



**UNIVERSIDAD AUTÓNOMA DE GUERRERO**

Unidad Académica de Matemáticas

---

---

Maestría en Matemáticas Aplicadas

**Modelación Bayesiana de datos  
multinivel en enfermedad  
periodontal**

**T E S I S**

PARA OBTENER EL GRADO DE:

**Maestra en Matemáticas Aplicadas**

PRESENTA:

**Edilberta Tino Salgado**

DIRECTORES DE TESIS:

MC. Cruz Vargas de León

Dr. Flaviano Godínez Jaimes

Diciembre de 2022.

(Página en blanco)

---

## **Dedicatoria**

---

Dedico este trabajo de tesis a mi misma por la entrega con que lo cree, a mis padres, hermanos, a la comunidad académica y a todas las personas que confiaron en mí.

---

## **Agradecimientos**

---

Al Consejo Nacional de Ciencia y Tecnología (CONACYT)

A la maestría en Matemáticas Aplicadas de la universidad y en especial a mis asesores de tesis por el apoyo otorgado para la elaboración de este trabajo

A mis padres y hermanos por inspirarme y apoyarme en cada proyecto que elijo

Y en general, a todos los que de alguna forma me apoyaron para hacer esto posible.

---

## Resumen

---

Los datos jerárquicos son un tipo de datos con una estructura jerárquica o anidada. Los modelos de regresión ordinaria de un nivel suponen independencia de las observaciones que componen un grupo, sin embargo casi siempre éste no es el caso. Es decir, se espera que las observaciones de un mismo grupo sean similares. Usar un modelo de regresión ordinaria (de un nivel) en datos jerárquicos puede conducir a inferencias estadísticas erróneas causadas por un incremento en la tasa de error tipo 1. El incremento de la tasa de error tipo 1 surge del hecho de no considerar las unidades de nivel más alto y solo considerar las mediciones del nivel más bajo donde es medida la variable de respuesta (se produce una subestimación de los errores estándar que hace que los parámetros sean significativos cuando en realidad no lo son).

Por el contrario, los modelos multinivel, modelos jerárquicos o modelos de efectos mixtos son modelos estadísticos para datos con estructura jerárquica o anidada debido a que suponen dependencia de las observaciones que componen un grupo. De esta manera, los modelos multinivel están compuestos por parámetros de efectos fijos y parámetros de efectos aleatorios. La proporción de varianza de la variable de respuesta en cada nivel de los datos es determinada por el coeficiente de correlación intraclase.

La estimación Bayesiana es una buena opción para la estimación de modelos complejos. En investigación dental, epidemiología y ciencias sociales los datos pueden tener una estructura jerárquica. En este trabajo se adapta la metodología bottom-up de un modelo multinivel de dos niveles a un modelo lineal generalizado multinivel de tres niveles y mediante estimación Bayesiana se determinan los factores clínicos que explican la profundidad de sondaje en pacientes adultos del estado de Guerrero, México que acudieron

al servicio de clínica odontológica externa de la facultad de odontología de la Universidad Autónoma de Guerrero en busca de tratamiento, durante el período comprendido entre agosto de 2015 y febrero de 2016.

Todas las distribuciones a priori de los parámetros del modelo fueron débilmente informativas, la convergencia de las MCMC a la distribución posterior se determinó por medio de los valores Rhat y diagnósticos gráficos. En la metodología bottom-up, para la comparación de modelos se usó el criterio Bayesiano de validación cruzada de dejar uno fuera. Para evaluar la adecuación del modelo se usó la densidad predictiva posterior. Adicionalmente, se realizó una selección de variables de nivel tres hacia delante para determinar las variables de nivel tres que estarían en el modelo.

Finalmente, sangrado (variable independiente de nivel uno) y cálculo (variable independiente de nivel tres) fueron factores que explican la profundidad de sondaje; mientras que, movilidad dental y fumar no fueron significativos.

---

## Abstract

---

Hierarchical data is a data type with a hierarchical or nested structure. One level ordinary regression models assume independence of the observations that make up a group, however this is almost always not the case. That is, it is expected that the observations of the same group are similar. Using an ordinary (one-level) regression model on hierarchical data can lead to abnormal statistical inferences caused by an increase in the type 1 error rate. The increase in the type 1 error rate arises from not considering the units of highest level and only consider the measurements of the lowest level where the response variable is measured (there is an underestimation of the standard errors that makes the parameters significant when in reality they are not).

On the contrary, multilevel models, hierarchical models or mixed effects models are statistical models for data with a hierarchical or nested structure because they assume dependence on the observations that make up a group. In this way, multilevel models are composed of parameters of fixed effects and parameters of random effects. The proportion of variance of the response variable at each data level is determined by the intraclass consequence coefficient.

Bayesian estimation is a good option for estimating complex models. In dental research, epidemiology and social sciences data may have a hierarchical structure. In this paper, the bottom-up methodology of a two-level multilevel model is adapted to a three-level multilevel generalized linear model and, using Bayesian estimation, the clinical factors that explain the depth of probing in adult patients from the state of Guerrero, Mexico are determined. who went to the external dental clinic service of the dental school of the Autonomous University of Guerrero in search of treatment, during the period between

August 2015 and February 2016.

All the prior distributions of the model parameters were weakly informative, the convergence of the MCMC to the posterior distribution was prolonged by means of the Rhat values and graphical diagnostics. In the bottom-up methodology, the Bayesian leave-one-out criterion of cross-validation was used to compare models. Posterior predictive density was used to assess model adequacy. Additionally, a forward selection of level three variables was performed to determine the level three variables that would be in the model.

Finally, bleeding (level one independent variable) and calculus (level three independent variable) were factors that explain probing depth; while dental mobility and smoking were not significant.



---

# Índice general

---

<b>1. Planteamiento del problema</b>	<b>1</b>
1.1. Enfermedad periodontal . . . . .	1
1.2. Descripción del problema . . . . .	5
1.3. Objetivos . . . . .	8
1.4. Antecedentes . . . . .	9
<b>2. Material y métodos</b>	<b>18</b>
2.1. Modelos multinivel . . . . .	18
2.1.1. Modelos multinivel de dos niveles . . . . .	19
2.1.2. Modelos multinivel de tres niveles . . . . .	21
2.1.3. Supuestos de los modelos multinivel . . . . .	24
2.1.4. Estimación de modelos multinivel . . . . .	25
2.1.5. Modelos lineales generalizados multinivel (MLGM) . . . . .	26
2.2. Inferencia Bayesiana . . . . .	28
2.2.1. Estimación Bayesiana . . . . .	29
2.2.2. Cadenas de Markov Monte Carlo . . . . .	30
2.2.3. Evaluación y comparación de modelos . . . . .	35
2.3. Metodología de los modelos multinivel . . . . .	38
2.3.1. Estrategia de construcción de modelos multinivel . . . . .	38
<b>3. Resultados</b>	<b>41</b>
3.1. Estimación Bayesiana . . . . .	43
3.2. Discusión . . . . .	48

<b>4. Conclusiones</b>	<b>52</b>
<b>Bibliografía</b>	<b>54</b>
<b>Anexos</b>	<b>57</b>
.1. Análisis univariado . . . . .	57
.2. Glosario de periodoncia . . . . .	63

---

## Índice de figuras

---

1.1. Diagnósticos periodontales en la dentición permanente. . . . .	2
3.1. Estructura de los datos de profundidad de sondaje para 1 de los 32 dientes de un paciente. . . . .	42
3.2. Supuestos del modelo de solo intercepto . . . . .	43
3.3. Distribución posterior de los parámetros y trazas de las MCMC del Modelo 3.5. . . . .	49
3.4. Densidad predictiva posterior del Modelo 3.5 con pendiente aleatoria para sangrado entre dientes y paciente. . . . .	50
1. Profundidad de sondaje . . . . .	57
2. Sangrado durante el sondaje . . . . .	58
3. Furcación dental . . . . .	58
4. Movilidad dental . . . . .	59
5. Edad . . . . .	59
6. Placa . . . . .	60
7. Cálculo dental . . . . .	60
8. Resistencia a la insulina . . . . .	61
9. Fumar . . . . .	61
10. Restos radiculares . . . . .	62
11. Restauraciones desajustadas . . . . .	62

---

## Índice de Cuadros

---

3.1. Datos de profundidad de sondaje . . . . .	42
3.2. Estimaciones Bayesianas de los modelos 3.1 - 3.6 . . . . .	46
3.3. Selección de variable hacia delante para las variables de nivel tres*. . . . .	51

## Planteamiento del problema

---

En este primer capítulo se describe la intención de este trabajo. Es decir, se describe el problema a estudiar, los objetivos a alcanzar y los antecedentes del tema de estudio.

### 1.1. Enfermedad periodontal

En esta sección se explican brevemente algunas ideas básicas que sirven para comprender la enfermedad periodontal, su clasificación, los factores asociados, el diagnóstico y el tratamiento a seguir.

De acuerdo a la Organización Mundial de la Salud (OMS), se estima que las periodontopatías graves afectan a cerca del 14 % de la población adulta, es decir, un equivalente a más de mil millones de casos en el mundo. La *Periodontitis* se define como una enfermedad inflamatoria de los tejidos de soporte de los dientes causada por microorganismos específicos o grupos de microorganismos específicos que provocan la destrucción progresiva del ligamento periodontal y el hueso alveolar con formación de profundidad de sondaje aumentada, recesión o ambas [1].

Signos clínicos de inflamación como contorno, color, consistencia y sangrado al sondeo no siempre son indicadores positivos de una pérdida de inserción. Sin embargo, la presencia de sangrado al sondaje durante visitas secuenciales ha mostrado ser un indi-

cador confiable de la presencia de inflamación y la posible pérdida de inserción posterior en el sitio de sangrado [1].

En 1999 la Asociación Americana de Periodontología (AAP, por sus siglas en inglés) clasificó las diferentes formas de periodontitis en tres manifestaciones clínicas generales: periodontitis crónica, periodontitis agresiva y periodontitis como manifestación de enfermedades sistémicas.

Las características que son comunes a pacientes con periodontitis crónica son: prevalencia en adultos pero puede observarse en niños, es asociada con la acumulación de placa y cálculo, tiene una tasa de progresión de la enfermedad de lenta a moderada, los aumentos en la tasa de progresión de la enfermedad pueden deberse a factores locales, sistémicos o ambientales. La periodontitis crónica se puede clasificar en base a la cantidad de pérdida de inserción clínica como leve: 1 a 2 mm, moderada: 3 a 4 mm o grave  $\geq 5$  mm. La Figura 1.1 muestra los diagnósticos periodontales que se distinguen en la dentición permanente [2].

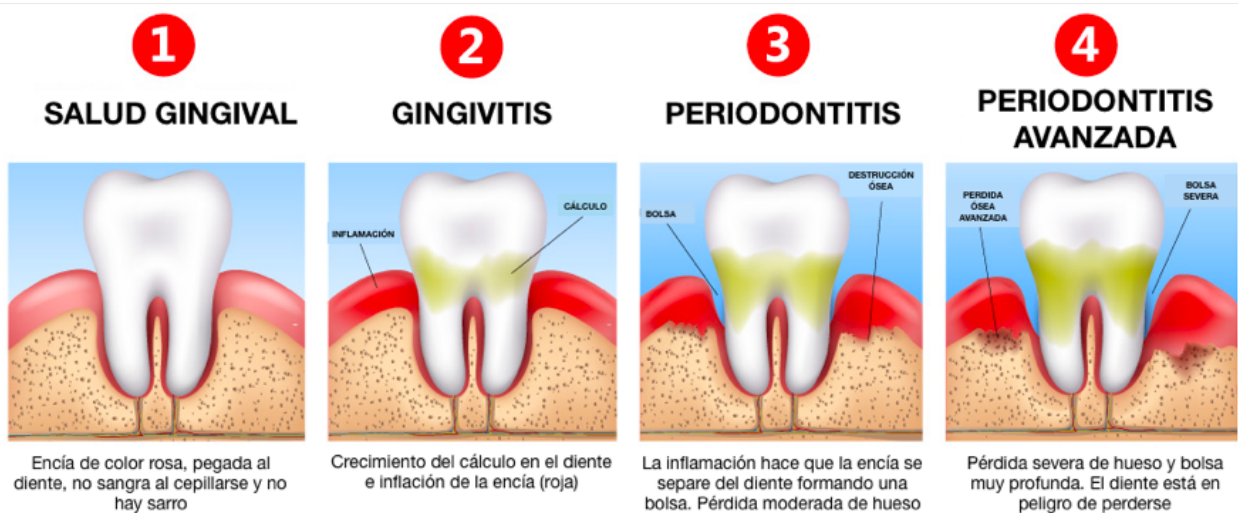


Figura 1.1: Diagnósticos periodontales en la dentición permanente.

Las características que son comunes a pacientes con periodontitis agresiva son: rápida tasa de progresión de la enfermedad periodontal, ausencia de grandes acumulaciones de placa y cálculo y antecedentes familiares de enfermedad periodontal agresiva que sugieren un rasgo genético.

La periodontitis como manifestación de enfermedades sistémicas es el diagnóstico utilizado cuando la enfermedad sistémica es el principal factor predisponente y los factores

locales como placa y cálculo no son evidentes. Estas condiciones sistémicas, principalmente trastornos hematológicos y genéticos tienen el efecto de producir alteraciones en los mecanismos de defensa del huésped.

Una de las características más importantes de enfermedad periodontal es la presencia de bolsas periodontales. Hay dos tipos de bolsas: *bolsa gingival* y *bolsa periodontal*. Una bolsa gingival se forma por el engrandecimiento gingival sin destrucción de los tejidos periodontales subyacentes. Por el contrario, la bolsa periodontal si produce destrucción de los tejidos periodontales subyacentes. Asimismo, una bolsa periodontal se clasifica como supraósea o intraósea si la parte inferior de la bolsa es coronal o apical al hueso alveolar subyacente.

La característica clínica que diferencia la periodontitis de la gingivitis es la presencia de pérdida de inserción clínicamente detectable, donde, el nivel de inserción es la distancia entre la base de la bolsa y un punto fijo en la corona como la unión amelocementaria.

Las bolsas periodontales pueden comprender una cara dental, dos o más caras dentales o una superficie diferente al sitio marginal de origen, así, son clasificadas como simples, compuestas y complejas, respectivamente. Las bolsas periodontales que requieren atención especializada son aquellas con una profundidad mayor a 3 milímetros [1]. Este tipo de bolsas requiere sesiones de curetajes para remover la placa bacteriana subgingival. Cuando la profundidad de las bolsas es menor o igual a 3 milímetros el tratamiento adecuado es la limpieza dental con ultrasonidos. En ambos casos, el raspado y alisado radicular es la técnica a seguir para tratar las bolsas periodontales.

En la primera visita el clínico obtiene la mayoría del historial médico y dental de forma que pueda establecer un diagnostico global del paciente. En la segunda visita el clínico realiza una examinación oral al paciente. La examinación del periodonto incluye la detección y evaluación de placa y cálculo supragingival y de bolsas periodontales. Es importante distinguir entre dos profundidades de una bolsa periodontal: la *profundidad biológica o histológica* y la *profundidad clínica o de sondaje*.

La profundidad biológica es la profundidad entre el margen gingival y la base de la bolsa mientras que, la profundidad de sondaje es la distancia que penetra una sonda en la bolsa.

A pesar de que el huésped tiene mecanismos protectores, la periodontitis es una

enfermedad multifactorial asociada con factores de riesgo sistémicos e indicadores de riesgo. Los factores de riesgo sistémicos (factores para los que hay evidencia suficiente) asociados con la periodontitis incluyen: fumar, diabetes mellitus, bacterias específicas (*P. Gingivalis*, *B. Forsythus*, *P. Intermedia*), género masculino y enfermedades periodontales preexistentes. Los indicadores de riesgo (factores relacionados con el desarrollo temprano de periodontitis) incluyen la osteoporosis, estrés, angustia y la baja ingesta de calcio [3].

El componente placa es el responsable del 100 % de los casos de enfermedad periodontal destructiva, fumar del 80 % y diabetes del 10 %. Por otra parte, fumar tabaco induce asociaciones causales espurias en otros factores de riesgo periodontal. Por ejemplo, fumar es un factor de riesgo para la periodontitis y para la diabetes, este hecho provoca que las asociaciones entre periodontitis y diabetes sean sensibles a sesgos [1].

Los primeros síntomas de enfermedad periodontal son: mal aliento, inflamación, enrojecimiento, dolor y sangrado de las encías. Otros síntomas de aparición posterior cuando la periodontitis ha llegado a un estado agudo son la aparición de espacios entre dientes, movilidad dental, encías retraídas que exponen la raíz del diente, dolor al masticar y exudado de pus. Estos síntomas varían en cada persona y pueden comprender alguno o todos los síntomas ya mencionados. En general, existen tres métodos para evaluar el daño de los tejidos periodontales: inspección visual, medición del nivel de inserción clínico (sondaje periodontal) y radiografías de pérdida ósea. Aquí, se profundiza en el sondaje periodontal.

La técnica de sondaje periodontal consiste en insertar la sonda paralela al eje vertical del diente y “caminar” alrededor de cada superficie de cada diente. El periodontograma es la representación gráfica del estado de las encías durante el sondaje periodontal, es indispensable para poder hacer una valoración sobre el grado de afectación de las encías y la magnitud de pérdida ósea. Además, permite establecer una comparación entre el primer sondaje periodontal al momento del diagnóstico y el sondaje periodontal después de concluir el tratamiento.

En la actualidad, existen diversos softwares de gráficos periodontales que brindan al paciente una experiencia más rápida y menos dolorosa, en tanto que ayuda a mejorar la productividad del odontólogo. *Florida Probe* es un sistema que sirve para el registro



de los valores de cada paciente durante el sondaje periodontal. Este software permite un examen personalizado acerca de la información incluida en el periodontograma. La precisión de la sonda Florida es de 0.2 mm.

El funcionamiento de este sistema en un primer paso consiste en configurar el periodontograma de acuerdo a la boca del paciente, calibrar la sonda periodontal digital e inalámbrica tipo Florida introduciéndola en una base plana y estéril y a continuación iniciar a sondear el primer cuadrante hasta sondear toda la boca, seguidamente, oprimir el botón de pie y de esta forma los datos numéricos quedan registrados en el periodontograma. Al finalizar el sondaje periodontal el sistema muestra un resumen del estado de salud bucal del paciente así como un diagnóstico.

## **1.2. Descripción del problema**

Kim y colaboradores [4] realizaron una revisión de 307 artículos publicados entre 1995 y 2009, seleccionados aleatoriamente de 10 revistas establecidas en ciencias dentales con el objetivo de evaluar la proporción de errores estadísticos en investigación dental así como, la principal fuente de errores estadísticos cometidos. Los resultados mostraron que la tasa de errores fue del 51.5%. En esa muestra se encontró que los métodos estadísticos más utilizados son: estadística descriptiva, prueba t (una muestra, dos muestras o pareada), diferentes tipos de ANOVA (uno, dos y tres factores y procedimientos de comparación múltiple) y pruebas no paramétricas. Los errores estadísticos más comunes comprendieron el no emplear una muestra aleatoria, el uso de una tabla de contingencia cuando el tamaño de muestra no es suficientemente grande y errores en el análisis de datos ordinales, como: calcular la media, desviación estándar y el coeficiente de correlación de Pearson, usar la prueba t, aplicar un ANOVA y usar indebidamente las escalas de medición ordinales o nominales.

Los datos jerárquicos se definen como un tipo de datos en el que las unidades del nivel más bajo son anidadas o agrupadas en las unidades de nivel más alto. La sociedad puede ser descrita por medio de estructuras jerárquicas pues los individuos componen diferentes tipos de grupos con diversos contextos sociales cada uno [5]. Hannigan y Lynch [6] describen que un error de los métodos estadísticos más utilizados en investigación

oral y dental es ignorar la agrupación de observaciones de múltiples dientes/superficies dentales/restauraciones/implantes dentro del mismo sujeto. Este hecho provoca que las observaciones de un mismo sujeto sean tratadas como independientes a pesar de no ser el caso. Estos autores explican que si se tienen 5 dientes de 20 sujetos y un total de 80 observaciones cuando se ajusta un modelo sin agrupación de los sujetos el número de observaciones es 80 en lugar de 20 lo que resulta en errores estándar más pequeños y en una mayor probabilidad de rechazar una hipótesis nula verdadera (inflación del error tipo 1).

