UNIVERSIDAD NACIONAL AUTÓNOMA DE NICARAGUA

UNAN – LEÓN

FACULTAD DE CIENCIAS Y TECNOLOGÍA

DEPARTAMENTO DE MATEMÁTICA Y ESTADÍSTICA

MONOGRAFÍA PARA OPTAR AL TÍTULO DE LICENCIADO EN CIENCIAS ACTUARIALES Y FINANCIERAS:

"ESTIMACIÓN DE LAS FRECUENCIAS TEÓRICAS OBSERVADAS UTILIZANDO LOS MODELOS DE DISTRIBUCIÓN DE PROBABILIDAD DISCRETAS Y LA PROGRAMACIÓN EN R."

AUTOR:

- **BR. MARTIN RAMÓN ALONSO MATAMOROS.**

TUTOR:

LIC. ÁLVARO ARÁUZ.

LEÓN, 10 DE OBTUBRE DEL 2018

"A LA LIBERTAD POR LA UNIVERSIDAD"

DEDICATORIA

A Dios:

Por darme la vida, el propósito y las fuerzas para siempre seguir adelante y cumplir con cada una de mis metas por lo cual fui creado. Él siempre ha sido la Luz que me guía y direcciona en todo momento de mi vida; y en Él siempre he confiado. A Él sea toda la Honra, la Gloria y el Poder.

Jesús, la luz del mundo:
Otra vez Jesús les habló, diciendo: Yo soy la luz del mundo; el que me sigue, no andará en tinieblas, sino que tendrá la luz de la vida.
Juan 8:12

A mis padres:

Por enseñarme, aconsejarme y guiarme en el camino correcto en el transcurso de mi vida, siempre ayudándome y apoyándome en cada uno de mis objetivos. Doy siempre gracias a Dios porque ellos han sido un gran ejemplo para mi vida. Ellos se han sacrificado mucho para que yo tenga una buena educación y que sea una persona de excelencia y buenos modales.

AGRADECIMIENTOS:

Tengo el deseo de mencionar a las personas que hicieron posible llevar a cabo este trabajo monográfico:

A Dios:

Por haber estado conmigo en todo momento, dándome sabiduría y entendimiento para poder terminar con éxitos mis estudios universitarios. En todo momento sentí su amor, su gracia y su favor que fue vital para mi vida.

A mis padres:

Por haber creído en mí, siempre apoyándome con su tiempo, sus recursos y su experiencia adquirida a lo largo de sus vidas. Que Dios siempre los guie, los guarde y les dé muchos años más de vida.

Al Lic. Álvaro Aráuz:

Quien fue mi tutor de este estudio monográfico; gracias por su tiempo y esmero que dedico para ayudarme y que esto llegara a ser posible. Usted es un profesor excelente en dar clases y que también nos ayuda a conocer más a Dios.

A mis maestros:

De la carrera de Ciencias Actuariales y financieras por haberme impartido todo su conocimiento día a día en las aulas de clases.

ÍNDICE

I. INTRODUCCIÓN

Las ciencias actuariales tienen por objeto el análisis de los modelos estocásticos que expliquen los fenómenos aleatorios denominados actuariales, que se enfrentan a una realidad que necesita de soluciones antes problemas suscitado en relación con los fenómenos que la integran, siniestralidad, mortalidad, ruina y supervivencia.

El uso de programas computarizados de una compañía de seguros, juega un rol muy importante ya que estos programas de computación dan un análisis completo y automatizado para conocer en su plenitud el riesgo en el que se están afrontando, permitiendo así tomar la mejor decisión en el proceso de tarificación ya que la prima debe cumplir con los requisitos de suficiencia, de no ser onerosa, estable y flexibles para poder ser eficientes y competentes en el mercado asegurador y así tener ganancias significativas para la empresa aseguradora.

La gran mayoría de las computadoras cuentan con el sistema operativo de Microsoft que utiliza imágenes y símbolos. Este sistema contiene varias aplicaciones informáticas en Microsoft office, como Word, Excel, Power Point, Access, entre otras. Siendo el Excel la aplicación más usada por los actuarios ya que este es utilizado para cálculos actuariales y financieros, así como manejo de grandes volúmenes de datos numéricos como clientes, proveedores, pasivos, nominas, inventarios, etc.

Es necesario que las compañías aseguradoras actualicen su sistema de cálculos y estimaciones, para no obtener perdidas numerosas y tener ganancias significativas. También estableciendo precios justos e indemnizaciones justas para que ambas partes tanto, así como la compañía aseguradora como la persona asegurada tengan partes equivalentes.

El propósito de este tema es realizar una estimación de las frecuencias teóricas a las observadas de una cartera de seguros mediante modelos de distribución de probabilidad discreta y del programa R. Además, se realizará una prueba de bondad de ajuste para conocer cual distribución de probabilidad discreta se adapta más a las frecuencias observadas. Así que se ha considerado transcendente realizar dicha investigación con el propósito de estimar a través del programa R y de los modelos de distribuciones de probabilidad discreta frecuencia teóricas que se ajusten a las frecuencias observadas de una cartera de seguros. El aporte de esta investigación es doble, es decir, está el aporte teórico y práctico acerca de estos modelos de distribución de probabilidad discreta y el uso de la programación en R; lo cual puede orientar a nuevos estudios actuariales como el estudio de tablas de mortalidad en R.

La investigación es viable, pues se dispone de la información necesaria de la cartera de seguro, así, como también de los recursos necesarios para llevarla a cabo. Cabe destacar que no se han hecho trabajos de monografía con modelos de distribución de probabilidad discretos que estimen el número de siniestros utilizando R; pero si existen trabajos con modelos de distribución de probabilidad continúas realizado por la Bra: Torres Corea Dalila mercedes, Vilchez Narváez Karen Walquiria cuyo objetivo era "*Aplicar la teoría de los modelos de distribución continuas que modelizan las cuantías de los siniestros al estudios de las situaciones más habituales en las ciencias actuariales y financieras utilizando la programación en R".*

Los aspectos más importantes a destacar en cada una de las partes que integran este trabajo son las siguientes:

La historia, definición y conceptos básicos de la probabilidad, características y propiedades de las modelos de distribución de probabilidad discreta que estiman el número de siniestros de una cartera de seguros son descritos y estudiados en el capítulo I.

En el capítulo II se presentan los conceptos básicos, descripciones, ambientes, características, propiedades y funcionamientos del programa R.

Por último, estimamos las frecuencias teóricas de los siniestros para las 5 carteras de seguros utilizando los modelos de probabilidad discretos y el programa en R. Además, la realización de la prueba de bondad de ajuste a los datos estimados con la intención de analizar qué distribución se adapta más a las frecuencias observadas de las distintas carteras de seguros.

II. OBJETIVOS

1. Objetivo General:

- Aplicar los modelos de distribución de probabilidades discretas y la programación en R, que estime el número de siniestros de una cartera de seguro.

2. Objetivos Específicos:

- Identificar los elementos matemáticos que caracterizan a las distribuciones de probabilidades discretas que son habituales en el ámbito de las empresas aseguradoras.

- Describir aspectos generales del programa R: Instalación, lenguaje, estructura, funciones, paquetes, etc.

- Describir y analizar situaciones sencillas de riesgos en problemas relacionados con el mundo actuarial haciendo uso de variables aleatorias discretas que modelizan el número de siniestros.

- Resolver problemas de casos reales para el tratamiento de riesgo, mediante fórmulas matemáticas basadas en los modelos de distribución de probabilidad discretas y la programación en R.

- Realizar una prueba de bondad de ajuste para saber qué modelo de distribución de probabilidad discreta se adapta mejor a los siniestros reales observados.

III. MARCO TEÓRICO

1. CAPITULO I: DISTRIBUCIONES DISCRETAS.

1.1. Historia y relevancia de la teoría de la probabilidad

Jacob Bernoulli (1654-1705), Abraham de Moivre (1667-1754), el reverendo Thomas Bayes (1702- 1761) y Joseph Lagrange (1736-1813) desarrollaron fórmulas y técnicas para el cálculo de la probabilidad. En el siglo XIX, Pierre Simon, marqués de Laplace (1749-1827), unificó todas estas ideas y compiló la primera teoría general de probabilidad.

La teoría de la probabilidad fue aplicada con éxito en las mesas de juego y, lo que es más importante en nuestro estudio, a problemas sociales y económicos. La industria de seguros, que surgió en el siglo XIX, requería un conocimiento preciso acerca de los riesgos de pérdida, con el fin de calcular las primas. Medio siglo más tarde, muchos centros de aprendizaje estaban estudiando la probabilidad como una herramienta para el entendimiento de los fenómenos sociales. En la actualidad, la teoría matemática de la probabilidad es la base para las aplicaciones estadísticas, tanto en investigaciones sociales como en la toma de decisiones.

La probabilidad constituye parte importante de nuestra vida cotidiana. En la toma de decisiones personales y administrativas, nos enfrentamos a la incertidumbre y utilizamos la teoría de la probabilidad, admitamos o no el uso de algo tan complejo. Cuando escuchamos una predicción de un 70% de posibilidades de lluvia, cambiamos nuestros planes de salir de día de campo y nos quedamos en casa divirtiéndonos con juegos de mesa. Cuando jugamos al bridge, hacemos algunas estimaciones de probabilidad antes de intentar una jugada arriesgada. Los administradores que se encargan de inventarios de ropa de moda para mujer deben preguntarse sobre las posibilidades de que las ventas alcancen o excedan un cierto nivel, y el comprador que adquiere una patineta considera la probabilidad de duración de su pasajero capricho.

Vivimos en un mundo incapaz de predecir el futuro con total certidumbre. Nuestra necesidad de encarar a la incertidumbre nos lleva a estudiar y utilizar la teoría de la probabilidad. En muchos casos, nosotros, como ciudadanos preocupados, tendremos algún conocimiento sobre los posibles resultados de una decisión. Al organizar esta información y considerarla de manera sistemática seremos capaces de reconocer nuestras suposiciones, comunicar nuestro razonamiento a otras personas y tomar una decisión más sólida que la que tomaríamos si sólo diéramos palos de ciego.

1.2. Probabilidad:

La probabilidad es una medida sobre la escala 0 a 1; correspondiendo el valor cero al suceso imposible, o sea el que no ocurre nunca, y el valor 1 suceso seguro. Para los restantes sucesos, daremos una probabilidad comprendida entre 0 y 1, de tal manera que será tanto más probable que ocurra un suceso cuanto mayor sea su probabilidad. Así pues, frecuentemente decimos que el hecho de que ocurra un accidente de automóvil es más probable en ciertas épocas del año que en otras.

Hechas estas indicaciones, nos surge la necesidad de dar un concepto de probabilidad, de tal forma que podamos asignar probabilidades a los diferentes sucesos de un experimento aleatorio. Este concepto de probabilidad no será único, ya que se pueden considerar diferentes enfoques o puntos de vista, así pues aquí expondremos el punto de vista **objetivo** y **subjetivo.** Dentro del enfoque objetivo se puede considerar una definición clásica o a priori de la probabilidad y otra frecuentista a posteriori.

1.2.1. Probabilidad Objetiva:

1.2.1.1. Definición Clásica de la probabilidad:

Consideremos un experimento aleatorio, cuyo correspondiente espacio muestral *E* está formado por un numero n, finito, de posibles resultados distintos y con la misma posibilidad de ocurrir $(e_1, e_2, e_3, \dots, e_n)$. Entonces si n_1 resultados constituyen el subconjunto o suceso A_1, n_1 resultados constituyen el suceso A_2, … y n_k resultados constituyen el suceso A_k de tal manera que:

$$n_1 + n_2 + \cdots + n_k = n$$

Y las probabilidades de los sucesos $A_1, A_2, \dots, A_k$ serán:

$$P(A_1) = \frac{n_1}{n}\ ; P(A_2) = \frac{n_2}{n}\ ; \dots ;\ (A_k) = \frac{n_k}{n}$$

Es decir, la probabilidad de cualquier suceso *A* es igual al cociente entre el número de resultados favorables o resultados que integran el suceso *A* y el número total de elementos posibles resultados del espacio muestral *E.* Luego una fórmula para calcular la probabilidad de un suceso cuando todos los posibles resultados tienen la misma probabilidad de ocurrir será:

$$P(A) = \frac{Número\ de\ casos\ favorables\ de\ A}{Número\ de\ casos\ posibles\ E}$$

Que se conoce con el nombre de regla Laplace para espacios muéstrales finitos.

No siempre resulta tan fácil y directa la aplicación de la regla Laplace, pues los sucesos del espacio maestral deben ser distintos y tener todos ellos la misma probabilidad de ocurrir. Para que se pueda aplicar la regla Laplace es necesario que todos los sucesos elementales sean equi-probables.

Observemos que la probabilidad verifica las siguientes condiciones:

- La probabilidad de cualquier suceso es siempre un número no negativo comprendido entre 0 y 1. En efecto, dicha probabilidad viene dada por $\frac{n_i}{n}$ en donde n_i es menor que n, y ambos son no negativos.
- La probabilidad del suceso seguro, *E,* vale 1, pues en este caso n_i será igual a n, ya que el suceso seguro *E* o espacio muestral contiene todos los posibles resultados y la probabilidad será $\frac{n}{n} = 1$. Análogamente la probabilidad del suceso imposible, ∅, es cero.
- La probabilidad de la unión de varios sucesos incompatibles o excluyentes es igual a la suma de las probabilidades de cada uno de los sucesos. En efecto, sean los sucesos $A_1, A_2, \dots, A_r$, compuestos cada uno de ellos por $n_1, n_2, \dots, n_r$ resultados elementales del espacio muestral *E* y tales que $A_i \cap A_j = \emptyset \; \forall \; i, j = 1, \dots, r$; entonces resulta que como
$$\frac{n_1 + n_2 + \cdots + n_r}{n} = \frac{n_1}{n} + \frac{n_2}{n} + \cdots + \frac{n_r}{n}$$
Tendremos
$$P(A_1 \cup A_2 \cup \dots \cup A_r) = P(A_1) + P(A_2) + \cdots + P(A_r)$$

Esta definición de la probabilidad clásica fue una de las primeras que se dieron, alrededor del año 1900, y se conoce con el nombre de la regla de Laplace ya que se le atribuye a él. También se le suele llamar probabilidad a priori, pues para calcularla es necesario conocer, antes de realizar el experimento aleatorio, el correspondiente espacio muestral y el número de resultados o sucesos elementales que entran a formar parte del suceso cuya probabilidad pretendemos determinar; pudiendo calcular la probabilidad de cualquier suceso antes de realizar el experimento aleatorio.

La aplicación de la definición clásica de la probabilidad puede presentar dificultades de aplicación en algunos casos. Concretamente, cuando el espacio muestral es infinito, o bien cuando los posibles resultados de un experimento no son igualmente probables.

Para resolver estos problemas se hace una extensión de la definición de la probabilidad, de manera que se pueda aplicar con menos restricciones, llegando a la definición de frecuentista de la probabilidad.

Ejemplo

Supongamos que si realizamos un experimento aleatorio que consiste en lanzar al aire dos monedas simultáneamente y estamos interesados en conocer la probabilidad de que aparezcan dos cruces. Para ello empezaríamos por obtener los posibles resultados que se pueden presentar al lanzar las dos monedas al aire que sería:

- Dos caras.
- Dos cruces.
- Una cara y una cruz.

Y como uno de estos resultados es el suceso, dos cruces, cuya probabilidad nos interesa, se podría decir que la probabilidad buscada es de 1/3. Pero esto no es cierto ya que los tres resultados no tienen la misma posibilidad de ocurrir, pues el resultado cara-cruz puede aparecer también como cruz-cara, siendo por tanto el espacio muestral o conjunto de posibles resultados,

$$E = \{HH, HT, TH, TT\}$$

Todos distintos con la misma posibilidad de ocurrir. La probabilidad correcta del suceso, dos cruces, seria ¼

1.2.1.2. Definición Frecuentista de la Probabilidad:

Dado dos sucesos incompatibles A_1 y A_2 tales que $A_1 \cup A_2 = A \subset E$ entonces cada vez que se presente el suceso A, se presentara necesariamente uno y solo uno de los sucesos A_1 o A_2, y si realizamos n repeticiones del experimento aparecerá n´ veces el suceso A de tal forma que

$$n\acute{} = n_1 + n_2$$

Siendo:

n_1: el numero de veces que aparece A_1 y

n_2: el numero de veces que aparece A_2

Luego tendremos que:

$$\frac{n\acute{}}{n} = \frac{n_1}{n} + \frac{n_2}{n} \rightarrow f_r(A) = f_r(A_1 \cup A_2) = f_r(A_1) + f_r(A_2)$$

Esta propiedad se puede generalizar al caso de n sucesos incompatibles.

La teoría frecuentista de la probabilidad asegura que existe el siguiente límite cuando n tiende al infinito:

$$\lim_{n\to\infty} \frac{n_i}{n} = P(A_i), \quad i = 1,2,3 \dots, k$$

Siendo $P(A_i)$ la probabilidad del suceso A_i

Luego la definición frecuentista de la probabilidad consiste en definir la probabilidad como el límite cuando n tiende al infinito de la proporción o frecuencia relativa del suceso.

En general, si realizamos un experimento aleatorio cuya correspondiente espacio muestral es E y designamos por A cualquier suceso perteneciente al espacio muestral E y repetimos en las mismas condiciones, n veces el experimento aleatorio, tendremos que la frecuencia relativa del suceso A será:

$$\frac{n(A)}{n}$$

En donde n(A) es el número de veces que ha aparecido el suceso A en las n repeticiones del experimento.

Cuando el número n de repeticiones del experimento se hace muy grande, o sea, cuando n tiende a infinito, la frecuencia relativa converge hacia un valor que llamaremos probabilidad del suceso A, P(A), o sea:

$$P(A) = \lim_{n\to\infty} \frac{n(A)}{n}$$

Pero como es imposible llegar a este límite, ya que no podemos repetir el experimento un número infinito de veces, lo que si podemos hacer es repetir el experimento muchas veces y observaríamos que las frecuencias relativas tienden a estabilizarse. Pero esta estabilización es relativa, pues en algunos casos las frecuencias relativas tienden a estabilizarse muy pronto, es decir con pocas repeticiones del experimento aleatorio se observa la estabilización sin embargo en otros casos la estabilización de las frecuencias relativas es más lenta, teniendo que repetir muchas veces el experimento aleatorio para que aparezca esa estabilización. Así pues, consideramos aleatorio que consiste en lanzar una moneda equilibrada al aire, o sea, una moneda muy perfecta, siendo los posibles resultados cara (H) o cruz (T); si repetimos 200 veces el experimento obtenemos los resultados siguientes:

Número de lanzamientos	Número de caras cada 10 lanzamientos	Suma acumulada de caras	Frecuencia relativas de caras
1	0	0	0
10	6	6	0.6
20	2	8	0.4
30	6	14	0.467
40	5	19	0.475
50	6	25	0.500
60	6	31	0.517
70	7	38	0.543
80	5	43	0.537
90	3	46	0.511
100	5	51	0.510
110	5	56	0.509
120	7	63	0.525
130	5	68	0.523
140	4	72	0.514
150	3	75	0.500
160	3	78	0.487
170	5	83	0.488
180	6	89	0.494
190	6	95	0.500
200	6	101	0.505

1.2.2. Interpretación subjetiva de la probabilidad:

Hemos visto que la definición clásica no era posible aplicarla en todas las ocasiones. Por otra parte, la definición Frecuentista de la probabilidad se basa en repeticiones del experimento aleatorio, pero hay muchos experimentos que no se pueden repetir bajo las mismas condiciones y, por tanto, tampoco se les podrá aplicar esta definición de la probabilidad. En consecuencia, habrá la probabilidad (clásica o Frecuentista), teniendo que recurrir a un punto de vista alternativo que consiste en considerar la probabilidad como un concepto subjetivo, que expresa el grado de creencia o confianza individual sobre la posibilidad de que el suceso ocurra. Es decir, la probabilidad subjetiva representa un juicio personal sobre el resultado de un experimento aleatorio, pudiendo ser diferente del juicio personal o probabilidad subjetiva asignada por otra persona. Luego la probabilidad subjetiva es la evaluación personal de la probabilidad de un fenómeno aleatorio.

Con el fin de aclarar lo que entendemos por grado de creencia o confianza individual consideramos el siguiente ejemplo. Que va a consistir en un partido de futbol entre dos equipos A y B que juegan por primera vez, no disponemos de información sobre resultados anteriores puesto que es la primera vez que van a jugar, solo se tiene información sobre algunos jugadores de ambos equipos. Como no han jugado en ocasiones anteriores no podemos atribuirle probabilidad objetiva al posible resultado del partido, es decir como no existen resultados de partidos anteriores, no podemos asignarle probabilidad de tipo frecuentista o a posteriori ni de tipo clásico o a priori a los tres posibles resultados del partido:

- Que gane el equipo A,
- Que gane el equipo B, o
- Que empaten.

Sin embargo, basándonos en el conjunto de la posible información que podemos tener acerca de los diferentes jugadores o aun sin ella, puede emitir un juicio personal o grado de creencia sobre el posible resultado, que sería la probabilidad subjetiva que se asigna a cada posible resultado. Así pues, un determinado observador puede asignar las siguientes probabilidades subjetivas.

Probabilidad de que gane el equipo A = 0.7

Probabilidad de que gane el equipo B = 0.2

Probabilidad de que empaten = 0.1

Otro observador puede asignar diferentes probabilidades subjetivas:

Probabilidad de que gane el equipo A = 0.3

Probabilidad de que gane el equipo B = 0.2

Probabilidad de que empaten = 0.5

Siendo, por tanto, posible que diferentes observadores tengan diferentes grados de creencias sobre los posibles resultados emitiendo juicios personales o probabilidades subjetivas diferentes e igualmente válidas.

1.3. Variables aleatorias

Una variable es aleatoria si toma diferentes valores como resultado de un experimento aleatorio. Esta variable aleatoria puede ser discreta o continua. Si puede tomar sólo un número limitado de valores, entonces es una variable aleatoria discreta. En el otro extremo, si puede tomar cualquier valor dentro de un intervalo dado, entonces se trata de una variable aleatoria continua.

