



Pontificia Universidad  
**JAVERIANA**  
Bogotá

MAESTRÍA EN   
**EPIDEMIOLOGÍA**  
CLÍNICA

**BIOESTADÍSTICA AVANZADA**

## **MÓDULO II**

**Semana 9**  
Regresión Logística

## Material elaborado por:

### **Nelcy Rodriguez Malagón**

Bioestadística, M.P.H

Profesora Titular

Departamento de Epidemiología Clínica y Bioestadística

Facultad de Medicina - Pontificia Universidad Javeriana

### **Kevin Maldonado-Cañón**

Médico Epidemiólogo, MD. MSc.

Estudiante - Doctorado en Epidemiología Clínica

Departamento de Epidemiología Clínica y Bioestadística

Facultad de Medicina - Pontificia Universidad Javeriana

# Modelo de Regresión Logística para la predicción de un evento dicotómico y evaluación de su desempeño: Sensibilidad, especificidad y curvas de características operativas (ROC)

## Introducción

Con alguna frecuencia, la pregunta de investigación se centra en seleccionar de un conjunto de factores o variables, aquellas que conjuntamente lleven a una predicción bastante precisa de un evento en salud. Esto significa la consideración de más de una variable que interactúan juntas en la predicción de ese evento. En este contexto, el modelo de regresión logística es una de las herramientas que puede ser usada para tal efecto. También dependiendo del contexto y la pregunta específica, se puede necesitar construir un modelo que responda a encontrar la probabilidad de ocurrencia de una enfermedad (modelo para diagnóstico, especialmente cuando el desenlace es dicotómico) o un modelo para pronóstico el cual busca determinar de la manera más precisa la probabilidad de condiciones futuras basadas en el perfil clínico del individuo y en un tiempo determinado. Por ejemplo, la ocurrencia de eventos específicos en un lapso específico de tiempo como muerte, complicaciones y cambios en la calidad de vida de vida. Los resultados de estos modelos resultan muy útiles como información de apoyo en la toma de decisiones sobre tratamiento de pacientes.

Cuando se habla de modelos de predicción, se tienen tres aspectos que son claves: El proceso de construcción del modelo hasta tener aquél que proporcione las estimaciones más precisas y que ojalá resulte ser el más parsimonioso y clínicamente el más significativo, la evaluación de su desempeño y su validación interna o en nuevos pacientes y el estudio de su impacto clínico. Al aplicar el modelo de regresión logística sea para diagnóstico o pronóstico, se tiene una serie de pasos que es importante considerar:

1. Requisitos
2. Evaluación descriptiva o exploratoria
3. Posible presencia de colinealidad
4. Construcción del modelo
5. Evaluación del Ajuste del modelo
6. Validación del modelo

En esta sesión haremos énfasis en el proceso de construcción de un modelo de predicción y la selección del mejor, así como de su evaluación de desempeño.

## Construcción del modelo de predicción

Cuando consideramos múltiples variables que actúan conjuntamente en la predicción de un evento y éste es de tipo dicotómico, el modelo de regresión logística es una opción a considerar como la estrategia de modelamiento matemático. El modelo de regresión logística considera la función

$$f(y) = \frac{1}{1 + e^{-y}}, \quad -\infty < y < +\infty$$

Donde  $y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k$

$$f(y) = Pr(D = 1 | x_1, x_2, \dots, x_k) = P(x)$$

$$\text{Entonces: } P(x) = \frac{1}{1 + e^{-(\alpha + \sum \beta_j x_j)}}$$

Corresponde a la probabilidad de desarrollar el evento dado, en un individuo que ha estado libre de él y que tiene  $k$  factores independientes. En este caso todos los factores o variables son considerados potenciales predictores y no se considera uno de ellos como un factor de exposición. La pregunta central es cuál o cuáles del conjunto de variable son los mejores predictores del desenlace.

El modelo logístico puede ser linealizado mediante la transformación logit:

$$\text{logit}[P(x)] = \alpha + \sum \beta_j x_j$$

Antes de construir un modelo de regresión logística para predicción se deben evaluar unos requisitos como los siguientes:

- **Los supuestos del modelo que han sido evaluados en las sesiones anteriores**
- **El número de factores a ser considerados o incluidos en el modelo.** Una regla general para determinar el número de predictores es:
  - $M/10$  donde  $M=3n_1n_2/n$ 
    - $n_1$ : Número de eventos o personas esperadas que desarrollarán el evento
    - $n_2$ : No eventos o personas esperadas que no desarrollarán el evento.

