UNIVERSIDAD NACIONAL AUTONOMA
DE MEXICO

DIVISIÓN DE ESTUDIOS DE POSGRADO

FACULTAD DE INGENIERIA

USO DE LA ESTIMACION DE LA DISTRIBUCION DE PROBABILIDAD PARA MUESTRAS PEQUEÑAS Y DE LA SIMULACION EN LA INFERENCIA DE CARTERAS DE SEGUROS

T E S I S

QUE PARA OBTENER EL GRADO DE

MAESTRO EN INGENIERÍA

(INVESTIGACION DE OPERACIONES)

P R E S E N T A :

JUAN CARLOS VARGAS AGUILAR

DRECTORA DE TESIS: M.I. ISABEL PATRICIA AGUILAR JUAREZ

CD. UNIVERSITARIA, MEXICO, D. F. JUNIO 2005.

A mis padres, por su cariño y amor.

A mis hermanos Elizabeth y Osvaldo, por todos los momentos felices que hemos pasado juntos.

AGRADECIMIENTOS

A Dios por darme la oportunidad de estudiar.

A la M. I. Isabel Patricia Aguilar Juárez mi directora de tesis, por compartirme sus conocimientos, su tiempo, su paciencia y su amistad.

A los sinodales, por su apoyo y comentarios que alentaron este trabajo que hoy se concreta.

Al Consejo Nacional de Ciencia y Tecnología (CONACYT), por el apoyo que me proporcionó durante el período que curse mis estudios de posgrado.

A la Universidad Nacional Autónoma de México, por haberme permitido ser alumno de esta Máxima Casa de Estudios.

RESEÑA

Título: Uso de la Estimación de la Distribución de Probabilidad para Muestras Pequeñas y de la Simulación en la Inferencia de Carteras de Seguros.

Objetivos:

Objetivo General

Mostrar las ventajas del uso de las técnicas de estimación de distribuciones empíricas para muestras pequeñas y de la simulación en la inferencia de carteras de seguros. Aportando así, un método para hacer inferencias de la cartera a pesar de que los datos de ésta no cumplan el supuesto de provenir de una muestra suficientemente grande. Finalmente, ilustrar la aplicación de la metodología propuesta mediante un ejemplo en el que se determina la cuota base para el seguro de Responsabilidad Civil Agentes de Seguros.

Objetivos Particulares

Presentar a los elementos necesarios que emplea la técnica propuesta por este trabajo:

- Estadística descriptiva.
- Distribuciones de probabilidad para variables aleatorias discretas y continuas.
- Construcción de distribuciones empíricas para muestras pequeñas.
- Pruebas de bondad de ajuste.
- Determinación del tamaño de muestra.
- Modelado de sistemas.
- Simulación.
- Ilustrar la metodología propuesta mediante un caso práctico.

Problemática en el Sector Asegurador: La probabilidad empírica o ley de los grandes números, uno de los conceptos más fuertes en la teoría del seguro, asume la existencia de una muestra suficientemente grande pero, ¿cómo debe proceder el asegurador cuando su información estadística es insuficiente? o ¿cómo corregir ésta deficiencia para hacer inferencias que causen impacto positivo en la toma de decisiones o en la elaboración de tarifas de seguros?

Importancia para el Sector Asegurador: El presente trabajo representa una alternativa de solución a la problemática detallada en el punto anterior, permitiendo el desarrollo de inferencias que causen impacto positivo en la toma de decisiones o en la elaboración de tarifas de seguros.

ÍNDICE

Página

Página

[1] Capítulo con numeración especial.

Página

INTRODUCCIÓN

Día a día en el campo del seguro surgen nuevas necesidades de protección, mismas que provocan que quienes laboran en esta industria tengan que trabajar con pocos datos y con base en éstos estudiar los riesgos a cubrir, tomar decisiones, planear y calcular tarifas que permitan determinar el costo o la prima del seguro en cuestión.

Sin embargo, si los profesionistas del seguro no utilizan bases técnicas para obtener el mejor provecho de estos datos insuficientes, el seguro se convierte en un juego de azar que puede colapsar la estabilidad financiera de la empresa y consecuentemente impedir las obligaciones de ésta con el público usuario de estos servicios.

En busca de dar solución a la problemática antes mencionada se elabora la presente tesis que conjunta el uso de dos técnicas, la estimación de la distribución de probabilidad para muestras pequeñas primero, y la simulación de eventos discretos después, mismas que permiten la elaboración de inferencias de carteras de seguros que causen impacto positivo a las entidades aseguradoras.

Previo a la realización de esta tesis, se observó que en nuestro país no existen escritos acerca de la aplicación de estos métodos en los campos del seguro, por lo que con el fin de hacer este trabajo accesible al mayor número de profesionistas del seguro, así como a los estudiantes de este tipo de técnicas, se decidió que este trabajo se desarrollara de manera gradual; tratando que los conceptos y los métodos sean claros y concisos, apoyándose en algunos casos en la exposición de ejemplos.

Este escrito se divide en cinco capítulos, en el primero se estudiará la colección, la agrupación y el análisis previo de datos, es decir lo que se conoce como estadística descriptiva.

En el segundo capítulo se hablará de la probabilidad, de las variables aleatorias y finalmente de las distribuciones de probabilidad, ya que sería muy difícil utilizarlas sin saber qué son.

En el tercer capítulo se estudiará el método del rango mediano; ya que éste permite estimar una distribución de probabilidad a partir de un conjunto de datos pequeño o limitado; y se mostrarán el uso de dos pruebas importantes en estadística que permiten comprobar la bondad de los ajustes de las distribuciones estimadas por este método; además se estudiará cómo determinar el tamaño de muestra ya que la simulación es una metodología que realiza experimentos de muestreo mediante un modelo del sistema.

En el cuarto capítulo se hablará del modelado de sistemas y de la simulación de eventos discretos.

En el quinto capítulo se ejemplificará la metodología propuesta que además se conjunta con la matemática del seguro mediante la elaboración de una nota técnica, en la que se determinarán las cuotas base para el seguro de responsabilidad civil agentes de seguros (personas físicas); producto del cual existe o existía muy poca información estadística y que además que en años anteriores a que se volviera un seguro obligatorio se observaba la falta de homogeneidad en el sector respecto a los precios calculados por diferentes compañías, lo que pone en tela de juicio las metodologías utilizadas por éstas para calcular sus tarifas ante la carencia de datos.

Finalmente, se realizarán las conclusiones de esta tesis.

Por lo tanto, este trabajo desea mostrar de manera gradual, clara y concisa las ventajas del uso de la estimación de distribuciones empíricas para muestras pequeñas y de la simulación en carteras de seguros. Aportando así, un método para hacer inferencias de la cartera a pesar de que los datos de ésta no cumplan el supuesto de provenir de una muestra suficientemente grande.

CAPÍTULO 1
ESTADÍSTICA DESCRIPTIVA

El estudio de la estadística se puede dividir en tres partes: reunir datos, analizarlos y, a partir de ellos, hacer inferencias. En este capítulo se tratará principalmente, el análisis de datos, es decir, la organización y el reporte de los mismos; ya que la elaboración de éste, es un requisito indispensable para hacer inferencias del conjunto de datos estudiado.

El presente capítulo, se integra de siete subcapítulos, los seis primeros explican algunos conceptos y métodos de la estadística descriptiva; mientras que en el último se ilustra mediante un ejemplo lo estudiado en los anteriores.

1.1 Estadística Descriptiva e Inferencia Estadística[1]

"La *estadística descriptiva* comprende las técnicas que se emplean para resumir y describir datos numéricos".[2] Estos métodos pueden ser gráficos o implicar el cálculo de medidas numéricas.

"La *inferencia estadística* trata de generalizaciones basadas en muestras de datos".[3]

1.2 Población y Muestra

Antes de estudiar descripciones estadísticas particulares, permitamos hacer la siguiente diferencia.

> La población es la colección de toda la posible información que caracteriza un fenómeno. En estadística, la población es un concepto mucho más general del que tiene la acepción común de esta palabra. En este sentido, una *población* es cualquier colección ya sea de un número finito de mediciones o una colección grande, virtualmente infinita, de datos acerca

[1] Algunas técnicas de inferencia estadística serán estudiadas en el capítulo 3.
[2] KAZMIER, Leonard J., Estadística Aplicada a la Administración y a la Economía, p. 1.
[3] MILLER, Irwin, John E. FREUD y Richard A. JOHNSON, Probabilidad y Estadística para Ingenieros, p. 2.

de algo de interés. Por otro lado, la *muestra* es un subconjunto representativo seleccionado de una población.[4]

Población		**Muestra**
Tamaño N	→	Tamaño n

$n < N$

1.3 Tipo de Datos

Los datos pueden ser *cuantitativos*, con valores expresados numéricamente, o *cualitativos* en cuyo caso se tabulan las características de las observaciones.[5]

1.4 Distribución de Frecuencias

Es una técnica usual en la estadística que permite el análisis de conjuntos grandes de datos.[6]

1.4.1 Tabla de Distribución de Frecuencias

"Una *tabla de distribución de frecuencias* es una clasificación de los datos (numéricos) en clases o categorías de acuerdo a sus valores".[7]

Los datos organizados en una distribución de frecuencias se llaman *datos agrupados*. En contraste con ello, en el caso de los *datos no agrupados* se enlistan todos los valores observados.[8]

Un ejemplo típico de una tabla de distribución de frecuencias es el que se muestra en el subcapítulo 1.7.

---

[4] CANAVOS C. George, Probabilidad y Estadística Aplicaciones y Métodos, p. 1.
[5] KAZMIER, *op. cit.* p. 1
[6] *Cfr.* AGUILAR Juárez, Isabel Patricia y Leonardo BAÑUELOS Saucedo, Notas del Curso Propedéutico de Probabilidad y Estadística, p. 1–2.
[7] *Ibid.* p. 2.
[8] KAZMIER, *op. cit.* p. 9.

El contenido de una tabla completa de distribución de frecuencias, así como los elementos necesarios para su elaboración se explican en el siguiente subcapítulo.

Si se acepta que en la construcción de una tabla de distribución de frecuencias se realiza una clasificación de datos, resulta indispensable contar primeramente con el criterio de clasificación a utilizar, mismo que se define a través de los límites de clase o mediante las fronteras de la clase.

1.4.2 Intervalos de Clase

"El *intervalo de clase* identifica el rango de valores incluidos dentro de una clase y puede determinarse restando el límite exacto de la clase superior el límite exacto de la clase inferior".[9]

Algunos autores manejan que el número, k, de intervalos que se deben considerar en una tabla de distribución de frecuencias, está dado por la fórmula empírica que a continuación se presenta:[10]

$$k \approx \begin{cases} \sqrt{n} & \text{si n no es muy grande.} \\ 1 + 3.22\log(n) & \text{en otro caso.} \end{cases} \qquad (1)$$

Para la construcción de una tabla de distribución de frecuencias es conveniente tomar en consideración las siguientes recomendaciones empíricas: [11]

1. La tabla de distribución de frecuencias constará de entre 5 y 20 clases.
2. Todas las clases serán de la misma longitud.

Para efectos de cálculo, considerando la recomendación 2, la siguiente fórmula puede emplearse para determinar el intervalo de clase aproximado por usar:[12]

---

[9] *Idem.*
[10] MARÍN Diazaraque, Juan Miguel, Apuntes de Estadística – Estadística Descriptiva, p. 8.
[11] AGUILAR Juárez, Isabel Patricia y Leonardo BAÑUELOS Saucedo, *op. cit.* p. 4.
[12] KAZMIER, *op. cit.* p. 9.

$$\text{Intervalo aproximado} = \frac{[\text{mayor valor en datos no agrupados}] - [\text{menor valor en datos no agrupados}]}{\text{número de clases deseadas}} \quad (2)$$

1.4.2.1 Límites Nominales de Clase

"Los *límites nominales de clase* inferior y superior indican los valores incluidos dentro de la clase".[13]

Los límites de clase nominales o simplemente de clase tendrán la misma aproximación que los datos y el límite superior de una clase diferirá del límite inferior de la clase siguiente, en una unidad de aproximación, es decir:[14]

Datos	**Límites**	**Diferencia** (lim. inf. de la clase siguiente – lim. sup. de la clase)
Enteros	Enteros	1
Décimas	Décimas	0.1
Centésimas	Centésimas	0.01

1.4.2.2 Fronteras de Clase

Las *fronteras de clase* o *límites exactos de clase*, son los puntos específicos que sirven para separar clases adyacentes en una escala de medición de variables continuas. Las fronteras de clase pueden determinarse identificando los puntos intermedios entre los límites nominales de clase superior e inferior, respectivamente de clases adyacentes.[15]

1.4.2.3 Longitud de la Clase (c)

"Se denota por c y es la diferencia entre la frontera superior y la inferior de la clase".[16]

---

[13] *Cfr.* AGUILAR Juárez, Isabel Patricia y Leonardo BAÑUELOS Saucedo, *op. cit.* p. 2.
[14] AGUILAR Juárez, Isabel Patricia y Leonardo BAÑUELOS Saucedo, *op. cit.* p. 2.
[15] *Cfr.* KAZMIER, *op. cit.* p. 9.
[16] AGUILAR Juárez, Isabel Patricia y Leonardo BAÑUELOS Saucedo, *op. cit.* p. 4.

1.4.2.4 Marca de Clase (x_i)

"Las *marcas de clase* de una distribución de frecuencias se obtienen promediando los límites nominales de clase consecutivos o las fronteras de clases sucesivas".[17]

1.4.3 Frecuencia (f_i)

Es el número de datos de la muestra que corresponden a la clase en cuestión. Para determinar la frecuencia de una clase, basta con realizar un conteo del número de observaciones en la muestra que se encuentra en ésta; ya sea por sus límites nominales o por sus fronteras, esto último es debido a que ambos (límites y fronteras) determinan exactamente la misma clasificación.[18]

1.4.4 Frecuencia Acumulada (F_i)

"Es el número de datos en la muestra cuyo valor no excede la frontera superior de la clase en cuestión. Para calcular F_i basta contabilizar las frecuencias observadas en la clase de interés y en las anteriores".[19]

1.4.5 Frecuencia Relativa (f_i^*)

"Es la proporción de los datos en la muestra que pertenecen a la clase en cuestión. Si denotamos por n al número de datos en la muestra y a i como al número de la clase, la frecuencia relativa se expresa como sigue":[20]

$$f_i^* = \frac{f_i}{n} = \frac{f_i}{\sum_{\forall i} f_i} \tag{3}$$

---

[17] *Cfr.* MILLER, Irwin, John E. FREUD y Richard A. JOHNSON, *op. cit.* p. 10.
[18] *Cfr.* AGUILAR Juárez, Isabel Patricia y Leonardo BAÑUELOS Saucedo, *op. cit.* p. 3.
[19] *Idem.*
[20] *Idem.*

1.4.6 Frecuencia Relativa Acumulada (F_i*)

"Es la proporción de datos en la muestra que no exceden la frontera de la clase en cuestión".[21]

$$F_i^* = \frac{F_i}{n} = \frac{F_i}{\sum_{\forall i} f_i} \tag{4}$$

1.5 Descripción Grafica de los Datos

Las propiedades de las distribuciones de frecuencias relacionadas con su forma se hacen más evidentes por medio de graficas, y en este subcapítulo introduciremos algunas formas más comunes de representar gráficamente la información proporcionada por los datos.

1.5.1 Histograma de Frecuencias

"El *histograma de frecuencias* se construye con rectángulos adyacentes, las alturas de los rectángulos representan las frecuencias de la clase y sus bases se extienden en fronteras de clases sucesivas".[22]

Las marcas sobre la escala horizontal pueden ser los límites de la clase, las fronteras de la clase, las marcas de clase (que resulta ser lo más común) o valores claves arbitrarios.

1.5.2 Polígono de Frecuencias

En el *polígono de frecuencias* las frecuencias de clase son trazadas sobre las marcas de clase, esto es, se dibujan los puntos (x_i, f_i) donde x_i es la marca de la clase de la i–ésima clase y f_i es la frecuencia correspondiente y los puntos sucesivos se unen por

---

[21] *Idem.*
[22] MILLER, Irwin, John E. FREUD y Richard A. JOHNSON, *op. cit.* p. 12.

medio de líneas rectas después de haber agregado clases con frecuencia cero en los puntos límite de la distribución.[23]

1.5.3 Curvas de Frecuencia

"Una curva de frecuencia es un polígono de frecuencias suavizado".[24]

1.5.4 Ojiva

"Las distribuciones acumuladas por lo general se representan gráficamente en forma de *ojivas*", [25] las cuales son similares a los polígonos de frecuencias acumuladas sobre las fronteras de clase en lugar de dibujar las frecuencias ordinarias sobre las marcas de clase.

1.6 Medidas Numéricas Descriptivas

En los subcapítulos anteriores se plantearon las técnicas gráficas para describir los patrones de distribución ocultos en un conjunto de datos. En éste se definen algunas medidas numéricas que se emplean comúnmente para describir conjuntos de datos.

En estadística, una medida descriptiva de una población, o parámetros de población, se representa por lo general con alguna de las letras del alfabeto griego, mientras que una medida descriptiva de una muestra, o estadística muestral, se representa por alguna de las letras del alfabeto latino. Esto lo notará el lector más adelante.

1.6.1 Medidas de Tendencia Central[26]

La *tendencia central* de un conjunto de datos es la disposición de éstos para agruparse ya sea alrededor del centro o de ciertos valores numéricos.

---

[23] *Ibid.* p. 13.
[24] KAZMIER, *op. cit.* p. 11.
[25] MILLER, Irwin, John E. FREUD y Richard A. JOHNSON, *op. cit.* p. 15.
[26] CANAVOS, *op. cit.* p. 12.

Existen principalmente tres medidas de tendencia central: la media aritmética, la mediana y la moda. Mismas que a continuación se detallan.

1.6.1.1 Media Aritmética[27]

a) Si $x_1, x_2, ..., x_n$ son los datos contenidos en la muestra y se encuentran sin agrupar, entonces la media aritmética será:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \tag{5}$$

Donde n es el tamaño de la muestra.

b) Si los datos se encuentran agrupados en una tabla de distribución de frecuencias, y se utiliza el mismo concepto que para los datos sin agrupar, se define la media aritmética como:

$$\bar{x} = \frac{\sum_{i=1}^{m} x_i f_i}{n} = \sum_{i=1}^{m} x_i f_i^{\,*} \text{ puesto que } \frac{f_i}{n} = f_i^{\,*} \tag{6}$$

Donde:

m = Número de clases
x_i = Marca de la clase i
f_i = Frecuencia de la clase i
n = Numero total de datos

1.6.1.2 Mediana

La *mediana* de un conjunto de observaciones es el valor para el cual, cuando todas las observaciones se ordenan de manera creciente, la mitad de éstas es menor que este valor y la otra mitad mayor.

Si el número de observaciones en el conjunto es impar, la mediana es el valor de la observación que se encuentra a la mitad del conjunto ordenado. Si el número es par se considera la mediana como el promedio aritmético de los valores de las dos observaciones que se encuentran a la mitad del conjunto ordenado.

[27] AGUILAR Juárez, Isabel Patricia y Leonardo BAÑUELOS Saucedo, *op. cit.* p. 9 – 10.

1.6.1.3 Moda

La *moda* de un conjunto de observaciones es el valor de la observación que ocurre con mayor frecuencia en el conjunto.

1.6.2 Medidas de Variabilidad

"Las medidas de variabilidad o dispersión, se ocupan de la descripción de la variabilidad entre los valores".[28]

Se dispone de diversas técnicas para medir el grado de variabilidad en un conjunto de datos. Las que describiremos en este subcapítulo son el rango, la varianza, la desviación estándar y el coeficiente de variación.

1.6.2.1 Rango

"El *rango* o R, es la diferencia entre los valores más alto y más bajo incluidos en el conjunto de datos".[29] Así, cuando My representa el mayor valor del grupo y Mn al menor, el rango de datos no agrupados es:

$$R = My - Mn \tag{7}$$

1.6.2.2 Varianza

"La *varianza* de las observaciones $x_1, x_2, \ldots, x_n$ es, en esencia, el promedio del cuadrado de las distancias entre cada observación y la media del conjunto de observaciones".[30] La varianza se denota por:[31]

$$\text{Varianza de la población:} \quad \sigma^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{N} \tag{8}$$

---

[28] KAZMIER, Leonard J., *op. cit.* p. 51.
[29] *Idem.*
[30] CANAVOS, *op. cit.* p. 15
[31] Estas fórmulas son aplicables a poblaciones finitas.

Varianza de la muestra: $$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} \qquad (9)$$

1.6.2.3 Desviación Estándar

La raíz cuadrada positiva de la varianza recibe el nombre de *desviación estándar*[32] y se denota por:[33]

Desviación estándar de la población: $$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{N}} \qquad (10)$$

Desviación estándar de la muestra: $$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} \qquad (11)$$

1.6.2.4 Coeficiente de Variación (CV)

"El *coeficiente de variación* indica la magnitud relativa de la desviación estándar en comparación con la media de la distribución de las medidas, expresada como porcentaje".[34] Así, las fórmulas son:

Coeficiente de Variación de la Población: $$\frac{\sigma}{\mu} \times 100 \qquad (12)$$

Coeficiente de Variación de la Muestra: $$\frac{s}{\bar{x}} \times 100 \qquad (13)$$

[32] *Idem.*
[33] Estas formulas son aplicables a poblaciones finitas.
[34] KAZMIER, *op. cit.* p.57.

1.7 Ejemplo de Estadística Descriptiva

Un gimnasio durante los 4 primeros meses del año recolectó datos de 55 clientes referentes a la frecuencia con que se presentaron a realizar algún tipo actividad física, el número de días que cada uno de ellos se presentó se lista a continuación:

4, 27, 100, 33, 32, 57, 77, 3, 45, 79, 10, 80, 4, 47, 35, 25, 90, 38, 6, 68, 57, 46, 40, 47, 87, 67, 66, 78, 80, 92, 68, 79, 78, 100, 76, 88, 89, 94, 48, 34, 90, 90, 22, 78, 56, 100, 91, 25, 80, 23, 5, 93, 30, 90 y 45.

Se pide lo siguiente:

1. Elaborar la tabla de distribución de frecuencias.
2. Elaborar la descripción grafica de los datos.
3. Calcular las medidas numéricas descriptivas.

