

**UNIVERSIDAD NACIONAL AGRARIA**

**LA MOLINA**

**Ciclo Optativo de Profesionalización en  
Marketing y Finanzas**



**“ANÁLISIS CLASIFICATORIO DE LAS GESTANTES SEGÚN  
VÍA DE CULMINACIÓN DEL PARTO APLICANDO REGRESIÓN  
LOGÍSTICA BINARIA”**

**Trabajo de Titulación para Optar el Título Profesional de:**

**INGENIERO ESTADÍSTICO E INFORMÁTICO**

**Dante Daniel Ballón Beltrán**

**Eduardo Bernabé Ponte**

**Lima – Perú  
2015**

# ÍNDICE

<b>I.</b>	<b>INTRODUCCIÓN.....</b>	<b>1</b>
1.1.	Fundamentación del problema de investigación.....	2
1.2.	Objetivos.....	3
1.2.1.	Objetivo principal.....	3
1.2.2.	Objetivo específico. ....	3
1.3.	Justificación.....	3
<b>II.</b>	<b>REVISIÓN DE LITERATURA.....</b>	<b>5</b>
2.1.	Antecedentes de la investigación.....	5
2.2.	Base teórica.....	7
2.2.1.	Distribución Binomial.....	7
2.2.1.1.	Proceso de Bernoulli.....	7
2.2.2.	Concepto de “ODDS RATIO” (OR) o razón de probabilidad.....	8
2.2.3.	Análisis de datos.....	9
2.2.4.	Análisis descriptivo.....	11
2.2.5.	Análisis de multicolinealidad.....	11
2.2.6.	Regresión logística.....	12
2.2.6.1.	Supuestos de la regresión logística.....	14
2.2.7.	Modelo matemático.....	14
2.2.7.1.	Estimación de los coeficientes.....	16
2.2.7.2.	Pruebas de significancia.....	18
a.	Evaluación estadística del coeficiente $\beta_1$ .....	18
b.	Estadístico WALD.....	19
2.2.7.3.	Bondad de ajuste del modelo.....	20

a. El incremento de estadístico $2\log L$ .....	20
b. Prueba de HosmerLemeshow.....	21
c. Prueba de Ómnibus.....	23
2.2.8. Modelo de Regresión Logística Múltiple.....	23
2.2.8.1. Estimación de los parámetros.....	24
a. Interpretación de los parámetros $\beta$ .....	26
2.2.8.2. Pruebas de significancia.....	29
a. Estadístico WALD.....	31
b. Pruebas de bondad de ajuste.....	32
2.3. Validación cruzada. ....	32
2.4. Medidas de la bondad del ajuste. ....	33
2.4.1. Evaluación de la bondad del ajuste.....	33
2.4.1.1. Coeficiente de determinación Pseudo $R^2$ .....	33
2.4.1.2. Tablas de clasificación (o matriz de confusión).....	34
2.4.1.2.1. Sensibilidad, especificidad y conceptos relacionados.....	35
2.4.1.2.3. Contraste de bondad del ajuste de Hosmer y Lemeshow.....	37
2.5. Análisis de la curva de COR.....	37
<b>III. MATERIALES Y MÉTODOS.....</b>	<b>39</b>
3.1. Lugar. ....	39
3.2. Materiales... ....	39
3.3. Metodología de la investigación.... ....	39
3.3.1. Diseño de investigación.....	39
3.3.2. Población y muestra. ....	40
3.3.3. Tipo de investigación.....	40
3.4. Formulación de hipótesis.....	41

3.5. Identificación de variables.....	41
<b>IV. RESULTADOS Y DISCUSIÓN.....</b>	<b>43</b>
4.1. Prueba binomial.....	44
4.2. Análisis descriptivo.....	44
4.3.Diagnóstico de multicolinealidad .....	53
4.4.Selección de variables.....	54
4.5.Verosimilitud del modelo.....	57
4.6.Prueba de bondad de ajuste.....	58
4.7.Validación cruzada.....	59
4.8.Cumplimiento de supuestos de Regresión Logística.....	65
<b>V.CONCLUSIONES.....</b>	<b>66</b>
<b>VI. RECOMENDACIONES.....</b>	<b>68</b>
<b>VII. REFERENCIAS BIBLIOGRÁFICAS.....</b>	<b>69</b>
<b>VIII. ANEXOS.....</b>	<b>71</b>

## ÍNDICE DE CUADROS

1. Cuadro 1: Prueba binomial.....	44
2. Cuadro 2: Estadística descriptiva de las variables en estudio.....	45
3. Cuadro 3: Diagnostico de multicolinealidad.....	53
4. Cuadro 4: Variables que ingresan al modelo.....	54
5. Cuadro 5: Prueba del coeficiente de determinación.....	55
6. Cuadro 6: Tabla de clasificación.....	56
7. Cuadro 7: Prueba del coeficiente de intercepción.....	57
8. Cuadro 8: Modelo.....	57
9. Cuadro 9: Prueba de bondad de ajuste.....	58
10. Cuadro 10: Prueba de Ómnibus.....	58
11. Cuadro 11: Prueba de validación cruzada - Primer modelo generado.....	59
12. Cuadro 12: Prueba de validación cruzada - Segundo modelo generado.....	60
13. Cuadro 13: Prueba de validación cruzada - Tercer modelo generado.....	60
14. Cuadro 14: Prueba de validación cruzada - Cuarto modelo generado.....	61
15. Cuadro 15: Prueba de validación cruzada - Quinto modelo generado.....	61
16. Cuadro 16: Prueba de validación cruzada - Sexto modelo generado.....	62
17. Cuadro 17: Prueba de validación cruzada - Séptimo modelo generado.....	62
18. Cuadro 18: Prueba de validación cruzada - Octavo modelo generado .....	63
19. Cuadro 19: Prueba de validación cruzada - Noveno modelo generado .....	63
20. Cuadro 20: Prueba de validación cruzada - Décimo modelo generado .....	64
21. Cuadro 21: Prueba T – La media de los errores es cero .....	65

## ÍNDICE DE FIGURAS

1. Figura 1: Clasificación de las gestantes según la vía de culminación de parto (Normal o por cesárea).....	46
2. Figura 2: Edad de la madre respecto a la vía de culminación de parto (Normal o porcesárea).....	47
3. Figura 3: Edad gestacional del bebe por nacer respecto a la vía de culminación de parto (Normal o por cesárea) de las madres gestantes.....	48
4. Figura 4: Número de bebés respecto a la vía de culminación de parto (Normal o por cesárea) de las madres gestantes.....	49
5. Figura 5: Peso del bebe por nacer de las madres gestantes respecto a la vía de culminación de parto (Normal o por cesárea).....	50
6. Figura 6: Talla del bebe por nacer de las madres gestantes respecto a la vía de culminación de parto (Normal o por cesárea).....	51
7. Figura 7: Perímetro cefálico del bebe por nacer de las madres gestantes respecto a la vía de culminación de parto (Normal o por cesárea).....	52
8. Figura 8: Grafica de la curva COR .....	56

## ÍNDICE DE ANEXOS

1. Anexo 1.....	72
2. Anexo 2 .....	75

## **RESUMEN**

En el presente estudio se tiene como objetivo primordial describir y predecir la vía de culminación de parto en gestantes (Normal o por cesárea) mediante un modelo logístico binario y como objetivo secundario identificar los factores asociados que permiten determinar si una gestante tendrá una vía de culminación de parto: normal o por cesárea. Luego se realizará el análisis de regresión logística binaria con el fin de encontrar el mejor modelo que explique la variable de interés en función a las siguientes variables independientes: complicaciones de la madre, edad de la madre, edad gestacional, número de bebés, peso del bebé por nacer, talla del bebé por nacer y perímetro cefálico del bebé por nacer. Los principales resultados que presentamos son: Tasa de clasificación según vía de culminación del parto en gestantes 88.5%; el coeficiente de determinación del modelo 65.9%, pruebas estadísticas de Hosmer y Lemeshow 0.760, lo cual verifica que nuestros datos se ajustan al modelo logístico binario, la tasa de mala clasificación del modelo validada mediante validación cruzada con 11.65%, siendo este un error de estimación bajo y aceptable.

Palabras clave: Regresión logística, tasa de clasificación, validación cruzada.

## **ABSTRACT**

The main purpose of the following study is to be able to describe and predict the appropriate delivery process in pregnant women. (Normal labor or Caesarean section/C-section) using an in depth binary logistic regression. A secondary objective will be to be able to identify different factors that help determine the type of delivery the expectant mother should receive. To obtain the desired results an in depth binary logistic regression analysis will be conducted. This is due to the fact that variable in question is dichotomous (two types of deliveries: normal labor and C-section). The objective of the investigation will be to find a specific model that successfully explains the relationship between the main variable and the following independent variables: complications with the mother, mother's age, gestational age of the newborn, number of

babies, newborn's weight and height and cephalic perimeter. The main results presented are as following: Rate of classification depending on the delivery process 88.5%; a 65.9% coefficient of determination of the model; In regards to the Hosmer-Lemeshow test we obtained a number of 0.760. This allows us to verify that our data is in line with the binary logistic regression model; finally we got a bad classification of the model rate validated through cross-validation of 11.65%. This makes it a low estimation rate and thus acceptable.

Key Words: Logistic Regression, Classification Rate, Cross Validation

## I. INTRODUCCIÓN

Según la Encuesta Demográfica y de Salud Familiar (ENDES) del Instituto Nacional de Estadística e Informática (INEI), en los últimos seis años, la tasa de cesáreas se ha incrementado casi 10 por ciento a nivel nacional. La proyección del 2013 revela que, en promedio, el 26.5 por ciento de los partos institucionales (aquellos que se realizan en establecimientos de salud) han sido cesáreas. Esta cifra agrupa las intervenciones del Ministerio de Salud, Essalud, Fuerzas Armadas y del sector privado.

Si bien, el sector privado es reducido para explicar cambios a nivel general (como en la ENDES) especialistas en salud reproductiva afirman que en comparación con la tasa en clínicas, la situación en hospitales es generalmente baja. Por ejemplo, durante el 2013, en Essalud se atendieron 41 mil 885 cesáreas, a diferencia de los 64 mil 149 partos naturales. (Berríos, M. 2014)

Por lo tanto, establecer mediante la regresión logística un modelo que permita realizar predicciones futuras sobre la vía de culminación de parto de una gestante apoyada en variables independientes sería un aporte para futuros estudios relacionados al tema.

La hipótesis principal es: La vía de culminación del parto de las gestantes se ajusta a una regresión logística binaria.

Por este motivo, el presente estudio busca recabar información útil y necesaria para todas las instituciones ligadas al cuidado de la salud, siendo un aporte a las numerosas investigaciones realizadas con el fin de determinar la mejor vía de culminación de parto.

Por ello, se buscará determinar la importancia de las variables en estudio y el grado de influencia en la determinación de la vía de culminación del parto de la gestante.

## **1.1. FUNDAMENTACIÓN DEL PROBLEMA DE INVESTIGACIÓN**

Los mecanismos fisiológicos implicados en el desencadenamiento del parto han sido materia de numerosas investigaciones; pero a pesar de la gran cantidad de evidencia acumulada, y de las numerosas hipótesis y teorías elaboradas, los procesos involucrados en el inicio del trabajo de parto en la especie humana permanecen aún elusivos. El conocer estos mecanismos no solo tiene importancia desde un punto de vista fisiológico y teórico, sino también desde uno clínico y práctico, por el gran impacto que presentan las patologías como el parto pretermino, cuya prevalencia se ha mantenido relativamente estable durante los últimos años, debido en gran medida al desconocimiento de factores asociados al comienzo de trabajo de parto.

Cada vez es más frecuente que la propia paciente pida la realización de una cesárea (cesárea a demanda), probablemente debido a la creencia, muy difundida de que la cesárea es una práctica igual de segura para la madre y el feto que el parto normal.

Si se practica una cesárea a demanda, teóricamente, la mujer y el feto asumen los riesgos, mientras que si no se hace los asume el médico (alerta, consecuencias médico-legales). Sin embargo, tal como se explica posteriormente, la mortalidad materna global (todo tipo de cesáreas) es aproximadamente seis veces mayor que en el parto vaginal. Estas diferencias son suficientemente relevantes como para que, en ausencia de otras evidencias (ensayos clínicos aleatorizados, controlados sobre diferentes tipos de parto), parezca inoportuno y quizá imprudente cuestionar cual es el modo de parto más seguro.

Cualquier decisión que incremente la cirugía mayor, con la referida mortalidad asociada, debe ser evaluada por todos los agentes involucrados. Los médicos que son los

agentes encargados de informar de este balance riesgo beneficio, no siempre disponen de la información adecuada bien por falta de evidencias científicas o por discrepancias. (Cabero, L; Saldívar, D; Cabrillo, E. 2007).

## **1.2 OBJETIVOS**

### **1.2.1. OBJETIVO PRINCIPAL**

Describir y predecir la vía de culminación de parto en gestantes (normal o por cesárea) mediante un modelo logístico binario.

### **1.2.2. OBJETIVO ESPECÍFICO**

Identificar los factores asociados que permiten determinar si una gestante es clasificada dentro del grupo de gestantes que tendrá una vía de culminación de parto: normal o por cesárea.

## **1.3. JUSTIFICACIÓN**

La vía de culminación del parto de una gestante determina la minimización de complicaciones o aumentos de estas según se determine la vía de parto, es por eso que hemos decidido realizar el presente trabajo para que pueda ser de utilidad al área de gineco-obstetricia de los centros de salud pública y privada, con el fin de elegir la vía de culminación de parto más favorable tanto para la madre como para el feto, así como en la toma de decisiones de los profesionales de la salud para determinar la resolución de parto, todo lo cual, hace de nuestra propuesta un proyecto de importancia que justifica su desarrollo.

La asistencia al parto juega un papel preventivo y útil de la versión externa, así como llegar a conocer la vía del parto vaginal o cesárea, aporta una menor morbilidad y mortalidad materno y perinatal, es decir una mejor calidad de vida materno-fetal.

Debemos reconocer que el incremento de cesáreas en estos últimos años, representa un valor significativo que afecta la economía del estado, en un país pobre como el nuestro, por lo que debemos tomar en cuenta en la paciente el riesgo benéfico. (Vera, DA. 2007).

Por estas razones el estudio busca recabar información útil y necesaria para todas las instituciones ligadas al cuidado de la salud, siendo un aporte a las numerosas investigaciones realizadas con el fin de determinar la mejor vía de culminación del parto.

## II. REVISIÓN DE LITERATURA

### 2.1 ANTECEDENTES DE LA INVESTIGACIÓN

Wayne W. Daniel (2006) indica que “la distribución binomial es una de las distribuciones utilizadas más ampliamente en estadística aplicada. La distribución se deriva de un procedimiento conocido como ensayo de Bernoulli, nombrado así en honor del matemático suizo James Bernoulli (1654-1705), quien realizó contribuciones importantes en el campo de la probabilidad, incluyendo, particularmente, la distribución binomial. Cuando en un proceso aleatorio experimento, llamado ensayo, puede ocurrir sólo uno de dos resultados mutuamente excluyentes, como vida o muerte, enfermo o sano, masculino o femenino, el ensayo se llama ensayo de Bernoulli”.

Según Joseph F. Hair, Jr. Rolph E. Anderson, Ronald L. Tatham y William C. Black (1999), “los casos atípicos son observaciones con una combinación única de características identificables que les diferencia claramente de las otras observaciones. Los casos atípicos problemáticos pueden distorsionar seriamente los test estadísticos”.

De acuerdo con Dallas E. Johnson (2000), “los investigadores deben aplicar sus propias opiniones expertas acerca de las poblaciones de las que se han tomado las muestras, así como las de otros expertos, cuando sea posible, para tomar decisiones subjetivas, pero bien informadas, sobre lo que hay que hacer con los datos outliers que se encuentren en los datos que se están analizando”.

Según Ezequiel Uriel Jiménez y Joaquín Aldas Manzano (2005), “el problema de la multicolinealidad hace referencia, en concreto, a la existencia de relaciones

aproximadamente lineales entre los regresores del modelo, cuando los estimadores obtenidos y la precisión de éstos se ven seriamente afectados”.

Dallas E. Johnson (2000) indica que “los métodos de regresión logística se pueden considerar en situaciones en las que las variables predictoras no estén distribuidas normalmente y en las que algunas o todas esas variables sean discretas o categóricas”.

Cesar Pérez López (2004) señala que “un modelo de regresión con variable dependiente binomial (modelo logístico o modelo de regresión logística) será un modelo que permita estudiar si dicha variable discreta depende o no, de otra u otras variables. Si una variable binomial de parámetros  $p$  es independiente de otra variable  $X$ , se cumple  $(p/X = x) = p$ , para cualquier valor  $x$  de la variable  $X$ . Por consiguiente, un modelo de regresión con variable dependiente binomial y una única variable independiente  $X$  se materializa en una función en la que  $p$  aparece dependiendo de  $X$  y de unos coeficientes cuya investigación permite abordar la relación de dependencia”.

Luis Carlos Silva Aycaguer e Isabel María Barroso Utra (2005) indican que “la regresión logística múltiple es una generalización del modelo simple, relaciona la probabilidad de que ocurra un determinado suceso independiente denotado por el vector  $X'=(x_1, \dots, x_p)$  con probabilidad condicional  $P(Y=1/X)$  en función de  $p$  variables independientes que pueden ser cuantitativas, cualitativas o ambos tipos según sea el tipo de diseño de estudio”.

De acuerdo con Pedret (2000), “la interpretación de un parámetro  $\beta_p$  no se efectuará directamente sino que se hará a partir de un ratio de probabilidad derivado (*odds ratio*). Dicha interpretación también dependerá del tipo de variable explicativa de que se trate: dicotómica o de más de dos categorías”.

R.Clifford Blair y Richard A. Taylor (2008) señalan que “las pruebas diseñadas para establecer la presencia o ausencia de alguna enfermedad rara vez son perfectas. El bueno o mal desempeño de una prueba en este aspecto puede evaluarse a través del cálculo de su sensibilidad, especificidad, valor predictivo positivo y valor predictivo negativo”.

R.Clifford Blair y Richard A. Taylor (2008), señalan que “la razón de probabilidad no es tan intuitiva como la tasa de riesgo, pero tiene la ventaja de ser aplicable en un rango más amplio de diseños de estudio. Como ocurre con RR, una razón de probabilidad con valor de uno implica que ambos grupos están al mismo nivel de riesgo. Cuando la prevalencia de enfermedad es relativamente pequeña, la razón de probabilidad provee un buen cálculo de la tasa de riesgo”.

## **2.2 BASE TEÓRICA**

### **2.2.1 DISTRIBUCIÓN BINOMIAL**

La distribución binomial es una de las distribuciones utilizadas más ampliamente en estadística aplicada. La distribución se deriva de un procedimiento conocido como ensayo de Bernoulli, nombrado así en honor del matemático suizo James Bernoulli (1654-1705), quien realizó contribuciones importantes en el campo de la probabilidad, incluyendo, particularmente, la distribución binomial. Cuando en un proceso aleatorio experimento, llamado ensayo, puede ocurrir sólo uno de dos resultados mutuamente excluyentes, como vida o muerte, enfermo o sano, masculino o femenino, el ensayo se llama ensayo de Bernoulli. (Wayne, D. 2006)

**2.2.1.1. Proceso de Bernoulli:** Una secuencia de ensayos de Bernoulli forma un *proceso de Bernoulli*, si se cumplen las siguientes condiciones:

1. En cada ensayo ocurre uno de dos posibles resultados, mutuamente excluyentes. Uno de los posibles resultados se denota (arbitrariamente) como un éxito y el otro, como fracaso.
2. La probabilidad de un éxito, denotado por  $p$ , permanece constante de un ensayo a otro, y la probabilidad de fracaso,  $1-p$ , se denota con  $q$ .
3. Los ensayos son independientes, es decir el resultado de algún ensayo en particular no es afectado por el resultado de cualquier otro ensayo.

### 2.2.2 CONCEPTO DE “ODDS RATIO” (OR) O RAZÓN DE PROBABILIDAD

En algunos entornos de investigación, la tasa de riesgo no aporta una comparación significativa del grupo expuesto y del no expuesto. En tales casos, se utiliza la razón de probabilidad para hacer comparaciones. Para comprender la razón de probabilidad, primero es necesario entender las posibilidades.

