

**UN CASO DE ESTUDIO SOBRE LAS APLICACIONES DE LOS MODELOS DE REGRESIÓN LINEAL: ANÁLISIS DE TRATAMIENTOS PARA POTABILIZACIÓN DEL AGUA MEDIANTE MODELOS LINEALES GENERALIZADOS, PARTE I**

**Isadore Nabi**

<b>I. INTRODUCCIÓN</b>	<b>2</b>
<b>II. GENERALIDADES</b>	<b>3</b>
<b>II.I. Configuración General del MLG</b>	<b>10</b>
II.I. I. Variables Independientes/Explicativas	10
III.I. II. Variables de Respuesta	10
<b>III. MODELOS LINEALES GENERALIZADOS OBTENIDOS</b>	<b>10</b>
<b>III.I. MLG 1: Dosis Óptima de Sulfato</b>	<b>10</b>
III.I. I. Forma General	10
III.I. II. Forma Particular	11
<b>III.II. MLG 2: Concentración Óptima de Sulfato</b>	<b>12</b>
III.II. I. Forma General	12
III.II. II. Forma Particular	12
<b>III.III. MLG 3: Dosis Óptima de Lima</b>	<b>13</b>
III.III. I. Forma General	13
III.III. II. Forma Particular	13
<b>IV. ANÁLISIS SOBRE LA VERIFICACIÓN Y CUMPLIMIENTO DE LOS SUPUESTOS DE LOS MLG EN (Cepeda &amp; Cepeda C., 2005)</b>	<b>13</b>
<b>IV.I. Supuesto 1: Todas las mediciones han sido realizadas bajo la misma escala física</b>	<b>13</b>
<b>IV.II. Supuesto 2: Las variables explicativas son linealmente independientes</b>	<b>14</b>
<b>IV.III. Supuesto 3: Las variables explicativas son independientes del valor esperado de la variable de respuesta</b>	<b>14</b>
<b>IV.IV. Supuesto 4: La variable de respuesta debe pertenecer a alguna familia exponencial</b>	<b>14</b>
<b>IV.V. Supuesto 5: Existe una relación lineal entre la variable de respuesta transformada (mediante la función enlace) y las variables independientes</b>	<b>14</b>
<b>IV.VI. Supuesto 6: Existe independencia entre los términos del error</b>	<b>14</b>
<b>IV.VII. Supuesto 7: La muestra es lo suficientemente grande</b>	<b>15</b>
<b>V. REFERENCIAS</b>	<b>16</b>

## I. INTRODUCCIÓN

La presente investigación es la primera parte de un conjunto dual de investigaciones que expresa la continuidad natural de (Nabi, Modelos Lineales Generalizados, 2021) y (Nabi, Supuestos del Modelo Clásico de Regresión Lineal y de los Modelos Lineales Generalizados, 2021), por cuanto en dicho conjunto se estudiarán las siguientes cuestiones:

1. La verificación de los supuestos de los modelos lineales generalizados en la práctica científica.
2. Su aplicación en la predicción de medias condicionales (los valores esperados de la variable de respuesta).
3. La estimación empírica de las probabilidades de ocurrencia de un determinado evento dado algún modelo lineal generalizado.

Esta primera parte de la investigación abordará los puntos 1 y 2 antes mencionados, mientras que en la segunda parte se abordará el punto 3. Para ello, en ambas investigaciones se tomará como referencia la investigación (Cepeda & Cepeda C., 2005), la cual versa sobre los tratamientos para la potabilización del agua en Colombia.

Finalmente, en cuanto a si los supuestos fueron cumplidos empíricamente hablando y/o si fueron verificados por los investigadores, debe decirse que en los casos en que no sea posible verificarlos empíricamente aquí por la falta de disponibilidad de datos, se planteará sintéticamente cómo realizar la verificación de cada uno de estos supuestos desde la teoría estadística.

## II. GENERALIDADES

Como se señala en (Cepeda & Cepeda C., 2005, pág. 233), en el tratamiento para obtener agua potable encontramos modelos lineales para explicar dosis y concentración de sulfato, dosis de cal, probabilidades de aplicación de cal y de polímero, incluyendo como variables explicativas algunas propiedades físicas y químicas del agua cruda (el agua que aún no ha sido procesada). El estudio fue desarrollado en una planta de tratamiento de agua localizada en Girardot (Colombia), cuya fuente hídrica es el río Magdalena. El tratamiento de agua comprende:

1. *Aireación del agua.* Como se verifica en (Organización Mundial de la Salud, 2009, pág. 1), es un proceso de tratamiento en el que el agua entra en contacto con el aire con el propósito primordial de incrementar su contenido de oxígeno. manganeso, para que formen precipitados, que se pueden retirar por decantación y filtración.
2. *Coagulación del agua.* Como se verifica en (Bermúdez Cabrera, Fleites Ramírez, & Contreras Moya, 2009, pág. 65), el proceso de coagulación-floculación consiste en añadir al agua o agua residual determinados aditivos químicos con el objetivo de favorecer la sedimentación de materia coloidal no sedimentable o aumentar la rapidez de sedimentación por la formación de flóculos. Los flóculos son una masa coagulada de partículas en un líquido. Los flóculos pueden existir de forma natural, pero a menudo se generan a partir de un sistema coloidal [como señala (Domènech & Peral, 2006, pág. 40) un sistema conformado por dos o más fases, normalmente una fluida -líquido o gas- y otra dispersa en forma de partículas generalmente sólidas muy finas, de diámetro comprendido entre  $10^{-9}$  y  $10^{-5}$  m] disperso al que se le adiciona una sustancia química floculante. Las partículas de arcilla y los polímeros en agua pueden flocularse para formar flóculos. Cabe resaltar que coagulación y floculación no son equivalentes, puesto que, como señala (Pérez de la Cruz & Urrea Mallebrera, 2021, pág. 21), el coagulante anula las fuerzas repulsivas entre las partículas coloidales, iniciando la formación de microflóculos, en cambio el floculante engloba estos microflóculos aumentando su tamaño y densidad de modo que sedimenten más fácil y rápidamente.
3. *Sedimentación.* La sedimentación es el proceso por el cual los sólidos que se encuentran en suspensión en el agua caen al fondo del recipiente donde el agua esté contenida<sup>1</sup>.