Solución:

1. Elaborar la tabla de distribución de frecuencias.
   a) Determinar el número k de intervalos:
   Para esto se utilizará la ecuación 1; además el número de datos n, es de 55.
   $$k \approx 1+3.22\log(n) = k = 1+3.22\log(55) = 6.60 = 7$$
   7 es un número de intervalos aceptable ya que se encuentra ente 5 y 20.
   b) Determinar el intervalo de clase aproximado:
   Para esto se utilizará la ecuación 2; además en los datos en la muestra se observa que el menor es 3 y el mayor 100.
   $$\text{Intervalo aproximado} = \frac{[\text{mayor valor en datos no agrupados}] - [\text{menor valor en datos no agrupados}]}{\text{número de clases deseadas}}$$
   entonces,
   $$\text{Intervalo aproximado} = \frac{100-3}{7} = 13.86$$
   c) Una vez lo anterior, se siguieron las definiciones proporcionadas en los subcapítulos 1.4.2.1 al 1.4.2.4 para calcular a las clases los límites nominales, las fronteras, su longitud y sus marca de clase; los conceptos 1.4.3 y 1.4.4 para calcular las frecuencias y las frecuencias absolutas, en cada una de las clases;

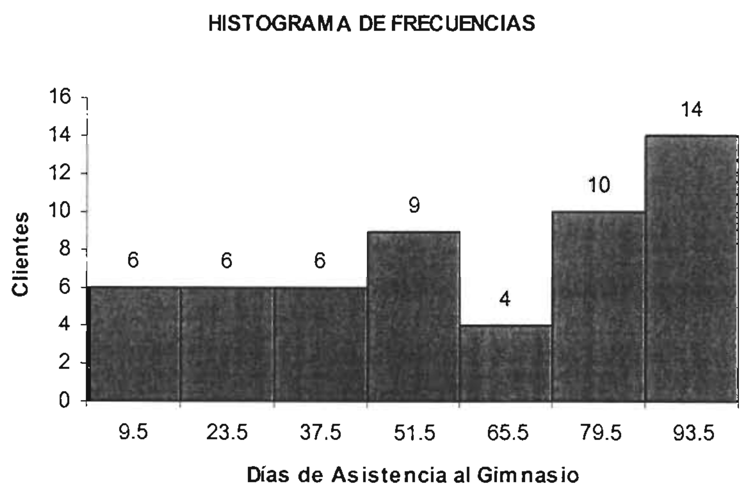
mientras que las ecuaciones 3 y 4 se utilizaron para calcularles su frecuencia relativa y su frecuencia relativa acumulada, respectivamente. Obteniéndose finalmente, la siguiente tabla de distribución de frecuencias:

TABLA DE DISTRIBUCIÓN DE FRECUENCIAS

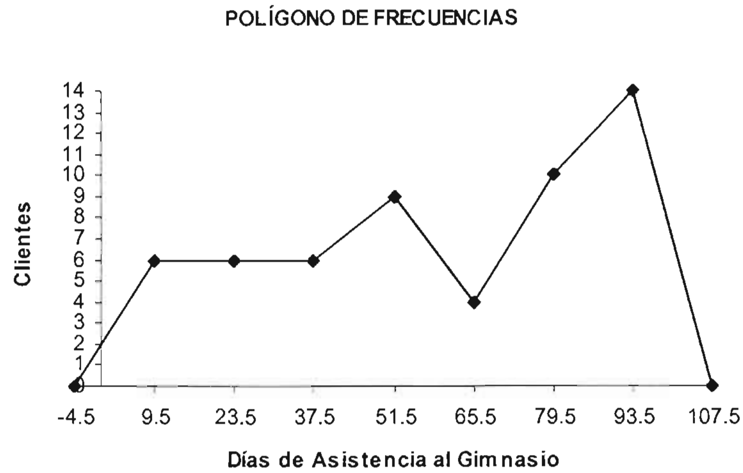
Intervalos de Clase				Marcas de Clase	Frecuencia	Frecuencia Acumulada	Frecuencia Relativa	Frecuencia Relativa Acumulada
Límites		Fronteras		x_i	f_i	F_i	fi*	F_{i^*}
3	16	2.5	16.5	9.5	6	6	0.10909	0.10909
17	30	16.5	30.5	23.5	6	12	0.10909	0.21818
31	44	30.5	44.5	37.5	6	18	0.10909	0.32727
45	58	44.5	58.5	51.5	9	27	0.16364	0.49091
59	72	58.5	72.5	65.5	4	31	0.07273	0.56364
73	86	72.5	86.5	79.5	10	41	0.18182	0.74545
87	100	86.5	100.5	93.5	14	55	0.25455	1.00000
					55		1.00000	

2. Elaborar la descripción grafica de los datos.

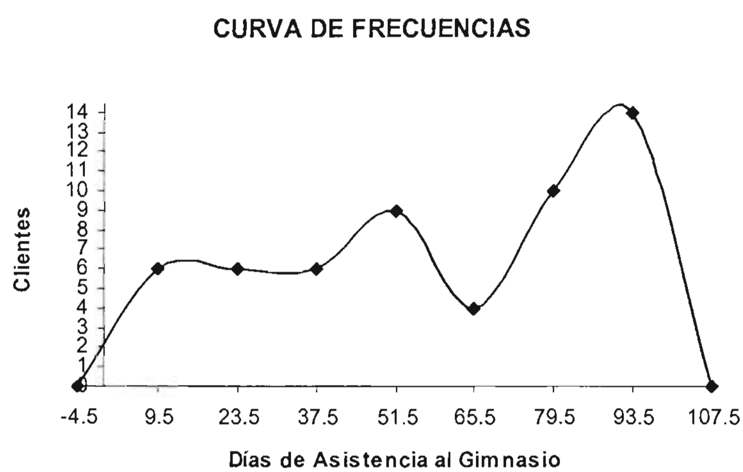
   a) Histograma de frecuencias



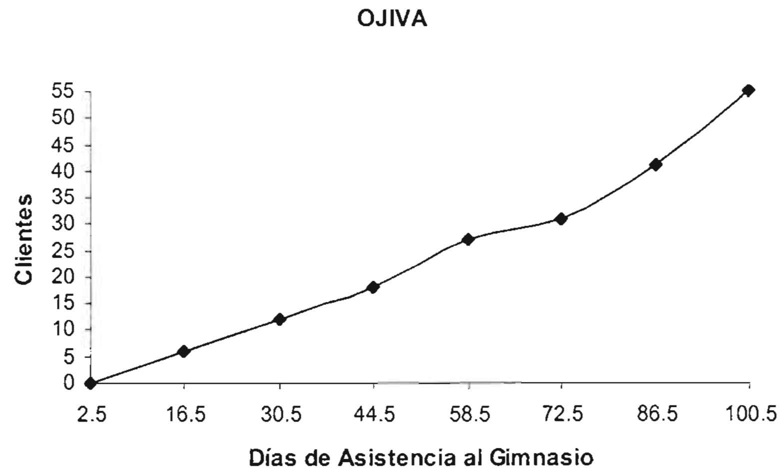
b) Polígono de frecuencias



c) Curva de Frecuencias



d) Ojiva



3. Calcular las medidas numéricas descriptivas.

a) Media Aritmética

Para calcularla se utilizó la ecuación 6, obteniéndose el siguiente resultado:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = 58.04$$

b) Mediana

Siguiendo el concepto presentado en el subcapítulo 1.6.1.2, se realiza lo siguiente: Se ordenan los datos en forma creciente, como se tiene un número de datos impar (55), la mediana será el valor de la observación que se encuentra a la mitad del conjunto ordenado es decir el dato 28:

Orden	1	2	3	4	5	6	7	8	9	10	11
Dato	3	4	4	5	6	10	22	23	25	25	27

Orden	12	13	14	15	16	17	18	19	20	21	22
Dato	30	32	33	34	35	38	40	45	45	46	47

Orden	23	24	25	26	27	**28**	29	30	31	32	33
Dato	47	48	56	57	57	**66**	67	68	68	76	77

Orden	34	35	36	37	38	39	40	41	42	43	44
Dato	78	78	78	79	79	80	80	80	87	88	89

Orden	45	46	47	48	49	50	51	52	53	54	55
Dato	90	90	90	90	91	92	93	94	100	100	100

Por lo tanto, la mediana es 66.

c) Moda

Siguiendo el concepto presentado en el subcapítulo 1.6.1.2, se obtuvo que ésta es 90, ya que es el dato con mayor frecuencia (4).

d) Rango

Para calcularlo se utilizó la ecuación 7, obteniéndose el siguiente resultado:

$$R = My - Mn = 100 - 3 = 97$$

e) Varianza

Para calcularla se utilizó la ecuación 9, obteniéndose el siguiente resultado:

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} = 898.11$$

f) Desviación Estándar

Para calcularla se utilizó la ecuación 11, obteniéndose el siguiente resultado:

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} = 29.97$$

g) Coeficiente de Variación

Para calcularlo se utilizó la ecuación 13, obteniéndose el siguiente resultado:

$$\frac{s}{\bar{x}} \times 100 = 51.64\%$$

Del ejercicio anterior, se puede decir que la representación gráfica de los datos es asimétrica, las medidas de tendencia central media aritmética, mediana y moda se calcularon en 58.04, 60 y 90, lo que confirma la carencia de simetría de los datos, además se observa dispersión alta en los datos, situación que se confirma con los resultados obtenidos por el rango, la desviación estándar y el coeficiente de variación.

En este subcapítulo se ejemplificó lo que se mostró de manera teórica en los anteriores, esperando que el lector al final de este capítulo pueda efectuar el análisis previo a la realización de inferencias; sin embargo para que éstas sean desarrolladas, es necesario profundizar en el conocimiento de las distribución de probabilidad y; por lo tanto, éste será el objetivo del capítulo siguiente.

CAPÍTULO 2
DISTRIBUCIONES DE PROBABILIDAD

En este capítulo se estudiará de manera breve qué es probabilidad, sus enfoques y su definición formal; esto porque la probabilidad tiene un papel crucial en la aplicación de la inferencia estadística.

De igual modo, se examinará qué es una variable aleatoria, concepto relacionado con eventos numéricos cuyo valor se determina por medio de un proceso aleatorio; así mismo, se diferenciará a las variables aleatorias: discretas y continuas; y a partir de estos conceptos se definirán las distribuciones de probabilidad y se estudiarán las características más importantes de éstas. Finalmente, se aprenderá sobre algunas distribuciones específicas de probabilidad: discretas y continuas; mismas que han demostrado empíricamente ser modelos útiles para diversos problemas prácticos y se discutirán algunas áreas de aplicación de cada modelo, con lo que se pretende proporcionar al lector una idea y comprensión suficiente para utilizar los modelos de manera apropiada.

2.1. Definiciones Básicas de Probabilidad[1]

Históricamente se han desarrollado tres enfoques conceptuales para definir la probabilidad y determinar valores de probabilidad: los enfoques clásico, empírico (de frecuencia relativa) y el subjetivo.

2.1.1 Definición Clásica de Probabilidad

"Dado un experimento cualquiera, que puede dar lugar a varios sucesos elementales «igualmente posibles» se define como *probabilidad* de un suceso, *A*, al cociente entre el número de sucesos favorables, m, y el número de sucesos elementales posibles n".[2]

---

[1] En este subcapítulo se manejarán intuitivamente las palabras suceso y/o evento, pero estas serán definidas de manera formal en el subcapítulo 2.2.

[2] ENCICLOPEDIA CIENTÍFICA CULTURAL – VOLUMEN ESTADÍSTICA p. 55.

$$P(A) = \frac{\text{número de sucesos favorables } A}{\text{número de sucesos posibles}} = \frac{m}{n}$$

Ejemplo 2.1.1 Sea el experimento aleatorio consistente en lanzar al aire una moneda legal, es decir que los sucesos elementales "que salga sol" y "que salga águila" son igualmente posibles.

Si se designa mediante la letra A, el suceso que salga sol, se tiene que:

$$P(A) = \frac{\text{número de sucesos favorables } A}{\text{número de sucesos posibles}} = \frac{1}{2} = 0.5$$

2.1.2 Definición Empírica de Probabilidad

Dado un experimento aleatorio cualquiera, que puede dar lugar a varios sucesos elementales, se define como probabilidad empírica de un suceso, A, a la frecuencia relativa de aparición de dicho suceso, cuando el número de observaciones crece indefinidamente.[3]

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

Ejemplo 2.1.2 Sea el experimento aleatorio del ejemplo anterior que consiste en lanzar una moneda al aire. Los dos sucesos posibles: son "que salga sol" y "que salga águila".

Si en 500 lanzamientos ha salido 263 veces sol, la probabilidad del suceso:

A = "que salga sol", será:

$$P(A) = f_i * (A) = \frac{263}{500} = 0.526$$

[3] *Cfr. Ibid*, p. 56

2.1.3 Definición Subjetiva de Probabilidad

La probabilidad de un evento es el *grado de certidumbre* que un individuo concede a la ocurrencia del evento, con base de todas las evidencias de que dispone.[4]

Ejemplo 2.1.3 Un comentarista deportivo considera que existe una probabilidad de 0.7 de que el equipo X gane el campeonato.

2.2. Desarrollo Axiomático de la Probabilidad[5]

Para formalizar la definición de probabilidad, a través de axiomas, se repasarán brevemente las definiciones de los conceptos siguientes:

Definición 2.2.1 El conjunto de todos los posibles resultados de un experimento aleatorio recibe el nombre de *espacio muestral*.

El conjunto de todos los posibles resultados puede ser finito, infinito numerable o infinito no numerable. Por ejemplo, el número de reservaciones sin cancelar de un vuelo comercial constituye un espacio muestral finito, dado que este número nunca excederá la capacidad del avión, que es finita. El número de llegadas al servicio constituye un espacio muestral infinito numerable porque es posible colocar los resultados en una correspondencia uno a uno con los enteros positivos. El peso del equipaje constituye un espacio muestral infinito innumerable. A continuación se proporcionan las siguientes definiciones.

Definición 2.2.2 Se dice que un espacio muestral es *discreto* si sus resultados pueden ponerse en correspondencia uno a uno con el conjunto de los enteros positivos.

Definición 2.2.3 Se dice que el espacio muestral es *continuo* si sus resultados consisten en un intervalo de números reales.

---

[4] *Cfr.* KAZMIER, Leonard J., Estadística Aplicada a la Administración y a la Economía, p. 65.
[5] CANAVOS C. George, Probabilidad y Estadística Aplicaciones y Métodos, p. 32 – 36.

Ejemplo 2.2.1 Considerando la representación de las fichas de un domino abajo mostrada, diga qué tipo de espacio muestral es éste.

$$\frac{0}{0} \quad \frac{0}{1} \quad \frac{0}{2} \quad \frac{0}{3} \quad \frac{0}{4} \quad \frac{0}{5} \quad \frac{0}{6}$$

$$\frac{1}{1} \quad \frac{1}{2} \quad \frac{1}{3} \quad \frac{1}{4} \quad \frac{1}{5} \quad \frac{1}{6}$$

$$\frac{2}{2} \quad \frac{2}{3} \quad \frac{2}{4} \quad \frac{2}{5} \quad \frac{2}{6}$$

$$\frac{3}{3} \quad \frac{3}{4} \quad \frac{3}{5} \quad \frac{3}{6}$$

$$\frac{4}{4} \quad \frac{4}{5} \quad \frac{4}{6}$$

$$\frac{5}{5} \quad \frac{5}{6}$$

$$\frac{6}{6}$$

Solución: Discreto.

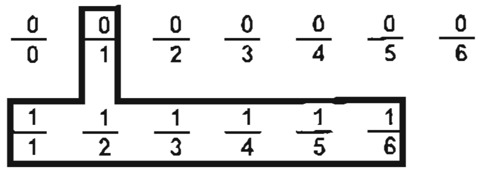
Con respecto a los resultados de un espacio muestral, se puede estar interesado en un subconjunto de éstos. De esta manera se tienen las siguientes definiciones.

Definición 2.2.4 Un *evento* del espacio muestral es un grupo de resultados contenidos en este espacio, cuyos miembros tiene una característica común.

Por característica común debe entenderse que únicamente un grupo de resultados en particular satisface la característica y los restantes, contenidos en el espacio muestral, no; Se dice que un evento ha ocurrido si los resultados del experimento aleatorio incluyen a algunos de los que definen al evento. En este contexto, el espacio muestral, evento en si mismo, puede entenderse como un evento seguro, puesto que se tiene un 100% de certidumbre de que ocurrirá un resultado del espacio muestral cuando el experimento se lleve a cabo.

Ejemplo 2.2.4 ¿De cuántos resultados consta el evento de que al seleccionar una ficha en el dominó del ejemplo 2.2.1 aparezca un uno?

Solución: De siete y son los siguientes:



Definición 2.2.5 El evento que contiene a ningún resultado del espacio muestral recibe el nombre de *evento nulo* o *vacío*.

El evento nulo o vacío se representa mediante el siguiente símbolo: Φ

Ejemplo 2.2.5 De un ejemplo de un evento nulo, considerando el domino del ejemplo 2.2.1.

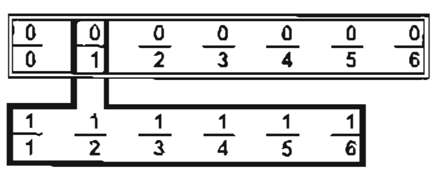
Solución: El vacio.

Sean E_1 y E_2 cualesquiera dos eventos que se encuentren en un espacio muestral dado denotado por S. Valiéndose de estos se indican las siguientes definiciones:

Definición 2.2.6 El evento formado por todos los posibles resultados en E_1 o E_2 o en ambos, recibe el nombre de la *unión* de E_1 y E_2 y se denota por $E_1 \cup E_2$.

Ejemplo 2.2.6 Considere el domino del ejemplo 2.2.1, ¿De cuántos elementos está formado $E_1 \cup E_2$ si E_1 es el conjunto de todas las fichas que tienen 0 y E_2 las fichas que tienen 1?

Solución: De 13 elementos y son los siguientes:



Definición 2.2.7 El evento formado por todos los resultados comunes tanto a E_1 como a E_2 o en ambos, recibe el nombre de la *intersección* de E_1 y E_2 y se denota por $E_1 \cap E_2$.

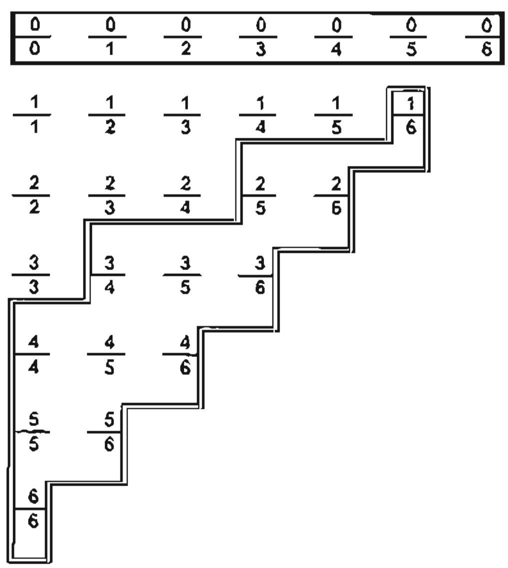
Ejemplo 2.2.7 ¿De cuántos elementos está formado $E_1 \cap E_2$ si E_1 y E_2 son los eventos que se describen en el ejemplo anterior?
Solución: De un elemento.

Definición 2.2.8 Se dice que los eventos E_1 y E_2 son *mutuamente excluyentes* o *disjuntos* si no tienen resultados en común; en otras palabras $E_1 \cap E_2 = \Phi \equiv$ evento vacío.

Ejemplo 2.2.8 Utilizando el dominó del ejemplo 2.2.1, diga dos eventos mutuamente excluyentes.
Solución:

E_1: El conjunto de todas las fichas con 0.

E_2: El conjunto de fichas en las que los puntos de cada una suman por lo menos 7.



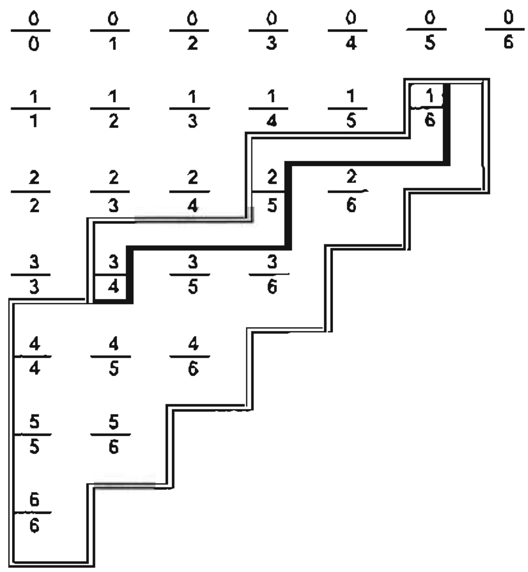
Definición 2.2.9 Si cualquier resultado de E_2 también es un resultado de E_1, se dice que el evento E_2 está *contenido* en E_1 y se denota por $E_2 \subset E_1$.

Ejemplo 2.2.9 Utilizando el dominó del ejemplo 2.2.1, dé dos eventos en los que uno esté contenido en otro.

Solución:

E_1: El conjunto de fichas en que los puntos de cada una suman por lo menos 7.

E_2: El conjunto de fichas en las que los puntos de cada una suman exactamente 7.



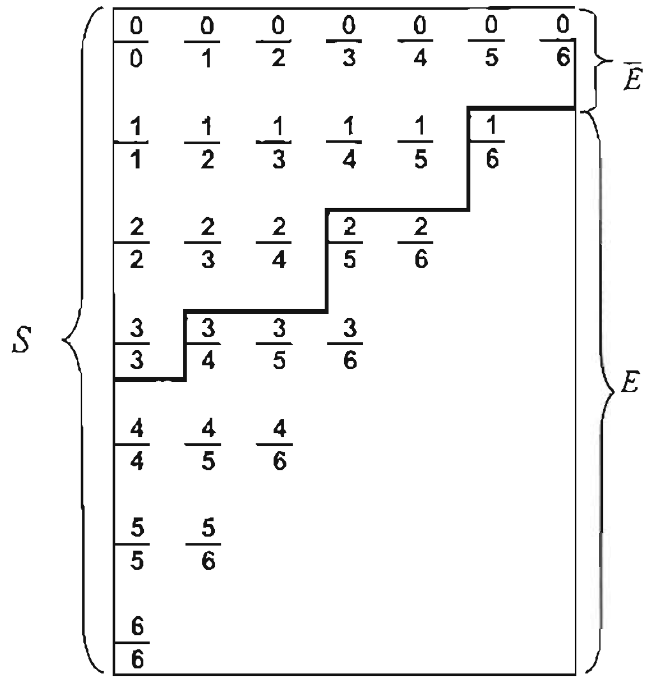
Definición 2.2.10 El *complemento* de un evento E con respecto al espacio muestral S, es aquel que contiene todos los resultados de S que no se encuentran en E, y se denota por $\overline{E}$.

Ejemplo 2.2.10 Utilizando el espacio muestral del ejemplo 2.2.1, dé un evento y el complemento de éste.

Solución:

E: El conjunto de fichas en las que los puntos de cada una suman por lo menos 7.

$\overline{E}$: El conjunto de fichas en las que los puntos de cada una suman menos de 7.



Antes de formalizar la definición de probabilidad, se dirá que ésta es un número real que mide la posibilidad colectiva, de ocurrencia, de los resultados del evento cuando se lleve a efecto el experimento. A continuación se da la definición axiomática de la probabilidad.

Definición 2.2.11 Sea S cualquier espacio muestral y E cualquier evento de éste. Se llamará función de probabilidad sobre el espacio muestral S a $P(E)$ si satisface los siguientes axiomas:

1. $P(E) \geq 0$
2. $P(S) = 1$
3. Si para los eventos E_1, E_2, E_3, ...

   $E_i \cap E_j = \Phi$ para toda $i \neq j$, entonces

   $P(E_1 \cup E_2 \cup \ldots) = P(E_1) + P(E_2) + \cdots$.

Los siguientes teoremas[6] son consecuencias de estos tres axiomas:

Teorema 1. $P(\Phi) = 0$.

Teorema 2.Para cualquier evento $E \subset S$, $0 \leq P(E) \leq 1$.

Teorema 3.Sea S un espacio muestral que contiene a cualesquiera dos eventos A y B; entonces,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

2.3 El Concepto de Variable Aleatoria[7]

El propósito de una variable aleatoria es transformar cada punto de un espacio muestral en un punto de un eje real, de tal manera que dicha transformación sea una función. De manera formal, este concepto se define en el siguiente párrafo.

Definición 2.3.1 Sea S un espacio muestral sobre el que se encuentra definida una función de probabilidad. Sea X una función de valor real definida sobre S, de manera que transforme los resultados de S en puntos sobre la recta de los reales. Se dice entonces que X es una *variable aleatoria*.

Ya que una variable aleatoria es una caracterización cuantitativa de los resultados de un espacio muestral, ésta posee intrínsecamente la naturaleza discreta o continua de este espacio.

Definición 2.3.2 Se dice que una variable aleatoria X es *discreta* si el número de valores que puede tomar es contable (ya sea finito o infinito), y si éstos pueden arreglarse en una secuencia que corresponde con los enteros positivos.

[6] Las demostraciones de estos teoremas pueden ser consultadas en la bibliografía citada para este subcapítulo.
[7] CANAVOS, *op. cit.* p. 52 – 53.

Los siguientes son ejemplos típicos de variables aleatorias discretas:

1. El número de tornillos defectuosos en una muestra de 10 extraída de una producción industrial.
2. El número de personas que esperan en la oficina de un doctor.

Definición 2.3.3 Se dice que una variable aleatoria X es *continua* si sus valores consisten en uno o más intervalos de la recta de los reales.

