

# Modelo de propensión de fuga para seguros de automóvil voluntarios

Evelyn Johanna Romero

Introducción a la ciencia de datos  
Maestría en Actuaría y Finanzas  
Universidad Nacional de Colombia

Septiembre 2021



# Contenido

- 1 Introducción
- 2 Metodología y datos
- 3 Resultados
- 4 Conclusiones
- 5 Bibliografía



- 1 La retención de clientes es trascendental para mantener e incluso optimizar la posición de una compañía en el mercado. Si bien para el crecimiento de un negocio es importante la captación de nuevos usuarios, lo es también el conservar los existentes
- 2 El 90.3 % de los hogares colombianos ha tenido algún tipo de esquema de protección o aseguramiento, sin embargo solo el 30.3 % adquieren un seguro voluntario [Óscar Vargas, 2018]
- 3 El 70 % de los vehículos que circula a nivel nacional no está asegurado frente a eventualidades como el robo [Asopartes, 2019]
- 4 Se requiere construcción de un modelo de propensión de fuga para la detección de deserción de clientes en una compañía que comercializa seguros de automóviles



## Estimación de Kaplan - Meier [Kaplan and Meier, 1958]

$$\hat{S}(t) = \prod_{j|t_j \leq t} \frac{n_j - d_j}{n_j} \quad (1)$$

Donde  $n_j$  representan los clientes en riesgo para el tiempo  $t_j$  y  $d_j$  es el número de clientes retirados al tiempo  $t_j$ .

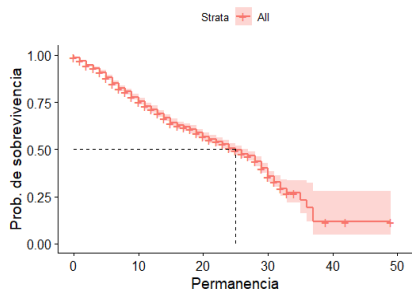
## Modelo de riesgos proporcionales de Cox (*Cox Proportional Hazard Model*) [Cox, 1972]

$$h(t) = h_0(t) \times \exp(b_1x_1 + b_2x_2 + \dots + b_px_p) \quad (2)$$

- $h(t)$  es la función de riesgo determinada por  $(x_1, x_2, \dots, x_p)$
- El término  $h_0$  se denomina *baseline hazard*. Corresponde al valor de riesgo o hazard si todos los  $x_i$  son iguales a cero ( $\exp(0)$  es igual a 1)

- ① Edad
  - ① Inferior a 40 años: Crecimiento
  - ② Entre 40 y 60 años: Apasionados
  - ③ Superior a 60 años: Expertos
- ② Género
- ③ Estado civil
- ④ Valor asegurado, marca y prima mensual → Gama de vehículo (Alta, media y baja)
- ⑤ Número de convenios
- ⑥ Modelo
  - ① Antiguos
  - ② Intermedios
  - ③ Modernos
- ⑦ Antigüedad y Salida
- ⑧ Salario (Medio, Medio alto y Alto)

# Resultados



time	n.risk	n.event	surv
0.00	3207.00	43.00	0.99
1.00	3127.00	56.00	0.97
2.00	2996.00	74.00	0.94
3.00	2747.00	42.00	0.93
6.00	2361.00	85.00	0.85
10.00	1762.00	56.00	0.75
12.00	1518.00	34.00	0.71



