

EVIDENCIAS EN PEDIATRÍA

Toma de decisiones clínicas basadas en las mejores pruebas científicas
www.evidenciasenpediatria.es

Fundamentos de medicina basada en la evidencia

Regresión logística binaria simple

Ortega Páez E¹, Ochoa Sangrador C², Molina Arias M³

¹UGC de Maracena. Distrito Granada-Metropolitano. Granada. España.

²Servicio de Pediatría. Hospital Virgen de la Concha. Zamora. España.

³Servicio de Gastroenterología. Hospital Universitario La Paz. Madrid. España.

Correspondencia: Eduardo Ortega Páez, edortegap@gmail.com

Palabras clave en español: estadística; inferencia estadística; regresión logística.

Palabras clave en inglés: statistics; statistical inference; logistic regression.

Fecha de recepción: 14 de febrero de 2022 • **Fecha de aceptación:** 28 de febrero de 2022

Fecha de publicación del artículo: 16 de marzo de 2022

Evid Pediatr. 2022;18:11.

CÓMO CITAR ESTE ARTÍCULO

Ortega Páez E, Ochoa Sangrador C, Molina Arias M. Regresión logística binaria simple. Evid Pediatr. 2022;18:11.

Para recibir Evidencias en Pediatría en su correo electrónico debe darse de alta en nuestro boletín de novedades en
<http://www.evidenciasenpediatria.es>

Este artículo está disponible en: <http://www.evidenciasenpediatria.es/EnlaceArticulo?ref=2022;18:11>.

©2005-22 • ISSN: 1885-7388

Regresión logística binaria simple

Ortega Páez E¹, Ochoa Sangrador C², Molina Arias M³

¹UGC de Maracena. Distrito Granada-Metropolitano. Granada. España.

²Servicio de Pediatría. Hospital Virgen de la Concha. Zamora. España.

³Servicio de Gastroenterología. Hospital Universitario La Paz. Madrid. España.

Correspondencia: Eduardo Ortega Páez, edortegap@gmail.com

La regresión logística (RL) es una técnica estadística que permite relacionar una variable dependiente cualitativa (Y) con una o mas variables independientes cuantitativas y/o cualitativas (X). Dependiendo del tipo de variable dependiente, tendremos la RL binaria (cualitativa dicotómica), la RL multinomial (cualitativa categórica, con más de dos categorías) y la RL ordinal (cualitativa ordinal). En este trabajo nos centraremos en el modelo más sencillo, la Regresión logística simple binaria.

FUNCIÓN LOGÍSTICA

La RL actualmente, como todas las técnicas de regresión, está englobada dentro del modelo lineal generalizado (MLG) que responde a la fórmula $Función(y) = a + bx_1 + cx_2 + \dots + nx_n + e$.

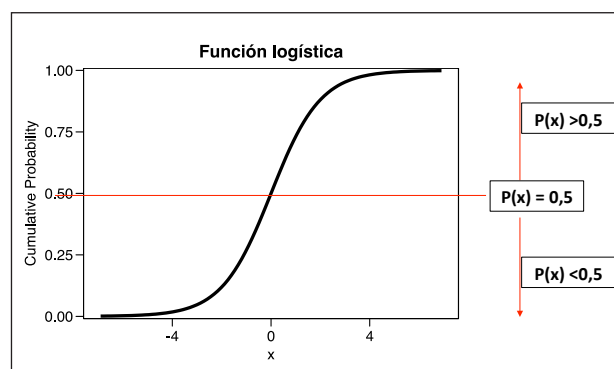
Si la variable dependiente es categórica, Y = 0 sería ausencia del carácter e Y = 1 presencia del mismo. A diferencia de la regresión lineal que predice los valores medios que toma la variable dependiente, la RL predice probabilidades (p) de ocurrencia del suceso en función del valor que toman las variables independientes, siendo sus valores desde 0 a 1. Matemáticamente es posible aplicar el modelo de regresión lineal de mínimos cuadrados, pero sus valores no serían creíbles, puesto que, para valores extremos del predictor, se obtienen valores de Y menores que 0 o mayores que 1. Esto contradice los valores que debe tomar una probabilidad (0,1). Para resolver este problema debemos buscar una función que transforme el valor devuelto por la regresión lineal en un resultado comprendido entre 0 y 1. Aunque hay varias funciones, la función logística o sigmoide (figura 1), que cumple estos requisitos, responde a la siguiente ecuación:

$$p(x) = \frac{1}{1 + e^{-z}}$$

- Si z tiene un valor muy grande positivo e^{-z} es aproximadamente 0, $p(x) = 1$.
- Si $z = 0$, $e^{-z} = 1$, $p(x) = 0,5$.
- Si z tiene un valor muy grande negativo, $e^{-z} = \infty$, $p(x) = 0$.

