

**UNIVERSIDAD DE EL SALVADOR**  
**FACULTAD DE CIENCIAS NATURALES Y MATEMÁTICA**  
**ESCUELA DE MATEMÁTICA**



**TESIS:**

**“CLASIFICACIÓN DE LA CONDICIÓN DE SALUD BUCAL UTILIZANDO  
REGRESIÓN LOGÍSTICA MULTINOMIAL CON LA POBLACIÓN  
SALVADOREÑA QUE ASISTIÓ A LA RED DE SALUD PÚBLICA EN EL AÑO  
2009”**

**PRESENTADO POR:**

**Br. AIDA MARICELA GÓMEZ RAMOS**

**Br. KARINA EVELYN CORNEJO RAMOS**

**PARA OPTAR AL GRADO DE:**

**LICENCIATURA EN ESTADÍSTICA**

**CIUDAD UNIVERSITARIA, DICIEMBRE 2014**

**UNIVERSIDAD DE EL SALVADOR**  
**FACULTAD DE CIENCIAS NATURALES Y MATEMÁTICA**  
**ESCUELA DE MATEMÁTICA**



**TESIS:**

**“CLASIFICACIÓN DE LA CONDICIÓN DE SALUD BUCAL UTILIZANDO  
REGRESIÓN LOGÍSTICA MULTINOMIAL CON LA POBLACION  
SALVADOREÑA QUE ASISTIÓ A LA RED DE SALUD PÚBLICA EN EL AÑO  
2009”**

**PRESENTADO POR:**

**Br. AIDA MARICELA GÓMEZ RAMOS**

**Br. KARINA EVELYN CORNEJO RAMOS**

**ASESORES**

**MSc. RENÉ ARMANDO PEÑ AGUILAR**  
**ESCUELA DE MATEMÁTICA**

**DRA. RUTH FERNANDEZ DE QUEZADA**  
**FACULTAD DE ODONTOLOGÍA**

**CIUDAD UNIVERSITARIA, DICIEMBRE 2014**

## **AUTORIDADES**

### **UNIVERSIDAD DE EL SALVADOR**

RECTOR UNIVERSITARIO: ING.MARIO ROBERTO NIETO LOVO

SECRETARIA GENERAL: DRA. ANA LETICIA ZA VALETA DE AMAYA

### **FACULTAD DE CIENCIAS NATURALES Y MATEMÁTICA**

DECANO: MSC. MARTÍN ENRIQUE GUERRA CÁCERES

SECRETARIO: LIC. CARLOS QUINTANILLA

### **ESCUELA DE MATEMÁTICA**

DIRECTOR: DR. JOSÉ NERYS FUNES TORRES

SECRETARIA: ALBA IDALIA CÓRDOVA CUÉLLAR

**UNIVERSIDAD DE EL SALVADOR**  
**FACULTAD DE CIENCIAS NATURALES Y MATEMÁTICA**  
**ESCUELA DE MATEMÁTICA**

**ASESORES:**

---

**MSc. RENÉ ARMANDO PEÑA AGUILAR**  
**ESCUELA DE MATEMÁTICA**

---

**DRA. RUTH FERNANDEZ DE QUEZADA**  
**FACULTAD DE ODONTOLOGÍA**

**CIUDAD UNIVERSITARIA, DICIEMBRE 2014**

## **AGRADECIMIENTO**

### **Karina Evelyn Cornejo Ramos,**

Le agradezco a Dios por haberme acompañado y guiado a lo largo de mi carrera, por ser mi fortaleza en los momentos de debilidad y por brindarme una vida llena de aprendizajes, experiencias y sobre todo felicidad.

Le doy gracias a mi Madre Amalia por apoyarme en todo momento, por los valores que me ha inculcado, y por haberme dado la oportunidad de tener una excelente educación en el transcurso de mi vida. Sobre todo por ser un excelente ejemplo de vida a seguir.

Le agradezco la confianza, apoyo y dedicación de tiempo a mis Asesores de tesis: Dra. Ruth De Quezada y a Msc. René Peña, por haber compartido conmigo sus conocimientos y sobre todo su amistad.

Finalmente te agradezco Aida por haber sido una excelente compañera de tesis y amiga, por haberme tenido la paciencia necesaria y por motivarme a seguir adelante en los momentos de desesperación y sobre todo por hacer de su familia, una familia para mí.

### **Aída Maricela Gómez Ramos,**

Primeramente a Dios agradezco la culminación de este proyecto ya que sin su fortaleza, provisión, sabiduría y guía sin duda no hubiera sido posible.

Agradezco además a mis asesores Dra. Ruth Fernández de Quezada y MSc. René Armando Peña Aguilar de quienes no puedo sentir más que un profundo respeto y admiración por su entrega y dedicación con la que emprendieron este reto que sin considerar sus ocupaciones aceptaron de forma desinteresada brindar la dirección de este proceso, a mi compañera y amiga Karina por permitirme trabajar con ella y por su apoyo.

Y especialmente agradezco a la persona que aun sin su presencia física cerca de mí sigue sembrando cosas buenas para mi vida, mi madre, Vilma Ramos de Gómez (Q.E.P.D) pues fue quien inculco desde las simientes de mi carrera el espíritu luchador y a no darme por vencida aun con las adversidades que la vida me presente, mostró un fiel ejemplo. Así también a mi hijo Aldo Efraín Martínez agradezco su apoyo que recibí de una manera inconciente cuando con sutileza comprendia que no era el momento de jugar o recibir una atención de mamá, a mi esposo René Osmin Martínez y a mi hermana Paulina Gómez por convertirse en mi incondicional apoyo. Asimismo agradezco a todos mis demás familiares y amigos que en algún momento me brindaron palabras de confianza y animo.

## **DEDICATORIA**

A mi madre.

Son únicos, los momentos que se comparten con una madre,  
gracias mamá, por darme tanto

**Karina Evelyn Cornejo Ramos**

A Dios y a la memoria de mi madre.

**Aída Maricela Gómez Ramos**

## ÍNDICE

<b>I. INTRODUCCIÓN .....</b>	<b>ix</b>
<b>II. PLANTEAMIENTO DEL PROBLEMA .....</b>	<b>xii</b>
<b>III. OBJETIVOS.....</b>	<b>xiii</b>
<b>IV. JUSTIFICACIÓN.....</b>	<b>xiv</b>
<b>CAPÍTULO I: FUNDAMENTOS TEÓRICOS. ....</b>	<b>1</b>
<b>1.1 DEFINICIONES Y CÁLCULOS DE ÍNDICES ODONTOLÓGICOS .....</b>	<b>5</b>
1.1.1. ÍNDICE DE PLACA BACTERIANA LÖE Y SILNESS. (IP).....	5
1.1.2. ÍNDICE GINGIVAL DE LÖE Y SILNESS (IG).....	7
1.1.3. ÍNDICE PERIODONTAL COMUNITARIO CPI (DE OMS) .....	8
1.1.4. ÍNDICE DE ESTÉTICA DENTAL (DAI).....	9
1.1.5. ÍNDICE DE CARIES CPO-D .....	12
<b>1.2. PRUEBA PARA VARIAS PROPORCIONES. ....</b>	<b>13</b>
<b>1.3. FUNDAMENTOS TEORICOS SOBRE MODELOS LINEALES GENERALIZADOS Y DEL MODELO DE REGRESIÓN LOGISTICA MULTINOMIAL. ....</b>	<b>17</b>
1.3.1. EL MODELADO ESTADISTICO .....	17
1.3.2. LOS MODELOS LINEALES GENERALIZADOS (MLGs) .....	20
1.3.3. LA ESTRUCTURA DEL MODELO .....	20
1.3.4. EL MODELO DE REGRESIÓN LOGISTICA MULTINOMIAL .....	22
1.3.4.1. COEFICIENTE DE PROBABILIDAD Y PROBABILIDAD COMPLEMENTARIA (ODDS) .....	30
1.3.4.2. RAZÓN DE ODDS .....	33
1.3.5. INTERPRETACIÓN DEL MODELO DE REGRESIÓN LOGISTICA MULTINOMIAL .....	34
1.3.5.1. VARIABLE INDEPENDIENTE DICOTÓMICA .....	34
1.3.5.2. VARIABLE INDEPENDIENTE POLITÓMICA .....	35
1.3.5.3. VARIABLE INDEPENDIENTE CUANTITATIVA .....	36
1.3.5.4. INTERACCIÓN .....	39
1.3.6. ESTIMACIÓN DE LOS PARÁMETROS EN UN MODELO DE REGRESIÓN LOGÍSTICA MULTINOMIAL POR MÁXIMA VEROSIMILITUD. ....	40
1.3.6.1 EL MÉTODO DE NEWTON-RAPHSON.....	45
1.3.7. SIGNIFICACIÓN DE LOS PARÁMETROS.....	46
1.3.7.1. CONTRASTES CONDICIONALES DE LA RAZÓN DE VEROSIMILITUD .....	47
1.3.7.2. EL TEST DE WALD .....	47
1.3.8. INFERENCIA EN REGRESIÓN LOGÍSTICA MULTINOMIAL.....	48
1.3.8.1. INTERVALOS DE CONFIANZA .....	49

1.3.9. BONDAD DE AJUSTE DEL MODELO .....	50
1.3.9.1. JI – CUADRADO DE PEARSON .....	51
1.3.9.2. TEST CHI-CUADRADO DE RAZÓN DE VEROSIMILITUDES, ESTADÍSTICO DE WILKS. DESVIANZA. ....	51
1.3.10. CAPACIDAD PREDICTIVA DEL MODELO .....	53
1.3.10.1. COEFICIENTE PSEUDO - $R^2$ DE MC-FADDEN. ....	53
1.3.10.2. COEFICIENTE PSEUDO - $R^2$ DE COX – SNELL.....	54
1.3.10.3 COEFICIENTE PSEUDO - $R^2$ DE NAGELKERKE.....	55
1.3.11. TASA DE CLASIFICACIONES CORRECTAS. ....	55
1.3.12. MÉTODOS DE SELECCIÓN DEL MODELO.....	56
1.3.12.1. SELECCIÓN HACIA ADELANTE.....	56
1.3.12.2. HACIA ATRÁS.....	57
1.3.12.3. STEPWISE (PASO A PASO) .....	57
<b>2.1. ANÁLISIS ESTADÍSTICO EXPLORATORIO. ....</b>	<b>59</b>
2.1.1. ANÁLISIS ESTADÍSTICO EXPLORATORIO DE LAS VARIABLES SOCIODEMOGRÁFICAS.....	60
2.1.2. ANÁLISIS DE LAS VARIABLES DE LOS ÍNDICES ODONTOLÓGICOS.....	63
<b>2.2. ANÁLISIS BIDIMENCIONAL.....</b>	<b>71</b>
<b>2.3. COMPARACIÓN DE VARIAS PROPORCIONES .....</b>	<b>75</b>
<b>2.4. APLICACIÓN DE REGRESIÓN LOGÍSTICA MULTINOMIAL .....</b>	<b>95</b>
2.4.1. PARTICIÓN DE LA MUESTRA. ....	97
2.4.2. ESTIMACIÓN DEL MODELO.....	98
<b>2.4.3. PARÁMETROS DEL MODELO .....</b>	<b>102</b>
2.4.4. BONDAD DEL AJUSTE.....	106
2.4.5. CALIDAD DEL AJUSTE DEL MODELO. ....	108
2.4.6. VALIDACIÓN DEL MODELO .....	109
<b>CONCLUSIONES .....</b>	<b>113</b>
<b>RECOMENDACIONES.....</b>	<b>116</b>
<b>REFERENCIAS BIBLIOGRÁFICAS .....</b>	<b>117</b>

## I.INTRODUCCIÓN

Al hablar de salud bucal se descubre que existen factores sociodemográficos que están directamente relacionados con la prevalencia e incidencia del surgimiento de las enfermedades bucales como son la edad, el sexo, el nivel de escolaridad, la ocupación, entre otros; de ahí que la medición del estado actual de la salud bucal se establece a través de indicadores científicos actualizados con un enfoque clínico y socioeconómico. La prevalencia de las enfermedades bucodentales podría ser más significativa con el aumento de la edad; debido a cambios en la dentición, cambios en la dieta diaria, el ritmo de vida más acelerado, las actividades sociales de moda, enfermedades graves y de rápida evolución y el stress potenciarían el incremento de tales problemáticas bucales [5].

Tal como lo muestra el estudio de La Fundación para el Crecimiento y Desarrollo de la Población Venezolana (FUNDACREDESA), se llevó a cabo una investigación sobre la condición de crecimiento y desarrollo llamada Proyecto Venezuela. En este estudio se consideró la salud bucal como componente fundamental de la salud del individuo. Se evaluó la caries dental en la dentición temporal y se consideraron las edades de 3, 7, 9 y 12 años por su grado de importancia. En la dentición permanente se aplicó el mismo criterio, resultando las edades de 7, 12, 18 y el grupo de 25 y más años los más afectados. Al igual en el 2008; en El Salvador, el Estudio Epidemiológico de Caries Dental y Fluorosis en Escolares de 5-6, 7-8, 12 y 15 años de Centros de Enseñanza Pública y Privada, determina la prevalencia de caries en escolares de 7 y 8 años es del 97.9%, y el CPO-D encontrado a los 12 años es de 1.45, ubicándose en niveles bajos de caries dental según ponderaciones de la OMS [23].

Por lo que esta investigación se enfocó en la aplicación de un modelo de regresión logística multinomial para la determinación de la probabilidad que tiene un individuo de tener cierto grado de salud bucal. Se utilizó el modelo de regresión logístico debido a que la variable dependiente es cualitativa la cual no es adecuada usarla para aplicar un análisis de regresión lineal; así mismo, desde su creación, los modelos de regresión

logísticos se han utilizado en el análisis de datos procedentes de investigaciones propias del ámbito de la ciencia de la salud [21].

La regresión logística es uno de los instrumentos estadísticos más expresivos y versátiles de que se dispone para el análisis de datos en clínica y epidemiología, su origen se remonta a la década de los sesenta, con el trascendente trabajo de CORNFIELD, GORDON Y SMITH 1961 sobre el riesgo de padecer una enfermedad coronaria y, ya en la forma en la que se conoce actualmente, con la contribución de WALKER Y DUNCAN (1967) en que se aborda el tema de estimar la probabilidad de ocurrencia de cierto acontecimiento en función de varias variables [21]. Ya que resulta conveniente, aplicar este método que permitirá llevar a cabo el debido tratamiento a la base de datos de la “Investigación de Diagnóstico Bucal de El Salvador” la cual fue desarrollada por la Facultad de Odontología de la Universidad de El Salvador con estudiantes en servicio social durante el año 2009.

De acuerdo a las características de los datos recolectados dan lugar a la búsqueda de un modelo de regresión logística; de este modo el objetivo principal será clasificar la condición de salud bucal: sana, daño leve, daño moderado y daño severo de acuerdo a los niveles de los índices CPO, Placa bacteriana (higiene oral), Gingivitis, Periodontitis y de Estética (DAI); los cuales son factores incidentes para analizar la condición bucal, así también se tendrán en cuenta las variables sociodemográficas de los pacientes que asistieron a las unidades de salud a nivel nacional en el año 2009.

Por otro lado, lo que se procura mediante la regresión logística es, en principio, expresar la probabilidad de que ocurran las condiciones: sano, daño leve, daño moderado y daño severo en función de ciertas variables que se presumen relevantes o influyentes. Al construir el modelo de regresión logístico las variables explicativas pueden ser de cualquier naturaleza: dicotómicas, ordinales, continuas o nominales, esta flexibilidad en cuanto a la información de entrada constituye uno de sus mayores atractivos.

Para el desarrollo de esta investigación fue precisa la siguiente organización:

En el Capítulo I, se describen los fundamentos teóricos necesarios sobre la construcción del modelo de regresión logística multinomial, es decir, una exposición de las principales herramientas estadísticas necesarias, y el contraste de hipótesis que se plantean en la investigación. También fué necesario realizar una aplicación sobre la comparación de varias proporciones para poder contrastar si existen diferencias en cuanto al género de los pacientes es decir el hecho de ser hombre o mujer puede determinar la condición de salud bucal, al igual que interesa conocer si según el ingreso económico mensual, según el nivel educativo, según la región de residencia son factores de igual manera influyentes en cuanto a condición de salud bucal.

Capítulo II, en este se presenta el análisis descriptivo, bidimensional y comparativo de las variables consideradas en el estudio, además se expone la aplicación del modelo de regresión logística multinomial donde muestra la estimación de los parámetros, interpretación del modelo, su bondad y calidad de ajuste y la validación del modelo encontrado. Y por último, se presentan las conclusiones de los principales resultados obtenidos. Para realizar el análisis estadístico que se presenta en esta investigación se utilizó el software estadístico SPSS 18 (PASWS Statistics 18).

## II. PLANTEAMIENTO DEL PROBLEMA

El método de regresión logística multinomial se utiliza para estimar las probabilidades de que un individuo pertenece a un determinado grupo, dado que las características de ese individuo han tomado ciertos valores concretos; de manera que la variable dependiente que dará respuesta a este problema se categorizará de la siguiente manera en los grupos: Condición sana, Condición leve, Condición moderada y Condición severa según sea su probabilidad de pertenencia. Este método permitirá llevar a cabo el debido tratamiento a la base de datos de la investigación de diagnóstico bucal de El Salvador la cual fue obtenida por la facultad de Odontología de la Universidad de El Salvador por estudiantes en Servicio Social durante el año 2009. En dicha base se ha recolectado información referente a la salud bucal de la población salvadoreña que hace uso de la atención brindada por las unidades de salud en El Salvador; así también información sociodemográfica de cada individuo.

En El Salvador se han elaborado pocas investigaciones sobre la salud bucal del país, esta investigación realizada por estudiantes en servicio social de la carrera “Doctorado en cirugía Dental” de la Universidad de El Salvador, ubicados en distintos municipios de los siguientes departamentos: Ahuachapán, San Salvador, Santa Ana, San Vicente, La paz, La Libertad, Cuscatlán, Cabañas, Morazán, Usulután y La Unión. Mostrando que el estado de la salud bucal de la población salvadoreña es preocupante debido a los altos valores que han tomado los diferentes índices bucales, por lo que ha sido de interés para esta investigación el hecho de indagar y confirmar la relación existente entre la edad, sexo, lugar de residencia, los ingresos percibidos mensualmente, entre otros factores y la prevalencia de las enfermedades bucales más comunes para así poder realizar una clasificación de la condición de salud bucal mediante el ajuste de un modelo de regresión logístico multinomial adecuado.

### **III.OBJETIVOS**

#### Objetivo General

Formular un modelo de regresión logística multinomial e identificar los factores determinantes que permitan clasificar la condición de salud bucal de los pacientes adultos que asistieron a la red de salud pública en el año 2009.

#### Objetivos específicos

- a) Determinar si el género de los pacientes influye sobre la condición de salud bucal.
- b) Determinar si el ingreso económico de los pacientes es determinante en cuanto la condición de su salud bucal.
- c) Determinar si el nivel de escolaridad alcanzado de los pacientes influye sobre la condición de su salud bucal.
- d) Determinar si la zona de residencia influye sobre la condición de su salud bucal.
- e) Determinar si el ingreso económico del paciente influye en la cantidad de dientes obturados.
- f) Determinar si el ingreso económico del paciente influye en la cantidad de dientes cariados.
- g) Determinar si el ingreso económico del paciente influye en la cantidad de dientes perdidos.
- h) Determinar si la edad del paciente influye en la cantidad de dientes perdidos.

#### IV.JUSTIFICACIÓN

Conscientes de la importancia y la necesidad que existe en el país que se realicen investigaciones con este tipo de aplicaciones con bases científicas y matemáticas que permitan dar a conocer las necesidades de cada individuo y que reciban las acciones preventivas y curativas según su condición de salud bucal, ya que las enfermedades bucales si no son tratadas a tiempo pueden llegar a agravar la condición del paciente y llegar a un estado de pérdidas de dientes afectando la función bucal, así como también su presentación estética. Para determinar el modelo fue necesario observar el comportamiento de los datos y el análisis de regresión logística multinomial es un método estadístico que permitió clasificar un individuo a un determinado grupo en función de varias variables relacionadas con el mismo. Cabe mencionar que en la actualidad el análisis de regresión logística es un método que está siendo utilizado con mayor frecuencia en diversos trabajos de investigación biomédica así como también en la epidemiología [21]. Un caso particular la realidad del país debido a que es muy poca la aplicación de este tipo de estudios relacionados con esta técnica estadística sobre todo para las áreas de las ciencias de la salud. El modelo de regresión logística multinomial aplicado en diferentes temáticas abordado por profesionales y estudiantes de la escuela de matemática de la Universidad de El Salvador, en la Maestría en Estadística, en el área de Modelos Lineales Generalizados, se han realizado los siguientes trabajos de investigación:

- 1.“Construcción de un Modelo de Regresión Logístico Sobre la Oferta Laboral a Jefes(as) de Hogares en El Salvador”, la que consiste en determinar las probabilidades de que un(a) jefe(a) de hogar se encuentre desocupado(a) a partir del conocimiento de variables socio-económicas y demográficas [15].
- 2.“Comparación entre el análisis discriminante y la regresión logística en la clasificación de una colonia de cangrejos herradura (*Limulus Polyphemus*)”, la que consiste en aplicar las técnicas (Regresión logística y análisis discriminante)

a un conjunto de variables que representan características cualitativas y cuantitativas de una muestra de 173 cangrejos de herradura [20].

3. “Métodos robustos aplicados a la clasificación del estado nutricional de la niñez salvadoreña (FESAL 2008)”, la que consiste en una aplicación de los modelos de regresión logística y del análisis discriminante con el fin de determinar el conjunto de variables más importantes en la estimación del estado nutricional de la niñez salvadoreña.

Temática abordada por estudiantes de la Licenciatura en Estadística:

1. “Modelación logística multinomial para clasificar los hogares de El Salvador por nivel de pobreza”, consiste en realizar una aplicación de los modelos de regresión logística multinomial para determinar las variables de mayor incidencia en la clasificación de los hogares de El Salvador por nivel de pobreza, [14].

Aplicaciones de regresión logística multinomial elaboradas en otros países en distintas áreas, por ejemplo:

1. “Regresión logística. Un ejemplo de su uso en Endocrinología. Cuba” [27].
2. “Aplicación de la regresión logística multinomial en la detección de factores económicos que influyen la productividad de los sectores industriales. Venezuela” [28].
3. “Empleo de regresión logística para la obtención de modelos de riesgo humano. Venezuela” [29].
4. “Desarrollo de un modelo de geoprocesamiento para la valoración productiva y tributaria de tierras agrícolas en Venezuela” [30]. A pesar que estas aplicaciones han sido para abordar problemáticas diferentes; en esta ocasión se ocupará la regresión logística multinomial para hacer una aplicación en el ámbito de la salud.

## **CAPÍTULO I: FUNDAMENTOS TEÓRICOS.**

### **INTRODUCCIÓN**

En este capítulo se explica toda la metodología que se ha utilizado para resolver el problema y alcanzar el objetivo general que sea planteado en esta investigación, se describen los diferentes enfoques de los modelo lineales generalizados así como también el multinomial y entre otras cosas, la formulación, los métodos de estimación, interpretación de parámetros, ajustes del modelo, validación, contrastes de hipótesis asociados. Pero antes se definen algunos términos clínicos propios de la odontología que son de importancia y que se utilizarán en el contexto de la investigación además de los fundamentos teóricos de la comparación de varias proporciones.

La Organización Mundial de la Salud (OMS) define salud bucodental (salud bucal) como la ausencia de enfermedades y trastornos que afectan boca, cavidad bucal y dientes, como cáncer de boca o garganta, llagas bucales, defectos congénitos como labio leporino o paladar hendido, enfermedades periodontales, caries dental, dolor orofacial crónico, entre otros. Por tanto, este concepto se refiere al estado de normalidad y funcionalidad eficiente de los dientes, estructuras de soporte y de los huesos, articulares, mucosas, músculos, todas las partes de la boca y cavidad bucal, relacionadas con la masticación, comunicación oral y músculo facial.

Adoptar medidas de prevención puede evitar sufrir graves dolencias bucodentales, así como gastos gigantescos en tratamientos de este tipo, a continuación se presentan algunas recomendaciones tan sencillas y simples que contribuyen a la conservación de esta:

- ✓ Mantener alimentación equilibrada, procurando alta ingesta de frutas, legumbres y verduras, que son indispensables para la buena salud bucodental, pues además de limpiar los dientes, los fortalecen y protegen.
- ✓ Reducir la ingesta de azúcares previene la caries dental y la pérdida prematura de dientes.

- ✓ Dejar de fumar y reducir el consumo de alcohol disminuyen el riesgo de cánceres de la cavidad bucal, periodontopatías y pérdida de dientes.
- ✓ Procurar constantemente alta concentración de fluoruro en la cavidad bucal, a largo plazo, reduce el número de caries tanto en los niños como en los adultos. Esto puede conseguirse mediante la fluoración del agua que se bebe a diario, la sal, o bien por el uso de enjuagues bucales o la aplicación profesional de fluoruros en el consultorio dental.
- ✓ Realizar limpieza bucal tres veces al día mediante cepillado, rutina que debe durar al menos tres minutos, usando alternadamente cepillo de dientes, hilo o instrumentos de limpieza interdental y 30 segundos de enjuague bucal.
- ✓ No comer dulces antes de dormir, ni dar biberón al niño con leche o bebidas azucaradas, ya que se fermentan rápido durante la noche y ponen en riesgo de caries dental. Si el pequeño tiene sed solo dar agua en su biberón.
- ✓ Consultar al especialista al menos dos veces al año, para revisión y/o tratamiento, así como para recibir orientación sobre la mejor forma de cuidar boca, cavidad bucal y dientes.

La buena salud bucodental repercute no sólo en el aspecto estético, si no también en la salud y bienestar en general de la persona, ya que facilita una mejor masticación de los alimentos y, por lo tanto, un eficaz aprovechamiento de los nutrientes por el organismo, entre otras funciones importantes. Basándonos en el concepto de salud bucal según la Organización mundial para la salud bucal se han considerado las siguientes variables para el desarrollo de esta investigación:

$x_{i1}$  = Edad del i – ésimo individuo

$x_{i2}$  = Sexo del i – ésimo individuo

$x_{i3}$  = Zona de residencia del i – ésimo individuo

$x_{i4}$  = Nivel educativo del i – ésimo individuo

La variable económica será:

$x_{i5}$  = Ingreso económico mensual del  $i$  – ésimo individuo

Y las variables índices odontológicos:

$x_{i6}$  = Índice de Gingivitis del  $i$  – ésimo individuo

$x_{i7}$  = Índice de Placa Bacteriana del  $i$  – ésimo individuo

$x_{i8}$  = Índice de Periodontitis del  $i$  – ésimo individuo

$x_{i9}$  = Número de dientes con caries del  $i$  – ésimo individuo

$x_{i10}$  = Número de dientes perdidos del  $i$  – ésimo individuo

$x_{i11}$  = Número de dientes obturados del  $i$  – ésimo individuo

En la siguiente Tabla 1.1 se presentan de forma detallada las variables que se consideran independientes y variable dependiente y sus respectivos indicadores. Como puede observarse la mayoría de las variables son cualitativas y además tienen diferentes tamaños en niveles.

**Tabla 1.1. Variables e indicadores.**

Variable	Indicador
Salud bucal (Variable Dependiente)	Condición sana y leve
	Codición moderada
	Condición severa
Variables independientes	
Índice de Placa bacteriana Løe y Silness	Buen estado
	Condición leve
	Condición moderada
	Condición severa
	No evaluable
Índice gingival Løe y Silness	Buen estado
	Inflamación leve
	Inflamación moderada
	Inflamación severa
	No evaluable

Índice Periodontal Comunitario CPI (OMS)	Sano, ausencia de signos
	Hemorragia al sondaje suave
	Cálculo supra o subgingival
	Bolsa de 4 -5 mm
	Bolsa igual o mayor de 6 mm
	No evaluable
Maloclusión, Índice DAI (de OMS)	Oclusión normal y mal oclusión mínima
	Mal oclusión definida
	Mal oclusión severa
	Mal oclusión muy severa o discapacitante
	No evaluable
sexo	FEMENINO
	MASCULINO
ingreso mensual	Menor o igual \$250
	Mayor a \$250
nivel de escolaridad	Sin escolaridad
	Educación básica
	Bachillerato
	Educación superior
Región Geográfica	Zona Oriental
	Zona Central
	Zona Paracentral
	Zona Occidental
CPO-D	Número de dientes con caries
	Número de dientes perdidos
	Número de dientes con obturaciones

Tradicionalmente las variables dependientes politómicas, han sido modeladas mediante análisis discriminante pero, cuando las variables independientes son cualitativas, tal como es la variable dependiente condición de salud bucal, entonces son más adecuados los modelos de regresión logística multinomial. La utilidad de estos

modelos es que son una herramienta poderosa para la clasificación y pueden implementarse en paquetes estadísticos como R y SPSS.

### **1.1 DEFINICIONES Y CÁLCULOS DE ÍNDICES ODONTOLÓGICOS**

Índice en epidemiología se define como un valor numérico que describe el estado relativo de una población respecto a una escala graduada con límites superiores e inferiores definidos y diseñados para presentar y facilitar la comparación con otras poblaciones clasificadas de acuerdo con los mismos criterios y métodos. (Young y Striffler)

Una proporción o coeficientes pueden verse como un índice que sirven para expresar la frecuencia con que ocurren ciertos hechos en la comunidad y que pueden incluir o no determinaciones del grado de severidad. Requisito que deben cumplir los índices:

1. Relación con el fenómeno que se desea estudiar, es decir pertenencia.
2. Ser confiable o lo que es lo mismo que mantenga su validez en el análisis estadístico y significación, es decir que brinde una idea comprensible del fenómeno estudiado.

Un índice puede describir la prevalencia de una enfermedad en una población y también puede describir la gravedad o la intensidad de la condición.

#### **1.1.1. ÍNDICE DE PLACA BACTERIANA LÖE Y SILNESS. (IP)**

Se llama placa bacteriana (biofilm oral o placa dental) a una acumulación heterogénea de una comunidad microbiana variada, aerobia y anaerobia, rodeada por una matriz intercelular de polímeros de origen salival y microbiano. Estos microorganismos pueden adherirse o depositarse sobre las paredes de las piezas dentarias. La placa dental se forma en la superficie de dientes, encía y restauraciones, y difícilmente puede observarse, a menos que esté teñida. Su consistencia es blanda, mate, color blanco-amarillo. Se forma en pocas horas y no se elimina con agua a presión. Varía de un

individuo a otro, y también varía su localización anatómica. Si la placa dental se calcifica, puede dar lugar a la aparición de cálculos o sarro tártaro. Su presencia puede estar asociada a hábitos de higiene, pero si los microorganismos consiguen los sustratos necesarios para sobrevivir y persisten mucho tiempo sobre la superficie dental, pueden organizarse y causar caries, gingivitis o enfermedad periodontal (enfermedades de las encías) [6].

En éste índice se examinará la clasificación de los dientes establecidos por Ramfjord, en los cuatro sitios; mesial, vestibular, distal y palatino o lingual. En caso de dientes ausentes se realiza la lectura en el diente vecino del mismo grupo dentario.