# Resultados

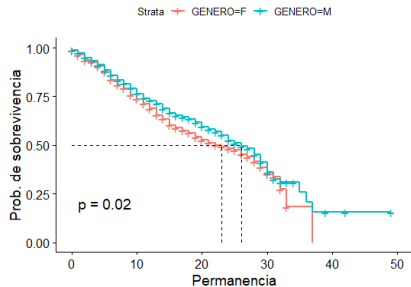


Figura: Curva de supervivencia de acuerdo al género

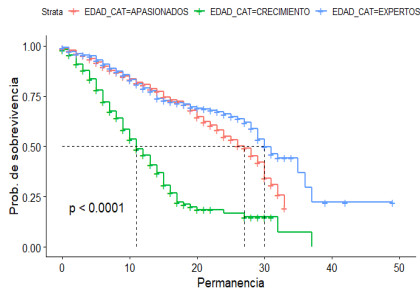


Figura: Curva de supervivencia de acuerdo a los grupos de edad



# Resultados

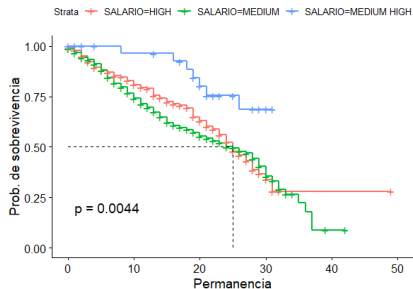


Figura: Curva de supervivencia de acuerdo al rango salarial

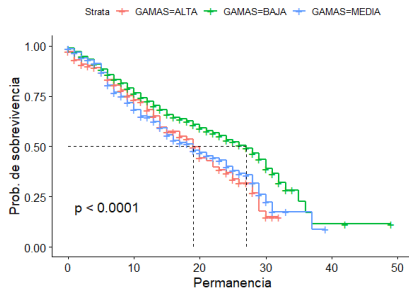


Figura: Curva de supervivencia de acuerdo a la gama del vehículo





# Análisis con dos factores

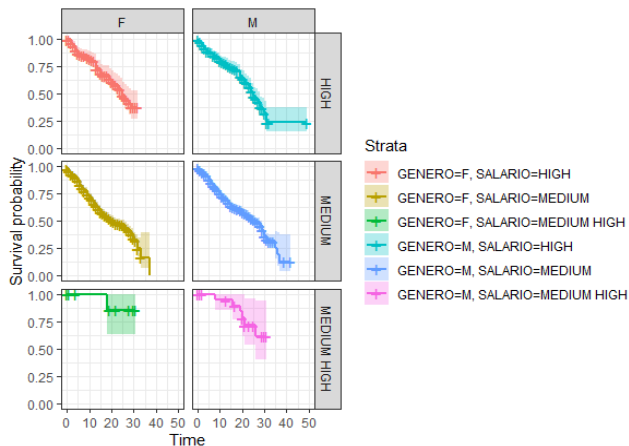


Figura: Curva de supervivencia de acuerdo al género y rango salarial



# Análisis con dos factores

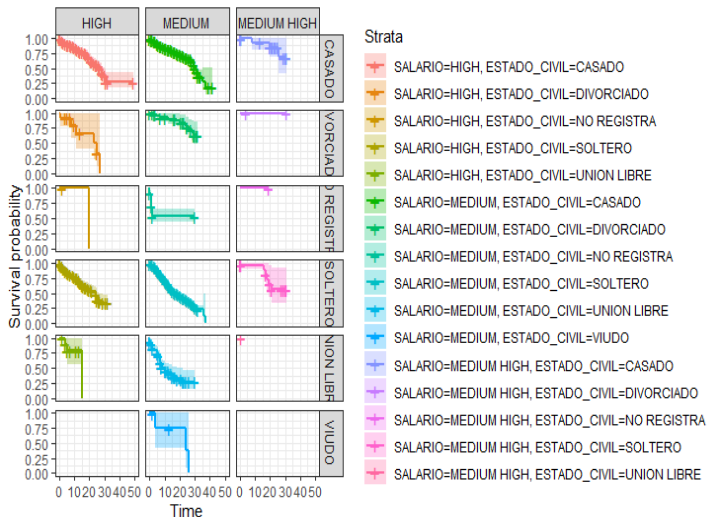


Figura: Curva de supervivencia de acuerdo al salario y estado civil



# Modelo de riesgos proporcionales de Cox

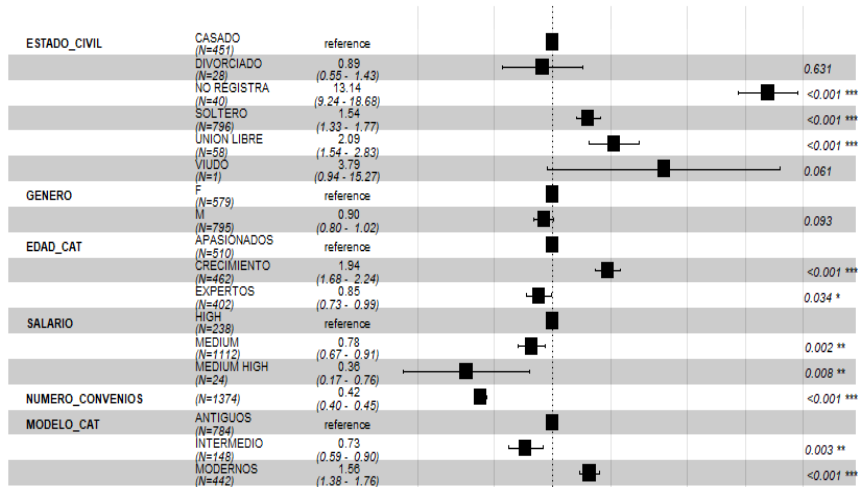
term	estimate	std.error	statistic	p.value
ESTADO_CIVILDIVORCIADO	0.89	0.24	-0.48	0.63
ESTADO_CIVILNO REGISTRA	13.14	0.18	14.34	0.00
ESTADO_CIVILSOLTERO	1.54	0.07	5.99	0.00
ESTADO_CIVILUNION LIBRE	2.09	0.15	4.77	0.00
ESTADO_CIVILVIUDO	3.79	0.71	1.87	0.06
GENEROM	0.90	0.06	-1.68	0.09
EDAD_CATCRECIMIENTO	1.94	0.07	8.92	0.00
EDAD_CATEXPERTOS	0.85	0.08	-2.12	0.03
SALARIOMEDIUM	0.78	0.08	-3.10	0.00
SALARIOMEDIUM HIGH	0.36	0.39	-2.67	0.01
NUMERO_CONVENIOS	0.42	0.03	-24.76	0.00
MODELO_CATINTERMEDIO	0.73	0.11	-2.99	0.00
MODELO_CATMODERNOS	1.56	0.06	7.13	0.00

Cuadro: Estimación de parámetros modelo 3



# Modelo de riesgos proporcionales de Cox

Hazard ratio

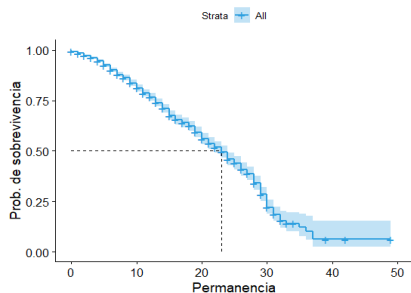


# Events: 1226; Global p-value (Log-Rank): 6.201e-250

AIC: 16637.77; Concordance Index: 0.78



# Modelo de riesgos proporcionales de Cox



**Figura:** Curva de supervivencia a nivel general para datos de test

time	n.risk	n.event	surv