Los siguientes son ejemplos típicos de variables aleatorias continuas:

1. La estatura de una persona.
2. La cantidad de azúcar en una mandarina.

2.4 Distribuciones de Probabilidad de Variables Aleatorias Discretas[8]

En general, una variable aleatoria discreta X representa los resultados de un espacio muestral en forma tal que por $P(X = x)$ se entenderá la probabilidad de que X tome el valor de x. De esta forma, al considerar los valores de una variable aleatoria es posible desarrollar una función matemática que asigne una probabilidad a cada *realización* x de la variable aleatoria X. Esta función recibe el nombre de *función de probabilidad*[9] de la variable aleatoria X. El Término más general, distribución de probabilidad, se refiere a la colección de valores de la variable aleatoria y a la distribución de probabilidades entre éstos. Sin embargo, hacer referencia a la distribución de probabilidad de X no sólo implica la existencia de la función de probabilidad, sino también la existencia de la *función de distribución acumulativa* de X.

Definición 2.4.1 Sea X una variable discreta. Se llamará a $p(x) \equiv P(X = x)$ *función de probabilidad* de la variable aleatoria X, si satisface las siguientes propiedades:

1. $p(x) \geq 0$ para todos los valores de x de X;
2. $\sum_x p(x) = 1$.

---

[8] *Ibid*, 53 – 57.

[9] El nombre completo de esta función es el de *función de masa de probabilidad* de una variable aleatoria discreta.

3. $p(x_i \leq X \leq x_j) = \sum_{X=x_i}^{x_j} p(x)$

Definición 2.4.2 La *función de distribución acumulativa* de la variable aleatoria X es la probabilidad de que X sea menor o igual a un valor específico x y está dada por:

$$F(x) \equiv P(X \leq x) = \sum_{x_i \leq x} p(x_i)$$

Por lo tanto, en el caso discreto, una variable aleatoria X está caracterizada por la función de probabilidad puntual $p(x)$, la cual determina la probabilidad puntual de que $X = x$, y por la función de distribución acumulativa de $F(x)$, la que representa la suma de las probabilidades puntuales hasta el valor de x de X inclusive.

En general, la función de distribución acumulativa $F(x)$ de una variable aleatoria discreta es una función no decreciente de los valores de X, de tal manera que

1. $0 \leq F(x) \leq 1$ para cualquier x;
2. $F(x_i) \geq F(x_j)$ si $x_i \geq x_j$;
3. $P(X > x) = 1 - F(x)$.

Además, puede establecerse que para variables aleatorias de valor entero se tiene que:

4. $P(X = x) = F(x) - F(x-1)$;
5. $P(x_i \leq X \leq x_j) = F(x_j) - F(x_i - 1)$

Ejemplo 2.4.1 El número de vagonetas solicitadas en renta a una agencia de alquiler en un período de 50 días se identifica en la siguiente tabla:

Demanda posible X	Número de Días	$p(x)$	$P(X \leq x)$
3	3	0.06	0.06
4	7	0.14	0.20
5	12	0.24	0.44
6	14	0.28	0.72
7	10	0.20	0.92
8	4	0.08	1.00
	50	1.00	

2.5 Distribuciones de Probabilidad de Variables Aleatorias Continuas[10]

En el caso continuo, la probabilidad de que una variable aleatoria X tome un valor específico x es cero.

La distribución de probabilidad de una variable aleatoria continua X está caracterizada por una función $f(x)$ que recibe el nombre de *función de densidad de probabilidad.* Esta función $f(x)$ no es la misma función de probabilidad que para el caso discreto. Como la probabilidad de que X tome el valor específico x es cero, la función de densidad de probabilidad no representa la probabilidad de que $X = x$. Más bien, ésta proporciona un medio para determinar la probabilidad de un intervalo $a \leq X \leq b$.

La función de densidad de probabilidad de una variable aleatoria continua X se define formalmente de la siguiente manera:

Si existe una función $f(x)$ tal que

1. $f(x) \geq 0, \ -\infty < x < \infty,$
2. $\int_{-\infty}^{\infty} f(x)dx = 1,$ y

[10] *Ibid*, p. 57 – 62.

3. $P(a \leq X \leq b) = \int_a^b f(x)dx$

para cualesquiera a y b, entonces $f(x)$ es la función de densidad de probabilidad de la variable aleatoria continua X.

Al igual que en el caso de una variable aleatoria discreta, la *función de distribución acumulativa* $F(x)$ de una variable aleatoria continua X es la probabilidad de que X tome un valor menor o igual a algún x especifico. Esto es,

$$P(X \leq x) = F(x) = \int_{-\infty}^{x} f(t)dt,$$

en donde t es una variable artificial de integración.

Dado que para cualquier variable aleatoria continua X,

$$P(X = x) = \int_x^x f(t)dt = 0,$$

entonces:

$$P(X \leq x) = P(X < x) = F(x).$$

La distribución acumulativa $F(x)$, es una función lisa no decreciente de los valores de la variable aleatoria con las siguientes propiedades:

1. $F(-\infty) = 0$;
2. $F(\infty) = 1$;
3. $P(a < X < b) = F(b) - F(a)$;
4. $dF(x)/dx = f(x)$.

Ejemplo 2.5.1 El tiempo requerido por los estudiantes para presentar un examen de una hora es una variable aleatoria con función de densidad dada por

$$f(x)=\begin{cases} cx^2+x & 0\leq x\leq 1 \\ 0 & \text{en otro caso} \end{cases}$$

a) Determinar el valor de c que hace de $f(x)$ una función de densidad.

b) Una vez calculada c, obtener $F(x)$

Solución:

a) $\int_0^1 (cx^2+x)dx = 1$, de donde

$$\frac{c}{3}+\frac{1}{2}=1 \Rightarrow c=\frac{3}{2}$$

b) $F(x)=\int_0^x \left(\frac{3}{2}t^2+t\right)dt=\frac{x^3}{2}+\frac{x^2}{2},\quad 0\leq x\leq 1$

Finalmente,

$$F(x)=\begin{cases} 0 & x<0 \\ \dfrac{x^3}{2}+\dfrac{x^2}{2} & 0\leq x\leq 1 \\ 1 & x>1 \end{cases}$$

2.6 Valor Medio y Varianza de Distribución

En lugar de usar la caracterización completa de una distribución, por lo regular es suficiente con describirla (aunque incompletamente), en términos de ciertas cantidades que caracterizan propiedades generales de dicha distribución. Las dos cantidades más importantes son el valor *medio* o *media* μ y la *varianza* σ^2; y a continuación se definen.

2.6.1 Valor Medio de una Distribución[11]

El valor medio o media de una distribución se representará por μ. En el caso de una distribución discreta se define como

$$E(x) = \mu = \sum_i x_i p(x_i)$$

y en el caso de una distribución continua por

$$E(x) = \mu = \int_{-\infty}^{\infty} x f(x) dx .$$

2.6.2 Varianza de una Distribución[12]

La varianza de una distribución se representa por σ^2. En el caso discreto se define según la fórmula

$$Var(x) = \sigma^2 = \sum_i (x_i - \mu)^2 p(x_i)$$

y en el caso continuo mediante la fórmula

$$Var(x) = \sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx.$$

2.7 Algunas Distribuciones de Probabilidad

Este subcapítulo está dedicado al conocimiento de las principales distribuciones de probabilidad, tanto discretas como continuas.

2.7.1 Distribuciones Discretas

Se considerarán tres distribuciones discretas que tienen mucha importancia en estadística. Éstas son, la distribución binomial, la de Poisson y la hipergeométrica.

---

[11] *Cfr.* KREYSZIG, Erwin, Introducción a la Estadística Aplicada Principios y Métodos, p. 97.

[12] *Cfr. Ibid*, p. 97.

2.7.1.1 Distribución Binomial

Es una de las distribuciones discretas de probabilidad más útiles. Sus áreas de aplicación incluyen inspección de calidad, ventas, mercadotecnia, medicina, investigación de opiniones y otras.[13]

La distribución binomial es un modelo aplicable para situaciones de toma de decisiones en las que puede suponerse que un proceso de muestreo responde a un proceso *Bernoulli*. Un proceso Bernoulli es un proceso de muestreo en el que:

1. En cada ensayo u observación sólo son posibles dos resultados mutuamente excluyentes. Por convención éstos resultados se llaman *éxito* y *fracaso*.
2. Los resultados de la serie de ensayos, u observaciones, constituyen *eventos independientes*.
3. La probabilidad de éxito de cada ensayo, indicada por p, es constante de un ensayo a otro. Esto es, el proceso es estacionario.

La distribución binomial puede servir para determinar la probabilidad de obtener un número establecido de éxitos en un proceso Bernoulli. Se requiere de tres valores: el número establecido de éxitos (x); el número de ensayos, u observaciones (n), y la probabilidad de éxito en cada ensayo (p).[14]

Para resolver problemas que satisfacen las condiciones enunciadas en el párrafo anterior, se utilizará una fórmula que se obtendrá de la siguiente forma: si p y $1-p$ son probabilidades de éxito y fracaso en cada ensayo, entonces la probabilidad de obtener x éxitos y $n-x$ fracasos en algún orden específico es $p^x(1-p)^{n-x}$; claramente, en este producto de las p's y de las $(1-p)$'s hay un factor p para cada éxito, un factor $1-p$ para cada fracaso, y los x factores p y los $n-x$ factores $1-p$ son multiplicados todos a la vez. En vista de que esta probabilidad se aplica a cualquier punto del espacio muestral que representa x éxitos y $n-x$ fracasos (en cualquier orden específico), sólo tenemos

[13] CANAVOS *op. cit.* p. 89.
[14] KAZMIER *op. cit.* p. 96.

que contar cuántos puntos de esta clase existen, y multiplicar entonces $p^x(1-p)^{n-x}$ por este número. Evidentemente, el número de formas en que podemos obtener x éxitos en n ensayos es el número de combinaciones de x objetos seleccionados de un conjunto de n objetos $\binom{n}{x}$ así llegamos al siguiente resultado.[15]

Función de Probabilidad	*Parámetros*
$p(x;n,p)=\binom{n}{x}p^x(1-p)^{n-x}$ $x=0,1,2,\ldots,n$	n, entero positivo p, $0\leq p\leq 1$
Media	*Varianza*
np	*npq*

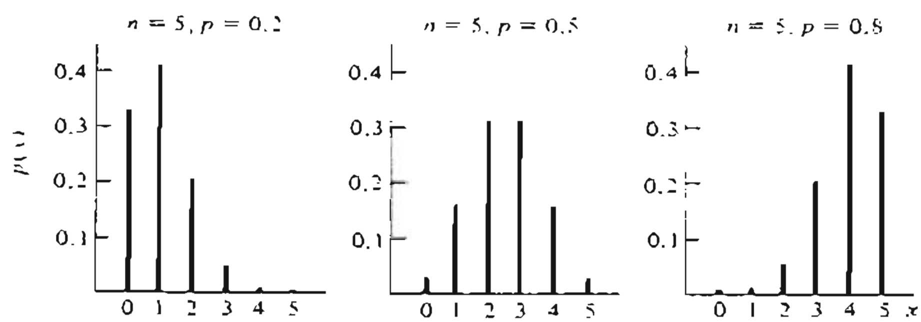


FIGURA 2.1. Gráficas de la función binomial de probabilidad.[16]

2.7.1.2 Distribución de Poisson

La distribución de Poisson puede utilizarse para determinar la probabilidad de ocurrencia de un número establecido de eventos cuando éstos ocurren en un *continuum* temporal o espacial. Este proceso se llama *proceso de Poisson*; aunque semejante al proceso Bernoulli,[17] se distingue de él en que los eventos ocurren a lo largo de un *continuum* (durante un intervalo temporal, por ejemplo) y en que no se dan ensayos propiamente dichos. Un ejemplo de un proceso de este tipo sería el número de personas que llegan a una tienda de autoservicio en un tiempo determinado, el número de bacterias en un cultivo, etcétera. Como en el caso del proceso Bernoulli, se supone que los eventos son

[15] MILLER, Irwin, John E. FREUND y Richard A. JOHNSON, Probabilidad y Estadística para Ingenieros, p. 93 – 94.
[16] CANAVOS *op. cit.* p. 91.
[17] *Vid.* 2.7.1.1

independientes y el proceso es estacionario.

Para determinar la probabilidad de ocurrencia de un número establecido de eventos en un proceso de Poisson sólo se requiere de un valor: el número medio de eventos a largo plazo en la dimensión temporal o espacial especifica de interés. Por lo general esta media se representa como λ, o en ocasiones como μ. La fórmula para determinar la probabilidad de un número establecido de éxitos x en una distribución de Poisson es:[18]

Función de Probabilidad	*Parámetro*
$p(x;\lambda)=\dfrac{e^{-\lambda}\lambda^{x}}{x!}$ $x=0,1,2,\ldots$	$\lambda>0$
Media	*Varianza*
λ	λ

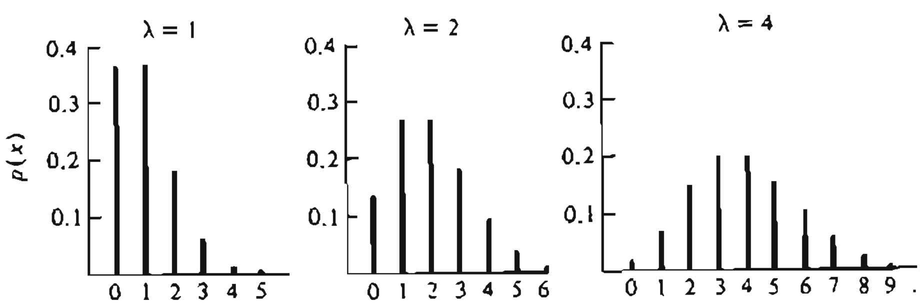


FIGURA 2.2 Gráficas de la función de probabilidad de Poisson.[19]

2.7.1.3 Distribución Hipergeométrica

Cuando el muestreo se realiza *sin reemplazo* de cada elemento muestreado tomado de una población finita de elementos, no se aplica el proceso Bernoulli, porque cuando se eliminan elementos de la población existe un cambio sistemático en la probabilidad de éxito. Cuando en una situación que de otro modo correspondería a un proceso Bernoulli, se hace uso del muestreo sin reemplazo, la distribución discreta adecuada es la distribución hipergeométrica.

[18] KAZMIER *op. cit.* p. 99.
[19] CANAVOS *op. cit.* p. 100.

Concediendo que x es el número establecido de éxitos, N el número total de elementos de la población, k el número total de éxitos incluidos en la población y n el número de elementos de la muestra, la fórmula para determinar probabilidades hipergeométricas es:[20]

Función de Probabilidad	*Parámetros*
$p(x;N,n,k)=\dfrac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$ $x=0,1,2,\ldots,n$ $x\le k,\ n-x\le N-k$	$N,\ n,\ k,$ enteros positivos $1\le n\le N;\ 1\le k\le N;$ $N=1,2,\ldots$
Media	*Varianza*
$\dfrac{nk}{N}$	$\dfrac{nk(N-k)(N-n)}{N^2(N-1)}$

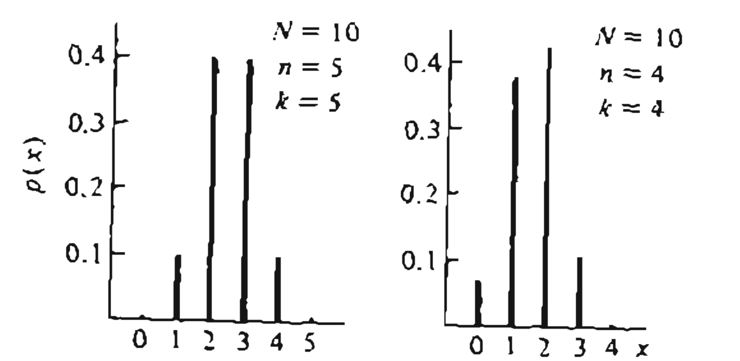


FIGURA 2.3 Gráficas de la función hipergeométrica de probabilidad.[21]

2.7.2 Distribuciones Continuas

De manera específica en este subcapítulo se estudiarán los siguientes modelos de probabilidad: normal, uniforme, gama, de Weibull, exponencial negativa y beta.

---

[20] KAZMIER *op. cit.* p. 98.
[21] CANAVOS *op. cit.* p. 110.

2.7.2.1 Distribución Normal

La distribución normal de probabilidad es una distribución de probabilidad continua. La curva de probabilidad que representa a la distribución normal de probabilidad tiene forma de campana, como lo ejemplifican las curvas de probabilidad de la siguiente figura.

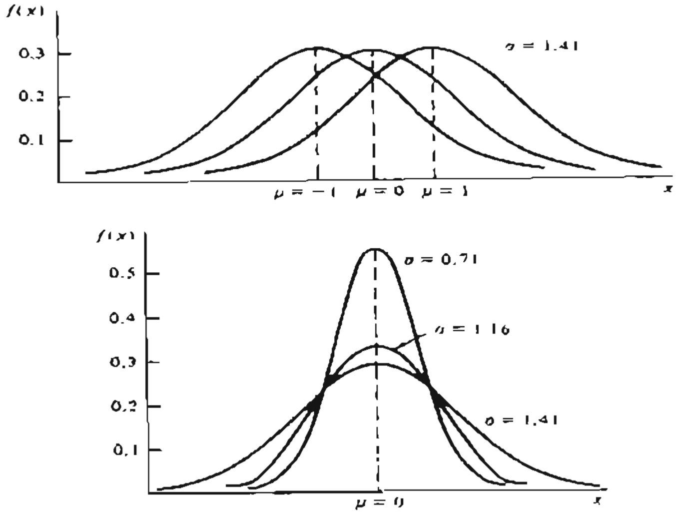


FIGURA 2.4 Gráficas de la función de densidad normal para diferentes valores de μ y σ.[22]

La distribución normal de probabilidad es importante para la inferencia estadística por tres razones:

1. Se sabe que las medidas obtenidas por muchos procesos aleatorios siguen esta distribución.
2. Las probabilidades normales suelen servir para aproximar otras distribuciones de probabilidad, como las distribuciones binomial y Poisson.
3. Las distribuciones estadísticas como la media muestral y la proporción muestral tiene distribución normal cuando el tamaño de muestra es grande, independientemente de la población origen.

[22] CANAVOS *op. cit.* p. 132.

Como sucede en todas las distribuciones continuas de probabilidad, un valor de probabilidad de una variable aleatoria continua sólo puede determinarse para un intervalo de valores.

Se dice que una variable aleatoria X se encuentra *normalmente distribuída* si su función de densidad de probabilidad está dada por:

Función de Densidad de Probabilidad	*Parámetros*	
$f(x;\mu,\sigma)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\left[(x-\mu)^2/2\sigma^2\right]}$	μ,	$-\infty<\mu<\infty$
$-\infty<x<\infty$	σ,	$\sigma>0$
Media	*Varianza*	
μ	σ^2	

Puesto que toda diferente combinación de μ y σ generaría una distribución normal de probabilidad diferente, las tablas de probabilidad normales se basan en una distribución en particular: la *distribución normal estándar*. Ésta es la distribución normal de probabilidad con $\mu=0$ y $\sigma=1$. Todo valor de x precedente de una población con distribución normal puede convertirse en el equivalente valor estándar de z mediante la fórmula:

$$z=\frac{x-\mu}{\sigma}$$

Un valor de z reformula el valor x original en términos del número de unidades de la desviación estándar por las cuales el valor original difiere de la media de la distribución. Un valor negativo z indicaría que el valor x original estaba por debajo de la media.[23]

[23] KAZMIER *op. cit.* p. 115.

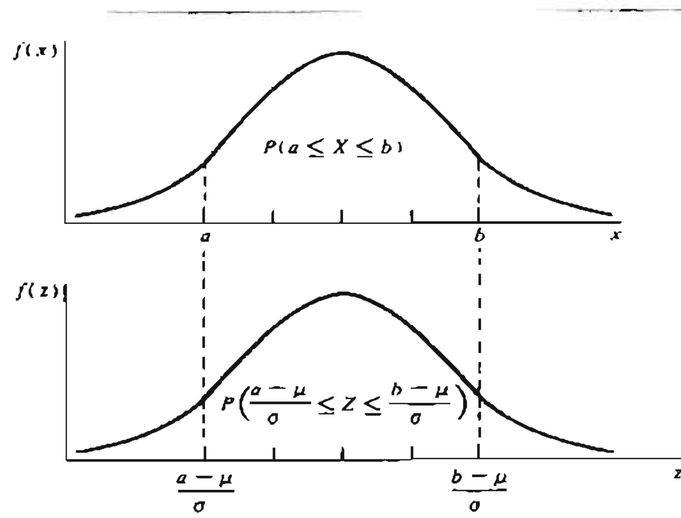


FIGURA 2.5 Correspondencia entre las probabilidades de X y de Z.[24]

2.7.2.2 Distribución Uniforme

Supóngase que ocurre un evento en que una variable aleatoria toma valores de un intervalo infinito, de manera que éstos se encuentran distribuidos igualmente sobre el intervalo. Esto es, la probabilidad de que la variable aleatoria tome un valor en cada subintervalo de igual longitud es la misma. Se dice entonces que la variable aleatoria se encuentra *distribuida uniformemente* sobre el intervalo.

La distribución uniforme proporciona una representación adecuada para redondear diferencias que surgen al medir cantidades físicas entre los valores observados y los reales. Por ejemplo, si el peso de un individuo se redondea al kilogramo más cercano.

Se dice que una variable aleatoria X está *distribuida uniformemente* sobre el intervalo (a,b) si su función de densidad está dada por:[25]

[24] *Ibid*, p. 137.
[25] *Ibid*, p. 143 – 147.

Función de Densidad de Probabilidad	Parámetros	
$f(x;a,b)=\begin{cases}\dfrac{1}{(b-a)} & a \le x \le b \\ 0 & \text{para cualquier otro caso}\end{cases}$	a, b,	$-\infty \prec a \prec \infty$ $-\infty \prec b \prec \infty$

Media	Varianza
$\dfrac{(a+b)}{2}$	$\dfrac{(b-a)^2}{12}$

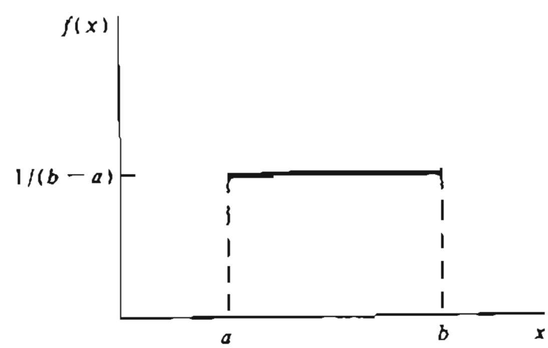


FIGURA 2.6 Gráfica de la función de densidad de probabilidad uniforme.[26]

2.7.2.3 Distribución Gama[27]

Esta distribución se emplea de manera extensa en una gran diversidad de áreas; por ejemplo, para representar el tiempo aleatorio de la falla sólo si de manera exacta los componentes fallan y la falla de cada componente ocurre a una frecuencia constante $\lambda = 1/\theta$ por unidad de tiempo. También se emplea en problemas de líneas de espera para representar el intervalo (de tiempo) total para completar una reparación si ésta se lleva a cabo en subestaciones; completar la reparación de cada subestación es un evento independiente que ocurre a una frecuencia constante igual a $\lambda = 1/\theta$. Existen algunos ejemplos que no siguen el patrón anterior pero se aproximan de manera adecuada mediante el empleo de la distribución gama, como los ingresos familiares y la edad del hombre al contraer matrimonio por primera vez.

---

[26] *Ibid*, p. 144.
[27] *Ibid*, p. 152 – 159.

Se dice que la variable aleatoria X tiene una *distribución gama* si su función de densidad de probabilidad está dada por:

Función de Densidad de Probabilidad	*Parámetros*	
$f(x;\alpha,\theta)=\begin{cases}\dfrac{1}{\Gamma(\alpha)\theta^{\alpha}}x^{\alpha-1}e^{-\frac{x}{\theta}} & x\succ 0\\ 0 & \text{para cualquier otro caso}\end{cases}$	α, θ,	$\alpha>0$ $\theta>0$
Media	*Varianza*	
$\alpha\theta$	$\alpha\theta^{2}$	

Donde: $\Gamma(\alpha)=\int_{0}^{\infty}x^{\alpha-1}e^{-x}dx,\quad \alpha>0$

Algunas propiedades de esta función son:

1. $\Gamma(1)=1$
2. $\Gamma(\alpha+1)=\alpha\Gamma(\alpha)$
3. $\Gamma(\alpha+1)=\alpha!$
4. $\Gamma\left(\frac{n}{2}\right)=\left(\frac{n}{2}-1\right)!$
5. $\Gamma\left(\frac{1}{2}\right)=\sqrt{\pi}$.[28]

La distribución gama es muy versátil puesto que exhibe varios perfiles que dependen del valor del parámetro α. En la siguiente figura se ilustran distintos perfiles de la función de densidad gama para distintos valores de α y θ. Como puede observarse, para $\alpha\leq 1$, la distribución gama tiene un perfil en forma de J transpuesta. Para $\alpha>1$, presenta un pico que ocurre en $x=\theta(\alpha-1)$. Para un valor fijo θ, el perfil básico de la distribución gama no se altera si el valor α cambia. Lo anterior da como resultado que las cantidades α y θ son los factores de forma y de escala, de la distribución gama.