Dándole los siguientes valores a cada superficie.

0 = Ausencia de Placa

1 = Placa no visible, pero que se extrae del tercio gingival del diente con ayuda de una sonda.

2 = Acumulación moderada de placa apreciable a simple vista.

3 = Placa abundante en esta misma zona (tercio gingival) e incluso cubriendo el diente adyacente.

Índice de placa bacteriana (I.P): se obtiene efectuando la sumatoria de los valores de todas las superficies entre el número de superficies examinadas.

$$I.P = \frac{\sum \text{de los valores asignados a cada superficie}}{\text{Nº de superficies examinadas}}$$

Resumen del índice para interpretación:

0.0 a 0.75 Buen estado

0.76 a 1.5 condición leve

1.51 a 2.25 condición moderada

2.26 a 3.00 condición severa

### **1.1.2. ÍNDICE GINGIVAL DE LÖE Y SILNESS (IG)**

Se conoce como gingivitis a la inflamación en encías enrojecidas o que sangren, pueden ser las primeras señales de una gingivitis. Si la enfermedad es ignorada, los tejidos que mantienen a los dientes en su lugar pueden comenzar a destruirse y al perder el soporte se pierden los dientes. La gingivitis se debe a los efectos de la placa bacteriana, si la placa dentobacteriana no se elimina se convierte en cálculo. Tanto la placa bacteriana como el cálculo irritan las encías produciendo inflamación [4], [6]. Cabe destacar que la encía es la que protege los dientes y recubre los maxilares. Está compuesta por un tejido conectivo y revestido de epitelio.

Además de la mala higiene dental es importante resaltar que existen factores que aumentan el riesgo a que alguien padezca de gingivitis pueden ser el tener diversas deficiencias de tipo nutricional, sufrir enfermedades como la diabetes, el tomar medicamentos tales como los anticoagulantes o los antibióticos, el estar sumido en una constante tensión emocional o el contar con desequilibrios hormonales debido a un embarazo o a la propia menstruación.

Entre los síntomas más claros, que determinan que una persona esté sufriendo dicha enfermedad se encuentran el mal aliento, el sangrado de las encías cuando se lleva a cabo el cepillado de los dientes o la hinchazón de las mismas, además también del color rojo de ellas.

La higiene bucal, en definitiva, es la mejor manera de prevenir la aparición y el desarrollo de gingivitis. Esta higiene debe incluir no sólo el cepillado de los dientes tras de cada comida, sino también el uso de hilo dental para quitar los restos que están en lugares inaccesibles para el cepillo. En caso que la gingivitis ya esté presente, se requiere de un tratamiento odontológico para revertir la inflamación y el sangrado.

Además del correcto cepillado, hay que destacar el llevar a cabo una alimentación sana y correcta, el ir periódicamente a la consulta del dentista y eliminar por completo una serie de hábitos como, por ejemplo, el tabaco [13].

El índice gingival de Löel y Silves, también elige los dientes que se tomaron para identificar la placa bacteriana y observar el color, tamaño, aspecto, consistencia, hemorragia de la encía correspondiente a cada diente examinado para poder determinar el estado y a cada superficie se asignan los siguientes valores:

0 = Encía Sana

1 = Inflamación leve sin hemorragia

2 = Inflamación moderada con hemorragia al sondaje

3 = Inflamación severa con hemorragia espontánea

Índice gingival (IG): se obtiene efectuando la sumatoria de los valores de todas las superficies entre el número de superficies examinadas, es decir,

$$I.G = \frac{\sum \text{de los valores asignados a cada superficie}}{\text{Nº de superficies examinadas}}$$

Resumen del índice para interpretación:

0.0 a 0.75 Buen estado

0.76 a 1.5 Inflamación leve

1.51 a 2.25 Inflamación moderado

2.26 a 3.00 Inflamación severa

### **1.1.3. ÍNDICE PERIODONTAL COMUNITARIO CPI (DE OMS)**

La periodontitis, denominada comúnmente piorrea, es una enfermedad crónica e irreversible que puede cursar con una gingivitis inicial, para luego proseguir con una pérdida de inserción del colágeno, recesión gingival, e incluso la pérdida de hueso; en el caso de no ser tratada, deja sin soporte óseo al diente; la pérdida de dicho soporte implica la pérdida irreparable del diente mismo. Es de etiología bacteriana afecta al periodonto (el tejido de sostén de los dientes, constituido por la encía, el hueso alveolar, el cemento radicular y el ligamento periodontal). Y se manifiesta más

comúnmente en adultos mayores de 35 años, pero puede iniciarse en edades más tempranas [25].

La enfermedad en su forma agresiva (periodontitis agresiva) puede aparecer en edades tempranas, evolucionando de manera rápida, lo que provoca la pérdida de piezas dentales en personas jóvenes.

Para determinar cómo se manifiesta la condición periodontal a nivel comunitario se utiliza el índice CPI y se obtiene de la siguiente manera:

- ✓ La cavidad bucal es seccionada en 6 partes, cada sextante debe contar por lo menos con 2 dientes no indicados para extracción. De lo contrario se considera como sextante excluido.
- ✓ A cada sextante, asigne un valor de acuerdo a:

Valores del Índice Periodontal Comunitario:

0 = Sano, ausencia de signos

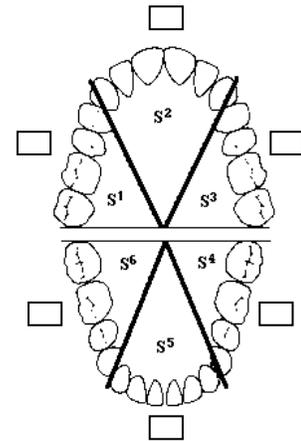
1 = Hemorragia al sondaje suave.

2 = Cálculo supra o subgingival

3 = Bolsa de 4 – 5 mm.

4 = Bolsa igual o mayor de 6 mm

**CPI = Al mayor valor encontrado.**



#### 1.1.4. ÍNDICE DE ESTÉTICA DENTAL (DAI).

El índice de estética dental conocido como DAI (Dental Aesthetic Index) por sus siglas en inglés fue creado con fines epidemiológicos y adoptado por la OMS para tal propósito. Este índice relaciona la apariencia estética de los dientes con posiciones oclusales determinadas, clasifica la gravedad de la mal oclusión y su necesidad de tratamiento a través de una ecuación de regresión estándar que tiene presente esos

componentes o características oclusales, con sus correspondientes coeficientes de regresión (ver Tabla 1.1). Cada característica fue multiplicada por su correspondiente peso y sumada la constante 13, para obtener el índice de cada uno de los individuos [10], [16].

Los pasos a seguir para obtener el valor del índice son:

1. En primer lugar este índice se aplica a individuos mayores de 12 años. (es decir individuos sólo con dentición permanente).
2. Se registra el dato de las diferentes situaciones clínicas en la casilla correspondiente en la columna (A); como se muestra en la Tabla 1.1 Componentes o situaciones clínicas.
3. Luego se multiplica cada valor numérico de la casilla (A) por el correspondiente coeficiente de regresión DAI estándar de la casilla (B) y se coloca el dato en la columna A x B.
4. Se suman los productos obtenidos en la columna A x B. El resultado de esta suma más 13, corresponde al valor de DAI de cada paciente.
5. Por último se coloca el resultado en la categoría correspondiente de acuerdo al resultado obtenido, para establecer la necesidad o no de tratamiento, como se muestra en la Tabla 1.2.

**Tabla 1.1. Componentes o situaciones clínicas**

Componentes o situaciones clínicas	Medición obtenida (A)	Regresión (B)	Producto A x B
Dientes visibles perdidos (incisivos, caninos, premolares, superiores e inferiores)		6	
Apiñamiento antero superior e inferior (de canino a canino) 0: No hay apiñamiento. 1: apiñamiento en un arco (superior o inferior)		1	

2: Apiñamiento en ambos arcos.			
Espaciamiento (falta de contacto interproximal) (de canino a canino) 0: No hay diastemas. 1: Diastema en un arco (superior o inferior) 2: Diastemas en ambos arcos.		1	
Diastema en línea media en mm (mayor medida)		3	
Mayor rotación maxilar en mm.		1	
Mayor rotación mandibular en mm.		1	
Overjet maxilar en mm		2	
Overjet mandibular (mordida cruzada) en mm		4	
Mordida abierta anterior en mm		4	
Relación oclusal molar (mesial o distal respecto de la normal) ya sea derecha o izquierda. 0: normal 1: media cúspide, ya sea mesial o distal. 2: una cúspide completa o más, ya sea mesial o distal.		3	
Constante			13
DAI = Suma del producto más 13:			DAI =

En la siguiente Tabla 1.2 esta detallado la necesidad de tratamiento según el valor obtenido en que tipo de oclusión o mal oclusión se encuentra el paciente

**Tabla 1.2. Necesidad de Tratamiento**

Oclusión normal y Mal oclusión mínima.	DAI menor o igual que 25
Mal oclusión definida	DAI entre 26 y 30
Mal oclusión severa	DAI entre 31 y 35
Mal oclusión muy severa o discapacitante.	DAI mayor o igual que 36

### 1.1.5. ÍNDICE DE CARIES CPO-D

El más utilizado y difundido de los indicadores de caries dental es sin duda es el Índice CPO, que principalmente refleja la experiencia de caries dental tanto presente como pasada en la dentición permanente éste indicador fue desarrollado por Klein, Palmer y Knutson en 1937 [12]. El Índice CPO-D considerada la historia de la patología en el individuo, ya que en su registro se incluyen datos sobre:

- ✓ Las piezas dentarias con lesión activa y clínicamente evidente (cariadas).
- ✓ Las piezas dentarias extraídas - perdidas por caries dental y aquellas que están indicadas para una extracción (perdidas).
- ✓ Las piezas que ya recibieron tratamiento para la caries dental (obturadas)

#### INDICE CPO-D

C = Dientes cariados

P = Dientes perdidos

O = Dientes obturados

#### ¿Cómo se determina el Índice CPO-D?

El Índice CPO-D se registra para cada individuo y toma en cuenta la dentición permanente, más específicamente las 28 piezas dentarias permanentes, sin contar las terceras molares. Se anota el número de dientes cariados, obturados y perdidos, incluyéndose las extracciones indicadas debido a caries dental tal como se ejemplifica en la Tabla 1.3, [8], [9]. Su valor constituye en realidad un promedio, que se calcula de la siguiente manera:

$$\text{CPO-D} = \frac{\text{sumatoria de dientes permanentes cariados, perdidos y obturados}}{\text{total de personas examinadas}}$$

**Tabla 1.3. Cuantificación de la OMS para el índice CPO-D de cada paciente.**

<b>Distribución de frecuencia de la condición del diente (CPO-D)</b>	<b>Frecuencia</b>
C	
P	
O	
<b>TOTAL CPO-D</b>	

En el Tabla 1.4 vemos la clasificación que la OMS hace del índice CPO-D; al dividir el TOTAL CPO-D entre el total de pacientes examinados se obtendrá la clasificación del grupo según dicho cuadro:

**Tabla 1.4. Clasificación del índice CPO-D**

Muy Bajo	0.0 a 1.1
Bajo	1.2 a 2.6
Moderado	2.7 a 4.4
Alto	4.5 a 6.5

Dejando de una vez claros los conceptos básicos sobre los índices odontológicos que forman parte de la base de datos utilizada para esta investigación, se relatan también las técnicas estadísticas que se aplicarán de acuerdo a las exigencias de los datos como se detallan a continuación:

## **1.2. PRUEBA PARA VARIAS PROPORCIONES.**

En la investigación biomédica se encuentran con frecuencia datos o variables de tipo cualitativo (nominal u ordinal), mediante las cuales se clasifican grupos de sujetos o individuos en dos o más categorías excluyentes entre ellas. Cuando pretendemos comparar grupos de sujetos con respecto a una variable categórica, los resultados se presentan mediante tablas de frecuencias de doble entrada, conocidas con el nombre de Tablas de contingencia. El método estadístico a utilizar dependerá del número de proporciones a comparar, es decir, del número de categorías de la variable que se quiere comparar.

El caso más sencillo y habitual de comparar dos variables cualitativas, es aquel en que ambas variables tienen dos posibles categorías de respuesta (estas variables reciben el nombre de variables dicotómicas), reduciéndose la tabla de contingencia a una tabla 2x2. Generalmente una de ellas corresponde al tratamiento y la otra al resultado (curación, éxito, muerte, entre otras).

Las pruebas estadísticas aplicables en la comparación de proporciones, ya sean dos o más, también difieren según se trate de comparar medidas realizadas en grupos independientes o de datos apareados (medidas realizadas en un mismo grupo de individuos en dos momentos distintos del tiempo). En el caso de comparar dos proporciones independientes, las pruebas más utilizadas son la prueba Z de comparación de proporciones y la prueba de Ji-cuadrado. En el caso de tratarse de datos apareados puede utilizarse la prueba de McNemar. En la comparación de más de dos proporciones independientes es posible aplicar la prueba de la Ji-cuadrado.

La estadística Ji Cuadrada para probar la homogeneidad también se aplica cuando se prueba la hipótesis de que  $k$  parámetros binomiales tienen el mismo valor. Esta es, por tanto, una extensión de la prueba donde se quiere determinar las diferencias entre dos proporciones a una prueba para determinar diferencias entre  $k$  proporciones, corresponde a la comparación de las proporciones de las columnas. A través de esta prueba se determina si existe una diferencia estadísticamente significativa entre los porcentajes obtenidos por cada una de las categorías de la variable ubicada en las columnas de acuerdo al cruce con cada una de las categorías de la variable ubicada en las filas; es decir, compara todos los porcentajes obtenidos para las categorías de la variable de columna con la primera, segunda, etc, categoría de la variable de fila.

Cuando queremos comparar una respuesta que se mide como una proporción entre dos o más niveles necesitamos pruebas que nos indiquen si hay diferencias entre estas proporciones, es decir, si se distribuyen homogéneamente entre los niveles de la

variable o por el contrario, existen diferencias. Por lo tanto, la hipótesis experimental es que las proporciones de ocurrencia de determinado evento medido en muestras independientes son diferentes. Por ejemplo, la comparación de medidas de respuesta tipo curación, fracaso y/o evolución en distintos tratamientos corresponden a este caso.

La prueba de la Ji-Cuadrado es una de las pruebas más frecuentemente utilizadas para el contraste de variables cualitativas, aplicándose para comparar si dos características cualitativas están relacionadas entre sí, si varias muestras de carácter cualitativo proceden de igual población o si los datos observados siguen una determinada distribución teórica. Para su cálculo se calculan las frecuencias esperadas (las que deberían haberse observado si la hipótesis de independencia fuese cierta), para compararlas con las observadas en la realidad. Se calcula el valor del estadístico  $\chi^2$  como:

$$\chi^2 = \sum \frac{|O_{ij} - E_{ij}|^2}{E_{ij}} \quad \chi^2_{(f-1)(c-1)}$$

Donde:

$O_{ij}$ : corresponde a las frecuencias observadas dentro de la casilla de la fila  $i$  y la columna  $j$ .

$E_{ij}$ : corresponde a las frecuencias esperadas o teóricas.

$f$ : es el número de las filas.

$c$ : el número de columnas.

$(f-1) \times (c-1)$ : corresponde a los grados de libertad de la distribución del estadístico de contraste.

El primer paso consiste en construir la tabla de contingencia asociada a las dos variables a analizar. A partir de ella se calculan las frecuencias esperadas en cada casilla bajo la suposición de que ambas variables sean independientes.

En el caso de una tabla de contingencia de f filas y c columnas, las frecuencias esperadas se pueden obtener de manera similar, como se describe en la siguiente Tabla 1.5:

**Tabla 1.5 Cálculo de las frecuencias esperadas en una tabla de contingencia**

	$A_1$	$A_2$	...	$A_c$	Total
$Y_1$	$E_{11} = \frac{f_{1.} * f_{.1}}{f}$	$E_{12} = \frac{f_{1.} * f_{.2}}{f}$		$E_{1c} = \frac{f_{1.} * f_{.c}}{f}$	$f_{1.}$
$Y_2$	$E_{21} = \frac{f_{2.} * f_{.1}}{f}$	$E_{22} = \frac{f_{2.} * f_{.2}}{f}$		$E_{2c} = \frac{f_{2.} * f_{.c}}{f}$	$f_{2.}$
...					
$Y_f$	$E_{f1} = \frac{f_{f.} * f_{.1}}{f}$	$E_{f2} = \frac{f_{f.} * f_{.2}}{f_{..}}$		$E_{fc} = \frac{f_{f.} * f_{.c}}{f_{..}}$	$f_{f.}$
Total	$f_{.1}$	$f_{.2}$		$f_{.c}$	$f_{..}$

Para obtener el valor de la Ji-cuadrado las frecuencias observadas se comparan con los valores esperados. Así, cuando mayor sea la diferencia entre los valores esperados y los valores observados mayor será el valor del estadístico, existiendo en este caso asociación entre las variables comparadas. El hecho de que las diferencias se eleven al cuadrado convierte cualquier diferencia en positiva, lo que indica si existe o no relación entre los factores pero no en que sentido se produce tal asociación.

Bajo la hipótesis nula de independencia, el estadístico Ji-cuadrado se distribuye según una distribución Ji-cuadrado con  $(f-1) \times (c-1)$  grados de libertad. En el caso de tablas 2x2 los grados de libertad son 1. Cuando el tamaño muestral no es demasiado grande, puede introducirse algún sesgo en los cálculos, ya que estos contrastes aproximan una distribución discreta por una continua. En caso de que más del 20% de las frecuencias esperadas sean menores de 5 o bien alguna celda tenga valores esperados inferiores a 2, se utiliza una corrección para eliminar este sesgo, conocida como la corrección de Yates para continuidad, aplicable en el caso de tablas 2x2.

Por ello, interesa probar  $H_0: p_{i1} = p_{i2} = \dots = p_{ik} \quad \forall i = 1, 2, \dots, r$  la hipótesis nula contra la hipótesis alternativa  $H_1$ , las proporciones de la población son distintas.

Al seleccionar la región crítica apropiada de la cola superior de la forma  $\chi^2 > \chi^2_\alpha$ , (con  $\alpha$  el nivel de tolerancia); se puede llegar ahora a una decisión con respecto a  $H_0$ .

### **1.3. FUNDAMENTOS TEORICOS SOBRE MODELOS LINEALES GENERALIZADOS Y DEL MODELO DE REGRESIÓN LOGISTICA MULTINOMIAL.**

En el presente tema se expone el orden en que se abordó la aplicación de los modelos lineales generalizados, además se realiza una exposición de la regresión logística simple y posteriormente se introduce el concepto de regresión multinomial con el objetivo de exponer de manera comprensible la metodología que se utiliza para explicar la razón del porque los datos se ajustan a un modelo de regresión logística multinomial y luego el proceso del tratamiento para obtener el modelo más parsimonioso.

#### **1.3.1. EL MODELADO ESTADISTICO**

“En las ciencias del hombre las teorías formuladas matemáticamente se llaman a menudo modelos matemáticos...” (Bunge, 1969). Anguera (1989) señala que, en Psicología, la utilización de los modelos son una de las estrategias más utilizadas de las últimas décadas para llevar a cabo representaciones conceptuales.

Arnu (1989) considera el modelo como una estructura conceptual a mitad de camino entre la explicación completamente teórica y los datos puramente empíricos, constituyen una intensa relación entre sistematización y descripción de lo real, en función de unos presupuestos teóricos.

Los modelos, en tanto resultan ser aproximaciones, no son la propia realidad, sino representaciones simbólicas de la realidad, por lo que en caso de estar bien elaborados, pueden ayudar a entenderla de una mejor manera (Llópiz, 1996).

La construcción de modelos estadísticos o modelos matemáticos aplicados se incluye dentro de lo que ha venido a denominarse modelado estadístico (“statistical modelling”). Como señala Lindsey (1993), con el modelado estadístico se pretende descubrir la variabilidad en los datos observados mediante procedimientos matemáticos.

El modelado estadístico consiste en una serie de procedimientos encaminados a explicar el comportamiento de una variable respuesta,  $Y$ , a través de una o más variables explicativas,  $X$ , mediante una función, en la cual se expresan la/s variable/s explicativa/s y los parámetros que la forman:  $Y = f(X)$ . Sin embargo, debido a que la explicación de la variable respuesta en función de la/s variable/s explicativa/s no es perfecta, se incluye un término denominado error o residual; por lo tanto el modelo se representa así:

$$Y = f(X) + \varepsilon \quad (1.1)$$

Siguiendo a Arnau (1989), y desde una perspectiva cualitativa, pueden darse tres posibilidades:

- Que la variabilidad observada en los datos sea totalmente explicada por el modelo, lo cual daría lugar a la consideración como un modelo absoluto determinístico y formal.
- Que la variabilidad observada en los datos está totalmente explicada por factores aleatorios, lo cual daría lugar a un modelo aleatorio, la base del cual puede encontrarse en la ejecución de errores de muestreo, de medida o, debido a las características inherentes del sistema observado.
- Que la variabilidad observada en los datos necesite de la participación de factores sistemáticos y aleatorios, lo que da lugar a los modelos probabilísticos o estadísticos.

En base a este contexto, por medio del modelado estadístico se pretende buscar el modelo más parsimonioso que sea capaz de explicar los datos con el mínimo error posible, las etapas del modelado estadístico pueden encontrarse bajo diferentes epígrafes (como por ejemplo, Arnau, 1981; Ato y López, 1996; McCullagh y Nelder, 1989), pero en esencia, vienen a englobar los mismos contenidos, pudiéndose obtener fundamentalmente de los estudios que tratan del modelado las siguientes fases:

1. Identificación/Selección. En esta primera etapa se han de estudiar y considerar diferentes modelos, así como cada uno de los componentes que pueda intervenir.
2. Estimación. Una vez se ha cumplido la primera etapa, es decir, se ha seleccionado e identificado el modelo que hipotéticamente ajusta los datos, se procede a la estimación de los parámetros, de manera que partiendo de los datos observados y, con un margen de incertidumbre determinado, se obtiene el valor de cada parámetro identificado, así como la significación estadística que alcanza cada uno de ellos.
3. Validación. Una vez estimados los parámetros el paso siguiente consiste en averiguar si se cumplen los supuestos que impone el modelo, así como el hecho de que los residuales que ha dejado el modelo no presentan ningún patrón que pueda resultar significativo. En caso contrario, convendría iniciar el proceso de selección e identificación de un nuevo modelo.
4. Metadiagnóstico. Una vez se ha comprobado que el modelo cumple los requisitos para su aceptación, se ha de tomar la decisión acerca de la bondad del modelo de acuerdo a los fines del análisis y al principio de parsimonia, en el caso que existan varios modelos en lo que se ajusten las observaciones. Un aspecto relacionado con los anteriores es el del uso y aplicaciones del modelo en función del cuál sea el objeto del análisis: pronóstico, evaluación de impactos, teórico, etc.

### 1.3.2. LOS MODELOS LINEALES GENERALIZADOS (MLGs)

En esta sección se aborda una extensión de los modelos lineales a una familia más general, propuesta por Nelder y Wedderburn (1972), denominada Modelos lineales generalizados (MLGs). En esta familia se incluyen tanto los modelos con variables respuesta numérica como categóricas, lo cual lleva a considerar otras distribuciones tales como las distribuciones de la familia exponencial.

Nelder y Wedderburn estudiaron los modelos lineales generalizados extendiendo la teoría de los modelos lineales, incorporando de esta manera la posibilidad de modelar variables respuestas continuas o categóricas con distribuciones del error no necesariamente homocedásticos. Los modelos log lineales, logit, probit, logístico y de regresión lineal son algunos modelos cuya variable respuesta tiene distribución perteneciente a la familia exponencial [3].

### 1.3.3. LA ESTRUCTURA DEL MODELO

Consideremos particularmente una variable aleatoria univariante  $Y$ , cuya distribución de probabilidades depende solamente del parámetro  $\theta$  y  $\phi$ , en los cuales cada uno de ellos respectivamente, representa el parámetro de dispersión que se supone conocido o se estima. Una distribución pertenece a la familia exponencial si puede escribirse de la forma:

$$f(y, \theta, \phi) = \exp\left(\frac{t(y)q(\theta) + r(\theta)}{p(\phi)} + s(y, \phi)\right)$$

Donde  $p(\cdot)$ ,  $t(\cdot)$ ,  $q(\cdot)$ ,  $r(\cdot)$  y  $s(\cdot)$  son funciones conocidas para cada variable aleatoria  $Y$ . Si  $t(y) = y$ , se dice que la distribución está escrita en la forma canónica o estándar y  $q(\theta)$  es llamado algunas veces el parámetro natural o canónico de la distribución.

Todos los MLGs tienen tres componentes (Agresti, 1990, 1996; Ato y López, 1996), un componente aleatorio, un componente sistemático y una función de enlace [3]:

1. **La componente aleatoria** que consiste en variables de respuesta  $Y_1, Y_2, \dots, Y_k$  independientes con función de densidad de la familia exponencial que son de la forma estándar.
2. **La componente sistemática** consiste en el conjunto de predictores lineales  $(\eta_1, \eta_2, \dots, \eta_k)$  tales que:

$$\eta_i = \sum_{j=1}^p \beta_j x_{ij}, \quad i = 1, \dots, n$$

Para  $p$  variables explicativas representadas en el vector

$$x_{ij} = (x_{i1}, \dots, x_{ip})$$

Y  $\beta_j$  es el vector de parámetros que acompaña a cada elemento de del vector  $x_{ij}$ .

3. **La tercera componente consiste en la función de enlace o nexa**,  $g(x)$ , que conecta las componentes aleatoria y sistemática. Sea  $\mu_i = E[Y_i]$ ,  $i = 1, 2, 3, \dots, n$ . El modelo enlaza  $\mu_i$  a  $\eta_i$  por  $\eta_i = g(\mu_i)$  es decir  $g(\mu_i) = \sum_{j=1}^p \beta_j x_{ij}$ ,  $i = 1, 2, 3, \dots, n$ . La función de enlace o nexa  $g(\cdot)$  es una función monótona y diferenciable, que transforma la media al parámetro natural y también es llamada el enlace canónico o nexa canónico. Esto es,  $g(\mu_i) = q(\theta_i)$  y  $q(\theta_i) = \sum_{j=1}^p \beta_j x_{ij}$ ,  $i = 1, \dots, n$ .

La relación puede adoptar distintas posibilidades, entre ellas, Agresti (1990, 1996), Ato y López (1996), McCullagh y Nelder (1989) proponen las que aparecen en la Tabla 1.6. Cada una de las funciones de enlace, mostradas en la tabla 1.6; junto a las características de los componentes aleatorio y sistemático, dan lugar a modelos lineales generalizados de mucha utilidad en el campo de la investigación que se observan en la Tabla 1.7

**Tabla 1.6. Clasificación de los modelos lineales generalizados.**

Componente Aleatoria	Función de Enlace	Componente Sistemático	Modelo
Normal	Identidad	Cuantitativo	Modelo de regresión normal
Binomial	Logit	Categorico	Respuesta binomial (Logit)
Binomial	Probit	Categorico	Respuesta binomial (Probit)
Binomial	Logit	Mixto	Respuesta logístico
Poisson	Logarítmica	Mixto	Log-lineal
Multinomial	Logit generalizado	Mixto	Respuesta multinomial

**Tabla 1.7. Enlaces canónicos de unos modelos lineales generalizados importantes.**

Modelo	Enlaces Canónicos
Regresión lineal	$\eta_i = g(\mu_i) = \mu_i$ El enlace es una identidad.
Modelo de Poisson o modelo de conteo	$\eta_i = g(\mu_i) = \ln(\mu_i)$ Por tanto, $\mu_i = \exp(\eta_i) = e^{\eta_i}$
Regresión Logística Binomial	$\eta_i = g(\pi_i) = \ln\left(\frac{\pi_i}{1 - \pi_i}\right)$ . Por tanto, $\pi_i = e^{\eta_i} / (1 + e^{\eta_i})$
Regresión Logística Multinomial	$\eta_{ij} = g(\pi_{ij}) = \ln\left(\frac{\pi_{ji}}{\pi_{rj}}\right)$ .Donde: $\eta_{ij} = \sum_{k=0}^p \beta_{jk} x_{ijk}$ con $x_{ij0} = 1$ $i = 1, \dots, n$ , N° de individuos  $j = 1, 2, \dots, K - 1$ , N° de categorías  $p = N^\circ$ de variables explicativas
Modelo Probit	$F_{z(0,1)}(\mu_i) = \eta_i$

### 1.3.4. EL MODELO DE REGRESIÓN LOGISTICA MULTINOMIAL

En el modelo de regresión logística se codifican los valores de la variable dependiente como 0 y 1, lo que se obtiene como resultado que la media de la variable represente la proporción de éxitos que ocurran (en el caso binomial) o bien, en una de sus múltiples categorías (en el caso multinomial). El valor predicho de la probabilidad por el modelo según la categoría, puede ser interpretado como la probabilidad de que un caso se

posicione en esa categoría. Un modelo lineal no se ajusta apropiadamente a variables binomiales, dado que los valores predichos de la variable dependiente con este modelo (ajustados mediante la ecuación de una recta), pueden tomar valores imposibles de probabilidad mayores que 1 o menores que 0, a pesar de que los valores observados estén entre 0 y 1. La misma situación se extiende a variables multinomiales.

El mejor modelo que linealiza la relación entre variable dependiente e independiente es el modelo con enlace canónico logit, construido a través de regresión logística. En una variable dependiente binomial ( $Y=0$ ;  $Y=1$ ), si se conoce la probabilidad de pertenecer a una clase ( $Y=0$ ), se puede conocer la probabilidad de pertenecer a la otra clase ( $Y=1$ ), es decir:

$$P(Y = 1 | X) = \pi(x); \quad P(Y = 0 | X) = 1 - \pi(x) \quad (1.2)$$

La media o valor esperado de la variable  $Y$  con distribución de probabilidad de Bernoulli será:

$$E(Y) = 1\pi(x) + 0[1 - \pi(x)] = \pi(x) \quad (1.3)$$

Y su varianza:

$$V(Y) = E(Y^2) - [E(Y)]^2 = \pi(x)[1 - \pi(x)] \quad (1.4)$$

Supóngase que  $\pi(x)$  depende de los valores que tome cada una de las variables independientes  $X = (X_1, X_2, \dots, X_p)$ , para reflejar esta dependencia, y simplificar la notación, se seguirá la explicación con una sola variable explicativa  $X$ .

Una posibilidad a la hora de estimar el efecto de la variable explicativa es la utilización de un modelo estándar de regresión lineal, según el cual el valor esperado de  $Y$ , es una función lineal de  $X$ , según la expresión:

$$P(Y = 1) = \beta_0 + \beta_1 X \quad (1.5)$$

El modelo resultante se denomina modelo de probabilidad lineal, puesto que la probabilidad de acierto cambia linealmente de  $X$  donde:

$P(Y = 1)$ , es la probabilidad de observar un éxito.

$X$ , es la variable independiente o predictora.

$\beta_j$ , son los parámetros de la población a ser estimados.

