

# Tarea Clasificación

Autor

**Sebastián Mena Aliaga**

Profesor

**Mario Guzmán**

Métodos Multivariados, Clasificación y Pronosticos

Magister en Estadística, PUCV

**Abril 2022**

# 1 Pregunta 1

Considerando  $N$  registros con características  $x$  en un espacio  $p$ -dimensional, y clasificados de forma binaria con codificación  $-N/N1$  (1),  $N/N2$  (2) con tamaño de clase  $N1$  y  $N2$  respectivamente. Se desea demostrar que la siguiente regla LDA clasifica la clase 2:

$$x^T \hat{\Sigma}^{-1} (\hat{\mu}_2 - \hat{\mu}_1) > \frac{1}{2} (\hat{\mu}_2 + \hat{\mu}_1)^T \hat{\Sigma}^{-1} (\hat{\mu}_2 - \hat{\mu}_1) - \ln \frac{N_2}{N_1}$$

Para la demostración se debe comenzar considerando la generalización de la aplicación del teorema de Bayes, que asigna la probabilidad que un conjunto de características  $x$  sea clasificada en algunos de los  $k$  grupos, dada por

$$\Pr(G = k \mid X = x) = \frac{f_k(x) \pi_k}{\sum_{l=1}^k f_l(x) \pi_l}$$

donde  $\pi_k = N_k/N$  es la probabilidad a priori del grupo  $k$ , y  $f_k(x)$  es la distribución de los datos del grupo  $k$  que sigue una ley gaussiana multivariada, anotada como

$$f_k(x) = \frac{1}{(2\pi)^{k/2} |\Sigma_k|} e^{-\frac{1}{2} (x - \hat{\mu}_k)^T \hat{\Sigma}^{-1} (x - \hat{\mu}_k)}$$

además, se debe considerar que el LDA sigue el supuesto que los grupos de clasificación poseen similaridad en sus matrices de varianza

$$\Sigma_k = \Sigma \quad \forall k$$

Dada las consideraciones anteriores, se comienza la demostración a partir de la siguiente aseveración. El ratio *odds*, expresado a continuación, da como resultado una mayor probabilidad de clasificar las características  $x$  en la clase 2, y en términos prácticos, se clasificará las características  $x$  en dicha clase

$$\frac{\Pr(G = 2 \mid X = x)}{\Pr(G = 1 \mid X = x)} > 1$$

si luego se aplica el logaritmo natural, se obtiene el log-*odds* o también llamado *logit*

$$\ln \frac{\Pr(G = 2 \mid X = x)}{\Pr(G = 1 \mid X = x)} > 0$$

se reexpresa las probabilidades por sus funciones

$$\ln \frac{f_2(x) \pi_2}{f_1(x) \pi_1} = \ln \frac{f_2(x)}{f_1(x)} + \ln \frac{\pi_2}{\pi_1} > 0$$

se denota la constante  $k = \frac{1}{(2\pi)^{k/2} |\Sigma_k|}$ , y se reexpresa las funciones desagregado por sus términos

$$\ln \frac{k e^{-\frac{1}{2} (x - \hat{\mu}_2)^T \hat{\Sigma}^{-1} (x - \hat{\mu}_2)}}{k e^{-\frac{1}{2} (x - \hat{\mu}_1)^T \hat{\Sigma}^{-1} (x - \hat{\mu}_1)}} + \ln \frac{N_2/N}{N_1/N} > 0$$

$$\ln e^{-\frac{1}{2} (x - \hat{\mu}_2)^T \hat{\Sigma}^{-1} (x - \hat{\mu}_2)} - \ln e^{-\frac{1}{2} (x - \hat{\mu}_1)^T \hat{\Sigma}^{-1} (x - \hat{\mu}_1)} + \ln \frac{N_2}{N_1} > 0$$

$$-\frac{1}{2} (x - \hat{\mu}_2)^T \hat{\Sigma}^{-1} (x - \hat{\mu}_2) - \left(-\frac{1}{2} (x - \hat{\mu}_1)^T \hat{\Sigma}^{-1} (x - \hat{\mu}_1)\right) + \ln \frac{N_2}{N_1} > 0$$

$$\frac{1}{2} \left( (x - \hat{\mu}_1)^T \hat{\Sigma}^{-1} (x - \hat{\mu}_1) - (x - \hat{\mu}_2)^T \hat{\Sigma}^{-1} (x - \hat{\mu}_2) \right) > -\ln \frac{N_2}{N_1}$$

por propiedad distributiva de la traspuesta

$$\frac{1}{2} \left( (x^T - \hat{\mu}_1^T) \hat{\Sigma}^{-1} (x - \hat{\mu}_1) - (x^T - \hat{\mu}_2^T) \hat{\Sigma}^{-1} (x - \hat{\mu}_2) \right) > -\ln \frac{N_2}{N_1}$$

por propiedad distributiva respecto de la suma de matrices por la izquierda

$$\frac{1}{2} \left( (x^T \hat{\Sigma}^{-1} (x - \hat{\mu}_1) - \hat{\mu}_1^T \hat{\Sigma}^{-1} (x - \hat{\mu}_1)) - (x^T \hat{\Sigma}^{-1} (x - \hat{\mu}_2) - \hat{\mu}_2^T \hat{\Sigma}^{-1} (x - \hat{\mu}_2)) \right) > -\ln \frac{N_2}{N_1}$$

