# UNIVERSIDAD NACIONAL AGRARIA
# LA MOLINA

## FACULTAD DE ECONOMÍA Y PLANIFICACIÓN

**“FACTORES DETERMINANTES EN EL PESO DEL RECIÉN NACIDO DE MADRES ADOLESCENTES EN LIMA A TRAVÉS DE LA REGRESIÓN LOGÍSTICA MULTINOMIAL”**

**TRABAJO DE SUFICIENCIA PROFESIONAL PARA OPTAR TÍTULO DE INGENIERA ESTADÍSTICA E INFORMÁTICA**

**JUANA RAFAELA TAPIA ALVA**

**LIMA – PERÚ**

**2022**

Ouriginal
by Turnitin

## Document Information

| | |
|---|---|
| **Analyzed document** | TSP-JUANA RAFAELA TAPIA ALVA.pdf (D144084553) |
| **Submitted** | 9/14/2022 7:00:00 PM |
| **Submitted by** | FERNANDO RENE ROSAS VILLENA |
| **Submitter email** | frosas@lamolina.edu.pe |
| **Similarity** | 7% |
| **Analysis address** | frosas.unalm@analysis.urkund.com |

## Sources included in the report

| | | |
|---|---|---|
| W | URL: https://sisbib.unmsm.edu.pe/bibvirtual/tesis/Basic/flores_ml/contenido.htm<br>Fetched: 9/14/2022 7:00:00 PM | 1 |
| SA | **TESIS DE JENNIFER_CEDEÑO_COYA.pdf**<br>Document TESIS DE JENNIFER_CEDEÑO_COYA.pdf (D138531083) | 27 |
| SA | **1598628330_Trabajo_de_titulación_alexandra_dos (3).docx**<br>Document 1598628330_Trabajo_de_titulación_alexandra_dos (3).docx (D78313899) | 2 |
| SA | **6153 paredes_lj.pdf**<br>Document 6153 paredes_lj.pdf (D32867029) | 1 |
| W | URL: https://ri.ues.edu.sv/id/eprint/9807/<br>Fetched: 9/14/2022 7:00:00 PM | 1 |

## Entire Document

1 UNIVERSIDAD NACIONAL AGRARIA LA MOLINA FACULTAD DE ECONOMIA Y PLANIFICACIÓN "FACTORES DETERMINANTES EN EL PESO DEL RECIÉN NACIDO DE MADRES ADOLESCENTES EN LIMA A TRAVÉS DE LA REGRESIÓN LOGÍSTICA MULTINOMIAL" TRABAJO DE SUFICIENCIA PROFESIONAL PARA OPTAR TÍTULO DE INGENIERO ESTADÍSTICO INFORMÁTICA JUANA RAFAELA TAPIA ALVA LIMA – PERÚ 2022 La UNALM es titular de los derechos patrimoniales de la presente investigación (Art. 24 - Reglamento de Propiedad Intelectual)

2 Resumen Dentro de los problemas más determinantes en la salud pública en los países subdesarrollados es la mortalidad neonatal que se presenta en mayor frecuencia en los nacidos con bajo peso (BPN) y la desnutrición crónica. El presente trabajo tiene como objetivo principal identificar los factores que determinan el peso del recién nacido de madres adolescentes que han dado a luz en los establecimientos de salud que se encuentran en Lima a través de la regresión logística multinomial, técnica usada con frecuencia en el área de salud. El modelo considera una variable dependiente, el peso del recién nacido de naturaleza el cual es de naturaleza categórica con 3 clases o categorías: bajo peso; peso normal y macrosómico; y variables independientes cuantitativas y cualitativas. El modelo de regresión logístico determina como factores influyentes la duración del embarazo, la condición del parto (normal, cesárea u otro), el nivel educativo de la madre, el sexo del recién nacido, así como el número de abortos que tuvo la madre. Esto último como resultado del análisis descriptivo, de la selección del modelo, del cálculo de los odds ratios e intervalos de confianza, contraste sobre los parámetros e interpretación del modelo, ajuste global del modelo y la validación del modelo.

# UNIVERSIDAD NACIONAL AGRARIA
# LA MOLINA

## FACULTAD DE ECONOMÍA Y PLANIFICACIÓN

"FACTORES DETERMINANTES EN EL PESO DEL RECIÉN NACIDO DE MADRES ADOLESCENTES EN LIMA A TRAVÉS DE LA REGRESIÓN LOGÍSTICA MULTINOMIAL"

**TRABAJO DE SUFICIENCIA PROFESIONAL PARA OPTAR TÍTULO DE INGENIERA ESTADÍSTICA E INFORMÁTICA**


Presentado por:

**JUANA RAFAELA TAPIA ALVA**


Sustentado y aprobado ante el siguiente Jurado:

| | |
|---|---|
| Dr. Jorge Chue Gallardo<br>PRESIDENTE | Dr. Fernando René Rosas Villena<br>ASESOR |
| Dr. Carlos López de Castilla Vásquez<br>MIEMBRO | Dra. Frida Rosa Coaquira Nina<br>MIEMBRO |

Lima – Perú

2022

# ÍNDICE GENERAL


I. INTRODUCCIÓN ........ 1
II. OBJETIVOS ........ 4
2.1. Objetivo general ........ 4
2.2. Objetivos específicos ........ 4
III. REVISIÓN DE LITERATURA ........ 5
3.1. Modelos de regresión logística ........ 5
3.1.1. Modelo de regresión logística binaria ........ 6
3.1.2. Modelo de regresión logística simple ........ 7
3.1.3. Modelo de regresión logística múltiple ........ 9
3.2. Regresión Logística Multinomial ........ 10
3.2.1. Formulación e interpretación del modelo ........ 11
3.2.2. Métodos de Estimación: Por máxima verosimilitud ........ 13
3.2.3. Contrastes sobre los parámetros del modelo ........ 16
3.2.4. Inferencia en regresión logística multinomial ........ 18
3.2.5. Bondad de ajuste del modelo ........ 21
3.2.6. Método de selección del modelo ........ 26
IV. DESARROLLO DEL TRABAJO ........ 28
4.1. Contribución en la solución del problema ........ 28
4.2. Análisis de la contribución en términos de competencias y habilidades ........ 28
4.3. Nivel de beneficio obtenido por el centro laboral ........ 28
V. RESULTADOS Y DISCUSIÓN ........ 29
5.1. Análisis descriptivo ........ 30
5.2. Metodología de la regresión logística múltiple ........ 33
5.2.1. Paso 1: Formulación y selección del modelo ........ 34
5.2.2. Cálculo de las odds ratios y los intervalos de confianza ........ 40
5.2.3. Contraste sobre los parámetros ........ 43
5.2.4. Ajuste Global del modelo ........ 43
5.2.5. Discusión de los resultados ........ 44
VI. CONCLUSIONES ........ 46
VII. RECOMENDACIONES ........ 47
VIII. REFERENCIAS BIBLIOGRAFICAS ........ 48
IX. ANEXOS ........ 51

# ÍNDICE DE TABLAS

Tabla 1. Descripción de la variable dependiente ................................................................ 31
Tabla 2. Descripción de las variables independientes ......................................................... 32
Tabla 3. Modelo inicial considerando todas las variables .................................................. 35
Tabla 4. Comparaciones de razón de verosimilitud del modelo en el primer momento ..... 36
Tabla 5. Comparaciones razón de verosimilitud del modelo con 1 variable vs modelo con 2 variables ............................................................................................................. 37
Tabla 6. Comparaciones razón de verosimilitud con el modelo con tres variables............. 38
Tabla 7. Comparaciones razón de verosimilitud con el retiro de una variable en el modelo…… ................................................................................................................ 38
Tabla 8. Comparaciones razón de verisimilitud del modelo de 5 a 6 variables .................. 39
Tabla 9. Parámetros del modelo seleccionado..................................................................... 40
Tabla 10. Coeficientes de las variables para el cálculo de los odds ratios .......................... 41
Tabla 11. Intervalos de confianza de los Odds Ratio para el modelo de la categoría del peso macrosómico ............................................................................................................ 41
Tabla 12. Intervalos de confianza de los odss ratio para el modelo de la categoría del peso NORMAL................................................................................................................. 42
Tabla 13. Contraste condicional de razón de verosimilitud ................................................ 43
Tabla 14. Cálculo del ajuste global del modelo................................................................... 44

# ÍNDICE DE FIGURAS

Figura 1. Función logística .......................................................................................... 8

# ÍNDICE DE ANEXOS

ANEXO 1. ESTIMACIÓN DE LOS PARÁMETROS ........................................................ 51
ANEXO 2. CÁLCULO DE LAS ODDS RATIOS Y LOS INTERVALOS DE CONFIANZA ........................................................................................................ 56
ANEXO 3. CONTRASTE SOBRE LOS PARÁMETROS ................................................. 57
ANEXO 4. AJUSTE GLOBAL DEL MODELO ................................................................. 58

# RESUMEN

Dentro de los problemas más determinantes en la salud pública en los países subdesarrollados es la mortalidad neonatal que se presenta en mayor frecuencia en los nacidos con bajo peso (BPN) y la desnutrición crónica. El presente trabajo tiene como objetivo principal identificar los factores que determinan el peso del recién nacido de madres adolescentes que han dado a luz en los establecimientos de salud que se encuentran en Lima a través de la regresión logística multinomial, técnica usada con frecuencia en el área de salud. El modelo considera una variable dependiente, el peso del recién nacido el cual es de naturaleza categórica con 3 clases o categorías: bajo peso; peso normal y macrosómico; y variables independientes cuantitativas y cualitativas. El modelo de regresión logístico determina como factores influyentes la duración del embarazo, la condición del parto (normal, cesárea u otro), el nivel educativo de la madre, el sexo del recién nacido, así como el número de abortos que tuvo la madre.



**Palabras clave:** Regresión Logística Multinomial, Odds Ratio,

# ABSTRACT

Among the most determining problems in public health in underdeveloped countries is neonatal mortality that occurs more frequently in low-weight births (LBW) and chronic malnutrition. The main objective of this study is to identify the factors that determine the weight of the newborn of adolescent mothers who have given birth in health facilities in Lima through multinomial logistic regression, a technique frequently used in the health area. The model considers a dependent variable, the weight of the newborn, which is categorical in nature with 3 classes or categories: low weight; normal and macrosomic weight; and quantitative and qualitative independent variables. The logistic regression model determines as influencing factors the duration of the pregnancy, the condition of the childbirth (normal, cesarean or other), the educational level of the mother, the sex of the newborn, as well as the number of abortions that the mother had.

**Keywords:** Regression Logistic Multinomial, Odds Ratio.

# I. INTRODUCCIÓN

Un recién nacido con bajo peso producto de una madre adolescente representa un alto riesgo y, por ende, requiere ser más extremados en su seguimiento. El desarrollo social y bienestar de la población, y específicamente la reducción de la anemia infantil en niños y niñas, es uno de los ejes de la política general del gobierno. El Plan Bicentenario diseñado por el Ministerio de Justicia y Derechos Humano (2021) promueve una visión estratégica que busca posicionar al Perú como un país desarrollado, democrático y cohesionado socialmente en un horizonte temporal, sustentándose en la Declaración Universal de los Derechos Humanos (DUDH) estableciendo diversos ejes temáticos. Uno de ellos plantea como objetivo concentrar el esfuerzo público en reducir de manera sustantiva y, de ser posible, erradicar la desnutrición crónica y la mortalidad infantil.

Un problema relacionado con la alimentación materna y la desnutrición infantil es la elevada incidencia de la anemia en las mujeres. En el año 2009, el 21,0% de las mujeres peruanas en edad fértil padecía de algún grado de anemia. Cárdenas *et al*. (2011) señala que la desnutrición infantil tiene secuelas irreversibles: baja talla para la edad; limitación de las capacidades físicas, emocionales o intelectuales; entre otras. A largo plazo, la desnutrición reducirá la productividad de la persona y perjudicará directamente el crecimiento económico de la familia. Es por ello, que combatir la desnutrición implica proteger el capital humano del país y prever este recurso para su participación en la economía y el desarrollo del país.

El presente trabajo se realizó en el órgano desconcentrado del Ministerio de Salud, Dirección de Redes Integradas de Salud Lima Este, Dirección Ejecutiva de Monitoreo y Gestión Sanitaria, oficina de Epidemiologia, área de Gestión de la Información. En el que se realizaron actividades de Registro y Análisis de Información Estadística. La Oficina de Intervenciones Sanitarias, parte de la dirección ejecutiva antes indicada, a través del equipo multidisciplinario, buscó mejorar las acciones para promover la nutrición adecuada en infantes y madres gestantes, así como reducir la tasa de desnutrición crónica infantil y la tasa de anemia de niños y mujeres en edad fértil de la jurisdicción.

El peso del recién nacido es determinado inmediatamente después del parto y su cuantificación en gramos permite su clasificación en intervalos desde 500 gramos a más. El bajo peso al nacer es cuando el nacido vivo pesa menos de 2,500 gramos, peso normal entre 2,500 a 3,999 gramos y peso macrosómico cuando pesa más de 4,000 gramos (Ministerio de Salud [MINSA], 2011). El peso del recién nacido en madres adolescentes en el Perú está asociado a un conjunto de factores o variables cuya capacidad explicativa es importante conocer para diseñar e implementar estrategias que permitan mejorar la salud de la madre y el niño.

Las variables explicativas corresponden a la base de datos de los registros de certificados de recién nacido vivo (SNV). Para ello, se tomaron los registros de certificados de nacidos vivos en los establecimientos de salud de Lima Metropolitana que son ingresados en el Sistema de Información desarrollado por el Ministerio de Salud. El Registro del Recién Nacido Vivo facilita la pronta identificación, ya que la misma persona tiene un código único, el cual está vinculado directamente con el Documento Nacional de Identidad [DNI] (MINSA, 2022).

Para determinar la capacidad explicativa que tienen las variables independientes sobre el peso del recién nacido de madres adolescentes se utilizó la regresión logística multinomial, modelo estadístico que se viene usando con mayor frecuencia cuando la medida del efecto o variable dependiente es categórica y se necesita establecer una relación con una o más variables de exposición o un conjunto de variables predictoras (Domínguez et al., 2011). Así, la regresión logística multinomial permite estudiar el efecto conjunto de la exposición a distintos factores. Los odds ratios, se derivan a partir de los coeficientes de la regresión logística (Franco et al., 2018).

Agbozo et al. (2016) evaluaron mediante la regresión logística multinomial la asociación entre factores maternos como la paridad, la edad de la madre y el tratamiento preventivo intermitente de la malaria, el sexo del bebé y el peso anormal al nacer en una muestra de 4,262 registros de partos de embarazos únicos de enero de 2013 a diciembre de 2014 vistos en el hospital municipal de Hohoe, región de Volta en Ghana. Sus hallazgos indicaron que, hubo un mayor riesgo de que las madres <20 años dieran a luz a bebés de bajo peso (RR; 1,46, IC; 1,11-1,93, p=0,007). De forma similar, en otro estudio realizado por Verropoulou & Basten, (2014), el análisis de regresión multinomial mostró que la edad materna menor a 20 años tenía una asociación significativa con todos los tipos peso del recién nacido.

En este estudio, la categoría del peso al nacer es considerada la variable objetivo, a través de la regresión logística multinomial se obtuvieron g-1 modelos, siendo g el número de categorías de la variable objetivo: bajo peso al nacer, normal y macrosómico; se obtuvieron 2 modelos para esta variable.

El objetivo general del presente trabajo consistió en responder a la interrogante ¿cuáles son los factores determinantes en el peso del recién nacido de madres adolescentes en Lima mediante la técnica de la regresión logística multinomial?

# II. OBJETIVOS

## 2.1. Objetivo general

- Identificar los factores determinantes en el peso del recién nacido de madres adolescentes en Lima mediante la técnica de regresión logística multinomial.

