

## Pricing

El pricing es el área dentro de la actuaria mediante la cual se establecen los precios de un producto o servicio. No es más que cuantificar **el mejor precio posible** a este tipo de elementos.

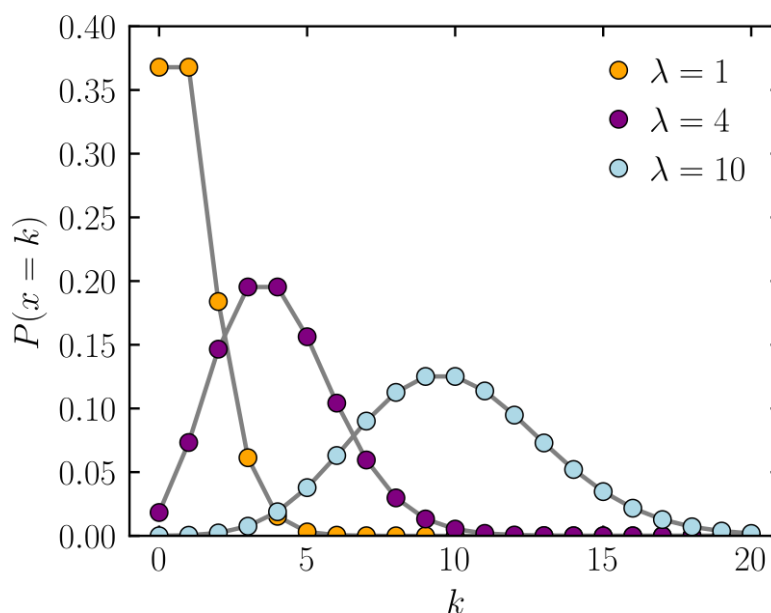
En sectores como el de manufactura, industria, con productos o servicios tangibles típicamente el precio se establece teniendo en cuenta dos aspectos.

1. Precio de materiales + precio de transformación + Rentabilidad exigida al producto.
2. ¿Es el precio competitivo con el mercado?, análisis de competidores, distribución del mercado, etc.

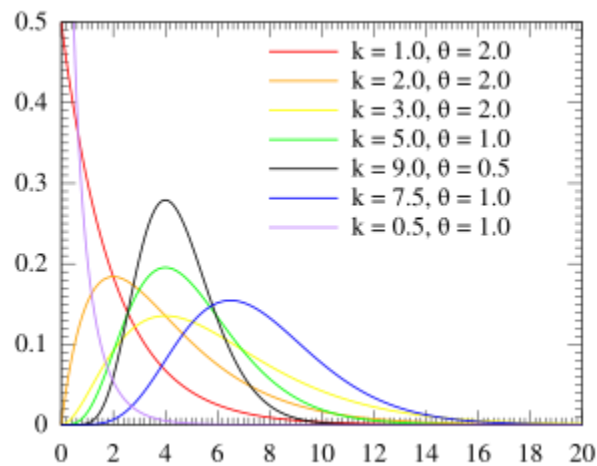
Sin embargo, en el sector seguros, si bien se quiere seguir esta misma metodología tenemos una situación compleja: **El precio se basa en un producto con condiciones claras, pero que no se materializa siempre, hay incertidumbre en el costo del producto.** ¿Sabes cuantos siniestros te va a dar un asegurado? ¿Sabes que tan graves serán estos siniestros? ->NO, y este es el reto, el objetivo es cuantificar el mejor precio posible, basándose en data histórica para estimar el numero de siniestros y el costo de cada uno de ellos.

Hay dos variables principales cuando cuantificamos el precio de un producto asegurador, la frecuencia y la severidad, se son respectivamente la cantidad de siniestros que dará un asegurado y dado que ha habido un siniestro entonces que costo tendrá.

