Teoría de la decisión con ciencia de datos[1]

Roberto Fortich Mesa, José Vergara Arrieta y Julio Hurtado Seferino

8 de abril de 2021

[1]Agradecemos a Carolina Barrios Laborda por las enriquecedoras discusiones sobre el contenido del programa. Alejandra Rojas Toro y Julián Caro Ortega brindaron asistencia de investigación en el tema de repaso de probabilidad. También agradecemos a Alba Cárdenas Escobar por comentar los recursos pedagógicos más útiles para el tema de simulaciones en hojas de cálculo.

Índice general

Lista de gráficos

Lista de cuadros

Nomenclatura

$\cap$	Intersección
$\cup$	Unión
$E(X)$	Media
$E^{\complement}$	Complemento de E
M	Millones de unidades
m	Miles de unidades

Capítulo 1

Repaso de probabilidad y estadística

Roberto Fortich Mesa y Julio Seferino Hurtado Márquez

1.1. Introducción

Este capítulo contiene una síntesis de las principales herramientas teóricas y metodológicas de estadística y probabilidad que son requisito para tomar cursos de Teoría de la decisión soportados en ciencia de datos. Está dirigido a estudiantes de pregrado en negocios y administración que estén avanzados en su formación y familiarizados con las nociones básicas de probabilidad y estadística.

Por esta razón, en este capítulo nos limitamos a presentar una selección de temas pensados para servir de referencia a los estudiantes que necesiten o quieran fortalecer posibles vacíos de estadística y probabilidad. Un aprendizaje completo de probabilidad y estadística está fuera del alcance de este capítulo, ante lo que es recomendable adquirir un buen libro de texto de probabilidad (e.g. Ross 2010; Webster 2000) para consulta de cabecera tanto en el resto de la carrera como en la vida profesional en la industria.

Para introducir el capítulo de repaso de probabilidad y estadística decidimos incluir un corto recuento de la historia de la probabilidad, que corresponde a la primera sección del capítulo. Aunque el estudio de la probabilidad y el riesgo aplicado al mundo de los negocios es reciente (c. s. XVIII, con el nacimiento de la corporación, las aseguradores y los mercados de valores) algunos de los aspectos que mejor conocemos de la incertidumbre germinaron del estudio de los juegos de azar en el renacimiento.

No obstante, y a diferencia de lo acostumbrado en muchos libros de texto de Teoría de la decisión que basan la mayoría de sus ejemplos en los experimentos clásicos asociados a algún juego de azar, en este libro redujimos a un mínimo el uso de analogías con ese tipo de actividades. En esta sección de recuento histórico de la probabilidad los juegos de azar son protagónicos pero en el resto de las notas de clase se sustituirán siempre por ejemplos aplicados a los negocios y las finanzas empresariales modernas.

1.2. Corta historia de la probabilidad

La noción de probabilidad tiene su raiz en la capacidad de mostrar comprensión acerca del azar que, hasta donde sabemos, es talento exclusivo del ser humano. Los seres humanos reconocemos que algo es fortuito cuando lo vemos, como cuando nos sorprende el "hoyo en uno" del golf, o cuando reímos al ver que alguien resbala tras pisar una cáscara de banano en el suelo. En esta usanza nuestra comprensión del riesgo suele ser instintiva y básica.

Si bien todos tenemos este instinto que nos hace reconocer que algo que ocurre es improbable, la medida en la que cada persona desarrolla su dominio de la probabilidad es muy variable, puesto que en algunas personas se manifiesta de forma precaria y rudimentaria (como los casos de los bebés apenas están descubriendo la noción de causa y efecto, o los adultos se convierten en jugadores irracionales.

Por otro lado, en todas las culturas y en todos los tiempos, se han encontrado individuos avezados en la medición, interpretación, y evaluación del azar. Su conocimiento, más refinado, quizá los hizo ser escépticos de asociar con el riesgo un origen sobrenatural. Quizá eran buenos jugadores de las primeras formas de dados y cartas. Es posible también que hayan diseñado los primeros préstamos a crédito o los primeros seguimientos sistemáticos del clima. Al desarrollo de estas habilidades de ir más allá del instinto en cuanto al riesgo y la probabilidad hoy se le llama *alfabetización cuantitativa o financiera.*

La fascinación humana con la probabilidad existe desde que se tienen registros históricos, con juegos de azar. Las primeras civilizaciones apostaban con una forma primitva de dado llamado *astrágalo*, que estaba fabricado con huesos del tobillo de las ovejas. También sabemos que ya en el 3.500 A.C. los egipcios jugaban con este objeto. Los árabes acuñaron el término *al zahr* para referirse al dado y de ahi derivamos nuestra raíz "azar", en el castellano. Ninguna de estas culturas dejó estudios escritos sobre la teoría de la probabilidad.

Las primeras civilizaciones no fueron las únicas en omitir el abordaje de la probabilidad. A pesar de ser la cuna del conocimiento de Occidente, la antigua Grecia desarrolló muy poco un enfoque cuantitativo de la probabilidad (Bernstein 1998), lo que sorprende dado lo avanzado de sus conocimientos en geometría. Sólo hasta la era del renacimiento, iniciaría los que hoy conocemos como Probabilidad en Italia y Francia, en primera instancia inspirados en el estudio de los juegos de azar.

Varios de los juegos de dados mencionados fueron llevados por las cruzadas a Europa. Allí surge una forma primitiva del pase inglés (*craps*, en inglés) que consiste en apostar al resultado que se obtendrá al lanzar dos dados en el tiro siguiente o en toda una ronda. De Francia, este juego viajaría a la región del rio Mississippi de Estados Unidos en algún momento del siglo XVIII, donde se popularizó.

Además de los dados, existe una gran variedad de juegos de azar. Es conveniente categorizar a estos juegos entre los que son de puro azar (e.g. la ruleta) y los que dependen de las habilidades del jugador (e.g. póker). La utilidad de esta clasificación radica en su extrapolación a fenómenos de los negocios. Por ejemplo, la bolsa de valores es considerada por algunos como algo casi igual a un casino de apuestas. Entre los estudiantes jóvenes suele despertar curiosidad las historias de personas que juegan blackjack o póker y logran derrotar a la casa en los casinos a partir del dominio de difíciles sistemas de conteo de cartas. La sola posibilidad de repetir este éxito estudiando la bolsa de valores es fascinante para muchos estudiantes.

La primera persona en la historia en documentar un profundo examen de la teoría de la probabilidad fue el médico italiano Girolamo Cardano (Bernstein 1998). Su tratado sobre probabilidad se llamó *Liber du ludo aleae* pero solo sería publicado póstumamente, en 1663.[1] Una de las principales contribuciones del libro de Cardano fue descubrir la utilidad de las *combinaciones* de números para descifrar las leyes de la probabilidad (Bernstein 1998, p. 51).

El primer tratado sobre probabilidad publicado y que ganó reconocimiento convirtiéndose en el estándar fue *De Ratiociniis in Ludo Aleae* de 1657, escrito por Christiaan Huygens (Hald 1988).

En 1687 se originó en Londres la primera aseguradora del mundo: Lloyd's. Inicialmente, era un lugar de reunión de capitanes de barcos ingleses en el que compartían información sobre el riesgo de enviar navíos por el mundo. Durante la segunda mitad del siglo XVII existía un ambiente fértil para el surgimiento de una aseguradora, por las guerras, la creciente clase empresarial, y el ascenso del comercio exterior. Edward Lloyd prestó en su cafetería, sin saberlo, un servicio de información sobre riesgos que hicieron de su local el punto al cual había que acudir para asegurar un navío antes de partir por el mundo a enfrentarse a tormentas y piratas en rutas de aventura. Los agentes de seguros, firmaban contratos en esta cafeteria en la parte de abajo, razón por la cual se les llamó *underwriters*. Con el tiempo, Lloyd's se convertiría en la aseguradora más famosa de la historia (Bernstein 1998, p. 91).

De los primeros esfuerzos de jugadores de azar se pasó a los aportes de Cardano y otros

[1] En ese momento, ya Blas Pascal había desarrollado de forma independiente un riguroso estudio de las leyes de probabilidad para investigar la eficiencia de las apuestas.

renacentistas. Luego la probabilidad avanzó en la astrofísica, con la determinación de las órbitas planetarias, o en las finanzas, con la creación de las primeras aseguradoras. En todo caso, se ha progresado mucho en la enseñanza de la probabilidad y la estadística. Los programas de negocios y administración de todas las universidades colombianas, cuentan actualmente en sus planes de estudio con cursos de enseñanza de estadística aplicada. En estos cursos se incluyen los fundamentos necesarios para navegar por un programa más ambicioso de alfabetización cuantitativa y financiera en el que las competencias finales permitan la administración del riesgo en la operación de un proyecto financiero y la toma de decisiones en presencia de incertidumbre. En la siguiente sección hemos recopilado un resumen de las nociones básicas de probabilidad, diseñadas precisamente para introducir al resto de capítulos que componen las notas de clase.

1.3. Definiciones y temas básicos de probabilidad

Los cuatro nociones que a nuestro parecer son centrales para adquirir una sólida fundamentación en probabilidad y riesgo son la aleatoriedad, las distribuciones de probabilidad, la variabilidad, y la comunicación en lengua natural de los grados de certeza. Esta sección está dedicada a desarrollar estos temas.

1.3.1. Aleatoriedad

Definición 1 (Aleatoriedad).
La aleatoriedad es una propiedad de un resultado de un experimento,[2] *tal que la disposición con la que salen los resultados luzca desordenada. Por ejemplo, que la lista de caras y sellos tras 10 lanzamientos de una moneda no tenga ningún patrón discernible.*

En caso de que esta disposición siguiese alguna regla general, se perdería la esencia de la aleatoriedad. Otra forma de ver que algo es aleatorio es que su proceso generador de datos no sea subyacente a ninguna causa determinística. Esto significa que se carece de una ecuación matemática con todas sus variables con valores predeterminados o con funciones de valores predeterminados.

La mejor forma de usar la aleatoriedad como herramienta de simulación, como se mostrará en los próximos capítulos, será con la incorporación de procesos generadores de datos, escritos por los estudiantes en programas de computador, en los que la aleatoriedad es consecuencia de que cada elemento del espacio muestral sea equiprobable.

[2]Un experimento o ensayo es toda acción bien definida que conlleva a un resultado único bien definido. Por ejemplo, un experimento es comprobar si un producto cumple con buenas especificaciones de fabricación, con resultado de *defectuoso* o *no defectuoso*.

De este modo conviene de aquí en adelante distinguir entre variables *puramente aleatorias* regidas por distribuciones (e.g. la uniforme) provenientes de algoritmos de computador generadores de números seudoaleatorios, y variables *aleatorias*, a secas, en las que un estadístico puede conocer su distribución de probabilidad tanto teórica como empírica y, de ese modo, simularla.

Definición 2 (Variable aleatoria).
Una variable aleatoria X es lo que ocurre si se asignan valores numéricos a los resultados de un experimento aleatorio.

Ejemplos de variables aleatorias son las siguientes:

- El precio del petróleo al final del año
- El precio del bitcoin
- El número de automóviles que serán multados este año
- La hora de llegada de un bus de Transcaribe a la estación
- La estatura de estudiantes universitarios

1.3.2. Distribución de probabilidad

El siguiente término que requiere explicación es el de *función de distribución de probabilidad* de una variable aleatoria X que, en su versión más simple, sólo toma valores discretos, ante lo cual se cambia la denominación de distribución por la de *masa* de probabilidad de la variable X, así:

Definición 3 (Función de masa de probabilidad).
Una función o distribución de masa de probabilidad de una variable discreta X, que denotaremos por $f(x)$ o $p(x)$ es una función que asigna a cada resultado x del espacio muestral un número $p(x)$, llamado su probabilidad, así:

$$p(a) = P\{X = a\}$$

Ejemplo:

$$p(0) = \tfrac{1}{6} \quad p(1) = \tfrac{3}{6} \quad p(2) = \tfrac{2}{6}$$

Donde 0, 1, y 2 son los *estados de la naturaleza.* Esta función tiene una representación gráfica en la que las alturas de la barra miden la probabilidad asociada a cada estado (Gráfico 1.2). A la representación de una realización de una función de masa de probabilidad (es decir, de su versión muestral) se le conoce como *histograma.*

Definición 4 (Estados de la naturaleza).
Los estados de la naturaleza son los diferentes resultados que pueden ocurrir tras un experimento.

El término "de la naturaleza" indica el carácter salvaje o primario del evento, pero no necesariamente tiene que tener relación con el medio ambiente. Al conjunto de los estados de la naturaleza se le llama *espacio muestral.*

La función de masa de probabilidad debe tomar valores no negativos y debe satisfacer que $\sum_{i=1}^{n} p(x_i) = 1$. Esto significa que los estados de la naturaleza deben ser mutuamente excluyentes y colectivamente exhaustivos.

Por un lado, un conjunto de estados de la naturaleza es *colectivamente exhaustivo* si contiene todos los posibles resultados del experimento. Se dice entonces que contiene todo el espacio muestral. Un contraejemplo de estados de la naturaleza que no son colectivamente exhaustivos es, *Color del semáforo en un momento determinado*:

- Verde
- Rojo

Estos estados de la naturaleza no incluyen todas las posibilidades porque excluye de la paleta a la opción *amarillo.*

Por otro lado, dos estados de la naturaleza son mutuamente excluyentes si no pueden ocurrir al mismo tiempo. Su intersección no contiene ningún elemento. Se representa como: $E \cap F = \emptyset$. Un contraejemplo de estados de la naturaleza que *no* se excluyen mutuamente es esta lista incompleta de *estados del clima*:

- Lluvia
- Sol

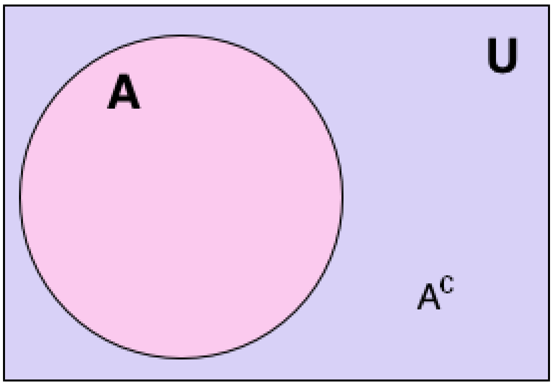
A veces conviene redefinir los estados de la naturaleza para que sean mutuamente excluyentes y colectivamente exhaustivos, porque si sólo son dos eventos, el experimento se convierte en una *distribución Bernoulli.* Luego, se puede usar la propiedad de $\sum_{i=1}^{n} p_i = 1$.

Hallar la probabilidad del *complemento* de un conjunto de estados permite, en algunos casos, mayor rapidez de cálculo.

Definición 5 (Complemento).
El complemento es el conjunto de todos los estados de la naturaleza que no estén en el estado de referencia, respecto al cual se define su complemento

El complemento de un estado de la naturaleza E se suele representar de varias formas: $E^{\complement}$, o E'. $E^{\complement}$ solo ocurre si y solo si E no ocurre, por eso $P(E^{\complement}) = 1 - P(E)$.

Gráfico 1.1: Complemento del estado de la naturaleza A

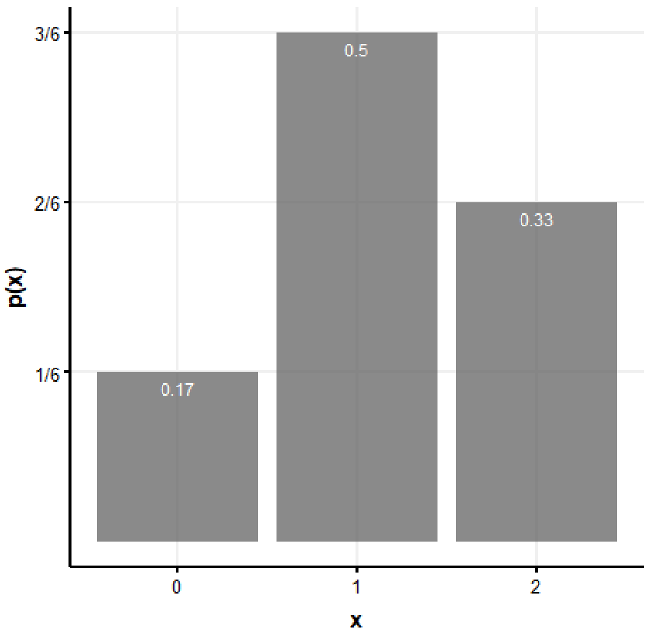


Fuente: Absolute complement [https://en.wikipedia.org/wiki/File:Absolute_complement.svg] por User:Brighterorange, 24 de enero de 2007. Dominio público.

Ejercicio 1 (Complemento de una probabilidad).
Suponga que la probabilidad de que una empresa de dulces lance su producto en menos de x fracción del año es igual a $\frac{x}{3}$ para todo $0 \leq x \leq 1$. Halle la probabilidad de que no lance el producto este año.

La respuesta se halla con el complemento del evento "la empresa lanza su producto en menos de 1 año", es decir, $1 - 1/3 = 2/3 = 66{,}6\,\%$

Gráfico 1.2: Una función de masa de probabilidad



Fuente: Elaboración propia.

Ejercicio 2 (Función de masa de probabilidad).

- *Una empresa de dulces desea lanzar un chocolate nuevo*
- *La demanda por el chocolate está proyectada como baja con un 20 % de probabilidad, media con un 40 % y alta con un 40 %*
- *Las ventas de la empresa de dulces serán de US$29 millones con demanda baja, US$36 millones con demanda media, y de US$37 millones con demanda alta*
- *¿Cuál es la función de masa de probabilidad de X, el valor de las ventas de la empresa?*

Para resolverlo, nótese que hay tres estados de la naturaleza que constituyen el dominio de la función: demanda baja, demanda media y demanda alta. La función de masa de probabilidad es:

- $p(baja) = 0{,}2$
- $p(media) = 0{,}4$
- $p(alta) = 0{,}4$

En términos de los valores monetarios, este ejemplo se puede expresar así:

- $p(29) = 0{,}2$
- $p(36) = 0{,}4$
- $p(37) = 0{,}4$

Estas funciones de masa de probabilidad le otorgan una probabilidad (rango) a un estado de la naturaleza (dominio).

Otra forma de derivar una probabilidad si no se conoce directamente el valor de la función de masa de probabilidad para un estado de la naturaleza en particular, consiste de usar el álgebra de probabilidades, sobre todo las operaciones de unión e intersección. Dos variables aleatorias tendrán estados de la naturaleza que pueden unirse e intersectarse, a partir de las siguientes definiciones:

Definición 6 (Unión).
La unión del evento E con el evento F, consiste de todos los resultados que están en E o en F, o en ambos. El evento $E \cup F$ ocurrirá si ocurre E o si ocurre F.

Definición 7 (Intersección).
La intersección de A con B es igual al conjunto de todos los elementos que estén tanto en A como en B. Se simboliza como $A \cap B$.

El estado de la naturaleza es entonces una noción fundamental para entender la probabilidad. En los ejercicios y aplicaciones de los libros de estadística, la mitad de la solución se logra mediante un planteamiento correcto de los estados de la naturaleza, que suelen estar fragmentados, requerir redefinirlos para que cumplan con propiedades deseables, agrupar para que se les pueda hallar su complemento, y sumar y restar para hallarles su unión o intersección.

1.3.3. Variabilidad

Si la aleatoriedad era el grado de dificultad en encontrar un patrón en un conjunto de valores que toma una variable, existe una noción muy cercana que la complementa: la variabilidad. Para entender por qué son términos cercanos, pensemos en que X es una constante: su variabilidad es nula y no es nada aleatoria, porque tiene el patrón más extremo en el que se repite siempre el mismo valor.

Del caso de una constante se puede pasar a introducir variabilidad a X, es decir, que cambie con muy pocos valores. Ahora, será muy probable que X sea una variable aleatoria. Donde comienzan a diferir las nociones de aleatoriedad y variabilidad es en el caso del lanzamiento de una moneda, que puede ser puramente aleatorio al tiempo que tiene poca variabilidad, entendida como la dispersión alrededor de su valor central. La relación se resume entonces en que el introducir variabilidad debe originar, en condiciones normales, menor probabilidad de que surjan patrones en los datos, pero se debe tener en cuenta que la variabilidad no es condición suficiente para la aleatoriedad.

La variabilidad de X se obtiene a partir de sus medidas de tendencia central y de dispersión. Podemos calcular el centro de una distribución de probabilidad usando tres indicadores:

- Media
- Moda
- Mediana

Definición 8 (Media).
La media de X es un promedio ponderado de los posibles valores que puede tomar X, en el que la ponderación de cada valor es igual a la probabilidad de que X asuma ese valor

La media también se conoce como el promedio, el valor esperado o la esperanza.[3] Los

[3] Algunos distinguen entre media aritmética simple y media ponderada, pero nosotros por conveniencia en estas notas de clase nos referiremos a la media ponderada como *media*, a secas. Cualquier otra versión de la media la denominaremos con el adjetivo que corresponda (i.e. aritmética simple, geométrica).

tres símbolos más usados para simbolizar la media son: $E(X), \mu$ y $\bar{X}$.

$$E(X) = \sum_{i=1}^{n} p_i x_i$$

Por ejemplo, con tres estados de la naturaleza, $n = 3$, y una función de masa de probabilidad $p(1) = 30\,\%, p(2) = 50\,\%, p(3) = 20\,\%$

$$\begin{aligned} E(X) &= \textstyle\sum_{i=1}^{3} p_i x_i \\ &= (0,3)\cdot(1) + (0,5)\cdot(2) + (0,2)\cdot(3) \\ &= 0,3 + 1 + 0,6 \\ &= 1,9 \end{aligned}$$

Ejercicio 3 (Media).
Suponga que la calificación del curso Finanzas 101 es calculada de la siguiente forma:

Actividad	Peso
Exposición	$30\,\%$
Quices	$20\,\%$
Trabajo	$25\,\%$
Examen final	$25\,\%$

Si un estudiante obtiene calificaciones de $4,5$ *en la exposición,* 4 *en los quices,* $4,2$ *en el trabajo, y* $4,6$ *en el examen final, ¿en cuánto le queda el curso?*
La media es igual a $E(\textit{calificación}) = \sum_{i=1}^{4} p_i x_i$

Actividad	Nota	Peso	Nota*peso	Subpuntaje
Exposición	4,5	0,3	4,5*0,3	1,35
Quices	4	0,2	4*0,2	0,8
Trabajo	4,2	0,25	4,2*0,25	1,05
Examen final	4,6	0,25	4,6*0,25	1,15
Nota final				4,35

Definición 9 (Moda).
Es el dato que ocurre con mayor frecuencia en la base de datos. En un gráfico de distribución de probabilidad se reconoce la moda porque es la barra con mayor altura.[4]

[4] Algunos histogramas muestran dos, o más picos, con lo que se les denomina bimodales y miltimodales, respectivamente.

Ejercicio 4 (Moda).
Para la siguiente variable aleatoria, halle la moda: $X = 3, 4, 3, 2, 3, 4, 3, 4, 1, 1, 1, 1, 3$

Definición 10 (Mediana).
Es el valor que separa a la mitad baja de los datos de la mitad alta. También puede verse como el valor medio de una base de datos, tras ordenarla de menor a mayor

La variabilidad de X se puede medir a partir de la distancia con la que sus valores están respecto a las anteriores medidas de tendencia central. A esta medición de variabilidad se le conoce como *dispersión*. Algunos de los indicadores más usados para medir la dispersión de una distribución de probabilidad son:

- Rango
- Desviación estándar
- Varianza
- Rango intercuartílico
- Coeficiente de variabilidad

Definición 11 (Rango).
Es la diferencia entre el mayor dato y el menor

El rango tiene como debilidad que es un indicador muy sensible a los valores extremos. Un conjunto de datos puede tener un rango muy alto pese a tener la gran mayoría de sus valores concentrados cerca de su tendencia central si le introducimos un valor atípico extremo en cualquiera de las colas.

Ejercicio 5 (Rango).
Si la base de datos tiene los valores 22, 40, 53, 57, 93, 98, 103, 108, 116, 121, 234, ¿cuál es el rango? La respuesta sale de restar los valores máximo y mínimo de la variable, $234 - 22$

La variabilidad es lo que subyace a cualquier tipo de investigación estadística. En esencia, es lo que motiva la probabilidad frecuentista que, recordemos, es la interpretación de la probabilidad que la define como el límite de su frecuencia relativa tras un número grande de ensayos. Medir con exactitud y comunicar con precisión la variabilidad de X es condición necesaria para una investigación de calidad científica.

Definición 12 (Desviación estándar).
Si X es una variable aleatoria con media μ, entonces la desviación estándar de X es:

$$sd(X) = \sqrt{E\Big((X - \mu)^2\Big)}$$

Donde la parte $E\Big((X - \mu)^2\Big)$ es conocida como la varianza de X, y sd es "standard deviation".

Ejercicio 6 (Desviación estándar).
Reescriba la fórmula de la desviación estándar usando los diferentes símbolos de la media $(E(X), \mu, \bar{X})$
Usando $E(X)$, se tiene: $sd = E\Big((X - E(X))^2\Big)$

La desviación estándar sirve para medir qué tan lejos respecto de la media estarán los valores de X, en promedio.[5] Es uno de los principales parámetros que definen la forma de una distribución de probabilidad y que, además, afecta el riesgo de las variables aleatorias y de los modelos financieros.

Ejercicio 7 (Desviación estándar).
Descargue los datos de cotización del dólar/peso y del bitcoin/dólar de los últimos diez dias. Halle la desviación estándar de cada serie e interprete cuál tiene mayor variabilidad.

1.3.4. Comunicación en lengua natural de los grados de certeza

Uno de los puntos más difíciles de entender para los estudiantes primerizos de la probabilidad es su significado preciso y la forma en que se traduce a expresiones de la lengua natural, es decir, el paso del lenguaje matemático al no matemático. Por esta razón, en este parte del capítulo recalcaremos los matices linguisticos con los que se estudia a la probabilidad.

Comencemos definiendo a la probabilidad como un número entre 0 y 100 que denota la frecuencia relativa con la que se produce un determinado estado de la naturaleza. A mayor probabilidad de ocurrencia, más próximo a 100 estará el porcentaje de probabilidad. También suele expresarse como una proporción que toma valores entre 0 y 1, como una fracción, o como una cuota.

[5] ¿Por qué se usa $(X - \mu)^2$ en vez de $(X - \mu)$? No es conveniente matemáticamente usar la diferencia sin elevar al cuadrado porque algunas veces la resta resultará en un número positivo (cuando el valor es mayor a la media) y otras veces en un número negativo (cuando el valor es menor a la media), entonces sacar la media de estas diferencias tenderá a ser igual a cero incluso si hay una lejanía promedio alta de los valores respecto a μ. Con el método de elevar al cuadrado, en cambio, la diferencia siempre será un número positivo que sí refleja la lejanía respecto a la media.

