



# ANÁLISIS DE SUPERVIVENCIA | PROYECTO

Elaborado por Alondra Elizabeth Matos Mendoza

2021

# APLICACIÓN DEL ANÁLISIS DE SUPERVIVENCIA EN LA FUGA DE CLIENTES DE UNA TELCO

## Introducción

Para una compañía en general, la habilidad de retener clientes es muy importante porque conlleva a un ahorro económico, pues es mucho más barato que el costo de adquirir nuevos clientes. Además, un cliente satisfecho es más propenso a seguir gastando dinero para mantener los servicios y comprar regularmente los productos que ofrece la empresa. Sin embargo, ganar la lealtad del cliente no es sencillo y requiere de una buena estrategia de retención. Por lo tanto, es indispensable estudiar el comportamiento de aquellos que han dejado de ser clientes. De esta forma, se detectarían las principales razones de fuga y esto permitiría a la empresa aplicar las mejoras pertinentes para evitar el abandono de futuros clientes.

Por consiguiente, el objetivo de este estudio es determinar un modelo de Cox que describa el efecto de ciertas covariables sobre el tiempo de vida del cliente de un proveedor de servicios de telecomunicaciones, entendiéndose con tiempo de vida como el tiempo en meses de permanencia de un cliente hasta su fuga (cancelación del contrato con la compañía).

La base de datos utilizada para realizar el modelo está compuesta de 7043 registros, de los cuales el 73.46% presenta censura por la derecha. La siguiente tabla resume las variables explicativas que se registraron.

	Atributo	Tipo de dato	Descripción	Categorías
Información demográfica	Gender	Cualitativo	Género del cliente	Masculino,Femenino
	SeniorCitizen	Cualitativo	Si el cliente está jubilado	Sí, No
	Partner	Cualitativo	Si el cliente tiene pareja	Sí, No
	Dependents	Cualitativo	Si el cliente tiene dependientes	Sí, No
Servicios de suscripción disponibles	PhoneService	Cualitativo	Si el cliente cuenta con servicio telefónico	Sí, No
	MultipleLines	Cualitativo	Si el cliente cuenta con líneas múltiples	Sí, No, Sin servicio telefónico
	InternetService	Cualitativo	Si el cliente cuenta con servicio de internet	DSL, Fibra óptica, No
	OnlineSecurity	Cualitativo	Si el cliente cuenta con seguridad en línea	Sí, No, Sin servicio de internet
	OnlineBackup	Cualitativo	Si el cliente cuenta con respaldo en línea	Sí, No, Sin servicio de internet
	DeviceProtection	Cualitativo	Si el cliente cuenta con protección del dispositivo	Sí, No, Sin servicio de internet
	TechSupport	Cualitativo	Si el cliente cuenta con soporte técnico	Sí, No, Sin servicio de internet
	StreamingTV	Cualitativo	Si el cliente cuenta con transmisión de TV	Sí, No, Sin servicio de internet
	StreamingMovies	Cualitativo	Si el cliente cuenta con transmisión de películas	Sí, No, Sin servicio de internet
Información de la cuenta del cliente	Contract	Cualitativo	Término del contrato	Mes a mes, 1 año ,2 años
	PaperlessBlling	Cualitativo	Si el cliente tiene facturación electrónica	Sí, No
	PaymentMethod	Cualitativo	Método de pago del cliente	Cheque enviado por correo(Mailed check)
		Cualitativo		Cheque electrónico (Electronic check)
		Cualitativo		Transferencia bancaria (Bank transfer)
		Cualitativo		Tarjeta de crédito (Credit card)
	MonthlyCharges	Cuantitativo	Cantidad mensual que se cobra al cliente	Valor numérico
	TotalCharges	Cuantitativo	Monto total cobrado al cliente	Valor numérico

Es importante mencionar que se unificaron en una sola categoría, manteniendo la etiqueta “No”, las clases “No” y “Sin servicio de internet” de los atributos *OnlineSecurity*, *OnlineBackup*, *DeviceProtection*, *TechSupport*, *StreamingTV* y *StreamingMovies*. Asimismo, para el factor *MultipleLines*, se juntaron “No” y “Sin servicio telefónico” en solamente la categoría “No”, con el fin de representar de una manera más amplia la ausencia del servicio.

