



DOCENTES DE ADMINISTRACIÓN FINANCIERA

**XXXVI Jornadas Nacionales de Administración Financiera
Septiembre 2016**

ANÁLISIS DETERMINÍSTICO VERSUS SIMULACIÓN PROBABILÍSTICA

Gabriel R. Feldman

Universidad Nacional de Tucumán

SUMARIO: 1. Introducción; 2. Marco conceptual; 3. Desarrollo práctico; 4. Conclusión.

Para comentarios: gfeldman@herrera.unt.edu.ar

Resumen

El análisis y toma de decisiones a través de métodos cuantitativos se basa en la elaboración de modelos matemáticos, que constituyen la base para la medición de riesgos. La simulación es una de las herramientas más valiosas y de mayor uso a la hora de la toma de decisiones estratégicas y la resolución de problemas. En el artículo se demuestra que la simulación Monte Carlo constituye una herramienta que aporta elementos de mayor robustez estadística que el análisis de sensibilidad y el análisis de escenarios.

1. Introducción

Un propósito básico de la administración financiera es contribuir a la toma de decisiones, y en general cuando se habla de toma de decisiones desde un punto de vista cuantitativo, hay una cuestión que está siempre involucrada: los modelos financieros. Estos constituyen los pilares de la toma de decisiones cuantitativas ya que son los elementos numéricos que permiten elaborar un esquema de análisis.

Una primera alternativa de análisis es la de estimación de punto único, es decir determinar un valor esperado para entradas específicas. Su ventaja es que es rápido y fácil, pero paralelamente su punto débil es que es solo aproximado y generalizado.

La segunda alternativa que se maneja para analizar valores futuros es el análisis de escenarios, es decir definir una posición optimista, pesimista y neutral. Si bien su ventaja es que pre-

sesta alternativas de respuestas, constituye solo 3 combinaciones posibles, no pudiéndose hacer inferencia en base a ello.

Una tercera manera de enfocar el análisis de datos para la toma de decisiones es el análisis de sensibilidad o análisis “y si”. Muestra rangos de posibilidades incrementales y marginales, permitiendo observar cual es el impacto marginal de una variable sobre otra. Pero tampoco constituye una posibilidad para realizar inferencia estadística.

La alternativa restante es la simulación. Es un concepto sencillo de implementar, pero requiere un software especializado. Las simulaciones nacen en la época de la segunda guerra mundial, en que se buscaba herramientas para poner a prueba el armamento, sin tener que utilizarlo en la práctica. En aquella oportunidad, en virtud de una necesidad bélica en se quería probar el efecto de la bomba atómica, un elemento costoso, complicado, con consecuencias reales desconocidas e impredecibles, que si bien se sabía que iba a ser un elemento trágico, se desconocía su real capacidad de destrucción.

Paralelamente con el desarrollo de las herramientas computacionales, se potencia el uso de la técnica de simulación Montecarlo, como elemento posible de implementar. Básicamente consiste en la generación de múltiples escenarios aleatorios¹ (múltiples significa números tan grandes como 1.000, o 10.000 o 100.000...), de forma tal que se van a obtener “n” resultados, los cuales son susceptibles de ser involucrados en ejercicios de inferencia estadística. Es decir, permite tomar decisiones sobre una población, a partir de una muestra, que es un subconjunto representativo de la población.

Se entiende así entonces por qué se utilizó este tipo de herramienta para evaluar el posible efecto de una bomba que es cara y que iba a tener efectos destructivos masivos. La base de una pequeña muestra, sirve de cimiento para soportar múltiples escenarios que pueden servir para saber qué debería esperarse sobre el objeto que se analiza. Ese “objeto” puede ser un proyecto, un nuevo negocio, el lanzamiento de un producto, el precio estimado para el futuro, la variación del tipo de cambio para el próximo mes, etc.

Es evidente que detrás del concepto de simulación está inserto el concepto de incertidumbre, es decir todo lo que se desconoce que va a ocurrir en el futuro. Toda decisión que implique datos de un contexto futuro, está sujeta a incertidumbre. En realidad de esta incertidumbre la parte que más interesa estudiar al campo de las finanzas es cuando los resultados salen mal, es decir cuando la incertidumbre puede afectar negativamente a la variable bajo estudio, y es la parte que se conoce como “riesgo”. Normalmente la parte de la incertidumbre en que los resultados futuros salen bien, se conoce como “oportunidades”. Es así que en general en finanzas se estudia la gestión del riesgo.

La simulación es un método no determinístico, ya que se fundamenta sobre la incertidumbre. En un escenario determinístico puede saberse *ex ante* qué es lo que va a ocurrir. En cambio la simulación se basa en números aleatorios, y es un concepto de aproximación a la realidad (como todos los modelos), que sirve de base para la toma de decisiones cuantitativas.

El propósito de este artículo es demostrar que la utilización de herramientas informatizadas científicas diseñadas para apoyar la toma de decisiones financieras, constituye un elemento que agrega valor a la tarea de análisis con modelos matemáticos. Como un subproducto, tener que construir un modelo preciso permite que quien hace el pronóstico amplíe y profundice la comprensión del proyecto.

¹ Se destaca el concepto de aleatoriedad porque justamente el nombre de la técnica proviene del nombre de la ciudad de Montecarlo, caracterizada por los juegos de azar. Al ser un ejercicio aleatorio y repetitivo, ello se asemeja al tipo de suerte que se puede tener al jugar en casinos.

2. Marco Conceptual

El proceso de análisis estadístico sintéticamente se divide en 4 fases:

- 1) Recolección de la información
- 2) Organización de la información
- 3) Análisis
- 4) Interpretación de los resultados

La población se refiere al conjunto de todas las variables de interés para el investigador. La muestra es un subconjunto representativo de dicha población.

Las variables pueden ser cuantitativas o cualitativas. Las primeras pueden ser discretas o continuas, y están expresadas en magnitudes. En cambio, las variables cualitativas están expresadas en categorías, y pueden ser de carácter:

- Nominal (por ejemplo el género), en las que el orden no interesa;
- Ordinal (por ejemplo: el extracto socio-económico, que puede ordenarse por un orden jerárquico, numéricamente).

Son variables continuas cuando entre un número A y un número B hay infinitos valores (generalmente representada por números decimales). Las variables son discretas cuando entre un número A y un número B hay finitos valores, generalmente representada por números enteros (positivos o negativos).

El analista no tiene la posibilidad de estudiar toda la población, ya que los limitantes son el tiempo y los recursos. Por ello se obtiene y analiza una muestra representativa de esta población. De la población se estudian los parámetros, que son medidas descriptivas de la variable en la población. De la muestra se estudian los estadísticos, que son medidas descriptivas de la variable en la muestra, y permiten estimar los parámetros poblacionales desconocidos.

Los parámetros más destacados son la media poblacional, la varianza poblacional, y la desviación estándar poblacional. Como no se cuenta con la población completa, se estudia la muestra y sus estadísticos más conocidos son: media muestral, varianza muestral y desviación muestral.

	<i>Poblacional</i>	<i>Muestral</i>
Media	μ	\bar{X}
Varianza	σ^2	S^2
Desviación Estándar	σ	S

Un experimento aleatorio es toda acción definida donde se conocen todos los resultados posibles, pero existe incertidumbre sobre la ocurrencia del resultado real. Cada evento o suceso es uno o más de los posibles resultados. Se denomina espacio muestral al conjunto de todos los resultados posibles de un experimento aleatorio. Está formado por “n” puntos muestrales, cada uno de los cuales tiene su probabilidad de ocurrencia.

Una variable aleatoria X es una función cuyos valores son números reales y dependen de una distribución de probabilidad. Una distribución de probabilidad describe el rango de valores que puede tomar una variable aleatoria y la probabilidad asignada a cada valor o rango de valores.