Una variable aleatoria es una especie de valor o magnitud que cambia de una ocurrencia a otra sin seguir una secuencia predecible. Por ejemplo, en una clínica para tratamiento del cáncer de mama no se tiene manera de saber con exactitud cuántas mujeres van a ser atendidas en un día cualquiera, de modo que el número de mujeres del día siguiente es una variable aleatoria. Los valores de una variable aleatoria son los valores numéricos correspondientes a cada posible resultado del experimento aleatorio. Si los registros diarios de la clínica indican que los valores de la variable aleatoria van desde 100 hasta 115 mujeres al día, entonces ésta es una variable aleatoria discreta.

1.3.1. Valor esperado de una variable aleatoria discreta:

1.3.1.1. Media

La media llamada también valor esperado, esperanza matemática o simplemente esperanza de una distribución de probabilidad discreta es la media aritmética ponderada de todos los resultados posibles en los cuales los pesos son las probabilidades respectivas de tales resultados. Se halla multiplicando cada resultado posible por su probabilidad y sumando los resultados. Se expresa mediante la siguiente fórmula:

$$\mu = E(x) = \sum (x_i * P(x_i))$$

Donde:

μ =Media, Valor Esperado, Esperanza Matemática o simplemente Esperanza

x_i= Posible resultado

$P(x_i)$ = Probabilidad del posible resultado

1.3.1.2. Varianza y desviación estándar de una variable aleatoria discreta:

La varianza es el promedio de las desviaciones al cuadrado con respecto a la media. La varianza mide la dispersión de los resultados alrededor de la media y se halla calculando las diferencias entre cada uno de los resultados y su media, luego tales diferencias se elevan al cuadrado y se multiplican por sus respectivas probabilidades, y finalmente se suman los resultados. Se expresa mediante la siguiente fórmula:

$$\sigma^2 = \sum (x_i - \mu)^2 * P(x_i)$$

Nota: La varianza se expresa en unidades al cuadrado, por lo que es necesario calcular la desviación estándar que se expresa en las mismas unidades que la variable aleatoria y que por lo tanto tiene una interpretación más lógica de la dispersión de los resultados alrededor de la media. La desviación estándar se calcula así: $\sigma = \sqrt{\sigma^2}$

1.4. Distribución de probabilidad:

La distribución de probabilidad de una variable aleatoria es una función que asigna a cada suceso definido sobre la variable la probabilidad de que dicho suceso ocurra. La distribución de probabilidad está definida sobre el conjunto de todos los sucesos y cada uno de los sucesos es el rango de valores de la variable aleatoria. También puede decirse que tiene una relación estrecha con las distribuciones de frecuencia. De hecho, una distribución de probabilidades puede comprenderse como una frecuencia teórica, ya que describe cómo se espera que varíen los resultados.

La distribución de probabilidad está completamente especificada por la función de distribución, cuyo valor en cada x real es la probabilidad de que la variable aleatoria sea menor o igual que x.

Las distribuciones de probabilidad son distribuciones de probabilidad continuas o distribuciones de probabilidad discretas, dependiendo de si definen probabilidades para variables continuas o discretas.

Distribuciones de Probabilidad

- **Distribuciones Discretas:**
  - Distribución degenerada en un punto.
  - Distribución uniforme discreta.
  - Distribución de Bernoulli.
  - Distribución geométrica.
  - Distribución binomial
  - Distribución binomial negativa.
  - Distribución de Poisson
  - Distribución Inversa Gaussiana
- **Distribuciones Continuas:**
  - Distribución uniforme continua.
  - Distribución normal.
  - Teorema Central del limite.
  - Distribución lagoritmo-normal
  - Distribución gamma
  - Distribución beta.
  - Distribución de Pareto
  - Distribución normal bivariante

1.5. Modelos de distribución de probabilidad discreta que estiman el número de siniestros:

1.5.1. Proceso de Bernoulli:

Sea un experimento aleatorio que da lugar, únicamente a dos posibles resultados que son mutuamente excluyentes y exhaustivos. Los dos posibles resultados los designaremos por:

- Suceso éxito A, que era el suceso en estudio, y
- Suceso fracaso, que es el complementario, Ā

Y son tales que

$$A, \bar{A} \subset E$$

$$A \cup \bar{A} = E$$

$$A \cap \bar{A} = \emptyset$$

A este tipo de experimentos aleatorios se les llama experimentos o pruebas de Bernoulli.

Asociado a este experimento o prueba de Bernoulli tenemos la variable aleatoria de Bernoulli *X* definida de tal manera que:

$$X = \begin{Bmatrix} 0 \; si \; ocurre \; el \; suceso \; \bar{A} \\ 1 \; si \; ocurre \; el \; suceso \; A \end{Bmatrix}$$

Si designamos por *p* la probabilidad del suceso éxito A, la probabilidad del suceso fracaso Ā será 1 – *p,* es decir:

$$P(A) = p, \qquad 0 \leq p \leq 1$$

$$P(\bar{A}) = 1 - p = q$$

La función de probabilidad de la variable X se puede expresar como:

$$P(X = 0) = P(\bar{A}) = 1 - p = q$$

$$P(X = 1) = P(A) = p$$

Diremos que una variable aleatoria X sigue una distribución de Bernoulli de parámetro *p* si su función de probabilidad es:

$$P(x) = P(X = x) = p^x.(1-p)^{1-x}, x = 0,1$$

Condiciones:

1. El experimento consta de ensayos repetidos.

2. Cada ensayo produce un resultado que se puede clasificar como éxito o fracaso.

3. La probabilidad de un éxito, que se denota con *p*, permanece constante de un ensayo a otro.

4. Los ensayos repetidos son independientes.

Parámetros:

p = probabilidad de éxito.

q = probabilidad de fracaso.

Significado de los términos:

p = probabilidad de éxito.

q = probabilidad de fracaso.

x = número de éxitos que se desean tener

Característica:

- **Función de Distribución**

$$F(x) = P(X \leq x) = \begin{Bmatrix} 0 & , & x < 0 \\ 1-p & , & 0 \leq x < 1 \\ 1 & , & x \geq 1 \end{Bmatrix}$$

- **Media**

$$E[X] = p$$

- **Varianza**

$$Var(X) = p.q$$

- **Función generatriz de momentos**

$$g(t) = q + pe^t$$

Ejercicio:

Una agente de seguros dedicado a la venta de seguros de vida, realiza vistas a posibles clientes con el fin de contratar un seguro de vida. Se sabe su trayectoria como agente que en el 60% de las visitas tienen éxito y contrata un seguro. Definir la variable aleatoria asociada a este experimento aleatorio y obtener la media y la varianza.

Solución: El experimento aleatorio o prueba de Bernoulli será el realizar la visita a un posible cliente, ya que los dos posibles resultados serán:

- Conseguir que contrate el seguro, Suceso éxito A.
- No conseguir que contrate el seguro, suceso fracaso Ā.

Luego la variable aleatoria X asociada al experimento toma los valores:

$$X = 1 \quad si \quad hace\ el\ seguro, ocurre\ el\ suceso\ A$$

$$X = 0 \quad si \quad no\ hace\ el\ seguro, ocurre\ el\ suceso\ \bar{A}$$

Su función de probabilidad será:

$$P(X = 1) = P(A) = 0.6 = P$$

$$P = (X = 0) = P(\bar{A}) = 0.4 = 1 - p = q$$

La media y la varianza son:

$$E[X] = P = 0.6$$

$$Var = p\,.(1 - p) = (0.6)(0.4) = 0.24$$

1.5.2. Distribución Binomial:

Una generalización importante de la distribución de Bernoulli se obtiene cuando el experimento o prueba de Bernoulli se repite varias veces. Así pues consideramos n repeticiones independientes de una prueba de Bernoulli, siendo la probabilidad de éxito, $P(A) = p$, constante en todas las repeticiones. Asociada a cada una de las n repeticiones del experimento o prueba de Bernoulli tenemos una variable aleatoria X_i, tal que

$$X_i = \begin{Bmatrix} 1 \ \ cuando\ en\ la\ repetición\ i\ ésima\ ocurre\ el\ suceso\ éxito, A \\ 0\ cuando\ en\ la\ repetición\ i\ ésima\ ocurre\ el\ suceso\ fracaso, \bar{A} \end{Bmatrix}$$

$$i = 1,2, \dots, n$$

Si definimos la variable aleatoria

$$X = X_1 + X_2 + \cdots + X_n = \sum_{i=1}^{n} X_i$$

Entonces como cada X_i puede tomar el valor 1 o cero dependiendo de si ocurre el suceso éxito o el fracaso en la repetición i ésima, la variable aleatoria X tomara valores enteros comprendidos entre 0 y n y nos indicara el número de éxitos en las n-repeticiones independientes del experimento o prueba de Bernoulli, y a estas variables aleatorias X se llama variable aleatoria binomial con parámetros n y p.

Definimos la variable aleatoria binomial X como el número de éxitos que tienen lugar cuando se realizan n-repeticiones independientes de un experimento o prueba de Bernoulli.

Para obtener la función de probabilidad de la variable binomial, tendremos que calcular la probabilidad de que la variable aleatoria X tome los diferentes valores x = 0, 1,…, n, es decir, la P(X=x) y para ello consideramos que al realizar las n repeticiones independientes del experimento o prueba de Bernoulli hemos obtenido en primer lugar x resultados éxitos, y los restantes n – x han sido fracasos, Ā. Si tenemos en cuenta que las n-repeticiones son independiente, la probabilidad de este suceso será

$$p^x.(1-p)^{n-x}$$

$$= p^x q^{n-x}$$

En donde q = 1 – p

Pero las posibles ordenaciones de los x-éxitos y los n – x fracasos son permutaciones n-arias con repetición de A y Ā, es decir

$$\frac{n!}{x!\,(n-x)!} = \binom{n}{x}$$

Y tendremos que multiplicar ese número de ordenaciones por la probabilidad del suceso anterior, para llegar a tener la **función de probabilidad**, es decir, será:

$$\binom{n}{x} p^x (1-p)^{n-x} = \binom{n}{x} p^x q^{n-x}$$

Diremos que una variable aleatoria X sigue una distribución binomial de parámetros n y p si su función de probabilidad es

$$P(x) = P(X = x) = \binom{n}{x} p^x.(1-p)^{n-x}, \quad x = 0,1,\dots,n$$

Para ver que esta función de probabilidad este bien definida todas las probabilidades tienen que ser mayores o iguales a cero, lo cual es evidente, además la suma de todas las probabilidades ha de ser igual a la unidad, en efecto, teniendo en cuenta el binomio de Newton

$$\sum_{x=0}^{n} P(X = x) = \sum_{x=0}^{n} \binom{n}{x} p^x.(1-p)^{n-x} = [p + (1-p]^n = 1$$

Condiciones:

1. La muestra se compone de un número fijo de observaciones n
2. Cada observación se clasifica en una de dos categorías, mutuamente excluyentes (los eventos no pueden ocurrir de manera simultánea. Ejemplo: Una persona no puede ser de ambos sexos) y colectivamente exhaustivos (uno de los eventos debe ocurrir).

Parámetros:

n = número de intentos que se realizan.

p = probabilidad de éxito.

Significado de los términos:

n = número de intentos que se realizan.

p = probabilidad de éxito.

q = probabilidad de fracaso.

x = número de éxitos que se desean tener.

Características

- **Función de distribución**

La expresión de la función e distribución de una B(n, p) es:

$$F(x) = P(X \leq x) = \left\{ \begin{array}{ll} 0, & x < 0 \\ \sum_{i \leq x} \binom{n}{i} p^i \, (1-p)^{n-i}, & 0 \leq x < n \\ 1, & x \geq n \end{array} \right\}$$

- **Media**

$$E[X] = n.p$$

- **Varianza**

$$Var(X) = n.p.q$$

- **Función generatriz de momentos**

$$gx^{(t)} = (q + pe^t)^n$$

- **Propiedad reproductiva:**

$$gx^t = (q + pe^t)^{n_1 + n_2}$$

- **Propiedad de simetría:**

$$P[X = x] = P[Y = n - x]$$

Ejercicio:

Dada una variable aleatoria X que se distribuye según una distribución binomial con parámetros n = 15 y p = 0.3. Obtener:

1. *P(X = 2)*
2. *P(X ≤ 2)*
3. *La media, varianza y función generatriz de momentos.*

Solución:

1. $P(X = 2)$
   $\binom{n}{x} p^x q^{n-x} = \binom{15}{2} (0.3^2)(0.7^{15-2}) = 0.0916$

2. $P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2)$
   $= 0.0047 + 0.0305 + 0.0916 = 0.1268$

3. *Media.*
   $E[X] = n.p$
   $= (15)(0.3) = 4.5$

   $Varianza.$
   $Var(X) = n.p.q$
   = (15)(0.3)(0.7) =3.15

   Función Generatriz de momentos:
   $gx^{(t)} = (q + pe^t)^n$
   $= (0.7 + 0.3e^t)^{15}$

1.5.3. Distribución Binomial Negativa:

Consideremos un experimento con las mismas propiedades de un experimento binomial, sólo que en este caso las pruebas se repetirán hasta que ocurra un número fijo de éxitos. Por lo tanto, en vez de encontrar la probabilidad de x éxitos en n pruebas, donde n es fija, ahora nos interesa la probabilidad de que ocurra el k-ésimo éxito en la x-ésima prueba. Los experimentos de este tipo se llaman experimentos binomiales negativos.

Función de probabilidad es:

$$P(X = x) = \binom{x + r - 1}{x} q^x . p^r \quad , \quad x = 0,1,2, \dots ; 0 < p \leq 1$$

Condiciones:

1. El proceso tiene un número indefinido de pruebas y este concluirá cuando se tenga un número determinado de resultados favorables *K*.
2. Las pruebas pueden dar dos resultados posibles y excluyentes a su vez, es decir, A y no A.
3. La probabilidad de obtener un resultado A en las pruebas es *p*, y la de conseguir no A es *q*, de forma que *p+q=1*.
4. *p* y *q* son constantes en cada prueba y a su vez estas son independientes.
5. Derivación de la distribución: si la variable aleatoria x es "el número de pruebas necesarias para conseguir K éxitos o resultados A " ; entonces x seguirá una distribución binomial negativa con parámetros *p* y *k.*

Parámetros:

r = número de intentos que se realizan.

p = probabilidad de éxito.

Significado de los términos:

r = número de intentos que se realizan.

p = probabilidad de éxito.

q = probabilidad de fracaso.

x = número de éxitos que se desean tener

Características:

- **Función de distribución:**

$$F(x) = P(X \leq x) = \left\{ \begin{matrix} \sum_{i=0}^{x} \binom{x+r-1}{i} q^x p^r & , & x \geq 0 \\ 0 & , & x < 0 \end{matrix} \right\}$$

- **Función Generatriz de momentos:**

$$g^t = \left(\frac{p}{1 - qe^t} \right)^r$$

- **Media:**

$$E[X] = \frac{rq}{p}$$

- **Varianza:**

$$Var(X) = \frac{rq}{p^2}$$

- **Propiedad reproductiva:**

$$gx^{(t)} \left(\frac{p}{1 - qe^t} \right)^{r_1 + r_2}$$

Ejercicio:

Un examen de estadística consta de 20 preguntas tipo test y se conoce de experiencias aleatorias que un alumno tiene probabilidad de 0.7 de contestar bien cada pregunta. Sabiendo que para aprobar el examen es necesario contestar bien 10 preguntas. ¿Cuál es la probabilidad de que aprueba al contestar la pregunta duodécima?

Solución:

X = número de preguntas que responde mal antes de que responda bien la r-ésima

Esta variable aleatoria sigue una distribución binomial negativa con r = 10 y p = 0.7

$$P(X = x) = \binom{x + r - 1}{x} q^x . p^r$$

$$P(X = 2) = \binom{2 + 10 - 1}{2} (0.3^2)(0.7^{10})$$

$$= \frac{11!}{2!\,9!}(0.09)(0.028) = 0.01398$$

1.5.4. Distribución de Poisson:

La distribución de Poisson debe su nombre a Siméon Denis Poisson (1781-1840), un francés que desarrolló la distribución a partir de los estudios que realizó durante la última parte de su vida.

La distribución de Poisson se utiliza para describir ciertos tipos de procesos, entre los que se encuentran la distribución de llamadas telefónicas que llegan a un conmutador, las solicitudes de pacientes que requieren servicio en una institución de salud, las llegadas de camiones y automóviles a una caseta de cobro, y el número de accidentes registrados en cierta intersección. Estos ejemplos tienen en común un elemento: pueden ser descritos mediante una variable aleatoria discreta que toma valores enteros (0, 1, 2, 3, 4, 5, etc). El número de pacientes que llegan al consultorio de un médico en un cierto intervalo será de 0, 1, 2, 3, 4, 5 o algún otro número entero. De manera parecida, si usted cuenta el número de automóviles que llegan a una caseta de cobro de alguna carretera durante un periodo de 10 minutos, el número será de 0, 1, 2, 3, 4, 5 y así consecutivamente.

La función de probabilidad es:

$$P(x) = P(X = x) = \frac{\lambda^x}{x!}\, e^{-\lambda} \quad , \quad x = 0,1 \ldots$$

Condiciones:

- El número de sucesos que ocurren en un intervalo de tiempo o en una región o en volumen especificado, son independientes del número de sucesos que ocurren en cualquier otro intervalo o región.
- La Probabilidad de que un solo suceso o resultado ocurra en un intervalo de tiempo muy pequeño o en una región muy pequeña, es proporcional a la longitud del intervalo de tiempo o al tamaño de la región, y no depende del número de sucesos que se produzcan fuera del intervalo o región considerada.
- La probabilidad de que ocurran más de un resultado o suceso en un intervalo de tiempo muy pequeño o en una región muy pequeña es prácticamente nula.

Parámetros:

λ : Número medio de veces que ocurre nuestro suceso en un determinado intervalo de tiempo.

Significado de los términos:

λ : Número medio de veces que ocurre nuestro suceso en un determinado intervalo de tiempo.

x: Probabilidad de que ocurra un determinado suceso un número de veces.

Características:

- **Función de Distribución:**

$$F(x) = P(X \leq x) = \begin{Bmatrix} 0 & , & x < 0 \\ \displaystyle\sum_{i=0}^{x} \frac{\lambda^i}{i!} e^{-\lambda} & , & x = 0,1,2 \ldots \end{Bmatrix}$$

- **Media:**

$$E[X] = \lambda$$

- **Varianza:**

$$Var(X) = \lambda$$

- **Función generatriz:**

$$g_x(t) = e^{\lambda(e^t-1)}$$

- **Función reproductiva:**

$$g_x(t) = e^{(\lambda_1+\lambda_2)(e^t-1)}$$

Ejercicio:

Durante un experimento de laboratorio el número promedio de partículas radiactivas que pasan a través de un contador en un milisegundo es 4. ¿Cuál es la probabilidad de que entren 6 partículas al contador en un milisegundo dado?

Solución: Al usar la distribución de Poisson con x = 6 y λt = 4

$$p(6;4) = \frac{e^{-4}4^6}{6!} = \sum_{x=0}^{6} p(x;4) - \sum_{x=0}^{5} p(x;4) = 0.8893 - 0.7851 = 0.1042$$

1.5.5. Distribución de poisson inversa gaussiana

Folks y Chhikara , hacen una breve reseña acerca de la distribución Inversa Gaussiana (IG) donde se remontan al año 1915 cuando Schrödinger da el primer paso en la construcción de esta distribución y mencionan la evolución que tuvo está en los años 1945 y 1947, en los que Tweedie y Wald, respectivamente, aportaron avances para llegar a lo que hoy conocemos como la distribución Inversa Gaussiana. La forma de la distribución está dada por:

$$g(\lambda; \mu, \sigma) = \left(\frac{\sigma}{2\pi\lambda^3}\right)^{1/2} Exp\left\{-\frac{\sigma}{2\mu^2\lambda}.(x-\mu)^2\right\} \quad , \lambda, \mu, \sigma > 0$$

Donde μ y σ son constantes positivas.

Si la variable aleatoria X se distribuye Inversa Gaussiana, entonces se puede expresar que X ∼ IG(μ, σ). Esta función de densidad pertenece a la familia exponencial y tiene las características de ser unimodal y sesgada positivamente, además de la particularidad que cuando μ = 1 la distribución obtenida es llamada la distribución de Wald. Como en la distribución Gamma, la suma de variables que se distribuyen Inversas Gaussianas es Inversa Gaussiana.

La media y la varianza de la distribución Inversa Gaussiana están dados por:

Media:

$$E[X] = \mu$$

Varianza:

$$V\,ar[X] = \frac{\mu^3}{\sigma}$$

Parámetros:

Los parámetros μ y σ son interpretados como los parámetros de localización y escala, respectivamente.

Seshadri, presenta características particulares de la distribución Inversa Gaussiana y algunas propiedades que entran más al detalle de esta distribución.

La función generadora de momentos de la distribución IG está definida como:

$$Mx(t) = \exp[\_\mu^{-1}\{1 - (1 - 2t\mu^2\sigma^{-1})^{1/2}\}]$$

2. CAPITULO II: ASPECTOS GENERALES DE LA PROGRAMACION EN R

2.1. Breve historia de programa R.

Fue desarrollado inicialmente por Robert Gentleman y Ross Ihaka del Departamento de Estadística de la Universidad de Auckland en 1993. Sin embargo, si se remonta a sus bases iniciales, puede decirse que inició en los Bell Laboratories de AT&T y ahora Alcatel-Lucent en Nueva Jersey con el lenguaje S. Este último, un sistema para el análisis de datos desarrollado por John Chambers, Rick Becker, y colaboradores diferentes desde finales de 1970. La historia desde este punto es prácticamente la del lenguaje S. Los diseñadores iniciales, Gentleman y Ihaka, combinaron las fortalezas de dos lenguajes existentes, S y Scheme. En sus propias palabras: "El lenguaje resultante es muy similar en apariencia a S, pero en el uso de fondo y la semántica es derivado desde Scheme". El resultado se llamó R "en parte al reconocimiento de la influencia de S y en parte para hacer gala de sus propios logros".

2.2. Instalación de R:

2.2.1. Pasos para descargar R:

Lo primero que tenemos que hacer es descargar el programa. En la página web http://www.r-project.org, seleccionamos en la parte izquierda la opción Download, Packages pinchando en CRAN

Accedemos entonces a una ventana en la que nos aparecen diferentes lugares desde los que descargar el programa. Podemos seleccionar, por ejemplo, la página de nuestro país (aunque si no funciona del todo bien o está sobrecargada podemos acceder a cualquier otro lugar para descargar el programa)

Spain
http://cran.es.r-project.org/ Spanish National Research Network, Madrid

Seleccionamos a continuación nuestro sistema operativo

Download and Install R

Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:

- Linux
- MacOS X
- Windows

Y a continuación seleccionamos el programa base con la opción:

R for Windows

This directory contains binaries for a base distribution and packages to run on i386/x64 Windows.