Cabe enfatizar que este tipo de estudios, en donde el evento de interés es la fuga del cliente, implica la necesidad de optar por un análisis de sobrevivencia a causa de la inevitable censura por la derecha; pues habrá información de clientes de los cuales se desconozca cuál exactamente es su tiempo de vida, debido a que podrían cancelar su contrato con la compañía posteriormente al tiempo de estudio.

## Análisis descriptivo

Debido a que una compañía telefónica se caracteriza principalmente por proporcionar servicios de telefonía e internet, se analizaron las variables *PhoneService* e *InternetService* mediante diversas gráficas que permitan comparar a los clientes que cancelaron el contrato contra los que no. Esto con el fin de detectar posibles signos de satisfacción sobre los servicios más primordiales.

**Distribución por servicio telefónico de los clientes que cancelaron el contrato**



**Distribución por servicio telefónico de los clientes que mantuvieron el contrato**



Por medio de las gráficas circulares que se presentan, se puede ver que, tanto para los clientes que mantienen el contrato como los que cancelaron, la proporción de quienes se encuentran suscritos al servicio telefónico y los que no, fue casi la misma. En consecuencia, las personas que dejaron de ser clientes no se clasifican de forma distinta bajo el atributo *PhoneService*.

En cambio, mediante la gráfica de barras que se encuentra a la derecha, se puede visualizar que los clientes que cancelaron el contrato se caracterizan por haber tenido internet de fibra óptica mientras que, para los que mantuvieron el contrato, predomina el servicio de internet DSL. Por tanto, los clientes más insatisfechos, según el servicio de internet suministrado, fueron los que recibieron internet de fibra óptica.

**Gráfica de barras del servicio de internet por tipo de cliente**

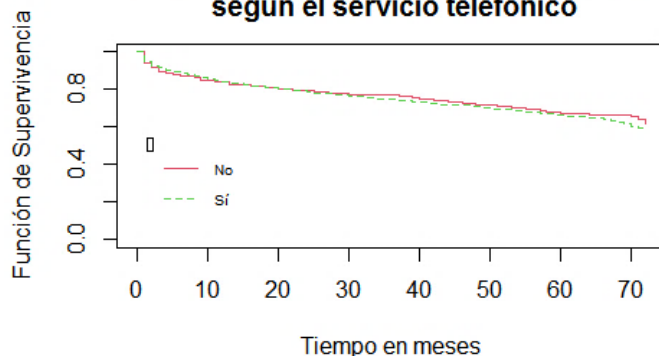


## Análisis estadístico sobre el tiempo de vida

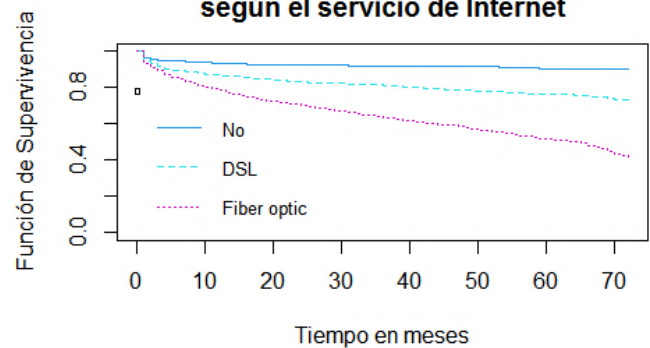
Antes de estudiar el efecto de las covariables en conjunto sobre el tiempo de vida de un cliente, resulta de interés analizar individualmente el impacto de las variables *PhoneService* e *InternetService*.

La gráfica de la función de supervivencia Kaplan Meier de *PhoneService* sugiere que las poblaciones que define este atributo tienen funciones de supervivencia con comportamientos similares entre sí. Por otro lado, de las poblaciones segregadas por *InternetService*, pareciera a primera vista, que los clientes con servicio de internet de fibra óptica tienen una mayor probabilidad de cancelar el contrato; mientras que, los clientes que no cuentan con el servicio de internet son los que tienen una menor probabilidad de fugarse. De hecho, la media del tiempo de vida de un cliente sin el servicio, con internet DSL y con internet de fibra óptica es de 66.5, 58.9 y 47 meses, respectivamente.