Como se puede ver en la figura 1, la curva logística se caracteriza porque valores de probabilidad de Y >0,5 se clasifican como 1 o presencia del evento y probabilidades Y <0,5 se

Figura 1. Función logística



clasifican como 0 o no presencia del mismo, siendo el $P(Y) = 0,5$ el punto de corte.

Despejando z en la ecuación logística tenemos:

$$p(x) + p(x)e^z = 1 \rightarrow p(x) + p(x)e^z = 1 \rightarrow p(x)e^z = 1 - p(x)$$

$$e^z = \frac{1-p(x)}{p(x)} \rightarrow e^z = \frac{p(x)}{1-p(x)} = Odds.$$

Si tomamos logaritmos neperianos (ln) a ambos de la igualdad tenemos:

$$\ln e^z = Z = \ln \frac{p(x)}{1-p(x)} = \beta_0 + \beta_1 x \rightarrow \ln(odds) = \beta_0 + \beta_1 x$$

Al final hemos conseguido expresar en ln de odds el modelo lineal, mucho más sencillo de manejar ya que sigue la distribución normal. A esta transformación se le llama logit.

MODELO DE REGRESIÓN LOGÍSTICA

Una vez vista la función logística, siendo la variable Y binaria (1/0), la media de los valores de Y para cada posible valor de X sigue una función logística, expresada en la siguiente ecuación

$$P(Y / X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X)}} \quad \epsilon = \frac{\exp(\beta_0 + \beta_1 X)}{1 + \exp(\beta_0 + \beta_1 X)}$$

β_1 y β_2 son los coeficientes del modelo de regresión y el término ϵ representa el error residual.

REQUISITOS

La regresión logística es una técnica robusta y necesita pocos requisitos, básicamente son:

1. Relación lineal entre el ln de las *odds ratios* (OR) y las variables independientes cuantitativas.
2. Independencia de las observaciones entre sí.
3. Ausencia de colinealidad entre variables independientes.
4. La variable dependiente debe ser cualitativa.

ESTIMACIÓN DE LOS PARÁMETROS

La estimación de los parámetros se realiza por el método de máxima verosimilitud, que consiste en calcular los valores de los coeficientes de regresión que dan a la muestra observada la máxima probabilidad posible. Este proceso es aproximado y se hace por varias iteraciones. Obviamos la explicación del proceso de estimación por exceder el objetivo de este documento. El lector interesado puede informarse en la bibliografía referenciada.

BONDAD DE AJUSTE DEL MODELO

Una vez estimados los coeficientes del modelo, debemos comprobar hasta qué punto explica la variación de la variable dependiente. Existen varios procedimientos.

Pruebas de bondad de ajuste

Se trata de conocer como reproduce el modelo la probabilidad pronosticada. Se realiza por dos tipos de contrastes, el contraste que analiza la bondad de ajuste global del modelo y el contraste de bondad de ajuste cuando se introduce la variable independiente.

La verosimilitud del modelo representada por l (*likelihood*), nos dice la probabilidad de reproducir los datos de la muestra a partir del modelo. Como este número es muy pequeño (entre 0 y 1), es más ventajoso trabajar con el logaritmo neperiano (LL), que como es un número negativo se suele expresar en $-2LL$, que es positivo. Desviación o lejanía (*deviance*) es el grado de ajuste del modelo a los datos y viene dada por el del estadístico $-2LL$ y cuanto más pequeño sea el valor mayor ajuste existirá.

Contrastes de bondad de ajuste global

Se realiza a partir la verosimilitud del modelo. También llamada prueba de Omnibus ($-2LR$). Se compara la verosimilitud del modelo nulo o *null* que sólo tiene la constante ($-2LL0$) con

el modelo que contiene la variable predictora ($-2LLM$) que sigue una distribución χ^2 con el número de grados de libertad equivalente al número de variables (p). Es similar a la prueba F global de la ANOVA utilizada en regresión lineal.

$$-2LR = -2LL0 - (-2LLM) \rightarrow \chi^2_{gl=p}$$

La hipótesis nula (H_0) asume que todos los coeficientes del modelo sin la constante son igual a cero. Si $p < 0,05$, rechazamos la hipótesis nula de igualdad de coeficientes y supondremos que el modelo estimado es significativo en cuanto a la relación entre ambas variables y presenta una buena bondad de ajuste.

