Análisis de Riesgo de una Compañía Aseguradora

Equipo

Federico Medina García Corral - A01721441 | Karen González Ugalde - A01411597 |
Ricardo de Jesús Balam Ek - A00831262 | Eduardo Emiliano Porcayo Arrieta - A01423285

MA2004B.101 ”Optimización Estocástica”

Abstract

En el presente, se resolvió la problemática de calcular la probabilidad de ruina de una compañía aseguradora, así como el tiempo aproximado en el que esto sucedería. Se utilizaron dos métodos, siendo el de Monte Carlo y Panjer, en donde este último seguía distribuciones exponencial y gamma. En el caso del método Monte Carlo se realizaron dos simulaciones utilizando diferentes primas para ver la comparación de la probabilidad de ruina mediante 10,000 iteraciones, en donde en más del 90 por ciento de nuestros casos se caía en ruina al menos una vez, debido a que en la primer simulación 9,425 de 10,000 llegaron a este fenómeno, y en la segunda, 9,378 de ellos. Además, se realizó una simulación con sólo 100 simulaciones, en donde se obtuvo un tiempo de ruina de 2335 días, equivalente a 6 años y 4 meses. Mientras que en el método de Panjer se obtuvo que la probabilidad más alta en donde la compañía caería en ruina era dentro de un intervalo de 60 y 80 días con sus dos respectivas distribuciones.

Keywords: Probabilidad de ruina, modelo de aproximación de Cramer-Lundberg, método de Monte-Carlo, método de Panjer distribución exponencial, distribución de poisson, distribución gamma.

1. Introducción

Toda compañía, independientemente del giro en el que estén involucradas tienen probabilidad, por mínima que sea, de estar en un momento de ruina. Consecuente de la naturaleza que conlleva el trabajo de una compañía aseguradora, se tiene el objetivo de examinar el comportamiento de sus ingresos y costes.

Las compañías aseguradoras toman un papel muy importante dentro de la estabilidad económica de los sistemas financieros debido a que son grandes inversoras dentro del mercado además de que cada vez aumenta la relación entre las aseguradoras y los bancos debido a que están salvaguardando la estabilidad financiera de muchos hogares y empresas.

La gravedad en la oscilación que existe en una compañía aseguradora depende mucho de la provisión que se tenga, es decir, que el capital con el que se cuente recaudado tanto por una inversión inicial, como del ingreso percibido por las suscripciones o abonos de los clientes afiliados debe ser pensado con el fin de que sea rentable para la empresa.

La manera en la que una compañía de seguros genera ingresos es mediante el cobro de una cierta cantidad de dinero al cliente durante un periodo de tiempo llamadas ”primas”, usualmente se manejan plazos mensuales y anuales, esto a cambio de la cobertura de seguro. Debido a que estos ingresos están expuestos al riesgo de inflación, la empresa reinvierte estas primas en activos seguros a corto plazo. Esto genera ingresos por intereses adicionales mientras espera posibles pagos de sus clientes. [1]

Con esta recaudación es cómo la compañía de seguros incrementa su capital. Luego, a la hora que hay algún accidente o se reclame el dinero asegurado, dicha compañía debe procesarlo, verificar su exactitud y desembolsar la cantidad que se negoció en el contrato con el cliente, menos un deducible. Este proceso de ajuste es necesario para filtrar reclamaciones fraudulentas y minimizar el riesgo de pérdida para la empresa.

Existen empresas que se dedican al reaseguro de compañías aseguradoras que las ayudan a mantenerse solventes en caso de que tengan pérdidas excesivas.

Si una empresa evalúa su riesgo de manera efectiva, debería generar más ingresos en primas de lo que gasta en pagos cubriendo siniestros. Es por esto que se debe de hacer un análisis profundo sobre los costos que se ofrecerán a los clientes, al igual que considerar los accidentes que puedan provocar un alto costo para la compañía. Una compañía de seguros solo debe emitir una póliza con un tope del 10 % de su valor, a menos que esté reasegurada. Esto funciona como un margen de referencia para las aseguradoras y en esto se basan para cobrar tarifas más altas por el seguro a los clientes individuales y luego obtienen tarifas más baratas al reasegurar estas pólizas a gran escala. [2]

Sin embargo, en el momento en que los gastos por reclamaciones (los momentos donde el usuario usa su seguro), sobrepasen con lo que la compañía aseguradora cuenta en el momento, entrará en ruina, lo cual no impide que salga de esa situación, pero no es lo ideal.

Si bien, no se tiene una certeza de cuántos accidentes ocurrirán, aunado a que el costo de cada uno varía, es importante prevenir que estos no sobrepasen nuestro presupuesto, lo que genera cuestionamientos referentes al sustento de las aseguradoras. Como consecuente, una herramienta aliada en empresas de este giro, es el modelo clásico de Cramer-Lundberg. Con este, se abre la posibilidad de conocer en qué momento la compañía estará en ruina, dando respuesta, desde el individualismo de cada empresa, si podrá solventarse en un tiempo determinado finito.

Cabe aclarar que en este artículo no se planea generar un método de análisis riesgo de propia autoría, sino que mostrar cómo se pueden elaborar ciertas estimaciones que podrían dar pie a la ruina de una compañía de seguros cuando se tienen en cuenta ciertos factores económicos. Las reglas y parámetros de la problemática serán utilizadas para seguir el modelo clásico de Cramer-Lundberg.

2. Descripción de la problemática

Para entender un poco sobre el contexto de la problemática, se debe de entender que las aseguradoras son un negocio, es decir, necesitan generar más dinero del que pueden llegar a gastar para sobrevivir. Sin embargo, como se explicó en la introducción, es imposible para una compañía predecir con exactitud el número de accidentes que habrán en un tiempo determinado. Es por esto que dentro de la industria aseguradora se tiene que utilizar probabilidades para sobrevivir y así poder tomar decisiones al reducir la probabilidad de riesgo que cada alternativa pueda tener.

Una de las decisiones que se deben de tomar es cuánto es el monto que un cliente debe pagar a la aseguradora en un tiempo determinado para que esta pueda cubrir los daños del cliente al igual que poder sobrevivir al mismo tiempo que aumentar su capital. Uno de los modelos utilizados para poder encontrar dicho flujo de ingreso es el modelo clásico de Cramer-Lundberg, también llamado modelo clásico de riesgo.