Dadas las características de las variables que se va a estudiar al hacer simulación, deben tenerse presente 2 conceptos que representan el eje central de la teoría de probabilidades:

- Ley de los grandes números: cuanto mayor sea el tamaño de la muestra, mayor será el ajuste entre la distribución teórica sobre la que se basa la muestra. La frecuencia relativa de los resultados de un cierto experimento aleatorio, tiende a estabilizarse en cier-

to número, que es precisamente la probabilidad, cuando el experimento se realiza muchas veces. Es decir, cuanto más se incrementa el tamaño de la muestra, más nos acercaremos a una probabilidad real. Cuando se incrementa el tamaño de la muestra, más nos a pareceremos a la población.

- Teorema central del límite: la media muestral de un conjunto de n variables muestradas en forma independiente a partir de una distribución $f(x)$ se ajusta a una distribución aproximada normal. En otras palabras, la distribución del promedio de un conjunto de variables aleatorias depende tanto de la cantidad de variables aleatorias promediadas como de la incertidumbre aportada por cada variable. Es decir que, independientemente de cómo se distribuya la población, si sacamos muestras repetidas y sacamos la media de cada muestra, y organizamos esa media de cada muestra (es decir formamos una distribución de medias), se va a comportar como una distribución Normal.

Hay 4 medidas que son muy importantes, y que se conocen como los 4 momentos de una distribución:

1) Momento 1: medidas de tendencia central. Es un valor representativo de un conjunto de datos. Se refiere al punto medio de una distribución. Las medidas de tendencia central se conocen también como medidas de posición.

a) La principal medida de tendencia central es la media aritmética, que es el valor esperado. Es el resultado que se esperaría, en promedio, si se tiene la posibilidad de ejecutar muchas veces el modelo. Se obtiene sumando todas las observaciones y dividiendo por el tamaño de la muestra.

También puede calcularse la media aritmética ponderada, teniendo en cuenta el peso asignado a cada una de las observaciones, pudiendo estimarse de manera relativa o absoluta. Permite calcular un promedio que toma en cuenta la importancia de cada valor respecto del total. El claro ejemplo se da en los portafolios de inversión, en que se asigna un porcentaje a cada activo.

A modo de síntesis, la media aritmética tiene las siguientes ventajas y desventajas:

Ventajas	Desventajas
1) Concepto familiar para la mayoría de las personas y es intuitivamente claro	1) Se puede ver afectada por valores extremos que no suelen ser representativos
2) Cada conjunto de datos tiene una media, es una medida que puede calcularse y es única debido a que cada conjunto de datos posee una y solo una media	2) Tedioso cuando la muestra es amplia
3) La media es útil para llevar procedimientos estadísticos como la comparación de medias de varios conjuntos de datos	

b) Mediana: es el valor central de un conjunto de datos ordenados. Es el elemento que está más al centro de este conjunto de números.

$$\mu_e = \frac{n+1}{2}; \text{ en caso de que sea par, se promedia la posición}$$

También se conoce como el cuartil 2, el percentil 50, y el decil 5.

Se utiliza la mediana cuando se quiere representar el valor esperado, pero se encuentran valores atípicos. Independientemente de los valores atípicos (positivos o negativos) siempre la mediana va a ser la misma.

c) Modo. Es la observación que se presenta con mayor frecuencia. Sus características son:

Ventajas	Desventajas
1) Datos cualitativos como cuantitativos	1) No se usa de manera frecuente
2) Los valores extremos no afectan independientemente a la moda	2) En ocasiones no existen modas
3) No importa la amplitud o que tan pequeños sean los valores del conjunto de datos e independientemente de cuál sea su dispersión.	3) Cuando existen más de dos modas es fácil interpretar y comparar la medida de tendencia central.

d) Media geométrica, que se va a encargar de obtener tasas de crecimiento promedio, cuando se trabaja con cantidades que cambian en cierto período y se necesita conocer una tasa promedio de cambio. Ejemplo: tasa promedio de inflación o tasa de interés promedio que obtiene una cartera.

2) Momento 2: Medidas de variabilidad o dispersión. Se refiere a la separación de los datos en una distribución, teniendo como criterio una medida de tendencia central (la medida de tendencia central a utilizar es la media).

Es importante medir la variabilidad para:

- Medir el control de la calidad estadística
- Proporciona información adicional para juzgar la confiabilidad de la medida de tendencia central
- Reconocer la dispersión de los datos
- Comparar la dispersión de diferentes muestras

El objetivo de medir la variabilidad o dispersión, es para ser utilizada como medida de riesgo; mientras que las medidas de tendencia central son indicadores de la rentabilidad.

Medidas de variabilidad:

a) Rango: es la diferencia ente el más alto valor y el más pequeño del conjunto de datos observados.

b) Rango interfractil: es una medida de dispersión entre dos fractiles de una distribución de frecuencia. Los fractiles tienen nombre dependiendo del número de partes iguales en que se dividen los datos (deciles, percentiles, cuartiles)

c) Rango intercuartil: la diferencia entre el cuartil 3² y el cuartil 1, mide la dispersión que hay en el 50% central de los datos.

d) Desviación estándar (o varianza)³: La desviación estándar indica lo que en promedio se desvía cada observación respecto de la medida de tendencia central (media)

e) Coeficiente de variación: Su fórmula expresa la desviación estándar como porcentaje de la media aritmética, mostrando una mejor interpretación porcentual del grado de variabilidad que la desviación estándar.

² La interpretación de cuartil 3 es que el conjunto de datos hasta allí representan el 75% del total.

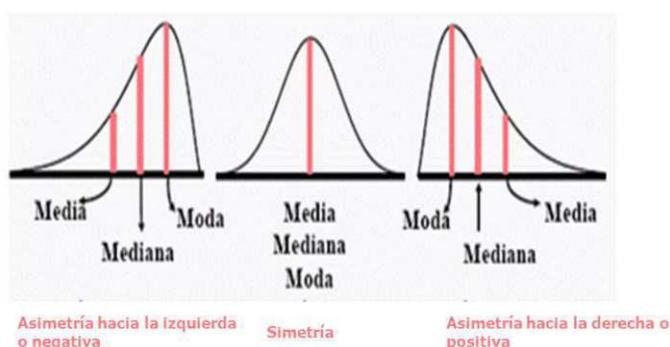
³ Está fuera del propósito del artículo expresar las particularidades y diferencias entre ambos conceptos.

Una importante regla empírica para interpretar la desviación estándar en las distribuciones de probabilidad que se distribuyen como normal es:

- Alrededor del 68% de las observaciones se encuentran en el intervalo $\mu \pm \sigma$
- Alrededor del 95% de las observaciones se encuentran en el intervalo $\mu \pm 2\sigma$
- Alrededor del 99% de las observaciones se encuentran en el intervalo $\mu \pm 3\sigma$

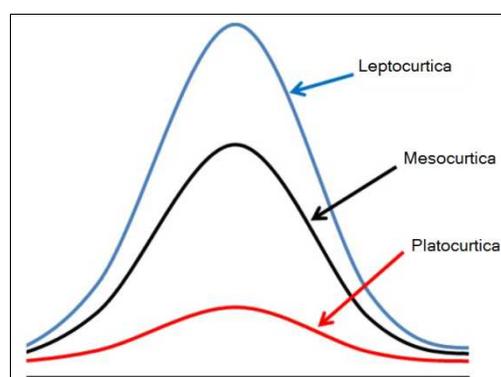
Esto ayudará a construir intervalos de confianza en la simulación, e interpretar la desviación estándar de sus observaciones.

3) Momento 3: asimetría. Las curvas que representan los datos puntuales de un conjunto de datos pueden ser simétricas o sesgadas. Indica hacia donde se cargan los resultados, hacia la derecha o la izquierda. Sería igual a cero para una distribución que sea totalmente simétrica, es decir que tenga las mismas posibilidades que ocurran los valores hacia la izquierda o la derecha de la media.