Este modelo de probabilidad presenta el problema en los valores predichos que pueden ser menores que cero o mayores que uno.

Un paso a la solución del problema es reemplazar la probabilidad [ $P(Y = 1)$ ] con el Odds ( $Y=1$ ), el cual se expresa como el cociente entre una probabilidad y la probabilidad complementaria, y se escribe de la siguiente forma:

$$odds(Y = 1) = \frac{P(Y = 1)}{[1 - P(Y = 1)]} \quad (1.6)$$

Como se deduce de la Ecuación 1.6, el Odds se interpreta como cuantas veces es más probable que ocurra un evento a que no ocurra, en particular  $odds(Y = 1)$  se interpreta como cuantas veces es más probable que ocurra un evento de éxito a que no ocurra, su valor puede variar entre 0 y  $+\infty$ , para valores de  $P$  entre 0 y 1, por lo que valores de Odds menores que 0 generen valores imposibles de  $P$  (el valor de una variable puede ser cualquier cifra entre  $-\infty$  y  $+\infty$ , sin embargo, en este modelo está restringida a valores entre 0 y  $+\infty$ ). Para evitar lo anterior se requiere otra transformación, conocida como logit ( $Y$ ) que se expresa:

$$\text{logit}(Y = 1) = \text{Ln} \left( \frac{P(Y = 1)}{[1 - P(Y = 1)]} \right) = \text{Ln} \left( \frac{\pi}{1 - \pi} \right) \quad (1.7)$$

El logit varía entre  $-\infty$  y  $+\infty$ , cuando el Odds varía entre 0 y  $+\infty$ , para valores de P entre 0 y 1, por lo que se elimina el problema que la probabilidad estimada puede exceder el máximo o el mínimo posible.

Así, la ecuación de la relación entre la variable dependiente y las variables independientes es la siguiente:

$$\text{logit}(Y = 1) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p \quad (1.8)$$

El logit de la variable dependiente se puede convertir a Odds mediante la ecuación:

$$\text{odds}(Y = 1) = e^{(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p)} \quad (1.9)$$

La función que mejor se ajusta a las características anteriores mencionadas adquiere la siguiente forma:

$$\pi(x) = \frac{1}{1 + e^{-z}} \quad (1.10)$$

Donde  $\pi(x)$  representa una probabilidad en función de  $x$  y  $Z$  una variable o una combinación de variables explicativas, según lo cual  $Z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p$ , que sustituido en la fórmula (1.10) se obtiene la expresión:

$$\pi(x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p)}} \quad (1.11)$$

O de forma alternativa:

$$\pi(x) = \frac{e^{(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p)}}{1 + e^{(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p)}} \quad (1.12)$$

Cuando la variable dependiente tiene más de dos categorías, se aplica la versión multinomial de la regresión logística. Para una variable dependiente con  $k$  categorías, la

regresión requiere k-1 ecuaciones logísticas, una para cada categoría, en relación a otra categoría tomada como referencia. La relación se representa por las siguientes funciones:

$$g_1(x_{i1}) = Ln \left[ \frac{P(Y=1|x_{i1})}{P(Y=k|X)} \right] = \beta_{10} + \beta_{11}x_{i11} + \beta_{12}x_{i12} + \dots + \beta_{1p}x_{i1p} \\ = x_{i1} \mathbf{B}_j \quad (1.13)$$

$$\text{donde } x'_{i1} = \begin{pmatrix} 1 \\ x_{i11} \\ x_{i12} \\ \vdots \\ x_{i1p} \end{pmatrix} \text{ y } \beta_j = \begin{pmatrix} \beta_{10} \\ \beta_{11} \\ \vdots \\ \beta_{1p} \end{pmatrix}; \text{ con } j = 1$$

$$g_2(x_{i2}) = Ln \left[ \frac{P(Y=2|x_{i2})}{P(Y=k|X)} \right] = \beta_{20} + \beta_{21}x_{i21} + \beta_{22}x_{i22} + \dots + \beta_{2p}x_{i2p} \\ = x_{i2} \mathbf{B}_j \quad (1.14)$$

$$\text{donde } x'_{i2} = \begin{pmatrix} 1 \\ x_{i21} \\ x_{i22} \\ \vdots \\ x_{i2p} \end{pmatrix} \text{ y } \beta_j = \begin{pmatrix} \beta_{20} \\ \beta_{21} \\ \vdots \\ \beta_{2p} \end{pmatrix}; \text{ con } j = 2$$

⋮

$$g_{K-1}(x_{i(k-1)}) = Ln \left[ \frac{P(Y=k-1|x_{i(k-1)})}{P(Y=k|X)} \right] = \beta_{(k-1)0} + \beta_{(k-1)1}x_{i(k-1)1} + \beta_{(k-1)2}x_{i(k-1)2} + \dots + \beta_{(k-1)p}x_{i(k-1)p} \\ = x_{i(k-1)} \mathbf{B}_j \quad (1.15)$$

$$\text{donde } \mathbf{x}_{i(k-1)}^t = \begin{pmatrix} 1 \\ x_{i(k-1)1} \\ x_{i(k-1)2} \\ \vdots \\ x_{i(k-1)p} \end{pmatrix} \text{ y } \beta_j = \begin{pmatrix} \beta_{(k-1)0} \\ \beta_{(k-1)1} \\ \vdots \\ \beta_{(k-1)p} \end{pmatrix}; \text{ con } j = (k-1)$$

Además:

$g_i(x)$ , es la función logit de la categoría  $i$  contra la categoría de referencia.

$x_{ij}$ , es el vector de variables independientes,  $x_{ij} = (1, x_{ij1}, \dots, x_{ijp})$ .

$\beta_j$ , es el vector de coeficientes, estimado para la categoría  $j$ .

Kleinbaum (1994) señala que la forma de la función logística es especialmente atractiva a los epidemiólogos, ya que si, por ejemplo, se trata de ver el efecto combinado de diferentes factores de riesgo  $Z$  sobre la probabilidad de padecer un enfermedad, ésta es mínima para valores bajos de  $Z$  hasta un determinado umbral, a partir del cual la probabilidad se eleva de forma rápida y, una vez  $Z$  ha crecido lo suficiente, la función permanece constante con valores altos, es decir próximos a 1.

Agresti (1989) señala que la elección de la función logística viene determinada por dos razones:

1. Desde el punto de vista matemático, es una función extremadamente fácil y flexible.
2. Permite una interpretación significativa desde el punto de vista biológico.

Considerando una variable respuesta poltómica  $Y$ , con más de dos categorías de respuestas que denotaremos por  $y_1, y_2, \dots, y_k$ .

Se pretende explicar la probabilidad de cada categoría de respuesta en función de un conjunto de covariables y/o factores  $X = \{x_1, x_2, \dots, x_p\}$  observados. Es decir, ajustar un

modelo de la forma  $p_{ij} = P(Y = y_j / X_{ij})$ ,  $\forall j = 1, \dots, k$ . Para cada vector  $X_{ij}$  de valores observados de las variables explicativas.

En el caso de una variable de respuesta binaria, la distribución condicionada a cada combinación de valores observados de las covariables sigue una distribución de Bernoulli. Cuando la variable de respuesta es politómica, la distribución de Bernoulli se convierte en una distribución multinomial de respuesta.

Es decir,  $(Y | X = x) \rightarrow M(n; \pi_1, \dots, \pi_k)$ , siendo  $\sum_{j=1}^k \pi_j = 1$ . En un modelo de regresión logística multinomial se tendrán  $\binom{k}{2}$  transformaciones logit al comparar cada categoría de la variable respuesta [1].

Dando como resultado K-1 elecciones con otra categoría de referencia, y se expresarían de la siguiente forma:

$$Ln \left[ \frac{\frac{\pi_i}{\pi_i + \pi_j}}{\frac{\pi_j}{\pi_i + \pi_j}} \right] = Ln \left[ \frac{\pi_i}{\pi_j} \right], \quad \forall i, j = 1, \dots, k (i \neq j) \quad (1.16)$$

Que representa el logaritmo de la ventaja de respuesta  $Y_i$  frente a  $Y_j$  condicionado a las observaciones de las variables independientes que caen en uno de ambos niveles. Pero para construir el modelo logit de respuesta multinomial bastaría con considerar  $(K - 1)$  transformaciones logit básicas, definidas con respecto a una categoría de referencia. Tomando como categoría de referencia la última  $Y_k$ . Así las transformaciones logit generalizadas se definen como  $logit_j(x_{ij}) = Ln \left[ \frac{\pi_{ij}}{\pi_{ik}} \right]$ ,  $\forall j = 1, \dots, (k - 1)$ , siendo  $logit_j(x)$  el logaritmo de la ventaja de respuesta  $Y_j$  dado que las observaciones de las variables independientes  $X$  que caen en la categoría  $Y_j$ .

El modelo lineal para cada una de las transformaciones logit generalizado, para  $p$  variables explicativas, es de la siguiente forma:

$$\text{logit}_j(x) = \sum_{r=0}^p b_{jr} X_{ijr} = X_{ij} b_j, \quad \forall j = 1, \dots, (k-1) \quad (1.17)$$

Para cada vector de valores observados de las variables explicativas  $x_{ij} = (1 \ x_{ij1} \ x_{ij2} \ \dots \ x_{ijp})$  y  $\beta_j = (\beta_{0j}, \beta_{1j}, \beta_{2j}, \dots, \beta_{pj})^t$  es el vector de parámetros asociados a la categoría  $Y_j$ .

Para las probabilidades de respuesta, podemos escribir el modelo de la siguiente forma:

$$\pi_{ij} = \frac{e^{\left(\sum_{r=0}^p \beta_{jr} x_{ijr}\right)}}{1 + \sum_{j=1}^{k-1} e^{\left(\sum_{r=0}^p \beta_{jr} x_{ijr}\right)}}, \quad \forall j = 1, \dots, (k-1) \quad (1.18)$$

Entonces para encontrar la probabilidad de respuesta para la categoría tomada como referencia  $\pi_{ik}$ , se ocupa la propiedad que  $\sum_{j=1}^k \pi_{ij} = 1$ .

Por tanto:

$$\begin{aligned} \pi_{ik} &= 1 - \sum_{j=1}^{k-1} \pi_{ij} \\ \pi_{ik} &= 1 - \left( \frac{e^{\left(\sum_{r=0}^p \beta_{j1} x_{ijr}\right)}}{1 + \sum_{j=1}^{k-1} e^{\left(\sum_{r=0}^p \beta_{jk} x_{ijr}\right)}} + \frac{e^{\left(\sum_{r=0}^p \beta_{j2} x_{ijr}\right)}}{1 + \sum_{j=1}^{k-1} e^{\left(\sum_{r=0}^p \beta_{jk} x_{ijr}\right)}} + \dots + \frac{e^{\left(\sum_{r=0}^p \beta_{j(k-1)} x_{ijr}\right)}}{1 + \sum_{j=1}^{k-1} e^{\left(\sum_{r=0}^p \beta_{jk} x_{ijr}\right)}} \right) \\ \pi_{ik} &= \frac{1 + \sum_{j=1}^{k-1} e^{\left(\sum_{r=0}^p \beta_{jk} x_{ijr}\right)}}{1 + \sum_{j=1}^{k-1} e^{\left(\sum_{r=0}^p \beta_{jk} x_{ijr}\right)}} - \left( \frac{e^{\left(\sum_{r=0}^p \beta_{j1} x_{ijr}\right)}}{1 + \sum_{j=1}^{k-1} e^{\left(\sum_{r=0}^p \beta_{jk} x_{ijr}\right)}} + \frac{e^{\left(\sum_{r=0}^p \beta_{j2} x_{ijr}\right)}}{1 + \sum_{j=1}^{k-1} e^{\left(\sum_{r=0}^p \beta_{jk} x_{ijr}\right)}} + \dots + \frac{e^{\left(\sum_{r=0}^p \beta_{j(k-1)} x_{ijr}\right)}}{1 + \sum_{j=1}^{k-1} e^{\left(\sum_{r=0}^p \beta_{jk} x_{ijr}\right)}} \right) \\ \pi_{ik} &= \frac{1 + \sum_{j=1}^{k-1} e^{\left(\sum_{r=0}^p \beta_{jk} x_{ijr}\right)} - e^{\left(\sum_{r=0}^p \beta_{j1} x_{ijr}\right)} - e^{\left(\sum_{r=0}^p \beta_{j2} x_{ijr}\right)} - \dots - e^{\left(\sum_{r=0}^p \beta_{j(k-1)} x_{ijr}\right)}}{1 + \sum_{j=1}^{k-1} e^{\left(\sum_{r=0}^p \beta_{jk} x_{ijr}\right)}} \end{aligned}$$

$$\pi_{ik} = \frac{1 + \sum_{j=1}^{k-1} e^{\left(\sum_{r=0}^p \beta_{jk} x_{ijr}\right)} - \sum_{j=1}^{k-1} e^{\left(\sum_{r=0}^p \beta_{jk} x_{ijr}\right)}}{1 + \sum_{j=1}^{k-1} e^{\left(\sum_{r=0}^p \beta_{jk} x_{ijr}\right)}}$$

Eliminando términos semejantes en la diferencia, se tiene que:

$$\pi_{ik} = \frac{1}{1 + \sum_{j=1}^{k-1} e^{\left(\sum_{r=0}^p \beta_{jk} x_{ijr}\right)}} \quad (1.19)$$

Por ejemplo, para una variable dicotómica supongamos que tanto la variable respuesta Y, como la variable independiente X son variables dicotómicas y codificadas con 1 y 0, además, en el caso de la variable respuesta, el valor 1 es la probabilidad de que ocurra el evento y el valor 0 la probabilidad de que no ocurra el evento. En la Tabla 1.8 aparecen las probabilidades de  $\pi(x)$  y  $1 - \pi(x)$ .

**Tabla 1.8 Valores del modelo de regresión logística cuando la variable independiente es dicotómica.**

Variable Respuesta Y	Variable Independiente X	
	x = 1	x = 0
y = 1	$\pi(1) = \frac{e^{\beta_0 + \beta_1}}{1 + e^{\beta_0 + \beta_1}}$	$\pi(0) = \frac{e^{\beta_0}}{1 + e^{\beta_0}}$
y = 0	$1 - \pi(1) = \frac{1}{1 + e^{\beta_0 + \beta_1}}$	$1 - \pi(0) = \frac{1}{1 + e^{\beta_0}}$
Total	1	1

#### 1.3.4.1. COEFICIENTE DE PROBABILIDAD Y PROBABILIDAD COMPLEMENTARIA (ODDS)

El coeficiente entre una probabilidad y la probabilidad complementaria se denomina odds, e indica, en este caso cuantas veces más probable es que ocurra un evento a que

no ocurra (Silva, 1995). Existe una relación simple entre probabilidades y odds. Si  $\pi(x)$  es la probabilidad de un evento y  $O$  es la odds del evento, entonces:

$$O = \frac{\pi(x)}{1 - \pi(x)} \quad (1.20)$$

$$\pi(x) = \frac{O}{1 + O} \quad (1.21)$$

La utilización de odds se debe a que es una escala más sensible para realizar comparaciones. Por ejemplo, si una persona A tiene una probabilidad de 0.30 de donar y una persona B tiene una probabilidad de 0.60, es razonable pensar que la probabilidad de la persona B es el doble que la de la persona A. Pero si la probabilidad de la persona A es 0.60, es imposible que la probabilidad de la persona B sea el doble, es decir 1.20. En este sentido no hay ningún problema si se trabaja con la escala de odds, ya que una probabilidad de 0,60 le corresponde una odds de  $0,60/0,40 = 1,5$ , cuyo doble 3. Volviendo este valor, la probabilidad nos da un valor de  $3/(1+3)=0.75$ , para efectos de observar la relación entre la probabilidad y la ODDS se presenta la siguiente Tabla 1.9.

**Tabla 1.9 Relación entre probabilidad y Odds.**

<b>Probabilidad</b>	<b>Odds</b>
<b>0.1</b>	0.12
<b>0.2</b>	0.25
<b>0.3</b>	0.43
<b>0.4</b>	0.67
<b>0.5</b>	1.00
<b>0.6</b>	1.50
<b>0.7</b>	2.33
<b>0.8</b>	4.00
<b>0.9</b>	9.00

La lectura de la Tabla 1.9 muestra que cuando un individuo tiene una probabilidad de donar de 0.1 resulta que el odds toma el valor de 0.12, lo que indica que el individuo

tiene poca probabilidad de donar ( $0.12 < 1$ ), ahora observemos el caso en el cual el individuo tiene una probabilidad de donar de 0.6, ocurre que ahora el número de veces que está dispuesto a donar ha aumentado a 1.5 veces más (que cuando tenía una probabilidad de donar de 0.1).

En el caso de la variable independiente dicotómica de la Tabla 1.8 las odds correspondientes son:

$$\text{la odds cuando } x=1 \text{ es } \frac{\pi(1)}{1-\pi(1)} = \frac{\frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}}{\frac{1}{1+e^{\beta_0+\beta_1}}} = e^{\beta_0+\beta_1} \quad (1.22)$$

$$\text{la odds cuando } x=0 \text{ es } \frac{\pi(0)}{1-\pi(0)} = \frac{\frac{e^{\beta_0}}{1+e^{\beta_0}}}{\frac{1}{1+e^{\beta_0}}} = e^{\beta_0} \quad (1.23)$$

Si  $\pi(1)=0,75$ , la odds de que ocurra el evento entre los sujetos con  $x=1$  será  $odds = 0,75 / 0,25 = 3$  (es decir, de cada cuatro personas con  $x=1$ , tres de ellas presentarán el estado  $y=1$ ).

Si  $\pi(1)=0,25$ , la odds de que ocurra el evento entre los sujetos con  $x=1$  será  $odds = 0,25 / 0,75 = 0,3$ .

Un inconveniente de la odds es que su rango oscila entre  $[0, +\infty]$ .

Si la odds  $> 1$ , el evento de un éxito es más probable en lugar de un fracaso. Si la odds  $< 1$ , el que se de el evento de un fracaso es más probable a que se de el evento de un éxito.

El valor de  $e^{\beta_1}$  es el factor por el que la odds cambia cuando la variable independiente se incrementa una unidad. Si  $\beta_1$  es positivo, el factor será mayor que 1, lo que significa que la odds aumenta; si  $\beta_1$  es negativo, el factor será menor que 1, en consecuencia la odds decrece. Cuando  $\beta_1$  es 0, el factor es igual a 1, lo que hace que la odds permanezca constante.

Tal y como se ha visto el logaritmo de la odds, llamado logit será igual a:

$$g(1) = \ln \left[ \frac{\pi(1)}{1 - \pi(1)} \right] = \beta_0 + \beta_1 \quad (1.24)$$

$$g(0) = \ln \left[ \frac{\pi(0)}{1 - \pi(0)} \right] = \beta_0 \quad (1.25)$$

#### 1.3.4.2. RAZÓN DE ODDS

La *razón de odds* (“odds ratio”), llámese  $\Psi$  al cociente de dos odds, por lo que siguiendo la nomenclatura indicada en el apartado anterior, Ato y López (1996) la traducen como *cociente de razones* y Díaz (1999) por *razón de razones*. Hosmer y Lemeshow (1989) señala que es una medida de asociación que se utiliza frecuentemente en epidemiología y su expresión es la siguiente:

$$\Psi = \frac{\pi(1) / [1 - \pi(1)]}{\pi(0) / [1 - \pi(0)]} \quad (1.26)$$

El logaritmo de una razón de odds o también *log odds*, es igual a:

$$\begin{aligned} \ln(\Psi) &= \ln \left[ \frac{\pi(1) / [1 - \pi(1)]}{\pi(0) / [1 - \pi(0)]} \right] \\ &= g(1) - g(0) \end{aligned} \quad (1.27)$$

El resultado de dicha ecuación es conocida como diferencia logit.

Si se desarrolla la ecuación (1.26) con las expresiones de la tabla 1.5 la razón de odds es igual a:

$$\Psi = \frac{\left( \frac{e^{\beta_0 + \beta_1}}{1 + e^{\beta_0 + \beta_1}} \right) \left( \frac{1}{1 + e^{\beta_0}} \right)}{\left( \frac{e^{\beta_0}}{1 + e^{\beta_0}} \right) \left( \frac{1}{1 + e^{\beta_0 + \beta_1}} \right)} = \frac{e^{\beta_0 + \beta_1}}{e^{\beta_0}} = e^{\beta_1} \quad (1.28)$$

Por lo que el *log odds* será:

$$\ln(\Psi) = \ln(e^{\beta_1}) = \beta_1 \quad (1.29)$$

### 1.3.5. INTERPRETACIÓN DEL MODELO DE REGRESIÓN LOGÍSTICA MULTINOMIAL

Para la interpretación de los parámetros del modelo de regresión logística simple y múltiple, se abordará en función de número y tipo de medida de la/s variable/s independiente/s.

#### 1.3.5.1. VARIABLE INDEPENDIENTE DICOTÓMICA

Tal y como se expone anteriormente, el valor de  $e^{\beta_1}$  es el factor por el que se multiplica la odds de la variable respuesta cuando la variable independiente se incrementa una unidad. Si  $\beta_1$  es positivo, el factor será mayor que 1, lo que significa que la odds aumenta; si  $\beta_1$  es negativo, el factor será menor que 1, en consecuencia la odds decrece. Cuando  $\beta_1$  es 0, el factor es igual a 1, lo que hace que la odds no cambie [11].

Hay que destacar que la interpretación de los coeficientes va a depender del tipo de estudio que se haya realizado:

- En un estudio de cohortes o de un ensayo clínico, las proporciones estiman incidencias acumuladas o riesgos, y el valor de  $e^{\beta_1}$  puede ser interpretado como una medida del efecto del factor  $X$  sobre la respuesta  $Y$ , en esta situación la razón de odds se denomina *razón de odds de riesgo*, ROR (“Risk Odds Ratio”).
- En un estudio de casos y controles, la razón de odds puede interpretarse como una estimación del *riesgo relativo* (RR) si la respuesta ( $Y=1$ ) tiene baja frecuencia. Para Kleinbaum (1994) esta forma de calcular el RR es una forma directa. Hosmer y Lemshow (1989) denominan a este parámetro como:

$$\eta = \frac{\pi(1)}{\pi(0)} \quad (1.30)$$

y, de acuerdo con la ecuación (1.28),  $\Psi \approx \eta$  en caso de que se de  $[1 - \pi(0)/1 - \pi(1)] \approx 1$ , lo cual ocurre cuando  $\pi(x)$  sea pequeño tanto para  $x = 1$  como  $x = 0$ .

- En un estudio transversal las odds estiman prevalencias, y en este caso la razón de odds se denomina *razón de odds de prevalencia*, POR (“Prevalece Odds Ratio”).

La estimación del parámetro de la razón de odds,  $\hat{\Psi}$ , suele tener distribución asimétrica; en teoría, si la muestra es suficientemente grande, se puede suponer que la distribución muestral del parámetro sigue una ley normal, debido a que el tamaño de la muestra debe ser alto, la alternativa consiste en inferir a partir de la distribución muestral,  $\ln(\hat{\Psi}) = \hat{\beta}_1$ . El intervalo de confianza determinado viene dado por la expresión:

$$\exp\left[\hat{\beta}_1 \pm (Z_{1-\alpha/2})SE(\hat{\beta}_1)\right] \quad (1.31)$$

### 1.3.5.2. VARIABLE INDEPENDIENTE POLITÓMICA

En el caso que algunas de las variables independientes sea categórica, con  $k > 2$  categorías, se debe transformar en un conjunto de  $k - 1$  variables ficticias o variables de diseño, también denominadas *dummy*, que introducidas simultáneamente en la ecuación aportan la misma información que la variable categórica (Pérez, 1996). Los métodos de codificación de las variables que más se utilizan en regresión logística son:

- Codificación respecto a una categoría de referencia. Se utiliza en el caso que tenga sentido clínico considerar una de las categorías como referencia o cuando se desea estimar razones de odds. El coeficiente de regresión de cada variable ficticia refleja el contraste entre la categoría que representa la variable ficticia y la categoría de

referencia. Hosmer y Lemeshow (1989) señalan que este es el método más utilizado en la regresión logística.

- Codificación mediante polinomios ortogonales. Se utiliza para variables ordinales. El coeficiente de regresión de la primera variable ficticia refleja la tendencia lineal del conjunto de las categorías, el coeficiente de regresión de la segunda variable ficticia refleja la tendencia cuadrática, y así sucesivamente.

Supongamos que tenemos la variable categórica  $A$  con categorías  $A_1, A_2, \dots, A_p$ , si de esta variable realizamos la transformación a variables de diseño mediante el método parcial que asigna un uno a la variable asociada a cada categoría y un cero al resto, y tomando como categoría de referencia la primera, obtenemos  $p-1$  variables que las denotaremos como  $X_{ij}^A = (x_{ij1}, x_{ij2}, \dots, x_{ij(p-1)})$ .

Así, el modelo de regresión logística multinomial generalizado que obtenemos sigue siendo un modelo lineal, como en los casos anteriores, para cada logit generalizado en función de esas variables de diseño procedentes de la variable  $X$  viene dado por:

$$\text{logit}(x_{ij}) = \ln \left[ \frac{\pi_{ij}}{\pi_{ik}} \right] = \beta_{0j} + \sum_{m=1}^p \tau_{jm}^A X_{ijm}^A ; i = 1, \dots, n; j = 1, \dots, k-1 \quad (1.32)$$

Siendo  $\pi_{ij} = P(Y = Y_j | A = X_{ij}^A)$  la probabilidad de respuesta  $Y_j$  en la categoría  $A_j$ .

### 1.3.5.3. VARIABLE INDEPENDIENTE CUANTITATIVA

Cuando en el modelo de regresión logística exista una variable independiente cuantitativa, la interpretación del coeficiente estimado va a depender de la forma en que es introducida en el modelo y de las unidades de medida de la variable [11].

Supóngase un modelo de regresión logística simple con una variable independiente cuantitativa,  $x$ . Si asumimos que el logit es lineal en la variable, el logit será  $g(x) = \beta_0 + \beta_1 x$ . El valor de  $e^{\beta_1}$  indica el cambio en la razón de odds por incremento de unidad en  $x$ , aumentando  $e^{\beta_1}$  si  $\beta_1$  es positivo y disminuyendo  $e^{\beta_1}$  si  $\beta_1$  es negativo. Otra interpretación en términos porcentuales sería a través de la transformación siguiente:

$$100 \cdot (e^{\beta_1}) \quad (1.33)$$

donde se obtiene el porcentaje de cambio en el logit para un incremento en una unidad en la variable independiente cuantitativa  $x$ .

Considérese el caso de un modelo de regresión logística múltiple, los cocientes de ventaja se definen incrementando una de las variables y manteniendo fijas las demás.

$$odds\_ratio_j(\Delta X_r = 1 | X_s = x_s, s \neq r) = \frac{\frac{P(Y = Y_j | X_r = x_r + 1, X_s = x_s, s \neq r)}{P(Y = Y_k | X_r = x_r + 1, X_s = x_s, s \neq r)}}{\frac{P(Y = Y_j | X_r = x_r, X_s = x_s, s \neq r)}{P(Y = Y_k | X_r = x_r, X_s = x_s, s \neq r)}}$$

Sabiendo que:

$$P(Y = Y_j | X_r = x_r + 1, X_s = x_s, s \neq r) = \frac{e^{\beta_{0j} + \beta_{1j} X_1 + \dots + \beta_{rj}(X_r + 1) + \beta_{sj} X_s + \dots + \beta_{n(k-1)j} X_n}}{1 + \sum_{j=0}^{k-1} e^{\beta_{0j} + \beta_{1j} X_1 + \dots + \beta_{rj}(X_r + 1) + \beta_{sj} X_s + \dots + \beta_{n(k-1)j} X_n}}$$

$$P(Y = Y_k | X_r = x_r + 1, X_s = x_s, s \neq r) = \frac{1}{1 + \sum_{j=0}^{k-1} e^{\beta_{0j} + \beta_{1j} X_1 + \dots + \beta_{rj}(X_r + 1) + \beta_{sj} X_s + \dots + \beta_{n(k-1)j} X_n}}$$

$$P(Y = Y_j | X_r = x_r, X_s = x_s, s \neq r) = \frac{e^{\beta_{0j} + \beta_{1j} X_1 + \dots + \beta_{rj} X_r + \beta_{sj} X_s + \dots + \beta_{n(k-1)j} X_n}}{1 + \sum_{j=0}^{k-1} e^{\beta_{0j} + \beta_{1j} X_1 + \dots + \beta_{rj} X_r + \beta_{sj} X_s + \dots + \beta_{n(k-1)j} X_n}}$$

$$P(Y = Y_k | X_r = x_r, X_s = x_s, s \neq r) = \frac{1}{1 + \sum_{j=0}^{k-1} e^{\beta_{0j} + \beta_{1j}X_1 + \dots + \beta_{rj}X_r + \beta_{sj}X_s + \dots + \beta_{n(k-1)}X_n}}$$

Luego:

$$\frac{P(Y = Y_j | X_r = x_r, X_s = x_s, s \neq r)}{P(Y = Y_k | X_r = x_r, X_s = x_s, s \neq r)} =$$

$$\frac{e^{\beta_{0j} + \beta_{1j}X_1 + \dots + \beta_{rj}X_r + \beta_{sj}X_s + \dots + \beta_{n(k-1)}X_n} \left( 1 + \sum_{j=0}^{k-1} e^{\beta_{0j} + \beta_{1j}X_1 + \dots + \beta_{rj}X_r + \beta_{sj}X_s + \dots + \beta_{n(k-1)}X_n} \right)}{1 + \sum_{j=0}^{k-1} e^{\beta_{0j} + \beta_{1j}X_1 + \dots + \beta_{rj}X_r + \beta_{sj}X_s + \dots + \beta_{n(k-1)}X_n}}$$

$$\frac{P(Y = Y_j | X_r = x_r + 1, X_s = x_s, s \neq r)}{P(Y = Y_k | X_r = x_r + 1, X_s = x_s, s \neq r)} =$$

$$\frac{e^{\beta_{0j} + \beta_{1j}X_1 + \dots + \beta_{rj}(X_r + 1) + \beta_{sj}X_s + \dots + \beta_{n(k-1)}X_n} \left( 1 + \sum_{j=0}^{k-1} e^{\beta_{0j} + \beta_{1j}X_1 + \dots + \beta_{rj}(X_r + 1) + \beta_{sj}X_s + \dots + \beta_{n(k-1)}X_n} \right)}{1 + \sum_{j=0}^{k-1} e^{\beta_{0j} + \beta_{1j}X_1 + \dots + \beta_{rj}(X_r + 1) + \beta_{sj}X_s + \dots + \beta_{n(k-1)}X_n}}$$

Por tanto:

$$\text{odds\_ratio}_j(\Delta X_r = 1 | X_s = x_s, s \neq r) = \frac{e^{\beta_{0j} + \beta_{1j}X_1 + \dots + \beta_{rj}(X_r + 1) + \beta_{sj}X_s + \dots + \beta_{n(k-1)}X_n}}{e^{\beta_{0j} + \beta_{1j}X_1 + \dots + \beta_{rj}X_r + \beta_{sj}X_s + \dots + \beta_{n(k-1)}X_n}}$$

$$= e^{\beta_{rj}}, \quad \forall j = 1, \dots, k-1, \quad s = 1, \dots, p, \quad r = 1, \dots, p, \quad s \neq r$$

Siendo  $\text{odds\_ratio}_j(\Delta X_r = 1 | X_s = x_s, s \neq r)$  el cociente de ventajas de respuesta  $Y_j$  frente a la última categoría  $Y_k$  cuando aumentan en una unidad la variable  $X_r$  y las demás se mantienen fijas.