Dentro del modelo clásico de Cramer-Lundberg, se obtendrá la representación del capital de la compañía en un tiempo t, el cual está dado por $X(t)$. De igual manera, la representación del capital inicial se definirá como $u \geq 0$, y las primas $c > 0$, que es el monto de dinero que recibe por unidad de tiempo a través de usuarios. Estos aspectos anteriormente mencionados, son únicamente entradas monetarias, sin embargo, se ocupa restar el tamaño de las reclamaciones Z_k, las cuales hay que tener siempre en cuenta que son variables aleatorias independientes una de otra, pero idénticamente distribuidas, que van desde 1 hasta el límite $N(t)$, que representa el número de reclamaciones que llegan a la compañía en un intervalo de $(0, t]$.

$$X(t) = u + ct - \sum_{k=1}^{N(t)} Z_k \tag{1}$$

El porqué del uso de esta metodología, es que es un modelo de entrada, el cual resulta fiable para poder estimar la probabilidad de ruina, en este caso de una compañía aseguradora. También es relevante conocer cuándo determinar que una compañía está en ruina. Intuitivamente se considera que este fenómeno ocurre cuando en cualquier $t \geq 0$, el capital de la compañía $X(t) \leq 0$.

3. Preguntas de Investigación

1. **¿Hasta qué punto se puede solventar la compañía?**

   Dada la pregunta anterior, se busca encontrar el punto donde, en base a diferentes parámetros que se le otorguen a la función dentro de las simulaciones realizadas, la compañía entre a la región de ruina en base al modelo clásico de Cramer-Lundberg.

2. **¿Cuál es la probabilidad de que la aseguradora se vaya a la ruina en un periodo de tiempo?**

   Mezclando ambas preguntas, lo que se busca en esta es, en caso de que haya una posibilidad de ruina de la compañía, encontrar la probabilidad de que dicha cosa suceda, igualmente basándose en los diferentes parámetros que la función utilizará en las simulaciones que se elaborarán.

4. Metodología utilizada

Para desarrollar una función que pueda describir el momento en que una compañía aseguradora entra en la región de ruina, se debe de entender el fundamento de los cálculos utilizados para desarrollar dicha función. Para esto, se utilizó el modelo clásico de Cramer-Lundberg y algunas aproximaciones para poder tener la mejor versión de la función para el comportamiento de una aseguradora con base en distintas variables y parámetros que definen el comportamiento de su probabilidad de ruina.

4.1. Modelo clásico de Cramer-Lundberg

Retomando lo que se mencionó en la descripción de la problemática, se sabe que $N(t)$ es el número de reclamaciones que llegan a la compañía en el intervalo $(0, t]$. $X(t)$ se llama modelo clásico de riesgo, o modelo de Cramer-Lundberg. Se dice que una compañía aseguradora está en ruina cuando para algún $t \geq 0$, $X(t)$ se encuentra por debajo de cero y el momento de ruina está dado por:

$$T = min\{t \geq 0 : X(t) < 0\} \tag{2}$$

Si $X(t) < 0$ para algún tiempo t finito, y $T = \infty$

Si $X(t) \geq 0$ para todo t.

Si se denota por $\varphi(u)$ la probabilidad de ruina de la compañía aseguradora, empezando con capital inicial u, entonces para algún $t > 0$:

$$\varphi(u) = P(X(t) < 0 | X(0) = u) \tag{3}$$

La probabilidad de no ruina o de supervivencia de la compañía será denotada por $\phi(u)$ y por lo tanto:

$$\phi(u) = 1 - \varphi(u)$$

Si se supone que $u = 0$ y usando la independencia entre $N(t)$ y $\{Z_1, Z_2, ...\}$, se deduce que:

$$E[X(t)] = t(c - \lambda\mu) \geq 0 \tag{4}$$

por lo que se supondrá que $c \geq \lambda\mu$.

Definición 1. La carga de seguridad de la compañía, denotada por ρ, se define como:

$$\rho = \frac{c - \lambda\mu}{\lambda\mu} > 0 \tag{5}$$

Esta condición es necesaria para asegurar la solvencia de la compañía por la observación anterior. [3]

4.2. Aproximaciones de la probabilidad de ruina.

1. **Aproximación de Cramer-Lundberg**

   Definición 2. Se define la función h:

   $$h(r) = \int_0^\infty e^{rz}\, dF(z) - 1, \tag{6}$$

   Se asume que existe $0 < r \leq \infty$ tal que $h(r) \uparrow +\infty$ cuando $r \uparrow r_\infty$.

   Con esto se sabe que F es de cola ligera, ya que para algún $r > 0$, la función generadora de momentos de Z es finita. La cola de F decrece más rápido que una función exponencial.
   Además, suponemos que que existe una constante $R > 0$, tal que:

   $$\frac{\lambda}{c}\int_0^\infty e^{Rz}[1 - F(z)]\, dz = 1, \tag{7}$$

   Cuando esta constante existe es única y se llama coeficiente de ajuste.

   Teorema 1. Cuando el coeficiente de ajuste R existe, se tiene la aproximación de Cramer-Lundberg:

   $$\lim_{u\to\infty} (e^{Ru}\varphi(u)) = \frac{\rho\mu}{h'(R) - \frac{c}{\lambda}} \tag{8}$$

   De aquí se tiene que para u grande:

   $$\varphi(u) = \frac{\rho\mu}{h'(R) - \frac{c}{\lambda}} e^{-Ru} \tag{9}$$

   Como consecuencia la probabilidad de ruina en los procesos de riesgo para los cuales existe R, decrece exponencialmente como función del capital inicial u, lo cual significa que la probabilidad de ruina es muy pequeña.

   Teorema 2. Cuando existe el coeficiente de ajuste R, para todo $u \geq 0$ se cumple la desigualdad,

   $$\varphi(u) < e^{-Ru} \tag{10}$$

   Debido a la desigualdad anterior, R se puede considerar una medida de riesgo para la compañía de seguros cuanto más grande es R, más pequeña es la probabilidad de ruina.