Otro posible error de no considerar la dependencia en en datos multinivel es que, de existir una correlación significativa entre las observaciones esta sea ignorada por el investigador y por consecuencia encuentre resultados estadísticamente significativos cuando en realidad no lo son. Por otro lado, considerar un modelo de las observaciones a nivel grupo reduce el tamaño de muestra y no toma en cuenta las variables de nivel individual.

Modelos estadísticos que toman en cuenta la agrupación de los datos incluyen a: *modelos de regresión de efectos mixtos, ecuaciones de estimación generalizada (GEE, por sus siglas en inglés)* y modelos de fragilidad [6].

La inferencia Bayesiana es una alternativa atractiva cuando: 1) tenemos información sobre los parámetros en el modelo, 2) el método de estimación frecuentista no converge, 3) el tamaño de la muestra es pequeño en el nivel más alto de los datos, o 4) se van a estimar funciones no lineales de los parámetros. A diferencia de la integración numérica vía la cuadratura Gaussiana, las estimaciones Bayesianas mediante integración Monte Carlo son computacionalmente intensivas y más que una convergencia numérica tienen una convergencia estocástica.

Es útil usar métodos de cadenas de Markov Monte Carlo (MCMC, por sus siglas en inglés) en lugar de mínimos cuadrados generalizados iterativos (IRLS, por sus siglas en inglés) para modelos lineales generalizados multinivel (MLGM) si: 1) Las matrices de efectos aleatorios son extremadamente grandes (salvo matrices diagonales o triangulares que pueden ser particionadas) como en el caso de problemas espaciales y problemas que involucran efectos aleatorios cruzados (efectos aleatorios entre niveles). 2) Si la estimación de los intervalos creíbles de los efectos aleatorios o hiperparámetros es de interés, debido a que en IRLS estas estimaciones no son calculadas [7, pág. 299].

Entre agosto de 2015 y febrero de 2016, Romero-Castro y colaboradores [8] llevaron a cabo un estudio con la finalidad de determinar la distribución de enfermedad periodontal en pacientes de 18 a 75 años residentes del estado de Guerrero y que acudieron al servicio de clínica odontológica externa de la facultad de odontología de la Universidad Autónoma de Guerrero (UAGro) en busca de tratamiento. Esta población de estudio excluyó a mujeres embarazadas y lactantes, pacientes con enfermedades sistémicas, personas que no concluyeron los estudios periodontales o de laboratorio y a quienes les faltaron datos de la encuesta o no regresaron para una revaloración. Otro objetivo de esta investigación fue evaluar el efecto conjunto de una intervención educativa de atención primaria junto a un tratamiento conservador en la mejora de parámetros clínicos periodontales.

Las pruebas diagnósticas incluyeron un examen bucal con espejo, sondaje periodontal e índice de placa y cálculo. Los autores excluyeron del estudio a los pacientes con periodontitis agresiva y clasificaron a 161 pacientes en 5 grupos de acuerdo a sus características periodontales como: sanos (18%), con gingivitis (21.1%), con periodontitis crónica leve (13%), moderada (19.9%) y severa (28.8%). Los pacientes se dividieron en dos grupos: sanos y con enfermedad periodontal (gingivitis o cualquier grado de periodontitis).

Para probar si había diferencia entre las medianas de variables cuantitativas en pacientes con o sin enfermedad periodontal se usó la prueba Mann Whitney (las variables no tenían distribución normal) y para probar si había diferencia entre las prevalencias de variables cualitativas se usó la prueba chi-cuadrado. Se concluyó que pertenecer a un grupo indígena, vivir en una zona rural o urbana, fumar, estrés, respiración bucal, hacinamiento y mordida abierta no tuvieron diferencias significativas diferente entre los grupos. Las variables que si tuvieron diferencias significativas entre grupos fueron: edad, niveles de glucosa, índice de placa, índice de cálculo, educación, raíces retenidas y restauraciones dentales.

Los autores usaron dos modelos de regresión logística multinomial (donde la variable dependiente, profundidad de sondaje, tuvo como categorías: saludable, gingivitis, periodontitis leve, periodontitis moderada y periodontitis severa) debido a que las variables independientes placa y cálculo se consideraron colineales. Cuando la variable placa fue categórica (punto de corte  $\leq 10\%$  para pacientes con poca o nada de placa y  $> 10\%$  pa-

ra niveles más altos de placa), el modelo de regresión logística multinomial ajustado por edad y educación mostró que los pacientes que tenían un aumento en el nivel de placa tenían 14.4% veces más riesgo de tener gingivitis y 68.5% veces más riesgo de tener periodontitis severa; en comparación con pacientes sanos. Cuando la variable placa se trató en su escala original (continua), el riesgo de tener gingivitis y periodontitis severa aumentó 1.08 y 1.09 veces, respectivamente, por cada punto porcentual de aumento de placa.

En el otro modelo de regresión logística multinomial (ajustado por edad y educación), el cálculo (como variable cuantitativa) mostró que por cada punto porcentual que aumenta la variable cálculo los riesgos son 1.6 y 2.1 veces mayor en pacientes con gingivitis y pacientes con periodontitis severa comparados con pacientes sanos.

Por otra parte, la prueba de Wilcoxon para datos pareados y el análisis de datos de panel estimaron un porcentaje de mejoría en los parámetros clínicos (profundidad de sondaje, sangrado al sondaje y movilidad dental) a los 6 meses de estar en tratamiento y recibir educación para la salud.

Si bien estos datos provenientes de un proceso de sondaje periodontal han sido analizados por medio de un modelo de regresión logística multinomial su estructura jerárquica hace que resulte adecuado analizar estos datos mediante un modelo lineal multinivel con un método de estimación Bayesiano.

### **1.3. Objetivos**

El objetivo general de este trabajo fue usar inferencia Bayesiana en un modelo multinivel para identificar los factores clínicos que explican la profundidad de sondaje en pacientes adultos del estado de Guerrero, México que acudieron al servicio de clínica odontológica externa de la facultad de odontología de la Universidad Autónoma de Guerrero (UAGro) en el período agosto de 2015 a febrero de 2016.

Para lograr este objetivo general, se plantearon los siguientes objetivos específicos:

1. Recolectar los datos de los expedientes electrónicos.
2. Adaptar la metodología frecuentista bottom-up al caso de tres niveles de jerarquía.

3. Identificar usando el enfoque Bayesiano el mejor modelo para el problema de estudio.

## 1.4. Antecedentes

En esta sección se detallan en orden descendente de publicación algunos de los trabajos precedentes en el uso de modelos multinivel aplicados al estudio de los factores y evolución de la enfermedad periodontal.

Song y colaboradores [9] publicaron un análisis multinivel acerca de los factores de riesgo de periodontitis crónica en la respuesta de curación. Este análisis comprendió una muestra de 54 pacientes que fueron remitidos al Departamento de Periodoncia de la Escuela de Estomatología de la Universidad Médica de China durante los años 2012 a 2014.

Los pacientes de 35 a 60 años recibieron tratamiento periodontal no quirúrgico y posteriormente fueron evaluados a los 3, 6 y 12 meses. Cuatro modelos multinivel fueron ajustados, dos modelos donde la variable dependiente fue la reducción de profundidad de sondaje (PD, por sus siglas en inglés) y dos modelos donde la variable dependiente fue la reducción de pérdida de inserción clínica (CAL, por sus siglas en inglés). Los autores analizaron 3 niveles; nivel 1: sitios, nivel 2: dientes y nivel 3: pacientes.

El 25.9 % de los participantes fueron fumadores, 57.4 % dijeron tener una experiencia de estrés y 79 % no tuvieron algún grado de movilidad. Al principio del estudio el 79.4 % de los pacientes presentaron placa a nivel sitio y el 66.1 % presentaron sangrado. Los modelos nulos de ambas variables dependientes demostraron que los tres niveles del modelo fueron significativos ( $p < 0.0001$ ). La variabilidad explicada de la reducción de PD y CAL en cada uno de estos modelos fue: 77-78 % y 70-80 % a nivel sitio, 11-12 % y 10-11 % a nivel diente y 10-11 % y 13-14 % a nivel paciente.

El segundo modelo donde la variable dependiente fue la reducción de PD mostró una disminución significativa ( $p < 0.05$ ) a los 3, 6 y 12 meses del grado de movilidad, presencia

de sangrado y porcentaje de sitios con  $PD \geq 5\text{mm}$ . Además, haber tenido experiencias de estrés tuvo un efecto negativo significativo ( $p < 0.05$ ) en la reducción de PD en comparación con pacientes que afirmaron no haber tenido alguna experiencia de estrés.

El segundo modelo donde la variable dependiente fue la reducción de CAL indicó una relación negativa significativa ( $p < 0.05$ ) entre género, pacientes fumadores y la reducción de CAL. Los dientes anteriores mostraron una reducción de CAL significativa ( $p < 0.05$ ) a los 3, 6 y 12 meses en comparación con los dientes molares. Además, el grado de movilidad mostró una mayor reducción significativa de CAL a los 3 y 12 meses. La presencia de sangrado al inicio del estudio y las superficies de mesioclusión/distoclusión mostraron una mayor reducción de CAL.

Mdala y colaboradores [10] examinaron a 187 pacientes de 20 a 79 años de edad que fueron monitoreados al inicio del estudio, 3, 6, 12, 18 y 24 meses para comparar la efectividad clínica sobre el tiempo (2 años) de 8 tipos de terapias periodontales.

El objetivo de estos autores fue comparar si los pacientes que solo recibieron una cirugía periodontal (SURG, por sus siglas en inglés) mantienen su estado clínico mejorado durante un periodo más corto de tiempo que los pacientes que fueron tratados con SURG y antibióticos sistémicos o locales. Otro objetivo de este estudio fue determinar si la reducción de PD o si la ganancia de CAL fue mayor a 1.5 mm en sitios o pacientes.

Dos modelos multinivel fueron planteados para evaluar los cambios en PD y en CAL (un modelo para cada variable) y para la efectividad clínica de la terapia se ajustaron dos modelos más. Un modelo donde la variable dependiente fue la reducción de  $PD > 1.5$  mm y un segundo modelo donde la variable dependiente fue la ganancia de  $CAL > 1.5$  mm. Ambas variables de respuesta binaria fueron modeladas por una regresión logística multinivel.

La estructura jerárquica de los datos fue conformada por 4 niveles, los pacientes conformaron las unidades de nivel más alto (nivel 4), 28 dientes excluyendo los 3 molares en

cada paciente fueron el nivel 3, los cuatro sitios de cada diente (mesiobucal, distobucal, distolingual y mesiolingual) comprendieron el nivel 2 y el nivel más bajo de los datos fue la variable tiempo.

Las estimaciones del modelo de reducción de PD mostraron que SURG + amoxicilina (AMOX) + metronidazol (MET) + tetraciclina (TET) tuvo estimaciones significativas más grandes en términos de valor absoluto después de los 6 meses, sugiriendo que esta terapia produce una mayor reducción de PD a través del periodo de estudio en comparación con el raspado y alisado radicular (SRP, por sus siglas en inglés). Además, el efecto fijo de edad y los efectos fijos de las variables independientes a nivel sitio (acumulación de placa, sangrado al sondaje (BOP, por sus siglas en inglés), enrojecimiento gingival y supuración) fueron significativos en la PD. Ex fumadores y actualmente fumadores incrementaron la PD significativamente por 0.19 y 0.23 mm, respectivamente.

El modelo de la reducción en el tiempo de  $PD \geq 1.5\text{mm}$  demostró que SURG combinado con AMOX, MET, y TET fue significativamente más probable ( $p < 0.05$ ) de reducir la PD más de 1.5 mm que las demás terapias. El modelo de los cambios medios de CAL estimó que los sitios que fueron tratados quirúrgicamente tuvieron una mayor pérdida de CAL en comparación con las condiciones al inicio del estudio y SURG combinado con AMOX, MET y TET tuvieron un aumento significativo ( $P < 0.05$ ) de CAL.

Jordao y colaboradores [11] analizaron datos de una encuesta de salud bucal de estudiantes de 12 años de la ciudad urbana de Goiana, con el objetivo de describir su estado de salud periodontal y su asociación con variables individuales y contextuales.

Los autores emplearon un modelo de regresión multinivel Poisson no ajustado con un estimador robusto de la varianza para estimar la razón de prevalencia como medida de efecto. El análisis fue no ajustado debido a la colinealidad entre las variables independientes y los intervalos de confianza del 95 % fueron obtenidos por la prueba de Wald.

La variable de interés fue la prevalencia de la condición periodontal, caracterizada

por la presencia/ausencia de cálculo y/o sangrado. Los niveles muestrales en el modelo fueron: estudiantes (efectos individuales), escuelas y distritos (efectos contextuales) donde, los siete distritos de salud fueron agrupados de acuerdo a sus características socioeconómicas en tres grupos: grupo I; con los mejores indicadores, grupo II; con indicadores intermedios y grupo III; con los peores indicadores.

Para probar la dependencia entre las variables se usó la prueba de Rao Scott (equivalente a la prueba Chi-cuadrado para muestras complejas). Los resultados correspondieron a 39 escuelas (24 públicas y 15 privadas) y 2075 estudiantes donde, 50.9% de los participantes eran hombres, 54.5% eran morenos y 71.2% eran estudiantes de escuelas públicas.

En el estudio las variables asociadas significativamente con la condición periodontal fueron: color de piel o raza y tipo de escuela. Los estudiantes del grupo II tuvieron 36.9% más probabilidad de tener una condición periodontal adversa comparados con los estudiantes de los distritos del grupo I. Por otra parte, los estudiantes de escuelas públicas presentaron una razón de momios de cálculo y sangrado 1.54 (1.14 - 2.08) veces mayor que los estudiantes de escuelas privadas ( $P < 0.004$ ) y a nivel individual, los estudiantes morenos mostraron una condición adversa de periodontitis 1.68 veces mayor que los estudiantes blancos ( $P < 0.05$ ).

Müller [12] expuso un ejemplo de análisis multinivel en el que analizó los datos de un grupo de 40 adultos jóvenes periodontalmente sanos (21 mujeres) con buena higiene bucal. El primer modelo propuesto fue una regresión ordinaria para investigar si el grosor gingival está relacionado con el ancho gingival del diente. Este modelo supuso independencia y los resultados mostraron que el ancho gingival tiene un efecto negativo significativo en el grosor gingival.

El siguiente modelo propuesto fue un modelo de dos niveles con un intercepto aleatorio para cada sujeto. Las estimaciones no fueron muy diferentes del modelo de regresión ordinaria y el efecto del ancho gingival (-0.036) con un error estándar de 0.011 fue signi-



ficativo para predecir el grosor gingival del diente. La prueba de razón de verosimilitud de este modelo y el modelo de regresión ordinaria (con un parámetro adicional, y por tanto con un grado de libertad) fue igual a 41.52 ( $p < 0.001$ ) y por tanto indicó un mejor ajuste del modelo con intercepto aleatorio.

Un tercer modelo de coeficiente aleatorio con diferentes predicciones de pendientes para los sujetos fue definido. En este modelo los interceptos para sujetos diferentes variaron significativamente ( $p = 0.004$ ). La media de los coeficientes del ancho gingival fue  $-0.028$  con un error estándar grande igual a  $0.019$  no fue significativa. La prueba de razón de verosimilitud para este modelo y el modelo de intercepto aleatorio (con dos grados de libertad) fue igual a 40.18 ( $p < 0.001$ ). Por tanto, el modelo con pendiente aleatoria describió mejor la situación.

El autor considerará una extensión del modelo de pendientes aleatorias que incluyó la PD como segunda variable independiente a nivel diente y una variable a nivel sujeto (género). Este modelo reveló que la PD periodontal fue el principal factor asociado con el grosor gingival con un incremento de  $0.03$  mm por cada incremento de un milímetro. El género y la varianza a nivel sujeto también resultaron significativos. La prueba de razón de verosimilitud (con cinco grados de libertad) para el modelo extendido de pendiente aleatoria y el modelo de pendiente aleatoria reveló que este último modelo describió mejor los datos ( $p < 0.001$ ).

Tomasi y colaboradores [13] investigaron los factores que pueden afectar la respuesta clínica a corto plazo del tratamiento periodontal no quirúrgico. Un total de 41 pacientes con periodontitis crónica moderadamente avanzada fueron asignados aleatoriamente a un tratamiento de una sesión de una hora de aseo quirúrgico de boca completa mediante un instrumento ultrasónico (FmUD, por sus siglas en inglés) o bien con instrumentos manuales con cuatro sesiones de raspado y alisado radicular (Q-SRP, por sus siglas en inglés).

En este estudio cuatro sitios fueron examinados por cada diente (mesial, bucal, dis-

tal y lingual). Los sitios incluidos en este estudio tuvieron una  $PD \geq 5$  mm. Los niveles jerárquicos para los modelos multinivel fueron paciente, diente y sitio del diente. Los autores construyeron un modelo de regresión logística multinivel para estimar los factores que afectan la variable dependiente (es decir, bolsas cerradas a los 3 meses). La estimación de parámetros fue por cuasi verosimilitud penalizada de segundo orden (en el paquete estadístico MLwiN 2.02) y la significancia de cada factor se determinó por medio de la prueba de Wald. Un segundo modelo (de regresión ordinaria) fue analizado para la PD a los 3 meses. Los coeficientes de regresión fueron estimados por IRLS.

El modelo nulo de regresión logística mostró que la probabilidad de que una bolsa haya cerrado al final de 3 meses después del desbridamiento inicial fue 0.63. El coeficiente de correlación intraclase mostró que el 17% de la variación en el cierre de las bolsas periodontales se debió a la variación entre pacientes, el 0% a la variación entre dientes y el 83% se debió a la variación entre los sitios dentales. En los modelos siguientes el nivel diente fue eliminado del análisis.

Luego de ajustar un modelo para cada una de las variables el modelo final de variable dependiente bolsas cerradas a los 3 meses incluyó los factores fumar, placa a nivel sitio y dientes con múltiples raíces. Este modelo estimó que el logaritmo de momios de la variable dependiente descendió 1.12, 0.54 y 0.85 en cada factor, respectivamente. El modelo explicó el 44% de la variabilidad total.

El primer modelo de variable dependiente PD a los 3 meses sin variables independientes estimó que el nivel sitio contenía el mayor porcentaje (81%) de varianza no explicada seguido del 14% de la varianza no explicada a nivel paciente y el 5% a nivel diente. Nuevamente, luego de ajustar un modelo para cada una de las variables el modelo final mostró que el tratamiento no fue significativo (no se encontraron diferencias significativas entre ambos tratamientos) ( $P < 0.05$ ) y fumar, placa a nivel sitio y dientes multiradiculares fueron factores asociados negativamente con la probabilidad de que una bolsa periodontal haya cerrado después de 3 meses.

D'Aiuto y colaboradores [14] realizaron un ensayo clínico longitudinal prospectivo con el objetivo de evaluar mediante un análisis multinivel la contribución relativa de los factores asociados con el paciente, el diente y el sitio para determinar los resultados clínicos del desbridamiento subgingival.

Los 94 pacientes participantes en el estudio presentaron periodontitis generalizada severa, es decir, su profundidad de sondaje fue  $\geq 5$ mm con sangrado durante el sondaje y una pérdida del hueso marinal alveolar  $> 30\%$  en al menos el 50% de la dentición. Ellos fueron reexaminados a los 2 y 6 meses después de completar el tratamiento. El análisis de varianza (ANOVA) o una prueba t pareada demostró una mejora de los parámetros puntajes de placa de boca completa, puntajes de sangrado de boca completa, PD y CAL ( $p < 0.001$ ).