#### 1.3.5.4. INTERACCIÓN

En los modelos múltiples, es preciso comprobar la existencia de interacción, la cual es un fenómeno que implica un efecto no lineal entre dos o más variables incluidas en el modelo. Por ejemplo, el modelo de regresión logística con dos variables sin interacción es:

$$P(Y = 1 / X) = \pi(x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2)}} \quad (1.34)$$

Si se estima conveniente la incorporación de una tercera variable, que represente el producto de las ya incluidas, el modelo resultante es:

$$\pi(x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \lambda x_1 x_2)}} \quad (1.35)$$

donde el coeficiente  $\lambda$  representa el efecto de la interacción de  $x_1$  y  $x_2$  sobre  $\pi(x)$ . En este caso, puede observarse que el coeficiente de  $x_1$  ya no es una constante, sino que depende de  $x_2$ , ya que:

$$\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \lambda x_1 x_2 = \beta_0 + (\beta_1 + \lambda x_2) x_1 + \beta_2 x_2$$

es decir, el grado en que influye el aumento de  $x_1$  en una unidad (la razón de odds asociada a  $x_1$ ) es igual a:  $\exp(\beta_1 + \lambda x_2)$ .

Como se puede suponer, esta idea se extiende a más variables, donde se incorporan términos que involucran a tres o más variables. Como regla general, si el ajuste de un modelo incluye un término de un cierto orden, se deben incluir todos los de orden inferior.

### 1.3.6. ESTIMACIÓN DE LOS PARÁMETROS EN UN MODELO DE REGRESIÓN LOGÍSTICA MULTINOMIAL POR MÁXIMA VEROSIMILITUD.

Para la estimación de los coeficientes del modelo y de sus errores estándar se utiliza la estimación por máxima verosimilitud, es decir, estimaciones que hagan máxima la probabilidad de obtener los valores de la variable dependiente y proporcionada por los datos de la muestra. Al contrario de lo que ocurre con la estimación de los coeficientes de regresión lineal múltiple que se utiliza el método de los mínimos cuadrados, los cálculos para las estimaciones de los coeficientes de la regresión logística multinomial no son directos, hay que llevar a cabo métodos iterativos, como por ejemplo el método de Newton-Raphson, que se explicará posteriormente.

Al aplicar estos métodos además de obtener las estimaciones de los coeficientes de regresión, se obtienen sus errores estándar y las covarianzas entre las covariables del modelo.

A continuación describimos el método de estimación de máxima verosimilitud para el cálculo de los coeficientes del modelo de regresión logística multinomial.

Primero se muestra un ejemplo en la Tabla 1.10 de cómo es la estructura de datos de un modelo logístico multinomial.

**Tabla 1.10 Representación Multinomial.**

Unidad	CATEGORÍA											$m_i$			
	1	J			K-1										
	$Y_1$	$X_{11}$	...	$X_{1p}$	...	$Y_j$	$X_{j1}$	...	$X_{jp}$	...	$Y_{k-1}$	$X_{(k-1)1}$	...	$X_{(k-1)p}$	
1	$y_{11}$	$X_{111}$	...	$X_{11p}$	...	$y_{1j}$	$X_{1j1}$	...	$X_{1jp}$	...	$y_{1(k-1)}$	$X_{1(k-1)1}$	...	$X_{1(k-1)p}$	$\sum_{j=1}^k y_{1j} = m_1$
.		.....		.....			.....		.....			.....		.....	
i	$y_{i1}$	$X_{i11}$	...	$X_{i1p}$	...	$y_{ij}$	$X_{ij1}$	...	$X_{ijp}$	...	$y_{i(k-1)}$	$X_{i(k-1)1}$	...	$X_{i(k-1)p}$	$\sum_{j=1}^k y_{ij} = m_i$
.		.....		.....			.....		.....			.....		.....	
n	$y_{n1}$	$X_{n11}$	...	$X_{n1p}$	...	$y_{nj}$	$X_{nj1}$	...	$X_{njp}$	...	$y_{n(k-1)}$	$X_{n(k-1)1}$	...	$X_{n(k-1)p}$	$\sum_{j=1}^k y_{nj} = m_n$

Los vectores  $y_i = (y_{i1}, \dots, y_{i(k-1)})'$   $\forall i = 1, \dots, n$  siguen una distribución de probabilidad

multinomial independiente  $M_k(m_i; \pi_i)$ , con  $\pi_i = \begin{pmatrix} \pi_{i1} \\ \vdots \\ \pi_{i(k-1)} \end{pmatrix}$ , siendo

$\pi_{ij} = P(Y = Y_j | X = x_{ij})$  y verificando que  $\sum_{j=1}^k \pi_{ij} = 1$ , también  $\sum_{j=1}^k y_{ij} = m_i$

observaciones independientes de la variable respuesta poltómica  $Y$

Su función de densidad viene dada por:

$$f(y; \pi) = \frac{m_i!}{y_{i1}! \dots y_{ik}!} \pi_{i1}^{y_{i1}} \dots \pi_{ik}^{y_{ik}} \quad (1.36)$$

Esta misma función se puede reescribir aplicando logaritmo y exponencial, quedando de la siguiente forma.

$$f(y_i; \pi_i) = \exp \left\{ \ln \left( \frac{m_i!}{y_{i1}! \dots y_{ik}!} \pi_{i1}^{y_{i1}} \dots \pi_{ik}^{y_{ik}} \right) \right\}$$

Aplicando propiedad de logaritmos y conociendo que

$$y_{ik} = m_i - \sum_{j=1}^{k-1} y_{ij}$$

Se tiene:

$$f(y_i; \pi_i) = \exp \left\{ \sum_{j=1}^{k-1} y_{ij} \ln(\pi_{ij}) + \ln(\pi_{ik}) \left( m_i - \sum_{j=1}^{k-1} y_{ij} \right) + \ln \left( \frac{m_i!}{y_{i1}! \dots y_{ik}!} \right) \right\}$$

$$f(y_i; \pi_i) = \exp \left\{ \sum_{j=1}^{k-1} y_{ij} \ln(\pi_{ij}) + m_i \ln(\pi_{ik}) - \sum_{j=1}^{k-1} y_{ij} \ln(\pi_{ik}) + \ln \left( \frac{m_i!}{y_{i1}! \dots y_{ik}!} \right) \right\}$$

$$f(y_i; \pi_i) = \exp \left\{ \sum_{j=1}^{k-1} y_{ij} \ln \left( \frac{\pi_{ij}}{\pi_{ik}} \right) + m_i \ln(\pi_{ik}) + \ln \left( \frac{m_i!}{y_{i1}! \dots y_{ik}!} \right) \right\} \quad (1.37)$$

Donde

$$\pi_{ik} = 1 - \sum_{j=1}^{k-1} \pi_{ij} \text{ e } y_{ik} = m_i - \sum_{j=1}^{j-1} y_{ij} \text{ y } E[y_{ij}] = \mu_{ij} = m_i \pi_{ij}$$

$$C(y_1, \dots, y_{k-1}) = \ln \left( \frac{m_i!}{y_{i1}! \dots y_{ik}!} \right)$$

Con:

$$\theta_{ij} = \ln \left( \frac{\pi_{ij}}{\pi_{ik}} \right) = \beta_{0j} + \sum_{r=1}^p \beta_{jr} x_{ijr} \quad ; \quad \text{para } j = 1, \dots, k-1 \quad (1.38)$$

Es el modelo lineal general asociado a la categoría j. Además, aplicando exponencial a  $\theta_{ij}$  se tiene que:

$$e^{\theta_{ij}} = \frac{\pi_{ij}}{\pi_{ik}} \Rightarrow \pi_{ij} = \pi_{ik} e^{\theta_{ij}} \quad (1.39)$$

También, dado que  $\sum_{j=1}^k \pi_{ij} = 1$ , se tiene:

$$1 - \pi_{ik} = \sum_{j=1}^{k-1} \pi_{ij} \text{ Por (1.39)}$$

$$\sum_{j=1}^{k-1} \pi_{ij} = \pi_{ik} \sum_{j=1}^{k-1} e^{\theta_{ij}} \quad (1.40)$$

Por transitividad dado (1.39) y (1.40)

$$\begin{aligned} 1 - \pi_{ik} = \pi_{ik} \sum_{j=1}^{k-1} e^{\theta_{ij}} &\Leftrightarrow \left( 1 + \sum_{j=1}^{k-1} e^{\theta_{ij}} \right) \pi_{ik} = 1 \\ \Leftrightarrow \pi_{ik} &= \frac{1}{\left( 1 + \sum_{j=1}^{k-1} e^{\theta_{ij}} \right)} \end{aligned} \quad (1.41)$$

Donde  $\theta_{ij}$  es como 1.38. Por tanto de (1.39) y (1.41)

$$\pi_{ik} = \frac{1}{\left(1 + \sum_{j=1}^{k-1} e^{\theta_{ij}}\right)} e^{\theta_{ij}} \quad (1.42)$$

Siendo  $\theta_{ij}$  función de los parámetros  $\beta$ 's definido como en (1.37) y

$$b\left(\theta_{i1}, \dots, \theta_{i(k-1)}\right) = -m_i \ln\left(\frac{1}{\left(1 + \sum_{j=1}^{k-1} e^{\theta_{ij}}\right)}\right)$$

$$b\left(\theta_{i1}, \dots, \theta_{i(k-1)}\right) = -m \ln(1) + m_i \ln\left(1 + \sum_{j=1}^{k-1} e^{\theta_{ij}}\right) = m_i \ln\left(1 + \sum_{j=1}^{k-1} e^{\theta_{ij}}\right)$$

Como los  $\theta_{ij}$  están en función de los  $\beta$ 's pero los  $\pi_{ij}$  están en función de los  $\theta_{ij}$  se

tiene:  $\pi_{ij} = \pi_{ij}(\beta)$ ,  $i=1, \dots, n$ ,  $j=1, \dots, k-1$ , donde  $\beta = \begin{pmatrix} \beta_{0j} \\ \vdots \\ \beta_{pj} \end{pmatrix}$ .

Y la función de verosimilitud para (1.36) viene dada por:

$$L(\beta; y_1, \dots, y_n) = \prod_{i=1}^n \exp\left\{\sum_{j=1}^{k-1} y_{ij} \ln\left(\frac{\pi_{ij}(\beta)}{\pi_{ik}(\beta)}\right) + m_i \ln(\pi_{ik}(\beta)) + \ln\left(\frac{m_i!}{y_{i1}! \dots y_{ik}!}\right)\right\} \quad (1.43)$$

La función de log-verosimilitud es

$$\begin{aligned}
l &= l(\beta; y_1, \dots, y_n) = \ln L(\beta; y_1, \dots, y_n) \\
&= \sum_{i=1}^n \left( \sum_{j=1}^{k-1} y_{ij} \ln \left( \frac{\pi_{ij}(\beta)}{\pi_{ik}(\beta)} \right) + m_i \ln(\pi_{ik}(\beta)) + \ln \left( \frac{m_i!}{y_{i1}! \dots y_{ik}!} \right) \right) \\
&= \sum_{i=1}^n \left[ \sum_{j=1}^{k-1} y_{ij} \ln \left( \frac{\pi_{ij}}{\pi_{ik}} \right) + m_i \ln(\pi_{ik}) + \ln \left( \frac{m_i!}{y_{ij}! \dots y_{ik}!} \right) \right]
\end{aligned} \tag{1.44}$$

$$\begin{aligned}
&= \sum_{i=1}^n \left[ \sum_{j=1}^{k-1} y_{ij} \theta_{ij} + m_i \ln \left( \frac{1}{1 + \sum_{j=1}^{k-1} e^{\theta_{ij}}} \right) + \ln \left( \frac{m_i!}{y_{ij}! \dots y_{ik}!} \right) \right] \\
&= \sum_{i=1}^n \left[ \sum_{j=1}^{k-1} y_{ij} (\beta_{0j} + \sum_{r=1}^p \beta_{pj} x_{ijr}) + m_i \left[ \ln(1) - \ln \left( 1 + \sum_{j=1}^{k-1} e^{\theta_{ij}} \right) \right] + \ln \left( \frac{m_i!}{y_{ij}! \dots y_{ik}!} \right) \right] \\
l &= \sum_{i=1}^n \left\{ \sum_{j=1}^{k-1} y_{ij} (x_{ij} \beta_j) - m_i \ln \left( 1 + \sum_{j=1}^{k-1} \exp \{ x_{ij} \beta_j \} \right) + \ln \left( \frac{m_i!}{y_{i1}! \dots y_{ik}!} \right) \right\}
\end{aligned} \tag{1.45}$$

Donde:

$$\pi_{ij} = \pi_{ij}(\beta), \quad i = 1, \dots, n \quad j = 1, \dots, k-1$$

Para obtener las ecuaciones de verosimilitud se tiene que derivar (1.44) ó (1.45) con respecto al parámetro de interés e igualar a cero, teniendo en cuenta que:

$$\theta_{ij} = \ln \left( \frac{\pi_{ij}}{\pi_{ik}} \right) = \beta_{0j} + \sum_{s=1}^p \beta_{sj} x_{isj}$$

Entonces se deducen las ecuaciones de verosimilitud para los parámetros respectivos y definido de  $\pi_{ik}$  como en (1.42):

$$O = U_{0j} = \frac{\partial l}{\partial \beta_{0j}} \sum_{i=1}^n \left[ y_{ij} - m_i \frac{\exp\{x_{ij}\beta_j\}}{1 + \sum_{r=1}^{k-1} \exp\{x_{ir}\beta_r\}} \right] = \sum_{i=1}^n (y_{ij} - m_i \pi_{ij}) \quad (1.46)$$

$$j = 1, 2, \dots, k-1$$

$$O = U_{pj} = \frac{\partial l}{\partial \beta_{pj}} \sum_{i=1}^n x_{ijp} \left[ y_{ij} - m_i \frac{\exp\{x_{ij}\beta_j\}}{1 + \sum_{r=1}^{k-1} \exp\{x_{ir}\beta_r\}} \right] = \sum_{i=1}^n x_{ijp} (y_{ij} - m_i \pi_{ij}) \quad (1.47)$$

$$j = 1, 2, \dots, k-1, \quad p = 1, 2, \dots, r$$

Para obtener los estimadores de máxima verosimilitud hay que resolver  $k-1$  sistemas de  $p+1$  ecuaciones no lineales, que no se pueden resolver explícitamente. Es necesario acudir a métodos numéricos para obtener aproximaciones de  $\beta_j = (\beta_{0j}, \beta_{1j}, \dots, \beta_{pj})$ .

### 1.3.6.1 EL MÉTODO DE NEWTON-RAPHSON

Una de las herramientas que ayuda a resolver este problema es el algoritmo de Newton – Raphson, se apoya en la serie de Taylor. Si  $\beta^*$  es solución de las ecuaciones de verosimilitud; es decir:

$$U(\beta^*) = 0 \quad (1.48)$$

Y  $\beta^{(0)}$  es un valor arbitrario de  $\beta_j$ , ( $j=1, 2, \dots, p$ ); entonces el desarrollo de Taylor de primer orden garantiza la siguiente aproximación:

$$U(\beta^*) = 0 \cong U(\beta^{(0)}) + H(\beta^{(0)})(\beta^* - \beta^{(0)}) \quad \text{donde } H = \frac{\partial U}{\partial \beta} = \left( \frac{\partial^2 l}{\partial \beta_p \partial \beta_j} \right) \text{ es la matriz}$$

Hessiana. Despejando  $\beta^*$  de la aproximación, se obtiene  $\beta^* \cong \beta^{(0)} - H^{-1}(\beta^{(0)})U(\beta^{(0)})$ , que sirve para plantear la ecuación recurrente:

$$\hat{\beta}^{(r)} \cong \hat{\beta}^{(r-1)} - H^{-1}(\hat{\beta}^{(r-1)})U(\hat{\beta}^{(r-1)}) \quad (1.49)$$

Donde  $U = \left( \frac{\partial l}{\partial \beta_1}, \dots, \frac{\partial l}{\partial \beta_j} \right)$ , son las derivadas parciales con respecto a cada

parámetro j-ésimo y  $\hat{\beta}_j^{(r)} = (\hat{\beta}_{0j}^{(r)}, \dots, \hat{\beta}_{pj}^{(r)})^t$ ,  $H = \left( \frac{\partial^2 l}{\partial \beta_{ij} \partial \beta_{nj}} \right)_{(j=1,2,\dots,k; i,n=0,1,2,\dots,p)}$

H es conocida como la matriz Hessiana de orden  $p \times p$  y está formada por las segundas derivadas parciales respecto a los parámetros betas.

Además,  $\hat{\beta}^{(r)}$  es el valor estimado de  $\hat{\beta}$  en la r-ésima iteración del algoritmo y  $H^{-1}(\hat{\beta}^{(r-1)})$ ,  $U(\hat{\beta}^{(r-1)})$  son  $H^{-1}$  y  $U$  evaluadas en  $(\hat{\beta}^{(r-1)})$ .

### 1.3.7. SIGNIFICACIÓN DE LOS PARÁMETROS

Una vez estimados los coeficientes hay que proceder a ver cuáles de ellos son significativamente diferentes de 0. Para analizar si cada uno de ellos es cero se utiliza el test de Wald. Si el objetivo es comprobar que el conjunto de covariables elegidas explica el fenómeno a estudio, es decir, contrastar si todos los coeficientes son iguales a 0 o hay alguno distinto, se puede utilizar la razón de verosimilitud.

Así que nos planteamos contrastar si un subconjunto de los parámetros del modelo de regresión logística multinomial, que denotaremos por  $\beta = (\beta_1, \beta_2, \dots, \beta_r)'$ , es nulo, es decir, que se realiza para determinar dicha afirmación es necesario el siguiente contraste de hipótesis:

$$H_0 : \beta = 0$$

$$H_1 : \beta \neq 0$$

Dos tipos de contrastes son los utilizados para decidir si rechazamos o no la hipótesis nula, que se detallan a continuación.

### 1.3.7.1. CONTRASTES CONDICIONALES DE LA RAZÓN DE VEROSIMILITUD

Se trata de ir contrastando cada modelo que surge de eliminar de forma aislada cada una de las covariables o factores frente al modelo completo. La ausencia de significación implica que el modelo sin la covariable no empeora respecto al modelo completo (Es decir, da igual su presencia que su ausencia), por lo que según la estrategia de obtención del modelo más reducido (principio de parsimonia) dicha covariable o factor debe ser eliminada del modelo ya que no aporta nada al mismo.

Supongamos que tenemos un modelo de regresión logística multinomial  $M_G$  que se ajusta y se desea contrastar si un subconjunto de parámetros es nulo. Sea  $\beta = \beta_1 \cup \beta_2$  donde la longitud del vector  $\beta_1$  es  $p_1$  y la longitud del vector  $\beta_2$  es  $p_2$  y  $p_1 + p_2 = p$  que es la longitud de  $\beta$ ; deseamos contrastar la siguiente hipótesis:

$$\begin{aligned}H_0: \beta_2 &= 0 \\H_1: \beta_2 &\neq 0\end{aligned}$$

Si asumimos que  $H_0$  es cierta, el estadístico del test de razón de verosimilitud para contrastar si  $H_0$  es:  $G^2(\beta_1 | \beta) = -2(L_{\beta_1} - L_{\beta}) = G^2(\beta_1) - G^2(\beta)$ , siendo  $L_{\beta_1}$  y  $L_{\beta}$  los máximos de la log-verosimilitud bajo la suposición de ser cierta  $H_0$ .

El estadístico  $G^2(\beta_1 | \beta)$  tiene distribución Chi – cuadrado con grados de libertad como la diferencia entre los grados de libertad de las distribuciones Chi – cuadrado asintóticas de,  $G^2(\beta_1)$  y  $G^2(\beta)$  es decir, los grados de libertad (*gl*) serían igual a la longitud de  $\beta$  menos la longitud de  $\beta_1$  es decir  $p - p_1$ .

Así que se rechaza la hipótesis nula al nivel de significancia  $\alpha$  cuando  $G^2(\beta_1 | \beta) \geq \chi^2_{p-p_1; \alpha}$

### 1.3.7.2. EL TEST DE WALD

Wald (1943; citado por Doménech, 1999), demostró que las distribuciones muestrales de las estimaciones máximo verosímiles de los parámetros, en el caso de muestras

grandes, se distribuyen según la curva normal. Por lo tanto, la significación de los parámetros puede estudiarse mediante el estadístico  $z = \hat{\beta}/SE(\hat{\beta})$ , el cual sigue una ley normal estandarizada bajo el supuesto de que  $H_0 : \beta = 0$ , o mediante el estadístico de Wald, es decir el cuadrado de ese cociente, el cual sigue una ley de ji-cuadrado, con 1 grado de libertad, según la fórmula:

$$\text{Wald} = \left[ \frac{\hat{\beta}}{SE(\hat{\beta})} \right]^2 \quad (1.50)$$

El estadístico de Wald tiene un gran problema respecto a la falta de potencia de esta prueba cuando el valor del parámetro  $\hat{\beta}$  se aleja de cero, es decir, al ser un valor cuadrático puede producir valores muy pequeños en presencia de coeficientes de regresión muy altos, por lo que diversos autores, como por ejemplo Ato y López (1996) recomiendan la utilización de la prueba de la razón de verosimilitud.

Así que se rechaza la hipótesis nula al nivel de significación  $\alpha$ , cuando el valor observado de este estadístico sea mayor o igual que el cuantil de orden  $(1-\alpha)$  de la distribución  $\chi_1^2$ .

### **1.3.8. INFERENCIA EN REGRESIÓN LOGÍSTICA MULTINOMIAL**

Lo principal que se pretende cuando se realiza un modelo estadístico a través de los datos procedentes de una muestra, es extrapolar los resultados muestrales a la población general, es por ello que, para este caso en particular después de haber estimado los parámetros del modelo de regresión logística multinomial, se pretende hacer inferencia.

### 1.3.8.1. INTERVALOS DE CONFIANZA

Basándonos en la normalidad asintótica de los estimadores de máxima verosimilitud se pueden construir intervalos de confianza asintóticos para cada uno de los parámetros del modelo, utilizando la distribución normal, y mediante las transformaciones correspondientes, intervalos de confianza para los odds\_ratio.

#### ✓ Construcción de los intervalos de confianza para los parámetros.

Se construye un intervalo de confianza  $(1-\alpha)$  para cada parámetro del modelo de regresión logística multinomial,  $\beta_{sj}$  con  $j=1,\dots,k-1$ ,  $s=0,\dots,p$ . La distribución asintótica de  $\hat{\beta}_{sj}$  es  $N(\beta_{sj}, \hat{\sigma}^2(\hat{\beta}_{sj}))$ , donde  $\hat{\sigma}^2(\hat{\beta}_{sj})$  es el valor correspondiente a la varianza del  $\hat{\beta}_{sj}$ .

Así tenemos que:  $P\left[-Z_{\alpha/2} \leq \frac{\hat{\beta}_{sj} - \beta_{sj}}{\hat{\sigma}(\hat{\beta}_{sj})} \leq Z_{\alpha/2}\right] = 1-\alpha$ . Por lo que obtenemos así el intervalo de confianza aproximado para  $\beta_{sj}$  al nivel  $(1-\alpha)$ :

$$IC_{1-\alpha}(\beta_{sj}) = \left(\hat{\beta}_{sj} \pm Z_{\alpha/2} \hat{\sigma}(\hat{\beta}_{sj})\right) \quad (1.51)$$

#### Construcción de intervalos de confianza para los odds\_ratio

Sabemos que los cocientes de ventajas vienen dados por:

$$odds\_ratio_j(\Delta X_r = 1 / X_s = x_s, s \neq r) = \exp(\beta_{rj}), \quad \forall j=1,\dots,k-1, \quad r=1,\dots,p$$

Por lo tanto, el intervalo de confianza para los cocientes de ventajas se calcula tomando exponenciales en el intervalo de confianza obtenido anteriormente para cada uno de los  $\beta_{pj}$  al nivel de confianza  $1-\alpha$ , el cual viene dado por:

$$IC_{1-\alpha}(Odds\ ratio_j) = IC_{1-\alpha}(\exp(\beta_{pj})) = \exp\left(\hat{\beta}_{pj} \pm Z_{\alpha/2} \hat{\sigma}(\hat{\beta}_{pj})\right) \quad (1.52)$$

### 1.3.9. BONDAD DE AJUSTE DEL MODELO

Uno de los primeros indicadores de importancia para apreciar el ajuste del modelo logístico multinomial es el doble logaritmo del estadístico de verosimilitud (likelihood), que se abordará posteriormente. Se trata de un estadístico que sigue una distribución similar a  $X^2$ .

Sea  $y_{ij}$  el número de observaciones que caen en la categoría de respuesta  $Y_j$   $\forall j = 1, \dots, k$  y sean las  $m_i$  observaciones correspondientes a la  $i$ -ésima combinación de valores de las variables explicativas.

Al denotar a  $\hat{d}_{ij}$  como la frecuencia esperada de la variable respuesta  $Y_j$  en la combinación  $X_i$  de valores observados de las variables predictoras, estimada bajo el modelo y definida como  $\hat{d}_{ij} = m_i \hat{\pi}_{ij}$ , siendo  $\hat{\pi}_{ij}$  el estimado por máxima verosimilitud de  $\pi_{ij}$ .

Así que, para contrastar la bondad del ajuste global del modelo, cuando el número de observaciones en cada combinación de valores de las variables explicativas es grande se utiliza el estadístico Chi-Cuadrado de Pearson y el estadístico de Wilks de razón de verosimilitudes. El test global de bondad de ajuste del modelo de regresión logística multinomial múltiple contrasta las siguientes hipótesis:

$$H_0 : \pi_{ij} = \frac{e^{\left( \sum_{r=0}^p \beta_{jr} x_{ijr} \right)}}{1 + \sum_{j=1}^{k-1} e^{\sum_{r=0}^p \beta_{jr} x_{ijr}}} \quad \forall i = 1, \dots, n ; \quad \forall j = 1, \dots, k$$

$$H_1 : \pi_{ij} \neq \frac{e^{\left( \sum_{r=0}^p \beta_{jr} x_{ijr} \right)}}{1 + \sum_{j=1}^{k-1} e^{\sum_{r=0}^p \beta_{jr} x_{ijr}}} \quad \text{para algún } i,r,j.$$

Con lo cual  $H_0$  trata de comprobar que los datos se pueden modelar con un ajuste de regresión logístico multinomial. Para la realización del contraste anterior se utilizan los siguientes test.

### 1.3.9.1. JI – CUADRADO DE PEARSON

El estadístico Chi-Cuadrado de Pearson de bondad de ajuste de un modelo de regresión logística multinomial M, viene dado por:

$$X^2(M) = \sum_{i=1}^n \sum_{j=1}^k \frac{(y_{ij} - m_i \hat{\pi}_{ij})^2}{m_i \hat{\pi}_{ij}} \quad (1.53)$$

Este estadístico tiene distribución asintótica Chi-Cuadrado con grados de libertad obtenidos como la diferencia entre el número de parámetros  $\pi_{ij}$  y el número de parámetros independientes en el modelo,  $(N-p)(k-1)$  [18]. Es decir,  $X^2(M) \xrightarrow{d} X^2_{(N-p)(k-1), 1-\alpha}$ , si  $m_i \rightarrow \infty$ .

Así que se rechaza la hipótesis nula con un nivel de significancia  $\alpha$  cuando  $X^2(M)_{obs} \leq X^2_{(N-p)(k-1), 1-\alpha}$  o equivalentemente se puede definir el p-valor del contraste como la probabilidad acumulada a la derecha del valor observado:  $p\text{-valor} = P[X^2(M)_{obs} \leq X^2_{(N-1)(k-1)}]$ , se rechaza la hipótesis nula cuando el p-valor  $\leq \alpha$ .

### 1.3.9.2. TEST CHI-CUADRADO DE RAZÓN DE VEROSIMILITUDES, ESTADISTICO DE WILKS. DESVIANZA.

Consiste en construir un test basado en la verosimilitud, para lo cual se calcula cuál es la lejanía o discrepancia que hay entre el modelo ajustado con la/s variable/s

predictora/s incluidas en el modelo  $l(\beta)$ , con el modelo saturado (o completo), es decir el que contiene todos los posibles efectos principales y todas las posibles combinaciones de las variables seleccionadas que lo componen y el modelo ajustado  $l(\hat{\beta})$  según la expresión siguiente:

$$\begin{aligned}
 D &= -2 \ln \left[ \frac{L(\hat{\beta})}{L(\beta)} \right] \\
 &= -2 \left[ \ln(L(\hat{\beta})) - \ln(L(\beta)) \right] \\
 &= -2 \left[ l(\hat{\beta}) - l(\beta) \right] \\
 &= 2 \left[ l(\beta) - l(\hat{\beta}) \right]
 \end{aligned} \tag{1.54}$$

La desviación ( $D$ ) sigue una distribución ji-cuadrado con tantos grados de libertad como la diferencia en grados de libertad entre ambos  $(N - p)(k - 1)$  [18]. Como medida de la discrepancia o bondad de ajuste existente entre los valores empíricos y ajustados, juega un papel similar en las estimaciones por máxima verosimilitud al de la suma de cuadrados de los residuos en las estimaciones por mínimos cuadrados. Al resultado del cociente entre corchetes, se le denomina *razón de verosimilitud*, hay una razón matemática en el cálculo del negativo de dos veces el logaritmo, y se debe a la necesidad de garantizar la distribución ji-cuadrado del test de contraste  $D$ , al test resultante se le denomina *prueba de la razón de verosimilitud*.

En la valoración de la significación de una variable independiente  $x_i$  se le compara el valor de  $D$  con y sin la variable independiente en el modelo, el cambio en  $D$  se debe a la exclusión de la variable independiente en el modelo y se obtiene con la siguiente fórmula:

$$G = -2 \ln \left[ \frac{l(\hat{\beta})}{l(\beta)} \right] \tag{1.55}$$

Bajo la hipótesis nula de que  $\hat{\beta}$  es igual a cero, el estadístico  $G$  sigue una distribución ji-cuadrado con un grado de libertad. Si el test es significativo indicará que la variable independiente añade información al modelo.

### 1.3.10. CAPACIDAD PREDICTIVA DEL MODELO

Además de los contrastes que se ha visto anteriormente, se puede calcular otras medidas que dan información sobre la calidad del modelo. Por ejemplo, en los modelos de regresión logística binaria, la calidad del ajuste se mide mediante coeficientes de determinación conocidos como Pseudo- $R^2$ , para la regresión logística multinomial también se utilizan otros coeficientes. De entre todos los que existen, los más usados son el de Mc-Fadden, el de Cox-Snell y el de Nagelkerke. A continuación se expone cómo se calcula cada uno de ellos.