Así por ejemplo si en el problema para el cual se pretende desarrollar el modelo, se espera que de 250 personas en el estudio, 50 desarrollen el evento, al reemplazar en la fórmula planteada, se esperaría que máximo se incluyeran en el modelo 12 predictores. Además, se debe tener en cuenta que más que del número de variables o predictores, se debería pensar en el número de parámetros.

- **El tamaño de muestra necesario.** En este caso se deben tener en cuenta elementos como:
  - La proporción esperada del evento:  $p$
  - El número de predictores incluidos:  $k$
  - El número de eventos por variable esperados:  $EPV$

Con estas consideraciones, el tamaño de muestra en general se calcula como:

$$N = (EPV \times k)/p$$

Sin embargo, existen otras consideraciones y fórmulas que incluyen el nivel de sobreajuste aceptable y la precisión en la estimación del riesgo de tener el desenlace o el desempeño esperado del modelo.

### Referencias:

1. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. *Journal of Clinical Epidemiology*. 1996;49(12):1373-1379.
  2. Riley RD, Ensor J, Snell KIE, Harrell FE Jr, Martin GP, Reitsma JB, et al. Calculating the sample size required for developing a clinical prediction model. *BMJ*. 2020;368:m441.
  3. Riley RD, Snell KIE, Ensor J, Burke DL, Harrell FE Jr, Moons KGM, Collins GS. Minimum sample size for developing a multivariable prediction model: PART II – binary and time-to-event outcomes.
- **La forma de codificación de las variables y los niveles de medición empleados.** Esto es importante y se debe tener en cuenta según el programa estadístico o software que se utilice en el proceso de construcción del modelo.

Con relación a los pasos 2 y 3, en las sesiones pasadas se revisó el proceso de análisis exploratorio y evaluación de la colinealidad especialmente en el caso de variables independientes de tipo numérico sean continuas o discretas. Aquí vale la pena mencionar que para explorar la relación entre el desenlace y cada uno de los predictores, es útil correr regresiones simples en cada caso.

En el proceso de construcción del modelo se puede seguir la siguiente estrategia:

1. Iniciar con un modelo que incluya todas las variables seleccionadas en pasos previos: Las que tienen una importancia clínica suficiente que no se pueden excluir, aquellas que mostraron aporte al desenlace en las regresiones simples, y en caso necesario, aquellas interacciones que se consideren importantes.
2. Correr un modelo con todas esas variables incluidas y compararlo con un modelo en el cual se incluya solamente el intercepto. Comparar estos dos modelos usando una prueba de razón de verosimilitud. Si esta prueba no resulta significativa, significa que las variables en su conjunto no contribuirán a la predicción del desenlace. Si por el contrario la prueba resulta significativa, una o más de las variables incluidas serán predictoras del desenlace.

3. Si la prueba fue significativa, correr el modelo completo versus uno sin las interacciones y de nuevo evaluar el resultado de la prueba de razón de verosimilitud para eliminar o no, las interacciones del modelo.
4. Utilizar criterio backward para ir eliminando factores que no resulten significativos en el modelo, pero siempre teniendo en cuenta el concepto clínico. Este debe primar en la selección de las variables que deben permanecer en el modelo. Recuerde además que no necesariamente, se debe eliminar de a un factor. Usted puede considerar la eliminación de más de un factor a la vez según su análisis de los modelos que compara y teniendo en cuenta el criterio clínico y el contexto del problema. Adicionalmente, se puede definir una probabilidad de entrada de las variables al modelo y una probabilidad de retención. Cuál nivel de esa probabilidad se define, depende del criterio del investigador de acuerdo al contexto del problema, pero en general se definen: probabilidad de entrada ( $p_e=0.15$ ) y probabilidad de retención de las variables ( $p_r=0.2$ ).
5. Repita el proceso hasta que obtenga el modelo final que ojalá sea el más preciso y parsimonioso. Recuerde: aunque las pruebas estadísticas digan que un factor no es un predictor importante, si el criterio clínico y el contexto del problema, señalan que es importante, no debe ser excluido del modelo final.

Recuerde además que, si entre sus factores independientes algunos son categóricos con más de dos categorías, debe generar variables dummy o indicadoras que implican más parámetros en el modelo y que deben permanecer juntas en el modelo.

Para ejemplificar el proceso usaremos la base de datos `icu` y la pregunta de interés es saber si el estado vital puede ser predicho usando las variables `age`, `sys`, `hra`, `inf`, `gender` y `type`.

