
- $W_2 = (1\ 1\ 2)$ → poca sospecha de Hg
- $W_3 = (1\ 2\ 2)$ → sospecha de Hg
- $W_4 = (2\ 2\ 2)$ → alta sospecha de Hg.

A continuación se presenta la tabla **5-3** codificada con los resultados de los eventos W_t , de esta manera podemos ver el comportamiento de cada uno de los ocho puntos de muestreo en sus distintas fechas, por ejemplo, se puede evidenciar como los puntos Pablo Balanta, La Escuela y Palo de Yuca, presentan muy poca sospecha de Hg por el instrumento. Mientras los puntos Erley, Gloria y Tres Quebradas presentan una mayor sospecha de Mercurio (Hg).

Tabla 5-3: Distribución de las combinaciones de resultados por sitio de muestreo

Sitio ID	Fechas de muestreo					
	Febrero 27	Marzo 7	Marzo 21	Abril 4	abril 18	Mayo 2
La Cristalina	w4	w1	w1	w1	w1	w2
Alcides	w2	w1	w2	w1	w1	w1
Pablo Balanta	w1	w1	w1	w2	w1	w1
Escuela	w2	w1	w1	w1	w1	w1
Palo de Yuca	w1	w1	w1	w1	w1	w1
Gloria	w2	w1	w2	w1	w1	w4
Erley	w2	w1	w2	w4	w1	w1
Tres Quebradas	w1	w1	w3	w1	w1	w4

La Tabla **5-4** presenta las estimaciones por el método de máxima verosimilitud en los ocho puntos de muestreo, de esta se concluye que las tirillas marcaron una tendencia a muy poca sospecha de Hg. Por ejemplo, los puntos de muestreo Alcides, Pablo Balanta, Escuela y Palo de Yuca, tienen una estimación muy poco probable de contener niveles considerables de Mercurio, según las tirillas. Mientras que los puntos Cristalina, Gloria y Erley presentan en algún momento indicios de Hg. Por su parte, Tres Quebradas presentó la más alta sospecha de Hg según los resultados de las tirillas (Sensafe).

Tabla 5-4: Estimación de máxima verosimilitud para cada punto de muestro

Sitio ID	n	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$\hat{\theta}_4$
Cristalina	6	$\frac{4}{6}$	$\frac{1}{6}$	0	$\frac{1}{6}$
Alcides	6	$\frac{4}{6}$	$\frac{2}{6}$	0	0
Pablo balanta	6	$\frac{5}{6}$	$\frac{1}{6}$	0	0
Escuela	6	$\frac{5}{6}$	$\frac{1}{6}$	0	0
Palo de Yuca	6	1	0	0	0
Gloria	6	$\frac{3}{6}$	$\frac{2}{6}$	0	$\frac{1}{6}$
Erley	6	$\frac{3}{6}$	$\frac{2}{6}$	0	$\frac{1}{6}$
Tres Quebradas	6	$\frac{4}{6}$	0	$\frac{1}{6}$	$\frac{1}{6}$

Resultados del estudio de simulación

Aplicando los tres pasos descritos en la metodología, haciendo uso del software R Core Team (2018), se obtuvo los valores para el vector de hiperparámetros $\alpha = (\alpha_0, \alpha_1, \alpha_2, \alpha_3)$ y las estimaciones Bayesianas de las probabilidades conjuntas en la Tabla 5-5, considerando los niveles hipotéticos de conocimiento de un especialista (Aproximados utilizando MEB, y denotados como alto, medio y bajo), un nivel de significancia y los tamaños de muestra que aparecen en la Tabla 4-2. Se puede decir, que para cada nivel de conocimiento se obtuvo una distribución a priori Dirichlet, razón por la que fue necesario utilizar un método de selección de modelos (FB), para obtener los resultados finales.

De la tabla 5-5 se puede observar que el vector de parámetros es muy similar para los diferentes niveles de experticia asignado al experto. Lo cual tiene sentido por la poca variabilidad que tienen los puntos.

Tabla 5-5: Hiperparámetros de la distribución Dirichlet, para los diferentes niveles de experto.

