

Factores asociados a los intentos suicidas en los adolescentes de Uruguay

Leidy Gómez Medina

15/12/2020

Resumen

Con el pasar de los años el suicidio se ha convertido en una de las 10 principales causas de muertes a nivel mundial. Lo que ha llevado a los gobiernos y ,sin ser una excepción, a Uruguay a prestarle mayor atención a este fenómeno que está afectando y aumentando en poblaciones cada vez más jóvenes. Siendo los adolescentes uruguayos el centro de atención del estudio que se llevará a cabo en este proyecto.

Palabras Clave: *Suicidio,adolescentes, logit, probit.*

Introducción

Debemos definir el suicidio como la acción orientada a causar la propia muerte de manera voluntaria, esta razón ubica al suicidio entre las 10 principales causas de muertes a nivel mundial. Se pueden tener varios factores a considerar en los motivos que una persona tiene para suicidarse, entre los más conocidos tenemos, psicopatológicos (depresión, psicosis, abuso de drogas), sociales (aislamiento social, situación laboral), biológicas (heredabilidad) y ambientales (sucesos estresantes, maltrato y abuso de cualquier tipo).

Por otro lado, el periodo de la adolescencia es estresante y está lleno de cambios principalmente hormonales, sentimentales y de pensamiento, además el estrés, la confusión, el miedo, la incertidumbre, la presión para tener éxito y la capacidad de pensar sobre cosas en nuevas formas, influyen de manera directa en las capacidades de un adolescente para resolver problemas y tomar decisiones. Así mismo, para algunos adolescentes es crítico el afrontar los cambios normales del desarrollo cuando se combinan con otros eventos como: el divorcio de los padres, mudarse a un lugar desconocido, perder amistades, problemas escolares, bullying, abuso de cualquier tipo, etc. Dado a lo anterior, los factores llegan a ser abrumadores y muy difíciles de superar lo que lleva a contemplar el suicidio como una solución definitiva para algunos de estos jóvenes.[1]

Uruguay es el país de América de Sur con la tasa de suicidios más alta y el segundo de América Latina después de Cuba, desde 1990 las cifras de suicidios vienen aumentando y en el año 2002 se tuvo el épico máximo con 683 suicidios, esto coincidiendo con una de las peores crisis económicas que ha sufrido el país en su historia.[3]

Descripción de los Datos

Se utilizó datos de la encuesta realizada por GSHS en el año de 2012 [2] del país de Uruguay, donde se utilizó un diseño de muestreo en dos etapas para producir una muestra representativa de los alumnos de 1º, 2º y 3º año del Ciclo Básico y 1er. año de Bachillerato, donde todos los alumnos de cada clase muestreada eran aptos para participar en la encuesta GSHS y cuyo tamaño de la muestra seleccionada fue de 3524 alumnos.

Variables

Se seleccionan 13 de las 126 variables que tiene la base de datos original, a continuación se mostrarán cuáles variables fueron seleccionadas y se dará una breve descripción de las mismas.

Variable	Descripción
Q1; Edad	Edad del estudiante (Organizados seis grupos)
Q2: Sexo	Sexo del estudiante (Femenino o Masculino)
Q4: Altura	Altura del estudiante en metros
Q5: Peso	Peso del estudiante en kilogramos
QN20: Bullying	Ha sufrido bullying en los últimos 30 días
QN22: Soledad	ha sentido solo en el último año
QN23: Sueño	La preocupación impidió dormir la mayor parte del tiempo
QN25: Plan	Hizo un plan para suicidarse durante el último año
QN27: Amigos	El estudiante tiene amigos cercanos
QN35: Alcohol	Ha ingerido alcohol en los últimos 30 días
QN44: Sexual	Ha tenido relaciones sexuales en algun momento
QN53: Clases	Ha faltado a clases sin permiso en los últimos 30 días
QN56: Padres	Los padres siempre entienden los problemas

Análisis Descriptivo Univariado

Nuestro objetivo es analizar que variables explican el intento de suicidio de los adolescentes en Uruguay, es por ello, que se comenzará con un análisis descriptivo univariado, donde podemos obtener las relaciones que hay entre las variable explicativas y la variable explicada.

A continuación se obtiene la matriz con los P valores de Chi-Squared test, que tiene como hipótesis nula la independencia entre las variables de estudio de tipo cualitativas.

En la anterior tabla cuando el p-valor sea menor al valor de significancia 0.05 nos va indicar que no hay independencia entre las variables implicadas. Por ejemplo, hay dependencia entre Q35: El estudiante ha consumido alcohol en los últimos 30 días y Q44: El estudiante ha tenido relaciones sexuales alguna vez en si vida. Se observa que hay un problema, porque no hay independencia entre la mayoría de las variables ya que su interacción representa un valor menor al valor de significancia 0.05.

Procedemos a analizar las variables continuas usando "Welch two sample t-test" la cual es una prueba de dos muestras que se utiliza para contrastar la hipótesis nula que hay independencia entre ambas variables contra la hipótesis que no hay independencia entre ambas variables. Estas variables a analizar son las cuantitativas Q4 y Q5 respectivamente contra la variable respuesta Q20.