## Modelo completo

```
## Cargamos nuestra base de datos  
base_icu <- aplotre3::icu
```

Se ajusta inicialmente un modelo de regresión logística con todas las covariables:

```
install.packages("DescTools")
library(DescTools)

modelo_completo <- glm(sta ~ age + sys + hra + inf + ser + gender + type,
  data = base_icu,
  family = binomial)

summary(modelo_completo)
##
## Call:
## glm(formula = sta ~ age + sys + hra + inf + ser + gender + type,
##     family = binomial, data = base_icu)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.726370   1.553241  -1.755  0.07921 .
## age          0.033943   0.011530   2.944  0.00324 **
## sys         -0.012278   0.006197  -1.981  0.04758 *
## hra         -0.009929   0.008007  -1.240  0.21495
## infYes       0.400791   0.424768   0.944  0.34540
## serSurgical -0.328270   0.446696  -0.735  0.46241
## genderFemale -0.107684   0.403419  -0.267  0.78952
## typeEmergency 2.179135   0.817365   2.666  0.00767 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 200.16  on 199  degrees of freedom
## Residual deviance: 165.40  on 192  degrees of freedom
## AIC: 181.4
##
## Number of Fisher Scoring iterations: 6
DescTools::PseudoR2(modelo_completo, which = "McFadden")
##   McFadden
## 0.1736449
```

- **Interpretación general:**

- El modelo es globalmente significativo ( $p < 0.001$ ).
- Variables como `age`, `sys` y `type` muestran asociación estadísticamente significativa con el estado vital.
- El pseudo  $R^2$  de McFadden ( $\sim 0.17$ ) sugiere una capacidad explicativa moderada.

## Modelo nulo

Se ajusta un modelo sin variables independientes (solo intercepto):

```
modelo_nulo <- glm(sta ~ 1,
                  data = base_icu,
                  family = binomial)
```

Comparación mediante test de razón de verosimilitud:

```
anova(modelo_nulo, modelo_completo, test = "LRT")
## Analysis of Deviance Table
##
## Model 1: sta ~ 1
## Model 2: sta ~ age + sys + hra + inf + ser + gender + type
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1      199      200.16
## 2      192      165.40  7   34.757 1.242e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- **Interpretación general:**

- El test de razón de verosimilitud es significativo ( $p < 0.001$ ), lo que indica que al menos una de las variables incluidas es un predictor relevante del estado del paciente.

## Selección de variables (Stepwise)

Para determinar el modelo final, utilizaremos la estrategia *stepwise*:

```

modelo_step <- step(modelo_completo, direction = "both")
## Start:  AIC=181.4
## sta ~ age + sys + hra + inf + ser + gender + type
##
##           Df Deviance    AIC
## - gender  1   165.48 179.48
## - ser     1   165.95 179.95
## - inf    1   166.29 180.29
## - hra    1   167.00 181.00
## <none>           165.40 181.40
## - sys    1   169.59 183.59
## - type   1   175.31 189.31
## - age    1   175.50 189.50
##
## Step:  AIC=179.48
## sta ~ age + sys + hra + inf + ser + type
##
##           Df Deviance    AIC
## - ser     1   166.06 178.06
## - inf    1   166.38 178.38
## - hra    1   167.08 179.08
## <none>           165.48 179.48
## + gender  1   165.40 181.40
## - sys    1   169.81 181.81
## - type   1   175.32 187.32
## - age    1   175.51 187.51
##
## Step:  AIC=178.06
## sta ~ age + sys + hra + inf + type
##
##           Df Deviance    AIC
  
```

```
## - inf      1   167.05 177.05
## - hra      1   167.28 177.28
## <none>     166.06 178.06
## + ser      1   165.48 179.48
## + gender   1   165.95 179.95
## - sys      1   170.42 180.42
## - age      1   176.16 186.16
## - type     1   180.70 190.70
##
## Step:   AIC=177.05
## sta ~ age + sys + hra + type
##
##           Df Deviance   AIC
## - hra      1   167.82 175.82
## <none>     167.05 177.05
## + inf      1   166.06 178.06
## + ser      1   166.38 178.38
## + gender   1   166.91 178.91
## - sys      1   172.63 180.63
## - age      1   179.24 187.24
## - type     1   183.24 191.24
##
## Step:   AIC=175.82
## sta ~ age + sys + type
##
##           Df Deviance   AIC
## <none>     167.82 175.82
## + hra      1   167.05 177.05
## + inf      1   167.28 177.28
## + ser      1   167.54 177.54
## + gender   1   167.70 177.70
## - sys      1   173.08 179.08
## - age      1   179.37 185.37
## - type     1   183.25 189.25
summary(modelo_step)
```