	Experticia alta $n = 206$			Experticia media $n = 268$			Experticia baja $n = 381$		
	Alfa	Media	Varianza	Alfa	Media	Varianza	Alfa	Media	Varianza
α_0	149.32	0.70	9.79e - 4	182.11	0.70	8.04e - 4	276.18	0.71	5.27e - 4
α_1	41.75	0.20	7.35e - 4	48.83	0.19	5.83e - 4	74.23	0.19	3.96e - 4
α_2	4.71	0.02	7.35e - 4	5.78	0.02	8.32e - 5	7.99	0.02	5.16e - 5
α_3	17.30	0.08	3.48e - 4	23.40	0.09	3.13e - 4	30.31	0.08	1.84e - 4

Luego de tener los parámetros de la distribución a priori Dirichlet, se procede a obtener las estimaciones de la distribución posteriori (ecuación 4-12), que resulta al multiplicar la verosimilitud de la distribución multinomial, por la distribución a priori Dirichlet con hiperparámetros expresadas en la tabla 5-5. De esta forma se obtiene una distribución a posteriori Dirichlet para cada punto de muestreo, con parámetros presentados en la tabla 5-6 y su respectiva región de credibilidad presentada en la tabla 5-7.

Tabla 5-6: Estimación a posteriori de la probabilidad conjunta de observar niveles de Mercurio asociado a una de las tres categorías en cada punto de muestreo

Punto	Nivel de experticia del experto	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$\hat{\theta}_4$
La Cristalina	Alto	0,689	0,179	0,010	0,122
	medio	0,683	0,177	0,011	0,129
	Bajo	0,685	0,181	0,011	0,123
Alcides	Alto	0,689	0,261	0,010	0,039
	medio	0,684	0,261	0,011	0,044
	Bajo	0,684	0,264	0,011	0,041
Pablo Balanta	Alto	0,771	0,179	0,010	0,039
	medio	0,768	0,176	0,011	0,044
	Bajo	0,765	0,182	0,011	0,042
La escuela	Alto	0,771	0,179	0,010	0,040
	medio	0,768	0,177	0,011	0,044
	Bajo	0,767	0,180	0,011	0,041
Palo de yuca	Alto	0,958	0,028	0,003	0,011
	medio	0,958	0,026	0,003	0,013
	Bajo	0,956	0,029	0,003	0,012
La gloria	Alto	0,605	0,262	0,010	0,122
	medio	0,599	0,261	0,011	0,129
	Bajo	0,602	0,264	0,011	0,122
Erley	Alto	0,607	0,261	0,011	0,122
	medio	0,599	0,262	0,011	0,129
	Bajo	0,603	0,263	0,011	0,123
Tres Quebradas	Alto	0,688	0,096	0,093	0,122
	medio	0,683	0,093	0,095	0,129
	Bajo	0,684	0,100	0,093	0,123