Note: CRAN does not have Windows systems and cannot check these binaries for viruses. Use the normal precautions with downloaded executables.

Subdirectories:

base	Binaries for base distribution (managed by Duncan Murdoch)
contrib	Binaries of contributed packages (managed by Uwe Ligges)

Por último pinchamos sobre la opción "Download R…" para que comience la descarga del programa:

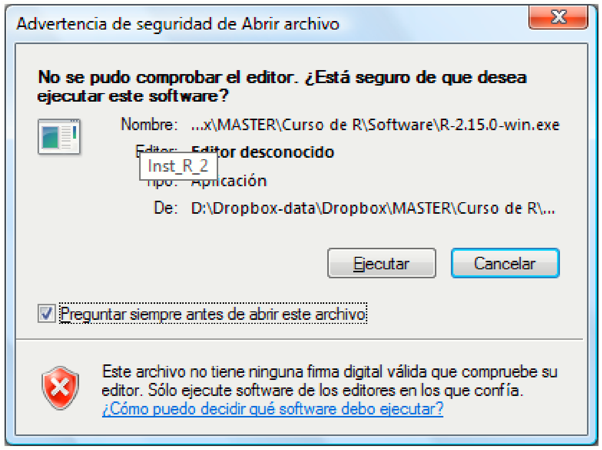
R-3.0.2 for Windows (32/64 bit)

Download R 3.0.2 for Windows (52 megabytes, 32/64 bit)

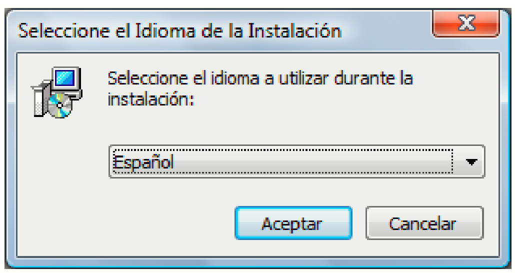
Installation and other instructions
New features in this version

Haciendo doble clic sobre el fichero comenzamos la instalación. Saldrá el típico aviso de Windows de que no se puede comprobar el editor, no le hacemos caso y pulsamos en ejecutar. A veces también sale otra ventana indicando que hay que permitir la ejecución del paquete ya que requiere privilegios de administrador.

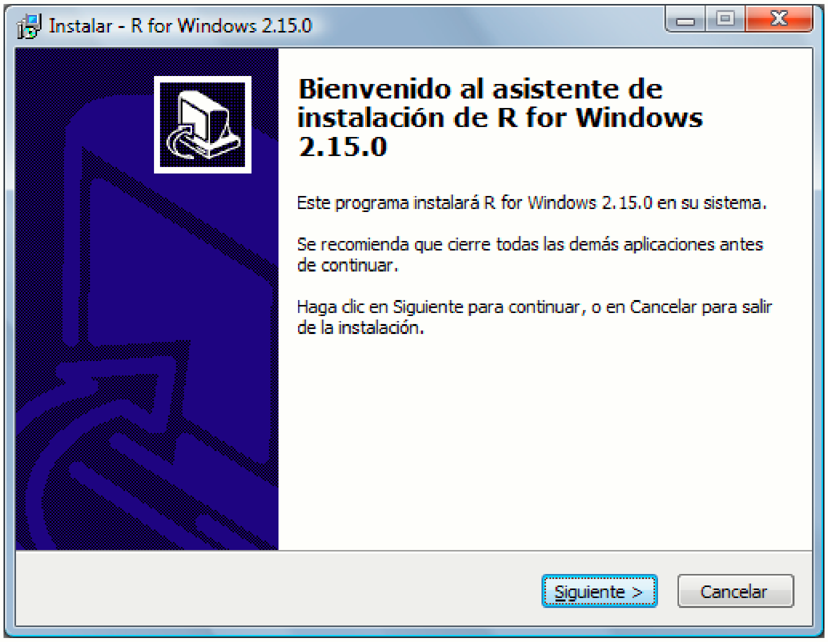
2.2.2. Pasos para Instalar R:



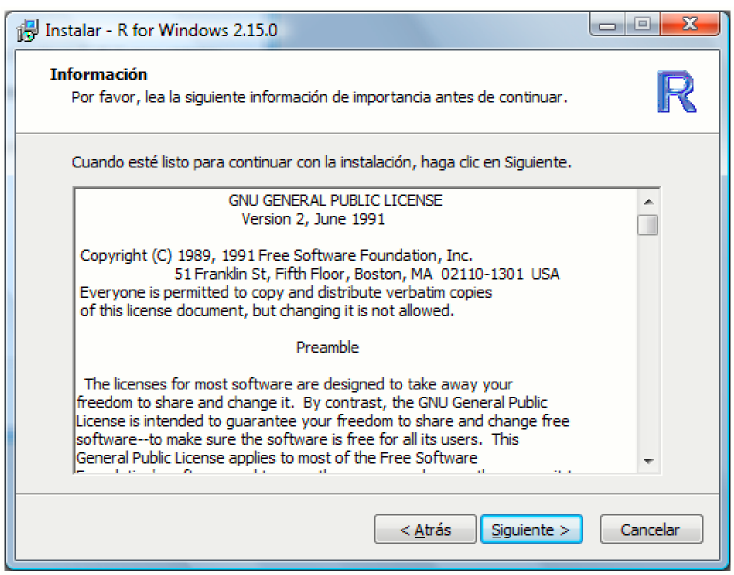
Después solicita el idioma de instalación



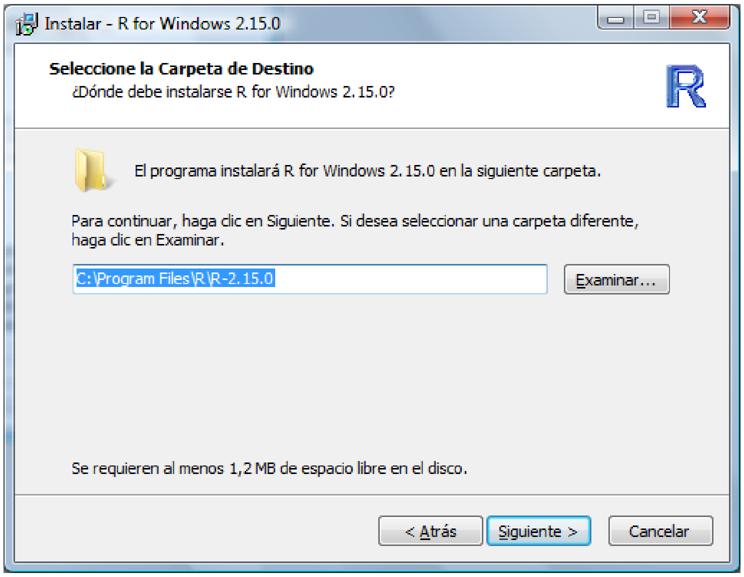
A continuación, arranca el Asistente de instalación, pulsar siguiente.



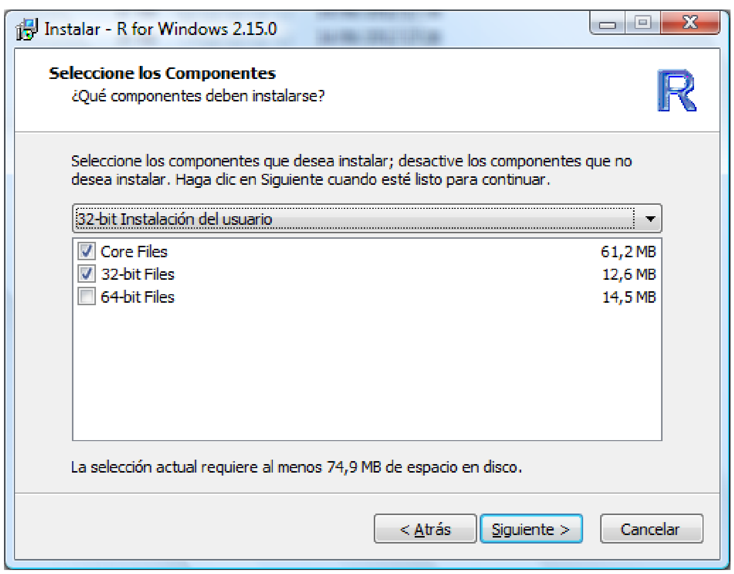
Después informa sobre el tipo de licencia, pulsar siguiente.



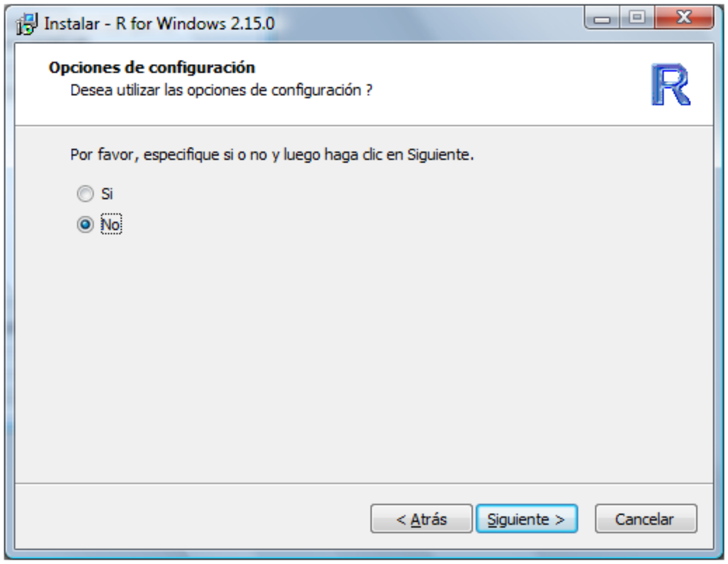
A continuación, nos indica la ruta de instalación, pulsar siguiente.



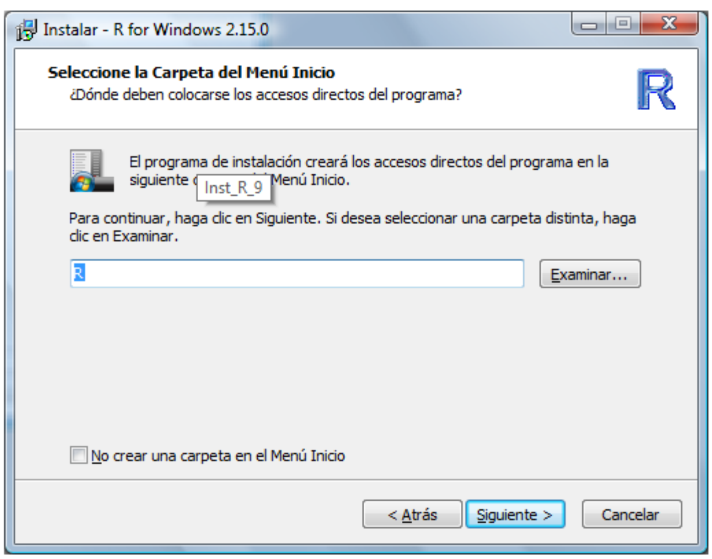
Ahora seleccionamos los paquetes a instalar, pulsar siguiente.



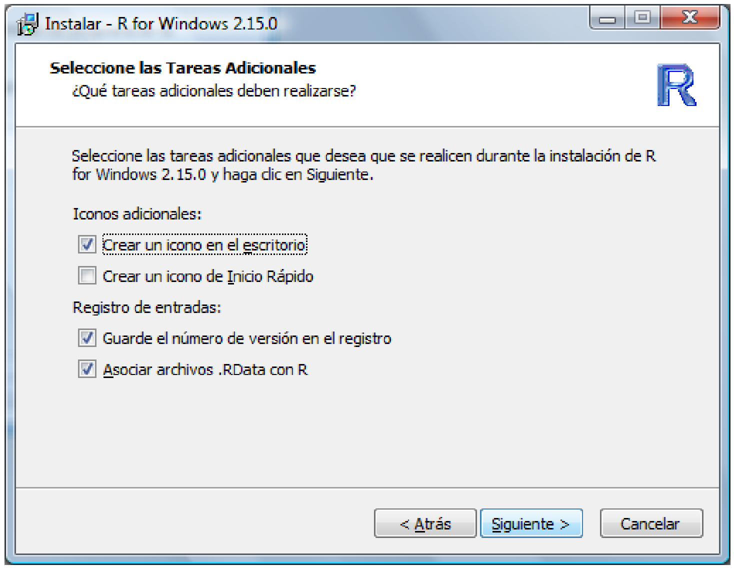
Después nos pregunta si queremos usar las opciones de configuración o no, por defecto aparece no y aquí tampoco nos vamos a complicar demasiado, pulsar siguiente.



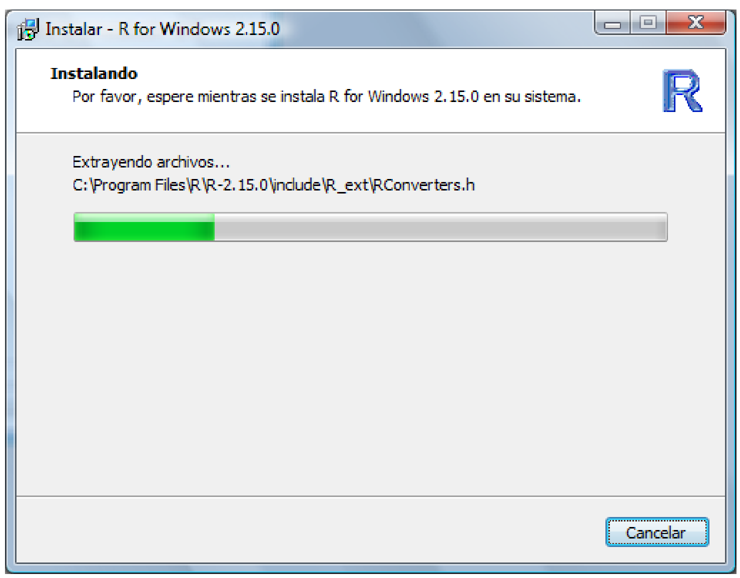
Elegir la carpeta del menú inicio donde colocar los accesos directos a los elementos del paquete, pulsar siguiente.



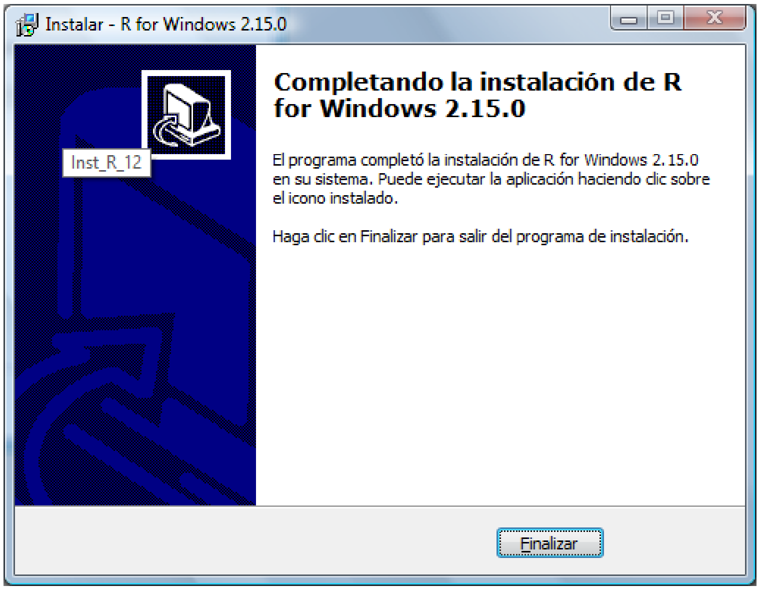
Selección de las tareas adicionales: crear icono en el escritorio, …, pulsar siguiente.



En este preciso instante se inicia el desempaquetado e instalación de la aplicación que tardará muy poco, en mi caso uno 50 segundos aproximadamente.



Una vez finalizada el proceso de instalación sale una ventana indicándolo. Pulsar Finalizar.



2.3. El entorno R

R es un conjunto integrado de programas para manipulación de datos, cálculo y gráficos.

Características y ventajas de la programación en R

- El lenguaje de programación R es un proyecto colaborativo y abierto, los desarrolladores pueden descargar el código de forma gratuita y modificarlo para incluir mejoras.
- Es un lenguaje interpretado, funciona mediante comandos.
- R proporciona una amplia gama de herramientas estadísticas que incluyen análisis de datos y generación de gráficos. Este lenguaje tiene capacidad de generar gráficos de alta calidad. Estas características lo convierten en una potente herramienta de cálculo.
- Gracias a este lenguaje de programación los Data Scientists pueden manejar grandes volúmenes de datos.

- Puede integrarse con distintas bases de datos. Una de las ventajas más importantes de R es que funciona con diferentes tipos de hardware y software (Windows, Unix, Linux…)
- El lenguaje R ofrece la posibilidad de cargar bibliotecas y paquetes con diversas funcionalidades lo que permite a los usuarios extender su configuración básica.
- La comunidad en torno a R es muy activa por lo que es sencillo encontrar soluciones rápidamente a los problemas que los usuarios se puedan encontrar.

El término “entorno" lo caracteriza como un sistema completamente diseñado y coherente, antes que como una agregación incremental de herramientas muy específicas e inflexibles, como ocurre frecuentemente con otros programas de análisis de datos.

R es en gran parte un vehículo para el desarrollo de nuevos métodos de análisis interactivo de datos. Como tal es muy dinámico y las diferentes versiones no siempre son totalmente compatibles con las anteriores. Algunos usuarios prefieren los cambios debido a los nuevos métodos y tecnología que los acompañan, a otros sin embargo les molesta ya que algún código anterior deja de funcionar. Aunque R puede entenderse como un lenguaje de programación, los programas escritos en R deben considerarse esencialmente efímeros.

2.4. Estadística con R:

Muchas personas utilizan R como un sistema estadístico. Nosotros preferimos describirlo como un entorno en el que se han implementado muchas técnicas estadísticas, tanto clásicas como modernas. Algunas están incluidas en el entorno base de R y otras se acompañan en forma de bibliotecas (packages). El hecho de distinguir entre ambos conceptos es fundamentalmente una cuestión histórica. Junto con R se incluyen ocho bibliotecas (llamadas bibliotecas estándar) pero otras muchas están disponibles a través de Internet en CRAN (http://www.r-project.org).

Como hemos indicado, muchas técnicas estadísticas, desde las clásicas hasta la última metodología, están disponibles en R, pero los usuarios necesitaran estar dispuestos a trabajar un poco para poder encontrarlas.

Existe una diferencia fundamental en la filosofía que subyace en R (o S) y la de otros sistemas estadísticos. En R, un análisis estadístico se realiza en una serie de pasos, con unos resultados intermedios que se van almacenando en objetos, para ser observados o analizados posteriormente, produciendo unas salidas mínimas. Sin embargo, en SAS o SPSS se obtendrá de modo inmediato una salida copiosa cualquier análisis, por ejemplo, una regresión o un análisis discriminante.

2.5. R en un sistema de ventanas

La forma más conveniente de usar R es en una estación de trabajo con un sistema de ventanas. Estas notas están escritas pensando en usuarios de estas características. En particular nos referiremos ocasionalmente a la utilización de R en un sistema X-window, aunque normalmente se pueden aplicar a cualquier implementación del entorno R.

Muchos usuarios encontraran necesario interactuar directamente con el sistema operativo de su ordenador de vez en cuando. En estas notas se trata fundamentalmente de la interacción con el sistema operativo UNIX. Si utiliza R bajo Microsoft Windows necesitara realizar algunos pequeños cambios.

El ajuste del programa para obtener el máximo rendimiento de las cualidades parametrizables de R es una tarea interesante, aunque tediosa y no se considerara en estas notas. Si tiene dificultades busque a un experto cercano a usted.

2.6. Como funciona R:

R es un lenguaje Orientado a Objetos: bajo este complejo término se esconde la simplicidad y flexibilidad de R. El hecho que R es un lenguaje de programación puede desaminar a muchos usuarios que piensan que no tienen “alma de programadores”. Esto no es necesariamente cierto por dos razones. Primero R es un lenguaje interpretado (como Java) y no compilado (como C, C++, Fortran, Pascal,...), lo cual significa que los comandos escritos en el teclado son ejecutados directamente sin necesidad de construir ejecutables. Como segunda medida, la sintaxis de R es muy simple e intuitiva. Por ejemplo, una regresión lineal se puede ejecutar con el comando lm(y ˜x). Para que una función sea ejecutada en R debe estar siempre acompañada de paréntesis, inclusive en el caso que no haya nada dentro de los mismos (por ej., ls()). Si se escribe el nombre de la función sin los paréntesis, R mostrar a el contenido (código) mismo de la función. En este documento, se escribirán los nombres de las funciones con paréntesis para distinguirlas de otros objetos, a menos que se indique lo contrario en el texto.

Orientado a Objetos significa que las variables, datos, funciones, resultados, etc., se guardan en la memoria activa del computador en forma de objetos con un nombre especıfico. El usuario puede modificar o manipular estos objetos con operadores (aritméticos, lógicos, y comparativos) y funciones (que a su vez son objetos). El uso y funcionamiento de los operadores es relativamente intuitivo. Los argumentos pueden ser objetos (“datos”, formulas, expresiones,...), algunos de los cuales pueden ser definidos por defecto en la función; sin embargo estos argumentos pueden ser modificados por el usuario con opciones. Una función en R puede carecer totalmente de argumentos, ya sea porque todos están definidos por defecto (y sus valores modificados con opciones), o porque la función realmente no tiene argumentos por ahora esta corta descripción es suficiente para entender el funcionamiento básico de R.