2. **Transformada de Laplace (cálculo exacto)**
   Definición 3. La transformada de Laplace, $L_a(s)$ de una función real no negativa $a(x)$ se define por medio de:

   $$L_a(s) = \int_0^\infty e^{-sx} a(x)\, dx, \tag{11}$$

   Teorema 3. Las transformadas de Laplace para $\phi(u)$ y $\varphi(u)$ denotadas por $L_{\phi(u)}(s)$ y $L_{\varphi(u)}(s)$

   $$L_{\phi(u)}(s) = \frac{c - \lambda\mu}{cs - \lambda(1 - l_F(s))}, s > 0 \tag{12}$$

   $$L_{\varphi(u)}(s) = \frac{1}{s} - \frac{c - \lambda\mu}{cs - \lambda(1 - l_F(s))}, s > 0 \tag{13}$$

   donde $l_F(s) = \int_0^\infty e^{-sy}\, dF(y)$

   Nótese que en los casos cuando se puede calcular la inversa de la transformada de Laplace, se obtiene una fórmula exacta para la probabilidad de ruina.

3. **Fórmula de Pollaczek-Khinchin**
   Definición 4. Sea Z una variable aleatoria no negativa con función de distribución F. Entonces la cola integrada de F, denotada por $F_Z^s(y)$ se define como:

   $$F_Z^s(y) = \begin{cases} \frac{1}{\mu}\int_0^\infty [1 - F_Z(y)]\, dy & para \quad x \geq 0 \\ 0 & o.c. \end{cases} \tag{14}$$

   Teorema 4. (Fórmula de Pollaczek-Khinchin). Para todo $u \geq 0$,

   $$\phi(u) = \left(1 - \frac{\lambda\mu}{c}\right)\sum_{n=1}^{\infty}\left(\frac{\lambda\mu}{c}\right)^n (F_Z^s)^{*n}(u) \tag{15}$$

   La fórmula es difícil de usarse por el cálculo que conlleva, pero con ella se puedan hacer aproximaciones a la probabilidad de ruina de la siguiente forma: $1-\phi(u)$ es la distribución de una variable aleatoria geométrica compuesta con parámetros $\left(\frac{\lambda\mu}{c}, F_Z^s\right)$.
   En efecto, sea N una variable aleatoria geométrica de parámetro $p = \frac{\lambda\mu}{c}$ y $X = \sum_{i=1}^n Z_i$ tal que $Z_1, Z_2, ...$, son variables aleatorias independientes con distribución F_Z, entonces:

   $$P(X \leq x) = \sum_{n=0}^{\infty} P(X \leq x | N = n) p(1-p)^n \tag{16}$$

   $$P(X \leq x) = \sum_{n=0}^{\infty} p(1-p)^n F^{*n} \tag{17}$$

Reemplazando F_Z por F_Z^s se obtiene que:

$$1-\varphi(u)=\sum_{n=1}^{\infty}\left(1-\frac{\lambda\mu}{c}\right)\left(\frac{\lambda\mu}{c}\right)^n\left(F_Z^s\right)^{*n}(u) \tag{18}$$

Con esta fórmula se puede obtener aproximaciones de las probabilidades de ruina usando el algoritmo de Panjer.

4. **Fórmula recursiva de Panjer**
Calcular o aproximar la función de distribución del número de reclamos acumulados en un intervalo de tiempo dado ha sido uno de los puntos centrales en matemática de seguros. Algunos ejemplos para la distribución del número de reclamos de procesos de riesgo en un intervalo de tiempo dado son:

- La distribución de Poisson. El proceso de riesgo clásico de Cramer-Lundberg supone que los reclamos se distribuyen de acuerdo a un proceso de Poisson.
- La distribución exponencial
- La distribución binomial negativa o distribución de Pascal.
- La distribución binomial.
- La distribución geométrica.

Estas distribuciones de probabilidad cumplen la relación recursiva de Panjer:

$$p_k=P(X=k)=\left(a+\frac{b}{k}\right)p_{k-1} \tag{19}$$

donde $a<1$ y b son constantes fijas.

Considérese la variable aleatoria compuesta

$$X=\sum_{i=1}^{N}Z_i \tag{20}$$

donde Z_1, Z_2,..., son variables aleatorias independientes e idénticamente distribuidas y N es una variable discreta no-negativa; se asume también que la sucesión Z_1, Z_2,..., es independiente del número de reclamos N. Se asume que los Z_i y N toman valores en los no negativos. Se denotará por q_k la distribución de probabilidad de Z_1, Z_2 ,..., y la distribución del número de reclamos por p_k.

Teorema 5. Sea $X=\sum_{i=1}^{N}Z_i$ y supóngase que la distribución de probabilidad de N satisface la relación recursiva de Panjer. Entonces,

$$r_j=\begin{cases} g_N(q_0) & j=0\\ (1-aq_0)^{-1}\sum_{k=1}^{j}(a+bkj^{-1})q_kr_{j-k} & j=1,\ldots \end{cases} \tag{21}$$

donde g_N es la función generadora de probabilidades de N.

5. Análisis de datos

En esta sección se demostrará qué tipo de distribución tienen dos de nuestras variables más importantes que son la cantidad de siniestros y el monto de cada uno de estos. En la base de datos proporcionada estas columnas están nombradas como “Reclamo de no-Cobertura” y “Monto del siniestro” respectivamente.

Antes que nada se debe hacer un pre-procesamiento de los datos para que sean más fáciles de manejar. Inicialmente la base de datos tenía la siguiente forma:

	Fecha del Siniestro	Tipo de auto	Modelo	Monto del siniestro	Aplica cobertura	Deducible	Reclamo de no-Cobertura	Pérdida total
0	01/12/20	Deportivo	2017	200000	Si	NaN	Si	Si
1	12/05/20	Austero	2016	100000	Si	NaN	Si	Si
2	11/01/20	compacto	2017	150000	Si	NaN	Si	Si
3	21/12/20	Subcompacto	2017	70000	Si	NaN	Si	Si
4	30/05/20	Subcompacto	2019	90000	Si	NaN	Si	Si
...	...	...	...	...	...	...	...	...
27093	08/07/20	Subcompacto	2017	10840	Si	2400.0	No	No
27094	23/12/20	De Lujo	2016	42280	Si	5600.0	No	No
27095	10/08/20	Subcompacto	2016	15700	Si	2600.0	No	No
27096	02/10/20	Austero	2016	8730	Si	1800.0	No	No
27097	16/07/20	Subcompacto	2017	16380	Si	2400.0	No	No

27098 rows × 8 columns

Figura 1. Base de datos inicial

Después se realizó una matriz de correlación para determinar que variables son significativas para el análisis:

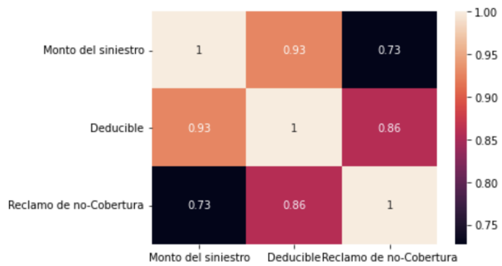

Figura 2. Matriz de Correlación

Se pudo observar que las variables con mayor relación lineal entre ellas fueron la cantidad de reclamos y el monto de los siniestros, por lo que se procedió a remover todas las columnas excepto las de “Fecha del Siniestro”, “Monto del siniestro” y “Reclamo de no-Cobertura” ya que son las variables significativas para nuestro análisis.