Algunas probabilidades vienen de leyes de la naturaleza. Por ejemplo, lanzar una moneda al aire, tiene una probabilidad de que caiga en cara de $50\,\%$. Se puede comprobar esta probabilidad si se lanza la moneda al aire una y otra vez y se cuenta cuántas veces cae en cara. Este es un ejemplo de probabilidad objetiva.

En cambio, las probabilidades subjetivas son aquellas probabilidades que reflejan creencias subjetivas sobre eventos con incertidumbre. Subjetivo se refiere a algo que pertenece a nuestro modo de pensar, y no del objeto en sí mismo.

Las probabilidades también se pueden expresar como una *cuota*.[6] Esta es la forma que se usa en las casas de apuestas deportivas, por ejemplo, $6 : 1$. Esta modalidad es conveniente porque es fácil de interpretar. Así, una cuota de $4/1$ produce al apostador una ganancia de 400 sobre una apuesta de 100.

Miremos otro caso. La cuota del partido de fútbol Colombia vs. Japón, en el Mundial de Fútbol de Rusia en 2018 fue de $7/10$. Por tanto, si uno apuesta 100 a que gana Colombia y acierta, entonces recibe 70. Nótese que entre más disparejos los números de la cuota, mayor potencial de altas ganancias tendrá el participante del juego. Un último caso más conveniente de interpretar por ser más extremo, fue la cuota que más ganancia potencial encerraba al inicio de Rusia 2018. Panamá vs. Bélgica tenía una cuota de $1071/50$. ¡Apostar 50 a que gana Panamá y acertar, implica ganar 1071!

Ejercicio 8 (Cuotas deportivas).

- *Suponga que la cuota de* $1/4$ *para un partido de Barcelona vs. Real Cartagena*
- *Esta cuota puede salir de dividir la probabilidad de que Barcelona gane* $80\,\%$ *sobre la probabilidad de que gane el Real Cartagena* $20\,\%$
- *¿Cómo averiguar la probabilidad de que Colombia le gane a Japón dado que la cuota es de* $7/10$*?*

El siguiente modo de expresar una probabilidad es el del cambio porcentual y los puntos porcentuales. Cuando un indicador sube de 80 a 88 unidades, se dice que aumentó un 10 %. En cambio, cuando una variable expresada en porcentajes sube de 80 % a 88 %, se dice que aumentó 8 puntos porcentuales. Esta diferencia suele ser confusa para algunos estudiantes.

Dado que en el campo de los negocios y la administración es común referirse a cambios en las tasas de interés, que siempre están expresadas en porcentajes, es muy común el uso del cambio expresado en puntos porcentuales por arriba o por debajo del valor de referencia.

[6] Las cuotas descritas aquí siguen el sistema de las *cuotas fraccionales* del Reino Unido. Otro sistema con una interpretación diferente es el de *cuotas decimales* usado en Europa continental y algunas otras regiones del mundo.

Ejercicio 9 (Cambio porcentual).
¿Cuánto es el cambio porcentual de una tasa de interés que sube de 3 % a 4 %?
La respuesta es igual a 1 punto porcentual.

La probabilidad suele expresarse a partir de su cantidad en una unidad especial denominada *puntos base*. Recordemos que las tasas de interés es natural que tengan cambios pequeños, por ejemplo, menores a un punto porcentual y que, a pesar de esto, estos cambios pueden tener una amplia repercusión sobre la economía. Para lidiar con esta característica de pequeña dimensionalidad, los inversionistas se refieren a estas variaciones con el lenguaje de *puntos base*. Un ejemplo: una tasa de interés que pasa de 3 % a 3,9 % se dice que subió 90 puntos base.

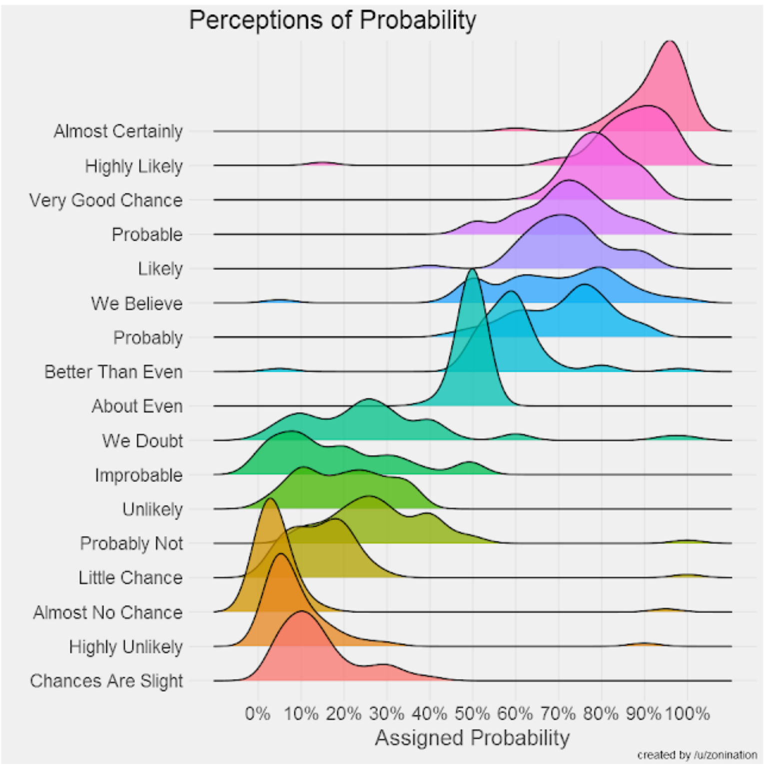
Cuando queremos referirnos en lengua natural a la certeza de un evento probabilístico usamos exprexiones como “estoy bastante seguro” o “es poco probable”. Denominaremos a estas oraciones *afirmaciones de certeza*. En algunos contextos, estas afirmaciones tienen cabida, como cuando cuando un médico dice: “estoy bastante seguro de que tu gripa no se tornará en algo grave”.

Sin embargo, existen situaciones en las que el uso de las afirmaciones de certeza deja un mensaje confuso porque ambos interlocutores carecen de herramientas para verificar con precisión las probabilidades subyacentes. Esto es, por ejemplo, lo que pasaría si el informe de una empresa encuestadora dijera que “confía en que es muy probable” que se cumplan sus predicciones de intención de voto de la población. En lugar de esto, la empresa se refiere al *margen de error* de la encuesta y lo cuantifica con cifras precisas.

La percepción que la gente tiene de los grados de certeza, como se leen en lengua natural no cuantitativa, ha resultado ser extremadamente imprecisa en múltiples investigaciones que han intentado medirla. Por ejemplo, una comunidad de aficionados a la estadística fue encuestada sobre cuáles valores numéricos le asignaban a un grupo de afirmaciones de certeza (i.e. probablemente, altamente probable) y sus respuestas tuvieron una dispersión muy alta para la gran mayoría de ellas (Gráfico 1.3).

La comunicación del grado de certeza no solo es errática por variar de persona en persona, sino que no existe un organismo que estandarice su uso. Sin embargo, entre la comunidad de científicos del clima, se han desarrollado estándares para la comunicación de la certeza de un hallazgo científico. Un buen ejemplo es la escala adoptada por el Panel Intergubernamental para el Cambio Climático (IPCC, por su sigla en inglés) que se basa en nueve categorías con una equivalencia exacta de rangos cuantitativos de probabilidad (Cuadro 1.1). Esta escala es extrapolable a otros contexto diferentes a los hallazgos de la ciencia del clima, y por eso la recomendamos en el campo de los negocios y la administración.

Gráfico 1.3: Percepciones de probabilidad de una muestra de aficionados a la estadística



Nota 1: Respuestas de los miembros de la comunidad "/r/samplesize" a la pregunta *¿Qué [probabilidad/número] le asignaría usted a la frase "[frase]"*
Nota 2: De "Perceptions of Probability and Numbers", por User: Zoni Nation, 2015, Recuperado el 8 de diciembre de 2019, del sitio web de Github: https://github.com/zonination/perceptions. Copyright 2015 Zoni Nation. Autorización del autor por la licencia del repositorio según el esquema "MIT License".

Cuadro 1.1: Lenguaje de incertidumbre calibrado por el IPCC

Afirmación	Rango
Virtualmente cierto	99 % – 100 %
Extremadamente probable	95 % – 100 %
Muy probable	90 % – 100 %
Probable	66 % – 100 %
Tanto de probable como de improbable	33 % – 66 %
Improbable	0 % – 33 %
Muy improbable	0 % – 10 %
Extremadamente improbable	0 % – 5 %
Excepcionalmente improbable	0 % – 1 %

Fuente: Mastrandrea y col. 2010, p. 3

Finalmente, tenemos una recomendación para la comunicación de probabilidades muy pequeñas. Si un evento tiene una probabilidad muy pequeña (i.e. 0,01 %) las personas con deficiente alfabetización cuantitativa y financiera tienen dificultad para interpretar y comprender la magnitud de esta frecuencia. En cambio, si el mismo evento se comunica como la probabilidad de elegir una persona entre 10.000, el mensaje será más efectivo y didáctico. Por esta razón, muchos indicadores que tienden a rondar áreas por debajo al 2 % suelen comunicarse en tasas del estilo de "camas por 10.000 habitantes", o "científicos por 10.000 habitantes".

1.4. Probabilidad de eventos múltiples

Hemos establecido que con una medición cuidadosa de la aleatoriedad y variabilidad de X, sumado al uso de un lenguaje calibrado con precisión, se puede tener una idea completa y satisfactoria del comportamiento de una variable. Sin embargo, estos son apenas casos que funcionan para variables individuales.

En esta sección definiremos e ilustraremos con ejemplos aplicados a los negocios, las principales herramientas estadísticas para examinar la probabilidad asociada a eventos múltiples. Usaremos tres situaciones durante la exposición: eventos conjuntos (simultáneos y secuenciales), eventos condicionales (simultáneos y secuenciales) y probabilidad total (simultánea y secuencial).

¿Por qué no presentamos estos temas en función de la noción de dependencia e independencia, como lo suelen hacer los libros de texto de probabilidad? Hemos optado por una presentación diferente, en la que explicaremos la dependencia en términos temporales en los que casi siempre eventos conjuntos secuenciales son dependientes (la excepción es cuando en la segunda etapa se tienen estados de la naturaleza idénticos tanto en una rama como en otra), mientras que casi siempre los eventos conjuntos simultáneos son independientes (la excepción siendo cuando se usan un mismo espacio muestral sin reemplazo). Creemos que omitir la dimensión de dependencia e independencia reemplazándola por una más intuitiva de temporalidad tiene alto valor didáctico.

1.4.1. Eventos conjuntos simultáneos

La primera probabilidad de eventos múltiples que explicaré es la de eventos conjuntos simultáneos. Comienzo definiendo la unidad natural de la que parten los eventos conjuntos estáticos: las variables categóricas.

Definición 13 (Variable categórica).
Son aquellas variables que representan atributos cualitativos de la unidad observada.[7]

Una variable categórica por construcción solo toma valores discretos. Su codificación original se puede hacer con números o con una descripción en palabras (i.e. "Categoría

[7]Una *unidad de observación* puede ser una persona, un hogar, una empresa, una ciudad, un país, entre otras. Es la escala con la que se realiza una medición y que determina cómo se interpretan las filas de una base de datos.

1", o "blanco"). Toda variable númerica puede agruparse en categorías y convertirse en variable categórica, pero lo contrario no es cierto. Las herramientas más didácticas para cruzar los valores de dos variables categóricas simultáneas son las *tabla de contingencia* y la *tabla de probabilidad.*

Definición 14 (Tablas de contingencia).
Es la tabulación en filas y columnas de dos variables categóricas

Cada celda i, j de una tabla de contigencia muestra la frecuencia de la muestra que pertenece a las categorías de la fila i y de la columna j.

Ejercicio 10 (Tabla de contigencia de eventos simultáneos).
Una empresa de dulces tiene un proceso de producción que es una variable categórica con dos categorías: endulzado natural y artificial. Simultáneamente, la empresa tiene una decisión de localización de sus plantas de producción, que es una variable categórica con dos categorías: Cartagena y Barranquilla. De las 10 plantas de producción de la empresa, 4 son de endulzado natural en Cartagena, 3 de endulzado artificial en Cartagena, 1 de endulzado natural en Barranquilla, y 2 de endulzado artificial en Barranquilla. La tabla de contingencia de esta situación se presenta en el Cuadro 1.2.

Cuadro 1.2: Tabla de contingencia simultánea de empresa de dulces

	Endulzado		
Ciudad	Natural	Artificial	Total
Cartagena	4	3	7
Barranquilla	1	2	3
Total	5	5	10

¿Cuál es la probabilidad de que en una selección aleatoria de una planta esta sea de Endulzado natural? Es igual a la frecuencia del total de la columna Natural dividida sobre la frecuencia del total de plantas. El total de plantas aparece en una tabla de contingencias en la esquina inferior derecha. La respuesta es $\frac{5}{10} = 50\,\%$

Cada celda i, j de una tabla de contingencia puede interpretarse como la ocurrencia conjunta del estado i de la variable que va en filas con el estado j de la variable en columnas.

Cuadro 1.3: Interpretación de una contingencia como ocurrencia conjunta

Ciudad	Endulzado	Ocurrencia conjunta
Cartagena (C)	Natural (N)	$C \cap N$

Cuadro 1.4: TABLA DE PROBABILIDAD SIMULTÁNEA DE EMPRESA DE DULCES

	Endulzado		
Ciudad	Natural	Artificial	Total
Cartagena	0,4	0,3	0,7
Barranquilla	0,1	0,2	0,3
Total	0,5	0,5	1

Definición 15 (Tablas de probabilidad).
Es una versión alterna de una tabla de contingencia en la que cada celda contiene la probabilidad conjunta de los eventos i y j.

Resulta de la división de la frecuencia absoluta del evento conjunto sobre el total. La tabla de probabilidad asociada al Cuadro 1.2 se presenta en el Cuadro 1.4.

1.4.2. Eventos conjuntos secuenciales

Ahora explicaremos otro tipo de eventos múltiples: los eventos conjuntos secuenciales. La herramienta más didáctica para representar los valores de dos variables categóricas secuenciales es el *árbol de probabilidad.*

Definición 16 (Árbol de probabilidad).
Es un gráfico que muestra una secuencia de eventos probabilísticos ubicados en nodos enlazados entre sí por líneas.

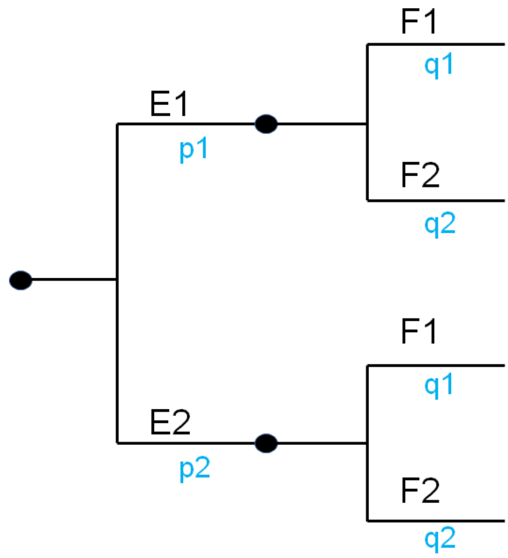
Ejercicio 11 (Árbol de probabilidad de eventos secuenciales).
Una empresa de dulces tiene 100 años de antiguedad. El 40 % ($p1$) de las plantas de producción que ha construido en su historia han sido de Endulzado natural ($E1$) y el 60 % ($p2$) de Endulzado artificial ($E2$). Después de fabricados los dulces y lanzados al mercado, puede tener una Demanda alta ($F1$) con una probabilidad de 70 % ($q1$) o una Demanda baja ($F2$) con una probabilidad de 30 % ($q2$). Nótese que la secuencia de los eventos implica dependencia estadística. El árbol de probabilidad de esta situación se muestra en el Gráfico 1.4.

¿Cuál es la probabilidad de que en una selección aleatoria de una planta esta sea de Endulzado natural y tenga Demanda alta? Es igual a $p1$ por $q1$, es decir, $(0,4) \cdot (0,7) = 0,28$

1.4.3. Eventos condicionales simultáneos

Otra probabilidad de eventos múltiples es la de eventos condicionales simultáneos. Requiere introducir primero una noción que suele generar confusión entre los estudiantes novatos de la probabilidad: la probabilidad condicional.

Gráfico 1.4: Árbol de probabilidad secuencial de empresa de dulces



Fuente: Elaboración propia.

Definición 17 (Probabilidad condicional).
Es la probabilidad de que una variable categórica sea de una categoría en particular condicionado a que, simultáneamente, otra variable categórica es de una categoría específica.

La probabilidad condicional de que la variable A sea de la categoría a condicionado a que simultáneamente la variable B es de la categoría b, es: $P\Big((A = a)|(B = b)\Big)$. El símbolo de barra vertical representa "condicionado a".

Si la variable A es de la categoría a, entonces, el hecho de que la variable B simultáneamente sea de la categoría b, implica que el evento condicionado sea un elemento tanto de a como de b. Es decir, debe estar dentro de $(A = a) \cap (B = b)$. $B = b$ será nuestro nuevo denominador sobre el cual se calcula la probabilidad, por lo que $P\Big((A = a)|(B = b)\Big) = \frac{P\big((A=a)\cap(B=b)\big)}{P(B=b)}$.

Ejercicio 12 (Condicionalidad de eventos simultáneos).
Regresemos al caso de la empresa de dulces del Cuadro 1.2. ¿Cuál es la probabilidad de que $Ciudad = C$ *condicionado a que* $Endulzado = N$*? Es igual a la frecuencia de plantas de endulzado natural en Cartagena dividida entre la frecuencia de Natural. . El total de plantas de endulzado natural aparece en una tabla de contingencias en el margen inferior.*[8] *La respuesta es* $\frac{4}{5} = 80\,\%$

En este punto los estudiantes suelen encontrar confusa la diferencia entre la probabilidad conjunta y la condicional de C y N. Para entender la diferencia conviene repasar la expresión en español que corresponde a cada una. Para el caso de la conjunta, la pregunta

[8]Por esta razón se le llama *probabilidad marginal.*

es: "en una selección aleatoria de una planta" queremos que caiga tanto en las categorías C como N, y por eso el punto de referencia son las 10 plantas totales del conjunto universal. En cambio, en el caso de la condicional la pregunta es "restringiéndonos sólo a la muestra de plantas de endulzado natural" queremos que caiga tanto en las categorías C como N y por eso el punto de referencia son sólo las 5 plantas de endulzado natural.

1.4.4. Eventos condicionales secuenciales

Ahora describiremos un caso en el que los eventos son condicionales porque uno de ellos ocurre cronológicamente primero que el otro, es decir, son eventos condicionales secuenciales.

Ahora explicaremos la probabilidad condicional en el contexto de eventos condicionales secuenciales.

Ejercicio 13 (Condicionalidad de eventos secuenciales).
Regresemos al caso de la empresa de dulces del Gráfico 1.4. ¿Cuál es la probabilidad de que la Demanda sea alta ($F1$) condicionado a que el Endulzado fue natural ($E1$)? Es igual a la probabilidad $q1$.

La interpretación en el caso secuencial de una condicionalidad es muy intuitiva. $F1|E1$ significa que el evento a la derecha de la barra vertical ocurre primero y, condicionado a que la empresa dulces sabe que ocurrió $E1$, ahora cuál es la probabilidad de que ocurra $F1$. Si ya se tiene el valor de $q1$ en el árbol de probabilidad, hallar la condicional no requiere ningún cálculo adicional.[9] $P(F1|E1) = q1$. Esto significa que una probabilidad condicional es la que aparece naturalmente en el árbol, y no requiere ningún cálculo adicional. Al aparecer en un nodo no inicial ya tiene implícita la condicionalidad sobre todos los nodos que le anteceden.

Ejercicio 14 (Probabilidad condicional secuencial).
En una fábrica de lentes de cámara de celular se muestrean 3 lentes de cámara de celular, que son secuencial y aleatoriamente elegidos de una caja. El objetivo es revisar cada lente y anotar si es frontal o trasero. Un ejemplo de esta revisión es:

1. *Primer lente: trasero*
2. *Segundo lente: frontal*
3. *Tercer lente: trasero*

[9]Sin embargo, la probabilidad inversa de $E1|F1$, que aparentemente viola la secuencia cronológica del árbol, requiere unos cálculos menos directos que se explicarán en el Capítulo 2 cuando se explique el *teorema de Bayes*.

Cuadro 1.5: Tabla de probabilidad simultánea de empresa de dulces

	Endulzado		
Ciudad	Natural	Artificial	Total
Cartagena	0,4	0,3	0,7
Barranquilla	0,1	0,2	0,3
Total		0,5	1

La caja que se está revisando contiene 2 lentes traseros (t) y 2 frontales (f), con $t + f \geq 3$. *Si añadimos una etapa previa a la elección del primer lente, en la que alguien nos informa cuántos de los lentes elegidos son frontales, este es un problema de probabilidad de eventos secuenciales condicionales. Condicional a que 1 de los 3 lentes elegidos es frontal, ¿Cual es la probabilidad de que el primer lente revisado sea frontal? Recalcule esta probabilidad si en la etapa previa le informan que 2 de los lentes elegidos son frontales. Finalmente, generalice la fórmula de esa probabilidad condicional, para cualquier valor de* k *y* n*, donde* k *es el número de lentes frontales en la muestra y* n *es el número de lentes muestreados*
Primero, si nos dicen que 1 de los 3 lentes elegidos es frontal, es porque estamos en alguno de estos espacios muestrales: ftt, tft, ttf*, por lo* $P(f|\text{un frontal}) = 1/3$.

1.4.5. Principio de probabilidad total para eventos simultáneos

Ahora explicaremos un caso especial de eventos conjuntos: el *principio de probabilidad total.* En
probabilidad marginal que, como mostramos, es la que aparece en los márgenes de las tablas de probabilidad.

Empecemos suponiendo que deseamos hallar la probabilidad marginal del endulzado natural en el contexto de eventos simultáneos, pero que la tabla de probabilidades está incompleta, como en el Cuadro 1.5.

Para hallar la probabilidad marginal de *Natural* se deben sumar las probabilidades conjuntas de $Natural \cap Cartagena$ con la de $Natural \cap Barranquilla$. De este modo, la probabilidad marginal hallada mediante el principio de probabilidad total es la suma de todas las formas posibles de que un elemento haga parte de *Natural.*

$$P(Natural) \quad = P(Natural \cap Cartagena) + P(Natural \cap Barranquilla)$$

1.4.6. Principio de probabilidad total para eventos secuenciales

En un árbol de probabilidad hay eventos probabilísticos que aparece en dos o más nodos de una etapa avanzada del árbol. Hallar la probabilidad total de que ocurra ese estado de la naturaleza en una segunda etapa implica reagrupar todos los casos, mediante la suma de las probabilidades conjuntas asociadas a cada aparición.

Ejercicio 15 (Probabilidad total en árboles de probabilidad).
Supongamos que deseamos calcular la probabilidad total del evento $F1$. Los caminos por los que se puede llegar a este evento en la segunda fase del árbol son $E1$ y $E2$. Por tanto su probabilidad total será la suma de ambos caminos:

$$\begin{aligned}
P(F1) &= P(F1 \cap E1) + P(F1 \cap E2) \\
P(F1) &= P(F1|E1) \cdot P(E1) + P(F1|E2) \cdot P(E2) \\
P(F1) &= q1 \cdot p1 + q1 \cdot p2 \\
P(F1) &= (0{,}7) \cdot (0{,}4) + (0{,}7) \cdot (0{,}6) \\
P(F1) &= 0{,}28 + 0{,}42 \\
P(F1) &= 0{,}7
\end{aligned}$$

Ejercicio 16 (Probabilidad total en árboles de decisión).
Un modelo del precio de una acción supone que cada día el precio de la acción sube una unidad con probabilidad p o disminuye una unidad con probabilidad $1-p$. El cambio en los diferentes días se supone que es independiente. ¿Cuál es la probabilidad de que luego de dos días la acción esté en su precio original? ¿Cuál es la probabilidad de que luego de tres días el precio de la acción haya subido una unidad? ¿Cuál es la probabilidad de que el tercer día la acción haya subido condicionado a que el primer día haya subido?
Para que luego de dos días la acción esté en su precio original se tiene: $\{sube, baja\} \vee \{baja, sube\}$

Con el conocimiento de las anteriores herramientas estadísticas los estudiantes podrán intentar solucionar problemas de incertidumbre y riesgo aplicando la teoría de la probabilidad asociada a eventos múltiples. La comprensión de la diferencia entre eventos conjuntos y condicionales, y su distinción entre los casos simultáneos y secuenciales, es una de las habilidades más difíciles de adquirir en un curso de introducción a la probabilidad, y por esa razón le dedicamos a esos temas una amplia cobertura. Si un estudiante entiende esa distinción y domina el cálculo de probabilidades totales, tendrá el camino sembrado para el estudio de problemas bayesianos y de árboles de decisión.

1.5. Distribuciones de probabilidad

Las variables aleatorias, es decir, las que ocurren cuando se asignan valores numéricos a los resultados de un experimento aleatorio, tienen frecuencias relativas que vienen en muchas formas y tamaños diferentes. Conocer las tipos comunes de distribuciones de probabilidad es una de las competencias más importantes en la estadística paramétrica.

Los objetivos de aprendizaje en este tema de la probabilidad incluyen: el estudio de los fenómenos de la vida real que se pueden asociar naturalmente con cada distribución, la representación gráfica de los parámetros (tanto en el gráfico de densidad marginal como en el de la acumulada), las relaciones que tienen los parámetros de diferentes distribuciones, y la obtención empírica de estimadores de los parámetros.[10]

Así como el conocimiento léxico de las principales palabras abre paso a todo un mundo de posibilidades del estudio de un idioma, en estadística el examen profundo y sólido de las principales funciones de distribución de probabilidad, discretas y continuas, es el vocabulario esencial que permitirá luego escribir simulaciones de Monte Carlo de procesos de incertidumbre y riesgo en las empresas.

Las distribuciones de probabilidad de una variable aleatoria pueden clasificarse en dos grupos:

- Discretas
- Continuas

Una variable aleatoria es *discreta* si puede tomar un número de posibles valores finito, o infinito pero enumerable. Las variables aleatorias discretas tienen funciones de masa de probabilidad, es decir listas en las que se asocia una probabilidad a cada valor de X.