El modelo nulo o modelo multinivel de componentes de varianza de variable dependiente la diferencia de PD en el principio del estudio y 6 meses después ( $\Delta PD$ ) estimó que la mayoría de la varianza fue atribuida a nivel sitio (80.4%), 11.6% a nivel diente y 8% a nivel paciente.

En el siguiente modelo fueron agregadas todas las variables a nivel paciente, diente y sitio. Los pacientes fumadores y los que presentan una copia o dos del alelo raro en el polimorfismo mostraron una respuesta menos favorable al tratamiento ( $p < 0.001$ ) en términos de reducción de PD. A nivel diente, la presencia de movilidad al inicio del estudio fue 0.07 menos favorable para reducir la PD ( $p < 0.05$ ). Los dientes incisores y caninos mostraron un decrecimiento mayor significativo en comparación de los dientes premolares o molares ( $p < 0.0001$ ). A nivel sitio, el efecto de los sitios mesial y distal fue 0.43 veces mayor que el efecto de los sitios facial y lingual ( $p < 0.0001$ ). Las demás variables y en particular, placa y presencia de sangrado durante el sondaje no fueron significativas.

Axtelius y colaboradores [15] analizaron el progreso de la enfermedad periodontal en 22 pacientes que fueron tratados por odontólogos experimentados. Para este fin, se consideraron dos variables dependientes, 1) El cambio en la profundidad de la bolsa perio-

dontal durante el sondaje determinado por la diferencia de la primera y la última medición y 2) La profundidad final de la bolsa periodontal.

Para cada variable dependiente se ajustaron cuatro modelos de tres niveles (un modelo modelo sin variables independientes y tres modelos más con las variables de cada nivel). Los niveles comprometidos en la inflamación periodontal fueron: sitio, diente e individuo. Los participantes tenían al menos dos cuadrantes con profundidad mayor a 6 mm en cada cuadrante.

Los resultados del modelo nulo con variable dependiente el cambio de profundidad de la bolsa en cada sitio estiman que el 80 % de la varianza se encuentra a nivel sitio, 7 % a nivel diente y 14 % a nivel individual. En el modelo nulo de la variable dependiente registro final de la profundidad de la bolsa en cada sitio se encontraron porcentajes de varianza iguales a: 70 % en el nivel sitio, 21 % en el nivel diente y 10 % en el nivel paciente.

El número de cigarrillos diarios consumidos, el consumo de alcohol y estado socio económico resultaron relacionados significativamente ( $P < 0.05$ ) con el cambio de profundidad de la bolsa periodontal pero no en el último registro de profundidad. El número de citas canceladas y de dentistas involucrados en el tratamiento se relacionaron significativamente ( $P < 0.05$ ) con ambas variables dependientes.

Albandar y Goldstein [16] ilustran un ejemplo de un modelo simple de 2 niveles para estudiar los predictores de la progresión periodontal evaluados en radiografías periapicales tomadas al inicio, 2 y 6 años de 142 sujetos de una planta industrial de Oslo, Noruega. Las medidas repetidas de cada sujeto correspondieron a 5,579 sitios.

Los autores propusieron dos modelos: un modelo de dos niveles que descompone la varianza en sitios y en sujetos específicos y un modelo de tres niveles que descompone la varianza en ocasiones específicas, sitios y sujetos. El modelo de dos niveles mostró que la altura inicial del hueso a nivel sitio estimada en 0.983 y los factores de retención de placa a nivel sitio estimados en 0.324 fueron factores significativos ( $p < 0.0001$ ) para

predecir la altura futura del hueso alveolar a nivel sitio. La varianza a nivel sujeto fue 0.22 y a nivel sitio fue 0.78.

En el modelo de tres niveles estos mismos factores y la edad estimada en 0.026 fueron factores significativos ( $p < 0.0001$ ) para predecir la altura del hueso a los 6 años. El 42% de la varianza total se debe a la varianza entre sujetos, 46% entre sitios y 12% entre ocasiones y cada componente de varianza fue significativamente mayor que cero.

Otros trabajos en los que se aplican modelos multinivel en investigación dental incluyen: Beek y colaboradores [17] estudiaron el cambio en las dimensiones faciales de niñas holandesas entre 7 y 14 años a través de dos modelos de dos niveles con un polinomio de grado cuatro y un polinomio de grado dos, respectivamente. Kopperud y colaboradores [18] analizaron dos modelos de regresión Cox multinivel. Un modelo para comparar la supervivencia de restauraciones compuestas por amalgama y las restauraciones compuestas por resina, y un modelo para analizar los factores relacionados a la supervivencia de las restauraciones compuestas de resina. Barrera y colaboradores [19] aplicaron un modelo de dos niveles longitudinal para evaluar cómo cambia el rendimiento masticatorio con la edad, si el rendimiento masticatorio difiere entre sexos y si los patrones de desempeño masticatorio difieren entre sujetos con varios tipos de maloclusión.

### Material y métodos

---

#### 2.1. Modelos multinivel

Los modelos multinivel hacen una partición de la varianza de la variable dependiente en los diferentes niveles de agrupación de los datos. Se distinguen dos tipos de varianza: *varianza dentro de grupos* o varianza a nivel individual (nivel uno),  $\sigma_w^2$ , y *varianza entre grupos*  $\sigma_b^2$ , que cuantifica la variación a nivel grupo. Al igual que en el modelo de VARCOMP, en los modelos multinivel la dependencia de las observaciones de un mismo grupo se mide con el *coeficiente de correlación intraclase (CCI)*. El CCI se define como un cociente de la *varianza entre grupos* y la *varianza total* (la suma de las varianzas dentro de grupos y entre grupos):

$$CCI = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_w^2} \quad (2.1)$$

El CCI varía entre 0 y 1 puesto que la varianza no puede ser negativa. Antes de usar un modelo multinivel se requiere determinar si el CCI es significativo en cada nivel de los datos. Para ello, mediante el modelo nulo (se define en la siguiente subsección) se determina si la varianza de los residuos de cada nivel es significativa, de ser así, el CCI también es significativo y esto quiere decir que a nivel individual las observaciones son dependientes y por tanto, es necesario usar un modelo multinivel en lugar de un modelo de regresión múltiple ordinaria [5, pág. 14-15].

### 2.1.1. Modelos multinivel de dos niveles

Sea  $y_{ij}$  la variable dependiente medida en el  $i$ -ésimo individuo en la  $j$ -ésima unidad del nivel 2 (por ejemplo, el  $j$ -ésimo grupo);  $i = 1, \dots, N$  donde  $N$  es el tamaño total de la muestra y  $j = 1, \dots, J$  para  $J$  unidades de nivel 2.

El modelo más simple de dos niveles es el *modelo nulo* (modelo de medias no condicional, modelo de solo intercepto o análisis de varianza de efectos aleatorios de un factor). Este modelo está definido por dos ecuaciones

$$y_{ij} = \beta_{0j} + e_{ij}$$

$$\beta_{0j} = \gamma_{00} + u_{0j}$$

$\beta_{0j}$  es la media de  $y$  en el grupo  $j$  que varía entre grupos,  $e_{ij}$  es la variación individual alrededor de esta media,  $\gamma_{00}$  es el intercepto general es decir, la gran media de  $y$ , y  $u_{0j}$  es la desviación de  $\beta_{0j}$  con respecto a  $\gamma_{00}$ .

La sustitución de  $\beta_{0j}$  en  $y_{ij}$  produce el modelo de ecuación única

$$y_{ij} = \gamma_{00} + u_{0j} + e_{ij} \quad (2.2)$$

La Ecuación 2.2 es compuesta por un efecto fijo o parte fija,  $\gamma_{00}$ , y por una parte aleatoria correspondiente a los dos efectos aleatorios  $u_{0j}$  y  $e_{ij}$ . Suponiendo que  $e_{ij} \sim N(0, \sigma^2)$ ,  $u_{0j} \sim N(0, \sigma_{u0}^2)$  y que  $e_{ij}$  y  $u_{0j}$  son independientes, la varianza de  $y_{ij}$  en la Ecuación 2.2 es

$$\begin{aligned} \text{var}(y_{ij}) &= \text{var}(\gamma_{00} + u_{0j} + e_{ij}) \\ &= \text{var}(u_{0j} + e_{ij}) \\ &= \sigma_{u0}^2 + \sigma^2 \end{aligned}$$

donde  $\sigma_{u0}^2 = \sigma_b^2$  es la varianza entre grupos y  $\sigma^2 = \sigma_w^2$  es la varianza dentro de grupos. Sustituyendo estos valores en la Ecuación 2.1 se obtiene el CCI.

Ahora, se considera un modelo de dos niveles con dos variables independientes de nivel uno,  $x_1$  con un efecto fijo  $\alpha_1$ , que no varía entre grupos, y  $x_2$  con un efecto aleatorio  $\beta_{2j}$ , que sí varía entre grupos.

Este modelo está definido por

$$y_{ij} = \beta_{0j} + \alpha_1 x_{1ij} + \beta_{2j} x_{2ij} + e_{ij}$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01} w_{1j} + u_{0j}$$

$$\beta_{2j} = \gamma_{10} + \gamma_{11} w_{1j} + u_{1j}$$

En la ecuación anterior, tanto  $\beta_{0j}$  como  $\beta_{2j}$  dependen de una variable independiente de nivel dos,  $w_1$  y  $u_{qj}$  es la desviación del efecto de la variable  $w_1$  sobre  $y_{ij}$  en el grupo  $j$  con respecto al efecto medio  $\gamma_{q0}$ ,  $q = 0, 1$ .

La sustitución de  $\beta_{0j}$  y  $\beta_{2j}$  en  $y_{ij}$  produce el modelo de ecuación única

$$y_{ij} = \gamma_{00} + \alpha_1 x_{1ij} + \gamma_{01} w_{1j} + \gamma_{10} x_{2ij} + \gamma_{11} w_{1j} x_{2ij} + (u_{0j} + u_{1j} x_{2ij} + e_{ij})$$

La generalización del modelo anterior de dos niveles al caso donde el nivel uno incluye  $P$  variables independientes  $x_p$  que tienen un efecto fijo,  $Q$  variables independientes  $x_q$  que tienen un efecto aleatorio, y  $M$  variables independientes de nivel dos  $w_m$  que también tienen un efecto fijo, es:

$$y_{ij} = \beta_{0j} + \sum_{p=1}^P \alpha_p x_{pij} + \sum_{q=1}^Q \beta_{qj} x_{qij} + e_{ij}$$

$$\beta_{0j} = \gamma_{00} + \sum_{m=1}^M \gamma_{0m} w_{mj} + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + \sum_{m=1}^M \gamma_{1m} w_{mj} + u_{1j}$$

⋮

$$\beta_{Qj} = \gamma_{Q0} + \sum_{m=1}^M \gamma_{Qm} w_{mj} + u_{Qj}$$

Cada efecto aleatorio de nivel 1,  $\beta_{qj}$ , es expresado como una función lineal de  $M$  variables independientes de nivel 2,  $w_{mj}$ .

La sustitución de los efectos aleatorios  $\beta_{qj}$  en  $y_{ij}$  produce el modelo de ecuación única



$$\begin{aligned}
y_{ij} = & \gamma_{00} + \sum_{m=1}^M \gamma_{0m} w_{mj} + \sum_{p=1}^P \alpha_p x_{pij} + \sum_{q=P+1}^{P+Q} \gamma_{q0} x_{qij} + \sum_{q=P+1}^{P+Q} \sum_{m=1}^M \gamma_{qm} w_{mj} x_{qij} \\
& + \left( u_{0j} + \sum_{q=P+1}^{P+Q} x_{qij} u_{qj} + e_{ij} \right)
\end{aligned} \tag{2.3}$$

Este último modelo está compuesto por: *efectos fijos* (los coeficientes  $\gamma$  y  $\alpha$ ) y *efectos aleatorios* (todos los términos entre paréntesis). Los modelos multinivel de dos niveles también pueden ser expresados en forma matricial mediante:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{W}\mathbf{u} + \mathbf{e} \tag{2.4}$$

donde  $\mathbf{Y}$  es el vector de mediciones de la variable dependiente,  $\mathbf{X}$  es la matriz diseño de los parámetros de efectos fijos dados por el vector  $\boldsymbol{\beta}$  (contiene a los efectos principales e interacciones),  $\mathbf{W}$  es la matriz diseño de los efectos aleatorios dados por el vector  $\mathbf{u}$  y  $\mathbf{e}$  es el vector de errores residuales de nivel 1.

### 2.1.2. Modelos multinivel de tres niveles

Sean  $i, j$  y  $k$  que indican las unidades de observación de los niveles 1, 2 y 3, respectivamente. Además, el nivel 3 tiene  $K$  unidades, el nivel 2 tiene  $J_k$  unidades de nivel 2 y la  $j$ -ésima unidad de nivel 2 en la  $k$ -ésima unidad de nivel 3 tiene  $n_{ijk}$  unidades de nivel 1. El modelo nulo es:

$$\begin{aligned}
y_{ijk} &= \beta_{0jk} + e_{ijk} \\
\beta_{0jk} &= \gamma_{00k} + u_{0jk} \\
\gamma_{00k} &= \xi_{000} + \nu_{00k}
\end{aligned}$$

En la primera ecuación  $\beta_{0jk}$  es el intercepto aleatorio de nivel 1 que varía entre grupos de nivel 2, es decir, es la media de  $y$  del grupo  $j$  y  $e_{ijk}$  es la varianza residual en el nivel 1 con respecto a  $\beta_{0jk}$ . En la segunda ecuación,  $\gamma_{00k}$  es el intercepto aleatorio de nivel 2 que varía entre las unidades de nivel 3, y  $u_{0jk}$  es la variación residual del grupo  $j$  con respecto a  $\gamma_{00k}$ . En la tercera ecuación  $\xi_{000}$ , es el intercepto general, es decir, la gran media de  $y$  y

$\nu_{0k}$  concentra la variación entre las medias de los grupos de nivel 3 (esto es, la desviación de la media del grupo  $k$  con respecto a la gran media).

Sustituyendo  $\gamma_{00k}$  en  $\beta_{0jk}$  y luego  $\beta_{0jk}$  en  $y_{ijk}$  se obtiene:

$$y_{ijk} = \xi_{000} + \nu_{00k} + u_{0jk} + e_{ijk} \quad (2.5)$$

Suponiendo que  $e_{ijk} \sim N(0, \sigma^2)$ ,  $u_{0jk} \sim N(0, \sigma_{u_0}^2)$ ,  $\nu_{00k} \sim N(0, \sigma_{\nu_0}^2)$  y que  $e_{ijk}$ ,  $u_{0jk}$  y  $\nu_{00k}$  son independientes, la varianza de  $y_{ij}$  en la Ecuación 2.5 es

$$var(y_{ijk}) = \sigma_{\nu_0}^2 + \sigma_{u_0}^2 + \sigma^2$$

Una forma de definir el CCI en los niveles dos y tres, atribuida a Davis y Scott en 1995 [20], es:

$$\rho_{\text{nivel 3}} = \frac{\sigma_{\nu_0}^2}{\sigma_{\nu_0}^2 + \sigma_{u_0}^2 + \sigma^2}$$

$$\rho_{\text{nivel 2}} = \frac{\sigma_{u_0}^2}{\sigma_{\nu_0}^2 + \sigma_{u_0}^2 + \sigma^2},$$

Ahora, se considera un modelo multinivel de tres niveles con dos variables independientes  $x_1$  y  $x_2$ . La primera tiene un efecto fijo y la segunda un efecto aleatorio:

$$y_{ijk} = \beta_{0jk} + \alpha_1 x_{1ijk} + \beta_{2jk} x_{2ijk} + e_{ijk} \quad (2.6)$$

Suponiendo que los coeficientes aleatorios  $\beta_{0jk}$  y  $\beta_{2jk}$  son explicados por una variable de segundo nivel  $w_1$  por las relaciones:

$$\beta_{0jk} = \gamma_{00k} + \gamma_{01k} w_{1jk} + u_{0jk} \quad (2.7)$$

$$\beta_{2jk} = \gamma_{10k} + \gamma_{11k} w_{1jk} + u_{1jk}$$

Y los coeficientes aleatorios en la Ecuación 2.7 son explicados por una variable de tercer nivel  $z_1$  por las ecuaciones:

$$\gamma_{00k} = \xi_{000} + \xi_{001} z_{1k} + \nu_{00k}$$

$$\gamma_{01k} = \xi_{010} + \xi_{011} z_{1k} + \nu_{01k} \quad (2.8)$$

$$\gamma_{10k} = \xi_{100} + \xi_{101} z_{1k} + \nu_{10k}$$

$$\gamma_{11k} = \xi_{110} + \xi_{111} z_{1k} + \nu_{11k}$$

Sustituyendo la Ecuación 2.8 en la Ecuación 2.7 y luego en la Ecuación 2.6, se tiene el modelo multinivel de tres niveles:

$$\begin{aligned}
y_{ijk} = & \xi_{000} + \alpha_1 x_{1ijk} + \xi_{001} z_{1k} + \xi_{010} w_{1jk} + \xi_{100} x_{2ijk} + \xi_{011} z_{1k} w_{1jk} \\
& + \xi_{101} z_{1k} x_{2ijk} + \xi_{110} w_{1jk} x_{2ijk} + \xi_{111} z_{1k} w_{1jk} x_{2ijk} \\
& + (u_{0jk} + \nu_{00k} + \nu_{10k} x_{2ijk} + u_{1jk} x_{2ijk} + \nu_{01k} w_{1jk} + \nu_{11k} w_{1jk} x_{2ijk} + e_{ijk})
\end{aligned} \tag{2.9}$$

donde los coeficientes de regresión  $\xi$  y  $\alpha$  son la parte fija del modelo, y los términos residuales de cada nivel contenidos entre paréntesis son la parte aleatoria.

La generalización del modelo de tres niveles de la Ecuación 2.8 al caso en que el nivel uno contiene  $P$  variables independientes  $x_p$  que tienen un efecto fijo  $\alpha_p$  y  $Q$  variables independientes  $x_q$ ,  $q = P + 1, \dots, P + Q$  que tienen un efecto aleatorio, donde; el nivel dos contiene  $M$  variables independientes  $w_m$ ,  $m = 1, \dots, M$  y el nivel tres contiene  $L$  variables independientes  $z_l$ ,  $l = 1, \dots, L$  es:

$$\begin{aligned}
y_{ijk} = & \xi_{000} + \sum_{p=1}^P \alpha_p x_{pijk} + \sum_{q=P+1}^{P+Q} \xi_{q00} x_{qijk} + \sum_{m=1}^M \xi_{0m0} w_{mjk} + \sum_{l=1}^L \xi_{00l} z_{lk} \\
& + \sum_{m=1}^M \sum_{l=1}^L \xi_{0ml} z_{lk} w_{mjk} + \sum_{q=P+1}^{P+Q} \sum_{l=1}^L \xi_{q0l} z_{lk} x_{qijk} + \sum_{q=P+1}^{P+Q} \sum_{m=1}^M \xi_{qm0} w_{mjk} x_{qijk} \\
& + \sum_{q=P+1}^{P+Q} \sum_{m=1}^M \sum_{l=1}^L \xi_{qml} z_{lk} w_{mjk} x_{qijk} + \left( \nu_{00k} + u_{0jk} + \sum_{q=P+1}^{P+Q} \nu_{10k} x_{qijk} \right. \\
& \left. + \sum_{q=P+1}^{P+Q} u_{qjk} x_{qijk} + \sum_{m=1}^M \nu_{0mk} w_{mjk} + \sum_{q=P+1}^{P+Q} \sum_{m=1}^M \nu_{qmk} w_{mjk} x_{qijk} + e_{ijk} \right).
\end{aligned} \tag{2.10}$$

Cuando el modelo de tres niveles incluye una pendiente aleatoria de nivel uno y una pendiente aleatoria de nivel dos, el modelo fácilmente incluye muchos parámetros (interacción y un efecto residual por cada coeficiente de pendiente aleatoria) que fácilmente causan problemas de convergencia, excepto para conjuntos de datos suficientemente grandes. Por lo tanto, la mayoría de los modelos de tres niveles tienen pocos coeficientes de pendiente aleatoria.