Si se estuviera estudiando una variable de ingresos, convendría la distribución con asimetría negativa, ya que brinda la posibilidad de ganar más. Tiene valores que se concentra con mayor frecuencia en valores altos. Por el contrario, si se estuviera estudiando una distribución de riesgos (por ejemplo catástrofes, siniestros) sería óptimo estar frente a una distribución de asimetría positiva, ya que concentra mayores frecuencias en valores bajos. La asimetría se mide a través del coeficiente de asimetría⁴. Este coeficiente es igual a cero para la distribución normal.

4) Momento 4: curtosis. Indica el nivel de apuntalamiento, y se utiliza para determinar la probabilidad de eventos extremos o catastróficos⁵.



⁴ En Excel se utiliza la función COEFICIENTE.ASIMETRIA

⁵ En Excel se utiliza la función CURTOSIS

En la distribución Leptocúrtica la curtosis es mayor que 3 (o el exceso de curtosis es mayor que cero), en la Mesocúrtica la curtosis es igual a 3 (o el exceso de curtosis es igual a cero) y en la Platocúrtica la curtosis es menor que 3 (o el exceso de curtosis es menor que cero).⁶

De estas distribuciones, la que sería preferible para distribución de los eventos de una empresa es la platocúrtica, ya que tienes “colas delgadas” lo que implica que hay pocas probabilidades de tener eventos en los extremos.⁷

En síntesis, al efectuar un análisis integral descriptivo debe basarse en estos 4 momentos de una distribución de probabilidad: i) Medida de tendencia central: indica rentabilidad; ii) Medida de dispersión: indica riesgo; iii) Coeficiente de asimetría: indica el sesgo; iv) Curtosis: indica cuál es la probabilidad de tener eventos extremos, o catastróficos.

3. Desarrollo práctico

El desafío para el analista es convertir un problema o situación de la vida real, que normalmente es de naturaleza descriptiva, en números.

El siguiente ejemplo base (caso A) muestra la determinación de la utilidad de una inversión, en que los valores actuales⁸ de ambos flujos de fondos son los siguientes:

	B	C	D
2		A	
3			
4	Ingresos	1000	
5	(-) Egresos	800	
6	Utilidad (VAN)	200	<-- =C4-C5

A priori es un proyecto conveniente ya que la utilidad (VAN) es positiva. El problema es que en cualquier actividad empresarial, los flujos de fondos futuros están sujetos a incertidumbre. Es decir, en este caso el modelo matemático es sencillo y claro, reportando un resultado positivo. No hay posibilidad de perder ni de ganar una suma diferente de \$200.

Aun así, lo que incomoda al analista es cuál es el argumento para asegurar que la realidad se va a comportar de dicha manera, es decir cuál es la seguridad de que el valor actual de los ingresos futuros sea de \$1.000 y de los egresos futuros de \$800. No hay forma de garantizarlo ya que todo contexto futuro está sujeto a incertidumbre.

Otra manera de plantearlo (caso B) es considerar que hay escenarios probables en el momento futuro, en el que se asume que los ingresos y egresos se comportan de manera incierta, tal como se describe en el siguiente cuadro:

	E	F	G	H
2		B		
3		Pesimista	Más Probable	Optimista
4	Ingresos	800	1000	1500
5	(-) Egresos	850	800	750
6	Utilidad (VAN)	-50	200	750

⁶ Se sustrae 3, que es la curtosis de la Normal, con objeto de generar un coeficiente que valga 0 para la Normal y tome a ésta como referencia de apuntamiento.

⁷ Tener en cuenta que en el eje Y se grafica la probabilidad, y en el eje X el valor de la variable.

⁸ El mismo análisis podría hacerse considerando que los flujos de fondos, tanto ingresos como egresos, ocurren en el momento cero, caso en el cual ya no haría falta mencionar que se trata de valores actuales.

Si bien este esquema incorpora variabilidad para el futuro, contempla solo una combinación de valores posibles, y la realidad de cualquier actividad empresarial es muy distante de ser así. En este ejemplo se planteó que en el escenario optimista, los ingresos aumentan y los costos disminuyen, mientras que en escenario pesimista los ingresos disminuyen y los costos aumentan. Pero en un escenario optimista también podría esperarse que los ingresos aumenten y que los costos se mantengan constantes, o cualquier otra combinación de las variables. Es decir que la generación de solo 3 escenarios es insuficiente para abarcar todas las posibilidades de interacciones/combinaciones entre las variables.

Considerando que existe variabilidad en los ingresos y egresos futuros, una tercera manera (caso C) de evaluar el contexto futuro es utilizar una herramienta de generación de números aleatorios de Excel. La función ALEATORIO.ENTRE permite generar números aleatorios dentro de un rango de valores, que en este caso serían los valores indicados para el escenario pesimista y optimista. Estos valores extremos deberían basarse en una justificación importante, en algún criterio razonable, una justificación que viene de la teoría, o de la experiencia, como puede ser la distribución histórica de los valores en la empresa.

Estadísticamente la generación de números aleatorios supone que la distribución proviene de una distribución uniforme, siendo que todos los valores tienen igual probabilidad de ocurrencia⁹.

	B	C	D
9		<u>C</u>	
10			
11	Ingresos	1121	<-- =ALEATORIO.ENTRE(800;1500)
12	(-) Egresos	800	
13	Utilidad (VAN)	321	

Es decir que de este modo se tiene la posibilidad de generar múltiples escenarios (tantas veces como se presione la tecla F9 para generar el nuevo ingreso aleatorio), no solo 3 como en el caso anterior.

Del mismo modo puede programarse que los egresos sean generados aleatoriamente, dentro del rango previamente definido, y utilizando la misma función que para los ingresos:

	B	C	D
9		<u>C</u>	
10			
11	Ingresos	1101	<-- =ALEATORIO.ENTRE(800;1500)
12	(-) Egresos	758	<-- =ALEATORIO.ENTRE(750;850)
13	Utilidad (VAN)	343	

Lógicamente ahora al presionar F9 se genera una nueva combinación de ingresos y egresos aleatorios, pudiendo simularse n escenarios. Por lo tanto el ejercicio consiste en ir acumulando cada uno de los resultados de la variable de interés (la utilidad en este caso), y hacer inferencia estadística sobre ello¹⁰.

⁹ En Excel al presionar F9 se genera un nuevo número aleatorio.

¹⁰ Una nota sobre la generación de números aleatorios en simuladores: se habla de números pseudo-aleatorios, si son generados a partir de un valor que se denomina semilla. Ello es recomendable solo si se quiere obtener el mismo resultado al replicar el ejercicio en otro momento (a fines académicos por ejemplo). Pero para correr la simulación (a fines empresariales) es recomendable trabajar sin valor semilla para que haya verdadera aleatoriedad. Por otra parte los números aleatorios responden a alguna

Para ejemplificar se calculan a continuación 20 escenarios, con las utilidades esperadas, para luego recurrir a las medidas estadísticas que resumen información:

	E	F
10	Escenario	Utilidad
11	1	201
12	2	251
13	3	358
14	4	635
15	5	84
16	6	564
17	7	450
18	8	38
19	9	158
20	10	290
21	11	676
22	12	669
23	13	59
24	14	346
25	15	571
26	16	205
27	17	374
28	18	341
29	19	56
30	20	284

Si bien en los 20 escenarios generados aleatoriamente no se observó una situación de pérdida, ello no implica que no exista esa posibilidad, solo que evidentemente es poco probable.

Las medidas que resumen información son las estadísticas descriptivas: media y desviación estándar. La primera indica cual es el valor esperado de los valores, y la segunda cual es el grado de dispersión en promedio.

	G	H	I
11	Media	330,50	<-- =PROMEDIO(F11:F30)
12	Desv. Std	208,26	<-- =DESVEST.M(F11:F30)

Observar que para la desviación estándar se utilizó la función que calcula la correspondiente a una muestra (y no a una población) ya que se trata de inferencia estadística.

