



ESTIMACIÓN DE LA PREVALENCIA DE CARIES INFANTIL EN CALI, USANDO MÉTODOS BAYESIANOS

Programa Académico de Estadística

Juan David Muñoz Espinosa

Universidad del Valle
Facultad de Ingeniería, Escuela de Estadística
Santiago de Cali, Colombia
2017

ESTIMACIÓN DE LA PREVALENCIA DE CARIES INFANTIL EN CALI, USANDO MÉTODOS BAYESIANOS

Programa Académico de Estadística

Juan David Muñoz Espinosa

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

Director:
José Rafael Tovar Cuevas

Universidad del Valle
Facultad de Ingeniería, Escuela de Estadística
Santiago de Cali, Colombia
2017

Agradecimientos

Quiero agradecer a Dios por haberme permitido culminar con este trabajo de grado. A mi mamá, mis hermanos y mi novia por brindarme siempre un apoyo moral y ánimo que me permitieron continuar siempre hacia adelante. Agradezco a la Dra. Judy Villavicencio por proporcionarme los datos que se utilizaron en esta investigación. Agradezco enormemente a mi director Jose Rafael Tovar Cuevas, ya que fue la persona que me orientó en sus asesorías con dedicación y paciencia.

Resumen

El presente trabajo de grado tiene como objetivo modelar la prevalencia de caries en niños con edades de 3 a 4 años residentes en barrios de escasos recursos de la ciudad de Cali, Colombia. Los datos corresponden a un grupo de niños vinculados en jornada completa a un centro de cuidado infantil bajo la supervisión del Instituto Colombiano de Bienestar Familiar (ICBF). La propuesta metodológica, considera la estimación de la prevalencia de caries tomando como variable respuesta la presencia-ausencia de la característica y la estimación de la prevalencia de caries cuando se tienen niveles de presencia. Para obtener los valores de los hiperparámetros asociados a las distribuciones a priori, se emplearon métodos empíricos de Bayes. De acuerdo con los resultados obtenidos el grado de pH, Unidades Formadoras de Colonias de Streptococcus Mutans y Unidades Formadoras de Colonias de Lactobacillus fueron factores de riesgo asociados a la prevalencia de caries.

Palabras claves: Distribución Binomial, Distribución Dirichlet, Distribución Multinomial, Distribución a Priori, Estadística Bayesiana, Prevalencia, Caries

Abstract

The objective of this thesis is to model the prevalence of caries in children aged 3-4 years living in low-income neighborhoods of the city of Cali, Colombia. The data correspond to a group of children linked full-time to a child care center under the supervision of the Colombian Family Welfare Institute (ICBF). Given that the data sample is not random, it was decided to use methods specific to Bayesian statistical inference, which are a very useful and reliable alternative in this type of situation, since, contrary to the classical paradigm, it is assumed that prevalence is a random variable and the sample data are observations already taken. The methodological proposal considers the estimation of the caries prevalence taking as a response variable the presence-absence of the characteristic and the estimation of the caries prevalence when there are levels of presence. To obtain the values of the hyperparameters associated with the a priori distributions, empirical Bayes methods were used. According to the results obtained, the degree of pH, Streptococcus Mutans Colonization Units, and Lactobacillus Colonies Formation Units were risk factors associated with the prevalence of caries.

Keywords: Binomial Distribution, Dirichlet Distribution, Multinomial Distribution, Prior Distribution, Bayesian Statistics, Prevalence, Caries

Contenido

Resumen	IV
Lista de Tablas	VII
1 Introducción	1
2 Planteamiento del Problema	2
2.1 Planteamiento del Problema Clínico	2
2.2 Planteamiento del Problema Estadístico	3
3 Justificación del Proyecto	4
3.1 Justificación Clínica	4
3.2 Justificación Estadística	4
4 Objetivos	5
4.1 Objetivo General	5
4.2 Objetivos Específicos	5
5 Antecedentes	6
5.1 Antecedentes Clínicos	6
5.2 Antecedentes Estadísticos	7
6 Marco Teórico	9
6.1 Marco Teórico Clínico	9
6.1.1 Caries Dental	9
6.1.2 Streptococcus Mutans y Lactobacillus	9
6.1.3 Capacidad Buffer y PH	10
6.2 Marco Teórico Estadístico	10
6.2.1 Distribución Binomial	10
6.2.2 Distribución Multinomial	10
6.2.3 Distribución Dirichlet	11
6.2.4 Inferencia Bayesiana	12
6.2.5 Regla de Bayes	13
6.2.6 Métodos Empíricos de Bayes	14
6.2.7 Distribución a Priori	14

6.2.8	Distribución a Priori no Informativa	15
6.2.9	Principio de Invariancia Jeffreys	15
6.2.10	Distribución a Priori Informativa	16
7	Metodología	17
7.1	Archivo de Datos	17
7.2	Variable de Respuesta	17
7.3	Variables Explicativas	18
7.4	Análisis Descriptivo de los Datos	18
7.5	Estimación de la Prevalencia de Caries	19
7.5.1	Experimento Estadístico	19
7.5.2	Modelo Estadístico	19
7.5.3	Grupos Binomiales	20
7.5.4	Grupos Binomiales Con Información Disponible	21
7.5.5	Distribución a Priori no Informativa	22
7.5.6	Elicitación de la Distribución a Priori Informativa	22
7.5.7	Distribución a Posteriori	23
7.5.8	Probabilidades Predictivas	24
7.6	Estimación de la Prevalencia de Caries Considerando el Grado de Severidad	24
7.6.1	Experimento Estadístico	24
7.6.2	Modelo Estadístico	24
7.6.3	Grupos Multinomiales Con Información Disponible	25
7.6.4	Distribución a Priori no Informativa	26
7.6.5	Elicitación de la Distribución a Priori Informativa	26
7.6.6	Distribución a Posteriori	27
8	Resultados	29
8.1	Análisis Descriptivo de los Datos	29
8.2	Proporción de Niños con Caries	30
8.2.1	Probabilidades Predictivas	33
8.3	Proporción de Niños Con Caries Considerando el Grado de Severidad	34
8.3.1	Distribución a Priori Informativa	35
9	Conclusiones y Recomendaciones	37
	Bibliografía	39

Lista de Tablas

7-1. Descripción de variables	18
7-2. Categorización de Variables	19
7-3. Grupos Binomiales Formados a Partir de las Categorías de las Covariables .	20
7-4. Clasificación de Niños	21
7-5. Grupos de Covariables	21
7-6. Grupos Finales de Estudio	22
7-7. Nueva Clasificación de Niños por Nivel de Caries	25
7-8. Tamaño Muestral Equivalente con un Nivel de Significancia α y un Porcentaje de Error d para Probabilidades Multinomiales (Rivera and Morales, 2015) .	27
8-1. Análisis Exploratorio	29
8-2. Distribución Conjunta de las Covariables y Variable de Respuesta	30
8-3. Proporción de Niños con Caries (A Priori No Informativas)	32
8-4. Proporción de Niños Con Caries (A Priori Informativa)	32
8-5. Probabilidades Predictivas Mediante la a Posteriori Informativa	33
8-6. Proporción de Niños con Nivel de Caries Asociado (Laplace)	35
8-7. Vector de Parámetros α con su Respectiva Media y Varianza para los Valores Elicitados.	36
8-8. Proporción de Niños con Nivel de Caries Asociado (A Priori Informativa) . .	36

1 Introducción

En el presente trabajo de grado, se desarrolló una propuesta metodológica para modelar la probabilidad de observar niños con presencia de caries. Los datos corresponden a un grupo de niños con edades entre 3 y 4 años vinculados en jornada completa a un centro de cuidado infantil bajo la supervisión del Instituto Colombiano de Bienestar Familiar (ICBF), en la ciudad de Cali. Se modeló la presencia de caries en al menos un diente como variable de respuesta binaria teniendo en cuenta la información contenida en tres covariables: Unidades Formadoras de Colonias Streptococcus Mutans, Unidades Formadoras de Colonias Lactobacillus y pH. También se modeló la probabilidad de niños con presencia de caries y la prevalencia del nivel de caries asociado. Dado que el tamaño de la muestra es reducido ($n=125$ niños) y no es aleatoria, se decidió utilizar métodos propios de la inferencia estadística bayesiana, los cuales son una alternativa bastante útil y confiable en este tipo de situaciones, pues, contrario al paradigma clásico, se asume que la prevalencia es una variable aleatoria y los datos de la muestra son observaciones ya tomadas. Se utilizaron distribuciones a priori no informativas o también llamadas de referencia. Las probabilidades fueron estimadas teniendo en cuenta la información adicional contenida en las tres covariables. De acuerdo a la estructura del documento se presentan diferentes capítulos donde se aborda el planteamiento del problema desde el punto de vista clínico y estadístico. La justificación argumentando la importancia de realizar el trabajo y los objetivos. Es importante mostrar estudios recientes que hayan abarcado el problema de la caries dental desde el punto de vista clínico y estadístico. Se presenta el marco teórico con el fin de proporcionar los conceptos mas relevantes que ayuden a la comprensión del trabajo de grado. La metodología desarrollada en este trabajo de grado permite tener claridad del paso a paso para obtener los resultados presentados para así poder concluir y hacer recomendaciones a futuros trabajos que se puedan realizar a partir de este.

2 Planteamiento del Problema

2.1. Planteamiento del Problema Clínico

La caries dental es la enfermedad con mayor prevalencia en el mundo, con pérdidas importantes por incapacidad, ausentismo escolar y dolor. Igualmente, su tratamiento es costoso para los servicios de salud y para quienes la padecen, tanto en países en desarrollo como industrializados (Sheiham, 2005; Organization et al., 2008; Schwendicke et al., 2015).

La caries dental puede aparecer durante cualquier etapa del ciclo vital, como consecuencia del desequilibrio en el intercambio mineral entre la superficie dental y la biopelícula, pues al consumir una dieta rica en azúcares y carbohidratos complejos, el balance ácido de la boca se pierde, producto del metabolismo de los microorganismos, acidificándose más el medio, lo que contribuye a la degradación dinámica del esmalte (Cameron et al., 1998).

Factores de riesgo medioambientales tales como la cantidad de la placa dental, la frecuencia de cepillado, la ingesta entre comidas de la dieta, la tasa de secreción salival, el pH, la capacidad buffer de la saliva, las bacterias y en especial los conteos de streptococcus mutans y los lactobacillus pueden influir en el desarrollo de la caries dental (García and Hicks, 2008). Sin embargo cuando se da una exposición del mismo nivel de factores de riesgo y del medio ambiente, algunos pacientes son más susceptibles o resistentes a la caries que otros. Estas diferencias pueden ser debido a factores genéticos en la etiopatogénesis de la caries dental (Selwitz et al., 2007).

De acuerdo con documentos del Ministerio de Salud y Protección Social, revela que en varios estudios han determinado una prevalencia de caries dental superior al 60 % en niños menores de seis años (Cadavid et al., 2010). El desarrollo de estas lesiones clínicamente visibles puede controlarse mediante la aplicación de algunas estrategias entre las que se encuentran la detección temprana de los factores de riesgo, la educación en higiene oral, el control de la dieta, la aplicación de fluoruros y sellantes, y el control biológico (González et al., 2009).

De acuerdo a lo anterior se presenta la necesidad de estimar la probabilidad de niños con dientes cariados, teniendo en cuenta diferentes factores clínicos presentes en los menores.

2.2. Planteamiento del Problema Estadístico

Un grupo de 125 niños con edades entre 3 y 4 años vinculados a un centro de protección supervisado por el Instituto Colombiano de Bienestar Familiar (ICBF) fueron evaluados con el propósito de estimar la probabilidad de los diferentes grados de severidad de caries. Se considera importante establecer una metodología estadística que permita identificar una variable de respuesta que se caracterice dentro del problema clínico y establecer asociaciones con variables explicativas, las cuales incidan en el riesgo del grado de severidad de la caries y a su vez sean utilizadas para fines predictivos en esta investigación.

