

Modelos de regresión en R para predicción de peso al nacer y tipo de parto

Regression models in R for predicting birth weight and type of delivery

Recibido: 23 de febrero de 2023

Aprobado: 28 de julio de 2023

Forma de citar: C.A. Mejía Rodríguez, M.A. Rincón Pinzón, M.U. Vargas Ramos, D.A. Diaz Vergel "Modelos de regresión en R para predicción de peso al nacer y tipo de parto", *Mundo Fesc*, vol 13, no. 27, pp. 152-167, 2023. <https://doi.org/10.61799/2216-0388.1349>

Carlos Alberto Mejía Rodríguez* 

Magíster en E-Learning
calbertomejia@unicesar.edu.co
Universidad Popular del Cesar
Aguachica, Colombia.

Miguel Alberto Rincón Pinzón 

Magíster en Gestión de la Tecnología Educativa
Eylen.arroyo@cecar.edu.co
Universidad Popular del Cesar
Aguachica, Colombia.

Mónica Urbina Vargas Ramos 

Magister en Educación, Especialista en Gerencia Financiera
mvargas262nab.edu.co
Montería, Colombia.

Deider Alfonso Diaz Vergel 

Magíster en Gestión de la tecnología Educativa
deideradiaz@unicesar.edu.co
Universidad Popular del Cesar
Aguachica, Colombia.

***Autor para correspondencia:**
calbertomejia@unicesar.edu.co



Modelos de regresión en R para predicción de peso al nacer y tipo de parto

Resumen

El propósito de esta investigación fue realizar un análisis estadístico predictivo de nacidos vivos en Guadalajara de Buga, Colombia, utilizando R. Se buscaba lograr una predicción acertada del peso al nacer en función de la longitud del feto, así como del tipo de parto basado en las semanas de gestación. La metodología incluyó la construcción de varios modelos de regresión, comenzando con una regresión lineal simple entre 'Peso' y 'Talla'. Posteriormente, se implementó una regresión multilínea para incorporar 'Tiempo de Gestación' como predictor adicional. Finalmente, se llevó a cabo una regresión logística para prever el tipo de parto según las semanas de gestación. Los resultados revelaron una correlación lineal positiva entre 'Talla' y 'Peso', indicando un modelo de regresión lineal viable. La ecuación resultante permitió pronosticar el aumento promedio de peso por cada centímetro adicional en 'Talla'. La regresión multilínea mostró un coeficiente de determinación del 62.8%, con 'Talla' y 'Tiempo de Gestación' como variables significativas. La regresión logística proporcionó información valiosa sobre cómo el 'Tiempo de Gestación' afecta la probabilidad de tener un parto por cesárea. La investigación estableció relaciones precisas entre las variables seleccionadas de nacimientos vivos, proporcionando valiosas herramientas predictivas a través de modelos de regresión y logísticos. Estos resultados tienen el potencial de impactar positivamente en la toma de decisiones clínicas y en la planificación obstétrica, contribuyendo significativamente a la mejora de la atención materno-neonatal.

Palabras clave: nacidos vivos, regresión lineal, regresión logística, regresión multilínea.

Regression models in R for predicting birth weight and type of delivery

Abstract

The purpose of this research was to perform a predictive statistical analysis of live births in Guadalajara de Buga, Colombia, using R. The aim was to achieve an accurate prediction of birth weight based on the length of the fetus, as well as the type of delivery based on the weeks of gestation. The methodology included the construction of several regression models, starting with a simple linear regression between 'Weight' and 'Height'. Subsequently, a multiline regression was implemented to incorporate 'Gestation Time' as an additional predictor. Finally, a logistic regression was carried out to predict the type of delivery according to the weeks of gestation. The results revealed a positive linear correlation between 'Height' and 'Weight', indicating a viable linear regression model. The resulting equation made it possible to predict the average weight gain for each additional centimeter in 'Height'. The multiline regression showed a coefficient of determination of 62.8%, with 'Height' and 'Gestation Time' as significant variables. Logistic regression provided valuable information on how 'Gestation Time' affects the probability of having a cesarean delivery. The research established precise relationships between selected live birth variables, providing valuable predictive tools through regression and logistic models. These results have the potential to positively impact clinical decision-making and obstetric planning, significantly contributing to the improvement of maternal-neonatal care.