Contraste de bondad de ajuste al introducir las variables en el modelo

Prueba de bondad de ajuste de Hosmer-Lemeshow. Valora la concordancia que existe entre las probabilidades observadas en la muestra y las predichas por el modelo, sigue una distribución χ^2 . Bajo la hipótesis nula de que no existen diferencias entre ambas probabilidades, una $p < 0,05$ concluye que rechazamos que exista una bondad de ajuste del modelo. Tiene el inconveniente, que cuando algún valor esperado de muy pequeño (< 5), no se computa el estadístico y toma el valor 0, pudiéndose concluir erróneamente que existe bondad de ajuste.

La alternativa a este estadístico es la prueba de verosimilitud vista en el epígrafe anterior.

Otra forma de evaluar la bondad del ajuste a partir de la idea de que un valor alto de la probabilidad predicha se asociará con un resultado de 1 en la variable dependiente y un valor bajo cercano al cero con el 0. Podríamos utilizar el modelo como clasificación de una prueba diagnóstica, calculando su sensibilidad, especificidad y exactitud. Un valor mayor del 75% nos diría que el modelo presenta un buen ajuste.

Coefficientes de determinación. PseudoR²

Para medir el grado de varianza explicada por el modelo ajustado, se han desarrollado métodos que por analogía con la R^2 de la regresión lineal se denominan genéricamente PseudoR². Toman valores entre 0 y 1, cuanto mayor es el valor, mayor es la varianza explicada. No se trata de una prueba para medir la bondad de ajuste propiamente dicha.

- PseudoR² de Mc McFadden. Compara al logaritmo de verosimilitud del modelo nulo con la constante sola ($L0$) con la verosimilitud del modelo con la variable predictora (LM). Es poco utilizado.

$$R^2_{MF} = 1 - \frac{l(0)}{l(M)}$$

- PseudoR² de Cox-Snell. Compara la verosimilitud del modelo nulo (L0) con la verosimilitud del modelo con la variable predictora (LM).

$$R^2_c = 1 - \left(\frac{l(0)}{l(M)} \right)^{\frac{2}{n}}; n = \text{tamaño muestral.}$$

Nos indica la proporción de varianza explicada por la variable predictora (independiente). Tiene el inconveniente que nunca toma el valor 1, aunque el modelo explique el 100% de la variabilidad de la variable de respuesta.

- PseudoR² de Nagelkerke. Corrige el caso anterior para que tome el valor 1 en el caso de que el modelo explique el 100% de la varianza. Es el más utilizado

$$R^2_N = 1 - \frac{R^2}{R^2_{\text{máx}}}; R^2_{\text{máx}} = 1 - l(0)^{\frac{2}{n}}.$$

EJEMPLO PRÁCTICO DE REGRESIÓN LOGÍSTICA BINARIA SIMPLE

Veamos un ejemplo utilizando un programa de acceso libre, el software estadístico R (<https://www.r-project.org/>) con el plugin RCommander (<https://estadistica-dma.ulpgc.es/cursoR4ULPGC/12-Rcommander.html>) y la base de datos GEA_ped.RData, disponible en la web de *Evidencias en Pediatría*. Si necesita saber cómo instalar RCommander, puede consultar el siguiente tutorial en línea (http://sct.uab.cat/estadistica/sites/sct.uab.cat.estadistica/files/instalacion_r_commander_0.pdf).

El ejercicio se ha realizado con la versión de RStudio Versión 1.1.463 para Mac y RCommander v 2.7.1 en inglés.