---

<sup>1</sup> La sedimentación, no como proceso de limpieza de sedimentos del agua sino como característica del agua cruda, puede ser medida en términos de los miligramos que poseen como masa las partículas varias que están suspendidas en el fluido (en este caso el agua). Existen aparatos

4. *Filtración*. Como se señala en (Organización Mundial de la Salud, 2009, pág. 2), la filtración es el paso de agua contaminada a través de un medio poroso (como la arena). El proceso usa el principio de limpieza natural del suelo.
5. *Desinfección*. Como se señala en (Organización Mundial de la Salud, 2009, pág. 4), es esencial que el agua de consumo esté libre de organismos nocivos. El almacenamiento, la decantación y la filtración del agua reducen el contenido de bacterias nocivas, aunque ninguno de ellos puede garantizar la completa remoción de los gérmenes. La desinfección es un proceso de tratamiento que garantiza que el agua potable quede libre de organismos o patógenos perjudiciales. Se recomienda que ésta sea la etapa final del tratamiento, pues muchos de los procesos de desinfección se dificultan por las partículas sólidas y el material orgánico suspendidos en el agua.

En la investigación se usan tanto compuestos químicos como físicos para la desinfección del agua<sup>2</sup>.

En este procesamiento del agua, las partículas coloidales son separadas del agua por medio de un proceso de coagulación química que consiste en la desestabilización de la carga de las partículas en suspensión mediante la adición de una sustancia coagulada, generalmente sulfato de aluminio. La coagulación química se realiza por varias razones, principalmente para eliminar la turbidez orgánica e inorgánica; color, verdadero y aparente; bacterias dañinas y otros patógenos; plancton y dureza. Como práctica común, el sulfato de aluminio se aplica en una cantidad determinada experimentalmente mediante la prueba de frascos.

Como señalan (Cepeda & Cepeda C., 2005, pág. 234), dado que el agua cruda se toma de un río, se necesitan mediciones frecuentes de variables como el nivel de turbidez, el color y la alcalinidad para realizar la prueba de los frascos para determinar la dosis de coagulado. Los costos de tiempo y operación hacen deseable utilizar un modelo estadístico para establecer las dosis óptimas de sulfato, cal y polímero a aplicar al agua, con base a turbidez, color, temperatura, pH, alcalinidad y dureza. Las variables en cuestión responden a los siguientes intervalos de variación:

---

especiales para medirla conocidos como tensiómetros, así como también una clasificación de los tipos de partícula suspendidos en las soluciones. Lo anterior puede verificarse en <https://www.biolinscientific.com/measurements/sedimentation> y [https://www.who.int/water\\_sanitation\\_health/resourcesquality/wqmchap13.pdf](https://www.who.int/water_sanitation_health/resourcesquality/wqmchap13.pdf), respectivamente.

<sup>2</sup> La clasificación de compuestos físicos y químicos en la desinfección del agua puede consultarse en <https://www.lenntech.es/procesos/desinfeccion/que-es-desinfeccion.htm>.

1. *Turbidez*: 200-1200 NTU<sup>3</sup>. El agua es considerada potable si NTU=1, mientras que es no-potable si NTU>5.
2. *Color*: 50-250 UC<sup>4</sup>. En aguas tratadas es ideal que UC=5 y si UC>15 entonces el agua se considera no-potable. Como se señala en (Abril Díaz, y otros, 2021, págs. 1-4), el color es una cualidad cuya determinación obedece a clasificarlo en función de una cantidad resultante tras realizar una medición determinada sobre una biomolécula, específicamente aquella medición relativa a la cantidad de luz que una biomolécula es capaz de absorber; los criterios de clasificación están determinados ex ante por el marco teórico de la Química relativo a ello. El proceso de medición del color requiere del método analítico (relativo al análisis de concentración de compuestos químicos) conocido como espectrofotometría UV-visible, el cual permite determinar la concentración de un compuesto en solución. Esto es posible por el hecho de que las moléculas absorben las radiaciones electromagnéticas y su estimación se simplifica por el hecho fisicoquímico de que la cantidad de luz absorbida depende de forma lineal de la concentración. Para hacer este tipo de medidas se emplea un espectrofotómetro, que es un instrumento de medición en el que se puede seleccionar la longitud de onda de la luz que pasa por una solución y medir la cantidad de luz absorbida por la misma<sup>5</sup>, pues de esta cantidad depende la clasificación del color (como cualidad<sup>6</sup>) de la biomolécula. Otra forma de medir el color es a través de sus unidades Pt-Co (Platino-Cobalto), sin embargo, este método es menos riguroso que el anteriormente mencionado,

---

<sup>3</sup> NTU significa unidad de turbidez nefelométrica, es decir, la unidad utilizada para medir la turbidez de un fluido o la presencia de partículas en suspensión en el agua. Cuanto mayor sea la concentración de sólidos en suspensión en el agua, más sucia se verá y mayor será la turbidez.