Keywords: Live births, Linear regression, Logistic regression, Multilinear regression.

Introducción

La somatometría del recién nacido se enfoca en medir el peso, longitud y perímetro craneal para evaluar si están dentro de los rangos normales, las desviaciones pueden indicar riesgo de complicaciones metabólicas y fisiológicas, pero también ayudar en la identificación de complicaciones obstétricas y de parto [1].

Esta investigación busca desarrollar y evaluar modelos de regresión lineal y logística para estimar el peso del recién nacido, así como predecir el tipo de parto utilizando variables como la talla y el tiempo de gestación. Se empleó el lenguaje de programación R para construir los modelos, destacándolo como una herramienta libre que, mediante la complementación de diversos paquetes, facilita comandos y funcionalidades similares a las de los principales softwares estadísticos comerciales [2].

La creación de modelos de regresión lineal involucra la evaluación de la relación lineal entre variables cuantitativas, seguida de la construcción de la recta de regresión para realizar predicciones con intervalos de confianza [3]. Por otro lado, los modelos de regresión logística se caracterizan por tener una variable de respuesta dicotómica, lo que significa que puede tomar solamente dos valores, comúnmente referidos como éxito y fracaso [4].

Estudios previos, respaldan la utilidad de los modelos de regresión como estadística práctica para ciencias de la salud [5]. De igual forma existen investigaciones que defienden la búsqueda de relaciones causa-efecto en la indagación médica y destacan el impacto del aprendizaje automático, así como la recopilación de datos en la práctica clínica [6]. Por otro lado, hay quienes enfatiza la importancia de identificar factores determinantes en el peso de los recién nacidos mediante la regresión logística multinomial [7]. La importancia de estos estudios radica en su potencial para lograr una mejor atención clínica utilizando la ciencia de datos.

El análisis de la relación entre variables es fundamental en diversas disciplinas. Por ejemplo, puede ser necesario comprender la relación entre los precios de un producto y sus ventas, o investigar cómo la capacidad de un motor afecta la eficiencia de combustible en un automóvil, estos estudios se llevan a cabo mediante análisis de regresión [8] lo que marca su importancia en cualquier área.

Regresión lineal simple

La regresión lineal satisface la demanda de modelos cuantitativos para establecer relaciones entre variables numéricas, facilitando la formulación de modelos efectivos para prever el comportamiento de sistemas y procesos en diversas áreas y actividades humanas. Este enfoque emerge como una herramienta valiosa en la construcción de relaciones estadísticas en una amplia gama de campos [9].

Para realizar un análisis de regresión lineal simple, se requiere un número adecuado de observaciones, variables numéricas, distribuciones normales y homogéneas, y que las predicciones estén dentro del rango conocido de la variable independiente. Además, los residuos deben seguir una distribución normal con media cero y desviación típica S . Estos criterios se evalúan mediante gráficos de residuos estandarizados en unidades Z [10].

Los modelos de regresión lineal simple [11] tienen como objetivo crear una ecuación que relacione la variable dependiente (Y) con la variable independiente (X) de manera lineal. En la ecuación $f(x) = \beta_0 + \beta_1x + \epsilon$, β_1 representa el coeficiente de regresión, indicando el cambio esperado en Y por unidad de aumento en X , mientras que β_0 denota el valor esperado de Y cuando X es cero. Esta ecuación se conoce como la ecuación regresiva estimada o de predicción para los valores de Y , y se presenta en la siguiente ecuación:

$$\hat{y} = b_0 + b_1X \quad \text{con } b_0 \text{ y } b_1 \text{ constantes.} \quad (1)$$

Para evaluar el modelo [11] se analiza la varianza de la variable dependiente. Si el modelo es adecuado, los valores predichos (\hat{y}_i) deberían aproximarse a los valores reales (y_i) de la variable Y , lo que implica que la diferencia ($y_i - \hat{y}_i = e_i$) tienda a cero. Esto se refleja en un coeficiente de determinación (R^2) cercano a uno.