$$\frac{1}{2} \left( (x^T \hat{\Sigma}^{-1} (x - \hat{\mu}_1) - \hat{\mu}_1^T \hat{\Sigma}^{-1} (x - \hat{\mu}_1) - x^T \hat{\Sigma}^{-1} (x - \hat{\mu}_2) + \hat{\mu}_2^T \hat{\Sigma}^{-1} (x - \hat{\mu}_2)) \right) > -\ln \frac{N_2}{N_1}$$

por propiedad distributiva respecto de la suma de matrices por la derecha

$$\begin{aligned} \frac{1}{2} \left( (x^T \hat{\Sigma}^{-1} x - x^T \hat{\Sigma}^{-1} \hat{\mu}_1) - (\hat{\mu}_1^T \hat{\Sigma}^{-1} x - \hat{\mu}_1^T \hat{\Sigma}^{-1} \hat{\mu}_1) \right. \\ \left. - (x^T \hat{\Sigma}^{-1} x - x^T \hat{\Sigma}^{-1} \hat{\mu}_2) + (\hat{\mu}_2^T \hat{\Sigma}^{-1} x - \hat{\mu}_2^T \hat{\Sigma}^{-1} \hat{\mu}_2) \right) > -\ln \frac{N_2}{N_1} \end{aligned}$$

$$\begin{aligned} \frac{1}{2} \left( x^T \hat{\Sigma}^{-1} x - x^T \hat{\Sigma}^{-1} \hat{\mu}_1 - \hat{\mu}_1^T \hat{\Sigma}^{-1} x + \hat{\mu}_1^T \hat{\Sigma}^{-1} \hat{\mu}_1 \right. \\ \left. - x^T \hat{\Sigma}^{-1} x + x^T \hat{\Sigma}^{-1} \hat{\mu}_2 + \hat{\mu}_2^T \hat{\Sigma}^{-1} x - \hat{\mu}_2^T \hat{\Sigma}^{-1} \hat{\mu}_2 \right) > -\ln \frac{N_2}{N_1} \end{aligned}$$

$$\frac{1}{2} \left( -x^T \hat{\Sigma}^{-1} \hat{\mu}_1 - \hat{\mu}_1^T \hat{\Sigma}^{-1} x + \hat{\mu}_1^T \hat{\Sigma}^{-1} \hat{\mu}_1 + x^T \hat{\Sigma}^{-1} \hat{\mu}_2 + \hat{\mu}_2^T \hat{\Sigma}^{-1} x - \hat{\mu}_2^T \hat{\Sigma}^{-1} \hat{\mu}_2 \right) > -\ln \frac{N_2}{N_1}$$

factorizando

$$\frac{1}{2} \left( x^T \hat{\Sigma}^{-1} (\hat{\mu}_2 - \hat{\mu}_1) - \hat{\mu}_1^T \hat{\Sigma}^{-1} x + \hat{\mu}_1^T \hat{\Sigma}^{-1} \hat{\mu}_1 + \hat{\mu}_2^T \hat{\Sigma}^{-1} x - \hat{\mu}_2^T \hat{\Sigma}^{-1} \hat{\mu}_2 \right) > -\ln \frac{N_2}{N_1}$$

$$\frac{1}{2} \left( x^T \hat{\Sigma}^{-1} (\hat{\mu}_2 - \hat{\mu}_1) + (\hat{\mu}_2 - \hat{\mu}_1)^T \hat{\Sigma}^{-1} x + \hat{\mu}_1^T \hat{\Sigma}^{-1} \hat{\mu}_1 - \hat{\mu}_2^T \hat{\Sigma}^{-1} \hat{\mu}_2 \right) > -\ln \frac{N_2}{N_1}$$

además, se puede demostrar que

$$(\hat{\mu}_2 - \hat{\mu}_1)^T \hat{\Sigma}^{-1} x = x^T \hat{\Sigma}^{-1} (\hat{\mu}_2 - \hat{\mu}_1)$$

y

$$\hat{\mu}_1^T \hat{\Sigma}^{-1} \hat{\mu}_1 - \hat{\mu}_2^T \hat{\Sigma}^{-1} \hat{\mu}_2 = (\hat{\mu}_2 + \hat{\mu}_1)^T \hat{\Sigma}^{-1} (\hat{\mu}_2 - \hat{\mu}_1)$$

entonces

$$\frac{1}{2} \left( 2 x^T \hat{\Sigma}^{-1} (\hat{\mu}_2 - \hat{\mu}_1) - (\hat{\mu}_2 + \hat{\mu}_1)^T \hat{\Sigma}^{-1} (\hat{\mu}_2 - \hat{\mu}_1) \right) > -\ln \frac{N_2}{N_1}$$

$$x^T \hat{\Sigma}^{-1} (\hat{\mu}_2 - \hat{\mu}_1) - \frac{1}{2} (\hat{\mu}_2 + \hat{\mu}_1)^T \hat{\Sigma}^{-1} (\hat{\mu}_2 - \hat{\mu}_1) > -\ln \frac{N_2}{N_1}$$

así llegamos a la expresión final

$$x^T \hat{\Sigma}^{-1} (\hat{\mu}_2 - \hat{\mu}_1) > \frac{1}{2} (\hat{\mu}_2 + \hat{\mu}_1)^T \hat{\Sigma}^{-1} (\hat{\mu}_2 - \hat{\mu}_1) - \ln \frac{N_2}{N_1}$$