**Función de Supervivencia Kaplan Meier según el servicio telefónico**



**Función de Supervivencia Kaplan Meier según el servicio de Internet**



Para cada una de las dos variables mencionadas, se realizó una prueba log-rank para probar las siguientes hipótesis:

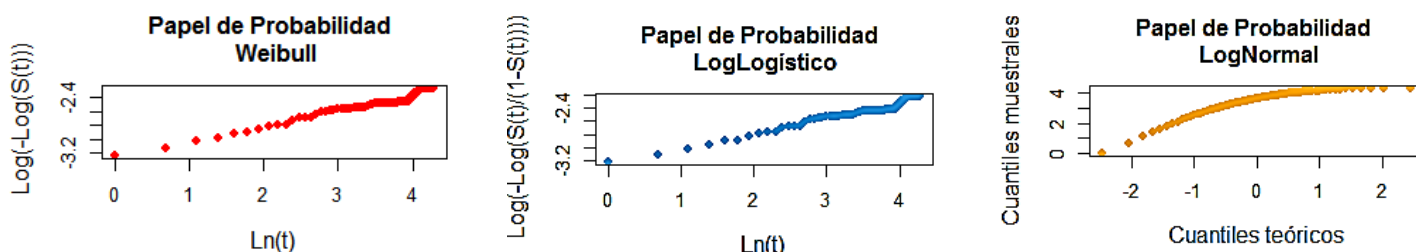
$$H_0: S_1(t) = \dots = S_k(t) \text{ para toda } t \quad VS \quad H_1: S_i(t) \neq S_j(t) \text{ para algún par } i, j \text{ y un tiempo } t_0$$

Donde  $S$  representa la función de supervivencia y  $K$  el número de poblaciones clasificadas según un atributo.

Respecto a *PhoneService*, se obtuvo un valor  $P$  de 0.5116 ( $>0.05$ ) Por tanto, a un nivel de confianza del 95%, no se puede rechazar la hipótesis nula, de modo que no existe evidencia suficiente que indique que el servicio telefónico afecte la probabilidad de un cliente de cancelar el contrato.

En cuanto a *InternetService*, el valor  $P$  obtenido fue menor que  $2e-16$  ( $<0.05$ ), lo cual permite rechazar la hipótesis nula, por lo que el servicio de internet afecta el comportamiento del tiempo de vida de un cliente. Ahora bien, para determinar las subpoblaciones con distintas funciones de supervivencia, se realizó la prueba log-rank mediante comparaciones múltiples y se obtuvo que todas provienen de distribuciones significativamente diferentes entre sí.

Para determinar un modelo estadístico que describa el comportamiento de la variable tiempo de vida para cada población generada según el tipo de servicio de internet, se realizaron papeles de probabilidad para las distribuciones Weibull, Log-Normal y Log-Logística. Las gráficas resultantes para la clientela sin servicio de internet se muestran a continuación.



Tanto en la gráfica del papel de probabilidad Weibull como en la Log-Logística, se aprecia una tendencia de línea recta; por tanto, existe la sospecha de que el tiempo de vida del cliente sin el servicio de internet siga alguna de estas distribuciones. Así pues, se ajustó un modelo para obtener los parámetros de las dos distribuciones mediante máxima verosimilitud y se obtuvo un AIC (Criterio de Información de Akaike) igual a 1456.012 para la Weibull; mientras que, para la Log-Logística, el AIC calculado fue de 1454.245. Entonces, se optó por el modelo con menor AIC asociado, que en este caso se obtuvo bajo la distribución Log-Logística. Análogamente, resultó que el mejor modelo para la población de clientes con servicio de internet DSL se consiguió bajo la distribución LogLogística; entre tanto, para la población de clientes con servicio de internet de fibra óptica, el mejor modelo que se tuvo fue bajo la distribución Weibull. Los parámetros estimados junto con los límites de los intervalos de confianza al 95% se muestran en la Tabla 1. Incluso, se puede notar que los intervalos de confianza entre ambas distribuciones Log-Logísticas no se intersectan.

Tabla 1

A un nivel del 95% de confianza, se infiere que la población de clientes sin servicio de internet sigue una distribución Log-Logística ( $\alpha = 0.44, \lambda = 7390$ ), la población de clientes con internet DSL una distribución Log-Logística ( $\alpha = 0.58, \lambda = 378.41$ ) y la población de clientes con internet de fibra óptica una distribución Weibull ( $\alpha = 0.72, \lambda = 100.1$ ).