	Q1	Q2	Q25	QN16	QN20	QN22	QN23	QN27	QN35	QN44	QN53	QN56
Q1	0.00	0.12	0.02	0.00	0.31	0.05	0.06	0.11	0.00	0.00	0.00	0.00
Q2	0.13	0.00	0.00	0.00	0.10	0.00	0.00	0.04	0.04	0.00	0.87	0.02
Q25	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.07	0.00	0.00	0.00	0.00
QN16	0.00	0.00	0.00	0.00	0.00	0.50	0.00	0.60	0.00	0.00	0.00	0.00
QN20	0.32	0.10	0.00	0.00	0.00	0.00	0.00	0.14	0.03	0.83	0.00	0.00
QN22	0.06	0.00	0.00	0.50	0.00	0.00	0.00	0.00	0.02	0.01	0.00	0.00
QN23	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.04	0.00
QN27	0.11	0.03	0.06	0.60	0.14	0.00	0.01	0.00	0.65	0.00	0.12	0.01
QN35	0.00	0.04	0.00	0.00	0.03	0.02	0.00	0.66	0.00	0.00	0.00	0.00
QN44	0.00	0.00	0.00	0.00	0.85	0.00	0.00	0.00	0.00	0.00	0.00	0.00
QN53	0.00	0.87	0.00	0.00	0.00	0.00	0.04	0.11	0.00	0.00	0.00	0.00
QN56	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00

Welch Two Sample t-test

data: Q4 by QN25

t = -2.6702, df = 355.81, p-value = 0.007928

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-0.022770064 -0.003454938

sample estimates:

mean in group Si mean in group No
1.623431 1.636543

Para este primer análisis se toman Q4 contra QN25. En este caso se rechaza la hipótesis nula de que hay independencia entre ambas variables, es decir, la altura puede ser una variable influyente en nuestro estudio. Observando que la altura media de los estudiantes en el último mes es de 1.63 metros mientras que en el otro grupo la altura media es de 1.64 metros.

Welch Two Sample t-test

data: Q5 by QN25

t = 0.18659, df = 342.27, p-value = 0.8521

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-1.346682 1.628969

sample estimates:

mean in group Si mean in group No
57.64599 57.50484

Para este segundo análisis se toman Q5 contra QN 25. En este caso se no se rechaza la hipótesis nula de que hay independencia entre variables de peso y el intento de suicidio, es decir, el peso del estudiante no afectará mucho al estudio.

Metodología

Debido a que la variable "Ha hecho un plan para suicidarse" que deseamos explicar es cualitativa utilizaremos modelos lineales generalizados, luego se hara una comparacion entre los modelos logit, probit y Cauchy donde veremos cual es el mas adecuado para los datos que estamos utilizando en este analisis, iniciaremos dando una breve introduccion teorica de cada uno de estos, para luego ver la implementación y resultados en nuestras variables seleccionadas del modelo adecuado.

Modelos lineales generalizados

En estadística, un modelo lineal generalizado (MLG) es una generalización flexible de la regresión lineal ordinaria que permite variables de respuesta con modelos de distribución de errores distintos de una distribución normal.

Un MLG generaliza la regresión lineal ya que permite que el modelo lineal se relacione con la variable de respuesta mediante una función de enlace y al permitir que la magnitud de la varianza de cada medición sea una función de su valor predicho.

Un MLG cumple los siguientes supuestos

- Se requiere que los errores distribuyan normal asintóticamente.
- Se requiere independencia entre las observaciones
- Debe haber poca o nula multicolinealidad entre las variables explicativas.
- Es necesaria independencia entre los errores y las variables explicativas.

Construcción del modelo

En este caso se utilizará el Modelo Log-lineal con distribución binomial dado al objetivo de nuestro estudio.

Un primer modelo

El primero modelo que vamos a usar es el modelo con todas las variables

$$Y_i = \frac{1}{1 + e^{-x_i^T \beta}} \quad (1)$$

Donde X son los valores de la matriz diseño y β un vector con los parámetros a estimar. teniendo en cuenta que las variables Q1, Q2, QN20, QN22, QN23, QN27, QN35, QN44, QN53 y QN56, son factores planteamos el siguiente modelo en R con todas las variables explicativas y sus interacciones, de este modelo utilizamos la funcion stepAIC para tomar las variables e interacciones significativas. De lo anterior, obtenemos el siguiente modelo.