$\therefore$  queda demostrado que la regla LDA clasifica a la clase 2

## 2 Pregunta 2

Se desea abordar el desafío de clasificar clientes que se fugaron de cierta compañía, para ello, se define la siguiente metodología de trabajo: Identificar modelos de clasificación a utilizar, realizar un análisis exploratorio de datos, evaluación de datos atípicos y/o outliers, entrenamiento y validación del modelo.

### 2.1 Análisis Exploratorio de Datos

El *dataset* cuenta con un total de 3333 registros, 20 variables y sin NA, cada registro se trata de un cliente de una cierta compañía, y presenta variables como el estado de donde proviene el cliente, si posee o no plan internacional, total de minutos diarios, cantidad de llamadas diarias, sus cargos diarios, entre otros. Además, el *dataset* cuenta con la variable de respuesta binaria *churn*, en la tabla 1 se observa que clasifica en *True* o *False* si el cliente a abandonado la empresa o no, ambas cuenta con 483 y 2850 registros respectivamente, mostrando un claro desbalance entre las clases.

<b>churn</b>	
False	2850
True	483

Figure 1: Variable *churn* de respuesta

#### 2.1.1 AED variables explicativas categóricas

El *dataset* cuenta con cuatro variables categóricas, descartando *phone.number* dado que es un registro único para cada cliente. En la tabla 2 observamos el conteo de las primeras 6 de 51 categorías de la variable *state*, la variable *area.code* cuenta con 3 categorías y las variables *international.plan* y *voice.mail.plan* son binarias de respuesta *yes* y *no*.

<b>state</b>		<b>area.code</b>			
WV	106				
MN	84				
NY	83	408	838		
AL	80	415	1655	<b>international.plan</b>	323 3010
OH	78	510	840	<b>voice.mail.plan</b>	922 2411
OR	78				

Figure 2: Tabla resumen de variable categoricas del *dataset*

### 2.1.2 AED variables explicativas cuantitativas

El *dataset* cuenta con 14 variables cuantitativas, de ellas se considera las cantidades totales de minutos, llamadas y su cargo, durante la mañana, tarde, noche e internacional, además de la cantidad total de llamadas a la mesa de ayuda y la cantidad total de mensajes virtuales.

El total de minutos de los clientes en la mañana, tarde, noche e internacional, tal como se observa en las figuras anexas A.1 a A.4, son variables continuas entre 0 y 395 minutos, se observa que la menor media en los minutos totales utilizados de los clientes son internacionales y la mayor en la tarde.

El total de llamadas de los clientes en la llamada, tarde, noche e internacional, tal como se observa en las figuras anexas A.5 y A.8, son variables numéricas discretas entre 0 y 175, la mayor mediana de llamadas totales se realizan en la mañana, y el máximo se observa durante la noche.

El total del cargo de los clientes en el día, tarde, noche e internacional, tal como se observa en las figuras anexas A.9 a A.12, son variables continuas entre 0 y 59.64, el mayor cargo total medio de los clientes es durante el día y el menor son las internacionales.

En todas las variables anteriormente señaladas se observan *outliers*, para los minutos se presencian 24, 25, 30 y 46 *outliers* durante el día, tarde, noche e internacional respectivamente, para las llamadas se presencia 25, 20, 22 y 78 respectivamente, y los cargos presentan 25, 24, 30 y 49 respectivamente, se observan cantidades similares para cada grupo, destacando las internacionales como las que presentan mayor cantidad de *outliers*.

La cantidad total de llamadas a la mesa de servicio *customer.service.calls*, tal como se observa en la figura anexa A.13, es una variable numérica discreta entre 0 y 9, con una mediana de 1 llamada realizada, además se observan 67 *outliers* entre los valores 4 y 9.

La cantidad total de mensajes virtuales *number.vmail.messages*, tal como se observa en la figura anexa A.14, es una variable cuantitativa discretas entre 0 y 51, el 77% de los registros cuenta con 0 mensajes virtuales realizados, además se observa un unico *outlier* con valor 51.

## 2.2 Modelo de Discriminante Lineal (LDA)

El Método de Discriminante Lineal (LDA) es utilizado cuando se tiene un conjunto de variables predictoras y se desea clasificar una variable de respuesta en dos o más clases, este método se engloba dentro de los métodos de clasificación supervisados. LDA hace uso del enfoque del teorema de Bayes para estimar la probabilidad que un conjunto de características pertenece a un determinado grupo, y su capacidad de clasificar correctamente depende de que tan buenas son las estimaciones a priori del grupo ( $\pi_k$ ) y la densidad de los datos ( $f_k(x)$ ). La linealidad del método se debe al hecho que la función discriminatoria es lineal respecto a las variables explicativas.