#### 1.3.10.1. COEFICIENTE PSEUDO - $R^2$ DE MC-FADDEN.

Al tener  $\Lambda = -2\ln(L(\beta; y_1, \dots, y_n))$ , se identifica  $\Lambda_0$  como el valor inicial de esta función, es decir, el mínimo  $\Lambda$  bajo el modelo saturado, que está dado solo por un término constante es de la siguiente forma:

$$\Lambda_0 = -2\ln(L_0(\beta; y_1, \dots, y_n)) = -2\sum_{i=1}^n \left\{ \sum_{j=1}^{k-1} y_{ij} \ln\left(\frac{\pi_{ij}(\beta)}{\pi_{ik}(\beta)}\right) + m_i \ln(\pi_{ik}(\beta)) + \ln\left(\frac{m_i!}{y_{i1}! \dots y_{ik}!}\right) \right\}$$

$$\text{con } \theta_{ij} = \ln\left(\frac{\pi_{ij}}{\pi_{ik}}\right) = \beta_{0j}$$

Además, sea  $\Lambda_f$ , el máximo de  $\Lambda$  bajo el modelo ajustado con todos los parámetros, el cual viene definido de la siguiente forma:

$$\Lambda_f = -2\ln(L(\beta; y_1, \dots, y_n)) = -2\sum_{i=1}^n \left\{ \sum_{j=1}^{k-1} y_{ij} \ln\left(\frac{\pi_{ij}(\beta)}{\pi_{ik}(\beta)}\right) + m_i \ln(\pi_{ik}(\beta)) + \ln\left(\frac{m_i!}{y_{i1}! \dots y_{ik}!}\right) \right\}$$

$$\text{con } \theta_{ij} = \ln\left(\frac{\pi_{ij}}{\pi_{ik}}\right) = \beta_{0j} + \sum_{r=1}^p \beta_{rj} x_{irj}$$

Con estos resultados se tiene la siguiente expresión reconocida como Mc-Fadden:

$$R_{MF}^2 = 1 - \frac{\Lambda_f}{\Lambda_0} \quad (1.56)$$

Siendo su rango teórico de valores  $0 \leq R_{MF}^2 \leq 1$  pero muy raramente su valor se aproxima a 1. Suele considerarse una buena calidad del ajuste, cuando  $0.2 \leq R_{MF}^2 \leq 0.4$  y excelente para valores superiores, siendo malo para valores menores a 0.2.

### 1.3.10.2. COEFICIENTE PSEUDO - $R^2$ DE COX – SNELL

En este caso se utiliza directamente la función de verosimilitud  $L(\beta; y_1, \dots, y_n)$ , y no la función auxiliar  $\Lambda$ . Por lo que se denota  $V_0 = \exp\left(-\frac{\Lambda_0}{2}\right)$  como el máximo de verosimilitud bajo el modelo saturado, dado sólo por el término constante y por  $V_f = \exp\left(-\frac{\Lambda_f}{2}\right)$  el máximo de verosimilitud bajo el modelo ajustado con todos los parámetros. Entonces definimos el coeficiente pseudo -  $R^2$  de Cox-Snel como:

$$R_{CS}^2 = 1 - \left(\frac{V_0}{V_f}\right)^{\frac{2}{N}} = 1 - \exp\left(\frac{V_f - V_0}{N}\right) \quad (1.57)$$

El rango teórico de valores para el coeficiente es  $0 \leq R_{CS}^2 \leq V_0^{\frac{2}{N}}$  lo que le hace poco interpretable al depender de  $V_0$ . Ya que puede ser próximo a cero cuando hay pocos datos. Por ello, es preferible utilizar el siguiente coeficiente como medida de bondad

de ajuste; además, por esta razón Nagelkerke (1991, citado en Doménech, 1999) ha propuesto el índice corregido que vale 1 en caso de que el modelo explique el 100% de la incertidumbre de los datos.

### 1.3.10.3 COEFICIENTE PSEUDO - $R^2$ DE NAGELKERKE

Viene definido por la expresión:

$$R_N^2 = \frac{R_{CS}^2}{1 - V_0^{\frac{2}{N}}} = \frac{1 - \exp\left(\frac{\Lambda_f - \Lambda_0}{N}\right)}{1 - \exp\left(\frac{-\Lambda_0}{N}\right)} \quad (1.58)$$

Y en este caso, su rango de valores está comprendido entre  $0 \leq R_N^2 \leq 1$ , por lo que puede interpretarse del mismo modo que el coeficiente de determinación de la regresión lineal clásica, aunque, es más difícil que alcance valores cercanos a 1.

### 1.3.11. TASA DE CLASIFICACIONES CORRECTAS.

Para cuantificar la bondad del ajuste global del modelo, se dispone también de otra medida, como es, la tasa de clasificaciones correctas. Es decir, a partir del modelo ajustado, se clasifica cada observación en la categoría más probable, construyendo así una matriz de clasificación: **observados-predichos** y se utiliza el porcentaje de clasificaciones correctas como una medida de calidad de predicción, del mismo modo que se hace en el análisis discriminante. Entonces la tasa de clasificación correcta se define como la proporción de individuos clasificados correctamente por el modelo y se calcula como el cociente entre el número de observaciones clasificadas correctamente y el tamaño muestral N.

Un individuo es clasificado correctamente por el modelo cuando su valor observado de la variable respuesta  $Y(Y_1, Y_2, \dots, Y_k)$  coincide con su valor estimado por el modelo.

### **1.3.12. MÉTODOS DE SELECCIÓN DEL MODELO.**

Una vez conocido el procedimiento de ajuste de modelos de regresión logística multinomial, el siguiente paso es el desarrollo de estrategias para seleccionar las variables que mejor explican a la variable de respuesta. Para ello, se adoptará el principio de parsimonia, que consiste en seleccionar el modelo que con menor número de parámetros se ajuste bien a los datos y lleve a una interpretación sencilla en términos de cocientes de ventajas.

Hay que tener especial atención a las covariables cualitativas que se transforman en varias variables dummies. Siempre que se incluya o excluya una de estas variables, todas las demás categorías deben ser incluidas o excluidas en bloque.

Si no se tiene en cuenta esta consideración, implicaría que se habría recodificado la variable, y por tanto la interpretación de la misma no sería la correcta. Además, hay que tener en cuenta la significación que pudiera tener cada variable dummy. No siempre todas las categorías de una covariable son significativas, o todas no significativas. Por lo que, cuando ocurra esta situación, es recomendable contrastar el modelo completo frente al modelo sin la covariable mediante la prueba de razón de verosimilitud, decidiendo incluir o excluir la covariable dependiendo del resultado de la prueba. Si se obtiene significación en este contraste, la variable permanecería en el modelo. Dentro de los métodos de selección de las variables para formar el modelo se tienen la selección hacia adelante, selección hacia atrás y la selección paso a paso; con el objetivo de puntualizar cada método, se describen a continuación.

#### **1.3.12.1. SELECCIÓN HACIA ADELANTE.**

1. Se inicia con un modelo vacío (sólo constante).
2. Se ajusta un modelo y se calcula el p-valor del contraste de razón de verosimilitud que resulta de incluir cada variable por separado.
3. Se selecciona el modelo con el p-valor más significativo.

4. Se ajusta de nuevo un modelo con la(s) variable(s) seleccionada(s) y se calcula el p-valor de añadir cada variable no seleccionada anteriormente por separado
5. Se selecciona el modelo con el más p-valor significativo.
6. Se repite 4 a 5 veces hasta que no queden variables significativas para incluir.

#### **1.3.12.2. HACIA ATRÁS**

1. Se inicia con un modelo con todas las variables candidatas.
2. Se eliminan, una a una, cada variable y se calcula la pérdida de ajuste al eliminar dichas variables.
3. Se selecciona para eliminar la variable menos significativa.
4. Se repite 2 a 3 veces hasta que todas las variables incluidas sean significativas y no pueda eliminarse ninguna sin que se pierda ajuste.

#### **1.3.12.3. STEPWISE (PASO A PASO)**

En este método se combinan los métodos adelante y atrás. Puede empezarse por el modelo vacío o por el completo, pero en cada paso se exploran las variables incluidas, por si deben salir y las no seleccionadas, por si deben entrar. Pero no todos los métodos llegan a la misma solución necesariamente. El método de stepwise, está basado en contrastes condicionales de razón de verosimilitudes.

Al partir del modelo vacío, sólo con la constante, este método consiste en partir de ese modelo inicial, y en cada paso se ajustarán todos aquellos modelos que resultan de incluir cada una de las variables explicativas que no están en el modelo seleccionado en el paso anterior. Entonces se llevan a cabo contrastes condicionales de razón de verosimilitudes que tiene en la hipótesis nula el modelo seleccionado en el paso anterior y en la hipótesis alternativa el modelo resultante de la inclusión de cada variable. De este modo, se seleccionarán las variables para las que el contraste sea

significativo, y se incluiría en el modelo, aquella variable asociada al mínimo p-valor de entre todos los menores o iguales que  $\alpha_1$ . La inclusión de variable mediante este método continua, hasta que ninguno de estos contrastes condicionales sea significativo.

Por otra parte, a la misma vez, se considera en cada paso la posibilidad de eliminar alguno de los parámetros del modelo seleccionado en el paso anterior (método hacia atrás). Pero no se puede eliminar en un paso la variable que acaba de entrar en el paso anterior, por lo que se fijará para la eliminación de variables un nivel de significación  $\alpha_2$  mayor que  $\alpha_1$  al igual que antes, para la eliminación de variables se realizarán contrastes condicionales de razón de verosimilitudes que tienen en la hipótesis nula el modelo que resulta de la eliminación de cada variable y en la hipótesis alternativa el modelo seleccionado en el paso anterior. Así, las variables candidatas a eliminar serán aquellas cuyo p-valor sea mayor de  $\alpha_2$  y se eliminará la variable con el mayor p-valor de éstos. La eliminación de variables continúa hasta que todos estos contrastes condicionales resulten significativos. Finalmente, se llegará a un paso en el que ninguno de los contrastes condicionales de introducción de variables sean significativos y todos los de eliminación de variables sean significativos, es decir que ya no será necesario introducir más variables en el modelo.

## **CAPÍTULO II. ANÁLISIS DE PROPORCIONES Y REGRESIÓN LOGÍSTICA MULTINOMIAL PARA CLASIFICAR LA CONDICIÓN DE SALUD BUCAL DE LA POBLACIÓN SALVADOREÑA QUE ASISTIÓ A LAS UNIDADES DE SALUD EN EL AÑO 2009.**

### **INTRODUCCIÓN**

En el Capítulo II, se ha llevado a la práctica la teoría explicada en el Capítulo I sobre la regresión logística multinomial. De esta manera, se realizó previamente un estudio exploratorio con el fin de analizar algunas variables importantes, posteriormente, se estudió la relación entre ellas para cuantificar su nivel de relación (Análisis bidimensional) y así también analizar algunos contrastes que nos proporcionen respuestas a las hipótesis planteadas y por tanto cumplir con los objetivos específicos de la investigación. Finalmente se llevó a cabo la regresión logística multinomial para la clasificación de la condición de salud bucal en virtud de las características clínicas y socioeconómicas del paciente como desarrollo del objetivo general, para ello se tuvo en cuenta varios aspectos previos para la preparación de la base de datos, entre ellos se incluyen la selección tanto de la variable dependiente como las variables independientes, el tamaño de la muestra para la estimación del modelo y la partición de la muestra con fines de validación.

#### **2.1. ANÁLISIS ESTADÍSTICO EXPLORATORIO.**

En este apartado se realizó un análisis estadístico univariado de cada una de las variables en el estudio, con el fin de tener una idea general de la información contenida en los datos, dado que tanto la variable dependiente y las variables independientes, tienen una escala de medida categórica, el análisis se basa en tablas de distribución de frecuencias. Se ha realizado en dos categorías este apartado.

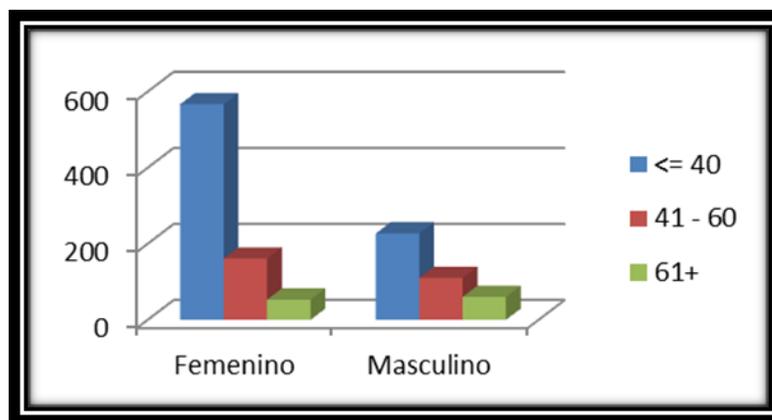
### 2.1.1. ANÁLISIS ESTADÍSTICO EXPLORATORIO DE LAS VARIABLES SOCIODEMOGRÁFICAS

En la Tabla de contingencia 2.1 se muestra la distribución de frecuencias de la relación entre las edades y el sexo, de los pacientes adultos que asistieron a las unidades de la red de salud pública de El Salvador. Se clasificaron en 3 intervalos de edades y se logró determinar que los adultos menores de 40 años son los que más acuden a recibir el servicio odontológico (67.4%). Además se puede contemplar en la lectura de la tabla 2.1, de cuantas mujeres y hombres recibieron el servicio y los resultados fueron 66.3% y 33.7% respectivamente (ver Tabla 2.1 y Grafico 1).

**Tabla 2.1 Tabla de contingencia por Grupos etarios y sexo de los pacientes**

Grupos de edad	Sexo					
	Femenino		Masculino		Total	
	Frecuencia	%	Frecuencia	%	Frecuencia	%
<= 40	567	48.10%	227	19.30%	794	67.40%
41 - 60	161	13.70%	110	9.30%	271	23.00%
61+	53	4.50%	60	5.10%	113	9.60%
Total	781	66.30%	397	33.70%	1178	100.00%

**Grafico 1. Sexo y grupos etarios.**



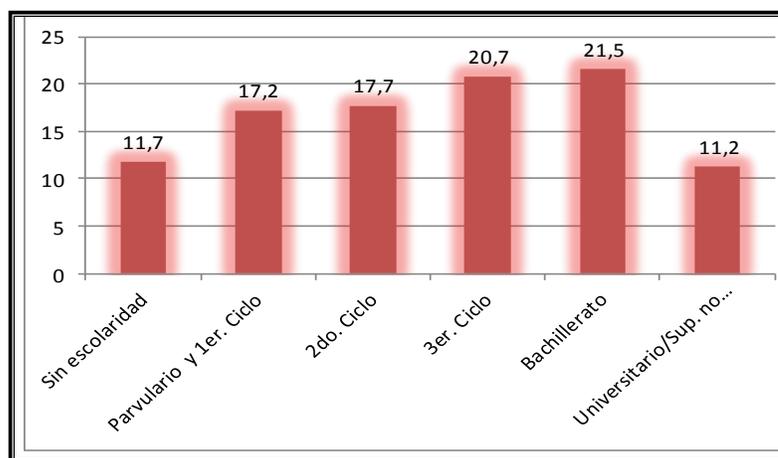
Otro detalle que se considera interesante es el nivel de escolaridad alcanzado por los pacientes; en la Tabla 2.2 se clasifican de acuerdo a la escolaridad y se observán

porcentajes similares, así se tiene que del 11.2% a 21.5% se ubicaron en el nivel de educación básica y bachillerato, con un claro descenso en el porcentaje en relación al nivel de educación superior (11.2%); además, cabe destacar que, el 11.7% de la población no tenía ningún tipo de estudio. Dentro de las unidades de salud que se consideraron, se tiene que la mayoría de los pacientes fueron los que acudieron a las ubicadas dentro de Facultad Multidisciplinaria de occidente y la FOUES, por lo que pudiera decirse que la razón del alto porcentaje de pacientes con estudio a nivel de bachillerato que se refleja en la tabla es alta, ya porque los pacientes que acuden a estos dos centros de salud son generalmente de la comunidad universitaria (ver Tabla 2.2 y Gráfico 2).

**Tabla 2.2. Tabla de distribución de frecuencias de Nivel de estudio obtenido**

Nivel de estudio obtenido	Frecuencia	Porcentaje
Sin escolaridad	138	11,7
Parvulario y 1er. Ciclo	203	17,2
2do. Ciclo	208	17,7
3er. Ciclo	244	20,7
Bachillerato	253	21,5
Universitario y Superior no universitario	132	11,2
<b>Total</b>	<b>1178</b>	<b>100,0</b>

**Gráfico 2. Nivel de estudio obtenido**



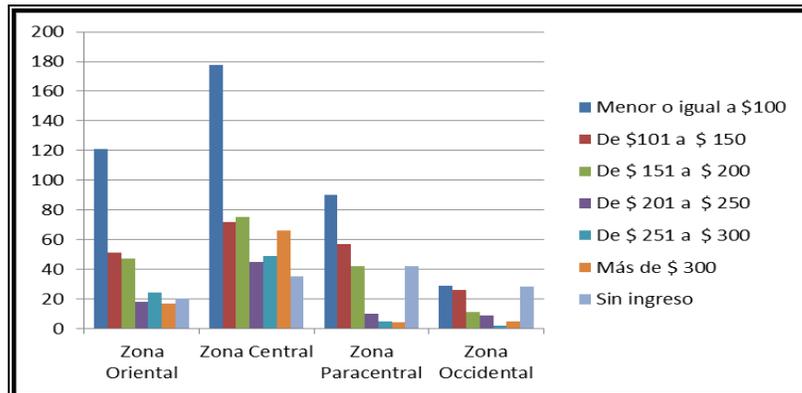
Los pacientes se clasificaron considerando el ingreso económico mensual y su ubicación geográfica de procedencia (Tabla 2.3) estos resultados nos muestran que el 67.9% de

los pacientes tuvieron un ingreso inferior a \$200, es decir, que un poco más del cincuenta por ciento de los pacientes tenían un ingreso inferior al salario mínimo; de los cuales 27.6% provienen de la zona central y el 16% de la zona paracentral que juntos hacen un total del 43.6% lo que curiosamente llama la atención el hecho de que coincide con el censo poblacional, que describe que la población salvadoreña está concentrada en el área metropolitana por lo se podría decir que la alta frecuencia se debe a que existe accesibilidad a las clínicas con servicio odontológico, el tiempo de traslado es menor, así como los costos por el mismo, etc. Cabe destacar también, que el 10.60% de los pacientes no tienen ingreso económico mensual. Además, de los pacientes que acudieron; se atendió el 25.3% de la zona oriental; 21.1% de la zona paracentral y de la zona occidental el 9.3% (ver Tabla 2.3 y Gráfico 3).

**Tabla 2.3. Tabla de contingencia Ingreso económico según Zona geográfica**

Ingreso económico mensual	Zona Geográfica									
	Zona Oriental		Zona Central		Zona Paracentral		Zona Occidental		Total	
	Frec	%	Frec	%	Frec	%	Frec	%	Frec	%
<b>Menor o igual a \$100</b>	121	10.30%	178	15.10%	90	7.60%	29	2.50%	418	35.50%
<b>De \$101 a \$150</b>	51	4.30%	72	6.10%	57	4.80%	26	2.20%	206	17.50%
<b>De \$151 a \$200</b>	47	4.00%	75	6.40%	42	3.60%	11	0.90%	175	14.90%
<b>De \$201 a \$250</b>	18	1.50%	45	3.80%	10	0.80%	9	0.80%	82	7.00%
<b>De \$251 a \$300</b>	24	2.00%	49	4.20%	5	0.40%	2	0.20%	80	6.80%
<b>Más de \$300</b>	17	1.40%	66	5.60%	4	0.30%	5	0.40%	92	7.80%
<b>Sin ingreso</b>	20	1.70%	35	3.00%	42	3.60%	28	2.40%	125	10.60%
<b>Total</b>	298	25.30%	520	44.10%	250	21.20%	110	9.30%	1178	100.00%

**Grafico 3. Zona de residencia e ingreso económico mensual.**



### 2.1.2. ANÁLISIS DE LAS VARIABLES DE LOS ÍNDICES ODONTOLÓGICOS.

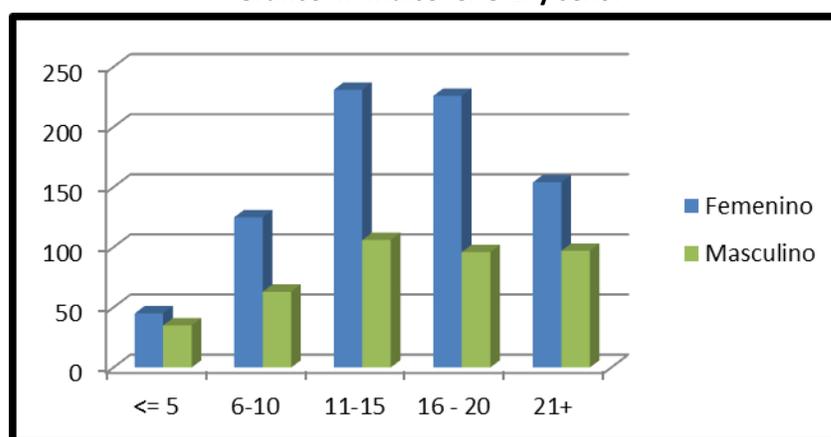
El índice CPO-D tiene un rol fundamental en los estudios odontológicos que se realizan para cuantificar la prevalencia de la Caries Dental, la cual nos señala la experiencia de caries tanto presente como pasada, pues toma en cuenta los dientes con lesiones de caries y con tratamientos previamente realizados.

En la Tabla 2.4 se presenta clasificación del índice de CPO-D agrupado en intervalos según el sexo de los pacientes, se observa que el 28.6% es la mayor población que tenía entre 11 y 15 dientes afectados por caries, de los cuales se distribuyen de forma que el 19.6% son mujeres y el 9% hombres; así también se tiene que el 27.3% de la población examinada presenta de 16 a 20 dientes afectados donde del total el 19.2% son mujeres y el 8.1% son hombres, por último, únicamente el 6.8% de los pacientes tienen 5 o menos dientes afectados según su historial de caries. Podemos señalar por esta información, de manera general, que este índice en la mayoría de la población es alto, esto es más del 90% de la población que ha sido atendida en las unidades de salud de la red pública (ver Tabla 2.4 y Gráfico 4).

**Tabla 2.4 Tabla de contingencia Índice CPO-D y sexo**

CPO-D	SEXO					
	Femenino		Masculino		Total	
	Frecuencia	%	Frecuencia	%	Frecuencia	%
<= 5	45	3.80%	35	3.00%	80	6.80%
6 -10	125	10.60%	63	5.30%	188	16.00%
11 - 15	231	19.60%	106	9.00%	337	28.60%
16 - 20	226	19.20%	96	8.10%	322	27.30%
21+	154	13.10%	97	8.20%	251	21.30%
Total	781	66.30%	397	33.70%	1178	100.00%

**Grafico 4. Índice CPO-D y sexo.**

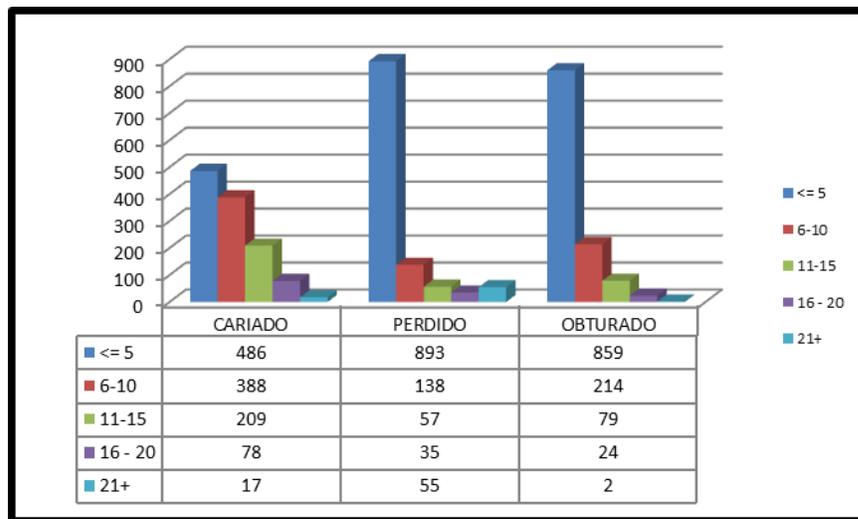


En la Tabla 2.5 se presentan las frecuencias y porcentajes del estado de los dientes en los pacientes y resulta alarmante el hecho que el 41.3% de los pacientes tienen 5 o menos dientes cariados, el 75.8% han perdido 5 o menos dientes y el 72.9% de los pacientes tienen 5 o menos obturaciones. Por otro lado, cabe mencionar también que en la categoría de más de 21 dientes afectados, únicamente 1.4% de los pacientes tienen dientes cariados, 4.7% perdidos y 0.2% obturados. También, para los pacientes que tienen más de 21 dientes afectados, 55 pacientes de ellos están edéntulos (ya no tiene dentadura propia/columna de perdido), 17 personas padecen de caries y solamente 2 personas tenían más de 21 dientes obturados (ver Tabla 2.5 y Gráfico 5).

**Tabla 2.5 Tabla de distribución de Frecuencias de dientes cariados, perdidos y obturados.**

Número de Dientes	CARIADO		PERDIDO		OBTURADO	
	Frecuencia	%	Frecuencia	%	Frecuencia	%
<= 5	486	41,3	893	75,8	859	72.9
6-10	388	32,9	138	11,7	214	18.2
11-15	209	17,7	57	4,8	79	6.7
16 – 20	78	6,6	35	3	24	2
21+	17	1,4	55	4,7	2	0.2
<b>Total</b>	<b>1178</b>	<b>100</b>	<b>1178</b>	<b>100</b>	<b>1178</b>	<b>100</b>

**Gráfico 5. Dientes cariados, perdidos y obturados.**



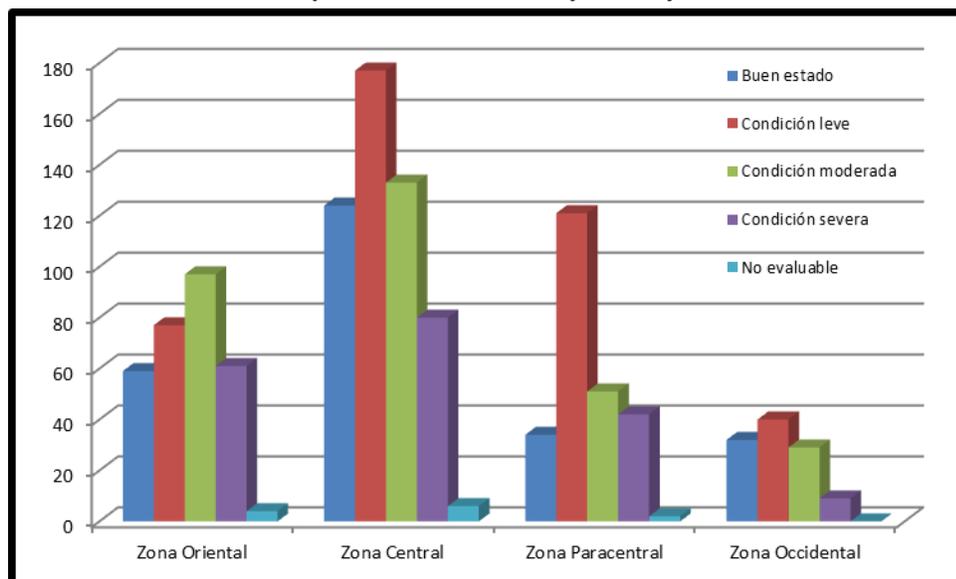
En la Tabla 2.6 se presenta el índice de placa bacteriana, la placa bacteriana no es más que el producto de no tener hábitos de higiene adecuados o simplemente que si el paciente los tiene, no se hace una correcta técnica de cepillado, por lo que los resultados son impresionantes, dado que en total 77.8% de la población sufren de placa bacteriana desde una condición leve hasta severa (ver Tabla 2.6 y Gráfico 6). Así también se muestra a cada individuo según la zona geográfica en la que el paciente reside, se observa que dentro de la clasificación se tiene que el 35.2% de la población se encontró en una condición leve, siendo el 15% residentes de la zona central y de la zona paracentral 10.3%; también, se muestra que el 26.3% de los pacientes se ubicaron

en una condición moderada, de los cuales el 11.3% provenían de la zona central y solamente el 2.5% eran de la zona oriental. Se encontró también que en el 21.1% de los pacientes en buen estado de placa bacteriana, el 10.5% fueron de la zona central y un 16.30% que se encontraban en condición severa, el 6.80% residían en la zona central.

**Tabla 2.6 Tabla de contingencia Índice de Placa Bacteriana según Zona geográfica**

Índice de Placa bacteriana Løe y Silness	Zona Geográfica									
	Z. Oriental		Z. Central		Z. Paracentral		Z. Occidental		Total	
	Frecuencia	%	Frecuencia	%	Frecuencia	%	Frecuencia	%	Frecuencia	%
Buen estado	59	5.00%	124	10.50%	34	2.90%	32	2.70%	249	21.10%
Condición leve	77	6.50%	177	15.00%	121	10.30%	40	3.40%	415	35.20%
Condición moderada	97	8.20%	133	11.30%	51	4.30%	29	2.50%	310	26.30%
Condición severa	61	5.20%	80	6.80%	42	3.60%	9	0.80%	192	16.30%
No evaluable	4	0.30%	6	0.50%	2	0.20%	0	0.00%	12	1.00%
<b>Total</b>	<b>298</b>	<b>25.30%</b>	<b>520</b>	<b>44.10%</b>	<b>250</b>	<b>21.20%</b>	<b>110</b>	<b>9.30%</b>	<b>1178</b>	<b>100.00%</b>

**Gráfico 6. Índice de placa bacteriana Loe y Silnes y zona de residencia**

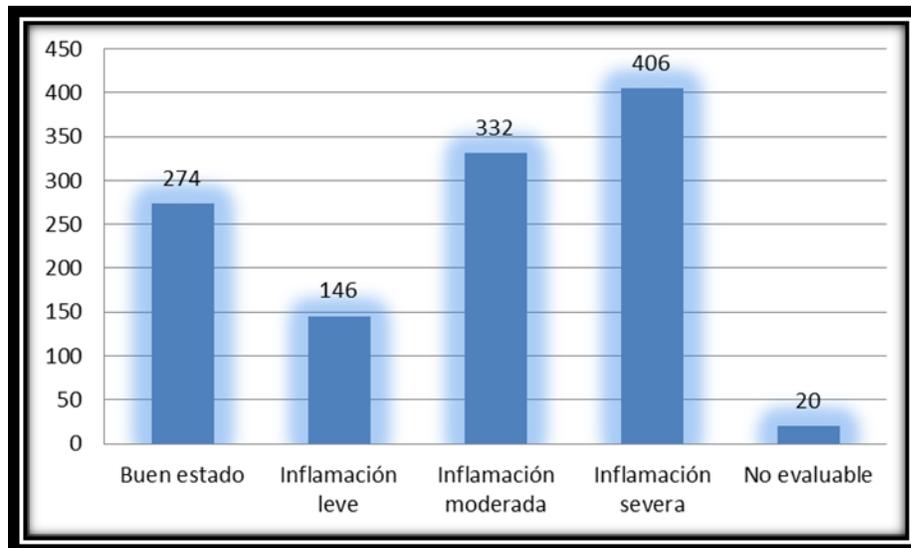


Otro de los problemas bucales que se contempló en los pacientes fue la gingivitis la cual es producida por altos índices de placa bacteriana, es un irritante local que puede ser removido al cepillarse los dientes pero cuando la gingivitis ya está establecida, en la mayoría de los casos para que pueda sanarse debe de ser tratada por el especialista y en varias consultas según el nivel que sufre la enfermedad el paciente. De acuerdo al tratamiento de la población con este problema, a continuación se presenta la clasificación de la gingivitis en la cual, la lectura de los resultados muestra que el 34.5% de los pacientes presentaron inflamación severa y el 28.2% inflamación moderada; el 23.3% de los pacientes no presentaban esta enfermedad infecciosa e inflamatoria. (Ver Tabla 2.7 y Gráfico 7).