Call:

```
glm(formula = QN25 ~ Q2 + QN20 + QN22 + QN23 + QN35 + QN44 +  
    QN53 + QN56 + Q2:QN20 + QN20:QN44 + QN22:QN44 + QN20:QN53 +
```

```
QN23:QN53 + QN35:QN53 + QN22:QN23, family = binomial(link = "logit"),
data = Datos3)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.8160	0.2355	0.3339	0.4617	1.5384

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	0.1431	0.5285	0.271	0.786594	
Q2Femenino	-0.9985	0.2789	-3.580	0.000343	***
QN20No	0.2476	0.3373	0.734	0.462945	
QN22No	-0.1192	0.4376	-0.272	0.785365	
QN23No	0.9640	0.4815	2.002	0.045270	*
QN35No	0.1568	0.2505	0.626	0.531305	
QN44No	0.3965	0.3978	0.997	0.318938	
QN53No	0.3617	0.4862	0.744	0.456947	
QN56No	-0.4665	0.1419	-3.287	0.001013	**
Q2Femenino:QN20No	0.6241	0.3273	1.907	0.056503	.
QN20No:QN44No	-0.4753	0.3242	-1.466	0.142587	
QN22No:QN44No	0.8684	0.3769	2.304	0.021223	*
QN20No:QN53No	0.7158	0.3303	2.167	0.030199	*
QN23No:QN53No	-0.8191	0.4356	-1.881	0.060015	.
QN35No:QN53No	0.5558	0.3073	1.809	0.070439	.
QN22No:QN23No	0.9496	0.4575	2.075	0.037942	*

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1761.9 on 2648 degrees of freedom
Residual deviance: 1481.3 on 2633 degrees of freedom
AIC: 1513.3

Donde encontramos que el modelo con menor AIC y con el que vamos a trabajar es

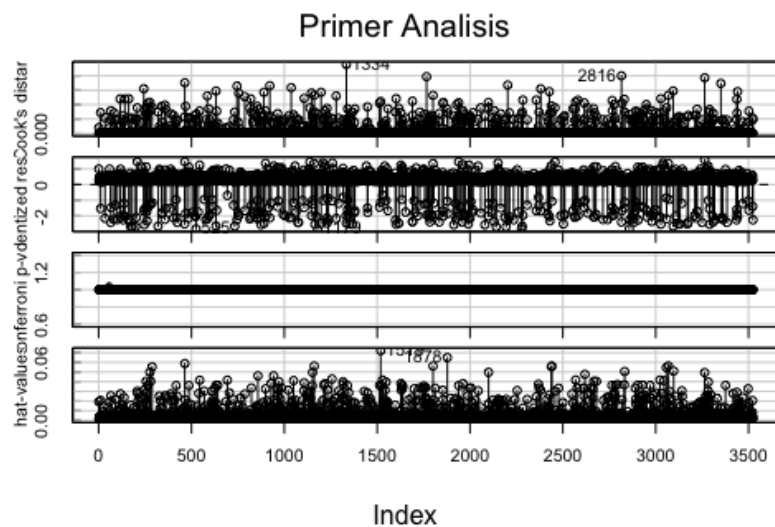
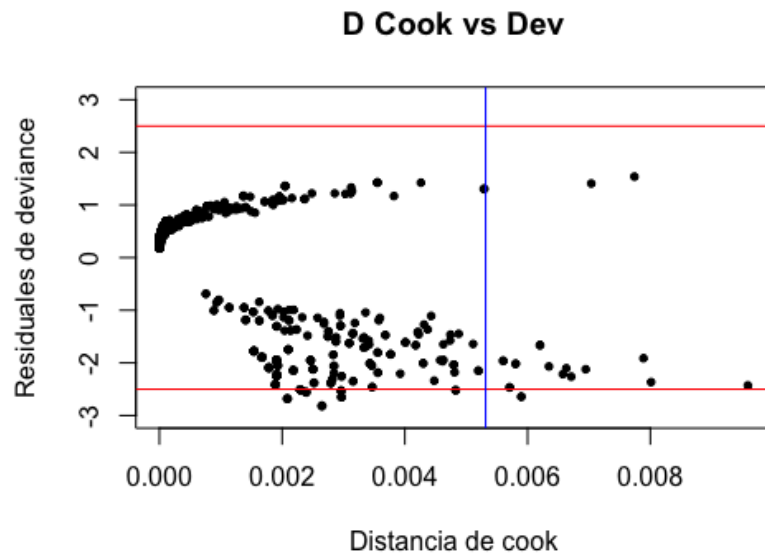
$$Y = \frac{1}{1 + e^{-(Z(X)_{ij})}}$$

con

$$\begin{aligned} Z(X)_{ij} = & \beta_0 + \beta_1 Q2_i + \beta_2 QN20_i + \beta_3 QN22_i + \beta_4 QN23_i + \beta_5 QN35_i \\ & + \beta_6 QN44_i + \beta_7 QN53_i + \beta_8 QN53_i + \beta_9 (Q2 * QN20)_{ij} \\ & + \beta_{10} (QN20 * QN44)_{ij} + \beta_{11} (QN22 * QN44)_{ij} \\ & + \beta_{12} (QN20 * QN53)_{ij} + \beta_{13} (QN23 * QN53)_{ij} \\ & + \beta_{14} (QN35 * QN53)_{ij} + \beta_{15} (QN22 * QN23)_{ij} \end{aligned}$$

Datos influyentes

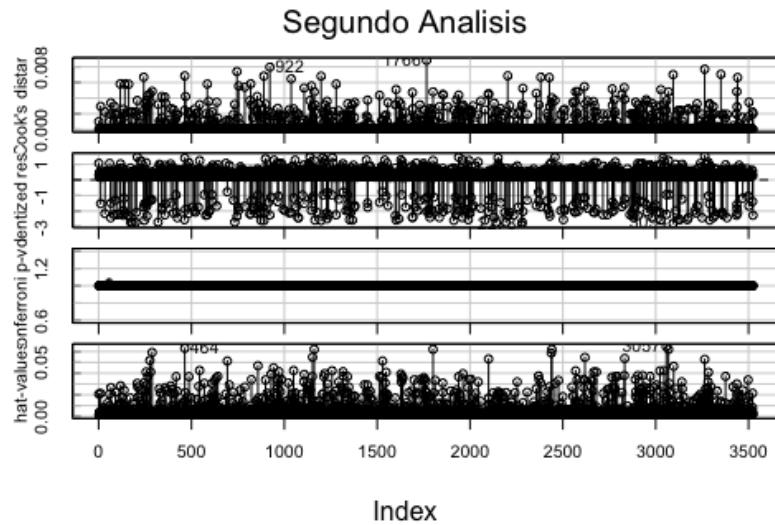
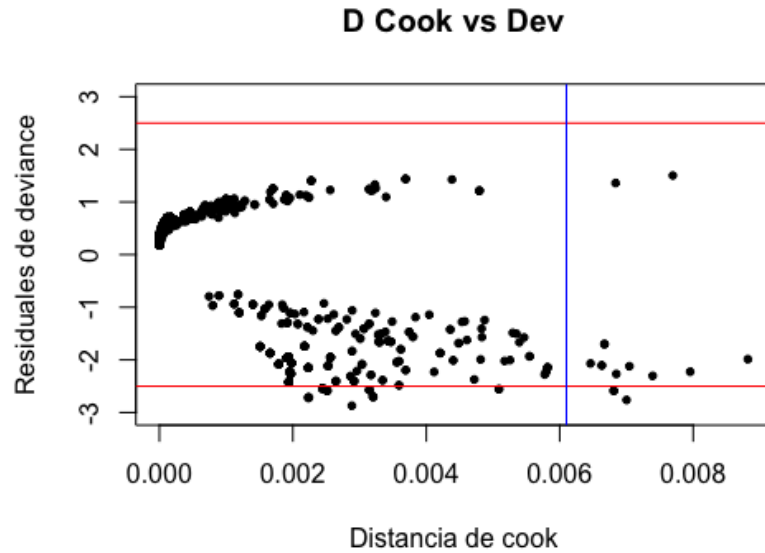
La idea ahora es lograr mejorar el AIC del modelo. Para esto se decidió realizar un análisis de datos influyentes a priori y a posteriori para las variables escogidas en la sección anterior.



En estas imágenes se empiezan a notar una clara influencia de algunas observaciones dentro del modelo. Para la figura 1 se analiza la distancia de Cook, se procede a omitir los datos 525,1179,1334,1519,2858,2812, 3088,1878,2816,464,632 y 1519.

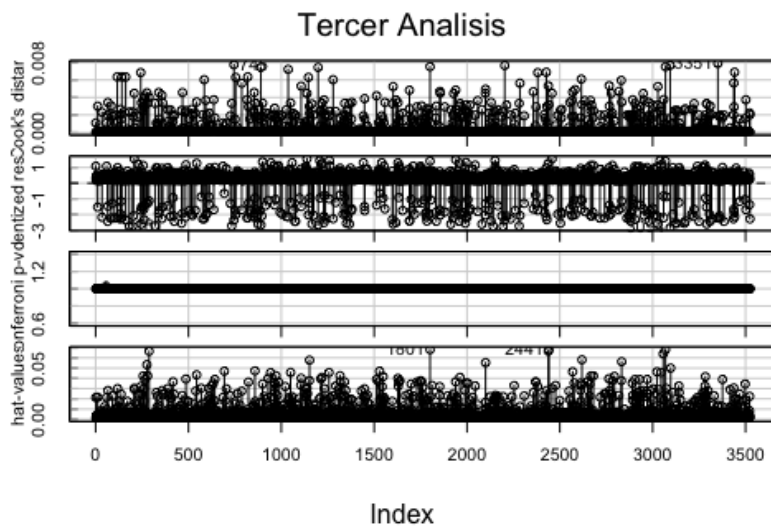
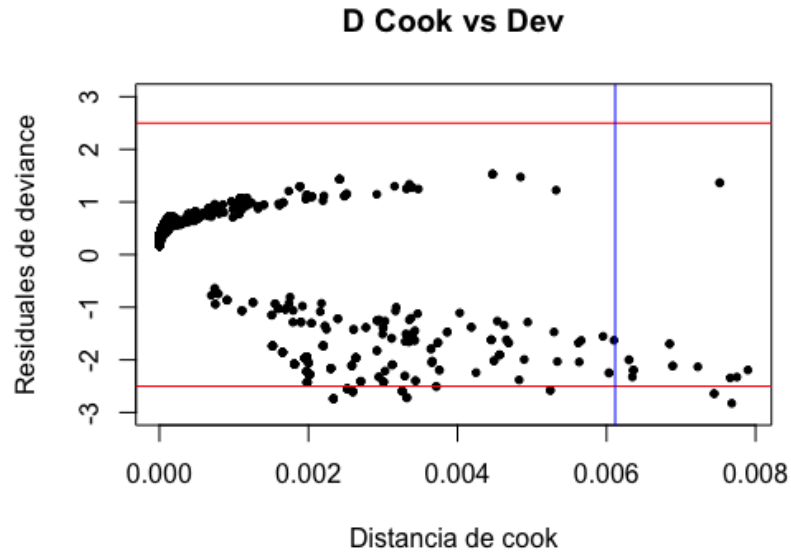
El resultado de esta omisión fue positivo y se logró reducir el AIC de 1513.3 a 1466.6.

Se realizó un proceso iterativo con un proceder análogo al anterior sin embargo se tuvo especial cuidado debido a que la base de datos de no cuenta con una gran cantidad de datos por lo que la omisión de cada observación debía ser, no solo cuidadosa, sino justificada por el valor de AIC que se obtenía.