Una variable aleatoria es *continua* si puede tomar un número de posibles valores infinito no enumerable. Las variables aleatorias continuas tienen funciones de distribución de probabilidad, que usualmente se expresan como una función matemática que resume en una función cómo transformar X en una probabilidad, es decir, mapea del estado de la naturaleza a los números reales. A esta ecuación se le llama la función de distribución de probabilidad.[11]

1.5.1. Funciones discretas

Algunas de las principales funciones de distribución de probabilidad discretas son:

- Bernoulli
- Binomial
- Poisson

[10]Un *estimador* es una regla o fórmula para calcular un valor estimado de un estadístico, en una muestra de datos, mientras que un *parámetro* es una cantidad que indexa una distribución de probabilidad y que define la característica númerica de esa distribución

[11]Otros autores denominan a la función de distribución de probabilidad, la *densidad* de probabilidad y a la función de distribución acumulada la llaman confusamente función de distribución. En nuestro caso cuando usemos función de distribución de probabilidad no nos estamos refiriendo a la acumulada sino a su derivada.

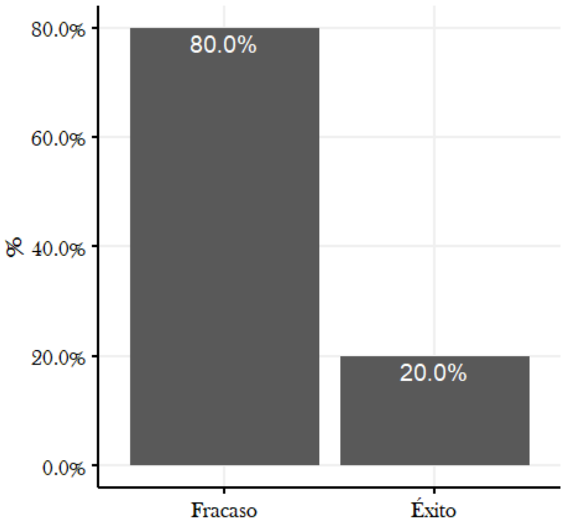
1.5.1.1. Distribución Bernoulli

Definición 18 (Distribución Bernoulli).
Una variable aleatoria X sigue una distribución Bernoulli si su función de masa de probabilidad es:

$$\begin{aligned} p(0) &= P\{X=0\} &= 1-p \\ p(1) &= P\{X=1\} &= p \end{aligned}$$

donde $p, 0 \leq p \leq 1$, es la probabilidad de que un ensayo resulte en un éxito.[12]

Gráfico 1.5: Ejemplo de una distribución Bernoulli con $p = 0{,}2$



Fuente: Elaboración propia.

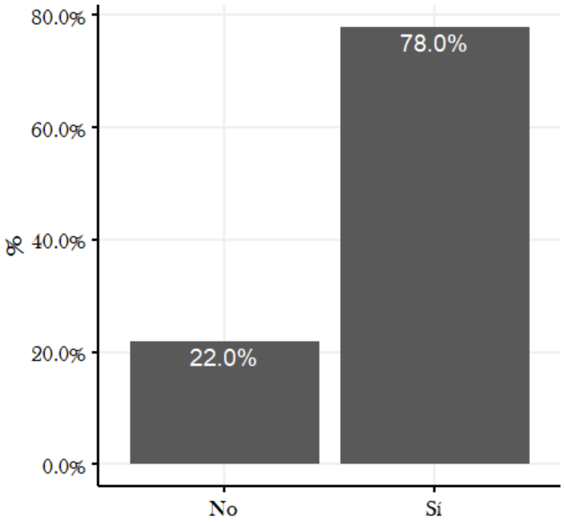
Esta distribución fue formulada por el matemático Jakob Bernoulli en el siglo XVII. Lo que la define es el poder dividir el espacio muestral en dos partes. La media de una Bernoulli es igual a la probabilidad de éxito, p, mientras que su varianza es: $p(1-p)$. Una Bernoulli puede obtenerse de una Binomial en la que $n = 1$, es decir, la Bernoulli es un caso especial de la Binomial.

Un ejemplo cotidiano de una variable que se distribuye como una Bernoulli es la respuesta a la pregunta "¿Ha visto videos del Youtuber Germán Garmendia? En 2019, una muestra de estudiantes de la UTB respondió a esta pregunta resultando en el histograma del Gráfico 1.6, mostrando una alta popularidad de Garmendia entre los encuestados.

Curiosamente, una variable que se distribuya como una Bernoulli con una alta probabilidad de éxito tendrá muy poca varianza, pero pese a esta escasa variabilidad, no pierde su carácter de ser una variable aleatoria. La aleatoriedad máxima de una Bernoulli, sin

[12] Los términos *éxito* y *fracaso* son arbitrarios e intercambiables. La intuición de un "éxito" es que un apostador obtiene el resultado deseado, pero no tiene que referirse a una apuesta

Gráfico 1.6: Proporción de estudiantes que han visto videos del Youtuber Germán Garmendia, 2019



Nota: la muestra fue de 18 estudiantes de la Facultad de Economía y Negocios de la UTB en el segundo semestre de 2019.
Fuente: trabajo de investigación de Alejandra Rojas Toro.

embargo, coincide con su estado de máxima varianza, y se alcanza cuando la probabilidad de éxito y fracaso son equiprobables, lo que resulta en una varianza de $0{,}25$.

Un fenómeno de la vida real interesante de modelar como una Bernoulli porque luego se conectará naturalmente con la distribución de Poisson, es la probabilidad de tomar al azar una página de un borrador de un libro y encontrar un error (gramatical u ortográfico). En el caso de la mayor parte de escritores competentes que producen borradores de libros, se espera que p sea muy bajo y que se distribuya como una Bernoulli.

Aunque parezca trivial modelar este ejemplo con una probabilidad de éxito diminuta (i.e. menor a 1/200), al pasar al problema agregado de hallar el número de errores cometidos en un libro de muchas páginas, nuevamente cobra interés realizar el modelo estadístico de esta situación.

1.5.1.2. Distribución Binomial

Definición 19 (Distribución Binomial).
Sirve para modelar el número de éxitos que ocurren en n ensayos Bernoulli. Formalmente, una variable X sigue una distribución Binomial si su función de masa de probabilidad es:

$$p(i) = \binom{n}{i} (p)^i (1-p)^{n-i} \quad ; i = 0, 1, \ldots, n$$

Donde $\binom{n}{i}$ representa el número de diferentes grupos de tamaño i que pueden ser seleccionadas de un conjunto de n objetos en el que el orden de selección es irrelevante. Este número de grupos se calcula con la fórmula de una combinación, que es igual a: $\frac{n!}{(n-i)!i!}$

Ejercicio 17 (Distribución Binomial).
Una empresa compra transistores en lotes. El evento que nos interesa medir es "se examinan tres transistores consecutivamente y sale un éxito". Cada ensayo se distribuye como una Bernoulli con un éxito ocurriendo con probabilidad de $20\,\%$ *y un fracaso ocurriendo con probabilidad de* $80\,\%$*. Para este ejemplo, éxito es igual a que el transistor salga dañado.*

- *"Se examinan tres transistores consecutivamente y sale un éxito" se escribe* $n = 3$*;* $i = 1$
- *Este evento significa que se halla un transistor dañado en alguno de tres transistores elegidos aleatoriamente para revisión.*

$$\begin{aligned}
p(i) &= \binom{n}{i}(p)^i(1-p)^{n-i} \qquad i = 0, 1, \ldots, n \\
p(1) &= \binom{3}{1}(0{,}2)^1(1-0{,}2)^{3-1} \\
p(1) &= \tfrac{3!}{(3-1)!1!}(0{,}2)(0{,}8)^2 \\
p(1) &= \tfrac{6}{2!(1)}(0{,}2)(0{,}64) \\
p(1) &= \tfrac{6}{2}(0{,}128) \\
p(1) &= 3(0{,}128) = 0{,}384 = 38,4\,\%
\end{aligned}$$

Si se examinan tres transistores, la probabilidad de encontrar exactamente uno dañado es igual a 38,4 %. En las distribuciones binomiales debe tenerse cuidado en no confundir la probabilidad de tener *exactamente un éxito* con la probabillidad de tener *al menos un éxito.*

La media de una variable aleatoria distribuida binomial es igual a np. Su varianza es $np(1-p)$.

Ejercicio 18 (Distribución Binomial).
Un repartidor de periódicos los compra a \$10 y los vende a \$15. Sin embargo, tiene prohibido devolver

los periódicos que se queden sin vender. Si la demanda diaria es una variable aleatoria binomial con $n = 10$, $p = \frac{1}{3}$, ¿Cuál es la probabilidad de que venda cuatro periódicos? ¿Cuál sería la media de su ganancia en ese caso? ¿Si compra 10 periódicos, cuál será la media y la varianza de periódicos vendidos?

Una variable Binomial con un parámetro n grande tiende a tener la misma distribución de probabilidad que una Normal con la misma media y varianza que la Binomial original.

1.5.1.3. Distribución Poisson

Definición 20 (Distribución Poisson).
Es aquella en la que la variable X describe un número de éxitos (i.e. errores, accidentes) en un intervalo de longitud L.

La función de distribución de probabilidad Poisson se describe así:

$$f(x) = \frac{\lambda^x exp(-\lambda)}{x!}$$

Donde λ es un parámetro positivo, y x es el número de éxitos.

Esta disribución suele emerger en procesos que ocurren de forma natural en intervalos de longitud L, que pueden ser de dimensión temporal o espacial.

Es la distribución de probabilidad que describe de forma alternativa experimentos binomiales en los que el número de repeticiones es muy grande y la probabilidad de éxito es muy pequeña. Por ejemplo, el número de erratas en un grupo de páginas de un libro se distribuye como una Poisson. Esta distribución es aproximadamente igual a una binomial en la que la probabilidad de error en una página individual sea muy pequeña y el número de páginas del libro sea muy alto.

Otro ejemplo cotidiano de una Poisson es el número de aviones que aterrizan en una pista de aeropuerto cada diez minutos, que se constituye en un intervalo de tiempo.[13] En el caso de un aeropuerto, la densidad de distribución de probabilidad luce como se muestra en el Gráfico 1.7. Un interpretación en este ejemplo, es que la mayoría de pistas de aeropuertos ven aterrizar un número de aviones dentro del intervalo entre 2 y 8.

Los intervalos que caracterizan a las variables distribuidas como una Poisson tienen que cumplir ciertas condiciones, o de lo contrario, la función no los representará fidedignamente.

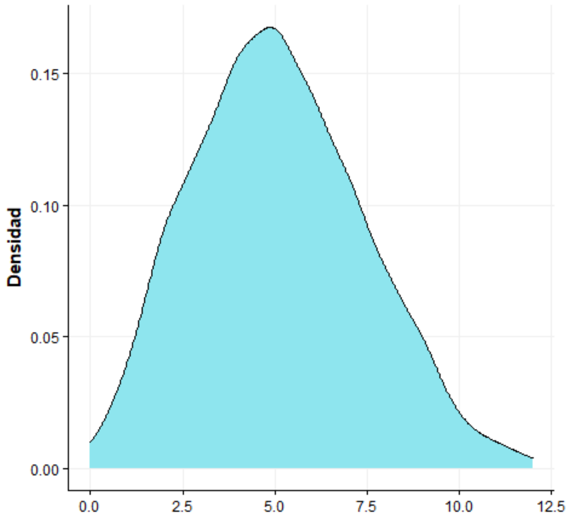
1. La probabilidad de que ocurra un éxito debe ser constante en todos los subintervalos

[13] Al menos en periodos homogéneos (i.e. en una hora valle).

2. No puede haber más de un éxito en cada subintervalo

3. Los éxitos deben ser independientes entre sí

Gráfico 1.7: Simulación de una densidad de distribución del número de aviones que aterrizan durante hora valle en una pista



Nota: Esta es apenas una de las formas que adquiere una Poisson, y representa el caso cuando $\lambda = 5$. Con valores de lambda menores la densidad se torna mucho más asimétrica.
Fuente: trabajo de investigación de Paola Hernández Jaraba.

Una distribución Poisson se relaciona con la distribución exponencial, porque esta describe un el tiempo que transcurre entre dos eventos Poisson.

1.5.2. Distribuciones continuas

Si la variable en nuestros datos del modelo es numérica en vez de categórica entonce se dice que es *continua*. Sin embargo, toda variable continua se puede convertir en discreta agrupándola en categorías según intervalos de ciertos valores, a esto se le llama *discretizar* la variable. Lo contrario, es decir, convertir a una variable discreta en continua, por lo general es imposible sin introducir algun supuesto arbitrario de desagregación.

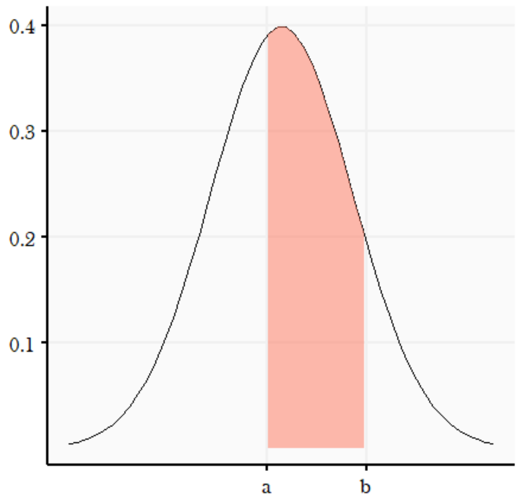
Algunas de las principales funciones de distribución de probabilidad continuas son:

- Uniforme
- Triangular
- Normal
- Chi cuadrado
- Logística

- Exponencial
- Gumbel

En distribuciones discretas, la altura de las cajas del histograma son iguales a la probabilidad asociada al valor de la variable x en el centro de la base de la caja. La suma de las alturas de varias cajas de un histograma, equivalen a la probabilidad de que x esté dentro un intervalo de valores. A medida que las cajas se acercan a un tamaño muy pequeño, la suma de sus alturas se aproxima al área debajo de la curva de la función de distribución de probabilidad (Gráfico 1.8).

Gráfico 1.8: Probabilidad de que la variable tome un valor entre a y b



Fuente: elaboración propia.

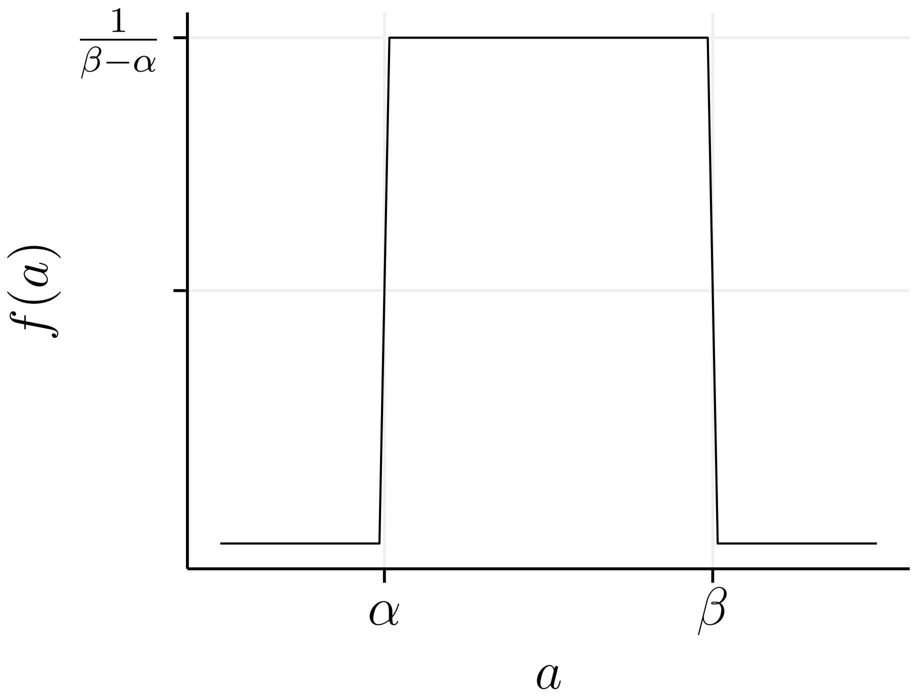
1.5.2.1. Distribución uniforme

Definición 21 (Distribución uniforme).
Es la distribución de probabilidad de una variable que es igual de probable de asumir cualquier valor en un intervalo.

Para el intervalo (α, β) la función de densidad de probabilidad de una uniforme es: $f(x) = \frac{1}{\beta - \alpha}$. La distribución uniforme produce para variables continuas el mismo efecto que las distribuciones equiprobables tienen en las variables discretas: que todos los estados de la naturaleza sean igualmente probables, lo que genera un gráfico plano horizontal (Gráfico 1.9). En ambientes de incertidumbre, una forma de convertir el problema en uno de riesgo es suponer total aleatoriedad con una distribución uniforme.

Si se define sobre el intervalo $(0, 1)$, la distribución uniforme es *unitaria*. En ese caso, la probabilidad de que X esté en algún subintervalo de $(0, 1)$ equivale a la longitud de ese subintervalo.

Gráfico 1.9: Gráfico de una distribución uniforme



Fuente: elaboración propia

La media de una uniforme es igual a $\frac{\beta+\alpha}{2}$ y su varianza es $\frac{(\beta-\alpha)^2}{12}$. La uniforme es la distribución que conjuntamente presenta mayor aleatoriedad/variabilidad de todas. Su función de distribución acumulada es: $F(a) = \frac{a-\alpha}{\beta-\alpha}$ y se representa como en el Gráfico 1.10 .[14]

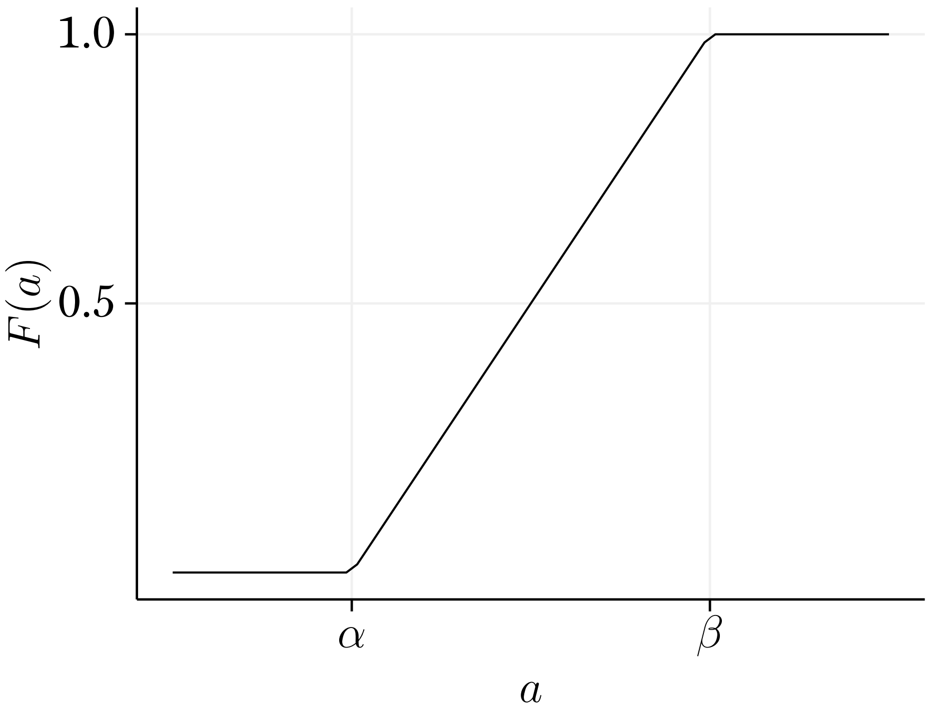
Ejercicio 19 (Distribución uniforme).
Sea X una variable aleatoria continua distribuida como una uniforme en el intervalo $(0, 10)$

- *Halle α y β*
- *Grafique la función de distribución de probabilidad*
- *Halle la fórmula de la distribución de probabilidad acumulada*
- *Encuentre la probabilidad de que $X < 3$*
- *Encuentre $P3 < 3X + 2 < 5$*

Los parámetros de la distribución en este ejemplo son: $\alpha = 0$ y $\beta = 10$. Su distribución de probabilidad se representa en el Gráfico 1.11. Su fórmula de distribución de probabilidad acumulada es igual a $F(a) = \frac{a}{10}$.

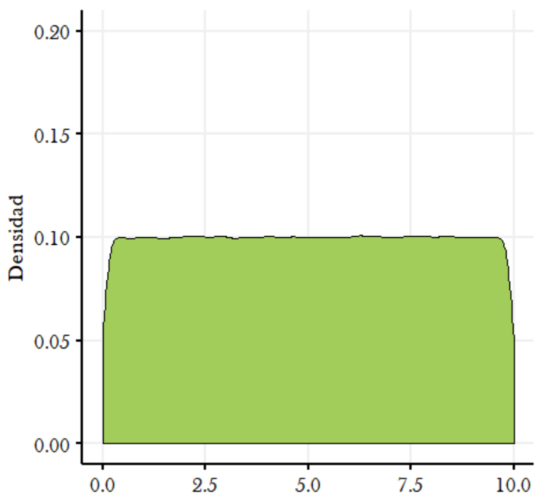
[14]Las minúsculas suelen servir de notación para las funciones de distribución de probabilidad, mientras que las mayúsculas para las funciones acumuladas.

Gráfico 1.10: Gráfico de la función de distribución acumulada de una uniforme



Fuente: elaboración propia.

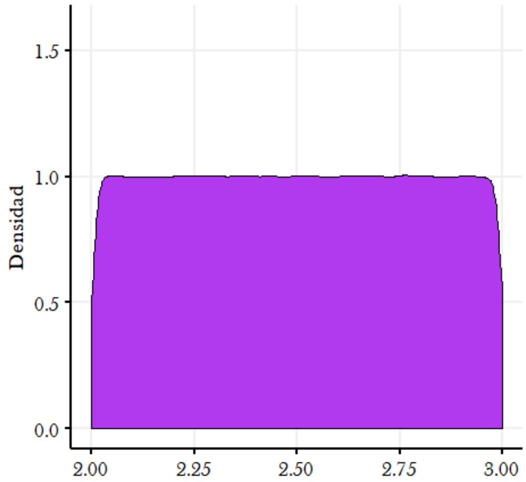
Gráfico 1.11: Función de distribución de probabilidad del ejercicio 19



Fuente: elaboración propia.

Un ejemplo de distribución uniforme lo muestra Liao (1990, p. 51) para modelar los costos fijos de una empresa fabricante de hardware, con un supuesto del rango de valores entre 2 y 3 millones de dólares anuales (Gráfico 1.12). Otro ejemplo de distribución uniforme , más cotidiano, es el tiempo de llegada del Transcaribe a una estación para un intervalo horario dado.

Gráfico 1.12: Gráfico de costos fijos distribuidos como una uniforme, en millones de dólares



Fuente: elaboración propia a partir de Liao (1990).

1.5.2.2. Distribución Triangular

Definición 22 (Distribución Triangular).
Se llama triangular porque el gráfico de su función de distribución de probabilidad es un triángulo (Gráfico 1.13). Tiene esta función de densidad de probabilidad:

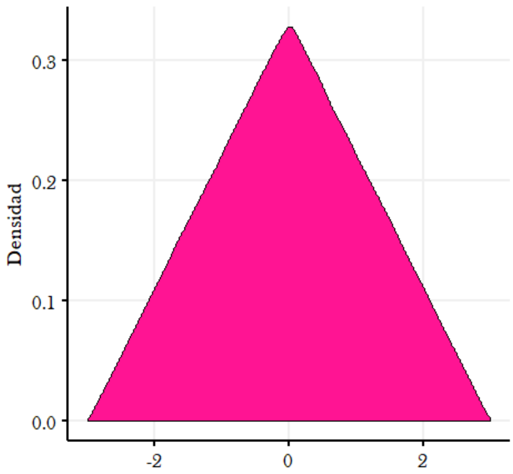
$$f(x) = \begin{cases} \frac{2(x-a)}{(b-a)(c-a)} & a \leq x \leq c \\ \frac{2(b-x)}{(b-a)(b-c)} & c \leq x \leq b \end{cases}$$

Una triangular tiene como ventaja frente a otras distribuciones que representa bien poblaciones con pocos datos. Funciona bien cuando se conoce un mínimo y un máximo, porque solo requiere añadir la moda, el cual es un valor que se puede tomar como una "suposición inspirada". Según Johnson (1997, p. 387), muchas cantidades inciertas tienen mínimos y máximos reconocibles, por lo que se torna conveniente modelar su riesgo con distribuciones con límites finitos.

También, puede resultar de la combinación de variables con otras distribuciones. Por ejemplo, si una variable es igual a la suma de dos variables aleatorias continuas distribuidas como una uniforme, entonces se distribuye como una triangular.

Ejercicio 20 (Distribución Triangular).
Un analista con experiencia puede ayudar a estimar el número de clientes que efectivamente llegarán a comprar el producto. A partir de preguntarle a esa persona cuáles son los números más bajos y más

Gráfico 1.13: Gráfico de una distribución triangular



Fuente: elaboración propia.

altos de clientes que el cree que puede haber, es suficiente para dibujar un triángulo y estimar la distribución de probabilidad triangular a partir de esa información.

La representación de la acumulada de una distribución triangular se muestra en el Gráfico xx.

1.5.2.3. Distribución Normal

Definición 23 (Distribución Normal).
Una variable es normal con parámetros μ y σ^2 si su densidad es:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \cdot exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)$$

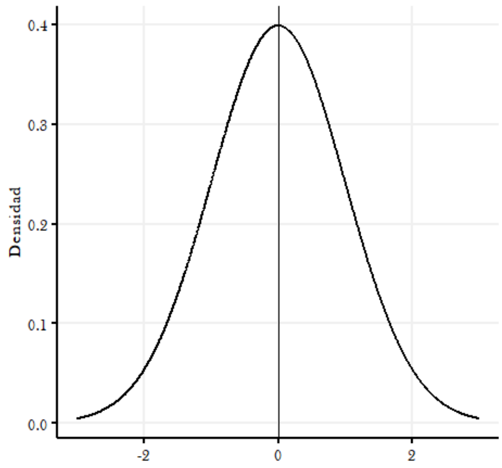
También se le nombra como distribución *gaussiana*, en honor a uno de los primeros matemáticos en describirla, Carl Gauss. Su gráfico tiene una forma de campana (Gráfico 1.14).

Muchos fenómenos aleatorios siguen aproximadamente una función de distribución de probabilidad normal (e.g. la estatura de los seres humanos). En el Gráfico 1.15 se muestra la distribución de probabilidad de la estatura de una muestra de estudiantes en la que se aprecia su forma similar a una Normal, excepto por la bimodalidad natural resultante de tener hombres y mujeres en la muestra.

Otras variables se distribuyen normal por agregación de un conjunto de variables independientes con el teorema del límite central.[15]

[15]Este teorema afirma que la suma de un número de variables aleatorias independientes e idénticamente

Gráfico 1.14: FUNCIÓN DE DENSIDAD NORMAL: $\mu = 0, \sigma = 1$



Fuente: elaboración propia.