```
##
## Call:
## glm(formula = sta ~ age + sys + type, family = binomial, data = base_icu)
##
## Coefficients:
##           Estimate Std. Error z value Pr(>|z|)
## (Intercept) -3.678554  1.306521  -2.816  0.00487 **
## age          0.033946  0.010865   3.124  0.00178 **
## sys         -0.013229  0.005985  -2.210  0.02709 *
## typeEmergency  2.287625  0.758168   3.017  0.00255 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 200.16  on 199  degrees of freedom
## Residual deviance: 167.82  on 196  degrees of freedom
## AIC: 175.82
##
## Number of Fisher Scoring iterations: 6
```

- **Interpretación general:**

- El modelo final incluye: age, sys, typ
- Variables eliminadas durante el proceso: sex, ser, inf, hra

Con este modelo, se puede determinar la probabilidad de morir en un paciente, reemplazando los valores correspondientes.

## Evaluación del modelo de predicción

Una vez se ha obtenido el modelo final, el siguiente paso corresponde a la evaluación del mismo. Esta evaluación considera dos aspectos: la calibración y la capacidad discriminativa.

### Evaluación de la calibración del modelo

La calibración de un modelo de predicción, se relaciona con la bondad de su ajuste. Existen algunas pruebas de bondad de ajuste para cuando el desenlace es dicotómico. Sin embargo, la prueba de bondad de ajuste de Hosmer-Lemeshow ha sido la más ampliamente usada. Esta prueba agrupa en deciles de riesgo, las probabilidades esperadas del evento obtenidas del modelo. La comparación del número de eventos esperados versus el número de eventos observados en los 10 grupos definidos, se hace mediante una prueba  $\chi^2$  con 8 grados de libertad (número de grupos - 2). Es una prueba que tiene una potencia baja si el tamaño de la muestra es pequeño. Como se mencionó antes, existen otras pruebas o medidas de calibración para desenlaces dicotómicos como por ejemplo la calibración de la pendiente, la estadística de E de Harrell o la prueba de calibración, pruebas que no se incluyen en este material (los interesados, lo pueden consultar en *Steyerberg EW. Clinical prediction models. 2nd ed. Cham: Springer; 2019. Capítulo 15.*).

La hipótesis nula que se evalúa en la prueba de Hosmer y Lemeshow es que el modelo tiene un buen ajuste a los datos versus la hipótesis alterna de que no se tiene este buen ajuste. Por tanto si se quiere un buen ajuste, es deseable que con los datos que se tienen, esta hipótesis no sea rechazada.

Usando la base de datos icu, ejemplificamos el uso de esta prueba.

Como vimos en la sección anterior el mejor modelo y más parsimonioso desarrollado incluyó las variables `age`, `sys` y `type` como predictores del estado vital (`sta`).

## Prueba de Hosmer Lemeshow

```
install.packages("ResourceSelection")  
library(ResourceSelection)  
  
ResourceSelection::hoslem.test(x = modelo_step$y, y = fitted(modelo_step), g = 10)  
  
##  
## Hosmer and Lemeshow goodness of fit (GOF) test  
##  
## data: modelo_step$y, fitted(modelo_step)  
## X-squared = 10.281, df = 8, p-value = 0.2459
```

Como se puede observar en los resultados anteriores, el modelo tiene un buen ajuste. Esto significa que su calibración es buena teniendo en cuenta esta prueba.

## Evaluación de la capacidad discriminativa del modelo

Un buen modelo de predicción debería poder discriminar entre quienes que tienen el evento de interés y quienes no lo tienen. Para evaluar esa discriminación de un modelo existen varias medidas estadísticas que dan cuenta de esa tarea.

La estadística *c* de concordancia es la medida más ampliamente usada para indicar la capacidad discriminativa en modelos lineales generalizados. Par el caso de modelos con evento binario, esta estadística es idéntica al área bajo la curva de características operativas (**ROC**).

La curva ROC es un gráfico que considera la **sensibilidad** (tasa de verdaderos positivos) en el eje de las ordenadas y **1-especificidad** (tasa de falsos positivos). Quiere esto decir que podemos entonces determinar tanto la sensibilidad como la especificidad del modelo.

En este caso, la **sensibilidad** del modelo está definida como la fracción de pacientes que este clasificó con el evento positivo entre todos aquellos que efectivamente tenían ese evento, y la **especificidad** es la fracción de pacientes que el modelo clasificó sin el evento entre aquellos que realmente no lo tenían. De esta forma, para evaluar el modelo en su capacidad discriminativa, usamos los elementos de una prueba diagnóstica.

En la curva ROC, para clasificar un paciente como positivo o negativo, se necesita utilizar un punto de corte de la probabilidad estimada. El punto de corte que se ha considerado y más usado es 0.5. En la gráfica corresponde a la línea que divide el área total en dos áreas de 45° cada una.

- Si la probabilidad estimada es igual a 0.5 o muy cercana a este valor, se considera una clasificación completamente aleatoria y entonces el modelo no clasifica muy bien a ese paciente.
- Si la probabilidad estimada es mayor que 0.5, el paciente es clasificado como positivo.
- Si la probabilidad estimada es menor que 0.5, el paciente es clasificado como negativo.

Ejemplificamos la obtención de estas estadísticas utilizando la base de datos icu.