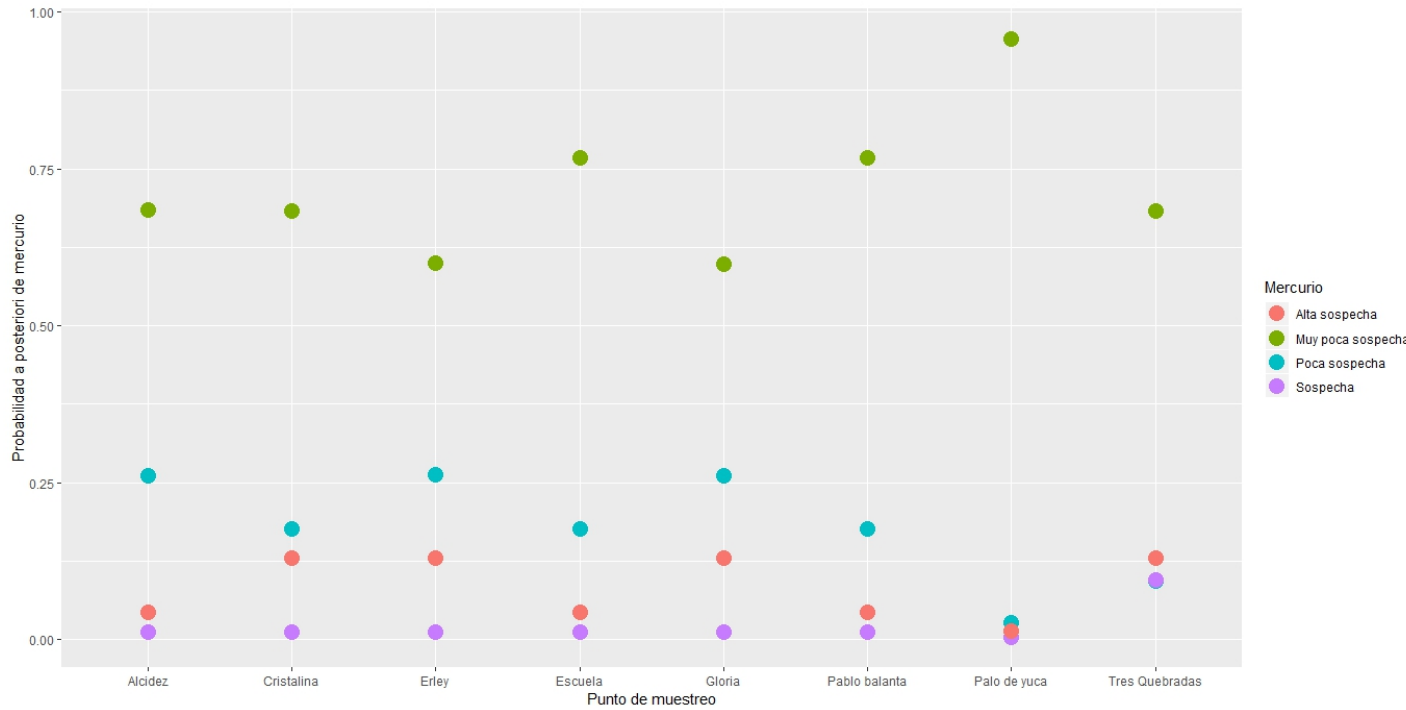
La tabla 5-6 y el gráfico 5-1 presentan las estimaciones de las probabilidades de observar los diferentes niveles de Hg en las fuentes de abastecimiento. Los resultados de los tres modelos evaluados llegan a resultados bastante similares. De acuerdo con los valores del FB obtenidos se decidió tomar el modelo cuya a priori se construyó asumiendo un experto con experticia hipotética media. Se puede pensar que los puntos La Cristalina, La Gloria, Erley, y Tres Quebradas son puntos que posiblemente presentan contaminación por Mercurio, dado que los valores estimados están por encima de 0.1 para la categoría identificada como alta sospecha de Mercurio reportada por las tirillas.

La tabla 5-7 presenta las regiones de credibilidad de 95 % para cada categoría, en cada uno de los puntos de muestreo, con la experticia media asignado al experto.

Tabla 5-7: Región de credibilidad para las estimaciones a posteriori de Hg para cada punto de muestreo.

ID sito	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$\hat{\theta}_4$
La Cristalina	(0,65-0,95)	(0,149-0,204)	(0,149-0,204)	(0,105-0,15)
Alcides	(0,65-0,715)	(0,23-0,292)	(0,23-0,292)	(0,03-0,058)
Pablo Balanta	(0,739-0,717)	(0,15-0,203)	(0,15-0,203)	(0,03-0,06)
Escuela	(0,735-0,796)	(0,15-0,204)	(0,15-0,204)	(0,03-0,06)
Palo de Yuca	(0,95-0,799)	(0,02-0,032)	(0,02-0,032)	(0,008-0,016)
Gloria	(0,563-0,965)	(0,229-0,294)	(0,229-0,294)	(0,104-0,155)
Erley	(0,562-0,636)	(0,23-0,293)	(0,23-0,293)	(0,105-0,152)
Tres Quebradas	(0,649-0,632)	(0,073-0,114)	(0,073-0,114)	(0,105-0,153)

El gráfico 5-1 muestra de forma visual la estimación de la probabilidad de Hg, para cada punto de muestreo según sea la clasificación de las tirillas.

**Figura 5-1:** Estimación a posteriori de Hg para cada punto de muestreo.

5.2. Colorimetría

La tabla **5-9** contiene los valores promedios calculados por el método de colorimetría mediante la ecuación 3-1, para cada punto y fecha de muestreo.