Regresión lineal múltiple

En esta sección, se aborda el cálculo de una ecuación que describe la relación entre una variable dependiente (Y) y múltiples variables independientes (X). Los coeficientes ($\beta_0, \beta_1, \beta_2, \dots, \beta_n$) se determinan mediante la resolución de ecuaciones o el uso de software estadístico especializado [12].

En un contexto médico, al evaluar la influencia de la edad del paciente (X_1), el historial de enfermedades previas (X_2), y el nivel de actividad física (X_3) en la predicción de resultados de salud, podemos crear una ecuación hipotética para capturar esta relación. Por ejemplo:

$$Y = \beta_0 + \beta_1 \cdot X_1 + \beta_2 \cdot X_2 + \beta_3 \cdot X_3 + \epsilon \quad (2)$$

Donde:

Y representa el resultado de salud o variable dependiente que se desea predecir.

- β_0 es el término de intersección.
- β_1, β_2 y β_3 son los coeficientes de regresión asociados a las variables X_1, X_2 y X_3 respectivamente.
- ϵ representa el término de error que captura las discrepancias no explicadas por el modelo.

Regresión logística

En algunos estudios, los eventos tienen solo dos resultados posibles, donde una variable Y se utiliza de manera dicotómica, con 1 representando la ocurrencia y 0 la no ocurrencia [13]. En un caso específico, en el estudio del tipo de parto en embarazadas, se analiza si las variables X_1 (tiempo de gestación) y X_2 (tamaño del feto) están relacionadas con la probabilidad de cesárea, usando 1 para cesárea y 0 para parto natural como categorías de la variable dependiente, a través de la regresión logística, se intenta estimar la probabilidad de que el evento de interés ocurra en relación con ciertas variables [13]. Por lo tanto, la ecuación de un modelo logístico basado en el ejemplo se podría expresar de la siguiente manera:

$$P(Y=1) = \frac{1}{1+e^{-(\beta_0+\beta_1X_1+\beta_2X_2)}} \quad (3)$$

Donde:

$P(Y=1)$ es la probabilidad de que el parto sea por cesárea.

- β_0 es el intercepto.
- β_1 y β_2 son los coeficientes asociados a las variables independientes X_1 (tiempo de gestación) y X_2 (tamaño del feto), respectivamente.
- e es la base del logaritmo natural.

Este modelo logístico estima la probabilidad de que el parto sea por cesárea en función de las variables independientes X_1 y X_2 , permitiendo analizar cómo influyen el tiempo de gestación y el tamaño del feto en la probabilidad de que ocurra una cesárea.

Materiales y métodos

El enfoque empleado en la investigación es principalmente cuantitativo, en este sentido se describieron las variables de estudios sin manipular los resultados [14] ante lo cual se busca la formulación de leyes generales, y para lograrlo, se recopilan grandes cantidades de datos de la población, eligiéndose muestras representativas para validar los descubrimientos [15]. El trabajo se centra en la construcción y aplicación de modelos de regresión, que son técnicas estadísticas cuantitativas para analizar y cuantificar las relaciones entre variables numéricas [10].

La investigación utiliza datos de nacidos vivos en Guadalajara de Buga, Colombia, del año 2016, para elaborar modelos estadísticos en R y prever resultados obstétricos. Se implementaron tres modelos: una regresión lineal simple para anticipar el peso al nacer a partir de la talla, una regresión multilíneal que incorpora el tiempo de gestación, y una regresión logística para evaluar la influencia del tiempo de gestación en la probabilidad de cesárea.

Los datos se obtuvieron del portal Datos Abiertos del gobierno colombiano, la tabla 1 proporciona un resumen de las variables utilizadas en el análisis.

Tabla 1. Resumen De Variables De La Base De Datos

Nombre	Descripción	Tipo de dato
Peso	Peso del recién nacido en gramos	Número
Talla	Estatura del recién nacido en centímetros	Número
Tiempo de Gestación	Número de semanas de embarazo que tenía la madre el día del nacimiento	Número
Tipo de Parto	Tipo de parto realizado	Texto simple

Resultados y discusión

R es un lenguaje de programación de código abierto para análisis estadísticos, se emplea RStudio Cloud, plataforma en línea que proporciona un entorno de desarrollo integral para R, simplificando la creación de scripts, manipulación de datos y visualización.