Las posibilidades de que ocurra un evento están indicadas por el cociente de la probabilidad de que el evento suceda y la probabilidad de que el evento no ocurra. Si se calculan las posibilidades de dos grupos y se convierten en un cociente, el resultado es, naturalmente, una razón de probabilidad. Si, usted compara un grupo expuesto y un grupo no expuesto, la razón de probabilidad (*odds ratio*, *OR*) sería:

$$OR = \frac{\frac{P(D|E)}{P(\bar{D}|E)}}{\frac{P(D|\bar{E})}{P(\bar{D}|\bar{E})}}$$

Que se simplifica a:

$$OR = \frac{P(D|E) P(\bar{D}|\bar{E})}{P(\bar{D}|E) P(D|\bar{E})}$$

Donde  $E$  y  $\bar{E}$  representan la exposición y la no exposición, respectivamente.

$D$  y  $\bar{D}$  representan la presencia y la no presencia de la característica en estudio (cesárea), respectivamente.

La razón de probabilidad no es tan intuitiva como la tasa de riesgo, pero tiene la ventaja de ser aplicable en un rango más amplio de diseños de estudio. Como ocurre con RR (Riesgo Relativo), una razón de probabilidad con valor de uno implica que ambos grupos están al mismo nivel de riesgo. Cuando la prevalencia de enfermedad es relativamente pequeña, la razón de probabilidad provee un buen cálculo de la tasa de riesgo. (Clifford, R; Taylor, R. 2008).

### 2.2.3 ANÁLISIS DE DATOS

El examen de los datos es un paso necesario, que lleva tiempo, y que habitualmente se descuida por parte de los analistas de datos. Un análisis cuidadoso de los datos conduce a una mejor predicción y a una evaluación más precisa de la dimensionalidad.

Los casos atípicos son observaciones con una combinación única de características identificables que les diferencia claramente de las otras observaciones. Los casos atípicos no pueden ser caracterizados categóricamente como benéficos o problemáticos sino que deben ser contemplados en el contexto del análisis y deben ser evaluados por los tipos de información que pueden proporcionar. Cuando son benéficos, los casos atípicos, aunque diferentes a la mayor parte de la muestra, pueden ser indicativos de las características segmento de la población que se llegarían a descubrir en el curso normal del análisis. Por

el contrario, los casos atípicos problemáticos no son representativos de la población y están en contra de los objetivos del análisis. Los casos atípicos problemáticos pueden distorsionar seriamente los test estadísticos. (Hair, J; Anderson Junior, R; Tatham, R; Black, W; 1999).

Cuando se encuentran esos datos, el investigador debe determinar, si es posible, si se cometieron errores de registro o en la entrada de datos. Si ocurrieron esos errores y pueden corregirse, entonces debe hacerse. Si se cometieron errores de registro o en la entrada de datos, pero no pueden corregirse, entonces esos datos deben eliminarse de todo análisis multivariado que deba efectuarse.

Saber cómo trabajar con los datos outliers cuando no hay errores aparentes de registro o en la entrada de datos es un problema difícil que no tiene una solución siempre aceptable. Ningún estadístico o técnica estadística puede decir a algún experimentador lo que tiene que hacer con los datos outliers que se encuentran en un archivo. Los investigadores deben aplicar sus propias opiniones expertas acerca de las poblaciones de las que se han tomado las muestras, así como las de otros expertos, cuando sea posible, para tomar decisiones subjetivas, pero bien informadas, sobre lo que hay que hacer con los datos outliers que se encuentren en los datos que se están analizando. (Johnson, D. 2000).

Los investigadores deben decidir por sí mismos si creen que una observación en particular los está conduciendo hacia la “verdad” o los está alejando de ella. Si creen que la observación los está llevando hacia la verdad, entonces la observación debe incluirse en los análisis estadísticos que se van a realizar; si creen que la observación los está alejando de la verdad, entonces esa observación debe eliminarse de todos los análisis estadísticos. (Johnson, D. 2000).

## 2.2.4 ANÁLISIS DESCRIPTIVO

Se construyeron tablas personalizadas entre cada una de las variables independientes continuas y la variable dependiente con el fin de obtener estadísticos descriptivos como la media, desviación estándar, máximos y mínimos de las variables independientes en función a la variable respuesta. Así mismo se realizó un análisis de dispersión con el diagrama de Cajas Boxplot donde se pudo visualizar la no presencia de puntos extremadamente influyentes en el estudio. También se realizó un gráfico de barras para las gestantes que tuvieron complicaciones durante el embarazo, en función a la vía de culminación del parto de la gestante (normal o por cesárea).

## 2.2.5 ANÁLISIS DE MULTICOLINEALIDAD

El problema de la multicolinealidad hace referencia, en concreto, a la existencia de relaciones aproximadamente lineales entre los regresores del modelo, cuando los estimadores obtenidos y la precisión de éstos se ven seriamente afectados. (Jiménez, E; Manzano, J. 2005).

Cuando existe una relación lineal fuerte entre los predictores de una regresión lineal se reduce la precisión de los coeficientes estimados, es decir se incrementa su varianza. La varianza estimada para un coeficiente  $B_k$  en regresión lineal es:

$$\hat{Var}(B_k) = \frac{\hat{\sigma}^2}{(n-1)S_k^2} * \frac{1}{1-R_k^2}$$

dónde  $\hat{\sigma}^2$  es la estimación de la varianza de los errores,  $S_k^2$  es la varianza muestral del predictor  $X_k$  y  $1/1-R_k^2$  es el factor de inflación de la varianza (VIF) de  $B_k$ .  $R_k^2$  es el coeficiente de correlación múltiple de la regresión de  $X_k$  sobre los otros predictores del modelo. Así, si una variable  $X_k$  está muy poco correlacionada con el resto, el valor del VIF es próximo a 1 (Cañadas, JL. 2013). Se denomina tolerancia al complemento de  $R_k^2$ ,  $1-R_k^2$ , cuando existe estrecha relación entre covariables la tolerancia tiende a ser 0, y por

tanto VIF tiende al infinito. Como regla general debería preocupar tolerancias menores de 0,1 y VIF mayores de 10, en consecuencia ambos indicadores ayudan a detectar la presencia de multicolinealidad.

Según Joseph F. Hair, Jr. Rolph E. Anderson, Ronald L. Tatham y William C. Black (1999) para evitar el problema de multicolinealidad en la regresión, usted debe examinar de antemano las correlaciones entre variables independientes. Si son demasiado altas ( $>0.50$ ), entonces debe considerar la posibilidad de eliminar una o más variables de las variables sumamente correlacionadas. Para hacerlo, puede elegir una de sus variables como variable representativa de las variables sumamente correlacionadas y utilizar solo esa en su modelo de regresión. No obstante hay otras opciones que pueden considerarse y que superan el problema de multicolinealidad. Una segunda opción es combinar las variables en una variable compuesta sumada. La tercera solución es aplicar un análisis factorial al conjunto de variables independientes.

## **2.2.6 REGRESIÓN LOGÍSTICA**

Los métodos de regresión logística se pueden considerar en situaciones en las que las variables predictoras no estén distribuidas normalmente y en las que algunas o todas esas variables sean discretas o categóricas. (Johnson, D. 2000).

La regresión logística es semejante a la regresión múltiple; la diferencia principal es que, en la logística, la variable dependiente suele ser binaria, es decir toma sólo dos valores posibles, en tanto que en la múltiple, esa variable dependiente es continua. En la regresión logística se intenta modelar la probabilidad de que  $Y$  sea igual a 1, dados valores observados de las variables predictoras.

Para emplear el modelo de regresión logística en nuestro estudio con fines de discriminación, sencillamente se calcula una estimación de la probabilidad de que  $Y$  sea

igual a 1, dados los valores de las variables predictoras para una gestante seleccionada. Si esta probabilidad estimada es mayor que 0.3, la gestante tendrá una vía de culminación de parto por cesárea; si la probabilidad estimada es menor que 0.3, la gestante tendrá una vía de culminación de parto normal.

El valor de punto de corte es 0.5 (por defecto en SPSS), cuando la probabilidad predicha de cada caso se compara cuando SPSS asigna casos a los grupos predichos. Cuando el modelo tiene éxito prediciendo la pertenencia a un grupo, los casos para lo cual el evento de interés a ocurrido deben estar a la derecha de 0.5, caso contrario a la izquierda de 0.5. Es decir, para determinar cómo clasifica el modelo a cada observación se elige un punto de corte, si la probabilidad predicha por el modelo es mayor que el punto de corte se clasifica como éxito y si es menor como fracaso. (SPSS, 2003, Análisis Estadístico Avanzado).

Ahora es posible aumentar la precisión predictiva cambiando el punto de corte a un valor diferente de 0.5. La justificación teórica del modelo está en que no siempre es adecuado utilizar el punto de corte 0.5 si contamos con datos reales no balanceados. Por el contrario sería conveniente que el punto de corte seleccionado trate de equilibrar la relación entre errores de clasificación de un tipo o de otro, hacia puntos intermedios normalmente. (SPSS, 2003, Análisis Estadístico Avanzado.).

Aunque muchas veces ese punto de corte se toma como 0.5, es más apropiado tomar como punto de corte la proporción de valores  $Y = 1$  en la muestra, e incluso más aconsejable es probar distintos puntos de corte y quedarnos con el que maximice la tasa de clasificaciones correctas. Por ejemplo si elegimos como punto de corte  $p = 0,5$  podemos calcular la tasa de clasificaciones correctas como los individuos correctamente clasificados entre el número total de individuos. El modelo clasifica correctamente un porcentaje de los individuos asociado a este punto de corte y si cambiamos ese punto de corte podría mejorar el porcentaje de individuos correctamente clasificados. (Salazar, A. 2011).

El motivo de la elección del punto de corte 0.3 es que nuestra data tiene en su mayoría a gestantes que tuvieron un parto normal, lo cual genera un sobreestimación de la tasa de clasificación hacia la vía de culminación de parto normal; es por ello la razón del cambio de punto de corte de 0.5 a 0.3.

Muchos investigadores en ciencias de la salud están motivados por el deseo de describir, comprender y hacer uso de las relaciones entre las variables independientes y una variable dependiente (o resultado) que sea discreta. Son particularmente abundantes las circunstancias en donde la variable que puede tomar sólo uno de dos valores mutuamente excluyentes. Estos valores, por lo general, se codifican como  $Y = 1$  para éxito y  $Y = 0$  para fracaso. Las variables dicotómicas incluyen aquellas cuyos dos posibles valores son categorías como: muerte o no-muerte; curado o no curado; enfermedad ocurrida o enfermedad no ocurrida; fumador o no fumador.

### **2.2.6.1 Supuestos de la Regresión Logística**

- Las variables independientes sean intervalos, de razón o dicotómicas.
- Que no haya auto correlación.
- El valor esperado del error es cero.
- Que no haya correlación entre el error y las variables independientes.
- Que no haya multicolinealidad entre las variables independientes.

### **2.2.7 MODELO MATEMÁTICO**

Un modelo de regresión con variable dependiente binomial (modelo logístico o modelo de regresión logística) será un modelo que permita estudiar si dicha variable discreta depende o no, de otra u otras variables. Si una variable binomial de parámetros  $p$  es independiente de otra variable  $X$ , se cumple  $(p/X = x) = p$ , para cualquier valor  $x$  de la variable  $X$ . Por consiguiente, un modelo de regresión con variable dependiente binomial y

una única variable independiente  $X$  se materializa en una función en la que  $p$  aparece dependiendo de  $X$  y de unos coeficientes cuya investigación permite abordar la relación de dependencia. (Pérez, C. 2004)

Para una única variable independiente  $X$ , el modelo de regresión logística toma la forma  $\ln p/q|X=x = \beta_0 + \beta_1 X$ , o de forma simplificada:  $\ln p/q = \beta_0 + \beta_1 X$ , donde  $\ln$  significa logaritmo neperiano,  $\beta_0$  y  $\beta_1$  son constantes y  $X$  una variable que puede ser aleatoria o no, continúa o discreta. Este modelo se puede fácilmente generalizar para  $k$  variables independientes, dando lugar al *modelo logístico múltiple*, que se expresa como sigue:

$$\ln p/q = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$$

Hemos definido el modelo logístico como el logaritmo de odds para el suceso que representa la variable aleatoria binomial puntual dependiente del modelo.

$$\text{Odds} = \text{Probabilidad de éxito} / \text{Probabilidad de fracaso} = p/q$$

Hay varias razones para plantear el modelo con el logaritmo de odds, en lugar de plantearlo simplemente con la probabilidad de éxito o con el odds. En primer lugar, el campo de variación de  $\ln(p/q)$  es todo el campo real (de  $-\infty$  a  $\infty$ ), mientras que, para  $p$  el campo es sólo de 0 a 1 y para  $p/q$  es de 0 a  $\infty$ . Por lo tanto, con el modelo definido en función del logaritmo de odds no hay que poner restricciones a los coeficientes que complicarían su estimación. Por otro lado, y más importante, en los modelos en función del logaritmo de odds los coeficientes son fáciles interpretables en términos de independencia o asociación entre las variables, como se verá más adelante.

El modelo logístico se puede describir de otras formas equivalentes que para ciertas aplicaciones son más cómodas de manejar. Tenemos:

$$\ln(p/q) = \beta_0 + \beta_1 X \Leftrightarrow \ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X \Leftrightarrow \frac{p}{1-p} = e^{\beta_0 + \beta_1 X} \Leftrightarrow$$

$$p = \frac{e^{\beta_0 + \beta_1 X}}{1 + e^{\beta_0 + \beta_1 X}} \Leftrightarrow p = \frac{1}{1 + e^{-\beta_0 + \beta_1 X}}$$

Estas dos últimas expresiones, si son conocidos los coeficientes  $\beta_0$  y  $\beta_1$ , permiten calcular directamente la probabilidad del proceso binomial para los distintos valores de la variable X. A la función:  $f(z) = \frac{1}{1 + e^{-z}}$  se le denomina función logística.

El modelo de regresión logística modeliza la probabilidad de un proceso binomial como la función logística de una combinación lineal de la variable dependiente.

El modelo de regresión logística múltiple tendrá la expresión:

$$p = \frac{1}{1 + e^{-\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k}}$$

### 2.2.7.1 Estimación de los coeficientes

El método de los mínimos cuadrados, clásico en la estimación de los coeficientes de los modelos de regresión no es aplicable al modelo logístico, ya que dicho método se basa en la normalidad de la variable dependiente, que en este caso no se cumple.

Por otra parte, cuando  $q = 0$ , es imposible calcular  $\ln(p/q)$ . Se tratará entonces de utilizar el método de máxima verosimilitud.

Comenzaremos considerando el caso más simple con una sola variable independiente X, tomamos una muestra de n observaciones  $(y_i, x_i)$  para la variable puntual binomial dependiente Y y para la variable independiente X. La variable Y toma valores  $y_i$  que sólo pueden ser 1 con probabilidad  $p_i$  o 0 con probabilidad  $1-p_i$ . Como  $x_i$  depende de  $p_i$  a través del modelo logístico tenemos:

$$P_i = \frac{e^{\beta_0 + \beta_1 X_i}}{1 + e^{\beta_0 + \beta_1 X_i}}$$

La función de verosimilitud para una variable binomial puntual es:

$$L P_i / y_i = P_i^{y_i} (1 - p_i)^{1 - y_i}$$

Y para n observaciones independientes la función de verosimilitud de la muestra será:

$$L P_1, \dots, P_n / y_1, \dots, y_n = \prod_{i=1}^n P_i^{y_i} (1 - p_i)^{1 - y_i}$$

Y al representar  $p_i$  por el modelo logístico tendremos ya la expresión de la función de verosimilitud para la muestra como función de los parámetros a estimar:

$$\begin{aligned} L P_1, \dots, P_n / y_1, \dots, y_n &= \prod_{i=1}^n \left( \frac{e^{\beta_0 + \beta_1 X_i}}{1 + e^{\beta_0 + \beta_1 X_i}} \right)^{y_i} \left( 1 - \frac{e^{\beta_0 + \beta_1 X_i}}{1 + e^{\beta_0 + \beta_1 X_i}} \right)^{1 - y_i} \\ &= \prod_{i=1}^n \left( \frac{e^{\beta_0 + \beta_1 X_i}}{1 + e^{\beta_0 + \beta_1 X_i}} \right)^{y_i} \left( \frac{1}{1 + e^{\beta_0 + \beta_1 X_i}} \right)^{1 - y_i} = \frac{e^{\beta_0 \sum_{i=1}^n y_i + \beta_1 \sum_{i=1}^n x_i y_i}}{\prod_{i=1}^n (1 + e^{\beta_0 + \beta_1 X_i})} = L(\beta_0, \beta_1) \end{aligned}$$

Y como suele ser usual en máxima verosimilitud, maximizaremos el logaritmo de la función  $L(\beta_0, \beta_1)$  en vez de la función misma. Los parámetros estimados del modelo serán los valores de  $\beta_0$  y  $\beta_1$ , que maximicen la función  $\ln L(\beta_0, \beta_1)$ .

### 2.2.7.2 Pruebas de significancia

Una vez estimado los coeficientes del modelo, se tienen que verificar si el modelo predice de manera adecuada a la variable dependiente en un nuevo individuo relacionado con la muestra, donde los valores de las variables explicativas son las probabilidades estimadas cuando  $P(Y=1)$  y  $P(Y=0)$ .

Para esto se formula y prueba la hipótesis estadística, para determinar si la variable independiente influye significativamente en la probabilidad del suceso del modelo relacionado a la variable del resultado.

$H_0$ : La variable independiente no influye sobre  $p_i$ .

$H_1$ : La variable independiente influye sobre  $p_i$ .

Donde:

$$p_i = p \quad x; \beta_0, \beta_1 = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}$$

El modelo de regresión logística es válido si  $\beta_1$ , es significativamente diferente de cero. Este coeficiente muestral es el estimador de la población que va a explicar la variable dependiente  $p_i$  y poder hacer predicciones con el modelo.

#### a. Evaluación estadística del coeficiente $\beta_1$

El coeficiente  $\beta_1$ , debe ser diferente de cero para que tenga influencia significativa en la variable dependiente en las siguientes hipótesis:

$H_0: \beta_1 = 0$

$H_1: \beta_1 \neq 0$

Estadístico de prueba:

$$t = \frac{\hat{\beta}_1 - \beta_1}{S_{\hat{\beta}_1}} \sim t_{n-p-1}$$

Donde  $S_{\beta_1}$  es el error estándar del coeficiente de regresión logística muestral, mientras que  $\beta_1$  es el coeficiente de regresión logística poblacional y  $\hat{\beta}_1$  es el coeficiente de regresión logístico muestral,  $\hat{\beta}_1 = 0$  por definición de la hipótesis y  $p = 1$  con nivel de significancia  $\alpha$ , entonces será:

$$t = \frac{\beta_1}{S_{\beta_1}} \sim t_{n-2, 1-\alpha/2}$$

La regla de decisión:

Si  $t > t_{(n-2, 1-\alpha/2)}$  se rechaza  $H_0$

## b. Estadístico WALD

Evalúa el coeficiente estimado en la población y se define como un cociente entre el coeficiente y el error estándar del coeficiente en la hipótesis.

$$H_0: \tilde{\beta}_1 = 0$$

$$H_1: \tilde{\beta}_1 \neq 0$$

Estadístico de prueba:

$$WALD = \frac{\hat{\beta}_1}{S_{\hat{\beta}_1}} \sim \chi^2_{1, 1-\alpha}$$

Decisión: si  $WALD > X^2_{(1,1-\alpha)}$  se *Rechaza*  $H_0$  con un nivel de significancia de  $\alpha$  y se concluye que la variable independiente influye en la probabilidad de las características de la variable dependiente. Si la variable independiente es cualitativa los grados de libertad es igual al número de categoría menos 1.

### 2.2.7.3 Bondad de ajuste del modelo

Para evaluar la bondad del modelo se utiliza el logaritmo del cociente de verosimilitud y la prueba de Hosmer-Lemeshow.

#### a. El incremento de estadístico $2\log L$

El estadístico  $-2\log L$  mide los cambios que se producen cuando se agrega o se quita una variable, donde  $L$  es la función de verosimilitud del modelo estudiado, puede oscilar entre 0 y 1, si el modelo se ajusta perfectamente los datos tiene una verosimilitud igual a 1, de allí que  $-2\log L = 0$ . Entonces diremos que el modelo se ajusta bien a los datos si tiene un valor pequeño de  $-2\log L$ , que es el logaritmo de la verosimilitud y se distribuye como una  $X^2$  (Chi-cuadrado), cuando el modelo incluye sólo la constante los grados de libertad es igual al número de casos menos uno ( $n-1$ ), y cuando se incluye la variable independiente sigue una distribución  $X^2$  con  $n-p-1$  grados de libertad, en el modelo de regresión logística simple es  $n-1$ , la diferencia entre esos dos valores de  $-2\log L$  se llama Devianza, esta prueba si la variable  $X_1$  es significativa, se define como: (Campomanes, F. 2010.)