[28] KREYSZIG *op. cit.* p. 164 – 165.

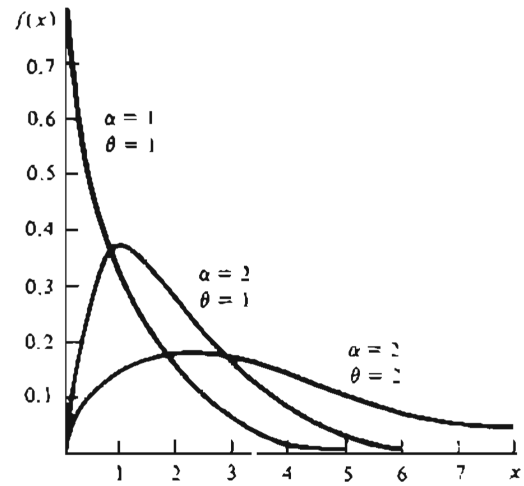


FIGURA 2.7 Gráficas de la función de densidad gama para distintos valores de α y θ.[29]

Cuando α es un entero positivo, la distribución gama se conoce también como *distribución de Erlang*; por otro lado, si $\alpha = 1$, la distribución Erlang (gama) se reduce a lo que se conoce como la *distribución exponencial negativa*.

Otro caso especial del modelo de probabilidad gama es la *distribución chi-cuadrado*[30] cuando $\alpha = \nu/2$ y $\theta = 2$; está distribución se encuentra caracterizada por un mismo parámetro ν, que recibe el nombre de grados de libertad.

---

[29] CANAVOS *op. cit.* p. 152.

[30] La función de densidad de probabilidad de una variable aleatoria chi–cuadrado se determina por:

$$f(x;\nu) = \begin{cases} \dfrac{1}{\Gamma(\nu/2)2^{\nu/2}} x^{\nu/2-1} e^{-\frac{x}{2}} & x \succ 0 \\ 0 & \text{para cualquier otro caso.} \end{cases}$$

Como se verá, esta distribución interviene en la inferencia estadística y de manera especial al hacer inferencia respecto a las varianzas.
Vid: *Ibid*, p. 158.

2.7.2.4 Distribución de Weibull[31]

En los últimos 25 años esta distribución se empleó como modelo para situaciones de tiempo–falla y con el objetivo de lograr una amplia variedad de componentes mecánicos y eléctricos.

Se dice que una variable aleatoria X tiene una *distribución de Weibull* si su función de densidad de probabilidad está dada por:

Función de Densidad de Probabilidad	*Parámetros*	
$f(x;\alpha,\theta)=\begin{cases}\dfrac{\alpha}{\theta^{\alpha}}x^{\alpha-1}e^{-\left(\frac{x}{\theta}\right)^{\alpha}} & x \succ 0 \\ 0 & \text{para cualquier otro caso}\end{cases}$	α, θ,	$\alpha \succ 0$ $\theta \succ 0$
Media	*Varianza*	
$\theta\,\Gamma\left(1+\dfrac{1}{\alpha}\right)$	$\theta^{2}\left[\Gamma\left(1+\dfrac{2}{\alpha}\right)-\Gamma^{2}\left(1+\dfrac{1}{\alpha}\right)\right]$	

La distribución Weibull es una familia de distribuciones que dependen de dos parámetros: el de forma, α y el de escala, θ. En la siguiente figura se muestran varias gráficas de la distribución Weibull para distintos valores de α y θ; y como puede observarse, esta distribución tiene distintos perfiles dependiendo del valor α. Por ejemplo, si $\alpha \prec 1$, tiene una forma de J transpuesta, y si $\alpha \succ 1$, la función de densidad presenta un pico único.

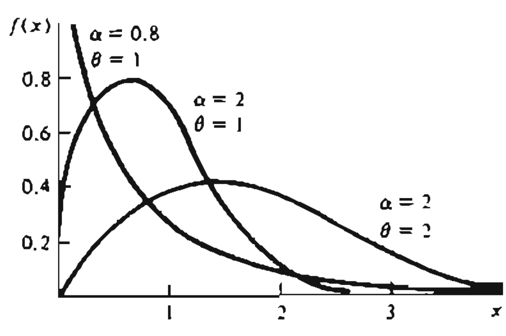


FIGURA 2.8 Gráficas de la función de densidad de Weibull para distintos valores de α y θ.[32]

[31] *Ibid*, p. 159 – 163.
[32] *Ibid*, p. 160.

Existen dos casos especiales en la distribución de Weibull que merecen mención especial. Cuando el parámetro de forma es igual a uno, la distribución de Weibull (al igual que la gama), se reduce a la distribución exponencial negativa. Cuando $\alpha = 2$ y el parámetro de escala θ se reemplaza por $\sqrt{2}\sigma$, la función de densidad de Weibull se reduce a lo que se conoce como *distribución de Rayleigh*.[33]

2.7.2.5 Distribución Exponencial Negativa[34]

Se ha notado con anterioridad que la distribución exponencial (negativa) es un caso especial de los modelos de Weibull y gama. Ya que es un caso especial de la distribución gama (Erlang), la variable aleatoria exponencial es el tiempo que transcurre hasta que se da el primer evento de Poisson. Es decir, la distribución exponencial puede modelar el lapso entre dos eventos consecutivos de Poisson que ocurren de manera independiente y a una frecuencia constante: esta distribución se emplea con bastante frecuencia con objeto de modelar problemas del tipo tiempo–falla y como modelo para el intervalo en problemas de líneas de espera.

Si una variable aleatoria X tiene una distribución exponencial, su función de densidad está dada por:

---

[33] La función de densidad de probabilidad de una variable Rayleigh se determina por:

$$f(x;\sigma^2)=\begin{cases}\dfrac{x}{\sigma^2}e^{-\frac{x^2}{2\sigma^2}} & x>0\\ 0 & \text{para cualquier otro caso.}\end{cases}$$

Esta distribución es utilizada en problemas de ruido asociados con sistemas de comunicación.

Vid: BILLINTON, Roy y Ronald N. ALLAN, Reliability Evaluation of Engineering Systems Concepts and Techniques, p. 192.

[34] CANAVOS *op. cit.* p. 163 – 167.

Función de Densidad de Probabilidad	Parámetro
$f(x;\theta)=\begin{cases}\frac{1}{\theta}e^{-\left(\frac{x}{\theta}\right)} & x \succ 0\\ 0 & \text{para cualquier otro caso}\end{cases}$	θ, $\theta \succ 0$
Media	**Varianza**
θ	θ^2

La distribución exponencial se caracteriza por un parámetro θ, que representa el lapso promedio de tiempo entre dos eventos independientes de Poisson. En el contexto de la confiabilidad, θ recibe el nombre de tiempo promedio entre fallas; y $1/\theta$ es la frecuencia de la falla.

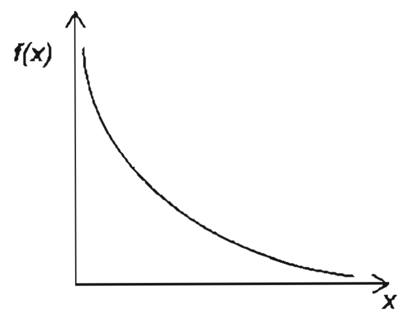


FIGURA 2.9 Gráfica de la función de densidad exponencial.[35]

2.7.2.6 Distribución Beta[36]

Una distribución que permite representar una gran variedad de perfiles es la distribución beta. Se ha utilizado para representar variables físicas cuyos valores se encuentran restringidos a un intervalo de longitud finita y para encontrar ciertas cantidades que se conocen como límites de tolerancia sin necesidad de la hipótesis de una distribución normal. Además, la distribución beta juega un gran papel en la estadística bayesiana.

Se dice que una variable aleatoria X posee una distribución beta si su función de densidad de probabilidad está dada por:

[35] *Crf.* MILLER, Irwin, John E. FREUND y Richard A. JOHNSON, *op. cit.* p. 161.
[36] CANAVOS *op. cit.* p. 147 – 151.

Función de Densidad de Probabilidad		Parámetros	
$f(x;\theta)=\begin{cases} \dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1} & 0 \prec x \prec 1 \\ 0 & \text{para cualquier otro caso} \end{cases}$		α, β,	$\alpha \succ 0$, $\beta \succ 0$
Media		**Varianza**	
$\dfrac{\alpha}{(\alpha+\beta)}$		$\dfrac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$	

Las cantidades α y β de la distribución beta son, ambas, parámetros de perfil. Valores distintos de α y β darán distintos perfiles para la función de densidad beta. Si tanto α como β son menores que uno, la distribución beta tiene un perfil en forma de U. Si $\alpha \prec 1$ y $\beta \geq 1$, la distribución tiene un perfil de J transpuesta; y si $\beta \prec 1$ y $\alpha \geq 1$, el perfil es una J. Cuando tanto α y β son ambos menores que uno, la distribución presenta un pico en $x=(\alpha-1)/(\alpha+\beta-2)$. Finalmente, la distribución beta es simétrica cuando $\alpha=\beta$. En la siguiente figura observe los perfiles de esta distribución.

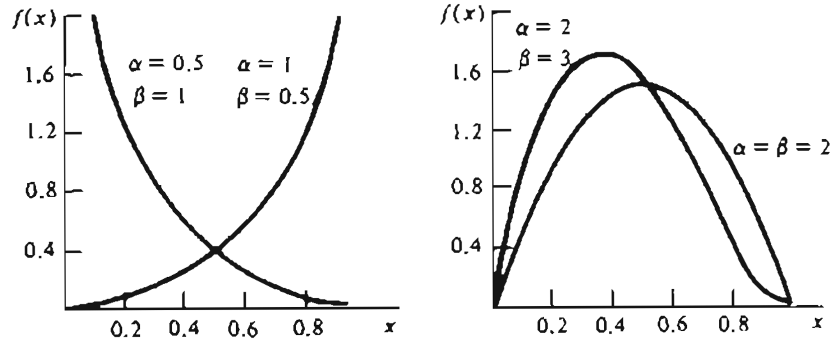


FIGURA 2.10 Gráfica de la función de densidad beta para distintos valores de α y β.[37]

En este capítulo se estudiaron las principales distribuciones de probabilidad y se discutieron algunas áreas de aplicación de cada modelo, sin embargo el lector debe de considerar que para un experimento aleatorio siempre se elegirá la distribución de probabilidad que mejor se ajuste a este, para tal elección en el siguiente capítulo se muestran algunas pruebas de bondad de ajuste así como algunos métodos de inferencia estadística.

[37] CANAVOS *op. cit.* p. 148.

CAPÍTULO 3

ESTIMACIÓN DE DISTRIBUCIONES DE PROBABILIDAD PARA DATOS EMPÍRICOS Y LA DETERMINACIÓN DEL TAMAÑO DE MUESTRA

En numerosas ocasiones la información con que se cuenta para analizar el comportamiento de cierto experimento, o para intentar hacer un pronóstico, es tan poca, que resulta insuficiente si se intenta utilizar las técnicas clásicas. Sin embargo, ante la cotidianeidad con que se presenta esta problemática se han desarrollado otras técnicas que permiten librar la situación.

Por esta razón, el objetivo de este capítulo es el de estudiar los métodos de estimación de la distribución de probabilidad para datos empíricos, entre los que se destaca el método del rango mediano que sirve para tal fin cuando se cuenta con una muestra pequeña o insuficiente; así como los métodos de determinación del tamaño de muestra, ya que ésta debe de ser lo suficientemente grande para que las inferencias elaboradas en una población o estudio, –como lo es el de simulación– se apeguen a la realidad.

3.1 Estimación de la Distribución de Probabilidad para Datos Empíricos[1]

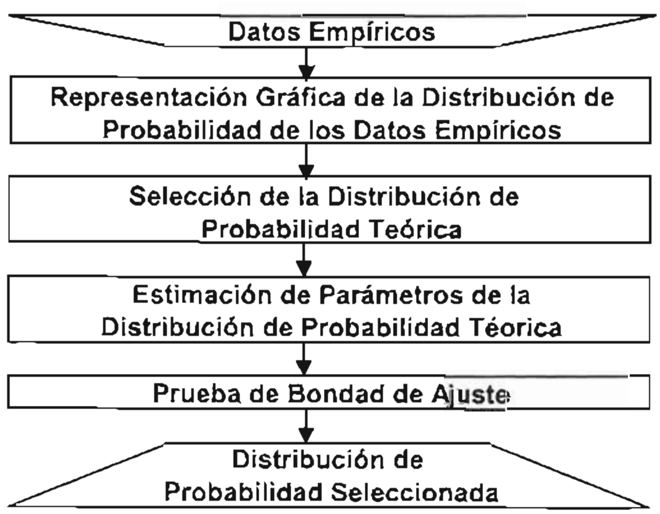


FIGURA 3.1 Algoritmo para la Estimación de la Distribución de Probabilidad para Datos Empíricos.[2]

[1] KNEZEVIC, Jezdimir, Reliability, Maintainability and Supportability a Probabilistic Approach p. 150 - 151.
[2] *Ibid.* p. 151.

El algoritmo anterior muestra los pasos a seguir para estimar la distribución de probabilidad de los datos empíricos obtenidos.

Es necesario hacer énfasis en que el término experimento abarca un gran número de actividades cuyo objetivo es generar y coleccionar datos (pruebas de laboratorio, datos de campo, reportes de usuario; y otros similares).

A continuación se explican los pasos del algoritmo anterior.

3.1.1 Representación de la Distribución de Probabilidad de los Datos Empíricos

Existen dos métodos para representar –o aproximar– la distribución de probabilidad de los datos empíricos dependiendo del número total de resultados disponibles y son los siguientes:

- Método del Rango Mediano.
- Método de Frecuencias Relevantes.

El primero es aplicable en casos donde el número de datos es menor a 50, mientras que el segundo se recomienda en el caso contrario. Ambos métodos son descritos para una variable aleatoria X.

3.1.1.1 Método del Rango Mediano para la Estimación de la Distribución de Probabilidad para Muestras Pequeñas[3]

De acuerdo con el teorema de Bernoulli, la probabilidad de ocurrencia de un evento se define por la razón de su frecuencia entre el número total de resultados, n, cuando ésta se aproxima a infinito, pero cuando el número de casos es insuficiente los datos no pueden ser agrupados y consecuentemente no se puede determinar de esta manera su distribución de probabilidad. Ante este problema se han desarrollado algunos métodos

[3] En este caso se considera muestra pequeña a aquella que consta de menos de 50 datos, esto porque existen unas tablas que proporcionan las probabilidades acumuladas de los datos (en orden ascendente) correspondientes al rango bajo consideración; estas probabilidades son calculadas mediante el método del rango mediano RM aquí explicado cuya aproximación es correcta al valor en tablas con 1% de error para 4 datos y de 0.1% para más de 50.

para estimar la distribución de probabilidad; entre éstos destaca el método del rango mediano el cual es el más utilizado y se presenta a continuación.

El rango mediano (RM) representa la proporción acumulada de los datos correspondientes al rango bajo consideración, la función de distribución acumulada de datos empíricos $F'(\cdot)$ puede determinarse para la variable aleatoria X, como sigue:

$$F'(x_i) = RM(x_i) = \frac{(i - 0.3)}{n + 0.4} \tag{3.1}$$

y la función de densidad de los datos empíricos $f'(\cdot)$:

$$f'(x_i) = \frac{1}{(n + 0.4) \times (x_{i+1} - x_i)} \tag{3.2}$$

Es necesario mencionar que para aplicar estas expresiones a los datos empíricos, estos deben de ser ordenados en forma ascendente, i.e. $i = 1$ corresponde al resultado con el valor numérico más bajo.[4]

Para aclarar este método y hacer los cálculos de manera fácil se presenta el siguiente ejemplo.

Ejemplo 1.3.1.1 El tiempo de falla de nueve dispositivos probados se lista a continuación:

85, 35, 38, 100, 52, 70, 27, 49 y 104.

Calcule y trace la gráfica de la distribución de probabilidad.

Solución: El primer paso es ordenar en forma ascendente los datos:

i	1	2	3	4	5	6	7	8	9
x_i	27	35	38	49	52	70	85	100	104

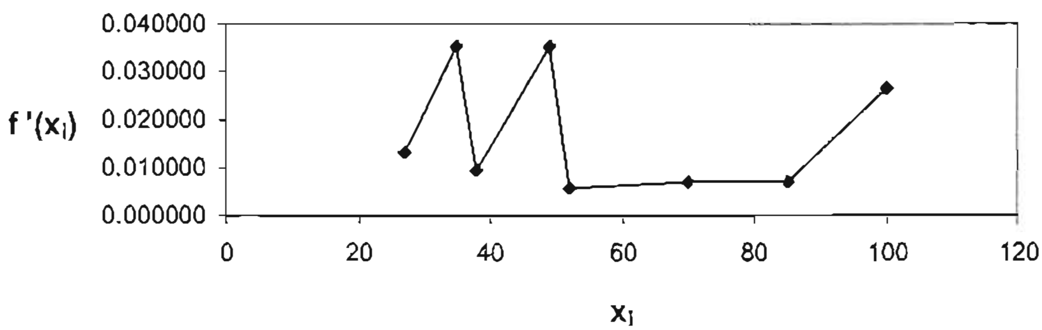
[4] *Ibid.* p. 151 -156.

Una vez ordenados los datos se calculan $F'(x_i)$ y $f'(x_i)$ con las ecuaciones 3.1 y 3.2 respectivamente, considerando que el tamaño de muestra es $n = 9$.

i	x_i	$F'(x_i)$	$f'(x_i)$
1	27	0.0745	0.013298
2	35	0.1809	0.035461
3	38	0.2872	0.009671
4	49	0.3936	0.035461
5	52	0.5000	0.005910
6	70	0.6064	0.007092
7	85	0.7128	0.007092
8	100	0.8191	0.026596
9	104	0.9255	

Finalmente, con los resultados de la tabla anterior se traza la gráfica de la distribución de probabilidad para los datos de este ejemplo.

DISTRIBUCIÓN DE PROBABILIDAD



3.1.1.2 Método de Frecuencias Relevantes para la Estimación de la Distribución de Probabilidad

Este método es el que se explicó en los subcapítulos 1.4 a 1.6.

Un ejemplo detallado de esta metodología se encuentra en el subcapítulo 1.7.

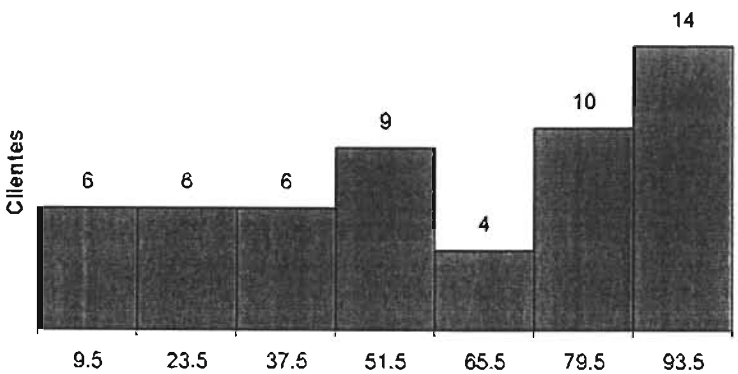
En el ejemplo mencionado los datos son el número de días que cada uno de 55 clientes asistió a un gimnasio durante los 4 primeros meses del año.

A partir de los datos, se obtuvo una tabla de distribución de frecuencias y un histograma de frecuencias, mismos que se muestran a continuación:

TABLA DE DISTRIBUCIÓN DE FRECUENCIAS

Intervalos de Clase Limites		Fronteras		Marcas de Clase x_i	Frecuencia f_i	Frecuencia Acumulada F_i	Frecuencia Relativa f_i^*	Frecuencia Relativa Acumulada F_i^*
3	16	2.5	16.5	9.5	6	6	0.10909	0.10909
17	30	16.5	30.5	23.5	6	12	0.10909	0.21818
31	44	30.5	44.5	37.5	6	18	0.10909	0.32727
45	58	44.5	58.5	51.5	9	27	0.16364	0.49091
59	72	58.5	72.5	65.5	4	31	0.07273	0.56364
73	86	72.5	86.5	79.5	10	41	0.18182	0.74545
87	100	86.5	100.5	93.5	14	55	0.25455	1.00000
					55		1.00000	

HISTOGRAMA DE FRECUENCIAS



De igual modo se calcularon las medidas numéricas descriptivas de los datos, algunas de éstas son las siguientes:

Media Aritmética: $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = 58.04$

Varianza: $s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1} = 898.11$

3.1.2 Selección de la Distribución de Probabilidad Teórica[5]

En esta fase, se debe seleccionar una distribución de probabilidad que represente tan cerca como sea posible a los datos empíricos. Utilizando lenguaje estadístico este proceso se conoce como planteamiento de la hipótesis.

Es difícil seleccionar una distribución de probabilidad teórica en particular que esté asociada a los datos empíricos, especialmente cuando la muestra es pequeña. El principal indicador será la gráfica de estos datos obtenida en el paso anterior, la cual puede compararse con las distribuciones de probabilidad que se estudiaron de manera breve en el capítulo 2.

Algunas recomendaciones para asistir el proceso de selección de la distribución de probabilidad teórica asociada a los datos empíricos son las siguientes:

- Si los valores de la desviación estándar y la media son relativamente cercanos, es un buen indicador de que la variable aleatoria obedece una distribución exponencial.
- Como la distribución normal es simétrica, i. e. la media, la mediana y la moda tienen el mismo valor, en los casos en que estás medidas centrales posean valores razonablemente cercanos, la distribución recomendada es la normal.
- Si las primeras condiciones no se cumplen, y la gráfica de la distribución de probabilidad empírica no es simétrica posiblemente se tenga una distribución de Weibull.
- Si ninguna de las condiciones anteriores se satisface posiblemente los datos no pertenece al mismo experimento o proceso, i. e. posiblemente se está tratando con una distribución mezclada. En este caso los datos empíricos deberán ser nuevamente examinados y si existen evidencias de una mezcla de datos será necesario separarlos; y repetir el procedimiento para cada conjunto de datos.

Una vez hecha la hipótesis sobre la distribución de probabilidad que siguen los datos, será necesario determinar los parámetros de ésta. El método que en el siguiente subcapítulo se presenta servirá para tal objetivo.

---

[5] *Ibid*. p. 166 – 184.

Ejemplo 3.1.2 Mediante la comparación del histograma del subcapítulo 1.7 (mismo que se muestra nuevamente en el subcapítulo 3.1.1.2) con las distribuciones de probabilidad continuas estudiadas en el subcapítulo 2.7.2 se formula la hipótesis de que tal conjunto de datos obedece una distribución gama.

3.1.3 Método de los Momentos para la Estimación de Parámetros de una Distribución de Probabilidad

Quizá este método sea el más antiguo para la estimación de parámetros. Éste consiste en igualar los momentos apropiados de la población con los correspondientes momentos muestrales para estimar un parámetro desconocido de la distribución.

Definición 3.1.3 Sea $X_1, X_2, \ldots X_n$ una muestra aleatoria de una distribución con función (masa) de probabilidad $f(x;\theta)$. El r–ésimo momento alrededor del cero se define como

$$M_r^{'} = \frac{1}{n}\sum_{i=1}^{n} X_i^r$$

Ejemplo 3.1.3 Con la media aritmética y la varianza calculadas en 1.7 (mismos que se muestra nuevamente en el subcapítulo 3.1.1.2) obtenga los parámetros[6] de una distribución gama, utilizando el método de los momentos.