<sup>4</sup> UC significa "Units of Color".

<sup>5</sup> La longitud de onda, como señala (Wikipedia, 2021), es la distancia que recorre una perturbación periódica que se propaga por un medio en un ciclo. La longitud de onda, también conocida como período espacial es la inversa de la frecuencia multiplicada por la velocidad de propagación de la onda en el medio por el cual se propaga.

<sup>6</sup> La cantidad acumulada genera la cualidad, pero a la vez, la cantidad está subordinada a la cualidad (puesto que aún en el escenario de ruptura topológica, simplemente estará subordinada a una nueva cualidad). Lo que parece una contradicción lógica se explica categóricamente por el hecho de que con independencia del salto de lo cuantitativo a lo cualitativo (por ejemplo, en las transiciones de fase), la cualidad es la cantidad organizada y las cantidades organizadas dependen de la totalidad que conforman colectivamente. Esto se expresa nítidamente en las distribuciones de probabilidad, en donde, aunque los números sean sus elementos conformantes, la localización de dichos números (i.e., su estado dentro del sistema de probabilidades) está determinada por la estructura de la distribución, precisamente a pesar de que esta estructura haya sido generada mediante números. La totalidad es, en distintos estados cualitativos de la materia, más que la conjunción simple de sus partes.

puesto que se realiza mediante una comparación meramente visual con una escala preestablecida<sup>7</sup>.

3. *pH*: 7.1-8.33. En aguas tratadas es ideal que el pH oscile entre 7 y 8.5. Sin embargo, un intervalo entre 6.5 y 9 también es considerado aceptable. El pH es una medida de alcalinidad o acidez de una solución, más específicamente el pH mide la cantidad de iones de hidrógeno que contiene una solución determinada. Sus siglas significan *potencial de Hidrogeniones (pH)*. Este indicador se ha convertido en una forma práctica de manejar cifras de alcalinidad, en lugar de otros métodos un poco más complicados. e indica si una sustancia es ácida o básica. Adicionalmente, nótese que es un indicador adimensional (sin unidades de medida); sin embargo, una escala de clasificación de sustancias para determinar su pH, como puede verificarse en (Asociación Española de Fabricantes de Agronutrientes, 2021).
4. *Alcalinidad*: 43-55. En aguas naturales, la alcalinidad total del agua expresa el contenido de hidróxidos (OH), carbonatos (CO<sub>3</sub>) y bicarbonatos (HCO<sub>3</sub>). En términos generales, el agua tratada no debe tener valores demasiado bajos para producir corrosión ni demasiado altos para producir incrustaciones. La alcalinidad se mide en mg/L de carbonato de calcio.
5. *Dureza*: 43-50. La dureza del agua para consumo doméstico debe estar entre 30 y 50 mg/L de CaCO<sub>3</sub>. En la investigación que aquí se estudia, la dureza del agua solo se incluye como variable explicativa en el modelo lineal que explica la concentración óptima de sulfato de aluminio. La dureza es también en mg/L, asumieron que venían de atrás las unidades.
6. *Temperatura*: 24-28 °C. Según la normativa francesa, la temperatura del agua debe estar entre 9 y 12 °C. Una temperatura superior a 15 °C favorece el desarrollo de microorganismos en los canales e intensifica los olores y sabores.

En la investigación analizada se utilizan las denominadas *variables de dosis*, que son las dosis óptimas de sulfato de aluminio, polímero y lima necesarias para potabilizar el agua, según un reporte presentado por la Asociación Estadounidense de Salud Pública en 1995.

- *Dosis óptima de sulfato de aluminio (D\_SUL)*

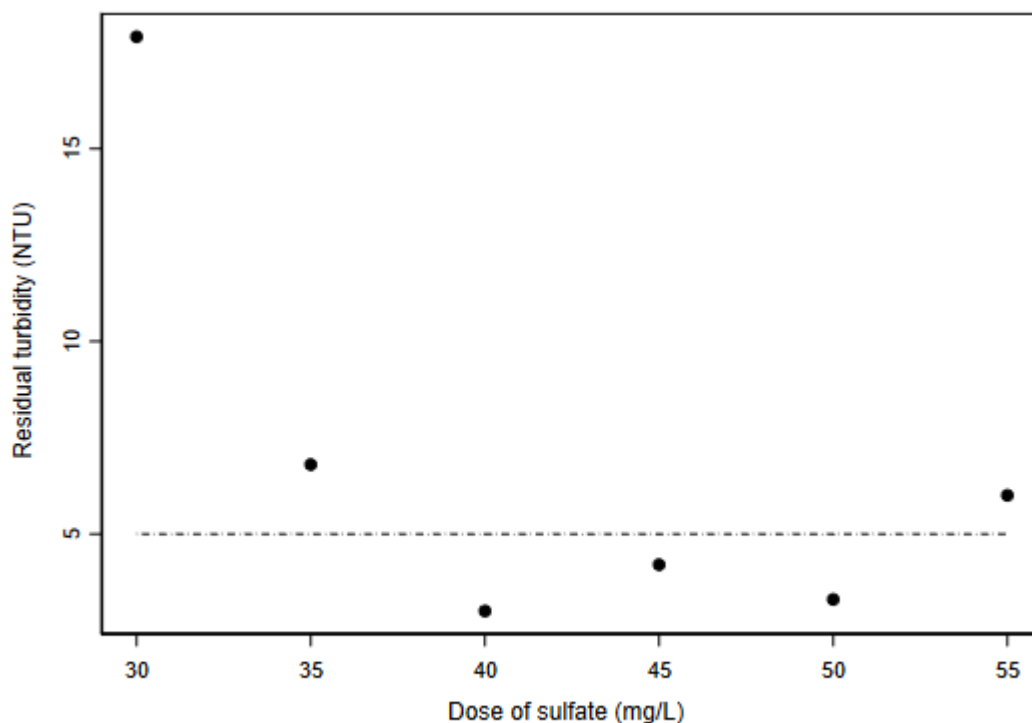
Para calcular la dosis requerida de coagulador, se deben realizar pruebas de laboratorio utilizando la prueba de frascos. Para medir la dosis óptima de sulfato de aluminio se utiliza un equipo con agitador mecánico de 4 a 6 paletas, que opera