Definición 24 (Distribución Normal general).
Si X se distribuye como una normal con parámetros μ y σ^2, entonces $Y = aX + b$ se distribuye normal con parámetros $\alpha\mu + b$ y $a^2\sigma^2$.

Este resultado implica que $Z = \frac{X-\mu}{\sigma}$ se distribuirá normal con parámetros 0 y 1. A la nueva variable Z se le conoce como *variable normal estándar* y se representa como en el Gráfico 1.16.

Una variable Binomial con un número de ensayos, n, muy grande, tiende a distribuirse como una Normal. Esta normal comparte la media y la varianza de la Binomial original. Los valores de los parámetros de ambas distribuciones que hacen que esta aproximación sea válida, son: $np \geq 5$ y $nq \geq 5$.

Ejercicio 21 (Distribución Normal).
Sea X una variable normal con parámetros $\mu = 3$ y $\sigma^2 = 9$, halle: $P\{2 < X < 5\}$

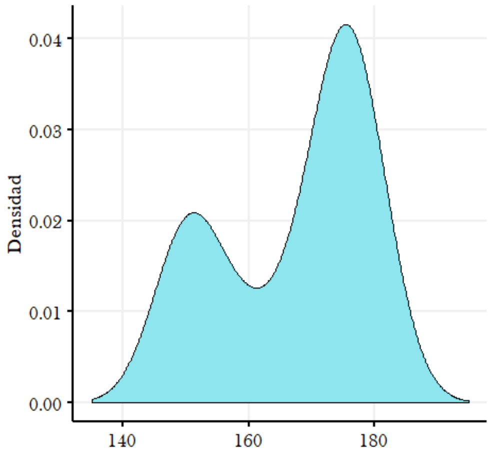
1.5.2.4. Distribución Chi Cuadrado

Definición 25.
Es cuando la probabilidad de una variable sigue la siguiente ecuación: $\frac{1}{2^{k/2}\Gamma(k/2)} x^{k/2-1} exp(-x/2)$. Donde Γ denota la función de distribución gamma.[16]

distribuidas con varianzas finitas tenderá a ser Normal a medida que crece el número de variables sumadas.

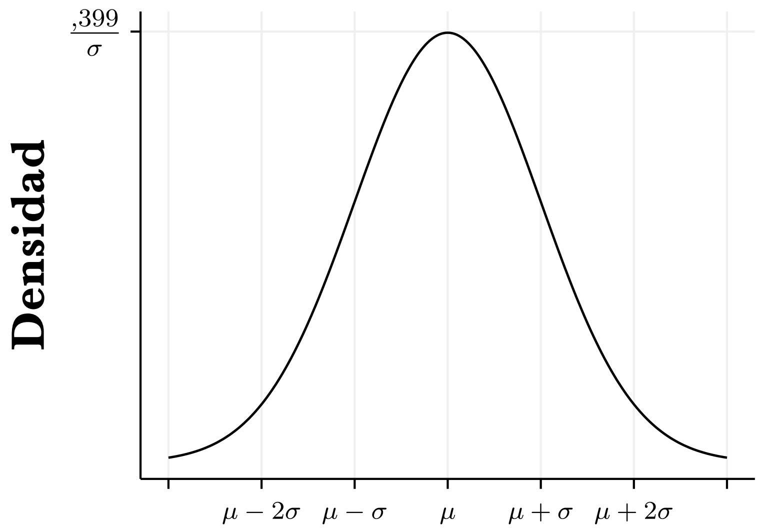
[16] Una distribución Gamma es la madre de varias distribuciones de probabilidad (i.e. Erlang, Exponencial, Chi Cuadrado), y es útil para trabajar con tasas (e.g. el tiempo de llegada de una llamada telefónica a un call center).

Gráfico 1.15: Densidad de distribución de la estatura de estudiantes, 2019



Nota: la muestra fue de 30 estudiantes de la Facultad de Economía y Negocios de la UTB en el segundo semestre de 2019.
Fuente: trabajo de investigación de Julián Caro Ortega.

Gráfico 1.16: Distribución normal para valores arbitrarios de μ y σ

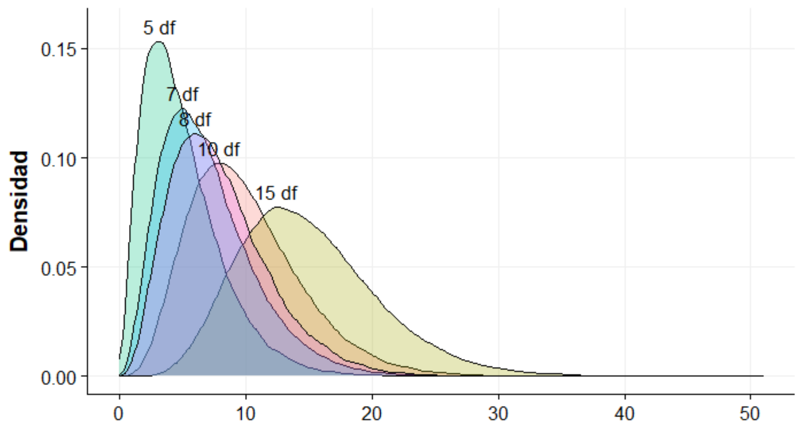


Fuente: elaboración propia.

Una distribución Chi Cuadrado resulta tras sumar los cuadrados de k variables independientes Normales estándar. Es muy usada para probar hipótesis en estadística inferencial y para representar la varianza muestral (s^2).

El gráfico de densidad de una Chi Cuadrado es siempre positivo y con asimetría positiva (i.e. hacia la derecha), aunque se torna más simétrica a medida que los grados de libertad suben (Gráfico 1.17).

Gráfico 1.17: Distribución Chi Cuadrado para diferentes grados de libertad



Fuente: elaboración propia.

Con esto concluímos el repaso de herramientas básicas de probabilidad y estadística. En últimas, los temas que hemos incluido son estándar en cualquier curso de teoría de la decisión para escuelas de negocios, y no constituyen sino una forma de los que Pierre Laplace llamaba "sentido común reducido a cálculos".

Capítulo 2

Aplicaciones a los negocios: árboles de decisión y teorema de Bayes

Roberto Fortich Mesa

2.1. Introducción

Para una buena planeación de los negocios es fundamental usar de de forma sistemática y científica información del desempeño pasado, propio y del entorno. Aunque en negocios de pequeña escala la intuición y el olfato para los negocios pueden bastar para planear hacia el futuro, a partir de un tamaño mayor es imprescindible reemplazar las heurísticas del gerente por la percepción profunda que ofrece la ciencia de datos.

Las competencias para esta exploración reciben el nombre de *business analytics* y una de sus áreas más sobresalientes es la descomposición de los gustos y hábitos de los consumidores en aplicaciones basadas en internet o redes sociales, porque es un ambiente en donde la riqueza de los datos es fácilmente convertible en estrategias de diseño de productos. Esta área de los negocios moderna aún no está siendo aprovechada por las empresas de Cartagena, quienes, en opinión de algunos, "poco saben usar inteligentemente los datos" (*El futuro de la información* 2018). Las personas encargadas de llevar procesos de *business analytics* en las empresas ocupan nuevos y modernos cargos denominados *management analyst* o *data analyst* y están siendo demandados con avidez por el sector privado.

En este capítulo se presentarán las herramientas de árboles de decisión y de teorema de Bayes, enfatizando su aplicación a los negocios, y recalcando su relación con términos y definiciones del Capítulo 1, como la probabilidad condicional y la probabilidad total.

2.2. Árbol de decisión

Cuando las empresas deben tomar decisiones secuenciales con incertidumbre pueden valerse de una herramienta gráfica desarrollada por expertos de investigación de operaciones en los 1960s llamada *árboles de decisión.* (Magee 1964)[1]

La principal ayuda de un árbol de decisión es la simplificación de escenarios que son complejos, pero que al representarlos con esta herramienta se tornan mucho más manejables. El ser humano no puede realizar cálculos matemáticos ni comparaciones cuantitativas sobre docenas de variables simultáneamente,[2] por lo que su representación en un árbol de decisión lo que hace es descomponer en partes más manejables de interpretar pero sin perder de vista la totalidad del problema sobre el que se debe decidir.

Definición 26 (Árbol de decisión).
Es un gráfico de un problema de inversión que facilita dilucidar las decisiones cuando están influidas por intricadas incertidumbres.

Representa en orden cronológico, de izquierda a derecha, las decisiones que se toman, y los resultados de eventos aleatorios que pueden ocurrir en el camino. Su valor no se limita a ser una herramienta auxiliar de planeación de un proyecto que un gerente realice de forma privada en el pizarrón de su oficina para ayudarse a sí mismo. Por el contrario, un árbol de decisión provee el lenguaje para que un equipo de personas asociadas con la toma de decisiones se comunique efectivamente. Pone sobre la mesa los posibles sesgos y visiones sobre los supuestos y la incertidumbre de un proyecto, añadiendo transparencia al proceso mismo de toma de decisiones.

La elaboración de los árboles de decisión tiene tres etapas principales: planteamiento, refinamiento, y solución.

2.2.1. Planteamiento

La dirección de una empresa depende fundamentalmente de la calidad de sus decisiones de inversión. Tradicionalmente estas decisiones son informadas por profesionales en finanzas a partir de indicadores como el valor presente neto. Sin embargo, en presencia de incertidumbre, la gerencia puede exigir a su equipo financiero que produzca pronósticos de la tendencia media de las principales variable aleatorias, o puede decidir usar su intuición personal para omitir escenarios que considera improbables, o puede incorporar la incertidumbre en el precio de sus productos o en la tasa de interés de sus proyectos.

[1] No confundir con la herramienta del mismo nombre usada en el campo del *machine learning.*

[2] Por ejemplo, es muy difícil memorizar más de nueve dígitos porque nuestra memoria inmediata está diseñada para trabajar con menores dimensiones.

En todo caso, cualesquiera que haya sido las medidas de planeación adoptadas, el equipo de dirección probablemente habrá acopiado una alta cantidad de datos que requiere procesar y resumir en formas convenientes, flexibles, ágiles e informativas para la toma de decisiones. Con el árbol de decisión, se requiere que haya entonces un planteamiento que incorpore la esencia de muchos de los riesgos, objetivos, beneficios económicos e información del proyecto de inversión. De esta preselección dependerá el que la herramienta funcione para dilucidar los mejores cursos de acción.

Para plantear compactamente los árboles de decisión, el gerente debe decidir qué elementos incluir, dado que no todas las minucias del proyecto serán relevantes para el planteamiento y, en cambio, lo que puede ocurrir al incluir cosas de más es convertir el problema en algo intratable. El analista siempre debe pensar en que lo que se incluye en el árbol deben ser decisiones y eventos que son importantes para el proyecto (i.e. que impliquen consecuencias que se deseen comparar).

Ninguno de los datos del proyecto que terminen dibujados en el árbol son novedosos porque son datos ya conocidos por la gerencia. Sin embargo, el disponerlos ordenamente dentro del árbol sirve para aplicar criterios lógicos de forma sistemática, lo que con los datos aislados e independientes hubiese sido muy difícil de implementar.

Los árboles de decisión son una serie de nodos y ramas.[3] Cada rama representa un curso de acción alternativo que puede desembocar en una decisión o en un nodo aleatorio. Al completarse cada curso de acción, es decir, en la parte derecha del árbol, se llega siempre a un pago monetario.

Siguiendo la convención propuesta por Magee (1964), las acciones se indican con nodos cuadrados mientras que los eventos aleatorios se indican con nodos circulares (Gráfico 2.1). Estos eventos escapan al control de la empresa a cargo de tomar decisiones y, en cambio, están sujetos al azar o a circunstancias incontrolables. En la práctica, los árboles terminan siendo eslabonamientos en los que se mezclan nodos de decisión y aleatorios y suele haber entonces intercalamientos entre ambos tipos.

En los árboles de decisión pueden existir hasta cuatro tipos de nodos:

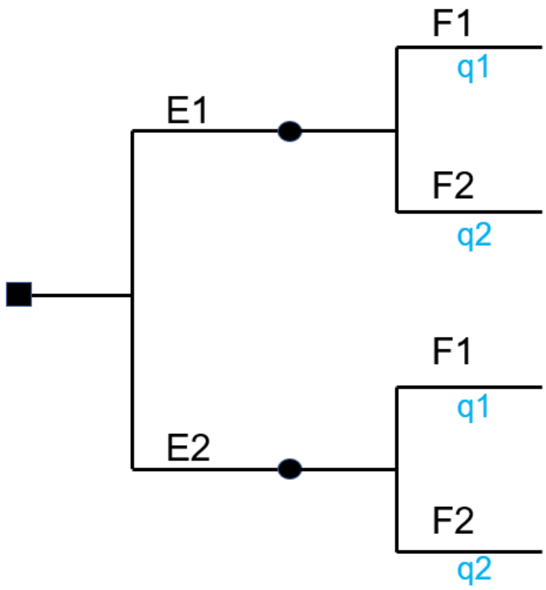
1. De decisión: se representan mediante un cuadrado y precede alternativas que no contienen probabilidades

2. Aleatorios: se representan mediante círculos y precede alternativas con probabilidades.

3. Inicial: es aquel nodo que no está precedido de ningún otro nodo

[3]Este nombre es un símil por su parecido con las bifurcaciones del tallo de un árbol.

Gráfico 2.1: Ejemplo de un árbol de decisión



Nota: En azul se muestran las probabilidades de las ramas originadas en nodos aleatorios.
Fuente: elaboración propia.

4. Terminales: son aquellos nodos que no están seguidos de ningún nodo subsiguiente. Algunos autores insertan triángulos al final de las ramas que se desprenden de estos nodos.

Todo árbol de decisión contiene hasta cuatro tipos de componentes. En primer lugar, están las *acciones*, que representan cursos de acción disponibles para quien decide. Se originan en nodos de decisión y se bifurcan según el número de posibles alternativas de elección. Los árboles de decisión siempre comienzan con un nodo de decisión. Las acciones del Gráfico 2.2 se muestran en rojo.

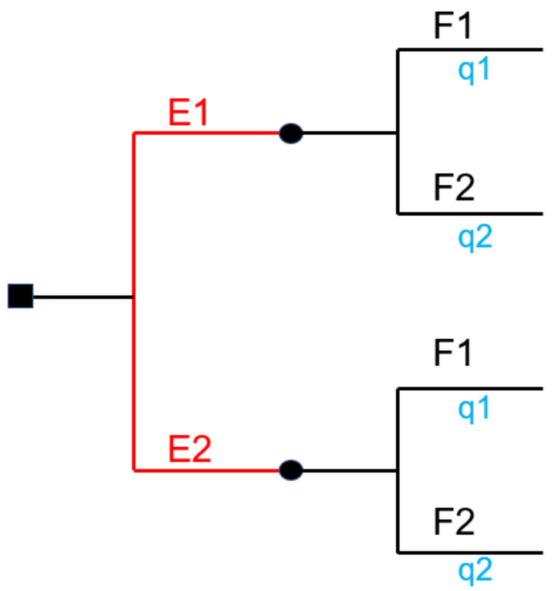
Segundo, las *ramas aleatorias*, que muestran los estados de la naturaleza bien definidos[4] que se originan en un nodo aleatorio y su probabilidad de ocurrencia. Esta probabilidad es, por construcción, condicional a que se haya llegado a ese punto del árbol. Las ramas aleatorias se muestran en rojo en un árbol de ejemplo en el Gráfico 2.3.

Tercero, los *valores económicos*, que pueden ser negativos (costos) o positivos (ingresos), y que se dibujan en la parte inferior de las ramas según la etapa del árbol en la que ocurren (Gráfico 2.4).

Finalmente, están los *valores económicos netos*, que sintetizan todo el recorrido desde el nodo inicial hasta el terminal en una sola cantidad.

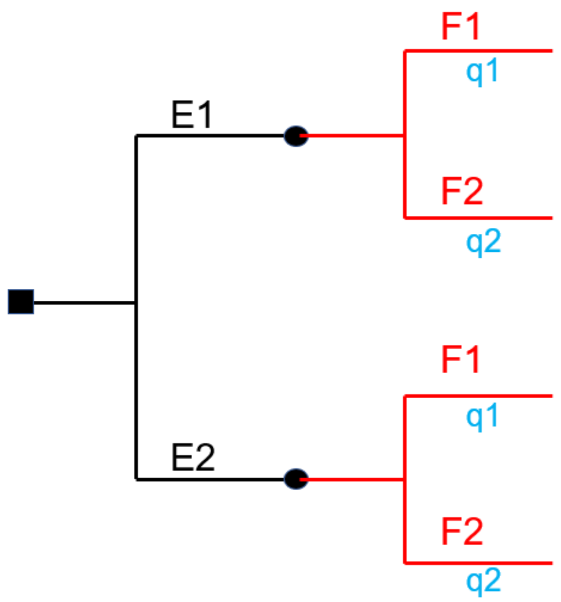
[4]Colectivamente exhaustivos y mutuamente excluyentes.

Gráfico 2.2: Acciones $E1$ y $E2$ del árbol de decisión de ejemplo



Fuente: elaboración propia.

Gráfico 2.3: Ramas aleatorias $F1$ y $F2$ del árbol de decisión de ejemplo



Fuente: elaboración propia.

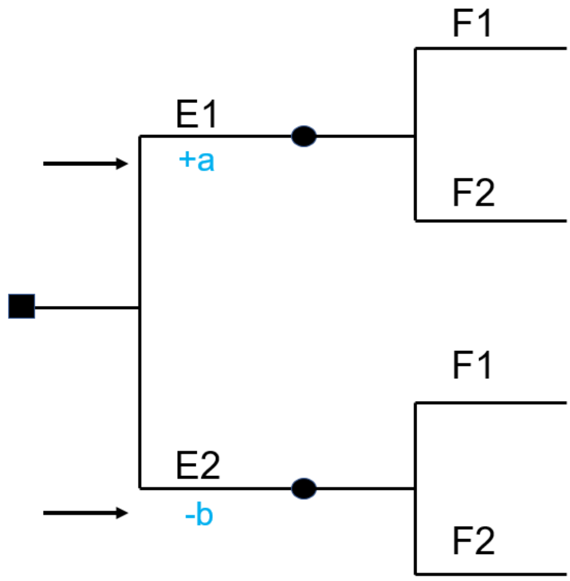
2.2.2. Refinamiento

Una vez planteado el árbol e incorporados los valores económicos a lo largo de sus ramas, se debe escribir en los nodos terminales, arriba y entre corchetes,[5] el valor económico neto (ingresos - egresos a lo largo del árbol).

Debe revisarse si existe un curso de acción que domine el resto, en el sentido de que presente mayores valores económicos netos independientemente del valor que tomen los

[5]La convención de usar corchetes es tomada de Castillo (2006).

Gráfico 2.4: Valores económicos asociados a las acciones de un árbol de ejemplo



Nota: a es un beneficio incurrido al elegir $E1$ y b es un costo incurrido al elegir $E2$.
Fuente: elaboración propia.

estados de la naturaleza. En ese raro caso, sugiero revisar que el árbol esté correctamente planteado, y esforzarse por "enlodar" el problema, en el sentido de incluir más datos relevantes que le introduzcan mayor riqueza y complejidad.

En el otro extremo, si el arbol de decisión queda muy "plano" en el primer ensayo de planteamiento, quiere decir que no existe una diferenciación muy marcada entre los valores económicos o las probabilidades asignadas a los estados de la naturaleza. Si esto es fiel reflejo de que la decisión de inversión tiene poca variabilidad, no es necesario cambiar nada. Pero con frecuencia, puede ser necesario recalibrar las probabilidades asignadas a los estados de la naturaleza para introducir al árbol la dispersión en su magnitud correcta.

Finalmente, sugiero revisar la extensión, medida en cantidad de etapas, elegida para el árbol, que incide en su tamaño horizontal. Esta preocupación no es estética sino funcional. En un árbol que tenga un horizonte demasiado amplio puede ser necesario ajustar los valores económicos para llevarlos todos a un tiempo común que los haga comparables. Los descuentos financieros más comunes son traer todo al presente o llevar todo a valor futuro. Si se elige este refinamiento, debe elegirse cuidadosamente la tasa de interés, y las técnicas de análisis de sensibilidad del Capítulo 3 pueden ayudar a garantizar que la incertidumbre sobre el valor de esta tasa no incida de forma muy crítica en cambiar el curso de acción. Usualmente, esa hipersensbilidad debe ser señal de alerta de que el modelo es poco confiable.

2.2.3. Solución

Solucionar el árbol es calcular la secuencia de decisiones que maximiza los pagos monetarios de quien toma las decisiones. Es similar al algoritmo de *inducción hacia atrás* de los cursos de teoría de juegos. Funciona gracias a que un nodo de varias ramas siempre se puede hacer equivalente a uno de una sola rama.

El algoritmo consiste en comenzar por la derecha del árbol, en los nodos terminales, y de ahí seguir los siguientes pasos. Primero, si el nodo terminal por el que se comienza es un nodo probabilístico, se calcula la media.[6] Si no es un nodo probabilístico, se marca la alternativa con el mayor pago.

Para simplificar el árbol, se escribe encima de cada nodo el valor de los pasos anteriores, y se repiten los pasos hasta llegar al nodo raíz. La alternativa óptima de decisión es la que, a partir del nodo raiz sigue las alternativas marcadas como mejores en cada nodo de decisión.[7]

El curso de acción óptimo es: construir la planta de caramelización. La media de valores económicos netos de hacer esto es de 144 (US $ 144 M). El procedimiento de inducción hacia atrás nos llevó a reconocer que esta era la ruta óptima.

Ejercicio 22 (Árbol de decisión).
Una empresa de dulces desea lanzar un chocolate nuevo y debe decir entre los procesos de producción de caramelización y endulzado natural. Si se elige la caramelización se debe pagar una fábrica que cuesta 30 mientras que si se elige endulzado natural se debe adquirir un contrato con el proveedor de azúcar artificial por valor de 4. La demanda por el chocolate está proyectada como baja con un 20 % de probabilidad y alta con un 80 %.

El Cuadro 2.1 muestra los beneficios que conseguirá la empresa de dulces en cada estado de la naturaleza para cada proceso productivo, en millones de dólares.

Cuadro 2.1: BENEFICIOS DE LA EMPRESA DE DULCE SEGÚN SUS DECISIONES Y DEMANDA

	Demanda	
Proceso	Baja	Alta
Caramelización	70	200
Natural	100	160

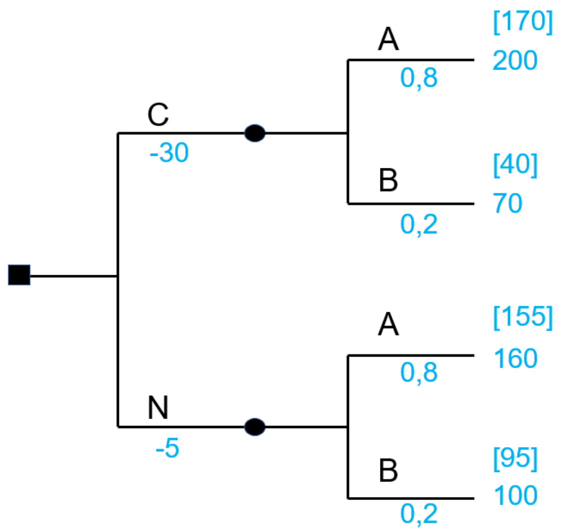
Represente la decisión de la empresa de dulces de cuál proceso productivo elegir con un árbol de decisión y halle la ruta óptima de solución.

[6] Esta es una media ponderada, y se calcula como se explicó en el Capítulo 1.

[7] En Hillier y Lieberman (2001, p. 767) se recomienda insertar una doble barra diagonal como una barrera atravesando cada rama rechazada.

El árbol que representa el problema de la empresa de dulces está en el Gráfico 2.5.

Gráfico 2.5: Árbol de decisión del Ejercicio 22



Fuente: elaboración propia.

Para hallar la ruta óptima nos ubicamos en el nodo terminal superior y hallamos la media, que es igual a 144. Repetimos el cálculo de la media para el nodo terminal inferior, encontrando un valor igual a 143. El siguiente paso consiste en decidir entre el mayor valor entre C y N, que corresponde al 144 de C.

Este procedimiento de solución, aunque consistió de movernos de derecha a izquierda en el árbol, en realidad debe ir seguido de un análisis posterior en el que se discutan las mejores opciones de izquierda a derecha, es decir, en el orden natural en el que se toman las decisiones.

Ejercicio 23 (Árbol de decisión).
La empresa Petróleos del Caribe posee un lote que puede contener petróleo. Un geólogo consultado por la empresa estimó que la probabilidad de hallar un yacimiento en el terreno es de 26 %. Dada la buena fortuna con la cuenta el terreno, otra empresa ha ofrecido comprarlo por $ 90 m. Sin embargo, Petróleos del Caribe está considerando usufructuar el lote ella misma, perforándolo para extraer el mineral. El costo de perforar es de $ 100 m. En caso de confirmarse que sí hay petróleo, se esperan ganancias de $ 800 m. Plantee y solucione el árbol de decisión de esta situación.

2.3. Teorema de Bayes

El teorema de Bayes[8] es un teorema de la probabilidad curiosamente rodeado de misterio. Para algunos tiene una alta efectividad como herramienta que mejora la estimación de una

[8]Bayes en inglés se pronuncia "beis" https://www.youtube.com/watch?v=PwyNPc8-uGw.

probabilidad, y para otros, es difícil de entender cuál su real valor. Se pueden implementar análisis bayesianos para aumentar la precisión de las probabilidades condicionales. Otros lo utilizan para otorgar propiedades cualitativas a las probabilidades (i.e. para modelar un proceso de aprendizaje), y en todo caso, sirve para transformar probabilidades gruesas y subjetivas que se tienen a priori en valores mas refinados.

Es como si el teorema de Bayes fuera entonces un híbrido entre la estadística descriptiva y la inferencial. Añade atractivo para esta ecuación que sigue siendo sorprendentemente vigente pese a haber sido desarrollado hace siglos (en el siglo XVIII) por el reverendo Thomas Bayes.

Mi visión sobre el alcance del teorema de Bayes es más conservador. Por un lado, en la aplicación a los negocios que introduciré, el teorema sí produce una ganancia de precisión en la probabilidad de que la demanda por un producto sea alta, usando la proyección de una empresa de consultoría y el conocimiento de su historial de aciertos y desatinos.

Pero por otro lado, desmitificaré el teorema al explicar que no es muy diferente de una probabilidad condicional simple y llana. En realidad es sólo una extensión pequeña del caso del Capítulo 1 de probabilidades múltiples secuenciales, en el que se aplica una probabilidad condicional con el denominador representando una probabilidad total. Su motivación vendrá dada en que, a diferencia de la probabilidad condicional ordinaria, la secuencia cronológica de los eventos está invertida, es decir, se condiciona un evento presente a la ocurrencia de uno futuro.