El modelo matricial equivalente para un modelo de tres niveles es:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{W}\mathbf{u} + \mathbf{Z}\boldsymbol{\nu} + \mathbf{e} \tag{2.11}$$

Nuevamente,  $\mathbf{Y}$  es el vector de mediciones de la variable dependiente,  $\mathbf{X}$  es la matriz de diseño del vector de parámetros de efectos fijos  $\beta$  (que contiene la gran media, los efectos principales y las interacciones),  $\mathbf{W}$  es la matriz de diseño de los efectos aleatorios dados por el vector  $\mathbf{u}$ ,  $\mathbf{Z}$  es la matriz de diseño de los efectos aleatorios dados por el vector  $\nu$ , y  $\mathbf{e}$  es el vector de errores residuales de nivel uno.

### 2.1.3. Supuestos de los modelos multinivel

Los supuestos estadísticos tales como distribución normal de los errores residuales e independencia entre errores en diferentes niveles han sido mencionados en la definición del modelo multinivel nulo. Se definen explícitamente en esta sección.

La dimensión del vector  $\mathbf{u}$  depende del número de coeficientes aleatorios en la ecuación de nivel uno, por ejemplo, en la Ecuación 2.7 la dimensión es dos. De manera similar, la dimensión del vector  $\nu$  depende del número de coeficientes aleatorios en la ecuación de nivel dos, por ejemplo, en la Ecuación 2.8 la dimensión es cuatro. Sea  $\mathbf{e} = (e_{111} \ e_{112} \ \dots)^T$ ,  $\mathbf{u} = (u_{0jk} \ u_{1jk} \ \dots \ u_{sjk})^T$  y  $\nu = (\nu_{00k} \ \dots \ \nu_{0tk} \ \dots \ \nu_{t0k} \ \dots \ \nu_{ttk})^T$  con dimensiones  $N$ ,  $s$  y  $t^2$  respectivamente.

Los supuestos de los modelos multinivel son:

$$\begin{pmatrix} \nu \\ \mathbf{u} \\ \mathbf{e} \end{pmatrix} \sim N \left( \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{D} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{R} \end{pmatrix} \right) \quad (2.12)$$

donde  $\mathbf{0}$  es el vector de ceros con la dimensión apropiada y

$$\mathbf{D} = \begin{pmatrix} \sigma_{\nu 0}^2 & \sigma_{\nu 01} & \dots & \sigma_{\nu 0t} \\ \sigma_{\nu 01} & \sigma_{\nu 1}^2 & \dots & \sigma_{\nu 1t} \\ & \dots & & \\ \sigma_{\nu 0t} & \sigma_{\nu 1t} & \dots & \sigma_{\nu t}^2 \end{pmatrix}, \quad \mathbf{G} = \begin{pmatrix} \sigma_{u0}^2 & \sigma_{u01} & \dots & \sigma_{u0s} \\ \sigma_{u01} & \sigma_{u1}^2 & \dots & \sigma_{u1s} \\ & \dots & & \\ \sigma_{u0s} & \sigma_{u1s} & \dots & \sigma_{us}^2 \end{pmatrix}, \quad \mathbf{R} = \sigma^2 \mathbf{I} \quad (2.13)$$

La Ecuación 2.12 dice:

- 1) Los errores de nivel uno,  $\mathbf{e}$ , son independientes, con una distribución normal idéntica con media cero y varianza  $\sigma^2$ .
- 2) Los errores de nivel dos,  $\mathbf{u}$ , siguen una distribución normal multivariante con media  $\mathbf{0}$  y

covarianza  $G$ .

3) Los errores de nivel tres,  $\nu$ , siguen una distribución normal multivariante con media 0 y covarianza  $D$ .

4) Los errores de nivel uno y nivel dos son independientes  $Cov(\mathbf{e}, \mathbf{u}) = \mathbf{0}$ .

5) Los errores de nivel uno y nivel tres son independientes  $Cov(\mathbf{e}, \nu) = \mathbf{0}$ .

6) Los errores de nivel dos y tres son independientes  $Cov(\mathbf{u}, \nu) = \mathbf{0}$ .

#### 2.1.4. Estimación de modelos multinivel

La estimación de un modelo lineal multinivel es compleja debido a que además de los residuos en el nivel individual en el modelo están presentes más términos residuales de interceptos y/o pendientes aleatorios de los niveles superiores. Es decir, deben estimarse simultáneamente los efectos fijos, y efectos aleatorios de las matrices  $D$ ,  $G$  y  $R$ . Cuando estas matrices son conocidas, la varianza de  $Y$  dado que las matrices  $D$ ,  $G$  y  $R$  son conocidas se expresa como:

$$\hat{V} = \mathbf{W}\mathbf{D}\mathbf{W}' + \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R} \quad (2.14)$$

usando la inversa de la matriz  $\hat{V}$  como una ponderación, la estimación de los coeficientes de regresión y los errores estándar del modelo pueden ser estimados usando mínimos cuadrados generalizados (GLS, por sus siglas en inglés) [5]. Sin embargo, las matrices  $D$ ,  $G$  y  $R$  son desconocidas. Un método para estimar los componentes de varianza-covarianza de las matrices  $D$ ,  $G$  y  $R$  es la estimación por máxima verosimilitud (ML, por sus siglas en inglés).

Considerando el modelo de dos niveles de la Ecuación 2.4, la verosimilitud se define como

$$L(\mathbf{Y}|\boldsymbol{\beta}) = \int f(\mathbf{Y}|\mathbf{u}, \boldsymbol{\beta})p(\mathbf{u}|\boldsymbol{\beta})d\mathbf{u}. \quad (2.15)$$

donde  $f(\mathbf{Y}|\mathbf{u}, \boldsymbol{\beta})$  es la distribución de probabilidad de  $Y$  en el nivel 1 y  $p(\mathbf{u}|\boldsymbol{\beta})$  es la distribución de los efectos aleatorios dados los parámetros. De manera análoga se define la verosimilitud de un modelo de tres niveles:

$$L(\mathbf{Y}|\boldsymbol{\beta}) = \int f(\mathbf{Y}|\boldsymbol{\nu}, \mathbf{u}, \boldsymbol{\beta})p(\boldsymbol{\nu}, \mathbf{u}|\boldsymbol{\beta})d\boldsymbol{\nu}d\mathbf{u}. \quad (2.16)$$

donde  $\boldsymbol{\nu}$  al igual que en la Ecuación 2.11 es el vector de efectos aleatorios de nivel 3. A pesar de que las estimaciones obtenidas por el método de ML tienen propiedades favorables en muestras grandes (son consistentes, asintóticamente normales y la función de verosimilitud puede ser usada para la evaluación del ajuste del modelo y para comparaciones de modelos), evaluar la integral de la Ecuación 2.15 (o de la Ecuación 2.16) y posteriormente maximizar su resultado es un problema que muchas veces no es sencillo [21].

Algunos de los algoritmos para maximizar la verosimilitud son EM (máximización de la esperanza) y Fisher scoring. El primero, tiene un enfoque ad hoc para la estimación con datos faltantes. Este algoritmo (EM) es compuesto por el paso E que calcula la verosimilitud esperada mediante la inclusión de variables latentes y por el paso M que maximiza la verosimilitud esperada del paso E. Este es un proceso iterativo, de forma que, las estimaciones obtenidas en el paso M son utilizadas para calcular el paso E de la iteración siguiente. Algunas de las desventajas del algoritmo EM son una convergencia lenta e incluso los pasos E y M se pueden volver intratables en problemas complejos de datos incompletos [22].

El algoritmo de Fisher scoring es una generalización del método iterativo de Newton que utiliza la matriz de información esperada de Fisher, es decir, inicia con un número pequeño de iteraciones del algoritmo EM que luego producen estimaciones en una vecindad de las estimaciones de ML para después calcular la estimación de Fisher. Si las estimaciones de Fisher se encuentran dentro del espacio de parámetros, se aceptan las estimaciones y se repiten el proceso hasta que las estimaciones de Fisher converjan. El algoritmo de Fisher scoring es equivalente a un algoritmo IGLS en el sentido de buscar minimizar los errores [21, pág. 444].

### **2.1.5. Modelos lineales generalizados multinivel (MLGM)**

Los modelos lineales generalizados multinivel son una extensión de los modelos lineales generalizados.

Un MLGM de tres niveles de variable dependiente  $Y$  condicionado en los efectos aleatorios  $\nu$  y  $\mathbf{u}$  se expresa como:

$$g[E(\mathbf{Y}|\nu, \mathbf{u})] = \boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} + \mathbf{W}\mathbf{u} + \mathbf{Z}\nu \quad (2.17)$$

donde  $g(\cdot)$  es una función de liga conocida, monótona y diferenciable y  $\boldsymbol{\eta}$  es el predictor lineal. Al igual que en un modelo lineal multinivel, se supone que los efectos aleatorios tienen una distribución normal con vector de medias cero y matriz de varianza/covarianza  $\mathbf{D}$  y  $\mathbf{G}$  respectivamente. Los modelos multinivel descritos en las Secciones 2.1.1 y 2.1.2 son un caso particular de los MLGM con  $g$  igual a la función identidad. La varianza de la variable dependiente  $\mathbf{Y}$  condicionada en los efectos aleatorios es:

$$Var(\mathbf{Y}|\nu, \mathbf{u}) = v(E(\mathbf{Y}|\nu, \mathbf{u}))\phi$$

donde,  $v(\cdot)$  es una función de varianza conocida o una función de la media condicional y  $\phi$  es un parámetro de dispersión, es decir, es la varianza residual. La estimación de estos modelos por el método de ML es aún más compleja que para los modelos multinivel, debido a que  $f(\mathbf{Y}|\mathbf{u}, \boldsymbol{\beta})$ , de la Ecuación 2.15, es no normal y por tanto no es posible obtener la verosimilitud de forma analítica. Una alternativa es elegir distribuciones a priori conjugadas para los efectos aleatorios. Sin embargo, elegir a priori's conjugadas es conveniente cuando cada unidad del nivel más alto tiene solo un efecto aleatorio y generalmente este no es el caso, es decir, en la mayoría de los casos se está interesado en efectos aleatorios multivariados [21]. Para resolver este problema de integración se recurre a métodos de aproximación numérica.

Una aproximación simple a la integral de la verosimilitud cuando el modelo no es normal es el algoritmo de cuasi verosimilitud (PQL, por sus siglas en inglés). Este algoritmo consiste en un primer paso tratar el modelo no normal como un modelo lineal multinivel estándar donde los errores tienen distribución normal y mediante algoritmos como EM y Fisher scoring estimar por máxima verosimilitud los componentes de varianza y covarianza y los efectos fijos a partir de estimaciones iniciales de  $\boldsymbol{\eta}$  y de la varianza de los errores residuales. En el paso 2, con las nuevas estimaciones se calculan los nuevos pesos y los nuevos valores de la variable linealizada  $y$ . Este es un proceso iterativo que se repite hasta alcanzar la convergencia [21].

El algoritmo de cuasi verosimilitud marginal (MQL, por sus siglas en inglés) es igual al algoritmo de PQL salvo que ahora la serie de Taylor es expandida alrededor de  $u = 0$  mas que alrededor de la media posterior  $u = u^*$ .

Otros algoritmos alternativos para tratar la no normalidad del modelo son la cuadratura Gauss-Hermite y la aproximación de Laplace de orden alto. La cuadratura Gauss-Hermite consiste en aproximar la integral de la Ecuación 2.15 tanto como se desee y después maximizar la integral aproximada. La aproximación de Laplace es un enfoque alternativo de aproximación que además corrige el sesgo de las estimaciones por PQL. Este enfoque tiene varias ventajas, entre ellas, la aproximación es precisa incluso para cualquier grado requerido, la convergencia es numérica más que estocástica y los cálculos son relativamente rápidos [21].

Más métodos de estimación usados en modelos multinivel cuando los supuestos de los modelos multinivel (Sección 2.1.3) no se cumplen, incluyen a: las *ecuaciones de estimación generalizadas (GEE)*, *métodos bootstrap* y *métodos Bayesianos* [20]. El método de ecuaciones de estimación generalizadas estima la matriz de varianzas y covarianzas directamente de los residuos y los efectos fijos son estimados por medio de GLS. Las dependencias de las unidades que componen un grupo es representada por una matriz de correlación de trabajo. Una desventaja de este método es que solo se aproxima a la estructura de los efectos aleatorios, de manera que los efectos aleatorios no se pueden analizar. El método bootstrap estima los parámetros del modelo y sus errores estándar a partir de una muestra. Por otra parte, la estimación Bayesiana se ha convertido en un método de estimación referente para tratar con tipos de datos no normales, con multicolinealidad y de tamaño de muestra pequeño.

## 2.2. Inferencia Bayesiana

La estimación Bayesiana se obtiene a partir de la distribución posterior que combina la nueva evidencia con la evidencia a priori. La estimación puntual de un parámetro desconocido es una medida de tendencia central de la distribución posterior de ese parámetro, por ejemplo, su media, mediana o moda. Un buen estimador puntual minimiza la distancia esperada entre el parámetro desconocido y su estimador.



Además, en inferencia Bayesiana el concepto de intervalo de confianza es sustituido por el *intervalo creíble* o *intervalo de densidad posterior más alto*. Una estimación del intervalo creíble de un parámetro define la probabilidad posterior de que el parámetro se encuentre en el intervalo [21]. La estimación del intervalo creíble de un parámetro es una forma de probar la significancia estadística del parámetro. Otra forma de probar la significancia de los parámetros es mediante el Factor de Bayes (FB) [20]. El FB se define como el cociente de dos verosimilitudes marginales, en la práctica puede ser interpretado como la evidencia favorable de la hipótesis nula vs la hipótesis alternativa. No obstante, no se recomienda en la práctica porque es muy sensible a los aspectos de un modelo Bayesiano y no se puede comprobar a partir de los datos [23]. Con esta motivación se definen algunos conceptos:

### 2.2.1. Estimación Bayesiana

**Definición 2.2.1** La **distribución de probabilidad conjunta** del parámetro  $\theta$  y los datos  $y$  es de la forma

$$p(\theta, y) = p(\theta)p(y|\theta) \quad (2.18)$$

donde,

$p(\theta)$  es la *distribución a priori* y

$p(y|\theta)$  es la *distribución de verosimilitud* (o *distribución de los datos*).

Las distribuciones a priori pueden ser de dos tipos: *informativas* y *no informativas*. Cuando el investigador tiene un alto grado de certeza sobre los valores que puede tomar  $\theta$  la distribución a priori especificada tendrá una varianza pequeña y será informativa. Si esto no sucede, es decir, el investigador no tiene mucha certeza acerca de los valores que puede tomar  $\theta$ , la distribución a priori especificada tendrá una varianza grande y será no informativa.

**Definición 2.2.2** Condicionando la distribución de probabilidad conjunta en un valor conocido y usando el Teorema de Bayes, la **densidad posterior**,  $p(\theta|y)$ , se define como [23],

$$p(\theta|y) = \frac{p(\theta, y)}{p(y)} = \frac{p(\theta)p(y|\theta)}{p(y)} \quad (2.19)$$

donde  $p(y) = \sum_{\theta} p(\theta)p(y|\theta)$  para  $\theta$  discreto (o  $p(y) = \int p(\theta)p(y|\theta)d\theta$  en el caso de  $\theta$  continuo) es llamada constante de normalización. El cálculo de esta constante es necesario para asegurar que el área bajo la curva de la Ecuación 2.19 es 1. Sin embargo, el cálculo de esta integral ha sido la mayor dificultad práctica en la inferencia Bayesiana, especialmente en altas dimensiones. Una expresión equivalente que omite la constante de normalización  $p(y)$  que no depende de  $\theta$  y que para un  $y$  fijo puede considerarse como una constante es la *distribución posterior no normalizada*:

$$p(\theta|y) \propto p(\theta)p(y|\theta) \quad (2.20)$$

Como la distribución a priori es un factor de la distribución posterior, una distribución a priori informativa tendrá mayor influencia en la estimación de la distribución posterior.

**Definición 2.2.3** La *esperanza posterior* de una función  $h(\theta)$  es:

$$E[h(\theta)|y] = \int h(\theta)p(\theta|y)d\theta \quad (2.21)$$

Cualquier característica de la distribución posterior (momentos, cuantiles, regiones de alta densidad posterior, etc.) puede escribirse en términos de una esperanza posterior.

El método más utilizado para obtener medias, medianas y cuantiles de la distribución posterior es usar la cadena de Markov Monte Carlo.

## 2.2.2. Cadenas de Markov Monte Carlo

**Definición 2.2.4** La *integración Monte Carlo* es un método de simulación que consiste en obtener muestras aleatorias  $\theta^s$  a partir de la distribución deseada  $p(\theta|y)$  y estimar la esperanza de cualquier función  $h(\theta)$  como

$$E(h(\theta)|y) = \int h(\theta)p(\theta|y)d\theta \approx \frac{1}{S} \sum_{s=1}^S h(\theta^s). \quad (2.22)$$

**Definición 2.2.5** Una *cadena de Markov a tiempo discreto* es una sucesión de variables aleatorias  $X_n, n \geq 1$  que toman valores en un conjunto finito o numerable  $\Omega$  y que satisface

$$p(X_{n+1} = j|X_0 = i_0, \dots, X_n = i_n) = p(X_{n+1} = j|X_n = i_n)$$

para todo  $n$  y cuales quiera estados  $i_0, \dots, i_n, j$  en  $\Omega$ .

**Definición 2.2.6** La cadena de Markov  $X_n$ , de la Definición 2.2.5; tiene una **distribución estacionaria**  $p(\theta|y)$ , si bajo condiciones de regularidad, la cadena gradualmente olvida sus estados  $i_0, \dots, i_n$  y  $p(X_{n+1} = j|X_0 = i_0, \dots, X_n = i_n)$  eventualmente converge a una única distribución  $\phi(\cdot)$  que no depende del estado actual  $X_n = j$  ni de los estados anteriores.

El quemado de una MCMC es el número de iteraciones  $m$  a eliminar, para que el resto muestre un comportamiento de muestras dependientes de la distribución estacionaria  $\phi(\cdot)$  [7]. Cuando el número de muestras de quemado es  $m$ , un estimador de la esperanza de  $f(X)$  para  $X$  un vector aleatorio es

$$E[\widehat{f(X)}] = \frac{1}{n - m} \sum_{t=m+1}^n f(X_t).$$

La eficacia de la MCMC no se debe a su propiedad en sí, sino a que las distribuciones aproximadas mejoran en cada paso de la simulación en el aspecto de converger a la distribución objetivo [23].

Luego de especificar el modelo jerárquico y elegir la teoría de estimación para hacer las inferencias estadísticas sobre los parámetros de la población es necesario elegir un algoritmo computacional que implemente la teoría de estimación [21, pág. 437]. Debido a que en este trabajo se sigue un método de estimación Bayesiano, a continuación se describen algunos de los algoritmos pertinentes para este tipo de estimación.