---

<sup>7</sup> Evidentemente, el problema no es la escala preestablecida, sino que la comparación se limite a ser visual. Una unidad de Platino-Cobalto es la que se produce al disolver en la sustancia (agua, en este caso) un miligramo de platino/L en forma de ion cloroplatinato, que es su forma de disolución química. Véase <http://legismex.mty.itesm.mx/normas/aa/aa045-01.pdf>.

a velocidades de 0 a 100 r.p.m., un iluminador flóculos es colocado en la base del agitador y vasos de precipitación de 2000 ml de capacidad. La dosis óptima de sulfato de aluminio se obtiene añadiendo coagulado al agua en dosis progresivas en cada vaso de precipitación. Las muestras se inyectan en los tarros agitadores mecánicos ajustados a 100 r.p.m. Si se requiere alcalinización del agua, debe realizarse antes de la prueba de coagulación. Después de 30 segundos a 2 minutos de mezcla rápida, comenzando con la aplicación del coagulado, la velocidad se reduce a la velocidad seleccionada como la condición óptima de floculación y luego, las muestras se floculan durante cinco a veinte minutos. Posteriormente se suspende la agitación, se retiran las paletas y se sedimenta el agua durante cinco a diez minutos. Se toman muestras de cada uno de los frascos y se mide la turbidez. La dosis óptima de sulfato de aluminio se obtendrá del frasco donde se encontraron los valores más pequeños de turbidez y color. Esta dosis, denotada por D\_SUL, se tomará como la respuesta cuantitativa para el análisis del modelo lineal.

Figura 6



Dose of sulfate effect on residual turbidity.

Fuente: (Cepeda & Cepeda C., 2005, pág. 236).

La figura 6 muestra el comportamiento de la turbidez residual en función de la dosis de sulfato de aluminio. El valor más bajo admisible de turbidez se obtiene mediante la aplicación de 40 mg/l de sulfato. Esta cantidad fue la mínima necesaria para que las condiciones de turbidez del agua sean aceptables para el consumo humano y se considere, para las condiciones iniciales del agua, la dosis óptima a aplicar como valor D\_SUL. Las condiciones iniciales del agua empleada para la determinación de los datos, fueron las siguientes: turbidez 200 NTU, color 100UC, pH 8.03, alcalinidad 40 mg / l CaCO<sub>3</sub> y temperatura 24 °C.

- *Dosis óptima de polímero (D\_POL)*

Una vez realizada la prueba para determinar la dosis óptima de sulfato de aluminio y elegido un polielectrolito, la dosis óptima de polielectrolito se obtiene mediante la siguiente prueba: Se reduce la dosis de sulfato de aluminio (por ejemplo en el 75% de la dosis óptima) y una cantidad variable de polielectrolito, teniendo en cuenta las condiciones técnicas y recomendaciones de las entidades de salud, o manteniendo constante la dosis de polielectrolito y cambiando la cantidad de sulfato de aluminio para obtener el menor costo para que la dosis de aluminio/polielectrolito dé al agua condiciones aceptables para el consumo humano. Si la calidad del agua en la planta cambia con el tiempo, las pruebas deben proceder a nuevas condiciones de turbidez y/o color.

La determinación del índice Willcomb<sup>8</sup>, el tiempo de aparición de flóculos, la turbidez residual, el color residual verdadero y el pH residual serán necesarios para determinar la dosis óptima del polímero, indicada por D\_POL.

- *Dosis óptima de cal (D\_CAL)*

Para determinar la dosis óptima de cal se tomó una muestra de 1000 ml de agua tratada con dosis óptimas tanto de sulfato de aluminio como de polímero. Si el pH es menor a 7, se aplica una solución de cal para obtener este valor. La cantidad de cal necesaria para obtener un pH igual a 7 se denomina dosis óptima de cal, indicada por D\_CAL. La concentración óptima de sulfato de aluminio (C\_SUL) se considera una variable de dosificación y sus valores se determinan experimentalmente. Se realiza una prueba de frascos en la forma convencional y la dosis óptima de sulfato se determina de acuerdo con el procedimiento experimental descrito anteriormente. Esta prueba se realiza utilizando una solución coagulada de 10 g/L (1%)<sup>9</sup> con pH entre 4 y 4,5. En cinco vasos de precipitado con la misma cantidad de agua de muestra, se agrega la dosis óptima

---

<sup>8</sup> Es el índice que evalúa el tamaño y consistencia de los flóculos.