De la columna “Reclamo de no-Cobertura” se removieron todos los “Si” ya que son los casos en los que los clientes no aplicaron la cobertura y por lo tanto no representaron un gasto para la compañía. Posteriormente se reemplazaron todos los “No” con el número 1 para poder contabilizar estos casos.

	Fecha del Siniestro	Monto del siniestro	Reclamo de no-Cobertura
0	01/01/20	6210	1
1	01/01/20	111290	1
2	01/01/20	21320	1
3	01/01/20	28610	1
4	01/01/20	21430	1
...	...	...	...
25799	31/10/20	36520	1
25800	31/10/20	13730	1
25801	31/10/20	3010	1
25802	31/10/20	14580	1
25803	31/10/20	16920	1

25804 rows × 3 columns

Figura 3. Base de datos final

5.1. Comportamiento de la Cantidad de Siniestros via Minitab

Primeramente para esta sección agrupamos los datos por fecha sumando así la cantidad de siniestros que es la columna de ”Reclamo no-Cobertura”para poder obtener el comportamiento de esta. Cabe mencionar que no se tomó en cuenta la columna de ”Monto del siniestro”dentro del análisis en el apartado de frecuencia ya que obtenía resultados que no iban de acuerdo a lo que se quería demostrar.

Fecha del Siniestro	Monto del siniestro	Reclamo de no-Cobertura
01/01/20	2672680	82
01/02/20	2399880	72
01/03/20	2430930	72
01/04/20	2635060	83
01/05/20	2672620	81
...	...	...
31/03/20	2045920	69
31/05/20	3302980	94
31/07/20	2297410	69
31/08/20	2206440	79
31/10/20	2336000	87

365 rows × 2 columns

Figura 4. Base de datos agrupada por fechas

Para encontrar el comportamiento de la Cantidad de Siniestros de los datos, se utilizó la herramienta de Ji-Cuadrada de Minitab en donde se establecieron dos distintas hipótesis para poder encontrar una solución con un nivel de significancia del 95 % . Para esto, se establecieron ambas hipótesis:

- H_0: Los datos de Cantidad de Siniestros siguen una distribución de Poisson

- H_1: Los datos de Cantidad del Siniestros no siguen una distribución de Poisson

Con esto, se encontró el p-valor, el cual es la probabilidad de que se obtenga algún valor sea posible suponiendo que la hipótesis nula es correcta.

Chi-Square Test

Null hypothesis H_0: Data follow a Poisson distribution

Alternative hypothesis H_1: Data do not follow a Poisson distribution

DF	Chi-Square	P-Value
21	28.1391	0.136

Figura 5. Prueba Ji-Cuadrada

Obteniendo estos resultados, se encontró que el valor del p-value de la prueba estadística utilizada es de 0.136. Tomando en cuenta la teoría de la interpretación de la prueba de **bondad de ajuste** de la ji cuadrada en el libro de ”Probabilidad y estadística para ingeniería y administración”de D. Montgomery [4] se utilizó el criterio de nivel de significancia que es igual a 0.05, lo cual significa que no se rechaza la hipótesis nula, dando a entender que los datos siguen una distribución de Poisson.

Sin embargo, para hacer el análisis de hipótesis más preciso, se encontró una gráfica donde se distribuyen los datos del problema y sus valores esperados, los cuales se observan como a continuación:

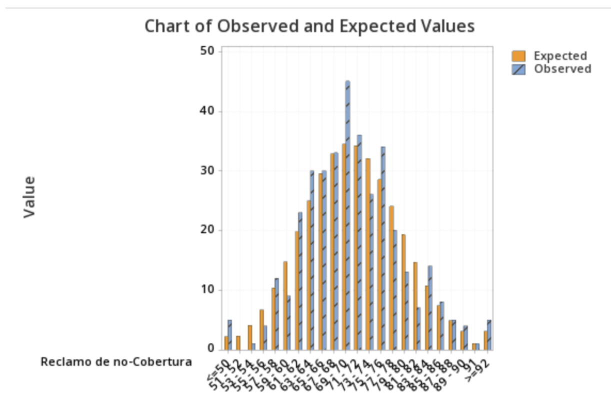

Figura 6. Gráfica de Valores Observados y Esperados

Con esto se encontró que los valores esperados y los valores reales son muy parecidos, lo cual significa que los valores reales sí siguen una distribución de Poisson [5]. Finalmente, utilizando ambas pruebas estadísticas, se concluye que los datos de la Cantidad de Siniestros efectivamente sigue una distribución de Poisson.

Por último, sabiendo que los datos siguen una distribución de Poisson, se puede verificar que el parámetro o la media para dichos datos tiene un valor de $\lambda = 70.696$ tal y como se muestra en la Figura 7 [6].

Descriptive Statistics

N	Mean
365	70.6959

Figura 7. Media de Cantidad de Siniestros y λ = 70.69

5.2. Comportamiento de Monto del Siniestro via Minitab

Para poder observar el comportamiento de la variable Monto del Siniestro, la cual cuenta la cantidad total que la compañía de seguros tuvo que desembolsar a causa de los accidentes o reclamaciones, se tuvieron que utilizar herramientas de calidad para asegurarse de cuál era la distribución que seguían los datos de esta variable.

Se tiene como hipótesis que el comportamiento de la variable de Monto del Siniestro seguirá una distribución exponencial, debido a que los datos tienen una propiedad de falta de memoria, es decir, que la probabilidad de tener que esperar un tiempo t no depende de un tiempo ya transcurrido. [7]

La base de datos tuvo que limpiarse, ya que se contaban con columnas que contenían datos nulos, los cuales iban a ser una distracción y obstáculo al momento de analizar. Se utilizó el software de Minitab para obtener las gráficas de distintas distribuciones con un criterio α = 0.05, y así, poder observar qué se ajusta mejor a los datos. Consecuentemente, se logrará rechazar alguna de las siguientes hipótesis:

- H_0: Los datos del monto del siniestro siguen una distribución exponencial
- H_1: Los datos del monto del siniestro no siguen una distribución exponencial

Se obtuvieron diferentes combinaciones de comportamientos, pero sólo se mostrarán los resultados de cómo se ajustaron con la distribución exponencial.