Solución:

media =	58.04	$=\alpha\theta$	$\hat{\theta}=$	15.4749
Varianza =	898.11	$=\alpha\theta^2$	$\hat{\alpha}=$	3.7503

[6] Cuando un parámetro es estimado, a éste se le coloca el símbolo ^.

3.1.4 Introducción a las Pruebas de Hipótesis Estadísticas[7]

Una hipótesis estadística es una afirmación con respecto a alguna característica desconocida de una población de interés. La esencia de probar una hipótesis estadística es el decidir si la afirmación se encuentra apoyada por la evidencia experimental que se obtiene a través de una variable aleatoria. En forma general, la afirmación involucra ya sea algún parámetro o a alguna forma funcional no conocida de la distribución de interés a partir de la cual se obtiene una muestra aleatoria. La decisión acerca de si los datos muestrales apoyan estadísticamente la afirmación se toma con base a la probabilidad, y, si esta es mínima, entonces será rechazada.[8]

Se han desarrollado tres procedimientos distintos para la prueba de hipótesis, todos los cuales conducen a las mismas decisiones cuando se emplean los mismos estándares de probabilidad (y riesgo). El *método del valor crítico* para la prueba de hipótesis, de acuerdo con este procedimiento, se determinan los así llamados valores críticos de la estadística de prueba que dictarían el rechazo de una hipótesis, tras de lo cual la estadística de prueba observada se compara con los valores críticos. Éste fue el primer método en desarrollarse, motivo por el cual buena parte de la terminología de las pruebas de hipótesis se deriva de él. Más recientemente, el *método del valor P* ha cobrado popularidad a causa de ser el más fácilmente aplicable a software de cómputo. Este método se basa en la determinación de la probabilidad condicional de que el valor observado de una estadística muestral pueda ocurrir al azar, dado que un supuesto particular sobre el valor del parámetro poblacional asociado sea en efecto correcto. Finalmente, *el método de intervalos de confianza* se basa en la observación de sí el valor supuesto de un parámetro poblacional está incluido en el rango de valores que define a un intervalo de confianza para ese parámetro.

Pero más allá del método de prueba de hipótesis que se use, debe hacerse notar que si un valor hipotético no se rechaza, y por lo tanto se acepta, ello no constituye una "prueba" de que sea correcto. La aceptación de un valor supuesto de un parámetro indica

[7] KAZMIER, Leonard J., Estadística Aplicada a la Administración y a la Economía, p. 164 - 165.
[8] CANAVOS C. George, Probabilidad y Estadística Aplicaciones y Métodos, p. 303.

simplemente que se trata de un valor verosímil, con base en el valor observado de la estadística muestral.

En este subcapítulo se describirá únicamente el método del valor crítico para la prueba de hipótesis, porque las pruebas de bondad de ajuste que se estudiaran en el siguiente utilizan esta metodología.

Los pasos básicos de la prueba de hipótesis con el Método del Valor Crítico son los siguientes:

Paso 1. *Formular la hipótesis nula y la hipótesis alternativa.* La *hipótesis nula* (H_0) es el valor paramétrico hipotético que se compara con el resultado muestral. Se le rechaza sólo si es poco probable que el resultado muestral haya ocurrido dado lo correcto de la hipótesis. La hipótesis alternativa (H_1) se acepta sólo si la hipótesis nula es rechazada. En muchos libros de texto la hipótesis alternativa también se designa como H_a.

Paso 2. *Especificar el nivel de significancia por aplicar.* El nivel de significancia es el estándar estadístico que se especifica para rechazar la hipótesis nula. Si se especifica un nivel de significancia de 5%, la hipótesis nula se rechaza sólo si el resultado muestral es tan diferente del valor hipotético que una diferencia por ese monto o un monto superior ocurriría al azar con una probabilidad de 0.05 o menos.

Nótese que si se usa el nivel de significancia de 5%, hay una probabilidad de 0.05 de rechazar la hipótesis nula aun siendo efectivamente cierta. Esto se llama *error tipo I.* La probabilidad del error tipo I siempre es igual al nivel de significancia empleado como estándar para rechazar la hipótesis nula; se le designa con la letra griega minúscula α (alfa), de modo que α designa también al nivel de significancia. Los niveles de significancia de uso más frecuente en la prueba de hipótesis son los de 5% y 1%.

Ocurre un *error tipo II* si la hipótesis nula no se rechaza, y es por lo tanto aceptada, cuando en realidad es falsa. En la siguiente tabla se resumen los tipos de decisiones y las posibles consecuencias de las decisiones tomadas en pruebas de hipótesis.

Decisión posible	**Estados Posibles**	
	Hipótesis nula cierta	**Hipótesis nula falsa**
No rechazo de la hipótesis nula	Aceptación correcta	Error tipo II
Rechazo de la hipótesis nula	Error tipo I	Rechazo correcto

TABLA 3.1 Consecuencias de decisiones en pruebas de hipótesis.[9]

Paso 3. *Selección de la estadística de prueba.* La estadística de prueba será ya sea la estadística muestral (el estimador insesgado del parámetro a prueba) o una versión estandarizada de la estadística muestral. Por ejemplo, para probar un valor hipotético de la media poblacional, la media de una muestra aleatoria tomada de esa población podría servir como la estadística de prueba. Sin embargo, si la distribución de muestreo de la media es normal, el valor de la media muestral se convierte usualmente en un valor z, el cual funge entonces como la estadística de prueba.

Paso 4. *Establecer el valor o valores críticos de la estadística de prueba.* Habiendo especificado la hipótesis nula, el nivel de significancia y la estadística de prueba por usar, se establece(n) entonces el(los) valor(es) crítico(s) de la estadística de prueba.[10] En cualquier caso, un *valor crítico* identifica el valor de la estadística de prueba requerido para rechazar la hipótesis nula.

Paso 5. *Determinar el valor de la estadística de prueba.* Por ejemplo, al probar un valor hipotético de la media poblacional, se recolecta una muestra aleatoria y se determina el valor de la media muestral. Si el valor crítico fue establecido como un valor z, la media muestral se convierte a un valor z.

Paso 6. *Tomar la decisión.* El valor observado de la estadística muestral se compara con el valor crítico de la estadística de prueba. Se rechaza o no entonces la hipótesis nula. Si la hipótesis nula es rechazada, no se rechazará la hipótesis alternativa.

---

9 KAZMIER, *op. cit.* p. 164.

10 Estos valores pueden ser uno o dos, dependiendo de si están implicadas las así llamadas pruebas unilaterales o bilaterales.

3.1.4.1 Pruebas de Bondad de Ajuste

"Las pruebas de bondad de ajuste[11] se emplean para decidir cuándo un conjunto de datos se apega a una distribución de probabilidad dada".[12]

3.1.4.1.1 Prueba de Bondad de Ajuste Chi–Cuadrado

Entre la distribución real y la que se ha supuesto existen siempre diferencias. Para que la sustitución de la primera por la segunda sea legítima, hay que averiguar si estas diferencias pueden ser debidas al azar.

Para ello se tomará en cuenta el hecho de que la variable suma de cuadrados de la diferencia de efectivos –frecuencia observada (f_i)– divididos por los efectivos teóricos –frecuencia teórica calculada (f_t)– es decir

$$R = \sum_{i=1}^{i=k} \frac{(f_i - f_t)^2}{f_t}$$

está distribuida según una ley de χ^2 (ley chi–cuadrado) cuando estas diferencias son aleatorias.

Para emplear las tablas χ^2, se fija previamente el error o umbral de probabilidad, de acuerdo con las necesidades del problema. Si el perjuicio que ocasiona tomar como malo un ajuste, no siéndolo, es pequeño, convendrá tomar un nivel de significancia α (probabilidad de rechazar un ajuste bueno) grande, 5%. Solamente si el perjuicio es grande, se tomará $\alpha = 1\%$ ó $\alpha = 0.1\%$.

Como grados de libertad ν (para manejar las tablas χ^2) se toma el número de pares de valores efectivos k (después de agrupar para que $f_t > 5$), disminuido en el número m de parámetros estimados, todo esto menos 1, es decir

$$\nu = k - m - 1.$$

---

[11] La calidad del ajuste de una curva se basa principalmente en el criterio del error cuadrático, el cual se define como la suma de $\{f_i * -f_t\}^2$ sobre todos los intervalos del histograma. El ajuste que tenga menor error cuadrático será el que se elija.

[12] CANAVOS, *op. cit.* p. 363.

Si el valor resultante R es inferior al valor de χ^2 correspondiente a α y a los grados de libertad ν, no se rechaza la hipótesis. Si es superior se rechaza la hipótesis y por consiguiente habrá de buscar otra distribución.[13]

Ejemplo 3.1.4.1.1 Con la tabla de distribución de frecuencias obtenida en 1.7 (misma que se muestra nuevamente en el subcapítulo 3.1.1.2); y por medio de la prueba de chi–cuadrado con $\alpha = 0.5\%$ demuestre que el ajuste obedece a una distribución gama con los parámetros estimados en el ejemplo anterior.

Solución:

A partir de lo anterior se plantea la siguiente hipótesis:

H_0: Los datos se distribuyen Gama con:			H_1: Los datos no se distribuyen Gama con:	
$\hat{\theta} =$	15.4749	vs	$\hat{\theta} =$	15.4749
$\hat{\alpha} =$	3.7503		$\hat{\alpha} =$	3.7503

Después, se construyó la siguiente tabla:

Intervalos de Clase Límites		Intervalos de Clase Fronteras		Marcas de Clase x_i	Frecuencia f_i	Frecuencia Acumulada F_i	Probabilidad Acumulada Teórica $F(x)$	Probabilidad Teórica $f(x)$	Frecuencia Teórica f_t
3	16	2.5	16.5	9.5	6	6	0.0336	0.0336	1.85
17	30	16.5	30.5	23.5	6	12	0.1724	0.1388	7.63
31	44	30.5	44.5	37.5	6	18	0.3760	0.2036	11.20
45	58	44.5	58.5	51.5	9	27	0.5747	0.1987	10.93
59	72	58.5	72.5	65.5	4	31	0.7315	0.1568	8.62
73	86	72.5	86.5	79.5	10	41	0.8401	0.1086	5.97
87	100	86.5	100.5	93.5	14	55	0.9090	0.0689	3.79
101	114	100.5	114.5	107.5	0	55	0.9501	0.0411	2.26
115	128	114.5	128.5	121.5	0	55	0.9734	0.0233	1.28
129	142	128.5	142.5	135.5	0	55	0.9862	0.0128	0.70
143	156	142.5	156.5	149.5	0	55	0.9930	0.0068	0.37
157	170	156.5	170.5	163.5	0	55	0.9965	0.0035	0.19
171	184	170.5	184.5	177.5	0	55	0.9983	0.0018	0.10
185	198	184.5	198.5	191.5	0	55	0.9992	0.0009	0.05
199	212	198.5	212.5	205.5	0	55	0.9996	0.0004	0.02
213	226	212.5	226.5	219.5	0	55	0.9998	0.0002	0.01
227	240	226.5	240.5	233.5	0	55	0.9999	0.0001	0.01
241	254	240.5	254.5	247.5	0	55	1.0000	0.0000	0.00
					55				55.00

[13] ENCICLOPEDIA CIENTÍFICA CULTURAL – VOLUMEN ESTADÍSTICA p. 133.

y a partir de ésta se construyó la siguiente:

Fronteras		Frecuencia f_i		Frecuencia Teórica f_t	$f_t>5$	$\frac{(f_i - f_t)^2}{f_t}$
2.5	16.5	6		1.85		
16.5	30.5	6	12	7.63	9.48	0.6679
30.5	44.5	6	6	11.20	11.20	2.4126
44.5	58.5	9	9	10.93	10.93	0.3404
58.5	72.5	4	4	8.62	8.62	2.4786
72.5	86.5	10	10	5.97	5.97	2.7135
86.5	100.5	14		3.79		
100.5	114.5	0		2.26		
114.5	128.5	0		1.28		
128.5	142.5	0		0.70		
142.5	156.5	0		0.37		
156.5	170.5	0		0.19		
170.5	184.5	0		0.10		
184.5	198.5	0		0.05		
198.5	212.5	0		0.02		
212.5	226.5	0		0.01		
226.5	240.5	0		0.01		
240.5	254.5	0	14	0.00	8.79	3.0865
		55		55.00		R = **11.6995**

Considerando que, $\nu = k - m - 1 = 6 - 2 - 1 = 3$ se obtiene que $\chi^2_{(3, 0.005)} = 12.8381$.

Como R < $\chi^2_{(3, 0.005)}$ la hipótesis nula H_0 no se rechaza.

Por lo tanto, se puede tomar este ajuste como bueno.

3.1.4.1.2 Prueba de Bondad de Ajuste de Kolmogorov

La prueba de χ^2 para la verificación de la legitimidad de un ajuste tiene el inconveniente ya mencionado de exigir frecuencias teóricas de cierto tamaño ($f_t > 5$), por lo que no siempre puede utilizarse por insuficiencia de los mismos.

A partir de la ley de Kolmogorov, que estudia las diferencias entre las frecuencias acumuladas para los valores de la muestra con respecto a las probabilidades acumuladas correspondientes a la función de distribución acumulativa de la variable teórica, se ha ideado una prueba que permite verificar la bondad de un ajuste independientemente del número de efectivos –frecuencia observada (f_i)– que se disponga.

La prueba consta de las siguientes fases:

Fase 1. Se calcula el valor D_n

$$D_n = \max |F_i^* - F(x)|$$

en donde F_t^* son las frecuencias relativas acumuladas para cada uno de los n valores de la muestra y $F(x)$[14] son las probabilidades teóricas acumuladas correspondientes a los mismos valores.

Fase 2. Se compara el valor de D_n encontrado en la fase anterior con el valor de $D_n(1-\alpha)$ límite proporcionado en las tablas de la prueba de Kolmogorov para un determinado nivel de confianza $1-\alpha$.

Fase 3. Si D_n encontrado < $D_n(1-\alpha)$ en tablas $\Rightarrow$ H_0, no se rechaza con un nivel de confianza $1-\alpha$.

Si D_n encontrado > $D_n(1-\alpha)$ en tablas $\Rightarrow$ H_0, se rechaza con un margen de error α (α, generalmente es igual a 5% ó 1%).[15]

Ejemplo 3.1.4.2 Con la tabla de distribución de frecuencias obtenida en 1.7 (misma que se muestra nuevamente en el subcapítulo 3.1.1.2); y por medio de la prueba de Kolmogorov con $\alpha = 5\%$ demuestre que el ajuste obedece a una distribución gama con los parámetros estimados en el ejemplo 3.1.3.

---

[14] Función de Distribución Acumulativa.
Vid: Subcapítulos 2.4 y 2.5.
[15] *Ibid.* p. 129.

Solución:

A partir de lo anterior se plantea la siguiente hipótesis:

H_0:			H_1:	
Los datos se distribuyen Gama con:			Los datos no se distribuyen Gama con:	
$\hat{\theta}$ =	15.4749	VS	$\hat{\theta}$ =	15.4749
$\hat{\alpha}$ =	3.7503		$\hat{\alpha}$ =	3.7503

Después, se construyó la siguiente tabla:

Intervalos de Clase Límites		Fronteras		Marcas de Clase x_i	Frecuencia f_i	Frecuencia Acumulada F_i	Frecuencia Relativa Acumulada F_i^*	Probabilidad Acumulada Teórica $F(x)$	$\max\lvert F_i^* - F(x)\rvert$
3	16	2.5	16.5	9.5	6	6	0.1091	0.0336	0.075
17	30	16.5	30.5	23.5	6	12	0.2182	0.1724	0.046
31	44	30.5	44.5	37.5	6	18	0.3273	0.3760	0.049
45	58	44.5	58.5	51.5	9	27	0.4909	0.5747	0.084
59	72	58.5	72.5	65.5	4	31	0.5636	0.7315	0.168
73	86	72.5	86.5	79.5	10	41	0.7455	0.8401	0.095
87	100	86.5	100.5	93.5	14	55	1.0000	0.9090	0.091
101	114	100.5	114.5	107.5	0	55	1.0000	0.9501	0.050
115	128	114.5	128.5	121.5	0	55	1.0000	0.9734	0.027
129	142	128.5	142.5	135.5	0	55	1.0000	0.9862	0.014
143	156	142.5	156.5	149.5	0	55	1.0000	0.9930	0.007
157	170	156.5	170.5	163.5	0	55	1.0000	0.9965	0.004
171	184	170.5	184.5	177.5	0	55	1.0000	0.9983	0.002
185	198	184.5	198.5	191.5	0	55	1.0000	0.9992	0.001
199	212	198.5	212.5	205.5	0	55	1.0000	0.9996	0.000
213	226	212.5	226.5	219.5	0	55	1.0000	0.9998	0.000
227	240	226.5	240.5	233.5	0	55	1.0000	0.9999	0.000
241	254	240.5	254.5	247.5	0	55	1.0000	1.0000	0.000
					55				Dn= 0.168

Considerando que, $n = 18$ se obtiene que $D_{18}(0.95) = 0.318$.

Como $D_{18} < D_{18}(0.95)$ la hipótesis nula H_0 no se rechaza.

Por lo tanto, se puede tomar este ajuste como bueno.

Tanto la prueba chi–cuadrado como la de Kolmogorov calificaron este ajuste como bueno, sin embargo posiblemente existan otras distribuciones de probabilidad que satisfagan las pruebas de bondad de ajuste, por lo que se elegirá a la distribución que tenga el menor error cuadrático, como se mencionó en 3.1.4.1.

3.2 Determinación del Tamaño de Muestra[16]

Un *estimador puntual* es el valor numérico de una estadística muestral empleada para estimar (o aproximar) el valor de un parámetro de la población o proceso. Una de las características más importantes de un estimador es que sea insesgado. Un *estimador insesgado* es una estadística muestral cuyo valor esperado es igual al parámetro por estimar. Un *valor esperado* es el promedio a largo plazo de la estadística muestral. La eliminación de todo sesgo sistemático está asegurada cuando la estadística muestral corresponde a una *muestra aleatoria* tomada de una población o a un *subgrupo racional* tomado de un proceso. Ambos métodos de muestreo garantizan que la muestra sea insesgada, aunque no eliminan la variabilidad del muestreo, o *error de muestreo.*

En la tabla siguiente se presentan algunos de los estimadores puntuales de parámetros de la población de uso más frecuente.

Estimadores Puntuales de Uso más frecuente	
Parámetros de la Población	**Estimador**
Media, μ	$\bar{X}$
Proporción, π	$\hat{p}$
Varianza, σ^2	s^2
Desviación estándar, σ	s

TABLA 3.1 Estimadores puntuales de uso más frecuente.[17]

3.2.1 Determinación del Tamaño de Muestra Requerido para la Estimación de la Media[18]

Supongamos que se especifican el tamaño deseado de un intervalo de confianza y el nivel de confianza asociado con él. Si σ es conocida o puede estimarse –a partir de los estudios similares por ejemplo–, el tamaño de muestra requerido con base en el uso de la distribución normal es

[16] KAZMIER, *op. cit.* p. 133.
[17] *Idem.*
[18] *Ibid.* p. 133 – 139.

$n = \left(\frac{z\sigma}{E}\right)^2$	Cuando lo que se busca determinar es el tamaño de la muestra, todo resultado fraccionario se redondea siempre al valor inmediato superior. Adicionalmente, todo tamaño de muestra calculado por debajo de 30 debe ser incrementado a 30 porque la fórmula se basa en la distribución normal.

Donde:

$z =$ Es el valor utilizado para el nivel de confianza especificado.
$\sigma =$ Es la desviación estándar de la población (o su estimación).
$E =$ Es el error de muestreo de "más o de menos" permitido en el intervalo.

Además, considérese lo siguiente:

z **(número de unidades de desviación estándar respecto a la media)**	**Proporción de área en el intervalo** $\mu \pm z\sigma$
1.645	0.90
1.96	0.95
2.58	0.99

Cuando el error de muestreo no se conoce este puede ser estimado con la fórmula que a continuación se presenta:

$$E \approx \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

En este caso n se refiere al número de elementos en la muestra *preliminar* al muestreo que se va realizar.

Ejemplo 3.2.1 Un administrador de recursos humanos desea estimar el número medio de horas de capacitación al año para los supervisores con un margen de error inferior a 2 horas (de más o de menos) y confianza de 90%. Con base en datos de años previos el administrador estima que la desviación estándar es $\sigma = 20$ horas. El tamaño de muestra requerido es:

$$n = \left(\frac{z\sigma}{E}\right)^2 = \left(\frac{(1.645)(20)}{2.5}\right)^2 = \left(\frac{32.9}{2.5}\right)^2 = 173.18 \cong 174$$

3.2.2 Determinación del Tamaño de Muestra Requerido para la Estimación de la Proporción[19]

El tamaño de muestra mínimo requerido puede determinarse especificando el nivel de confianza requerido y el error de muestreo aceptable y haciendo una estimación inicial (subjetiva) de π, la proporción poblacional desconocida:

$n = \dfrac{z^2\pi(1-\pi)}{E^2}$	Cuando lo que se busca determinar es el tamaño de la muestra, todo resultado fraccionario se redondea siempre al valor inmediato superior. Adicionalmente, todo tamaño de muestra calculado por debajo de 100 se debe de incrementar a 100 porque la formula se basa en la distribución normal.

Donde:

$z =$ Es el valor utilizado para el nivel de confianza especificado.
$\pi =$ Es la estimación inicial de la proporción poblacional
$E =$ Es el error de muestreo "de más o de menos" permitido en el intervalo (siempre la mitad del intervalo de confianza completo).

Si no es posible determinar un estimado inicial de π, se le deberá estimar en 0.50. Esta estimación es conservadora en tanto que representa el valor para el que se requeriría el tamaño de muestra mayor.

Cuando el error de muestreo no se conoce este puede ser estimado con la fórmula del error estándar que a continuación se presenta:

$$E \approx s_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

En este caso n se refiere al número de elementos en la muestra *preliminar* al muestreo que se va realizar.

Ejemplo 3.2.2 El director administrativo de una universidad desea estimar, con un margen de error inferior a ± 0.05 y confianza del 90%, la proporción de estudiantes inscritos en programas de maestria en ingeniería de sistemas hayan cursado la carrera de actuaría. ¿Cuál sería el tamaño de muestra mínimo si los datos e información previos indican que la proporción no será mayor a 0.30?

[19] *Ibid*, p. 152 – 153.

Solución:

$$n = \frac{z^2\pi(1-\pi)}{E^2} = \frac{(1.645)^2(0.30)(0.70)}{(0.05)^2} = 227.31 \cong 228$$

Con este ejemplo, se concluye la teoría necesaria para efectuar el análisis de datos de un fenómeno o de varios; y con el cual se obtienen conjeturas y/o características de éstos, mismas que pueden ser representadas con un modelo del sistema; y con el uso de otra metodología conduce a la elaboración de inferencias del sistema, la metodología que esta tesis presenta es la de simulación y se expone en el siguiente capítulo.

CAPÍTULO 4
SIMULACIÓN

El modelado de simulaciones se basa principalmente en la computación, la probabilidad y la estadística, las bases necesarias de éstas dos últimas fueron presentadas en los capítulos previos, por esta razón en este capítulo se complementa a estos elementos, con el modelado de sistemas, esto para poder realizar experimentos de muestreo sobre un modelo del sistema, es decir realizar simulación.

4.1 Teoría General de Sistemas

La teoría General de Sistemas (TGS) se presenta como una forma sistemática y científica de aproximación y representación de la realidad y, al mismo tiempo, como una orientación hacia una práctica estimulante para formas de trabajo multidisciplinarias.

En la TGS lo importante son las relaciones y los conjuntos que a partir de ellas emergen; ofreciendo un ambiente adecuado para la interrelación y comunicación entre especialistas y especialidades.

Los objetivos originales de la Teoría General de Sistemas son los siguientes:

- Impulsar el desarrollo de una terminología general que permita describir las características, funciones y comportamientos sistémicos.
- Desarrollar un conjunto de leyes aplicables a todos estos comportamientos.
- Promover una formalización (matemática) de estas leyes.[1]

[1] RINCON, Juana, Cooperación del Personal Académico: Mecanismo para la Integración del Sistema Universitario Nacional. http://members.tripod.com/~gepsea/sistema.htm

4.2 El Sistema y Otros Conceptos Relacionados con éste[2]

En este subcapítulo se expondrán de manera conceptual las relaciones y los conjuntos importantes de un sistema.