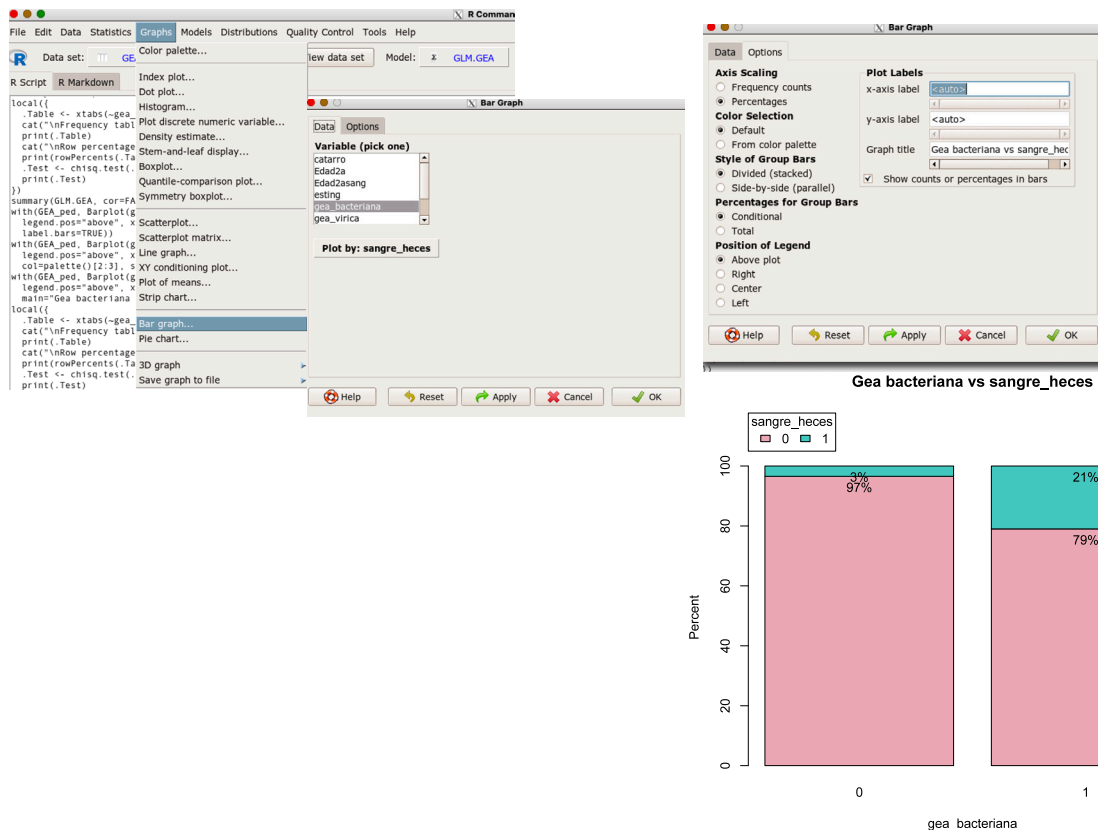
En la base de datos se recoge un estudio transversal de 256 niños sobre algunos factores de riesgo asociados con diarrea aguda (bacteriana o no bacteriana). Nos centraremos en estudiar si la variable sangre en heces (sangre_heces) codificada como dicotómica (0 = no, 1 = sí) está asociada a gastroenteritis bacteriana (gea_bacteriana) codificada como dicotómica (0 = no, 1 = sí).

En primer lugar, veremos la distribución de frecuencias de sangre en heces en las gastroenteritis agudas (GEA) bacterianas y no bacterianas. Para ello seleccionamos la opción Graph (gráficos) → bar graph (gráfico de barras). En la ventana emergente seleccionamos en la pestaña de data, gea_bacteriana y plot by: sangre_heces (estratificado por sangre en heces), en la pestaña options (opciones) percentages en la primera casilla y el resto por defecto. Obtenemos que el 21% de las GEA bacterianas presentan sangre en heces frente a no bacterianas (figura 2). Podemos deducir que puede existir una relación entre sangre en heces y GEA bacteriana.

Construcción del modelo de regresión logística

En primer lugar, tenemos que activar la base de datos (GEA_ped) una vez cargada en la pestaña “No active data set” (base de datos no activada) y seleccionarla en la ventana emergente

Figura 2. Gráfico de frecuencias entre GEA bacteriana y sangre en heces



inmediatamente nos aparecerá la base activa. Seguidamente en la pestaña statistics (estadísticos) → fit models (ajuste de modelos) → Generalized linear model (modelo lineal generalizado) y aceptamos (figura 3). Como ya se dijo, R Commander incluye dentro de los modelos lineales generalizados todo modelo de regresión. Seguidamente nos aparece un cuadro de diálogo común para cualquier modelo de regresión, en la que primeramente tenemos las variables a seleccionar, la fórmula de regresión y por último la función del modelo. En nuestro caso, seleccionamos `gea_bacteriana` como variable dependiente en la primera casilla de la fórmula, `sangre_heces` en la casilla de la derecha, en family → binomial (la distribución de probabilidad de la variable dependiente es binomial) y en link function → logit, le damos un nombre al modelo (GLM.GEA) para poder después identificarlo, aceptamos y nos aparece el modelo activo en la pestaña model (figura 4).

En la ventana de salida de datos (figura 5) R nos muestra en primer lugar la fórmula utilizada para hacer la regresión, función `glm`, la variable dependiente primero y la independiente después, la distribución (binomial), la función (logit) y la base de datos. Posteriormente los residuos de las desviaciones (*deviance residuals*) por cuantiles, interpretados como la contribución de cada dato a la *deviance* total. En segundo lugar, los coeficientes de regresión (*coeficcients*), el termino constante (*Intercept*), y la variable independiente (`sangre_heces`, con la categoría analizada presencia de sangre en heces -T.1) que es la única que interpretamos, ya que es la que contribuye a la explicación del modelo. De izquierda a derecha tenemos el valor puntual estimado (2.01) que al ser positivo nos dice que existe una relación

positiva, el error estándar (0,49), la prueba de significación de Wald (z: 4,049), que en es este caso se ha realizado por aproximación a la normal, al ser un tamaño muestral suficientemente grande, y, por último, el nivel de significación ($p = 0,00005$); el programa nos señala con tres asteriscos que la $p < 0,0001$.