$D = -2\log L$  (verosimilitud del modelo sin la variable/verosimilitud del modelo con la variable).

$$D = -2 \sum_{i=1}^n \left[ y_i \log \left( \frac{\hat{P}_i}{y_i} \right) + 1 - y_i \log \left( \frac{1 - \hat{P}_i}{y_i} \right) \right]$$

Las hipótesis son:

$H_0$ : El modelo ajustado no es significativo

$H_1$ : El modelo ajustado es significativo

Estadístico de prueba:

$$D \sim \chi^2 \text{ con } n - p - 1 \text{ grados de libertad}$$

La regla de decisión:

Si  $D > \chi^2_{n-p-1, 1-\alpha}$  se Rechaza  $H_0$ . El modelo ajustado es significativo.

#### **b. Prueba de Hosmer Lemeshow**

En este contraste se dividen todos los casos en decilas basadas en las probabilidades predichas, es decir la primera decila es en la que se encuentra los casos con probabilidades predichas más elevadas.

En consecuencia, cuando rechazamos la hipótesis para un determinado nivel de significación, esto implicará que el modelo ajustado no es el adecuado para ese nivel de significación.

Evalúa la bondad del modelo construyendo una tabla de contingencia, divide la muestra en aproximadamente  $j$  grupos iguales a partir de las probabilidades estimadas, para comparar las frecuencias observadas con las esperadas en cada uno de estos grupos a través de la prueba  $\chi^2$  con  $j-2$  grados de libertad, en donde  $j$  es el número de grupos formados.

Se calculan los deciles de las probabilidades estimadas  $\hat{p}_i; i = 1, \dots, n$   $D_1, \dots, D_9$  que son los deciles observados divididos en 10 grupos dados por:

$$A_j = \{ i \in \{1, \dots, n\} / P_i \in [D_{j-1}, D_j] \} \quad j = 1, \dots, 10$$

Donde:

$$D_0 = 0, D_{10} = 1$$

Sean:

$n_j =$  Número de casos en  $A_j; j = 1, \dots, 10$

$o_j =$  Número de  $y_i = 1$  en  $A_j; j = 1, \dots, 10$

$$\bar{P}_j = \frac{1}{n_j} \sum_{i \in A_j} \hat{P}_i; j = 1, \dots, 10$$

La hipótesis a contratar:

$H_0:$  El modelo es adecuado

$H_1:$  El modelo no es adecuado

Estadístico de prueba:

$$x^2 = \sum_{j=1}^10 \frac{(o_j - n_j \bar{P}_j)^2}{P_j n_j (1 - P_j)} - x^2_{j-2, 1-\alpha}$$

La regla de decisión:

Si  $x^2 > x^2_{j-2, 1-\alpha}$  se Rechaza  $H_0$ . El modelo no es adecuado.

### c. Prueba de Ómnibus

Comprende tres entradas, la del modelo con el estadístico Chi-Cuadrado es la diferencia entre el valor de  $-2LL$  para el modelo sólo con la constante y el valor de  $-2LL$  para el modelo actual que prueba la hipótesis nula de que los coeficientes para todos los términos del modelo son 0. Si se rechazara la hipótesis nula se concluiría que el grupo de variables mejora la predicción del logaritmo natural de las oportunidades. Se ajusta al modelo los resultados de la prueba de Chi-Cuadrado con los grados de libertad correspondientes en función del número de variables o categorías.

La entrada por bloque, es el cambio en  $-2LL$  entre bloques de entrada sucesivos durante la construcción del modelo. Si como es habitual en la práctica se introducen las variables en un solo bloque, la Chi-Cuadrado del bloque es el mismo que la Chi-Cuadrado del modelo.

La entrada por paso, es la correspondiente al cambio de verosimilitud (de  $-2LL$ ) entre pasos sucesivos en la construcción del modelo.

Se debe señalar que las entradas bloque y paso tienen los mismos valores que los estadísticos del modelo cuando se ingresa solamente en un bloque las variables y no se usa la selección por pasos del modelo, en ese caso, mostraría los resultados del último bloque o paso.

## 2.2.8 MODELO DE REGRESIÓN LOGÍSTICA MÚLTIPLE

Consideremos entonces la variable dicotómica de respuesta  $Y$  (que puede valer 1 o 0 en dependencia de que se presente o no la condición que se estudia) y un conjunto de variables independientes  $X_1, X_2, \dots, X_k$ . Siguiendo el patrón del modelo simple, en el caso múltiple se establece que:

$$\text{Logit}(P) = \ln \frac{P}{1-P} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k ,$$

Dónde:  $\beta_0$  y los  $\beta_i$  son los parámetros (en principio desconocidos) del modelo. Es fácil demostrar que una forma equivalente de expresar este mismo modelo es:

$$P_{Y=1} = \frac{\exp \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k}{1 + \exp \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k}$$

Y también:

$$P_{Y=1} = \frac{1}{1 + \exp -\beta_0 + \beta_1 X_1 - \dots - \beta_k X_k}$$

Es una generalización del modelo simple, relaciona la probabilidad de que ocurra un determinado suceso independiente denotado por el vector  $X'=(x_1, \dots, x_p)$  con probabilidad condicional  $P(Y=1/X)$  en función de  $p$  variables independientes que pueden ser cuantitativas, cualitativas o ambos tipos según sea el tipo de diseño de estudio. (Silva L. y Barroso I., 2005).

### 2.2.8.1 Estimación de los parámetros

Cálculo de los parámetros del modelo de regresión logística (método de máxima verosimilitud).

Supongamos que la variable explicativa  $y_i$  tiene los niveles (1 si la mujer tuvo un parto por cesárea; y 0 si su parto fue normal).

El modelo de regresión logística asume que:

$$y_i = 1, \text{ con Probabilidad } P_i = \frac{e^{\sum_{j=1}^p \beta_j x_{ij}}}{1 + e^{\sum_{j=1}^p \beta_j x_{ij}}}$$

$$y_i = 0, \text{ con Probabilidad } 1 - P_i = \frac{1}{1 + e^{\sum_{j=1}^p \beta_j x_{ij}}}$$

Donde  $x_{i0} = 1$  y  $p$  es el número de variables explicativas.

Los parámetros desconocidos del modelo son:  $\beta_0, \beta_1, \dots, \beta_p$ .

La función de verosimilitud de la muestra contendrá factores del tipo (1) si  $y_i = 1$ , y del tipo (2) si  $y_i = 0$ , con lo que la función de verosimilitud  $L$  será:

$$L(\beta_0, \beta_1, \dots, \beta_p) = \frac{\prod_{i=1}^n e^{y_i \sum_{j=1}^p \beta_j x_{ij}}}{\prod_{i=1}^n \left( 1 + e^{\sum_{j=1}^p \beta_j x_{ij}} \right)} = \frac{e^{\sum_{j=1}^p \beta_j t_j}}{\prod_{i=1}^n \left( 1 + e^{\sum_{j=1}^p \beta_j x_{ij}} \right)}$$

Con  $t_j = \sum_{i=1}^n x_{ij} y_i$

El  $\ln$  con  $L$  queda como:

$$\ln \left[ L(\beta_0, \beta_1, \dots, \beta_p) \right] = \sum_{j=0}^p \beta_j t_j - \sum_{j=1}^p \sum_{i=1}^n \ln \left[ 1 + e^{\sum_{j=0}^p \beta_j x_{ij}} \right]$$

Los  $\beta_j$  que maximizan  $L(\cdot)$ , también maximizarán  $\ln[L(\cdot)]$  y se obtienen resolviendo es siguiente sistema de  $p+1$  ecuaciones no lineales:

$$I_{j_1 j_2}^* = - \sum_{i=0}^n \frac{x_{ij} e^{\sum_{j=0}^p \beta_j x_{ij}}}{1 + e^{\sum_{j=0}^p \beta_j x_{ij}}} = 0 \quad j = 0, 1, \dots, p$$

Para resolver este sistema se usa el método de Newton-Raphson que necesita de la matriz de segundas derivadas  $I_{j_1 j_2}^*$  que viene dada por:

$$I_{j_1 j_2}^* = - \sum_{i=1}^n \frac{x_{ij_1} - x_{ij_2} e^{\sum_{j=0}^p \beta_j x_{ij}}}{1 + e^{\sum_{j=0}^p \beta_j x_{ij}}} = 0 \quad j_1, j_2 = 0, 1, \dots, p$$

Llamando  $I_{j_1 j_2}$  se tiene que  $I_{j_1 j_2}^{-1}$  es la matriz asintótica de varianzas y covarianzas de los  $\beta_j$ 's.

Un IC  $(1-\alpha)$  para los  $\beta_j$  se puede obtener usando:

$$\beta_j \pm Z(1 - \frac{\alpha}{2}) \sqrt{I_{jj}^{-1}}$$

(Cortinas, P; Juez, P; Muñoz, A; Santos, J. 2003).

### a. Interpretación de los parámetros $\beta$

En el caso de una regresión lineal, el valor obtenido en la estimación de un parámetro  $\beta_p$  nos dice cuál va a ser la variación en la variable a explicar cuando la variable  $X_p$  asociada experimenta un incremento en una unidad, siempre y cuando las demás variables explicativas se mantengan constantes. En este contexto podemos definir el parámetro  $\beta_p$  como sigue:

$$\beta_p = f(x_p + 1) - f(x_p)$$

Siendo:

$$f(x_p + 1) = \beta_0 + \beta_1 x_1 + \dots + \beta_p (x_p + 1) + \dots + \beta_p x_p$$

$$f(x_p) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p + \dots + \beta_p x_p$$

La interpretación de un parámetro  $\beta_p$  en un modelo LOGIT no es tan evidente. Teniendo en cuenta la relación que existe entre la variable a explicar y las variables explicativas, un parámetro  $\beta_p$  quedaría definido de la siguiente forma:

$$\beta_p = f(x_p + 1) - f(x_p)$$

Siendo:

$$f(x_p + 1) = \ln \left( \frac{\prod x_p + 1}{1 - \prod x_p + 1} \right)$$

$$f(x_p) = \ln \left( \frac{\prod x_p}{1 - \prod x_p} \right)$$

$\beta_p$  representa, pues, el cambio que experimentaría la relación entre la probabilidad de que un fenómeno ocurra y la probabilidad de que no ocurra, cuando la variable  $X_p$  aumente en una unidad y las demás variables explicativas se mantengan constantes.

La interpretación de un parámetro  $\beta_p$  no se efectuará directamente sino que se hará a partir de un ratio de probabilidad derivado (*odds ratio*). Dicha interpretación también dependerá del tipo de variable explicativa de que se trate: dicotómica o de más de dos categorías. (Pedret, 2000).

*Deducción del intervalo de confianza para una probabilidad puntual obtenida mediante regresión logística.*

Consideremos la función logística multivariada con parámetros  $\beta_0, \beta_1, \dots, \beta_k$ :  $p(t) = \left[ 1 + e^{\mu^* t} \right]^{-1}$  donde  $\mu^* t = -\beta_0 - \sum_{i=1}^k t_i \beta_i$ ;  $t = t_1, t_2, \dots, t_k$

Supongamos que  $\beta_0^*, \beta_1^*, \dots, \beta_k^*$  es el estimador máximo verosímil de  $\beta_0, \beta_1, \dots, \beta_k$  basado en una muestra aleatoria de tamaño suficientemente grande como para admitir que dicho vector de estimaciones sigue una distribución normal multivariada.

Se cumple entonces que, asintóticamente, para todo vector  $t$  fijo, la variable aleatoria  $\mu^* t = -\beta_0^* - \sum_{i=1}^k t_i \beta_i^*$  tienen distribución normal con media  $m^*(t)$  y varianzas iguales a:

$$\sigma^2 t = \text{Var } \beta_0^* + \sum_{i=1}^k t_i^2 \text{Var } \beta_i^* + 2 \sum_{i=1}^k t_i \text{Cov } \beta_0^*, \beta_i^* + 2 \sum_{1 \leq i < j} t_i t_j \text{Cov } \beta_i^*, \beta_j^*$$

así como que

$$s^2 t = \hat{\text{V ar}} \beta_0^* + \sum_{i=1}^k t_i^2 \hat{\text{V ar}} \beta_i^* + 2 \sum_{i=1}^k t_i \hat{\text{C ov}} \beta_0^*, \beta_i^* + 2 \sum_{1 \leq i < j} t_i t_j \hat{\text{C ov}} \beta_i^*, \beta_j^*$$

es un estimador insesgado de  $\sigma^2$ .

Teniendo ahora en cuenta la expresión del intervalo de confianza para la media de la distribución normal:

$$\Pr \left[ \mu^* t - Z_{\alpha} s t \leq \mu t \leq \mu^* t + Z_{\alpha} s t \right] = 1 - \alpha$$

donde  $\alpha$  es un número tal que  $0 < \alpha < 1$  y  $Z_{\alpha}$  es el percentil  $\left(1 - \frac{\alpha}{2}\right)$  de la distribución  $t$  de

Student con  $n-1$  grados de libertad, se deduce que:

$$\Pr \left[ e^{\mu^* t + Z_{\alpha} s t} \leq e^{\mu t} \leq e^{\mu^* t - Z_{\alpha} s t} \right] = 1 - \alpha$$

y, finalmente, que:

$$\Pr \left[ \frac{1}{e^{\mu^* t + Z_{\alpha} s t}} \leq p t \leq \frac{1}{1 + e^{\mu^* t - Z_{\alpha} s t}} \right] = 1 - \alpha$$

### 2.2.8.2. Pruebas de significancia

Para contrastar la significatividad global en los modelos logit o probit no se pueden construir estadísticos  $F$  que tengan distribución  $F$  de Snedecor, debido a que son modelos no lineales. En su lugar usamos el estadístico de razón de verosimilitud (RV).

Como en el caso univariante, se prueba la significancia de las variables independientes del modelo mediante la prueba de verosimilitud con la significancia de los  $K+1$  parámetros, para determinar si las variables independientes influyen significativamente en la probabilidad del suceso del modelo relacionado a la variable en estudio del siguiente modo:

Concretamente, para contrastar la significatividad global del modelo las hipótesis nula y alternativa son las siguientes:

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0$$

$$H_1 : \text{No } H_0$$

mientras que la RV se calcula de la siguiente forma:

$$RV_0 = -2 \ln L_0 - \ln L$$

donde  $\ln L$  es el logaritmo de la función de verosimilitud que se ha obtenido al estimar el modelo completo, mientras que  $\ln L_0$  es el logaritmo de la función de verosimilitud que se ha obtenido al estimar el modelo con sólo el término independiente. Se verifica que este estadístico se distribuye asintóticamente, es decir cuando el tamaño de la muestra es grande, aunque estrictamente es cuando el tamaño de la muestra tiende a  $\infty$  como una Chi cuadrado con  $k-1$  grados de libertad, siendo  $k$  el número de regresores del modelo incluido el término independiente. Por tanto,

$$RV_0 : X_{k-1}^2$$

En general, cuando se desee contrastar la significatividad de un subconjunto de coeficientes, el estadístico a aplicar es el siguiente:

$$RV_0 = -2 \ln L_R - \ln L$$

Donde  $\ln L_R$  es el logaritmo de la función de verosimilitud que se obtiene al estimar el modelo en el que están excluidos los  $r$  términos cuya significatividad se desea contrastar, Bajo la hipótesis nula el estadístico anterior se distribuye como:

$$RV_0 : X_r^2 \text{ grados de libertad}$$

Una vez encontrado el mejor conjunto de variantes explicativas que predicen la variable  $Y$ , se debe evaluar mediante Wald cada coeficiente para determinar cuál o cuáles ingresan al modelo.

### a. Estadístico WALD

Existe otra vía para probar, en la regresión logística, si una variable independiente hace un aporte estadísticamente significativo a la “explicación” de la variable de respuesta sin tener que usar la razón de verosimilitudes. Este método es frecuentemente llamado “prueba de Wald”. Esta prueba permite contrastar si un coeficiente  $\beta$  es significativamente diferente de 0.

El estadígrafo para realizar la *prueba de Wald* se obtiene dividiendo la estimación del coeficiente de interés por su error estándar:

$$Z_{wald} = \frac{\hat{B}}{se \hat{B}}$$

Donde se  $(\hat{B})$  es el error estándar de  $\hat{B}$ .

El estadígrafo  $Z_{wald}$  sigue una distribución normal estándar. Por lo general los paquetes estadísticos proveen este último y su valor de  $p$  asociado.

La razón de verosimilitudes y la *prueba de Wald* brindan resultados similares en muestras grandes, por ello en estudios con suficiente tamaño de muestra no importa cuál de las dos pruebas sea la usada. Sin embargo, en muestras no muy grandes, dichas pruebas producen diferentes resultados. Por lo general se recomienda el uso de la razón de verosimilitudes.

Evalúa la significancia de los coeficientes, se define como el vector matriz de los coeficientes estimados, según hipótesis:

$$H_0: B_i = 0$$

$$H_1: \text{Para algún } B_i \neq 0; i = 1, \dots, k$$

Como se puede apreciar, la hipótesis viene a ser una extensión de la regresión logística simple, puesto que se trabajará con  $k$  parámetros debido a que nuestro trabajo comprende más de 1 variable con su respectivo coeficiente  $B$ .

La regla de decisión será con un nivel de significación fijado  $\alpha$ , se concluye que la variable independiente influye en la probabilidad del suceso.

### **b. Pruebas de bondad de ajuste**

Para evaluar la bondad de ajuste del modelo se utiliza la probabilidad de Hosmer-Lemeshow, que consiste en calcular para cada observación del conjunto de datos las probabilidades esperadas y se compara con las frecuencias observadas mediante una prueba  $X^2$  con  $j-2$  grados de libertad, don  $j$  es el número de grupos formados como se explicó en el modelo simple. El modelo se ajusta bien si no hay evidencia para rechazar la hipótesis nula.

## **2.3. VALIDACIÓN CRUZADA**

Normalmente se utiliza cuando no se dispone de otro conjunto de datos para poder validar el modelo. Una alternativa, si se disponen de las suficientes observaciones es dividir la muestra en dos submuestras, de forma que una se utilice para ajustar el modelo y la otra para evaluación del mismo.

La validación cruzada parte de esta última alternativa, considerando dos variantes. En la primera, la llamada K-Fold cross-validation, se divide la muestra en  $K$  submuestras, de forma que se Utilizan  $K-1$  para estimar el modelo y la restante como submuestra de evaluación, este proceso se repite  $K$  veces, de forma que cada submuestra es utilizada una vez para evaluar el modelo y  $K-1$  veces para el ajuste. Como medida de validación se suele utilizar la media de las tasas de clasificaciones correctas, o su complementario que sería la

tasa de clasificaciones incorrectas. A la segunda aproximación se la conoce como leave-one-out cross-validation (LOOCV) y consiste en tomar  $K$  igual al número de observaciones, este método exige un mayor coste computacional, debido a que hay que ajustar tantos modelos como observaciones. (Cañadas, JL. 2013)

## 2.4 MEDIDAS DE LA BONDAD DEL AJUSTE

Vamos a examinar distintos tipos de medidas para la bondad del ajuste, en concreto examinaremos las siguientes: coeficientes Pseudo  $R^2$ , tablas de clasificación y el contraste de bondad del ajuste de Hosmer y Lemeshow.

### 2.4.1 EVALUACIÓN DE LA BONDAD DEL AJUSTE

Una vía relativamente simple de evaluar la calidad del ajuste en este caso es aquella cuyos pasos se delinearán a continuación.

#### 2.4.1.1 Coeficiente de determinación Pseudo $R^2$

Para evaluar la capacidad explicativa de estos modelos se pueden utilizar medidas de bondad del ajuste y contrastes de significatividad.

En los modelos logit y probit no tiene aplicación el coeficiente de determinación utilizado en la regresión. En su lugar se utilizan los denominados Pseudos  $R^2$ . El primero que apareció fue el debido a Mac Fadden, que lo definió de la siguiente forma:

$$PR_{MacFadden}^2 = 1 - \frac{\ln L}{\ln L_0}$$

Donde  $\ln L$  y  $\ln L_0$ , tienen el significado que hemos visto previamente.