En **R**, es un lenguaje orientado a objetos con tipos de dato de características distintas, que incluyen datos numéricos que abarcan números reales y complejos, datos de caracteres que representan cadenas de texto y datos lógicos que solo pueden ser verdaderos o falsos [16]. Es así que en R, la documentación de scripts mediante comentarios es esencial para garantizar la replicabilidad de resultados [17]. A continuación, se presentan los comandos clave y los resultados obtenidos al construir los modelos estadísticos propuestos.

Modelo 1: Regresión lineal simple

En el primer modelo, se emplean las variables "Peso" y "Talla" del recién nacido debido a su alta correlación. El objetivo es predecir el peso a partir de la talla. Una vez construido el modelo, su aplicación será práctica, ya que permitirá estimar el peso a partir de la talla obtenida mediante ecografía antes del nacimiento.

En **R**, la importación de datos comienza con la lectura de un archivo externo que debe seguir un formato específico [18] el archivo debe tener los nombres de las variables en la primera línea y los valores de las variables en líneas subsiguientes, facilitando así su lectura y procesamiento. A continuación, se muestra el código para importar el dataset.

```

# Se importa el dataset
data = read.csv("Nacidos_Vivos.csv")

#se presenta un resumen estadístico de las variables seleccionadas
summary(data [, c("Peso", "Talla")])
##   Peso      Talla
## Min.   :815   Min.   :31.00
## 1st Qu.:2905  1st Qu.:49.00
## Median :3180  Median :50.00
## Mean   :3146  Mean   :49.93
## 3rd Qu.:3460  3rd Qu.:52.00
## Max.   :4775  Max.   :59.00
  
```

Se observa que la variable "Peso" tiene un rango que va desde 815 hasta 4775 gramos, con una media de 3180 gramos. Mientras en la variable "Talla" (en centímetros), los valores oscilan entre 31 y 59, con una media de 50. Ahora procedemos al cálculo de la correlación.

```

Correlation<-cor(data$Peso,data$Talla)
correlation
## [1] 0.7572996
  
```

Se evidencia una correlación lineal positiva fuerte entre las variables. En el contexto del caso, se considera a la variable "Talla" como la variable independiente, que influiría en los valores de "Peso". Se procedemos entonces a construir el modelo de regresión lineal simple.

El comando básico para realizar modelos de regresión en R es "lm()". Su sintaxis básica implica especificar la variable dependiente seguida de "~" y el predictor, y luego se especifica la fuente de datos [19].

La función summary () proporciona un resumen del modelo ajustado, incluyendo la fórmula, los residuales, estimaciones de parámetros, error estándar, estadísticas t y p-valores para probar hipótesis, estimación de σ y grados de libertad, R2 habitual y ajustado, y estadístico F con sus grados de libertad y p-valor [20].

```

# Creación del modelo de regresión lineal simple
regresionlineal<-lm(Peso ~ Talla, data = data)

# Visualización básica de los valores estimados de  $\beta_0$  y  $\beta_1$ 
regresionlineal
##
## Call:
## zlm (formula = Peso ~ Talla, data = data)
  
```

```

##
## Coefficients:
## (Intercept)    Talla
##   -3329.3      129.7
# Uso de la función summary() para obtener un resumen detallado del modelo
summary(regresionlineal)
##
## Call:
## lm(formula = Peso ~ Talla, data = data)
##
## Residuals:
##   Min     1Q   Median     3Q    Max
## -1192.80 -214.70  -4.92  195.15 2489.63
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) -3329.342   147.916  -22.51 <2e-16 ***
## Talla       129.697     2.958   43.85 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 324.3 on 1430 degrees of freedom
## Multiple R-squared:  0.5735, Adjusted R-squared:  0.5732
## F-statistic: 1923 on 1 and 1430 DF, p-value: < 2.2e-16

```

Análisis del Modelo 1

El análisis de regresión involucra conceptos matemáticos fundamentales, especialmente en relación con las funciones que describen la naturaleza de la relación entre variables. Además, implica una serie de procedimientos de diagnóstico, como el análisis de residuos y la evaluación de la bondad del ajuste, que proporcionan información sobre la estabilidad y adecuación del modelo [21].