Categoría	Distribución	Parámetro	Estimación	L95%	U95%	SE
Sin internet	Loglogística	Forma ( $\alpha$ )	0.44	0.38	0.52	3.80E-02
		Escala ( $\lambda$ )	7390.00	2630.00	20800.00	3.90E+03
DSL	Loglogística	Forma ( $\alpha$ )	0.5845	0.539	0.63	0.0241
		Escala ( $\lambda$ )	378.41	288.6536	496.08	52.274
Fibra Óptica	Weibull	Forma ( $\alpha$ )	0.72	0.6821	0.7511	0.0176
		Escala ( $\lambda$ )	100.10	91.6738	109.3082	4.4929

## Proceso de ajuste del modelo de Cox

A continuación, se describirá el procedimiento que se llevó a cabo en la selección de variables para ajustar un modelo de Cox, con el fin de estimar la función de riesgo.

### Paso 1:

Primeramente, se crearon variables tipo “dummy” para poder medir el efecto del factor cualitativo. En general, las variables binarias, correspondientes a los factores con categorías de “Sí” y “No”, toman el valor de 1 para “Sí” y 0 de otra forma. Por otro lado, para los factores con más de dos categorías, se indica a continuación los grupos base.

Factor	InternetService	Contract	PaymentMethod
Grupo base	Clientes sin servicio de internet	Clientes con contrato tipo mes a mes	Clientes cuyo método de pago es mediante un cheque por correo

Luego, para cada una de las variables de estudio, se consideró un modelo de riesgos proporcionales de Cox y su significancia se probó bajo las siguientes hipótesis:

$$H_0: \beta = 0 \quad VS \quad H_1: \beta \neq 0, \quad \beta \text{ es el coeficiente de la variable}$$

Siendo el estadístico de prueba  $-2 \ln(\hat{L})$ , donde  $\hat{L}$  es la verosimilitud máxima

A continuación, se presenta junto con el valor P obtenido, las variables consideradas en esta etapa. A un nivel del 95% de confianza, las variables que no resultaron significativas se exhiben con color rojo.

<b>Gender</b> ValorP=0.472>0.05	<b>SeniorCitizen</b> ValorP< 2e-16<0.05	<b>Partner</b> ValorP<2e-16<0.05	<b>Dependents</b> ValorP<2e-16<0.05	<b>PhoneService</b> ValorP= 0.51>0.05
<b>MultipleLines</b> ValorP= 0.0205<0.05	<b>InternetServiceFiberOptic</b> ValorP <2e-16<0.05	<b>InternetServiceDSL</b> ValorP<2e-16<0.05	<b>OnlineSecurity</b> ValorP= 2e-16<0.05	<b>OnlineBackup</b> ValorP <2e16<0.05
<b>Contract1 (1 Año)</b> ValorP<2e-16<0.05	<b>Contract2 (2 Años)</b> ValorP<2e-16<0.05	<b>PaperlessBilling</b> ValorP<2e-16<0.05	<b>MonthlyCharges</b> ValorP=3.7e-15<0.05	<b>TotalCharges</b> ValorP<2e-16<0.05
<b>PaymentMethod1</b> Bank Transfer ValorP <2e-16<0.05	<b>PaymentMethod2</b> Credit Card ValorP <2e-16<0.05	<b>PaymentMethod3</b> Electronic Check ValorP <2e-16<0.05	<b>InternetService1</b> Fiber Optic ValorP <2e-16<0.05	<b>InternetService2</b> DSL ValorP<2e-16<0.05
<b>DeviceProtection</b> ValorP<2e-16<0.05	<b>TechSupport</b> ValorP<2e-16<0.05	<b>StreamingTV</b> ValorP= 0.0451<0.05	<b>StreamingMovies</b> ValorP=0.022<0.05	

Paso 2:

Se ajustó un modelo con todas las variables significativas y se obtuvo que particularmente, a un nivel del 95% de confianza, no se puede rechazar la hipótesis que indica que sea igual a 0 el coeficiente de las siguientes variables:

<b>SeniorCitizen</b> ValorP= 0.57>0.05	<b>Dependents</b> ValorP= 0.17>0.05	<b>MultipleLines</b> ValorP= 0.13>0.05	<b>DeviceProtection</b> ValorP= 0.1>0.05
<b>StreamingTV</b> ValorP= 0.16>0.05	<b>StreamingMovies</b> ValorP= 0.19>0.05	<b>PaymentMethod3 (Electronic Check)</b> ValorP= 0.06>0.05	

Por tanto, no hay evidencia de que alguna de las variables mostradas contribuya significativamente en el modelo, dado que los otros atributos ya están considerados en él. Por ello, estas variables fueron descartadas en la estimación de la función de riesgo.