En 1989 Cox y Snell propusieron otra medida de bondad del ajuste que tenía en cuenta el tamaño de la muestra, y cuya expresión es la siguiente:

$$PR_{Cox-Snell}^2 = 1 - e^{-\frac{2}{N} \ln L_0 - \ln L} = 1 - e^{-\frac{RV_0}{N}}$$

La medida anterior no puede llegar, aunque el ajuste fuera perfecto, a tomar el valor 1. Por ello, Nagelkerke en 1991 propuso una medida alternativa que se obtiene dividiendo la medida de Cox y Snell por el valor máximo que puede tomar en caso de que el ajuste fuera perfecto. Su expresión, por tanto, es la siguiente:

$$PR_{Nagelkerke}^2 = \frac{PR_{Cox-Snell}^2}{PR_{MAX Cox-Snell}^2} = \frac{PR_{Cox-Snell}^2}{1 - e^{-\frac{2}{N} \ln L_0}}$$

Las tres medidas anteriores suelen aparecer en muchos de los paquetes estadísticos.

#### 2.4.1.2 Tablas de clasificación (o matriz de confusión)

Para determinar la bondad del ajuste se suele utilizar la tabla de clasificación que muestra la distribución de valores observados y pronosticados. Los valores pronosticados se obtienen a partir del modelo estimado. Si se trata de un modelo logit, los valores pronosticados se obtienen a partir de las probabilidades estimadas, que adoptan la siguiente forma:

$$\hat{P}_i = \frac{1}{1 + e^{-Z_i}} = A\left(\hat{Z}_i\right) \quad ; \text{ Donde } Z_i = x_i^1 \beta$$

Por tanto, para cada observación tenemos una probabilidad predicha. Ahora bien, esta probabilidad la tenemos que convertir en una variable discreta que tome los valores 0 y 1. Para ello tenemos que tomar decisiones de este tipo:

$$\hat{Y}_i = 1 \text{ si se verifica que } \hat{P}_i = P^*$$

Donde  $P^*$  es un valor predeterminado. Una regla razonable es hacer  $P^* = 0,5$ .

Una vez realizadas las estimaciones de las  $\hat{Y}_i$  podemos construir una tabla de doble entrada que permita comparar los valores observados con los pronosticados.

#### **2.4.1.2.1 Sensibilidad, especificidad y conceptos relacionados**

Las pruebas diseñadas para establecer la presencia o ausencia de alguna enfermedad rara vez son perfectas. Por ejemplo, nos gustaría que una prueba médica para establecer la presencia o ausencia de alguna enfermedad en particular fuera positiva para quienes tengan la enfermedad y negativa para quienes no la tengan. Por desgracia, en algunas ocasiones una persona enferma recibe un resultado negativo o una persona sana obtiene un resultado positivo. (Clifford, R; Taylor, R; 2008)

El bueno o mal desempeño de una prueba en este aspecto puede evaluarse a través del cálculo de su sensibilidad, especificidad, valor predictivo positivo y valor predictivo negativo. (Clifford, R; Taylor, R; 2008)

Se utiliza  $D$  para indicar enfermedad y  $\bar{D}$  para representar la ausencia de enfermedad.

La sensibilidad de una prueba es la probabilidad de que una persona con la enfermedad de un resultado positivo o:

$$\text{Sensibilidad} = P (+/D)$$

La especificidad de una prueba es la probabilidad de que una persona que no tiene la enfermedad (sana) dé un resultado negativo o

$$\text{Especificidad} = P (-/\bar{D})$$

El valor predictivo positivo es la probabilidad de que una persona que da positivo tenga la enfermedad o

$$\text{PPV} = P (D/+)$$

El valor predictivo negativo es la probabilidad de que una persona que da negativo no tenga la enfermedad o

$$\text{NPV} = P (\bar{D}/-)$$

Finalmente, la prevalencia es simplemente la probabilidad de enfermedad o

$$\text{Prevalencia} = P (D)$$

### 2.4.1.3 Contraste de bondad del ajuste de Hosmer y Lemeshow

En este contraste se dividen todos los casos en deciles basados en las probabilidades predichas, es decir la primera decila es en la que se encuentra los casos con probabilidades predichas más elevadas. A partir de estas probabilidades predichas y de los datos observados se construye el siguiente estadístico:

$$HL = \sum_{i=1}^{10} \frac{\left[ O_i - N_i \bar{\pi}_i \right]^2}{N_i \bar{\pi}_i (1 - \bar{\pi}_i)}$$

Donde:

$O_i$  : es el número de unos en la decila  $i$ -ésima.

$\bar{\pi}_i$  : es la media de las probabilidades predichas en la decila  $i$ -ésima.

$N_i$  : es el número de observaciones en la decila  $i$ -ésima.

Bajo la hipótesis nula de que no existen diferencias entre los valores observados y predichos, se tiene que HL se distribuye como una Chi-cuadrado con ocho grados de libertad, es decir

$$HL : X_8^2$$

En consecuencia, cuando rechazamos la hipótesis para un determinado nivel de significación, esto implicará que el modelo ajustado no es el adecuado para ese nivel de significación.

## 2.5 ANÁLISIS DE LA CURVA COR

Otra forma de evaluar un modelo es a través de la curva COR. La curva COR representa en un gráfico la fracción de falsos positivos definida como  $FP/(FP + VN)$ , frente a la fracción de verdaderos positivos definida como  $VP/(VP + FN)$ . Así, se puede evaluar un modelo según si el interés sea maximizar la fracción de verdaderos positivos, minimizar la

fracción de falsos positivos u obtener la mejor tasa de clasificación correcta. (Cañadas, JL. 2013)

**VP:** Verdadero Positivo --Predicciones Correctas

**VN:** Verdadero Negativo --Predicciones Correctas

**FP:** Falso Positivo --Predicciones Incorrectas

**FN:** Falso Negativo --Predicciones Incorrectas

Otra forma de evaluar el modelo, es mediante el análisis de la curva COR, comparándola con las curvas COR de otros modelos.

La curva COR es una gráfica que permite evaluar la capacidad del modelo para discriminar. El área bajo la curva COR representa la probabilidad de que un individuo enfermo elegido al azar tenga mayor probabilidad estimada de padecer la enfermedad que un individuo no enfermo elegido también al azar. Por tanto, lo deseable es que esta medida sea lo más alta posible, considerándose que el modelo es preciso y tiene alta capacidad de discriminación cuando el área es al menos 0.7. (Salazar, A. 2011).

## **III. MATERIALES Y MÉTODOS**

### **3.1. LUGAR**

Los datos se obtuvieron en el centro médico Asociación Taller de los Niños ubicado en el distrito de San Juan de Lurigancho, durante el periodo comprendido entre los meses de Junio 2010 hasta Octubre 2012.

### **3.2 MATERIALES**

- Pc-Portátil
- Útiles de escritorio
- Movilidad local
- Programas: Excel 2013, Word 2013, Spss18.

### **3.3 METODOLOGÍA DE LA INVESTIGACIÓN**

#### **3.3.1 DISEÑO DE INVESTIGACIÓN**

Se cuenta con datos de las gestantes con una culminación de parto normal o por cesárea, que asistieron al centro médico de la Asociación Taller de los Niños en el periodo de junio del 2010 hasta octubre del 2012.

### **3.3.2 POBLACIÓN Y MUESTRA**

#### **a. Población**

Está constituida por todos los partos que se registran en el centro médico de la Asociación Taller de los Niños a partir de junio del 2010.

#### **b. Muestra**

La muestra está constituida por 1950 partos observados en el centro médico de la Asociación Taller de los Niños.

### **3.3.3 TIPO DE INVESTIGACIÓN**

La presente investigación es de tipo descriptivo y correlacional.

Es descriptivo porque presenta mediciones y gráficos que permitan comparar la vía de culminación del parto en gestantes con respecto a la Comp.Gestación (complicaciones de la madre), edad de la madre, edad gestacional (en semanas), número de bebés, peso del bebe por nacer, talla del bebe por nacer, Pc (perímetro cefálico) y vía de culminación del parto (normal o cesárea).

Es correlacional porque se desea encontrar un modelo que permita determinar la vía de culminación del parto en gestantes, en función a los factores asociados: Comp.Gestación (complicaciones de la madre), edad de la madre, edad gestacional (en semanas), número de bebés, peso del bebe por nacer, talla del bebe por nacer, Pc (perímetro cefálico).

### **Unidad de investigación:**

Un parto registrado en el centro médico de la Asociación Taller de los Niños comprendido entre junio de 2010 hasta octubre de 2012.

### **3.4 FORMULACIÓN DE HIPÓTESIS**

En el presente trabajo, la predicción de la vía de culminación del parto de gestantes será modelada mediante una regresión logística binaria.

#### **Hipótesis principal del estudio**

H<sub>0</sub>: Los datos sobre la vía de culminación del parto de las gestantes se ajustan a una regresión logística binaria.

H<sub>1</sub>: Los datos sobre la vía de culminación del parto de las gestantes no se ajustan a una regresión logística binaria.

#### **Hipótesis secundaria del estudio**

H<sub>0</sub>: La variable independiente no realiza un aporte significativo a la explicación de la variable respuesta.

H<sub>1</sub>: Para alguna variable independiente se realiza un aporte significativo a la explicación de la variable respuesta.

### **3.5 IDENTIFICACIÓN DE VARIABLES**

Para determinar la vía de culminación del parto de las gestantes trabajaremos con variables independientes.

### **a. Variable dependiente**

La variable dependiente es vía de culminación del parto de las gestantes (Tipo de parto):

0 Normal

1 Cesárea

### **b. Variables independientes**

- Comp.Gestación (complicaciones de la madre durante la gestación: El código 0 para aquellas gestantes que no tuvieron complicaciones durante el embarazo y 1 para las que sí tuvieron complicaciones durante el embarazo).
- Edad de la madre.
- Edad gestacional del bebé por nacer (semanas).
- Número de bebés.
- Peso del bebé por nacer (peso en kg en la semana gestacional).
- Talla del bebé por nacer (talla en cm en la semana gestacional).
- Perímetro cefálico del bebé por nacer (cm).

#### **IV. RESULTADOS Y DISCUSIÓN**

Como primer paso se realizó la identificación de las variables provenientes de los registros del centro médico de la Asociación Taller de los Niños. Las variables que hemos utilizado en el presente estudio, son variables designadas por los especialistas en el área de salud.

Se realizó un análisis de los datos, para verificar la no presencia de datos ausentes, y puntos extremadamente influyentes. Luego se procedió a realizar un análisis descriptivo y un análisis de multicolinealidad con el fin de determinar la no existencia de multicolinealidad entre las variables predictores.

Se realizó un análisis de regresión logística entre las variables predictores y la variable dependiente con todas las variables en estudio y se observó que todas las variables eran significativas al modelo.

Se procedió entonces a buscar el mejor conjunto de variables que ayuden a predecir la probabilidad de la vía de culminación de parto en gestantes que tuvieron parto normal o por cesárea, mediante el método de regresión logística binaria.

El método utilizado es el de selección directa que contrasta la entrada basándose en la significación del estadístico de puntuación y contrasta la eliminación basándose en la probabilidad del estadístico de Wald, se evaluó también la Bondad de Ajuste, la confiabilidad del modelo mediante el  $R^2$ , el test de Hosmer-Lemeshow y el test de Wald y se planteó la validación del modelo mediante validación cruzada.

A continuación se muestra el detalle de lo mencionado.

#### 4.1 PRUEBA BINOMIAL

La prueba binomial analiza variables dicotómicas y compara las frecuencias observadas en cada categoría con las que cabría esperar una distribución binomial de parámetro especificado de 0.5 en la hipótesis nula: Es decir en el presente cuadro 1 se contrasta la proporción de gestantes que tuvieron una vía de culminación de parto normal o por cesárea sean iguales con un parámetro de probabilidad especificado de 0.5. Se indica el rechazo de la proporción similar de la vía de culminación del parto en gestantes.

**Cuadro 1: Prueba binomial**

Prueba binomial						
	Categoría	N	Proporción observada	Prop. de prueba	Sig. asintót. (bilateral)	
TIPO DE PARTO	Normal	,00	1422	,73	,50	,000 <sup>a</sup>
	Cesárea	1,00	528	,27		
	Total		1950	1,00		

a. Basado en la aproximación Z.

Fuente: Elaboración propia

#### 4.2 ANÁLISIS DESCRIPTIVO

En el Cuadro 2 se observa que 1422 gestantes tuvieron una vía de culminación de parto normal, mientras 528 tuvieron por cesárea.

De las gestantes que tuvieron una vía de culminación de parto normal, el peso promedio del bebe por nacer es de 3,36 kg, la talla promedio del bebe por nacer es de 50,06 cm, y la edad máxima de la madre es de 39 años de edad. De las gestantes que tuvieron una vía de culminación de parto por cesárea el peso promedio del bebe por nacer

es de 3,45 kg, la talla promedio del bebe por nacer es de 49,89 cm, y la edad máxima de la madre es de 38 años.

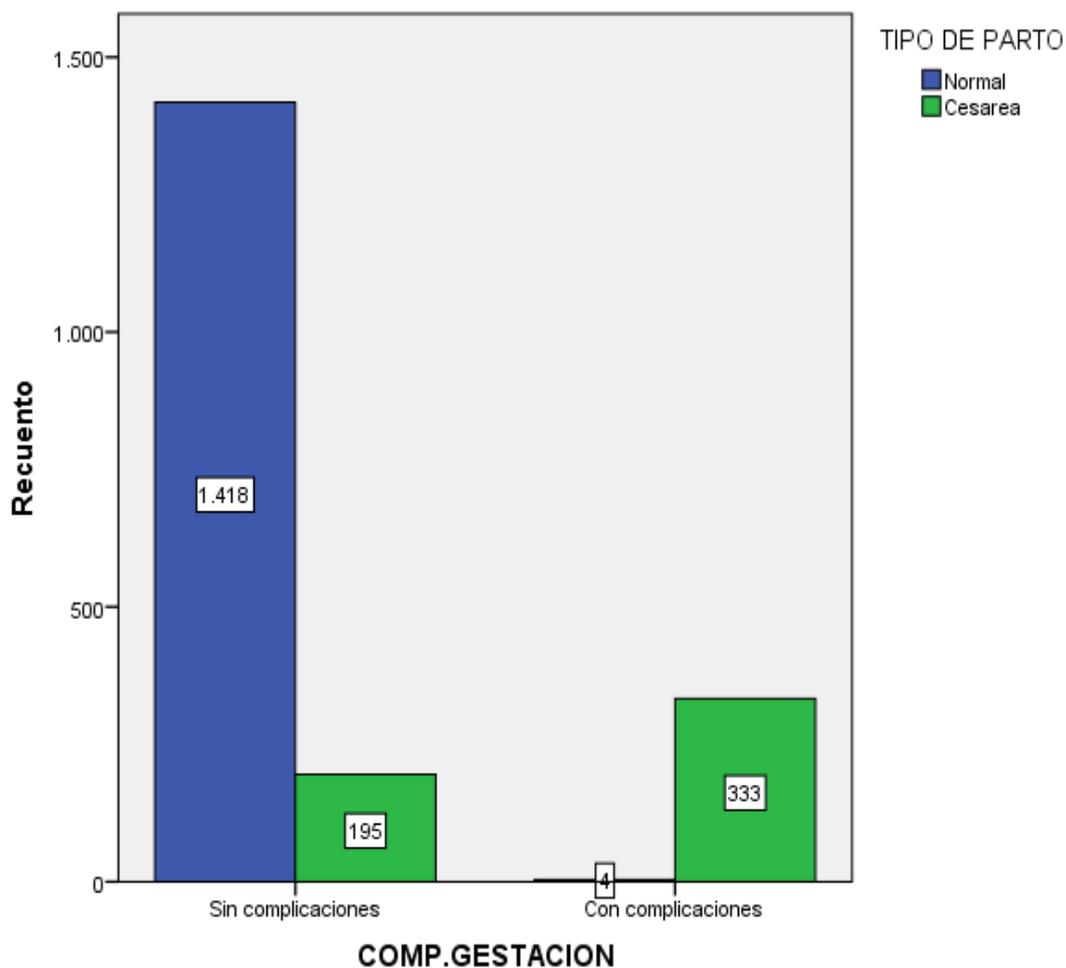
**Cuadro 2: Estadística descriptiva de las variables en estudio**

	TIPO DE PARTO									
	Normal					Cesárea				
	Recuento	Media	Desviación típica	Mínimo	Máximo	Recuento	Media	Desviación típica	Mínimo	Máximo
<b>Edad de la madre</b>	1422	25,68	5,39	18,00	39,00	528	26,76	5,37	18,00	38,00
<b>Edad gestacional</b>	1422	39,19	,80	36,00	41,00	528	39,00	,99	36,00	41,00
<b>Número de bebés</b>	1422	1,82	,87	1,00	4,00	528	1,74	,85	1,00	4,00
<b>Peso del bebe por nacer</b>	1422	3,36	,35	2,59	4,27	528	3,45	,37	2,62	4,28
<b>Talla del bebe por nacer</b>	1422	50,06	1,60	44,00	54,00	528	49,89	1,70	43,00	54,00
<b>Pc del bebe por nacer</b>	1422	34,24	1,27	30,00	39,50	528	34,98	1,30	30,00	39,80

Fuente: Elaboración propia

En la Figura 1, se muestra la cantidad de gestantes que tuvieron una vía de culminación de parto normal o por cesárea según las complicaciones que tuvieron en la etapa de gestación, el 82,72 por ciento de las gestantes no tuvieron complicaciones durante la etapa de gestación y el 17,28 por ciento de mujeres gestantes si tuvieron complicaciones durante la etapa de gestación, y el 72,92 por ciento de gestantes tuvieron una vía de culminación de parto normal, mientras que el 27,08 por ciento de mujeres gestantes tuvieron una vía de culminación de parto por cesárea.

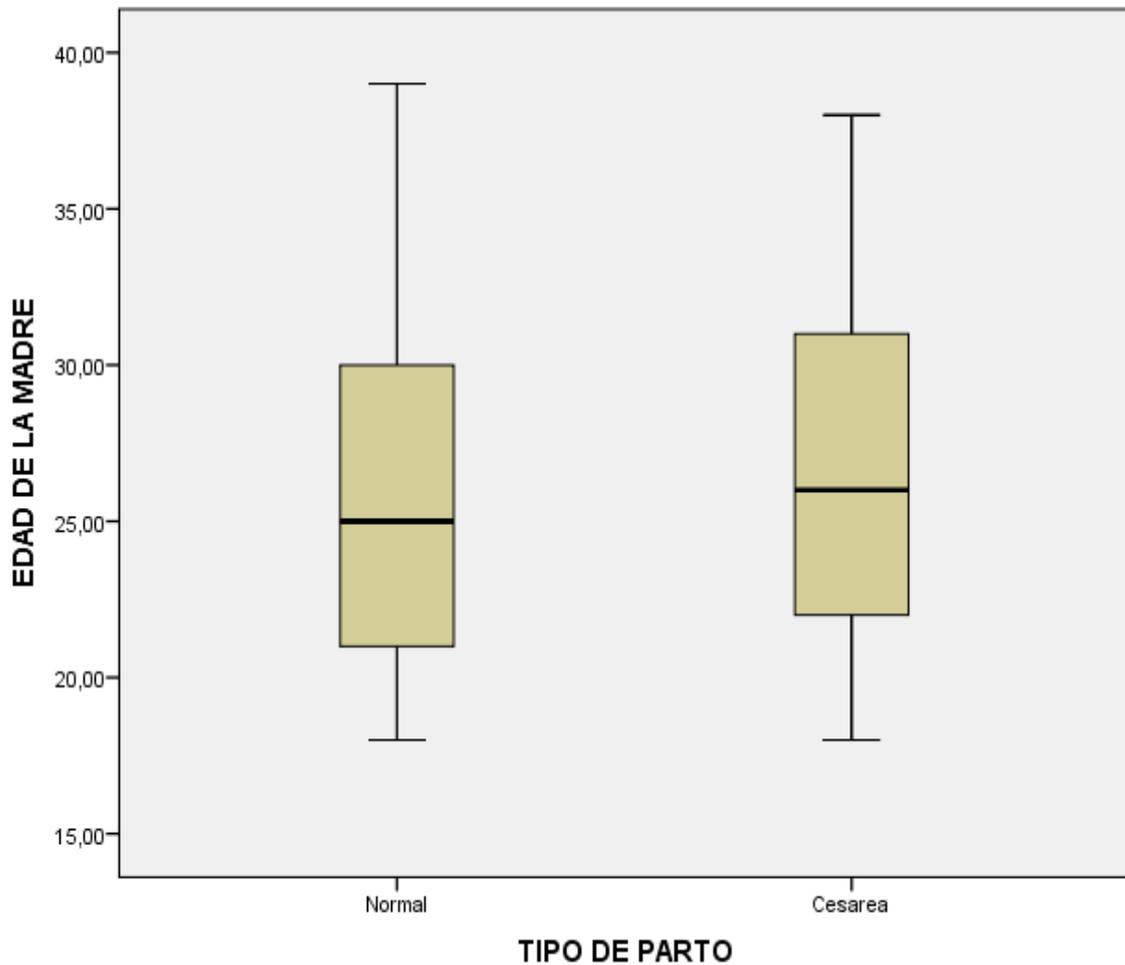
**Figura 1: Clasificación de las gestantes según la vía de culminación de parto (normal o por cesárea)**



Fuente: Elaboración propia

En la Figura 2, se observa que existe una similar variabilidad en el 50 por ciento central de los datos de la edad de la madre que ha tenido una vía de culminación de parto normal o por cesárea.