Los resultados muestran un modelo de regresión lineal simple con Talla como variable independiente y Peso como variable dependiente. El bajo p-valor de Talla indica una relación significativa. El coeficiente R^2 de 0.5735 explica el 57.35% de la variabilidad en el Peso, indicando una fuerte relación positiva. La ecuación de la recta es $y = -3329.342 + 129.697x$, lo que significa que, en promedio, por cada centímetro adicional en Talla, se espera un aumento de 129.697 gramos en Peso. Se proporciona el siguiente código y la Figura 1 la cual muestra los residuos del modelo, ofreciendo una representación visual de cómo las predicciones se ajustan a los datos en relación con las variables "Talla" y "Peso".

código que genera un gráfico del modelo y sus residuales.

```
data$predicciones <- predict(regresionlineal)
ggplot(data, aes(x=Talla, y=Peso)) +
  geom_smooth(method="lm", se=FALSE, color="lightgrey") +
  geom_segment(aes(xend=Talla, yend=predicciones), col='red', lty='dashed') +
  geom_point() +
  geom_point(aes(y=predicciones), col='red') +
  theme_light()
```

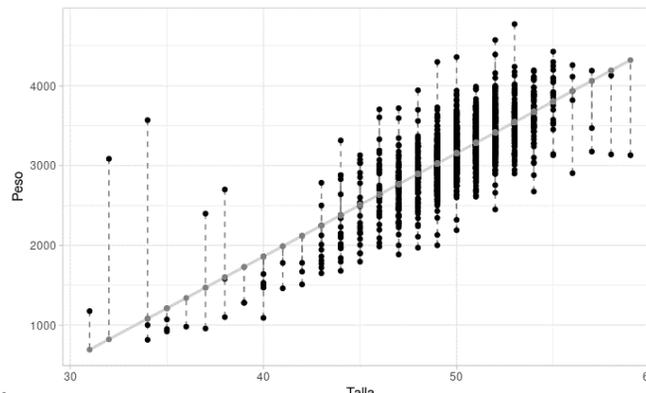


Figura 1. Gráfico De Residuales y Ajuste En Un Modelo De Regresión Lineal Simple

La gráfica muestra una línea de regresión positiva que confirma el aumento en el "Peso" con la "Talla", siguiendo el coeficiente positivo del modelo de regresión. Líneas punteadas conectan puntos de datos con predicciones basadas en "Talla" y "Peso". Puntos sobre la línea representan las estimaciones del modelo, respaldando la relación positiva y significativa entre "Talla" y "Peso".

Modelo 2: Regresión multilínea

Se aplica una regresión lineal múltiple para analizar la influencia conjunta de "Talla" y "Tiempo de Gestación" en el "Peso" de recién nacidos. A continuación, se proporciona el código para generar un resumen estadístico de las variables o columnas seleccionadas en el modelo.

Se presenta un resumen estadístico de las variables seleccionadas

```
summary(data[, c("Peso", "Talla", "Tiempo.de.Gestación")])
##   Peso   Talla  Tiempo.de.Gestación
## Min.  :815   Min.  :31.00   Min.  :24.00
## 1st Qu.:2905 1st Qu.:49.00   1st Qu.:38.00
## Median :3180 Median :50.00   Median :39.00
## Mean   :3146 Mean   :49.93   Mean   :38.34
## 3rd Qu.:3460 3rd Qu.:52.00   3rd Qu.:39.00
## Max.   :4775 Max.   :59.00   Max.   :42.00
```

Se aprecia que el rango de tiempo de gestación es de 24 a 42 semanas, con una mediana

de 39 semanas. Ahora para explorar la relación entre varias variables, se puede usar el paquete "scatterplot3d" en R. Se presenta a continuación el código y la Figura 2, que ofrece una vista de cómo las variables se distribuyen en un espacio tridimensional.

```
# Creación de un gráfico 3D para visualizar la relación entre las variables
library(scatterplot3d)
attach(data)
scatterplot3d(x=Talla, y=Tiempo.de.Gestación, z=Peso, pch=16, cex.lab=1,
highlight.3d=TRUE, type="h", xlab='Talla (centímetros)',
ylab='Tiempo de Gestación (semanas)', zlab='Peso (gramos)')
```