Para utilizar LDA es necesario tomar en cuenta que las variables predictoras deben ser de tipo cuantitativas, deben seguir una distribución gaussiana multivariante para cada grupo a

predecir y se debe asumir homogeneidad en sus varianzas, dicho lo anterior, para utilizar LDA se excluyen las variables categóricas *state*, *area.code*, *international.plan* y *voice.mail.plan* para el modelo de fuga de cliente.

Al realizar el test de Mardia para resolver la hipótesis de normalidad multivariante para todas las variables continuas, la curtosis de mardia aporta un  $p$ -value de 0.091 y la asimetría de mardia  $p$ -value cercana a 0, por lo que no hay evidencia para aseverar que se distribuye normal. Interesantemente, al remover todos los *outliers* de las variables continuas ambos estadísticos entregan un  $p$ -value cercano a cero, rechazando su normalidad.

Al realizar el test M de homocedasticidad de la matriz de covarianza entre ambos grupos de la variable de respuesta *churn*, considerando todas las variables cuantitativas, se observa un  $p$ -value  $< 2.2e - 16$  rechazando la hipótesis nula de homocedasticidad, sin embargo, al retirar las variables *number.vmail.messages*, *customer.service.calls*, *total.intl.charge*, *total.intl.minutes*, *total.day.minutes*, *total.day.charge*, el test entrega un  $p$ -value = 0.3184 se puede asumir la homocedastidad de la matriz.

Luego de definir las variables a utilizar, se define un muestreo aleatorio de entrenamiento (70%) y testeo (30%), los datos de entrenamiento se utiliza para estimar los parámetros del modelo y luego el subconjunto de testeo para revisar el *performance* del modelo estimado. Las predicciones realizadas con el modelo estimado aporta un  $accuracy = 0.841$ ,  $recall = 1$ , y  $F1 = 0.9136$ , si bien los estadísticos aportan buenos resultados de predicción, podemos observar en la tabla 3 de confusión que el modelo no predice ningún verdadero ni falso positivo, sin embargo, entendiendo la naturaleza del problema, existe más valor poder clasificar la categoría positiva debido a que son los clientes que si se fugan de la empresa.

	False	True
False	841	159
True	0	0

Figure 3: Matriz de Confusión LDA

Para resolver el problema anterior, se realiza un balanceo de los datos por técnica de *oversampling*, sin embargo, este modelo aporta los estadísticos  $accuracy = 0.5465$ ,  $recall = 0.5347$ , y  $F1 = 0.5353$ , y tal como se obseta en la tabla 4 y la curva ROC de la figura, esta nueva estimación del modelo posee un muy bajo *performance*, teniendo una predicción no tan lejana a la de lanzar una moneda. Queda como trabajo revisar nuevas combinaciones variables predictoras dentro del modelo y nuevos métodos de balanceo para encontrar un modelo con un mejor rendimiento.

	False	True
False	447	387
True	389	488

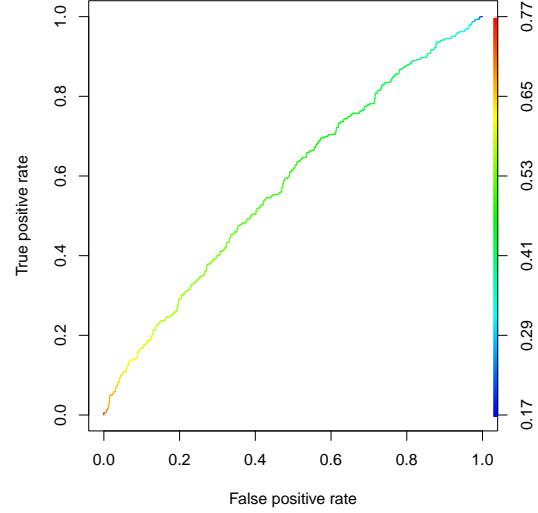


Figure 4: Matriz de Confusión LDA balanceada y curva ROC

## 2.3 Regresión Logística

La Regresión Logística es un método de regresión que permite estimar la probabilidad de una variable cualitativa binaria en función de una variable cuantitativa, dicha regresión es una extensión de una regresión lineal con uso de la función sigmoide para predecir variables categóricas. Se debe tener en consideración los supuestos de independencia entre las variables predictoras y baja colinealidad entre los predictores.

$$Pr(G = k|X = x) = \frac{\exp(\beta_k 0 + \beta_k^T x)}{1 + \exp(\beta_k 0 + \beta_k^T x)}$$

Las variables consideradas son tanto de tipos cualitativas y cuantitativas, las variables cualitativas se han transformado a tipos *dummys* para ser incorporada a la regresión, y las variables cuantitativas se han estandarizado según el método normal estándar para evitar problemas de escalamiento. Posteriormente se seleccionará el modelo utilizando métodos *forward*, *backward* y *stepwise*, aportando los valores  $AIC = 1554.27$ ,  $AIC = 1554.85$  y  $AIC = 1554.27$  respectivamente, luego de las iteraciones realizadas. Por lo tanto, el modelo óptimo seleccionado por el método *forward* para la estimación de *churn* considera las siguientes variables:

$\text{churn} \sim \text{total.day.charge} + \text{international.plan.no} + \text{customer.service.calls} + \text{total.eve.minutes} + \text{voice.mail.plan.no} + \text{total.intl.charge} + \text{total.intl.calls} + \text{total.night.charge} + \text{state.MT} + \text{state.VA} + \text{state.TX} + \text{state.CA} + \text{state.WA} + \text{state.MI} + \text{state.NJ} + \text{state.SC} + \text{state.MS} + \text{state.RI} + \text{state.HI} + \text{state.MA}$

Luego se realiza la estimación del modelo de Regresión Logística de las variables seleccionadas, los coeficientes están anexas en la tabla A.15.