## 2.2. Objetivos específicos

- Determinar si la duración del embarazo es una característica determinante en el peso del recién nacido de madres adolescentes en Lima.
- Determinar si la condición de parto es una característica determinante en el peso del recién nacido de madres adolescentes en Lima.
- Determinar si el número de abortos es una característica determinante en el peso del recién nacido de madres adolescentes en Lima.
- Determinar si el nivel educativo de la madre es una característica determinante en el peso del recién nacido de madres adolescentes en Lima.
- Determinar si el sexo del recién nacido es una característica determinante en el peso del recién nacido de madres adolescentes en Lima.

# III. REVISIÓN DE LITERATURA

## 3.1. Modelos de regresión logística

La regresión logística es una de las aplicaciones estadísticas más utilizadas en diversos sectores de la investigación, entre ellos, el sector salud, la misma estudia la relación entre una variable categoría y un conjunto de variables independientes o predictoras (Stoltzfus, 2011).

Este modelo de regresión logística puede ser de carácter estimativo, por ejemplo, para evaluar cuál es la mejor relación de las variables independientes con la variable dependiente, generalmente, en estudios y/o investigaciones de causas de cierta cualidad de evento que ocasionan la probabilidad de que se produzca un suceso determinado. Asimismo, puede ser de carácter predictivo, es decir, por medio de las variables independientes se trata de predecir la variable dependiente, modelo típicamente dicotómico (ordena el valor de la variable objetivo o de respuesta como 1 cuando el evento se presenta y con un valor de 0, cuando está ausente); sin embargo, también puede ser politómica cuando se desea evaluar las diferentes probabilidades de cada uno de los valores que pueda tener la variable objetivo o respuesta (Sperandei, 2014).

El uso de este método sería importante en los casos que la estrategia de regresión lineal no es material, es decir, cuando no es aplicable, pues la variable de respuesta tiene sólo dos cualidades (caso dicotómico), por ejemplo, puede usarse para reconocer los elementos de riesgo y los de prevención de enfermedades en futuras muestras (Sperandei, 2014).

Asimismo, Hosmer y Lemeshow (1989), en su libro sobre regresión logística aplicada, analizan la conexión entre la probabilidad de evento de los resultados de una variable de reacción dicotómica (en su mayor parte llamada la variable dependiente) y las variables lógicas y explicativas categóricas o constantes (conocidas como variables independientes). El pensamiento esencial consiste en establecer una conexión directa entre las variables

explicativas (o algunos cambios de éstos) y un cambio, llamado logit, de la variable de reacción o de respuesta.

Estos autores identifican el par (Y, X), donde Y viene a ser la variable binaria dependiente que toma dos cualidades potenciales llamadas 0 y 1; por otro lado, $X= (X_1,...,X_{k)}$ las variables independientes a estudiar, las cuales predecirían el valor de Y. es así que, la finalidad del modelo es conocer $P[Y=1/X_1,...,X_k]$, donde P indica la probabilidad, de tal manera que:

$$P[Y=0/X1,...,Xk] = 1-P[Y=1/X1,...,Xk]) \qquad (1)$$

A partir de ello, se el modelo que se estructura es el siguiente:

$$P[Y=1/X_1,...,X_k] = p(X_1,...,X_k ; ß) \qquad (2)$$

Donde $p(X_1,...,X_k ; ß)$: $R^k$ [0,1] y es denominada función de enlace o función de probabilidad cuyo valor depende de un vector límite $ß = (ß_1,...,ß_k)'$

A continuación, se presentan los tres principales modelos de regresión logística, según Stoltzfus (2011).

### 3.1.1. Modelo de regresión logística binaria

Estableciendo como $p (X_1,...,X_k; ß) = G(ß_1X_1 + ... + ß_kX_k)$, donde:

$$G(x) = \frac{e^x}{1+e^x} \qquad (3)$$

viene a ser la función de densidades acumuladas que es la función logística, siendo el modelo que regularmente se conoce:

$$\log\left(\frac{p(x_1,\ldots,x_k;\beta)}{1-p(x_1,\ldots,x_k;\beta)}\right)=\beta_0+\beta_1 x_1+\ldots+\beta_k x_k \tag{4}$$

también llamado logit (Stoltzfus, 2011).

En el momento en que la variable cualitativa toma el valor 1 en la ecuación:

$$\frac{p\left[Y=1/X_1,\ldots,X_k\right]}{p\left[Y=0/X_1,\ldots,X_k\right]}=\frac{p(x_1,\ldots,x_k;\beta)}{1-p(x_1,\ldots,x_k;\beta)} \tag{5}$$

Generalmente se denomina factor de riesgo en el área de la salud, mientras que la variable Y muestra la presencia de una enfermedad específica que se está analizando y que toma el valor de 0 cuando está ausente.

### 3.1.2. Modelo de regresión logística simple

Para realizar el modelo es importante tener datos numéricos, las cuales se obtienen pensando en la probabilidad de ocurrencia de un evento dado P(Y), correspondiente a la dependencia de que esta probabilidad no ocurra 1-P(Y).

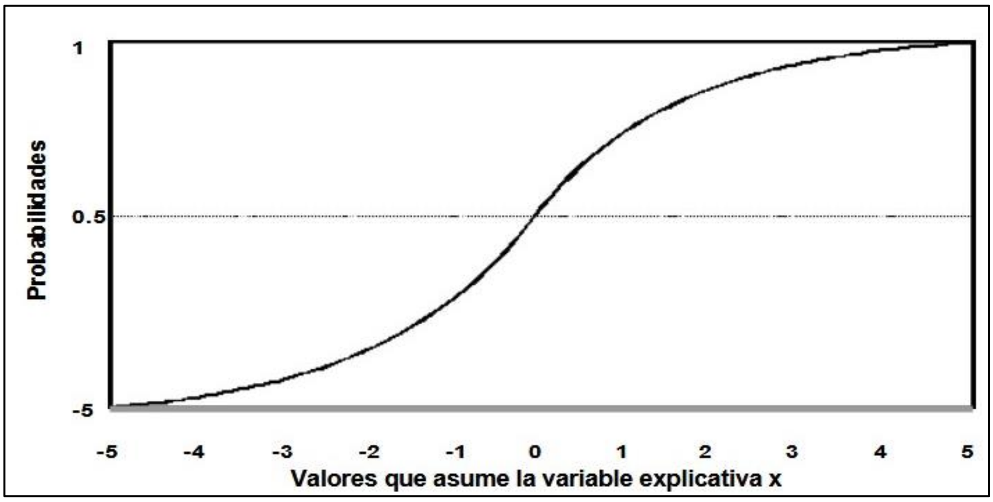
Se explica la ecuación teniendo en la primera parte a P y, en la segunda parte, la conexión útil con la mediación de las variables independientes que son las variables de interés. La probabilidad es un número cuyo valor varía en el rango de 0 y 1, permitiendo predicciones consistentes o estables y cuyos resultados no son difíciles de interpretar en razón de probabilidades "odds ratio" (OR) odds (Stoltzfus, 2011).
Sea la función:

$$f(x)=\frac{1}{1+e^{-x}} \tag{6}$$

que se grafica como lo muestra la Figura 1 y es denominada función logística:

**Figura 1.**

*Función logística*




*Nota*. Adaptado de Flores (2002).

Para solo una variable independiente X, el modelo de regresión logística se muestra de la siguiente forma:

$$p_i = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{(\beta_0 + \beta_1 x)}} \tag{7}$$

el modelo Logit será:

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 X \tag{8}$$

O facilitando la notación:

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 X \tag{9}$$

Donde log es logaritmo en base diez, $\beta_0$ y $\beta_1$ son constantes y X es una variable independiente que puede ser continua o discreta. El campo de variación de log ($p_i/(1-p_i)$) es todo el campo real, mientras que para p el campo es sólo de 0 a 1 y para $p_i/(1-p_i)$ de 0 a $\infty$. Se puede afirmar que lo principal en este modelo es que los coeficientes son efectivamente interpretables con respecto a la independencia o la relación entre los variables; por lo cual, no es necesario limitar los coeficientes ya que sólo complicarían su evaluación (Stoltzfus, 2011).

### 3.1.3. Modelo de regresión logística múltiple

Es una extensión del modelo simple que relaciona la probabilidad de un evento independiente identificado por el vector X'= ($x_1$,...,$x_k$) con la probabilidad condicional P(Y=1/X) con relación de k cantidad de variables independientes, las cuales pueden ser cualitativas, cuantitativas o combinadas, dependiendo del tipo diseño del estudio (Stoltzfus, 2011).

El modelo logístico múltiple es:

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 x_1 + ..... + \beta_k x_k \tag{10}$$

O también:

$$p_i = \frac{e^{\beta_0 + \beta_1 x_1 + ..... + \beta_k x_k}}{1 + e^{\beta_0 + \beta_1 x_1 + ..... + \beta_k x_k}} \tag{11}$$

En la regresión logística la medida de afiliación o asociación más utilizada es "odds ratio" (OR), ya que "e" es el fundamento de los logaritmos neperianos y elevados a un coeficiente de regresión logística del factor, cuando sea mayor que 1 indica que mayor es el factor de riesgo.

En el caso de que el modelo de regresión logística sea significativo y una de las variables independientes sea dicotómica, con valores de 0 y 1, el número "e" elevado al coeficiente de regresión logística calcula el OR, llamado factor de riesgo que sugiere un incremento unitario de la variable independiente. Si se trata con una variable cuantitativa, "e" elevado a $ß_1$ viene a ser la cantidad de veces que aumenta la probabilidad de experimentar una enfermedad por cada incremento unitario de la variable independiente o, en otras palabras, cuantas veces es más frecuente que un individuo que presenta efectos secundarios relacionados con la infección está destinado a experimentar los efectos nocivos de la enfermedad (Sperandei, 2014).

Como refiere Franco *et al*. (2018) la regresión logística múltiple es uno de los instrumentos estadísticos más expresivos y versátiles disponibles para el análisis de datos en las esferas clínica y epidemiológica, así como en la salud pública. Por ello, para el cumplimiento de los objetivos del presente trabajo de investigación, el modelo de regresión logística múltiple es el método a utilizar. Por tal motivo, se profundizará acerca de dicho modelo en la siguiente sección.

## 3.2. Regresión Logística Multinomial

Los modelos de modelo de regresión logística son modelos estadísticos en los que se trata de explicar la conexión entre una variable dependiente cualitativa, dicotómica (regresión logística binaria o binomial), con más de dos condiciones (regresión logística multinomial) y entre variables aclarativas independientes, las cuales pueden ser cualitativas o cuantitativos (Pando y San Martín, 2004).

Las covariables cualitativas dicotómicas deben codificarse como 0 para una de las clases o para su inasistencia y como 1 para la otra clasificación o para su presencia (esta codificación es significativa en vista de que alguna otra codificación podría provocar cambios en la comprensión del modelo). Sin embargo, si la covariable cualitativa tiene varias clases, se realiza un cambio para ingresar en el modelo. Este cambio consiste en hacer dicotómicas a unas variables cualitativas, llamadas factores **dummies**, de modo que una de las variables se tomaría como una clase de perspectiva y cada una de las variables realizadas entraría en el modelo por separado. Como regla general, si la covariable cualitativa tiene c categoría, deben hacerse c-1 covariables ficticias (Silva & Barroso, 2004).

La regresión logística multinomial se utiliza en modelos que tienen una variable dependiente nominal politómica (con múltiples características) y es una expansión multivariada de la regresión logística binaria clásica. Las variables independientes pueden ser consistentes (covariables) o categóricas (factores) (Silva & Barroso, 2004).

Estos modelos se evalúan eligiendo una clasificación como referencia para la variable de reacción y, a la vez, se van formando algunas ecuaciones, una para cada una de las clases relativas a la de referencia.

### 3.2.1. Formulación e interpretación del modelo

Según Debella-Gilo *et al*. (2007), en el modelo de regresión logística las clases de la variable dependiente se codifican como 0 y 1, lo que hace que la media de la variable se refiera a la proporción de casos que ocurren en una de sus dos clases (en el caso binomial) o en una de sus numerosas clasificaciones (en el caso multinomial). El valor estimado de la probabilidad por el modelo indicado por la clase puede descifrarse como la probabilidad de que un caso se presente en esa clasificación. Un modelo lineal no se acopla adecuadamente a las variables binomiales, ya que los valores predichos de la variable dependiente con este modelo (ajustado por la ecuación de una línea) pueden tomar valores de probabilidad inimaginables superiores a 1 o por debajo de 0, a pesar de que las valores observados están en el rango de 0 y 1. El mejor modelo que linealiza la conexión entre la variable dependiente e independiente es el modelo logit, el cual se desarrolla a través de la regresión logística.

En una variable respuesta dicotómica (Y=0; Y=1), si se conoce la probabilidad de ser parte de una categoría del suceso (Y=0), se puede conocer la probabilidad de ser parte de la otra categoría (Y=1), por ejemplo: $\mathrm{P}(Y = 1) = [1 - \mathrm{P}(Y = 0)]$. Se puede tratar de emplear el modelo lineal de probabilidad, mencionado como:

$$\mathrm{P}(Y = 1) = \beta 0 + \beta 1 X.$$

Dónde:

- $\mathrm{P}(Y = 1)$. Es la probabilidad asociada a la variable predicha (Y)
- X. Es la variable independiente
- $\beta$. Son los parámetros de la población a estimar.

Este modelo de probabilidad presenta la dificultad de ser no lineal, con valores predichos que pueden ser menores a cero o mayores a 1.

La regresión logística multinomial se utiliza en modelos que tienen una variable dependiente nominal politómica (con múltiples características) y es una expansión multivariada de la regresión logística binaria clásica. Las variables independientes pueden ser consistentes (covariables) o categóricas (factores) (Silva & Barroso, 2004). La misma puede utilizarse para examinar problemas en los que hay tres o más categorías desordenadas en la variable dependiente (Britt & Weisburd, 2009).

El modelo de regresión logística multinomial se suele formular con las siguientes funciones, que brindan la probabilidad de pertenencia a las primeras k-1 clases y la probabilidad para la última clase k se obtiene por diferencia a la unidad (Valencia & Bonifaz, 2018). Dichas fórmulas son:

$$\pi_{in} = \frac{e^{Z_{in}}}{1+e^{Z_{i1}}+e^{Z_{i2}}+\ldots+e^{Z_{ik-1}}} \tag{12}$$

$$Z_{in} = \beta_{n0} + \beta_{n1}X_{i1} + \beta_{n2}X_{i2} + \cdots \beta_{nj}X_{ij} \tag{13}$$

Donde

$\pi_{in}$ es la probabilidad de pertenencia del caso i a la clase n

$Z_{in}$ es el valor de la variable dependiente Z, correspondiente a la clase en el caso i;

$\beta_{nj}$ es el coeficiente de la variable independiente j para la clase n;

$X_{ij}$ es el valor de la variable independiente j en el caso i.

En la regresión logística multinomial se considera a menudo un análisis atractivo porque no asume la normalidad, la linealidad ni la homocedasticidad. Sin embargo, una alternativa a la regresión logística multinomial es el análisis de función discriminante, pero que requiere que se cumplan estos supuestos. Así, la regresión logística multinomial se utiliza con más frecuencia que el análisis de función discriminante porque el análisis no tiene esos supuestos (Kwak & Clayton, 2002).

Fernández & Fernández, (2004) refieren que, gracias a los modernos métodos de cálculo, los modelos de regresión logística multinomial pueden encontrarse en programas estadísticos populares como SPSS, SAS, asimismo en la actualidad posee muchos adeptos debido a la mejor interpretabilidad de sus resultados (Fernández & Fernández, 2004).