Con estos valores, si se asume que la distribución de resultados tiene la forma acampanada de la distribución normal, podría decirse con un 95% de confianza que la utilidad esperada es igual a la media (\$330,5) más (o menos) 2 desviaciones estándar ($2 \cdot 208,26$). Es decir, debajo de una curva acampanada, un intervalo que está entre la media y ± 2 desviaciones involucra el 95% de la población. Mientras que si se quiere estimar con un 99% de confianza los valores que asumirá la utilidad, en una distribución normal se afirma que entre la media y ± 3 desviaciones estándar se ubica el 99% de los valores¹¹.

función de distribución, que en este caso es una distribución uniforme (en realidad responde a la función de distribución que se le asigne a la variable).

¹¹ La determinación del nivel de confianza con el que se trabajará depende de la preferencia del investigador o de la normativa vigente (en particular para el sistema financiero en que las normas de control marcan la pauta al respecto)

Hasta este momento podemos predecir el rango de valores en que se encontrará el resultado, pero aún no conocemos la probabilidad de que el proyecto resulta en, por ejemplo, una pérdida.

	G	H	I	J
11	Media	330,50	<-- =PROMEDIO(F11:F30)	
12	Desv. Std	208,26	<-- =DESVEST.M(F11:F30)	
13				
14		Límite inferior	Límite superior	
15		-86,03	747,03	<-- =H11+2*H12

Se va aclarando con el avance del ejemplo lo que es hacer una simulación: generar múltiples escenarios, y sobre ellos poder inferir. Iremos incorporando elementos que permiten tomar decisiones más certeras, y comprender interrogantes como ser: ¿porque se dieron cada uno de los valores?, ¿porque se asume que es una distribución normal?, ¿porque se tomaron solo 20 escenarios?, etc.

De todos modos, hasta este momento ya se puede apreciar la ventaja del caso C para la toma de decisiones respecto de los casos A y B, dada la posibilidad de contar con múltiples escenarios y hacer inferencia estadística. En el caso B también se hubiera podido calcular la media aritmética y la desviación estándar, obteniendo los siguientes resultados:

	E	F	G	H	I
2		B			
3		Pesimista	Más Probable	Optimista	
4	Ingresos	800	1000	1500	
5	(-) Egresos	850	800	750	
6	Utilidad (VAN)	-50	200	750	
7					
8			Media	300,00	<-- =PROMEDIO(F6:H6)
9			Desv. Std	409,27	<-- =DESVEST.M(F6:H6)

La relevancia estadística de los resultados que provienen de sólo 3 valores es menos relevante que si provienen de 20 (o 100 o 1.000 o 10.000 ... valores). Una muestra pequeña tiene mayor error al intentar efectuar una aproximación sobre la población. En cambio los múltiples escenarios brindan mayor precisión en cuanto a lo que pueda inferirse sobre la población.

Simulación Montecarlo con Software Científico

Cuando en un modelo se aplica la técnica de simulación Montecarlo, es porque dentro del modelo hay algunas variables que son inciertas. La justificación de la simulación Montecarlo es ofrecer la posibilidad a un modelo de incorporarle incertidumbre, es decir pasar de modelos que son estáticos a modelos dinámicos, que tienen la posibilidad de tener múltiples respuestas.

Simulación Montecarlo, en su forma más sencilla, es la generación de números aleatorios que son útiles para predicción, estimación y análisis de riesgo. La simulación calcula numerosos escenarios de un modelo mediante la repetición de muestras a partir de una distribución de probabilidades predefinida para la variable incierta, y utilizando dichos valores para el modelo. Todos esos escenarios producen resultados asociados en un modelo, y cada escenario implica una predicción. Cada predicción es un evento definido como una salida importante del modelo. Cada resultado es registrado y tabulado, generando un resultado probable del modelo.

En la simulación Montecarlo se generan valores aleatorios para las variables en función de su distribución de probabilidad, que son totalmente independientes, es decir que los valores generados para un evento no tienen efectos sobre los próximos valores que se generen.

Continuando con el ejemplo desarrollado previamente, la idea es ampliar el número de escenarios para obtener mayor precisión en la inferencia estadística. Si se quiere que el número de muestras sea suficientemente grande, como 100.000 observaciones, la utilización de Excel sin una herramienta especializada resulta ineficiente.

A tal efecto, en el desarrollo del ejemplo será utilizado un soft específico para simulación financiera, para ejecutar una simulación con 100.000 repeticiones¹², y valores de ingreso y egreso dentro del rango previamente indicado, tal como muestran las siguientes imágenes:

Ingresos

K	L	M	N	O	P
	Caso D				
	n Escenarios (con simulador)				
	Ingresos	1000			
	(-) Egresos	800			
	Utilidad (VAN)	200			

Simulador de Riesgo Supuestos
 Nombre = M5: Ingresos
 Triangular
 Mínimo = 800
 MostLikely = 1000
 Máximo = 1500

Egresos

	L	M	N	O	P
	Caso D				
	n Escenarios (con simulador)				
	Ingresos	1000			
	(-) Egresos	800			
	Utilidad (VAN)	200			

Simulador de Riesgo Supuestos
 Nombre = M6: (-) Egresos
 Triangular
 Mínimo = 750
 MostLikely = 800
 Máximo = 850

Al correr la simulación se obtienen los siguientes resultados:

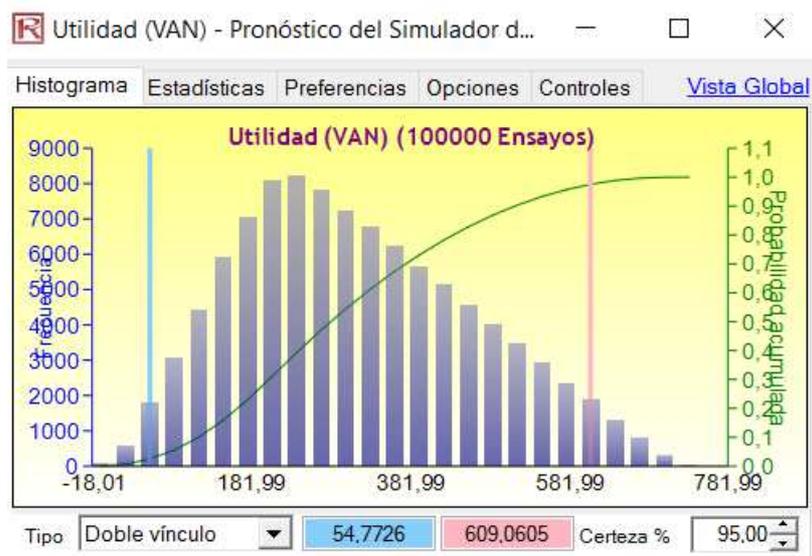
¹² El número mínimo recomendable para correr una simulación es 10.000 repeticiones

Estadísticas	Resultado
Número de simulaciones	100000
Media	299,8155
Mediana	282,0962
Desviación Estándar	148,7935
Variación	22.139,4983
Coficiente de Variación	0,4963
Máximo	732,9636
Mínimo	-38,0609
Rango	771,0245
Asimetría	0,3657
Curtosis	-0,5767
25% Percentil	185,7408
75% Percentil	404,6912
Precisión de Error al 95% de Confian...	0,3076%

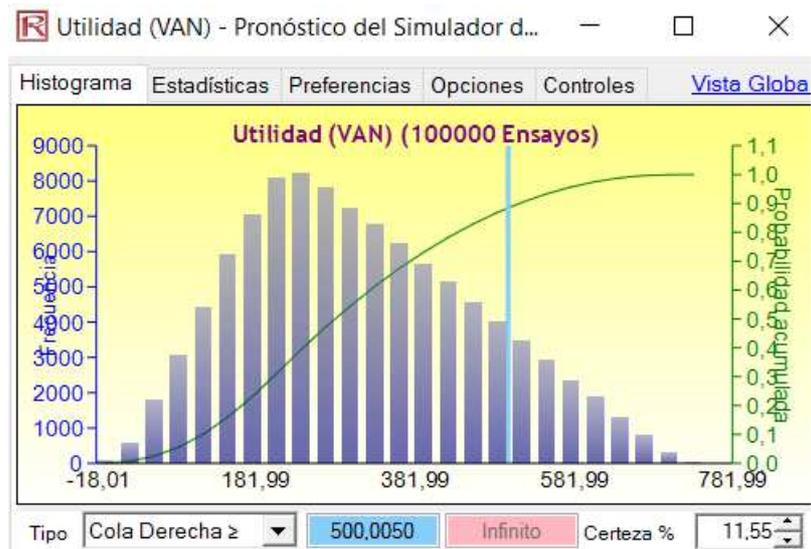
Como puede apreciarse, la media es \$299,81 la cual no es muy distinta de la calculada antes, pero la desviación estándar es muy inferior a la que se había calculado tanto con 3 observaciones como con 20 observaciones.