0.00	3207.00	43.00	0.99
1.00	3127.00	56.00	0.99
2.00	2996.00	74.00	0.97
3.00	2747.00	42.00	0.96
10.00	1762.00	56.00	0.81
15.00	1215.00	47.00	0.68
35.00	7.00	1.00	0.12

**Cuadro:** Estimación de probabilidades a nivel general para los datos de test



# Conclusiones

- Para este caso en particular variables de la población activa y retirada como la edad, género, salario, estado civil, el modelo del vehículo asegurado e incluso el número de productos de seguro que tienen los usuarios impactan en el retiro que se pueda presentar en una póliza de automóvil voluntaria, independientemente si son factores de riesgo o de reducción contra la deserción.
- Se logró identificar que las personas solteras o en unión libre pueden incrementar el riesgo de retiro, también los usuarios con edad en el grupo crecimiento (inferior a los 40 años) y autos modernos, caso contrario de clientes de género masculino, con más de 60 años, afiliados a otros productos o pólizas y con salarios medios y medio altos.
- A futuro se espera adaptar estos resultados a un entorno más comercial en un dashboard dinámico, recibir feedback y realizar cambios para robustecer el modelo, además de extrapolar a otros ramos.





Asopartes (2019).

El 70 % de los autos en colombia no está asegurado.



Cox, D. R. (1972).

Regression Models and Life-Tables.

*Journal of the Royal Statistical Society. Series B (Methodological)*,  
34(2):187–220.



Kaplan, E. and Meier, P. (1958).

Nonparametric estimation from incomplete observations.

*Journal of the American Statistical Association*, 53(282):457–481.



Óscar Vargas (2018).

Seguros voluntarios: la gran oportunidad para aumentar la protección de las familias.

*Revista Fasecolda*, pages 80–85.