Estos modelos se evalúan eligiendo una clasificación como referencia para la variable de reacción y, a la vez, se van formando algunas ecuaciones, una para cada una de las clases relativas a la de referencia. Por ello, un paso importante en una regresión multinomial es la definición de dicha categoría de referencia. Esto es necesario porque tenemos que decidir qué categoría queremos utilizar como línea de base para todas las comparaciones. Es una decisión arbitraria sobre qué categoría se designa como categoría de referencia, pero en la medida en que podamos hacer una elección que tenga alguna relación teórica o que haga que la interpretación de los resultados sea más sencilla, esa sería la elección preferida (Britt & Weisburd, 2009).

### 3.2.2. Métodos de Estimación: Por máxima verosimilitud

La evaluación de la mayor probabilidad o de máxima verosimilitud, es utilizada para estimar tanto los coeficientes del modelo como de sus errores estándar; se trata de estimaciones que aumentan la probabilidad de obtener los valores de la variable dependiente dada por la información de la muestra. Para este caso, los cálculos para la evaluación de los coeficientes de regresión logística multinomial no son inmediatos, por lo cual deben completarse técnicas iterativas como la estrategia de Newton-Raphson.

Para la regresión logística, la estimación por mínimos cuadrados no es capaz de producir estimadores insesgados de varianza mínima para los parámetros reales. En su lugar, se utiliza la estimación de máxima verosimilitud para resolver los parámetros que mejor se ajustan a los datos (Czepiel, 2002).

Bajo el supuesto de que se contraste una muestra arbitraria de tamaño N bajo Q distintas combinaciones de valores de las variables independientes $X_1,...,X_n$; cada combinación de valores de las variables explicativas se representa por $x_q = (x_{q0}, x_{q1}, \ldots, x_{qn})'$ con $x_{q0} = 1$ $\forall$ $q$=1,..., En cada una de estas, se obtiene una muestra al azar de $dq$ observaciones independientes de la variable dependiente politómica Y, de entre las cuales, $yj/q$ vendría a ser la cantidad de percepciones que entran en la clase dependiente $Y_j$ $\forall$ j = 1, ... . , k. Por lo tanto, se verifica la ecuación (Aguilera del Pino, 2002):

$$\sum_{j=1}^{k} y_{j/q} = d_q \text{ y } \sum_{q=1}^{Q} d_q = N. \tag{14}$$

Los vectores $(y_{1/q}, \dots, x_{k/q})' \; \forall$ q =1,..., Q tienen una distribución de probabilidad multinomial libre (independiente), $M(d_q; p_{1/q}, \dots, p_{k/q})$ , donde $p_{j/q} = P[Y = Y_j / X = x_q]$ y comprobando que:

$$\sum_{j=1}^{k} y_{j/q} = 1 \tag{15}$$

De modo que, la siguiente ecuación expresa la función de máxima verosimilitud:

$$V = \prod_{q=1}^{q} \left( \frac{d_q!}{\prod_{j=1}^{k} \left( y_{j/q} \right)!} \prod_{j=1}^{k} P_{j/q}^{y_{j/q}} \right) \tag{16}$$

En efecto, la pieza de la log-probabilidad es:

$$K = \sum_{q=1}^{Q} \sum_{j=1}^{k} y_{j/q} \ln(P_{j/q}) \tag{17}$$

Generalmente, la función alterna:

$$\Lambda = -2\ln(V) \tag{18}$$

es la que se utiliza en lugar de la verosimilitud. Por lo cual, ampliar la verosimilitud sería equivalente a minimizar la función alterna. Considerando la ecuación del modelo logit multinomial y reemplazando la ecuación anterior se obtiene la ecuación adjunta del bit de la log-probabilidad (Aguilera del Pino, 2002):

$$K = \sum_{q=1}^{Q} \sum_{j=1}^{k} y_{j/q} \left( \sum_{s=0}^{n} b_{sj} x_{qs} \right) - \sum_{q=1}^{Q} \left( \sum_{j=1}^{k} y_{j/q} \right) ln \left( \sum_{j=1}^{k} exp \left( \sum_{s=0}^{n} b_{sj} x_{qs} \right) \right)$$

$$= \sum_{q=1}^{Q}\sum_{j=1}^{k} y_{j/q}\left(\sum_{s=0}^{n} b_{sj}x_{qs}\right) - \sum_{q=1}^{q} n_q ln\left(\sum_{j=1}^{k} exp\left(\sum_{s=0}^{n} b_{sj}x_{qs}\right)\right) \quad (19)$$

Haciendo la derivación de los parámetros respectivos, se obtiene:

$$\frac{\Delta K}{b_{sj}} = \sum_{q=1}^{Q} y_{jq}x_{qs} - \sum_{q=1}^{q} n_q x_{qs}\frac{\exp\left(\sum_{s=0}^{n} b_{sj}x_{qs}\right)}{\sum_{j=1}^{k}\exp\left(\sum_{s=0}^{n} b_{sj}x_{qs}\right)} \quad (20)$$

En consecuencia, se consiguen las ecuaciones de probabilidad con estructura matriacal: $X_{((n+1)xQ)'}\ y_{j(Qx1)} = X'_{((n+1)xQ)}\ \hat{m}_{j(Qx1)}\ \forall\ j=1,\ldots,-1$, siendo $y_j = (y_{j/1}, \ldots, y_{j/q})'$ $y$ $\hat{m}_j = (\hat{m}_{j/1}, \ldots, \hat{m}_{j/Q})'$ con $\hat{m}_{j/q}$ la recurrencia normal de la reacción $Y_j$ en la mezcla $xq$ de los valores notados de los factores indicadores, evaluados bajo el modelo y caracterizados como $\hat{m}_{j/q} = d_q\hat{p}_{j/q}$. Para hallar los evaluadores de máxima probabilidad o verosimilitud, se hace uso de la estrategia iterativa de Newton-Raphson, de tal manera que se consigue el evaluador de los parámetros, el cual es una cuadrícula de aspecto (n+1)x(k-1) formada por $b\hat{}= (b\hat{}'_1, b\hat{}'_2, \ldots, b\hat{}'_{k-1})'$ , siendo $b\hat{}'_j$ el evaluador de máxima probabilidad o verosimilitud del vector límite relacionado con la clasificación de la variable de reacción $Y_j$ ((Aguilera del Pino, 2002).

Con esto, lo que se obtiene es la matriz de covarianza, que viene a ser la inversa de la matriz de datos de Fisher, el marco de covarianza de cada vector límite $b\hat{}'$ . Por esta razón, hay que determinar las segundas derivadas de K con r≠s (Aguilera del Pino, 2002).

$$\frac{\Delta^2 K}{\Delta b_{rj}\Delta b_{sj}} = \sum_{q=1}^{Q} n_q x_{qs}x_{qr}\frac{exp\left(\sum_{s=0}^{n} b_{sj}x_{qs}\right)\left[\sum_{j=1}^{k} exp\left(\sum_{s=0}^{n} b_{sj}\, x_{qs}\right) - exp\left(\sum_{s=0}^{n} b_{sj}\, x_{qs}\right)\right]}{\left[\sum_{j=1}^{k} exp\left(\sum_{s=0}^{n} b_{sj}\, x_{qs}\right)\right]^2} \quad (21)$$

De este modo, la matriz de covarianza viene dado por  (Aguilera del Pino, 2002).

$$cov(\hat{b}_j) = \left[-E\left(\frac{\Delta^2 K}{\Delta b_{rj}\Delta b_{sj}}\right)\right]^{-1} = \left[X' diag\left[d_q p_{j/q}\left(1 - p_{j/q}\right)\right]X\right]^{-1} \quad (22)$$

Las matrices de covarianza cruzadas entre cada conjunto de evaluadores $\hat{b}_j$ y $\hat{b}_i$ ($i \neq j$).
Por ello, se determinan las siguientes segundas derivadas de K con r≠s y j≠i:

$$\frac{\Delta^2 K}{\Delta b_{ri} \Delta b_{sj}} = \sum_{q=1}^{Q} n_q x_{qs} x_{qr} \frac{-exp\left(\sum_{s=0}^{n} b_{sj} x_{qs}\right) exp\left(\sum_{s=0}^{n} b_{si}\, x_{qs}\right)}{\left[\sum_{j=1}^{k} exp\left(\sum_{s=0}^{n} b_{sj}\, x_{qs}\right)\right]^2} \tag{23}$$

Se obtiene, la ecuación adjunta para la matriz de covarianza

$$cov(\hat{b}_j, \hat{b}_i) = \left[-E\left(\frac{\Delta^2 K}{\Delta b_{ri} \Delta b_{sj}}\right)\right]^{-1} = \left[-X' diag\left[d_q p_{j/q} p_{i/q}\right] X\right]^{-1} \tag{24}$$

Por último, se tiene que la matriz de covarianzas del estimador es:

$$cov(\hat{b}) = \begin{pmatrix} cov(\hat{b}_1) & cov(\hat{b}_1, \hat{b}_2) & \dots & cov(\hat{b}_1, \hat{b}_{k-1}) \\ cov(\hat{b}_1, \hat{b}_2) & cov(\hat{b}_2) & \dots & cov(\hat{b}_j, \hat{b}_{k-1}) \\ cov(\hat{b}_1, \hat{b}_{k-1}) & cov(\hat{b}_2, \hat{b}_{k-1}) & \dots & cov(\hat{b}_{k-1}) \end{pmatrix} \tag{25}$$

### 3.2.3. Contrastes sobre los parámetros del modelo

Silva & Barroso (2004) refieren que, tras construir el modelo, ajustarlo y una vez obtenidas las evaluaciones; la siguiente etapa consiste en comprobar el significado factual de cada uno de los coeficientes de regresión en el modelo. Lo antes mencionado se logrará utilizando, fundamentalmente, dos técnicas para los modelos de regresión logística multinomial: el contraste de Wald y el contraste condicional de razón de verosimilitud. Entonces, lo que se busca es diferenciar si un subconjunto de los parámetros del modelo de regresión logística multinomial, mostrado por ß =( ß 1, ..., ß r)′, es nulo. Para ello, la hipótesis será:

- $H_0$: ß =0
- $H_1$: ß ≠0

Como ya se mencionó, la verificación o comprobación de la significación estadística de cada uno de los coeficientes de regresión del modelo se hace a través de 2 métodos:

### *3.2.3.1. Contraste de Wald.*

Evalúa la significancia de los coeficientes de cada una de las variables independientes. Esto permite determinar si la variable independiente hace un aporte estadísticamente significativo a la explicación de la variable respuesta sin tener que usar la razón de verosimilitud (Aguilera del Pino, 2002).

Autores como Silva & Barroso (2004) mencionan que el valor de su coeficiente de regresión se calcula como el cociente entre el valor del coeficiente y su correspondiente valor del error estándar; es decir, se va a contrastar con la siguiente hipótesis:

$H_0 : ß_i = 0$

$H_1 : ß_i \neq 0$

El estadístico de *Wald*, será:

$$Z_{Wald} = \frac{\hat{\beta}_i}{\hat{\sigma}^2(\hat{\beta}_i)} \tag{26}$$

El estadístico de *Wald*, tiene una distribución chi-cuadrado asintótica con un grado de libertad cuando la muestra es grande, llevando a rechazar la hipótesis nula con un nivel de confianza 1 – α si: $Z_{Wald} \geq X_{1}^{2}{}_{,\alpha}$ , lo que indica que el coeficiente de regresión es estadísticamente significativo, pues es diferente de cero y merece estar considerado en el modelo. Sin embargo, el estadístico de Wald no se recomienda cuando se trabaja con variables de diseño y tampoco con modelos con errores estándar grandes, ya que puede mostrar que el coeficiente no es significativo, cuando en realidad sí lo es. Para estos casos, lo recomendable es usar el test de razón de verosimilitudes (Silva & Barroso, 2004).

#### *3.2.3.2. Contrastes Condicionales de Razón de Verosimilitud.*

Examina cada modelo que surge al eliminar de forma aislada cada una de las variables independientes frente al modelo completo. La no significancia indicaría que el modelo sin la variable independiente no empeora respecto al modelo completo, es decir, no se ve afectado por su presencia o su ausencia. Por ello, según la estrategia de obtención del modelo más reducido (principio de parsimonia), dicha variable independiente debe ser eliminada del modelo, pues no aporta nada al mismo (Silva & Barroso, 2004).

### 3.2.4. Inferencia en regresión logística multinomial

Aguilera del Pino (2002) afirma que, la razón principal para reproducir un modelo medible utilizando la información de un ejemplo es extrapolar los resultados de una muestra a la población general, que es la razón por la que se evalúan los parámetros del modelo de regresión logística multinomial y se utiliza para hacer la conjetura.

#### *3.2.4.1. Intervalo de Confianza.*

A la luz de la normalidad asintótica de los evaluadores de probabilidad más extremos (o de máxima verosimilitud), se pueden desarrollar intervalos de confianza asintóticos para cada uno de los parámetros del modelo, utilizando la distribución de Gauss e intervalos de confianza para las proporciones de las odds ratio (OR), haciendo las transformaciones respectivas (Sperandei, 2014).

Los odds ratio son considerados la razón de probabilidades, la probabilidad de ocurrencia de un evento o suceso P(y) en función de la dependencia que dicha probabilidad no ocurra 1-P(y).

Cerda *et al*. (2013), define odds ratio (OR) como la medida de la razón entre dos odds, es así que un odds es la probabilidad de que un evento de interés ocurra o que una exposición se presente. Este concepto es generalmente usado en salud para informar los resultados de una investigación dada.

Aguilera del Pino (2002) señala que, una vez ajustado el modelo, se debe dar una interpretación y, para ello, se calculan las odds ratios. Para obtener las odds ratios de los parámetros del modelo con del software R se usará la función "exp", ya que están definidas como la exponencial de los parámetros:

Exp(x)

Donde:

x: es un número o vector. En este caso, será el vector que incluye los coeficientes del modelo.

- ***Intervalos de Confianza para los Parámetros.*** Para cada parámetro del modelo de regresión logística multinomial, $\beta_{sj}$ con j=1,...,k, se elaborará un intervalo de confianza con nivel 1- α. La dispersión asintótica de $\hat{b}_{sj}$ es $N$ ($\beta_{sj}$, $\hat{\sigma}^2(\hat{b}_{sj})$), donde $\hat{\sigma}(\hat{b}_{sj})$ es el valor relativo al error estándar del evaluador del parámetro $\beta_{sj}$ (Aguilera del Pino, 2002).

  La ecuación que surge sería:

$$P[-Z_{\alpha/2} \leq \frac{\hat{b}_{sj}-b_{sj}}{\hat{\sigma}(\hat{b}_{sj})} \leq Z_{\alpha/2}] = 1-\alpha \tag{27}$$

  Siendo el intervalo de confianza supuesto para $\beta_{sj}$ en el nivel 1–b$\alpha$:

$$IC(\beta_{sj}) = (\hat{b}_{sj} \pm Z_{\frac{\alpha}{2}}\hat{\sigma}(\hat{b}_{sj}) \tag{28}$$

- ***Intervalos de Confianza para las Odds Ratio.*** Para este caso, según Aguilera del Pino (2002) se tiene que los cocientes de las ventajas son:

$$\theta_j(\Delta X_r = 1/X_s = x_s, s \neq r) = \exp(b_{sj}) \ \forall \ \mathrm{r} = 1, \dots, \mathrm{n}; \ \forall \ \mathrm{j} = 1, \dots, \mathrm{k}-1) \tag{29}$$

  En consecuencia, el intervalo de confianza para los cocientes de ventajas se determina tomando exponenciales en el tramo de confianza obtenido anteriormente para cada uno de los parámetros $\beta_{sj}$. Así, el intervalo de confianza para ($\beta_{sj}$) en el nivel de confianza 1-$\alpha$, sería:

$$IC(\exp(\beta_{rj})) = exp(\hat{b}_{sj} \pm Z_{\frac{\alpha}{2}}\hat{\sigma}(\hat{b}_{sj})) \quad (30)$$

#### 3.2.4.2. *Validación del Modelo.*

Después de haber utilizado la prueba de chi-cuadrado de Pearson $X^2$, debe contemplarse la bondad del ajuste de cada percepción para demostrar si una percepción es convincente. Para esto, una de las estrategias para el estudio de este modelo es la investigación de los residuales que contemplan el número de éxitos observados en cada mezcla de los valores de las variables predictoras con su valor cambiado por el modelo. Los tipos de residuos más reconocidos en vista de las medidas $X^2$ y $G^2$ se caracterizan para cada combinación de $x_q$ lados superiores de los variables explicativas (Aguilera del Pino, 2002).