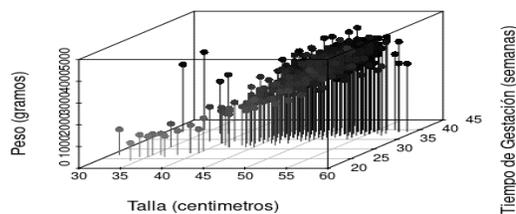


Figura 2. Gráfico Exploratorio De La Relación Entre Varias Variable

En la figura anterior se ve claramente que a medida que aumenta la Talla y el Tiempo de Gestación el Peso de los recién nacidos tiende a ser mayor.

En la construcción del modelo, se representa el modelo ajustado como sigue:

$\mu_i = \beta_0 + \beta_1 * Talla + \beta_2 * Tiempo$, con una varianza constante (σ^2).

Se procede a definir el modelo de regresión lineal múltiple utilizando el software R.

```
# Creación del modelo de regresión lineal múltiple
mod2 <- lm(Peso ~ Talla + Tiempo.de.Gestación, data=data)
# Uso de la función summary() para obtener un resumen detallado del modelo
summary(mod2)
##
## Call:
## lm(formula = Peso ~ Talla + Tiempo.de.Gestación, data = data)
##
## Residuals:
##   Min     1Q   Median     3Q    Max
## -1029.05 -196.26 -15.52  189.45 1832.71
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -4776.415   170.652  -27.99 <2e-16 ***
## Talla           96.853     3.578   27.07 <2e-16 ***
```

```
## Tiempo.de.Gestación 80.518 5.570 14.46 <2e-16 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 303 on 1429 degrees of freedom
## Multiple R-squared: 0.6279, Adjusted R-squared: 0.6274
## F-statistic: 1206 on 2 and 1429 DF, p-value: < 2.2e-16
```

Análisis del Modelo 2

Este modelo de regresión lineal múltiple usa "Talla" y "Tiempo de Gestación" para predecir "Peso". El coeficiente de determinación múltiple (R^2) es 0,6279, explicando alrededor del 62,8% de la variabilidad en "Peso". Ambas variables tienen coeficientes de regresión positivos y altamente significativos, indicando que un aumento en ambas se relaciona con un aumento en "Peso". El modelo es estadísticamente significativo, y su representación es:

Peso = -4776.415 + 96.853 * Talla + 80.518 * Tiempo_Gestacion, con $\sigma^2 = 303$.

A continuación, el código y la Figura 3 permiten visualizar la incorporación de un plano de regresión para visualizar cómo se ajustan los datos al modelo.

```
# Crear un gráfico 3D con puntos y un plano de regresión
mi_3d <- scatterplot3d(x=Talla, y=Tiempo.de.Gestación, z=Peso, pch=16, cex.lab=1,highlight.3d=TRUE,
type="h", xlab='Talla (cm)', ylab='T. Gestación (semanas)', zlab='Peso (gramos)')
## Agregar un plano de regresión al gráfico utilizando el modelo mod2
mi_3d$plane3d(mod2, lty.box="solid", col='mediumblue')
```

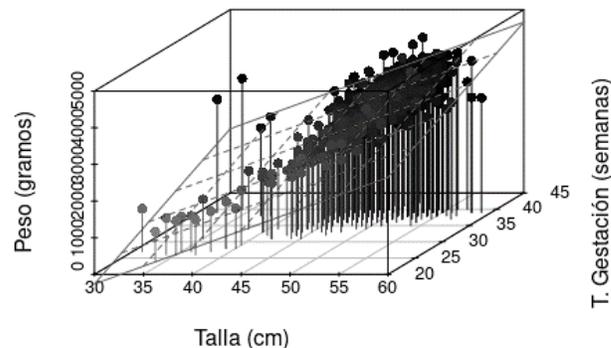


Figura 3. Plano De Regresión Que Representa El Modelo Multilínea

El plano en el gráfico 3D refleja el modelo de regresión múltiple ajustado, inclinándose en las direcciones positivas y altamente significativas de los coeficientes de regresión de Talla y Tiempo de Gestación. Esto ilustra el aumento del Peso de los recién nacidos con valores más altos de Talla y Tiempo de Gestación.