Paso 3:

Para cada una de las variables que resultaron no significativas en el paso 1, se ajustó un modelo junto con las variables que no fueron desestimadas en el paso 2. Respecto a *Gender*, se obtuvo que su valor P es 0.37(>0.05). Consecuentemente, no se puede rechazar la hipótesis que señala que el coeficiente de esta variable es igual a 0, es decir, no existe evidencia suficiente que señale que *Gender* contribuye significativamente en el modelo, dada la presencia de las otras variables consideradas en él. De esta forma se justifica que el atributo *Gender* no sea tomado en cuenta como parte del modelo. Por el contrario, en cuanto a *PhoneService*, se rechaza la hipótesis nula de que el coeficiente de la variable sea 0 ya que su valor P es 0.01(>0.05). Es por esto que, no hay evidencia suficiente para la eliminación de esta variable en el modelo.

Posteriormente, se procedió a probar la incorporación de interacciones o términos mixtos al modelo. En primer lugar, se obtuvo que la interacción entre *OnlineSecurity* y *TechSupport* es significativa a un nivel de confianza del 95% pues su valor P es 0.03(>0.05); por tanto, no hay evidencia suficiente para quitarla del modelo. Asimismo, resultaron significativos los términos mixtos para cada categoría de *Contract* junto a la variable *MonthlyCharges*, porque el valor P fue menor que 0.05 en ambos casos.



## Modelo de Cox estimado

En la Tabla 2, se presentan las covariables incluidas en el modelo definitivo junto la estimación del coeficiente y el valor P asociado. La tercera columna de la tabla representa la cantidad en la que aumenta o disminuye el riesgo en un tiempo  $t$ , según la característica del cliente.

El modelo de Cox estimado se expresa de la siguiente manera:

$$h(t, X) = h_0 \exp (0.457\textit{PhoneService} - 0.196\textit{Partner} + 0.823\textit{InternetService1} \\ + 1.209\textit{InternetService2} - 0.453\textit{OnlineSecurity} - 0.210\textit{OnlineBackup} \\ - 0.347\textit{TechSupport} - 6.266\textit{Contract1} - 10.12\textit{Contract2} + 0.153\textit{PaperlessBilling} \\ - 0.385\textit{PaymentMethod1} - 0.375\textit{PaymentMethod2} + 0.037\textit{MonthlyCharges} \\ - 0.002\textit{TotalCharges} + 0.317\textit{OnlineSecurity} * \textit{TechSupport} + 0.07\textit{Contract} \\ * \textit{MonthlyCharges} + 0.101 \textit{Contract2} * \textit{MonthlyCharges})$$

Donde  $X$  es el vector con las covariables de interés y  $h_0$  es la función de riesgo de los individuos con valor 0 en todas las variables tipo “dummy” y con el mínimo valor en las variables cuantitativas.

En particular, el grupo base se caracteriza por ser clientes sin pareja, con contrato tipo “mes a mes” sin servicio telefónico ni de internet, sin seguridad ni soporte técnico, y sin respaldo en línea. No cuentan con facturación electrónica. El monto mensual cobrado es de \$ 18.25 y monto total de \$18.8. Efectúan el pago mediante cheque por correo.

## Interpretación del modelo de Cox (bajo el supuesto de validez)

Tabla 2

El modelo de Cox especifica el cambio, mediante los coeficientes asociados a las variables, de la función de riesgo base respecto de aquellos con variables predictoras diferentes de cero. De la Tabla 2, se puede observar que un contrato de 2 años es lo que más disminuye el riesgo de fuga del cliente. De hecho, este tipo de contrato aumenta el riesgo en 0.00004 veces que un contrato “mes a mes”. Contrariamente, lo que incrementa más el riesgo de fuga, es la contratación del servicio de internet de fibra óptica; esto sugiere que la empresa debe aplicar mejoras sobre tal servicio en particular; pues los clientes que cuentan con internet de fibra óptica tienen un riesgo 3.35 veces mayor que uno sin internet, y un riesgo con diferencia de 1.07 respecto a un cliente con internet DSL.