Para un ultimo análisis de los datos influyentes se nota en las figuras, se omiten adicionalmente los datos 1766,2278,3264,172,464,922 y 1161 se refleja de nuevamente en la reducción del AIC (de 1466.6. a 1435.5).

Con esto se concluye el análisis de datos influyentes ya que, aunque sigan apareciendo más datos aparentemente influyentes, ya se obtuvo un AIC óptimo con la omisión de 19 datos.



Modelo logit

la regresión logística es un tipo de análisis de regresión utilizado para predecir el resultado de una variable categórica en función de las variables independientes o predictoras. Es útil para modelar la probabilidad de un evento ocurriendo como función de otros factores. El análisis de regresión logística se enmarca en el conjunto de Modelos Lineales Generalizados que usa como función de

enlace la función logit. Las probabilidades que describen el posible resultado de un único ensayo se modelan, como una función de variables explicativas, utilizando una función logística.

El logit del modelo de regresión logística se presenta por la siguiente ecuación:

$$z(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + e$$

Donde $\beta_0, \beta_1, \dots, \beta_p$ son parámetros del modelo de regresión logística múltiple y e es el error asociado al modelo. En este caso el modelo de regresión logística es:

$$\pi(x) = p_j = \frac{e^{z(x)}}{1 + e^{z(x)}}$$

Para valores de $z(x)$ muy grandes positivos, el valor de $e^{z(x)}$ es aproximadamente 0 por lo que el valor de la función es 1. Para valores de $z(x)$ muy grandes negativos, el valor $e^{z(x)}$ tiende a infinito por lo que el valor de la función es 0. El valor p_j puede interpretarse como: la probabilidad de que la variable categórica Y adquiera el valor k , dado que el predictor X tiene el valor x .

Una de las principales aplicaciones de un modelo de regresión logística es clasificar la variable cualitativa en función de valor que tome el predictor. Una herramienta para evaluar un modelo de regresión logística binaria es su capacidad para predecir con precisión una respuesta binaria.

Modelo probit

En el modelo Probit, Las probabilidades que describen el posible resultado de un único ensayo se modelan, como una función de variables explicativas, utilizando la función probabilidad de la distribución normal estándar.

El modelo probit presenta la siguiente ecuación:

$$g(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + e$$

donde $\beta_0, \beta_1, \dots, \beta_p$ son parámetros del modelo de regresión probit múltiple y e es el error asociado al modelo. En este caso el modelo de regresión probit es:

$$\pi(x) = p_j = \Phi(g(x)) = \int_{-\infty}^{g(x)} \varphi(t) dt$$

donde el símbolo Φ es simplemente la función de distribución acumulada estándar, mientras que el símbolo en minúscula, φ como antes, representa la función de densidad normal estándar.

Cauchy

Este enlace es aplicado a la probabilidad acumulada de cada categoría de la respuesta. Es apropiado para la distribución multinomial.

El modelo probit presenta la siguiente ecuación:

$$\tan = (\pi(z(x) - 0,5))$$

Elección del modelo

Mejor enlace

El enlace o link es un elemento fundamental del modelo a construir ya que de este depende no sólo del ajuste del modelo sino también del contexto . Es así que, se hizo un estudio de estos enlaces obteniendo los siguientes resultados.

Modelo	G.L.	AIC
Logit	16	1435.504
Probit	16	1433.245
Cauchy	16	1473.225

Obtuvimos que el de mejor AIC es el modelo Probit, sin embargo, para propósitos de simplicidad y contexto se hará uso del modelo Logit.

Hipótesis a juzgar

Se desea juzgar ahora la hipótesis global, que esta definida de la siguiente manera:

$$H_o : \beta_i = 0 \quad \text{para todos los } i = 1, 2, \dots, 15$$

$$H_a : \beta_i \neq 0 \quad \text{para algún } j = 1, 2, \dots, 15$$

Para contrastar esta hipótesis realizaremos un test χ^2 recordando que

$$D_0 - D_r \quad \text{Distribuye } \chi_k^2$$

Donde

$$D_0 = \text{devianza nula} \quad \text{y} \quad D_r = \text{devianza residual}$$

En nuestro modelo obtenemos los siguientes resultados

$$\begin{array}{c|c|c} \text{Devianza nula} & \text{Devianza residual} & P(>\chi_{15}^2) \\ \hline 1706.0 & -1403.5 & 0e-15 \end{array}$$

Por lo cual hay suficiente evidencia estadística para rechazar la hipótesis nula de que $\beta_i = 0$ para todos los $i = 1, 2, \dots, 15$ es decir, al menos una de las 8 variables explicativas o de las 7 interacciones del modelo es relevante para explicar nuestra variable respuesta. Esto motiva ahora a juzgar una hipótesis individual para cada variable explicativa y para las interacciones.

Primero revisaremos las interacciones, para esto debemos juzgar la siguiente hipótesis nula:

$$H_o : \beta_i = 0 \quad \text{para todos los } i = 9, 10, \dots, 15$$

$$H_a : \beta_i \neq 0 \quad \text{para algún } i = 9, 10, \dots, 15$$

Al realizar la prueba χ^2 obtenemos los siguientes resultados:

	estimado	σ	$P(>\chi_1^2)$
Q2:QN20	6.823	2622	0.0090
QN20:QN44	0.553	2621	0.4572
QN22:QN44	7.175	2620	0.0073
QN20:QN53	7.502	2619	0.0061
QN23:QN53	6.697	2618	0.0096
QN35:QN53	3.752	2617	0.0527
QN22:QN23	3.837	2616	0.0501