Modelo 3: Regresión logística

La regresión logística es un algoritmo de clasificación que estima la probabilidad de pertenecer a una clase. En el conjunto de datos actual, se busca crear un modelo para predecir el tipo de parto (cesárea o espontáneo) basado en las semanas de gestación. Para comenzar, se realizará un resumen de la distribución de valores en la variable "Tipo de Parto".

#Muestra la distribución de valores en la variable "Tipo de Parto"

```
table(data$Tipo.de.Parto)
```

```
##
```

```
##   CESÁREA   ESPONTÁNEO
```

```
##    497     935
```

El análisis de la variable "Tipo de Parto" revela que hay 935 casos de partos espontáneos y 497 casos por cesáreas. Se procede a crear el modelo de regresión logística

#convertir una variable categórica de dos valores en variable numérica 1 y 0

```
data$variable_bool <- as.logical(data$Tipo.de.Parto=="CESÁREA")
```

```
data$tipo_parto <- as.numeric(data$variable_bool)
```

#De esta forma CESÁREA=1 y ESPONTÁNEO=0

```
table(data$tipo_parto)
```

```
##
```

```
## 0 1
```

```
## 935 497
```

#Se crea el modelo

```
modelo3 <- glm(data$tipo_parto ~ data$Tiempo.de.Gestación, data = data, family = "binomial")
```

#Imprimimos el modelo

```
summary(modelo3)
```

```
##
```

```
## Call:
```

```
## glm(formula = data$tipo_parto ~ data$Tiempo.de.Gestación, family = "binomial",
```

```
##   data = data)
```

```
##
```

```
## Coefficients:
```

```
##           Estimate Std. Error z value Pr(>|z|)
```

```
## (Intercept)      7.94099   1.26990   6.253 4.02e-10 ***
```

```
## data$Tiempo.de.Gestación -0.22391 0.03313 -6.758 1.40e-11 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 1849.0 on 1431 degrees of freedom
## Residual deviance: 1796.1 on 1430 degrees of freedom
## AIC: 1800.1
##
## Number of Fisher Scoring iterations: 4
```

Análisis del Modelo 3

La variable categórica “Tipo de Parto” se convierte en numérica (1 para cesárea y 0 para espontáneo). Se crea un modelo de regresión logística con “Tipo de Parto” como variable dependiente y “Tiempo de Gestación” como independiente. El resumen del modelo muestra que, a mayor tiempo de gestación, disminuye la probabilidad de cesárea (coeficiente: -0.22391). El p-valor confirma la significancia. La ecuación resultante del modelo de regresión logística es:

$$P(\text{Cesárea}) = \frac{1}{1 + e^{-(7.94099 - 0.22391 \times \text{Tiempo_de_Gestación})}}$$

Conclusiones

Con los resultados obtenidos de los modelos de regresión desarrollados en R, se puede concluir que se alcanzó con éxito el objetivo planteado de realizar un análisis estadístico predictivo de nacimientos vivos en Guadalajara de Buga, Colombia. El enfoque metodológico, que abarcó desde una regresión lineal simple hasta modelos más complejos, proporcionó una comprensión profunda de las relaciones entre las variables clave: la longitud del feto, el tiempo de gestación, el peso al nacer y el tipo de parto. En el Modelo 1, la regresión lineal simple reveló una correlación positiva fuerte entre la talla del feto y su peso al nacer. La ecuación resultante permitió prever con precisión el aumento promedio de peso por cada centímetro adicional en la talla, estableciendo un modelo viable y práctico para la predicción de pesos neonatales.

El Modelo 2, una regresión multilínea, llevó la investigación un paso más allá al incorporar el tiempo de gestación como predictor adicional. La visualización tridimensional proporcionó una representación clara de cómo la combinación de la talla y el tiempo de gestación influye en el peso de los recién nacidos.