Lo que comienza a observarse es la importancia de las estadísticas descriptivas para la toma de decisiones, ya que en el primer escenario determinista no se contaba con ninguna información para toma de decisiones, y luego al ir incrementando el número de observaciones se tiene la posibilidad de hacer inferencia estadística. Con 3 escenarios ya se tiene alguna flexibilidad para las proyecciones, pero la muestra sigue siendo muy reducida. Incluso puede observarse que en ese caso la desviación estándar es mayor que el valor esperado del proyecto.

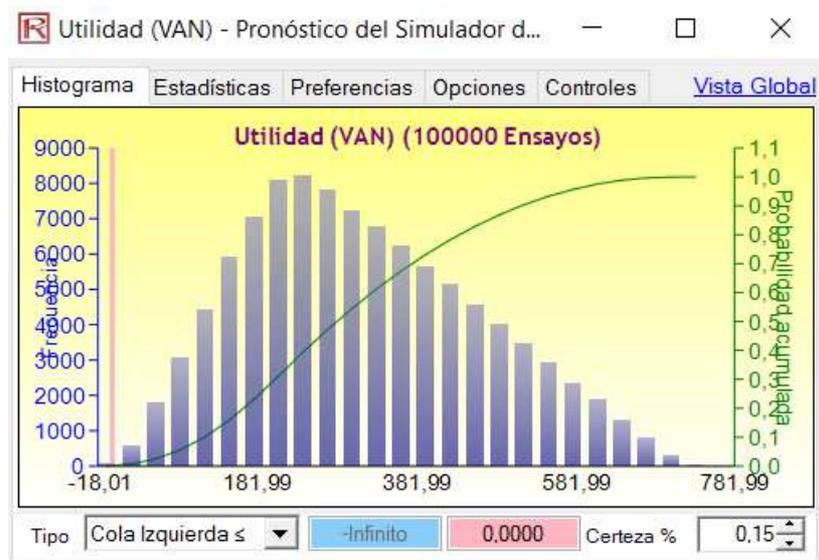
Contando con el resultado de la simulación, puede graficarse a través de un gráfico de frecuencia, y observar por ejemplo que con un 95% de certeza las utilidades del proyecto estarán entre \$54 y \$609, tal como indica el gráfico siguiente:



Puede también apreciarse en el gráfico siguiente, cuál es la probabilidad de que el proyecto genere utilidades de \$500 o más: 11,55%



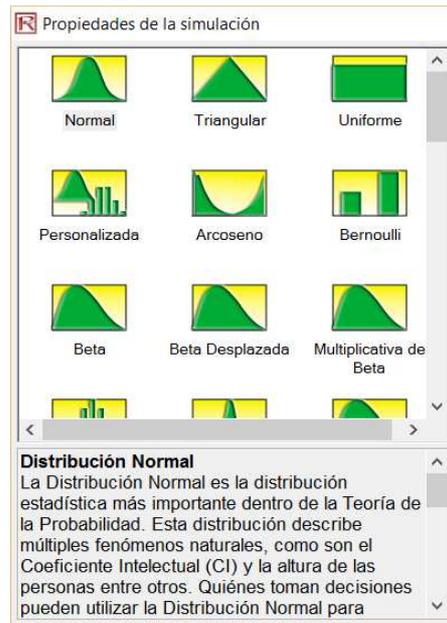
Otra variante es observar cuál es la probabilidad de que el proyecto genere pérdidas, que es una probabilidad muy baja en este caso: 0,15%



Distribuciones de Probabilidad

Los software científicos tienen programadas funciones de distribución para utilizarlas en la predicción estadística. Es decir, intentan identificar a cual función de distribución se asemejan los datos históricos de la empresa, o sea cual es la que mejor describe la situación que se quiere modelar, aunque en realidad podrían existir tantas funciones de distribución como casos se analice.

La imagen siguiente expone la selección de distribuciones de probabilidad que deber hacer el usuario al definir la simulación:



Todas están en orden alfabético, menos las primeras que son las más utilizadas e intuitivamente son las más fáciles de comprender.

Graficar los datos constituye una guía para la elección de la distribución de probabilidad. La primera pauta para identificar la distribución es listar todo lo que se conoce sobre las condiciones que rodean la variable. Reunir la mayor cantidad de información importante sobre la variable incierta a partir de los datos históricos. Si la información histórica no está disponible, puede utilizarse el propio juicio de valor, basado en la experiencia, para seleccionar aquella distribución cuyas características mejor describan el comportamiento de la variable.

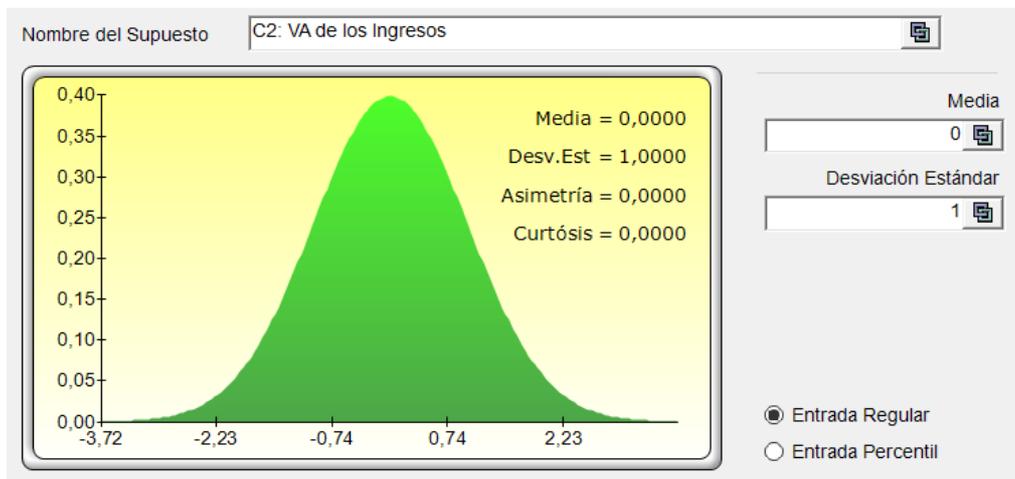
Distribución Uniforme

Es también conocida como la *distribución de resultados igualmente probables*, en la que la distribución cuenta con un conjunto de N elementos que recaen entre los mínimos y máximos y suceden con la misma probabilidad. Los valores mínimos y máximos son fijos, representando los dos parámetros de la distribución.