Atributo	Coeficiente	Exp(coef)	Valor P
PhoneService	0.457	1.57900	8.60E-05
Partner	-0.196	0.82210	0.000123
InternetService1 (DSL)	0.823	2.27800	5.02E-08
InternetService2 (Fiber optic)	1.209	3.35100	2.85E-09
OnlineSecurity	-0.453	0.63570	1.55E-08
OnlineBackup	-0.210	0.81080	0.00023
TechSupport	-0.347	0.70710	1.17E-05
Contract1 (1 año)	-6.266	0.00190	2.00E-16
Contract2 (2 años)	-10.120	0.00004	2.00E-16
PaperlessBilling	0.153	1.16500	0.006309
PaymentMethod1 (BankTransfer)	-0.385	0.68050	4.80E-08
PaymentMethod2 (Credit Card)	-0.375	0.68770	3.23E-07
MonthlyCharges	0.037	1.03800	2.00E-16
TotalCharges	-0.002	0.99800	2.00E-16
OnlineSecurity:TechSupport	0.317	1.37300	0.030421
Contract1:MonthlyCharges	0.070	1.07300	2.00E-16
Contract2:MonthlyCharges	0.101	1.10700	2.00E-16

Por otro lado, existe un efecto entre el servicio de seguridad en línea y el de soporte técnico. El hecho de que un cliente tenga contratado ambos servicios aumenta el riesgo de fuga en 1.37 veces más, en comparación a un cliente con solo alguno de estos servicios.

Con respecto a *MonthlyCharges* y *TotalCharges*, dichas variables continuas aumentan el riesgo de fuga a medida que aumentan su valor. Además, para cada valor del monto mensual, el efecto del riesgo de fuga es diferente dependiendo de la longitud del contrato. El riesgo sufre un incremento mayor, debido al monto mensual, cuando el cliente tiene un contrato de 2 años.

## Comparación de subpoblaciones

En este caso, el modelo de Cox está en función de un vector de la forma:

$$X = (\textit{PhoneService}, \textit{Partner}, \textit{InternetService1}, \textit{InternetService2}, \textit{OnlineSecurity}, \textit{OnlineBackup}, \\ \textit{TechSupport}, \textit{Contract1}, \textit{Contract2}, \textit{PaperlessBilling}, \textit{PaymentMethod1}, \textit{PaymentMethod2}, \\ \textit{MonthlyCharges}, \textit{TotalCharges}, \textit{OnlineSecurity} * \textit{TechSupport}, \textit{Contract1} \\ * \textit{MonthlyCharges}, \textit{Contract2} * \textit{MonthlyCharges})$$

Suponiendo que se tienen 2 grupos de clientes, cuyas características se describen a continuación:

Grupo 1	Grupo 2
Clientes sin pareja, con contrato tipo “mes a mes” que cuentan con servicio telefónico y de internet DSL, con seguridad y soporte técnico, pero sin respaldo en línea. No cuentan con facturación electrónica. El monto mensual cobrado es de \$72 y el monto total de \$564. Efectúan el pago mediante cheque por correo.	Clientes sin pareja, con contrato tipo “mes a mes” que cuentan con servicio telefónico y de internet fibra óptica, sin seguridad ni soporte técnico, y sin respaldo en línea. Además, cuentan con facturación electrónica. El monto mensual cobrado es de \$72 y el monto total de \$80. Efectúan el pago mediante tarjeta de crédito.
$X_1 = (1, 0, 1, 0, 1, 0, 1, 0, 0, 0, 0, 0, 53.75, 545.75, 1, 0, 0)$	$X_2 = (1, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 1, 53.75, 61.75, 0, 0, 0)$

El factor de proporcionalidad entre los riesgos de ambos grupos es:  $\frac{h(t, X_1)}{h(t, X_2)} = \frac{5.640}{27.827} = 0.203 < 1$

Por ende, el grupo 1 tiene un riesgo inferior al grupo 2. De hecho, el riesgo en un tiempo  $t$  para el grupo 1 es 0.203 veces el riesgo del grupo 2. Esto implica que la probabilidad de que una persona del grupo 1 deje de ser un cliente luego de un tiempo  $t$ , es mayor que la probabilidad de cancelar el contrato por parte de un sujeto del grupo 2 después de un tiempo  $t$ .