Una vez estimado el modelo, se realiza el test Kolmogorov-Smirnov (KS) para evaluar el desempeño del modelo, nos permite compara si el modelo teórico posee alguna diferencia significativa con la empírica. En este caso, el test aporta un  $p - value = 0.5125$ , rechazando la hipótesis de diferencias significativas, por lo que se podría estar en presencia de un buen modelo.

Además, se puede observar los estadísticos de *performance* del modelo luego de realizar las predicciones y contrastarlas con los resultados originales dato de testing, los estadísticos tales como  $accuracy = 0.877$ ,  $recall = 0.9896$  y  $F1 = 0.9896$ , además, observando la matriz de confusión y curva ROC de la figura 5, se concluye que el modelo entrenado posee un buen desempeño para predecir las fugas de clientes de la empresa según las características que se han reconocido como significantes al momento de clasificar.

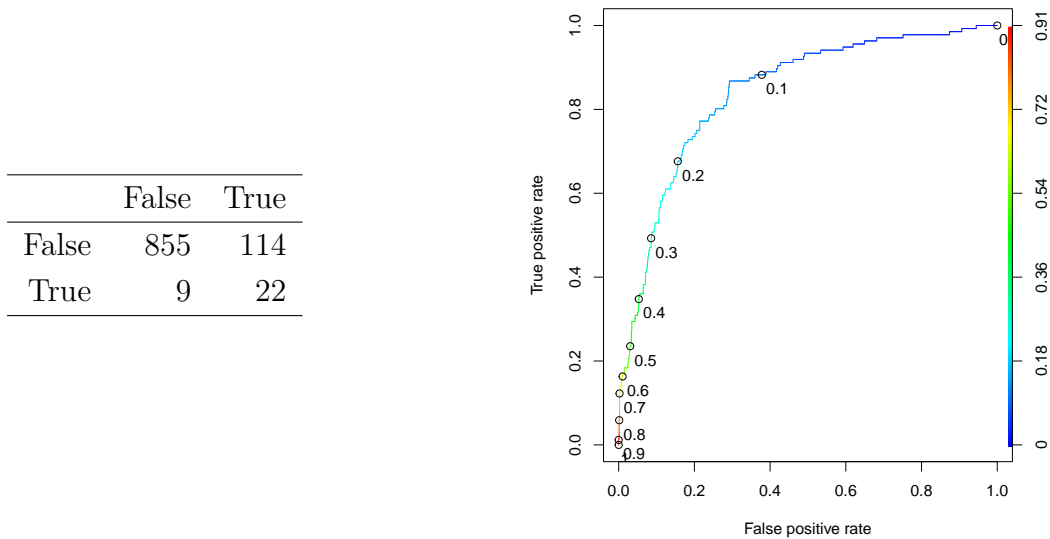


Figure 5: Matriz de Confusión de la Regresión Logística y curva ROC

# Anexo

	total.day.minutes
X	Min. : 0.0
X.1	1st Qu.:143.7
X.2	Median :179.4
X.3	Mean :179.8
X.4	3rd Qu.:216.4
X.5	Max. :350.8

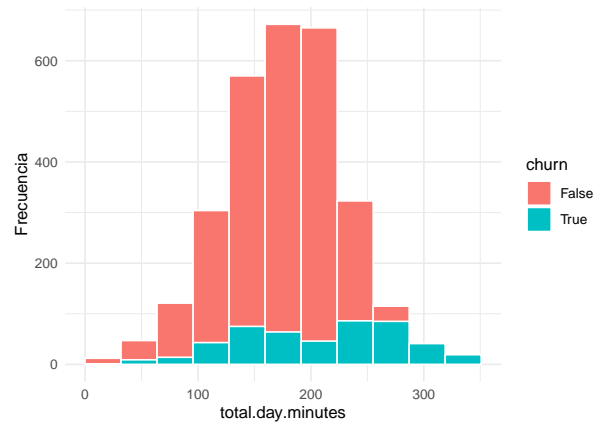


Figure A.1: Resumen e histograma de *total.day.minutes*

	total.eve.minutes
X	Min. : 0.0
X.1	1st Qu.:166.6
X.2	Median :201.4
X.3	Mean :201.0
X.4	3rd Qu.:235.3
X.5	Max. :363.7