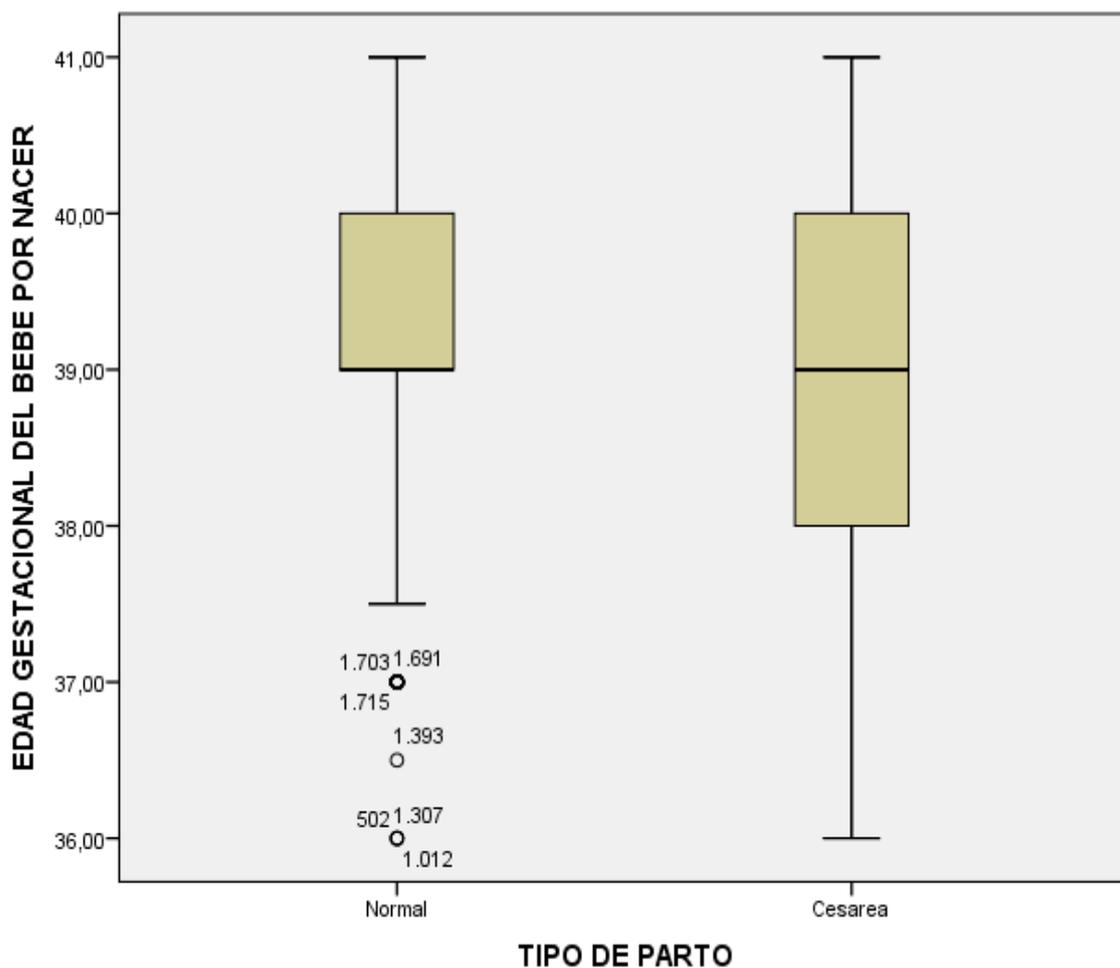
**Figura 2: Edad de la madre respecto a la vía de culminación de parto (normal o por cesárea).**



Fuente: Elaboración propia

En la Figura 3, se observa que existe mayor variabilidad en el 50 por ciento central de los datos de la edad gestacional con respecto a las gestantes que tuvieron una vía de culminación de parto por cesárea con respecto a las que tuvieron una vía de culminación de parto normal, aunque las que tuvieron una vía de culminación de parto normal presentan varios puntos influyentes que están fuera del límite inferior por ejemplo (observación 1012 = 36 semanas y observación 1307 = 36 semanas) y sus observaciones en los bigotes se encuentren más dispersos.

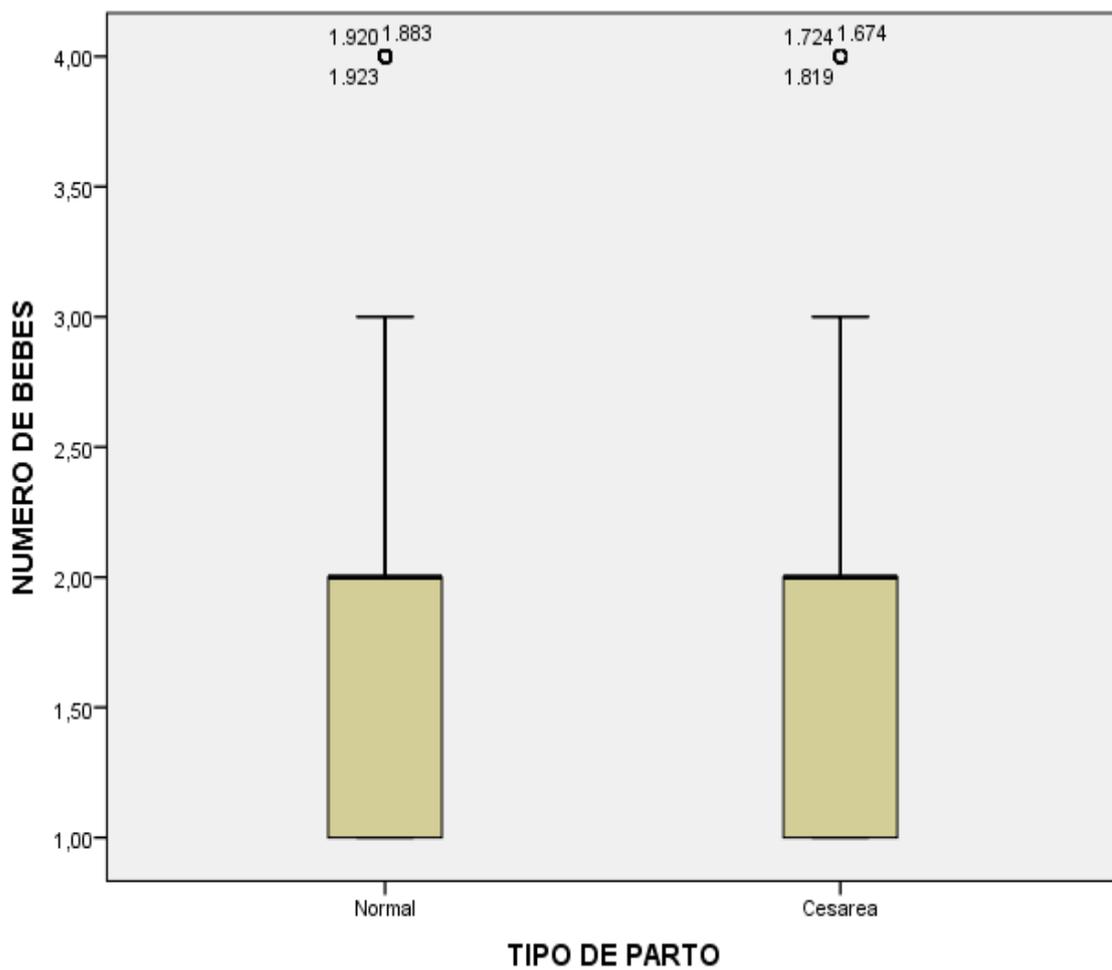
**Figura 3: Edad gestacional del bebe por nacer respecto a la vía de culminación de parto (normal o por cesárea) de las madres gestantes**



Fuente: Elaboración propia

En la Figura 4, se puede apreciar que el número de bebés de las gestantes que tuvieron parto normal o por cesáreas presentan varias observaciones que están fuera del límite superior en ambos casos por ejemplo (observación 1883 = 4 bebés y observación 1920 = 4 bebés para las gestantes que tuvieron parto normal) y (observación 1724 = 4 bebés y observación 1819 = 4 bebés para las gestantes que tuvieron parto por cesárea).

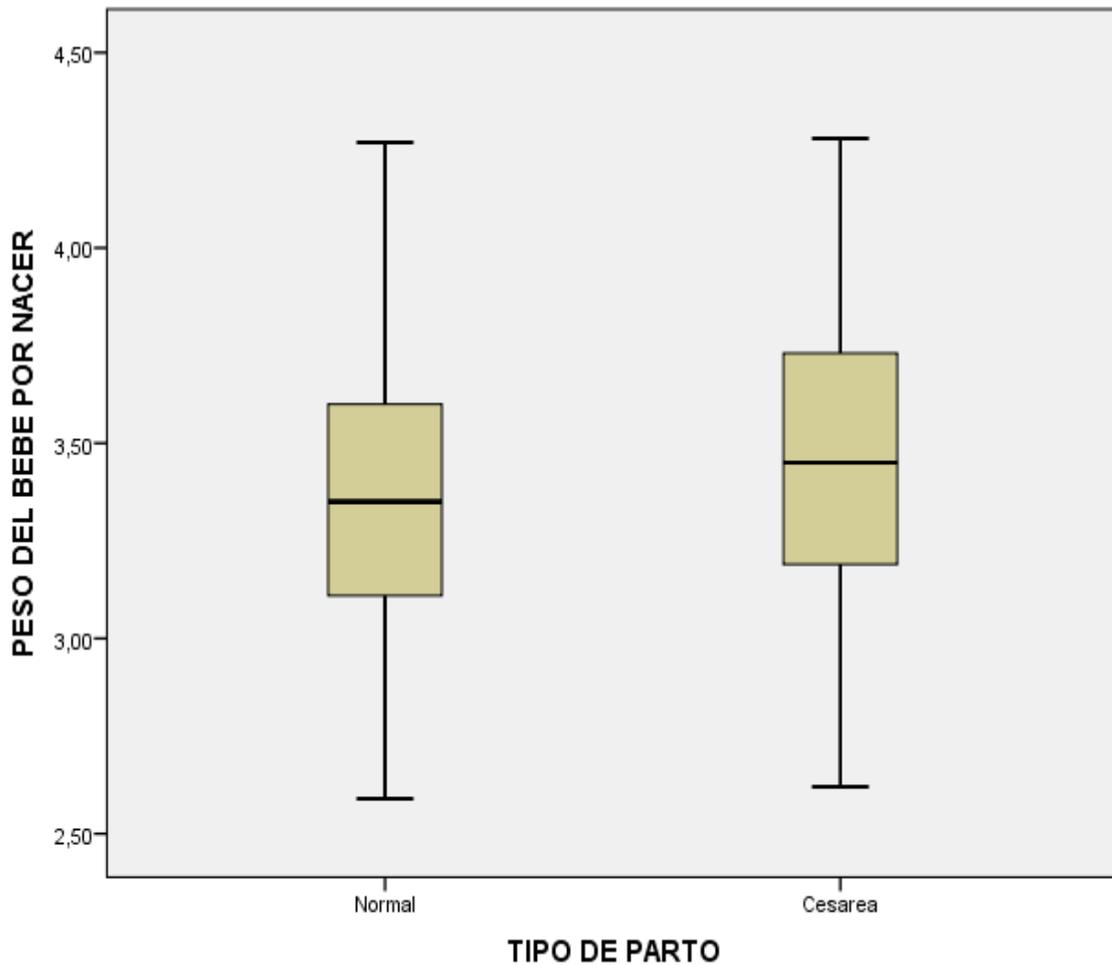
**Figura 4: Número de bebés respecto a la vía de culminación de parto (normal o por cesárea) de las madres gestantes**



Fuente: Elaboración propia

En la Figura 5, podemos observar que existe una mayor variabilidad en el 50 por ciento central de los datos del peso al nacer del bebe de las gestantes que han tenido una vía de culminación de parto por cesárea; también se observa la no existencia de puntos influyentes.

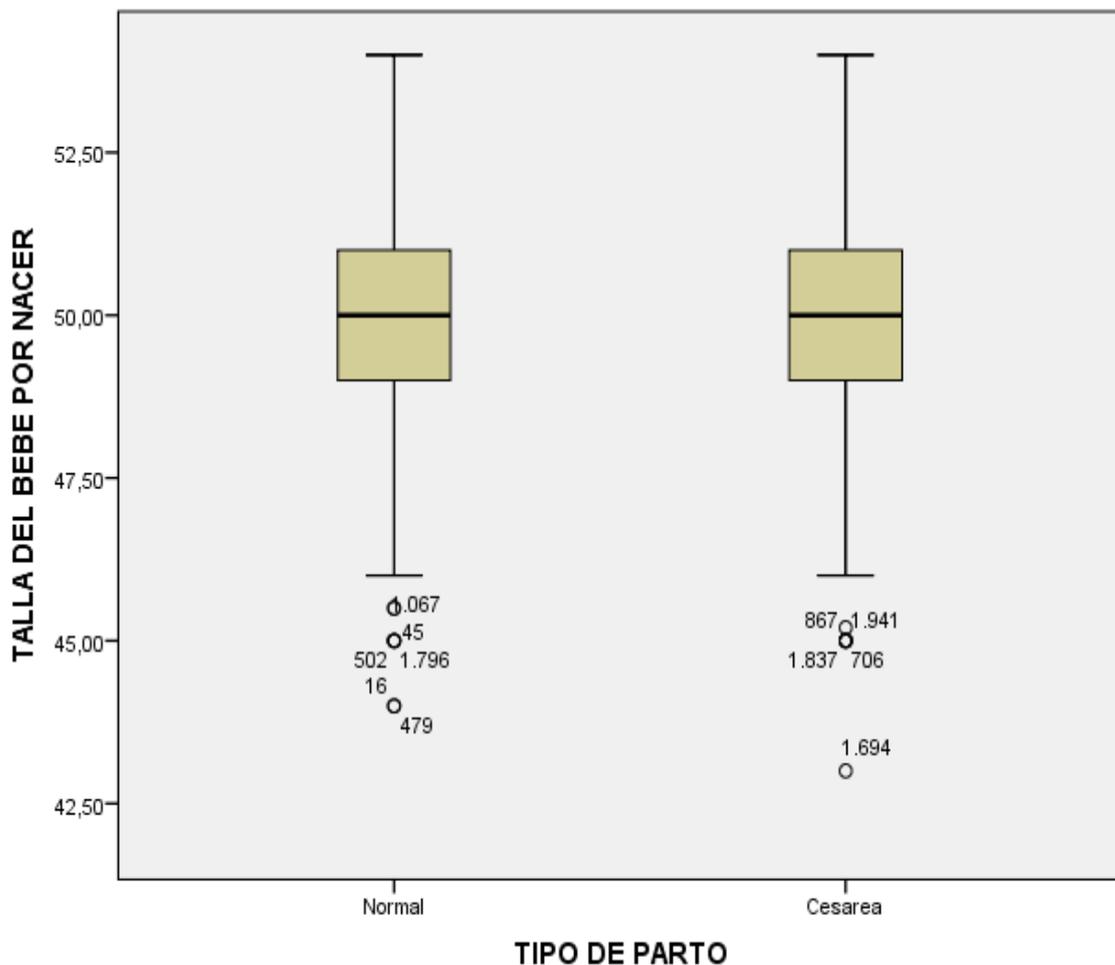
**Figura 5: Peso del bebe por nacer de las madres gestantes respecto a la vía de culminación de parto (normal o por cesárea)**



Fuente: Elaboración propia

En la Figura 6, podemos observar que existe una similar variabilidad en el 50 por ciento central de los datos de la talla al nacer del bebe de las gestantes que han tenido una vía de culminación de parto normal o por cesárea; también se observa la existencia de puntos influyentes. Presentan varias observaciones que están por debajo del límite inferior por ejemplo: (observación 16 = 44 cm y observación 479 = 44 cm en la talla del bebe por nacer para las gestantes que tuvieron una vía de culminación de parto normal) y (observación 1694 = 43 cm y observación 1941 = 45 cm en la talla del bebe por nacer para las gestantes que tuvieron una vía de culminación de parto por cesárea).

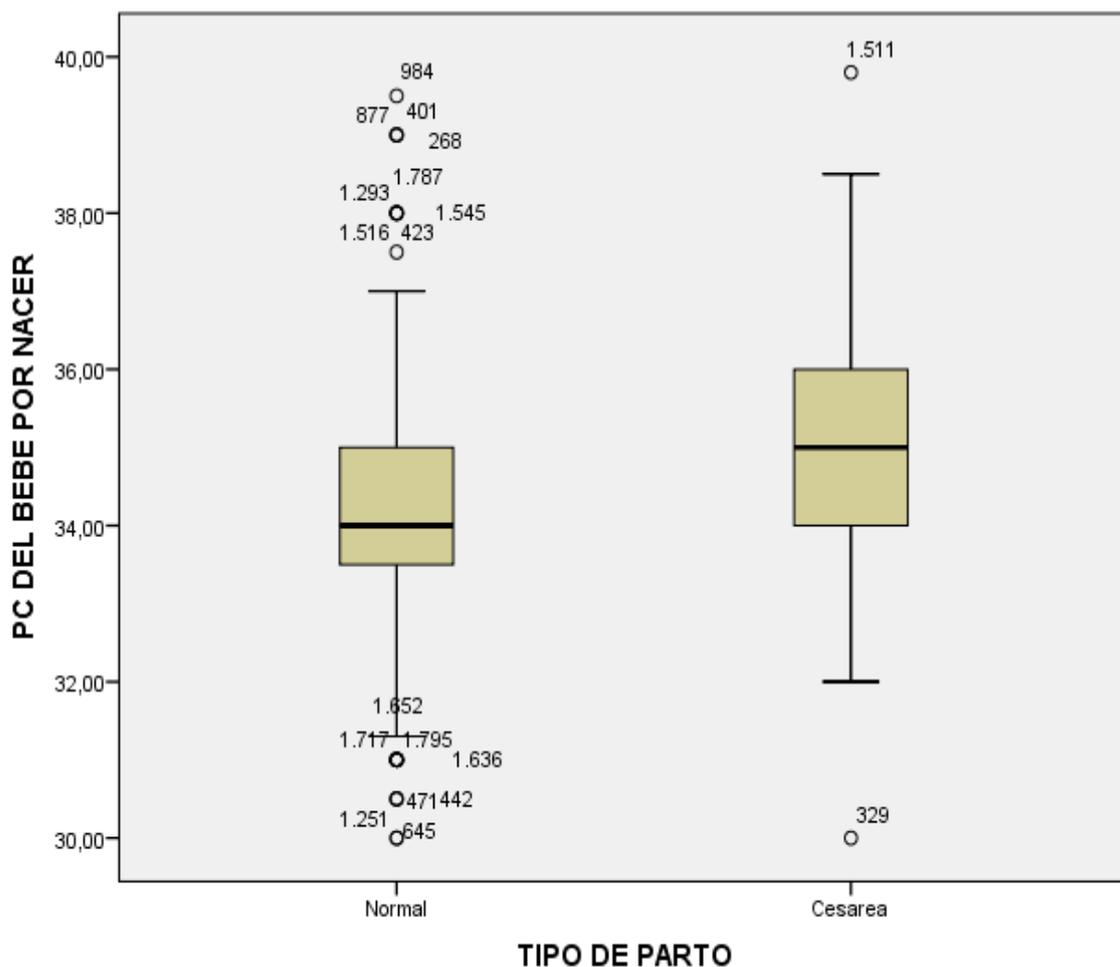
**Figura 6: Talla del bebe por nacer de las madres gestantes respecto a la vía de culminación de parto (normal o por cesárea)**



Fuente: Elaboración propia

En la Figura 7, podemos observar que existe una mayor variabilidad en el 50 por ciento central de los datos del perímetro cefálico del bebe de las gestantes que han tenido una vía de culminación de parto por cesárea en comparación con las que tuvieron una vía de culminación de parto normal, con respecto a los puntos influyentes podemos apreciar en el grafico que la dispersión de los puntos influyentes que sobrepasan el límite superior e inferior es mucho mayor en las gestantes que tuvieron una vía de culminación de parto normal con respecto a las que tuvieron una vía de culminación de parto por cesárea.