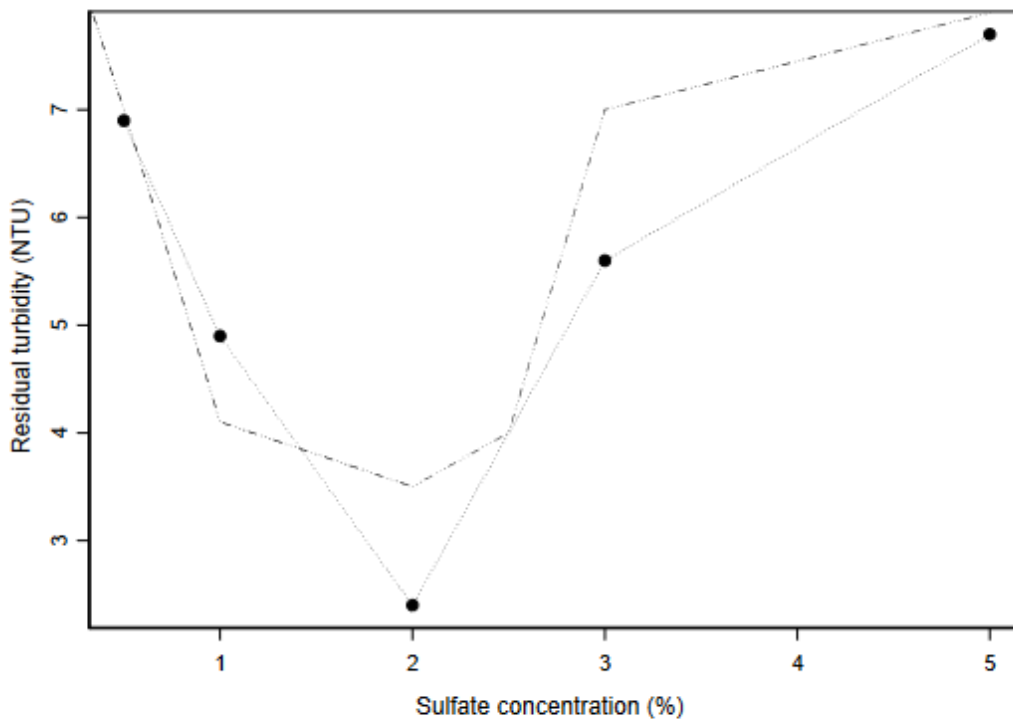
<sup>9</sup> 10 g/L es equivalente a 1g/ 0.1L, 1g/100 ml o 1 %



determinada en la prueba de frascos, con diferentes concentraciones de coagulador, tales como 10%, 5%, 1%, 0.5% y 0.1%.

Después de la mezcla, el agua se flocula y se deposita. Durante este proceso se evalúan tanto el tamaño del flóculo (índice de Willcomb) como la velocidad de reacción (tiempo inicial de formación del flóculo). A continuación, se determinan la turbidez y el color residual en el agua clarificada. La concentración aplicada en el frasco con la menor turbidez y color residual se llama concentración óptima.

Figura 7



Sulfate concentration effect on turbidity.

Fuente: (Cepeda & Cepeda C., 2005, pág. 237).

La figura 7 muestra el valor C\_SUL para algunas características del agua cruda. En esta figura también se muestra cómo la concentración en la aplicación de la dosis óptima de sulfato de aluminio influye significativamente en el comportamiento de la turbidez residual y cómo, para algunos valores de concentración de C\_SUL, en cada caso, la turbidez residual es menor. Esto corresponde al vidrio de la concentración óptima de aplicación de sulfato de aluminio. Las condiciones iniciales fueron: turbidez 500 NTU, alcalinidad 58 mg / l CaCO<sub>3</sub>, pH 7,38, temperatura 27 ° C.

En la planta de tratamiento de agua de Girardot, la medición de variables físicas, químicas y de dosis se realiza después de un proceso de sedimentación, en el que los materiales pesados se separan del agua por gravitación. A lo largo de 1997<sup>10</sup> los autores de la investigación midieron la turbidez, color, alcalinidad, pH, temperatura y dosis óptima de sulfato. Para determinar el modelo estadístico para explicar la dosis óptima de sulfato, seleccionamos medidas de 40 muestras de agua cruda, considerando la variabilidad de las propiedades de los datos del agua cruda. En 20 de estas muestras medimos la concentración óptima para la aplicación de sulfato. Durante el mismo año realizamos 33 experimentos para determinar la dosis óptima de cal para el tratamiento del agua. En cada uno de estos experimentos también medimos la turbidez, el color, la alcalinidad, el pH y la temperatura del agua cruda. Para determinar la probabilidad de aplicación de cal y polímero, se tomaron 60 observaciones de los registros de la planta de tratamiento de agua.