Para obtener la variable de respuesta se realizó el examen clínico para el registro de caries con base en los criterios del Sistema Internacional de Detección y Evaluación de Caries (ICDAS). Este sistema de diagnóstico diferencia las lesiones tempranas no cavitadas de las cavitadas o dentinales, y permite establecer los diferentes niveles de avance de las lesiones de caries dental (Banting et al., 2005). Esta separación, útil en el ámbito clínico, conlleva a un problema estadístico puesto que es necesario establecer un modelo que permita estudiar la relación entre la severidad de la caries dental considerando una estructura ordinal.

Para obtener las diferentes variables explicativas en este estudio, se tomó como referencia los trabajos presentados por Selwitz et al. (2007); Peres et al. (2016); García and Hicks (2008), quienes identificaron algunos aspectos clínicos y otros asociados directamente con los individuos que pueden estar relacionados con la presencia de caries dental.

El archivo de datos empleado quedó conformado por 125 registros de niños, en los que se contaba con información de las características físicas, clínicas y mediciones del perfil microbiano salival. Para desarrollar la propuesta metodológica estadística, se consideraron algunos aspectos como que la variable de respuesta es de tipo ordinal y tener un número reducido de observaciones, lo cual implica limitaciones con el proceso de estimación de los parámetros, y el riesgo de tener intervalos de confianza demasiado anchos y poco informativos.

En una primera aproximación, se modeló la presencia-ausencia de caries, operacionalizando la variable Número de dientes con caries como: Éxito cuando se presentaba al menos un diente con caries y fracaso en caso contrario. Posteriormente, se realizó una aproximación mas fina al asunto modelando el grado de severidad de la caries usando un modelo multinomial de probabilidades.

3 Justificación del Proyecto

3.1. Justificación Clínica

Actualmente, se hace necesario encontrar alternativas que hagan más efectivas las intervenciones diagnósticas y terapéuticas que ayuden en la toma de decisiones al médico tratante. Todo cuanto se realice a favor de caracterizar estas prácticas ayudaría a los niños con presencia de caries a un seguimiento oportuno y adecuado de una posible eventualidad (Morera, 2007).

Por lo anterior, esta investigación esta orientada a identificar factores asociados con la presencia de caries lo que conlleva a conocer las características propias de la población nacional y a promover herramientas útiles de predicción, contemplando alternativas de control del proceso de caries dental; estas alternativas serían fáciles de implementar y sus beneficios se reflejarían en una disminución de la prevalencia de caries dental (Min. and Protección, 2013).

Se espera que los resultados de este estudio sirvan como insumo para las personas encargadas de tomar decisiones en áreas como la salud pública, pues contar con información precisa sobre la prevalencia de la enfermedad (Caries Dental) en los niños permite diseñar planes de manejo, control y prevención.

3.2. Justificación Estadística

Actualmente, se requiere alternativas estadísticas mediante la modelación de tipo predictivo, lo cual es una herramienta metodológica clínica (Aycaguer, 1994), para la toma de decisiones, criterios y estrategias a emplear para el seguimiento, la intervención y la confirmación de los niños con presencia de caries.

Por otra parte, en la mayoría de los estudios clínicos, epidemiológicos y biometricos el número reducido de observaciones es muy común, debido a las dificultades que se presentan al momento de la recolección de datos, por lo tanto la información externa brindada puede llegar a ser muy útil desde el ámbito clínico y estadístico. En estos casos, la inferencia bayesiana es una buena alternativa, ya que permite incluir información externa a través de la distribución a priori, es flexible a muestras pequeñas y no se corre el riesgo de tener problemas en el proceso de estimación.

4 Objetivos

4.1. Objetivo General

Modelar la prevalencia de caries en niños con edades de 3 a 4 años residentes en barrios de escasos recursos de la ciudad de Cali, Colombia.

4.2. Objetivos Específicos

- Estimar la prevalencia de caries en niños con edades de 3 a 4 años en Cali, Colombia.
- Estimar la prevalencia de los diferentes grados de severidad de caries en niños.
- Realizar las estimaciones de las prevalencias utilizando métodos propios de inferencia bayesiana.

5 Antecedentes

5.1. Antecedentes Clínicos

Se tienen los siguientes antecedentes sobre estudios clínicos donde se estableció el efecto del consumo de productos con probióticos, para determinar si existen mejorías en la presencia de caries en los dientes de los pacientes.

Näse et al. (2001) realizó un estudio donde *Lactobacillus rhamnosus* GG, ATCC (LGG), demostró antagonismo con muchas bacterias, incluyendo los estreptococcus mutans. Este estudio de intervención aleatorizado, doble ciego, controlado con placebo, fue diseñado para examinar si la leche que contiene LGG tiene un efecto sobre la caries y el riesgo de caries en los niños en comparación con la leche normal. Se incluyeron 594 niños de 1 a 6 años de 18 guarderías municipales. Los niños recibieron la leche con comidas de envases codificados 5 días a la semana en las guarderías durante 7 meses. La salud bucal de los niños se registró al inicio y al final, utilizando los criterios de la OMS. El riesgo de caries se calculó sobre la base de datos clínicos y microbiológicos, comprendiendo los niveles de estreptococos mutans de la placa dental y la saliva. El riesgo se clasificó como alto si el niño tenía un dmft / DMFT o una puntuación inicial de caries mayor que 0, y un recuento de estreptococos mutans mayor o igual a 10⁵ UFC/ml. Los resultados mostraron menos caries dentales en el grupo LGG y menores recuentos de estreptococos mutans al final del estudio. LGG se encontró para reducir el riesgo de caries significativamente (OR = 0,56, p = 0,01, controlado por edad y sexo, OR = 0,51, p = 0,004). El efecto fue particularmente claro en los niños de 3 a 4 años. Por lo tanto, la leche que contiene las bacterias LGG probióticas puede tener efectos beneficiosos sobre la salud dental de los niños.

Glavina et al. (2012) realizó un estudio para establecer el efecto del consumo de yogur que (disponible comercialmente durante 14 días) contenía *Lactobacillus rhamnosus* ATCC53103 - LGG (Bioaktiv LGG, Dukat, Croacia) en *Streptococcus mutans* y *Lactobacillus* spp salivales en niños. Veinticinco pacientes de 6 a 10 años de edad participaron en el estudio. En la inclusión del estudio se evaluó el riesgo de caries para cada paciente. Las muestras de saliva se ensayaron con kits de lado de silla para la capacidad buffer de la saliva (CRT buffer, Vivadent, Schaan, Liechtenstein), *Streptococcus Mutans* y *Lactobacillus* (prueba de bacterias CRT, Vivadent, Schaan, Liechtenstein). Siete, catorce y treinta días después de las muestras de saliva de consumo de yogur se probaron nuevamente con pruebas de CRT buffer

y pruebas CRT de bacterias. Los datos obtenidos se analizaron mediante pruebas de ceo-d y Kruskal-Wallis. Los resultados mostraron un aumento significativo en la capacidad buffer de la saliva treinta días después del consumo de yogur. Los recuentos salivales de *Streptococcus Mutans* disminuyeron significativamente después de treinta días. No se observaron diferencias significativas en los recuentos de *Lactobacillus*. Se podría concluir que el consumo diario de yogur que contiene LGG tiene un efecto inhibitor sobre las bacterias patógenas orales y puede ser beneficioso en la prevención de caries.

5.2. Antecedentes Estadísticos

Casanova et al. (2005) determinaron la prevalencia de pérdida del primer molar permanente (PMP) en niños escolares de 6 a 13 años de edad, e investigaron la relación entre la pérdida del PMP y factores sociodemográficos, socioeconómicos, conductuales y clínicos. Los autores realizaron un estudio transversal en 1517 escolares de la ciudad de Campeche, México. Un cuestionario estructurado dirigido a las madres fue enviado a estas para explorar las variables sociodemográficas, socioeconómicas y conductuales. En los niños se efectuó un examen bucal clínico. En el análisis se realizaron estadísticas descriptivas y un modelo multivariado de regresión logística binaria. Los resultados arrojaron que la prevalencia de sujetos con al menos un PMP perdido fue de 7,5 % (n=114). De los 6,068 PMP examinados, 2.1 % (n=130) se clasificaron como perdidos. Los PMP de la arcada inferior se perdieron con mayor frecuencia (70 %, n=91). Las variables asociadas a la pérdida del PMP que permanecieron en el modelo final fueron: mayor edad y la higiene bucal regular e inadecuada. Así como una interacción entre los defectos estructurales del esmalte y el sexo, considerando 1) el efecto del defecto estructural del esmalte en los niños, y 2) el efecto del defecto estructural del esmalte en las niñas. En el estudio concluyeron que la prevalencia de pérdida del PMP fue relativamente alta considerando el grupo de edad bajo investigación (6-13 años de edad). Se encontraron variables de diversos tipos asociadas a la pérdida del PMP. Los datos presentados sirven como un indicador del estado de salud bucal y podrían ser empleados como línea basal para la evaluación de programas de salud dental preventivos.

García et al. (2011) se plantearon como objetivo identificar y cuantificar la relación empírica entre las características socioeconómicas y demográficas de las familias y el consumo de alimentos fuera del hogar (AFH), con base a la información provista por la Encuesta Nacional de Gasto de Hogares (ENGH). Para ello especificaron un modelo multinomial de utilidad aleatoria sobre la base de respuestas politómicas en función de los motivos o de las circunstancias prevalecientes al momento en que las familias deciden consumir AFH, sea en horas de esparcimiento, por motivos de trabajo o estudio, o por ambas razones.

Rivera and Morales (2015) desarrollaron un método de elicitación que permite estimar la distribución del vector de parámetros π de la distribución Multinomial bajo un método indirecto de elicitación y, adicionalmente implementarlo en el software R, de manera

que el analista pueda llevar registro de todas estimaciones hechas por el experto y brindarle retroalimentación continua durante todo el proceso. Inicialmente revisaron los enfoques y desarrollos actuales para la cuantificación de la incertidumbre o creencias de un “experto” sobre un parámetro Binomial y sobre el vector de parámetros Multinomial por medio de un proceso de elicitación. Posteriormente revisaron los software más comunes que se han propuesto hasta ahora para la elicitación de juicios de expertos; una vez revisados, identificaron las ventajas, desventajas y posibles limitaciones que tanto los métodos de elicitación como los software podían tener, de tal forma que sirvan como base teórica para el método propuesto y la implementación en R. Finalmente el método y la implementación en R fueron usados en una aplicación real en el campo de la urología.

6 Marco Teórico

En esta sección, se presentan los conceptos teóricos que se utilizaran en la realización del presente trabajo. Una Primera parte donde se presenta un marco conceptual en el cual se definen algunos conceptos sobre el perfil microbiano salival de la caries presente en los niños. En la segunda parte se presentará el sustento teórico de las técnicas estadísticas para dar solución al problema en cuestión.

6.1. Marco Teórico Clínico

6.1.1. Caries Dental

Una de las enfermedades de mayor prevalencia e incidencia en los niños preescolares y escolares es la caries dental. Esta enfermedad infecciosa tiene el potencial de producir cavitación del esmalte dental y posteriormente daño a la dentina y a la pulpa, culminando con la destrucción localizada de los tejidos duros del diente si la desmineralización es desenfrenada. Además, como consecuencia de ese deterioro, el proceso puede causar dolor, pérdida temprana de piezas dentales, ausentismo escolar, y requerir tratamientos costosos. Estas características hacen que se le considere un problema significativo de salud pública. Se concentra en poblaciones pobres tanto en países en vías de desarrollo como en los industrializados (Medina et al., 2006).

6.1.2. Streptococcus Mutans y Lactobacillus

Se ha reconocido que los lactobacillus tienen una baja afinidad por los dientes. Ellos requieren retención mecánica para poder jugar un rol en la colonización de las superficies dentales. Por lo tanto son los streptococcus mutans y otras bacterias orales acidogénicas las que crean inicialmente unos nichos precarios o lesiones iniciales en el esmalte de los dientes, que le da la posibilidad al lactobacillus de tener un nicho retentivo, el cual es un prerequisite para su colonización. Este nicho estancado provee un contenedor físico que soporta un pH bajo y un medio ambiente anaerobio. Adicionalmente, se da en ellos el atrapamiento mecánico de comida, con una fuente de carbohidratos. La evidencia indica que la presencia de lactobacillus en la cavidad oral desde la niñez hasta la edad adulta es un indicador de que un individuo desarrolle caries o no (Badet and Thebaud, 2008; Parisotto et al., 2010).