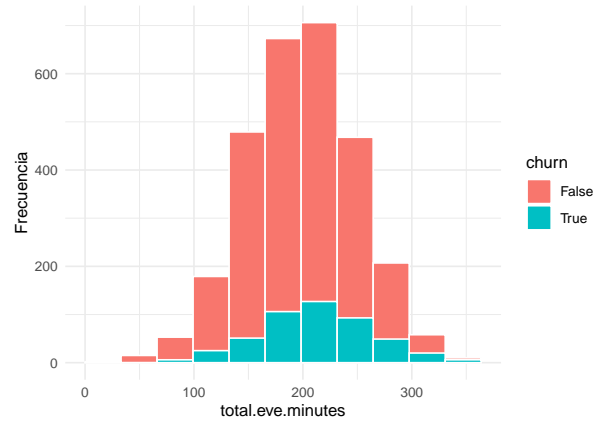


Figure A.2: Resumen e histograma de *total.eve.minutes*

	total.night.minutes
X	Min. : 23.2
X.1	1st Qu.:167.0
X.2	Median :201.2
X.3	Mean :200.9
X.4	3rd Qu.:235.3
X.5	Max. :395.0

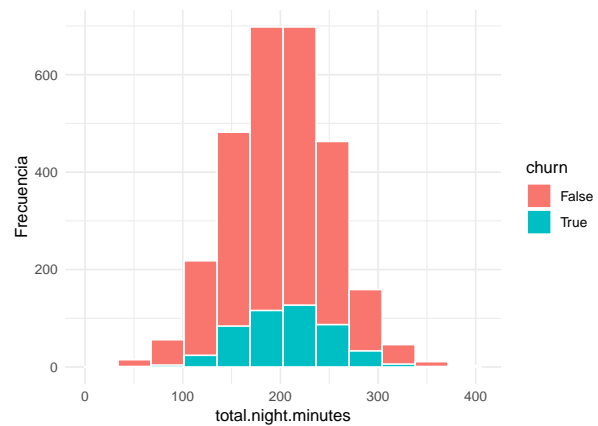


Figure A.3: Resumen e histograma de *total.night.minutes*

total.intl.minutes	
X	Min. : 0.00
X.1	1st Qu.: 8.50
X.2	Median :10.30
X.3	Mean :10.24
X.4	3rd Qu.:12.10
X.5	Max. :20.00

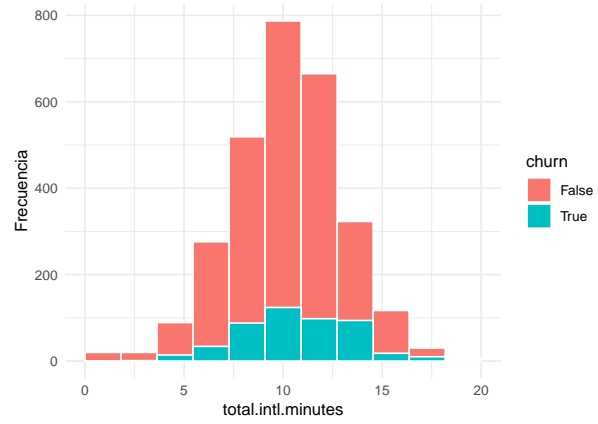


Figure A.4: Resumen e histograma de *total.intl.minutes*

total.day.calls	
X	Min. : 0.0
X.1	1st Qu.: 87.0
X.2	Median :101.0
X.3	Mean :100.4
X.4	3rd Qu.:114.0
X.5	Max. :165.0

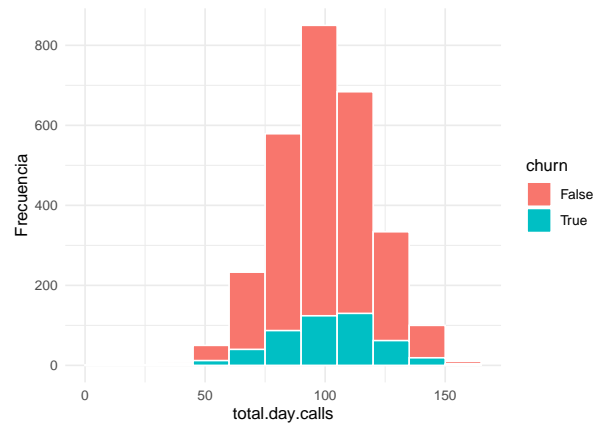


Figure A.5: Resumen e histograma de *total.day.calls*

total.eve.calls	
X	Min. : 0.0
X.1	1st Qu.: 87.0
X.2	Median :100.0
X.3	Mean :100.1
X.4	3rd Qu.:114.0
X.5	Max. :170.0

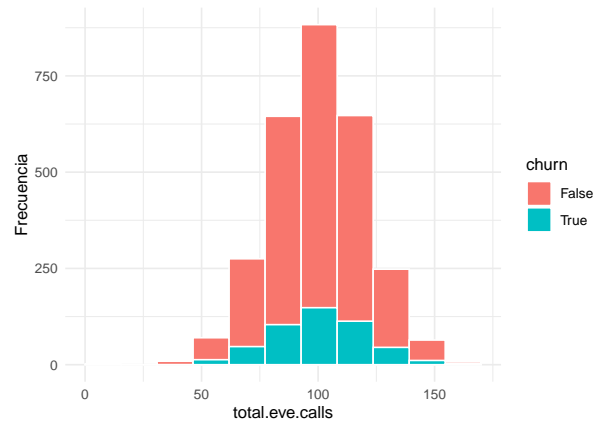


Figure A.6: Resumen e histograma de *total.eve.calls*

	total.night.calls
X	Min. : 33.0
X.1	1st Qu.: 87.0
X.2	Median :100.0
X.3	Mean :100.1
X.4	3rd Qu.:113.0
X.5	Max. :175.0