```
install.packages("caret")
library(caret)
```

Con este código, separamos la variable que nos dice el estado real de cada paciente (*sta*) como factor y obtenemos las probabilidades predichas del modelo (`predict(..., type = "response")`). Luego, estas probabilidades se convierten en una clasificación binaria usando un punto de corte de 0.5 ( $\geq 0.5 = \text{"Died"}$ ,  $< 0.5 = \text{"Lived"}$ ). Con esto, se construye una matriz de confusión mediante `caret::confusionMatrix`, especificando que el evento de interés es "Died", lo que permite calcular métricas como sensibilidad, especificidad, valores predictivos y *accuracy*.

```
# Convertir a factor
real <- factor(base_icu$sta, levels = c("Lived", "Died"))

# Probabilidades predichas
prob_pred <- predict(modelo_step, type = "response")

# Clasificación (punto de corte 0.5)
pred_clase <- ifelse(prob_pred >= 0.5, 1, 0)
pred <- factor(pred_clase, levels = c(0,1), labels = c("Lived","Died"))

caret::confusionMatrix(pred, real, positive = "Died")
```

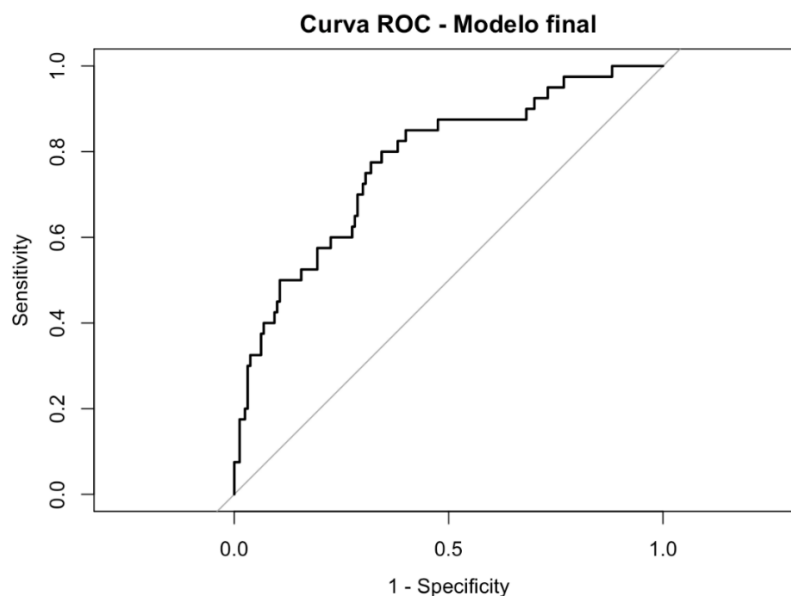
```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction Lived Died
##      Lived   158   33
##      Died     2    7
##
##           Accuracy : 0.825
##           95% CI : (0.7651, 0.875)
##      No Information Rate : 0.8
##      P-Value [Acc > NIR] : 0.2151
##
##           Kappa : 0.2291
##
## Mcnemar's Test P-Value : 3.959e-07
##
##           Sensitivity : 0.1750
##           Specificity : 0.9875
##           Pos Pred Value : 0.7778
##           Neg Pred Value : 0.8272
##           Prevalence : 0.2000
##           Detection Rate : 0.0350
##      Detection Prevalence : 0.0450
##           Balanced Accuracy : 0.5813
##
##           'Positive' Class : Died
##
install.packages("pROC")
library(pROC)
```

Adicionalmente, evaluaremos la capacidad discriminativa del modelo utilizando la curva ROC (`pROC::roc`), que, como se introdujo anteriormente, compara todas las posibles combinaciones de sensibilidad y especificidad para distintos puntos de corte. A partir de esta curva se calcula el área bajo la curva (AUC), que resume la capacidad del modelo para diferenciar entre pacientes que mueren y sobreviven, y finalmente se visualiza la curva ROC