Figura 8. Gráfica de probabilidad para Monto del siniestro

Como puede observarse, además de que se percibe a simple vista que se ajusta a la distribución exponencial, se cuenta para esta un valor p = 0.083, lo cual es mayor que el criterio α de significancia establecido a priori, por lo que no rechazamos la H_0.

Así mismo se obtuvieron las estadísticas descriptivas de la variable y un valor de lambda de 0.256.

Estadísticas descriptivas

N	N*	Media	Desv.Est.	Mediana	Mínimo	Máximo	Asimetría	Curtosis
27098	0	30778.3	31240.5	21280	10	341510	2.09878	6.77879

transformación de Box-Cox: λ = 0.256187

Figura 9. Estadísticas descriptivas y valor de μ = 30778.3

Prueba de bondad del ajuste

Distribución	AD	P	LRT P
Normal	1296.121	<0.005	
Transformación Box-Cox	7.479	<0.005	
Lognormal	306.363	<0.005	
Lognormal de 3 parámetros	90.971	*	0.000
Exponencial	1.132	0.083	
Exponencial de 2 parámetros	1.184	0.076	0.000
Weibull	0.433	>0.250	
Weibull de 3 parámetros	0.366	0.454	0.000
Valor extremo más pequeño	3063.694	<0.010	
Valor extremo por máximos	490.793	<0.010	
Gamma	0.614	0.138	
Gamma de 3 parámetros	0.474	*	0.000
Logística	842.737	<0.005	
Loglogística	184.394	<0.005	
Loglogística de 3 parámetros	145.303	*	0.000

Figura 10. Prueba de bondad de ajuste

Cabe mencionar igual que, para la prueba de bondad de ajuste, Minitab utiliza el estadístico de Anderson-Darling para calcular el valor p. El valor p es una probabilidad que mide la evidencia en contra de la hipótesis nula de que los datos siguen la distribución.

Generalmente, valores substancialmente menores del estadístico de Anderson-Darling indican que los datos siguen una distribución más de cerca. Sin embargo, se debe evitar comparar directamente los valores AD en diferentes distribuciones cuando los valores AD son cercanos, porque los estadísticos AD se distribuyen de manera diferente para distintas distribuciones. Para comparar mejor el ajuste de diferentes distribuciones, se debe utilizar criterios adicionales, tales como las gráficas de probabilidades, los valores p y el conocimiento del proceso. En el caso para la exponencial, el valor de AD es relativamente bajo, pero tomando los criterios adicionales del valor p y las gráficas, se puede ver que la hipótesis nula “los datos del monto del siniestro siguen una distribución exponencial” no se rechaza, como se había visto anteriormente.

Cabe mencionar que al buscar si había una dependencia entre el Número de Reclamos y el Monto del Siniestro, se elaboraron diferentes estadísticas descriptivas en Minitab. Con esto, se encontró que ambas variables no tienen una relación entre si, por lo que son **independientes**. Igualmente, al razonar un poco sobre ambas, se intuyó que ambas debían ser independientes ya que, teniendo el número de choques no se puede encontrar el monto a pagar gracias a diferentes factores como la gravedad del accidente, el valor de los carros involucrados, los daños externos ocacionados, etc.

Finalmente, los parámetros importantes encontrados para la elaboración del código son:

- Media de Poisson: $\lambda = 70.69$
- Media de Exponencial: $\mu = 30778.30$

6. Método de Monte Carlo

Este es un método estadístico utilizado para resolver problemas matemáticos a través de la generación de variables aleatorias haciendo tantos intentos hasta tener los números más cercanos [8]. Consiste en una 'simulación' donde se repite o duplica las características y comportamientos de un sistema real. Se busca imitar el comportamiento de variables reales para analizar o predecir cómo van a evolucionar.

Algunas de sus aplicaciones son:

- Crear, valorar y analizar carteras de inversión.
- Valorar productos financieros complejos.
- Creación de modelos de gestión de riesgo.

6.1. Elaboración de la Simulación del Método de Monte Carlo

Para elaborar dicha simulación sobre la probabilidad de ruina de una compañía aseguradora, se tuvo que hacer un proceso de limpieza de datos ya que se elaboraron 10,000 simulaciones con un total de 5000 entradas para cada uno, es decir, se tenían que simular por lo menos 50,000,000 iteraciones. Para reducir la potencia que la computadora debía utilizar, se borraron todas las columnas que se consideraban innecesarias para obtener los resultados que se buscaban, las cuales crearon un nuevo DataFrame con las columnas:

- Fecha del Siniestro
- Monto del Siniestro: Z_n representa sus valores y μ representa su media
- Reclamo no cobertura: N_n representa sus valores al sumar todos sus valores por día y λ representa su media

Para hacer la simulación, se tuvieron que encontrar los valores de los parámetros anteriormente mencionados. Estos se encontraron tal y como se explica en la sección 5.1 y 5.2, siendo el valor de μ la media de una distribución exponencial y λ la media de una distribución Poisson. Con esto, se buscó encontrar un número al azar que siga una distribución Poisson con parámetro $\lambda = 70.6959$ (valor explicado en la sección 5.1) para simular el número de siniestros que ocurren en un día dado. Igualmente se buscó encontrar la cantidad de monto del siniestro utilizando un número encontrado al azar en una distribución exponencial con media $\mu = 30778.30$ (como se muestra en la sección 5.2).

En seguida, faltaba definir los valores de la capital inicial, representada por u y el valor de las primas de la aseguradora, representada por c. El valor de u se escogió eligiendo un monto que se consideró "justo" para iniciar una compañía, siendo 20 millones de pesos. Por otro lado, el valor de c se encontró al buscar el costo promedio de asegurar un carro en Estados Unidos [2], siendo 30,500 pesos anuales. Sin embargo, para utilizar la c se debía cambiar las unidades de "por persona anual" a "total de los clientes por día", lo cual se tuvo que dividir entre los 365 días del año y multiplicar por la longitud de la base de datos, lo cual cumple con $c > \lambda\mu$ [3].