#### 3.2.4.3. *Residuos de Pearson o Residuos Estandarizados.*

Los residuos de Pearson tienen la siguiente ecuación (Fagerland *et al.*, 2018).:

$$r_{j/q} = \frac{y_{j/q} - d_q\hat{p}_{j/q}}{[d_q\hat{p}_{j/q}]^{\frac{1}{2}}} \quad (31)$$

A partir de ello, se puede caracterizar la medida chi-cuadrado de Pearson como:

$$X^2 = \sum_{q=1}^{Q}\sum_{j=1}^{k} r_{j/q}^2 \quad (32)$$

Para diferenciar el significado medible de los residuos se representa lo siguiente:

- $H0$: $r$j/q =0.
- $H1$: $r$j/q ≠0

Teniendo a la hipótesis nula $r_{j/q}$ , esta presenta una distribución normal asintótica con media cero y fluctuación evaluada $\hat{\sigma}^2(\hat{b}_{sj}) < 1$, en otras palabras, los residuos tienen una variabilidad menor que una variable irregular estándar; sin embargo, se tratan típicamente como normales estándar, siendo vistos como enormes cuando sus cualidades absolutas son más prominentes que dos (ausencia de ajuste). Para evitar esto, los residuos de Pearson ajustados se caracterizan por tener dispersiones típicas asintóticas estándar (Aguilera del Pino, 2002).

$$r_{j/q}^{s} = \frac{r_q}{\hat{\sigma}(r_{j/q})} \tag{33}$$

Se puede, igualmente, tomar el cuadrado de $r^s_q$ que tiene distribución chi-cuadrado con un nivel de oportunidad. Así que la hipótesis nula se descarta con un nivel de importancia α cuando:

$$|r_{j/q}^{s}| \geq Z_{\alpha/2} \tag{34}$$

### 3.2.5. Bondad de ajuste del modelo

#### *3.2.5.1. Contrastes de Bondad de Ajuste del Modelo.*

Muy posiblemente el signo de significación más temprano para ver el valor en el ajuste del modelo logístico multinomial es el logaritmo doble de la medida de probabilidad o verosimilitud, que se examinará más adelante, cuya dispersión es χ2 (Fagerland *et al*., 2018). Cabe resaltar que todo el apartado 3.2.5. se basará en dichos autores.

Sea $yj/q$ la cantidad de percepciones que entran en la clasificación de reacción Yj ∀ j = 1, . . ., k. y sean las dq percepciones relacionadas con la q-ésima mezcla de valores de los variables explicativas.

Se entiende por $\hat{m}_{j/q}$ la recurrencia normal de la reacción Yj en la mezcla xq de valores notados de los variables predictoras, evaluada según el modelo y caracterizada como $\hat{m}_{j/q} = d_q p\hat{1}/q\hat{p}1/q$.

Para probar la bondad de ajuste del modelo cuando la cantidad de percepciones en cada mezcla de valores de las variables explicativas es enorme, se utiliza la medida de chi-cuadrado de Pearson y la medida de proporción de probabilidad de Wilks.

El test de bondad de ajuste del modelo de regresión logística multinomial pone a prueba la hipótesis adjunta (Aguilera del Pino, 2002):

$$H_0: p_{j/q} = \frac{exp\left(\sum_{s=0}^{n} b_{sj}x_{qs}\right)}{1 + exp\left(\sum_{s=0}^{n} b_{sj}x_{qs}\right)} \;\; \forall \text{ q} = 1,..,\text{Q}; \; \forall \text{ j} = 1, \dots, \text{k} \tag{35}$$

$$H_1: p_{j/q} \neq \frac{exp\left(\sum_{s=0}^{n} b_{sj}x_{qs}\right)}{1 + exp\left(\sum_{s=0}^{n} b_{sj}x_{qs}\right)} \;\; para\ algún\ q\ y\ j \tag{36}$$

**a.** ***Test Chi-cuadrado de Pearson*****.** La medida chi-cuadrado de Pearson de la integridad de ajuste a un modelo de regresión lógica multinomial M, de la estructura anterior viene dada por:

$$X^2(M) \sum_{q=1}^{Q} \sum_{j=1}^{k} \frac{\left(y_{j/q} d_q \hat{p}_{j/q}\right)^2}{d_q \hat{p}_{j/q}} \tag{37}$$

siendo $\hat{p}_{j/q}$ el indicador de probabilidad más extremo de $p_{j/q}$.

Esta medida tiene una difusión asintótica de chi-cuadrado con niveles de oportunidad adquiridos como la distinción entre la cantidad de parámetros $p_{j/q}$ y la cantidad de parámetros independientes en el modelo, Q-(n+1) x (k-1). Es decir, $X^2_{Q-(n+1)\,x\,(k-1)}$ , $si$ dq→ ∞

Por ello, la hipótesis nula se descarta en el nivel de importancia α cuando $X^2(M)_{Obs} \geq X^2_{Q-(n-1)x(k-1);\alpha}$ . O, se puede caracterizar el valor p de la diferenciación como la probabilidad agregada a un lado del valor observado: $p$ - $valor$ = $[X^2(M) \geq X^2(M)_{Obs}]$, la hipótesis nula se descarta cuando p-valor $\leq \alpha$ (Aguilera del Pino, 2002).

**b.** ***Test Chi-cuadrado de Razón de Verosimilitudes. Estadístico de Wilks. Devianza.*** La medida de la proporción de probabilidad de Wilks para el ensayo de bondad de ajuste del modelo de regresión logística multinomial M se obtiene como menos dos veces el logaritmo del cociente entre el supremo de la probabilidad (o verosimilitud) bajo la hipótesis nula y el supremo de la probabilidad en la población. A partir de esta expresión, por trabajo, se adquiere la declaración de esta medida que viene dada por (Aguilera del Pino, 2002).:

$$G^2(M) = 2\left[\sum_{q=1}^{Q}\sum_{j=1}^{k} y_{j/q} ln\left(\frac{y_{j/q}}{\widehat{m}_{j/q}}\right)\right] \tag{38}$$

Esta medida presenta distribución asintótica chi-cuadrado donde la distinción entre la dimensión del espacio paramétrico y el espacio de la hipótesis nula representan los grados de libertad. En un modelo de regresión logística multinomial estos se obtienes a partir de la diferencia la cantidad de límites parametros $p_{1/q}$ y la cantidad de límites $\beta_{sj}$ bajo el modelo, es decir, Q- (n+1) x (k 1) grados de libertar $G^2(M)^d \longrightarrow X^2{}_{Q\text{-}(n+1)\ x\ (k\text{-}1)}$, si dq $\rightarrow \infty$. A lo que un $p$ - $valor = [G^2(M) \geq G^2(M)_{Obs}] \leq \alpha$ o con valor $\alpha$ de nivel de significación cuando: $G^2(M)_{bs} \geq X^2{}_{Q}\text{-}(n+1)\ x\ (k\text{-}1)\ \alpha$ se rechaza la hipótesis nula. La medida (estadístico) de Wilk, $G^2(M)$, se denomina devianza.

#### *3.2.5.2 Calidad del Ajuste.*

Existen otras medidas que pueden calcularse y que proporcionan datos sobre la naturaleza del modelo, como en la regresión lineal donde se utiliza la medida $R^2$. En los modelos de regresión logística binaria, la calidad del ajuste se estima mediante coeficientes de seguridad conocidos como Pseudo-$R^2$; para la regresión logística multinomial se utilizan además estos coeficientes. Los coeficientes más utilizados son los de Mc-Fadden, Cox Snell y Nagelkerke (Fernández & Fernández, 2004).

a. ***Coeficiente Pseudo-R2 de Mc-Fadden.*** Suponiendo que se tiene Λ=-2ln ($V$), distinguimos que Λ0 es el valor subyacente de esta capacidad; en otras palabras, el menor valor de Λ bajo el modelo inválido dado, exclusivamente, por un término consistente y por $\Lambda f$, la base de Λ bajo el modelo ajustado con todos los parámetros, se obtiene la expresión que a continuación se muestra de la pseudo-R2 de Mc-Fadden (Fernández & Fernández, 2004):

$$R_{MF}^2 = 1 - \frac{\Lambda_f}{\Lambda_0}. \tag{39}$$

Siendo su ámbito teórico de valores $0 \leq R^2{}_{MF} \leq 1$; sin embargo, es poco frecuente que su valor se acerque a 1. Generalmente, se considera una naturaleza de ajuste decente cuando $0{,}1 \leq R^2{}_{MF} \leq 0{,}4$ y fenomenal para valores superiores.

b. ***Coeficiente Pseudo-R2 de Cox-Snell.*** Para esta situación se aplica la capacidad de probabilidad (o verosimilitud) V y no la capacidad de asistencia o función auxiliar Λ. Por lo tanto, mediante $V_0$= exp(- $\Lambda_0$/2), la probabilidad más extrema bajo el modelo nulo, dado simplemente por un término estable y mediante Vf = exp(- $\Lambda f$/2) la mayor probabilidad (o verosimilitud) bajo el modelo ajustado con todos los parámetros, se determina el coeficiente pseudo-$R^2$ de Cox-Snell como (Fernández & Fernández, 2004):

$$R_{cs}^2 = 1 - \left(\frac{V_0}{V_f}\right)^{\frac{2}{N}} = 1 - exp\left(\frac{\Lambda_f - \Lambda_0}{N}\right) \tag{40}$$

El ámbito teórico de valores para el coeficiente es $0 \leq R^2cs \leq 1$, por lo que, al depender de $V_0$, reduce su capacidad de ser interpretado debido a que se acercaría al valor de cero cuando hay poca información. Por lo tanto, lo mejor es utilizar el coeficiente de acompañamiento como medida de bondad del ajuste (Fernández & Fernández, 2004).

***c. Coeficiente Pseudo-R2 de Nagelkerke.*** Representado por la *ecuación siguiente*:

$$R_N^2 = \frac{R_{cs}^2}{1 - V_0^{\frac{2}{N}}} = \frac{1 - exp\left(\frac{\Lambda_f - \Lambda_0}{N}\right)}{1 - exp\left(\frac{-\Lambda_0}{N}\right)} \tag{41}$$

Además, para esta situación, su ámbito de valores es $0 \leq R^2{}_N \leq 1$, por lo que tiende a ser interpretado de forma similar al coeficiente de aseguramiento de la regresión lineal clásica, a pesar de que es más desafiante, pues rara vez llega a tener valores cercanos a 1. Para contrastar los modelos de regresión logística multinomial y diversos números de variables indicadores, se suelen presentar coeficientes Pseudo-$R^2$ modificados. El más popular es el coeficiente de Mc Fadden, caracterizado como (Fernández & Fernández, 2004):

$$Adj - R_{MF}^2 = 1 - \frac{0.5\Lambda_f + n + 1}{0.5\Lambda_0 + 1} \tag{42}$$

Donde n es la cantidad de indicador.

### *3.2.5.2. Tasa de Clasificaciones Correctas.*

Consiste en que, a partir del modelo ajustado, cada percepción se ordena en la clasificación más probable, desarrollando así una red de agrupaciones notadas anticipadas donde el nivel de caracterizaciones correctas se usa como una proporción de la naturaleza de la anticipación, de forma similar que en el examen discriminante. La tasa de clasificaciones correcta es definida como la extensión de elementos ordenados con precisión por el modelo y se determina como la proporción entre la cantidad de percepciones ordenadas con precisión y el tamaño de la muestra N. Un elemento está ordenado con precisión por el modelo cuando su valor notado de la variable de reacción Y ($Y_1$, $Y_2$, ..., $Y_k$) coincide con su valor evaluado por el modelo (Fernández & Fernández, 2004).

### 3.2.6. Método de selección del modelo

Después de haber identificado cómo realizar el ajuste de modelos de regresión logística multinomial, se debe desarrollar las estrategias para la selección de las variables que expliquen la variable respuesta bajo la premisa de parsimonia, es decir, aquel modelo que con el mínimo número de parámetros o variables independientes se ajuste a los datos y permita una interpretación sencilla respecto a cocientes de ventaja (Silva & Barroso, 2004).

Antes de continuar, se debe prestar especial atención a las covariables cualitativas que se transforman en variables falsas. Siempre que se incorpore o, por el contrario, se retire una de estas variables, todas las demás clasificaciones deberían incorporarse o retirarse en bloque; de lo contrario, la variable habría sido recodificada y, en consecuencia, la comprensión de la variable no sería correcta. Asimismo, hay que considerar la importancia que pueden tener las variables falsas, ya que no todas las clases de una covariable tienen la misma significancia; cuando se de esta circunstancia, es prudente revisar el modelo total frente al modelo sin la covariable haciendo uso de la prueba de proporción de probabilidad o (verosimilitud), eligiendo incorporar o retirar la covariable en función del resultado experimental y del interés clínico de la covariable. Suponiendo que se obtenga significancia en esta prueba, la variable se mantendrá en el modelo, en el caso de que no se adquiera importancia y la covariable sea de interés clínico, su consideración en el modelo queda a criterio del investigador (Silva & Barroso, 2004).

De acuerdo a Silva & Barroso (2004) el modelo puede ajustarse por diferentes métodos:

#### *3.2.6.1. Selección Hacia Adelante.*

1. Partimos de un modelo vacío (solo la constante).
2. Se realiza el ajuste de un modelo y se determina el valor p del contraste de proporciones de probabilidad (o verosimilitud) que se produce al incluir cada variable de forma independiente.
3. Se elige el modelo cuyo p-valor tenga mayor significancia.
4. Nuevamente, se realiza el ajuste de nuevo un modelo con la(s) variable(s) elegida(s) y se

determina el p-valor de añadir cada variable no elegida recientemente de forma independiente.

5. Se elige el modelo con el p-valor que tenga mayor significancia.
6. Se repite el proceso 4 y 5 hasta que no queden variables significativas por incorporar.

#### *3.2.6.2. Selección Hacia Atrás.*

1. Se parte de un modelo con todos los variables ascendentes.
2. Se excluye cada variable individualmente y se determina la deficiencia de ajuste mientras se elimina.
3. Se escoge la variable con menor significancia de tamaño.
4. Se repite el proceso 2 y 3 hasta que cada una de las variables que se incluyan sea crítica y no se pueda eliminar ninguna sin pérdida de ajuste.

#### *3.2.6.3. Selección Stepwise (Paso a Paso).*

En esta estrategia se consolidan las técnicas de adelante y de atrás. Se puede comenzar con el modelo vacío o con el modelo total, pero en todo el proceso se investigan las variables incluidas, en el caso de que deban retirarse, y los no elegidos, en el caso de que deban incorporarse.

El método Stepwise utiliza procedimientos estadísticos para seleccionar las variables que más contribuyen a la predicción de la variable de salida, o la entrada forzada, en la que el investigador determina qué variables se incluyen basándose en la teoría o la práctica (Petrucci, 2009).