Posteriormente R nos muestra las desviaciones que ahora veremos, el valor del criterio de información de Akaike (AIC), que valora el mejor modelo ajustado cuando existen varios modelos a valorar (que será el de menor AIC), y el valor puntual de la OR de sangre en heces = 7,48. Para conocer los intervalos de confianza del 95% (IC 95) de la OR debemos ir a la pestaña models (modelos) → confidence intervals (intervalos de confianza) y se nos despliega una ventana en la que podemos elegir el método para construir el IC 95; recomendándose el de Wald para muestras grandes y el *likelihood ratio* para muestras pequeñas. Elegimos el de Wald, el más utilizado, ya que ambos darán resultados muy similares. Obtenemos un IC 95 de la OR de 2,9 a 21,5 (figura 6).

La interpretación sería, que existe relación con un nivel muy alto de significación entre la presencia de sangre en heces y la GEA bacteriana, esto supone que por término medio es 7 veces más probable presentar GEA bacteriana si existe sangre en heces que si no existe (OR = 7,48, IC 95: 2,9 a 21,5)

R no nos ofrece la sensibilidad, especificidad y exactitud el modelo, para conocerla tenemos que recurrir a comandos de sintaxis. Conociendo que el punto de corte es 0,5 ($cp = 0,5$), clasificamos (clasificación) todos los valores mayores de 0,5

Figura 3. Construcción de un modelo de regresión logística (I)