Definición 27 (Teorema de Bayes).

$$P(E|F) = \frac{P(F|E) \cdot P(E)}{P(F|E) \cdot P(E) + P(F|E^{\complement}) \cdot P(E^{\complement})}$$

Ejercicio 24 (Árbol de decisión con teorema de Bayes).
Suponga que la empresa de dulces del Ejercicio 22 desea consultar con la UTB sobre cuál será la demanda por su producto. La UTB cobra honorarios de 35 por la consultoría.

La empresa de dulces busca el curso de acción óptimo, que incluye consultar o no con la UTB, y decidir cuál proceso productivo implementar.

La UTB estima que la demanda es baja o alta con alguna probabilidad. Los archivos de los resultados históricos de las estimaciones de la UTB muestran la tasa de aciertos (Cuadro 2.2).

Por ejemplo, el 20 % de las veces que la demanda fue baja, la UTB había estimado que sería alta.

Cuadro 2.2: Resultado histórico de las proyecciones de demanda de la UTB

	Demanda real	
Estimación UTB	Baja	Alta
Baja	0,8	0,3
Alta	0,2	0,7

El principal ejemplo de invertir un árbol de decisión lo tenemos con este lanzamiento de un producto con demanda incierta. A partir del historial de contrataciones de la empresa de estudios de mercado, se hace un árbol en dos etapas (Gráfico 2.6). Primero ocurre el estado de la naturaleza de demanda alta o baja.[9] Luego ocurre el estado de la naturaleza de acierto o yerro del estudio de mercado.

Una vez planteada la intervención del consultor en el problema de decisión original, vemos que equivale a realizar el experimento de ensayar nueva información. Los ejercicio de árboles de decisión pueden tomar dos caminos respecto a las interveneciones de consultores o, en general, a experimentos de introducir nueva información: primero, se puede calcular su valor potencial. Y segundo, se puede resolver el árbol para hallar la ruta óptima.

En estas notas nos enfocaremos en la solución del árbol. Para invertir este árbol auxiliar se calculan las probabilidades de la primera etapa. $P(\text{UTB alta})$ se halla computando su probabilidad total en el árbol sin invertir.[10] De igual modo se calcula $P(\text{UTB baja})$. Al pasar a la segunda etapa del árbol invertido, las probabilidades se hallan mediante el teorema de Bayes. Por ejemplo, $P(\text{Real alta}|\text{UTB alta})$, se calcula dividiendo la probabilidad conjunta de ambos eventos sobre la probabilidad total de *UTB alta*. A estas probabilidades de la segunda etapa del árbol invertido se les llama *a posteriori*.

Del árbol invertido obtenemos las probabilidad que incluiremos en el árbol definitivo del problema completo, como se muestra en la rama superior del Gráfico 2.8.

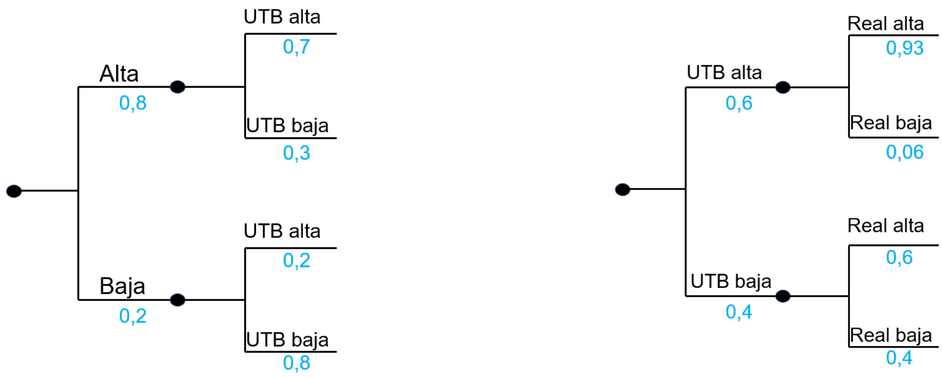
La interpretación de la "magia" del teorema de Bayes es como sigue. Condicionado a que la UTB estimó que la demanda sería baja, el teorema mejoró en 20 puntos porcentuales la precisión de la probabilidad de que la demanda fuera baja, que a priori era igual a 20 % y a posteriori fue de 40 %. Si la empresa de dulces no hubiese "actualizado" su opinión sobre la demanda habría subestimado significativmente la posibilidad de una baja demanda.

[9] A las probabilidades de este nodo aleatorio se les llama *a priori* en la literatura Bayesiana.

[10] El procedimiento para hallar una probabilidad total para eventos múltiples secuenciales se explicó en el Capítulo 1.

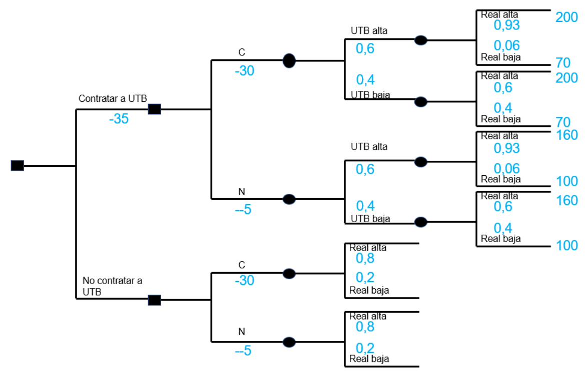
Gráfico 2.6: Árboles auxiliares del teorema de Bayes para el Ejercicio 24

(a) Árbol sin invertir (b) Árbol invertido



Fuente: elaboración propia.

Gráfico 2.8: Árbol de decisión del Ejercicio 24



Fuente: elaboración propia.

2.4. Extensión: funciones de utilidad

Los valores económicos que hemos usado hasta aquí en los árboles de decisión son pagos monetarios. Siguiendo la teoría de la microeconomía, sin embargo, las personas puede que tomen decisiones en la escala de la *utilidad*. Típicamente estas son funciones cóncavas que transforman una cantidad de dinero X en una cantidad de utilidad $u(X)$. La concavidad de la función refleja la aversión al riesgo de la persona.[11]

Una vez se ha elegido una función de utilidad para el empresario, se debe incorporar al árbol de decisión. El procedimiento de solución es idéntico al caso expuesto con valores económicos como pagos monetarios, excepto porque ahora estos se sustituirán por su equivalente en la escala de utilidad. Debo advertir que, según Hillier y Lieberman (2001, p. 773), muchos tomadores de decisión no están lo suficientemente cómodos con la noción de utilidad en el sentido microeconómico, de modo que es probable que esta aplicación caiga en desuso.

[11]Aunque también se pueden modelar personas con funciones lineales o convexas, pero representaría actitudes de neutralidad y propensión al riesgo.

En este capítulo mostramos que los árboles de decisión y el teorema de Bayes son herramientas que se usan conjuntamente para modelar decisiones secuenciales de inversión en presencia de incertidumbre. La exposición usó la didáctica de extender el teorema de Bayes como una versión ligeramente diferente de la probabilidad condicional que se mostró en el Capítulo 1 y su integración al árbol de decisión se hizo mediante la inversión de un árbol auxiliar que permitió refinar las probabilidades con las que una empresa de consultoría acertará en sus predicciones. Con el conocimiento de estas dos técnicas el estudiante adquiere la competencia de modelar y comunicar de forma simple escenarios de inversión con complejidad en las dimensiones de actualización de probabilidades subjetivas y de intercalamiento de múltiples decisiones con eventos aleatorios.

Capítulo 3

Simulaciones en hojas de cálculo

Roberto Fortich Mesa, José Vergara Arrieta y Pedro Castilla Ávila

Anque los analistas financieros suelen expresar la información de los proyectos de inversión con ecuaciones matemáticas, o con softwares contables, financieros y estadísticos, el expresarlos en hojas de cálculo es uno de los medios más populares. El uso de hojas de cálculo en las finanzas tiene un largo historial.[1]

La ventaja principal de las hojas de cálculo es su abundante disponiblidad, porque virtualmente todo negocio con equipos de computación personal tiene instalado alguna versión de ellas (LeBlanc y Grossman 2008; *Google apps vs. Office 365: Which suite reigns supreme?* 2016). Una ventaja secundaria es que su curva de aprendizaje es sencilla y permite a un estudiante primerizo enfrentarse de forma "sucia y rápida" con un modelo financiero.

La conveniencia de las hojas de cálculo se concentra entonces en el campo de los profesionales de las finanzas que se desempeñan en la industria, dado que los investigadores suelen preferir realizar sus simulaciones con softwares estadísticos especializados tales como R, como se explicará en detalle en el Capítulo 4 sobre simulaciones de Monte Carlo. Los profesores pueden enseñar simulaciones en hojas de cálculo como puerta de entrada a hacerlo en lenguajes de programación con menos énfasis en lo visual y más acento en la eficiencia y reproducibilidad.

En síntesis, simular usando el lenguaje de las hojas de cálculo aunque no es óptimo, facilita mostrar simultáneamente los datos proyectados de múltiples variables financieras, así como aplicarles algoritmos, funciones y fórmulas, con el método de arrastre horizontal o vertical.

[1] Los primeros antecedentes son Liao (1990) y Seila y Banks (1990). Un pionero en América Latina de proponer la enseñanza de las finanzas con hojas de cálculo es Vélez, cuyo texto más centrado en el riesgo y la incertidumbre es Vélez (2003).

En este capítulo explicaremos las técnicas de análisis de sensibilidad con cambios diminutos, tablas de una y dos variables, y buscar objetivo. La versión de hoja de cálculo que elegimos para ejemplificar estas técnicas fue Microsoft Excel para Office 365.[2]

3.1. Análisis de sensibilidad

El análisis de sensibilidad mide cómo el cambio en una variable afecta un resultado de un modelo. Aunque tiene aplicaciones en cualquier contexto (i.e. modelos científicos, climáticos, sociológicos, epidemiológicos) en este capítulo abordaremos su implementación a ejemplos financieros.

En finanzas uno de los principales usos de las modelos es para documentar una inversión de capital. En esta aplicación, el objetivo del modelo es describir las interacciones en un sistema usando ecuaciones que incluyen variables de entrada, algoritmos,[3] y variables de salida. Las variables de entrada para una empresa industrial pueden incluir, por ejemplo, el costos de los equipos, la fuerza laboral, las actividades de planeación y análisis, entre otras. Las variables de salida de mayor interés suelen ser las de criterios de decisión (i.e. valor presente neto, tasa interna de retorno).

Un análisis de sensibilidad busca las variables más críticas dados diferentes supuestos, por ejemplo, al ensayar con diferentes precios de venta para ver qué ocurre con los resultados financieros de una empresa comercial. Otro ejemplo, es ensayar que pasaría con el desempeño de la empresa tras ampliar alguna línea de sus productos. Como los modelos en hojas de cálculo se entrelazan con fórmulas de actualización automática, basta con introducir un cambio de, digamos, 5 % en una variable de entrada, para inmediatamente determinar qué efecto produce sobre alguna variable de resultado financiera en otra celda de la hoja.

La diferencia entre el Valor Presente Neto (VPN) luego de cambiar un parámetro con respecto a su valor original –en el caso base– es la *sensibilidad*. Esta diferencia es de las medidas de variabilidad del resultado más simples de implementar e interpretar, por lo que será la que expondremos en este capítulo. Otras medidas de variabilidad para los análisis de sensibilidad están descritas en Saltelli (2002).

Entonces la idea de la sensibilidad es medir qué tanto puede cambiar el problema respecto a la incertidumbre que se tiene de alguna variable. Si a consecuencia de modificar alguna variable el VPN reacciona volátilmente, entonces es señal de que si se permite que ese parámetro tenga un rango amplio hay riesgo de desestabilizar el modelo. En otras palabras,

[2] Los autores trabajamos con copias licenciadas de Excel.

[3] En el contexto de hojas de cálculo, un algoritmo en realidad es una ecuación de Excel escrita por el usuario en la que enlazan unas variables de entrada con una de salida.

tener muy alta sensibilidad respecto a una variable puede impedir predecir con certeza el valor del VPN.

Aunque muchas variables en un modelo financiero son aleatorias, con frecuencia se conocen estimativos de cuál es su rango de incertidumbre. Estos estimativos suelen provenir de múltiples fuentes tales como mediciones, opinión de expertos o analogías con otros procesos comparables.[4]

El análisis de sensibilidad en hojas de cálculo usa valores *ad hoc* de la grilla de la variable de entrada, a diferencia de las simulaciones de Monte Carlo del Capítulo 4, en las que se investiga metódicamente el rango completo de riesgo asociado con cada variable de entrada. Por eso en este capítulo las técnicas no se alimentan con distribuciones de probabilidad simuladas, sino que adoptan valores arbitrarios para las variables aleatorias. Un buen ejemplo en el que se adoptan simulaciones de Monte Carlo y generación de números aleatorios en hojas de cálculo es Liao (1990) y un popular complemento comercial que simula en Excel es *Crystal Ball*.

En este capítulo explicaremos dos tipos de variación a los modelos: cuando los cambios de los parámetros son diminutos (e.g.del 1 %), y cuando se establece una grilla de valores equidistantes. Los cambios diminutos son convenientes para tener una medida estandarizada que pueda ser comparada para cada una de las variables de entrada y facilita la generación de gráficos de tornado, en los que se ordenan de mayor a menor las variables críticas. La grilla de valores equidistantes permite explorar cómo reacciona una variable de resultado tras un mayor rango de valores de un parámetro en particular o, máximo, de dos parámetros (para este caso se requiere usar la técnica de *tablas de dos variables*).

3.1.1. Repaso del VPN

El VPN es uno de los principales indicadores disponibles dentro del menú de criterios de decisiones de inversión porque permite valorar proyectos de forma simple y rápida. Dado que el objetivo de las simulaciones es ayudar al gerente a tomar decisiones que beneficien a la empresa, conviene hacer un digresión en la que se repase el signficado del valor de la empresa y su medición con el VPN.

El valor de un proyecto se halla usualmente con el método del flujo de caja descontado. Requiere que los flujos de caja futuros se estimen y descuenten con el costo del capital para obtener así sus valores presentes. La teoría que soporta hallar el valor de un proyecto con este enfoque consiste en suponer que lo que genera valor para un inversionista es la expectativa de recibir utilidades en un futuro en combinación con el riesgo asociado con

[4]La incertidumbre no tiene por qué introducirse de forma individual, sino que también se puede extender al caso de distribuciones conjuntas de probabilidad que cobijen varias variables con incertidumbre intrínseca interdependientes.

el proyecto. Una corriente de flujos futuros entonces se puede traer al presente si se elige bien la tasa de descuento que refleja correctamente el riesgo del proyecto. Con los flujos presentes el VPN es simplemente el resultado de sopesar costos y beneficios en la escala nueva que, ahora sí, es comparable.

El VPN sirve para evaluar y comparar proyectos de inversión disímiles entre sí, así tengan flujos de caja desperdigados a lo largo del tiempo como, por ejemplo, cuando se tienen planificados préstamos, inversiones y amortizaciones que son afectadas de forma diferencial por el valor del dinero en el tiempo. El VPN resulta de sumar los valores descontados de los flujos de caja de los proyectos.

En la práctica, la interpretación de este indicador es: un VPN positivo resulta en un proyecto que merece ser ejecutado porque genera más utilidades para la empresa que la opción de inversión de reserva con riesgo equivalente, mientras que uno negativo refleja que no es viable invertir.

3.1.2. Cambios diminutos a los parámetros

El análisis de sensibilidad busca medir cómo reaccionan las variables de salida del modelo tras cambios de los parámetros. La razón de que al estandarizar la magnitud de los cambios se elija hacerlo con valores muy pequeños es por su conveniencia. Si los cambios introducidos fuesen significativamente grandes, sería fácil encontrar que sí hacen reaccionar a las variables de salida, lo que trivializaria el experimento. En cambio, estandarizarlos para que sean minúsculos a lo largo de todos los parámetros, implica que solamente resaltarán aquellos que terminen repercutiendo más sobre el modelo.

Los cambios se examinan a partir de un caso de referencia llamado *caso base*, que es el modelo determinístico en el que se predeterminan valores para todos los parámetros. Usualmente este caso se plantea con certeza o, de haber incertidumbre, supone que se producen los valores puntuales más probables de cada variable aleatoria (Raychaudhuri 2008, p. 91).

Ejercicio 25 (VPN con certeza).
Memorias Bolívar es una empresa que ha desarrollado la tecnología para fabricar un accesorio para computadores (e.g. discos duros) y su gerente financiero desea proyectar su estado de resultados financieros para ver sus perspectivas de rentabilidad de mediano y largo plazo. La empresa hizo una inversión inicial de contruir un edificio que costó US $ 10 millones (M) en el año cero. Las ventas anuales de Memorias Bolívar son de US $ 8,4 M y se producirán durante tres años.

Si la tasa de descuento es igual a 0,07 (en notación de propoción, que equivale a 7 % en porcentaje) halle el VPN de este proyecto cuando se tiene certeza de todas sus variables.

Los flujos de caja son: -US$ 10 M en el año cero, +US$ 8,4 M en el año uno, +US $ 8,4 M en el año dos, y +US $ 8,4 M en el año tres.

Para calcular los valores presentes, traemos al presente los flujos de caja futuros a una tasa de descuento de 7 %.[5]

$VP_t = \frac{C_t}{1+i}$ con $t \in 1, 2, 3$

De este modo el VPN, con el método de traer los flujos al año cero, resulta de sumar[6] la inversión inicial con los valores presentes de los flujos de caja, así:

$VPN = -US\$10M + US\$7,85M + US\$7,34M + US\$6,86M = US\$12,05M$

El VPN del proyecto es positivo, lo que significa que es rentable realizarlo.

Ahora mostraremos los cálculos de este VPN usando una hoja de cálculo, con pantallazos de la secuencia de pasos en Excel (Gráficos 3.1, y 3.2).

Gráfico 3.1: PLANTEAMIENTO DEL EJEMPLO DE CÁLCULO DEL VPN EN UNA HOJA DE CÁLCULO

	A	B	C
1	**Tasa de descuento=**	0,07	
2	**Periodos**	**Flujos de caja**	**Valor Presente**
3	0	-10,00 M	-10,00 M
4	1	8,40 M	7,85 M
5	2	8,40 M	7,34 M
6	3	8,40 M	6,86 M
7			12,05 M
8		Otra alternativa	12,04 M
9			

Fuente: elaboración propia

Ejercicio 26 (Análisis de sensibilidad con cambios diminutos).
Suponga ahora que la empresa del Ejercicio 25 tiene datos del precio de su producto en el año 1, del volumen de ventas, y de la tasa de variación del precio. También suponga que ahora la duración del proyecto es de cuatro años y que la tasa de descuento es de 10 %. El resumen de los datos del proyecto con certeza se muestra en el Cuadro 3.1.

Para obtener el valor del proyecto se calcula el flujo de caja en cada año y se halla el VPN. Las variables de ingreso y salida del modelo deben quedar escritas en la hoja de cálculo.

[5] Traer a valor presente también se conoce como *descontar*, y a la tasa de descuento también se le llama *costo de oportunidad del capital* o tasa de interés.

[6] Suma algebraica, implica que los valores negativos se restan

Gráfico 3.2: Fórmulas del ejemplo de cálculo del VPN en una hoja de cálculo

	A	B	C
1	**Tasa de descuento=**	0,07	
2	**Periodos**	**Flujos de caja**	**Valor Presente**
3	0	-10	=+VA($B$1;A3;;-B3)
4	1	8,4	=REDONDEAR(VA($B$1;A4;;-B4);2)
5	2	=+B4	=REDONDEAR(VA($B$1;A5;;-B5);2)
6	3	=+B5	=REDONDEAR(VA($B$1;A6;;-B6);2)
7			=+SUMA(C3:C6)
8		Otra alternativa	=REDONDEAR(VNA(B1;B4:B6)+B3;2)

Nota: Usamos la fórmula *Redondear* (para el caso, configurada a dos decimales) para asegurar que el resultado sea exactamente igual al de la ecuación del VPN. A pesar de este recurso, al escribir la fórmula alternativa *VNA* (celda C8), también con redondeo a dos decimales, el resultado es ligeramente diferente. Esta discrepancia por redondeo no debe preocupar al lector.
Fuente: elaboración propia

Cuadro 3.1: Variables típicas de un Flujo de caja para el caso base de Memorias Bolívar, en dólares

Variable	Valor
Futuro cambio de precio	-2,5 % anual
Precio (año 1)	120
Volumen de ventas (todos los años)	70 m
Inversión inicial	10 M
Tasa de descuento	10 %

El valor de las ventas se calcula en cada año con el precio proyectado según su futuro cambio de precio (Cuadro 3.2). En este ejemplo, por tratarse de un producto de tecnología, se proyecta que su precio decrece anualmente. El valor presente se obtiene dividiendo cada valor de la fila “Flujo de caja” por el factor de descuento, $(1 + i)^n$, donde n es el número de periodos entre el periodo del flujo de caja y el año cero.

Cuadro 3.2: Proyección del precio del caso base, en dólares

	Año				
Variable	0	1	2	3	4
Precio		120	117	114	111
Flujo de caja	-10 M	8,4 M	8,19 M	7,98 M	7,78 M
Valor presente	-10 M	7,63 M	6,76 M	6 M	5,31 M
VPN = 15,72 M					

El VPN del caso base sale de sumar los valores de la fila “Valor presente” del Cuadro 3.2, y es igual a US$ 15,72 M.

Cuadro 3.3: PROYECCIÓN DEL PRECIO TRAS UN CAMBIO DIMINUTO AL PRECIO INICIAL, EN DÓLARES

	Año				
	0	1	2	3	4
Precio		121	118	115	112
Flujo de caja	-10 M	8,48 M	8,27 M	8,06 M	7,8 M
Valor presente	-10 M	8,4 M	7,44 M	6,6 M	5,85 M
VPN = 15,98 M					

Impongamos un cambio en el precio inicial para estudiar cómo reaccionará el VPN de la empresa. Introduciremos un cambio muy pequeño al primer parámetro: el precio inicial[7]. Aumentar en 1 % el precio inicial, es decir, cambiarlo de su valor de 120 en el caso base a uno de 121,2, provoca que los flujos de caja del proyecto reaccionen aumentando según los valores mostrados en el Cuadro 3.3.

En últimas, el cambio al precio inicial desencadena un aumento de 1,65 % en el VPN del proyecto, como se muestra con el nuevo VPN de US $ 15,98 M del Cuadro 3.3. Esta es la sensibilidad del VPN ante cambios en el precio inicial.

Ahora mostraremos los cálculos de este análisis de sensibilidad usando una hoja de cálculo, con pantallazos de la secuencia de pasos en Excel (Gráficos 3.3, 3.4, 3.5, y 3.6).

Gráfico 3.3: PLANTEAMIENTO DEL EJEMPLO DE ANÁLISIS DE SENSIBILIDAD DEL VPN RESPECTO AL PRECIO EN UNA HOJA DE CÁLCULO

	A	B	C	D	E	F
1	**Variable**	**Valor**				
2	Futuro cambio de precio	-2,5%				
3	Precio (año 1)	120				
4	Volumen de ventas	70 m				
5	Inversión inicial	10,00 M				
6	Tasa de descuento	0,1				
7						
8		**Año**				
9		**0**	**1**	**2**	**3**	**4**
10	**Precio**		120,00	117,00	114,08	111,22
11	**Flujo de caja**	-10,00 M	8,40 M	8,19 M	7,99 M	7,79 M
12	**Valor presente**	-10,00 M	7,64 M	6,77 M	6,00 M	5,32 M
13	**VPN=**	15,72 M				
14	**VPN=**	15,72 M				

Fuente: elaboración propia

[7]El valor del 1 % sigue la convención sugerida en Vélez (2003, p. 242), quién además propone otras magnitudes de cambio basadas en la desviación estándar o en significado económico.

Gráfico 3.4: Fórmulas del ejemplo de análisis de sensibilidad del VPN en una hoja de cálculo

	A	B	C	D	E	F
1	Variable	Valor				
2	Futuro cambio de precio	-0,025				
3	Precio (año 1)	120				
4	Volumen de ventas	70				
5	Inversión inicial	10				
6	Tasa de descuento	0,1				
7						
8		Año				
9		0	=+B9+1	=+C9+1	=+D9+1	=+E9+1
10	Precio		=+B3	=+C10*(1+$B$2)	=+D10*(1+$B$2)	=+E10*(1+$B$2)
11	Flujo de caja	=-B5	=C10*$B$4/1000	=D10*$B$4/1000	=E10*$B$4/1000	=F10*$B$4/1000
12	Valor presente	=+VA($B$6;B9;;-B11)	=+VA($B$6;C9;;-C11)	=+VA($B$6;D9;;-D11)	=+VA($B$6;E9;;-E11)	=+VA($B$6;F9;;-F11)
13	VPN=	=+SUMA(B12:F12)				
14	VPN=	=+SUMA(VNA(B6;C11:F11);B11)				

Nota: para simplificar los resultados, decidimos expresar los flujos de caja en miles de dólares, para lo cual, desde las celdas $C11$ a $F11$ se dividió entre mil. Además en la hoja de cálculo usamos la fórmula *VNA* (celda $B14$), como una alternativa de cálculo del VPN.

Fuente: elaboración propia

Gráfico 3.5: Pantallazo de la hoja de cálculo que modela el VPN tras un cambio diminuto del precio inicial, en dólares

	A	B	C	D	E	F
1	Variable	Valor				
2	Futuro cambio de precio	-2,5%				
3	Precio (año 1)	121,2				
4	Volumen de ventas	70 m				
5	Inversión inicial	10,00 M				
6	Tasa de descuento	0,1				
7						
8		Año				
9		0	1	2	3	4
10	Precio		121,20	118,17	115,22	112,34
11	Flujo de caja	-10,00 M	8,48 M	8,27 M	8,07 M	7,86 M
12	Valor presente	-10,00 M	7,71 M	6,84 M	6,06 M	5,37 M
13	VPN=	15,98 M				
14	VPN=	15,98 M				
15	VPN (Original)=	15,72 M				
16	Cambio en el VPN (%)=	1,65%				

Fuente: elaboración propia

A manera de ejercicio, el lector puede jugar a alterar otros parámetros. Por ejemplo, subir en 0,1 puntos porcentuales, la tasa de descuento, es decir, que suba de 10 % a 10,1 %. Recalculando la variable resultado del modelo, esto provoca que el VPN reaccione cayendo un 0,35 %. Esta es la sensibilidad del VPN ante variaciones de la tasa de descuento, y los pantallazos que enseñan el procedimiento en Excel se muestran en los Gráficos 3.7, y 3.8.