Todas las acciones en R se realizan con objetos que son guardados en la memoria activa del ordenador, sin usar archivos temporales. La lectura y escritura de archivos solo se realiza para la entrada y salida de datos y resultados (graficas,...). El usuario ejecuta las funciones con la ayuda de comandos definidos. Los resultados se pueden visualizar directamente en la pantalla, guardar en un objeto o escribir directamente en el disco (particularmente para gráficos). Debido a que los resultados mismos son objetos, pueden ser considerados como datos y analizados como tal. Archivos que contengan datos pueden ser leídos directamente desde el disco local o en un servido remoto a través de la red. Las funciones disponibles están guardadas en una librería localizada en el directorio R HOME/library (R HOME es el directorio donde R está instalado). Este directorio contiene paquetes de funciones, las cuales a su vez están estructuradas en directorios. El paquete denominado base constituye el núcleo de R y contiene las funciones básicas del lenguaje para leer y manipular datos, algunas funciones gráficas y algunas funciones estadísticas (regresión lineal y análisis de varianza). Cada ´ paquete contiene un directorio denominado R con un archivo con el mismo nombre del paquete (por ejemplo, para el paquete base, existe el archivo R HOME/library/base/R/base). Este archivo esta en formato ASCII y contiene todas las funciones del paquete. El comando más simple es escribir el nombre de un objeto para visualizar su contenido. Por ejemplo, si un objeto n contiene el valor 10:

```
> n
[1] 10
```

El dıgito 1 indica que la visualización del objeto comienza con el primer elemento de n. Este comando constituye un uso implícito de la función print, y el ejemplo anterior es similar a print(n) (en algunas situaciones la función print debe ser usada explícitamente, como por ejemplo dentro de una función o un bucle). El nombre de un objeto debe comenzar con una letra (A-Z and a-z) y puede incluir letras, dígitos

(0-9), y puntos (.). R discrimina entre letras mayúsculas y minúsculas para el nombre de un objeto, de tal manera que x y X se refiere a objetos diferentes (inclusive bajo Windows).

2.7. La ayuda en línea

La ayuda en línea de R proporciona información muy útil de cómo utilizar las funciones. La ayuda se encuentra disponible directamente para una función dada. Por ejemplo:

```
> ?lm
```

Al llamar la ayuda, se abre una ventana o página (esto depende del sistema operativo) con información general sobre la función en la primera línea, tal como el nombre del paquete donde se encuentra la función u operador. Después viene el tıtulo, seguido de secciones con información específica acerca de la misma:

- **Description:** descripción breve.
- **Usage:** para una función, proporciona el nombre de la misma con todos sus argumentos y los posibles valores por defecto (opciones); para un operador describe su uso típico.
- **Arguments:** para una función, describe en detalle cada uno de sus argumentos.
- **Details:** descripción detallada.
- **Value:** si se aplica, el tipo de objeto retornado por la función o el operador.
- **See Also:** otras páginas de ayuda con funciones u operadores similares.
- **Examples:** algunos ejemplos que generalmente pueden ser ejecutados sin abrir la ayuda con la función examples()

2.8. Clases de objetos:

Los vectores son el tipo básico de objeto en R, pero existen más tipos que veremos de modo formal posteriormente.

- **Las matrices** o, más generalmente, variables indexadas (Arrays) son generalizaciones Multidimensionales de los vectores. De hecho, son vectores indexados por dos o más índices y que se imprimen de modo especial.
- **Los factores** sirven para representar datos categóricos.
- **Las listas** son una forma generalizada de vector en las cuales los elementos no tienen por qué ser del mismo tipo y a menudo son a su vez vectores o

listas. Las listas permiten devolver los resultados de los cálculos estadísticos de un modo conveniente.

- **Las hojas de datos** (data frames) son estructuras similares a una matriz, en que cada columna puede ser de un tipo distinto a las otras. Las hojas de datos son apropiadas para describir `matrices de datos' donde cada fila representa a un individuo y cada columna una variable, cuyas variables pueden ser numéricas o categóricas. Muchos experimentos se describen muy apropiadamente con hojas de datos: los tratamientos son categóricos pero la respuesta es numérica.
- **Las funciones** son también objetos de R que pueden almacenarse en el espacio de trabajo, lo que permite extender las capacidades de R fácilmente.

2.9. Objetos más complejos en R: de vectores a listas

2.9.1. Vectores en R

La forma más natural de definir y almacenar más de un valor en R es crear un vector, que es probablemente el objeto R más simple. Esto se puede hacer usando la función c () (para concatenar o combinar)

```
> x <- c (-1,0,2)
> x
[1] -1 0 2
> y <- c (0,2 ^ x)
> y
[1] 0.0 0.5 1.0 4.0
```

Aquí, [1] indica que la respuesta comienza en el primer elemento de un vector: cuando se muestra un vector, R enumera los elementos, de izquierda a derecha, usando (posiblemente) varias filas (dependiendo del ancho de la pantalla) Observe que si un objeto es seguido por el operador de asignación <-, entonces se le asignará un valor (o un marco de datos, o una función, etc.) al objeto. Si simplemente escribimos el nombre del objeto, y luego ingresamos, el valor del objeto aparecerá en la consola (o el código de la función, si el objeto es una función). Cada nueva fila incluye el índice del valor que comienza esa fila, es decir,

```
> u <- 1:50
> u
[1] 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17
[18] 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34
[35] 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50
```

2.9.2. Matrices

Una matriz es solo un vector de datos, junto con un vector adicional, accesible por la función dim (), que contiene las dimensiones (es decir, número de filas nrow y columnas ncol).

```
> M <- matrix(U,nrow=5,ncol=4)> M
[, 1] [, 2] [, 3] [, 4]
[1,] 0.2655087 0.89838968 0.2059746 0.4976992
[2,] 0.3721239 0.94467527 0.1765568 0.7176185
[3,] 0.5728534 0.66079779 0.6870228 0.9919061
[4,] 0.9082078 0.62911404 0.3841037 0.3800352
[5,] 0.2016819 0.06178627 0.7698414 0.7774452
```

La dimensión de la matriz M se obtiene usando

```
> dim (M)
[1] 5 4
```

2.10. Lectura de datos de un archivo

Los datos suelen leerse desde archivos externos y no teclearse de modo interactivo. Las capacidades de lectura de archivos de R son sencillas y sus requisitos son bastante estrictos cuando no inflexibles. Se presupone que el usuario es capaz de modificar los archivos de datos con otras herramientas, por ejemplo, con editores de texto, para ajustarlos a las necesidades de R. Generalmente esta tarea es muy sencilla.

La función read.fwf() puede utilizarse para leer un archivo con campos de anchura fija no delimitados.(Esta función utiliza un programa perl para transformar el archivo en otro adaptado para su lectura con read.table.) La función count.fields() cuenta el número de campos por línea de un archivo de campos delimitados. Estas dos funciones pueden resolver algunos problemas elementales, pero en la mayoría de los casos es mejor preparar el archivo a las necesidades de R antes de comenzar el análisis.

Si los datos se van a almacena en hojas de datos, método que recomendamos, puede leer los datos correspondientes a las mismas con la función read.table(). Existe también una función más genérica, scan(), que puede utilizar directamente.

2.11. Paquetes R:

Un paquete es un conjunto de funciones relacionadas, que incluye archivos de ayuda y datos, que han sido unidos y se comparte entre la comunidad R. Esos paquetes son similares a bibliotecas en C / C ++ y clases en Java. Para obtener la lista de paquetes cargados de manera predeterminada, se puede usar el comando getOption:

```
> getOption ("defaultPackages")
[1] "datasets" "utils" "grDevices" "graphics" "stats" "methods"
> (.packages (all.available = TRUE))
[1] "AER" "evd" "sandwich" "lmtest" "nortest"
```

y muchos más paquetes, previamente instalados en la máquina. Todos estos paquetes están disponibles; solo tienes que cargarlos. Pero antes, tenemos que instalar un paquete de Internet (por ejemplo, quantreg para ejecutar regresiones de cuantiles) que usamos

```
> install.packages("quantreg", dependencies=TRUE)
```

y luego seleccionamos un sitio espejo para descargar. La opción dependencias = VERDADERO se usa porque el paquete quantreg puede estar usando funciones provenientes de otros paquetes (aquí, también se debe instalar el paquete MatrixModels). Tenga en cuenta que si un paquete no se ha cargado, no es posible llamar a las funciones asociadas:

```
> ajuste <- rq (Y ~ X1 + X2, datos = base, tau = .9)
Error: no se pudo encontrar la función "rq"
```

Para cargar un paquete en R, uno debe usar el comando library () o require (),

```
> library(quantreg)
```

2.12. Distribuciones Probabilísticas en R.

R contiene un amplio conjunto de tablas estadísticas. Para cada distribución soportada, hay funciones que permiten calcular la función de distribución, F(x) = P(X = x), la función de distribución inversa, la función de densidad y generar números pseudoaleatorios de la distribución. Las distribuciones son las siguientes:

Distribución	nombre en R	argumentos adicionales
beta	beta	shape1, shape2, ncp
binomial	binom	size, prob
Cauchy	cauchy	location, scale
ji cuadrado	chisq	df, ncp
exponencial	exp	rate
F de Snedecor	f	df1, df1, ncp
gamma	gamma	shape, scale
geométrica	geom	prob
hipergeométrica	hyper	m, n, k
log-normal	lnorm	meanlog, sdlog
logística	logis	location, scale
binomial negativa	nbinom	size, prob
normal	norm	mean, sd
Poisson	pois	lambda
t de Student	t	df, ncp
uniforme	unif	min, max
Weibull	weibull	shape, scale
Wilcoxon	wilcox	m, n

Para construir el nombre de cada función, utilice el nombre de la distribución precedido de “d” para la función de densidad, “p” para la función de distribución, “q” para la función de distribución inversa, y “r” para la generación de números pseudoaleatorios. El primer Argumento es x para la función de densidad, q para la función de distribución, p para la función de distribución inversa, y n para la función de generación de números pseudoaleatorios (excepto en el caso de rhyper y rwilcox, en los cuales es nn). En el momento de escribir este manual, el parámetro ncp solo está disponible prácticamente en las funciones de distribución. Para conocer donde puede usarlo utilice la ayuda interactiva.

IV. DISEÑO METODOLOGICO

1. Tipo de Estudio:

Es de tipo cuantitativo, descriptivo y de corte transversal.

a. Cuantitativo: Se realizó el cálculo de las frecuencias observadas del número de siniestros siguiente el procedimiento del programa R y los modelos de distribución de probabilidad discretas que permite obtener estimaciones del número de siniestros.
b. Descriptivo: Se describieron todos los datos obtenidos en la operación del cálculo del número de siniestro atreves de sus medias de tendencias y de dispersión y de representación gráfica con el fin de analizar sus resultados.
c. Corte Transversal: Se realizó en los meses de enero – Junio del 2018.

2. Área de Estudio:

Las operaciones del seguro privado.

3. Población:

Cincos carteras de seguros con 2,243,917 pólizas compuesta por, 639,950 pólizas suscritas en el ramo de Hogar, 1,044,454 pólizas suscritas en el ramo de Automóvil, 9,461 pólizas suscritas en el ramo de Robo, 149,473 pólizas suscritas en el ramo de Responsabilidad Civil y 400,579 pólizas suscritas en el ramo de Incendio.

4. Criterio de Inclusión:

- Cartera de seguro de Hogar
- Cartera de seguro de Automóvil
- Cartera de seguro de Responsabilidad Civil
- Cartera de seguro de Robo
- Cartera de seguro de Incendio

5. Criterio de Exclusión:

- Cartera de seguro de vida y Accidentes Personales
- Cartera de Fianzas.
- Cartera de Microseguro.
- Cartera de seguro de Transporte

6. Tipo de Variables:

a. Variables Dependientes:

- El cálculo de las frecuencias teóricas ajustadas de pólizas.
- Prueba de bondad de ajuste.

b. Variables independientes:

- Modelo de Distribución de probabilidad discreta.
- Los parámetros de los modelos de probabilidad discreta
- Número de siniestros observados en cada cartera de seguro.

7. Fuente de la recolección de la información:

Secundaria: Base de datos estadísticos de una compañía de seguros, Libro informe y documento de internet.

8. Procesamiento de la información:

Se introdujeron los datos de la información utilizando el método electrónico computarizado para el procesamiento de datos, Mediante el uso del programa de R, Word y Power Point año 2013, en la cual se hizo por cada variable de estudio.

9. Análisis de los datos:

Los datos se analizaron por cada modelo de distribución de probabilidad discreta que estima el número de siniestros presentándolo en tabla y gráficos.

V. RESULTADOS

La base de datos para realizar este trabajo contiene la información necesaria para estimar el número de siniestros a través de los modelos de distribuciones de probabilidad discretas y de la programación en R. La base de datos contiene el número de sinestros acaecidos y el número de pólizas que presentaron 0, 1, 2,…,8 reclamaciones en 5 ramos de seguros de daños en un año en particular. A continuación, se muestra la tabla con la siguiente distribución empírica de datos sobre el número de siniestros y número de pólizas para un determinado periodo:

Tabla 1: Datos sobre el número de siniestro por pólizas y número de pólizas para un determinado periodo, extraído de 5 carteras de seguros de daños de una empresa de seguros.

k (siniestros)	***Seguro de Robos***	***Seguro de Automóviles***	***Seguro de Hogar***	***Seguro de Responsabilidad Civil***	***Seguro de Incendio***
0	7,840	881,705	565,664	122,618	371,481
1	1,317	142,217	68,714	21,686	26,784
2	239	18,088	5,177	4,014	2,118
3	42	2,118	365	832	174
4	14	273	24	224	18
5	4	53	6	68	2
6	4	0	0	17	2
7	1	0	0	7	0
>=8	0	0	0	7	0
Total	9,461	1,044,454	639,950	149,473	400,579

A partir de esta información ajustaremos los datos observados a tres distribuciones de probabilidad discreta, es decir calcularemos las frecuencias absolutas teóricas (número de pólizas), que se obtendrían si la probabilidad de que ocurra 0, 1, 2, ..., n siniestros se distribuyese según el modelo poisson, binomial negativa o poisson - inversa gaussiana.

1. Estimación de las frecuencias absolutas teóricas según poisson mediante el lenguaje de programación R.

Consideraremos que el número de siniestros sigue una distribución $P_o(\lambda)$, con $\hat{\lambda} = \bar{x}$, siendo la $\bar{x}$ de la muestra de datos sobre el número de siniestros. A continuación mostraremos el cálculo en R de la media muestral, que utilizaremos como parámetro $\hat{\lambda}$, así como las probabilidades de que el número de siniestros tome valores 0, 1, 2,…,8 según la distribución de poisson.

Pasos:

> # 1) Cargar la base de datos de las 5 carteras de seguros la cual contiene el número de siniestros acaecidos y el número de pólizas con 0, 1, 2…8 reclamaciones.

```
> cartera = read.csv(file="C:/Users/Martin Alonso/Desktop/Monografia/Datos/cartera.csv", header=TRUE)

> attach(cartera)
```

> # 2) Calculamos la media ponderada de las 5 carteras de seguros:

```
> med.s.d.r = weighted.mean(k,s.d.r) ; med.s.d.r
[1] 0.2143537
> med.s.d.a = weighted.mean(k,s.d.a) ; med.s.d.a
[1] 0.1781831
> med.s.d.h = weighted.mean(k,s.d.h) ; med.s.d.h
[1] 0.1254614
> med.s.d.r.c = weighted.mean(k,s.d.r) ; med.s.d.r.c
[1] 0.2143537
> med.s.d.i = weighted.mean(k,s.d.i) ; med.s.d.i
[1] 0.07897568
```

> # 3) Se Calcula la probabilidad de poisson para cada una de las carteras:

```
> prpois.s.d.r = dpois(k,med.s.d.r) ; prpois.s.d.r
[1] 8.070629e-01 1.729969e-01 1.854126e-02 1.324796e-03 7.099369e-05
[6] 3.043552e-06 1.087327e-07 3.329609e-09 8.921423e-11
> prpois.s.d.a = dpois(k,med.s.d.a) ; prpois.s.d.a
[1] 8.367892e-01 1.491017e-01 1.328369e-02 7.889764e-04 3.514555e-05
[6] 1.252468e-06 3.719477e-08 9.467826e-10 2.108758e-11
> prpois.s.d.h = dpois(k,med.s.d.h) ; prpois.s.d.h
[1] 8.820898e-01 1.106682e-01 6.942291e-03 2.903298e-04 9.106293e-06
[6] 2.284976e-07 4.777936e-09 8.563520e-11 1.342989e-12
> prpois.s.d.r.c = dpois(k,med.s.d.r.c) ; prpois.s.d.r.c
[1] 8.070629e-01 1.729969e-01 1.854126e-02 1.324796e-03 7.099369e-05
[6] 3.043552e-06 1.087327e-07 3.329609e-09 8.921423e-11
> prpois.s.d.i = dpois(k,med.s.d.i) ; prpois.s.d.i
[1] 9.240624e-01 7.297846e-02 2.881762e-03 7.586303e-05 1.497834e-06
[6] 2.365849e-08 3.114075e-10 3.513375e-12 3.468390e-14
```

>\# Ley de la probabilidad total: la suma de las probabilidades tiene que ser igual a 1.

```
> sum(prpois.s.d.r)
[1] 1
> sum(prpois.s.d.a)
[1] 1
> sum(prpois.s.d.h)
[1] 1
> sum(prpois.s.d.r.c)
[1] 1
> sum(prpois.s.d.i)
```

[1] 1

> # 4) Calculamos las frecuencias teóricas observadas de las carteras de seguros:

> f.t.pois.s.d.r = prpois.s.d.r*sum(s.d.r) ; f.t.pois.s.d.r

[1] 7.635622e+03 1.636724e+03 1.754188e+02 1.253389e+01 6.716713e-01

[6] 2.879504e-02 1.028720e-03 3.150143e-05 8.440558e-07

> f.t.pois.s.d.a = prpois.s.d.a*sum(s.d.a) ; f.t.pois.s.d.a

[1] 8.739879e+05 1.557298e+05 1.387421e+04 8.240495e+02 3.670792e+01

[6] 1.308146e+00 3.884823e-02 9.888709e-04 2.202500e-05

> f.t.pois.s.d.h = prpois.s.d.h*sum(s.d.h) ; f.t.pois.s.d.h

[1] 5.644934e+05 7.082211e+04 4.442719e+03 1.857965e+02 5.827572e+00

[6] 1.462270e-01 3.057640e-03 5.480225e-05 8.594456e-07

> f.t.pois.s.d.r.c = prpois.s.d.r.c*sum(s.d.r.c) ; f.t.pois.s.d.r.c

[1] 1.206341e+05 2.585836e+04 2.771418e+03 1.980212e+02 1.061164e+01

[6] 4.549288e-01 1.625261e-02 4.976866e-04 1.333512e-05

> f.t.pois.s.d.i = prpois.s.d.i*sum(s.d.i) ; f.t.pois.s.d.i

[1] 3.701600e+05 2.923364e+04 1.154373e+03 3.038914e+01 6.000007e-01

[6] 9.477094e-03 1.247433e-04 1.407384e-06 1.389364e-08

> # 5) Exportamos la cartera de seguros al Excel para visualizar los datos en formato de números.

> cp=data.frame(k,s.d.r, f.t.pois.s.d.r, s.d.a, f.t.pois.s.d.a, s.d.h, f.t.pois.s.d.h, s.d.r.c, f.t.pois.s.d.r.c, s.d.i, f.t.pois.s.d.i)

> write.csv(cp, file="carterapoisson.csv", row.names=FALSE)

En la tabla 2 se muestran los resultados de los ajustes de la distribución de probabilidad aplicada, utilizando los datos de las 5 cartera de seguros de la tabla 1.

Tabla 2: Ajuste de las frecuencias teóricas de las 5 carteras de seguros mediante la distribución de poisson.