## ***II.I. Configuración General del MLG***

### *II.I. I. Variables Independientes/Explicativas*

*Turbidez:* TURB

*Color:* COL

*Alcalinidad:* ALCA

*pH:* PH

*Temperatura:* TEMP

### *III.I. II. Variables de Respuesta*

*Dosis de Sulfato:* D\_SUL

*Concentración Óptima de Sulfato:* C\_SUL

## **III. MODELOS LINEALES GENERALIZADOS OBTENIDOS**

### ***III.I. MLG 1: Dosis Óptima de Sulfato***

#### *III.I. I. Forma General*

$$D\_SUL = \beta_0 + \beta_1TURB + \beta_2COL + \beta_3ALCA + \beta_4PH + \beta_5TEM + \epsilon$$

En la expresión anterior,  $\epsilon \sim N(0, I\sigma^2)$ . Esto significa que la distribución de probabilidad del término de error de estimación se aproxima a una distribución normal multivariada con un vector de medias de magnitud nula y una matriz de covarianzas  $\sigma^2$ .

---

<sup>10</sup> Nótese que la investigación fue publicada, como se verifica en las referencias, en el año 2005.

### III.I. II. Forma Particular

(Cepeda & Cepeda C., 2005, pág. 238) establecen el modelo para los valores de los parámetros obtenidos por MLE mediante aproximaciones numéricas vía IRLS<sup>11</sup> que se presenta a continuación:

$$\widehat{D}_{SUL} = -13.571 + 0.085TURB + 1.592TEMP, R^2 = 0.991 \quad (1)$$

La razón por la cual no aparecen en la ecuación de regresión las demás variables obedece a que los autores usaron una significancia 0.01 para eliminar variables como parte del ajuste final del modelo, es decir, eliminaron todas aquellas variables cuyos parámetros no eran estadísticamente significativos a un nivel de confianza de 0.99 y se volvió a construir el modelo con las variables restantes; este proceso de regresoras se conoce *eliminación hacia atrás* o *backward elimination*<sup>12</sup>. Esto se justifica a juicio de los autores por el hecho que, además de que los valores p de los parámetros de las variables descartadas son todos menores que 0.01, no se pierde información estadísticamente relevante cuando se eliminan COL, ALCA y pH del modelo. Sin embargo, los autores no especifican bajo qué criterio concreto establecen que no existe pérdida de información al omitir tales variables.

Los valores p son todos menores que 0.01 y no se pierde información cuando se eliminan COL, ALCA y pH del modelo. El resultado (1) muestra que la dosis óptima de sulfato a aplicar en el tratamiento del agua depende básicamente de la turbidez y de las bajas temperaturas, en cuanto estas favorecen la acción coagulante. Si la temperatura tiene valores bajos, se requiere una dosis menor de sulfato.

Tabla 1

Table 1: Simple correlations

	TURB	COL	ALCA	PH	TEMP
D-SUL	0.978**	0.517**	-0.196	0.133	-0.266
TURB	1	0.485**	-0.119	0.156	-0.337*
COL		1	-0.448**	0.048	-0.046
ALCA			1	-0.207	0.181
PH				1	-0.408*

Fuente: (Cepeda & Cepeda C., 2005, pág. 239).

<sup>11</sup> Véase (Nabi, Modelos Lineales Generalizados, 2021).

<sup>12</sup> Véase <https://www.javatpoint.com/backward-elimination-in-machine-learning>.

La tabla 1 muestra las correlaciones lineales de Pearson entre las variables incluidas en el estudio. Los niveles de significancia bilateral (“a dos colas”) 0.01 y 0.05 están indicados por \*\* y \*, respectivamente. La tabla muestra la correlación más alta entre la dosis de sulfato y la turbidez con un valor de 0.978.

### **III.II. MLG 2: Concentración Óptima de Sulfato**

#### *III.II. I. Forma General*

$$C\_SUL = \beta_0 + \beta_1TURB + \beta_2COL + \beta_3ALCA + \beta_4PH + \beta_5TEM + \beta_6DUR + \epsilon$$

Con un  $R^2 = 0.530$  y con valores p mayores que 0.05 para la hipótesis parcial  $\beta_k = 0, 1, \dots, 6$ .

Casi seguramente el lector se estará preguntando en estos momentos en qué consiste una prueba de hipótesis parcial. Como se verifica en (Rahim, Budiantara, & Permatasari, 2019, pág. 41), se denominan como pruebas de hipótesis parcial a aquellas pruebas de hipótesis que, en contraste con las pruebas de hipótesis globales que prueban conjuntamente la significancia de los parámetros de un modelo, estiman la significancia de los parámetros tomados individualmente. Esto encuentra su fundamento teórico en nivel teórico en la investigación (Atil Husni, Nyoman Budiantara, & Zain, 2018, pág. 1), puesto que es en esta investigación teórica del Instituto Estadounidense de Física (AIP, por sus siglas en inglés) en que parece acuñarse este concepto teórico<sup>13</sup>. Posterior a la investigación del AIP aparecería la de Rahim, Budiantara y Permatasari antes citada y también la investigación (Li, Duan, Zhang, Li, & Ming, 2009), realizada por la Asociación de Lingüística Computacional (ACL).