Estos cambios minúsculos se repiten para todos los parámetros del modelo porque el que esté asociada con una mayor sensibilidad se considera una *variable crítica* con alta importancia para un gerente que desee minimizar la volatilidad del valor del proyecto. Se

Gráfico 3.6: Pantallazo de las fórmulas de la hoja de cálculo que modela el VPN tras un cambio diminuto del precio inicial, en dólares

	A	B	C	D	E	F
1	Variable	Valor				
2	Futuro cambio de precio	-0,025				
3	Precio (año 1)	=120*1,01				
4	Volumen de ventas	70				
5	Inversión inicial	10				
6	Tasa de descuento	0,1				
7						
8		Año				
9		0	=+B9+1	=+C9+1	=+D9+1	=+E9+1
10	Precio		=+B3	=+C10*(1+$B$2)	=+D10*(1+$B$2)	=+E10*(1+$B$2)
11	Flujo de caja	=-B5	=C10*$B$4/1000	=D10*$B$4/1000	=E10*$B$4/1000	=F10*$B$4/1000
12	Valor presente	=+VA($B$6;B9;;-B11)	=+VA($B$6;C9;;-C11)	=+VA($B$6;D9;;-D11)	=+VA($B$6;E9;;-E11)	=+VA($B$6;F9;;-F11)
13	VPN=	=+SUMA(B12:F12)				
14	VPN=	=+SUMA(VNA(B6;C11:F11);B11)				
15	VPN (Original)=	15,72				
16	Cambio en el VPN (%)=	=+B14/B15-1				

Fuente: elaboración propia

Gráfico 3.7: Planteamiento del ejemplo de análisis de sensibilidad del VPN respecto a la tasa de descuento en una hoja de cálculo

	A	B	C	D	E	F
1	Variable	Valor				
2	Futuro cambio de precio	-2,5%				
3	Precio (año 1)	120				
4	Volumen de ventas	70 m				
5	Inversión inicial	10,00 M				
6	Tasa de descuento	10,10%				
7						
8		Año				
9		0	1	2	3	4
10	Precio		120,00	117,00	114,08	111,22
11	Flujo de caja	-10,00 M	8,40 M	8,19 M	7,99 M	7,79 M
12	Valor presente	-10,00 M	7,63 M	6,76 M	5,98 M	5,30 M
13	VPN=	15,67 M				
14	VPN=	15,67 M				
15	VPN (Original)=	15,72 M				
16	Cambio en el VPN (%)=	-0,35%				

Fuente: elaboración propia

debe cambiar cada parámetro, uno a la vez, manteniendo todo lo demás constante.[8]

Cuando se tienen calculados los impactos de cada parámetro, se suelen ordenar de mayor a menor en un gráfico que los representa como barras a la derecha si son positivos, o a la izquierda si son negativos. Por la forma de embudo de esta visualización se le conoce como un *gráfico de tornado.*

Ejercicio 27 (Análisis de sensibilidad aplicado a un negocio inmobiliario).
La constructora Concreto y Mar desarrolló un proyecto residencial en la Zona Norte de Cartagena

[8]Gracias al respeto del principio de mantener lo demás constante, al análisis de sensibilidad se le considera una forma de experimentación.

Gráfico 3.8: Pantallazo de las fórmulas de la hoja de cálculo que modela el VPN tras un cambio diminuto de la tasa de descuento, en dólares

	A	B	C	D	E	F
1	**Variable**	**Valor**				
2	Futuro cambio de precio	-0,025				
3	Precio (año 1)	120				
4	Volumen de ventas	70				
5	Inversión inicial	10				
6	Tasa de descuento	=0,1+0,001				
7						
8		**Año**				
9		**0**	**=+B9+1**	**=+C9+1**	**=+D9+1**	**=+E9+1**
10	**Precio**		=+B3	=+C10*(1+$B$2)	=+D10*(1+$B$2)	=+E10*(1+$B$2)
11	**Flujo de caja**	=-B5	=C10*$B$4/1000	=D10*$B$4/1000	=E10*$B$4/1000	=F10*$B$4/1000
12	**Valor presente**	=+VA($B$6;B9;;-B11)	=+VA($B$6;C9;;-C11)	=+VA($B$6;D9;;-D11)	=+VA($B$6;E9;;-E11)	=+VA($B$6;F9;;-F11)
13	**VPN=**	=SUMA(B12:F12)				
14	**VPN=**	=SUMA(VNA(B6;C11:F11);B11)				
15	**VPN (Original)=**	=+Ejercicio_25!B13				
16	**Cambio en el VPN (%)=**	=+B14/B15-1				
17						

Fuente: elaboración propia

que costó US $ 4 M. El edificio recibirá unos inquilinos que pagarán arriendos estimados de US $ 0,5 M en el primer año y que luego crecerán a la tasa de inflación que se proyecta que será de 0,02. La tasa de descuento es de 7 %. Escriba el modelo del VPN de este proyecto en una hoja de cálculo, primero con certeza para obtener el caso base, y luego con la técnica de introducir cambios enanos a los parámetros.

3.2. Tabla de una variable

La desventaja de perturbar el modelo en tan sólo un 1 % es que no revela todo el rango de variación de las variables de entrada, porque se concentra en estandarizar un cambio en una magnitud comparable para cada una. En un análisis de sensibilidad con *Tabla de una variable*, en cambio, se buscar hallar cuáles son las variables que hacen que el proyecto deje de ser conveniente. La principal pregunta que surge de la Tabla de una variable aplicada a proyectos de inversión de capital es: ¿El VPN del proyecto se vuelve negativo en un análisis de sensibilidad?

Aunque un análisis de sensibilidad se puede ejecutar "manualmente" con la técnica de cambios diminutos, en Excel existe una forma más rápida y eficiente de implementarlo. La herramienta *Tabla de una variable* estudia una gama de valores que pueden adoptar los parámetros. Se pueden explorar varios posibles resultados al modificar la variable de entrada. La ruta para implementar esta ténica en Excel se encuentra en el menú de *Datos*, luego en *Análisis de hipótesis* y finalmente en *Tabla de datos.*

Aplicaremos la técnica de Tabla de una variable al siguiente modelo de Flujo de caja, que es la continuación del Ejercicio 26.

Ejercicio 28 (Análisis de sensibilidad para una grilla de valores del parámetro variable).

Memorias Bolívar fabrica discos duros para computadores y ha proyectado su flujo de caja teniendo en cuenta que hizo una inversión inicial para construir un edificio por valor de US $ 10 M en el año cero y tendrá ventas anuales de 70 m unidades. El horizonte temporal es de cuatro años. Si la tasa de descuento es de 7 %, halle el VPN de este proyecto cuando el parámetro precio de venta unitario se cambia con la grilla de posibles valores: 100, 110, 120, 130 y 140.

Es útil comenzar a resolver este ejercicio organizando los datos de entrada (Gráfico 3.9). Luego podemos calcular el VPN del modelo en un caso base (Gráfico 3.10), como se explicó en el Ejercicio 26.

Gráfico 3.9: Datos de entrada del ejemplo de tabla de una variable en una hoja de cálculo

	A	B
1	**Datos de Entrada**	**Valor**
2	Inversión inicial	US 10,00 M
3	Precio venta unitario - PVU	100
4	Volumen de ventas	70 m
5	Variación grilla del PVU	10
6	Tasa de descuento	0,07

Fuente: elaboración propia

Gráfico 3.10: Modelo para el cálculo del VPN en una hoja de cálculo

	A	B	C	D	E	F
8		**Año**				
9		**0**	**1**	**2**	**3**	**4**
10	**Precio**		100,00	100,00	100,00	100,00
11	**Flujo de caja**	-10,00 M	7,00 M	7,00 M	7,00 M	7,00 M
12	**Valor presente**	-10,00 M	6,54 M	6,11 M	5,71 M	5,34 M
13	**VPN=**	13,71 M				

Fuente: elaboración propia

En la primera columna de una Tabla de una variable se escribe el rango de valores posibles que toma el parámetro. Los valores de la columna “Precio de Venta Unitario (PVU)” son escritos arbitrariamente según la grilla que se desee ingresar (Celdas $A17 - A21$).

El paso más importante y difícil de recordar para insertar una Tabla de una variable es escribir (en la celda $B16$) el enlace a la celda donde se calcula el VPN (la celda $B13$ del Gráfico 3.11).

Luego, se debe seleccionar el rango de celdas completo $\Big($desde Precio Venta Unitario (celda $A16$) hasta la esquina inferior derecha (celda $B21$)$\Big)$ como se muestra en el Gráfico 3.12.[9]

[9]El valor de $ 13,71 M en la celda $B16$ no fue digitado sino que está enlazado a la celda del modelo

Gráfico 3.11: Ejemplo de grilla para el PVU en un modelo de VPN

	A	B	C	D	E	F
8		Año				
9		0	1	2	3	4
10	Precio		100,00	100,00	100,00	100,00
11	Flujo de caja	-10,00 M	7,00 M	7,00 M	7,00 M	7,00 M
12	Valor presente	-10,00 M	6,54 M	6,11 M	5,71 M	5,34 M
13	VPN=	13,71 M				
14						
15	Precio Venta	VPN				
16		=+B13				
17	100					
18	110					
19	120					
20	130					
21	140					

Fuente: elaboración propia

Gráfico 3.12: Pantallazo de la selección requerida antes de usar Tabla de una variable

	A	B	C	D	E	F
8		Año				
9		0	1	2	3	4
10	Precio		100,00	100,00	100,00	100,00
11	Flujo de caja	-10,00 M	7,00 M	7,00 M	7,00 M	7,00 M
12	Valor presente	-10,00 M	6,54 M	6,11 M	5,71 M	5,34 M
13	VPN=	13,71 M				
14						
15	Precio Venta Unitario	VPN				
16		US 13,71 M				
17	100					
18	110					
19	120					
20	130					
21	140					

Fuente: elaboración propia.

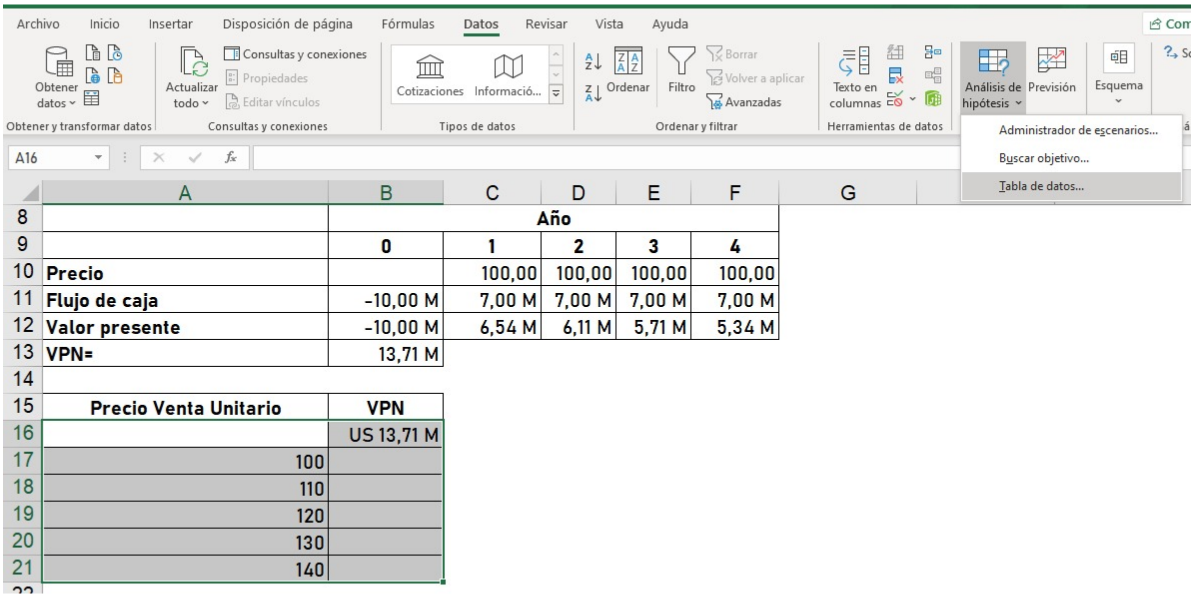
Con estas celdas seleccionadas, se debe hacer click en la pestaña *Datos*, *Análisis de hipótesis* y *Tabla de datos* (Gráfico 3.13). En el cuadro de diálogo que aparece se debe escribir el enlace a la coordenada de la celda de entrada en el campo *celda de entrada de columna*. En este caso se elige entrada de columna porque los datos de la variable de entrada están apilados como columna, si estuvieran escritos sobre una fila se debe marcar la opción de entrada de fila.

En el Gráfico 3.14 se muestra el pantallazo final de la Tabla de Datos en el que se debe introducir la celda de entrada.

Del resultado de la Tabla de una variable se generan (en nuestro ejemplo, en la columna B de Excel) los diferentes valores de VPN del proyecto de Memorias Bolívar, correspondientes

financiero en la misma hoja en la que está ubicada la variable de salida. Esto se logra escribiendo "=" en la celda que se desea enlazar y digitando la celda destino (e.g. "B13").

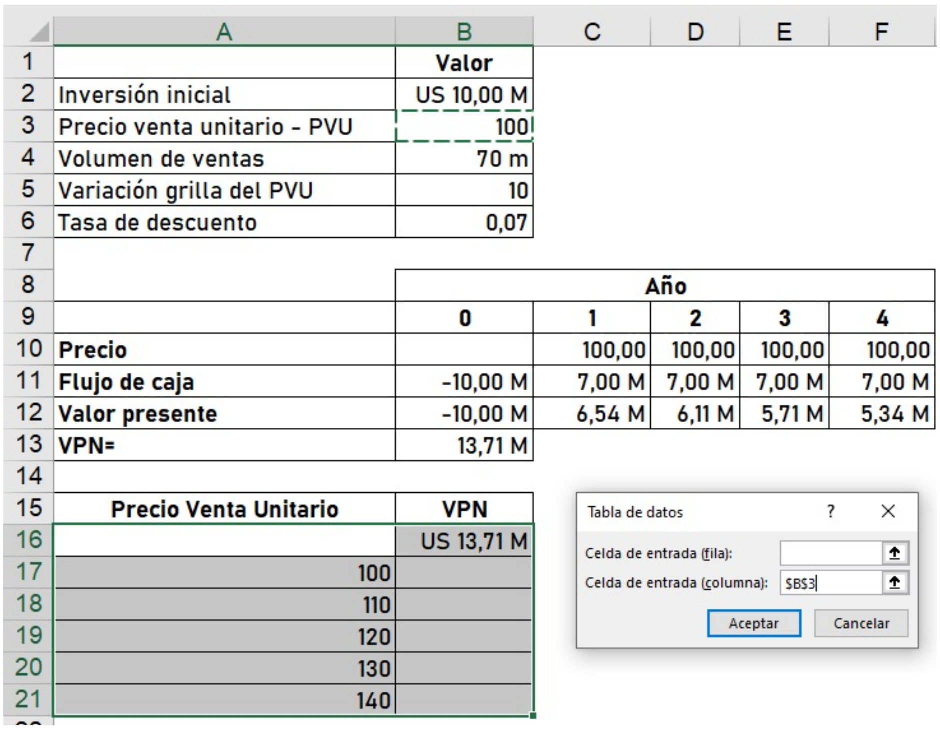
Gráfico 3.13: Pantallazo ubicando la herramienta Tabla de datos dentro del menú de la hoja de cálculo



	A	B	C	D	E	F
8		Año				
9		0	1	2	3	4
10	Precio		100,00	100,00	100,00	100,00
11	Flujo de caja	-10,00 M	7,00 M	7,00 M	7,00 M	7,00 M
12	Valor presente	-10,00 M	6,54 M	6,11 M	5,71 M	5,34 M
13	VPN=	13,71 M				
14						
15	Precio Venta Unitario	VPN				
16		US 13,71 M				
17	100					
18	110					
19	120					
20	130					
21	140					

Fuente: elaboración propia

Gráfico 3.14: Pantallazo ubicando la celda de entrada de la herramienta Tabla de datos



	A	B	C	D	E	F
1		Valor				
2	Inversión inicial	US 10,00 M				
3	Precio venta unitario - PVU	100				
4	Volumen de ventas	70 m				
5	Variación grilla del PVU	10				
6	Tasa de descuento	0,07				
7						
8		Año				
9		0	1	2	3	4
10	Precio		100,00	100,00	100,00	100,00
11	Flujo de caja	-10,00 M	7,00 M	7,00 M	7,00 M	7,00 M
12	Valor presente	-10,00 M	6,54 M	6,11 M	5,71 M	5,34 M
13	VPN=	13,71 M				
14						
15	Precio Venta Unitario	VPN				
16		US 13,71 M				
17	100					
18	110					
19	120					
20	130					
21	140					

Fuente: elaboración propia

a suponer que el PVU toma los distintos valores de la grilla. El resultado de simular una grilla de valores de la variable de entrada es que para un PVU de 100 corresponde un VPN de US$ 13,71 M, y de ahí en adelante el VPN aumenta hasta llegar a uno de US$ 23,19 M, correspondiente a un PVU de 140 (Gráfico 3.15).

Gráfico 3.15: Pantallazo del resultado de implementar Tabla de una variable

	A	B	C	D	E	F
1		**Valor**				
2	Inversión inicial	US 10,00 M				
3	Precio venta unitario - PVU	100				
4	Volumen de ventas	70 m				
5	Variación grilla del PVU	10				
6	Tasa de descuento	0,07				
7						
8		**Año**				
9		**0**	**1**	**2**	**3**	**4**
10	**Precio**		100,00	100,00	100,00	100,00
11	**Flujo de caja**	-10,00 M	7,00 M	7,00 M	7,00 M	7,00 M
12	**Valor presente**	-10,00 M	6,54 M	6,11 M	5,71 M	5,34 M
13	**VPN=**	13,71 M				
14						
15	**Precio Venta Unitario**	**VPN**				
16		US 13,71 M				
17	100	US 13,71 M				
18	110	US 16,08 M				
19	120	US 18,45 M				
20	130	US 20,82 M				
21	140	US 23,19 M				

Fuente: elaboración propia

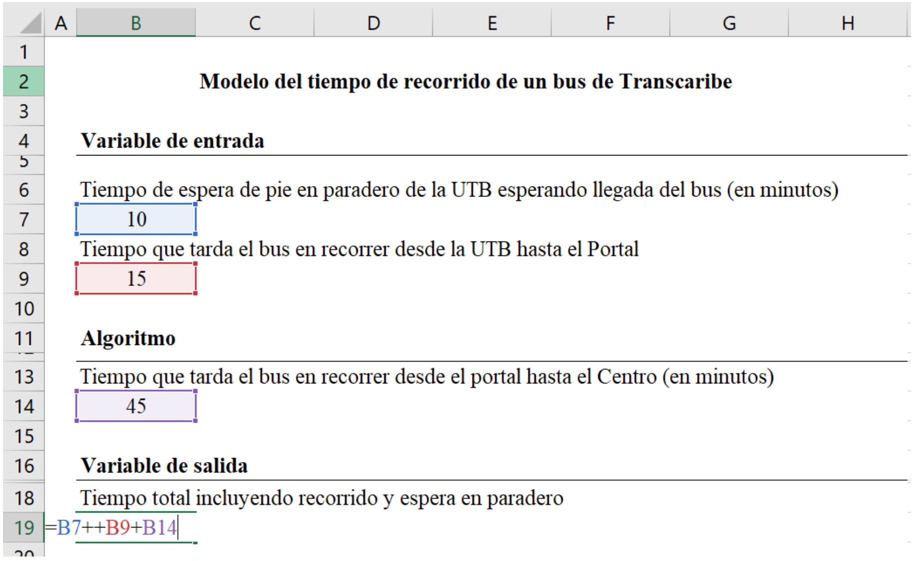
En nuestro modelo la relación entre las variables de entrada y salida fue monotónica, pero en muchos casos de modelos en los que las variables interactúan de formas complejas la relación no siempre será monotónica y existirán óptimos locales que dependerán de cómo se defina la grilla de la variable de entrada.

Ejercicio 29 (Tabla de una variable aplicado a tiempo de llegada de un Transcaribe). *Escriba en una hoja de cálculo un modelo en el que las variables de entrada sean el tiempo que se demore en llegar el Transcaribe a la estación de la UTB, y el tiempo que tarda desde allí hasta el portal. Establezca un algoritmo en forma de ecuación tal que se le sume a los anteriores tiempos un recorrido de 45 minutos dentro del bus para llegar al destino en el Centro de Cartagena. Luego, calcule como variable de salida la duración total del trayecto hasta la estación Centro.*

Por ejemplo, un valor para el caso base puede suponer unos tiempos de 10 minutos para la llegada del bus, de 15 minutos para el recorrido hasta el portal, y una duración de recorrido de 45 minutos adicionales hasta el centro, con lo que la variable de salida, el tiempo total, sería de 70 minutos (Gráfico 3.16). Use la Tabla de una variable para simular una grilla de cinco valores diferentes de la variable de entrada Tiempo de espera (que en vez de los 10 minutos del caso base, el tiempo de llegada oscile entre 0 y 20 minutos en intervalos de 2 minutos) y su impacto sobre el modelo.

Ejercicio 30 (Análisis de sensibilidad para una grilla de valores del parámetro variable). *Industria La Mesa comercializa partes de enfriadores industriales en toda América Latina. Considere los siguientes datos para la elaboración del Estado de Resultados del año 1: ventas por $ 9 m; costo de* *ventas del 60 % sobre las ventas; gastos generales por $ 2,35 m; gastos por comisiones del 2 % sobre las ventas; tasa impositiva del 30 %. Usted debe realizar un análisis de sensibilidad para una variable: porcentaje del costo de ventas. Tome los siguientes valores: 30 %, 40 %, 50 %, 60 %, 70 %, y 80 %. Debe hallar los efectos que el cambio en el porcentaje del costo de ventas conlleva sobre las siguientes variables de salida: utilidad bruta, utilidad antes de impuestos, impuesto de renta y utilidad neta.*

Gráfico 3.16: Pantallazo de la hoja de cálculo con el modelo del Ejercicio 29



Fuente: elaboración propia.

3.3. Tabla de dos variables

Excel también permite examinar cómo reacciona el VPN ante cambios introducidos a dos parámetros, simultáneamente, con la opción de Tabla de dos variables.

En este caso, la primera columna de la tabla se dedica a una grilla del primer parámetro, y la primera fila a la del segundo.

Usted puede aplicar como ejercicio esta técnica al modelo de Flujo de caja del Ejercicio 26, de tal forma que se introduzcan cambios, al mismo tiempo, al precio inicial y a la tasa de descuento.

Ejercicio 31 (Tabla de dos variables).
A partir del enunciado del Ejercicio 30, usted debe ahora aplicar la técnica de Tabla de dos variables. Para realizar el análisis de sensibilidad, deberá utilizar dos variables: el porcentaje del costo de ventas y los gastos por comisiones sobre ventas. Para el Porcentaje de costo de ventas, tome los siguientes valores: 30 %, 40 %, 50 %, 60 %, 70 %, y 80 %; y para los Gastos por comisiones sobre ventas, tome los siguientes valores: 0 %, 10 %, 20 %, 30 %, 40 % y 50 %. Debe hallar los efectos que el cambio conlleva sobre la utilidad neta.

3.4. Buscar objetivo

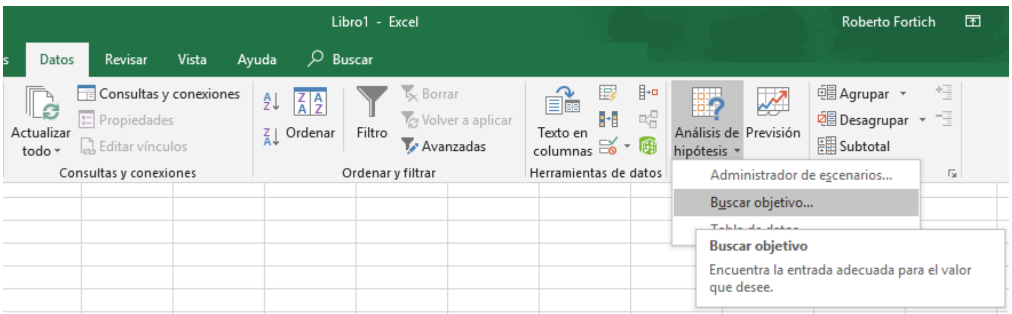
Hasta aquí hemos mostrado cómo introducir experimentos cambiando valores de los parámetros, pero con frecuencia el analista puede no estar interesado en cualquier valor, sino sólo en los concernientes a un valor predeterminado del VPN. Buscar qué valores de un parámetro hacen que la variable de resultado reaccione con un valor específico se obtiene con la opción *buscar objetivo* de Excel. Es un proceso que opera en reversa porque va de la función objetivo a la función insumo.

Esta función requiere iniciar con un valor arbitrario de la variable de entrada, que en este caso es cualquier número que se le ocurra al usuario, porque mediante un proceso iterativo el valor cambiará y se ajustará hasta llegar al valor correcto. En una operación de buscar objetivo el valor correcto de la variable de entrada es aquel que, al ingresarlo al modelo resulta en el valor deseado de la variable de resultado.

La herramienta se localiza en el menú de “Datos” al que el usuario debe hacer click desde la celda en la que está la variable de salida. La ruta para llegar a “Buscar objetivo” se muestra en el Gráfico 3.17.

Las dos celdas que le debemos proporcionar al cuadro de diálogo de la hoja de cálculo, son: *con el valor*, que debe ser igual al valor que queremos que adopte la variable de salida, y *cambiando la celda*, que debe contener la coordenada de la celda de la variable de entrada cuyo valor queremos encontrar.

Gráfico 3.17: Pantallazo de la ubicación de buscar objetivo en el menú de Excel



Ejercicio 32 (Buscar objetivo).
Para el modelo del Ejercicio 26 añada un costo total anual de US $ 2,5 M. Luego halle el valor del precio inicial que hace que el VPN sea igual a cero, usando la función Buscar objetivo de Excel.

Con la adición del costo total como nuevo parámetro, las variables de entrada del ejercicio son como se indican en el Cuadro 3.4.

A partir de esta información, los flujos de caja se plantean en Excel junto con su proyección a futuro y el cálculo del VPN del proyecto, como se muestra en el Gráfico 3.18.