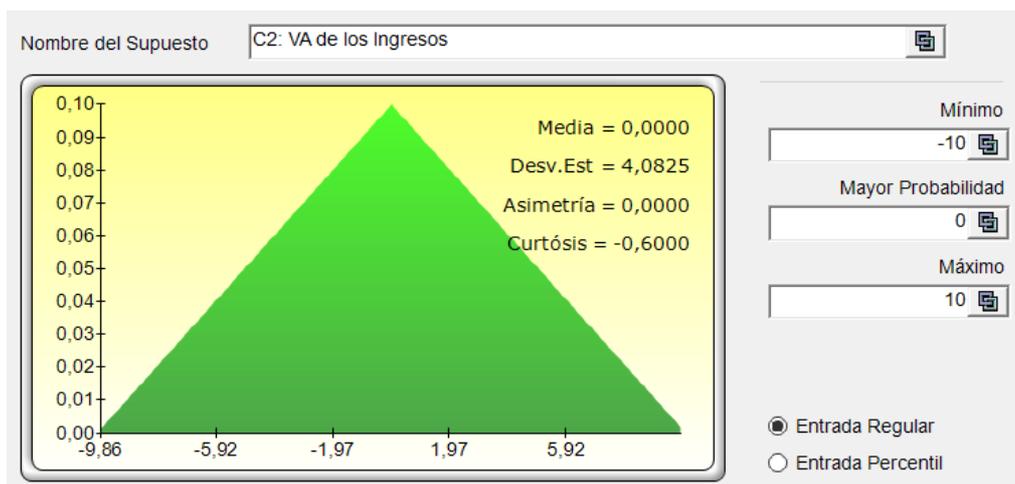
Distribución Normal

Es la distribución estadística más importante dentro de la teoría de probabilidades. Esta distribución describe múltiples fenómenos naturales, como son el coeficiente intelectual y la altura de las personas, entre otras. Quienes toman decisiones pueden utilizar la distribución normal para describir el comportamiento esperado de variables aleatorias como la tasa de inflación o el precio futuro esperado del combustible. Respecto a esta distribución existe una mayor probabilidad de que los valores de las variables aleatorias tengan cierto valor determinado (la media o el promedio de la distribución), y también es más probable que este valor se encuentre más cercano que lejano de la media o el promedio de los datos. Además la forma en la que se distribuyen los datos sigue una estructura uniforme, es decir, tiene una forma simétrica respecto a la media o el promedio de los datos tanto a la derecha como a la izquierda. La media y la desviación estándar son parámetros de la distribución.



Distribución Triangular

La distribución triangular describe una situación donde se conoce los valores mínimos, máximos y los que con mayor probabilidad pueden suceder. Por ejemplo, podría describir el número de unidades vendidas por semana de un producto, cuando las ventas anteriores muestran el número mínimo, máximo y el número habitual o más probable de unidades vendidas. Los números mínimos y máximos son fijos, el número más habitual se encuentra entre ellos, formando una distribución en forma triangular, la cual muestra que los valores que se aproximan al mínimo y al máximo tienden a ocurrir con menos frecuencia que aquellos que se encuentran cerca del valor más probable. El valor mínimo, el de mayor probabilidad y el máximo son los parámetros de la distribución.



Distribución Personalizada no paramétrica

La distribución personalizada no paramétrica es una distribución práctica basada en información disponible, es decir en alguna situación observada. Esto significa que se utiliza información comparable fidedigna disponible para definir la distribución personalizada, la simulación no tiene distribución y por lo tanto no requiere de ningún parámetro de entrada (no paramétrico). Significa que se le permite a la información que defina la distribución y no ajustar deliberadamente una distribución con base en la información disponible. La información es probada repetidamente con reemplazos utilizando el Teorema Central del Límite.

CASO 1

Se presenta a continuación un ejemplo en que se puede observar las alternativas para tratar un caso en forma determinística, y luego incorporarle variantes con la herramienta de simulación: La administración de la empresa XYZ está considerando introducir un nuevo producto. El costo fijo de producción es de \$30.000 y el costo variable \$20 por unidad. El producto se venderá a \$50 por unidad y la demanda se estima en 1.200 unidades.

- 1) ¿Cuál es el valor esperado de la utilidad del proyecto?
- 2) ¿Cuál es la probabilidad de que el proyecto termine en una pérdida?
- 3) ¿Cuál es su recomendación respecto a la introducción del nuevo producto en el mercado?

Desarrollo

Puede apreciarse que desde un punto de vista financiero, la variable de interés es la utilidad, más allá de consideraciones estratégicas que se puedan considerar. Hasta aquí es un modelo determinístico, ya que las variables no están sujetas a incertidumbre. Pero puede introducirse la incertidumbre ya que en la vida real este tipo de situaciones está sujeta a incertidumbre.

Dado que en este ejemplo no hay gastos financieros, la utilidad antes de intereses e impuestos sería la variable de interés para la toma de decisiones. Como se ve en la siguiente imagen, no es mucho lo que puede evaluarse en términos de toma de decisiones en este contexto:

	A	B	C	D	E
2	Precio	\$50			a. Valor esperado de la utilidad del proyecto: \$6.000
3	Cantidad	1.200			b. Posibilidad de pérdida: cero
4	Ingresos	\$60.000	<-- =B2*B3		c. Recomendación: la respuesta dependerá de la tasa de Rdto esperada de c/u
5					
6					
7	Costo fijo	\$30.000			
8	CV unitario	\$20			
9	CV totales	\$24.000	<-- =B8*B3		
10	Egreso	\$54.000	<-- =B7+B9		
11					
12	Utilidad	\$6.000	<-- =B4-B10		

Lógicamente en la realidad las cosas no son tan ciertas como en el modelo determinístico mencionado, por lo que se agregan a continuación supuestos de incertidumbre:

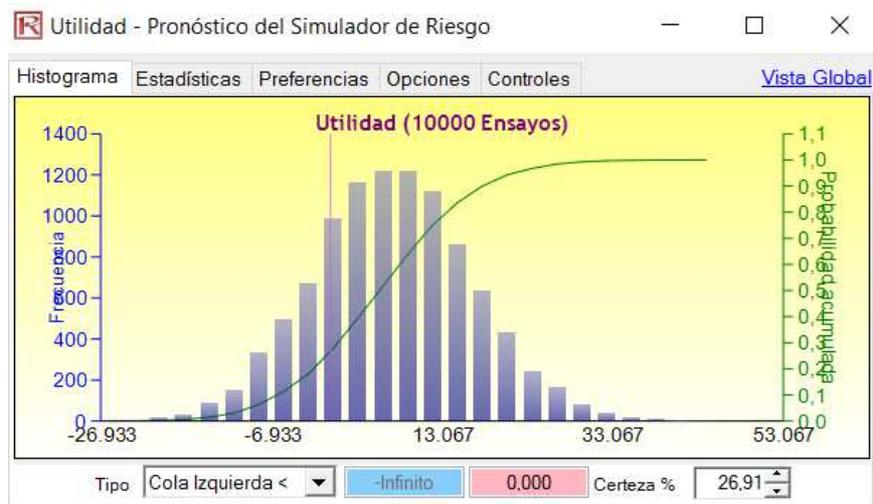
- El costo variable unitario oscila entre \$16 y \$24 (distribución uniforme)
- Demanda → media de 1.200 unidades y desviación estándar de 300 (distribución normal)

En base a estos nuevos supuestos, se vuelve a elaborar la respuesta, para 10.000 repeticiones, obteniendo el siguiente resultado de simulación:

Estadísticas	Resultado
Número de simulaciones	10000
Media	5.823,9245
Mediana	5.752,9641
Desviación Estándar	9.330,4283
Variación	87.056.892,7621
Coficiente de Variación	1,6021
Máximo	43.990,5247
Mínimo	-28.826,0627
Rango	72.816,5875
Asimetría	0,0771
Curtosis	0,0801
25% Percentil	-454,6481
75% Percentil	11.946,8979
Precisión de Error al 95% de Confian...	3,1400%

Respuesta 1: Puede apreciarse que, efectuando 10.000 repeticiones del ejercicio, el valor esperado de la utilidad será de \$5.823,92 (representado por la media). En el análisis determinístico era de \$6.000.