#### *III.II. II. Forma Particular*

Como se señala (Cepeda & Cepeda C., 2005, pág. 239), después de realizar el mismo proceso de eliminación antes expuesto (no especifican los autores si esto lo realizaron también bajo un 0.05 o si realizaron la eliminación bajo un criterio de 0.01 como con MLG 1), los autores arriban a la siguiente ecuación de regresión

$$\hat{C}\_SUL = -1.597 + 0.445PH, \quad R^2 = 0.413$$

Este modelo tiene un valor p de intersección = 0.133 y un valor p de pendiente igual a 0.003. Podríamos considerar un modelo de intersección cero, pero los autores en el lugar citado expresaron su preferencia por el modelo con intercepto

---

<sup>13</sup> Aunque no se ha hecho una investigación lo suficientemente exhaustiva como para aseverar con total confianza que la génesis del concepto de prueba parcial de hipótesis se encuentra en esta investigación, sí se ha investigado lo suficiente para decir que casi seguramente esa investigación fue al menos una de las investigaciones pioneras.

no nulo, argumentando que es mejor predecir la concentración para un valor particular de pH en el intervalo [7.1, 8.33].

### **III.III. MLG 3: Dosis Óptima de Lima**

#### *III.III. I. Forma General*

$$D_{CAL} = \beta_0 + \beta_1TURB + \beta_2COL + \beta_3ALCA + \beta_4PH + \beta_5TEMP + \epsilon$$

#### *III.III. II. Forma Particular*

Como señalan los autores en el mismo lugar dentro de la fuente referida, tras realizar una eliminación hacia atrás como en los MLG 1 y 2, encontraron el siguiente modelo:

$$\hat{D}_{CAL} = 59.374 + 0.01262TURB + 0.07746COL - 8.31PH, \quad R^2 = 0.915$$

Para la ecuación de regresión anterior, todos los valores p fueron menores que 0.01, reportan los investigadores. Así, la dosis óptima de cal a aplicar en el tratamiento del agua depende básicamente de la turbidez, el color y el pH.

Todos los valores p son menores que 0.01. Así, la dosis óptima de cal a aplicar en el tratamiento del agua depende básicamente de la turbidez, el color y el pH.

## **IV. ANÁLISIS SOBRE LA VERIFICACIÓN Y CUMPLIMIENTO DE LOS SUPUESTOS DE LOS MLG EN (Cepeda & Cepeda C., 2005)**

Anteriormente, en la sección II.III de esta investigación, se estudiaron los supuestos de los MLG. En esta sección se analizará si en la investigación estudiada tales supuestos fueron cumplidos y/o verificados<sup>14</sup>, en caso de ser posible determinarlo.

### ***IV.I. Supuesto 1: Todas las mediciones han sido realizadas bajo la misma escala física***

Si bien esto no es rigurosamente cierto en esta investigación, tampoco lo es en ninguna otra. Sin embargo, no es un problema a causa de dos razones:

1. Existe un marco teórico-científico que establece, bajo determinadas condiciones, la existencia de equivalencia métrica entre las unidades de estudio.
2. A causa de lo anterior, las fórmulas utilizadas para realizar los cálculos conjuntos poseen una estructura formal que, en conjunto con las unidades de medición de las variables en cuestión (y las que no tienen unidades),

---

<sup>14</sup> Con excepción del supuesto 6, que no es propiamente un supuesto sino la expresión de que un supuesto usualmente requerido (y usualmente incumplido) por el MRLC ya no es necesario en el contexto de los MLG.

permiten la cancelación de términos de tal forma que, en última instancia, no existen problemas dimensionales con las variables de estudio.

***IV.II. Supuesto 2: Las variables explicativas son linealmente independientes***

Esto significa que alguna de ellas, cualquiera, no puede ser expresada como combinación lineal de las variables restantes (o un subconjunto de estas). Este supuesto no es verificado por los autores en ninguno de los modelos elaborados y los autores de la investigación no proporcionan los datos utilizados para que se verifiquen en esta investigación mediante una prueba chi-cuadrado (que es una prueba no-paramétrica) o mediante un análisis de correlaciones lineales de Pearson (que un indicador paramétrico) u otro tipo de correlación lineal (como las de Spearman y Kendall, que son no-paramétricas).

***IV.III. Supuesto 3: Las variables explicativas son independientes del valor esperado de la variable de respuesta***

Este supuesto no es verificado por los autores en ninguno de los modelos elaborados y estos no proporcionan los datos utilizados para que se verifiquen en esta investigación mediante prueba chi-cuadrado o mediante un análisis de correlaciones lineales de Pearson u otro tipo de correlación lineal (como las de Spearman y Kendall).

***IV.IV. Supuesto 4: La variable de respuesta debe pertenecer a alguna familia exponencial***

Este supuesto no es verificado por los autores en ninguno de los modelos elaborados y estos no proporcionan los datos utilizados para que se verifiquen en esta investigación mediante ajustes de distribución empíricos y mediante la prueba chi-cuadrado (que es utilizada también para determinar de forma más rigurosa si un conjunto de datos se ajusta o no a tal o cual distribución).

***IV.V. Supuesto 5: Existe una relación lineal entre la variable de respuesta transformada (mediante la función enlace) y las variables independientes***

Este supuesto no es verificado por los autores en ninguno de los modelos elaborados y estos no proporcionan los datos utilizados para que se verifiquen en esta investigación mediante, por ejemplo, un contraste RESET de Ramsey.

***IV.VI. Supuesto 6: Existe independencia entre los términos del error***

Este supuesto no es verificado por los autores en ninguno de los modelos elaborados y estos no proporcionan los datos utilizados para que se verifiquen en esta investigación ni proporcionan los datos del término de error obtenido en cada uno de los modelos para realizar por cuenta propia tal verificación.