La frecuencia se modela típicamente con una distribución de Poisson.  
**(DISCRETA)**



La severidad se modela con una distribución Gamma. (Continua- cola larga)



Así pues, el precio mínimo, basado en el costo será

$$P_{\text{perfil}} = E[f]_{\text{perfil}} * E[S|f]_{\text{perfil}}$$

$E[f]_{\text{perfil}} \rightarrow$  Frecuencia estimada para un perfil  $x$

$E[S|f]_{\text{perfil}} \rightarrow$

*Dado que hubo un siniestro que costó tendrá este siniestro para el perfil  $x$*

Dependiendo de la ubicación geográfica la regulación es restrictiva en el tipo de modelo que se puede implementar para calcular el precio base o el precio técnico.

En Europa y Latinoamérica la mayoría de las compañías emplean GLM como instrumento, en contraste si bien en estados unidos la mayoría emplea GLM, otras aseguradoras más modernas, no capacidad técnica alta se han inclinado por complementar la modelación de GLM con otros algoritmos de ML como el LGBM, que no es más que otra regresión basada en árboles de decisiones, la cual ha demostrado excelentes resultados, **el reto es poder explicar perfectamente a la entidad reguladora el significado de las variables del modelo en la definición del precio.**

**GLM (Generalized Linear Model)**

¿Por qué un GLM y no una regresión lineal simple? La regresión lineal típicamente se usa para modelar variables respuesta que se comportan como una distribución normal, más o menos simétricas respecto a un centro, incluso podrían ser variables negativas. En contraste, un GLM permite modelar variables de respuesta con distribuciones diferentes a la normal, usualmente

**distribuciones sesgadas que toman valores positivos, como es el caso de la poisson o la severidad.**

$$g(\mathbb{E}[Y|X]) = X\beta$$

donde:

1.  $Y$  es la **variable de respuesta**.
2.  $X$  representa la **matriz de predictores** o variables independientes.
3.  $\beta$  es el vector de **parámetros** (incluyendo el intercepto y los coeficientes de cada predictor).
4.  $\mathbb{E}[Y|X]$  es la **media condicional** de la variable de respuesta  $Y$ , dado el valor de los predictores  $X$ .
5.  $g(\cdot)$  es la **función de enlace**, que transforma la media condicional para hacerla lineal en los predictores.



$$g(\mathbb{E}[Y|X]) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p$$

**Recuerde, en el caso de la frecuencia  $Y$  se distribuye Poisson y en el caso de la severidad  $Y$  se distribuye Gamma.**

**La función  $g(\cdot)$  se conoce como función de enlace, y en el caso de la frecuencia y la severidad  $g(\cdot)$  es la función logarítmica.**

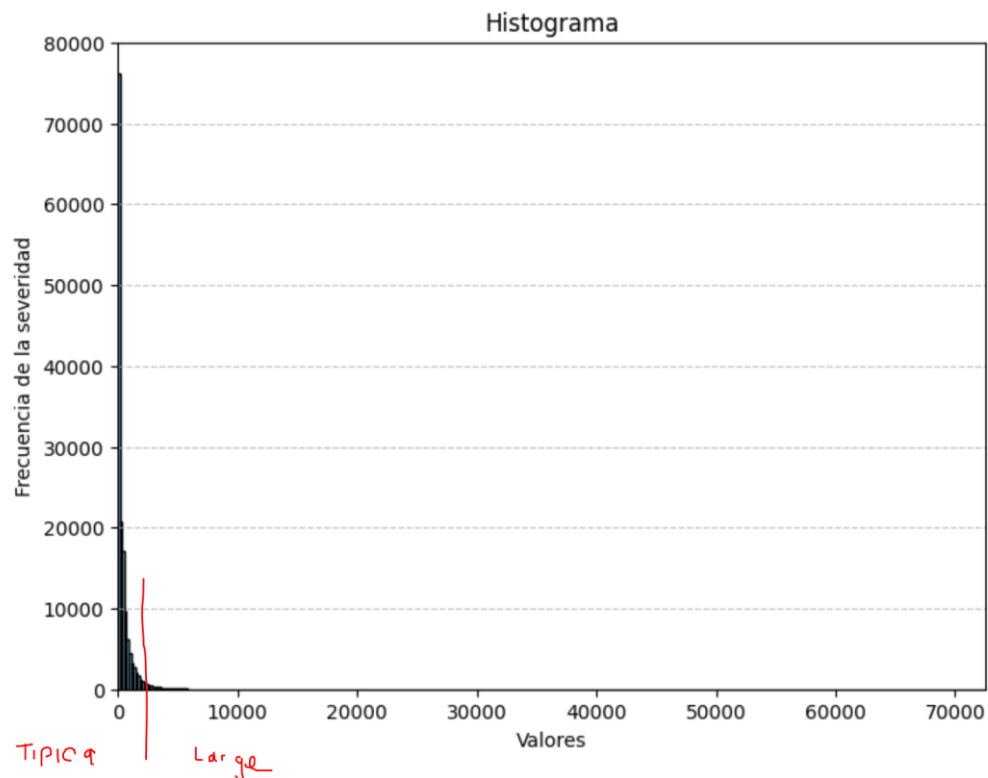
$$g(\mathbb{E}[Y|X]) = \log(\mathbb{E}[Y|X]) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p$$