Sistema:	Conjunto de entidades caracterizadas por algunos atributos que tienen relaciones entre sí y que están localizadas en un ambiente de acuerdo con un cierto objetivo.
Entidad:	Objeto concreto o abstracto de interés en el sistema sobre el que se recoge información que se representa en el sistema.
Atributos:	Características necesarias para describir completamente a cada tipo de entidad.
Relación:	Asociación natural entre dos o más entidades o entre sus atributos.
Ambiente:	Condiciones o circunstancias que favorecen o no al sistema.
Objetivo:	Actividad proyectada o planeada que se ha seleccionado antes de su ejecución y está basada tanto en apreciaciones subjetivas como en razonamientos técnicos de acuerdo con las características que posee el sistema.
Sistemas Abiertos:	Son aquellos que importan y procesan elementos de sus ambientes.[3]
Sistemas Cerrados:	Son aquellos en los cuales ningún elemento de afuera entra y ninguno sale del sistema.[4]
Subsistemas:	Se entiende por éstos a un conjunto de elementos y relaciones que responden a estructuras y funciones especializadas dentro de un sistema mayor. En términos generales, los subsistemas tienen las mismas propiedades que los sistemas.
Modelo del Sistema:	Es una representación del sistema o de una parte de éste.

---

[2] *Cfr. Ibid.*

[3] Ésta es una característica propia de todos los seres vivos. Que un sistema sea abierto significa que establece intercambios permanentes con su ambiente, intercambios que determinan su equilibrio, capacidad reproductiva o continuidad, es decir su viabilidad o permanencia en el entorno.

[4] Éstos alcanzan su estado de máximo equilibrio al igualarse con el medio. En ocasiones el término sistema cerrado es también aplicado a sistemas que se comportan de una manera fija, rítmica o sin variaciones.

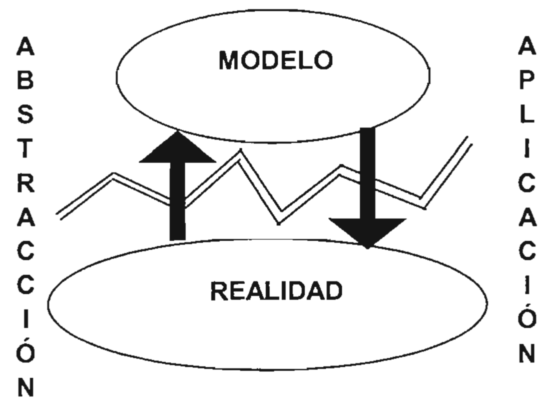


FIGURA 4.1 Rompimiento entre la Realidad y el Modelo.[5]

La figura anterior ilustra la relación entre el modelo y la realidad donde se aprecia un rompimiento entre ellos, este rompimiento se da en el proceso de abstracción al seleccionar los "elementos más importantes", y en la determinación de las reglas o leyes de relación entre dichos elementos, de ahí el sumo cuidado que el investigador debe procurar para la construcción del modelo.[6]

4.3 Simulación

Desde el punto de vista de la investigación de operaciones, la simulación es una *técnica de muestreo estadístico controlada* para *estimar* el desempeño de sistemas estocásticos complejos cuando los modelos analíticos no son suficientes. Más que describir el comportamiento global de un sistema directamente, el *modelo de simulación* describe la operación del mismo en términos de los *eventos individuales* de cada una de las componentes del sistema. En particular, el sistema se divide en elementos cuyo comportamiento se puede predecir, al menos en términos de distribuciones de probabilidad para cada uno de los diferentes *estados posibles del sistema* y sus insumos. También se construyen dentro del modelo las interrelaciones entre esos elementos.

Entonces, la simulación proporciona un medio para dividir el trabajo de construir un modelo en componentes más pequeñas que se pueden formular con mayor facilidad (por ejemplo, una

[5] SAUTTO Vallejo, José Maclovio, La técnica de la Simulación Digital en la Investigación de Operaciones. p. 10.
[6] *Idem.*

componente puede ser un sistema de líneas de espera sencillo) y después combina estas componentes en su orden natural. Una vez construido el modelo se activa usando números aleatorios para generar los eventos simulados a través del tiempo, de acuerdo con las distribuciones de probabilidad apropiadas. El resultado es una simulación de la operación del sistema real a través del tiempo en la que se puede registrar su comportamiento agregado. Si este proceso se repite para las diferentes configuraciones del diseño y las políticas de operación del sistema y se compara su desempeño, se pueden identificar las configuraciones que más prometen. Debido al error estadístico, es imposible garantizar que la configuración que produzca el mejor comportamiento simulado sea la óptima pero al menos debe ser cercana a la óptima si el experimento de simulación estuvo bien diseñado.

Así, la simulación, en general, no es otra cosa que la técnica de realizar *experimentos de muestreo* sobre el modelo del sistema. Los experimentos se llevan a cabo en el modelo en lugar de hacerlo en el propio sistema real ya que esto último resultaría inconveniente, muy caro y muy tardado. Por lo demás, los experimentos simulados deben considerarse en esencia iguales a los experimentos estadísticos comunes, por lo que también deberán tener como fundamento la teoría estadística formal.[7]

4.4 Etapas de un Estudio de Simulación[8]

Resultaría limitante o redundante dar una serie de pasos a seguir al resolver problemas de simulación, sin embargo algunos autores que escriben al respecto proponen ciertos pasos de acuerdo a su experiencia personal, a continuación se compilan algunos de estos no como una secuencia a seguir, sino como una referencia de la cual partir:

1. Definición del Sistema. Para tener una definición exacta del sistema, se debe de hacer un análisis en el que se determine, cuáles son las entidades del mismo, consecuentemente sus atributos, relaciones y el ambiente.
2. Colección y Análisis de Datos. Una vez definido lo anterior, se deben recolectar datos sobre cada uno de los componentes del sistema; después proceder a un análisis estadístico de éstos; ya que los resultados de éste facilitaran el siguiente paso.

---

[7] HILLER, Frederick S. y Gerald J. LIEBERMAN, <u>Introducción a la Investigación de Operaciones</u>, p. 900–902.
[8] *Cfr.* COSS Bu, Raúl, <u>Simulación un Enfoque Práctico</u>, p. 12–14.
SAUTTO *op. cit.* 12–14.

3. Formulación del Modelo. Definir y construir el modelo con el cual se espera obtener los resultados deseados.
4. Implementar el Modelo en la Computadora. En este paso el modelo del sistema se lleva a la computadora, ya sea que éste se programe en algún lenguaje de computación o utilizando un paquete.
5. Validación y Calibración. En este paso el modelo se alimenta con ciertas condiciones conocidas del sistema, se comparan los resultados obtenidos con los esperados, en caso de que estos difieran, se mueven los parámetros de calibración del sistema hasta obtener los resultados esperados. Este proceso conduce con frecuencia a un ciclo que puede conducir al paso 3.
6. Experimentación. Consiste en generar los datos deseados. En esta etapa se contestan las siguientes preguntas: ¿Qué grado de precisión requiere? y consecuentemente ¿Tamaño de Muestra?
7. Interpretación. En esta etapa se interpretan los resultados que arroja la simulación y se elaboran las sugerencias.
8. Implementación de la Solución. Una vez aprobadas las sugerencias se procede a su implementación.

4.5 Ejemplo de Simulación

Considere el modelo M/M/1 de teoría de colas (entradas Poisson, tiempo de servicio exponencial y un solo servidor). En particular, suponga que los valores de la tasa de llegadas λ y la tasa de servicio μ son:

$$\lambda = 3 \text{ por hora} \quad \mu = 5 \text{ por hora}$$

Para resumir la operación física del sistema, los clientes que llegan se unen a una cola, eventualmente son servidos y después se van. Es necesario que el modelo de simulación describa y sincronice la llegada de los clientes y el servicio de los mismos.

Comenzando en el tiempo 0, el reloj simulador registra el tiempo (simulado) t que ha transcurrido hasta ahora durante la corrida de simulación. La información sobre el sistema de colas que define el estado actual, es decir, el estado del sistema es

$N(t) =$ Número de clientes en el sistema en el tiempo *t*.

Los eventos que cambian el estado del sistema son la llegada de un cliente o la terminación del servicio del cliente que se está atendiendo en este momento (si lo hay). El mecanismo generador de eventos se describirá un poco más adelante. El mecanismo de transición de estados es

$$\text{Restablecer } N(t) = \begin{cases} N(t)+1, & \text{Si ocurre una llegada en el tiempo } t. \\ N(t)-1, & \text{Si termina un servicio en el tiempo } t. \end{cases}$$

Existen métodos para hacer que avance el reloj simulador y registrar la operación del sistema. Ahora se describirá e ilustrarán de *incrementos de tiempo fijo* que sirve para avanzar el tiempo.

En el mecanismo de incrementos de tiempo fijo se usa repetidamente el siguiente procedimiento de dos pasos:

Resumen de Incrementos de Tiempo Fijo

1. Se avanza el tiempo una cantidad fija pequeña.
2. Se actualiza el sistema determinando los eventos que ocurrieron durante este lapso y el estado del sistema que resulta. También se registra la información deseada sobre el comportamiento del sistema.

En el caso del modelo de líneas de espera que se analiza ahora, sólo pueden ocurrir dos tipos de eventos durante cada uno de estos intervalos, a saber, una o más llegadas y una o más terminaciones de servicio. Es más, la probabilidad de que ocurran dos o más llegadas o dos o más terminaciones de servicio durante una unidad de tiempo es despreciable para este modelo si la unidad de tiempo es relativamente pequeña. Entonces, los únicos dos eventos posibles en esa unidad de tiempo, que se tienen que investigar son la llegada de un cliente y la terminación de servicio de un cliente. Cada uno de estos eventos tiene una probabilidad conocida.

A manera de ejemplo se usará 0.1 horas (6 minutos) como la cantidad fija pequeña en que se avanza el reloj cada vez. (Normalmente se usaría un intervalo mucho más chico para que en realidad sea despreciable la probabilidad de llegadas o terminaciones de servicio múltiples, pero esta elección creará más movimientos con el fin de ejemplificar). Debido a que tanto los tiempos entre llegadas como los de servicio tienen una distribución exponencial, la probabilidad P_{Ll} de que un intervalo de 0.1 horas incluya una *llegada* es

$$P_{Ll} = 1 - e^{-3/10} = 0.259,$$

y la probabilidad P_S de que incluya una salida (terminación de servicio), dado que se estaba sirviendo un cliente al inicio del intervalo es

$$P_{Ll} = 1 - e^{-5/10} = 0.393.$$

Para generar aleatoriamente cualquiera de los dos eventos de acuerdo a estas probabilidades. La computadora se usa para generar un *número aleatorio uniforme* en el intervalo (0,1), es decir, se tiene una observación aleatoria a partir de una distribución uniforme entre 0 y 1. Si este número aleatorio se denota por r_{Ll},

$r_{Ll} < 0.259 \Rightarrow$ ocurre una llegada
$r_{Ll} \geq 0.259 \Rightarrow$ no ocurre una llegada,

De igual manera, con otro número aleatorio uniforme r_S,

$r_S < 0.393 \Rightarrow$ ocurre una salida
$r_S \geq 0.393 \Rightarrow$ no ocurre una salida,

dado que había un cliente en servicio al principio del intervalo. Sin clientes en servicio (es decir, sin clientes en el sistema) se supone que no pueden ocurrir salidas durante el intervalo aunque haya ocurrido una llegada.

La siguiente tabla, muestra los resultados al usar este enfoque para 10 iteraciones del procedimiento de *incrementos de tiempo fijo* comenzando sin clientes en el sistema y usando minutos como unidad de tiempo.

t tiempo (min)	r_{LI}	¿Llegada en el intervalo?	r_S	¿Salida en el intervalo?	*N(t)*
0	-	-	-	-	0= 0
6	0.722	No	-	-	0+0-0= 0
12	0.577	No	-	-	0+0-0= 0
18	0.138	Si	0.389	Si	0+1-1= 0
24	0.410	No	-	-	0+0-0= 0
30	0.845	No	-	-	0+0-1= 0
36	0.931	No	-	-	0+0-0= 0
42	0.014	Si	0.768	No	0+1-0= 1
48	0.026	Si	0.097	Si	1+1-1= 1
54	0.557	No	0.210	Si	1+0-1= 0
60	0.228	Si	0.689	No	0+1-0= 0

Como se observó en la tabla, el siguiente paso del procedimiento (actualizar el sistema) incluye el registro de las medidas de desempeño deseadas sobre el comportamiento agregado al sistema durante este intervalo.[9]

En el ejemplo anterior se apreció que en simulación es necesario tener conocimiento de lo siguiente:

1. Generación de números aleatorios uniformes en el intervalo (0,1).
2. Generación de variables aleatorias no–uniformes.
3. Conocer la notación empleada en teoría de colas.

Por esta razón, en los siguientes subcapítulos se desarrollan estos temas.

4.6 Método Congruencial Mixto para la Generación de Números Aleatorios con Distribución Uniforme en el Intervalo (0,1)

Los generadores congrueneciales lineales generan una secuencia de números pseudoaleatorios en la cual el próximo número pseudoaleatorio es determinado a partir del último número generado, es decir el número pseudoaleatorio X_{n+1} es derivado a partir del número pseudoaleatorio X_n.

[9] HILLER, *op. cit.* p. 904-906.

Para el caso particular del generador congruencial mixto, la relación de recurrencia es la siguiente:

$$X_{n+1} = (aX_n + c) \bmod m$$

donde:

$X_0 =$ la semilla $(X_0 > 0)$
$a =$ el multiplicador $(a > 0)$
$c =$ constante aditiva $(c > 0)$
$m =$ El módulo $(m > X_0, m > a \text{ y } m > c)$

Esta relación de recurrencia dice que X_{n+1} es el residuo de dividir $aX_n + c$ entre el módulo. Lo anterior significa que los valores posibles de X_{n+1} son 0, 1, 2, ..., $m-1$, es decir m representa el número posible de valores diferentes que pueden ser generados.

Con el propósito de ilustrar la generación de números pseudoaleatorios a través de este método, suponga que se tiene un generador en el cual los valores de los parámetros son:

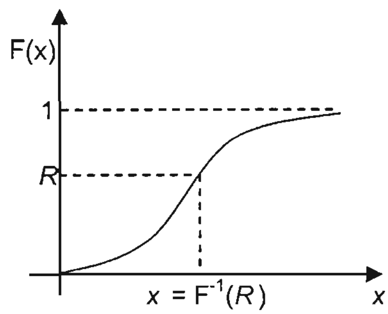
$$a = 5,\ c = 7,\ X_0 = 4 \text{ y } m = 8.$$

Para estos valores, la secuencia de números pseudoaleatorios y números uniformes (X_{n+1} / m) son mostrados en la siguiente tabla, el período del generador es 8.[10]

n	**X_n**	**$(5X_n+7)/8$**	**X_n+1**	**Números Uniformes**
0	4	3 + 3/8	3	3/8
1	3	2 + 6/8	6	6/8
2	6	4 + 5/8	5	5/8
3	5	4 + 0/8	0	0
4	0	0 + 7/8	7	7/8
5	7	5 + 2/8	2	2/8
6	2	2 + 1/8	1	1/8
7	1	1 + 4/8	4	4/8

[10] COSS, *op. cit.* p. 19 – 29.

4.7 Método de la Transformada Inversa para la Generación de Variables Aleatorias No–Uniformes[11]



El método de la transformada inversa utiliza la distribución acumulada $F(x)$ de la distribución que se va a simular. Puesto que $F(x)$ está definida en el intervalo (0,1), se puede generar un numero aleatorio uniforme R y tratar de determinar el valor de la variable aleatoria para la cual su distribución acumulada es igual a R, es decir, el valor simulado de la variable aleatoria que sigue una distribución de probabilidad $f(x)$, se determina al resolver la siguiente ecuación:

$$F(x) = R \Rightarrow x = F^{-1}(R)$$

La dificultad principal de este método descansa en el hecho de que en algunas ocasiones es difícil encontrar la transformada inversa.

Ejemplo 4.7 Distribución exponencial

Se desea generar números al azar que sigan la siguiente distribución de probabilidad:

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases}$$

La distribución acumulada de esta distribución es:

$$F(x) = \int_0^x \lambda e^{-\lambda x} dt = 1 - e^{-\lambda x}$$

[11] *Ibid.* p. 49 – 53.

y utilizando el método de la transformada inversa, es decir, igualando la distribución acumulada con el número uniforme *R*, se obtiene:

$$1 - e^{-\lambda x} = R$$

$$e^{-\lambda x} = 1 - R$$

Pero si R sigue una distribución uniforme, entonces $1 - R$ también sigue esta distribución. Por consiguiente:

$$x = -\frac{1}{\lambda} Ln\, R\,.$$

4.8 Notación Utilizada en Teoría de Colas

Debido a que la simulación es una técnica que se aplica frecuentemente en teoría de colas, se proporciona la notación empleada en esta área:

Característica	Símbolo	Explicación
Distribución de tiempo entre llegadas (A)	M D Ek G1	Exponencial o Poisson Determinística Erlang General independiente
Distribución de tiempo de servicio (B)	M D Ek G	Exponencial o Poisson Determinística Erlang General
Número de estaciones paralelas (X)	1, 2, …, X	
Restricciones de capacidad del sistema (Y)	1, 2, …, X	
Disciplina de espera (Z)	FIFO LIFO SIRO PRI GD	Primero en entrar Primero en salir Último en entrar Último en salir Servicio en forma aleatoria Prioridad Disciplina General

La representación de una cola es de la siguiente forma A/B/X/Y/Z.

Normalmente se utilizan los primeros tres símbolos, omitiéndose los símbolos Y y Z. Así M/D/2 representa un sistema de espera con tiempo entre llegadas exponencial, servicio determinístico, dos estaciones de servicio, ningún limite en la capacidad del servicio y como disciplina primero en llegar primero en servicio.[12]

En los capítulos previos así como en este, se han desarrollado y analizado los conceptos y herramientas requeridas para el uso de la estimación de la distribución de probabilidad y de la simulación; por consiguiente en el capítulo precedente se avanzará con una aplicación en el que se conjunten éstas técnicas.

[12] SAUTTO, *op. cit.* p. 143.

CAPÍTULO 5

CASO PRÁCTICO: DETERMINACIÓN DE LAS CUOTAS DE TARIFA PARA EL SEGURO DE RESPONSABILIDAD CIVIL AGENTES DE SEGUROS (PERSONAS FÍSICAS)

En el presente capítulo se conjunta a los elementos de la Matemática del Seguro con los que se presentaron en las páginas previas, para determinar sobre bases técnicas, las primas netas de riesgo a fin de garantizar con un elevado grado de certidumbre, el cumplimiento de las obligaciones que contraigan con sus asegurados las instituciones de seguros, estas bases deben de ser presentadas en una nota técnica, como lo mencionan los artículos 36 y 36-A entre otros de la Ley General de Instituciones y Sociedades Mutualistas de Seguros[1].

Se entiende por nota técnica a "la documentación que integra el conjunto de procedimientos técnicos esenciales en el diseño de un seguro para efectos de su registro; ya sea que se trate de un seguro nuevo o de la sustitución de uno previamente registrado".[2]

En nuestro país es la Comisión Nacional de Seguros y Fianzas (C.N.S.F.) el organismo encargado de conocer y resolver sobre los registros, control, revisión, supervisión, etcétera de cualquier actividad aseguradora; entre éstos el procedimiento para el registro de un seguro o nota técnica.

La CNSF hace del conocimiento de los sectores supervisados, los consumidores y el público en general, las disposiciones especificas que emite con base en un marco jurídico aplicable, a través de circulares que se publican en el Diario Oficial de la Federación (DOF) y dentro de éstas, se encuentran las que regulan el registro de Tarifas y Documentación Contractual, entre las que se destaca a la circular S–8.1[3]; en la que se señala la forma y términos para el registro de productos de seguros, la disposición décima segunda menciona los datos que éstas deberán contener según apliquen para la operación, ramo y seguro que se trate. Atendiendo ésas disposiciones se elaboró el presente capitulo el cual corresponde a una nota técnica para el Seguro de Responsabilidad Civil Agentes de Seguros (Personas Físicas), misma que a continuación se presenta.

---

[1] Ley General de Instituciones y Sociedades Mutualistas de Seguros.
[2] VARGAS Aguilar, Juan Carlos, Fundamentos para el Desarrollo de Productos en Daños, p. 96.
[3] Diario Oficial de la Federación, 20 de Febrero de 2004.

NOTA TÉCNICA DE LA TARIFA PARA EL SEGURO DE RESPONSABILIDAD CIVIL AGENTES DE SEGUROS (PERSONAS FÍSICAS)

ÍNDICE[4]

[4] En este capítulo la numeración cambia para no confundir al lector con la de la nota técnica.

i CARACTERISTICAS GENERALES DEL PLAN

i.1 Nombre Comercial del Plan

El plan que se envía a ustedes para efectos de registro y vigilancia se denomina **SEGURO DE RESPONSABILIDAD CIVIL PARA AGENTES DE SEGUROS (Personas Físicas).**

i.2 Descripción de la Cobertura Básica

La Compañía se obliga a reparar o, a su elección, indemnizar los daños y perjuicios que el Asegurado cause al público, conforme a la legislación en materia de responsabilidad civil vigente en los Estados Unidos Mexicanos, por actos negligentes o imperitos (acciones u omisiones); resultantes de su actividad de agente de seguros, ocurridos y reclamados al Asegurado durante la vigencia de esta póliza, amparándose las siguientes responsabilidades:

1. La responsabilidad por daños directos al patrimonio:

   Se entenderán bajo este concepto. Los daños en sentido estricto (pérdidas o menoscabos sufridos en el patrimonio), así como los perjuicios (privaciones de ganancia lícita que necesariamente se hubiese obtenido de no haberse presentado los daños) que se causen a los clientes del Asegurado, sin incluir daños corporales o daños materiales a sus propiedades o un perjuicio consecuencial a ambos.

   La responsabilidad de la Compañía empieza después de la aplicación del deducible a cargo del Asegurado.

2. La responsabilidad por pérdida o destrucción de documentos:

   La responsabilidad civil legal por daños y perjuicios que sufran los clientes del Asegurado por daños materiales destrucción o pérdidas de documentos que los clientes le hayan entregado para el desarrollo de las actividades encomendadas.

No se incluyen bajo el concepto de documentos, dinero, moneda extranjera, otros signos pecuniarios, metales amonedados, títulos de crédito, valores, cualquier otro título representativo de dinero, mercancías, valores o promesas, por último, archivos o almacenamientos para el procesamiento electrónico de datos.

Las reclamaciones contra este concepto procederán de acuerdo con el orden siguiente:

Primero

En el caso de pérdida de documentos, el Asegurado procederá a una búsqueda diligente y de ella levantará una constancia circunstanciada.

Segundo

En los demás casos la Compañía determinará si procede efectuar gestiones extrajudiciales, judiciales o técnicas para obtener la restauración o la reposición de documentos; los gastos y costos serán por cuenta de la Compañía, después de aplicar el deducible a cargo del Asegurado.

Tercero

Después de agotar las posibilidades de encuentro, restauración o reposición, la Compañía determinará la procedencia y extremos de la responsabilidad civil legal, para proceder a indemnizar al cliente del Asegurado; después de la aplicación del deducible a cargo de éste.

3. La responsabilidad civil legal personal de los empleados y trabajadores del Asegurado frente al público (quienes se consideraran Asegurados), derivada del ejercicio de la actividad materia de este seguro.

i.3 Temporalidad del Plan

La vigencia de este plan puede ser menor o igual a un año.[5]

i.4 Operación y Ramo en el que se Registrará

Este plan se utilizará dentro de las operaciones de **DAÑOS**, en el ramo de **RESPONSABILIDAD CIVIL Y RIESGOS PROFESIONALES.**

ii HIPÓTESIS ESTADÍSTICAS Y FINANCIERAS

La Compañía cuenta con una cartera escasa para este tipo de seguros, ya que tiempo antes no era obligatorio para los agentes de seguros contratar una póliza por errores u omisiones, por este motivo cuando se han suscrito negocios de este tipo se ha hecho siguiendo las medidas recomendadas por nuestro reasegurador (reaseguro facultativo) en el ramo de responsabilidad civil.

ii.1 Hipótesis Estadísticas

La información de nuestra cartera para los años 1998, 1999 y 2000 misma que se muestra en el anexo 1; aunque escasa permite hacer inferencias con impacto positivo, con el uso de los métodos que se detallan en procedimientos técnicos.

ii.2 Hipótesis Financieras

ii.2.1 Utilidad Técnica

La utilidad técnica mínima esperada es del 5%.[6]

---

[5] Generalmente en los seguros para operaciones en daños se otorga la cobertura menor o igual a un año.