6.1.3. Capacidad Buffer y PH

La desmineralización del esmalte está asociada a una disminución en el pH salival que se da por la fermentación del azúcar por medio de las bacterias presentes en boca. Los cambios en el patrón del pH salival en pacientes que tienen alto riesgo para la caries descienden en comparación con aquellos pacientes que tienen un bajo riesgo. Hay una relación entre el pH, una baja capacidad buffer de la saliva y la experiencia de caries dental. El pH alto y una capacidad buffer salival alta se asocia a niños con caries disminuida (Witjaksono, 2013).

Según Díaz (2002) la capacidad buffer es la capacidad que debe tener una solución amortiguadora de pH y en términos de número de moles (o equivalentes) de ácido fuerte o de base fuerte necesarios para disminuir o aumentar una unidad de pH respectivamente, por litro de solución buffer.

6.2. Marco Teórico Estadístico

6.2.1. Distribución Binomial

En el modelo Binomial simple, el objetivo es estimar la proporción de población desconocida a partir de los resultados de una secuencia de *ensayos de Bernoulli*, es decir, de datos $y_1, \dots, y_n$, cada uno de los cuales es 0 o 1. Este problema proporciona un punto de partida relativamente simple, pero importante para la discusión de la inferencia bayesiana.

La distribución Binomial ofrece un modelo natural para los datos que surgen de una secuencia de ensayos intercambiables de una gran población, donde cada ensayo da lugar a uno de dos resultados posibles, convencionalmente etiquetados *éxito* y *fracaso*. Debido a la intercambiabilidad, los datos se puede resumir por el número total de éxitos en n ensayos, que denotamos aquí por y . La conversión de una formulación en términos canjeables se logra con ensayos uno a uno usando variables aleatorias independientes e idénticamente distribuidas (*v.a.i.i.d*) naturalmente, al permitir que el parámetro θ represente la proporción de éxitos en la población o, equivalentemente, la probabilidad de éxito en cada ensayo. De acuerdo con Gelman et al. (2014) el modelo de muestreo binomial es:

$$p(\theta | y) = \text{Bin}(y | n, \theta) = \binom{n}{y} \theta^y (1 - \theta)^{n-y} \quad (6-1)$$

Donde en el lado izquierdo suprimimos la dependencia de n , ya que es considerada como parte del diseño experimental que se considera fija, todas las probabilidades discutidas para este problema se supone que están condicionadas a n (Gelman et al., 2014).

6.2.2. Distribución Multinomial

La distribución Multinomial atrae considerablemente la atención de numerosos investigadores teóricos en el área de las distribuciones multivariantes discretas. Generalmente se usa en las

mismas situaciones en las que se podría utilizar una distribución Binomial, cuando hay múltiples categorías de eventos en lugar de una simple dicotomía (Johnson et al., 2005; Elfadaly and Garthwaite, 2012).

Considere una serie de n ensayos independientes, en donde tan solo uno de los k eventos mutuamente excluyentes de $E_1, E_2, \dots, E_k$ debe ser observado, y en el cual la probabilidad de ocurrencia del evento k , en cualquier ensayo es igual a π (con $\pi_1 + \pi_2 + \dots + \pi_k = 1$). Dejemos que $n_1, n_2, \dots, n_k$ denoten una variable aleatoria del número de ocurrencias de los eventos $E_1, E_2, \dots, E_k$ respectivamente en esos n ensayos, con $\sum_{i=1}^k n_i = n$. Entonces la función de masa de probabilidad de $n_1, n_2, \dots, n_k$ está dada por:

$$P(n_1, n_2, \dots, n_k | \pi_1, \pi_2, \dots, \pi_k) = \begin{cases} \frac{n!}{n_1! n_2! \dots n_k!} \pi_1^{n_1} \pi_2^{n_2} \dots \pi_k^{n_k}, & \text{Cuando } \sum_{i=1}^k n_i = n \\ 0, & \text{Otro Caso} \end{cases} \quad (6-2)$$

6.2.3. Distribución Dirichlet

Un ejemplo de una tarea de elicitación multivariada donde se correlacionan necesariamente las cantidades, es elicitación del conocimiento del experto sobre un conjunto de proporciones que deben sumar 1. Un enfoque habitual es asumir que el conocimiento del experto puede ser representado adecuadamente por una distribución Dirichlet, ya que esta es una distribución multivariable muy simple que es apropiada para dicho conjunto de proporciones. La elección de una a priori Dirichlet tiene algunas ventajas con respecto a la tratabilidad matemática y a la representación de una rica clase de creencias. La a priori Dirichlet también permite evaluaciones analíticas de la distribución predictiva y de las distribuciones de características numéricas estadísticamente relevantes de una medida de probabilidad aleatoria (Regazzini and Sazonov, 1999). Adicional a lo anterior, esta distribución es la más conveniente cuando el conocimiento a priori del experto es combinado con una muestra Multinomial, ya que la Dirichlet es la distribución conjugada (DeGroot, 2004; Zapata et al., 2014).

Considere el caso de elicitación de las creencias del experto sobre un conjunto de cantidades inciertas $\underline{\pi} = \pi_1 + \pi_2 + \dots + \pi_k = 1$ las cuales están restringidas a caer en la categoría $k-1$ (ya que la categoría k queda determinada por defecto), es decir, $\pi_i \geq 0$ para $i = 1, \dots, k$ y $\sum_{i=1}^k \pi_i = 1$. Entonces podemos decir que π tiene una distribución Dirichlet con parámetros $\alpha = (\alpha_1, \dots, \alpha_k)$, denotada por $\pi \sim Di(\alpha)$ si su función de densidad está dada por:

$$f(\pi | \alpha) = c(\alpha) \prod_{i=1}^k \pi_i^{\alpha_i - 1} \quad (6-3)$$

donde $c(\alpha) = \Gamma(\sum_{i=1}^k \alpha_i) / \prod_{i=1}^k \Gamma(\alpha_i)$. La restricción de que la $\sum_{i=1}^k \pi_i = 1$, hace que los π_i 's sean inevitablemente correlacionados.

6.2.4. Inferencia Bayesiana

La inferencia bayesiana está basada en el análisis de la distribución a posteriori porque ésta contiene toda la información sobre el parámetro a ser estimado θ condicional a los datos observados y . En general en inferencia bayesiana es útil resumir la información que está contenida en la distribución a posteriori, y estos resúmenes típicamente toman la forma de esperanzas de funciones particulares de los parámetros. El problema de calcular este tipo de esperanzas es que es muy complicado o imposible evaluarlas de forma analítica. Incluso técnicas numéricas de cuadratura u otras para aproximar podrían presentar problemas, mas aun si el parámetro θ es multidimensional. Por ese motivo se hace necesario utilizar métodos aproximados para obtener las integrales de la expresión (6-11). En los últimos años la inferencia bayesiana ha experimentado un gran avance debido a la introducción de técnicas de simulación que permiten en forma relativamente simple obtener muestra de una distribución objetivo (Bazán and Bayes, 2010). En particular, los métodos conocidos como Cadenas de Markov de Monte Carlo (MCMC) son actualmente muy utilizados. La racionalidad de estos métodos subyace en diseñar iterativamente una cadena de Markov para θ de tal manera que $p(y | \theta)$ sea su distribución ergódica estacionaria. Empezando en algún estado inicial θ_0 la idea es simular un numero suficientemente grande M de transiciones bajo la cadena de Markov y registrar los correspondientes estados simulados θ_j . El reto de los métodos MCMC consiste entonces en precisar una cadena de Markov adecuada con la distribución a posteriori $p(y | \theta)$ como su distribución estacionaria y decidir cuando detener la simulación (Bazán and Bayes, 2010).

Teorema de Bayes

Según la definición de probabilidad condicional dicha por Bayes (1763), se tiene que:

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \quad (6-4)$$

Se sabe que la probabilidad marginal de evento A se encuentra sumando las probabilidades de sus partes disyuntas. Desde entonces $A = (A \cap B) \cup (A \cap B')$ y con claridad $(A \cap B)$ y $(A \cap B')$ son disyuntos (Bayes, 1763).

$$P(A) = P(A \cap B) + P(A \cap B') \quad (6-5)$$

Se sustituye la ecuación 6-5 en la definición de la probabilidad condicional para conseguir:

$$P(B|A) = \frac{P(A \cap B)}{P(A \cap B) + P(A \cap B')} \quad (6-6)$$

Ahora, se usa la regla de multiplicación para encontrar cada una de estas probabilidades conjuntas dadas para un solo evento.

$$P(B|A) = \frac{P(A|B) * P(B)}{P(A|B) * P(B) + P(A|B') * P(B')} \quad (6-7)$$

6.2.5. Regla de Bayes

Con el fin de hacer declaraciones acerca de $P(\theta|y)$, se empieza con proporcionar una distribución de probabilidad conjunta para θ y y . La función de masa o densidad conjunta se puede escribir como un producto de dos densidades, que se refiere a menudo como la distribución a priori $p(\theta)$ y la distribución de los datos (verosimilitud) $p(y | \theta)$, respectivamente (Gelman et al., 2014):

$$p(\theta, y) = p(\theta)p(y|\theta) \quad (6-8)$$

Utilizando la propiedad básica de probabilidad condicional conocida como la regla de Bayes (1763), se obtiene la densidad posterior:

$$p(\theta|y) = \frac{p(\theta, y)}{p(y)} = \frac{p(\theta)p(y|\theta)}{p(y)} \quad (6-9)$$

Donde $p(y) = \sum_{\theta} p(\theta)p(y|\theta)$, en caso de θ discreta y $p(y) = \int p(\theta)p(y|\theta)d\theta$ en el caso de θ continua. Por lo tanto, es una forma equivalente de la ecuación (6-9) que omite el factor $p(y)$, que no depende de θ , de este modo puede considerarse una constante, dando la densidad posterior, como (Gelman et al., 2014):

$$p(\theta|y) \propto p(\theta)p(y|\theta) \quad (6-10)$$

El segundo término en esta expresión, $p(y | \theta)$, se toma como una función de θ . Estas fórmulas simples encapsulan el núcleo técnico de la inferencia bayesiana. La tarea primordial de cualquier aplicación específica es desarrollar el modelo de $p(\theta, y)$ y realizar los cálculos para resumir $p(\theta, y)$ de manera apropiada. Para hacer inferencias acerca de un θ desconocido o a menudo llamadas inferencias observables predictivas. Es decir, la distribución de lo desconocido, pero observable y es:

$$p(y) = \int p(\theta, y)d\theta = \int p(\theta)p(y|\theta)d\theta \quad (6-11)$$

La ecuación (6-11) llamada distribución marginal de y , no está condicionada a una observación previa del proceso y predictivo, ya que es la distribución de una cantidad que es observable (Gelman et al., 2014).

Después de que se han observado los datos de y , se puede predecir un observable desconocido, y^* , desde el mismo proceso. Por ejemplo, $y = (y_1, \dots, y_n)$ puede ser el vector de pesos registrados de un objeto pesado n veces en una escala, $\theta \sim (\mu, \sigma^2)$ Puede ser el peso verdadero desconocido del objeto y la varianza de medición de la escala, y y^* puede ser un vector de valores que aún no se registra el peso. La distribución de y^* se llama la distribución posterior predictiva; posterior, ya que está condicionada a la observada y y predictiva porque es una predicción para un y^* observable (Gelman et al., 2014):

$$p(y^*|y) = \int p(y^*, \theta|y)d\theta = \int p(y^*|\theta)p(\theta|y)d\theta \quad (6-12)$$

6.2.6. Métodos Empíricos de Bayes

El método de Bayes se plantea como una metodología alterna al método clásico. Dicho método se fundamenta en el conocimiento de una distribución sobre el parámetro θ , $f(\theta)$, denominada distribución a priori. Tal distribución se supone completamente conocida, no tiene parámetros desconocidos e involucra probabilidades subjetivas. El método de Bayes Empírico depende también de la existencia de una distribución a priori $f(\theta)$ a la cual se le puede dar una interpretación frecuencial y se puede estimar usando observaciones apropiadas. Así el método de Bayes Empírico puede ser esencialmente no bayesiano, en el sentido de no involucrar probabilidades subjetivas. El Bayes Empírico concuerda con el método de Bayes, pero no especifica valores de los parámetros de la distribución a priori $f(\theta)$, estima tales parámetros a partir de datos auxiliares (pasados o presentes). Para concluir se puede decir, que la diferencia fundamental entre el método de Bayes y el método de Bayes Empírico, radica en que el método de Bayes supone el conocimiento total de la distribución a priori $f(\theta)$ y el método de Bayes Empírico estima la distribución a priori $f(\theta)$ (Prieto et al., 1995).