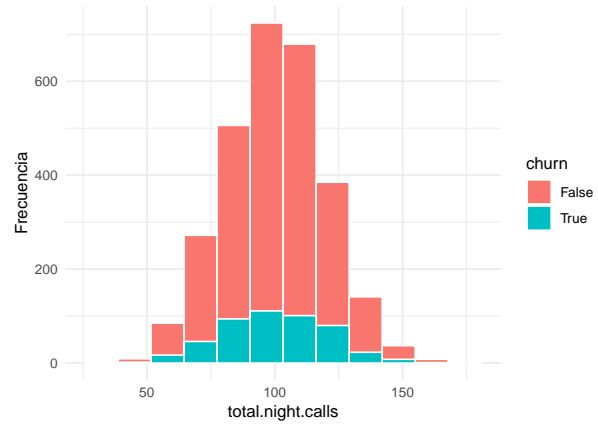


Figure A.7: Resumen e histograma de *total.night.calls*

	total.intl.calls
X	Min. : 0.000
X.1	1st Qu.: 3.000
X.2	Median : 4.000
X.3	Mean : 4.479
X.4	3rd Qu.: 6.000
X.5	Max. :20.000

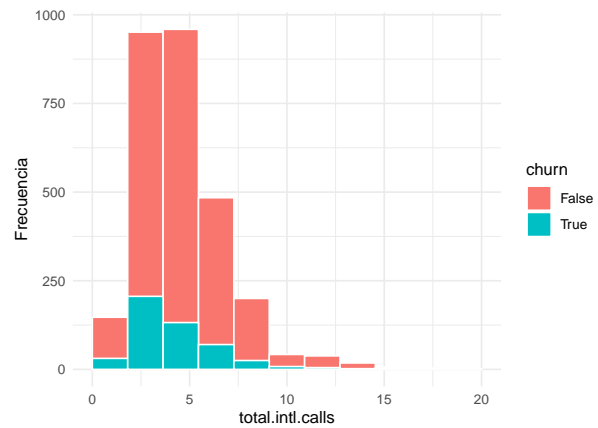


Figure A.8: Resumen e histograma de *total.intl.calls*

	total.day.charge
X	Min. : 0.00
X.1	1st Qu.:24.43
X.2	Median :30.50
X.3	Mean :30.56
X.4	3rd Qu.:36.79
X.5	Max. :59.64

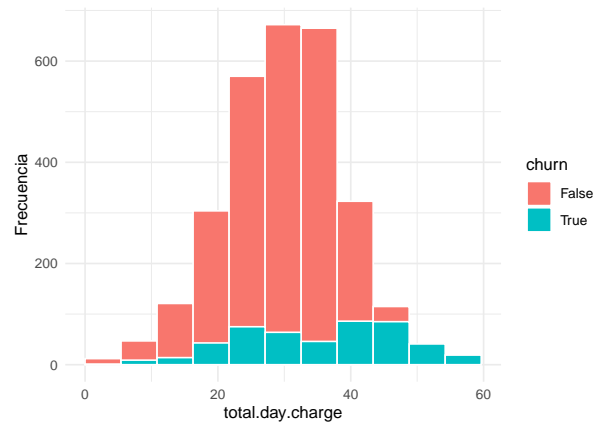


Figure A.9: Resumen e histograma de *total.day.charge*

	total.eve.charge
X	Min. : 0.00
X.1	1st Qu.:14.16
X.2	Median :17.12
X.3	Mean :17.08
X.4	3rd Qu.:20.00
X.5	Max. :30.91

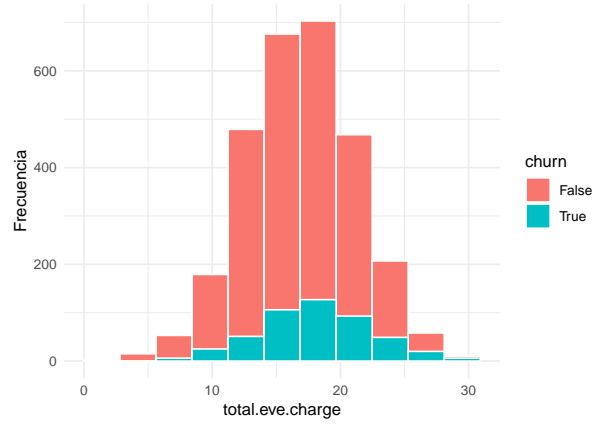


Figure A.10: Resumen e histograma de *total.eve.charge*

	total.night.charge
X	Min. : 1.040
X.1	1st Qu.: 7.520
X.2	Median : 9.050
X.3	Mean : 9.039
X.4	3rd Qu.:10.590
X.5	Max. :17.770