[6] Es la utilidad que espera (planea y desea) obtener el asegurador después del pago de los siniestros y de los gastos de administración y de adquisición.

iii PROCEDIMIENTOS TÉCNICOS

iii.1 Gastos de Administración

Los Gastos de Administración considerados serán del 15%.

iii.2 Gastos de Adquisición

Los Gastos de Adquisición considerados serán del 9%.

iii.3 Gastos Totales

Gastos Totales = Gastos de Administración + Gastos de Adquisición + Utilidad Técnica.

iii.4 Procedimientos para la Generación de Datos Mediante Simulación Estocástica

Debido a que los datos de la cartera no constituyen una muestra suficientemente grande, estos serán utilizados como referencia para generar más datos con el supuesto de provenir de la misma función de distribución, para tal objetivo los procedimientos utilizados serán:

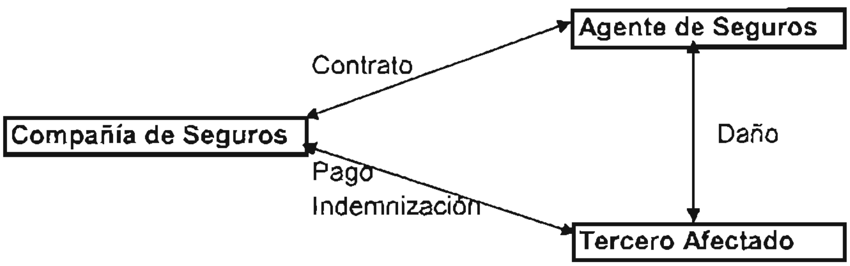
- Definición del Sistema.
- Análisis de Datos:
    - Estimar la Distribución de Probabilidad del Volumen de Prima Intermediada por Agente.
    - Estimar la Distribución de Probabilidad Empírica de los Siniestros utilizando el Método del Rango Mediano para Muestras Pequeñas.
    - Determinar las Probabilidades de Siniestro para Diferentes Montos de Prima Intermediada.
- Formulación del Modelo.
- Implementar el Modelo en la Computadora.
- Experimentación.
    - Determinar el Tamaño de Muestra Requerido.

- Generación de Carteras del Seguro de Responsabilidad Civil Agentes de Seguros (Personas Físicas) Utilizando Simulación Estocástica.

➢ Interpretación de los Resultados.

iii.4.1 Definición del Sistema[7]

MODELO DEL SISTEMA



En la figura anterior se muestra un modelo del sistema, en este mismo se aprecia que el contrato de seguro cubrirá al asegurado por errores u omisiones que éste cause a terceros en su actividad de agente de seguros.[8]

Para entender mejor el sistema observe la siguiente tabla:

RELACIONES EN EL SISTEMA		
Entidad	**Atributo**	**Relaciones con:**
Compañía de Seguros	Entidad Financiera que asegura al Agente de seguros mediante el pago de una prima y consecuentemente pagará los daños que éste cause a un tercero hasta el límite de responsabilidad contratado.	Agente de Seguros Tercero Afectado
Agente de Seguros	Intermediario (para asesoría y venta de seguros) entre la Compañía y el público en general.	Compañía de Seguros Tercero Afectado
Tercero Afectado	Persona a quien le causa un daño el asegurado (agente de seguros) y que se conoce al momento del siniestro.	Agente de Seguros Compañía de Seguros

[7] *Vid*: 4.2 El Sistema y Otros Conceptos Relacionados con éste.
[8] *Vid*: i.2 Descripción de la Cobertura Básica.

iii.4.2 Análisis de Datos

iii.4.2.1 Estimar la Distribución de Probabilidad del Volumen de Prima Intermediada por Agente[9]

Se utilizó el software Arena para estimar la Distribución de Probabilidad que mejor se ajusta a los datos del Anexo 2 obteniéndose los siguientes resultados:

Distribution Summary

Distribution: Exponential[10]
Expression: 0.999 + EXPO(1.32e+006)
Square Error: 0.011007

Chi Square Test[11]
Number of intervals = 4
Degrees of freedom = 2
Test Statistic = 8.07
Corresponding p-value = 0.0193

Kolmogorov-Smirnov Test[12]
Test Statistic = 0.499
Corresponding p-value < 0.01

Data Summary

Number of Data Points = 168
Min Data Value = 1
Max Data Value = 1.42e+007
Sample Mean = 1.32e+006
Sample Std Dev = 2.38e+006

Histogram Summary

Histogram Range = 0.999 to 1.42e+007
Number of Intervals = 12

Function	Sq Error
Erlang	0.011
Exponential	0.011
Beta	0.0152
Weibull	0.0442
Gamma	0.159
Normal	0.237
Lognormal	0.278
Triangular	0.329
Uniform	0.388

[9] *Vid*: Capítulo 1 Estadística Descriptiva.
[10] *Vid*: 2.2.7.5 Distribución Exponencial Negativa.
[11] *Vid*: 3.1.4.1.1 Prueba de Bondad de Ajuste Chi–Cuadrado.
[12] *Vid*: 3.1.4.1.2 Prueba de Bondad de Ajuste de Kolmogorov.

Es decir que la Distribución de Probabilidad que mejor aproxima el Volumen de Prima Intermediada por Agente es:

$$f(x;\theta) = 0.999 + \begin{cases} \frac{1}{\theta} e^{-\left(\frac{x}{\theta}\right)} & x \succ 0 \\ 0 & \text{para cualquier otro caso} \end{cases} \quad \text{donde: } \theta = 1'320{,}000$$

iii.4.2.2 Estimar la Distribución de Probabilidad Empírica de los Siniestros Utilizando el Método del Rango Mediano para Muestras Pequeñas[13]

En la siguiente tabla se muestran los siniestros presentados en los años de 1998 al 2000 y cuyas cifras están en moneda nacional:

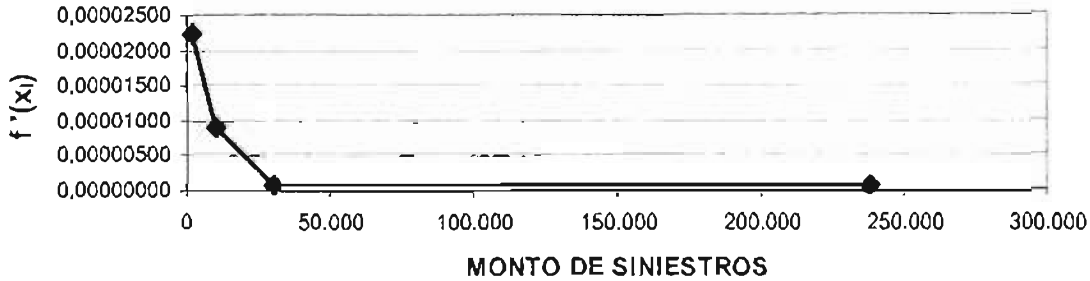
SINIESTRO	
AÑO EN QUE SE PRESENTO	MONTO
1998	2,000.00
1998	10,250.40
1999	427,845.73
2000	238,501.00
2000	30,404.61

A estos datos se les aplicó el método del rango mediano para muestras pequeñas obteniéndose lo siguiente:

SINIESTRO			
ÍNDICE	MONTO	$F'(x_i)$	$f'(x_i)$
1	2,000.00	0.1796	0.00002245
2	10,250.40	0.3648	0.00000919
3	30,404.61	0.5500	0.00000089
4	238,501.00	0.7352	0.00000098
5	427,845.73	0.9204	

[13] *Vid*: 3.1.1.1 Método del Rango Mediano para la Estimación de la Distribución de Probabilidad para Muestras Pequeñas.

DISTRIBUCIÓN EMPÍRICA DE SINIESTROS



Una vez que a los datos se les aplicó la metodología para muestras pequeñas y se obtuvo el gráfico arriba mostrado, después se utilizó el software Arena para determinar qué función de distribución se aproximaba más; el resultado fue una distribución gama, sin embargo esto se tiene que probar.

Distribution Summary	Function	Sq Error
Distribution: Gamma[14]	Gamma	0.292
Expression: -0.001+GAMM(8.29e+003,	Exponential	0.298
0.121)	Erlang	0.298
Square Error: 0.291965	Uniform	0.3
	Weibull	0.306
Kolmogorov-Smirnov Test[15]	Lognormal	0.345
Test Statistic = 1.23e+003	Normal	0.364
Corresponding p-value < 0.01	Triangular	0.416
	Beta	-1.#J
Data Summary		
Number of Data Points = 2		
Min Data Value = 0		
Max Data Value = 2e+003		
Sample Mean = 1e+003		
Sample Std Dev = 1.41e+003		
Histogram Summary		
Histogram Range = -0.001 to 2e+003		
Number of Intervals = 5		

[14] *Vid:* 2.7.2.3 Distribución Gama.

[15] *Vid:* 3.1.4.1.2 Prueba de Bondad de Ajuste de Kolmogorov.

A partir de lo anterior se plantea la siguiente hipótesis:

H_0: Los siniestros se distribuyen Gama	vs	H_1: Los siniestros no se distribuyen Gama

Para probar lo anterior debemos estimar los parámetros de esta distribución, recordemos lo siguiente:

Se dice que la variable aleatoria X tiene una distribución gama si su función de densidad de probabilidad está dada por:

Función de Densidad de Probabilidad	Parámetros
$f(x;\alpha;\theta)=\frac{1}{\Gamma(\alpha)\theta^{\alpha}}x^{\alpha-1}e^{\left(-x/\theta\right)}$	$\alpha \succ 0$ $\theta \succ 0$ $x \succ 0$
Media	**Varianza**
$\alpha\theta$	$\alpha\theta^2$

Utilizando el método de los momentos[16] se estimaron los parámetros que a continuación se presentan:

$$\text{media} = \quad \bar{X} = \quad 141{,}800.35 \ = \alpha\theta \qquad\qquad \hat{\theta} = 247{,}586.69$$

$$\text{Varianza} = \quad s^2 = \quad 35{,}107'878{,}306 \ = \alpha\theta^2 \quad \Rightarrow \quad \hat{\alpha} = 0.5727301$$

Con los parámetros anteriores se obtienen las Probabilidades Teóricas Acumuladas $F(x_i)$, y aplicando la prueba de Kolmogorov – Smirnov[17] se llega a la siguiente tabla:

PRUEBA DE HIPÓTESIS

Siniestro	F'(xi)	F(xi)	Dn
2.000,00	0,17962963	0,07086374	0,10876589
10.250,40	0,36481481	0,17850845	0,18630636
30.404,61	0,55000000	0,32321123	**0,22678877**
238.501,00	0,73518519	0,80490556	0,06972038
427.845,73	0,92037037	0,92327200	0,00290163

[16] *Vid:* 3.1.3 Método de los Momentos para la Estimación de Parámetros de una Distribución de Probabilidad.
[17] *Vid:* 3.1.4.1.2 Prueba de Bondad de Ajuste de Kolmogorov.

De la tabla anterior, se tiene que:

Dn encontrado	0.227
Dn(α) de Kolmogorov – Smirnov α=5% (1-α=95%)	0.565

Como Dn encontrado < Dn(α) de Kolmogorov – Smirnov, no se rechaza la hipótesis nula.

Por lo tanto, la Distribución de Probabilidad estimada de los siniestros para este seguro es:

$$f(x;\alpha,\theta)=\begin{cases}\dfrac{1}{\Gamma(\alpha)\theta^{\alpha}}x^{\alpha-1}e^{-\frac{x}{\theta}} & x \succ 0\\ 0 & \text{para cualquier otro caso}\end{cases} \qquad \text{donde: } \begin{array}{l}\hat{\alpha}=0.5727301\\ \hat{\theta}=247{,}586.69\end{array}$$

iii.4.2.3 Determinar las Probabilidades de Siniestro para Diferentes Montos de Prima Intermediada

Se ordenaron los datos de las tres carteras de mayor a menor con el supuesto de que el Monto de Suma Asegurada es igual (o muy semejante) al Monto de Prima Intermediada, se calculó la frecuencia de siniestros y se considero ésta como probabilidad. Obteniéndose los siguientes resultados.

Monto Promedio de Prima Intermediada	Probabilidad de Siniestro
Hasta $1'500,000.00	0.0217
Mayor a $1'500,000.00	0.0732

iii.4.3 Formulación del Modelo[18]

Con los datos obtenidos en el subcapítulo iii.4.2 Análisis de Datos se creó el modelo de las carteras de la siguiente forma:

1. Se generarán tres números aleatorios a, b, y c; mismos que siguen una Distribución de Probabilidad Uniforme, y su uso es el siguiente:
   a. Determinar el Volumen de Prima Promedio Intermediada por Agente.
   b. Determinar si se presentó o no un siniestro.
   c. Determinar el Monto del Siniestro.
2. El Límite de Responsabilidad que se otorgará será de \$100,000.00 por esta razón se hará constante para todos los elementos en la cartera.
3. El Volumen de Prima Promedio Intermediada por Agente se obtendrá a partir del número aleatorio a, el cual se asume como probabilidad y por medio de la función inversa de la siguiente distribución de probabilidad:

$$f(x;\theta) = 0.999 + \begin{cases} \frac{1}{\theta} e^{-\left(\frac{x}{\theta}\right)} & x \succ 0 \\ 0 & \text{para cualquier otro caso} \end{cases} \quad \text{donde: } \theta = 1'320,000.$$

4. La existencia o no de siniestro se determinará utilizando la siguiente regla:

$$\text{Siniestro} = \begin{cases} 1 & \text{si } b \leq 0.0217 \text{ y el Volumen Promedio de Prima Intermediada} \leq \$1'500,000.00 \\ 1 & \text{si } b \leq 0.0732 \text{ y el Volumen Promedio de Prima Intermediada} > \$1'500,000.00 \\ 0 & \text{e.o.c.} \end{cases}$$

5. El Monto del Siniestro se obtiene a partir del número aleatorio c, el cual se asume como probabilidad y por medio de la de la función inversa de la siguiente distribución de probabilidad:

$$f(x;\alpha,\theta) = \begin{cases} \frac{1}{\Gamma(\alpha)\theta^{\alpha}} x^{\alpha-1} e^{-\frac{x}{\theta}} & x \succ 0 \\ 0 & \text{para cualquier otro caso} \end{cases} \quad \text{donde: } \begin{matrix} \alpha = 0.5727301 \\ \theta = 247,586.69 \end{matrix}$$

[18] *Vid:* 4.4 Etapas de un Estudio de Simulación – 3.
4.5 Ejemplo de Simulación.

iii.4.4 Implementar el Modelo en la Computadora[19]

En este caso el modelo se ejecutará en hoja de cálculo, atendiendo que entre sus funciones se encuentran las necesarias para implementar el modelo y permite almacenar los datos en un archivo electrónico.

iii.4.5 Experimentación[20]

iii.4.5.1 Determinar el Tamaño de Muestra[21]

Para determinar el tamaño de muestra, se toma como base la siguiente información:[22]

	Año		
	1998	**1999**	**2000**
Siniestros	1	2	2
Elementos en la cartera	37	39	57
Frecuencia	0,02703	0,05128	0,03509
Frecuencia más Probable	0,03780	≈	2/53

Con la siguiente formula calcularemos el error estándar estimado de la proporción es,

$$E = S_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$\hat{p} = 0.03780$

n = 53

Sustituyendo estos valores se obtiene lo siguiente:

E = 0.02617498

---

[19] *Vid:* 4.4 Etapas de un Estudio de Simulación – 4.

[20] *Vid:* 4.4 Etapas de un Estudio de Simulación – 6.

[21] *Vid:* 3.2.2 Determinación del Tamaño de Muestra Requerido para la Estimación de la Proporción.

[22] En la matemática del seguro, la frecuencia más probable es el promedio de las frecuencias.

Para determinar el tamaño de muestra se utilizará la siguiente fórmula:

$$n = \frac{z^2 \pi (1 - \pi)}{E^2}$$

Donde, z es el valor usado para el intervalo de confianza especificado, π es la estimación inicial de la proporción poblacional y E es el error de muestreo "de más o de menos" permitido en el intervalo.

z = 1.96 (para un intervalo del 95% de confianza)
$\pi = 0.03780$
E = 0.02617498

Sustituyendo estos valores se obtiene lo siguiente:

$$\mathbf{n = 203.6048 \approx 204}$$

Es decir que, empleando la frecuencia más probable como proporción se determinó el tamaño de muestra requerido para un intervalo de confianza del 95%, el cual fue de 204 carteras con 53 elementos cada una y el número esperado de siniestros (de las 204 carteras) será de 2.

iii.4.5.2 Generación de Carteras del Seguro de Responsabilidad Civil Agentes de Seguros (Personas Físicas) Utilizando Simulación Estocástica

Las 204 carteras (muestras) fueron generadas y una parte de una de éstas se presenta como ejemplo:[23]

Corrida : 10

	Número Aleatorio			Límite de	Volumen de Prima	Siniestro	
Dato	a	b	c	Responsabilidad	Promedio Intermediada	Sí=1; No=0	Monto
1	0.60701	0.28362	0.87833	$100,000.00	$1,232,846.78	0	$0.00
2	0.01593	0.49260	0.67002	$100,000.00	$21,202.37	0	$0.00
3	0.61458	0.74905	0.69245	$100,000.00	$1,258,500.62	0	$0.00
4	0.87556	0.36999	0.95287	$100,000.00	$2,750,841.19	0	$0.00
5	0.22355	0.93993	0.13336	$100,000.00	$333,984.90	0	$0.00
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
13	0.32089	0.90674	0.24424	$100,000.00	$510,812.80	0	$0.00
14	0.83043	0.85914	0.71354	$100,000.00	$2,342,363.40	0	$0.00
15	0.68382	0.42538	0.90787	$100,000.00	$1,519,921.82	0	$0.00
16	0.74583	0.09648	0.04296	$100,000.00	$1,808,095.98	0	$0.00
17	0.70071	0.66475	0.04963	$100,000.00	$1,592,365.31	0	$0.00
18	0.60689	0.00075	0.86659	$100,000.00	$1,232,427.17	1	$313,740.97
19	0.07316	0.63342	0.44348	$100,000.00	$100,284.62	0	$0.00
20	0.65819	0.79660	0.69581	$100,000.00	$1,417,032.29	0	$0.00
21	0.23762	0.31606	0.32421	$100,000.00	$358,123.59	0	$0.00
22	0.14671	0.77476	0.84020	$100,000.00	$209,419.53	0	$0.00
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
47	0.81860	0.91850	0.50501	$100,000.00	$2,253,323.60	0	$0.00
48	0.28000	0.56115	0.70373	$100,000.00	$433,622.41	0	$0.00
49	0.98254	0.71406	0.07092	$100,000.00	$5,342,789.70	0	$0.00
50	0.00521	0.67292	0.68969	$100,000.00	$6,891.30	0	$0.00
51	0.80741	0.84845	0.34504	$100,000.00	$2,174,278.30	0	$0.00
52	0.65123	0.01114	0.76253	$100,000.00	$1,390,398.55	1	$200,926.36
53	0.33116	0.98586	0.90498	$100,000.00	$530,911.49	0	$0.00
TOTAL				$5,300,000.00		2	$514,667.33

[23] El lector podrá ver las 204 carteras en el anexo electrónico.

iii.4.6 Interpretación de los Resultados[24]

Siguiendo los pasos anteriores se simularon 204 carteras tal y como se determinó en el subcapítulo iii.4.5.1; cada cartera consta de 53 elementos (pólizas) y con base a los datos generados se creó la siguiente tabla que proporciona un resumen de éstos:

Clase de Siniestros		Número Acumulado	Número Efectivos	Monto Efectivos	Monto Acumulado
$0	$10.000	399	71	$286.092,73	$57.401.254,00
$10.000	$20.000	328	36	$547.608,06	$57.115.161,27
$20.000	$30.000	292	24	$593.402,98	$56.567.553,21
$30.000	$40.000	268	20	$721.006,39	$55.974.150,23
$40.000	$50.000	248	24	$1.100.897,71	$55.253.143,83
$50.000	$60.000	224	17	$936.760,75	$54.152.246,12
$60.000	$70.000	207	12	$766.512,16	$53.215.485,36
$70.000	$80.000	195	9	$674.462,47	$52.448.973,21
$80.000	$90.000	186	9	$764.009,36	$51.774.510,74
$90.000	$100.000	177	7	$662.161,26	$51.010.501,38
$100.000	$200.000	170	62	$8.765.547,13	$50.348.340,12
$200.000	$300.000	108	43	$10.515.121,22	$41.582.793,00
$300.000	$400.000	65	29	$10.013.586,28	$31.067.671,78
$400.000	$500.000	36	11	$4.909.816,98	$21.054.085,49
$500.000	$600.000	25	14	$7.559.375,76	$16.144.268,51
$600.000	$700.000	11	5	$3.244.993,69	$8.584.892,75
$700.000	$800.000	6	3	$2.263.908,39	$5.339.899,06
$800.000	$900.000	3	1	$855.760,57	$3.075.990,68
$900.000	$1.000.000	2	0	$0,00	$2.220.230,10
$1.000.000	$1.100.000	2	1	$1.053.123,47	$2.220.230,10
$1.100.000	$1.200.000	1	1	$1.167.106,63	$1.167.106,63

Estos resultados serán utilizados para el cálculo de cuotas mismo que se explica en el siguiente punto.

[24] *Vid:* 4.4 Etapas de un Estudio de Simulación – 7.

iii.5 Procedimientos para Cálculo de Cuotas

Para el cálculo de cuotas se utilizarán las formulas que a continuación se presentan.

Monto en Primas Puras = P

Suma de Siniestros = $\sum_{i=1}^{m} s_i$

De donde,

$$P = \sum_{i=1}^{m} s_i \tag{1}$$

Por otro lado, la Probabilidad de Siniestro q se calculará con la siguiente formula:

$$q = \frac{\sum_{i=1}^{n} \alpha_i}{n} \tag{2}$$

Donde,

Grado del Daño[25]= α_i

[25] Una gran parte de los objetos asegurados afectados por siniestros sufre sólo una pérdida parcial, es decir no se destruye v (límite de responsabilidad o suma asegurada, dependiendo del seguro) en su totalidad, sino sólo una parte de v, la cual suele ser expresada, αv, en donde α es la fracción destruida de v; y es llamada *grado del daño*.

Sí $\alpha v = s$ es la fracción de la pérdida económica causada por un siniestro, tendremos casos donde $s = v$ y la relación entre el daño y el valor del objeto será $\frac{s}{v} = 1$, quedando así definida la pérdida total. En la mayoría de los casos será $s = 0$ y también $\frac{s}{v} = 0$. En los casos intermedios, $s < v$, por lo que $\frac{s}{v}$ será una expresión decimal entre 0 y 1, quedando así definida la pérdida parcial.

La Desviación Estándar del Gasto Total por Siniestros σ es:

$$\sigma = v\left(\sqrt{nq\left(\frac{\sum_{i=1}^{n}\alpha_i^2}{\sum_{i=1}^{n}\alpha_i} - q\right)}\right) \qquad (3)$$

Las Primas de Riesgo P_t se calculan de la siguiente forma:

$$P_t = P + \varphi\sigma \qquad (4)$$

En nuestro caso $\varphi = 1$.