6.2.7. Distribución a Priori

Se consideran dos interpretaciones básicas que se pueden dar a las distribuciones previas (a priori). En la interpretación de la población, la distribución a priori representa una población de posibles parámetros o valores, del que se ha sacado el θ de interés actual. En el estado más subjetivo de la interpretación del conocimiento, el principio rector es que debemos expresar nuestro conocimiento (e incertidumbre) sobre θ como si su valor podría ser considerado como una realización aleatoria de la distribución a priori. Para muchos problemas, tales como la estimación de la probabilidad de fallo en un nuevo proceso industrial, no hay ninguna población perfectamente relevante de θ desde el cual el θ actual se ha dibujado, excepto en la contemplación hipotética. Por lo general, la distribución a priori debe incluir todos los valores plausibles de θ , pero la distribución no necesita ser realista concentrada en torno al valor verdadero, porque muchas veces la información sobre θ esta contenida en los datos que superan con creces cualquier especificación o probabilidad previa razonable (Gelman et al., 2014). En primer lugar, perseguimos el modelo binomial con más detalle en el uso de una familia paramétrica de distribuciones, que incluye la Uniforme como un caso especial. Por conveniencia matemática, se construyó una familia de densidades anteriores que conducen a densidades posteriores simples. Considerado como una función de θ , la probabilidad (6-1) es de la forma (Gelman et al., 2014).

$$p(y | \theta) \propto \theta^a (1 - \theta)^b \quad (6-13)$$

Por lo tanto, si la densidad anterior es de la misma forma, con sus propios valores a y b , entonces la densidad posterior también será de esta forma. Vamos a parametrizar una densidad tal como antes.

$$p(\theta) \propto \theta^{\alpha-1} (1 - \theta)^{\beta-1} \quad (6-14)$$

Que es una distribución Beta con parámetros α y β donde $\theta \sim \text{Beta}(\alpha, \beta)$. Comparando $p(\theta)$ y $p(y | \theta)$ sugiere que la densidad anterior es equivalente a $\alpha - 1$ éxitos anteriores y $\beta - 1$ fracasos anteriores. Los parámetros de la distribución anterior se refieren a menudo como *hiperparámetros*. La distribución previa Beta está indexada por dos hiperparámetros, que significa que podemos especificar una distribución anterior en particular mediante la fijación de dos características de la distribución, por ejemplo, su media y la varianza. La densidad posterior de θ es:

$$p(y | \theta) \propto \theta^y (1 - \theta)^{n-y} \theta^{\alpha-1} (1 - \theta)^{\beta-1} = \text{Beta}(\theta | \alpha + y, \beta + n - y) \quad (6-15)$$

La propiedad que hace referencia en cuanto a que la distribución posteriori sigue la misma forma paramétrica de la distribución a priori, se llama conjugación. La a priori de distribución Beta es una familia conjugada para la probabilidad Binomial. La media posterior de θ se puede interpretar como la probabilidad posterior de éxito para un futuro sorteo de la población. (Gelman et al., 2014).

$$E(\theta | y) = \frac{\alpha + y}{\beta + n - y} \quad V(\theta | y) = \frac{(\alpha + y)(\beta + n - y)}{(\beta + n + \alpha)^2 (\beta + n + \alpha + 1)} \quad (6-16)$$

6.2.8. Distribución a Priori no Informativa

Para realizar inferencia estadística desde el enfoque bayesiano, se requiere contar con información externa al proceso de muestreo que dio origen a los datos observados, la cual puede tener como fuente el conocimiento subjetivo de un especialista en el tema, datos históricos recolectados sobre el tema de interés, resultados publicados sobre el mismo asunto en diferentes fuentes, etc. Una apriori no informativa, o también llamada apriori de referencia, son distribuciones que representan la mínima cantidad posible de información a priori disponible, generalmente útiles cuando no se cuenta con un especialista en el tema o se tiene tan poca información publicada sobre el comportamiento natural de las cantidades aleatorias de interés. Generalmente, las a priori de referencia se utilizan para tener una línea de base con la cual comparar otras distribuciones con los resultados obtenidos usando métodos clásicos (Gelman et al., 2014).

6.2.9. Principio de Invariancia Jeffreys

El principio de Jeffreys conduce a la definición de la densidad a priori no informativa como $h(\theta) \propto [J(\theta)]^{1/2}$, donde $J(\theta)$ es la información de Fisher para θ (Gelman et al., 2014).

$$J(\theta) = -E \left(\frac{d^2 \ln[l(\theta|x)]}{d\theta^2} \right) \quad (6-17)$$

La distribución a priori de jeffreys sera $h(\theta) = \sqrt{J(\theta)}$

Por lo tanto:

$$\sqrt{J(\theta_j)} = \sqrt{\frac{n_j}{\theta_j(1 - \theta_j)}} = \sqrt{n_j} \theta_j^{-1/2} (1 - \theta_j)^{-1/2} \quad (6-18)$$

Dado lo anterior se concluye que la distribución a priori de Jeffrey es:

$$h(\theta_j) = \text{Beta}(1/2, 1/2) \quad (6-19)$$

6.2.10. Distribución a Priori Informativa

Se presenta cuando se tiene información externa del evento de interés, dicho conocimiento se puede dar por la experiencia del especialista o investigaciones previas. El hecho de expresarla en forma de una distribución de probabilidades, se ha denominado proceso de elicitación a priori el cual es un proceso complejo con el que se formalizan las creencias que se tiene acerca del comportamiento natural de un fenómeno (Tovar, 2015).

De acuerdo con Gelman et al. (2014), cuando no se cuenta con fuentes externas de información, es posible usar una parte de los datos o realizar procesos con los mismos que permitan tener una primera aproximación a lo que podría ser el comportamiento de los parámetros sin llegar a ser concluyentes en nuestras apreciaciones. Los autores denominaron a dicho proceder como métodos empíricos de Bayes o Bayes Empíricos (Gelman et al., 2014).

7 Metodología

7.1. Archivo de Datos

Los datos corresponden a un grupo de escolares con edades entre 3 y 4 años, provenientes de un Centro de Desarrollo Infantil de Cali, Valle del Cauca, Colombia. Como criterio de exclusión se consideró el uso de antimicrobianos sistémicos a la fecha de la inclusión y los niños con trastornos sistémicos que necesitan cuidados especiales o que tienen intolerancia a las bebidas lácteas o alergia a alguno de los componentes de la bebida experimental y placebo y los niños que están bajo tratamiento con antibióticos permanente. Los criterios de inclusión fueron: niños de 3 a 4 años escolarizados, que sus padres firmaron el consentimiento informado. Finalmente la muestra estuvo conformada por 125 niños, entre los cuales 70 niños presentaban caries y 55 no.

7.2. Variable de Respuesta

En cada visita del grupo de investigadores, los niños debían cepillarse los dientes antes de realizar la revisión de caries dental cumpliendo con las condiciones del criterio del Sistema Internacional de Detección y Evaluación de Caries (ICDAS) para evaluar la severidad de la caries en niños (Banting et al., 2005). Se tomó como componente cariado, todo tipo de lesión de caries desde 2 al 6, clasificándose como: Leve (lesión ICDAS 2) lesión en mancha blanca/café en superficie húmeda, moderada (ICDAS 3-4) discontinuidad de la superficie y sombra subyacente y severa (ICDAS 5-6) lesión cavitada de menos del 50 % del diente y más del 50 % de compromiso de la superficie dental. Se determinó excluir las lesiones 1 de ICDAS por su dificultad en el diagnóstico en estas etapas y también los dientes que no estuvieran completamente erupcionados, no se evaluaban. Por tanto, al realizar la clasificación se obtuvieron 3 subgrupos de los niños que presentaban caries, uno con 27 casos de niños con caries leve, otro con 7 casos de niños con caries moderada y otro con 36 casos de niños con caries severa. Hubo 55 casos de niños sin presencia de caries en sus dientes.

De acuerdo a la clasificación anterior y teniendo en cuenta el objetivo del proyecto de grado, la variable de respuesta presenta una estructura binaria para la estimación de la prevalencia de caries:

$$Y = \begin{cases} 0, & \text{Ausencia de Caries: 55 Niños} \\ 1, & \text{Presencia de Caries: 70 Niños} \end{cases} \quad (7-1)$$

Y una estructura ordinal con cuatro posibles estados para la estimación de la prevalencia de caries considerando el grado de severidad:

$$Y = \begin{cases} 1, & \text{Sin presencia de caries: 55 Niños} \\ 2, & \text{Caries Leve: 27 Niños} \\ 3, & \text{Caries Moderada: 7 Niños} \\ 4, & \text{Caries Severa: 36 Niños} \end{cases} \quad (7-2)$$

Finalmente, la muestra del estudio quedo conformado por 125 niños.

7.3. Variables Explicativas

Con el fin de proponer variables explicativas o covariables, se tomó como referencia los trabajos presentados por Selwitz et al. (2007); García and Hicks (2008); Peres et al. (2016). Al revisar los diferentes estudios se identificaron tres (3) variables (unidades formadoras de colonias lactobacillus (UFCLB), unidades formadoras de colonias streptococcus mutans (UFCSM) y pH. Se contaba con otras variables que tenían poca asociación clínica y estadística con la variable de respuesta y por ende se omitieron. En la tabla **7-1**, se describen las variables propuestas.

Tabla 7-1: Descripción de variables

Variable	Descripción	Típo	Rango
UFCSM	Unidades Formadoras de Colonias Streptococcus Mutans	Cuantitativa	0-2125000 UFC/ml
UFCLB	Unidades Formadoras de Colonias LactoBacillus	Cuantitativa	0-3900000 UFC/ml
PH	pH de la saliva	Cuantitativa	2,25-7,73

7.4. Análisis Descriptivo de los Datos

El análisis estadístico comenzó con la realización de un análisis exploratorio con las variables de interés propuestas en la Tabla **7-1**. Adicionalmente, en la Tabla **7-2** se categorizaron las variables explicativas UFCLB y UFCSM mediante los estadísticos de posición como los cuantiles (mediana, tercer cuantil) en caso de asimetría de los datos y celdas en cero, la variable pH se categorizó mediante el criterio clínico dado en el estudio de Liso et al. (2000).