K	Seguro de robo	Frec.teo.pois. Seguro de robo	Seguro de automóvil	Frec.teo.pois. Seguro de automóvil.	Seguro de hogar	Frec.teo.pois. Seguro de hogar
0	7,840	7,636	881,705	873,988	565,664	564,493
1	1,317	1,637	142,217	155,730	68,714	70,822
2	239	175	18,088	13,874	5,177	4,443
3	42	12	2,118	824	365	186
4	14	1	273	37	24	6
5	4	0	53	1	6	0
6	4	0	0	0	0	0
7	1	0	0	0	0	0
>=8	0	0	0	0	0	0
Total	**9,461**	**9,461**	**1,044,454**	**1,044,454**	**639,950**	**639,950**

K	Seguro de responsabilidad civil.	Frec.teo.pois. Seguro de responsabilidad civil.	Seguro de incendio.	Frec.teo.pois. Seguro de incendio.
0	122,618	120,634	371,481	370,160
1	21,686	25,858	26,784	29,234
2	4,014	2,771	2,118	1,154
3	832	199	174	30
4	224	11	18	1
5	68	0	2	0
6	17	0	2	0
7	7	0	0	0
>=8	7	0	0	0
Total	**149,473**	**149,473**	**400,579**	**400,579**

2. Estimación de las frecuencias absolutas teóricas según binomial negativa.

Supongamos que el número de siniestros sigue una distribución $BN(a,b)$, es decir que depende de dos parámetros, a y b, cuyos estimadores se obtienen a partir de la media ($\bar{x}$) y varianza (S^2) de los datos muestrales:

$$\hat{a} = \frac{\bar{x}^2}{S^2 - \bar{x}}$$

$$\hat{b} = \frac{S^2 - \bar{x}}{\bar{x}}$$

Igual que en el caso anterior estimaremos los parámetros de la distribución y calcularemos las probabilidades teóricas a partir de la distribución binomial negativa:

Pasos:

> # 1) Cargar la base de datos de las 5 carteras de seguros la cual contiene el número de siniestros acaecidos y el número de pólizas con 0, 1,2...8 reclamaciones.

```
> cartera = read.csv(file="C:/Users/Martin Alonso/Desktop/Monografia/Datos/cartera.csv", header=TRUE);
> attach(cartera)
```

> # 2) Calculamos la media ponderada y la varianza de las 5 carteras de seguros:

```
> med.s.d.r = weighted.mean(k,s.d.r) ; med.s.d.r
[1] 0.2143537
> med.s.d.a = weighted.mean(k,s.d.a) ; med.s.d.a
[1] 0.1781831
> med.s.d.h = weighted.mean(k,s.d.h) ; med.s.d.h
[1] 0.1254614
> med.s.d.r.c = weighted.mean(k,s.d.r) ; med.s.d.r.c
[1] 0.2143537
> med.s.d.i = weighted.mean(k,s.d.i) ; med.s.d.i
[1] 0.07897568
```

```
> var.s.d.r = sum((k-med.s.d.r)^2*s.d.r)/sum(s.d.r); var.s.d.r
[1] 0.2889008
> var.s.d.a = sum((k-med.s.d.a)^2*s.d.a)/sum(s.d.a); var.s.d.a
[1] 0.1973887
> var.s.d.h = sum((k-med.s.d.h)^2*s.d.h)/sum(s.d.h); var.s.d.h
[1] 0.1299599
> var.s.d.r.c = sum((k-med.s.d.r.c)^2*s.d.r.c)/sum(s.d.r.c); var.s.d.r.c
[1] 0.2967601
> var.s.d.i = sum((k-med.s.d.i)^2*s.d.i)/sum(s.d.i); var.s.d.i
[1] 0.0867083
```

> # 3) Calcular los parámetros de la distribución binomial negativa que son alfa y beta de cada una de las carteras de seguros:

```
> alfa.bineg.s.d.r = med.s.d.r^2/(var.s.d.r-med.s.d.r); alfa.bineg.s.d.r
[1] 0.6163546
> alfa.bineg.s.d.a = med.s.d.a^2/(var.s.d.a-med.s.d.a); alfa.bineg.s.d.a
[1] 1.653117
> alfa.bineg.s.d.h = med.s.d.h^2/(var.s.d.h-med.s.d.h); alfa.bineg.s.d.h
[1] 3.499046
> alfa.bineg.s.d.r.c = med.s.d.r.c^2/(var.s.d.r.c-med.s.d.r.c); alfa.bineg.s.d.r.c
[1] 0.5575717
> alfa.bineg.s.d.i = med.s.d.i^2/(var.s.d.i-med.s.d.i); alfa.bineg.s.d.i
[1] 0.8066035
> beta.bineg.s.d.r = (var.s.d.r-med.s.d.r)/med.s.d.r; beta.bineg.s.d.r
[1] 0.3477765
> beta.bineg.s.d.a = (var.s.d.a-med.s.d.a)/med.s.d.a; beta.bineg.s.d.a
[1] 0.1077861
> beta.bineg.s.d.h = (var.s.d.h-med.s.d.h)/med.s.d.h; beta.bineg.s.d.h
[1] 0.03585588
```

```
> beta.bineg.s.d.r.c = (var.s.d.r.c-med.s.d.r.c)/med.s.d.r.c; beta.bineg.s.d.r.c
[1] 0.3844415
> beta.bineg.s.d.i = (var.s.d.i-med.s.d.i)/med.s.d.i; beta.bineg.s.d.i
[1] 0.09791141
```

> # 4) Calcular la probabilidad binomial negativa de cada una de las carteras de seguros:

```
> prbineg.s.d.r = dnbinom(k,alfa.bineg.s.d.r,1/(1+beta.bineg.s.d.r)); prbineg.s.d.r
[1] 8.319734e-01 1.323191e-01 2.759380e-02 6.209679e-03 1.448647e-03
[6] 3.451232e-04 8.336039e-05 2.033119e-05 4.994619e-06
> prbineg.s.d.a = dnbinom(k,alfa.bineg.s.d.a,1/(1+beta.bineg.s.d.a)); prbineg.s.d.a
[1] 8.443241e-01 1.358062e-01 1.752883e-02 2.076835e-03 2.350677e-04
[6] 2.585935e-05 2.789963e-06 2.967875e-07 3.123453e-08
> prbineg.s.d.h = dnbinom(k,alfa.bineg.s.d.h,1/(1+beta.bineg.s.d.h)); prbineg.s.d.h
[1] 8.840298e-01 1.070724e-01 8.337370e-03 5.290006e-04 2.975136e-05
[6] 1.544557e-06 7.573276e-08 3.557351e-09 1.616023e-10
> prbineg.s.d.r.c = dnbinom(k,alfa.bineg.s.d.r.c,1/(1+beta.bineg.s.d.r.c)); prbineg.s.d.r.c
[1] 8.341214e-01 1.291474e-01 2.792925e-02 6.611826e-03 1.632942e-03
[6] 4.133233e-04 1.063113e-04 2.765540e-05 7.254840e-06
> prbineg.s.d.i = dnbinom(k,alfa.bineg.s.d.i,1/(1+beta.bineg.s.d.i)); prbineg.s.d.i
[1] 9.274239e-01 6.671206e-02 5.374069e-03 4.483622e-04 3.805158e-05
[6] 3.262173e-06 2.815424e-07 2.441419e-08 2.124616e-09
```

># Ley de la probabilidad total: la suma de las probabilidades tiene que ser igual a 1.

```
> sum(prbineg.s.d.r)
[1] 0.9999984
> sum(prbineg.s.d.a)
[1] 1
> sum(prbineg.s.d.h)
```

```
[1] 1
> sum(prbineg.s.d.r.c)
[1] 0.9999974
> sum(prbineg.s.d.i)
[1] 1
```

En este ejemplo, a diferencia del ajuste mediante la distribución de poisson, en el caso de la distribución binomial negativa la suma de las probabilidades teóricas hasta n = 8 es significativamente distinta de 1. Por lo tanto,

> # 5) Calcularemos la probabilidad de que n > 8, así como la frecuencia absoluta teórica correspondiente:

```
> prbineg.s.d.r = c(prbineg.s.d.r,1-sum(prbineg.s.d.r)); prbineg.s.d.r
 [1] 8.319734e-01 1.323191e-01 2.759380e-02 6.209679e-03 1.448647e-03
 [6] 3.451232e-04 8.336039e-05 2.033119e-05 4.994619e-06 1.641688e-06
> prbineg.s.d.a = c(prbineg.s.d.a,1-sum(prbineg.s.d.a)); prbineg.s.d.a
 [1] 8.443241e-01 1.358062e-01 1.752883e-02 2.076835e-03 2.350677e-04
 [6] 2.585935e-05 2.789963e-06 2.967875e-07 3.123453e-08 3.636295e-09
> prbineg.s.d.h = c(prbineg.s.d.h,1-sum(prbineg.s.d.h)); prbineg.s.d.h
 [1] 8.840298e-01 1.070724e-01 8.337370e-03 5.290006e-04 2.975136e-05
 [6] 1.544557e-06 7.573276e-08 3.557351e-09 1.616023e-10 7.469914e-12
> prbineg.s.d.r.c = c(prbineg.s.d.r.c,1-sum(prbineg.s.d.r.c)); prbineg.s.d.r.c
 [1] 8.341214e-01 1.291474e-01 2.792925e-02 6.611826e-03 1.632942e-03
 [6] 4.133233e-04 1.063113e-04 2.765540e-05 7.254840e-06 2.608960e-06
> prbineg.s.d.i = c(prbineg.s.d.i,1-sum(prbineg.s.d.i)); prbineg.s.d.i
 [1] 9.274239e-01 6.671206e-02 5.374069e-03 4.483622e-04 3.805158e-05
 [6] 3.262173e-06 2.815424e-07 2.441419e-08 2.124616e-09 2.031726e-10
```

```
># Ley de la probabilidad total: la suma de las probabilidades tiene que ser igual a 1.
> sum(prbineg.s.d.r)
[1] 1
> sum(prbineg.s.d.a)
[1] 1
> sum(prbineg.s.d.h)
[1] 1
> sum(prbineg.s.d.r.c)
[1] 1
> sum(prbineg.s.d.i)
[1] 1
> # 6) Calcular las frecuencias teóricas observadas para cada cartera de seguros:
> f.t.bineg.s.d.r = prbineg.s.d.r*sum(s.d.r); f.t.bineg.s.d.r
 [1] 7.871300e+03 1.251871e+03 2.610649e+02 5.874978e+01 1.370565e+01
 [6] 3.265210e+00 7.886726e-01 1.923534e-01 4.725409e-02 1.553201e-02
> f.t.bineg.s.d.a = prbineg.s.d.a*sum(s.d.a); f.t.bineg.s.d.a
 [1] 8.818577e+05 1.418433e+05 1.830806e+04 2.169159e+03 2.455174e+02
 [6] 2.700891e+01 2.913988e+00 3.099808e-01 3.262303e-02 3.797943e-03
> f.t.bineg.s.d.h = prbineg.s.d.h*sum(s.d.h); f.t.bineg.s.d.h
 [1] 5.657349e+05 6.852099e+04 5.335500e+03 3.385339e+02 1.903938e+01
 [6] 9.884390e-01 4.846518e-02 2.276527e-03 1.034174e-04 4.780371e-06
> f.t.bineg.s.d.r.c = prbineg.s.d.r.c*sum(s.d.r.c); f.t.bineg.s.d.r.c
 [1] 1.246786e+05 1.930405e+04 4.174669e+03 9.882895e+02 2.440807e+02
 [6] 6.178067e+01 1.589066e+01 4.133735e+00 1.084403e+00 3.899691e-01
> f.t.bineg.s.d.i = prbineg.s.d.i*sum(s.d.i); f.t.bineg.s.d.i
 [1] 3.715065e+05 2.672345e+04 2.152739e+03 1.796045e+02 1.524266e+01
 [6] 1.306758e+00 1.127800e-01 9.779811e-03 8.510765e-04 8.138667e-05
```

> # 7) Exportamos la cartera de seguros al Excel para visualizar los datos en formato de números.

```
> cbineg = data.frame(f.t.bineg.s.d.r, f.t.bineg.s.d.a, f.t.bineg.s.d.h, f.t.bineg.s.d.r.c, f.t.bineg.s.d.i)
```

```
> write.csv(cbineg, file="carterabinega.csv", row.names=FALSE)
```

Tabla 3: Ajuste de las frecuencias teóricas de las 5 carteras de seguros mediante la distribución binomial negativa.

k	Seguro de robo.	Frec.teo.bineg. Seguro de robo.	Seguro de automovil.	Frec.teo.bineg. Seguro de automóvil.	Seguro de hogar.	Frec.teo.bineg. Seguro de hogar.
0	7,840	7,871	881,705	881,858	565,664	565,735
1	1,317	1,252	142,217	141,843	68,714	68,521
2	239	261	18,088	18,308	5,177	5,335
3	42	59	2,118	2,169	365	339
4	14	14	273	246	24	19
5	4	3	53	27	6	1
6	4	1	0	3	0	0
7	1	0	0	0	0	0
>=8	0	0	0	0	0	0
Total	**9,461**	**9,461**	**1,044,454**	**1,044,454**	**639,950**	**639,950**

K	Seguro de responsabilidad civil.	Frec.teo.bineg. Seguro de responsabilidad civil.	Seguro de incendio .	Frec.teo.bineg . Seguro de incendio.
0	122,618	124,679	371,481	371,507
1	21,686	19,304	26,784	26,723
2	4,014	4,175	2,118	2,153
3	832	988	174	180
4	224	244	18	15
5	68	62	2	1
6	17	16	2	0
7	7	4	0	0
>=8	7	1	0	0
Total	**149,473**	**149,473**	**400,579**	**400,579**

3. Estimación de las frecuencias absolutas teóricas según poisson-inversa gaussiana.

Se considera que el número de siniestros sigue una distribución $IG(\mu,\beta)$, con parámetros µ y β, que pueden estimarse a partir de las siguientes expresiones:

$$\hat{\mu} = \bar{x}$$

$$\hat{\beta} = \frac{S^2}{\bar{x}} - 1$$

Para calcular las probabilidades teóricas a partir de la distribución poisson-inversa gaussiana utilizaremos la función $dPIG$ de la librería gamlss.dist. se muestra, a continuación, la ejecución de dicha función y los resultados obtenidos:

Pasos:

> # 1) Subimos al R la librería de gamlss.dist. Esta librería tiene la formula y el código para la distribución de poisson inversa gaussiana

```
> library(gamlss.dist)
Loading required package: MASS
```

> # 2) Cargar la base de datos de las 5 carteras de seguros la cual contiene el número de siniestros acaecidos y el número de pólizas con 0,1,2...8 reclamaciones.

```
> cartera = read.csv(file="C:/Users/Martin Alonso/Desktop/Monografia/Datos/cartera.csv", header=TRUE)
> attach(cartera)
```

> # 3) Calculamos la media ponderada y la varianza de las 5 carteras de seguros:

```
> med.s.d.r = weighted.mean(k,s.d.r) ; med.s.d.r
[1] 0.2143537
> med.s.d.a = weighted.mean(k,s.d.a) ; med.s.d.a
[1] 0.1781831
> med.s.d.h = weighted.mean(k,s.d.h) ; med.s.d.h
[1] 0.1254614
> med.s.d.r.c = weighted.mean(k,s.d.r) ; med.s.d.r.c
```

```
[1] 0.2143537
> med.s.d.i = weighted.mean(k,s.d.i) ; med.s.d.i
[1] 0.07897568
> var.s.d.r = sum((k-med.s.d.r)^2*s.d.r)/sum(s.d.r); var.s.d.r
[1] 0.2889008
> var.s.d.a = sum((k-med.s.d.a)^2*s.d.a)/sum(s.d.a); var.s.d.a
[1] 0.1973887
> var.s.d.h = sum((k-med.s.d.h)^2*s.d.h)/sum(s.d.h); var.s.d.h
[1] 0.1299599
> var.s.d.r.c = sum((k-med.s.d.r.c)^2*s.d.r.c)/sum(s.d.r.c); var.s.d.r.c
[1] 0.2967601
> var.s.d.i = sum((k-med.s.d.i)^2*s.d.i)/sum(s.d.i); var.s.d.i
[1] 0.0867083
```

> # 4) Calculamos el parámetro beta de la distribución poisson inversa gaussiana

```
> beta.poig.s.d.r = var.s.d.r/med.s.d.r ; beta.poig.s.d.r
[1] 1.347777
> beta.poig.s.d.a = var.s.d.a/med.s.d.a ; beta.poig.s.d.a
[1] 1.107786
> beta.poig.s.d.h = var.s.d.h/med.s.d.h ; beta.poig.s.d.h
[1] 1.035856
> beta.poig.s.d.r.c = var.s.d.r.c/med.s.d.r.c ; beta.poig.s.d.r.c
[1] 1.384441
> beta.poig.s.d.i = var.s.d.i/med.s.d.i ; beta.poig.s.d.i
[1] 1.097911
```

> # 5) Calculamos la probabilidad de poisson inversa gaussiana para cada cartera de seguros:

```
> prpoig.s.d.r = dPIG(k,med.s.d.r,beta.poig.s.d.r/med.s.d.r); prpoig.s.d.r
```

```
[1] 0.8635548167 0.0962899807 0.0229269546 0.0085610438 0.0039265446
[6] 0.0020101496 0.0011012864 0.0006317462 0.0003746433
> prpoig.s.d.a = dPIG(k,med.s.d.a,beta.poig.s.d.a/med.s.d.a); prpoig.s.d.a
[1] 0.8802200221 0.0874638324 0.0194113972 0.0068312861 0.0029577513
[6] 0.0014299235 0.0007399008 0.0004008945 0.0002245605
> prpoig.s.d.h = dPIG(k,med.s.d.h,beta.poig.s.d.h/med.s.d.h); prpoig.s.d.h
[1] 0.9128739673 0.0653477438 0.0133573709 0.0045602411 0.0019279842
[6] 0.0009113968 0.0004613471 0.0002445900 0.0001340750
> prpoig.s.d.r.c = dPIG(k,med.s.d.r.c,beta.poig.s.d.r.c/med.s.d.r.c); prpoig.s.d.r.c
[1] 0.8643726104 0.0954389345 0.0227979186 0.0085683869 0.0039574935
[6] 0.0020404375 0.0011258935 0.0006505033 0.0003885428
> prpoig.s.d.i = dPIG(k,med.s.d.i,beta.poig.s.d.i/med.s.d.i); prpoig.s.d.i
[1] 0.9449149295 0.0417440665 0.0080925749 0.0027937482 0.0012010416
[6] 0.0005779305 0.0002978965 0.0001608486 0.0000898060
```

># Ley de la probabilidad total: la suma de las probabilidades tiene que ser igual a 1 para cada una de las carteras.

```
> sum(prpoig.s.d.r)
[1] 0.9993772
> sum(prpoig.s.d.a)
[1] 0.9996796
> sum(prpoig.s.d.h)
[1] 0.9998187
> sum(prpoig.s.d.r.c)
[1] 0.9993407
> sum(prpoig.s.d.i)
[1] 0.9998728
```

> #) por lo anterior, nos damos que la suma de las probabilidades de las carteras no es igual a 1 dado que tiene más de 8 reclamos; por lo que se debe realizar un pequeño ajuste de la siguiente manera:

```
> prpoig.s.d.r = c(prpoig.s.d.r,1-sum(prpoig.s.d.r));prpoig.s.d.r
 [1] 0.8635548167 0.0962899807 0.0229269546 0.0085610438 0.0039265446
 [6] 0.0020101496 0.0011012864 0.0006317462 0.0003746433 0.0006228341
> prpoig.s.d.a = c(prpoig.s.d.a,1-sum(prpoig.s.d.a)); prpoig.s.d.a
 [1] 0.8802200221 0.0874638324 0.0194113972 0.0068312861 0.0029577513
 [6] 0.0014299235 0.0007399008 0.0004008945 0.0002245605 0.0003204315
> prpoig.s.d.h = c(prpoig.s.d.h,1-sum(prpoig.s.d.h));prpoig.s.d.h
 [1] 0.9128739673 0.0653477438 0.0133573709 0.0045602411 0.0019279842
 [6] 0.0009113968 0.0004613471 0.0002445900 0.0001340750 0.0001812839
> prpoig.s.d.r.c = c(prpoig.s.d.r.c,1-sum(prpoig.s.d.r.c));prpoig.s.d.r.c
 [1] 0.8643726104 0.0954389345 0.0227979186 0.0085683869 0.0039574935
 [6] 0.0020404375 0.0011258935 0.0006505033 0.0003885428 0.0006592790
> prpoig.s.d.i = c(prpoig.s.d.i,1-sum(prpoig.s.d.i));prpoig.s.d.i
 [1] 0.9449149295 0.0417440665 0.0080925749 0.0027937482 0.0012010416
 [6] 0.0005779305 0.0002978965 0.0001608486 0.0000898060 0.0001271577
```

># EL resultado del ajuste nos indica que la suma de las probabilidades es igual a 1.

```
> sum(prpoig.s.d.r)
[1] 1
> sum(prpoig.s.d.a)
[1] 1
> sum(prpoig.s.d.h)
[1] 1
> sum(prpoig.s.d.r.c)
[1] 1
> sum(prpoig.s.d.i)
```

```
[1] 1
```

> # 7) Calculamos las frecuencias teóricas observadas de cada cartera de seguro:

```
> f.t.poig.s.d.r = prpoig.s.d.r*sum(s.d.r); f.t.poig.s.d.r
 [1] 8170.092121  910.999508  216.911918   80.996036   37.149038   19.018025
 [7]   10.419271    5.976951    3.544500    5.892633
> f.t.poig.s.d.a = prpoig.s.d.a*sum(s.d.a); f.t.poig.s.d.a
 [1] 919349.3230  91351.9496  20274.3115   7134.9641   3089.2351   1493.4894
 [7]    772.7923    418.7158    234.5432    334.6760
> f.t.poig.s.d.h = prpoig.s.d.h*sum(s.d.h); f.t.poig.s.d.h
 [1] 584193.69535  41819.28867   8548.04951   2918.32627   1233.81348
 [6]    583.24839    295.23905    156.52540     85.80129    116.01261
> f.t.poig.s.d.r.c = prpoig.s.d.r.c*sum(s.d.r.c); f.t.poig.s.d.r.c
 [1] 129200.36719  14265.54386   3407.67328   1280.74249    591.53843
 [6]    304.99032    168.29068     97.23268     58.07666     98.54441
> f.t.poig.s.d.i = prpoig.s.d.i*sum(s.d.i); f.t.poig.s.d.i
 [1] 378513.07755  16721.79640   3241.71556   1119.11687    481.11206
 [6]    231.50683    119.33108     64.43256     35.97440     50.93670
```

> # 8) Exportamos la cartera de seguros al Excel para visualizar los datos en formato de números.

```
> cpoig = data.frame(f.t.poig.s.d.r, f.t.poig.s.d.a, f.t.poig.s.d.h, f.t.poig.s.d.r.c, f.t.poig.s.d.i)
> write.csv(cpoig, file="carterapoissoninversa.csv", row.names=FALSE)
```

Tabla 4: Ajuste de las frecuencias teóricas de las 5 carteras de seguros mediante la distribución poisson-inversa gaussiana.

K	Seguro de Robo	Frec.teo.poig. Seguro de Robo	Seguro de Automóvil	Frec.teo.poig. Seguro de Automóvil	Seguro de Hogar	Frec.teo.poig. Seguro de Hogar
0	7,840	8,170	881,705	919,349	565,664	584,194
1	1,317	911	142,217	91,352	68,714	41,819
2	239	217	18,088	20,274	5,177	8,548
3	42	81	2,118	7,135	365	2,918
4	14	37	273	3,089	24	1,234
5	4	19	53	1,493	6	583
6	4	10	0	773	0	295
7	1	6	0	419	0	157
>=8	0	9	0	569	0	202
Total	**9,461**	**9,461**	**1,044,454**	**1,044,454**	**639,950**	**639,950**

K	Seguro de Responsabilidad Civil	Frec.teo.poig. Seguro de Responsabilidad Civil	Seguro de Incendio	Frec.teo.poig. Seguro de incendio
0	122,618	129,200	371,481	378,513
1	21,686	14,266	26,784	16,722
2	4,014	3,408	2,118	3,242
3	832	1,281	174	1,119
4	224	592	18	481
5	68	305	2	232
6	17	168	2	119
7	7	97	0	64
>=8	7	157	0	87
Total	**149,473**	**149,473**	**400,579**	**400,579**

4. Prueba de Bondad de Ajuste:

A veces nos preguntamos cual es el modelo (poisson, binomial negativa, poisson-inversa gaussiana) más adecuado para una serie de datos de frecuencias teóricas de una cartera de seguro en particular. Una forma de decidirse por un modelo es el test chi-cuadrado de bondad de ajuste a distribuciones que puede aplicarse de la siguiente forma:

Si disponemos de los datos completos de la muestra, como en este caso, podremos comprobar el ajuste del modelo propuesto, es decir, de la distribución elegida. Si le llamamos O_i a la frecuencia observada en la clase i, e_i, el número de elementos que teóricamente que debería caer en esa clase (el número esperado de elementos contenidos en cada clase debe ser al menos 5) entonces el estadístico de contraste es:

$$x^2 = \sum_{i=0}^{k-1} \frac{(O_i - e_i)^2}{{e_i}^2} \sim {x^2}_{n-p-1}$$

Donde:

$p = es\ el\ número\ de\ parametros\ de\ la\ distribución\ propuesta\ que\ se\ hayan$

$estimado\ por\ medio\ de\ estadisticos\ muestrales.$

$e_i\ = n * p_i = Frecuancias\ Teóricas\ Esperadas.$

$$n = \sum_{i=1}^{k} O_i = Tamaño\ de\ la\ muestra.$$

$p_i = Probabilidad\ de\ que\ la\ variable\ aleatoria\ tome\ valores\ dentro\ de\ un$

$intervalo\ determinado, según\ la\ distribucion\ elegida$

$O_i = Número\ de\ observaciones\ reales.$

El valor del x^2 del estadístico no puede ser negativo. Sera nulo si hay un acuerdo perfecto entre los valores experimentales y teóricos. Es obvio que los valores teóricos se aproximan mejor a los experimentales mientras más cerca de 0 este valor x^2

Rechazamos la distribución propuestas al nivel de significación $\propto$ si el estadístico $x^2 > x^2_{\alpha,n-p-1}$

Si el número de los elementos no es suficiente para realizar este test puede emplearse el test de bondad de ajuste de Kolmogorov Smirnov. Este test de chi-cuadrado puede encontrarse implementando en la mayor parte de los paquetes estadísticos.