#### ***IV.VII. Supuesto 7: La muestra es lo suficientemente grande***

Como se adelantó en (Nabi, Supuestos del Modelo Clásico de Regresión Lineal y de los Modelos Lineales Generalizados, 2021, pág. 25), la regla heurística para determinar ello es que no más del 20% de las unidades muestrales deben contener dentro de sí menos de 5 observaciones. Sin embargo, para verificar esto se requeriría conocer la lista de las unidades de muestreo (*i.e.*, el marco muestral). La información detallada del marco muestral implica un nivel de especificidad categóricamente mayor que el de las variables utilizadas para la regresión, puesto que, por ejemplo, se requeriría conocer con cierto nivel de detalle el diseño del muestreo de aguas que realizaron los investigadores. Por consiguiente, no fue posible verificar si la muestra era lo suficientemente grande mediante, una vez conociendo sus características concretas, determinar si es lo suficientemente representativa de la población.

## V. REFERENCIAS

- Abril Díaz, N., Bárcena Ruiz, A., Fernández Reyes, E., Galván Cejudo, A., Jorrín Novo, J., Peinado Peinado, J., . . . Túñez Fiñana, I. (6 de Julio de 2021). *Espectrofotetría: Espectros de absorción y cuantificación colorimétrica de biomoléculas*. Obtenido de Departamento de Bioquímica y Biología Molecular, Universidad de Córdoba:  
[https://www.uco.es/dptos/bioquimica-biol-mol/pdfs/08\\_ESPECTROFOTOMETRIA.pdf](https://www.uco.es/dptos/bioquimica-biol-mol/pdfs/08_ESPECTROFOTOMETRIA.pdf)
- Asociación Española de Fabricantes de Agronutrientes. (26 de Enero de 2021). *Clasificación del pH*. Obtenido de Glosario de términos útiles en Agronutrición: <https://aefa-agronutrientes.org/glosario-de-terminos-utiles-en-agronutricion/clasificacion-del-ph>
- Atil Husni, I., Nyoman Budiantara, I., & Zain, I. (2018). *Partial hypothesis testing of truncated spline model in nonparametric regression*. College Park, Maryland: American Institute of Physics Conference Proceedings. Obtenido de <https://aip.scitation.org/doi/pdf/10.1063/1.5062798>
- Bermúdez Cabrera, X., Fleites Ramírez, M., & Contreras Moya, A. M. (Septiembre-diciembre de 2009). ESTUDIO DEL PROCESO DE COAGULACIÓN-FLOCULACIÓN DE AGUAS RESIDUALES DE LA EMPRESA TEXTIL "DESEMBARCO DEL GRANMA" A ESCALA DE LABORATORIO. *Revista de Tecnología Química*, XXIX(3), 64-73. Obtenido de <https://www.redalyc.org/pdf/4455/445543760009.pdf>
- Cepeda, Z., & Cepeda C., E. (2005). Application of Generalized Linear Models to Data Analysis in Drinking Water Treatment. *Revista Colombiana de Estadística*, XXVIII(2), 233-242.
- Domènech, X., & Peral, J. (2006). *Química Ambiental de sistemas terrestres*. (S. REVERTÉ, Ed.) Barcelona.
- Li, M., Duan, N., Zhang, D., Li, C.-H., & Ming, Z. (2009). *Collaborative Decoding: Partial Hypothesis Re-ranking Using Translation Consensus between Decoders*. Suntec, Singapore: Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP. Obtenido de <https://aclanthology.org/P09-1066.pdf>
- Nabi, I. (27 de Agosto de 2021). *Modelos Lineales Generalizados*. Obtenido de El Blog de Isadore Nabi:



<https://marxianstatistics.files.wordpress.com/2021/08/modelos-lineales-generalizados-isadore-nabi.pdf>

Nabi, I. (21 de Septiembre de 2021). *Supuestos del Modelo Clásico de Regresión Lineal y de los Modelos Lineales Generalizados*. Obtenido de El Blog de Isadore Nabi: <https://marxianstatistics.com/2021/09/24/supuestos-del-modelo-clasico-de-regresion-lineal-y-de-los-modelos-lineales-generalizados/>

Organización Mundial de la Salud. (Mayo de 2009). *Tratamiento de emergencia del agua potable en el lugar de consumo*. Obtenido de <http://bvspcr.paho.org/share/ETRAS/AyS/texcom/desastres/opsguia5.pdf>

Pérez de la Cruz, F. J., & Urrea Mallebrera, M. A. (21 de Enero de 2021). *ABASTECIMIENTO DE AGUAS. Coagulación y floculación*. Obtenido de Universidad Politécnica de Cartagena: [https://ocw.bib.upct.es/pluginfile.php/6019/mod\\_resource/content/1/Tema\\_06\\_COAGULACION\\_Y\\_FLOCULACION.pdf](https://ocw.bib.upct.es/pluginfile.php/6019/mod_resource/content/1/Tema_06_COAGULACION_Y_FLOCULACION.pdf)

Rahim, F., Budiantara, N., & Permatasari, E. O. (Marzo de 2019). Spline Truncated Nonparametric Regression Modeling for Maternal Mortality Rate in East Java. *Jurnal Penelitian Sosial Keagamaan*, II(1), 39-44. Obtenido de <https://media.neliti.com/media/publications/323488-spline-truncated-nonparametric-regressio-fae11742.pdf>