Por último, se aplicó el modelo de Cramer-Lundberg para encontrar el número de simulaciones que caían en ruina en algún punto dentro de los 5000 días que se simulaban. Los resultados de dicha simulación se muestran en la Figura 10 y en la Figura 11.

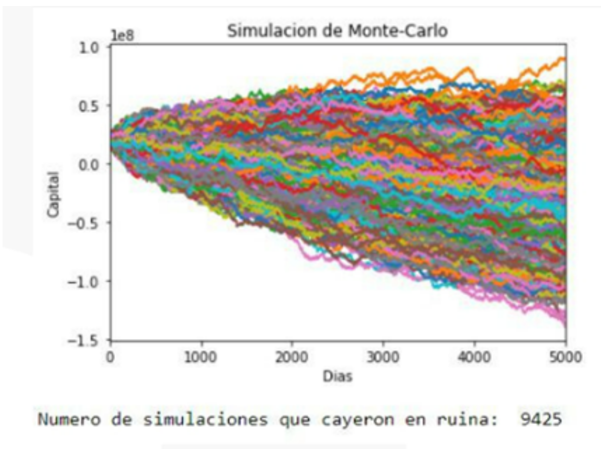

Figura 11. Simulación No. 1 de Monte-Carlo

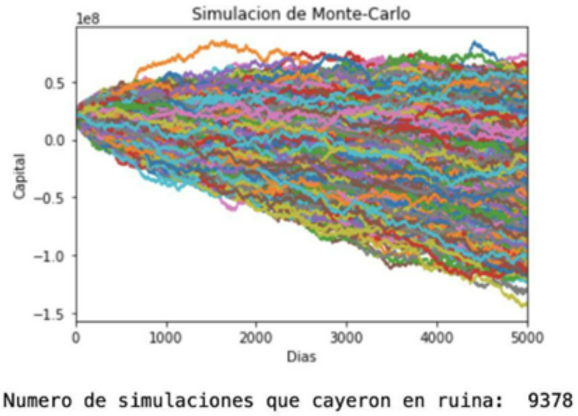

Figura 12. Simulación No. 2 de Monte-Carlo

Como se puede observar con los resultados de las simulaciones, el número de casos que cayeron en ruina utilizando los parámetros anteriormente mencionados fueron de 9425 y de 9378 casos, siendo el equivalente al 94.25 % y 93.78 % de ruina respectivamente.

Para estas simulaciones calculamos el error como se muestra en la Figura 13 tomando en cuenta un intervalo de confianza del 95 %. Este parámetro se consiguió a través de la fórmula $(LS - LI)/2$, donde se obtuvo un total de 0.001447 = 0.1447 %. Debido a que este error es mínimo y esta debajo del 0.5 %, quiere decir que es un modelo confiable para el caso específico que se está tratando, y no será necesario que se realice una reducción de varianza para buscar disminuir esa incertidumbre.

```
valor_esperado = 1 / _lambda
print("La media es: ", valor_esperado)

valor_esperado_c = valor_esperado - valor_esperado / (len(datos2) - 1)
print("La media corregida es: ", valor_esperado_c)
```

```
La media es:  0.014145093783909 47
La media corregida es:  0.014106233636151476
```

```
""" Maximum likelihood 95% confidence"""
_epsylon = valor_esperado_c * 1.96/(len(datos2)**0.5)
print("Lambda con intervalo de confianza: {0:8f} +/- {1:8f}".format(valor_esperado_c, _epsylon))
print("Intervalos de confianza: ({0:8f}, {1:9f})".format(valor_esperado_c - _epsylon, valor_esperado_c + _epsylon))
print("El error es de: {0:8f}".format(_epsylon))
```

```
Lambda con intervalo de confianza: 0.014106 +/- 0.001447
Intervalos de confianza: (0.012659,  0.015553)
El error es de: 0.001447
```

Figura 13. Error con intervalo de confianza del 95 %

Así mismo, se realizó un estudio con 100 simulaciones, capital inicial u de \$20,000,000, recibiendo una prima c de \$30,500 para obtener un valor aproximado de en qué día caería en ruina la compañía.

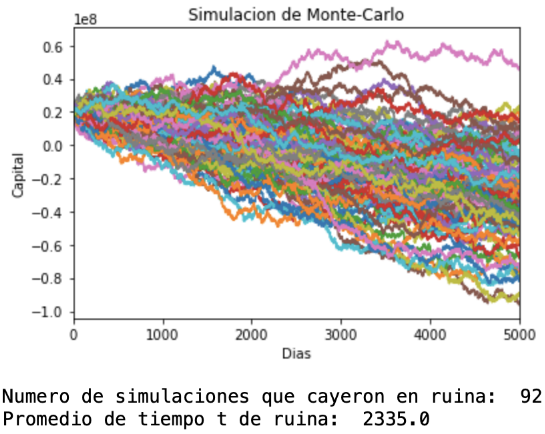

Figura 14. Simulación No. 3 para calcular el tiempo de ruina

Con estos resultados determinamos que bajo estas circunstancias la compañía caería en ruina en 2,335 días, es decir, 6 años y 4 meses aproximadamente.

Sin embargo, para buscar otro resultado, se decidió incrementar el precio de las primas que recibe la compañía, subiendo el costo de 30500 pesos anuales a 31500 anuales, es decir, un incremento de 1000 pesos anuales, equivalente a 2.74 pesos diarios. Con esto, se volvió a hacer una simulación y se obtuvo la gráfica mostrada en la Figura 12.

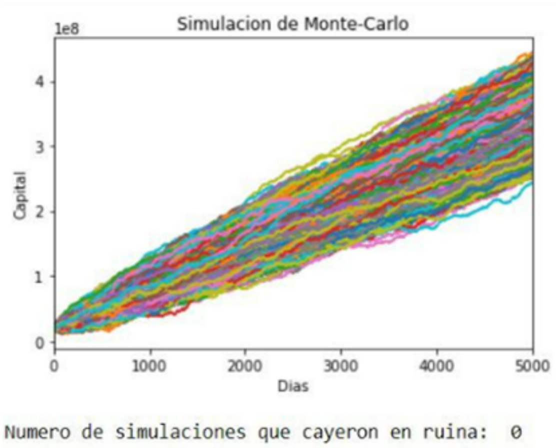

Figura 15. Simulación No. 4 con incremento de primas

Al observar dichos resultados, se puede ver que, increíblemente, al incrementar el valor de la prima por un valor de 1000 pesos anuales, el número de casos que cayeron en ruina fueron 0 de las 10,000 simulaciones elaboradas, constando de un porcentaje equivalente al 0 % de ruina.