```
# Curva ROC
roc_obj <- pROC::roc(base_icu$sta, prob_pred)

# Área bajo la curva (AUC)
pROC::auc(roc_obj)
## Area under the curve: 0.7744

# Visualización de la curva ROC
plot(roc_obj,
      main = "Curva ROC - Modelo final",
      legacy.axes = TRUE)
```



- **Interpretación general:**

En términos de **calibración**, el modelo presentó un buen ajuste global, evidenciado por un test de Hosmer-Lemeshow no significativo ( $\chi^2 = 10.28$ ;  $p = 0.246$ ), lo que sugiere que las probabilidades predichas son consistentes con las observadas en la muestra.

Respecto a la **capacidad discriminativa**, el modelo alcanzó un área bajo la curva ROC (AUC) de aproximadamente 0.77, lo que indica una discriminación aceptable entre pacientes que fallecen y aquellos que sobreviven. Esto implica que, en promedio, el modelo asigna mayores probabilidades de muerte a los pacientes que efectivamente fallecen.

Además, se observó una **exactitud global** del 82.5%, cercana al No Information Rate (80%), lo que sugiere una mejora limitada respecto a un modelo trivial. *En otras palabras, si simplemente se predijera que todos los pacientes sobreviven, sin usar ningún modelo, ya se obtendría un 80% de aciertos.* Por tanto, el modelo solo mejora marginalmente ese desempeño (82.5% vs 80%), lo que indica que su capacidad real de clasificación es limitada y que la alta exactitud está influenciada por el desbalance de clases, más que por una verdadera capacidad predictiva sólida.

Por otro lado, el modelo mostró una **sensibilidad** muy baja (17.5%), indicando una pobre capacidad para identificar pacientes que fallecen, mientras que la **especificidad** fue muy alta (98.7%), reflejando una excelente capacidad para identificar pacientes que sobreviven. El valor predictivo positivo fue de 77.8% y el negativo de 82.7%. Adicionalmente, la **exactitud balanceada** balanced accuracy (el promedio entre la sensibilidad y la especificidad) fue de 58.1% y el **índice Kappa** de 0.23 (grado de concordancia entre las predicciones del modelo y los valores reales, ajustando por la concordancia que ocurriría por azar), lo que indica una concordancia débil.

Aunque el modelo presenta una adecuada calibración y una capacidad discriminativa aceptable, su utilidad clínica es limitada, principalmente debido a su baja sensibilidad. Esto implica que no es adecuado como herramienta de detección temprana de pacientes en alto riesgo de muerte, aunque podría ser útil en contextos donde se requiera confirmar bajo riesgo o estimar probabilidades a nivel poblacional.

## Desempeño global del modelo

Finalmente, existen unas medidas que evalúan el desempeño global del modelo entre las cuales en la actualidad son usadas más ampliamente la estadística R<sup>2</sup> de Nagelkerke y el score Brier.

Para los interesados en profundizar en estas dos estadísticas, se recomienda revisar la referencia:

- Steyerberg EW. Clinical prediction models. 2nd ed. Cham: Springer; 2019, capítulo 15

## Lecturas complementarias:

- Steyerberg EW. Clinical prediction models. 2nd ed. Cham: Springer; 2019. Capítulo 15.
- Kleinbaum DG, Klein M. Logistic regression: a self-learning text. 3rd ed. New York: Springer; 2010. Capítulos 6 y 10
- Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. *Journal of Clinical Epidemiology*. 1996;49(12):1373-1379.
- Riley RD, Ensor J, Snell KIE, Harrell FE Jr, Martin GP, Reitsma JB, et al. Calculating the sample size required for developing a clinical prediction model. *BMJ*. 2020;368:m441.
- Riley RD, Snell KIE, Ensor J, Burke DL, Harrell FE Jr, Moons KGM, Collins GS. Minimum sample size for developing a multivariable prediction model: PART II – binary and time-to-event outcomes.