Cuadro 3.4: Variables típicas de un Estados de Resultados para el caso base de Memorias Bolívar, en dólares

Variable	Valor
Futuro cambio de precio	-2,5 % anual
Precio (año 1)	120
Volumen de ventas (año 1)	70 m
Costo total anual	2,5 M
Inversión inicial	10 M
Tasa de descuento	0,1

Gráfico 3.18: Modelo para el cálculo del VPN con *buscar objetivo* en una hoja de cálculo

Año	0	1	2	3	4
		$83.62	$81.53	$79.49	$77.51
		70,000	70,000	70,000	70,000
		$2,500,000	$2,500,000	$2,500,000	$2,500,000
inv0	($10,000,000.00)				
ingresos		$5,853,616.37	$5,707,275.97	$5,564,594.07	$5,425,479.21
vp ingresos		$5,321,469.43	$4,716,757.00	$4,180,761.88	$3,705,675.31
vp costos		$2,272,727.27	$2,066,115.70	$1,878,287.00	$1,707,533.64
vpn		$0.00			
tasa descuento	10%				

Fuente: elaboración propia

Finalizado este planteamiento, el lector debe verificar con la herramienta de *buscar objetivo* que el valor del precio inicial que hace que el VPN reaccione volviéndose cero, es igual a $83,62$.

3.5. Apéndice

Pasos para usar la Tabla de una variable:

1. Tener un modelo de inversión de capital en Excel
2. Escribir cerca del anterior modelo una grilla de posibles valores del parámetro como se indica en el Gráfico 3.12
3. Escribir el enlace a la celda de VPN en la parte que dice $B13$ del Gráfico 3.12
4. Seleccionar las celdas como se indica en el Gráfico 3.13
5. Buscar por menús de Excel la herramienta de *Tabla de datos*
6. En el menú emergente, proporcionar, en el campo *celda de entrada de columna* el enlace

a la coordenada de la celda con el valor del parámetro a experimentar

7. Con los anteriores pasos la Tabla de una variable autocompletará la columna B con los valores de la variable de salida buscados, como se muestra en el Gráfico 3.15

Capítulo 4

Simulaciones de Monte Carlo en R

Roberto Fortich Mesa y Leonardo Castellanos Acuña

En ciencias de la computación existe un principio poderoso que afirma que podemos conocer el comportamiento de una variable aleatoria de la vida real estudiando las propiedades de sus datos simulados, siempre y cuando usemos muchas repeticiones. La idea es buscar imitar a la variable aleatoria creando un proceso generador de datos seudo aleatorios que preserve su esencia estocástica. Así como los diseñadores de aviones prueban prototipos en túneles de viento, el equivalente en el mundo de la estadística son las simulaciones de Monte Carlo.

Definición 28 (Simulación de Monte Carlo).
Es la generación artificial de datos que son extraídos en muestras aleatorias de una función de distribución de probabilidad y que luego se simulan y se combinan para generar una o varias variables de resultado

En relación con los temas del Capítulo 3, puede decirse que una simulación de Monte Carlo es una versión extendida del análisis de sensibilidad, en la que en lugar de cambios pequeños o grillas de valores equidistantes, se le asigna a la variable de entrada un valor aleatorio. Pero este valor no es puramente aleatorio, sino que sigue una función de distribución de probabilidad con parámetros previamente calibrados.

Los elementos comunes a las simulaciones de Monte Carlo incluyen:

1. Generar en un computador números seudo aleatorios

2. Escribir comandos que repitan una secuencia de instrucciones

3. Almacenar y procesar los resultados de la simulación para evaluar sus propiedades

Las simulaciones de Monte Carlo se justifican porque la incertidumbre de las variables puede interactuar de formas imprevisibles: así como se pueden contrarrestar sus efectos, también pueden multiplicarse, lo que aumenta exponencialmente la exposición al riesgo de la empresa.

Uno de los primeros trabajos en advertir que las empresas que no simulen sus variables son muy vulnerables a la incertidumbre fue Hertz (1964), quien citó el caso de una empresa de alimentos hipotética que debía decidir si le convenía lanzar un producto al mercado. Hertz supuso que solo cinco variables aleatorias conformaban las variables de entrada del modelo del éxito del nuevo producto, y conjeturó que era una cantidad suficiente para que, al componerlas y calcular el riesgo agregado, dañase todo intento de pronóstico preciso de su TIR. Específicamente, habría apenas un 8 % de probabilidad de que se diesen de forma conjunta los valores medios esperados para cada una de las cinco variables, pese a que individualmente tenían una mayor pronosticabilidad.

Este estimativo se obtuvo usando valores medios de las variables aleatorias, asignando conservadoramente un 60 % de probabilidad de acierto y un 40 % de yerro a la estimación individual. Hertz (1964, p. 170) concluyó que:

> "This simple example illustrates that the rate of return actually depends on a specific combination of values of a great many different variables"

El incorporar incertidumbre con métodos de Monte Carlo paradójicamente no apunta a reducir la variabilidad del estimador puntual del VPN porque, como lo señaló Hertz (1964), de poco sirve un estimador de rango estrecho que se sabe que no se cumplirá casi nunca. Por el contrario, las simulaciones buscan ampliar el abanico de posibles valores del VPN y acompañarlas de una mayor plausibilidad de que el resultado final caiga en el rango pronosticado. La principal virtud de este análisis es que, de caer en un rango positivo, implica que los inversionistas pueden tener tranquilidad de que el proyecto no es riesgoso.

El nombre de "Monte Carlo" se refiere al espíritu prágmatico y acientífico de esta técnica.[1] Es un método que no presume querer conocer la forma funcional exacta del algoritmo complejo que gobierna la variable de salida de un modelo. Ese esfuerzo es abandonado a priori y, en cambio, es reemplazado por el procedimiento práctico de generar un muy alto número de muestreos y examinar las proporciones relativas de los posibles resultados al final de la caja negra.

[1] Monte Carlo es una región del principado de Mónaco que es célebre porque allí está ubicado uno de los casinos más tradicionales y lujosos del mundo. En Metropolis (1987) se rastrea el origen del nombre del método de Monte Carlo.

4.1. Introducción a R

Las simulaciones de riesgo de modelos financieros son un campo de las finanzas que congrega a dos grandes disciplinas: los profesionales de las finanzas y los estadísticos. Como se mencionó en el Capítulo 3 los profesionales de las finanzas se desenvuelven en la industria principalmente usando hojas de cálculo y, cuando la sofisticación del problema lo requiere, complementos de Excel y softwares patentados.

En cuanto a los profesionales en estadística aplicada, la tendencia en años recientes ha llevado a la popularización del uso de R y Python (*Left for dead, R surges again* 2020). El software R es un ambiente para computador que hace cálculos estadísticos y gráficos y que es ideal para computar simulaciones de Monte Carlo. Python tiene similares capacidades aunque es más un lenguaje de programación general. En este capítulo explicaremos las aplicaciones usando R.

R fue desarrollado en 1993 en la Universidad de Auckland (Nueva Zelanda) con la figura de software libre. Está basado en el lenguaje de programación S creado en los Laboratorios Bell en los 1970s.

R posee dos principales fortalezas. En primer lugar, posee amplia variedad de paquetes de complemento que han sido aportados por su comunidad de usuarios. Es uno de los softwares estadísticos que más se pueden personalizar, dado que el código fuente es abierto a ser mejorado por cualquier usuario, y a que otorga muchas opciones de versatilidad y flexibilización (*New GUIs aim to make R open source analytics language more accessible* 2010).

Segundo, R permite crear visualizaciones convincentes, útiles y atractivas con uno de sus paquetes de complemento estrella, "ggplot2". Este paquete ha sido tan exitoso que ha sido descargado más de diez millones de veces (*All hail ggplot2—The code powering all those excellent charts is 10 years old* 2017).

Una desventaja de R que le ha restado adeptos es su mayor complejidad respecto a las hojas de cálculo y otros softwares estadísticos[2] porque tiene una curva de aprendizaje moderadamente empinada.

En R los programas se escriben en archivos de secuencias de comandos (*scripts*). Las secciones de una secuencia de comandos se encabezan mediante títulos antecedidos del símbolo #. El símbolo # sirve para hacer comentarios en R.

Los archivos de secuencias de comandos de R quedan almacenados con la extensión $.r$ y pueden ejecutarse completa o parcialmente. Algunos programadores usan editores de texto especializados para editar los archivos $.r$ que contienen los scripts de un programa de R.

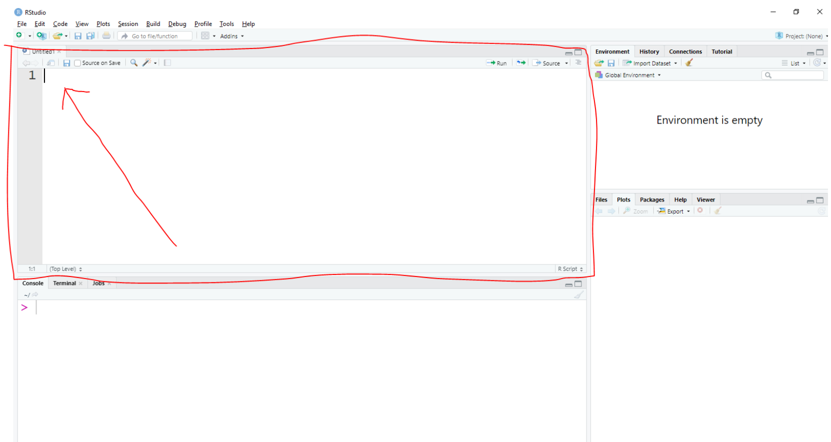
[2]SAS, SPSS y Stata son los principales competidores de R en la industria de la consultoría estadística.

Explicaremos todas las instrucciones de aquí en adelante usando la interfaz RStudio, que es un popular Entorno de Desarrollo Integrado, creado en 2009, que asiste al usuario de formas didácticas, tales como resaltar con colores ciertos tipos de comandos o autocompletar instrucciones.

4.2. Interfaz de RStudio

La interfaz de RStudio se compone de cuatro paneles y una barra superior. Los comandos deben trabajarse en el panel de Editor o archivo de secuencia de comando que aparece en la esquina superior izquierda de RStudio (Gráfico 4.1).

Gráfico 4.1: Pantallazo que muestra el panel del editor de RStudio



Fuente: elaboración propia

Los otros paneles que componen la interfaz son:

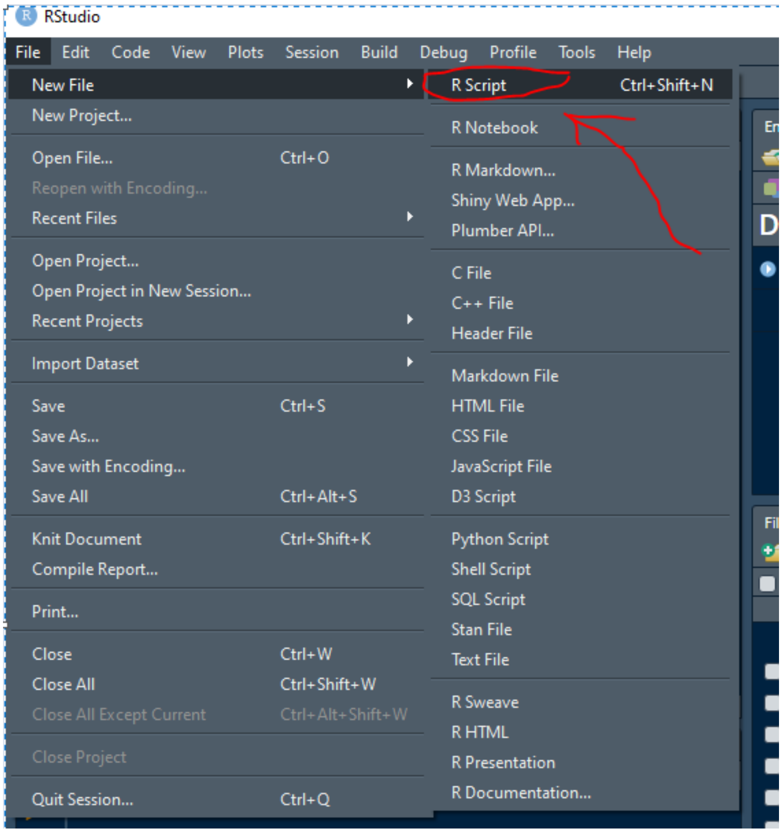
- Entorno
- Consola
- Visor misceláneo

4.3. Introducción al entorno de trabajo

Para crear un archivo se debe hacer click en *File: New Script* en la barra de comandos de RStudio (Gráfico 4.2).

Para ejecutar un comando o un conjunto de líneas de comando, se debe hacer click en *Run line*, que aparece en la barra de herramientas de RStudio.

Gráfico 4.2: Pantallazo de creación de un nuevo archivo en RStudio



Fuente: elaboración propia

4.4. Aprendizaje de R

Las principales competencias que un estudiante de R debe aprender de primero, siguiendo a Dobrow (2016), son:

1. R como una calculadora
2. Atajos de teclado en R
3. Buscar información de ayuda específica de un comando o su sintaxis
4. Creación de vectores e indexación de sus elementos

Las habilidades intermedias que sugerimos que un estudiante de R debe dominar, incluyen: ciclos for, funciones, indexaciones anidadas de elementos de vectores, gestión de bases de datos con los paquetes *dplyr* y *tidyr*, comprensión de la gramática que soporta a los verbos del paquete *ggplot2*, la exportación del código en archivos PDF o HTML por medio de *markdown*, y el control de versiones en *Github*.[3]

El primer uso que se debe aprender de R consiste en usarlo como calculadora, es decir,

[3]Esta lista no pretende ser exhaustiva y puede requerir añadir nuevas técnicas dado que probablemente aparecerán paquetes creados por usuarios con potencial de volverse indispensables.

escribir en la consola de R operaciones matemáticas y oprimir *Enter* para solucionarlas.

En segundo lugar, conviene conocer los principales atajos de teclado en R. Por ejemplo, la tecla de flecha arriba permite recuperar el último comando ejecutado, lo que ayuda enormemente a ahorrar tiempo porque a menudo uno se halla escribiendo unos comandos que son leves modificaciones de otros escritos anteriormente en la sesión. Con la combinación "CTRL+A" se selecciona todo, con "CTRL+L" se limpia la consola y con "CTRL+ENTER" se ejecuta un script.

Tercero, dado que el aprendizaje de R es un proceso continuo y de empinada curva de aprendizaje, es útil familiarizarse con las distintas formas de desatascar un script que no corre. Una característica esencial es la ayuda, que consiste en anteponer el signo ? al comando. Por ejemplo: `?mean`, que llama a la ayuda del comando *mean*.

Sin embargo, a nuestro parecer estos archivos de ayuda son un poco confusos para un usuario que no sea programador profesional. Por ese motivo recomendamos además el aprender a buscar respuestas en motores especializados.[4] También ayuda que cada usuario novato recopile sus propios códigos de ejemplo, escritos con cuidado de que sean de fácil navegación, y que contengan explicaciones detalladas de las distintas partes del programa.

Finalmente, para la creación de vectores[5] es fundamental desarrollar destrezas para crear listas consecutivas de números, combinar elementos, usar el operador de asignación ($< -$), e indexar elementos usando corchetes.

Ejercicio 33 (Creación de vectores).
Calcule en R los cuadrados de los primeros nueve enteros. La respuesta es:

```
(1:9)^2
```

Ejercicio 34 (Creación de vectores).
Genere una variable mundial que contenga los años en que se han realizado los últimos tres mundiales de fútbol.

4.5. Generación de números seudo aleatorios

En el corazón de las simulaciones de Monte Carlo yacen los números aleatorios. Se requiere una rutina de computador que los genere de forma apropiada. A pesar de que una máquina obviamente no tiene capacidad de libre albedrío, los expertos en ciencias de computación

[4] Uno de los más populares es el foro *Stack Overflow*.

[5] En programación un vector es una lista de números, que equivalen a tener una variable en una columna de excel, con los elementos del vector igual a las celdas de la columna y el nombre del vector igual al nombre de la variable.

han ideado formas en que pueden programar un computador para que genere números que parezca que hubiesen sido generados aleatoriamente.

Los números puramente aleatorios son aquellos que tienen igual probabilidad de aparecer, sin tener en cuenta el número de veces que hayan aparecido antes. Los números aleatorios, en cambio, no necesariamente tienen igual probabilidad, sino que son elegidos por muestreo de una función de distribución de probabilidad.

Una forma de entender la diferencia entre las dos nociones, es que los números puramente aleatorios salen de muestrear un número aleatoriamente de entre una distribución uniforme, mientras que los números aleatorios salen de aplicar una uniforme a un conjunto de ensayos de un Proceso Generador de Datos (PGD) que simula una función de distribución de probabilidad acumulada de alguna de las familias expuestas en el Capítulo 1.

El generador de números aleatorios de los software estadísticos en realidad no es aleatorio, sino que aprovecha la propiedad matemática del operador módulo que crea listas desordenadas de números de un dígito. Esta lista de números es determinística, finita y se repite periódicamente. Sin embargo, para efectos prácticos sirven para representar el papel de un número aleatorio, y han superado las procedimientos que los estadísticos han preparado para evaluar el cumplimiento de unos mínimos de aleatoriedad (i.e. que sean equiprobables e idéntica e independientemente distribuidos).

El valor inicial del ciclo de esta lista se llama *semilla*. El valor por defecto de la semilla está asociado con el reloj interno del computador, pero puede ser ajustado por el usuario. Dado que con frecuencia se busca reproducir los resultados de una simulación, es útil casi siempre incluir un valor de semilla antes de generar cualquier valor aleatorio.

El comando *sample* genera una muestra aleatoria de cierto tamaño a partir de los elementos de un vector. Equivale a aplicar una distribución uniforme a un rango de números enteros que siempre comienza en uno, con la conveniencia de que permite que el muestreo sea con o sin reemplazo. Un ejemplo es:

```r
sample(1:10,1)
```

La sintaxis de las instruciones dentro del paréntesis, indica que los argumentos se deben especificar en el siguiente orden: *vector,tamaño,opción de reemplazo*. Para simular, por ejemplo, diez lanzamientos de un dado, con reemplazo, se escribe:

```r
sample(1:6,10,replace=TRUE)
```

Ejercicio 35 (Simulación de números aleatorios).
Simule con sample el lanzamiento de un dado virtual

4.6. Simulación de distribuciones de probabilidad

Las funciones de generación de números aleatorios que incorpora R sirven para evaluar distribuciones de probabilidad en cuatro distintas formas:

1. La densidad
2. La función de distribución acumulada, $P(X \leq x)$
3. La función cuantil (dado q, el x tal que $P(X \leq x) = q$)
4. Simular valores, x, de la distribución

Los comandos de simulación en R se componen de dos partes (prefijo+raiz). La primera corresponde a un prefijo que se refiere a las anteriores formas, así:

1. *d* = densidad
2. *p* = función de distribución de probabilidad acumulada, $P(X \leq x)$
3. *q* = función cuantil (dado q, el x tal que $P(X \leq x) = q$)
4. *r* = valores simulados, x, de la distribución

Las raíces son la segunda parte de un comando de simulación en R, y deben seguirse a la siguiente equivalencia:

- *binom* = Binomial
- *pois* = Poisson
- *norm* = Normal
- *unif* = Uniforme
- *triangle* = Triangular
- *chisq* = Chi cuadrado

4.6.1. Bernoulli

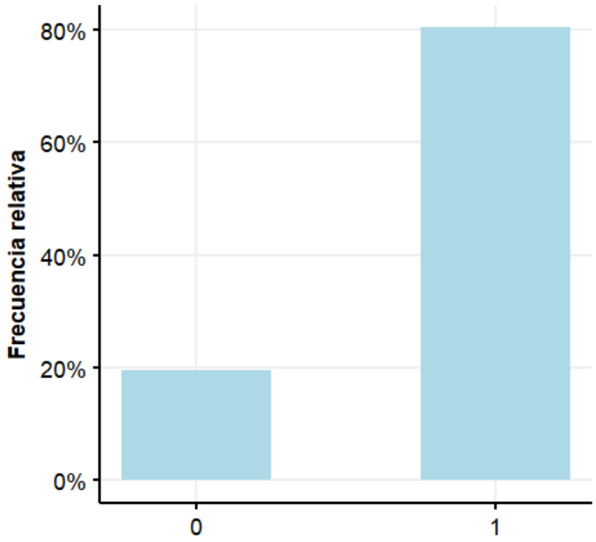
La distribución Bernoulli debe ser descargada del paquete *Rlab* porque no viene predeterminada con R base.[6]

[6] A la versión de R que aún no se le ha añadido ningún paquete de complemento se le llama R base. La instalación de un paquete del internet requiere los comandos *install.packages()* y *library()*.

Alternativamente, se pueden simular valores de una distribución de Bernoulli si se usa su equivalencia con una Binomial en la que el número de ensayos es igual a uno (size=1).[7]

Un ejemplo del gráfico de una distribución de Bernoulli con p igual a 0,8 es el del Gráfico 4.3.

Gráfico 4.3: Números aleatorios generados de acuerdo a una distribución de Bernoulli



Fuente: elaboración propia.

4.6.2. Binomial

Entonces, la sintaxis para lograr una Binomial en R es:

```
rbinom(n=m, size=q, prob=p)
```

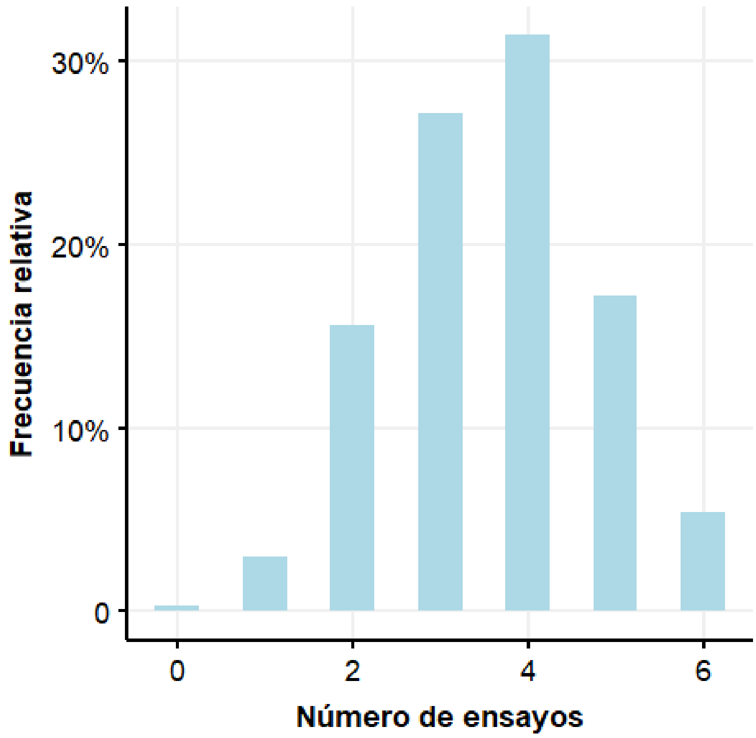
Donde m es el argumento de cuántos números deseamos generar, q es el número de ensayos que se repetirán, y p es la probabilidad de éxito en cada ensayo. Cuando $q = 1$ se están en el caso especial de una Bernoulli.

En el Gráfico 4.4 se representa un ejemplo de valores simulados que se distribuyen como una Binomial con parámetro p de 0.6 y seis ensayos.

Ejercicio 36 (Simular valores de una Binomial).
Ejecute 1.000 simulaciones de una Binomial con 6 ensayos y probabilidad de éxito $p = 0,6$

[7]La explicación de la teoría de probabilidad de la Bernoulli y su relación con otras distribuciones se mostró en el Capítulo 1.

Gráfico 4.4: Números aleatorios generados de acuerdo a una distribución Binomial



Fuente: elaboración propia

```
rbinom(n=1000,size=6,prob=0.6)
```

Ejercicio 37 (Densidad de una Binomial).
Una empresa compra transistores en lotes. ¿Cuál es la probabilidad de que se examinen cuatro transistores consecutivamente y salgan dos éxitos exactamente? Cada ensayo se distribuye como una Bernoulli con un éxito ocurriendo con probabilidad de $20\,\%$ *y un fracaso con probabilidad de* $80\,\%$. *Para este ejemplo, éxito es igual a que el transistor salga dañado.*

Como este ejercicio nos pide una probabilidad, usaremos la forma *densidad* de las simulaciones de R, que requiere el prefijo *d*, y usaremos como párametros un x igual a 2, un *size* de 4, y un *prob* de 0,2.

```
dbinom(x=2,size=4,prob=0.2)
```

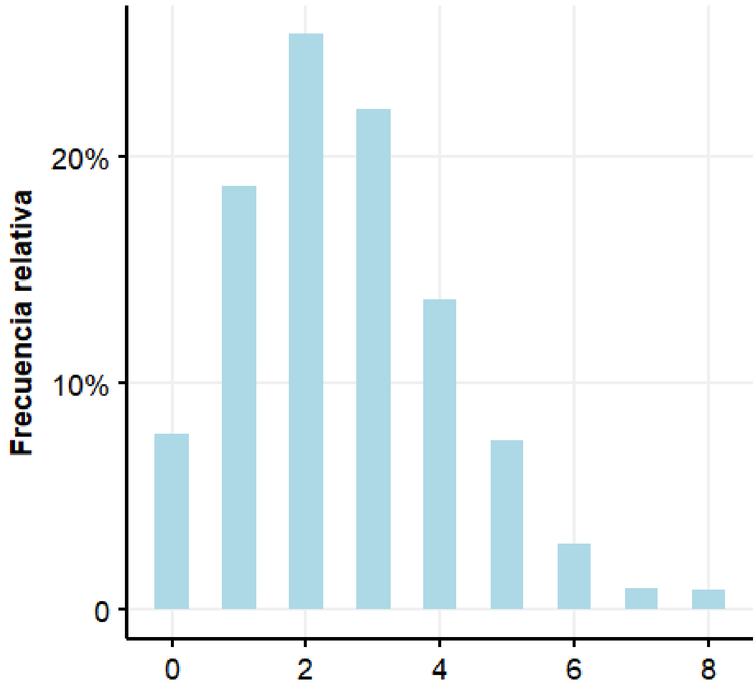
```
0.1536
```

Existe una probabilidad de 15,36 % de que en cuatro exámenes a transistores se obtengan exactamente 2 éxitos, es decir, 2 transistores defectuosos.

4.6.3. Poisson

En el gráfico 4.5 se muestra el gráfico de la distribución de probabilidad de algunos valores simulados que se comportan como una función Poisson con parámetro λ igual a 2.6.