Tabla 5-8: Clasificación para los colores de las celdas

ID colores
W1 = Muy poca sospecha de Hg.
W2 = Poca sospecha de Hg.
W3= Moderada sospecha de Hg.
W4= Alta sospecha de Hg.

Tabla 5-9: Valores de colorimetría con clasificación de colores para cada W_i

Sitio ID	feb-27	mar - 07	mar - 21	abr - 04	abr - 18	may - 02
Tiras en blanco	0	0	0	0	0	0
La Cristalina	9,88	13,00	7,56	6,63	11,70	6,56
Alcides	9,58	12,98	6,51	13,67	1,93	3,73
Pablo Balanta	9,47	3,22	2,74	5,17	1,71	4,67
Escuela	2,70	2,11	6,99	11,25	1,72	2,80
Palo de Yuca	7,81	6,92	7,67	0,86	1,94	2,57
Gloria	9,24	9,33	8,09	2,40	7,96	4,90
Erley	7,86	11,20	7,35	9,10	10,96	4,80
Tres Quebradas	8,23	10,20	5,68	6,46	3,52	9,00

Los resultados encontrados se dividieron en dos grupos. El grupo 1 conformado por los puntos y fechas que marcaron muy poca y poca sospecha de Hg con las tirillas Sensafe, el grupo 2 conformado con las tirillas que marcaron sospecha y alta sospecha de Hg. Dicha división se realizó porque una categoría (sospecha de Hg) contaba con un solo valor y para que las pruebas de significancia de medianas sean robustas se necesitan en cada grupo de niveles de Mercurio mínimo cinco observaciones. A cada grupo se le realizó un análisis descriptivo con los valores obtenidos de la colorimetría ver tabla **5-10**, en esta podemos analizar que los coeficientes de variación en ambos grupos son muy altos, lo cual refleja la alta variabilidad que estas muestras presentaron con el método de colorimetría, especialmente el grupo 1, dado que presentó el cambio de color más pequeño (0.8634) y el más alto (13.6713). Sin embargo, la media y el tercer cuartil en ambos grupos son similares.

Tabla 5-10: Estadísticas descriptivas grupos 1 y 2

Grupo	Grupo 1	Grupo 2
Mínimo	0.8634	4.901
Q_1	3.008	5.678
Median	6.869	8.997
Mean	6.5958	7.710
Q_3	9.2823	9.098
Max	13.6713	9.875
Var	12.794	5.073
CV	54.231	29.213

Después de analizar los datos descriptivos, se planteó la prueba no paramétrica U de Mann Withney, para contrastar si la mediana en los grupos son diferentes, es decir, evidenciar estadísticamente si estas diferencias de color son lo suficientemente grandes para determinar o no la presencia de Mercurio. Para ello se estableció la siguiente hipótesis con una significancia del 5 %.

H_0 : La estimación mediana de Hg en el grupo 1 es igual a la del grupo 2.

H_1 : La estimación mediana de Hg en el grupo 1 es diferente a la del grupo 2.

Esta prueba arroja un valor-p de 0,4694. Lo que permite llegar a la conclusión de que no existe suficiente evidencia para rechazar la hipótesis nula con una significancia de $\alpha = 5\%$, es decir, que las diferencias de los grupos 1 y 2 no son significativas. Afirmando así que la colorimetría no permitió identificar si existe o no presencia de Mercurio de acuerdo a las cantidades presentadas en la tabla **5-3**.

Teniendo en cuenta los resultados previos podemos denotar que la colorimetría en este caso, no es un aporte relevante para detectar los niveles de Hg en las tirillas. Dado que para el punto Tres Quebradas, que es un caso comprobado en laboratorio de la presencia de Mercurio, no se encontró una tendencia de color significativa que lo diferenciara de los demás puntos.

5.3. Espectrofotometría de absorción atómica por vapor frío.

Las 10 muestras analizadas con este método fueron los puntos: Cristalina, Alcidez, Gloria, Erley, y Tres Quebradas en las fechas 27 de febrero y 21 de marzo del 2018. Estas muestras fueron escogidas aleatoriamente, y se encontraron los siguientes resultados.