## Validación del modelo

Se determinará si el modelo de Cox ajustado representa adecuadamente los datos mediante la verificación del supuesto de riesgos proporcionales, el análisis sobre la presencia de datos influyentes y la relación lineal entre el logaritmo de la función de riesgo y las covariables.

### Supuesto de Riesgos Proporcionales.

Para verificar este supuesto, se efectuó una prueba de hipótesis basado en los residuales de Schoenfeld, donde la hipótesis nula corresponde al cumplimiento del supuesto de riesgos proporcionales.

#### Hipótesis de la prueba de riesgos proporcionales:

$H_0$ : Existen riesgos proporcionales VS  $H_1$ : No existen riesgos proporcionales

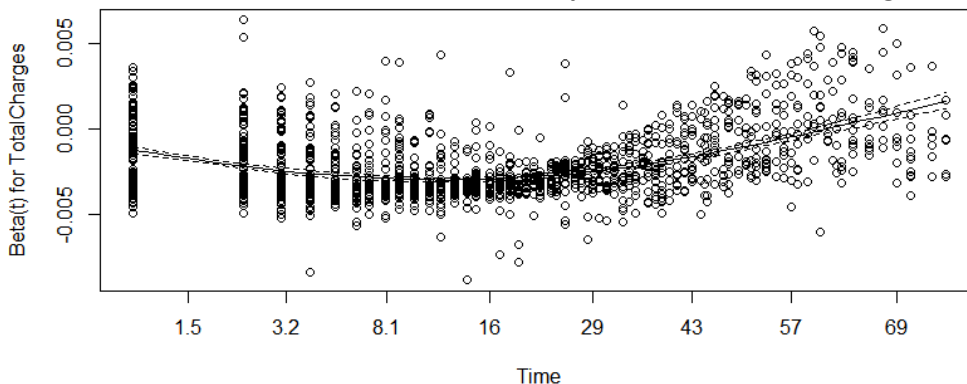
Los resultados se muestran en la Tabla 3 e indican que se viola este supuesto a un nivel de confianza del 95%, pues se rechaza la hipótesis nula para casi todas las covariables y como consecuencia, también globalmente en el modelo. Por tanto, el modelo de riesgos proporcionales de Cox ajustado no es válido.

Tabla 3

Atributo	Valor P
PhoneService	2.00E-16
Partner	0.03
InternetService	2.00E-16
OnlineSecurity	0.032
OnlineBackup	2.20E-13
TechSupport	1.70E-05
Contract	1.10E-08
PaperlessBilling	7.10E-12
PaymentMethodBank	0.209
PaymentMethodCredit	0.24
MonthlyCharges	2.00E-16
TotalCharges	2.00E-16
OnlineSecurity:TechSupport	0.072
Contract:MonthlyCharges	2.00E-16
GLOBAL	2.00E-16

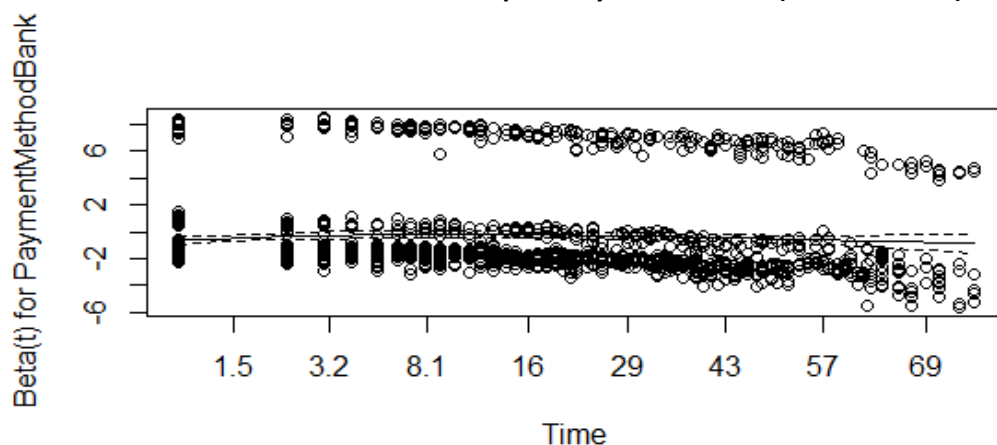
Otra forma de verificar si se cumple el criterio de riesgos proporcionales es mediante la gráfica de los residuales de Shoenfeld. Como ejemplo, se presentan a continuación las respectivas gráficas de residuales para las variables *TotalCharges* y *PaymentMehod (Bank Transfer)*, las cuales son gráficas de dispersión de los residuos asociados a cada variable y el tiempo.