**Figura 7: Perímetro cefálico del bebe por nacer de las madres gestantes respecto a la vía de culminación de parto (normal o por cesárea)**



Fuente: Elaboración propia

### 4.3 DIAGNOSTICO DE MULTICOLINEALIDAD

#### Regla empírica (Kleinbaum):

Existen problemas de colinealidad si algún  $VIF$  es  $>10$ .  $VIF = 1/1 - R_j^2$  donde  $R_j^2$  es el coeficiente de determinación múltiple para la regresión de  $x_j$  en función de las otras variables explicativas. Otra forma de determinar si existe multicolinealidad es cuando la Tolerancia  $< 0.1$ , la cual es la inversa del VIF.

**Cuadro 3: Diagnostico de multicolinealidad**

Coeficientes			
Modelo		Estadísticos de Colinealidad	
		Tolerancia	VIF
1	Edad de la madre	,705	1,418
	Edad gestacional	,878	1,138
	Numero de bebes	,704	1,421
	Peso del bebe por nacer	,449	2,227
	Talla del bebe por nacer	,601	1,665
	PC del bebe por nacer	,669	1,494
	Comp.Gestacion	,943	1,060

Fuente: Elaboración propia

En la presente cuadro, se muestra que todas las variables poseen un VIF menor a 10 y una tolerancia mayor a 0.4.

#### 4.4 SELECCIÓN DE VARIABLES

Se realizó la estimación inicial del modelo con ocho variables predictoras, sin embargo, aplicando el método directo que brinda el SPSS por defecto, se determinó que todas las variables son significativas.

En el Cuadro 4, se observa todas las variables significativas para el modelo:

- Comp.Gestación: Complicaciones durante la gestación
- Edad de la madre: Edad de la madre
- Edad gestacional: Edad gestacional del bebe por nacer
- Número de bebés: Número de bebés
- Peso al nacer: Peso del bebe por nacer (peso en la semana gestacional)
- Talla al nacer: Talla del bebe por nacer (talla en la semana gestacional)
- Pc: Perímetro cefálico del bebe por nacer.

**Cuadro 4: Variables que ingresan al modelo**

		Variables en la ecuación						I.C. 95% para	
		B	E.T.	Wald	Gl	Sig.	Exp(B)	EXP(B)	
								Inferior	Superior
Paso 1 <sup>a</sup>	Comp.gestación(1)	6,515	,517	158,731	1	,000	675,270	245,078	1860,586
	Edad de la madre	,078	,017	21,119	1	,000	1,081	1,046	1,117
	Edad gestacional	-,397	,098	16,475	1	,000	,672	,555	,814
	Número de bebés	-,590	,118	25,095	1	,000	,554	,440	,698
	Peso del bebe por nacer	,998	,321	9,646	1	,002	2,712	1,445	5,091
	Talla del bebe por nacer	-,266	,062	18,177	1	,000	,766	,678	,866
	Pc	,518	,072	52,073	1	,000	1,678	1,458	1,931
	Constante	4,536	4,646	0,953	1	,329	93,317		

a. Variable(s) introducida(s) en el paso 1: Comp. Gestación, Edad de la madre, Edad gestacional, Numero de bebes, Peso del bebe por nacer, Talla del bebe por nacer, Pc.

Fuente: Elaboración propia

En el Cuadro 5, se observa el coeficiente de determinación. Para nuestro modelo el coeficiente de determinación es de 65.9 por ciento. Es decir un 65.9 por ciento de la variación de la variable tipo de parto (*vía de culminación de parto normal o por cesárea*) es explicada por las variables incluidas en el modelo que son Comp.Gestación, Edad de la madre, Edad gestacional del bebe por nacer, Número de bebés, Peso del bebe por nacer, Talla del bebe por nacer, Pc.

**Cuadro 5: Prueba del coeficiente de determinación**

**Resumen del modelo**

Paso	-2 log de la verosimilitud	R cuadrado de Cox y Snell	R cuadrado de Nagelkerke
1	1096,933 <sup>a</sup>	,454	,659

- a. La estimación ha finalizado en el número de iteración 7 porque las estimaciones de los parámetros han cambiado en menos de ,001.

Fuente: Elaboración propia

En el Cuadro 6, se observa la tabla de clasificación, para el modelo se obtiene una tasa de buena clasificación del 88.5 por ciento. En la tabla de clasificación se comprueba que el modelo tiene una especificidad alta (94.8%) y una sensibilidad de (71.6%). Con las variables predictoras: Comp.Gestación, Edad de la madre, Edad gestacional del bebe por nacer, Número de bebés, Peso del bebe por nacer, Talla del bebe por nacer, Pc se comprueba que el modelo tiene una efectividad de 88.5 por ciento, es decir posee una buena capacidad predictora con las variables incluidas en el modelo. Hay que tomar en cuenta que el punto de corte es de 0,3. (Ver Cuadro 6).

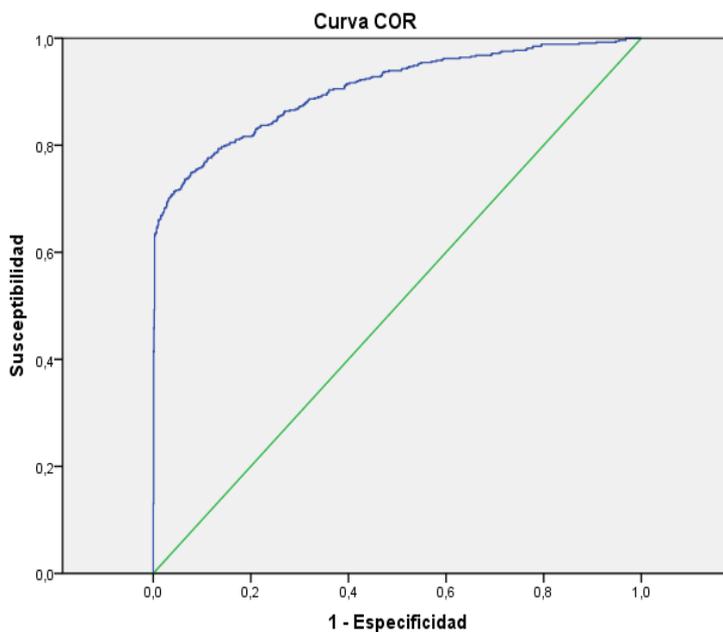
**Cuadro 6: Tabla de clasificación**

Observado			Pronosticado		
			Tipo de parto		Porcentaje correcto
			,00	1,00	
Paso 1	Tipo de parto	,00	1348	74	94,8
		1,00	150	378	71,6
		Porcentaje global			88,5

a. El valor de corte es 0,300

Fuente: Elaboración propia

En la Figura 8 se observa que nuestro modelo representado por la curva azul está por encima de la curva aleatoria cubriendo un área de 0,905.



**Figura 8: Grafica de la curva COR**

Área bajo la curva
Área
,905

Si el área que encierra la Curva COR es 1, se tendría una clasificación perfecta y conforme nos alejamos de este valor la clasificación disminuye; entonces como la curva COR es 0.9 hace de este un buen modelo para clasificar a las gestantes. Con punto de corte 0.3.

**VP:** La vía de culminación del parto de una gestante es por cesárea

Predicción Correcta 71.6%

**VN:** La vía de culminación del parto de una gestante es normal-Predicción Correcta 94.8%

**FP:** La vía de culminación del parto de una gestante es normal-Predicción Incorrecta 5.2%

**FN:** La vía de culminación del parto de una gestante es por cesárea-

Predicción Incorrecta 28.4%

#### 4.5 VEROSIMILITUD DEL MODELO

En el Cuadro 7, se observa la significación estadística con la prueba de Wald, la constante del modelo es significativa, por lo tanto, la constante se debe incluir en el modelo porque produce un cambio sobre la variable Tipo de parto (vía de culminación de parto normal o por cesárea).

#### Cuadro 7: Prueba del coeficiente de intercepción

		Variables en la ecuación					
		B	E.T.	Wald	G1	Sig.	Exp(B)
Paso 0	Constante	-,991	,051	377,923	1	,000	,371

Fuente: Elaboración propia

En el Cuadro 8, al evaluar la hipótesis nula ( $B_i=0$ ) la significancia estadística asociada, y el valor de la OR( $\exp(B)$ ), las variables que contribuyen al modelo son Comp.Gestación, Edad de la madre, Edad gestacional del bebe por nacer, número de bebés, Peso del bebe por nacer, Talla del bebe por nacer, Pc a un nivel de significación del 5 por ciento.

#### Cuadro 8: Modelo

Por lo tanto; el modelo será:

$$\ln(\text{odds})=4,536+6,515*\text{Comp.Gestacion}+0,078*\text{Edad de la madre}-0,397*\text{Edad gestacional del bebe por nacer}-0,590*\text{Número de bebés}+0,998*\text{Peso del bebe por nacer}-0,266*\text{Talla del bebe por nacer}+ 0,518*\text{Pc del bebe por nacer}$$

Fuente: Elaboración propia

#### 4.6 PRUEBA DE BONDAD DE AJUSTE

En el Cuadro 9, se obtiene la prueba de bondad de ajuste del modelo, a un nivel de significación de 5 por ciento, el modelo es el adecuado, es decir el modelo de regresión logística se ajusta a los datos ( $p$ -valor=0,760).

##### Cuadro 9: Prueba de bondad de ajuste

**Prueba de Hosmer y Lemeshow**

Paso	Chi cuadrado	gl	Sig.
1	4,978	8	,760

Fuente: Elaboración propia

En el Cuadro 10, se tiene la prueba Ómnibus:

En la primera fila se contrasta la  $H_0$  de que los coeficientes de las variables añadidas son cero. Con la significancia de 0.000 se rechaza la  $H_0$ , es decir al menos uno de los coeficientes de las variables es diferente de cero.

Indica que el modelo con las variables pertenecientes al modelo tiene un ajuste significativo.

##### Cuadro 10: Prueba de Ómnibus

**Pruebas ómnibus sobre los coeficientes del modelo**

		Chi cuadrado	Gl	Sig.
Paso 1	Paso	1180,754	7	,000
	Bloque	1180,754	7	,000
	Modelo	1180,754	7	,000

Fuente: Elaboración propia

## 4.7 VALIDACIÓN CRUZADA

Tomando en cuenta que cada submuestra está conformada por 195 observaciones, realizaremos la validación cruzada  $k=10$  veces hasta obtener la media de la tasa de mala clasificación.

En los siguientes cuadros, se observa la tasa de mala clasificación aplicando validación cruzada en Excel.

### **Cuadro 11: Prueba de validación cruzada - Primer modelo generado (Ver anexo 1)**

El presente modelo ha sido generado a partir de restar el total de observaciones (1950) menos las 195 décimas observaciones que forman la submuestra número 1. Es decir el modelo fue generado con 1755 datos. Tomar en cuenta que la submuestra es la décima parte del total de datos.

Constante	Edad de la madre	Edad gestacional	NBB	Peso	Talla	PC	Comp. Gestación
2,887	0,075	-0,397	-0,527	1,01	-0,246	0,532	6,502

Fuente: Elaboración propia

**TMC: 13.92%**

El valor de 13.92 por ciento, representa la proporción de casos en la submuestra número 1 que no fueron asignados al grupo correcto después de realizada la predicción.

### **Cuadro 12: Prueba de validación cruzada - Segundo modelo generado**

El presente modelo ha sido generado a partir de restar el total de observaciones (1950) menos las 195 novenas observaciones que forman la submuestra número 2. Es decir el modelo fue generado con 1755 datos. Tomar en cuenta que la submuestra es la décima parte del total de datos.

Constante	Edad de la madre	Edad gestacional	NBB	Peso	Talla	PC	Comp. Gestación
5,056	0,076	-0,411	-0,616	1,071	-0,246	0,484	6,411

Fuente: Elaboración propia

**TMC: 8.76%**

El valor de 8.76 por ciento, representa la proporción de casos en la submuestra número 2 que no fueron asignados al grupo correcto después de realizada la predicción.

### **Cuadro 13: Prueba de validación cruzada- Tercer modelo generado**

El presente modelo ha sido generado a partir de restar el total de observaciones (1950) menos las 195 octavas observaciones que forman la submuestra número 3. Es decir el modelo fue generado con 1755 datos. Tomar en cuenta que la submuestra es la décima parte del total de datos.

Constante	Edad de la madre	Edad gestacional	NBB	Peso	Talla	PC	Comp. Gestación
1,691	0,081	-0,352	-0,605	0,929	-0,264	0,55	6,396

Fuente: Elaboración propia

**TMC: 12.89%**

El valor de 12.89 por ciento, representa la proporción de casos en la submuestra número 3 que no fueron asignados al grupo correcto después de realizada la predicción

#### **Cuadro 14: Prueba de validación cruzada - Cuarto modelo generado**

El presente modelo ha sido generado a partir de restar el total de observaciones (1950) menos las 195 séptimas observaciones que forman la submuestra número 4. Es decir el modelo fue generado con 1755 datos. Tomar en cuenta que la submuestra es la décima parte del total de datos.

Constante	Edad de la madre	Edad gestacional	NBB	Peso	Talla	PC	Comp. Gestación
4,49	0,077	-0,412	-0,574	0,963	-0,268	0,541	6,382

Fuente: Elaboración propia

**TMC: 10.82%**

El valor de 10.82 por ciento, representa la proporción de casos en la submuestra número 4 que no fueron asignados al grupo correcto después de realizada la predicción.

#### **Cuadro 15: Prueba de validación cruzada - Quinto modelo generado**

El presente modelo ha sido generado a partir de restar el total de observaciones (1950) menos las 195 sextas observaciones que forman la submuestra número 5. Es decir el modelo fue generado con 1755 datos. Tomar en cuenta que la submuestra es la décima parte del total de datos.

Constante	Edad de la madre	Edad gestacional	NBB	Peso	Talla	PC	Comp. Gestación
5,276	0,077	-0,399	-0,624	0,935	-0,268	0,508	7,068

Fuente: Elaboración propia

**TMC: 12.37%**

El valor de 12.37 por ciento, representa la proporción de casos en la submuestra número 5 que no fueron asignados al grupo correcto después de realizada la predicción.

### **Cuadro 16: Prueba de validación cruzada - Sexto modelo generado**

El presente modelo ha sido generado a partir de restar el total de observaciones (1950) menos las 195 quintas observaciones que forman la submuestra número 6. Es decir el modelo fue generado con 1755 datos. Tomar en cuenta que la submuestra es la décima parte del total de datos.

Constante	Edad de la madre	Edad gestacional	NBB	Peso	Talla	PC	Comp. Gestación
2,281	0,07	-0,358	-0,573	0,828	-0,249	0,535	6,417

Fuente: Elaboración propia

**TMC: 11.34%**

El valor de 11.34 por ciento, representa la proporción de casos en la submuestra número 6 que no fueron asignados al grupo correcto después de realizada la predicción.

### **Cuadro 17: Prueba de validación cruzada-Séptimo modelo generado**

El presente modelo ha sido generado a partir de restar el total de observaciones (1950) menos las 195 cuartas observaciones que forman la submuestra número 7. Es decir el modelo fue generado con 1755 datos. Tomar en cuenta que la submuestra es la décima parte del total de datos.

Constante	Edad de la madre	Edad gestacional	NBB	Peso	Talla	PC	Comp. Gestación
6,352	0,076	-0,4	-0,576	1,161	-0,289	0,485	6,424

Fuente: Elaboración propia

**TMC: 12.37%**

El valor de 12.37 por ciento, representa la proporción de casos en la submuestra número 7 que no fueron asignados al grupo correcto después de realizada la predicción.

### **Cuadro 18: Prueba de validación cruzada - Octavo modelo generado**

El presente modelo ha sido generado a partir de restar el total de observaciones (1950) menos las 195 terceras observaciones que forman la submuestra número 8. Es decir el modelo fue generado con 1755 datos. Tomar en cuenta que la submuestra es la décima parte del total de datos.

Constante	Edad de la madre	Edad gestacional	NBB	Peso	Talla	PC	Comp. Gestación
5,275	0,079	-0,395	-0,586	1,086	-0,274	0,497	7,112

Fuente: Elaboración propia

**TMC: 9.28%**

El valor de 9.28 por ciento, representa la proporción de casos en la submuestra número 8 que no fueron asignados al grupo correcto después de realizada la predicción.

### **Cuadro 19: Prueba de validación cruzada - Noveno modelo generado**

El presente modelo ha sido generado a partir de restar el total de observaciones (1950) menos las 195 segundas observaciones que forman la submuestra número 9. Es decir el modelo fue generado con 1755 datos. Tomar en cuenta que la submuestra es la décima parte del total de datos.

Constante	Edad de la madre	Edad gestacional	NBB	Peso	Talla	PC	Comp. Gestación
4,505	0,085	-0,41	-0,586	0,794	-0,245	0,518	6,293

Fuente: Elaboración propia

**TMC: 10.31%**

El valor de 10.31 por ciento, representa la proporción de casos en la submuestra número 9 que no fueron asignados al grupo correcto después de realizada la predicción.

## Cuadro 20: Prueba de validación cruzada - Décimo modelo generado (Ver anexo 2).

El presente modelo ha sido generado a partir de restar el total de observaciones (1950) menos las 195 primeras observaciones que forman la submuestra número 10. Es decir el modelo fue generado con 1755 datos. Tomar en cuenta que la submuestra es la décima parte del total de datos.

Constante	Edad de la madre	Edad gestacional	NBB	Peso	Talla	PC	Comp. Gestación
7,353	0,083	-0,445	-0,638	1,203	-0,313	0,535	6,504

Fuente: Elaboración propia

**TMC: 14.43%**

El valor de 14.43 por ciento, representa la proporción de casos en la submuestra número 10 que no fueron asignados al grupo correcto después de realizada la predicción.

### Ejemplo de Predicción-Modelo Global

Vamos a calcular la probabilidad de que una madre tenga parto por cesárea dado que tiene 30 años, no tuvo complicaciones durante la gestación, la edad gestacional del bebe por nacer es de 40 semanas, la madre tiene 1 hijo, el peso del bebe por nacer es de 3.17 kilogramos, la talla del bebe por nacer es de 44 cm, el perímetro cefálico del bebe por nacer es de 35cm.

$$prob(evento) = \frac{1}{1 + e^{-\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k}}$$

$$prob(evento) = \frac{1}{1 + e^{-4.536 + 6.515 X_1 + 0.078 X_2 - 0.397 X_3 - 0.590 X_4 + 0.998 X_5 - 0.266 X_6 + 0.518 X_7}} = 0.498915$$

La probabilidad de que esta madre en particular, tenga parto por cesárea es de 0,498915 y como hemos optado por utilizar un punto de corte de 0.3 con el fin de maximizar la tasa de clasificaciones correctas podemos concluir que dicha madre se clasifica en el grupo de madres que tendrán parto por cesárea.

**Conclusiones de la validación cruzada:** El 11.69 por ciento, representa la media de los errores de todas las submuestras, es decir indica la media de la tasa de mala clasificación realizada para determinar la vía de culminación del parto en gestantes que fueron asistidas en el centro de salud de la Asociación Taller de los Niños.