Tabla 5-11: Resultados del patrón de oro vs las tirillas (Sensafe)

		Espectrofotometría		
Muestra de agua		Si	No	
Tirillas	Si	1	1	2
	No	1	7	8
		2	8	10

De esta se observa que de las 10 muestras, las tirillas coincidieron con la espectrofotometría 1 de 2 para las muestras que marcaron cantidades de Mercurio superior a 1ppb, que es lo establecido por la norma, y coincidieron 7 de 8 para las muestras que marcaron cantidades de Mercurio inferior a lo establecido por la norma Colombiana.

Ahora, haciendo uso de la estimación Bayesiana presentada en la metodología de la sección 4.3, se tiene las siguientes distribuciones beta para las tirillas que coincidieron con el patrón de oro, que son las variables A_1 y A_4 :

$$\begin{cases} \text{Beta}(2, 2) & \text{para la variable } A_1 \\ \text{Beta}(8, 2) & \text{para la variable } A_4 \end{cases} \quad (5-1)$$

De estas distribuciones beta en 5-1 se generaron 1000 muestras aleatorias y se obtuvieron las proporciones esperadas de tirillas que clasificaron correctamente cuando hay Mercurio iónico en un punto de muestreo, que fue de 49.7% con una región de credibilidad del 95% de (0.14 - 0.83).

Mientras que la proporción esperada de tirillas que clasifican correctamente cuando no hay Mercurio iónico en un punto de muestreo, es de 79.8% con una región de credibilidad del 95% de (0.57 - 0.95). Esto quiere decir que de 100 veces que se tome muestra en un punto, la tirilla y el patrón de oro determinarán que la presencia de Mercurio es inferior a la norma 79 veces.

Con esto se concluye que las tirillas tienen una buena capacidad para identificar lugares donde no hay Mercurio. Sin embargo, la capacidad de identificar lugares con Mercurio es reducida.

Finalmente se estimo la proporción de fuentes de agua de la zona con niveles de Mercurio mayor a los aceptados en Colombia para consumo humano, como fue descripto en la metodología de la sección 4.3, obteniendo así una proporción de fuentes estimadas con altos niveles de Mercurio de 24.7% y una región de credibilidad del 95% de (0.1 - 0.46).

Las dos muestras del río Tres Quebradas marcaron una cantidad de Hg superior a lo establecido por la resolución 2115 (2007), mientras que los otros cuatro puntos no evidenciaron Mercurio superior a la norma según la prueba de patrón de oro.

Capítulo 6

Conclusiones y recomendaciones

En este trabajo se desarrolló una propuesta que permite obtener métodos de clasificación en escala ordinal (niveles de Presencia - Ausencia) cuando se cuenta con instrumentos que poseen una muy baja capacidad de detección obligando al uso de k -plicados. El método propuesto utiliza conceptos básicos de la teoría de la probabilidad y métodos Bayesianos para aproximar el nivel de presencia de la característica de interés en las unidades muestrales.

En muchas situaciones es difícil obtener resultados de una prueba Gold Standard porque la misma puede ser muy costosa en su aplicación. En esos casos, el método propuesto podría funcionar de manera adecuada aumentando el número de los k -plicados.

El uso de la metodología propuesta fue ilustrado con la evaluación del desempeño en campo de unas tirillas colorimétricas, cuyo uso comercial fue adaptado por ingenieros de alimentos e ingenieros sanitarios para la detección de Hg en las fuentes de agua.

Al no contar con un experto o estudios estadísticos anteriores sobre la contaminación de Hg en aguas superficiales se propuso utilizar el Método Empírico de Bayes. Sin embargo, la poca variabilidad obtenida en cada punto de muestreo dificultó el procedimiento de simulación, dado que el MEB utiliza los datos para encontrar información, por tal motivo la credibilidad asignada al experto en los niveles alto, medio y bajo presentó hiperparámetros similares en esta investigación.