El procedimiento para hacer un test de chi-cuadrado de bondad de ajuste, para comprobar si los datos pueden considerarse procedentes de una distribución normal:

- Obtener la medida de tendencia central y dispersión.
- Agrupar los intervalos para que en ninguno de las casillas (intervalos de clases) haya una frecuencia menor que 5.
- Se obtiene la columna de las frecuencias teóricas, ya sea de la distribución Poisson, Binomial Negativa y Poisson Inversa Gaussiana.
- Se obtiene el valor del estadístico x^2
- Se compara el valor del estadístico x^2 y el P-valué de cada una de las distribuciones para saber cuál es la distribución más cercana a las frecuencias observadas.
- Se toma la decisión de cual modelo de distribución de probabilidad discreta se ajusta mejor a los siniestros acaecidos observados tomando en cuenta aquel que tenga el x^2 más bajos, es decir aquel que sea más cercano a cero, y el P-value más alto.

Para poder realizar el contraste de bondad de ajuste con los datos obtenidos según las distribuciones Poisson, Binomial Negativa y Poisson-Inversa Gaussiana, ejecutaremos la función que se ha programado en R para dicho fin. Esta función, que se denomina *funcion.ajuste,* utiliza la función *pchisq* de R, de manera que para su ejecución es necesario disponer de los vectores con los valores de las frecuencias absolutas reales y teóricas, convenientemente agrupados para que se cumpla que $n.pi \geq 5$, además del número de parámetros que se han estimado en cada distribución. Recordemos que la distribución de Poisson sólo depende de un parámetro, mientras que las distribuciones Binomial-Negativa y Poisson-Inversa Gaussiana dependen de dos parámetros.

```
> # 1) Se Crea una función para poder calcular el chi-cuadrado y el p-value
> funcion.ajuste = function(s.d.x.analisis,f.t.pois.s.d.x.analisis,par) {
+ D = sum((s.d.x.analisis-f.t.pois.s.d.x.analisis)^2/f.t.pois.s.d.x.analisis)
+ grad = length(s.d.x.analisis)-par-1
+ pvalue = 1-pchisq(D,grad)
```

```
+ print(D)
+ print(pvalue)
+ }
```

> # 2) Análisis de Ajuste de Bondad para la distribución de poisson

> # Se reagrupa la cartera de las frecuencias teóricas poisson de manera de que el número de las muestras sea mayor o igual a 5, creando una nueva variable

```
> f.t.pois.s.d.r.analisis = c(f.t.pois.s.d.r[1:3],sum(f.t.pois.s.d.r[4:9])); f.t.pois.s.d.r.analisis
[1] 7635.62216 1636.72357  175.41885   13.23542
> f.t.pois.s.d.a.analisis = c(f.t.pois.s.d.a[1:4],sum(f.t.pois.s.d.a[5:9])); f.t.pois.s.d.a.analisis
[1] 873987.86328 155729.82372  13874.20754    824.04954     38.05592
> f.t.pois.s.d.h.analisis = c(f.t.pois.s.d.h[1:4],sum(f.t.pois.s.d.h[5:9])); f.t.pois.s.d.h.analisis
[1] 5.644934e+05 7.082211e+04 4.442719e+03 1.857965e+02 5.976912e+00
> f.t.pois.s.d.r.c.analisis = c(f.t.pois.s.d.r.c[1:4],sum(f.t.pois.s.d.r.c[5:9])); f.t.pois.s.d.r.c.analisis
[1] 120634.11386  25858.36412   2771.41753    198.02117     11.08333
> f.t.pois.s.d.i.analisis = c(f.t.pois.s.d.i[1:3],sum(f.t.pois.s.d.i[4:9])); f.t.pois.s.d.i.analisis
[1] 370159.99008  29233.63792   1154.37326     30.99874
```

> # Se reagrupa la cartera de las frecuencias observadas poisson de manera de que el número de las muestras sea mayor o igual a 5, creando una nueva variable

```
> s.d.r.pois.analisis = c(s.d.r[1:3],sum(s.d.r[4:9])); s.d.r.pois.analisis
[1] 7840 1317  239   65
> s.d.a.pois.analisis = c(s.d.a[1:4],sum(s.d.a[5:9])); s.d.a.pois.analisis
[1] 881705 142217  18088   2118    326
> s.d.h.pois.analisis = c(s.d.h[1:4],sum(s.d.h[5:9])); s.d.h.pois.analisis
[1] 565664  68714   5177    365     30
```

> s.d.r.c.pois.analisis = c(s.d.r.c[1:4],sum(s.d.r.c[5:9])); s.d.r.c.pois.analisis

[1] 122618 21686 4014 832 323

> s.d.i.pois.analisis = c(s.d.i[1:3],sum(s.d.i[4:9])); s.d.i.pois.analisis

[1] 371481 26784 2118 196

> # Se utiliza la función de ajuste para poder conocer el p-value y el chi cuadrado de la distribución de Poisson

> analisis.ajuste.pois.s.d.r = funcion.ajuste(s.d.r.pois.analisis,f.t.pois.s.d.r.analisis,1)

[1] 293.4263

[1] 0

> analisis.ajuste.pois.s.d.a = funcion.ajuste(s.d.a.pois.analisis,f.t.pois.s.d.a.analisis,1)

[1] 6730.937

[1] 0

> analisis.ajuste.pois.s.d.h = funcion.ajuste(s.d.h.pois.analisis,f.t.pois.s.d.h.analisis,1)

[1] 455.9387

[1] 0

> analisis.ajuste.pois.s.d.r.c = funcion.ajuste(s.d.r.c.pois.analisis,f.t.pois.s.d.r.c.analisis,1)

[1] 12070.93

[1] 0

> analisis.ajuste.pois.s.d.i = funcion.ajuste(s.d.i.pois.analisis,f.t.pois.s.d.i.analisis,1)

[1] 1892.656

[1] 0

> # 3) Análisis de Ajuste de Bondad para la distribución Binomial Negativa

> # Se reagrupa la cartera de las frecuencias teóricas Binomial Negativa de manera de que el número de las muestras sea mayor o igual a 5, creando una nueva variable

> f.t.bineg.s.d.r.analisis = c(f.t.bineg.s.d.r[1:4],sum(f.t.bineg.s.d.r[5:9])); f.t.bineg.s.d.r.analisis

[1] 7871.29991 1251.87073 261.06491 58.74978 18.01467

```
> f.t.bineg.s.d.a.analisis = c(f.t.bineg.s.d.a[1:5],sum(f.t.bineg.s.d.a[6:9]));
f.t.bineg.s.d.a.analisis

[1] 881857.6509 141843.3474  18308.0560   2169.1591    245.5174     30.2693

> f.t.bineg.s.d.h.analisis = c(f.t.bineg.s.d.h[1:4],sum(f.t.bineg.s.d.h[5:9]));
f.t.bineg.s.d.h.analisis

[1] 565734.89617  68520.99157   5335.49965    338.53394     20.07867

> f.t.bineg.s.d.r.c.analisis = c(f.t.bineg.s.d.r.c[1:7],sum(f.t.bineg.s.d.r.c[8:9]));
f.t.bineg.s.d.r.c.analisis

[1] 1.246786e+05 1.930405e+04 4.174669e+03 9.882895e+02 2.440807e+02

[6] 6.178067e+01 1.589066e+01 5.608107e+00

> f.t.bineg.s.d.i.analisis = c(f.t.bineg.s.d.i[1:4],sum(f.t.bineg.s.d.i[5:9]));
f.t.bineg.s.d.i.analisis

[1] 371506.53226  26723.45127   2152.73906    179.60449     16.67291
```

> # Se reagrupa la cartera de las frecuencias observadas Binomial Negativa de manera de que el número de las muestras sea mayor o igual a 5, creando una nueva variable

```
> s.d.r.bineg.analisis = c(s.d.r[1:4],sum(s.d.r[5:9])); s.d.r.bineg.analisis

[1] 7840 1317  239   42   23

> s.d.a.bineg.analisis = c(s.d.a[1:5],sum(s.d.a[6:9])); s.d.a.bineg.analisis

[1] 881705 142217  18088   2118    273     53

> s.d.h.bineg.analisis = c(s.d.h[1:4],sum(s.d.h[5:9])); s.d.h.bineg.analisis

[1] 565664  68714   5177    365     30

> s.d.r.c.bineg.analisis = c(s.d.r.c[1:7],sum(s.d.r.c[8:9])); s.d.r.c.bineg.analisis

[1] 122618  21686   4014    832    224     68     17     14

> s.d.i.bineg.analisis = c(s.d.i[1:4],sum(s.d.i[5:9])); s.d.i.bineg.analisis

[1] 371481  26784   2118    174     22
```

> # Se utiliza la función de ajuste para poder conocer el p-value y el chi cuadrado de la distribución Binomial Negativa

```
> vanalisis.ajuste.bineg.s.d.r =
funcion.ajuste(s.d.r.bineg.analisis,f.t.bineg.s.d.r.analisis,2)
```

[1] 11.5328

[1] 0.003131009

> analisis.ajuste.bineg.s.d.a = funcion.ajuste(s.d.a.bineg.analisis,f.t.bineg.s.d.a.analisis,2)

[1] 25.00823

[1] 1.53794e-05

> analisis.ajuste.bineg.s.d.h = funcion.ajuste(s.d.h.bineg.analisis,f.t.bineg.s.d.h.analisis,2)

[1] 12.23247

[1] 0.002206753

> analisis.ajuste.bineg.s.d.r.c = funcion.ajuste(s.d.r.c.bineg.analisis,f.t.bineg.s.d.r.c.analisis,2)

[1] 373.7825

[1] 0

> analisis.ajuste.bineg.s.d.i = funcion.ajuste(s.d.i.bineg.analisis,f.t.bineg.s.d.i.analisis,2)

[1] 2.576452

[1] 0.2757595

> # 4) Análisis de Ajuste de Bondad para la distribución Poisson Inversa Gaussiana

> # Se reagrupa la cartera de las frecuencias teóricas Poisson Inversa Gaussiana de manera de que el número de las muestras sea mayor o igual a 5, creando una nueva variable

> f.t.poig.s.d.r.analisis = c(f.t.poig.s.d.r[1:6],sum(f.t.poig.s.d.r[7:9])); f.t.poig.s.d.r.analisis

[1] 8170.09212 910.99951 216.91192 80.99604 37.14904 19.01802 25.83336

> f.t.poig.s.d.a.analisis = c(f.t.poig.s.d.a[1:5],sum(f.t.poig.s.d.a[6:9])); f.t.poig.s.d.a.analisis

[1] 919349.323 91351.950 20274.312 7134.964 3089.235 3254.217

> f.t.poig.s.d.h.analisis = c(f.t.poig.s.d.h[1:5],sum(f.t.poig.s.d.h[6:9])); f.t.poig.s.d.h.analisis

[1] 584193.695 41819.289 8548.050 2918.326 1233.813 1236.827

```
> f.t.poig.s.d.r.c.analisis = f.t.poig.s.d.r.c; f.t.poig.s.d.r.c.analisis
[1] 129200.36719  14265.54386   3407.67328   1280.74249    591.53843
[6]    304.99032    168.29068     97.23268    156.62107
> f.t.poig.s.d.i.analisis = c(f.t.poig.s.d.i[1:4],sum(f.t.poig.s.d.i[5:9])); f.t.poig.s.d.i.analisis
[1] 378513.0775  16721.7964   3241.7156   1119.1169    983.2936
```

> # Se reagrupa la cartera de las frecuencias observadas Poisson Inversa Gaussiana de manera de que el número de las muestras sea mayor o igual a 5, creando una nueva variable

```
> s.d.r.poig.analisis = c(s.d.r[1:6],sum(s.d.r[7:9])); s.d.r.poig.analisis
[1] 7840 1317  239   42   14    4    5
> s.d.a.poig.analisis = c(s.d.a[1:5],sum(s.d.a[6:9])); s.d.a.poig.analisis
[1] 881705 142217  18088   2118    273     53
> s.d.h.poig.analisis = c(s.d.h[1:5],sum(s.d.h[6:9])); s.d.h.poig.analisis
[1] 565664  68714   5177    365     24      6
> s.d.r.c.poig.analisis = s.d.r.c; s.d.r.c.poig.analisis
[1] 122618  21686   4014    832    224     68     17      7      7
> s.d.i.poig.analisis = c(s.d.i[1:4],sum(s.d.i[5:9])); s.d.i.poig.analisis
[1] 371481  26784   2118    174     22
```

> # Se utiliza la función de ajuste para poder conocer el p-value y el chi cuadrado de la distribución Poisson Inversa Gaussiana

```
> analisis.ajuste.poig.s.d.r = funcion.ajuste(s.d.r.poig.analisis,f.t.poig.s.d.r.analisis,2)
[1] 258.3863
[1] 0
> analisis.ajuste.poig.s.d.a = funcion.ajuste(s.d.a.pois.analisis,f.t.pois.s.d.a.analisis,2)
[1] 6730.937
[1] 0
> analisis.ajuste.poig.s.d.h = funcion.ajuste(s.d.h.pois.analisis,f.t.pois.s.d.h.analisis,2)
```

```
[1] 455.9387

[1] 0

> analisis.ajuste.poig.s.d.r.c =
funcion.ajuste(s.d.r.c.pois.analisis,f.t.pois.s.d.r.c.analisis,2)

[1] 12070.93

[1] 0

> analisis.ajuste.poig.s.d.i = funcion.ajuste(s.d.i.pois.analisis,f.t.pois.s.d.i.analisis,2)

[1] 1892.656

[1] 0
```

5. Resumen de la prueba de bondad de ajuste aplicada a las 5 carteras de los 3 modelos de probabilidad discretas.

Tabla 5: Tabla de salida de la prueba de bondad de ajuste con sus 5 carteras y sus 3 distribuciones de probabilidades discretas.

Cartera de Seguros	χ^2 Valor de Probabilidad	Poisson	Binomial Negativa	Poisson-inversa Gaussiana
Seguro de Robo	D	293.43	12	258.3863
	P-Value	0	0.003131	0
Seguro de Autómovil	D	6,730.94	25	6730.9370
	P-Value	0	0.000015	0
Seguro de Hogar	D	455.94	12	455.9387
	P-Value	0	0.002207	0
Seguro de Responsabilidad Civil	D	12,070.93	374	12070.9300
	P-Value	0	0.000000	0
Seguro de Incendio	D	1,892.66	3	1892.6560
	P-Value	0	0.275760	0

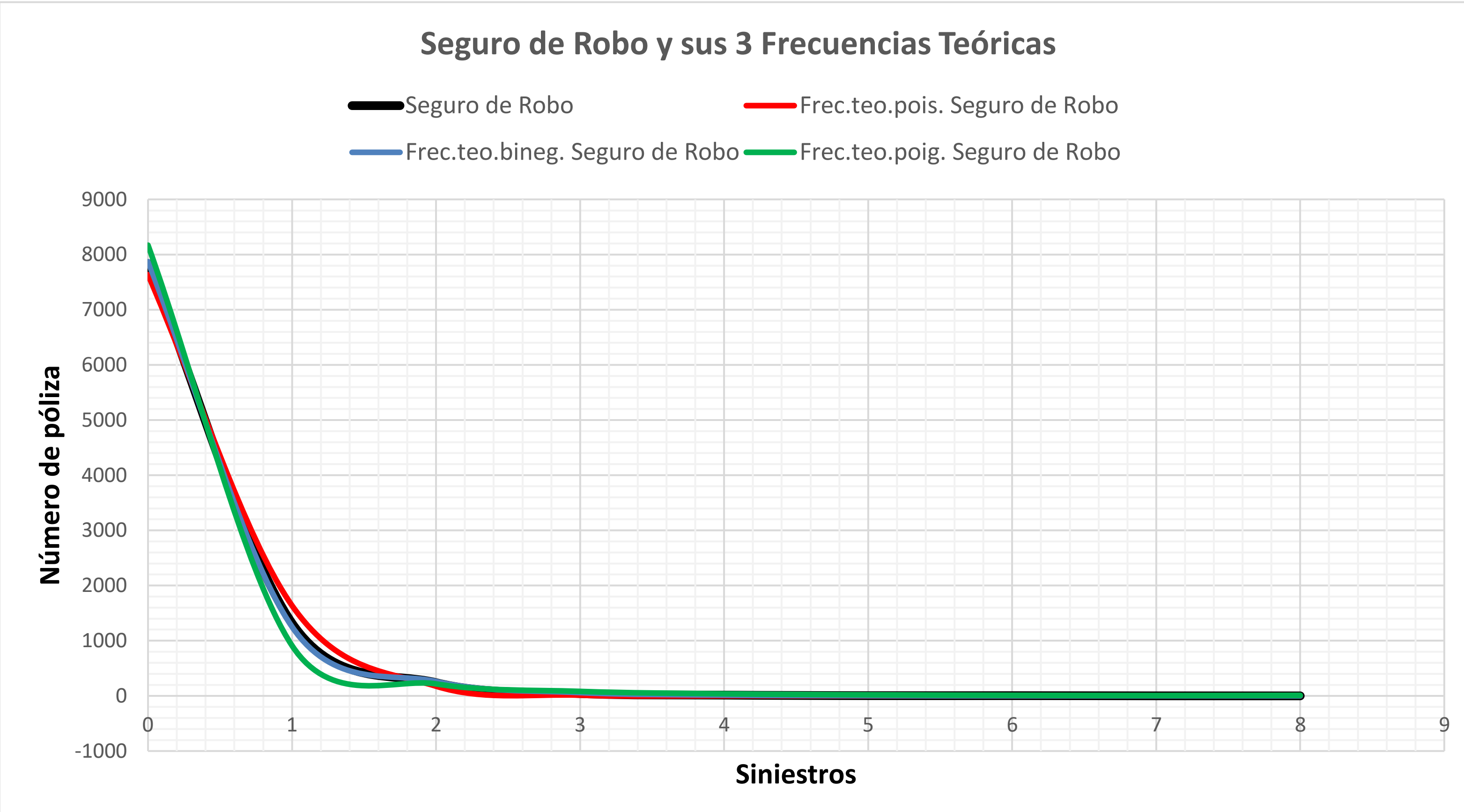
6. Discusión de los resultados:

6.1. Resultados No 1 de la cartera de Robo.

Tabla 6: Estimación de las frecuencias teóricas mediante los tres modelos de distribución de probabilidad (poisson, binomial negativa y poisson-inversa gaussiana) para la cartera de Seguro de Robo.

K	Seguro de Robo	Frec.teo.pois. Seguro de Robo	Frec.teo.bineg. Seguro de Robo	Frec.teo.poig. Seguro de Robo
0	7,840	7,636	7,871	8,170
1	1,317	1,637	1,252	911
2	239	175	261	217
3	42	13	59	81
4	14	1	14	37
5	4	0	3	19
6	4	0	1	10
7	1	0	0	6
>=8	0	0	0	9
Total	**9,461**	**9,461**	**9,461**	**9,461**

Gráfico 1: Líneas suavizadas del seguro de robo con sus tres Frecuencias Teóricas



Fuente: *Compañía de seguros.*

El gráfico 1 y la Tabla 6 muestra el modelo de distribución de probabilidad que mejor se ajusta a los siniestros acaecidos observados. Mediante la prueba de bondad de ajuste se pudo determinar, que la que mejor se ajusta a la cartera de Robo es la distribución binomial negativa dado que presenta un menor valor de x^2 y un P-value más grande tal como se muestra a continuación:

Tabla 7: Bondad de ajuste de la cartera de seguro de robo con sus 3 distribuciones discretas

Seguro de Robo			
	Poisson	**Binomial Negativa**	**Poisson-inversa Gaussiana**
D	293.43	12	258.3863
P-value	0	0.003131	0

En cuanto al gráfico 1 podemos ver como la línea Azul que es la frecuencia teórica binomial negativa sigue muy de cerca a la línea negra que son los siniestros acaecidos observados de la cartera de Robo, siendo la más lejana la frecuencia teórica Poisson-inversa gaussiana dado que ésta presenta un buen número de póliza con más de seis siniestros.