#### 4.8 CUMPLIMIENTO DE SUPUESTOS DE REGRESION LOGISTICA

**Cuadro 21: Prueba T - La media de los errores es cero.**

Estadísticos para una muestra				
	N	Media	Desviación típ.	Error típ. de la media
Diferencia entre las probabilidades pronosticadas y observadas	1950	0,0000000000	0,28532909	0,00646143

##### - No autocorrelación

Dada la naturaleza de las variables lo que ocurre con una madre no tiene nada que ver con lo que ocurra con la siguiente madre; por ello no hay autocorrelación; pero sí podría haber autocorrelación si fuese la misma madre.

##### - Multicolinealidad

Se detectó a través de los diagnósticos de colinealidad, la no existencia de multicolinealidad debido a que no tuvimos un  $VIF > 10$  ni una tolerancia menor a 0.1.

## V. CONCLUSIONES

Al iniciar el Análisis de Regresión Logística se empleó inicialmente ocho variables consideradas importantes para el estudio (comp.gestacional, edad de la madre, edad gestacional del bebe por nacer, numero de bebes, peso del bebe por nacer, talla del bebe por nacer y perímetro cefálico).

Se debe señalar que se utilizó el método directo (método por defecto que brinda el software SPSS para la selección de variables) en el cual el modelo es generado usando todas las variables independientes al mismo tiempo para que todas las variables predictoras entran juntas (Ver cuadro 4). En ese sentido encontramos que todas las variables resultan ser significativas para el modelo.

Se determinó el modelo adecuado para predecir la vía de culminación del parto en gestantes del centro médico de la Asociación Taller de los Niños, con una efectividad de 88.5 por ciento. Existe 5.2 por ciento de casos de cometer el error de que la gestante tenga una culminación de parto por cesárea cuando la gestante tiene una culminación de parto normal y 28.4 por ciento de casos de cometer el error de que la gestante tenga una culminación de parto normal cuando la gestante tiene una culminación de parto por cesárea.

En la tabla de clasificación, 1348 gestantes presentan una vía de culminación de parto normal que significa que 94.8 por ciento están correctamente clasificados, mientras que 378 gestantes presentan una vía de culminación de parto por cesárea (71.6%) correctamente clasificados y 224 gestantes se encuentran mal clasificadas.

El 65.9 por ciento de la variación de la vía de culminación del parto en gestantes es explicada por todas las variables que ingresaron al modelo: Comp. Gestación, edad de la

madre, edad gestacional del bebe por nacer, numero de bebes, peso del bebe por nacer, talla del bebe por nacer y perímetro cefálico del bebe por nacer.

El error de la prueba de regresión logística, estimado mediante validación cruzada es de 11.69 por ciento, siendo un indicador bueno al estudio.

El ODDS Ratio para la variable Comp.Gestación es 675.270, ello quiere decir que manteniendo constantes las demás variables, el hecho de que la gestante tenga un parto por cesárea es 675.270 más ventajoso entre las gestantes que han tenido complicaciones durante el embarazo en comparación con quienes no la han tenido. En síntesis la gestante que ha tenido complicaciones durante el embarazo tiene más ventaja de tener un parto por cesárea.

El ODDS Ratio para la variable Edad de la madre es 1.081, para la variable Peso del bebe por nacer es 2.712 y Pc es 1.678, todos ellos son mayores a la unidad; por lo tanto son factores que incrementan la probabilidad de ocurrencia de tener una vía de culminación de parto por cesárea.

El ODDS Ratio para la variable Edad Gestacional es 0.672, para la variable Numero de bebes es 0.554 y para la Talla del bebe por nacer es 0,766, todos ellos presentan un ODDS Ratio menor que uno; por lo tanto dichas variable son factores que disminuyen la probabilidad de ocurrencia de tener una vía de culminación de parto por cesárea.

## VI. RECOMENDACIONES

1. Un paso previo al análisis estadístico es la “limpieza de datos”, es decir detectar la ausencia de datos faltantes, datos ilógicos etc. Parte de este análisis es también la detección de outliers (o valores extremos) que no contribuyen a obtener un buen el modelo y ante ello se deben tomar las mejores decisiones con el fin de encontrar el resultado deseado.
2. Para un estudio futuro, se recomienda utilizar más variables predictoras para poder tener un mejor modelo, pueden ser variables independientes como controles prenatal, variedad de presentación del parto, etc. y operacionales como las medidas del cráneo, hombros, cadera, duración de las contracciones, frecuencia de contracciones, intensidad de las contracciones, duración del parto, anomalías del vientre de la madre, etc. Adicionalmente se debe realizar un análisis a mayor profundidad de los datos que no han sido clasificados correctamente, con el fin de disminuir el grado de error y aumenta la tasa de clasificación para ello se pueden usar otros métodos de predicción como arboles de clasificación, análisis discriminante, etc.
3. El modelo es como una alerta para los especialistas de la salud por las principales variables que podrían definir una cesárea y en nuestro estudio todas las variables resultaron significativas; por ello se recomienda a los centros de salud tener en cuenta las variables significativas en nuestro modelo con el fin de planificar sus actividades para que tengan nacimientos sin problemas. Por ejemplo cuando la edad de la madre aumenta en un año manteniendo constantes las demás variables la probabilidad de que ocurra un parto por cesárea es mayor a que ocurra un parto normal; por lo cual tanto el centro de salud como los especialistas deben tomar las medidas preventivas ante tales eventos; porque ello nos indica que las cesáreas se incrementan a medida que aumenta la edad, por ello se debe prestar atención a la edad de la madre al momento de ingresar al centro de salud.

## VII. REFERENCIAS BIBLIOGRÁFICAS

1. Berríos, M. 2014. En 10% se incrementaron los casos de cesáreas en los últimos 6 años. (en línea). La República, Lima, PE. Consultado 23 de agosto de 2014. Disponible en <http://www.larepublica.pe/12-05-2014/en-10-se-incrementaron-los-casos-de-cesareas-en-los-ultimos-6-anos>
2. Cabero, L; Saldívar, D; Cabrillo, E. 2007. Obstetricia Y Medicina Materno Fetal. Editorial Médica Panamericana.
3. Campomanes, F. 2010. Clasificación de personas según su nivel de obesidad aplicando regresión logística binaria a los clientes del gimnasio Golds Gym en Lima Metropolitana. Modalidad Examen Profesional. Lima Perú.
4. Cañadas, JL. 2013. Regresión logística. Tratamiento computacional con R. (en línea). Granada, España. Consultado 15 de agosto de 2014. Disponible en [http://masteres.ugr.es/moea/pages/tfm-1213/tfm\\_caaadasreche\\_jluis/!](http://masteres.ugr.es/moea/pages/tfm-1213/tfm_caaadasreche_jluis/)
5. Clifford, R; Taylor, R. 2008. Bioestadística. Editorial Pearson Educación.
6. Cortinas, P; Juez, P; Muñoz, A; Santos, J. 2003. Diseño de encuestas para estudios de mercado- Técnicas de muestreo y análisis multivariante. Editorial Centro De Estudios Ramón Areces, S.A.
7. Hair, J; Anderson Junior, R; Tatham, R; Black, W; 1999. Análisis Multivariantes. Quinta Edición. Editorial Prentice Hall International, Inc.
8. Jiménez, E; Manzano, J. 2005. Análisis Multivariado Aplicado. Multicolinealidad. Editorial Thomson. 238 p
9. Johnson, D. 2000. Métodos Multivariados Aplicados al Análisis de Datos. International Thomson Editores, S.A.

10. Pedret, R. 2000. Herramientas para segmentar mercados y posicionar productos. Ediciones Deusto S.A.
11. Pérez, C. 2004. Técnicas estadísticas con spss. Editorial Pearson Educación, S.A
12. Salazar, A. 2011. Modelos de respuesta discreta con R y aplicación con datos reales. Granada, España. Consultado 15 de agosto de 2014. Disponible en <http://masteres.ugr.es/moea/pages/tfm1011/modelosderespuestadiscretaenryaplicacion>.
13. Silva L. y Barroso I., 2005. Regresión logística. Construcción y evaluación de modelos. Editorial La Muralla S.A.
14. SPSS. 2003. Análisis Estadístico Avanzado. Regresión Logística Binaria 3-1,3-2p.
15. Vera, D.2007. Parto podálico vía vaginal vs vía cesárea: Hospital María Auxiliadora – 2005. Tesis Médico especialista en ginecología y obstetricia. Lima, PE
16. Wayne, D. 2006. Bioestadística. Base para el análisis de las ciencias de la salud. Editorial Limusa, S.A.

## **VIII. ANEXOS**

## ANEXO 1

ID	Edad de la madre	Edad gestacional	NBB	PESO	Talla	PC	Comp. Gestacion	Tipo de parto	Y predecido	Pcorte (0.3)	%Predicciones correctas	Verdadero	Falso	% TBC	% TMC
1	28	39	2	3.25	50	35	0	1	0.10	0	Falso	167	28	85.64	14.36
2	29	40	3	4.2	52	36	0	0	0.11	0	Verdadero				
3	20	39	2	3.83	51	35	0	0	0.10	0	Verdadero				
4	24	38	2	3.26	49	34	0	0	0.11	0	Verdadero				
5	34	40	2	2.79	48	33	0	0	0.04	0	Verdadero				
6	21	39	1	2.74	48	33	0	0	0.05	0	Verdadero				
7	26	38	2	3.4	49	33	0	0	0.09	0	Verdadero				
8	24	40	1	3.1	48	35	0	0	0.11	0	Verdadero				
9	39	39	4	3.4	49	33	0	0	0.06	0	Verdadero				
10	31	41	2	3.1	49	32	0	0	0.02	0	Verdadero				
11	35	39	2	3.26	50	34	1	1	0.99	1	Verdadero				
12	28	39	2	3.41	52	34	0	0	0.06	0	Verdadero				
13	23	39	1	3.28	51	34	0	0	0.07	0	Verdadero				
14	30	39	4	3.57	50	33	0	0	0.03	0	Verdadero				
15	18	39	1	4.06	52	36	0	0	0.21	0	Verdadero				
16	30	40	3	3.55	49	36	0	0	0.16	0	Verdadero				
17	33	40	3	3.59	52	32	0	0	0.01	0	Verdadero				
18	36	40	2	3.25	48	34	1	1	0.99	1	Verdadero				
19	19	40	1	3.1	49	32	0	0	0.02	0	Verdadero				
20	37	40	3	3.94	52	36	0	0	0.19	0	Verdadero				
21	20	40	1	3.16	49	35	0	0	0.07	0	Verdadero				
22	27	39	2	3.92	52	35	0	0	0.14	0	Verdadero				
23	26	39	2	3.24	49	36	0	0	0.17	0	Verdadero				
24	20	40	1	3.57	50	34	0	0	0.07	0	Verdadero				
25	32	38	1	3.2	48	33	0	1	0.23	0	Falso				
26	26	39	2	3.06	49	33	0	1	0.04	0	Falso				
27	25	41	1	3.35	51	33	0	0	0.03	0	Verdadero				
28	22	40	1	3.42	50	34	0	0	0.07	0	Verdadero				
29	35	40	2	3.36	50	34	0	0	0.10	0	Verdadero				
30	24	40	1	2.97	49	34	0	0	0.06	0	Verdadero				
31	33	39	1	3.79	51	35	0	1	0.35	1	Verdadero				
32	27	40	2	3.87	54	38	0	0	0.23	0	Verdadero				
33	26	39	1	2.93	48	33	0	0	0.08	0	Verdadero				
34	29	39	1	3.56	51	34	1	1	0.99	1	Verdadero				
35	29	39	2	3.3	49	34	0	0	0.11	0	Verdadero				
36	22	38	1	3.16	51	32	0	0	0.03	0	Verdadero				
37	22	37	1	2.7	47	33	0	1	0.12	0	Falso				
38	18	39	1	3.55	51	35	0	0	0.11	0	Verdadero				
39	26	39	1	3.01	49	33	0	0	0.07	0	Verdadero				
40	36	40	2	2.9	48	31	0	0	0.02	0	Verdadero				
41	27	39	1	3.5	45	33	0	0	0.25	0	Verdadero				
42	24	39	2	3.52	51	35	0	0	0.10	0	Verdadero				
43	29	41	2	3.6	51	35	0	0	0.07	0	Verdadero				
44	28	40	1	3.76	52	37	0	1	0.33	1	Verdadero				
45	35	39	2	4.11	51	36	0	0	0.43	1	Falso				
46	25	39	1	2.99	48	33	0	0	0.08	0	Verdadero				
47	18	39	1	3.25	50	33	0	0	0.04	0	Verdadero				
48	21	39	1	3.5	50	35	0	0	0.16	0	Verdadero				
49	34	40	1	3.15	51	33	0	0	0.06	0	Verdadero				
50	28	40	1	3.43	49	36	0	1	0.30	0	Falso				
51	22	40	1	3.38	49	34	0	1	0.08	0	Falso				
52	18	39	1	3	51	33	0	0	0.02	0	Verdadero				
53	22	40	2	3.33	49	34	0	0	0.05	0	Verdadero				
54	19	40	1	3.25	49	35	0	0	0.10	0	Verdadero				
55	26	39	1	3.51	48	35	0	1	0.31	1	Verdadero				
56	30	39	2	3.58	51	34	0	0	0.10	0	Verdadero				
57	22	39	2	2.83	48	34	0	0	0.06	0	Verdadero				
58	24	39	1	3.28	52	34	0	0	0.06	0	Verdadero				
59	25	38	3	3.32	47	35	1	1	0.99	1	Verdadero				
60	35	38	2	3.04	49	33	0	0	0.11	0	Verdadero				
61	25	39	3	3.19	48	33	0	0	0.03	0	Verdadero				
62	29	39	2	3.45	49	35	0	0	0.16	0	Verdadero				
63	18	39	1	3.25	48	32	0	0	0.04	0	Verdadero				
64	26	39	4	3.42	51	35	1	1	0.96	1	Verdadero				
65	19	39	2	3.22	49	34	0	0	0.05	0	Verdadero				
66	31	39	2	3.6	49	34	0	0	0.16	0	Verdadero				
67	23	40	1	3.86	53	35	1	1	0.99	1	Verdadero				
68	27	41	1	3.8	53	36	0	1	0.13	0	Falso				
69	26	40	2	3.29	48	35	0	0	0.10	0	Verdadero				
70	24	40	1	3.6	50	33	0	0	0.06	0	Verdadero				

ID	Edad de la madre	Edad gestacional	NBB	PES O	Talla	PC	Comp. Gestacion	Tipo de parto	Y predecido	Pcorte (0.3)	%Prediccion es correctas	Verdadero	Falso	% TBC	% TMC
71	37	38	4	3.72	52	36	0	0	0.23	0	Verdadero				
72	24	39	2	3.05	49	35	0	0	0.10	0	Verdadero				
73	32	39	2	3.3	50	33	0	0	0.07	0	Verdadero				
74	24	39	1	3.37	51	35	0	0	0.12	0	Verdadero				
75	20	40	1	3.22	49	34	0	0	0.06	0	Verdadero				
76	35	40	2	3.5	51	35	1	1	0.99	1	Verdadero				
77	25	39	2	3.55	50	36	1	1	0.99	1	Verdadero				
78	22	40	2	3.73	45	37	1	1	1.00	1	Verdadero				
79	33	40	3	3.84	50	36	1	1	0.99	1	Verdadero				
80	27	39	2	3.56	51	36	0	1	0.17	0	Falso				
81	20	39	2	3.1	50	35	0	0	0.05	0	Verdadero				
82	19	37	1	2.89	45	33	1	1	0.99	1	Verdadero				
83	35	39	3	3.2	50	34	0	0	0.07	0	Verdadero				
84	32	39	1	4	54	38	1	1	1.00	1	Verdadero				
85	20	39	1	3.3	50	34	0	0	0.08	0	Verdadero				
86	29	40	2	3.98	53	35	0	0	0.08	0	Verdadero				
87	21	38	2	3.12	50	35	0	0	0.10	0	Verdadero				
88	22	39	2	2.99	47	34	0	0	0.08	0	Verdadero				
89	24	39	1	3.16	50	33	0	0	0.05	0	Verdadero				
90	20	40	1	3.24	49	35	0	0	0.10	0	Verdadero				
91	18	39	1	3.42	50	33	0	0	0.04	0	Verdadero				
92	31	41	1	4.11	53	36	0	0	0.17	0	Verdadero				
93	29	38	2	2.87	48	32	0	0	0.05	0	Verdadero				
94	19	37	1	2.71	47	33	0	1	0.08	0	Falso				
95	24	39	1	3.6	51	35	0	0	0.17	0	Verdadero				
96	22	40	2	3.45	50	34	0	0	0.04	0	Verdadero				
97	29	40	3	3.48	52	35	0	0	0.04	0	Verdadero				
98	23	40	1	4	51	36	0	1	0.20	0	Falso				
99	24	38	1	3.14	52	34	0	0	0.06	0	Verdadero				
100	27	39	2	3.07	50	33	0	0	0.04	0	Verdadero				
101	37	39	2	3.65	51	35	0	0	0.20	0	Verdadero				
102	26	41	1	3.34	52	35	0	1	0.06	0	Falso				
103	24	38	3	3.62	51	36	1	1	0.99	1	Verdadero				
104	33	38	2	3.1	50	32	0	0	0.05	0	Verdadero				
105	23	39	2	2.84	49	34	1	1	0.97	1	Verdadero				
106	36	40	3	3.99	50	37	0	0	0.33	1	Falso				
107	20	38	1	3.19	49	34	0	1	0.12	0	Falso				
108	20	40	2	3.74	51	35	0	0	0.06	0	Verdadero				
109	24	39	2	3.2	50	34	0	1	0.06	0	Falso				
110	35	40	3	4.2	51	36	0	1	0.25	0	Falso				
111	20	40	1	3.58	52	35	0	0	0.07	0	Verdadero				
112	20	37	1	3.18	51	33	0	0	0.05	0	Verdadero				
113	24	40	2	3.75	52	37	0	0	0.17	0	Verdadero				
114	31	39	2	3.11	49	34	0	0	0.10	0	Verdadero				
115	19	39	1	2.69	48	32	0	0	0.02	0	Verdadero				
116	36	39	4	3.75	52	35	0	0	0.08	0	Verdadero				
117	38	40	3	2.93	49	32	0	0	0.02	0	Verdadero				
118	35	39	3	3.53	48	35	1	1	1.00	1	Verdadero				
119	32	38	3	3.98	51	34	0	1	0.14	0	Falso				
120	25	40	1	3.46	51	34	0	0	0.07	0	Verdadero				
121	19	39	1	3.3	50	35	0	0	0.13	0	Verdadero				
122	29	39	3	3.76	51	36	0	0	0.19	0	Verdadero				
123	24	39	3	3.14	48	34	0	0	0.04	0	Verdadero				
124	27	37	1	3.18	49	35	1	1	1.00	1	Verdadero				
125	24	37	2	3.14	52	35	1	1	0.99	1	Verdadero				
126	30	40	2	3.18	50	35	1	1	0.98	1	Verdadero				
127	23	39	1	3.1	48	33	0	0	0.07	0	Verdadero				
128	38	39	4	3.56	50	36	0	0	0.16	0	Verdadero				
129	20	39	1	4.15	53	35	0	0	0.11	0	Verdadero				
130	18	40	1	3.79	50	35	0	0	0.09	0	Verdadero				
131	28	38	3	3.22	50	36	1	1	0.99	1	Verdadero				
132	24	39	2	3.87	52	36	0	0	0.19	0	Verdadero				
133	24	39	1	2.79	51	32	0	0	0.02	0	Verdadero				
134	28	39	2	3.6	50	37	1	1	1.00	1	Verdadero				
135	21	40	1	3.37	50	35	0	0	0.10	0	Verdadero				
136	35	41	2	3.79	51	36	0	1	0.20	0	Falso				
137	21	39	1	3.1	51	34	0	0	0.04	0	Verdadero				
138	23	39	2	3.96	53	35	0	0	0.07	0	Verdadero				
139	26	39	1	2.89	49	33	0	0	0.06	0	Verdadero				
140	31	40	1	3.11	50	34	0	0	0.09	0	Verdadero				