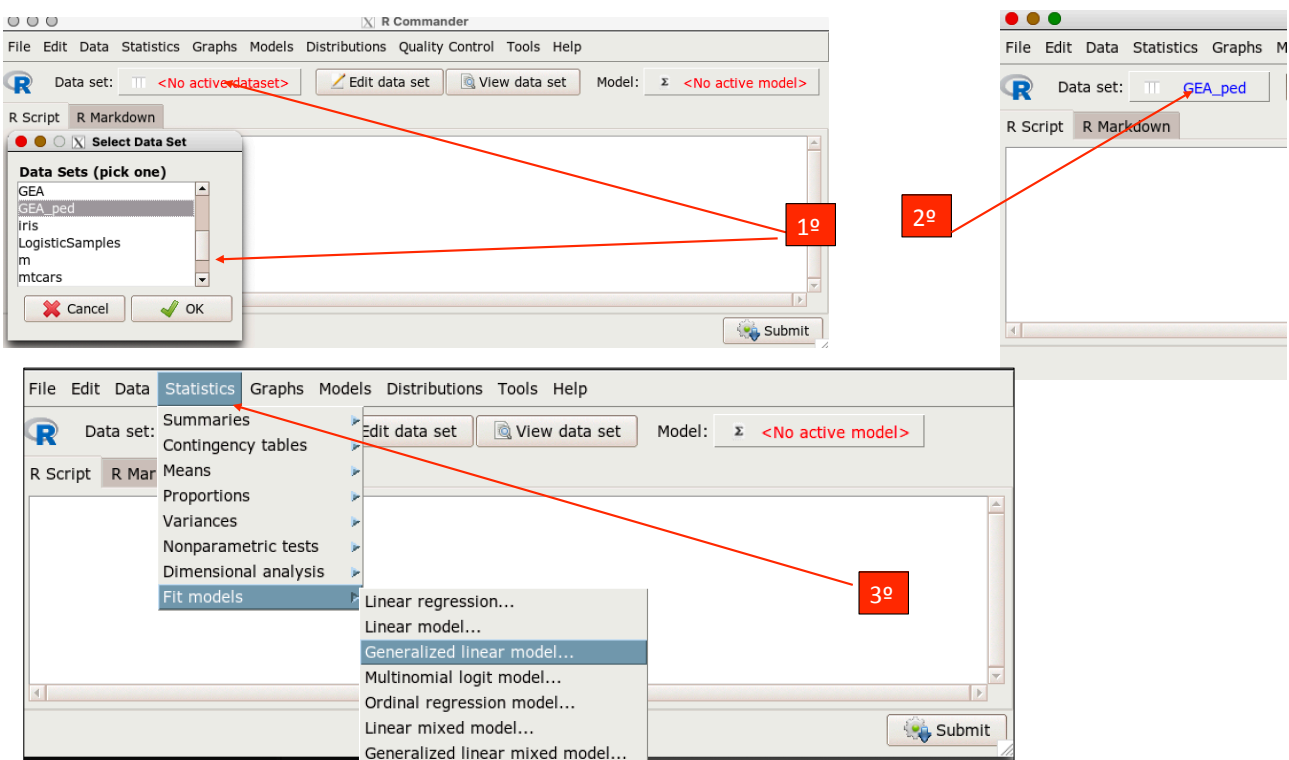
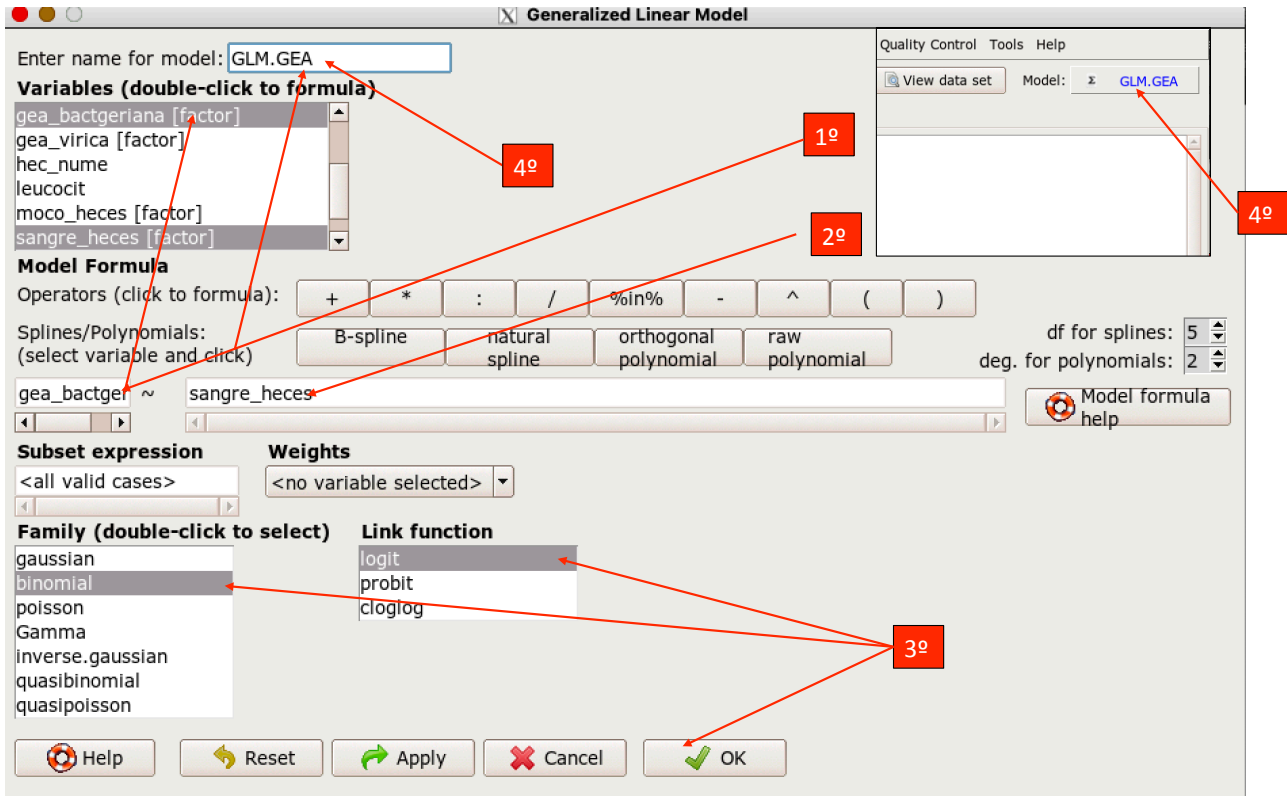


Figura 4. Construcción de un modelo de regresión logística (II)



como 1 ($cp > 0,1$). Colocamos en la variable clasificación y en la variable `gea_bacteriana` los 1 para que estén en la columna 1 y los 0 en la columna 2. Aplicamos la función “`confusionMatrix`” de la librería `caret` (figura 7). Como podemos ver, el modelo clasifica correctamente al 72,6% de los valores (*accuracy*: 0,726; IC 95: 0,66 a 0,78) con una sensibilidad 20% y una especificidad del 96%. El modelo nos dice que la presencia de

`sangre` en heces predice escasamente la diarrea bacteriana, su ausencia prácticamente la descarta.