La Cuota de Riesgo τ se obtiene como sigue:

$$\tau = \frac{P_t}{\sum_{i=1}^{n} v_i} \qquad (5)$$

Mientras que la Cuota de Tarifa τ_π se calcula de la siguiente forma:

$$\tau_\pi = \frac{\tau}{1-(\varepsilon_1 + \varepsilon_2 + \varepsilon_3 + \varepsilon_4)} \qquad (6)$$

Donde,

$\varepsilon_1 =$ Gastos de Administración

$\varepsilon_2 =$ Gastos de Adquisición

$\varepsilon_3 =$ Otros Gastos

$\varepsilon_4 =$ Utilidad

iii.5.1 Cálculo de Cuotas sin Deducible

Con base a la siguiente estadística se calculan P, P_t, τ y τ_π considerando que el valor total del riesgo es de $\sum_{i=1}^{n} v_i = 1{,}081'200{,}000.00$, donde v =100,000 y n = 10,812.[26]

Clase		$\sum_{i=1}^{m} m_i$	m_i	$\sum_{i=1}^{m} s_i$	Acumulados $\sum_{i=1}^{m} s_i$	$\alpha_i m$	$\alpha_i^2 m$
\$0	\$10.000	399	71	\$286.092,73	\$24.023.088,61	2,8609	0,1886
\$10.000	\$20.000	328	36	\$547.608,06	\$23.736.995,88	5,4761	0,8703
\$20.000	\$30.000	292	24	\$593.402,98	\$23.189.387,81	5,9340	1,4914
\$30.000	\$40.000	268	20	\$691.181,12	\$22.595.984,83	6,9118	2,4025
\$40.000	\$50.000	248	24	\$1.100.897,71	\$21.904.803,71	11,0090	5,0675
\$50.000	\$60.000	224	17	\$936.760,75	\$20.803.905,99	9,3676	5,1779
\$60.000	\$70.000	207	12	\$766.512,16	\$19.867.145,24	7,6651	4,9032
\$70.000	\$80.000	195	9	\$674.462,47	\$19.100.633,08	6,7446	5,0643
\$80.000	\$90.000	186	9	\$764.009,36	\$18.426.170,62	7,6401	6,4921
\$90.000	\$100.000	177	177	\$17.662.161,26	\$17.662.161,26	176,6216	176,2686
						240,2309	**207,9262**

De la tabla anterior y de las ecuaciones presentadas en iii.5 se obtiene:

$$P = \sum_{i=1}^{m} s_i = 24'023.088.61 \qquad q = \frac{\sum_{i=1}^{n} \alpha_i}{n} = \frac{240.2309}{10{,}812} = 0.0222 \qquad \frac{\sum_{i=1}^{n} \alpha_i^2}{\sum_{i=1}^{n} \alpha_i} = \frac{207.9262}{240.2309} = 0.8655$$

$$\sigma = v\left(\sqrt{nq\left(\frac{\sum_{i=1}^{n} \alpha_i^2}{\sum_{i=1}^{n} \alpha_i} - q\right)}\right) = 100{,}000\sqrt{(10{,}812)(0.0222)(0.8655 - 0.0222)} = 1'423{,}336.15$$

[26] Para elaborar ésta tabla, se consideró que en el caso de los siniestros mayores a \$100,000.00 sólo se paga ésta cantidad ya que el exceso queda a cargo del asegurado; en el caso de tener un siniestro menor a esta cantidad se paga el valor del siniestro. Lo anterior porque en el Ramo de Responsabilidad Civil los seguros se otorgan a primer riesgo.

"El seguro a *primer riesgo* es una modalidad que consiste en asegurar un bien por un valor determinado hasta el cual queda cubierto el riesgo".

$$P_1 = P + \sigma = 24'023{,}088.61 + 1'423{,}336.15 = 25'446{,}424.76$$

$$\tau = \frac{P_1}{\sum_{i=1}^{n} v_i} = \frac{25'446{,}424.76}{1{,}081'200{,}000.00} = 23.54\ ‰$$

Con base a la siguiente tabla se calcula la cuota de tarifa:

Gastos de Administración =	ε_1 =	15%
Gastos de Adquisición =	ε_2 =	9%
Utilidad Mínima Esperada =	ε_4 =	5%

$$\tau_\pi = \frac{\tau}{1-(\varepsilon_1+\varepsilon_2+\varepsilon_4)} = \frac{23.54}{1-(0.15+0.09+0.05)} = 33.15\ ‰$$

Finalmente, se construyó una escala de peligrosidad haciendo uso de las probabilidades encontradas en iii.4.2.3, obteniéndose la Tarifa siguiente:

Monto Promedio de Prima Intermediada	**Cuota Aplicable al Límite de Responsabilidad $100,000.00 M. N.**	**Prima en Pesos M. N.**
Hasta $1'500,000.00	19.23‰	$1,923.00
Mayor a $1'500,000.00	64.64‰	$6,464.00

Las cuotas obtenidas, se aplicaron a las carteras que previamente se simularon, los resultados de dicha aplicación causaron impacto positivo; y se concluye que permiten la sana operación de éste seguro.

Vid: VARGAS, *op. cit.* p. 140.

iii.5.2 Cálculo de Cuotas con Deducible

Con base a la siguiente estadística en la que se maneja un deducible de $7,500.00 se calculan P, P_t, τ y τ_π considerando que el valor total del riesgo es de $\sum_{i=1}^{n} v_i$ = 1,081'200,000.00, donde v =100,000 y n = 10,812.

Clase Después de Aplicar Deducible		$\sum_{i=1}^{m} m_i$	m_i	$\sum_{i=1}^{m} s_i$	Acumulados $\sum_{i=1}^{m} s_i$	$a_i m$	$a_i^2 m$
$0	$2.500	341	13	$16.985,62	$21.293.981,50	0,1836	0,0034
$2.500	$12.500	328	36	$277.608,06	$21.276.995,88	3,0012	0,2938
$12.500	$22.500	292	24	$413.402,98	$20.999.387,81	4,4692	0,8605
$22.500	$32.500	268	20	$541.181,12	$20.585.984,83	5,8506	1,7276
$32.500	$42.500	248	24	$920.897,71	$20.044.803,71	9,9557	4,1503
$42.500	$52.500	224	17	$809.260,75	$19.123.905,99	8,7488	4,5211
$52.500	$62.500	207	12	$676.512,16	$18.314.645,24	7,3136	4,4657
$62.500	$72.500	195	9	$606.962,47	$17.638.133,08	6,5618	4,7955
$72.500	$82.500	186	9	$696.509,36	$17.031.170,62	7,5298	6,3074
$82.500	$92.500	177	177	$16.334.661,26	$16.334.661,26	176,5909	176,2116
						230,2052	**203,3368**

De la tabla anterior y de las ecuaciones presentadas en iii.5 se obtiene:

$$P=\sum_{i=1}^{m} s_i = 21'293,981.50 \qquad q=\frac{\sum_{i=1}^{n}\alpha_i}{n}=\frac{230.2052}{10,812}=0.0213 \qquad \frac{\sum_{i=1}^{n}\alpha_i^2}{\sum_{i=1}^{n}\alpha_i}=\frac{203.3368}{230.2052}=0.8833$$

$$\sigma = v\left(\sqrt{nq\left(\frac{\sum_{i=1}^{n}\alpha_i^2}{\sum_{i=1}^{n}\alpha_i}-q\right)}\right)=100,000\sqrt{(10,812)(0.0213)(0.8833-0.0213)}=1'408,670.92$$

$$P_t = P+\sigma = 21'293,981.50+1'408,670.92=22'702,652.42$$

$$\tau=\frac{P_t}{\sum_{i=1}^{n} v_i}=\frac{22'702,652.42}{1,081'200,000.00}=21.00\ ‰$$

Con base a la siguiente tabla se calcula la cuota de tarifa:

$$\text{Gastos de Administración} = \varepsilon_1 = 15\%$$
$$\text{Gastos de Adquisición} = \varepsilon_2 = 9\%$$
$$\text{Utilidad Mínima Esperada} = \varepsilon_4 = 5\%$$

$$\tau_\pi = \frac{\tau}{1-(\varepsilon_1+\varepsilon_2+\varepsilon_4)} = \frac{21.00}{1-(0.15+0.09+0.05)} = 29.57\text{‰}$$

Finalmente, se construyó una escala de peligrosidad haciendo uso de las probabilidades encontradas en iii.4.2.3, obteniéndose la Tarifa siguiente:

Monto Promedio de Prima Intermediada	**Cuota Aplicable al Límite de Responsabilidad $100,000.00 M. N.**	**Prima en Pesos M. N.**
Hasta $1'500,000.00	17.10‰	$1,710.00
Mayor a $1'500,000.00	57.56‰	$5,756.00

Las cuotas obtenidas, se aplicaron a las carteras que previamente se simularon, los resultados de dicha aplicación causaron impacto positivo; y se concluye que permiten la sana operación de éste seguro.

La aplicación del deducible propuesto permite que exista un descuento de casi el 11% respecto a las cuotas manejadas sin deducible.

Los resultados de la metodología propuesta en este trabajo, resultaron ser satisfactorios para la problemática de este seguro, ya que permitirán a la compañía tener una tarifa justa y no tener problemas de insuficiencia de primas para el pago de siniestros; así como hacer frente a los gastos de administración y de adquisición; además la institución tendrá la posibilidad de obtener utilidad.

Con esta nota técnica se ejemplificó, en un caso real las ventajas del uso de la Estimación de la Distribución de Probabilidad para Muestras Pequeñas y de la Simulación en la Inferencia de Carteras de Seguros. Aportando así, un método que permite hacer inferencias de la cartera a pesar de que los datos de ésta no cumplan el supuesto de provenir de una muestra lo suficientemente grande.

Finalmente, en las siguientes páginas se procederá a exponer las conclusiones de la presente tesis.

ANEXO 1
DE NOTA TÉCNICA

CARTERA 1998
RESPONSABILIDAD CIVIL AGENTES DE SEGUROS
CIFRAS EN MONEDA NACIONAL

POLIZA	LIMITE DE RESPONSABILIDAD	SINIESTROS NUMERO	SINIESTROS MONTO
1	29,998,500	0	0
2	25,392,900	0	0
3	8,816,400	0	0
4	3,000,000	0	0
5	8,068,100	0	0
6	175,000	0	0
7	100,000	0	0
8	16,136,200	0	0
9	1,000,000	0	0
10	5,000,000	0	0
11	4,244,900	0	0
12	100,000	0	0
13	902,000	0	0
14	150,000	0	0
15	12,000,000	0	0
16	1,019,330	0	0
17	162,000	0	0
18	19,999,000	0	0
19	1,000,000	0	0
20	250,000	0	0
21	2,000,000	0	0
22	75,000	0	0
23	75,000	0	0
24	100,000	0	0
25	300,000	0	0
26	75,000	0	0
27	1,000,000	0	0
28	250,000	0	0
29	1,150,000	1	2,000
30	250,000	0	0
31	500,000	0	0
32	75,000	0	0
33	75,000	0	0
34	75,000	0	0
35	1,000,000	0	0
36	592,000	0	0
37	500,000	0	0
TOTAL	**145,606,330**	**1**	**2,000**

CARTERA 1999
RESPONSABILIDAD CIVIL AGENTES DE SEGUROS
CIFRAS EN MONEDA NACIONAL

POLIZA	LIMITE DE RESPONSABILIDAD	SINIESTROS NUMERO	SINIESTROS MONTO
1	28,296,000	0	0
2	28,332,300	0	0
3	9,432,000	0	0
4	9,687,200	0	0
5	100,000	0	0
6	418,000	0	0
7	420,000	0	0
8	100,000	0	0
9	5,000,000	0	0
10	4,622,350	0	0
11	2,000,000	0	0
12	1,366,292	0	0
13	1,000,000	0	0
14	150,000	0	0
15	11,000,000	0	0
16	9,385,000	0	0
17	934,830	0	0
18	1,000,000	0	0
19	162,000	0	0
20	19,250,000	1	427,846
21	420,000	0	0
22	2,000,000	0	0
23	1,000,000	0	0
24	175,000	0	0
25	75,000	0	0
26	200,000	0	0
27	1,000,000	0	0
28	1,000,000	0	0
29	3,000,000	0	0
30	75,000	0	0
31	3,000,000	0	0
32	1,150,000	0	0
33	250,000	0	0
34	75,000	0	0
35	1,000,000	0	0
36	125,000	0	0
37	1,000,000	0	0
38	1,481,246	1	10,250
39	500,000	0	0
TOTAL	**150,182,218**	**2**	**438,096**

CARTERA 2000
RESPONSABILIDAD CIVIL AGENTES DE SEGUROS
CIFRAS EN MONEDA NACIONAL

POLIZA	LIMITE DE RESPONSABILIDAD	SINIESTROS NUMERO	SINIESTROS MONTO
1	100,000	0	0
2	9,498,600	0	0
3	28,246,500	0	0
4	9,415,500	0	0
5	9,511,000	0	0
6	3,000,000	0	0
7	9,511,000	0	0
8	49,121,500	0	0
9	3,000,000	0	0
10	100,000	0	0
11	100,000	0	0
12	100,000	0	0
13	500,000	0	0
14	5,000,000	0	0
15	1,000,000	0	0
16	500,000	0	0
17	4,706,350	0	0
18	2,500,000	1	30,405
19	500,000	0	0
20	1,000,000	0	0
21	150,000	0	0
22	5,000,000	0	0
23	2,000,000	1	238,501
24	200,000	0	0
25	944,590	0	0
26	500,000	0	0
27	19,128,400	0	0
28	162,000	0	0
29	500,000	0	0
30	420,000	0	0
31	175,000	0	0
32	1,000,000	0	0
33	2,000,000	0	0
34	500,000	0	0
35	300,000	0	0
36	500,000	0	0
37	1,000,000	0	0
38	75,000	0	0
39	500,000	0	0
40	500,000	0	0
41	250,000	0	0
42	1,370,000	0	0
43	75,000	0	0
44	100,000	0	0
45	1,150,000	0	0
46	75,000	0	0
47	5,000,000	0	0
48	1,000,000	0	0
49	200,000	0	0
50	200,000	0	0
51	200,000	0	0
52	125,000	0	0
53	1,000,000	0	0
54	1,766,586	0	0
55	500,000	0	0
56	250,000	0	0
57	500,000	0	0
TOTAL	**186,727,026**	**2**	**268,906**

ANEXO 2

DE NOTA TÉCNICA

ESTADÍSTICA 1999 - 2002
Cifras en Moneda Nacional

Volumen de Prima Promedio Intermediada	Número de Agentes
40,781	1,253
85,088	842
116,947	681
164,614	493
212,054	417
260,076	356
305,364	250
334,692	215
406,709	182
424,805	153
444,697	144
538,323	110
618,191	104
628,205	76
623,741	80
829,021	39
784,463	47
791,900	56
716,144	40
1,004,473	42
965,833	30
929,940	27
1,102,811	31
1,280,641	30
962,514	27
1,125,519	20
1,070,391	19
1,411,220	20
1,229,467	16
1,170,140	12
1,715,853	10
1,346,394	11
1,284,556	9
1,651,606	10
1,778,352	8
1,191,824	5
1,461,960	11
1,636,713	6
1,657,805	7
1,877,224	7
2,573,612	5
2,188,574	5
2,768,045	5
2,152,758	9
3,211,147	3

ESTADÍSTICA 1999 - 2002
Cifras en Moneda Nacional

	Volumen de Prima Promedio Intermediada	Número de Agentes
	2,743,185	5
	2,748,725	3
	1,806,382	3
	1,922,538	3
	2,131,275	1
	1,803,425	1
	2,336,643	4
	2,265,370	2
	1,214,540	1
	1,732,305	1
	2,994,966	5
	3,458,190	2
	4,101,368	3
	3,152,309	5
	6,719,605	1
	2,269,985	3
	4,023,990	1
	5,321,472	3
	2,548,482	3
	3,803,252	3
	1,454,335	1
	3,384,760	1
	7,333,795	1
	3,574,620	1
	2,012,715	1
	3,013,220	1
	10,868,885	1
	1,465,768	2
	14,245,380	1
	1,624,933	2
	6,149,430	1
	3,236,355	1
	3,102,965	1
	5,378,095	1
	8,836,350	1
	6,981,670	1
	7,949,690	1
	12,755,430	1
	9,510,265	1
Total	221,051,850	5,997

ANEXO ELECTRÓNICO

CONCLUSIONES Y RECOMENDACIONES

En el presente trabajo se estudió de manera previa el tema de estadística descriptiva ya que el paso inicial de toda investigación es coleccionar, agrupar y analizar los datos con los que se desea trabajar o simplemente con los que se trabajará.

Una vez hecho lo anterior, el siguiente paso consiste en hacer inferencias sobre el comportamiento de los datos; motivo por el cual esta tesis dedicó el segundo capítulo al tema de distribuciones de probabilidad, ya que a través de estas se pueden hacer conjeturas sobre los datos cuando éstos forman un conjunto suficientemente grande.

Ante el problema de trabajar con pocos datos, situación común en el campo del seguro, en este trabajo se explicó el método del rango mediano ya que este permite estimar una distribución de probabilidad a pesar de que los datos no sean de una muestra suficientemente grande; además se mostraron dos pruebas útiles para acreditar que los datos se ajustan a la distribución estimada.

Así mismo, para calcular el número de veces que se trabajará con el modelo simulado se mostró como determinar el tamaño de muestra.

También, se habló de la simulación y para entenderla mejor se estudió el modelado de sistemas porque con esta técnica se simula realmente un modelo del sistema.

Finalmente, la metodología expuesta por esta tesis se ejemplifico con una nota técnica en la que se calculan las cuotas base para el seguro de responsabilidad civil agentes de seguros, producto que carecía de homogeneidad respecto a los precios calculados por diferentes compañías aseguradoras y consecuentemente se observaba la carencia de técnica para extraer el mayor provecho como sea posible a pocos datos.

Como dato adicional, cuando se calcularon las cuotas para seguro del que se habla en el capítulo cinco, se contaba con la tarifa que el reasegurador entregó a la compañía que proporcionó los datos, los costos para diferentes límites de responsabilidad publicados por otra compañía de seguros (información de dominio público) y con los pocos datos de la cartera proporcionada por la compañía.

Mediante el uso de la metodología aquí propuesta, se observó que la tarifa entregada por el reasegurador y la publicada por la otra compañía llevan a números rojos, mientras que la que se desarrolla en este trabajo con el uso de las técnicas propuestas proporciona las consecuencias positivas implícitas de un buen cálculo, hacer frente a los siniestros, utilidad para la empresa y el pago de los gastos de administración y adquisición, observándose que la construcción del modelo y los supuestos del mismo son buenos.

Se recomienda a las personas que deseen emplear esta técnica tomen en cuenta lo siguiente:

1. La construcción del modelo juega un papel importante en los resultados que se obtengan por el uso de éstas técnicas.
2. Esta metodología sólo puede emplearse por cada riesgo del que se tenga información, aunque ésta sea escasa.
3. Está metodología sólo puede aplicarse a cada grupo o clase de riesgos de los que se tenga información.
4. Siempre será mejor trabajar con una muestra lo suficientemente grande, claro que ante la carencia de datos esta metodología es una buena opción.
5. Independientemente de la aplicación final de los datos obtenidos con estas técnicas, tal aplicación debe de ser supervisada una vez puesta en práctica para evitar desvíos respecto a lo planeado y en caso contrario nuevamente estudiar el riesgo para diseñar programas correctivos; esta recomendación constituye un paso obligado independientemente de las técnicas utilizadas.

La realidad, es que el uso de nuevas técnicas en campos especializados y con innovadoras exigencias, –como lo es el del seguro– permite la consolidación de vínculos entre estos, cultivando lazos fructíferos en este manantial inextinguible de siempre nuevos conocimientos por aprender.

NOTA FINAL

La compañía de seguros que proporcionó su información para el cálculo de la tarifa del Seguro de Responsabilidad Civil Agentes de Seguros (Persona Físicas), determinaba las primas para este riesgo con una tarifa proporcionada por el resaegurador, el método de cotización que éste propone se conoce como *pay back* además de un descuento en el caso de deducible, el pay back se da en un intervalo permisible que indica el número de pólizas que se deben de vender para hacer frente a un siniestro. A los elementos de las carteras simuladas se les aplicó esta metodología para hacer comparación con el propuesto por esta tesis, el resultado fue que el pay back debería de ser menor; lo que implica que las primas deberían de ser mayores; además de que haciendo uso del deducible no se podía dar el descuento propuesto, por lo tanto hacer uso de la tarifa propuesta por el reasegurador para este riesgo llevará a números rojos a la compañía de seguros.

La tarifa publicada por una empresa competidora es más baja que la propuesta por esta tesis, lo que implica que no permite pagar los siniestros que se presenten, además tampoco permite hacer frente a los gastos de administración, los de adquisición y la compañía de seguros no obtendrá utilidad, por lo tanto las primas de la competencia llevaran a números rojos a la empresa competidora.

Finalmente, la tarifa calculada con la metodología propuesta por esta tesis garantiza la solvencia del asegurador, ya que el uso de ésta permite hacer frente a todos los compromisos que se adquieren con la venta de seguros para este riesgo.

BIBLIOGRAFÍA

AGUILAR Juárez, Isabel Patricia y Leonardo BAÑUELOS Saucedo
Notas del Curso Propedéutico de Probabilidad y Estadística.
México. Trabajo No publicado. 2002. 97 p.

BILLINTON, Roy y Ronald N. ALLAN
Reliability Evaluation of Engineering Systems Concepts and Techniques. 2nd Edition.
United States of America. Plenum Publishing Corporation. 1992. 453 p.

CANAVOS C. George
Probabilidad y Estadística Aplicaciones y Métodos. 1a. Edición.
[Tr: Edmundo Gerardo Urbina Medal]
[R. T. Gustavo Javier Valencia Ramírez]
México. Mc Graw Hill. 1988. 651 p.

COMISIÓN NACIONAL DE SEGUROS Y FIANZAS (CNSF)
Circular S – 8.1
México. Diario Oficial de la Federación.
Febrero 20, 2004.

COSS Bu, Raúl
Simulación un Enfoque Práctico. 1a Edición.
México. Editorial Limusa. 1988. 158 p.

ENCICLOPEDIA CIENTÍFICA CULTURAL
Volumen estadística.
[Tx: Juan Luis Gutiérrez Duchos]
España. Cultural S. A. de Ediciones. 1980. 243 p.

HILLER, Frederick S. y Gerald J. LIEBERMAN
Introducción a la Investigación de Operaciones. 4a. Edición
[Tr. Marcia A. González Osuna]
[R. T. José Humberto Cantú Delgado
Perla J. Fernández Reyna
Marco A. Montufar Benítez]
México. Mc Graw Hill – Interamericana Editores. 1999. 998 p.

KAZMIER, Leonard J.
Estadística Aplicada a la Administración y a la Economía. 3a. Edición.
[R. T. Alejandro Alegría Hernández]
México. Mc Graw Hill. 1998. 416 p.

KNEZEVIC, Jezdimir
Reliability, Maintainability and Supportability a Probabilistic Approach
United Kingdom. Mc. Graw Hill Book Company Europe. 1993. 291 p.

KREYSZIG, Erwin
Introducción a la Estadística Aplicada Principios y Métodos. 1a. Edición.
[Versión española: Arturo Galán Martínez]
[R. Octavio A. Rascon Chávez]
México. Editorial Limusa. 1979. 505 p.

MARÍN Diazaraque, Juan Miguel
Apuntes de Estadística – Estadística Descriptiva, p. 333
España. Universidad Carlos III de Madrid. 2004.
http://halweb.uc3m.es/

MILLER, Irwin, John E. FREUD y Richard A. JOHNSON
Probabilidad y Estadística para Ingenieros. 4a. Edición.
[Tr: Virgilio González Pozo]
[R. T. Graciela Rojas de Finck]
México. Prentice – Hall Hispanoamericana. 1992. 624 p.

RINCON, Juana
Cooperación del Personal Académico: Mecanismo para la Integración del Sistema Universitario Nacional.
San Fernando de Apure, Venezuela. Universidad Simón Rodríguez. 1998.
http://members.tripod.com/~gepsea/sistema.htm

SAUTTO Vallejo, José Maclovio
La técnica de la Simulación Digital en la Investigación de Operaciones.
México. Tesis UNAM. 1991. 146 p.

SECRETARÍA DE HACIENDA Y CRÉDITO PÚBLICO
Reglas que Establecen las Orientaciones de Política General Aplicables a los Agentes y Apoderados de Seguros y Fianzas.
México. Diario Oficial de la Federación.
Febrero 27, 2004.

VARGAS Aguilar, Juan Carlos
Fundamentos para el Desarrollo de Productos en Daños
México. Tesis UNAM. 2003. 164 p.

Ley General de Instituciones y Sociedades Mutualistas de Seguros.
Archivo de la CNSF 2001.
México. Diario Oficial de la Federación.
Enero 16, 2002.