...continuación

ID	Edad de la madre	Edad gestacional	NBB	Peso	Talla	PC	Comp. Gestacion	Tipo de parto	Y predecido	Pcorte (0.3)	%Predicciones correctas	Verdadero	Falso	% TBC	% TMC
141	30	40	4	3.93	52	36	1	1	0.98	1	Verdadero				
142	35	39	2	3.78	52	35	0	0	0.20	0	Verdadero				
143	18	38	1	2.69	48	32	0	0	0.03	0	Verdadero				
144	27	40	3	3.54	51	35	0	0	0.05	0	Verdadero				
145	22	39	1	3.59	51	36	0	1	0.22	0	Falso				
146	25	39	1	3.84	50	36	1	1	1.00	1	Verdadero				
147	19	40	1	3.93	53	37	0	0	0.19	0	Verdadero				
148	22	39	1	3.53	52	35	0	0	0.11	0	Verdadero				
149	19	39	2	4	52	35	0	0	0.07	0	Verdadero				
150	21	40	1	3.56	50	34	1	1	0.98	1	Verdadero				
151	25	40	1	3.39	49	34	0	1	0.10	0	Falso				
152	20	39	2	3.3	49	34	0	0	0.06	0	Verdadero				
153	19	40	1	3.72	52	35	0	0	0.08	0	Verdadero				
154	28	40	2	3.08	49	32	0	1	0.02	0	Falso				
155	21	39	1	3.43	50	35	1	1	0.99	1	Verdadero				
156	22	41	1	4.08	53	37	0	0	0.18	0	Verdadero				
157	26	40	1	3.63	52	35	0	0	0.12	0	Verdadero				
158	18	39	1	2.7	48	33	0	0	0.04	0	Verdadero				
159	29	39	2	3.45	48	34	0	0	0.15	0	Verdadero				
160	22	40	1	3.07	49	33	0	0	0.04	0	Verdadero				
161	25	38	3	3.49	50	36	0	1	0.18	0	Falso				
162	24	40	1	3.26	49	33	0	0	0.05	0	Verdadero				
163	20	39	2	4.06	53	35	0	0	0.08	0	Verdadero				
164	20	39	1	3.63	51	35	0	0	0.10	0	Verdadero				
165	27	39	4	3.76	51	34	0	0	0.03	0	Verdadero				
166	38	40	3	3.04	51	34	0	0	0.04	0	Verdadero				
167	28	38	2	2.72	49	32	0	1	0.03	0	Falso				
168	20	39	4	3.77	52	36	0	0	0.05	0	Verdadero				
169	19	39	1	3.57	50	36	0	0	0.20	0	Verdadero				
170	31	38	1	3.8	50	37	0	1	0.69	1	Verdadero				
171	38	39	2	3.27	48	34	1	1	0.99	1	Verdadero				
172	32	40	1	3.42	48	35	0	1	0.30	0	Falso				
173	28	37	1	3.57	48	37	1	1	1.00	1	Verdadero				
174	28	39	2	3.79	53	36	0	0	0.17	0	Verdadero				
175	34	38	2	3.37	52	36	0	0	0.28	0	Verdadero				
176	28	39	2	4.08	52	37	0	0	0.31	1	Falso				
177	27	38	1	3.08	50	33	0	1	0.09	0	Falso				
178	31	40	1	3.87	53	36	0	0	0.24	0	Verdadero				
179	31	39	2	3.09	49	34	1	1	0.99	1	Verdadero				
180	27	39	2	3.64	51	35	0	0	0.13	0	Verdadero				
181	25	38	1	3.5	50	33	0	0	0.11	0	Verdadero				
182	32	36	3	3.34	50	35	1	1	1.00	1	Verdadero				
183	21	40	1	3.52	51	35	0	0	0.07	0	Verdadero				
184	25	39	4	2.98	49	34	0	0	0.02	0	Verdadero				
185	23	37	1	2.65	48	32	1	1	0.98	1	Verdadero				
186	20	38	1	3.25	45	34	1	1	1.00	1	Verdadero				
187	18	38	2	4.17	51	37	0	1	0.39	1	Verdadero				
188	20	39	1	3.06	49	34	0	0	0.08	0	Verdadero				
189	19	39	1	3.62	51	35	1	1	0.99	1	Verdadero				
190	26	40	1	3.67	51	36	1	1	0.99	1	Verdadero				
191	30	40	3	3.08	51	35	0	0	0.04	0	Verdadero				
192	19	40	1	2.9	50	33	0	0	0.02	0	Verdadero				
193	34	40	1	3.24	49	34	1	1	0.99	1	Verdadero				
194	25	39	2	3.57	52	36	0	1	0.14	0	Falso				
195	21	40	1	3.44	50	36	0	1	0.17	0	Falso				

## ANEXO 2

ID	Edad de la madre	Edad gestacional	NBB	Peso	Talla	PC	Comp. Gestacion	Tipo de parto	Y predecido	Pcorte (0.3)	%Predicciones correctas	Verdadero	Falso	% TBC	% TMC
1	30	39	3	3.1	49	35	0	0	0.09	0	Verdadero	167	28	85.64	14.36
2	32	40	3	3.85	51	33	0	0	0.03	0	Verdadero				
3	21	37	2	3.55	50	35	0	0	0.17	0	Verdadero				
4	24	38	1	2.87	50	32	1	1	0.96	1	Verdadero				
5	26	40	1	3.05	50	34	0	0	0.06	0	Verdadero				
6	29	39	3	2.98	46	34	0	0	0.12	0	Verdadero				
7	22	40	2	3.57	50	36	0	0	0.09	0	Verdadero				
8	35	39	1	3.06	50	34	0	0	0.18	0	Verdadero				
9	21	38	1	4.17	52	35	0	0	0.27	0	Verdadero				
10	29	38	1	3.47	51	35	0	1	0.30	0	Falso				
11	31	37	1	3.4	50	34	0	0	0.37	1	Falso				
12	21	38	2	2.87	48	34	0	1	0.08	0	Falso				
13	34	40	4	4.01	53	34	0	0	0.02	0	Verdadero				
14	29	39	2	3.2	50	35	0	0	0.10	0	Verdadero				
15	29	39	1	3.05	49	34	0	0	0.15	0	Verdadero				
16	38	40	2	2.69	44	32	0	0	0.12	0	Verdadero				
17	29	40	2	3.44	49	35	0	0	0.14	0	Verdadero				
18	21	39	1	3.71	51	34	0	0	0.10	0	Verdadero				
19	29	40	3	3.41	51	34	0	0	0.03	0	Verdadero				
20	28	40	3	3.19	49	35	1	1	0.97	1	Verdadero				
21	27	40	1	3.29	48	36	0	1	0.29	0	Falso				
22	33	39	2	3.54	50	36	0	0	0.34	1	Falso				
23	34	40	3	3.99	54	36	0	1	0.08	0	Falso				
24	19	39	1	3.23	51	34	0	0	0.05	0	Verdadero				
25	27	40	1	3.7	50	35	1	1	0.99	1	Verdadero				
26	19	40	1	4.16	52	36	0	0	0.18	0	Verdadero				
27	20	39	1	2.85	50	34	0	0	0.05	0	Verdadero				
28	27	37	2	3.07	51	33	0	0	0.06	0	Verdadero				
29	27	39	3	3.22	50	34	0	0	0.04	0	Verdadero				
30	25	39	2	3.89	52	35	0	0	0.11	0	Verdadero				
31	26	39	2	3.81	50	34	0	0	0.12	0	Verdadero				
32	18	39	1	3.29	49	34	0	1	0.09	0	Falso				
33	19	38	1	3.42	51	35	0	0	0.12	0	Verdadero				
34	35	40	3	3.4	50	36	0	0	0.15	0	Verdadero				
35	22	40	1	3.21	48	35	0	0	0.18	0	Verdadero				
36	23	39	1	3.59	49	35	0	0	0.26	0	Verdadero				
37	21	40	1	2.98	50	34	0	0	0.04	0	Verdadero				
38	37	40	3	3.21	50	35	0	1	0.09	0	Falso				
39	19	39	1	3.71	52	35	0	0	0.10	0	Verdadero				
40	22	40	1	3.02	49	36	0	0	0.15	0	Verdadero				
41	19	40	1	2.98	46	33	0	0	0.05	0	Verdadero				
42	20	39	2	3.6	51	34	0	0	0.04	0	Verdadero				
43	21	39	2	3.43	50	34	0	0	0.04	0	Verdadero				
44	29	40	3	4.15	52	35	0	0	0.06	0	Verdadero				
45	21	40	1	3.04	45	33	0	0	0.11	0	Verdadero				
46	28	38	3	3.24	49	35	1	1	0.99	1	Verdadero				
47	21	40	1	3.42	51	35	0	0	0.08	0	Verdadero				
48	33	40	2	3.7	51	33	0	0	0.06	0	Verdadero				
49	34	40	3	3.28	49	34	0	0	0.06	0	Verdadero				
50	18	39	1	3.38	49	33	1	1	0.98	1	Verdadero				
51	20	38	1	2.85	50	32	0	0	0.03	0	Verdadero				
52	36	37	2	2.95	50	35	1	1	1.00	1	Verdadero				
53	24	39	2	3.3	50	33	0	0	0.04	0	Verdadero				
54	25	38	2	3.23	50	35	0	0	0.14	0	Verdadero				
55	35	39	2	4.14	53	35	1	1	0.99	1	Verdadero				
56	29	39	2	3.09	48	34	0	1	0.12	0	Falso				
57	30	39	1	3.68	51	35	1	1	1.00	1	Verdadero				
58	21	39	1	2.71	46	33	0	0	0.07	0	Verdadero				
59	38	39	4	3.25	51	34	0	0	0.04	0	Verdadero				
60	21	39	1	2.87	50	34	0	0	0.05	0	Verdadero				
61	26	38	1	2.96	48	33	0	0	0.14	0	Verdadero				
62	18	38	2	3.52	51	36	1	1	0.99	1	Verdadero				
63	25	39	1	4.12	50	35	0	0	0.37	1	Falso				
64	33	40	2	3.88	50	36	0	0	0.33	1	Falso				
65	20	39	2	3	50	35	0	0	0.05	0	Verdadero				
66	25	39	3	3.5	52	34	0	0	0.02	0	Verdadero				
67	33	40	2	3.82	50	35	1	1	0.99	1	Verdadero				
68	37	39	1	3.02	49	35	0	0	0.37	1	Falso				
69	25	39	3	3.73	51	36	0	0	0.13	0	Verdadero				
70	23	39	1	3.5	51	36	0	0	0.23	0	Verdadero				

...continuación

ID	Edad de la madre	Edad gestacional	NBB	Peso	Talla	PC	Comp. Gestacion	Tipo de parto	Y predecido	Pcorte (0.3)	% Predicciones correctas	Verdadero	Falso	% TBC	% TMC
71	32	39	3	3.75	51	35	0	0	0.12	0	Verdadero				
72	20	40	1	3.99	53	37	0	1	0.22	0	Falso				
73	27	39	1	3.78	53	34	0	0	0.10	0	Verdadero				
74	19	39	1	2.96	48	32	0	0	0.02	0	Verdadero				
75	32	39	2	3.07	49	33	0	0	0.07	0	Verdadero				
76	18	40	1	3.46	53	32	0	0	0.01	0	Verdadero				
77	34	38	3	3.7	50	37	0	1	0.49	1	Verdadero				
78	33	40	1	3.45	52	35	1	1	0.99	1	Verdadero				
79	35	40	1	3.32	50	35	0	0	0.25	0	Verdadero				
80	26	39	1	3.51	52	34	0	1	0.09	0	Falso				
81	33	38	4	2.79	49	33	0	0	0.02	0	Verdadero				
82	31	39	2	4.21	51	37	0	0	0.55	1	Falso				
83	22	38	1	3.1	50	34	0	0	0.11	0	Verdadero				
84	20	39	1	2.98	48	33	0	0	0.06	0	Verdadero				
85	38	40	4	4.08	54	36	0	0	0.05	0	Verdadero				
86	20	39	1	3.5	50	34	1	1	0.99	1	Verdadero				
87	21	40	2	3.74	51	34	0	0	0.04	0	Verdadero				
88	34	38	4	3.27	49	37	1	1	1.00	1	Verdadero				
89	20	39	1	3.47	47	32	0	0	0.08	0	Verdadero				
90	19	40	2	3.16	49	36	0	1	0.08	0	Falso				
91	26	38	2	3.19	52	33	0	0	0.03	0	Verdadero				
92	21	38	1	2.74	48	33	0	0	0.07	0	Verdadero				
93	38	40	3	2.82	48	35	0	0	0.11	0	Verdadero				
94	31	39	3	3.86	49	35	0	0	0.17	0	Verdadero				
95	19	39	1	3.54	53	35	0	0	0.05	0	Verdadero				
96	18	40	1	3.43	51	37	0	0	0.15	0	Verdadero				
97	26	40	2	3.26	51	34	0	1	0.02	0	Falso				
98	27	39	2	3.4	51	34	0	0	0.06	0	Verdadero				
99	20	38	2	3.36	50	34	0	0	0.05	0	Verdadero				
100	24	38	2	2.98	50	37	0	0	0.25	0	Verdadero				
101	27	39	1	3.47	50	35	0	0	0.24	0	Verdadero				
102	27	39	2	3.35	50	35	0	0	0.14	0	Verdadero				
103	21	39	2	3.52	50	34	0	0	0.05	0	Verdadero				
104	21	40	2	3.63	49	35	1	1	0.99	1	Verdadero				
105	22	40	2	4.2	50	36	0	0	0.25	0	Verdadero				
106	25	40	1	3.28	49	35	0	0	0.16	0	Verdadero				
107	31	39	2	3.05	49	34	0	0	0.10	0	Verdadero				
108	32	40	3	3.1	49	34	0	0	0.04	0	Verdadero				
109	29	39	2	3.2	50	35	0	0	0.10	0	Verdadero				
110	27	40	3	3.94	51	36	0	0	0.11	0	Verdadero				
111	22	38	2	3.38	50	34	0	0	0.08	0	Verdadero				
112	25	39	1	3.68	52	35	0	1	0.16	0	Falso				
113	18	39	2	3.64	48	35	0	0	0.15	0	Verdadero				
114	19	40	1	3.28	51	34	0	0	0.03	0	Verdadero				
115	22	40	1	3.37	51	35	0	0	0.06	0	Verdadero				
116	19	39	1	2.93	48	34	0	0	0.10	0	Verdadero				
117	26	38	3	4.15	53	35	0	0	0.12	0	Verdadero				
118	23	38	2	3.12	50	33	0	0	0.04	0	Verdadero				
119	32	39	2	3.9	52	37	0	0	0.40	1	Falso				
120	26	40	1	4.06	50	36	0	1	0.39	1	Verdadero				
121	27	40	1	3.66	51	34	0	0	0.08	0	Verdadero				
122	18	40	1	3.89	52	36	0	0	0.10	0	Verdadero				
123	22	39	1	3.44	50	33	0	0	0.06	0	Verdadero				
124	31	38	3	3.14	47	36	0	1	0.30	1	Verdadero				
125	27	39	4	3.41	49	35	0	0	0.06	0	Verdadero				
126	27	39	3	2.82	46	33	0	0	0.05	0	Verdadero				
127	24	40	1	3.26	52	36	0	0	0.08	0	Verdadero				
128	30	39	2	3.75	50	37	0	0	0.40	1	Falso				
129	30	40	2	4.01	53	37	0	0	0.23	0	Verdadero				
130	19	39	1	2.99	49	31	0	0	0.01	0	Verdadero				
131	25	39	1	3.61	49	36	1	1	1.00	1	Verdadero				
132	23	39	2	3.04	50	34	0	0	0.04	0	Verdadero				
133	30	40	1	2.99	50	35	1	1	0.99	1	Verdadero				
134	21	40	2	3.72	51	35	0	0	0.06	0	Verdadero				
135	35	39	2	3.28	50	35	0	0	0.17	0	Verdadero				
136	22	38	1	3.54	52	36	0	0	0.24	0	Verdadero				
137	36	39	2	3.5	51	36	0	0	0.31	1	Falso				
138	28	38	3	3.27	49	35	0	0	0.14	0	Verdadero				
139	29	40	3	3.95	52	35	0	0	0.06	0	Verdadero				
140	21	39	1	3.18	48	34	0	0	0.13	0	Verdadero				

...continuación

ID	Edad de la madre	Edad gestacional	NBB	Peso	Talla	PC	Comp. Gestacion	Tipo de parto	Y predicho	Pcorte (0.3)	%Predicciones correctas	Verdadero	Falso	% TBC	% TMC
141	24	40	2	3.64	51	35	0	0	0.07	0	Verdadero				
142	28	40	4	3.95	50	35	0	0	0.05	0	Verdadero				
143	25	40	1	3.83	51	35	0	0	0.16	0	Verdadero				
144	23	40	2	4.05	51	37	0	0	0.25	0	Verdadero				
145	36	39	2	3	49	32	0	0	0.05	0	Verdadero				
146	22	40	1	3.44	51	34	0	0	0.04	0	Verdadero				
147	21	39	2	3.48	51	36	0	0	0.11	0	Verdadero				
148	19	39	1	3.18	50	34	0	0	0.05	0	Verdadero				
149	23	40	1	3.46	50	33	0	0	0.05	0	Verdadero				
150	20	39	1	3.05	51	33	0	0	0.02	0	Verdadero				
151	24	39	1	3.16	50	35	0	0	0.15	0	Verdadero				
152	25	38	2	3.36	52	34	0	1	0.06	0	Falso				
153	23	40	3	3.47	50	34	0	0	0.02	0	Verdadero				
154	27	40	2	3.7	51	36	0	0	0.15	0	Verdadero				
155	32	38	2	2.91	49	35	1	1	0.99	1	Verdadero				
156	33	38	2	3.34	51	34	0	0	0.14	0	Verdadero				
157	35	40	2	3.16	48	34	0	0	0.14	0	Verdadero				
158	23	40	1	3.47	51	35	0	1	0.08	0	Falso				
159	37	38	4	3.47	51	36	0	0	0.14	0	Verdadero				
160	29	40	1	4.03	53	38	0	1	0.45	1	Verdadero				
161	20	39	1	3.15	49	37	0	0	0.32	1	Falso				
162	23	38	1	3.44	49	34	0	0	0.22	0	Verdadero				
163	29	39	3	4.02	52	37	0	0	0.19	0	Verdadero				
164	19	39	2	3.21	51	34	0	0	0.03	0	Verdadero				
165	27	39	1	3.5	50	35	0	0	0.25	0	Verdadero				
166	31	39	1	3.1	51	35	0	0	0.17	0	Verdadero				
167	33	38	3	2.95	48	34	0	0	0.12	0	Verdadero				
168	25	39	2	3.36	50	34	0	0	0.07	0	Verdadero				
169	25	39	1	2.83	50	33	0	0	0.04	0	Verdadero				
170	25	40	1	2.62	46	34	0	0	0.09	0	Verdadero				
171	31	40	1	3.1	49	33	1	1	0.98	1	Verdadero				
172	28	40	1	3.45	50	35	1	1	0.99	1	Verdadero				
173	21	38	2	3.14	50	32	0	0	0.02	0	Verdadero				
174	19	39	1	2.9	47	32	0	0	0.04	0	Verdadero				
175	35	39	2	3.31	50	36	1	1	1.00	1	Verdadero				
176	21	40	1	3.74	53	35	0	1	0.06	0	Falso				
177	24	39	3	3.76	52	37	0	0	0.10	0	Verdadero				
178	26	39	1	3.16	50	35	0	0	0.17	0	Verdadero				
179	34	39	4	3.16	49	34	0	1	0.04	0	Falso				
180	25	39	3	3.47	51	35	0	0	0.05	0	Verdadero				
181	22	39	2	3.34	52	33	0	1	0.02	0	Falso				
182	29	40	3	3.6	51	36	1	1	0.98	1	Verdadero				
183	20	40	1	3.52	50	35	0	0	0.11	0	Verdadero				
184	28	39	2	3.06	51	33	0	0	0.03	0	Verdadero				
185	24	40	1	3.86	51	34	1	1	0.99	1	Verdadero				
186	24	40	4	3.94	52	37	0	0	0.05	0	Verdadero				
187	26	40	4	3.8	50	35	0	0	0.04	0	Verdadero				
188	28	40	2	3.34	50	35	0	0	0.09	0	Verdadero				
189	23	40	2	3.84	51	37	1	1	0.99	1	Verdadero				
190	25	38	2	3.56	49	37	0	0	0.50	1	Falso				
191	25	38	1	3.5	50	34	0	0	0.20	0	Verdadero				
192	26	40	3	3.98	50	37	0	0	0.18	0	Verdadero				
193	23	38	1	3.76	52	33	0	0	0.08	0	Verdadero				
194	24	39	3	3.48	51	33	0	0	0.02	0	Verdadero				
195	23	38	1	3.42	52	35	0	0	0.15	0	Verdadero				