En esta tabla logramos evidenciar que hay 3 interacciones cuyo p-valor es mayor a $\alpha = 0,05$, en las dos ultimas este valor es solo un poco mayor al $\alpha = 0,05$, mientras que para la interaccion "QN20:QN44" su valores mucho mayor, por lo cual podemos concluir que no se rechaza la hipotesis nula para este β y esta interaccion no es significativa, luego de realizar diferentes estudios tomando en cuenta y sin tomar en cuenta esta interaccion nos damos cuenta que aunque no sea significativa nos resuelve ciertos inconvenientes como la heterocedasticidad entre grupos, por lo cual se mantendra en el analisis. Respecto a las otras interacciones notamos que hay suficiente evidencia estadistica para rechazar la hipótesis nula y concluir que son significativas.

Por ultimo lo realizamos para cada una de las variables, para ello se hizo uso del estadístico de wald para contrastar la hipótesis

$$H_o : \beta_j = \beta_0 \quad \text{para todos los } j = 1, 2, \dots, 8$$

$$H_a : \beta_j \neq \beta_0 \quad \text{para algún } j = 1, 2, \dots, 8$$

Recordando que

$$w_j = \frac{(\hat{\beta}_j - \beta_0)^2}{Var(\hat{\beta}_j)} \quad \text{Distribuye } \chi_1^2$$

Para nuestro caso el valor de interés es $\beta_0 = 0$ con lo que se obtienen los siguientes resultados:

	estimado	σ	$P(>\chi_1^2)$
Q2	27.571	2630	1.514e-07
QN20	42.803	2629	6.053e-11
QN22	64.077	2628	1.196e-15
QN23	29.670	2627	5.122e-08
QN35	40.139	2626	2.365e-10
QN44	41.853	2625	9.842e-11
QN53	6.601	2624	0.0101947
QN56	13.402	2623	0.0002514

En esta tabla se evidencia que cada una de las variables explicativas escogidas es significativa para un valor $\alpha = 0,05$. Es decir, hay suficiente evidencia estadística para rechazar la hipótesis nula de que $H_o : \beta_j = 0$ para $j = 1, 2, \dots, 8$.

Descripción modelo

Gracias al análisis anterior concluimos que el modelo mas adecuado es el Logit, por lo cual especificaremos los resultados obtenidos con este modelo, iniciando por la estimación de los parámetros

con los cuales llegamos al siguiente modelo:

$$\begin{aligned}\hat{Z}(X)_i = & 0,431 - 1,386(Q2Femenino) - 0,081(QN20No) - 0,171(QN22No) \\ & + 1,099(QN23No) + 0,194(QN35No) + 0,466(QN44No) + 0,328(QN53No) \\ & - 0,515(QN56No) + 0,926(Q2Femenino)(QN20No) - 0,654(QN20No)(QN44No) \\ & + 1,044(QN22No)(QN44No) + 0,948(QN20No)(QN53No) - 1,009(QN23No)(QN53No) \\ & + 0,630(QN35No)(QN53No) + 0,958(QN22No)(QN23No)\end{aligned}$$

Para su interpretación recordemos la definición de Odds ratio o también conocido como "riesgo relativo" para el enlace logit, por lo cual la transformación del puntaje $\hat{Z}(X)$ a través de la función logística sería de la siguiente manera:

$$ODDs = \frac{P(Y = 1)}{P(Y = 0)} = \frac{\hat{\pi}}{1 - \hat{\pi}} = e^{\hat{Z}(X)} \quad (2)$$

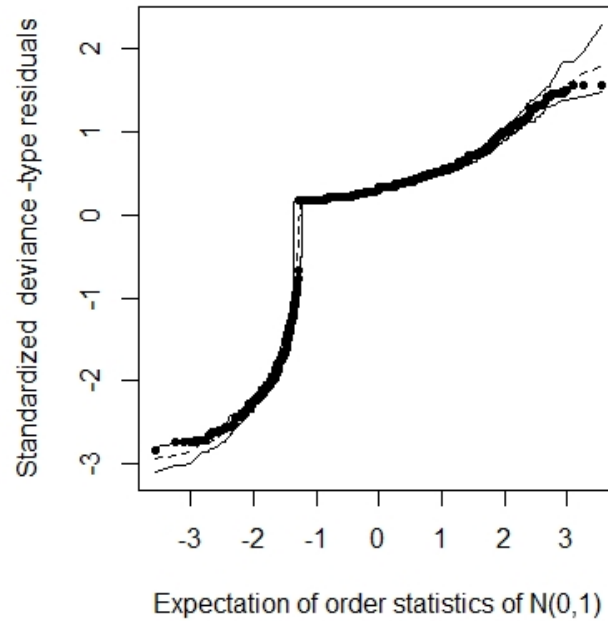
y obtenemos los siguientes resultados

Variable	Incremento