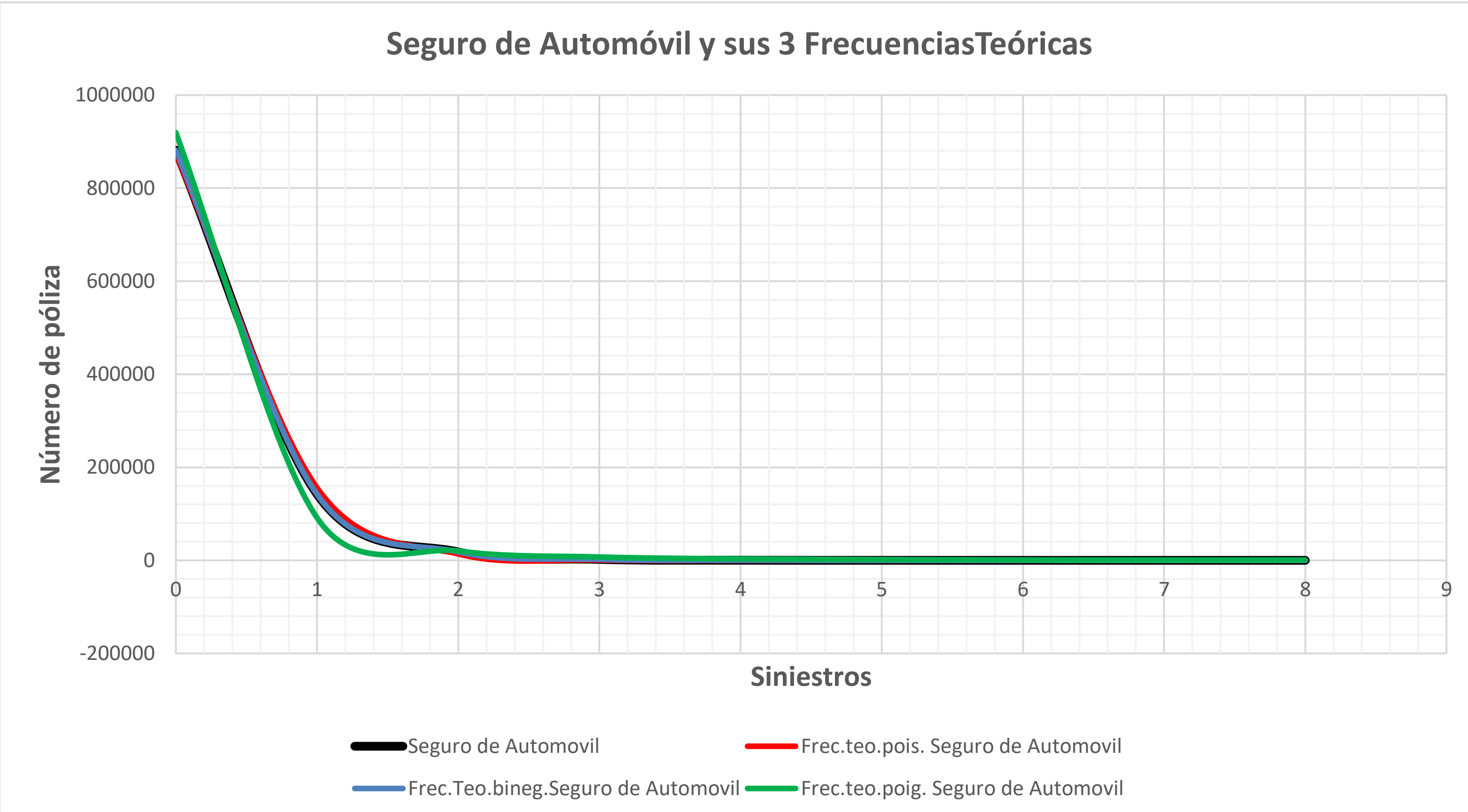
Debido a que la media es mayor que la mediana y la moda, la distribución de los valores observados tiene un sesgo positivo, es decir, hacia la derecha. Como el coeficiente de sesgo es mayor que cero (esto es P = 0.39880125> 0) no tiene una distribución normal. Esto se debe que en la cola del extremo derecho encontramos pocas pólizas con pocos siniestros.

6.2. Resultados No. 2 de la cartera de automóvil.

Tabla 8: Estimación de las frecuencias teóricas mediante los tres modelos de distribución de probabilidad (poisson, binomial negativa y poisson-inversa gaussiana) para la cartera de Seguro de automóvil.

K	Seguro de Automóvil	Frec.teo.pois. Seguro de Automóvil	Frec.Teo.bineg.Seguro de Automóvil	Frec.teo.poig. Seguro de Automóvil
0	881,705	873,988	881,858	919,349
1	142,217	15,5730	141,843	91,352
2	18,088	1,3874	18,308	20,274
3	2,118	824	2,169	7,135
4	273	37	246	3,089
5	53	1	27	1,493
6	0	0	3	773
7	0	0	0	419
>=8	0	0	0	569
Total	**1,044,454**	**1,044,454**	**1,044,454**	**1,044,454**

Gráfico 2: Líneas suavizada del Seguro de Automóvil con sus 3 *Frecuencias Teóricas*



***Fuente**: Compañía de seguro*

El gráfico 2 y la Tabla 8 muestran que el modelo que mejor se adapta a la cartera de seguro de automóvil es la distribución binomial negativa. Según el análisis de la prueba de bondad de ajuste la distribución binomial negativa tiene el x^2 más bajo y el P-value más alto tal y como se muestra en la siguiente tabla:

Tabla 9: Bondad de ajuste de la cartera de automóvil de robo con sus 3 distribuciones discretas

Seguro de Automóvil			
	Poisson	**Binomial Negativa**	**Poisson-inversa Gaussiana**
D	6,730.94	25	6730.9370
P-value	0	0.000015	0

En cuanto al gráfico 2 podemos ver como la línea Azul que es la frecuencia teórica binomial negativa sigue muy de cerca a la línea negra que son los siniestros acaecidos observados de la cartera de Automóvil, siendo la más lejana la frecuencia teórica poisson-inversa gaussiana.

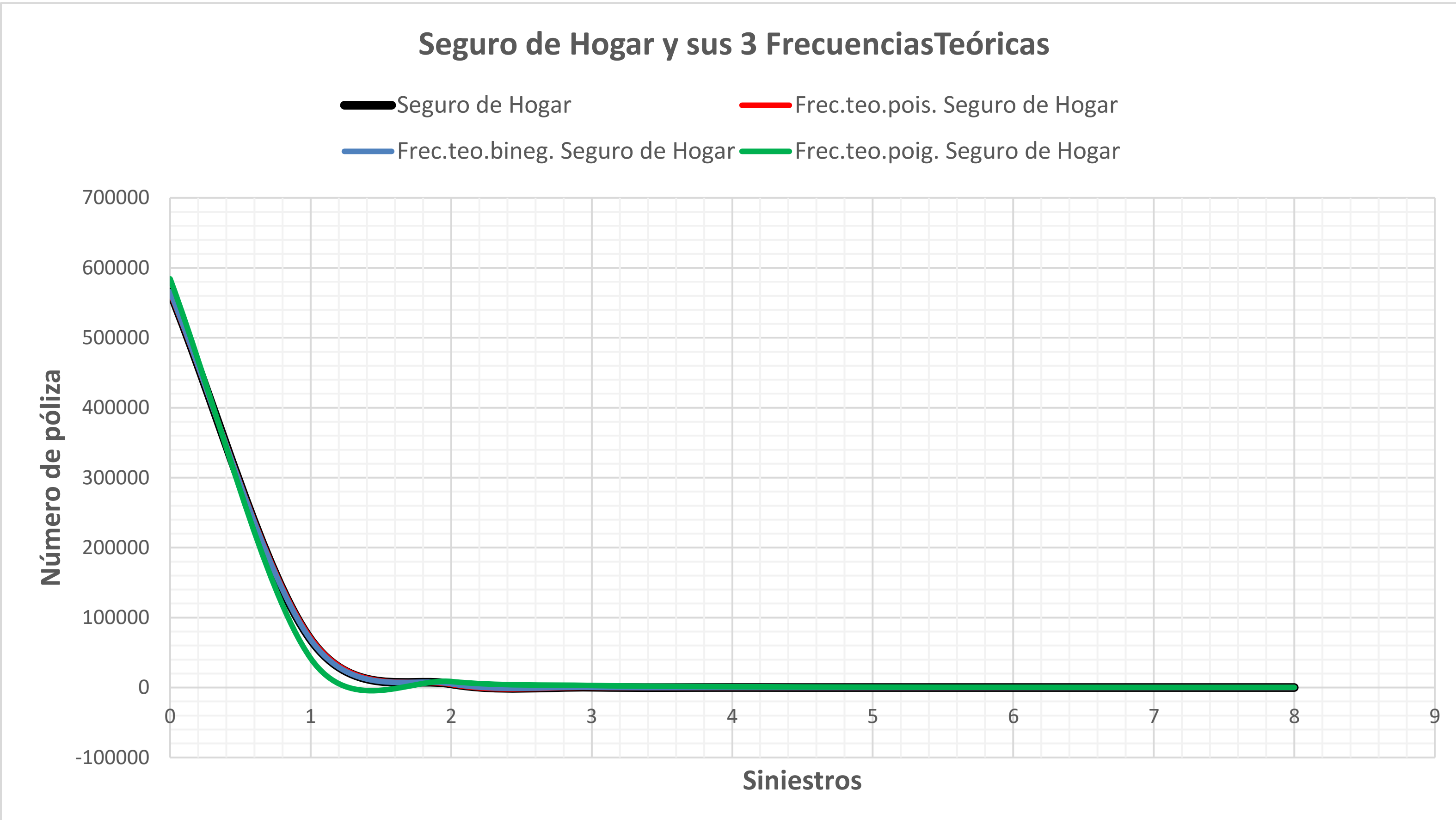
Debido a que la media es mayor que la mediana y la moda, la distribución de los valores observados tiene un sesgo positivo, es decir, hacia la derecha. Como el coeficiente de sesgo es mayor que cero (esto es $P = 0.40105621 > 0$) no tiene una distribución normal. Esto se debe que en la cola del extremo derecho encontramos pocas pólizas con pocos siniestros.

6.3. Resultados No 3 de la cartera de Hogar.

Tabla 10: Estimación de las frecuencias teóricas mediante los tres modelos de distribución de probabilidad (poisson, binomial negativa y poisson-inversa gaussiana) para la cartera de Seguro de Hogar.

K	Seguro de Hogar	Frec.teo.pois. Seguro de Hogar	Frec.teo.bineg. Seguro de Hogar	Frec.teo.poig. Seguro de Hogar
0	565,664	564,493	565,735	58,4194
1	68,714	70,822	68,521	41,819
2	5,177	4,443	5,335	8,548
3	365	186	339	2,918
4	24	6	19	1,234
5	6	0	1	583
6	0	0	0	295
7	0	0	0	157
>=8	0	0	0	202
Total	**639,950**	**639,950**	**639,950**	**639,950**

Gráfico 3: Líneas suavizada del Seguro de Hogar con sus 3 Frecuencias Teóricas



Fuente: *Compañía de seguro.*

El gráfico 3 y la Tabla 10 muestra el modelo de distribución de probabilidad que mejor se ajusta a los siniestros acaecidos observados y que a través de la prueba de bondad de ajuste se obtiene que la que mejor se ajusta a la cartera de Hogar es la distribución binomial negativa dado que presenta un menor valor de x^2 y un P-value más grande tal como se muestra a continuación:

Tabla 11: Bondad de ajuste de la cartera de seguro de Hogar con sus 3 distribuciones discretas

Seguro de Hogar			
	Poisson	**Binomial Negativa**	**Poisson-inversa Gaussiana**
D	455.94	12	455.9387
P-value	0	0.002207	0

La línea Azul es la frecuencia teórica binomial negativa y esta sigue muy de cerca, con respecto a las demás, a la línea negra que son los siniestros acaecidos observados de la cartera de Hogar siendo la más lejana la frecuencia teórica poisson-inversa gaussiana.

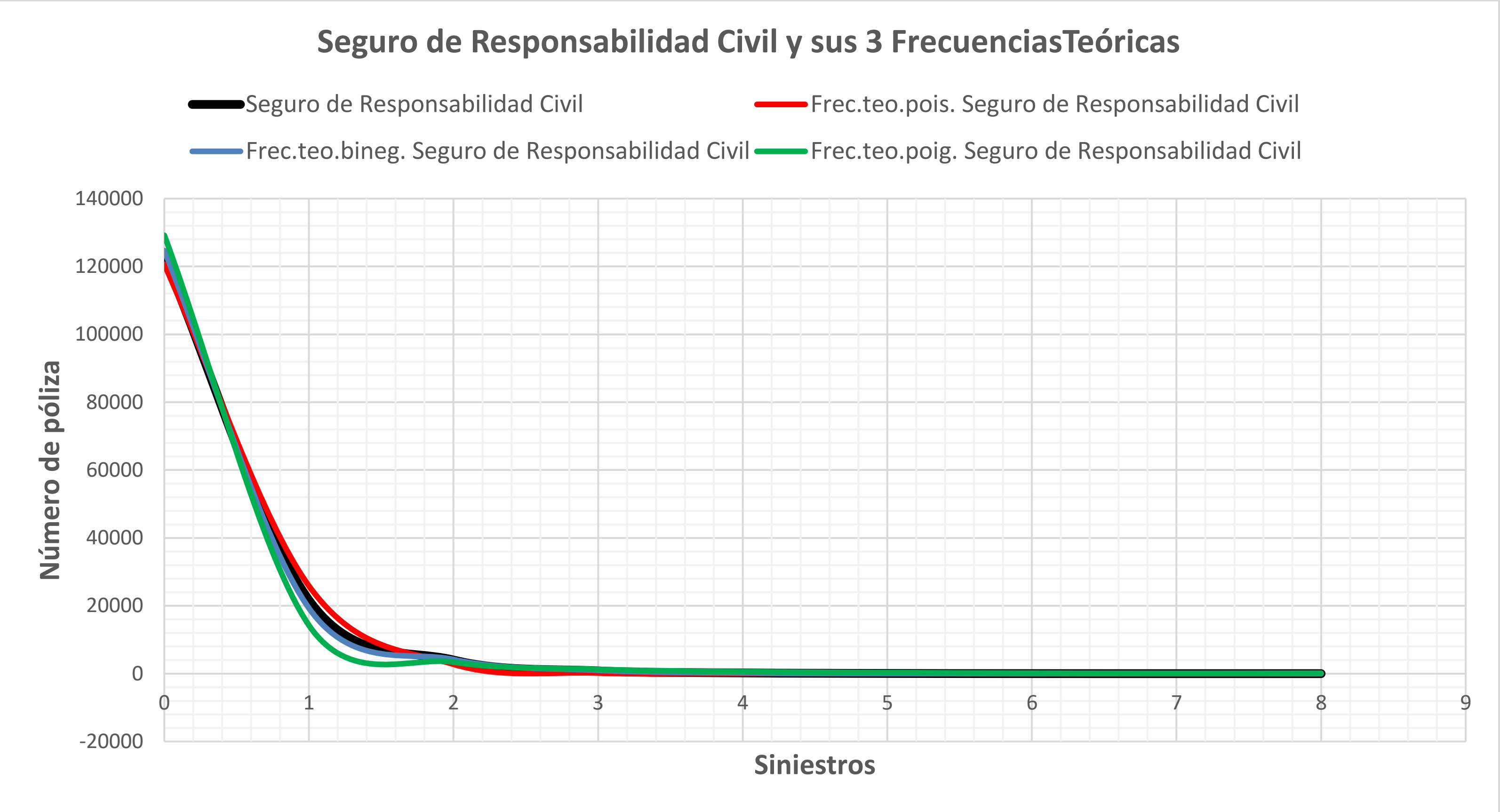
Debido a que la media es mayor que la mediana y la moda, la distribución de los valores observados tiene un sesgo positivo, es decir, hacia la derecha. Como el coeficiente de sesgo es mayor que cero (esto es P = 0.34802091 > 0) no tiene una distribución normal. Esto se debe que en la cola del extremo derecho encontramos pocas pólizas con pocos siniestros.

6.4. Resultados No 4 de la cartera de Responsabilidad Civil.

Tabla 12: Estimación de las frecuencias teóricas mediante los tres modelos de distribución de probabilidad (poisson, binomial negativa y poisson-inversa gaussiana) para la cartera de Seguro de Responsabilidad Civil.

K	Seguro de Responsabilidad Civil	Frec.teo.pois. Seguro de Responsabilidad Civil	Frec.teo.bineg. Seguro de Responsabilidad Civil	Frec.teo.poig. Seguro de Responsabilidad Civil
0	122,618	120,634	124,679	129,200
1	21,686	25,858	19,304	14,266
2	4,014	2,771	4,175	3,408
3	832	198	988	1,281
4	224	11	244	592
5	68	0	62	305
6	17	0	16	168
7	7	0	4	97
>=8	7	0	1	157
Total	**149,473**	**149,473**	**149,473**	**149,473**

Gráfico 4: Líneas suavizada del Seguro de Responsabilidad Civil con sus 3 Frecuencias Teóricas



Fuente: *Compañía de seguro.*

El gráfico 4 y la Tabla 12 muestra el modelo de distribución de probabilidad que mejor se ajusta a los siniestros acaecidos observados, mediante la prueba de bondad de ajuste se determina que la que mejor se ajusta para la cartera de Responsabilidad civil es la distribución binomial negativa dado que presenta un menor valor de x^2 y un P-valué igual a cero, que aunque el P-valué no es mayor que cero, su x^2 es menor que el x^2 de las otras distribuciones, por lo tanto decimos, que la distribución binomial negativa es la que mejor se ajusta a los datos observados reales tal como se muestra a continuación:

Tabla 13: Bondad de ajuste de la cartera de seguro de responsabilidad civil con sus 3 distribuciones discretas

Seguro de Responsabilidad Civil			
	Poisson	**Binomial Negativa**	**Poisson-inversa Gaussiana**
D	12,070.93	374	12070.93
P-value	0	0	0

En cuanto al gráfico 4 podemos ver como la línea Azul que es la frecuencia teórica binomial negativa sigue muy de cerca a la línea negra que son los siniestros acaecidos observados de la cartera de Responsabilidad civil, siendo la más lejana la frecuencia teórica poisson-inversa gaussiana, por el cumulo de siniestros que tiene en lo extremo de su cola.

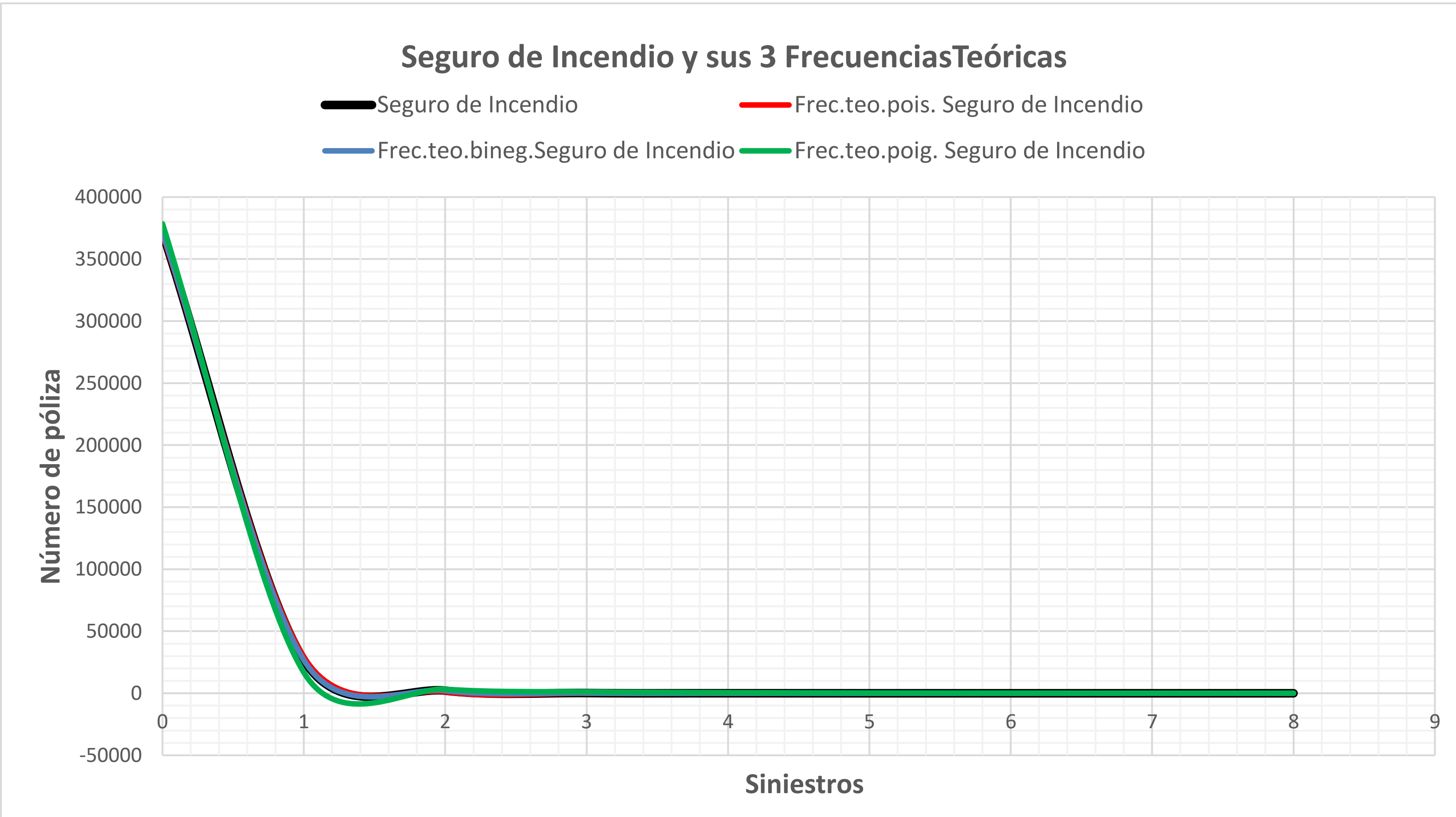
Debido a que la media es mayor que la mediana y la moda, la distribución de los valores observados tiene un sesgo positivo, es decir, hacia la derecha. Como el coeficiente de sesgo es mayor que cero (esto es $P = 0.41337433 > 0$) no tiene una distribución normal. Esto se debe que en la cola del extremo derecho encontramos pocas pólizas con pocos siniestros.

6.5. Resultados No 5 de la cartera de Incendio.

Tabla 14: Estimación de las frecuencias teóricas mediante los tres modelos de distribución de probabilidad (poisson, binomial negativa y poisson-inversa gaussiana) para la cartera de Seguro de Incendio.