La prueba de laboratorio espectrofotometría de absorción atómica por el método de vapor frío, que puede detectar Mercurio iónico (divalente), mostró ser el mejor aliado al momento de confirmar la presencia de Mercurio en las fuentes de agua, debido a que solo cuenta con censura por izquierda de 1ppb lo cual concuerda con el límite permitido por Minambiente para el consumo humano. Esto ayuda de forma positiva al momento de determinar cuáles son las fuentes de agua que pudieran contener Mercurio, sin necesidad de anular una posible presencia menor de Mercurio en la sensibilidad del método. Sin embargo, hay que reconocer que la comunidad no cuenta con los recursos necesarios para acceder a este método pese a

ser muy efectivo.

Se considera que el método propuesto presenta una limitación cuando se pretende hacer validación del procedimiento de clasificación utilizando datos de triplicados de muestras obtenidas en campo. Cuando estos triplicados son asumidos como triplas de observaciones Bernoulli, puede ocurrir que se requieran muchas repeticiones de muestreos para obtener vectores con la variabilidad suficiente como para simular datos binarios con estructura de dependencia. En la aplicación con los datos de Hg observamos que con seis fechas de muestreos y $k = 3$ (Triplicados) no fue posible construir vectores con la variabilidad requerida. Consideramos que esta limitación podría ser manejada aumentando el número de replicas para una misma muestra ($k > 3$) aunque este resultado podría estar mas asociado a la baja capacidad que tenían las tirillas para establecer la presencia de mercurio en las fuentes de agua.

Apéndice A

Anexo: Código R

```
1 rm(list = ls())
3 set.seed(1234)
  alpha=c(.20,.40,.60)
5 nmc= c(381, 268, 206)

7 ##proporciones paso 2: vectores de alfa
  w1=34/48
9 w2=9/48
  w3=1/48
11 w4=4/48

13 ws=c(w1,w2,w3,w4)

15 ##Paso 3:

17 sim1= rmultinom(1000, 381,ws)/381
  sim2= rmultinom(1000, 268,ws)/268
19 sim3=rmultinom(1000, 206,ws)/206

21 #sim~Dirichlet (alpha_1, alpha_2, alpha_3, alpha_4)

23 #estimaciones
  set.seed(1234)
25 #de sim1
  medias1<-(rowMeans(sim1)/sum(rowMeans(sim1)))
27 varianzas1<-sapply(1:4, function(x) var(sim1[x,]))

29 alfas1<- (((medias1 -medias1^2)/varianzas1)-1)*medias1
  media.alfa1<-alfas1/sum(alfas1)
31 var.alfa1<-(alfas1*(sum(alfas1)-alfas1))/((sum(alfas1)+1)*sum(alfas1)^2)
  resultados1<-data.frame(Alfa_1=alfas1,Media_1=media.alfa1,Varianza_1=var.alfa1
    )
```

```

33 #de sim2
35 medias2<-rowMeans(sim2)/sum(rowMeans(sim2))
   varianzas2<-sapply(1:4, function(x) var(sim2[x,]))
37
   alfaso2<-(((medias2 -( medias2^2))/varianzas2)-1)
39 alfas2<- (((medias2 -( medias2^2))/varianzas2)-1)*medias2
   media.alfa2<-alfas2/sum(alfas2)
41 var.alfa2<-(alfas2*(sum(alfas2)-alfas2))/((sum(alfas2)+1)*sum(alfas2)^2)
   resultados2<-data.frame(Alfa_2=alfas2,Media_2=media.alfa2,Varianza_2=var.alfa2
   )
43
45 #de sim3
   medias3<-rowMeans(sim3)/sum(rowMeans(sim3))
47 varianzas3<-sapply(1:4, function(x) var(sim3[x,]))
49
   alfaso3<-(((medias3 -( medias3^2))/varianzas3)-1)
   alfas3<- (((medias3 -( medias3^2))/varianzas3)-1)*medias3
51 media.alfa3<-alfas3/sum(alfas3)
   var.alfa3<-alfas3*(sum(alfas3)-alfas3)/((sum(alfas3)+1)*sum(alfas3)^2)
53 resultados3<-data.frame(Alfa_3=alfas3,Media_3=media.alfa3,Varianza_3=var.alfa3
   )
   resultados2
55 resultados1
   data.frame(alfas1,alfas2,alfas3)
57
#####Puntos
59 library(coda)
   require(MCMCpack)
61 library(MCMCpack)
   z1=34
63 z2=9
   z3=1
65 z4=4
   zs<-c(z1,z2,z3,z4)