Bondad de ajuste del modelo

Vimos anteriormente que en la figura 5 se ofrecen los resultados de la *deviance* del modelo solo con la constante, LLO (*null deviance*)

Figura 5. Resultados de un modelo de regresión logística

```
Rcmdr> GLM.GEA <- glm(gea_bacteriana ~ sangre_heces, family=binomial(logit),
Rcmdr+ data=GEA_ped)

Rcmdr> summary(GLM.GEA)

Call:
glm(formula = gea_bacteriana ~ sangre_heces, family = binomial(logit),
data = GEA_ped)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.6394  -0.8014  -0.8014   0.9850   1.6076

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -0.9710    0.1468  -6.616 3.70e-11 ***
sangre_heces[T.1]  2.0125    0.4970   4.049 5.14e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 319.55 on 255 degrees of freedom
Residual deviance: 300.34 on 254 degrees of freedom
AIC: 304.34

Number of Fisher Scoring iterations: 4

Rcmdr> exp(coef(GLM.GEA)) # Exponentiated coefficients ("odds ratios")
(Intercept) sangre_heces[T.1]
  0.3786982      7.4817708
```

AIC, criterio de información Akaike proporciona una medida del ajuste de un modelo que penaliza al modelo que contiene más variables predictoras, pudiendo así comparar dos modelos.

-2LL(0)

-2LL(B)

Figura 6. Intervalos de confianza para la OR

Intervalos de confianza del 95% para la OR

```

Rcmdr> Confint(GLM.GEA, level=0.95, type="LR")
              Estimate      2.5 %      97.5 %
(Intercept) -0.9710156 -1.265452 -0.6890522
sangre_heces[T.Si] 2.0124695 1.086967 3.0685655

Rcmdr> Confint(GLM.GEA, level=0.95, type="LR", exponentiate=TRUE)
Exponentiated Coefficients and Confidence Bounds
              Estimate      2.5 %      97.5 %
(Intercept)  0.3786982 0.2821117 0.5020517
sangre_heces[T.Si] 7.4817708 2.9652653 21.5110237
    
```

319,55 y la deviance del modelo con la variable independiente LLB (residual deviance) 300,24. Ambos valores son mediocres, ya que están alejados de 0 que sería el ajuste perfecto, pero existe una disminución del modelo LLB lo que indica un mejor ajuste al introducir la variable sangre_heces en el modelo.

Para conocer si la bondad de ajuste es significativa realizamos la prueba de bondad de ajuste global u ómnibus del modelo. R ofrece esta prueba bajo el nombre ANOVA table, por simi-

litud con la F de la ANOVA de regresión lineal pero una prueba χ^2 . Una vez seleccionado el modelo (GLM.GEA), en la pestaña models (modelos) tenemos una serie de opciones para estudiar el ajuste del modelo. Seleccionamos Hypothesis test (test de hipótesis) → ANOVA table (tabla ANOVA) y en la ventana emergente seleccionamos la segunda opción y aceptamos. R nos ofrece en la ventana de resultados el valor de la prueba ($\chi^2 = 19,12$; $gl = 1$) con una significación muy alta ($p = 0,0000117$) (figura 8).

Figura 7. Sensibilidad, especificidad y exactitud de un modelo logístico

Sintaxis

```

>cp = 0.5
>clasificacion = ifelse(GLM.GEA$fitted.values > cp, 1, 0)
>install.packages("caret")
>library(caret)
>clasificacion_f = factor(clasificacion, levels = 1:0)
>gea_bacteriana_f = factor(gea_bacteriana, levels = 1:0)
(>confMat = confusionMatrix(clasificacion_f, gea_bacteriana_f))
    
```

Ventana sintaxis

Ventana resultados

Confusion Matrix and Statistics

		Reference	
Prediction		1	0
1		17	6
0		64	169

Accuracy : 0.7266
95% CI : (0.6676, 0.7802)

No Information Rate : 0.6836
P-Value [Acc > NIR] : 0.07779

Kappa : 0.2174

Mcnemar's Test P-Value : 9.572e-12

Sensitivity : 0.20988
Specificity : 0.96571

Pos Pred Value : 0.73913
Neg Pred Value : 0.72532

Prevalence : 0.31641
Detection Rate : 0.06641

Detection Prevalence : 0.08984
Balanced Accuracy : 0.58780

'Positive' Class : 1

Si queremos conocer la prueba de Bondad de ajuste global de Hosmer-Lemeshow, debemos hacerla mediante sintaxis, ya que R Commander no la tiene implementada, usando la función `blr_test_hosmer_lemeshow` de la librería `blorr`, como se detalla a continuación (figura 9). Nos ofrece resultados parecidos a la prueba de ómnibus, $p < 0,00001$.

Para obtener las pseudo R^2 , igualmente hay que recurrir a los comandos de sintaxis con la librería `DescTools` y la función `PseudoR2` (figura 10).