Gráfico 4.5: Números aleatorios generados de acuerdo a una distribución de Poisson



Fuente: elaboración propia

Ejercicio 38 (Simular valores de una Poisson).
Ejecute 1.000 simulaciones de una Poisson con parámetro $\lambda = 0,2$

```
rpois(n=1000,lambda = 0.2)
```

4.6.4. Uniforme

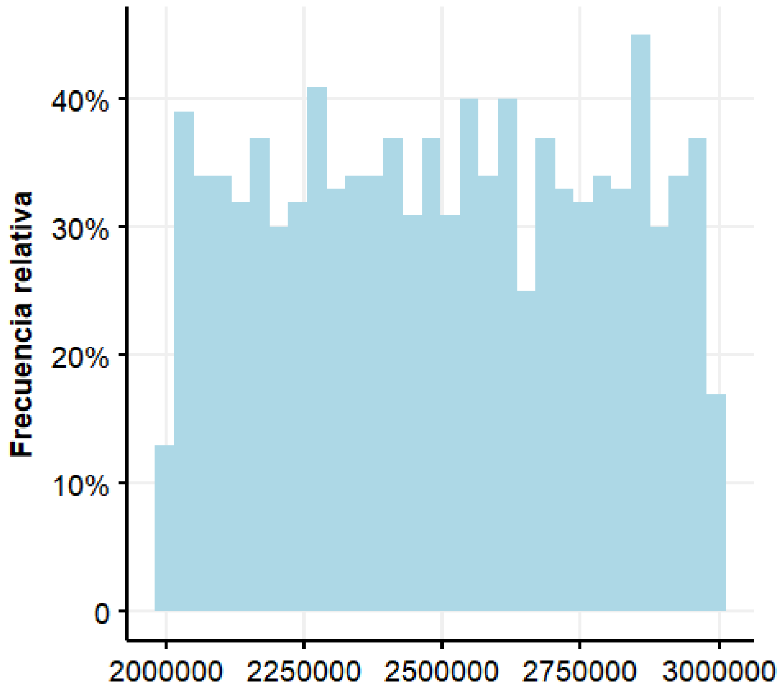
En el Gráfico 4.6 se muestra la representación de una simulación de valores que se distribuyen como una Uniforme.

La sintaxis para generar valores simulados de una uniforme en R es:

```
runif(m,min=alpha,max=beta)
```

Donde m es el argumento de cuántos números deseamos generar, α es el límite inferior del intervalo, y β el superior

Gráfico 4.6: Números aleatorios generados de acuerdo a una distribución Uniforme



Fuente: elaboración propia

Ejercicio 39 (Simular valores de una Uniforme).
Ejecute 1.000 simulaciones del valor de los costos fijos anuales de una empresa suponiendo que se distribuyen como una Uniforme entre US$2 y US$3 M

```
cfix <- runif(n=1000,min=2e+06,max=3e+06)
```

Ejercicio 40 (Densidad de una Uniforme).
Si la vida de un proyecto se distribuye como una Uniforme entre 4 y 6 años, calcule la densidad asociada a exactamente 5 años

```
proy_life <- dunif(x=5,min=4,max=6)
```

```
0.5
```

La densidad de todos los valores de una uniforme entre 4 y 6 es igual a 0,5. Esta densidad se multiplica por el intervalo que se desee examinar para obtener la probabilidad de ocurrencia. En funciones de distribución de probabilidad continuas, la probabilidad de un valor puntual es igual a 0, de modo que la densidad sólo cobra sentido si se multiplica por algún intervalo. Por ejemplo, podemos aproximar la ocurrencia de una vida del proyecto de 5 años como el intervalo entre 4,9 y 5,1 años, que resulta de multiplicar la densidad por el intervalo, $(0,5)(0,2) = 0,1$.

Este mismo resultado lo pudimos obtener con la función de distribución de probabilidad acumulada en R, así:

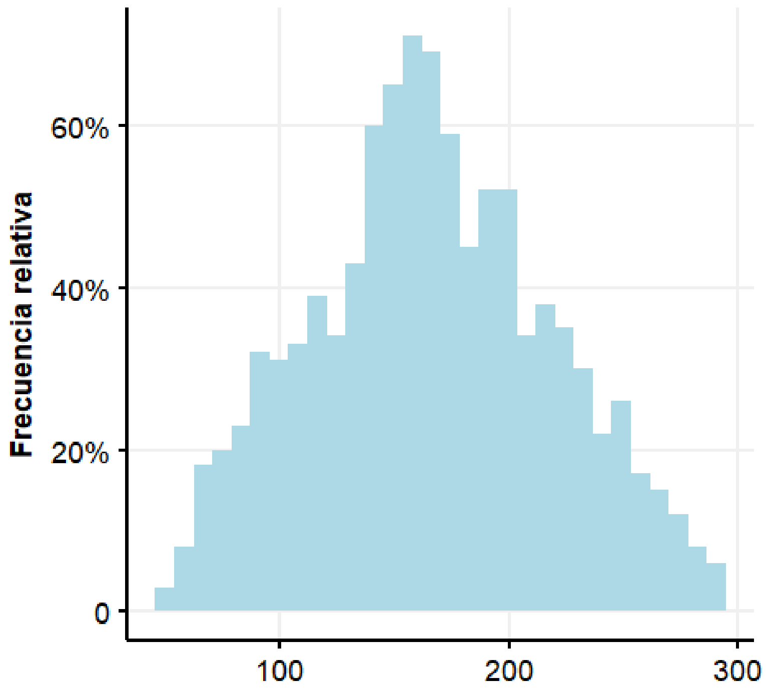
```
proy_life <- punif(q=5.1,min=4,max=6)-punif(q=4.9,min=4,max=6)
proy_life
```

```
0.1
```

4.6.5. Triangular

En el Gráfico 4.7 se muestran valores simulados extraidos de una distribución Triangular.

Gráfico 4.7: Números aleatorios generados de acuerdo a una distribución Triangular



Fuente: elaboración propia

La distribución triangular debe obtenerse del paquete *triangle*.

Ejercicio 41 (Simular valores de una triangular).
Un analista con experiencia puede ayudar a estimar el número de clientes que efectivamente llegarán a comprar el producto. Al responder la pregunta de cuáles son los números más bajos y más altos de clientes que el cree que puede haber en una semana, se obtuvieron valores de 40 y 300, respectivamente.

Halle los valores de a, b y c, y ejecute 1.000 simulaciones del número de clientes por semana.

Los valores solicitados siguen la notación del Capítulo 1 para una Triangular, en la que a es el límite inferior, que en este caso es igual a 40; b es el límite superior, que aquí es de 300; y c es la moda. En este ejemplo, el enunciado no informa cuál es la moda y, por tanto, lo reemplazaremos con el punto medio.[8]

Ejercicio 42 (Función de distribución acumulada de una Triangular).
Para el Ejercicio 41 halle la probabilidad de que el número de clientes sea menor a 100.

```
library(triangle)
ptriangle(q=100,a=40,b=300,c=170)
```

```
0.1065089
```

Existe una probabilidad de 10,65 % de que el número de clientes en una semana sea menor a 100.

4.6.6. Normal

En el Gráfico 4.8 se muestra un ejemplo de una función de distribución para un grupo de valores simulados que siguen una Normal.

Ejercicio 43 (Simular valores de una normal).
Ejecute 1.000 simulaciones de una normal con parámetros $\mu = 3$ *y* $\sigma^2 = 9$

```
rnorm(n=1000,mean=3,sd=3)
```

Ejercicio 44 (Función de distribución de probabilidad acumulada de una Normal).
Para el Ejercicio 43, halle la probabilidad de que la variable tome un valor menor o igual a 5

```
pnorm(q=5,mean=3,sd=3)
```

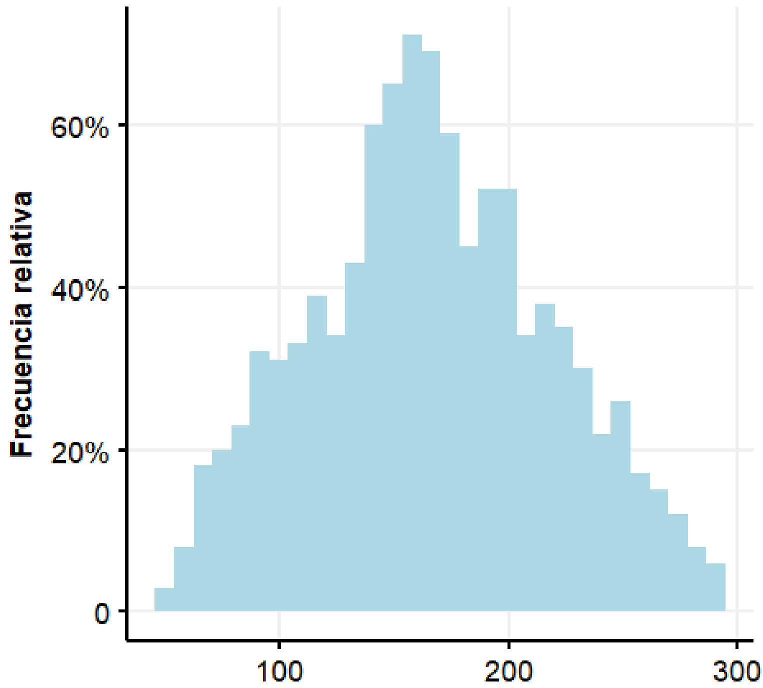
```
0.7475075
```

Existe un 74,75 % de probabilidad de que la variable tenga un valor menor a 5.

Ejercicio 45 (Probabilidad entre dos valores).
Demuestre usando R que la probabilidad de que una variable (distribuida Normal estándar) tenga un valor entre -1 y +1 desviaciones estándar, es igual al 68,3 %

[8] Sin embargo, una Triangular no necesariamente tiene que tener una moda igual al punto medio, porque puede ser un triángulo asimétrico.

Gráfico 4.8: Números aleatorios generados de acuerdo a una distribución normal



Fuente: elaboración propia

Ejercicio 46 (Cuantil asociado a una probabilidad).
Si la estatura de un grupo de estudiantes se distribuye como una Normal con media de 169 cms y desviación estándar de 9 cms, halle el cuantil 95, es decir, la estatura a partir de la cual se debería ubicar el 5 % más alto de la muestra

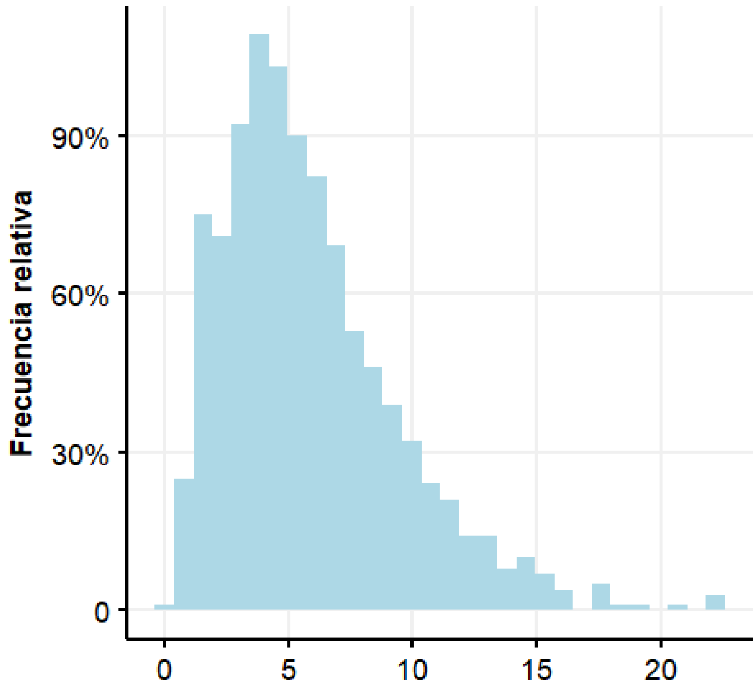
4.6.7. Chi cuadrado

El Gráfico 4.9 muestra un conjunto de valores simulados que se distribuyen de acuerdo con una Chi cuadrado.

Ejercicio 47 (Simular valores de una Chi cuadrado).
Ejecute 1.000 simulaciones de una Chi cuadrado con 6 grados de libertad

```
rchisq(n=1000,df=6)
```

Gráfico 4.9: Números aleatorios generados de acuerdo a una distribución Chi cuadrado



Fuente: elaboración propia

4.7. Configuración de una simulación

Una simulación será útil en la medida en que se acerque a la realidad en cuanto a los supuestos sobre sus variables de entrada, parámetros de distribución de probabilidad, algoritmos. Estas condiciones son necesarias pero no suficientes para una simulación exitosa porque además se requiere que haya un alto número de repeticiones del Monte Carlo.

Partiendo de una elección apropiada de la distribución de probabilidad, es decir, que se use la que mejor se amolde al comportamiento natural de la variable, es necesario suponer un valor para sus parámetros estadísticos. Los parámetros de las distribuciones que se programen en R se suelen obtener a partir de los datos usando el método de máxima verosimilitud. En el caso de la Normal, este método equivale a usar los momentos muestrales (i.e. media y desviación estándar muestral) como su contraparte poblacional.

4.8. Aplicación a modelos financieros

El objetivo principal de hacer simulaciones de Monte Carlo para modelos de inversión de capital es conocer la probabilidad de que el VPN sea menor a cero, lo que mide de foma aproximada qué tan riesgoso sería decidir invertir en el proyecto.

Para Hertz 1964 un Monte Carlo requiere recopilar información de parámetros en las siguientes categorías:

1. Análisis de mercado
2. Análisis de costo de la inversión
3. Costos variables y fijos

La parametrización de los parámetros seleccionados de las anteriores categorías (i.e. tamaño de mercado, vida útil de los activos) se puede calibrar a partir de entrevistas con expertos.

Ejercicio 48 (Monte Carlo aplicado).
Memorias Bolívar es una empresa que ha desarrollado la tecnología para fabricar un accesorio para computadores (e.g. discos duros) y su gerente financiero desea examinar sus perspectivas de rentabilidad de mediano y largo plazo. La empresa hizo una inversión inicial de contruir un edificio que costó US $ 10 M en el año cero. El precio inicial se distribuye como una Triangular con límite inferior de 100, superior de 140 y moda de 120. La tendencia de cambio anual del precio se distribuye como una uniforme entre -10 % y +5 %. El volumen de ventas se distribuye como una triangular con mínimo 50 mil, máximo 80 mil y moda de 70 mil. Los costos fijos serán de US$2,5 M cada año. Los costos variables serán de US$32 por unidad. Halle el VPN del proyecto estimando las tres variables aleatorias con simulaciones de Monte Carlo dada una tasa de descuento del 10 % y un horizonte de proyección de seis años.

El preámbulo del script en R es el siguiente:

```
# Modelo financiero

# Preambulo
library(triangle)
set.seed(1)
```

La primera línea contiene un título para el programa. Luego se carga el paquete de complemento de las distribuciones triangulares y se fija un valor semilla. El valor de 1 en este caso fue arbitrario.

```
# Variables de entrada

t <- 6 # numero de periodos
r <- 0.1 # tasa de descuento
```

```
inv0 <- 10e+06 # Inversion inicial
nsimul <- 2000 # Numero de simulaciones
```

Esta sección contiene el valor de los parámetros, es decir, las variables de entrada que se conocen con certeza y no requieren ser simuladas. También se específica una variable con el número de simulaciones, asignando un valor alto como es estándar en todo Monte Carlo.

```
# Inicializacion de vectores
vpn <- vector() # Valor presente neto
p <- vector() # precio
volumen <- vector() # volumen de ventas
vp <- vector() # valor presente de los flujos de caja
```

A continuación en el script va una sección de inicialización de vectores. Aquí se agrupan los comandos que crean vectores vacíos para las variables que serán creadas por el algoritmo del modelo financiero. En programación los vectores suelen inicializarse porque esto ahorra tiempo de cómputo dado que se separa la cantidad de memoria justa para el tamaño requerido.

```
# Algoritmos

for (i in 1:nsimul) { # Ciclo que calcula vpn

# Proyectar ingresos y costos

volumen[1] <- rtriangle(n=1,a=50e+03,b=80e+03,c=70e+03) # incertidumbre
del volumen

for (j in 2:t) {
  volumen[j] = volumen[j-1]
}

p[1] <- rtriangle(n=1,a=100,b=140,c=120) # incertidumbre del precio

for (j in 2:6) {
  p[j] = p[j-1]*(1+runif(n=1,min=-0.1,max=0.05)) # incertidumbre del
#cambio en el precio
}
```

```
# Hallar las utilidades
ing <- volumen*p # Ingresos operacionales

costos <- 2.5e+06+32*volumen

utilidad <- ing - costos # Ingreso operacional

# Traer al presente las utilidades

vp0 <- -inv0

for (j in 1:t) {
  vp[j] <- utilidad[j]/((1+r)^j)
}

# Hallar el VPN
vpn[i] <- vp0 + sum(vp)
```

Esta sección contiene el algoritmo en el que se transforman las variables de entrada del modelo, que en este caso se hace proyectando en el tiempo a partir de tasas de crecimiento, calculando los flujos de caja, y descontándolos a valor presente. Nótese, por un lado, el uso de un *ciclo for* para la proyección del precio a seis años; en programación es muy usual el uso de estos ciclos para escribir con eficiencia códigos de un modelo. Por otro lado, en esta sección se encuentran las simulaciones de cada variable con incertidumbre, que generan un número aleatorio, y que se repiten con cada *ciclo for* externo según el valor asignado a *nsimul*.

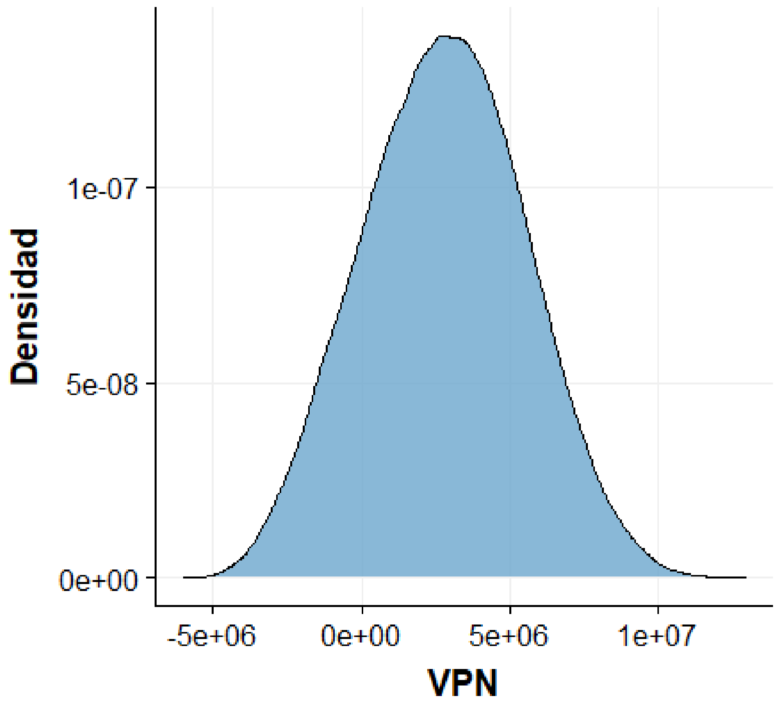
```
# Variable de salida
hist(vpn)
```

Finalmente, el script contiene el código que calcula la visualización de la variable de salida, en un histograma, el cual se puede embellecer con algo de trabajo adicional para que luzca como en el Gráfico 4.10.

Con el resultado de la simulación se puede calcular la probabilidad de que el proyecto tenga un VPN positivo, porque el resultado se distribuye Normal con una media y desviación estándar que se puede calcular con la variable de resultado que nos botó el Monte Carlo.

Ejercicio 49 (Mejora de tecnología con incertidumbre).
La empresa Jaguar Blanco está considerando comprar una nueva máquina que reemplace la actual.

Gráfico 4.10: VPN tras la simulación de Monte Carlo



Fuente: elaboración propia.

Los costos de mantenimiento de la nueva máquina se distribuyen Normal con media de US$ 30 m anuales y una desviación estándar igual a US$ 1 m multiplicados por la edad de la máquina. Por ejemplo, tres años adentro, la desviación estándar será de US$ 3 m. Su precio es de US$ 190 m y tiene una vida útil de 5 años sin valor de salvamento. La máquina vieja ha estado en servicio durante tres años y le quedan cuatro más sin valor de salvamento, con un costo de mantenimiento anual que se distribuye como una Uniforme entre US$ 70 m $\pm$ US$ 1,5 m $\times$ la edad de la máquina vieja. Finalmente, la tasa de descuento para cada año se distribuye como una Triangular con un mínimo de 11 %, una moda de 15 %, y un máximo de 19 %. Jaguar Blanco desea elegir la máquina con mayor VPN, ¿es la nueva o la vieja?

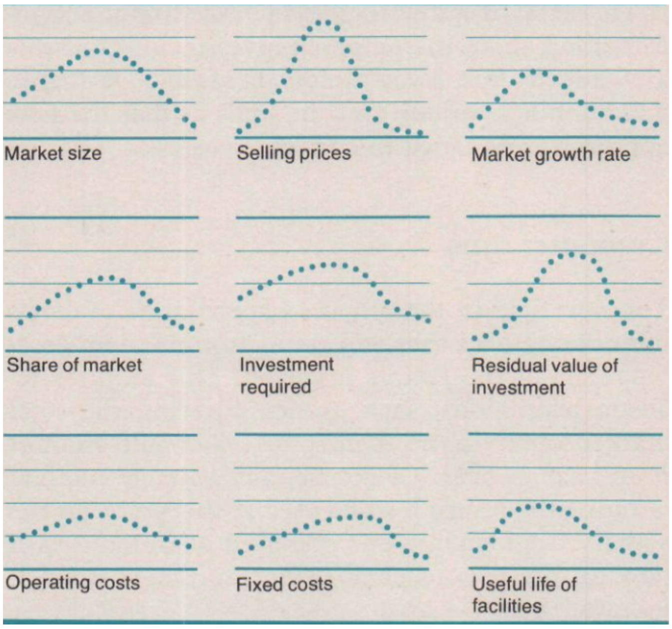
Ejercicio 50 (Ejercicio clásico de Hertz 1964).
Una empresa de químicos está evaluando la conveniencia de ampliar su planta de procesamiento inviertiendo US$ 10 M. El tiempo de vida útil de la instalación se estima que será de 10 años. Los ingenieros esperan usar 250 m toneladas de material procesado a un costo de procesamiento medio de $ 435 por tonelada. Los principales parámetros que la gerencia he decidido usar son el tamaño del mercado, los precios de venta, la tasa de crecimiento del mercado, la participación de mercado (que resulta en un volumen físico de ventas), la inversión requerida, el valor residual de la inversión, los costos operativos, los costos fijos, y la vida útil da las instalaciones. Los parámetros se calibran así:

- *Tamaño de mercado: media = 250 m, rango = (100 m, 340 m)*

- *Precio de venta: media = 510, rango = (385, 575)*
- *Tasa de crecimiento del mercado: media = 3 %, rango = (0 %, 6 %)*
- *Tasa de participación de mercado eventual: media = 12 %, rango = (3 %, 17 %)*
- *Inversión total requerida: media = 9,5 M, rango = (7 M, 10,5 M)*
- *Vida útil de las instalaciones: media = 10, rango = (5, 15)*
- *Valor terminal (a los 10 años): media = 4,5 M, rango = (3,5 M, 5 M)*
- *Costos variables por unidad: media = 435, rango = (370, 545)*
- *Costos fijos: media = 300 m, rango = (250 m, 375 m)*

La asimetría de la distribución de probabilidad que se asigna a cada variable se estima como en el Gráfico 4.11.

Gráfico 4.11: Forma esperada de las simulaciones de variables de entrada del modelo



Nota: De "Risk analysis in capital investment", por D. Hertz, 1964, *Harvard Business Review, 42(1).* Copyright 1964 Harvard Business Review. Autorización pendiente.

Estime el VPN del proyecto incorporando la incertidumbre en las variables anteriores, primero con 3.600 simulaciones (como en el artículo original) y luego con 100 mil.

Referencias

Left for dead, R surges again (julio de 2020). URL: https://www.datanami.com/2020/07/10/left-for-dead-r-surges-again/.

El futuro de la información (septiembre de 2018). URL: https://www.eluniversal.com.co/opinion/columna/el-futuro-de-la-informacion-15168-GUEU405399.

Google apps vs. Office 365: Which suite reigns supreme? (Mayo de 2016). URL: https://www.informationweek.com/cloud/google-apps-vs-office-365-which-suite-reigns-supreme/d/d-id/1325658.

All hail ggplot2—The code powering all those excellent charts is 10 years old (junio de 2017). URL: https://qz.com/1007328/all-hail-ggplot2-the-code-powering-all-those-excellent-charts-is-10-years-old/.

New GUIs aim to make R open source analytics language more accessible (septiembre de 2010). URL: https://searchbusinessanalytics.techtarget.com/news/2240022538/New-GUIs-aim-to-make-R-open-source-analytics-language-more-accessible.

Bernstein, Peter L. (1998). *Agains the gods: The remarkable story of risk.* Hoboken, NJ: Wiley.

Castillo, Mario (2006). *Toma de decisiones en las empresas: entre el arte y la técnica.* Bogotá, Colombia: Ediciones Uniandes.

Dobrow, Robert (2016). *Introduction to stochastic processes with R.* Hoboken, NJ: Wiley.

Hald, Anders (1988). *A history of probability and statistics and their applications before 1750.* Hoboken, NJ: Wiley-Interscience.

Hertz, David (1964). «Risk analysis in capital investment». En: *Harvard Business Review* 42.1, pp. 96-106.

Hillier, Frederick y Gerald Lieberman (2001). *Introduction to operations research.* Nueva York, NJ: McGraw-Hill.

Johnson, David (1997). «The triangular distribution as a proxy for the beta distribution in risk analysis». En: *The Statistician* 46.3, pp. 387-398.

LeBlanc, Larry y Thomas Grossman (2008). «The use of spreadsheet software in the application of management science and operations research». En: *Interfaces* 38.4, pp. 225-227.

Liao, Shu (1990). «Spreadsheet-based simulation modeling for risk analysis». En: *Journal of Financial Education* 19, pp. 49-58.

Magee, John (1964). «Decision trees for decision making». En: *Harvard Business Review* July, pp. 35-48.

Mastrandrea, Michael D. y col. (2010). *Guidance note for lead authors of the IPCC Fifth Assesment Report on consistent treatment of uncertainties.*

Metropolis, Nicholas (1987). «The beginning of the monte carlo method». En: *Los Alamos Science Special Issue* 15, pp. 125-130.

Raychaudhuri, Samik (2008). «Introduction to monte carlo simulation». En: *2008 Winter simulation conference.* IEEE, pp. 91-100.

Ross, Sheldon (2010). *A first course in probability.* Londres: Pearson.

Saltelli, Andrea (2002). «Sensitivity analysis for important assessment». En: *Risk Analysis* 22.3, pp. 579-590.

Seila, Andrew y Jerry Banks (1990). «Spreadsheet risk analysis using simulation». En: *Simulation* 55.3, pp. 163-170.

Vélez, Ignacio (2003). *Decisiones empresariales bajo riesgo e incertidumbre.* Bogotá, Colombia: Norma.

Webster, Allen (2000). *Estadistica aplicada a los negocios y la economia.* Nueva York: McGraw-Hill.