# IV. DESARROLLO DEL TRABAJO

## 4.1. Contribución en la solución del problema

El área de Gestión de Información en atención al requerimiento del avance de los indicadores que periódicamente miden la anemia, tomó como necesario no solo contar con información descriptiva de las prestaciones que se vienen realizando sino agregar un mayor nivel de análisis. Por ello, a través de la técnica estadística Regresión Logística Multinomial se permitió determinar los factores que determinan el peso de recién nacido en madres adolescente en Lima y con ello las áreas encargadas del seguimiento y monitoreo de las prestaciones de salud que ofrecen a los pacientes en los establecimientos consideren dentro de sus estrategias y monitoreo estos factores, representado en campañas, planes y otros enfocados en el binomio madre-niño para mejorar los indicadores evaluados por la gestión.

## 4.2. Análisis de la contribución en términos de competencias y habilidades

Los conocimientos obtenidos en la etapa universitaria permitieron identificar el estado de los datos desde su originan en el nivel operativo y con ello proponer el uso de diferentes técnicas estadísticas como la regresión logística y los resultados aportar al análisis técnico sobre las prestaciones que realizan en los establecimientos y hospitales y tomar decisiones descritas en reuniones de gestión.

## 4.3. Nivel de beneficio obtenido por el centro laboral

Con lo presentado se fortaleció a capacidad del área de trabajo de poder realizar análisis más complejos y brindar mayor información a los responsables de las intervenciones en salud, con ello la dirección general enfatiza el trabajo interdisciplinario y complementar con el sector educación a través de promoción de la salud en acciones preventivas a través de capacitaciones, campañas, etc., en las etapas de gestión de la madre.

# V. RESULTADOS Y DISCUSIÓN

La estructura del trabajo comprende 2 fases: descriptiva y aplicativa. En la fase descriptiva se caracterizan las variables mediante recuentos, porcentajes de frecuencia, estadísticos de resumen y de dispersión. En la fase aplicativa se identifican los factores determinantes en el peso del recién nacido de madres adolescentes en Lima Metropolitana y se emplea la técnica de regresión logística multinomial. Esta técnica es una extensión de los modelos logísticos en el caso de analizar una variable dependiente categórica con más de dos posibles respuestas.

En la aplicación de la Regresión Logística Multinomial se utilizó como variable dependiente: Y==Peso del recién nacido (se fijaron k=3 clases: 1= bajo peso, 2=normal, 3= macrosómico). Este modelo se puede representar con las siguientes funciones, que brindan la probabilidad de pertenencia a las primeras k-1 clases de peso de niño al nacer:

$$\pi_{in} = \frac{e^{Z_{in}}}{1+e^{Z_{i1}}+e^{Z_{i2}}+\ldots+e^{Z_{ik}-1}} \tag{43}$$

$$Z_{in} = \beta_{n0} + \beta_{n1}X_{i1} + \beta_{n2}X_{i2} + \cdots \beta_{nj}X_{ij} \tag{44}$$

Donde la expresión $\pi in$ es la probabilidad de pertenencia del caso i a la clase de peso n; Zin es el valor de la variable dependiente Z, correspondiente a la clase peso n en el caso i; $\beta nj$ es el coeficiente de la variable independiente j para la clase n; Xij es el valor de la variable independiente j para el caso i cono se detalla a continuación:

Xi1= Edad de la madre (Años).
Xi2= Número de hijos nacidos vivos que ha tenido la madre incluido el recién nacido (Un hijo, dos hijos y tres hijos).
Xi3= Número de hijos nacidos vivos que fallecieron (cero y uno).
Xi4= Número de abortos (Cero, uno y dos).

Xi5= Duración del embarazo (Entre 24 y 28 semanas, entre 30 y 34 semanas, entre 35 y 38 semanas, entre 39 y 42 semanas).

Xi6= Estado civil de la madre (Soltera, casada y conviviente).

Xi7= Lugar de procedencia de la madre (Lima, Provincia).

Xi8= Nivel educativo de la madre (Primaria, secundaria, superior universitaria y superior no universitaria).

Xi9= Sexo del recién nacido vivo (Masculino y femenino).

Xi10= Financiador (Usuario, ESSALUD, Privados y Sanidad EP).

Xi11= Tipo de parto (Único y doble).

Xi12= Condición de parto (cesárea, espontaneo, instrumentada u otro)

Xi13= Tipo de documento de identidad de la madre (Acta de nacimiento y DNI)

Xi14= Ocupación de la madre (Ama de casa, trabaja).

## 5.1. Análisis descriptivo

### a. Población y muestra

En el análisis de considera población a los registros de certificados de nacido vivo que se encontraban registrados en el Sistema de Información de Nacimientos, de los cuales se tomó los 1655 registros de nacido vivo (nacimientos) ocurridos en el periodo del 01 de enero al 31 de diciembre 2016 en los establecimientos de salud de Lima Metropolitana de madres adolescentes. La muestra de registros considerados validos en el Sistema de Información de Nacimientos son los 1655 que ocurren en el periodo indicado.

### b. Resultados descriptivos

La variable peso del recién nacido, se le considera la variable de estudio, los valores encontrados se encuentran descritos en la siguiente Tabla 1, las siguientes 14 variables son las predictivas en la Tabla 2 se detallan los valores que se tiene de cada una de ellas.

Las variables cualitativas se codificaron de manera numérica con 0 y 1, generalmente, con lo cual se posteriormente se trabajará en el modelo de regresión logística multinomial.

De la Tabla 1, los valores que toma la variable categoría del peso del recién nacido son: bajo peso al nacer (menor de 2,500 gramos), que se presentan en el 4.89% de los registros; peso normal (2,500 a 3,999 gramos), presente en el 90.94% de los registros; y peso macrosómico (mayor a 4,000 gramos), en el 4.17% de las madres.

**Tabla 1.**

*Descripción de la variable dependiente*

| Variable | Nombre de la variable | Valores | | | Descriptivo |
|---|---|---|---|---|---|
| Categoría del Peso del Recién Nacido | CATGPESO | Bajo peso al nacer | < 2,500 gramos | 0=BPN | 81(4.89%) |
| | | Peso normal | 2,500 g a 3,999 gramos | 1=Normal | 1,505(90.94%) |
| | | Peso macrosómico | >4,000 gramos | 2=Macrosómico | 69(4.17%) |

FUENTE: Elaboración propia basado en los resultados del análisis.

La Tabla 2 lista los valores que toman cada una de las variables predictivas, además de mostrar de manera general la media, máximo y mínimo de la única variable continua (edad de la madre) y el número de los casos y porcentajes, de cada una de las variables cualitativas. Se detalla la frecuencia de cada uno de los valores de las variables presentes en la base de datos; la edad promedio de la madre al momento de parto es de 16.85 años, aunque algunos de los casos muestran que la edad de la puérpera varía entre los 12 a 17 años de edad. De la base de datos se tiene que el 93.47% de las madres adolescentes indican que este es su primer hijo vivo, el 6.34% indica que es el segundo hijo vivo al momento del parto y se presenta que el 0.18% indica que es el tercer hijo vivo nacido. Con respecto al número de hijos nacidos fallecidos, el 99.27% no tuvieron ninguno, pero se presentó que 0.73% si tuvo 1 hijo fallecido y, del número de abortos que han tenido las madres, el 92.63% no ha tenido ninguno, el 7.25% ha tenido al menos y solo el 0.12% ha tenido 2 abortos. El número de semanas de gestación es la variable de duración del embarazo con valores de 39 a 42 semanas se presentaron en el 69.7% de los registros, 35 a 38 semanas alcanzaron el 27.7% de los registros, de 30 a 34 semanas alcanzan el 2.3% y de 24 a 28 semanas solo se presentó en el 0.4% de los registros. El estado civil de las madres muestra al 99.40% indicar condición de soltera, 0.42% condición de conviviente y solo el 0.18% condición de casada. De las madres que han dado a luz en los establecimientos de salud ubicados en Lima Metropolitana, 87.60% procede de la región Lima y el 12.40% procede de las diferentes regiones del país. El nivel educativo de la madre llego a ser: secundaria (88.22%), primaria (7.85%), superior no universitario (2.18%) y superior universitaria (1.75%). El 50.57% de los registros indican sexo masculino del recién nacido vivo y el 49.43%, sexo femenino. El financiador predominantemente de estos partos fue el SIS con 78.07%, 15.5% por el mismo usuario, 6.22% por ESSALUD y el 0.66% a través de privados y/o la sanidad de las fuerzas armadas. El tipo de parto fue único en el 99.58% de los casos, solo el 0.42% fue parto doble, la condición de parto fue 71.9% por parto espontaneo, 27.73% cesárea, 0.24% instrumentado y 0.13% otros. El 99.20% de las madres contó con DNI como tipo de documento de identidad, mientras que el 0.80% solo contó con acta de nacimiento; el 99.70% indicó ama de casa como ocupación y solo el 0.3% indicó trabajar fuera de casa.

**Tabla 2.**

*Descripción de las variables independientes*

| Variable | Nombre de la variable | Valores | Descriptivo |
|---|---|---|---|
| Edad De La Madre | EDAD_MADRE | Años | Media: 16.85<br>Desviación Estándar: 0.57<br>Mínimo: 12<br>Máximo: 17 |
| Número de Hijos nacidos vivos que ha tenido la madre incluido el recién nacido vivo | NRO_HIJS_VIVOS | Un Hijo = valor 0<br>Dos Hijos = valor 1<br>Tres Hijos = valor 2 | 1547(93.47%)<br>105(6.34%)<br>3(0.18%) |
| Número De Hijos Fallecidos | NRO_HIJS_FALLEC | Cero=0 Uno=1 | 1643(99.27%)<br>12(0.73%) |
| Número De Abortos | NRO_ABORT | Cero=0<br>Uno=1<br>Dos=2 | 1533(92.63%)<br>120(7.25%)<br>2(0.12%) |
| Duración del embarazo (número de semanas de gestación) | SS_GEST | Entre 24 Y 28 Semanas= valor 1<br>Entre 30 Y 34 Semanas= valor 2<br>Entre 35 Y 38 Semanas= valor 3<br>Entre 39 Y 42 Semanas= valor 4 | 6(0.4%)<br>38(2.3%)<br>458(27.7%)<br>1153(69.7%) |
| Estado civil de la Madre | ESTAD_CIVIL_MADRE | Soltera=0<br>Casada=1<br>Conviviente=2 | 1645(99.40%)<br>3(0.18%)<br>7(0.42%) |
| Lugar de procedencia de la Madre | ORIG_MADRE | Lima=0<br>Provincia=1 | 1450(87.60%)<br>205(12.40%) |
| Nivel Educativo De La Madre | NIVL_INST_MADRE | Primaria=1<br>Secundaria=2<br>Superior Universitario=3<br>Superior No Universitario=4 | 130(7.85%)<br>1460(88.22%)<br>29(1.75%)<br>36(2.18%) |
| Sexo del recién nacido vivo | SEX_RN | Masculino=0<br>Femenino=1 | 818 (49.43%)<br>837(50.57% |

| Variable | Nombre de la variable | Valores | Descriptivo |
|---|---|---|---|
| Financiador | FINANCIADOR | SIS=0<br>Usuario=1<br>ESSALUD=2<br>Privados=3<br>Sanidad EP=4 | 1292(78.07%)<br>249(15.05%)<br>103(6.22%)<br>10(0.60%)<br>1(0.06%) |
| Tipo de Parto | TIPO_PARTO | Único=0<br>Doble=1 | 1648(99.58%)<br>7(0.42%) |
| Condición de Parto | COND_PARTO | Cesárea=0<br>Espontaneo=1<br>Instrumentado=2<br>Otro=3 | 459(27.73%)<br>1190(71.90%)<br>2(0.13%)<br>4(0.24%) |
| Tipo de documento de Identidad de la Madre | DOC_MADRE | Acta De Nacimiento=0<br>DNI=1 | 14(0.80%)<br>1641(99.20%) |
| Ocupación de la Madre | OCUP_MADRE | Ama De Casa=0<br>Trabaja=1 | 1650(99.70%)<br>5((0.30%) |

FUENTE: Elaboración propia basado en los resultados del análisis.

Las variables predictivas número de hijos nacidos vivos que ha tenido la madre incluido el recién nacido vivo, número de hijos fallecidos, número de abortos, duración del embarazo (número de semanas de gestación), tipo de parto y condición de parto son consideradas como factores obstétricos, la edad de la madre, el estado civil de la madre, lugar de procedencia de la madre, nivel educativo de la madre, tipo de documento de identidad de la madre y ocupación de la madre como factores socio demográficos y por último la variable sexo del recién nacido vivo como factores neonatales.

## 5.2. Metodología de la regresión logística múltiple

La metodología de presentación de resultados empleada en la investigación es una adaptación de la propuesta Dueñas (2013) y Roque (2018), la cual considera 4 pasos: (1) Formulación y Selección del Modelo con el método de Stepwise, (2) Cálculo de las Odds Ratios y los Intervalos de Confianza, (3) Contraste sobre los Parámetros, (4) Ajuste Global del Modelo mediante el Contraste condicional de razón de verosimilitud, y los psuedo R2 de Cox-Snell, Nagelkerke y MCFadden.

Para el procesamiento de datos se utilizó el programa R.4.1.3. El ajuste de los modelos de regresión logística multinomial se realizó a través de la función “multinom”, de la librería “nnet”. Para ello, se cargó previamente el paquete “nnet”, considerando en la función: multinom(formula, data, weights, subset, Hess = FALSE, model = FALSE, ...).

Las pruebas fueron bilaterales y el nivel de significación estadística fue de $p<0.05$.

### 5.2.1. Paso 1: Formulación y selección del modelo

El peso al nacer es considerado la variable objetivo, a través de la regresión logística multinomial se obtuvieron 2 modelos para esta variable; es decir, g-1 modelos, siendo g el número de categorías de la variable objetivo: Bajo Peso al Nacer, Normal y Macrosómico.

#### *5.2.1.1. Formulación.*

En el modelo de Regresión Logística Multinomial se considera una variable dependiente de naturaleza categórica representada por el peso del recién nacido, esta variable toma las siguientes clases o categorías: 0= bajo peso BPN, 1=normal, 2= macrosómico. También, se consideran 14 variables independientes: Xi1= Edad de la madre (Años), Xi2= Número de hijos nacidos vivos que ha tenido la madre incluido el recién nacido (Un hijo, dos hijos y tres hijos), Xi3= Número de hijos nacidos vivos que fallecieron (cero y uno), Xi4= Número de abortos (Cero, uno y dos), Xi5= Duración del embarazo (Entre 24 y 28 semanas, entre 30 y 34 semanas, entre 35 y 38 semanas, entre 39 y 42 semanas), Xi6= Estado civil de la madre (Soltera, casada y conviviente), Xi7= Lugar de procedencia de la madre (Lima, Provincia). Xi8= Nivel educativo de la madre (Primaria, secundaria, superior universitaria y superior no universitaria), Xi9= Sexo del recién nacido vivo (Masculino y femenino), Xi10= Financiador (Usuario, ESSALUD, Privados y Sanidad EP), Xi11= Tipo de parto (Único y doble), Xi12= Condición de parto (cesárea, espontaneo, instrumentada u otro), Xi13= Tipo de documento de identidad de la madre (Acta de nacimiento y DNI) y Xi14= Ocupación de la madre (Ama de casa, trabaja)

Las 14 variables ingresan al modelo como posibles factores, en el caso de estas variables en la Tabla 2 se indica la categoría de referencia (valor 0) de cada una de estas formulando así las variables dummies. En el caso de la variable edad, esta es considerada discreta. El modelo inicial, si se consideraran todas las variables, sería el que se muestra en la Tabla 3. Allí se puede ver los coeficientes de cada variable predictora y el valor de AIC (834.7061).