**Tabla 2.7 Tabla de distribución de frecuencias de Índice gingival Løe y Silness**

Niveles de la enfermedad	Frecuencia	Porcentaje
Buen estado	274	23.3
Inflamación leve	146	12.4
Inflamación moderada	332	28.2
Inflamación severa	406	34.5
No evaluable	20	1.7
<b>Total</b>	<b>1178</b>	<b>100</b>

**Gráfico 7. Índice gingival Loe y Silness**

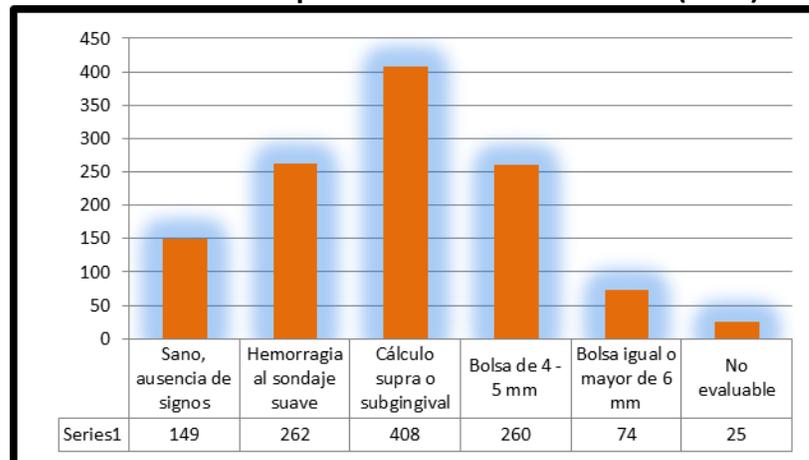


En cuanto a la gravedad del problema bucal, en la Tabla 2.8 está reflejada la distribución de frecuencias de los niveles de la enfermedad que se calcula considerando el índice periodontal, donde el 34.6% de los pacientes presentaron cálculo supra o subgingival, 22.2% hemorragia al sondaje suave, 22.1% Bolsa de 4 -5 mm y los no evaluables por causa de faltantes de dientes al seccionar en sextantes (seis partes) la cavidad bucal fueron un total de 2.1%. Cabe destacar que solo el 12.6% de los pacientes estaban en un nivel de ausencia de esta enfermedad, es decir, sanos en cuanto a condición periodontal y resulta preocupante el hecho de que aproximadamente el 85% de los pacientes están sufriendo dicha enfermedad, y por consideración del especialista es necesario un tratamiento inmediato (ver Tabla 2.8 y Gráfico 8).

**Tabla 2.8 Tabla de distribución de frecuencias de Índice Periodontal Comunitario CPI.**

Niveles de la enfermedad	Frecuencia	Porcentaje
Sano, ausencia de signos	149	12.6
Hemorragia al sondaje suave	262	22.2
Cálculo supra o subgingival	408	34.6
Bolsa de 4 -5 mm	260	22.1
Bolsa igual o mayor de 6 mm	74	6.3
No evaluable	25	2.1
<b>Total</b>	<b>1178</b>	<b>100</b>

**Gráfico 8. Índice periodontal comunitario CPI (OMS)**

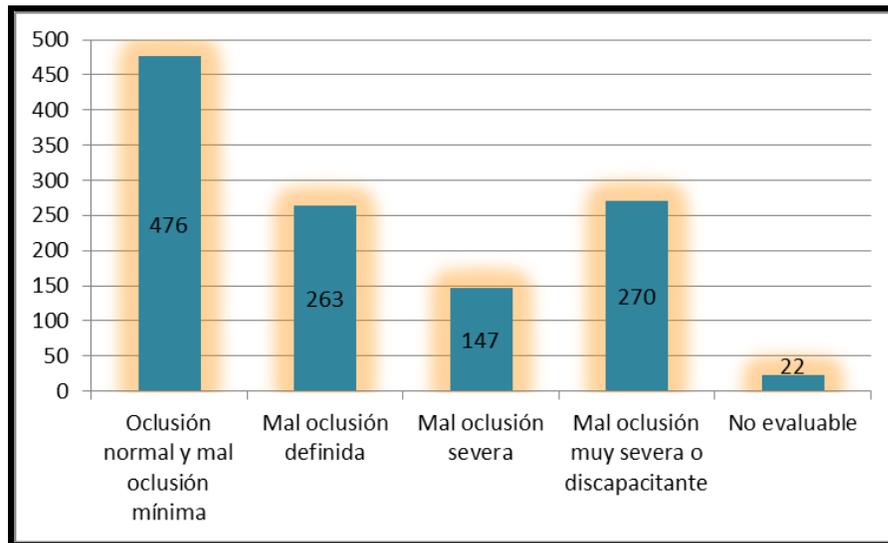


Índice de estética dental (DAI) establece la estética y la posición adecuada que deben de tener los dientes en los arcos maxilar y mandibular de la boca. La Tabla 2.9 nos muestra los niveles de afectación de este problema y su lectura nos pone de manifiesto que el 40.4% de los pacientes se encontraban bajo la condición de Oclusión normal y mal oclusión mínima, o sea, este paciente no requiere tratamiento de ortodoncia porque la posición de sus dientes no está afectada y así la estética podría considerarse como adecuada; seguidamente, el 22.3% estaban en condición de Mal oclusión definida y el 1.9% de los pacientes fueron no evaluables, éstos son los pacientes edéntulos, aspectos que reflejan la preocupación de las entidades de salud en cuanto a este fenómeno es que el 22.9% de la población estuvo evaluada con mal oclusión severa o incapacitante éstos pacientes en dicho nivel de enfermedad; sugieren los odontólogos, que requerían un tratamiento de carácter urgente debido a la condición de gravedad en el paciente y debe de plantearse a la comunidad la necesidad y gran importancia de la prevención para evitar la pérdida de dientes anteriores (dientes de adelante, ver Coeficiente de regresión Tabla 1.2) se pone en evidencia que este requiere de rehabilitación protésica urgente, si no se trata, el paciente puede experimentar los siguientes aspectos: estética inaceptable, reducción de la función masticatoria, situación traumática que predispone destrucción de tejido, afectación en la pronunciación, falta de estabilidad en la oclusión (ver Tabla 2.9 y Gráfico 9).

**Tabla 2.9 Tabla de distribución de frecuencias Mal oclusión Índice DAI (de OMS)**

Niveles de la enfermedad	Frecuencia	Porcentaje
Oclusión normal y mal oclusión mínima	476	40.4
Mal oclusión definida	263	22.3
Mal oclusión severa	147	12.5
Mal oclusión muy severa o discapacitante	270	22.9
No evaluable	22	1.9
<b>Total</b>	<b>1178</b>	<b>100</b>

**Gráfico 9. Índice de Mal oclusión DAI.**

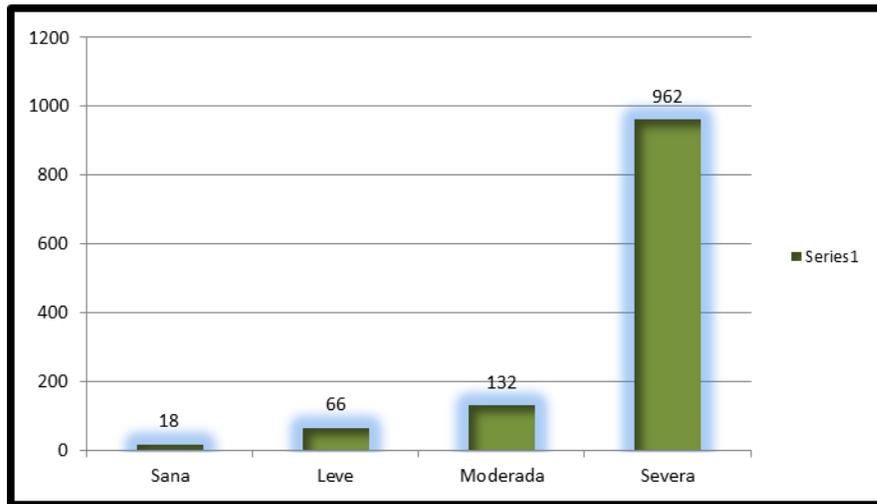


A continuación en la Tabla 2.10 se muestra la distribución de frecuencias para la variable respuesta Condición de salud bucal. Los pacientes mostraron diferentes niveles de enfermedad bucal, la cual es la conjunción de los índices afectados en la población; en ella se observa que el 81.7% revelan tener condición en el nivel de salud bucal severa, dicho porcentaje de pacientes es considerablemente mayor; mientras que solamente el 1.5% tienen condición ubicada en el nivel sana, el 5.6% y 11.2% se encuentran en condición de salud bucal leve y moderada respectivamente (ver Tabla 2.10 y Gráfico 10).

**Tabla 2.10 Tabla de distribución de frecuencias Condición de Salud Bucal**

Niveles de la enfermedad	Frecuencia	Porcentaje
Sana	18	1,5
Leve	66	5,6
Moderada	132	11,2
Severa	962	81,7
<b>Total</b>	<b>1178</b>	<b>100,0</b>

**Gráfico 10. Condición de salud bucal.**



## **2.2. ANÁLISIS BIDIMENCIONAL.**

Para efectuar un buen nivel de revisión y profundización en el estudio condición de salud bucal es necesario cuantificar la relación que presentan dos características importantes de la investigación, por medio de test paramétricos; con ello se establecen hipótesis para realizar dicho proceso.

A continuación, se presenta el análisis para determinar si existe relación de algunas variables explicativas ( $X_i$ ), con la variable respuesta Condición de salud bucal ( $Y$ ), para el estudio se han considerado las variables ( $X_i$ ,  $Y$ ) construyendo la respectiva tabla de contingencia y luego la aplicación del test de prueba el coeficiente Chi – Cuadrado de Pearson. El objetivo principal de este análisis es estudiar la Hipótesis nula  $H_0$ : ( $X_i$ ,  $Y$ ) si su estudio muestra una relación de independencia versus la Hipótesis alternativa  $H_1$ : ( $X_i$ ,  $Y$ ) las variables no son independientes (es decir que están asociadas). Veamos a continuación:

Con el fin de establecer si el estado de la condición de salud bucal tiene algún tipo de relación con los índices placa bacteriana, gingivitis, periodontitis, estética (DAI), dientes

perdidos, se presenta la Tabla 2.11, teniendo en cuenta que para la elaboración de esta tabla no se ha considerado la categoría “No evaluable” de las variables independientes es decir Índice de Placa bacteriana Løe y Silness, gingivitis, periodontitis, estética (DAI), debido a que cierta cantidad de casillas presentaban un porcentaje de los datos que tenían una frecuencia esperada inferior a 5 (más del 20% de frecuencias inferiores a 5) lo que plantea estadísticamente, es la falta de significancia para la prueba, por lo que sólo se contemplan las categorías restantes.

Mediante la prueba de Chi-cuadrado de Pearson se elaboraron hipótesis en las que, se determinará si existe algún tipo de influencia entre ambas variables, en consecuencia se plantean las siguientes hipótesis:

1)H<sub>0</sub>: “El índice de placa bacteriana no influye en la condición de salud bucal” vs.  
H<sub>1</sub>: “El índice de placa bacteriana influye en la condición de salud bucal”.

2)H<sub>0</sub>: “El índice de gingival no influye en la condición de salud bucal”. vs.  
H<sub>1</sub>: “El índice gingival influye en la condición de salud bucal”.

3)H<sub>0</sub>: “El índice periodontal comunitario no influye en la condición de salud bucal”  
vs. H<sub>1</sub>: “El índice periodontal comunitario influye en la condición de salud bucal”.

4)H<sub>0</sub>: “La mal oclusión no influye en la condición de salud bucal” vs.  
H<sub>1</sub>: “La mal oclusión influye en la condición de salud bucal”

Además, en la misma tabla fácilmente se puede verificar que a medida los niveles de cada una de las enfermedades presenta un nivel grave, también también se presenta el nivel de desmejora de la condición de salud bucal, observando luego de efectuar los contrastes que en la columna de p-valor todos los niveles de significancia son inferiores a 0.05, con lo que se garantiza que la probabilidad de equivocarnos al aceptar la hipótesis altermaniva es de 0.00 que el planteado, 0.05. En conclusión, están

íntimamente relacionadas las enfermedades de la cavidad oral con la condición de salud bucal de los consultados, en otras palabras, a mayor índice de placa bacteriana detectada en la boca sugiere condición de salud bucal en estado severo, por consiguiente las enfermedades de la que se hacen referencia son: inflamación severa en las encías, índices periodontales con bolsas de 4-5 mm, mal oclusión discapacitante o severa que determinará la condición de salud bucal severa.

**Tabla 2.11 Relación entre enfermedades bucales y la condición de salud bucal**

Enfermedad	Niveles de la enfermedad	Condición de Salud Bucal										p-valor
		Condición Sana		Condición Leve		Condición Moderada		Condición Severa		Total		
		Frec.	%	Frec.	%	Frec.	%	Frec.	%	Frec.	%	
Índice de Placa bacteriana Løe y Silness	Buen estado	14	1.20%	28	2.40%	47	4.10%	156	13.50%	245	21.20%	0.00*
	Condición leve	4	0.30%	27	2.30%	48	4.10%	335	28.90%	414	35.80%	
	Condición moderada	0	0.00%	7	0.60%	20	1.70%	281	24.30%	308	26.60%	
	Condición severa	0	0.00%	4	0.30%	17	1.50%	170	14.70%	191	16.50%	
	Total	18	1.60%	66	5.70%	132	11.40%	942	81.30%	1158	100.00%	
Índice gingival Løe y Silness	Buen estado	17	1.50%	33	2.80%	56	4.80%	168	14.50%	274	23.70%	0.00*
	Inflamación leve	0	0.00%	9	0.80%	22	1.90%	115	9.90%	146	12.60%	
	Inflamación moderada	0	0.00%	6	0.50%	21	1.80%	305	26.30%	332	28.70%	
	Inflamación severa	1	0.10%	18	1.60%	33	2.80%	354	30.60%	406	35.10%	
	Total	18	1.60%	66	5.70%	132	11.40%	942	81.30%	1158	100.00%	
Índice Periodontal Comunitario CPI (OMS)	Sano, ausencia de signos	16	1.40%	24	2.10%	27	2.30%	82	7.10%	149	12.90%	0.00
	Hemorragia al sondaje suave	2	0.20%	25	2.20%	38	3.30%	197	17.10%	262	22.70%	
	Cálculo supra o subgingival	0	0.00%	11	1.00%	44	3.80%	353	30.60%	408	35.40%	
	Bolsa de 4 -5 mm y Bolsa mayor o igual a 6 mm	0	0.00%	6	0.50%	23	2.00%	305	26.50%	334	29.00%	
	Total	18	1.60%	66	5.70%	132	11.40%	937	81.30%	1153	100.00%	

<b>Maloclusión, Índice DAI (de OMS)</b>	<b>Oclusión normal y mal oclusión mínima</b>	17	1.50%	47	4.10%	85	7.40%	327	28.30%	476	41.20%	0.00*
	<b>Mal oclusión definida</b>	1	0.10%	16	1.40%	23	2.00%	223	19.30%	263	22.80%	
	<b>Mal oclusión severa</b>	0	0.00%	1	0.10%	14	1.20%	132	11.40%	147	12.70%	
	<b>Mal oclusión muy severa o discapacitante</b>	0	0.00%	2	0.20%	10	0.90%	258	22.30%	270	23.40%	
	<b>Total</b>	18	1.60%	66	5.70%	132	11.40%	940	81.30%	1156	100.00%	

p-valor: Test de independencia de Chi-cuadrado. Nivel de significación 0.05.

### 2.3. COMPARACIÓN DE VARIAS PROPORCIONES

En esta sección se pretende estudiar las variables con el fin de proporcionar la solución de los objetivos específicos planteados a la luz de esta investigación. Para este abordaje fue necesario usar el software estadístico SPSS, en el menú Analizar>Tablas personalizadas> Estadísticos de contraste, seguidamente, se selecciona la opción: “Comparar las proporciones de columna (pruebas z)” con el objeto de establecer la relación entre las variables en estudio [26].

A través de esta prueba se planteó determinar si existía diferencia estadísticamente significativa entre los porcentajes obtenidos entre los niveles de la condición bucal relacionando con cada una de las categorías de la variable ubicada en las filas; por ejemplo en Tabla 2.12, en las que se observa la comparación de los porcentajes obtenidos para cada una de las categorías de la variable de columna (Condición de salud bucal) con condición sana, leve, moderada y severa con las categorías de la variable de fila (Sexo).

Las pruebas estadísticas que se incluyen en el procedimiento son Tablas personalizadas, que permiten determinar si existe una relación entre las variables de fila (Sexo, para este caso) y columna (Condición de salud bucal); para esta ocasión utilizamos el test de Chi-cuadrado y luego plantearse la Comparación de proporciones debido a que las variables son categóricas.

Las hipótesis que se plantearon para cada una de las comparaciones de proporciones son de la siguiente forma:

$$H_0 : p_{i1} = p_{i2} = p_{i3} = p_{i4} \quad \forall i=1,2,\dots,r$$
$$H_1 : p_{ij} \neq p_{ik} \quad \text{para al menos } i \neq k$$

Dónde:  $p_{ij}$ , representa la proporción de población de la variable Condición de salud bucal (columna), en las categorías: condición sana, condición leve, condición moderada y condición severa ( $j=1,2,3,4$ ); según la población de las categorías masculino y

femenino de la variable Sexo que se ubicaran en la fila  $i$  con  $r$  categorías  $\forall i = 1, 2, \dots, r$  ; es decir, para el caso,  $r = 2$  [24].

Los resultados se reflejan en las siguientes tablas para cada uno los objetivos, tal como a continuación se presentan.

**Objetivo específico 1:**

***“Determinar si el sexo de los pacientes atendidos influye sobre la condición de salud bucal”***

Hipótesis a validar:

$H_0$  : La proporción de los pacientes atendidos no son diferentes en su condición de salud bucal según su sexo.

$H_1$  : La proporción de los pacientes atendidos son diferentes en su condición de salud bucal según su sexo.

La lectura de los resultados obtenidos, según la Tabla 2.12 podemos decir que de la población de pacientes atendidos con mayor porcentaje de asistencia al área odontológica son del sexo femenino, se puede también observarse que el 66.3% de pacientes del sexo femenino atendidas se encuentran bajo una condición de salud bucal severa y que el 33.7% de de pacientes masculinos también se encuentran bajo esta misma condición, contrariamente en la Tabla 2.13 el estadístico de prueba de Chi-cuadrado de Pearson nos indica que no existen suficientes criterios para sostener que el sexo y la condición de salud bucal están plenamente relacionadas (existencia de independencia) según el nivel de significación establecido, dado que su  $p > 0.05$ . Por lo que se concluye bajo el nivel de confianza del 95% que la comparación de proporciones de las categorías de la variable “Condición de salud bucal” ubicada en la columna para esta relación, no son significativamente diferentes en sentido estadístico, por tanto, se acepta la hipótesis nula en la que se plantea que no existe diferencias significativas entre las proporciones.

**Tabla 2.12 Tabla de contingencia condición de salud bucal según sexo.**

Sexo	Condición de Salud Bucal								TOTAL
	Condición Sana		Condición Leve		Condición Moderada		Condición Severa		
	Frecuencia	%	Frecuencia	%	Frecuencia	%	Frecuencia	%	
<b>Femenino</b>	12	66.7	45	68.2	86	65.2	638	66.3	781
<b>Masculino</b>	6	33.3	21	31.8	46	34.8	324	33.7	397
<b>Total</b>	18	100	66	100	132	100	962	100	1178

**Tabla 2.13 Tabla de Pruebas de chi-cuadrado de Pearson de la variable sexo relacionada con Condición de salud bucal**

SEXO		Condición de Salud Bucal
	Chi cuadrado	,184
	gl	3
	Sig.	,980

**Objetivo específico 2:**

***“Determinar si el ingreso económico mensual de los pacientes es un factor determinante sobre la Condición de salud bucal”***

Hipótesis a validar:

$H_0$ : La proporción de los pacientes atendidos según la condición de salud bucal no son diferentes en cuanto a su ingreso económico mensual.

$H_1$ : La proporción de pacientes atendidos según la condición de salud bucal es diferente en cuanto a su ingreso económico mensual.

Una vez dado tratamiento a los datos los resultados que se obtuvieron, según la Tabla 2.14 diremos que donde se presenta el ingreso económico mensual familiar según la condición de salud bucal, se establece que la población con ingresos menores o iguales a \$250.00 es del 77% y con ingresos mayores a \$250 resultar ser un 23% de los pacientes; además según la Tabla 2.15, se realiza el test para establecer el criterio de asociación entre el ingreso económico mensual del paciente y la condición de salud bucal se observa que con un nivel de confianza del 95% se acepta la existencia de

diferencias entre proporciones estadísticamente significativas según el nivel de significación establecido para su comparación;  $0.001 < 0.05$  (Tabla 2.15).

Para verificar los resultados anteriores se realiza la tabla de comparación (Tabla 2.16) en la que se observa que se confirma lo mencionado anteriormente; en las filas se tiene las categorías de la variable Ingreso económico mensual. Se muestra que *el ingreso mensual mayor a \$250* casilla que se relaciona con la columna *condición sana (A)*, aparece la letra D esta letra indica que el porcentaje de columna para la categoría condición sana es estadísticamente mayor que el porcentaje de la categoría condición severa y así mismo se interpreta el porcentaje para la categoría condición de salud bucal moderada es estadísticamente mayor que la condición severa.

Además, fácilmente se pueden comparar los porcentajes mostrados en la Tabla 2.16 en la que para la categoría *Condición Severa* relacionada con los ingresos mensuales menores o iguales a \$250, el valor corresponde a 77%, mientras que el porcentaje para la categoría condición sana y moderada es de 50% y 64% respectivamente la relación existente de los pacientes que tienen ingresos mensuales mayor a \$250 bajo una condición sana el porcentaje es 50% y los que presentan una condición severa es del 23%, ahora, bajo la condición moderada el porcentaje es 36% vrs. 23% de la severa presentan un ingreso mayor a \$250. Desde luego estas aseveraciones no se pueden realizar hasta generar la prueba de comparaciones reflejadas en la misma Tabla 2.16.

Debido a los resultados comentados y anteriormente estudiados, podemos concluir que la condición de salud bucal que gozan los pacientes con ingresos superiores al ingreso mínimo es mejor que en pacientes con ingresos inferiores. Cabe mencionar que cuando alguno de los porcentajes es cero, las pruebas de comparación no pueden realizarse.

**Tabla 2.14 Tabla de contingencia del ingreso económico mensual según Condición de Salud Bucal**

Ingreso Mensual	Condición de Salud Bucal								Total
	Condición Sana		Condición Leve		Condición Moderada		Condición Severa		
	Frec.	%	Frec.	%	Frec.	%	Frec.	%	
Menor o igual \$250	9	50%	47	71%	84	64%	741	77%	881
Mayor a \$250	9	50%	19	29%	48	36%	221	23%	297
<b>Total</b>	<b>18</b>	<b>100%</b>	<b>66</b>	<b>100%</b>	<b>132</b>	<b>100%</b>	<b>962</b>	<b>100%</b>	<b>1178</b>

**Tabla 2.15 Pruebas de chi-cuadrado de Pearson de ingreso económico mensual relacionado con la Condición de Salud Bucal**

Ingreso Mensual	Condición de Salud Bucal	
	Chi cuadrado	17,577
	gl	3
	Sig.	,001*
Los resultados se basan en filas y columnas no vacías de cada subtabla más al interior.		
*. El estadístico de chi-cuadrado es significativo en el nivel 0.05.		

**Tabla 2.16 Comparaciones de proporciones de columnas<sup>a</sup> del ingreso económico mensual relacionado con la Condición de Salud Bucal**

Ingreso Mensual	Condición de Salud Bucal			
	Condición Sana	Condición Leve	Condición Moderada	Condición Severa
	(A)	(B)	(C)	(D)
Menor o igual \$250				A C
Mayor a \$250	D		D	
Los resultados se basan en pruebas bilaterales con un nivel de significación 0.05. Para cada par significativo, la clave de la categoría con la proporción de columna menor aparece debajo de la categoría con mayor proporción de columna.				
a. Utilizando la corrección de Bonferroni, se han ajustado las pruebas para todas las comparaciones por pares dentro de una fila para cada subtabla situada más al interior.				

**Objetivo específico 3:**

***“Determinar si el nivel de escolaridad alcanzado del pacientes atendidos influye sobre la condición de salud bucal”***

Hipótesis a validar:

$H_0$  : La proporción de los pacientes atendidos según su condición de salud bucal no son diferentes en cuanto al nivel de escolaridad alcanzado.

$H_1$ : La proporción de los pacientes atendidos según la condición de salud bucal es diferente en cuanto al nivel de escolaridad alcanzado.

Se investigó el nivel de escolaridad de los pacientes atendidos, para ello se elaboró la Tabla 2.17, en la que se presenta la información y según el porcentaje de la relación entre las variables Nivel de escolaridad y la Condición de salud bucal y en la Tabla 2.18 se presenta el test de chi-cuadrado de Pearson, en el que se pretende explicar que en el nivel de escolaridad está asociado en algún grado con la condición de salud bucal del paciente ( $p < 0.05$ ). Al observar la Tabla 2.19 en la fila *Sin escolaridad*, la casilla que se relaciona con la columna *Condición severa* (Condición de salud bucal), en las que aparecen las letra B y C; éstas indican que el porcentaje de columna para la categoría Condición severa es estadísticamente mayor que los porcentajes de las categorías Condición leve y Condición moderada. Notese que en la categoría *Sin escolaridad* de la variable de filas; es decir, que en los pacientes sin escolaridad las condiciones de salud bucal Leve, moderada y severa se describen diferencias porcentuales estadísticamente significativas. Ahora, al observar la fila el nivel de educación; educación básica se distingue que, en la casilla que se relaciona con la columna Condición severa, aparece la letra C lo que indica que el porcentaje de columna para la categoría condición severa es estadísticamente mayor que los porcentajes de las categorías de condición moderada. Para los pacientes que tiene un nivel de escolaridad básico las condiciones de salud bucal moderada y severa cuentan con diferencias porcentuales estadísticamente significativas, observándose que el porcentaje de la relación Educación básica y Condición severa es 58% mientras que para la condición moderada es de 45%.

**Tabla 2.17 Tabla de contingencia de Nivel de escolaridad según Condición de salud bucal**

Nivel de escolaridad	Condición de Salud Bucal								Total
	Condición Sana		Condición Leve		Condición Moderada		Condición Severa		
	Frec.	%	Frec.	%	Frec.	%	Frec.	%	
Sin escolaridad	0	0%	2	3%	6	5%	130	14%	138
Educación básica	6	33%	33	50%	59	45%	557	58%	655
Bachillerato	6	33%	18	27%	33	25%	196	20%	253
Educación superior	6	33%	13	20%	34	26%	79	8%	132
<b>Total</b>	<b>18</b>	<b>100%</b>	<b>66</b>	<b>100%</b>	<b>132</b>	<b>100%</b>	<b>962</b>	<b>100%</b>	<b>1178</b>

**Tabla 2.18 Pruebas de chi-cuadrado de Pearson de Nivel de escolaridad relacionado con la Condición de salud bucal**

Nivel de escolaridad	Condición de Salud Bucal	
	Chi cuadrado	68,790
gl	9	
Sig.	,000*	
Los resultados se basan en filas y columnas no vacías de cada subtabla más al interior.		
*. El estadístico de chi-cuadrado es significativo en el nivel 0.05.		

**Tabla 2.19 Comparaciones de proporciones de columnas<sup>b</sup> de Nivel de escolaridad relacionado con la Condición de salud bucal**

Nivel de escolaridad	Condición de Salud Bucal			
	Condición Sana	Condición Leve	Condición Moderada	Condición Severa
	(A)	(B)	(C)	(D)
Sin escolaridad	<sup>a</sup>			B C
Educación básica				C
Bachillerato				
Educación superior	D	D	D	
Los resultados se basan en pruebas bilaterales con un nivel de significación 0.05. Para cada par significativo, la clave de la categoría con la proporción de columna menor aparece debajo de la categoría con mayor proporción de columna.				
a. Esta categoría no se utiliza en las comparaciones porque su proporción de columna es igual a cero o uno.				
b. Utilizando la corrección de Bonferroni, se han ajustado las pruebas para todas las comparaciones por pares dentro de una fila para cada subtabla situada más al interior.				

**Objetivo específico 4:**

***“Determinar si la zona de el lugar de residencia de los pacientes atendidos influye sobre la condición de su salud bucal”***

Hipótesis a validar

$H_0$ : La proporción de los apacientes atendidos según su condición de salud bucal no son diferentes en cuanto a su zona de residencia.

$H_1$ : La proporción de los pacientes atendidos según su condición de salud bucal es diferente en cuanto a su zona de residencia.

Una vez seleccionadas las variables se elaboró la Tabla 2.20 en la que se relaciona la información de la zona de residencia, se presentan frecuencias, porcentajes de la relación de las variables y de la región geográfica de donde proceden con su respectiva condición de salud bucal, además en la Tabla 2.21 se presenta el test chi-cuadrado de Pearson con la que se establece la existencia del grado de asociación entre ambas condiciones ( $p < 0.05$ ).

Con el objetivo de establecer si existe diferencia estadísticamente significativa de la condición de salud bucal según la región geográfica de residencia del paciente mediante la Tabla 2.22 para determinar y efectuar la comparación de proporciones tenemos que:

- Al observar la relación entre las categorías de éstas variables, zona de residencia central de los atendidos (variable fila) y la condición de salud bucal moderada (variable columna), muestra que la proporción es estadísticamente mayor en los pacientes que presentan una condición moderada con un 55% (C) y de un 42% (D) de pacientes con una condición de salud bucal severa. En conclusión diremos que los pacientes que necesitan de una atención médica directa se encuentran, en primer lugar, los pacientes con una condición de salud bucal moderada y posteriormente los que presentan condición de salud bucal severa, además, todos ellos pertenecen a la zona central del país.