Gráfica de residuales de Shoenfeld para la variable TotalCharges



En esta gráfica, se percibe que la curva estimada no tiene un comportamiento lineal. Por tanto, para *TotalCharges*, se presenta una violación del supuesto de riesgo proporcional.

**Gráfica de residuales de Shoenfeld para PaymentMethod (Bank Transfer)**

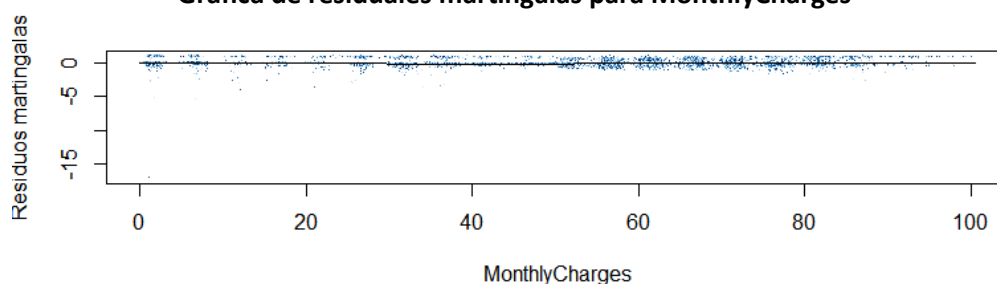


Mediante esta gráfica se puede observar una curva estimada con tendencia clara de línea recta horizontal. Esto indica que se cumple el supuesto del riesgo proporcional para *PaymentMehod* (Bank Transfer).

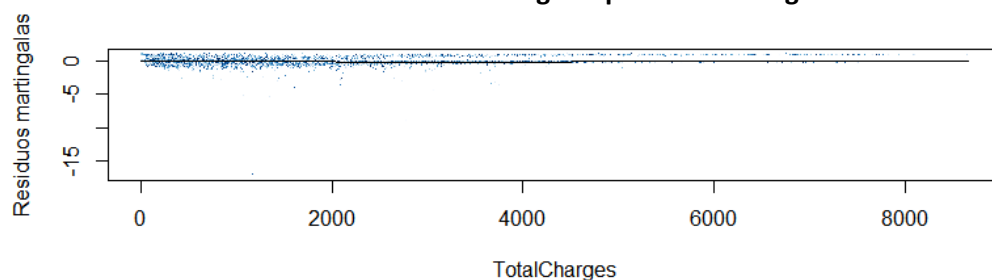
## Linealidad

Para cada una de las covariables continuas, se realizó el gráfico de los residuos de martingala versus el valor correspondiente de la covariable, con el fin de verificar si debe incluirse de forma lineal en el modelo. Para ambas covariables *MonthlyCharges* y *TotalCharges*, la nube de puntos plasmada muestra una tendencia de línea recta, por lo que existe una relación lineal entre el logaritmo de la función y tales covariables. De esta manera, se establece que no hay necesidad de incorporar alguna transformación en dichas variables predictoras.

**Gráfica de residuales martingalas para MonthlyCharges**



**Gráfica de residuales martingalas para TotalCharges**



## Presencia de datos influyentes

Para determinar si alguna observación en particular tiene un efecto potencial sobre las inferencias basadas en el modelo, se realiza una index plot, que resulta ser una gráfica de dispersión de los residuos contra el número de observación. Como ejemplificación, se muestra la gráfica para el caso de *TotalCharges*, en la cual se puede observar unos cuantos datos alejados del 0 y uno particularmente extremo. Para evitar la influencia de estos datos, se recomienda tratarlos.