## Muestreo Gibbs

El muestreador de Gibbs, también conocido como *algoritmo de muestreo condicional alterno* es el más comúnmente utilizado en muchos problemas multidimensionales. Suponga que el vector de parámetros  $\theta$  ha sido dividido en  $d$  componentes o subvectores,  $\theta = (\theta_1, \dots, \theta_d)$ . Por tanto, la iteración  $t$  tiene  $d$  pasos. En cada iteración se elige un orden de los  $d$  subvectores y cada iteración se muestrea de la distribución condicional al valor de todos los demás subvectores. Es decir, el subvector  $j$  en la iteración  $t$ , denotado como  $\theta_j^t$  es muestreado de la distribución condicional  $p(\theta_j|\theta_{-j}^{t-1}, y)$  donde  $\theta_{-j}^{t-1}$  representa todas las componentes de  $\theta$ , excepto  $\theta_j$  en sus valores actuales [23],

$$\theta_{-j}^{t-1} = (\theta_1^t, \dots, \theta_{j-1}^t, \theta_{j+1}^{t-1}, \dots, \theta_d^{t-1})$$

## Algoritmo Metropolis

Es una adaptación de una caminata aleatoria con una regla de aceptación/rechazo para converger a la distribución objetivo. El algoritmo es implementado en 2 pasos: 1) Muestrear un punto inicial  $\theta^0$  para el que  $p(\theta^0|y) > 0$ , a partir de una distribución inicial  $p_0(\theta)$ . 2) Para  $t = 1, 2, \dots$ ; (a) Muestrear un parámetro  $\theta^*$  de la distribución propuesta en el tiempo  $t$ ,  $J_t(\theta^*|\theta^{t-1})$ . La distribución propuesta o también conocida como distribución de salto debe ser simétrica ( $J_t(\theta_a|\theta_b) = J_t(\theta_b|\theta_a)$  para todo  $\theta_a, \theta_b$  y  $t$ ). (b) Calcular la razón  $r$  de densidades,  $r = \frac{p(\theta^*|y)}{p(\theta^{t-1}|y)}$ . (c) Definir  $\theta^t = \theta^*$  con probabilidad igual a  $\min(r, 1)$  y  $\theta^t = \theta^{t-1}$  en otro caso. Para poder elegir una de estas opciones se requiere generar un número aleatorio uniforme [23].

## Hamiltoniano de Monte Carlo (HMC)

El muestreo de Gibbs y la caminata aleatoria Metropolis son métodos que convergen en distribución a la distribución objetivo, sin embargo, para modelos complicados con un gran número de parámetros pueden requerir un tiempo inaceptablemente largo para alcanzar la convergencia. Este problema en gran parte es ocasionado por caminatas aleatorias ineficientes que estiman el espacio de parámetros.

El algoritmo Hamiltoniano de Monte Carlo o algoritmo híbrido de Monte Carlo elimina las caminatas aleatorias por medio de variables de momento que transforman el problema de muestreo de la distribución objetivo en el problema de simulación de la dinámica Hamiltoniana. El integrador “leapfrog” (de pasos de salto) de Störmer-Verlet es usado para simular la evolución en el tiempo de este sistema.

Dada una muestra  $m$ , un tamaño de paso  $\epsilon$  y un número de pasos  $L$  el algoritmo HMC consiste en volver a muestrear las variables de momento,  $r_d$ , de una distribución normal multivariada estándar (puede verse como una actualización del muestro de Gibbs), después aplicar  $L$  actualizaciones “leapfrog” a las variables de posición y momento ( $\theta$  y  $r$ ) para generar un par de variables propuestas de posición y momento ( $\tilde{\theta}, \tilde{r}$ ) que se definen como  $\theta^m = \tilde{\theta}$  y  $r^m = \tilde{r}$ , y serán aceptadas o rechazadas de acuerdo al algoritmo Metropolis. Para más detalles ver [24]. En general, especificar los parámetros de tamaño de paso

( $\epsilon$ ) y número de pasos ( $L$ ) es bastante difícil cuando la trayectoria es demasiado corta, larga o recta.

### Muestreo No-U-Turn (NUTS)

El muestreo No-U-Turn (NUTS, por sus siglas en inglés) es una extensión del algoritmo HMC que elimina la necesidad de especificar un valor fijo  $L$  de número de pasos y la especificación del tamaño de paso se basa en el algoritmo de promedio dual. Un criterio para determinar si la dinámica Hamiltoniana ha sido simulada un tiempo suficiente (es decir, si al ejecutar la simulación para más pasos la distancia entre el valor propuesto  $\tilde{\theta}$  y el valor inicial  $\theta$  ya no aumenta) se basa en el producto punto entre el momento actual y el vector desde la posición inicial hasta la posición actual, es decir, es el producto punto de la derivada con respecto al tiempo del cuadrado de la distancia entre la posición inicial y la posición final:

$$\frac{d}{dt} \frac{(\tilde{\theta} - \theta) \cdot (\tilde{\theta} - \theta)}{2} = (\tilde{\theta} - \theta) \cdot \frac{d}{dt} (\tilde{\theta} - \theta) = (\tilde{\theta} - \theta) \cdot \tilde{r} \quad (2.23)$$

En otras palabras, si las simulaciones se corren un número infinitesimal de tiempo entonces esta cantidad es proporcional al progreso que se obtiene desde el punto de partida. Esto sugiere que el número de pasos de salto a ejecutarse se fija cuando la última igualdad de la Ecuación 2.23 se vuelve menor que cero, es decir, hasta que la posición propuesta  $\tilde{\theta}$  comienza a moverse hacia la posición inicial  $\theta$ . Esta acción es llamada *reversibilidad* pero este algoritmo no la garantiza y por tanto, no se garantiza la convergencia a la distribución objetivo. NUTS por el contrario, preserva la reversibilidad ejecutando la simulación Hamiltoniana hacia adelante y hacia atrás en el tiempo. El muestreo No-U-Turn inicia introduciendo una “variable slice”  $u$  con una distribución condicional  $p(u|\theta, r)$  uniforme ( $u; [0, \exp[\mathcal{L}(\theta) - \frac{1}{2}r \cdot r]]$ ) que produce una distribución condicional  $p(\theta, r|u)$  uniforme  $[\theta, r; \{\theta', r' | \exp[\mathcal{L}(\theta) - \frac{1}{2}r \cdot r] \geq u\}]$ . Este paso no es estrictamente necesario pero simplifica la derivación e implementación de NUTS. Después mediante el integrador leapfrog se traza una trayectoria ficticia hacia adelante y hacia atrás. Primero, avanza o retrocede un paso, avanza y retrocede dos pasos, en seguida avanza o retrocede 4 pasos y así sucesivamente. Este proceso genera un árbol binario cuyos nodos corresponden a los estados de posición - movimiento. El proceso de duplicación se termina cuando la subtrayectoria

del nodo más a la izquierda hasta el nodo más a la derecha de cualquier subárbol equilibrado del árbol binario empieza a duplicarse, es decir,  $\theta$  comienza a dar vuelta en forma de U [24].

## El paquete brms

Varios softwares hacen posible la aplicación de algoritmos de MCMC como: WinBUGS, openBUGS, JAGS, MCMCgllmm y Stan [25]. A excepción de este último, todos estos programas usan muestreo Gibbs y actualizaciones Metropolis-Hastings, algunos incorporan muestreo slice. Sin embargo, un problema de estos softwares es su convergencia lenta en modelos de altas dimensiones con parámetros correlacionados. Por el contrario, Stan implementa el algoritmo HMC y su extensión, el muestreador NUTS. Este lenguaje de programación probabilístico demanda menor tiempo para la convergencia, incluso en modelos de altas dimensiones y las distribuciones a priori no son conjugadas necesariamente. El paquete brms ajusta modelos multinivel a través del lenguaje de programación probabilístico Stan [26].

## Diagnósticos de convergencia

Después de un número suficientemente grande de iteraciones, la MCMC eventualmente converge a la distribución posterior. Se necesita un estadístico de diagnóstico para determinar si la MCMC ya ha convergido a la distribución estacionaria o si se necesitan más iteraciones. Se han propuesto varios estadísticos de diagnóstico, pero en este trabajo son de interés el diagnóstico de Gelman & Rubin y los diagnósticos gráficos.

**Diagnóstico de Gelman & Rubin (GR)** [27]. Este diagnóstico utiliza varias cadenas  $\{X_{i0}, \dots, X_{in-1}\}$ ,  $i = 1, \dots, m$  extraídas de una densidad sobredispersa con respecto a la densidad objetivo  $\phi(\cdot)$ . En 1992, Gelman y Rubin definieron dos estimadores de la varianza de  $X$  cuando  $X \sim \phi(\cdot)$ :

1) La *varianza dentro de la cadena*:  $W = \sum_{i=1}^m \sum_{j=0}^{n-1} (X_{ij} - \bar{X}_i)^2 / (m(n-1))$ , y

2) La *varianza combinada*:  $\hat{V} = ((n-1)/n)W + B/n$

donde  $B/n = \sum_{i=1}^m (\bar{X}_i - \bar{X}_{..})^2 / (m-1)$  es la estimación de *varianza entre cadenas*,  $\bar{X}_i$  es la media de la cadena  $i$ ,  $i = 1, \dots, m$ , y  $\bar{X}_{..}$  es la media general. El factor de reducción de escala potencial (PSRF, por sus siglas en inglés) o Rhat se define por:

$$\hat{R} = \frac{\hat{V}}{W}$$

La varianza en el numerador de  $\hat{R}$  sobrestima la varianza objetivo, mientras que la varianza en el denominador la subestima. Este hecho produce  $\hat{R}$  mayor que 1. Un criterio para detener la simulación MCMC es que  $\hat{R} \approx 1$  o  $\hat{R} < 1.1$ . El diagnóstico GR se implementa en el paquete *coda* [28] del paquete R [29].

**Diagnóstico gráfico.** Los gráficos de trazas de MCMC son los gráficos de diagnóstico más utilizados para determinar la convergencia. Son series de tiempo que muestran el comportamiento de las cadenas de Markov alrededor de su espacio de estado y sus realizaciones en cada iteración. Cuando las tendencias visibles muestran cambios en la dispersión de la traza de la cadena, la MCMC no ha alcanzado un estado estacionario. Por el contrario, cuando se observa una buena mezcla, se dice que la MCMC converge a la distribución objetivo [27].

### 2.2.3. Evaluación y comparación de modelos

#### Evaluación del modelo

Cualquier análisis Bayesiano debe incluir una verificación de la adecuación del ajuste del modelo postulado a los datos. La adecuación del ajuste de un modelo se mide por qué tan bien se aproxima la distribución del modelo propuesto a la distribución de los datos; cuanto mejor sea el ajuste del modelo postulado a los datos, mejor será el modelo. Pero, si este ajuste es pobre no significa que el modelo sea malo sino más bien que contiene deficiencias que podrían mejorar. En esta sección se explica un método de evaluación del modelo basado en la distribución predictiva posterior.

Sea  $y^{rep}$  los datos replicados que podrían ser observados mañana si el experimento que produjo los datos actuales  $y$  se replicara mañana con el mismo modelo y los mismos valores de  $\theta$  que produjeron los  $y$ . La distribución de  $y^{rep}$  dados los datos actuales  $y$  se llama *distribución predictiva posterior* y se define como [23]

$$y^{rep}|y = \int p(y^{rep}|\theta)p(\theta|y)d\theta \quad (2.24)$$

Si el modelo es preciso, es decir, tiene un ajuste razonablemente bueno se espera que

las réplicas hipotéticas sean similares a los datos observados. Una forma de realizar la evaluación predictiva posterior es graficar las replicas con los datos observados, mediante densidades o gráficas de dispersión.

## Comparación de modelos

La **log densidad predictiva puntual** (lppd, por sus siglas en inglés) del modelo ajustado a los datos para un parámetro desconocido  $\theta$  se define como

$$lppd = \log \prod_{i=1}^n p(\theta|y_i) = \sum_{i=1}^n \log \int p(y_i|\theta)p(\theta|y) d\theta \quad (2.25)$$

La precisión predictiva se usa para medir el rendimiento del modelo y para la comparación de modelos [23]. Hay diversos métodos disponibles para calcular la esperanza predictiva fuera de la muestra. El primer método consiste en usar muestras de  $p(\theta|y)$  obtenidas con  $\theta^s$ ,  $s = 1, \dots, S$  simulaciones suficientes MCMC. Así, la *lppd calculada* (clppd) se define como

$$clppd = \sum_{i=1}^n \log \left( \frac{1}{S} \sum_{s=1}^S p(y_i|\theta^s) \right)$$

Un segundo método para evaluar la esperanza predictiva posterior es la *precisión predictiva dentro de la muestra ajustada* que consiste en una corrección del sesgo de la estimación de la lppd a partir de criterios de información como el criterio de información de Akaike, el criterio de información de devianza o el criterio de información de Watanabe-Akaike.

Un tercer método para evaluar la esperanza predictiva posterior es la *validación cruzada*, que captura el error predictivo fuera de la muestra a través de ajustar el modelo a los datos de entrenamiento y evaluar el ajuste predictivo en los datos reserva [23]. Se explica este método en detalle:

**Validación cruzada de dejar uno fuera** (LOO-CV, por sus siglas en inglés). Funciona con  $n$  particiones en las que cada conjunto reservado tiene solo una observación, lo que genera  $n$  inferencias diferentes  $p_{post(-i)}$  obtenido mediante  $S$  simulaciones posteriores,  $\theta^{is}$ .



La estimación LOO-CV Bayesiana del ajuste predictivo fuera de la muestra es

$$lppd_{loo-cv} = \sum_{i=1}^n \log p_{post(-i)}(y_i) \approx \sum_{i=1}^n \log \left( \frac{1}{S} \sum_{s=1}^S p(y_i | \theta^{is}) \right)$$

Cada predicción está condicionada en  $n-1$  puntos de datos, lo que subestima el ajuste predictivo. Para  $n$  grande, la diferencia es insignificante; sin embargo, para  $n$  pequeños, se puede usar una corrección de sesgo de primer orden  $b = lppd - \overline{lppd}_{-i}$ , donde

$$\overline{lppd}_{-i} = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \log p_{post(-i)}(y_j) \approx \frac{1}{n} \sum_{i=1}^S \sum_{j=1}^n \log \left( \frac{1}{S} \sum_{s=1}^S p(y_j | \theta^{is}) \right)$$

El LOO-VC Bayesiano corregido por sesgo es

$$lppd_{eloo-cv} = lppd_{loo-cv} + b \quad (2.26)$$

Una estimación del número de parámetros efectivos es

$$p_{loo-cv} = lppd - lppd_{loo-cv}$$

Al comparar dos modelos ajustados se puede estimar la diferencia en su precisión predictiva esperada por la diferencia en  $elppd_{loo-cv}$ . El error estándar de la diferencia se puede calcular usando una estimación pareada para aprovechar el hecho de que se usa el mismo conjunto de  $n$  puntos de datos para ajustar ambos modelos.

Suponga que se comparan los modelos I y II, con las medidas de ajuste correspondientes  $elppd_{loo-cv}^I$  y  $elppd_{loo-cv}^{II}$ , entonces la diferencia y su error estándar son

$$\begin{aligned} elppd\_diff &= elppd_{loo-cv}^I - elppd_{loo-cv}^{II} \\ se\_diff &= se \left( elppd_{loo-cv}^I - elppd_{loo-cv}^{II} \right) = \sqrt{nV_i^n \left( elppd_{loo,i}^I - elppd_{loo,i}^{II} \right)} \end{aligned}$$

Cuando se comparan dos modelos mediante el estadístico LOO-CV, el que tiene el valor más bajo de este estadístico se declara el mejor modelo. Si se utiliza `elppd_diff` con la función `loo_compare` del paquete `brms` [26], el valor de la diferencia se informa en el mejor modelo acompañado de su `se_diff`. Hay más evidencia de la superioridad de un modelo sobre otro cuando `elppd_diff` es mayor que el `se_diff`.

## 2.3. Metodología de los modelos multinivel

Para proponer un modelo multinivel, es necesario determinar qué variables estarán en la parte fija, cuáles en la parte aleatoria y las interacciones entre niveles. Esta tarea puede ser compleja, por lo que es necesaria una estrategia para construir el modelo.

### 2.3.1. Estrategia de construcción de modelos multinivel

En esta subsección, se muestra una adaptación de la estrategia bottom-up para especificar un modelo multinivel de tres niveles. La estrategia bottom-up se utiliza en el enfoque frecuentista [20]. Esta propuesta de adaptación utiliza el LOO-CV Bayesiano del paso dos al paso siete para la comparación de modelos.

*Paso 1.* Ajustar del modelo de solo intercepto:

$$y_{ijk} = \xi_{000} + \nu_{00k} + u_{0jk} + e_{ijk}, \quad (2.27)$$

Este modelo da una línea base para comparar con los modelos siguientes.

*Paso 2.* Agregar todas las variables independientes fijas de nivel uno:

$$y_{ijk} = \xi_{000} + \sum_{p=1}^P \alpha_p x_{pijk} + \sum_{q=P+1}^{P+Q} \xi_{q00} x_{qijk} + (\nu_{00k} + u_{0jk} + e_{ijk}) \quad (2.28)$$

Debe determinarse qué variable de nivel uno tuvo un efecto significativo en  $y$ . Se supondrá que todas las variables de nivel uno son estadísticamente significativas. Los Modelos 2.28 y 2.27 deben ser comparados.

*Paso 3.* Agregar las variables independientes fijas de nivel dos:

$$y_{ijk} = \xi_{000} + \sum_{p=1}^P \alpha_p x_{pijk} + \sum_{q=P+1}^{P+Q} \xi_{q00} x_{qijk} + \sum_{m=1}^M \xi_{0m0} w_{mjk} + (\nu_{00k} + u_{0jk} + e_{ijk}) \quad (2.29)$$

Debe determinarse qué variable de nivel dos tuvo un efecto significativo en  $y$ . Si las variables  $w_m$  explican la variabilidad de  $y$ , el Modelo 2.29 deberá ser superior al Modelo 2.28. Se supondrá que todas las variables independientes de nivel dos son estadísticamente significativas.

*Paso 4.* Agregar las variables independientes fijas de nivel tres:

$$y_{ijk} = \xi_{000} + \sum_{p=1}^P \alpha_p x_{pijk} + \sum_{q=P+1}^{P+Q} \xi_{q00} x_{qijk} + \sum_{m=1}^M \xi_{0m0} w_{mjk} + \sum_{l=1}^L \xi_{00l} z_{lk} + (\nu_{00k} + u_{0jk} + e_{ijk}) \quad (2.30)$$

Debe determinarse qué variable de nivel tres tuvo un efecto significativo en  $y$ . Si las variables  $z_l$  explican la variabilidad de  $y$ , el Modelo 2.30 deberá ser superior al Modelo 2.29. Nuevamente, se supondrá que todas las variables independientes de nivel tres son estadísticamente significativas.

Los pasos 1-4 consideran la especificación de la parte fija del multinivel de tres niveles. Ahora se especifica la parte aleatoria del modelo.

*Paso 5.* Evaluar si alguna de las pendientes de las variables independientes en el nivel uno tiene un componente de varianza significativo entre los grupos en el nivel dos o en el nivel tres.

$$y_{ijk} = \xi_{000} + \sum_{p=1}^P \alpha_p x_{pijk} + \sum_{q=P+1}^{P+Q} \xi_{q00} x_{qijk} + \sum_{m=1}^M \xi_{0m0} w_{mjk} + \sum_{l=1}^L \xi_{00l} z_{lk} + \left( \nu_{00k} + u_{0jk} + \sum_{q=P+1}^{P+Q} \nu_{q0k} x_{qijk} + \sum_{q=P+1}^{P+Q} u_{qjk} x_{qijk} + e_{ijk} \right) \quad (2.31)$$

donde  $u_{qjk}$  son los residuos de nivel dos de las pendientes de las variables independientes de nivel uno  $x_q$ , y  $\nu_{q0k}$  son los residuos de nivel tres de las pendientes de las variables independientes de nivel dos  $w_m$ .

Las variables independientes de nivel uno que no tienen una pendiente significativa pueden tener una pendiente aleatoria significativa. Este paso y el siguiente deben realizarse con cuidado, ya que el modelo puede sobreparametrizarse fácilmente y/o tener problemas como la falta de convergencia o cálculos extremadamente lentos. Es recomendable evaluar la significancia de las pendientes variable por variable. A continuación, se formula el modelo con todas las variables con pendientes aleatorias significativas. Si el Modelo 2.31 no es mejor que el Modelo 2.30, el procedimiento para especificar un modelo multinivel de tres niveles se detiene.