Tabla 7-2: Categorización de Variables

X_1	X_2	X_3
pH	UFCLB	UFCSM
$1 = pH < 7$	$1 = UFCLB < 625000 \text{ UFC/ml}$	$1 = UFCSM < 20000 \text{ UFC/ml}$
$2 = pH = 7$	$2 = UFCLB = 625000 \text{ UFC/ml}$	$2 = UFCSM = 20000 \text{ UFC/ml}$
$3 = pH > 7$	$3 = UFCLB > 625000 \text{ UFC/ml}$	$3 = UFCSM > 20000 \text{ UFC/ml}$

7.5. Estimación de la Prevalencia de Caries

7.5.1. Experimento Estadístico

Se tomó un niño de un Centro de Desarrollo Infantil de la ciudad de Cali, con el fin de observar cuantos dientes cariados presenta el menor, obteniendo el siguiente evento asociado:

$$A = \{\text{El niño tiene al menos un diente con caries}\} \quad (7-3)$$

7.5.2. Modelo Estadístico

Sea Y la v.a. que identifica al evento A tal que:

$$Y = \begin{cases} 0, & \text{Ocurre } A' \\ 1, & \text{Ocurre } A \end{cases} \quad (7-4)$$

Por lo tanto $Y \sim \text{Bernoulli}(\theta)$. Cuya función de masa es:

$$f(y) = \theta^y (1 - \theta)^{1-y} \quad \theta \in (0, 1) \quad y \in \{0, 1\} \quad (7-5)$$

Se procede a formar grupos, los cuales se denotaron como: C_j , estos se conformaron teniendo en cuenta las combinaciones de todas las variables, es decir, cada covariable presenta tres estados (1, 2, 3). Por ende, se pueden formar 3^3 grupos. Por lo tanto C_j es la variable que identifica al j -ésimo grupo, tal que:

$$C_j = \{x_1, x_2, x_3\} \quad j = 1, \dots, 27 \quad (7-6)$$

Teniendo en cuenta los grupos conformados en (7-6), el paso a seguir es identificar los niños que presentan caries en cada grupo. Por lo tanto W_{ij} es la v.a tal que:

$$W_{ij} = \begin{cases} 1, & \text{El } i\text{-ésimo niño del } j\text{-ésimo grupo tiene caries} \\ 0, & \text{El } i\text{-ésimo niño del } j\text{-ésimo grupo no tiene caries} \end{cases} \quad (7-7)$$

Sea r_j la variable asociada a la variable W_{ij} tal que $r_j = \sum_{i=1}^{n_j} W_{ij}$, por lo tanto

$$r_j \sim \text{Binomial}(n_j, \theta_j) \quad \theta_j \in (0, 1) \quad j = 1, 2, \dots, 27 \quad (7-8)$$

Cuya función de masa es:

$$f(r_j) = K_j \theta_j^{r_j} (1 - \theta_j)^{n_j - r_j} \quad \theta \in (0, 1) \quad r_j \in \{0, 1, \dots, n_j\} \quad (7-9)$$

K_j es: $\binom{n_j}{r_j}$, donde n_j es el número de niños en el j -ésimo grupo y r_j es el número de niños con caries en el j -ésimo grupo.

7.5.3. Grupos Binomiales

En la sección 7.5.2 se conformaron grupos denominados Binomiales, con el fin de contar el número r_j de niños con caries dentro de cada grupo de n_j niños, caracterizados por la j -ésima combinación de categorías de las covariables. En la Tabla **7-3** se presenta la conformación de dichos grupos.

Tabla 7-3: Grupos Binomiales Formados a Partir de las Categorías de las Covariables

Grupo	pH	UFCLB	UFCSM	Grupo	pH	UFCLB	UFCSM
1	1	1	1	15	2	2	3
2	1	1	2	16	2	3	1
3	1	1	3	17	2	3	2
4	1	2	1	18	2	3	3
5	1	2	2	19	3	1	1
6	1	2	3	20	3	1	2
7	1	3	1	21	3	1	3
8	1	3	2	22	3	2	1
9	1	3	3	23	3	2	2
10	2	1	1	24	3	2	3
11	2	1	2	25	3	3	1
12	2	1	3	26	3	3	1
13	2	2	1	27	3	3	1
14	2	2	2				

Dada la anterior conformación de grupos Binomiales en la Tabla **7-3**, el siguiente paso fue clasificar los niños de cada grupo que tiene presencia de caries, es decir: Número de niños, Presencia de caries, Sin presencia de caries para cada grupo.

Tabla 7-4: Clasificación de Niños

Grupo	Niños por Grupo	Con Caries	Sin Caries	Grupo	Niños por Grupo	Con Caries	Sin Caries
1	41	15	26	15	0	0	0
2	1	0	1	16	1	0	1
3	11	7	4	17	0	0	0
4	0	0	0	18	0	0	0
5	0	0	0	19	7	5	2
6	0	0	0	20	0	0	0
7	36	23	13	21	0	0	0
8	2	1	1	22	0	0	0
9	16	12	4	23	0	0	0
10	2	0	2	24	1	1	0
11	0	0	0	25	5	4	1
12	0	0	0	26	0	0	0
13	0	0	0	27	2	2	0
14	0	0	0				

7.5.4. Grupos Binomiales Con Información Disponible

En la Tabla 7-4 se puede observar que se construyeron $3^3 = 27$ grupos, teniendo en cuenta las diferentes combinaciones de las covariables. En la base de datos se tienen 125 registros, por lo tanto hay grupos en los que no hubo niños, por esta razón en esta nueva etapa solo se tuvieron en cuenta los grupos para los cuales habían niños. Por lo tanto, después de realizar esta clasificación quedaron finalmente 12 grupos, distribuidos de acuerdo a la Tabla 7-5 y clasificados de acuerdo a la Tabla 7-6

Tabla 7-5: Grupos de Covariables

Grupo	pH	UFCLB	UFCSM
1	1 :< 7	1 :< 625000UFC/ml	1 :< 20000UFC/ml
2	1 :< 7	1 :< 625000UFC/ml	2 := 20000UFC/ml
3	1 :< 7	1 :< 625000UFC/ml	3 :> 20000UFC/ml
7	1 :< 7	3 :> 625000UFC/ml	1 :< 20000UFC/ml
8	1 :< 7	3 :> 625000UFC/ml	2 := 20000UFC/ml
9	1 :< 7	3 :> 625000UFC/ml	3 :> 20000UFC/ml
10	2 := 7	1 :< 625000UFC/ml	1 :< 20000UFC/ml
16	2 := 7	3 :> 625000UFC/ml	1 :< 20000UFC/ml
19	3 :> 7	1 :< 625000UFC/ml	1 :< 20000UFC/ml
24	3 :> 7	2 := 625000UFC/ml	3 :> 20000UFC/ml
25	3 :> 7	3 :> 625000UFC/ml	1 :< 20000UFC/ml
27	3 :> 7	3 :> 625000UFC/ml	3 :> 20000UFC/ml

Tabla 7-6: Grupos Finales de Estudio

Grupo	Niños por Grupo	Con Caries	Sin Caries
1	41	15	26
2	1	0	1
3	11	7	4
7	36	23	13
8	2	1	1
9	16	12	4
10	2	0	2
16	1	0	1
19	7	5	2
24	1	1	0
25	5	4	1
27	2	2	0

7.5.5. Distribución a Priori no Informativa

Una distribución a priori no informativa, no cuenta con información externa por lo contrario deja que los datos hablen por si solos. También es llamada a priori de referencia.

Distribución a Priori, Principio de Laplace

Dado que no se cuenta con información del comportamiento de θ_j , una excelente alternativa es asociar una distribución conjugada para la verosimilitud Binomial. En este caso se asocio una distribución Beta, con parámetros ($a_j = 1; b_j = 1$) asumiendo que todos los valores de θ_j tienen la misma probabilidad de ocurrir.

Distribución a Priori de Jeffrey

El principio de Jeffrey conduce a la definición de la densidad a priori no informativa como $h(\theta) \propto [J(\theta)]^{1/2}$, donde $J(\theta)$ es la información de Fisher para θ (Gelman et al., 2014). Dado lo anterior se puede concluir que la distribución a priori de Jeffrey es:

$$h(\theta_j) = \text{Beta}(1/2, 1/2) \quad (7-10)$$

7.5.6. Elicitación de la Distribución a Priori Informativa

Se define como θ_j la proporción de interés en el j –ésimo grupo, la cual es una cantidad aleatoria, con una distribución de probabilidad desconocida $\Pi(\theta_j)$. Una opción natural para $\Pi(\theta_j)$ es la distribución Beta con hiperparámetros a_j y b_j . Usando la información disponible

se obtiene un intervalo θ_{1j} y θ_{2j} , que explica en un $(1 - \alpha) \%$ de los valores de θ_j que caen dentro del intervalo $(\theta_{1j}, \theta_{2j})$, en segunda instancia se estima la media del intervalo dado, y se iguala a la media de la distribución Beta (Tovar, 2012).

$$\theta_{0j} = \frac{\theta_{1j} + \theta_{2j}}{2} \quad (7-11)$$

Como la media $\theta_{0j} = E(\theta_j)$ y la varianza $\sigma_j^2 = V(\theta_j)$ se pueden escribir como funciones de los hiperparámetros de la Beta a priori, entonces el problema se convierte en una cuestión de solucionar un sistema de dos ecuaciones con dos incógnitas (Tovar, 2012)

$$\begin{aligned} w_j &= \frac{\theta_{0j}}{1 - \theta_{0j}} \\ a_j &= w_j b_j \\ b_j &= \frac{w_j - [(w_j + 1)^2] \sigma_j^2}{(w_j^3 + 3w_j^2 + 3w_j + 1) \sigma_j^2} \end{aligned} \quad (7-12)$$

Teniendo en cuenta la mención que hace el Ministerio de Salud y Protección Social donde revela que en varios estudios han determinado una prevalencia de caries dental superior al 60 % en niños menores de seis años (González et al., 2009; Cadavid et al., 2010), podemos hacer uso de esta información para la construcción de nuestra distribución a priori informativa. De acuerdo al método de elicitación propuesto por Tovar (2012) y mediante métodos de Bayes, se estimó la media a posteriori, es decir, la proporción de niños con caries para cada grupo.

7.5.7. Distribución a Posteriori

La distribución a posteriori de las proporciones de niños con caries o probabilidad de tener caries dada la combinación de categorías de covariables, se obtiene actualizando la información expresada por la distribución a priori con la información contenida en la muestra de datos de cada grupo Binomial. Entonces, sea $f(\theta_j|r_j)$ la distribución de probabilidad a posteriori de la proporción de niños con caries que cumplen con la j -ésima combinación, tal que:

$$\begin{aligned} f(\theta_j|r_j) &\propto f(\theta_j)L(\theta_j|r_j) \quad j = 1, 2, \dots, 27 \\ &\propto \frac{\Gamma(a_j + b_j)}{\Gamma(b_j)} \theta^{a_j} (1 - \theta)^{b_j-1} \theta^{r_j} (1 - \theta)^{n_j-r_j} \\ &\propto \frac{\Gamma(a_j + b_j + n_j)}{\Gamma(a_j + r_j) \Gamma(r_j + b_j + n_j)} \theta^{a_j+r_j-1} (1 - \theta)^{b_j+n_j-r_j-1} \end{aligned} \quad (7-13)$$

Entonces $f(\theta_j|r_j) \sim \text{Beta}(\alpha, \beta)$. Donde $\alpha = a_j + r_j$ y $\beta = b_j + n_j - r_j$ son los parámetros de la distribución a posteriori de la proporción de niños con caries con las características propias

del grupo j , con $j = 1, 2, \dots, 27$. Finalmente n_j es la cantidad de niños en el respectivo grupo. Dado que la distribución Beta es conjugada para la verosimilitud Binomial, las distribuciones a posteriori obtenidas pertenecen a la familia Beta, la cual es una distribución bastante conocida, fácilmente programable en la mayoría de los software estadísticos existentes en el mercado. Se usaron métodos de Monte Carlo para aproximar las estimaciones posteriores y las regiones de credibilidad en R.

7.5.8. Probabilidades Predictivas

Se calcularon las probabilidades predictivas para cada grupo de covariables, es decir se calculó la probabilidad de tener caries de un nuevo niño, el cual se agrega al estudio y presenta alguna de las combinaciones de factores de riesgo que se tienen.

$$P(R_j \geq r_j | \theta_j) \approx \binom{n_j}{r_j} \theta_j^{r_j} (1 - \theta_j)^{n_j - r_j} \quad (7-14)$$

Donde $\hat{\theta}_{B_j}$ es el estimador de Bayes. La probabilidad presentada en 7-14, es la probabilidad de que un nuevo niño R_j presente caries ya que presenta la j -ésima combinación de covariables, tomando como la probabilidad de éxito la probabilidad a posteriori para este grupo.