7. Método de Panjer

Retomando la teoría vista en la sección 4.2 sobre la fórmula recursiva de Panjer se diseñó un algoritmo para obtener el intervalo de días en donde es más probable que la compañía aseguradora caiga en ruina.

Fórmula recursiva de Panjer:

$$p_k = P(X = k) = (a + \frac{b}{k})p_{k-1} \tag{22}$$

Sabiendo que el número de reclamos esta distribuido por un proceso de Poisson, los valores para los parámetros a,b y p_0 se obtuvieron del artículo referenciado en [9]:

- $a = 0$
- $b = \lambda$
- $p_0 = e^{-\lambda}$

$$p_k = (\frac{\lambda}{k})(p_{k-1}) = (\frac{70{,}69}{k})(p_{k-1}) \tag{23}$$

Para las siguientes distribuciones, se tuvieron que discretizar los datos ya que Panjer solo funciona con variables discretas.

7.1. Distribución Exponencial

Para este caso se consideró que la compañía empieza con un capital inicial de $u = 20,000,000$ y que el número de reclamos están distribuidos de manera exponencial con parámetro:

- $\mu = 30,778$.

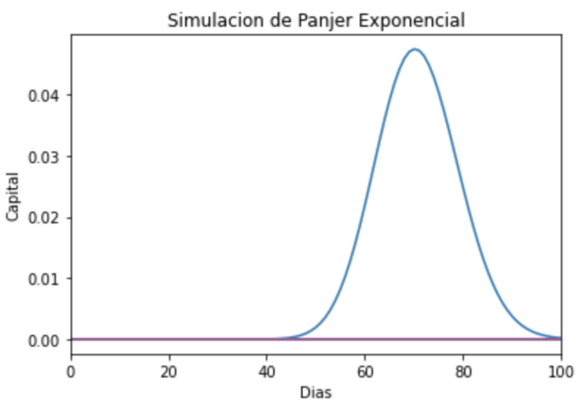

Figura 16. Simulación del método de Panjer con distribución exponencial

Analizando la gráfica se pudo concluir que la compañía aseguradora caería en ruina en el intervalo de 60 y 80 días.

7.2. Distribución Gamma

Para este caso se realizó lo mismo que la sección anterior, un capital inicial de $u = 20,000,000$ pero ahora considerando que el número de reclamos tiene una distribución gamma con parámetros α y β.

Para obtener el valor de estos parámetros se utilizaron como base las formulas del sitio web referenciadas en [10] [11] [12]:

$$\alpha = \frac{\lambda^2}{\sigma^2} \quad , \quad \beta = \lambda \tag{24}$$

La media del número de reclamos λ su desviación estándar σ se obtuvieron con la herramienta de Minitab en el apartado de estadísticas descriptivas. Entonces sustituyendo estos valores, tenemos que:

$$\alpha = \frac{\lambda^2}{\sigma^2} = \frac{(70{,}69)^2}{(8{,}24)^2} = 73{,}59 \tag{25}$$

$$\beta = \lambda = 70{,}69 \tag{26}$$

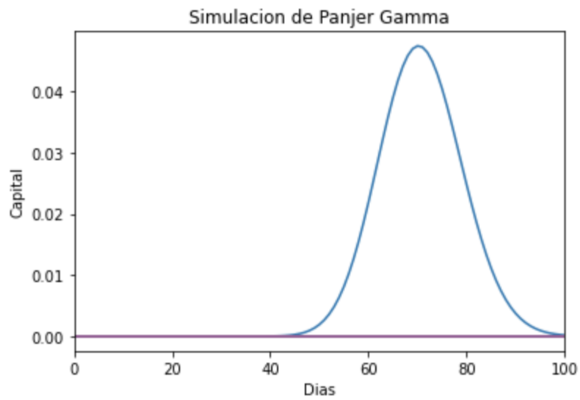

Figura 17. Simulación del método de Panjer con distribución gamma

Al igual que la gráfica anterior, se pudo concluir que la compañía aseguradora caería en ruina en el intervalo de 60 y 80 días.

8. Conclusión

Con base en los resultados obtenidos es de gran importancia determinar el comportamiento de los datos de cada una de las variables para tener un mejor entendimiento del funcionamiento de la compañía aseguradora.

Gracias al análisis realizado con ayuda de Minitab, fue posible encontrar información que, a simple vista, no hubiera sido sencillo de recolectar. Se realizaron pruebas de hipótesis para corroborar que las distribuciones de los datos son tanto Poisson como exponencial; con esto, se facilita encontrar una función para calcular la probabilidad de ruina de una compañía aseguradora con el modelo de Cramer-Lundberg y así poder tomar decisiones con base en estadísticos confiables, esperados y reales. La relevancia de la interpretación de este comportamiento tiene su auge en donde se estima si existe una probabilidad de ruina, de lo cual ninguna empresa quiere llegar. Por lo tanto, fue de vital importancia tener un buen entendimiento de los datos para posteriormente continuar con el análisis e interpretaciones correspondientes.

Cabe mencionar que el modelo de Cramer-Lundberg se seleccionó por los parámetros que engloba para su cálculo, el cual considera muchos factores relevantes en la vida real.

Con ello se hicieron simulaciones según el método de Monte Carlo, con un total de 10,000 simulaciones en un máximo de 5,000 días, en el cual se observó cómo el cambio de la prima constante, aunque fuera mínimo, generaba una gran alteración en el comportamiento de las simulaciones, cambiando notablemente el número de simulaciones que cayeron en ruina. Por otra parte, también se trabajó con la fórmula recursiva de Panjer, la cual utiliza una distribución para el cálculo de la probabilidad. En este caso, se usaron las distribuciones exponencial y gamma, validándolas con pruebas de hipótesis. Con Panjer se logró estimar la sensibilidad de la solución, ver cómo se distribuye la probabilidad del Monto del siniestro y ver en qué intervalo de tiempo es más probable que la aseguradora caiga en ruina, para que con ello, sea posible tomar decisiones preventivas dentro de la compañía para evitar caer en ruina.