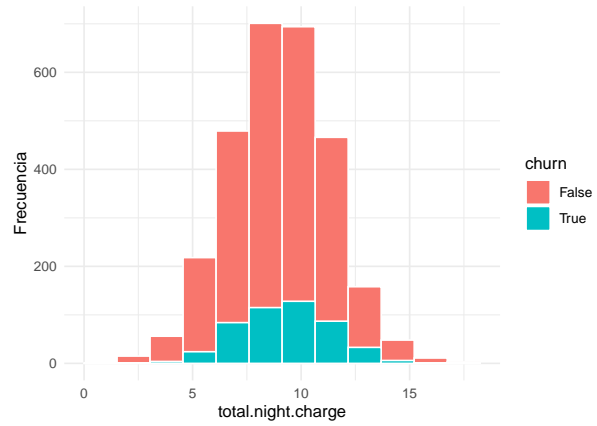


Figure A.11: Resumen e histograma de *total.night.charge*

	total.intl.charge
X	Min. :0.000
X.1	1st Qu.:2.300
X.2	Median :2.780
X.3	Mean :2.765
X.4	3rd Qu.:3.270
X.5	Max. :5.400

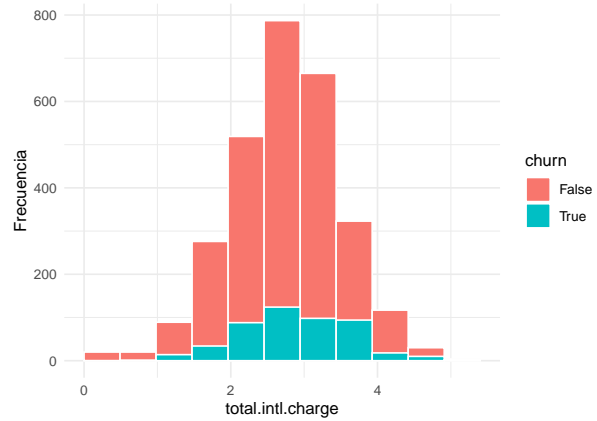


Figure A.12: Resumen e histograma de *total.intl.charge*

customer.service.calls	
X	Min. :0.000
X.1	1st Qu.:1.000
X.2	Median :1.000
X.3	Mean :1.563
X.4	3rd Qu.:2.000
X.5	Max. :9.000

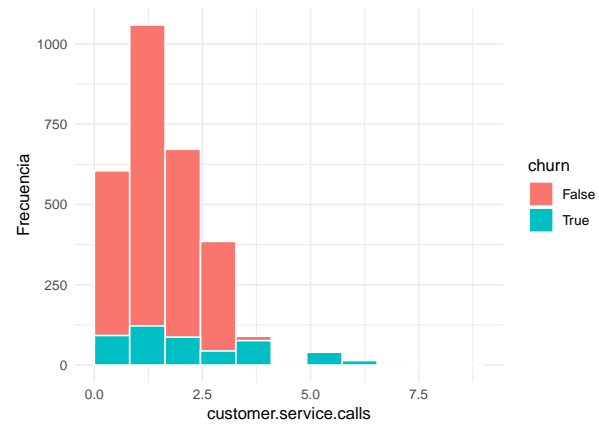


Figure A.13: Resumen e histograma de *customer.service.calls*

number.vmail.messages	
X	Min. : 0.000
X.1	1st Qu.: 0.000
X.2	Median : 0.000
X.3	Mean : 8.099
X.4	3rd Qu.:20.000
X.5	Max. :51.000

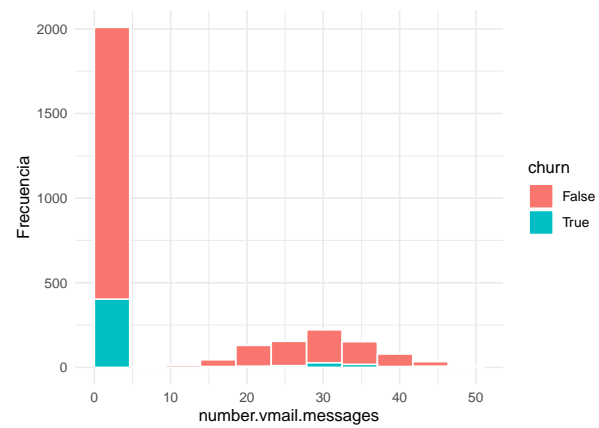


Figure A.14: Resumen e histograma de *number.vmail.messages*

	x
(Intercept)	-1.13
total.day.charge	0.72
international.plan.no	-2.03
customer.service.calls	0.67
total.eve.minutes	0.37
voice.mail.plan.no	0.75
total.intl.charge	0.28
total.intl.calls	-0.23
total.night.charge	0.20
state.MT	1.21
state.VA	-1.14
state.TX	0.98
state.CA	1.40
state.WA	0.98
state.MI	0.80
state.NJ	0.76
state.SC	0.86
state.MS	0.74
state.RI	-0.86
state.HI	-1.41
state.MA	0.73

Figure A.15: Coeficientes del modelo de regresión logística optimo po método *forward*