Respuesta 2: Puede apreciarse en el gráfico que la posibilidad de que el proyecto genere una pérdida es del 26,91%



Respuesta 3: es la pregunta más importante, ¿conviene o no emprender el proyecto?

El factor que incidirá en la respuesta es el grado de aversión al riesgo de cada agente, es decir que tan proclive está cada individuo a aceptar riesgo o no. Siendo el valor esperado de \$5.823 y la desviación estándar de 9.330, habrá personas para las cuales la posibilidad del 26% de pérdida no sea muy alta y decidan emprender el proyecto. La simulación permite brindar objetividad a los resultados en base a inferencia estadística.

CASO 2

El siguiente ejemplo describe un caso que permite apreciar la ventaja del método de simulación para toma de decisiones:

Una empresa es responsable del mantenimiento, la venta y la operación diaria de un complejo de departamentos. Actualmente está evaluando el reemplazo de los compresores de aire acondicionado (AA), por lo que desea simular el número de fallas anuales de los compresores durante los siguientes 20 años. Con los datos de un edificio similar, establece una tabla de frecuencias relativas de las fallas durante un año, como se observa a continuación:

Número de Fallas de compresores de AA	Probabilidad (frecuencia relativa)
0	0,06
1	0,13
2	0,25
3	0,28
4	0,20
5	0,07
6	0,01

Pregunta: ¿En un período de 20 años, es común tener 3 o más años consecutivos de operación con 2 o menos fallas anuales de los compresores? Es decir, se quiere conocer cuánto habrá que gastar en mantenimiento.

Desarrollo

A efectos de modelar con un simulador, debe tenerse presente qué tipo de forma se asumirá para la distribución de los datos. Hay eventos para los cuales se conoce la forma que asume la distribución, y otros para los que no se conoce de antemano¹³. Entonces, hay eventos para los cuales se sabe cuál es la distribución, hay eventos para los cuales estimo la forma de la distribución porque hay alguna teoría que lo pueda explicar, hay eventos para los cuales en base a la experiencia previa se puede estimar la distribución que debería tener. Puede ocurrir también que no se sepa qué forma tiene la distribución, pero se cuenta con información histórica suficiente para que el soft efectúe un ajuste de distribución. Si no se cuenta con ninguna información ni se conoce en forma precisa la distribución, tendrá que aplicarse el sentido común. Es por ello que las distribuciones uniforme, normal, triangular y personalizada encabezan el listado de distribuciones para generar aleatoriedad, ya que intuitivamente pueden estimarse sus parámetros.

En resumen, si se conoce la forma de la distribución, se la utiliza directamente. Si no se conoce con certeza, será el sentido común el que prevalezca en cuanto al procedimiento para determinarla. Es así ya que si se recurre a información histórica, puede que ella brinde un ajuste impreciso por no contar con la cantidad de datos suficientes, o porque el período analizado esté sesgado por alguna “estacionalidad” que no se repetirá en el futuro.

Si no se cuenta con información ni se puede estimar en forma precisa la forma, quizás corresponda utilizar la distribución normal, ya que es la forma en que normalmente ocurren los eventos. Aunque una distribución triangular puede ser incluso más sencilla, ya que no necesita conocerse la desviación estándar que requiere la normal.

Volviendo al ejemplo práctico, se efectuará una distribución personalizada con la información disponible, según se aprecia en la siguiente imagen:

¹³ Si se quiere modelar el resultado del lanzamiento de una moneda, deberá utilizarse la distribución de Bernoulli, que es una distribución diversificada con 2 resultados (cara o cruz, éxito o fracaso), por lo que también se conoce como distribución Si/No. El lanzamiento de un dado, en cambio, tendrá una distribución uniforme discreta, ya que todos los números tienen igual probabilidad de salir.



De este modo se ha utilizado una distribución personalizada, para definir el comportamiento de la variable aleatoria “fallos”. Luego puede hacerse un ajuste para identificar a que distribución “conocida” se ajusta el comportamiento.

Recordar que la incógnita es determinar, en un período de 20 años, que tan probable es que en 3 años consecutivos se dañen 2 o menos veces por año los compresores.

En este modelo el supuesto de entrada y el pronóstico de salida son la misma variable, ya que se están estudiando las fallas, y se quiere concluir sobre ello.

Corriendo la simulación para 1.000 ensayos, puede observarse el siguiente resultado:

Estadísticas	Resultado
Número de simulaciones	1000
Media	2,6930
Mediana	3,0000
Desviación Estándar	1,3246
Variación	1,7545
Coefficiente de Variación	0,4919
Máximo	6,0000
Mínimo	0,0000
Rango	6,0000
Asimetría	0,0000
Curtosis	-0,4030
25% Percentil	2,0000
75% Percentil	4,0000
Precisión de Error al 95% de Confian...	3,0485%

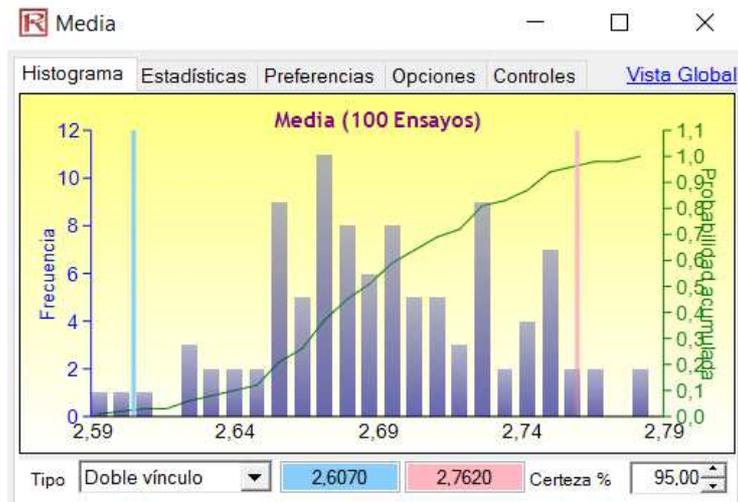
Ver que en promedio se espera que se dañen 2,69 veces los compresores en un año. Dado que se corrió la simulación para 1.000 ensayos (en este caso cada ensayo corresponde a un año), se habrá obtenido el resultado promedio para 1.000 años. Separando el análisis en tramos de 20 años se obtendría el siguiente esquema:

	A	B	C	D	E	F
1		E2: Fallos:	suma c/3 años	Prueba lógica		
2	1	2,00				
3	2	1,00				
4	3	3,00	6,00		<-- =SI(C4<=2;"ok";"")	
5	4	4,00	8,00			
6	5	3,00	10,00			
7	6	2,00	9,00			
8	7	2,00	7,00			
9	8	5,00	9,00			
10	9	1,00	8,00			
11	10	3,00	9,00			
12	11	0,00	4,00			
13	12	2,00	5,00			
14	13	2,00	4,00			
15	14	5,00	9,00			
16	15	3,00	10,00			
17	16	3,00	11,00			
18	17	2,00	8,00			
19	18	3,00	8,00			
20	19	3,00	8,00			
21	20	3,00	9,00			

Puede apreciarse que es muy baja la cantidad de veces que se verifica la consigna (3 o más años con 2 o menos fallas) por lo que aplicando un filtro se obtiene el resumen, solo 11/1000 es la cantidad de veces que se verifica la situación buscada, tal como indica el cuadro siguiente:

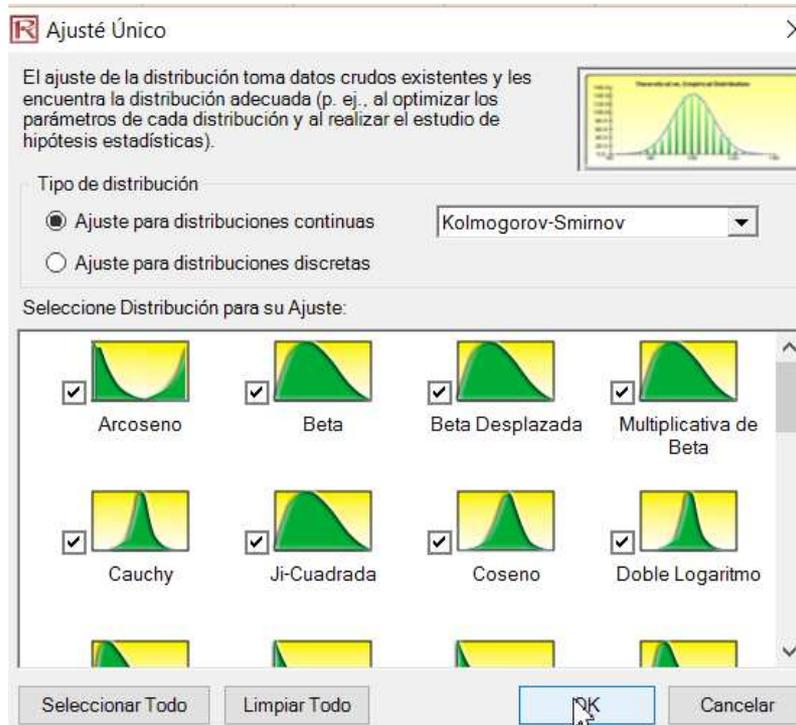
	A	B	C	D
1		E2: Fallos:	suma c/3 años	Prueba lógica
127	6	0,00	1,00	ok
128	7	1,00	1,00	ok
129	8	1,00	2,00	ok
263	2	1,00	2,00	ok
277	16	0,00	2,00	ok
278	17	1,00	2,00	ok
354	13	0,00	2,00	ok
355	14	2,00	2,00	ok
587	6	0,00	2,00	ok
991	10	1,00	2,00	ok
992	11	0,00	2,00	ok

La técnica deboot straping (estadísticas para recopilación o prueba de autosuficiencia no paramétrica) es una técnica que permite afirmar que tan confiable es un indicador estadístico obtenido en la simulación, la exactitud estadística del pronóstico, siendo que por ejemplo en este caso el valor esperado es 2,69 fallas por año, y el riesgo de que ese valor no se dé es la desviación estándar de 1,32. Se obtiene como resultado una simulación, para los valores esperados de la simulación anterior. Puede apreciarse en la imagen siguiente, que el 95% de las veces el valor esperado de las fallas anuales va a estar entre 2,60 y 2,76.

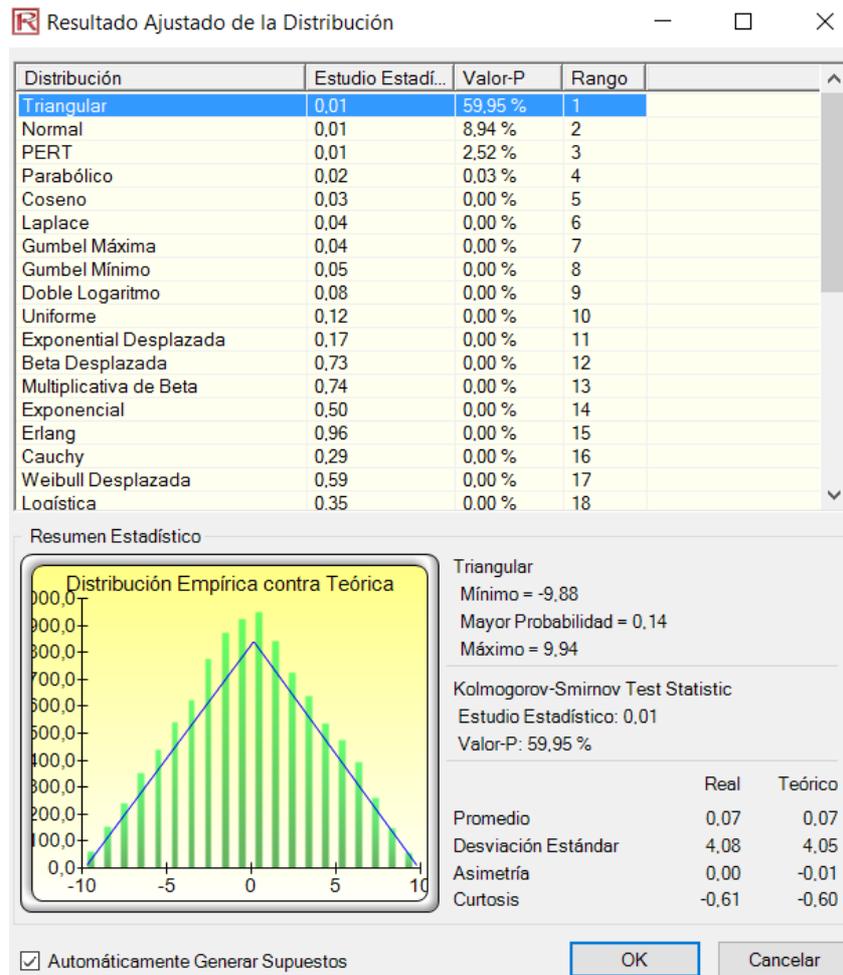


Ello indica que el resultado de la estadística tiene alto grado de confiabilidad, ya que la amplitud de resultados es pequeña. Por lo tanto se tiene otro argumento para afirmar que sería muy difícil que en un año se verifiquen menos de 2 fallas.

Por otro lado, los software de simulación permiten efectuar un ajuste de distribución, es decir, contando con la información histórica, el soft determina la forma estadística a la que más se ajusta la distribución.



Observar que el soft permite seleccionar si se trata de una variable discreta (si estuviéramos hablando de número de proyectos, etc) o continua (precios, etc). Lo que hace el soft es comparar los datos con todas las distribuciones que vienen programadas.



Lo que hace el análisis de Kolmogorov-Smirnov es observar en qué tanto se parecen los momentos de la distribución teórica con lo realmente observado. En este caso dice que la distribución que mejor explica los datos es una triangular¹⁴, indicando incluso los parámetros, aunque la normal también podría explicarlo pero con mucho menor precisión.

Puede apreciarse que el valor-P (valor probable para la prueba de hipótesis) es significativo para la distribución triangular, y no así para las restantes. Por lo tanto no se rechaza esa hipótesis. Generalizando: la regla afirma que si es mayor al 5% entonces no es rechazada la hipótesis nula.

Si la cantidad de observaciones fuera reducida, la confiabilidad en torno a la inferencia que se hace no va a ser muy buena. Es así que si se hubiera intentado efectuar el ajuste de distribución antes mencionado, para una cantidad reducida de información histórica, quizás el resultado obtenido sería erróneo.

4. Conclusión

Lo importante para toma de decisiones es utilizar la información estadística combinando varias técnicas, para apoyo al análisis con sentido común. Interviene tanto la teoría, como la evidencia empírica, o contar con información y probabilidades para determinar la técnica a

¹⁴ En este ejemplo se generaron los datos con información correspondiente a una distribución triangular, para ejemplificar.

emplear, pero en todos los casos la utilización de software científico aporta elementos en pos de apoyar con robustez estadística la tarea de análisis.

REFERENCIAS

Benninga Simon, *Principios de Finanzas con Excel*. IC Editorial 2015

Brealey, Richard; Myers, Stewart y Marcus Alan, *Fundamentos de Finanzas Corporativas*. 5ª Ed. McGraw Hill 2007

Mayes, Timothy y Shank, Todd, *Análisis Financiero con Microsoft Excel*. 5ª Ed, Cengage Learning 2009

Ross, Stephen; Westerfield, Randolph y Jaffe, Jeffrey, *Finanzas Corporativas*. 7ª Ed. McGraw Hill 2005