*Paso 6.* Evaluar si alguna de las pendientes de la variable independiente de nivel dos tiene un componente de varianza significativo entre los grupos de nivel tres.

$$\begin{aligned}
y_{ijk} = & \xi_{000} + \sum_{p=1}^P \alpha_p x_{pijk} + \sum_{q=P+1}^{P+Q} \xi_{q00} x_{qijk} + \sum_{m=1}^M \xi_{0m0} w_{mjk} + \sum_{l=1}^L \xi_{00l} z_{lk} \\
& + \left( \nu_{00k} + u_{0jk} + \sum_{q=P+1}^{P+Q} \nu_{q0k} x_{qijk} + \sum_{q=P+1}^{P+Q} u_{qjk} x_{qijk} + \sum_{m=1}^M \nu_{0mk} w_{mjk} + e_{ijk} \right)
\end{aligned} \tag{2.32}$$

donde  $\nu_{0mk}$  son los residuales de nivel tres de las pendientes de las variables independientes de nivel dos  $w_m$ .

La evaluación de las pendientes aleatorias de las variables de nivel dos se debe realizar variable por variable y, luego, todas estas variables con pendiente aleatoria significativa se deben incluir en el modelo para evaluar la mejora con respecto al Modelo 2.31.

*Paso 7.* Agregar interacciones entre las variables independientes de nivel tres y las variables independientes de nivel uno y nivel dos que tenían una varianza de pendiente significativa en el Paso 5 y el Paso 6. Esto produce el modelo completo:

$$\begin{aligned}
y_{ijk} = & \xi_{000} + \sum_{p=1}^P \alpha_p x_{pijk} + \sum_{q=P+1}^{P+Q} \xi_{q00} x_{qijk} + \sum_{m=1}^M \xi_{0m0} w_{mjk} + \sum_{l=1}^L \xi_{00l} z_{lk} \\
& + \sum_{m=1}^M \sum_{l=1}^L \xi_{0ml} z_{lk} w_{mjk} + \sum_{q=P+1}^{P+Q} \sum_{l=1}^L \xi_{q0l} z_{lk} x_{qijk} + \sum_{q=P+1}^{P+Q} \sum_{m=1}^M \xi_{qm0} w_{mjk} x_{qijk} \\
& + \sum_{q=P+1}^{P+Q} \sum_{m=1}^M \sum_{l=1}^L \xi_{qml} z_{lk} w_{mjk} x_{qijk} + \left( \nu_{00k} + u_{0jk} + \sum_{q=P+1}^{P+Q} \nu_{q0k} x_{qijk} \right. \\
& \left. + \sum_{q=P+1}^{P+Q} u_{qjk} x_{qijk} + \sum_{m=1}^M \nu_{0mk} w_{mjk} + \sum_{q=P+1}^{P+Q} \sum_{m=1}^M \nu_{qmk} w_{mjk} x_{qijk} + e_{ijk} \right).
\end{aligned} \tag{2.33}$$

Al explicar las varianzas de las pendientes aleatorias en términos de variables contextuales, el modelo automáticamente incluye términos de interacción entre niveles que componen la parte fija del modelo. Se recomienda agregar variables que expliquen la varianza de los coeficientes de pendiente aleatoria una por una y no como se muestra en este paso (esto se hizo aquí para evitar especificar más ecuaciones).

Cuando se trata de un MLGM, la metodología cambia ligeramente, es decir, en lugar de definir modelos en términos de  $y$ , los modelos se definen en términos de  $g(\mu_{ijk}) = g(E[Y_{ijk} | \nu, u])$ , y ya no se especifican los errores residuales a nivel individual. A continuación se ilustra un ejemplo de esta metodología para un MLGM.

### Resultados

---

En este capítulo, se usa un modelo lineal generalizado multinivel para datos de un estudio transversal realizado por Romero-Castro y colaboradores [8].

Treinta y dos dientes fueron examinados en cada uno de los 116 pacientes (se eliminaron los registros de pacientes que tuvieron un valor perdido en un nivel de los datos). La profundidad de sondaje se registró en seis sitios en cada diente, es decir, en las ubicaciones *mesiobucal*, *mediobucal*, *distobucal*, *mesiolingual*, *mediolingual* y *distolingual* de cada diente. La profundidad de la bolsa se registró mediante el uso de la sonda Florida en los seis sitios. La variable de respuesta fue la *profundidad de sondaje* medida en milímetros, es decir, la profundidad de sondaje es una variable continua y mayor que cero ( $> 0$ ). El conjunto de datos constó de 18,780 observaciones.

Las variables independientes, excepto la edad, fueron todas dicotómicas: *sangrado*, *movilidad*, *placa (punto de corte igual a 100)*, *cálculo (punto de corte igual a 100)*, *resistencia a la insulina (glucosa plasmática en ayunas  $> 100$  mg/dL)*, *fumar*, *restos radiculares* y *restauraciones desajustadas*, donde 0 indica ausencia y 1 presencia. En el Anexo .1 se muestran las frecuencias de estos factores (el factor furcación no se incluyó en el análisis debido a que el 99 % de los pacientes no tuvieron ningún grado de furcación).

La Figura 3.1 y el Cuadro 3.1 muestran los tres niveles de los datos y las variables en cada nivel. El primer nivel correspondió a los sitios de sondaje donde se midió la variable independiente sangrado y la variable respuesta profundidad al sondaje. El nivel

dos correspondió a la pieza dentaria, es decir, los dientes que solo tenían la variable independiente movilidad, y el nivel tres correspondió a los pacientes, donde se midieron las variables independientes edad, placa, cálculo, resistencia a la insulina, fumar, restos radiculares y restauraciones desajustadas.

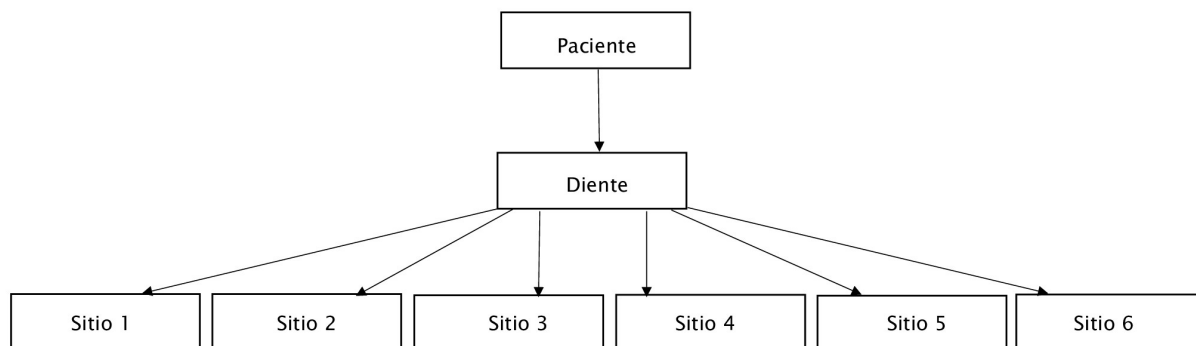


Figura 3.1: Estructura de los datos de profundidad de sondaje para 1 de los 32 dientes de un paciente.

Cuadro 3.1: Datos de profundidad de sondaje

Niveles	VARIABLES DE NIVEL
Nivel 3: Paciente	Edad, Placa, Cálculo, Resistencia a la insulina, Fumar, Restos radiculares Restauraciones desajustadas
Nivel 2: Diente	Movilidad
Nivel 1: Sitios de sondaje	Sangrado Profundidad de sondaje

Se realizó un primer análisis de datos a partir de un modelo multinivel de tres niveles donde se asumió una distribución normal para la profundidad de sondaje. El ajuste frecuentista tuvo dos problemas, los residuos no tenían una distribución normal (Figura 3.2) y el método numérico para obtener las estimaciones no convergía.

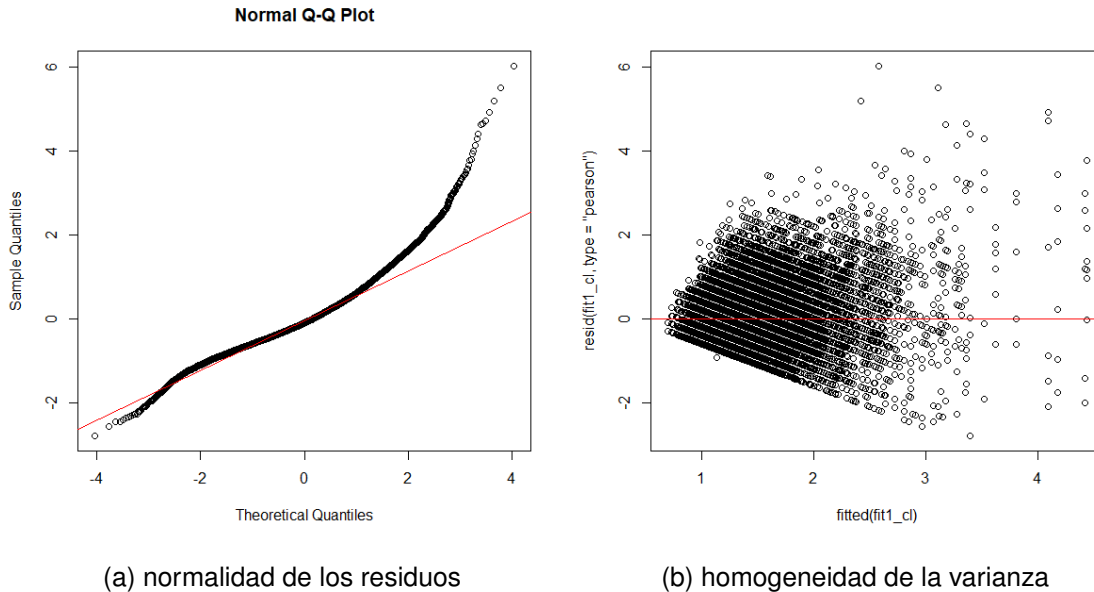


Figura 3.2: Supuestos del modelo de solo intercepto

La profundidad de sondaje mínima fue de 0.2 mm, Q1 de 0.8 mm, Q2 de 1.2 mm, Q3 de 1.8 y la máxima de 9 mm. Además, su distribución era asimétrica a la derecha (sesgo=1.6 y curtosis=8.0). Por lo tanto, se asumió que la profundidad de sondaje tenía una distribución gamma con media  $\mu$  y varianza  $\mu^2/\alpha$ :

$$f(y) = \frac{(\alpha/\mu)^\alpha}{\Gamma(\alpha)} y^{\alpha-1} \exp\left(-\frac{\alpha y}{\mu}\right)$$

Es bien sabido que la regresión gamma pertenece a la familia de modelos lineales generalizados. Pero, como los datos estudiados son de naturaleza jerárquica, el modelo adecuado es el modelo lineal generalizado multinivel. Dado que la variable de respuesta no es negativa, la función de enlace utilizada fue el logaritmo natural para obtener una profundidad de sondaje esperada mayor que cero.

### 3.1. Estimación Bayesiana

**Verosimilitud:** Se asumió que la profundidad de sondaje sigue una distribución gamma.

**Distribución a priori:** Se definió como el producto de distribuciones a priori marginales para cada componente de  $\beta$  en el Modelo 2.17.  $\beta$  está compuesto por la media general, los efectos principales y las interacciones:  $\xi_{000}, \xi_{q00}, \xi_{0m0}, \xi_{00l}, \xi_{0ml}, \xi_{q0l}, \xi_{qm0}, \xi_{qml}$  y todos tenían una a priori  $N(0, 10^2)$ . La función *brms* usa una parametrización especial para las matrices  $D$  y  $G$  en la Ecuación 2.14. Esta parametrización es  $G = F(\sigma_k)\Omega_k F(\sigma_k)$ , donde  $F(\sigma_k)$  es una matriz diagonal con elementos diagonales  $\sigma_k$  [26]. Las a priori para  $D$  y  $G$  solo necesitaron especificar las a priori para  $\sigma_k$  y  $\Omega_k$  que fueron  $\sigma_k \sim HalfCauchy(10)$  y  $\Omega_k \sim CorrLKJ(1)$ . Finalmente, el hiperparámetro de forma fue  $forma \sim Gamma(0.01, 0.01)$ .

El análisis de este modelo se realizó con el paquete *brms* [26, 30] que utiliza el lenguaje de programación probabilística *Stan* [25] en el entorno de *Software R 4.0.5*.

**Simulación:** Todas las MCMC tenían cuatro cadenas, el número de iteraciones y de quemado no fue el mismo para los modelos estudiados, pero todos usaron una muestra final de 4000.

Las MCMC de los Modelos (2.27, 2.28 y 2.29), es decir, nulo, con variables de nivel uno y con nivel dos, se obtuvieron con 4000 iteraciones y un período de quema de 3000. El Modelo 2.30 usó 5000 iteraciones y un período de quema de 4000, el Modelo 2.31 usó 7000 iteraciones y un período de quema de 6000, y el Modelo 2.33 usó 8000 iteraciones y una quema de 7000.

**Estimadores Bayesianos:** Como estimador Bayesiano se utilizó la media de la distribución posterior, esto está relacionado con la minimización de la función de pérdida cuadrática.

**Modelo estudiado:** Se estudió un modelo lineal generalizado multinivel de tres niveles, donde  $i$  representaba las unidades de nivel uno,  $j$  las unidades de nivel dos y  $k$  las unidades de nivel tres. Aunque no se muestran los valores de los Rhats, todas las MCMC de los modelos estudiados convergieron ya que todos los Rhats fueron como máximo 1.01.

**Paso 1.** El modelo nulo es

$$\log(\mu_{ijk}) = \xi_{000} + (\nu_{00k} + u_{0jk}) \quad (3.1)$$

Las Columnas 2 y 3 del Cuadro 3.2 muestran las estimaciones Bayesianas del modelo nulo. Los intervalos creíbles no contienen al cero, por lo que la varianza a nivel de diente



y a nivel de paciente es significativa. Esto apoya el uso del MLGM.

**Paso 2.** El modelo con variables de nivel uno, sangrado, es

$$\log(\mu_{ijk}) = \xi_{000} + \alpha_1 \text{sangrado}_{1ijk} + (\nu_{00k} + u_{0jk}) \quad (3.2)$$

Las estimaciones Bayesianas del modelo mostraron que el coeficiente de sangrado es significativo (Columnas 4 y 5 del Cuadro 3.2). La comparación de los Modelos 3.1 y 3.2, mediante el LOO-CV, indica que el modelo que incluye las variables de nivel uno es mejor (penúltimo renglón y Columna 5 en el Cuadro 3.2).

**Paso 3.** El modelo con variable de nivel dos, movilidad, es

$$\log(\mu_{ijk}) = \xi_{000} + \alpha_1 \text{sangrado}_{1ijk} + \xi_{010} \text{movilidad}_{1jk} + (\nu_{00k} + u_{0jk}) \quad (3.3)$$

El efecto fijo de movilidad no fue significativo (Columnas 6 y 7 del Cuadro 3.2); sin embargo, se retuvo en el modelo porque era la única variable de nivel dos y para obtener estimaciones del efecto de las variables independientes de tercer nivel ajustadas por la variable de efecto de nivel dos. El criterio LOO-CV indica que este modelo es ligeramente mejor (penúltimo renglón y Columna 7 en el Cuadro 3.2).

Hay siete variables contextuales de nivel tres (Cuadro 3.1), antes de especificar el modelo que contiene solo las variables significativas de nivel tres, se realizó una selección de variables hacia delante para evitar tener un modelo sobreparametrizado. El Cuadro 3.3 muestra el procedimiento de selección de variables donde cada modelo contiene la variable de nivel 1, sangrado, y la variable de nivel 2, movilidad. La comparación de modelos LOO-CV indica que el modelo que incluye las variables cálculo y fumar es el mejor modelo.

Cuadro 3.2: Estimaciones Bayesianas de los modelos 3.1 - 3.6

	Modelo 3.1		Modelo 3.2		Modelo 3.3		Modelo 3.4		Modelo 3.5		Modelo 3.6	
	Coef	(ICr 95%)	Coef	(ICr 95%)	Coef	(ICr 95%)	Coef	(ICr 95%)	Coef	(ICr 95%)	Coef	(ICr 95%)
<b>Efectos a nivel de grupo:</b>												
Paciente (116 niveles)												
sd(Intercepto)	0.20	(0.17, 0.23)	0.20	(0.17, 0.23)	0.20	(0.17, 0.23)	0.19	(0.17, 0.22)	0.19	(0.17, 0.22)	0.19	(0.17, 0.22)
sd(sangrado)							0.12	(0.02, 0.21)	0.13	(0.02, 0.23)		
Paciente:Diente (3131 niveles)												
sd(Intercepto)	0.23	(0.22, 0.24)	0.22	(0.21, 0.23)	0.22	(0.21, 0.23)	0.22	(0.21, 0.23)	0.22	(0.21, 0.23)	0.22	(0.21, 0.23)
sd(sangrado)							0.25	(0.16, 0.33)	0.25	(0.16, 0.34)		
<b>Efectos a nivel de población:</b>												
Intercepto	0.34	(0.30, 0.37)	0.33	(0.29, 0.37)	0.33	(0.29, 0.36)	0.29	(0.24, 0.34)	0.29	(0.25, 0.34)	0.29	(0.25, 0.34)
Sangrado			0.15	(0.11, 0.19)	0.15	(0.10, 0.19)	0.15	(0.10, 0.19)	0.13	(0.07, 0.20)	0.11	(0.02, 0.19)
Movilidad					0.04	(-0.00, 0.08)	0.03	(-0.01, 0.07)	0.03	(-0.01, 0.07)	0.03	(-0.01, 0.08)
Cálculo							0.10	(0.03, 0.18)	0.10	(0.03, 0.18)	0.10	(0.03, 0.18)
Fumar							-0.02	(-0.14, 0.09)	-0.02	(-0.14, 0.10)	-0.02	(-0.13, 0.09)
Sangrado:Cálculo									0.06			(-0.07, 0.19)
<b>Parámetros específicos:</b>												
shape	4.50	(4.41, 4.60)	4.51	(4.41, 4.61)	4.51	(4.41, 4.61)	4.51	(4.42, 4.61)	4.55	(4.45, 4.64)	4.54	(4.45, 4.64)
elpd_diff (se_diff)			-16.2	(8.2) <sup>†</sup>	-2.3	(2.0) <sup>‡</sup>	-3.4	(1.5) <sup>◇</sup>	-10.7	(7.9) <sup>‡</sup>	-1.3	(1.8) <sup>*</sup>

Coef: Coeficiente, \*ICr 95%: Intervalo Creíble del 95%. Comparaciones <sup>‡</sup> Modelo 3.1 vs 3.2, <sup>†</sup> Modelo 3.2 vs 3.3, <sup>◇</sup> Modelo 3.2 vs 3.4, <sup>‡</sup> Modelo 3.3 vs 3.4, <sup>‡</sup> Modelo 3.4 vs 3.5, \* Modelo 3.5 vs 3.6.

**Paso 4.** El modelo con variables de nivel tres, cálculo y fumar, es

$$\begin{aligned} \log(\mu_{ij|k}) = & \xi_{000} + \alpha_1 \text{sangrado}_{1ijk} + \xi_{010} \text{movilidad}_{1jk} \\ & + \xi_{001} \text{cálculo}_{1k} + \xi_{002} \text{fumar}_{2k} + (\nu_{00k} + u_{0jk}) \end{aligned} \quad (3.4)$$

El Cuadro 3.2 en las Columnas 8 y 9 muestra que el factor fumar no es significativo; sin embargo, el modelo que contiene el factor fumar es mejor que los demás. El modelo 3.4 es mejor que el modelo 3.3 (penúltimo renglón y Columna 9 en el Cuadro 3.2).