Referencias

[1] D. Nissim, *Analysis and valuation of insurance companies*. Center for Excellence in Accounting and Security Analysis, Jan 2011.

[2] S. Ross, "What Is the Main Business Model for Insurance Companies?," July 2021.

[3] E. T. Kolkovska, J. C. Jiménez Hernández, and A. D. Maldonado Santiago, "Probabilidad de ruina en el modelo clásico de Cramer-Lundberg," *Universidad Tecnológica de la Mixteca — Centro de Investigación en Matemáticas*, vol. 15, 12 2011.

[4] D. C. Montgomery and W. W. Hines, *Probabilidad y estadística para ingeniería y administración*. CONTINENTAL, S.A. DE C.V. MÉXICO, 2 ed., 1996.

[5] D. Sharpe, *Chi-square test is statistically significant: Now what?* Practical Assessment, Research, and Evaluation: Vol. 20, Article 8., 2015.

[6] M. L. Cruz D., *Distribucion Poisson-Pascal generalizada utilizando el algoritmo de Panjer.*, vol. 3. Universidad Autónoma de Colombia, 2010.

[7] R.-M. L. Martín Pliego FJ, *Estadística I: Probabilidad*. 1997.

[8] EALDE, *En qué consiste el método de simulación de Monte Carlo*. EALDE Business School, 8 2020.

[9] C. Escalante Coterio, "Distribuciones clase (a, b): estimación y generación de números aleatorios," *Comunicaciones en Estadística*, vol. 10, p. 145, 5 2017.

[10] Editor, "How to find alpha and beta from a gamma distribution?," 4 2018.

[11] Editores, *4.2.4 Gamma Distribution*. Introduction to Probability, Statistics and Random Processes, 2022.

[12] S. ADJEMIAN, *PRIOR DISTRIBUTIONS IN DYNARE*. École normale supérieure, 2007.

9. Apéndice

9.1. Codigo con método de Monte-Carlo

```
# Entradas
simulaciones = 10000
dias_maximos = 5000

_lambda = 0
for i in range(len(datos2)):
    _lambda += datos2["Reclamo de no-Cobertura"][i]
_lambda = _lambda/len(datos2) #parametro de poisson

miu = 0
for i in range(len(datos)):
    miu += datos["Monto del siniestro"][i]
miu = miu/len(datos) #parametro exponencial

ganancia_clientes = (30500/365)*len(datos1) #ganancia diaria por todos los clientes
ganancia_clientes
```

```
capital_final = []
numero_ruinas = 0

plt.figure()
plt.title("Simulacion de Monte-Carlo")
plt.xlabel("Dias")
plt.ylabel("Capital")
plt.xlim([0, dias_maximos])

for i in range(simulaciones):
    dinero = [20000000]
    dias_trabajados = [0]
    dia_ruina = []
    while dias_trabajados[-1] < dias_maximos:
        n = int(poisson.rvs(_lambda))
        costos = sum([np.random.exponential(scale = miu) for r in range(n)])
        dinero.append(dinero[-1] + ganancia_clientes - costos)
        dias_trabajados.append(dias_trabajados[-1] + 1)
    for r in range(len(dinero)):
        if dinero[r] <= 0:
            numero_ruinas += 1
            dia_ruina.append(r)
            break


    capital_final.append(dinero[-1])
    plt.plot(dias_trabajados, dinero)

plt.show()
dia_ruina_cuando = sum(dia_ruina)/len(dia_ruina)

print("Numero de simulaciones que cayeron en ruina: ", numero_ruinas)
print("Promedio de tiempo t de ruina: ", dia_ruina_cuando)
```

```
valor_esperado = 1 / _lambda
print("La media es: ", valor_esperado)

valor_esperado_c = valor_esperado - valor_esperado / (len(datos2) - 1)
print("La media corregida es: ", valor_esperado_c)
```

```
""" Maximum likelihood 95% confidence"""
_epsylon = valor_esperado_c * 1.96/(len(datos2)**0.5)
print("Lambda con intervalo de confianza: {0:8f} +/- {1:8f}".format(valor_esperado_c, _epsylon))
print("Intervalos de confianza: ({0:8f}, {1:9f})".format(valor_esperado_c - _epsylon, valor_esperado_c + _epsylon))
print("El error es de: {0:8f}".format(_epsylon))
```

9.2. Codigo de Panjer Exponencial

```
capital_final = []
numero_ruinas = 0

plt.figure()
plt.title("Simulacion de Panjer Exponencial")
plt.xlabel("Dias")
plt.ylabel("Capital")
plt.xlim([0, 100])
a = 0
b = _lambda
k = 0


for i in range(simulaciones):
    dinero = [20000000]
    dias_trabajados = [0]
    dia_ruina = []
    pk = [np.exp(-_lambda)]
    while dias_trabajados[-1] < dias_maximos:
        k += 1
        n = int(expon.rvs(miu))
        costos = sum([np.random.exponential(scale = miu) for r in range(n)])
        pk.append((a+(b/k))*pk[-1])
        dinero.append(dinero[-1] + ganancia_clientes - costos)
        dias_trabajados.append(dias_trabajados[-1] + 1)
    for r in range(len(pk)):
        if dinero[r] <= 0:
            numero_ruinas += 1
            dia_ruina.append(r)
            break


    capital_final.append(dinero[-1])
    plt.plot(dias_trabajados, pk)

plt.show()
```

9.3. Codigo de Panjer Gamma

```
capital_final = []
numero_ruinas = 0

plt.figure()
plt.title("Simulacion de Panjer Gamma")
plt.xlabel("Dias")
plt.ylabel("Capital")
plt.xlim([0, 100])
a = 0
b = _lambda
k = 0

for i in range(simulaciones):
    dinero = [20000000]
    dias_trabajados = [0]
    dia_ruina = []
    pk = [np.exp(-_lambda)]
    while dias_trabajados[-1] < dias_maximos:
        k += 1
        n = int(gamma.rvs(_lambda))
        costos = sum([np.random.exponential(scale = miu) for r in range(n)])
        pk.append((a+(b/k))*pk[-1])
        dinero.append(dinero[-1] + ganancia_clientes - costos)
        dias_trabajados.append(dias_trabajados[-1] + 1)
    for r in range(len(pk)):
        if dinero[r] <= 0:
            numero_ruinas += 1
            dia_ruina.append(r)
            break


    capital_final.append(dinero[-1])
    plt.plot(dias_trabajados, pk)

plt.show()
```