Finalmente, el Modelo 3, basado en una regresión logística, abordó la predicción del tipo de parto según las semanas de gestación. La significancia de la variable de tiempo

de gestación en la probabilidad de cesárea destacó la utilidad clínica de este enfoque predictivo.

La investigación permitió demostrar relaciones importantes entre variables en nacimientos vivos, proporcionando herramientas predictivas valiosas a través de modelos de regresión y logísticos. Estos hallazgos pueden tener un impacto positivo significativo en la toma de decisiones clínicas y en la planificación obstétrica, señalando resultados que tienen el potencial de mejorar sustancialmente la atención materno-neonatal en Guadalajara de

Referencias

- [1] R. Sesmero, *Principios de medicina: materno fetal*. España, Editorial de la Universidad de Cantabria, 2018
- [2] B. Oñate, E. Morales, C. Jiménez, y P. Méndez, "Una alternativa a Stata: usando R para estimación de Modelos de Regresión", *ACI Avances En Ciencias e Ingenierías*, Vol. 11, no 2, 2019
- [3] F. Nedel, B. Morina, y M. Utzet, *Introducción a la estadística para ciencias de la salud con R-Commander*. Universitat Autònoma de Barcelona, 2016
- [4] A. García, *Estadística aplicada avanzada con R*. UNED - Universidad Nacional de Educación a Distancia, 2022
- [5] W. López, *Estadística práctica para ciencias de la salud y enfermería*. San Juan, Puerto Rico: Universidad de Puerto Rico, 2021
- [6] J. Pineda, "Modelos predictivos en salud basados en aprendizaje de maquina (machine learning)", *Revista Médica Clínica Las Condes*, Vol. 33, no 6, pp. 583–590, 2022
- [7] J. Tapia, "Factores determinantes en el peso del recién nacido de madres adolescentes en lima a través de la regresión logística multinomial", 2022
- [8] H. Llinás, *Estadística Inferencial*. Barranquilla, Colombia: Universidad del Norte, 2018
- [9] M. Hernández, M. Tapia, y S. Hernández, *Estadística inferencial 2: aplicaciones para ingeniería*. Ciudad de México, México: Grupo Editorial Patria, 2019
- [10] C. Viedma, *Estadística descriptiva e inferencial*. Madrid, España: Ediciones IDT, 2018
- [11] M. Díaz, *Estadística inferencial aplicada*. Universidad del Norte, 2019
- [12] W. Proaño, *Estadística descriptiva e inferencial*. Universidad del Azuay, 2020

- [13] L. Díaz, M. Morales, y L. León, *Análisis estadístico de datos categóricos*. Bogotá, Colombia: Editorial Universidad Nacional de Colombia, 2018
- [14] J. H. Ávila-Toscano, L. J. Vargas-Delgado, C. D. Jiménez-Yejas y D. J. Ortiz-Mejía, "Statistical accessibility in articles published in scientific education journals (Europe-Latin America) included in Scimago Journal Ranking", *Eco Matemático*, vol. 13, no. 1, pp. 81-92, 2022
- [15] K. Trejo, *Fundamentos de metodología para la realización de trabajos de investigación*. Editorial Parmenia, Universidad La Salle México, 2021
- [16] M. Pujol, y D. Liviano, *Análisis cuantitativo con R: matemáticas, estadística y econometría*. Editorial UOC, 2017
- [17] A. López, M. Rojas, y A. Tunja, *Herramientas para el análisis estadístico de datos biológicos en R*. 1. Editorial UPTC, 2022
- [18] J. Gil, *Aplicaciones de R en estadística básica y textual*. UNED - Universidad Nacional de Educación a Distancia, 2020
- [19] J. Freijo, *El paquete estadístico R (2a. ed.)*. CIS - Centro de Investigaciones Sociológicas, 2013
- [20] C. Tallerez, *Modelos estadísticos lineales: con aplicaciones en R*. Ediciones de la U, 2016
- [21] M. Contenido, *Estadística con aplicaciones en R*. Editorial Utadeo, 2019