**Tabla 3.**

*Modelo inicial considerando todas las variables*

| **Categoría de peso** | **Coeficientes** | |
|---|---|---|
| | Normal | Macrosomico |
| (Intercept) | -8.004342 | -17.283106 |
| EDAD_MADRE | 0.2196379 | 0.3369799 |
| NRO_HIJS_VIVOS | -0.1129246 | 0.5181094 |
| NRO_HIJS_FALLEC | 0.06444158 | -29.1582647 |
| NRO_ABORT | 0.9731967 | 1.7408492 |
| SS_GEST | 3.338355 | 5.31566 |
| ESTAD_CIVIL_MADRE | 19.935488 | -5.987809 |
| ORIG_MADRE | -0.4000752 | -0.6435016 |
| NIVL_INST_MADRE | -0.2122572 | -0.3853553 |
| SEX_RN | -0.1439232 | -0.9488023 |
| FINANCIADOR | 0.03967595 | 0.08042277 |
| TIPO_PARTO | -17.62908 | -11.31005 |
| COND_PARTO | 0.7182609 | -0.7050191 |
| DOC_MADRE | -0.3169777 | -13.0205182 |
| OCUP_MADRE | 17.575337 | -7.993207 |

| | |
|---|---|
| Residual Deviance | 834.7601 |
| AIC | 834.7601 |

FUENTE: Elaboración propia basado en los resultados del análisis.

### *5.2.1.2. Selección del Modelo.*

El resultado de la selección del modelo es aquel que sea explicativo con el menor número de variables. A través del método de Stepwise, se inicia con el modelo que contenga solo la constante, esta se contrasta con el ingreso o eliminación de cada una de las variables predictoras al modelo. Esta comparación de los modelos se realizará a través de los contrastes condiciones de razón de verosimilitud, comparando el valor de la devianzas que alcanza cada posible modelo.

Para la selección del modelo se utiliza el método de Stepwise el cual permite ir generando los diferentes contrastes de los modelos que se van originando al ingreso o retiro de una variable.

En cada uno de los diferentes momentos que se contrastan los modelos se responde la hipótesis que el modelo es el adecuado con las variables actuales, se inicia con el modelo que considera solo el valor del intercepto o constante:

$H_o$: el modelo es adecuado solo con la constante.
$H_1$: el modelo no es adecuado solo con la constante.

Se inicia definiendo cada uno de los modelos en cada momento y se obtienen los coeficientes, error estándar, estadístico de Wald, devianza y el parámetro AIC. Dentro del Anexo 1 se encuentra los comandos.

Se procede a comparar el modelo solo con el intercepto con los diferentes modelos que tienen solo una variable, a través del ANOVA se compara las devianzas de cada par de modelo y se selecciona aquel cuyo valor de prueba sea más significativo. Esto se replica en los siguientes modelos evaluando agregar o eliminar alguna variable.

En el primer momento, al comparar el modelo solo con el intercepto (Y = constante) con los modelos con una sola variable predictora (Y = b Xi), por medio del contraste de razón de verosimilitud y considerando un nivel de significación ($\alpha$=0.05), es la variable duración del embarazo (SS_GEST) la más significativa porque el p-valor=0 es menor a un $\alpha$=0.05 y, además, produce el mayor cambio en la devianza de 1,213.2338 a 844.0997 (ver Tabla 4).

**Tabla 4.**

*Comparaciones de razón de verosimilitud del modelo en el primer momento*

| | | **Resid. Df** | **Resid. Dev** | **Test** | **Df** | **LR stat.** | **Pr(Chi)** |
|---|---|---|---|---|---|---|---|
| modelo 0 | + intercepto | 3308 | 1213.2338 | | NA | NA | NA |
| modelo (con 1 variable) | +SS_GEST | 3306 | 844.0997 | 1 vs 2 | 2 | 443.3014 | 0 |

FUENTE: Elaboración propia basado en los resultados del análisis.

La primera variable que ingresa al modelo es duración del embarazo: Categoría del peso del recién nacido = intercepto + duración del embarazo (SS_GEST).

Las comparaciones del modelo solo intercepto con cada uno de los demás modelos y los contrastes de razón de verosimilitud se encuentran en el Anexo 1.

En el segundo momento, se compara el modelo con la primera variable ingresada (duración del embarazo) con los diferentes modelos con 2 variables, incorporando una siguiente variable a la variable duración del embarazo.

Al comparar los modelos mediante el contraste de razón de verosimilitud, resulta el modelo que considera a la variable COND_PARTO el más significativo porque el pvalor=1.177079e-07 es menor a un α=0.05 y, además, produce el mayor cambio en la devianza de 844.0997 a 801.1197, como se puede apreciar en la Tabla 5.

**Tabla 5.**

*Comparaciones razón de verosimilitud del modelo con 1 variable vs modelo con 2 variables*

| | | Resid. Df | Resid. Dev | Test | Df | LR stat. | Pr(Chi) |
|---|---|---|---|---|---|---|---|
| modelo (con 1 variable) | +SS_GEST | 3306 | 844.0997 | | NA | NA | NA |
| modelo (con 2 variables) | +SS_GEST + COND_PARTO | 3300 | 801.1197 | 1 vs 2 | 6 | 42.98006 | 1.177079E-07 |

FUENTE: Elaboración propia basado en los resultados del análisis.

La segunda variable que ingresa al modelo es la condición del parto (COND_PARTO): Categoría del peso del recién nacido = intercepto + duración del embarazo (SS_GEST)+ condición del parto (COND_PARTO).

Se evalúa el ingreso de una tercera variable, mediante el contraste de razón de verosimilitud resulta ser este último, que considera a la variable NIVL_INST_MADRE, ser el más significativo porque el p-valor=0.02233868 es menor a un α=0.05 y, además, produce el mayor cambio en la devianza de 801.1197 a 786.3751. En la Tabla 6 se puede ver estos resultados:

**Tabla 6.**

*Comparaciones razón de verosimilitud con el modelo con tres variables*

| | | Resid. Df | Resid. Dev | Test | Df | LR stat. | Pr(Chi) |
|---|---|---|---|---|---|---|---|
| modelo (con 2 variables) | +SS_GEST + COND_PARTO | 3300 | 801.1197 | NA | NA | NA | NA |
| modelo (con 3 variables) | +SS_GEST + COND_PARTO + NIVL_INST_MADRE | 3294 | 786.3751 | 1 vs 2 | 6 | 14.74463 | 2.233868E-02 |

FUENTE: Elaboración propia basado en los resultados del análisis.

Categoría del peso del recién nacido = Constante + SS_GEST + COND_PARTO+NIVL_INST_MADRE.

En este siguiente momento se tiene el modelo que solo considera a las variabes: COND_PARTO y NIVL_INST_MADRE y, mediante el contrastare de razón de verosimilitud, se determina si es significativo o no eliminar la variable SS_GEST, comparando con el modelo que considera a las siguientes variables: SS_GEST, COND_PARTO y NIVL_INST_MADRE (ver Tabla 7).

**Tabla 7.**

*Comparaciones razón de verosimilitud con el retiro de una variable en el modelo*

| | | Resid. Df | Resid. Dev | Test | Df | LR stat. | Pr(Chi) |
|---|---|---|---|---|---|---|---|
| modelo (con 2 variables) | + COND_PARTO + NIVL_INST_MADRE | 3296 | 1,134.8621 | | NA | NA | NA |
| modelo (con 3 variables) | +SS_GEST + COND_PARTO + NIVL_INST_MADRE | 3294 | 786.3751 | 1 vs 2 | 2 | 348.487 | 0 |

FUENTE: Elaboración propia basado en los resultados del análisis.

Mediante el contraste de razón de verosimilitud, resulta ser este último, que considera a la variable SS_GEST, ser el más significativo porque el p-valor=0 es menor a un $\alpha$=0.05 y además produce el mayor cambio en la devianza de 1,134.8621 a 786.3751.

Se continúa realizando las comparaciones con el retiro e ingreso de variables al modelo, llegando hasta al momento en el que el ingreso de una nueva variable no es significativo.

**Tabla 8.**

*Comparaciones razón de verisimilitud del modelo de 5 a 6 variables*

| | | Resid. Df | Resid. Dev | Test | Df | LR stat. | Pr(Chi) |
|---|---|---|---|---|---|---|---|
| modelo (con 5 variables) | +SS_GEST + COND_PARTO + NIVL_INST_MADRE + SEX_RN + NRO_ABORT | 3290 | 769.9324 | | NA | NA | NA |
| modelo (con 6 variables) | modelo c/5 variables + ORIG_MADRE | 3288 | 768.2753 | 1 vs 2 | 2 | 1.657183 | 4.367E-01 |
| | modelo c/5 variables + DOC_MADRE | 3288 | | 1 vs 2 | 2 | 0.8384965 | 6.575E-01 |
| | modelo c/5 variables + EDAD_MADRE | 3288 | | 1 vs 2 | 2 | 0.5261296 | 7.687E-01 |
| | modelo c/5 variables + ESTAD_CIVIL_MADRE | 3286 | | 1 vs 2 | 4 | 1.69434 | 7.917E-01 |
| | modelo c/5 variables + OCUP_MADRE | 3288 | | 1 vs 2 | 2 | 0.5901066 | 7.445E-01 |
| | modelo c/5 variables + NRO_HIJS_VIVOS | 3288 | | 1 vs 2 | 2 | 2.576713 | 2.757E-01 |
| | modelo c/5 variables + NRO_HIJS_FALLEC | 3288 | | 1 vs 2 | 2 | 0.7568789 | 6.849E-01 |
| | modelo c/5 variables + FINANCIADOR | 3282 | | 1 vs 2 | 8 | 9.128929 | 3.315E-01 |
| | modelo c/5 variables + TIPO_PARTO | 3288 | 766.6065 | 1 vs 2 | 2 | 3.325917 | 1.896E-01 |

FUENTE: Elaboración propia basado en los resultados del análisis.

De esta prueba, se decide detener la comparación del modelo hasta 5 variables con los diferentes modelos que consideran una variable adicional, ya que, como resultado del contraste de razón de verosimilitud los p-valores son mayores a $\alpha=0.05$ para cada uno de los modelos, lo cual nos indica que ya no se incluirá ninguna variable. Por lo tanto, el modelo final está conformado por las siguientes variables:

Categoría del peso del recién nacido = Constante+ SS_GEST+ COND_PARTO + NIVL_INST_MADRE + NRO_ABORT + SEX_RN.

Categoría del peso del recién nacido = intercepto + duración del embarazo + condición del parto + nivel educativo de la madre + número de abortos + sexo del recién nacido, o:
CATGPESO = intercepto + SS_GEST + COND_PARTO + NIVL_INST_MADRE + NRO_ABORT + SEX_RN.

El modelo ajustado considera variables cualitativas condición de parto y nivel educativo de la madre con 4 categorías por lo que se crean 4 - 1 variables dummy (dicotómicas), tomando como referencia las categorías cesáreas de la variable condición de parto y nivel primario de variable instrucción de la madre.

Se presentan los parámetros del modelo seleccionado en la Tabla 9:

**Tabla 9.**

*Parámetros del modelo seleccionado*

| | Coeficientes | | Error Estándar | | Wald | |
|---|---|---|---|---|---|---|
| | **Peso macrosómico** | **Peso normal** | **Peso macrosómico** | **Peso normal** | **Peso macrosómico** | **Peso normal** |
| intercepto | -73.649220 | -43.075490 | 6.570399 | 3.842062 | -11.20925 | -11.21156 |
| SS_GEST | 1.878009 | 1.182087 | 0.167775 | 0.102226 | 11.19363 | 11.56342 |
| COND_PARTOEspontáneo | -0.357227 | 1.006915 | 0.423017 | 0.337574 | -0.84447 | 2.98280 |
| COND_PARTOInstrumentado | -5.088109 | 7.359121 | 0.000000 | 0.000097 | -16,771,869.05 | 75,501.18000 |
| COND_PARTOOtro | -12.365887 | -2.755208 | 0.000084 | 1.352900 | -147,201.30000 | -2.03652 |
| NIVL_INST_MADRESecundaria | 1.598161 | 1.227738 | 0.695297 | 0.446302 | 2.29853 | 2.75091 |
| NIVL_INST_MADRESuperior No Universitario | -0.125493 | -0.072314 | 1.448510 | 0.894648 | -0.08664 | -0.08083 |
| NIVL_INST_MADRESuperior Universitario | -10.493223 | -0.218714 | 0.000075 | 0.975212 | -140,041.00000 | -0.22427 |
| SEX_RNMasculino | 0.956031 | 0.124567 | 0.434052 | 0.338906 | 2.20257 | 0.36756 |
| NRO_ABORT | 1.739663 | 0.939748 | 0.849636 | 0.769973 | 2.04754 | 1.22049 |

Residual Deviance: 769.9324
AIC: 809.9324

FUENTE: Elaboración propia basado en los resultados del análisis.

### 5.2.2. Cálculo de las odds ratios y los intervalos de confianza

Habiendo identificado el conjunto de variables que se incluyen en el mejor modelo, es decir, se ha identificado los factores que determinan el peso del recién nacido; normal o macrosomico, se calculan los odds ratios y sus intervalos de confianza.

**Cálculo de los Odds Ratios:** Los odds ratios se expresan con el exponencial de los coeficientes que se muestran en la Tabla 10:

**Tabla 10.**

*Coeficientes de las variables para el cálculo de los odds ratios*

| | **Peso macrosómico exp (coef.)** | **Peso normal exp (coef.)** |
|---|---|---|
| intercepto | 1.03E-32 | 1.96E-19 |
| SS_GEST | 6.540471 | 3.261173 |
| COND_PARTOEspontáneo | 0.6996138 | 2.7371431 |
| COND_PARTOInstrumentado | 6.17E-03 | 1.57E+03 |
| COND_PARTOOtro | 4.26E-06 | 6.36E-02 |
| NIVL_INST_MADRESecundaria | 4.943932 | 3.413499 |
| NIVL_INST_MADRESuperior No Universitario | 0.8820622 | 0.9302389 |
| NIVL_INST_MADRESuperior Universitario | 2.77E-05 | 8.04E-01 |
| SEX_RNMasculino | 2.601352 | 1.132658 |
| NRO_ABORT | 5.695424 | 2.559336 |

FUENTE: Elaboración propia basado en los resultados del análisis.

Las variables duración del embarazo, nivel educativo de la madre (secundaria), sexo del recién nacido (masculino) y el número de abortos; presentan cocientes de ventaja mayores a 1; por lo tanto, estas variables actúan como factores de riesgo para el peso macrosómico de un recién nacido incrementando la probabilidad de ocurrencia del peso del recién nacido macrosómico.

En el caso del recién nacido con peso normal, las variables: duración del embarazo, nivel educativo de la madre, condición de parto (espontaneo), condición de parto (instrumentado), nivel educativo de la madre (secundaria), sexo del recién nacido (masculino); los cocientes de ventaja son mayores que 1; por lo tanto, estas variables actúan como factores de riesgo para el peso normal de un recién nacido.

**Cálculo de los Intervalos de Confianza:** Los intervalos de confianza al 95% de los odds ratios se calculan el intervalo de los coeficientes y posterior de calcula el exponencial.

**Tabla 11.**

*Intervalos de confianza de los Odds Ratio para el modelo de la categoría del peso macrosómico*

| **MACROSOMICO:** | **2.50%** | **97.50%** |
|---|---|---|
| intercepto | 2.64129E-38 | 4.04840E-27 |
| SS_GEST | 4.70759E+00 | 9.08697E+00 |
| COND_PARTOEspontáneo | 3.05341E-01 | 1.60299E+00 |

| MACROSOMICO: | 2.50% | 97.50% |
|---|---|---|
| COND_PARTOInstrumentado | 6.16967E-03 | 6.16968E-03 |
| COND_PARTOOtro | 4.26081E-06 | 4.26221E-06 |
| NIVL_INST_MADRESecundaria | 1.26542E+00 | 1.93157E+01 |
| NIVL_INST_MADRESuperior No Universitario | 5.15852E-02 | 1.50825E+01 |
| NIVL_INST_MADRESuperior Universitario | 2.77196E-05 | 2.77278E-05 |
| SEX_RNMasculino | 1.11105E+00 | 6.09066E+00 |
| NRO_ABORT | 1.07725E+00 | 3.01119E+01 |

FUENTE: Elaboración propia basado en los resultados del análisis.