7.6. Estimación de la Prevalencia de Caries Considerando el Grado de Severidad

7.6.1. Experimento Estadístico

Para una población de niños de 3 a 4 años vinculados al Instituto Colombiano de Bienestar Familiar (ICBF) en la ciudad de Cali, Colombia, se seleccionó una muestra de 125 niños. Se desea conocer el número de niños con el k -ésimo nivel de caries asociado (Sin Caries, Leve, Moderada, Severa), obteniendo el siguiente evento asociado:

$$B = \{\text{El niño presenta la } k\text{-ésima característica}\} \quad k=1,2,3,4 \quad (7-15)$$

7.6.2. Modelo Estadístico

Sea S_k la v.a. que identifica al evento B tal que

$$S_k = \begin{cases} 0, & \text{Ocurre } B \\ 1, & \text{Ocurre } B' \end{cases} \quad (7-16)$$

De acuerdo a la categorización de las variables en la tabla 7-2 y teniendo en cuenta los grupos conformados en 7-6, el paso a seguir es identificar la cantidad de niños que presentan

algún nivel de caries asociado en cada grupo. Sea entonces $Z_{jk} = \sum_{i=1}^{n_j} s_{ijk}$, por lo tanto Z_{jk} es la v.a tal que:

$$Z_{jk} = \begin{cases} 1, & \text{Número de niños sin caries pertenece al } j\text{-ésimo grupo} \\ 2, & \text{Número de niños con caries leve pertenece al } j\text{-ésimo grupo} \\ 3, & \text{Número de niños con caries moderada pertenece al } j\text{-ésimo grupo} \\ 4, & \text{Número de niños con caries severa pertenece al } j\text{-ésimo grupo} \end{cases} \quad (7-17)$$

De modo que $Z_{jk} \sim Multinomial(n_j, \pi_{j1}, \pi_{j2}, \pi_{j3}, \pi_{j4})$. Cuya función de masa es:

$$P(Z_{j1} = z_{j1}, \dots, Z_{j4} = z_{j4} | \pi_{j1}, \dots, \pi_{j4}) = \frac{n}{z_{j1}! z_{j2}! z_{j3}! z_{j4}!} \prod_{k=1}^4 \pi_{jk}^{z_{jk}} \quad (7-18)$$

Sujeto a:

$$\pi_{jk} \in (0, 1) \quad \sum_{k=1}^4 z_{jk} = n_j \quad \sum_{k=1}^4 \pi_{jk} = 1 \quad (7-19)$$

7.6.3. Grupos Multinomiales Con Información Disponible

De acuerdo a los 12 grupos finalmente conformados en las Tablas 7-5 y 7-6, se formaron grupos multinomiales. En la Tabla 7-7 aparece la distribución de los niños teniendo en cuenta el nivel de caries que tienen.

Tabla 7-7: Nueva Clasificación de Niños por Nivel de Caries

Grupo	Niños por Grupo	Con Caries	Sin Caries	Caries Leve	Caries Moderada	Caries Severa
1	41	15	26	7	1	7
2	1	0	1	0	0	0
3	11	7	4	2	1	4
7	36	23	13	12	3	8
8	2	1	1	0	0	1
9	16	12	4	0	0	12
10	2	0	2	0	0	0
16	1	0	1	0	0	0
19	7	5	2	3	1	1
24	1	1	0	0	0	1
25	5	4	1	2	1	1
27	2	2	0	1	0	1

7.6.4. Distribución a Priori no Informativa

Distribución a Priori, Principio de Laplace

Dado que no se cuenta con información del comportamiento de π_{jk} , una excelente alternativa es asociar una distribución conjugada para la verosimilitud Multinomial. En este caso se asocio una distribución Dirichlet, con parámetros $\alpha_{jk} = 1$.

7.6.5. Elicitación de la Distribución a Priori Informativa

Rivera and Morales (2015) proponen una metodología de elicitación para una distribución multinomial, la cual consta de una mezcla entre las técnicas de elicitación HFS (muestras hipotéticas futuras) y método de la ruleta. El método se ejecuta a través de tres etapas:

Etapas 1: Estimación del n-equivalente: Si no se cuenta con información a priori sobre la probabilidad de cada categoría, se puede estimar el valor del n-equivalente fijando un valor para d y dejando el conocimiento del experto sobre la variable de interés en términos del nivel de significancia α . Por ejemplo, el facilitador puede asignar un valor de $\alpha = 0,20$ si considera que el conocimiento del experto es alto, asignar un valor de $\alpha = 0,40$ si considera que el nivel de conocimiento del experto es medio o asignar un valor de $\alpha = 0,60$ si considera que el nivel de conocimiento del experto bajo. En la Tabla 7-8 se muestran los valores estimados del n-equivalente para diferentes valores de α , d y k , donde, k corresponde al número de categorías de la variable de interés.

Etapas 2: Estimación de la frecuencia relativa: En esta etapa a diferencia del método propuesto por el autor, se hace uso de métodos empíricos de bayes, donde se calcularon las probabilidades de ocurrencia de cada categoría para el grupo que contenía información, y se determinaron las probabilidades mínimas distintas de cero para cada categoría, quedando evidencia de las probabilidades usadas en esta etapa.

Etapas 3: Simulación: Una vez el facilitador ha estimado el valor del n-equivalente en el paso 1 y el vector de probabilidad típica en el paso 2 se procede con la simulación de una distribución Multinomial por medio de la función *rmultinom* del software R. La función *rmultinom* recibe tres parámetros. El primero corresponde al número de simulaciones que deseamos realizar. El segundo es el tamaño de muestra equivalente (estimado en el paso 1). El tercero es el vector de probabilidades de cada categoría (estimado en el paso 2).

Por ultimo se hace la estimación del vector de parámetros $\alpha^* = [\alpha_1, \dots, \alpha_4]$ de la distribución Dirichlet. Dado que cada z_{jk} simulado en el paso 3 sigue una distribución Multinomial, entonces $Y_{jk} = Z_{jk}/n$ tiene una distribución Dirichlet con un vector de parámetros $\alpha^* = [\alpha_{j1}, \dots, \alpha_{j4}]$, donde n es el tamaño muestral estimado en el paso 1.

Tabla 7-8: Tamaño Muestral Equivalente con un Nivel de Significancia α y un Porcentaje de Error d para Probabilidades Multinomiales (Rivera and Morales, 2015)

	α								
	0,20			0,40			0,60		
	d			d			d		
k	0,05	0,075	0,10	0,05	0,075	0,10	0,05	0,075	0,10
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
11	553	243	134	434	191	106	366	161	89
12	568	249	138	449	197	109	381	167	93
13	582	256	141	462	203	112	394	173	96
14	595	261	145	475	209	116	407	179	99
15	607	267	147	487	214	118	418	184	102

7.6.6. Distribución a Posteriori

La distribución a posteriori de las proporciones de niños con algún nivel de caries o probabilidad de tener caries de algún tipo dada la combinación de categorías de covariables, se obtiene actualizando la información expresada por la distribución a priori con la información contenida en la muestra de datos de cada grupo Multinomial. Entonces, sea $f(\pi_{jk}|z_{jk})$ la distribución de probabilidad a posteriori de la proporción de niños con algún nivel de caries asociado que cumplen con la j –ésima combinación, tal que:

$$\begin{aligned}
 f(\pi_{jk}|z_{jk}) &\propto f(\pi_{jk})L(\pi_{jk}|z_{jk}) \quad j = 1, 2, \dots, 27 \\
 &\propto \left(\prod_{k=1}^4 \pi_{jk}^{\alpha_{jk}-1}\right) \left(\prod_{k=1}^4 \pi_{jk}^{z_{jk}}\right) \\
 &\propto \prod_{k=1}^4 \pi_{jk}^{z_{jk}+\alpha_{jk}-1}
 \end{aligned} \tag{7-20}$$

Entonces $f(\pi_{jk}|z_{jk}) \sim \text{Dirichlet}(\alpha^*)$. Donde $\alpha^* = (z_{j1} + \alpha_{j1}, \dots, z_{j4} + \alpha_{j4})$ son los parámetros de la distribución a posteriori de la proporción de niños con algún nivel de caries asociado,

teniendo en cuenta las características propias del grupo j , con $j = 1, 2, \dots, 27$.

Dado que la distribución Dirichlet es conjugada para la verosimilitud Multinomial, las distribuciones a posteriori obtenidas pertenecen a la familia Dirichlet, la cual es una distribución programable en los software estadísticos utilizados en este trabajo de grado. Se usaron métodos de Monte Carlo para aproximar las estimaciones posteriores y las regiones de credibilidad.

8 Resultados

8.1. Análisis Descriptivo de los Datos

Se realizó un análisis exploratorio de las variables de interés. Inicialmente, la distribución de la variable respuesta se compone de dos grupos de niños, uno con 70 (56 %) casos de presencia de caries y otro de 55 (44 %) que no presentan caries, 60 del género femenino (48 %) y 65 del género masculino (52 %). Para el caso donde la variable de respuesta presenta cuatro categorías, 55 (44 %) niños no presentaron caries, 27 (21,6 %) niños presentaron caries leve, 7 (5,6 %) niños presentaron caries moderada y 36 (28,8 %) niños presentaron caries severa.

Tabla 8-1: Análisis Exploratorio

Variable	Mínimo	Máximo	Media	Mediana	Desviación Estándar	Coefficiente de variación
pH	2,25	7,73	6,42	6,43	0,6569	10,23 %
UFCLB	0 UFC/ml	3900000 UFC/ml	764400 UFC/ml	625000 UFC/ml	631679,5 UFC/ml	83,63 %
UFCSM	0 UFC/ml	2125000 UFC/ml	92880 UFC/ml	0 UFC/ml	339502,3 UFC/ml	365,52 %

En la Tabla **8-1** aparece el resumen descriptivo para las variables pH, las Unidades Formadoras de Colonias de Streptococcus Mutans en recuentos de UFC por mililitro de saliva y Unidades Formadoras de Colonias de Lactobacillus en recuentos de UFC por mililitro de saliva. El pH medio fue de 6,42 con una desviación estándar de 0,6569 fluctuando en un rango entre 2,25 y 7,73, el 50 % de los niños presentan pH inferior a 6,43, y la variabilidad es de 10,23 %. La media de UFCLB es de 764400 UFC/ml con una desviación estándar de 631679,5 UFC/ml fluctuando en un rango entre 0 UFC/ml y 3900000 UFC/ml, el 50 % de los niños presentan UFCLB inferiores a 625000 UFC/ml, y la variabilidad relativa es de 83,63 %. La media de UFCSM es de 92880 UFC/ml con una desviación estándar de 339502,3 UFC/ml fluctuando en un rango entre 0 UFC/ml y 2125000 UFC/ml, el 50 % de los niños presentan UFCSM de 0 UFC/ml, y una variabilidad relativa de 365,52 %.

Para la variable cuantitativa pH se tomó de referencia las categorías (< 7 , $= 7$ y > 7) utilizadas en el estudio realizado por Liso et al. (2000).

Para la variable cuantitativa UFCLB se evaluó mediante el segundo cuantil (mediana) la distribución con respecto a la variable de respuesta, obteniéndose una adecuada distribución de los datos, sin presentar ceros en las celdas de niños con caries, por tanto, en este caso se utilizó el segundo cuantil (mediana) para la categorización.

Para la variable cuantitativa UFCSM se evaluó mediante el tercer cuantil la distribución con respecto a la variable de respuesta, obteniéndose una adecuada distribución de los datos, sin celdas en cero para los niños con caries, por tanto, en este caso se utilizó el tercer cuantil para la categorización.

Finalmente, estas covariables se categorizan de acuerdo a los estadísticos de posición. Por tanto, mediante la Tabla 8-2 se muestra la distribución conjunta de las covariables y la variable de respuesta, como también los resultados de los *valores p* mediante las pruebas de chi-cuadrado para analizar las relaciones entre ellas. Como se puede observar todas las covariables presentan valores inferiores al nivel de significancia 0.05, por lo tanto, hay evidencia estadística suficiente para rechazar la hipótesis nula de no asociación entre las variables, es decir, que la variable de respuesta parece estar relacionada con el pH, UFCLB y UFCSM.

Tabla 8-2: Distribución Conjunta de las Covariables y Variable de Respuesta

Variable		Presencia Caries				Valor-p
		Si		No		
		Niños	%	Niños	%	
pH	< 7	58	54,2 %	49	45,8 %	0,024
	≥ 7	12	66,7 %	3	33,3 %	
UFCLB	< 625000	27	43,5 %	35	56,5 %	0,017
	≥ 625000	43	68,3 %	20	31,7 %	
UFCSM	< 20000	47	51,1 %	45	48,9 %	0,075
	≥ 20000	23	69,7 %	10	30,3 %	

8.2. Proporción de Niños con Caries

En esta sección se presentan estimaciones de la proporción de niños con presencia de caries, obtenidas mediante métodos propios de la estadística bayesiana como es la distribución a posteriori, utilizando una distribución a priori informativa elicitada mencionada en la sección (7.5.6) y distribuciones a priori no informativas. A continuación, se muestran las distribuciones a priori utilizadas en este estudio, junto con sus respectivos parámetros.