**Paso 5.** El modelo con pendiente aleatoria para la variable sangrado.

En la Ecuación 3.5 se agrega una pendiente aleatoria para la variable sangrado que varía a nivel de pacientes y dientes, es decir, la relación entre la profundidad de sondaje y el sangrado varió entre pacientes y entre dientes.

$$\begin{aligned} \log(\mu_{ij|k}) = & \xi_{000} + \alpha_1 \text{sangrado}_{1ijk} + \xi_{010} \text{movilidad}_{1jk} \\ & + \xi_{001} \text{cálculo}_{1k} + \xi_{002} \text{fumar}_{2k} \\ & + \nu_{10k} \text{sangrado}_{1ijk} + u_{1jk} \text{sangrado}_{1ijk} + (\nu_{00k} + u_{0jk}) \end{aligned} \quad (3.5)$$

Las Columnas 10 y 11 del Cuadro 3.2 muestran que la pendiente aleatoria de sangrado es significativa a nivel de pacientes y dientes. Nuevamente, este modelo se comparó con el Modelo 3.4 mediante el criterio LOO-CV, y el mejor modelo es el Modelo 3.5 con pendiente aleatoria (penúltimo renglón y Columna 11 en el Cuadro 3.2).

Finalmente, en el siguiente modelo se agregaron términos de interacción basados en signos que ocurren en la enfermedad periodontal.

**Paso 7.** El modelo con interacciones entre niveles es

$$\begin{aligned} \log(\mu_{ij|k}) = & \xi_{000} + \alpha_1 \text{sangrado}_{1ijk} + \xi_{010} \text{movilidad}_{1jk} \\ & + \xi_{001} \text{cálculo}_{1k} + \xi_{002} \text{fumar}_{2k} + \xi_{101} \text{sangrado}_{1ijk} \text{cálculo}_{1k} \\ & + \nu_{10k} \text{sangrado}_{1ijk} + u_{1jk} \text{sangrado}_{1ijk} + (\nu_{00k} + u_{0jk}) \end{aligned} \quad (3.6)$$

La Ecuación 3.6 tiene una interacción entre el cálculo variable de nivel tres con el sangrado variable de nivel uno. Las Columnas 12 y 13 del Cuadro 3.2 muestran que la interacción no es significativa (su intervalo creíble contiene al cero). La comparación de

modelos indica que el mejor modelo es el Modelo 3.5 con pendiente aleatoria de sangrado (último renglón y Columna 13 en el Cuadro 3.2). Entonces, este modelo es interpretado.

La Figura 3.3 muestra la distribución posterior de los parámetros y las trazas de las MCMC del Modelo 3.5. Se muestra que las cadenas se mezclan bien lo que indica convergencia a la distribución posterior.

La Figura 3.4 muestra la evaluación predictiva posterior del Modelo 3.5 a los datos. Los datos replicados se representan en un color claro y los datos observados se representan en negro. Como ambas curvas concuerdan muy bien, la densidad predictiva posterior se ajusta muy bien a la distribución de la profundidad de sondaje. Ambas distribuciones claramente no son simétricas y parecen seguir una distribución gamma. Definitivamente, la distribución normal no era una suposición apropiada para la profundidad de sondaje. En conclusión, el Modelo 3.5 de pendiente aleatoria tuvo un buen ajuste.

## 3.2. Discusión

En este modelo para estudiar la profundidad de sondaje, la varianza a nivel de diente (1.59) y la varianza a nivel de paciente (1.49) fueron significativas (Cuadro 3.2), es decir, la media de la variable dependiente varió entre dientes anidados en pacientes. Este hallazgo prueba que es apropiado usar un modelo multinivel para estos datos de profundidad de sondaje. Además, la pendiente aleatoria de sangrado fue significativa entre dientes y pacientes, es decir, hubo una relación positiva entre la profundidad de sondaje y el sangrado que varió entre pacientes y entre dientes en los pacientes (si el sitio sangró, en promedio, la profundidad de sondaje entre dientes aumentó 1.28 mm y entre pacientes aumentó 1.13 mm). Por otro lado, el cálculo es una forma de placa dental endurecida. En el modelo de pendiente aleatoria (Modelo 3.5), el sangrado y el cálculo fueron parámetros significativos que estimaron que, en promedio, la profundidad de los sitios de sondaje sangrantes fue 1.14 mm mayor que los sitios que no mostraron sangrado y, en promedio, la profundidad de sondaje de los pacientes que tenían cálculo en cualquiera de los dientes fue 1.11 mm mayor que en los pacientes que no tenían cálculo. Los intervalos creíbles para sangrado y cálculo fueron (1.07,1.22) y (1.03,1.20), respectivamente.

En el modelo de pendiente aleatoria, la movilidad y el fumar no se asociaron signifi-

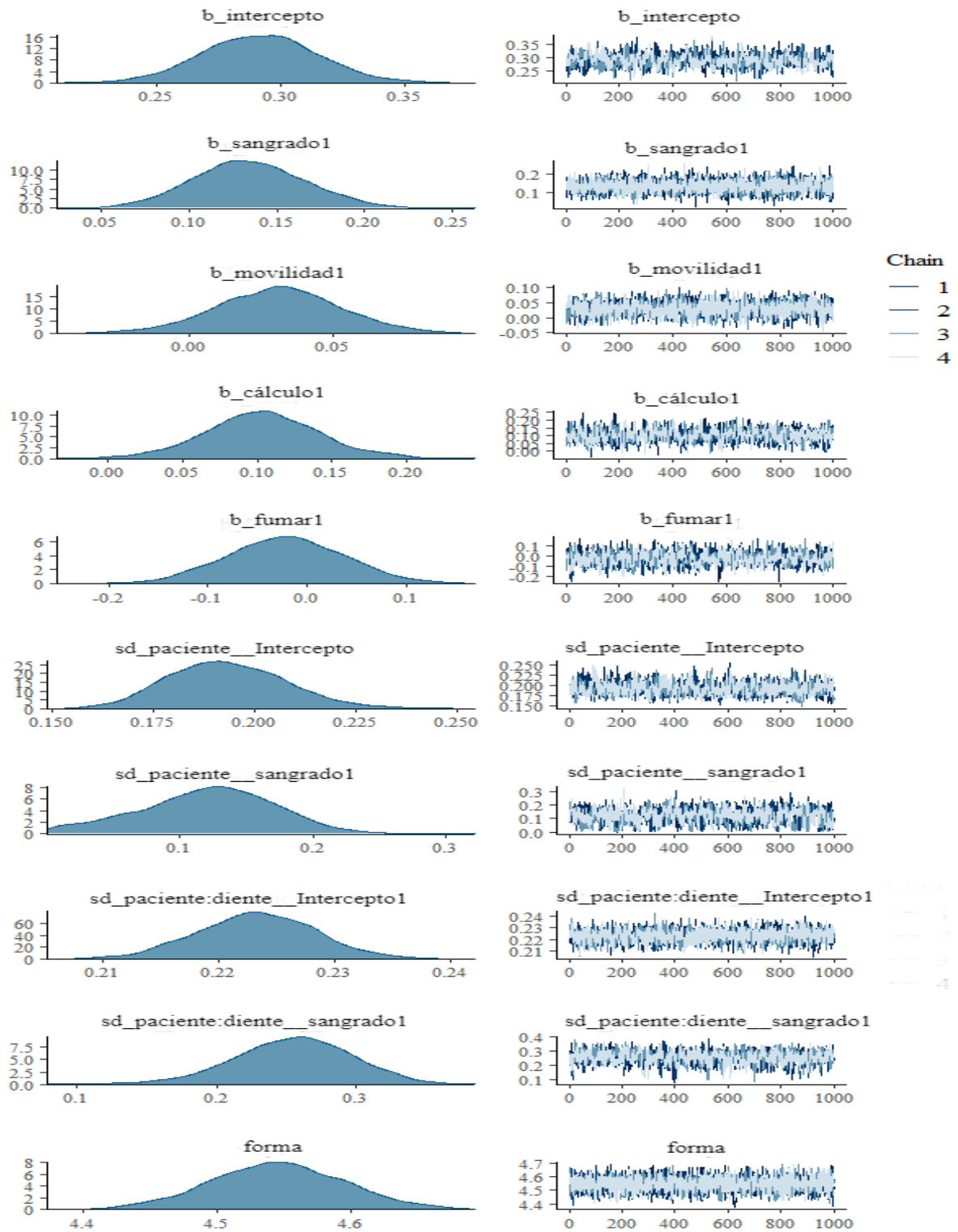


Figura 3.3: Distribución posterior de los parámetros y trazas de las MCMC del Modelo 3.5.

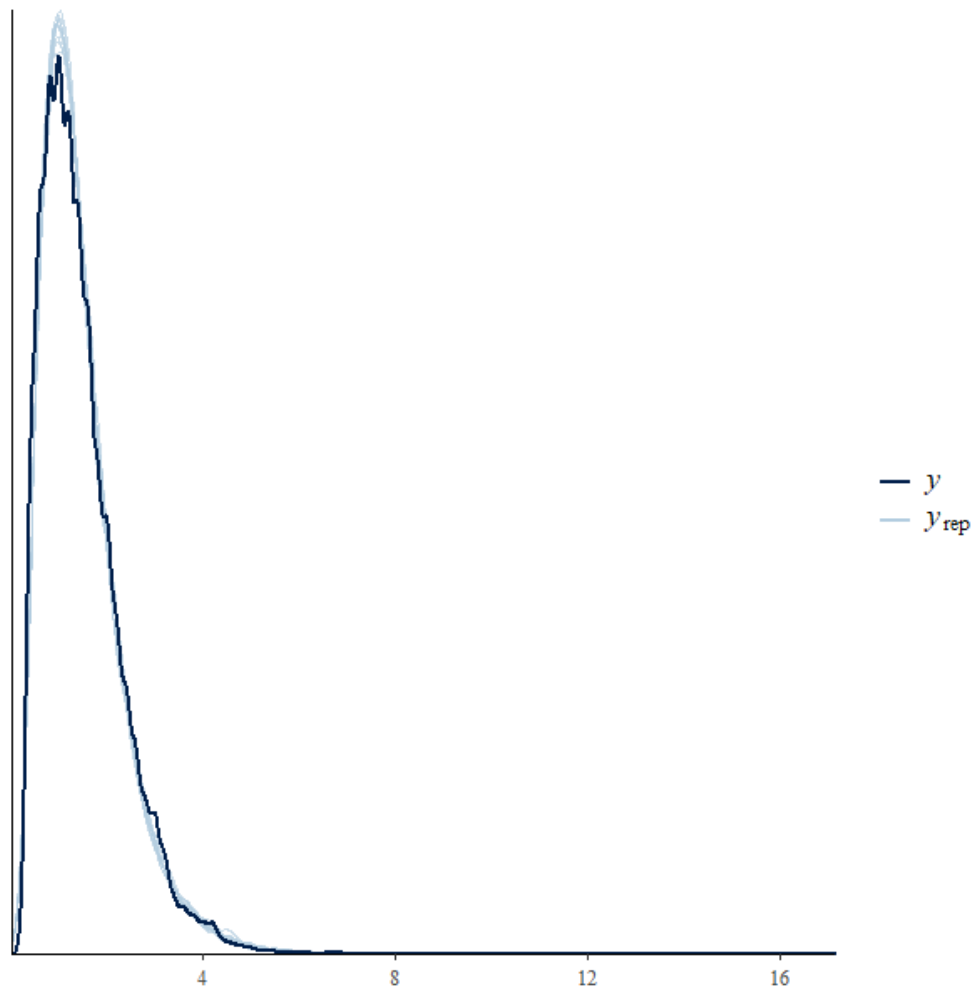


Figura 3.4: Densidad predictiva posterior del Modelo 3.5 con pendiente aleatoria para sangrado entre dientes y paciente.

cativamente con la profundidad de sondaje, pero si se decide dar una interpretación, se puede decir que, en promedio, la profundidad de sondaje de los pacientes que presentaron movilidad dentaria fue 1.03 mm mayor que la profundidad de sondaje de pacientes que no tenían movilidad dentaria. Del mismo modo, los pacientes fumadores tenían, en promedio, una profundidad de sondaje 0.98 mm mayor que los pacientes no fumadores. Se podrían obtener resultados e interpretaciones diferentes a partir de medir las variables independientes en niveles distintos a los dados en este análisis. Específicamente, la variable cálculo se puede medir a nivel diente. Se pueden encontrar más ejemplos prácticos usando el *Software R* en [31].

Cuadro 3.3: Selección de variable hacia delante para las variables de nivel tres\*.

Modelo	elpd_diff	se_diff
Comparación de modelos con una variable independiente de nivel tres		
$\xi_{001}$ cálculo <sub>1k</sub>	0.0	0.0
$\xi_{003}$ resistencia a la insulina <sub>3k</sub>	-0.1	1.6
$\xi_{004}$ restos radiculares <sub>4k</sub>	-0.8	1.5
$\xi_{005}$ placa <sub>5k</sub>	-1.5	1.6
$\xi_{002}$ fumar <sub>2k</sub>	-1.7	1.6
$\xi_{006}$ edad <sub>6k</sub>	-1.8	1.5
$\xi_{007}$ restauraciones desajustadas <sub>7k</sub>	-2.5	1.5
Comparación de modelos con dos variables independientes de nivel tres		
$\xi_{001}$ cálculo <sub>1k</sub> + $\xi_{002}$ fumar <sub>2k</sub>	0.0	0.0
$\xi_{001}$ cálculo <sub>1k</sub> + $\xi_{003}$ resistencia a la insulina <sub>3k</sub>	-2.1	1.5
$\xi_{001}$ cálculo <sub>1k</sub> + $\xi_{006}$ edad <sub>6k</sub>	-2.6	1.5
$\xi_{001}$ cálculo <sub>1k</sub> + $\xi_{005}$ placa <sub>5k</sub>	-2.9	1.5
$\xi_{001}$ cálculo <sub>1k</sub> + $\xi_{007}$ restauraciones desajustadas <sub>7k</sub>	-4.4	1.5
$\xi_{001}$ cálculo <sub>1k</sub> + $\xi_{004}$ restos radiculares <sub>4k</sub>	-5.2	1.5
Comparación de los mejores modelos con una y dos variables independientes de nivel tres		
$\xi_{001}$ cálculo <sub>1k</sub> + $\xi_{002}$ fumar <sub>2k</sub>	0.0	0.0
$\xi_{001}$ cálculo <sub>1k</sub>	-2.6	1.5
Comparación de modelos con tres variables independientes de nivel tres		
$\xi_{001}$ cálculo <sub>1k</sub> + $\xi_{002}$ fumar <sub>2k</sub> + $\xi_{004}$ restos radiculares <sub>4k</sub>	0.0	0.0
$\xi_{001}$ cálculo <sub>1k</sub> + $\xi_{002}$ fumar <sub>2k</sub> + $\xi_{006}$ edad <sub>6k</sub>	-0.2	1.6
$\xi_{001}$ cálculo <sub>1k</sub> + $\xi_{002}$ fumar <sub>2k</sub> + $\xi_{003}$ resistencia a la insulina <sub>3k</sub>	-0.6	1.5
$\xi_{001}$ cálculo <sub>1k</sub> + $\xi_{002}$ fumar <sub>2k</sub> + $\xi_{005}$ placa <sub>5k</sub>	-1.2	1.6
$\xi_{001}$ cálculo <sub>1k</sub> + $\xi_{002}$ fumar <sub>2k</sub> + $\xi_{007}$ restauraciones desajustadas <sub>7k</sub>	-1.6	1.5
Comparación de los mejores modelos con dos y tres variables independientes de nivel tres		
$\xi_{001}$ cálculo <sub>1k</sub> + $\xi_{002}$ fumar <sub>2k</sub>	0.0	0.0
$\xi_{001}$ cálculo <sub>1k</sub> + $\xi_{002}$ fumar <sub>2k</sub> + $\xi_{004}$ restos radiculares <sub>4k</sub>	-2.9	1.5

\* Todos los modelos tienen la estructura:  $\log(\mu_{ijk}) = \xi_{000} + \alpha_1 \text{sangrado}_{1ijk} + \xi_{010} \text{movilidad}_{1jk} + \text{var1} + \text{var2} + \text{var3} + (\nu_{00k} + u_{0jk})$  donde var1 es la variable independiente que produce el mejor ajuste entre los siete modelos con una variable independiente. Del mismo modo var2 es la segunda variable independiente que produce el mejor ajuste entre los seis modelos, teniendo var1 en común, con dos variables independientes, y así sucesivamente para var3.

### Conclusiones

---

En ciertos diseños de investigación clínica los datos tienen una estructura anidada (es decir, una estructura jerárquica). Los datos que componen una estructura anidada se modelan mediante modelos multinivel porque estiman simultáneamente los efectos de las variables a nivel individual y los efectos de las variables contextuales o variables a nivel de grupo.

Un CCI significativo determina si es necesario utilizar un modelo multinivel. Si el CCI no es significativo, un modelo de regresión ordinario es suficiente para modelar los datos anidados. Una desventaja de los modelos multinivel es que fácilmente contienen una gran cantidad de parámetros a estimar. Por otro lado, modelar los niveles de datos por separado incurre en un gran error de tipo 1 incluso cuando el CCI es pequeño. Este hecho hace que las inferencias sean incorrectas.

La estimación de máxima verosimilitud de los parámetros de un modelo multinivel requiere que se cumplan los supuestos de la distribución. Métodos más generales, como la estimación Bayesiana, permiten estimar los parámetros sin exigir que se satisfagan los supuestos de los modelos multinivel. Además, la estimación Bayesiana es robusta para un tamaño de muestra pequeño, situación que es más probable que ocurra en observaciones de mayor nivel y, en general, es capaz de lidiar con problemas técnicos como la multicolinealidad de los datos.

En este trabajo de tesis, se adaptó la estrategia bottom-up para especificar un modelo



multinivel en el enfoque frecuentista del enfoque Bayesiano. La propuesta fue utilizar el LOO-CV Bayesiano entre los diferentes pasos para la comparación de modelos. El criterio de información de devianza (DIC) también podría usarse en lugar de LOO-CV Bayesiano.

Dos factores tuvieron una asociación significativa con la profundidad de sondaje. Sangrado (variable independiente a nivel del sitio) y cálculo dental (variable independiente a nivel del paciente). A nivel del diente no se encontró un factor asociado a la profundidad de sondaje.

La metodología expuesta en esta tesis puede ser aplicada a otras áreas de las ciencias de la salud con datos con estructura jerárquica y variable de respuesta numérica.

---

## Bibliografía

---

- [1] Newman MG, Takei H, Klokkevold PR, Carranza FA. Newman and Carranza's clinical periodontology. 11th ed. Elsevier Health Sciences; 2011.
- [2] clinicacarlosmunguia. PERIODONCIA;. Available from: <https://www.clinicacarlosmunguia.com/especialidades/periodoncia/>.
- [3] Genco RJ, Borgnakke WS. Risk factors for periodontal disease. *Periodontology* 2000. 2013;62(1):59–94.
- [4] Kim JS, Kim DK, Hong SJ. Assessment of errors and misused statistics in dental research. *International Dental Journal*. 2011;61(3):163–167.
- [5] Wang J, Xie H, Fisher JF. *Multilevel Models: Applications using SAS*. Berlin, Boston: de Gruyter; 2011.
- [6] Hannigan A, Lynch CD. Statistical methodology in oral and dental research: pitfalls and recommendations. *Journal of Dentistry*. 2013;41(5):385–392.
- [7] Gilks WR, Richardson S, Spiegelhalter D. *Markov chain Monte Carlo in practice*. CRC press; 1995.
- [8] Romero Castro NS, Castro Alarcon N, Reyes Fernández S, Flores Alfaro E, Parra Rojas I. Periodontal disease distribution, risk factors, and importance of primary healthcare in the clinical parameters improvement. *International Journal of Odontostomat*. 2020;14(2):183–190.
- [9] Song J, Zhao H, Pan C, Li C, Liu J, Pan Y. Risk factors of chronic periodontitis on healing response: a multilevel modelling analysis. *BMC Medical Informatics and Decision Making*. 2017;17(1):1–9.