- También en la relación de las categorías zona occidental y condición de salud sana (A) la proporción de pacientes es estadísticamente mayor que la de pacientes que están en condición modera (C) y en condición severa, 10%. O sea que los pacientes que residen en la zona occidental están bajo condición de salud bucal sana y severa, de igual forma es necesaria la atención médica.

**Tabla 2.20. Tabla de contingencia de Zona geográfica de residencia según Condición de salud bucal.**

Zona Geográfica	Condición de Salud Bucal								Total
	Condición Sana		Condición Leve		Condición Moderada		Condición Severa		
	Frec	%	Frec	%	Frec	%	Frec	%	
<b>Zona Oriental</b>	3	17%	15	23%	31	23%	249	26%	298
<b>Zona Central</b>	11	61%	33	50%	73	55%	404	42%	521
<b>Zona Paracentral</b>	1	6%	15	23%	25	19%	208	22%	249
<b>Zona Occidental</b>	3	17%	3	5%	3	2%	101	10%	110
<b>Total</b>	18	100%	66	100%	132	100%	962	100%	1178

**Tabla 2.21 Pruebas de chi-cuadrado de Pearson de Zona geográfica de residencia relacionado con la Condición de salud bucal**

		Condición de Salud Bucal
<b>Zona Geográfica</b>	<b>Chi cuadrado</b>	21.065
	<b>gl</b>	9
	<b>Sig.</b>	.012*
Los resultados se basan en filas y columnas no vacías de cada subtabla más al interior.		
*. El estadístico de chi-cuadrado es significativo en el nivel 0.05.		

**Tabla 2.22 Comparaciones de proporciones de columnas<sup>a</sup> de Zona geográfica de residencia relacionado con la Condición de salud bucal**

Zona Geográfica	Condición de Salud Bucal			
	Condición Sana	Condición Leve	Condición Moderada	Condición Severa
	(A)	(B)	(C)	(D)
Zona Oriental				
Zona Central			D	
Zona Paracentral				
Zona Occidental	C			C
Los resultados se basan en pruebas bilaterales con un nivel de significación 0.05. Para cada par significativo, la clave de la categoría con la proporción de columna menor aparece debajo de la categoría con mayor proporción de columna.				
a. Utilizando la corrección de Bonferroni, se han ajustado las pruebas para todas las comparaciones por pares dentro de una fila para cada subtabla situada más al interior.				

Otro aspecto que cabe destacar es sobre el ingreso económico relacionado con la condición de salud bucal en cuanto al deterioro de los dientes (cariado, perdido y obturado) de los pacientes. Para el abordaje seguimos la misma metodología de comparación de proporciones, para ello se ha tenido a bien tomar solamente los pacientes que tienen un ingreso económico mensual haciendo un total de 1,053 pacientes; los detalles a continuación:

**Objetivo específico 5:**

***“Determinar si el ingreso económico del pacientes influye en la condición cantidad de dientes obturados”***

Hipótesis a validar:

$H_0$  : La proporción de los pacientes atendidos según el ingreso económico mensual no son diferentes en cuanto a la condición de la cantidad de dientes obturados.

$H_1$  : La proporción de los pacientes atendidos según el ingreso económico mensual es diferente en cuanto a la condición de la cantidad de dientes obturados.

Los resultados que se obtuvieron de relacionar ambas variables se tienen en la Tabla 2.23, se presentan los datos cantidad de dientes obturados agrupados en intervalos y el ingreso económico de los pacientes, cabe destacar que el 73% de los pacientes que se han realizado alguna obturación tienen menos o igual a 5 obturaciones y que únicamente el 9% presenta más de 11 obturaciones. La Tabla 2.24 muestra los resultados sobre la prueba de chi-cuadrado de Pearson en la que presenta el indicador que se acepta que existe relación de dependencia entre ambas variables.

De la Tabla 2.25 al revisar la relación entre las variables en estudio podemos obtener las siguientes conclusiones:

- En la categoría de los ingresos menores o igual a \$100 (fila) y la categoría menos o igual a 5 dientes obturados (columna), indica que la proporción de pacientes que tienen 5 obturaciones o menos es mayor a la proporción de pacientes que tienen de 6-10 obturaciones y los que tienen más de 11 obturaciones cuando tienen ingresos de \$100 o menos.
- Algo similar ocurre en la categoría de ingresos de \$101 a \$150 en los que se observa que la proporción que es estadísticamente menor la de los pacientes que tienen más de 11 obturaciones.
- Aunque resulta de mucho interés considerar lo que ocurre cuando los ingresos son mayores a \$300 ya que los pacientes muestran que la cantidad de dientes obturados ha aumentado; de manera que la proporción de pacientes con más de 11 obturaciones es estadísticamente mayor que la proporción de pacientes del resto de las categorías de las obturaciones planteadas.

**Tabla 2.23. Tabla de contingencia de las variables ingreso económico mensual familiar y frecuencia de dientes obturados.**

INGRESO ECONOMICO MENSUAL	Dientes obturados							
	<= 5		6 - 10		11 +		Total	
	Frec	%	Frec	%	Frec	%	Frec	%
<b>Menor o igual a \$100</b>	330	31.3%	60	5.7%	28	2.7%	418	39.7%
<b>De \$101 a \$150</b>	161	15.3%	37	3.5%	8	.8%	206	19.6%
<b>De \$151 a \$200</b>	127	12.1%	31	2.9%	17	1.6%	175	16.6%
<b>De \$201 a \$250</b>	43	4.1%	33	3.1%	6	.6%	82	7.8%
<b>De \$251 a \$300</b>	58	5.5%	14	1.3%	8	.8%	80	7.6%
<b>Más de \$300</b>	50	4.7%	14	1.3%	28	2.7%	92	8.7%
<b>Total</b>	769	73.0%	189	18.0%	95	9.0%	1053	100.0%

**Tabla 2.24. Pruebas de chi-cuadrado de Pearson de las variables ingreso económico mensual familiar y Frecuencia de dientes obturados**

		Dientes obturados
INGRESO ECONOMICO MENSUAL	Chi cuadrado	93.754
	gl	10
	Sig.	.000*
Los resultados se basan en filas y columnas no vacías de cada subtabla más al interior.		
*. El estadístico de chi-cuadrado es significativo en el nivel 0.05.		

**Tabla 2. 25 Comparaciones de proporciones de columnas<sup>a</sup> de las variables de ingreso económico mensual familiar y frecuencia de dientes obturados**

INGRESO ECONOMICO MENSUAL	Dientes obturados		
	<= 5	6 - 10	11 +
	(A)	(B)	(C)
Menor o igual a \$100	B C		
De \$101 a \$ 150	C	C	
De \$ 151 a \$ 200			
De \$ 201 a \$ 250		A C	
De \$ 251 a \$ 300			
Más de \$ 300			A B
Los resultados se basan en pruebas bilaterales con un nivel de significación 0.05. Para cada par significativo, la clave de la categoría con la proporción de columna menor aparece debajo de la categoría con mayor proporción de columna.			
a. Utilizando la corrección de Bonferroni, se han ajustado las pruebas para todas las comparaciones por pares dentro de una fila para cada subtabla situada más al interior.			

**Objetivo específico 6:**

***“Determinar si el ingreso económico mensual del pacientes atendidos influye en la condición de la cantidad de dientes cariados”***

Hipótesis a validar:

$H_0$  : La proporción de los pacientes atendidos según el ingreso económico mensual no son diferentes en cuanto a la condición de la cantidad de dientes cariados.

$H_1$  : La proporción de los pacientes atendidos según su ingreso económico mensual es diferente en cuanto a la condición de la cantidad de dientes cariados.

En la Tabla 2.26 se presentan los resultados de relacionar las variables ingreso económico mensual familiar y la cantidad de dientes afectados por caries. Los resultados muestran que el 41.2% de los pacientes presentaron caries en 5 dientes o menos y que el 25.8% presentaron caries en 11 dientes o más, al pretender establecer

relación entre estas condiciones se efectúa la prueba de contraste (Tabla 2.27), la prueba de Chi-cuadrado de Pearson, en la que los resultados arrojan que existe dependencia entre ambas (dado que  $p < 0.05$ ), es decir que el factor económico es relevante en el estado de salud de sus dientes. De la Tabla 2.28, se obtiene la lectura que los pacientes que tienen ingresos menores o iguales a \$100 tienden a sufrir más problemas de caries activa (la casilla que cruza los ingresos antes mencionados con más de 11 dientes con caries es estadísticamente mayor que las proporciones de las casillas A y B) y en un estado inversamente proporcional se podría comentar (según la Tabla 2.28) lo que ocurre de la relación entre las categorías ingreso económico de \$201 a \$250, dado que los dientes con caries en estos pacientes no son más de 5 dientes, con lo que se podría argumentar que a medida que los ingresos aumentan el tratamiento de los pacientes en relación a las caries disminuye por la posibilidad de poder adquirir los insumos de higiene.

**Tabla 2.26 Tabla de contingencia de las variables ingreso económico mensual familiar y frecuencia de dientes cariados.**

INGRESO ECONOMICO MENSUAL	Dientes cariados							
	<= 5		6 - 10		11 +		Total	
	Frec	%	Frec	%	Frec	%	Frec	%
<b>Menor o igual a \$100</b>	151	14.3%	133	12.6%	134	12.7%	418	39.7%
<b>De \$101 a \$ 150</b>	78	7.4%	75	7.1%	53	5.0%	206	19.6%
<b>De \$ 151 a \$ 200</b>	74	7.0%	60	5.7%	41	3.9%	175	16.6%
<b>De \$ 201 a \$ 250</b>	46	4.4%	25	2.4%	11	1.0%	82	7.8%
<b>De \$ 251 a \$ 300</b>	38	3.6%	25	2.4%	17	1.6%	80	7.6%
<b>Más de \$ 300</b>	47	4.5%	29	2.8%	16	1.5%	92	8.7%
<b>Total</b>	434	41.2%	347	33.0%	272	25.8%	1053	100.0%

**Tabla 2.27 Pruebas de Chi-cuadrado de Pearson de las variables ingreso económico mensual familiar y frecuencia de dientes cariados.**

INGRESO ECONOMICO MENSUAL		Dientes cariados
		Chi cuadrado
	gl	10
	Sig.	.003*
Los resultados se basan en filas y columnas no vacías de cada subtabla más al interior.		
*. El estadístico de chi-cuadrado es significativo en el nivel 0.05.		

**Tabla 2.28 Comparaciones de proporciones de columnas<sup>a</sup> de las variables ingreso económico mensual familiar y frecuencia de dientes cariados.**

INGRESO ECONOMICO MENSUAL	Dientes cariados		
	<= 5	6 - 10	11 +
	(A)	(B)	(C)
Menor o igual a \$100			A B
De \$101 a \$ 150			
De \$ 151 a \$ 200			
De \$ 201 a \$ 250	C		
De \$ 251 a \$ 300			
Más de \$ 300			
Los resultados se basan en pruebas bilaterales con un nivel de significación 0.05. Para cada par significativo, la clave de la categoría con la proporción de columna menor aparece debajo de la categoría con mayor proporción de columna.			
a. Utilizando la corrección de Bonferroni, se han ajustado las pruebas para todas las comparaciones por pares dentro de una fila para cada subtabla situada más al interior.			

**Objetivo específico 7:**

***“Determinar si el ingreso económico mensual del pacientes influye en la cantidad de dientes perdidos”***

Hipótesis a validar:

$H_0$  : La proporción de los pacientes atendidos según el ingreso económico mensual no son diferentes en cuanto a la condición de la cantidad de dientes perdidos.

$H_1$  : La proporción de los pacientes atendidos según el ingreso económico mensual es diferente en cuanto a la condición de la cantidad de dientes perdidos.

Siguiendo la misma metodología para el tratamiento de la información en la Tabla 2.29 se presentan las frecuencias y porcentajes de la relación entre las variables ingreso económico mensual y la cantidad de dientes perdidos. Se observa en los resultados globales que el porcentaje de pacientes que han perdido 5 dientes o menos es de 76.2% y un total de 12.3% de pacientes que han perdido más de 11 dientes de su dentadura normal. Al igual que en los últimos dos casos investigados, interesa establecer si existe algún tipo de relación entre el ingreso económico de los pacientes y la condición de pérdidas de los dientes. Según la Tabla 2.30 al efectuar el proceso nos muestra que existe suficiente evidencia para establecer una relación de dependencia ( $p < 0.05$ ).

Para fundamentar los resultados anteriores se observa en la Tabla 2.31 que a medida aumentan los ingresos económicos del paciente la pérdida de dientes disminuye, como ocurre con la categoría de los ingresos menores a \$100. Las proporciones en la columna B y C son estadísticamente menores que la proporción de la columna A, lo que indica que pacientes en la categoría ingresos menores a \$100 tiene más de 6 dientes perdidos, caso contrario ocurre en la categoría ingresos de \$201 a \$250 la cantidad de dientes perdidos ha disminuido. Finalmente diremos que la columna A es estadísticamente mayor que la columna C, por lo tanto, la cantidad de pacientes que han perdido 5 dientes o menos es mayor que los pacientes que han perdido más de 11 dientes.

**Tabla 2.29. Tabla de contingencia de las variables ingreso económico mensual familiar y frecuencia de dientes perdidos**

INGRESO ECONOMICO MENSUAL	Dientes perdidos							
	<= 5		6 - 10		11 +		Total	
	Frec	%	Frec	%	Frec	%	Frec	%
<b>Menor o igual a \$100</b>	280	26.6%	60	5.7%	78	7.4%	418	39.7%
<b>De \$101 a \$ 150</b>	158	15.0%	22	2.1%	26	2.5%	206	19.6%
<b>De \$ 151 a \$ 200</b>	147	14.0%	14	1.3%	14	1.3%	175	16.6%
<b>De \$ 201 a \$ 250</b>	74	7.0%	5	.5%	3	.3%	82	7.8%
<b>De \$ 251 a \$ 300</b>	68	6.5%	10	.9%	2	.2%	80	7.6%
<b>Más de \$ 300</b>	75	7.1%	10	.9%	7	.7%	92	8.7%
<b>Total</b>	802	76.2%	121	11.5%	130	12.3%	1053	100.0%

(F=Frecuencia; %= Porcentaje total de la tabla)

**Tabla 2.30 Pruebas de chi-cuadrado de Pearson de las variables ingreso económico mensual familiar y frecuencia de dientes perdidos.**

		Dientes perdidos
INGRESO ECONOMICO MENSUAL	Chi cuadrado	45.633
	gl	10
	Sig.	.000*
Los resultados se basan en filas y columnas no vacías de cada subtabla más al interior.		
*. El estadístico de chi-cuadrado es significativo en el nivel 0.05.		

**Tabla 2.31 Comparaciones de proporciones de columnas<sup>a</sup> de las variables ingreso económico mensual familiar y Frecuencia de dientes perdidos.**

INGRESO ECONOMICO MENSUAL	Dientes perdidos		
	<= 5	6 - 10	11 +
	(A)	(B)	(C)
Menor o igual a \$100		A	A
De \$101 a \$ 150			
De \$ 151 a \$ 200			
De \$ 201 a \$ 250	C		
De \$ 251 a \$ 300	C	C	
Más de \$ 300			
Los resultados se basan en pruebas bilaterales con un nivel de significación 0.05. Para cada par significativo, la clave de la categoría con la proporción de columna menor aparece debajo de la categoría con mayor proporción de columna.			
a. Utilizando la corrección de Bonferroni, se han ajustado las pruebas para todas las comparaciones por pares dentro de una fila para cada subtabla situada más al interior.			

**Objetivo específico 8:**

***“Determinar si la edad del paciente influye en la cantidad de dientes perdidos”***

Hipótesis a validar:

$H_0$  : La proporción de los pacientes atendidos según la edad no son diferentes en cuanto a la cantidad de dientes perdidos.

$H_1$  : La proporción de los pacientes atendidos según la edad es diferente en cuanto a la cantidad de dientes perdidos.

Los resultados obtenidos se presenta la Tabla 2.32, se obtiene que el 67.40% de los pacientes con 40 o menos años de edad ya perdieron dientes, también cabe destacar que a medida aumenta la edad la frecuencia de pacientes que asisten a la atención odontológica en las unidades de salud familiar disminuye. El resultado del contraste realizado de la relación entre ambas condiciones (edad y perdida de dientes) se presenta en la Tabla 2.33 en la que muestra la dependencia entre ambas variables, por

lo que se dice que existen pruebas suficientes, ya que es claro que a medida que aumenta la edad el nivel de la salud bucal está en riesgo. De acuerdo a la Tabla 2.34 en la cantidad de dientes perdidos menores o igual a 5 comparada con la columna B y C afirma que es estadísticamente menor que la columna A, lo que se interpreta que las personas con 40 años o menos han perdido a lo sumo 5 dientes. También es claro pensar y evidente según los resultados, que a medida se incrementa la edad de igual manera se incrementa la pérdida de dientes, y es lo que se evidencia en las columnas B y C que son estadísticamente mayor que la columna A.

**Tabla 2.32 Comparaciones de proporciones de columnas<sup>a</sup> de las variables cantidad de dientes perdidos y grupo etario.**

Cantidad de dientes perdidos	Grupos de edad							
	<= 40		41 - 60		61+		Total	
	Frec	%	Frec	%	Frec	%	Frec	%
<= 5	709	60.20%	158	13.40%	26	2.20%	893	75.80%
6 -10	61	5.20%	58	4.90%	19	1.60%	138	11.70%
11 - 15	13	1.10%	27	2.30%	17	1.40%	57	4.80%
16 - 20	7	0.60%	13	1.10%	15	1.30%	35	3.00%
21+	4	0.30%	15	1.30%	36	3.10%	55	4.70%
<b>Total</b>	794	67.40%	271	23.00%	113	9.60%	1178	100.00%

**Tabla 2.33 Pruebas de chi-cuadrado de Pearson dientes perdidos y grupos etarios**

		Edad por grupos
PERDIDO (agrupado)	Chi cuadrado	426.59
	gl	8
	Sig.	.000*
Los resultados se basan en filas y columnas no vacías de cada subtabla más al interior.		
*. El estadístico de chi-cuadrado es significativo en el nivel 0.05.		

**Tabla 2.34 Comparaciones de proporciones de columnas<sup>a</sup> de las variables dientes perdidos y grupos etarios.**

Cantidad de dientes perdidos	Grupos de edad		
	<= 40	41 - 60	61+
	(A)	(B)	(C)
<= 5	B C	C	
6 -10		A	A
11 - 15		A	A
16 - 20		A	A B
21+		A	A B
Los resultados se basan en pruebas bilaterales con un nivel de significación 0.05. Para cada par significativo, la clave de la categoría con la proporción de columna menor aparece debajo de la categoría con mayor proporción de columna.			
a. Utilizando la corrección de Bonferroni, se han ajustado las pruebas para todas las comparaciones por pares dentro de una fila para cada subtabla situada más al interior.			

## 2.4. APLICACIÓN DE REGRESIÓN LOGÍSTICA MULTINOMIAL.

Según la Organización Mundial para la Salud (OMS), se define la salud como “Un estado de completo bienestar físico, mental y social, y no solamente la ausencia de afecciones o enfermedades” [19].

La cita procede del Preámbulo de la Constitución de la OMS, que fue adoptada por la Conferencia Sanitaria Internacional, celebrada en Nueva York del 19 de junio al 22 de julio de 1946, firmada el 22 de julio de 1946 por los representantes de 61 Estados y entró en vigor el 7 de abril de 1948. La definición no ha sido modificada desde entonces.

En esta investigación se aborda este tema considerando dos componentes esenciales por el comportamiento de la información tanto sociodemográficos como enfermedades del aparato estomatognático. Basados en pruebas y resultados bajo una base científica matemática, se considera la condición de salud bucal; parte del completo estado de bienestar del individuo. Por tanto, resulta de importancia el aporte de esta misma de la comunidad por lo que la metodología de investigación, como se detalla a continuación:

La condición de los pacientes con las características de Placa bacteriana, gingivitis, periodontitis y DAI, en los pacientes que muestran la pérdida las 28 piezas dentales se denota como **“No evaluable”** debido a que según el cálculo de cada uno de los índices, requiere de ciertas condiciones para establecer la condición de salud bucal entre ellas, que se encuentre el diente a evaluar o el diente vecino, aunque esto no sugiere que el paciente este bajo condición de salud bucal sana, al contrario la condición del paciente es severa y por ella no quedan excluidos de la muestra. Pero para el caso contrario, la clasificación de la condición de salud bucal de cada paciente se realiza de acuerdo a lo establecido en la Tabla 2.35. **Por ejemplo, al seleccionar al i-ésimo paciente y éste presenta las siguientes condiciones:**

Placa bacteriana: Buen estado o Condición leve.

Número de dientes con caries: 0-1; es decir ningún diente con caries (0) y posee un (1) diente con caries.

Número de dientes perdidos: 0-1; es decir ningún diente con caries (0) y posee un (1) diente con caries.

Gingivitis: Buen estado

Índice periodontal: nivel sano

Oclusión: Normal y mal oclusión mínima

Todas las anteriores variables explicativas (Xi) dan paso para generar la variable respuesta (Yi) resultando

**Condición de salud bucal del paciente: Condición sana**

De igual manera, se realizó el mismo procedimiento de acuerdo a las características de cada uno de los individuos con el propósito de lograr la óptima clasificación.

**Tabla 2.35 criterios para la clasificación de la variable dependiente.**

Índices \ Condición de salud	Condición Sana	Condición Leve	Condición Moderada	Condición Severa
Placa Bacteriana	Buen estado / o Condición Leve	Condición Leve	Leve o moderada	Moderada o Severa
Cariado - Perdido	0 - 1	1 - 2	3-6	7-28
Gingivitis	Buen estado (0)	Leve (1)	Leve o moderada (1 ó 2)	Inflamación moderada C/ hemorragia al sondaje - Inflamación severa C/ hemorragia espontánea.
Periodontitis	Sano	Hemorragia al sondaje suave	Hemorragia al sondaje suave / Cálculo supra o subgingival	Cálculo supra o subgingival / Bolsa de 4 -5 mm / bolsa igual o mayor de 6 mm
DAI	Oclusión normal Y mal oclusión mínima.	Oclusión normal Y mal oclusión mínima / o Maloclusión definida	Maloclusión definida /o Maloclusión severa / o Maloclusión muy severa o discapacitante	Puede presentar cualquier categoría

### 2.4.1. PARTICIÓN DE LA MUESTRA.

En la presente sección una vez estudiada la información y observado el comportamiento de los datos, se seleccionó la técnica que se ajustó a ella. Por las características de los datos fue importante aplicar el ajuste de un modelo de regresión logístico multinomial, según el objetivo principal planteado al inicio de esta investigación, para el cual fue necesario partir la muestra en dos submuestras.

La metodología para aplicar la técnica de regresión logística fue:

1. Partir la población de pacientes atendidos en la red de salud pública en dos submuestras.
2. Luego, se utilizó el 20% de la población que asciende a 254 pacientes de la base de datos y el 80% se utilizó para modelar la ecuación de regresión logística el número de pacientes asciende a 924 los cuales serán utilizados para la estimación de los parámetros del modelo de los cuales se distribuyen según la Tabla 2.36 en dicha tabla se presentan las categorías en las que se clasifica la condición bucal de los pacientes puede identificarse el 81.6% de los pacientes presentan condición de salud bucal severa y en contraste 1.6% de pacientes poseen condición sana.

El procedimiento que se siguió fue basado en la técnica del muestreo estratificado con afijación proporcional que viene dado por la ecuación:

$$n_k = \sum_{j=1}^k \frac{N_k}{N} n$$

Donde N es la muestra total (1178),  $N_k=224$ , es el 20% del tamaño de la muestra y  $n_k$  son los tamaños de los distintos estratos (K=4, categorías de la variable dependiente), la distribución de tamaños muestrales se presentan en la Tabla 2.36, [7].

**Tabla 2.36 Distribución de frecuencias después de calcular la submuestra**

Condición de Salud Bucal	Frecuencia (N)	Porcentaje marginal
Condición Sana	15	1.6%
Condición Leve	46	5.0%
Condición Moderada	109	11.8%
Condición Severa	754	81.6%
Válidos	924	100.0%
a. La variable dependiente sólo tiene un valor observado en 252 (100.0%) subpoblaciones.		

#### 2.4.2. ESTIMACIÓN DEL MODELO.

Para ello fue necesario elaborar la **Tabla 2.37** en la que se muestra las variables que se utilizaron para ajustar el modelo, en su mayoría son categóricas:

**Tabla 2.37 Lista de variables que intervienen para la construcción del modelo.**

Variable	Categoría
Índice de Placa bacteriana Løe y Silness	Buen estado
	Condición leve
	Condición moderada
	Condición severa
	No evaluable
Índice gingival Løe y Silness	Buen estado
	Inflamación leve
	Inflamación moderada
	Inflamación severa
	No evaluable
Índice Periodontal Comunitario CPI (OMS)	Sano, ausencia de signos
	Hemorragia al sondaje suave
	Cálculo supra o subgingival
	Bolsa de 4 -5 mm
	Bolsa igual o mayor de 6 mm
	No evaluable
Maloclusión, Índice DAI (de OMS)	Oclusión normal y mal oclusión mínima
	Mal oclusión definida
	Mal oclusión severa
	Mal oclusión muy severa o discapacitante

	No evaluable
sexo	FEMENINO
	MASCULINO
ingreso mensual	Menor o igual \$250
	Mayor a \$250
nivel de escolaridad	Sin escolaridad
	Educación básica
	Bachillerato
	Educación superior
Región Geográfica	Zona Oriental
	Zona Central
	Zona Paracentral
	Zona Occidental
Edad del paciente	Edad en años cumplidos
Cariado	Número de dientes con caries
Perdido	Número de dientes perdidos
Obturado	Número de dientes con obturaciones

Por las variables que intervienen en este estudio la estructura del modelo multinomial con efectos principales es:

$$\ln\left(\frac{\pi_{ij}}{\pi_{i3}}\right) = \beta_{j0} + \beta_{j1}X_{ij1} + \beta_{j2}X_{ij2} + \dots + \beta_{jp}X_{ijp}$$

Donde:

$X_{ijk}$  : es la k-esima variable explicativa del individuo i en la categoría j

$\forall i=1,2,\dots,924$   $j=1,2$  dado que la variable dependiente tiene 3 categorías la tercera es de referencia.

Por ejemplo el modelo de la categoría 1 es:

$$\ln\left(\frac{\pi_{i1}}{\pi_{i3}}\right) = \beta_{10} + \beta_{11}X_{i11} + \beta_{12}X_{i12} + \dots + \beta_{1p}X_{i1p}$$

El modelo de la categoría 2 sería:

$$\ln\left(\frac{\pi_{i2}}{\pi_{i3}}\right) = \beta_{20} + \beta_{21}X_{i21} + \beta_{22}X_{i22} + \dots + \beta_{2p}X_{i2p}$$

El modelo de la categoría 3 es:

$$\pi_{i3} = \frac{1}{1 + \sum_{j=1}^2 e^{\beta_{j0} + \beta_{j1}X_{ij1} + \beta_{j2}X_{ij2} + \dots + \beta_{jp}X_{ijp}}}$$

Para la creación del modelo y tratarlo con el software SPSS se consideraron los índices de la condición de salud bucal de los pacientes además de las variables sociodemográficas planteados en la tabla anterior (**Tabla 2.37**). Una vez definido en el software se siguió la ruta *Analizar > Regresión > Logística-Multinomial*. Para ello el modelo se especificó de la siguiente forma:

- a) Dejando como categoría de referencia “Condición de salud severa”.
- b) Método: Personalizada/pasos sucesivos.
- c) Los términos se construyen bajo la condición de interacción.
- d) Se utiliza el método de inclusión por pasos: Entrada hacia adelante, para seleccionar cada uno de los términos y
- e) Los pasos se basan en estimaciones de máxima verosimilitud.

El software ejecutó las instrucciones siguiendo un orden de entrada de las variables más importantes y a la vez estudia si estas han ingresado en un paso anterior y siguen siendo importantes según el método que se menciona en el literal d).

Se tiene como resultado en primer lugar la **Tabla 2.38**, se puede ver en resumen los efectos de la entrada de variables del método antes descrito. Donde resultó que algunas de las variables no presentan un aporte significativo por lo que se estima que

no son necesarias, no tienen efecto para construir el modelo para cada categoría de la variable respuesta, es decir, que son redundantes y no ayudan a clasificar los diferentes niveles de la condición de salud bucal. Para eliminar dichas variables que no aportan al modelo, es conveniente ir eliminando de una a una las que no son significativas ( $p > 0.05$ ), para seleccionar solo aquellas que aportan y son significativas en el modelo (ver Tabla 2.38).

En primer lugar se ajustó el modelo sin variables, que fue el modelo básico (Columna Modelo, fila 0, Tabla 2.38) frente al que se contrastaron los nuevos modelos. Posteriormente, se ajustó otro modelo para cada una de las variables consideradas, incluidas las variables de diseño, y que mediante el test condicional de razón de verosimilitudes, se decide seleccionar cuál de las variables deberían entrar. Atendiendo al principio de parsimonia, finalmente el modelo final que mejor explica la variabilidad de la condición de salud bucal con un menor número de parámetros es el que incluye las variables: “CARIADO” y “PERDIDO”.

**Tabla 2.38 Resumen de pasos mediante el método de selección: Entrada hacia adelante**

Modelo	Acción	Efecto(s)	Criterio de ajuste del modelo	Contrastes de selección de efectos		
			-2 log verosimilitud	Chi-cuadrado <sup>a</sup>	gl	Sig.
0	Introducido	Intersección	764,359			
1	Introducido	CARIADO	480,620	283,740	2	,000
2	Introducido	PERDIDO	196,699	283,921	2	,000
Método por pasos: Entrada hacia adelante						
a. El valor de chi-cuadrado para su inclusión se basa en la prueba de la razón de verosimilitudes.						

### 2.4.3. PARÁMETROS DEL MODELO

Una vez establecidas las variables del modelo, se muestra la Tabla 2.39 que presenta los resultados de las estimaciones de los parámetros del modelo, desviación estándar (E.T.), Estadístico de Wald, grados de libertad(gf), Nivel de significancia(Sig.) y los valores de la razón de ventaja (Exp(B)) y sus respectivos intervalos de confianza al 95% (I.C. 95% para Exp(B)), con lo que se valoran dos aspectos importantes el primero es la significancia de los parámetros (estadístico de Wald) que acompañaran a la variable explicativa en el logit y el segundo aspecto es el signo que acompaña al parámetro.