Q2Femenino	3.99
QN20No	1.084
QN22No	1.186
QN23No	3.00
QN35No	1.21
QN44No	1.59
QN53No	1.38
QN56No	1.67
Q2Femenino:QN20No	2.52
QN20No:QN44No	1.92
QN22No:QN44No	2.84
QN20No:QN53No	2.58
QN23No:QN53No	2.74
QN35No:QN53No	1.87
QN22No:QN23No	2.60

Validación de supuestos

Para poder examinar los supuestos de normalidad, homocedasticidad, multicolinealidad y autocorrelación, se hace uso del conjunto de funciones presentes en el script *Macros* del profesor Luis Vanegas disponible en *R*. En principio, se encuentran y omiten algunos datos atípicos e influyentes a posteriori en función de la distancia de *Cook* y de los residuales *Deviance*.

Normalidad asintótica



Se puede apreciar que los residuos de la muestra generados por este modelo y los valores esperados de una distribución normal guardan cierta similitud.

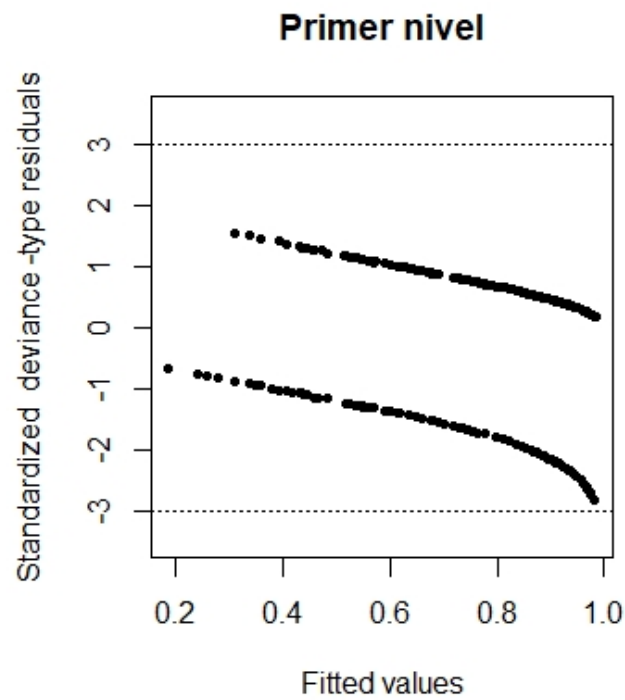
Multicolinealidad

	df	$\text{GVIF}^{1/(2 \cdot \text{df})}$
Q2	1.00	1.20
QN20	1.00	1.19
QN22	1.00	2.32
QN23	1.00	1.16
QN35	1.00	1.08
QN44	1.00	1.47
QN53	1.00	3.14
QN56	1.00	1.03
Q2:QN20	1.00	1.26
QN20:QN44	1.00	1.19
QN22:QN44	1.00	1.36
QN20:QN53	1.00	1.23
QN23:QN53	1.00	3.00
QN35:QN53	1.00	1.27
QN22:QN23	1.00	2.39

Se descartan problemas de multicolinealidad empleando el factor de inflación generalizado de la varianza *GVIF*, dado que ninguna de las variables presentes en el modelo, manifiesta un *GVIF* por encima de 5.

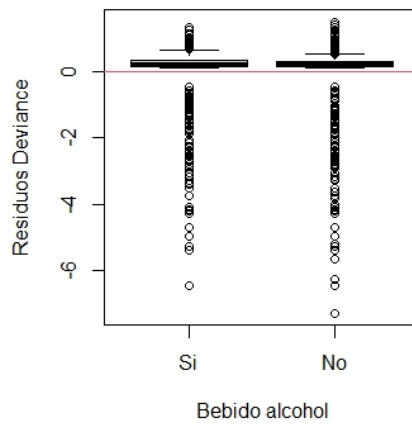
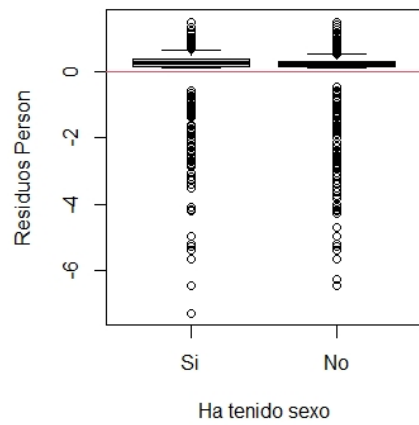
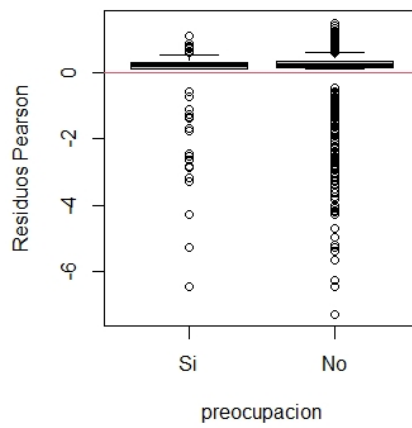
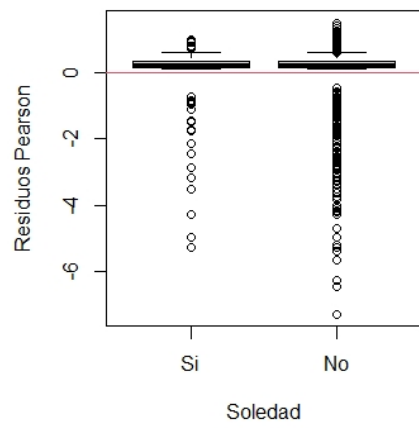
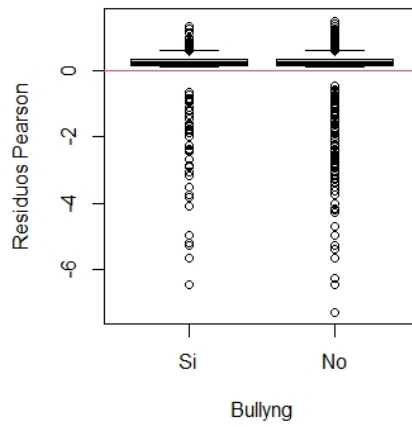
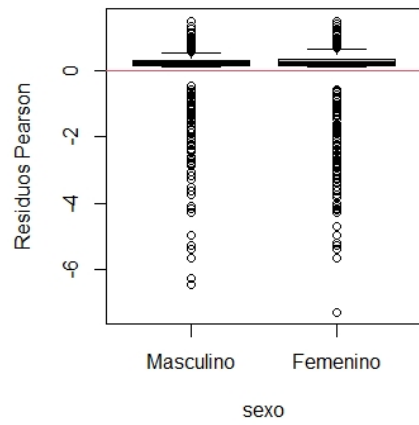
Residuales

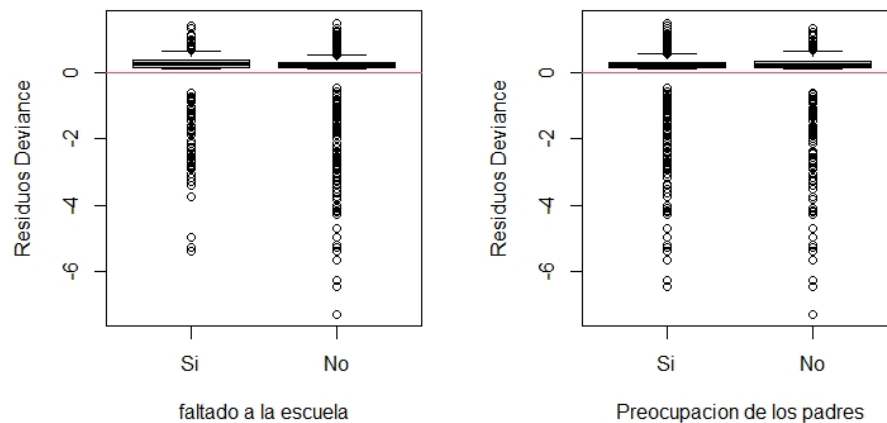
Se observa que los residuales tienen un comportamiento deseable ya que se encuentran en un intervalo dado entre -3 y 3, por lo cual podemos concluir que no hay problema respecto a los datos atípicos.



Correlación entre los residuales y las variables explicativas

Para revisar este supuesto analizaremos la matriz de correlaciones y los gráficos de los cruces de los residuales de Deviance con todas las variables respectivamente.





Notamos de las anteriores gráficas que no hay un inconveniente respecto a la correlación entre los residuos y las variables explicativas. Sin embargo, se observan datos atípicos en gran mayoría en valores bajos de los residuales.

Autocorrelación

Para este supuesto utilizaremos el estadístico de Durbin Watson donde se juzga la hipótesis $\rho = 0$ vs $\rho \neq 0$, donde obtenemos que no se rechaza la hipótesis nula con una significancia de $\alpha = 0,05$, por lo tanto, no hay autocorrelación de primer orden en el modelo.

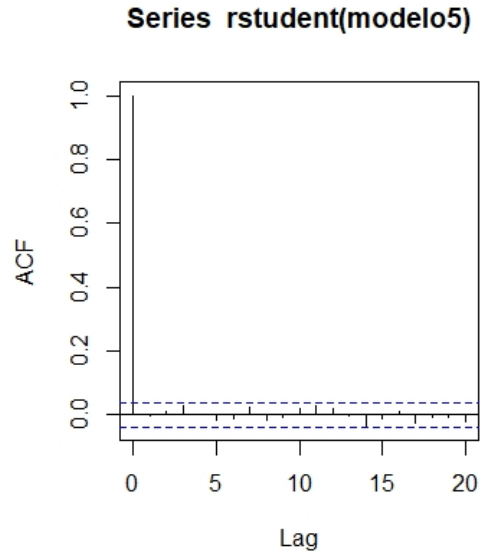
Durbin-Watson test