Distribución a Priori no Informativas

Se utilizaron dos tipos de a priori no informativas conjugadas:

- **Distribución a priori principio de Laplace:** Se utilizó una distribución Beta con parámetros ($a_j = 1, b_j = 1$) asumiendo que todos los valores de θ_j tienen la misma probabilidad de ocurrir.
- **Distribución a priori de Jeffreys:** Se utilizó una distribución Beta con parámetros ($a_j = 1/2, b_j = 1/2$) teniendo en cuenta la información de Fisher para θ_j .

Distribución a Priori Informativa

Para la construcción de esta distribución se tomó en cuenta la metodología planteada en la sección 7.5.6, donde se tuvo en cuenta que el 60 % de los niños menores de seis años cuentan con presencia de caries. Dado lo anterior se puede construir un intervalo (0,60 - 1) el cual se toma de referencia para la construcción de una a priori informativa.

Tabla 8-3: Proporción de Niños con Caries (A Priori No Informativas)

Grupo	Estimación por Método de Bayes a Priori Laplace	Intervalos de Credibilidad	Estimación por Método de Bayes a Priori Jeffrey	Intervalos de credibilidad
1	0,37	0,241 - 0,516	0,37	0,237 - 0,515
2	0,33	0,010 - 0,830	0,26	0,000 - 0,864
3	0,61	0,361 - 0,846	0,62	0,374 - 0,867
7	0,63	0,490 - 0,777	0,64	0,489 - 0,785
8	0,49	0,098 - 0,901	0,50	0,066 - 0,937
9	0,72	0,497 - 0,906	0,74	0,517 - 0,920
10	0,25	0,010 - 0,683	0,15	0,000 - 0,619
16	0,33	0,012 - 0,850	0,25	0,000 - 0,852
19	0,67	0,340 - 0,904	0,69	0,338 - 0,934
24	0,66	0,157 - 0,987	0,75	0,143 - 1,000
25	0,71	0,351 - 0,957	0,76	0,360 - 0,978
27	0,75	0,277 - 0,993	0,83	0,347 - 1,000

En la Tabla 8-3, se muestran las probabilidades de niños con caries para cada grupo, mediante métodos de Bayes usando una distribución a priori de Laplace y una distribución a priori de Jeffrey. Las probabilidades más altas de niños con caries teniendo en cuenta las respectivas combinaciones de covariables en cada grupo, se encuentran en negrilla.

Tabla 8-4: Proporción de Niños Con Caries (A Priori Informativa)

Grupo	Estimación por Método de Bayes a Priori Informativa	Intervalos de Credibilidad
1	0,37	0,241 - 0,520
2	0,34	0,001 - 0,925
3	0,59	0,380 - 0,807
7	0,64	0,498 - 0,790
8	0,58	0,103 - 0,976
9	0,75	0,538 - 0,930
10	0,21	0,001 - 0,767
16	0,36	0,002 - 0,937
19	0,72	0,381 - 0,952
24	0,91	0,331 - 1,000
25	0,80	0,392 - 0,990
27	0,95	0,555 - 1,000

La Tabla 8-4 contiene las probabilidades de niños con presencia de caries para cada grupo, usando una distribución a priori informativa. Se puede observar que las probabilidades calculadas utilizando la distribución a priori informativa, son mas altas que las a priori no informativas. Por lo que se podría decir que, estas combinaciones de factores de riesgo son muy influyentes al momento de que un niño presente caries en sus dientes.

Las distribuciones a priori no informativas construidas por el método de Laplace y Jeffreys tienen como objetivo demostrar, que los métodos propios de Bayes son mas eficientes e informativos en este tipo de casos en donde el tamaño de muestra es muy pequeño.

8.2.1. Probabilidades Predictivas

Desde el punto de vista médico es importante tener una referencia en caso de la llegada de un nuevo niño con alguna de las combinaciones de factores de riesgo mencionados anteriormente, con el fin de poder conocer cual es la probabilidad de que un niño tenga presencia de caries, para esto se cuenta con las probabilidades predictivas las cuales tienen precisamente este significado. En las Tablas 8-4, se muestran estos valores predictivos para cada grupo teniendo en cuenta el uso de la distribución a priori informativa.

Tabla 8-5: Probabilidades Predictivas Mediante la a Posteriori Informativa

Grupo	Predictiva
1	0,128
2	0,656
3	0,230
7	0,137
8	0,486
9	0,225
10	621
16	0,645
19	0,318
24	0,912
25	0,410
27	0,894

De acuerdo a las Tabla **8-5**, se realizó un resumen con las probabilidades predictivas.

- Grupo 1, probabilidad predictiva: 0,128. La cual es una probabilidad baja, dicho grupo incluye los factores de riesgo: $PH < 7$, $UFCLB < 625000 UFC/ml$ y $UFCSM < 20000 UFC/ml$.
- Grupo 16, probabilidad predictiva: 0,645. Incluye los factores de riesgo: $PH = 7$, $UFCLB > 625000 UFC/ml$ y $UFCSM < 20000 UFC/ml$.
- Grupo 24, probabilidad predictiva: 0,912. La cual es una probabilidad alta, dicho grupo incluye los factores de riesgo: $PH > 7$, $UFCLB = 625000 UFC/ml$ y $UFCSM > 20000 UFC/ml$.
- Grupo 27, probabilidad predictiva: 0,894. La cual es una probabilidad relativamente alta, dicho grupo incluye los factores de riesgo: $PH > 7$, $UFCLB > 625000 UFC/ml$ y $UFCSM > 20000 UFC/ml$.
- Haciendo una comparación entre los grupos 24 y 27, se puede decir que tener $PH > 7$ y $UFCSM > 20000 UFC/ml$ juega un papel importante en la presencia de caries de un niño, ya que un niño con $PH > 7$, $UFCLB = 625000 UFC/ml$ y $UFCSM > 20000 UFC/ml$ tiene una probabilidad de presencia de caries alta, similar a la de un niño con $PH > 7$, $UFCLB > 625000 UFC/ml$ y $UFCSM > 20000 UFC/ml$.
- De la comparación entre los grupos 1 y 16 se puede decir que los niños con $PH < 7$ y $UFCLB < 625000 UFC/ml$ tienden a presentar menos probabilidad de presentar caries.

8.3. Proporción de Niños Con Caries Considerando el Grado de Severidad

En esta sección se presentan estimaciones de la proporción de niños con presencia de algún nivel caries, obtenidas mediante métodos propios de la estadística bayesiana como es la distribución a posteriori, utilizando una distribución a priori informativa elicitada y una distribución a priori no informativa. A continuación, se muestran las distribuciones a priori utilizadas en este estudio, junto con sus respectivos parámetros.

Distribución a Priori no Informativa

Se utilizó solo un tipo de a priori no informativa conjugada:

- **Distribución a priori principio de Laplace:** Se utilizó una distribución Dirichlet con parámetros $\alpha_j = 1$.

Tabla 8-6: Proporción de Niños con Nivel de Caries Asociado (Laplace)

Grupo	Estimación por Metodo de Bayes a Priori Laplace							
	Sin Caries		Caries Leve		Caries Moderada		Caries Severa	
	Estimación	RC	Estimación	RC	Estimación	RC	Estimación	RC
1	0,60	0,461 - 0,742	0,18	0,078 - 0,289	0,04	0,005 - 0,118	0,18	0,083 - 0,303
2	0,40	0,079 - 0,800	0,19	0,007 - 0,616	0,21	0,011 - 0,586	0,20	0,008 - 0,589
3	0,33	0,132 - 0,572	0,20	0,046 - 0,408	0,13	0,016 - 0,333	0,34	0,128 - 0,577
7	0,35	0,202 - 0,494	0,33	0,193 - 0,474	0,10	0,029 - 0,218	0,22	0,107 - 0,360
8	0,34	0,059 - 0,731	0,17	0,005 - 0,513	0,17	0,006 - 0,552	0,33	0,048 - 0,714
9	0,25	0,098 - 0,456	0,05	0,002 - 0,161	0,05	0,001 - 0,173	0,65	0,435 - 0,842
10	0,50	0,131 - 0,831	0,16	0,005 - 0,518	0,17	0,006 - 0,528	0,17	0,005 - 0,509
16	0,39	0,060 - 0,793	0,20	0,006 - 0,599	0,20	0,004 - 0,613	0,20	0,007 - 0,595
19	0,27	0,070 - 0,551	0,36	0,121 - 0,661	0,18	0,025 - 0,460	0,18	0,024 - 0,437
24	0,20	0,007 - 0,625	0,20	0,010 - 0,590	0,20	0,006 - 0,570	0,40	0,071 - 0,817
25	0,23	0,029 - 0,543	0,33	0,080 - 0,661	0,22	0,030 - 0,503	0,22	0,039 - 0,500
27	0,17	0,004 - 0,513	0,33	0,057 - 0,680	0,17	0,005 - 0,523	0,34	0,053 - 0,728

De acuerdo a los datos presentados en la tabla 7-7, podemos realizar las estimaciones, teniendo en cuenta que se hace para cada grupo, por medio de métodos bayesianos, usando una a priori no informativa. En las tablas 8-6 se presentan los resultados obtenidos.

8.3.1. Distribución a Priori Informativa

Para la construcción de esta distribución se tomo en cuenta la metodología planteada en la sección 7.6.5, donde se elicitaron los parámetros del vector α de la distribución dirichlet, mediante el método propuesto por Rivera and Morales (2015).

De forma hipotética, con base en su cercanía con el área de odontología, se considera a una odontóloga como un experto con alto nivel de experiencia por lo que se le asigna un n-equivalente de 381, donde se distribuyen las probabilidades de ocurrencia mínimas (obtenidas con métodos empíricos de bayes) en las diferentes categorías de la variable a elicitar $\pi_{jkmin} = (0,66, 0,17, 0,02, 0,14)$. El n-equivalente corresponde a un $\alpha = 0,20$ y un error de estimación igual al 5 % (ver Tabla 7-8; note que aquí α tiene que ver con el nivel de significancia en la estimación del tamaño muestral y no con el vector de parámetros de la distribución Multinomial).

Una vez estimado el valor del n-equivalente y el vector de probabilidad típica, se lleva a cabo el proceso de simulación y estimación del vector de parámetros $\vec{\alpha}^* = (\alpha_1, \dots, \alpha_4)$. Para dicho proceso solo es necesario introducir los valores estimados dentro del software R, el cual se encarga de realizar la simulación y entregar los valores estimados de $\alpha_1, \alpha_2, \alpha_3, \alpha_4$.

Las estimaciones obtenidas para la distribución Dirichlet son:

Tabla 8-7: Vector de Parámetros α con su Respectiva Media y Varianza para los Valores Elicitados.