$$\mathbb{E}[Y|X] = e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p}$$

$$\mathbb{E}[Y|X] = e^{\beta_0} \cdot e^{\beta_1 X_1} \cdot e^{\beta_2 X_2} \cdot \dots \cdot e^{\beta_p X_p}$$

### **Separación severidad típica de la severidad Large**

**Habíamos comentado que la severidad, que se distribuye Gamma es una distribución de cola larga, en ciertos casos es muy muy larga.**



### Perils / Coverturas y cantidad de modelos

COBERTURAS Y SERVICIOS
Responsabilidad Civil Obligatoria y Voluntaria
Seguro de ocupantes del vehículo
Asistencia en viaje
Defensa jurídica y reclamación de daños
Rotura de lunas
Robo
Incendio
Daños propios del vehículo
Valor a nuevo 3 años
Vehículo de sustitución
Servicio gratuito de traslado para pasar la ITV
Asistencia en viaje ampliada
Abogado de referencia
Reconocimiento médico y trámites para la renovación del carné de conducir gratuitos
Asesoría jurídica familiar

Para cada peril se modela la frecuencia y la severidad, de frecuencia hay un dolo modelo pero de severidad hay 3, así:


$$SEV\_total = SEV\_TIPICA + Propension(exceso) * SEV\_LARGE$$

La propensión sería la probabilidad de que dado que existe un siniestro, este sea large.

## Formatos y gestión de variables categóricas.

### One hot encoder

id	color			
1	red			
2	blue			
3	green			
4	blue			



id	color_red	color_blue	color_green
1	1	0	0
2	0	1	0
3	0	0	1
4	0	1	0

### Categorical encoding

Customer feedback - 5 point Likert scale	
Feedback	Assign numerical code
Poor	1
Fair	2
Good	3
Very Good	4
Excellent	5

### MZ, Reducción de la dimensionalidad

Una entidad bancaria o aseguradora posee mucha información de sus asegurados, entre estos datos de información social y demográfica, típicamente asociados a un código postal u otro nivel de granularidad. Para hacer los modelos una entidad de estas puede contar con fácilmente 400, o 500 variables de este tipo, las cuales en su mayoría están altamente correlacionadas, así, entonces ¿cuál es la mejor variable de todas? ¿Debería usarlas todas?

La microzonificación no es más que una técnica de reducción de dimensionalidad, por medio de algoritmos como el PCA, BYM (Besag York

Mollie) entre otras, para “Combinar” todas estas variables en una sola que tenga un mayor efecto predictor para los modelos.

### Loadings/recargos

Una vez tenemos el precio estimado, este se recarga con ciertos factores para llegar al precio técnico.

$$P_{Tecnico} = P \prod_{i=1}^n Loading_i$$

Loadings: Inflación, Factor de siniestros incurridos pero no reportados, Factor siniestros no incurridos.