**Tabla 12.**

*Intervalos de confianza de los odss ratio para el modelo de la categoría del peso NORMAL*

| NORMAL: | 2.50% | 97.50% |
|---|---|---|
| intercepto | 1.05240E-22 | 3.65527E-16 |
| SS_GEST | 2.66906E+00 | 3.98464E+00 |
| COND_PARTOEspontáneo | 1.41239E+00 | 5.30446E+00 |
| COND_PARTOInstrumentado | 1.57016E+03 | 1.57076E+03 |
| COND_PARTOOtro | 4.48578E-03 | 9.01610E-01 |
| NIVL_INST_MADRESecundaria | 1.42333E+00 | 8.18640E+00 |
| NIVL_INST_MADRESuperior No Universitario | 1.61090E-01 | 5.37181E+00 |
| NIVL_INST_MADRESuperior Universitario | 1.18826E-01 | 5.43395E+00 |
| SEX_RNMasculino | 5.82937E-01 | 2.20078E+00 |
| NRO_ABORT | 5.65881E-01 | 1.15752E+01 |

FUENTE: Elaboración propia basado en los resultados del análisis.

Para la categoría del peso macrosómico del recién nacido, en la Tabla 11 se observa que los intervalos de confianza de las variables: duración del embarazo, condición del parto (instrumentado), condición de parto (otro), nivel educativo de la madre (secundaria), nivel educativo de la madre (superior universitario), sexo del recién nacido (masculino) y el número de abortos; no contienen al 1. Esto afirma la significación de los parámetros por el test de Wald.

Para la categoría del peso normal del recién nacido, en la Tabla 12 se observa que los intervalos de confianza de las variables: duración del embarazo, condición del parto (espontáneo), condición del parto (instrumentado), condición de parto (otro) y nivel educativo de la madre (secundaria); no contienen al 1. Esto afirma la significación de los parámetros por el test de Wald.

### 5.2.3. Contraste sobre los parámetros

Aquí se realiza el Contraste condicional de razón de verosimilitud para comprobar el efecto de las variables dependientes que quedan en el modelo, se compara las devianzas del modelo considerando solo el intercepto o constante con el valor de la devianza del modelo final. La prueba de hipótesis para la significación de cada variable de interés con coeficiente βk es la siguiente:

$$Ho: \beta k=0$$
$$H1: \beta k\neq 0$$

**Tabla 13.**

*Contraste condicional de razón de verosimilitud*

| | | Resid. Df | Resid. Dev | Test | Df | LR stat. | Pr(Chi) |
|---|---|---|---|---|---|---|---|
| modelo 0 | + intercepto | 3308 | 1213.234 | | NA | NA | NA |
| | +SS_GEST + COND_PARTO modelo + NIVL_INST_MADRE + 0 final SEX_RN + NRO_ABORT | 3290 | 769.9324 | 1 vs 2 | 2 | 443.3014 | |

FUENTE: Elaboración propia basado en los resultados del análisis.

El contraste condicional de razón de verosimilitud que se muestra en la Tabla 13 sigue una distribución chi-cuadrado con 18 grados de libertad, con un p-valor=0 menor a α=0.05 resulta ser significativo, lo que indica que el modelo final presenta un buen ajuste, rechazando así la hipótesis nula de que todos los coeficientes del modelo, a excepción de la constante, sean cero.

### 5.2.4. Ajuste Global del modelo

A un nivel de significación del 5% (0.05) se realiza el cálculo del ajuste global del modelo, como se puede ver en la Tabla 14:

**Tabla 14.**

*Cálculo del ajuste global del modelo*

| | | valor |
|---|---|---|
| | **X-squared** | 0.9999966 |
| | **tasa de clasificaciones correctas** | 0.9335347 |
| **Calidad de Ajsute del modelo** | **R2 de Mc - Fadden** | 0.3653883 |
| | **R2 de Cox - Snell** | 0.2349819 |
| | **R2 de Nagelkerke** | 0.4522631 |

FUENTE: Elaboración propia basado en los resultados del análisis.

La prueba de hipótesis para la prueba de bondad de ajuste es:

Ho: el modelo se ajusta adecuadamente a los datos.

H1: el modelo no se ajusta adecuadamente a los datos.

Con un p-valor=0.9999966 mayor a $\alpha=0.05$, no se rechaza la hipótesis nula. Esto indica que el modelo es adecuado para el ajuste de los datos.

La tasa de clasificaciones correctas, indica que el 93.35% de los casos se clasificó de manera correcta mediante el modelo.

De la Calidad del ajuste del modelo:

- El valor del R2 de Mc – Fadden (0.3653883) se encuentra en el rango de [0.2,0.4], por lo que se puede decir que el modelo presenta un buen ajuste.
- El valor del R2 de Cox – Snell (0.2349819), es mayor que 0, al igual que el caso anterior, nos indica que el modelo presenta un buen ajuste.
- El valor del R2 de Nagelkerke (0.4522631) indica que el modelo presenta un buen ajuste.

### 5.2.5. Discusión de los resultados

Los hallazgos del modelo seleccionado permitieron identificar que las semanas de gestación de la madre adolescentes de Lima es un factor predictivo del peso macrosómico en recién nacidos de madres adolescentes, coincidiendo con lo encontrado por Barros *et al*. (2014) que una adecuada edad gestacional es un factor predictivo del peso macrosómico en bebes de madres adolescentes ($p<0.001$).

Asimismo, se encontró que el parto espontaneo una característica determinante del peso macrosómico y del peso normal. De igual forma, el parto instrumentado resultó un factor asociado del peso macrosómico y peso normal del neonato

Del mismo modo, los resultados indican que el nivel educativo de la madre adolescente de Lima resultó un factor significativo para predecir la categoría del peso del recién nacido, semejante con lo reportado por Mengesha *et al*. (2017) quienes encontraron en una muestra de madres adolescentes en que no terminar la secundaria resultó un factor significativo ($p<0.001$) obstante, Barros *et al*. (2014) encontró que el nivel de estudio no es un factor predictivo del peso macrosómico en bebes de madres adolescentes.

Se obtuvo que, el sexo masculino como factor neonatal puede duplicar la posibilidad de alcanzar un peso macrosómico (OR=2.601352) y elevar 1.13 veces la posibilidad de ocurrencia de un peso normal en madres adolescentes (OR=1.132658).

Se halló que, el número de abortos previos de las madres adolescentes puede influir en la posibilidad de alcanzar un peso macrosómico y peso normal del neonato.

## VI. CONCLUSIONES

- Se identificaron tres factores determinantes del peso del recién nacido de madres adolescentes en Lima: obstétricos, socio demográficos y neonatales que se resumen en duración del embarazo (en semanas gestacionales), condición del parto, número de abortos, nivel educativo de la madre y sexo del recién nacido.

- Se determinó que de las tres características asociadas al factor obstétrico; duración del embarazo (en semanas gestacionales), condición del parto y número de abortos que explican el peso de los neonatos; el más determinante es la duración del embarazo.

- La condición de parto es una característica determinante en el peso del recién nacido de madres adolescentes en Lima.

- Se logró distinguir que, el número de abortos previos conforma una característica determinante en el peso de los bebes recién nacidos de madres adolescentes en Lima.

- Como factor socio demográfico el nivel educativo de la madre adolescente es una característica determinante en el peso del recién nacido.

- Se determinó que el sexo del neonato es una característica determinante en el peso de los recién nacidos de madres adolescentes en Lima.

# VII. RECOMENDACIONES

- El estudio de los factores determinantes en el peso del recién nacido de madres adolescentes en Lima, puede ser complementada mediante la inclusión de nuevas variables ligadas al aspecto fisiológico y metabólico de la madre.

- Dado que la duración del embarazo conformó un factor determínate en el peso del recién nacido de madres adolescentes, se sugiere continuar el proceso de investigación incluyendo los neonatos de bajo peso.

- En función de los hallazgos de significancia ligados a la condición de parto, se sugiere ampliar la muestra de datos a otros hospitales donde pueda reflejarse con mayor prevalencia las diferentes condiciones de alumbramiento.

- Dado las evidencias ligadas al nivel educativo de la madre, y al ser una variable de difícil modificación corto plazo, es pertinente desarrollar acciones que permitan facilitar las transferencias de conocimientos sobre las prácticas de cuidado del embarazo que favorecen la consecución de un peso normal en los neonatos.

- Se sugiere realizar revisiones de literatura sistemáticas y complementarias para corroborar el resultado del sexo del neonato como factor significativo del peso al nacer.

- Se recomienda el uso de otras técnicas o herramientas estadísticas que validen o confronten los resultados obtenidos respecto a los abortos previos como factores significativos del peso del recién nacido.

# VIII. REFERENCIAS BIBLIOGRAFICAS


Aguilera del Pino, A. (2002). *Modelos de Respuesta Discreta*. Granada, España: Copias Coca, Dep. Legal GR-11554-02.

Agbozo, F.; Abubakari, A.; Der, J. & Jahn, A. (2016). Prevalence of low birth weight, macrosomia and stillbirth and their relationship to associated maternal risk factors in Hohoe Municipality, Ghana. *Midwifery*, 40, 200-206. https://cutt.ly/RBa0lo5

Barros, D.C.D.; Saunders, C.; Santos, M.M.A.D.S.; Líbera, B.D.; Gama, S.G.N.D. & Leal, M.D.C. (2014). The performance of various anthropometric assessment methods for predicting low birth weight in adolescent pregnant women. *Revista Brasileira de Epidemiologia*, *17*, 761-774.

Cárdenas, M.; Pérez, B.; Landaeta, M. & Vásquez, M. (2011). Consumo de alimentos y estado nutricional según estrato socioeconómico en una población infantil de Caracas. *Archivos Venezolanos de Puericultura y Pediatría*, *74*(2), 002-009. http://ve.scielo.org/scielo.php?script=sci_arttext&pid=S0004-06492011000200002

Cerda, J.; Vera C. y Rada G. (2013). Odds ratio: aspectos teóricos y prácticos. Medicina basada en evidencias. *Revista Médica de Chile*, 141 (Número 10), 1329-1335. https://www.scielo.cl/scielo.php?script=sci_arttext&pid=S0034-98872013001000014

Czepiel, S.A. (2002). *Maximum likelihood estimation of logistic regression models: theory and implementation*. https://czep.net/stat/mlelr.pdf

Debella-Gilo, M.; Etzelmuller, B. y Klakegg, O. (2007). *Digital soil mapping using Digital Terrain analysis and statistical modelling integrated into GIS: Examples from Vestfold County of Norway*. ScanGIS´2007 - Proceedings of the 11th Scandinavian Research Conference on Geographical Information Sciences. Ås-Noruega, University of Life Sciences. 237-253.

Dueñas, M. (2013). *Modelos de respuesta discreta en r y aplicación con datos reales*. (Tesis de maestría, Universidad de Granada). DocPlayer.

Fagerland, M.W.; Hosmer, D.W. & Bofin, A.M. (2008). Multinomial goodness-of-fit tests for logistic regression models. *Statistics in medicine, 27*(21), 4238-4253. https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.3202

Fernández, V.P. & Fernández, R.S.M. (2004). Regresión logística multinomial. *Cuadernos de la Sociedad Española de Ciencias Forestales,* (18), 323-327. https://dialnet.unirioja.es/descarga/articulo/2981898.pdf

Flores, L. (2002). *Análisis estadístico de los factores de riesgo que influyen en la enfermedad angina de pecho*. (Trabajo monográfico, Universidad Mayor de San Marcos). Sistema de bibliotecas de la Universidad Mayor de San Marcos. https://sisbib.unmsm.edu.pe/bibvirtual/tesis/Basic/flores_ml/contenido.htm

Franco, J.; Tun, M.; Hernández, J.R. & Serralta, L. (2018). Risk factors for low birth weight according to the multiple logistic regression model. A retrospective cohort study in José María Morelos municipality, Quintana Roo, Mexico. *Medwave, 18*(01).

Hosmer, D.W. and Lemeshow, S. (1989). *Applied Logistic Regression*. John Wiley & Sons, Inc., New York.

Mengesha, H.G.; Wuneh, A.D.; Weldearegawi, B. & Selvakumar, D.L. (2017). Low birth weight and macrosomia in Tigray, Northern Ethiopia: who are the mothers at risk?. *BMC pediatrics*, 17(1), 1-9. https://link.springer.com/article/10.1186/s12887-017-0901-1

Ministerio de Justicia y Derechos Humanos. (2021). *Plan Nacional de Derechos Humanos 2018-2021*. https://cdn.www.gob.pe/uploads/document/file/1539318/PLAN-NACIONAL-2018-2021.pdf.pdf

Ministerio de Salud. (2011). *Norma técnica de salud para el control del crecimiento y desarrollo de la niña y el niño menor de cinco años*. http://bvs.minsa.gob.pe/local/MINSA/2197.pdf

Ministerio de Salud. (2022). S*istema en Línea del Registro del Certificado de Nacido Vivo - CNV*. https://www.minsa.gob.pe/cnv/

Pando, V. y San Martín, R. (2004). Regresión logística multinomial. *Cuadernos de la Sociedad Española de Ciencias Forestales*, 000, (Número 18), 323–327. https://dialnet.unirioja.es/servlet/articulo?codigo=2981898

Roque, M.J. (2018). Modelos de regresión logística multinomial de la calidad de fibra de alpaca huacaya en función de sus características: sexo y edad - Corani, Carabaya, Puno – 2017. (Tesis de titulación, Universidad Nacional del Altiplano).

Silva, L.C. & Barroso, I. (2004). *Regresión logística*. Editorial La Muralla.

Sperandei, S. (2014). Understanding logistic regression analysis. *Biochemia medica, 24*(1), 12-18. https://hrcak.srce.hr/file/171128

Stoltzfus, J.C. (2011). Logistic regression: a brief primer. *Academic emergency medicine*, *18*(10), 1099-1104.