```
data: modelo5
```

```
DW = 1.9967, p-value = 0.4627
```

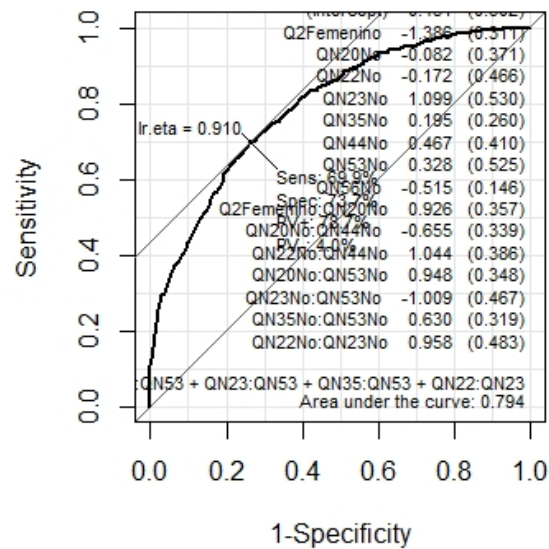
```
alternative hypothesis: true autocorrelation is greater than 0
```

Lo podemos comprobar con la siguiente grafica



Validación del modelo

Curva de ROC



Podemos observar que la grafica de ROC contiene la representación gráfica de la sensibilidad frente a la especificidad para el clasificador binario ,según se varía el umbral de discriminación.

Aquí tenemos un resultado de 0.795. Además, tenemos las medidas de bondad de ajuste del modelo.

Medidas de Bondad de Ajustes

Para contrastar la hipótesis nula:

$$H_0 : E(\mu|\mathbf{x}_i) = \frac{1}{1+e^{-x_i^T \beta}} \text{ vs.}$$

$$H_a : E(\mu|\mathbf{x}_i) \neq \frac{1}{1+e^{-x_i^T \beta}}$$

	df	Value
Pearson's Chi-squared	2616	2465.56051
Deviance	2616	1403.50383
Likelihood-based R-squared		0.17729
-2logLikelihood		1403.50383
AIC		1435.50383
BIC		1529.51182

Matriz de confusion

Es una herramienta que permite la visualización del desempeño de un algoritmo que se emplea en aprendizaje supervisado. Cada columna de la matriz representa el número de predicciones de cada clase, mientras que cada fila representa a las instancias en la clase real.

Total Observations in Table: 2649

	predict.fit		
QN25	1	0	Row Total
Si	38	236	274
	0.014	0.089	
No	25	2350	2375
	0.009	0.887	
Column Total	63	2586	2649

Notamos que el modelo predice bastante bien respecto a los que dicen que no, que son la mayoría de las personas a las que se realiza la encuesta, incluso notamos que hay cierto sesgo hacia esta opción, pero no es algo significativo por lo cual podemos concluir que el modelo realiza buenas predicciones.

Resultados

Recordamos el modelo y los Odds ratio que son los siguientes

$$\begin{aligned}\hat{Z}(X)_i = & 0,431 - 1,386(Q2Femenino) - 0,081(QN20No) - 0,171(QN22No) \\ & + 1,099(QN23No) + 0,194(QN35No) + 0,466(QN44No) + 0,328(QN53No) \\ & - 0,515(QN56No) + 0,926(Q2Femenino)(QN20No) - 0,654(QN20No)(QN44No) \\ & + 1,044(QN22No)(QN44No) + 0,948(QN20No)(QN53No) - 1,009(QN23No)(QN53No) \\ & + 0,630(QN35No)(QN53No) + 0,958(QN22No)(QN23No)\end{aligned}$$

variable	incremento
Q2Femenino	3.99
QN20No	1.084
QN22No	1.186
QN23No	3.00
QN35No	1.21
QN44No	1.59
QN53No	1.38
QN56No	1.67
Q2Femenino:QN20No	2.52
QN20No:QN44No	1.92
QN22No:QN44No	2.84
QN20No:QN53No	2.58
QN23No:QN53No	2.74
QN35No:QN53No	1.87
QN22No:QN23No	2.60

De la tabla anterior obtuvimos los siguientes resultados principales

- El hecho de ser mujer hace que sea 4 veces menos probable planear un suicidio durante los últimos 12 meses
- Los adolescentes que no han podido dormir bien debido a las preocupaciones en el ultimo año son 3 veces mas propensos a planear un suicidio
- El hecho de no recibir bullying reduce en un 8 % las probabilidades de planear un suicidio
- Quienes no reciben bullying y no han tenido relaciones sexuales son un 92 % menos propensos a planear un suicidio

Conclusiones

- A partir del entendimiento de cuáles variables realmente ayudan a predecir el intento de suicidio se podrán planear nuevos programas preventivos para así disminuir al tasa de suicidios en adolescentes.
- A futuro se espera que se puedan incluir interacciones en modelos más complejos para reducir el patrón observado en los residuales Pearson versus las variables explicativas.
- Como se ha observado a través de los años más recientes, el suicidio a una temprana edad ha ido aumentando y más aún la contemplación y la planeación del mismo. Esto debido a diversas variables seleccionadas para el modelo utilizado.
- Para futuros estudios se recomienda añadir variables que se relacionen más con las dinámicas del mundo moderno y la forma en que el contacto físico y tangible entre seres humanos está pasando a un segundo plano.

Referencias

- [1] Stanford Children's Health
Recuperado de <https://www.stanfordchildrens.org/es/topic/default?id=teen-suicide-90-P05694>
- [2] Global School-based Student Health Survey (GSHS)
Recuperado de <https://www.cdc.gov/gshs/countries/americas/uruguay.htm>
- [3] Revista Médica del Uruguay
Recuperado de <http://www.scielo.edu.uy/scielo.php?script=sciarttext&pid=S1688-03902004000200001>
- [4] Borja Manuel Fernández Félix. «Validación interna de modelos predictivos de regresión logística. Comando Validation(Stata)». En: (2018).