Por motivo de tener muy pocos casos la categoría de la variable respuesta, condición sana (con 18 casos), se tomó a bien agrupar las categorías “sano” y “leve”, por lo que el resto del proceso se ejecutará con 3 categorías. Como ya se había mencionado; además, se toma como referencia la categoría “**Condición de salud bucal severa**”; por lo que se construyen dos modelos para las dos categorías restantes “**Condición de salud bucal sana o leve**” y “**Condición de salud bucal moderada**” que presentamos a continuación:

$$g_1 = \left( \frac{\pi_{SL}}{\pi_S} \right) = \ln \left( \frac{\pi_{SL}}{\pi_S} \right) = 3.822 - 1.790 * PERDIDO - 1.293 * CARIADO$$

$$g_2 = \left( \frac{\pi_M}{\pi_S} \right) = \ln \left( \frac{\pi_M}{\pi_S} \right) = 2.453 - 0.551 * PERDIDO - 0.623 * CARIADO$$

Dónde:

$\pi_{SL}$ , como la probabilidad estimada que tiene el paciente “i-ésimo” de pertenecer a la categoría “**Condición de salud bucal sana o leve**” dada la información de las variables predictoras en cuestión.

$\pi_M$ , como la probabilidad estimada que tiene el paciente “i-ésimo” de pertenecer a la categoría “**Condición de salud bucal moderada**” dada la información de las variables predictoras en cuestión.

$\pi_S$ , como la probabilidad estimada que tiene el paciente “i-ésimo” de pertenecer a la categoría “**Condición de salud bucal severa**” dada la información de las variables predictoras en cuestión.

Obteniéndose de esta manera cada una de las siguientes probabilidades estimadas:

$$\pi_{SL} = \frac{e^{g_1\left(\frac{\pi_{SL}}{\pi_S}\right)}}{1 + e^{g_1\left(\frac{\pi_{SL}}{\pi_S}\right)} + e^{g_2\left(\frac{\pi_M}{\pi_S}\right)}}, \quad \pi_M = \frac{e^{g_2\left(\frac{\pi_M}{\pi_S}\right)}}{1 + e^{g_1\left(\frac{\pi_{SL}}{\pi_S}\right)} + e^{g_2\left(\frac{\pi_M}{\pi_S}\right)}} \quad \text{y} \quad \pi_{SL} = \frac{1}{1 + e^{g_1\left(\frac{\pi_{SL}}{\pi_S}\right)} + e^{g_2\left(\frac{\pi_M}{\pi_S}\right)}}$$

Además en el programa SPSS en la opción Analizar>Regresión>Logística-Multinomial permite establecer el contraste de significación de los coeficientes bajo los supuestos que:

1.El contraste de significación de los coeficientes se realiza bajo el supuesto:

$$H_0 : \beta_{ij} = 0$$

Que significa probar la hipótesis de que cada coeficiente que acompaña a la variable correspondiente en el modelo logístico multinomial para la categoría j y la variable i asume el valor de cero. Esta prueba como ya se mencionó, se puede realizar con el Estadístico de Wald; rechazando si el valor de este estadístico es mayor que el valor en tablas para una Chi-cuadrado con un grado de libertad a un nivel de significancia de  $\alpha$  ( $p > 0.05$ ).

Al revisar este criterio se observó en la Tabla 2.39, que todos los parámetros son significativas para el modelo logit **condición de salud bucal sano y leve** y también para el modelo logit **condición de salud bucal moderado**, el p-valor de cada parámetro ( $\beta_{ij}$ )

es significativo, está claro que rechazamos la hipótesis nula, concluyéndose así, que cada parámetro estimado es distinto de cero y que si son adecuados para la construcción del modelo.

2. También otra consideración interesante a realizarse sobre la estimación de los parámetros es el signo de estos, observándose que cuando  $\beta_{ij} < 0$  significa que el odds ratio decrece y que cuando  $\beta_{ij} > 0$  la odds ratio crece, a continuación se realizan las siguientes interpretaciones para cada uno de los logit:

- Para la categoría **“Condición de salud bucal sana o leve”**  $\beta_{11}, \beta_{21} < 0$  por lo tanto si el paciente presente dientes perdidos y dientes con caries son factores que disminuyen la probabilidad que el paciente sea clasificado bajo **“Condición de salud sana y leve”** en comparación a que el paciente presente **“Condición de salud severa”**.

- Ahora para la categoría **“Condición de salud bucal moderada”** los parámetros  $\beta_{12}, \beta_{22} < 0$  interpretándose como el hecho que cuando el paciente presente dientes perdidos y dientes con caries son factores que disminuyen la probabilidad que el paciente sea clasificado bajo **“Condición de salud moderada”** en comparación a que el paciente presente **“Condición de salud severa”**.

**Tabla 2.39. Estimaciones de los parámetros**

Salud bucal		B	Error tip.	Wald	gl	Sig.	Exp (B)	Intervalo de confianza al 95% para Exp(B)	
								Límite inferior	Límite superior
sano y leve	Intersección	3,822	,451	71,931	1	,000			
	PERDIDO	-1,790	,381	22,121	1	,000	,167	,079	,352
	CARIADO	-1,293	,146	78,180	1	,000	,274	,206	,365
moderado	Intersección	2,453	,336	53,241	1	,000			
	PERDIDO	-,551	,077	51,850	1	,000	,576	,496	,670
	CARIADO	-,623	,064	95,212	1	,000	,536	,473	,608

a. La categoría de referencia es: severo.

Además en la Tabla 2.39 como ya se había mencionado tiene los siguientes detalles:

El Odds ratio ( $OR = \text{Exp}(B)$ ) que es un concepto ampliamente utilizado en la investigación biomédica se ha traducido de múltiples formas al español: razón de oportunidades, razón de posibilidades, oportunidad relativa, razón de probabilidades o razón de productos cruzados, e incluso algo tan extraño como «razón de momios». Una buena opción que sirve para evitar confusiones y se ha hecho mayoritaria es incorporar directamente el término inglés y decir siempre “Odds ratio”.

Una OR es el cociente entre dos Odds, carece de unidades de medida y la Odds del grupo de interés se debe colocar siempre en el numerador y la de referencia en el denominador. O sea que es el cociente de ventaja de respuesta  $Y_j$  frente a la última categoría  $Y_k$  cuando aumenta en una unidad una de las variables predictoras y las demás se mantienen fijas.

**Exp(B):** Este estadístico indica cuanto aumenta (o disminuye) la probabilidad de éxito/fracaso luego de que X aumenta una unidad. El valor de 1 indica que la variable no influye, valores superiores a 1 indican aumento y valores inferiores a 1 indican disminución.

**Límite inferior y superior:** Son los límites del intervalo de confianza para el verdadero valor de  $\text{Exp}(B)$ , con 95% de probabilidad. Por lo tanto, si este intervalo de confianza incluye 1, no se descarta la hipótesis nula de que estos “odds ratio” no cambian al variar X. Debido a que como ninguno de los intervalos contienen el uno rechazamos la hipótesis nula (ver Tabla 2.39).

Además al interpretar cada uno de los Odds Ratio (Columna  $\text{Exp}(B)$ , Tabla 2.39) se obtuvo que cuando el paciente tiene dientes perdidos y dientes con caries la posibilidad que este sea clasificado bajo **condición de salud bucal sano/leve** es nada

más 16.7% y 27.4% respectivamente es claro que cuando el paciente presenta este tipo de afecciones la posibilidad de gozar de una condición de salud bucal sana o leve es baja.

Asimismo cuando la condición de salud bucal es moderada y el paciente presenta dientes perdidos y dientes con caries 57.6% y 53.6% respectivamente, son las posibilidades que este tiene de ser clasificado bajo condición moderada es decir que se asemeja más a la realidad de poseer también condición bucal severa.

#### 2.4.4. BONDAD DEL AJUSTE.

En este bloque inicial se calcula la verosimilitud de un modelo que sólo tiene el término constante  $\beta_0$ . Puesto que la verosimilitud  $L$  es un número muy pequeño (comprendido entre 0 y 1), se suele ofrecer el logaritmo neperiano de la verosimilitud (LL), que es un número negativo, o menos dos veces el logaritmo neperiano de la verosimilitud (-2LL), que es un número positivo.

El estadístico (-2LL) mide hasta qué punto un modelo se ajusta bien a los datos. El resultado de esta medición recibe también el nombre de desviación. Cuanto más pequeño sea el valor, mejor será el ajuste; por lo que resulta ser significativo el modelo final ( $p < 0.05$ ). En general, la razón de verosimilitudes (RV = -2LL) es útil, para determinar si hay una diferencia significativa entre incluir en el modelo todas las variables y no incluir ninguna, dicho de otro modo, la RV sirve para evaluar si las variables tomadas en conjunto, contribuyen efectivamente a 'explicar' las modificaciones que se producen en  $\pi_{ij}$  o las probabilidades que un individuo sea clasificado bajo una condición de salud bucal cuando en realidad su condición es diferente.

Se plantean las siguientes hipótesis con el objetivo de probar la bondad de ajuste:

$$H_0: \pi_{ij} = \frac{e^{\left(\sum_{q=0}^p \beta_{qj} x_{ijq}\right)}}{1 + \sum_{j=1}^k e^{\left(\sum_{q=0}^p \beta_{qj} x_{ijq}\right)}} \quad \forall i = 1, \dots, n ; \quad \forall j = 1, \dots, k$$

$$H_1 : \pi_{ij} \neq \frac{e^{\left(\sum_{q=0}^p \beta_{qj} \cdot x_{ijq}\right)}}{1 + \sum_{j=1}^{k-1} e^{\left(\sum_{q=0}^p \beta_{qj} \cdot x_{ijq}\right)}} \text{ para algún } i, j.$$

La bondad y ajuste de este modelo logístico multinomial se puede contrastar con la información de la siguiente **Tabla 2.40** donde el valor de la RV cuando el modelo es contemplado con solo la intersección tiene un valor alto (764,359) en comparación al valor de la RV del modelo final (196,699) reflejándose de esta manera, que existe una diferencia significativa entre incluir en el modelo todas las variables y no incluir ninguna, con un valor de  $\chi^2 = 567.661$ , 4 grados de libertad, basando los resultados en  $p < 0.05$ , de manera que se concluye que las variables que han sido seleccionadas para presentar los modelos logit en conjunto contribuyen efectivamente a explicar las modificaciones que se producen en las probabilidades que un individuo sea clasificado bajo una condición de salud bucal según sea la de este.

**Tabla 2.40 Información del ajuste del modelo**

Modelo	Criterio de ajuste del modelo	Contrastes de la razón de verosimilitud		
	-2 log verosimilitud	Chi-cuadrado	gl	Sig.
Sólo la intersección	764,359			
Final	196,699	567,661	4	,000

Otros estadísticos que son muy utilizados para probar la bondad de ajuste del modelo son los presentados en la **Tabla 2.41** donde los estadísticos Pearson y Desviación toman valores de 151.461 y 127.457 respectivamente, con  $(N - p)(k - 1) = (252 - 3)(3 - 1) = 498$  grados de libertad y  $p > 0.05$ .

Donde N representa las clases covariantes (252, dato que brinda el programa SPSS), p es el número de parámetros en el modelo y k el número de categorías de la variable dependiente, de este modo no se podría rechazar la hipótesis nula, y se concluye que ambos estadísticos Pearson y Desviación indican que el modelo se ajusta adecuadamente a las observaciones.

**Tabla 2.41 Bondad de ajuste**

	Chi-cuadrado	gl	Sig.
Pearson	151,461	498	1,000
Desviación	127,457	498	1,000

También para cuantificar la bondad del ajuste global del modelo se utiliza la tasa de clasificaciones correctas, que permite clasificar cada observación en la categoría más probable, construyendo así una tabla de clasificación de los valores observados y predichos

#### **2.4.5. CALIDAD DEL AJUSTE DEL MODELO.**

Para medir la calidad del ajuste del modelo se utilizan los coeficientes Pseudo-R<sup>2</sup> de Mc-Fadden, de Cox-Snell y de Nagelkerke. El cálculo de éstos depende del valor de las funciones de log-verosimilitud del modelo final y del modelo inicial con solo la constante

Interpretación por separado de los coeficientes Pseudo-R<sup>2</sup>:

- ✓ La R cuadrado de Mc-Fadden, muy raramente alcanza valores cercanos a uno, suele considerarse una buena calidad del ajuste cuando oscila entre 0.2 y 0.4 y es excelente calidad de ajuste para valores superiores y malo para valores menores a 0.2.

- ✓ La R cuadrado de Cox y Snell es un coeficiente de determinación generalizado que se utiliza para estimar la proporción de varianza de la variable dependiente explicada por las variables predictoras (independientes). Los valores oscilan entre 0 y 1.
- ✓ La R cuadrado de Nagelkerke es una versión corregida de la R cuadrado de Cox y Snell. La R cuadrado de Cox y Snell tiene un valor máximo inferior a 1, incluso para un modelo "perfecto". La R cuadrado de Nagelkerke corrige la escala del estadístico para cubrir el rango completo de 0 a 1

Seguidamente en la **Tabla 2.42** se presentan las tres medidas resumen de los modelos mencionadas anteriormente, para evaluar de forma global su validez. Los coeficientes de determinación tienen valores muy adecuados, indicando que el 47.4%, 68% o el 53.7% de la variación de la variable dependiente es explicada por las variables incluidas en el modelo.

Se puede observar en la **Tabla 2.42** los coeficientes pseudo-R<sup>2</sup>, alcanzan valores moderadamente cercanos a 1, por lo que se puede concluir que el modelo presenta buena calidad de ajuste.

**Tabla 2.42 Pseudo R-cuadrado**

Cox y Snell	.474
Nagelkerke	.680
McFadden	.537

#### **2.4.6. VALIDACIÓN DEL MODELO**

Una vez encontrado el mejor modelo, tiene que ser validado. En un modelo predictivo se trata de ver si el modelo vaticina bien la variable dependiente en un nuevo paciente encuestado o a un conjunto de datos nuevos.

Se había mencionado anteriormente que la muestra es N=1178 casos válidos observados, para este análisis fue dividida en dos submuestras, una para el análisis y la otra con fines de validación, la primera con un tamaño de 924 casos/paciente y la segunda con 254 casos/pacientes.

En la **Tabla 2.43** se presentan algunos casos de ejemplo de la clasificación que realiza el modelo cuando las probabilidades  $\hat{\pi}_{SL}$ ,  $\hat{\pi}_M$  y  $\hat{\pi}_S$  son calculadas. Se define los siguientes criterios de clasificación:

1. Si se cumple la condición que  $\hat{\pi}_{SL} \geq \hat{\pi}_M$  y  $\hat{\pi}_{SL} > \hat{\pi}_S$  como consecuencia se clasificaría al paciente bajo **condición de salud bucal sana/leve**.
2. Si se cumple la condición que  $\hat{\pi}_M > \hat{\pi}_{SL}$  y  $\hat{\pi}_M \geq \hat{\pi}_S$  como consecuencia se clasificaría al paciente bajo **condición de salud bucal moderada**.
3. Si se cumple la condición que  $\hat{\pi}_S > \hat{\pi}_{SL}$  y  $\hat{\pi}_S > \hat{\pi}_M$  como consecuencia se clasificaría al paciente bajo **condición de salud bucal severa**.

Se observa claramente que el paciente de la posición 13 a sido clasificado bajo condición de salud bucal moderada cuando la condición real es severa y caso parecido ocurre con el paciente de la posición 15 que se ha clasificado bajo condición de salud severa cuando en realidad su condición es moderada, por lo que se concluye que de 15 paciente tomados de forma al azar 13 se han logrado clasificar correctamente.

**Tabla 2.43 Probabilidades de clasificación del modelo, Condición de salud real y pronosticada del paciente.**

Paciente	$\hat{\pi}_{SL}$	$\hat{\pi}_M$	$\hat{\pi}_S$	Condición de salud Pronosticada	Condición de salud Real
1	0	0.2	0.8	Condición Severa	Condición Severa
2	0	0.5	0.5	Condición Severa	Condición Severa
3	0	0.1	0.9	Condición Severa	Condición Severa
4	0	0.2	0.8	Condición Severa	Condición Severa
5	0	0.4	0.6	Condición Severa	Condición Severa
6	0	0.2	0.8	Condición Severa	Condición Severa
7	0	0.7	0.3	Condición Moderada	Condición Moderada
8	0	0.3	0.7	Condición Severa	Condición Severa
9	0	0.1	0.9	Condición Severa	Condición Severa
10	0	0.1	0.9	Condición Severa	Condición Severa
11	0	0.7	0.3	Condición Moderada	Condición Moderada
12	0	0.6	0.4	Condición Moderada	Condición Moderada
13	0	0.6	0.4	Condición Moderada	Condición Severa
14	0	0.4	0.6	Condición Severa	Condición Severa
15	0	0.4	0.6	Condición Severa	Condición Moderada

Otra manera de poder validar el modelo encontrado es por medio de las Tablas de Clasificación Tabla 2.44 y 2.45 muestran lo bien que el modelo clasifica correctamente los casos tanto para la muestra del 80% que se utilizó para estimar el modelo como también para el 20% utilizado para la validación; en este caso el 87.4% y 83.9% respectivamente, un modelo perfecto sería el que por medio de las tablas muestre únicamente valores en la diagonal y en el resto de las casillas cero reflejando así que clasifica correctamente todos los casos se puede concluir acerca de la eficacia predictiva del modelo y es evidente que posee una muy buena predicción o sea los casos analizados logran ser correctamente clasificados, algunos autores mencionan que un modelo es adecuado cuando clasifica de manera inadecuada al menos el 20% para este caso clasifica menos del 14% de forma inadecuada.

**Tabla 2.44. Tabla de Clasificación para el 80% de datos utilizados para la realización del modelo de regresión logístico.**

Observado	Pronosticado			
	Condición sano y leve	Condición moderado	Condición severo	Porcentaje correcto
Condición sano y leve	41	17	3	67.2%
Condición moderado	11	42	56	38.5%
Condición severo	10	19	725	96.2%
Porcentaje global	6.7%	8.4%	84.8%	87.4%

**Tabla 2.45. Tabla de Clasificación para el 20% de datos utilizados para la validación del modelo de regresión logístico encontrado.**

Observado	Pronosticado			
	Condición Sana y Leve	Condición Moderada	Condición Severa	Porcentaje correcto
Condición Sana y Leve	12	2	9	52,2%
Condición Moderada	5	1	17	4,3%
Condición Severa	7	1	200	96,2%
Porcentaje global	9,4%	1,6%	89,0%	83,9%

## CONCLUSIONES

- 1.El sexo del paciente no es un factor que está directamente relacionado con la condición de salud bucal, aunque sea el sexo femenino el que asistió con más frecuencia (66.3%) que los hombres, a la consulta odontológica en la red de salud pública durante el año 2009.
- 2.El ingreso económico mensual percibido por el paciente es un factor directamente relacionado con la condición de salud bucal del mismo, con una significancia estadística menor a 0.05 ( $p= 0.001$ ). El 77% de los pacientes que perciben ingresos inferiores a \$250.00 mensuales presentan una condición severa de salud bucal.
- 3.El nivel académico alcanzado por los pacientes es estadísticamente significativo ( $p < 0.05$ ) al relacionarlo con la condición de salud bucal; dado que sólo el 8% de pacientes que alcanzan educación superior presentan una condición severa en su salud bucal, caso contrario cuando se alcanza únicamente educación básica el 58% posee condición severa.
- 4.La región de residencia es otro factor relacionado a la calidad de salud bucal en los salvadoreños, esto se debe a que en la zona central se encuentra la capital y la mayor concentración de la población, al igual que el conglomerado de consultorios odontológicos de la red pública, así los que asistieron a la consulta odontológica presentan en un 61% condición sana y un 42% condición severa; por lo que resultó estadísticamente significativo ( $p < 0.05$ ).
- 5.El ingreso económico de los pacientes que acuden a los servicios de la salud pública se relaciona con la cantidad de dientes obturados del paciente pues estadísticamente es significativamente mayor la proporción (39.7%) de pacientes con ingresos menores a \$100 dólares que acuden a estas instituciones a recibir este tipo de tratamiento. Y solo el 8.7% de los que perciben ingresos superiores a

\$300.00 dólares; así el ingreso económico mensual resultó ser estadísticamente significativo ( $p < 0.05$ ).

6.A medida los ingresos económicos mensuales aumentan las lesiones de caries activas disminuyen de forma significativa estadísticamente ya que la mayor proporción de pacientes con ingresos menores a \$100 tiene más de 11 dientes con caries (12.7%) que los que perciben ingresos mayor a \$300.00 pues sólo el 1.5 % de estos presentan más de 11 dientes con lesión cariosa, por lo que la prueba resultó estadísticamente significativo ( $p < 0.05$ ).

7.El ingreso económico mensual también resulta estar relacionado con la pérdida de dientes, evidenciando que cuando los ingresos son menores o iguales a \$100 el 26.6% ha perdido entre 1 y 5 dientes; y solo el 7.1% de los que tienen ingresos mayores a \$300 han sufrido igual pérdida, así la prueba es estadísticamente significativa ( $p < 0.05$ ).

8.La edad resulta ser otro factor relacionado con la pérdida de dientes, ya que de los 55 pacientes que perdieron más de 21 dientes el 7.2 % (4 pacientes) son del grupo etario de 20 a 40 años; incrementándose el porcentaje en los mayores de 61 años con un 65.45% (36 pacientes), por tanto fue estadísticamente significativa la prueba ( $p < 0.05$ ).

9. Las variables que resultaron ser significativas para estimar el modelo más adecuado que representa la condición de salud bucal, según el método de selección entrada hacia adelante son: número de dientes cariados, número de dientes perdidos, Zona de residencia del paciente.

10.Utilizando el Test de Wald para comparar si el conjunto de covariables elegidas explica la condición de salud bucal de esa población de estudio, resultó que, los coeficientes son significativamente distintos de cero para cada logit, Condición de salud sana/leve y Condición de salud moderada. Además los intervalos de

confianza de la razón de ventaja excluyen el valor 1 indicando de esta manera que las variables son de interés en el modelo.

11. Los datos pueden modelarse con un ajuste de regresión logístico multinomial los test de Chi cuadrado de Pearson y el de Chi cuadrado de razón de verosimilitud (Wills) confirman que la hipótesis  $H_0$  se acepta. Es decir que el modelo clasifica adecuadamente el 87.4% la condición de salud bucal que posee el paciente según la condición real del mismo, únicamente el 12.6% no es clasificado correctamente.
12. La calidad de ajuste del modelo se presenta con los coeficientes Pseudo- $R^2$  siendo, Cox y Snell el coeficiente de determinación generalizado, que se utiliza para estimar la proporción de varianza de la variable dependiente, explicada por las variables predictoras (independientes) cuyo valor es de 47.4% y Nagelkerke (es el coeficiente corregido de Cox y Snell) explica el 68% de la variabilidad de la variable dependiente, así mismo el coeficiente de McFadden explica el 53.7%; de esta manera la calidad variabilidad explicada es adecuada.
13. Una vez encontrado el mejor modelo como ha ocurrido en esta investigación se procede a validarlo, se observa que al probar el mismo modelo encontrado, en los datos reservados para la validación, se obtiene que éste clasifica de manera correcta el 83.9% de los casos, dejando de forma clara que el modelo es adecuado.

## RECOMENDACIONES

1. Se recomienda la implementación del modelo en el sistema de salud pública, para que así, se puedan focalizar los recursos de manera que lleguen al paciente según sea su condición de salud bucal.
2. A la red de salud pública se recomienda reforzar la atención, educación, promoción y motivación de la salud bucal a los pacientes en general.
3. A la Facultad de Ciencias Naturales y Matemática de la Universidad de El Salvador, que siga fomentando la investigación, para fortalecer al profesional que egresa con el fin de dar respuesta a diferentes temáticas de carácter social, epidemiológicas y biológicas.
4. La realización de esta misma temática considerando la participación de más población sana en la muestra.
5. La implementación de este modelo en el sector de salud público para priorizar el tratamiento que el paciente necesite forma inmediata.

## REFERENCIAS BIBLIOGRÁFICAS

- [1] Agresti A. (2002), *Categorical Data Analysis. Second Edition* ed. New York: Wiley. Charter 7, Pág 267.
- [2] Aguilera del Pino, A. M. (2002). *Modelos de respuesta Discreta*. Granada: Copias Coca, Dep. Legal GR-11554-02.
- [3] Annette J. Dobson, "An introduction to generalized linear models", second edition, CHAPMAN & HALL/CRC A CRC Press Company Boca Raton London New York Washington, D.C. ISBN 1-58488-165-8, QA276 .D589 2001, 3.2 Exponential family of distributions.
- [4] Argentina Luretagoyena, Marcelo Alberto, Diciembre 2012, *Diagnóstico Gingival, Salud dental para todos* (en línea. Disponible en: [www.sdpt.net/PER/diagnosticogingival.html](http://www.sdpt.net/PER/diagnosticogingival.html)
- [5] Barquero Cea Elizabeth Marylu, 2010, *Diagnostico de la condición bucal de la población que asiste a las unidades de salud de santa rosa de lima, agua fría y san alejo del departamento de la unión durante el año 2009*. Trabajo de graduación para optar al grado de: Doctorado en cirugía dental. Universidad de El Salvador, San Salvador.
- [6] Bordoni N, Doño R, Mischi C. *Precon Org. Panamivición de la Salud 1992* Disponible en: [www.sdpt.net//ID/indiceloessilness.html](http://www.sdpt.net//ID/indiceloessilness.html)
- [7] César Pérez López, "Técnicas de muestreo estadístico, teoría, practica y aplicaciones", ISBN 970 - 15 – 0445 -3, pág. 231
- [8] CPO - D Luis Arturo de León Saldaña (Dr. Odontopediatra), 2011, "Curso: Odontología preventiva y social I", [En línea] visto por última vez el 29 de Julio/2014, Disponible en: [http://www.google.com/sv/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0CCAQFjAA&url=http%3A%2F%2Fodontologia42012.files.wordpress.com%2F2011%2F01%2Fdocumentocpo.doc&ei=h\\_iRU9nuHujQsATi7oCIDQ&usg=AFQjCNG3JUWkwbweyGIPllyqi m-3RbG1g&sig2=NCsyY5DIFPTWXHsdcDMxpg&bvm=bv.68445247,d.b2U](http://www.google.com/sv/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0CCAQFjAA&url=http%3A%2F%2Fodontologia42012.files.wordpress.com%2F2011%2F01%2Fdocumentocpo.doc&ei=h_iRU9nuHujQsATi7oCIDQ&usg=AFQjCNG3JUWkwbweyGIPllyqi m-3RbG1g&sig2=NCsyY5DIFPTWXHsdcDMxpg&bvm=bv.68445247,d.b2U)
- [9] CPO – S Marcelo Alberto Iruretagoyena, "Salud dental para todos: CPOD, CEOD, CPOS", [En línea] visto por última vez el 29 de Julio/2014, Disponible en: <http://www.sdpt.net/ID/cpodcposceod.htm>
- [10] Ana Cristina Mafla, David Alexander Barrera, Ginna Mabel Muñoz, 2011, "Mal oclusión y necesidad de tratamiento ortodóntico en adolescentes de Pasto, Colombia", [En línea] visto por última vez el 29 de Julio/2014, Disponible en:

<http://aprendeonline.udea.edu.co/revistas/index.php/odont/article/viewFile/7308/8880>

[11] Dueñas Rodríguez, María Ángeles, 2006, Modelos de respuesta discreta en R y aplicación con datos reales, Universidad de Granada, España.

[12] Fernández Prats, M J, [Fecha de acceso 1 de junio 2013], “Índices epidemiológicos para medir la caries dental”. Disponible en:  
[www.estsocial.sid.cud/publicaciones/indicesepidemiologicosparamedirlacariesdental.pdf](http://www.estsocial.sid.cud/publicaciones/indicesepidemiologicosparamedirlacariesdental.pdf)

[13] Gingivitis-Med line Plus [En línea] Febrero 2012 Fecha de acceso [Mayo 3013] Disponible en:  
<http://www.nlm.nih.gov/medlineplus/spanish/ency/article/001056.html>

[14] Gómez Hernández Sandra Elizabeth, 2013, “Modelación logística multinomial para clasificar los hogares de el salvador por nivel de pobreza”. Disponible en: Biblioteca de la facultad de Ciencias Naturales y Matemáticas de la Universidad de El Salvador, San Salvador.

[15] Lemus Gómez Oscar Hernán, Lemus Gómez Rolando, 2005. “Construcción de un modelo de regresión logístico sobre la oferta laboral a jefes(as) de hogares en El Salvador”. Disponible en: Biblioteca de la facultad de Ciencias Naturales y Matemáticas de la Universidad de El Salvador, San Salvador.

[16] Llodra Calvo JC, Bravo Pérez M, 1997, “Encuesta de Salud Bucodental de la OMS”, cuarta edición, pg 191-192 [Fecha de acceso: 1 de junio 2013] Disponible en:  
[www.org.es/~placa/PRencuestasaludbucodental/02e600099f4106e023/prac12.pdf](http://www.org.es/~placa/PRencuestasaludbucodental/02e600099f4106e023/prac12.pdf)

[17] Morales González, Domingo, 2004, Modelos Lineales Generalizados, Departamento de Estadística y Matemática Aplicada, Universidad Miguel Hernández de Elche, España.

[18] Multinomial Logistic Regresion Model, STAT S44, Lecture 19, Pág.

[19] Official Records of the World Health Organization, N° 2, p. 100

[20] Peña Aguilar René Armando, 2009, “Comparación entre el análisis discriminante y la regresión logística en la clasificación de una colonia de cangrejos herradura (Limulus polyphemus)”. Anteproyecto de Tesis, Disponible en: Biblioteca de la facultad de Ciencias Naturales y Matemáticas de la Universidad de El Salvador, San Salvador.

[21] Silva Ayçaguer LC., 1995, “Excursión a la regresión logística en ciencias de la salud” Madrid: Díaz de Santos.

[22] "The community periodontal Index of treatment needx (CPITN) procedure for population groups and individuality". Dental research unit, medical research of New Zealand, Wellington internacional Dental Journal [1987; 37(4):222-233

[23] Unidad de Salud Bucal del Ministerio de Salud Pública y Asistencia Social, República de El Salvador, 2012, "Estudio Epidemiológico de Caries Dental y Fluorosis en Escolares de 5-6, 7-8, 12 y 15 años de Centros de Enseñanza Pública y Privada de El Salvador 2008".

[24] W. J. Conover, "Practical Nonparametric Statistic" 1971, ISBN 0471 16851 3, JOHN WILEY & SONS, New York Pág 150.

[25] Wikipedia [En línea], 13 mayo 2013 [Fecha de acceso: 1 de junio 2013] "Periodontitis", Zenón Agustín "Nueva clasificación de las enfermedades periodontales" Artículo original. Revista ADM Vol. LVIII. Nº 1. Enero 2001 pág 16-20 Disponible en: [es.wikipedia.org/wiki/periodontitis](http://es.wikipedia.org/wiki/periodontitis)

[26] <http://www.spssfree.com/spss/tablas26.html>

[27] <http://www.redalyc.org/articulo.oa?id=70712293003>

[28] [http://bvs.sld.cu/revistas/end/vol12\\_1\\_01/end07101.htm](http://bvs.sld.cu/revistas/end/vol12_1_01/end07101.htm)

[29] [http://age-tig.es/docs/XII\\_1/038%20%20Vilar%20del%20Hoyo.pdf](http://age-tig.es/docs/XII_1/038%20%20Vilar%20del%20Hoyo.pdf)

[30] [http://oa.upm.es/3363/2/OSCAR\\_ABARCA.pdf](http://oa.upm.es/3363/2/OSCAR_ABARCA.pdf)