```

Referencias

- Barajas, L. E. N. (2018). *Métodos estadísticos bayesianos* (Programa de clases).
- Bromaghin, J. (1993). Sample size determination for interval estimation of multinomial probabilities. *The American Statistician*, 47, 203-206(3).
- Connor, R., y Mosimann, J. (1969). Conceptos de independencia para proporciones con una generalización de la distribución de dirichlet. *Diario de la Asociación Americana de Estadística*, 64(325), 194–206.
- Correa, J. C., y Barrera, C. J. (2019). *Introducción a la estadística bayesiana* (1.^a ed.). Instituto Tecnológico Metropolitano: Fondo editorial ITM.
- CRC. (2017). *Informe técnico en atención a acciones de protección y vigilancia de los recursos naturales y ambiente, atención a denuncia impuesta por la secretaria de gobierno del municipio de suarez ©, a través de oficio sgm-01-17-02-2017, presunta contaminación por cianuro en la quebrada yolombó que desemboca en el río ovejas y luego al río cauca.* (Informe técnico).
- Flórez, A. F. (2014). *Elicitación de una distribución subjetiva del vector de parámetros Π de la distribución multinomial* (Tesis).
- Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A., y Rubin, D. (2004). *Bayesian data analysis* (3.^a ed.).
- González-E, M., Bodas-Pinedo, M.-G., A., J., M., Trasobares-Iglesias, E. M., P., Bermejo Barrera, y Ordóñez-Iriarte, . P., J. M. (2014). *Metilmercurio: recomendaciones existentes. métodos de análisis e interpretación de resultados. evaluación económica. nutrición hospitalaria* (Informe técnico).
- Granada, N., y Zertuche, I. (2019). *Caracterización de la contaminación por mercurio en fuentes hídricas usando dispositivos de detección in-situ como herramienta de alerta social temprana en el corregimiento de la toma, departamento del cauca* (Tesis). Universidad del Valle.
- Hossain, M., Habib, I., y Andersson, K. (2017). A belief rule based expert system to diagnose dengue fever under uncertainty. En *2017 computing conference* (p. 179-186).
- Icontec. (2017). *Calidad del agua, método para determinar mercurio por espectrofotometría de absorción atómica con vapor frío* (Manual).
- Juarez, F. (2016, 01). Illegal mining in colombia: a conflict of narratives. , 16, 135-146.
- Kuboyama, K., Sasaki, N., Nakagome, Y., y Kataoka. (2005). Digestión húmeda. *Analytical Chemistry* 184-191.
- Mann, H., y Whitney, D. (1947). On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics*, 18(1), 50–60.
- Ministerio de la protección social y Ministerio de ambiente , vivienda y desarrollo territorial. (2007). *Resolución 2115: Por medio de la cual se señalan características, instrumentos básicos y frecuencias del sistema de control y vigilancia para la calidad del agua para*

- consumo humano* (Resolución gubernamental).
- Méndez. (2017). *Propuesta metodologica para construir reglas de clasificacion: Caso de aplicacion dengue. director: Rafael tovar* (Trabajo de grado).
- Naranjo, J. L. M. (2017). *Uso de la distribución Lévy para ajustar datos con marcada asimetría y valores extremos* (Tesis). Universidad Nacional de Colombia.
- Oakley, Daneshkhah, y O'Hagan. (2010). Nonparametric prior elicitation using the roulette method. School of Mathematics and Statistics, University of Sheffield, UK.
- Park, C., Park, T., y Shini, D. (1996). A simple method for generating correlated binary variates. the american statistician. *The American Statistician*, 50(4)(306-310).
- Quiñones, K., y Suárez, L. (2018). *Validación externa de prototipos clasificadores en un caso aplicado: Dengue* (Tesis). Escuela de Estadística, Universidad del Calle, Colombia.
- R Core Team. (2018). R: A language and environment for statistical computing [Manual de software informático]. Vienna, Austria.
- Ruiz Chaves, I. (2016). *Metodologías analíticas utilizadas actualmente para la determinación de mercurio en músculo de pescado* (Pensamiento actual, 16, No.26.)
- Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin* 1(80-83).