K	Seguro de Incendio	Frec.teo.pois. Seguro de Incendio	Frec.teo.bineg.Seguro de Incendio	Frec.teo.poig. Seguro de Incendio
0	371,481	370,160	371,507	378,513
1	26,784	29,234	26,723	16,722
2	2,118	1,154	2,153	3,242
3	174	30	180	1,119
4	18	1	15	481
5	2	0	1	232
6	2	0	0	119
7	0	0	0	64
>=8	0	0	0	87
Total	**400,579**	**400,579**	**400,579**	**400,579**

Gráfico 5: Líneas suavizada del Seguro de Incendio con sus 3 frecuencias teóricas



Fuente: *Compañía de seguro.*

El gráfico 5 y la Tabla 14 muestra el modelo de distribución de probabilidad que mejor se ajusta a los siniestros acaecidos observados, mediante la prueba de bondad de ajuste se determinó, que la que mejor se ajusta para la cartera de Incendio casi de manera perfecta, es la distribución binomial negativa dado que presenta un menor valor de x^2 y un P-value más grande tal como se muestra a continuación:

Tabla 15: Bondad de ajuste de la cartera de seguro de incendio con sus 3 distribuciones discretas

Seguro de Incendio			
	Poisson	**Binomial Negativa**	**Poisson-inversa Gaussiana**
D	1,892.66	3	1892.6560
P-value	0	0.275760	0

En cuanto al gráfico 5 podemos ver como la línea Azul que es la frecuencia teórica binomial negativa sigue muy de cerca a la línea negra que son los siniestros acaecidos observados de la cartera de Incendio, seguida de Poisson y siendo la más lejana la frecuencia teórica Poisson-inversa Gaussiana.

Debido a que la media es mayor que la mediana y la moda, la distribución de los valores observados tiene un sesgo positivo, es decir, hacia la derecha. Como el coeficiente de sesgo es mayor que cero (esto es $P = 0.26820264 > 0$) no tiene una distribución normal. Esto se debe que en la cola del extremo derecho encontramos pocas pólizas con pocos siniestros.

VI. CONCLUSIÓN

Al finalizar este trabajo monográfico obtuvimos las siguientes conclusiones que:

- Los modelos de distribución de probabilidad sirven para estimar, modelizar y cuantificar el número de siniestros y su costo de una cartera de seguros. Las distribuciones de probabilidades aplicadas al ámbito de los seguros se estudian en dos campos diferentes: en el campo continuo, el costo de las reclamaciones, en el campo discreto, el número de siniestros.

- Las distribuciones de probabilidades no cuantifican ni calcula el costo o valor de una cartera de seguro, pero sí estiman y modelizan el número de siniestros de una cartera de seguros.

- El lenguaje de programación R, que consta de un sistema base y de librerías adicionales, cubre las necesidades descritas como herramienta de cálculo gratuita y de libre acceso, siendo además un software muy dinámico, ya que las funciones y las librerías se van ampliando gracias a las aportaciones de usuarios que pasan un estricto proceso de elaboración y contrastación. Estamos hablando, además, de un software ampliamente utilizado en el campo actuarial, siendo constante la aparición de librerías especializadas en temas actuariales como el paquete gamlss.dist, que se utilizó en la monografía para calcular las frecuencias teóricas de la distribución Poisson inversa-gausseana.

- El programa R nos ayuda a guardar base de datos y poder analizarla y manipularla, en nuestro caso, poder estimar la cartera de seguros con las distribuciones discretas y hacer el análisis de ajuste de bondad. Los actuarios y la tecnología podemos adaptarnos para generar mayor aprovechamiento del tiempo y reducción de las cargas de trabajo. Que la ausencia algunos conocimientos no es limitación cuando se tiene interés y dedicación, y que podemos dirigir o colaborar en trabajos tan pequeños como este o grandes con el fin de brindar soluciones a tareas repetitivas de cálculos actuariales.

- Al utilizar la prueba de bondad de ajuste nos dimos cuentas que el modelo de distribución de probabilidad que mejor ajuste realiza a los cincos carteras de seguros estudiada es la distribución binomial negativa dado a que su P-value (valor de probabilidad) es mayor que la de los otros ramos y un D (chi-cuadrado) más cercano a cero, lo que nos indican que los riesgos suscritos en las carteras analizadas son heterogéneos, esto es, presentan comportamiento probabilístico diferente, y la que peor ajuste realiza es la Poisson inversa-gaussiana.

- A pesar de los grandes esfuerzos que realizan las compañías de seguros en clasificar sus carteras en ramos con objetivo de homogeneizar los riesgos clasificándolo por ramo, esto en la práctica no se logra debido a que cada riesgo se comporta de manera diferente en cuanto a su frecuencia e intensidad. Es por esto que quizás la más importante aplicación de la distribución binomial negativa, al menos desde el punto de vista de los seguros generales, es la que se refiere a la distribución de la frecuencia de siniestralidad cuando los riesgos no son homogéneos.

VII. RECOMENDACIONES

Luego de cumplir con los objetivos consideramos que este trabajo puede ser mejorado, es por eso que damos las siguientes recomendaciones:

A los estudiantes:

- Que trabajen con el programa R ya que con este programa se pueden hacer distintos cálculos afines a la carrera, utilizándose más en las clases de estadísticas, demografía, estadísticas actuariales y matemática actuarial. Ya que en este programa se pueden hacer cálculo de prima de seguros de vidas cálculos de rentas vitalicias, probabilidades de supervivencia y de fallecimiento, etc., Además, es un programa gratuito que todos podemos adquirir.

- Que el error más común en el programa R es escribir mal un nombre de una variable creada, por eso recomendamos hacer una lista de las variables creadas con sus significados para así no equivocarse al momento de programar.

- Es necesario que al momento de programar cuiden la ortografía, dado que R distingue entre mayúsculas y minúsculas; así como en la aplicación de tildes, ya que al hacer caso omiso a estas indicaciones le produciría errores a la hora de correr el programa.

A las Compañías Aseguradoras:

- Actualizar y capacitar a su área técnica en el manejo del programa R, ya que este evita realizar cálculos repetitivos y cansados, dado que los actuarios y la tecnología podemos adaptarnos para generar mayor aprovechamiento del tiempo y reducción de las cargas de trabajo.

- Que las compañías de seguros utilicen los modelos de distribución de probabilidad discretas para estimar en número de siniestros esperados y puedan además establecer sus respectivas reservas. Además, que esta distribución consta de pocos parámetros de estimación.

A la comisión curricular:

- Romper la zona de confort y no limitarse a lo ya aprendido e incentivarse en la adquisición de nuevos conocimientos para generar nuevas soluciones. Diseñar y proponer un componente de programación en R y VBA de Excel o cualquier otro paquete de programación, como componente electivo y si fuese obligatorio mejor.

VIII. BIBLIOGRAFIA

- Casas Sánchez, José Miguel; Estadística I: Probabilidad y Distribución, Editorial Centro de Estudios Ramón Areces, S.A, Madrid, 2000.

- Levin, Rubin Balderas, Del Valle, Gómez; Estadísticas para Administradores y Economistas, VII Edición, Pearson Educación, México 2004.

- Walple Ronald, Myers Raymond, Myers Sharon, Probabilidad y Estadísticas para Ingeniera y Ciencias, IX Edición, Pearson Educación, México, 2012.

- Hossack, J. Pollard, B. Zehnwirth, Introducción a la estadística con aplicaciones a los seguros generales, Editorial Mapfre S.A, Madrid, 2001.

- Charpentier Arthur, Computational Actuarial Science With R, I Edición, Editorial CRC Press, Canada, 2015.

- Gonzalez Andres y Gonzalez Silvia, Introducción a R, R development core team, Versión 1.0.1, 2000-05-16

Referencias electrónicas:

- Urcera J., publicado el 15 de Junio del 2012, floss square in freedom city, R y RStudio, instalación y primeros pasos. http://blog.urcera.com/wordpress/?p=242

- Casco Fernández, Ignacio; 2009, Modelos de distribuciones discretas y continuas, Universidad Caros III de Madrid. http://halweb.uc3m.es/esp/personal/personas/icascos/esp/resumen_distribuciones.pdf

- Garcia Martha, 2006, Teoría de la probabilidad. http://www.dm.uba.ar/materias/estadistica_Q/2006/2/teoricas/EstadQuimProbabilidad.pdf

- Bates, Douglas; R Foundation. Fundado por los miembros del R Core Development Team. https://www.r-project.org/about.html

IX. ANEXOS

1. Variables Creadas

1.1. Tablas

- **cartera** = Tabla de las 5 carteras de seguros
- **cp** = Tabla de las 5 carteras de seguros con las frecuencias teóricas de poisson
- **cbineg** = Tabla de las 5 carteras de seguros con las frecuencias teóricas binomial negativa
- **cpoig** = Tabla de las 5 carteras de seguros con las frecuencias teóricas poisson inversa gaussiana

1.2. Media

- **med.s.d.r** = Media de la cartera de seguro de robo
- **med.s.d.a** = Media de la cartera de seguro de automóvil
- **med.s.d.h** = Media de la cartera de seguro de hogar
- **med.s.d.r.c** = Media de la cartera de seguro de responsabilidad civil
- **med.s.d.i** = Media de la cartera de seguro de incendio

1.3. Varianza

- **var.s.d.r** = Varianza de la cartera de seguro de robo
- **var.s.d.a** = Varianza de la cartera de seguro de automóvil
- **var.s.d.h** = Varianza de la cartera de seguro de hogar
- **var.s.d.r.c** = Varianza de la cartera de seguro de responsabilidad civil
- **var.s.d.i** = Varianza de la cartera de seguro de incendio

1.4. Alfa Binomial Negativa

- **alfa.bineg.s.d.r** = Alfa de la distribución Binomial Negativa de la cartera de seguro de robo
- **alfa.bineg.s.d.a** = Alfa de la distribución Binomial Negativa de la cartera de seguro de automóvil
- **alfa.bineg.s.d.h** = Alfa de la distribución Binomial Negativa de la cartera de seguro de hogar
- **alfa.bineg.s.d.r.c** = Alfa de la distribución Binomial Negativa de la cartera de seguro de responsabilidad civil
- **alfa.bineg.s.d.i** = Alfa de la distribución Binomial Negativa de la cartera de seguro de incendio

1.5. Beta Binomial Negativa

- **beta.bineg.s.d.r** = Beta de la distribución Binomial Negativa de la cartera de seguro de robo
- **beta.bineg.s.d.a** = Beta de la distribución Binomial Negativa de la cartera de seguro de automóvil

- **beta.bineg.s.d.h** = Beta de la distribución Binomial Negativa de la cartera de seguro de hogar
- **beta.bineg.s.d.r.c** = Beta de la distribución Binomial Negativa de la cartera de seguro de responsabilidad civil
- **beta.bineg.s.d.i** = Beta de la distribución Binomial Negativa de la cartera de seguro de incendio

1.6. Beta Poisson Inversa Gaussiana

- **beta.poig.s.d.r** = Beta de la distribución poisson inversa gaussiana de la cartera de seguro de robo
- **beta.poig.s.d.a** = Beta de la distribución poisson inversa gaussiana de la cartera de seguro de automóvil
- **beta.poig.s.d.h** =Beta de la distribución poisson inversa gaussiana de la cartera de seguro de hogar
- **beta.poig.s.d.r.c** = Beta de la distribución poisson inversa gaussiana de la cartera de seguro de responsabilidad civil
- **beta.poig.s.d.i** = Beta de la distribución poisson inversa gaussiana de la cartera de seguro de incendio.

1.7. Probabilidad Poisson

- **prpois.s.d.r** = Probabilidad de la distribución de poisson para la cartera de seguro de robo
- **prpois.s.d.a** = Probabilidad de la distribución de poisson para la cartera de seguro de automóvil
- **prpois.s.d.h** = Probabilidad de la distribución de poisson para la cartera de seguro de hogar
- **prpois.s.d.r.c** = Probabilidad de la distribución de poisson para la cartera de seguro de responsabilidad civil
- **prpois.s.d.i** = Probabilidad de la distribución de poisson para la cartera de seguro de incendio

1.8. Probabilidad Binomial Negativa

- **prbineg.s.d.r** = Probabilidad de la distribución binomial negativa para la cartera de seguro de robo
- **prbineg.s.d.a** = Probabilidad de la distribución binomial negativa para la cartera de seguro de automóvil
- **prbineg.s.d.h** = Probabilidad de la distribución binomial negativa para la cartera de seguro de hogar
- **prbineg.s.d.r.c** = Probabilidad de la distribución binomial negativa para la cartera de seguro de responsabilidad civil
- **prbineg.s.d.i** = Probabilidad de la distribución binomial negativa para la cartera de seguro de incendio

1.9. Probabilidad Poisson Inversa Gaussiana

- **prpoig.s.d.r** = Probabilidad de la distribución poisson inversa gaussiana para la cartera de seguro de robo
- **prpoig.s.d.a** = Probabilidad de la distribución poisson inversa gaussiana para la cartera de seguro de automóvil
- **prpoig.s.d.h** = Probabilidad de la distribución poisson inversa gaussiana para la cartera de seguro de hogar
- **prpoig.s.d.r.c** = Probabilidad de la distribución poisson inversa gaussiana para la cartera de seguro de responsabilidad civil
- **prpoig.s.d.i** = Probabilidad de la distribución poisson inversa gaussiana para la cartera de seguro de incendio

1.10. Frecuencia Teórica Poisson

- **f.t.pois.s.d.r** = Frecuencia teórica de la distribución de poisson para la cartera de seguro de robo
- **f.t.pois.s.d.a** = Frecuencia teórica de la distribución de poisson para la cartera de seguro de automóvil
- **f.t.pois.s.d.h** = Frecuencia teórica de la distribución de poisson para la cartera de seguro de hogar
- **f.t.pois.s.d.r.c** = Frecuencia teórica de la distribución de poisson para la cartera de seguro de responsabilidad civil
- **f.t.pois.s.d.i** = Frecuencia teórica de la distribución de poisson para la cartera de seguro de incendio

1.11. Frecuencia Teórica Binomial Negativa

- **f.t.bineg.s.d.r** = Frecuencia teórica de la distribución binomial negativa para la cartera de seguro de robo
- **f.t.bineg.s.d.a** = Frecuencia teórica de la distribución binomial negativa para la cartera de seguro de automóvil
- **f.t.bineg.s.d.h** = Frecuencia teórica de la distribución binomial negativa para la cartera de seguro de hogar
- **f.t.bineg.s.d.r.c** = Frecuencia teórica de la distribución binomial negativa para la cartera de seguro de responsabilidad civil
- **f.t.bineg.s.d.i** = Frecuencia teórica de la distribución binomial negativa para la cartera de seguro de incendio

1.12. Frecuencia Teórica Poisson Inversa Gaussiana

- **f.t.poig.s.d.r** = Frecuencia teórica poisson inversa gaussiana para la cartera de seguro de robo
- **f.t.poig.s.d.a** = Frecuencia teórica poisson inversa gaussiana para la cartera de seguro de automóvil
- **f.t.poig.s.d.h** = Frecuencia teórica poisson inversa gaussiana para la cartera de seguro de hogar

- **f.t.poig.s.d.r.c** = Frecuencia teórica poisson inversa gaussiana para la cartera de seguro de responsabilidad civil
- **f.t.poig.s.d.i** = Frecuencia teórica poisson inversa gaussiana para la cartera de seguro de incendio

1.13. Reagrupación de las frecuencias observadas de poisson

- **s.d.r.pois.analisis** = Reagrupación de las frecuencias observadas del seguro de robo para el análisis ajuste de bondad
- **s.d.a.pois.analisis** = Reagrupación de las frecuencias observadas del seguro de automovil para el análisis ajuste de bondad
- **s.d.h.pois.analisis** = Reagrupación de las frecuencias observadas del seguro de hogar para el análisis ajuste de bondad
- **s.d.r.c.pois.analisis** = Reagrupación de las frecuencias observadas del seguro de responsabilidad civil para el análisis ajuste de bondad
- **s.d.i.pois.analisis** = Reagrupación de las frecuencias observadas del seguro de incendio para el análisis ajuste de bondad

1.14. Reagrupación de las frecuencias observadas binomial negativa

- **s.d.r.bineg.analisis** = Reagrupación de las frecuencias observadas del seguro de robo para el análisis ajuste de bondad
- **s.d.a.bineg.analisis** = Reagrupación de las frecuencias observadas del seguro de automóvil para el análisis ajuste de bondad
- **s.d.h.bineg.analisis** = Reagrupación de las frecuencias observadas del seguro de hogar para el análisis ajuste de bondad
- **s.d.r.c.bineg.analisis** = Reagrupación de las frecuencias observadas del seguro de responsabilidad civil para el análisis ajuste de bondad
- **s.d.i.bineg.analisis** = Reagrupación de las frecuencias observadas del seguro de incendio para el análisis ajuste de bondad

1.15. Reagrupación de las frecuencias observadas poisson inversa gaussiana

- **s.d.r.poig.analisis** = Reagrupación de las frecuencias observadas del seguro de robo para el análisis ajuste de bondad
- **s.d.a.poig.analisis** = Reagrupación de las frecuencias observadas del seguro de automóvil para el análisis ajuste de bondad
- **s.d.h.poig.analisis** = Reagrupación de las frecuencias observadas del seguro de hogar para el análisis ajuste de bondad
- **s.d.r.c.poig.analisis** = Reagrupación de las frecuencias observadas del seguro de responsabilidad civil para el análisis ajuste de bondad
- **s.d.i.poig.analisis** = Reagrupación de las frecuencias observadas del seguro de incendio para el análisis ajuste de bondad

1.16. Reagrupación de las frecuencias teóricas de poisson

- **f.t.pois.s.d.r.analisis** = Reagrupación de las frecuencias teóricas poisson del seguro de robo para el análisis de ajuste de bondad

- **f.t.pois.s.d.a.analisis** = Reagrupación de las frecuencias teóricas poisson del seguro de automóvil para el análisis de ajuste de bondad
- **f.t.pois.s.d.h.analisis** = Reagrupación de las frecuencias teóricas poisson del seguro de hogar para el análisis de ajuste de bondad
- **f.t.pois.s.d.r.c.analisis** = Reagrupación de las frecuencias teóricas poisson del seguro de responsabilidad civil para el análisis de ajuste de bondad
- **f.t.pois.s.d.i.analisis** = Reagrupación de las frecuencias teóricas poisson del seguro de incendio para el analisis de ajuste de bondad.

1.17. Reagrupación de las frecuencias teóricas binomial negativa

- **f.t.bineg.s.d.r.analisis** = Reagrupación de las frecuencias teóricas binomial negativa del seguro de robo para el análisis de ajuste de bondad
- **f.t.bineg.s.d.a.analisis** = Reagrupación de las frecuencias teóricas binomial negativa del seguro de automóvil para el análisis de ajuste de bondad
- **f.t.bineg.s.d.h.analisis** = Reagrupación de las frecuencias teóricas binomial negativa del seguro de hogar para el análisis de ajuste de bondad
- **f.t.bineg.s.d.r.c.analisis** = Reagrupación de las frecuencias teóricas binomial negativa del seguro de responsabilidad civil para el análisis de ajuste de bondad
- **f.t.bineg.s.d.i.analisis** = Reagrupación de las frecuencias teóricas binomial negativa del seguro de incendio para el análisis de ajuste de bondad

1.18. Reagrupación de las frecuencias teóricas poisson inversa gaussiana

- **f.t.poig.s.d.r.analisis** = Reagrupación de las frecuencias teóricas poisson inversa gaussiana del seguro de robo para el análisis de ajuste de bondad
- **f.t.poig.s.d.a.analisis** = Reagrupación de las frecuencias teóricas poisson inversa gaussiana del seguro de automóvil para el análisis de ajuste de bondad
- **f.t.poig.s.d.h.analisis** = Reagrupación de las frecuencias teóricas poisson inversa gaussiana del seguro de hogar para el análisis de ajuste de bondad
- **f.t.poig.s.d.r.c.analisis** = Reagrupación de las frecuencias teóricas poisson inversa gaussiana del seguro de responsabilidad civil para el análisis de ajuste de bondad
- **f.t.poig.s.d.i.analisis** = Reagrupación de las frecuencias teóricas poisson inversa gaussiana del seguro de incendio para el análisis de ajuste de bondad

1.19. Análisis de Ajuste de Bondad para la distribución de poisson

- **analisis.ajuste.pois.s.d.r** = Análisis de ajuste de bondad de la distribución de poisson para la cartera de seguro de robo
- **analisis.ajuste.pois.s.d.a** = Análisis de ajuste de bondad de la distribución de poisson para la cartera de seguro de automóvil
- **analisis.ajuste.pois.s.d.h** = Análisis de ajuste de bondad de la distribución de poisson para la cartera de seguro de hogar
- **analisis.ajuste.pois.s.d.r.c** = Análisis de ajuste de bondad de la distribución de poisson para la cartera de seguro de responsabilidad civil
- **analisis.ajuste.pois.s.d.i** = Análisis de ajuste de bondad de la distribucion de poisson para la cartera de seguro de incendio

1.20. Analisis de Ajuste de Bondad para la distribución binomial negativa

- **analisis.ajuste.bineg.s.d.r** = Análisis de ajuste de bondad de la distribución binomial negativa para la cartera de seguro de robo
- **analisis.ajuste.bineg.s.d.a** = Análisis de ajuste de bondad de la distribución binomial negativa para la cartera de seguro de automóvil
- **analisis.ajuste.bineg.s.d.h** = Análisis de ajuste de bondad de la distribución binomial negativa para la cartera de seguro de hogar
- **analisis.ajuste.bineg.s.d.r.c** = Análisis de ajuste de bondad de la distribución binomial negativa para la cartera de seguro de responsabilidad civil
- **analisis.ajuste.bineg.s.d.i** = Análisis de ajuste de bondad de la distribución binomial negativa para la cartera de seguro de incendio

1.21. Análisis de Ajuste de Bondad para la distribución poisson inversa gaussiana

- **analisis.ajuste.poig.s.d.r** = Análisis de ajuste de bondad de la distribución poisson inversa gaussiana para la cartera de seguro de robo
- **analisis.ajuste.poig.s.d.a** = Análisis de ajuste de bondad de la distribución poisson inversa gaussiana para la cartera de seguro de automóvil
- **analisis.ajuste.poig.s.d.h** = Análisis de ajuste de bondad de la distribución poisson inversa gaussiana para la cartera de seguro de hogar
- **analisis.ajuste.poig.s.d.r.c** = Análisis de ajuste de bondad de la distribución poisson inversa gaussiana para la cartera de seguro de responsabilidad civil
- **analisis.ajuste.poig.s.d.i** = Análisis de ajuste de bondad de la distribución poisson inversa gaussiana para la cartera de seguro de incendio.

1.22. Función de Ajuste de Bondad

```
funcion.ajuste = function(s.d.x.analisis,f.t.pois.s.d.x.analisis,par) {

D = sum((s.d.x.analisis-f.t.pois.s.d.x.analisis)^2/f.t.pois.s.d.x.analisis)

grad = length(s.d.x.analisis)-par-1

pvalue = 1-pchisq(D,grad)

print(D)

print(pvalue)

}
```