# IX. ANEXOS

## ANEXO 1. ESTIMACIÓN DE LOS PARÁMETROS

Modelo de Regresión Logística Multinomial COMANDOS CON EL R STUDIO PARA AGREGAR AL ANEXO

```
datos<-
read.delim("clipbo
ard") attach(datos)
head(datos)
```

## MOSTRANDO EL MODELO ORIGINAL CON TODAS LAS VARIABLES

```
 modelo_inicial<- multinom( CATG_PESO~
EDAD_MADRE+NRO_HIJS_VIVOS+NRO_HIJS_FALLEC+NRO_ABORT+SS_GEST+ESTA
D_CIVIL_MADRE+ORIG_MADRE+NIVL_INST_MADRE+SEX_RN+FINANCIADOR+TIPO
_PARTO+COND_PARTO+DOC_MADRE+OCUP_MADRE, data= datos, trace=FALSE)

summary(modelo_inicial,cor=FALSE,Wald=TRUE)
```

### SELECCIÓN DEL MODELO

### a. Paso N° 1: contrastes de razón de verosimilitud con el modelo0 hasta el modelo14

```
library(nnet)

modelo0 <- multinom( CATG_PESO~ 1, data= datos, trace=FALSE) modelo1 <-
multinom( CATG_PESO~ORIG_MADRE, data= datos, trace=FALSE) modelo2
<- multinom( CATG_PESO~ DOC_MADRE, data= datos, trace=FALSE) modelo3
<- multinom( CATG_PESO~ EDAD_MADRE, data= datos, trace=FALSE)
modelo4 <- multinom( CATG_PESO~ ESTAD_CIVIL_MADRE, data= datos,
trace=FALSE) modelo5 <- multinom( CATG_PESO~ OCUP_MADRE, data=
datos, trace=FALSE) modelo6 <- multinom( CATG_PESO~ NRO_HIJS_VIVOS,
data= datos, trace=FALSE) modelo7 <- multinom( CATG_PESO~
NRO_HIJS_FALLEC, data= datos, trace=FALSE) modelo8 <- multinom(
CATG_PESO~ NRO_ABORT, data= datos, trace=FALSE) modelo9 <- multinom(
CATG_PESO~ NIVL_INST_MADRE, data= datos, trace=FALSE) modelo10 <-
multinom( CATG_PESO~ SEX_RN, data= datos, trace=FALSE) modelo11 <-
multinom( CATG_PESO~ COND_PARTO, data= datos, trace=FALSE) modelo12
<- multinom( CATG_PESO~ FINANCIADOR, data= datos, trace=FALSE)
modelo13 <- multinom( CATG_PESO~ TIPO_PARTO, data= datos,
trace=FALSE) modelo14 <- multinom( CATG_PESO~ SS_GEST, data= datos,
trace=FALSE)
```

**Contraste de razón de verosimilitud**

anova(modelo0,modelo1)
anova(modelo0,modelo2)
anova(modelo0,modelo3)
anova(modelo0,modelo4)
anova(modelo0,modelo5)
anova(modelo0,modelo6)
anova(modelo0,modelo7)
anova(modelo0,modelo8)
anova(modelo0,modelo9)
anova(modelo0,modelo10)
anova(modelo0,modelo11)
anova(modelo0,modelo12)
anova(modelo0,modelo13)
anova(modelo0,modelo14)

**b. Paso N° 2: contrastes de razón de verosimilitud con el modelo1 hasta el modelo14**

modelo1<- multinom( CATG_PESO~ SS_GEST, data= datos, trace=FALSE) modelo2<- multinom( CATG_PESO~ SS_GEST+ORIG_MADRE, data= datos, trace=FALSE) modelo3<- multinom( CATG_PESO~ SS_GEST+DOC_MADRE, data= datos, trace=FALSE) modelo4<- multinom( CATG_PESO~ SS_GEST+EDAD_MADRE, data= datos, trace=FALSE) modelo5<- multinom( CATG_PESO~ SS_GEST+ESTAD_CIVIL_MADRE, data= datos, trace=FALSE) modelo6<- multinom( CATG_PESO~ SS_GEST+OCUP_MADRE, data= datos, trace=FALSE) modelo7<- multinom( CATG_PESO~ SS_GEST+NRO_HIJS_VIVOS, data= datos, trace=FALSE) modelo8<- multinom( CATG_PESO~ SS_GEST+NRO_HIJS_FALLEC, data= datos, trace=FALSE) modelo9<- multinom( CATG_PESO~ SS_GEST+NRO_ABORT, data= datos, trace=FALSE) modelo10<- multinom( CATG_PESO~ SS_GEST+NIVL_INST_MADRE, data= datos, trace=FALSE) modelo11<- multinom( CATG_PESO~ SS_GEST+SEX_RN, data= datos, trace=FALSE) modelo12<- multinom( CATG_PESO~ SS_GEST+COND_PARTO, data= datos, trace=FALSE) modelo13<- multinom( CATG_PESO~ SS_GEST+FINANCIADOR, data= datos, trace=FALSE) modelo14<- multinom( CATG_PESO~ SS_GEST+TIPO_PARTO, data= datos, trace=FALSE)

**Contraste de razón de verosimilitud**

anova(modelo1,modelo2)
anova(modelo1,modelo3)
anova(modelo1,modelo4)
anova(modelo1,modelo5)
anova(modelo1,modelo6)
anova(modelo1,modelo7)
anova(modelo1,modelo8)
anova(modelo1,modelo9)
anova(modelo1,modelo10)
anova(modelo1,modelo11)
anova(modelo1,modelo12)
anova(modelo1,modelo13)
anova(modelo1,modelo14)

## c. Paso N° 3: inclusión de variables, contrastes de razón de verosimilitud del modelo1 hasta el modelo 13

modelo1<- multinom( CATG_PESO~ SS_GEST+COND_PARTO, data= datos, trace=FALSE)
modelo2<- multinom( CATG_PESO~ SS_GEST+COND_PARTO+ORIG_MADRE, data= datos, trace=FALSE)
modelo3<- multinom( CATG_PESO~ SS_GEST+COND_PARTO+DOC_MADRE, data= datos, trace=FALSE)
modelo4<- multinom( CATG_PESO~ SS_GEST+COND_PARTO+EDAD_MADRE, data= datos, trace=FALSE)
modelo5<- multinom( CATG_PESO~ SS_GEST+COND_PARTO+ESTAD_CIVIL_MADRE, data= datos, trace=FALSE)
modelo6<- multinom( CATG_PESO~ SS_GEST+COND_PARTO+OCUP_MADRE, data= datos, trace=FALSE)
modelo7<- multinom( CATG_PESO~ SS_GEST+COND_PARTO+NRO_HIJS_VIVOS, data= datos, trace=FALSE)
modelo8<- multinom( CATG_PESO~ SS_GEST+COND_PARTO+NRO_HIJS_FALLEC, data= datos, trace=FALSE)
modelo9<- multinom( CATG_PESO~ SS_GEST+COND_PARTO+NRO_ABORT, data= datos, trace=FALSE) modelo10<- multinom( CATG_PESO~ SS_GEST+COND_PARTO+NIVL_INST_MADRE, data= datos, trace=FALSE)
modelo11<- multinom( CATG_PESO~ SS_GEST+COND_PARTO+SEX_RN, data= datos, trace=FALSE) modelo12<- multinom( CATG_PESO~ SS_GEST+COND_PARTO+FINANCIADOR, data= datos, trace=FALSE)
modelo13<- multinom( CATG_PESO~ SS_GEST+COND_PARTO+TIPO_PARTO, data= datos, trace=FALSE)

### Contraste de razón de verosimilitud

anova(modelo1,modelo2)
anova(modelo1,modelo3)
anova(modelo1,modelo4)
anova(modelo1,modelo5)
anova(modelo1,modelo6)
anova(modelo1,modelo7)
anova(modelo1,modelo8)
anova(modelo1,modelo9)
anova(modelo1,modelo10)
anova(modelo1,modelo11)
anova(modelo1,modelo12)
anova(modelo1,modelo13)

### Modelo15 solo considerando las variables condición de parto (COND_PARTO) y el nivel educativo de la madre (NIVL_INST_MADRE)

modelo15<- multinom( CATG_PESO~ COND_PARTO+NIVL_INST_MADRE, data= datos, trace=FALSE)

### Modelo10 considerando las variables número de semanas gestacionales o duración del embarazo (SS_GEST), la condición de parto (COND_PARTO) y el nivel educativo de la madre (NIVL_INST_MADRE).

modelo10<- multinom( CATG_PESO~ SS_GEST+COND_PARTO+NIVL_INST_MADRE, data= datos, trace=FALSE)

**CONTRASTE DE RAZÓN DE VEROSIMILITUD**

anova(modelo15,modelo10)

### d. Paso N° 4: contrastes de razón de verosimilitud con el modelo0 hasta el modelo12

modelo1<- multinom( CATG_PESO~ SS_GEST+COND_PARTO+NIVL_INST_MADRE, data= datos, trace=FALSE)
modelo2<- multinom( CATG_PESO~ SS_GEST+COND_PARTO+NIVL_INST_MADRE+ORIG_MADRE, data= datos, trace=FALSE)
modelo3<- multinom( CATG_PESO~ SS_GEST+COND_PARTO+NIVL_INST_MADRE+DOC_MADRE, data= datos, trace=FALSE)
modelo4<- multinom( CATG_PESO~ SS_GEST+COND_PARTO+NIVL_INST_MADRE+EDAD_MADRE, data= datos, trace=FALSE)
modelo5<- multinom( CATG_PESO~ SS_GEST+COND_PARTO+NIVL_INST_MADRE+ESTAD_CIVIL_MADRE, data= datos, trace=FALSE) modelo6<- multinom( CATG_PESO~ SS_GEST+COND_PARTO+NIVL_INST_MADRE+OCUP_MADRE, data= datos, trace=FALSE)
modelo7<- multinom( CATG_PESO~ SS_GEST+COND_PARTO+NIVL_INST_MADRE+NRO_HIJS_VIVOS, data= datos, trace=FALSE) modelo8<- multinom( CATG_PESO~ SS_GEST+COND_PARTO+NIVL_INST_MADRE+NRO_HIJS_FALLEC, data= datos, trace=FALSE) modelo9<- multinom( CATG_PESO~ SS_GEST+COND_PARTO+NIVL_INST_MADRE+NRO_ABORT, data= datos, trace=FALSE)
modelo10<- multinom( CATG_PESO~ SS_GEST+COND_PARTO+NIVL_INST_MADRE+SEX_RN, data= datos, trace=FALSE)
modelo11<- multinom( CATG_PESO~ SS_GEST+COND_PARTO+NIVL_INST_MADRE+FINANCIADOR, data= datos, trace=FALSE)
modelo12<- multinom( CATG_PESO~ SS_GEST+COND_PARTO+NIVL_INST_MADRE+FINANCIADOR+TIPO_PARTO, data= datos, trace=FALSE)

**Contraste de razón de verosimilitud**

anova(modelo1,modelo2)
anova(modelo1,modelo3)
anova(modelo1,modelo4)
anova(modelo1,modelo5)
anova(modelo1,modelo6)
anova(modelo1,modelo7)
anova(modelo1,modelo8)
anova(modelo1,modelo9)
anova(modelo1,modelo10)
anova(modelo1,modelo11)
anova(modelo1,modelo12)

### e. Paso N° 5: contrastes de razón de verosimilitud con el modelo1 hasta el modelo10

modelo1<- multinom( CATG_PESO~ SS_GEST+COND_PARTO+NIVL_INST_MADRE+SEX_RN+NRO_ABORT, data= datos, trace=FALSE) modelo2<- multinom( CATG_PESO~ SS_GEST+COND_PARTO+NIVL_INST_MADRE+SEX_RN+NRO_ABORT+ORIG_MADRE, data= datos, trace=FALSE) modelo3<- multinom( CATG_PESO~ SS_GEST+COND_PARTO+NIVL_INST_MADRE+SEX_RN+NRO_ABORT+DOC_MADRE, data= datos, trace=FALSE) modelo4<- multinom( CATG_PESO~ SS_GEST+COND_PARTO+NIVL_INST_MADRE+SEX_RN+NRO_ABORT+EDAD_MADRE, data= datos, trace=FALSE) modelo5<- multinom( CATG_PESO~ SS_GEST+COND_PARTO+NIVL_INST_MADRE+SEX_RN+NRO_ABORT+ESTAD_CIVIL_MADRE, data= datos, trace=FALSE) modelo6<- multinom( CATG_PESO~ SS_GEST+COND_PARTO+NIVL_INST_MADRE+SEX_RN+NRO_ABORT+OCUP_MADRE, data= datos, trace=FALSE) modelo7<- multinom( CATG_PESO~ SS_GEST+COND_PARTO+NIVL_INST_MADRE+SEX_RN+NRO_ABORT+NRO_HIJS_VIVOS, data= datos, trace=FALSE) modelo8<- multinom( CATG_PESO~ SS_GEST+COND_PARTO+NIVL_INST_MADRE+SEX_RN+NRO_ABORT+NRO_HIJS_FALLEC, data= datos, trace=FALSE) modelo9<- multinom( CATG_PESO~ SS_GEST+COND_PARTO+NIVL_INST_MADRE+SEX_RN+NRO_ABORT+FINANCIADOR, data= datos, trace=FALSE) modelo10<- multinom( CATG_PESO~ SS_GEST+COND_PARTO+NIVL_INST_MADRE+SEX_RN+NRO_ABORT+TIPO_PARTO, data= datos, trace=FALSE)

### Contraste de razón de verosimilitud

```
anova(modelo1,modelo2)
anova(modelo1,modelo3)
anova(modelo1,modelo4)
anova(modelo1,modelo5)
anova(modelo1,modelo6)
anova(modelo1,modelo7)
anova(modelo1,modelo8)
anova(modelo1,modelo9)
anova(modelo1,modelo10)
```

### PARÁMETROS DEL MODELO FINAL

```
modelo1<- multinom( CATG_PESO~
SS_GEST+COND_PARTO+NIVL_INST_MADRE+SEX_RN+NRO_ABORT,
data= datos, trace=FALSE)
summary(modelo1,cor=FALSE,Wald=TRUE)
```

## ANEXO 2. CÁLCULO DE LAS ODDS RATIOS Y LOS INTERVALOS DE CONFIANZA

Cálculo de las Odds Ratios y los Intervalos de Confianza modelo1<- multinom( CATG_PESO~

SS_GEST+COND_PARTO+NIVL_INST_MADRE+SEX_RN+NRO_ABORT, data= datos,
trace=FALSE) modelo1
summary(modelo1,cor=FALSE,Wald=TRUE)

**Odds Ratios**

coeficientes<-coef(modelo1)
exp(coeficientes)

**Intervalos de confianza al 95%** ic<-
confint(modelo1)
exp(ic)

# ANEXO 3. CONTRASTE SOBRE LOS PARÁMETROS

## a. Contraste de Wald

```
wald<-matrix(c(-11.20925,11.19363,-0.8444739,-16771869.05,-147201.3,2.298529,-0.08663568,-140041,
2.2025743,2.047539,-11.21156,11.56342,2.9827989,75501.18,-2.036521,2.750910,-0.08082933,0.2242728,0.3675559,1.220494),nrow=10,ncol=2)  wald
pnorm(wald[1,],mean=0,sd=1,lower.tail = TRUE)
pnorm(wald[2:10,],mean=0,sd=1,lower.tail = FALSE)
```

## b. Contraste condicional de razón de verosimilitud

```
modelo0<-multinom( CATG_PESO~ 1, data= datos, trace=FALSE)
modelo1<- multinom( CATG_PESO~
SS_GEST+COND_PARTO+NIVL_INST_MADRE+SEX_RN+NRO_ABORT, data= datos, trace=FALSE)
anova(modelo0, modelo1)
```

# ANEXO 4. AJUSTE GLOBAL DEL MODELO

## a. Ajuste global del modelo

```
modelo1<- multinom( CATG_PESO~
SS_GEST+COND_PARTO+NIVL_INST_MADRE+SEX_RN+NRO_ABORT, data= datos,
trace=FALSE) modelos<- multinom( CATG_PESO~
SS_GEST*COND_PARTO*NIVL_INST_MADRE*SEX_RN*NRO_ABORT, data= datos,
trace=FALSE)
deviance(modelo1)-deviance(modelos)
pchisq(deviance(modelo1)-deviance(modelos),df=12)
```

## b. Tasa de Clasificaciones correctas

```
obs<-datos$CATG_PESO
pre<-
predict(modelo1,type="class")
cont=0
for(i in 1:1655){if (pre[i]==obs[i])cont=cont+1 else
cont=cont} tcc<-cont/1655 tcc
```

## c. Calidad del ajuste del

## modelo R2 de Mc – Fadden

```
modelo0<-multinom( CATG_PESO~ 1, data= datos, trace=FALSE)
modelo1<- multinom( CATG_PESO~
SS_GEST+COND_PARTO+NIVL_INST_MADRE+SEX_RN+NRO_ABORT, data= datos,
trace=FALSE)
dv1<-
deviance(mod
elo1) dv0<-
deviance(mod
elo0) mf<-1-
(dv1/dv0) mf
```

**R2 de Cox –**
**Snell**

```
cs<-1-
exp((dv1-
dv0)/1655) cs
```

**R2 de Nagelkerke**

```
n<-(1-exp((dv1-dv0)/1655))/(1-exp(-
dv0/1655)) n
```