- [10] Mdala I, Haffajee AD, Socransky SS, de Blasio BF, Thoresen M, Olsen I, et al. Multilevel analysis of clinical parameters in chronic periodontitis after root planing/scaling, surgery, and systemic and local antibiotics: 2-year results. *Journal of Oral Microbiology*. 2012;4(1):17535.
- [11] Jordao LMR, Vasconcelos DN, da Silveira Moreira R, Freire MdCM. Individual and contextual determinants of periodontal health in 12-year-old schoolchildren in a Brazilian capital city. *International Journal of Dentistry*. 2012;2012.
- [12] Müller HP. Dealing with hierarchical data in periodontal research. *Clinical Oral Investigations*. 2009;13(3):273–278.
- [13] Tomasi C, Leyland AH, Wennström JL. Factors influencing the outcome of non-surgical periodontal treatment: a multilevel approach. *Journal of Clinical Periodontology*. 2007;34(8):682–690.
- [14] D’Aiuto F, Ready D, Parkar M, Tonetti MS. Relative contribution of patient-, tooth-, and site-associated variability on the clinical outcomes of subgingival debridement. I. Probing depths. *Journal of Periodontology*. 2005;76(3):398–405.
- [15] Axtelius B, Söderfeldt B, Attström R. A multilevel analysis of factors affecting pocket probing depth in patients responding differently to periodontal treatment. *Journal of Clinical Periodontology*. 1999;26(2):67–76.
- [16] Albandar JM, Goldstein H. Multi-level statistical models in studies of periodontal diseases. *Journal of Periodontology*. 1992;63(8):690–695.
- [17] Beek Mvd, Hoeksma J, Prah-Andersen B. Vertical facial growth: a longitudinal study from 7 to 14 years of age. *The European Journal of Orthodontics*. 1991;13(3):202–208.
- [18] Kopperud SE, Tveit AB, Gaarden T, Sandvik L, Espelid I. Longevity of posterior dental restorations and reasons for failure. *European Journal of Oral Sciences*. 2012;120(6):539–548.
- [19] Barrera LM, Buschang PH, Throckmorton GS, Roldán SI. Mixed longitudinal evaluation of masticatory performance in children 6 to 17 years of age. *American Journal of Orthodontics and Dentofacial Orthopedics*. 2011;139(5):e427–e434.
- [20] Hox JJ, Moerbeek M, Van de Schoot R. *Multilevel analysis: Techniques and applications*. 3rd ed. Routledge; 2017.

- [21] Raudenbush SW, Bryk AS. Hierarchical linear models: Applications and data analysis methods. vol. 1. sage; 2002.
- [22] Ng SK, Krishnan T, McLachlan GJ. The EM algorithm. In: Handbook of Computational Statistics. Springer; 2012. p. 139–172.
- [23] Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. Bayesian data analysis. 1st ed. New York: CRC press; 2013.
- [24] Hoffman MD, Gelman A. The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. Journal of Machine Learning Research. 2014;15(1):1593–1623.
- [25] Carpenter B, Gelman A, Hoffman MD, Lee D, Goodrich B, Betancourt M, et al. Stan: A probabilistic programming language. Journal of Statistical Software. 2017;76(1):1–32.
- [26] Bürkner PC. brms: An R package for Bayesian multilevel models using Stan. Journal of Statistical Software. 2017;80(1):1–28.
- [27] Roy V. Convergence diagnostics for markov chain monte carlo. Annual Review of Statistics and Its Application. 2020;7:387–412.
- [28] Plummer M, Best N, Cowles K, Vines K. CODA: Convergence Diagnosis and Output Analysis for MCMC. R News. 2006;6(1):7–11. Available from: <https://journal.r-project.org/archive/>.
- [29] R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria; 2021. Available from: <https://www.R-project.org/>.
- [30] Bürkner PC. Advanced Bayesian multilevel modeling with the R package brms. arXiv preprint arXiv:170511123. 2017;.
- [31] Finch WH, Bolin JE, Kelley K. Multilevel modeling using R. 2nd ed. Crc Press; 2019.

# Anexo 1

## .1. Análisis univariado

La profundidad promedio de los sitios fue 1.24 mm. El mínimo de las profundidades fue 0.2 mm. y el máximo fue 9 mm. El 25% de las profundidades fue menor o igual 0.8 mm, el 50% menor o igual a 1.2mm. y el 75% de las profundidades fue menor o igual a 1.8 mm.

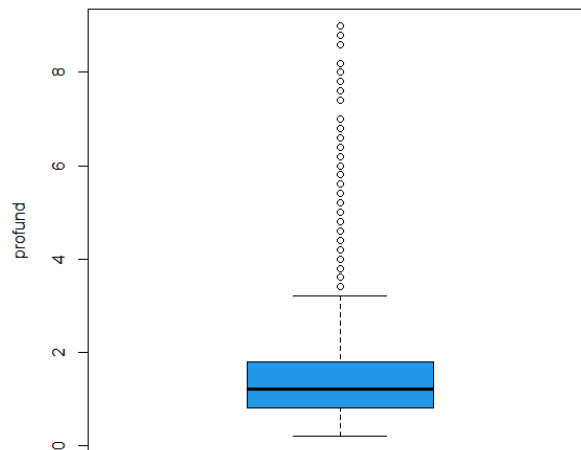


Figura 1: Profundidad de sondaje

El 97% (18,229) sitios no sangraron durante el sondaje periodontal y el 3% (551) de los sitios si presentaron sangrado en el sondaje.

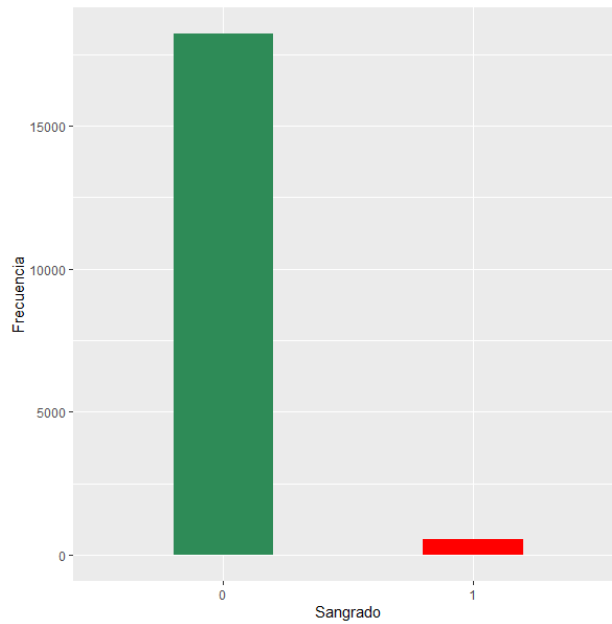


Figura 2: Sangrado durante el sondaje

El 99% (18,583) de los pacientes no presentaron grado de furcación y un 1% (197) si tuvo algún grado de furcación.

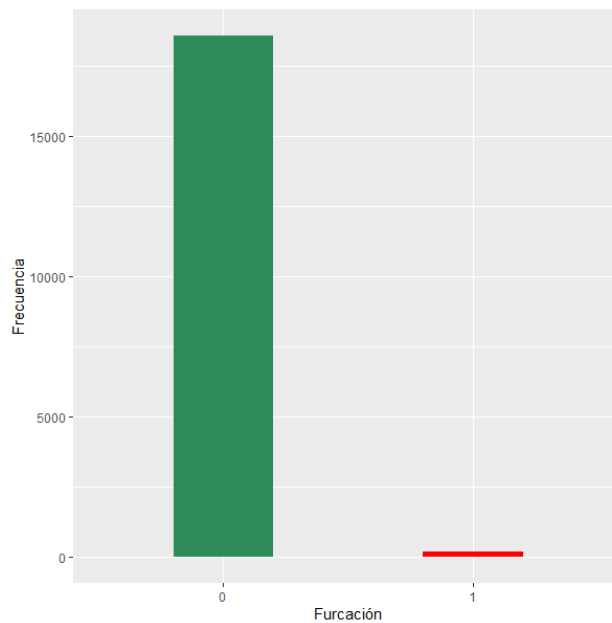


Figura 3: Furcación dental

El 92% (2,882) de dientes no presentaron ningún grado de movilidad y el 8% (249) tuvo un grado de movilidad dental

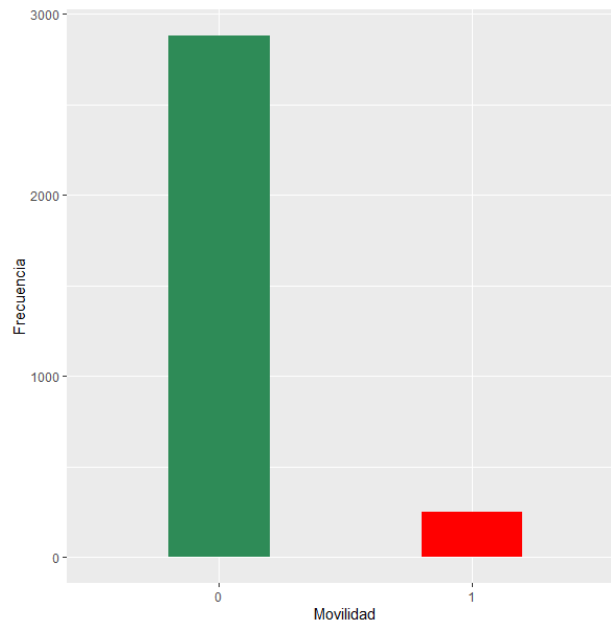


Figura 4: Movilidad dental

La edad promedio de los pacientes fue 40 años. La edad mínima fue 18 y la máxima fue 75. El 25% de las edades fue menor o igual 24 años, el 50% menor o igual a 43 y el 75% de las edades fue menor o igual a 50 años.

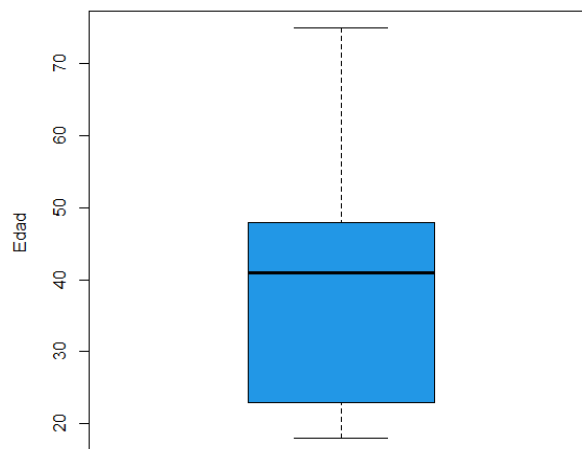


Figura 5: Edad

El 25% (29) de pacientes no presentaron placa bacteriana y el 75% (87) si presentaron placa bacteriana.

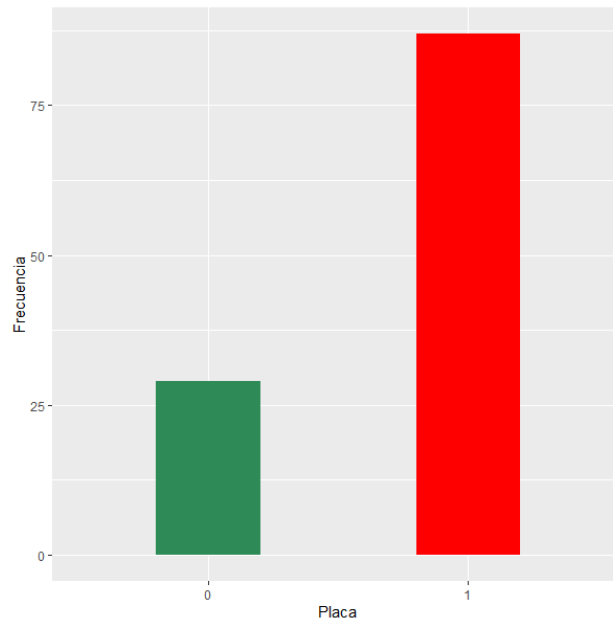


Figura 6: Placa

El 63% (73) de pacientes no presentaron cálculo dental y el 37% (43) si presentaron cálculo dental.

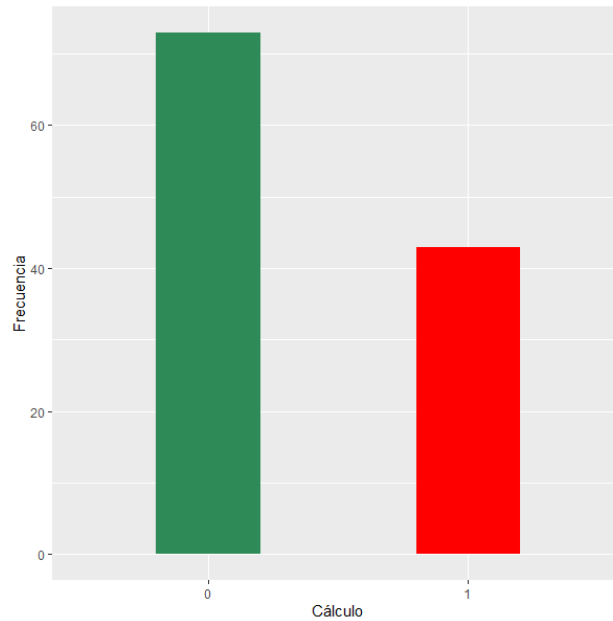


Figura 7: Cálculo dental

El 82% (95) de pacientes no presentaron resistencia a la insulina y el 18% (21) si presentaron resistencia a la insulina.



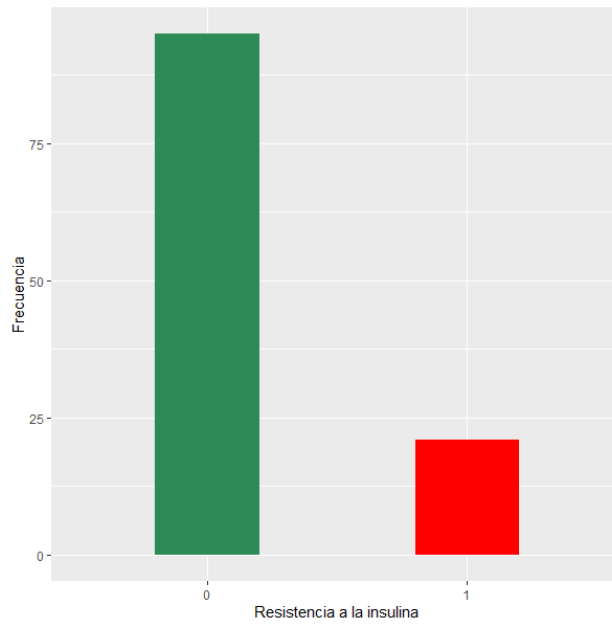


Figura 8: Resistencia a la insulina

El 89 % (103) de los pacientes no fuman y el 11 % (13) si fuman.

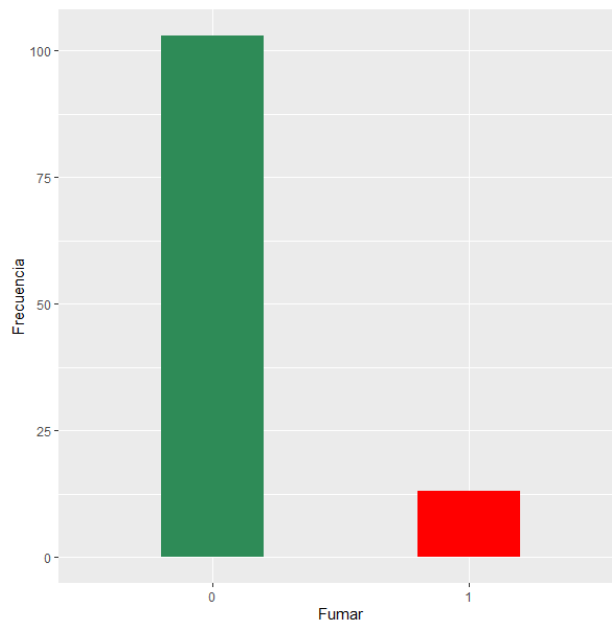


Figura 9: Fumar

El 81 % (94) de los pacientes no presentaron restos radiculares y el 19 % (22) si presentaron restos radiculares.

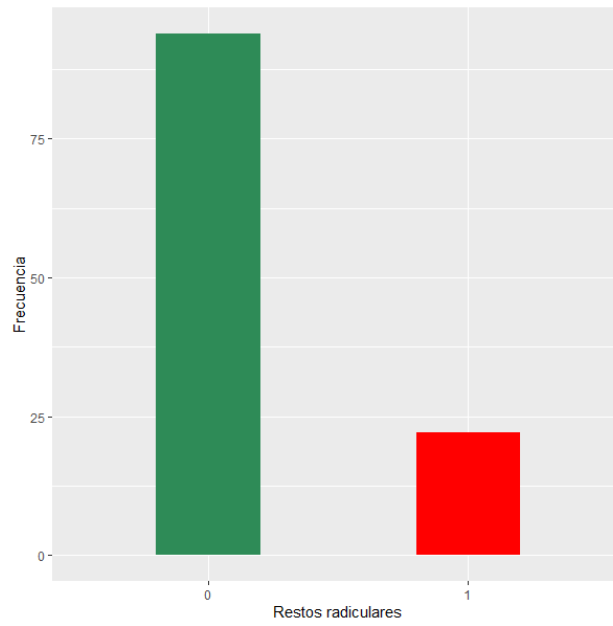


Figura 10: Restos radiculares

El 87% (101) de los pacientes no tuvieron restauraciones desajustadas y el 13% (15) si tuvieron restauraciones desajustadas.

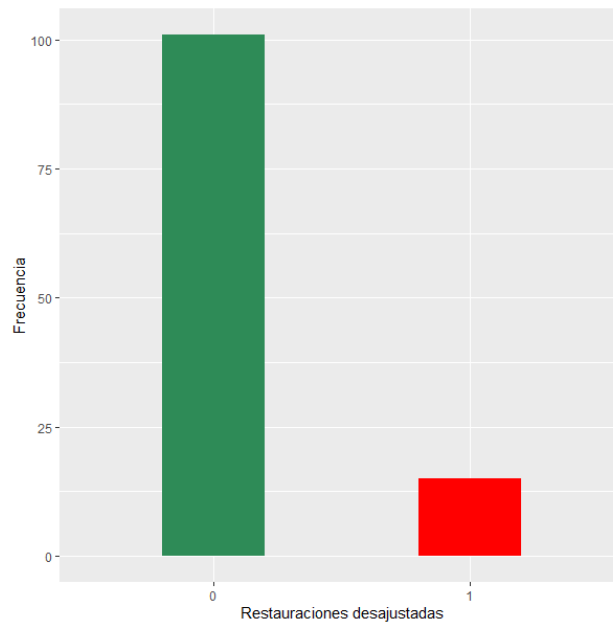


Figura 11: Restauraciones desajustadas

## .2. Glosario de periodoncia

### B

*Bolsa gingival:* Surco que se forma entre el diente y la encía que aumenta su profundidad a medida que se acumulan bacterias debajo de la encía.

### C

*cálculo dental:* También conocido como sarro, es un depósito compuesto mayormente por sales de calcio y fósforo y minerales que se adhiere a las superficies dentales. Es originado por la mineralización de la placa bacteriana.

*curataje dental:* También conocido como raspado y alisado radicular. Es un procedimiento llevado a cabo para acceder a las bolsas periodontales y limpiarlas.

### E

*Exudado:* Líquido extravasado en una inflamación por alteración de la permeabilidad vascular, rico en elementos del plasma sanguíneo, incluyendo elementos formes (eritrocitos).

### F

*Furca dental:* Zona anatómica que se encuentra delimitada por las raíces de dientes multiradiculares.

*Furcación:* Lesión producida en la furca dental.

### R

*Recesión gingival:* Es el desplazamiento del margen gingival apical a la unión cemento - esmalte que expone la superficie radicular.

*Restauraciones dentales:* Rellenos con los que se cubre cualquier cavidad dental, carillas o coronas para restaurar el diente.

*Restos radiculares:* Son fragmentos de raíz dental de dientes fracturados que se encuentran en la cavidad oral sin realizar una función. También pueden ser originados por una condición de caries.