α	Media	Varianza
255,20	0,66	0,00058
64,13	0,17	0,00036
6,96	0,018	0,00046
59,18	0,15	0,00034

Tabla 8-8: Proporción de Niños con Nivel de Caries Asociado (A Priori Informativa)

	Estimación por Metodo de Bayes a Priori Informativa							
	Sin Caries		Caries Leve		Caries Moderada		Caries Severa	
Grupo	Estimación	RC	Estimación	RC	Estimación	RC	Estimación	RC
1	0,66	0,610 - 0,704	0,17	0,131 - 0,203	0,02	0,008 - 0,033	0,16	0,122 - 0,192
2	0,66	0,613 - 0,706	0,17	0,133 - 0,204	0,02	0,007 - 0,033	0,15	0,120 - 0,193
3	0,65	0,604 - 0,700	0,17	0,132 - 0,206	0,02	0,008 - 0,035	0,16	0,123 - 0,199
7	0,64	0,589 - 0,683	0,18	0,145 - 0,220	0,02	0,011 - 0,039	0,16	0,128 - 0,197
8	0,66	0,617 - 0,708	0,17	0,130 - 0,206	0,02	0,007 - 0,034	0,16	0,123 - 0,193
9	0,65	0,599 - 0,689	0,16	0,127 - 0,197	0,02	0,007 - 0,031	0,18	0,138 - 0,214
10	0,66	0,618 - 0,709	0,17	0,132 - 0,201	0,02	0,007 - 0,033	0,15	0,118 - 0,192
16	0,66	0,615 - 0,713	0,17	0,130 - 0,206	0,02	0,008 - 0,033	0,15	0,120 - 0,190
19	0,66	0,609 - 0,702	0,17	0,137 - 0,209	0,02	0,009 - 0,036	0,15	0,119 - 0,189
24	0,66	0,617 - 0,705	0,17	0,129 - 0,204	0,02	0,007 - 0,034	0,16	0,123 - 0,195
25	0,66	0,609 - 0,702	0,17	0,134 - 0,208	0,02	0,008 - 0,036	0,15	0,121 - 0,192
27	0,66	0,606 - 0,703	0,17	0,134 - 0,207	0,02	0,008 - 0,033	0,16	0,120 - 0,192

La Tabla **8-8** contiene las probabilidades de niños con presencia de caries de algún tipo para cada grupo, usando una distribución a priori informativa. Se puede observar que las probabilidades calculadas utilizando la distribución a priori informativa, son mas uniformes que las a priori no informativas, que presentan mayor variabilidad.

9 Conclusiones y Recomendaciones

Los estudios clínicos de seguimientos de pacientes son importantes para el especialista, ya que puede estar informado de los posibles sucesos de salud a los que puede estar sometido el implicado. La metodología aplicada en el estudio mediante métodos estadísticos bajo el enfoque bayesiano fue pertinente, dado que el tamaño de muestra y la selección de la misma no permitía usar métodos clásicos, además del poder tener información de la cantidad aleatoria de interés por medio de una a priori informativa.

Realizar todas las posibles combinaciones de covariables fue oportuno, ya que esto permitió abarcar todos los posibles estados que puede presentarse en un niño teniendo en cuenta solo los factores de riesgo mencionados anteriormente. Por lo tanto, la metodología aplicada en el estudio dio como solución o información las probabilidades estimadas de niños con presencia de caries y también del nivel de caries asociado para cada grupo, lo cual es útil desde el punto de vista médico. Además para el caso donde la variable de respuesta fue de tipo binario, el hecho de calcular las probabilidades predictivas para un nuevo individuo que se incluya en el estudio es importante, ya que de acuerdo a qué tan grande sea la probabilidad, se tomará medidas de prevención en el seguimiento del paciente.

En cuanto a la obtención de los resultados, se puede decir que los grupos que mayor probabilidad presentan al tener al menos un diente cariado, involucran factores de riesgo tales como: $pH > 7$ y $UFCSM > 20000$. Por otra parte, hay grupos que tienen probabilidades relativamente bajas que incluyen algunos de estos factores de riesgo. Una explicación para esto puede ser que se observó los factores de riesgo de forma conjunta (las combinaciones entre ellos) y no de manera marginal (factor por factor).

Cuando se observan los factores de riesgo de manera conjunta, tener $pH > 7$ y $UFCSM > 20000$ hace que la probabilidad de que un niño presente caries en al menos un diente sea alta, a diferencia de un niño que no presente estos factores de riesgo. Si un niño tiene $pH = 7$ y $UFCSM < 20000$ hace que la probabilidad de que un niño presente caries en al menos un diente disminuya notablemente.

Cuando la variable de respuesta tiene una estructura ordinal, los resultados usando la distribución a priori informativa obtenida a través de métodos empíricos de bayes, permitió establecer mejores resultados, más precisos pese a que para los niveles de caries las estimaciones para cada grupo no difieren notoriamente. Se recomienda establecer distribuciones a priori teniendo en cuenta la estructura de cada grupo.

Uno de los aspectos importantes del trabajo de grado y de la metodología implementada en

el mismo, es su generalidad y sus limitaciones. Además de que tan eficaz y propicio puede ser para otro estudio similar. En este sentido, su generalidad es muy complicada si se tiene en cuenta la misma metodología, ya que los resultados obtenidos sólo pueden ser de ayuda desde el punto de vista clínico, para los niños que cuentan con alguna de las 12 combinaciones finalmente seleccionadas. Por lo tanto, los niños que no cuenten con alguna combinación no le será útil los resultados obtenidos en el trabajo.

Debido a la estructura ordinal de la variable de respuesta, se podría construir un modelo de regresión ordinal, tomando los tres factores de riesgos como covariables. Para esto, se debe establecer una metodología que permita realizar la modelación a través de Modelos Lineales Generalizados (GLM), lo cual puede ser una gran propuesta para un posterior trabajo de grado, y de gran ayuda desde el punto de vista clínico.

Por otro lado, el trabajo puede ser aplicable en otros estudios similares, dado que al momento de implementar la metodología estadística lo que se desea es que otras personas puedan usarla y aplicarla, en este caso lo es, debido a que los pacientes en estudio presentan características medicas o fisiológicas distintas, permitiendo en este caso la construcción y desarrollo del trabajo de grado actual.

Bibliografía

- Ayçaguer, L. C. S. (1994). *Excursión a la regresión logística en ciencias de la salud*. Ediciones Díaz de Santos.
- Badet, C. and Thebaud, N. (2008). Ecology of lactobacilli in the oral cavity: a review of literature. *The open microbiology journal*, 2:38.
- Banting, D., Eggertsson, H., Ekstrand, K., Zandoná, A. F., Ismail, A., Longbottom, C., Pitts, N., Reich, E., Ricketts, D., Selwitz, R., et al. (2005). Rationale and evidence for the international caries detection and assessment system (icdas ii). *Ann Arbor*, 1001:48109–1078.
- Bayes, T. (1763). *A letter from the late Reverend Mr. Thomas Bayes, FRS to John Canton, MA and FRS*. Philosophical Transactions (1683-1775).
- Bazán, J. and Bayes, C. (2010). Inferencia bayesiana en modelos de regresión binaria usando brmuw.
- Cadavid, A. S., Lince, C. M. A., and Jaramillo, M. C. (2010). Dental caries in the primary dentition of a colombian population according to the icdas criteria. *Brazilian oral research*, 24(2):211–216.
- Cameron, A. C., Widmer, R. P., and Martínez, A. B. (1998). *Manual de odontología pediátrica*. Harcourt Brace.
- Casanova, A., Medina, C., Casanova, J. F., Vallejos, A., Maupomé, G., and Kageyama, M. d. l. L. (2005). Factores asociados a la pérdida del primer molar permanente en escolares de campeche, méxico: Pérdida del primer molar permanente. *Acta Odontológica Venezolana*, 43(3):268–275.
- DeGroot, M. H. (2004). *Optimal statistical decisions*, volume 82. John Wiley & Sons.
- Díaz, A. C. (2002). *Fundamentos de química analítica: equilibrio iónico y análisis químico*, volume 4. Univ. Nacional de Colombia.
- Elfadaly, F. and Garthwaite, P. (2012). On eliciting some prior distributions for multinomial models. *Department of Mathematics and Statistics, The Open University, UK*.

- García, F. and Hicks, M. J. (2008). Maintaining the integrity of the enamel surface: the role of dental biofilm, saliva and preventive agents in enamel demineralization and remineralization. *The Journal of the American Dental Association*, 139:25S–34S.
- García, R., Depetris Guiguet, E., and Rossini, G. (2011). Un modelo de elección multinomial de consumo de alimentos fuera del hogar con datos de encuesta de hogares. *Cuadernos del CIMBAGE*, (13).
- Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2014). *Bayesian data analysis*. Chapman & Hall Book, Boca Raton, FL, USA.
- Glavina, D., Goršeta, K., Škrinjaric, I., Negovetić Vranić, D., Mehulić, K., and Kožul, K. (2012). Effect of lgg yoghurt on streptococcus mutans and lactobacillus spp. salivary counts in children. *Collegium antropologicum*, 36(1):129–132.
- González, F., Sánchez, R., and Carmona, L. (2009). Indicadores de riesgo para la caries dental en niños preescolares de la boquilla, cartagena. *Revista de salud pública*, 11(4):620–630.
- Johnson, N. L., Kemp, A. W., and Kotz, S. (2005). *Univariate discrete distributions*, volume 444. John Wiley & Sons.
- Liso, M. R. J., de Manuel Torres, E., García, F. G., and López, F. S. (2000). La utilización del concepto de ph en la publicidad y su relación con las ideas que manejan los alumnos: aplicaciones en el aula. *Enseñanza de las ciencias: revista de investigación y experiencias didácticas*, 18(3):451–461.
- Medina, C. E., Maupomé, G., Pelcastre, B., Avila, L., Vallejos, A. A., and Casanova, A. J. (2006). Desigualdades socioeconómicas en salud bucal: caries dental en niños de seis a 12 años de edad. *Revista de investigación clínica*, 58(4):296–304.
- Min., d. S. and Protección, S. (2013). Plan decenal de salud pública pdsp, 2012-2021: La salud en colombia la construyes tú.
- Morera, R. G. (2007). Plan nacional de salud pública 2007-2010. *Medicina*, 29(3):149–150.
- Näse, L., Hatakka, K., Savilahti, E., Saxelin, M., Pönkä, A., Poussa, T., Korpela, R., and Meurman, J. H. (2001). Effect of long-term consumption of a probiotic bacterium, lactobacillus rhamnosus gg, in milk on dental caries and caries risk in children. *Caries research*, 35(6):412–420.
- Organization, W. H. et al. (2008). Commission on the social determinants of health. *Geneva: World Health Organization*.

- Parisotto, T. M., Steiner-Oliveira, C., Silva, C. M. S. E., Rodrigues, L. K. A., and Nobre-dos Santos, M. (2010). Early childhood caries and mutans streptococci: a systematic review. *Oral health & preventive dentistry*, 8(1).
- Peres, M., Sheiham, A., Liu, P., Demarco, F., Silva, A., Assunção, M., Menezes, A., Barros, F., and Peres, K. (2016). Sugar consumption and changes in dental caries from childhood to adolescence. *Journal of dental research*, 95(4):388–394.
- Prieto, V. H., Quintero, C., Rodríguez, I., et al. (1995). Análisis de bayes empírico mediante un ejemplo. *Revista Colombiana de Estadística; Vol. 16, núm. 31-32 (1995) Revista Colombiana de Estadística; Vol. 16, núm. 31-32 (1995) 0120-1751*.
- Regazzini, E. and Sazonov, V. V. (1999). Approximation of laws of multinomial parameters by mixtures of dirichlet distributions with applications to bayesian inference. *Acta Applicandae Mathematicae*, 58(1):247–264.
- Rivera, A. F. F. and Morales, J. C. C. (2015). Una propuesta metodológica para elicitar el vector de parámetros π de la distribución multinomial. *Comunicaciones en Estadística*, 8(1):81–97.
- Schwendicke, F., Paris, S., and Tu, Y.-K. (2015). Effects of using different criteria for caries removal: a systematic review and network meta-analysis. *Journal of dentistry*, 43(1):1–15.
- Selwitz, R. H., Ismail, A. I., and Pitts, N. B. (2007). Dental caries. *The Lancet*, 369(9555):51–59.
- Sheiham, A. (2005). Oral health, general health and quality of life. *Bulletin of the World Health Organization*, 83(9):644–644.
- Tovar, J. R. C. (2012). Eliciting beta prior distributions for binomial sampling. *Rev. Bras. Biom.*
- Tovar, J. R. C. (2015). Inferencia bayesiana e investigación en salud: un caso de aplicación en diagnóstico clínico. *Revista Medica de Risaralda*.
- Witjaksono, W. (2013). Saliva ph changes in patients with high and low caries risk after consuming organic (sucrose) and non-organic (maltitol) sugar. *The International Medical Journal of Malaysia*, 12(2).
- Zapata, R. E., O’Hagan, A., and Soares Bastos, L. (2014). Eliciting expert judgements about a set of proportions. *Journal of Applied Statistics*, 41(9):1919–1933.