Podemos observar que el modelo GLM.GEA explica el 10% de la varianza de la GEA bacteriana (R^2 Nagelkerke).

Figura 8. Prueba de ajuste global u Omnibús

The figure shows the R Commander interface with the 'ANOVA Table' dialog box open. The dialog box has 'Partial obeying marginality ("Type II")' selected. Below the dialog box, the R console shows the following output:

```
Rcmdr> Anova(GLM.GEA, type="II", test="LR")
Analysis of Deviance Table (Type II tests)

Response: gea_bacteriana
          LR Chisq Df Pr(>Chisq)
sangre_heces 19.212  1 0.0000117 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figura 9. Prueba de bondad de ajuste de Hosmer-Lemeshow

Sintaxis

```
> install.packages("blorr")
> library("blorr")
> blr_test_hosmer_lemeshow(GLM.GEA)
```

Goodness of Fit Test

Chi-Square	DF	Pr > ChiSq
0.0000	8	1.0000

Figura 10. Pseudo R^2

Sintaxis

```
> install.packages("DescTools")
> Library("DescTools")
> PseudoR2(GLM.GEA, c("McFadden", "Nagel", "CoxSnell"))
```

McFadden	Nagelkerke	CoxSnell
0.06011958	0.10140051	0.07229823

PREDICCIONES

Una de las utilidades del modelo logístico es realizar predicciones sobre la presencia o ausencia del suceso en la variable dependiente, siempre y cuando el diseño epidemiológico sea adecuado (longitudinal). En nuestro ejemplo, podemos calcular la probabilidad de presentar gastroenteritis bacteriana en presencia de sangre en heces.

$$P(\text{GEA bacteriana/sangre en heces}) = \frac{1}{1+e^{-(0,97+2,01)}} = \frac{1}{1+e^{-1,04}} = \frac{1}{1+0,35} = 0,74.$$

El modelo predice que en presencia de sangre en heces existe una probabilidad del 74% de presentar GEA bacteriana.

CONCLUSIONES

Como resumen, podemos distinguir varias etapas en la RL:

1. Recodificar las variables independientes categóricas u ordinales y la variable dependiente en 0 y 1.
2. Ejecutar el modelo de Regresión logística, seleccionando la variable dependiente y la independiente, modelo binomial y función logit.
3. Analizar la significación del estadístico de Wald para la variable independiente. Un valor $p < 0,05$ significa que la introducción de la variable es significativa.
4. Analizar la fuerza y sentido y significación de los coeficientes mediante la OR ($e^{\beta 1}$).
5. Analizar el ajuste general de modelo observando la *deviance null* y residual junto con la prueba Omnibus. Una $p < 0,05$ nos habla a favor de un ajuste adecuado.
6. Analizar el poder discriminativo del modelo mediante la sensibilidad, especificidad y exactitud del modelo.
7. Analizar la capacidad predictiva del modelo si el diseño lo hace posible.

BIBLIOGRAFÍA

- Amat Rodrigo J. Correlación lineal y regresión lineal simple. En: Estadística con R [en línea] [consultado el 28/02/2022]. Disponible en https://github.com/JoaquinAmatRodrigo/Estadistica-con-R/blob/master/PDF_format/24_Correlaci%C3%B3n_y_Regresi%C3%B3n_lineal.pdf
- Botella-Rocamora P, Martínez-Beneito MA, Alacr eu-García M. Regresión logística en R. Coomander [en línea] [consultado el 28/02/2022]. Disponible en <https://www.pinae.es/wp-content/uploads/2012/09/introduccionalaregresionlogisticaconr-commander.pdf>
- Dom enech Massons JM. An alisis Multivariante: Modelos de Regresión Unidad did actica 8: Modelos de regresión logística. Editor: Domenech JM. Barcelona. Signo. 2000.
- Logistic regression in R using blorr package [en línea] [consultado el 28/02/2022]. Disponible en <https://www.r-bloggers.com/2019/02/logistic-regression-in-r-using-blorr-package/>
- Molina Arias M, Ochoa Sangrador C, Ortega P aez E. Correlaci on. Modelos de regresión. Evid Pediatr. 2021;17:25.
- Package Descstool [en línea] [consultado el 28/02/2022]. Disponible en <https://cran.r-project.org/web/packages/DescTools/DescTools.pdf>
- S anchez-Villegas A, Mart ın-Calvo N, Mart ınez-Gonz alez MA. Correlaci on y regresión lineal simple. En: Mart ınez MA, S anchez-Villegas A, Toledo EA, Faulin A (eds.). Bioestadística amigable. 3.ª ed. Barcelona: Elsevier; 2014. p. 269-326.